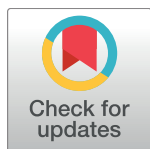# On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs

Felipe Gerhard[1], Moritz Deger[2,3], Wilson Truccolo[1,4,5]*

1 Department of Neuroscience, Brown University, Providence, Rhode Island, United States of America,
2 School of Computer and Communication Sciences and School of Life Sciences, Brain Mind Institute, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, 3 Institute for Zoology, Faculty of Mathematics and Natural Sciences, University of Cologne, Cologne, Germany, 4 Institute for Brain Science, Brown University, Providence, Rhode Island, United States of America, 5 Center for Neurorestoration & Neurotechnology, U. S. Department of Veterans Affairs, Providence, Rhode Island, United States of America

* Wilson_Truccolo@brown.edu

Abstract

Point process generalized linear models (PP-GLMs) provide an important statistical framework for modeling spiking activity in single-neurons and neuronal networks. Stochastic stability is essential when sampling from these models, as done in computational neuroscience to analyze statistical properties of neuronal dynamics and in neuro-engineering to implement closed-loop applications. Here we show, however, that despite passing common goodness-of-fit tests, PP-GLMs estimated from data are often unstable, leading to divergent firing rates. The inclusion of absolute refractory periods is not a satisfactory solution since the activity then typically settles into unphysiological rates. To address these issues, we derive a framework for determining the existence and stability of fixed points of the expected conditional intensity function (CIF) for general PP-GLMs. Specifically, in nonlinear Hawkes PP-GLMs, the CIF is expressed as a function of the previous spike history and exogenous inputs. We use a mean-field quasi-renewal (QR) approximation that decomposes spike history effects into the contribution of the last spike and an average of the CIF over all spike histories prior to the last spike. Fixed points for stationary rates are derived as self-consistent solutions of integral equations. Bifurcation analysis and the number of fixed points predict that the original models can show stable, divergent, and metastable (fragile) dynamics. For fragile models, fluctuations of the single-neuron dynamics predict expected divergence times after which rates approach unphysiologically high values. This metric can be used to estimate the probability of rates to remain physiological for given time periods, e.g., for simulation purposes. We demonstrate the use of the stability framework using simulated single-neuron examples and neurophysiological recordings. Finally, we show how to adapt PP-GLM estimation procedures to guarantee model stability. Overall, our results provide a stability framework for data-driven PP-GLMs and shed new light on the stochastic dynamics of state-of-the-art statistical models of neuronal spiking activity.

Author summary

Earthquakes, gene regulatory elements, financial transactions, and action potentials produced by nerve cells are examples of sequences of discrete events in space or time. In many cases, such events do not appear independently of each other. Instead, the occurrence of one event changes the rate of upcoming events (e.g, aftershocks following an earthquake). The nonlinear Hawkes process is a statistical model that captures these complex dependencies. Unfortunately, for a given model, it is hard to predict whether stochastic samples will produce an event pattern consistent with observations. In particular, with positive feedback loops, the process might diverge and yield unrealistically high event rates. Here, we show that an approximation to the mathematical model predicts dynamical properties, in particular, whether the model will exhibit stable and finite rates. In the context of neurophysiology, we find that models estimated from experimental data often tend to show metastability or even unstable dynamics. Our framework can be used to add constraints to data-driven estimation procedures to find the optimal model with realistic event rates and help to build more robust models of single-cell spiking dynamics. It is a first step towards studying the stability of large-scale nonlinear spiking neural network models estimated from data.

Introduction

Point-process generalized linear models (PP-GLMs) have become an important approach in the statistical modeling of neurophysiological responses from single nerve cells and their interactions in neural circuits [1–8]. A specific class of PP-GLMs are nonlinear Hawkes processes [9, 10]. In this case, each action potential (spike) modulates the firing intensity of the neurons in the future. Nonlinear Hawkes PP-GLMs can capture the major canonical dynamics of single neurons [11–13] and as phenomenological models avoid the many issues that arise in the specification of biophysically detailed neuronal models [14, 15]. In this way, nonlinear Hawkes PP-GLMs are also important phenomenological models for the simulation and study of large-scale neuronal network models of brain function. However, nonlinear Hawkes PP-GLMs also lead to non-renewal point process spike train statistics because contributions to the intensity from many previous spikes can accumulate over arbitrary time scales [16, 17]. This raises the question of whether such models will produce stable, stationary dynamics in simulations, or whether firing rates will diverge or settle into unphysiological rates depending on a specified absolute refractory period.

For the linear Hawkes point process model, stability can be assessed by calculating the integral of the spike-history kernel, i.e., the effect that each spike has on subsequent activity of the same cell [18–20]. However, for the prevalent nonlinear case, no such practical criterion is currently available. Main stability results established by Brémaud and Massoulié are too restrictive for our applications [9, 21]. Furthermore, model parameters are typically estimated from data using maximum-likelihood methods [2]. For linear autoregressive processes, it is well known that maximum-likelihood estimates can lead to unstable dynamics [22, 23]. We expect this to be even more severe in nonlinear models.

Here, we first show that PP-GLMs estimated from physiological data might not generate spike train realizations that match even simple statistics such as mean firing rates of the original data. Instead, firing rates tend to diverge to the maximum firing rate that is allowed in the presence of an absolute refractory period. Firing patterns like this would typically be

considered unphysiological. This can happen in spite of the models passing commonly used goodness-of-fit tests based on, for example, the time-rescaling theorem [24, 25].

To address the above stability issues, we propose an approximative framework to derive stochastic stability conditions for PP-GLMs. For a stochastically stable point process, the state of the point process stochastically evolves in time, but can be described by a stable and time-invariant distribution of "states" (sample paths), resulting in a stationary point process [9]. In contrast, a process that is not stochastically stable may show similar stochastic dynamics for some time, but eventually its state may diverge and never return. In neural point processes, this scenario is typically associated with a divergence of the firing rate. In case of an actual divergence there is no stationary distribution of states. Note, however, that when considering absolute refractory periods, the divergence of the firing rate and associated internal states of the point process are limited. Firing with an inter-event interval equal to the refractory period is nonphysiological, and the state distribution in this mode of firing is singular. We call this a diverged state in slight deviance from the usual terminology. Our use of the term stochastic stability, however, is in line with the definition from stochastic dynamical systems, but generalized to stochastic point processes (see also [9 Remark 4]).

Our approach to derive stability conditions for PP-GLMs is based on a recently developed mean-field theory of neural dynamics [26, 27]. The approach relies on the following steps. First, we use a quasi-renewal (QR) approximation that decomposes spike history effects into the contribution of the last spike and an average of the conditional intensity function (CIF) over all spike histories prior to the most recent spike. Second, after truncation of a moment-based expansion, this decomposition leads to a tractable expression for the approximated CIF. Third, under stationarity conditions, fixed points can be derived as self-consistent solutions of an integral equation, which correspond to expected steady-state firing rates of the neuron. Fourth, depending on the number and stability of these fixed points, each single-neuron model can be unambiguously classified into one of three types: stable, divergent, or fragile. The latter corresponds to metastability which results from stochastic fluctuations perturbing the dynamics in the presence of multiple stable fixed points and when the upper fixed point corresponds to a stable but unphysiologically high firing rate. Fluctuations around the low-rate fixed point of the network dynamics predict an expected time horizon until rates will converge to the high-rate state. This expected time metric can be used to estimate the probability of firing rates to remain finite for a given time period. Examination of the stability of the fixed points and how it depends on the shape of the spike-history filter not only determines the stability of the stochastic dynamics, but also leads to general stability constraints on PP-GLM parameters.

In the following sections, we present our framework to assess the stability of a specific neuron model in detail. We validate the QR approximation in comparison to results of numerical simulations for a large range of artificial neuron models that are neurophysiologically plausible. We then apply the method to real-world data sets. Finally, we demonstrate and discuss how parameter estimation procedures could be adapted to ensure stability of estimated models.

As stated above, stability of neuron models is particularly important when numerical simulations are desired or spike trains are to be generated from the model. Generated spike trains can be used to assess model goodness-of-fit and to perform forecasting of neural activity over longer time scales. Our results are a first step towards ensuring stability for recurrently connected neural network models. These models can be put in the framework of multivariate nonlinear Hawkes models, and our classification framework conceptually translates to the multivariate case. Stable (or stabilized) neuron models play an important role in the field of computational neuroscience, especially in the simulation of large-scale models of brain

function. They are also important in neuroengineering, where neuron models are embedded in hybrid and closed-loop applications [28, 29].

Results

Estimated data-driven PP-GLMs can diverge, despite passing goodness-of-fit tests

The nonlinear Hawkes process is a point process model that is commonly used to describe neurophysiological responses. It defines the conditional intensity function (instantaneous firing rate) of a neuron as a nonlinear function of previous spiking activity (Fig 1A):

$$\lambda(t|\mathcal{H}_t) = \phi(h(t)), \tag{1}$$

where $\mathcal{H}_t$ denotes the spiking history up to time t and $\phi(x)$ is a non-negative nonlinear function. The term $h(t)$ consists of a constant offset I_0 and a convolution of the spike train $S(t)$ with (temporal) spike-history kernels or filters $\eta(s)$:

$$h(t) = I_0 + [\eta * S](t) = I_0 + \sum_{k=1}^{K} \eta(t - t_k), \tag{2}$$

where the $\{t_k\} \in \mathcal{H}_t$ correspond to the previous spike times (see "Materials and Methods" for details). Based on both theoretical and empirical arguments [2, 7], we set $\phi(x) = \exp(x)$ to
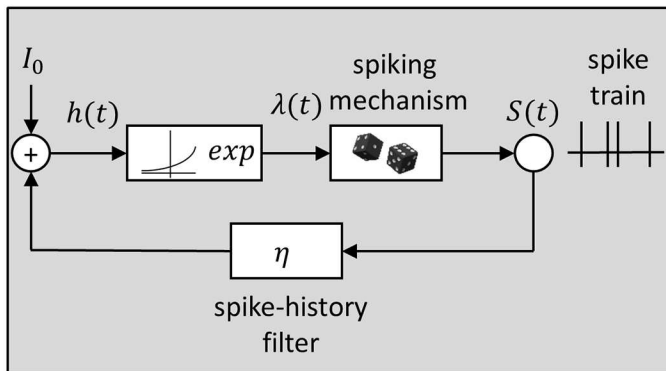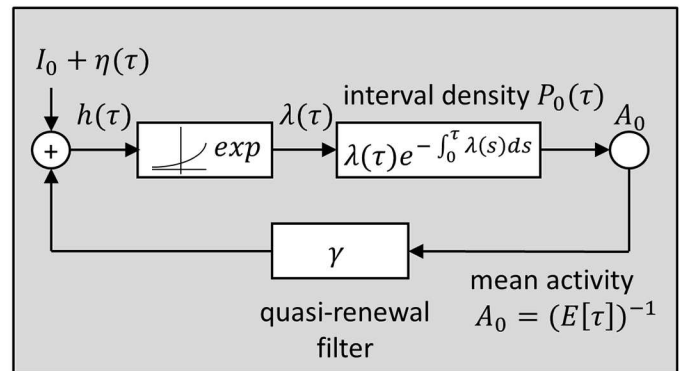

Fig 1. Schematic overview of the nonlinear Hawkes process and the quasi-renewal (QR) approximation. The quasi-renewal approximation can be used to semi-analytically obtain steady-state firing rates of general, nonlinear Hawkes processes (PP-GLMs). Left: In the nonlinear Hawkes process, the conditional intensity of the point process, $\lambda(t|\mathcal{H}_t)$, is a function of the whole spiking history (see Eq (3)). It is modeled as a nonlinear function (here, an exponential function) of a linear convolution of the previous spike history with a spike-history filter $\eta(s)$ plus a constant offset I_0. The dependence of the instantaneous firing rate on all previous spikes results in a non-renewal process. There are no closed-form solutions for even the first-order statistics of general, nonlinear Hawkes processes. Right: In the quasi-renewal approximation, the conditional intensity is modeled as a combination of the effect of the most recent spike t_i and a term involving the average over the whole spike history before the most recent spike (see Eq (5)). This term includes the average firing activity in the past $A_0 \equiv A(t - s)$, which is filtered with the quasi-renewal filter $\gamma(s)$ and added to the spike-history filter $\eta(\tau)$ of the most recent spike at $t - \tau$. This predicts the instantaneous inter-spike interval density $P_0(\tau)$ from which the steady-state firing rate can be obtained as the inverse of the expected inter-spike interval $E[\tau]$. The self-consistent solutions for which an assumed average history of A_0 leads to an equivalent predicted steady-state rate are fixed points of the transfer function defined in the quasi-renewal approximation.

arrive at:

$$\lambda(t|\mathcal{H}_t) = \exp\left(I_0 + \sum_{k=1}^{K} \eta(t - t_k)\right) = c \exp\left(\sum_{k=1}^{K} \eta(t - t_k)\right), \quad (3)$$

with $c = e^{I_0} > 0$.

Every previous spike contributes a spike-history kernel, and effects of all previous spikes accumulate. This leads to, in general, a non-renewal point process model. The model parameters that describe the kernel $\eta(s)$ and the baseline firing rate c can be estimated using maximum-likelihood optimization within the framework of generalized linear models (GLMs) [2, 4].

As stated earlier, these point-process GLMs (PP-GLMs) were recently shown to be able to describe all major canonical dynamics of single neurons [11–13] and, thus, can serve as a canonical class of mathematically tractable models to describe general single-neuron spiking activity. For example, in [6], we analyzed spiking data from the stomatogastric nervous system of the crab. Neurons that are part of the pyloric network fire in stereotypical, rhythmic activity patterns (Fig 2A). Estimated PP-GLMs from this physiological dataset pass common goodness-of-fit tests such as based on residual analysis or the time-rescaling theorem. We created stochastic realizations of spike trains based on the model. These spike trains reproduce the observed burst pattern of the training data, and in a complete network simulation, the relative phases of the overall pyloric rhythm [6].

However, such simulations of spike trains from data-driven PP-GLMs do not always result in physiological spiking patterns. For example, when estimating PP-GLMs from single-unit data recorded from the neocortex of human epileptic patients [30], goodness-of-fit tests are generally passed, but simulated activity tends to diverge to unphysiologically high firing rates. One data set and estimated model are shown in Fig 2B.

In a more comprehensive analysis, we fitted PP-GLMs to spiking data from neurons recorded from motor-related cortical areas in the monkey [31]. We restricted data to a one-second steady-state movement preparation period of the trial. This period was roughly stationary since it did not include, by design, firing rate transients driven by sensory stimuli or movement execution. For 35 out of the 99 data-driven models, we find that simulated spike trains have finite divergence times. Some of these models diverge in simulations even when goodness-of-fit tests are passed. We show two examples in Fig 2C and 2D. Qualitatively similar results are obtained for all other models.

Instability is also observed when the simulation is performed using a nonlinearity that grows less rapidly than the exponential. We generated stochastic realizations of spike trains using two additional nonlinearities (a linear rectifier, and $f(x) = \log(1 + e^x)$, a smooth interpolation between an exponential and linear function, S1 Fig). Both functions are globally dominated by the exponential function (see S1A Fig). Potentially, nonstable behavior could be observed when simulating with the exponential nonlinearity but not with the two less rapidly accelerating nonlinearities. However, we find that even in these cases, firing rates diverged for the same data sets as presented in Fig 2B and 2C.

In summary, while PP-GLMs estimated from data may serve well in encoding and decoding analysis that require one-step spike prediction conditioned on actually observed spike history and may pass goodness-of-fit tests, they tend to be poor generative models because of the lack of stochastic stability. The use of PP-GLMs as generative models, however, is essential when statistical analyses of spike trains generated by the model are required, or when long-term prediction of future spiking states in single-neuron and neuronal networks is used in neural decoding or closed-loop interventions.
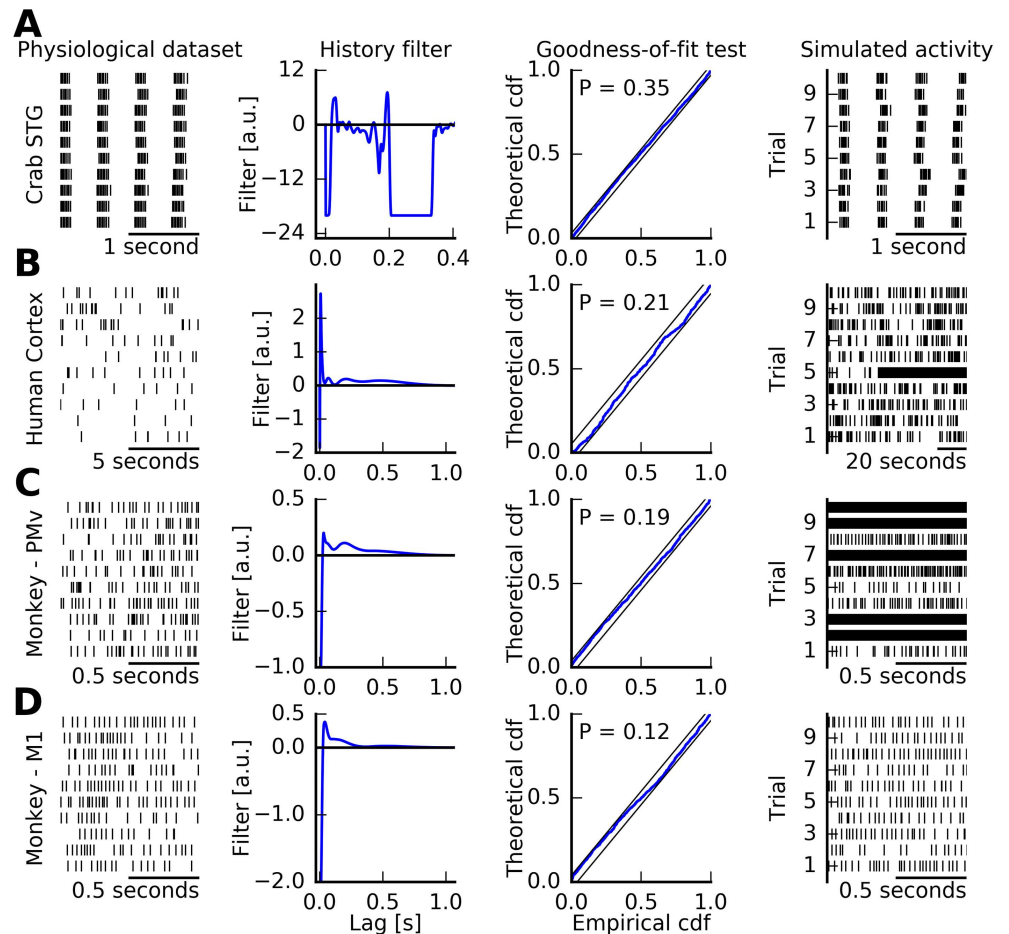
Fig 2. Point-process models estimated from physiological data can pass common goodness-of-fit tests, but simulated activity may diverge. (A) Neurons in the stomatogastric ganglion (STG) of the crab show rhythmic bursts of spike patterns. Each line shows a random 2-second segment of the data from one neuron aligned to the first spike of a burst. The spike-history filter is estimated following the procedure in [6]. The neuron model passes commonly used goodness-of-fit tests, such as those based on the time-rescaling theorem [24, 25]. Here, the Kolmogorov-Smirnov test is shown for rescaled inter-spike intervals to come from an exponential distribution with unit mean. The null hypothesis that observed spikes are coming from the estimated model is not rejected ($P > 0.05$). When sampling spike trains from the model, the model regenerates the rhythmic, bursty activity that is qualitatively matched to the training data. (B) Similar analysis for single-unit activity from neocortical recordings in a person with pharmacologically intractable focal epilepsy [30]. Each line corresponds to a random ten-second segment of spontaneous activity during interictal periods, i.e., outside seizures. The estimated spike-history filter shows a refractory period and an excitatory rebound. The model passes commonly used goodness-of-fit tests ($P > 0.05$). When stochastic samples are generated from the model, spiking activity diverges to a periodic firing pattern at the maximally allowed frequency given the absolute refractory period (here, 2 ms). For some sampled realizations, this divergence can happen very early in the simulated trial (e.g., trial 5). Therefore, simulated activity from the model is unphysiological. It does not match statistics of the spike train in the training data (mean firing rate, inter-spike interval statistics) despite passing the goodness-of-fit test. (C, D) Additional examples of single-unit activity from monkey cortex, areas PMv and M1 [31, 32]. Each line represents a steady-state movement preparation period preceding visual cues leading to execution of reach and grasp actions. Although spike-history filters appear typical in both examples, and goodness-of-fit tests are passed, simulated activity diverges into unphysiological firing rates in one case (first example) and remains physiological in the other.

doi:10.1371/journal.pcbi.1005390.g002

To our knowledge, the stability of PP-GLMs estimated from data has not been systematically examined before. In the next sections, we will develop a framework to assess the dynamics and stability of stochastic spiking neuron models.

A framework to assess stability and dynamics of stochastic spiking neuron models

PP-GLMs have post-spike filters which typically make the spike train probability depend on many previous spikes. These dynamics are in contrast to the conditional intensity function of a renewal point process which depends only on the very last spike time. Therefore, PP-GLMs are generally not renewal processes. When assessing dynamics and stability of PP-GLMs, we are interested in the behavior of the corresponding firing rates. For such general PP-GLMs, however, there are no closed-form solutions for even simple statistical features, such as expected mean firing rates or second-order statistics. Here, to obtain estimates of such statistical features for a given nonlinear Hawkes process, we employ an approximation based on a recently introduced quasi-renewal approximation [26, 27].

The quasi-renewal approximation (Fig 1B) consists of approximating the (non-renewal) PP-GLM by a process which is nearly a renewal-process (hence, "quasi-renewal") that depends on the last spike time and on the average firing rate in the past [26]. Specifically, we consider the steady-state conditional intensity $\lambda_0(t, \hat{t})$ at time t as the average intensity over all possible spike histories that share the most recent spike at time $\hat{t}$:

$$
\begin{aligned}
\lambda_0(t, \hat{t}) &= \langle \lambda(t|\mathcal{H}_t) \rangle_{S(t<\hat{t})} \\
&= c\exp(\eta(t - \hat{t}))\langle \exp([\eta * S](t)) \rangle_{S(t<\hat{t})}.
\end{aligned}
\tag{4}
$$

The first term explicitly models the effect of the most recent spike only, and the second term represents the average of the spiking activity prior to the time of the last spike in the steady-state regime. It can be approximated by (see "Materials and Methods" for details):

$$
\langle \exp([\eta * S](t)) \rangle_{S(t<\hat{t})} \approx \exp \left\{ A_0 \int_{t-\hat{t}}^{\infty} \underbrace{(e^{\eta(u)} - 1)}_{\gamma(u)} du \right\},
\tag{5}
$$

with $\gamma(u) = e^{\eta(u)} - 1$ for the exponentiated spike-history kernel. Here, A_0 is the steady-state firing rate of the process. Intuitively, the convolution of the actual spike train $S(t)$ with $\eta(s)$ is replaced by the convolution of a homogeneous Poisson process spike history of intensity $A(t - u) \equiv A_0$ with an effective filter $\gamma(u)$ (Fig 1B). Since the convolution is applied to a constant A_0, the term reduces to a product of A_0 and the integral of $\gamma(u)$ with the lower bound dependent on $\tau = t - \hat{t}$. Using this approximation, we obtain a quasi-renewal CIF by combining Eqs (4) and (5). Given the QR-CIF, Eq (4), we can then as for ordinary renewal processes, derive the steady-state survivor function $\mathcal{S}_0$ as:

$$
\mathcal{S}_0(\tau) = \exp \left\{ -\int_0^{\tau} \lambda_0(u) du \right\}.
\tag{6}
$$

$\mathcal{S}_0$ then yields the steady-state probability density P_0 of the inter-spike intervals:

$$
P_0(\tau) = \mathcal{S}_0(\tau)\lambda_0(\tau).
\tag{7}
$$

The inverse of the expected inter-spike interval must equal the firing rate f which thus is an

implicit function of A_0 through Eq (5):

$$f(A_0) = \left[\int_0^\infty \tau\, P_0(\tau)d\tau \right]^{-1}. \tag{8}$$

In effect, the QR theory derives a transfer function $f(A_0) > 0$ that links an assumed average spike history to a predicted firing rate. Assuming stationarity, $f(A_0)$ has to match A_0 which leads to a fixed-point equation. Intersections of $f(A_0)$ with the identity correspond to expected fixed points of the dynamics. Stable fixed points in the quasi-renewal approximation predict steady-state firing rates of nonlinear Hawkes processes (PP-GLMs).

   We first show how the number and stability of fixed points of the derived transfer function $f(A_0)$ for the nonlinear Hawkes process endowed with an absolute refractory period can be used to classify the dynamical behavior of the single-neuron model (Fig 3):

- **No fixed point**: If there is no stable fixed point (i.e., $f(A_0) > A_0$ for all A_0), activity will diverge after a finite transient. However, this cannot happen in the presence of an absolute
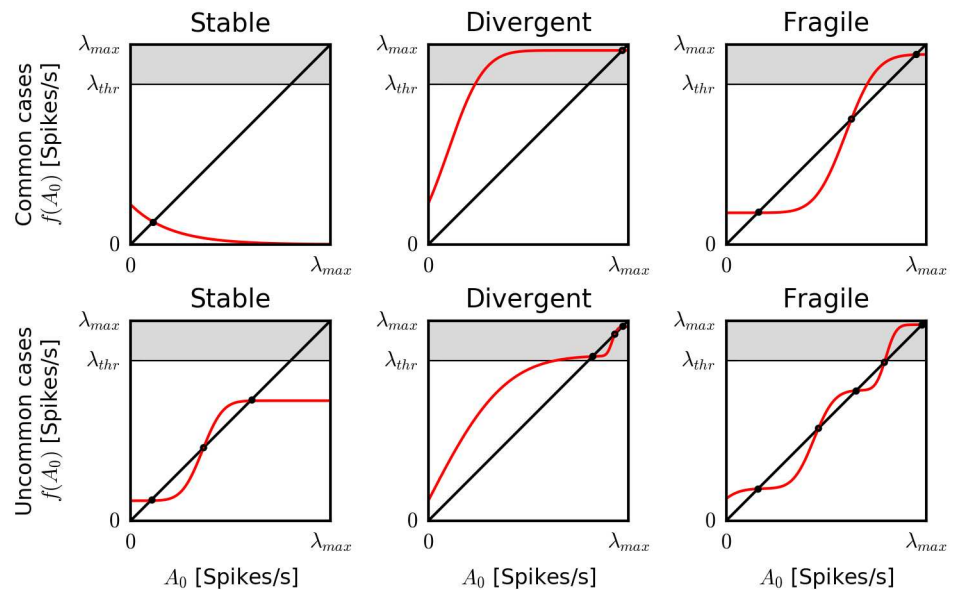
Fig 3. Stochastic dynamics of nonlinear Hawkes PP-GLMs endowed with absolute refractory periods: Classification based on transfer functions in the quasi-renewal approximation. The quasi-renewal approximation provides a predicted firing rate of a neuron model $f(A_0)$ based on an assumed average firing rate in the past A_0 (see "Materials and Methods"). This defines an iterative equation whose fixed points represent the steady-state firing rates. A qualitative classification of the dynamical behaviors is based on the location and stability of the fixed points. Note that throughout the study, we are assuming that there exists an absolute refractory period. Thereby, the maximum firing rate of any model is limited by a maximal firing rate λ_{max}. We define a steady-state firing rate to be unphysiological if it exceeds $\lambda_{thr} = 0.9 \times \lambda_{max}$ (gray area). Given an absolute refractory period, there always exists at least one stable fixed point. If it is the only one and below λ_{thr}, the model is classified as stable (top, left). If the only stable fixed point is above λ_{thr}, the model is divergent (top, center). If there are two stable fixed points, one above and one below λ_{thr}, the model is classified as "fragile" (metastable), indicating that the (physiological) low-rate fixed point is only transiently stable. Expected divergence times $E[T_{div}]$ will depend on the distance between the fixed points. To provide a complete classification framework, we also need to consider the case of two or more stable fixed points, although the latter case seems to be rarely encountered in our experience. In case of two or more stable fixed points below the threshold, the model is classified as stable (bottom, left). Its dynamics is predicted to be multi-stable with steady-state rates fluctuating around two fixed points. If all stable fixed points lie above the threshold the model is considered "divergent" (bottom, center). Any case for which there are multiple stable fixed points both below and above λ_{thr} are considered "fragile" (bottom, right).

refractory period τ_{ref}. The maximal firing rate of the neuron is bounded by $\lambda_{max} = 1/\tau_{\text{ref}}$. In that case, there is always at least one stable fixed point. Throughout the rest of the paper, we assume the existence of such an absolute refractory period and set $\tau_{\text{ref}} = 2$ ms.

- **One stable fixed point**: Any perturbation in the firing rate will eventually decay back to the steady-state rate. If the steady-state firing rate predicted by the fixed point is close to the mean firing rate of the training data, the rate is considered physiological and we classify the neuron model as "stable" (Fig 3 top, left). Conversely, if the fixed point is above $\lambda_{thr} = 0.9 \times \lambda_{max}$, we define here the rate as unphysiologically high and classify the model as "divergent" (Fig 3 top, center).

- **Two stable fixed points**: More than one stable fixed point leads to multi-stable dynamics. Due to the continuity of the transfer function, the first fixed point is necessarily stable, followed by an unstable fixed point, and so on, in alternating fashion. In general, in stochastic multi-stable dynamics, activity will remain around one state for some time before fluctuations drive it towards a different stable fixed point (metastability). Depending on the location of the stable fixed points, such multi-stable activity can be classified as "stable" (both fixed points smaller than λ_{thr}; Fig 3 bottom, left) or "divergent" (both fixed points larger than λ_{thr}; Fig 3 bottom, center). Interesting dynamics emerges when one fixed point is within the range of physiological rates and the other one is close to the "divergent", maximal rate (Fig 3 top, right). In that case, if activity is initialized in the lower state, it will remain in the lower state for a finite time before switching to the unphysiological, high-rate state. We call these metastable dynamics "fragile" to emphasize that such models may produce realistic firing rates in simulations when initialized in the low state but will ultimately escape the stable fixed point and diverge to unphysiological high-firing activity. By $\mathrm{E}[T_{\text{div}}]$, we denote the expected time to transition from the low-rate state to the high-rate state for the first time. Expected divergence times $\mathrm{E}[T_{\text{div}}]$ will depend on the distance between the fixed points and the location of the unstable fixed point.

- **Three or more stable fixed points**: Dynamics can be classified analogously to the previous cases based on how many fixed points are above and below λ_{thr} (Fig 3 bottom, right). However, more than two stable fixed points seem rare in our datasets and simulations. Although they might be constructed by a deliberate choice of the spike-history filter $\eta(s)$, all models that we estimated from physiological data turned out to have at most two stable fixed points.

Overall, the above classification of the qualitative stochastic dynamics suggests a general framework to assess stability and dynamics of stochastic spiking neuron models (Fig 4). In the particular case of data-driven models, training data are used to estimate parameters of a nonlinear Hawkes model (PP-GLM) through (regularized) maximum-likelihood optimization. As shown above, we find empirically that simulating spike trains from these models often yields unphysiological spiking patterns, and firing rates may diverge (Figs 2 and 4, top). We can use the quasi-renewal approximation to analyze the stability of the estimated neuron model. This approximation predicts the dynamics of the neuron model and distinguishes three qualitatively different dynamical behaviors.

Stable or fragile models with high expected divergence times can be safely used to generate stochastic samples from the model. For divergent models or fragile models with low expected divergence times, stabilization constraints can be added to the maximum-likelihood optimization problem to constrain the feasible parameter space to non-divergent models. In any of the three cases, the model (or its stabilized variant) is evaluated based on standard model selection and goodness-of-fit tests before any inference is made (Fig 4, bottom).
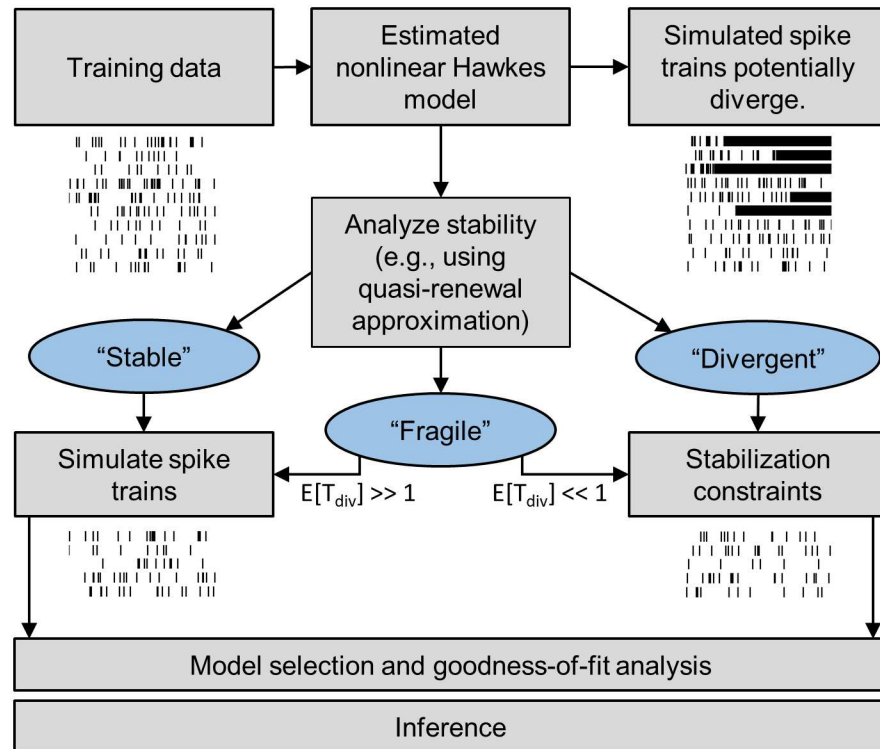
Fig 4. A framework to assess stability and dynamics of stochastic spiking neuron models. Stability of models estimated from physiological data is analyzed using the quasi-renewal approximation. When stable models are desired for sampling or simulations, stability constraints can be included in model estimation. Three types of dynamics can be distinguished: First, "stable" models have steady-state firing rates that are in the physiological range. Spike trains can be safely generated from the model. Second, "divergent" models have a steady-state firing rate that is very close to the maximally allowed firing rate. In this case, stabilization constraints can be added to the maximum-likelihood optimization problem to constrain the feasible parameter space to non-divergent models. Finally, model dynamics can be classified as "fragile", indicating metastable dynamics. While there is a steady-state firing rate at physiological firing rates, there are additional steady-state rates at unphysiological high rates. A simulation that is started with physiological initial conditions may remain in the low-rate regime for a while, but will ultimately visit the unphysiological rate. The framework may provide an estimate of the expected escape or "divergence" time, $E[T_{div}]$. Depending on $E[T_{div}]$, the model can be effectively treated as "stable" or "divergent" based on the typical time scales that would be relevant for simulation. In any of the three cases, the model (or its stabilized variant) is evaluated based on standard model selection and goodness-of-fit tests before any inference is made.

In the next sections, we present the application of this framework to simulated and real data to demonstrate its validity and utility in modeling electrophysiological responses. First, we demonstrate the proposed method for PP-GLMs that have a spike-history filter that consists of either a single exponential or a sum of two exponentials before moving on to filters estimated from neurophysiological data.

Predicted stability reflects simulation outcomes for exponential spike-history filters

We start with the analysis of a simple PP-GLM with a spike-history filter given by a single exponential and an absolute refractory period. The complete model is given by:

$$\lambda(t|\mathcal{H}_t) = ce^{[\eta * S](t)},$$
$$\eta(s) = J\theta(s)e^{-s/\tau} + J_{ref}\theta(s)\theta(\tau_{ref} - s),$$
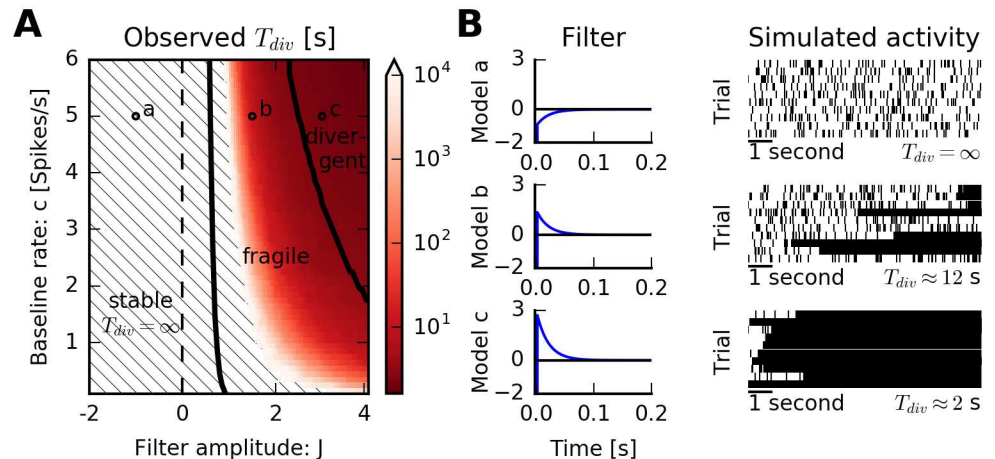
Fig 5. Predicted stability reflects simulation outcomes: Exponential spike-history filter. (A) Spike trains are simulated from a nonlinear Hawkes model with baseline firing rate c and an auto-history kernel $\eta(s) = J \exp(-s/\tau)$ with amplitude J, time constant $\tau = 0.02$ s, and absolute refractory period $\tau_{ref} = 2$ ms. The QR approximation qualitatively predicts three classes of dynamics (separated by thick lines). Color indicates average divergence times estimated from simulations. In the dashed region, no finite divergence times were observed. (B) An example of a spike-history filter for each of the dynamics is shown together with a spike raster of simulated activity. Purely inhibitory filters produce stable dynamics (top, $J = -1$, $c = 5$ s^{-1}). For fragile models, after an episode of irregular firing, dynamics switches into a tonic firing mode close to the limit frequency $1/\tau_{ref}$ (middle, $J = 1$, $c = 5$ s^{-1}). For divergent models, the only stable fixed point is close to the limit frequency (bottom, $J = 3$, $c = 5$ s^{-1}). After a brief transient, the model switches to the tonic firing mode at limit frequency. All dynamics observed in simulations are in accordance with the prediction from the QR approximation.

with Heaviside function $\theta(x)$ and parameters $\tau_{ref} = 2$ ms, $J_{ref} = -\infty$ (-10^{12} in numerical simulations), $\tau = 20$ ms, and amplitude J.

We scanned the two-dimensional parameter space given by the amplitude J of the filter and the baseline firing rate c (Fig 5A; $-2 \leq J \leq 4$ with 121 equally spaced samples and $0.1 \leq c \leq 6.0$ s^{-1} with 60 equally-spaced samples). The QR approximation predicts three regimes of dynamical activity: For slightly positive and negative kernel amplitude J, the model is stable. Indeed, in simulations, we observe finite and stable rates (Fig 5B, top row). For higher amplitudes, the dynamics are predicted to be fragile and ultimately divergent (for large J and c). As expected, the average divergence time estimated from numerical simulations gradually decreases with increasing J and c (color-coded in Fig 5A). Divergent models are almost instantly diverging (Fig 5B, bottom row), while for fragile models, a whole spectrum of divergence times is observed (Fig 5B, middle row). Within the variance given by the finite number of simulated models, we did not observe any discrepancies between the behavior predicted by the QR approximation and the numerical results.

For all models of the parameter range that were classified as stable, we compared the predicted steady-state firing rate to the one observed in numerical simulations (Fig 6A). In this case, the QR approximation provides an excellent prediction of mean firing rates (Pearson's correlation coefficient $\rho > 0.999$).

A major feature of the QR approximation is to predict (an upper bound on) the expected divergence times for fragile models. In practice, this is relevant for model sampling via simulation where it is important to classify fragile models as "effectively stable" or divergent (see Fig 4). For high firing rates close to $A_0 = \tau_{ref}^{-1}$, the regular spike train with inter-spike intervals around τ_{ref} is the only possible spike train realization. Therefore, one way to estimate $E[T_{div}]$ is
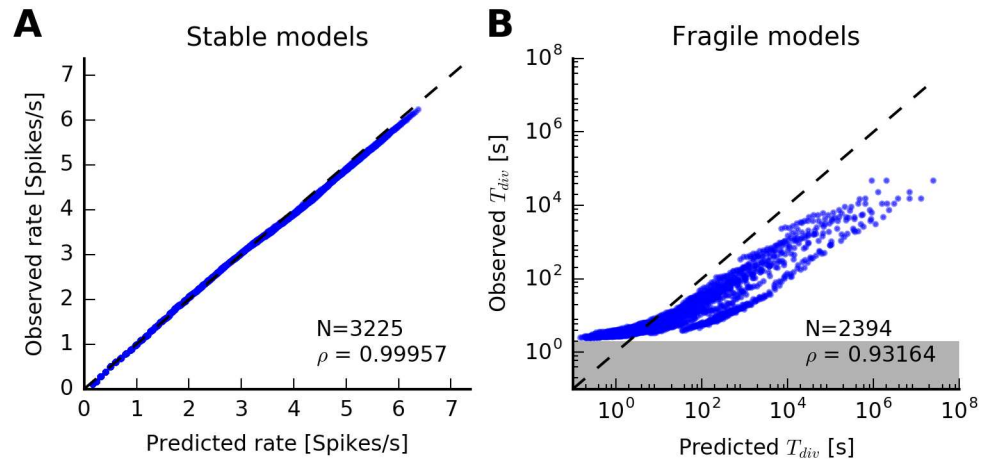
Fig 6. Rate and time to divergence: Exponential spike-history filter. (A) Observed steady-state firing rate versus predicted steady-state firing rate for every model classified as "stable" in Fig 5. Each dot corresponds to a model. Dashed line indicates equality. Linear correlation coefficient is $\rho = 0.9996$. (B) Observed versus predicted divergence times for all models classified as "fragile" in Fig 5. Note the logarithmic axes. The QR approximation provides an approximation of the divergence times. The divergence time of a simulation was defined as the end of the first two-second interval in which the average rate exceeded λ_{thr}. Therefore, estimated divergence times cannot be below 2 s (gray area).

to consider periodic spike histories with different frequencies that would lead to self-sustained periodic firing at maximal rate with high probability (see "Materials and Methods" for details).

Fig 6B compares predicted versus observed divergence times for all fragile models. The predicted T_{div} provide an upper bound on the observed divergence times. The divergence time of a simulation was defined as the end of the first two-second interval in which the average rate exceeded λ_{thr}. For this reason, estimated divergence times cannot be below 2 s (Fig 6B, gray area). Therefore, small estimated divergence times do not obey the predicted bound. However, there seems to be a reasonable (power-law) dependence between predicted and observed T_{div} (Pearson's correlation coefficient $\rho = 0.925$).

We now look at more complex PP-GLMs to test the validity of our proposed framework. We consider spike-history filters consisting of a sum of two exponentials with amplitudes J_r, J_a, and corresponding time constants τ_r and τ_a:

$$\lambda(t|\mathcal{H}_t) = ce^{[\eta*S](t)},$$
$$\eta(s) = J_r\theta(s)e^{-s/\tau_r} + J_a\theta(s)e^{-s/\tau_a} + J_{\text{ref}}\theta(s)\theta(\tau_{\text{ref}} - s).$$

Depending on the signs of the amplitudes, this model resembles many plausible single-neuron behaviors: $J_r < 0$ indicates a relative refractory period beyond the absolute 2 ms refractory period while $J_r > 0$ promotes bursty dynamics. Similarly, $J_a \neq 0$ can be interpreted as inhibitory or facilitating adaptation (e.g., spike-frequency adaptation [16, 17, 33]).

We evaluated models on a wide range of combinations of amplitudes J_r and J_a (Fig 7A; $-11 \leq J_r \leq 11$ with 100 equally spaced samples and $-3 \leq J_a \leq 3$ with 75 equally-spaced samples), for $\tau_r = 20$ ms and $\tau_a = 100$ ms, respectively, and for fixed $c = 5\ s^{-1}$. As expected, negative and positive but small values of J_r and J_a lead to stable dynamics. For a narrow band, models are expected to be fragile (Fig 7B, top row). This observation is consistent with divergence times estimated from numerical simulations. Finally, larger values of either J_a or J_r lead to divergent models, although in an asymmetric way (Fig 7A and 7B, bottom row). Qualitatively similar results are obtained for other values of the baseline c (S2 and S3 Figs).
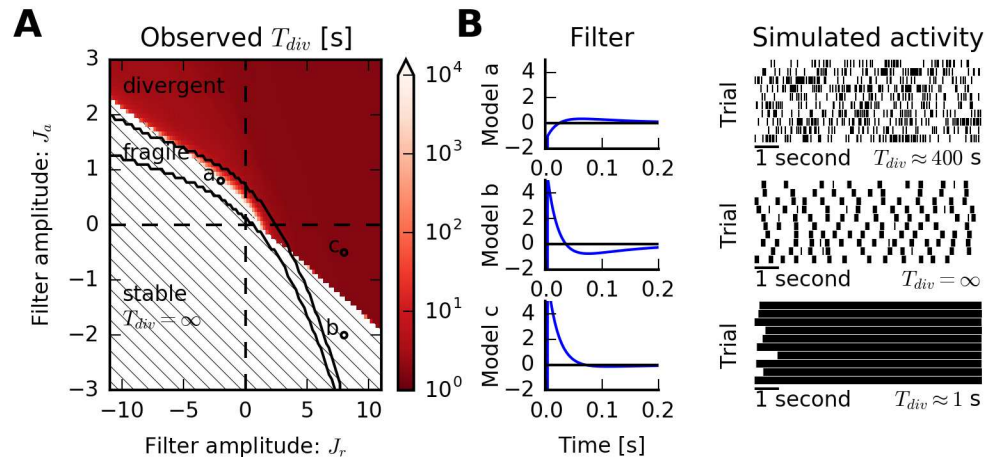
Fig 7. Predicted stability reflects simulation outcomes: Sum-of-exponentials spike-history filter. (A) Spike trains are simulated from a nonlinear Hawkes model with fixed baseline $c = 5$ s^{-1} and an auto-history kernel consisting of two exponentials with amplitudes J_r, J_a, and corresponding time constants $\tau_r = 0.02$ s and $\tau_a = 0.1$ s. Observed divergence times for simulated spike trains are color-coded (same scale as in Fig 5). In the dashed region, no finite divergence times were observed. (B) Auto-history kernels for three different parameter values (models a, b and c), displaying irregular, bursty and divergent dynamics, respectively, consistent with the prediction of the QR-approximation except for model b (see main text).

doi:10.1371/journal.pcbi.1005390.g007

Estimated divergence times are generally consistent with the qualitative prediction of the QR approximation with one exception: For multiphasic spike-history filters, i.e. either strongly refractory neurons ($J_r \ll 0$) with excitatory rebounds ($J_a \gg 0$) or the opposite ($J_r \gg 0$ and $J_a \ll 0$), the QR approximation predicts divergent models, but simulations indicate that rates remain below the threshold λ_{thr} to be classified as divergent (Fig 7A, upper left and lower right corners; Fig 7B, middle row). Spike trains generated from models with these parameters tend to produce intermittent bursts. This is a condition for which the quasi-renewal approximation is known to become invalid [26]. The dynamics lead to the divergent state where the model neuron fires initially at maximally allowed firing rate (hence, unphysiological). However, the dynamics escape this high-firing rate fixed point after a finite number of such high-rate bursts, and the activity reverts back to the low-rate state. Averaged over a longer time period, the mean activity stays well below λ_{thr} and therefore, $\hat{T}_{\mathrm{div}} = \infty$ in contrast to the dynamics predicted by the QR approximation. The reason for this discrepancy is that the QR approximation assumes homogeneous, Poisson-like firing prior to the last spike time, while the only way to achieve firing rates close to the maximally allowed rate by the absolute refractory period is to have a highly regular spike train. For highly regular spike histories, the QR approximation does not provide valid fixed points of the dynamics. However, these cases can be captured by an analysis of the regular spiking limit (see "Bursting and the regular spiking limit") which does not explicitly depend on the QR approximation.

A less severe limitation of the QR approximation is visible in the comparison between predicted and simulated steady-state firing rates for models classified as stable (Fig 8A). While most rates are accurately estimated (points near the diagonal line), very bursty neurons have higher firing rates in simulations than predicted by the QR approximation (stable models with $J_r \gg 0$ and $J_a \ll 0$). This is due to dependencies beyond the last spike that are ignored in the approximation but are non-negligible for burst firing. Loosely speaking, in this case the QR approximation predicts the rate of isolated spikes and *bursts*, but not the number of total spikes. The burst duration can be predicted from a simple criterion based on the conditional
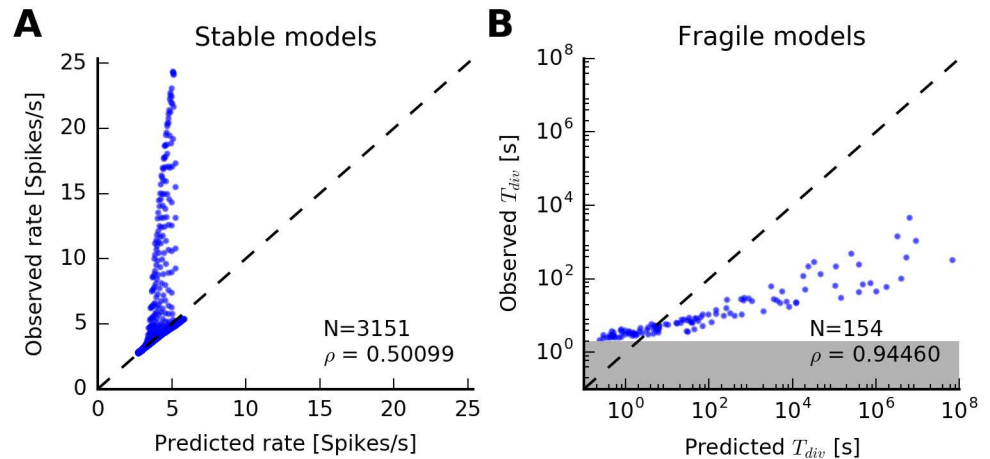
Fig 8. Rate and time to divergence: Sum-of-exponentials spike-history filter. (A) Observed steady-state firing rate versus predicted steady-state firing rate for every model classified as "stable" in Fig 7. Each dot corresponds to a model. Dashed line indicates equality. Linear correlation coefficient is $\rho = 0.50$. While steady-state rates are correctly predicted for most models, discrepancies arise due to limiting assumptions of the QR approximations when $J_r \gg 0$ and $J_a \ll 0$. However, the qualitative prediction of the dynamics as stable remains correct. (B) Observed versus predicted divergence times for all models classified as "fragile" in Fig 7. Note the logarithmic axes. Dashed line indicates diagonal. The QR approximation provides an approximation of the divergence times. The divergence time of a simulation was defined as the end of the first two-second interval in which the average rate exceeded λ_{thr}. Therefore, estimated divergence times cannot be below 2 s (gray area). The data suggest a power-law dependence between predicted and observed divergence times (Pearson's correlation coefficient $\rho = 0.94$).

doi:10.1371/journal.pcbi.1005390.g008

intensity function of the PP-GLM, independent of the QR approximation (see "Bursting and the regular spiking limit").

Finally, predicted and observed divergence times for all fragile models are well approximated by the QR approximation (Fig 8B) except for small divergence times whose estimation is biased due to the finite time window to detect divergence in numerical simulations.

In summary, the QR approximation yielded remarkably accurate predictions of the dynamical behavior of PP-GLMs for most parameter settings. When the steady-state rate was not accurately predicted for bursty neurons, the qualitative prediction was still consistent with simulations. For extreme parameter values, we observed intermittent burst activity that was incorrectly predicted to be divergent. Although not divergent according to the definition of Fig 3, the resulting spiking pattern would nevertheless be considered unphysiological and undesirable in modeling applications. Thus, in this case, the discrepancy between the QR prediction and the simulation does not play a significant role in practice.

So far, we have studied parametric spike-history filters in the form of a single exponential or sum of two exponential terms. In the next section, we will show that the validity of the QR approximation extends to physiological PP-GLM spike-history filters as they are typically obtained in the context of data-driven model estimation.

The quasi-renewal approximation predicts stability for complex (physiologically plausible) model parameters

We applied the QR approximation to models estimated from actual neuronal recordings, specifically, multi-electrode single-unit recordings in monkey cortex (see "Materials and Methods"). Of the nonlinear Hawkes PP-GLMs estimated from 99 recorded single units, 11 were

predicted to be stable, 86 were predicted as "fragile" with varying degrees of expected divergence times, and 2 were predicted to be divergent.

In all cases, the predictions were consistent with numerical simulations: For all models predicted to be stable, none of the $N = 48$ simulations of length $T = 1000$ s diverged ($\hat{T}_{div} = \infty$), and both divergent models showed finite divergence times ($\hat{T}_{div} = 86$ s and $\hat{T}_{div} = 2$ s, respectively). Fragile models did not diverge in our simulations of length $T = 1000$ s in 53 out of the 86 cases, while the other maximum-likelihood models diverged with varying degrees of observed divergence times ($\hat{T}_{div} = 4 - 45\,000$ s).

We examined in detail the stability predictions based on the QR approximation for a divergent neuron model in Fig 9. Spike-sorted single-unit activity (Fig 9A) was used to estimate a nonlinear Hawkes process with ten basis functions for the spike-history filter consisting of raised cosines [4, 5]. The resulting maximum-likelihood estimate (MLE) displays a relative refractory period followed by an excitatory rebound (Fig 9B). The corresponding transfer function of the QR approximation shows a single stable fixed point close to the maximally allowed firing rate (Fig 9C). Therefore, this model is classified as divergent.

We then explored the neighboring parameter space by varying the baseline rate parameter c and using scaled versions of the MLE spike-history filter. The qualitative predictions (separated by thick lines in Fig 9D) were overall consistent with numerical simulations of the model (Fig 9D). Here, the color scale represents the estimated divergence time in simulations based on 48 independent simulations of $T = 1000$ s each.
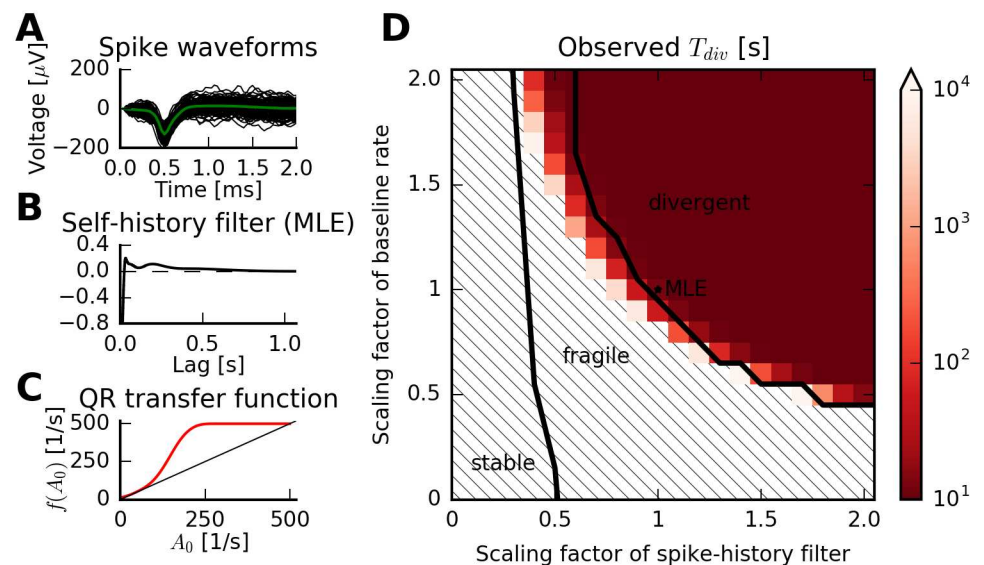
Fig 9. The quasi-renewal approximation predicts stability for complex (physiologically plausible) model parameters. (A) Single-unit activity (SUA) from multi-electrode recordings from monkey cortical area PMv [31] (same unit as in Fig 2C). Spike waveforms are shown (mean waveform in green) and indicate well-sorted SUA. (B) Estimated spike-history kernel using maximum-likelihood estimation. The kernel exhibits a relative refractory period followed by an excitatory rebound. (C) The transfer function predicted by the QR approximation. There is a single stable fixed point at ≈ 500 s^{-1}. The model is therefore classified as "divergent". (D) The QR approximation predicts the maximum-likelihood estimate (MLE) (center dot) to be divergent. Color indicates average divergence times in simulations for variations of the baseline rate c (y-axis) and scaled versions of the filter relative to the integral of $e^{\eta(s)} - 1$ (x-axis). Thick lines indicate the separation between areas for which the QR approximation predicts stability, fragility, or divergence. Overall, estimated divergence times from simulations agree with the qualitative predictions.

doi:10.1371/journal.pcbi.1005390.g009

In sum, the QR approximation not only predicted correctly the stability of the data-driven neurophysiological models, but also the stability of parameter variations around the MLE model.

Neuron models can be stabilized with constrained maximum-likelihood estimation

We have shown that fitting PP-GLMs to electrophysiological data can lead to divergent and fragile models (Figs 2 and 9). The QR approximation not only offers a way to predict stability, but also to find stable models. As stated before, stability is an important feature when the goal is to sample from the model or to obtain data-driven models for simulations. Conceptually, we can constrain the parameter search for the maximum-likelihood solution to the parameter space for which the QR approximation predicts stable models. We will now provide a proof of concept of this approach by demonstrating how this constraint can be implemented in practice.

We search for the maximum-likelihood estimate (MLE) under the additional constraint that the model is predicted to be stable by the QR approximation. This can be implemented by minimizing the cost function consisting of the negative log-likelihood of the data plus a penalty term that is infinity whenever the model is predicted to be fragile or divergent and zero otherwise. We use a gradient-free numerical optimization scheme (see "Materials and Methods"). We initialize the parameter values with the unconstrained MLE for which all positive coefficients are set to zero. This corresponds to a non-positive spike-history filter and ensures that the initial evaluation of the cost function is finite. We call the constrained solution the "stabilized MLE".

For illustration, in Fig 10A, we use the same data from monkey electrophysiological recordings as in Fig 9. We find that the spike-history filter of the stabilized MLE resembles a regularized version of the (unconstrained) MLE (Fig 10A), i.e., its coefficients are slightly biased towards zero. The MLE itself passes a goodness-of-fit test on training and test data and yields a substantial power in predicting spiking in 1 ms time bins (Fig 10B–10D). However, sampling from the MLE leads to divergent and unphysiological spike trains (Fig 10E). In contrast, sampling from the stabilized model yields firing rates comparable to the training data, and spike trains that are qualitatively similar (Fig 10E).

We quantified similarity of spike train statistics up to second order using the inter-spike interval (ISI) distribution (Fig 10F) and spike train auto-correlation function (Fig 10G). We evaluated the spiking pattern using three metrics (local ISI variability lv, shape ($\log \kappa$) and scale parameter ($\log \alpha$) of the best-fitting Gamma distribution, following the methodology presented in [35, 36]. Our data from monkey region PMv are consistent with reported values in [36, Table 1, row 15]: $lv^{(\text{data})} = 0.54 \pm 0.18$, $\log \kappa^{(\text{data})} = 0.80 \pm 0.38$, $\log \alpha^{(\text{data})} = 3.16 \pm 0.26$ (mean ± standard deviation over segments of 20 consecutive ISIs). Metrics for spike trains generated from the stabilized MLE model are well aligned to those of the physiological data: $lv^{(\text{stab.MLE})} = 0.66 \pm 0.22$, $\log \kappa^{(\text{stab.MLE})} = 0.69 \pm 0.35$, and $\log \alpha^{(\text{stab.MLE})} = 3.26 \pm 0.26$. In addition, the spike trains generated from the stabilized MLE also reproduce the serial correlations between consecutive ISIs in the physiological data (Fig 10H).

Although this was not a direct optimization criterion, the stabilized MLE is almost as good in predicting spiking activity in 1 ms time bins conditioned on observed spiking history as the MLE (Fig 10B) and scores only marginally worse on the goodness-of-fit tests (Fig 10D). The cross-validated log-likelihood of the stabilized model is within 80% of the log-likelihood score of the MLE (2.43 bits/s versus 2.90 bits/s relative to the prediction of a homogeneous Poisson process with correct spiking rate, Fig 10C).
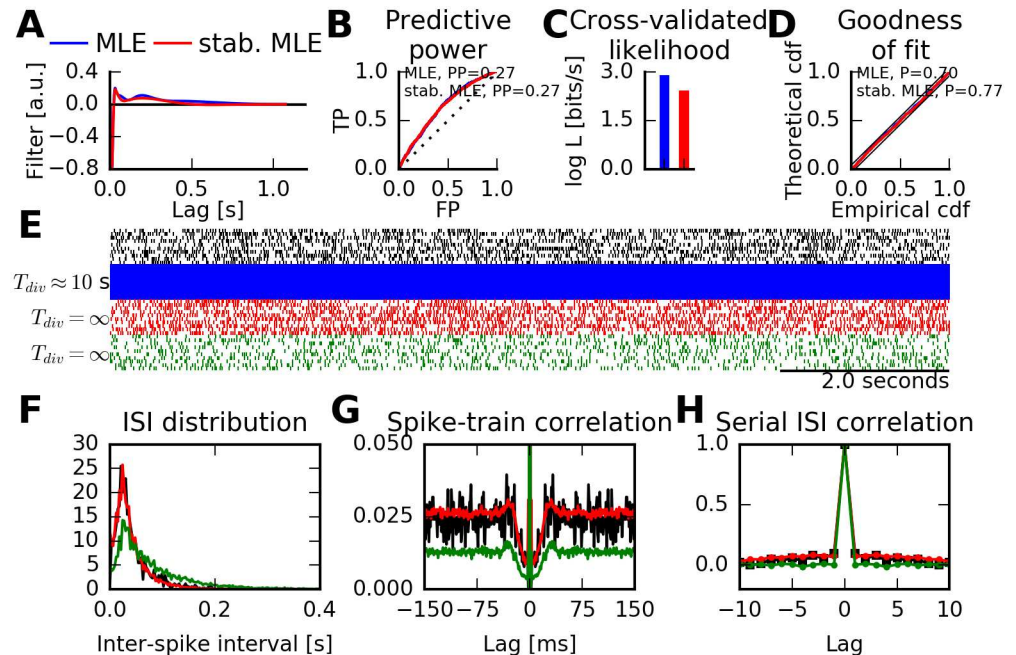
Fig 10. Neuron models can be stabilized with constrained maximum-likelihood estimation. (A) Estimated spike-history kernels using maximum-likelihood estimation (blue) and the constrained (stabilized) maximum-likelihood estimation (red). Data are from single-unit activity (SUA) recordings from monkey cortical area PMv [31]. (B) Power of predicting spiking activity on test data for both the maximum-likelihood estimate (MLE) and the stabilized MLE. The receiver operating characteristic (ROC) curve is shown for predicting spikes in 1 ms time bins with false positive rate (FP, x-axis) and true positive rates (TP, y-axis). Diagonal line indicates chance-level prediction. Predictive power is defined as PP = 2 · AUC − 1 with AUC being the area under the curve. Perfect spike prediction corresponds to PP = 1. Both models predict spikes equally well. (C) Log-likelihood evaluated on test data. Both MLE and stabilized MLE models preserve information about spike times. Model log-likelihoods (in bits per second) are relative to a homogeneous Poisson process with the correct spiking rate. (D) Kolmogorov-Smirnov test of rescaled inter-spike intervals following the time-rescaling theorem. Both MLE and the stabilized MLE pass the goodness-of-fit test ($P > 0.05$). (E) Recorded spike trains and simulated spike trains from estimated models. Top: Randomly selected 10 s intervals of neural activity part of the training data (black). Simulating from the unconstrained MLE model (blue) leads to quickly diverging firing rates ($T_{\text{div}} \approx 10$ s). Spiking activity from the stabilized MLE (red) remains finite and physiological. Simulating from the MLE with a reset condition after each spike (green, [34]) leads to non-divergent firing rates, but the firing rate of the training data is not matched. (F) Inter-spike interval statistics of real and simulated spiking activity for which rates were non-divergent. Same colors as in (D). The stabilized MLE qualitatively reproduces the training data ISI distribution. (G) Autocorrelation of recorded and simulated activity. (H) Serial ISI correlations of real and simulated spiking activity. The stabilized MLE accurately reproduces correlations in the training data. Simulating the MLE with a reset condition (green) leads to a renewal process, hence vanishing correlations at non-zero lags.

Spike train statistics were not shown for the MLE because physiological spike trains could not be obtained. However, we explored the possibility of a simple modification in the simulation procedure for the MLE solution that guarantees stability of the generated spike train. In the "reset condition" [34], the conditional intensity of the point process at time t is not calculated using the original CIF (Eq (3)) but instead using only the spike-history effect stemming from the most recent spike at $\hat{t}$: $\lambda(t|\mathcal{H}_t) = c \exp\left(\eta(t - \hat{t})\right)$. This yields a renewal process with guaranteed finite rate in simulations.

We found that models simulated with the reset condition do produce finite firing rates and spike trains that look plausible on first sight (Fig 10D). However, the firing rate and autocorrelation function are not matched with the training data. Most notably, neither ISI shape (Fig 10F and 10G, green line; $lv^{(\text{reset})} = 0.72 \pm 0.21$, $\log \kappa^{(\text{reset})} = 0.51 \pm 0.31$,

log $\alpha^{(\text{reset})}$ = 2.57 ± 0.20) nor ISI correlations are reproduced due to the renewal property of the modified model (Fig 10H, green line).

Discussion

We have presented a framework to predict the stability and dynamics of a general class of stochastic neural point process models, specifically nonlinear Hawkes processes and point process GLMs. This framework is based on a quasi-renewal approximation of the exact conditional intensity function model. The assessment of stability can serve as an additional goodness-of-fit test along with other approaches, such as tests based on the time-rescaling theorem. We have also shown that simulated activity from point-process models estimated from neurophysiological data tends to exhibit unphysiologically high firing rates. This behavior results from lacking stochastic stability of the estimated PP-GLMs. When sampling and simulations based on these PP-GLMs is desired, our framework can be used to derive stability constraints which can be included into standard parameter estimation techniques such as (regularized) maximum-likelihood optimization. Furthermore, our framework provides a way to determine and classify the qualitative types of stochastic dynamics exhibited by PP-GLMs, specifically nonlinear Hawkes processes with absolute refractory periods.

Empirically, we have shown that many point-process GLMs estimated from neurophysiological data tend to show fragile or even divergent dynamics. Several reasons could explain this finding. One potential scenario is that of the actual recorded neural dynamics being close to instability. Even assuming a correctly specified model, finite data will lead to finite variance in the parameter sets estimated via maximum likelihood. If the true parameters are close to the boundary between stability and divergence, the estimated MLE from finite data may lie in the unstable (divergent) region of the parameter space. In this way, the simulated model would show qualitatively different dynamics than the underlying data set from which it was derived (i.e., it would diverge).

Another potential reason is the extent of model misspecification. Constraining the true data-generating process into the form of the nonlinear Hawkes process, as assumed here, might bias the parameter estimation towards the parameter space for which model dynamics is unstable. An example for such model misspecification are non-stationary firing rates. In the present models, we assumed a constant baseline firing rate. Maximum-likelihood estimates have shown to be systematically biased if the non-stationarity is not accounted for in the model [37]. Detecting and accounting for non-stationarity in neuronal spiking data is, however, a nontrivial task. The stability framework proposed here can be used as a tool for complementary analyses in this regard.

The quasi-renewal approximation can be used alongside other goodness-of-fit tests to assess model adequacy. It would be a reasonable demand for a model to be classified as stable or fragile with high expected divergence times if it is to be used within the context of simulations or closed-loop applications. Even outside the context of simulations, a model classified as divergent is failing a goodness-of-fit test as it is not able to reproduce the statistics of the training data. Conversely, a model classified as stable might produce finite steady-state firing rates but may still fail with regards to other statistics of the training data, such as first- and second-order firing statistics, interval correlations, or burst properties. Our work exemplifies that it is important to check goodness-of-fit using a comprehensive battery of tests to check for different types of model misspecifications.

The derivation of the quasi-renewal approximation involves a series expansion which here we truncated after the first order. As it was previously stated, this approximation becomes less valid with non-vanishing second- and higher-order spike-spike correlations [26]. This is the

case, e.g., for strongly bursting neurons. We have observed these limitations in the exploration of the parameter space over which the QR approximation is valid (Figs 5 and 7). In principle, second- and higher-order correlation terms can be included in the expansion. In particular, Naud et al. [26] truncate the series after the second order. In this case, a self-consistent solution has to be found in terms of the steady-state firing rate and the steady-state spike train auto-correlation.

An additional limitation is inherent to the reduction of the spiking dynamics to a one-dimensional description of a transfer or gain function. For high firing rates, spike trains are constrained by the absolute refractory period and exhibit strong regularity. The transfer function can be qualitatively different when point process histories are assumed to be periodic spike trains of a particular rate rather than homogeneous Poisson-like firing with vanishing higher-order correlations. Hence, the description of the dynamics using a single transfer functions becomes less valid for high firing rates. In practice, this happens when activity diverges beyond which the usefulness of the QR approximation becomes limited.

Relaxing assumptions and extensions

We now discuss possible relaxing assumptions and extensions to the quasi-renewal framework to assess stability and dynamics of neural point process models.

First, the ability to separate the effect of the most recent spike from all previous spikes and to use the moment-generating functional are unique to the exponential nonlinearity. Possibly, for other specific choices of the nonlinearity a similar manipulation or approximation of Eq (11) may be conceivable, or the firing rate transfer function may be estimated by numerical simulation of the model. However, we consider such an extension of the theory beyond the scope of the present work. Using a different nonlinearity will likely require to invoke some type of Lipschitz condition and alternative ways of studying the stability properties (see, e.g., [9]).

Second, the nonlinear Hawkes process can be formulated as a multivariate process to describe an ensemble of coupled neurons [4, 5, 21]. The corresponding extension of the quasi-renewal framework is possible [27] and may be used to study stability of such networks of heterogeneous neurons. Local linear stability analysis of derived fixed points for the neuronal network dynamics can then be readily implemented based on the spectral radius of a coupling matrix (obtained from coupling coefficients, history filters' integrals, and the nonlinearity's first derivative) computed at the fixed point locations [9, 21].

Third, we assumed no (time-varying) exogenous input. The framework can be easily extended to accommodate non-stationary inputs, such as stimulus drive, by allowing the baseline firing rate c in Eq (3) to be time-dependent and performing the QR stability analysis for the supremum of c as long as such a bound exists. If stability is predicted for this dominating model, the model with time-varying exogenous input will be stable as long as the exogenous drive is independent of the firing rate of the neuron itself [21]. This generalization allows the stability analysis to be performed, e.g., for state-space models such as linear dynamic systems with conditionally Poisson observations, where spike-history effects are combined with neural couplings to a low-dimensional latent state whose dynamics is stable itself [2, 38–42].

Fourth, throughout this study, we assumed an absolute refractory period of 2 ms. Because the absolute length of such a refractory period does not impact the theoretical analysis, a refractory period of arbitrary length may be used. Although the existence of refractory periods is pervasive in most physical applications, the assumption of an absolute refractory period might not be justified in some cases. The refractory period leads to a finite support of the transfer function and a bounded firing rate. In the limit of a vanishing refractory period, there

might be additional alignments of fixed points to those outlined in Fig 3, such as a single stable and a single unstable fixed point, which can be similarly classified into stable, fragile, and divergent dynamics.

Finally, we provided an approximation of the divergence time for fragile, metastable models. A more direct estimation of the divergence time seems desirable. One alternative to estimate the divergence time would be a fluctuation analysis in analogy to Brownian motion in a potential. In this case, the divergence time would correspond to the escape time of the particle from the potential given around the low-rate stable fixed point. In first order, noise in this process may be described by white noise, scaled with an intensity that should be proportional (if not equal) to the square root of the rate (Poisson statistics). A more insightful or useful estimate of the fluctuations may also be derived from the power spectral density of the activity in the metastable state [27, 43]. Alternatively, a periodic perturbation approach as in [7] may be attempted to calculate the time scales of expected divergences.

Alternative stabilization methods

We provided a proof-of-principle of how our proposed quasi-renewal framework can be incorporated into a maximum-likelihood parameter estimation procedure in order to guarantee the stability of models estimated from physiological data. We saw that the ordinary MLE provided unphysiological spike trains while the stabilized version matched first- and second-order statistics of the training data with only marginal loss in predictive power and goodness-of-fit scores.

In our current implementation, we maximize the likelihood of the model under the constraint that the model is predicted to be stable by the QR approximation. The determination of stability based on the number and location of fixed points—essentially a bifurcation analysis—consists of multiple steps and is highly nonlinear. We were not able to differentiate the penalty term to exploit more efficient gradient-based optimization schemes. However, empirically, we have observed that the space of admissible parameters seems to form a single connected, possibly even convex, set. From a computational point of view, the determination of stability involves one-dimensional (scalar) arithmetic that allows fast evaluation of many candidate parameter sets during the optimization procedure. If necessary, additional speed-ups could be obtained by parallelizing the computation of the penalty term when evaluating different local search directions.

We restricted the attainable parameter space to all models that are classified as stable, thereby explicitly excluding fragile models with long expected divergence times. In practice, stable models and fragile models are both candidates for physiological dynamics as long as simulation times are shorter than typical divergence times. Therefore, a more refined cost function than the one used here (see Eq (35)) could involve a penalty proportional to the expected divergence rate $\mathrm{E}[T_{\mathrm{div}}]^{-1}$, weighted by a corresponding regularization parameter. The calculation of the expected divergence rate is more computationally expensive than the determination of stability itself but would provide a continuous and potentially smooth penalty function that could be superior to the all-or-nothing penalty term of the proposed optimization scheme.

Alternatively, a computationally efficient regularization is the L1-regularized maximum-likelihood estimate. It provides a convex optimization problem that can be efficiently solved [44–46]. For stronger regularization parameters, estimated coefficients tend towards zero. Nonlinear Hawkes processes with vanishing spike-history filter (coefficients tending towards zero) are always stable. This implies the existence of an optimally L1-regularized solution that

is predicted to be stable. Therefore, strong L1-regularization might be an alternative approach to model stabilization.

In addition, a more parametric description of the spike-history kernel could facilitate stability. For example, the addition of an adapting (inhibitory) power-law component is likely to prevent any runaway-excitation and has been observed experimentally under certain conditions [17]. A parametric form like this will have to be accompanied by model selection and appropriate goodness-of-fit tests.

A simple way to ensure stability of nonlinear Hawkes processes in simulations is to implement a "reset condition" following each spike when the previous spiking history is forgotten. This leads to a renewal process with a well-defined stationary and unique solutions [47]. However, by definition, second-order statistics such as ISI correlations cannot be reproduced and spike trains generated with this condition are not realizations from the original nonlinear Hawkes model. Nevertheless, it might be an easy fix in certain applications where stable network simulations are desired without explicitly reproducing physiological spike train statistics [34].

Finally, we note that while previous work in the mathematical domains covers uniqueness and existence theorems for stationary point processes [9, 21], it does not provide predictions of dynamics, metastability, or whether steady-state rates are in a physiological regime. In the case of Lipschitz-continuous nonlinearities, existing conditions for stability are overly conservative and of little relevance for neurophysiologically plausible spike-history filters. We are currently working on relaxations of these conditions that would allow easier characterizations of stability of neuron models and neuronal networks, and we hope to report such results in the future.

Importance of stable point process models for applications

The ability to predict stochastic stability of a given point process model has merit in its own right and is therefore a main contribution of our study. While stability in simulations is neither a necessary nor sufficient condition for the model fitting procedure itself, and a trade-off between stability constraints and other standard regularized MLE approaches should be considered case by case, data-driven models that are guaranteed to be stable are of major importance for many applications. We will conclude by giving a few examples.

First, to understand brain circuits may mean to be able to rebuild them using artificial components. PP-GLMs offer a direct and relatively well-understood method to derive neuron models from data. But their use in rebuilding brain circuits in simulation is limited if stochastic stability is uncertain. For the simple example of a PP-GLM fit of a neuronal network with an embedded "synfire chain", Zaytsev et al. ensured stability of the network by adding a reset mechanism of the membrane after each spike to prevent run-away excitation [34].

Second, PP-GLMs are used in the context of (closed-loop) brain machine interfaces. Present-day experiments may interface brain tissue to virtual actuators that in turn provide feedback signals to the brain (e.g., [48]). Future applications of such technology may replace damaged neural tissue by simulated neural circuits which are connected bidirectionally to the brain. Such circuits could potentially be made using PP-GLMs or related models, fitted to the system that is being substituted. Stability of the model is essential in this case to exclude major system malfunction.

A final application is to make spike-timing predictions from neuron models: Given parallel recordings of neuronal activity, precise timing of single spikes can already be predicted using PP-GLMs [5]. However, such predictions have been limited to the very near future (on the order of milliseconds), but typically cannot be used for extended time differences into the

future because of model instability. Ensuring stability of PP-GLMs potentially extends the time horizon of spike timing prediction. This may enable applications in epilepsy treatment, seismology, finance, and other fields that study self-exciting dynamical processes.

Materials and methods

Ethics statement

For the non-human primate data, all procedures were in accordance with Brown University Institutional Animal Care and Use Committee approved protocols and the Guide for the Care and Use of Laboratory Animals. Approval for the human studies was granted by local Institutional Review Boards (Partners Human Research Committee, Massachusetts General Hospital, Harvard Medical School), and the participant was enrolled after obtaining informed consent.

Experimental details

Full experimental details for the electrophysiological data from the crab can be found in [6], for the non-human primate data in [31], and for the human data in [30, 49].

The nonlinear Hawkes process

The nonlinear Hawkes process is a flexible class of self-exciting and/or self-inhibiting point process models [9]. For a stochastic point process, the conditional intensity function is given by [10]:

$$\lambda(t|\mathcal{H}_t) = \lim_{\Delta \to 0} \frac{P(\text{spike in}(t, t + \Delta]|\mathcal{H}_t)}{\Delta}, \tag{9}$$

where $\mathcal{H}_t$ is the history of the process (i.e., all K spikes at times t_k up to time t).

In the nonlinear Hawkes model, $\lambda(t|\mathcal{H}_t)$ is assumed to take the following form:

$$\lambda(t|\mathcal{H}_t) = \phi([\eta * S](t)) = \phi\left(\sum_{k=1}^{K} \eta(t - t_k)\right), \tag{10}$$

where $\phi(x) : \mathbb{R} \to \mathbb{R}^+$ is a nonlinearity that maps the convolution of the spike train S with a causal auto-history kernel $\eta(s) \in \mathbb{R}$ onto a non-negative conditional intensity $\lambda(t|\mathcal{H}_t)$.

Here, we consider $\phi(x) = c \exp(x) = \exp(I_0 + x)$ with $c = \exp(I_0) > 0$. The exponential nonlinearity implies that modulations from previous spikes interact in a multiplicative way [50]. The choice for the exponential nonlinearity has both theoretical [2, 7] and empirical support, e.g., from electrophysiological experiments [51, 52]. We only consider the single-neuron (univariate) case although nonlinear Hawkes processes can be similarly defined for coupled neuronal ensembles with the corresponding matrix of auto- and cross-history kernels $\eta_{ij}(s)$.

Furthermore, we assume constant exogenous input, i.e., $c \equiv const$. For time-varying inputs $c(t)$ or other (potentially non-stationary) exogenous inputs, a conservative stability analysis can be performed by using $c_{\sup} = \sup c(t)$ as long as such a bound exists.

Hence, the stochastic process is completely determined by two parameters: c (or equivalently I_0) and the causal auto-history kernel $\eta(s)$.

To simplify analysis, we make certain assumptions about $\eta(s)$. One is the introduction of an absolute refractory period τ_{ref} which indicates that the next spike can occur no closer than τ_{ref} to the last spike. It can be modeled by setting $\eta(s) = -\infty$ for $s < \tau_{\text{ref}}$. Given that action potentials, the events that are modeled with the point process model, have an extent of around 1 ms, we assume $\tau_{\text{ref}} = 2$ ms.

Quasi-renewal approximation

In the nonlinear Hawkes model, the effects of previous spikes can accumulate. This leads, in general, to a non-renewal process. For this specific class of point process models, there are no closed-form formulas to predict mean intensities, inter-spike interval (ISI) distributions, power spectra, or other properties of the process. We are interested in whether a nonlinear Hawkes process with given parameters permits stable, finite steady-state firing rates.

To make progress, we need to approximate Eq (10). Our approach is based on the recently introduced quasi-renewal approximation [26, 27]. To obtain an estimate of the mean rate of the process A_0, we average $\lambda(t|\mathcal{H}_t)$ over all possible spike trains $S(t)$ prior to the last spike $\hat{t}$:

$$
\begin{aligned}
\lambda_0(t, \hat{t}) &= \langle \lambda(t|\mathcal{H}_t) \rangle_{S(t<\hat{t})} \\
&= c \exp\left(\eta(t - \hat{t})\right) \langle \exp\left([\eta * S](t)\right) \rangle_{S(t<\hat{t})}.
\end{aligned}
$$

We identify the second term with the moment-generating functional of S that can be expanded in a series of moments [26, 53] which we truncate after the first order:

$$
\lambda_0(t, \hat{t}) \approx c \exp\left(\eta(t - \hat{t})\right) \exp\left\{ \int_{-\infty}^{\hat{t}} (e^{\eta(t-t')} - 1) \underbrace{\langle S(t') \rangle_{S(t<\hat{t})}}_{A_0} dt' \right\}, \tag{11}
$$

with $A_0 = \langle S(t') \rangle_{S(t<\hat{t})}$ being the first moment of the averaged spike train, which corresponds to a constant by assuming stationarity. A_0 is the steady-state firing rate which for now remains unknown.

We can rewrite the integration bounds and introduce $\tau = t - \hat{t}$ as the time since the last spike to obtain:

$$
\lambda_0(\tau) = c \exp\left\{ \eta(\tau) + A_0 \int_{\tau}^{\infty} \underbrace{(e^{\eta(u)} - 1)}_{\gamma(u)} du \right\}, \tag{12}
$$

with $\gamma(u) = e^{\eta(u)} - 1$ for the exponentiated kernel and $\Gamma = \int_0^{\infty} \gamma(u) du$. Γ is used in the rescaling of the filter for the analysis presented in Fig 9.

Using the quasi-renewal (QR) conditional intensity of Eq (12), we obtain the steady-state survivor function $\mathcal{S}_0$ and inter-spike interval (ISI) density P_0 as:

$$
\mathcal{S}_0(\tau) = \exp\left\{ -\int_0^{\tau} \lambda_0(u) du \right\}, \tag{13}
$$

$$
P_0(\tau) = \mathcal{S}_0(\tau) \lambda_0(\tau), \tag{14}
$$

which, in turn, predict the firing rate:

$$
f(A_0) = \left[\int_0^{\infty} \tau P_0(\tau) d\tau \right]^{-1}. \tag{15}
$$

Because $\frac{d}{d\tau}\mathcal{S}_0(\tau) = -P_0(\tau)$ and integrating by parts:

$$
\begin{aligned}
f(A_0) &= \left[-\int_0^\infty \tau \frac{d}{d\tau}\mathcal{S}_0(\tau)d\tau \right]^{-1} \\
&= \left[-[\tau \mathcal{S}_0(\tau)]_0^\infty + \int_0^\infty \mathcal{S}_0(\tau)d\tau \right]^{-1} \\
&= \left[\int_0^\infty \mathcal{S}_0(\tau)d\tau \right]^{-1},
\end{aligned}
\tag{16}
$$

which is continuous and differentiable in A_0.

Stability based on the transfer function

Eq (16) defines an average input-output mapping for the single neuron, known as transfer or gain function, which maps an assumed mean input rate A_0 to the mean output rate of the process $f(A_0)$. The dynamics of the model can be characterised based on the properties of the transfer function. Fixed points of this map, $\hat{A}_0 = f(\hat{A}_0)$, can be locally stable or unstable. To detect the fixed points reliably, we search for the zero crossings of the function $g(A_0) = f(A_0) - A_0$.

A fixed point $\hat{A}_0$ is locally stable if $g'(\hat{A}_0) = \frac{d}{dA_0}g(A_0)|_{A_0 = \hat{A}_0} < 0$ and unstable if $g'(\hat{A}_0) \geq 0$. Based on the number and location of fixed points of the mean firing rate map (Eq (16)) we can then classify the model (see Results).

Prediction of divergence rate for fragile models

For fragile (metastable) models, we may ask whether there is an inter-spike interval x, which if several spikes occur repeatedly with this interval, causes a divergence of the firing rate. If such an x exists, we can compute the probability of this event. Although other routes to a divergent rate are possible, this one yields an explicit value for its rate of occurrence and can be used as a lower bound for the divergence rate of a metastable model.

Let t_1 be the time of a spike of the process. Then the next spike occurs within the interval x with probability $1 - \mathcal{S}(t_1 + x, t_1)$, where $\mathcal{S}$ is the survivor function. We may now iterate this argument to compute the probability that, following t_1, there is a sequence of K spikes with intervals smaller or equal to x, as:

$$
\begin{aligned}
p_{\text{reg}}(x) &= (1 - \mathcal{S}(t_1 + x, t_1))(1 - \mathcal{S}(t_1 + 2x, t_1 + x))\ldots \\
&= \prod_{k=1}^K \underbrace{[1 - \mathcal{S}(t_1 + kx, t_1 + (k-1)x)]}_{p(x,k)}
\end{aligned}
\tag{17}
$$

To evaluate Eq (17) and compute $\mathcal{S}$, we need to approximate the intensity function $\lambda(t|\mathcal{H}_t)$ for the case that up to t_1 we do not have information about the spike history apart from the rate A_0, but from t_1 on it is defined as the regular firing case with spike times $t_k = t_1 + (k-1)x$ for

$k \geq 1$. This gives rise to a similar quasi-renewal approximation as for Eq (12):

$$
\begin{aligned}
\lambda_{\mathrm{reg}}(t, x) &= \langle \lambda(t|\mathcal{H}_t) \rangle_{S(t < t_1)} \\
&= c \exp\left\{ \sum_{k=1}^{\infty} \eta(t - t_k) \right\} \langle \exp\{ [\eta * S](t) \} \rangle_{S(t < t_1)} \\
&\approx c \exp\left\{ \sum_{k=1}^{\infty} \eta(t - t_1 - (k-1)x) \right\} \exp\left\{ \int_{-\infty}^{t_1} (e^{\eta(t-t')} - 1) A_0 dt' \right\} \\
&= c \exp\left\{ \sum_{k=1}^{\infty} \eta(t - t_1 - (k-1)x) \right\} \exp\left\{ A_0 \int_{t-t_1}^{\infty} \gamma(s) ds \right\}.
\end{aligned}
\tag{18}
$$

Inserted into Eq (17), we then first check whether the sequence $p(x, k)$ increases monotonically towards 1, setting $t_1 = 0$. If for a given k we have $p(x, k) > p(x, k + 1)$, we terminate the iteration because x does not seem to lead to the regular divergence and return $p_{\mathrm{reg}} = 0$. If, in contrast, for some value of k, $p(x, k)$ is close to 1, we have found a divergent case that occurs with probability $p_{\mathrm{reg}}(x)$, as given by Eq (17) with $K = k$.

This procedure is performed for all $x \in [\tau_{\mathrm{ref}}, A_0^{-1}]$, and the maximum $\max_x p_{\mathrm{reg}}(x)$ is returned. As $p_{\mathrm{reg}}(x)$ is the probability of the regular divergence with intervals x or shorter to occur after any spike of the process, the rate of divergence is thus bounded from below by:

$$
r_{\mathrm{div}} \geq A_0 \max_x p_{\mathrm{reg}}(x). \tag{19}
$$

This provides an upper bound for $T_{\mathrm{div}} = r_{\mathrm{div}}^{-1}$.

Bursting and the regular spiking limit

Apart from the dynamic stability that we have discussed so far, a particular limit of the space of possible spike trains is of special interest. In case of a divergent firing rate, which occurs in unstable or fragile models, the analysis of the gain function predicts that the firing rate saturates at the limit given by the inverse of the refractory period, $A_0 = \tau_{\mathrm{ref}}^{-1}$. However, there is only one spike train that can realize this firing rate, which is the regular spike train:

$$
S_x(t) = \sum_{k=-\infty}^{\infty} \delta(t - kx), \tag{20}
$$

with inter-spike-interval $x = \tau_{\mathrm{ref}}$. Here θ denotes the Heaviside function and δ denotes the Dirac delta function.

For the regular spike train S_x, with $x > \tau_{\mathrm{ref}}$ being close to the refractory period, to be a possible mode of firing of the model, it is necessary that the conditional intensity of the neuron (Eq (10)), evaluated at time x after the last spike (at $t - x$),

$$
\lambda_{\mathrm{reg}}^K(x) = c \exp\left[\sum_{k=1}^{K} \eta(kx) \right], \tag{21}
$$

reaches a sufficiently high value, so that the rate x^{-1} can be maintained, when the regular spike train extends into the past forever, $\lambda_{\mathrm{reg}}(x) = \lim_{K \to \infty} \lambda_{\mathrm{reg}}^K(x)$. But what precisely is that sufficiently high value of λ_{reg}? Since after τ_{ref} the refractory period is over, and because the conditional intensity changes approximately on the time scale of the filter η that is much greater than the remaining interval $x - \tau_{\mathrm{ref}}$, we may approximate the mean output inter-spike-interval

of the process by:

$$\mu_{\mathrm{reg}}(x) = \tau_{\mathrm{ref}} + \lambda_{\mathrm{reg}}^{-1}(x) \,. \tag{22}$$

Now we can formulate a condition on the divergent firing model class: If the expected interval μ_{reg} in the regular firing case is smaller or equal to the input interval x, the regular firing state can be maintained. From Eqs (22) and (21), we obtain the condition:

$$\sum_{k=1}^{\infty} \eta(kx) \geq -\ln\left[c(x - \tau_{\mathrm{ref}})\right] \,. \tag{23}$$

This condition is instructive in two ways: First, to maintain the interval $x = \tau_{\mathrm{ref}}$, the series on the left hand side (LHS) has to diverge to positive infinity quicker than the logarithm on right hand side (RHS). Second, in case the series on the LHS of Eq (23) converges for all $x \geq \tau_{\mathrm{ref}}$, there is a minimum value of x for which Eq (23) is still fulfilled. Then x^{-1} is the peak firing rate that this model can maintain close to the regular spiking limit.

Many models, even ones with an upper unstable fixed point, might not fulfill Eq (23). These models are fragile, but also cannot maintain the regular firing mode. Nonetheless we may ask for how many regular spikes they can maintain the tonic activity. This can be addressed by a modified condition like Eq (21) considering $K < \infty$. By analogous reasoning as above, we arrive at the condition:

$$\sum_{k=1}^{K} \eta(kx) \geq -\ln\left[c(x - \tau_{\mathrm{ref}})\right] \,. \tag{24}$$

For a given tonic firing interval x close to τ_{ref} (e.g., defined as $x^{-1} = 0.9 \times \tau_{\mathrm{ref}}^{-1}$), the maximum K_{max} for which Eq (24) is fulfilled yields a good approximation of the duration $K_{\mathrm{max}} x$ of the intermittent regular spiking episodes of the model.

If the spike-history filter is a sum of two exponential terms, the condition in Eq (24) takes the specific form of a geometric series:

$$
\begin{aligned}
-\ln[c(x - \tau_{\mathrm{ref}})] \;\leq\; & \sum_{k=1}^{K} \eta(kx) = \sum_{k=1}^{K} \left[J_{\mathrm{r}} e^{-kx/\tau_{\mathrm{r}}} + J_{\mathrm{a}} e^{-kx/\tau_{\mathrm{a}}} \right] \\
=\; & J_{\mathrm{r}} e^{-x/\tau_{\mathrm{r}}} \sum_{k=0}^{K} e^{-kx/\tau_{\mathrm{r}}} + J_{\mathrm{a}} e^{-x/\tau_{\mathrm{a}}} \sum_{k=0}^{K} e^{-kx/\tau_{\mathrm{a}}} \\
=\; & J_{\mathrm{r}} e^{-x/\tau_{\mathrm{r}}} \frac{1 - e^{-Kx/\tau_{\mathrm{r}}}}{1 - e^{-x/\tau_{\mathrm{r}}}} + J_{\mathrm{a}} e^{-x/\tau_{\mathrm{a}}} \frac{1 - e^{-Kx/\tau_{\mathrm{a}}}}{1 - e^{-x/\tau_{\mathrm{a}}}} \\
=\; & J_{\mathrm{r}} \frac{1 - e^{-Kx/\tau_{\mathrm{r}}}}{e^{x/\tau_{\mathrm{r}}} - 1} + J_{\mathrm{a}} \frac{1 - e^{-Kx/\tau_{\mathrm{a}}}}{e^{x/\tau_{\mathrm{a}}} - 1} \,,
\end{aligned}
\tag{25}
$$

which for $K \to \infty$ becomes condition Eq (23), which here is:

$$-\ln[c(x - \tau_{\mathrm{ref}})] \leq J_{\mathrm{r}}(e^{x/\tau_{\mathrm{r}}} - 1)^{-1} + J_{\mathrm{a}}(e^{x/\tau_{\mathrm{a}}} - 1)^{-1} \,. \tag{26}$$

The boundary defined by Eq (26) in the (J_r, J_a) space is a line: For every J_r there is a maximum $J_{\mathrm{a}}^{\mathrm{max}}$ from which on Eq (26) is true. For $J_{\mathrm{a}} < J_{\mathrm{a}}^{\mathrm{max}}$, in contrast, we are assured that the regular firing mode with interval x is unstable. Models for which Eq (26) is not fulfilled will show intermittent bursting activity (compare with Fig 7).

Estimation of average divergence time from simulations

We estimate the average divergence time of a given neuron model by simulating $N = 48$ independent neurons for $T = 1000$ s each. A neuron is said to have diverged at time t if its average firing rate in the interval $[t - 1, t + 1]$ seconds exceeds $\lambda_{thr} = 0.9 \times \tau_{\mathrm{ref}}^{-1}$. Alternatively, the firing rate may stay below λ_{thr} until the end of the simulation ("censored observation").

We can now derive the maximum-likelihood estimate of the divergence time. We assume that neurons diverge randomly with rate $r = 1/T_{\mathrm{div}}$. This seems to be justified in practice based on our simulations. Then, the likelihood of observing a divergence time y smaller than T is $\frac{1}{T_{\mathrm{div}}} e^{-y/T_{\mathrm{div}}}$ and the probability to observe a censored observation of length T is given by:

$$\int_T^\infty \frac{1}{T_{\mathrm{div}}} e^{-y/T_{\mathrm{div}}} dy = e^{-T/T_{\mathrm{div}}} . \tag{27}$$

If we denote the observed divergence times as $y_1, y_2, \ldots, y_k$, and we have $N_c = N - k$ censored observations, the overall log-likelihood function is given by:

$$\log L = \sum_{i=1}^k \log \left(\frac{1}{T_{\mathrm{div}}} e^{-y_i/T_{\mathrm{div}}} \right) + N_c \log \left(e^{-T/T_{\mathrm{div}}} \right) \tag{28}$$

$$= -(N - N_c) \log \left(T_{\mathrm{div}} \right) - \frac{N_c T}{T_{\mathrm{div}}} - \frac{\sum y_i}{T_{\mathrm{div}}} . \tag{29}$$

At the maximal (log-)likelihood estimate $\hat{T}_{\mathrm{div}}$, the gradient with respect to T_{div} has to vanish:

$$\left. \frac{\partial \log L}{\partial T_{\mathrm{div}}} \right|_{T_{\mathrm{div}} = \hat{T}_{\mathrm{div}}} = -\frac{N - N_c}{\hat{T}_{\mathrm{div}}} + \frac{N_c T}{\hat{T}_{\mathrm{div}}^2} + \frac{\sum y_i}{\hat{T}_{\mathrm{div}}^2} \overset{!}{=} 0 , \tag{30}$$

$$\hat{T}_{\mathrm{div}}(N - N_c) = N_c T + \sum y_i , \tag{31}$$

$$\hat{T}_{\mathrm{div}} = \frac{N_c T + \sum y_i}{N - N_c} . \tag{32}$$

This assumes that there was at least one non-censored observation ($N_c < N$). Otherwise, we set $\hat{T}_{\mathrm{div}} = \infty$. Note that if $N_c > 0$, $\hat{T}_{\mathrm{div}}$ may be larger than T.

Model estimation via maximum-likelihood optimization

If $\eta(s)$ in Eq (10) is parameterized through a set of basis functions $\{B_i(s)\}$ with linear coefficients $\{\beta_i\}$, then $\eta(s) = \Sigma_i \beta_i B_i(s)$. All model parameters $\{I_0, \vec{\beta}\}$ can be estimated via the statistical framework of generalized linear models (GLMs) [2]. We discretize the spike train to obtain a series of spike counts n_i in each time window of length $\Delta = 1$ ms. The expected spike count is given by the discrete-time approximation of Eq (10) as $\mathrm{E}[n_i] = \lambda_i \Delta$. The log-likelihood is then proportional to:

$$\log L \propto \sum_i (n_i \log \left(\lambda_i \Delta \right) - \lambda_i \Delta). \tag{33}$$

For the estimation of physiologically plausible model parameters (Fig 9), we used 10 raised cosine functions [4] with logarithmically spaced peaks up to 400 ms as basis functions $\{B_i(s)\}$ for the spike-history filter. In addition, an absolute refractory period of $\tau_{\mathrm{ref}} = 2$ ms was enforced. To improve numerical convergence and to ensure finite parameters for very sparse

data sets, we added a small L2-penalty term to the log-likelihood function so that the maximum-likelihood estimate (MLE) corresponds to the minimum of the cost function:

$$C(I_0, \vec{\beta}) = -\log L(I_0, \vec{\beta}) + \alpha \sum_{i=1}^{10} \beta_i^2, \quad (34)$$

with regularization parameter $\alpha = 5 \cdot 10^{-4}$.

Stabilization-constrained maximum-likelihood estimation

For the stabilization procedure (Fig 10), we performed the maximum-likelihood estimation (see previous section) under the additional constraint that the model is predicted to be stable by the QR approximation. That is, we optimized the cost function:

$$C(I_0, \vec{\beta}) = \begin{cases} -\log L(I_0, \vec{\beta}) + \alpha \sum_{i=1}^{10} \beta_i^2 & \text{if the model } \{I_0, \vec{\beta}\} \text{ is predicted to be stable,} \\ \infty & \text{otherwise.} \end{cases} \quad (35)$$

We used a gradient-free optimization algorithm (Nelder-Mead) with a convergence criterion on the change in parameter values ($\frac{\|\Delta\vec{\beta}\|}{\|\vec{\beta}\|} < 10^{-4}$).

The initial condition was chosen as the MLE solution for which positive parameter coefficients were set to zero. Because basis functions $B_i(s)$ are non-negative, this corresponds to a non-positive spike-history filter and ensures that the initial evaluation of the cost function is finite. Then, the optimization starts from a region with finite cost and allows the algorithm to descend to a (local) minimum.

Simulation

All simulations with the spike-history filter consisting of one or two exponentials (Figs 5 to 8) were performed using NEST [54], with neuron model "pp_psc_delta" in time steps of 0.5 ms. All other spike train simulations were performed with custom-written MATLAB software with a time discretization of 0.2 ms. Analysis and optimization were performed in MATLAB and Python.

Supporting information

S1 Fig. Less-rapidly growing nonlinearities do not prevent instability. (A) In addition to the exponential nonlinearity used in Fig 2 (blue), we also simulated spike trains using two less rapidly growing nonlinearities: First, a linear-rectifier function, i.e., $f(x) = [x + 1]_+$ which is $x + 1$ for $x > -1$ and 0 otherwise. The offset is chosen so that the function matches the exponential nonlinearity at $x = 0$ (green). In addition, we used $f(x) = \log(1 + e^x)$ (red), i.e., a smooth interpolation between the exponential for small x with linear asymptotic behavior for large x. (B) Simulated spike trains for two additional nonlinearities for the two data sets that were shown to diverge in simulations (Fig 2B and 2C).
(TIF)

S2 Fig. Same as Fig 7A with $c = 2$ s^{-1}. Spike trains are simulated from a nonlinear Hawkes model with fixed baseline $c = 2$ s^{-1} and an auto-history kernel consisting of two exponentials with amplitudes J_r, J_a, and corresponding time constants $\tau_r = 0.02$ s and $\tau_a = 0.1$ s. Observed divergence times for simulated spike trains are color-coded (same scale as in Fig 5). In the dashed region, no finite divergence times were observed.
(TIF)

S3 Fig. Same as Fig 7A with c = 10 s^{-1}. Spike trains are simulated from a nonlinear Hawkes model with fixed baseline c = 10 s^{-1} and an auto-history kernel consisting of two exponentials with amplitudes J_r, J_a, and corresponding time constants τ_r = 0.02 s and τ_a = 0.1 s. Observed divergence times for simulated spike trains are color-coded (same scale as in Fig 5). In the dashed region, no finite divergence times were observed.
(TIF)

Author Contributions

Conceptualization: FG MD WT.

Data curation: FG.

Formal analysis: FG MD.

Funding acquisition: WT.

Investigation: FG MD.

Methodology: FG MD.

Project administration: WT.

Resources: WT.

Software: FG MD.

Supervision: WT.

Visualization: FG.

Writing – original draft: FG MD.

Writing – review & editing: FG MD WT.

References

1. Brown EN, Kass RE, Mitra PP. Multiple neural spike train data analysis: state-of-the-art and future challenges. Nature Neuroscience. 2004 may; 7(5):456–461. doi: 10.1038/nn1228 PMID: 15114358

2. 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 feb; 93(2):1074–89. doi: 10.1152/jn.00697.2004 PMID: 15356183

3. Stevenson IH, Rebesco JM, Miller LE, Körding KP. Inferring functional connections between neurons. Current Opinion in Neurobiology. 2008 dec; 18(6):582–588. doi: 10.1016/j.conb.2008.11.005 PMID: 19081241

4. Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature. 2008 jul; 454(7207):995–999. doi: 10.1038/nature07140 PMID: 18650810

5. Truccolo W, Hochberg LR, Donoghue JP. Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. Nature Neuroscience. 2010 jan; 13(1):105–111. doi: 10.1038/nn.2455 PMID: 19966837

6. Gerhard F, Kispersky T, Gutierrez GJ, Marder E, Kramer M, Eden U. Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone. PLOS Computational Biology. 2013 jan; 9(7):e1003138. doi: 10.1371/journal.pcbi.1003138 PMID: 23874181

7. Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal Dynamics. Cambridge: Cambridge University Press; 2014.

8. Takahashi K, Kim S, Coleman TP, Brown KA, Suminski AJ, Best MD, et al. Large-scale spatiotemporal spike patterning consistent with wave propagation in motor cortex. Nature Communications. 2015; 6(May):7169. doi: 10.1038/ncomms8169 PMID: 25994554

9. Brémaud P, Massoulié L. Stability of nonlinear Hawkes processes. The Annals of Probability. 1996 jul; 24(3):1563–1588. doi: 10.1214/aop/1065725193

10. Daley DJ, Vere-Jones D. An Introduction to the Theory of Point Processes, Volume 1 (2nd ed.). New York: Springer; 2003.

11. Mensi S, Naud R, Gerstner W. From Stochastic Nonlinear Integrate-and-Fire to Generalized Linear Models. In: Shawe-Taylor J, Zemel R, Bartlett P, Pereira F, Weinberger K, editors. NIPS; 2011. p. 1377–1385.

12. Mensi S. A new Mathematical Framework to Understand Single Neuron Computations [PhD Thesis]. EPFL (Lausanne, Switzerland); 2014.

13. Weber AI, Pillow JW. Capturing the dynamical repertoire of single neurons with generalized linear models. arXiv. 2016 feb;p. 1–19.

14. Prinz AA, Bucher D, Marder E. Similar network activity from disparate circuit parameters. Nature Neuroscience. 2004 dec; 7(12):1345–1352. doi: 10.1038/nn1352 PMID: 15558066

15. Marder E, Goaillard JM. Variability, compensation and homeostasis in neuron and network function. Nature Reviews Neuroscience. 2006 jul; 7(7):563–574. doi: 10.1038/nrn1949 PMID: 16791145

16. Benda J, Herz AVM. A universal model for spike-frequency adaptation. Neural Computation. 2003; 15:2523–2564. doi: 10.1162/089976603322385063 PMID: 14577853

17. Pozzorini C, Naud R, Mensi S, Gerstner W. Temporal whitening by power-law adaptation in neocortical neurons. Nature Neuroscience. 2013 jul; 16(7):942–948. doi: 10.1038/nn.3431 PMID: 23749146

18. Hawkes AG. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. Biometrika. 1971 apr; 58(1):83. doi: 10.1093/biomet/58.1.83

19. Hawkes AG. Point Spectra of Some Mutually Exciting Point Processes. Journal of the Royal Statistical Society Series B. 1971; 33(3):438–443.

20. Hawkes AG, Oakes D. A Cluster Process Representation of a Self-Exciting Process. Journal of Applied Probability. 1974 sep; 11(3):493. doi: 10.1017/S0021900200096273

21. Massoulié L. Stability results for a general class of interacting point processes dynamics, and applications. Stochastic Processes and their Applications. 1998 jun; 75(1):1–30.

22. Priestley M B. Spectral analysis and time series. London: Academic Press; 1982.

23. Percival DB, Walden AT. Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. vol. 38. Cambridge: Cambridge University Press; 1993.

24. Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The Time-Rescaling Theorem and Its Application to Neural Spike Train Data Analysis. Neural Computation. 2002 feb; 14(2):325–346. doi: 10.1162/08997660252741149 PMID: 11802915

25. Gerhard F, Haslinger R, Pipa G. Applying the Multivariate Time-Rescaling Theorem to Neural Population Models. Neural Computation. 2011 jun; 23(6):1452–1483. doi: 10.1162/NECO_a_00126 PMID: 21395436

26. Naud R, Gerstner W. Coding and Decoding with Adapting Neurons: A Population Approach to the Peri-Stimulus Time Histogram. PLOS Computational Biology. 2012 oct; 8(10):e1002711. doi: 10.1371/journal.pcbi.1002711 PMID: 23055914

27. Deger M, Schwalger T, Naud R, Gerstner W. Fluctuations and information filtering in coupled populations of spiking neurons with adaptation. Physical Review E. 2014 dec; 90(6):062704. doi: 10.1103/PhysRevE.90.062704

28. Brookings T, Grashow R, Marder E. Statistics of Neuronal Identification with Open- and Closed-Loop Measures of Intrinsic Excitability. Frontiers in Neural Circuits. 2012 jan; 6(April):1–14.

29. Potter SM, El Hady A, Fetz EE. Closed-loop neuroscience and neuroengineering. Frontiers in Neural Circuits. 2014 sep; 8:115. doi: 10.3389/fncir.2014.00115 PMID: 25294988

30. Truccolo W, Ahmed OJ, Harrison MT, Eskandar EN, Cosgrove GR, Madsen JR, et al. Neuronal Ensemble Synchrony during Human Focal Seizures. Journal of Neuroscience. 2014 jul; 34(30):9927–9944. doi: 10.1523/JNEUROSCI.4567-13.2014 PMID: 25057195

31. Vargas-Irwin CE, Franquemont L, Black MJ, Donoghue JP. Linking objects to actions: encoding of target object and grasping strategy in primate ventral premotor cortex. Journal of Neuroscience. 2015; 35(30):10888–10897. doi: 10.1523/JNEUROSCI.1574-15.2015 PMID: 26224870

32. Rule ME, Vargas-Irwin C, Donoghue JP, Truccolo W. Contribution of LFP dynamics to single-neuron spiking variability in motor cortex during movement execution. Frontiers in Systems Neuroscience. 2015 jun; 9(June):1–16. https://doi.org/10.3389/fnsys.2015.00089

33. Mensi S, Naud R, Pozzorini C, Avermann M, Petersen CCH, Gerstner W. Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. Journal of Neurophysiology. 2012 mar; 107(6):1756–1775. doi: 10.1152/jn.00408.2011 PMID: 22157113

34. Zaytsev YV, Morrison A, Deger M. Reconstruction of recurrent synaptic connectivity of thousands of neurons from simulated spiking activity. Journal of Computational Neuroscience. 2015; 39(1):77–103. doi: 10.1007/s10827-015-0565-5 PMID: 26041729

35. Shinomoto S, Shima K, Tanji J. Differences in spiking patterns among cortical neurons. Neural Computation. 2003 dec; 15(12):2823–2842. doi: 10.1162/089976603322518759 PMID: 14629869

36. Mochizuki Y, Onaga T, Shimazaki H, Shimokawa T, Tsubo Y, Kimura R, et al. Similarity in Neuronal Firing Regimes across Mammalian Species. Journal of Neuroscience. 2016; 36(21):5736–47. doi: 10.1523/JNEUROSCI.0230-16.2016 PMID: 27225764

37. Nadkarni D. The Incidental Parameter Problem in Network Analysis for Neural Spiking Data [PhD Thesis]. Brown University (Providence, RI, USA); 2015.

38. Smith AC, Brown EN. Estimating a State-Space Model from Point Process Observations. Neural Computation. 2003; 15:965–991. doi: 10.1162/089976603765202622 PMID: 12803953

39. Macke JH, Buesing L, Cunningham JP, Yu BM, Shenoy KV, Sahani M. Empirical models of spiking in neural populations. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ, editors. Advances in Neural Information Processing Systems 24. Curran Associates, Inc.; 2011. p. 1350–1358.

40. Buesing L, Macke JH, Sahani M. Learning stable, regularised latent models of neural population dynamics. Network (Bristol, England). 2012 jan; 23(1–2):24–47.

41. Aghagolzadeh M, Truccolo W. Latent state-space models for neural decoding. Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 aug;2014:3033–3036.

42. 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. doi: 10.1109/TNSRE.2015.2470527 PMID: 26336135

43. Schwalger T, Deger M, Gerstner W. Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size. Preprint: arXiv:161100294 [q-bioNC]. 2016;.

44. Koh K, Kim SJ, Boyd S. An Interior-Point Method for Large-Scale L1 -Regularized Logistic Regression. The Journal of Machine Learning Research. 2007; 8:1519–1555.

45. Chen Z, Putrino DF, Ghosh S, Barbieri R, Brown EN. Statistical inference for assessing functional connectivity of neuronal ensembles with sparse spiking data. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2011; 19(2):121–135. doi: 10.1109/TNSRE.2010.2086079 PMID: 20937583

46. Zhao M, Batista A, Cunningham JP, Chestek C, Rivera-Alvidrez Z, Kalmar R, et al. An L1-regularized logistic model for detecting short-term neuronal interactions. Journal of Computational Neuroscience. 2012; 32(3):479–497. doi: 10.1007/s10827-011-0365-5 PMID: 22038503

47. Galves A, Löcherbach E. Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets. Journal of Statistical Physics. 2013; 151(5):896–921. doi: 10.1007/s10955-013-0733-9

48. O'Doherty JE, Lebedev MA, Ifft PJ, Zhuang KZ, Shokur S, Bleuler H, et al. Active tactile exploration using a brain-machine-brain interface. Nature. 2011 oct; 479(7372):228–31. doi: 10.1038/nature10489 PMID: 21976021

49. Truccolo W, Donoghue J, Hochberg L, Eskandar E, Madsen J, Anderson W, et al. Single-neuron dynamics in human focal epilepsy. Nature Neuroscience. 2011 may; 14(5):635–641. doi: 10.1038/nn.2782 PMID: 21441925

50. Cardanobile S, Rotter S. Multiplicatively interacting point processes and applications to neural modeling. Journal of Computational Neuroscience. 2010 apr; 28(2):267–284. doi: 10.1007/s10827-009-0204-0 PMID: 20052525

51. Carandini M. Amplification of trial-to-trial response variability by neurons in visual cortex. PLoS Biology. 2004 sep; 2(9):E264. doi: 10.1371/journal.pbio.0020264 PMID: 15328535

52. Jolivet R, Rauch A, Lüscher HR, Gerstner W. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. Journal of Computational Neuroscience. 2006 aug; 21(1):35–49. doi: 10.1007/s10827-006-7074-5 PMID: 16633938

53. Van Kampen NG. Stochastic processes in physics and chemistry. Amsterdam: Elsevier; 1992.

54. Gewaltig MO, Diesmann M. NEST (NEural Simulation Tool). Scholarpedia. 2007; 2(4):1430. doi: 10.4249/scholarpedia.1430