Current Opinion in
Neurobiology

Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings

Lea Duncker[1,2] and Maneesh Sahani[1]

Abstract

The question of how the collective activity of neural populations gives rise to complex behaviour is fundamental to neuroscience. At the core of this question lie considerations about how neural circuits can perform computations that enable sensory perception, decision making, and motor control. It is thought that such computations are implemented through the dynamical evolution of distributed activity in recurrent circuits. Thus, identifying dynamical structure in neural population activity is a key challenge towards a better understanding of neural computation. At the same time, interpreting this structure in light of the computation of interest is essential for linking the time-varying activity patterns of the neural population to ongoing computational processes. Here, we review methods that aim to quantify structure in neural population recordings through a dynamical system defined in a low-dimensional latent variable space. We discuss advantages and limitations of different modelling approaches and address future challenges for the field.

Addresses

[1] Gatsby Computational Neuroscience Unit, University College London, London, UK
[2] Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

Corresponding author: Sahani, Maneesh (maneesh@gatsby.ucl.ac.uk)

Introduction

Survival often depends on computations that are explicitly dynamical in structure. Consider navigating to a remembered shelter, integrating ambiguous or noisy cues to detect a predator, or planning and executing a precise movement to catch or avoid becoming prey. In each case, an estimate of current state — reflecting relative location, partial information about the threat, or pose of the body in relation to the target — must be retained and updated to determine the best course of future action.

Dynamical computations, and the associated maintenance of dynamical state, are likely to be supported by the exuberantly recurrent circuitry found in much of the central nervous system. Even rapid responses to brief stimuli may depend in part on recurrent dynamics. One signature of dynamical computation is that it takes time. Although observers can assign a broad category to the visual object that appears in a flashed image with a speed that seems consistent with feedforward processing alone, more detailed categorisation requires longer processing of the same briefly presented stimulus [1]. Indeed, almost all sensory decisions benefit (to a point) from greater processing time, even when the duration of the sensory stimulus itself remains fixed. This is the ubiquitous phenomenon of speed-accuracy tradeoff [2]. The hypothesis that this gain in accuracy results from greater time for dynamical computation is supported by studies of backward masking, in which performance may be limited by 'stopping' such processing prematurely [3].

An emphasis on dynamics in neural computation is increasingly bringing to the fore empirical methods that characterise the temporal evolution of neural activity in recurrent circuits. Many such methods focus on visualisation, for example, using dimensionality reduction techniques, but by themselves, such methods cannot identify the dynamical rule that dictates how the current state and inputs combine to shape the evolution of activity. It is this dynamical rule that embodies the computational algorithm implemented by the network, and thus, it is important for analytic methods to move beyond mere visualisation of temporal patterns to direct statistical identification of dynamics.

Here, we review recent work that seeks to combine the characterisation of a low-dimensional manifold explored by population activity with the estimation of the dynamical rules responsible for generating the relevant activity patterns. We begin with a brief overview of dimensionality reduction, connecting those ideas to a dynamical systems framework for studying neural computation. Next, we explore methods that seek to capture dynamical laws of increasing complexity on the

low-dimensional manifolds underlying population data. We close by considering limitations and opportunities for future work in the field.

Characterising neural activity manifolds

The time-varying activity of a population of neurons can be described in terms of a high-dimensional coordinate system with axes corresponding to the firing rate of each neuron. The collective activity of all neurons in the population determines a single point in this high-dimensional coordinate system — the *population state*. As the activity unfolds over time, the population state traces out a *trajectory* in the high-dimensional space. In some situations, the dimensionality of the space explored by these population trajectories for a given computation may be much smaller than the number of neurons in the circuit that maintains the relevant representation. In this case, computational activity will be confined to a low-dimensional subset or *manifold* contained within this high-dimensional *ambient* space (Figure 1) [4]. Here, each state value corresponds to one (or few) pattern(s) of population activity, and the set of all activity patterns encountered during computation will, up to intrinsic neuronal noise, reflect this dimensionality rather than the size of the network.

Computational manifolds with a range of geometries have been observed in neural activity. These include linear or affine subspaces (Figure 1a) observed in cortical activity before and during deliberate reaching [5–8], decision making [9,10] and motor timing [11,12], where the activity manifold lies almost flat. Nonlinear examples include representations of space in the hippocampus and entorhinal cortex [13–15], such as a ring topology in the head-direction cell system [16,17] (Figure 1b), or more generally curved manifolds [18]. In

other settings, the manifold may be star-shaped (Figure 1c), as when sparse groups of neurons encode different state values [19] or comprise a union of disjoint submanifolds as in the collection of synchronous firing states of a synfire chain [20].

In many of these examples, the dimensionality of the computational manifold is closely related to the dimensionality of an encoded state variable, such as location in two-dimensional or three-dimensional space or head direction. The dynamical computation that updates this state variable in response to self-motion, sensory input, or perhaps as part of memory or action planning will then appear to move activity along the neutrally stable manifold of valid encodings giving rise to a computation that is largely confined within this manifold. More generally, if the complexity of the underlying computation is low — as is often the case in tightly controlled experimental tasks — then only a few computational variables need to be encoded in the circuit, and the dimensionality of the associated activity patterns may hence be low as well [21]. Where representational demands are more elaborate, for instance, when extracting features from a high-dimensional set of natural stimuli, then neural activity reflects these added dimensions [22]. Even in this case, however, the temporal evolution of activity in response to any one image may remain confined to a lower dimensional space [23].

When relevant, the low-dimensional manifold of activity, as well as the trajectories by which activity evolves along that manifold, may be found by applying dimensionality reduction techniques to population data recorded while the computation is performed [24]. In effect, these methods use instantaneous relationships between the activity in different neurons

Figure 1
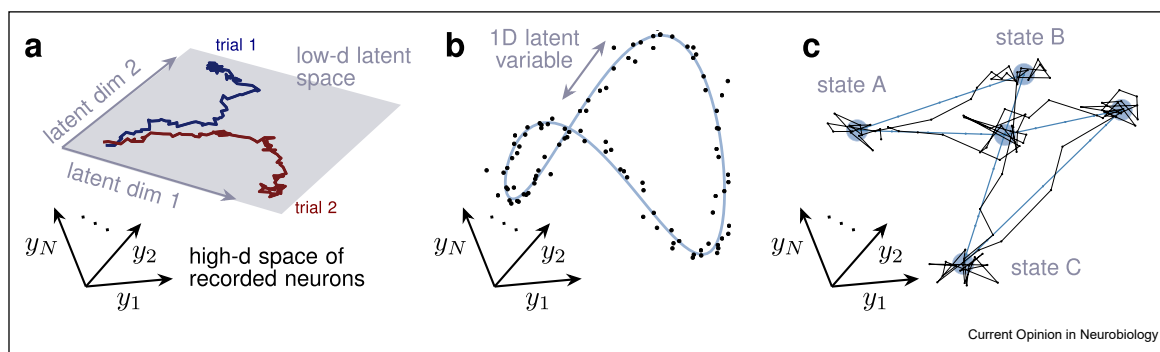
(**a**): Population activity evolves in a high dimensional coordinate space, where each axis corresponds to the firing rate of a recorded unit. Although this space is high-dimensional, neural activity may only explore a low-dimensional subspace. Intertrial differences in task outcome may be reflected in intertrial differences in the evolution of the neural population state over time. Dimensionality reduction techniques can be applied to obtain descriptions of the high-dimensional population activity in terms of a low-dimensional latent variable. (**b**): Schematic illustration of a one-dimensional manifold embedded nonlinearly in the high-dimensional space of recorded neurons. Similar topologies have been identified in the head-direction cell system [16,17]. (**c**): Illustration of a manifold representing different state values, where each state engages different groups of neurons.

[5,16,19,25–28] or between that activity and externally measurable quantities [9,29,30] to estimate the low-dimensional structure in the data. Often, these statistical relationships in observed activity are captured by assuming dependence on a shared but unobserved *latent* variable. In such cases, the latent variable provides a direct estimate of the representation of the underlying computational state. Otherwise, a further step is necessary to establish a coordinate system for the recovered manifold and so provide a variable that represents the state [16].

The manifold of population states is one part of the link between neural circuits and dynamical computation. However, activity manifolds are only the signature of an ongoing computational process and do not by themselves provide a description of the relevant computation. Although at least one recent approach seeks to identify low-dimensional projections most compatible with dynamics [31], a full connection between the observed population activity patterns and the underlying circuit computation depends on explicit identification of the dynamical rule that dictates how neural activity evolves over time.

From data to dynamics to computation

The way in which temporal evolution of population activity implements computation is formalised by viewing the network as a dynamical system [6,32,33], defined by a state and a dynamical law — the system *dynamics* — that governs the future evolution of that state based on its current value and inputs. This dynamics, ultimately implemented by the connections and integration properties of the neurons in the network, then captures the computational transformation of inputs to outputs over time. Its features, such as the location and stability of fixed points or limit cycles, their basins of attraction, and properties of input-driven orbits in the state space, provide insight into the structure of the computation.

Consider, for example, a network involved in a simple perceptual decision-making task, such as reporting the direction of motion of a random dots stimulus [34]. To solve this task, the network would need to update a decision variable representing leftward or rightward motion based on perceptual evidence arriving as inputs to the system. We may hence observe that activity in this network evolves gradually towards a different pattern of firing for each choice made (Figure 1a). Such trajectories

Figure 2

Current Opinion in Neurobiology
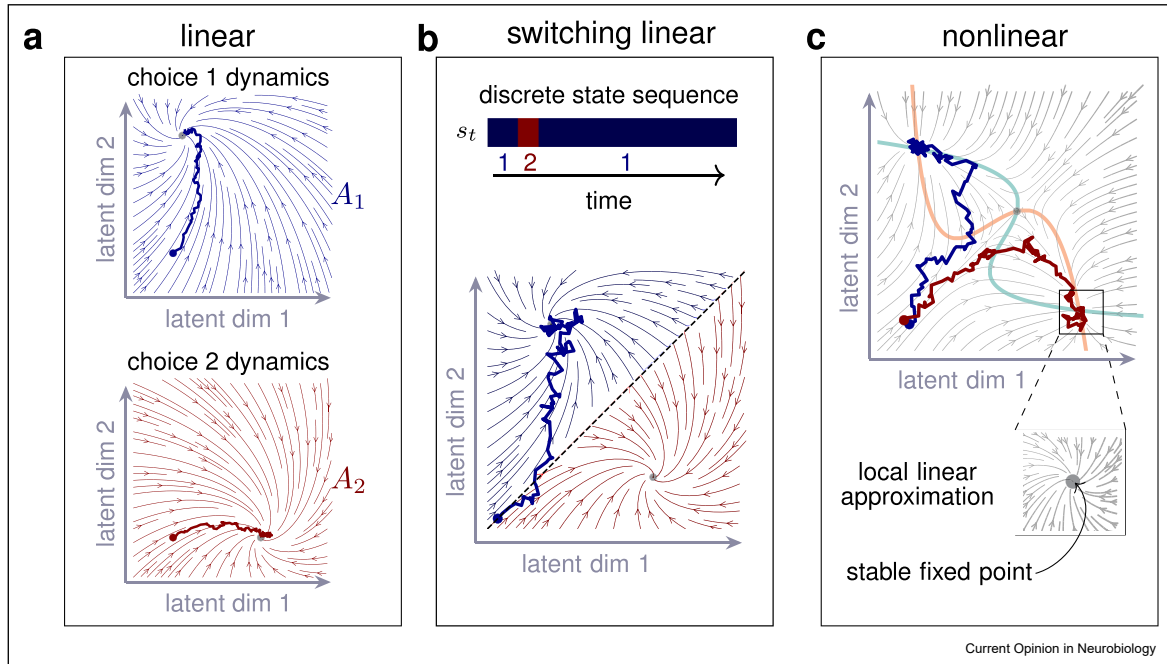
Description of the temporal evolution of the population activity in Figure 1a in terms of different dynamical systems defined on the low-dimensional latent variable. The dynamics are illustrated through a flow field. (**a**): Linear dynamics with a single fixed point fit to trials corresponding to the same choice. (**b**): Switching dynamical system composed of two linear subsystems and a discrete variable indicating the system that dictates the evolution of the latent variable at each time point. When the discrete state variable is dependent on the current value of the latent variable, one subsystem is more likely to be active in a given region of state-space. (**c**): Nonlinear dynamical system with three fixed points. Coloured lines show the system's nullclines, where the temporal derivative of one of the latent dimensions is zero. The path of the latent variables converges to one of the two stable fixed points. The local behaviour of the system can be analysed via linear approximations.

are consistent with a simple dynamical system in which two (or more) attractor states are positioned such that transient inputs that reflect sensory evidence nudge the state from a unstable or marginally stable initial point towards one or the other attractor (Figure 2c). Indeed, from a theoretical point of view, spiking neural network models that achieve such integrative decision making [35] can be reduced to just such a two-dimensional dynamical description in terms of decision-related variables [36] (Figure 3c). In this case, the link between circuit computation and dynamics is made explicit.

Is it possible to extract an analogous description in terms of low-dimensional latent variables directly from data recorded during a perceptual decision making task? Data analysis methods that facilitate such a description could offer ways to refine and test hypotheses around computation in recurrent circuits (Figure 3a), and, thus, represent a path towards a data-driven understanding of the biological computational strategy.

Identifying dynamical systems from data

To study neural computation, we are interested in estimating and analysing the dynamics that govern the temporal evolution of the underlying computational state of the network. A general class of latent dynamical, or *state-space* models, facilitate this through the description of the evolution of neural activity in terms of a low-dimensional latent computational state (here written $x(t)$ or x_t in discretised time) that relates linearly or nonlinearly to recorded neural activity and evolves in time with dynamics of an assumed functional form. Given recorded neural activity, the models can be fit to data to infer a distribution over the path of the latent variable and estimate the dynamics that describes this evolution.

Linear models

The simplest class of latent dynamical models incorporates a linear dynamical system, modelling the evolution of the latent state as

Figure 3
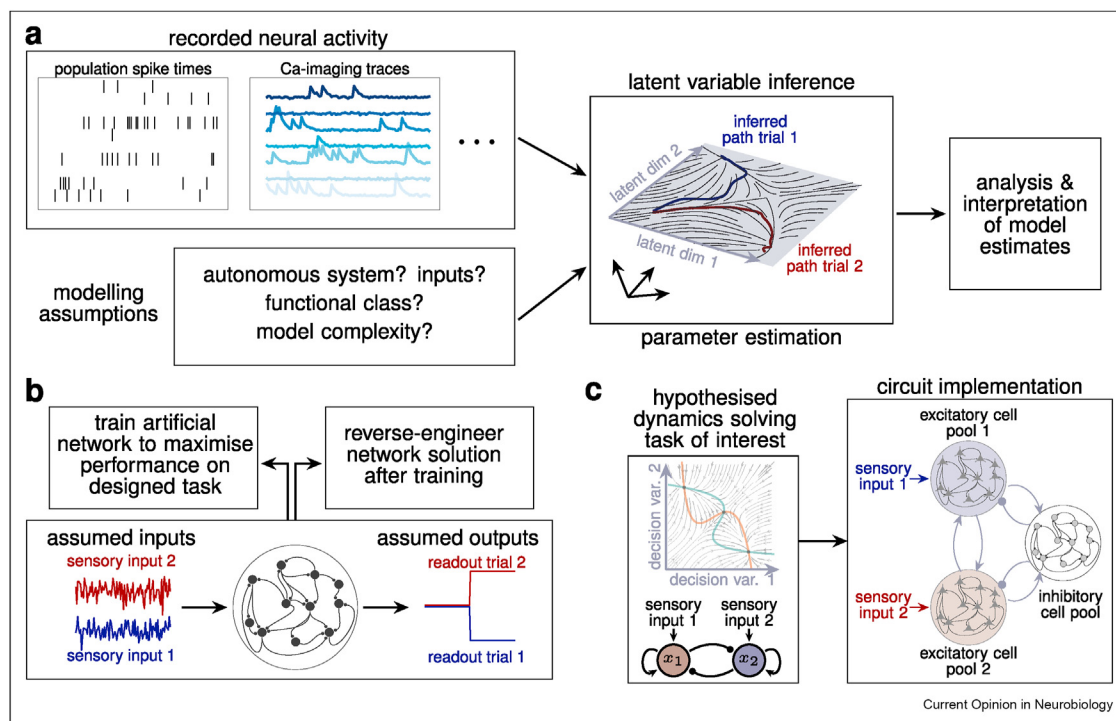
Different approaches towards studying neural population dynamics and computation. (**a**): Data analysis approaches require making assumptions to specify a parametric model of the underlying population dynamics. Different modalities of recorded neural activity can be used for model fitting to obtain a description of low-dimensional structure in the neural population and estimates of the model parameters. The recovered dynamical system can then be analysed in light of the task and computation of interest. (**b**): Task-optimised artificial network models learn a prespecified input−output transformation. Everything about the system is observable and can be analysed in detail to extract insight about the resulting dynamical system solving the task. However, the solution will depend on assumptions about the task and network design, as well as the cost function used for optimisation. (**c**): Mechanistic models are designed to implement a hypothesised solution to a given task, often taking biological constraints like cell-type structure into account. The design is often set up to capture selected qualitative features of neural data.

$$x_{t+1} = Ax_\mathbf{t} + Bu_t + \boldsymbol{\eta}_t \qquad (1)$$

in terms of a dynamics matrix A, inputs $B\mathbf{u}_\mathbf{t}$, and (typically random Gaussian) state innovations $\boldsymbol{\eta}_\mathbf{t}$.

Linear systems are made particularly attractive by the fact that their asymptotic behaviour can be understood in terms of the eigenvalues and eigenvectors of A. For stable systems and in the absence of inputs, activity will eventually decay towards a single fixed point at zero. Depending on the orientation of the eigenvectors of A, the system may also exhibit transient, non-normal amplification on shorter timescales [37,38]. Analysis of control-theoretic objects like controllability and observability Gramians provide further insight into the system's behaviour in response to inputs [39].

Linear dynamical systems have been applied to brain-machine interface control [40] and to the study of the mechanisms underlying contextual evidence integration in prefrontal cortex [41]. They have also provided a theoretical framework within which to understand features of neural responses during motor preparation and execution [42,43], in primary visual cortex [38] and in working memory tasks [37]. These applications have provided insight, although linear dynamical transformations of input are unlikely to capture the full richness of neural dynamics. In part, this is because linear systems provide parsimonious *local* descriptions (for a single experimental condition, a single task epoch, or for restricted ranges of neural activity) of potentially more complex circuit dynamics (Figure 2a). Local linear approximations also represent an important building block for the analysis of the behaviour of more complex, nonlinear dynamical systems (Figure 2c) [44].

Switching linear models

Switching linear dynamical systems mix multiple linear updates under the control of a discrete switching variable. The simplest model takes the form:

$$\begin{aligned} s_{t+1} &\sim p(s_{t+1}|s_t), \quad s_t \in \{1, ..., M\} \\ x_{t+1} &= A_{s_t}x_t + B_{s_t}u_t + \boldsymbol{\eta}_t, \end{aligned} \qquad (2)$$

with the discrete Markov variable s_t selecting the linear evolution A_i and input mapping B_i ($i = 1, …, M$) that act on the computational state at time t. Switching models are more expressive than a single linear model: switches in dynamics may reflect the influence of multiple fixed points, or capture changes in dynamics with time, or under the control of unobserved inputs. More generally, arbitrary nonlinear dynamics may be approximated by a piecewise linear form (Figure 2b). However, while the simple form of Eq. (2) allows transitions between dynamical regimes to be estimated from neural data, it lacks the capacity to select linear dynamics based on the value of the computational state. Thus, a richer model includes dependence of s_t (and thus the local linear dynamics) on the current computational state x_t [45–48]. One advantage to such an approximation is that by retaining explicit linearity in each subsystem, dynamics within each region can still be analysed and understood using classic tools from linear systems analysis.

Applications of switching linear dynamical systems include the analysis of neural population activity during motor planning and execution [48,49] and the identification of latent activity patterns in recordings from *C. elegans* [46,48]. In each case, models were able to identify relationships between behavioural states and the discrete state sequence inferred in the model.

Nonlinear models

Theoretical studies have suggested essential roles for nonlinear network dynamics in computation [35,50,51], with outputs depending on multiple fixed points, limit cycles or line attractors, or bifurcations with changes in inputs. Many of these features cannot be expressed in linear systems, and may be difficult to approximate well with switching linear systems without overwhelming complexity. Consequently, there has also been substantial interest in identifying nonlinear models of the circuit dynamics from neural data. In these models, the dynamics are expressed very generally through a parametrised function $f(\cdot)$:

$$x_{t+1} = f(x_t, u_t) + \boldsymbol{\eta}_t \qquad (3)$$

which describes the state evolution as a function of both current latent state x and any inputs arriving to the system u.

Choosing flexible function classes for $f(\cdot)$ makes it possible to describe richer sets of dynamics than linear or piecewise-linear models. Such flexible choices include nonparametric approaches like Gaussian Process State Space Models [52–54], representation of the dynamics in terms of eigenfunctions of the Hermitian operator of the dynamical system [55], or parameterisations in terms of a neural network [56,57].

The local behaviour of nonlinear dynamics can still be understood in terms of linear approximations given by the Jacobian matrix evaluated at fixed points of the system (Figure 2c) [44]. However, identifying fixed points and Jacobians of the nonlinear system typically requires a second stage of estimation [58] or the specification of models to include these features directly as additional parameters [54].

Outstanding challenges and future directions

While dynamical systems offer a compelling framework for studying computation in recurrent circuits, connecting such ideas to neural data is met with a number of challenges in practice.

These include challenges associated with fitting dynamical models to data, like choosing good criteria for model selection when the goal is interpretation and scientific discovery [55], and adapting statistical models to accommodate neural response features, which can lead to intractabilities in performing statistical inference.

Furthermore, inputs to the local population are generally unknown, which makes it difficult to determine which features of neural activity reflect the dynamics of the local circuit or dynamics of inputs from other areas. Recent work has begun to investigate the use of single trial activity to tease apart these contributions [59]. Still, existing approaches for studying the dynamics of a local population generally have to make assumptions about these components in order to arrive at a model description that is identifiable from data (Figure 3a). So far, it has been difficult to validate such assumptions, yet their impact on the inferred dynamics may be large.

A further challenge relates to the interpretability of latent dynamical models with respect to the goal of better understanding computation in biological circuits. The low-dimensional dynamics evolve in an abstract latent space, which is typically identified in an unsupervised way and may not easily be related to variables relevant to the computation or task of interest. Deriving explicit statements about mechanistic circuit computation from the estimated model fits is hard since any inference about characteristics of dynamics in latent space do not directly translate to dynamics defined through a connectivity matrix in neural space. Known constraints on the dynamics of the biological circuit, such as excitatory and inhibitory cell-type structure and Dale's law, are typically not taken into account in methods that learn dynamics from data. However, such properties might influence features of the neural responses used for fitting. This discrepancy in levels of analysis also makes it difficult to derive testable experimental predictions based on any insights data analysis methods may provide.

To address these challenges, it is important to highlight that there are other complementary approaches towards studying network dynamics and computation. These include reverse-engineering task-optimised artificial neural network models [44,60] (Figure 3b) and designing mechanistic network implementations of a particular computation [35] (Figure 3c). Such network models are fully observable and can incorporate biological constraints, yet their connection to data is often only qualitative in nature. Furthermore, they too rely on model assumptions, such as network architecture, cost function, and structure in network inputs and target outputs.

Drawing connections between mechanistic models of network computations, task-optimised artificial neural networks, and data analysis will help to better inform model assumptions and provide more quantitative links between theoretical network models and observed neural data. Such connections will also aid model interpretability and ultimately inform the design of experiments to test and refine the hypotheses different models may generate. Indeed, recent work has begun to formulate statistical models in ways to establish more direct connections between theory and data analysis [54,61] or use experimentally observed responses to constrain activity in recurrent network models [62]. In addition to this, establishing connections between latent variable models and mechanistic modelling approaches might aid the interpretability of model parameters that are learned from data. In going forward, it will be important to establish clearer links between theoretical descriptions of network computation and data-driven latent dynamical models.

Conflict of interest statement

Nothing declared.

Acknowledgements

References

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
•• of outstanding interest

1. Fabre-Thorpe Michèle: **The characteristics and limits of rapid visual categorization**. *Front Psychol* 2011, **2**:243.

2. Wickelgren Wayne A: **Speed-accuracy tradeoff and information processing dynamics**. *Acta Psychol* 1997, **41**:67−85.

3. Adam Reeves: **An analysis of visual masking, with a defense of 'stopped processing'**. *Adv Cognit Psychol* 2008, **3**:57−65.

4. Gao Peiran, Ganguli Surya: **On simplicity and complexity in the brave new world of large-scale neuroscience**. *Curr Opin Neurobiol* 2015, **32**:148−155.

5. Byron M Yu, Cunningham John P, Santhanam Gopal, Ryu Stephen I, Shenoy Krishna V, Sahani Maneesh: **Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity**. *J Neurophysiol* 2009, **102**:614−635.

6. Shenoy Krishna V, Sahani Maneesh, Churchland Mark M: **Cortical control of arm movements: a dynamical systems perspective**. *Annu Rev Neurosci* 2013, **36**:337−359.

7. Churchland Mark M, Cunningham John P, Kaufman Matthew T, Foster Justin D, Paul Nuyujukian, Ryu Stephen I, Shenoy Krishna V: **Neural population dynamics during reaching**. *Nature* 2012, **487**:51–56.

8. Elsayed Gamaleldin F, Lara Antonio H, Kaufman Matthew T, Churchland Mark M, Cunningham John P: **Reorganization between preparatory and movement population responses in motor cortex**. *Nat Commun* 2016, **7**:1–15.

9. Mante Valerio, Sussillo David, Shenoy Krishna V, Newsome William T: **Context-dependent computation by recurrent dynamics in prefrontal cortex**. *Nature* 2013, **503**:78–84.

10. Machens Christian K, Romo Ranulfo, Brody Carlos D: **Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex**. *J Neurosci* 2010, **30**:350–360.

11. Remington Evan D, Narain Devika, Hosseini Eghbal A, Jazayeri Mehrdad: **Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics**. *Neuron* 2018, **98**:1005–1019.

12. Sohn Hansem, Narain Devika, Meirhaeghe Nicolas, Jazayeri Mehrdad: **Bayesian computation through cortical latent dynamics**. *Neuron* 2019, **103**:934–947.

13. O'Keefe John, Dostrovsky Jonathan: **The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat**. *Brain Res* 1971, **34**:171–175.

14. Taube Jeffrey S, Muller Robert U, Ranck James B: **Head-direction cells recorded from the postsubiculum in freely moving rats. I. description and quantitative analysis**. *J Neurosci* 1990, **10**:420–435.

15. Fyhn Marianne, Molden Sturla, Witter Menno P, Moser Edvard I, Moser May-Britt: **Spatial representation in the entorhinal cortex**. *Science* 2004, **305**:1258–1264.

16. Chaudhuri Rishidev, Gerçek Berk, Pandey Biraj, Peyrache Adrien, Fiete Ila: **The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep**. *Nat Neurosci* 2019, **22**:1512–1520.

17. Jensen Kristopher, Kao Ta-Chu, Tripodi Marco,
• Hennequin Guillaume: **Manifold GPLVMs for discovering non-euclidean latent structure in neural data**. In Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H. *Advances in neural information processing systems*, vol. 33. Curran Associates, Inc.; 2020:22580–22592.
Jensen et al. develop a novel approach to infer a latent variable specified on a manifold of a given geometry, such as a sphere, ring or torus. The approach recovers improved latent state estimates when the underlying manifold is taken into account, rather than assuming a Euclidean latent space. In cases where the manifold topology is unknown, cross-validation can be used for model selection across candidate topologies.

18. Low Ryan J, Lewallen Sam, Aronov Dmitriy, Nevers Rhino, Tank David W: *Probing variability in a cognitive map using manifold inference from neural dynamics*. bioRxiv; 2018:418939.

19. Rubin Alon, Sheintuch Liron, Brande-Eilat Noa, Pinchasof Or, Rechavi Yoav, Geva Nitzan, Ziv Yaniv: **Revealing neural correlates of behavior without behavioral measurements**. *Nat Commun* 2019, **10**:1–14.

20. Abeles Moshe: *Local cortical circuits an electrophysiological study*. Springer; 1982.

21. Gao Peiran, Trautmann Eric, Yu Byron, Gopal Santhanam, Ryu Stephen, Shenoy Krishna, Ganguli Surya: *A theory of multineuronal dimensionality, dynamics and measurement*. bioRxiv; 2017:214262.

22. Stringer Carsen, Pachitariu Marius, Steinmetz Nicholas, Carandini Matteo, Harris Kenneth D: **High-dimensional geometry of population responses in visual cortex**. *Nature* 2019, **571**:361–365.

23. Jeffrey S Seely, Kaufman Matthew T, Ryu Stephen I, Shenoy Krishna V, Cunningham John P, Churchland Mark M: **Tensor analysis reveals distinct population structure that parallels the different computational roles of areas M1 and V1**. *PLoS Comput Biol* 2016, **12**, e1005164.

24. Cunningham John P, Yu Byron M: **Dimensionality reduction for large-scale neural recordings**. *Nat Neurosci* 2014, **17**:1500–1509.

25. Kato Saul, Kaplan Harris S, Schrödel Tina, Skora Susanne, Lindsay Theodore H, Yemini Eviatar, Lockery Shawn, Zimmer Manuel: **Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans***. *Cell* 2015, **163**:656–669.

26. Wu Anqi, Pashkovski Stan, Datta Sandeep R, Pillow Jonathan W: **Learning a latent manifold of odor representations from neural responses in piriform cortex**. In *Advances in neural information processing systems*; 2018:5378–5388.

27. Gallego Juan A, Perich Matthew G, Miller Lee E, Solla Sara A: **Neural manifolds for the control of movement**. *Neuron* 2017, **94**:978–984.

28. Keeley Stephen, Aoi Mikio, Yu Yiyi, Smith Spencer, Pillow Jonathan W: **Identifying signal and noise structure in neural population activity with Gaussian process factor models**. In *Advances in neural information processing systems*, vol. 33; 2020.

29. Kobak Dmitry, Brendel Wieland, Constantinidis Christos, Feierstein Claudia E, Kepecs Adam, Mainen Zachary F, Qi Xue-Lian, Romo Ranulfo, Uchida Naoshige, Machens Christian K: **Demixed principal component analysis of neural population data**. *Elife* 2016, **5**, e10989.

30. Aoi Mikio C, Mante Valerio, Pillow Jonathan W: **Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making**. *Nat Neurosci* 2020:1–11.

31. Rutten Virginia MS, Bernacchia Alberto, Sahani Maneesh, Hennequin Guillaume: **Non-reversible Gaussian processes for identifying latent dynamical structure in neural data**. In Larochelle Hugo, Ranzato Marc'Aurelio, Hadsell Raia, Balcan Maria Florina, Lin Hsuan-Tien. *Advances in neural information processing systems*, vol 33. Curran Associates, Inc.; 2020:9622–9632.

32. Driscoll Laura N, Golub Matthew D, Sussillo David: **Computation through cortical dynamics**. *Neuron* 2018, **98**:873–875.

33. Vyas Saurabh, Golub Matthew D, Sussillo David, Shenoy Krishna V: **Computation through neural population dynamics**. *Annu Rev Neurosci* 2020, **43**:249–275.

34. Roitman Jamie D, Shadlen Michael N: **Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task**. *J Neurosci* 2002, **22**:9475–9489.

35. Wang Xiao-Jing: **Probabilistic decision making by slow reverberation in cortical circuits**. *Neuron* 2002, **36**:955–968.

36. Wong Kong-Fatt, Wang Xiao-Jing: **A recurrent network mechanism of time integration in perceptual decisions**. *J Neurosci* 2006, **26**:1314–1328.

37. Goldman Mark S: **Memory without feedback in a neural network**. *Neuron* 2009, **61**:621–634.

38. Murphy Brendan K, Miller Kenneth D: **Balanced amplification: a new mechanism of selective amplification of neural activity patterns**. *Neuron* 2009, **61**:635–648.

39. Kao Ta-Chu, Hennequin Guillaume: **Neuroscience out of control: control-theoretic perspectives on neural circuit dynamics**. *Curr Opin Neurobiol* 2019, **58**:122–129.

40. Kao Jonathan C, Nuyujukian Paul, Ryu Stephen I, Churchland Mark M, Cunningham John P, Shenoy Krishna V: **Single-trial dynamics of motor cortex and their applications to brain-machine interfaces**. *Nat Commun* 2015, **6**:1–12.

41. Magraner Joana Soldado: *Linear dynamics of evidence integration in contextual decision making*. PhD thesis. UCL (University College London); 2018.

42. Hennequin Guillaume, Vogels Tim P, Gerstner Wulfram: **Optimal control of transient dynamics in balanced networks supports generation of complex movements**. *Neuron* 2014, **82**:

1394−1406.

43. Kao Ta-Chu, Sadabadi Mahdieh S, Hennequin Guillaume: **Optimal anticipatory control as a theory of motor preparation: a thalamo-cortical circuit model**. *Neuron* 2021.

44. Sussillo David, Barak Omri: **Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks**. *Neural Comput* 2013, **25**:626−649.

45. Scott Linderman, Johnson Matthew, Miller Andrew, Adams Ryan, Blei David, Paninski Liam: **Bayesian learning and inference in recurrent switching linear dynamical systems**. In *Artificial intelligence and statistics*; 2017:914−922.

46. Scott Linderman, Nichols Annika, Blei David, Zimmer Manuel,
• Paninski Liam: *Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in* C. elegans. bioRxiv; 2019:621540.
Linderman et al. introduce a hierarchical switching dynamical system to model neural activity in *C. elegans*. Their hierarchical approach allows them to use data from different partial recordings and different animals to estimate model structure that is shared or unique to each animal. They demonstrate that the interred time-course of discrete states in the model is behaviourally meaningful.

47. Nassar Josue, Scott Linderman, Bugallo Monica, Park Il Memming: **Tree-structured recurrent switching linear dynamical systems for multi-scale modeling**. In *International conference on learning representations*; 2019.

48. Glaser Joshua, Whiteway Matthew, Cunningham John P,
•• Paninski Liam, Scott Linderman: **Recurrent switching dynamical systems models for multiple interacting neural populations**. In Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H. *Advances in neural information processing systems*, vol. 33. Curran Associates, Inc.; 2020:14867−14878.
Glaser et al. use a switching linear dynamical system to model the dynamics of interacting neural populations. The model allows for switches in the dynamics of each population, as well as in the interactions across populations. The authors apply their method to data from cortical motor areas in nonhuman primates and neural activity in *C. elegans*.

49. Petreska Biljana, Yu Byron M, Cunningham John P, Santhanam Gopal, Ryu Stephen I, Shenoy Krishna V, Sahani Maneesh: **Dynamical segmentation of single trials from population neural data**. In *Advances in neural information processing systems*; 2011:756−764.

50. Machens Christian K, Romo Ranulfo, Brody Carlos D: **Flexible control of mutual inhibition: a neural model of two-interval discrimination**. *Science* 2005, **307**:1121−1124.

51. Mastrogiuseppe Francesca, Ostojic Srdjan: **Linking connectivity, dynamics, and computations in low-rank recurrent neural networks**. *Neuron* 2018, **99**:609−623.

52. Wang Jack, Hertzmann Aaron, Fleet David J: **Gaussian process dynamical models**. In *Advances in neural information processing systems*, vol. 18; 2006:1441−1448.

53. Turner Ryan, Deisenroth Marc, Rasmussen Carl: **State-space inference and learning with Gaussian processes**. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*; 2010:868−875.

54. Duncker Lea, Bohner Gergo, Boussard Julien, Sahani Maneesh:
•• **Learning interpretable continuous-time models of latent stochastic dynamical systems**. In *Proceedings of the 36th*

international conference on machine learning, volume 97 of proceedings of machine learning research. Edited by Chaudhuri Kamalika, Salakhutdinov Ruslan, Long Beach, California, USA: PMLR; 09−15 Jun 2019:1726−1734.
The authors develop an approach to learn a low-dimensional nonlinear dynamical system directly from the spike times of simultaneously recorded neurons. The nonlinear dynamics are described via a Gaussian Process, conditioned on fixed points and local Jacobian matrices around them. These interpretable features of the dynamics are hence directly accessible for further analysis instead of requiring a second stage of estimation to identify fixed points of the dynamics.

55. Genkin Mikhail, Engel Tatiana A: **Moving beyond generalization**
• **to accurate interpretation of flexible models**. *Nat Mach Intell* 2020.
Genkin et al. develop an approach for maximum likelihood learning of a latent nonlinear dynamical system from spike train observations. The authors also discuss issues relating to model selection based on generalisation performance and the interpretation of the resulting dynamics. They propose a new complexity metric to select interpretable models across a range of models with similar predictive performance.

56. Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, Trautmann EM, Kaufman MT, Ryu SI, Hochberg LR, Henderson JM, Shenoy KV, Abbott LF, Sussillo D: **Inferring single-trial neural population dynamics using sequential auto-encoders**. *Nature Methods* 2018, **15**:805−815.

57. Hernandez Daniel, Moretti Antonio K, Saxena Shreya, Wei Ziqiang, Cunningham John, Paninski Liam: **Nonlinear evolution via spatially-dependent linear dynamics for electrophysiology and calcium data**. *Neurons, Behav, Data Anal Theory* 2020, **3**:13476.

58. Golub Matt, Sussillo David: **Fixedpointfinder: a Tensorflow toolbox for identifying and characterizing fixed points in recurrent neural networks**. *J Open Source Software* 2018, **3**: 1003.

59. Galgali Aniruddh R, Sahani Maneesh, Mante Valerio: *Residual*
•• *dynamics resolves recurrent contributions to neural computation*. bioRxiv; 2021.
Galgali et al. develop a method for identifying local recurrent contribution to observed neural activity patterns by fitting time-varying linear dynamics to single-trial residuals around a condition of average trajectory. The authors apply their approach to population activity from the macaque prefrontal cortex to study its role in perceptual decision making.

60. Barak Omri: **Recurrent neural networks as versatile tools of neuroscience research**. *Curr Opin Neurobiol* 2017, **46**:1−6.

61. Zoltowski David M, Pillow Jonathan W, Linderman Scott W: **A**
• **general recurrent state space framework for modeling neural dynamics during decision-making**. In *Proceedings of the 37th international Conference on machine learning, volume 119 of Proceedings of machine learning research*. PMLR; 13−18 Jul 2020:11680−11691. Virtual.
Zoltowski et al. express different theoretical models of decision-making as specific instantiations of a switching dynamical system. This allows them to quantitatively evaluate and compare the different theoretical models in neural data. The inferred latent states have a direct theoretical interpretation in their modelling framework.

62. Perich Matthew G, Arlt Charlotte, Soares Sofia, Young Megan E, Mosher Clayton P, Minxha Juri, Carter Eugene, Rutishauser Ueli, Rudebeck Peter H, Harvey Christopher D, Rajan Kanaka: *Inferring brain-wide interactions using data-constrained recurrent neural network models*. bioRxiv; 2020.