Why deconvolve Ca^{2+}?

Mathew H. Evans

June 30, 2017

Abstract

2-photon fluorescence imaging of somatic calcium (Ca^{2+} for short) is an increasingly popular technique for recording the activity of large groups of neurons (cite reviews). Given that Ca^{2+} is an indirect measurement of neural activity, and signal quality considerations (background noise, slow kinetics and heterogeneity in expression across cells) the fluorescence time-series is typically *deconvolved* prior to tuning or correlation analysis. Recent efforts have seen an explosion in the number of available methods for Ca^{2+} deconvolution, with seemingly impressive results (Theis, spikefinder). Here we reconsider this progress showing (i) a popular metric - Pearson correlation coefficient (PCC) - is poorly conditioned (results change a great deal with small changes in parameters), and is heavily biased towards overestimating firing rates (see also Ganmor), (ii) when applied to data collected at large-scale-recording resolution (with sampling rates of 2-10Hz) performance is much poorer than expected from published results, (iii) the purported advantages of deconvolution do not outweigh the disadvantages.

1 Introduction

why deconvolve
- signal noise
- temporal resolution
- normalisation of expression across neurons

how to deconvolve
- many methods (not the point)
- comparing methods involves metrics
- metrics are important

the cautionary tale
- if the signal isn't perfect our metrics and methods lead to biased estimates of neural properties (FR, tuning, correlations)
- this bias is worse than the benefits we're supposed to get when deconvolving

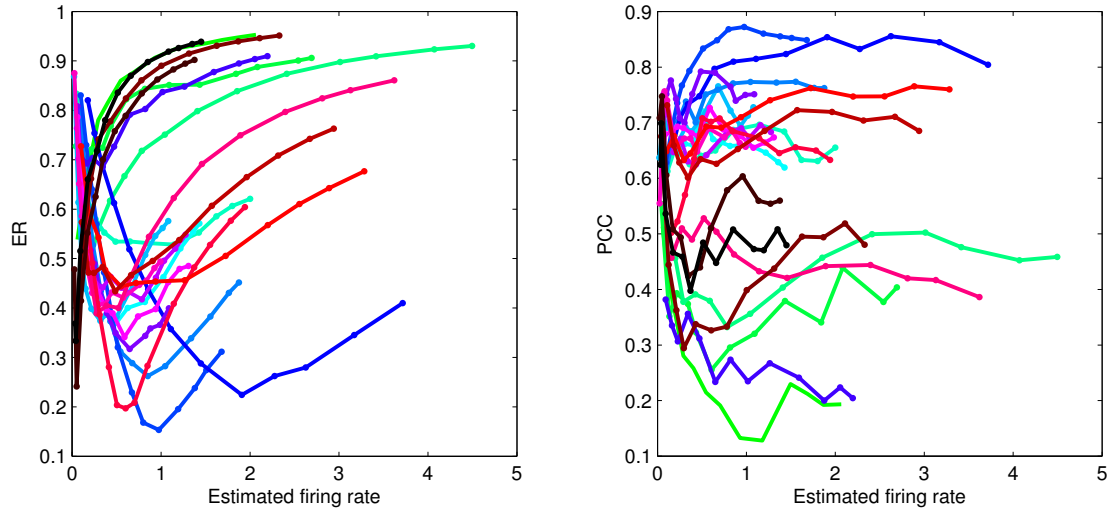
2 Results

Pearson correlation coefficient over-estimates firing rate

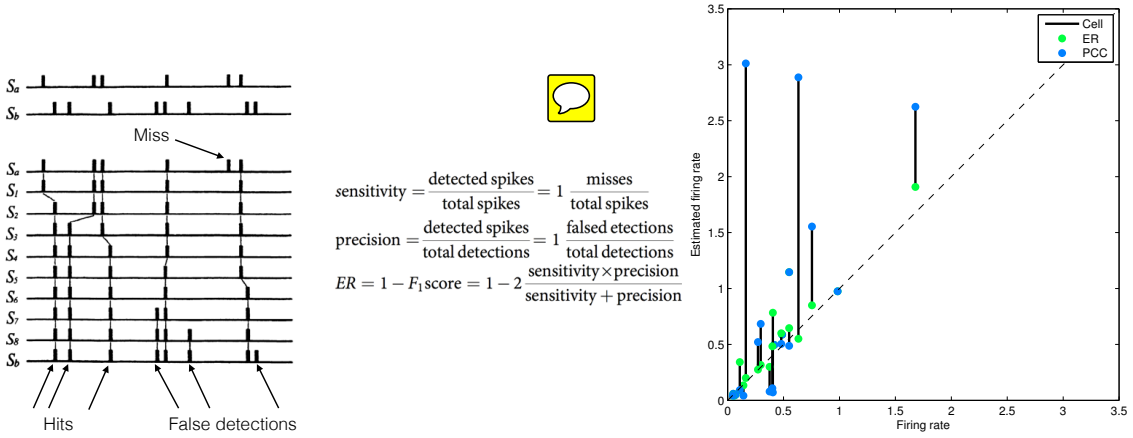
To assess the performance of deconvolution algorithms at estimating spike trains from Ca^{2+} signals, methods are tested on *ground truth* datasets - where the spiking activity of a cell was recorded simultaneously with Ca^{2+} imaging, ideally using high-signal-to-noise techniques such as juxtacellular recording.

Figure 1 (a) shows the results from decorating ground truth datasets using Suite2P (Pachitariu et al) using a range of a threshold parameter which trades off misses vs false detections. Results assessed by PCC are on the right. In general PCC increases as estimated firing rate increases. However, this leads to overestimation of firing rate when comparing the true firing rate to the estimated firing rate using the 'best' (highest PCC) parameters (Figure 1 (c)). Another downside of PCC as a metric is it does not change smoothly with parameter changes, as can be seen in the lines for individual cells in Figure 1 (a).

To address the weaknesses of PCC, we implemented the Error Rate (ER) spike distance metric of Victor & Purpura - using code from Deneaux et al 2016) - which is described in Figure 1 (b). ER returns a normalised score which is 0 for a perfect match, and 1 when all the spikes are missed. When applied to deconvolution of ground truth data ER is lowest for an intermediate estimated firing rate, suggesting that estimates closer to the true firing rate are rewarded with good scores (Figure 1 (a)). This intuition is shown to be true when comparing the 'best' estimate (lowest ER) to true firing rate for each cell (Figure 1 (c)). In addition, unlike for PCC, individual cell results in Figure 1 (a) show ER varies smoothly with deconvolution parameters.



(a)



(b) (c)

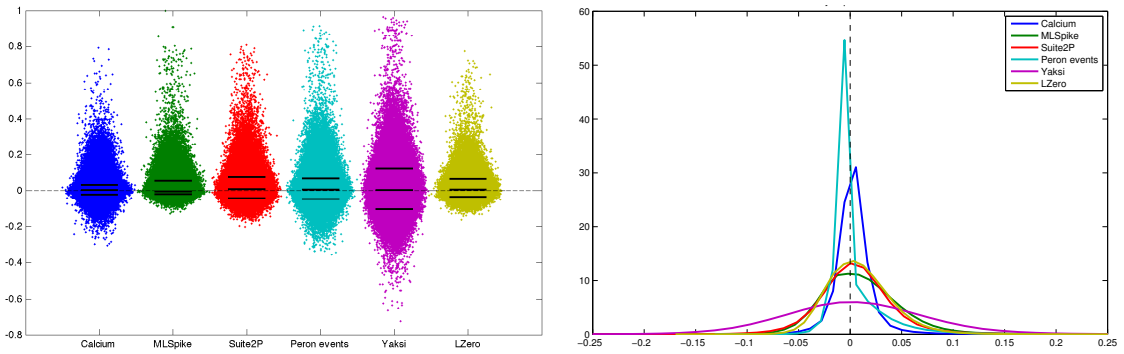
Figure 1: Differences in spike metrics. (a) Metric score as a function of estimated firing rate. As deconvolution parameters are changed to increase the estimated firing rate - to trade off hits at the cost of false positive errors - ER is penalised while PCC continues to improve. ER also changes smoothly with deconvolution parameters, whereas PCC is more stochastic, leading to noisy estimates of the best parameters. Colours are different cells. (b) Left: ER. Victor and Purpura (1996) proposed a spike metric to compare spike trains. This metric is generated by determining the number of elementary operations (shift, addition, or deletion of individual spikes - depicted as rows in here) required to match two spike trains, up to some temporal precision (here 0.5s). Right: In Deneaux et al 2016 the Error Rate (ER) is similarly computed as a normalised ratio of sensitivity vs precision in spike detection. Detections are counted to within 0.5s. (c) Estimated firing rate for 'best' deconvolution parameters versus real firing rate. Best parameters are taken as the lowest (left) or highest (right) points in (a). PCC both over and underestimates firing rate. ER also overestimates firing rate but to a lesser degree.

Deconvolution leads to bias

Always a trade-off between false positives and misses.
Miss spikes
Overestimate background rates (false positive) (see Ganmor), therefore
- tuning estimates are weakened
- correlation estimates are noisier

Tuning - POSS. TO DO compute number of cells significantly touch/whisking/trial selective

Correlations



(a) (b)

Figure 2: Pairwise correlations. MLSpike and Suite2P - using parameters tuned to increase ER on ground truth data - have correlation distributions more similar to noise than using raw Ca^{2+}. The Peron events have a large number of PCCs below zero, suggesting that choosing parameters that penalise false-positives may be actively decorrelating the data.

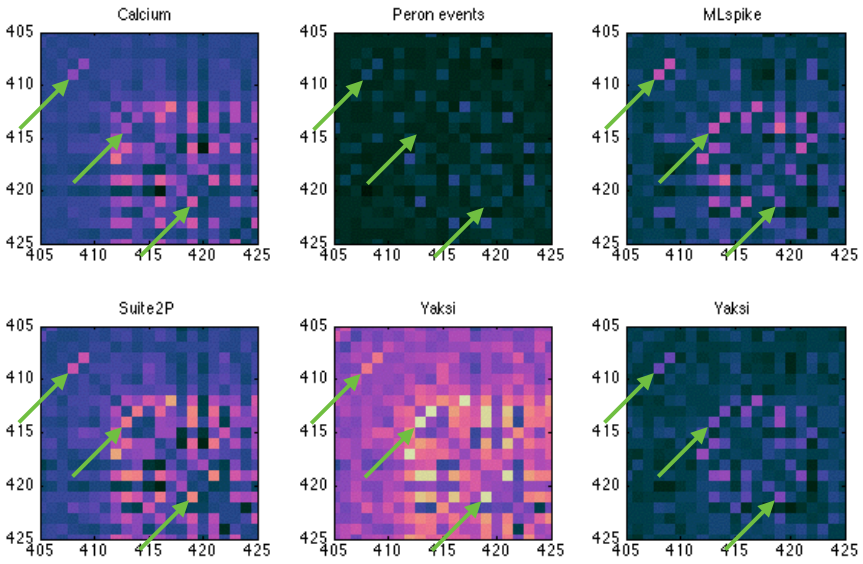


Figure 3: Peron events may actively decorrelate neurons. Each panel shows a subset of pairwise correlations between neurons from one session of data from Peron et al 2015, one panel for each method. Some pairs (arrows) showing high correlation across methods are missing (middle arrow) for the Peron events method.

Deconvolution does not improve temporal resolution under realistic conditions

Comparison of estimates when deconvolving vs not

- **Tuning**

PSTHs

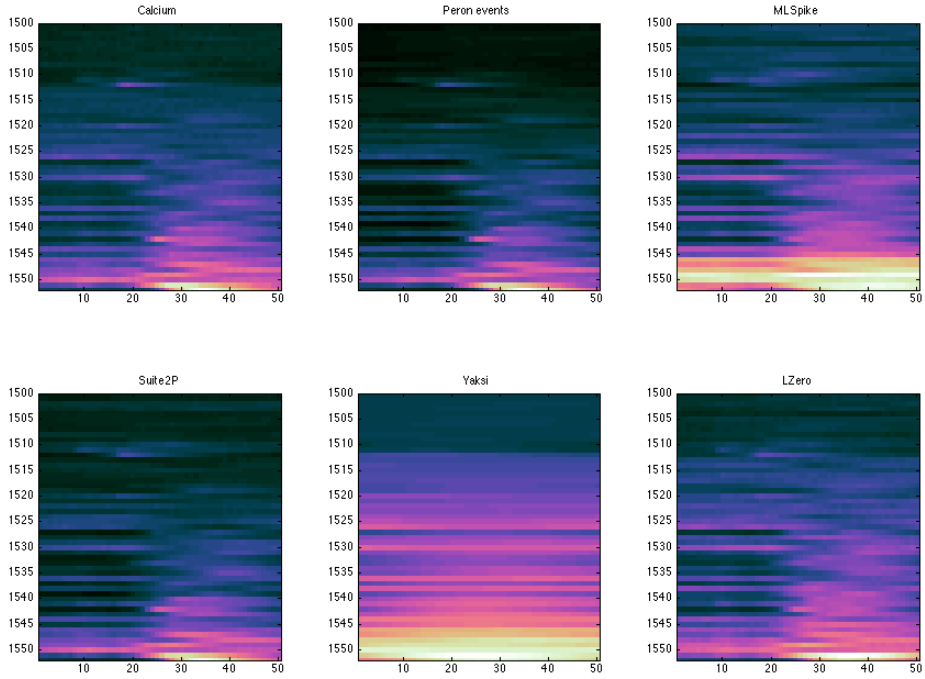


Figure 4: PSTHs do not show increase in temporal sharpness following deconvolution.

- **Temporal correlations**

Pick small number of cells e.g. of touch tuned vs some other parameter, and compute their correlations over time. Yaksi and Friedrich showed an advantage in their hands.

We would expect to do a better job of distinguishing touch vs delay activity or delay vs reward. Is this true?

3 Methods

Spike train metrics

Pearson correlation coefficient - down sampled or gaussian convolved. Victor and Purpura 1996 Error Rate.

List of deconvolution methods

Suite2P

`Suite2P` (https://github.com/cortex-lab/Suite2P) is actively developed by Marius Pachitariu and members of the cortexlab (Kenneth Harris and Matteo Carandini) at UCL. Suite2P's USP is it's application to large scale 2-photon imaging analysis, with an emphasis on end-to-end processing (images to neural event time series) and speed. A preprint describing the toolbox is available here:

http://biorxiv.org/content/early/2016/06/30/061507,

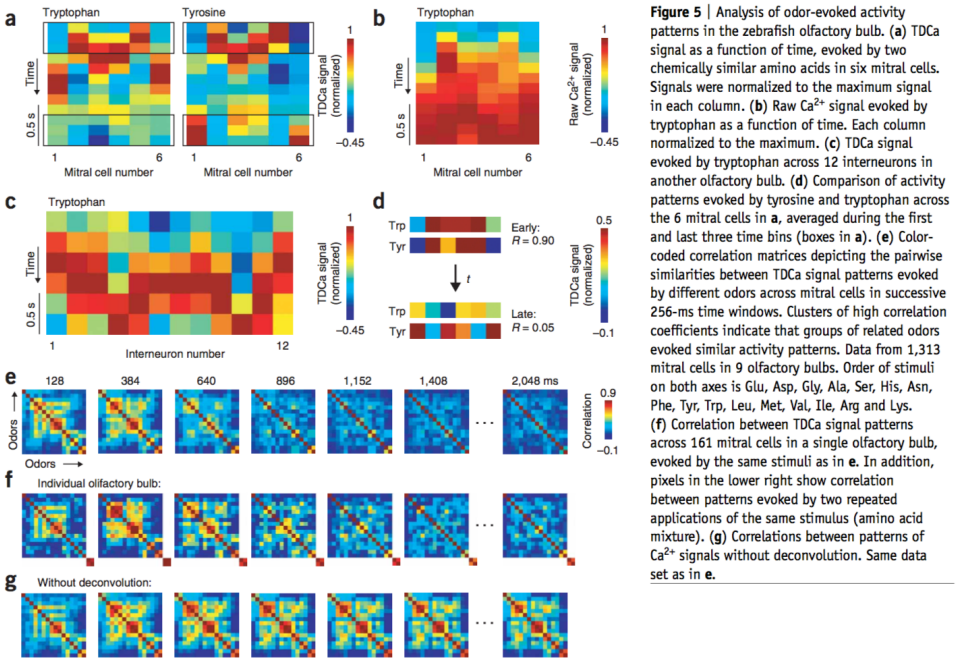


Figure 5 | Analysis of odor-evoked activity patterns in the zebrafish olfactory bulb. (**a**) TDCa signal as a function of time, evoked by two chemically similar amino acids in six mitral cells. Signals were normalized to the maximum signal in each column. (**b**) Raw Ca^{2+} signal evoked by tryptophan as a function of time. Each column normalized to the maximum. (**c**) TDCa signal evoked by tryptophan across 12 interneurons in another olfactory bulb. (**d**) Comparison of activity patterns evoked by tyrosine and tryptophan across the 6 mitral cells in **a**, averaged during the first and last three time bins (boxes in **a**). (**e**) Color-coded correlation matrices depicting the pairwise similarities between TDCa signal patterns evoked by different odors across mitral cells in successive 256-ms time windows. Clusters of high correlation coefficients indicate that groups of related odors evoked similar activity patterns. Data from 1,313 mitral cells in 9 olfactory bulbs. Order of stimuli on both axes is Glu, Asp, Gly, Ala, Ser, His, Asn, Phe, Tyr, Trp, Leu, Met, Val, Ile, Arg and Lys. (**f**) Correlation between TDCa signal patterns across 161 mitral cells in a single olfactory bulb, evoked by the same stimuli as in **e**. In addition, pixels in the lower right show correlation between patterns evoked by two repeated applications of the same stimulus (amino acid mixture). (**g**) Correlations between patterns of Ca^{2+} signals without deconvolution. Same data set as in **e**.

Figure 5: Deconvolution resolves the fine timescale of pairwise correlations, fig from Yaksi and Friedrich (2006).

and our own notes on the spike detection algorithm are here:

https://drive.google.com/open?id=1NeQhmoRpS-x8R0e84w3TqkUR1PNMXiem6ZIjJta-U7A.

MLSpike

`MLSpike` (https://github.com/mlspike) was developed by Thomas Deneux at INT, CRNS Marseille, France. A model-based probabilistic approach, `MLSpike` was developed to recover spike trains in calcium imaging data by taking baseline fluctuations and cellular properties into account. A comprehensive explanation of the algorithm and its benefits can be found in the paper:

Deneux, Thomas, Attila Kaszas, Gergely Szalay, Gergely Katona, Tamás Lakner, Amiram Grinvald, Balázs Rózsa, and Ivo Vanzetta. "Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo." Nature Communications 7 (2016).

Link: https://www.nature.com/articles/ncomms12190

LZero

The method we refer to as `LZero` was developed by Sean Jewell and Daniela Witten from U.Washington, Seatle, USA. The goal for this implementation was to cast spike detection as a change-point detection problem, which could be solved with an existing $l0$ optimization algorithm. In their paper Jewell and Witten show that the $l0$ solution is better than previously implemented $l2$ solutions, with results much closer to the real spike train ($l2$ solutions tend to overestimate the true firing rate). Details can be found in the paper:

Jewell, Sean, and Daniela Witten. "Exact Spike Train Inference Via ℓ_0 Optimization." arXiv preprint arXiv:1703.08644 (2017).

Link: https://arxiv.org/abs/1703.08644

Yaksi

`Yaksi` refers to the 'vanilla' deconvolution of Yaksi and Friedrich (2006). This is to be used as a baseline for comparison with more sophisticated methods. ***NOTE 8.6.17*** my implementation results in signals that are more temporally smooth (as opposed to more temporally sharp) than the calcium signal, indicating the filtering has not been performed properly.

The method is detailed in the paper:
Yaksi, Emre, and Rainer W. Friedrich. "Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging." Nature Methods 3, no. 5 (2006): 377-383.

Peron events

`Peron events` refer to the extracted events detailed in the original *Peron et al. 2015* paper. It is a version of the 'peeling' algorithm tuned to generate a low number of false positive detections (a rate of 0.01Hz) on ground truth data, leading to a hit rate of 54%
Peron, Simon P., Jeremy Freeman, Vijay Iyer, Caiying Guo, and Karel Svoboda. "A cellular resolution map of barrel cortex activity during tactile behavior." Neuron 86, no. 3 (2015): 783-799.