Statistical Modelling of Home Range and Larvae Movement Data

Christopher Richard McLellan

Doctor of Philosophy
University of Edinburgh
2013

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Christopher Richard McLellan*)

Acknowledgements

I would like to thank my main supervisor, Dr. Bruce Worton, for his advice and support over the last four years. I am also grateful to my secondary supervisor Dr. Natalia Bochkina, and to William Deasy and Professor A. Nicholas E. Birch at The James Hutton Institute for providing data to analyse. Thanks must go to the EPSRC for providing the studentship that supported my research.

I wish to thank Fiona Manson, Holly Curless and Jacqueline Brown for their assistance, and to thank my parents for always supporting me.

Abstract

In this thesis, we investigate two different approaches to animal movement modelling; finite mixture models, and diffusion processes. These models are considered in two different contexts, firstly for analysis of data obtained in home range studies, and then, on a much smaller scale, modelling the movements of larvae. We consider the application of mixture models to home range movement data, and compare their performance with kernel density estimators commonly used for this purpose. Mixtures of bivariate normal distributions and bivariate t distributions are considered, and the latter are found to be good models for simulated and real movement data. The mixtures of bivariate t distributions are shown to provide a robust parametric approach. Subsequently, we investigate several measures of overlap for assessing site fidelity in home range data.

Diffusion processes for home range data are considered to model the tracks of animals. In particular, we apply models based on a bivariate Ornstein-Uhlenbeck process to recorded coyote movements. We then study modelling in a different application area involving tracks. Diffusion models for the movements of larvae are used to investigate their behaviour when exposed to chemical compounds in a scientific study. We find that the fitted models represent the movements of the larvae well, and correctly distinguish between the behaviour of larvae exposed to attractant and repellent compounds. Mixtures of diffusion processes and Hidden Markov models provide more flexible alternatives to single diffusion processes, and are found to improve upon them considerably. A Hidden Markov model with 4 states is determined to be optimal, with states accounting for directed movement, localized movement and stationary observations. Models incorporating higher-order dependence are investigated, but are found to be less effective than the use of multiple states for modelling the larvae movements.

Contents

Abstract **4**

1 Introduction to movement studies **9**

1.1 Introduction 9
1.2 Use of GPS telemetry for animal movement studies 10
1.3 Overview of animal movement modelling 11
1.4 Cabbage root fly larvae bioassays 13
1.5 Scope of thesis 16

2 Analysis of home range data **18**

2.1 Introduction 18
2.2 Review of kernel and mixture models 19
2.2.1 Nonparametric kernel methods 19
2.2.2 Parametric mixture modelling 22
2.2.3 Mixtures of bivariate normal distributions 23
2.2.4 Mixtures of bivariate t distributions 24
2.3 Simulation study 29
2.4 Model assessment using integrated squared error 34
2.4.1 Calculation for kernel and mixture models 34
2.4.2 Model comparison 39
2.4.3 Calculation for non-circular distributions 44
2.4.4 Model comparison for non-circular distributions 47
2.5 Application of kernel and mixture methods 51
2.5.1 Fixed kernel method 52
2.5.2 Adaptive kernel method 54
2.5.3 Mixture of bivariate normal components 56
2.5.4 Mixture of bivariate t components 59
2.5.5 Discussion 63
2.6 Bayesian analysis using mixture models 64
2.6.1 Prior distributions for the parameters of mixture models . 64

2.6.2 Application to mule deer data . . . 68
2.7 Discussion . . . 73

3 Fidelity **75**
3.1 Introduction . . . 75
3.2 Distance measures . . . 76
3.2.1 Defining distance measures . . . 76
3.2.2 Calculation of distance measures from density estimates . . . 79
3.3 Simulation study . . . 82
3.3.1 Overview of study . . . 82
3.3.2 Identical bivariate normal distributions . . . 83
3.3.3 Bivariate normal distributions with coordinate shift . . . 87
3.3.4 Mixture of t distributions I . . . 91
3.3.5 Mixture of t distributions II . . . 96
3.4 Expectation and variance of the product measure . . . 101
3.5 Bias approximation for product measure . . . 110
3.6 Application to mule deer movement data . . . 118
3.7 Discussion . . . 127

4 Diffusion modelling of home range data **129**
4.1 Introduction . . . 129
4.2 Coyote movement data . . . 130
4.3 The bivariate Ornstein-Uhlenbeck process . . . 132
4.4 Application to coyote movement data . . . 133
4.5 A mixture of two BOU components . . . 137
4.6 Discussion . . . 140

5 Diffusion modelling of bioassay data **141**
5.1 Introduction . . . 141
5.1.1 Background . . . 141
5.1.2 Data collection . . . 141
5.2 Diffusion modelling of larvae tracks . . . 145
5.2.1 A simple diffusion process . . . 145
5.2.2 Likelihood approach . . . 146
5.2.3 Model comparison using BIC . . . 150
5.2.4 Bayesian approach . . . 150
5.2.5 Model comparison and discussion . . . 161
5.3 Simulation and diagnostics . . . 162
5.3.1 Simulation study . . . 162

5.3.2 Diagnostic plots using Mahalanobis distance 163
5.3.3 Simulation plots 165
5.3.4 Discretization error 171
5.4 Mixture model 174
5.4.1 Mixture of two diffusion processes 174
5.4.2 Application to bioassay data 175
5.4.3 Model comparison between mixture and simple diffusion process 175
5.4.4 Simulation plots 179
5.5 Sensitivity analysis 185
5.6 Discussion 190

6 Further modelling of bioassay data 193
6.1 Introduction 193
6.2 Hidden Markov models for larva movements 194
6.3 Application to bioassay data 195
6.3.1 2-state HMM 196
6.3.2 3-state HMM 198
6.3.3 4-state HMM 201
6.3.4 Model comparison 203
6.4 Higher order dependence 210
6.4.1 Autocorrelation in bioassay data 210
6.4.2 Higher order diffusion processes 214
6.4.3 Mixture model with higher order dependence 218
6.5 Discussion 230

7 Conclusions 232
7.1 Review 232
7.2 Mixture models 233
7.3 Diffusion models 234
7.4 Further work 237

A R code for mixture of bivariate t distributions 241

B BUGS code for diffusion process 248

C BUGS code for HMM with three states 249

D Summary statistics for 4-state HMM 252

References **258**

Chapter 1

Introduction to movement studies

1.1 Introduction

In this thesis, models for animal movement are investigated in several different contexts. Home range studies of animal movement are a subject of continued interest, and a variety of models have been used to analyse animal location data (Worton, 1987; Laver and Kelly, 2008; Cagnacci et al., 2010; Kie et al., 2010). Kernel density estimators are routinely used for such data, and provide a nonparametric method which is very flexible but gives only a description of the observed data without attempting to produce an estimate of an underlying density function. Finite mixture models are an alternative parametric method, and our first topic is an investigation of mixture models for animal home range data and comparison of their effectiveness with commonly used kernel methods. The aim is to construct mixtures that provide useful and robust models of animal movements. We also investigate the use of several measures of overlap for estimating home range fidelity.

A different approach to movement modelling is also considered in the form of diffusion processes. Diffusion modelling accounts for the correlation frequently present in movement data and models the tracks of animals, whereas the mixtures considered in the earlier chapters of the thesis provide an estimate of an animal's overall use of space. We initially apply diffusion models to home range data. However, they are primarily investigated in a different context, modelling the movements of cabbage root fly larvae exposed to chemical compounds in a series of bioassays. Our aim is to devise a diffusion model that successfully characterizes the movements of the larvae and identifies attraction to or repulsion from the

chemical compounds used. We investigate mixtures and Hidden Markov models of multiple diffusion processes and the possibility of higher-order dependence to determine the most optimal approach for the data.

1.2 Use of GPS telemetry for animal movement studies

The study of animal movements is a field of ongoing interest, and the July 2010 issue of the *Philosophical Transactions of the Royal Society B* was dedicated entirely to the subject of GPS technology for animal movement modelling. Studies are primarily focused on identifying patterns that represent movement in a deterministic or mechanistic way (Nathan et al., 2008; Smouse et al., 2010). A standard method of collecting movement data is to attach GPS or Argos tracking devices to animals (see Figure 1.1) to record location data and transmit it to a satellite. The advent of widespread GPS use for recording animal movements has allowed for the collection of large quantities of movement data, as the performance of GPS systems greatly surpasses that of conventional very high-frequency (VHF) radio-tracking systems in terms of accuracy and repeatability (Coelho et al., 2007; Soutullo et al., 2007; Tomkiewicz et al., 2010; Urbano et al., 2010). VHF methods involve collecting telemetry locations in a biased fashion, as data collection is restricted by the ability of observers to collect it (Hebblewhite and Haydon, 2010). GPS telemetry provides relatively unbiased data compared to VHF, and also greatly increases the precision of location observations. Furthermore, methods to correct for GPS bias are commonly implemented, whereas bias correction has been largely ignored for VHF methods (Frair et al., 2010). The resulting availability of relatively unbiased data on habitat use by animals has improved the ability of researchers to assess relationships between habitat and performance (Gaillard et al., 2010), and has allowed for models of an animal's movement process to be dependent on the habitat available at a particular time rather than merely considering the resources available within the entire area of study (Beyer et al., 2010). GPS telemetry has also enabled improved estimation of kill rates by carnivores due to the greater quantity of data available (Merrill et al., 2010). This technology has also provided valuable results on the subject of animal migration by combining movement data with remotely sensed resource availability data (Hebblewhite, 2009) to generate evidence for migration on scales that would be impractical without the use of GPS technology (Hebblewhite and Haydon, 2010). Many animal populations traverse areas that are much larger than previously

shown by conventional studies of their movement, and GPS location data has given researchers more information on the ecology of wide-ranging species such as fin whales (Littaye et al., 2004), leatherback sea turtles (James et al., 2005) and wandering albatross (Fritz et al., 2003; Weimerskirch et al., 2007).

There are certain drawbacks to GPS technology. The increased equipment cost over VHF systems frequently necessitates smaller sample sizes, resulting in poorer population-level inference (Rodgers, 2001; Cagnacci et al., 2010). Furthermore, GPS data are often recorded at short time intervals, frequently at a rate of one observation per second, resulting in highly correlated data sets (Boyce et al., 2010; Fieberg et al., 2010) which require appropriate statistical modelling, and in some cases the fine-scale knowledge gained is of limited relevance to overall patterns of movement. However, GPS telemetry is nonetheless a more effective method of data collection than VHF systems.

Figure 1.1: An American oystercatcher (*Haematopus palliatus*) with an attached GPS tracking device. This image has been reproduced from `http://www.migratoryconnectivityproject.org/satellite-telemetry/`.

1.3 Overview of animal movement modelling

Animal movement modelling has classically taken two distinct approaches, Lagrangian and Eulerian (Smouse et al., 2010). The Lagrangian approach involves the use of stochastic processes such as Markovian diffusion processes to model the tracks of individual animals, while the Eulerian approach models the overall expected use of space by an individual or population (Okubo, 1980) and frequently includes estimation of home range.

An animal's home range is the area of space it uses while searching for food and mates, and raising young (Burt, 1943). Early attempts at home range analysis used the method of minimum convex polygons (Blair, 1940; Odum and Kuenzler,

1955), which attempted to describe the area without assuming a probability density generating the animal's movemements. However, it is more convenient to consider an idealized probabilistic model of the way the animal uses its home range, under which the animal's position has a bivariate distribution over the plane for a specified time period. This distribution is known as the utilization distribution, and its probability density is the utilization density (Van Winkle, 1975).

In wildlife studies of the home range behaviour of animals, large quantities of high quality data can be collected using radio telemetry (White and Garrott, 1990). Home range studies are frequently used to describe an animal's use of space, and the utilization density can be estimated in various ways using independent observations of the animal's position (Worton, 1987). Kernel density estimators have become the standard method for estimating the utilization density (Worton, 1995a; Kernohan et al., 2001; Laver and Kelly, 2008; Kie et al., 2010), and provide a nonparametric density estimate dependent only on the data. A kernel density estimator for a bivariate data set $\{\mathbf{x}_1 = (x_{1,1}, x_{1,2})^T, \ldots, \mathbf{x}_n = (x_{n,1}, x_{n,2})^T\}$ is of the form

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x_i}}{h}\right),$$

where K is a unimodal symmetric bivariate probability density function, and the bandwidth h determines the level of smoothing, with a larger value of h resulting in a more heavily smoothed density estimate. The properties of kernel density estimators are further discussed in Chapter 2, with consideration also given to the use of finite mixture models for home range data.

Wildlife telemetry data are frequently correlated. There are many different causes of such correlation, including constraints imposed by an animal's maximum speed, behavioural processes and extrinsic forces such as temperature or prey and resource availability (Fieberg et al., 2010). GPS telemetry data in particular is usually comprised of observations recorded at short intervals, resulting in strong autocorrelation. The autocorrelation can be accounted for either by attempting to remove it by subsampling the original data to give a new data set with observations separated by much longer intervals, or by incorporating it into the statistical models used to analyse the data. The classic Lagrangian approach for highly correlated data uses stochastic differential equations to describe the movement steps of an animal at time t, with coordinates $\mathbf{r}(t) = (x, y, t)$ (Brillinger

et al., 2004; Preisler et al., 2004; Smouse et al., 2010), of the form

$$\begin{Bmatrix} \mathrm{d}x(t) \\ \mathrm{d}y(t) \end{Bmatrix} = \begin{Bmatrix} \mu_x(\mathbf{r}(t), t) \\ \mu_y(\mathbf{r}(t), t) \end{Bmatrix} \mathrm{d}t + \mathbf{D}\{\mathbf{r}(t), t\} \begin{Bmatrix} \mathrm{d}\psi_x(t) \\ \mathrm{d}\psi_y(t) \end{Bmatrix},$$

where $\mathrm{d}x(t)$ and $\mathrm{d}y(t)$ are incremental step sizes along the x-axis and y-axis respectively, $\boldsymbol{\mu} = (\mu_x, \mu_y)^T$ is the vector of drift parameters, $\mathbf{D}$ is a diffusion matrix, and ψ_x and ψ_y are random processes with expectations of zero. The general model includes several special cases. When the drift is zero and the diffusion terms along the x-axis and y-axis are independent, the model is an uncorrelated random walk, whereas if the diffusion terms are not independent then it is a correlated random walk. If the drift terms are non-zero, the model is a biased random walk which favours movement in a particular direction. In this thesis, attention is primarily given to the special case of attraction towards a particular point rather than in a direction, in which the model is a bivariate Ornstein-Uhlenbeck (BOU) process, and the conditional distribution of position $\mathbf{x}_{s+t}$ at time $s+t$ given previous position $\mathbf{x}_s$ at time s is as follows,

$$\begin{aligned} \mathbf{x}_{s+t}|\mathbf{x}_s &\sim \mathrm{N}\{\mu + e^{\mathbf{B}t}(\mathbf{x}_s - \boldsymbol{\mu}), \boldsymbol{\Lambda} - e^{\mathbf{B}t}\boldsymbol{\Lambda}e^{\mathbf{B}^T t}\}, \quad t > 0, \\ &= \mathrm{N}\{\boldsymbol{\Gamma}\mathbf{x}_s + (\mathbf{I} - \boldsymbol{\Gamma})\boldsymbol{\mu}, \boldsymbol{\Phi}\}, \end{aligned} \tag{1.1}$$

where $\boldsymbol{\mu}$ is a point of attraction, $\boldsymbol{\Phi}$ is the covariance matrix of movement over the time interval t, and $\boldsymbol{\Gamma} = e^{\mathbf{B}t}$ determines the strength of attraction towards $\boldsymbol{\mu}$. The BOU process is applied to home range data in Chapter 4. In Chapters 5 and 6, we explore the application of diffusion processes to movements on a smaller scale, namely the movements of cabbage root fly larvae in a series of bioassays.

1.4 Cabbage root fly larvae bioassays

The larva of the cabbage root fly (*Delia radicum* L.), illustrated in Figure 1.2, is a serious pest that causes damage to *Brassica* plants by feeding on their roots (Figure 1.3). The first generation of flies each year appears from April to early June, emerging from pupae which have overwintered in the soil. The females lay 1mm eggs in close proximity to the stems of suitable host plants, and the larvae must then move through the soil to make their way into the stems and roots of the hosts and begin feeding. The larvae are fully grown after a period of about three weeks, and leave the plant at this point. They then move a short distance into the soil and become small brown pupae, as shown in Figure 1.4. After a

pupation period of around one week, adult flies emerge. Three generations of fly normally hatch in each summer, with the pupae of the eggs laid by the third generation remaining in the soil until the next year. Due to the insect's rapid life cycle, illustrated in Figure 1.5, it is a problem for host plants throughout the spring and summer.

Figure 1.2: Larvae of the cabbage root fly. Image is reproduced from `http://www.allotment-garden.org.uk/vegetable/cabbage-root-fly.php`, and is a modification of an image uploaded to `nl.wikipedia.org`.

The use of organophosphate insecticide for controlling the larvae is restricted to a single pre-planting application, and novel alternative treatments are currently being investigated. One of the potential treatments is to develop a way of manipulating the behaviour of the larvae in their attempts to find suitable hosts. Studies suggest that the larvae respond to the odour of *Brassica* plants, and use the chemicals excreted by the roots of these plants to locate suitable hosts (Städler, 1978; Koštál, 1992; Roessingh et al., 1992; Baur et al., 1998; van Leur et al., 2008; Deasy, 2011). The larvae have also been shown to be repelled by sufficiently high concentrations of plant-specific chemicals (Ross and Anderson, 1992; Ewan, 2011). If such chemicals can be identified, and their effects on the behaviour of the larvae understood, it will be possible to develop a control system using the appropriate plant extracts as soil amendments to act as deterrents.

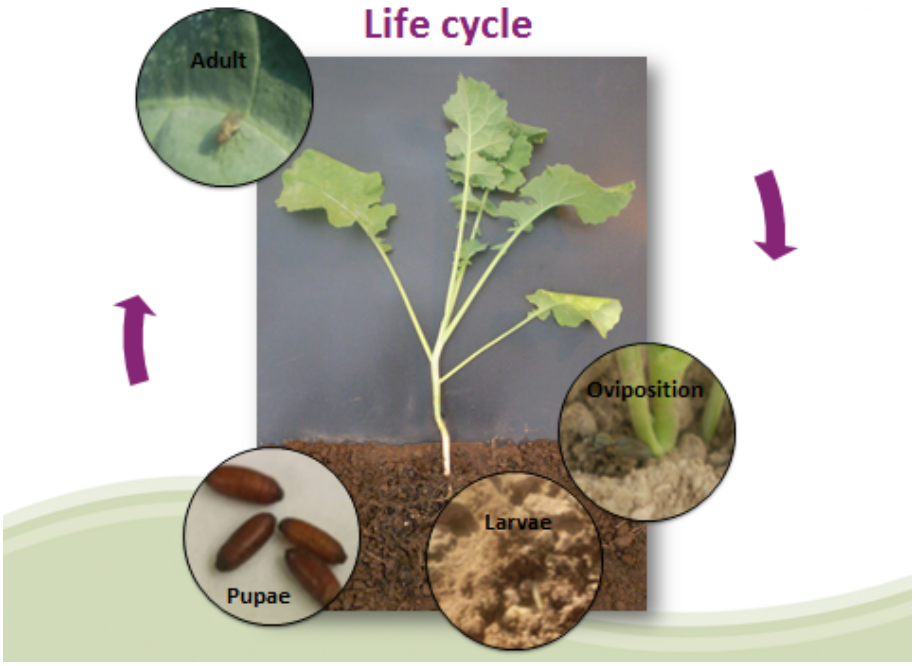
In Chapters 5 and 6 of the thesis, we will use diffusion processes to describe the

Figure 1.3: Damage caused to cabbage roots by the cabbage root fly. Image reproduced from `http://www7.inra.fr/hyppz/IMAGES/7031592.jpg`.

Figure 1.4: Pupae of the cabbage root fly (*Delia radicum* L.). Image reproduced from `http://www7.inra.fr/hyppz/IMAGES/7031593.jpg`.

movements of cabbage root fly larvae exposed to chemical compounds in a series of bioassays. We initially propose the use of a single diffusion process similar to a BOU process, and then consider mixtures of multiple diffusion processes and Hidden Markov models as well as the possibility of including higher-order dependence in our models.

Figure 1.5: A figure illustrating the life cycle of the cabbage root fly. The fly lays eggs near to a suitable host plant, and the larvae then move onto the host and begin feeding. After around three weeks, the larvae are fully grown and leave the host to pupate in the soil. Pupation lasts for around one week, after which adult flies emerge. Three generations are normally hatched in each summer. This image was provided by William Deasy.



1.5 Scope of thesis

An overview of the field of animal home range studies is given in the previous sections of this chapter, as is a description of the cabbage root fly larva and its effects on agriculture. The former motivates the modelling in Chapters 2–4, while the latter is relevant in Chapters 5 and 6 as described below. In Chapter 2, we present an overview of the kernel density estimators commonly used for analysing animal movement data, and consider mixture models fitted using a Bayesian implementation of the EM algorithm as a possible alternative. Kernel and mixture methods are compared by simulation studies and application to real data. We ensure the robustness of the mixtures by choosing appropriate densities for the components and prior distributions for the model parameters. We then consider the related subject of home range fidelity in Chapter 3. Several measures of fidelity are introduced and compared to one another by a simulation study, and we then apply these measures to real data, assessing fidelity between successive years.

In Chapter 4, we move on to the topic of diffusion modelling of animal tracks. A bivariate-Ornstein Uhlenbeck (BOU) process is considered as a model for the data, and we obtain maximum likelihood estimates of the parameters. Bayesian Markov chain Monte Carlo (MCMC) methods are used to apply the BOU process to animal movement data, and a mixture of multiple BOU processes is then used. In Chapter 5, more flexible diffusion models are applied in a different context, to the movements of cabbage root fly larvae in a series of bioassays. Mixtures of multiple diffusion processes are also considered as a potentially improved model that accounts for localized movements present in the data.

In Chapter 6, we introduce Hidden Markov models as an alternative to the simpler mixtures previously considered. Additional states are considered to investigate various aspects of the data such as stationary observations due to discretization error, and to allow for varying strengths of attraction and repulsion.

Finally, in Chapter 7, we conclude the thesis by providing an overall discussion and offer suggestions for possible further research on the work presented in this thesis.

Chapter 2

Analysis of home range data

2.1 Introduction

This chapter is concerned with methods for utilization density estimation in home range studies. Two approaches for estimation are the use of kernel estimators (Silverman, 1981, 1986; Scott, 1992; Wand and Jones, 1995), which is now a standard method for analysis of home range data, and the use of finite mixtures of probability distributions.

Kernel methods are nonparametric, as a kernel density estimator is dependent only on the data and on one or more smoothing parameters. Several methods exist for the selection of smoothing parameters (Bowman, 1984; Silverman, 1986; Park and Marron, 1990; Hall et al., 1991; Sheather and Jones, 1991; Terrell and Scott, 1992), and objective selection procedures such as least-squares cross-validation give estimates with low bias at the cost of high variance (Seaman and Powell, 1996; Seaman et al., 1999). Kernels are more effective than the alternative methods of the minimum convex polygon and harmonic mean (Naef-Daenzer, 1993; Worton, 1995a; Swihart and Slade, 1997), and can estimate densities of any shape effectively, provided that the degree of smoothing is appropriate. However, the superiority of kernels has been questioned (Robertson et al., 1998; Girard et al., 2002). Hemson et al. (2005) found that kernels are of limited use for data sets containing large numbers of points in close proximity, and that least-squares cross-validation may fail altogether in such cases.

Mixture models offer a flexible way to describe data structures (Frühwirth-Schnatter, 2006, pp. 3–13; Marin and Robert, 2007), and the properties of finite mixtures of multivariate normal distributions are well documented (Peters and Walker, 1977; Lo et al., 2001). The use of maximum likelihood estimates for the parameters is recommended (Hasselblad, 1966; Day, 1969), and the EM algorithm

and similar procedures are commonly used to obtain the estimates. Mixtures of t distributions provide a more robust alternative to normal distributions (Wang et al., 2004; Lin et al., 2007; McLachlan et al., 2007; Lin, 2010) and can more easily model data sets with heavy tails, or containing atypical observations (McLachlan and Peel, 2000). However, they have received less attention than normal mixture models at this time (Andrews and McNicholas, 2011). While mixtures have been the focus of important statistical work (Akaike, 1973; McLachlan and Basford, 1988; Celeux and Govaert, 1995; Richardson and Green, 1997; McLachlan and Peel, 2000; McGrory and Titterington, 2007; McNicholas and Murphy, 2008), they have been little used in comparison with kernel methods for analysing home range data.

In the following sections, we review the properties of the bivariate kernel density estimator and of finite mixtures of bivariate normal densities, and obtain parameter estimates for the mixture model using the EM algorithm. We consider a mixture of t distributions as an alternative to the normal mixture, and investigate its properties in a simulation study. Further studies are conducted to compare the effectiveness of kernel and normal mixture models, using integrated squared error as a criterion for model comparison. We then apply the models to real data, and compare the resulting density estimates. We introduce prior distributions for the parameters of the mixtures of t distributions in order to increase robustness by weighting the estimated degrees of freedom towards small values.

2.2 Review of kernel and mixture models

2.2.1 Nonparametric kernel methods

The current standard method for quantifying animals' use of space in home range studies is to estimate the probability density of their utilization of space using a kernel estimator (Worton 1989, 1995b). In the case of a bivariate data set $\{\mathbf{x}_1 = (x_{1,1}, x_{1,2})^T, \ldots, \mathbf{x}_n = (x_{n,1}, x_{n,2})^T\}$, the estimator takes the following form (Scott, 1992; Silverman, 1986, pp. 100–106; Wand and Jones, 1995),

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right),$$

where kernel K is a unimodal symmetric bivariate probability density function, and $h > 0$ is a smoothing parameter, also called the bandwidth, that can be

varied. When the bivariate normal density kernel is used, the resulting estimate is given by

$$\hat{f}_h(\mathbf{x}) = \frac{1}{2\pi n h^2} \sum_{i=1}^{n} \exp\left\{ -\frac{(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)}{2h^2} \right\}. \tag{2.1}$$

The choice of smoothing parameter h determines the variance of each component of the above estimate and therefore the level of detail that will be retained from the data set. A small value of h results in little smoothing and preserves much of the detail, whereas a larger value results in a more heavily smoothed density estimate that reflects only the most prominent features of the data.

An alternative to the fixed kernel shown in (2.1) is the adaptive kernel method, which varies the smoothing parameter so that areas with fewer data points have higher values of h, and are therefore smoothed more heavily. Consequently the adaptive kernel smooths the tails of the density more than the fixed method, and is less influenced by outlying points. The adaptive kernel density estimator takes the form

$$\hat{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^2} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_i} \right).$$

Here, $h_i = h\lambda_i$, where h is a global smoothing parameter and the λ_i's are local smoothing parameters defined by

$$\lambda_i = \left\{ \frac{\tilde{f}(\mathbf{x}_i)}{g} \right\}^{-\frac{1}{2}},$$

where $\tilde{f}(\mathbf{x}_i)$ is a pilot estimate of $f(\mathbf{x})$ evaluated at the data point $\mathbf{x}_i$, and

$$g = \frac{1}{n} \sum_{i=1}^{n} \log \tilde{f}(\mathbf{x}_i).$$

A simple ad-hoc method of choosing h is to use the optimum value for a standard distribution such as the normal. For a fixed bivariate normal kernel with covariance matrix of the form $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the 2×2 identity matrix, the optimum choice for a large sample size n is $h = \sigma n^{-1/6}$. As the value of σ will generally be unknown, h must be estimated, with an obvious choice of estimate being $\hat{h} = \hat{\sigma} n^{-1/6}$, where $\hat{\sigma}^2 = \frac{1}{2}(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$, and where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the sample variances of the data sets $\{x_{1,1}, \ldots, x_{n,1}\}$ and $\{x_{1,2}, \ldots, x_{n,2}\}$ respectively. Often, in practice, robust versions of sample variances are used due

to outliers. However, animals' utilization distributions are frequently non-normal, with multiple modes. This ad-hoc choice of h treats the utilization distribution as bivariate normal and gives a value that produces the amount of smoothing appropriate for such a density, which can result in oversmoothing for a utilization distribution that strongly departs from normality by being multimodal (Naef-Daenzer, 1993; Seaman and Powell, 1996; Seaman et al., 1999; Powell, 2000).

The method of least-squares cross-validation (Bowman, 1984; Silverman, 1986; Worton, 1989; McLachlan, 1992) provides an objective means of estimating the smoothing parameter, and is often considered a default selection method for home range estimation with kernels (Hooge and Eichenlaub, 1997; Gitzen and Millspaugh, 2003). The value of h is chosen to minimize the mean integrated squared error (MISE), which is defined as

$$MISE(h) = E\left[\int \{\hat{f}_h(\mathbf{x}) - f(\mathbf{x})\}^2 \mathrm{d}\mathbf{x}\right],$$

where $f(\mathbf{x})$ is the true underlying density of the data set evaluated at $\mathbf{x}$ (Azzalini and Bowman, 1997). When using fixed kernel density estimation with a bivariate normal kernel, the following approximate equality holds,

$$MISE(h) = E\left[\int \{\hat{f}_h(\mathbf{x}) - f(\mathbf{x})\}^2 \mathrm{d}\mathbf{x}\right] \approx E\left\{M(h)\right\} + \int f^2(\mathbf{x})\mathrm{d}\mathbf{x},$$

where

$$M(h) = \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K^*\left(\frac{\mathbf{x}_i - \mathbf{x}_j}{h}\right) + 2n^{-1}h^{-2}K(\mathbf{0}).$$

Here, $\mathbf{0}$ is defined as $(0,0)^T$, and $K^* = K^{(2)} - 2K$, where $K^{(2)}$ is a bivariate normal density with covariance matrix $2\mathbf{I}$. The parameter h is therefore chosen to minimize $M(h)$, with the intention of minimizing $MISE(h)$.

In the case of the adaptive kernel method with a bivariate normal kernel, $MISE(h)$ is instead minimized by minimizing

$$M(h) = \int \hat{f}_h^{\,2}(\mathbf{x})\mathrm{d}\mathbf{x} - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(\mathbf{x}_i),$$

where $\hat{f}_{-i}$ is the density estimate constructed from all data points except $\mathbf{x}_i$, given

by

$$\hat{f}_{-i}(\mathbf{x}) = \frac{1}{n-1}\sum_{j\neq i}\frac{1}{h^2\lambda_j^2}K\left(\frac{\mathbf{x}-\mathbf{x}_j}{h\lambda_j}\right),$$

and

$$\int \hat{f}_h^{\,2}(\mathbf{x})\mathrm{d}\mathbf{x} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K^{(2)}(\mathbf{x}_i-\mathbf{x}_j;\lambda_i h,\lambda_j h),$$

where $K^{(2)}(\mathbf{x}_i-\mathbf{x}_j;\lambda_i h,\lambda_j h)$ is the bivariate normal density with covariance matrix $(\lambda_i^2+\lambda_j^2)h^2\mathbf{I}$. Other methods exist for smoothing parameter selection, such as the bootstrap method (Taylor, 1989). This entails choosing h to minimize

$$\int\left\{\hat{f}_h^*(\mathbf{x})-\hat{f}_h(\mathbf{x})\right\}^2\mathrm{d}\mathbf{x},$$

where $\hat{f}_h^*(\mathbf{x})$ is the kernel estimate using a resampled data set $\{\mathbf{x}_1^*,\ldots,\mathbf{x}_n^*\}$, where each $\mathbf{x}_i^*$ is a realization of a random variable $\mathbf{X}_i^*$ with probability density function $\hat{f}_h(\mathbf{x})$. A realization of $\mathbf{X}_i^*$ can be generated as follows (Silverman, 1986, p. 143; Taylor, 1989),

- Choose an integer I with equal probability from $\{1,\ldots,n\}$,
- Generate a random variable $\phi\sim K(\mathbf{x})$,
- Set $\mathbf{X}_i^* = \mathbf{X}_i + h\phi$.

Another method of smoothing parameter selection is the direct plug-in rule (Wand and Jones, 1994), which minimizes the asymptotic mean squared error of the density estimate and is described in Chapter 3.

2.2.2 Parametric mixture modelling

While the standard method for analysis of home range data is to estimate the utilization density by using a kernel or convex hull based estimator (Worton 1987, 1995b), the former of which has been discussed in Section 2.2.1, finite mixture modelling has attractive properties when applied to this area of research as it can be used to build in known features of movement, for example using known den locations as the means of mixture components (Don and Rennolls, 1983). The mixture approach provides an improvement on the simple parametric bivariate normal based estimate of Jennrich and Turner (1969), which was extensively

used in the 1970s. However, Don and Rennolls assumed circular bivariate normal mixture components, which are too restrictive in many cases. Here, we consider more flexible finite mixture modelling (McLachlan and Peel, 2000).

2.2.3 Mixtures of bivariate normal distributions

The R package `mclust` (Fraley and Raftery, 2003, 2006, 2007) contains the procedure "em", which fits parametrized multivariate normal mixture models to data sets. It also contains the procedure "mclust", which can be used to select the optimal multivariate normal mixture for a data set, and displays the parameter values of the selected model. The package assesses the optimality of mixtures of spherical, diagonal and ellipsoidal distributions of equal and unequal volume, shape (in the case of diagonal and ellipsoidal models), and orientation (in the case of ellipsoidal models only). It uses the Bayesian Information Criterion (BIC) as the selection criterion for the optimal model. BIC was introduced by Schwarz (1978), and is defined as

$$\mathrm{BIC} = -2 \log L + k \log n,$$

where $\log L$ is the log-likelihood of the model, k is the number of parameters, and n is the number of observations.

The package `mclust` uses the Expectation-Maximization (EM) algorithm to determine the mean, covariance matrix and weight of each component distribution. The EM algorithm is used for maximum likelihood estimation in cases where estimation would be easy given the presence of additional data (Dempster et al., 1977; Titterington et al., 1985), and its use in the context of finite mixture models is well documented (McLachlan and Peel, 2000; Frühwirth-Schnatter, 2006). The formulation we use here is given in McLachlan and Krishnan (2008, pp. 18–20), and treats the observed data vector $\mathbf{x}$ as an incomplete version of a complete data set whose analysis would be straightforward (Davison, 2003, p. 210). The vector $\mathbf{x}$ is postulated to have a density of the form $f(\mathbf{x}, \theta)$, where θ is a vector of unknown parameters from a parameter space Θ.

Denoting the complete data vector by $\mathbf{z}$, the complete-data log-likelihood for θ is

$$\log L_c(\theta) = \log f_c(\mathbf{z}, \theta) = \log f(\mathbf{x}, \theta) + \log f(\mathbf{z}|\mathbf{x}, \theta).$$

The vectors $\mathbf{x}$ and $\mathbf{z}$ belong to sample spaces X and Z respectively, and $\mathbf{x}$ is considered as a function of $\mathbf{z}$. Therefore, the observed data vector $\mathbf{x}$ is used

instead of the complete vector $\mathbf{z}$, and it is apparent that the density of $\mathbf{x}$ is given by

$$f(\mathbf{x}, \theta) = \int_{Z(\mathbf{x})} f_c(\mathbf{z}, \theta) \mathrm{d}\mathbf{z},$$

where $Z(\mathbf{x})$ is the subset of Z such that $\mathbf{x} = \mathbf{x}(\mathbf{z})$. The EM algorithm finds the maximum likelihood estimate of θ by iteration of $\log L_c(\theta)$. As the data set $\mathbf{z}$ is partially unobservable, the complete-data log-likelihood cannot be calculated, so it is replaced by its conditional expectation given $\mathbf{x}$, using the current estimate for θ. Given an initial estimate $\theta^{(0)}$, the k'th iteration of the algorithm (starting from $k = 1$) is as follows (McLachlan and Krishnan, 2008, p. 19).

First, the conditional expectation

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}} \{\log L_c(\theta) | \mathbf{x}\}$$

is calculated. This is referred to as the E-step of the algorithm. Then, $\theta^{(k+1)}$ is chosen to maximize $Q(\theta, \theta^{(k)})$, so that

$$Q(\theta^{(k+1)}, \theta^{(k)}) \geq Q(\theta, \theta^{(k)}),$$

for all $\theta \in \Theta$. This is the M-step. These steps are repeated until the difference in likelihoods,

$$L(\theta^{(k+1)}) - L(\theta^{(k)}),$$

falls below a set threshold of convergence. A threshold for $\theta^{(k+1)} - \theta^{(k)}$ can be used instead if using the likelihood difference is impractical. It can be shown that

$$L(\theta^{(k+1)}) \geq L(\theta^{(k)}),$$

so the likelihood is non-decreasing with successive iterations, and convergence must be reached with a sequence of likelihood values that are bounded above. However, convergence may be to a local maximum rather than a global maximum.

2.2.4 Mixtures of bivariate t distributions

Mixtures of bivariate normal distributions fitted to animal movement data may be excessively influenced by outlying points. Such outliers are a common feature of home range data sets, as animals will occasionally visit areas that are not

in their home range in order to investigate them. Mixtures of heavy-tailed components such as bivariate t densities have been considered for modelling data sets containing atypical observations (Wang et al., 2004; Lin et al., 2007; Lin, 2010). Here, the bivariate t distributions should be less strongly affected by small numbers of outliers, and can be used in an attempt to create a more robust model.

A p-dimensional random variable $\mathbf{x}$ is said to have a multivariate t distribution $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ with mean $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, and ν degrees of freedom, if it has density function

$$f_p(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(\pi\nu)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)\{1+\delta(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})/\nu\}^{\frac{\nu+p}{2}}},$$

where

$$\delta(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

is the Malahanobis squared distance between $\mathbf{x}$ and $\boldsymbol{\mu}$ with $\boldsymbol{\Sigma}$ as the covariance matrix.

When bivariate t distributions are used in place of normal distributions in the mixture, the EM algorithm can similarly be used to determine the properties of the components as shown below (McLachlan and Peel, 2000; McLachlan and Krishnan, 2008, pp. 58–61, 176–177). Given an observed random sample $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ from a g-component mixture of p-dimensional t distributions, the p.d.f. of the jth observation from the sample is given by

$$f(\mathbf{x}_j, \boldsymbol{\Psi}) = \sum_{i=1}^{g} w_i f(\mathbf{x}_j, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \nu_i),$$

where $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ and ν_i are the mean, scale matrix and degrees of freedom of the ith mixture component respectively, $\{w_1, \ldots, w_g\}$ are the weights of the g components, and the vector of parameters $\boldsymbol{\Psi}$ is defined as

$$\boldsymbol{\Psi} = (w_1, \ldots, w_{g-1}, \theta^T, \boldsymbol{\nu}^T)^T,$$

where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_g)$ and θ contains the distinct elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The observed data $\mathbf{x}$ is viewed as incomplete. For the purposes of the EM algorithm, the complete-data vector is given by

$$\mathbf{z} = (\mathbf{x}, \mathbf{z}_1^T, \ldots, \mathbf{z}_n^T, \mathbf{u}^T)^T,$$

where $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ are the component-label vectors defining the component of origin of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ respectively, and $z_{ij} = (\mathbf{z}_j)_i$ is equal to 1 if $\mathbf{x}_j$ belongs to the ith mixture component and 0 if it does not. The missing data $\mathbf{u} = (u_1, \ldots, u_n)^T$ are defined such that

$$(\mathbf{x}_i | u_i, z_{ij} = 1) \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}/u_j)$$

and

$$(u_i | z_{ij} = 1) \sim \Gamma\left(\frac{\boldsymbol{\nu}}{2}, \frac{\boldsymbol{\nu}}{2}\right)$$

independently for $i = 1, \ldots, n$.

The complete-data likelihood function can be factored into the product of the marginal densities of $\mathbf{z}_j$, and the conditional densities of u_j given $\mathbf{z}_j$ and $\mathbf{x}_j$ given u_j and $\mathbf{z}_j$. Accordingly, the complete-data log-likelihood can be written as

$$\log L_c(\boldsymbol{\Psi}) = \log L_{1c}(\mathbf{w}) + \log L_{2c}(\boldsymbol{\nu}) + \log L_{3c}(\theta), \tag{2.2}$$

where

$$\log L_{1c}(\mathbf{w}) = \sum_{i=1}^{g} \sum_{j=1}^{n} z_{ij} \log w_i, \tag{2.3}$$

$$\log L_{2c}(\boldsymbol{\nu}) = \sum_{i=1}^{g} \sum_{j=1}^{n} z_{ij} \left\{ -\log \Gamma\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2} \log\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2}(\log u_j - u_j) - \log u_j \right\}, \tag{2.4}$$

$$\log L_{3c}(\theta) = \sum_{i=1}^{g} \sum_{j=1}^{n} z_{ij} \left\{ -\frac{p}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \frac{1}{2} u_j (\mathbf{x}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}_j - \boldsymbol{\mu}_i) \right\}. \tag{2.5}$$

The E-step on the $(k+1)$th iteration of the algorithm requires the calculation of $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(k)}) = E_{\theta^{(k)}} \{\log L_c(\boldsymbol{\Psi}) | \mathbf{x}\}$. This can be found by first taking the expectation of $\log L_c(\boldsymbol{\Psi})$ conditional on $\mathbf{x}$ and $\mathbf{z}$, and then over $\mathbf{z}$ given $\mathbf{x}$. It can be seen from (2.3), (2.4) and (2.5) that to do so it is necessary to calculate $E_{\boldsymbol{\Psi}}^{(k)}(z_{ij}|\mathbf{x}_j)$, $E_{\boldsymbol{\Psi}}^{(k)}(u_j|\mathbf{x}_j, \mathbf{z}_j)$, and $E_{\boldsymbol{\Psi}}^{(k)}(\log u_j|\mathbf{x}_j, \mathbf{z}_j)$ for $i = 1, \ldots, g$ and $j = 1, \ldots, n$. These expectations above take the following forms,

$$E_{\boldsymbol{\Psi}}^{(k)}(z_{ij}|\mathbf{x}_j) = \tau_{ij}^{(k)},$$

$$E_{\boldsymbol{\Psi}}^{(k)}\left(u_j|\mathbf{x}_j,\mathbf{z}_j\right)=u_{ij}^{(k)},\text{ and}$$
$$E_{\boldsymbol{\Psi}}^{(k)}\left(\log u_j|\mathbf{x}_j,\mathbf{z}_j\right)=\log u_{ij}^{(k)}+\phi\left(\frac{\nu_i^{(k)}+p}{2}\right)-\log\left(\frac{\nu_i^{(k)}+p}{2}\right).$$

Here, $\tau_{ij}^{(k)}$ is the posterior probability that $\mathbf{x}_j$ belongs to the ith mixture component using the current fit $\boldsymbol{\Psi}^{(k)}$ for $\boldsymbol{\Psi}$, and is given by

$$\tau_{ij}^{(k)}=E_{\boldsymbol{\Psi}^{(k)}}(z_{ij}|\mathbf{x}_j)=\frac{w_i^{(k)}f(\mathbf{x}_j,\boldsymbol{\mu}_i^{(k)},\boldsymbol{\Sigma}_i^{(k)},\nu_i^{(k)})}{f(\mathbf{x}_j,\boldsymbol{\Psi}^{(k)})},$$

following from the definition of $\mathbf{z}_j$. Note that

$$u_{ij}^{(k)}=\frac{\nu_i^{(k)}+p}{\nu_i^{(k)}+\delta(\mathbf{x}_j,\boldsymbol{\mu}_i^{(k)},\boldsymbol{\Sigma}_i^{(k)})},$$

and that ϕ is the Digamma function, defined by

$$\phi(s)=\{\partial\Gamma(s)/\partial s\}/\Gamma(s).$$

Following the approach of McLachlan and Peel (2000), $Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(k)})$ can be calculated as

$$Q(\boldsymbol{\Psi},\boldsymbol{\Psi}^{(k)})=Q_1(\mathbf{w},\boldsymbol{\Psi}^{(k)})+Q_2(\boldsymbol{\nu},\boldsymbol{\Psi}^{(k)})+Q_3(\theta,\boldsymbol{\Psi}^{(k)}),$$

where

$$Q_1(\mathbf{w},\boldsymbol{\Psi}^{(k)})=\sum_{i=1}^{g}\sum_{j=1}^{n}\tau_{ij}^{(k)}\log w_i,$$
$$Q_2(\boldsymbol{\nu},\boldsymbol{\Psi}^{(k)})=\sum_{i=1}^{g}\sum_{j=1}^{n}\tau_{ij}^{(k)}Q_{2j}(\nu_i,\boldsymbol{\Psi}^{(k)}),\text{ and}$$
$$Q_3(\theta,\boldsymbol{\Psi}^{(k)})=\sum_{i=1}^{g}\sum_{j=1}^{n}\tau_{ij}^{(k)}Q_{3j}(\theta_i,\boldsymbol{\Psi}^{(k)}),$$

where, ignoring terms not involving ν_i,

$$Q_{2j}(\nu_i,\boldsymbol{\Psi}^{(k)})=-\log\Gamma\left(\frac{\nu_i}{2}\right)+\frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right)$$
$$+\frac{\nu_i}{2}\left\{\log u_{ij}^{(k)}-u_{ij}^{(k)}+\phi\left(\frac{\nu_i^{(k)}+p}{2}\right)-\log\left(\frac{\nu_i^{(k)}+p}{2}\right)\right\},$$

and

$$Q_{3j}(\theta_i, \boldsymbol{\Psi}^{(k)}) = \left\{ -\frac{p}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \frac{p}{2}\log u_{ij}^{(k)} \right. \\ \left. -\frac{1}{2}u_{ij}^{(k)}(\mathbf{x}_j - \boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j - \boldsymbol{\mu}_i) \right\}.$$

On the M-step at the $(k+1)$th iteration, it is apparent that $\mathbf{w}^{(k+1)}$, $\boldsymbol{\nu}^{(k+1)}$ and $\theta^{(k+1)}$ can be computed independently by consideration of $Q_1(\mathbf{w}, \boldsymbol{\Psi}^{(k)})$, $Q_2(\boldsymbol{\nu}, \boldsymbol{\Psi}^{(k)})$ and $Q_3(\theta, \boldsymbol{\Psi}^{(k)})$ respectively. The solutions for $\mathbf{w}^{(k+1)}$ and the elements of $\theta^{(k+1)}$ exist in closed form and are as follows,

$$\begin{aligned} w_i^{(k+1)} &= \sum_{j=1}^{n} \tau_{ij}^{(k)}/n, \\ \boldsymbol{\mu}_i^{(k+1)} &= \frac{\sum_{j=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)} \mathbf{x}_j}{\sum_{j=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)}}, \\ \boldsymbol{\Sigma}_i^{(k+1)} &= \frac{\sum_{j=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)} (\mathbf{x}_j - \boldsymbol{\mu}_i^{(k+1)})(\mathbf{x}_j - \boldsymbol{\mu}_i^{(k+1)})^T}{\sum_{j=1}^{n} \tau_{ij}^{(k)}}. \end{aligned}$$

If the degrees of freedom are fixed for each component, the M-step therefore exists in closed form. If they are not fixed, the solution for ν_i does not exist in closed form but can be computed iteratively at each step by calculating the left hand side of the equations

$$\partial Q_2(\nu_i; \boldsymbol{\Psi}^{(k)})/\partial\nu_i = 0,$$

from which it follows that $\nu_i^{(k+1)}$ is a solution of the equation

$$-\phi\left(\frac{\nu_i}{2}\right) + \log\frac{\nu_i}{2} + 1 + \frac{\sum_{j=1}^{n} \tau_{ij}^{(k)}(\log u_{ij}^{(k)} - u_{ij}^{(k)})}{\sum_{j=1}^{n} \tau_{ij}^{(k)}} + \phi\left(\frac{\nu_i^{(k)} + p}{2}\right) \\ - \log\left(\frac{\nu_i^{(k)} + p}{2}\right) = 0.$$

We created 2 procedures using R to fit mixtures of bivariate t components to data, using the iterative parameter estimates given in this subsection for the means, covariance matrices and weights of the components,

- MIX-T-FIX, which uses fixed degrees of freedom ν,
- MIX-T-VAR, which estimates ν as detailed above.

Both MIX-T-FIX and MIX-T-VAR fit mixtures of bivariate t distributions with up to a selected number of components, and then select the optimal fitted model using BIC.

2.3 Simulation study

We conducted a simulation study to explore the properties of the procedures for fitting mixtures of bivariate t components introduced in Section 2.2.5. We simulated 200 data sets from the following density, a mixture of 2 bivariate normal distributions,

$$\frac{1}{2}\mathrm{N}\left\{(0,0)^T, \begin{pmatrix} 5 & -2 \\ 2 & 5 \end{pmatrix}\right\} + \frac{1}{2}\mathrm{N}\left\{(5,5)^T, \begin{pmatrix} 15 & -1 \\ -1 & 2 \end{pmatrix}\right\}. \tag{2.6}$$

Each simulated data set consisted of 100 observations sampled from the density above, and an additional point at the coordinates $(12, -3)$, a clear outlier, for a total of 101 observations. Outliers such as this are commonplace in home range data, and it is therefore important to test the procedures' ability to correctly deal with their presence. We used MIX-T-FIX and MIX-T-VAR to fit mixtures of 1, 2, 3 and 4 t distributions to the simulated data sets.

For MIX-T-FIX, all of the t distributions were given $\nu = 2$, providing a highly robust estimator. This number is ad hoc, but is necessarily low. A large value of ν, such as $\nu = 25$, would result in a model very similar in nature to a mixture of normal distributions and would defeat the purpose of using the t distribution instead.

In the case of MIX-T-VAR, experimentation revealed that the algorithm is of limited effectiveness in estimating ν for each distribution. When the model is applied to a data set with outliers suggesting a departure from normality, and the initial values of ν are high for all components, it will output a low value of ν for at least one of the components. However, when fitted to data with little or no indication of a non-normal distribution and the initial values of ν are low, they will not increase substantially. For this reason, we set the initial degrees of freedom to a high value, specifically $\nu = 100$ for this study.

Another potential limitation of MIX-T-VAR is that when high initial degrees of freedom are used, the presence of outliers will frequently result in a low estimate for the degrees of freedom of one component distribution, while the estimates for the other components are close to the initial values. This behaviour suggests that most or all outliers are assigned to the same component. When a point is far away

from the estimated modes of all of the components, there does not appear to be a realistic way of assessing the component to which the point belongs.

Table 2.1 gives the means and standard deviations of the BIC values of models fitted to the simulated data sets, with $m = 1, 2, 3$ and 4 components. The numbers of simulated data sets for which the optimal model had m components are also included. The table shows that the 2 component model is correctly selected for approximately 90% of the simulated data sets by MIX-T-FIX and MIX-T-VAR, indicating that the procedures identify the true number of components with a high degree of success. The 2 component mixture has the lowest mean BIC for both procedures. Contour plots of the fitted densities for the selected models appear similar for the simulated data sets, and plots for one such data set are displayed in Figure 2.1.

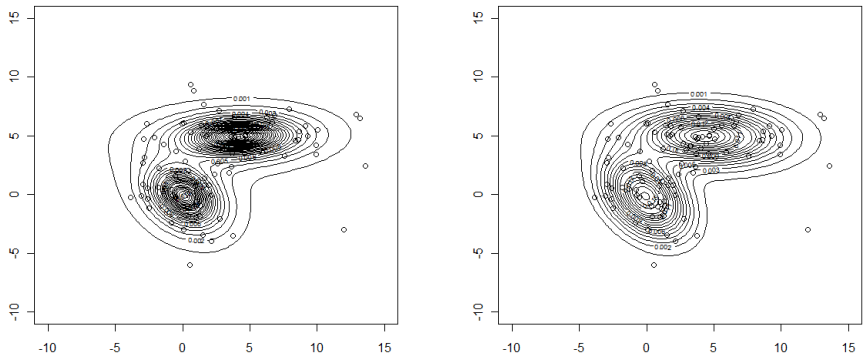
Table 2.1: Means and standard deviations (SD) of BIC values for the mixtures of bivariate t distributions with m components fitted to data sets simulated from the mixture of bivariate normal distributions given in (2.6) with an outlier at $(12, -3)$ using MIX-T-FIX and MIX-T-VAR. Also given are the numbers N of simulated data sets for which the optimal model, using BIC as a selection criterion, had m components.

Estimate		**Number of components**			
		1	**2**	**3**	**4**
MIX-T-FIX	Mean BIC	1131.6	1096.2	1108.5	1124.2
	SD	16.7	15.8	14.0	14.5
	N	2	189	9	0
MIX-T-VAR	Mean BIC	1104.5	1087.3	1106.0	1131.0
	SD	15.6	14.3	19.0	16.3
	N	3	190	7	0

When Figure 2.1 is compared with a contour plot of the true probability density, which is shown in Figure 2.2, it is clear that the models reflect the true density quite closely. MIX-T-VAR appears to provide a slightly superior fit to MIX-T-FIX, which is to be expected as the data are normally distributed with the exception of the single outlier. The degrees of freedom for the model fitted by MIX-T-VAR were $\nu_1 = 6$, $\nu_2 = 116$. The lower value for ν_1 is due to the presence of the outlying point, which has been assigned to that distribution.

A similar study was carried out using a mixture of bivariate t distributions to simulate data sets, resulting in simulations with more outlying points. The

Figure 2.1: Contour plots of mixtures of bivariate t distributions fitted to a data set generated from the mixture of normal distributions given in (2.6) with an outlier at $(12, -3)$. The left and right panels show fitted densities obtained using MIX-T-FIX and MIX-T-VAR respectively.



simulation density used was

$$\frac{1}{4}t_4\{(5,5)^T, \mathbf{I}\} + \frac{1}{2}t_4\{(0,0)^T, \mathbf{I}\} + \frac{1}{4}t_4\{(0,10)^T, \mathbf{I}\}, \tag{2.7}$$

where $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a bivariate t distribution with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom. We simulated 200 data sets of 100 observations each from the density given in (2.7), and MIX-T-FIX and MIX-T-VAR were applied to each simulated data set. As was the case for the simulation study detailed previously in this section, mixtures of 1, 2, 3 and 4 components were fitted, $\nu = 2$ for MIX-T-FIX, initial values of $\nu = 100$ were used for all components of MIX-T-VAR, and BIC was used for model selection. The means and standard deviations of the BIC values for the fitted models, and the number of simulated data sets for which the optimal model had each possible number of components, are shown in Table 2.2.

Table 2.2 shows that the 3 component mixture model has the lowest mean BIC for MIX-T-FIX and MIX-T-VAR, and the correct number of components is selected for about 80% of the simulated data sets using MIX-T-FIX and MIX-T-VAR. This is a lower percentage than was observed for data sets simulated from a mixture of normal distributions (Table 2.1), which suggests that the mixture of t distributions is more difficult to model well using mixture methods than the less variable mixture of normals. However, the correct number of components is still identified for the majority of the simulated data sets.

Figure 2.2: Contour plot of the probability density function given in (2.6), shown with a sample data set generated from the density.

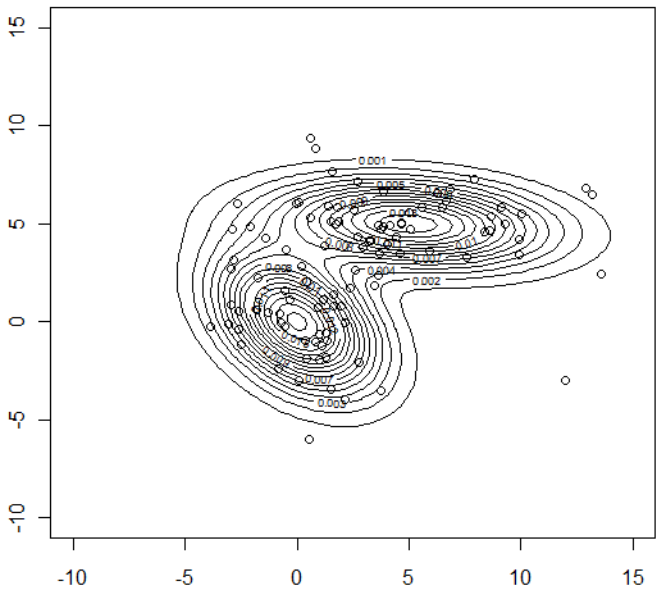


Table 2.2: Means and standard deviations (SD) of BIC values for the mixtures of bivariate t distributions with m components fitted to data sets simulated from the mixture of bivariate t distributions given in (2.7) using MIX-T-FIX and MIX-T-VAR. Also given are the numbers N of simulated data sets for which the optimal model, using BIC as a selection criterion, had m components.

Estimate		**Number of components**			
		1	**2**	**3**	**4**
MIX-T-FIX	Mean BIC	1108.596	988.978	941.722	951.382
	SD	21.915	36.455	23.454	23.311
	N	0	3	171	26
MIX-T-VAR	Mean BIC	1079.251	986.501	949.982	969.144
	SD	16.387	33.352	28.596	38.365
	N	0	19	172	9

Figure 2.3: Contour plots of mixtures of bivariate t distributions fitted to a data set generated from the mixture of bivariate t distributions given in (2.7). The left and right panels show fitted densities obtained using MIX-T-FIX and MIX-T-VAR respectively.

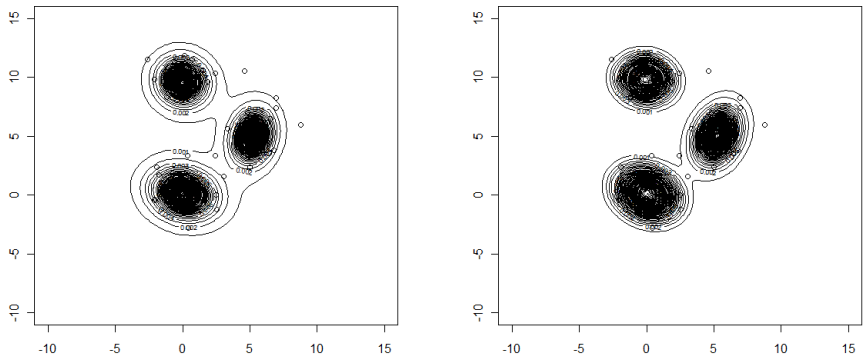
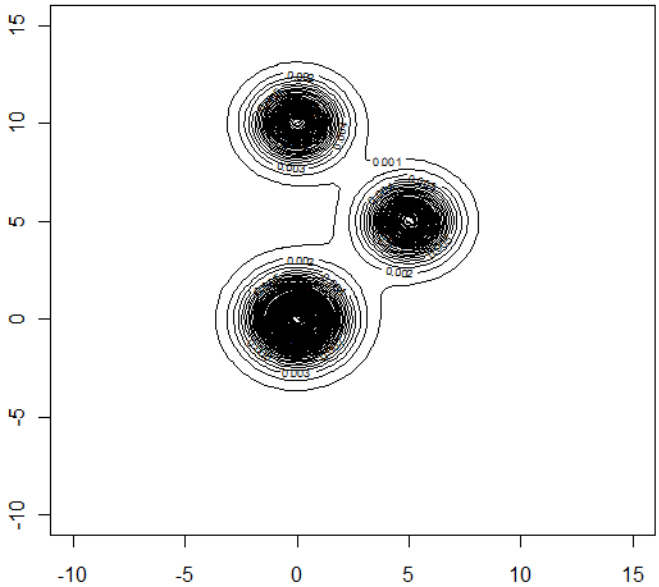


Figure 2.4: Contour plot of the probability density function given in (2.7).



All 3 components of the true probability distribution have only 4 degrees of freedom. However, the 3 component models fitted by MIX-T-VAR frequently consist of 1 or 2 distributions with few degrees of freedom, with the others having higher degrees of freedom. The contour plots of the 3 component models fitted by MIX-T-FIX and MIX-T-VAR appear similar to one another, and an example from one of the simulated data sets is presented in Figure 2.3. Both of the plots in the figure resemble the true density function (Figure 2.4) quite closely. In the case of the particular data set used to generate the contour plots in Figure 2.3, MIX-T-VAR estimated the degrees of freedom of the 3 components as $\nu_1 = 106$, $\nu_2 = 4$, $\nu_3 = 98$, again demonstrating the behaviour of assigning outliers to the same component.

2.4 Model assessment using integrated squared error

When data sets are simulated from known probability distributions as in Section 2.3, it is possible to calculate the integrated squared error (ISE) of an estimated probability density for the data. The ISE is defined as

$$\text{ISE} = \int \left\{ f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right\}^2 \mathrm{d}\mathbf{x},$$

where $f(\mathbf{x})$ and $\hat{f}(\mathbf{x})$ are the true and estimated underlying density of the data respectively. The ISE provides a measure of difference between a fitted distribution and the true distribution, and therefore a method of assessing the suitability of a fitted model.

2.4.1 Calculation for kernel and mixture models

In the general case of a kernel density estimator such as those described in Section 2.2.1, when the true density is a mixture of k components,

$$f(\mathbf{x}) = \sum_{i=1}^{k} w_i f_i(\mathbf{x}),$$

where w_i is the weight assigned to component f_i, and assuming a data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the ISE takes the following form,

$$\begin{aligned}
\text{ISE} &= \int \left\{ \sum_{i=1}^{k} w_i f_i(\mathbf{x}) - \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right) \right\}^2 \mathrm{d}\mathbf{x} \\
&= \int \left\{ \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j f_i(\mathbf{x}) f_j(\mathbf{x}) + \frac{1}{n^2h^4} \sum_{i=1}^{n}\sum_{j=1}^{n} K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right) K\left(\frac{\mathbf{x}-\mathbf{x}_j}{h}\right) \right. \\
&\qquad \left. - \frac{2}{nh^2} \sum_{i=1}^{k}\sum_{j=1}^{n} w_i f_i(\mathbf{x}) K\left(\frac{\mathbf{x}-\mathbf{x}_j}{h}\right) \right\} \mathrm{d}\mathbf{x} \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} \int f_i(\mathbf{x}) f_j(\mathbf{x}) \mathrm{d}\mathbf{x} + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{n^2h^4} \int K\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right) K\left(\frac{\mathbf{x}-\mathbf{x}_j}{h}\right) \mathrm{d}\mathbf{x} \\
&\qquad - 2\sum_{i=1}^{k}\sum_{j=1}^{n} \frac{w_i}{nh} \int f_i(\mathbf{x}) K\left(\frac{\mathbf{x}-\mathbf{x}_j}{h}\right) \mathrm{d}\mathbf{x}.
\end{aligned}$$

The ease with which the ISE can be calculated analytically is clearly dependent on the true probability density and the kernel density estimator used. A tractable case occurs when the true probability density is a mixture of circular bivariate normal density functions, and a bivariate normal density kernel is used. Under these assumptions, the ISE can be calculated as follows. For convenience we write the true probability density as

$$f(\mathbf{x}) = \sum_{i=1}^{k} w_i \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}_i, \sigma_i^2 \mathbf{I}) = \sum_{i=1}^{k} w_i \mathrm{N}_i,$$

where $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}_i, \sigma_i^2\mathbf{I})$ is the bivariate normal probability density function corresponding to the ith mixture component, with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})^T$ and covariance matrix $\sigma_i^2\mathbf{I}$, evaluated at $\mathbf{x}$, and is concisely referred to as N_i.

The bivariate normal kernel density estimate can be expressed as follows,

$$\begin{aligned}
\hat{f}(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2\pi h^2} \exp\left\{ -\frac{(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)}{2h^2} \right\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathrm{N}_{\mathbf{x}}(\mathbf{x}_i, h^2\mathbf{I}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathrm{N}}_i, \qquad (2.8)
\end{aligned}$$

where $\mathrm{N}_{\mathbf{x}}(\mathbf{x}_i, h_i^2\mathbf{I})$ is the bivariate normal density with mean $\mathbf{x}_i$ and covariance

matrix $h_i^2\mathbf{I}$, referred to concisely as $\hat{\mathrm{N}}_i$. The ISE is therefore equal to

$$\int \left(\sum_{i=1}^{k} w_i \mathrm{N}_i - \frac{1}{n} \sum_{i=1}^{n} \hat{\mathrm{N}}_i \right)^2 \mathrm{d}\mathbf{x}$$
$$= \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \int \mathrm{N}_i \mathrm{N}_j \mathrm{d}\mathbf{x} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \hat{\mathrm{N}}_i \hat{\mathrm{N}}_j \mathrm{d}\mathbf{x} - \frac{2}{n} \sum_{i=1}^{k} \sum_{j=1}^{n} w_i \int \mathrm{N}_i \hat{\mathrm{N}}_j \mathrm{d}\mathbf{x}. \tag{2.9}$$

To calculate the ISE, it is therefore necessary to determine the integral of the product of two circular bivariate normal densities. The densities $\mathrm{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $\mathrm{N}(\boldsymbol{\nu}, \tau^2\mathbf{I})$ are considered, with mean vectors $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and $\boldsymbol{\nu} = (\nu_1, \nu_2)^T$ respectively, and respective covariance matrices $\sigma^2\mathbf{I}$ and $\tau^2\mathbf{I}$. Multiplying the probability density functions together, evaluated at $\mathbf{x} = (x_1, x_2)^T$, gives

$$\begin{aligned}
&\mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I}) \\
&\quad = \left(2\pi\sigma^2\right)^{-1} \exp\left\{ -\frac{\sum_{i=1}^{2} (x_i - \mu_i)^2}{2\sigma^2} \right\} (2\pi\tau^2)^{-1} \exp\left\{ -\frac{\sum_{i=1}^{2} (x_i - \nu_i)^2}{2\tau^2} \right\} \\
&\quad = (4\pi^2\sigma^2\tau^2)^{-1} \exp\left\{ -\frac{\sum_{i=1}^{2}(x_i - \mu_i)^2}{2\sigma^2} - \frac{\sum_{i=1}^{2}(x_i - \nu_i)^2}{2\tau^2} \right\} \\
&\quad = (4\pi^2\sigma^2\tau^2)^{-1} \exp\left\{ -\frac{1}{2\sigma^2\tau^2} \sum_{i=1}^{2} (x_i^2\tau^2 - 2x_i\mu_i\tau^2 + \mu_i^2\tau^2 + x_i^2\sigma^2 - 2x_i\nu_i\sigma^2 \right. \\
&\qquad \left. + \nu_i^2\sigma^2) \right\} \\
&\quad = (4\pi^2\sigma^2\tau^2)^{-1} \exp\left[-\frac{1}{2\sigma^2\tau^2} \sum_{i=1}^{2} \left\{ (\sigma^2 + \tau^2)x_i^2 - 2(\mu_i\tau^2 + \nu_i\sigma^2)x_i \right\}^2 \right] \\
&\qquad \times \exp\left\{ -\frac{1}{2\sigma^2\tau^2} \sum_{i=1}^{2} \left(\mu_i^2\tau^2 + \nu_i^2\sigma^2 \right) \right\} \\
&\quad = (4\pi\sigma^2\tau^2)^{-1} \exp\left\{ -\frac{(\sigma^2 + \tau^2)}{2\sigma^2\tau^2} \sum_{i=1}^{2} \left(x_i - \frac{\mu_i\tau^2 + \nu_i\sigma^2}{\sigma^2 + \tau^2} \right)^2 \right\} \\
&\qquad \times \exp\left[\frac{1}{2\sigma^2\tau^2} \sum_{i=1}^{2} \left\{ \frac{(\mu_i\tau^2 + \nu_i\sigma^2)^2}{\sigma^2 + \tau^2} - \mu_i^2\tau^2 - \nu_i^2\sigma^2 \right\} \right] \\
&\quad = (4\pi^2\sigma^2\tau^2)^{-1} \exp\left\{ -\frac{(\sigma^2 + \tau^2)}{2\sigma^2\tau^2} \sum_{i=1}^{2} \left(x_i - \frac{\mu_i\tau^2 + \nu_i\sigma^2}{\sigma^2 + \tau^2} \right)^2 \right\}
\end{aligned}$$

$$\times \exp\left\{\frac{1}{2\sigma^2\tau^2} \times \sum_{i=1}^{2}\left(\frac{\mu_i^2\tau^4+\nu_i^2\sigma^4+2\mu_i\nu_i\sigma^2\tau^2-\mu_i^2\tau^4-\mu_i^2\sigma^2\tau^2-\nu_i^2\sigma^4-\nu_i^2\sigma^2\tau^2}{\sigma^2+\tau^2}\right)\right\}.$$

Cancelling out terms and factorising in the argument of the second exponential in the expression given above gives the following,

$$\begin{aligned}
&(4\pi\sigma^2\tau^2)^{-1}\exp\left\{-\frac{(\sigma^2+\tau^2)}{2\sigma^2\tau^2}\sum_{i=1}^{2}\left(x_i-\frac{\mu_i\tau^2+\nu_i\sigma^2}{\sigma^2+\tau^2}\right)^2\right\}\\
&\quad\times\exp\left\{-\frac{1}{2(\sigma^2+\tau^2)}\sum_{i=1}^{2}(\mu_i-\nu_i)^2\right\}\\
&\quad=\{2\pi(\sigma^2+\tau^2)\}^{-1}\exp\left\{-\frac{\sum_{i=1}^{2}(\mu_i-\nu_i)^2}{2(\sigma^2+\tau^2)}\right\}\mathrm{N}_{\mathbf{x}}\left(\frac{\boldsymbol{\mu}\tau^2+\boldsymbol{\nu}\sigma^2}{\sigma^2+\tau^2},\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\mathbf{I}\right).
\end{aligned}$$

Calculating the integral of the product of the two densities is now possible, as we observe that

$$\begin{aligned}
&\int\left[\{2\pi(\sigma^2+\tau^2)\}^{-1}\exp\left\{-\frac{\sum_{i=1}^{2}(\mu_i-\nu_i)^2}{2(\sigma^2+\tau^2)}\right\}\mathrm{N}_{\mathbf{x}}\left(\frac{\boldsymbol{\mu}\tau^2+\boldsymbol{\nu}\sigma^2}{\sigma^2+\tau^2},\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\mathbf{I}\right)\right]\mathrm{d}\mathbf{x}\\
&\quad=\{2\pi(\sigma^2+\tau^2)\}^{-1}\exp\left\{-\frac{-\sum_{i=1}^{2}(\mu_i-\nu_i)^2}{2(\sigma^2+\tau^2)}\right\}\\
&\qquad\times\int\mathrm{N}_{\mathbf{x}}\left(\frac{\boldsymbol{\mu}\tau^2+\boldsymbol{\nu}\sigma^2}{\sigma^2+\tau^2},\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\mathbf{I}\right)\mathrm{d}\mathbf{x}\\
&\quad=\{2\pi(\sigma^2+\tau^2)\}^{-1}\exp\left\{-\frac{\sum_{i=1}^{2}(\mu_i-\nu_i)^2}{2(\sigma^2+\tau^2)}\right\},
\end{aligned}$$

as the bivariate normal distribution in the expression above integrates to unity. Using the above derivation, we obtain the ISE for a bivariate normal kernel density estimator shown in (2.9) which can be written as

$$\begin{aligned}
\mathrm{ISE}=&\frac{1}{2\pi}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{w_iw_j}{\sigma_i^2+\sigma_j^2}\exp\left\{-\frac{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}{2(\sigma_i^2+\sigma_j^2)}\right\}\\
&+\frac{1}{4\pi n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{x}_j)^T(\mathbf{x}_i-\mathbf{x}_j)}{4h^2}\right\}\\
&-\frac{1}{n\pi}\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{w_i}{\sigma_i^2+h^2}\exp\left\{-\frac{(\boldsymbol{\mu}_i-\mathbf{x}_j)^T(\boldsymbol{\mu}_i-\mathbf{x}_j)}{2(\sigma_i^2+h^2)}\right\},
\end{aligned}\tag{2.10}$$

and is therefore easy to calculate. When an adaptive kernel is used, the fixed smoothing parameter h in the equations above is replaced with individual parameters h_i for each observation. The calculations are similar, resulting in the following expression for the ISE,

$$\begin{aligned}\text{ISE} = &\frac{1}{2\pi}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{w_i w_j}{\sigma_i^2+\sigma_j^2}\exp\left\{-\frac{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}{2(\sigma_i^2+\sigma_j^2)}\right\}\\ &+\frac{1}{2\pi n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{h_i^2+h_j^2}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{x}_j)^T(\mathbf{x}_i-\mathbf{x}_j)}{2(h_i^2+h_j^2)}\right\}\\ &-\frac{1}{n\pi}\sum_{i=1}^{k}\sum_{j=1}^{n}\frac{w_i}{\sigma_i^2+h_j^2}\exp\left\{-\frac{(\boldsymbol{\mu}_i-\mathbf{x}_j)^T(\boldsymbol{\mu}_i-\mathbf{x}_j)}{2(\sigma_i^2+h_j^2)}\right\}.\end{aligned}$$

The ISE can also be calculated when a bivariate normal mixture is used in place of a bivariate normal kernel density estimator. If a mixture of m circular bivariate normal distributions is used to estimate the underlying density of a data set, the resulting density estimate takes the following form,

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{m}\hat{w}_i \mathrm{N}_{\mathbf{x}}\left(\hat{\boldsymbol{\mu}}_i, \hat{\sigma}^2\mathbf{I}\right) = \sum_{i=1}^{m}\hat{w}_i\hat{\mathrm{N}}_i,$$

where $\hat{w}_i$ is the estimated weight of the ith mixture component and $\hat{\mathrm{N}}_i = \mathrm{N}_{\mathbf{x}}(\hat{\boldsymbol{\mu}}_i, \hat{\sigma}^2\mathbf{I})$ is the normal density using the corresponding estimated mean $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, \hat{\mu}_{i2})^T$ and covariance matrix $\hat{\sigma}_i^2\mathbf{I}$, considered as a function of $\mathbf{x}$. Assuming, as for the normal kernel density estimators, that the true probability density is a mixture of k circular bivariate normal components, it is therefore the case that

$$\begin{aligned}\text{ISE} &= \int\left(\sum_{i=1}^{k}w_i\mathrm{N}_i - \sum_{i=1}^{m}\hat{w}_i\hat{\mathrm{N}}_i\right)^2 \mathrm{d}\mathbf{x}\\ &= \sum_{i=1}^{k}\sum_{j=1}^{k}w_i w_j\int \mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x} + \sum_{i=1}^{m}\sum_{j=1}^{m}\hat{w}_i\hat{w}_j\int\hat{\mathrm{N}}_i\hat{\mathrm{N}}_j\mathrm{d}\mathbf{x}\\ &\quad - 2\sum_{i=1}^{k}\sum_{j=1}^{m}w_i\hat{w}_j\int\mathrm{N}_i\hat{\mathrm{N}}_j\mathrm{d}\mathbf{x}.\end{aligned}$$

Using the same reasoning that we employed for the normal kernel density estimate, this expression can be written as

$$\begin{aligned}\text{ISE} = &\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{w_i w_j}{2\pi(\sigma_i^2+\sigma_j^2)}\exp\left\{-\frac{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}{2(\sigma_i^2+\sigma_j^2)}\right\}\\ &+\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\hat{w}_i\hat{w}_j}{2\pi(\hat{\sigma}_i^2+\hat{\sigma}_j^2)}\exp\left\{-\frac{(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)^T(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)}{2(\hat{\sigma}_i^2+\hat{\sigma}_j^2)}\right\}\\ &-\sum_{i=1}^{k}\sum_{j=1}^{m}\frac{w_i\hat{w}_j}{\pi(\sigma_i^2+\hat{\sigma}_j^2)}\exp\left\{-\frac{(\boldsymbol{\mu}_i-\hat{\boldsymbol{\mu}}_j)^T(\boldsymbol{\mu}_i-\hat{\boldsymbol{\mu}}_j)}{2(\sigma_i^2+\hat{\sigma}_j^2)}\right\}. \qquad (2.11)\end{aligned}$$

2.4.2 Model comparison

In Section 2.2, BIC was used as a criterion for model selection. ISE provides an alternative approach, and here we apply the results of Section 2.4.1 in a simulation study to compare bivariate normal kernel density estimators and mixture models.

We simulated 1000 data sets of 100 observations each from the following density,

$$\frac{1}{2}\text{N}\{(0,0)^T, 5\mathbf{I}\} + \frac{1}{2}\text{N}\{(5,5)^T, 2\mathbf{I}\}. \qquad (2.12)$$

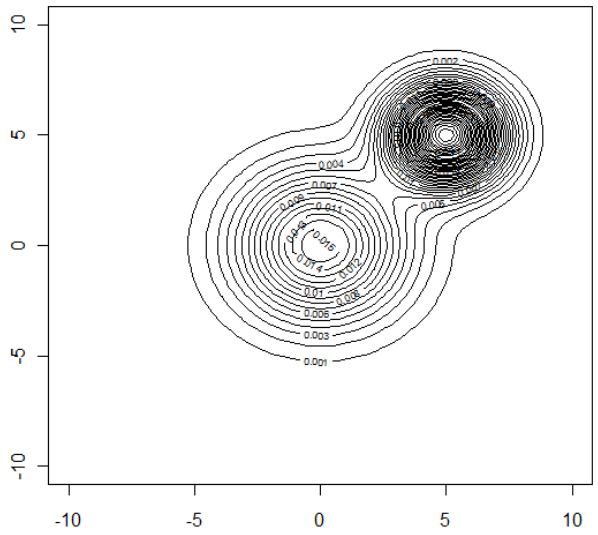
As the density in (2.12) is a mixture of circular bivariate normal distributions, we can calculate the ISE for normal kernel and normal mixture models using the methods described in the previous subsection, as seen in (2.10) and (2.11). A contour plot of the simulation density is displayed in Figure 2.5 as a point of reference. We used fixed and adaptive kernel methods to obtain density estimates for each data set, using least-squares cross-validation to select the smoothing parameters. The `em` procedure contained in the `mclust` package was used to fit a mixture of circular bivariate normal distributions to each data set.

While `mclust` is capable of determining the most suitable mixture of normal distributions to be fitted to a given data set, in this case it is necessary to calculate the ISE of the fitted model analytically. The `em` procedure fits a mixture of a specified number of distributions, and we determined the appropriate number by first using `mclust` to determine the optimal number of components for each simulated data set.

The means and standard deviations of the ISE for each of the modelling methods used are presented in Table 2.3. They are each expressed as a percentage of the total squared integral of the true density, $\int f^2(\mathbf{x})\mathrm{d}\mathbf{x}$, for convenience. This squared integral has a value of 0.01425 to 5 decimal places.

Similarly, we obtained mean ISE values for the models fitted to the simulated

Figure 2.5: Contour plot of the probability density function given in (2.12)



data sets by the fixed and adaptive kernel methods. The value for the adaptive kernel is slightly lower than for the fixed kernel, but the difference is small and unlikely to be indicative of any substantial difference in goodness of fit. However, the adaptive kernel clearly results in somewhat more variation of the ISE between different data sets, as the standard deviation is higher than that for the fixed kernel. The mixture model results in fitted models with a considerably lower mean ISE than either of the kernel methods, suggesting that it produces density estimates closer to the true simulation density.

Further support for the conclusion that the mixture model performs better

Table 2.3: Means and standard deviations (SD) of the ISE for fixed kernel, adaptive kernel, and bivariate normal mixture models fitted to data sets simulated from the mixture of bivariate normal distributions given in (2.12). Values are expressed as a percentage of the total squared integral of the true density of the simulated data sets.

Model	**Mean ISE %**	**ISE SD %**
Fixed kernel	11.0	4.1
Adaptive kernel	10.0	5.5
Normal mixture	4.6	3.0

than the kernels is provided by the frequencies of the models providing the best fit in terms of ISE for each of the 1000 data sets, as recorded in Table 2.4. This shows that the normal mixture model results in the lowest ISE for the great majority of cases (907 out of 1000 data sets). The fixed kernel method has the highest ISE, and thus provides the worst fit by this criterion, in 768 cases, which is a clear majority of the data sets. As such, the adaptive kernel is superior to the fixed kernel according to the ISE criterion.

Table 2.4: Number of data sets for which the fixed kernel, adaptive kernel and mixture of circular bivariate normal distributions produced the density estimate with the lowest and highest ISE, shown in the columns titled Lowest ISE and Highest ISE respectively. The 1000 data sets of 100 observations each were simulated from (2.12).

Model	**Lowest ISE**	**Highest ISE**
Fixed kernel	32	768
Adaptive kernel	61	183
Normal mixture	907	49

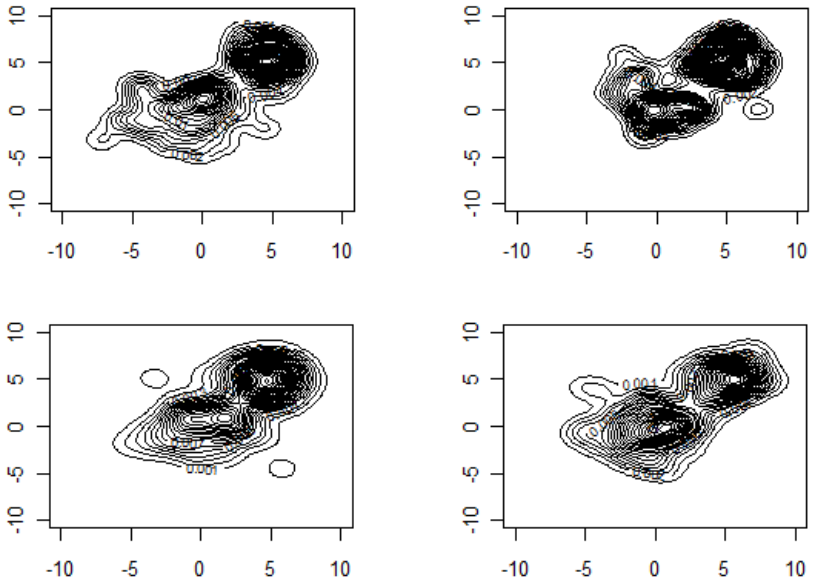
Example contour plots of density estimates obtained by the fixed kernel, adaptive kernel and mixture of circular bivariate normal distributions for 4 of the simulated data sets are displayed in Figures 2.6, 2.7 and 2.8 respectively. When the fixed kernel density estimates in Figure 2.6 are compared to Figure 2.5, it is apparent that while the basic shape of the true density has been recovered in each case, the fitted models are under-smoothed and suggest the presence of additional small modes that do not in fact exist. The densities of the fitted models also show substantial differences from one another.

The adaptive kernel density estimator has produced density estimates, presented in Figure 2.7, that are more heavily smoothed than those obtained using the fixed kernel and considerably closer to the true density. This is to be expected, as the adaptive kernel will generally result in greater smoothing of the outlying areas of the estimated density. The densities estimated by the adaptive kernel are also more similar to one another than those of the fixed kernel.

The densities fitted by the normal mixture model are extremely similar to the true density, and clearly provide a superior fit to either of the kernel density estimators. While there are slight differences between the plots, all of them are very similar to one another.

The plots shown in Figures 2.6 to 2.8 corroborate the results of the ISE calculations, as shown in Table 2.3, in suggesting that the normal mixture model

Figure 2.6: Fixed kernel density estimates for 4 data sets simulated from the density given in (2.12), with smoothing parameters chosen using least-squares cross-validation.



results in a better fit than the kernel density estimators. As the data sets are simulated from a mixture of circular bivariate normal distributions, they are extremely well suited to being modelled by mixtures of circular bivariate normal distributions, whereas for real data analysis this may not be the case. However, this is evidence that in some circumstances, a mixture model provides a superior density estimate for data to kernel methods.

Figure 2.7: Adaptive kernel density estimates for 4 data sets simulated from the density given in (2.12), with smoothing parameters chosen using least-squares cross-validation

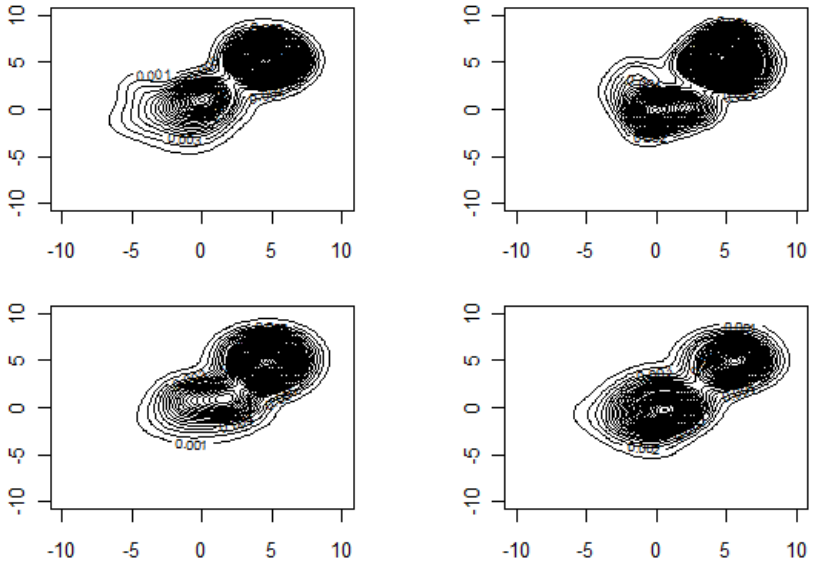
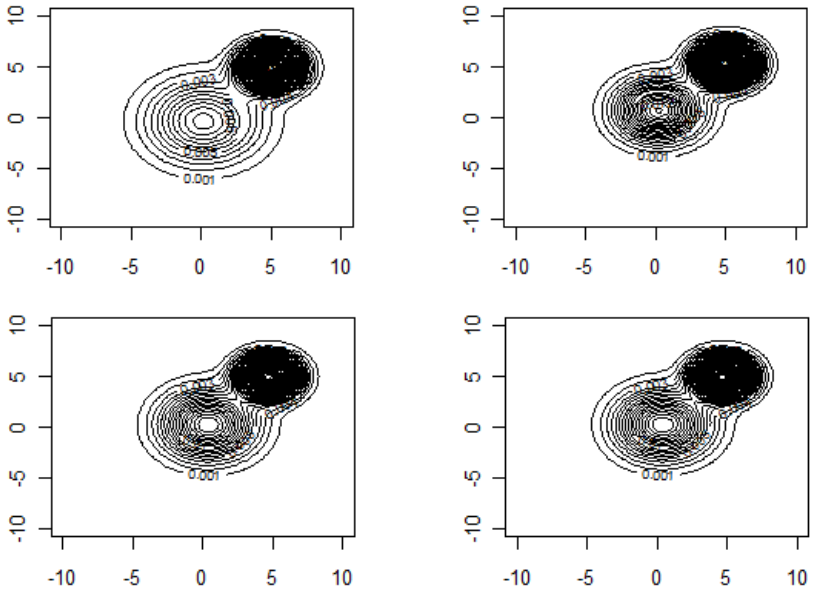


Figure 2.8: Fitted circular bivariate normal mixture models for 4 data sets simulated from the density given in (2.12), fitted using `mclust`.



2.4.3 Calculation for non-circular distributions

Throughout Section 2.4.1, the integrated squared error was calculated analytically for normal kernel and normal mixture models applied to data produced by a mixture of circular bivariate normal distributions, using (2.10) and (2.11). Here, the calculations are extended to the case of data generated using a mixture of k non-circular bivariate normal distributions. It is now assumed that the true probability density of the data takes the following form,

$$f(\mathbf{x}) = \sum_{i=1}^{k} w_i \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \sum_{i=1}^{k} w_i \mathrm{N}_i,$$

where the bivariate normal density at $\mathbf{x}$ for the ith mixture component, with mean vector and covariance matrix $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ respectively, is referred to as N_i as in Section 2.4.1. When a bivariate normal kernel is used to produce a density estimate for the data set, the resulting estimator can be expressed as shown in (2.8), and the ISE can therefore be written as

$$\sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j \int \mathrm{N}_i \mathrm{N}_j \mathrm{d}\mathbf{x} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \int \hat{\mathrm{N}}_i \hat{\mathrm{N}}_j \mathrm{d}\mathbf{x} - \frac{2}{n}\sum_{i=1}^{k}\sum_{j=1}^{n} w_i \int \mathrm{N}_i \hat{\mathrm{N}}_j \mathrm{d}\mathbf{x}.$$

It is therefore necessary to determine the integral of the product of two bivariate normal distribution functions, without the assumption of circularity made in Section 2.4.1.

Consider the bivariate normal densities $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \boldsymbol{\Omega})$ with mean vectors $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and $\boldsymbol{\nu} = (\nu_1, \nu_2)^T$ respectively, and covariance matrices

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$$

respectively. Multiplying the two densities together yields

$$\begin{aligned} \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \boldsymbol{\Omega}) = (2\pi)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \\ \times (2\pi)^{-1}|\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\nu})^T\boldsymbol{\Omega}^{-1}(\mathbf{x}-\boldsymbol{\nu})\right\}, \end{aligned}$$

which can be more simply expressed as

$$(4\pi^2)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left\{(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})+(\mathbf{x}-\boldsymbol{\nu})^T\boldsymbol{\Omega}^{-1}(\mathbf{x}-\boldsymbol{\nu})\right\}\right]. \tag{2.13}$$

Now note that when $\mathbf{x}$, $\mathbf{a}$ and $\mathbf{b}$ are $k \times 1$ vectors, and $\mathbf{A}$ and $\mathbf{B}$ are $k \times k$ symmetric matrices such that the inverse $(\mathbf{A}+\mathbf{B})^{-1}$ exists, then

$$\begin{aligned}&(\mathbf{x}-\mathbf{a})^T\mathbf{A}(\mathbf{x}-\mathbf{a})+(\mathbf{x}-\mathbf{b})^T\mathbf{B}(\mathbf{x}-\mathbf{b})\\ &\quad=(\mathbf{x}-\mathbf{c})^T(\mathbf{A}+\mathbf{B})(\mathbf{x}-\mathbf{c})+(\mathbf{a}-\mathbf{b})^T\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}(\mathbf{a}-\mathbf{b}),\end{aligned} \tag{2.14}$$

(Box and Tiao, 1973) where

$$\mathbf{c}=(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{A}\mathbf{a}+\mathbf{B}\mathbf{b}). \tag{2.15}$$

Application of (2.14) to the product of probability distributions in (2.13) gives the following.

$$\begin{aligned}&(4\pi^2)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left\{(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})+(\mathbf{x}-\boldsymbol{\nu})^T\boldsymbol{\Omega}^{-1}(\mathbf{x}-\boldsymbol{\nu})\right\}\right]\\ &\quad=(4\pi^2)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left\{(\boldsymbol{\mu}-\boldsymbol{\nu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1})^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\nu})\right\}\right]\\ &\qquad\times\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{c})^T(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1})(\mathbf{x}-\mathbf{c})\right\}\\ &\quad=(4\pi^2)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1}|^{-\frac{1}{2}}\\ &\qquad\times\exp\left[-\frac{1}{2}\left\{(\boldsymbol{\mu}-\boldsymbol{\nu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1})^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\nu})\right\}\right]\\ &\qquad\times|\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1}|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{c})^T(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1})(\mathbf{x}-\mathbf{c})\right\}\\ &\quad=(2\pi)^{-1}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\boldsymbol{\Omega}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1}|^{-\frac{1}{2}}\\ &\qquad\times\exp\left[-\frac{1}{2}\left\{(\boldsymbol{\mu}-\boldsymbol{\nu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}-\boldsymbol{\nu})\right\}\right]\\ &\qquad\times\mathrm{N}_{\mathbf{x}}\{\mathbf{c},(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1})^{-1}\},\end{aligned} \tag{2.16}$$

where

$$\mathbf{c}=(\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Omega}^{-1})^{-1}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}+\boldsymbol{\Omega}^{-1}\boldsymbol{\nu}).$$

The integral of (2.16) can be calculated as follows,

$$\begin{aligned}
&\int \Bigg((2\pi)^{-1} |\boldsymbol{\Sigma}^{-\frac{1}{2}}||\boldsymbol{\Omega}|^{-\frac{1}{2}} |\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^{-1}|^{-\frac{1}{2}} \\
&\quad \times \exp\left[-\frac{1}{2} \left\{ (\boldsymbol{\mu} - \boldsymbol{\nu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \boldsymbol{\nu}) \right\} \right] \\
&\quad \times \mathrm{N}_{\mathbf{x}}\{\mathbf{c}, (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^{-1})^{-1}\} \Bigg) \mathrm{d}\mathbf{x} \\
&= (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} |\boldsymbol{\Omega}|^{-\frac{1}{2}} |\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^{-1}|^{-\frac{1}{2}} \\
&\qquad \times \exp\left[-\frac{1}{2} \left\{ (\boldsymbol{\mu} - \boldsymbol{\nu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Omega}^{-1})^{-1} \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \boldsymbol{\nu}) \right\} \right],
\end{aligned}$$

as the bivariate normal distribution integrates to unity. Using the calculations above in combination with the formula given in (2.9), the ISE can be expressed as follows,

$$\begin{aligned}
\mathrm{ISE} = &\sum_{i=1}^{k} \sum_{j=1}^{k} \Bigg[\frac{w_i w_j}{2\pi} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_j|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1})^{-1} \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \right\} \Bigg] \\
&+ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{4\pi h^2} \exp\left\{ -\frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{4h^2} \right\} \\
&- \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n} \Bigg[\frac{w_i}{\pi h^2} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_i + (h^2 \mathbf{I})^{-1}|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu}_i - \mathbf{x}_j)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Sigma}_i^{-1} + (h^2 \mathbf{I})^{-1})^{-1} (h^2 \mathbf{I})^{-1} (\boldsymbol{\mu}_i - \mathbf{x}_j) \right\} \Bigg].
\end{aligned} \tag{2.17}$$

Equation (2.17) may be used to evaluate the ISE exactly. When an adaptive kernel density estimator is used in place of a fixed kernel, the smoothing parameter h is replaced with individual smoothing parameters h_i for each observation in the data set. The calculations detailed above are similar, resulting in the following ISE,

$$\begin{aligned}
\mathrm{ISE} = &\sum_{i=1}^{k} \sum_{j=1}^{k} \Bigg[\frac{w_i w_j}{2\pi} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_j|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j|^{-\frac{1}{2}} \\
&\quad \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1})^{-1} \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \right\} \Bigg]
\end{aligned}$$

$$+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2\pi(h_i^2+h_j^2)}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{x}_j)^T(\mathbf{x}_i-\mathbf{x}_j)}{2(h_i^2+h_j^2)}\right\}$$
$$-\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n}\left[\frac{w_i}{\pi h_j^2}|\boldsymbol{\Sigma}_i|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_i^{-1}+(h_j^2\mathbf{I})^{-1}|^{-\frac{1}{2}}\right.$$
$$\left.\times\exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_i-\mathbf{x}_j)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Sigma}_i^{-1}+(h_j^2\mathbf{I})^{-1})^{-1}(h_j^2\mathbf{I})^{-1}(\boldsymbol{\mu}_i-\mathbf{x}_j)\right\}\right].$$

If a mixture of bivariate normal distributions is used to produce a density estimate, the ISE can be calculated in a similar manner. Note that there is no requirement for the mixture components to be circular. Given such a mixture of m components, the estimated probability density function fitted by the model is

$$f(\mathbf{x})=\sum_{i=1}^{m}\hat{w}_i\mathrm{N}_{\mathbf{x}}(\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\Sigma}}_i)=\sum_{i=1}^{m}\hat{w}_i\hat{\mathrm{N}}_i,$$

where $\hat{w}_i$ is the estimated weight of the ith component, and $\hat{\mathrm{N}}_i=\mathrm{N}(\hat{\boldsymbol{\mu}}_i,\hat{\boldsymbol{\Sigma}}_i)$ is the normal density with the corresponding estimates for the mean and covariance matrix, which are $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ respectively. Following the same technique that led to (2.17), the ISE is therefore equal to

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\left[\frac{w_iw_j}{2\pi}|\boldsymbol{\Sigma}_i|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_j|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{\Sigma}_j^{-1}|^{-\frac{1}{2}}\right.$$
$$\left.\times\exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{\Sigma}_j^{-1})^{-1}\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)\right\}\right]$$
$$+\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\frac{\hat{w}_i\hat{w}_j}{2\pi}|\hat{\boldsymbol{\Sigma}}_i|^{-\frac{1}{2}}|\hat{\boldsymbol{\Sigma}}_j|^{-\frac{1}{2}}|\hat{\boldsymbol{\Sigma}}_i^{-1}+\hat{\boldsymbol{\Sigma}}_j^{-1}|^{-\frac{1}{2}}\right.$$
$$\left.\times\exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\mu}}_i-\hat{\boldsymbol{\mu}}_j)^T\hat{\boldsymbol{\Sigma}}_i^{-1}(\hat{\boldsymbol{\Sigma}}_i^{-1}+\hat{\boldsymbol{\Sigma}}_j^{-1})^{-1}\hat{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)\right\}\right]$$
$$-\sum_{i=1}^{k}\sum_{j=1}^{m}\left[\frac{w_i\hat{w}_j}{\pi}|\boldsymbol{\Sigma}_i|^{-\frac{1}{2}}|\hat{\boldsymbol{\Sigma}}_j|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_i+\hat{\boldsymbol{\Sigma}}_j^{-1}|^{-\frac{1}{2}}\right.$$
$$\left.\times\exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_i-\hat{\boldsymbol{\mu}}_j)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Sigma}_i^{-1}+\hat{\boldsymbol{\Sigma}}_j^{-1})^{-1}\hat{\boldsymbol{\Sigma}}_j^{-1}(\boldsymbol{\mu}_i-\hat{\boldsymbol{\mu}}_j)\right\}\right].$$

2.4.4 Model comparison for non-circular distributions

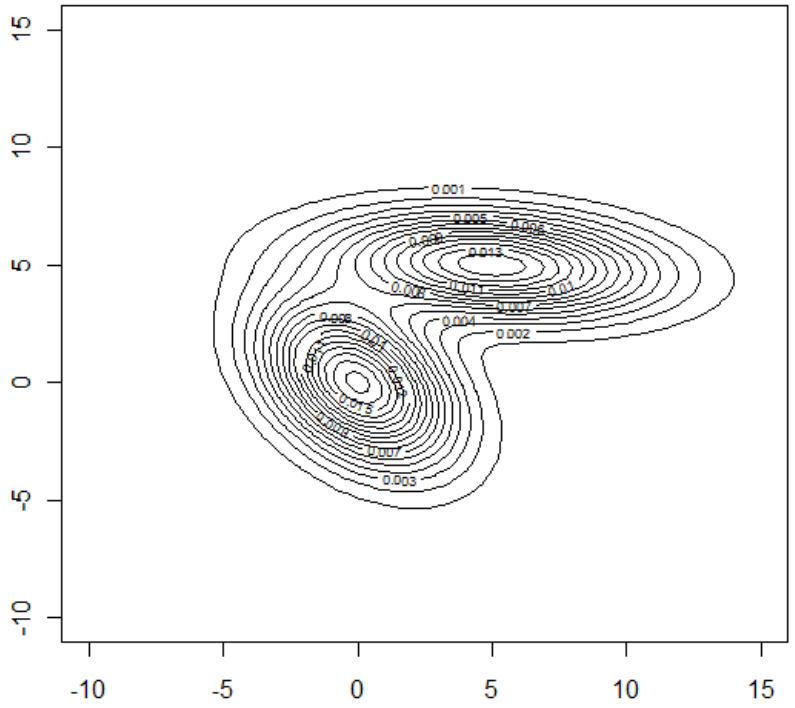
The simulation study in Section 2.4.2 used a simple mixture of circular bivariate normal distributions with 2 components as the simulation density. For the purpose of further comparison between kernel and mixture methods, a second

simulation study was conducted using simulations from a mixture of non-circular bivariate normal densities,

$$\frac{1}{2}\mathrm{N}\left\{(0,0)^T, \begin{pmatrix} 5 & -2 \\ -2 & 5 \end{pmatrix}\right\} + \frac{1}{2}\mathrm{N}\left\{(5,5)^T, \begin{pmatrix} 15 & -1 \\ -1 & 2 \end{pmatrix}\right\}. \tag{2.18}$$

A contour plot of the density in (2.18) is displayed in Figure 2.9 for reference.

Figure 2.9: Contour plot of the probability density function given in (2.18)



We simulated 1000 data sets of 100 observations each, and the underlying density of each data set was estimated using fixed and adaptive bivariate normal kernels and mixtures of bivariate normal distributions. However, the mixture model was not limited to circular components. Means and standard deviations of the ISE for each method are given in Table 2.5. They are each expressed as a percentage of the total squared integral of the true density, $\int f^2(\mathbf{x})\mathrm{d}\mathbf{x}$, for convenience. This squared integral has a value of 0.00833 to 5 decimal places.

Table 2.5: Means and standard deviations (SD) of the ISE for fixed kernel, adaptive kernel, and bivariate normal mixture models fitted to data sets simulated from the mixture of bivariate normal distributions given in (2.18). Values are expressed as a percentage of the total squared integral of the true density of the simulated data sets.

Model	**Mean ISE %**	**ISE SD %**
Fixed kernel	11.5	4.0
Adaptive kernel	12.3	5.4
Mixture of bivariate normal distributions	6.9	3.6

The models fitted to the simulated data sets by the fixed and adaptive normal kernel methods have similar mean ISE values, and the adaptive kernel results in ISE values with standard deviation higher than for the fixed kernel. The mixture model results in fitted models with lower mean ISE than either of the kernel methods, as was also the case for the simulation study in Section 2.4.2. This again suggests that the mixture model has produced density estimates closer to the true underlying density of the simulated data than either kernel method.

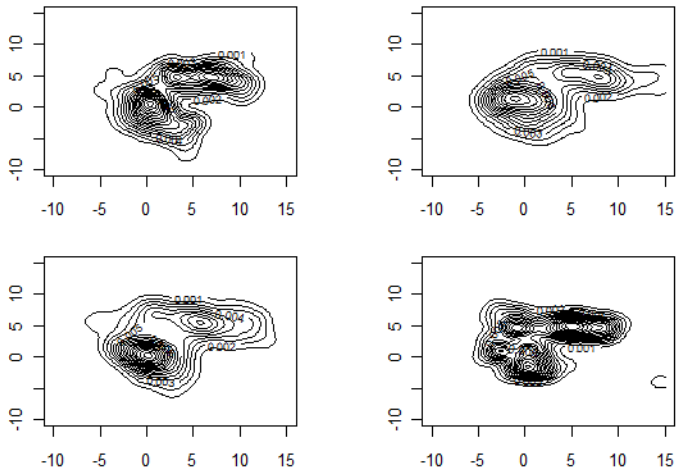
Table 2.6: Number of data sets for which the fixed kernel, adaptive kernel and mixture of circular bivariate normal distributions produced the density estimate with the lowest and highest ISE, shown in the columns titled Lowest ISE and Highest ISE respectively. The 1000 data sets of 100 observations each were simulated from (2.18).

Model	**Lowest ISE**	**Highest ISE**
Fixed kernel	107	342
Adaptive kernel	87	522
Normal mixture	806	136

The frequencies with which each model provided the best or worst fit to the simulated data sets in terms of ISE are given in Table 2.6. The normal mixture has the lowest ISE in the large majority of cases (806 out of 1000), providing further evidence that it gives superior density estimates to the kernel methods. The fixed kernel appears to produce slightly better results than the adaptive kernel, as the latter method gives the worst fit by the criterion of ISE in 522 out of 1000 cases. However, the difference in effectiveness between the kernel methods is small when compared to the difference between kernel and mixture methods.

Contour plots of density estimates obtained for 4 of the simulated data sets

Figure 2.10: Fixed kernel density estimates for 4 data sets simulated from the density given in (2.18), with smoothing parameters chosen using least-squares cross-validation.



using the fixed kernel, adaptive kernel and normal mixture are displayed in Figures 2.10, 2.11 and 2.12 respectively. The estimates obtained using the fixed kernel are undersmoothed and vary substantially between the different data sets, as was the case for data simulated from a mixture of circular bivariate normal distributions (Figure 2.6). The adaptive kernel results in density estimates that are more heavily smoothed and appear to improve upon the fixed kernel, but they still vary considerably in shape between different simulated data sets and are somewhat different from the true probability density of the data as seen in Figure 2.9. The mixture model produces estimates closer to the true density than either of the kernel methods.

In both of the simulation studies carried out in Section 2.4, the normal mixture model produces fitted density estimates with lower ISE than the fixed or adaptive kernel for the majority of simulated data sets. The mean ISE over all simulated data sets is also lower for the mixture, and plots of the density estimates show that the fitted models resemble the true density of the data more closely than those obtained using the kernel methods. The results of the studies therefore strongly indicate that the mixture model is more effective than the kernel density estimators for modelling the distributions used to generate the data.

Figure 2.11: Adaptive kernel density estimates for 4 data sets simulated from the density given in (2.18), with smoothing parameters chosen using least-squares cross-validation.

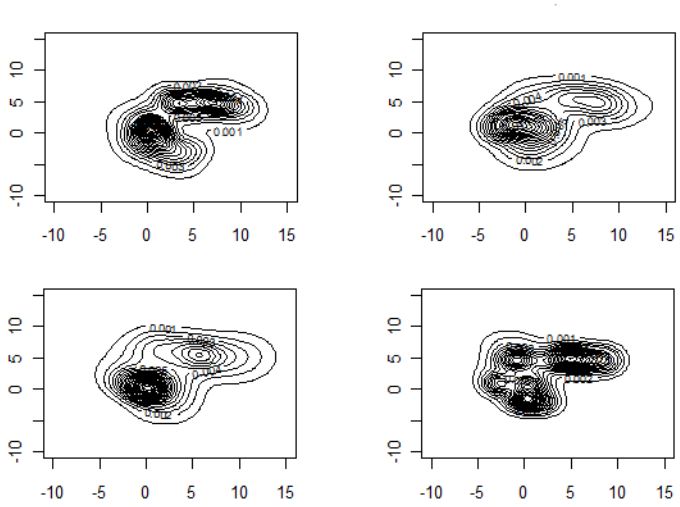
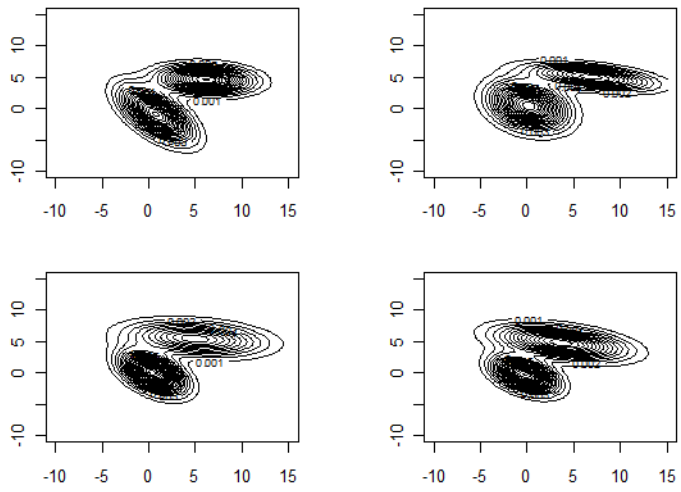


Figure 2.12: Fitted bivariate normal mixture models for 4 data sets simulated from the density given in (2.18), fitted using `mclust`.

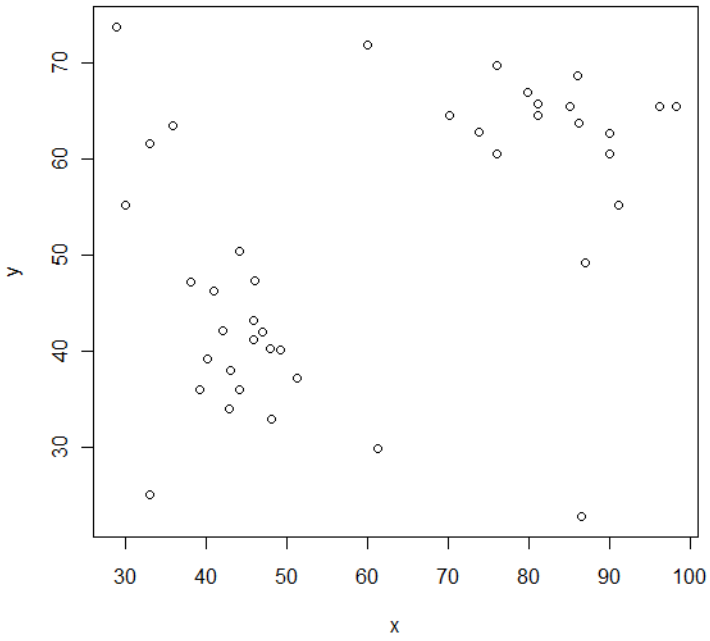


2.5 Application of kernel and mixture methods

A home range data set is taken from the ecology paper by Dixon and Chapman (1980), and consists of 41 points of location data for an individual juvenile female brush rabbit observed near Corvallis, Oregon. This data set will be used for kernel

and mixture modelling in this section. It is referred to henceforth as the DC data set, and is displayed in Figure 2.13. The DC data set is not easily represented by simple methods such as a bivariate normal distribution.

Figure 2.13: Plot of the DC data set. This data set consists of 41 location observations of a juvenile female brush rabbit observed near Corvallis, Oregon. Coordinates are in metres.



Also considered is the artificial data set used in Worton (1989), which consists of 100 points simulated from the following mixture of bivariate normal densities,

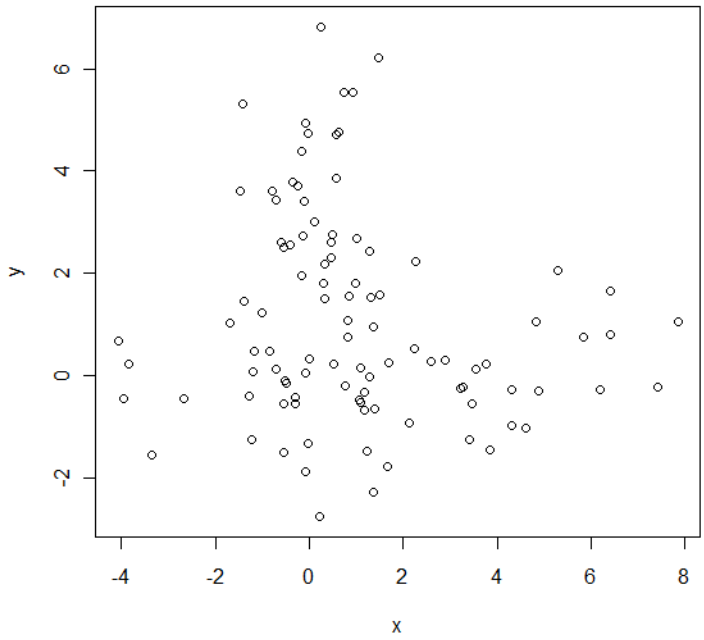
$$\frac{1}{2}\mathrm{N}\left\{(2.5,0)^T, \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}\right\} + \frac{1}{2}\mathrm{N}\left\{(0,2.5)^T, \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}\right\}. \tag{2.19}$$

It is referred to as the SIM data set, and is shown in Figure 2.14. The kernel and mixture methods that were applied to simulated data in Section 2.3 are used to model the DC and SIM data sets.

2.5.1 Fixed kernel method

The fixed kernel method was used to estimate the utilization densities for the DC and SIM data sets. The smoothing parameters were estimated using the ad-hoc

Figure 2.14: Plot of the SIM data set. This data set consists of 100 observations simulated from the mixture of bivariate normal distributions shown in (2.19)



approach detailed in Section 2.2.1, and also by least-squares cross-validation. The ad-hoc estimates were $h = 9.85$ for the DC data set and $h = 1.02$ for the SIM data set, while least-squares cross-validation gave $h = 4.66$ for the former and $h = 0.77$ for the latter. Contour plots of the fixed kernel density estimates using the smoothing parameters obtained by the ad-hoc and cross-validation methods are presented in Figures 2.15 and 2.16 respectively.

The density estimates for the SIM data set obtained using the ad-hoc approach and least-squares cross-validation appear similar to one another. It can also be seen that they have roughly the same shape as the true underlying density function, a contour plot of which is shown in Figure 2.17. The results for the DC data set in each case show 2 obvious modes, both of which are apparent from visual examination of the data set. However, the ad-hoc choice of h, as shown in Figure 2.15, results in substantially more smoothing than the value chosen using least-squares cross-validation, and has lost much of the fine detail of the original data. This is because the underlying density is clearly non-normal. The density estimate constructed using least-squares cross-validation, displayed in Figure 2.16, gives a better picture of the tails of the density function and

Figure 2.15: Contour plots of the fixed kernel density estimates for the DC and SIM data sets using ad-hoc smoothing parameters, shown in the left and right panels respectively.

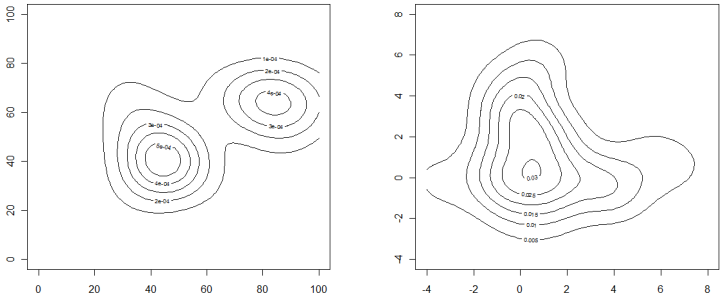
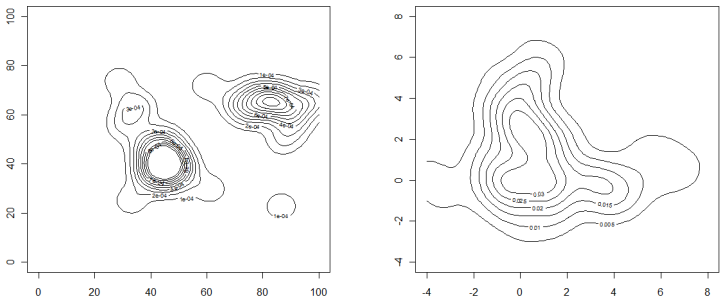


Figure 2.16: Contour plots of the fixed kernel density estimates for the DC and SIM data sets using least-squares cross-validation to choose the smoothing parameters, shown in the left and right panels respectively.

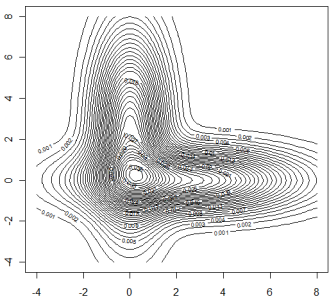


suggests the possible presence of 2 more modes. One of the modes is above the left cluster of data points. The other is removed from the larger modes and has coordinates close to $(80, 20)$. However, it is clear upon examination of the data set in Figure 2.13 that the further removed mode is caused by an outlying point and is not truly indicative of a feature of the density function.

2.5.2 Adaptive kernel method

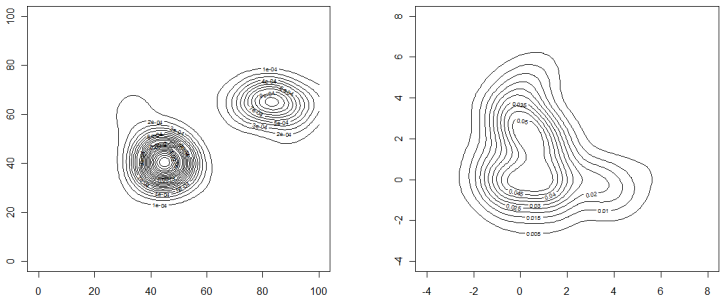
Density estimates were constructed for the DC and SIM data sets using the adaptive kernel method, with values for the smoothing parameters selected using least-squares cross-validation, giving $h = 5.74$ and $h = 0.79$ for the DC and SIM data sets respectively. The pilot estimates $\tilde{f}(\mathbf{x}_i)$ required by the adaptive kernel

Figure 2.17: Contour plot of the probability density function from which the SIM data set was generated.



were set equal to the fixed kernel density estimates at the points $\mathbf{x}_i$. Contour plots of the adaptive kernel density estimates are presented in Figure 2.18.

Figure 2.18: Contour plots of the adaptive kernel density estimates for the DC and SIM data sets using least-squares cross-validation to choose the smoothing parameter values, shown in the left and right panels respectively.



The adaptive kernel results in greater smoothing in areas with few data points than the fixed kernel, and is therefore less affected by outliers. Consequently, the adaptive kernel estimate for the DC data set in Figure 2.18 does not show the bump that was visible when the fixed kernel was used (in Figure 2.16), and gives a simpler picture of the shape of the underlying density. The tail of the contours around the left cluster of data points suggests the possibility of a third mode above the cluster, though not as strongly as the fixed kernel. For the SIM data, the density estimated using the adaptive kernel is considerably smoother than that obtained using the fixed kernel, and is closer to the true underlying density as shown by comparison with Figure 2.17.

2.5.3 Mixture of bivariate normal components

The `mclust` procedure introduced in Section 2.2.3 was used to fit mixtures of bivariate normal distributions to the SIM and DC data sets as an alternative to the kernel methods. For the SIM data set, the BIC values for different potential models indicate that the optimal normal mixture is a diagonal model with 2 components of equal volume and varying shape, referred to as model EVI. The models are referred to by the designations given to them by the MCLUST package; a reference key is included in Table 2.7. Each model has a designation of 3 letters, where the first, second and third letters describe the volume, shape and orientation of the mixture components respectively. The letter V means that the property varies between components, and E means that it is the same for all components. The letter I for the orientation signifies a diagonal model where the axes of the ellipsoidal density function are parallel to the x and y axes, and the letter I for both orientation and shape implies that the components are spherical. The BIC values for all potential models are tabulated in Table 2.8.

Table 2.7: Reference key for the mixtures of normal distributions used by the `mclust` package.

Model name	**Properties**
EII	Spherical, equal volume
VII	Spherical, unequal volume
EEI	Diagonal, equal volume and shape
VEI	Diagonal, varying volume, equal shape
EVI	Diagonal, equal volume, varying shape
VVI	Diagonal, varying volume and shape
EEE	Ellipsoidal, equal volume, shape and orientation
EEV	Ellipsoidal, equal volume and shape, varying orientation
VEV	Ellipsoidal, equal shape, varying volume and orientation
VVV	Ellipsoidal, varying volume, shape and orientation

The estimated density contour plot and point classification of the fitted model EVI are displayed in Figure 2.19. When compared to the contour plot of the true density in Figure 2.17, it is apparent that the fitted model recovers the true density well. Parameter estimates for the means and covariance matrices of the 2 components of the fitted model EVI are presented in Table 2.9, where the means are referred to as

$$\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12})^T, \quad \boldsymbol{\mu}_2 = (\mu_{21}, \mu_{22})^T,$$

Table 2.8: BIC values of mixtures of normal distributions fitted to the SIM data set using the `mclust` package. Table 2.7 provides a reference key for the models used.

Components	**EII**	**VII**	**EEI**	**VEI**	**EVI**
1	895.1	895.1	898.0	898.0	898.0
2	892.4	899.5	892.5	888.7	852.6
3	880.7	883.2	883.7	887.2	862.0
4	894.2	890.3	890.9	890.8	879.0
5	886.5	892.5	889.6	896.7	897.2
6	881.9	894.0	886.4	898.3	906.2
7	891.1	908.4	894.7	913.1	909.1
8	898.2	922.7	899.7	926.6	924.3
9	904.7	934.0	909.1	936.8	935.1

Components	**VVI**	**EEE**	**EEV**	**VEV**	**VVV**
1	898.0	900.2	900.2	900.2	900.2
2	853.7	897.0	856.7	857.1	860.8
3	867.6	887.0	864.6	873.7	882.5
4	888.9	895.2	880.7	886.8	899.7
5	905.4	892.2	894.4	906.1	923.9
6	924.0	885.4	906.7	919.4	941.7
7	932.8	888.4	897.0	923.9	948.2
8	952.3	900.8	914.9	939.2	967.1
9	968.3	905.8	931.6	NA	NA

and the covariance matrices as

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{12} \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} \sigma_{21} & 0 \\ 0 & \sigma_{22} \end{pmatrix}.$$

The estimates in Table 2.9 roughly correspond to the true values given in Section 2.5, based on the sample of 100 observations.

However, the SIM data set is more suitable for this analysis than real animal location data, as it is simulated from a mixture of bivariate normal densities. More information can be obtained by application to the DC data set, for which `mclust` indicates that the optimal choice using BIC as the selection criterion is an unconstrained ellipsoidal model with 3 components designated VVV, as detailed in Table 2.7. The BIC values for all potential models are given in Table 2.10.

A graphical representation of the optimal model is displayed in Figure 2.20, and it is clear that the presence of a small number of outlying points in the

Figure 2.19: Point classification and density contour plot for the mixture of normal distributions EVI fitted to the SIM data set by `mclust`, shown in the left and right panels respectively. The mixture EVI is the optimal selection using BIC.

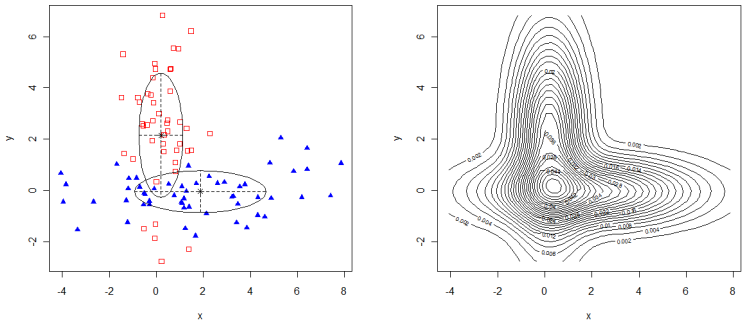


Table 2.9: Parameter estimates for the normal mixture model fitted to the SIM data set. The selected model using BIC is a mixture of 2 diagonal components with equal volume and different shape, designated EVI.

Parameters							
μ_{11}	μ_{12}	μ_{21}	μ_{22}	σ_{11}	σ_{12}	σ_{21}	σ_{22}
1.87	-0.05	0.19	2.15	7.78	0.66	0.88	5.87

data set has had disproportionate influence on the shape of the fitted density. This influence has given rise to difficulties in determining the correct number of mixture components, leading to the selection of a 3 component model for data more adequately represented by 2, with one of the components visible in the left panel of Figure 2.20 as a line joining 2 outliers with a point in the left cluster of observations. It is apparent upon visual examination of the data set that the presence of this component is dubious, and the outliers are exerting undue influence upon the model. As mentioned in Section 2.3, outliers such as this are common in home range data sets, and this behaviour will therefore occur frequently when using mixture models with normal components to analyse animal location data, indicating that care must be taken when using such models for this purpose.

Table 2.10: BIC values of mixtures of normal distributions fitted to the DC data set using the `mclust` package. Table 2.7 provides a reference key for the models used.

Components	EII	VII	EEI	VEI	EVI
1	720.4	720.4	716.6	716.6	716.6
2	682.2	682.0	681.9	685.2	693.0
3	681.7	668.8	683.3	672.2	680.2
4	670.6	674.5	672.1	681.0	686.5
5	672.0	685.1	675.7	688.8	675.8
6	678.0	693.8	682.6	697.3	680.7
7	687.5	NA	692.5	NA	NA
8	697.0	NA	700.0	NA	NA
9	691.1	NA	694.0	NA	NA

Components	VVI	EEE	EEV	VEV	VVV
1	716.6	710.5	710.5	710.5	710.5
2	686.1	680.7	679.7	677.2	680.9
3	686.5	678.4	670.5	675.2	647.2
4	674.6	673.5	676.7	683.9	647.8
5	681.0	678.5	685.8	694.6	654.2
6	701.6	686.3	698.3	700.3	653.5
7	710.4	696.2	707.3	710.7	NA
8	NA	696.7	707.8	NA	NA
9	NA	709.4	705.9	NA	NA

2.5.4 Mixture of bivariate t components

The mixture of normal distributions used by `mclust` is strongly influenced by the presence of outliers in the DC data set, as demonstrated in Section 2.5.3. The heavy-tailed bivariate t distributions used as the components of the procedures MIX-T-FIX and MIX-T-VAR introduced in Section 2.2.5 should be less susceptible to the influence of a small number of outlying points. MIX-T-FIX (with $\nu = 2$ for all components) and MIX-T-VAR were applied to the DC data set. In each case, mixtures of 1, 2, 3 and 4 t distributions were fitted, and BIC was used for model selection. The BIC values are tabulated in Table 2.11.

As shown in Table 2.11, the optimal model has 2 components whether MIX-T-FIX or MIX-T-VAR is used, and MIX-T-FIX is selected over MIX-T-VAR. Clearly, the excessive influence by the outlier that was noticeable when fitting a mixture of normal distributions is not present here. Density contour plots of the optimal models are shown in Figure 2.21. The density estimates obtained using

Figure 2.20: Point classification and density contour plot for the mixture of normal distributions VVV fitted to the DC data set by `mclust`, shown in the left and right panels respectively. The mixture VVV is the optimal selection using BIC.

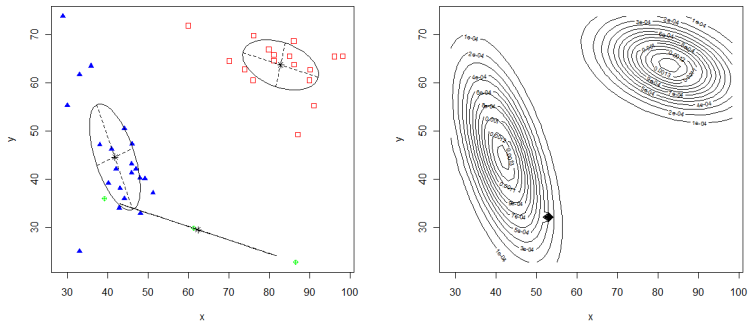


Table 2.11: BIC values for mixtures of t distributions fitted to the DC data set using MIX-T-FIX and MIX-T-VAR.

	Number of components			
	1	**2**	**3**	**4**
BIC for MIX-T-FIX	722.3	666.2	680.3	686.0
BIC for MIX-T-VAR	714.7	673.6	691.7	704.7

MIX-T-FIX and MIX-T-VAR appear similar to one another, and both fit the data well. The fact that the presence of the outlier does not have a strong effect on the shape of the density estimates suggests that mixtures of t distributions provide a substantial improvement over mixtures of normal distributions for modelling the DC data set.

One potential limitation of MIX-T-FIX and MIX-T-VAR is influence by the initial values used for the means of the component distributions. The initial values are chosen ad hoc based on visual inspection of the data set. For example, if MIX-T-FIX and MIX-T-VAR are used to fit a mixture of 3 t distributions to the DC data set, with the following initial values for the means of the distributions,

$$\boldsymbol{\mu}_1 = (40, 40)^T, \quad \boldsymbol{\mu}_2 = (70, 90)^T, \quad \boldsymbol{\mu}_3 = (30, 20)^T, \tag{2.20}$$

then the resulting density contour plots are as seen in Figure 2.22.

The fitted models in Figure 2.22 appear appropriate, and are not overly influenced by outlying points. Many other sets of initial values give rise to the

Figure 2.21: Density contour plots of the optimal mixtures of bivariate t distributions fitted to the DC data set by MIX-T-FIX and MIX-T-VAR, shown in the left and right panels respectively.

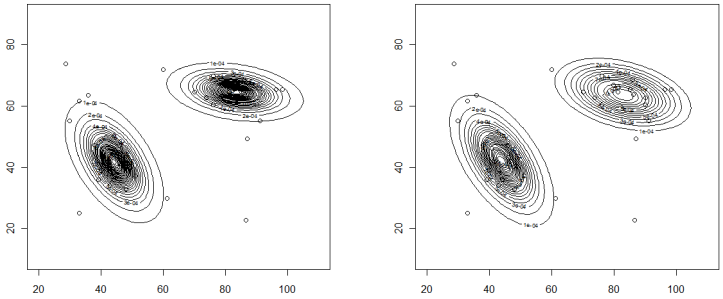
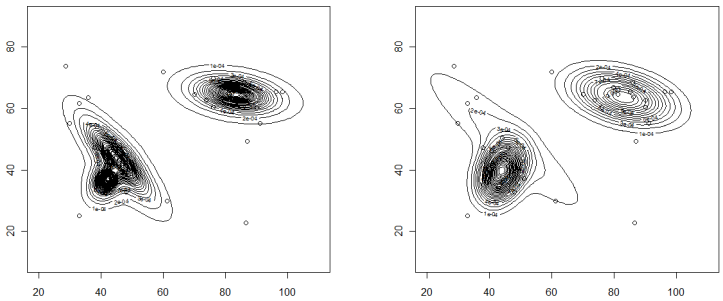


Figure 2.22: Density contour plots of a mixture of 3 bivariate t distributions fitted to the DC data set by MIX-T-FIX and MIX-T-VAR, shown in the left and right panels respectively. The initial values of the means of the mixture components were as given in (2.20).



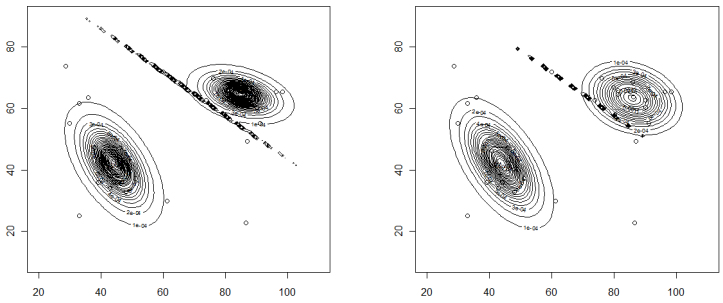
same fitted densities. However, there are initial values that will yield very different results. For example, if the following initial values are used,

$$\boldsymbol{\mu}_1 = (40, 40)^T, \quad \boldsymbol{\mu}_2 = (70, 90)^T, \quad \boldsymbol{\mu}_3 = (60, 10)^T, \tag{2.21}$$

then the density contour plots are as presented in Figure 2.23.

It is obvious upon visual examination that the fitted densities in Figure 2.22 are superior to those in Figure 2.23, but the BIC values are only very slightly better. It was found that initial values for the means of the mixture components that lie within the range of the data points resulted in fitted densities the same as those shown in Figure 2.22, and only certain choices that lie outwith this range

Figure 2.23: Density contour plots of a mixture of 3 bivariate t distributions fitted to the DC data set by MIX-T-FIX and MIX-T-VAR, shown in the left and right panels respectively. The initial values of the means of the mixture components were as given in (2.21).



will lead to fitted models such as that in Figure 2.23. As such, the use of a large number of different random starting points gives a density estimate corresponding to the global maximum log likehood.

MIX-T-FIX and MIX-T-VAR were used to fit mixtures of 1, 2, 3 and 4 bivariate t distributions to the SIM data set. The BIC values in Table 2.12 indicate that a 2 component model is optimal for MIX-T-FIX and MIX-T-VAR. The true probability density is known to be a mixture of 2 bivariate normal distributions, so the number of components selected is correct. Table 2.12 indicates that MIX-T-VAR is more appropriate than MIX-T-FIX as the BIC for the optimal 2 component model is superior.

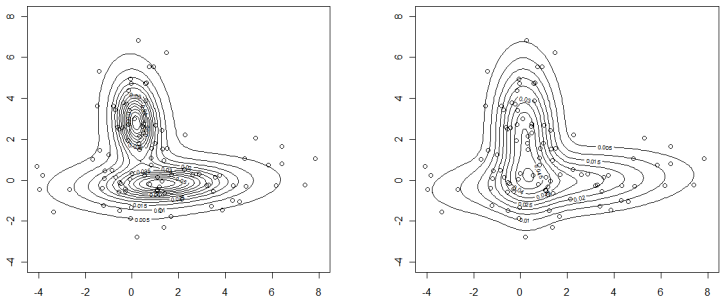
Table 2.12: BIC values for mixtures of t distributions fitted to the SIM data set using MIX-T-FIX and MIX-T-VAR.

	Number of components			
	1	**2**	**3**	**4**
BIC for MIX-T-FIX	925.0	887.0	904.3	923.6
BIC for MIX-T-VAR	905.0	870.5	896.6	919.3

Contour plots of the optimal densities fitted using MIX-T-FIX and MIX-T-VAR are displayed in Figure 2.24, and are similar to each other. Comparison to the true probability density in Figure 2.17 reveals that MIX-T-VAR produces a better fit than MIX-T-FIX, which is in agreement with model selection using BIC. This is to be expected, as for MIX-T-FIX the degrees of freedom of the

mixture components are set to $\nu = 2$, whereas the true density is a mixture of bivariate normal distributions which are better represented by higher values of ν.

Figure 2.24: Density contour plots of the optimal mixtures of bivariate t distributions fitted to the SIM data set by MIX-T-FIX and MIX-T-VAR, shown in the left and right panels respectively.



2.5.5 Discussion

The results obtained in Section 2.5 reveal that mixtures of bivariate normal distributions applied to animal home range data are susceptible to disproportionate influence from outlying data points. As such outliers are commonplace in home range data sets, this sensitivity to their presence is likely to result in frequent errors and suggests that bivariate normal mixtures are of limited suitability for the purpose of modelling such data. We therefore use mixtures with more robust component densities.

Mixtures of bivariate t distributions are not so influenced by outliers as the normal mixtures. In particular, the mixture of bivariate t distributions applied to the DC data set in Section 2.5.4 provided an appropriate density estimate, whereas the mixture of normal distributions returned a clearly erroneous fitted model with a spurious mixture component created by the presence of outliers. The simulation study conducted in Section 2.3 also demonstrated that mixtures of bivariate t distributions selected the most suitable model in the majority of cases, using BIC as the model selection criterion. The mixtures provide a robust approach for density estimation of home range data and compare favourably to kernel density estimators.

2.6 Bayesian analysis using mixture models

2.6.1 Prior distributions for the parameters of mixture models

In Section 2.5, mixtures of bivariate t distributions were shown to be a more robust alternative to mixtures of bivariate normal distributions for modelling data with outlying points. The parameters of the mixture components were determined using the EM algorithm, which chooses the parameter values using maximum likelihood estimation.

However, it is possible to adopt a Bayesian approach by introducing prior probability distributions for the parameters into the algorithm (Lin et al., 2004). Appropriate selection of priors has been shown to produce an appropriate posterior density estimate even when the number of components in the mixture is larger than the true number, as the extra components will be empty (Rousseau and Mengersen, 2011). The prior distributions for the means, covariance matrices and weights of the mixture components are selected in accordance with Lin et al. (2004).

Given a data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ modelled by fitting a mixture of bivariate t distributions with m components, the mean vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m$ of the components are given bivariate normal prior distributions,

$$\boldsymbol{\mu}_i \sim \mathrm{N}(\boldsymbol{\xi}, \boldsymbol{\kappa}^{-1}), \quad i = 1, \ldots, m,$$

where $\boldsymbol{\xi}$ is chosen so that each of its elements is the midpoint of the corresponding coordinates of the data, and $\boldsymbol{\kappa}$ is taken to be a diagonal matrix, with each of its elements equal to the inverse square range of the corresponding coordinates of the data. These specifications provide weak prior information for the $\boldsymbol{\mu}_i$s.

The scale matrices of the components, $\{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m\}$, are given Inverse Wishart prior distributions,

$$\boldsymbol{\Sigma}_i \sim \text{Inverse Wishart}(\rho, \boldsymbol{\Omega}), \quad i = 1, \ldots, m,$$

where Inverse Wishart$(\rho, \boldsymbol{\Omega})$ is an Inverse Wishart distribution with 2×2 scale matrix $\boldsymbol{\Omega}$ and ρ degrees of freedom. $\boldsymbol{\Omega}$ is taken to be the sample covariance matrix of the data, providing weak prior information, and ρ is set equal to 4, so that $\mathrm{E}(\boldsymbol{\Sigma}_i) = \boldsymbol{\Omega}$, $i = 1, \ldots, m$. The Inverse Wishart distribution is a probability distribution defined on real-valued positive definite matrices, with probability

density function

$$f(\mathbf{X}|\mathbf{\Psi},\nu) = \frac{|\mathbf{\Psi}|^{\frac{\nu}{2}}}{2^{\frac{\nu p}{2}}\Gamma_p\left(\frac{\nu}{2}\right)}|\mathbf{X}|^{-\frac{\nu+p+1}{2}}\exp\left\{-\frac{1}{2}\text{tr}(\mathbf{\Psi}\mathbf{X}^{-1})\right\},$$

where the random variable $\mathbf{X}$ is a $p \times p$ real-valued positive definite matrix, $\mathbf{\Psi}$ is the $p \times p$ positive definite scale matrix of the degrees of freedom, and ν is the degrees of freedom of the distribution. Γ_p is the real-valued multivariate gamma function

$$\Gamma_p(x) = \int_{\mathbb{G}} e^{-\text{tr}(S)}|S|^{x-\frac{1}{2}(p+1)}dS,$$

where $\mathbb{G}$ is the set of all $p \times p$ real-valued positive definite symmetric matrices S (James, 1964). The function $\Gamma_p(x)$ can also be expressed in terms of the gamma function as follows,

$$\Gamma_p(x) = \pi^{p(p-1)/4}\Pi_{i=1}^{p}\Gamma\left(x + \frac{1-i}{2}\right).$$

The Inverse Wishart distribution is commonly used as a conjugate prior for the covariance matrix of a multivariate normal distribution (Leonard and Hsu, 1999; Gelman et al., 2004; Carlin and Louis, 2008). Further discussion of the distribution's properties can be found in Box and Tiao (1973) and Press (1982). Here, we assume a mixture of bivariate t distributions and thus use $p = 2$. A Dirichlet prior distribution is used for the vector of weights of the components, $\mathbf{w} = (w_1, \ldots, w_m)^T$, such that

$$(w_1, \ldots, w_m) \sim \text{Dirichlet}(\alpha, \ldots, \alpha),$$

where $\text{Dirichlet}(\alpha, \ldots, \alpha)$ is the Dirichlet distribution on the space $\{(w_1, \ldots, w_m)|\sum_{i=1}^{m} w_i = 1\}$ with density function proportional to

$$w_1^{\alpha-1}w_2^{\alpha-1}, \ldots, w_m^{\alpha-1}.$$

The choice of $\alpha = 1$ gives a uniform prior over this space. Prior distributions are used for the degrees of freedom of the components, $\{\nu_1, \ldots, \nu_m\}$. Lin et al. (2004) used a continuous uniform distribution over a fixed range $[0, h]$ for large h, following the suggestions of Vounatsou and Smith (1997). While this is suitable when weak prior information is desired, the use of mixtures of bivariate t distributions in place of bivariate normal distributions in this particular context

arose from the need for increased robustness. Accordingly, a prior distribution is used that will weight the estimate towards lower degrees of freedom and thus attain the desired robustness. Specifically, a chi-squared distribution is used,

$$\nu_i \sim \chi^2_r,$$

where the degrees of freedom r can be assigned in accordance with the degree of robustness required. Defining the vector of degrees of freedom for the components as $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)^T$, the full prior probability density therefore takes the form

$$\begin{aligned}
p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}, \boldsymbol{\nu}) &= \prod_{i=1}^{m} f(\boldsymbol{\mu}_i) f(\boldsymbol{\Sigma}_i) f(w_i) f(\nu_i) \\
&= \prod_{i=1}^{m} \mathrm{N}(\boldsymbol{\xi}, \boldsymbol{\kappa}^{-1}) \text{Inverse Wishart}(\rho, \boldsymbol{\Omega}) \chi^2_r \\
&\propto \prod_{i=1}^{m} (2\pi)^{-1} |\boldsymbol{\kappa}|^{\frac{1}{2}} \left[\exp\left\{ -\frac{1}{2} (\boldsymbol{\mu}_i - \boldsymbol{\xi})^T \boldsymbol{\kappa} (\boldsymbol{\mu}_i - \boldsymbol{\xi}) \right\} \right. \\
&\qquad \times \frac{|\boldsymbol{\Omega}|^{\frac{\rho}{2}}}{2^{\rho} \Gamma_2\left(\frac{\rho}{2}\right)} |\boldsymbol{\Sigma}_i|^{-\frac{\rho+3}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}(\boldsymbol{\Omega} \boldsymbol{\Sigma}_i^{-1}) \right\} \\
&\qquad \left. \times \frac{1}{2^{\frac{r}{2}} \Gamma\left(\frac{r}{2}\right)} \nu_i^{\frac{r-2}{2}} e^{-\frac{\nu}{2}} \right].
\end{aligned}$$

Taking the logarithm of the above expression, substituting in $\rho = 4$, and adding it to the complete-data log-likelihood shown in (2.2) gives the following log posterior density,

$$\log p(\boldsymbol{\Psi}) = \log L_{1c}(\mathbf{w}) + \log L_{2c}(\boldsymbol{\nu}) + \log L_{3c}(\theta),$$

where

$$\begin{aligned}
\log L_{1c}(\mathbf{w}) &= \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} \log w_i, \\
\log L_{2c}(\boldsymbol{\nu}) &= \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} \left\{ -\log \Gamma\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2} \log\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2} (\log u_j - u_j) - \log u_j \right\} \\
&\qquad + \sum_{i=1}^{m} \left\{ -\frac{r}{2} \log 2 - \log \Gamma\left(\frac{r}{2}\right) + \left(\frac{r-2}{2}\right) \log \nu_i - \frac{\nu_i}{2} \right\},
\end{aligned}$$

$$\log L_{3c}(\theta) = \sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij}\left\{-\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}u_j(\mathbf{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)\right\}$$
$$-\frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{\mu}_i-\boldsymbol{\xi})^T\boldsymbol{\kappa}(\boldsymbol{\mu}_i-\boldsymbol{\xi}) - \frac{7}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{m}\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_i^{-1}).$$

The term θ is defined as in Section 2.2.3, u_j and z_{ij} are defined as in Section 2.2.4, and the E-step on the $(k+1)$th iteration of the EM algorithm is as detailed on pages 23 and 24. Following the same approach that was used without the prior distributions yields the following conditional expectation $Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(k)})$,

$$Q(\boldsymbol{\Psi}, \boldsymbol{\Psi}^{(k)}) = Q_1(\mathbf{w}, \boldsymbol{\Psi}^{(k)}) + Q_2(\boldsymbol{\nu}, \boldsymbol{\Psi}^{(k)}) + Q_3(\theta, \boldsymbol{\Psi}^{(k)}),$$

where

$$Q_1(\mathbf{w}, \boldsymbol{\Psi}^{(k)}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\tau_{ij}^{(k)}\log w_i,$$

with τ_{ij} defined as in Section 2.2.4, and ignoring terms not involving the ν_i,

$$Q_2(\boldsymbol{\nu}, \boldsymbol{\Psi}^{(k)}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\tau_{ij}^{(k)}\left\{-\log\Gamma\left(\frac{\nu_i}{2}\right) + \frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right)\right\}$$
$$+\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\tau_{ij}^{(k)}\nu_i}{2}\left\{\log u_{ij}^{(k)} - u_{ij}^{(k)} + \phi\left(\frac{\nu_i^{(k)}+2}{2}\right)\right.$$
$$\left.-\log\left(\frac{\nu_i^{(k)}+2}{2}\right)\right\}$$
$$+\sum_{i=1}^{m}\left\{\left(\frac{r-2}{2}\right)\log\nu_i - \frac{\nu_i}{2}\right\},$$
$$Q_3(\theta, \boldsymbol{\Psi}^{(k)}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\tau_{ij}^{(k)}\left\{-\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \log u_{ij}^{(k)}\right.$$
$$\left.-\frac{1}{2}u_{ij}^{(k)}(\mathbf{x}_j-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_j-\boldsymbol{\mu}_i)\right\}$$
$$-\frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{\mu}_i-\boldsymbol{\xi})^T\boldsymbol{\kappa}(\boldsymbol{\mu}_i-\boldsymbol{\xi}) - \frac{7}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{m}\operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_i^{-1}),$$

where the $u_{ij}^{(k)}$ terms are defined as in Section 2.2.4. The M-step results in the

following parameter estimates at the $(k+1)$th step (Lin et al., 2004),

$$w_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n}\tau_{ij}^{(k)}, \tag{2.22}$$

$$\boldsymbol{\mu}_i^{(k+1)} = \left\{\sum_{j=1}^{n}\tau_{ij}^{(k)}u_{ij}^{(k)}(\boldsymbol{\Sigma}_i^{(k)})^{-1} + \boldsymbol{\kappa}\right\}^{-1} \times \left\{(\boldsymbol{\Sigma}_i^{(k)})^{-1}\sum_{j=1}^{n}\tau_{ij}^{(k)}u_{ij}^{(k)}\mathbf{x}_j + \boldsymbol{\kappa}\boldsymbol{\xi}\right\}, \tag{2.23}$$

$$\boldsymbol{\Sigma}_i^{(k+1)} = \frac{\sum_{i=1}^{n}\tau_{ij}^{(k)}u_{ij}^{(k)}(\mathbf{x}_j - \boldsymbol{\mu}_i^{(k+1)})(\mathbf{x}_j - \boldsymbol{\mu}_i^{(k+1)})^T + \boldsymbol{\Omega}}{\sum_{j=1}^{n}\tau_{ij}^{(k)} + 7}. \tag{2.24}$$

The degrees of freedom are computed at each step by calculating the left hand side of the equations

$$\partial Q_2(\boldsymbol{\nu}; \boldsymbol{\Psi}^{(k)})/\partial\nu_i = 0,$$

and it follows that $\nu_i^{(k+1)}$ is a solution of the equation

$$\begin{aligned} &-\phi\left(\frac{\nu_i}{2}\right) + \log\left(\frac{\nu_i}{2}\right) + 1 + \frac{\sum_{i=1}^{n}\tau_{ij}^{(k)}(\log u_{ij}^{(k)} - u_{ij}^{(k)})}{\sum_{j=1}^{n}\tau_{ij}^{(k)}} + \phi\left(\frac{\nu_i^{(k)}+p}{2}\right) \\ &\quad - \log\left(\frac{\nu_i^{(k)}+p}{2}\right) + \frac{1}{\sum_{j=1}^{n}\tau_{ij}^{(k)}}\left(\frac{r-2}{\nu_i} - 1\right) = 0. \end{aligned} \tag{2.25}$$

The following procedure was created using R,

- MIX-T-BAYES: This procedure fits mixtures of bivariate t distributions to data while placing prior distributions on the parameters of the mixture components as detailed in this section, resulting in the iterative parameter estimates shown in (2.22), (2.23), (2.24) and (2.25).

and the R code is given in Appendix A. As the degrees of freedom of the components are estimated rather than fixed, MIX-T-BAYES is essentially equivalent to MIX-T-VAR introduced in Section 2.2.1 with the addition of the prior distributions for the parameters mentioned above.

2.6.2 Application to mule deer data

In this section, home range data sets are studied that consist of location observations of a female mule deer in the Piceance Basin in Western Colorado, radio-tracked in the summers of 1984 and 1985 (White and Garrott, 1990). The observations from 1984 and 1985 are referred to as the WG1 and WG2 data sets

respectively. Plots of the data are given in Figure 2.25. The observations are recorded in chronological order, but they are not taken at regular time intervals, and an examination of the time-ordered data, in Figure 2.26, does not display any obvious trends over time. Here it is assumed that the data set as a whole is time-independent.

Figure 2.25: Plots of the WG1 and WG2 data sets. These data sets consist of location observations of a female mule deer in the Piceance Basin, Western Colorado. The left panel shows observations recorded during the summer of 1984, and the right panel shows observations recorded during the summer of 1985. Coordinates are in metres.

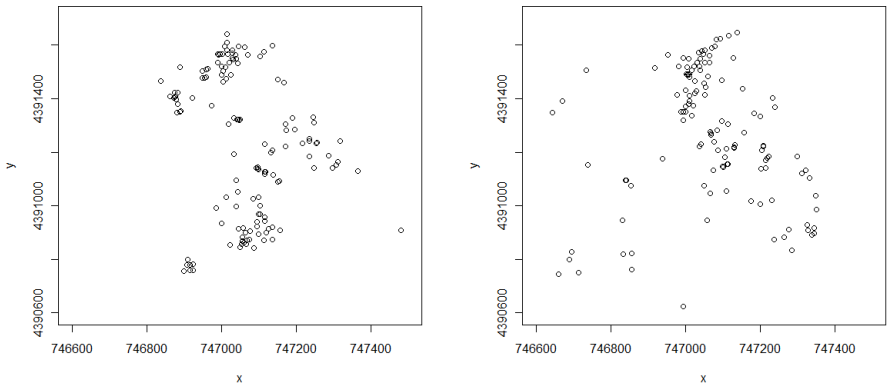
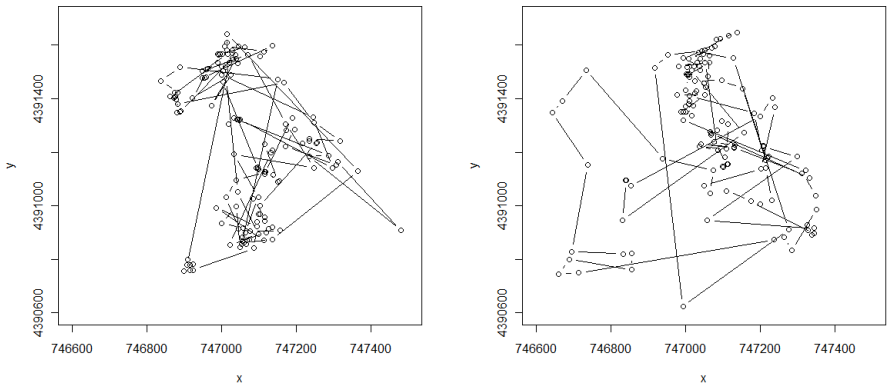


Figure 2.26: Time-ordered plots of the WG1 and WG2 data sets. These data sets consist of location observations of a female mule deer in the Piceance Basin, Western Colorado. The left panel shows observations recorded during the summer of 1984, and the right panel shows observations recorded during the summer of 1985.



Three mixture models were applied to the WG1 and WG2 data sets, the mixture of bivariate normal distributions fitted by `mclust`, MIX-T-VAR, and MIX-T-BAYES. Graphical representation of the models fitted using `mclust` are shown in Figures 2.27 and 2.28. The fitted mixture selected for the WG2 data set using BIC has 9 components, while the selected mixture for the WG1 data set has 8. The results indicate that large numbers of observations at in close proximity have resulted in unnecessary additional components, yielding an overly complicated fitted model that is of limited use.

Figure 2.27: Point classification and density contour plot for the mixture of normal distributions fitted to the WG1 data set by `mclust`, shown in the left and right panels respectively.

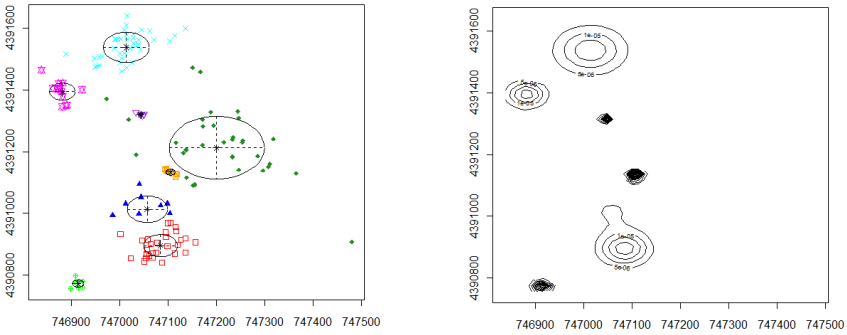
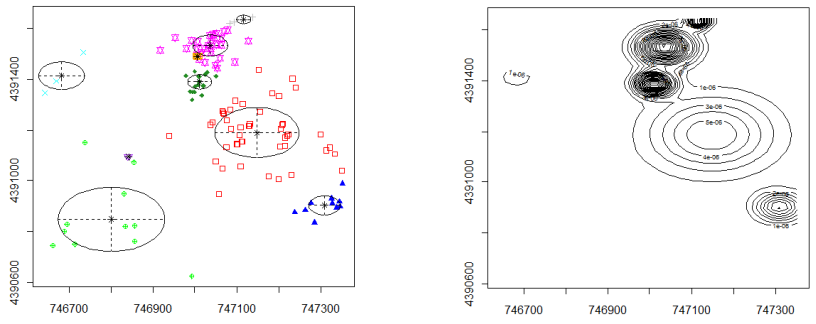


Figure 2.28: Point classification and density contour plot for the mixture of normal distributions fitted to the WG2 data set by `mclust`, shown in the left and right panels respectively.



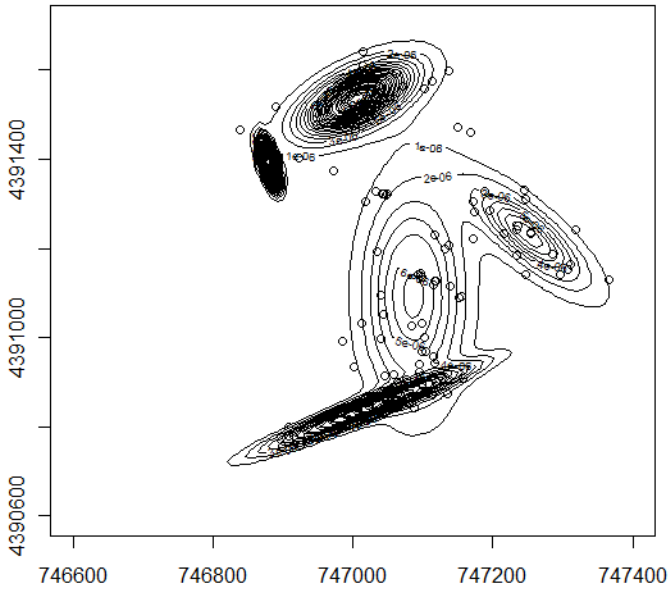
MIX-T-VAR was fitted to the WG1 data set, and the number of mixture components was selected by model comparison using BIC. Table 2.13 shows the

BIC values for mixtures of 1 to 6 components for the data. The optimal model has 5 components, suggesting that the use of the more robust t distributions in place of normal distributions has slightly alleviated the influence of small clusters of outliers. However, a contour plot of the density estimate, displayed in Figure 2.29, shows some signs of overfitting due to these clusters.

Table 2.13: BIC values for mixtures of t distributions fitted to the WG1 data set using MIX-T-VAR.

	Number of components					
	1	**2**	**3**	**4**	**5**	**6**
BIC	2082.7	1969.4	1958.5	1944.5	1932.0	1940.6

Figure 2.29: Density contour plot of the optimal mixture of bivariate t distributions fitted to the WG1 data set using MIX-T-VAR.



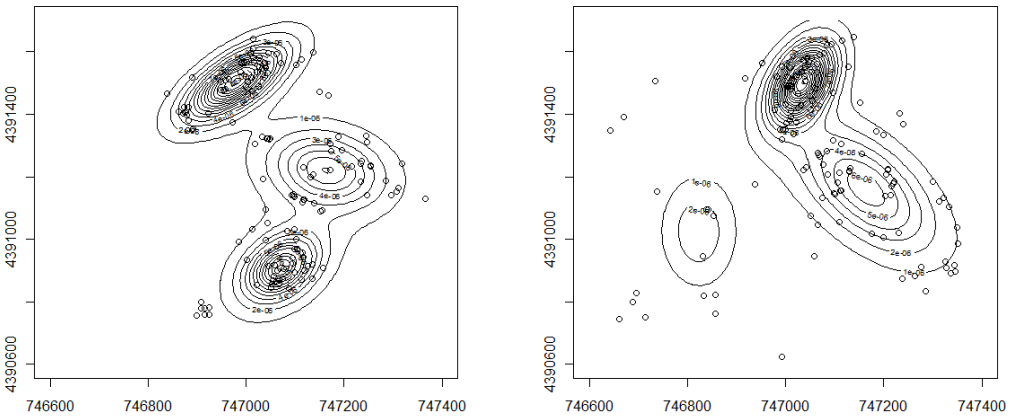
MIX-T-VAR was applied to the WG2 data set, but the parameter estimates did not successfully converge. However, MIX-T-BAYES was fitted to the WG1 and WG2 data sets without problems. The BIC values for the different numbers of mixture components are shown in Table 2.14, and it can be seen that the

optimal models for both data sets have 3 components. This is a reduction from the 5 components selected for the WG1 data set by MIX-T-VAR, and suggests that the overfitting observed in that case has been eliminated. Contour plots of the density estimates are shown in Figure 2.30, and the fitted model for the WG1 data set in particular appears to represent the data well. For the WG2 data set, the 2 larger modes of the fitted model match up well with the largest clusters of observations. All of the sparse observations towards the left of the animal's home range are represented as outliers of a single, smaller, mixture component. This appears adequate as a model.

Table 2.14: BIC values for mixtures of t distributions fitted to the WG1 and WG2 data sets using MIX-T-BAYES. For both data sets, the optimal model has 3 components.

	Number of components					
Data set	**1**	**2**	**3**	**4**	**5**	**6**
WG1	2092.7	1986.3	1976.4	1989.4	2005.7	2024.6
WG2	2010.5	1940.8	1925.9	1939.3	1957.7	1982.6

Figure 2.30: Density contour plots of the optimal mixtures of bivariate t distributions fitted to the WG1 and WG2 data sets using MIX-T-BAYES, shown in the left and right panels respectively.



2.7 Discussion

In this chapter, the effectiveness of kernel and mixture methods for modelling animal home range data is compared by their application to simulated and real data sets. The simulation studies in Section 2.4 compare the performance of fixed and adaptive kernel density estimators and mixtures of bivariate normal distributions using ISE as a criterion for model comparison. The results indicate that mixture models provide a more appropriate alternative to kernels for the simulation densities considered, as the fitted mixture models have lower ISE values than the kernel density estimates for most of the simulated data sets, and considerably lower mean ISE values across all simulations. The density estimates obtained using mixtures were more consistent between simulated data sets than those of the kernels.

As kernel methods are nonparametric and simply provide a description of the data rather than attempting to describe an underlying density, they can potentially model data regardless of its true distribution, but an accompanying drawback is that they can clearly produce substantially different density estimates for different data sets that are in fact simulated from the same density. With knowledge of the true underlying densities of the simulated data sets, it can be seen that the density estimates obtained using the fixed kernel are undersmoothed and suggest the presence of modes where none truly exist. The adaptive kernel density estimator resulted in density estimates with more heavily smoothed tails, but performed slightly worse than the fixed kernel in terms of ISE values, in that it had a higher mean ISE and also produced the density estimate with the highest ISE for more of the data sets.

The mixture of normal distributions performed well when applied to the simulated data sets in Section 2.3, and to the SIM data set in Section 2.5. However, application to real data with substantial departure from normality suggests that the density estimates obtained are susceptible to disproportionate influence from outlying data points. The mixture of normal distributions fitted to the DC data set using `mclust` in Section 2.5 was excessively influenced by outliers and produced a clearly erroneous fitted model with a third mixture component created by their presence. When applied to the WG1 and WG2 data sets, the normal mixture demonstrated both the aforementioned susceptibility to outliers and vulnerability to influence by observations in very close proximity to one another. The existence of small clusters of data points resulted in the presence of additional components in the fitted mixture model, giving unrealistic density estimates that failed to adequately describe the overall shape of the home range.

As outliers appear frequently in home range data sets, errors of this nature are likely to occur frequently when normal mixture models are fitted, suggesting that they are somewhat unsuitable.

The application of mixtures of bivariate t distributions to real and simulated data sets in Sections 2.3, 2.5 and 2.6 indicates that they are not so likely to be influenced by outlying points as mixtures of bivariate normal distributions. In particular, the results of the simulation study in Section 2.3 demonstrated that when mixtures of bivariate t distributions were fitted to data sets simulated from a mixture of 2 bivariate normal components with an added outlying point, the density estimates were not noticeably influenced by the outliers and the correct number of components was selected (using BIC as a selection criterion) in the great majority of cases. Simulated data sets generated from a mixture of t distributions were also modelled well. The estimated degrees of freedom for the fitted models in the fitted t distributions were frequently quite different from the true values, indicating that outliers were frequently assigned to a single component. However, the shape of the density estimates was generally very close to that of the true distribution. The mixtures of t distributions used to model the DC data set resulted in reasonable density estimates, in notable contrast with the normal mixture.

The introduction of prior distributions for the parameters of the t distributions successfully increases robustness to outliers, by assigning priors that weight the estimated degrees of freedom towards lower values. In Section 2.6, mixtures of t distributions fitted to the WG1 and WG2 data sets without priors produced a density estimate for the WG1 data set with components clearly shaped to an undesirable extent by small clusters of observations close together, although still superior to that obtained using the normal mixture model. The fitted mixture of normal distributions had 9 components, and included several small modes caused by small clusters of observations. The mixture of t distributions had 5 components, considerably fewer, but the shape of the density estimate was still strongly influenced by small groups of observations close together. The procedure failed to converge for the WG2 data set. However, a mixture of t distributions using the priors described in the text produced good density estimates for the WG1 and WG2 data sets, each with 3 components, and appears to provide a robust model for home range data.

Chapter 3

Fidelity

3.1 Introduction

The purpose of this chapter is to investigate different measures of fidelity in home range data, which estimate the fidelity of an animal to a particular site by measuring the overlap of its utilization densities estimated at different times. These measures are used to estimate the overlap between the WG1 and WG2 data sets introduced in Chapter 2. We consider several measures of overlap, namely the product measure, square root product measure, and OVL, and define estimators that can be used to obtain values for these measures of overlap between 2 data sets when the underlying densities are unknown. A simulation study is conducted to investigate the properties of the estimators, where the true densities of the simulated data sets are known and thus the estimated values of the overlap measures can be compared to the true values. Subsequently we derive expressions for the expectation and variance of the estimators as a function of the underlying density, and compare the expressions to the true overlap measures of various density functions. An approximation is obtained for the bias of the product measure, and we similarly compare the resulting approximate bias values to the true values for data sets simulated from various densities. We then apply the overlap measures to the WG1 and WG2 data sets. The density estimates used for the data sets are those obtained using the kernel and mixture models discussed in Chapter 2, and the OVL values obtained for the different models are used for model comparison.

3.2 Distance measures

3.2.1 Defining distance measures

The quantification of home range overlap for the same animal at different times, or between animals, is a subject of ongoing interest (Millspaugh et al., 2004; Fieberg and Kochanny, 2005; Dillon and Kelly, 2008; Frère et al., 2010), and there are several distance measures that can be used to assess the extent of the difference between 2 probability densities (Titterington et al., 1985). Each distance measure can be written as

$$\delta(f_1, f_2),$$

where f_1 and f_2 are either the probability density functions or the probability distribution functions of the two densities. The function δ is defined such that

$$\delta(f_1, f_2) \geq \delta(f_1, f_1) \text{ for any } f_1,\ f_2,$$

with equality only if $f_1(\mathbf{x}) = f_2(\mathbf{x})$ almost everywhere, and $\delta(f, f) = 0$ for all densities f. A number of different measures are provided in Titterington et al. (1985). We intend to apply these measures to the mixtures of t distributions used in Section 2.6. Here, we consider three different measures as described below.

The cumulative distribution function of a mixture of t distributions cannot be expressed in analytical form, and therefore some distance measures such as the L_2 norm with distribution functions cannot be calculated analytically. The L_2 norm with density functions can be used and has the form

$$\delta(f_1, f_2) = \int \{f_1(\mathbf{x}) - f_2(\mathbf{x})\}^2 \, \mathrm{d}\mathbf{x} \tag{3.1}$$

where f_1 and f_2 are the probability density functions of the two distributions. Note that the L_2 norm has a lower limit of 0 over all $\{f_1, f_2\}$, when $f_1(\mathbf{x}) = f_2(\mathbf{x})$ almost everywhere, and that

$$\{f_1(\mathbf{x}) - f_2(\mathbf{x})\}^2 = f_1^2(\mathbf{x}) + f_2^2(\mathbf{x}) - 2f_1(\mathbf{x})f_2(\mathbf{x}) \leq f_1^2(\mathbf{x}) + f_2^2(\mathbf{x}),$$

as $f_1(\mathbf{x})$, $f_2(\mathbf{x}) \geq 0$ everywhere. The L_2 norm for densities given in (3.1) therefore cannot exceed

$$\int \left\{f_1^2(\mathbf{x}) + f_2^2(\mathbf{x})\right\} \mathrm{d}\mathbf{x}, \tag{3.2}$$

and dividing (3.1) by (3.2) gives a scaled measure of fidelity that takes values between 0 and 1, with 0 representing complete fidelity and 1 signifying a complete lack of fidelity.

We refer to the integral

$$\int f_1(\mathbf{x}) f_2(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{3.3}$$

as the product measure of distance between the two densities. The product measure is directly related to the L_2-norm by the equation,

$$\int f_1(\mathbf{x}) f_2(\mathbf{x}) \mathrm{d}\mathbf{x} = \frac{1}{2}\left[\int f_1^2(\mathbf{x}) \mathrm{d}\mathbf{x} + \int f_2^2(\mathbf{x}) \mathrm{d}\mathbf{x} - \int \left\{f_1(\mathbf{x}) - f_2(\mathbf{x})\right\}^2 \mathrm{d}\mathbf{x}\right].$$

Dividing (3.3) by

$$\frac{\int \left\{f_1^2(\mathbf{x}) + f_2^2(\mathbf{x})\right\} \mathrm{d}\mathbf{x}}{2},$$

results in a scaled product measure which takes values between 0 and 1 (Morisita, 1959). The scaled product measure of two densities is equal to 0 if and only if their scaled L_2-norm obtained by dividing (3.1) by (3.2) is equal to 1, and vice versa.

Another distance measure we consider is the Hellinger distance (Beran, 1977), defined as

$$\delta(f_1, f_2) = \left[\int \left\{\sqrt{f_1(\mathbf{x})} - \sqrt{f_2(\mathbf{x})}\right\}^2 \mathrm{d}\mathbf{x}\right]^{\frac{1}{2}}.$$

The Hellinger distance can be scaled in the same manner as the L_2 norm, as it takes values between 0, when $\sqrt{f_1(\mathbf{x})} = \sqrt{f_2(\mathbf{x})}$ almost everywhere, and $\int \left\{f_1(\mathbf{x}) + f_2(\mathbf{x})\right\} \mathrm{d}\mathbf{x} = 2$. We refer to

$$\int \sqrt{f_1(\mathbf{x})} \sqrt{f_2(\mathbf{x})} \mathrm{d}\mathbf{x} \tag{3.4}$$

as the square root product measure of distance between the two densities (Bhattacharyya, 1943; Matusita, 1955). The square root product measure is related to the Hellinger distance by the equation,

$$\int \sqrt{f_1(\mathbf{x})}\sqrt{f_2(\mathbf{x})}\mathrm{d}\mathbf{x} = \frac{1}{2}\left[\int f_1(\mathbf{x})\mathrm{d}\mathbf{x} + \int f_2(\mathbf{x})\mathrm{d}\mathbf{x} - \int \left\{\sqrt{f_1(\mathbf{x})} - \sqrt{f_2(\mathbf{x})}\right\}^2 \mathrm{d}\mathbf{x}\right],$$

and is alternatively referred to as the Bhattacharyya measure of affinity. The square root product measure takes values between 0 and 1, and is equal to 1 when $f_1(\mathbf{x}) = f_2(\mathbf{x})$ are identical almost everywhere.

The overlapping coefficient or OVL (Weitzman, 1970; Inman and Bradley, 1989; Seidel, 1992), also referred to as the volume of intersection statistic or VI, is an alternative measure of overlap that takes the form

$$\int \min\left\{f_1(\mathbf{x}), f_2(\mathbf{x})\right\} \mathrm{d}\mathbf{x}. \tag{3.5}$$

Like the scaled distance measures defined above, the OVL takes values between 0 and 1. It is equal to 1 when $f_1(\mathbf{x}) = f_2(\mathbf{x})$ are identical almost everywhere, and equal to 0 when $\min\left\{f_1(\mathbf{x}), f_2(\mathbf{x})\right\} = 0$ almost everywhere (Clemons and Bradley, 2000).

The home range overlap proportions

$$HR_{i,j} = A_{i,j}/A_i,$$
$$HR_{j,i} = A_{i,j}/A_j$$

are the proportion of home range i that is overlapped by home range j and the proportion of home range j that is overlapped by i respectively (Kernohan et al., 2001). A_i and A_j are the respective areas of home ranges i and j, and $A_{i,j}$ is the area of overlap between home ranges i and j. The home range overlap proportions can be applied to animal movement data without requiring a density estimate. However, as the animals' utilization distributions are ignored, these measures may give high overlap estimates even when the parts of the animal's home ranges that overlap are infrequently used and so the probability of both animals occupying the same general area is small (Fieberg and Kochanny, 2005). The home range overlap proportions will not be further used in this chapter. In Section 3.3, we will use the other three distance measures described here to assess the site fidelity demonstrated by the mule deer observed in the WG1 and WG2 data sets defined in Section 2.6.2, and to assess the suitability of the mixtures of t distributions that were fitted to these data sets in Section 2.6.2 by comparing

the distance measures for the data to those calculated for simulations from the density estimates for the data.

3.2.2 Calculation of distance measures from density estimates

In Section 3.2.1, we introduced three distance measures, all of which require the calculation of integrals. Numerical calculation of these integrals is computationally expensive, and we would prefer to obtain values for the integrals analytically. Given two data sets, $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$, we can obtain a measure of overlap from the data by using the product measure (3.3) of the bivariate normal kernel density estimates obtained from the data sets, with smoothing parameters h_1 and h_2. The product measure can be calculated using a similar approach to that used to derive the ISE of kernel density estimators (2.17) in Section 2.4, and is given by the following expression,

$$\frac{1}{2\pi n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |h_1^2\mathbf{I}|^{-\frac{1}{2}} |h_2^2\mathbf{I}|^{-\frac{1}{2}} |(h_1^2\mathbf{I})^{-1} + (h_2^2\mathbf{I})^{-1}|^{-\frac{1}{2}} \\ \times \exp\left[-\frac{1}{2} (\mathbf{x}_i - \mathbf{y}_j)^T (h_1^2\mathbf{I})^{-1} \left\{ (h_1^2\mathbf{I})^{-1} + (h_2^2\mathbf{I})^{-1} \right\}^{-1} (h_2^2\mathbf{I})^{-1} (\mathbf{x}_i - \mathbf{y}_j) \right].$$

Similarly, if f_1 and f_2 are mixtures of k_1 and k_2 bivariate normal distributions respectively, such that

$$f_1(\mathbf{x}) = \frac{1}{2\pi} \sum_{i=1}^{k_1} w_i |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\},$$
$$f_2(\mathbf{x}) = \frac{1}{2\pi} \sum_{i=1}^{k_2} v_i |\boldsymbol{\Omega}_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\nu}_i)^T \boldsymbol{\Omega}_i^{-1} (\mathbf{x} - \boldsymbol{\nu}_i) \right\},$$

then the product measure can be calculated using the same methods and is equal to

$$\frac{1}{2\pi} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} w_i v_j |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} |\boldsymbol{\Omega}_j|^{-\frac{1}{2}} |(\boldsymbol{\Sigma}_i)^{-1} + (\boldsymbol{\Omega}_j)^{-1}|^{-\frac{1}{2}} \\ \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu}_i - \boldsymbol{\nu}_j)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Omega}_j^{-1})^{-1} \boldsymbol{\Omega}_j^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\nu}_j) \right\}.$$

We can therefore obtain product measures for certain density estimates without numerical integration. This approach relies on the use of bivariate normal kernel and mixture models and therefore is not applicable when using the mixtures of

bivariate t distributions discussed in Sections 2.5 and 2.6. It is, however, possible to estimate the value of the product measure of two densities as follows. We consider two data sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$, which are generated from distributions with probability density functions f_1 and f_2 respectively. Given respective density estimates $\hat{f}_1$ and $\hat{f}_2$ for f_1 and f_2, we define

$$\hat{I}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \hat{f}_2(\mathbf{x}_i),$$
$$\hat{I}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \hat{f}_1(\mathbf{y}_i).$$

We observe that

$$\mathrm{E}\left(\hat{I}_1\right) = \int f_1(\mathbf{x}) \hat{f}_2(\mathbf{x}) \mathrm{d}\mathbf{x},$$
$$\mathrm{E}\left(\hat{I}_2\right) = \int \hat{f}_1(\mathbf{y}) f_2(\mathbf{y}) \mathrm{d}\mathbf{y},$$

and that $\hat{I}_1$ and $\hat{I}_2$ are therefore unbiased estimators of these integrals. The weighted average of the two estimators,

$$\hat{I} = \frac{n_1 \hat{I}_1 + n_2 \hat{I}_2}{n_1 + n_2},$$

can be used as an estimator for (3.3). Under the assumptions that $\hat{f}_1$ and $\hat{f}_2$ are asymptotically unbiased estimators of f_1 and f_2 respectively, $\hat{I}$ is an asymptotically unbiased estimator of (3.3).

The estimators $\hat{I}_1$, $\hat{I}_2$ and $\hat{I}$ can be used in cases where the product measure cannot easily be calculated analytically, such as for the mixtures of bivariate t distributions. We then define

$$\begin{aligned}
\hat{J}_1 &= \frac{2\hat{I}_1}{\frac{1}{n_1}\sum_{i=1}^{n_1} \hat{f}_1(\mathbf{x}_i) + \frac{1}{n_2}\sum_{i=1}^{n_2} \hat{f}_2(\mathbf{y}_i)}, \\
\hat{J}_2 &= \frac{2\hat{I}_2}{\frac{1}{n_1}\sum_{i=1}^{n_1} \hat{f}_1(\mathbf{x}_i) + \frac{1}{n_2}\sum_{i=1}^{n_2} \hat{f}_2(\mathbf{y}_i)}, \\
\hat{J} &= \frac{n_1 \hat{J}_1 + n_2 \hat{J}_2}{n_1 + n_2} = \frac{2\hat{I}}{\frac{1}{n_1}\sum_{i=1}^{n_1} \hat{f}_1(\mathbf{x}_i) + \frac{1}{n_2}\sum_{i=1}^{n_2} \hat{f}_2(\mathbf{y}_i)},
\end{aligned} \tag{3.6}$$

as estimators for the scaled product measure.

Similar estimators can be used for the square root product measure and OVL. For the former distance measure, we can use

$$\hat{K}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\sqrt{\frac{\hat{f}_2(\mathbf{x}_i)}{\hat{f}_1(\mathbf{x}_i)}}, \quad \hat{K}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2}\sqrt{\frac{\hat{f}_1(\mathbf{y}_j)}{\hat{f}_2(\mathbf{y}_j)}},$$

and the weighted average

$$\hat{K} = \frac{n_1\hat{K}_1 + n_2\hat{K}_2}{n_1 + n_2}. \tag{3.7}$$

We observe that

$$\mathrm{E}(\hat{K}_1) = \int \frac{f_1(\mathbf{x})}{\sqrt{\hat{f}_1(\mathbf{x})}}\sqrt{\hat{f}_2(\mathbf{x})}\mathrm{d}\mathbf{x},$$
$$\mathrm{E}(\hat{K}_2) = \int \frac{f_2(\mathbf{y})}{\sqrt{\hat{f}_2(\mathbf{y})}}\sqrt{\hat{f}_1(\mathbf{y})}\mathrm{d}\mathbf{y}.$$

$\hat{K}_1$, $\hat{K}_2$ and $\hat{K}$ are of use as estimators of the square root product measure, and have expectation equal to (3.4) in the ideal case where $\hat{f}_1 = f_1$, $\hat{f}_2 = f_2$. For the OVL, we can use

$$\hat{L}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\min\left\{\frac{\hat{f}_2(\mathbf{x}_i)}{\hat{f}_1(\mathbf{x}_i)}, 1\right\},$$
$$\hat{L}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2}\min\left\{\frac{\hat{f}_1(\mathbf{y}_j)}{\hat{f}_2(\mathbf{y}_j)}, 1\right\},$$

(Schmid and Schmidt, 2006; Ridout and Linkie, 2009), and the weighted average

$$\hat{L} = \frac{n_1\hat{L}_1 + n_2\hat{L}_2}{n_1 + n_2}. \tag{3.8}$$

$\hat{L}_1$ and $\hat{L}_2$ have expectations equal to

$$\int \min\left\{\frac{\hat{f}_2(\mathbf{x})}{\hat{f}_1(\mathbf{x})}f_1(\mathbf{x}), f_1(\mathbf{x})\right\}\mathrm{d}\mathbf{x},$$

and

$$\int \min\left\{\frac{\hat{f}_1(\mathbf{y})}{\hat{f}_2(\mathbf{y})} f_2(\mathbf{y}), f_2(\mathbf{y})\right\} \mathrm{d}\mathbf{y},$$

respectively, which are both equal to (3.5) when $\hat{f}_1 = f_1$ and $\hat{f}_2 = f_2$. The estimators for the scaled product measure and OVL take values between 0 and 1, as do the true distance measures. However, the estimators $\hat{J}_1$ and $\hat{J}_2$ used for the square root product measure can take values greater than 1, which will occur if the density estimate for a data set has lower likelihood given the data than the density estimate obtained from the other data set.

3.3 Simulation study

3.3.1 Overview of study

We conducted a simulation study to assess the performance of the overlap estimators. Pairs of data sets were simulated from a given pair of probability distributions, one data set from each distribution. For each pair of simulated data sets, the estimators $\hat{J}$, $\hat{K}$ and $\hat{L}$) defined in (3.6), (3.7) and (3.8) were used to obtain estimates for the values of the three measures of overlap. Density estimates were fitted using fixed kernel estimators, mixtures of bivariate normal distributions using `mclust`, and mixtures of bivariate t distributions using the Bayesian procedure MIX-T-BAYES defined in Section 2.6. Results for the scaled product measure, scaled square root product measure and OVL are presented in Tables 3.1, 3.2 and 3.3 respectively.

We used ad-hoc smoothing parameter selection for the kernel density estimators, as this method of selection is asymptotically optimal. As an alternative, we used a plug-in bandwidth selector in accordance with the direct plug-in (DPI) rule (Wand and Jones, 1994). This entails the selection of a 2×2 matrix of smoothing parameters $\mathbf{H}$, chosen to minimize the asymptotic mean squared error (AMISE) of the kernel density estimator K, which is given as a function of $\mathbf{H}$ by

$$\mathrm{AMISE}\left\{\hat{f}(\mathbf{H})\right\} = n^{-1}|\mathbf{H}|^{-\frac{1}{2}} R(K) + \frac{1}{4}\mu_2(K)^2 \int \mathrm{tr}^2\left\{\mathbf{H} H_f(\mathbf{x})\right\} \mathrm{d}\mathbf{x},$$

where

$$R(K) = \int K^2(\mathbf{x})\mathrm{d}\mathbf{x},$$
$$\mu_2(K) = \int \mathbf{x}^2 K(\mathbf{x})\mathrm{d}\mathbf{x},$$

and H_f is the Hessian matrix of f. The R package `ks` (Duong, 2007) can fit kernel density estimates to data using a plug-in bandwidth selector.

3.3.2 Identical bivariate normal distributions

The initial study included 200 pairs of data sets of 100 observations each, all simulated from a standard bivariate normal distribution. As the data sets were all simulated from the same distribution, the true values of the scaled product measure, scaled square root product measure and OVL are all 1. Density estimates were obtained for each simulated data set using kernel density estimators with asymptotically optimal and plug-in smoothing parameters, mixtures of bivariate normal distributions, and mixtures of bivariate t distributions. Summary statistics for the scaled product measure, scaled square root product measure and OVL are presented in Tables 3.1, 3.2 and 3.3 respectively.

Bias in the estimates is to be expected, as for all three of the distance measures an estimated value equal to the true value of 1 will occur only if the two data sets simulated from the two distributions are identical. For the scaled product measure, as shown in Table 3.1, the mean and median values appear appropriate, with the mixture models giving estimates closer to the true values than the kernels. For the square root product measure (Table 3.2), the mixtures also produce more accurate estimates than the kernels, and the estimates given by both methods are closer to the true values than for the scaled product measure. The OVL (Table 3.3) is more heavily biased than the other two distance measures, and in contrast to the results in Tables 3.1 and 3.2, the mixture models do not generally produce more accurate estimates than kernels. The mixture of bivariate t distributions gives estimates further from the true values than the kernel density estimator with asymptotically optimal smoothing parameter, and the mixture of normal distributions improves on the kernel slightly but to a lesser extent than for the other distance measures.

The results given here suggest that the square root product measure is most suitable for estimating the overlap when the true extent of overlap is very high, and that the OVL is the least suitable of the three distance measures in such

cases. For the square root product measure and OVL, the estimated values obtained using the kernel methods have higher standard deviation than those for the mixture models. For the scaled product measure, there is no substantial difference between the two types of model.

Table 3.1: Summary statistics of the values of the estimated scaled product measures $\hat{J}$ shown in (3.6) obtained for 200 pairs of data sets of 100 observations each simulated from a standard bivariate normal distribution. Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.91	0.89
Median	0.92	0.89
SD	0.03	0.03
Bias	−0.09	−0.11
Lower 5% point	0.86	0.83
Upper 5% point	0.95	0.94

	Normal mixture	t **mixture**
Mean	0.98	0.98
Median	0.99	0.99
SD	0.03	0.03
Bias	−0.02	−0.02
Lower 5% point	0.95	0.95
Upper 5% point	1.00	1.00

Table 3.2: Summary statistics of the values of the estimated square root product measures $\hat{K}$ shown in (3.7) obtained for 200 pairs of data sets of 100 observations each simulated from a standard bivariate normal distribution. Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.93	0.92
Median	0.94	0.93
SD	0.01	0.02
Bias	−0.07	−0.08
Lower 5% point	0.91	0.89
Upper 5% point	0.95	0.95

	Normal mixture	t **mixture**
Mean	0.99	0.99
Median	0.99	0.99
SD	0.01	0.01
Bias	−0.01	−0.01
Lower 5% point	0.98	0.98
Upper 5% point	1.00	1.00

Table 3.3: Summary statistics of the values of the estimated OVLs $\hat{L}$ shown in (3.8) obtained for 200 pairs of data sets of 100 observations each simulated from a standard bivariate normal distribution. Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.82	0.80
Median	0.82	0.80
SD	0.03	0.04
Bias	−0.18	−0.20
Lower 5% point	0.77	0.74
Upper 5% point	0.87	0.86

	Normal mixture	t **mixture**
Mean	0.83	0.80
Median	0.83	0.80
SD	0.02	0.03
Bias	−0.17	−0.20
Lower 5% point	0.79	0.75
Upper 5% point	0.86	0.83

3.3.3 Bivariate normal distributions with coordinate shift

A simulation study similar to the one in Section 3.3.2 was carried out using bivariate normal distributions with densities f_1 and f_2, where

$$f_1 \text{ is } \mathrm{N}\{(0,0)^T, \mathbf{I}\}, \quad f_2 \text{ is } \mathrm{N}\{(1,0)^T, \mathbf{I}\}. \tag{3.9}$$

We simulated 200 pairs of data sets of 100 observations each, with one data set in each pair simulated from f_1 and the other from f_2. In this case, the true scaled product measure is equal to $\exp\left(-\frac{1}{4}\right) \approx 0.78$, and the true square root product measure and OVL were numerically calculated to be 0.88 and 0.62 respectively. The estimators $\hat{J}$, $\hat{K}$ and $\hat{L}$ were again used to estimate the three distance measures, and density estimates were fitted using the same methods used for the previous simulations.

Results for the scaled product measure, scaled square root product measure and OVL are presented in Tables 3.4, 3.5 and 3.6 respectively. For the product measure (Table 3.4), the mixture models result in mean and median estimated values considerably closer to the true values than those produced by the kernels. The results for the square root product measure (Table 3.5) closely resemble those obtained previously for two identical bivariate normal distributions. The means and medians of the estimated values of the product measure for the two mixture models are extremely close to the true value, particularly for the mixture of t distributions, while the corresponding values for the kernel density estimators have a slight negative bias. For the OVL (Table 3.6), the mean and median of the estimated OVL values are close to the true OVL for all estimators, particularly for the kernel with asymptotically optimal smoothing parameter.

In the case of two standard bivariate normal distributions considered in Section 3.3.1, we found that the estimator for the square root product measure gave the most accurate estimates, followed by the scaled product measure and finally the OVL. Here, this is also the case for the two mixture models. However, for the kernels, the OVL produces the least biased estimates, and the square root product measure the most. We observe that the estimates obtained for all three distance measures are closer to the true values than in Section 3.3.1.

Table 3.4: Summary statistics of the values of the estimated scaled product measures $\hat{J}$ shown in (3.6) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate normal distributions shown in (3.9). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.73	0.71
Median	0.73	0.71
SD	0.06	0.06
Bias	−0.05	−0.07
Lower 5% point	0.62	0.60
Upper 5% point	0.83	0.81

	Normal mixture	t **mixture**
Mean	0.77	0.78
Median	0.77	0.78
SD	0.07	0.07
Bias	−0.01	0.00
Lower 5% point	0.60	0.61
Upper 5% point	0.87	0.88

Table 3.5: Summary statistics of the values of the estimated square root product measures $\hat{K}$ shown in (3.7) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate normal distributions shown in (3.9). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.82	0.81
Median	0.82	0.81
SD	0.03	0.04
Bias	−0.06	−0.07
Lower 5% point	0.77	0.76
Upper 5% point	0.87	0.87

	Normal mixture	**t mixture**
Mean	0.88	0.88
Median	0.88	0.88
SD	0.03	0.03
Bias	0.00	0.00
Lower 5% point	0.83	0.83
Upper 5% point	0.93	0.93

Table 3.6: Summary statistics of the values of the estimated OVLs $\hat{L}$ shown in (3.8) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate normal distributions shown in (3.9). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.61	0.60
Median	0.61	0.60
SD	0.05	0.05
Bias	−0.01	−0.02
Lower 5% point	0.51	0.51
Upper 5% point	0.69	0.68

	Normal mixture	t **mixture**
Mean	0.65	0.63
Median	0.65	0.63
SD	0.04	0.04
Bias	0.03	0.02
Lower 5% point	0.59	0.56
Upper 5% point	0.71	0.68

3.3.4 Mixture of t distributions I

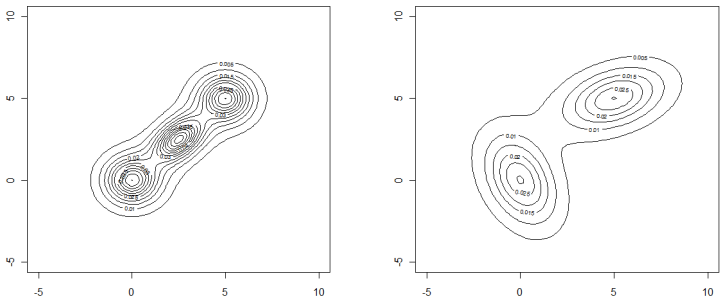
We considered a more complex case, using two mixtures of bivariate t distributions. Here, the two distributions used were defined by

$$f_1 \text{ is } \frac{1}{3}t_4\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right\}+\frac{1}{3}t_4\left\{\begin{pmatrix}2.5\\2.5\end{pmatrix},\begin{pmatrix}1&0.5\\0.5&1\end{pmatrix}\right\}$$
$$+\frac{1}{3}t_4\left\{\begin{pmatrix}5\\5\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right\}, \tag{3.10}$$

$$f_2 \text{ is } \frac{1}{2}t_4\left\{\begin{pmatrix}5\\5\end{pmatrix},\begin{pmatrix}4&1\\1&2\end{pmatrix}\right\}+\frac{1}{2}t_4\left\{\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}2&-1\\-1&4\end{pmatrix}\right\}. \tag{3.11}$$

Density contour plots of f_1 and f_2 are displayed in Figure 3.1. Unlike the densities used in Section 5.3.1 and 5.3.2, the true scaled product measure for these two densities cannot be expressed analytically, and numerical integration gives a value of 0.77. Numerical integration gives values of 0.89 for the square root product measure and 0.65 for the OVL. Summary statistics for the scaled product measure, square root product measure and OVL are presented in Tables 3.7, 3.8 and 3.9 respectively.

Figure 3.1: Density contour plots of the mixtures of t distributions defined as f_1 and f_2 in (3.10) and (3.11), shown in the left and right panels respectively.



For the scaled product measure (Table 3.7), the estimates obtained for the kernel with asymptotically optimal selection of the smoothing parameter have the lowest bias, while those for the kernel with plug-in smoothing parameter have by far the highest. All estimates except those using the plug-in kernel are reasonably close to the true value. We observe that the estimates for the mixture models have slightly higher standard deviations than those for the kernels.

The estimates of the square root product measure (Table 3.8) for the kernel

method with asymptotically optimal smoothing parameter and the mixture of t distributions have lower bias than the estimates using the normal mixture and plug-in kernel. In the case of the normal mixture model, the median of the estimates is close to the true value but the mean of the estimates was found to be extremely large. This value for the mean is considered to be erroneous and is not displayed. The upper 5% point of the estimated values is much smaller than the mean, indicating that the large mean values are caused by a small number of cases in which the estimated square root product measure is extremely large. Further examination of the contribution of individual data points to the value of the estimated square root product measure reveals that these results are caused by the presence of outliers in one data set which are very close to a mode of the density estimate for the other data set. This effect is not observed for the mixture of bivariate t distributions as, due to the heavier tails of the mixture components, the value of the density estimate at such outliers will not be nearly as small, and therefore $\hat{K}$ will not be so large. The kernel density estimators similarly have heavier tails than the mixture of normal distributions and avoid this problem, although as mentioned previously the estimate obtained using plug-in kernel is more heavily biased than the estimate using the mixture of t distributions.

In Table 3.9, the plug-in kernel results in mean and median estimated OVL values closest to the true OVL of the four methods, and the corresponding estimates for the mixture of t distributions are also close to the true value. The normal mixture model has more strongly underestimated the OVL, while the kernel density using the asymptotically optimal selection of h has overestimated it to an even greater extent. Again, the mixture of bivariate t distributions is evidently a more reliable model for the estimation of overlap than the normal mixture.

As shown in Tables 3.7, 3.8 and 3.9, the estimates of overlap for the mixture of t distributions and the kernel with asymptotically optimal smoothing parameter have low bias for the scaled product measure, square root product measure and OVL, whereas the kernel with plug-in smoothing parameter produces considerably more biased estimates for the scaled product measure and square root product measure, and the estimates obtained using the mixture of normal distributions in the case of the square root product measure are extremely biased. None of the distance measure estimators is less biased than the others for all estimators used. The OVL estimates have the lowest bias when the mixture of t distributions or the plug-in kernel is used, the square root product measure estimates for the kernel with asymptotically optimal smoothing parameter, and the scaled product measure estimates for the mixture of normal distributions.

Table 3.7: Summary statistics of the values of the estimated scaled product measures $\hat{J}$ shown in (3.6) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate t distributions shown in (3.10) and (3.11). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.79	0.62
Median	0.80	0.62
SD	0.05	0.05
Bias	0.02	−0.15
Lower 5% point	0.70	0.54
Upper 5% point	0.87	0.70

	Normal mixture	t **mixture**
Mean	0.74	0.73
Median	0.74	0.74
SD	0.08	0.07
Bias	−0.03	−0.04
Lower 5% point	0.61	0.60
Upper 5% point	0.89	0.85

Table 3.8: Summary statistics of the values of the estimated square root product measures $\hat{K}$ shown in (3.7) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate t distributions shown in (3.10) and (3.11). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.88	0.81
Median	0.88	0.81
SD	0.02	0.03
Bias	−0.01	−0.08
Lower 5% point	0.83	0.75
Upper 5% point	0.91	0.86

	Normal mixture	t **mixture**
Mean	NA	0.91
Median	0.90	0.91
SD	NA	0.03
Bias	NA	0.02
Lower 5% point	0.81	0.86
Upper 5% point	16.71	0.96

Table 3.9: Summary statistics of the values of the estimated OVLs $\hat{L}$ shown in (3.8) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate t distributions shown in (3.10) and (3.11). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.69	0.65
Median	0.69	0.64
SD	0.04	0.05
Bias	0.04	−0.01
Lower 5% point	0.62	0.56
Upper 5% point	0.77	0.73

	Normal mixture	t **mixture**
Mean	0.61	0.64
Median	0.61	0.64
SD	0.05	0.03
Bias	−0.04	−0.01
Lower 5% point	0.52	0.59
Upper 5% point	0.60	0.69

3.3.5 Mixture of t distributions II

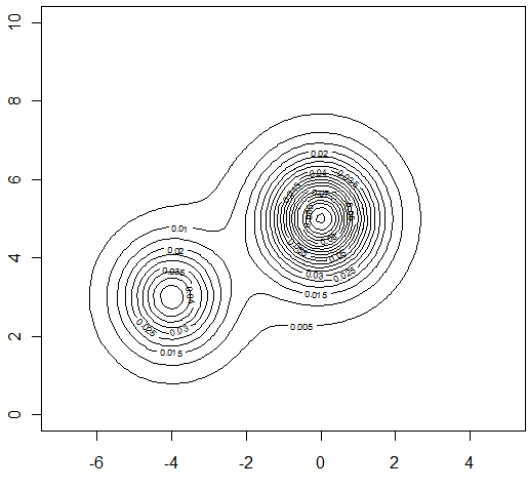
We considered a pair of mixtures of t distributions with densities f_1 and f_2 where

$$f_1 \text{ is } \frac{1}{3}t_4\left\{\begin{pmatrix} -4 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\} + \frac{2}{3}t_4\left\{\begin{pmatrix} 0 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}, \tag{3.12}$$

$$f_2 \text{ is } \frac{1}{2}t_4\left\{\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}\right\} + \frac{1}{2}t_4\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix}\right\}. \tag{3.13}$$

Here, f_1 is the same density used in Section 3.3.4, and a density contour plot of f_2 is displayed in Figure 3.2. The extent of overlap is much lower than in Section 3.3.4, as numerical integration gives values of 0.13 for the scaled product measure, 0.46 for the square root product measure and 0.18 for the OVL. Summary statistics for the scaled product measure, square root product measure and OVL are presented in Tables 3.10, 3.11 and 3.12 respectively.

Figure 3.2: Density contour plot of the mixture of t distributions defined as f_2 in (3.13).



The mixture of t distributions gives estimates for the scaled product measure (Table 3.10) and square root product measure (Table 3.11) with lower bias than the kernel methods. The mixture of normal distributions also performs well using the scaled product measure, but its estimates of the square root measure are clearly affected by the same problem as described in Section 3.3.4, as the mean value was inappropriately large and is not displayed. The OVL estimates

for both mixture models (Table 3.12) are more biased than those for the kernels. The kernel with plug-in smoothing parameter returns an OVL closest to the numerically calculated value.

In general, the estimator $\hat{L}$ for the OVL appears to be most accurate in cases of moderate overlap as seen in Sections 3.3.3 and 3.3.4, whereas for high or low levels of overlap, as in Sections 3.3.2 and 3.3.5, the estimates are considerably more biased. For the scaled product measure, there is not such an obvious relationship between the extent of overlap and the accuracy of the estimator $\hat{J}$. The estimates obtained for the two different bivariate normal distributions in Table 3.4, where the true value of the scaled product measure is 0.78, are less biased than the estimates for the two identical bivariate normal distributions in Table 3.1. However, the estimates for the two mixtures of t distributions in Section 3.3.4 (Table 3.7), with a similar true value of 0.77, are considerably more biased for all modelling methods other than the kernel with asymptotically optimal smoothing parameter. For mixtures of t distributions in Section 3.3.5 (Table 3.10), which have much lower overlap and a true scaled product measure of 0.13, the estimates obtained using the fixed kernel with asymptotically optimal smoothing parameter are more biased than in Section 3.3.4. The estimates produced by the other three methods are less biased, but those for the mixture models are more biased than the values obtained in Section 3.3.3. For the square root product measure there is similarly no obvious relationship between overlap and accuracy of the estimator $\hat{K}$.

Inspection of the bias values in the tables throughout Section 3.3 indicates that no one measure is more accurate than the others over all of the simulated data sets. However, the estimated OVL is more biased than the estimated scaled product measure and square root product measure when overlap is extremely high as seen in Tables 3.1 to 3.3, and the estimated square root product measure can fail to give a reasonable value when the estimates are obtained from fitted mixtures of normal distributions, as seen in Tables 3.8 and 3.11. This behaviour is not exhibited by the kernels or the mixtures of t distributions due to their heavier tails.

Table 3.10: Summary statistics of the values of the estimated scaled product measures $\hat{J}$ shown in (3.6) obtained for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate t distributions shown in (3.12) and (3.13). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.20	0.16
Median	0.29	0.16
SD	0.05	0.04
Bias	0.07	0.03
Lower 5% point	0.13	0.10
Upper 5% point	0.28	0.24

	Normal mixture	t **mixture**
Mean	0.14	0.11
Median	0.13	0.11
SD	0.05	0.04
Bias	0.01	−0.02
Lower 5% point	0.07	0.06
Upper 5% point	0.24	0.19

Table 3.11: Summary statistics of the values of the estimated square root product measures $\hat{K}$ shown in (3.7) for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate t distributions shown in (3.12) and (3.13). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.42	0.35
Median	0.42	0.35
SD	0.05	0.06
Bias	−0.04	−0.11
Lower 5% point	0.34	0.27
Upper 5% point	0.51	0.44

	Normal mixture	t **mixture**
Mean	NA	0.44
Median	0.45	0.44
SD	NA	0.06
Bias	NA	−0.02
Lower 5% point	0.34	0.34
Upper 5% point	0.93	0.55

Table 3.12: Summary statistics of the values of the estimated OVLs $\hat{L}$ shown in (3.8) for 200 pairs of data sets of 100 observations each, simulated from the 2 bivariate t distributions shown in (3.12) and (3.13). Presented summary statistics include the mean, median, standard deviation (SD), bias and upper and lower 5% points. Fitted models include kernel density estimators with plug-in and asymptotically optimal (asy opt) smoothing parameter selection, and mixtures of bivariate normal and bivariate t distributions.

	Kernel (asy opt)	**Kernel (plug-in)**
Mean	0.21	0.17
Median	0.21	0.17
SD	0.04	0.03
Bias	0.03	−0.01
Lower 5% point	0.15	0.11
Upper 5% point	0.26	0.23

	Normal mixture	t **mixture**
Mean	0.25	0.28
Median	0.24	0.28
SD	0.05	0.04
Bias	0.07	0.10
Lower 5% point	0.16	0.22
Upper 5% point	0.35	0.34

3.4 Expectation and variance of the product measure

In this section, we calculate the expectation and variance of the product measure of two data sets simulated from known probability distributions f_1 and f_2 respectively, where $\hat{f}_1$ and $\hat{f}_2$ are density estimates for f_1 and f_2 respectively. We note that

$$\begin{aligned}\int f_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x} &= \int \left[\mathrm{E}\{\hat{f}_1(\mathbf{x})\} - \mathrm{Bias}\{\hat{f}_1(\mathbf{x})\}\right] \hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x} \\ &\approx \int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x} - \int \mathrm{Bias}\{\hat{f}_1(\mathbf{x})\}\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}.\end{aligned}$$

In the case where $\hat{f}_1$ and $\hat{f}_2$ are bivariate normal kernel density estimators with smoothing parameters h_1 and h_2 respectively, obtained from respective data sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$, we recall that the integral of the product of two circular bivariate normal distributions $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})$ is

$$\int \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x} = \frac{1}{2\pi(\sigma^2+\tau^2)} \exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)}\right\},$$

as shown in Section 2.4. It follows that

$$\begin{aligned}&\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x} \\ &\quad= \int \frac{1}{n_1}\sum_{i=1}^{n_1} \mathrm{N}_{\mathbf{x}}(\mathbf{x}_i, h_1^2\mathbf{I}) \frac{1}{n_2}\sum_{j=1}^{n_2} \mathrm{N}_{\mathbf{x}}(\mathbf{y}_j, h_2^2\mathbf{I})\mathrm{d}\mathbf{x} \\ &\quad= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \int \mathrm{N}_{\mathbf{x}}(\mathbf{x}_i, h_1^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\mathbf{y}_j, h_2^2\mathbf{I})\mathrm{d}\mathbf{x} \\ &\quad= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \hat{g}(\mathbf{x}_i, \mathbf{y}_j),\end{aligned}$$

where

$$\hat{g}(\mathbf{x}_i, \mathbf{y}_j) = \frac{1}{2\pi(h_1^2+h_2^2)} \exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{2(h_1^2+h_2^2)}\right\}.$$

When we consider the data sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$ as random variables with densities f_1 and f_2 respectively rather than known values, the expectation of the product measure of the density estimates for an unknown pair of simulated data sets is given by

$$\mathrm{E}\left\{\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\} = \mathrm{E}\left\{\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}.$$

In the simple case where f_1 and f_2, the true underlying densities of $\{\mathbf{x}_1,\ldots,\mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1,\ldots,\mathbf{y}_{n_2}\}$ respectively, are independent circular bivariate normal densities $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, given by $\mathrm{N}(\boldsymbol{\mu},\sigma^2\mathbf{I})$ and $\mathbf{N}(\boldsymbol{\nu},\tau^2\mathbf{I})$ respectively, the above expectation can be expressed as

$$\begin{aligned}\mathrm{E}\left\{\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\} &= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathrm{E}\left\{\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}\\ &= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int\int \hat{g}(\mathbf{x}_i,\mathbf{y}_j)f_{\mathbf{x},\mathbf{y}}(\mathbf{x}_i,\mathbf{y}_j)\mathrm{d}\mathbf{x}_i\mathbf{y}_j,\end{aligned}$$

where $f_{\mathbf{x},\mathbf{y}}$ is the joint probability density function of a pair of observations, one generated from f_1 and one from f_2. As f_1 and f_2 are independent, $f_{\mathbf{x},\mathbf{y}} = f_1 f_2$, and so the expectation above is equal to

$$\begin{aligned}&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int \hat{g}(\mathbf{x}_i,\mathbf{y}_j)f_1(\mathbf{x}_i)f_2(\mathbf{y}_j)\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j\\ &\quad= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int\int\left[\frac{1}{2\pi(h_1^2+h_2^2)}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{2(h_1^2+h_2^2)}\right\}\right.\\ &\qquad\times\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{(\mathbf{x}_i-\boldsymbol{\mu})^T(\mathbf{x}_i-\boldsymbol{\mu})}{2\sigma^2}\right\}\\ &\qquad\left.\times\frac{1}{2\pi\tau^2}\exp\left\{-\frac{(\mathbf{y}_j-\boldsymbol{\nu})^T(\mathbf{y}_j-\boldsymbol{\nu})}{2\tau^2}\right\}\right]\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j,\end{aligned}$$

which is equal to

$$\begin{aligned}&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int\int \mathrm{N}_{\mathbf{x}_i}\{\mathbf{y}_j,(h_1^2+h_2^2)\mathbf{I}\}\mathrm{N}_{\mathbf{x}_i}(\boldsymbol{\mu},\sigma^2\mathbf{I})\mathrm{N}_{\mathbf{y}_j}(\boldsymbol{\nu},\tau^2\mathbf{I})\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j\\ &\quad= \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int \mathrm{N}_{\mathbf{y}_j}(\boldsymbol{\nu},\tau^2\mathbf{I})\left[\int \mathrm{N}_{\mathbf{x}_i}\{\mathbf{y}_j,(h_1^2+h_2^2)\mathbf{I}\}\mathrm{N}_{\mathbf{x}_i}(\boldsymbol{\mu},\sigma^2\mathbf{I})\mathrm{d}\mathbf{x}_i\right]\mathrm{d}\mathbf{y}_j,\end{aligned}$$

where an expression of the form $\mathbf{N}_{\mathbf{x}}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is equal to the probability density function of a bivariate normal random variable $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance

matrix $\boldsymbol{\Sigma}$.

Using the expression for the integral of a product of bivariate normal distributions calculated in Section 2.4.1, we see that this is equal to

$$\begin{aligned}
&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int\left[\mathrm{N}_{\mathbf{y}_j}(\boldsymbol{\nu},\tau^2\mathbf{I})\frac{1}{2\pi(\sigma^2+h_1^2+h_2^2)}\right.\\
&\qquad\left.\times\exp\left\{-\frac{(\mathbf{y}_j-\boldsymbol{\mu})^T(\mathbf{y}_j-\boldsymbol{\mu})}{2(h_1^2+h_2^2+\sigma^2)}\right\}\right]\mathrm{d}\mathbf{y}_j\\
&=\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\int\mathrm{N}_{\mathbf{y}_j}(\boldsymbol{\nu},\tau^2\mathbf{I})\mathrm{N}_{\mathbf{y}_j}\{\boldsymbol{\mu},(h_1^2+h_2^2+\sigma^2)\mathbf{I}\}\mathrm{d}\mathbf{y}_j\\
&=\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{2\pi(h_1^2+h_2^2+\sigma^2+\tau^2)}\exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(h_1^2+h_2^2+\sigma^2+\tau^2)}\right\}\\
&=\frac{1}{2\pi(h_1^2+h_2^2+\sigma^2+\tau^2)}\exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(h_1^2+h_2^2+\sigma^2+\tau^2)}\right\}. \qquad (3.14)
\end{aligned}$$

For example, when f_1 and f_2 are given by $\mathrm{N}\{(0,0)^T,\mathbf{I}\}$ and $\mathrm{N}\{(d,0)^T,\mathbf{I}\}$ respectively, for some distance d, then

$$\mathrm{E}\left\{\int\hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\}=\frac{1}{2\pi(h_1^2+h_2^2+2)}\exp\left\{-\frac{d^2}{2(h_1^2+h_2^2+2)}\right\},$$

and it is therefore possible to calculate the above expectation for varying distances d between the two densities. The results can be compared with the true values of $\int f_1(\mathbf{x})f_2(\mathbf{x})\mathrm{d}\mathbf{x}$ for the same distributions.

Similarly, we can determine an expression for the variance of the estimated product measure. Under the assumption that the observations in each of the two data sets are independent and identically distributed, we have

$$\begin{aligned}
&\mathrm{Var}\left\{\int\hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\}\\
&\qquad=\mathrm{Var}\left\{\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}\\
&\qquad=\frac{1}{n_1^2 n_2^2}\mathrm{Var}\left\{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}\\
&\qquad=\frac{1}{n_1^2 n_2^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathrm{Var}\left\{\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}.
\end{aligned}$$

To calculate $\mathrm{Var}\left\{\hat{g}\left(\mathbf{x}_i,\mathbf{y}_j\right)\right\}$, we can find $\mathrm{E}\left\{\hat{g}(\mathbf{x}_i,\mathbf{y}_j)^2\right\}$, as $\left[\mathrm{E}\left\{\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}\right]^2$ follows immediately from (3.14). We note that $\mathrm{E}\left\{\hat{g}(\mathbf{x}_i,\mathbf{y}_j)^2\right\}$ is equal to

$$
\begin{aligned}
&\mathrm{E}\left(\left[\frac{1}{2\pi(h_1^2+h_2^2)}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{2(h_1^2+h_2^2)}\right\}\right]^2\right)\\
&\quad=\mathrm{E}\left[\frac{1}{4\pi^2(h_1^2+h_2^2)^2}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{h_1^2+h_2^2}\right\}\right]\\
&\quad=\int\int\frac{1}{4\pi^2(h_1^2+h_2^2)^2}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{h_1^2+h_2^2}\right\}f_{\mathbf{x},\mathbf{y}}(\mathbf{x}_i,\mathbf{y}_j)\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j\\
&\quad=\int\int\frac{1}{4\pi^2(h_1^2+h_2^2)^2}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{h_1^2+h_2^2}\right\}f_1(\mathbf{x}_i)f_2(\mathbf{y}_j)\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j\\
&\quad=\frac{1}{4\pi(h_1^2+h_2^2)}\int\int\frac{1}{2\pi\times\frac{1}{2}(h_1^2+h_2^2)}\exp\left\{-\frac{(\mathbf{x}_i-\mathbf{y}_j)^T(\mathbf{x}_i-\mathbf{y}_j)}{2\times\frac{1}{2}(h_1^2+h_2^2)}\right\}\\
&\qquad\times f_1(\mathbf{x}_i)f_2(\mathbf{y}_j)\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j\\
&\quad=\frac{1}{4\pi(h_1^2+h_2^2)}\int\int\mathrm{N}_{\mathbf{x}_i}\left\{\mathbf{y}_j,\frac{1}{2}(h_1^2+h_2^2)\right\}\mathrm{N}_{\mathbf{x}_i}\left(\boldsymbol{\mu},\sigma^2\mathbf{I}\right)\mathrm{N}_{\mathbf{y}_j}\left(\boldsymbol{\nu},\tau^2\mathbf{I}\right)\mathrm{d}\mathbf{x}_i\mathrm{d}\mathbf{y}_j,
\end{aligned}
$$

and similar calculations to those used for $\mathrm{E}\left(\hat{g}\{\mathbf{x}_i,\mathbf{y}_j\}\right)$ show that this is equal to

$$
\left\{4\pi^2(h_1^2+h_2^2)(h_1^2+h_2^2+2\sigma^2+2\tau^2)\right\}^{-1}\exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{h_1^2+h_2^2+2\sigma^2+2\tau^2}\right\}.
$$

When f_1 is given by $\mathrm{N}\{(0,0)^T,\mathbf{I}\}$ and f_2 by $\mathrm{N}\{(d,0)^T,\mathbf{I}\}$, this expectation is equal to

$$
\left\{4\pi^2(h_1^2+h_2^2)(h_1^2+h_2^2+4)\right\}^{-1}\exp\left(-\frac{d^2}{h_1^2+h_2^2+4}\right).
$$

Furthermore, we know that

$$
\begin{aligned}
\left\{\mathrm{E}(\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\right\}^2&=\left[\frac{1}{2\pi(h_1^2+h_2^2+2)}\exp\left\{-\frac{d^2}{2(h_1^2+h_2^2+2)}\right\}\right]^2\\
&=\frac{1}{4\pi^2(h_1^2+h_2^2+2)^2}\exp\left(-\frac{d^2}{h_1^2+h_2^2+2}\right),
\end{aligned}
$$

and therefore that

$$
\begin{aligned}
\mathrm{Var}\{\hat{g}(\mathbf{x}_i,\mathbf{y}_j)\}&=\left\{4\pi^2(h_1^2+h_2^2)(h_1^2+h_2^2+4)\right\}^{-1}\exp\left(-\frac{d^2}{h_1^2+h_2^2+4}\right)\\
&\quad-\left\{4\pi^2(h_1^2+h_2^2+2)^2\right\}^{-1}\exp\left(-\frac{d^2}{h_1^2+h_2^2+2}\right).
\end{aligned}
$$

It follows that

$$\begin{aligned}
&\mathrm{Var}\left(\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right) \\
&\quad = \frac{1}{n_1 n_2}\left[\left\{4\pi^2(h_1^2+h_2^2)(h_1^2+h_2^2+4)\right\}^{-1}\exp\left(-\frac{d^2}{h_1^2+h_2^2+4}\right)\right. \\
&\qquad \left. -\left\{4\pi^2(h_1^2+h_2^2+2)^2\right\}^{-1}\exp\left(-\frac{d^2}{h_1^2+h_2^2+2}\right)\right].
\end{aligned}$$

We simulated data sets of 100 observations each from f_1 and f_2, using $d = 0, 1, 2, 3, 4, 5$, and used fixed kernels with asymptotically optimal smoothing parameter selection to obtain density estimates for each simulation. This method of smoothing parameter selection is asymptotically optimal here, as the data are generated from bivariate normal distributions. We calculated the expectation and standard deviation of the product measure of distance between the two data sets for each value of d using the methods detailed in this section. The true value of the product measure can be found easily in this case and was also recorded. The results are presented in Table 3.13, and are scaled by division by

$$\frac{1}{2}\left\{\int f_1^2(\mathbf{x})\mathrm{d}\mathbf{x} + \int f_2^2(\mathbf{x})\mathrm{d}\mathbf{x}\right\} = \frac{1}{4\pi}$$

so that the scaled product measure ranges from 0 in the case of no overlap to 1 for complete overlap.

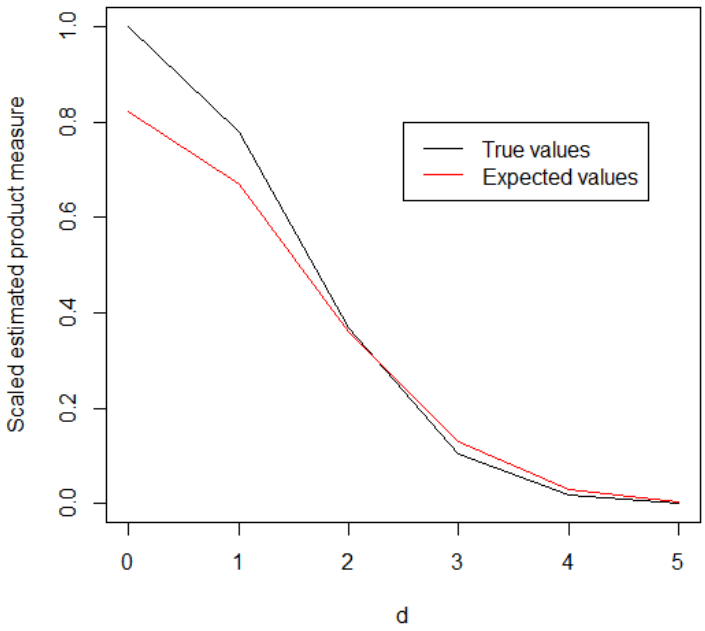
Table 3.13: Summary statistics of the scaled product measure distances between the densities $N\{(0,0)^T, \mathbf{I}\}$ and $N\{(d,0)^T, \mathbf{I}\}$. Results are recorded for $d = 0, 1, 2, 3, 4, 5$, and obtained using fixed kernel density estimators with asymptotically optimal smoothing parameter selection. Presented summary statistics include the true value of the scaled product measure and the expectation, standard deviation, and bias based given the simulation of 100 observations from each density.

	Summary statistics of scaled product measure			
d	**Expectation**	**True value**	**Bias**	**Standard deviation**
0	0.8227	1.0000	−0.1773	0.0119
1	0.6698	0.7788	−0.1090	0.0111
2	0.3614	0.3670	−0.0065	0.0085
3	0.1292	0.1054	0.0238	0.0051
4	0.0306	0.0183	0.0123	0.0024
5	0.0048	0.0019	0.0029	0.0009

The expected values of the scaled product measure have negative bias when

the two overlapping distributions are close together, and positive bias when the two distributions are further apart. This can be seen in Figure 3.3, which displays the expected and true values of the scaled product measure with increasing distance. The scaled bias and standard deviation of the scaled product measure are displayed in Figure 3.4, and Figure 3.5 shows their values divided by the expected scaled product measure. It is apparent from these two figures that while the overall magnitude of the scaled bias of the estimated product measure decreases as the distance between the two distributions increases, with the bias negative for small distances between distributions and positive for larger distances, the bias increases with d as a proportion of the expectation of the estimated product measure. The scaled standard deviation also decreases with increasing distance between distribution, but increases as a proportion of the expectation of the estimated product measure.

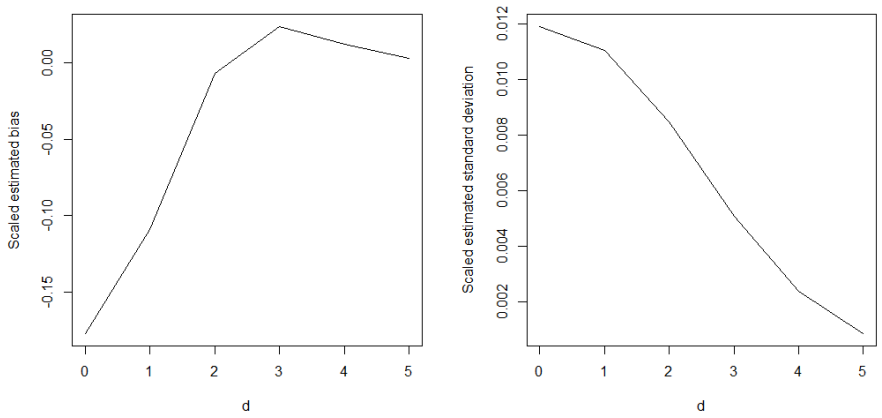
Figure 3.3: Expected and true values of the scaled product measure of the densities f_1 and f_2 given by $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$ respectively, for $d = 0, 1, 2, 3, 4, 5$.



The mean squared error (MSE) of the estimated distribution in each case is equal to

$$\left\{\mathrm{Bias}\hat{f}(\mathbf{x})\right\}^2 + \mathrm{Var}\{\hat{f}(\mathbf{x})\},$$

Figure 3.4: Bias and standard deviation of the scaled product measure of the densities f_1 and f_2 given by $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$ respectively, for distances $d = 0, 1, 2, 3, 4, 5$. The bias is displayed in the left panel, and the standard deviation in the right.



and Table 3.14 displays the values of the MSE for different distances between the two distributions. The MSE behaves as displayed in Figure 3.6 with increasing d. It decreases rapidly as d increases from 0, attains a local maximum between $d = 3$ and $d = 4$, and subsequently decreases with further increasing distance. The selection of values for the smoothing parameters that minimize MSE may not be optimal for estimation of overlap. To investigate whether the selected values of h minimize the MSE, we view plots of the MSE as a function of the smoothing parameter, which are shown in Figure 3.7. Here we assume a common smoothing parameter for both fitted distributions and a fixed sample size of 100 for both data sets. Lines corresponding to the asymptotically optimal choice of $h = 100^{-\frac{1}{6}}$ have been added to each of the plots.

The selected values of h do not correspond exactly with the values that minimise the MSE of the product measure, but they are fairly close in terms of MSE for moderate overlap. In some cases, the MSE varies very little over a comparatively large range for h. It should be noted that when h is close to 0, the variance is by far the larger component of the MSE, with the squared bias extremely small. The squared bias is, of course, equal to 0 at $h = 0$. As h increases, the variance decreases and the squared bias increases, eventually to the point where the variance is negligible in comparison to the squared bias, which becomes by far the larger component of the MSE.

Figure 3.5: Bias and variance of the scaled product measure of the densities f_1 and f_2 given by $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$ respectively, for $d = 0, 1, 2, 3, 4, 5$, expressed as a proportion of the expected value of the scaled product measure. The bias is displayed in the left panel, and the standard deviation in the right.

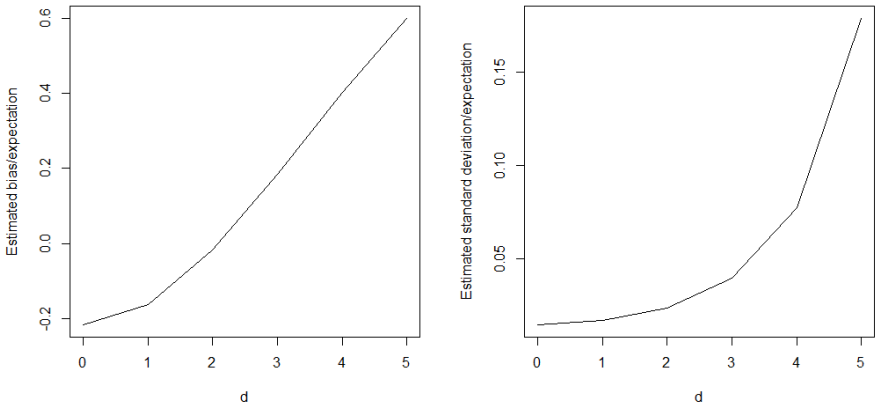


Figure 3.6: Scaled mean squared error (MSE) of the product measure of the densities f_1 and f_2 given by $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$ respectively, for distances $d = 0, 1, 2, 3, 4, 5$. The left panel displays the MSE, and the right panel the MSE as a proportion of the expectation of the scaled product measure.

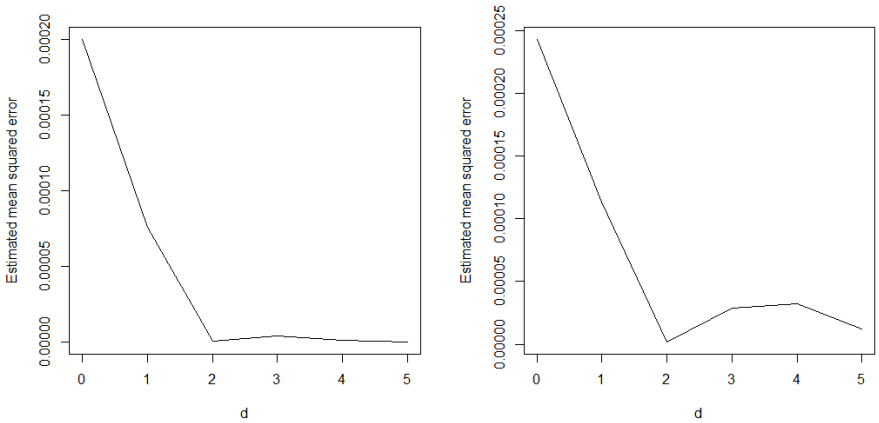
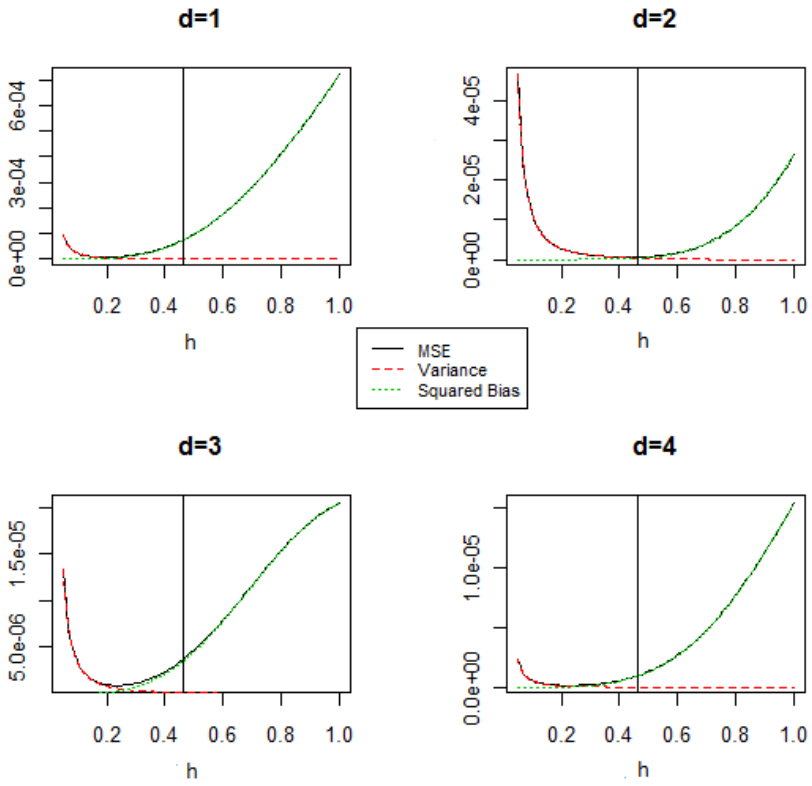


Table 3.14: Mean squared error (MSE) of the scaled product measure distances between two hypothetical data sets of 100 observations each. One data set consists of points simulated from $N\{(0,0)^T, \mathbf{I}\}$ and the other of points simulated from $N\{(d,0)^T, \mathbf{I}\}$. Results are recorded for $d = 0, 1, 2, 3, 4, 5$, using fixed kernel density estimators with asymptotically optimal smoothing parameters.

d	**MSE**
0	1.999×10^{-4}
1	7.603×10^{-5}
2	7.236×10^{-7}
3	3.755×10^{-6}
4	9.941×10^{-7}
5	5.716×10^{-8}

Figure 3.7: Scaled mean squared error (MSE) of the product measure of the densities f_1 and f_2 given by $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$ respectively. The mean squared error is considered as a function of the smoothing parameter h. Plots for $d = 1, 2, 3, 4$ are shown in the top left, top right, bottom left and bottom right panels respectively.



3.5 Bias approximation for product measure

When bivariate normal density kernels $\hat{f}_1$ and $\hat{f}_2$ are used to model data sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$ generated from the distributions f_1 and f_2 respectively, the bias of

$$\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}$$

can be calculated analytically. However, this is not necessarily possible for all kernel density estimators. Under the assumption that $\hat{f}_1$ and $\hat{f}_2$ are independent,

$$\begin{aligned}
\mathrm{E}\left\{\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\} &= \int \mathrm{E}\left\{\hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&= \int \mathrm{E}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{E}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&= \int \left[f_1(\mathbf{x}) + \mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\right]\left[f_2(\mathbf{x}) + \mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\right]\mathrm{d}\mathbf{x} \\
&= \int f_1(\mathbf{x})f_2(\mathbf{x})\mathrm{d}\mathbf{x} + \int f_1(\mathbf{x})\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&\quad + \int f_2(\mathbf{x})\mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&\quad + \int \mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x},
\end{aligned}$$

and it follows that

$$\begin{aligned}
\mathrm{Bias}\left\{\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\} &= \mathrm{E}\left\{\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\} - \int f_1(\mathbf{x})f_2(\mathbf{x})\mathrm{dx} \\
&= \int f_1\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x} + \int f_2\mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&\quad + \int \mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x}. \qquad (3.15)
\end{aligned}$$

In cases where direct calculation is not possible, we can estimate the values of the integrals in (3.15) using the bias approximations

$$\begin{aligned}
\mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\} &\approx \frac{1}{2}\alpha_1 h_1^2 \nabla^2 f_1(\mathbf{x}), \qquad (3.16) \\
\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\} &\approx \frac{1}{2}\alpha_2 h_2^2 \nabla^2 f_2(\mathbf{x}),
\end{aligned}$$

where h_1 and h_2 are the smoothing parameters of $\hat{f}_1$ and $\hat{f}_2$ respectively (Silverman, 1986). In the case where f_1 and f_2 are mixtures of k_1 and k_2 bivariate normal distributions respectively, such that the ith component of f_1

has mean vector $\boldsymbol{\mu}_i = (\mu_{i_1}, \mu_{i_2})^T$, covariance matrix $\sigma_i^2\mathbf{I}$ and weight w_i, and the ith component of f_2 has mean vector $\boldsymbol{\nu}_i = (\nu_{i_1}, \nu_{i_2})^T$, covariance matrix $\tau_i^2\mathbf{I}$ and weight v_i, then (Silverman, 1986),

$$\nabla^2 f_1(\mathbf{x}) = \sum_{i=1}^{k_1} \frac{w_i \left\{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i) - 2\sigma_i^2\right\}}{2\pi\sigma_i^6} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)}{2\sigma_i^2}\right\},$$

$$\nabla^2 f_2(\mathbf{x}) = \sum_{i=1}^{k_2} \frac{v_i \left\{(\mathbf{x}-\boldsymbol{\nu}_i)^T(\mathbf{x}-\boldsymbol{\nu}_i) - 2\sigma_i^2\right\}}{2\pi\tau_i^6} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\nu}_i)^T(\mathbf{x}-\boldsymbol{\nu}_i)}{2\tau_i^2}\right\}.$$

We observe that

$$\begin{aligned}
&\int f_1(\mathbf{x})\text{Bias}\left\{\hat{f}_2(\mathbf{x})\right\} d\mathbf{x} \\
&\quad\approx \int \left[\sum_{i=1}^{k_1} \frac{w_i}{2\pi\sigma_i^2} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)}{2\sigma_i^2}\right\}\right. \\
&\qquad\times \frac{\alpha_2 h_2^2}{2} \sum_{j=1}^{k_2} \frac{v_j\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j) - 2\tau_j^2\}}{2\pi\tau_j^6} \\
&\qquad\left.\times \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j)}{2\tau_j^2}\right\}\right] d\mathbf{x} \\
&\quad= \int \left[\sum_{i=1}^{k_1} w_i \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}_i, \sigma_i^2\mathbf{I})\right. \\
&\qquad\left.\times \sum_{j=1}^{k_2} \frac{\alpha_2 h_2^2 v_j\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j) - 2\tau_j^2\}}{2\tau_i^4} \mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}_j, \tau_j^2\mathbf{I})\right] d\mathbf{x} \\
&\quad= \int \left[\sum_{i=1}^{k_1} w_i \mathrm{N}_i \sum_{j=1}^{k_2} \frac{\alpha_2 h_2^2 v_j\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j) - 2\tau_j^2\}}{2\tau_j^4} \mathrm{N}_j\right] d\mathbf{x} \\
&\quad= \frac{\alpha_2 h_2^2}{2} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \int \frac{w_i v_j\{(x_1-\nu_{j_1})^2 + (x_2-\nu_{j_2})^2 - 2\tau_j^2\}}{\tau_j^4} \mathrm{N}_i\mathrm{N}_j d\mathbf{x} \\
&\quad= \frac{\alpha_2 h_2^2}{2} \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} \frac{w_i v_j}{\tau_j^4} \left\{\int x_1^2 \mathrm{N}_i\mathrm{N}_j d\mathbf{x} + \int x_2^2 \mathrm{N}_i\mathrm{N}_j d\mathbf{x} - 2\nu_{j_1}\int x_1\mathrm{N}_i\mathrm{N}_j d\mathbf{x}\right. \\
&\qquad\left. -2\nu_{j_2}\int x_2\mathrm{N}_i\mathrm{N}_j d\mathbf{x} + (\nu_{j_1}^2 + \nu_{j_2}^2 - 2\tau_j^2)\int \mathrm{N}_i\mathrm{N}_j d\mathbf{x}\right\},
\end{aligned}$$

where N_i and N_j are the ith and jth components of f_1 and f_2 respectively, and h_1 and h_2 are the smoothing parameters of $\hat{f}_1$ and $\hat{f}_2$ respectively. The evaluation of the above expression therefore requires similar calculations so those carried out in Section 2.4 to obtain the integral of the product of two bivariate normal distributions. Consider the bivariate normal densities $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})$,

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and $\boldsymbol{\nu} = (\nu_1, \nu_2)^T$. Similar working to that in Section 2.4, and calculation using Maple, gives

$$\begin{aligned}
&\int x_1 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x} \\
&\quad = \frac{\sigma^2\nu_1 + \tau^2\mu_1}{2\pi(\sigma^2+\tau^2)^2} \exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)}\right\}, \qquad (3.17) \\
&\int x_2 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x} \\
&\quad = \frac{\sigma^2\nu_2 + \tau^2\mu_2}{2\pi(\sigma^2+\tau^2)^2} \exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)}\right\}, \\
&\int x_1^2 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x} \\
&\quad = \frac{1}{2\pi}\left\{\frac{(\mu_1\tau^2 + \nu_1\sigma^2)^2}{(\sigma^2+\tau^2)^3} + \frac{\sigma^2\tau^2}{(\sigma^2+\tau^2)^2}\right\} \exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{\sigma^2+\tau^2}\right\}, \\
&\int x_2^2 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x} \\
&\quad = \frac{1}{2\pi}\left\{\frac{(\mu_2\tau^2 + \nu_2\sigma^2)^2}{(\sigma^2+\tau^2)^3} + \frac{\sigma^2\tau^2}{(\sigma^2+\tau^2)^2}\right\} \exp\left\{-\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{\sigma^2+\tau^2}\right\}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\int f_1(\mathbf{x})\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&\quad = \frac{\alpha_2 h_2^2}{4\pi}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\frac{w_i v_j}{\tau_j^4}\left\{\frac{(\mu_{i_1}\tau_j^2 + \nu_{j_1}\sigma_i^2)^2}{(\sigma_i^2+\tau_j^2)^3} + \frac{\sigma_i^2\tau_j^2}{(\sigma_i^2+\tau_j^2)^2} + \frac{(\mu_{i_2}\tau_j^2 + \nu_{j_2}\sigma_i^2)^2}{(\sigma_i^2+\tau_j^2)^3}\right. \\
&\qquad + \frac{\sigma_i^2\tau_j^2}{(\sigma_i^2+\tau_j^2)^2} - \frac{2\nu_{j_1}(\mu_{i_1}\tau_j^2 + \nu_{j_1}\sigma_i^2)}{\sigma_i^2+\tau_j^2} - \frac{2\nu_{j_2}(\mu_{i_2}\tau_j^2 + \nu_{j_2}\sigma_i^2)}{\sigma_i^2+\tau_j^2} \\
&\qquad \left. + \frac{(\nu_{j_1}^2 + \nu_{j_2}^2 - 2\tau_j^2)}{\sigma_i^2+\tau_j^2}\right\} \times \exp\left\{-\frac{(\boldsymbol{\mu}_i - \boldsymbol{\nu}_j)^T(\boldsymbol{\mu}_i - \boldsymbol{\nu}_j)}{2(\sigma_i^2+\tau_j^2)}\right\},
\end{aligned}$$

and likewise

$$\begin{aligned}
&\int f_2(\mathbf{x})\mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&\quad = \frac{\alpha_1 h_1^2}{4\pi}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\frac{w_i v_j}{\tau_j^4}\left\{\frac{(\mu_{i_1}\tau_j^2 + \nu_{j_1}\sigma_i^2)^2}{(\sigma_i^2+\tau_j^2)^3} + \frac{\sigma_i^2\tau_j^2}{(\sigma_i^2+\tau_j^2)^2} + \frac{(\mu_{i_2}\tau_j^2 + \nu_{j_2}\sigma_i^2)^2}{(\sigma_i^2+\tau_j^2)^3}\right. \\
&\qquad + \frac{\sigma_i^2\tau_j^2}{(\sigma_i^2+\tau_j^2)^2} - \frac{2\mu_{j_1}(\mu_{i_1}\tau_j^2 + \nu_{j_1}\sigma_i^2)}{\sigma_i^2+\tau_j^2} - \frac{2\mu_{j_2}(\mu_{i_2}\tau_j^2 + \nu_{j_2}\sigma_i^2)}{\sigma_i^2+\tau_j^2} \\
&\qquad \left. + \frac{(\mu_{j_1}^2 + \mu_{j_2}^2 - 2\sigma_i^2)}{\sigma_i^2+\tau_j^2}\right\} \times \exp\left\{-\frac{(\boldsymbol{\mu}_i - \boldsymbol{\nu}_j)^T(\boldsymbol{\mu}_i - \boldsymbol{\nu}_j)}{2(\sigma_i^2+\tau_j^2)}\right\}.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&\int \mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\}\mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\}\mathrm{d}\mathbf{x} \\
&\quad\approx \int\left[\sum_{i=1}^{k_1}\frac{w_i\alpha_1 h_1^2\{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)-2\sigma_i^2\}}{4\pi\sigma_i^6}\exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)}{2\sigma_i^2}\right\}\right. \\
&\qquad\times\sum_{j=1}^{k_2}\frac{v_j\alpha_2 h_2^2\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j)-2\tau_j^2\}}{4\pi\tau_j^6} \\
&\qquad\left.\times\exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j)}{2\tau_j^2}\right\}\right]\mathrm{d}\mathbf{x} \\
&\quad= \int\left[\sum_{i=1}^{k_1}\frac{w_i\alpha_1 h_1^2\{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)-2\sigma_i^2\}}{2\sigma_i^4}\mathrm{N}_{\mathbf{x}}\left(\boldsymbol{\mu}_i,\sigma_i^2\mathbf{I}\right)\right. \\
&\qquad\left.\times\sum_{j=1}^{k_2}\frac{v_j\alpha_2 h_2^2\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j)-2\tau_i^2\}}{2\tau_j^4}\mathrm{N}_{\mathbf{x}}\left(\boldsymbol{\nu}_j,\tau_j^2\mathbf{I}\right)\right]\mathrm{d}\mathbf{x} \\
&\quad= \int\left[\sum_{i=1}^{k_1}\frac{w_i\alpha_1 h_1^2\{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)-2\sigma_i^2\}}{2\sigma_i^4}\mathrm{N}_i\right. \\
&\qquad\left.\times\sum_{j=1}^{k_2}\frac{v_j\alpha_2 h_2^2\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j)-2\tau_i^2\}}{2\tau_j^4}\mathrm{N}_j\right]\mathrm{d}\mathbf{x} \\
&\quad= \frac{\alpha_1\alpha_2 h_1^2 h_2^2}{4}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\frac{w_i v_j}{\sigma_i^4\tau_j^4}\int\left[\{(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)-2\sigma_i^2\}\right. \\
&\qquad\left.\times\{(\mathbf{x}-\boldsymbol{\nu}_j)^T(\mathbf{x}-\boldsymbol{\nu}_j)-2\tau_j^2\}\mathrm{N}_i\mathrm{N}_j\right]\mathrm{d}\mathbf{x},
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
&\frac{\alpha_1\alpha_2 h_1^2 h_2^2}{4}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\frac{w_i v_j}{\sigma_i^4\tau_j^4}\left\{\int x_1^4\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x}+\int x_2^4\mathrm{N}_i\mathrm{N}_j d\mathbf{x}\right. \\
&\quad-2(\mu_{i_1}+\nu_{j_1})\int x_1^3\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x}-2(\mu_{i_2}+\nu_{j_2})\int x_2^3\mathrm{d}\mathbf{x}+2\int x_1^2x_2^2\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x} \\
&\quad-2(\mu_{i_2}+\nu_{j_2})\int x_1^2x_2\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x}-2(\mu_{i_1}+\nu_{j_1})\int x_1x_2^2\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x} \\
&\quad+4(\mu_{i_1}\nu_{j_2}+\mu_{i_2}\nu_{j_1})\int x_1x_2\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x} \\
&\quad\left.+(4\mu_{i_1}\nu_{j_1}+\mu_{i_1}^2+\mu_{i_2}^2+\nu_{j_1}^2+\nu_{j_2}^2-2\sigma_i^2-2\tau_j^2)\int x_1^2\mathrm{N}_i\mathrm{N}_j\mathrm{d}\mathbf{x}\right.
\end{aligned}
$$

$$
\begin{aligned}
&+ (4\mu_{i_2}\nu_{j_2} + \mu_{i_1}^2 + \mu_{i_2}^2 + \nu_{j_1}^2 + \nu_{j_2}^2 - 2\sigma_i^2 - 2\tau_j^2) \int x_2^2 \mathrm{N}_i \mathrm{N}_j \mathrm{d}\mathbf{x} \\
&- 2(\mu_{i_1}\nu_{j_1}^2 + \mu_{i_1}\nu_{j_2}^2 + \mu_{i_1}^2\nu_{j_1} + \mu_{i_2}^2\nu_{j_1} - 2\mu_{i_1}\tau_j^2 - 2\nu_{j_1}\sigma_i^2) \int x_1 \mathrm{N}_i \mathrm{N}_j \mathrm{d}\mathbf{x} \\
&- 2(\mu_{i_2}\nu_{j_1}^2 + \mu_{i_2}\nu_{j_2}^2 + \mu_{i_1}^2\nu_{j_2} + \mu_{i_2}^2\nu_{j_2} - 2\mu_{i_2}\tau_j^2 - 2\nu_{j_2}\sigma_i^2) \int x_2 \mathrm{N}_i \mathrm{N}_j \mathrm{d}\mathbf{x} \\
&+ (\mu_{i_1}^2\nu_{j_1}^2 + \mu_{i_1}^2\nu_{j_2}^2 + \mu_{i_2}^2\nu_{j_1}^2 + \mu_{i_2}^2\nu_{j_2}^2 - 2\mu_{i_1}^2\tau_j^2 - 2\mu_{i_2}^2\tau_j^2 - 2\nu_{j_1}^2\sigma_i^2 - 2\nu_{j_2}^2\sigma_i^2 \\
&+ 4\sigma_i^2\tau_j^2) \int \mathrm{N}_i \mathrm{N}_j \mathrm{d}\mathbf{x} \Bigg\}. \qquad (3.18)
\end{aligned}
$$

We use Maple to calculate

$$
\begin{aligned}
&\int x_1^4 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I}) \mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I}) \mathrm{d}\mathbf{x} \\
&\quad = \left\{ \frac{(\mu_1\tau^2 + \nu_1\sigma^2)^4}{(\sigma^2+\tau^2)^5} + \frac{6\sigma^2\tau^2(\mu_1\tau^2 + \nu_1\sigma^2)^2}{(\sigma^2+\tau^2)^4} + \frac{3\sigma^4\tau^4}{(\sigma^2+\tau^2)^3} \right\} \\
&\qquad \times \frac{1}{2\pi} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&\int x_2^4 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I}) \mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I}) \mathrm{d}\mathbf{x} \\
&\quad = \left\{ \frac{(\mu_2\tau^2 + \nu_2\sigma^2)^4}{(\sigma^2+\tau^2)^5} + \frac{6\sigma^2\tau^2(\mu_2\tau^2 + \nu_2\sigma^2)^2}{(\sigma^2+\tau^2)^4} + \frac{3\sigma^4\tau^4}{(\sigma^2+\tau^2)^3} \right\} \\
&\qquad \times \frac{1}{2\pi} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&\int x_1^2 x_2^2 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I}) \mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I}) \mathrm{d}\mathbf{x} \\
&\quad = \left\{ \frac{(\mu_1\tau^2 + \nu_1\sigma^2)^2(\mu_2\tau^2 + \nu_2\sigma^2)^2}{(\sigma^2+\tau^2)^5} + \frac{(\mu_1\tau^2 + \mu_2\tau^2 + \nu_1\sigma^2 + \nu_2\sigma^2)\sigma^2\tau^2}{(\sigma^2+\tau^2)^4} \right. \\
&\qquad \left. + \frac{\sigma^4\tau^4}{(\sigma^2+\tau^2)^3} \right\} \times \frac{1}{2\pi} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\}.
\end{aligned}
$$

We also calculate

$$
\begin{aligned}
&\int x_1^3 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I}) \mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I}) \mathrm{d}\mathbf{x} \\
&\quad = \left\{ \frac{(\mu_1\tau^2 + \nu_1\sigma^2)^3}{(\sigma^2+\tau^2)^4} + \frac{3\sigma^2\tau^2(\mu_1\tau^2 + \nu_1\sigma^2)}{(\sigma^2+\tau^2)^3} \right\} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\},
\end{aligned}
$$

and

$$\int x_2^3 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x}$$
$$= \left\{ \frac{(\mu_2\tau^2 + \nu_2\sigma^2)^3}{(\sigma^2+\tau^2)^4} + \frac{3\sigma^2\tau^2(\mu_2\tau^2+\nu_2\sigma^2)}{(\sigma^2+\tau^2)^3} \right\} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\},$$

and

$$\int x_1^2 x_2 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x}$$
$$= \left\{ \frac{(\mu_1\tau^2 + \nu_1\sigma^2)^2(\mu_2\tau^2+\nu_2\sigma^2)}{(\sigma^2+\tau^2)^4} + \frac{\sigma^2\tau^2(\mu_2\tau^2+\nu_2\sigma^2)}{(\sigma^2+\tau^2)^3} \right\}$$
$$\times \frac{1}{2\pi} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\},$$

as well as

$$\int x_1 x_2^2 \mathrm{N}_{\mathbf{x}}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x}$$
$$= \left\{ \frac{(\mu_1\tau^2 + \nu_1\sigma^2)(\mu_2\tau^2+\nu_2\sigma^2)^2}{(\sigma^2+\tau^2)^4} + \frac{\sigma^2\tau^2(\mu_1\tau^2+\nu_1\sigma^2)}{(\sigma^2+\tau^2)^3} \right\}$$
$$\times \frac{1}{2\pi} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\},$$

and

$$\int x_1 x_2 \mathrm{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})\mathrm{N}_{\mathbf{x}}(\boldsymbol{\nu}, \tau^2\mathbf{I})\mathrm{d}\mathbf{x}$$
$$= \frac{(\mu_1\tau^2 + \nu_1\sigma^2)(\mu_2\tau^2+\nu_2\sigma^2)}{2\pi(\sigma^2+\tau^2)^3} \exp\left\{ -\frac{(\boldsymbol{\mu}-\boldsymbol{\nu})^T(\boldsymbol{\mu}-\boldsymbol{\nu})}{2(\sigma^2+\tau^2)} \right\}.$$

Formulae for the other integral terms in (3.18) are given in the collection of expressions beginning with (3.17), and an approximate value for

$$\int \mathrm{Bias}\left\{\hat{f}_1(\mathbf{x})\right\} \mathrm{Bias}\left\{\hat{f}_2(\mathbf{x})\right\} \mathrm{d}\mathbf{x}$$

can therefore be calculated in the same way. Note that in the case where $\hat{f}_1(\mathbf{x})$ and $\hat{f}_2(\mathbf{x})$ are bivariate normal kernel density estimators, α_1 and α_2 in (3.16) are equal to 1 (Silverman, 1986).

It is feasible that the bias approximation may not be accurate enough to remain useful when multiplied by the probability density function and integrated. When f_1 is given by $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and f_2 is given by $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$, the bias of

$\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}$ can easily be calculated directly, and we used this case to test the accuracy of the bias approximation. For distances $d = 0, 1, 2, 3, 4, 5$, 100 pairs of data sets of 100 observations each were simulated, with one data set in each pair simulated from f_1 and the other from f_2. Fixed bivariate normal kernels with asymptotically optimal smoothing parameter selection were used to obtain density estimates $\hat{f}_1$ and $\hat{f}_2$ for f_1 and f_2 respectively. Bias $\left\{\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})\mathrm{d}\mathbf{x}\right\}$ was calculated both directly, and by using the approximations detailed above and the relationship given in (3.15) and (3.16). The mean values of the exact and approximate bias for each value of d are presented in Table 3.15.

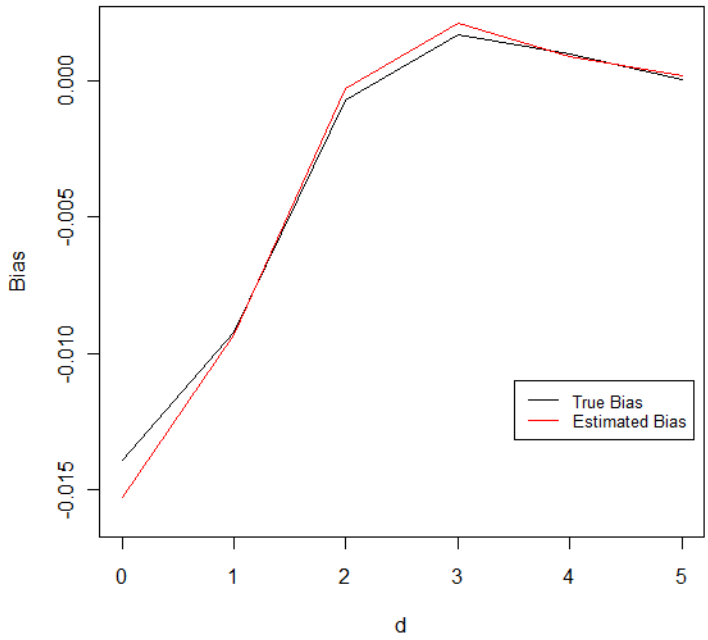
Table 3.15: Mean bias values of the product measure of the bivariate normal fixed kernel density estimates of 100 pairs of data sets of 100 observations each, with asymptotically optimal smoothing parameter selection. In each pair, one data set was simulated from $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and the other from $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$. Results are given for $d = 0, 1, 2, 3, 4, 5$, for the exact bias and the approximation using (3.15) and (3.16).

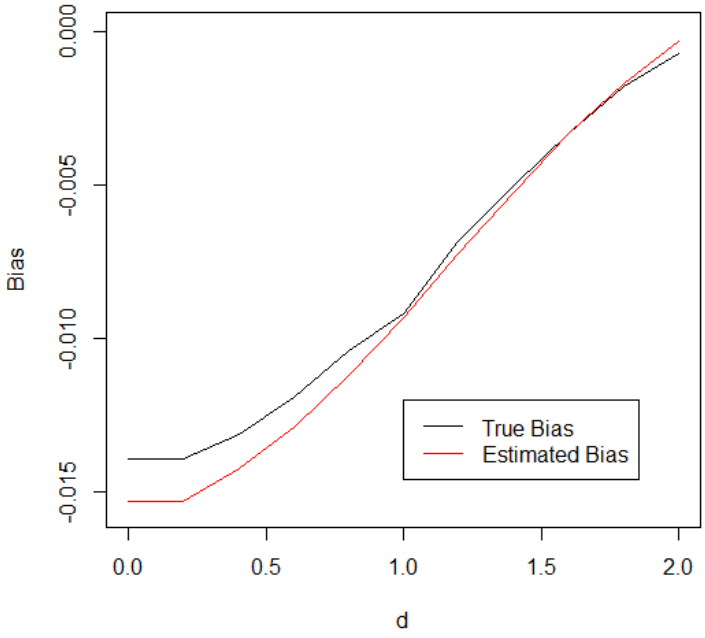
d	**Exact bias**	**Approximate bias**
0	−0.0139	−0.0153
1	−0.0092	−0.0093
2	−0.0007	−0.0003
3	0.0017	0.0021
4	0.0010	0.0009
5	0.0001	0.0002

The approximation appears to be accurate at all values of d, with the partial exception of $d = 0$, where it is slightly further from the true value. When $d = 2$ and $d = 5$, the difference between the true and approximated bias is large in comparison to the value of the bias itself. However, in absolute terms the difference is small. The top panel of Figure 3.8 plots the exact bias values against those provided by the estimator, and it is clear that the approximation has estimated the exact values well. These results suggest that the bias approximation is accurate enough that it still gives good approximations when used for the product measure of distance $\int \hat{f}_1(\mathbf{x})\hat{f}_2(\mathbf{x})d\mathbf{x}$ rather than for $\hat{f}(\mathbf{x})$.

As the bias approximation appears to be less accurate for $d = 0$ than for greater distances, the same calculations are carried out for smaller increments of d to gain a clearer picture of its behaviour when the level of overlap is very high. The values are shown in the bottom panel of Figure 3.8. It is clear that the approximation is close to the exact bias even for values of d that result in high levels of overlap between the two densities.

Figure 3.8: Plot of mean bias values of the product measure of the bivariate normal fixed kernel density estimates of 100 pairs of data sets of 100 observations each, with asymptotically optimal smoothing parameter selection. In each pair, one data set was simulated from $\mathrm{N}\{(0,0)^T, \mathbf{I}\}$ and the other from $\mathrm{N}\{(d,0)^T, \mathbf{I}\}$. Results for the exact bias and the approximate bias calculated using (3.15) are given for $d = 0, 1, 2, 3, 4, 5$ in the top panel, and for $d = 0, 0.2, \ldots, 2$ in the bottom panel.





3.6 Application to mule deer movement data

The mule deer data sets introduced in Chapter 2 and referred to as WG1 and WG2 contain location observations of the same animal from two consecutive summers, and we can measure the fidelity of the animal to its home range between the two years. This entails choosing a suitable measure of distance between the density estimates for the WG1 and WG2 data sets. A contour plot of the difference between the density estimates obtained for the 1984 and 1985 data sets using mixtures of t distributions in Section 2.6, as seen in Figure 2.30, is shown in Figure 3.9, to give an impression of the level of overlap. Figure 3.9 suggests that there are shifts between years in the areas most frequented, but that the animal has remained within the same area. The noticeable differences between the two density estimates are all located in areas common to the animal's home range over both years. Based on the plot in Figure 3.9, it is clear that the two densities differ from one another substantially, and that the data from the two years do not share a common distribution. A density contour plot of the minimum of the two density estimates is displayed in Figure 3.10, and suggests that the two densities do overlap to a great extent.

Figure 3.9: Contour plot of the difference between density estimates obtained using mixtures of t distributions for the WG1 and WG2 data sets.

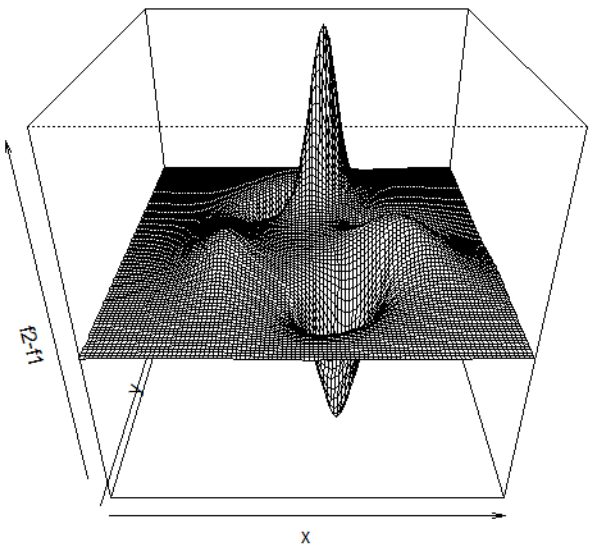
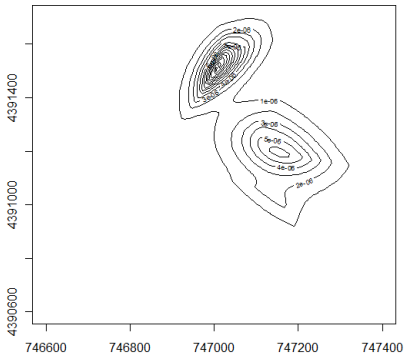


Figure 3.10: Contour plot of the minimum of the density estimates for the WG1 and WG2 data sets using mixtures of t distributions.



We can use the OVL to assess the goodness of fit of the mixture model by comparing the OVL value for the mixture density estimates to nonparametric estimates of the OVL derived from the data. Clemons and Bradley (2000) used naive kernel density estimators to provide nonparametric estimates of the OVL of pairs of data sets. In one dimension, the naive kernel estimator is given by

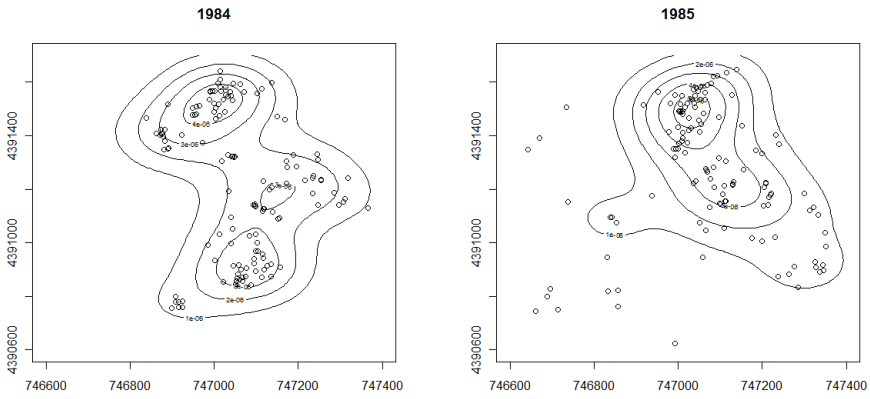
$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h},$$

where F_n is the empirical cumulative distribution function of the data, and h is the bandwidth of the kernel estimator. In order to obtain an estimate of the OVL of two data sets, Clemons and Bradley fitted naive kernel density estimators to each data set, and calculated the jump points of each empirical cumulative density function. The jump points for both data sets were combined into a single set sorted in ascending order, and the intervals between consecutive points calculated. The estimator of the OVL was then calculated by summing the area under the smaller curve over each interval. However, the extension of this method to multi-dimensional data is not straightforward, as the jump points cannot be sorted into ascending order, and a suitable ordering for them is not apparent. It is more practical to use the bivariate normal kernel density estimator introduced in Section 2.2.1.

The fixed kernel method was used to estimate the utilization densities for the WG1 and WG2 data sets, with smoothing parameters initially chosen using the asymptotically optimal approach outlined in Section 2.2.1. Contour plots of the resulting density estimates are presented in Figure 3.11, and a plot

of the difference between them in Figure 3.12. The kernel density estimates appear somewhat similar to the mixture densities in Figure 2.30. However, the asymptotically optimal choice of smoothing parameter is not a particularly rigorous method, and we use least-squared cross-validation to obtain alternate density estimates. Contour plots are shown in Figure 3.13.

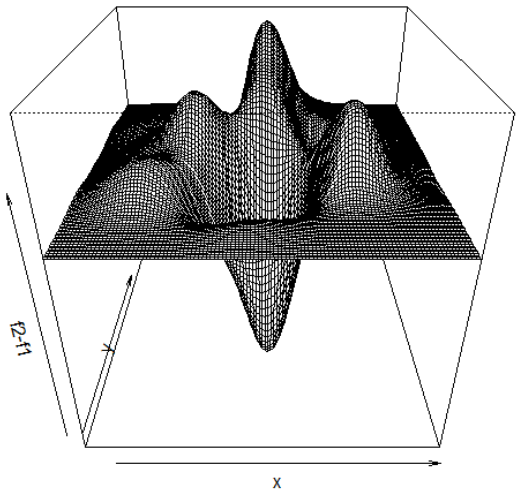
Figure 3.11: Contour plots of the fixed kernel density estimates for the WG1 and WG2 data sets using asymptotically optimal smoothing parameters.



The estimates obtained using least-squares cross-validation are clearly extremely undersmoothed, and do not provide a good approximation of the underlying densities. The adaptive kernel density estimator was also used, and the resulting density estimates displayed in Figure 3.14. However, the density estimates obtained by the adaptive kernel appear to be less appropriate than the fixed kernel. The density has been smoothed to nothing in any area outside the immediate proximity of an observation, resulting in density estimates that are effectively indistinguishable from the data sets. It is apparent from the plots in Figure 3.13 and Figure 3.14 that the bivariate normal kernels fitted using least-squares cross-validation are not appropriate for the data, and their estimates for the OVL are unreasonably low.

As mentioned in Chapter 1, Hemson et al. (2005) observed that least-squares cross-validation can fail for data sets with large numbers of points in close proximity. This may be the case here, as it can be seen in Figure 2.25 that the WG1 and WG2 data sets contain many instances of such points. We altered the positions of the observations slightly, replacing each data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

Figure 3.12: Contour plot of the difference between density estimates for the WG1 and WG2 data sets using fixed kernel density estimators with asymptotically optimal smoothing parameters.



with $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ defined as

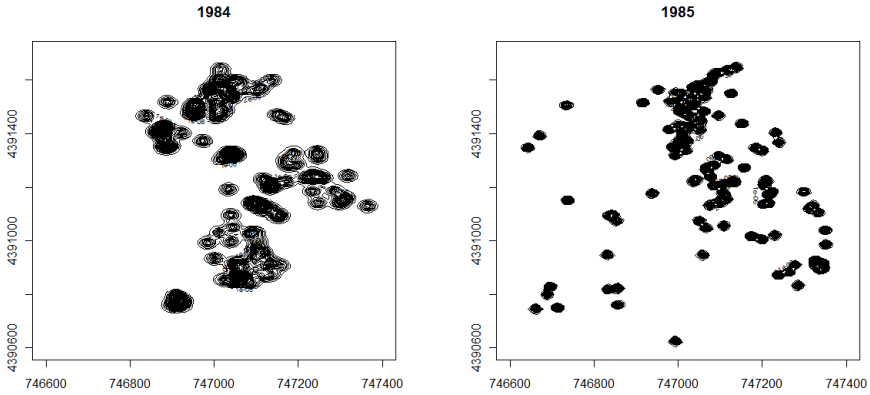
$$\mathbf{y}_i = \mathbf{x}_i + (u_{i1}, u_{i2}), \tag{3.19}$$

where

$$u_{i1}, u_{i2} \sim \text{Uniform}(-10, 10), \quad i = 1, \ldots, n.$$

The distribution of u_{i1} and u_{i2} was chosen to have small variance compared to the scale of the data sets. This alteration ensures that there are no longer identical observations in the data set. Contour plots of fixed and adaptive kernel density estimates for the altered data sets using least-squares cross-validation are displayed in Figures 3.15 and 3.16 respectively. While the estimates in Figures 3.15 and 3.16 are more reasonable than those in Figures 3.13 and 3.14, they still appear to be noticeably undersmoothed due to clusters of points. Plots of the differences between the two densities in Figure 3.17 also show obvious undersmoothing. Contour plots of the kernel density estimates for the WG1 and WG2 data sets using the plug-in rule to determine the smoothing parameter are presented in Figure 3.18, and appear to provide a more appropriate model than

Figure 3.13: Contour plots of the fixed kernel density estimates for the WG1 and WG2 data sets using smoothing parameters selected by least-squares cross-validation, shown in the left and right panels respectively.



the use of least-squares cross-validation.

The estimated scaled product measure, square root product measure and OVL for the density estimates obtained using the methods employed in this section are presented in Table 3.16, as well as the values of the measures calculated by numerical integration. The kernel density estimator using least-squares cross-validation is considered to be inappropriate and not used here. It is clear from Table 3.16 that the estimates are generally closest to the numerically calculated values for the mixture of bivariate t distributions. The estimates of the OVL are also accurate for both of the kernel density estimators. However, the estimates for the scaled product measure and square root product measure are further from the numerically calculated values. We additionally observe that the estimates and numerically calculated values obtained for the mixture model and the plug-in kernel are closer to one another than to the kernel with asymptotically optimal smoothing parameter. The numerically calculated OVL for the two density estimates is equal to 0.57, which is very close to the overlapping coefficient for the estimates obtained using the mixture model. This suggests that the mixture model correctly estimates the degree to which the two distributions overlap. The value of 0.57 appears reasonable on examination of the plots of the data shown in 2.25.

Figure 3.14: Contour plots of the adaptive kernel density estimates for the WG1 and WG2 data sets using smoothing parameters selected by least-squares cross-validation. shown in the left and right panels respectively.

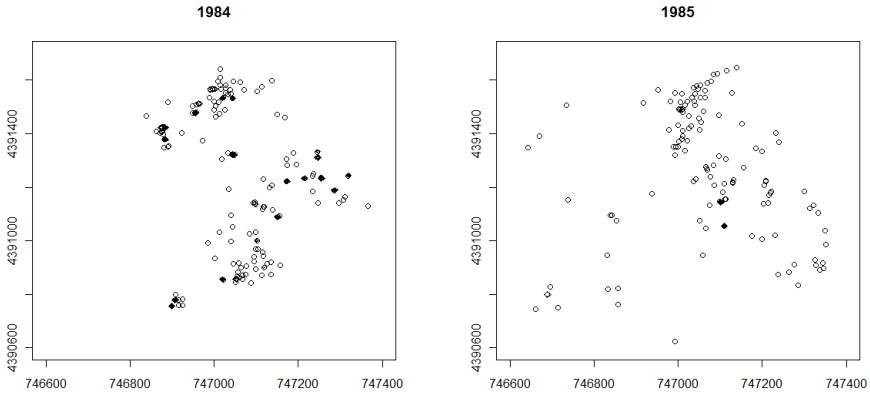


Table 3.16: Estimated and numerically calculated values of the scaled product measure (referred to as SP), square root product measure (referred to as SRP) and OVL of the density estimates obtained for the WG1 and WG2 data sets. Estimates for the SP, SRP and OVL were calculated using the estimators shown in (3.6), (3.7) and (3.8) respectively.

Model		**SP**	**SRP**	**OVL**
Kernel (asy opt)	Estimate	0.79	0.84	0.70
	Numerically calculated value	0.87	0.92	0.74
Kernel (plug-in)	Estimate	0.64	0.70	0.55
	Numerically calculated value	0.73	0.79	0.58
Mixture of t	Estimate	0.61	0.75	0.60
distributions	Numerically calculated value	0.61	0.81	0.55

We assessed the mixture of bivariate t distributions as a model for the WG1 and WG2 data sets by comparing the estimated overlap measures for the data sets to the corresponding values for pairs of data sets simulated from the mixture models fitted to the data. We simulated 200 pairs of data sets with each pair consisting of one set of 144 observations simulated from the mixture of t distributions fitted to the WG1 data set and one set of 135 observations simulated from the mixture fitted to the WG2 data set. The sizes of the data sets were chosen to be equal to those of the WG1 and WG2 data sets respectively. For each pair, the OVL, product measure and square root product measure were estimated using the estimators shown in (3.6), (3.7) and (3.8) respectively. Summary statistics are presented in Table 3.17. For all three overlap measures used, the estimated values for the mixture of t distributions in Table 3.16 are between the upper and lower 5% points. The results indicate that the true data sets are not unusual in terms of overlap when compared to the majority of the simulations, and thus suggest that the fitted mixture produces simulated data sets similar to the true data and is an appropriate model.

Figure 3.15: Contour plots of the fixed kernel density estimates using smoothing parameters selected by least-squares cross-validation, with the points in the data sets separated using (3.19). In each row, the left panel shows the WG1 data set, and the right panel shows the WG2 data set.

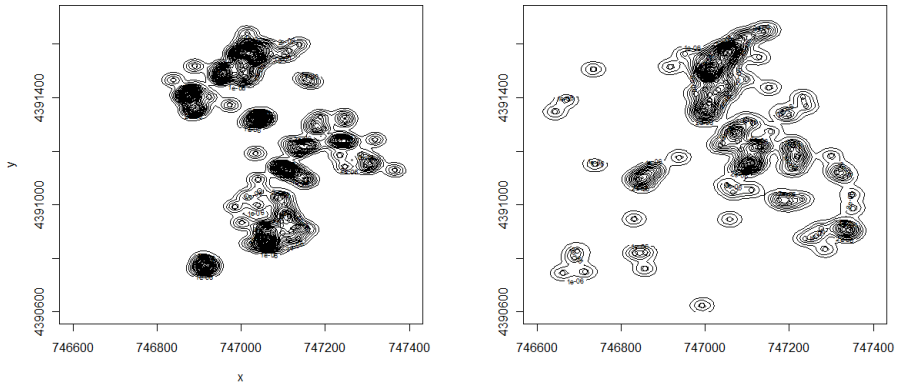


Figure 3.16: Contour plots of the adaptive kernel density estimates using smoothing parameters selected by least-squares cross-validation, with the points in the data sets separated using (3.19). In each row, the left panel shows the WG1 data set, and the right panel shows the WG2 data set.

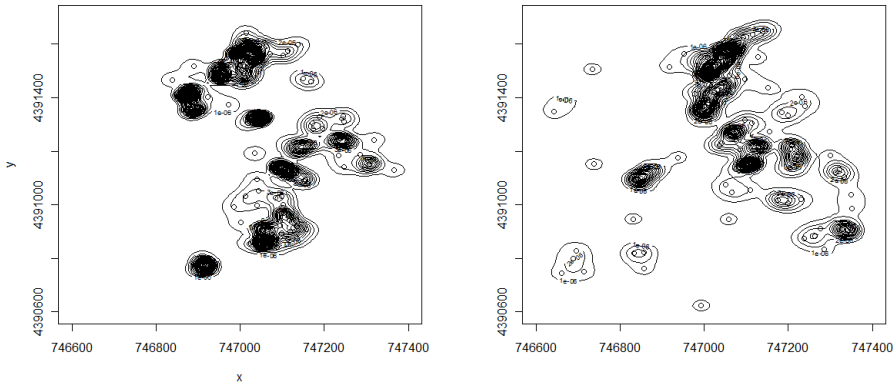


Figure 3.17: Contour plots of the difference between density estimates obtained using fixed and adaptive kernel density estimators with smoothing parameters selected by least-squares cross-validation for the WG1 and WG2 data sets. The left panel shows the fixed kernel density estimates, and the right panel the adaptive kernel density estimates. The points in the data sets have been separated using (3.19).

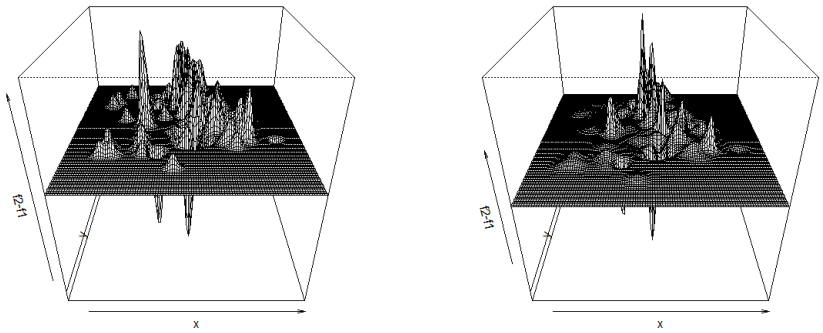


Figure 3.18: Plug-in kernel density estimates for the WG1 and WG2 data sets, displayed in the left and right panels respectively.

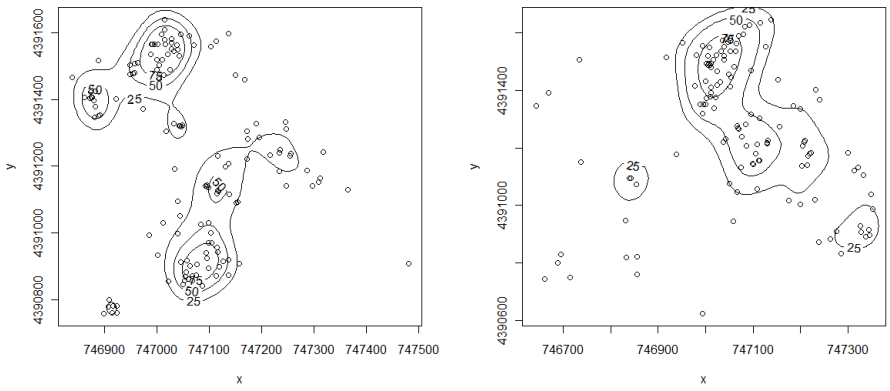


Table 3.17: Summary statistics of the estimated overlap scaled product measure (SP), square root product measure (SRP) and OVL between 200 pairs of data sets. Each pair consists of a set of 144 observations simulated from a mixture of t distributions fitted to the WG1 data set and a set of 135 observations simulated from a mixture of t distributions fitted to the WG2 data set. Presented summary statistics include the mean, standard deviation (SD), bias and upper and lower 5% points. The SP, SRP and OVL were estimated using the estimators shown in (3.6), (3.7) and (3.8) respectively.

	SP	**SRP**	**OVL**
Mean	0.60	0.77	0.58
Median	0.61	0.77	0.58
Standard Deviation	0.067	0.03	0.02
Lower 5% point	0.58	0.75	0.57
Upper 5% point	0.70	0.82	0.62

3.7 Discussion

Several measures of home range fidelity have been compared with one another. The simulation study conducted in Section 3.3 uses the scaled product measure, the square root product measure and the OVL to assess the level of overlap between pairs of data sets simulated from various sample densities. We found that the OVL was more biased than the other two measures in cases of very high or very low overlap, and that for the scaled product measure and square root product measure there was no obvious relationship between accuracy and level of overlap. However, when the simulated data sets were generated by a mixture of bivariate t distributions, and a mixture of bivariate normal distributions was used to produce a density estimate, the estimated square root product measures for the fitted densities were extremely high. These values were clearly incorrect, and indicate that the mixture of normal distributions fails to represent the data well. This corroborates our findings in Chapter 2. With this exception, the estimators used for the distance measures give estimated values with low bias.

The scaled product measure has the advantage of giving rise to tractable calculations to determine its properties. In particular, we were able to calculate the expectation and variance of the scaled product measure for a hypothetical data set generated by a known underlying density. The comparison of expected and true values for pairs of bivariate normal distributions separated by a varying distance, as shown in Section 3.4, indicates substantial bias when the level of overlap is high. For low levels of overlap, the bias is much smaller, but remains non-negligible as a proportion of the expected and true values.

We can use the bias approximation introduced in Section 3.5 to estimate the bias in cases where explicit calculation is more difficult. This approximation provides a suitable estimation for the bias of a density estimate, but it does not necessarily follow that it will remain sufficiently accurate when multiplied by another density and integrated, as is necessary to obtain an estimate for the bias of the scaled product measure. However, applying the resulting bias estimator for the scaled product measure to pairs of data sets simulated from different bivariate normal distributions shows that the approximation is highly accurate for all but extremely high levels of overlap.

In Section 3.6, we applied kernel and mixture models to the WG1 and WG2 data sets and used the previously defined estimators for the three distance measures of interest. The estimated values for the OVL were close to the true values obtained by numerical integration in all cases, and were far more quickly calculated. The estimators for the scaled product measure and square root

product measure were also accurate for the mixture model, but somewhat less so for the kernels. The kernel using least-squares cross-validation failed to produce appropriate results, and the resulting density estimate was effectively equal to the data points. The kernel density estimator with plug-in smoothing parameter and the mixture of bivariate t distributions used provided suitable models for the data.

Chapter 4

Diffusion modelling of home range data

4.1 Introduction

In Chapter 2, we used mixtures of bivariate normal and t distributions for modelling home range data. We found that mixtures of bivariate t distributions provide a useful model for animal movement data, and are more robust to outliers than bivariate normal mixtures. However, these mixture models are to some extent a simplification of the underlying movement process. In particular, each observation is assumed to be generated from the same mixture of densities, irrespective of the position of the previous observation. This is not problematic when locations are recorded at sufficiently long time intervals, but when the time between observations is shorter then the presence of substantial autocorrelation is inevitable, and the mixture model does not take this into account. As such, while the model gives an approximation of the overall area utilized by an animal, it does not describe the path of its movement. In practice, animal movement data are frequently correlated (Fieberg et al., 2010), due to physical constraints on speed and location (Verwaijen and van Damme, 2008) as well as correlated behavioural states (Cushman et al., 2005; Fryxell et al., 2008; Haydon et al., 2008). Autocorrelation can be accounted for using continuous-time models such as a bivariate Ornstein-Uhlenbeck (BOU) process, which is shown in equation (1.1), (Dunn and Gipson, 1977; Worton, 1995a; Blackwell, 1997; Nations and Anderson-Sprecher, 2006) and correlated random walks (Turchin, 1998; Jonsen et al., 2005; Johnson et al., 2008a). Here, we consider modelling the tracks of animals using a BOU process.

We first introduce a data set comprised of the movements of a coyote, recorded

in multiple subsets of observations separated by 15-minute intervals, with longer time intervals between subsets. The short time intervals between points indicate strong autocorrelation. We apply the mixture models described in Chapter 2 to this data set, and comment on the results. The BOU process is then defined and applied to the data. Maximum likelihood estimators are obtained for the parameters of the model, and we use MCMC methods to obtain posterior distributions for the parameters. It is observed that the coyote frequently remains stationary or moves a very short distance between observations, while much longer movements are also common. As an alternative model, we propose a mixture of 2 BOU processes, one corresponding to stationary observations and small movements and the other to larger movements. Posterior distributions are obtained using MCMC methods, and the results are discussed.

4.2 Coyote movement data

The data set considered here consists of the movements of a 5-year old male coyote equipped with a transmitter in the spring of 1970. Nine "bursts" of data were recorded, each comprised of observations separated by intervals of 15 minutes. This data set was previously examined by Dunn and Gipson (1977) and is here referred to as the DG data set. The optimal mixture of normal distributions fitted to the DG data set with `mclust` and a mixture of t distributions fitted using the procedure MIX-T-VAR defined in Chapter 2 are shown in Figure 4.1. The mixture of normal distributions has selected an overly complex fitted model, with mixture components of excessively low variance created by small clusters of data points. The mixture fitted using MIX-T-VAR has avoided this problem and appears appropriate, with 3 components. However, the positions of the components may be misleading. While the mixture components do correspond to the areas in which most observations were recorded, the animal may have spent significant time in other areas while its position was not being recorded.

The mixture of t distributions may not be suitable for the DG data set, as the short time intervals between observations in each burst suggest significant autocorrelation which the mixture model does not account for. Furthermore, as the data are recorded in short bursts separated by longer time intervals, the data set does not necessarily describe the animal's overall use of the area. In effect, the animal's position is only recorded for a small portion of the total time period of the study, so its presence is only recorded in the areas it was utilizing at the particular times of the observations, which may not cover its entire home range.

Figure 4.1: Density contour plots for the mixtures of normal distributions fitted to the DG data set by `mclust` and MIX-T-VAR, shown in the left and right panels respectively. This data set consists of location observations of a male coyote in the spring of 1970. Nine "bursts" of observations were recorded, with successive observations in each burst separated by 15 minutes.

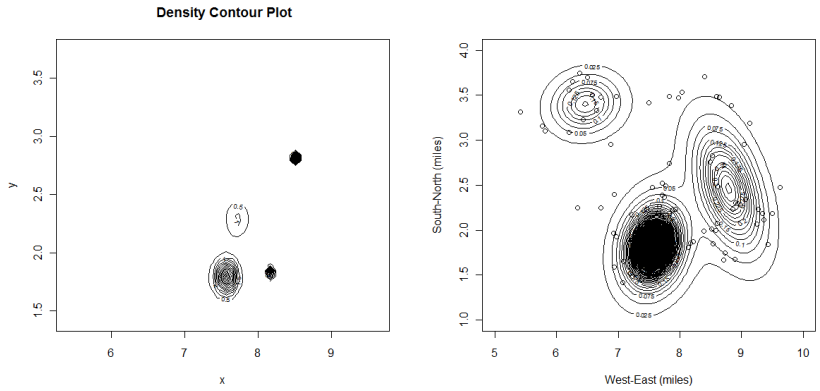
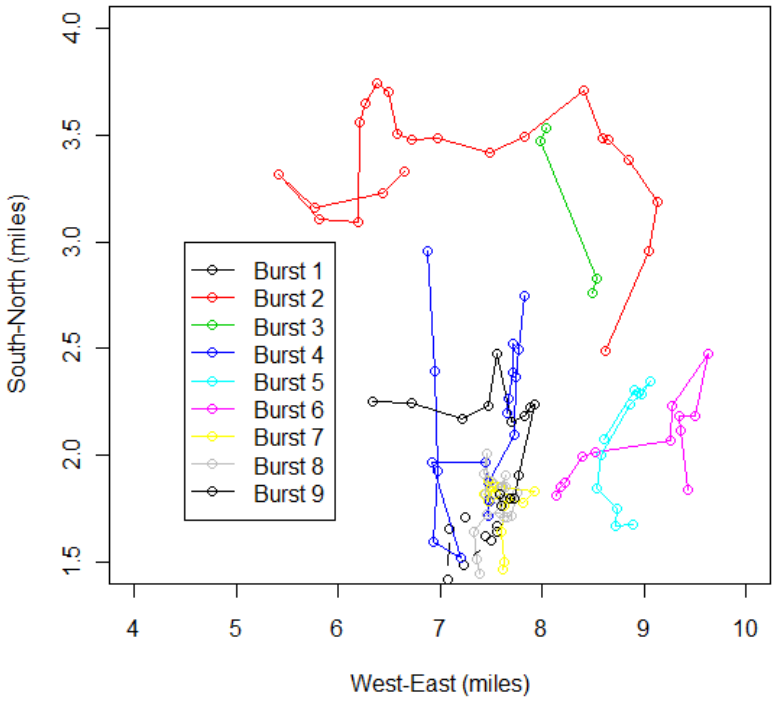


Figure 4.2: Points in the DG data set, with successive observations within each burst linked by lines.



A rough estimate of the animal's movements, constructed from the observed data points, is displayed in Figure 4.2, and it is apparent that there are areas within the range of its movements that it is not recorded passing through, but may well have done so over the time period of the study.

4.3 The bivariate Ornstein-Uhlenbeck process

We assume that the animal's movements follow a BOU process as defined by (1.1). Given a data set consisting of n bursts of observations, $\{\mathbf{x}_{1,0}, \ldots, \mathbf{x}_{1,m_1}\}, \ldots, \{\mathbf{x}_{n,0}, \ldots, \mathbf{x}_{n,m_n}\}$, let $\mathbf{x}_{i,j}$ denote the jth observation from the ith burst of data. The first observation in each burst of data is assumed to have the distribution

$$\mathbf{x}_{i,0} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}),$$

and each subsequent observation is assumed to have a distribution of the form

$$\begin{aligned} \mathbf{x}_{i,j}|\mathbf{x}_{i,j-1} &\sim \mathrm{N}\{\boldsymbol{\mu} + e^{\mathbf{B}t}(\mathbf{x}_{i,j-1} - \boldsymbol{\mu}), \boldsymbol{\Lambda} - e^{\mathbf{B}t}\boldsymbol{\Lambda}e^{\mathbf{B}^T t}\} \\ &= \mathrm{N}\{\boldsymbol{\Gamma}\mathbf{x}_{i,j-1} + (\mathbf{I} - \boldsymbol{\Gamma})\boldsymbol{\mu}, \boldsymbol{\Phi}\}, \end{aligned} \tag{4.1}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is a point of attraction. The parameters $\boldsymbol{\Lambda}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are defined as

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} = e^{\mathbf{B}t}, \ \boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = \boldsymbol{\Lambda} - \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T.$$

Here, $\boldsymbol{\Gamma}$ is a centralization matrix determining the strength of attraction to $\boldsymbol{\mu}$, and $\boldsymbol{\Phi}$ is the covariance matrix of the animal's movement at each step. When $\boldsymbol{\Gamma} = \mathbf{I}$, the movement is independent of $\boldsymbol{\mu}$, and at the other extreme, when $\boldsymbol{\Gamma} = 0$, the movement is dependent only on $\boldsymbol{\mu}$ and not on the previous observation. As the time interval between observations in a burst is constant at $t = 15$ minutes, the parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are constant for all observations.

4.4 Application to coyote movement data

The approach of Dunn and Gipson (1977) was to develop the likelihood equations for the BOU process and solve them to yield maximum likelihood estimators. If it is assumed that $\mathbf{x}_{i,0}$ is a known constant for each burst, the log likelihood of the data, up to an additive constant, is equal to

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{-\frac{1}{2}\ln|\boldsymbol{\Phi}| - \frac{1}{2}\left(\mathbf{x}_{i,j}-\boldsymbol{\nu}-\boldsymbol{\Gamma}\mathbf{x}_{i,j-1}\right)^T\boldsymbol{\Phi}^{-1}\left(\mathbf{x}_{i,j}-\boldsymbol{\nu}-\boldsymbol{\Gamma}\mathbf{x}_{i,j-1}\right)\right\}, \quad (4.2)$$

where $\boldsymbol{\nu} = (\mathbf{I}-\boldsymbol{\Gamma})\,\boldsymbol{\mu}$. Maximum likelihood estimates for the parameters can be shown to be

$$\hat{\boldsymbol{\Gamma}} = \left\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{x}_{i,j}\mathbf{x}_{i,j-1}^T - \frac{1}{n}\left(\sum_{i=1}^{m}\sum_{j=1}^{m_i}\mathbf{x}_{i,j}\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{m_i}\mathbf{x}_{i,j-1}\right)^T\right\} \times\left\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathbf{x}_{i,j-1}\mathbf{x}_{i,j-1}^T - \frac{1}{n}\left(\sum_{i=1}^{n}\sum_{j=1}^{m_1}\mathbf{x}_{i,j-1}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{m_1}\mathbf{x}_{i,j-1}\right)^T\right\}^{-1}, \quad (4.3)$$

$$\hat{\boldsymbol{\nu}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\mathbf{x}_{i,j}-\hat{\boldsymbol{\Gamma}}\mathbf{x}_{i,j-1}\right), \quad (4.4)$$

$$\hat{\boldsymbol{\mu}} = (\mathbf{I}-\hat{\boldsymbol{\Gamma}})^{-1}\hat{\boldsymbol{\nu}}, \quad (4.5)$$

$$\hat{\boldsymbol{\Phi}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\mathbf{x}_{i,j}-\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\Gamma}}\mathbf{x}_{i,j-1}\right)\left(\mathbf{x}_{i,j}-\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\Gamma}}\mathbf{x}_{i,j-1}\right)^T, \quad (4.6)$$

$$\mathrm{vec}(\hat{\boldsymbol{\Lambda}}) = \left\{\mathbf{I}-(\hat{\boldsymbol{\Gamma}}\otimes\hat{\boldsymbol{\Gamma}})\right\}^{-1}\mathrm{vec}(\hat{\boldsymbol{\Phi}}), \quad (4.7)$$

following the derivation given in Anderson (1971, pp. 183–185). In the case where $\mathbf{x}_{i,0}$ is not a known constant, Dunn and Gipson assumed that it was sampled from the equilibrium distribution

$$\mathbf{x}_{i,0} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}).$$

The addition of the probability density function of $\mathbf{x}_{i,0}$ modifies the log likelihood shown in (4.2) by the addition of the following term,

$$\sum_{i=1}^{n}\left\{-\frac{1}{2}\ln|\boldsymbol{\Lambda}| - \frac{1}{2}\left(\mathbf{x}_{i,0}-\boldsymbol{\mu}\right)^T\boldsymbol{\Lambda}^{-1}\left(\mathbf{x}_{i,0}-\boldsymbol{\mu}\right)\right\},$$

which as described by Dunn and Gipson results in the following likelihood

equations,

$$\boldsymbol{\Lambda}^{-1}\sum_{i=1}^{n}(\mathbf{x}_{i,0}-\boldsymbol{\mu})+(\mathbf{I}-\boldsymbol{\Gamma}^{T})\boldsymbol{\Phi}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}=0, \quad (4.8)$$

$$-\sum_{i=1}^{n}\left[\sum_{j=1}^{m_i}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}^{T}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda}$$
$$+\boldsymbol{\Gamma}\boldsymbol{\Lambda}\sum_{i=1}^{n}m_i+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})^{T}=0, \quad (4.9)$$

$$-\boldsymbol{\Gamma}^{T}\boldsymbol{\Phi}^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}^{T}\right.$$
$$\left.-\sum_{i=1}^{n}m_i\boldsymbol{\Phi}\right]\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}$$
$$+\boldsymbol{\Phi}^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}\{\mathbf{x}_{i,j}-\boldsymbol{\mu}-\boldsymbol{\Gamma}(\mathbf{x}_{i,j-1}-\boldsymbol{\mu})\}^{T}\right.$$
$$\left.-\sum_{i=1}^{n}m_i\boldsymbol{\Phi}\right]\boldsymbol{\Phi}^{-1}$$
$$+\boldsymbol{\Lambda}^{-1}\left\{\sum_{i=1}^{n}(\mathbf{x}_{i,0}-\boldsymbol{\mu})(\mathbf{x}_{i,0}-\boldsymbol{\mu})^{T}-n\boldsymbol{\Lambda}\right\}\boldsymbol{\Lambda}^{-1}=0. \quad (4.10)$$

Dunn and Gipson then used a composite estimation procedure which required the calculation of the information matrix generated by each observation with respect to the vector of parameters $\{\boldsymbol{\mu}^{T},\mathrm{vec}(\boldsymbol{\Gamma}^{T}),\mathrm{vec}(\boldsymbol{\Lambda}_{T}^{T})\}$, where $\boldsymbol{\Lambda}_{T}$ is the upper triangular portion of $\boldsymbol{\Lambda}$. The information matrix generated by an initial observation of a burst is

$$F_I=\begin{pmatrix}F_{I\boldsymbol{\mu}} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & F_{I\boldsymbol{\Lambda}}\end{pmatrix},$$

where

$$F_{I\boldsymbol{\mu}}=\boldsymbol{\Lambda}^{-1},$$
$$F_{I\boldsymbol{\Lambda}}=\frac{1}{2}T(\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda})^{-1}T^{T},$$

where $\otimes$ is the Kronecker product (Neudecker, 1969) and

$$T=\frac{\partial\mathrm{vec}(\boldsymbol{\Lambda})}{\partial\mathrm{vec}(\boldsymbol{\Lambda}_{T})}.$$

For successive observations, the information matrix is

$$F_S = \begin{pmatrix} F_{S\boldsymbol{\mu}} & 0 & 0 \\ 0 & F_{S\boldsymbol{\Gamma}} & F_{S\mathbf{J}}^T \\ 0 & F_{S\mathbf{J}} & F_{S\boldsymbol{\Lambda}} \end{pmatrix},$$

where

$$\begin{aligned}
F_{S\boldsymbol{\mu}} &= (\mathbf{I} - \boldsymbol{\Gamma}^T)\boldsymbol{\Phi}^{-1}(\mathbf{I} - \boldsymbol{\Gamma}), \\
F_{S\boldsymbol{\Gamma}} &= \boldsymbol{\Lambda} \otimes \boldsymbol{\Phi}^{-1} + (\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda} \otimes \boldsymbol{\Phi}^{-1}) + (\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Lambda} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma}^T\boldsymbol{\Phi}^{-1})_{(p)}, \\
F_{S\mathbf{J}} &= -T\left\{\mathbf{I} - (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\right\}^T (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Lambda} \otimes \mathbf{I}), \\
F_{S\boldsymbol{\Lambda}} &= \frac{1}{2}T\left\{\mathbf{I} - (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\right\}^T (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1} \left\{\mathbf{I} - (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\right\} T^T,
\end{aligned}$$

and where the symbol $A_{(p)}$ indicates a p-cycle permutation of the rows of the matrix A (Tracy and Dwyer, 1969; Dunn and Gipson, 1977).

The estimator $\hat{\boldsymbol{\Gamma}}$ in (4.3) is taken as is, as the initial observations in each burst provide no further information about $\boldsymbol{\Gamma}$. The estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Lambda}}$ in (4.5) and (4.7) are now referred to as $\hat{\boldsymbol{\mu}}_S$ and $\hat{\boldsymbol{\Lambda}}_S$. The unbiased estimators from the initial observations are

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_I &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_{i,0}, \\
\hat{\boldsymbol{\Lambda}}_I &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} \mathbf{x}_{i,0}\mathbf{x}_{i,0}^T - \frac{1}{n}\left(\sum_{i=1}^{n} \mathbf{x}_{i,0}\right)\left(\sum_{i=1}^{n} \mathbf{x}_{i,0}\right)^T\right\},
\end{aligned}$$

giving rise to the weighted estimator

$$\mathrm{vec}(\hat{\boldsymbol{\Lambda}}_T) = (nF_{I\boldsymbol{\Lambda}} + F_{S\boldsymbol{\Lambda}}\sum_{i=1}^{n} m_i)^{-1}\left\{nF_{I\boldsymbol{\Lambda}}\mathrm{vec}(\hat{\boldsymbol{\Lambda}}_{I_T}) + F_{S\boldsymbol{\Lambda}}\mathrm{vec}(\hat{\boldsymbol{\Lambda}}_{S_T})\sum_{i=1}^{n} m_i\right\}.$$

Here, $F_{I\boldsymbol{\Lambda}}$ and $F_{S\boldsymbol{\Lambda}}$ have been iterated until convergence is attained. Given the resulting estimate $\hat{\boldsymbol{\Lambda}}$ of $\boldsymbol{\Lambda}$, we obtain a corresponding estimate $\hat{\boldsymbol{\Phi}}$ for $\boldsymbol{\Phi}$ using the relation $\boldsymbol{\Phi} = \boldsymbol{\Lambda} - \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$, and define

$$\hat{F}_{S\boldsymbol{\mu}} = \left(\mathbf{I} - \hat{\boldsymbol{\Gamma}}^T\right)\hat{\boldsymbol{\Phi}}^{-1}\left(\mathbf{I} - \hat{\boldsymbol{\Gamma}}\right).$$

Given $\hat{\boldsymbol{\Lambda}}$ and $\hat{F}_{S\boldsymbol{\mu}}$, (4.8) can be solved to obtain the following weighted estimate

of $\boldsymbol{\mu}$ (Dunn and Gipson, 1977),

$$\hat{\boldsymbol{\mu}} = (n\hat{\boldsymbol{\Lambda}}^{-1} + \hat{F}_{S\boldsymbol{\mu}} \sum_{i=1}^{n} m_i)^{-1}(n\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{\mu}}_I + \hat{F}_{S\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}_S \sum_{i=1}^{n} m_i).$$

The resulting parameter estimates for the DG data set are shown in Table 4.1. However, we can instead implement a Bayesian approach by using Markov Chain Monte Carlo (MCMC) methods to obtain posterior distributions for the parameters of the model (Casella and George, 1992, Gilks et al., 1996; West and Harrison, 1997; Leonard and Hsu, 1999; Robert and Casella, 2004). For simplicity, it is now assumed that $\boldsymbol{\Gamma}$ is equal to a scalar multiple of $\mathbf{I}$.

Table 4.1: Parameter estimates of the parameters of the BOU process for the DG data set obtained using the composite estimation procedure of Dunn and Gipson (1977).

Parameter	**Estimate**
μ_1	7.7540
μ_2	2.1310
λ_{11}	0.4256
λ_{12}	0.1361
λ_{22}	0.3585
γ_{11}	0.9478
γ_{12}	0.0416
γ_{21}	−0.0059
γ_{22}	0.9690
ϕ_{11}	0.0347
ϕ_{12}	0.0000
ϕ_{22}	0.0234

Posterior distributions for the parameters of the BOU process fitted to the DG data set were estimated with a MCMC approach using WinBUGS and OpenBUGS (Lunn et al., 2000). The following independent prior distributions were used for the parameters of the model,

$$\boldsymbol{\mu} \sim \mathrm{N}\left\{(0,0)^T, 10^2\mathbf{I}\right\}, \tag{4.11}$$

$$\boldsymbol{\Gamma} = \gamma\mathbf{I}, \tag{4.12}$$

$$\gamma \sim \mathrm{N}\left(1, 10^2\right), \tag{4.13}$$

$$\boldsymbol{\Phi} \sim \text{Inverse Wishart}\left(10^{-2}\mathbf{I}, 2\right), \tag{4.14}$$

and provide vague prior information.

Summary statistics for the posterior distributions of the model parameters for the data are presented in Table 4.2. The posterior means of μ_1 and μ_2 are very similar to the estimates in Table 4.1. The posterior means of ϕ_{11} and ϕ_{22}, and thus of λ_{11} and λ_{22}, are slightly greater than the estimates in Table 4.1, suggesting somewhat larger movements, while the posterior mean of γ is quite close to the estimates of γ_{22} and, to a lesser extent, γ_{11}.

Table 4.2: Summary statistics of posterior distributions of the parameters of the BOU process for the DG data set. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (4.11), (4.12), (4.13) and (4.14).

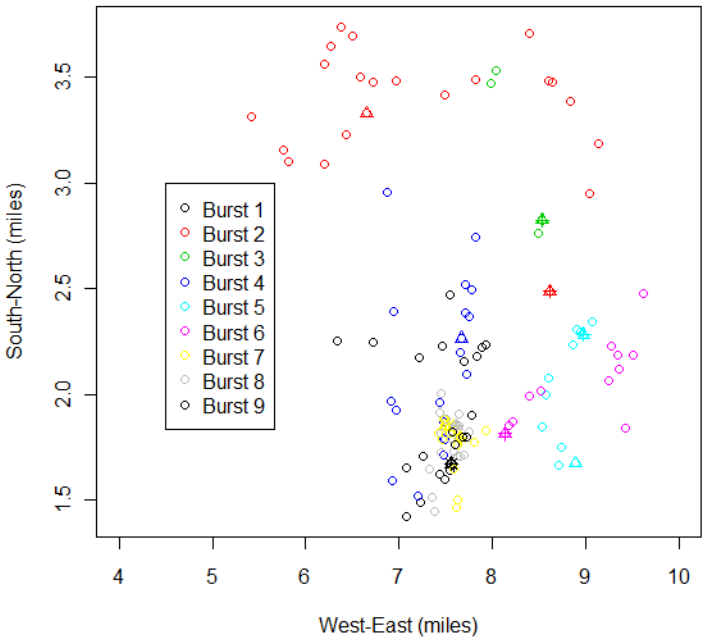
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
μ_1	7.8100	7.8108	0.2283	7.3590	8.2613
μ_2	2.1647	2.1642	0.1852	1.8051	2.5389
λ_{11}	0.5601	0.5282	0.1642	0.3381	0.9826
λ_{12}	−0.0001	−0.0001	0.0384	−0.0756	0.0781
λ_{22}	0.3886	0.3661	0.1163	0.2313	0.6796
γ	0.9673	0.9676	0.0083	0.9497	0.9825
ϕ_{11}	0.0338	0.0334	0.0039	0.0271	0.0422
ϕ_{12}	0.0000	0.0000	0.0022	−0.0044	0.0044
ϕ_{22}	0.0234	0.0232	0.0027	0.0188	0.0292

4.5 A mixture of two BOU components

The BOU process implemented in Section 4.2 assumes that the movements of the animal are all generated by a single diffusion process, and therefore will always have the same expected step length. However, this assumption will not be correct if the animal has multiple distinct behavioural states, for example, if it spends some time intervals resting and others hunting. Furthermore, the frequent occurrence of consecutive observations at the same location is apparent upon inspection of Figure 4.3, which denotes repeated observations at the same location with different point types.

An alternative model would be a mixture with at least 2 components, one "rest" state corresponding to observations at which the animal remained stationary or moved a very short distance, and one "active" state for observations at which it did not. We assume 2 states, and that the animal's movements in state $k \in \{1, 2\}$ follow a BOU process with common point of attraction $\boldsymbol{\mu}$, and state-dependent parameters $\boldsymbol{\Gamma}_k$, $\boldsymbol{\Lambda}_k$ and $\boldsymbol{\Phi}_k$. The first observation in each burst

Figure 4.3: Points in the DG data set. Observations at the same location are denoted with a different point type for every such successive observation after the first.



of data is therefore assumed to have the distribution

$$\mathbf{x}_{i,0} \sim \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k),$$

and each subsequent observation is assumed to have the distribution

$$\mathbf{x}_{i,j}|\mathbf{x}_{i,j-1} \sim \mathrm{N}\{\boldsymbol{\mu}_k + \boldsymbol{\Gamma}_k(\mathbf{x}_{i,j-1} - \boldsymbol{\mu}_k), \boldsymbol{\Phi}_k\},$$

where k is unknown and must be estimated. The model includes a parameter $0 < \pi < 1$ representing the probability of an observation being generated from component $k = 1$.

Summary statistics for the mixture of 2 BOU processes fitted to the DG data set are presented in Table 4.3. The following vague priors were used,

$$\boldsymbol{\mu}_k \sim \mathrm{N}\left\{(0,0)^T, 10^2\mathbf{I}\right\}, \quad k = 1, 2, \tag{4.15}$$

$$\Gamma_k = \gamma_k \mathbf{I}, \quad k = 1, 2, \tag{4.16}$$

$$\gamma_k \sim \mathrm{N}\left(1, 10^2\right), \quad k = 1, 2, \tag{4.17}$$

$$\mathbf{\Phi}_k \sim \text{Inverse Wishart}\left(10^{-2}\mathbf{I}, 2\right), \quad k = 1, 2, \tag{4.18}$$

$$(\pi, 1-\pi) \sim \text{Dirichlet}\left(\frac{1}{2}, \frac{1}{2}\right). \tag{4.19}$$

The estimated point of attraction $\boldsymbol{\mu}$ is very close to the corresponding point for the single BOU process shown in Table 4.2. The posterior means of γ_1 and γ_2 indicate that component 2 of the model contains the movements that are more strongly attracted towards $\mathbf{a}$, while component 1 consists of more localized movements. These localized movements are the stationary observations and small steps in the data set, as shown by the posterior means of the elements of Φ_1 which are much smaller than those of the elements of Φ_2. The value of π shows that approximately 40% of the animal's movements are assigned to component 1.

Table 4.3: Summary statistics of posterior distributions of the parameters of the mixture of 2 BOU processes for the DG data set. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and π are as given in (4.15), (4.16), (4.17), (4.18) and (4.19).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
$\mu_{1,1}$	7.8240	7.8258	0.1679	7.5190	8.9140
$\mu_{1,2}$	2.1647	2.1642	0.1711	1.7170	2.2560
$\lambda_{1,11}$	0.0635	0.0600	0.0355	0.0004	0.1458
$\lambda_{1,12}$	0.0508	0.0466	0.0290	0.0000	0.1179
$\lambda_{1,22}$	0.1129	0.1083	0.0584	0.0003	0.2487
$\lambda_{2,11}$	0.6630	0.6170	0.2059	0.3793	1.1800
$\lambda_{2,12}$	−0.0125	−0.0115	0.0607	−0.1324	0.1083
$\lambda_{2,22}$	0.4498	0.4222	0.1419	0.2511	0.8122
γ_1	0.9909	0.9910	0.0043	0.9826	1.0000
γ_2	0.9534	0.9542	0.0118	0.9287	0.9764
$\phi_{1,11}$	0.0011	0.0011	0.0006	0.0000	0.0025
$\phi_{1,12}$	0.0009	0.0009	0.0004	0.0000	0.0017
$\phi_{1,22}$	0.0019	0.0020	0.0008	0.0000	0.0034
$\phi_{2,11}$	0.0569	0.0559	0.0108	0.0374	0.0818
$\phi_{2,12}$	−0.0012	−0.0011	0.0052	−0.0126	0.0094
$\phi_{2,22}$	0.0386	0.0380	0.0073	0.0256	0.0545
π	0.3984	0.4128	0.0850	0.1557	0.5124

4.6 Discussion

Diffusion processes such as the BOU process provide a way to model the tracks of animals, as opposed to simply estimating the utilization density as was done in Chapter 2. However, the single BOU process introduced in Section 4.3 is potentially overly simplistic, as it assumes that all of the animal's movements are generated by the same process. We can account for different modes of movement by using a mixture of multiple diffusion processes. In particular, the 2 component mixture introduced in Section 4.5 models the frequent stationary observations and extremely small movements in the DG data set, as well as the much larger movements present in the data.

The results presented in Table 4.3 suggest that the mixture represents the data well, as the posterior means obtained for the elements of the centralization matrices and covariance matrices of the 2 components indicate that component 1 consists of the stationary observations and very small movements present in the data set, while component 2 accounts for the frequent larger movements. In Chapters 5 and 6, we will conduct a far more extensive investigation of more complex and flexible diffusion processes, for the purpose of application to larger and more strongly correlated data sets.

Chapter 5

Diffusion modelling of bioassay data

5.1 Introduction

5.1.1 Background

As described in Section 1.4, the larva of the cabbage root fly (*Delia radicum* L.) is a pest that damages *Brassica* plants by feeding on their roots (an example of this damage is shown in Figure 5.1), and novel treatment methods are currently being investigated, including the use of repellent chemical compounds to protect *Brassica* crops from the larvae. Often, studies involve collecting data on the locations of larvae after a given time and using standard statistical analysis of circular data, but here the more challenging problem of modelling the tracks of the larvae is considered. The former method gives only one locational observation per bioassay, whereas the number of observations obtained by the latter approach is in the thousands and thus new approaches are required to model these highly correlated observations. Here, suitable models for such data are investigated, together with appropriate methods of statistical analysis. The aim is then to use the parameters of the models as a way of summarizing the complex patterns of the tracks, and thus give a greater understanding of the behaviour of the larvae.

5.1.2 Data collection

Bioassays were conducted at The James Hutton Institute in a research project concerned with developing novel approaches to pest management of cabbage root fly. In each bioassay, a newly hatched neonate cabbage root fly larva was placed in an arena within a 9cm diameter petri dish half filled with agar, with a zone of

Figure 5.1: Plant damage caused by the larvae of the cabbage root fly (*Delia radicum* L). This image has been reproduced from `http://www.dunchurchallotments.org.uk/other-news.html`.

chemical compound on one side and a control zone on the other. The positions of the larva were then detected by infrared camera (Sanyo) and recorded using the EthoVision 3.1 software system (Noldus et al., 2001) at intervals of 0.2 seconds for 30 minutes, giving a nominal total of 9000 observations for each bioassay. The variables recorded at each observation were the position of the larva, distance from the origin, distance to the border of the zone of chemical compound, total distance moved, turning angle, velocity, angular velocity and meander (amount of turning per unit of distance moved). If the larva entered one of the zones, the bioassay was terminated. The infrared camera was placed in a long covered pipe to minimize background infrared readings, and the data were collected in the dark for the same reason. The camera and covered pipe are shown in Figure 5.2.

Figure 5.3 displays plots of the tracks for 5 bioassays. In Bioassays 1–3 and 5, the upper solid circle referred to in the plots as **a** corresponds to a zone of damaged broccoli roots. As a suitable host plant, these roots are hypothesised to act as an attractant. In Bioassay 4, **a** is a zone of allyl isothiocyanate from which the larva is repelled. The lower open circle in each plot referred to as **b** is the control zone.

Figure 5.2: The infrared camera used to record the cabbage root fly larva bioassay data shown in Figure 5.3, placed in a covered pipe to minimize background infrared readings and connected to a PC with EthoVision software. This image was provided by William Deasy.

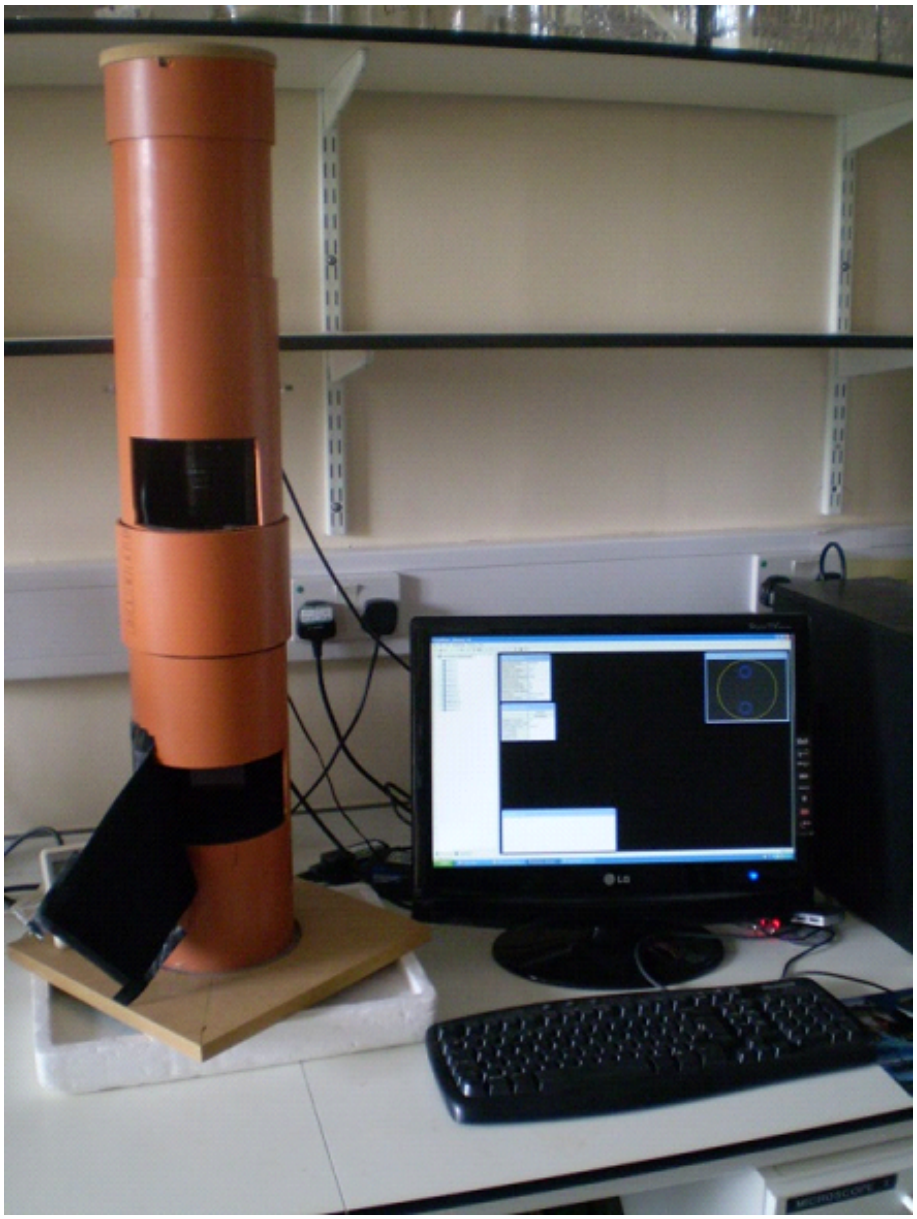
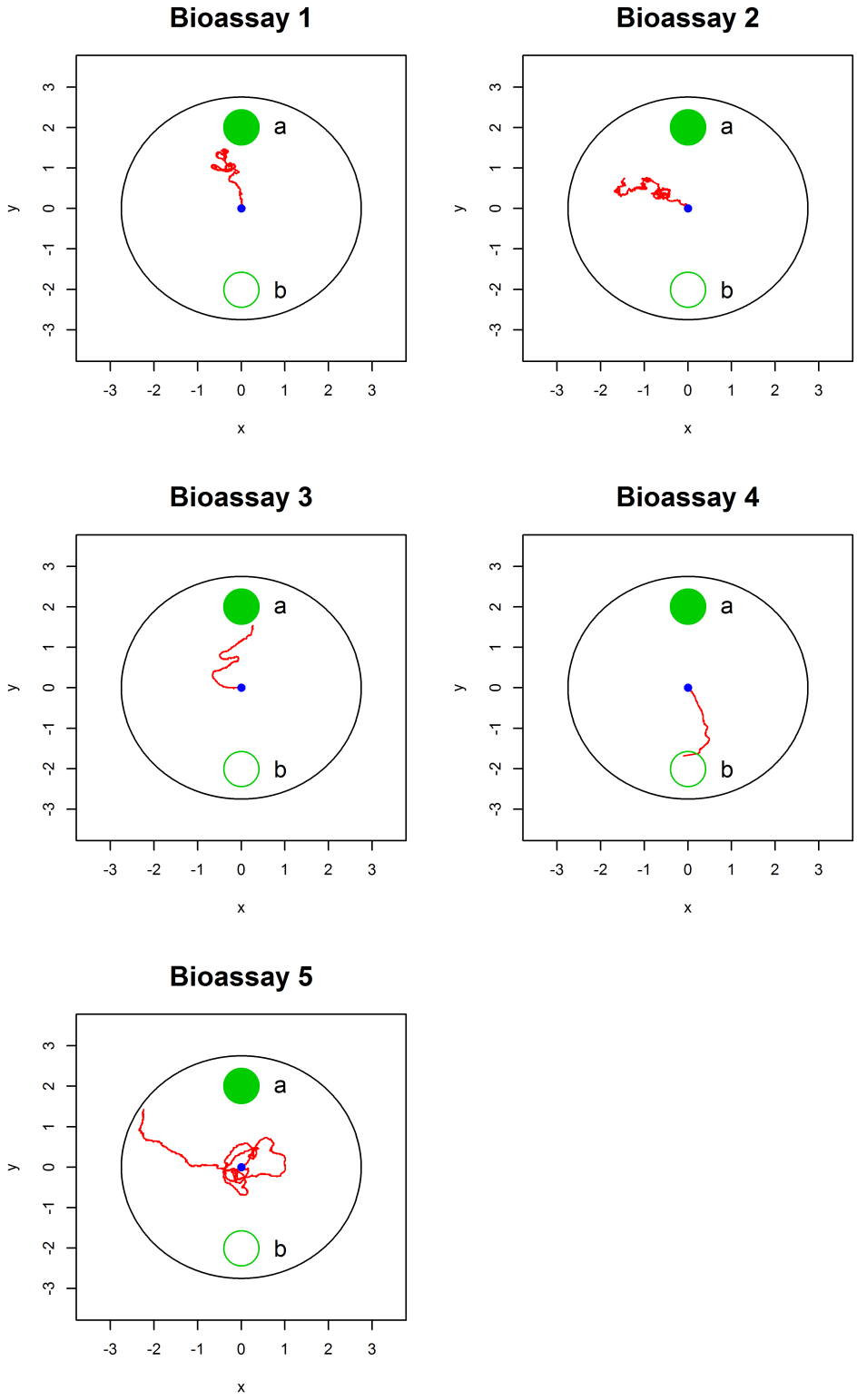


Figure 5.3: Tracks of cabbage root fly larvae for 5 bioassays. Each track starts at the origin (small dot) and the location of the larva is recorded every 0.2 seconds for 30 minutes using EthoVision 3.1. Each bioassay has a nominal 9,000 observations. The outer circle is the arena, and the upper solid circles denoted by **a** and lower open circles denoted by **b** are the attractant (repellent for Bioassay 4) and control regions respectively.



5.2 Diffusion modelling of larvae tracks

5.2.1 A simple diffusion process

We initially propose a diffusion process using a bivariate normal distribution. Under this process, the conditional distribution of the larva's position $\mathbf{X}_{s+t}$ at time $s+t$ given its previous position $\mathbf{X}_s$ at time s is given by

$$\mathbf{X}_{s+t}|\mathbf{X}_s \sim \mathrm{N}\left\{\boldsymbol{\mu} + \boldsymbol{\Gamma}\left(\mathbf{X}_s - \boldsymbol{\mu}\right), \boldsymbol{\Phi}\right\}, \tag{5.1}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is a point of attraction or repulsion, and the matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are defined as

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}, \ \boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

Here, $\boldsymbol{\Gamma}$ is a centralization matrix determining the strength of attraction to or repulsion from $\boldsymbol{\mu}$, and $\boldsymbol{\Phi}$ is the covariance matrix of the larva's movement at each observation step. As the larva is placed in the arena by the experimenter at the start of each bioassay, its initial position is a known constant.

Note that diffusion process (5.1) is similar to the BOU process (4.1) introduced in Chapter 4. However, process (4.1) assumes that $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are linked by the relation $\boldsymbol{\Phi} = \boldsymbol{\Lambda} - \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$, where $\boldsymbol{\Lambda}$ is the covariance matrix of the bivariate normal equilibrium distribution of $\mathbf{X}_s$. In the current context, an equilibrium distribution may not exist and this assumption is therefore relaxed within the modelling framework.

Under the BOU process, $\boldsymbol{\Gamma}$ is dependent on the time t between successive observations, as $\boldsymbol{\Gamma} = e^{\mathbf{B}t}$ for a matrix $\mathbf{B}$, where $e^{\mathbf{B}t}$ is the matrix exponential. The matrix $\boldsymbol{\Phi}$ is dependent on t as well, but as the time between observations is fixed at 0.2 seconds $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are constant throughout each bioassay and thus dependence on time is not explicitly included in the notation above. As the time interval between observations is constant, the conditional distribution of an observation $\mathbf{x}_{i+1}$ given the previous observation $\mathbf{x}_i$ is

$$\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\left\{\boldsymbol{\mu} + \boldsymbol{\Gamma}\left(\mathbf{x}_i - \boldsymbol{\mu}\right), \boldsymbol{\Phi}\right\}. \tag{5.2}$$

When $\boldsymbol{\Gamma}$ is equal to the identity matrix, the conditional distribution of $\mathbf{x}_{i+1}$ given $\mathbf{x}_i$ is not dependent on $\boldsymbol{\mu}$ and the process corresponds to bivariate Brownian motion. If $\boldsymbol{\Gamma}$ is not equal to the identity matrix and the non-diagonal elements of $\boldsymbol{\Gamma}$ are equal to 0, a value of γ_{11} that is less than unity indicates that the x-

coordinates of the larva's position move towards the x-coordinate of $\boldsymbol{\mu}$, while a value greater than unity implies repulsion. The parameter γ_{22} is related to the y-coordinates of the larva's position and of $\boldsymbol{\mu}$ in the same way, and the determinant $|\boldsymbol{\Gamma}|$ describes the overall strength of influence of $\boldsymbol{\mu}$. Non-zero values for the non-diagonal elements of $\boldsymbol{\Gamma}$ introduce additional complexity, but the above statements remain applicable provided that the non-diagonal elements are small.

In each bioassay, the arena is homogeneous with the exception of the zone of chemical compound. As such, there is no apparent physical basis for attraction to or repulsion from any specific point $\boldsymbol{\mu}$ which lies outside the zone. With this in mind, an alternative model is considered which replaces the parameter $\boldsymbol{\mu}$ with a constant $\mathbf{a}$ equal to the position of the centre of the zone of chemical compound. Under this model, the conditional distribution of observation $\mathbf{x}_{i+1}$ given the previous observation $\mathbf{x}_i$ is

$$\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\left\{\mathbf{a} + \boldsymbol{\Gamma}\left(\mathbf{x}_i - \mathbf{a}\right), \boldsymbol{\Phi}\right\}, \tag{5.3}$$

where $\mathbf{a}$ is a known point of attraction or repulsion as indicated in Figure 5.3 and $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as defined for diffusion process (5.2). Note that the points $\mathbf{a}$ have x-coordinates close to 0 for all 5 bioassays. The strength of attraction to or repulsion from $\mathbf{a}$ is therefore primarily determined by the parameter γ_{22}.

5.2.2 Likelihood approach

In the bioassays, each larva is released at the origin, $\mathbf{x}_0$, and generates a subsequent observed sample path, $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, with a constant time interval between observations of 0.2 seconds. The log likelihood for diffusion process (5.2), up to an additive constant, is given by

$$-\frac{n}{2}\ln|\boldsymbol{\Phi}| - \frac{1}{2}\sum_{i=1}^{n}\left\{\mathbf{x}_i - \boldsymbol{\mu} - \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \boldsymbol{\mu})\right\}^T \boldsymbol{\Phi}^{-1}\left\{\mathbf{x}_i - \boldsymbol{\mu} - \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \boldsymbol{\mu})\right\}.$$

In simple cases such as diffusion processes (5.2) and (5.3) it is possible to obtain explicit maximum likelihood estimates of the parameters, but for more complex mixed processes the likelihood may be difficult to express analytically. The maximum likelihood estimates of the parameters of diffusion process (5.2) can be shown to be

$$\hat{\boldsymbol{\Gamma}} = \left\{ \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_{i-1} - \frac{1}{n} \left(\sum_{i=1}^{n} \mathbf{x}_i \right) \left(\sum_{i=1}^{n} \mathbf{x}_{i-1} \right)^T \right\}$$
$$\times \left\{ \sum_{i=1}^{n} \mathbf{x}_{i-1} \mathbf{x}_{i-1} - \frac{1}{n} \left(\sum_{i=1}^{n} \mathbf{x}_{i-1} \right) \left(\sum_{i=1}^{n} \mathbf{x}_{i-1} \right)^T \right\}^{-1},$$
$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \left(\mathbf{I} - \hat{\boldsymbol{\Gamma}} \right)^{-1} \sum_{i=1}^{n} \left(\mathbf{x}_i - \hat{\boldsymbol{\Gamma}} \mathbf{x}_{i-1} \right),$$
$$\hat{\boldsymbol{\Phi}} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbf{x}_i - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\Gamma}} (\mathbf{x}_{i-1} - \hat{\boldsymbol{\mu}}) \right\} \left\{ \mathbf{x}_i - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\Gamma}} (\mathbf{x}_{i-1} - \hat{\boldsymbol{\mu}}) \right\}^T,$$

following the derivation given in Anderson (1971, pp. 183–184). Note that the estimates shown above are equivalent to those for the BOU process shown in (4.3), (4.5) and (4.6) in the particular case of a single burst of observations.

For diffusion process (5.3), as $\mathbf{a}$ is a known constant, it is simpler to obtain maximum likelihood estimates than for (5.2) and by following a similar derivation they can be shown to be

$$\hat{\boldsymbol{\Gamma}} = \left\{ \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{a})(\mathbf{x}_{i-1} - \mathbf{a})^T \right\} \left\{ \sum_{i=1}^{n} (\mathbf{x}_{i-1} - \mathbf{a})(\mathbf{x}_{i-1} - \mathbf{a})^T \right\}^{-1},$$
$$\hat{\boldsymbol{\Phi}} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbf{x}_i - \mathbf{a} - \hat{\boldsymbol{\Gamma}} (\mathbf{x}_{i-1} - \mathbf{a}) \right\} \left\{ \mathbf{x}_i - \mathbf{a} - \hat{\boldsymbol{\Gamma}} (\mathbf{x}_{i-1} - \mathbf{a}) \right\}^T.$$

Tables 5.2 and 5.3 give the maximum likelihood estimates of the parameters of diffusion processes (5.2) and (5.3) respectively for Bioassays 1 to 5. In all cases, the data have been transformed so that the initial position of the larva is at the origin. The transformed positions of the centres $\mathbf{a}$ of the zones of chemical compound for each bioassay are given in Table 5.1 for comparison with the maximum likelihood estimates $\hat{\boldsymbol{\mu}}$.

For Bioassays 1–3, $\hat{\boldsymbol{\mu}}$ is in roughly the same direction from the larva's initial position as $\mathbf{a}$, whereas for Bioassay 4 it is in the opposite direction. The position of $\hat{\boldsymbol{\mu}}$ in Bioassay 5 is very close to the larva's initial position. For Bioassays 1, 2 and 4, both diagonal elements of $\hat{\boldsymbol{\Gamma}}$ in Table 5.2 are less than unity, indicating attraction towards $\hat{\boldsymbol{\mu}}$. These results support the hypothesis that the larva is attracted towards the damaged broccoli roots and repelled by allyl isothiocyanate. However, for Bioassay 3, $\hat{\gamma}_{22}$ in Table 5.2 is greater than unity, implying repulsion from $\hat{\boldsymbol{\mu}}$. The results do not agree with the hypothesis of attraction towards $\mathbf{a}$, and

upon visual examination of Figure 5.3, the parameter estimates do not suitably reflect the nature of the observed track. For Bioassay 5, $\hat{\gamma}_{11}$ is greater than unity and $\hat{\gamma}_{22}$ is smaller, indicating repulsion along the x-axis and attraction along the y-axis. As the x-coordinate of $\hat{\boldsymbol{\mu}}$ is close to 0 for Bioassay 5, the fact that the larva moves steadily away from the origin, as shown in Figure 5.3, can be interpreted as repulsion from $\hat{\boldsymbol{\mu}}$ along the x-axis, regardless of the direction of movement. It is obvious that the results obtained for Bioassay 5 are not necessarily truly indicative of repulsion from the chemical compound.

For Bioassays 1–3, both diagonal elements of $\hat{\boldsymbol{\Gamma}}$ in Table 5.3 are less than unity, indicating attraction towards $\mathbf{a}$ for the larvae used in the bioassays. The results shown in Table 5.3 differ from those in Table 5.2, in which only Bioassays 1 and 2 exhibited attraction. Visual examination of Figure 5.3 shows that the larva in Bioassay 3 does move towards the zone of damaged broccoli roots, and therefore that the results obtained in this case are to be preferred. For Bioassay 4, $\hat{\gamma}_{22}$ is greater than unity, suggesting repulsion from $\mathbf{a}$, and this matches the track seen in Figure 5.3. For Bioassay 5, $\hat{\gamma}_{11}$ is greater than unity, implying repulsion from the x-coordinate of $\mathbf{a}$. This result is to be expected for the reason given when discussing the values shown in Table 5.2.

Table 5.1: Positions of the centres of the zones of chemical compound, $\mathbf{a}$, for Bioassays 1–5.

Bioassay	**Position of a**
1	$(-0.084, -1.917)$
2	$(0.000, -1.805)$
3	$(-0.021, -1.830)$
4	$(-0.232, -1.900)$
5	$(-0.214, -1.979)$

Table 5.2: Maximum likelihood estimates of the parameters of diffusion process (5.2) for Bioassays 1–5.

	Bioassays				
Parameter	**1**	**2**	**3**	**4**	**5**
μ_1	-0.4790	-2.2798	4.3373	-0.3471	-0.0186
μ_2	-1.2882	-0.8360	-5.2221	2.1258	-0.0232
γ_{11}	0.9999	0.9998	0.9974	0.9985	1.0006
γ_{12}	0.0004	0.0004	-0.0026	-0.0010	-0.0015
γ_{21}	-0.0001	0.0003	0.0007	0.0008	0.0007
γ_{22}	0.9996	0.9990	1.0005	0.9993	0.9998
$\phi_{11}^{\dagger}$	4.2530	4.1334	7.6062	3.4474	5.5888
$\phi_{12}^{\dagger}$	0.0375	0.2367	0.2979	0.4632	0.3289
$\phi_{22}^{\dagger}$	4.5123	3.9575	4.7003	5.1895	5.0342

† Values multiplied by 10^5

Table 5.3: Maximum likelihood estimates of the parameters of diffusion process (5.3) for Bioassays 1–5.

	Bioassays				
Parameter	**1**	**2**	**3**	**4**	**5**
γ_{11}	0.9996	0.9999	0.9989	0.9994	1.0003
γ_{12}	-0.0001	-0.0001	-0.0003	0.0020	-0.0002
γ_{21}	-0.0004	0.0000	-0.0001	0.0015	0.0007
γ_{22}	0.9997	0.9999	0.9994	1.0002	0.9999
$\phi_{11}^{\dagger}$	4.2554	4.1339	7.6834	3.4868	5.5888
$\phi_{12}^{\dagger}$	0.0369	0.2328	0.2602	0.4936	0.3289
$\phi_{22}^{\dagger}$	4.5125	3.9591	4.7178	5.2129	5.0342

† Values multiplied by 10^5

5.2.3 Model comparison using BIC

The BIC was used to compare the competing diffusion processes. BIC values for the models fitted to the bioassays are presented in Table 5.4. Diffusion process (5.3) has lower BIC values for all 5 bioassays, indicating that it is preferable for all of them. It is therefore considered to be an improvement over process (5.2). However, for Bioassays 3 and 5, the BIC values for diffusion process (5.3) are only very slightly lower than those for (5.2).

Table 5.4: BIC values of diffusion processes (5.2) and (5.3) fitted to Bioassays 1–5, and the differences Δ BIC between the two.

Bioassay	**Process (5.2)**	**Process (5.3)**	**Δ BIC**
1	-129381	-129412	-31
2	-130848	-130880	-32
3	-25376	-25379	-3
4	-16665	-16677	-12
5	-126008	-126009	-1

5.2.4 Bayesian approach

While maximum likelihood estimates can be obtained easily for diffusion process (5.2) and (5.3), this is less straightforward for more complex models and we consider a Bayesian approach. Within a Bayesian framework, it is possible to obtain posterior estimates for the parameters of diffusion process (5.3) analytically. Given the diffusion process

$$\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\left\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_i - \mathbf{a}), \boldsymbol{\Phi}\right\},$$

we can reparametrize by defining $\mathbf{y}_i = \mathbf{x}_i - \mathbf{a}$. It follows that

$$\mathbf{y}_{i+1}|\mathbf{y}_i \sim \mathrm{N}\left(\boldsymbol{\Gamma}\mathbf{y}_i, \boldsymbol{\Phi}\right).$$

Given an observed data set $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\} = \{\mathbf{x}_1 - \mathbf{a}, \ldots, \mathbf{x}_n - \mathbf{a}\}$, where $\mathbf{y}_i = (y_{i1}, y_{i2})^T$ for $i = 0, \ldots, n$, the above can be rewritten in the form of linear regression as follows,

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{n2} \end{bmatrix} = \begin{bmatrix} y_{01} & y_{02} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ y_{(n-1)1} & y_{(n-1)2} & 0 & 0 \\ 0 & 0 & y_{01} & y_{02} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & y_{(n-1)1} & y_{(n-1)2} \end{bmatrix} \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{2n} \end{bmatrix},$$

where

$$(\epsilon_1, \ldots, \epsilon_{2n})^T \sim \mathrm{N}(0, \boldsymbol{\Phi} \otimes \mathbf{I}_n).$$

and $\otimes$ is the Kronecker product (Neudecker, 1969). This linear regression can be expressed as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

$$\mathbf{Z} = \begin{bmatrix} y_{01} & y_{02} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ y_{(n-1)1} & y_{(n-1)2} & 0 & 0 \\ 0 & 0 & y_{01} & y_{02} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & y_{(n-1)1} & y_{(n-1)2} \end{bmatrix}, \quad \boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{\Gamma}^T), \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{2n} \end{bmatrix}.$$

The likelihood function for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ takes the form

$$L(\boldsymbol{\Gamma}, \boldsymbol{\Phi} | \mathbf{Y}, \mathbf{Z}) \propto |\boldsymbol{\Phi}|^{-n} \exp\left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})^T \boldsymbol{\Phi}^{-1} \otimes \mathbf{I}_n (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}) \right\}.$$

The maximum likelihood estimates for the likelihood function can be obtained easily and are as follows,

$$\hat{\boldsymbol{\Gamma}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1} \mathbf{Z}^T\mathbf{Y},$$
$$\hat{\boldsymbol{\Phi}} = \left(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}\right)^T \left(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}\right),$$

where

$$\hat{\boldsymbol{\beta}} = \mathrm{vec}(\hat{\boldsymbol{\Gamma}}^T).$$

The likelihood function can be rewritten as a function of the maximum likelihood estimators of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ as follows,

$$L(\mathbf{\Gamma}, \mathbf{\Phi}|\mathbf{Y}, \mathbf{Z}) \propto |\mathbf{\Phi}|^{-n} \exp\left\{ -\frac{1}{2}\text{tr}(\mathbf{\Phi}^{-1}\hat{\mathbf{\Phi}}) \, , \right.$$
$$\left. -\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{\Phi}^{-1} \otimes \mathbf{Z}^T\mathbf{Z}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}.$$

We assume the following invariant prior distribution (Tiao and Zellner, 1964; Jeffreys, 1961),

$$\begin{aligned} p(\mathbf{\Gamma}, \mathbf{\Phi}) &= p(\mathbf{\Gamma})p(\mathbf{\Phi}), \\ p(\mathbf{\Gamma}) &\propto \mathbf{I}, \\ p(\mathbf{\Phi}) &\propto |\mathbf{\Phi}|^{-\frac{3}{2}}. \end{aligned}$$

The posterior distribution of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ using this prior can be written as

$$\begin{aligned} &p(\mathbf{\Gamma}, \mathbf{\Phi}|\mathbf{Y}, \mathbf{Z}) \\ &\quad = L(\mathbf{\Gamma}, \mathbf{\Phi}|\mathbf{Y}, \mathbf{Z})p(\mathbf{\Gamma})p(\mathbf{\Phi}) \\ &\quad \propto |\mathbf{\Phi}|^{-n-\frac{3}{2}} \exp\left\{ -\frac{1}{2}\text{tr}(\mathbf{\Phi}^{-1}\hat{\mathbf{\Phi}}) \right. \\ &\qquad \left. -\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{\Phi}^{-1} \otimes \mathbf{Z}^T\mathbf{Z}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\} \\ &\quad = |\mathbf{\Phi}|^{-2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{\Phi}^{-1} \otimes \mathbf{Z}^T\mathbf{Z}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\} \\ &\qquad \times |\mathbf{\Phi}|^{-n+\frac{1}{2}} \exp\left\{ -\frac{1}{2}\text{tr}(\mathbf{\Phi}^{-1}\hat{\mathbf{\Phi}}) \right\}. \end{aligned}$$

This posterior distribution may be expressed as

$$p(\mathbf{\Gamma}, \mathbf{\Phi}|\mathbf{Y}, \mathbf{Z}) = p(\mathbf{\Gamma}|\mathbf{\Phi}, \mathbf{Y}, \mathbf{Z})p(\mathbf{\Phi}|\mathbf{Y}, \mathbf{Z}),$$

where

$$\begin{aligned} p(\mathbf{\Gamma}|\mathbf{\Phi}, \mathbf{Y}, \mathbf{Z}) &\propto |\mathbf{\Phi}|^{-2} \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{\Phi}^{-1} \otimes \mathbf{Z}^T\mathbf{Z}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}, \\ p(\mathbf{\Phi}|\mathbf{Y}, \mathbf{Z}) &\propto |\mathbf{\Phi}|^{-n+\frac{1}{2}} \exp\left\{ -\frac{1}{2}\text{tr}(\mathbf{\Phi}^{-1}\hat{\mathbf{\Phi}}) \right\}. \end{aligned}$$

Therefore, it follows that the posterior distribution of $\boldsymbol{\beta}$ and $\mathbf{\Phi}$ is given by

$$\boldsymbol{\beta}|\boldsymbol{\Phi}, \mathbf{Y}, \mathbf{Z} \sim \mathrm{N}\{\hat{\boldsymbol{\beta}}, \boldsymbol{\Phi} \otimes (\mathbf{Z}^T\mathbf{Z})^{-1}\},$$
$$\boldsymbol{\Phi}|\mathbf{Y}, \mathbf{Z} \sim \text{Inverse Wishart}(2n-4, \hat{\boldsymbol{\Phi}}).$$

The posterior distribution shown here requires the assumption of an improper prior for $\boldsymbol{\Gamma}$, and the calculations become extremely challenging when additional components are introduced. As discussed in Chapter 4, within a Bayesian framework, Markov Chain Monte Carlo (MCMC) methods can be used to obtain posterior distributions for the parameters in such cases. Posterior distributions for the parameters of diffusion processes (5.2) and (5.3) fitted to the bioassay data were estimated with a MCMC approach, using WinBUGS and OpenBUGS. The BUGS code used for diffusion process (5.3) is given in Appendix B. The following independent prior distributions were used for diffusion process (5.2),

$$\boldsymbol{\mu} \sim \mathrm{N}\left\{(0,0)^T, 10^2\mathbf{I}\right\}, \tag{5.4}$$
$$\gamma_{ij} \sim \mathrm{N}\left(1, 10^2\right), \quad i,j = 1,2, \tag{5.5}$$
$$\boldsymbol{\Phi} \sim \text{Inverse Wishart}\left(10^{-10}\mathbf{I}, 2\right), \tag{5.6}$$

and represent vague prior information. For diffusion process (5.3), the priors for γ_{ij} and $\boldsymbol{\Phi}$ shown in (5.5) and (5.6), were used.

Summary statistics for the posterior distributions of the model parameters for diffusion process (5.2) for Bioassays 1–5 are presented in Tables 5.5 to 5.9. For Bioassay 1, the posterior means of μ_1 and μ_2 shown in Table 5.5 are similar to the corresponding maximum likelihood estimates in Table 5.2. However, the posterior standard deviations are very high, and the intervals between the 2.5% and 97.5% percentiles are large enough to encompass movement in any direction from the origin. As such, these results provide little information about the location of $\boldsymbol{\mu}$. The posterior means of γ_{11} and γ_{22} are less than unity, indicating attraction towards $\boldsymbol{\mu}$. The 97.5% percentiles for γ_{11} and γ_{22} are not greater than unity, corroborating the hypothesis of attraction towards $\boldsymbol{\mu}$. Posterior means of the elements of $\boldsymbol{\Phi}$ are close to the maximum likelihood estimates. Note that the the intervals between the 2.5% and 97.5% percentiles for γ_{12}, γ_{21}, and ϕ_{12} include 0, suggesting that the off-diagonal elements of $\boldsymbol{\Gamma}$ or $\boldsymbol{\Phi}$ are zero.

For Bioassay 2, as seen in Table 5.6, the posterior means of μ_1 and μ_2 differ more from the maximum likelihood estimates than in the case of Bioassay 1. The posterior standard deviations are very large, indicating that the results provide extremely limited information about $\boldsymbol{\mu}$. The summary statistics for the elements

of $\boldsymbol{\Gamma}$ suggest attraction towards $\boldsymbol{\mu}$, and the posterior means of the elements of $\boldsymbol{\Phi}$ are similar to the maximum likelihood estimates, as was found for Bioassay 1. The 2.5% and 97.5% percentiles support the hypothesis that $\boldsymbol{\Gamma}$ is diagonal. However, the hypothesis that $\boldsymbol{\Phi}$ is diagonal is not supported.

For Bioassay 3 (Table 5.7), the posterior means of μ_1 and μ_2 differ considerably from the maximum likelihood estimates, but both location estimates for $\boldsymbol{\mu}$ are in approximately the same direction from the origin. As the larva does not reach either set of coordinates during the bioassay, the effect of each point of attraction on the larva's movements is similar. The posterior standard deviations for the elements of $\boldsymbol{\mu}$ are again very high. The posterior means for the elements of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are all close to the corresponding maximum likelihood estimates. While the summary statistics for the elements of $\boldsymbol{\Gamma}$ indicate attraction towards $\boldsymbol{\mu}$ along the x-axis and repulsion along the y-axis, the 2.5% percentile for γ_{22} is less than unity, suggesting that there is not strong evidence for the repulsion. The parameter estimates in Table 5.7 have replicated the repulsion from $\boldsymbol{\mu}$ implied by the maximum likelihood estimates in Table 5.2, which does not appear to represent the data well. The results indicate that $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are not diagonal.

For Bioassay 4 (Table 5.8), the posterior means of μ_1 and μ_2 are different from the maximum likelihood estimates, though the posterior standard deviations are very large, and $\hat{\boldsymbol{\mu}}$ lies between the 2.5% and 97.5% percentiles for $\boldsymbol{\mu}$. In contrast with Bioassay 3, the estimated location of $\boldsymbol{\mu}$ is not in the same direction from the origin as $\hat{\boldsymbol{\mu}}$. The posterior means for the elements of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are fairly close to the maximum likelihood estimates, and the 2.5% and 97.5% percentiles for ϕ_{12} suggest that $\boldsymbol{\Phi}$ is not diagonal.

For Bioassay 5 (Table 5.9), the posterior means of the elements of $\boldsymbol{\mu}$ are fairly close to the maximum likelihood estimates, and their posterior standard deviations are much smaller than for Bioassays 1–4, implying that much stronger information about the position of $\boldsymbol{\mu}$ has been obtained for this bioassay. However, the estimated location of $\boldsymbol{\mu}$ is close to the origin, and does not correspond even roughly to the direction of the centre of the zone of damaged broccoli roots. The posterior means of the elements of $\boldsymbol{\Gamma}$ indicate repulsion from $\boldsymbol{\mu}$ along the x-axis and attraction along the y-axis. The posterior means for the elements of $\boldsymbol{\Phi}$ are quite close to the maximum likelihood estimates, as is the case for all of the bioassays. The 2.5% and 97.5% percentiles for γ_{21} suggest that $\boldsymbol{\Gamma}$ is not diagonal.

The corresponding summary statistics for diffusion process (5.3) are presented in Tables 5.10 to 5.14. For Bioassay 1, as shown in Table 5.10, the posterior means of the elements of $\boldsymbol{\Gamma}$ are identical to the maximum likelihood estimates in Table 5.3, whereas for diffusion process (5.2) the posterior means shown in

Table 5.5 are close to but not the same as the maximum likelihood estimates. The 97.5% percentiles for γ_{11} and γ_{22} are not greater than unity, supporting the hypothesis of attraction towards $\mathbf{a}$. The posterior means of the elements of $\mathbf{\Phi}$ are close to the maximum likelihood estimates. The 2.5% and 97.5% percentiles of the elements of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ provide support for the hypotheses that these matrices are diagonal.

The results for Bioassay 2 (Table 5.11) resemble those for Bioassay 1 in that the posterior means of the elements of $\mathbf{\Gamma}$ are identical to the maximum likelihood estimates, and the corresponding values for the elements of $\mathbf{\Phi}$ are close but not identical. The 97.5% percentiles for γ_{11} and γ_{22} are not greater than unity, implying attraction towards $\mathbf{a}$. The 2.5% and 97.5% percentiles for ϕ_{12} suggest that $\mathbf{\Phi}$ is not diagonal.

The posterior means of the parameters for Bioassay 3 (Table 5.12) resemble those for Bioassays 1 and 2 in terms of their similarity to the maximum likelihood estimates, and suggest attraction towards $\mathbf{a}$. The 2.5% and 97.5% percentiles for the elements of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ support the hypotheses that both of these matrices are diagonal.

For Bioassay 4 (Table 5.13), the posterior means of the parameters are similar to the maximum likelihood estimates, and indicate attraction towards $\mathbf{a}$ along the x-axis and repulsion along the y-axis. This corresponds to movement away from $\mathbf{a}$, as is evident in Figure 5.3. The 2.5% and 97.5% percentiles for ϕ_{12} indicates that $\mathbf{\Phi}$ is not diagonal.

For Bioassay 5 (Table 5.14), the posterior means are again similar to the maximum likelihood estimates. The posterior mean of γ_{11} is slightly higher than unity, and that of γ_{22} is lower. These values indicate movement towards $\mathbf{a}$ along the y-axis and repulsion along the x-axis, which results in weak attraction. This bioassay differs from the others in that the 2.5% and 97.5% percentiles for γ_{21} suggests that $\mathbf{\Gamma}$ is not diagonal.

Table 5.5: Summary statistics of posterior distributions of the parameters of diffusion process (5.2) for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (5.4), (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
μ_1	-0.4850	-0.4998	1.9193	-4.0062	3.6172
μ_2	-1.3238	-1.3520	3.3342	-8.7322	7.3021
γ_{11}	0.9990	0.9990	0.0006	0.9978	1.0000
γ_{12}	0.0003	0.0003	0.0003	-0.0002	0.0009
γ_{21}	-0.0002	-0.0002	0.0005	-0.0012	0.0008
γ_{22}	0.9997	0.9997	0.0003	0.9992	1.0000
$\phi_{11}^{\dagger}$	4.2568	4.2570	0.0641	4.1330	4.3830
$\phi_{12}^{\dagger}$	0.0367	0.0379	0.0457	-0.0532	0.1258
$\phi_{22}^{\dagger}$	4.5140	4.5150	0.0674	4.3860	4.6480

† Values multiplied by 10^5.

Table 5.6: Summary statistics of posterior distributions of the parameters of diffusion process (5.2) for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (5.4), (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
μ_1	-1.2277	-1.5640	5.3022	-11.8503	11.4200
μ_2	-0.5388	-0.5933	2.2252	-4.8780	4.5210
γ_{11}	0.9998	0.9999	0.0002	0.9950	1.0000
γ_{12}	0.0003	0.0002	0.0005	-0.0007	0.0011
γ_{21}	0.0002	0.0002	0.0002	-0.0001	0.0005
γ_{22}	0.9993	0.9993	0.0005	0.9982	1.0000
$\phi_{11}^{\dagger}$	4.1366	4.1350	0.0610	4.0170	4.2560
$\phi_{12}^{\dagger}$	0.2342	0.2339	0.0428	0.1580	0.3180
$\phi_{22}^{\dagger}$	3.9607	3.9600	0.0587	3.8460	4.0760

† Values multiplied by 10^5.

Table 5.7: Summary statistics of posterior distributions of the parameters of diffusion process (5.2) for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (5.4), (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
μ_1	2.9295	3.0544	3.4032	−2.8421	11.4800
μ_2	−3.5712	−3.5934	3.4593	−11.9700	2.3501
γ_{11}	0.9975	0.9975	0.0011	0.9952	0.9994
γ_{12}	−0.0003	−0.0003	0.0006	−0.0037	−0.0014
γ_{21}	0.0009	0.0008	0.0006	−0.0004	0.0022
γ_{22}	1.0051	1.0050	0.0006	0.9995	1.0100
$\phi_{11}^{\dagger}$	7.6263	7.6230	0.2519	7.1450	8.1350
$\phi_{12}^{\dagger}$	0.2954	0.2990	0.1403	0.0233	0.5710
$\phi_{22}^{\dagger}$	4.7109	4.7060	0.1540	4.4200	5.0230

† Values multiplied by 10^5.

Table 5.8: Summary statistics of posterior distributions of the parameters of diffusion process (5.2) for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (5.4), (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
μ_1	1.1354	1.2764	3.8411	−6.3756	10.6100
μ_2	0.5205	0.5323	5.3796	−12.6400	11.4900
γ_{11}	0.9991	0.9990	0.0019	0.9954	1.0030
γ_{12}	−0.0010	−0.0011	0.0005	−0.0019	0.0000
γ_{21}	−0.0006	−0.0006	0.0018	−0.0035	0.0040
γ_{22}	0.9995	0.9996	0.0005	0.9984	1.0000
$\phi_{11}^{\dagger}$	3.4634	3.4631	0.1457	3.1930	3.7600
$\phi_{12}^{\dagger}$	0.3650	0.3644	0.1248	0.2029	7.0971
$\phi_{22}^{\dagger}$	5.2125	5.1985	0.2159	4.8020	5.6491

† Values multiplied by 10^5.

Table 5.9: Summary statistics of posterior distributions of the parameters of diffusion process (5.2) for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (5.4), (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
μ_1	-0.0189	-0.0193	0.1491	-0.3144	0.2779
μ_2	-0.2395	-0.2327	0.0991	-0.4442	-0.0684
γ_{11}	1.0008	1.0010	0.0004	1.0000	1.0010
γ_{12}	-0.0014	-0.0014	0.0002	-0.0018	-0.0010
γ_{21}	-0.0007	-0.0007	0.0001	-0.0005	0.0009
γ_{22}	0.9998	0.9998	0.0002	0.9994	1.0000
$\phi_{11}^{\dagger}$	5.5718	5.5690	0.0834	5.4120	5.7400
$\phi_{12}^{\dagger}$	0.3280	0.3278	0.0567	0.2176	0.4367
$\phi_{22}^{\dagger}$	5.0366	5.0370	0.0742	4.8890	5.1850

† Values multiplied by 10^5.

Table 5.10: Summary statistics of posterior distributions of the parameters of diffusion process (5.3) for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9996	0.9996	0.0003	0.9991	1.0000
γ_{12}	-0.0001	-0.0001	0.0001	-0.0003	0.0000
γ_{21}	-0.0004	-0.0004	0.0003	-0.0010	0.0002
γ_{22}	0.9997	0.9997	0.0001	0.9996	0.9999
$\phi_{11}^{\dagger}$	4.2574	4.2570	0.0638	4.1330	4.3840
$\phi_{12}^{\dagger}$	0.0369	0.0368	0.0458	-0.0507	0.1259
$\phi_{22}^{\dagger}$	4.5142	4.5140	0.0675	4.3840	4.6480

† Values multiplied by 10^5.

Table 5.11: Summary statistics of posterior distributions of the parameters of diffusion process (5.3) for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9999	0.9999	0.0001	0.9997	1.0000
γ_{12}	−0.0001	−0.0001	0.0001	−0.0003	0.0000
γ_{21}	0.0000	0.0000	0.0001	−0.0003	0.0002
γ_{22}	0.9999	0.9999	0.0001	0.9998	1.0000
$\phi_{11}^{\dagger}$	4.1356	4.1350	0.0616	4.0170	4.2580
$\phi_{12}^{\dagger}$	0.2333	0.2333	0.0427	0.1498	0.3163
$\phi_{22}^{\dagger}$	3.9611	3.9610	0.0590	3.8480	4.0780

† Values multiplied by 10^5.

Table 5.12: Summary statistics of posterior distributions of the parameters of diffusion process (5.3) for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9989	0.9989	0.0010	0.9970	1.0010
γ_{12}	−0.0003	−0.0003	0.0003	−0.0008	0.0002
γ_{21}	−0.0001	−0.0001	0.0008	−0.0016	0.0014
γ_{22}	0.9994	0.9994	0.0002	0.9990	0.9998
$\phi_{11}^{\dagger}$	7.6981	7.6940	0.2525	7.2180	8.2060
$\phi_{12}^{\dagger}$	0.2613	0.2616	0.1401	−0.0135	0.5351
$\phi_{22}^{\dagger}$	4.7289	4.7260	0.1557	4.4340	5.0410

† Values multiplied by 10^5.

Table 5.13: Summary statistics of posterior distributions of the parameters of diffusion process (5.3) for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9995	0.9995	0.0020	0.9956	1.0030
γ_{12}	0.0002	0.0002	0.0004	−0.0001	0.0010
γ_{21}	0.0016	0.0016	0.0024	−0.0031	0.0061
γ_{22}	1.0001	1.0000	0.0005	0.9992	1.0010
$\phi_{11}^{\dagger}$	3.4981	3.4940	0.1458	3.2240	3.7940
$\phi_{12}^{\dagger}$	0.4955	0.4943	0.1260	0.2486	0.7432
$\phi_{22}^{\dagger}$	5.2317	5.2280	0.2171	4.8240	5.6720

† Values multiplied by 10^5.

Table 5.14: Summary statistics of posterior distributions of the parameters of diffusion process (5.3) for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (5.5) and (5.6).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0001	1.0000	0.0002	1.0000	1.0010
γ_{12}	−0.0002	−0.0002	0.0000	−0.0003	−0.0001
γ_{21}	0.0007	0.0007	0.0001	0.0005	0.0009
γ_{22}	0.9999	0.9999	0.0000	0.9998	1.0000
$\phi_{11}^{\dagger}$	5.5908	5.5900	0.0833	5.4300	5.7550
$\phi_{12}^{\dagger}$	0.3291	0.3299	0.0557	0.2186	0.4371
$\phi_{22}^{\dagger}$	5.0376	5.0370	0.0754	4.8910	5.1870

† Values multiplied by 10^5.

5.2.5 Model comparison and discussion

BIC values for diffusion processes (5.2) and (5.3) for the 5 bioassays are presented in Table 5.15, using the posterior means for the parameter values. The conclusions are the same as those obtained when using the maximum likelihood estimates (Table 5.4). Diffusion process (5.3) has lower BIC values for all 5 bioassays and is therefore assumed to be superior to process (5.2). In contrast to Table 5.4, the BIC values of diffusion process (5.3) for Bioassays 3 and 5 are much lower than those of process (5.2). The results in Table 5.15 appear reasonable, as diffusion process (5.3) is a more physically realistic model than (5.2). As the arenas used in each bioassay are homogeneous other than the zones of chemical compound, attraction to or repulsion from a point $\boldsymbol{\mu}$ outside the zone is not considered plausible. Diffusion process (5.3) correctly suggests attraction towards $\mathbf{a}$ for Bioassays 1–3 and 5, and repulsion for Bioassay 4.

Table 5.15: BIC values of diffusion processes (5.2) and (5.3) fitted to Bioassays 1–5, and the differences Δ BIC between the two.

Bioassay	**Process (5.2)**	**Process (5.3)**	**Δ BIC**
1	-129380	-129412	-32
2	-130844	-130879	-35
3	-17891	-25378	-7487
4	-16636	-16675	-39
5	-125756	-125996	-240

We observed that the posterior standard deviations for the elements of $\boldsymbol{\mu}$ for diffusion process (5.2) are consistently large other than for Bioassay 5. Furthermore, the posterior means for $\boldsymbol{\mu}$ differ considerably from the vector of maximum likelihood estimates $\hat{\boldsymbol{\mu}}$ for Bioassays 3 and 4. This suggests that process (5.2) is not very successful at identifying the location of a particular point of attraction or repulsion. Logically, diffusion process (5.3), which does not use the parameter $\boldsymbol{\mu}$ and instead uses the known position of the zone $\mathbf{a}$, appears to be a more suitable alternative. The posterior means of the elements of $\boldsymbol{\Gamma}$ for diffusion process (5.3) are also closer to the maximum likelihood estimates than when diffusion process (5.2) is used.

The posterior means of γ_{11} and γ_{22} for diffusion process (5.3) applied to Bioassays 1–3 are both less than unity, indicating attraction towards $\mathbf{a}$. For Bioassay 4, the mean of γ_{11} is less than unity while the mean of γ_{22} is greater, resulting in repulsion from $\mathbf{a}$. For Bioassay 5, the mean of γ_{11} is greater than

unity and the mean of γ_{22} is less, signifying lateral movement away from $\mathbf{a}$ and vertical movement towards it. On visual examination of the bioassay data plots in Figure 5.3, all of these conclusions appear correct.

5.3 Simulation and diagnostics

5.3.1 Simulation study

The properties of diffusion process (5.3) were explored using a simulation study. We simulated 200 data sets from the diffusion process, with parameters equal to the posterior means for Bioassay 1 as given in Table 5.10. Each data set contained 9000 observations, the same number as in the bioassay data. Without loss of generality, the value of the initial observation $\mathbf{x}_0$ was set equal to the origin, and each subsequent observation $\mathbf{x}_i$ was simulated from the density

$$\mathrm{N}\left\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \mathbf{a}), \boldsymbol{\Phi}\right\}.$$

Diffusion process (5.3) was fitted to each of the simulated data sets, with prior distributions for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ as shown in (5.3). Posterior distributions for the parameters were estimated using WinBUGS. Summary statistics for the posterior means of the parameters are shown in Table 5.16. The values displayed include the mean, standard deviation (SD), bias and root mean squared error (RMSE).

Table 5.16: Summary statistics of estimated posterior means obtained when fitting diffusion process (5.3) to 200 data sets of 9000 observations each simulated from diffusion process (5.3) with parameter values equal to the posterior means shown in Table 5.10. Presented summary statistics include the mean, standard deviation (SD), bias and root mean squared error (RMSE).

Parameter	**Mean**	**SD**	**Bias**	**RMSE**
γ_{11}	0.9995	0.0002	−0.0001	0.0002
γ_{12}	−0.0004	0.0001	−0.0003	0.0003
γ_{21}	−0.0001	0.0002	0.0003	0.0004
γ_{22}	0.9997	0.0001	0.0000	0.0001
$\phi^{\dagger}_{11}$	4.2611	0.0634	0.0037	0.0635
$\phi^{\dagger}_{12}$	0.0787	0.0462	0.0418	0.0623
$\phi^{\dagger}_{22}$	4.5237	0.0675	0.0095	0.0682

† Values multiplied by 10^5.

The mean values of the posterior means of the elements of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are close to

the true values in the simulation density. The standard deviations of the posterior means are comparatively small, indicating that the values obtained do not vary greatly between simulated data sets. These results imply that the fitted diffusion process produces very similar parameter estimates for the simulated data sets to those produced for the bioassay data, indicating that the simulations resemble the original data. This provides support for the diffusion process as a suitable model.

5.3.2 Diagnostic plots using Mahalanobis distance

The conditional distributions of $\mathbf{x}_i$ given $\mathbf{x}_{i-1}$ for a data set $\{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$ can be used to further assess how realistic a model the diffusion process provides for the data. Under diffusion process (5.3), the conditional distribution of $\mathbf{x}_i$ given $\mathbf{x}_{i-1}$ is

$$\mathbf{x}_i|\mathbf{x}_{i-1} \sim \mathrm{N}\left\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \mathbf{a}), \boldsymbol{\Phi}\right\}.$$

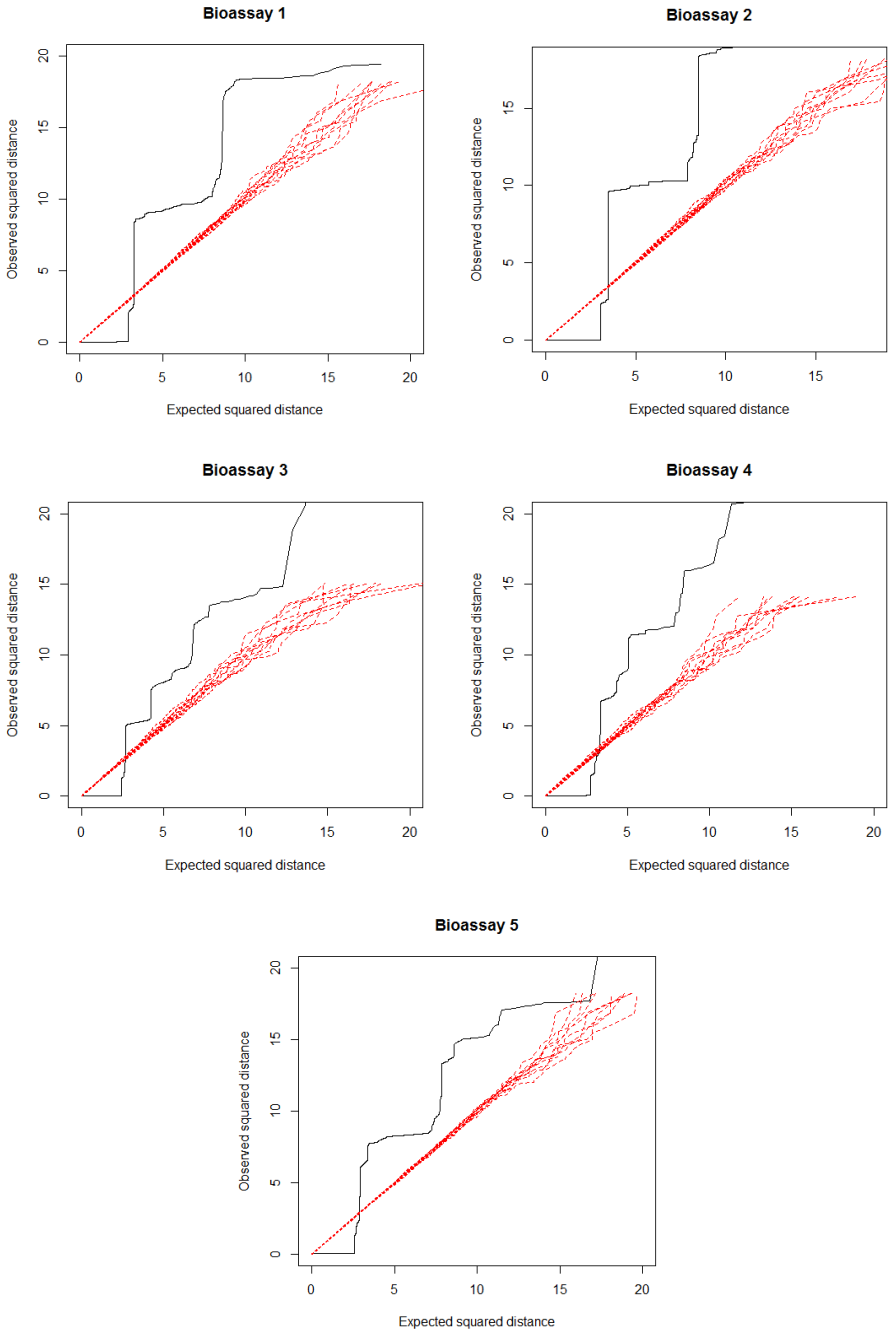
If the bioassay data are generated by a bivariate normal diffusion process, the squared Mahalanobis distances

$$d_i^2 = \left[\mathbf{x}_i - \left\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \mathbf{a})\right\}\right]^T \boldsymbol{\Phi}^{-1} \left[\mathbf{x}_i - \left\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \mathbf{a})\right\}\right], \quad i = 0, \ldots, n,$$

for the data should follow a χ_2^2 distribution. The true values of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$, and by extension the true d_i^2 values, are unknown, but can be estimated using the posterior means for the relevant bioassay in Tables 5.10 to 5.14.

Plots of the ordered estimated d_i^2 values against the expected values for Bioassays 1–5 are shown in Figure 5.4. Also included for comparison are the corresponding values for 10 data sets simulated from diffusion process (5.3) fitted to each bioassay. The estimated d_i^2 values for the simulated data sets are similar to the expected values, which is to be expected as the data sets are known to be generated by a bivariate normal diffusion process. However, the values for the bioassay data are very different from the expected values in all cases, suggesting that the data are not generated by bivariate normal diffusion. Furthermore, most of the observed estimated d_i^2 values belong to one of several subsets of similar values, suggesting that the larva alternates between movements of several different approximate lengths. This behaviour could be caused by discretization error, as the observations in the bioassays are rounded to 2 decimal places.

Figure 5.4: Ordered observed estimated d_i^2 values against expected values for Bioassays 1–5, displayed as black lines. In each plot, the 10 red lines correspond to ordered observed estimated d_i^2 values from 10 data sets simulated from diffusion process (5.3), with parameters equal to the posterior means obtained by fitting the process to the relevant bioassay.



5.3.3 Simulation plots

Plots of the bioassay data are shown in Figures 5.5 to 5.9. Each plot also includes 5 data sets simulated from diffusion process (5.3) with parameter values equal to the posterior means obtained when fitting the diffusion process to the appropriate bioassay. In Figure 5.5, the simulated data sets consist of movement in the same direction as the bioassay data for Bioassay 1. However, the simulations appear to contain larger random movements than the real data set, which consists of movement along a more consistent path with smaller localized random movements. The same is true of Figure 5.6. It should be noted that the path of the larva in Bioassay 2 includes lateral movement with respect to $\mathbf{a}$. In the absence of any attraction to another point in the arena, it is possible that this feature of the data is due to weaker attraction to $\mathbf{a}$, and that the direction of movement along the x-axis is not determined by the position of $\mathbf{a}$. If this is the case, the fact that the simulated data sets consist of movement in the same direction as the bioassay data indicates that the fitted model has replicated a feature of the original data set which is not in fact determined by the true underlying nature of the larva's movements.

For Bioassay 3 (Figure 5.7), the simulated data sets do not resemble the bioassay data as strongly as those for Bioassays 1 and 2. The simulated data sets have roughly the same overall direction of movement as the bioassay data. However, the larva in Bioassay 3 changes direction noticeably less often than the larvae in Bioassays 1 and 2. The simulated data sets do not reflect this fact and include many more changes of direction than the real data. The direction of the larva's movement appears to be heavily dependent on the previous direction of movement, a feature currently not accounted for by the diffusion process.

The simulated data sets for Bioassay 4 (Figure 5.8) have similar directions of movement to the real data, but appear to be more localized and generally remain close to the origin, whereas the bioassay data set consists of strongly directed movement and reaches a greater distance from the origin. The opposite is true of Bioassay 5 (Figure 5.9), for which the simulated data sets move much further from the origin than the real data set. The larva in Bioassay 5 spends much of the bioassay moving around the origin before eventually moving away, whereas it appears that the simulated data sets have reproduced only the latter feature of the data. Furthermore, the simulated tracks move away from the origin more quickly than the larva does, and in a different direction. The diffusion process fitted to Bioassay 5 is clearly not a realistic model of the larva's movements.

Figure 5.5: Plot of Bioassay 1 and 5 data sets of 9000 observations simulated from diffusion process (5.3), with parameter values equal to the posterior means shown in Table 5.10.

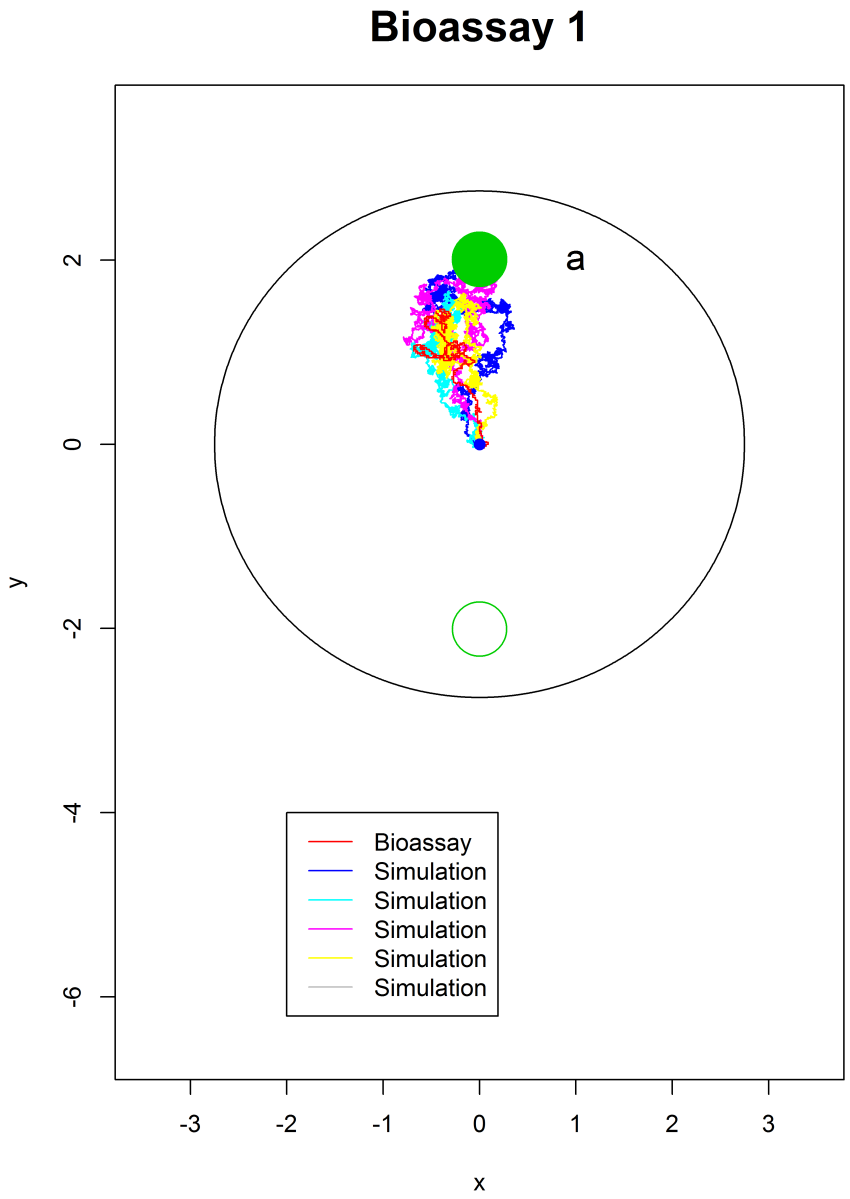


Figure 5.6: Plot of Bioassay 2 and 5 data sets of 9000 observations simulated from diffusion process (5.3), with parameter values equal to the posterior means shown in Table 5.11.

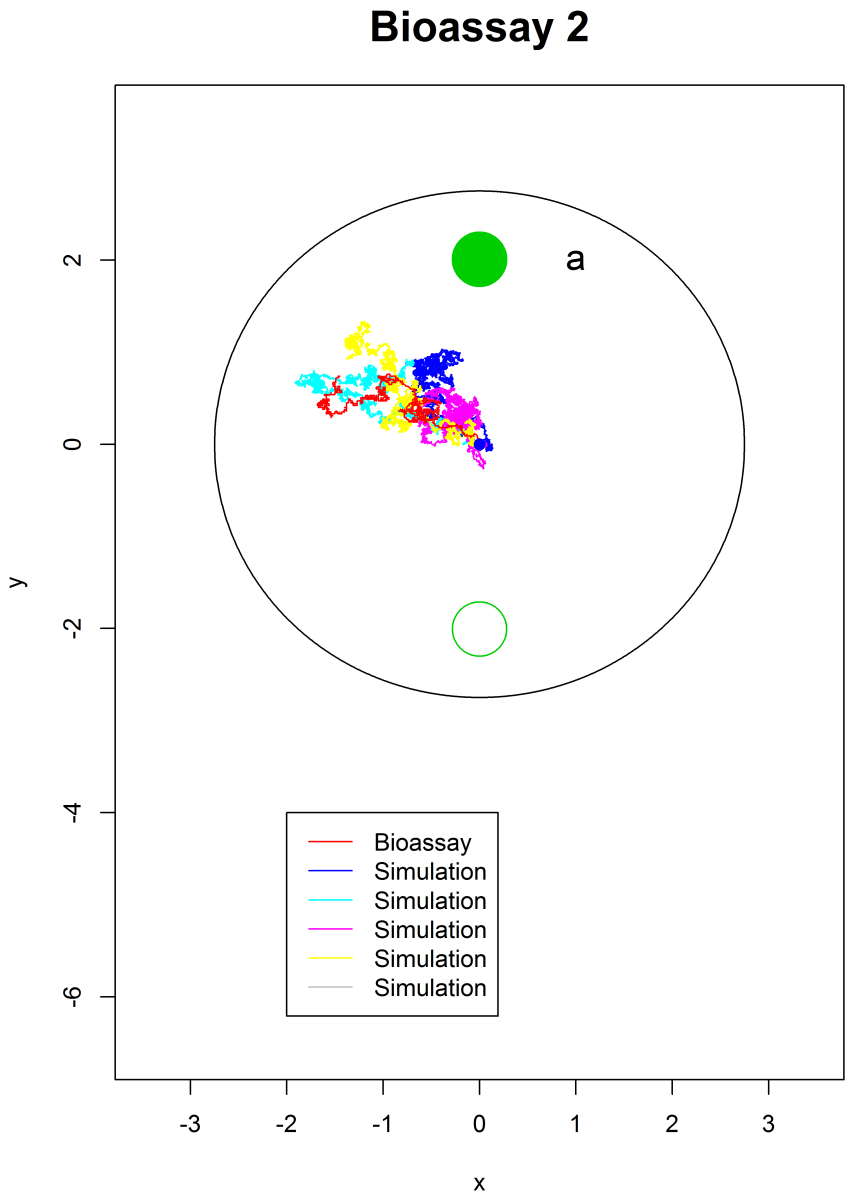


Figure 5.7: Plot of Bioassay 3 and 5 data sets of 9000 observations simulated from diffusion process (5.3), with parameter values equal to the posterior means shown in Table 5.12.

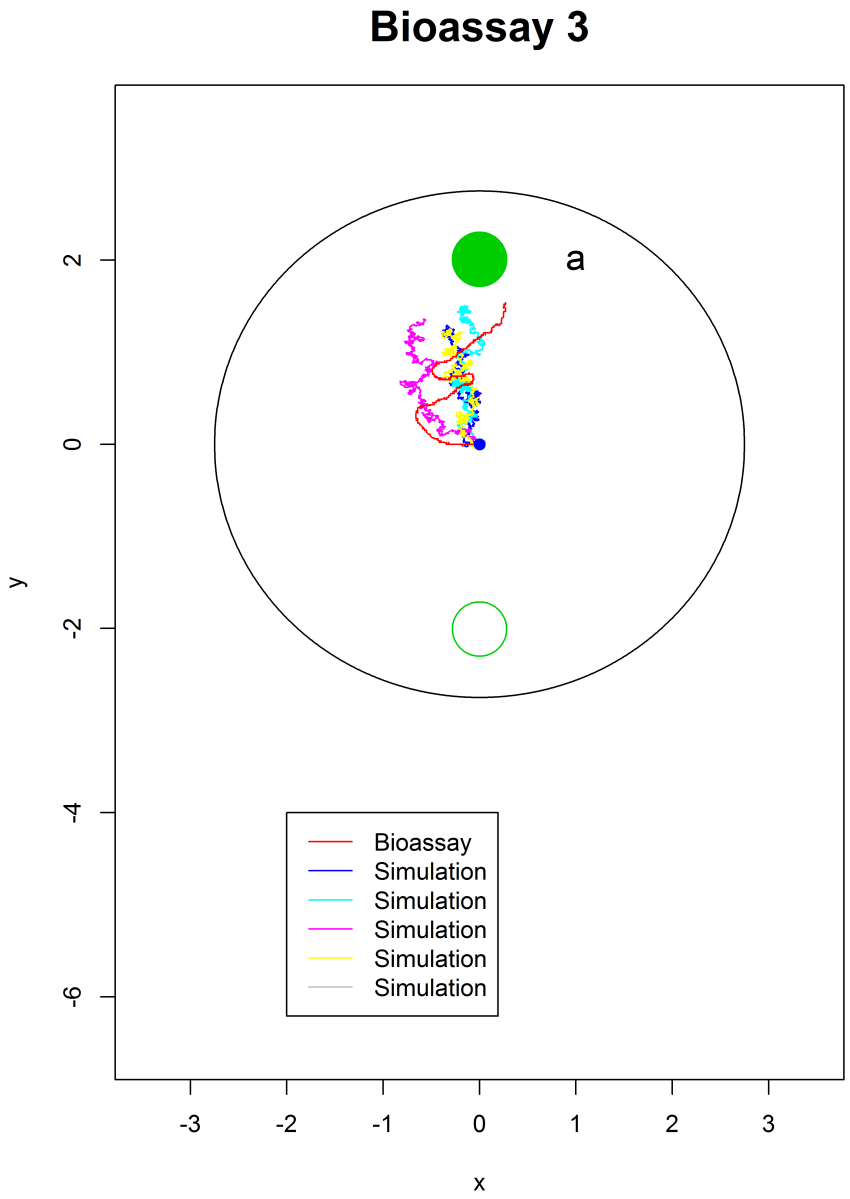


Figure 5.8: Plot of Bioassay 4 and 5 data sets of 9000 observations simulated from diffusion process (5.3), with parameter values equal to the posterior means shown in Table 5.13.

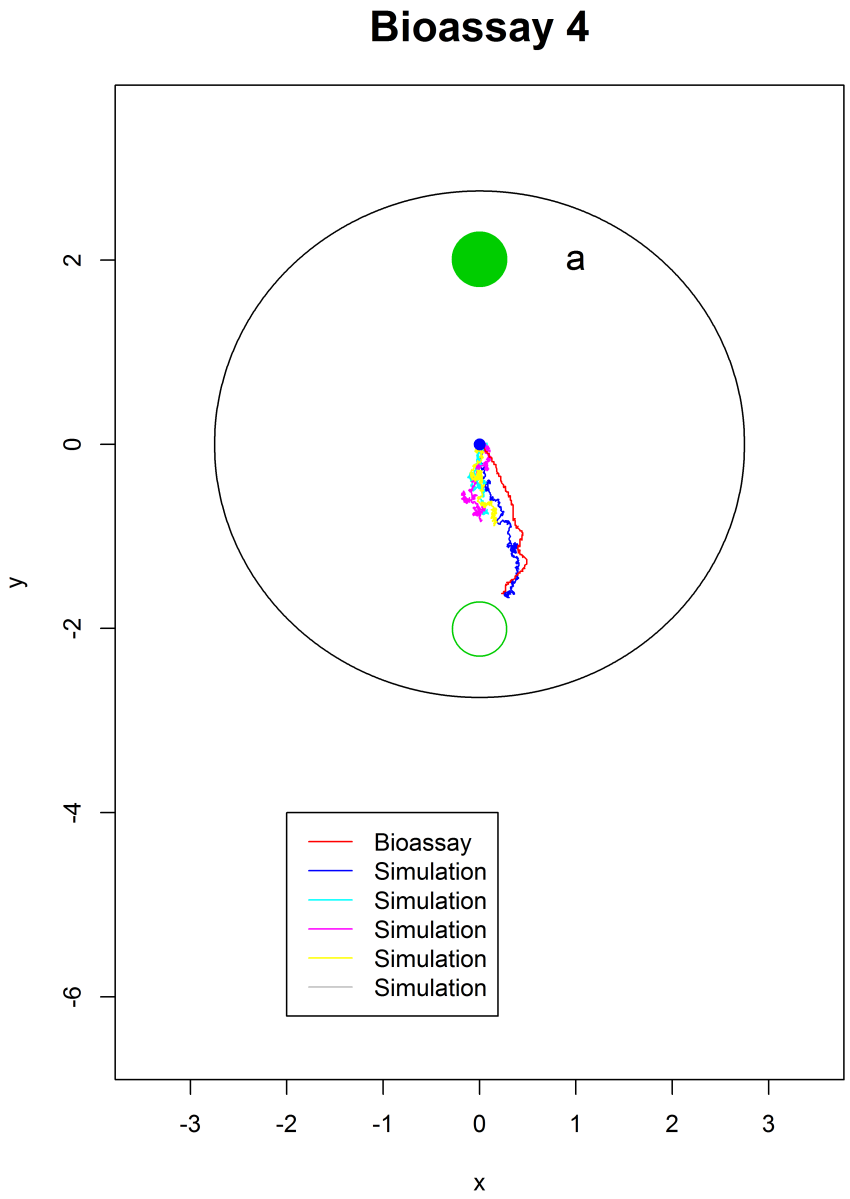
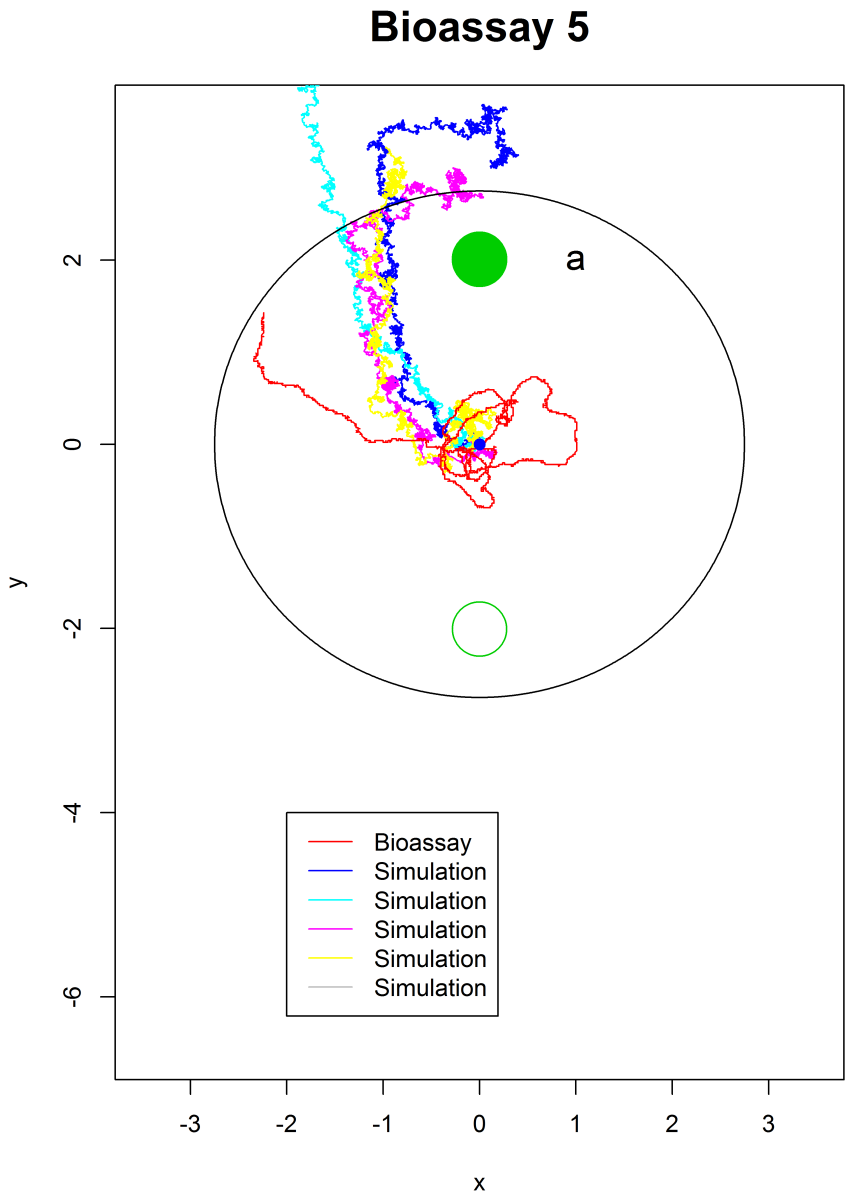


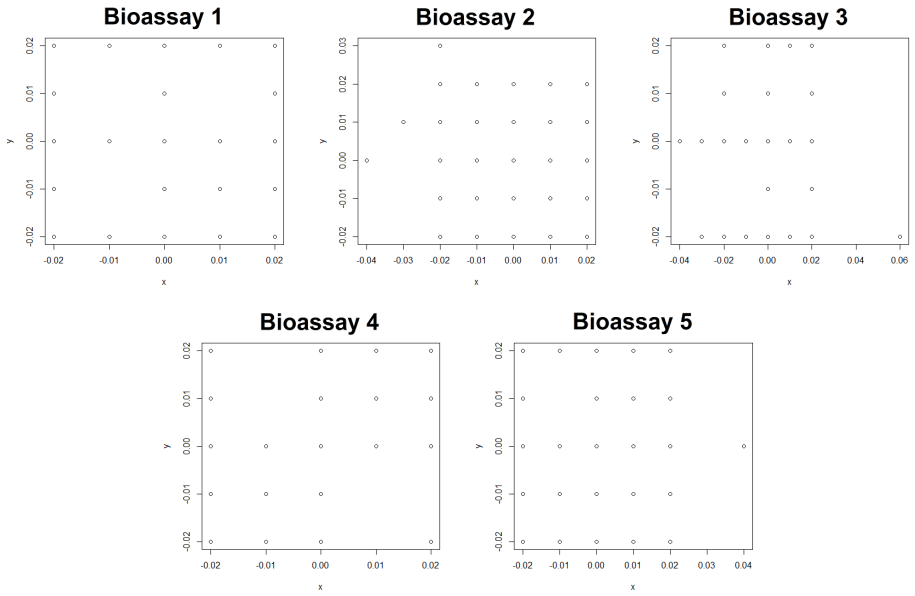
Figure 5.9: Plot of Bioassay 5 and 5 data sets of 9000 observations simulated from diffusion process (5.3), with parameter values equal to the posterior means shown in Table 5.14.



5.3.4 Discretization error

In the bioassays, the larva's position can only be measured to a certain degree of precision, and is effectively rounded to the nearest position on a grid. The length of each grid square is equal to 0.01 in the coordinates used for the bioassays. The effects of the discretization can be seen in plots of the differences between successive observations in the bioassays, displayed in Figure 5.10. It is clear that the differences are of the same order of magnitude as the grid size, and as a result they are heavily affected by discretization error. However, the data sets simulated in Sections 5.3.1 to 5.3.3 are not affected by measurement error, as each observation is simulated from a continuous distribution. The effects of discretization error can be investigated by rounding each observation from the simulated data sets to 2 decimal places. It is then possible to examine the differences between the results obtained from fitting diffusion process (5.3) to the rounded data sets and those previously obtained using the unrounded simulated data.

Figure 5.10: Plots of the differences between successive observations in Bioassays 1–5.



To investigate the effects of discretization error, 200 data sets of 9000 observations each were simulated from diffusion process (5.3), with parameters equal to the posterior means for Bioassay 1 as given in Table 5.10. The observations in each data set were rounded to 2 decimal places, and diffusion process (5.3) was fitted to each of the simulated data sets with prior distributions

for the parameters as shown in (5.5) and (5.6). Summary statistics for the posterior means of the parameters are shown in Table 5.17. The mean values of the posterior means for the elements of $\mathbf{\Gamma}$ are very similar to those obtained for the unrounded data sets in Table 5.16, while the corresponding values for the elements of $\mathbf{\Phi}$ are all larger.

Plots of the ordered observed estimated d_i^2 values for Bioassays 1–5 against the expected values for data generated by diffusion process (5.3) with parameter values equal to the posterior means for the relevant bioassay are shown in Figure 5.11. For each bioassay, the estimated d_i^2 values for 10 data sets simulated from diffusion process (5.3) with parameters equal to the posterior means obtained by fitting the diffusion process to the appropriate bioassay are also displayed. Observations in the simulated data sets are rounded to 2 decimal places.

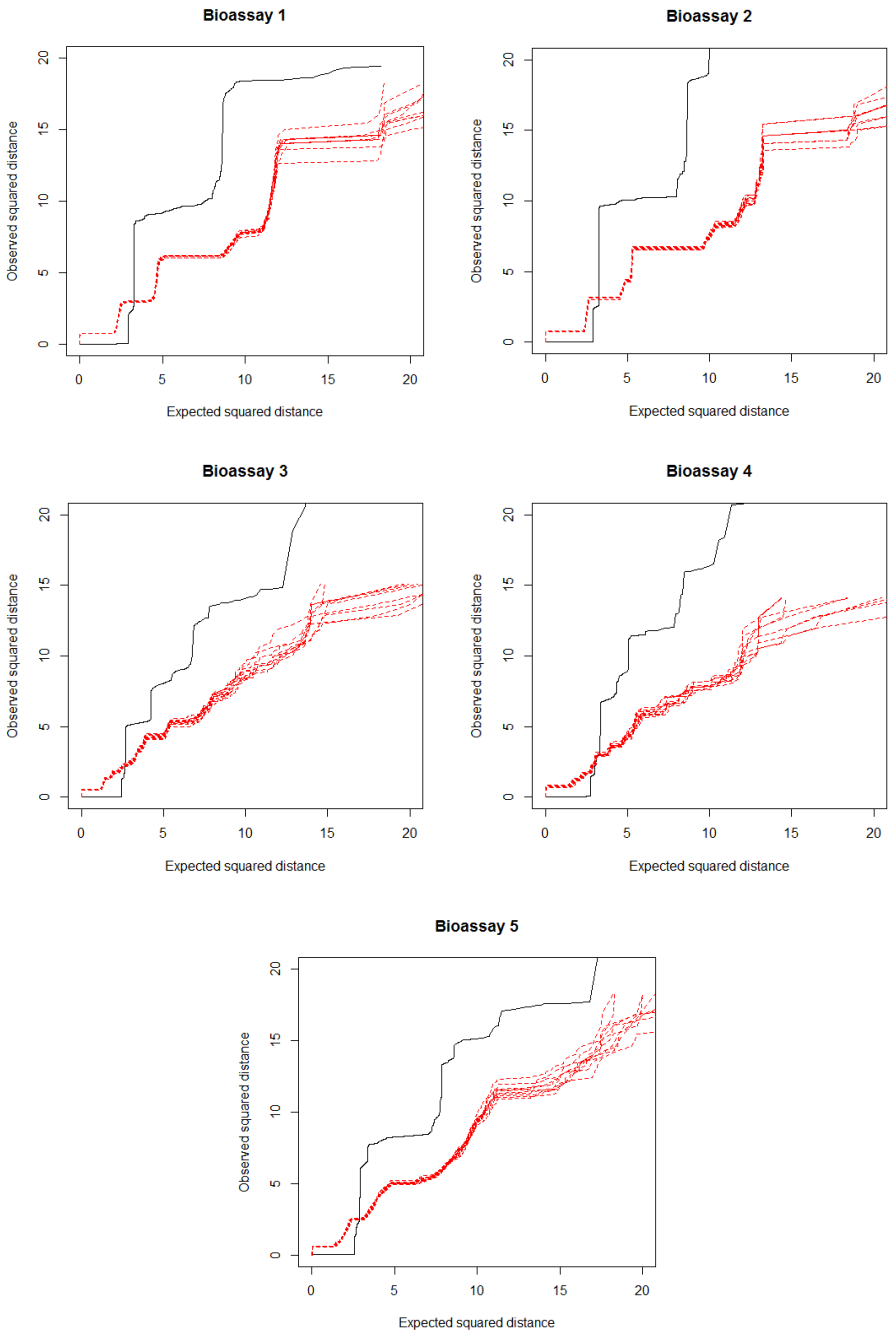
The plots of estimated d_i^2 values against expected values for the simulated data sets resemble the plots for the bioassay data considerably more than those obtained for unrounded simulated data sets in Figure 5.4. However, clear differences remain. Many of the estimated d_i^2 values for the simulated data sets appear to belong to one of several subsets of similar values, as is the case for the bioassay data. However, for all bioassays, the estimated d_i^2 values for the bioassay data are for the most part higher than those for the simulated data sets, and the smallest of the estimated values for the bioassays are smaller than those for the simulations. Discretization error is clearly responsible for some of the differences between the bioassay data and the simulated data sets, but it is not the sole reason for discrepancies between the two.

Table 5.17: Summary statistics of estimated posterior means obtained when fitting diffusion process (5.3) to 200 data sets of 9000 observations each simulated from diffusion process (5.3) with parameter values equal to the posterior means shown in Table 5.10, and rounded to 2 decimal places. Presented summary statistics include the mean and standard deviation (SD).

Parameter	**Mean**	**SD**
$\gamma_{1,1}$	0.9995	0.0002
$\gamma_{1,2}$	−0.0005	0.0001
$\gamma_{2,1}$	−0.0001	0.0002
$\gamma_{2,2}$	0.9997	0.0001
$\phi_{1,1}^{\dagger}$	5.8223	0.0881
$\phi_{1,2}^{\dagger}$	0.0599	0.0635
$\phi_{2,2}^{\dagger}$	6.6170	0.0921

† Values multiplied by 10^5.

Figure 5.11: Ordered observed estimated d_i^2 values against expected values for Bioassays 1–5, displayed as black lines. In each plot, the 10 red lines correspond to ordered observed estimated d_i^2 values from 10 data sets simulated from diffusion process (5.3), with parameters equal to the posterior means obtained by fitting the process to the relevant bioassay. Observations in the simulated data sets are rounded to 2 decimal places.



5.4 Mixture model

5.4.1 Mixture of two diffusion processes

We now consider using mixtures of multiple diffusion processes to obtain more flexible and complex models. The comparison of the observed estimated d_i^2 distances with the expected values for diffusion process (5.3) indicate that the process has limited ability to suitably represent the data. Discretization error is not solely responsible, as data sets simulated from the diffusion process and rounded to the same precision of the bioassay data still have observed estimated d_i^2 values very different from those of the bioassay data. Upon examination of the plots of simulated data sets alongside the bioassay data shown in Figures 5.5 to 5.9, it is apparent that the simulated data sets include larger random movements, whereas the movements of the larvae in the bioassays follow a more clearly defined path with much smaller localized movements. The frequent small movements are likely a product of the recording method. In the bioassays, the position recorded at each interval is that of the larva's head, and as the larvae move by expansion and contraction of their bodies, this results in frequent localized movements which are unrelated to the overall direction. The two distinct types of movement, directed and localized, can be represented by a mixture consisting of

(i) a component related to diffusion process (5.3),

$$\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_i - \mathbf{a}), \boldsymbol{\Phi}\},$$

and

(ii) a component accounting for localized movements resulting from body contractions,

$$\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}(\mathbf{x}_i, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

with a parameter $0 < \pi < 1$ representing the probability of an observation being generated from component (i). The parameter π is the proportion of time that the larva spends making movements dependent on $\mathbf{a}$, while $1 - \pi$ is the proportion of time that it spends making localized movements. The use of this

parameter allows the observations at which the larva remains stationary or makes very small movements to be modelled by a different diffusion process from its larger movements towards or away from **a**.

5.4.2 Application to bioassay data

Posterior distributions for the parameters of the mixture model defined in Section 5.4.1 fitted to the bioassay data were estimated using WinBUGS. The following independent vague prior distributions were used for $\mathbf{\Sigma}$ and π,

$$\mathbf{\Sigma} \sim \text{Inverse Wishart}\left(10^{-10}\mathbf{I}, 2\right), \tag{5.7}$$

$$(\pi, 1-\pi) \sim \text{Dirichlet}\left(\frac{1}{2}, \frac{1}{2}\right). \tag{5.8}$$

The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ were as shown in (5.5) and (5.6), giving vague prior information.

Summary statistics for the posterior distributions of the model parameters for Bioassays 1–5 are presented in Tables 5.18 to 5.22. For all 5 bioassays, the posterior means for σ_{11} and σ_{22} are several orders of magnitude smaller than those for ϕ_{11} and ϕ_{22}. This implies that component (ii) of the mixture has successfully captured the localized movements of the larvae, while state (i) represents larger directed movements.

The proportion of movements assigned to component (i) is similar for all bioassays, ranging from around 20–30%. For Bioassays 1–3 (Tables 5.18 to 5.20), the posterior means of γ_{11} and γ_{22} are less than unity, indicating attraction towards **a**. For Bioassay 4 (Table 5.21), the posterior mean of γ_{22} is greater than unity, which corresponds to repulsion. This is consistent with the path in Figure 5.3, which moves directly away from **a**. For Bioassay 5 (Table 5.22), the posterior mean of γ_{11} is greater than unity while that of γ_{22} is lower. These values are consistent with the corresponding path in Figure 5.3, which shows a track moving closer to **a** along the y-axis and further away along the x-axis.

5.4.3 Model comparison between mixture and simple diffusion process

The BIC values for the mixture model fitted to the 5 bioassays are shown in Table 5.23, along with the BIC values for diffusion process (5.3), which were given in Table 5.15. The mixture has lower BIC values for all bioassays, indicating that it is a more appropriate model for the data than process (5.3).

Table 5.18: Summary statistics of posterior distributions of the parameters of the 2 component mixture model defined in Section 5.4.1 for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are as given in (5.5), (5.6), (5.7) and (5.8).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9983	0.9983	0.0012	0.9960	1.0010
γ_{12}	−0.0005	−0.0005	0.0003	−0.0011	0.0002
γ_{21}	−0.0013	−0.0013	0.0013	−0.0038	0.0012
γ_{22}	0.9990	0.9990	0.0003	0.9983	0.9997
$\phi_{11}^{\dagger}$	1.8295	1.8270	0.5710	1.7210	1.9450
$\phi_{12}^{\dagger}$	0.0142	0.1376	0.4127	−0.0677	0.0950
$\phi_{22}^{\dagger}$	1.9351	1.9340	0.5962	1.8230	2.0540
$\sigma_{11}^{\ddagger}$	1.4489	1.4490	0.0243	1.4020	1.4980
$\sigma_{12}^{\ddagger}$	0.0003	0.0003	0.0174	−0.0343	0.0338
$\sigma_{22}^{\ddagger}$	1.4487	1.4480	0.0248	1.4010	1.4990
π	0.2328	0.2328	0.0044	0.2243	0.2416

† Values multiplied by 10^4.
‡ Values multiplied by 10^{10}.

Table 5.19: Summary statistics of posterior distributions of the parameters of the 2 component mixture model defined in Section 5.4.1 for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are as given in (5.5), (5.6), (5.7) and (5.8).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9993	0.9993	0.0006	0.9980	1.0010
γ_{12}	−0.0010	−0.0010	0.0005	−0.0018	−0.0001
γ_{21}	−0.0004	−0.0004	0.0007	−0.0016	0.0008
γ_{22}	0.9994	0.9994	0.0005	0.9986	1.0000
$\phi_{11}^{\dagger}$	1.9198	1.9190	0.0617	1.8030	2.0450
$\phi_{12}^{\dagger}$	0.1046	0.1045	0.0428	0.0204	0.1881
$\phi_{22}^{\dagger}$	1.8424	1.8410	0.0592	1.7290	1.9620
$\sigma_{11}^{\ddagger}$	1.4155	1.4150	0.0239	1.3690	1.4630
$\sigma_{12}^{\ddagger}$	−0.0001	−0.0002	0.0168	−0.0330	0.0328
$\sigma_{22}^{\dagger}$	1.4159	1.4160	0.0238	1.3700	1.4630
π	0.2151	0.2151	0.0043	0.2067	0.2235

† Values multiplied by 10^4.
‡ Values multiplied by 10^{10}.

Table 5.20: Summary statistics of posterior distributions of the parameters of the 2 component mixture model defined in Section 5.4.1 for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are as given in (5.5), (5.6), (5.7) and (5.8).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9955	0.9955	0.0036	0.9885	1.0030
γ_{12}	−0.0012	−0.0012	0.0009	−0.0030	0.0005
γ_{21}	0.0000	0.0000	0.0028	−0.0054	0.0055
γ_{22}	0.9981	0.9981	0.0007	0.9968	0.9994
$\phi_{11}^{\dagger}$	2.5991	2.5940	0.1578	2.3110	2.9230
$\phi_{12}^{\dagger}$	0.0822	0.0817	0.0856	−0.0856	0.2543
$\phi_{22}^{\dagger}$	1.5609	1.5560	0.0947	1.3880	1.7903
$\sigma_{11}^{\ddagger}$	7.6844	7.6790	0.3023	7.1150	8.2958
$\sigma_{12}^{\ddagger}$	0.0020	0.0005	0.2127	−0.4078	0.4155
$\sigma_{22}^{\ddagger}$	7.6722	7.6780	0.2985	7.1120	8.2840
π	0.2966	0.2967	0.0105	0.2761	0.3176

† Values multiplied by 10^4.
‡ Values multiplied by 10^9.

Table 5.21: Summary statistics of posterior distributions of the parameters of the 2 component mixture model defined in Section 5.4.1 for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are as given in (5.5), (5.6), (5.7) and (5.8).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9965	0.9963	0.0081	0.9812	1.0130
γ_{12}	0.0006	0.0005	0.0015	−0.0024	0.0037
γ_{21}	0.0061	0.0062	0.0090	−0.0120	0.0235
γ_{22}	1.0010	1.0010	0.0018	0.9974	1.0040
$\phi_{11}^{\dagger}$	1.3897	1.3833	0.1159	1.1808	1.6337
$\phi_{12}^{\dagger}$	0.1796	0.1781	0.0949	−0.0043	0.3688
$\phi_{22}^{\dagger}$	1.8620	1.8544	0.1550	1.5823	2.1911
$\sigma_{11}^{\ddagger}$	1.1557	1.1539	0.0567	1.0521	1.2710
$\sigma_{12}^{\ddagger}$	0.0000	0.0000	0.0404	−0.0760	0.0772
$\sigma_{22}^{\dagger}$	1.1557	1.1540	0.0569	1.0519	1.2711
π	0.2539	0.2538	0.0129	0.2289	0.2794

† Values multiplied by 10^4.
‡ Values multiplied by 10^8.

Table 5.22: Summary statistics of posterior distributions of the parameters of the 2 component mixture model defined in Section 5.4.1 for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are as given in (5.5), (5.6), (5.7) and (5.8).

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0014	1.0014	0.0004	1.0010	1.0020
γ_{12}	−0.0007	−0.0007	0.0002	−0.0010	−0.0004
γ_{21}	0.0028	0.0028	0.0004	0.0021	0.0035
γ_{22}	0.9997	0.9997	0.0001	0.9994	0.9999
$\phi_{11}^{\dagger}$	2.0160	2.0150	0.0573	1.9080	2.1310
$\phi_{12}^{\dagger}$	0.0993	0.0992	0.0385	0.0240	0.1755
$\phi_{22}^{\dagger}$	1.7976	1.7974	0.0512	1.7010	1.9010
$\sigma_{11}^{\ddagger}$	1.5329	1.5326	0.0027	1.4820	1.5860
$\sigma_{12}^{\ddagger}$	−0.0002	−0.0002	0.0019	−0.0364	0.0374
$\sigma_{22}^{\ddagger}$	1.5332	1.5333	0.0027	1.4820	1.5870
π	0.2752	0.2752	0.0046	0.2662	0.2842

† Values multiplied by 10^4.
‡ Values multiplied by 10^{10}.

Table 5.23: BIC values of diffusion process (5.3) and the 2 component mixture model fitted to Bioassays 1–5, and the differences Δ BIC between the two models.

Bioassay	**Diffusion**	**Mixture**	**Δ BIC**
1	−129412	−301536	−172124
2	−130879	−307090	−176211
3	−25378	−57859	−32481
4	−16675	−30479	−13804
5	−125996	−288701	−162705

5.4.4 Simulation plots

While the BIC values shown in Table 5.23 indicate that the mixture model is preferable to diffusion process (5.3), they provide little further information on how well the mixture characterizes the bioassay data. To investigate the fitted models further, plots of data sets simulated from the mixtures fitted to the bioassays are shown alongside the corresponding bioassay data in Figures 5.12 to 5.16.

The data sets simulated from the mixtures for Bioassays 1 and 2 (Figures 5.12 and 5.13 for Bioassays 1 and 2 respectively) follow more consistent paths than those simulated from the diffusion process (5.3) (Figures 5.5 and 5.6), without the frequent large random movements present in the latter. As a result, the data sets generated by the mixture bear a stronger resemblance to the bioassay data. Note that the different data sets simulated from the mixture model fitted to Bioassay 2 consist of movement in different directions, unlike those simulated from diffusion process (5.3) which all moved in the same direction as the bioassay data. In this case, the mixture model appears more realistic, as the paths generated by diffusion process (5.3) imply consistent attraction towards a point away from **a**. As the arena is homogeneous outside of the zones, a physical basis for such an attraction is not apparent.

For Bioassay 3, as shown in Figure 5.14, the simulated data sets consist of movement in the same overall direction as the bioassay data, but do not replicate its strong directional persistence. When compared to the simulations using diffusion process (5.3) (Figure 5.7), the mixture model produces data sets with an average direction closer to that of the bioassay data. However, the simulations from the mixture do not reach as great a distance from the origin as the real data set or those generated by diffusion process (5.3).

The data sets simulated using the mixture model for Bioassay 4 (Figure 5.15) more closely resemble the bioassay data than those generated by diffusion process (5.3) (Figure 5.8). Most of the simulations from the mixture remain close to the origin, whereas the simulations from diffusion process (5.3) consistently move in the same direction as the larva in the bioassay.

For Bioassay 5, the mixture model (Figure 5.16) clearly produces simulated data sets closer to the bioassay data than diffusion process (5.3) (Figure 5.9). While both neither model has reproduced the larva's movement around the origin prior to heading away from it, the data sets simulated using the mixture model move away from the origin in the same direction as the larva. Furthermore, these data sets clearly show more directional persistence and fewer large random movements than the data sets simulated by diffusion process (5.3).

Figure 5.12: Plot of Bioassay 1 and 5 data sets of 9000 observations simulated from the 2 component mixture model, with parameter values equal to the posterior means shown in Table 5.18.

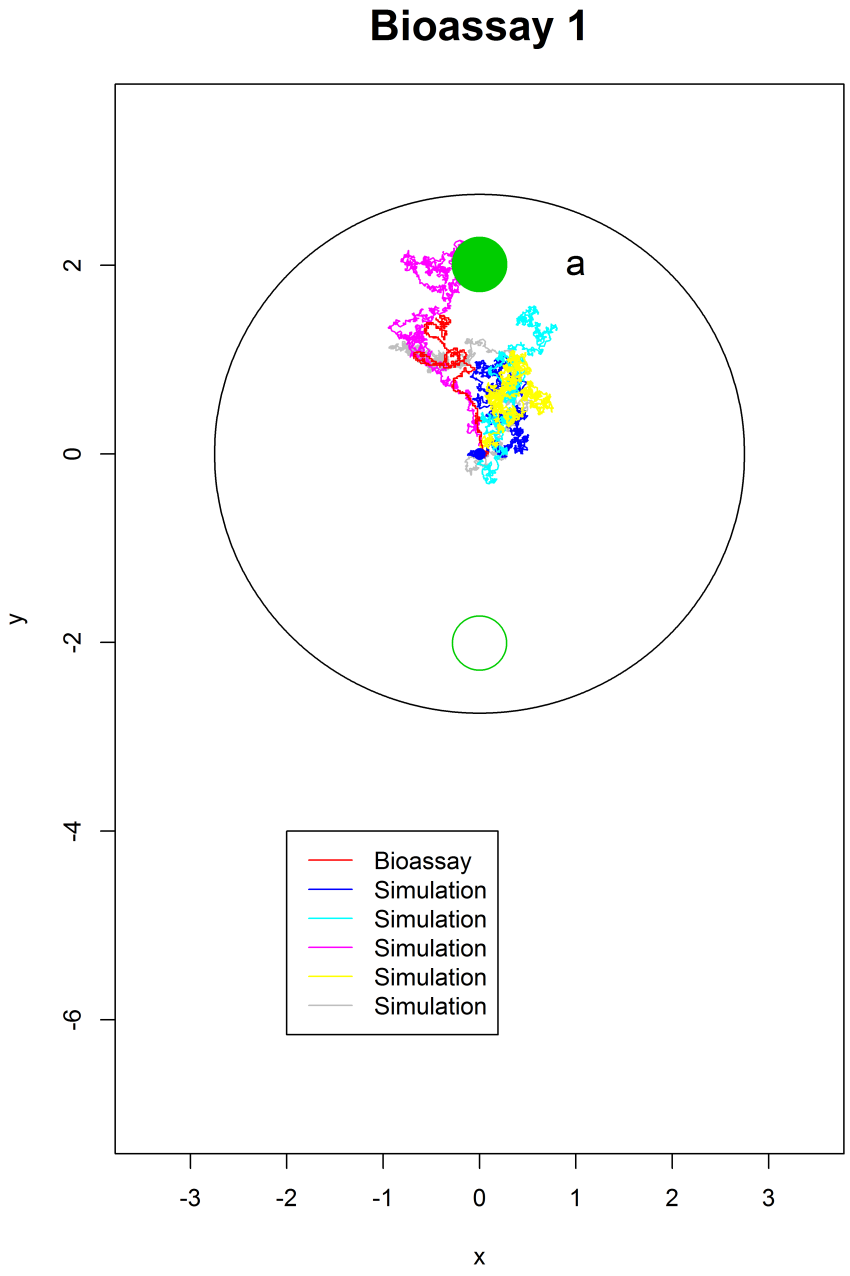


Figure 5.13: Plot of Bioassay 2 and 5 data sets of 9000 observations simulated from the 2 component mixture model, with parameter values equal to the posterior means shown in Table 5.19.

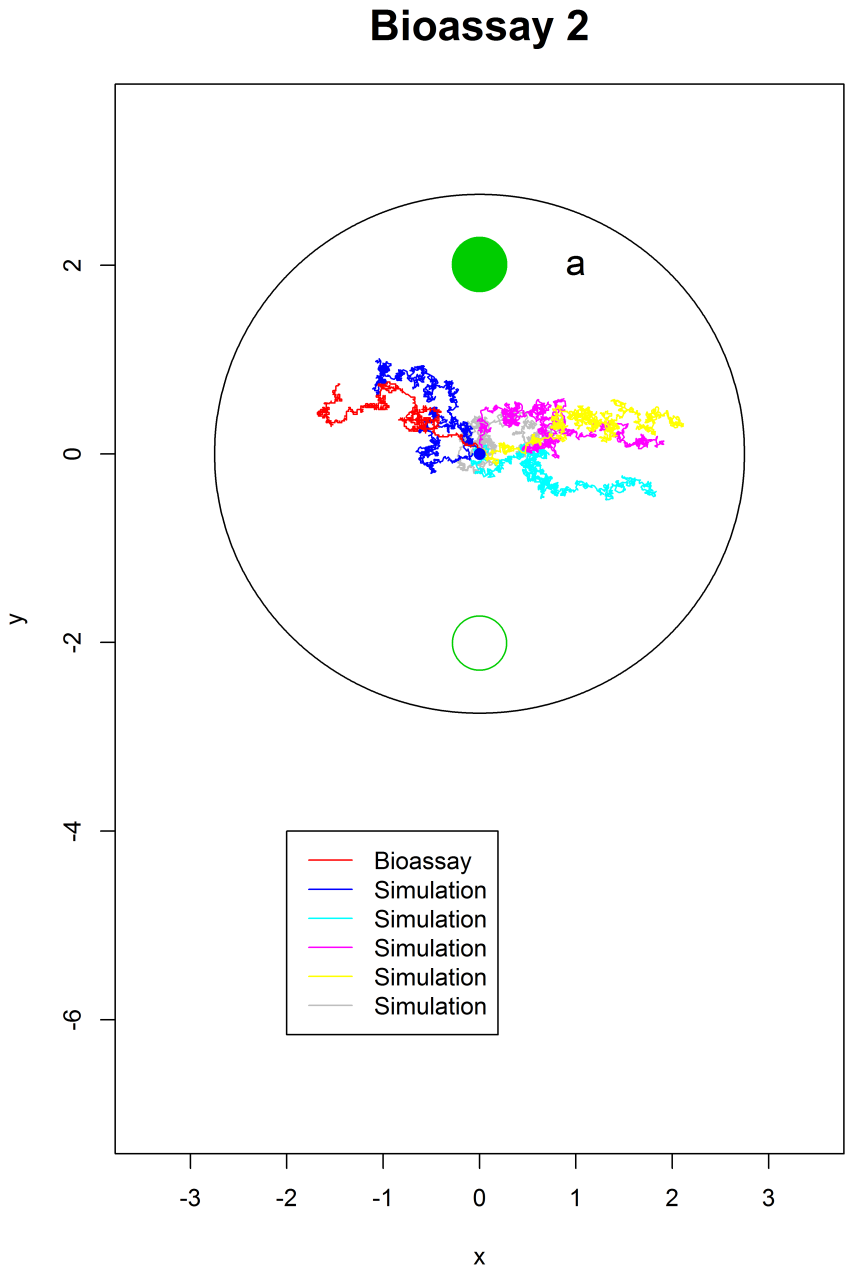


Figure 5.14: Plot of Bioassay 3 and 5 data sets of 9000 observations simulated from the 2 component mixture model, with parameter values equal to the posterior means shown in Table 5.20.

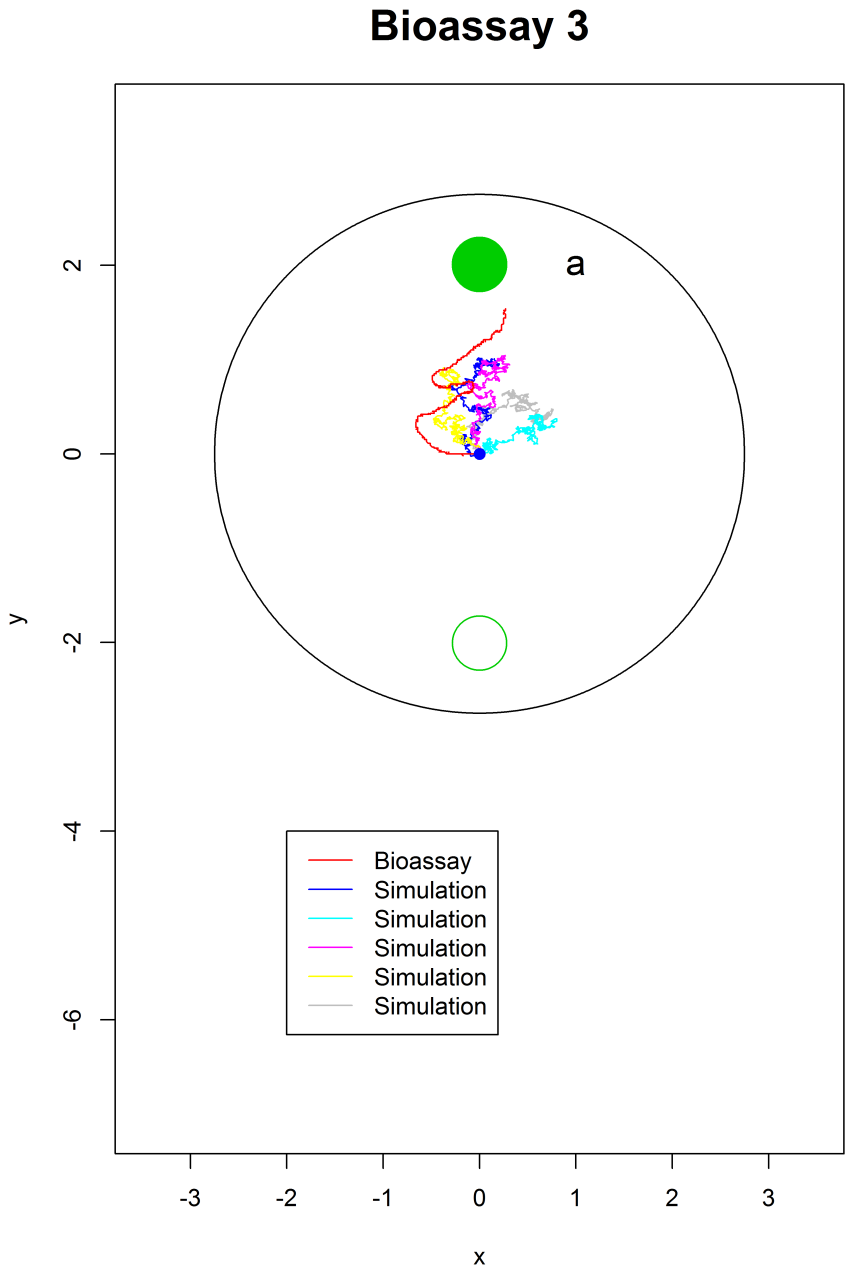


Figure 5.15: Plot of Bioassay 4 and 5 data sets of 9000 observations simulated from the 2 component mixture model, with parameter values equal to the posterior means shown in Table 5.21.

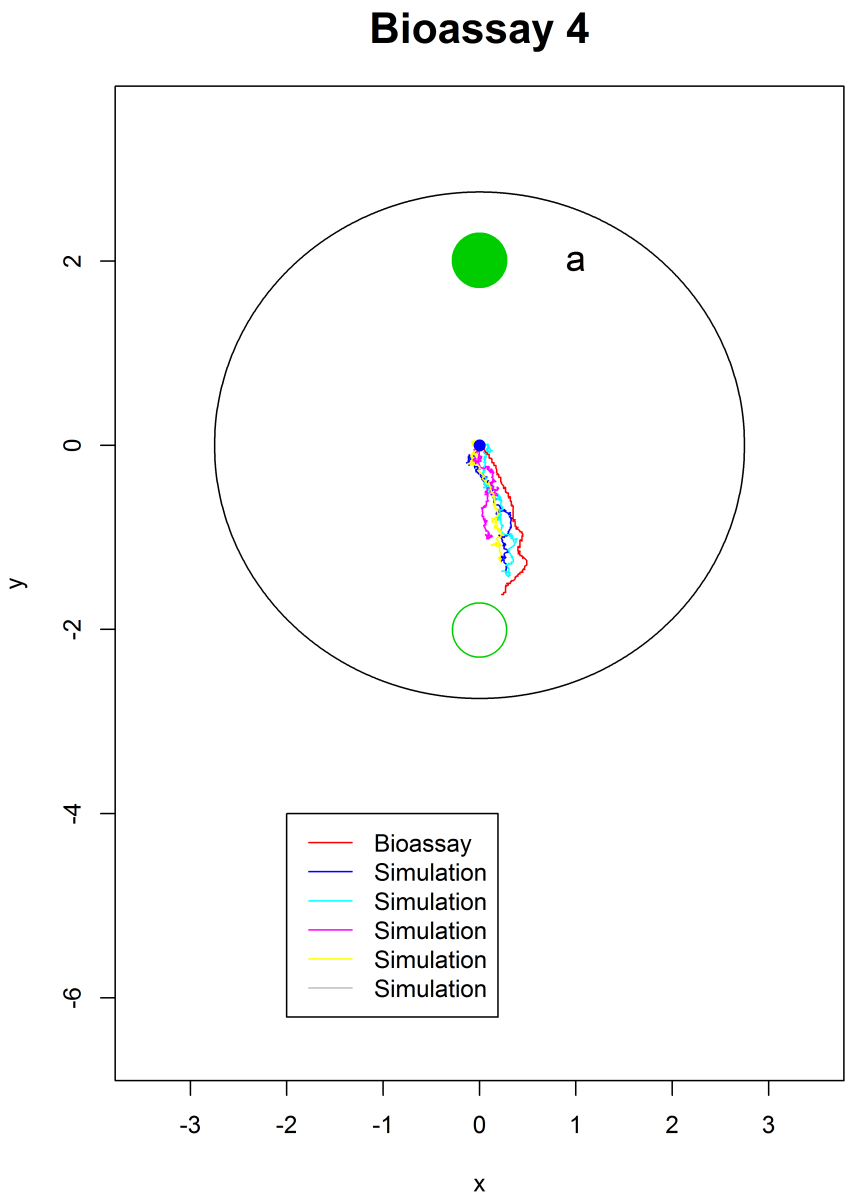
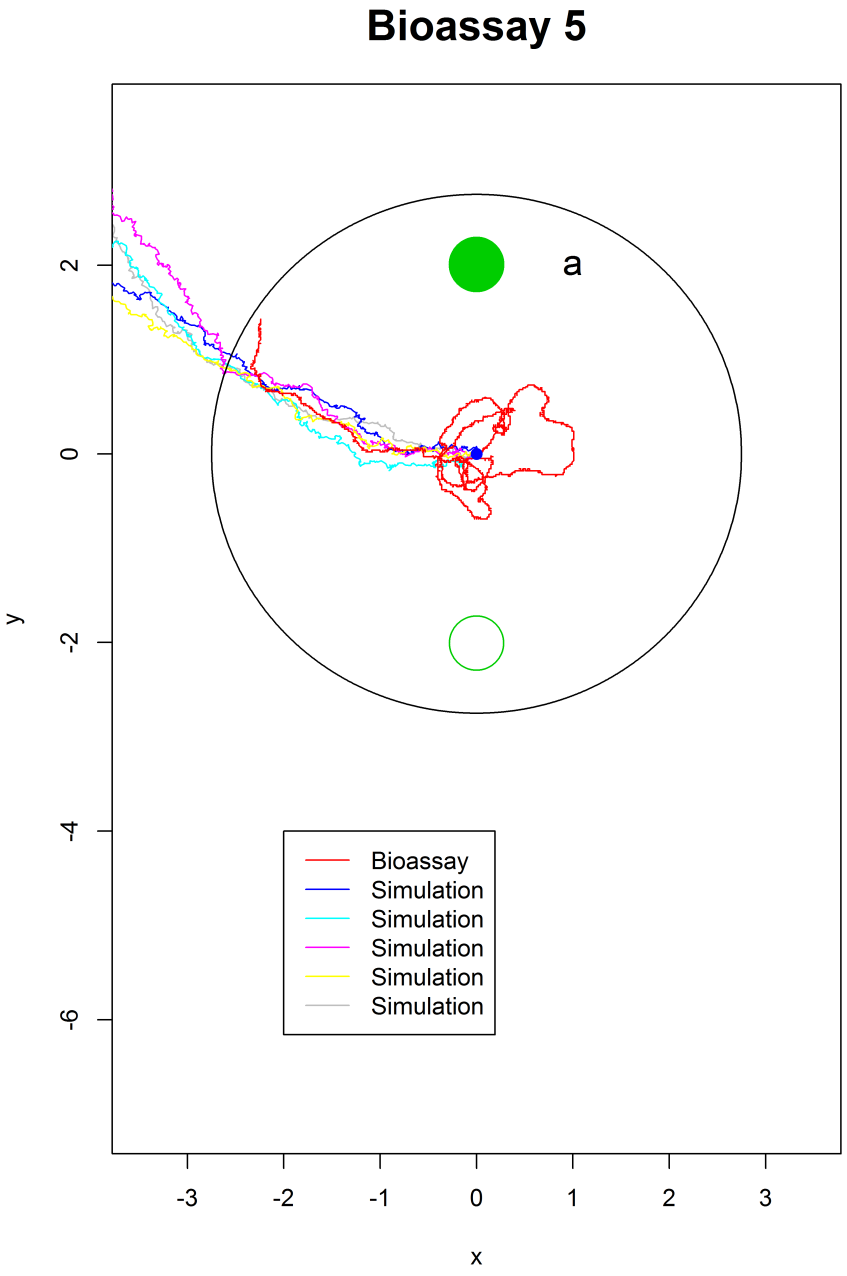


Figure 5.16: Plot of Bioassay 5 and 5 data sets of 9000 observations simulated from the 2 component mixture model, with parameter values equal to the posterior means shown in Table 5.22.



5.5 Sensitivity analysis

In Sections 5.2 to 5.4, we have used MCMC methods to fit diffusion models to the bioassay data introduced in Section 5.1 as well as various simulated data sets. The prior distributions used for the parameters of the models, given in (5.5) and (5.6), provide vague prior information and should have little effect on the posterior densities of the parameters. To investigate the sensitivity of the models to the priors, and ensure that the prior distributions used are sufficiently vague, we re-fitted the diffusion models to the bioassay data with different values for the hyperparameters.

Initially, the scale matrix in the prior distribution of the covariance matrix $\mathbf{\Phi}$ for the diffusion process (5.3) was given varying values, leaving the other hyperparameters constant as follows,

$$\begin{aligned} \gamma_{ij} &\sim \mathrm{N}(0, 10^2), \quad i, j = 1, 2, \\ \mathbf{\Phi} &\sim \text{Inverse Wishart}\left(\Omega\mathbf{I}, 2\right), \end{aligned}$$

where Ω was given a series of different values. Plots of the posterior means of selected parameters for the diffusion process fitted to Bioassay 1 using the values $\Omega = 10^{-20}, 10^{-19}, \ldots, 10^{0}$ are shown in Figures 5.17 and 5.18. The values that were used in Sections 5.2 and 5.3 are signified by red vertical lines.

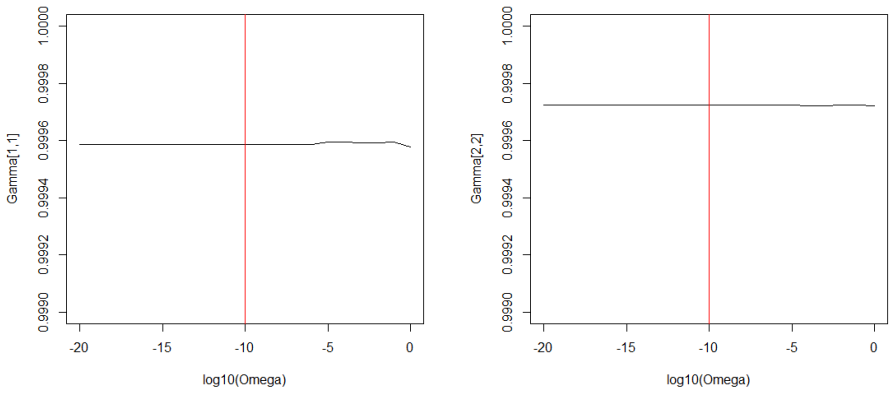
The value of Ω used has little effect on the posterior means obtained for γ_{11} and γ_{22}. The posterior means of ϕ_{11} and ϕ_{22} initially decrease quickly as Ω decreases from 1, but are almost constant for values of Ω less than or equal to 10^{-4}. Figure 5.19 shows the results on a finer scale. Note that the scale of the coordinates for ϕ_{11} is extremely small, indicating that any value of Ω equal to 10^{-5} or less will result in essentially the same posterior means, and as such is sufficiently vague.

The number of degrees of freedom in the prior distribution of $\mathbf{\Phi}$ was altered in the same way, giving prior distributions as follows,

$$\begin{aligned} \gamma_{ij} &\sim \mathrm{N}(0, 10^2), \quad i, j = 1, 2, \\ \mathbf{\Phi} &\sim \text{Inverse Wishart}\left(10^{-10}\mathbf{I}, \nu\right), \end{aligned}$$

for $\nu = 2, \ldots, 20$. Plots of posterior means are displayed in Figures 5.20 and 5.21. The posterior means do not appear to have any relationship to ν. The values obtained for ϕ_{11} and ϕ_{22}, in particular, are the same for all values of ν used.

Figure 5.17: Estimated posterior means of γ_{11} and γ_{22} for diffusion process (5.3) for Bioassay 1, using different values of Ω in the prior distribution of $\mathbf{\Phi}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.



Finally, the variance in the prior distributions of the elements of $\mathbf{\Gamma}$ for diffusion model (5.3) was given different values as follows,

$$\begin{aligned} \mathbf{\Gamma}_{i,j} &\sim \mathrm{N}(0, G)\ i, j = 1, 2, \\ \mathbf{\Phi} &\sim \text{Inverse Wishart}(10^{-10}\mathbf{I}, 2), \end{aligned}$$

where $G = 1^2, 2^2, \ldots, 20^2$. The resulting posterior mean plots are presented in Figures 5.22 and 5.23. It is clear that all values of G used result in the same posterior means, and the value of 10^2 used in previous sections of Chapter 5 is therefore sufficiently vague.

Figure 5.18: Estimated posterior means of ϕ_{11} and ϕ_{22} for diffusion process (5.3) for Bioassay 1, using different values of Ω in the prior distribution of $\mathbf{\Phi}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.

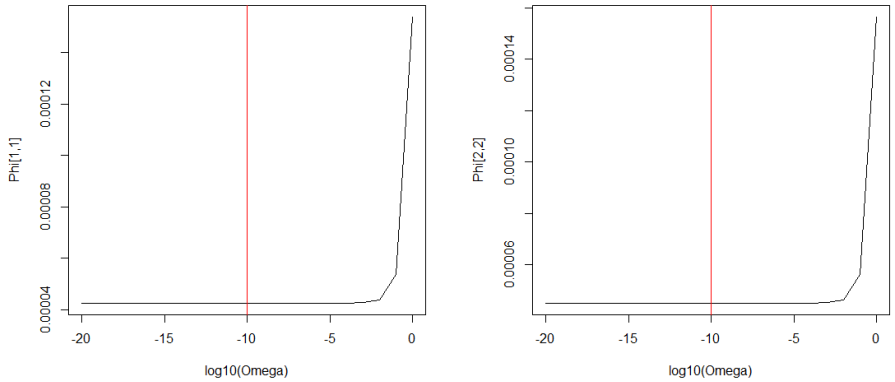


Figure 5.19: Estimated posterior means of ϕ_{11} and ϕ_{22} for diffusion process (5.3) for Bioassay 1, using different values of Ω in the prior distribution of $\mathbf{\Phi}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.

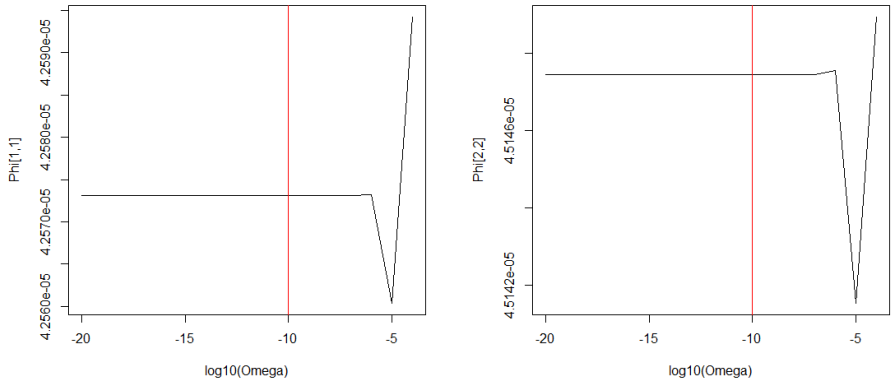


Figure 5.20: Estimated posterior means of γ_{11} and γ_{22} for diffusion process (5.3) for Bioassay 1, using different values of ν in the prior distribution of $\mathbf{\Phi}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.

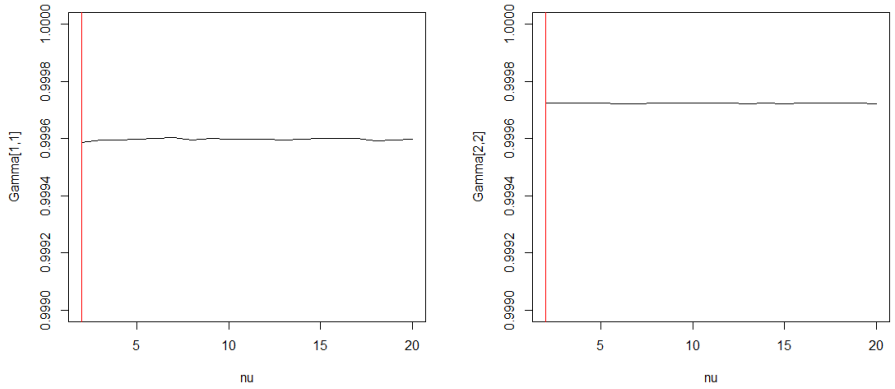


Figure 5.21: Estimated posterior means of ϕ_{11} and ϕ_{22} for diffusion process (5.3) for Bioassay 1, using different values of ν in the prior distribution of $\mathbf{\Phi}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.

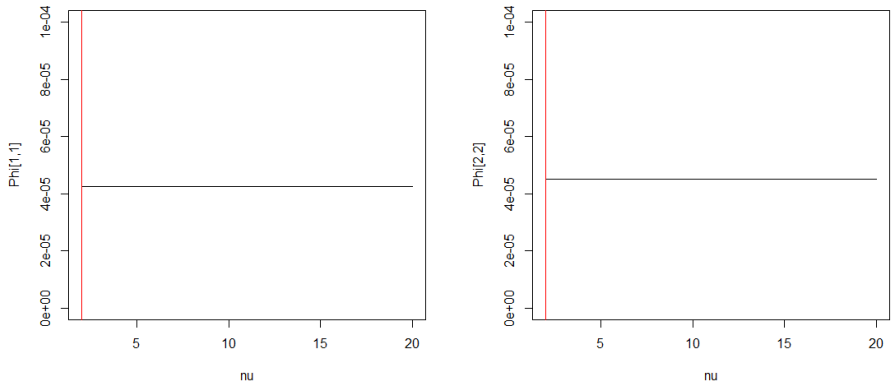


Figure 5.22: Estimated posterior means of γ_{11} and γ_{22} for diffusion process (5.3) for Bioassay 1, using different values of G in the prior distribution of $\mathbf{\Gamma}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.

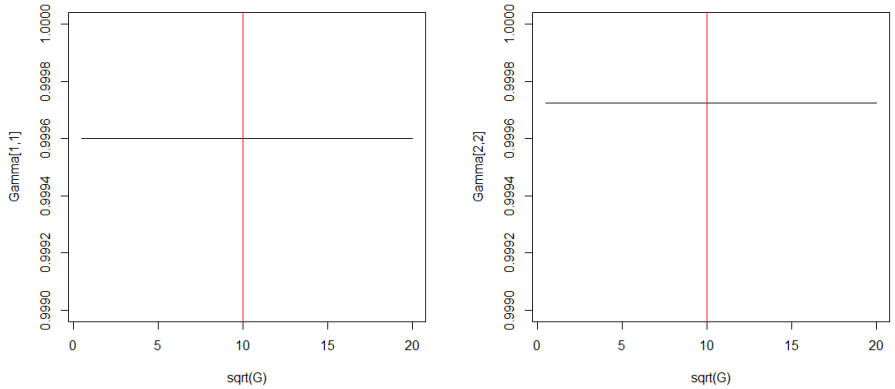
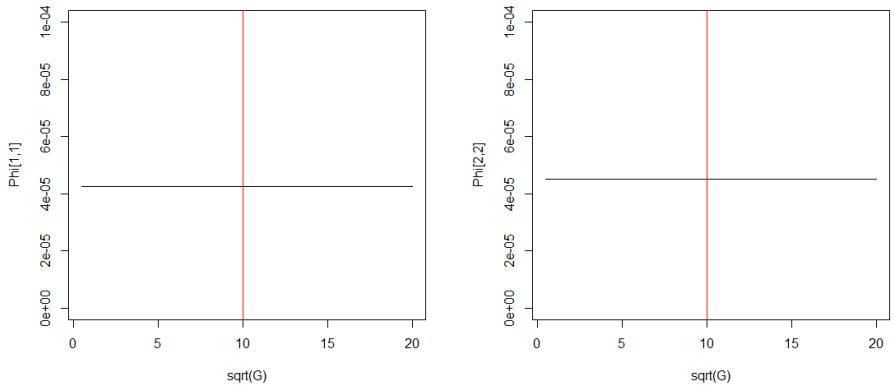


Figure 5.23: Estimated posterior means of ϕ_{11} and ϕ_{22} for diffusion process (5.3) for Bioassay 1, using different values of G in the prior distribution of $\mathbf{\Gamma}$. The values that were used for the modelling throughout Chapter 5 are signified by red vertical lines.



5.6 Discussion

We have shown that diffusion processes may be used to model the tracks of cabbage root fly larvae, and to gain greater understanding of the processes that determine their movements. It is clear even from the plots of the bioassay data displayed in Figure 5.3 that the behaviour of the larvae in the presence of allyl isothiocyanate (Bioassay 4) differs strongly from that in the case of damaged broccoli roots (Bioassays 1–3 and 5), and we can use the modelling to characterize the movements more quantitatively. The bivariate normal diffusion process (5.2) introduced in Section 5.2 successfully indicates that the larvae in Bioassays 1 and 2 are attracted towards the zones of damaged broccoli roots denoted by $\mathbf{a}$, and that the larva in Bioassay 5 moves in the region of the origin. For the larva in Bioassay 4, the parameter estimates suggest attraction in roughly the opposite direction from the allyl isothiocyanate, which is consistent with repulsion from $\mathbf{a}$. However, for Bioassay 3, diffusion process (5.2) does not appear to provide an adequate model for the data. The parameter estimates in this case imply repulsion from $\mathbf{a}$, which is clearly not consistent with the data set.

Model comparison using BIC reveals that diffusion process (5.3), which uses a fixed point of attraction $\mathbf{a}$ in place of the parameter $\boldsymbol{\mu}$, is an improvement over process (5.2). Process (5.3) also suggests attraction towards $\mathbf{a}$ for Bioassays 1 and 2 and repulsion for Bioassay 4, and it produces more appropriate results for Bioassay 3, indicating attraction towards $\mathbf{a}$.

The simulation study conducted in Section 5.3 reveals that data sets simulated from diffusion process (5.3) using the parameter estimates obtained for Bioassay 1 result in very similar parameter estimates to those of the model fitted to the bioassay data. This suggests that data sets simulated from the fitted diffusion process are similar to the real data and thus that the process is a suitable model. However, comparison of the ordered observed estimated squared Mahalanobis (d_i^2) distances with the expected squared distances for diffusion process (5.3) indicates that it is not a realistic model of the bioassay data. If the data were in fact generated by a bivariate normal diffusion process, the observed d_i^2 distances would follow a χ_2^2 distribution. The distances do not follow this distribution, and are divided into subsets of distances with similar values. Further investigation shows that discretization error is responsible for the latter behaviour, as the observed d_i^2 distances for simulated data sets rounded to the same precision as the bioassay data also divide into subsets with similar values. However, plots of the observed d_i^2 distances still demonstrate noticeable differences from the bioassay data, indicating that discretization error is not the sole cause of discrepancies

between the real data and the simulations.

Comparison of the bioassay data to the simulated data sets indicates that the latter contain larger random movements and fail to replicate the systematic localized movements present in the former. The 2 component mixture model introduced in Section 5.4 is intended to improve on the single diffusion process by including a component that corresponds to localized movements independent of the point of attraction or repulsion **a**. In this respect it is successful, as the posterior means for the parameter values for all bioassays indicate that one component consists of larger movements influenced by the position of **a**, and the other is comprised of much smaller movements independent of **a**. Model comparison using BIC shows that the mixture is superior to the simpler diffusion process.

Plots of real and simulated data for Bioassays 1, 2 and 4 show that the mixture captures the features of the data more successfully than diffusion process (5.3), as the simulated data sets strongly resemble the real movements of the larvae. For Bioassay 3, the mixture gives similar results to the single diffusion process. Both models correctly indicate that the larva is attracted towards **a**, and produce simulated data sets that move in the same overall direction as the larva but make more frequent turns. For Bioassay 5, the mixture improves on the single diffusion process, but the resulting simulated data sets are still very different from the real data.

It is clear that diffusion models provide a useful tool for describing the movements of larvae, and comparing their responses to different chemical compounds. By doing so, we can gain a greater understanding of the underlying processes governing their movements. Although it is possible to collect extensive sets of data with the currently available experimental equipment, care is needed to incorporate important features of the movement into our models of larvae paths. In this respect, the mixture improves on the single diffusion process by accounting for the frequent small localized movements present in the data sets, and thus provides a more realistic model of larva movement.

While the mixture used here is successful at identifying attraction to and repulsion from **a** and provides a suitable model for the bioassays, further refinement would be useful in cases of strong directional persistence. The plots of Bioassays 3 and 4 in particular suggest that the movements of the larvae in these bioassays are strongly dependent on the current direction of movement, whereas the mixture used here only involves first order dependence and so is fairly limited in its ability to model this behaviour. One natural development would be to incorporate higher-order dependence into the model. Another possibility is to use

an alternative model where the probability that an observation belongs to a given mixture component is dependent on the component that generated the previous observation. This would require replacing the parameter π with a probability transition matrix, which would allow for different probabilities of switching from each state and result in a far more flexible model. Additional components can also be added to account for features of the data as necessary.

Chapter 6

Further modelling of bioassay data

6.1 Introduction

We propose more complex alternatives to the diffusion model and 2 component mixture of diffusion processes used in Chapter 5 for modelling the cabbage root fly larva bioassay data. Previously we observed that the mixture model could be improved by using a probability transition matrix for switching between different states, rather than the single parameter π used by the mixture, which results in a Hidden Markov Model (HMM). HMMs offer a flexible way to describe data structures (Frühwirth-Schnatter, 2006), and have been used to model data in various fields such as finance, meteorology, medicine and speech recognition (Juang and Rabiner, 1991; Robert et al., 2000; Altman, 2007; Langrock and Zucchini, 2011). They have previously been used to model the movements of animals (MacDonald and Raubenheimer, 1995; Blackwell, 2003; Franke et al., 2004, 2006; Patterson et al., 2009; Langrock et al., 2012; Schliehe-Diecks et al., 2012). Animal movement often consists of different movement types corresponding to different behavioural states (Morales et al., 2004; Benhamou, 2007), which can be modelled as the different states of a HMM. However, the bioassay data sets consist of movements on a much smaller scale than these examples.

We have observed that the movements of the larvae are highly dependent on the previous movement direction, whereas the mixture models used previously include only first-order dependence. Therefore, we consider the use of diffusion processes with higher-order dependence, which can more adequately represent the directional persistence of the bioassay data.

A series of HMMs with 2, 3, 4 and 5 states are defined. Maximum likelihood estimates are not easily obtained for HMMs, and MCMC methods are normally used to obtain parameter estimates (Robert et al., 1993; Chib, 1996). We therefore obtain posterior distributions using MCMC, and the results are discussed. Second-order and third-order diffusion processes are also introduced and fitted to the data, and the competing models are compared using BIC. A model incorporating both higher-order dependence and multiple states is then proposed. The model is a mixture of 2 diffusion processes, one of which has second-order dependence. BIC values for this model applied to the bioassays are compared to the results for the diffusion process and 2 component mixture model defined in Chapter 5, and the second-order diffusion process used in this chapter.

6.2 Hidden Markov models for larva movements

We propose the use of an HMM as a more flexible alternative to the mixture model introduced for the purpose of modelling cabbage root fly larva movements in Chapter 5. Initially, we consider a model with the following 2 states,

(i) $\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_i - \mathbf{a}), \boldsymbol{\Phi}\}$, and

(ii) $\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}(\mathbf{x}_i, \boldsymbol{\Sigma})$.

States (i) and (ii) are identical to the components of the mixture model introduced in Chapter 5. The HMM has a probability transition matrix

$$\boldsymbol{P} = \begin{pmatrix} \pi_{11} & \dots & \pi_{1n} \\ \vdots & \vdots & \vdots \\ \pi_{n1} & \dots & \pi_{nn} \end{pmatrix},$$

where π_{ij} is defined by

$$Pr\left(\text{Observation belongs to state } j | \text{Previous observation belongs to state } i\right).$$

In the transition matrix $\mathbf{P}$, n is equal to the number of states in the HMM, and in this case $n = 2$. The HMM is more computationally intensive than the simpler mixture, but can model a wider range of larvae behaviour. The inclusion of further states allows for still greater flexibility in the modelling. We observe that the bioassay data sets introduced in Chapter 5 include many identical successive observations due to discretization error and the short time interval between observations. These stationary observations can be represented by the

addition of a third "null" state to the HMM in which an observation is identical to the previous one,

(iii) $\mathbf{x}_{i+1}$ is the same as $\mathbf{x}_i$.

When considering the addition of a fourth state, we note that movement towards the zone of chemical compound **a** can be interpreted as repulsion from the control zone **b**, and similarly that movement away from **a** can be explained by attraction towards **b**. Attraction to or repulsion from the control zone can be represented by a fourth state

(iv) a diffusion process with point of attraction or repulsion **b**,

$$\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\{\mathbf{b} + \mathbf{\Lambda}(\mathbf{x}_i - \mathbf{b}), \mathbf{\Psi}\},$$

where $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are defined as

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}, \ \mathbf{\Psi} = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}.$$

It is possible that the strength of attraction towards **a** varies. However, using only state (i), we are unable to represent this adequately. By using a further state of a similar form to (i), we can incorporate this feature as part of the HMM. The additional state takes the form

(v) $\mathbf{x}_{i+1}|\mathbf{x}_i \sim \mathrm{N}\{\mathbf{a} + \mathbf{\Gamma}^*(\mathbf{x}_i - \mathbf{a}), \mathbf{\Phi}^*\}$,

where

$$\mathbf{\Gamma}^* = \begin{pmatrix} \gamma^*_{11} & \gamma^*_{12} \\ \gamma^*_{21} & \gamma^*_{22} \end{pmatrix}, \ \mathbf{\Phi}^* = \begin{pmatrix} \phi^*_{11} & \phi^*_{12} \\ \phi^*_{21} & \phi^*_{22} \end{pmatrix}.$$

6.3 Application to bioassay data

Posterior distributions for the parameters of the HMMs fitted to the bioassay data were estimated using WinBUGS, using the following independent vague

prior distributions,

$$\gamma_{ij} \sim \mathrm{N}\left(0, 10^2\right), \quad i, j = 1, 2, \tag{6.1}$$

$$\lambda_{ij} \sim \mathrm{N}\left(0, 10^2\right), \quad i, j = 1, 2, \tag{6.2}$$

$$\gamma^*_{ij} \sim \mathrm{N}\left(0, 10^2\right), \quad i, j = 1, 2, \tag{6.3}$$

$$\mathbf{\Phi} \sim \text{Inverse Wishart}\left(10^{-5}\mathbf{I}, 2\right), \tag{6.4}$$

$$\mathbf{\Sigma} \sim \text{Inverse Wishart}\left(10^{-5}\mathbf{I}, 2\right), \tag{6.5}$$

$$\mathbf{\Psi} \sim \text{Inverse Wishart}\left(10^{-5}\mathbf{I}, 2\right), \tag{6.6}$$

$$\mathbf{\Phi^*} \sim \text{Inverse Wishart}\left(10^{-5}\mathbf{I}, 2\right), \tag{6.7}$$

$$(\pi_{i1}, \ldots, \pi_{in}) \sim \text{Dirichlet}\left(\frac{1}{2}, \ldots, \frac{1}{2}\right), \quad i = 1, \ldots, n, \tag{6.8}$$

for those of the parameters present in each HMM. Summary statistics for the posterior distributions of the models obtained for Bioassays 1–5 using the HMMs with 2, 3 and 4 states are presented in Sections 6.3.1 to 6.3.3.

6.3.1 2-state HMM

The summary statistics obtained for the 2-state HMM consisting of states (i) and (ii) defined in Section 6.2 fitted to the bioassay data are shown in Tables 6.1 to 6.5. We can also calculate π, the probability that an observation is generated from state (i), which is given by

$$\pi = \frac{1 - \pi_{22}}{2 - \pi_{11} - \pi_{22}}. \tag{6.9}$$

For Bioassay 1 (Table 6.1), the posterior means of π_{11} and π_{22} and equation (6.9) suggest that the larva spends about 13% of the time in state (i), which is considerably lower than the estimate of 23% obtained for the mixture model using the same states in Table 5.18. The posterior mean of γ_{22} is less than unity, which as mentioned in Section 5.2.1 corresponds to attraction towards $\mathbf{a}$, though the smaller proportion of movements assigned to state (i) in comparison to the mixture model means that the overall influence of $\mathbf{a}$ is weaker. The posterior means of σ_{11} and σ_{22} reveal that state (ii) consists primarily of movement along the x-axis, as σ_{11} is several orders of magnitude larger than σ_{22}. However, both are much smaller than the posterior means of ϕ_{11} and ϕ_{22}. The fitted 2-state HMM consists of one diffusion process containing larger movements influenced by $\mathbf{a}$, and one with smaller localized movements, as was the case for the mixture model in Chapter 5.

The results for Bioassay 2 (Table 6.2) are similar to those for Bioassay 1 in that the model consists of switching between one diffusion process with larger movements towards **a** and a second with small localized movements. However, in this case the posterior means of π_{11} and π_{22} give an estimate 21% of the time spent in state (i). This is much closer to the corresponding value for the mixture model (Table 5.19) than the corresponding estimates for Bioassay 1.

For Bioassay 3 (Table 6.3), the posterior mean of σ_{11} is several orders of magnitude larger than that of σ_{22} and is of similar magnitude to the posterior means of ϕ_{11} and ϕ_{22}. These results indicate that state (ii) consists almost entirely of movement along the x-axis. The model therefore involves 2 diffusion processes of similar order of magnitude, whereas the summary statistics for the corresponding mixture model (Table 5.20) suggested one diffusion process consisting of larger movements and the other of smaller movements. The posterior mean of γ_{22} implies attraction towards **a**, and the posterior means of π_{11} and π_{22} indicate that the larva spends 13% of the time in state (i). This estimate is considerably lower than that for the mixture model applied to the same bioassay.

The summary statistics for Bioassay 4 (Table 6.4) indicate that the larva is repelled from **a**, as the posterior mean of γ_{22} is greater than unity. As was found for Bioassays 1 and 2, the posterior means of σ_{11} and σ_{22} are much smaller than those of ϕ_{11} and ϕ_{22}, indicating that one component is comprised of larger movements away from **a** and the other of smaller undirected movements. The posterior means of π_{11} and π_{22} give an estimate of 25% of the time spent in state (i), similar to that for the mixture.

For Bioassay 5 (Table 6.5), the posterior mean for γ_{22} is less than unity and that of γ_{11} is greater, indicating that the larva moves towards **a** along the y-axis and away along the x-axis. This behaviour was also suggested by the summary statistics for the mixture model in Table 5.22. Similarly to Bioassays 1–2 and 4, the posterior means for ϕ_{11}, ϕ_{22}, σ_{11} and σ_{22} indicate larger movements influenced by **a** and smaller movements independent of **a**. The posterior means of π_{11} and π_{22} give an estimate of 27% of the time spent in state (i), which is again close to the corresponding value for the fitted mixture model.

For all bioassays, the HMM provides similar estimates of attraction to or repulsion from **a** to the mixture model. The fitted mixture models for all bioassays are mixtures of 2 diffusion processes, one of comparatively large movements with attraction to or repulsion from **a**, and one of smaller localized movements. The fitted HMMs for Bioassays 1, 2, 4 and 5 also have such components. However, for Bioassay 3, both states of the fitted HMM consist of movements with similar order of magnitude. In general, the use of a probability transition matrix in place

of the single parameter π in the mixture model allows the HMM to model more complex behaviour patterns in the data. For Bioassays 1 and 3, the estimates of the times spent in states (i) and (ii) given by the HMM differ considerably from those for the mixture model, suggesting that the greater flexibility of the HMM results in a noticeably improved model. The mixture and HMM give similar estimates for Bioassays 2, 3 and 5 in this respect.

Table 6.1: Summary statistics of posterior distributions of the parameters of the 2-state HMM defined in Section 6.3.1 for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0000	1.0000	0.0008	0.9984	1.0020
γ_{12}	−0.0001	−0.0001	0.0002	−0.0006	0.0003
γ_{21}	−0.0023	−0.0022	0.0022	−0.0046	0.0020
γ_{22}	0.9983	0.9983	0.0006	0.9971	0.9994
$\phi^{\dagger}_{11}$	0.5110	0.5104	0.0211	0.4710	0.5533
$\phi^{\dagger}_{12}$	0.0279	0.0272	0.0392	−0.0456	0.1068
$\phi^{\dagger}_{22}$	3.4539	3.4510	0.1440	3.1840	3.7470
$\sigma^{\ddagger}_{11}$	0.4132	0.4131	0.0659	0.4005	0.4264
$\sigma^{\ddagger}_{12}$	−0.0104	−0.0132	0.0010	−0.0161	0.0169
$\sigma^{\ddagger}_{22}$	0.0013	0.0013	0.0012	0.0010	0.0013
π_{11}	0.3919	0.3918	0.0143	0.3639	0.4202
π_{22}	0.9090	0.9090	0.0033	0.9024	0.9152

† Values multiplied by 10^4.
‡ Values multiplied by 10^{10}.

6.3.2 3-state HMM

The 3-state HMM consists of states (i), (ii) and (iii) defined in Section 6.2. It is therefore similar to the 2-state HMM, with the addition of a third "null" state in which each observation is the same as the previous one. The summary statistics for the 3-state model fitted to the bioassays are presented in Tables 6.6 to 6.10. The BUGS code used to fit the model is given in Appendix C. As was done for the 2-state HMM, we can calculate the overall state probabilities π_1, π_2 and π_3 for states 1, 2 and 3 respectively. To do so, we observe that the vector $\pi = (\pi_1, \pi_2, \pi_3)$

Table 6.2: Summary statistics of posterior distributions of the parameters of the 2-state HMM defined in Section 6.3.1 for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9993	0.9993	0.0006	0.9981	1.0010
γ_{12}	−0.0010	−0.0010	0.0004	−0.0019	−0.0002
γ_{21}	−0.0004	−0.0004	0.0006	−0.0016	0.0009
γ_{22}	0.9994	0.9994	0.0004	0.9986	1.0002
$\phi^{\dagger}_{11}$	1.9819	1.9170	0.0621	1.8010	2.0450
$\phi^{\dagger}_{12}$	0.1034	0.1044	0.0430	0.0177	0.1860
$\phi^{\dagger}_{22}$	1.8412	1.8400	0.0604	1.7260	1.9590
$\sigma^{\ddagger}_{11}$	1.4161	1.4160	0.0242	1.3690	1.4630
$\sigma^{\ddagger}_{12}$	−0.0001	−0.0002	0.0169	−0.0315	0.0326
$\sigma^{\ddagger}_{22}$	1.4157	1.4160	0.0240	1.3710	1.4640
π_{11}	0.3919	0.3912	0.0112	0.3691	0.4135
π_{22}	0.8331	0.8331	0.0044	0.8245	0.8414

† Values multiplied by 10^4.
‡ Values multiplied by 10^9.

must satisfy the equation

$$\pi = \pi \mathbf{P}, \tag{6.10}$$

where $\mathbf{P}$ is the probability transition matrix of the HMM. The posterior mean estimates obtained for the elements of $\mathbf{P}$ with MCMC are used, and (6.10) is solved in R to obtain estimates for π_1, π_2 and π_3.

For Bioassays 1, 2 and 3, as shown in Tables 6.6, 6.7 and 6.8 respectively, the posterior means of γ_{11} and γ_{22} are less than unity, implying attraction towards $\mathbf{a}$. For Bioassay 4 (Table 6.9), the posterior mean of γ_{11} is equal to unity while that of γ_{22} is greater, resulting in movement away from $\mathbf{a}$. Table 6.10 shows the summary statistics for Bioassay 5, for which the posterior mean of γ_{11} is greater than unity and that of γ_{22} is lower, indicating movement towards $\mathbf{a}$ along the y-axis and away along the x-axis. These findings concur with those for the 2-state HMM and mixture model, and appear to successfully describe the data sets as seen in Figure 5.3.

For Bioassays 1 and 2, the posterior means of σ_{11} and σ_{22} reveal that state (ii) consists primarily of movement along the y-axis, as σ_{22} is much larger than

Table 6.3: Summary statistics of posterior distributions of the parameters of the 2-state HMM defined in Section 6.3.1 for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0050	1.0050	0.0038	0.9976	1.0130
γ_{12}	0.0013	0.0013	0.0011	−0.0009	0.0035
γ_{21}	−0.0073	−0.0074	0.0064	−0.0195	0.0053
γ_{22}	0.9933	0.9934	0.0018	0.9896	0.9968
$\phi_{11}^{\dagger}$	1.3044	1.3010	0.1216	1.0850	1.5470
$\phi_{12}^{\dagger}$	0.2207	0.2270	0.1397	−0.0714	0.4973
$\phi_{22}^{\dagger}$	3.3395	3.3295	0.3070	2.7760	3.9740
$\sigma_{11}^{\dagger}$	0.6888	0.6883	0.0235	0.6445	0.7356
$\sigma_{12}^{\ddagger}$	−0.0287	−0.0132	0.0010	−0.0161	0.0169
$\sigma_{22}^{\ddagger}$	0.0006	0.0006	0.0000	0.0001	0.0007
π_{11}	0.3406	0.3406	0.0307	0.2835	0.3992
π_{22}	0.9002	0.9004	0.0077	0.8849	0.9153

† Values multiplied by 10^4.
‡ Values multiplied by 10^{13}.

σ_{11}. This differs from the 2-state HMM, where the movements in state (ii) were much smaller than those in state (i). For Bioassay 4, ϕ_{22} is larger than ϕ_{11} and σ_{11} is larger than σ_{22}, so the localized movements are primarily along the x-axis and the directed movements along the y-axis. The posterior means of σ_{11} for Bioassays 3 and 5 are of the same order of magnitude as σ_{22}, again suggesting 2 states with movements of roughly similar distance plus a null state. In general, the estimated sizes of the movements in state (ii) are larger than those obtained for the 2-state HMM, which may be interpreted as stationary observations that were assigned to state (ii) by the 2-state model instead being assigned to the null state in by the 3-state model.

The posterior means of the elements of π for the bioassays are displayed in Table 6.11, and indicate that the larvae spend the majority of the time in state (iii) for all bioassays, that is, that the majority of observations are identical to the previous one. State (iii) clearly accounts for a large part of the behaviour exhibited in the data sets, and its addition to the third-state HMM should therefore result in a greatly improved model over the 2-state alternative.

Table 6.4: Summary statistics of posterior distributions of the parameters of the 2-state HMM defined in Section 6.3.1 for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0013	1.0010	0.0080	0.9876	1.0160
γ_{12}	−0.0001	0.0000	0.0016	−0.0031	0.0027
γ_{21}	0.0039	0.0045	0.0018	−0.0013	0.0020
γ_{22}	1.0011	1.0010	0.0018	0.9837	1.0050
$\phi^{\dagger}_{11}$	1.3890	1.3820	0.1136	1.1840	1.6311
$\phi^{\dagger}_{12}$	0.1777	0.1750	0.0959	−0.0010	0.3741
$\phi^{\dagger}_{22}$	1.8520	1.8430	0.1542	1.5760	2.1790
$\sigma^{\ddagger}_{11}$	1.1578	1.1540	0.0560	1.0570	1.2750
$\sigma^{\ddagger}_{12}$	−0.0004	−0.0001	0.0394	−0.0783	0.0743
$\sigma^{\ddagger}_{22}$	1.1528	1.1510	0.0540	1.0520	1.2601
π_{11}	0.4427	0.4420	0.0286	0.3870	0.4962
π_{22}	0.8106	0.8111	0.0136	0.7827	0.8362

† Values multiplied by 10^4.
‡ Values multiplied by 10^{13}.

6.3.3 4-state HMM

Summary statistics for the HMM with states (i), (ii), (iii) and (iv) as defined in Section 6.2 are presented in Appendix D due to occupying a large amount of space. The posterior means of γ_{11} and γ_{22} indicate repulsion from **a** for Bioassay 4 and attraction for the other bioassays. For Bioassay 1, the posterior means of λ_{11} and λ_{22} suggest repulsion from **b**, which also results in movement towards **a**. However, for Bioassays 2–5, the corresponding values imply attraction towards **b**. The attraction towards **b** is weaker than attraction towards **a** for Bioassays 2–5, although only slightly so in the case of Bioassay 5. These results are consistent with the fact that attraction towards **a** appears strongest in Bioassay 1 and upon inspection of Figure 5.3. The larva in Bioassay 5 shows less attraction towards **a** than those in Bioassays 1–3, which is also consistent with the values obtained here for the elements of $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$.

For Bioassay 1, the posterior mean of σ_{11} is much larger than that of σ_{22}, suggesting that the localized movements of state (ii) are almost entirely along the x-axis. This stands in contrast with the 3-state HMM (Table 6.6), where the movements in state (ii) were primarily along the y-axis, and a plausible

Table 6.5: Summary statistics of posterior distributions of the parameters of the 2-state HMM defined in Section 6.3.1 for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0014	1.0010	0.0005	1.0010	1.0020
γ_{12}	−0.0007	−0.0007	0.0001	−0.0010	−0.0004
γ_{21}	0.0028	0.0028	0.0004	0.0020	0.0035
γ_{22}	0.9997	0.9997	0.0001	0.9994	0.9999
$\phi^{\dagger}_{11}$	2.0157	2.0150	0.0570	1.9070	2.1310
$\phi^{\dagger}_{12}$	0.0977	0.0988	0.0379	0.0184	0.1686
$\phi^{\dagger}_{22}$	1.7968	1.7960	0.0504	1.7040	1.8980
$\sigma^{\ddagger}_{11}$	1.5325	1.5320	0.0274	1.4800	1.5286
$\sigma^{\ddagger}_{12}$	0.0003	0.0003	0.0193	−0.0372	0.0386
$\sigma^{\ddagger}_{22}$	1.5324	1.5320	0.0272	1.4800	1.5870
π_{11}	0.4019	0.4111	0.0099	0.3911	0.4307
π_{22}	0.7763	0.7764	0.0051	0.7663	0.7858

† Values multiplied by 10^4.
‡ Values multiplied by 10^{10}.

interpretation is that vertical movements assigned to state (ii) by the 3-state model have here been allocated to state (iv).

The summary statistics for Bioassays 3, 4 and 5 likewise indicate that the movements in state (ii) are primarily along the x-axis. For Bioassays 3 and 4, the posterior means of ϕ_{22} are much larger than those of ϕ_{11}, showing that state (i) consists mainly of movement along the y-axis. This is also true of state (iii) for Bioassay 3. These results imply that for Bioassay 3 in particular, and to a lesser extent for Bioassay 4, vertical movement is for the most part influenced by $\mathbf{a}$, while horizontal movement is not. For Bioassay 2, the posterior mean of σ_{22} is larger than for the other bioassays which implies more undirected vertical movement and less consistent attraction to or repulsion from $\mathbf{a}$. This appears consistent with the plots in Figure 5.3, when the larva's track is compared to the more strongly directed movement seen in Bioassays 1 and 4 in particular.

Estimates of the overall state probabilities are obtained with (6.10), using the posterior mean estimates for the elements of the probability transition matrix $\mathbf{P}$ for each bioassay. The estimates are shown in Table 6.12, and the larvae in all of the bioassays spend the majority of time in the null state (iii), as was also the case for the 3-state HMM.

Table 6.6: Summary statistics of posterior distributions of the parameters of the 3-state HMM defined in Section 6.3.2 for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9967	0.9966	0.0024	0.9951	1.0010
γ_{12}	−0.0002	−0.0002	0.0002	−0.0007	0.0001
γ_{21}	0.0007	0.0007	0.0007	−0.0009	0.0024
γ_{22}	0.9996	0.9996	0.0003	0.9991	1.0001
$\phi^{\dagger}_{11}$	3.5161	3.5100	0.1572	3.2170	3.8321
$\phi^{\dagger}_{12}$	0.3020	0.3030	0.2534	−0.2175	0.8867
$\phi^{\dagger}_{22}$	0.5411	0.5398	0.0235	0.4969	0.5896
$\sigma^{\ddagger}_{11}$	0.0997	0.0997	0.0045	0.0916	0.1088
$\sigma^{\ddagger}_{12}$	0.0112	0.0108	0.0009	−0.0105	0.0157
$\sigma^{\dagger}_{22}$	3.4643	3.4610	0.1553	3.1700	3.7670
π_{11}	0.4108	0.4108	0.0148	0.3819	0.4409
π_{12}	0.0694	0.0693	0.0079	0.0548	0.0860
π_{13}	0.5198	0.5199	0.0151	0.4900	0.5493
π_{21}	0.0613	0.0610	0.0074	0.0473	0.0764
π_{22}	0.3707	0.3704	0.0150	0.3418	0.3996
π_{23}	0.5680	0.5683	0.0155	0.5371	0.5978
π_{31}	0.0844	0.0845	0.0033	0.0780	0.0911
π_{32}	0.0806	0.0856	0.0032	0.0745	0.0869
π_{33}	0.8440	0.8350	0.0043	0.8263	0.8433

† Values multiplied by 10^4.
‡ Values multiplied by 10^8.

6.3.4 Model comparison

The Bayesian Information Criterion (BIC) was used to compare the competing HMMs. BIC values for the 2-state, 3-state and 4-state HMMs fitted to the bioassays are presented in Table 6.13, along with the values for the simple diffusion process and 2 component mixture model previously shown in Table 5.23. The mixture and the 2-state HMM are improvements over the single diffusion process, and Table 6.13 shows that the fitted 2-state HMMs for Bioassays 1,3, 4 and 5 have improved BIC over the simpler mixture models, but for Bioassay 2 the BIC is worse. However, the 3-state HMMs incorporating null movement result in greatly improved BIC values for all 5 bioassays, and are therefore to be preferred over both 2 component models. The addition of a fourth state describing influence

Table 6.7: Summary statistics of posterior distributions of the parameters of the 3-state HMM defined in Section 6.3.2 for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9986	0.9986	0.0012	0.9964	1.0010
γ_{12}	−0.0005	−0.0005	0.0006	−0.0018	0.0008
γ_{21}	−0.0009	−0.0009	0.0003	−0.0015	−0.0002
γ_{22}	0.9998	0.9998	0.0003	0.9992	1.0004
$\phi_{11}^{\dagger}$	3.4421	3.4360	0.1492	3.1760	3.7561
$\phi_{12}^{\dagger}$	−0.3034	−0.3044	0.0995	−0.4964	−0.1002
$\phi_{22}^{\dagger}$	0.5811	0.5804	0.0249	0.5334	0.6318
$\sigma_{11}^{\ddagger}$	0.1168	0.1167	0.0056	0.1063	0.1280
$\sigma_{12}^{\ddagger}$	−0.0121	−0.0121	0.1033	−0.1463	0.2236
$\sigma_{22}^{\dagger}$	3.4310	3.4240	0.1687	3.1280	3.7900
π_{11}	0.2850	0.2847	0.0136	0.2584	0.3122
π_{12}	0.1011	0.1007	0.0092	0.0845	0.1199
π_{13}	0.6139	0.6142	0.0144	0.5865	0.6414
π_{21}	0.1267	0.1262	0.0111	0.1054	0.1483
π_{22}	0.2715	0.2713	0.0152	0.2427	0.3016
π_{23}	0.6018	0.6015	0.0165	0.5686	0.6326
π_{31}	0.0938	0.0938	0.0035	0.0871	0.1006
π_{32}	0.0732	0.0732	0.0029	0.0676	0.0790
π_{33}	0.8330	0.8330	0.0043	0.8244	0.8413

† Values multiplied by 10^4.
‡ Values multiplied by 10^7.

by the control zone **b** provides further improvement over the 3-state model for Bioassays 1, 3 and 5, though the difference in BIC is smaller than the difference between the 2-state and 3-state models. For Bioassays 2 and 4, the 3-state model is superior. The 5-state model with states (i) to (v) described in Section 6.2 is also fitted to the bioassays, but is not selected for any of them, and is considered to be unnecessarily complex for the data.

The strength of the influence that the position of **a** has on the larva's movements is determined by $|\mathbf{\Gamma}|$. However, in Section 5.2.1 we observed that the strength of attraction to or repulsion from **a** is primarily determined by the value of the parameter γ_{22}, with a value less than unity indicating attraction and a value greater than unity corresponding to repulsion. Posterior density plots of γ_{22} and $|\mathbf{\Gamma}|$ for Bioassays 1–5 are displayed in Figure 6.1. It is apparent

Table 6.8: Summary statistics of posterior distributions of the parameters of the 3-state HMM defined in Section 6.3.2 for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9931	0.9933	0.0043	0.9845	1.0020
γ_{12}	−0.0107	−0.0087	0.0010	−0.0126	−0.0087
γ_{21}	0.0116	0.0114	0.0066	0.0001	0.0240
γ_{22}	0.9971	0.9968	0.0020	0.9941	1.0019
$\phi_{11}^{\dagger}$	0.8116	0.7974	0.1085	0.6337	1.0410
$\phi_{12}^{\dagger}$	−0.6948	−0.6904	0.0980	−0.9024	−0.5078
$\phi_{22}^{\dagger}$	1.1770	1.1790	0.1508	0.8796	1.4890
$\sigma_{11}^{\dagger}$	2.7320	2.7160	0.2226	2.3300	3.2180
$\sigma_{12}^{\dagger}$	0.0525	0.0526	0.1128	−0.1809	0.2722
$\sigma_{22}^{\dagger}$	1.5340	1.5210	0.1333	1.2970	1.8210
π_{11}	0.0030	0.0012	0.0041	0.0000	0.0153
π_{12}	0.4544	0.4547	0.0439	0.3682	0.5381
π_{13}	0.5426	0.5419	0.0436	0.4611	0.6283
π_{21}	0.2340	0.2340	0.0447	0.1486	0.3169
π_{22}	0.2159	0.2147	0.0475	0.1260	0.3085
π_{23}	0.5501	0.5505	0.0296	0.4909	0.6041
π_{31}	0.0886	0.0883	0.0115	0.0659	0.1115
π_{32}	0.1424	0.1422	0.0126	0.1189	0.1666
π_{33}	0.7690	0.7692	0.0116	0.7458	0.7903

† Values multiplied by 10^4.

that the majority of the posterior densities of γ_{22} and $|\mathbf{\Gamma}|$ are above unity for Bioassay 4 and below unity for the other bioassays. We also observe that the standard deviations of γ_{22} for Bioassays 1, 2 and 5 are much smaller than those for Bioassays 3 and 4. The models have successfully distinguished between the attraction exhibited by larvae exposed to damaged broccoli roots and repulsion from allyl isothiocyanate. The difference in variance between the bioassays is less pronounced in the case of $|\mathbf{\Gamma}|$, and we consider γ_{22} to be a more helpful measure of attraction and repulsion.

Table 6.9: Summary statistics of posterior distributions of the parameters of the 3-state HMM defined in Section 6.3.2 for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0000	1.0000	0.0001	0.9998	1.0000
γ_{12}	−0.0005	−0.0004	0.0013	−0.0030	0.0024
γ_{21}	0.0026	0.0026	0.0021	−0.0019	0.0065
γ_{22}	1.0025	1.0025	0.0035	0.9957	1.0095
$\phi_{11}^{\ddagger}$	0.6232	0.6182	0.0711	0.5004	0.7806
$\phi_{12}^{\ddagger}$	0.1825	0.1826	0.0854	0.0097	0.3422
$\phi_{22}^{\dagger}$	2.6028	2.5890	0.2917	2.0394	3.2302
$\sigma_{11}^{\dagger}$	3.1366	3.1025	0.3898	2.4370	3.9681
$\sigma_{12}^{\ddagger}$	−0.0056	−0.0056	0.0277	−0.0790	0.0554
$\sigma_{22}^{\ddagger}$	0.8225	0.8167	0.1035	0.6434	1.0411
π_{11}	0.3295	0.3289	0.0364	0.2592	0.4017
π_{12}	0.1329	0.1318	0.0260	0.0849	0.1887
π_{13}	0.5377	0.5380	0.0389	0.4588	0.6118
π_{21}	0.0606	0.0575	0.0223	0.0248	0.1212
π_{22}	0.3626	0.3616	0.0483	0.2800	0.4450
π_{23}	0.5767	0.5769	0.0437	0.4925	0.6628
π_{31}	0.1178	0.1173	0.0110	0.0981	0.1402
π_{32}	0.0714	0.0712	0.0088	0.0550	0.0893
π_{33}	0.8107	0.8110	0.0130	0.7848	0.8347

† Values multiplied by 10^4.
‡ Values multiplied by 10^7.

Table 6.10: Summary statistics of posterior distributions of the parameters of the 3-state HMM defined in Section 6.3.2 for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.4) and (6.5).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0010	1.0010	0.0005	0.9998	1.0020
γ_{12}	−0.0055	−0.0055	0.0002	−0.0059	−0.0051
γ_{21}	0.0019	0.0019	0.0005	0.0008	0.0029
γ_{22}	0.9955	0.9955	0.0002	0.9950	0.9959
$\phi_{11}^{\dagger}$	0.9577	0.9553	0.0492	0.8684	1.0630
$\phi_{12}^{\dagger}$	−0.9144	−0.9119	0.0479	−1.0110	−0.8249
$\phi_{22}^{\dagger}$	0.9925	0.9906	0.0510	0.8987	1.0980
$\sigma_{11}^{\dagger}$	2.0260	2.0230	0.0724	1.8830	2.1730
$\sigma_{12}^{\dagger}$	0.1833	0.1823	0.0472	0.0921	0.2786
$\sigma_{22}^{\dagger}$	1.8810	1.8810	0.0678	1.7500	2.0200
π_{11}	0.0013	0.0006	0.0019	0.0000	0.0063
π_{12}	0.4001	0.3999	0.0201	0.3617	0.4392
π_{13}	0.5987	0.5986	0.0200	0.5598	0.6365
π_{21}	0.1872	0.1877	0.0130	0.1609	0.2129
π_{22}	0.2293	0.2293	0.0144	0.2028	0.2582
π_{23}	0.5835	0.5735	0.0128	0.5581	0.6081
π_{31}	0.0787	0.0788	0.0040	0.0710	0.0867
π_{32}	0.1450	0.1449	0.0002	0.1355	0.1551
π_{33}	0.7764	0.7763	0.0051	0.7663	0.7869

† Values multiplied by 10^4.

Table 6.11: Estimates of the overall state probabilities for the 3-state HMM defined in Section 6.3.2 for Bioassays 1–5. The estimates were obtained using (6.10) with the posterior mean estimates for the elements of the probability transition matrix $\mathbf{P}$ for each bioassay.

	Bioassays				
Parameter	**1**	**2**	**3**	**4**	**5**
π_1	0.1215	0.1199	0.1070	0.1412	0.0916
π_2	0.1116	0.0955	0.1897	0.1130	0.1839
π_3	0.7669	0.7847	0.7032	0.7458	0.7247

Table 6.12: Estimates of the overall state probabilities for the 4-state HMM defined in Section 6.3.3 for Bioassays 1–5. The estimates were obtained using (6.10) with the posterior mean estimates for the elements of the probability transition matrix $\mathbf{P}$ for each bioassay.

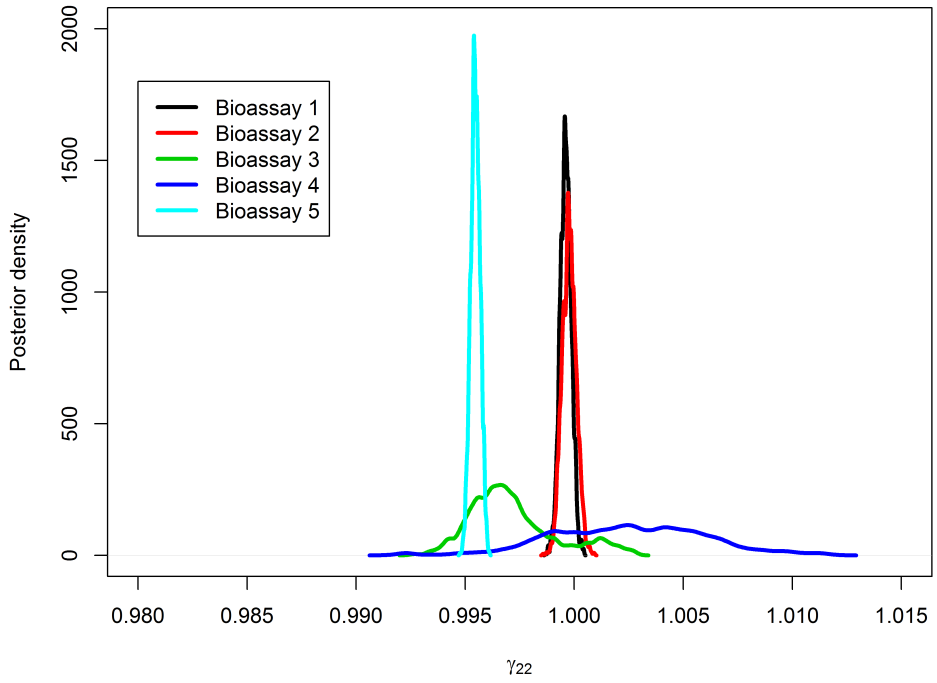
	Bioassays				
Parameter	**1**	**2**	**3**	**4**	**5**
π_1	0.0631	0.0469	0.2185	0.1408	0.1970
π_2	0.1261	0.1315	0.1552	0.0430	0.1252
π_3	0.7687	0.7845	0.5655	0.7440	0.6033
π_4	0.0421	0.0371	0.0608	0.0721	0.0745

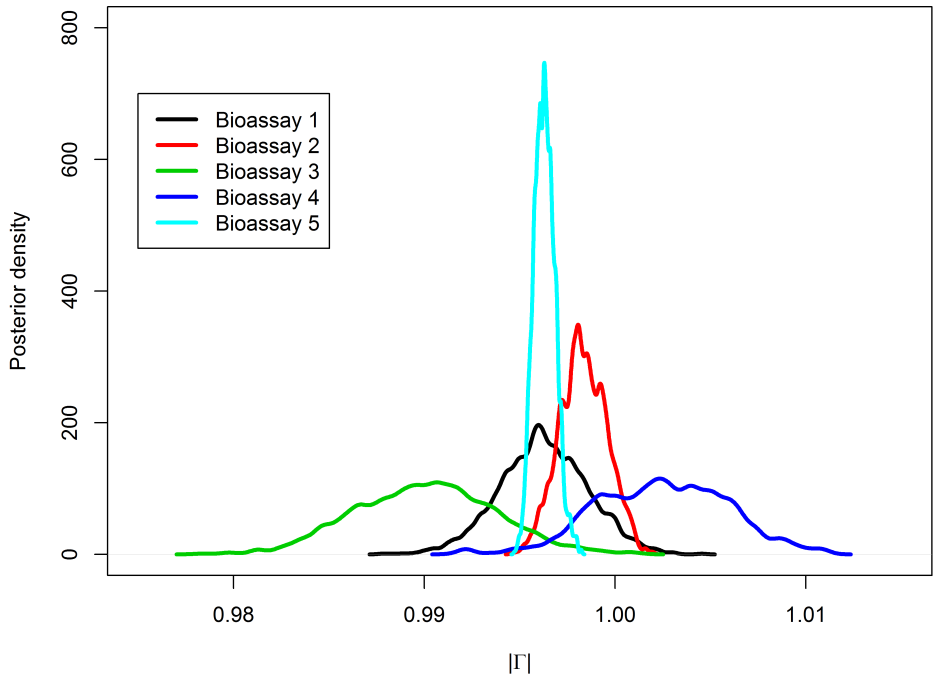
Table 6.13: BIC values of competing models fitted to Bioassays 1–5. The models used are diffusion process (5.3), the 2 component mixture model described in Section 5.4, and the HMMs with 2, 3 and 4 states introduced in Section 6.2.

	Number of	**Bioassay**		
Model	**parameters**	**1**	**2**	**3**
Single diffusion process	7	−129412	−130879	−25378
2 component mixture	11	−301536	−307090	−57859
2-state HMM	12	−307090	−284115	−59500
3-state HMM	16	−581769	−590949	−115580
4-state HMM	29	−588411	−588236	−116282
5-state HMM	44	−583972	−584099	−113369

	Number of	**Bioassay**	
Model	**parameters**	**4**	**5**
Single diffusion process	7	−16675	−125996
2 component mixture	11	−30479	−288701
2-state HMM	12	−51800	−302091
3-state HMM	16	−73337	−547732
4-state HMM	29	−73224	−561436
5-state HMM	44	−72851	−551699

Figure 6.1: Posterior density plots of the parameters γ_{22} (upper panel) and $|\mathbf{\Gamma}|$ (lower panel) for the 3-state HMMs fitted to Bioassays 1–5 in Section 6.3.2.





6.4 Higher order dependence

6.4.1 Autocorrelation in bioassay data

Visual examination of Figure 5.3 strongly suggests that the direction of movement in Bioassays 3–5 in particular has some dependence on the previous direction. This feature of the data is not accounted for by the models used thus far, which assume only first-order dependence. The R function `pacf` was used to produce plots of the partial correlation coefficients for each of the 5 bioassays, which are displayed in Figures 6.2 to 6.6.

Figure 6.2 shows evidence of significant correlation at up to sixth order for the x-coordinates of the data, and at up to third order for the y-coordinates. The spikes corresponding to first order correlation are much larger than subsequent spikes, and the correlation coefficients at sixth and third order for the x and y-coordinates respectively are only just significant at the 5% level. Nevertheless, there is evidence of higher-order dependence for Bioassay 1.

For Bioassay 2, there is evidence of significant correlation at up to fourth order for the x-coordinates of the data, and up to fifth for the y-coordinates. Again, the spikes corresponding to first order correlation are much larger than at subsequent lags, particularly for the x-coordinates, for which the spike for fourth order correlation is only just significant at the 5% level, and the spike for third order correlation is not.

There is less evidence of dependence upon previous observations in Bioassay 3 than was observed for Bioassays 1 and 2. For the x-coordinates the only significant spikes are for first and second order dependence, and the former is much larger than the latter. For the y-coordinates the only significant spike is for first order dependence. There is no evidence of anything beyond first-order dependency for the y-coordinates, and limited evidence in the case of the x-coordinates when compared to Bioassays 1 and 2. It is not obvious from the data plots that Bioassay 3 should exhibit less dependence on previous observations than Bioassays 1 and 2. In fact, the larva in Bioassay 3 appears to maintain a more consistent direction.

The results for Bioassay 4 are similar to those for Bioassay 3. There are significant spikes for first and second order dependence for the x-coordinates, with the former much larger than the latter, and only for first order dependence for the y-coordinates. For Bioassay 5, there is evidence of significant correlation at up to eighth order for the x-coordinates of the data, and up to seventh for the y-coordinates. However, while the spikes for seventh and eighth order for

the x-coordinates are significant at the 5% level, those for fourth, fifth and sixth order are not. Similarly, the spikes at fifth and sixth order for the y-coordinates are not significant at the 5% level. However, the plots still provide evidence of higher-order dependence in the bioassay data.

Directional dependence in animal movement data is frequently accounted for by random walk models (Kareiva and Shigesada, 1983; Bovet and Benhamou, 1988; Bergman et al., 2000; Byers, 2001; Jonsen et al., 2005; Johnson et al., 2008a, 2008b) with the step length and turning angle between successive observations as parameters. In the case of a diffusion process, it can be incorporated into the model by introducing higher order dependence into state (i) so that the conditional distribution of an observation $\mathbf{x}_i$ given previous observations $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i-2}$ takes the form

$$\mathbf{x}_i|\mathbf{x}_{i-1},\ \mathbf{x}_{i-2} \sim \mathrm{N}\left\{\mathbf{a} + \mathbf{\Gamma}_1(\mathbf{x}_{i-1} - \mathbf{a}) + \mathbf{\Gamma}_2(\mathbf{x}_{i-2} - \mathbf{a}), \mathbf{\Phi}\right\}.$$

Figure 6.2: Partial correlation plots for Bioassay 1.

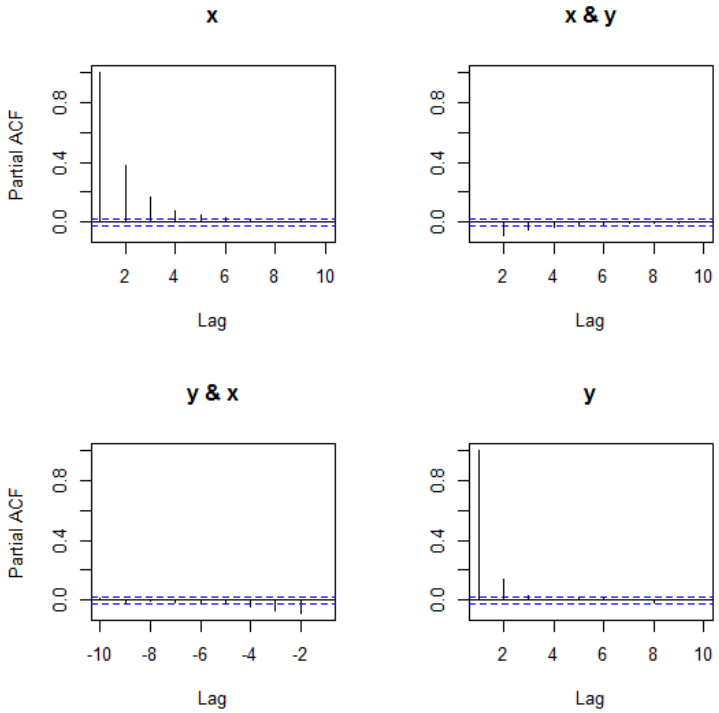


Figure 6.3: Partial correlation plots for Bioassay 2.

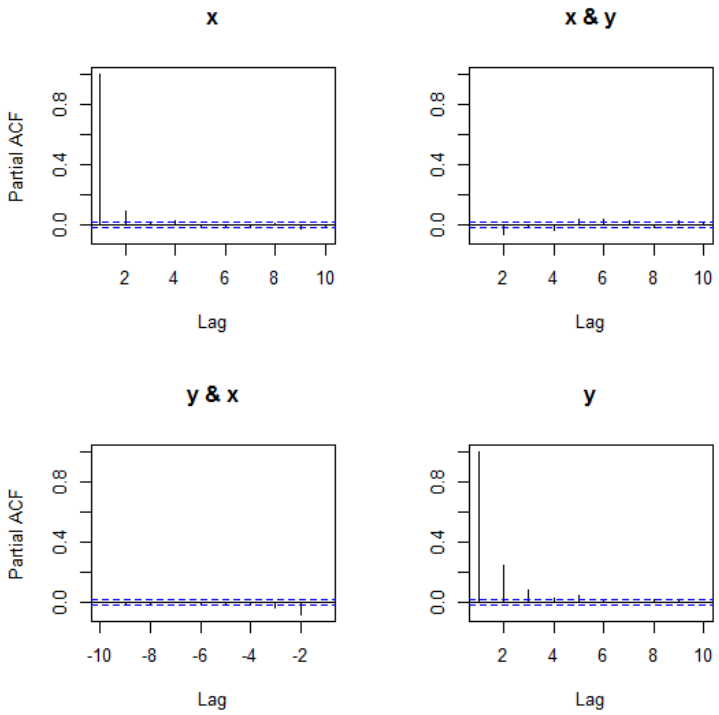


Figure 6.4: Partial correlation plots for Bioassay 3.

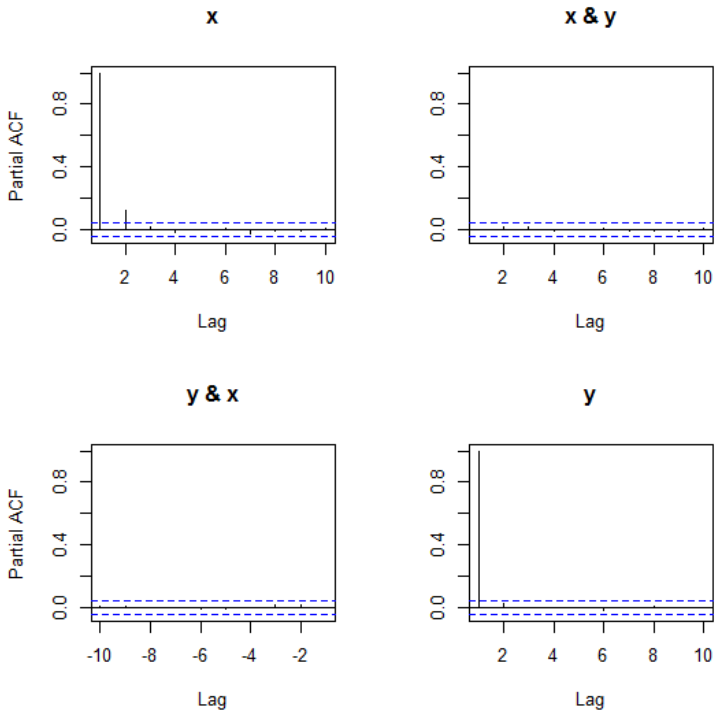


Figure 6.5: Partial correlation plots for Bioassay 4.

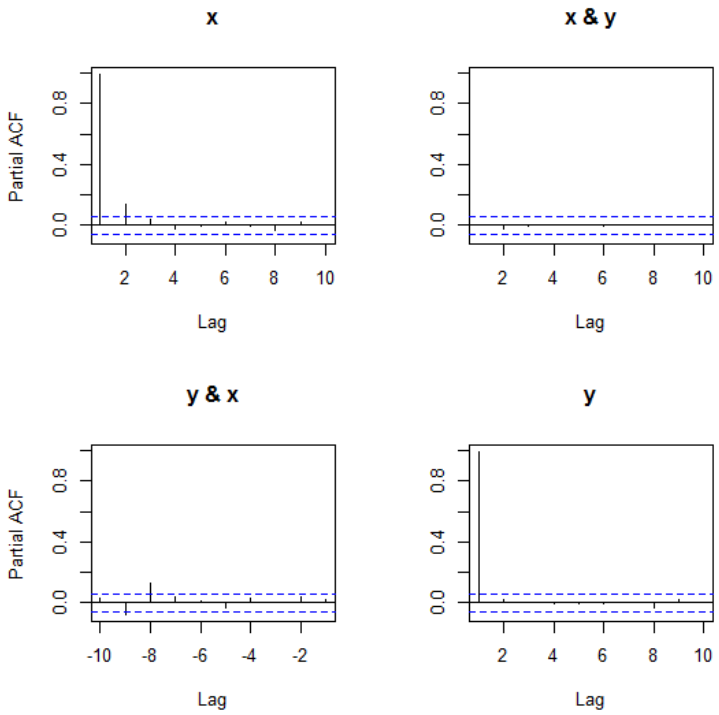
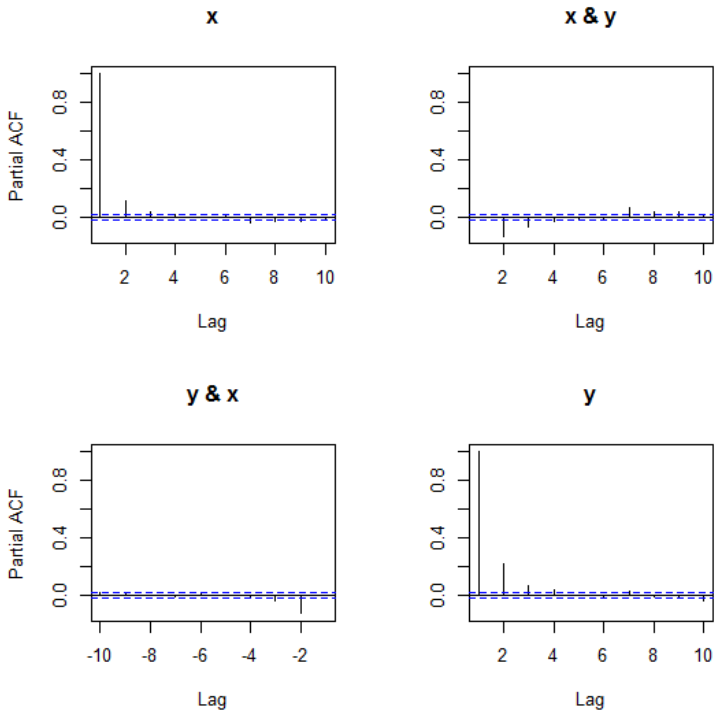


Figure 6.6: Partial correlation plots for Bioassay 5.



6.4.2 Higher order diffusion processes

We introduce a second-order diffusion process, under which the distribution of each observation $\mathbf{x}_i$ given its previous 2 positions $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i-2}$ is

$$\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{x}_{i-2} \sim \mathrm{N}\left\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \mathbf{a}) + \boldsymbol{\Delta}(\mathbf{x}_{i-2} - \mathbf{a}), \boldsymbol{\Phi}\right\}, \tag{6.11}$$

where

$$\boldsymbol{\Delta} = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}.$$

This model is similar to the diffusion process (5.3) introduced in Section 5.2. Summary statistics for the posterior distributions of the model parameters for diffusion process (6.11) for Bioassays 1–5 are presented in Tables 6.14 to 6.18. Results were obtained using WinBUGS and OpenBUGS. The prior distributions used for $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ were as given in (6.1) and (6.4). The prior distributions for the elements of $\boldsymbol{\Delta}$ were the same as those for the elements of $\boldsymbol{\Gamma}$.

For Bioassay 1, as shown in Table 6.14, the posterior means of the elements of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$ indicate that dependence on the immediately previous observation is, as expected, stronger than dependence on the observation prior to it. For Bioassays 2 and 3 (Tables 6.15 and 6.16 respectively), the ratios of the posterior means of the elements of $\boldsymbol{\Gamma}$ to those of the elements of $\boldsymbol{\Delta}$ are similar to that of Bioassay 1. However, for Bioassay 4 (Table 6.17), the posterior mean of δ_{11} is larger than that of γ_{11}. The posterior mean of δ_{22} is smaller than that of γ_{22}, but larger in comparison to it than is the case for Bioassays 1–3. These results suggest that the higher-order dependence is stronger for Bioassay 4 than the other bioassays. Conversely, the posterior means of δ_{11} and δ_{22} obtained for Bioassay 5 (Table 6.18) are smaller in comparison to those of γ_{11} and γ_{22} than for the other bioassays, indicating less pronounced higher-order dependence. Upon examination of Figure 5.3, directional persistence indeed appears strongest in Bioassay 4. The track for Bioassay 5 consists largely of movement around the origin. Although the movements in the latter part of the bioassay do show strong directional persistence, it appears that the movements around the origin have contributed more to the posterior distributions of the model parameters.

The sums of the posterior means of $\gamma_{11} + \delta_{11}$ and $\gamma_{22} + \delta_{22}$ for Bioassays 1–3 are less than unity, implying attraction towards $\mathbf{a}$. The corresponding values for Bioassay 4 indicate repulsion from $\mathbf{a}$, and those for Bioassay 5 suggest attraction to $\mathbf{a}$ along the y-axis and repulsion along the x-axis. These conclusions all agree

with those for the first-order diffusion process.

We also observe that for all bioassays, the posterior means of ϕ_{11} and ϕ_{22} are smaller than those for the first-order diffusion process (Table 5.10). The second-order diffusion processes fitted to the bioassay data therefore all consist of slightly smaller movements than the corresponding first-order diffusion processes.

Table 6.14: Summary statistics of posterior distributions of the parameters of the second-order diffusion process (6.11) for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6088	0.6094	0.0021	0.5980	0.6168
γ_{12}	0.0117	0.0107	0.0024	0.0090	0.0171
γ_{21}	0.0001	0.0000	0.0053	−0.0076	0.0118
γ_{22}	0.6349	0.6345	0.0021	0.6319	0.6402
δ_{11}	0.3909	0.3903	0.0048	0.3828	0.4018
δ_{12}	−0.0118	−0.0108	0.0024	−0.0172	−0.0091
δ_{21}	−0.0011	−0.0001	0.0053	−0.0123	0.0072
δ_{22}	0.3647	0.3652	0.0021	0.3595	0.3678
$\phi_{11}^{\dagger}$	3.5550	3.5550	0.0538	3.4530	3.6620
$\phi_{12}^{\dagger}$	0.0391	0.0392	0.0387	−0.0370	0.1158
$\phi_{22}^{\dagger}$	3.7930	3.7930	0.0568	3.6840	3.9050

† Values multiplied by 10^5.

We also consider the addition of third-order dependence, resulting in the diffusion process

$$\mathbf{x}_i|\mathbf{x}_{i-1},\mathbf{x}_{i-2},\mathbf{x}_{i-3} \sim \mathrm{N}\left\{\mathbf{a}+\mathbf{\Gamma}(\mathbf{x}_{i-1}-\mathbf{a})+\mathbf{\Delta}(\mathbf{x}_{i-2}-\mathbf{a})+\mathbf{\Upsilon}(\mathbf{x}_{i-3}-\mathbf{a}),\mathbf{\Phi}\right\}, \tag{6.12}$$

where

$$\mathbf{\Upsilon} = \begin{pmatrix} \upsilon_{11} & \upsilon_{12} \\ \upsilon_{21} & \upsilon_{22} \end{pmatrix}.$$

Again, summary statistics of the posterior distributions of the parameters obtained for the bioassays are displayed (Tables 6.19 to 6.23). The prior distributions used for $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ were as shown in (6.1) and (6.4). The prior distributions for the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ were the same as those for the elements of $\mathbf{\Gamma}$.

Table 6.15: Summary statistics of posterior distributions of the parameters of the second-order diffusion process (6.11) for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6368	0.6371	0.0086	0.6205	0.6501
γ_{12}	0.1065	0.1066	0.0008	0.1045	0.1080
γ_{21}	−0.0080	−0.0096	0.0044	−0.0139	0.0000
γ_{22}	0.6126	0.6131	0.0019	0.6098	0.6164
δ_{11}	0.3632	0.3629	0.0086	0.3499	0.3794
δ_{12}	−0.1066	−0.1067	0.0008	−0.1082	−0.1047
δ_{21}	0.0080	0.0095	0.0044	0.0000	0.0139
δ_{22}	0.3873	0.3868	0.0019	0.3835	0.3901
$\phi_{11}^{\dagger}$	3.9140	3.9130	0.0594	3.8000	4.0310
$\phi_{12}^{\dagger}$	0.1234	0.1233	0.0400	0.0438	0.2015
$\phi_{22}^{\dagger}$	3.6680	3.6680	0.0550	3.5620	3.7770

† Values multiplied by 10^5.

For Bioassay 1 (Table 6.19), the posterior mean of γ_{11} is smaller than that obtained using the second-order diffusion process, while the posterior mean of γ_{22} is larger. The posterior mean of δ_{11} is smaller than that of γ_{11}, but larger than that of υ_{11}, indicating that the dependence of each observation on a previous observation decreases with increasing time separating the two. This appears reasonable and agrees with the partial correlation plots shown in Figure 6.2. However, the posterior mean of υ_{22} is larger than that of δ_{22}, which is extremely small. This implies that in terms of movement along the y-axis, an observation $\mathbf{x}_i$ is more strongly dependent on $\mathbf{x}_{i-3}$ than $\mathbf{x}_{i-2}$.

Similar behaviour can be observed in the results for Bioassays 2 and 3 (Tables 6.20 and 6.21 respectively). Again, the posterior mean of γ_{11} is larger than that of δ_{11}, which in turn is larger than that of υ_{11}, but the posterior mean of υ_{22} is larger than that of δ_{22}.

For Bioassay 4 (Table 6.22), the posterior mean of υ_{22} is again larger than that of δ_{22}. Additionally, the posterior mean of δ_{11} is larger than that of γ_{11}, further indication that the estimated dependence on previous observations does not decrease with increasing time separating them from the dependent observation.

The results for Bioassay 5 (Table 6.23) differ from the other bioassays in that the posterior means of δ_{11} and δ_{22} are negative. This suggests that the location of an observation $\mathbf{x}_i$ is correlated positively with the location of $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i-3}$, but

Table 6.16: Summary statistics of posterior distributions of the parameters of the second-order diffusion process (6.11) for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\boldsymbol{\Delta}$ are the same as those for the elements of $\boldsymbol{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6010	0.5990	0.0105	0.5845	0.6209
γ_{12}	0.0131	0.0096	0.0088	0.0006	0.0306
γ_{21}	−0.0084	−0.0134	0.0141	−0.0266	0.0163
γ_{22}	0.6062	0.6070	0.0058	0.5960	0.6154
δ_{11}	0.3981	0.4000	0.0105	0.3782	0.4147
δ_{12}	−0.0134	−0.0100	0.0087	−0.0309	−0.0010
δ_{21}	0.0084	0.0135	0.0014	−0.0165	0.0266
δ_{22}	0.3930	0.3922	0.0058	0.3839	0.4032
$\phi_{11}^{\dagger}$	6.4170	6.4140	0.2127	6.0130	6.8410
$\phi_{12}^{\dagger}$	0.1931	0.1924	0.1220	−0.0495	0.4328
$\phi_{22}^{\dagger}$	4.2310	4.2290	0.1392	3.9670	4.5100

† Values multiplied by 10^5.

negatively with the location of $\mathbf{x}_{i-2}$. This does not appear physically plausible, and suggests that the fitted model may not be a good representation of the data.

We observe that the sum of posterior means of $\gamma_{22} + \delta_{22} + \upsilon_{22}$ for Bioassay 1 is greater than unity, suggesting repulsion from $\mathbf{a}$. This contradicts the results of the first-order and second-order diffusion models, and is in conflict with the behaviour observed in Figure 5.3. For Bioassays 2–4, the corresponding sums of posterior means lead to the same conclusions as for the first-order and second-order models.

BIC values for the first-order, second-order and third-order models are presented in Table 6.24. The second-order model is an improvement over the first-order model for all 5 bioassays. However, the third-order model improves on the second-order model only for Bioassays 3 and 4, performing worse for Bioassays 1, 2 and 5. Furthermore, the BIC values of the fitted second-order diffusion processes are still greatly inferior to those of the fitted mixture models and HMMs used previously, as shown in Table 6.13. The BIC values therefore suggest that the introduction of higher order dependence is less effective for modelling the data than the introduction of additional states explored in Sections 5.4 and 6.2.

Table 6.17: Summary statistics of posterior distributions of the parameters of the second-order diffusion process (6.11) for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\boldsymbol{\Delta}$ are the same as those for the elements of $\boldsymbol{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.4543	0.4502	0.0109	0.4417	0.4785
γ_{12}	−0.0684	−0.0694	0.0031	−0.0719	−0.0612
γ_{21}	−0.1609	−0.1597	0.0205	−0.1930	−0.1211
γ_{22}	0.5459	0.5464	0.0018	0.5402	0.5476
δ_{11}	0.5467	0.5509	0.0108	0.5223	0.5583
δ_{12}	0.0683	0.0693	0.0031	0.0613	0.0720
δ_{21}	0.1629	0.1618	0.0021	0.1228	0.1956
δ_{22}	0.4544	0.4539	0.0018	0.4527	0.4606
$\phi_{11}^{\dagger}$	3.1660	3.1650	0.1330	2.9180	3.4330
$\phi_{12}^{\dagger}$	0.4805	0.4779	0.1131	0.2602	0.7069
$\phi_{22}^{\dagger}$	4.4910	4.4880	0.1883	4.1350	4.8630

† Values multiplied by 10^5.

6.4.3 Mixture model with higher order dependence

We have observed that the second-order diffusion process (6.11) is an improvement over its first-order counterpart. However, model selection using BIC indicates that the 2 component mixture model introduced in Section 5.4 provides a much better fit to the data. We likewise modify the mixture model to include second-order dependence, giving a mixture of 2 components

(i) a component related to second-order diffusion process (6.11)

$$\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{x}_{i-2} \sim \mathrm{N}\{\mathbf{a} + \boldsymbol{\Gamma}(\mathbf{x}_{i-1} - \mathbf{a}) + \boldsymbol{\Delta}(\mathbf{x}_{i-2} - \mathbf{a}), \boldsymbol{\Phi}\},$$

and

(ii) a component consisting of localized movements,

$$\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{x}_{i-2} = \mathbf{x}_i|\mathbf{x}_{i-1} \sim \mathrm{N}(\mathbf{x}_{i-1}, \boldsymbol{\Sigma}),$$

with a parameter $0 < \pi < 1$ equal to the probability of an observation being generated from component (i).

Again, summary statistics of the posterior distributions of the parameters obtained for the bioassays are displayed (Tables 6.25 to 6.29). The prior

Table 6.18: Summary statistics of posterior distributions of the parameters of the second-order diffusion process (6.11) for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.7134	0.7132	0.0040	0.7064	0.7194
γ_{12}	−0.0251	−0.0252	0.0014	−0.0272	−0.0224
γ_{21}	−0.0162	−0.0174	0.0048	−0.0227	−0.0053
γ_{22}	0.9034	0.9045	0.0036	0.8980	0.9107
δ_{11}	0.2871	0.2873	0.0039	0.2811	0.2942
δ_{12}	0.0248	0.0250	0.0014	0.0221	0.0270
δ_{21}	0.0170	0.0181	0.0048	0.0060	0.0235
δ_{22}	0.0964	0.0954	0.0036	0.0893	0.1019
$\phi_{11}^{\dagger}$	4.9490	4.8489	0.0749	4.8070	5.0980
$\phi_{12}^{\dagger}$	0.2385	0.2388	0.0514	0.1371	0.3406
$\phi_{22}^{\dagger}$	4.7810	4.7810	0.0720	4.6410	4.9230

† Values multiplied by 10^5.

distributions used for $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ were as shown in (6.1), (6.4) and (6.5), the prior distribution for π was as shown in (5.8), and the prior distributions for the elements of $\mathbf{\Delta}$ were the same as those for the elements of $\mathbf{\Gamma}$.

The mean estimates of π are extremely close to 1 for all of the bioassays, indicating that the model has reduced to the single second-order diffusion process (6.11). Parameter estimates for the elements of $\mathbf{\Gamma}$, $\mathbf{\Delta}$, and $\mathbf{\Phi}$ are very similar to those shown in Tables 6.14 to 6.18. As such, the second-order mixture model is effectively equal to process (6.11), and does not provide a meaningful improvement.

BIC values for the second-order mixture models for the bioassays are shown in Table 6.30. Values for the second-order single diffusion processes used in Section 6.4.2 and the first-order mixture models used in Section 5.4 are also shown for comparison. As mentioned above, the second-order mixture model has reduced to a single diffusion process, and provides a worse fit to the data than the first-order mixture.

Table 6.19: Summary statistics of posterior distributions of the parameters of the third-order diffusion process (6.12) for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.5072	0.5073	0.0046	0.4982	0.5162
γ_{12}	0.0354	0.0351	0.0018	0.0323	0.0390
γ_{21}	0.0120	0.0112	0.0054	0.0024	0.0226
γ_{22}	0.7637	0.7658	0.0085	0.7482	0.7758
δ_{11}	0.2772	0.2809	0.0092	0.2605	0.2887
δ_{12}	−0.0756	−0.0742	0.0043	−0.0825	−0.0705
δ_{21}	−0.0597	−0.0601	0.0041	−0.0663	−0.0506
δ_{22}	0.0109	0.0116	0.0190	−0.0226	0.0367
υ_{11}	0.2153	0.2158	0.0093	0.2006	0.2282
υ_{12}	0.0401	0.0401	0.0043	0.0322	0.0466
υ_{21}	0.0472	0.0493	0.0073	0.0281	0.0558
υ_{22}	0.2502	0.2231	0.0109	0.2118	0.2469
$\phi_{11}^{\dagger}$	3.4444	3.4440	0.0507	3.43460	3.5430
$\phi_{12}^{\dagger}$	0.1116	0.1112	0.0397	0.0315	0.1889
$\phi_{22}^{\dagger}$	3.9457	3.9460	0.0668	3.8190	4.0741

† Values multiplied by 10^5.

Table 6.20: Summary statistics of posterior distributions of the parameters of the third-order diffusion process (6.12) for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6864	0.6863	0.0014	0.6838	0.6895
γ_{12}	0.1542	0.1532	0.0038	0.1495	0.1624
γ_{21}	−0.0499	−0.0502	0.0018	−0.0531	−0.0454
γ_{22}	0.6290	0.6297	0.0031	0.6235	0.6333
δ_{11}	0.1742	0.1764	0.0072	0.1620	0.1841
δ_{12}	−0.1905	−0.1893	0.0043	−0.1977	−0.1851
δ_{21}	0.0118	0.0114	0.0019	0.0094	0.0166
δ_{22}	0.0941	0.0943	0.0286	0.0885	0.0982
υ_{11}	0.1393	0.1388	0.0066	0.1297	0.1491
υ_{12}	0.0361	0.0361	0.0011	0.0342	0.0379
υ_{21}	0.0380	0.0383	0.0018	0.0335	0.0407
υ_{22}	0.2767	0.2760	0.0059	0.2695	0.2882
$\phi_{11}^{\dagger}$	3.9390	3.9380	0.0580	3.8270	4.0350
$\phi_{12}^{\dagger}$	0.1709	0.1710	0.0407	0.0898	0.2503
$\phi_{22}^{\dagger}$	3.6961	3.6950	0.0554	3.5910	3.8060

† Values multiplied by 10^5.

Table 6.21: Summary statistics of posterior distributions of the parameters of the third-order diffusion process (6.12) for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.5454	0.5453	0.0169	0.5141	0.5717
γ_{12}	0.0077	0.0066	0.0053	−0.0008	0.0224
γ_{21}	−0.0300	−0.0310	0.0134	−0.0501	−0.0022
γ_{22}	0.6390	0.6392	0.0027	0.6333	0.6439
δ_{11}	0.3064	0.3095	0.0115	0.2819	0.3236
δ_{12}	−0.0816	−0.0805	0.0052	−0.0929	−0.0739
δ_{21}	−0.0264	−0.0279	0.0087	−0.0413	−0.0119
δ_{22}	0.1186	0.1177	0.0055	0.1096	0.1271
υ_{11}	0.1474	0.1491	0.0145	0.1134	0.1721
υ_{12}	0.0735	0.0738	0.0026	0.0674	0.0776
υ_{21}	0.0565	0.0550	0.0148	0.0304	0.0790
υ_{22}	0.2414	0.2424	0.0042	0.2341	0.2476
$\phi_{11}^{\dagger}$	6.2688	6.2540	0.2058	5.8730	6.6710
$\phi_{12}^{\dagger}$	0.1066	0.1055	0.1206	−0.1330	0.3388
$\phi_{22}^{\dagger}$	4.1604	4.1580	0.1364	3.9030	4.4380

† Values multiplied by 10^5.

Table 6.22: Summary statistics of posterior distributions of the parameters of the third-order diffusion process (6.12) for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.3768	0.3732	0.0176	0.3527	0.4168
γ_{12}	−0.0498	−0.0476	0.0044	−0.0578	−0.0444
γ_{21}	−0.2680	−0.2658	0.0182	−0.3024	−0.2438
γ_{22}	0.4325	0.4339	0.0040	0.4251	0.4389
δ_{11}	0.4761	0.4755	0.0169	0.4427	0.5004
δ_{12}	0.0498	0.0497	0.0050	0.0418	0.0576
δ_{21}	0.3176	0.3153	0.0163	0.2879	0.3428
δ_{22}	0.2179	0.2189	0.0051	0.2069	0.2236
υ_{11}	0.1488	0.1491	0.0044	0.1409	0.1574
υ_{12}	−0.0002	−0.0003	0.0014	−0.0028	0.0026
υ_{21}	−0.0463	−0.0454	0.0064	−0.0592	−0.0350
υ_{22}	0.3498	0.3495	0.0025	0.3462	0.3548
$\phi_{11}^{\dagger}$	3.0737	3.0710	0.1288	2.8360	3.3380
$\phi_{12}^{\dagger}$	0.6079	0.6085	0.1180	0.3737	0.8377
$\phi_{22}^{\dagger}$	4.6563	4.6560	0.1944	4.2890	5.0501

† Values multiplied by 10^5.

Table 6.23: Summary statistics of posterior distributions of the parameters of the third-order diffusion process (6.12) for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are as given in (6.1) and (6.4), and the prior distributions of the elements of $\boldsymbol{\Delta}$ and $\boldsymbol{\Upsilon}$ are the same as those for the elements of $\boldsymbol{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.8718	0.8698	0.0193	0.8453	0.9080
γ_{12}	−0.0402	−0.0402	0.0009	−0.0422	−0.0382
γ_{21}	0.0739	0.0070	0.0588	0.0588	0.0830
γ_{22}	1.1066	1.1070	0.0018	1.1030	1.1090
δ_{11}	−0.0351	−0.0293	0.0286	−0.0910	0.0036
δ_{12}	0.0218	0.0219	0.0011	0.0193	0.0234
δ_{21}	−0.0340	−0.0356	0.0148	−0.0560	−0.0062
δ_{22}	−0.7955	−0.7948	0.0150	−0.8219	−0.6189
Υ_{11}	0.1638	0.1603	0.0095	0.1513	0.1830
Υ_{12}	0.0182	0.0184	0.0012	0.0160	0.0204
Υ_{21}	−0.0388	−0.0375	0.0085	−0.0518	−0.0250
Υ_{22}	0.6888	0.6868	0.0163	0.6685	0.7196
$\phi_{11}^{\dagger}$	5.2788	5.2750	0.1054	5.0850	5.4960
$\phi_{12}^{\dagger}$	0.4119	0.4125	0.0780	0.2596	0.5632
$\phi_{22}^{\dagger}$	8.1226	8.1170	0.1697	7.8100	8.4530

† Values multiplied by 10^5.

Table 6.24: BIC values of competing models fitted to Bioassays 1–5. The models used are first-order diffusion process (5.3), second-order diffusion process (6.11) and third-order diffusion process (6.12).

Model	**Number of parameters**	**Bioassay 1**	**2**	**3**
First-order diffusion	7	−129412	−130879	−25378
Second-order diffusion	11	−132600	−132100	−25917
Third-order diffusion	15	−131261	−132001	−26067

Model	**Number of parameters**	**Bioassay 4**	**5**
First-order diffusion	7	−16675	−125996
Second-order diffusion	11	−16792	−127600
Third-order diffusion	15	−17039	−122317

Table 6.25: Summary statistics of posterior distributions of the parameters of the second-order mixture model defined in Section 6.4.3. for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and π are as given in (6.1), (6.4), (6.5) and (5.8), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6014	0.6011	0.0050	0.5937	0.6112
γ_{12}	0.0262	0.0257	0.0016	0.0254	0.0294
γ_{21}	0.0223	0.0235	0.0062	0.0110	0.0325
γ_{22}	0.6400	0.6403	0.0065	0.6300	0.6530
δ_{11}	0.3983	0.3986	0.0051	0.3885	0.4061
δ_{12}	−0.0263	−0.0258	0.0016	−0.0295	−0.0241
δ_{21}	−0.0227	−0.0240	0.0062	−0.0329	−0.0114
δ_{22}	0.3596	0.3594	0.0065	0.3466	0.3697
$\phi_{11}^{\dagger}$	3.5584	3.5580	0.0538	3.4530	3.6640
$\phi_{12}^{\dagger}$	0.0426	0.0426	0.0380	−0.0308	0.1177
$\phi_{22}^{\dagger}$	3.7940	3.7940	0.0570	3.6850	3.9070
$\sigma_{11}^{\dagger}$	0.0462	0.0000	2.4608	0.0000	0.0095
$\sigma_{12}^{\dagger}$	0.0173	0.0000	0.7196	−0.0013	0.0015
$\sigma_{22}^{\dagger}$	0.0359	0.0000	1.4931	0.0000	0.0084
π	1.0000	0.9999	0.0001	0.9997	1.0000

† Values multiplied by 10^5.

Table 6.26: Summary statistics of posterior distributions of the parameters of the second-order mixture model defined in Section 6.4.3. for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and π are as given in (6.1), (6.4), (6.5) and (5.8), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6333	0.6322	0.0077	0.6230	0.6450
γ_{12}	0.1158	0.1159	0.0006	0.1148	0.1170
γ_{21}	0.0190	0.0194	0.0025	0.0141	0.0229
γ_{22}	0.6174	0.6174	0.0013	0.6156	0.6196
δ_{11}	0.3667	0.3667	0.0077	0.3550	0.3770
δ_{12}	−0.1160	−0.1160	0.0006	−0.1171	−0.1149
δ_{21}	−0.0190	−0.0194	0.0025	−0.0229	−0.0141
δ_{22}	0.3825	0.3825	0.0011	0.3803	0.3843
$\phi_{11}^{\dagger}$	3.9276	3.9270	0.0596	3.8120	4.4042
$\phi_{12}^{\dagger}$	0.1125	0.1123	0.0392	0.0371	0.1908
$\phi_{22}^{\dagger}$	3.6647	3.6650	0.0549	3.5580	3.7740
$\sigma_{11}^{\dagger}$	0.0462	0.0000	2.4608	0.0000	0.0095
$\sigma_{12}^{\dagger}$	0.0173	0.0000	0.7196	−0.0013	0.0015
$\sigma_{22}^{\dagger}$	0.0359	0.0000	1.4931	0.0000	0.0084
π	1.0000	0.9999	0.0001	0.9997	1.0000

† Values multiplied by 10^5.

Table 6.27: Summary statistics of posterior distributions of the parameters of the second-order mixture model defined in Section 6.4.3. for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and π are as given in (6.1), (6.4), (6.5) and (5.8), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.6234	0.6256	0.0152	0.5965	0.6478
γ_{12}	-0.0164	-0.0162	0.0125	-0.0333	0.0004
γ_{21}	-0.0146	-0.0161	0.0099	-0.0304	0.0046
γ_{22}	0.6152	0.6161	0.0079	0.6030	0.6297
δ_{11}	0.3757	0.3733	0.0153	0.3511	0.4027
δ_{12}	0.0160	0.0158	0.0125	-0.0007	0.0329
δ_{21}	0.0146	0.0161	0.0099	-0.0046	0.03049
δ_{22}	0.3841	0.3831	0.0079	0.3695	0.3962
$\phi_{11}^{\dagger}$	6.2707	6.2640	0.2063	5.8840	6.6840
$\phi_{12}^{\dagger}$	0.2369	0.2376	0.1183	0.0033	0.4676
$\phi_{22}^{\dagger}$	4.2215	4.2170	0.1405	3.9560	4.4990
σ_{11}	0.0739	0.0164	0.7430	0.0036	0.3535
σ_{12}	-0.5901	-0.1307	5.9344	-2.8230	-2.6040
σ_{22}	4.7126	1.0435	47.3918	0.2080	22.5405
π	0.9991	0.9992	0.0007	0.9972	0.9999

† Values multiplied by 10^5.

Table 6.28: Summary statistics of posterior distributions of the parameters of the second-order mixture model defined in Section 6.4.3. for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and π are as given in (6.1), (6.4), (6.5) and (5.8), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.5149	0.5138	0.0143	0.4755	0.5363
γ_{12}	−0.0498	−0.0498	0.0011	−0.0521	−0.0476
γ_{21}	−0.1966	−0.1937	0.0103	−0.2165	−0.1819
γ_{22}	0.5451	0.5457	0.0020	0.5414	0.5476
δ_{11}	0.4864	0.4876	0.0146	0.4660	0.5157
δ_{12}	0.0496	0.0496	0.0012	0.0488	0.0520
δ_{21}	0.1991	0.1954	0.0106	0.1836	0.2193
δ_{22}	0.4551	0.4544	0.0020	0.4524	0.4589
$\phi_{11}^{\dagger}$	3.0936	3.0890	0.1323	2.8490	3.3560
$\phi_{12}^{\dagger}$	0.4617	0.4597	0.1022	0.2491	0.6856
$\phi_{22}^{\dagger}$	4.5314	4.5240	0.1836	4.1740	4.9320
$\sigma_{11}^{\dagger}$	0.0291	0.0000	1.2417	0.0000	0.0086
$\sigma_{12}^{\dagger}$	−0.0049	0.0000	0.2664	−0.0016	0.0017
$\sigma_{22}^{\dagger}$	0.0216	0.0000	0.7605	0.0000	0.0084
π	0.9996	0.9998	0.0006	0.9978	1.0000

† Values multiplied by 10^5.

Table 6.29: Summary statistics of posterior distributions of the parameters of the second-order mixture model defined in Section 6.4.3. for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of $\mathbf{\Gamma}$, $\mathbf{\Phi}$ and π are as given in (6.1), (6.4), (6.5) and (5.8), and the prior distributions of the elements of $\mathbf{\Delta}$ and $\mathbf{\Upsilon}$ are the same as those for the elements of $\mathbf{\Gamma}$.

Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.7037	0.7031	0.0070	0.6897	0.7159
γ_{12}	−0.0149	−0.0161	0.0028	−0.0181	−0.0010
γ_{21}	0.0101	0.0078	0.0068	0.0019	0.0220
γ_{22}	0.9356	0.9350	0.0054	0.9267	0.9433
δ_{11}	0.2968	0.2902	0.0070	0.2845	0.3108
δ_{12}	0.0146	0.0115	0.0028	0.0097	0.0179
δ_{21}	−0.0094	−0.0071	0.0068	−0.0213	−0.0012
δ_{22}	0.0643	0.0593	0.0054	0.0566	0.0732
$\phi_{11}^{\dagger}$	4.9485	4.8980	0.0740	4.8070	5.0980
$\phi_{12}^{\dagger}$	0.2584	0.2227	0.0528	0.1564	0.3598
$\phi_{22}^{\dagger}$	4.8607	4.8110	0.0739	4.7160	5.0030
$\sigma_{11}^{\dagger}$	0.0523	0.0000	1.7065	0.0000	0.0101
$\sigma_{12}^{\dagger}$	0.0155	0.0000	0.7598	−0.0014	0.0021
$\sigma_{22}^{\dagger}$	0.0375	0.0000	1.2087	0.0000	0.0102
π	1.0000	0.9999	0.0001	0.9997	1.0000

† Values multiplied by 10^5.

Table 6.30: BIC values of competing models fitted to Bioassays 1–5. The models used are first-order diffusion process (5.3), second-order diffusion process (6.11), the first-order mixture model defined in Section 5.4 and the second-order mixture defined in Section 6.4.3.

	Number of	**Bioassay**		
Model	**parameters**	**1**	**2**	**3**
First-order diffusion	7	−129412	−130879	−25378
Second-order diffusion	11	−132600	−132100	−25917
First-order mixture	11	−301536	−307090	−57859
Second-order mixture	15	−132563	−131963	−25946

	Number of	**Bioassay**	
Model	**parameters**	**4**	**5**
First-order diffusion	7	−16675	−125996
Second-order diffusion	11	−16792	−127600
First-order mixture	11	−30479	−288701
Second-order mixture	15	−16959	−127394

6.5 Discussion

We have shown that there are advantages to using an HMM approach in place of the mixture models implemented in Chapter 5. Due to their greater flexibility, the HMMs produced more biologically realistic descriptions of the bioassay data which resulted in improved BIC values for the models. The 2-state HMM captures the salient features of the larva movements and clearly differentiates between the attraction towards **a** present in Bioassays 1–3, the repulsion in Bioassay 4, and the lateral movement in Bioassay 5. The addition of a third "null" state, representing no movement, results in a far more appropriate model that accounts for the successive identical observations present in the data. Model comparison using BIC indicates that the 3-state model is greatly superior to the 2-state HMM and mixture. The 4-state model implemented in this chapter introduces attraction towards or repulsion from the control zone **b**, and results in further improvement in terms of BIC for 3 of the 5 bioassays. However, the BIC differences between the 3-state and 4-state models are much smaller than those between the 2-state and 3-state models. The 5-state model adds a second state of attraction to or repulsion from **a**, thus allowing for attraction or repulsion of varying strength, but is inferior to the 4-state model for all bioassays in terms of BIC. The 3-state HMM is also preferred to the 5-state model for 3 out of 5 bioassays.

We have considered alternative models including more pronounced directional persistence. Partial correlation plots of the bioassay data support the existence of higher than first-order dependence, and we have attempted to account for this using second-order and third-order models which are similar to the first-order diffusion process introduced in Section 5.2 except that dependence on the previous observations is included. The second-order model identifies attraction to and repulsion from **a** for the bioassays as the first-order model does. Furthermore, the results obtained appear to indicate the presence of second-order dependence for all bioassays, and improve on the first-order diffusion process in terms of BIC. The third-order model suggests repulsion from **a** for Bioassay 1, which contradicts the first-order and second-order models and does not adequately describe the observed larva track. For 3 out of 5 bioassays, the third-order model is inferior to the second-order model using BIC. The results indicate second-order dependence in the data. However, the improvement in BIC is much smaller than that attained by using mixtures in place of a single diffusion process.

Finally, we have attempted to combine higher-order dependence with multiple components by using a 2 component mixture model similar to the one described in Section 5.4, with second-order dependence introduced to component (i). The

resulting models fitted to the bioassays are extremely similar to the corresponding second-order diffusion processes, as the posterior means of π are very close to 1 which indicates that almost all observations are assigned to component (i). This model is less successful than the first-order mixture for all bioassays. While the bioassay data show evidence of higher-order dependence, incorporating this dependence into the diffusion models is clearly a less effective approach than the increased flexibility of the mixture and HMMs.

Chapter 7

Conclusions

7.1 Review

The main objectives of this thesis were

- To implement finite mixture models for home range movements that successfully model real data and provide a flexible parametric alternative to kernel density estimators.
- To investigate the application of measures of overlap to home range data for the purpose of estimating site fidelity.
- To investigate diffusion processes as an approach to home range analysis, for modelling animal tracks rather than providing an estimate of their overall space usage.
- To apply diffusion processes in a different context to the movements of larvae, and propose more complex models to account for particular features of the data such as small localized movements.

Here we present an overall discussion of the material presented in the previous chapters of the thesis, organized into two sections. The first relates to the work on kernel and mixture models, including the application of overlap measures to the resulting density estimates. The second relates to diffusion models for both home range data and larva movements. We conclude by outlining possible avenues for future research on the topics covered.

7.2 Mixture models

We have considered a number of different models of animal movement data. We initially investigated the use and effectiveness of kernel and mixture models for modelling home range data sets, and in subsequent chapters introduced mixtures of diffusion processes as a more complex and flexible alternative. We also covered the topic of fidelity in home range data, introducing several measures of overlap and comparing their effectiveness when applied to sample data sets.

In Chapter 2 we observed that the kernel methods routinely used for modelling home range data provided effective nonparametric estimates of home range area. Our comparison of fixed and adaptive kernels using ISE for model comparison in Section 2.4 indicated that fixed kernels were generally to be preferred. However, kernel density estimators merely provide a description of the surveyed animal's movements and do not describe the underlying utilization density generating these movements. For this reason we proposed the use of mixtures as an alternative to approximate the utilization density. We initially used mixtures of bivariate normal distributions, fitted using the `mclust` R package. While the normal mixture models produced reasonable density estimates when applied to simple simulated data sets, they were susceptible to the presence of outlying points in real home range data sets. The presence of outliers resulted in overfitting, with unnecessary additional mixture components corresponding to small clusters of observations. Outliers are common in home range data sets, suggesting that normal mixtures are not sufficiently robust models, and indeed such results were observed for the DC, WG1, WG2 and DG data sets defined throughout the thesis.

We then considered mixtures of bivariate t distributions as an alternative with more robustness to outliers, and designed the MIX-T-FIX and MIX-T-VAR procedures in R for fitting such models to home range data. Application to simulated data sets in Section 2.3 showed that the mixtures of bivariate t distributions were less strongly influenced by outlying points as desired, and the density estimates obtained for various simulated data sets contained the correct number of components in the great majority of cases. While the estimated degrees of freedom for the mixture components indicated that outliers were often assigned to different components than those which generated them, the density estimates were nonetheless close to the true simulation densities. In Section 2.5, the mixtures of t distributions were also fitted to the same real data sets as the mixture of normal distributions, and produced far more appropriate density estimates, again demonstrating reduced susceptibility to outliers in comparison with the normal mixtures.

While MIX-T-FIX and MIX-T-VAR generally produce good estimates of home range area, the density estimates for certain data sets, in particular WG1, still show unwanted influence by clusters of outlying points. This influence was overcome in Section 2.6 by refining the models to include prior distributions for the parameters of the mixture components. We used χ^2 priors with low degrees of freedom for the degrees of freedom of the components, in order to weight the estimated degrees of freedom towards lower values and thus increase robustness to outliers. The resulting procedure, MIX-T-BAYES, gave a greatly improved density estimate for the WG1 data set. We have found that a mixture of bivariate t distributions with appropriate priors for the model parameters is a robust and effective model of home range data.

We examined measures of fidelity for the mixture models in Chapter 3. We considered the scaled product measure, square root product measure and OVL, and compared their effectiveness by applying them to pairs of simulated data sets where the true values of overlap between the simulation densities for each measure were known. The OVL was found to be more biased than the other two measures when the level of overlap was very high or very low. Our application of these measures of overlap to the fitted kernel and mixture models that had been discussed in Chapter 2 corroborated our conclusion that mixtures of bivariate t distributions provide an appropriate model for an animal's utilization of space. In particular, we considered the WG1 and WG2 data sets, which include the movements of a coyote over 2 successive years. The measures of overlap applied to the fitted mixtures of t distributions for these data sets therefore provide a measure of the animal's fidelity to the same home range. We found that the measures of overlap for the models fitted to the data were not unusual when compared to the corresponding values for data sets simulated from the fitted models, indicating that the fitted models for the 2 data sets result in simulated data sets with overlap similar to the true data, and thus that the fitted models accurately describe the degree of fidelity demonstrated by the animal.

7.3 Diffusion models

The mixtures of bivariate t distributions implemented in this thesis are useful models of animal movement data for the purpose of estimating the utilization density of animal movements and the shape of animals' home ranges. As mentioned in Chapter 4, however, the density estimates produced by the mixture models do not take any autocorrelation in the data into account. As such, they

only provide an approximation of overall use of space, and do not describe the paths of movement followed by the animals. We have addressed the latter problem by modelling animal tracks using diffusion processes. Initially, we considered a BOU process. The results produced in Chapter 4 upon application to a sample data set of coyote tracks suggested that the model is a useful representation of animal movements.

However, the BOU process requires certain assumptions that may not be sufficiently flexible for real data analysis. In particular, the link between the centralization and covariance matrices at each step requires the existence of an equilibrium distribution as the time step between observations tends to infinity. Such an equilibrium distribution will not exist in all cases, and clearly does not exist for the cabbage root fly larva tracks considered in Chapter 5. When modelling the bioassay data, we therefore relaxed this constraint on the diffusion process, assuming a more flexible model without a link between the centralization and covariance matrices. In Section 5.2, we showed that the resulting process, when assuming a fixed point of attraction or repulsion determined by the position of the chemical compound in each bioassay, gives a satisfactory representation of the data. With our simulation study in Section 5.3, we found that the fitted models for data sets simulated from the model fitted to the bioassay had parameter estimates very close to those of the original fitted model. Furthermore, plots of real and simulated data sets indicated that the simulations correctly replicated the overall direction of movement of the real data sets in all cases.

However, there were indications that further improvements could be made to the diffusion process. Data sets simulated from the models fitted to the bioassay data consisted of larger random movements than the real data sets, and did not include the common small localized movements. Comparison of ordered squared Mahalanobis distances for the bioassay data with the expected distances assuming a bivariate normal diffusion process showed that the observed distances differed strongly from the expected distances and divided into several subsets of similar values, whereas the observed values for data sets simulated from the fitted models were close to the expected values. The rounding present in the bioassay data was found to be responsible for some of these discrepancies, as observed squared Mahalanobis distances for the simulated data sets rounded to the same precision also belonged to several subsets of similar values. Nonetheless, the resulting distances still differed considerably from those of the bioassay data.

We accounted for the presence of small localized movements and larger movements in the bioassay data, and more generally for the possibility of multiple types of movement or behavioural states in animal movement data, by using a

mixture of multiple diffusion processes. In the case of the cabbage root fly larva movements, we considered a mixture of 2 processes, one consisting of movements attracted to or repelled by a point, and another comprised of undirected localized movements. Comparison of real and simulated data sets in Section 5.4 suggests that the mixture models fitted to the bioassay data are a more realistic description of the larvae's behaviour than the single diffusion process, and model comparison using BIC also indicates that the mixture is preferable. The parameter estimates for the fitted mixture models show that one of the 2 states contains small undirected movements, and the other contains larger movements attracted to or repelled by **a**, as was hypothesised for the data. The mixture model is successful in modelling the bioassay data, and provides a flexible model for movement data in general. It successfully distinguishes between the effects of attractant and repellent compounds when applied to the bioassays, and is therefore of use for determining strategies for the protection of crops from cabbage root fly larvae.

While the mixture of diffusion processes is successful in modelling the bioassay data, we have improved upon it further in Chapter 6 by using a similar HMM, with a probability transition matrix in place of a single weight parameter. We initially considered a 2-state HMM similar to the mixture model, and in Section 6.3 we compared the results to those obtained for a 3-state, 4-state and 5-state model, where the third state introduced consisted of no movement, the fourth of movement influenced by a control zone **b**, and the fifth of a second component of movement influenced by **a**. Application of the HMMs to the bioassay and comparison using BIC reveals that the 2-state HMM is preferable to the mixture model, and that the 3-state HMM is far superior to both. The 4-state HMM is slightly superior to the 3-state HMM in terms of BIC, as it has improved BIC values for 3 of the 5 bioassays. It is thus the preferred model of all those considered, while the 5-state HMM is inferior to the 4-state model for all bioassays and also to the 3-state model for the majority of them.

We have also explored the incorporation of higher-order dependence into the diffusion models as a potential alternative to the addition of extra states resulting in a mixture or HMM. In Section 6.4 we used a second-order diffusion process similar to the first-order model, except that the position at each time step is dependent on the positions at the 2 previous observations rather than just the immediately previous one, and a third-order process resembling the second-order process in the same way. Model comparison using BIC selects the second-order process over the first-order process. However, the mixture and HMMs were superior to the second-order model. Furthermore, the third-order model was inferior to the second-order model, suggesting that the inclusion of higher-order

dependence is less effective in improving the diffusion models than the inclusion of additional types of movement, even where higher-order dependence is clearly strong, as is the case for the bioassay data.

The conclusion that representing features of the movement data with more states in an HMM is more effective than allowing for higher-order dependence is corroborated by the results obtained for the second-order mixture models fitted to the bioassay data. For all bioassays, the second-order mixture essentially reduces to its first-order counterpart, as almost all observations are assigned to the same component of the model. The model thus provides no advantage over the single second-order diffusion process, and is not as useful as the HMMs previously considered. We therefore recommend the first-order 3-state and 4-state HMMs for use in modelling larva movements. These models are the most effective of those considered at modelling the responses of the larvae to different chemical compounds, and thus for identifying compounds which can be used as deterrents for the purpose of crop protection.

7.4 Further work

To conclude, we discuss further refinements of the models investigated in this thesis. While the mixtures of bivariate t distributions discussed in Chapters 2 and 3 provide an effective model for home range data, we observe that there are continued developments in mixture modelling using multivariate t or multivariate skew t distributions (Lee and McLachlan, 2011, 2013a, 2013b; Lin, 2010; Lin et al., 2013; Mengersen et al., 2011) which would be a possible alternative for home range data where substantial skewness is present. Software has been developed recently to fit mixtures of multivariate skew t distributions using the EM algorithm and could be used for this purpose (Lee and McLachlan, 2013c; Wang et al., 2009). More significantly, we note that the diffusion models used in Chapters 4–6 currently assume a bivariate normal distribution at each step, with a centralization matrix and covariance matrix as parameters. Both of these matrices are assumed to be constant. However, we could consider a diffusion model with time-dependent centralization and covariance matrices. The single diffusion process initially considered in Chapter 5, for example, could be altered to take the form

$$\mathbf{X}_{t+1}|\mathbf{X}_t \sim \mathrm{N}\left\{\mathbf{a} + \boldsymbol{\Gamma}(t+1)(\mathbf{X}_t - \mathbf{a}), \boldsymbol{\Phi}(t+1)\right\},$$

with the other states described for the mixture and Hidden Markov models altered similarly. This alteration would allow the inclusion of behavioural features not accounted for by the current models. For example, in the case of the bioassay data introduced in Chapter 5, we could consider the possibility of the larvae losing energy over time, and thus assume the covariance matrix $\mathbf{\Phi}(t)$ to be a decreasing function of overall time t since the beginning of the bioassay. We could instead consider the possibility of dependence on proximity to the point $\mathbf{a}$, giving a process of the form

$$\mathbf{X}_{t+1}|\mathbf{X}_t \sim \mathrm{N}\left\{\mathbf{a} + \mathbf{\Gamma}(t+1)(\mathbf{X}_t - \mathbf{a}), \mathbf{\Phi}(|\mathbf{X}_t - \mathbf{a}|)\right\}.$$

This model could assume, for instance, that larvae are more active when exposed to an attractant. The elements of the matrix $\mathbf{\Phi}$ could be defined as decreasing functions of distance from $\mathbf{a}$, or as step functions to represent attraction towards or repulsion from $\mathbf{a}$ when within a particular distance of it and undirected movement otherwise. Dependence on distance from $\mathbf{a}$ and on elapsed time t can both be included in the same model, giving the similar formulation

$$\mathbf{X}_{t+1}|\mathbf{X}_t \sim \mathrm{N}\left\{\mathbf{a} + \mathbf{\Gamma}(t+1)(\mathbf{X}_t - \mathbf{a}), \mathbf{\Phi}(|\mathbf{X}_t - \mathbf{a}|, t)\right\}.$$

Parameter estimates could be obtained using MCMC methods with a similar approach to the one that was employed for the diffusion models in Chapters 5 and 6. The diffusion models could be applied in different contexts. In this thesis, they have mostly been used to model the movements of larvae on a small scale. However, in Chapter 4 we more briefly considered a mixture of BOU processes for application to coyote home range data. The more complex diffusion models that we investigated in subsequent chapters could potentially be applied to home range movements on this scale, and we could compare the results to those of the simpler models as was done for the bioassay data.

Furthermore, the diffusion models could be compared with alternative models for animal tracks. One other approach, given a data set of 2-dimensional observations $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, is to consider the step lengths $\{r_1, \ldots, r_n\}$ and turning angles $\{\theta_1, \ldots, \theta_n\}$ of the n observations. The distance and angle of each observation can then be modelled as dependent on the distance and angle at the previous observation. For example, we could use the following model,

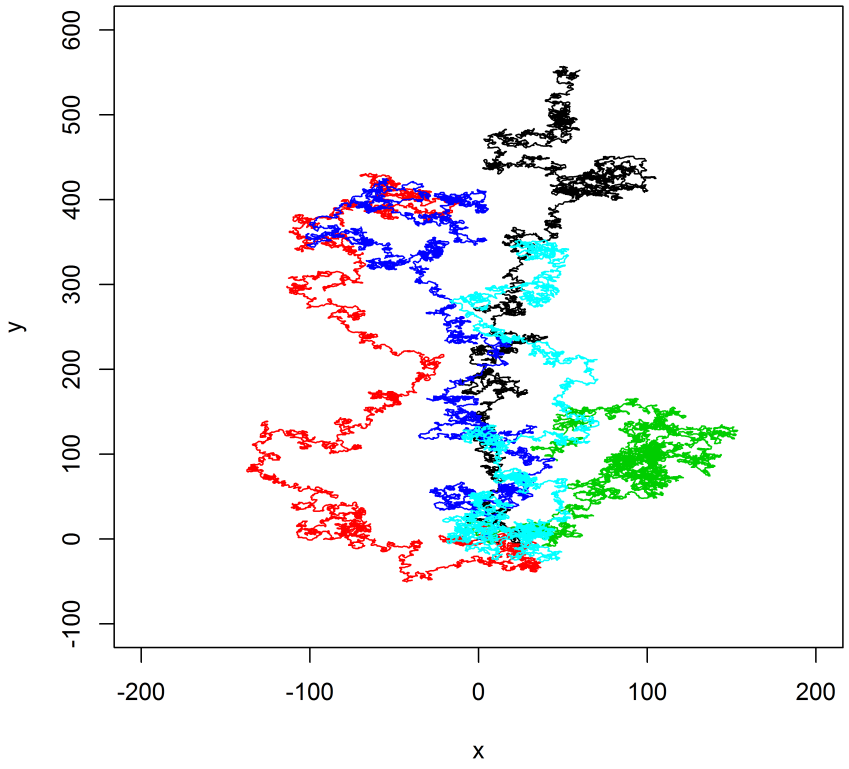
$$
\begin{aligned}
r_i|r_{i-1} &\sim \mathrm{N}_{[0,\infty]}(\gamma r_{i-1}, \sigma^2), \quad i = 2, \ldots, n, \\
\theta_i|\theta_{i-1} &\sim \text{Von Mises}(\mu + \delta\theta_{\mathbf{a}_{i-1}} + \epsilon\theta_{i-1}, k),
\end{aligned} \tag{7.1}
$$

where $\mathrm{N}_{[0,\infty]}$ is a truncated normal distribution taking only non-negative values and $\theta_{\mathbf{a}_{i-1}}$ is the angle from the point (r_{i-1}, θ_{i-1}) to a fixed point of attraction $\mathbf{a}$. The model therefore combines attraction towards $\mathbf{a}$ with dependence on the previous direction of movement, with the parameters δ and ϵ determining the respective strengths of these influences. The parameter μ is a mean turning angle in the absence of either influence, and the assumption $\mu = 0$ seems realistic. Other alterations to this model are possible, such as using a wrapped Cauchy distribution in place of the von Mises distribution for the turning angle. We have not currently attempted to fit models of this type to the bioassay data, though we have generated simulations using a variety of selected parameter values to compare the appearance of the simulated tracks with those of the larvae. For example, when the following parameters are used in the model described above,

$$
\begin{aligned}
k &= 1, \\
\delta &= 0.95, \\
\epsilon &= 0.05, \\
\mathbf{a} &= (200, 1000)^T,
\end{aligned} \tag{7.2}
$$

then the resulting simulated tracks are as shown in Figure 7.1. Further exploration of step length and turning angle models and comparison to the diffusion models would be a potential avenue for future research.

Figure 7.1: Plot of 5 data sets of 9000 observations each simulated from the model given in (7.1), with parameter values as shown in (7.2).



Appendix A

R code for mixture of bivariate t distributions

The R function `VarFit` given below fits the mixture of bivariate t distributions MIX-T-BAYES (defined on page 68) to a data set. It should be used repeatedly with a large number of different starting points to ensure convergence to the density estimate with global maximum likelihood.

```
delta<-function(y,mu,S){
(t(y-mu))%*%(ginv(S))%*%(y-mu)
}
#The function delta gives the Mahalanobis distance of a data point
#y from a mean mu given the covariance matrix S.

VarFit<-function(data,mu,S,w,v,r,p){
#data is the data set
#mu is the initial estimate of the means
#S is the initial estimate of the scale matrices
#w is the initial vector of weights
#v is the initial estimate of the degrees of freedom
#r is the degrees of freedom for the Chi-squared prior
#for the degrees of freedom
#p is the dimension of the data
n<-length(data[1,])
q<-length(w)
D<-1
niter<-0
#D is the tolerance variable . The algorithm stops once it
```

```
#decreases beyond a set value.

PMMean<-list(0)
for (i in 1:q){
PMMean[[i]]<-matrix(0,ncol=1,nrow=p)
for (j in 1:p){
PMMean[[i]][j,1]<-mean(data[j,])
}
}

PMVar<-list(0)
for (i in 1:q){
PMVar[[i]]<-matrix(0,ncol=p,nrow=p)
for (j in 1:p){
PMVar[[i]][j,j]<-((max(data[j,])-min(data[j,]))^2)
}
}
#Prior for means is normal
PVP<-list(0)
PVPar<-list(0)
for (i in 1:q){
PVP[[i]]<-matrix(0,ncol=p,nrow=p)
for (j in 1:p){
for (k in 1:p){
PVP[[i]][j,k]<-sum((data[j,]-mean(data[j,]))*
(data[k,]-mean(data[k,])))
}
}
PVPar[[i]]<-(1/(n-1))*PVP[[i]]
}
PVDeg<-p+2
#Prior for covariance matrices is Inverse Wishart
PDDeg<-r
#Prior for degrees of freedom is Chi-squared

#The parameters for the prior distributions

d<-matrix(0,ncol=n,nrow=q)
```

```
for (i in 1:q){
for (j in 1:n){
d[i,j]<-delta(data[,j],mu[,i],S[,(((i-1)*p+1):(i*p))])
}
}

f<-function(i,j){
((gamma((v[i]+p)/2))*((det(S[,(((i-1)*p+1):(i*p))]))^-0.5))
/(((pi*v[i])^(p/2))*gamma(v[i]/2)*(1+d[i,j]/v[i])^((v[i]+p)/2))
}

likcont<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
likcont[i,j]<-w[i]*f(i,j)
}
}

likmat<-matrix(0,ncol=n,nrow=1)
for (i in 1:n){
likmat[i]<-sum(likcont[,i])
}

loglikmat<-log(likmat)

loglik<-sum(loglikmat)
#Initial log-likelihood estimate.

logdif<-100
#This is the difference between the log-likelihood of
#the current iteration and the previous one.  It is
#initally set to an arbitrary number larger than the
#threshold for convergence.

while (D>0.001){
#this condition can be replaced with
#"while(logdif>0.001)" if it is desired to use
#log-likelihood as a measure of convergence.
```

```
d<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
d[i,j]<-delta(data[,j],mu[,i],S[,(((i-1)*p+1):(i*p))])
}
}
#This matrix contains the Malahanobis distances

u<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
u[i,j]<-(v[i]+p)/(v[i]+d[i,j])
}
}
#This matrix contains the conditional expectations of the
#"missing data" u which are needed to calculate
#estimates of mu and S

f<-function(i,j,v,S){
((gamma((v[i]+p)/2))*((det(S[,(((i-1)*p+1):(i*p))]))^-0.5))/
(((pi*v[i])^(p/2))*gamma(v[i]/2)*(1+d[i,j]/v[i])^((v[i]+p)/2))
}
#This function calculates the probability density of the j'th
#data entry according to the i'th t distribution

fm<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
fm[i,j]<-f(i,j,v,S)
}
}
#Matrix of probabilities

tau<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
ps<-sum(w*fm[,j])
```

```
tau[i,j]<-w[i]*fm[i,j]/ps
}
}
#This is the matrix of tau values - conditional expectations of
#the indicators for the various distributions, given the data.

for (i in 1:q){
w[i]<-(1/n)*sum(tau[i,])
}
#The posterior probabilities of each distribution, calculated
#using the EM algorithm.

nu<-mu
#The old matrix of means is kept for comparison
#with the new one.

for (i in 1:q){
muu<-matrix(0,ncol=1,nrow=p)
for (j in 1:p){
muu[j]<-sum(tau[i,]*u[i,]*data[j,])
}
SV<-S[,(2*i-1):(2*i)]
SV<-ginv(SV)
PMV<-ginv(PMVar[[i]])
A<-SV%*%muu+PMV%*%PMMean[[i]]
B<-(sum(tau[i,]*u[i,]))*SV+PMV
B<-ginv(B)
mu[,i]<-B%*%A
}
#The posterior means, calculated
#using the ECM algorithm.

T<-S
#The old covariance matrix is kept for comparison
#with the new one.

for (i in 1:q){
SS<-matrix(0,ncol=p,nrow=p)
```

```
for (k in 1:p){
for (l in 1:p){
SS[k,l]<-(sum(tau[i,]*u[i,]*(data[k,]-mu[k,i])*(data[l,]-mu[l,i])))
}
}
A<-SS+PVPar[[i]]
B<-(sum(tau[i,]))+PVDeg+p+1
S[,(p*(i-1)+1):(p*i)]<-A/B
}

#The posterior covariance matrix, calculated
#using the EM algorithm.

for (i in 1:q){
df<-function(vv){
-digamma(vv/2)+log(vv/2)+1+(1/sum(tau[i,]))
*sum(tau[i,]*(log(u[i,])-u[i,]))+digamma((v[i]+p)/2)
-log((v[i]+p)/2)+(2/sum(tau[i,]))
*((((PDDeg/2)-1)*(1/vv))-(1/2))
}
sol<-uniroot.all(df,c(1,1000))
#v[i]<-round(sol$root,digits=0)
v[i]<-sol[1]
}
#The degrees of freedom are updated here.

oldloglik<-loglik

d<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
d[i,j]<-delta(data[,j],mu[,i],S[,(p*i-1):(p*i)])
}
}

f<-function(i,j,v,S){
((gamma((v[i]+p)/2))*((det(S[,(((i-1)*p+1):(i*p))]))^-0.5))/
(((pi*v[i])^(p/2))*gamma(v[i]/2)*(1+d[i,j]/v[i])^((v[i]+p)/2))
```

```
}

likcont<-matrix(0,ncol=n,nrow=q)
for (i in 1:q){
for (j in 1:n){
likcont[i,j]<-w[i]*f(i,j,v,S)
}
}

likmat<-matrix(0,ncol=n,nrow=1)
for (i in 1:n){
likmat[i]<-sum(likcont[,i])
}

loglikmat<-log(likmat)

loglik<-sum(loglikmat)

logdif<-(loglik-oldloglik)

#Calculation of the log-likelihood (which requires updating
#the Malahanobis distances and the function f first)
niter<-niter+1

D<-sum(abs(mu-nu))+sum(abs(T-S))
#A sum of differences of absolute values between
#the new and old means, variances and covariances.
#The algorithm stops once this becomes small enough.
}
return(list(mu,S,w,v,loglik,niter))
}
```

Appendix B

BUGS code for diffusion process

The BUGS code given below fits a diffusion process with a point of attraction $\mathbf{a} = (a_1, a_2)^T$ to a data set $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

```
model{
for (i in 2:N){
x[i,1:2]~dmnorm(psi[i,1:2],P[,])
psi[i,1]<-a1+Gamma[1,1]*(x[(i-1),1]-a1)+Gamma[1,2]*(x[(i-1),2]-a2)
psi[i,2]<-a2+Gamma[2,1]*(x[(i-1),1]-a1)+Gamma[2,2]*(x[(i-1),2]-a2)
}

P[1:2,1:2]~dwish(Omega[,],2)

Phi[1:2,1:2]<-inverse(P[,])

Gamma[1,1]<-gamma11
Gamma[1,2]<-gamma12
Gamma[2,1]<-gamma21
Gamma[2,2]<-gamma22
DetGamma<-gamma11*gamma22-gamma12*gamma21

gamma11~dnorm(0,0.1)
gamma12~dnorm(0,0.1)
gamma21~dnorm(0,0.1)
gamma22~dnorm(0,0.1)
}
```

Appendix C

BUGS code for HMM with three states

The BUGS code given below fits an HMM with 3 states to a data set $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. The 3 states are a diffusion process with a point of attraction $\mathbf{a} = (a_1, a_2)^T$, a diffusion process with no point of attraction, and a "null" state under which an observation is identical to the previous one. HMMs with other numbers of states are similar.

```
model{
T[1]~dcat(R[1:3])
for (i in 2:N){
x[i,1:2]~dmnorm(psi[i,1:2],P[T[i],,])
psi[i,1]<-a1+Gammaf[T[i],1,1]*(x[(i-1),1]-a1)+Gammaf[T[i],1,2]
*(x[(i-1),2]-a2)
psi[i,2]<-a2+Gammaf[T[i],2,1]*(x[(i-1),1]-a1)+Gammaf[T[i],2,2]
*(x[(i-1),2]-a2)
T[i]~dcat(Q[T[i-1],1:3])
}

Q[1,1:3]~ddirch(prob1[])
Q[2,1:3]~ddirch(prob2[])
Q[3,1:3]~ddirch(prob3[])
Pi[1]<-Q[1,1]
Pi[2]<-Q[2,2]
Pi[3]<-Q[3,3]
R[1:3]~ddirch(prob4[])
```

```
P[1,1,1]<-P1[1,1]
P[1,1,2]<-P1[1,2]
P[1,2,1]<-P1[2,1]
P[1,2,2]<-P1[2,2]
P[2,1,1]<-P2[1,1]
P[2,1,2]<-P2[1,2]
P[2,2,1]<-P2[2,1]
P[2,2,2]<-P2[2,2]
P[3,1,1]<-1000000000*1000000000
P[3,1,2]<-0
P[3,2,1]<-0
P[3,2,2]<-1000000000*1000000000

P1[1:2,1:2]~dwish(Omega1[,],2)
P2[1:2,1:2]~dwish(Omega2[,],2)

Phi[1:2,1:2]<-inverse(P[1,,])
Sigma[1:2,1:2]<-inverse(P[2,,])

Gammaf[1,1,1]<-gamma11
Gammaf[1,1,2]<-gamma12
Gammaf[1,2,1]<-gamma21
Gammaf[1,2,2]<-gamma22
DetGamma<-gamma11*gamma22-gamma12*gamma21

Gamma[1,1]<-Gammaf[1,1,1]
Gamma[1,2]<-Gammaf[1,1,2]
Gamma[2,1]<-Gammaf[1,2,1]
Gamma[2,2]<-Gammaf[1,2,2]

gamma11~dnorm(0,0.1)
gamma12~dnorm(0,0.1)
gamma21~dnorm(0,0.1)
gamma22~dnorm(0,0.1)

Gammaf[2,1,1]<-1
Gammaf[2,1,2]<-0
Gammaf[2,2,1]<-0
```

```
Gammaf[2,2,2]<-1

Gammaf[3,1,1]<-1
Gammaf[3,1,2]<-0
Gammaf[3,2,1]<-0
Gammaf[3,2,2]<-1

phi11<-Phi[1,1]
phi12<-Phi[1,2]
phi21<-Phi[2,1]
phi22<-Phi[2,2]

sigma11<-Sigma[1,1]
sigma12<-Sigma[1,2]
sigma21<-Sigma[2,1]
sigma22<-Sigma[2,2]

pi11<-Q[1,1]
pi12<-Q[1,2]
pi13<-Q[1,3]
pi21<-Q[2,1]
pi22<-Q[2,2]
pi23<-Q[2,3]
pi31<-Q[3,1]
pi32<-Q[3,2]
pi33<-Q[3,3]
}
```

Appendix D

Summary statistics for 4-state HMM

Summary statistics for the HMM with states (i), (ii), (iii) and (iv) as defined in Section 6.3 fitted to Bioassays 1–5 are displayed in Tables D.1 to D.5.

Table D.1: Summary statistics of posterior distributions of the parameters of the 4-state HMM defined in Section 6.3.3 for Bioassay 1. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.2), (6.4), (6.5), (6.6) and (6.8).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9984	0.9984	0.0012	0.9961	1.0010
γ_{12}	−0.0008	−0.0008	0.0003	−0.0013	−0.0001
γ_{21}	0.0229	0.0229	0.0007	0.0215	0.0242
γ_{22}	0.9878	0.9878	0.0002	0.9875	0.9882
λ_{11}	0.9972	0.9973	0.0022	0.9920	1.0010
λ_{12}	−0.0002	−0.0002	0.0005	−0.0010	0.0001
λ_{21}	0.0045	0.0044	0.0015	0.0018	0.0078
λ_{22}	1.0041	1.0040	0.0005	1.0030	1.0050
$\phi_{11}^{\dagger}$	0.5518	0.5502	0.0311	0.4957	0.6149
$\phi_{12}^{\dagger}$	−0.0013	−0.0013	0.0129	−0.0260	0.0240
$\phi_{22}^{\dagger}$	0.1893	0.1892	0.0110	0.1680	0.2103
$\sigma_{11}^{\dagger}$	3.5039	3.4970	0.1612	3.1990	3.8450
$\sigma_{12}^{\ddagger}$	0.0106	0.0212	0.6272	−1.1710	1.2850
$\sigma_{22}^{\ddagger}$	0.1086	0.1085	0.0050	0.0990	0.1187
$\psi_{11}^{\dagger}$	0.4611	0.4598	0.0282	0.4094	0.5179
$\psi_{12}^{\dagger}$	0.0059	0.0062	0.0142	−0.0230	0.0328
$\psi_{22}^{\dagger}$	0.1536	0.1530	0.0093	0.1469	0.1737
π_{11}	0.0008	0.0004	0.0011	0.0000	0.0037
π_{12}	0.0675	0.0668	0.0102	0.0496	0.0900
π_{13}	0.5722	0.5720	0.0196	0.5330	0.6098
π_{14}	0.3595	0.3596	0.0189	0.3233	0.3948
π_{21}	0.0506	0.0503	0.0075	0.0374	0.0668
π_{22}	0.3725	0.3718	0.0158	0.3426	0.4943
π_{23}	0.5356	0.5362	0.0165	0.5040	0.5670
π_{24}	0.0413	0.0408	0.0067	0.0297	0.0555
π_{31}	0.0504	0.0504	0.0025	0.0455	0.0557
π_{32}	0.0740	0.0740	0.0031	0.0681	0.0798
π_{33}	0.8349	0.8352	0.0044	0.8264	0.8433
π_{34}	0.0408	0.0408	0.0024	0.0363	0.0455
π_{41}	0.4264	0.4265	0.0210	0.3864	0.4707
π_{42}	0.0503	0.0502	0.0093	0.0337	0.0699
π_{43}	0.5524	0.5037	0.0212	0.4820	0.5618
π_{44}	0.0009	0.0004	0.0013	0.0000	0.0045

† Values multiplied by 10^4.
‡ Values multiplied by 10^7.

Table D.2: Summary statistics of posterior distributions of the parameters of the 4-state HMM defined in Section 6.3.3 for Bioassay 2. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.2), (6.4), (6.5), (6.6) and (6.8).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	0.9990	0.9990	0.0008	0.9975	1.0010
γ_{12}	−0.0014	−0.0014	0.0005	−0.0024	−0.0005
γ_{21}	0.0035	0.0035	0.0001	0.0033	0.0038
γ_{22}	0.9880	0.9880	0.0001	0.9878	1.0040
λ_{11}	1.0015	1.0020	0.0012	0.9992	1.0040
λ_{12}	−0.0011	−0.0011	0.0004	−0.0019	−0.0004
λ_{21}	0.0020	0.0020	0.0001	0.0017	0.0023
λ_{22}	0.9902	0.9902	0.0003	0.9897	0.9908
$\phi_{11}^{\dagger}$	0.6952	0.6904	0.0477	0.6098	0.7916
$\phi_{12}^{\dagger}$	−0.0084	−0.0085	0.0060	−0.0199	0.0046
$\phi_{22}^{\dagger}$	0.0247	0.0246	0.0017	0.0216	0.0283
$\sigma_{11}^{\dagger}$	2.9251	2.9230	0.1219	2.6910	3.1632
$\sigma_{12}^{\dagger}$	0.0106	0.0108	0.0219	−0.0306	0.0538
$\sigma_{22}^{\dagger}$	0.1942	0.1942	0.0084	0.1787	0.2109
$\psi_{11}^{\dagger}$	0.5295	0.5281	0.0387	0.4588	0.6109
$\psi_{12}^{\dagger}$	0.3020	0.303 0	0.2534	−0.2175	0.8867
$\psi_{22}^{\dagger}$	0.0195	0.0195	0.0014	0.0169	0.0225
π_{11}	0.0121	0.0113	0.0052	0.0042	0.0240
π_{12}	0.0971	0.0961	0.0148	0.0702	0.1307
π_{13}	0.5677	0.5666	0.0237	0.5215	0.6143
π_{14}	0.3231	0.3238	0.0224	0.2780	0.3673
π_{21}	0.0434	0.0434	0.0059	0.0329	0.0558
π_{22}	0.3120	0.3122	0.0138	0.2847	0.6447
π_{23}	0.6177	0.6175	0.0144	0.5893	0.6447
π_{24}	0.0270	0.0266	0.0049	0.0181	0.0375
π_{31}	0.0383	0.0383	0.0022	0.0342	0.0427
π_{32}	0.0961	0.0960	0.0036	0.0893	0.1033
π_{33}	0.8331	0.8333	0.0044	0.8245	0.8415
π_{34}	0.0325	0.0325	0.0020	0.0285	0.0367
π_{41}	0.2838	0.2828	0.0227	0.2408	0.3289
π_{42}	0.0857	0.0854	0.0146	0.0582	0.1149
π_{43}	0.6219	0.6223	0.0246	0.5276	0.6670
π_{44}	0.0085	0.0077	0.0046	0.0021	0.0199

† Values multiplied by 10^4.

Table D.3: Summary statistics of posterior distributions of the parameters of the 4-state HMM defined in Section 6.3.3 for Bioassay 3. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.2), (6.4), (6.5), (6.6) and (6.8).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0000	1.0000	0.0001	0.9997	1.0000
γ_{12}	0.0000	0.0000	0.0000	0.0000	0.0001
γ_{21}	−0.0233	−0.0234	0.0031	−0.0290	−0.0165
γ_{22}	0.9789	0.9788	0.0010	0.9771	0.9810
λ_{11}	1.0000	1.0000	0.0002	0.9996	1.0000
λ_{12}	0.0000	0.0000	0.0000	−0.0001	0.0001
λ_{21}	−0.0068	−0.0068	0.0002	−0.0098	−0.0036
λ_{22}	0.9935	0.9935	0.0002	0.9931	0.9939
$\phi_{11}^{\ddagger}$	0.9494	0.9298	0.1384	0.7186	1.2510
$\phi_{12}^{\ddagger}$	0.1051	0.1027	1.8260	−3.6060	3.7150
$\phi_{22}^{\dagger}$	0.3972	0.3909	0.0626	0.2900	0.5395
$\sigma_{11}^{\dagger}$	3.7040	3.6850	0.2729	3.2330	4.2920
$\sigma_{12}^{\dagger}$	0.1272	0.2916	0.0847	−0.0308	0.2916
$\sigma_{22}^{\ddagger}$	0.7276	0.7259	0.0519	0.6318	0.8361
$\psi_{11}^{\ddagger}$	1.9810	1.9200	0.4153	1.3310	2.9640
$\psi_{12}^{\ddagger}$	0.0119	−0.0008	1.6910	−3.4200	3.2680
$\psi_{22}^{\dagger}$	0.0692	0.0715	0.0291	0.0234	0.1248
π_{11}	0.0128	0.0099	0.0113	0.0005	0.0420
π_{12}	0.1344	0.1314	0.0356	0.0746	0.2111
π_{13}	0.6543	0.6554	0.0448	0.5634	0.7393
π_{14}	0.1985	0.1962	0.0397	0.1266	0.2817
π_{21}	0.0302	0.0296	0.0091	0.0004	0.0498
π_{22}	0.4181	0.4168	0.0025	0.3686	0.4703
π_{23}	0.5138	0.5148	0.0251	0.4658	0.5613
π_{24}	0.0379	0.0368	0.0098	0.0219	0.0584
π_{31}	0.3667	0.3659	0.0630	0.2367	0.4893
π_{32}	0.1014	0.0980	0.0400	0.0358	0.1928
π_{33}	0.5236	0.5232	0.0650	0.3994	0.6508
π_{34}	0.0082	0.0039	0.0111	0.0000	0.0417
π_{41}	0.0594	0.0593	0.0066	0.0472	0.0732
π_{42}	0.1573	0.1573	0.0102	0.1382	0.1776
π_{43}	0.7688	0.7689	0.0120	0.7446	0.7928
π_{44}	0.0146	0.0145	0.0032	0.0086	0.0021

† Values multiplied by 10^4.
‡ Values multiplied by 10^7.

Table D.4: Summary statistics of posterior distributions of the parameters of the 4-state HMM defined in Section 6.3.3 for Bioassay 4. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.2), (6.4), (6.5), (6.6) and (6.8).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0000	1.0000	0.0000	1.0000	1.0000
γ_{12}	0.0000	0.0000	0.0000	0.0000	0.0000
γ_{21}	0.0028	0.0026	0.0126	−0.0222	0.0270
γ_{22}	1.0029	1.0030	0.0028	0.9970	1.0080
λ_{11}	1.1159	1.1160	0.0064	1.1030	1.1290
λ_{12}	0.0619	0.0712	0.0021	0.0335	0.0924
λ_{21}	−0.0128	−0.0141	0.0012	−0.0153	−0.0004
λ_{22}	0.9943	0.9943	0.0030	0.9877	0.9996
$\phi^{\ddagger}_{11}$	0.0006	0.0006	0.0001	0.0005	0.0008
$\phi^{\ddagger}_{12}$	−0.0367	−0.0169	0.9916	−2.1520	1.8260
$\phi^{\dagger}_{22}$	2.6102	2.5765	0.2905	2.1009	3.2275
$\sigma^{\dagger}_{11}$	3.0735	3.0345	0.5251	2.2239	4.2332
$\sigma^{\dagger}_{12}$	0.4596	0.4517	0.1509	0.4251	0.7920
$\sigma^{\ddagger}_{22}$	0.4810	0.4667	0.1012	0.3226	0.7197
$\psi^{\dagger}_{11}$	0.2846	0.2744	0.0971	0.1259	0.5070
$\psi^{\dagger}_{12}$	0.0560	0.0545	0.0482	−0.0132	0.0951
$\psi^{\dagger}_{22}$	0.9318	0.9035	0.2012	0.6107	1.4071
π_{11}	0.3312	0.3311	0.0356	0.2613	0.4018
π_{12}	0.0491	0.0467	0.0199	0.0168	0.0930
π_{13}	0.5313	0.5316	0.0375	0.4561	0.6052
π_{14}	0.0884	0.0864	0.0257	0.0431	0.1443
π_{21}	0.1075	0.1025	0.0397	0.0438	0.1944
π_{22}	0.0629	0.0529	0.0529	0.0002	0.1831
π_{23}	0.5516	0.5534	0.0692	0.4098	0.6811
π_{24}	0.2780	0.2717	0.0771	0.1385	0.4425
π_{31}	0.1181	0.1181	0.0109	0.0982	0.1399
π_{32}	0.0423	0.0420	0.0083	0.0273	0.0601
π_{33}	0.8107	0.8108	0.0135	0.7841	0.8357
π_{34}	0.0289	0.0283	0.0074	0.0160	0.0435
π_{41}	0.0231	0.0157	0.0246	0.0000	0.0887
π_{42}	0.3823	0.3794	0.0690	0.2483	0.5219
π_{43}	0.5865	0.5912	0.0692	0.4487	0.7166
π_{44}	0.0081	0.0040	0.0110	0.0000	0.0381

† Values multiplied by 10^4.
‡ Values multiplied by 10^9.

Table D.5: Summary statistics of posterior distributions of the parameters of the 4-state HMM defined in Section 6.3.3 for Bioassay 5. Presented statistics include the mean, median, standard deviation (SD), and 2.5% and 97.5% percentiles. The prior distributions of the parameters are as given in (6.1), (6.2), (6.4), (6.5), (6.6) and (6.8).

	Posterior summary statistics				
Parameter	**Mean**	**Median**	**SD**	**2.5%**	**97.5%**
γ_{11}	1.0000	1.0000	0.0004	0.9993	1.0010
γ_{12}	−0.0010	−0.0010	0.0002	−0.0013	−0.0006
γ_{21}	0.0026	0.0026	0.0002	0.0022	0.0030
γ_{22}	0.9904	0.9904	0.0001	0.9902	0.9906
λ_{11}	1.0000	1.0000	0.0005	0.9994	1.0010
λ_{12}	−0.0005	−0.0005	0.0002	−0.0009	−0.0002
λ_{21}	0.0010	0.0010	0.0003	0.0004	0.0016
λ_{22}	0.9905	0.9905	0.0001	0.9902	0.9907
$\phi_{11}^{\dagger}$	0.6541	0.6536	0.0360	0.5878	0.7299
$\phi_{12}^{\dagger}$	−0.0520	−0.0519	0.0156	−0.0820	−0.0212
$\phi_{22}^{\dagger}$	0.2380	0.2372	0.0134	0.2133	0.2652
$\sigma_{11}^{\dagger}$	3.4600	3.4490	0.1443	3.1960	3.7700
$\sigma_{12}^{\ddagger}$	0.0809	0.1042	4.6870	−9.1850	9.4650
$\sigma_{22}^{\ddagger}$	0.8257	0.8242	0.0356	0.7621	0.8950
$\psi_{11}^{\dagger}$	0.6922	0.6911	0.0397	0.6170	0.7738
$\psi_{12}^{\dagger}$	0.0228	0.0224	0.0171	−0.0091	0.0573
$\psi_{22}^{\dagger}$	0.2432	0.2429	0.0138	0.2178	0.2721
π_{11}	0.0007	0.0003	0.0011	0.0000	0.0368
π_{12}	0.0930	0.0927	0.0113	0.0719	0.1169
π_{13}	0.6121	0.6122	0.0184	0.5767	0.6491
π_{14}	0.2942	0.2935	0.0176	0.2600	0.3296
π_{21}	0.0596	0.0592	0.0070	0.0469	0.0739
π_{22}	0.2997	0.3001	0.0128	0.2756	0.3255
π_{23}	0.5769	0.5771	0.0143	0.5491	0.6037
π_{24}	0.0638	0.0634	0.0069	0.0514	0.0784
π_{31}	0.3060	0.3063	0.0188	0.2691	0.3450
π_{32}	0.1070	0.1062	0.0123	0.0842	0.1320
π_{33}	0.5845	0.5846	0.0202	0.5426	0.6242
π_{34}	0.0024	0.0019	0.0019	0.0002	0.0069
π_{41}	0.0643	0.0643	0.0030	0.0585	0.0706
π_{42}	0.1106	0.1106	0.0038	0.1034	0.1183
π_{43}	0.7762	0.7763	0.0051	0.7656	0.7860
π_{44}	0.0490	0.0490	0.0026	0.0436	0.0541

† Values multiplied by 10^4.
‡ Values multiplied by 10^8.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Czáki (eds), *Second International Symposium on Information Theory*, pp. 267–281. Budapest: Akadémiai Kiadó.

Altman, R.M. (2007). Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, **102**, 201–210.

Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* New York: Wiley.

Andrews, J.L. and McNicholas, P.D. (2011). Extending mixtures of multivariate t-factor analyzers. *Statistics and Computing*, **21**, 361–373.

Azzalini, A. and Bowman, A.W. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford: Oxford University Press.

Baur, R., Städler, E., Monde, K., and Takasugi, M. (1998). Phytoalexins from *Brassica* (Cruciferae) as oviposition stimulants for the cabbage root fly, *Delia radicum*. *Chemoecology*, **8**, 163–168.

Benhamou, S. (2007). How many animals really do the Lévy walk? *Ecology*, **88**, 1962–1969.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, **5**, 445–463.

Bergman, C.M., Schaefer, J.A., and Luttich, S.N. (2000). Caribou movement as a correlated random walk. *Oecologia*, **123**, 364–374.

Beyer, H.L., Haydon, D.T., Morales, J.M., Frair, J.L., Hebblewhite, M., Mitchell, M., and Matthiopoulos, J. (2010). The interpretation of habitat preference metrics under use-availability designs. *Philosophical Transactions of the Royal Society B*, **365**, 2245–2254.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of Calcutta Mathematical Society*, **35**, 99–109.

Blackwell, P.G. (1997). Random diffusion models for animal movement. *Ecological Modelling*, **100**, 87–102.

Blackwell, P.G. (2003). Bayesian inference for Markov processes with diffusion and discrete components. *Biometrika*, **90**, 613–627.

Blair, W.F. (1940). Notes on home ranges and populations of the short-tailed shrew. *Ecology*, **21**, 284–288.

Bovet, P. and Benhamou, S. (1988). Spatial analysis of animals' movements using a correlated random walk model. *Journal of Theoretical Biology*, **131**, 419–433.

Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Massachusetts: Addison-Wesley.

Boyce, M.S., Pitt, J., Northrup, J.M., Morehouse, A.T., Knopff, K.H., Cristescu, B., and Stenhouse, G.B. (2010). Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data. *Philosophical Transactions of the Royal Society B*, **365**, 2213–2219.

Brillinger, D.R., Preisler, H.K., Ager, A.A., and Kie, J.G. (2004). An exploratory data analysis (EDA) of the paths of moving animals. *Journal of Statistical Planning and Inference*, **122**, 43–63.

Burt, W.H. (1943). Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*, **24**, 346–352.

Byers, J.A. (2001). Correlated random walk equations of animal dispersal resolved by simulation. *Ecology*, **82**, 1680–1690.

Cagnacci, F., Boitani, L., Powell, R.A., and Boyce, M.S. (2010). Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B*, **365**, 2157–2162.

Carlin, B.P. and Louis, T.A. (2008). *Bayesian Methods for Data Analysis*, Third Edition. London: Chapman and Hall/CRC.

Casella, G. and George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167–174.

Celeux, G. and Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition*, **28**, 781–793.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, **75**, 79–97.

Clemons, T.E. and Bradley, E.L. (2000). A nonparametric measure of the overlapping coefficient. *Computational Statistics and Data Analysis*, **34**, 51–61.

Coelho, C.M., de Melo, L.F.B., Sábato, M.A.L., Rizel, D.N., and Young, R.J. (2007). A note on the use of GPS collars to monitor wild maned wolves *Chrysocyon brachyurus* (Illiger 1815) (Mammalia, Canidae). *Applied Animal Behaviour Science*, **105**, 259–264.

Cushman, S.A., Chase, M., and Griffin, C. (2005). Elephants in space and time. *Oikos*, **109**, 331–341.

Davison, A.C. (2003). *Statistical Models*. Cambridge: Cambridge University Press.

Day, N.E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, **56**, 463–474.

Deasy, W. (2011). Novel approaches for the management of cabbage root fly. *Horticultural Development Company (HDC) Technical Seminar: Protecting Your Field Veg Crop*, 30th June 2011, Stockbridge Technology Centre, North Yorkshire, UK.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Dillon, A. and Kelly, M.J. (2008). Ocelot home range, overlap and density: comparing radio telemetry with camera trapping. *Journal of Zoology*, **275**, 391–398.

Dixon, K.R. and Chapman, J.A. (1980). Harmonic mean measure of animal activity areas. *Ecology*, **61**, 1040–1044.

Don, B.A.C. and Rennolls, K. (1983). A home range model incorporating biological attraction points. *Journal of Animal Ecology*, **52**, 69–81.

Dunn, J.E. and Gipson, P.S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, **33**, 85–101.

Duong, T. (2007). ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, **21**, 1–16.

Ewan, A. (2011). How to put root fly larvae off their food. *HDC News Field Vegetables Review*, Annual Supplement, Second Edition, 11.

Fieberg, J. and Kochanny, C.O. (2005). Quantifying home-range overlap: the importance of the utilization distribution. *The Journal of Wildlife Management*, **69**, 1346–1359.

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M.S., and Frair, J.L. (2010). Correlation and studies of habitat selection: problem, red herring or opportunity? *Philosophical Transactions of the Royal Society B*, **365**, 2233–2244.

Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J., and Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transactions of the Royal Society B*, **365**, 2187–2200.

Fraley, C. and Raftery, A.E. (2003). Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *Journal of Classification*, **20**, 263–286.

Fraley, C. and Raftery, A.E. (2006). MCLUST version 3: an R package for normal mixture modeling and model-based clustering. Technical Report No. 504, *Department of Statistics, University of Washington, Seattle, USA*.

Fraley, C. and Raftery, A.E. (2007). Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification*, **24**, 155–181.

Franke, A., Caelli, T., and Hudson, R.J. (2004). Analysis of movements and behavior of caribou (*Rangifer tarandus*) using hidden Markov models. *Ecological Modelling*, **173**, 259–270.

Franke, A., Caelli, T., Kuzyk, G., and Hudson, R.J. (2006). Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov models. *Ecological Modelling*, **197**, 237–246.

Frère, C.H., Krützen, M., Mann, J., Watson-Capps, J.J., Tsai, Y.J., Patterson, E.M., Connor, R., Bejder, L., and Sherwin, W.B. (2010). Home range overlap, matrilineal and biparental kinship drive female associations in bottlenose dolphins. *Animal Behaviour*, **80**, 481–486.

Fritz, H., Said, S., and Weimerskirch, H. (2003). Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society B*, **270**, 1143–1148.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* New York: Springer.

Fryxell, J.M., Hazell, M., Börger, L., Dalzie, B.D., Haydon, D.T., Morales, J.M., McIntosh, T., and Rosatte, R.C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19114–19119.

Gaillard, J.M., Hebblewhite, M., Loison, A., Fuller, M., Powell, R., Basille, M., and van Moorter, B. (2010). Habitat-performance relationships: finding the right metric at a given scale. *Philosophical Transactions of the Royal Society B*, **365**, 2255–2265.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis*, Second Edition. London: Chapman and Hall/CRC.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall/CRC.

Girard, I., Ouellet, J.P., Courtois, R., Dussault, C., and Breton, L. (2002). Effects of sampling effort based on GPS telemetry on home-range size estimations. *The Journal of Wildlife Management*, **66**, 1290–1300.

Gitzen, R.A. and Millspaugh, J.J. (2003). Comparison of least-squares cross-validation bandwidth options for kernel home-range estimation. *Wildlife Society Bulletin*, **31**, 823–831.

Hall, P., Sheather, S.J., Jones, M.C., and Marron, J.S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, **78**, 263–269.

Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics*, **8**, 431–444.

Haydon, D.T., Morales, J.M., Yott, A., Jenkins, D.A., Rosatte, R., and Fryxell, J.M. (2008). Socially informed random walks: incorporating group dynamics into models of population spread and growth. *Proceedings of the Royal Society B*, **275**, 1101–1109.

Hebblewhite, M. (2009). Linking wildlife populations with ecosystem change: state-of-the-art satellite technology for national-park science. *Park Science*, **26**, 1–14.

Hebblewhite, M. and Haydon, D.T. (2010). Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B*, **365**, 2303–2312.

Hemson, G., Johnson, P., South, A., Kenward, R., Ripley, R., and MacDonald, D. (2005). Are kernels the mustard? Data from global positioning system (GPS) collars suggests problems for kernel home-range analyses with least-squares cross-validation. *Journal of Animal Ecology*, **74**, 455–463.

Hooge, P.N. and Eichenlaub, B. (1997). Animal movement. *Extension to ArcView*. Alaska Biological Science Center, US Geological Survey, Anchorage.

Inman, H.F. and Bradley, E.L. (1989). The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Communications in Statistics - Theory and Methods*, **18**, 3581–3874.

James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, **35**, 475–501.

James, M.C., Myers, R.A., and Ottensmeyer, C.A. (2005). Behaviour of leatherback sea turtles, *Dermochelys coriacea*, during the migratory cycle. *Proceedings of the Royal Society B*, **272**, 1547–1555.

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press.

Jennrich, R.I. and Turner, F.B. (1969). Measurement of non-circular home range. *Journal of Theoretical Biology*, **22**, 227–237.

Johnson, D.S., London, J.M., Lea, M.A., and Durban, J.W. (2008a). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, **89**, 1208–1215.

Johnson, D.S., Thomas, D.L., ver Hoef, J.M., and Christ, A. (2008b). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, **64**, 968–976.

Jonsen, I.D., Flemming, J.M., and Myers, R.A. (2005). Robust state-space modeling of animal movement data. *Ecology*, **86**, 2874–2880.

Juang, B.H. and Rabiner, L.R. (1991). Hidden Markov models for speech recognition. *Technometrics*, **33**, 251–272.

Kareiva, P.M. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia*, **56**, 234–238.

Kernohan, B.J., Gitzen, R.A., and Millspaugh, J.J. (2001). Analysis of animal space use and movements. In J.J. Millspaugh and J.M. Marzluff (eds), *Radio Tracking and Animal Populations*, pp. 125–166. San Diego: Academic Press.

Kie, J.G., Matthiopoulos, J., Fieberg, J., Powell, R.A., Cagnacci, F., Mitchell, M.S., Gaillard, J.M., and Moorcroft, P.R. (2010). The home-range concept: are traditional estimators still relevant with modern telemetry technology? *Philosophical Transactions of the Royal Society B*, **365**, 2221–2231.

Košťál, V. (1992). Orientation behavior of newly hatched larvae of the cabbage maggot, *Delia radicum* (L.) (Diptera: Anthomyiidae), to volatile plant metabolites. *Journal of Insect Behavior*, **5**, 61–70.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J.M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, **93**, 2336–2342.

Langrock, R. and Zucchini, W. (2011). Hidden Markov models with arbitrary state dwell-time distributions. *Computational Statistics and Data Analysis*, **55**, 715–724.

Laver, P.N. and Kelly, M.J. (2008). A critical review of home range studies. *The Journal of Wildlife Management*, **72**, 290–298.

Lee, S.X. and McLachlan, G.J. (2011). On the fitting of mixtures of multivariate skew t-distributions via the EM algorithm. *arXiv: 1109.4706*.

Lee, S. and McLachlan, G.J. (2013a). Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing*. To appear (DOI: 10.1007/s11222-012-9362-4).

Lee, S.X. and McLachlan, G.J. (2013b). On mixtures of skew normal and skew t-distributions. *Advances in Data Analysis and Classification*. To appear (DOI: 10.1007/s11634-013-0132-8).

Lee, S.X. and McLachlan, G.J. (2013c). EMMIX-uskew: an R package for fitting mixtures of multivariate skew t-distributions via the EM algorithm. *Journal of Statistical Software*. To appear.

Leonard, T. and Hsu, J.S.J. (1999). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers.* Cambridge: Cambridge University Press.

Lin, T.-I. (2010). Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing*, **20**, 343–356.

Lin, T.-I., Ho, H.J., and Lee, C.-R. (2013). Flexible mixture modelling using the multivariate skew-t-normal distribution. *Statistics and Computing*. To appear (DOI: 10.1007/s11222-013-9386-4).

Lin, T.-I., Lee, J.C., and Hsieh, W.J. (2007). Robust mixture modeling using the skew t distribution. *Statistics and Computing*, **17**, 81–92.

Lin, T.-I., Lee, J.C., and Ni, H.F. (2004). Bayesian analysis of mixture modelling using the multivariate t distribution. *Statistics and Computing*, **14**, 119–130.

Littaye, A., Gannier, A., Laran, S., and Wilson, J.P.F. (2004). The relationship between summer aggregation of fin whales and satellite-derived environmental conditions in the northwestern Mediterranean Sea. *Remote Sensing of Environment*, **90**, 44–52.

Lo, Y., Mendell, N.R., and Rubin, D.B. (2001). Testing the number of components in a normal mixture. *Biometrika*, **88**, 767–778.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

MacDonald, I.L. and Raubenheimer, D. (1995). Hidden Markov models and animal behaviour. *Biometrical Journal*, **37**, 701–712.

Marin, J.M. and Robert, C.P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics.* New York: Springer.

Matusita, K. (1955). Decision rules, based on the distance, for problems of fit, two samples, and estimation. *The Annals of Mathematical Statistics*, **26**, 631–640.

McGrory, C.A. and Titterington, D.M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics and Data Analysis*, **51**, 5352–5367.

McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* New York: Wiley.

McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering.* New York: Dekker.

McLachlan, G.J., Bean, R.W., and Jones, L.B.T. (2007). Extension of the mixture of factor analyzers to incorporate the multivariate t-distribution. *Computational Statistics and Data Analysis*, **51**, 5327–5338.

McLachlan, G.J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, Second Edition. Hoboken, New Jersey: Wiley.

McLachlan, G.J. and Peel, D. (2000). *Finite Mixture Models.* New York: Wiley.

McNicholas, P.D. and Murphy, T.B. (2008). Parsimonious Gaussian mixture models. *Statistics and Computing*, **18**, 285–296.

Mengersen, K.L., Robert, C.P., and Titterington, D.M. (2011). *Mixtures: Estimation and Applications*. Chichester: Wiley.

Merrill, E., Sand, H., Zimmermann, B., McPhee, H., Webb, N., Hebblewhite, M., Wabakken, P., and Frair, J.L. (2010). Building a mechanistic understanding of predation with GPS-based movement data. *Philosophical Transactions of the Royal Society B*, **365**, 2279–2288.

Millspaugh, J.J., Gitzen, R.A., Kernohan, B.J., Larson, M.A., and Clay, C.L. (2004). Comparability of three analytical techniques to assess joint space use. *Wildlife Society Bulletin*, **32**, 148–157.

Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., and Fryxell, J.M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85**, 2436–2445.

Morisita, M. (1959). Measuring interspecific association and similarity between communities. *Memoirs of the Faculty of Science of Kyushu University, Series E (Biology)*, **3**, 65–80.

Naef-Daenzer, B. (1993). A new transmitter for small animals and enhanced methods of home-range analysis. *The Journal of Wildlife Management*, **57**, 680–689.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadom, R., Saltz, D., and Smouse, P.E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 19052–19059.

Nations, C.S. and Anderson-Sprecher, R.C. (2006). Estimation of animal location from radio telemetry data with temporal dependencies. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 87–105.

Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association*, **64**, 953–963.

Noldus, L.P.J.J., Spink, A.J., and Tegelenbosch, R.A.J. (2001). EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments and Computers*, **33**, 398–414.

Odum, E.P. and Kuenzler, E.J. (1955). Measurement of territory and home range size in birds. *The Auk*, **72**, 128–137.

Okubo, A. (1980). *Diffusion and Ecological Problems: Mathematical Models*. Basel, Switzerland: Springer.

Park, B.U. and Marron, J.S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, **85**, 66–72.

Patterson, T.A., Basson, M., Bravington, M.V., and Gunn, J.S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, **78**, 1113–1123.

Peters, B.C. and Walker, H.F. (1977). An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions. *SIAM Journal on Applied Mathematics*, **35**, 362–378.

Powell, R.A. (2000). Animal home ranges and territories and home range estimators. In L. Boitani and T.K. Fuller (eds), *Research Techniques in Animal Ecology: Controversies and Consequences*, pp. 65–110. New York: Columbia University Press.

Preisler, H.K., Ager, A.A., Johnson, B.K., and Kie, J.G. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, **15**, 643–657.

Press, S.J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, Second Edition. Malabar, Florida: Krieger.

Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, **59**, 731–792.

Ridout, M.S. and Linkie, M. (2009). Estimating overlap of daily activity patterns from camera trap data. *Journal of Agricultural, Biological and Environmental Statistics*, **14**, 322–337.

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Second Edition. New York: Springer.

Robert, C.P., Celeux, G., and Diebolt, J. (1993). Bayesian estimation of hidden Markov chains: a stochastic implementation. *Statistics and Probability Letters*, **16**, 77–83.

Robert, C.P., Rydén, T., and Titterington, D.M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Series B*, **62**, 57–75.

Robertson, P.A., Aebischer, N.J., Kenward, R.E., Hanski, I.K., and Williams, N.P. (1998). Simulation and jack-knifing assessment of home-range indices based on underlying trajectories. *Journal of Applied Ecology*, **35**, 928–940.

Rodgers, A.R. (2001). Recent telemetry technology. In J.J. Millspaugh and J.M. Marzluff (eds), *Radio Tracking and Animal Populations*, pp. 79–121. San Diego: Academic Press.

Roessingh, P., Städler, E., Fenwick, G.R., Lewis, J.A., Nielsen, J.K., Hurter, J., and Ramp, T. (1992). Oviposition and tarsal chemoreceptors of the cabbage root fly are stimulated by glucosinolates and host plant extracts. *Entomologia Experimentalis et Applicata*, **65**, 267–282.

Ross, K.T.A. and Anderson, M. (1992). Larval responses of three vegetable root fly pests of the genus *Delia* (Diptera: Anthomyiidae) to plant volatiles. *Bulletin of Entomological Research*, **82**, 393–398.

Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society, Series B*, **73**, 689–710.

Schliehe-Diecks, S., Kappeler, P.M., and Langrock, R. (2012). On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus*, **2**, 180–189.

Schmid, F. and Schmidt, A. (2006). Nonparametric estimation of the coefficient of overlapping - theory and empirical application. *Computational Statistics and Data Analysis*, **50**, 1583–1596.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: Wiley.

Seaman, D.E., Millspaugh, J.J., Kernohan, B.J., Brundige, G.C., Raedeke, K.J., and Gitzen, R.A. (1999). Effects of sample size on kernel home range estimates. *The Journal of Wildlife Management*, **63**, 739–747.

Seaman, D.E. and Powell, R.A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, **77**, 2075–2085.

Seidel, K.S. (1992). *Statistical Properties and Applications of a New Measure of Joint Space Use for Wildlife*. M.Sc. Thesis, University of Washington, Seattle. USA.

Sheather, S.J. and Jones, M.C. (1991). A reliable data-based bandwidth selection methods for kernel density estimation. *Journal of the Royal Statistical Society, Series B*, **53**, 683–690.

Silverman, B.W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B*, **43**, 97–99.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall/CRC.

Smouse, P.E., Focardi, S., Moorcroft, P.R., Kie, J.G., Forester, J.D., and Morales, J.M. (2010). Stochastic modelling of animal movement. *Philosophical Transactions of the Royal Society B*, **365**, 2201–2211.

Soutullo, A., Cadahia, L., Urios, V., Ferrer, M., and Negro, J.J. (2007). Accuracy of lightweight satellite telemetry: a case study in the Iberian peninsula. *The Journal of Wildlife Management*, **71**, 1010–1015.

Städler, E. (1978). Chemoreception of host plant chemicals by ovipositing females of *Delia (Hylemya) brassicae*. *Entomologia Experimentalis et Applicata*, **24**, 711–720.

Swihart, R.K. and Slade, N.A. (1997). On testing for independence of animal movements. *Journal of Agricultural, Biological and Environmental Statistics*, **2**, 48–63.

Taylor, C.C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, **76**, 705–712.

Terrell, G.R. and Scott, D.W. (1992). Variable kernel density estimation. *The Annals of Statistics*, **20**, 1236–1265.

Tiao, G.C. and Zellner, A. (1964). On the Bayesian estimation of multivariate regression. *Journal of the Royal Statistical Society, Series B*, **26**, 277–285.

Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.

Tomkiewicz, S.M., Fuller, M.R., Kie, J.G., and Bates, K.K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B*, **365**, 2163–2176.

Tracy, D.S. and Dwyer, P.S. (1969). Multivariate maxima and minima with matrix derivatives. *Journal of the American Statistical Association*, **64**, 1576–1594.

Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sunderland, Massachusetts: Sinauer Associates.

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A., and Neteler, M. (2010). Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society B*, **365**, 2177–2185.

van Leur, H., Raaijmakers, C.E., and van Dam, N.M. (2008). Reciprocal interactions between the cabbage root fly (*Delia radicum*) and two glucosinolate phenotypes of *Barbarea vulgaris*. *Entomologia Experimentalis et Applicata*, **128**, 312–322.

Van Winkle, W. (1975). Comparison of several probabilistic home-range models. *The Journal of Wildlife Management*, **39**, 118–123.

Verwaijen, D. and Van Damme, R. (2008). Foraging mode and locomotor capabilities in Lacertidae. *Amphibia-Reptilia*, **29**, 197–206.

Vounatsou, P. and Smith, A.F.M. (1997). Simulation-based Bayesian inferences for two-variance components linear models. *Journal of Statistical Planning and Inference*, **59**, 139–161.

Wand, M.P. and Jones, M.C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, **9**, 97–116.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

Wang, H.X., Zhang, Q.B., Luo, B., and Wei, S. (2004). Robust mixture modelling using multivariate t-distribution with missing information. *Pattern Recognition Letters*, **25**, 701–710.

Wang, K., McLachlan, G.J., Ng, S.K., and Peel, D. (2009). EMMIX-skew: EM Algorithm for Mixture of Multivariate Skew Normal/t Distributions. URL: `http://www.maths.ed.uq.edu.au/gjm/mixsoft/EMMIX-skew`, R package version 1.0–12.

Weimerskirch, H., Pinaud, D., Pawlowski, F., and Bost, C.A. (2007). Does prey capture induce area-restricted search? A fine-scale study using GPS in a marine predator, the wandering albatross. *The American Naturalist*, **170**, 734–743.

Weitzman, M.S. (1970). Measures of overlap of income distribution of white and negro families in the United States. Technical Report No. 22, *US Department of Commerce, USA*.

West, M., and Harrison, W. (1997). *Bayesian Forecasting and Dynamic Models*, Second Edition. New York: Springer.

White, G.C. and Garrott, R.A. (1990). *Analysis of Wildlife Radio-Tracking Data*. London: Academic Press.

Worton, B.J. (1987). A review of models of home range for animal movement. *Ecological Modelling*, **38**, 277–298.

Worton, B.J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, **70**, 164–168.

Worton, B.J. (1995a). Modelling radio-tracking data. *Environmental and Ecological Statistics*, **2**, 15–23.

Worton, B.J. (1995b). A convex hull-based estimator of home-range size. *Biometrics*, **51**, 1206–1215.