Tracking fast and slow changes in synaptic weights from simultaneously observed pre- and postsynaptic spiking

Ganchao Wei1\* and Ian H. Stevenson2,3

1Department of Statistics, University of Connecticut

2Department of Psychological Sciences, University of Connecticut

3Department of Biomedical Engineering, University of Connecticut

\*Corresponding Author: ganchao.wei@uconn.edu

# Abstract

Synapses change on multiple timescales, ranging from milliseconds to minutes, due to a combination of both short- and long-term plasticity. Here we develop an extension of the common Generalized Linear Model to infer both short- and long-term changes in the coupling between a pre- and post-synaptic neuron based on observed spiking activity. We model short-term synaptic plasticity using additive effects that depend on the presynaptic spike timing, and we model long-term changes in both synaptic weight and baseline firing rate using point process adaptive smoothing. Using simulations, we first show that this model can accurately recover time-varying synaptic weights 1) for both depressing and facilitating synapses, 2) with a variety of long-term changes (including realistic changes, such as due to STDP), 3) with a range of pre and post-synaptic firing rates, and 4) for both excitatory and inhibitory synapses. We also show the importance of tracking fast changes in synaptic weights, slow changes in synaptic weights, and unexplained variations in baseline firing simultaneously. Omitting any one of these factors can lead to spurious inferences for the others. Altogether, this model provides a flexible framework for tracking short- and long-term variation in spike transmission.

# 1 Introduction

Synapses play an important role in neurophysiology, although detecting and characterizing these synapses is not easy. The extracellular spiking activities of neurons provide a way for synapses detection, and it is usually done by analyzing the cross-correlogram1,2. The cross-correlogram is a histogram of postsynaptic spiking times forward and backward to presynaptic spikes, which provides an estimate of cross-correlation between pre- and post-synaptic neurons3. When a monosynaptic occurs, the postsynaptic spiking probability will often increase or decrease after a presynaptic spike, for an excitatory or inhibitory synapse. This corresponds to a transient and short-latency peak or trough in the cross-correlogram. However, investigating the cross-correlogram directly may not be enough for modern neural data analysis. By using multielectrode arrays, the extracellular spiking of hundreds of neurons can be recorded simultaneously. These large-scale recorded neurons can have tens of thousands of potential synapses between them, but both spiking and synapses can be sparse. This makes it difficult for synapses detection by simply plotting the cross-correlograms. Some recent research tried to solve this problem by modeling spike cross-correlations to handle the sparse information for detection4,5.

In cases where synapses can be reliably identified from spikes, one possibility is that these recordings can be used to examine changes in synaptic strength over time. Changes in synaptic strength occur over multiple timescales and due to different biophysical mechanisms6. For instance, on timescales of a few milliseconds, the synaptic strength may decrease by consumption of neurotransmitters, or may increase by influx of calcium into presynaptic axon terminal, which increases neurotransmitters release probability. In contrast, increases or decreases in receptor density or structural changes occur on timescales of minutes to hours. To understand learning rules and to make sure our models generalize it is important to separate the short- and long-timescale effects. It is also important to note that fluctuations in the presynaptic rate can create fluctuations in synaptic strength on much longer timescales. For instance, higher presynaptic rate will produce a weaker synaptic strength, if the synapse has short-term depression7.

The short- and long-term dynamics in synapses can be roughly studied by investigating the cross-correlations or postsynaptic spiking probability, under different presynaptic inter-spike interval (ISI)7,8 and different recording time respectively. Several authors have proposed models for estimating either long-term9,10 or short-term11–13 changes in synaptic weights. Unlike English et al.12, the short-term plasticity estimation is conditioned on the previous presynaptic inter-spike intervals (ISIs) in Ghanbari et al.14, and this provides only an incomplete picture for synaptic strength inference. Here we introduce a statistical model that estimates both long- and short-term changes simultaneously. The unexplained changes in postsynaptic neuron are treated as the baseline firing rate. The long-term changes in both synaptic weight and baseline is estimated by the point process adaptive smoothing15, while the short-term plasticity is modeled by additive effects that depend on the presynaptic spike timing10. By several simulations, we show that this model provides a flexible framework for tracking short- and long-term variation in spike transmission, and it’s necessary to estimate both long-term and short-term effects simultaneously.

# 2 Methods

Here we introduce an extension of a Poisson GLM that aims to describe the coupling between a pre- and postsynaptic neuron16,17. While many previous studies9–13 have modeled static coupling between neurons, our goal here is to describe a time-varying synaptic strength with both short- and long-term changes.

We model the postsynaptic spiking in discrete time as a doubly stochastic Poisson process with time-varying parameters. Partitioning the total recording time T into evenly-spaced bins , such that with time steps , we denote the total number of presynaptic spikes in as and represents the number of spikes observed in . and denote the same calculations for the postsynaptic neuron’s spikes. Make small enough such that both and can only take 0 or 1, and hence we can view and as spike indicators at time .

Previous models of static coupling between neurons are modeled by including both past spiking history of both pre- and post-synaptic neuron directly.

|  |  |  |
| --- | --- | --- |
|  |  | (2.1) |
|  |  | (2.2) |
|  |  | (2.3) |
|  |  | (2.4) |

where is the conditional intensity of postsynaptic neuron at , given spike history of postsynaptic neuron and presynaptic neuron , and model parameters. defines the fixed baseline firing rate and the history effects are weighted by and **,** for post- and pre-synaptic neurons respectively.

Here we model the postsynaptic neuron’s firing rate as

|  |  |  |
| --- | --- | --- |
|  |  | (2.5) |
|  |  | (2.6) |

In our model, we don’t include the spike history of postsynaptic neuron explicitly and the effect is captured by the unstructured time-varying baseline to avoid biased inference18. In the static coupling model, the main goal is to accurately infer the shapes of the filters, here we assume that the shape is fixed , but is weighted by an additional factor that varies over time. Additionally, we aim to partition the variations in into a long-term component and a short-term component that scales the long-term synaptic weight multiplicatively. Here we model transient increases/decreases in as a function of the presynaptic spike timing to account for short-term synaptic facilitation/depression. In the absence of presynaptic activity, returns to a base value of . To simplify notations, we further write , , and as , , and . These two sets of notations will both be used in the remaining of this paper for convenience.

## Estimating the Synaptic Filter

We estimate the shape of the static presynaptic filter by modeling the cross-correlogram between the pre- and postsynaptic spikes directly, as in Ren et al4. The shape of synaptic connection is described by an alpha function , with a latency and time constant . The parameters are estimated by modeling the cross-correlogram, with consideration of both the slow effects of background fluctuations, and the fast, transient effect of the synaptic connection. Briefly, the model is described as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.7) |

where denotes the expected rate of coincidences from the cross-correlogram with baseline and a linear combination of smooth basis functions accounting for slow changes in cross-correlation. The fast effect is described as , where is the connection strength from pre-synaptic neuron to post-synaptic neuron. The fast effect is estimated by maximizing a penalized Poisson log-likelihood. Since the objective function is non-convex, the function is optimized by several random restarts.

We assume that the latency and time constant of the synaptic filter are fixed, and not affected by the long- or short-term changes in synaptic strength. This simplifying assumption allows us to model rescaling of the basic presynaptic input, given by .

## Estimation of baseline firing rate and synaptic weight

The baseline firing rate and synaptic plasticity (, and ) are changing in two different time- scales: long-term changes ( and ) and short-term changes ().

Different methods are used to estimate changes in these two timescales. The long-term effects are estimated by point process adaptive smoothing15,19, while the short-term effect is estimated by additive effects depending on presynaptic inter-spike intervals (ISIs)10. The estimation of STP gives a generalized bilinear model (GBLM). These two sets of estimations are combined by alternating optimization, i.e. holding one set of parameters fixed while updating the other and alternating between the two optimizations.

### Long-Term Effects Estimation: Point Process Adaptive Smoothing

Denote the joint parameters for long-term effects as and covariates as . Therefore, the conditional intensity function can be re-written as: . The mean and covariance for are estimated by adaptive smoothing, which has two steps: 1) forward algorithm: adaptive filtering15, and 2) backward algorithm: Rauch-Tung-Striebel (RTS) smoothing19. The model for parameters evolution is as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (2.8) |

where is a system evolution matrix and represents Gaussian noise with covariance at . The evolution equation (x.4) gives model prediction (prior) for parameters, at each time step.

Step One: Forward Algorithm (Adaptive Filtering)

Let and be the model predicted mean and covariance, i.e. prior estimations, of parameters at . Let and be the adaptive filtering estimated mean and covariance after observations, i.e. posterior estimations, at . Since , and . Therefore, the adaptive filtering is:

|  |  |  |
| --- | --- | --- |
|  |  | (2.9) |
|  |  | (2.10) |
|  |  | (2.11) |
|  |  | (2.12) |
|  |  | (2.13) |

Equations (2.9) – (2.11) are prior predictions, while equations (2.12) and (2.13) are posterior corrections.

Step Two: Backward Algorithm (RTS Smoothing)

Let and be RTS smoothing estimates for mean and covariance at . The RTS smoothing is:

|  |  |  |
| --- | --- | --- |
|  |  | (2.14) |
|  |  | (2.15) |
|  |  | (2.16) |

To make the algorithm numerically stable, the estimation of mean can be equivalently written as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.17) |

In this analysis, is set as a time constant. To give unbiased estimations, is set as the identity matrix.

Selection of

The performance of adaptive smoothing is highly affected by , and improper will even make the algorithm diverge. Usually, the is chosen based on previous knowledge. When there’s no sufficient knowledge for , it is usually adjusted by manual trial-and-error approaches. Although we can use the EM algorithm20 to estimate , the convergence is notoriously slow, even with an accelerator21.

Here we choose to estimate by maximizing prediction likelihood, i.e. likelihood under . The is assumed to be diagonal, i.e. noise for () and ()are unconditionally independent. Besides estimating by direct two-dimensional optimization, the optimized can be approximated by sequential one-dimensional optimization. Although and are assumed to be unconditionally independent, they are not independent conditioning on data. Therefore, we need to be careful to the order of one-dimensional approximation. Since both synaptic connection () and STP () are small, and most synaptic connections are 0, the values of is small. Therefore, the values of has negligible influence on estimation. Based on this observation, we can do one-dimensional approximation as follows: 1) fix and get MLE ; 2) fix as and get MLE .

### Estimation of Short-Term Effect: Additive Effects Depending on presynaptic ISIs

Since the short-term effects are changing too fast, it’s impossible to track them by adaptive smoothing. Here, we model the short-term synaptic plasticity by the following linear model10:

|  |  |  |
| --- | --- | --- |
|  |  | (2.18) |
|  |  | (2.19) |

where is the presynaptic firing time and the inter-spike interval (ISI) between and spikes is . is a nonlinear function of presynaptic ISI, which gives amplitude of short-term effect under different ISIs. The STP is modeled by convolution of on presynaptic spikes, with exponentially decaying of rate . The nonlinearity of is captured by linear combinations of raised-cosine basis functions , which are functions of presynaptic ISI.

To simplify notations, define , and hence . The short-term modification function is defined as . The variances for STP and modification function can be calculated as:

|  |  |  |
| --- | --- | --- |
|  |  | (2.20) |
|  |  | (2.21) |

Altogether, we have the following parameters: , , , and , with the hyper-parameters and parameters defining raised-cosine basis . We estimate the synaptic latency and time-constant directly from the cross-correlogram. Then optimize the remaining parameters by alternating between fitting the short-term parameters assuming , fixed, and fitting the long-term parameters , assuming fixed.

## Simulations

Here we validate the model using simulated pre- and postsynaptic spiking. Presynaptic spike times are generated by homogeneous Poisson process with a firing rate of 5Hz, if not specified. The postsynaptic neuron is simulated as a conditionally Poisson process defined in equation 2.5 and 2.6. The observed time length is 20min if not specified, with binsize be 1ms.

For simulating spike-timing-dependent plasticity (STDP), a special and more realistic model for LTP, the modification function is set as a double-exponential function, which can accurately model STDP in cortical and hippocampal slices. In this case, each pair of pre- and post-synaptic spikes modifies the synapse by

|  |  |  |
| --- | --- | --- |
|  |  | (2.22) |

In our case, , , and . We further add an additional long-term decaying factor that pushes the synaptic weights back to 1, as in Stevenson et al10. Namely,

|  |  |  |
| --- | --- | --- |
|  |  | (2.23) |

where is the indicator function, and we set . When fitting the model without tuning of , it is set as , if not specified.

# 3 Results

In this study, 1) we implemented commonly used methods (i.e. cross-correlogram and efficacy plots) to analysis the simulated data. 2) Then we implemented our estimation method into several simulations, to see the fitting results under different situation. 3) We further discussed influence of presynaptic firing rates, model misspecification and hyper-parameter selection for long-term effects tracking (i.e. in point process adaptive smoothing, see details in Method).

## Traditional Methods for Synaptic Weight Inference

In this section, the baseline firing rate for postsynaptic neuron is constantly 20Hz. The LTP is step-changing at 10min, while keeping synaptic weight constantly as 3 before 10min and 5 after 10min. The STP is depression. Traditionally, the baseline firing rate and LTP are detected by splitting cross-correlogram by recording time, and the STP is detected by splitting cross-correlogram according to presynaptic inter-spike intervals (ISIs). Plotting post-synaptic firing efficacy against presynaptic ISIs can also show STP. Figure 1 shows the analysis results by implementing these traditional methods. These methods are easy to implement and can give powerful insight about synaptic weights. However, through traditional we cannot give accurate and detailed inferences on the synaptic weights. Also, the interval estimation is not possible for traditional methods.

|  |
| --- |
|  |
| **Fig 1. Traditional methods for synaptic weights analysis. (A)** In this example the simulated LTP is step-changing at middle recording time (T/2). **(B)** The change of LTP can be visualized by splitting the cross-correlogram, before and after T/2. The cross-correlation between pre- and postsynaptic spiking is higher after T/2, which suggests a stronger coupling. **(C)** The simulated STP can be viewed by plotting the postsynaptic neuron efficacy against presynaptic inter-spike interval (ISI). Shorter presynaptic ISIs correspond to lower postsynaptic efficacies, which shows that the STP is depression here. The efficacy after T/2 consistently dominates the efficacy before T/2, and this shows abruptly increase of LTP after T/2. **(D)** By splitting cross-correlogram according to T and presynaptic ISI, we can show LTP and STP simultaneously. The cross-correlation is higher for later recording period (>T/2) and longer presynaptic ISIs (> median of simulated ISI). The baselines are the same across all split plots, which suggest a constant postsynaptic firing rate. |

## Simulations

In this section, we use several simulations to see the inference results under different situations, i.e. synapses with 1) different LTP, 2) different STP, 3) different postsynaptic baseline firing rates and 4) different synaptic types.

Inference for long-term and short-term plasticity

First, we use two simulated examples to show inference of linear- and sinusoidal-changing LTP. The baseline firing rate for postsynaptic neuron is constantly 20Hz. The synaptic weight for linear-changing LTP changes linearly from 1.5 to 3.5, and the weight for sinusoidal-changing LTP oscillates around 2.5 with period of 10min. The STP is depression. The LTPs are further shown by splitting cross-correlogram for quartiles of recording time. The left two panels in Figure 2 (A and B) show fitting results for these two simulations.

|  |
| --- |
|  |
| **Figure 2. Inference of different long- and short-term plasticity.** These plots show fitting results for different LTP (orange) and STP (blue). The STP is shown by modification function. The dashed lines show standard error from point estimations. **(A)** The fitting results under different types of LTP, i.e. linear- and sinusoidal-changing. The postsynaptic baseline firing rates are constant and STPs are depression. **(B)** These two LTPs and fitted values can also be visualized by splitting cross-correlogram for quartiles of recording time. **(C)** The fitting results under different types of STP, i.e. depression, facilitation and no STP. The postsynaptic baseline firing rates are constant and LTPs are step-changing at middle recording time. **(D)** Split cross-correlograms for quartiles of presynaptic ISIs show these STPs and fitted values. |

However, the linear- and sinusoidal-changing LTPs are far from real situation. To investigate the fitting results for more realistic LTPs, we further simulate an LTP as the spike-timing dependent plasticity (STDP) and implement our methods for estimation. The baseline firing rate for postsynaptic neuron is constantly 15Hz. The modification function for LTP is a traditional double-exponential function that can accurately model STDP in cortical and hippocampal slices. The STP is depression. Similarly, the LTP is shown by the split cross-correlogram by recording length. Figure 3 shows the fitting results.

|  |
| --- |
|  |
| **Figure 3. Inference under STDP generated LTP.** These plots show one example with LTP generated by spike-timing dependent plasticity (STDP) model. The modification function for LTP in STDP model is a double-exponential function. The baseline is constant and STP is depression, and STP is shown by modification function. **(A), (B) and (C)** show fitted results for baseline, LTP and STP. **(D)** The STP can be visualized by splitting cross-correlogram for quartiles of presynaptic ISIs. |

Then we see the fitting performance under different STPs, i.e. depression, facilitation and no STP cases. Again, the baseline firing rate for postsynaptic neuron is constantly 20Hz. The LTP is step-changing at 10min, and the STPs are set as depression, facilitation and constantly one (no STP) accordingly. The STPs are further shown by splitting cross-correlogram for quartiles of presynaptic ISIs. The right two panels in Figure 2 (C and D) show fitting results for these three simulations.

Inference for postsynaptic baseline firing rates

The long-term dynamics of postsynaptic firing rates can be caused by unexplained factors. To build the unbiased model, we put the unexplained long-term dynamics into baseline line firing rate and make inference on it. Here, we show the fitting results for linear- and sinusoidal-changing postsynaptic baseline firing rates. To see the influence of LTP on baseline estimations, we further simulated step- and linear-changing LTP in each case.

The baseline firing rates for postsynaptic neuron 1) linearly change from 7Hz to 55Hz or 2) oscillate sinusoidally around 20Hz with period of 10min in log-space. The step- and linear-changing LTPs are considered for each case. The STP is depression. The LTPs and baselines are further shown by splitting cross-correlogram for quartiles of recording time. Figure 4 shows fitting results for these four simulations. By these simulations, the baselines are accurately recovered, and dynamics in LTP will not influence baseline estimations. Therefore, we can separate LTP and unexplained long-term dynamics successfully.

|  |
| --- |
|  |
| **Figure 4. Inference of different postsynaptic baseline firing rates.** These plots show fitting results of postsynaptic baseline firing rates (yellow). **(A)** The fitting results under different types of baseline, i.e. linear- and sinusoidal-changing. The LTPs are step-changing at middle recording time and STPs are depression. **(C)** The same simulation settings as (A), except that the LTP changes linearly. **(B) and (D)** Split cross-correlograms for quartiles of recording time show dynamics of baseline and LTP. |

Inference for inhibitory synapse

The previous simulations are all for excitatory synapses. Here, we show one example for inhibitory synapse. The setting is similar to previous examples, i.e. constant baseline, step-changing LTP and depression STP, but the LTP weight changes from -9 to -3 in this case. Figure 5 shows the fitting results. We can see the fitting results are generally good but a bit worse than excitatory synapses, since we will observe less postsynaptic firing in inhibitory synapses.

|  |
| --- |
|  |
| **Figure 5. Inference of inhibitory synapse.** These plots show one example for estimations on inhibitory synapse, and the STP is shown by modification function. In this example, the baseline firing rate is constant, the LTP jumps at mid-recording-time and STP is depression. **(A), (B) and (C)** show fitted results for baseline, LTP and STP. By splitting cross-correlogram for quartiles of recording time **(D)** and presynaptic ISIs **(E)**, we can show corresponding LTP and STP. |

## Influence of Presynaptic Firing Rates

Since the LTP effect shows only when there is a presynaptic spike, the presynaptic firing rates will influence LTP estimation accuracy. The larger the firing rates, the more information for LTP estimation and therefore the estimation of LTP will be more accurate. Although STP effect also depends on presynaptic spikes, the STP depends directly on presynaptic ISIs but not the spiking activities itself. In other words, STP is estimated by combined information at different ISIs and presynaptic firing rates will not influence the estimation a lot. Figure 6 shows one example. In this example, the baseline and LTP are constant, and STP is facilitation to show more significant results. The recording time is divided into three parts evenly, and different presynaptic firing rates are assigned to each part respectively (i.e. 3Hz-15Hz-3Hz and 5Hz-0Hz-5Hz), to show influence of presynaptic firing rates. The results show that the more presynaptic spikes, the more accurate the LTP estimation is. The depression STP has similar pattern, but the results are not as significant as facilitation STP.

|  |
| --- |
|  |
| **Figure 6. Influence of presynaptic firing rates on LTP estimation.** In these two examples, the baseline firing rates and LTP are constant. The STP is facilitation to show more significant results. To show the influence of presynaptic firing rates, the recording time is divided into three parts evenly and different presynaptic firing rates are assigned to each part respectively. **(A)** In this simulation, the presynaptic firing rates are 3Hz-15Hz-3Hz. When the firing rate increase, the estimation of LTP is more accurate. **(B)** Now, the presynaptic firing rates are 5Hz-0Hz-5Hz. Lack of the presynaptic spikes leads to a variated LTP estimation. The variation in presynaptic firing rates will not influence estimation of baseline and STP a lot. |

## Model Misspecification

The model can be mis-specified by omitting any one of these three pieces, i.e. baseline, LTP and STP. Although the mis-specified model is good for prediction generally, it will lead to spurious inferences for the specific synaptic weights. Figure 7 shows one example about influence of omitting baseline or STP on LTP estimation. In this example, there are fluctuations in baseline and STP, but the LTP is constant. When all three effects (baseline, LTP and STP) are estimated simultaneously, the model is fitted well. However, when we miss to estimate baseline, the fluctuations in baseline will flow into and even be enlarged in LTP estimation. Similarly, if we miss to estimate STP, the LTP estimation will capture the fluctuations in STP.

|  |
| --- |
|  |
| **Figure 7. Miss to estimate the baseline or STP causes spurious LTP estimations.** These plots show one example when missing estimations on baseline or STP. The postsynaptic baseline firing rate fluctuated around exp(3) = 20 Hz, and the presynaptic firing rate fluctuated around 8Hz. The STP is depression. **(A)** Fitting results when all three effects (baseline, LTP and STP) are estimated simultaneously. **(B)** When we miss to estimate the baseline, the fluctuation in baseline will flow and enlarge in LTP estimation. **(C)** Similarly, the LTP estimation will capture STP fluctuation when we miss to estimate STP. |

## Selection of Hyper-parameter in Adaptive Smoothing

In long-term effects (baseline and LTP) estimations, the covariance for Gaussian noise defines the time scales for tracking. If is too small, the estimations are oversmoothed; if is too large, the estimations are noisy. Here, we estimated the hyper-parameter by maximizing the prediction likelihood (See details in method). The first two panels in figure 8 (A and B) show one example. Both baseline and LTP have Gaussian noise with variance be . Besides doing full 2D optimization (bounded gradient descent in this example), 1D approximated optimization results are also shown in the plot. The slices of prediction log-likelihood show that for baseline has large influence on LTP likelihood, but baseline likelihoods are nearly the same under different for LTP. This shows that 1D approximation is appropriate. As shown in this example, the 2D optimization and 1D approximation give similar results, and they are all close to true .

The panel at right-bottom (C) show estimation performance under different combinations of true . Overall, the estimation of for baseline is better than that for LTP, since LTP fluctuations are shrink by small in the model. The observed time length is 10min for quick simulations. Generally, the longer the recording time, the more accurate the estimation is.

|  |
| --- |
|  |
| **Figure 8. Selection and influence of in point process adaptive smoothing. (A)** The heatmap shows prediction log-likelihood under different . The red upward-pointing triangle and dashed lines represent true . The orange dot represents the maximum prediction likelihood estimate (MLE) . The blue dot is and the green dot is . The one-dimensional approximated MLE is shown in blue downward-pointing triangle. The slices of prediction log-likelihood for LTP and baseline is shown besides the heatmap. **(B)** Corresponding fitted baselines and LTPs, for , and . When is too small, the estimations are over-smoothed; when is too large, the estimations are too noisy. **(C)** The performance of MLE (2D) under different combinations of true , shown by plotting MLE values against true values. Each combination has 5 replicates. The simulated recording time lengths are 10 min in panel C. |

# 4 Discussion

Here we examined a statistical model that aims to model short- and long-term synaptic changes simultaneously. By simulation examples, we show that our model can successfully recover time-varying synaptic weights for both excitatory and inhibitory synapses with a variety of STPs and LTPs, within reasonable range of pre- and post-synaptic firing rates. The LTP estimation accuracy depends on presynaptic firing rates: the more presynaptic neuron fires, the more accurate the LTP estimation. We found that the three pieces (unexplained variations in baseline, LTP and STP) should be estimated simultaneously. If we omit any one of these factors, the inferences for others will be biased. Finally, we provide two efficient way, i.e. 2D MLE and 1D approximation, to choose the appropriate hyper-parameters in adaptive smoothing.

In GLMs, omitted variables can result in biased parameter estimates for the effect that are included in the model18. Some previous research for synaptic strength inference had tried to tackle this problem by including postsynaptic history effects10,11,14,22. However, the postsynaptic history may not capture all effects that have impact on neural activity, and the history effects can also be biased by omitted variables. Therefore, we allow the postsynaptic baseline firing rates to be time-varying to compensate for slowly varying omitted variables. By conditioning on the adaptive baseline, the inference on synaptic strengths will be unbiased.

The neurons are usually connected in a network in vivo, and each single neuron may receive multiple inputs from an ensemble of neurons. Some previous methods can be used to analyze the functional connectivity in this situation. Our model can also extend to multiple input cases easily. Assume there are presynaptic neurons, denote the synaptic plasticity and synaptic connection between the pre- and post-synaptic neurons at as and . The synaptic plasticity can be further divided as LTP and STP, i.e. . Therefore, we can write the conditional intensity as . By setting and in adaptive smoothing and expanding parameter dimension for STP estimation, the model can be fitted by the same techniques.

Several alternative statistical and biophysical models were used to describe LTP (e.g. STDP and Ca-based LTP/LTD23) and STP (e.g. TM24). However, the examples here highlight the need for inferring short- and long-timescale effects simultaneously. Ignoring the long-term effects of short-term plasticity or ignoring fluctuations baseline firing can lead to spurious estimates of the synaptic strength. Also, recent studies show that STP and LTP can together tune sensitive range of synaptic plasticity, and this suggest the necessity for considering effects in different timescale at the same time.

This model provides a flexible framework for tracking short- and long-term variation in spike transmission, and more detailed modeling can be done based on the framework. For example, the LTP is currently modeled in an unstructured way to provide an unbiased reference. We can replace it by some more biophysical meaningful models, such as STDP and Ca-based LTP/LTD23, to make a more detailed inference.

# References

1. Barthó, P. *et al.* Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J. Neurophysiol.* **92**, 600–608 (2004).

2. Fetz, E., Toyama, K. & Smith, W. Synaptic Interactions between Cortical Neurons. in 1–47 (Springer, Boston, MA, 1991). doi:10.1007/978-1-4615-6622-9\_1.

3. Perkel, D. H., Gerstein, G. L. & Moore, G. P. Neuronal Spike Trains and Stochastic Point Processes: II. Simultaneous Spike Trains. *Biophys. J.* **7**, 419–440 (1967).

4. Ren, N., Ito, S., Hafizi, H., Beggs, J. M. & Stevenson, I. H. Model-based detection of putative synaptic connections from spike recordings with latency and type constraints. *J. Neurophysiol.* (2020) doi:10.1152/jn.00066.2020.

5. Kobayashi, R. *et al.* Reconstructing neuronal circuitry from parallel spike trains. *Nat. Commun.* **10**, 4468 (2019).

6. Zucker, R. S. & Regehr, W. G. Short-Term Synaptic Plasticity. *Annu. Rev. Physiol.* **64**, 355–405 (2002).

7. Swadlow, H. A. & Gusev, A. G. The impact of ‘bursting’ thalamic impulses at a neocortical synapse. *Nat. Neurosci.* **4**, 402–408 (2001).

8. Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsáki, G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* **11**, 823–833 (2008).

9. Linderman, S., Stock, C. H. & Adams, R. P. A framework for studying synaptic plasticity with neural spike train data. in *Advances in Neural Information Processing Systems* (eds. Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. & Weinberger, K. Q.) vol. 27 2330–2338 (Curran Associates, Inc., 2014).

10. Stevenson, I. H. & Kording, K. Inferring spike-timing-dependent plasticity from spike train data. in *Advances in Neural Information Processing Systems* (eds. Shawe-Taylor, J., Zemel, R. S., Bartlett, P., Pereira, F. C. N. & Weinberger, K. Q.) vol. 24 (2011).

11. Ghanbari, A., Malyshev, A., Volgushev, M. & Stevenson, I. H. Estimating short-term synaptic plasticity from pre- and postsynaptic spiking. *PLOS Comput. Biol.* **13**, e1005738 (2017).

12. English, D. F. *et al.* Pyramidal Cell-Interneuron Circuit Architecture and Dynamics in Hippocampal Networks. *Neuron* **96**, 505-520.e7 (2017).

13. Chan, R. H. M., Song, D. & Berger, T. W. Tracking temporal evolution of nonlinear dynamics in hippocampus using time-varying volterra kernels. in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* vol. 2008 4996–4999 (2008).

14. Ghanbari, A. *et al.* Modeling the Short-Term Dynamics of in Vivo Excitatory Spike Transmission. *J. Neurosci.* **40**, 4185–4202 (2020).

15. Eden, U. T., Frank, L. M., Barbieri, R., Solo, V. & Brown, E. N. Dynamic Analysis of Neural Encoding by Point Process Adaptive Filtering. *Neural Comput.* **16**, 971–998 (2004).

16. Brillinger, D. R. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybern.* **59**, 189–200 (1988).

17. Brillinger, D. R. Nerve Cell Spike Train Data Analysis: A Progression of Technique. *J. Am. Stat. Assoc.* (1992) doi:10.2307/2290256.

18. Stevenson, I. H. Omitted Variable Bias in GLMs of Neural Spiking Activity. *Neural Comput.* **30**, 3227–3258 (2018).

19. RAUCH, H. E., TUNG, F. & STRIEBEL, C. T. Maximum likelihood estimates of linear dynamic systems. *AIAA J.* **3**, 1445–1450 (1965).

20. Ananthasayanam, M. R., Mohan, M. S., Naik, N. & Gemson, R. M. O. A heuristic reference recursive recipe for adaptively tuning the Kalman filter statistics part-1: formulation and simulation studies. *Sādhanā* **41**, 1473–1490 (2016).

21. Du, Y. & Varadhan, R. SQUAREM: An R Package for Off-the-Shelf Acceleration of EM, MM and Other EM-Like Monotone Algorithms. *J. Stat. Software; Vol 1, Issue 7*  (2020) doi:10.18637/jss.v092.i07.

22. Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P. & Brown, E. N. A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *J. Neurophysiol.* **93**, 1074–1089 (2005).

23. Graupner, M. & Brunel, N. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proc. Natl. Acad. Sci.* **109**, 3991–3996 (2012).

24. Costa, R., Sjostrom, P. J. & van Rossum, M. Probabilistic inference of short-term synaptic plasticity in neocortical microcircuits. *Front. Comput. Neurosci.* **7**, 75 (2013).