Modelling larval movement data from individual bioassays

Citation for published version:

Digital Object Identifier (DOI):
10.1002/bimj.201400035

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Biometrical Journal

Publisher Rights Statement:
© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Modelling larval movement data from individual bioassays

Chris R. McLellan 1, Bruce J. Worton∗,1, William Deasy 2,3, and A. Nicholas E. Birch 3

1 School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh, James Clerk Maxwell Building, King’s Buildings, Mayfield Road, Edinburgh, EH9 3JZ, UK
2 SRUC, West Mains Road, Edinburgh, EH9 3JG, UK
3 The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, UK

We consider modelling the movements of larvae using individual bioassays in which data are collected at a high-frequency rate of 5 observations per second. The aim is to characterize the behaviour of the larvae when exposed to attractant and repellent compounds. Mixtures of diffusion processes, as well as Hidden Markov models, are proposed as models of larval movement. These models account for directed and localized movements, and successfully distinguish between the behaviour of larvae exposed to attractant and repellent compounds. A simulation study illustrates the advantage of using a Hidden Markov model rather than a simpler mixture model. Practical aspects of model estimation and inference are considered on extensive data collected in a study of novel approaches for the management of cabbage root fly.

Key words: Diffusion; Hidden Markov models; Mixtures; Model comparison.

Supporting Information for this article is available from the authors or on the WWW under http://dx.doi.org/10.1022/bimj.XXXXXXX

1 Introduction

The larva of the cabbage root fly, Delia radicum L. (Diptera: Anthomyiidae), is a serious pest that causes damage to Brassica host plants by feeding on their roots. The use of organophosphate insecticide for controlling larvae is restricted to a single pre-planting application and novel alternative treatments are currently being investigated. One of these potential treatments is to develop a way of manipulating the behaviour of the larvae in their attempts to find host plants. Studies suggest that the larvae respond to the odour of Brassica plants, and use the presence of specific chemicals excreted by the roots of these plants to locate suitable hosts (Ross and Anderson, 1992; Johnson and Gregory, 2006; Deasy, 2011). The larvae have also been shown to be repelled by sufficiently high concentrations of plant-specific chemicals (Finch, 1977; Koštál, 1992; Ewan, 2011). If these chemicals can be identified, and their effects upon the behaviour of the larvae understood, it may be possible to develop a control system using the appropriate plant extracts as soil amendments to act as deterrents.

Often, studies involve collecting data on locations of larvae after a given time and using standard statistical analysis of circular data, but here we consider the more challenging problem of modelling the tracks of the larvae. 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 data. We investigate suitable models for such data, 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.

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
We initially propose a model of larval movement with a general form defined by the conditional distribution of the position of the larva \( X_{s+t} \) at time \( s + t \) given its position \( X_s \) at time \( s \). In particular, we assume a bivariate normal distribution,

\[
X_{s+t} | X_s \sim N \{ a + \Gamma (X_s - a), \Phi \},
\]

(1)

where \( a \) is a known point of attraction or repulsion as indicated in Figure 1, and the matrices \( \Gamma \) and \( \Phi \) are defined as

\[
\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \quad \text{and} \quad \Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.
\]

Here, \( \Gamma \) is a centralization matrix determining the strength of attraction to, or repulsion from, the point \( a \), and \( \Phi \) is the covariance matrix of the larva’s movement at each observation step. Both of these matrices are
dependent on the time \( t \) between successive observations, but as this is fixed at 0.2 seconds the parameters are constant throughout each bioassay and thus dependence on time is not explicitly included in the notation above. We note that in the modelling the observed locations of the larvae are limited to the arena within the Petri dish, as indicated in Figure 1, and if the larvae were to move outside the arena then the data collection would be terminated at the boundary of the arena.

Diffusion process (1) is similar to a bivariate Ornstein-Uhlenbeck diffusion process, which has been used in a different context to model the movements of radio tracked animals (Dunn and Gipson, 1977; Worton, 1995; Blackwell, 1997; Nations and Anderson-Sprecher, 2006). However, the latter process assumes that \( \Gamma \) and \( \Phi \) are linked by the relation \( \Phi = \Lambda - \Gamma \Delta \Gamma^T \), where \( \Lambda \) is the covariance matrix of the bivariate normal equilibrium distribution of \( X_s \). In the current context, an equilibrium distribution may not exist and we therefore relax this assumption within our modelling framework.

If \( \Gamma \) is a diagonal matrix with positive diagonal elements, a value of \( \gamma_{11} \) that is less than unity indicates that the \( x \)-coordinate of the larva’s position is attracted towards the \( x \)-coordinate of \( a \), while a value greater than unity implies repulsion. The parameter \( \gamma_{22} \) is similarly related to the \( y \)-coordinates of the larva’s position and of \( a \). Having non-zero values for the non-diagonal elements of \( \Gamma \) introduce additional complexity, but the above statements remain roughly applicable and still give a guide to presence of attraction or repulsion provided that the non-diagonal elements are small in comparison to the diagonal elements. In the case of the larval data sets we are particularly interested in the parameter \( \gamma_{22} \) as attraction is associated with movement in the \( y \) direction towards \( a \) from the initial starting position of the larvae.

In the bioassays, each larva is released at the origin, \( x_0 \), and generates a subsequent observed sample path, \( \{x_1, \ldots, x_n\} \), with a constant time interval between observations of 0.2 seconds. The distribution of each observation under diffusion process (1) is given by

\[
x_{i+1} | x_i \sim N \{a + \Gamma(x_i - a), \Phi\}, \quad i = 0, \ldots, n - 1,
\]

and the log likelihood for the data, up to an additive constant, is given by

\[
-\frac{n}{2} \ln |\Phi| - \frac{1}{2} \sum_{i=1}^{n} \{x_i - a - \Gamma(x_{i-1} - a)\}^T \Phi^{-1} \{x_i - a - \Gamma(x_{i-1} - a)\}.
\]

In simple cases such as this it is possible to obtain explicit MLEs of the parameters, but in more complex mixed processes the likelihood may be difficult to express analytically. As \( a \) is known, the MLEs of the parameters of diffusion process (1) can be shown to be

\[
\hat{\Gamma} = \left\{ \sum_{i=1}^{n} (x_i - a)(x_{i-1} - a)^T \right\}^{-1} \left\{ \sum_{i=1}^{n} (x_{i-1} - a)(x_{i-1} - a)^T \right\},
\]

\[
\hat{\Phi} = \frac{1}{n} \sum_{i=1}^{n} (x_i - a - \hat{\Gamma}(x_{i-1} - a)) \{x_i - a - \hat{\Gamma}(x_{i-1} - a)\}^T,
\]

using a derivation similar to Anderson (1971, pp. 183–184).

While MLEs can be obtained easily for the simple diffusion process above, this is less straightforward for more complex models. Within a Bayesian framework, it is possible to obtain posterior estimates for the parameters of diffusion process (1) analytically but again this becomes extremely challenging when additional components are introduced. However, in general, Markov Chain Monte Carlo (MCMC) methods can be used to obtain posterior distributions for the parameters in such cases.

We note that the literature on statistical modelling of animal movement, which is related to the current problem, has been developed beyond (1) into much more flexible approaches (Blackwell, 2003; McClintock et al., 2012; Harris and Blackwell, 2013). However, there are some significant differences between the current type of bioassay data and wildlife tracking data. In particular, the larvae data are collected
under controlled experimental conditions and automatic monitoring is at a high-frequency collection rate of 5 locational observations per second. Each bioassay thus produces a very fine-scale track of locations to analyse and model, with thousands of observations collected over a relatively short time period of the duration of the bioassay.

3 Mixture and Hidden Markov models

3.1 Mixture modelling

We now consider a more flexible and complex model using mixtures of multiple diffusion processes. The bioassays described in Section 1 contain frequent localized movements, which occur because the position recorded for each observation is that of the larva’s head, and the larvae move by expansion and contraction of their bodies. The two distinct types of movement, directed and localized, can be represented by a mixture consisting of

- **S1** a component related to diffusion process (1) with attraction towards (or repulsion from) \(a\),
  \[
  x_{i+1}|x_i \sim N\{a + \Gamma(x_i - a), \Phi\},
  \]
  and
- **S2** a component accounting for localized movements resulting from body contractions (without any attraction/repulsion involving \(a\)),
  \[
  x_{i+1}|x_i \sim N(x_i, \Sigma),
  \]

where

\[\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},\]

is a covariance matrix, with a parameter \(0 < \pi < 1\) representing the probability of an observation being generated from component **S1**; for \(0 \leq i \leq n - 1\).

3.2 Hidden Markov modelling

The mixture model defined in Section 3.1 assumes that the component generating each observation is independent of previous observations. This may be too much of a simplification if a larva’s movement is more likely to remain in the same component than to switch. To allow for such possibilities, we consider the use of a Hidden Markov model (HMM; Frühwirth-Schnatter, 2006), which may offer a more appropriate approach for representing the behaviour of the larvae. We consider an HMM with states **S1** and **S2** as defined in Section 3.1, and a \(2 \times 2\) probability transition matrix

\[P = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix},\]

where \(\pi_{s_1,s_2}\) is the probability that an observation belongs to state \(s_2\) given that the previous one belongs to state \(s_1\). The HMM is more computationally intensive than the simpler mixture, but can model a wider range of larval behaviour. We also consider the inclusion of further states to gain greater flexibility in our model of larval path movement.

The bioassay data sets introduced in Section 1 include many identical successive points due to the short time interval between observations. These can be represented by a third state in the HMM in which an observation is identical to the previous one,
S3 no observed movement, $x_{i+1}$ is the same as $x_i$, and the HMM has a $3 \times 3$ transition matrix $P = (\pi_{s_1s_2})$.

To account for attraction towards (or repulsion from) the control zone $b$ in the experiment, we may include a fourth state,

S4 a diffusion process with point of attraction or repulsion $b$, $x_{i+1} \mid x_i \sim N(b + \Omega(x_i - b), \Psi)$, where $\Omega = (\omega_{kl})$ and $\Psi = (\psi_{kl})$, and in this HMM the matrix $P = (\pi_{s_1s_2})$ is $4 \times 4$.

4 Simulation study

A simulation study was conducted to explore the properties of an estimation approach for the two-state mixture and HMM models given in Sections 3.1 and 3.2. We simulated 100 sets of 1000 observations each from the mixture model, with the following parameter values

\begin{align*}
\mathbf{a} &= (0, 20)^T, \quad \Gamma = 0.95\mathbf{I}, \quad \Phi = 0.01\mathbf{I}, \quad \Sigma = 0.001\mathbf{I}, \quad \pi = 0.2. \tag{2}
\end{align*}

This process has one component which produces comparatively large movements influenced by the position of $a$, and a second component consisting of smaller localized movements. We let the starting point $x_0 = (0, 0)^T$ for all simulations. The two-component mixture and two-state HMM were both fitted to the simulated data sets. The following independent prior distributions were used for both of these models,

\begin{align*}
\gamma_{kl} &\sim \text{Normal}(1, 10^2), \quad k, l = 1, 2, \tag{3} \\
\Phi &\sim \text{Inverse-Wishart}\left(10^{-5}\mathbf{I}, 2\right), \tag{4} \\
\Sigma &\sim \text{Inverse-Wishart}\left(10^{-5}\mathbf{I}, 2\right), \tag{5}
\end{align*}

and represent vague prior information. 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., 2003; Carlin and Louis, 2008). Further discussion of the distribution’s properties can be found in Box and Tiao (1973) and Press (1982). For the mixture model, the prior distribution used for $\pi$ was

\begin{equation}
(\pi, 1 - \pi) \sim \text{Dirichlet}\left(\frac{1}{2}, \frac{1}{2}\right), \tag{6}
\end{equation}

i.e. a Beta distribution. The prior distributions used for each row $(\pi_{s_11}, \ldots, \pi_{s_1m})$ of the probability transition matrix $P$ for the HMMs were

\begin{equation}
(\pi_{s_11}, \ldots, \pi_{s_1m}) \sim \text{Dirichlet}\left(\frac{1}{2}, \ldots, \frac{1}{2}\right), \quad s_1 = 1, \ldots, m, \tag{7}
\end{equation}

where $m$ is the number of states, and in the particular case considered in the simulation study we have $m = 2$, so are using a Beta prior. Again these represent vague prior information.

Posterior distributions for the parameters were estimated with a Markov chain Monte Carlo approach using WinBUGS (Lunn et al., 2000). Summary statistics for the posterior means of selected parameters for the mixture and HMM models fitted to the simulated data sets are given in Table 1; note that the Monte Carlo errors associated with the values in the table are low relative to the values presented. These results include the mean and standard deviation, as well as the bias and root mean squared error (RMSE). The non-diagonal elements of $\Gamma$, $\Phi$ and $\Sigma$ are close to 0 and are not displayed.

The means of the posterior means for $\phi_{11}$ and $\phi_{22}$ for the fitted two-component mixture model, displayed in Table 1, are lower than the true values, and the corresponding values for $\sigma_{11}$ and $\sigma_{22}$ are higher. This indicates that some observations generated by component S1 have been assigned to component S2 and vice versa. Note that $\pi$ is estimated well, so the overall proportion of observations produced by each component has been estimated correctly. The means for $\gamma_{11}$ and $\gamma_{22}$ are close to the true values.
Table 1  Summary statistics of estimated posterior means of selected parameters obtained when fitting the two-component mixture model defined in Section 3.1, and the two-state HMM defined in Section 3.2, to 100 data sets of 1000 observations simulated from the two-component mixture model with the parameter values shown in (2). Presented summary statistics include the mean, standard deviation (SD), bias and root mean squared error (RMSE). The prior distributions of the parameters are given in (3), (4), (5), (6) and (7).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mixture</th>
<th>HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9614</td>
<td>0.9378</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9700</td>
<td>0.9528</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.0064</td>
<td>0.0065</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.0075</td>
<td>0.0064</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.0056</td>
<td>0.0053</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.0034</td>
<td>0.0030</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.2027</td>
<td>0.2412</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.2412</td>
<td>0.2412</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.7616</td>
<td>0.7616</td>
</tr>
</tbody>
</table>

The summary statistics obtained from fitting the two-state HMM to the same simulated data sets are also presented in Table 1. As for the mixture model, the means for $\gamma_{11}$ and $\gamma_{22}$ are close to the true values. However, with the HMM the means for $\phi_{11}$ and $\phi_{22}$ are much closer to the true values than in the mixture model, and the same is true to a lesser extent for $\sigma_{11}$ and $\sigma_{22}$. It is interesting to note that the HMM provides a satisfactory model even though the data sets were generated from a mixture model. The means for $\pi_{11}$ and $\pi_{22} = 1 - \pi_{22}$ are similar, indicating that the probability that an observation is generated from state $S_1$ is approximately the same regardless of the state of the previous observation and correctly reflecting the true stochastic process, i.e. the fitted HMM may be viewed as being equivalent to the simpler mixture model.

We now present the results of a simulation study that was conducted involving data sets generated from an HMM. In this study, 100 data sets of 1000 observations were simulated from the two-state HMM defined in Section 3.2, with parameter values

$$ a = (0, 20)^T, \quad \Gamma = 0.95 I, \quad \Phi = 0.01 I, \quad \Sigma = 0.001 I, \quad \pi_{11} = 0.5, \quad \pi_{22} = 0.9, $$

and the two-component mixture model and two-state HMM were both fitted to the simulated data sets. Note that the value of $\pi$, the probability that an observation is generated from state $S_1$, is equal to $1/6$ as given by

$$ \pi = \frac{1 - \pi_{22}}{2 - \pi_{11} - \pi_{22}}. $$

Summary statistics for the posterior means of selected parameters of the fitted mixtures and HMMs are shown in Table 2.

The means of posterior for $\gamma_{11}$ and $\gamma_{22}$ shown in Table 2 for the mixture model are close to the true values. The means for $\phi_{11}$ and $\phi_{22}$ are lower than the true values and the means for $\sigma_{11}$ and $\sigma_{22}$ are higher, suggesting that some observations have been assigned to the incorrect component of the mixture. The mean for $\pi$ has large bias, indicating that the mixture does not adequately model data sets generated by the HMM. Clearly, the mixture model has failed to capture certain features of the HMM due to overly simplistic assumptions made by the former about switching between states. However, as we saw above, the HMM was able to model the mixture data in a satisfactory manner at the expense of some slight extra variation.
Table 2  Summary statistics of estimated posterior means of selected parameters obtained when fitting the two-component mixture model defined in Section 3.1, and the two-state HMM defined in Section 3.2, to 100 data sets of 1000 observations simulated from the two-state HMM with the parameter values shown in (8). Presented summary statistics include the mean, standard deviation (SD), bias and root mean squared error (RMSE). The prior distributions of the parameters are given in (3), (4), (5), (6) and (7).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mixture</th>
<th>HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9560</td>
<td>0.0213</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9660</td>
<td>0.0012</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.0072</td>
<td>0.0011</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.0072</td>
<td>0.0012</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.0046</td>
<td>0.0007</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.0048</td>
<td>0.0085</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.3553</td>
<td>0.0191</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.8764</td>
<td>0.0168</td>
</tr>
</tbody>
</table>

The means of posterior means for $\gamma_{11}$ and $\gamma_{22}$ in Table 2 for the HMM are close to the true values, while the means for $\phi_{11}$ and $\phi_{22}$ and for $\sigma_{11}$ and $\sigma_{22}$ are closer to the true values than for the mixture model in Tables 1 and 2. This indicates that fewer observations are assigned to the wrong component by the HMM than by the mixture model. The means for $\pi_{11}$ and $\pi_{22}$ are both quite close to the true values. Overall, the results of the simulation study indicate that the fitted HMMs are similar to the true models used to generate the simulated data sets, but the mixture model approach has some limitations with regard to the flexibility of the models possible.

5  Application to larval movement data

In this section, we apply the methodology outlined in Section 3 to the larval data sets introduced in Section 1, and displayed in Figure 1. We initially use our simplest single component model from Section 2, but then consider the mixture model and HMM to account for the complex features of the data. Finally, we compare the models to assess which is the most appropriate. In the modelling we will use the priors defined in Section 4 as expert belief is vague. Use of informative prior information would be highly desirable for this problem but at the moment is difficult to specify.

5.1  Single component model

Diffusion process (1) was applied to the bioassays in Figure 1, with the prior distributions of $\Gamma$ and $\Phi$ as shown in (3) and (4). Table 3 presents the summary statistics of the posterior distributions of selected parameters. The parameters $\gamma_{12}$ and $\gamma_{21}$, which are not displayed, are close to 0 for all bioassays, so $\gamma_{11}$ and $\gamma_{22}$ describe the strength of attraction to or repulsion from a in the x direction and y direction respectively. The values of $\gamma_{12}$ and $\gamma_{21}$ are also close to 0 for the other models used throughout Section 5. For Bioassays 1 and 2, the posterior means of $\gamma_{11}$ and $\gamma_{22}$ are positive but less than unity, signifying attraction towards a, which is confirmed by visual inspection of Figure 1. However, the results for Bioassay 3 differ in that the posterior mean of $\gamma_{22}$ is greater than unity. This corresponds to movement away from a, as is evident in Figure 1.
Table 3  Summary statistics of posterior distributions of selected parameters for diffusion process (1) for Bioassays 1–3. The prior distributions of $\Gamma$ and $\Phi$ are as given in (3) and (4).

<table>
<thead>
<tr>
<th>Bioassay</th>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\gamma_{11}$</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.0003</td>
<td>0.9991</td>
<td>1.0001</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{22}$</td>
<td>0.9997</td>
<td>0.9997</td>
<td>0.0001</td>
<td>0.9996</td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>$\phi_{11}$</td>
<td>4.2576</td>
<td>4.2567</td>
<td>0.0641</td>
<td>4.1354</td>
<td>4.3882</td>
</tr>
<tr>
<td></td>
<td>$\phi_{22}$</td>
<td>4.5144</td>
<td>4.5142</td>
<td>0.0668</td>
<td>4.3875</td>
<td>4.6448</td>
</tr>
<tr>
<td>2</td>
<td>$\gamma_{11}$</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.0001</td>
<td>0.9997</td>
<td>1.0002</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{22}$</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.0001</td>
<td>0.9998</td>
<td>1.0001</td>
</tr>
<tr>
<td></td>
<td>$\phi_{11}$</td>
<td>4.1360</td>
<td>4.1350</td>
<td>0.0623</td>
<td>4.0174</td>
<td>4.2629</td>
</tr>
<tr>
<td></td>
<td>$\phi_{22}$</td>
<td>3.9608</td>
<td>3.9608</td>
<td>0.0586</td>
<td>3.8481</td>
<td>4.0774</td>
</tr>
<tr>
<td>3</td>
<td>$\gamma_{11}$</td>
<td>0.9994</td>
<td>0.9994</td>
<td>0.0020</td>
<td>0.9957</td>
<td>1.0035</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{22}$</td>
<td>1.0002</td>
<td>1.0002</td>
<td>0.0005</td>
<td>0.9993</td>
<td>1.0011</td>
</tr>
<tr>
<td></td>
<td>$\phi_{11}$</td>
<td>3.4986</td>
<td>3.4916</td>
<td>0.1466</td>
<td>3.2313</td>
<td>3.8028</td>
</tr>
<tr>
<td></td>
<td>$\phi_{22}$</td>
<td>5.2357</td>
<td>5.2284</td>
<td>0.2189</td>
<td>4.8198</td>
<td>5.6868</td>
</tr>
</tbody>
</table>

† Values multiplied by $10^5$.

5.2 Mixture model

The mixture model described in Section 3.1 was applied to the bioassays, and selected results of the analysis are presented in Table 4. For Bioassays 1 and 2, the posterior means of $\sigma_{11}$ and $\sigma_{22}$ are five orders of magnitude smaller than those for $\phi_{11}$ and $\phi_{22}$. This means that component $S_2$ of the mixture has captured the localized movements of the larva, while state $S_1$ represents larger directed movements. For both of these bioassays, about 20% of movements are assigned to component $S_1$, and the posterior means of $\gamma_{11}$ and $\gamma_{22}$ are positive but less than unity, indicating attraction towards $a$. For Bioassay 3, the posterior means of $\phi_{11}$, $\phi_{22}$, $\sigma_{11}$ and $\sigma_{22}$ indicate that component $S_1$ consists of the larva’s larger directed movements while state $S_2$ has accounted for localized movements. However, the posterior mean of $\gamma_{22}$ is greater than unity, and this is consistent with the path in Figure 1, which shows movement away from $a$.

5.3 Hidden Markov models

The two-state HMM described in Section 3.2 was applied to the bioassays, as it contains desirable features of attraction to (or repulsion from) the point $a$ as well as allowing for localized movements due to body contraction, but also has the transition matrix $P$. Summary statistics for selected parameters are displayed in Table 5, and describe a process with some differences from the mixture. For Bioassay 1, the posterior mean of $\tau_{11}$ and $\tau_{22}$ suggest that the larva spends about 13% of the time in state $S_1$ as given by (9), considerably less than the 23% obtained for the mixture model. The posterior means of $\gamma_{11}$ and $\gamma_{22}$ indicate attraction towards $a$, in the $y$ direction. However, the smaller proportion of movements assigned to state $S_1$ means that this influence is weaker than suggested by the mixture model. For Bioassays 2 and 3, the results are similar to those obtained for the mixture, in that the posterior means show attraction towards $a$ for Bioassay 2 and repulsion for Bioassay 3, and the larva spends approximately 22% and 25% of the time in state $S_1$ respectively for these two bioassays.

The three-state model was fitted to each of the bioassays to investigate whether a third state $S_3$ might improve on possible inadequacies of the two-state HMM by incorporating a null move state into the HMM. For Bioassays 1 and 2, as shown in Tables 6 and 7 respectively, the posterior means of $\tau_{11}$ and $\tau_{22}$ again imply attraction towards $a$. The posterior means of $\sigma_{11}$ and $\sigma_{22}$ reveal that state $S_2$ consists almost entirely...
Biometrical Journal 00 (2014) 00

Table 4  Summary statistics of posterior distributions of selected parameters for the two-component mixture model defined in Section 3.1 for Bioassays 1–3. The prior distributions of the parameters are given in (3), (4), (5) and (6).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior summary statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Bioassay 1</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9984</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9990</td>
</tr>
<tr>
<td>$\phi_{11}^1$</td>
<td>1.8312</td>
</tr>
<tr>
<td>$\phi_{22}^1$</td>
<td>1.9346</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.1449</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.1447</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.2328</td>
</tr>
<tr>
<td>Bioassay 2</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9993</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9994</td>
</tr>
<tr>
<td>$\phi_{11}^1$</td>
<td>1.9181</td>
</tr>
<tr>
<td>$\phi_{22}^1$</td>
<td>1.8430</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.1416</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.1417</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.2150</td>
</tr>
<tr>
<td>Bioassay 3</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9986</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>1.0011</td>
</tr>
<tr>
<td>$\phi_{11}^1$</td>
<td>1.3869</td>
</tr>
<tr>
<td>$\phi_{22}^1$</td>
<td>1.8642</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1.1583</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>1.1555</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0.2535</td>
</tr>
</tbody>
</table>

† Values multiplied by $10^4$.
‡ Values multiplied by $10^8$.

of movement in the $y$ direction, as $\sigma_{22}$ is several orders of magnitude larger than $\sigma_{11}$. The posterior means of $\phi_{11}$ and $\phi_{22}$ are of roughly the same order of magnitude as that of $\sigma_{22}$, implying that the model consists of two components with movements in the $y$ direction of the same order of magnitude plus the “null” state S3 of stationary observations, whereas the mixture model and two-state HMM only suggested one diffusion process consisting of larger movements and the other of much smaller movements. This difference may be interpreted as stationary observations that were assigned to one of the two diffusion processes by the two-state model instead being assigned to the null state S3 by the three-state model. Therefore, the three-state HMM with account taken for lack of movement at certain steps, appears to provide a more satisfactory description of the observed data.
Table 5  Summary statistics of posterior distributions of selected parameters for the two-state HMM defined in Section 3.2 for Bioassays 1–3. The prior distributions of the parameters are given in (3), (4), (5) and (7).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior summary statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Bioassay 1</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>1.0001</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9983</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>0.5118</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>3.4438</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.4131</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.1279</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.3929</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.9090</td>
</tr>
<tr>
<td>Bioassay 2</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9993</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9995</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>1.9170</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>1.8445</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.1416</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.1415</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.3916</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.8332</td>
</tr>
<tr>
<td>Bioassay 3</td>
<td></td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9993</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>1.0013</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>1.3869</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>1.8557</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1.1553</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>1.1539</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.4439</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.8108</td>
</tr>
</tbody>
</table>

† Values multiplied by $10^4$.
‡ Values multiplied by $10^8$. 
Table 6  Summary statistics of posterior distributions of selected parameters for the three-state HMM defined in Section 3.2 for Bioassay 1. The prior distributions of the parameters are given in (3), (4), (5) and (7).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9966</td>
<td>0.9966</td>
<td>0.0023</td>
<td>0.9921</td>
<td>1.0011</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.0003</td>
<td>0.9991</td>
<td>1.0001</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>3.5111</td>
<td>3.5086</td>
<td>0.1502</td>
<td>3.2375</td>
<td>3.8238</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.5403</td>
<td>0.5400</td>
<td>0.0236</td>
<td>0.4961</td>
<td>0.5881</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.9978</td>
<td>0.9964</td>
<td>0.0448</td>
<td>0.9145</td>
<td>1.0887</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>3.4700</td>
<td>3.4679</td>
<td>0.1526</td>
<td>3.1766</td>
<td>3.7815</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.4112</td>
<td>0.4107</td>
<td>0.0149</td>
<td>0.3829</td>
<td>0.4404</td>
</tr>
<tr>
<td>$\pi_{12}$</td>
<td>0.0691</td>
<td>0.0688</td>
<td>0.0077</td>
<td>0.0548</td>
<td>0.0842</td>
</tr>
<tr>
<td>$\pi_{13}$</td>
<td>0.5197</td>
<td>0.5198</td>
<td>0.0149</td>
<td>0.4902</td>
<td>0.5486</td>
</tr>
<tr>
<td>$\pi_{21}$</td>
<td>0.0618</td>
<td>0.0616</td>
<td>0.0077</td>
<td>0.0476</td>
<td>0.0774</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.3704</td>
<td>0.3701</td>
<td>0.0152</td>
<td>0.3416</td>
<td>0.4006</td>
</tr>
<tr>
<td>$\pi_{23}$</td>
<td>0.5678</td>
<td>0.5675</td>
<td>0.0157</td>
<td>0.5363</td>
<td>0.5978</td>
</tr>
<tr>
<td>$\pi_{31}$</td>
<td>0.0843</td>
<td>0.0843</td>
<td>0.0033</td>
<td>0.0787</td>
<td>0.0908</td>
</tr>
<tr>
<td>$\pi_{32}$</td>
<td>0.0808</td>
<td>0.0808</td>
<td>0.0034</td>
<td>0.0743</td>
<td>0.0879</td>
</tr>
<tr>
<td>$\pi_{33}$</td>
<td>0.8349</td>
<td>0.8349</td>
<td>0.0045</td>
<td>0.8260</td>
<td>0.8437</td>
</tr>
</tbody>
</table>

† Values multiplied by $10^4$.
‡ Values multiplied by $10^8$.

Table 7  Summary statistics of posterior distributions of selected parameters for the three-state HMM defined in Section 3.2 for Bioassay 2. The prior distributions of the parameters are given in (3), (4), (5) and (7).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{11}$</td>
<td>0.9986</td>
<td>0.9985</td>
<td>0.0012</td>
<td>0.9963</td>
<td>1.0009</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>0.9998</td>
<td>0.9998</td>
<td>0.0003</td>
<td>0.9992</td>
<td>1.0004</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>3.4403</td>
<td>3.4376</td>
<td>0.1502</td>
<td>3.1513</td>
<td>3.7494</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>0.5810</td>
<td>0.5805</td>
<td>0.0254</td>
<td>0.5335</td>
<td>0.6334</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.1167</td>
<td>0.1165</td>
<td>0.0057</td>
<td>0.1063</td>
<td>0.1284</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>3.4241</td>
<td>3.4196</td>
<td>0.1661</td>
<td>3.1162</td>
<td>3.7734</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.2855</td>
<td>0.2858</td>
<td>0.0139</td>
<td>0.2588</td>
<td>0.3124</td>
</tr>
<tr>
<td>$\pi_{12}$</td>
<td>0.1008</td>
<td>0.1003</td>
<td>0.0092</td>
<td>0.0840</td>
<td>0.1196</td>
</tr>
<tr>
<td>$\pi_{13}$</td>
<td>0.6137</td>
<td>0.6136</td>
<td>0.0147</td>
<td>0.5848</td>
<td>0.6431</td>
</tr>
<tr>
<td>$\pi_{21}$</td>
<td>0.1271</td>
<td>0.1268</td>
<td>0.0116</td>
<td>0.1052</td>
<td>0.1509</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.2710</td>
<td>0.2711</td>
<td>0.0154</td>
<td>0.2413</td>
<td>0.3007</td>
</tr>
<tr>
<td>$\pi_{23}$</td>
<td>0.6019</td>
<td>0.6012</td>
<td>0.0169</td>
<td>0.5694</td>
<td>0.6372</td>
</tr>
<tr>
<td>$\pi_{31}$</td>
<td>0.0936</td>
<td>0.0936</td>
<td>0.0034</td>
<td>0.0871</td>
<td>0.1001</td>
</tr>
<tr>
<td>$\pi_{32}$</td>
<td>0.0732</td>
<td>0.0732</td>
<td>0.0031</td>
<td>0.0672</td>
<td>0.0795</td>
</tr>
<tr>
<td>$\pi_{33}$</td>
<td>0.8332</td>
<td>0.8333</td>
<td>0.0044</td>
<td>0.8244</td>
<td>0.8417</td>
</tr>
</tbody>
</table>

† Values multiplied by $10^4$.
‡ Values multiplied by $10^7$. 
For Bioassay 3, the results, given in Table 8, are quite different from Bioassays 1 and 2. The posterior means of $\gamma_{11}$ and $\gamma_{22}$ imply repulsion from $a$, and the posterior means of $\phi_{11}$ and $\phi_{22}$ indicate that state $S_1$ consists primarily of movement in the $y$ direction. This confirms our visual impression from Figure 1, since the larva moves almost directly away from $a$, which results in such movement. For all bioassays, the posterior means of $\pi_{13}$, $\pi_{23}$ and $\pi_{33}$ highlight the importance of including state $S_3$ in the model.

Table 8 Summary statistics of posterior distributions of selected parameters for the three-state HMM defined in Section 3.2 for Bioassay 3. The prior distributions of the parameters are given in (3), (4), (5) and (7).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Posterior summary statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>$\gamma_{11}$</td>
<td>1.0000</td>
</tr>
<tr>
<td>$\gamma_{22}$</td>
<td>1.0025</td>
</tr>
<tr>
<td>$\phi_{11}^\dagger$</td>
<td>0.6227</td>
</tr>
<tr>
<td>$\phi_{22}^\dagger$</td>
<td>2.6030</td>
</tr>
<tr>
<td>$\sigma_{11}^\dagger$</td>
<td>3.1305</td>
</tr>
<tr>
<td>$\sigma_{22}^\dagger$</td>
<td>0.8225</td>
</tr>
<tr>
<td>$\pi_{11}$</td>
<td>0.3308</td>
</tr>
<tr>
<td>$\pi_{12}$</td>
<td>0.1314</td>
</tr>
<tr>
<td>$\pi_{13}$</td>
<td>0.5377</td>
</tr>
<tr>
<td>$\pi_{21}$</td>
<td>0.0615</td>
</tr>
<tr>
<td>$\pi_{22}$</td>
<td>0.3623</td>
</tr>
<tr>
<td>$\pi_{23}$</td>
<td>0.5762</td>
</tr>
<tr>
<td>$\pi_{31}$</td>
<td>0.1183</td>
</tr>
<tr>
<td>$\pi_{32}$</td>
<td>0.0718</td>
</tr>
<tr>
<td>$\pi_{33}$</td>
<td>0.8100</td>
</tr>
</tbody>
</table>

$\dagger$ Values multiplied by $10^4$.
$\ddagger$ Values multiplied by $10^7$.

The four-state model was applied to the bioassays to incorporate the control zone $b$ (see Figure 1) into the model, using the following independent vague priors,

$$\omega_{kl} \sim \text{Normal}(0, 10^2), \ k, l = 1, 2,$$

$$\Psi \sim \text{Inverse-Wishart}(10^{-5} \cdot I, 2),$$

with priors for the other parameters as shown in (3), (4), (5) and (7). For these fitted models the posterior means of $\gamma_{11}$ and $\gamma_{22}$ indicate attraction towards $a$ for Bioassays 1 and 2, and repulsion for Bioassay 3, as was the case for the other models used. However, we found that the inclusion of the fourth state in the HMM added very little in terms of the feature of the behaviour in which we are primarily interested. In addition, the four-state model has a possible disadvantage that it may have some difficulty distinguishing between state $S_1$ and state $S_4$ as they can be viewed as being similar in nature, i.e. movement towards/away from the chemical/control zones in the $y$ direction. These considerations suggest that the three-state HMM may be sufficient for our modelling.

5.4 Model comparison

The Bayesian Information Criterion (BIC) was used to give some comparison of the competing models. BIC values for the models fitted to the bioassays are presented in Table 9. The mixture and the two-state HMM are improvements over the single diffusion process. However, the three-state HMMs incorporating
null movement result in much lower BIC values for all three bioassays, and are therefore to be preferred over both two-component models. The addition of a fourth state describing influence by the control zone provides further improvement over the three-state model for Bioassay 1, though the difference in BIC is not as large as the difference between the two-state and three-state models. For Bioassays 2 and 3, the three-state model is superior. The BIC values give us a guide, and together with our previous discussion above, suggest that the three-state model provides a compromise between accounting for the important features of the larval track data without incorporating unnecessary detail.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of parameters</th>
<th>Bioassay 1</th>
<th>Bioassay 2</th>
<th>Bioassay 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single diffusion process</td>
<td>7</td>
<td>-129469</td>
<td>-130936</td>
<td>-16719</td>
</tr>
<tr>
<td>Two-component mixture model</td>
<td>11</td>
<td>-279611</td>
<td>-284020</td>
<td>-31864</td>
</tr>
<tr>
<td>Two-state HMM</td>
<td>12</td>
<td>-216830</td>
<td>-284018</td>
<td>-31861</td>
</tr>
<tr>
<td>Three-state HMM</td>
<td>16</td>
<td>-581770</td>
<td>-590949</td>
<td>-73338</td>
</tr>
<tr>
<td>Four-state HMM</td>
<td>29</td>
<td>-588411</td>
<td>-588236</td>
<td>-73224</td>
</tr>
</tbody>
</table>

### 5.5 Plots of posterior densities

The strength of the influence that the position of a has on the larva’s movements is determined by $|\Gamma|$. However, as the position of a in each bioassay has an $x$-coordinate very close to the $x$-coordinate of the larva’s starting position, the strength of attraction to or repulsion from a is primarily determined by the parameter $\gamma_{22}$. As noted above, a non-negative value of $\gamma_{22}$ less than unity suggests movement in the direction of a, which corresponds to attraction, and a value greater than unity indicates repulsion. As this is the aspect of movement of primary interest, we focus our attention on the parameter $\gamma_{22}$ as well as $|\Gamma|$. Posterior density plots of $\gamma_{22}$ and of $|\Gamma|$ are shown in Figure 2 for the three-state HMM, and it is apparent that the majority of the densities of both parameters are below unity for Bioassays 1 and 2 and above it for Bioassay 3. Furthermore, in the case of $\gamma_{22}$ the dispersion of the density obtained for Bioassay 3 is clearly different from the other two bioassays, having much higher variance. The fitted three-state 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 Bioassay 3 and the other two bioassays is less pronounced in the case of $|\Gamma|$. We note that a $\gamma_{22}$ value of unity corresponds to no attraction, but not actually repulsion.

### 5.6 MCMC diagnostics and posterior predictive checks

Care is needed when applying MCMC techniques to ensure the simulations are providing a true picture of the posterior distribution. With all the models considered in this paper MCMC diagnostics were used to investigate any problems with model fitting. In addition, particular care is needed with mixture type problems with regard to potential ‘label switching’ (Stephens, 2000). Fortunately, the MCMC convergence was excellent for all models considered and also there were no issues with label switching. Figure 3 gives an illustration of the type of trace plots obtained in the MCMC in the case of the three-state HMM for the $\pi_{11}$ parameter, but trace plots for the other parameters have acceptable properties. Fortunately, the structure and nature of the components (with/without attraction) are sufficiently different so that the model fitting does not lead to label switching. Note that even in models with the null state that may have potentially lead to problems, the trace plots did not show any significant issues.

With regard to sensitivity to the prior information we investigated varying the priors and found that the results were essentially unchanged unless we used very extreme changes of prior information. This may
reflect the large sample sizes involved in the analyses. As noted at the beginning of Section 5, it would be desirable to use informative prior information in the Bayesian analysis, if it were available.

Posterior predictive checks were conducted to assess the adequacy of the models (Gelman et al., 2003). Simulated tracks generated from the three-state HMM using values simulated from the posterior distribution were compared with the observed track. In particular, we used the final observation as an assessment. This is the observation that is collected in non-tracking experiments so provides a natural assessment of the adequacy of the model. Figure 4 shows the estimated posterior predictive distributions of the distance between the final location of the track and the point \( \mathbf{a} \). For Bioassay 1 the observed value is towards the lower end of the distribution, but there is no significant evidence of disagreement with the model. Similarly, for Bioassays 2 and 3, the three-state HMM seems adequate for our modelling.

6 Conclusions

We have seen that diffusion based processes may be used to model the tracks of cabbage root fly larvae. By doing this we can gain a greater understanding of the processes that govern their movements. It is clear, even from the plots of the bioassay data in Figure 1, that behaviour of the larvae in the presence of allyl isothiocyanate (Bioassay 3) is very different from that in the case of damaged broccoli roots (Bioassays 1 and 2), but we can use the modelling to characterize the movements more quantitatively. The mixture of bivariate normal diffusion processes successfully indicates that the larvae in Bioassays 1 and 2 are moving towards the zone of damaged broccoli roots, and that the larva in Bioassay 3 is repelled from the allyl isothiocyanate. However, we found that there are advantages of using an HMM approach, and due to their increased complexity the HMMs produced more biologically realistic descriptions.

Although it is possible to collect extensive sets of data with the currently available experimental equipment and software, care is needed to incorporate important features of the movement into our models of larval paths. The two-state HMM captures the salient features of the data and clearly differentiates between the attraction towards \( \mathbf{a} \) present in Bioassays 1 and 2 and the repulsion in Bioassay 3. The addition of a third “null” state, representing no movement, results in a more appropriate model that accounts for the successive identical observations present in the data. Model comparison using BIC indicates that the three-state model is far superior to the two-state HMM and mixture, and it is also preferred to the four-state model for two of three bioassays.

While the models used here are successful at identifying attraction to and repulsion from \( \mathbf{a} \) and provide a suitable model for the bioassays, we conclude the article by considering possible alternatives and refinements. Further improvement to the models might involve modification to include more pronounced directional persistence. The data plot for Bioassay 3 in particular, presented in Figure 1, suggests that the
movements of the larva are highly dependent on the current direction of movement, whereas the HMM used here only involves first-order dependence and so is fairly limited in its ability to model this behaviour. As such, a natural development would be to incorporate higher-order dependence. This may result in models even more appropriate for the bioassay data, and allow further inference to be made about the behaviour of the larvae. Nevertheless, the models we consider in this paper have worked well in extracting important behavioural responses with regard to the study of novel approaches for the management of cabbage root fly.

In this paper we have investigated modelling a small number of tracks from individual bioassays so that although the number of observations per bioassay is very large the sample size in terms of tracks is very

Figure 3  MCMC trace plots for the parameter $\pi_{11}$ for the three-state HMMs: Bioassay 1 (upper panel), Bioassay 2 (middle panel) and Bioassay 3 (lower panel).
Figure 4  Posterior predictive distributions of the distance between final location of the track and point a with the vertical line indicating the observed value: Bioassay 1 (left panel), Bioassay 2 (middle panel) and Bioassay 3 (right panel).

small. As more data become available there will be the opportunity to extend the modelling to simultaneously model a large number of larvae tracks. However, we might expect to have to include in such models individual variation with regard to each larva. By increasing the sample size it will be possible to obtain a fuller understanding of the population behaviour of the larvae with regard to the attraction/repulsion.

Acknowledgements  The authors would like to thank the Editor Dr. Lutz Edler, the Associate Editor and an anonymous referee for their extremely valuable comments and suggestions on an earlier version of this paper, which substantially improved the content and clarity of the article. We also thank Dr. Benjamin Hofner, the Reproducible Research Editor, for very helpful comments concerning the code. The experimental study was conducted at The James Hutton Institute, Invergowrie, Dundee. Chris McLellan is supported by an EPSRC studentship, and William Deasy is supported by an HDC studentship (FV 364 Novel approaches for the management of cabbage root fly).

Conflict of Interest  The authors have declared no conflict of interest.

References


