**Dynamic modeling of spike count data with Conway-Maxwell 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

Here we consider a dynamic GLM with Conway-Maxwell Poisson (CMP) observations to describe time-varying spike counts. We first introduce the model. Although the CMP distribution allows us to flexibly model non-Poisson variability, one major challenge with using this model is that there are not closed-form solutions for the CMP likelihood. Here, we fit the model using a global Gaussian approximation, and we discuss several additional technical challenges that arise when using the CMP distribution with a dynamic GLM.

## Dynamic Conway-Maxwell Poisson Model

A count observation , such as the spike count for a neuron, is assumed to follow the CMP distribution, with parameters and . The probability mass function (pmf) of CMP is

, where is the normalizing constant. The parameter controls different dispersion patterns, i.e. equi- (), over- () or under-dispersion (). Three common distributions occur as special cases: 1) the Poisson (), 2) geometric (), and 3) Bernoulli ().

For multiple observations up to steps, such as simultaneous spike counts from neurons, denote the counts at time bin as , for The corresponding CMP parameters at are and . Previous work has examined the CMP-GLM (Chatla and Shmueli 2018; Sellers and Shmueli 2010), and here we focus on the dynamic version of this GLM. The CMP parameters at are modeled by two log-linear models, and , with and , and and denote known predictors. Under the CMP-GLM, the parameters are static. Here, we assume that they progress linearly with a Gaussian noise.

The observations follow conditionally independent CMP distributions, given the state vector .

While the state vector evolves linearly with Gaussian noise:

Given the initial state mean , covariance , linear dynamics and process covariance .

## Inference by Gaussian approximation

To fit the model to data we need to estimate the time-varying state vector In this section, we first assume **F** and **Q** are known. Since the observations are CMP distributed, we cannot estimate in closed form. Instead, here we approximate it by a multivariate Gaussian distribution, , with . The parameters of this Gaussian are found by a global Laplace approximation, i.e., and . The log-posterior is given by:

, where is the log-likelihood. The log-posterior is concave (Gupta et al. 2014), and the Markovian structure of the state vector dynamics makes it possible to optimize by Newton-Raphson (NR) in time (Paninski et al. 2010). After the Newton update, we can further quantify the uncertainty for the CMP parameters and the underlying rates (see Appendix).

There are several technical challenges involved with performing the Newton update with CMP observations. Firstly, in order to find the gradient and Hessian we need to calculate moments of and , which have no closed forms (Shmueli et al. 2005). We can calculate these moments by truncated summation. However, when and , truncated summation is computationally costly since we need many steps for accurate approximation. In this case, we approximate the moments using previous (Chatla and Shmueli 2018; Gaunt et al. 2019) asymptotic results (see Appendix). A second challenge is that the Hessian is not robust to outliers. Outliers often result in the Hessian being close to singular or even positive-definite (see details in Appendix). To ensure robustness, we use Fisher scoring where the observed information is replaced by the expected information. Finally, a third challenge is that the Newton updates take a long time to converge if the initial state estimate is far from the maximum of the posterior, especially when is large. To resolve this issue, we use a smoothing estimate with local Gaussian approximation as a “warm start”. Forward filtering for a dynamic Poisson model has been previously described (Eden et al. 2004), and, here we implement CMP filtering following the same rationale. Let and be the mean and variance for the one-step prediction density and and be mean and variance for the posterior density, then the filtering update for step is given by

Here, to again ensure robustness, we use Fisher scoring when updating the state covariance. We then find smoothed estimates using a backward pass ((Rauch et al. 1965)). Although doing smoothing is fast, the estimates can be inaccurate, especially when there are large changes in the state vector. In the forward filtering stage, the Gaussian approximation at each step is conducted locally at the recursive prior . This will be statistically inefficient when the recursive prior is too far away from the posterior mode, or when there is a large change in the state vector. Moreover, Fisher scoring reduces the efficiency of the smoother even further. The smoother provides reasonable initial estimates, but estimation accuracy is substantially improved by using Newton’s method to find the global Laplace approximation for the posterior.

## Estimating process noise

For the applications to neural data examined here, we assume that . However, we still need to estimate the process noise . When is small, especially when , different values will have a substantial influence on estimation. One possible way to estimate is to use an Expectation Maximization (EM) algorithm (Macke et al. 2011). However, using the Laplace approximation for during the E-step breaks the usual guarantee of non-decreasing likelihoods in EM, and, hence, may lead to divergence. To avoid that, we could sample the posterior directly by MCMC. However, the lack of closed-form moments for the CMP distribution makes sampling computationally intensive. Here, to estimate robustly and quickly, we instead assume is diagonal and estimate it by maximizing the prediction likelihood in the filtering stage, as in (Wei and Stevenson 2021).

## Neural Data

[need to add section here with description and citations for V1 and HC data, similar to Neural Comp paper or can also look at the J Comp Neuro CMP paper. This should also include details on and ]

I prefer to move this section to application part (reason: this is the data we use, but the not the method we propose. If we collect the data by ourself, surely we should write it here. But we are just using it…)

# Results

## Figure 1

Code: [figure1\_singleNu\_dirShift.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/figure1/figure1_singleNu_dirShift.m)

Chart, histogram

Description automatically generated with medium confidence

Figure 1. Fitting a simulated neuron with a shifting firing pattern. (A) The tuning curve of the neuron shifts over time, with the preferred stimulus orientation changing and the response amplitude increasing. (B) At the same time, the variability in spiking changes from being over-dispersed relative to a Poisson distribution to under-dispersed. This leads to a decreasing Fano factor from 1.94 to 0.38 overall. (C) To illustrate the shifts we show the tuning curve at two time points: Trial 20 (blue) and Trial 80 (red). Dots denote… Dashed lines denote… and solid lines denote… (D) To illustrate the shift in dispersion over time we show the true (solid) and estimated (dashed) Fano Factor for two specific stimuli as a function of time. The dispersion for the initial preferred orientation is shown in cyan, while the dispersion for the final preferred orientation is shown in yellow.

The neural responses are nonstationary over time, not only in mean but also in variance. There is growing evidence that the neural variability also changes over time, depending on an animal’s alertness or motivation, as well as, the specific stimuli or behavior. To track changes in both mean response properties (e.g. tuning) and variability, here we examine a dynamic Conway-Maxwell-Poisson model.

To illustrate how the CMP-DGLM can track both time-varying mean and dispersion, here we simulated a neuron with time-varying tuning curve, where the response to hypothetical visual stimuli shifts over 100 trials. The response amplitude also increases along the time at the same time (panel A and C in Figure 1). Meanwhile, the dispersion pattern also changes from over-dispersion to under-dispersion. We then fitted the simulated data, with the same generated predictors: be 10-knots cubic spline basis expansion of the orientation and , at trial . The panel C and D in figure 1 show the fitting results for the selected slices.

By using CMP observations, rather than Poisson or NB observations, the Fano Factor can be both <1 and >1. Since these fits are model-based… we avoid some technical difficulties with previous approaches to estimating dispersion directly (e.g. mean-matching from Churchland et al. Nat Neuro).

## Simulation

Ideas…

* Q determines the timescales in the state estimates… we can’t track arbitrarily fast changes. Changes in dispersion are somewhat hard to track.
* (Fig 2) Practical issue – most neuroscientists only look at the mean, but you could imagine the dispersion changing while the mean is constant… (Scott … Pillow Neurips – NB-LDS constant dispersion param?). Also illustrate how CMP-DGLM can track over- and under-dispersion.

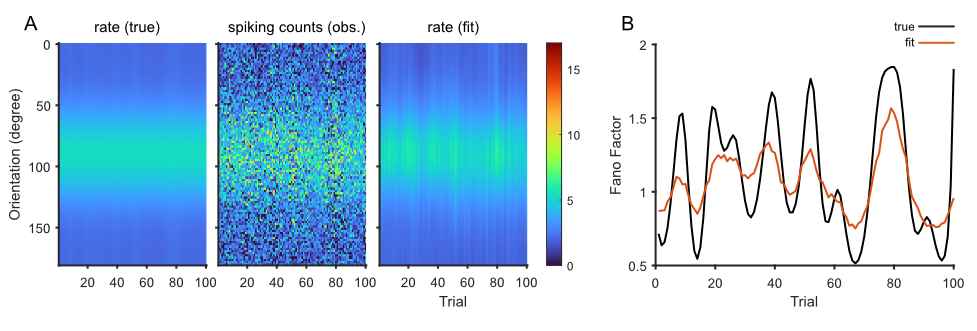


Figure 2. The mean firing rate is controlled to be constant across the trial, but the Fano Factor varies.

Chart

Description automatically generated with low confidence

Figure 3.The over-dispersed data is generated by dynamic CMP, with 2 knots modeling and keep across the trial. After holding 80% data out in the speckled pattern, we fit dynamic CMP and dynamic Poisson models. (A) The estimations are different. In this case, the estimated place field in dynamic CMP is better (it’s very anecdotal, just for this seed). (B) When evaluating the max response in each trial, the dynamic Poisson estimates are biased (underestimated) and the estimation error is also underestimated. The solid line gives the MAP estimates of mean firing rate, and the dashed lines show one standard deviations from MAP. The standard error of dynamic Poisson is calculated as in appendix, using the truncated summation.

nanmean((CMP\_mean\_fit\_trans - spk\_test).^2, 'all') = 5.2135

nanmean((POI\_mean\_fit\_trans - spk\_test).^2, 'all') = 5.3114



## Application

We next applied our method to two datasets: 1) Utah array extracellular recordings of visually evoked activity from anesthetized macaque primary visual cortex (referred as the V1 dataset), and 2) recordings from different rat hippocampal regions while the rat was performing linear maze task (referred as the HC data).

V1 data

In the V1 dataset (Kohn and Smith 2016), the macaque monkeys were anesthetized by a continuous intravenous infusion of sufentanil citrate. To minimize eye movements, the animals were paralyzed with a continuous intravenous infusion of vecuronium bromide (0.1 mg/kg/h). The neural activities were recorded by “Utah” array. The waveform segments were sorted by hand with modified competitive mixture decomposition methods. (Cite Shoham et al, 2003).

We use the data when presenting a movie of sinusoidal gratings with ~100 different drift directions (300ms each, 30s movie in total). The trial was replicated 120 times. For further details on how the data were obtained, see (Kelly et al. 2010; Smith and Kohn 2008).

In this analysis, we set bin size to be 0.3s, but ignoring the spike counts for the first 0.05s for each bin to […?]. Considering the sequential measurement effect, we analyze the data with single neuron input. The predictors and are circular cubic spline basis expansion of the grating directions. The basis number for is 5, while it varies between 1 and 3 for model comparisons.

Graphical user interface

Description automatically generated with medium confidence

Hippocampus data

In the HC dataset, the rat was running back and forth along the 250cm linear track. The recording holds up to ~66 min. The recorded spikes are further sorted by program KlustaKwick (<https://github.com/klustateam/klustakwik>) for automatic spike sorting, then by Klusters, http://klusters.sourceforge.net/ for manual adjustment. The input predictors are 10-knots cubic spline basis expansion of positions, and the for all .

Code: [hc\_comparison\_v2.m](https://github.com/weigcdsb/COM_POISSON/blob/main/demo/hc/hc_comparison_v2.m);

I’m still confused about the following plot (middle & bottom panels). Although the upper panel can clearly show 2 place fields, the empirical mean (mid) and Fano factor (bottom) shouldn’t be calculated without considering directions. When showing the fitted vs. observation mean & FF (especially FF), we still need to separate 2 directions out.

Basically, if we include the upper panel, we need to go back and forth between “direction-on” and “direction-off”.



I Prefer to just show this



BTW, I’m kind of afraid the hippocampus example is a bit redundant for an applied stats paper. Although previously you persuade me that this additional example is valuable for neuroscience, **the story we tell and the figures here (as far as I can imagine) should be very similar to figure 4 (V1 data)**. If we really need to show this, maybe we need to emphasize something different?

E.g. the place fields are sharper here. Maybe we can try to see if something like **Figure 3** occurs here, by extracting, say 1/2 data or even more, out?

But I don’t know if this is appropriate. This is an analysis for real data: is there any practical motivation to throw 1/2 data away?

# 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.

[Some references are not correct, but they are fixed in LaTeX]

# 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 AA**, **Muller RU**. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proc Natl Acad Sci U S A* 95: 3182–3187, 1998.

**Gaunt RE**, **Iyengar S**, **Olde Daalhuis AB**, **Simsek B**, **Robert Gaunt BE**. An asymptotic expansion for the normalizing constant of the Conway-Maxwell-Poisson distribution. *Ann Inst Stat Math* 71: 163–180, 2019.

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

**Gupta RC**, **Sim SZ**, **Ong SH**. Analysis of discrete data by Conway–Maxwell Poisson distribution. *AStA Adv Stat Anal* 98: 327–343, 2014.

**Kelly RC**, **Smith MA**, **Kass RE**, **Lee TS**. Local field potentials indicate network state and account for neuronal response variability. *J Comput Neurosci* 29: 567–579, 2010.

**Kohn A**, **Smith MA**. Utah array extracellular recordings of spontaneous and visually evoked activity from anesthetized macaque primary visual cortex (V1) [Online]. *CRCNS.org* 2016.http://dx.doi.org/10.6080/K0NC5Z4X.

**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.

**Macke JH**, **Buesing L**, **Cunningham JP**, **Yu BM**, **Shenoy K V.**, **Sahani M**. Empirical models of spiking in neural populations. *Adv Neural Inf Process Syst* 24, 2011.

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

**Paninski L**, **Ahmadian Y**, **Ferreira DG**, **Koyama S**, **Rahnama Rad K**, **Vidne M**, **Vogelstein J**, **Wu W**. A new look at state-space models for neural data. *J. Comput. Neurosci.* 29Springer: 107–126, 2010.

**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.

**Sellers KF**, **Shmueli G**. A flexible regression model for count data. *https://doi.org/101214/09-AOAS306* 4: 943–961, 2010.

**Shmueli G**, **Minka TP**, **Kadane JB**, **Borle S**, **Boatwright P**. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *J R Stat Soc Ser C (Applied Stat* 54: 127–142, 2005.

**Smith MA**, **Kohn A**. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28: 12591–12603, 2008.

**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**, **van Daal RJJ**, **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.

**Wei G**, **Stevenson IH**. Tracking Fast and Slow Changes in Synaptic Weights From Simultaneously Observed Pre- and Postsynaptic Spiking. *Neural Comput* 33: 2682–2709, 2021.

# Appendix

## Quantifying Uncertainties

After convergence, we have an approximation of the log-posterior , and we can use this approximation to quantify he uncertainty about the CMP parameters, as well as about the mean rate at each time.

The CMP parameters are log-normal distributed. Let , then . Denote the variance of CMP parameters as . The variance can be easily found in Wiki. If I write it out, I need to define some useless notations. . and .

The conditional mean firing rate is , whose variance can be calculated by the Delta method:

We can calculate the moments as in the first section of the appendix, or we can use a simpler approximation when or . Then and .

## Moments approximation for Conway-Maxwell Poisson distribution

To estimate the state-vector for the dynamic CMP model, we need to find first and second moments for and . For ,

Generally, these moments can be calculated by truncated summation.

However, when and , we need many steps for accurate approximation. In this case, we make use of a previous asymptotic results for efficient calculation. Let , and

Then the moments are (The highlights can be found in the reference, the covariance is calculated by myself):

## Gradient and Hessian of the log-posterior

We estimate the state vector by maximizing the log-posterior with Newton-Raphson updates. Denote , the -th update of NR algorithm is .

The gradient is:

The Hessian:

, where

When , the Hessian may be ill-conditioned or even positive-definite. To ensure the robustness, do Fisher scoring, i.e. replace the observed information by the expected information , so that .