Why deconvolve Ca^{2+}?

Mathew H. Evans

October 9, 2017

Abstract

Two-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. We conclude that the purported advantages of deconvolution may not outweigh the disadvantages of losing information and introducing spurious noise.

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

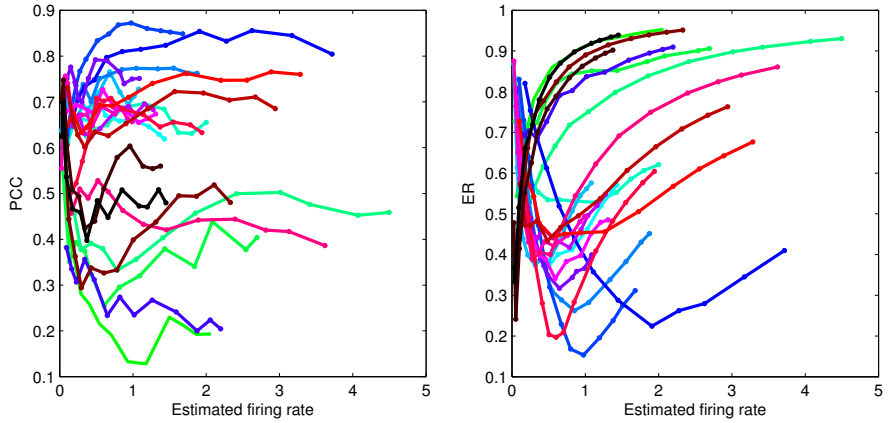
The point of this section is that if you are going to do deconvolution, PCC is a biased metric so use something else e.g. ER

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.

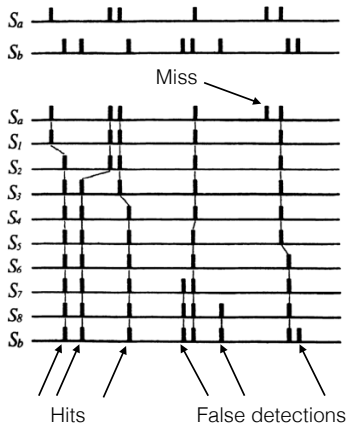
Figure 1 (a) shows the results from decorrelating ground truth datasets using Suite2P (Pachitariu et al) using a range of a threshold parameter which trades off misses vs false detections. Results assessed by PCC are on the right. In general PCC increases as estimated firing rate increases. However, this can lead to extreme overestimation of the true 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 Deneaux et al 2016, a summary measure based on the distance measure of Victor & Purpura (1996). ER (described in Figure 1 (c)) 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 'best' (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 of firing rate to the true firing rate for each cell (Figure 1 (d)). Over the population PCC overestimates firing rate, in some cases severely (Figure 1 (e)). Though ER also overestimates firing rate, it does so to a much smaller degree (mean error 0.1Hz). In addition, unlike for PCC, individual cell results in Figure 1 (b) show ER varies smoothly with deconvolution parameters.

TO DO: add a results for other methods to Fig 1 (e). Also - important re application to large-scale imaging like Peron - repeat the analysis using degraded data (lower sampling rate to 7Hz in the first instance).



(a) (b)

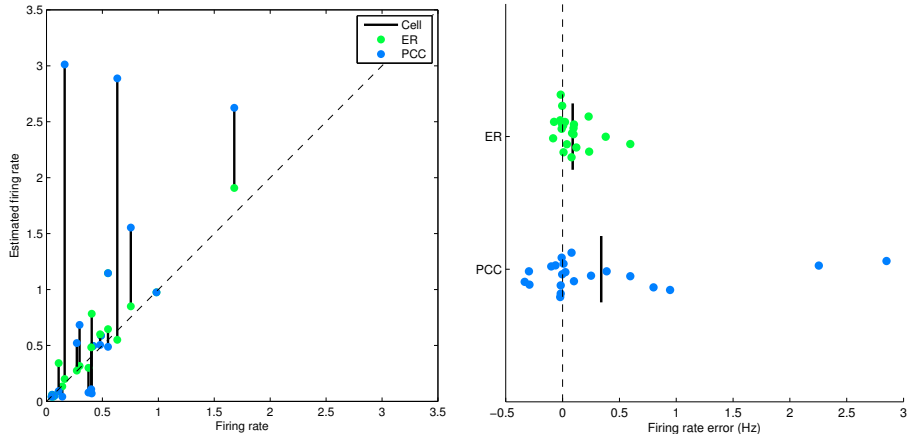


$$\text{sensitivity} = \frac{\text{detected spikes}}{\text{total spikes}} = 1\ \frac{\text{misses}}{\text{total spikes}}$$

$$\text{precision} = \frac{\text{detected spikes}}{\text{total detections}} = 1\ \frac{\text{falsed etections}}{\text{total detections}}$$

$$ER = 1 - F_1\text{score} = 1 - 2\frac{\text{sensitivity} \times \text{precision}}{\text{sensitivity} + \text{precision}}$$

(c)



(d) (e)

Figure 1: Differences in spike metrics. (a) PCC as a function of estimated firing rate. As deconvolution parameters are changed to incrementally increase the estimated firing rate - to trade off hits at the cost of false positive errors PCC continues to improve. (b) In contrast, ER is penalised for overestimating firing rate. ER changes smoothly with gradual changes in deconvolution parameters, whereas PCC is more stochastic, leading to noisy estimates of the best parameters. Colours are different cells. (c) 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. (d) Estimated firing rate for 'best' deconvolution parameters versus real firing rate. Best parameters are taken as the lowest or highest points in (a) and (b), respectively. (e) Firing rate error (estimated FR - true FR). PCC both over and underestimates firing rate. ER also overestimates firing rate but to a lesser degree. Lines are means

2.1 Deconvolution leads to biased and noisy estimates of pairwise correlation

Deconvolution is always a trade-off between false positives and misses, meaning you get both
Miss real spikes + overestimate background rates (see also Ganmor), therefore
- correlation estimates are noisier (due to false positives) or biased (due to misses/decorrelation)
- tuning estimates are weakened (next section)

To determine the effect of deconvolution on population level measures such as pairwise correlations (to infer functional networks) and tuning distributions (how many cells are/are not tuned to particular stimuli) we applied each deconvolution method to data from a single experiment from Peron et al 2015. In Peron et al 2015 two-photon Ca^{2+} imaging was used to record neural activity from up to 2000 neurons simultaneously at 7Hz from superficial barrel (Layer 2/3 somatosensory) cortex as mice performed a head-fixed tactile localisation task with their whisker. For the results presented here 1552 neurons were recorded for a total of just over 56 minutes (23559 time points). This relatively long recording ensured good estimation of the measured parameters (TO DO: quantify this).

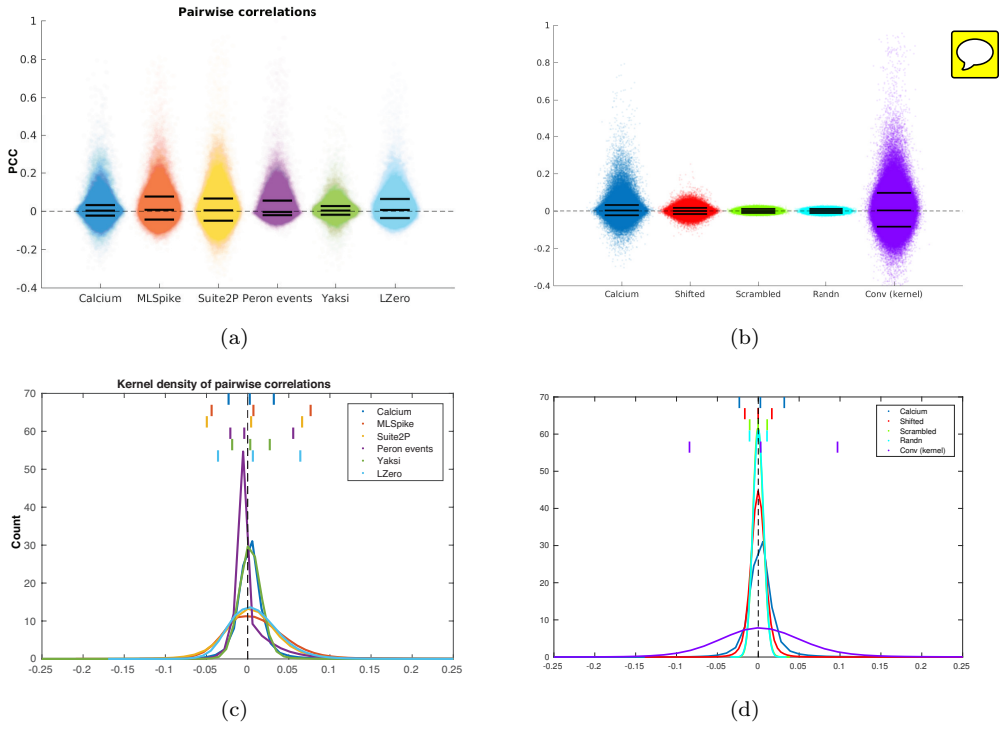


Figure 2: Pairwise correlations. (a) For each method (colours) each dot is the Pearson correlation coefficient between each pair of neurons (1203576 unique pairs, excluding self-correlations). X-axis jitter added for clarity. Black lines correspond to 5th, 50th (median) and 95th percentiles of the data. (c) Kernel density plots of data in (a). Lines above density plots correspond to 5th, 50th (median) and 95th percentiles of the data, coloured to match the distributions. (b) and (d) are the same as (a) and (c) but with different permutations or surrogates of the data. Calcium - data as in (a). Shifted - the fluorescence time series for each cell was randomly shifted in time (using Matlab's circshift function) by up to 10000 frames. Scrambled - elements of the original N x T data matrix were sampled randomly (without replacement) to generate a new data matrix. Randn - pseudorandom values drawn from a normal distribution. Conv (kernel) - original data convolved with an exponentially decaying kernel as is used in the MLSpike and Suite2P deconvolution methods. MLSpike and Suite2P - using parameters tuned to increase ER on ground truth data - have broader distributions more similar to that resulting from smoothing the 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 decorrellating the data.

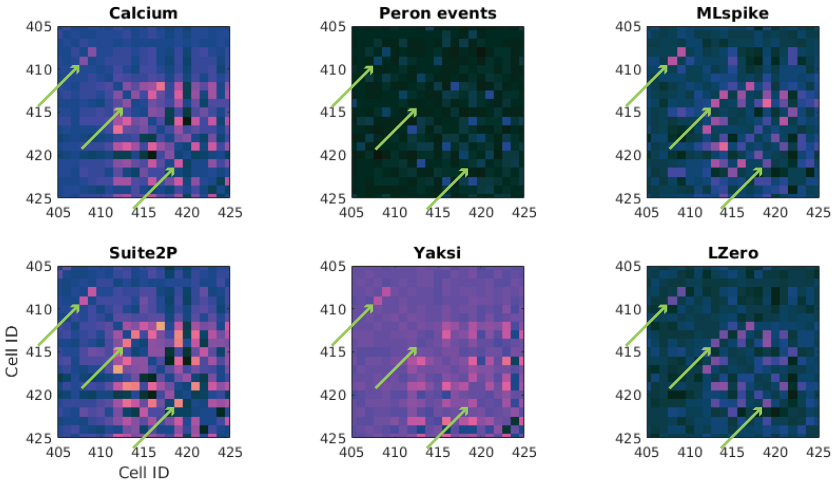


Figure 3: Peron events and Yaksi may actively decorrelate neurons. Each panel shows a subset of pairwise correlations between neurons from one session of data from Peron et al 2015, one panel for each method. Some pairs (arrows) showing high correlation across methods are missing (middle arrow) for the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

Deconvolution does not improve the precision of temporal resolution under realistic conditions

Comparison of estimates when deconvolving vs not

- signal to noise

- temporal resolution (rise/decay time)

- TO DO: 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.

To determine whether deconvolution improves the temporal precision of analyses such as tuning curves we computed the touch-triggered average for all cells in the example dataset. In the somatosensory system, touch onset is a salient sensory signal known to drive a subset of neurons with short latency and low jitter (O'Connor et al 2010, Hires et al 2014).

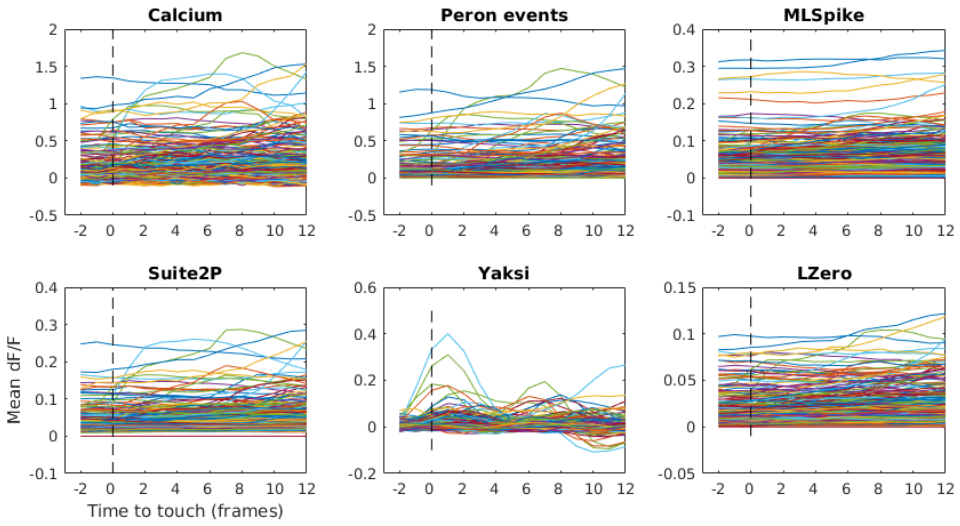


Figure 4: Comparing touch-triggered average (mean deconvolved FR per imaging frame) from different deconvolution methods for one all cells. Touch occurs at time zero. Only the Yaksi method shows a temporally sharp touch response.

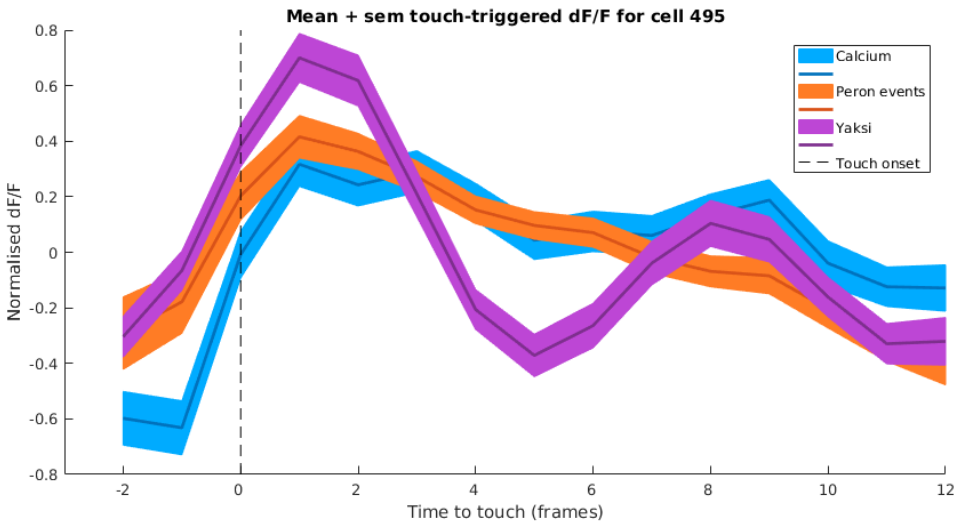
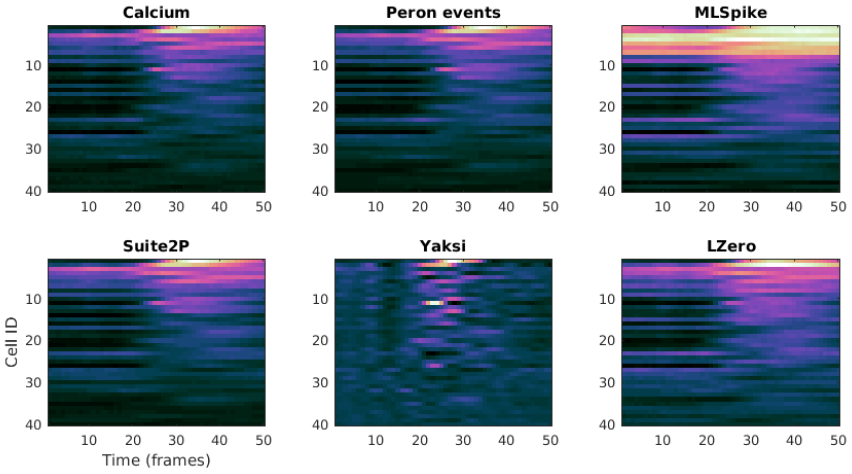
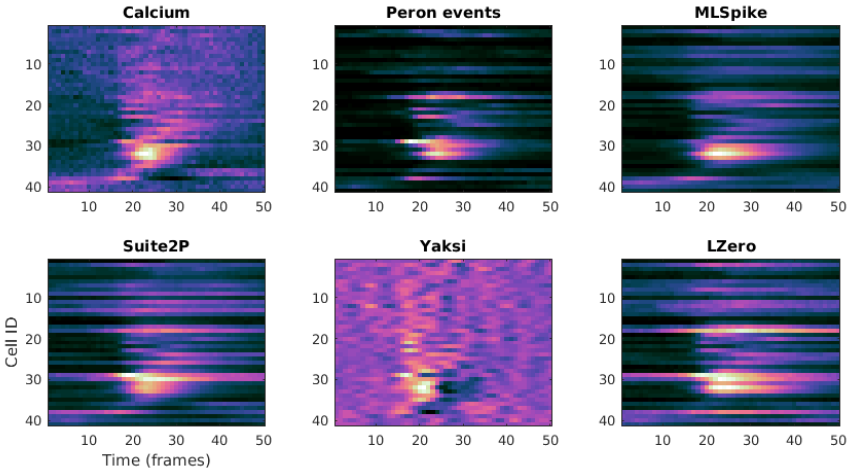


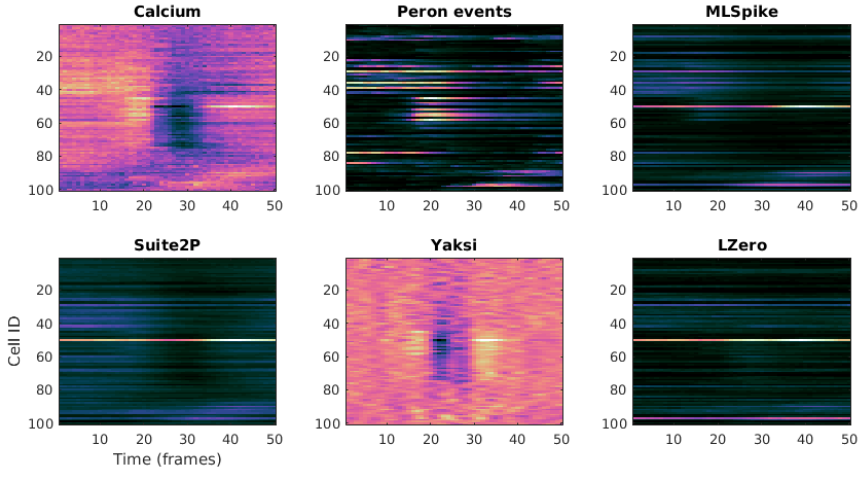
Figure 5: Comparing touch-triggered average (mean and s.e.m) from different deconvolution methods for one cell. Touch occurs at time zero. While the Yaksi method shows a temporally precise touch response, decaying within 1s (<7 frames), the Calcium and Peron events show a sustained response.



(a)



(b)



(c)

Figure 6: PSTHs do not show obvious increase in temporal sharpness following deconvolution, but do show signs of missing important features of the data. Each panel shows a different cluster of neurons sorted by t-SNE ordering of PSTHs derived from Calcium data. Pole up is at frame 9, pole down at frame 17, and reward cue is at frame 22. NOTE: Group 3 is aligned to the (likely lick artefact induced) dip in df/f. If this is indeed artifactual, the change in signal still results in a 'rebound' response using Peron (and others, when looking at raw data, figures not shown here)

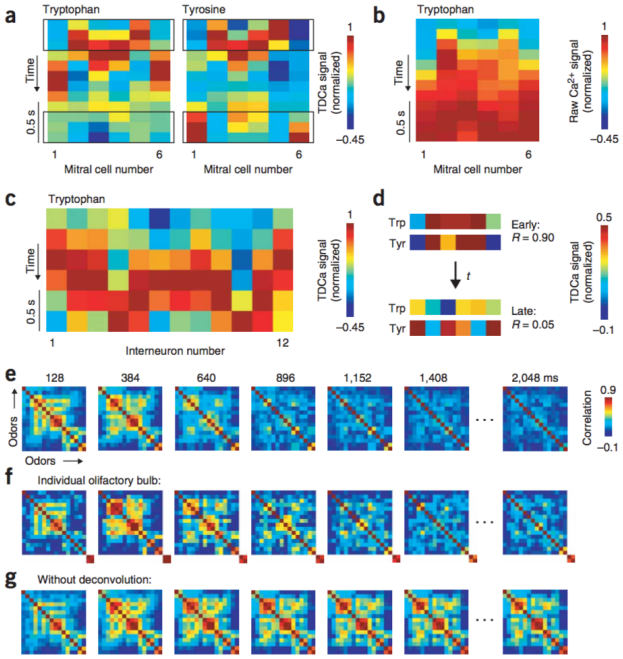


Figure 5 | Analysis of odor-evoked activity patterns in the zebrafish olfactory bulb. (**a**) TDCa signal as a function of time, evoked by two chemically similar amino acids in six mitral cells. Signals were normalized to the maximum signal in each column. (**b**) Raw Ca^{2+} signal evoked by tryptophan 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 7: Deconvolution resolves the fine timescale of pairwise correlations, fig from Yaksi and Friedrich (2006).

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,

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 l_0 optimization algorithm. In their paper Jewell and Witten show that the l_0 solution is better than previously implemented l_2 solutions, with results much closer to the real spike train (l_2 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.

4 Supplemental

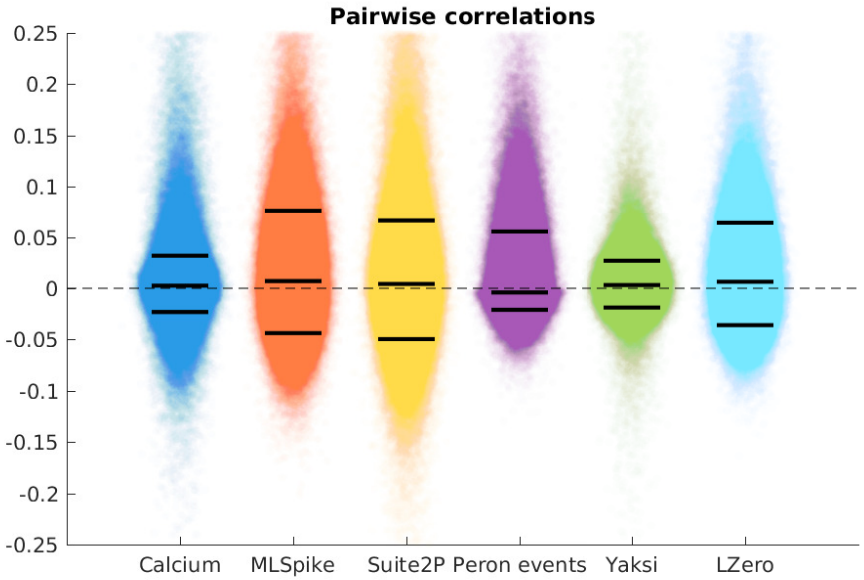


Figure 8: Zoomed in version of 2

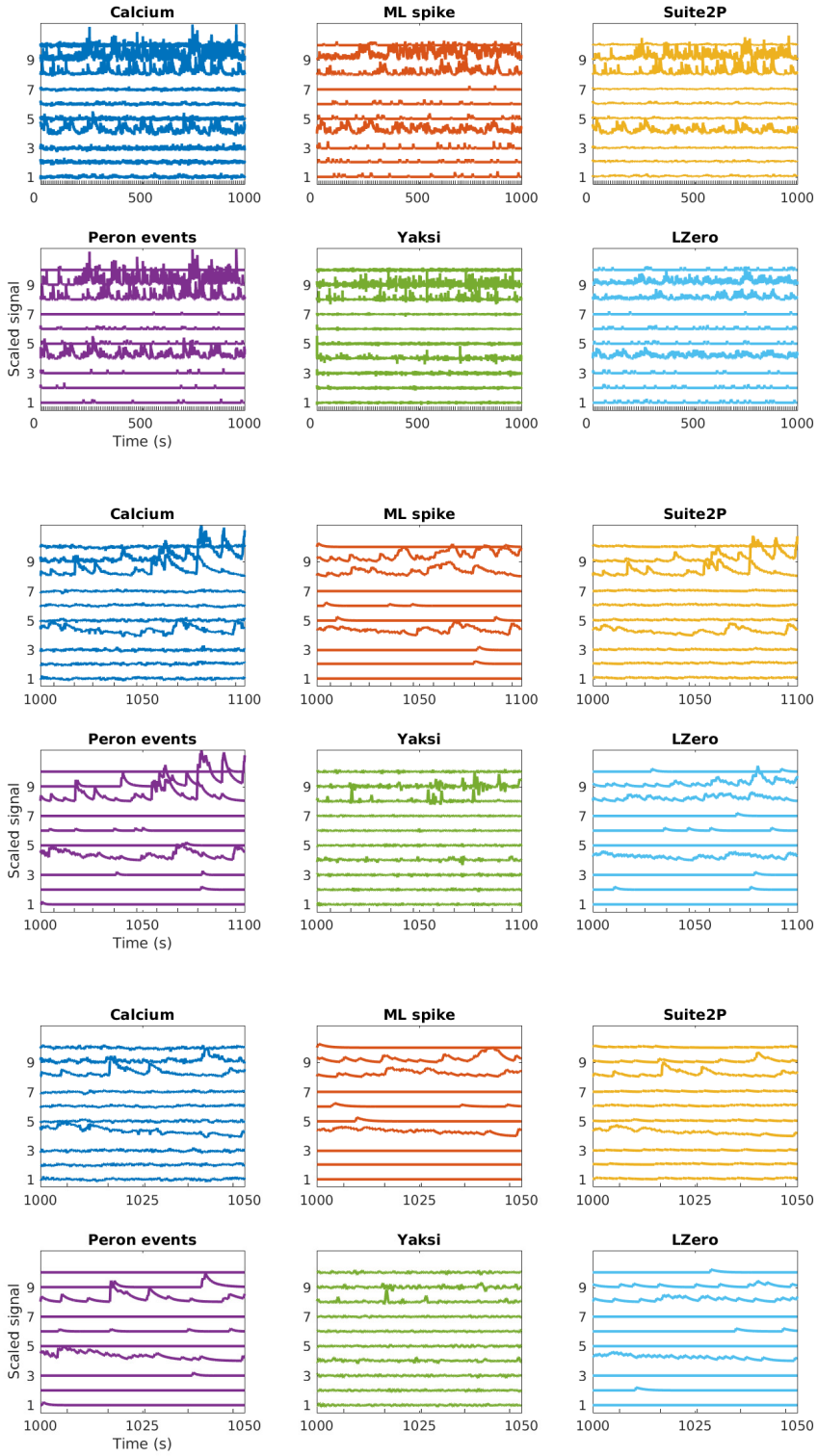


Figure 9: Example deconvolution for 10 cells using each method. Different figures are the same data at increasing levels of zoom.

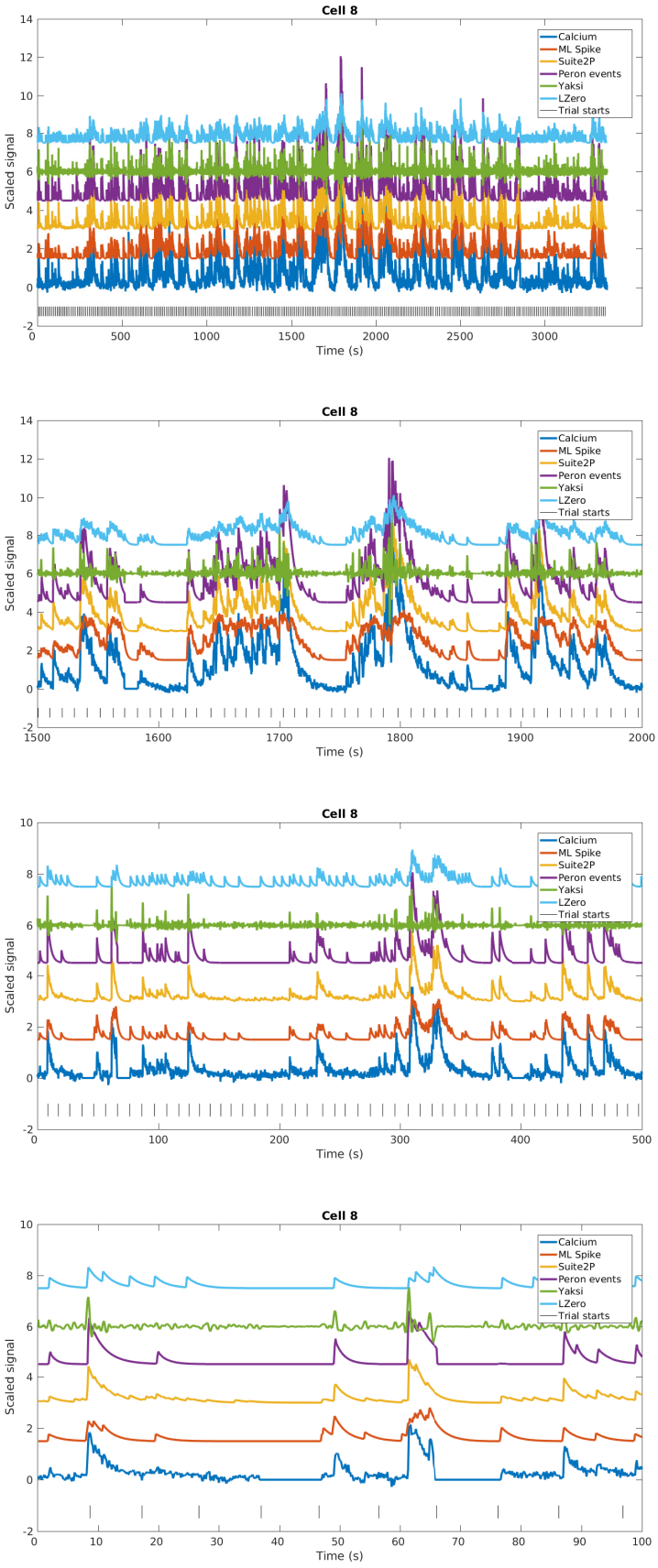


Figure 10: Example deconvolution for one cell using each method. Different figures are the same data at increasing levels of zoom.