Noname manuscript No.
(will be inserted by the editor)

Stability of Point Process Spiking Neuron Models

Yu Chen* · **Qi Xin*** · **Valérie Ventura** · **Robert E. Kass**

Received: date / Accepted: date

Abstract Point process regression models, based on generalized linear model (GLM) technology, have been widely used for spike train analysis, but a recent paper by Gerhard et al. described a kind of instability, in which fitted models can generate simulated spike trains with explosive firing rates. We analyze the problem by extending the methods of Gerhard et al. First, we improve their instability diagnostic and extend it to a wider class of models. Next, we point out some common situations in which instability can be traced to model lack of fit. Finally, we investigate distinctions between models that use a single filter to represent the effects of all spikes prior to any particular time t, as in a 2008 paper by Pillow et al., and those that allow different filters for each spike prior to time t, as in a 2001 paper by Kass and Ventura. We re-analyze the data sets used by Gerhard et al., introduce an additional data set that exhibits bursting, and use a well-known model described by Izhikevich to simulate spike trains from various ground truth scenarios. We conclude that models with multiple filters tend to avoid instability, but there are unlikely to be universal rules. Instead, care in data fitting is required and models need to be assessed for each unique set of data.

Keywords Generalized linear model · outlier trials · point process regression · spike train

Y. Chen(✉) · V. Ventura · R. Kass
Carnegie Mellon University, Pittsburgh, PA 15213, USA
E-mail: yuc2@andrew.cmu.edu

Q. Xin
University of Science and Technology of China, Hefei, Anhui 230026, China

* Two authors have contributed equally to this paper.

1 Introduction

Point process regression models based on the framework of generalized linear models (GLMs) have been applied to a wide variety of spiking neuron data (Kass et al. (2014), Weber and Pillow (2017), and references therein). These models, which may be considered nonlinear Hawkes processes (Chen et al., 2017; Eichler et al., 2017), allow neural firing rates to depend on spiking history. Recently, however, Gerhard et al. (2017) reported that models fitted to real data sets could be unstable in the sense that their firing rates could evolve to become arbitrarily large, generating unrealistic spike trains, even when standard goodness-of-fit tests fail to identify lack of fit (see Figure 1 for two examples). In this paper we identify several factors that can lead to this problem, we provide additional analysis for diagnosing it, and we present methods to improve model stability.

In some circumstances, causes of instability are easy to identify and easy to fix. The problem of stability, however, leads naturally to an interesting detail in GLM-type modeling of spike trains. When spike trains are modeled as point processes, the firing rate is defined by the conditional intensity function

$$\lambda(t|H_t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\big(\Delta N_{(t,t+\Delta t]} = 1 | H_t\big)}{\Delta t} \tag{1}$$

where H_t is the set of spikes prior to time t, known as the spiking history up to time t, and $\Delta N_{(t,t+\Delta t]}$ is the number of spikes in the interval $(t, t+\Delta t]$. This succinct representation can also incorporate stimulus effects and coupling effects and its implementation can take advantage of a large body of knowledge about generalized regression models (Kass et al., 2014). Here we only consider the history effects without external stimulus and

coupling neurons. There are many ways to capture the effects of the history H_t on the intensity. Letting t_{j*} be the jth spike time counting backwards prior to time t, a concise and intuitive assumption, for steady-state scenarios (where the baseline rate is constant), takes the intensity to have the form

$$\log \lambda(t|H_t) = \beta_0 + \sum_j h(t - t_{j*}) \tag{2}$$

where $h(u)$ is a smooth function, and the summation extends to all spikes that precede time t (within a given trial, if there are trials). This is the form used by Pillow et al. (2008) and by Gerhard et al. (2017). Pillow et al. (2008) referred to $h(u)$ as a post-spike filter. An alternative model, used by Kass and Ventura (2001), instead allows the effects of each previous spike to be different:

$$\log \lambda(t|H_t) = \beta_0 + \sum_{j=1}^{k} h_j(t - t_{j*}). \tag{3}$$

If each function h_j involves separate free parameters, then the model in (3) would typically have more parameters than the model in (2). A main contribution of this paper is to describe situations under which this additional flexibility can be useful. In particular, we suggest that, in realistic scenarios, models of the form (3) tend to be stable.

One issue in using (3) is that the number of terms k must be selected. In theory the number could be infinite as long as a suitable condition is placed on the functions h_j, such as $\sum M_j < \infty$, where $M_j = \max_u h_j(u)$, but in practice Kass and Ventura (2001) selected k by applying the likelihood ratio test; here, in Section 4, we will suggest another criterion, based on stability. Similarly, in practice, the summation in (2) extends over a fixed window of time preceding t, having length we label T_h (at most, the total length of the experiment), leading to the alternative representation

$$\log \lambda(t|H_t) = \beta_0 + \sum_{t_{j*} \in (t-T_h, t]} h(t - t_{j*}). \tag{4}$$

We refer to a model described by (4) as Fixed Length Filter (FLF), and those described by (3) as Fixed Number Filter (FNF), where the fixed number refers to the fixed number of spikes. In our analysis we have found it useful to further categorize FNF models by considering the special case in which $h_j(u) = h_1(u)$, for all j and all u. These models we write these as $\mathrm{FNF_S}$, where S stands for single filter. The more general case we write as $\mathrm{FNF_M}$, with M for multiple. Note that $\mathrm{FNF_S}$ differs from FLF in that the number of spikes is fixed rather than the length of the time interval, but both models use a single filter while $\mathrm{FLF_M}$ uses multiple filters.

We begin, in Section 2, by giving some analytical stability results along the lines of those in Gerhard et al. (2017). In Section 3 we identify several kinds of model mis-specification that lead to instability, and we note potential solutions. In Section 4 we focus on FNF models and the variation that replaces the constant β_0 with a time-varying function $\beta(t)$. An analytical diagnostic method is then used to select the number of previous spikes to be considered in the model, i.e., the number of terms k in (3). In section 5, we compare FLF and FNF models. We close in Section 6 with advices on the use of these models.

2 Stability analysis

Figure 1 shows two examples of unstable simulations from FLF models (equation 4) fitted to the *Monkey-PMv* and *Human-Cortex* datasets described in Table 1, Appendix A, using the smooth basis method of Pillow et al. (2008). The fitted FLF models pass the original and discrete KS goodness of fit tests (Brown et al., 2002; Haslinger et al., 2010) but they are unstable, in the sense that the firing rates of some or all the simulated spike trains evolve to become arbitrarily large, generating unrealistic spike trains. We emphasize that instability is not a matter of extrapolation to unseen data outside the experimental range of time. Rather, if a model is unstable, in simulations it can evolve to producing firing rates far in excess of those seen in real data, which makes it patently unrealistic as a representation of neural physiology.

Gerhard et al. (2017) argue that a reliable diagnostic of model instability can be obtained from the relationship between the firing rate A_0 before the last spike at t_{*1} in an interval, and the firing rate after t_{*1}. They approximate A_0 with the average firing rate in the interval $(t - T_h, t_{1*})$. Then they rewrite the fitted FLF model (Equation 4) as

$$\log \lambda(t|H_t) = \\ \beta_0 + h(t - t_{1*}) + \sum_{t_{j*} \in (t-T_h, t_{1*})} h(t - t_{j*}),$$

and approximate the summation by its expectation under the assumption that the point process in the interval $(t - T_h, t_{1*})$ is homogeneous Poisson, yielding

$$\log \lambda(t|H_t) \approx \\ \beta_0 + h(t - t_{1*}) + A_0 \int_{t-t_{1*}}^{T_h} \big(e^{h(u)} - 1\big) \mathrm{d}u. \tag{5}$$

Gerhard et al. (2017) then use Equation 5 to derive the approximate PDF of the future ISI $t_* - t_{1*}$, where t_* is

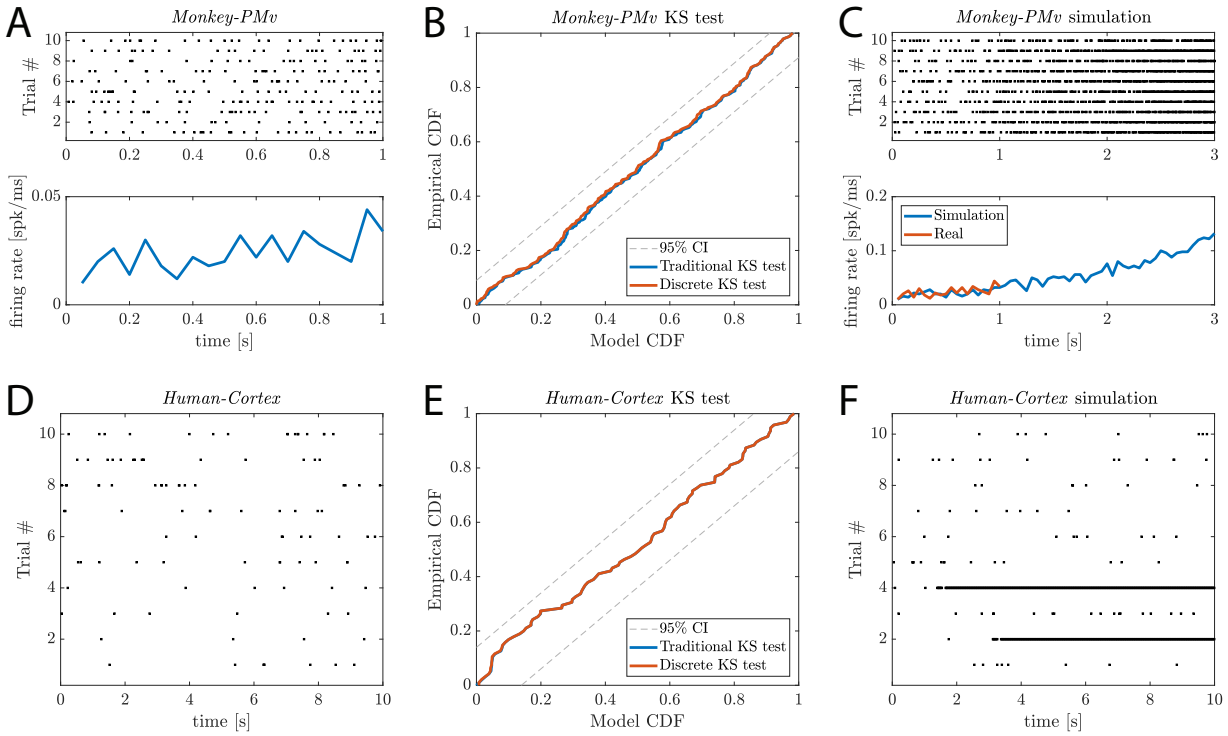

Fig. 1: Simulation divergence of FLF models (equation 4) fitted to data. (A,D) Spike time raster plot and PSTH for datasets *Monkey-PMv* and *Human-Cortex*. (B,E) The fitted FLF models pass the original and discrete KS tests of Brown et al. (2002) and Haslinger et al. (2010); the two tests overlap so they are hard to distinguish. (C,F) Spike time raster plot and PSTH of spike trains simulated from the fitted FLF models, using algorithm 2 in Appendix C. (C) If the simulation lasts longer than the training session, the firing rate keeps growing to produce ISIs that are shorter than the refractory period. (F) The simulated spike trains resemble the observed data except for trials 2 and 4, which have many more spikes than the observed spike trains.

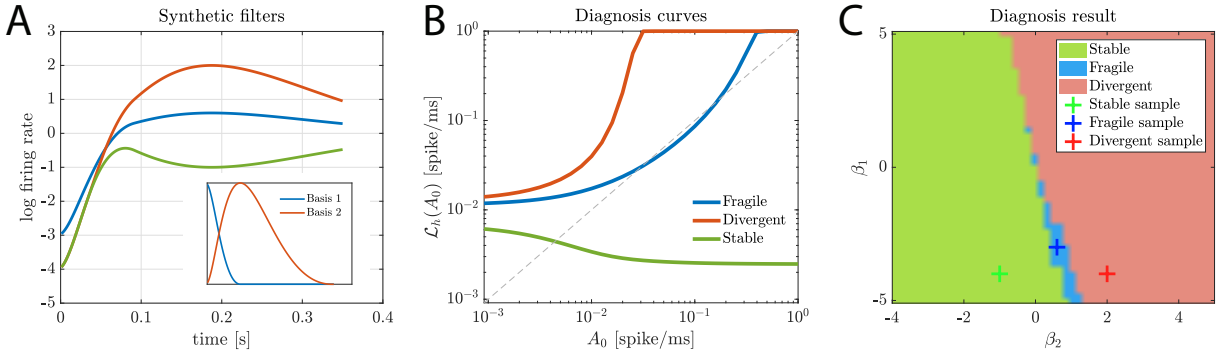

Fig. 2: (B) Three diagnostic curves corresponding to the three FLF models (equation 4) with baseline rates $\beta_0 = -4$ and filters $h(t)$ shown in (A) in matching colors, where $h(t) = \beta_1 \cdot \mathrm{B}_1(t) + \beta_2 \cdot \mathrm{B}_2(t)$, $\mathrm{B}_1(t) = e^{-t/0.02}$ and $\mathrm{B}_2(t) = e^{-t/0.1}$ are the smooth basis functions used in Pillow et al. (2008) and shown in the insert in (A), and $T_h = 0.35$ sec. The values of (β_1, β_2) for the three models are marked as crosses in (C). (C) Diagnostic map for the above model as β_1 and β_2 vary.

the time of the next spike after t_{1*}, and calculate the firing rate after t_{1*} as the reciprocal of the mean future ISI:

$$\mathcal{L}_h(A_0) = \frac{1}{\mathbb{E}[t_* - t_{1*}]}. \tag{6}$$

Equation 6 is a function of A_0 because Equation 5 is a function of A_0. The instability diagnostic is thus obtained by plotting $\mathcal{L}_h(A_0)$ versus A_0, as in Figure 2B, for $A_0 \in [0, \lambda_{max}]$, where λ_{max} is the maximum possible firing rate. Without loss of generality, in this paper we do not build a refractory period in the models we consider, except to reproduce Gerhard et al. (2017) Figure 4 (see Appendix B Figure 9), so λ_{max} is our simulation resolution of 1000 spikes per second. We examine intersections of the diagnostic curve with the line $\mathcal{L}_h(A_0) = A_0$, referring to them as *cross points*.

A model is deemed

- **divergent** if all cross points exceed λ_{thr} or $\mathcal{L}_h(A_0)$ is always larger than A_0 (e.g. Figure 2B, red curve), where λ_{thr} is a threshold rate judged too high physiologically; here we used $\lambda_{thr} = 0.9 \cdot \lambda_{max}$ spikes/sec;
- **stable** if the number of cross points is odd and they are all below λ_{thr} (green curve);
- **fragile** otherwise (blue curve).

Note that the diagnostic green curve in Figure 2B exceeds the first sequent for small values of A_0, which is desirable because otherwise the firing rate would eventually decrease down to zero. A divergent model yields unstable simulations, whereas spike trains simulated from a fragile model might first look stable and then degenerate. Therefore the difference between divergent and fragile models is the duration it takes for spike trains to become unstable. Without loss of generality of our results, we do not distinguish between divergent and fragile models, and consider them both unstable.

Gerhard et al. (2017) validated their model stability diagnostic against spike train data: they considered a model family F_θ parametrized by θ, and for each value of θ in a range, they (i) simulated a 10 sec. long spike train from F_θ (e.g. using algorithm 2 in Appendix C), and deemed the spike train unstable if the model generated over $0.9 \cdot \lambda_{max}$ spikes in the last second, (ii) produced the diagnostic curve and determined from it if the model was stable, fragile, or divergent and (iii) plotted θ against the outcomes in (i) and (ii). Such plots are displayed in Appendix B Figure 9 for two model families F_θ. Figure 9A shows the same stability map as in Gerhard et al. (2017) Figure 4, where F_θ is an FLF model (Equation 4) with baseline rate $\beta_0 = -5.3$ and filter $h(t) = \beta_1 \cdot \mathrm{B}_1(t) + \beta_2 \cdot \mathrm{B}_2(t) + \mathrm{Dip(t)}$, $\theta = (\beta_1, \beta_2)$, $\mathrm{B}_1(t) = e^{-t/0.02}$ and $\mathrm{B}_2(t) = e^{-t/0.1}$ are the smooth basis functions used in Pillow et al. (2008) and shown in the insert in Figure 2A, Dip(t) is a negative window function modeling a 2 msec. refractory period, and the filter length is $T_h = 0.2$ sec. In Figure 9C, F_θ is an FLF model with baseline rate $\beta_0 = -4$, filter $h(t) = \beta_1 \cdot \mathrm{B}_1(t) + \beta_2 \cdot \mathrm{B}_2(t)$ with $B_1(t)$ and $B_2(t)$ defined above, and filter length $T_h = 0.35$ sec.

The stability maps in Figure 9A,C suggest that the diagnostic is mostly reliable, except in small regions of the parameter spaces. This happens because Gerhard et al. (2017) replaced $h(u)$ by the Taylor expansion $(\exp h(u) - 1)$ in Equation 5, which is accurate only when $h(u)$ is small. Without this approximation, Equation 5 becomes

$$\log \lambda(t|H_t) \approx \beta_0 + h(t - t_{1*}) + A_0 \int_{t-t_{1*}}^{T_h} h(u)\mathrm{d}u, \tag{7}$$

and the diagnostic is still tractable, as shown in Appendix C. Figure 9B,D show the updated stability maps based on Equation 7. The agreement between diagnostic and simulation is very close, and closer than in Figures 9A,C, so we use the updated diagnostic in the rest of the paper. Figure 9D is reproduced in Figure 2C.

To solve the stability problem when a model is found to be divergent or fragile, Gerhard et al. (2017) suggested stabilizing it by refitting to the data with the constraint that its parameters lie in the stable region of the parameter space. In the next section, we identify three data features that might lead to unstable simulation models, namely small sample size, time varying firing rates, and trial to trial variability or outlier trials, and we provide alternative suggestions for stabilization: collecting more data, fitting inhomeogeneous rate models, and removing outliers, respectively.

3 Special cases of FLF model instability

A feature that might lead to unstable models is a small sample size. Indeed fitting a model to a small dataset yields parameter estimates that have large variances and, therefore, that could lie in unstable regions of the parameter space by chance, even if the true parameters lie in stable regions. Collecting more data, if possible, would reduce the variability of parameter estimates and stabilize the model.

Next, consider the *Monkey-PMv*, shown in Figure 1A. An FLF model fitted to the data satisfies the KS goodness-of-fit tests (Figure 1B), yet simulations from the model diverge (Figure 1C). Because the peri-stimulus time histogram (PSTH) in Figure 1A appears to increase, we fit a time-varying baseline rate $\beta(t)$ in place of β_0 in Equation 4. That model fits the data somewhat better according to a likelihood ratio test ($p = 0.046$), and data simulated from it do not diverge (Figure 3C). (To simulate data past the maximum experimental time of one second, we set $\beta(t) = \beta(1)$ for $t \geq 1$ sec.) Figure 3D,E,F shows a similar outcome when we apply the same analysis to synthetic data generated from an inhomogeneous *Izhikevich* model (algorithm 1, Appendix B). Hence, in the presence of a time-varying trial-averaged rate, fitting a constant rate term can produce instability and fitting a time-varying rate can rectify the problem.

Finally, consider the *Human-Cortex* data displayed in Figure 1D. An FLF model fitted to the data satisfies the KS goodness-of-fit tests (Figure 1E) but two

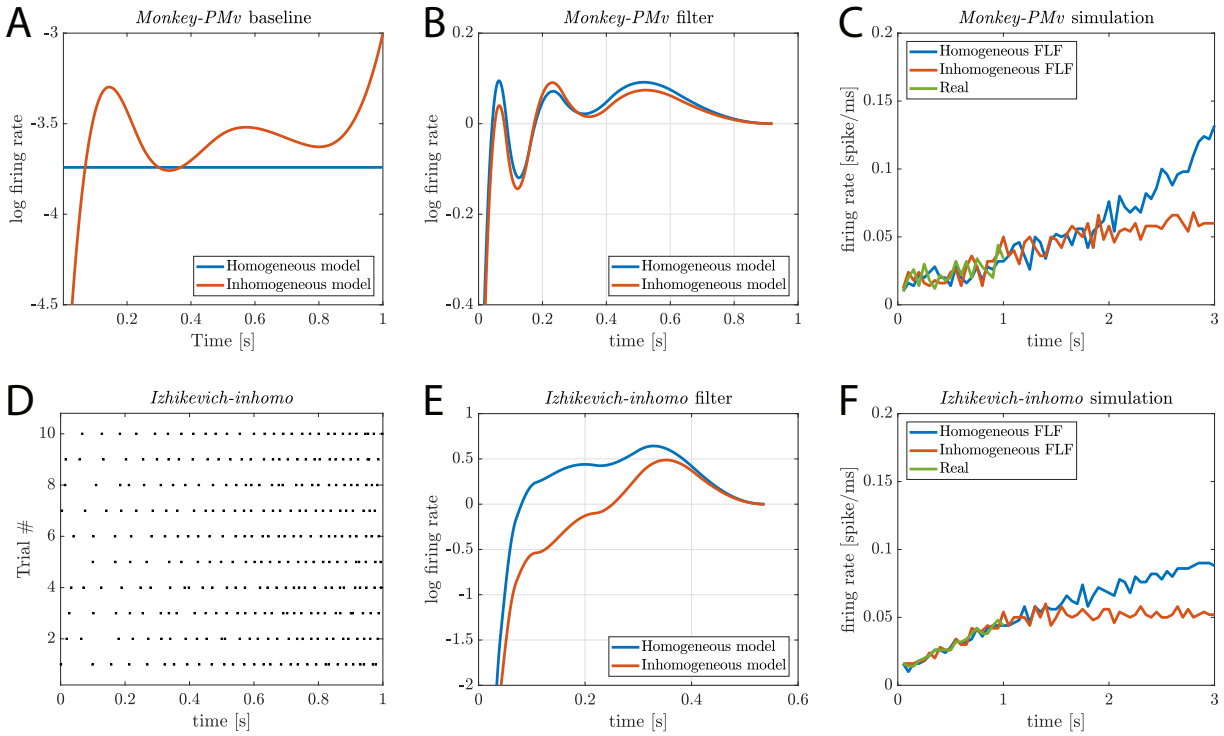

Fig. 3: Stability of FLF models (equation 4) with constant and time varying baseline firing rates. (A) Constant (blue) and time varying (red) baseline rates of FLF models fitted to the *Monkey-PMv* data: the baseline appears to vary. A likelihood ratio test confirms that the time varying baseline model fits the data significantly better (p = 0.046). (B) Fitted filters of the constant and time-varying baseline models. (C) PSTHs of the observed data and of data simulated from the fitted homogeneous and inhomogeneous FLF models: the homogeneous model simulates unstable spike trains; the inhomogeneous model is simulation stable. (D, E, F) Same analysis applied to artificial data generated from an inhomogeneous *Izhikevich* model (algorithm 1 in Appendix C). (D) Data raster plot. (E) Fitted filters of the homogeneous and inhomogeneous baseline FLF models: the latter is mostly below the former, which may reduce the chance of unstable simulated spike trains. (F) Indeed, the homogeneous model produces data whose rate diverges; the inhomogeneous model appears to be stable. Note that the inhomogeneous model fits the data significantly better according to a likelihood ratio test ($p \ll 0.001$).

out of 10 spike trains simulated from it diverge (Figure 1F). Figure 4A show that trials 8, 9, and 10 have rather large spike counts compared to the others, so there might be excess trial-to-trial variability or outlier trials that might cause the instability. To examine the extent to which some trials may be unusually different than others, compute the distance of each spike train from a central spike train $\overline{ST}$ (defined below) based on a spike train metric devised by Wu and Srivastava (2011). This metric measures the discrepancy between two spike trains by counting the number of spikes in one spike train that can be matched by spikes in the other spike train using a smooth deformation of time, or "time-warping function." If there are N_1 and N_2 spikes in two spike trains ST_1 and ST_2, the distance between the two spike trains is defined as

$$d(ST_1, ST_2) = \inf_{\gamma \in \Gamma} \left(N_1 + N_2 - 2 \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} I_{[t_i = \gamma(s_j)]} + \eta \int_0^T \left(1 - \sqrt{\gamma'(t)}\right)^2 \mathrm{d}t \right), \tag{8}$$

where I is the indicator function, t_i and s_j are spike times from ST_1 and ST_2, respectively, and $\Gamma(t)$ is the set of all continuous and piecewise differentiable time warping functions γ such that $\gamma(0) = 0$, $\gamma(T) = T$, and $0 < \gamma'(T) < \infty$. In practice γ is approximated by a piecewise linear function from $[0, 0]$ to $[T, T]$ in a discrete grid and the tuning parameter η is set to $(N_1 + N_2) \cdot c/2T$, with $5 \leq c \leq 25$ (Wu and Srivastava, 2011). The choice of c in this range has little impact on results. The first term on the right hand side of Equation 8 measures how close ST_1 is to the time wrapped ST_2, and the second penalizes the deviation

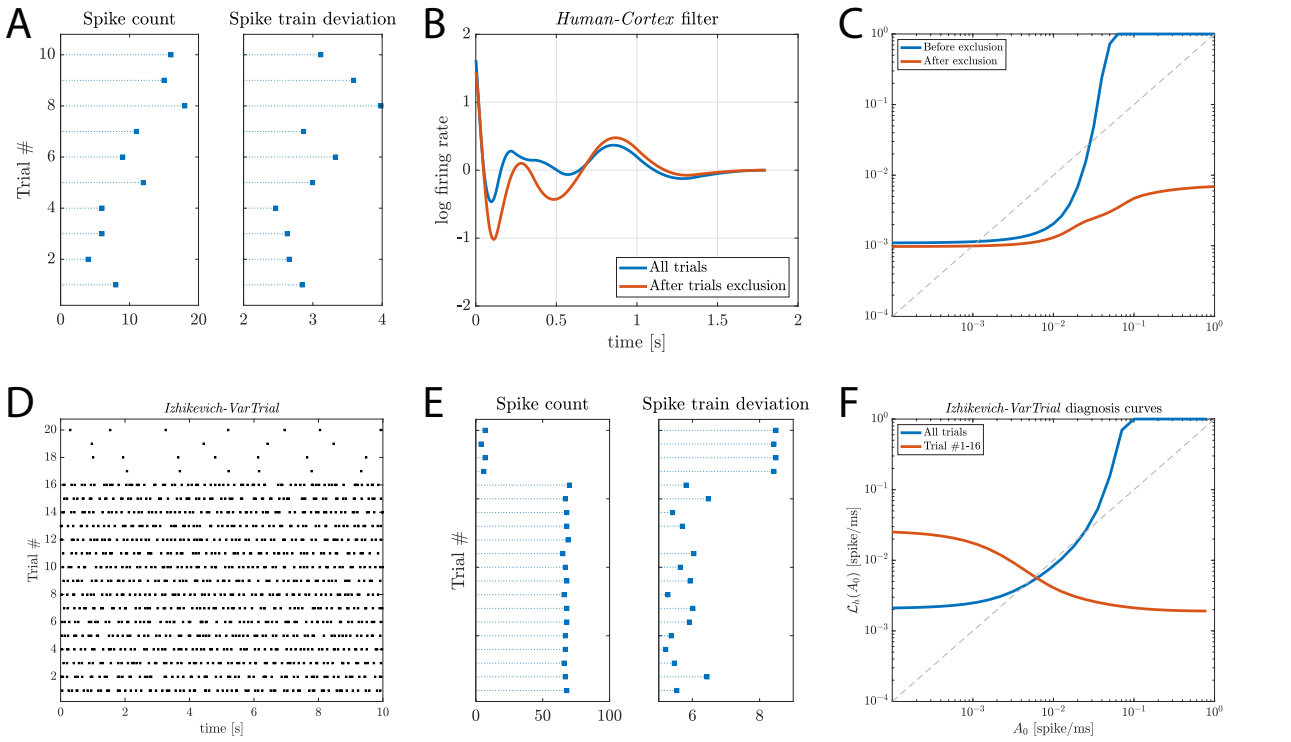

Fig. 4: Trial-to-trial variability affects simulation stability. (A) Spike train counts and deviations from the central spike train for the *Human-Cortex* dataset. Trials 8 and 9 might be outliers. (B) Filters of FLF models (Equation 4) fitted to data with and without suspected outliers. The latter remains mostly below the former for all t, which reduces the possibility of simulating unstable spike trains. (C) Diagnostic curves for models fitted before and after removing the suspected outliers: the model becomes stable after removal. Synthetic spike trains simulated from that more are indeed stable (not shown). (D,E,F) Same analysis applied to spike trains generated from a two-rate *Izhikevich* model. (D) The dataset is composed of 16 spike trains with a high firing rate, and four with a low firing rate. (E) spike counts and deviations from the central spike train clearly identify two groups of spike trains. (F) Diagnostic curves of FLF models fitted to the full datasets (blue), and to the dataset after the four unusual spike trains are removed. The former diagnoses an unstable model, the latter a stable model. Spike trains simulated from the latter model are indeed stable (not shown).

of the time warping transformation from the identity function $\gamma(t) = t$. The central spike train $\overline{ST}$, is defined as

$$\overline{ST} = \arg\min_{C \in \mathcal{S}} \sum_{i=1}^{n} d(ST_i, C),$$

where $\mathcal{S}$ is the set of all spike trains. We then compute each distance

$$d_i = d(\overline{ST}, ST_i), \qquad i = 1, \ldots, n$$

and use d_i to identify unusually discrepant trials.

The deviations d_i for the *Human-Cortex* spike trains are shown in Figure 4A. Trials 8 and 9 have the largest values of d_i, and they also have large spike counts. After removing them, the fitted FLF model becomes stable, according to the diagnostic plot in Figure 4C. Figure 4B shows that the filter fitted after excluding the outliers lies mostly below the initial filter, which reduces the chance of simulation divergence. We note that if we remove only one of these trials the fitted model is again unstable. Furthermore, if we remove any other 2 trials the fitted model is unstable. Figure 4D,E,F shows that a similar analysis applied to data generated from *Izhikevich* models with two different firing rates – 16 spike trains have a large firing rate and four have a small firing rate – yields similar conclusions: that is, outlier trials can destabilize models, and careful data pre-processing to remove them might improve stability.

Different kinds of outliers may have to be treated differently. Outlier trials resulting from bad recordings should be removed. But absent such experimental difficulties it remains important to consider unusual features of the data, and to avoid models that fail to account for those features. In the synthetic two-rate *Izhikevich* dataset, for example, four trials are noticeably sparse, making a common rate model fit poorly. Methods based on models that allow for excess trial-to-trial variability are available (Ventura et al., 2005).

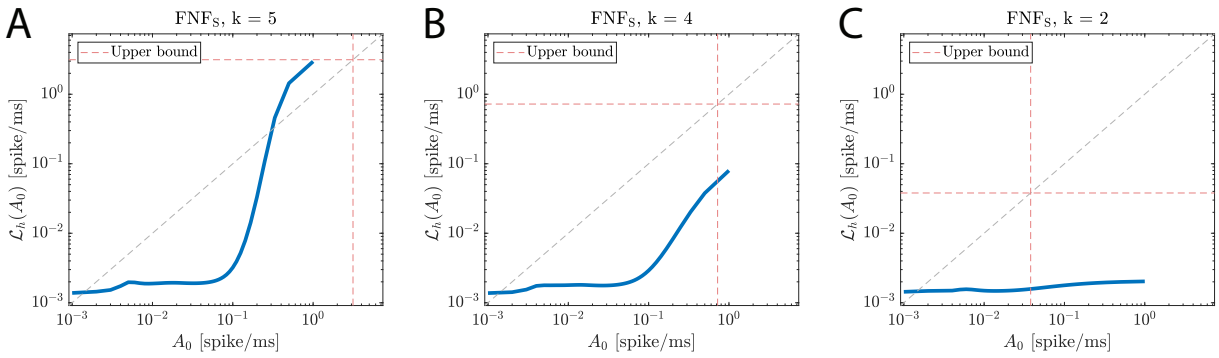

Fig. 5: Diagnostic curves of FNF$_S$ models with $k = 2, 4, 5$ fitted to the *Human-Cortex* dataset. The dashed red lines are the models' maximum firing rates. The largest model with $k = 5$ is unstable because the diagnostic curve is above the first sequent after it last intersects it. The two other models are stable.

4 Stability of FNF models

To evaluate the performance of FNF models, we extend the method of Gerhard et al. (2017) to obtain a stability diagnostic, further allowing the baseline rate to be a time-varying function $\beta(t)$, as in Kass and Ventura (2001). If we approximate the firing rate before the last spike t_{1*} with the reciprocal of the mean ISI, $A_0 = 1/\tau$, and replace the ISIs by their expectation in Equation 6, we obtain:

$$\log \lambda(t|H_t) \approx \beta(t) + \sum_{k=1}^{k} h_j(t - t_{1*} + (k-1)\tau). \quad (9)$$

As in Section 2, we then use this approximation to derive the approximate PDF of the future ISI $t_* - t_{1*}$, and calculate the firing rate $\mathcal{L}_h(A_0)$ after t_{1*} as the reciprocal of the mean future ISI (see Equation 6). The diagnostic for a fitted model is again based on a plot of $\mathcal{L}_h(A_0)$ against A_0, and its stability determined using the rules in the boxed text in Section 2. For example, Figure 5 shows the diagnostics of three FNF$_S$ models fitted to the *Human-Cortex* dataset described in Table 1, using the smooth basis in (Pillow et al., 2008). The largest model (panel A) is unstable, in the sense that spike trains generated from that model could have unrealistically large number of spikes; the other two models are stable.

Just as in Gerhard et al. (2017), we can validate our FNF model stability diagnostic against spike train data. For example, Figure 5 shows the stability map for the FNF$_S$ family of models with firing rate $\log \lambda(t|H_t) = \beta_0 + \sum_{j=1}^{5} h(t - t_{j*})$, where $\beta_0 = -4$, $h(t) = \beta_1 \mathrm{B}_1(t) + \beta_2 \mathrm{B}_2(t)$, and $\mathrm{B}_1(t) = e^{-t/0.02}$ and $\mathrm{B}_2(t) = e^{-t/0.1}$ are the basis functions shown in the inset of Figure 2A. For each value of (β_1, β_2), we (i) simulated a 10 sec. long spike train from the model and deemed the model divergent if it generated over 900 spikes in the last second, (ii) produced the diagnostic curve and determined from it if the model was stable, fragile, or divergent, and (iii) plotted (β_1, β_2) against the outcomes in (i) and (ii): the two match, which suggests that our diagnostic is reliable. The many FNF$_M$ models we investigated also suggest that the diagnostic is reliable; we did not provide an example diagnostic map here because all these models were stable across the entire parameter space.

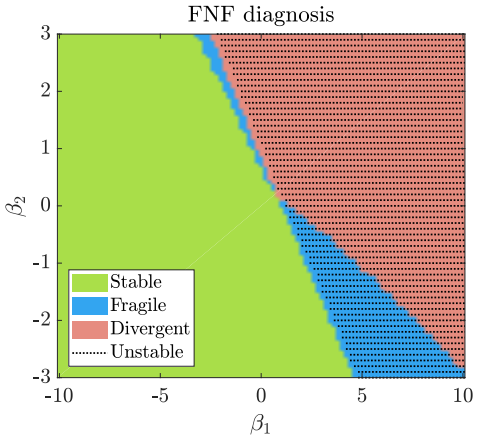

Fig. 6: Stability map for FNF$_S$ models with firing rates $\log \lambda(t|H_t) = \beta_0 + \sum_{j=1}^{5} h(t - t_{j*})$, where $\beta_0 = -4$, $h(t) = \beta_1 \mathrm{B}_1(t) + \beta_2 \mathrm{B}_2(t)$, $\mathrm{B}_1(t) = e^{-t/0.02}$ and $\mathrm{B}_2(t) = e^{-t/0.1}$. For each value of (β_1, β_2), we (i) simulated a 10 sec. long spike train from the model and deemed the model unstable if it generated over 900 spikes in the last second, (ii) produced the diagnostic curve and determined from it if the model was stable, fragile, or divergent, and (iii) plotted (β_1, β_2) against the outcomes in (i), with unstable simulations indicated by black dots, and (ii), in colors. Our diagnostic is reliable because it matches the simulation well.

Figures 5 and 7C,F show that all FNF$_S$ models with $k \leq 4$ and all FNF$_M$ models fitted to the *Human-Cortex* dataset are stable. They also all pass the two KS tests so

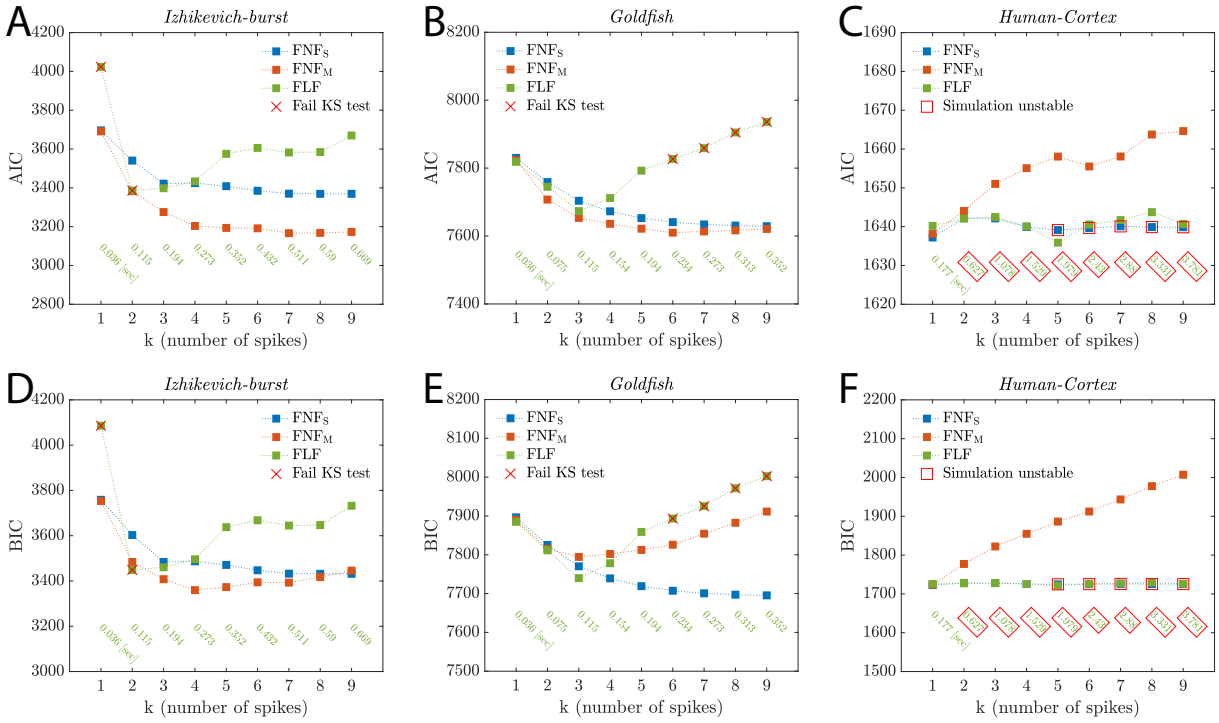

Fig. 7: (A,B,C) AIC and (D,E,F) BIC values for several FLF, FNF$_S$, and FNF$_M$ models fitted to three datasets. (FNF$_S$ and FLF values are equal in panel F, and the latter mask the former.) Simulation unstable models and models that failed the original and/or the discrete KS test are indicated by red squares and crosses. Many models are simulation stable; the FNF$_M$ models fitted here are all simulation stable. Stable models that achieve a desirable criterion, e.g. low AIC, could be chosen.

they are not obviously deficient. With many simulation stable models available, it may be desirable to choose one that also fits the data best according to some criterion. Figures 7C,F show the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for all these models. Small AIC and BIC values are desirable because AIC is an estimate of prediction risk, and BIC is inversely related to the posterior probability of fitting the correct model, in an asymptotic sense; BIC tends to prefer models with fewer parameters Kass et al. (2014). Both criteria suggest that the FNF$_S$ model with $k = 1$ fits best; it is also simulation stable. Figure 7 shows results from two additional examples.

Among the stable models, the FNF$_M$ model with $k = 6$ filters fits best based on AIC, and FNF$_S$ model with $k = 9$ fits best based on BIC. The synthetic *Izhikevich-burst* dataset is bursty as well; see Figure 10A. Its best simulation stable models are FNF$_M$ models with $k = 7$ and $k = 4$ according to AIC and BIC, respectively.

To summarize, our general strategy is to fit FNF models for several values of k and choose a model that is simulation stable and also fits the data well according to criteria such as the KS tests, and AIC or BIC. (Note that because AIC and BIC are obtained from a finite data sample, they have variability, so that similar values should be considered equal.) If no stable model can be found, a model that provides a good fit may be used after constraining its parameters to lie in the parameter subspace corresponding to stability, as Gerhard et al. (2017) suggest.

5 Comparison of FNF and FLF models

A key feature of FLF models is that they sum the effect of all spikes in the filter window of length T_h, which puts no limitation on the number of past spikes influencing the firing rate at t, even if T_h is short. Therefore, if a fitted intensity function has a rising trend, an increasing number of spikes could fall within the filter window, and this could increase the firing rate, eventually yielding unstable simulated spike trains. In contrast, FNF models (Equation (3) and also the extension to time-varying baseline rates) model history with a fixed number of spikes, k, and if the baseline rate $\beta(t)$ and all of the individual filters are bounded above, the firing rate will be bounded. Furthermore, by allowing multiple filters, FNF$_M$ models can diminish the effects of multiple spikes that occur, somewhat infrequently, in close temporal proximity. Thus, in principle, FNF$_M$ models tend to be stable, and we did not find any cases in which FNF$_M$ models were unstable. See, for exam-

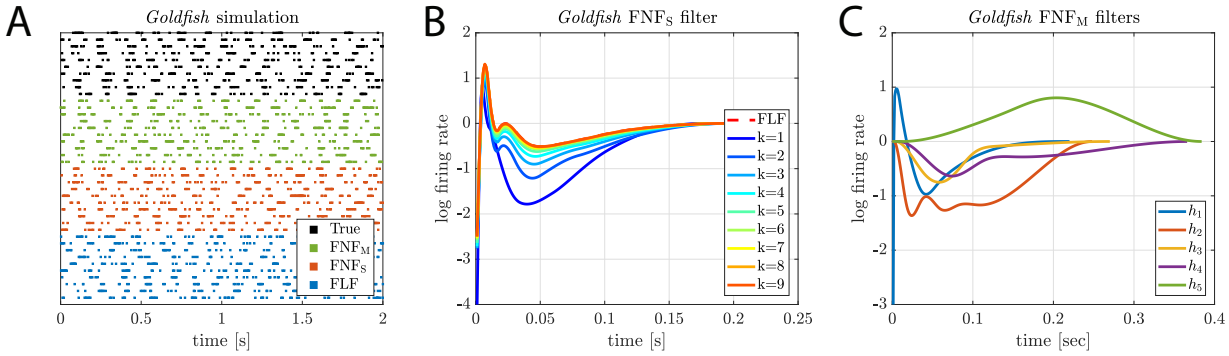

Fig. 8: (A) *Goldfish* dataset spike trains and simulated spike trains from FLF, FNF$_S$ ($k = 5$), and FNF$_M$ ($k = 5$) models fitted to the dataset: despite modeling the effects of past spiking differently, both type of models can generate busts similar to those in the dataset. (B) Filters of the FNF$_S$ models fitted to the *Goldfish* dataset with past number of spikes $k = 1, ..., 9$. When $k \geq 5$, the filters are almost identical, suggesting that effects of spikes prior to 5 spikes back do not contribute much to the fits. The fitted FLF filter has length $T_h = 0.19$ sec., which contains 4 past spikes on average; it overlaps with the FNF$_S$ filters with $k \geq 5$, suggesting that these models are functionally similar (although FNF$_S$ models are more likely to be stable). (C) Fitted filters of an FNF$_M$ model with $k = 5$: the filters are substantially different, suggesting that burst behavior is captured by differentially weighing the contribution of previous spikes according to their timing and ordering. A likelihood ratio test comparing the FNF$_S$ and FNF$_M$ models with $k = 5$ strongly favors the latter ($p \ll 0.001$).

ple, the results in Figure 7. However, we have also seen that, in some cases, the fitted FNF$_S$ firing rates can be large enough to become unstable (see Figures 5A and panels C and F of Figure 7), and for that reason we developed a stability diagnostic and a strategy to stabilize a divergent FNF model in Section 4. We could apply a similar strategy to FLF models, using several filter window lengths T_h in place of several values of k. Gerhard et al. (2017) fitted an FLF model with $T_h = 0.35$ sec to the *Human-Cortex* dataset, which was unstable. Figure 7C,F shows the stability status, AIC, and BIC values of fitted FLF models for several values of T_h. (FLN and FNF$_S$ models have similar AIC and BIC values that are hard to distinguish from one another on the plots.) The FLF model with a very short filter length of $T_h = 0.177$ fits the data well (it passes both KS tests and has smallest AIC and BIC) and is stable. Thus, while FNF$_M$ models seem to be inherently less likely to be unstable, we can not make any universal comparative statement about stability, and, importantly, fitting with either type of model requires care. A remaining issue is whether there are interesting cases in which the additional flexibility of FNF$_M$ models is useful. We now present a few additional comparative results.

Figure 7 provides a summary of fits for the *Human-Cortex* dataset. We see that FNF$_S$ models have AIC and BIC values similar to, or smaller than FLF models. The FNF$_M$ models have higher AIC, presumably because the additional flexibility of using several filters is not needed to fit the data well yet it increases complexity. On the other hand, the FNF models fitted to the *Izhikevich-burst* data set have smaller AIC and BIC values than the FLF models, and FNF$_M$ models have smaller AIC and nearly all smaller BIC values than FNF$_S$ models, presumably because they are bursty and thus are not fitted adequately with simpler models. We also used a real data set, labeled *Goldfish*, which consists of recordings from retinal ganglion cells *in vitro* that exhibit bursting firing (Levine, 1991; Tokdar et al., 2010); see Table 1 and Figure 8A. The AIC and BIC values are again smaller for most FNF models.

These comparisons are substantiated in Figures 8 and 10. Figure 8A displays the *Goldfish* data spike trains together with simulated spike trains from a FLF model fitted to the data, having filter length $T_h = 0.19$ seconds containing five past spikes on average, as well as from fitted FNF$_S$ and FNF$_M$ models with $k = 5$: despite modeling the effects of past spiking differently, both type of models can generate busts similar to those observed in the data. Figure 8B displays the filters of FNF$_S$ models fitted to the *Goldfish* dataset with past number of spikes $k = 1$ to 9. When $k \geq 5$, the filters are almost identical, suggesting that effects of spikes prior to 5 spikes back do not contribute much to the fits. The overlayed fitted FLF filter with $T_h = 0.19$ sec. overlaps with the FNF$_S$ filters with $k \geq 5$, suggesting that these models are functionally similar (although, based on our previous analysis, FNF$_S$ models are more likely to be stable).

Models with a single filter, as in Figure 8B, assume that the effect of any past spike t_{j*} on the firing rate at time t depends only on the elapsed time $t - t_{j*}$ without considering the number of spikes that may have occurred between t_{j*} and t. For the *Goldfish* data, this is a questionable assumption: a likelihood ratio test (LRT) comparing FNF$_S$ and FNF$_M$ models with $k = 5$

strongly favors the latter ($p \ll 0.001$). Furthermore, the fitted filters of the FNF_M model, shown in Figure 8C, are substantially different, suggesting that burst behavior is captured better by differentially weighing the contribution of previous spikes according to their timing and ordering. We may interpret these multiple distinct filters by observing several characteristics of the data (see Tokdar et al. (2010)): the average burst length is roughly 25ms, a burst ISI is around 7ms, the median number of spikes in a burst is 4, and bursts occur, on average, roughly every 200 ms. With these in mind, the narrowness and height of the first filter suggests that the effect of the first spike back is strongly influenced by bursting, i.e., when a spike occurs less than 25 ms in the past it is likely that the cell is in a bursting state and the probability of spiking is increased; filters 2 to 4, corresponding to the 2nd to 4th spikes back, diminish the firing rate starting around 25 ms in the past, which presumably signals that when these multiple spikes back are spaced further than 25 ms in the past, the neuron has transitioned to a "down" state; the effect of the 5th spike back is to increase the firing rate after a longer duration, peaking around 200 ms in the past, reflecting an expectation that the neuron has already finished its pause after a burst and has now returned to the bursting state. Figure 10 in the appendix contains the corresponding plots for the synthetic *Izhikevich-inhomo* dataset, from which similar conclusions can be drawn.

In summary, FNF_M models do, sometimes, provide better fits than FLF or FNF_S models, but this is an empirical question that must be answered for each set of data separately. We should also note that models that incorporate hidden burst and non-burst states, as in Tokdar et al. (2010), may provide even better descriptions of bursting spike train data.

6 Discussion

We have attempted to provide a thorough analysis of the instability phenomenon identified by Gerhard et al. (2017). We improved the diagnostic of Gerhard et al. (2017) and extended it to FNF models; we noted that time-varying baseline rates and excess trial-to-trial variability can cause instability of models that do not account for these effects; we introduced a method to detect outlier trials and illustrated its use; and we compared FNF with FLF models in several examples.

It is perhaps worth emphasizing that FNF_M models, with sufficiently large k, were always stable in the examples we investigated, regardless of whether there were stable FLF or FNF_S models. See, for example, Figure 7. That figure, together with Figures 8 and 10, also illustrate the way differing variations in spiking behavior may suggest different numbers of filters to use in an FNF model, according to standard model-fitting procedures.

Overall, we concluded that FNF_M models tend to avoid instability, and can provide helpful flexibility in some cases, but we cautioned that selection among the different FLF and FNF models must be done carefully based on the unique characteristics of particular data sets.

Our code is available on `https://github.com/AlbertYuChen/Divergent_Spiketrain_public.git`.

Appendix A Datasets

Data	*Monkey*	*Human*	*Goldfish*
Feature	inhomogeneous	trial-to-trail variability	bursty
# Trials	10	10	1
Trial duration (sec)	1	10	30
Mean firing rate (Hz)	24.0	1.0	32.4
Source	Gerhard et al.	Gerhard et al.	Tokdar et al.

Table 1: Details of the datasets used in this paper. The *Monkey* and *Human* datasets (Gerhard et al., 2017) consist of single unit recordings from monkey cortex PMv and M1 areas, and from the neocortex of a person with a pharmacologically intractable focal epilepsy, respectively. The *Goldfish* dataset (Levine, 1991; Tokdar et al., 2010) consists of recordings from retinal ganglion cells in vitro that exhibit bursting firing.

Appendix B Diagnostic Maps

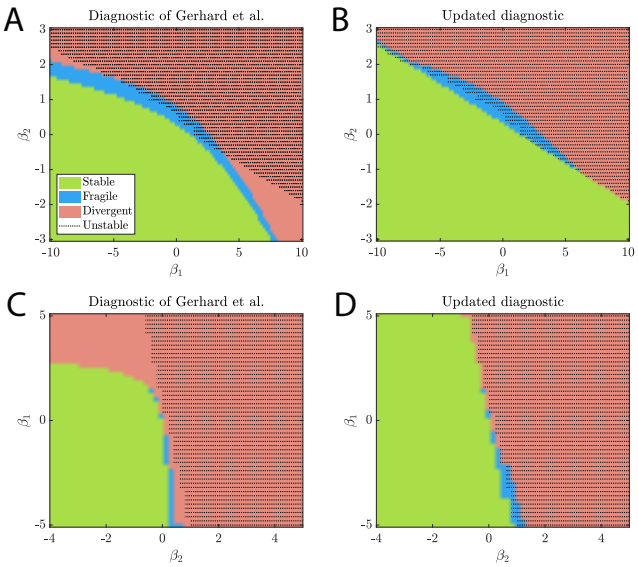

Fig. 9: Stability maps for two FLF models (Equation 4) F_θ using the diagnostic of Gerhard et al. (2017) and our updated diagnostic. For each value of θ, we (i) simulated a 10 sec. long spike train from F_θ, and deemed the model unstable if it generated over 900 spikes in the last second, (ii) produced the diagnostic curve and determined from it if the model was stable/fragile/divergent, and (iii) plotted θ against the outcomes in (i) and (ii). (A) Reproduction of the stability map in Gerhard et al. (2017) Figure 4, where F_θ is an FLF model with $\beta_0 = -5.3$ and $h(t) = \beta_1 \cdot \mathrm{B}_1(t) + \beta_2 \cdot \mathrm{B}_2(t) + \mathrm{Dip(t)}$, where $\mathrm{B}_1(t) = e^{-t/0.02}$, $\mathrm{B}_2(t) = e^{-t/0.1}$ and Dip(t) is a negative window function modeling a 2 msec. refractory period, $\theta = (\beta_1, \beta_2)$, and filter length $T_h = 0.2$ sec. The maps suggest that the diagnostic is mostly reliable, except in small regions of the parameter spaces. (B) Our updated diagnostic for the same model matches the simulation better. (C) Stability map using the diagnostic of Gerhard et al. (2017) for F_θ an FLF model with with $\beta_0 = -4$, $h(t) = \beta_1 \cdot \mathrm{B}_1(t) + \beta_2 \cdot \mathrm{B}_2(t)$. Basis $\mathrm{B}_1(t)$ and $\mathrm{B}_2(t)$ are the same as Figure 2A. $\theta = (\beta_1, \beta_2)$, and filter length $T_h = 0.35$ sec. (D) Our updated diagnostic for the same model matches the simulation better.

Appendix C Simulation algorithms

Dataset in form *Izhikevich-xx* are synthesized by algorithm 1. FLF and FNF model simulation are algorithm 2 and 3.

Algorithm 1: Izhikevich simulation algorithm. he Izhikevich dynamical model can generate a rich family of biophysically realistic spike patterns (Izhikevich, 2003, 2004), including time varying rate pike trains, tonic spikes, bursts and etc. A list of parameters to produce various effects is given in Weber and Pillow (2017).

Input: Parameters a, b, c, d. Time resolution Δ. $I(t)$.
Initial: $u(0) = 0$, $v(0) = 0$, $\mathbf{S} = \{0\}$
for $t = 0$ *to* T **do**
  $\mathrm{d}v = 0.04v(t)^2 + 5v(t) + 140 - u(t) + I(t)$
  $\mathrm{d}u = a(bv(t) - u(t))$
  $v(t+1) = v(t) + \mathrm{d}v \cdot \Delta$
  $u(t+1) = u(t) + \mathrm{d}u \cdot \Delta$
  if $v(t+1) > 30$ **then**
    $v(t+1) = c$
    $u(t+1) = u(t) + d$
    $\mathbf{S} = \mathbf{S} \cup \{t+1\}$
  else
    continue
  end
end
Output: $\mathbf{S}$

Algorithm 2: FLF model (Equation 2) simulation algorithm. This algorithm simulates ISIs from a unit rate exponential distribution and inverts them using the time rescaling theorem (Brown et al., 2002; Kass et al., 2014) to obtain the past spike times.

Input: time resolution Δ, baseline $\beta(t), t \in [0, T]$, and post-spike filter h with length L;
define $f(s_1, s_2) := \sum_{t=s_1}^{s_2} \lambda(t|H_t)\Delta$, $0 \le s_1 < s_2 \le T$, assume $\lambda(t|H_t) \ge 0$;
Initial: $\mathbf{S} = \emptyset$ be the set of spike time points;
$\lambda(t|H_t) = \beta(t)$
while *TURE* **do**
  draw one sample $Z \sim \mathrm{Exp}(1)$;
  if $f(t_{1*}, T) < Z$ **then**
    return
  else
    $s = \arg\min_\tau \{f(t_{1*}, \tau) \ge Z\}$
    $\mathbf{S} = \mathbf{S} \cup \{s\}$
    Update the firing rate function by adding the impact of the spike to the future firing rate: $\log \lambda(\tau + t|H_t) = \log \lambda(\tau + t|H_t) + h(\tau)$, for all $\tau \in [t, \min(t + L, T)]$
  end
end
Output: $\mathbf{S}$, $\lambda(t|H_t)$

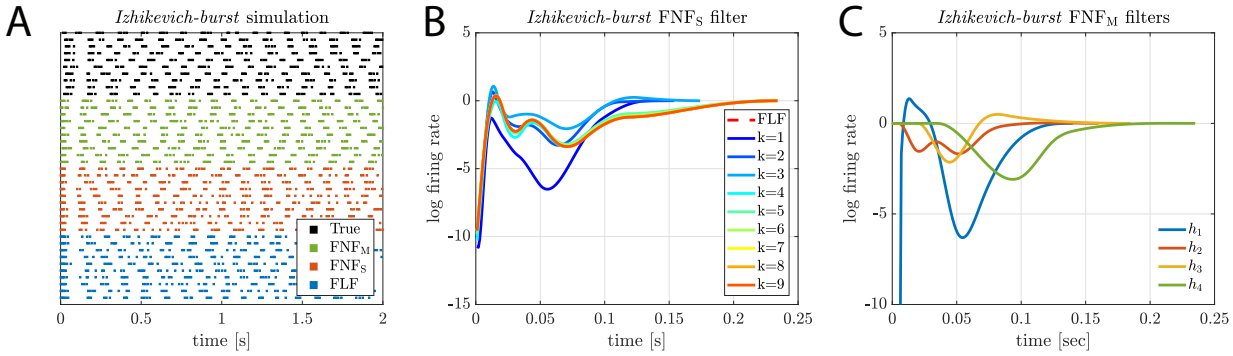

Fig. 10: (A) *Izhikevich-burst* synthetic dataset spike trains and simulated spike trains from FLF, FNF_S(k=4), FNF_M(k=4). Both type of models can generate busts similar to those in the dataset. (B) Fitted filters of the FNF_S models with number of spikes $k = 1, ..., 9$. When $k > 3$, the filters are very close to each other since further spikes will not make too much contribution to the future firing rate, thus will only affect the filter shape slightly. The fitted FLF filter overlaps with the FNF_S filters with $k \geq 5$, suggesting that these models are functionally similar. (C) Fitted filters of an FNF_M model with $k = 4$: the filters are substantially different, which suggests that past spikes of different order have different effects on the firing rate. A likelihood ratio test comparing the FNF_S(k=5) and FNF_M(k=5) models favors the FNF_M model ($p \ll 0.001$).

Algorithm 3: FNF model (Equation 3) simulation algorithm.

Input: time resolution Δ, baseline $\beta(t), t \in [0, T]$, and k post-spike filter h_i with length L_i
define $f(s_1, s_2) := \sum_{t=s_1}^{s_2} \lambda(t|H_t)\Delta$, $0 \leq s_1 < s_2 \leq T$, $\lambda(t|H_t) \geq 0$ is the total firing rate
Initial: $\mathbf{S} = \emptyset$ be the set of spike time points
$\lambda(t|H_t) = \beta(t)$
while *TURE* **do**
  draw one sample $Z \sim \text{Exp}(1)$
  Update the firing rate function $\log \lambda(t|H_t) = \beta(t) + \sum_{i=1}^{\min(k,|\mathbf{S}|)} h_i(t - t_{i*})$, if $t > L_i$, $h_i(t) = 0$. t_{i*} are the last i'th spike. **if** $f(t_{1*}, T) < Z$ **then**
    return
  else
    $s = \arg\min_{\tau}\{f(t_{1*}, \tau) \geq Z\}$
    $\mathbf{S} = \mathbf{S} \cup \{s\}$
  end
end
Output: $\mathbf{S}$, $\lambda(t|H_t)$

Appendix D Misc. results

Derivation of Equation 7

$$\sum_{t_{j*} \in (t-T_h, t_{1*})} h(t - t_{j*}) \approx \mathop{\mathbb{E}}_{N} \left[\int_{t-T_h}^{t_{1*}} h(t-\tau) \mathrm{d}N_{(t-T_h,\tau]} \right] \tag{10}$$

$$= \mathop{\mathbb{E}}_{N_\Delta} \left[\mathop{\mathbb{E}}_{N|N_\Delta} \left[\int_{t-T_h}^{t_{1*}} h(t-\tau) \mathrm{d}N_{(t-T_h,\tau]} \middle| N_\Delta \right] \right] \tag{11}$$

$$\overset{t-\tau=u}{=} \mathop{\mathbb{E}}_{N_\Delta} \left[\frac{N_\Delta}{t_{1*} - t + T_h} \int_{t-t_{1*}}^{T_h} h(u) \mathrm{d}u \right] \tag{12}$$

$$= A_0 \int_{t-t_{1*}}^{T_h} h(u) \mathrm{d}u \tag{13}$$

where $N_\Delta = N_{(t-T_h,t_{1*})}$ is the number of spikes in $(t - T_h, t_{1*})$, and A_0 is the mean firing rate in that time window. In equation 11, the inner expectation is taken over spike count conditioned on a fixed number of spikes in the interval $(t - T_h, t_{1*})$. If the filter $h(u)$ is estimated by $e^{h(u)} - 1$ in (Gerhard et al., 2017), the error will be larger if $h(u)$ is not close to 0. Because the point process itself is unknown, the firing rate function is approximated under the assumption that it is a homogeneous Poisson process. For homogeneous Poisson process, if the number of events is fixed, they distribute evenly in the interval, which leads to equation 12.

The time rescaling theorem Let $Z_i = \int_{t_{i-1}}^{t_i} \lambda_0(t)dt$, where t_i are spike times, Z_i are time integral transformed intervals. Time rescaling theorem states that if $\lambda_0(t)$ is the firing rate of the true model, then Z_i are iid and $Z_i \sim \text{Exp}(1)$. The goodness-of-fit test checks how close the distribution of transformed intervals from estimated model is to the unit exponential distribution.

References

Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM (2002) The time-rescaling theorem and its application to neural spike train data analysis. Neural computation 14(2):325–346

Chen S, Shojaie A, Shea-Brown E, Witten D (2017) The multivariate hawkes process in high dimensions: Beyond mutual excitation. arXiv preprint arXiv:170704928

Eichler M, Dahlhaus R, Dueck J (2017) Graphical modeling for multivariate hawkes processes with nonparametric link functions. Journal of Time Series Analysis 38(2):225–242

Gerhard F, Deger M, Truccolo W (2017) On the stability and dynamics of stochastic spiking neuron models: Nonlinear hawkes process and point process glms. PLoS computational biology 13(2):e1005,390

Haslinger R, Pipa G, Brown E (2010) Discrete time rescaling theorem: determining goodness of fit for discrete time statistical models of neural spiking. Neural computation 22(10):2477–2506

Izhikevich EM (2003) Simple model of spiking neurons. IEEE Transactions on neural networks 14(6):1569–1572

Izhikevich EM (2004) Which model to use for cortical spiking neurons? IEEE transactions on neural networks 15(5):1063–1070

Kass RE, Ventura V (2001) A spike-train probability model. Neural computation 13(8):1713–1720

Kass RE, Eden UT, Brown EN (2014) Analysis of neural data, vol 491. Springer

Levine MW (1991) The distribution of the intervals between neural impulses in the maintained discharges of retinal ganglion cells. Biological cybernetics 65(6):459–467

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky E, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454(7207):995–999

Tokdar S, Xi P, Kelly RC, Kass RE (2010) Detection of bursts in extracellular spike trains using hidden semi-markov point process models. Journal of computational neuroscience 29(1-2):203–212

Ventura V, Cai C, Kass RE (2005) Trial-to-trial variability and its effect on time-varying dependency between two neurons. Journal of neurophysiology 94(4):2928–2939

Weber AI, Pillow JW (2017) Capturing the dynamical repertoire of single neurons with generalized linear models. Neural Computation

Wu W, Srivastava A (2011) An information-geometric framework for statistical inferences in the neural spike train space. Journal of Computational Neuroscience 31(3):725–748