Gaussian Processes for Prediction of Homing Pigeon Flight Trajectories


*Department of Engineering Science, University of Oxford
†Microsoft Research, Cambridge
**Department of Zoology, University of Oxford
‡Department of Zoology, University of Sheffield

Abstract. We construct and apply a stochastic Gaussian Process (GP) model of flight trajectory generation for pigeons trained to home from specific release sites. The model shows increasing predictive power as the birds become familiar with the sites, mirroring the animal’s learning process. We show how the increasing similarity between successive flight trajectories can be used to infer, with increasing accuracy, an idealised route that captures the repeated spatial aspects of the bird’s flight. We subsequently use techniques associated with reduced-rank GP approximations to objectively identify the key waypoints used by each bird to memorise its idiosyncratic habitual route between the release site and the home loft.

Keywords: GP, Gaussian Process, Pigeon, Avian, Navigation

INTRODUCTION

The navigational abilities of birds have inspired scientific study for over a century. In this time the domestic pigeon (*Columba livia*) has become the canonical experimental species. The past decade has seen an experimental paradigm shift in the study of pigeon homing since the development of micro-GPS logging devices small enough to be carried by a bird in flight [1, 2]. The use of these devices has enabled researchers to obtain very high spatial and temporal resolution data about the bird’s position during flight and has revealed hitherto unsuspected phenomena. These include the propensity to follow roads and other strong linear features in the landscape [2, 3, 4, 5] and the tendency to form idiosyncratic habitual routes back to the loft when released repeatedly from a single site [3, 6].

Experiments suggest that pigeons have a very robust loyalty to their habitual routes once formed. Pigeons displaced up to 1.5 km perpendicular from their habitual route before release are observed to recapitulate the established habitual route, rapidly rejoining the original path before returning home along their habitual route [3]. This implies a non-compass based orientation since a compass bearing from the new location would direct them toward the home loft rather than directly back to the habitual route. The habitual route has also been shown to be robust under manipulation of the compass. Birds whose sun-compass mechanism has been disturbed by using clock-shifting techniques will perceive a effective ‘compass’ that is rotated relative to reality. Birds navigating
by a clock-shifted compass should fly at a displaced bearing to their non-clock-shifted flights. However, experiments show that clock-shifted birds are able to faithfully follow the habitual route they have previously formed with only minor average displacements [7].

Loyalty to the habitual route has also been observed and exploited in the study of group behaviour [8]. Biro et al. trained individual pigeons to return home from a fixed release site. These individuals were then released in pairs in a series of experiments. Comparison of the recorded flight paths during paired flight with the individual training flights revealed a clear pattern between the instinctive desire to remain as a pair and the conflicting loyalty to an idiosyncratic habitual route, dissimilar to the partner’s. When conflict is too great between these aims the birds were found to make a decisive choice between remaining with the partner and abandoning their habitual route or remaining on the route and leaving the partner. This is further evidence against the use of the compass in the familiar area since individuals relying on a compass bearing to the home loft should be broadly in agreement in their choice of direction and should be able to improve their bearing through information sharing.

These discoveries are persuasive evidence that orientation in the familiar area is controlled principally by visual recognition. Pilotage has been posited as the most likely homing mechanism [3], implying a memorised route between the release site and the home loft encoded by the locations of fixed geographical waypoints. These waypoints are likely to be associated with recognisable visual features – landmarks. Although several studies have observed pigeons following particular features [3, 4] further understanding of how information from the landscape controls flight behaviour has remained out of reach because experimental manipulation of landmarks on such a scale is not practical. Determining the locations of these waypoints in an objective manner is a crucial step towards discovering the visual features birds use to orient themselves in familiar environments. We therefore approached the problem of remote sensing visual attention in a free-flying bird by developing an algorithm for detecting those elements of the flight path that are most important in a navigational sense.

In this study we will show how we can use cutting edge statistical techniques to quantitatively predict the future flight paths of a trained individual bird based on observations of its previous flights. By isolating the points during the past flights that allow for the best predictions of future flights we will demonstrate an objective algorithm for automatically detecting navigational waypoints. Moreover we will show that the model of individual flight can be extended to predict the distinctive behaviour of pair-released birds. Using this extended model we will show that the relationship between conflicting aims, divergence, dominance and compromise can be accurately predicted by considering only the past flight behaviour of each bird as an individual.

To construct an appropriate model for this we will use a mathematical toolset known as Gaussian Processes.
Gaussian Processes

Gaussian Processes (GPs) are a powerful and flexible framework for performing inference over functions, \( f(\cdot) \) [9]. The distribution is specified by a mean function, \( m(\cdot) \), and a covariance kernel, \( k(\cdot, \cdot) \). Any finite number of function values, \( f(t) \), evaluated at a set of inputs, \( t \), has a multi-variate Gaussian distribution.

\[
f(t) \sim N(m(t), k(t, t)).
\]  

(1)

Where \( k(t, t) \) indicates the matrix evaluation of \( k(t, t') \) for all pairs of input values. The prior mean function is chosen to accurately represent the prior knowledge of the function (typically being specified by a symmetry in the model). The covariance kernel is chosen to represent prior beliefs about the dynamical structure of the function – how the function values change with varying inputs. In this paper we will only consider the following stationary Matérn covariance,

\[
k(\Delta t) \equiv \lambda^2 \left( 1 + \sqrt{3 \Delta t} \right) \exp \left( -\sqrt{3 \Delta t} \right).
\]  

(2)

The kernel, \( k(\Delta t) \), represents the strength of correlation between function values for inputs separated by \( \Delta t \equiv |t - t'| \). It is a decreasing function of \( \Delta t \) such that closely spaced values of the function are highly correlated and widely spaced values are roughly independent. The adjustable parameters, \( \lambda \) and \( \sigma \) specify the output scale – the scale of variation of the function values – and the input scale – the characteristic separation between inputs over which the function values become uncorrelated – respectively.

So far we have specified a prior distribution over the function before we make any observations. Now assume that we have observations, \( f_D \), at inputs, \( t_D \). Furthermore we are interested in making predictions about the value of the function, \( f_* \), at inputs \( t_* \). The posterior distribution of these values is given as

\[
f_* | f_D, t_D \sim N(\mu, C),
\]  

(3)

with updated mean and covariance matrices:

\[
\mu = m(t_*) + k(t_*, t_D) k(t_D, t_D)^{-1} (f_D - m(t_D))
\]  

(4)

\[
C = k(t_*, t_*) - k(t_*, t_D) k(t_D, t_D)^{-1} k(t_D, t_*).
\]  

(5)

MODEL

A pigeon’s loyalty to its habitual route makes it predictable. We suggest that observed flight trajectories represent imperfect attempts to replicate an unseen and never seen idealised habitual route. Moreover in this study we assume that variation around the idealised habitual route is non-predictable and simply represents correlated noise. Therefore we aim to learn about the structure of the underlying habitual route and the scale of variation around it from these imperfect observations. Each flight trajectory, \( x_i(t) \), is a
two-dimensional function of time. In our model an observed flight trajectory, \( x_i \equiv x_i(t_i) \), represents a sample from a Gaussian Process, with observations at times, \( t_i \), with a mean, \( h(t_i) \), that represents the habitual route, and a covariance, \( k_\theta(t_i, t'_i) \), that determines the scale of variation around the habitual route and the smoothness of the trajectory (parametrised by \( \phi \equiv \{ \lambda, \sigma \} \)). Multiple trajectories may be assumed to be generated from a common idealised habitual path – mathematically this means they are identically and independently distributed from this Gaussian Process, sharing a common mean function, \( h(t) \), representing the idealised habitual route. As these are real observations they may also be subject to a level of ‘white’ Gaussian observation noise, \( \eta \).

\[
x_i | h(t), \phi, \eta \sim N (h(t_i), k_\phi(t_i, t_i) + \eta^2 \delta(t_i, t_i))
\]  

(6)

We place a Gaussian Process prior over the value of \( h(t) \). This is centred on the straight beeline, \( s(t) \), between the release site and the loft to which the bird will return and has its own dynamical structure parametrised by the covariance kernel, \( k_\theta(t, t) \), with dynamical parameters, \( \theta \). This is an unobserved process and thus includes no observation noise.

\[
h(t)|s(t), \theta \sim GP (s(t), k_\theta(t, t')) .
\]  

(7)

Since the habitual path is never observed we integrate over all possible values to obtain a distribution over sets of trajectories sharing a common, unknown \( h(t) \). We drop the explicit dependence on parameters \( \theta, \phi \) and \( \eta \) for simplicity.

\[
\begin{bmatrix}
x_1 \\
\vdots \\
x_n
\end{bmatrix}
\sim N \left( \begin{bmatrix} s(t_1) \\
\vdots \\
s(t_n) \end{bmatrix}, \Sigma \right).
\]  

(8)

With a combined covariance \( \Sigma \):

\[
\Sigma = \begin{bmatrix}
k_\phi(t_1, t_1) + k_\theta(t_1, t_1) + \eta^2 \delta(t_1, t_1) & k_\theta(t_1, t_2) & \cdots & k_\theta(t_1, t_n) \\
k_\theta(t_2, t_1) & k_\theta(t_2, t_2) & \ddots & k_\theta(t_2, t_n) \\
\vdots & \vdots & \ddots & k_\theta(t_n, t_{n-1}) \\
k_\theta(t_n, t_1) & k_\theta(t_n, t_2) & \cdots & k_\theta(t_n, t_n) + \eta^2 \delta(t_n, t'_n)
\end{bmatrix}
\]  

(9)

The distribution over the mean can then be obtained by application of Bayes’ Rule. Let \( t \) be any chosen vector of time inputs.

\[
h(t)| \begin{bmatrix}
x_1 \\
\vdots \\
x_n
\end{bmatrix}
\sim N (m_h(t), \Sigma_h(t, t))
\]  

(10)

\[
m_h(t) = s(t) + k_\theta \left( t, \begin{bmatrix} t_1 \\
\vdots \\
t_n \end{bmatrix} \right) \Sigma^{-1} \begin{bmatrix} x_1 \\
\vdots \\
x_n \end{bmatrix}
\]  

(11)

\[
\Sigma_h(t, t) = k_\theta(t, t) - k_\theta \left( t, \begin{bmatrix} t_1 \\
\vdots \\
t_n \end{bmatrix} \right) \Sigma^{-1} k_\theta \left( \begin{bmatrix} t_1 \\
\vdots \\
t_n \end{bmatrix}, t \right)
\]  

(12)
We can also calculate the distribution of future flights

\[
x_{n+1}(t) | \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \sim N \left( m_h(t), \Sigma_h(t,t) + k \phi(t,t) + \eta^2 \delta(t,t) \right)
\]  

(13)

Note that we are free to choose any set of time-location triplets where we have observations for each flight. Equation 13 will allow us to test the predictions made based on a putative set of waypoints against the subsequent experimental data.

We use Bayesian marginalisation (through Monte Carlo integration [10]) for an honest propagation of the uncertainty associated with the hyperparameters, \(\phi\), \(\theta\) and \(\eta\). Ten thousand samples were generated from the prior distribution of the GP hyperparameters. Model learning consisted of calculating the likelihood of each hyperparameter sample based on the training data (past flights) from equation 8. The appropriately weighted samples were then used to numerically integrate over the hyperparameters in calculating the conditional probability of the test data (future flights) from equation 13 and the posterior distribution of the habitual route from equation 10.

**RESULTS**

First we establish that our model is capable of accurate prediction. We collated previous collected data [6, 11] from 31 birds during training flights from four distinct sites around the Oxford Field Station. Each bird was released 20 times from its selected release site and its flight home recorded using a GPS logger. Using equation 13 we took each flight and used it to predict its immediate successor. We compared this with the prior probability of the successor path to define a metric of predictability using the Marginal Information Gain (MIG),

\[
MIG = \log p(x_{i+1} | x_i, M) - \log p(x_{i+1} | M)
\]  

(14)

Where \(M\) represents our model. Values of MIG above zero indicate predictable behaviour - the flight is more likely in the light of observations than it was a priori. Figure 1 (left) shows the MIG averaged over the 31 birds as a function of the flight number being predicted. The clear trend is for increasing predictability which corresponds to our belief that the bird is initially naive and forms a memorised route. Figure 1 (right) shows that this is accompanied by a reduction in the corresponding uncertainty on the habitual route inferred from successive pairs of flights as defined by equation 12. These results demonstrate that after twenty releases flight paths are predictable. We use a forward selection algorithm to determine an optimal subset of previous observations, using the observations at the same points in time from each previous flight. The forward selection aims to pick a subset of observations that maximises the marginal likelihood of the set of paths, using the same formalism as the MIG criterion in Equation 7. Let \(x_i^m\) refer to \(m\) observations of flight \(i\) at times \(t^m\). At each iteration we add another observation to the subset to maximise MIG

\[
MIG^m = \sum_i \log p(x_i | x_i^m, M) - \log p(x_i | M)
\]  

(15)
Figure 1. Left: Marginal Information Gain in predicting each flight path from its immediate predecessor. Right: Average (RMS) spatial uncertainty for the location of the idealised habitual route, as inferred from consecutive pairs of flights.

Where $\bar{i}$ indicates all considered paths except path $i$. We permute the calculation over all paths so as to ensure the minimum impact of outliers. We note here that in this algorithm we 'pre-train' the model, learning the dynamical parameters from the full data set. We are interested in the spatial information each observation can provide, rather than the potential to learn dynamical properties.

The number of waypoints can be pre-specified or one can apply an information criterion with respect to the number of waypoints to automate model order selection. We have elected to show the landmarks selected according to the Bayesian Information Criterion (BIC) [12] which provides a parsimonious model, suggesting we can be very confident about the identified points. We also show the remainder of the first ten identified waypoints along with the value of $\text{MIG}_m$ for comparison. Because we limit possible waypoint locations to be a subset of the observed data and because it is combinatorially impractical to find the global optimum subset we expect the BIC to be somewhat overly conservative in model selection.

Figures 2 and 3 indicate the discovered waypoints from 2 birds released at different sites. The red circles indicate the mean position of the paths at each selected time, whilst the yellow polygons indicate the minimum convex hull enclosing the positions at that time on each path. Where the convex hulls overlap they have been combined for clarity. The size of the hulls is partly indicative of the size of the waypoint. We have subjectively inspected the areas identified and noted any prominent visual features present. Algorithmic image analysis of the identified areas could provide a means of objectively determining the characteristics of visual landmarks in the future.

Inspection of the identified waypoints appears to indicate that man-made features are frequently used as landmarks, both on the smallest scale (a single building) and the much larger scale (an entire village). Other waypoints appear to correspond to distinct environmental boundaries – between fields and forest or at river crossings for example. Here we have shown a pair of case studies to indicate the output of this algorithm. Further work will use this method to begin a systematic comparison of identified waypoints, in terms of location and quantity, between different individuals.
Figure 2. (Left) Identified waypoints based on the final 5 training flights (black line) from a bird released at the Horspath site. White circles indicate points selected according to the Bayesian Information Criterion (BIC). Black circles indicate the remaining first ten selected points. Prominent landscape features near waypoints are labelled. (Right) The log-likelihood as a function of increasing number of selected waypoints. The vertical line indicates the optimum number of waypoints selected by BIC.

Figure 3. (Left) Identified waypoints based on the final 5 training flights (black line) from a bird released at the Bladon Heath site. White circles indicate points selected according to the Bayesian Information Criterion (BIC). Black circles indicate the remaining first ten selected points. Prominent landscape features near waypoints are labelled. (Right) The log-likelihood as a function of increasing number of selected waypoints. The vertical line indicates the optimum number of waypoints selected by BIC.

and between different release sites.

DISCUSSION

The Gaussian Process model we have presented provides an easily extensible and adjustable model for making quantitative predictions about future flight paths based on observations of the past. By providing a probability distribution over flight paths it provides a framework to compare hypotheses by rephrasing the comparison as a model selection problem. In this paper we have shown how this approach can be used to iden-
tify navigational waypoints, selecting both the number of waypoints and their location by maximising the predictive power of a subset of the observed data. This provides an algorithmic and objective mechanism for identifying salient locations which may correspond to visual landmarks.

The waypoints identified in our case studies suggest pigeons use prominent man-made features as an element of their route memorisation. Most identified waypoints correspond to significant discontinuities in the landscape, for example the presence of isolated buildings in a large meadow (Figure 2) or a village in an otherwise agricultural landscape (Figure 3). Further case studies from the Horspath release site (Figure 2) show a consistent absence of waypoints during flight over the urban centre of Oxford, suggesting that man-made features are not intrinsically salient, but that contrast with the surrounding environment is the most important factor in the birds’ selection of memorisable waypoints. This supports work by Wiltschko et al [13] who found that pigeons showed no evidence of habitual route formation over the information rich but low contrast environment of urban Frankfurt (Germany).

This technique could be more widely applied as a general method for locating important regions in animal movement paths, for example in locating and tracking food sources in the ocean through observing foraging trips by seabirds.

ACKNOWLEDGMENTS

The authors would like to thank the EPSRC, the Royal Society, the ALADDIN Project and Microsoft Research for their funding and support.

REFERENCES