

RESEARCH ARTICLE

Neural field models for latent state inference: Application to large-scale neuronal recordings

Michael E. Rule^{1*}, David Schnoerr², Matthias H. Hennig³, Guido Sanguinetti³

1 Department of Engineering, University of Cambridge, Cambridge, United Kingdom, **2** Theoretical Systems Biology, Imperial College London, London, United Kingdom, **3** Department of Informatics, University of Edinburgh, Edinburgh, United Kingdom

* mrule7404@gmail.com



Abstract

Large-scale neural recording methods now allow us to observe large populations of identified single neurons simultaneously, opening a window into neural population dynamics in living organisms. However, distilling such large-scale recordings to build theories of emergent collective dynamics remains a fundamental statistical challenge. The neural field models of Wilson, Cowan, and colleagues remain the mainstay of mathematical population modeling owing to their interpretable, mechanistic parameters and amenability to mathematical analysis. Inspired by recent advances in biochemical modeling, we develop a method based on moment closure to interpret neural field models as latent state-space point-process models, making them amenable to statistical inference. With this approach we can infer the intrinsic states of neurons, such as active and refractory, solely from spiking activity in large populations. After validating this approach with synthetic data, we apply it to high-density recordings of spiking activity in the developing mouse retina. This confirms the essential role of a long lasting refractory state in shaping spatiotemporal properties of neonatal retinal waves. This conceptual and methodological advance opens up new theoretical connections between mathematical theory and point-process state-space models in neural data analysis.

OPEN ACCESS

Citation: Rule ME, Schnoerr D, Hennig MH, Sanguinetti G (2019) Neural field models for latent state inference: Application to large-scale neuronal recordings. *PLoS Comput Biol* 15(11): e1007442. <https://doi.org/10.1371/journal.pcbi.1007442>

Editor: Jonathan David Touboul, Brandeis University, UNITED STATES

Received: May 9, 2019

Accepted: September 27, 2019

Published: November 4, 2019

Copyright: © 2019 Rule et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets were previously published in "Maccione A, Hennig MH, Gandolfo M, Muthmann O, Copenhagen J, Eglen SJ, et al. Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. *The Journal of physiology* 2014;592(7):1545–1563". The data underlying the results presented in the study are available from Dr. Evelyne Sernagor (evelyne.sernagor@ncl.ac.uk). A subset of these datasets are available online (damtp.cam.ac.uk/user/sje30/waverepo). Demonstrate code is available on Github (github.com/michaelerule/NFCP_release).

Author summary

Developing statistical tools to connect single-neuron activity to emergent collective dynamics is vital for building interpretable models of neural activity. Neural field models relate single-neuron activity to emergent collective dynamics in neural populations, but integrating them with data remains challenging. Recently, latent state-space models have emerged as a powerful tool for constructing phenomenological models of neural population activity. The advent of high-density multi-electrode array recordings now enables us to examine large-scale collective neural activity. We show that classical neural field approaches can yield latent state-space equations and demonstrate that this enables inference of the intrinsic states of neurons from recorded spike trains in large populations.

Funding: Funding was provided by EPSRC EP/L027208/1 "Large scale spatio-temporal point processes: novel machine learning methodologies and application to neural multi-electrode arrays". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Neurons communicate using electrical impulses, or spikes. Understanding the dynamics and physiology of collective spiking in large networks of neurons is a central challenge in modern neuroscience, with immense translational and clinical potential. Modern technologies such as high-density multi-electrode arrays (HDMEA) enable the simultaneous recording of the electrical activity of thousands of interconnected neurons, promising invaluable insights into neural dynamics at the network level. However, the resulting data is high-dimensional and frequently exhibits complex, non-linear dynamics, presenting formidable statistical challenges.

Due to the complexity of the data, most analyses of neuronal population activity take a descriptive approach, adopting methods from statistical signal processing such as state-space models (SSM; [1–7]) or autoregressive generalized-linear point-process models (PP-GLM; [8–11]). Such methods capture the population statistics of the system, but fail to provide mechanistic explanations of the underlying neural dynamics. While this phenomenological description is valuable and can aid many investigations, the inability to relate microscopic single-neuron properties to emergent collective dynamics limits the scope of these models to extract biological insights from these large population recordings.

Connecting single-neuron dynamics with population behavior has been the central focus of research within the theoretical neuroscience community over the last four decades. Neural field models [12–15] have been crucial in understanding how macroscopic firing dynamics in populations of neurons emerge from the microscopic state of individual neurons. Such models have found diverse applications including working memory (see [16] for a review), epilepsy (e.g. [17–20]), and hallucinations (e.g. [21–23]), and have been successfully related to neuroimaging data such as Electroencephalography (EEG; [24–26]), Magnetoencephalography (MEG; [24]), Electromyography (EMG; [27]), and Functional Magnetic Resonance Imaging (fMRI; [25]), which measure average signals from millions of neurons. Nevertheless, using neural-field models to model HDMEA spiking data directly remains an open statistical problem: HDMEA recordings provide sufficient detail to allow modeling of individual neurons, yet the large number of neurons present prevents the adoption of standard approaches to non-linear data assimilation such as likelihood free inference.

In this paper, we bridge the data-model divide by developing a statistical framework for Bayesian modeling in neural field models. We build on recent advances in stochastic spatio-temporal modeling, in particular a recent result by Schnoerr et al. [28] which showed that a spatiotemporal agent-based model of reaction-diffusion type, similar to the ones underpinning many neural field models, can be approximated as a spatiotemporal point process associated with an intensity (i.e. density) field that evolves in time. Subsequently, Rule and Sanguinetti [29] illustrated a moment-closure approach for mapping stochastic models of neuronal spiking onto latent state-space models, preserving the essential coarse-timescale dynamics. Here, we demonstrate that a similar approach can yield state-space models for neural fields derived directly from a mechanistic microscopic description. This enables us to leverage large-scale spatiotemporal inference techniques [30, 31] to efficiently estimate an approximate likelihood, providing a measure of fit of the model to the data that can be exploited for data assimilation. Our approach is in spirit similar to latent variable models such as the Poisson Linear Dynamical System (PLDS; [5, 32, 33]), with the important difference that the latent variables reflects non-linear neural field dynamics that emerge directly from a stochastic description of single-neuron activity [34–36].

We apply this approach to HDMEA recordings of spontaneous activity from ganglion cells in the developing mouse retina [37], showing that the calibrated model effectively captures the

non-linear excitable phenomenon of coordinated, wave-like patterns of spiking [38] that have been considered in both discrete [39] and continuous neural-field models before [40].

Results

High level description of the approach

We would like to explain large-scale spatiotemporal spiking activity in terms of the intrinsic states of the participating neurons, which we cannot observe directly. Latent state-space models (SSMs) solve this problem by describing how the unobserved states of neurons relate to spiking observations, and predict how these latent states evolve in time. In this framework, one estimates a distribution over latent states from observations, and uses a forward model to predict how this distribution evolves in time, refining the latent-state estimate with new observations as they become available. This process is often called ‘data assimilation’. However, in order to achieve statistical tractability, SSMs posit simple (typically linear) latent dynamics, which cannot be easily related to underlying neuronal mechanisms. Emergent large-scale spatiotemporal phenomena such as traveling waves typically involve multiple, coupled populations of neurons and nonlinear excitatory dynamics, both of which are difficult to incorporate into conventional state-space models.

Fortunately, mathematical neuroscience has developed methods for describing such dynamics using neural field models. Neural field models map microscopic dynamics to coarse-grained descriptions of how population firing rates evolve. This provides an alternative route to constructing latent state-space models for large-scale spatiotemporal spiking datasets. However, neural field models traditionally do not model statistical uncertainty in the population states they describe, which makes it difficult to deploy them as statistical tools to infer the unobserved, latent states of the neuronal populations. A model of statistical uncertainty is important for describing the uncertainty in the estimated latent states (posterior variance), as well as correlations between states or spatial regions. As we will illustrate, work over the past decades to address noise and correlations in neural field models also provides the tools to employ such models as latent SSMs in data-driven inference.

At a high level then, our approach follows the usual derivation of neural field models, starting with an abstract description of single-neuron dynamics, and considers how population averages evolve in time. Rather than deriving a neural-field equation for the population mean rate, we instead derive two coupled equations for the mean and covariance of population states. We interpret these two moments as a Gaussian-process estimate of the latent spatiotemporal activity, and derive updates for how this distribution evolves in time and how it predicts spiking observations. This provides an interpretation of neural-field dynamics amenable to state-space inference, which allows us to infer neural population states from spiking observations.

Neural field models for refractoriness-mediated retinal waves

Although Wilson and Cowan [41, 42] considered refractoriness, most subsequent applications consider only two states: neurons may be either actively spiking (*A* state), or quiescent (*Q* state). In general, voltage and calcium gated conductances typically lead to refractory states, which can be short following individual spikes, or longer after more intensive periods of activity. An excellent example of the importance of a refractory mechanism is found in the developing retina, where a slow afterhyperpolarization (sAHP) current mediates the long-timescale refractory effects that strongly shapes the spatiotemporal dynamics of spontaneous retinal waves [43]. To address this, we explicitly incorporate additional refractory (*R*) states into our neural field model (e.g. [44, 45]; Fig 1). In the following, we first outline a non-spatial model

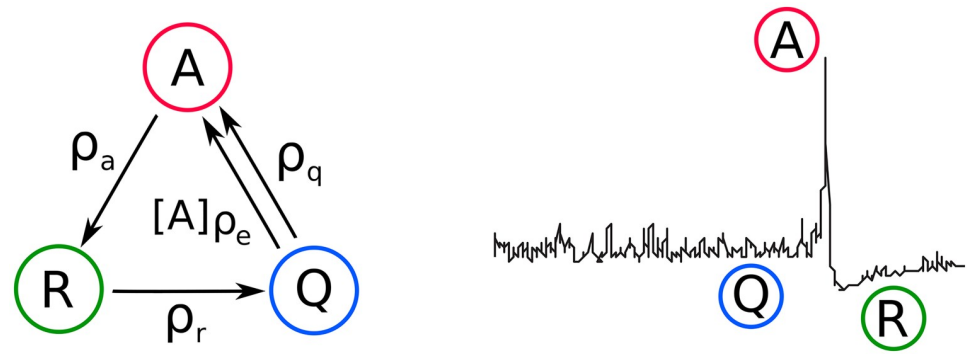


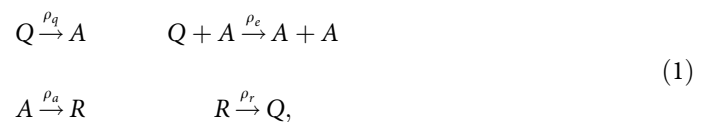
Fig 1. 3-state Quiescent-Active-Refractory (QAR) neural-field model. Cells in the developing retina are modeled as having three activity states. Active cells (A; red) fire bursts of action potentials, before becoming refractory (R; green) for an extended period of time. Quiescent (Q; blue) cells may burst spontaneously, or may be recruited into a wave by other active cells. These three states are proposed to underlie critical multi-scale wave dynamics [43].

<https://doi.org/10.1371/journal.pcbi.1007442.g001>

for such system, before extending it to a spatial setting with spatial couplings. Finally, we develop a Bayesian inference scheme for inferring latent states from observational data.

A stochastic three-state neural mass model

We now consider the neural field model with three states as a generic model of a spiking neuron (Fig 1), where a neuron can be in either an actively spiking (A), refractory (R), or quiescent (Q) state. We assume that the neurons can undergo the following four transitions:



i.e. quiescent neurons transition spontaneously to the active state; active neurons excite quiescent neurons; active neurons become refractory, and refractory neurons become quiescent. The $\rho_{(\cdot)}$ denote corresponding rate constants.

For illustration, we first consider the dynamics of a local (as opposed to spatially-extended) population of neurons. In this case the state of the system is given by the non-negative number counts Q , A and R of the respective neuron types (we slightly abuse notation here and use Q , A , and R both as symbols for the neuron states and as variables counting the neurons in the corresponding states; see Fig 2 for an illustration). The time evolution of the corresponding probability distribution to be in a state (Q, A, R) at a certain time point is then given by a master equation ([34, 44, 46]; *Methods: Moment-closure for a single population*). Due to the nonlinear excitatory interaction $Q + A \rightarrow A + A$ in Eq (1), no analytic solutions to the master equation are known. To get an approximate description of the dynamics, we employ the Gaussian moment closure method which approximates the discrete neural counts (Q, A, R) by continuous variables, and assumes a multivariate normal distribution (Fig 2B; [29, 34, 35, 47–50]). This allows one to derive a closed set of ordinary differential equations for the mean and covariance of the approximate process which can be solved efficiently numerically (*Methods: Moment-closure for a single population; Fig 2*).

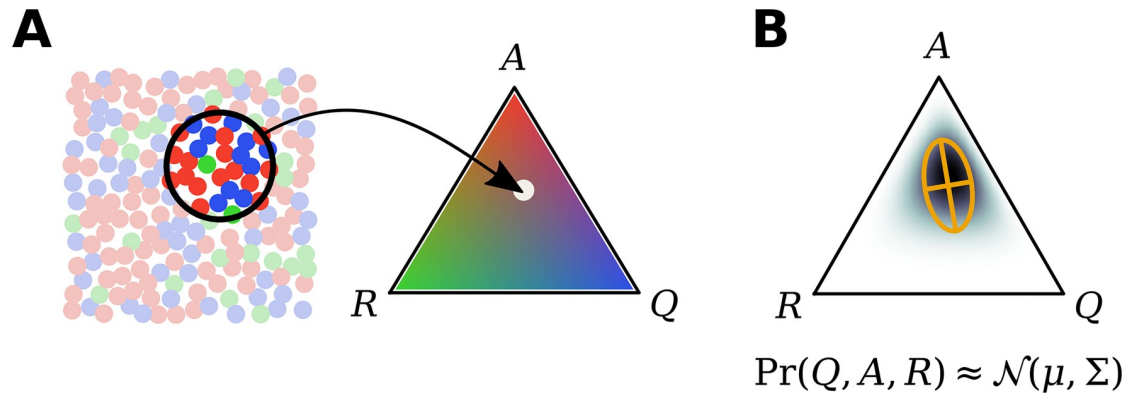


Fig 2. Summarizing estimated neural state as population moments. (A) The activity within a local spatial region (encircled, left) can be summarized by the fraction of cells (represented by colored dots) in the quiescent (blue), active (red), and refractory (green) states (Q, A, R; right). (B) An estimate of the population state can be summarized as a probability distribution $\Pr(Q, A, R)$ over the possible proportions of neurons in each state. A Gaussian moment-closure approximates this distribution as Gaussian, with given mean and covariance (orange crosshairs).

<https://doi.org/10.1371/journal.pcbi.1007442.g002>

Applying this procedure to our system leads to the following evolution equations of the first moments (mean concentrations):

$$\begin{aligned} \partial_t \langle Q \rangle &= r_{rq} - r_{qa} & r_{qa} &= \rho_q \langle Q \rangle + \rho_e \langle AQ \rangle \\ \partial_t \langle A \rangle &= r_{qa} - r_{ar} & r_{ar} &= \rho_a \langle A \rangle \\ \partial_t \langle R \rangle &= r_{ar} - r_{rq} & r_{rq} &= \rho_r \langle R \rangle, \end{aligned} \tag{2}$$

where the rate variables $r_{(\cdot)(\cdot)}$ describe the rates of the different transitions in Eq (1), and $\langle \cdot \rangle$ denotes expected-value with respect to the distribution over population states. Intuitively, Eq (2) says that the mean number of neurons in each state evolves according to the difference between the rate that neurons enter, and the rate that neurons leave, said state. For spontaneous (Poisson) state transitions, these rates are linear and depend only on the average number of neurons in the starting state. The transition from Q to A, however, has both a spontaneous and excito-excitatory component. The latter depends on the expected product of active and quiescent cells $\langle AQ \rangle$, which is a second moment and can be expressed in terms of the covariance: $\langle AQ \rangle = \langle A \rangle \langle Q \rangle + \Sigma_{AQ}$. We obtain similar equations for the covariance of the system (Eq 6; Methods: Moment-closure for a single population). These can be solved jointly with Eq (2) forward in time to give an approximation of the system’s dynamics.

Generalization to spatial (neural field) system

So far we have considered a single local population. We next extend our model to a two-dimensional spatial system. In this case the mean concentrations become density or mean fields (‘neural fields’) that depend on spatial coordinates $\mathbf{x} = (x_1, x_2)$, e.g. $\langle Q \rangle$ becomes $\langle Q(\mathbf{x}) \rangle$. Similarly, the covariances become two-point correlation functions. For example, $\Sigma_{QA}(\mathbf{x}, \mathbf{x}')$ denotes the covariance between the number of neurons in the quiescent state at location \mathbf{x} and the number of neurons in the active state at location \mathbf{x}' (see Methods: Extension to spatial system for details).

By replacing the mean concentrations and covariances accordingly in Eqs (2) and (6), we obtain spatial evolution equations for these space-dependent quantities. The terms arising from the linear transitions in Eq (1) (i.e. r_{rq} , r_{aq} and the first term in r_{qa} in Eq 2) do not introduce any spatial coupling and hence do not need to be modified (note also that neurons do not

diffuse or move otherwise, which is why we do not obtain a dynamic term in the resulting equations). The nonlinear excitatory interaction $Q + A \rightarrow A + A$ in Eq (1), however, introduces a coupling which we need to specify further in a spatial setting. We assume that each quiescent neuron experiences an excitatory drive from nearby active neurons, and that the interaction strength can be described as a function of distance $\|\Delta\mathbf{x}\|$ by a Gaussian interaction kernel:

$$k(\Delta\mathbf{x}) \propto \exp(-\|\Delta\mathbf{x}\|^2/2\sigma_e^2), \tag{3}$$

where σ_e the standard deviation determining the length scale of the interaction, which decays exponentially as a function of distance squared. This kernel introduces a spatial coupling between the neurons, which could be mediated by synaptic interactions, diffusing neurotransmitters, gap junction coupling, or combinations thereof. With this coupling, the transition rate (compare to Eq (2)) from the quiescent to active state at position \mathbf{x} becomes the following integral:

$$r_{qa}(\mathbf{x}) = \rho_q \langle \mathbf{Q}(\mathbf{x}) \rangle + \rho_e \int k(\mathbf{x} - \mathbf{x}') \langle \mathbf{Q}(\mathbf{x}) \mathbf{A}(\mathbf{x}') \rangle d\mathbf{x}', \tag{4}$$

where the integral runs over the whole volume of the system (Methods: Extension to spatial system).

We thus obtain a ‘second-order’ neural field in terms of the mean fields and two-point correlation functions. We simulated the spatially-extended system by sampling. Fig 3 shows that it is indeed capable of producing multi-scale wave-like phenomena similar to the waves observed in the retina (Methods: Sampling from the model).

Neural field models as latent-variable state-space models

The equations for the mean fields and correlations can be integrated forward in time and used as a state-space model to explain population spiking activity (Fig 4; Methods: Bayesian filtering). In extracellular recordings, we do not directly observe the intensity functions $\langle \mathbf{Q}(\mathbf{x}) \rangle$, $\langle \mathbf{A}(\mathbf{x}) \rangle$, and $\langle \mathbf{R}(\mathbf{x}) \rangle$. Instead, we observe the spikes that active neurons emit, or in the case of developmental retinal waves recorded via a HDMEA setup, we observe the spikes of retinal ganglion cells which are driven by latent wave activity. The spiking intensity should hence

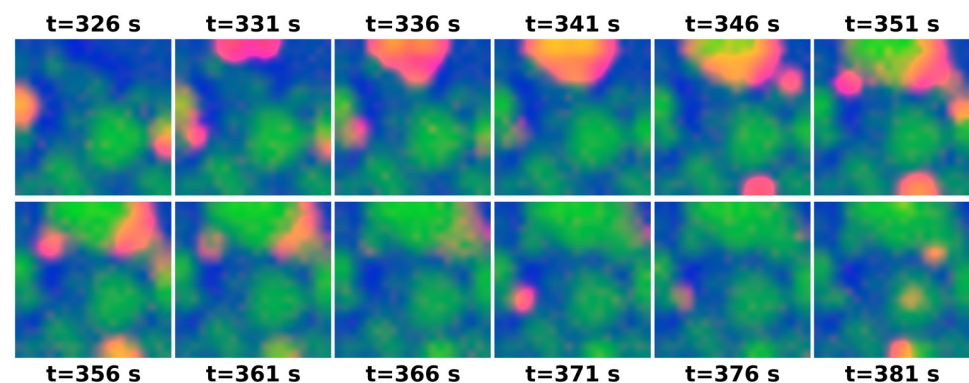


Fig 3. Spatial 3-state neural-field model exhibits self-organized multi-scale wave phenomena. Simulated example states at selected time-points on a $[0, 1]^2$ unit interval using a 20×20 grid with effective population density of $\rho = 50$ cells per unit area, and rate parameters $\sigma = 0.075$, $\rho_a = 0.4$, $\rho_r = 3.2 \times 10^{-3}$, $\rho_e = 0.028$, and $\rho_q = 0.25$ (Methods: Sampling from the model). As, for instance, in neonatal retinal waves, spontaneous excitation of quiescent cells (blue) lead to propagating waves of activity (red), which establish localized patches in which cells are refractory (green) to subsequent wave propagation. Over time, this leads to diverse patterns of waves at a range of spatial scales.

<https://doi.org/10.1371/journal.pcbi.1007442.g003>

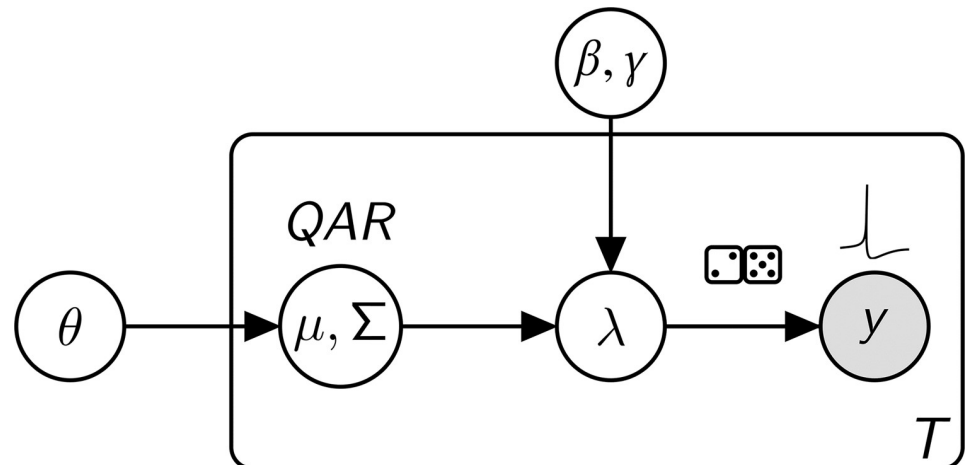


Fig 4. Hidden Markov model for latent neural fields. For all time-points T , state transition parameters $\theta = (\rho_q, \rho_a, \rho_r, \rho_e, \sigma)$ dictate the evolution of a multivariate Gaussian model μ, Σ of latent fields Q, A, R . The observation model (β, γ) is a linear map with adjustable gain and threshold, and reflects how field A couples to firing intensity λ . Point-process observations (spikes) y are Poisson with intensity λ .

<https://doi.org/10.1371/journal.pcbi.1007442.g004>

depend on the density $\mathbf{A}(\mathbf{x})$ of active neurons. Here, we assume that neural firing is a Poisson process conditioned on the number of active neurons, which allows us to write the likelihood of point (i.e. spike) observations in terms of $\mathbf{A}(\mathbf{x})$ ([10, 11, 51]; **Methods: Point-process measurement likelihood**).

The combination of this Poisson-process observation model with the state-space model derived in previous sections describes how hidden neural field states evolve in time and how these states drive neuronal spiking. Given spatiotemporal spiking data, the latent neural field states and correlations can then be inferred using a sequential Bayesian filtering algorithm. The latter uses the neural field model to predict how latent states evolve, and updates this estimate at each time point based on the observed neuronal spiking (**Methods: Bayesian filtering**). This provides estimates of the unobserved physiological states of the neurons.

We verified that this approach works using simulated data. We first simulated observations from the neural field equations (Fig 3; **Methods: Sampling from the model**), which generated waves qualitatively similar to those seen in the developing retina. We then sampled spiking as a conditionally-Poisson process driven by the number of active neurons in each location, with a baseline rate of $\beta = 0$ and gain of $\gamma = 15$ spikes/second per simulation area. We then applied Bayesian filtering to these spiking samples in order to recover a Gaussian estimate of the latent neural field states (**Methods: Bayesian filtering**). Fig 5 illustrates the latent states recovered via filtering using the known ground-truth model parameters, and shows that filtering can recover latent neural field states from the spiking observations. Overall, this indicates that moment-closure of stochastic neural field equations can yield state-space models suitable for state inference from spiking data. In the next section, we illustrate this approach applied to waves recorded from the developing retina.

State inference in developmental retinal waves

Having developed an interpretation of neural field equations as a latent-variable state-space model, we next applied this model to the analysis of spatiotemporal spiking data from spontaneous traveling wave activity occurring in the neonatal vertebrate retina (e.g. Fig 6; [37–39, 52–55]).

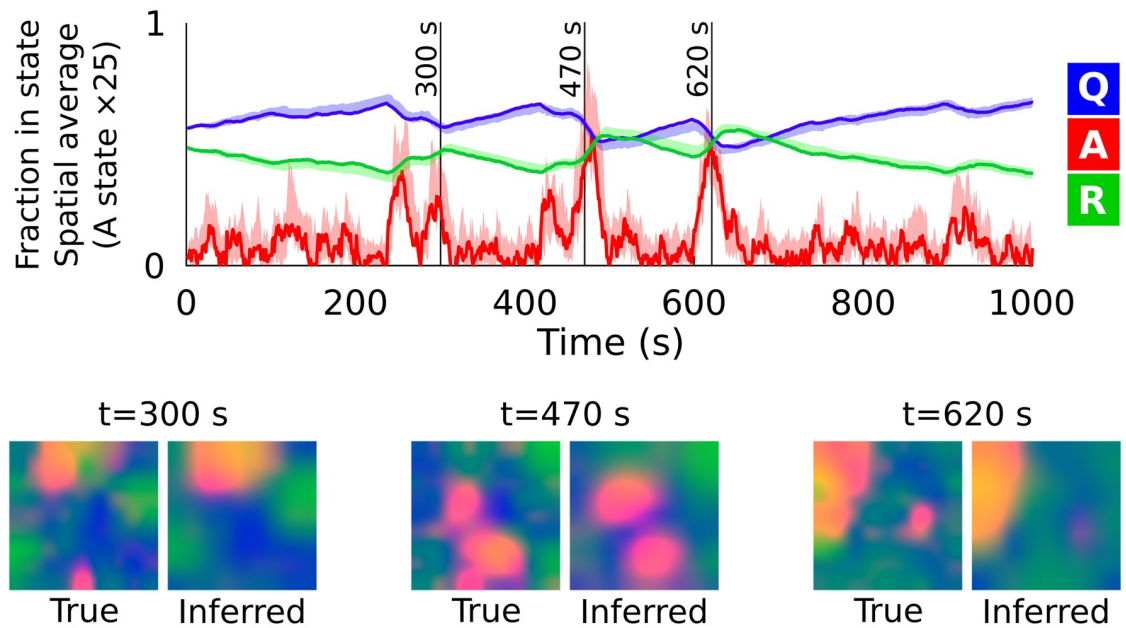


Fig 5. Filtering recovers latent states in ground-truth simulated data. Spatially averaged state occupancy (blue, red, and green: Q, A, and R) (vertical axis) is plotted over time (horizontal axis). Solid lines represent true values sampled from the model, and shaded regions represent the 95% confidence interval estimated by filtering. The active (A) state density has been scaled by $\times 25$ for visualization. Colored plots (below) show the qualitative spatial organization of quiescent (blue), active (red), and refractory (green) neurons. Model parameters are the same as Fig 3, with the exception of the spatial resolution, which has been reduced to 9×9 .

<https://doi.org/10.1371/journal.pcbi.1007442.g005>

During retinal development, the cell types that participate in wave generation change [37, 52, 54], but the three-state model globally describes dynamics in the inner retina at all developmental stages (Fig 6). The Active (A) state describes a sustained bursting state, such as the depolarization characteristic of starburst amacrine cells (Fig 6a) during acetylcholine-mediated early-stage (Stage 2) waves between P0 and P9 [54, 55], and late-stage (Stage 3) glutamate-dependent waves [54, 56]. For example, Fig 6c and 6d illustrates spontaneous retinal wave activity recorded from a postnatal day 6 mouse pup (Stage 2). In addition, at least for cholinergic waves, the slow refractory state R is essential for restricting wave propagation into previously active areas [57]. We note that the multi-scale wave activity exhibited in the three-state neural field model (e.g. Fig 3) recapitulates the phenomenology of retinal wave activity explored in the discrete three-state model of Hennig et al. [43].

Using RGC spikes recorded with a 4,096-electrode HDMEA (Fig 6), we demonstrate the practicality of latent-state inference using heuristic rate parameters and illustrate an example of inference for a retinal wave dataset from postnatal day 11 (Stage 3; Fig 7). For retinal wave inference, we normalize the model by population-size (Methods: System-size scaling) so that the gain and bias do not depend on the local neuronal population size.

The state inference (‘data assimilation’) procedure uses new observations to correct for prediction errors. Because of this, many different model parameters may give similar state estimates. Nevertheless, it is important that the rate parameters approximately match the data. The rate of excitation (ρ_e) should be fast, and the rate at which active cells become refractory (ρ_a) should match the typical wave duration. Likewise, it is important that the recovery rate ρ_r matches the inter-wave interval timescale. In Fig 7, model parameters were set based on observed timescales, and then adjusted such that the simulated model dynamics match those recovered during state inference ($\rho_e = 10$, $\rho_a = 1.8 \rho_r = 0.1$, and $\sigma = 0.1$). These parameters were

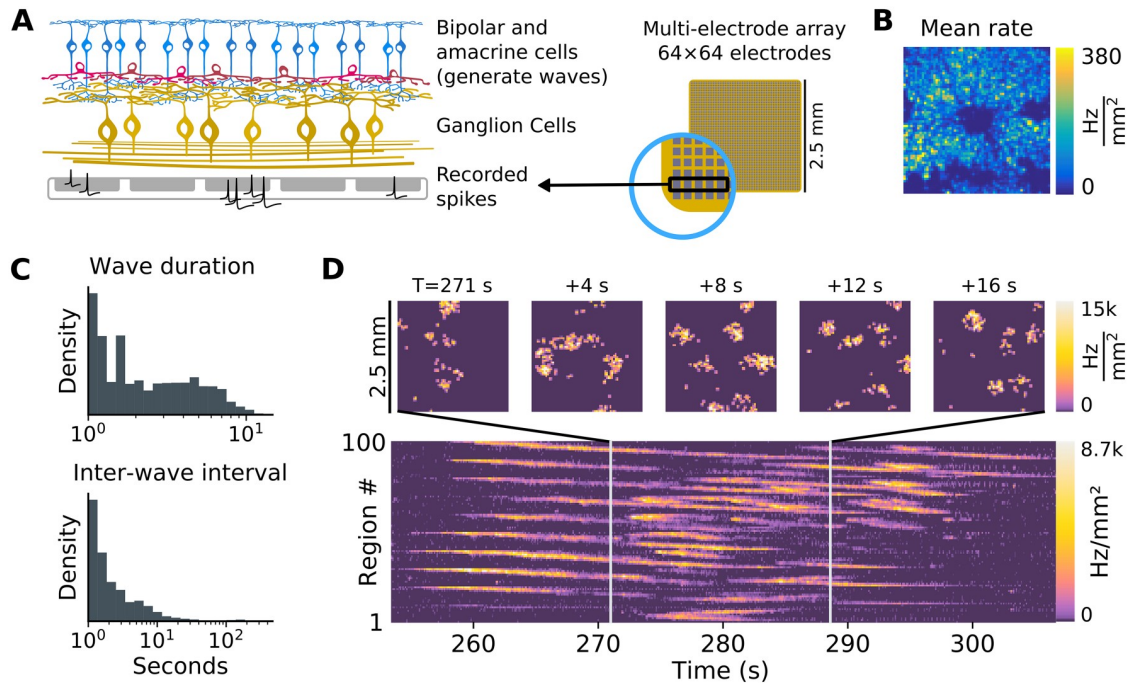


Fig 6. Retinal waves recorded via high-density multi-electrode arrays. (A) Spontaneous retinal waves are generated in the inner retina via laterally interacting bipolar (blue) and amacrine (red) cells, depending on the developmental age. These waves activate retinal ganglion cells (yellow), the output cells of the retina. Electrical activity is recorded from the neonatal mouse retina via a 64×64 -electrode array with $42 \mu\text{m}$ spacing. (B) Average spiking rate recorded across the retina (the central region devoid of recorded spikes is the optic disc). This example was recorded on postnatal day 6. (C) Spikes were binned at 100 ms resolution, and assigned to 10×10 spatial regions for analysis. Spiking activity on each electrode was segmented into “up” states (during wave activity) and “down” states (quiescent) using a two-state hidden Markov model with Poisson observations. In this example, most waves and inter-wave intervals lasted between one and ten seconds. (D) Example wave event, traveling across multiple spatial regions and lasting for a duration of 16-20 seconds.

<https://doi.org/10.1371/journal.pcbi.1007442.g006>

held fixed during subsequent state inference. The interaction radius $\sigma = 0.15$ and excitation strength ρ_e interact to determine how excitable the system is and how quickly waves propagate. The overall excitability should be small enough so that the system is stable, and does not predict wave events in the absence of spiking observations. As in Lansdell et al. [40], lateral interactions in our model reflect an effective coupling that combines both excitatory synaptic interactions and the putative effect of diffusing excitatory neurotransmitters, which has been shown to promote late-stage glutamatergic wave propagation [53].

The moment-closure system does not accurately approximate the rare and abrupt nature of wave initiation. We therefore model spontaneous wave-initiation events as an extrinsic noise source, and set the spontaneous excitation rate ρ_q to zero in the neural field model that defines our latent state-space. The Poisson noise was re-scaled to reflect an effective population size of 16 neurons/mm^2 , significantly smaller than the true population density [58]. However, due to the recurrent architecture and correlated neuronal firing, the effective population size is expected to be smaller than the true population size. Equivalently, this amounts to assuming supra-Poisson scaling of fluctuations for the neural population responsible for retinal waves.

Bayesian filtering recovers the expected features of the retinal waves (Fig 7): the excitatory transition $Q + A \rightarrow A + A$ and the onset of refractoriness $A \rightarrow R$ are rapid compared to the slow refractory dynamics, and therefore the A state is briefly occupied and mediates an effective $Q \rightarrow R$ transition during wave events. The second-order structure provided by the covariance is essential, as it allows us to model posterior variance (shaded regions in Fig 7),

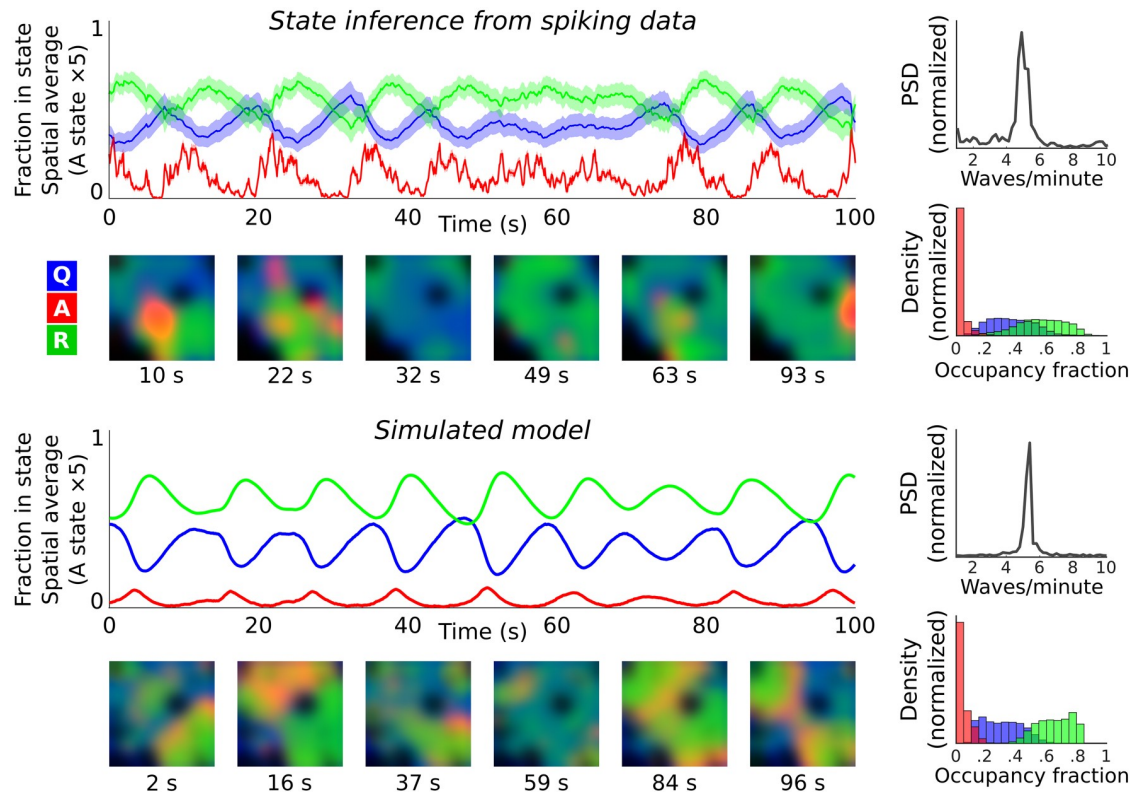


Fig 7. State inference via filtering: Retinal datasets. We apply a calibrated model to spiking observations from retinal waves (postnatal day 11) to infer latent neural-field states. In all plots, red, green, and blue indicate (normalized) densities of active, refractory, and quiescent cells. **(top)** Solid lines indicate inferred spatial means, and shaded regions the 95% confidence bound. The the A state has been scaled-up by $\times 5$. Example time slices are shown in the colored plots below. Dark regions indicate areas absent from the recording. Summary statistics are shown on the right, with power spectra (averaged over all included regions and states) indicating periodic ~ 5 waves/min, and the typical fraction of Q/A/R states, pooled over all times and regions, summarized in histograms below. **(bottom)** Forward simulation of the calibrated model without data recapitulates the retinal wave activity. Solid lines indicate sampled spatial means. Colored plots show example time slices. Wave frequency is comparable to the data (~ 5 waves/min), and occupancy statistics are similar. The model was initialized with 70% of cells quiescent and 30% refractory, with a 25 s burn-in to remove initial transients.

<https://doi.org/10.1371/journal.pcbi.1007442.g007>

while also capturing strong anti-correlations due to the conservation of reacting agents, and the effect of correlated fluctuations on the evolution of the means. Furthermore, spatial correlations allow localized RGC spiking events to be interpreted as evidence of regional (spatially-extended) latent neuronal activity.

Open challenges in model identification

So far, we have demonstrated good recovery of states when the true rate parameters are known (Fig 5), and shown that plausible latent-states can be inferred from neural point-process datasets using *a priori* initialized parameters (Fig 7). A natural question then is whether one can use the Bayesian state-space framework to estimate a posterior likelihood on the rate parameter values, and infer model parameters directly from data. Presently, model inference remains challenging for four reasons: under-constrained parameters, computational complexity, numerical stability, and non-convexity in the joint posterior. It is worth reviewing these challenges as they relate to important open problems in machine learning and data assimilation.

First, the effective population size, the typical fraction of units in quiescent vs. refractory states, and the gain parameter mapping latent activations to spiking, are all important to

setting appropriate rates, and are not accessible from observation of RGC spiking alone. Recovering a physiologically realistic model would require direct measurement or appropriate physiological priors on these parameters. In effect, many equivalent systems can explain the observed RGC spiking activity, a phenomenon that has been termed “sloppiness” in biological systems [59, 60]. Indeed, Hennig et al. [61] show that developmental waves are robust to pharmacological perturbations, suggesting that the retina itself can use different configurations to achieve similar wave patterns. *Second*, although state inference is computationally feasible, parameter inference requires many thousands of state-inference evaluations. A Matlab implementation of state-inference running on a 2.9 GHz 8-core Xeon CPU can process ~ 85 samples/s for a 3-state system on a 10×10 spatial basis. For a thirty-minute recording of retinal wave activity, state inference is feasible, but repeated state inference for parameter inference is impractical. *Third*, model likelihood must be computed recursively, and is subject to loss of numerical accuracy due to back-propagation through time [62–64]. In other words, small errors in the past can have large effects in the future owing to the nonlinear and excitable nature of the system. *Fourth* and finally, the overall likelihood surface need not be convex, and may contain multiple local optima. Additionally, regions of parameters space can exhibit vanishing gradient for one or model parameters. This can occur when the value of one parameter makes others irrelevant. For example, if the excito-excitatory interaction ρ_e is set to a low value, the interaction radius σ_e for excitation becomes irrelevant since the overall excitation is negligible.

Overall, parameter inference via Bayesian filtering presents a formidable technical challenge. Presently, it seems that traditional methods, based on mathematical expertise and matching observable physical quantities (e.g. wavefront speed, c.f. [40]), remain the best-available approach to model estimation. Despite these challenges, the innovation presented here, of applying moment-closure methods for data assimilation, is important *per se*, because it provides a snapshot of the activity of unobserved states which can greatly aid scientific investigation. The state-space formulation of neural field models enables Bayesian state inference from candidate neural field models, and opens the possibility of likelihood-based parameter inference in the future.

Discussion

In this work, we showed that classical neural-field models, which capture the activity of large, interacting neural populations, can be interpreted as state-space models, where we can explicitly model microscopic, intrinsic dynamics of the neurons. This is achieved by interpreting a second-order neural field model as defining equations on the first two moments of a latent-variable process, which is coupled to spiking observations. In the state-space model interpretation, latent neural field states can be recovered from Bayesian filtering. This allows inferring the internal states of neuronal populations in large networks based solely on recorded spiking activity, information that can experimentally only be obtained with whole cell recordings.

We demonstrated successful state inference for simulated data, where the correct model and parameters were known. Next, we applied the model to large-scale recordings of developmental retinal waves. Here the correct latent-state model is unknown, but a relatively simple three-state model with slow refractoriness is well-motivated by experimental observations [57]. Previous works [39, 57, 65, 66] predict that activity-dependent refractoriness is important for restricting the spatial spreading of waves. Intuitively, one should expect the refractory time constant to be a highly sensitive parameter: very long refractory constants will impede the formation of waves, while short constants might lead to interference phenomena. These intuitions were borne out empirically by our simulation studies; additionally, we observed that long

refractory constants led to ineffective data assimilation, as the model prior is too dissimilar from the data it is trained upon. In contrast to phenomenological latent state-space models, the latent states here are motivated by an (albeit simplified) description of single-neuron dynamics, and the state-space equations arise directly from considering the evolution of collective activity as a stochastic process.

In the example explored here, we use Gaussian moment-closure to arrive at a second-order approximation of the distribution of latent states and their evolution. In principle, other distributional assumptions may also be used to close the moment expansion. Other mathematical approaches that yield second-order models could also be employed, for example the linear noise approximation [67], or defining a second cumulant in terms of the departure of the model from Poisson statistics [35]. The approach applied here to a three-state system can generally be applied to systems composed of linear and quadratic state transitions. Importantly, systems with only linear and pairwise (quadratic) interactions can be viewed as a locally-quadratic approximation of a more general smooth nonlinear system [68], and Gaussian moment closure therefore provides a general approach to deriving approximate state-space models in neural population dynamics.

The state-space interpretation of neural field models opens up future work to leverage the algorithmic tools of SSM estimation for data assimilation with spiking point-process datasets. However, challenges remain regarding the retinal waves explored here, and future work is needed to address these challenges. Model likelihood estimation is especially challenging. Despite this, the connection between neural-field models and state-space models derived here will allow neural field modeling to incorporate future advances in estimating recursive, nonlinear, spatiotemporal models. We also emphasize that some of the numerical challenges inherent to high-dimensional spatially extended neural field models do not apply to simpler, low-dimensional neural mass models, and the moment-closure framework may therefore provide a practical avenue to parameter inference in such models.

In summary, this report connects neural field models, which are grounded in models of stochastic population dynamics, to latent state-space models for population spiking activity. This connection opens up new approaches to fitting neural field models to spiking data. We expect that this interpretation is a step toward the design of coarse-grained models of neural activity that have physically interpretable parameters, have physically measurable states, and retain an explicit connection between microscopic activity and emergent collective dynamics. Such models will be essential for building models of collective dynamics that can predict the effects of manipulations on single-cells on emergent population activity.

Materials and methods

Data acquisition and preparation

Example retinal wave datasets are taken from Maccione et al. [37]. Spikes were binned at 100 ms resolution for analysis. Spikes were further binned into regions on a 20×20 spatial grid. For the three-state model, this resulted in a 1200-dimensional spatiotemporal system, which provided an acceptable trade-off between spatial resolution and numerical tractability.

Spiking activity in each region was segmented into wave-like and quiescent states using a two-state hidden Markov model with a Poisson observations. To address heterogeneity in the Retinal Ganglion Cell (RGC) outputs, the observation model was adapted to each spatial region based on firing rates. Background activity was used to establish per-region biases, defined as the mean activity in a region during quiescent periods. The scaling between latent states and firing rate (gain) was adjusted locally based on the mean firing rate during wave

events. The overall (global) gain for the observation model was then adjusted so that no wave events exhibited a fraction of cells in the active (A) state greater than one.

Moment-closure for a single population

To develop a state-space formalism for inference and data assimilation in neural field models, we begin with a master equation approach. This approach has been used before to analyze various stochastic neural population models, often as a starting point to derive ordinary differential equations for the moments of the distribution of population states, as we do here [34–36, 44, 46, 69]. In our case, we examine a three-state system of the kind proposed in Buice and Cowan [44, 45], and use a Gaussian moment-closure approach similar to Bressloff [34].

The master equation describes how the joint probability distribution of neural population states (in our example the active, quiescent and refractory states) evolves in time. However, modelling this full distribution is computationally prohibitive for a spatially-extended system, since the number of possible states scales exponentially with the number of neural populations. Instead, we approximate the time evolution of the moments of this distribution.

In principle, an infinite number of moments are needed to describe the full population activity. To limit this complexity, we consider only the first two moments (mean and covariance), and use a moment-closure approach to close the series expansion of network interactions in terms of higher moments ([47–50]; for applications to neuroscience see [29, 34–36, 69, 70]). Using this strategy, we obtain a second-order neural field model that describes how the mean and covariance of population spiking evolve in time, and recapitulates spatiotemporal phenomena when sampled.

We may describe the number of neurons in each state in terms of a probability distribution $\Pr(Q, A, R)$ (Fig 2A), where we slightly abuse notation and use $Q, A,$ and R both as symbols for the neuron states and as variables counting the neurons in the corresponding states, i.e. non-negative integers. The time evolution of this probability distribution captures stochastic population dynamics, and is represented by a master equation that describes the change in density for a given state $\{Q, A, R\}$ when neurons change states. Accordingly, the master equation describes the change in probability of a given state $\{Q, A, R\}$ in terms of the probability of entering, minus the probability of leaving the state:

$$\begin{aligned} \partial_t \Pr(Q, A, R) &= \Pr(Q, A + 1, R - 1)\rho_a(A + 1) && \text{(transition } A \rightarrow R) \\ &+ \Pr(Q - 1, A, R + 1)\rho_r(R + 1) && \text{(transition } R \rightarrow Q) \\ &+ \Pr(Q + 1, A - 1, R)[\rho_q + \rho_e(A - 1)](Q + 1) && (Q \rightarrow A \text{ and } A + Q \rightarrow A + A) \\ &- \Pr(Q, A, R)[(\rho_e A + \rho_q)Q + \rho_a A + \rho_r R] && \text{(outgoing transitions)} \end{aligned} \tag{5}$$

Even in this simplified non-spatial scenario, no analytic solutions are known for the master equation. However, from Eq (5) one can derive equations for the mean and covariance of the process.

The approach, generally, is to consider expectations of individual states, e.g. $\langle Q \rangle$ (first moments, i.e. means), or $\langle QA \rangle$ (second moments), taken with respect to the probability distribution $\Pr(Q, A, R)$ described by the master Eq (5). Differentiating these moments in time, and substituting in the time-evolution of the probability density as given by the master equation, yields expressions for the time-evolution of the moments. However, in general these expressions will depend on higher moments and are therefore not closed.

For our system, the nonlinear excitatory interaction $Q + A \rightarrow A + A$ couples the evolution of the means to the covariance Σ_{AQ} , and the evolution of the covariance is coupled to the third moment, and so on. The moment equations are therefore not closed, and require an infinite

number of moments to describe the evolution of the mean and covariance. To address this complexity, we approximate $\Pr(Q, A, R)$ with a multivariate normal distribution at each time-point (Fig 2B), thereby replacing counts of neurons with continuous variables. This Gaussian moment-closure approximation sets all cumulants beyond the variance to zero, yielding an expression for the third moment in terms of the mean and covariance, leading to closed ordinary differential equations for the means and covariances [47–50].

For our model with transitions given in Eq (1) this leads to the system of ODEs for the mean values given in Eq (2) in the main text. For the evolution of the covariance we obtain

$$\begin{aligned} \partial_t \Sigma &= \mathbf{J}\Sigma + \Sigma\mathbf{J}^T + \Sigma_{\text{noise}}, \\ \Sigma_{\text{noise}} &= \begin{bmatrix} r_{qa} + r_{rq} & -r_{qa} & -r_{rq} \\ -r_{qa} & r_{qa} + r_{ar} & -r_{ar} \\ -r_{rq} & -r_{ar} & r_{ar} + r_{qa} \end{bmatrix} \\ \mathbf{J} &= \begin{bmatrix} -\rho_q - \rho_e \langle A \rangle & -\rho_e \langle Q \rangle & \rho_r \\ \rho_q + \rho_e \langle A \rangle & \rho_e \langle Q \rangle - \rho_a & 0 \\ 0 & \rho_a & -\rho_r \end{bmatrix} \end{aligned} \tag{6}$$

where \mathbf{J} is the Jacobian of the equations for the deterministic means in Eq (2), and the Σ_{noise} fluctuations are Poisson and therefore proportional to the mean reaction rates (Eq 2). Intuitively, the Jacobian terms \mathbf{J} describe how the covariance of the state distribution ‘stretches’ or ‘shrinks’ along with the deterministic evolution of the means, and the additional Σ_{noise} reflects added uncertainty due to the fact that state transitions are stochastic. Each state experiences Poisson fluctuations with variance equal to the mean transition rates, due to the sum of transitions into and away from the state. Because the number of neurons is conserved, a positive fluctuation into one state implies a negative fluctuation away from another, yielding off-diagonal anticorrelations in the noise.

Together, Eqs (2) and (6) provide approximate equations for the evolution of the first two moments of the master equation (Eq 5), expressed in terms of ordinary differential equations governing the mean and covariance of a multivariate Gaussian distribution. Here, we have illustrated equations for a 3-state system, but the approach is general and can be applied to any system with spontaneous and pairwise state transitions.

Extension to spatial system

To extend the moment Eqs (2) and (6) to a neural field system, we consider a population of neurons at each spatial location. In this spatially-extended case, we denote the intensity fields as \mathbf{Q} , \mathbf{A} , and \mathbf{R} , which are now vectors with spatial indices (or, in the spatially-continuous case: scalar functions of coordinates \mathbf{x}). In the spatially-extended system, active (\mathbf{A}) neurons can excite nearby quiescent (\mathbf{Q}) neurons. We model the excitatory influence of active cells as a weighted sum over active neurons in a local neighborhood, defined by a coupling kernel $K(\Delta\mathbf{x})$ that depends on distance (Eq 4). To simplify the derivations that follow, denote the

convolution integral in Eq (4) as a linear operator \mathbf{K} such that

$$\mathbf{KA} = K(\Delta\mathbf{x}) * A(\mathbf{x}). \tag{7}$$

In this notation, one can think of \mathbf{K} as a matrix that defines excitatory coupling between nearby spatial regions. Using the notation of Eq (7), the rate that active cells excite quiescent ones is given by the product

$$\rho_e(\mathbf{KA}) \circ \mathbf{Q} = \rho_e \text{Diag}(\mathbf{KAQ}^\top), \tag{8}$$

where \circ denotes element-wise (in the spatially-continuous case: function) multiplication. For the time evolution of the first moment (mean intensity) of \mathbf{Q} in the spatial system, one therefore considers the expectation $\langle \mathbf{KAQ}^\top \rangle$, as opposed to $\langle A\mathbf{Q} \rangle$ in the non-spatial system. Since \mathbf{K} is a linear operator, and the extension of the Gaussian state-space model over the spatial domain \mathbf{x} is a Gaussian process, the second moment of the nonlocal interactions \mathbf{KA} with \mathbf{Q} can be obtained in the same way as one obtains the correlation for a linear transformation of a multivariate Gaussian variable:

$$\begin{aligned} \langle \mathbf{KAQ}^\top \rangle &= \mathbf{K} \langle A\mathbf{Q}^\top \rangle \\ &= \mathbf{K}(\Sigma_{A,Q} + \langle A \rangle \langle Q \rangle^\top). \end{aligned} \tag{9}$$

The resulting equations for the spatial means are similar to the nonspatial system (Eq 2), with the exception that we now include spatial coupling in the rate at which quiescent cells enter the active state:

$$\begin{aligned} r_{qa} &= \rho_q \langle Q \rangle + \rho_e \text{Diag}[\langle \mathbf{KAQ}^\top \rangle] \\ &= \rho_q \langle Q \rangle + \rho_e \text{Diag}[\mathbf{K}(\Sigma_{A,Q} + \mathbf{K} \langle A \rangle \langle Q \rangle^\top)] \\ &= \rho_q \langle Q \rangle + \rho_e [\text{Diag}(\mathbf{K}\Sigma_{A,Q}) + \mathbf{K} \langle A \rangle \circ \langle Q \rangle]. \end{aligned} \tag{10}$$

The numbers of neurons in the quiescent versus active states are typically anti-correlated, because a neuron entering the active state implies that one has left the quiescent state. Therefore, the expected number of interactions between quiescent and active neurons is typically smaller than what one might expect from the deterministic mean field alone. The influence of correlations $\text{Diag}(\mathbf{K}\Sigma_{A,Q})$ on the excitation is therefore important for stabilizing the excitatory dynamics.

To extend the equations for the second moment to the neural field case, we consider the effect of spatial couplings on the Jacobian (Eq 6). The spontaneous first-order reactions remain local, and so the linear contributions are similar to the non-spatial case. However, nonlocal interaction terms emerge in the nonlinear contribution to the Jacobian:

$$\mathbf{J}_{\text{nonlinear}} = \rho_e \begin{matrix} \underbrace{\hspace{10em}}_{\text{Gradient in Q}} & \underbrace{\hspace{10em}}_{\text{Gradient in A}} & \\ \begin{bmatrix} -\text{Diag}(\mathbf{K} \langle A \rangle) & -\text{Diag}(\langle Q \rangle \mathbf{K}) & 0 \\ \text{Diag}(\mathbf{K} \langle A \rangle) & \text{Diag}(\langle Q \rangle \mathbf{K}) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \end{matrix} \tag{11}$$

where here the “Diag” operation refers to constructing a diagonal matrix from a vector. Intuitively, the first column of Eq (11) reflects the fact that the availability of quiescent cells

modulates the excitatory effect of active cells, and the second column reflects the fact that the density active of neurons in nearby spatial volumes contribute to the rate at which quiescent cells become active.

Basis projection

The continuous neural field equations are simulated by projection onto a finite spatial basis B . Each basis element is an integral over a spatial volume. Means for each basis element are defined as an integral over this volume, and correlations are defined as a double integral. For example, consider the number of quiescent neurons associated with the i^{th} basis function B_i , which we will denote as Q_i . The mean $\langle Q_i \rangle$ and covariance Σ_{QA}^{ij} between the quiescent and active states are given by the projections:

$$\begin{aligned} \langle Q_i \rangle &= \int B_i(\mathbf{x}) \mathbf{Q}(\mathbf{x}) \, d\mathbf{x} \\ \Sigma_{QA}^{ij} &= \iint B_i(\mathbf{x}) B_j(\mathbf{x}') \Sigma_{QA}(\mathbf{x}, \mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}', \end{aligned} \tag{12}$$

where \mathbf{x} and \mathbf{x}' range over spatial coordinates as in Eqs (3) and (4). When selecting a basis B , assumptions must be made about the minimum spatial scale to model. A natural choice is the radius of lateral (i.e. spatially nonlocal) interactions in the model σ_e (Eq 3), since structure below this scale is attenuated by the averaging over many nearby neurons in the dendritic inputs.

Sampling from the model

For ground-truth simulations, we sample from a hybrid stochastic model derived from a Langevin approximation to the three-state neural field equation. In this approximation, the deterministic evolution of the state is given by the mean-field equations (Eq 2) for a local system, Eq (10) for the neural field system), and the stochastic noise arising from Poisson state transitions is approximated as Gaussian as given by second-order terms (i.e. Σ_{noise} in Eq (6); see also [50, 71]). Spontaneous wave initiation events are too rare to approximate as Gaussian, and instead are sampled as Poisson (shot) noise, giving us a hybrid stochastic model:

$$r_q(t) \sim \text{Poisson}(\rho_q \cdot dt) \cdot \delta(t), \tag{13}$$

where $\delta(t)$ is a Dirac delta (impulse). To avoid uniform spontaneous excitation, the excitatory reaction rate is adjusted by a small finite threshold ϑ , i.e. $r_{qa} \leftarrow \max(0, r_{qa} - \vartheta)$ in Eq (10). For our simulations (e.g. Fig 3), we let $\vartheta = 8 \times 10^{-3}$. For the non-spatial system, the hybrid stochastic differential equation is:

$$\begin{bmatrix} dQ \\ dA \\ dR \end{bmatrix} = \left(\begin{bmatrix} -r_q(t) & 0 & \rho_r \\ r_q(t) & -\rho_a & 0 \\ 0 & \rho_a & -\rho_r \end{bmatrix} \begin{bmatrix} Q \\ A \\ R \end{bmatrix} + \rho_e \begin{bmatrix} -QA \\ QA \\ 0 \end{bmatrix} \right) dt + \Sigma_{\text{noise}}^{1/2} dW, \tag{14}$$

where Σ_{noise} is the fluctuation noise covariance as in Eq (6) (with ρ_q excluded, as it is addressed by the shot noise, Eq 13), and dW is the derivative of a multidimensional standard Wiener process, i.e. a spherical (white) Gaussian noise source. The deterministic component of (14) equation can be compared to Eq (2) for the means of the non-spatial system in the moment-closure system (without the covariance terms).

The stochastic differential equation for the spatial system is similar, consisting to a collection of local populations coupled through the spatial interaction kernel (Eqs 3 and 4), and follows the same derivation used when extending the moment-closure to the spatial case (Methods: *Extension to spatial system*, Eqs 7–10). When applying the Euler-Maruyama method to the spatiotemporal implementation, fluctuations were scaled by $\sqrt{\Delta t \Delta \mathbf{x}}$, where $\Delta \mathbf{x}$ is the volume of the spatial basis functions used to approximate the spatial system (see Methods: *System-size scaling* for further detail). The Euler-Maruyama algorithm samples noise from a Gaussian distribution, and can therefore create negative intensities due to discretization error. We addressed this issue by using the complex Langevin equation [72], which accommodates transient negative states.

Point-process measurement likelihood

Similarly to generalized linear point-process models for neural spiking [10, 11, 51], we model spikes as a Poisson process conditioned on a latent intensity function $\lambda(\mathbf{x}, t)$, which characterises the probability of finding a certain number of spikes k in a small spatiotemporal interval $\Delta \mathbf{x} \times \Delta t$ as:

$$\Pr \left(\int_{t_0}^{t_0+\Delta t} \int_{\mathbf{x}_0}^{\mathbf{x}_0+\Delta \mathbf{x}} y(\mathbf{x}, t) d\mathbf{x} dt = k \right) \sim \text{Poisson} \left(k; \int_{t_0}^{t_0+\Delta t} \int_{\mathbf{x}_0}^{\mathbf{x}_0+\Delta \mathbf{x}} \lambda(\mathbf{x}, t) d\mathbf{x} dt \right). \quad (15)$$

In (15), $y(\mathbf{x}, t)$ denotes the experimentally-observed spiking output, and is a sum over Dirac delta distributions corresponding to each spike with an associated time t_i and spatial location \mathbf{x}_i , i.e. $y(\mathbf{x}, t) = \sum_{i \in 1..N} \delta(\mathbf{x}_i) \delta(t_i)$. We use a linear Poisson likelihood for which the point-process intensity function

$$\lambda(\mathbf{x}, t) = \gamma(\mathbf{x}) \mathbf{A}(\mathbf{x}, t) + \beta(\mathbf{x}) \quad (16)$$

depends linearly on the number of active neurons $\mathbf{A}(\mathbf{x}, t)$ with spatially-varying gain $\gamma(\mathbf{x})$ and bias $\beta(\mathbf{x})$. In other words, the observed firing intensity in a given spatiotemporal volume should be proportional to the number of active neurons, with some additional offset or bias β to capture background spiking unrelated to the neural-field dynamics.

Bayesian filtering

Having established an approach to approximate the time-evolution of the moments of a neural field system, we now discuss how Bayesian filtering allows us to incorporate observations in the estimation of the latent states. Suppose we have measurements y_0, \dots, y_N of the latent state x at time t_0, \dots, t_N , given by a measurement process $\Pr(y_i | x_i)$, which in our case is given by the point-process likelihood (Eq 16). Bayesian filtering allows us to recursively estimate the *filtering distribution* $\Pr(x_i | y_i, \dots, y_0)$ at time t_i , i.e. the posterior state probability at time t_i given the current and all previous observations. The procedure works by the following iterative scheme: i) suppose we know the filtering distribution $\Pr(x_i | y_i, \dots, y_0)$ at time t_i . Solving the model dynamics forward in time up to t_{i+1} gives the predictive distribution $\Pr(x_i | y_i, \dots, y_0)$ for all times $t_i < t \leq t_{i+1}$. ii) at the time t_{i+1} the measurement y_{i+1} needs to be taken into account which can be done by means of the Bayesian update:

$$\Pr(x_{i+1} | y_{i+1}, \dots, y_0) = \frac{\Pr(y_{i+1} | x_{i+1}) \Pr(x_{i+1} | y_i, \dots, y_0)}{\Pr(y_{i+1} | y_i, \dots, y_0)}, \quad (17)$$

where we have used the Markov property and $\Pr(y_{i+1} | x_{i+1}, y_i, \dots, y_0) = \Pr(y_{i+1} | x_{i+1})$ to obtain

the right hand side. Eq (17) gives the filtering $\Pr(x_{t_{i+1}} | y_{i+1}, \dots, y_0)$ at time t_{i+1} which serves as the input of the next i step. Performing steps i) and ii) iteratively hence provides the filtering distribution for all times $t_0 \leq t \leq t_n$.

For our neural field model we must compute both steps approximately: to obtain the predictive distribution in step i) we integrate forward the differential equations for mean and covariance derived from moment-closure (Eqs 2–6 and Methods: Extension to spatial system). In practice, we convert the continuous-time model to discrete time. If $F_{\partial t}$ denotes the local linearization of the mean dynamics in continuous time such that $\partial_t \mu(t) = F_{\partial t} \mu(t)$, then the approximated discrete-time forward operator is

$$F_{\Delta t} = \exp(F_{\partial t} \Delta t) \approx I + F_{\partial t} \Delta t. \tag{18}$$

We update the covariance using this discrete-time forward operator, combined with an Euler integration step for the Poisson fluctuations. A small constant diagonal regularization term Σ_{reg} can be added, if needed, to improve stability. The resulting equations read:

$$\begin{aligned} \mu_{t+\Delta t} &= F_{\Delta t} \mu_t \\ \Sigma_{t+\Delta t} &= F_{\Delta t} \Sigma_t F_{\Delta t}^T + \Sigma_t^{\text{noise}} \cdot \Delta t + \Sigma_{\text{reg}}. \end{aligned} \tag{19}$$

This form is similar to the update for a discrete-time Kalman filter [73, 74], the main difference being that the dynamics between observation times are taken from the nonlinear moment equations.

Consider next the measurement update of step ii) in Eq (17). Since the Gaussian model for the latent states x is not conjugate with the Poisson distribution for observations y , we approximate the posterior $\Pr(x_{t+1} | y_{i+1}, \dots, y_0)$ using the Laplace approximation (c.f. [1, 32]). The Laplace-approximated measurement update is computed using a Newton-Raphson algorithm. The measurement update is constrained to avoid negative values in the latent fields by adding a ϵ/x potential (compare to the log-barrier approach; [27]), which ensures that the objective function gradient points away from this constraint boundary, where x is the intensity of any of the three fields. The gradients and Hessian for the posterior measurement log-likelihood $\ln \mathcal{L}$ are

$$\begin{aligned} -\ln \mathcal{L} &= \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) + \nu(\gamma x + \beta) - y \ln(\gamma x + \beta) \\ -\frac{\partial \ln \mathcal{L}}{\partial x} &= \Sigma^{-1} (x - \mu) + \nu \gamma - y \left(\frac{\gamma}{\gamma x + \beta} \right) \\ -\frac{\partial^2 \ln \mathcal{L}}{\partial x^2} &= \Sigma^{-1} + y \left(\frac{\gamma}{\gamma x + \beta} \right)^2, \end{aligned} \tag{20}$$

where x is the latent state with prior mean μ and covariance Σ , and couples to point-process observations y linearly with gain γ and bias β as in Eq (16). The parameter $\nu = \Delta x^2 \cdot \Delta t$ is the spatiotemporal volume of the basis function or spatial region over which the counts are observed.

System-size scaling

For clarity, the derivations in this paper are presented for a population of neurons with a known size, such that the fields $\mathbf{Q}(\mathbf{x})$, $\mathbf{A}(\mathbf{x})$, and $\mathbf{R}(\mathbf{x})$ have units of *neurons*. In practice, the population size Ω of neurons is unknown, and it becomes expedient to work in normalized intensities, where $\mathbf{Q}(\mathbf{x})$, $\mathbf{A}(\mathbf{x})$, and $\mathbf{R}(\mathbf{x})$ represent the *fraction* of neurons in a given state

between 0 and 1, and are constrained such that $\mathbf{Q}(\mathbf{x}) + \mathbf{A}(\mathbf{x}) + \mathbf{R}(\mathbf{x}) = 1$. In this normalized model for population size Ω , quadratic interaction parameters (like ρ_e) as well as the gain are multiplied by Ω , to reflect the re-scaled population. In contrast, noise variance should be *divided* by Ω to account for the fact that the coefficient of variation decreases as population size increases. Although rescaling by Ω is well-defined for finite-sized populations, the infinitesimal neural-field limit for the second-order model is not. This is because, while the mean-field equations scale with the population size $\mathcal{O}(\Omega)$, the standard deviation of Poisson fluctuations scales with the square root of the population size $\mathcal{O}(\sqrt{\Omega})$. The ratio of fluctuations to the mean (coefficient of variation) therefore scales as $\mathcal{O}(1/\sqrt{\Omega})$, which diverges as $\Omega \rightarrow 0$.

This divergence is not an issue in practice as all numerical simulations are implemented on a set of basis functions with finite nonzero volumes, and each spatial region is therefore associated with finite nonzero population size. Even in the limit where fluctuations would begin to diverge, one can treat the neural field equations as if defined over a continuous set of overlapping basis functions with nonzero volume. Conceptually, this can be viewed as setting a minimum spatial scale for the neural field equations, which is defined by spatial extent of each local population. If the model is defined over a set of overlapping spatial regions, then these populations experience correlated fluctuations. Consider Poisson fluctuations as entering with some rate-density $\sigma^2(\mathbf{x})$ per unit area. The observed noise variances and covariances, projected onto basis functions $B_i(\mathbf{x})$ and $B_j(\mathbf{x})$, are:

$$\Sigma_{ij}^{\text{noise}} = \int B_i(\mathbf{x})B_j(\mathbf{x})\sigma^2(\mathbf{x})d\mathbf{x} \quad (21)$$

If the neuronal population density is given as $\rho(\mathbf{x})$ per unit area, then the effective population size for a given basis function is:

$$\Omega_i = \int B_i(\mathbf{x})\rho(\mathbf{x})d\mathbf{x} \quad (22)$$

If the population density is uniform, and if basis functions have a constant volume ν , we can write this more simply as $\Omega = \nu\rho$. In the system-size normalized model, the contributions of basis function volume cancel and the noise variance should be scaled simply as $1/\rho$.

Acknowledgments

We thank Gerrit Hilgen for important discussions in establishing biologically-plausible parameter regimes for the three-state model. We thank Evelyne Sernagor for the retinal wave datasets, as well as ongoing advice and invaluable feedback on the manuscript. We also thank Dimitris Milios and Botond Cseke for helpful technical discussions.

Author Contributions

Conceptualization: Michael E. Rule, David Schnoerr, Matthias H. Hennig, Guido Sanguinetti.

Formal analysis: Michael E. Rule.

Funding acquisition: Matthias H. Hennig, Guido Sanguinetti.

Investigation: Michael E. Rule, Matthias H. Hennig, Guido Sanguinetti.

Methodology: Michael E. Rule, David Schnoerr, Matthias H. Hennig, Guido Sanguinetti.

Project administration: Matthias H. Hennig, Guido Sanguinetti.

Software: Michael E. Rule.

Supervision: Matthias H. Hennig, Guido Sanguinetti.

Writing – original draft: Michael E. Rule.

Writing – review & editing: Michael E. Rule, David Schnoerr, Matthias H. Hennig, Guido Sanguinetti.

References

1. Paninski L, Ahmadian Y, Ferreira DG, Koyama S, Rad KR, Vidne M, et al. A new look at state-space models for neural data. *Journal of Computational Neuroscience*. 2010; 29(1-2):107–126. <https://doi.org/10.1007/s10827-009-0179-x> PMID: 19649698
2. Yuan Zhao, Il Memming Park. Variational latent gaussian process for recovering single-trial dynamics from population spike trains. *Neural computation*. 2017; 29(5):1293–1316. MIT Press.
3. Yuan Zhao, Il Memming Park. Variational joint filtering. *arXiv preprint arXiv:1707.09049*. 2017;.
4. Sussillo D, Jozefowicz R, Abbott L, Pandarinath C. LFADS-Latent Factor Analysis via Dynamical Systems. *arXiv preprint arXiv:160806315*. 2016;.
5. Aghagolzadeh M, Truccolo W. Inference and decoding of motor cortex low-dimensional dynamics via latent state-space models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2016; 24(2):272–282. <https://doi.org/10.1109/TNSRE.2015.2470527> PMID: 26336135
6. Linderman SW, Tucker A, Johnson MJ. Bayesian latent state space models of neural activity. *Computational and Systems Neuroscience (Cosyne) Abstracts*. 2016;.
7. Gao Y, Archer EW, Paninski L, Cunningham JP. Linear dynamical neural population models through nonlinear embeddings. In: *Advances in Neural Information Processing Systems*; 2016. p. 163–171.
8. Paninski L. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*. 2004 jan; 15(4):243–262. https://doi.org/10.1088/0954-898X_15_4_002
9. Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky E, et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*. 2008; 454(7207):995. <https://doi.org/10.1038/nature07140> PMID: 18650810
10. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*. 2005; 93(2):1074–1089. <https://doi.org/10.1152/jn.00697.2004> PMID: 15356183
11. Truccolo W. From point process observations to collective neural dynamics: Nonlinear Hawkes process GLMs, low-dimensional dynamics and coarse graining. *Journal of Physiology-Paris*. 2016; 110(4):336–347. <https://doi.org/10.1016/j.jphysparis.2017.02.004>
12. Amari Si. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*. 1977; 27(2):77–87. <https://doi.org/10.1007/bf00337259> PMID: 911931
13. Wilson HR, Cowan JD, Baker T, Cowan J, van Drongelen W. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal*. 1972 jan; 12(1):1–24. [https://doi.org/10.1016/S0006-3495\(72\)86068-5](https://doi.org/10.1016/S0006-3495(72)86068-5) PMID: 4332108
14. Cowan J. A Personal Account of the Development of the Field Theory of Large-Scale Brain Activity from 1945 Onward. In: *Neural Fields*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 47–96.
15. Bressloff PC. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical*. 2012 jan; 45(3):033001. <https://doi.org/10.1088/1751-8113/45/3/033001>
16. Durstewitz D, Seamans JK, Sejnowski TJ. Neurocomputational models of working memory. *Nature Neuroscience*. 2000; 3(11s):1184. <https://doi.org/10.1038/81460> PMID: 11127836
17. Zhang H, Xiao P. Seizure Dynamics of Coupled Oscillators with Epileptor Field Model. *International Journal of Bifurcation and Chaos*. 2018; 28(03):1850041. <https://doi.org/10.1142/S0218127418500414>
18. Proix T, Jirsa VK, Bartolomei F, Guye M, Truccolo W. Predicting the spatiotemporal diversity of seizure propagation and termination in human focal epilepsy. *Nature Communications*. 2018; 9(1):1088. <https://doi.org/10.1038/s41467-018-02973-y> PMID: 29540685
19. González-Ramírez L, Ahmed O, Cash S, Wayne C, Kramer M. A biologically constrained, mathematical model of cortical wave propagation preceding seizure termination. *PLoS Computational Biology*. 2015; 11(2):e1004065–e1004065. <https://doi.org/10.1371/journal.pcbi.1004065> PMID: 25689136
20. Martinet LE, Fiddym G, Madsen J, Eskandar E, Truccolo W, Eden UT, et al. Human seizures couple across spatial scales through travelling wave dynamics. *Nature Communications*. 2017; 8:14896. <https://doi.org/10.1038/ncomms14896> PMID: 28374740

21. Ermentrout GB, Cowan JD. A mathematical theory of visual hallucination patterns. *Biological Cybernetics*. 1979; 34(3):137–150. <https://doi.org/10.1007/bf00336965> PMID: 486593
22. Bressloff PC, Cowan JD, Golubitsky M, Thomas PJ, Wiener MC. Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2001; 356(1407):299–330. <https://doi.org/10.1098/rstb.2000.0769> PMID: 11316482
23. Rule M, Stoffregen M, Ermentrout B. A model for the origin and properties of flicker-induced geometric phosphenes. *PLoS Comput Biol*. 2011; 7(9):e1002. <https://doi.org/10.1371/journal.pcbi.1002158>
24. Moran R, Pinotsis DA, Friston K. Neural masses and fields in dynamic causal modeling. *Frontiers in Computational Neuroscience*. 2013; 7:57. <https://doi.org/10.3389/fncom.2013.00057> PMID: 23755005
25. Bojak I, Oostendorp TF, Reid AT, Kötter R. Connecting Mean Field Models of Neural Activity to EEG and fMRI Data. *Brain Topography*. 2010 jun; 23(2):139–149. <https://doi.org/10.1007/s10548-010-0140-3> PMID: 20364434
26. Pinotsis DA, Moran RJ, Friston KJ. Dynamic causal modeling with neural fields. *Neuroimage*. 2012 jan; 59(2):1261–1274. <https://doi.org/10.1016/j.neuroimage.2011.08.020> PMID: 21924363
27. Nazarpour K, Ethier C, Paninski L, Rebesch JM, Miall RC, Miller LE. EMG Prediction From Motor Cortical Recordings via a Nonnegative Point-Process Filter. *IEEE Transactions on Biomedical Engineering*. 2012 jul; 59(7):1829–1838. <https://doi.org/10.1109/TBME.2011.2159115> PMID: 21659018
28. Schnoerr D, Grima R, Sanguinetti G. Cox process representation and inference for stochastic reaction-diffusion processes. *Nature Communications*. 2016 may; 7:11729. <https://doi.org/10.1038/ncomms11729> PMID: 27222432
29. Rule ME, Sanguinetti G. Autoregressive Point-Processes as Latent State-Space Models: a Moment-Closure Approach to Fluctuations and Autocorrelations. *Neural Computation*. 2018; 30(10):2757–2780. https://doi.org/10.1162/neco_a_01121 PMID: 30148704
30. Cseke B, Zammit-Mangion A, Heskes T, Sanguinetti G. Sparse approximate inference for spatio-temporal point process models. *Journal of the American Statistical Association*. 2016; 111(516):1746–1763. <https://doi.org/10.1080/01621459.2015.1115357>
31. Zammit-Mangion A, Dewar M, Kadirkamanathan V, Sanguinetti G. Point process modelling of the Afghan War Diary. *Proceedings of the National Academy of Sciences of the United States of America*. 2012 jul; 109(31):12414–9. <https://doi.org/10.1073/pnas.1203177109> PMID: 22802667
32. Macke JH, Buesing L, Cunningham JP, Byron MY, Shenoy KV, Sahani M. Empirical models of spiking in neural populations. In: *Advances in Neural Information Processing Systems*; 2011. p. 1350–1358.
33. Smith AC, Brown EN. Estimating a state-space model from point process observations. *Neural Computation*. 2003; 15(5):965–991. <https://doi.org/10.1162/089976603765202622> PMID: 12803953
34. Bressloff PC. Stochastic neural field theory and the system-size expansion. *SIAM Journal on Applied Mathematics*. 2009; 70(5):1488–1521. <https://doi.org/10.1137/090756971>
35. Buice MA, Cowan JD, Chow CC. Systematic fluctuation expansion for neural network activity equations. *Neural Computation*. 2010; 22(2):377–426. <https://doi.org/10.1162/neco.2009.02-09-960> PMID: 19852585
36. Touboul JD, Ermentrout GB. Finite-size and correlation-induced effects in mean-field dynamics. *Journal of computational neuroscience*. 2011; 31(3):453–484. <https://doi.org/10.1007/s10827-011-0320-5> PMID: 21384156
37. Maccione A, Hennig MH, Gandolfo M, Muthmann O, Copenhagen J, Eglen SJ, et al. Following the ontogeny of retinal waves: pan-retinal recordings of population dynamics in the neonatal mouse. *The Journal of physiology*. 2014; 592(7):1545–1563. <https://doi.org/10.1113/jphysiol.2013.262840> PMID: 24366261
38. Meister M, Wong R, Baylor DA, Shatz CJ. Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science*. 1991; 252(5008):939–943.
39. Hennig MH, Adams C, Willshaw D, Sernagor E. Early-stage waves in the retinal network emerge close to a critical state transition between local and global functional connectivity. *Journal of Neuroscience*. 2009; 29(4):1077–1086. <https://doi.org/10.1523/JNEUROSCI.4880-08.2009> PMID: 19176816
40. Lansdell B, Ford K, Kutz JN, Rokhsar D, Pleniz D. A Reaction-Diffusion Model of Cholinergic Retinal Waves. *PLOS Computational Biology*. 2014 dec; 10(12):e1003953. <https://doi.org/10.1371/journal.pcbi.1003953> PMID: 25474327
41. Wilson HR, Cowan JD. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*. 1972; 12(1):1–24. [https://doi.org/10.1016/S0006-3495\(72\)86068-5](https://doi.org/10.1016/S0006-3495(72)86068-5) PMID: 4332108
42. Wilson HR, Cowan JD. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*. 1973; 13(2):55–80. <https://doi.org/10.1007/BF00288786> PMID: 4767470

43. Hennig MH, Adams C, Willshaw D, Sernagor E. Early-Stage Waves in the Retinal Network Emerge Close to a Critical State Transition between Local and Global Functional Connectivity. *Journal of Neuroscience*. 2009; 29(4). <https://doi.org/10.1523/JNEUROSCI.4880-08.2009>
44. Buice MA, Cowan JD. Field-theoretic approach to fluctuation effects in neural networks. *Physical Review E*. 2007; 75(5):051919. <https://doi.org/10.1103/PhysRevE.75.051919>
45. Buice MA, Cowan JD. Statistical mechanics of the neocortex. *Progress in Biophysics and Molecular Biology*. 2009 feb; 99(2-3):53–86. <https://doi.org/10.1016/j.pbiomolbio.2009.07.003> PMID: 19695282
46. Ohira T, Cowan JD. Master-equation approach to stochastic neurodynamics. *Physical Review E*. 1993; 48(3):2259. <https://doi.org/10.1103/PhysRevE.48.2259>
47. Goodman LA. Population growth of the sexes. *Biometrics*. 1953; 9(2):212–225. <https://doi.org/10.2307/3001852>
48. Whittle P. On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1957; 19(2):268–281.
49. Gomez-Uribe CA, Verghese GC. Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *The Journal of chemical physics*. 2007; 126(2):024109. <https://doi.org/10.1063/1.2408422> PMID: 17228945
50. Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *Journal of Physics A: Mathematical and Theoretical*. 2017; 50(9):093001. <https://doi.org/10.1088/1751-8121/aa54d9>
51. Truccolo W, Hochberg LR, Donoghue JP. Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nature Neuroscience*. 2010; 13(1):105–111. <https://doi.org/10.1038/nn.2455> PMID: 19966837
52. Sernagor E, Young C, Eglén SJ. Developmental modulation of retinal wave dynamics: shedding light on the GABA saga. *Journal of Neuroscience*. 2003; 23(20):7621–7629. <https://doi.org/10.1523/JNEUROSCI.23-20-07621.2003> PMID: 12930801
53. Blankenship AG, Ford KJ, Johnson J, Seal RP, Edwards RH, Copenhagen DR, et al. Synaptic and extrasynaptic factors governing glutamatergic retinal waves. *Neuron*. 2009; 62(2):230–241. <https://doi.org/10.1016/j.neuron.2009.03.015> PMID: 19409268
54. Zhou ZJ, Zhao D. Coordinated transitions in neurotransmitter systems for the initiation and propagation of spontaneous retinal waves. *Journal of Neuroscience*. 2000; 20(17):6570–6577. <https://doi.org/10.1523/JNEUROSCI.20-17-06570.2000> PMID: 10964962
55. Feller MB, Wellis DP, Stellwagen D, Werblin FS, Shatz CJ. Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. *Science*. 1996; 272(5265):1182–1187. <https://doi.org/10.1126/science.272.5265.1182> PMID: 8638165
56. Bansal A, Singer JH, Hwang BJ, Xu W, Beaudet A, Feller MB. Mice lacking specific nicotinic acetylcholine receptor subunits exhibit dramatically altered spontaneous activity patterns and reveal a limited role for retinal waves in forming ON and OFF circuits in the inner retina. *Journal of Neuroscience*. 2000; 20(20):7672–7681. <https://doi.org/10.1523/JNEUROSCI.20-20-07672.2000> PMID: 11027228
57. Zheng J, Lee S, Zhou ZJ. A transient network of intrinsically bursting starburst cells underlies the generation of retinal waves. *Nature Neuroscience*. 2006; 9(3):363. <https://doi.org/10.1038/nn1644> PMID: 16462736
58. Jeon CJ, Strettoi E, Masland RH. The major cell populations of the mouse retina. *Journal of Neuroscience*. 1998; 18(21):8936–8946. <https://doi.org/10.1523/JNEUROSCI.18-21-08936.1998> PMID: 9786999
59. Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, Sethna JP. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*. 2015; 143(1):07B201_1. <https://doi.org/10.1063/1.4923066>
60. Panas D, Amin H, Maccione A, Muthmann O, van Rossum M, Berdondini L, et al. Sloppiness in spontaneously active neuronal networks. *Journal of Neuroscience*. 2015; 35(22):8480–8492. <https://doi.org/10.1523/JNEUROSCI.4421-14.2015> PMID: 26041916
61. Hennig MH, Grady J, van Copenhagen J, Sernagor E. Age-dependent homeostatic plasticity of GABAergic signaling in developing retinal networks. *Journal of Neuroscience*. 2011; 31(34):12159–12164. <https://doi.org/10.1523/JNEUROSCI.3112-11.2011> PMID: 21865458
62. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. In: *International Conference on Machine Learning*; 2013. p. 1310–1318.
63. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*. 1994; 5(2):157–166. <https://doi.org/10.1109/72.279181> PMID: 18267787

64. Hochreiter S, Bengio Y, Frasconi P, Schmidhuber J, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. *A field guide to dynamical recurrent neural networks*. IEEE Press; 2001.
65. Feller MB, Butts DA, Aaron HL, Rokhsar DS, Shatz CJ. Dynamic processes shape spatiotemporal properties of retinal waves. *Neuron*. 1997; 19(2):293–306. [https://doi.org/10.1016/s0896-6273\(00\)80940-x](https://doi.org/10.1016/s0896-6273(00)80940-x) PMID: [9292720](https://pubmed.ncbi.nlm.nih.gov/9292720/)
66. Godfrey KB, Swindale NV. Retinal wave behavior through activity-dependent refractory periods. *PLoS Computational Biology*. 2007; 3(11):e245. <https://doi.org/10.1371/journal.pcbi.0030245> PMID: [18052546](https://pubmed.ncbi.nlm.nih.gov/18052546/)
67. Van Kampen NG. *Stochastic processes in physics and chemistry*. vol. 1. Elsevier; 1992.
68. Ale A, Kirk P, Stumpf MP. A general moment expansion method for stochastic kinetic models. *The Journal of Chemical Physics*. 2013; 138(17):174101. <https://doi.org/10.1063/1.4802475> PMID: [23656108](https://pubmed.ncbi.nlm.nih.gov/23656108/)
69. El Boustani S, Destexhe A. A master equation formalism for macroscopic modeling of asynchronous irregular activity states. *Neural computation*. 2009; 21(1):46–100. <https://doi.org/10.1162/neco.2009.02-08-710> PMID: [19210171](https://pubmed.ncbi.nlm.nih.gov/19210171/)
70. Ly C, Tranchina D. Critical analysis of dimension reduction by a moment closure method in a population density approach to neural network modeling. *Neural computation*. 2007; 19(8):2032–2092. <https://doi.org/10.1162/neco.2007.19.8.2032> PMID: [17571938](https://pubmed.ncbi.nlm.nih.gov/17571938/)
71. Riedler MG, Buckwar E. Laws of large numbers and langevin approximations for stochastic neural field equations. *The Journal of Mathematical Neuroscience*. 2013; 3(1):1. <https://doi.org/10.1186/2190-8567-3-1> PMID: [23343328](https://pubmed.ncbi.nlm.nih.gov/23343328/)
72. Schnoerr D, Sanguinetti G, Grima R. The complex chemical Langevin equation. *The Journal of Chemical Physics*. 2014; 141(2):07B606_1. <https://doi.org/10.1063/1.4885345>
73. Kalman RE, et al. Contributions to the theory of optimal control. *Boletín de la Sociedad Matemática Mexicana*. 1960; 5(2):102–119.
74. Kalman RE, Bucy RS. New results in linear filtering and prediction theory. *Journal of Basic Engineering*. 1961; 83(1):95–108. <https://doi.org/10.1115/1.3658902>