

Gaussian Process Segmentation of Co-Moving Animals

Steven Reece^{*}, Richard Mann[†], Iead Rezek^{**} and Stephen Roberts^{*}

^{}Department of Engineering Science, Oxford University, Oxford, UK*

[†]Department of Mathematics, Uppsala University, Uppsala, Sweden

*^{**}Department of Clinical Neuroscience, Imperial College, London, UK*

Abstract. The analysis of how groups of animals move collectively is the focus of much current research. This paper offers a Bayesian model of animal co-movement based on a mixture of Gaussian processes.

Keywords: Gaussian processes, Variational Bayes, Tracking, Bayesian model order selection

PACS: 02.50.Cw, 89.75.Fb

INTRODUCTION

The analysis of how groups of animals move collectively and how they effectively align their movements is the focus of much current research [1]. A key research question is how individuals transfer information to retain group cohesion whilst achieving both collective and individual goals. The potential existence of distinct subgroups, the members of which are all more closely connected to each other than to other members of the collective, suggest a somewhat partitioned graph of inter-individual interactions. This would have implications for the speed and reliability of information transfer within the group and thus the effectiveness of the group's response to external factors such as the presence of predators and its stability as a cohesive unit.

Numerical simulation models of collective behaviour usually propose simple individual rules that allow each animal to interact with others within a certain 'interaction radius', normally defined by a euclidean metric (e.g. [2]), or allow for an interaction of adjustable strength depending on the individuals' spatial separation (e.g. [3]). Recently the fine structure of these interactions has attracted increased attention. Studies have used the average correlations in the positions or movements of different individuals to infer the existence of interactions and, as a result, have proposed alternative models for how these are generated, suggesting for example a topological model [4] or a hierarchical model [5] as alternatives to the standard euclidean metric.

We propose to formalise these correlation based approaches through the use of Gaussian processes (GP) to model the distribution of individual movement paths. GPs have been used previously to describe the distribution of pigeon flight paths that are potentially highly correlated [6]. Our approach not only forms groups based on individuals' spatial separation, but it uses their full temporal behaviours. Thus, our method is able to discern individual groups even when these groups coincide spatially.

We apply Variational Bayes to a Gaussian process mixture model and are thus able to discern the number of groups within a collective as well as the group compositions

efficiently. Our approach exploits Gaussian process interpolation to group animals even when the data is of poor quality which may involve noisy, sparsely and asynchronously observed paths. Further, animals often suffer disorientation and their paths deviate temporarily from a smooth route. We exploit non-stationary Gaussian processes to model animals when they suffer temporary disorientation. We demonstrate the efficacy of our method on simulated data sets and real homing pigeon data.

MODEL

We assume that the data comprises N sparsely and asynchronously observed paths drawn from up to M Gaussian mixture components (or groups). Let y_n (with $1 \leq n \leq N$) be the observations made of path n at times X_n . Also, let $X = \bigcup_{i=1}^N X_n$ be the accumulated set of times at which some path was observed and define C_n and C as the cardinality of X_n and X , respectively. Finally, define the $C_n \times C$ *observation indicator matrix*, H_n , so that $X_n = H_n X$. The indicator matrix is key to fusing asynchronous path observations.

Each path, y_n , belonging to group g , is assumed to be a multi-variate Gaussian random variable:

$$y_n \sim N(\mu_g(X_n), K_p(X_n, X_n)^{-1})$$

where $\mu_g(X_n)$ is the Gaussian group mean and $K_p(X_n, X_n)^{-1}$ is its precision. Similarly, the group mean, $\mu_g(X)$, over all times, X , is a multi-variate Gaussian random variable:

$$\mu_g(X) \sim N(0, K_g(X, X)^{-1}).$$

The covariance matrices, $K_p(X_n, X_n)$ and $K_g(X, X)$, are derived from GP covariance functions, K_p and K_g , respectively. We use the Matérn covariance function [7], $K_{\text{Matérn}}$, with the Bessel function order set to $3/2$. This choice of covariance function imposes few assumptions on the form of the data and requires only that the data generating process is continuous and first order differentiable. The group mean covariance function is thus:

$$K_g(x_i, x_j; l_g) = K_{\text{Matérn}}(x_i, x_j; l_g)$$

where l_g is the group input scale. However, we need to modify the Matérn kernel slightly to allow it to model discontinuous paths. We assume that it is possible during pre-processing to identify the times, Δ , when paths are discontinuous. Each contiguous path segment is assumed to be generated by the Matérn. However, the segments are assumed to be uncorrelated across segment boundaries. Since each path will exhibit discontinuities at different times then a distinct covariance function is defined for each path n [8]:

$$K_p(x_i, x_j; l_p, \Delta_n) = \begin{cases} K_{\text{Matérn}}(x_i, x_j; l_p) & \text{when } (x_i - \delta)(x_j - \delta) \geq 0 \quad (\forall \delta \in \Delta_n), \\ 0 & \text{otherwise.} \end{cases}$$

Each covariance is scaled by a positive factor, S , which is the *output scale*. For example, we may use $K_g(x_i, x_j; l_g)/S_g$ in place of $K_g(x_i, x_j; l_g)$ as the group covariance.

Throughout this paper we consider a simplified model in which the group input and output scales are identical for all groups and similarly for the path input and output scales.

METHOD FOR INFERRING GROUP MEMBERSHIP

This section outlines our approach to determining the number of groups within the data set as well as the group compositions. Given data, D , the appropriate number of groups can be determined by setting the mixing weights, π , to maximize the marginal log-likelihood:

$$P(D | \pi, \rho) = \sum_{i=1}^M \pi_i N(D | \rho_i) .$$

A fully Bayesian treatment involves marginalising over prior distributions for the mixing weights and the hidden variables and their parameters, ρ . Following [9], our approach uses the familiar variational method to approximate the marginal. Variational learning [9] aims to minimise the so-called Kullback-Leiber (KL) divergence between the (intractable) model posterior P and a simpler (analytic) approximating distribution Q . Given a set of hidden variables $\rho = \{\rho_1, \dots, \rho_T\}$, the variational approximation assumes that the Q-distribution factorises, $Q(\rho) = \prod_{i=1}^T Q(\rho_i)$, with the additional constraint that $\int Q(\rho_i) d\rho_i = 1$.

In our model the set of hidden variables consists of the path mean, μ , for each group, the variables, s , indicating membership of a path to a particular group, and the GP covariance output scale parameters, S_p and S_g , for the paths and path means, respectively. The joint distribution of the hidden variables conditioned on the mixture model weights, π , the GP kernel length scales, l_p and l_g , and the path discontinuity points, Δ , is:

$$P(D, S_p, S_g, s | \pi, \Delta, l_p, l_g) = P(D | \mu, l_p, S_p, \Delta, s) P(\mu | l_g, S_g) P(S_p) P(S_g) P(s | \pi) . \quad (1)$$

The factors are:

$$\begin{aligned} P(D | \mu, l_p, S_p, \Delta, s) &= \prod_{n=1}^N \prod_{i=1}^M N(y_n; \mu_i(X_n), S_p K_p(X_n, X_n; l_p, \Delta_n)^{-1})^{s_{in}} , \\ P(\mu | l_g, S_g) &= \prod_{i=1}^M N(\mu_i(X); 0, S_g K_g(X, X; l_g)^{-1}) , \\ P(s | \pi) &= \prod_{n=1}^N \prod_{i=1}^M \pi_i^{s_{in}} , \quad P(S_p) = \text{Ga}(S_p; k_p^0, \theta_p^0) , \quad P(S_g) = \text{Ga}(S_g; k_g^0, \theta_g^0) \end{aligned}$$

where $\mu_i(X)$ is the mean trajectory of group i and θ and k are the gamma distribution scale and shape parameters for the GP output scales, respectively. The binary scalar, s_{in} , is unity if path n belongs to group i and zero otherwise.

We marginalise $\{\mu, S_p, S_g, s\}$ within the variational framework and, in keeping with [9], infer the most likely mixing weights. Unfortunately, no general, tractable variational factor exists which would allow marginalisation of the length scales within

the variational framework. Consequently, we sample the length scales via Monte-Carlo and find the most likely $\{l_p, l_g\}$. Our algorithm iteratively samples both group and path length scales, l_p and l_g , respectively, from a uniform distribution and then applies the variational treatment to the posterior conditioned on each sample pair. The following exposition of our variational equations uses nomenclature of [9] and the reader is invited to consult this publication for further details of the variational approach.

The posterior joint distribution for our variational approximation is:

$$Q(\mu, S_p, S_g, s) = Q_\mu(\mu) Q_{S_p}(S_p) Q_{S_g}(S_g) Q_s(s) \quad (2)$$

where the factor distributions are of the form:

$$Q_s(s) = \prod_{i=1}^M \prod_{n=1}^N p_{in}^{s_{in}}, \quad Q_\mu(\mu) = \prod_{i=1}^M N(\mu_i | m_\mu^{(i)}, S_p T_\mu^{(i)}),$$

$$Q_{S_p}(S_p) = \text{Ga}(S_p | k_p, \theta_p), \quad Q_{S_g}(S_g) = \text{Ga}(S_g | k_g, \theta_g).$$

The variational update equations which minimise the KL divergence between the true posterior, (1), and the approximate posterior, (2), are:

$$\begin{aligned} \log \tilde{p}_{in} &= \log \pi_i + \frac{1}{2} [C_n [\langle \log S_p \rangle - \log 2\pi] - \log |K_p(X_n, X_n; l_p, \Delta_n)| \\ &\quad - \langle S_p \rangle \text{Tr} \left\{ K_p(X_n, X_n; l_p, \Delta_n)^{-1} \left(y_n y_n^T - \langle \mu_i(X_n) \rangle y_n^T - y_n \langle \mu_i(X_n) \rangle^T + \langle \mu_i(X_n) \mu_i(X_n)^T \rangle \right) \right\}], \\ T_\mu^{(i)} &= \langle S_g \rangle K_g(X, X; l_g)^{-1} + \langle S_p \rangle \sum_{n=1}^N \langle s_{in} \rangle H_n^T K_p(X_n, X_n; l_p, \Delta_n)^{-1} H_n, \\ m_\mu^{(i)} &= \langle S_p \rangle T_\mu^{(i)-1} \sum_{n=1}^N \langle s_{in} \rangle H_n^T K_p(X_n, X_n; l_p, \Delta_n)^{-1} y_n, \quad k_p = k_p + \frac{1}{2} \sum_{i=1}^M \sum_{n=1}^N \langle s_{in} \rangle C_n, \\ \frac{1}{\theta_p} &= \frac{1}{\theta_p^0} + \frac{1}{2} \sum_{i=1}^M \sum_{n=1}^N \langle s_{in} \rangle \text{Tr} \left\{ K_p(X_n, X_n; l_p, \Delta_n)^{-1} \left(y_n y_n^T - \langle \mu_i(X_n) \rangle y_n^T - y_n \langle \mu_i(X_n) \rangle^T + \langle \mu_i(X_n) \mu_i(X_n)^T \rangle \right) \right\}, \\ \frac{1}{\theta_g} &= \frac{1}{\theta_g^0} + \frac{1}{2} \sum_{i=1}^M \text{Tr} \left\{ K_g(X, X; l_g)^{-1} \langle \mu_i(X) \mu_i(X)^T \rangle \right\}, \quad k_g = k_g^0 + \frac{MC}{2} \end{aligned}$$

where:

$$\begin{aligned} \langle \mu_i(X_n) \mu_i(X_n)^T \rangle &= H_n T_\mu^{(i)-1} H_n^T + m_\mu^{(i)}(X_n) m_\mu^{(i)}(X_n)^T, \\ \langle s_{in} \rangle &= p_{in}, \quad \langle \mu_i \rangle = m_\mu^{(i)}, \\ \langle S_p \rangle &= k_p \theta_p, \quad \langle \log S_p \rangle = \psi(k_p) + \log \theta_p, \\ \langle S_g \rangle &= k_g \theta_g, \quad \langle \log S_g \rangle = \psi(k_g) + \log \theta_g \end{aligned}$$

and ψ is the digamma function. The update equation for the probability, p_{in} , that path n belongs to group i is as in [9], $p_{in} = \tilde{p}_{in} / \sum_{i=1}^M \tilde{p}_{in}$, as well as the update equation for the mixture weights, $\pi_i = \sum_{n=1}^N p_{in} / N$.

The score function which we need to maximise is a lower bound for the log marginal likelihood, $\log P(D | \pi, l_p, l_g, \Delta)$:

$$\begin{aligned} L(Q) &= \langle \log P(D | \mu, l_p, S_p, \Delta, s) \rangle + \left\langle \log \frac{P(\mu | l_g, S_g)}{Q_\mu(\mu)} \right\rangle \\ &\quad + \langle \log P(S_p) / Q_{S_p}(S_p) \rangle + \langle \log P(S_g) / Q_{S_g}(S_g) \rangle + \langle \log P(s | \pi) / Q_s(s) \rangle. \end{aligned}$$

From the set of Q-distributions which maximise the score $L(Q)$ over $\{\pi, l_p, l_g\}$ we determine the path to group assignment probabilities, p_{in} . Using these probabilities, we assign path n to group g where $g = \operatorname{argmax}_i \{p_{in}\}$.

ILLUSTRATIVE EXAMPLES

We demonstrate the efficacy of our algorithm on simulated data here and then real data in the next section. In each of the following examples the path data is drawn from two groups and the algorithm is presented with a five component GP mixture model. The modified Matérn kernel is used throughout and 20 length scale sample pairs are generated for each problem. The length scales both here and in the next section are sampled from the range $[1, 60]$. The gamma priors are vague with shape and scale parameters set to 10 and 1000, respectively.

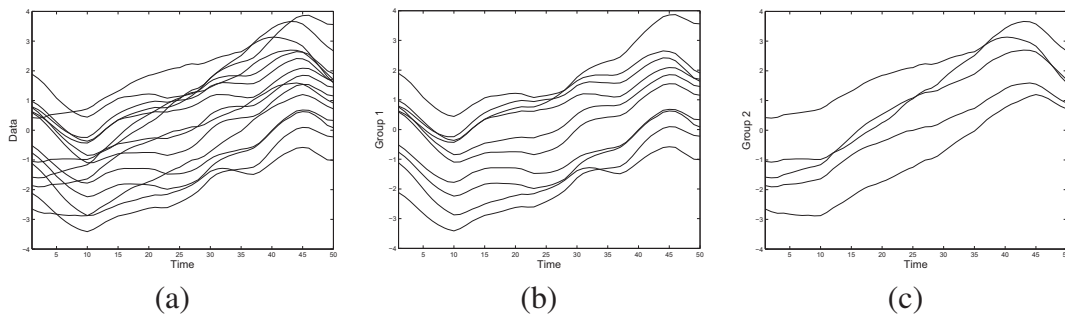


FIGURE 1. (a) Two overlapping sets of intra-group correlated paths. Panes (b) and (c) show their successful segmentation.

Figure 1 (a) shows two sets of paths which overlap spatially. These paths are generated from two different groups, each group has its own mean. Although the paths occupy the same spatial region their dynamics differ sufficiently for our algorithm to distinguish the group membership (see Figure 1, (b) and (c)).

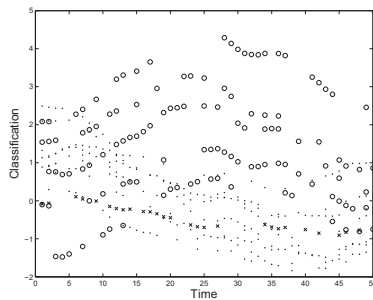


FIGURE 2. Two groups of paths which are intermittently observed. The groups are distinguished by a dot and circle and these symbols also denote when the path was observed. The path marked by crosses is incorrectly assigned to its own group.

In Figure 2 the targets are intermittently and asynchronously observed this time. The algorithm performs well on this data set and, in this case, only misclassifies one path. This path is mistakenly assigned its own solo group.

Figure 3 shows two groups of paths, some of these paths (marked in bold) are discontinuous. Again, when supplied with pre-processed information which identifies the discontinuous paths and where the discontinuities occur, the algorithm successfully allocates the paths to their groups. It is, perhaps, interesting to note that, when the

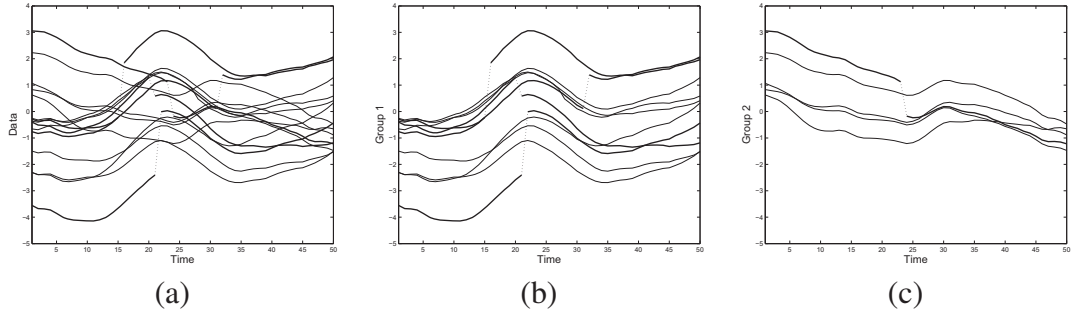


FIGURE 3. (a) Two groups of paths. Discontinuous paths are marked in bold. The algorithm successfully allocates the paths to their groups, shown in (b) and (c).

algorithm assumes that all paths are continuous then all continuous paths are mistakenly assigned to the same group and all discontinuous paths are assigned to their own solo groups.

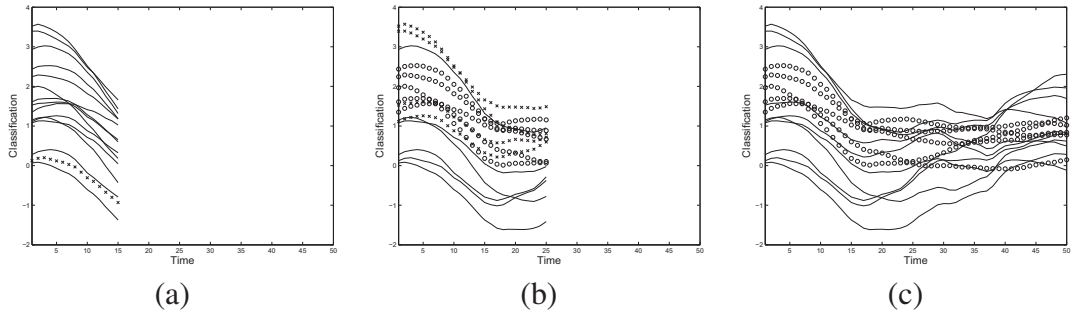


FIGURE 4. (a) A single group of co-moving targets splits at time $t = 20$ (b) and separates into two distinct groups (c).

Finally, in Figure 4 a single group of targets are co-moving before, at $t = 20$, splitting into two groups. We ran the algorithm at epochs $t = 15$, $t = 20$ and $t = 50$, each time using all data obtained up to and including the epoch. The algorithm successfully identified a single group at $t = 15$ and two distinct groups at $t = 50$.

APPLICATION

In this section we apply our method to determine the number of groups and group structure of co-moving birds. By inferring the optimum number of mean flight paths that correspond to strongly correlated groups we can segment the larger collective into subsets of strongly interacting individuals.

Since there is no ground truth data available for pre-segmented co-moving birds we consider sets of flight paths from multiple birds as if they were simultaneous flights from a large group of birds. We assume that strongly correlated groups within a larger collective show small variation around a common mean in a similar fashion to the manner in which an individual bird shows small variation around its idiosyncratic memorised route. Thus, we exploit data from individual homing pigeon flight paths [10] and, since

we know the identities of each of the real birds, we have a pre-defined ground truth for the correct group segmentation. The data comprises flight paths for two birds. Each bird flew 20 flights from the same site.

The data is subject to some pre-processing. Firstly, data for disorientated segments in each flight path are discarded leaving gaps in the data stream. Then, the remaining data is projected onto two axes, one aligned with the pigeon release and home vector (the X-axis) and an axis perpendicular to this (the Y-axis). The data for each path is then binned along the X-axis to reduce its dimensionality (the Y-axis component is averaged in each bin).

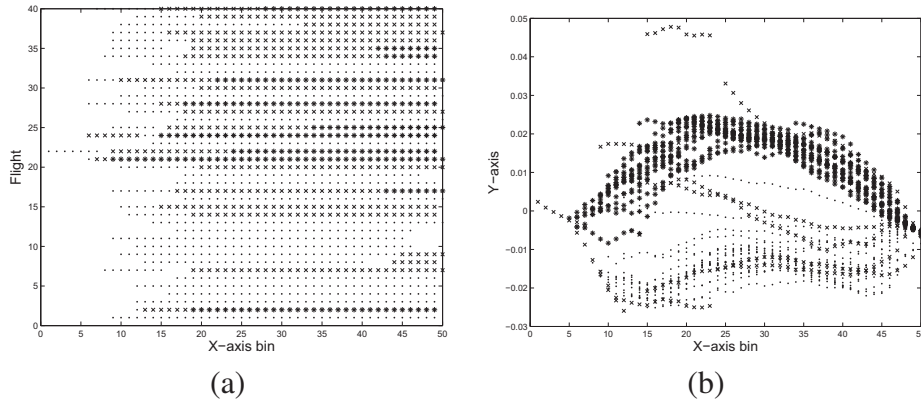


FIGURE 5. Data from 40 flight paths after pre-processing. (a) The presence of a '.', 'x' or '*' symbol denotes data available from that part of the flight path. A sequence of similar symbols denotes a contiguous flight segment during which a pigeon does not become disorientated. (b) Flight paths grouped into two classes. The '.' and '*' symbols denote correctly grouped paths. The 'x' denotes paths which have been misclassified.

Figure 5 (a) shows where each flight is visible after pre-processing and Figure 5 (b) shows how our algorithm has allocated flights to birds. Our method correctly identifies the presence of two birds and correctly classifies 87% of their flights. It is the earlier flights undertaken by both birds which are misclassified. These flights can be seen as exploratory and follow distinctly different paths to the birds' eventual habitual routes. As shown by [6] the flights are initially highly disordered and uncorrelated, but over successive releases become more strongly correlated and show substantial increase in cohesion.

We used our method to determine the number of exploratory flights undertaken by the birds. We applied our algorithm to sets of five paths from each of the two birds. The sets of five paths were chosen successively, so we tested the first five flights of each bird, then the 2nd flight to the 6th and so on until the final five flights (see Figure 6 (a)). Each time our method correctly identifies two groups except for flight sets 4, 5 and 6 for which an additional group was mistakenly identified.

The classification success is more variable and is shown as a function of flight set in Figure 6 (b). As can be seen in Figure 6 (b), there is a significant rate of misclassification initially. This accurately represents the lack of cohesion in these early paths. However, the classification success rate improves until we have a 100% success rate in the final 3 sets which correspond to the final 7 flights. As well as a successful test of our classification method our results provide a quantitative analysis of the emergence of

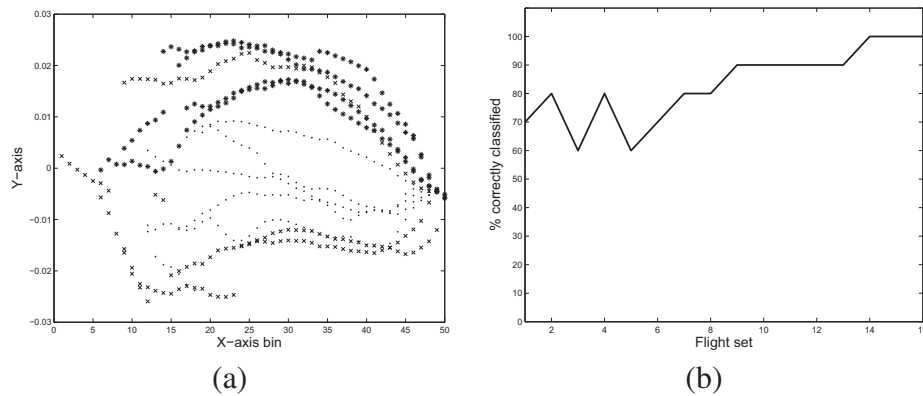


FIGURE 6. (a) Example classification of pigeon flight paths (low cohesive paths) and (b) the classification success as a function of flight set.

individual navigational strategies in learning animals.

ACKNOWLEDGMENTS

This research was undertaken as part of the ALADDIN (Autonomous Learning Agents for Decentralised Data and Information Networks) project which is jointly funded by a BAE Systems and EPSRC strategic partnership (EP/C548051/1).

REFERENCES

1. D. J. T. Sumpter, *Philosophical Transactions of the Royal Society of London: Series B* **361**, 5–22 (2006).
2. I. Couzin, J. Krause, N. Franks, and S. Levin, *Nature* **433**, 513–516 (2005).
3. D. Biro, D. Sumpter, J. Meade, and T. Guilford, *Current Biology* **16**, 2123–2128 (2006).
4. M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, et al., *Proceedings of the National Academy of Sciences* **105**, 1232 (2008).
5. M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, *Nature* **464**, 890–893 (2010).
6. R. Mann, R. Freeman, M. Osborne, R. Garnett, J. Meade, C. Armstrong, D. Biro, T. Guilford, and S. Roberts, “Gaussian Processes for Prediction of Homing Pigeon Flight Trajectories,” in *American Institute of Physics Conference Proceedings* 1193, 2009, pp. 360–367.
7. C. E. Rasmussen, and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2006.
8. R. Garnett, M. A. Osborne, S. Reece, A. Rogers, and S. J. Roberts, *The Computer Journal, Section C: Computational Intelligence* (2010), doi:10.1093/comjnl/bxq003.
9. A. Corduneanu, and C. M. Bishop, “Variational Bayesian Model Selection for Mixture Distributions,” in *Proc. Eighth International Conference on Artificial Intelligence and Statistics*, 2001, pp. 27–34.
10. J. Meade, D. Biro, and T. Guilford, *Proceedings of the Royal Society B: Biological Sciences* **272**, 17–23 (2005).