**Dynamic filtering of spike count data with non-Poisson variability**

Ganchao Wei1\* and Ian H. Stevenson2-4

*1 Department of Statistics, University of Connecticut*

*2 Department of Psychological Sciences, University of Connecticut*

*3 Department of Biomedical Engineering, University of Connecticut*

*4 Connecticut Institute for Brain and Cognitive Science, University of Connecticut*

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

# Abstract

# Introduction

Drift…

(Rokni et al. 2007)

(Chestek et al. 2007)

(Stevenson et al. 2011)

(Tomko and Crapper 1974)

Stability…

(Dickey et al. 2009)

(Steinmetz et al. 2021) [shows that correcting for electrode drift reduces some instability]

[some amount of drift is functional] adaptation/plasticity…

(Lesica et al. 2007)

Variability…

(Ghanbari et al. 2019)

(Stevenson 2016)

(Fenton and Muller 1998)

(Barbieri et al. 2001)

(Maimon and Assad 2009)

(Churchland et al. 2010)

(Churchland et al. 2011)

(Eden et al. 2004)

(Brown et al. 2001)

(DeWeese and Zador 1998)

# Methods

## Conway-Maxwell Poisson Point Process Adaptive Filter (CMP-PPAF)

Or a fancier acronym?

Partition the total recording time T into N bins, such that . Assume there are neurons at time bin *k*. The bin size should large enough for multiple spikes. The number of spikes for neuron at bin k is , and all spike counts is , with . Further denote all observations up to k as .

Assume follows a Conway-Maxwell Poisson (CMP) distribution with parameters and , i.e. . (May add some property about CMP, e.g. dispersion performance for different ). Assume parameters are governed by latent state vector and history of the spiking process as and . If each observation at step is independent, the likelihood () and log-likelihood () are:

, where .

Assume the state vector evolves linearly with gaussian error (state equation) as , where is a system evolution matrix and represents gaussian noise with covariance . By Bayes’ theorem, the posterior of the state vector at time bin k is . The prior is gaussian distributed by the state equation but the likelihood is not gaussian, so there’s no closed form for the posterior distribution. In this paper, we use Laplace approximation, i.e. approximate posterior at each time bin by a gaussian distribution, followed by the rationale in Poisson adaptive filtering (Eden et al. 2004).

Denote the posterior mean and variance as and . According to the state equation, the prior mean and variance are and . After gaussian approximation for the posterior, we can obtain a recursive expression for the posterior mean and variance as follows:

We call this recursive algorithm as Conway-Maxwell Poisson point process adaptive filter (CMP-PPAF). If we further replace the observed Fisher information by expected Fisher information , the posterior update for variance is:

, which is called as “Fisher scoring” CMP-PPAF (fs-CMP-PPAF). The moments in the algorithm can be calculated by truncated summation, or calculate by asymptotic approximation of normalizing constant (Chatla and Shmueli 2018). The details of moments computation can be found in the appendix (TODO).

Given the posterior estimates from (fs)-CMP-PPAF, we can further pass these estimates backward by a RTS smoother (Rauch et al. 1965).

(TODO: write down details of RTS smoother)

Call the algorithms as (fs)-CMP-PPAS, where “S” means “smoother”.

TODO: Estimate Q

## Linear Model for CMP Parameters

In the remainder of this paper, we jointly model two CMP parameters by linear models with log-link, which is a special case of the above algorithm. Let , and . Then , , and . The update for posterior mean is not changed, but the update for the posterior variance in CMP-PPAF can be further simplified as:

However, the update is not robust to the outliers. Take for example: when , will be very small or even to a negative value, which will correspondingly lead to a large or even negative-definite . Because of the factorial, the “outliers” are common. When using the fs-CMP-PPAF, the issue caused by outliers will be resolved by sacrificing the efficiency. The update for posterior variance is:

## Improving Sensitivity for Small Sample Size

Information about variance for spike counts is essential for state vector estimation. However, when

the sample size is small, the information about variance is rare and the posterior estimation will largely lead by the recursive prior. This will make the posterior estimation insensitive to a large change. When using “Fisher scoring” (fs-CMP-PPAF) for robustness, the sensitivity will even be worse.

Solution: window

TODO: write down details and make a better cartoon…

Chart

Description automatically generated

Call the “window” with a new name? e.g. (fs)-wCMP-PPAF

## Quantifying Uncertainties

After estimation from the smoother, the posterior of is . Under linear models for and , . The covariance matrix of two CMP parameters, denoted as , can be calculated by property of multivariate log-normal distribution.

The conditional mean firing rate, is . The approximation works well when or . By delta method, we can calculate the variance of the mean firing rate, (kind of like VarCE (Churchland et al. 2011), but here is just the variance of estimates… In the following, I still mistakenly call it VarCE), as follows:

The direct calculation of and can be found in the appendix.

# Results

## Simulations (three cases)

Stimulus onset quenches neural variability is a widespread cortical phenomenon (Churchland et al. 2010). Motivated by this, we simulate single neuron records in the following three cases. The recording length is 10s, with bin size equals to 5ms.

(The simulations are about quenching neural variability, so we only have under-dispersion in these 3 cases)

In the first case, the mean firing rate increases and then decays exponentially, while the variance is the same as mean (Poisson neuron). In the second case, the mean firing rate holds constant, while the variance decreases and goes back exponentially (under-dispersion). In the third case, the mean firing rate changes as case 1, but the variance is less than mean at first (still under-dispersion). In all these three cases, two CMP parameters are governed by single state vector each. Here, we show the fitting results with fs-wCMP-PPAF and fs-wCMP-PPAS. The Qs are for all three cases, and the window size are 50, 100 and 50.

|  |  |  |
| --- | --- | --- |
| Case 1 | Case 2 | Case 3 |
|  |  |  |
|  |  |  |

To show the effect of “Fisher scoring” and “window”, we use case 3 as an example:

|  |  |  |
| --- | --- | --- |
|  | No window | window |
| No Fisher |  |  |
| Fisher |  |  |

“Fisher scoring”: (1) guarantee the robustness (in “no Fisher + window” still can see some jumps) but (2) sacrifice sensitivity/ efficiency (can be somewhat remedy by “window”).

“window”: (1) improve robustness when no “Fisher”; (2) improve sensitivity (important for single neuron tracking)

The algorithm is not very sensitive to window size. Can search over a coarse grid or pick up by convenience.

## Optimizing Process Noise Covariance

Optimizing prediction likelihood

No window. When turn on the “window”, Q will be overestimated.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Heatmap | Q\_lam  Y = estimated log10(Q\_lam)  X = true log10(Q\_lam)  Color: yellow = small Q\_nu; black = large Q\_nu | Q\_nu  Y = estimated log10(Q\_nu)  X = true log10(Q\_nu)  Color: purple = small Q\_lam; black = large Q\_lam | fitting |
| Fisher |  |  |  |  |
| No Fisher |  |  |  |  |

## Asymmetry

Mean constant, but variance switch upward and downward. The results for fs-wCMP-PPAF and fs-wCMP-PPAS. The optimized Q = diag(3.845\*1e-4 2.500\*1e-4), and window size is 20.

|  |  |
| --- | --- |
|  |  |
|  |  |

To show the asymmetry effect, we plot as (DeWeese and Zador 1998). Still no Fisher/ no window is fitted for comparison. The optimized Q for no “Fisher” is diag(4.267\*1e-4 1.613\*1e-4).

|  |  |  |
| --- | --- | --- |
|  | No window | window |
| No Fisher |  |  |
| Fisher |  |  |

Top-left panel:

When we don’t use the Fisher, the CMP-PPAF will assign large weight to the under-expected spike counts but assign small weight to the over-expected observations (See argument for robustness in the method part). This effect dominates the asymmetry argument by (DeWeese and Zador 1998), and makes CMP-PPAF more sensitive to downward switch. When pass the estimation into smoother backwardly, downward switch is essentially an upward switch, and there’s no further CMP-specific weight. So, by the asymmetry argument in DeWeese, the asymmetry effect will even be larger.

Bottom-left panel:

When use the Fisher, the weight/ variance doesn’t depend on observation at k. So as expected, downward switch is less sensitive than upward switch now.

Second column:

Using the window improve the sensitivity/ accuracy during the switch. The asymmetry effect is smaller.

## Application 1: V1

Not well tuned neuron: neuron = 13



Fit by full data (TODO: in the following fit, 7 basis for both lambda and nu)



Use 50% for training, 50% for test…

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | nBasis\_X = 10  nBasis\_G = 5 | nBasis\_X = 10  nBasis\_G = 1 | nBasis\_X = 10  constant nu | nBasis\_X = 10  nu = 1 |
| Train llhd/trial | -85.2017 | -85.3573 | -85.6872 | -87.7028 |
| Test llhd/trial | -87.8935 | -87.8327 | -87.8811 | -89.6786 |

|  |  |  |
| --- | --- | --- |
| nBasis\_X = 10  nBasis\_G = 5 | nBasis\_X = 10  nBasis\_G = 1 | nBasis\_X = 10  constant nu |
|  |  |  |

TODO: heatmap?

Well tuned neuron: neuron = 11



TODO: use full data

50% for training…

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | nBasis\_X = 10  nBasis\_G = 5 | nBasis\_X = 10  nBasis\_G = 1 | nBasis\_X = 10  constant nu | nBasis\_X = 10  nu = 1 |
| Train llhd/trial | -67.3465 | -65.9934 | -70.0691 | -87.5171 |
| Test llhd/trial | -71.0509 | -75.0462 | -73.0102 | -87.9039 |

|  |  |  |
| --- | --- | --- |
| nBasis\_X = 10  nBasis\_G = 5 | nBasis\_X = 10  nBasis\_G = 1 | nBasis\_X = 10  constant nu |
|  |  |  |

## Application 2: Place Cell in Hippocampus

Neuron = 12

nBasis\_X = 20, nBasis\_G = 1

when plotting

colIdx = find(max(heatmap\_mean, [], 1) < max(spk\_raw(:))\*2);

heatmap\_mean = heatmap\_mean(:, colIdx);

heatmap\_var = heatmap\_var(:, colIdx);

heatmap\_t = heatmap\_t(colIdx);

Graphical user interface

Description automatically generated

# Discussion

Omitted variables can increase the apparent variability of observations via the law of total variance. For example, in the hippocampus, place cell firing is highly variable on different passes through the field (cite Fenton). This may be partially due to joint selectivity to position, speed, and head direction, as well as the influence of theta phase. Here, rather than model these distinct covariates assuming Poisson observations, we allow the variability to be non-Poisson.

# Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1931249.

# References

**Barbieri R**, **Quirk MC**, **Frank LM**, **Wilson MA**, **Brown EN**. Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *J Neurosci Methods* 105: 25–37, 2001.

**Brown EN**, **Nguyen DP**, **Frank LM**, **Wilson MA**, **Solo V**. An analysis of neural receptive field plasticity by point process adaptive filtering. *Proc Natl Acad Sci* 98: 12261–12266, 2001.

**Chatla SB**, **Shmueli G**. Efficient estimation of COM–Poisson regression and a generalized additive model. *Comput Stat Data Anal* 121: 71–88, 2018.

**Chestek CA**, **Batista AP**, **Santhanam G**, **Yu BM**, **Afshar A**, **Cunningham JP**, **Gilja V**, **Ryu SI**, **Churchland MM**, **Shenoy K V**. Single-neuron stability during repeated reaching in macaque premotor cortex. *J Neurosci* 27: 10742–10750, 2007.

**Churchland AK**, **Kiani R**, **Chaudhuri R**, **Wang XJ**, **Pouget A**, **Shadlen MN**. Variance as a Signature of Neural Computations during Decision Making. *Neuron* 69: 818–831, 2011.

**Churchland MM**, **Yu BM**, **Cunningham JP**, **Sugrue LP**, **Cohen MR**, **Corrado GS**, **Newsome WT**, **Clark AM**, **Hosseini P**, **Scott BB**, **Bradley DC**, **Smith M a**, **Kohn A**, **Movshon JA**, **Armstrong KM**, **Moore T**, **Chang SW**, **Snyder LH**, **Lisberger SG**, **Priebe NJ**, **Finn IM**, **Ferster D**, **Ryu SI**, **Santhanam G**, **Sahani M**, **Shenoy K V**. Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat Neurosci* 13: 369–378, 2010.

**DeWeese M**, **Zador A**. Asymmetric dynamics in optimal variance adaptation. *Neural Comput* 10: 1179–1202, 1998.

**Dickey AS**, **Suminski A**, **Amit Y**, **Hatsopoulos NG**. Single-unit stability using chronically implanted multielectrode arrays. *J Neurophysiol* 102: 1331–1339, 2009.

**Eden UT**, **Frank LM**, **Barbieri R**, **Solo V**, **Brown EN**. Dynamic Analysis of Neural Encoding by Point Process Adaptive Filtering. *Neural Comput* 16: 971–998, 2004.

**Fenton A**, **Muller R**. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc Natl Acad Sci U S A* 95, 1998.

**Ghanbari A**, **Lee CM**, **Read HL**, **Stevenson IH**. Modeling stimulus-dependent variability improves decoding of population neural responses. *J Neural Eng* 16, 2019.

**Lesica NA**, **Jin J**, **Weng C**, **Yeh CI**, **Butts DA**, **Stanley GB**, **Alonso JM**. Adaptation to Stimulus Contrast and Correlations during Natural Visual Stimulation. *Neuron* 55: 479–491, 2007.

**Maimon G**, **Assad J a.** Beyond Poisson: Increased Spike-Time Regularity across Primate Parietal Cortex. *Neuron* 62: 426–440, 2009.

**Rauch HE**, **Tung F**, **Striebel CT**. Maximum likelihood estimates of linear dynamic systems. *AIAA J* 3: 1445–1450, 1965.

**Rokni U**, **Richardson AG**, **Bizzi E**, **Seung HS**. Motor learning with unstable neural representations. *Neuron* 54: 653–666, 2007.

**Steinmetz NA**, **Aydin C**, **Lebedeva A**, **Okun M**, **Pachitariu M**, **Bauza M**, **Beau M**, **Bhagat J**, **Böhm C**, **Broux M**, **Chen S**, **Colonell J**, **Gardner RJ**, **Karsh B**, **Kloosterman F**, **Kostadinov D**, **Mora-Lopez C**, **O’Callaghan J**, **Park J**, **Putzeys J**, **Sauerbrei B**, **Daal RJJ van**, **Vollan AZ**, **Wang S**, **Welkenhuysen M**, **Ye Z**, **Dudman JT**, **Dutta B**, **Hantman AW**, **Harris KD**, **Lee AK**, **Moser EI**, **O’Keefe J**, **Renart A**, **Svoboda K**, **Häusser M**, **Haesler S**, **Carandini M**, **Harris TD**. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science (80- )* 372, 2021.

**Stevenson IH**. Flexible models for spike count data with both over- and under- dispersion. *J Comput Neurosci* 41: 29–43, 2016.

**Stevenson IH**, **Cherian A**, **London BM**, **Sachs NA**, **Lindberg E**, **Reimer J**, **Slutzky MW**, **Hatsopoulos NG**, **Miller LE**, **Kording KP**. Statistical assessment of the stability of neural movement representations. *J Neurophysiol* 106: 764–774, 2011.

**Tomko GJ**, **Crapper DR**. Neuronal Variability - Nonstationary Responses to Identical Visual-Stimuli. *Brain Res* 79: 405–418, 1974.

# Appendix

## 1. Derivation for CMP-PPAF

First, make a gaussian approximation to the posterior:

Then, taking the log of both sides gives

, where R is a constant. Now, differentiate once and twice with respect to :

If the gaussian approximation is valid, this relation holds approximately for all values of . Evaluating the two derivatives above at yields:

Notice the gradient is the score function and the negative hessian is the observed Fisher information.

To help with derivation, denote . The transpose of score function is:

And the observed Fisher information is:

## 2. Calculation for CMP Moments

TODO