

Finding the dimension of slow dynamics in a rhythmic system

Shai Revzen^{1,*} and John M. Guckenheimer²

¹Department of Integrative Biology, University of California, 3060 Valley Life Sciences, Building 3140, Berkeley, CA 94720, USA

²Department of Mathematics, Cornell University, Ithaca, NY 14853, USA

Dynamical systems with asymptotically stable periodic orbits are generic models for rhythmic processes in dissipative physical systems. This paper presents a method for reconstructing the dynamics near a periodic orbit from multivariate time-series data. It is used to test theories about the control of legged locomotion, a context in which time series are short when compared with previous work in nonlinear time-series analysis. The method presented here identifies appropriate dimensions of reduced order models for the deterministic portion of the dynamics. The paper also addresses challenges inherent in identifying dynamical models with data from different individuals.

Keywords: phase; dynamical systems; locomotion; cockroach; floquet theory; eigenvalues

1. INTRODUCTION

This paper stems from an effort to find broadly applicable data-driven methods for constructing and validating reduced order models of animal motion. In 1999, Full & Koditschek [1] presented the *Templates and Anchors Hypotheses* (TAH). They suggest that a low-dimensional attracting invariant manifold¹ (a *template*) is present in any high-dimensional system modelling the detailed dynamics of a specific animal's motion (an *anchor*), and that this same template is the natural mathematical expression of the statement that animals with dissimilar morphologies follow similar gaits. Numerous templates for terrestrial locomotion have been studied, including sagittal plane inverted pendulum and spring–mass templates [4–11] and horizontal plane spring–mass templates [12–14]. These templates describe the centre of mass motion and ground reaction forces of two-, four-, six- and eight-legged animals during steady-state running [15,16] and in response to perturbations [17,18] using physical models for the underlying dynamics. The low dimension of these templates was obtained based on mechanical reasoning. In contrast, our method establishes the presence and dimension of a template directly from data, with no intervening modelling step.

1.1. Oscillators

This section presents some of the mathematical technicalities for the sequel. We represent steady running

*Author for correspondence (shrevzen@umich.edu).

¹An ‘attracting invariant manifold’ is, loosely speaking, a surface to which trajectories of the system are drawn, and from which they do not escape [2,3].

gaits by periodic solutions of an n -dimensional vector field

$$\frac{d}{dt}x = F(x; \mu) \quad (1.1)$$

$$F : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}^n, \quad x \in \mathbb{R}^n \text{ and } \mu \in \mathbb{R}^k,$$

with k parameters μ . A periodic orbit $\vartheta = \{o(t) | t \in \mathbb{R}\}$ with period τ is a solution of this equation that satisfies $o(t + \tau) = o(t)$; it is asymptotically stable if all nearby trajectories approach ϑ as $t \rightarrow \infty$. The set of initial conditions evolving on trajectories with this property is B , the *basin of attraction* of ϑ . In general, both ϑ and B are functions of μ .

We use the *return (Poincaré) map* [2,3] from a cross section back to itself to study the stability of a periodic orbit. A *cross section* Σ is an $n - 1$ -dimensional manifold intersecting the vector field transversely. The return map $\sigma : \Sigma \rightarrow \Sigma$ is defined by setting $\sigma(x)$ to be the first intersection of the trajectory starting at x with Σ . If Σ intersects the periodic orbit ϑ at the single point o_0 , then o_0 is a fixed point of σ ; it is asymptotically stable if and only if ϑ is asymptotically stable. The derivative $D\sigma|_{o_0}$ and the corresponding derivative of the *time- τ* map along trajectories of the vector field are central to our data analysis. When $\Lambda := D\sigma|_{o_0}$ has eigenvalues smaller than one in magnitude, we say that ϑ is an *oscillator*. Oscillators are asymptotically stable.

There is a well-developed theory of *normal forms* for oscillators, based on the concepts of *isochrons* [19] and *Floquet coordinates* (figure 1). Oscillators have isochrons: cross sections that are mapped into themselves by the time- τ -map. For isochrons, the return map and the time τ map coincide; for other cross sections they differ. If x is in the isochron of $o_0 \in \vartheta$, the distance between $x(t)$ and $o_0(t)$ tends to 0 at an exponential

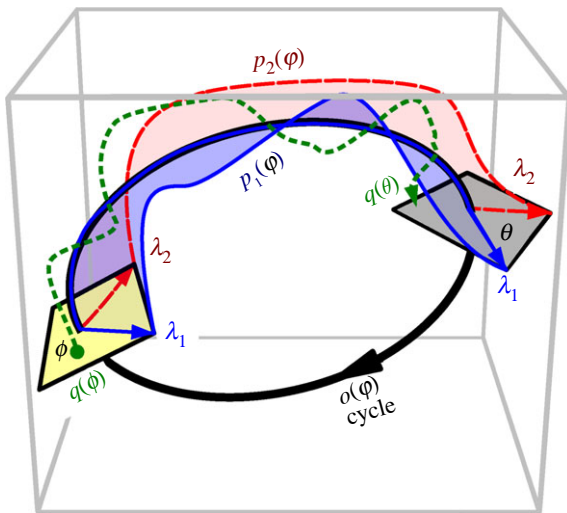


Figure 1. Illustration of Floquet structure. An oscillator is given with a limit cycle (thick dark looping arrow, $o(\varphi)$). Empirically observed trajectories (e.g. $q(\cdot)$, dense dashed arrow, green) cycle around nearby the limit cycle. At any phase ϕ in that cycle, if perturbed in certain directions, the recovery may have a special property similar to that of an eigenvector. For example, on the section representing states with the phase ϕ (square; yellow), a perturbation in the direction of the solid (blue) arrow (marked with λ_1 , representing an eigenvector) will leave the state within the surface whose direction tangent to the cycle is swept by the Floquet axis $p_1(\cdot)$ associated with λ_1 (blue). The perturbed state will return to phase ϕ a cycle later with its distance from the periodic cycle changed by a factor of λ_1 (the eigenvalue, or Floquet multiplier, of the ϕ return map). A similar property holds for the dashed (red) arrow, with respect to the surface of λ_2 (red). At another phase θ (section represented by grey square), each of the Floquet axes maintains the same eigenvalues (λ_1 for solid, blue; λ_2 for dashed, red) but intersects the phase section at a different set of eigenvectors (i.e. the Floquet axis threads through related eigenvectors at all sections). The phase-dependent coordinate frame comprising all Floquet axes defines the Floquet coordinates with respect to which the system takes on the form equations (1.2). Each axis (or pair of axes, in the case of complex conjugate eigenvalues) defines a different, linearly independent invariant surface tangent to the cycle. (Online version in colour.)

rate as $t \rightarrow \infty$. There is a natural parametrization of \mathcal{D} by S^1 that maps $o_0(t)$ to $t/\tau \in S^1 = \mathbb{R}/\mathbb{Z}$ (the circle with circumference 1). The map $\varphi: B \rightarrow S^1$ that sends isochrons to their intersection with \mathcal{D} defines a phase coordinate on the basin of attraction B of \mathcal{D} . We have $\varphi(x(t)) = \varphi(x(0)) + t/\tau$.

We use Floquet theory to make a further change of coordinates that establishes a normal form for an oscillator. If we write the derivative A of the return map of the isochron $o_0 \in \mathcal{D}$ in the form $A = \exp(A)$, then we seek a coordinate system for B , and a rescaling of time, in which the system becomes

$$\text{and } \left. \begin{aligned} \frac{d}{dt} \varphi &= 1 \\ \frac{d}{dt} y &= Ay. \end{aligned} \right\} \quad (1.2)$$

Floquet established the existence of such coordinate systems for the linearized flow near an oscillator [20]. Smooth linearization without approximations is possible

in the absence of resonance conditions on eigenvalues [3]. We ignore the technical difficulties posed by resonances and assume models of the form equations (1.2) for (deterministic) oscillators. We call the associated coordinate systems Floquet coordinates. The eigenvalues of A are the Floquet multipliers. As A is a linear operator, each of its eigenvalues is associated with an invariant subspace spanned by k generalized eigenvectors. Generically, $k = 1$ for each (real) eigenvalue and the action of A reduces to multiplication by the eigenvalue in its one-dimensional eigenspace. The Floquet structure allows each of these eigenvectors to be extended into a Floquet mode—a trajectory of the system, which in addition to being a solution of the ordinary differential equation, has an eigenvector-like property—after one cycle its offset from the limit cycle is scaled by a constant factor equal to the Floquet multiplier.

We give a definition of the template concept in the setting of oscillators as a low(er)-dimensional system compatible with the dynamics of the vector field equation (1.1). Formally, a template consists of an invariant m -dimensional manifold S with an invariant transverse foliation. Stated differently, there is a family of $n - m$ -dimensional manifolds parametrized by S that are mapped into each other by evolving the system t units of time. An oscillator and its isochrons are an example of a template, but we seek templates of intermediate dimensions that represent more of the dynamics and can be used to test the TAH.

Templates of intermediate dimension appear in systems with multiple time scales, for example in fast–slow systems of the form

$$\varepsilon \frac{d}{dt} x = f(x, y, \varepsilon) \quad (1.3)$$

and

$$\frac{d}{dt} y = g(x, y, \varepsilon). \quad (1.4)$$

$\varepsilon > 0$ is assumed small, x is the fast variable and y is the slow variable. These equations possess a template structure comprising a slow manifold and its fast foliation, under the hypothesis that the layer equations $(d/dt)x = f(x, y)$ have negative eigenvalues along the critical manifold C defined by $f = 0$. See the study of Arnold *et al.* [21] for a more complete treatment. In a fast–slow model of animals, the fast attraction of such a template implies that we are only likely to observe animals near the low-dimensional slow manifold (figure 2). This is an example of the posture principle of Full & Koditschek [1].

The theory of normal hyperbolicity establishes the existence of invariant manifolds with fast foliations in more general settings than fast–slow systems. Still, this theory relies upon rate conditions that require convergence rates towards an invariant manifold S be faster than convergence/divergence of trajectories within S . The smoothness of S depends upon the ratio of these rates. Application of these theories to stable periodic orbits yields templates when there is a sufficiently large spectral gap that separates large and small Floquet multipliers of the periodic orbit. Such gaps are not a necessary condition for the existence of

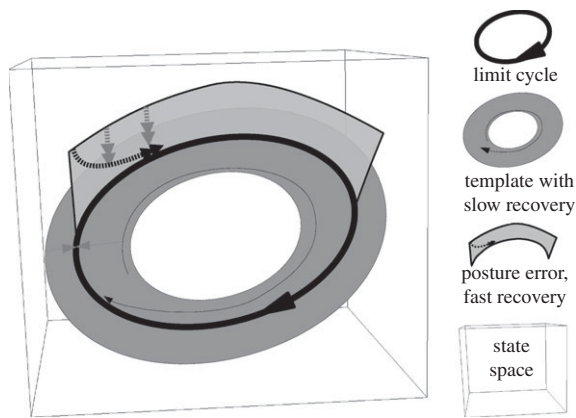


Figure 2. Slow–fast dynamics as a template. The periodic limit cycle is shown as a thick dark loop with an arrow. The template (a slow manifold) is shown as a grey oval band representing a family of trajectories and is the target for all perturbed states. Perturbations (shown by a light grey vertical wall perpendicular to the template band) that generate states that are not part of the template manifold collapse quickly (double arrows, dashed line) to the template. Perturbations that generate valid template states (are on a slow manifold) collapse back to the cycle much more slowly (thin spiral arrow on oval template band).

templates but they do affect their persistence when a system is perturbed [22].

Our approach will focus on the distribution of eigenvalues of periodic orbits in the presence of *measurement noise*: uncertainty owing to measurement accuracy, and *system noise*: uncertainty deriving from unmodelled disturbances applied to the system. Invariant spaces of A in equations (1.2) that are governed by small eigenvalues of A yield empirical trajectory data similar to the noise itself. We regard the *noise dynamics* subspace to be the maximal subspace where this happens. Its complementary subspace contains *significant dynamics*—dynamics that require a model which is not noise-like. We consider the case of significant dynamics governed by a small number of large eigenvalues to be evidence of a template.

1.2. Stochastic oscillators

There is a modest literature on *attractor reconstruction* based upon time-series analysis [23], but these methods do not recover the dynamics *off* of an attractor as we seek to do here. Observation of dynamics off of a periodic orbit requires that the system is perturbed from the periodic orbit time to time, either by an external influence or by stochastic elements within the system. We assume that the motion we observe is stochastic and use a stochastic differential equation (SDE) to model the system. We investigate an SDE with stochastic components that are small enough that the system can be viewed as random perturbations of an underlying deterministic system. Our goal is to reconstruct the characteristics of the underlying deterministic system along with information about the probability distributions of the SDE from typical sample paths drawn from the solution space. This problem occurs much more broadly than the setting of animal locomotion.

Our search for templates has three steps: (i) reconstruct the periodic orbit from time-series data; (ii) generate a

randomized ensemble of estimates of the Jacobian of a return map; and (iii) partition the eigenvalue spectra of the estimated Jacobians into a noise-like part and its complement. We have encountered several mathematical questions that we have been unable to answer rigorously, so we adopted an empirical approach and used numerical experiments to test our methods.

The stochastic models we study assume that the underlying system has an oscillator, and that it has been transformed to Floquet coordinates. We analyse equations of the form

$$\text{and } \left. \begin{aligned} \frac{d}{dt} \varphi &= 1 + \nu_{\varphi} \\ \frac{d}{dt} y &= Ay + \nu_y, \end{aligned} \right\} \quad (1.5)$$

where ν_{φ} and ν_y are Wiener processes whose variances are state-dependent.

Our first task is to compute a stochastic return map for a chosen cross section of the system given by equations (1.5). We expect this return map to take the approximate form $\sigma(y) = Ay + \eta$, where η is a Gaussian random variable and A a matrix. Here, we assume that the coordinate system is centred at the intersection of the cross section with the periodic orbit. Our data analysis seeks to fit σ directly from time-series measurements. However, we also seek to estimate confidence limits on A , so we resort to resampling methods to estimate a distribution of A matrices that plausibly fit the data.

Our null hypothesis was that the A we found is owing to ‘chance alone’, which we take to be: (i) the mean of the distribution of A matrices being the zero matrix and the entries in these matrices being identically distributed (i.i.d) Gaussian random variables;² (ii) no correlation between consecutive intersections of stochastic trajectories with the cross section; (iii) absence of any preferred directions in the covariance of intersection points, i.e. a scalar covariance matrix. This null hypothesis is not a statement about the covariance matrix of η in the model; instead, it expresses limits to our ability to estimate A from the data. We use this simple scalar noise model because of our poor understanding of the interplay between $\text{cov}(\eta)$ and the distribution of A estimates, especially given meager data.

Our task in this paper requires us to produce an estimated spectrum of A from a distribution of estimates of A and to partition that spectrum into template-derived and non-template-derived subsets. Here, we venture into the theory of Gaussian random matrices. Let $\mathfrak{M}_d(s)$ be the space of real-valued matrices of dimension d , whose entries are independently and i.i.d Gaussians $N(0,s)$ with mean 0 and variance s .² Our null hypothesis is that trajectories intersect a cross section at i.i.d Gaussian random points. This leads³ to matrices whose eigenvalue distribution satisfies the *circular law* [24], which states that the eigenvalues of $\mathfrak{M}_d(1/\sqrt{d})$

²Also known as being a *Gaussian random matrix*.

³It is not obvious that given a Gaussian $N(0,1)$ distribution of vector inputs x_i and vector outputs y_i of a map, the distribution of least-squares regression estimates M of the matrix mapping $Mx_i = y_i$ would have the same (up to scale) spectrum as a (real, Gaussian) random matrix. Our numerical experimentation show that this is true at least to a very close approximation.

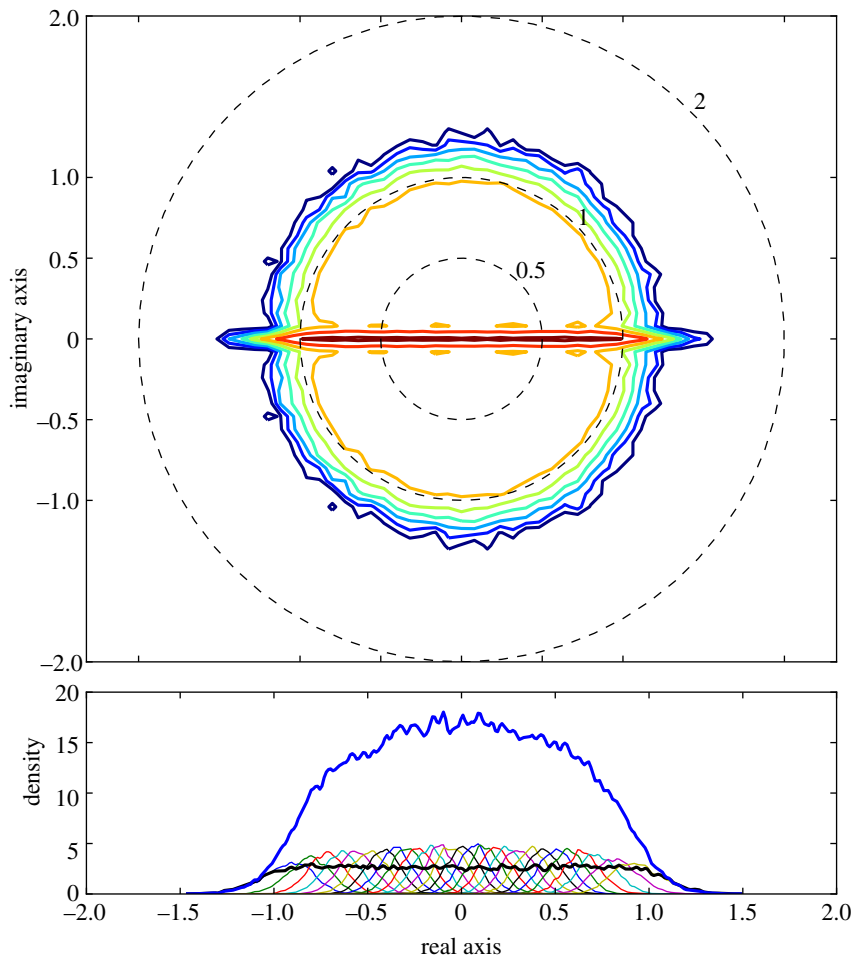


Figure 3. The circular law approximated by sampling 4096 matrices of size 27×27 with entries taken from the standard Gaussian $N(0, 1/\sqrt{27})$. Density contours on the complex plane (a) show that distribution is similar to a uniform density disc of radius 1, except for added mass on the real line. (b) Shows the distribution of real parts of real eigenvalues (lower thick line, black) and the distribution of real parts of all eigenvalues (upper thick line, blue). We also plotted the distribution of the real part of eigenvalues binned by the ordinal position of their real part (thin lines); this shows that eigenvalues are distributed throughout the disc for each matrix, leading the ordinals of the real part to be fairly well-localized on the real line. Each of the bell-shaped curves in (b) shows the density of real parts of a specific ordinal of eigenvalue (e.g. the leftmost shows the distribution of the real parts of the eigenvalue with the smallest real part). (Online version in colour.)

approach the uniform distribution in a complex disc of radius 1 around 0 as $d \rightarrow \infty$.

We expect that invariant subspaces associated with small eigenvalues of A will satisfy the circular law and hope that there will be a complementary subspace that yields a template. We define a *statistically significant template* to be a slow stable manifold whose eigenvalues are inconsistent with the circular law. Our statistical test for a template is thus formulated as a one-sided confidence interval on the spectrum of eigenvalue magnitudes of the linearized return map A . It should be noted that the fact that any given mode is observed to belong to the template, i.e. to exist with a slower time-constant than expected by noise dynamics, does not immediately imply that it plays any role in controlling some variable of interest to the investigator (e.g. centre of mass velocities). Relating the mode to state variable modulations requires recovery of the Floquet multipliers and not only the Floquet coordinates (eigenvalues)—a topic for a future publication.

Figure 3 illustrates the circular law for matrices of the dimension derived from our experimental data

with two-dimensional histograms $\#_{x,y}$ of the eigenvalue points, presented using contours of $\log(1 + \#_{x,y})$. The log-transformation converts Gaussian distributions to paraboloids that are easy to identify visually and thus makes density plots easier to interpret. The eigenvalue distribution shown in the figure (up to scale) is used as one of our null hypothesis in the sequel—it is the eigenvalue distribution of the return map of noise dynamics expected with no template at all. Because of the special role the real line plays in the circular law distribution, we also visualized the density of real parts of the eigenvalues, and the density of real eigenvalues.

1.2.1. The cumulative spectral density

Looking at the isochron at phase ϕ , we could have estimated a single return map matrix from all the available data. However, such a computation would not have provided us with confidence bounds on the matrix and its eigenvalues, and as Hurmuzlu & Basdogan [25] already noted for human gait data, these eigenvalues (the Floquet multipliers) vary between Poincaré sections in

empirical data in a manner counter to what the theory for deterministic and noiseless dynamical systems leads us to expect.

To address these concerns, we obtained $\mathfrak{P}[\phi]$ —the empirical distribution of return map matrices at phase ϕ —by using a resampling method described in detail in the appendix A. We searched for a means of quantitatively characterizing this empirical $\mathfrak{P}[\phi]$ so that we may identify the template dimension and associated invariant subspace of the return map. We required that the method provide consistent answers at sections with different values of ϕ . In both experiment and simulation, we observed that while the density of eigenvalues on the complex plane seems to vary with ϕ , variations in the distribution of eigenvalue magnitudes are much less pronounced.

Given a matrix M , we define $\rho_k(M)$ to be the magnitude of the k th eigenvalue of M , in increasing order. Any distribution of matrices $M \sim \mathfrak{P}$ induces a cumulative distribution $h(k, r, \mathfrak{P})$ that gives the probability $\rho_k(M) < r$ for $M \sim \mathfrak{P}$. We call h the *cumulative spectral density of \mathfrak{P}* . If $h(k, r, \mathfrak{P}) > 1 - p$, the chance of observing any of the first k eigenvalues being larger than r is less than p . Choosing a confidence level p , we invert $h(\cdot)$ obtaining a function $\hat{r}(k, \mathfrak{P})$ (p being understood from the context), such that $h(k, \hat{r}(k, \mathfrak{P}), \mathfrak{P}) = 1 - p$. The function $\hat{r}(k, \mathfrak{P})$ determines the confidence level $1 - p$ in terms of a threshold on the k th eigenvalue magnitude. In other words, with confidence of error less than p , we expect that matrices in the distribution \mathfrak{P} have their k th eigenvalue magnitude smaller than $\hat{r}(k, \mathfrak{P})$. A template with eigenvalues larger than some r_T must, in terms of statistical confidence p , have dimension $n - k$ for any k for which $\hat{r}(k, \mathfrak{P}) < r_T$. We used the functions $\hat{r}(\cdot)$ to provide an upper bound on template dimension.

Our family of null models is constructed from the cumulative spectral densities of $\mathfrak{M}_d(s)$ for dimensions d and scales s , expressing the eigenvalue magnitude spectra of random matrices. We know that for any scale factor β , $h(k, \beta r, \mathfrak{M}_d(\beta)) = h(k, r, \mathfrak{M}_d(1))$, since multiplying a matrix by a scalar scales all eigenvalues of the matrix by that scalar. Similarly, $\hat{r}(k, \mathfrak{M}_d(s)) = s\hat{r}(k, \mathfrak{M}_d(1))$. Thus, these models reduce to the one parameter family $h_d(k, r) = h(k, r, \mathfrak{M}_d(1))$ taken as a function of the parameter d . We abbreviate $\hat{r}_d(k) = \hat{r}(k, \mathfrak{M}_d(1))$.

In terms of cumulative spectral density, the circular law states:

$$\lim_{d \rightarrow \infty} h_d([\beta d], \beta\sqrt{d}) = \begin{cases} \beta & 0 \leq \beta \leq 1 \\ 1 & \beta > 1. \end{cases} \quad (1.6)$$

Taking d to be the dimension of the noise dynamics subspace and \mathfrak{P} to be the sampled distribution of return map matrices, we expect that for $k \leq d$ the empirical $\hat{r}(k, \mathfrak{P})$ and the null hypothesis $\hat{r}_d(k)$ match up to a scale. We developed a test that finds an optimal d for which such a correspondence holds. The complement of this d -dimensional subspace contains dynamics that are unexpectedly slow with respect to the noise model, satisfying our requirements for a statistically significant template.

2. DATASETS

We used simulated data to assess the properties of our algorithm, which was also applied to experimental kinematic data from running cockroaches.

2.1. Stochastic simulation

We constructed an SDE simulation of a system exhibiting a template structure. The system we chose comprised 12 coupled Van Der Pol oscillators (VDPs hereon; see figure 4 for an example trajectory), which are coupled along a bi-directional chain with stronger forward coupling and weaker backward coupling. Our intuition that such a coupling structure would lead to phase-locking and allow us to easily manipulate the larger eigenvalues was born out of the numerical experimentation.

2.1.1. Simulation details

The parameters of the oscillators and their coupling were selected at random and then easily tweaked to achieve phase-locked limit cycle oscillations with three large eigenvalues by allowing strong two-way coupling for the first three oscillators.

The SDE for this system is:

$$\left. \begin{aligned} dx_k &= \mu_k \left(x_k - \frac{1}{3} x_k^3 - y_k \right) dt + dw_k, \\ dy_k &= \mu_k^{-1} x_k dt + dv_k + \sum_j \nu_{j,k} (y_j - y_k) dt \end{aligned} \right\} \text{and } dw, dv \sim N(0, \eta^2), \quad (2.1)$$

where the vector of parameters μ_k gives the standard parameter for VDPs, and governs the convergence rate to the limit cycle in x_k, y_k , when coupling is taken to zero. ν is a tri-diagonal matrix, expressing the coupling between adjacent VDPs (its diagonal elements are ignored). It is best described as two vectors: the forward coupling coefficients ν upper, and the backward coupling coefficients ν lower. Finally, Wiener processes w_k, v_k with isotropic variance η^2 introduce system noise. The values of all of the parameters used for our simulation are in table 1.

As our Poincaré section, we used the positive zero crossing of the mean of y_k for all 12 VDPs.

2.2. Cockroach data

We chose cockroaches as our study animal because template models have been developed for cockroach running (Holmes *et al.* [14] and references therein). We hope that by providing a numerical means for obtaining a template, we provide an empirical anchor-point to which existing and new theoretical models may be related. Cockroaches are large, easy to raise and state information in the form of leg posture is readily measured, suggesting cockroaches as a natural target for predictive models based on leg state.

We used two datasets collected with different experimental systems. Most of our analysis was done on kinematic data of cockroaches running on a transparent treadmill, and photographed from below through

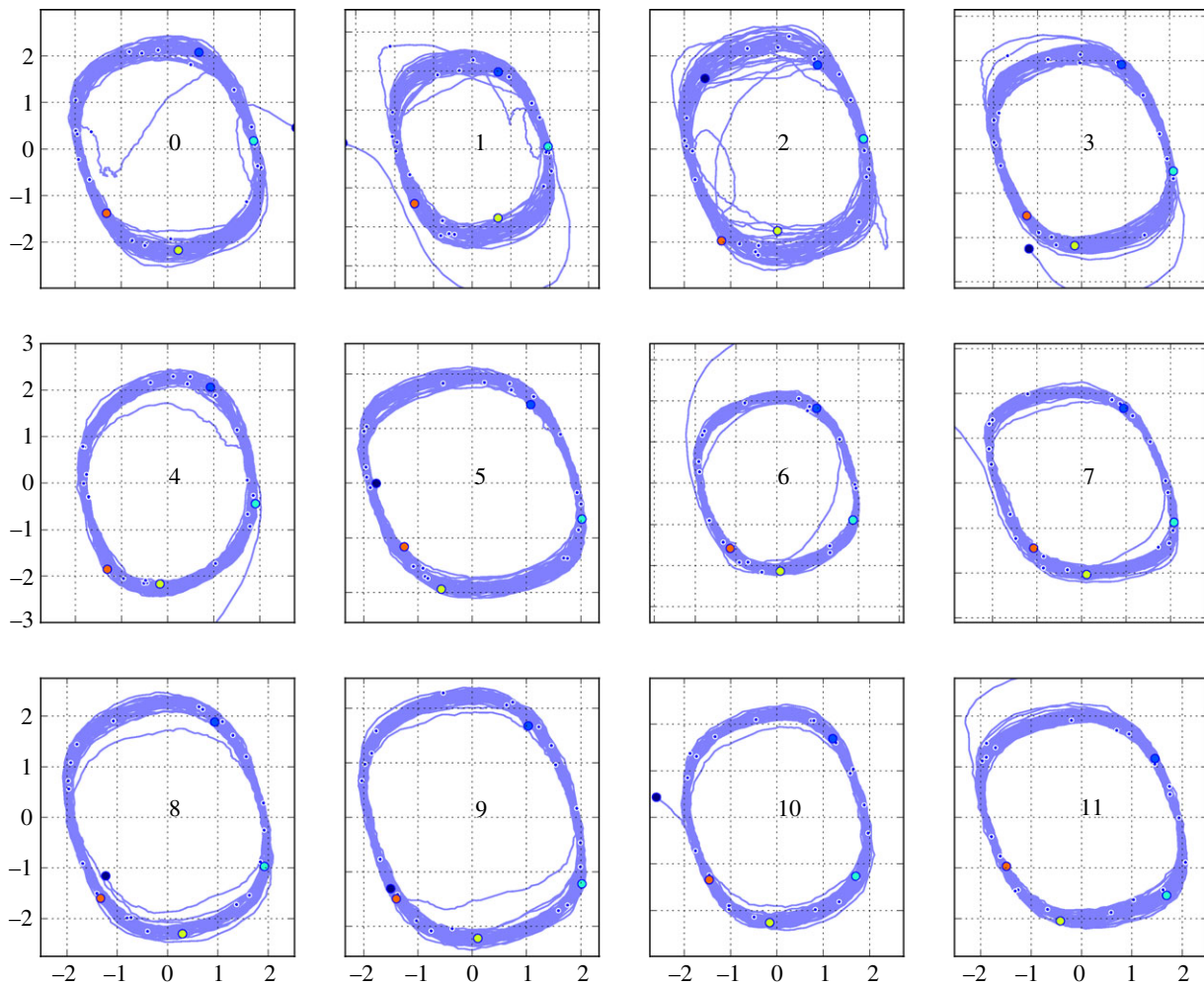


Figure 4. State-space trajectory of our SDE simulation. Each sub-figure plots one of the Van Der Pol oscillators. Circles and dots indicate corresponding times on the different trajectories. (Online version in colour.)

Table 1. Parameter values for the stochastic differential equation simulation.

μ	-0.019	-0.022	-0.063	0.005	-0.215	0.159
	-0.043	0.221	-0.076	-0.04	-0.113	0.097
ν upper	0.97	0.93	0.77	0.39	0.7	0.08
	0.71	0.58	0.73	0.65	0.42	
ν lower	-1	-0.925	-0.445	-0.0085	-0.011	-0.0075
	-0.0405	-0.0435	-0.0445	-0.0075	-0.028	

the treadmill belt [26,27]. Animals ran in the transparent arena of size 80×200 cm, while a robot tracked their movements using a high-speed camera with a zoom lens.

We used 34 adult *Blaberus discoidalis* cockroaches of both sexes (mass 3.3 ± 0.34 g; body length 49 ± 2.6 mm) in the treadmill experiment. The dataset consists of 45 132 frames of 500 fps video, each with body position and orientation, and tarsal claw (tip of the foot) position for all six legs in two dimensions. These data contain about 800 cycles of running, tracking six legs and two body-axis points, giving a 16-dimensional time-series of measurements $y_k^{(n)}$. The data were converted to a 27-dimensional series of state estimates $x_k^{(n)}$ comprising positions and velocities of tarsi in the

body frame ($6 \times 2 \times 2 = 24$ dimensions), velocity of the centre of mass (two dimensions) and angular velocity of the body in the world frame (one dimension).

We were concerned that results from aggregated data from multiple individuals may not be representative of a single animal because of the potential for different values of the μ in equation (1.1) interacting adversely with the *unbalanced experimental design* inherent in differences of trial length. The problem with such a design is that the sample variance of estimates based on pooled data from different animals may not decrease with increasing amounts of data even when the underlying distributions do have convergent second moments. This results from a perennial problem in biostatistics—the problem of controlling for *individual variation*.

A common approach to dealing with individual variation in a statistic is to generate one such statistic per individual, and then combine the individual statistics. This is often done with the sample mean of a normal distribution where each individual's measurements have Gaussian measurement errors: given different numbers of samples from different individuals, the sample mean of each sample is an unbiased estimator of the individual's value, and the mean of the means is an unbiased estimator of the sought-after population mean.⁴ Such an approach was impossible in our case because the statistic we wish to obtain is a return-map estimate, which cannot be computed from short trials at all.

To address this limitation, we developed a bootstrap-like approach for constructing a balanced (re-)sample from the data which can then be used to estimate a return map. The approach is a *multiple imputation* method. Multiple imputation is an approach for dealing with incomplete data in statistical samples using a statistical model. One of its advantages is that at the same time as it addresses the incomplete data, multiple imputation also provides bounds on the effects it introduces into the result. The main innovation in our multiple imputation approach is that the statistical model we use is itself non-parametric, and thus makes weak assumptions about the underlying distributions that are being simulated. Details of our multiple imputation approach are provided in appendix A.

Because of the novel nature of the statistical tools employed in the treadmill experiment, we developed an alternative experimental system—the arena—for validation and control. The arena allowed us to collect enough data from a single individual animal to estimate a return map. We hoped that by finding a return map similar to that found through the multiple imputation procedure, we would support the notion that the statistical methods we employed on the treadmill did not grossly distort our results.

For the arena experiment, we used three adult cockroaches (mass 3.1 ± 0.12 g; body length 47 ± 0.8 mm). After checking that the arena kinematics are typical with respect to previously collected treadmill data, we selected the individual with the most available data for further analysis. Arena data for that individual comprised 15 trials totaling 8565 frames with over 280 cycles.

3. ALGORITHM

This section describes an idealized Floquet analysis algorithm and modifications we made in applying it to our simulation and cockroach data.

3.1. Idealized algorithm

The idealized algorithm outlined in algorithm 1 is suited to modelling equation (1.1) with very large datasets having small measurement noise and the same constant μ for all trials. It assumes that data consist of N trials (unrelated to n of the previous section), each with T

⁴We also draw the reader's attention to the fact that the confidence intervals for the population mean obtained in this way are somewhat pessimistic, as the individual measurement noise can 'leak in' through the smaller samples.

Algorithm 1. Estimating template dimension from measurements y (idealized version)

- 1: $\hat{x} \leftarrow N \times T$ state estimates of dimension d_{state} derived from $N \times T$ values y of dimension d_{measured}
- 2: $\hat{\phi} \leftarrow N \times T$ phase estimates derived from \hat{x}
- 3: $\hat{\omega} \leftarrow$ estimate of the cycle $\hat{\omega}(\hat{\phi})$ derived by taking \hat{x} as a Fourier series in $\hat{\phi}$
- 4: choose phase ϕ for use as a Poincaré section
- 5: interpolate \hat{x} to obtain the $N \times K$ values $\hat{x}(\phi)$, e.g. using linear interpolation
- 6: $\hat{q}(\phi) \leftarrow$ offsets from cycle $\hat{x}(\phi) - \hat{\omega}(\phi)$
- 7: $S \leftarrow N \times (K - 1)$ input–output pairs $\langle \hat{q}(\phi), \hat{q}(\phi + 2\pi) \rangle$ of the return map, each of dimension $2 \times d_{\text{state}}$
- 8: **for** as many samples as wanted $j = 1, \dots, J$ **do**
- 9: bootstrap S to obtain \tilde{S}_j , a $N \times (K - 1)$ set of input–output pairs taken with replacement from S
- 10: estimate a return map matrix $\mathfrak{M}_j[\phi, \phi + 2\pi]$ by least-squares regression of \tilde{S}_j
- 11: **end for**
- 12: $\mathfrak{P}[\phi] \leftarrow \{\mathfrak{M}_j[\cdot, \cdot]\}_{j \in J}$ the (empirical) distribution of matrices at the Poincaré section ϕ .
- 13: find maximal dimension d such that the smallest d eigenvalues of $M \sim \mathfrak{P}[\phi]$ follow the circular law, i.e. the distribution of eigenvalues of random matrices $\mathfrak{M}_d(s)$ for some scale s .
- 14: **return** the codimension of d as the statistically significant template dimension

measurements $y_k^{(n)}$, $n \in 1, \dots, N$, $k \in 1, \dots, T$ at equally spaced intervals over exactly $K = T/\tau$ cycles. We assume a method to reconstruct a state estimate $\hat{x}_k^{(n)}$ from $y_k^{(n)}$ whose errors satisfy (i) $\|F(\hat{x}; \mu) - F(x; \mu)\|$ is small when compared with the magnitude of the stochastic term in the evolution of x , and (ii) $\|\hat{x} - x\|$ is small relative to the diameter of the cycle ϑ .

The SDE simulation employed only a single Poincaré section, and thus applied only steps 4 and above, with the assumption that $\hat{\omega}(\phi)$ is the sample mean $\langle \hat{x}(\phi) \rangle$.

3.2. State reconstruction (step 1)

Since we expected our cockroach systems to function as a mechanical oscillator (see Holmes *et al.* [14] for possible models), we computed velocity estimates to accompany the position measurements of the cockroach and used both to represent the complete state of the animal. The velocity estimates were generated in tandem with an improved position estimate by applying a Kalman smoother [28] to the position measurements.

The Kalman smoother requires knowledge of measurement error covariance and system noise covariance. We tuned both these values from a calibration trial that used a dead animal moved by a positioning stage to generate a baseline. We identified the measurement noise by comparing the stage position with the recovered pixel positions. We chose the system noise level to be the scalar value that produced the best fit in the calibration trial, using a nonlinear line search. While such an approach disregards the system noise introduced by animal-produced motions, it accounts

for uncertainty in motions of the body and tarsi introduced by the moving belt.

3.3. Estimating phase (step 2)

In our animal data, the range of motion of individual legs shifts slowly in individual trials. We estimated phase directly from the raw measurements, rather than our state reconstruction. For each trial, we detrended the $\{y_i\}_{i=1}^{n_k}$ time series by subtracting a baseline $\{b_i\}_{i=1}^{n_k}$ obtained by lowpass smoothing the data with a four cycle cut-off order 3 Butterworth filter in a zero-phase lag (forward-backward) configuration [29]. The resulting zero mean position time series were fed into the Phaser phase-estimation algorithm of Revzen & Guckenheimer [30].

Taken as a number in \mathbb{R}/\mathbb{Z} , phase on a stable cycle is defined up to an arbitrary additive constant which can be taken as a choice of which cycle point has zero phase. Phaser uses a ‘zero Poincaré section function’ $\zeta(\cdot)$ to define zero phase consistently across trials with different stable cycles $\mathcal{D}(\mu)$. It takes the zero crossing point $o(t) \in \mathcal{D}(\mu)$ with $\zeta(o(t)) = 0$ increasing to be the (unique) point with phase zero on $\mathcal{D}(\mu)$. We used the sum of the fore-aft positions of the tarsi with alternating signs going around the body, a distinctly oscillatory quantity for the tripod gait adopted by running cockroaches, as our zero Poincaré section function. Phase zero therefore corresponds to mid-stance.

We used the phase estimates produced from Phaser $\{\hat{\varphi}_i^{(k)}\}_{i=1}^{n_k}$ $k = 1, \dots, N$ to generate Poincaré sections and to compute an orbit model.

3.4. Orbit model (step 3)

We constructed the orbit model by taking the \hat{x} time series as functions of phase, and fitting an order 11 Fourier series model of the orbit $\hat{o}(\varphi)$ using trapezoidal integration. The order of the Fourier series was chosen such that incremental changes in the model order produce changes that are an order of magnitude smaller than the measurement errors. It is unclear how o depends on μ , and thus how to compute \hat{o} from the different trials. At one extreme, a single \hat{o} can be computed from the entire dataset, by taking $\hat{o}^{(n)}(\varphi)$ of trial n and averaging over the index n . This implies an assumption that o does not change as a function of the sample values of μ , and such averaging is therefore an estimator of the true o . At the other extreme, a different $\hat{o}^{(n)}(\varphi)$ can be used for the orbit in each trial. This requires each trial to be sufficiently long to allow a good estimate of the orbit to be computed. As neither assumption is valid for our data, we chose an intermediate solution. We shifted and scaled $\hat{o}(\varphi)$ to have the same mean and standard deviation as the $\hat{o}^{(n)}(\varphi)$ fitted to the individual trial, thereby adapting two of the 23 Fourier series coefficients on a trial-by-trial basis.

3.5. Poincaré sections (step 4)

We sectioned a \hat{q} trajectory at a phase ϕ by linearly interpolating between the surrounding samples. When $\hat{\varphi}_j \leq \phi \leq \hat{\varphi}_{j+1}$, $\hat{q}(\phi)$ was computed from \hat{q}_j and \hat{q}_{j+1} .

Algorithm 2. Non-parametric multiple imputation.

- 1: select parameters m and c
- 2: skip all short trials, i.e. any trial k such that $L_k < m/c$
- 3: **for** each trial k with $L_k \geq m/c$ **do**
- 4: let $\hat{Y}^{(k)}$ be an m -tuple of measurements chosen at random, with replacement.
- 5: **end for**
- 6: the concatenation of all $\hat{Y}^{(k)}$ is the required imputed balanced sample.

We computed intersections of the observed trajectories with cross sections at 20 equally spaced phases around the cycle. At 50 samples a cycle, more sections would have introduced spurious correlations between adjacent sections.

3.6. Bootstrap sampling (steps 7 and 9)

If we just computed a return map from the sample S , we would have a single matrix, with no ability to identify which part of this matrix is significant. Instead we bootstrapped S to obtain many \tilde{S}_j using the multiple imputation algorithm 2, allowing multiple return map estimates to be computed in the following steps.

3.7. Linear regression (step 10)

The return map matrix $\mathfrak{M}_j[\phi, \phi + 2\pi]$ maps a $d - 1$ -dimensional Poincaré section to itself, but does so in the d -dimensional state-space coordinates. Because the input (\hat{q}_i^-) and output (\hat{q}_i^+) points used for the regression are restricted to $d - 1$ -dimensional subspaces, these data do not determine the $d \times d$ matrix produced by least-squares linear regression of the output states with respect to the input states in the direction normal to the section surfaces. However, the behaviour of the time- τ map has an eigenvalue 1 with eigenvector along the flow. Moreover, the time- τ map sends the phase-derived Poincaré section of the cockroach data to itself because it is an isochron.

Using principal component analysis (PCA), we obtained a $d \times d - 1$ orthogonal matrix R whose columns span the tangent space to the section. R comprises the basis for all but the least significant component of the PCA. We performed the linear regression solving for \hat{A} such that

$$R^T \hat{q}_i^+ \approx \hat{A} R^T \hat{q}_i^-. \quad (3.1)$$

Then the matrix $M = R \hat{A} R^T$ is an approximation of the linearized time- τ map except that it has an eigenvalue 0 rather than 1 along the vector field. The collection of these matrices was taken to be a sample from the desired distribution of matrices $\mathfrak{P}[\phi]$.

3.8. Empirical distributions of matrices (step 12)

Using the multiple imputation procedure, we generated a sample from $\mathfrak{P}[\phi]$ —the distribution of return maps matrices at section ϕ . The imputed sample size we

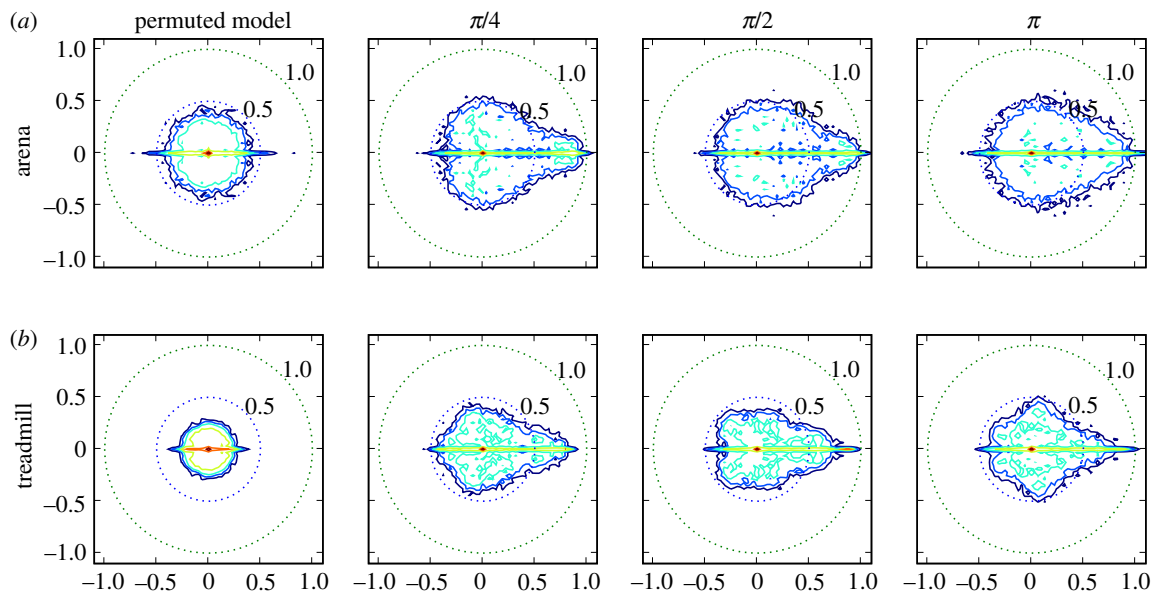


Figure 5. Eigenvalue visualizations for experimentally derived distributions of return map matrices $\mathfrak{P}[\phi]$ at phases $\phi = \pi/4, \pi/2, \pi$ (columns 2,3,4), from two datasets (rows; §2). The visualization shows probability contours for the distribution of eigenvalues obtained from the matrices produced by algorithm 1 step 9; if all matrices generated were identical, the distribution would consist of 27 delta functions. The leftmost column shows eigenvalues for surrogate based control $\mathfrak{P}_S[\phi]$ (see §3.8.1 for details). While these controls seem similar to a circular law distribution, map data show distinct structures with large real eigenvalues. Despite some superficial similarities, the eigenvalue distributions at different phases are clearly distinct. (Online version in colour.)

chose was $N_R = 7290$ (i.e. $27 \times 27 \times 10$), which we felt to be sufficiently large to characterize the local structure of a distribution in the 729-dimensional space of 27×27 size matrices we studied. We tested the resulting distributions to determine whether they gave consistent and predictive models of the return map dynamics.

The eigenvalues of return maps from different sections should have the same eigenvalues, but the density plots in figure 5 show that this property does not hold in practice. The estimated eigenvalue distribution varies from section to section, especially in the eigenvalues of intermediate magnitude. We believe that this variation is partly owing to differences in measurement noise at different phases. The phase (in)dependence of eigenvalue estimates is important, and bears further investigation.

3.8.1. Surrogate data

Following a similar method to the multiple imputation generating $\mathfrak{P}[\phi]$, we generated a distribution of matrices $\mathfrak{P}_S[\phi]$ from *surrogate data* in the sense of Theiler *et al.* [31] and Schreiber & Schmitz [32]. These data consisted of return-map input–output pairs $(\hat{q}_i^+, \hat{q}_{\sigma(i)}^-)$ for a random permutation σ —meaning that each input was associated with an output selected at random from all outputs. Because the permutation destroys the temporal and causal relationships between inputs and outputs, we did not expect the matrices in the $\mathfrak{P}_S[\phi]$ distributions to have a template structure.

3.8.2. Variance reduction

As an additional test of predictive ability of the return maps in $\mathfrak{P}[\phi]$, we examined the distribution of ratios of the variance of the residuals $\hat{q}_i^+ - M \hat{q}_i^-$ with the pre-prediction variance of \hat{q}_i^+ (figure 7). Perfect prediction would make this ratio 0, and simple numerical

experimentation attempting to predict data from a random number generator produces ratios larger than 1. We ran this test on validation data that were not used for the regression of M . Each set of validation data consisted of input–output pairs that were not used in constructing the model; such pairs arise naturally in the multiple imputation algorithm. Thus, the only information the model could have on the validation data comes from the correspondences between the validation data and the training data, and thus through the model being a good representation of a systematic structure in both the training and the validation data.

3.8.3. Density of eigenvalues

The circular distribution for finite-dimensional matrices [24] predicts a mixture of real eigenvalues and complex conjugate pairs. Figure 3 provides insight into how these eigenvalues are distributed for noise dynamics \mathfrak{M}_{27} . We examined the eigenvalues of $\mathfrak{P}[\phi]$ with the same visualization algorithm (figure 7), plotting the distribution of eigenvalues in the complex plane as density contours and the separation of eigenvalue real parts as kernel-smoothed histograms of distinct eigenvalues. The visualization helped guide our investigations; in particular, by illustrating that no obvious spectral gaps were present and that the empirical eigenvalue distributions do depend on ϕ .

3.9. Model selection (Step 13)

The previous steps of the algorithm produced distributions of approximate return map matrices. We wanted to find the largest number d of ‘small’ eigenvalues that span a noise dynamics subspace whose eigenvalue distribution fits that predicted by the circular law. This *model selection problem* is complicated by the unknown scale

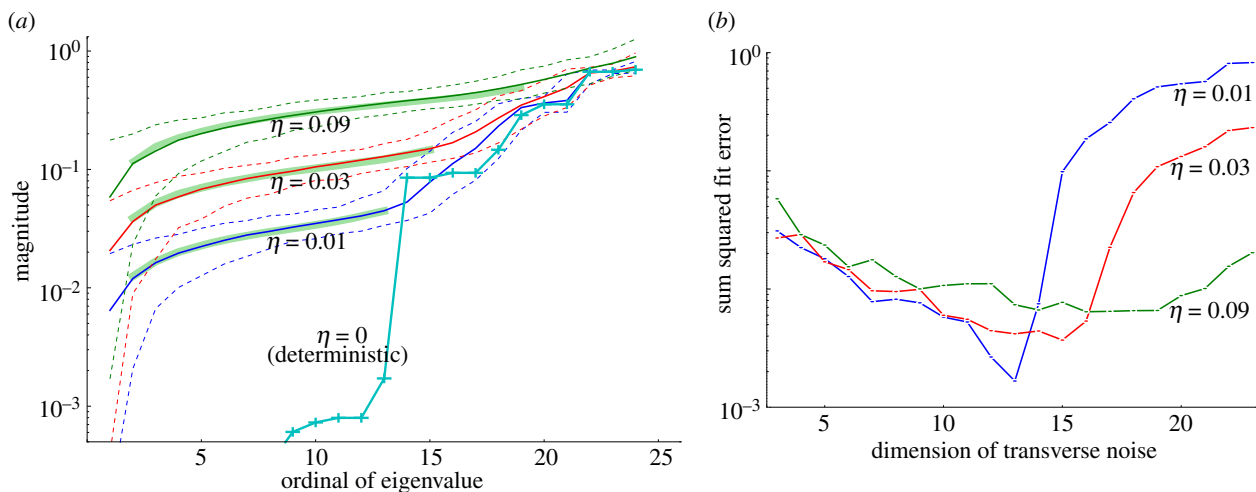


Figure 6. (a) Eigenvalue magnitude distributions and the noise dynamics spectra fit to them for three values of $\eta = 0.01, 0.03, 0.09$. We also plotted the eigenvalue magnitudes of the underlying deterministic system (cross symbols). Each simulation lasted 4086 cycles. For each value of η , we plotted the median (solid) and 1st and 99th percentiles (dashed) of the eigenvalue magnitude distributions obtained from a bootstrap of the input–output pairs of the return map. We also used our model selection criteria to find the dimension and scale of the circular law (noise dynamics) model that best fits each spectrum (thick band, green). In (b), we plotted the sum-of-squares of the fitting error for a circular law model of each dimension (see equation (3.2)). Results show (in (a)) that our model selection criteria identifies a $6 = 5 + 1$ -dimensional template at $\eta = 0.09$, a 10-dimensional one at $\eta = 0.03$ and the full 12-dimensional template at $\eta = 0.01$. The magnitudes of the largest three eigenvalues is estimated correctly at the lower noise levels. From (b), we see that at $\eta \leq 0.03$ the model dimension is very easy to identify. (Online version in colour.)

factor s in the family $\mathfrak{M}_d(s)$ of null models, which would render the natural choice of ‘sum-squared error’ $\sum_k (\hat{r}(k, \mathfrak{P}) - \hat{r}_d(k))^2$ useless, since $\hat{r}_d(k)$ all scale with the unknown s .

We formulated an alternative model selection criterion by constructing a goodness-of-fit statistic Ξ_d that is independent of s .

$$\Xi_d := \sum_{k=1}^{d-1} \xi_d^2(k) \quad (3.2)$$

and

$$\xi_d(k) := (\log \hat{r}(k+1, \mathfrak{P}) - \log \hat{r}(k, \mathfrak{P})) - (\log \hat{r}_d(k+1) - \log \hat{r}_d(k)). \quad (3.3)$$

The values $\xi_d(k)$ do not depend on the s : the magnitudes \hat{r} and \hat{r}_d scale with s , therefore their logarithms have $\log s$ as an additive term that cancels out of the difference sequences. The statistic Ξ_d measures how well the (logarithms of) ratios of magnitudes of consecutive eigenvalues match between \hat{r} and \hat{r}_d .

Our model selection criterion is extremely simple in that it has no information criterion correcting for model order, unlike classical model identification criteria such as the Akaike Information Criterion [33]. Nonetheless, our criterion produced conclusive results when applied to both cockroach data and simulation data.

4. RESULTS

4.1. Stochastic differential equation simulation results

We compared the eigenvalue magnitude distributions obtained from our simulation at various values of η to the eigenvalues obtained when the simulation was deterministic ($\eta = 0$). The results are presented in figure 6.

The spectrum of the time-T map of the deterministic system (marked with crosses) exhibits a clear separation of time scales, with 13 small eigenvalues less than 2×10^{-3} in magnitude, one of which (not shown) is zero to numerical precision and associated with Poincaré section plane itself.

Our simulation results are presented in figure 6. These show that our model selection criteria identifies the signature of large non-random eigenvalues even at levels of noise for which the magnitude spectrum looks smooth to the eye (figure 6a, $\eta = 0.09$). The magnitudes of the largest eigenvalues were correctly identified for $\eta \leq 0.03$, and even at $\eta = 0.09$, the values of the deterministic system fell within the confidence intervals. This suggests that our method would not only identify the presence and dimension of statistically significant dynamical structure but also allow the dynamics of the slowest modes to be estimated.

The plot in figure 6b shows that at moderate levels of noise, the model-fitting error from equation (3.2) shows a rapid increase beyond a certain dimension—thus, the dimension of statistically significant dynamics seems to be uniquely characterized.

From our simulations, we conclude that the method presented here does indeed enable us to detect the presence and dimension of a template in a stochastic system with the prerequisite properties.

4.2. Animal eigenvalue distributions

Our analysis attempted to reconstruct linearized return maps directly from time series. We compared results from our two datasets and from different cross sections when inferring the dimension of the statistically significant template. A preliminary scan of the variance reduction values at different choices of ϕ showed that the overall variance reduction is similar: the norm of

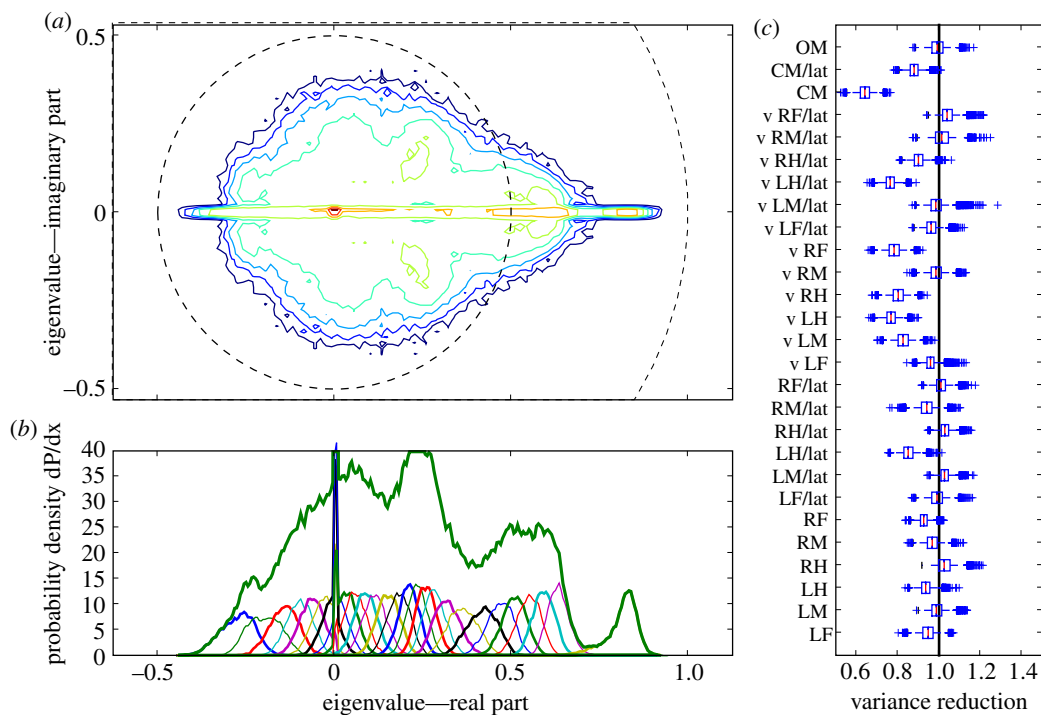


Figure 7. Study of a return map: the eigenvalue distribution on the complex plane (a) shows a distinct lobe on the positive real axis. The distribution of real parts of the eigenvalues (b) shows both individual eigenvalues (sorted by real part) and the total density for all eigenvalues (thick line). From (b), we conclude that the lobe in (a) comprises six or seven eigenvalues, with real parts ranging from 0.4 to 0.9. Tukey box-plots (centre line median; box between lower and upper quartiles; whiskers extend to the nearest point in the upper and lower deciles; crosses show outliers) in (c) show the variance reduction (ratio of residual variance to initial variance, see §3.8.2) that was achieved in each coordinate by using the return map as a predictor. Coordinate labels are a combination of F, M, H for ‘front’, ‘middle’ and ‘hind’; R, L for ‘right’ and ‘left’; v for velocity; /lat for lateral motions; CM for ‘centre of mass’ and OM for ‘omega’, i.e. rotation rate. For example, v LM/lat is the lateral velocity of the left middle tarsus of the cockroach. (Online version in colour.)

residuals was consistently 80 per cent of the norm of the data—suggesting that the return maps have predictive value. However, at certain choices of ϕ , these predictions were exceptionally good for a few coordinates. In particular, this occurred for phases $(\pi/4)(2n + 1)$, corresponding to states where both velocities and positions were equally far from zero—suggesting that better numerical conditioning of the data was a contributing factor to the improvement.

Figure 7 shows an analysis of the return maps $\mathfrak{P}[\pi/4]$ for treadmill data. Comparison of figures 7 and 3 shows distinct structures in the eigenvalue distribution of the return maps that are not noise-like. However, a central disc of eigenvalues with magnitude <0.3 may be noise-like. Figure 5 shows density plots similar to figure 7a at additional phases, for both datasets, for both $\mathfrak{P}[\phi]$ and the $\mathfrak{P}_S[\phi]$ control. The resemblance of the $\mathfrak{P}_S[\phi]$ eigenvalue distribution to the circular law shown in figure 3 is obvious.

4.3. Cumulative spectral density

Figure 8 presents the cumulative spectral densities of our data. It is apparent that the cumulative spectral density plots are similar to one another at different phases. As an example of the interpretation of figure 8, we examine the point at ordinal 20, eigenvalue magnitude 0.6 of the arena data in figure 8b. It implies that when sorted by magnitude, 95 per cent of the 20th

eigenvalues are smaller than 0.6. Thus, (20,0.6) means that only in at most 5 per cent of the imputed return maps do we find more than seven of the 27 return map eigenvalues larger than 0.6. Consequently, the dimension of a template with eigenvalues slower than 0.6 is eight or less, with statistical confidence $p < 0.05$. The template dimension implied in this example is eight and not seven because we must always include the direction of the cycle in the template, and it corresponds to eigenvalue 0.

The more of the dynamics we choose to attribute to the template, the faster the fastest recovery rates in the template will be. The graphs in figure 8 directly represent the trade-off between the dimension of a purported template and the fastest rate of recovery allowed for modes in that template, as computed from return maps at the sections $\pi/4$, $\pi/2$ and π .

The return maps of surrogate data $\mathfrak{P}_S[\phi]$ test the effect of breaking down the relationship between return map initial states and final states. Instead of taking the state of an animal in a particular trajectory as an initial state, and mapping it to the state this same animal had a stride later, a surrogate maps that initial state to a randomly chosen state the animal had at the same phase. The differences of eigenvalue magnitude between $\hat{r}(k, \mathfrak{P})$ (marked coloured lines) and $\hat{r}(k, \mathfrak{P}_S)$ (unmarked coloured lines) are substantial. Surrogate \hat{r} values are less than 1/2 of the corresponding eigenvalue magnitude in the unmodified treadmill data (figure 8a)

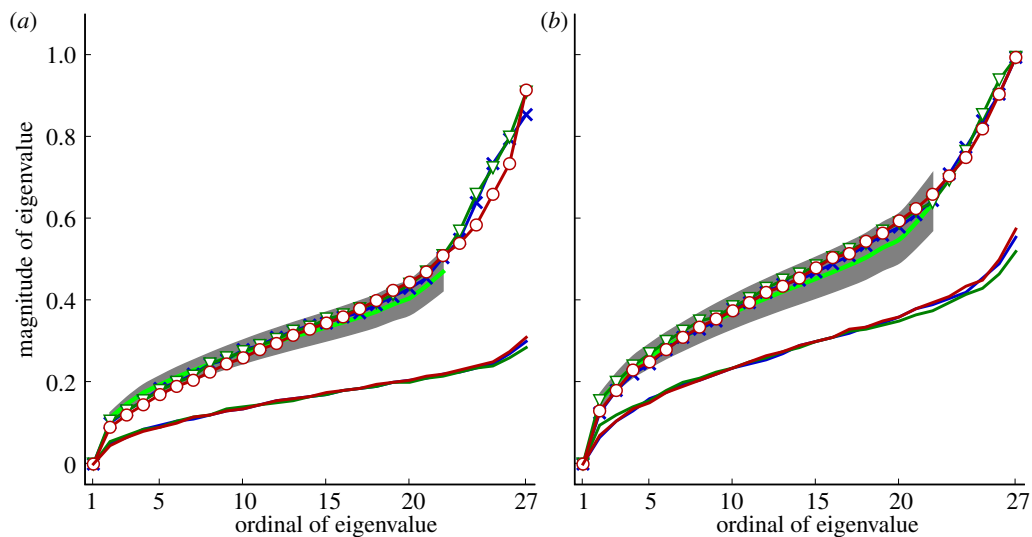


Figure 8. Cumulative spectral density of the empirically obtained distributions of matrices \mathfrak{P} . The abscissa represents eigenvalue ordinal when sorted by magnitude; the ordinate represents eigenvalue magnitude. Lines with markers plot $\hat{r}(k, \mathfrak{P})$ for $p = 0.05$ with k the abscissa; lines without markers plot $\hat{r}(k, \mathfrak{P}_S)$. To demonstrate the consistency of $\hat{r}(k, \mathfrak{P})$ across choices of Poincaré section, we conducted analysis at three phase sections $\pi/4$, $\pi/2$, π (represented by marked line) for both datasets ((a) treadmill and (b) arena)—the same sections as in figure 5. The thick grey band with central light band (green) shows the $r_d(k)$ random matrix null hypothesis model that most closely fit the \hat{r} from the experiment when computed with equation (3.2); the optimal choice was $d = 21$ (figure 9). This 21-value null-model curve runs eigenvalues 2 through 22 because the eigenvalue of 0 at ordinate 1 remaps an eigenvalue equal to 1 associated with the flow direction (see §3.7). The width of the grey band shows a 1 s.d. range around r_d . Our results show that the five largest eigenvalues are not accounted for by the null (random matrix) model; we conclude the template is of dimension six. Lines with cross symbols at $\varphi = \pi/4$; lines with inverted triangles at $\varphi = \pi/2$; lines with open circles at $\varphi = \pi$. (Online version in colour.)

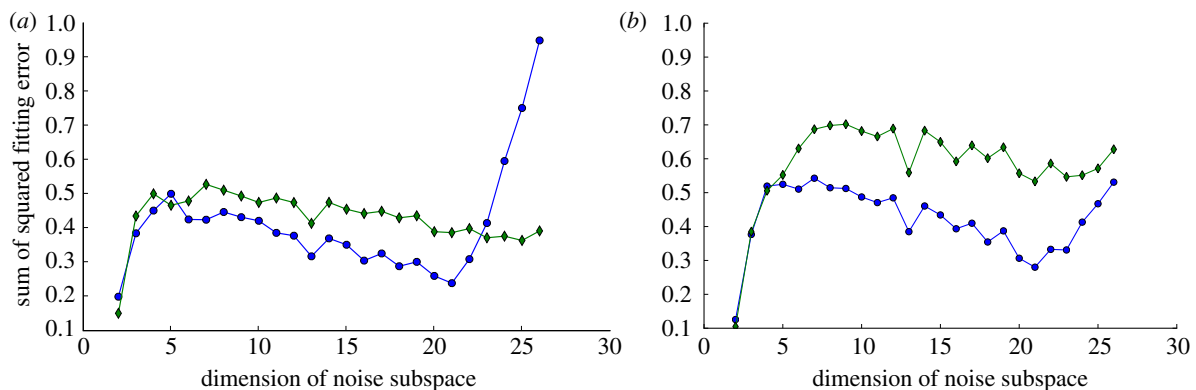


Figure 9. Model selection based on plots of sum-of-squared fitting error (Ξ_d in equation (3.2)) for animal data (circle markers, blue; \mathfrak{P}) and surrogate data (diamond markers, green; \mathfrak{P}_S) for both (a) treadmill and (b) arena. These results show the best-fit noise dynamics dimension is 21, implying that a template contained in the complementary subspace to the noise has a dimension of $6 = 27 - 21$. (Online version in colour.)

and less than 2/3 of the corresponding eigenvalue magnitude in the unmodified arena data (figure 8a). This demonstrates that eigenvalue magnitudes are strongly tied to the causal relationship between the animals' states in consecutive cycles.

4.4. Model selection

We decided the dimension of the template by solving the model selection problem as posed in equation (3.2). We compare the noise spectrum described by \hat{r}_d (see §1.2.1) with the spectrum of the d smallest eigenvalues from the return map estimates, leaving a problem of selecting the best value of d . In terms of figure 8, the

model selection criterion we used consisted of taking the best \hat{r}_d model at each order and computing its fitting error Ξ_d with the animal data (lines with markers), to select the best fit (green line in grey band is \hat{r}_{21} , matching eigenvalues 2 through 22). Merely minimizing a sum-of-squares of the fitting error Ξ_d is typically insufficient because the error grows as models are fit to more and more data—here increasing d to fit a subspace with a larger dimension to a bigger fraction of the measurements—leaving the investigator to decide on some rational means of penalizing models fit to fewer constraints. In our case, the larger d models fit the data so much better than those with small d that Ξ_d decreased with the increase in d

(figure 9), only to sharply increase when the noise subspace dimension grew beyond 21, so the question of model selection criterion was moot.

We conclude that the animal data has a 21-dimensional subspace in which the dynamics are noise-like, and the remaining six dimensions, unaccounted for by the null model, must be attributed to the template. Such a structure is not evident in the surrogates, implying that it is owing to the stride-to-stride causal structure in the animal data.

From figure 8, we obtained an estimated lower bound on the magnitude of the eigenvalues of the template. Rounding to one decimal place, the largest five eigenvalues are larger than 0.5 in 95 per cent of the imputed return maps. The dimension of the noise subspace in surrogate data was that of the full space. We therefore conclude that the six-dimensional dynamics we found represent a causal effect allowing the animals' state to be predicted stride-to-stride.

We observed the same excursion towards larger eigenvalues in the five slowest modes of the arena control data (figure 8*b*) as we did in the treadmill data (figure 8*a*). The similarity between the eigenvalue magnitude distribution found in a single animal (figure 8*b*) and that found from data combined from multiple animals (figure 8*a*) supported the conclusion that the results in figure 8*a* are a consequence of individual animals having such an eigenvalue structure, rather than the observed structure being a computational artefact of inter-animal variability. The similarity of these results is the evidence that our multiple imputation approach does indeed allow us to combine information from multiple subjects in a useful way.

5. DISCUSSION

Comparing our results with those expected from templates such as the Lateral Leg Spring (LLS) [12,13], we observe that our dimension matches the expected dimension of six. However, we observe one key difference—whereas the LLS predicts a strong stabilization of the yaw rate (angular velocity, the time derivative of the yaw angle), we observe little variance reduction in that quantity. It may be that angular velocity is reduced so quickly that we only ever observe system noise in the return map—i.e. this variable is too 'fast' for its dynamics to appear in a statistically significant template as we defined the term. It may also (or alternatively) be that angular velocity does not vary sufficiently to show a change larger than the measurement errors, and thus appears to have no dynamical structure. Finally, it may be that angular velocity changes are not predicted by state variables which we collected. We consider this last hypothesis less plausible, given the mechanical relationship of the predicted and predictor variables—but it cannot be conclusively ignored without further analysis.

More broadly speaking, a reliable method for obtaining reduced dimensional 'template' models of rhythmic systems can have broad utility in many areas. For example, many engineered systems are designed for

periodic operation. The rhythmic motions of legged robots are a specific case where the construction of a data-driven Floquet model along the lines of equations (1.5) provides an approximation of the robot dynamics without requiring a mechanical model. Such an approximation can be used as a target for planning and control, whereby the robot initiates a manoeuvre by transiently perturbing its state away from the limit cycle, in a manner similar to that which has been proposed for cockroach turning [34]. The Floquet modes and the associated eigenvalues provide insight into the forms of persistent instability such a robot may exhibit, and may also be used to validate the robot design. A potential advantage of our approach is that we model the dynamics of the animal or robot in, and with, its environment as one complete system. Because we make no distinction between closed or open loop control, we can capture indirect and emergent consequences of dynamics and passive structures that are often missed when identifying an 'open loop' model and then 'closing the loop' by placing it in the environment. We are currently pursuing an investigation of robot dynamics with data-driven Floquet methods, applied to both centipede-inspired [35] and cockroach-inspired [36] robots. In a follow-up publication, we plan to discuss the reconstruction of Floquet modes.

The trade-offs that govern the effects of trial length, measurement noise, system noise (v_y or equations (1.5)) and phase noise (v_ϕ of equations (1.5)) on the accuracy of eigenvalue estimates are unknown. Our ability to understand the factors that govern the recovery of the Floquet structure of systems is severely limited by lack of such theoretical results. For example, we want to determine the distribution of the spectra of $A + \delta$, $\delta \in \mathfrak{M}_n(\gamma)$ for a matrix A with distinct eigenvalues $\lambda_1, \dots, \lambda_n$. For $\gamma \gg \max|\lambda_j|n^{-1/2}$, the answer is given by the circular law. For γ small enough, the answer is given by sensitivity of the eigenvalues to perturbations. In the intermediate range relevant to many practical problems, little is known.

From a computational standpoint, the most obvious improvement for our current algorithm would be simultaneous estimation of return maps at multiple phases. The variability in eigenvalue estimates obtained from different Poincaré sections suggests that the differences between the sections make eigenvalues susceptible to different biases. A return map estimation method that simultaneously solves for the representation at multiple sections may well be considerably more robust to measurement errors.

This work was funded by NSF Frontiers for Integrative Biology Research (FIBR); grant no. 0425878-Neuromechanical Systems Biology. The authors would like to thank D. E. Koditschek and M. Maus for their comments.

APPENDIX A. MULTIPLE IMPUTATION APPROACH

We faced several obstacles in deriving estimates of return map matrices. The number of cycles in each trial was different, yet we wish all trials to be equally represented in our estimate. This unbalanced trial

length could not be resolved by reducing each trial to a return map estimate and combining the estimates because many of the trials were too short to determine even a single return map matrix. Similarly, we could not reduce the information provided by a trial regarding the return map matrix to some summary statistics because the underlying distribution in the space of matrices is unknown, and has no obvious parametrization.

Consequently, we employed a procedure for constructing a balanced (re-)sample from the data—one for which the longest and shortest trials are not too far apart in length—and then generated large samples of matrices from the data, using the samples themselves to represent the distribution in a non-parametric way. The approach is a form of *multiple imputation*, which is a simulation-based approach to deal with incomplete data [37,38]. In *imputation*, the investigator tries to fill in values for missing measurements of some variable Z , with guesses conditioned on a variable T that is never missing from any measurement. In multiple imputation, the investigator tries multiple possible imputations to allow the effects of imputation on the statistics of interest to be evaluated. It is convenient to define an auxiliary variable M expressing what is sometimes referred to as *missingness*— $M = 1$ when Z is measured and $M = 0$ when it is missing.

By choosing a common length m for trial results, we converted our data to the standard form used in imputation methods. Our data indices had the range $k = 1, \dots, N, i = 1, \dots, m$. The known measurement $T_{i,k}$ was a categorical variable holding the trial identifier. The sometimes missing measurements $Z_{i,k}$, $i = 1, \dots, L_k$ consisted of input–output pairs of the Poincaré section map. $M_{i,k} = 1[i \leq L_k]$ is the indicator variable that is 1 for i values included in the trial and 0 whenever $i > L_k$. We performed a non-parametric imputation step, filling in the missing measurements by resampling the available measurements in each trial until we have m measurements in each.

We departed from the standard approaches used in imputation in two ways. First, we discarded all trials shorter than some minimal length m/c , for some $c > 1$ which we termed the *leverage*. This expresses our belief that with very few measurements, the sample in a given trial cannot adequately reflect the relevant properties of the distribution for that trial, and the entire trial is better discarded. Second, we re-sampled trials longer than m with replacement down to m measurements—an operation akin to the bootstrap. The integrity of the maps was preserved under this resampling because each measurement consisted of a map input and its consequent output. Re-sampling ensured that with enough different iterations of the multiple imputation procedure, every measurement taken in the longer trials contributes to the results. Our multiple imputation procedure is shown in algorithm 2.

The procedure tends to give more weight to measurements from short trials with L_k close to m/c , and does not use all the measurements, and thus under-represents trials with $L_k \gg m$. We chose m by an optimization procedure. First, we selected the leverage c . Leverages larger than e (the natural base) imply that short trials are almost certainly duplicated. We

selected $c = 2$, so as to allow a fairly broad range of trial lengths while at the same time avoiding the near-certain duplication of data points inherent in larger leverages. Armed with the choice of leverage and knowing the length distribution of our trials, we computed the number of input–output pairs that would remain usable for each choice of m , and selected the m that would admit the maximal number of such pairs while still obeying the leverage limit.

REFERENCES

- Full, R. J. & Koditschek, D. E. 1999 Templates and anchors: neuromechanical hypotheses of legged locomotion on land. *J. Exp. Biol.* **202**, 3325–3332.
- Abraham, R. & Marsden, J. E. 1978 *Foundations of mechanics: nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Reading, MA: Addison-Wesley.
- Guckenheimer, J. & Holmes, P. 1983 *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Berlin, Germany: Springer.
- Alexander, R. M. 1988 Why mammals gallop. *Am. Zool.* **28**, 237–245.
- Alexander, R. M. 1989 Elastic mechanisms in the locomotion of vertebrates. *Neth. J. Zool.* **40**, 93–105. (doi:10.1163/156854289X00200)
- Alexander, R. M. 1992 A model of bipedal locomotion on compliant legs. *Phil. Trans. R. Soc. Lond. B* **338**, 189–198. (doi:10.1098/rstb.1992.0138)
- Alexander, R. M. 1995 Leg design and jumping technique for humans, other vertebrates and insects. *Phil. Trans. R. Soc. Lond. B* **347**, 235–248. (doi:10.1098/rstb.1995.0024)
- Blickhan, R. 1989 The spring mass model for running and hopping. *J. Biomech.* **22**, 1217–1227. (doi:10.1016/0021-9290(89)90224-8)
- Cavagna, G. A., Heglund, N. C. & Taylor, C. R. 1977 Mechanical work in terrestrial locomotion—2 basic mechanisms for minimizing energy-expenditure. *Am. J. Physiol.* **233**, R243–R261.
- McGeer, T. 1990 Passive bipedal running. *Proc. R. Soc. Lond. B* **240**, 107–134. (doi:10.1098/rspb.1990.0030)
- McMahon, T. A. & Cheng, G. C. 1990 The mechanics of running: how does stiffness couple with speed. *J. Biomech.* **23**, 65–78. (doi:10.1016/0021-9290(90)90042-2)
- Schmitt, J. & Holmes, P. 2000 Mechanical models for insect locomotion: dynamics and stability in the horizontal plane. I. Theory. *Biol. Cybern.* **83**, 501–515. (doi:10.1007/s004220000181)
- Schmitt, J. & Holmes, P. 2000 Mechanical models for insect locomotion: dynamics and stability in the horizontal plane. II. Application. *Biol. Cybern.* **83**, 517–527. (doi:10.1007/s004220000180)
- Holmes, P., Full, R. J., Koditschek, D. E. & Guckenheimer, J. M. 2006 The dynamics of legged locomotion: models, analyses, and challenges. *SIAM Rev.* **48**, 207–304. (doi:10.1137/S003614450445133)
- Blickhan, R. & Full, R. J. 1993 Similarity in multilegged locomotion: bouncing like a monopode. *J. Comp. Physiol. A* **173**, 509–517. (doi:10.1007/BF00197760)
- Farley, C. T., Glasheen, J. & McMahon, T. A. 1993 Running springs: speed and animal size. *J. Exp. Biol.* **185**, 71–86.
- Jindrich, D. L. & Full, R. J. 2002 Dynamic stabilization of rapid hexapedal locomotion. *J. Exp. Biol.* **205**, 2803–2823.

- 18 Daley, M. A., Usherwood, J. R., Felix, G. & Biewener, A. A. 2006 Running over rough terrain: guinea fowl maintain dynamic stability despite a large unexpected change in substrate height. *J. Exp. Biol.* **209**, 171–187. (doi:10.1242/jeb.01986)
- 19 Guckenheimer, J. 1975 Isochrons and phaseless sets. *J. Math. Biol.* **1**, 259–273. (doi:10.1007/BF01273747)
- 20 Floquet, G. 1883 Sur les équations différentielles linéaires à coefficients périodiques. *Annales Scientifiques de l'École Normale Supérieure, Sér 2*, 12.
- 21 Arnold, L., Mischaikow, K., Raugel, G. & Jones, C. 1995 Geometric singular perturbation theory. In *Dynamical systems*, vol. 1609. Lecture Notes in Mathematics, pp. 44–118. Heidelberg/Berlin, Germany: Springer. (doi:10.1007/BFb0095239)
- 22 Hirsch, M. W., Pugh, C. C. & Shub, M. 1977 *Invariant manifolds*. Lecture Notes in Mathematics. Berlin/New York: Springer.
- 23 Kantz, H. & Schreiber, T. 2004 *Nonlinear time series analysis*, 2nd edn. Cambridge, MA: Cambridge University Press.
- 24 Edelman, A. 1997 The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.* **60**, 203–232.
- 25 Hurmuzlu, Y. & Basdogan, C. 1994 On the measurement of dynamic stability of human locomotion. *J. Biomech. Eng. Trans. ASME* **116**, 30–36. (doi:10.1115/1.2895701)
- 26 Revzen, S., Koditschek, D. E. & Full, R. J. 2006 Testing feedforward control models in rapid running insects using large perturbations (abstract only). *Integr. Comp. Biol.* **46**, e1–e162. (doi:10.1093/icb/icl056)
- 27 Revzen, S. 2009 Neuromechanical control architectures in arthropod locomotion. PhD thesis, Department of Integrative Biology, University of California, Berkeley.
- 28 Briers, M., Doucet, A. & Maskell, S. 2009 *Smoothing algorithms for state-space models*. *Ann. Inst. Stat. Math.* **62**, 61–89. (doi:10.1007/s10463-009-0236-2)
- 29 Gustafsson, F. 1996 Determining the initial states in forward-backward filtering. *Signal Process. IEEE Trans.* **44**, 988–992. (doi:10.1109/78.492552)
- 30 Revzen, S. & Guckenheimer, J. M. 2008 Estimating the phase of synchronized oscillators. *Phys. Rev. E* **78**, 051907. (doi:10.1103/PhysRevE.78.051907)
- 31 Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. 1992 Testing for nonlinearity in time series: the method of surrogate data. *Phys. D Nonlinear Phenom.* **58**, 77–94. (doi:10.1016/0167-2789(92)90102-S)
- 32 Schreiber, T. & Schmitz, A. 2000 Surrogate time series. *Phys. D Nonlinear Phenom.* **142**, 346–382. (doi:10.1016/S0167-2789(00)00043-9)
- 33 Akaike, H. 1974 A new look at the statistical model identification. *IEEE Trans Automatic Control* **19**, 716–723. (doi:10.1109/TAC.1974.1100705)
- 34 Proctor, J. & Holmes, P. J. 2008 Steering by transient destabilization in piecewise-holonomic models of legged locomotion. *Regul. Chaotic Dyn.* **13**, 267–282. (doi:10.1134/S1560354708040047)
- 35 Sastra, J., Bernal-Heredia, W. G., Clark, J. & Yim, M. 2008 A biologically-inspired dynamic legged locomotion with a modular reconfigurable robot. In *ASME Dynamic Systems and Control Conf., Ann Arbor, Michigan, USA, October*. See http://modlab.seas.upenn.edu/publications/2008_DSCC_Centipede.pdf.
- 36 Saranlı, U., Buehler, M. & Koditschek, D. E. 2001 RHex: a simple and highly mobile hexapedal robot. *Int. J. Robot. Res.* **20**, 616–631. (doi:10.1177/02783640122067570)
- 37 Harel, O. & Zhou, X.-H. 2007 Multiple imputation: review of theory, implementation and software. *Stat. Med.* **26**, 3057–3077. (doi:10.1002/sim.2787)
- 38 Efron, B. 1994 Missing data, imputation, and the bootstrap. *J. Am. Stat. Assoc.* **89**, 463–475. See <http://www.jstor.org/stable/2290846>.