

# On the use of calcium deconvolution algorithms in practical contexts

Mathew H. Evans, Rasmus S. Petersen & Mark D. Humphries [add affiliations]

April 25, 2019

## Abstract

Calcium imaging is a powerful tool for capturing the simultaneous activity of large populations of neurons. Studies using it to address scientific questions of population dynamics and coding often use the raw time-series of changes in calcium fluorescence at the soma. But somatic calcium traces are both contaminated with multiple noise sources and are non-linearly related to spiking. A suite of methods are available to further process the raw calcium signal: denoising to recover spike-evoked calcium; or deconvolving the spikes themselves. Here we explore the extent to which our choice of raw, denoised, or spike-deconvolved calcium time-series affects the scientific inferences we can draw. Our results show the choice qualitatively changes the potential scientific inferences we draw about neural activity, coding, and correlation structure. We show that a substantial fraction of the processing methods fail to recover simple features of population activity in barrel cortex already established by electrophysiological recordings. Raw calcium time-series contain an order of magnitude more cells tuned to task features; yet there is also qualitative disagreement between deconvolution methods on which neurons are tuned. Finally, we show that raw and processed calcium time-series qualitatively disagree on the structure of correlations within the population and the dimensionality of its joint activity. We suggest that results obtained from population calcium-imaging be verified across multiple forms of the calcium time-series.

## 1 Introduction

Calcium imaging is a wonderful tool for high yield recordings of large neural populations [Harris; Stringer; Ahrens; Orger; others]. Many pipelines are available for moving from pixel intensity across frames of video to a time-series of calcium fluorescence in the soma of identified neurons [cite loads; including van Rossum latest Sci Report paper].

But raw calcium fluorescence is nonlinearly related to spiking, and contains noise from a range of sources. These issues have inspired a wide range of deconvolution algorithms [cite Theis benchmarking; Stringer Curr Opin], which attempt to turn raw somatic calcium into something more closely approximating spikes. We address here the question facing any systems neuroscientist using calcium imaging: do we use the raw calcium, or attempt to clean it up?

We distinguish here two classes of deconvolution algorithms. The first is denoising, which attempts to extract all calcium events that are spike-related [e.g. Yaksi and Ganmor?]. A further step is spike deconvolution, inferring individual spikes, usually starting from the denoised traces [loads of examples... Stringer]. These options raise the further question of the extent to which we should proceed with cleaning up the raw calcium signals. Thus our aim here is to understand if our choice matters: how do our scientific inferences depend on our choice of raw, de-noised, or spike-deconvolved calcium time-series.

We proceed in two stages. In order to use spike deconvolution algorithms, we need to choose their parameters. We'd like to know whether it is worth taking this extra step: how good can these algorithms be in principle, and how sensitive their results are to the choice of parameter values. We thus first evaluate qualitatively different spike deconvolution algorithms, by optimising their parameters against ground truth data with known spikes. With our understanding of their parameters in hand, we then turn to our main question, by analysing a large-scale population recording from the barrel cortex of a mouse performing a whisker-based decision task. We compare the scientific inferences about population coding and correlations we obtain using either raw calcium signals, denoised signals, or deconvolved spikes.

We find contrasting answers. A substantial fraction of the methods used here fail to recover basic features of population activity in barrel cortex established from electrophysiology. The inferences we draw about coding qualitatively differ between raw and deconvolved calcium signals. In particular, coding analyses based on raw calcium signals detect an order of magnitude more cells tuned to task features. Yet there is also qualitative disagreement between deconvolution methods on which neurons are tuned. The inferences we draw about correlations between neurons do not distinguish between raw and deconvolved calcium signals, but can qualitatively differ between spike and non-spike methods. Our results thus suggest care is needed in drawing inferences from population recordings of somatic calcium, and that one solution is to replicate all results in both raw and deconvolved calcium signals.

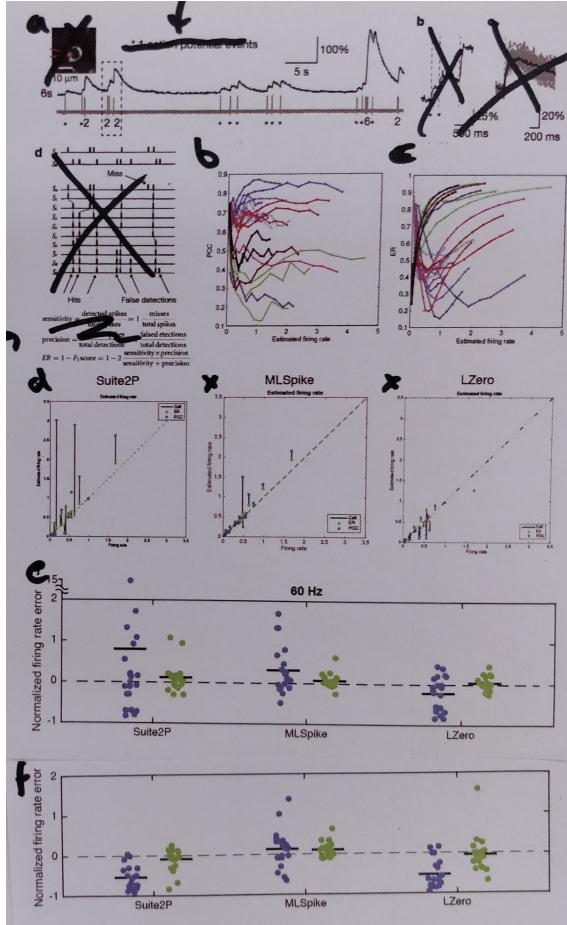
## 2 Results

### 2.1 Performance of spike deconvolution algorithms on ground-truth data-sets

We select here three spike deconvolution algorithms, each an example of the state of the art in qualitatively different approaches to the problem: Suite2p ([Pachitariu et al.](#)), a peeling algorithm that matches a kernel to the calcium signal to detect spike-triggered calcium events; LZero ([Jewell and Witten, 2017](#)), a change-point detection algorithm, which find the step-like changes in the calcium signal the imply spikes [Check]; and MLspike ([Deneux et al., 2016](#)), a forward model, which fits an explicit model of the spike-to-calcium dynamics in order to find spike-evoked changes in the calcium signal. We emphasize that these methods were chosen as exemplars of their approaches, and are each innovative takes on the problem; we are not here critiquing individual methods, but using an array of methods to illustrate the problems and decisions facing the experimentalist when using calcium imaging data.

We first ask if spike deconvolution methods work well in principle. We fit the parameters of each method to a data-set of 21 ground-truth recordings ([Chen et al., 2013](#)), where the spiking activity of a cell is recorded simultaneously with 60 Hz  $\text{Ca}^{2+}$  imaging using high-signal-to-noise juxtacellular recording techniques (Figure ??a). To fit the parameters for each recording, we sweep each method's parameter space to find the parameter value(s) with the best match between the true and inferred spike train.

The best-fit parameters depend strongly on how we evaluate the match between true and inferred spikes. The Pearson correlation coefficient between the true and inferred spike train is a common choice ([Brown et al., 2004; Paiva et al., 2010; Theis et al., 2016; Reynolds et al., 2017; Berens et al., 2018](#)). However, we find that choosing parameters to maximise the correlation coefficient can create systematic errors, where the inferred spike



**Figure 1: Ground truth data analysis.**

- (a) Example simultaneous recording of somatic voltage and calcium activity imaged at 60Hz from (Chen et al., 2013). Spikes are marked with asterisks.
- (b) The correlation coefficient between true and inferred spikes as a function of the firing rate estimated from the inferred spikes. One line per recording. Spikes inferred here using Suite2P; each data-point on a line is a different value of Suite2P's threshold parameter. [Correct range] [Here and throughout: lose the box axes (upper and right)]
- (c) as in (b), but using Error Rate between the true and inferred spikes. [b-c to become last panels?]
- (d) Comparison of estimated firing rates for the optimised parameters against the true firing rate for each neuron. For each spike deconvolution method, we plot two estimated firing rates per recording (linked symbols), one for parameters optimised to maximise the correlation coefficient (PCC), the other for parameters optimised to minimise the error rate (ER).
- (e) Relative error in the estimate for true firing rate for all recordings, across all three methods. Horizontal black bars are means. Error is computed relative to the true firing rate:  $(Rate_{true} - Rate_{estimated})/Rate_{true}$ .
- (f) As for (e), but with the somatic calcium down-sampled to 7Hz before optimising parameters for the spike deconvolution methods.

84 trains from Suite2p and MLspike have too many spikes and those from LZero have too few  
85 spikes (Fig ??d-e, blue symbols). The consequent errors in estimating firing rates are large,  
86 being on average 79.47% for Suite2p, 31.72% for MLSpikes, and -21.14% for LZero (Fig  
87 ??e). We attribute these errors to the noisy relationship between the correlation coefficient  
88 and the number of inferred spikes, which we illustrate in Figure ??b for Suite2p. For many  
89 recordings, there is both an overall increase in correlation coefficient with the number of  
90 spikes, and no well-defined maximum coefficient, and thus the choice of deconvolution  
91 parameter(s) is highly sensitive.

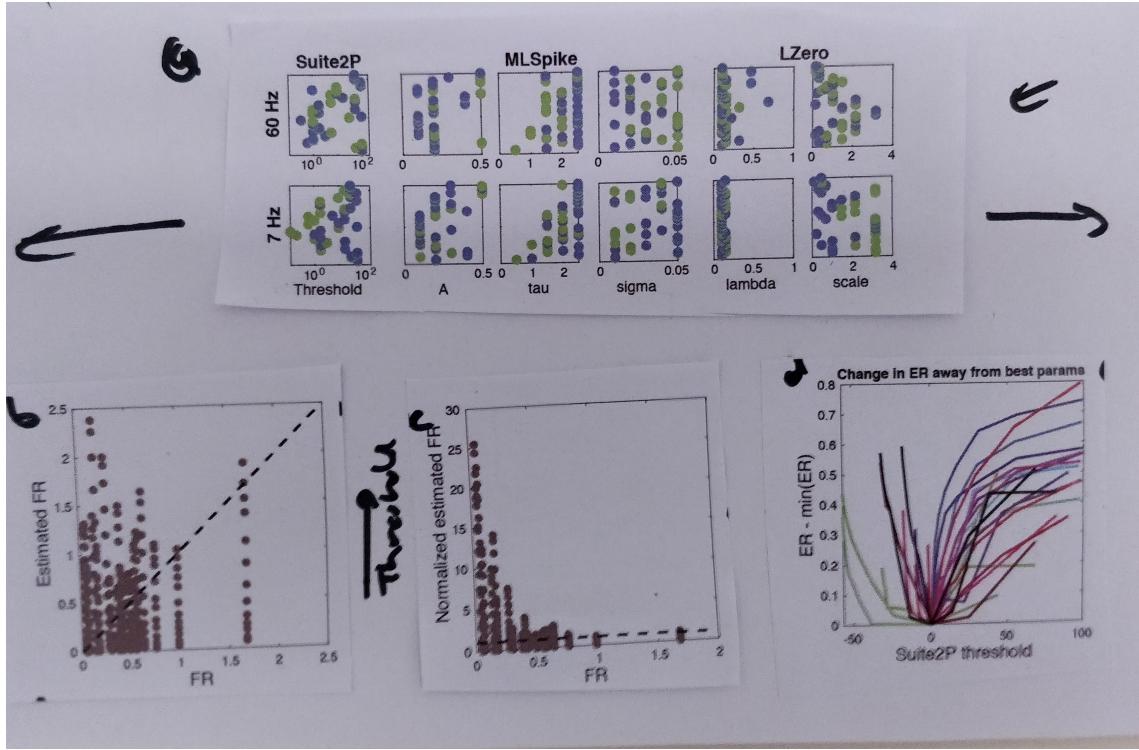
92 To address the weaknesses of the Pearson correlation coefficient, we instead optimise  
93 parameters using the Error Rate metric of Deneux et al. (2016). Error Rate returns a  
94 normalised score between 0 for a perfect match between two spike trains, and 1 when  
95 all the spikes are missed. Choosing parameters to minimise the Error Rate between the  
96 true and inferred spike-trains results in excellent recovery of the true number of spikes for  
97 all three deconvolution methods (Fig ??d-e, green symbols), with mean errors of 12% for  
98 Suite2P, 7.3% for MLSpikes, and 5% for LZero. As we show in Figure ??c for Suite2p, the  
99 Error Rate has a well-defined minima for almost every recording. Consequently, all spike  
100 deconvolution methods can, in principle, accurately recover the true spike-trains given an  
101 appropriate choice of parameters.

102 A potential caveat here is that the ground-truth data are single neurons imaged at a  
103 frame-rate of 60Hz, an order of magnitude greater than is typically achievable in popula-  
104 tion recordings (Peron et al., 2015a). Such a high frame-rate could allow for more accurate  
105 recovery of spikes than is possible in population recordings. To test this, we downsample  
106 the ground-truth data to a 7Hz frame-rate, and repeat the parameter sweeps for each  
107 deconvolution method applied to each recording. As we show in Figure ??f, optimising  
108 parameters using the minimum Error Rate still results in excellent recovery of the true  
109 spike rate (and interestingly for some recordings reduces the error when using the cor-  
110 relation coefficient). Lower frame-rates need not then be an impediment to using spike  
111 deconvolution methods.

## 112 2.2 Parameters optimised on ground-truth are widely distributed and 113 sensitive

114 What might be an impediment to using deconvolution methods on population record-  
115 ings is that the best parameter values vary widely between cells. Figure ??a-b plots the  
116 best-fit parameter values for each recording across spike deconvolution methods and sam-  
117 pling rates. Each method has at least one parameter with substantial variability across  
118 recordings, [SOMETHING HERE ON RANGE: are these ranges a high proportion of the  
119 total range of that parameter?]. This suggests that the best parameters for one cell may  
120 perform poorly for another cell.

121 The problem of between-cell variation in parameter values would be compensated  
122 somewhat if the quality of the inferred spike trains is robust to changes in those values.  
123 However, we find performance is highly sensitive to changes in some parameters. Figure  
124 ??b-c shows that the estimated firing rate can be out by up to a factor of 25 when varying  
125 Suite2p's threshold parameter across the best-fit range [CHECK: is this what was done?].  
126 Such firing rate errors are largest for low firing rate neurons (Fig. ??c), a particular  
127 concern for population recording data from the cortex of awake animals where low firing  
128 rates are the norm (O'Connor et al., 2010; Wohrer et al., 2013). Figure ??c shows that for  
129 most recordings the quality of the inferred spike train also abruptly worsens with small  
130 increases or decreases in the best parameter. [SOMETHING HERE ON THE OTHERS:  
131 MLSpike and LZero]. Thus using spike deconvolution algorithms on population recordings



**Figure 2: Variation in best-fit spike deconvolution parameters across ground-truth recordings.**

(a) Distributions of optimised parameter values across recordings. In each panel, we plot parameter values on the x-axis against the recording ID on the y-axis (in an arbitrary but consistent order). Parameter values are plotted for those optimised using the Error Rate (green). Top row: fits to the original 60 Hz frame-rate data; bottom row: fits to data down-sampled to 7 Hz. [1: DITCH Cxy] [2: link recordings of the same cell] [3: add density histogram to top of each plot, to show distribution of parameters]

(b) Example of variability in spike deconvolution performance with changes in a algorithm parameter. We plot here the estimated firing rates obtained for each recording as we vary the threshold parameter for Suite2p (between X and X). Each symbol is one tested parameter value, and each column is one recording, plotted against its true firing rate. The dashed black line indicates the correct estimated firing rate. All results in panels c-e are from Suite2P applied to downsampled data.

(c) as (b), but plotting the estimated firing rate as a multiple of the true firing rate (Estimated FR/true FR), to show how changes in the threshold parameter have a proportionally larger effect for lower-rate neurons.

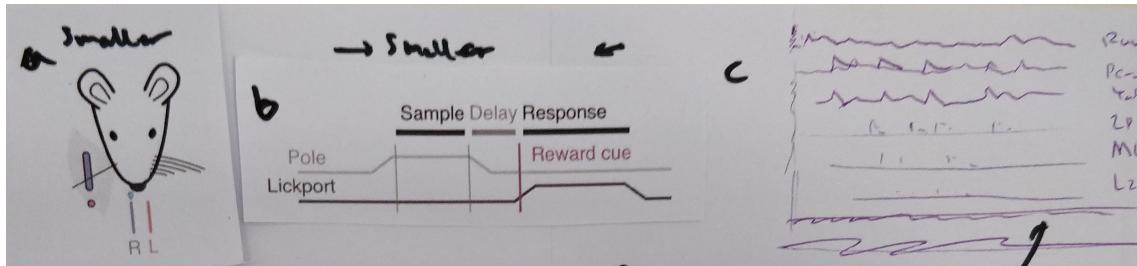
(d) Change in the error rate with change in the threshold parameters value away from it's optimum for each recording [label x-axis correctly "Change in threshold"]. One line per recording. (e) [Quantify (f) for all 3 methods, for all parameters: what is slope of ER change from minima with variation? [e.g. Take approx derivative]]

132 comes with potential issue that parameters can be both sensitive and vary considerably  
133 across cells.

### 134 2.3 Deconvolution of population imaging in barrel cortex during a de- 135 cision task

136 We turn now to seeing if and how these issues play out when analysing a large-scale  
137 population recording with no ground-truth. The data we use are two-photon calcium

138 imaging time-series from a head-fixed mouse performing a whisker-based two-alternative  
 139 decision task (Fig. 3a-b), from the study of (Peron et al., 2015a). We analyse here a single  
 140 session with 1552 simultaneously recorded pyramidal neurons in L2/3 of a single barrel  
 141 in somatosensory cortex, imaged at 7 Hz for just over 56 minutes, giving 23559 frames in  
 142 total across X trials of the task.



**Figure 3: Experimental data from (Peron et al., 2015a).**

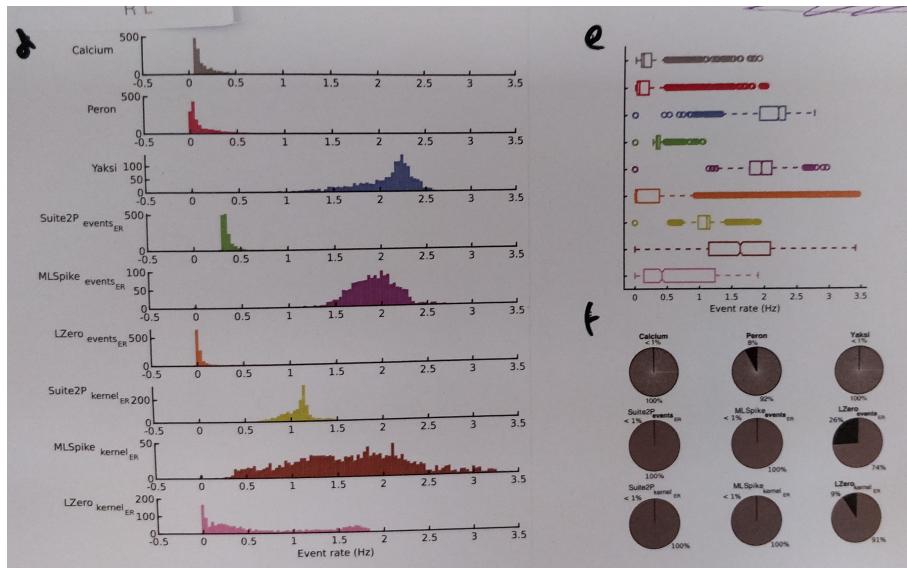
- (a) Schematic of task set-up. A pole was raised within range of the single left-hand whisker; its position, forward (red) or backward (purple) indicated whether reward would be available from the left or right lick-port. [Redraw]
- (b) Schematic of trial events. The pole was raised and lowered during the sample period; a [light?] cue indicated the start of the response period. [Redraw]
- (c) All deconvolution methods applied to one raw calcium signal from the same neuron.

143 Our primary goal is to understand how the choices of deconvolution of these calcium-  
 144 imaging data alter the scientific inferences we can draw. As our baseline, we use the  
 145 “raw”  $\Delta F/F$  time-series of changes in calcium indicator fluorescence. We use the three  
 146 chosen spike deconvolution methods to extract spike counts per frame. As examples of  
 147 simpler denoising methods, we explore a range of options from simple to complex: we use  
 148 Yaksci and Friedrich (2006)’s simple deconvolution of the raw calcium with a kernel of the  
 149 calcium response to a single spike (spike-response kernel); we use (Peron et al., 2015a)’s  
 150 own version of denoised calcium time-series, created by detecting calcium event onsets  
 151 by a threshold and convolving events with a spike-response kernel; and finally we create  
 152 denoised versions of the spike-deconvolution methods, by convolving their recovered spikes  
 153 with a spike-response kernel. Figure 3c show an example raw calcium time-series for one  
 154 neuron, and the result of applying each of these 8 processing methods. We thus repeat all  
 155 analyses on 9 different sets of time-series extracted from the same population recording.

156 We choose the algorithm parameters as follows. Yaksci and Friedrich (2006) is just  
 157 kernel of the calcium response to a single spike; here is GCAMP6s[?], so parameterise  
 158 kernel to that. For Peron’s own time-series, threshold was tuned to obtain calcium events  
 159 at suitable rate and sparseness across whole data-set [CHECK]; kernel was the same as  
 160 used for Yaksci [CHECK]. Three spike deconvolution methods, chose modal values of best-  
 161 fit parameters given by Error Rate; most consistently good values, and taken from L2/3  
 162 neurons in same species in similar primary sensory area (V1). [FILL IN DETAILS AND  
 163 TIDY]

## 164 2.4 Deconvolution methods disagree on estimates of simple neural statistics

166 We first check how well each approach recovers the basic statistics of neural activity event  
 167 rates in L2/3 of barrel cortex. Electrophysiology has shown that the distribution of firing  
 168 rates across neurons in a population is consistently long-tailed, and often log-normal, all  
 169 across rodent cortex (Wohrer et al., 2013); and L2/3 neurons in barrel cortex are no



**Figure 4: Estimates of population-wide event rates vary qualitatively across deconvolution methods.**

(a) The distribution of event rate per neuron across the recorded population, according to each deconvolution method. For raw calcium and the five denoising methods (upper 6 panels), events are detected as fluorescence transients greater in magnitude than three standard deviations of background noise. Spike deconvolution methods (lower 3 panels) return a spike count per time bin.

(b) Proportion of active (gray) and silent (black) cells for each method. Silent cells are defined following (Peron et al., 2015b) as those with an event rate less than 0.0083Hz.

170 different (O'Connor et al., 2010), with median firing rates less than 1 Hz, and a long right-  
171 hand tail of rarer high-firing neurons. We thus expect the calcium event rates (obtained  
172 by thresholding the calcium time-series) or spike rates from our time-series would follow  
173 such a distribution.

174 Figure 4a shows that the raw calcium and two of the spike deconvolution methods  
175 (Suite2p, LZero) have qualitatively correct distributions of event or spike rates (median  
176 near zero, long right-hand tails). The Peron time-series also have the correct distribution  
177 of event rates, which is unsurprising as it was tuned to do so. All other methods give qual-  
178 itatively wrong distributions of spike rates (MLSpike) or event rates (all other methods).  
179 There is also little overlap in the distributions of spike rates between the three spike de-  
180 convolution methods. Applying a kernel to their inferred spikes shifts rather than smooths  
181 the firing rate distributions (Suite2P<sub>kernel</sub>, MLSpike<sub>kernel</sub>, LZero<sub>kernel</sub>), suggesting noise  
182 in the spike inference process is amplified through the additional steps of convolution with  
183 a kernel and thresholding.

184 Cell-attached recordings in barrel cortex have shown that ~26% of L2/3 pyramidal  
185 cells are silent during a similar pole localisation task, with silence defined as emitting  
186 fewer than one spike every two minutes O'Connor et al. 2010. For the nine approaches  
187 we test here, six estimated the proportion of silent cells to be less than 1%, including two  
188 of the spike deconvolution methods (Figure 4c). For raw calcium and denoising methods,  
189 raising the threshold for defining events will lead to more silent cells, but at the cost of  
190 further shifting the event rate distributions towards zero. Even for simple firing statistics  
191 of neural activity, the choice of time-series gives widely differing, and sometimes wrong,  
192 results.

193 2.5 Inferences of single cell tuning differ widely between raw calcium  
194 and deconvolved methods

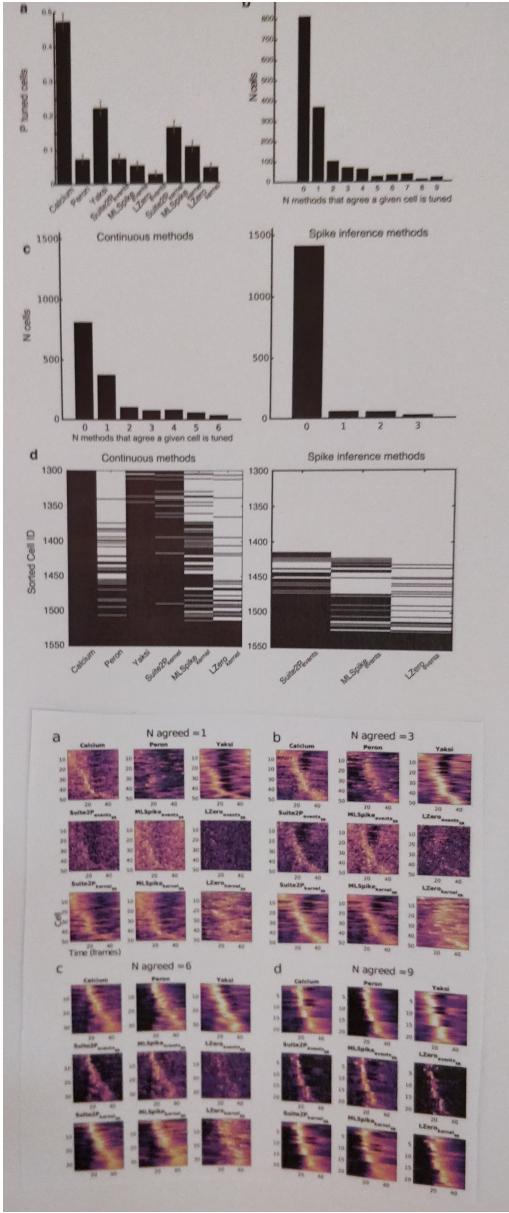
195 We turn now to what we can infer about simple properties of neural coding, and how our  
196 choice of deconvolution method can alter those inferences. The decision task facing the  
197 mouse (Fig. 3a) requires that it moves its whisker back-and-forth to detect the position  
198 of the pole, delay for a second after the pole is withdrawn, and then make a choice of  
199 the left or right lick-port based on the pole's position (Fig. 3b). As the imaged barrel  
200 corresponds to the single spared whisker (on the contralateral side of the face), so the  
201 captured population activity during each trial likely contains neurons tuned to different  
202 aspects of the task. We show here that the number and identity of such task-tuned neurons  
203 in the population differ widely between deconvolution methods.

204 Following Peron et al. 2015a, we define a task-tuned cell as one for which the peak  
205 in its trial-averaged histogram of activity exceeds the predicted upper limit from shuffled  
206 data (Fig.5a0; see Methods). When applied to the raw calcium time-series, close to half  
207 the neurons are tuned (Fig.5a). This is more than double the proportion found for the  
208 next nearest method (Yaksi's simple deconvolution), and at least a factor of 5 greater than  
209 the proportion of tuned neurons resulting from any spike deconvolution method, which  
210 each report less than 10% of the neurons are tuned.

211 Worse, few neurons are detected as tuned in time-series resulting from multiple meth-  
212 ods (Fig.5b). Only XXX neurons are labelled as tuned in at least two sets of time-series,  
213 and just 21 (X %) are labelled as tuned in all nine. Even separately considering the  
214 continuous and spike time-series, we find only 38 cells are tuned across all six continuous  
215 methods, and 25 neurons for all three spike deconvolution methods (Fig.5c). Figure 5d  
216 illustrates the diversity of detected tuning even amongst the neurons with the greatest  
217 agreement between methods.

218 These results suggest that raw calcium alone over-estimates tuning in the population,  
219 but also that there can be substantial disagreement between deconvolution methods. One  
220 solution for robust detection of tuned neurons is to find those agreed between the raw  
221 calcium time-series and more than one deconvolution method. In Figure 5e-h, we show how  
222 increasing the number of methods required to agree on a neuron's tuned status creates clear  
223 agreement between time-series processed with all methods, even if a particular method  
224 did not reach significance for that cell. Even requiring agreement between the raw calcium  
225 and just two other methods is enough to see tuning of many cells. The identification of  
226 unambiguously task-tuned cells could thus be achieved by triangulating the raw calcium  
227 with the output of multiple deconvolution methods.

228 In the pole detection task considered here, neurons tuned to pole contact are poten-  
229 tially crucial to understanding the sensory information used to make a decision. Touch  
230 onset is known to drive a subset of neurons to spike with short latency and low jitter  
231 (O'Connor et al., 2010; Hires et al., 2015). Detecting such rapid, precise responses in the  
232 slow kinetics of calcium imaging is challenging, suggesting spike-deconvolution methods  
233 might be necessary to detect touch-tuned neurons. To test this, in each of the 9 sets of  
234 time-series we identify touch-tuned neurons by a significant peak in their tough-triggered  
235 activity (Fig ??a). Figure 6b shows that, while all data-sets have touch-tuned neurons,  
236 the number of such neurons differs substantially between them. And rather than being  
237 essential, spike deconvolution methods disagree strongly on touch-tuning, with MLSpike  
238 (events) finding 45 touch-tuned neurons and LZero (events) finding one. Thus our infer-  
239 ences of the coding of task-wide or specific sensory events crucially depends on our choice  
240 of calcium imaging time-series.



**Figure 5: Inferences of single cell tuning show poor agreement between raw calcium and deconvolution methods, and between methods.**

(a0) Examples of a tuned (left) and non-tuned (right) cell from the raw calcium time-series. X: Data; Y: upper 95% interval from shuffled data.

(a) Number of tuned cells per deconvolution method. Error bars are 95% binomial confidence intervals.

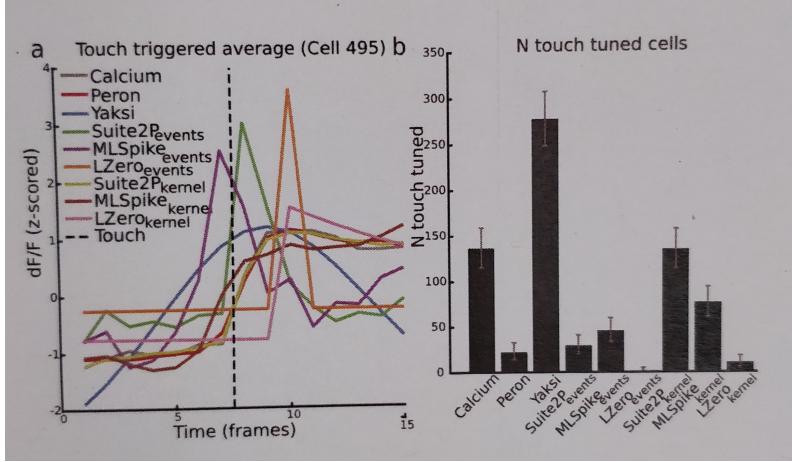
(b) Agreement between methods. For each neuron, we count the number of methods (including raw calcium) for which it is labelled as tuned. Bars show the number of cells classified as tuned by exactly  $N$  methods.

(c) Similar to (b), but breaking down the cells into: agreement between methods (raw or denoising) resulting in continuous signals (left panel); and agreement between spike deconvolution methods (right panel).

(d) Comparison of cell tuning across methods. Each row shows whether that cell is tuned (black) or not (white) under that deconvolution method. Cells are ordered from bottom to top by the number of methods that classify that cell as tuned.

(e-h) Identifying robust cell tuning. Panel groups (e) to (h) show cells classed as tuned by increasing numbers of deconvolution methods. Each panel within a group plots one cell's normalised (z-scored) trial-average histogram per row, ordered by the time of peak activity. The first panel in a group of 9 shows histograms from raw calcium signals; each of the 8 subsequent panel shows trial-average histograms for the same cells, but following processing by each of the eight deconvolution methods.

[Q: How were the neurons chosen?]



**Figure 6: Touch-triggered neuron responses.**

(a) Touch-triggered average activity from one neuron, across all deconvolution methods. The dotted line is the imaging frame in which the whisker touched the pole.

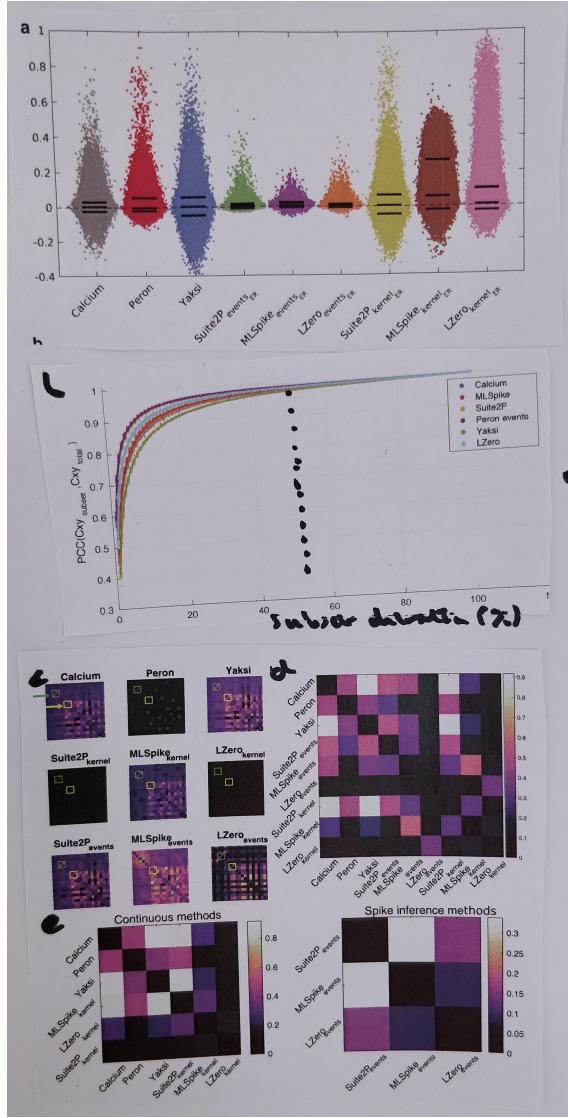
(b) Number of touch-tuned cells across deconvolution methods. A cell is classed as touch-tuned if its peak touch-triggered activity is significantly greater than shuffled data. Error bars are Jeffreys intervals

**241 2.6 Inconsistent recovery of population correlation structure across de-  
242 convolution approaches**

243 The high yield of neurons from calcium imaging is ideal for studying the dynamics and  
244 coding of neural populations [cite some examples here]. Many analyses of populations start  
245 from pairwise correlations between cells, whether as measures of a population’s synchrony  
246 or joint activity, or as a basis for further analyses like clustering and dimension reduction  
247 [Cunningham]. We now show how our inferences of population correlation structure also  
248 depend strongly on the choice of deconvolution method.

249 Figure 7a shows that the distributions of pairwise correlations qualitatively differ be-  
250 tween the sets of time-series we derived from the same calcium imaging data. Some  
251 distributions are approximately symmetric, with broad tails; some asymmetric with nar-  
252 row tails; the correlation distribution from the Peron method time-series is the only one  
253 with a median below zero. These qualitative differences are not due to noisy estimates  
254 of the pairwise correlations: for all our sets of time-series the correlations computed on  
255 a sub-set of time-points in the session agree well with the correlations computed on the  
256 whole session (Figure 7b). Thus pairwise correlation estimates for each method are stable,  
257 but their distributions differ between methods.

258 Looking in detail at the full correlation matrix shows that even for methods with simi-  
259 lar distributions, their agreement on correlation structure is poor. Some neuron pairs that  
260 appear correlated from time-series processed by one deconvolution method are uncorre-  
261 lated when processed with another method (Figure 7c). Over the whole population, the  
262 correlation structure obtained from the raw calcium, Yaksi and Suite2p (kernel) time-series  
263 all closely agree, but nothing else does (Figure 7d): the correlation structure obtained from  
264 LZero agrees with nothing else; and the spike deconvolution methods all generate dissi-  
265 milar correlation structures (Figure 7e). Our inferences about the extent and identity  
266 of correlations within the population will differ qualitatively depending on our choice of  
267 imaging time-series.



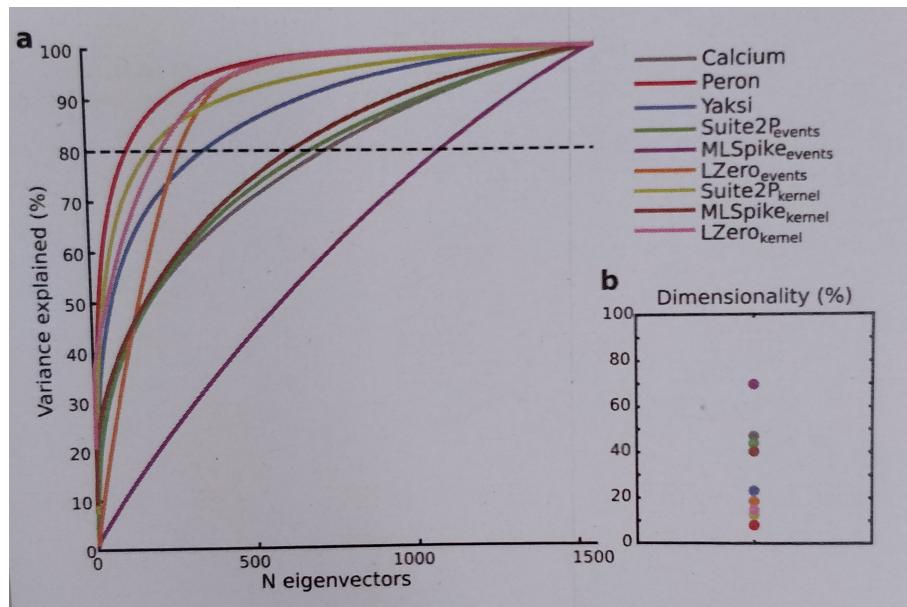
**Figure 7: Effects of deconvolution on pairwise correlations between neurons.**

- (a) Distributions of pairwise correlations between all cells, for each deconvolution method (one dot per cell pair, x-axis jitter added for clarity). Solid black lines are 5th, 50th and 95th percentiles.
- (b) Stability of correlation structure in the population. We quantify here the stability of the pairwise correlation estimates, by comparing the correlation matrix constructed on the full data ( $C_{xy\text{total}}$ ) to the same matrix constructed on a subset of the data ( $C_{xy\text{subset}}$ ). Each data-point is the correlation between  $C_{xy\text{total}}$  and  $C_{xy\text{subset}}$ ; one line per deconvolution method.
- (c) Examples of qualitatively differing correlation structure across methods. Each panel plots the pairwise correlations for the same 50 neurons on the same colour scale. As examples, we highlight two pairs of cells: one consistently correlated across different methods (green arrow and boxes); the other not (yellow arrow and boxes).
- (d) Comparison of pairwise correlation matrices between deconvolution methods. Each square is the correlation between the full-data correlation matrix for that pair of methods. [What is color-scale, Pearson's?]
- (e) as in (d), but split to show continuous methods (left) or spike deconvolution methods (right). [Check all labels: clearly something inconsistent between (d) and (e)]

268 **2.7 Deconvolution methods show the same population activity is both**  
 269 **low and high dimensional**

270 Dimensionality reduction techniques, like principal components analysis (PCA), allow re-  
 271 searchers to make sense of large scale neuroscience data ([Cunningham and Yu, 2014](#)) [more  
 272 here], by reducing the data from  $N$  neurons to  $d < N$  dimensions. Key to such analyses  
 273 is the choice of  $d$ , a choice guided by how much of the original data we can capture. To  
 274 assess such inferences of population dimensionality, we apply PCA to our 9 sets of imaging  
 275 time-series to estimate the dimensionality of the imaging data (which for PCA is the  
 276 variance explained by each eigenvector of the data's covariance matrix).

277 Figure 8a plots for each deconvolution method the cumulative variance explained when  
 278 increasing the number of retained dimensions. Most deconvolution methods qualitatively  
 279 disagree with the raw calcium data-set on the relationship between dimensions and vari-  
 280 ance. This relationship is also inconsistent across deconvolution methods; indeed the spike  
 281 deconvolution methods result in the shallowest ( $\text{MLSpike}_{\text{events}}$ ) and amongst the steepest  
 282 ( $\text{LZero}_{\text{events}}$ ) relationships between increasing dimensions and variance explained. The  
 283 number of dimensions required to explain 80% of the variance in the data ranges from  
 284  $d = 125$  (Peron) to  $d = 1081$  ( $\text{MLSpike}_{\text{events}}$ ), a jump from 8% to 70% of all possible  
 285 dimensions (Fig 8b). Thus we could equally infer that the same L2/3 population activity  
 286 is low dimensional (<10% dimensions required to explain 80% of the variance) or high-  
 287 dimensional (>50% of dimensions required) depending on our choice of imaging time-series.



**Figure 8: Dimensionality of population activity.**

(a) Cumulative variance explained by each dimension of the data's covariance matrix, one line per deconvolution method. Dimensions are obtained from principal components analysis, and are ordered by decreasing contribution to the total variance explained. Dashed line is the 80% threshold used in panel (b).

(b) Proportion of dimensions required to explain 80% of the variance in the data.

288 **3 Discussion**

289 Imaging of somatic calcium is for capturing the simultaneous activity of hundreds to thou-  
290 sands of neurons. But the time-series of each neuron's calcium fluorescence is inherently  
291 noisy and non-linearly related to its spiking. We sought here to address how our choice of  
292 corrections to these time-series – to use them raw, denoise them or deconvolve them into  
293 spikes – affect the quality and reliability of the scientific inferences drawn.

294 Our results show the choice qualitatively changes the potential scientific inferences  
295 we draw about neural activity, coding, and correlation structure. We consistently ob-  
296 serve that the analysis results differ sharply between the raw calcium and most, if not  
297 all, of the processed time-series. However, the deconvolved time-series also consistently  
298 disagreed with each other, even between methods of the same broad class (denoising or  
299 spike deconvolution).

300 **3.1 Accurate spike deconvolution is possible, but sensitive**

301 We find much that is encouraging. In fitting spike deconvolution methods to ground-  
302 truth data, we found they can in principle accurately recover neural activity. A caveat  
303 here is that the choice of metric for evaluation and fitting of parameters is of critical  
304 importance. The widely-used Pearson correlation coefficient is a poor choice of metric  
305 as it returns inconsistent results with small changes in algorithm parameters, and leads  
306 to poor estimates of simple measures such as firing rate when used across methods and  
307 sampling rates. By contrast, the Error Rate metric ([Deneux et al., 2016](#); [Victor and Purpura, 1996](#)) resulted in excellent recovery of ground-truth spike trains. Other recently  
308 developed methods for comparing spike-trains based on information theory ([Theis et al., 2016](#))  
309 or fuzzy set theory ([Reynolds et al., 2017](#)), may also be appropriate.

310 However, while good estimates of neural activity can be achieved with modern spike  
311 deconvolution methods [cf Spikefinder, Pitcharriu 2018 JNS], the best parameters vary  
312 substantially between cells, and small changes in analysis parameters result in poor spike  
313 inference performance. This variation and sensitivity of parameters played out as widely-  
314 differing results between the three spike deconvolution methods in analyses of neural ac-  
315 tivity, coding, and correlation structure.

317 **3.2 Choosing parameters for deconvolution methods**

318 A potential limitation of our study is that we use a single set of parameter values for each  
319 spike deconvolution method applied to the population imaging data from barrel cortex.  
320 But then our situation is the same as that facing any experimentalist: in the absence of  
321 ground-truth, how do we set the parameters? Our solution here was to use the modal  
322 parameter values from ground-truth fitting, as these values are candidates for the most  
323 general solutions. We also felt these were a reasonable choice for the population imaging  
324 data from barrel cortex, given that the ground-truth recordings came from the same species  
325 (mouse) in the same layer (2/3) of a different bit of primary sensory cortex (V1).

326 Rather than use the most general parameters values, another solution would be to  
327 tune the parameters to obtain known gross statistics of the neural activity. This was the  
328 approach used by Peron and colleagues [cite] to obtain the denoised Peron time-series we  
329 included here. But as we've seen, this approach can lead to its own problems: for example,  
330 in the Peron time-series, it created a distributions of correlations that differed from any  
331 other set of time-series. Indeed, finding good parameter values may be an intractable  
332 problem, as it is possible each neuron requires individual fitting, to reflect the combination

333 of its expression of fluorescent protein, and its particular non-linearity between voltage and  
334 calcium.

### 335 3.3 Ways forward

336 The simplest solution to the inconsistencies between different forms of time-series is to  
337 triangulate them, and take the consensus across their results. For example, our finding of  
338 a set of tuned neurons across multiple methods is strong evidence that neurons in L2/3  
339 of barrel cortex are responsive across the stages of the decision task. Further examples  
340 of such triangulation in the literature are rare; Klaus and colleagues (Klaus et al., 2017)  
341 used two different pipelines to derive raw  $\Delta F/F$  of individual neurons from one-photon  
342 fibre-optic recordings in the striatum, and replicated all analyses using the output of both  
343 pipelines. Our results encourage the further use of triangulation to create robust inference:  
344 obtaining the same result in the face of wide variation increases our belief in its reliability  
345 (Munaf and Davey Smith, 2018).

346 There are caveats to using triangulation. For single neuron analyses, triangulation  
347 inevitably comes at the price of reducing the yield of neurons to which we can confidently  
348 assign roles. A further problem for triangulation is how to combine more complex analyses,  
349 such as pairwise correlations; the alternative is to rely on qualitative comparisons.

350 Many studies use the raw calcium signal as the basis for all their analyses [cite], perhaps  
351 assuming this is the least biased approach. Our results show this is not so: the discrepancy  
352 between raw and deconvolved calcium on single neuron coding suggests an extraordinary  
353 range of possible results, from about half of all neurons tuned to the task down to less  
354 than 5 percent. The qualitative conclusion – there is coding – is not satisfactory. Thus our  
355 results should not be interpreted as a call to abandon deconvolution methods; rather they  
356 serve to delimit how we can interpret their outputs.

357 Instead, we need deconvolution solved: as sensors with faster kinetics (though funda-  
358 mentally limited by kinetics of calcium release itself) and higher signal-to-noise ratios are  
359 developed [cite GCAMP7; red-shifted], so the accuracy and robustness of de-noising and  
360 deconvolution should improve; and as the neuron yield continues to increase [Ahrens], so  
361 the potential for insights from inferred spikes grows. Developing further advanced decon-  
362 volution algorithms will harness these advances, but are potentially always limited by the  
363 lack of ground-truth to fit their parameters. Our results may provide impetus for a differ-  
364 ent direction of research, focussing on how we can get the output of different algorithms  
365 to agree, and thus provide robust scientific inferences about neural populations.

366 **4 Methods**

367 **Ground truth data**

368 Ground truth data was accessed from [crcns.org](http://crcns.org) ([Svoboda, 2015](#)), and the experiments have  
369 been described previously [Chen et al. \(2013\)](#). Briefly, mouse visual cortical neurons ex-  
370 pressing the fluorescent calcium reporter protein GCamP6s were imaged with two-photon  
371 microscopy at 60Hz. Loose-seal cell-attached recordings were performed simultaneously  
372 at 10kHz. The data-set contains twenty one recordings from nine cells.

373 **Spike train metrics**

374 Pearson correlation coefficient was computed between the ground truth and inferred spike  
375 trains following convolution with a gaussian kernel (61 samples wide, 1.02 seconds).

376 Error Rate was computed between the ground truth and inferred spike trains using  
377 the [Deneux et al. \(2016\)](#), implementation of normalised error rate, derived from [Victor](#)  
378 and [Purpura \(1996\)](#) Error Rate (code available <https://github.com/MLspike>). Briefly, the  
379 error rate is  $1 - F1\text{-score}$ , where the  $F1\text{-score}$  is the harmonic mean of sensitivity (number  
380 of missed spikes divided by total spikes) and precision (number of falsely detected spikes  
381 divided by total detected spikes),

[write this in full, defining sensitivity and precision]

$$\text{ErrorRate} = 1 - 2 \frac{\text{sensitivity} \times \text{precision}}{\text{sensitivity} + \text{precision}}.$$

382 Hits, misses and false detections were counted with a temporal precision of 0.5 seconds.  
383 [Firing rate: how computed for true and inferred?]

384 For normalised estimation of errors in firing rate, we computed (estimated FR - true  
385 FR / true FR)

386 **Parameter fitting**

387 For each method the best parameters for each cell were determined by brute force search  
388 over an appropriate range. [Give the ranges. How was “appropriate” defined?]

389 The modal best parameters, as determined using Error Rate on downsampled data,  
390 were then fixed for the population imaging data analysis.

391 In additional tests (*TO DO Supplemental figs, subscript ‘PCC’*), the modal best pa-  
392 rameters when assessed using correlation coefficient were used, and in the case of ML Spike  
393 parameters were additionally hand-tuned using a built-in GUI (*TO DO Supplemental figs,*  
394 *subscript ‘hand’*).

395 **Downsampling**

396 Ground truth calcium data was downsampled from 60Hz to 7Hz in Matlab by up-sampling  
397 by 7 - `interp(ca,7)` and downsampling the resultant signal by 60, as Matlab’s downsam-  
398 pling must be done in integer steps [??? how was the downsampling done then? By taking  
399 every Nth frame from the interpolated data? Or averaging over every approx 8.5 frames  
400 of the 60Hz signal?].

401 **Population imaging data description**

402 Population imaging data was accessed from [crcns.org](http://crcns.org) and have been described previously  
403 (Peron et al., 2015b). Briefly, volumetric two photon calcium imaging of primary  
404 somatosensory cortex (S1) was performed in awake head-fixed mice performing a whisker-  
405 based object localisation task. In the task a metal pole was presented on one of two loca-  
406 tions and mice were motivated with fluid reward to lick at one of two lick ports depending  
407 on the location of the pole following a brief delay. Two photon imaging of GCaMP6s  
408 expressing neurons in superficial S1 was performed at 7Hz. Images were motion corrected  
409 and aligned, before regions of interest were manually set and neuropil-subtracted. A single  
410 recording from this dataset was used for population analysis. The example session had  
411 1552 neurons recorded for a total of 23559 frames (56 minutes).

412 **4.1 Event rate estimation**

413 Spike inference methods ( $\text{Suite2P}_{events}$ ,  $\text{MLSpike}_{events}$ ,  $\text{LZero}_{events}$ ) return estimated spikes  
414 per frame which were converted into mean event rates (Hz) per cell. Event rate for contin-  
415 uous methods (Calcium, Peron, Yaksi,  $\text{Suite2P}_{kernel}$ ,  $\text{MLSpike}_{kernel}$ ,  $\text{LZero}_{kernel}$ ) for each  
416 cell was determined by counting activity/fluorescence events greater [what was greater?] than  
417 three standard deviations of the background noise. Background noise was calculated  
418 by taking a four-point moving average [missing object in sentence: average of what?] and  
419 subtracting this [what is "this"?] from the activity/fluorescence trace. Event rate was  
420 then computed in Hz.

421 [what does this all mean? That each data-point for the background noise was an  
422 average over 4 adjacent frames, and the standard deviation of the noise was computed  
423 from those data-points? Why smooth the noise? And what segment of the data was  
424 treated as "background"? (i.e. how many data-points)? Or does this mean that the whole  
425  $\text{Ca}^{2+}$  trace for each cell was smoothed using a 4-frame average (shifting 1 frame?), and  
426 the SD of that smoothed signal was used as an estimate of background noise?]

427 Silent cells were defined as cells with event rates below 0.0083Hz (or fewer than one  
428 spike per two minutes of recording) as in (O'Connor et al., 2010).

429 **4.2 Task-tuned cells**

430 Task-tuning was determined for each neuron using the model-free approach of Peron et al.  
431 (2015b). Neurons were classed as task-tuned if their peak trial-average activity exceeded  
432 the 95th percentile of a distribution of trial-average peaks from shuffled data (10000 shuf-  
433 fles). The shuffle test was done separately for correct lick-left and lick-right trials and cells  
434 satisfying the tuning criteria in either case were counted as task-tuned.

435 Tuned cell agreement was calculated as the number of methods that agreed to the  
436 tuning status of a given cell, for all methods and separately for continuous and spike  
437 inference methods.

438 **4.3 Touch-related responses**

439 Touch-tuned cells were determined by computing touch-triggered average activity for each  
440 cell, before calculating whether the data distribution of peak touch-induced activity ex-  
441 ceeds the expected activity of shuffled data. In more detail, the time of first touch -  
442 between the mouse's whisker and the metal pole - on each trial was recorded. For each  
443 touch time, one second of activity (seven data samples) was extracted before and after the  
444 frame closest to touch (15 samples total); taking the mean of these gave the average touch

445 response for the cell. The time of peak touch-triggered average activity was calculated,  
446 and a ranksum test (bonferroni corrected) between the true data distribution at peak time  
447 and a matched random sample of data from the same cell.

448 [Figure legend said: (Mann-Whitney U test, Benjamini Hochberg corrected); which is  
449 it? And what is  $\alpha$  here?]

450 [unclear what the pairwise tests were between; the mean activity at the peak time, and  
451 a N-length vector of mean activity as the same time obtained from N shuffled datasets?]

452 [Shuffled how many times? Shuffled how?].

#### 453 4.4 Pairwise correlations

454 Pairwise correlations were calculated for all pairs of neurons in Matlab (corrcoef) at the  
455 data sampling rate (7Hz). Correlations between correlation matrices (Fig. ??) were com-  
456 puted between the unique pairwise correlations from each method (i.e. `CXY(find(triu(CXY)))`).

457 For spike-inference methods, how was this done? On spike-count vectors?

458 Stability of correlation. For each deconvolution method, we computed the pairwise  
459 correlation matrix using the entire sessions data, as above. We then sampled a subset of  
460 time-points (1%-100%) of the full dataset at random without replacement and compute  
461 its matrix of pairwise correlations. We compute the similarity between the [Total] and  
462 [Subset] matrices using Pearsons correlation coefficient. [Finish this])

#### 463 4.5 Dimensionality

464 To determine the dimensionality of each dataset we performed eigendecomposition of the  
465 covariance matrix of each dataset. Computed as per the pairwise correlations above?  
466 The resultant eigenvalues were sorted into descending order, and the variance explained  
467 (`cumsum(egs)/sum(egs)`) plotted.

#### 468 List of deconvolution methods

##### 469 Suite2P

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

475

476 [Explain: the basic method, and its free parameters]

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

478

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

480

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

##### 482 ML Spike

483 ML Spike (<https://github.com/mlspike>) was developed by Thomas Deneux at INT, CRNS  
484 Marseille, France. A model-based probabilistic approach, ML Spike was developed to re-  
485 cover spike trains in calcium imaging data by taking baseline fluctuations and cellular

486 properties into account. A comprehensive explanation of the algorithm and its benefits  
487 can be found in the paper ([Deneux et al., 2016](#)).

488 [Explain: the basic method, and its free parameters]

489 ML Spike can return a maximum a posteriori spike train, or a spike probability per  
490 time step. (*TO DO: We show results for both denoted ML Spike<sub>events</sub> and ML Spike<sub>pspike</sub>  
491 in Supplement*)

## 492 **LZero**

493 The method we refer to as LZero was developed by Sean Jewell and Daniela Witten  
494 from U.Washington, Seattle, USA. The goal for this implementation was to cast spike  
495 detection as a change-point detection problem, which could be solved with an existing  
496  $l_0$  optimization algorithm. In their paper Jewell and Witten show that the  $l_0$  solution is  
497 better than previously implemented  $l_2$  solutions, with results much closer to the real spike  
498 train ( $l_2$  solutions tend to overestimate the true firing rate). Details can be found in the  
499 paper ([Jewell and Witten, 2017](#)).

500 [Explain: the basic method, and its free parameters]

501 Link: <https://arxiv.org/abs/1703.08644>

## 502 **Yaksi**

503 Yaksi refers to the ‘vanilla’ deconvolution of Yaksi and Friedrich (2006). This is to be used  
504 as a baseline for comparison with more sophisticated methods. The method is detailed in  
505 the paper: ([Yaksi and Friedrich, 2006](#)).

506 [Explain: the basic method]

## 507 **Peron events**

508 Peron events refer to the extracted events detailed in the original [Peron et al. \(2015b\)](#)  
509 paper. It is a version of the ‘peeling’ algorithm ([Lütcke et al., 2013](#)) tuned to generate a  
510 low number of false positive detections (a rate of 0.01Hz) on ground truth data, leading  
511 to a hit rate of 54%.

512 [Explain: the basic method, and its free parameters]

## 513 **Events + kernel versions**

514 Where a spike inference method returns spike rates per time point, these are plotted as  
515 Method<sub>events</sub>. To compare to other methods that return a de-noised dF/F or firing rate  
516 estimates, these events are convolved with a calcium kernel and plotted as Method<sub>kernel</sub>.

517 [Explain: what are the parameters of the kernel]

## 518 **References**

519 Philipp Berens, Jeremy Freeman, Thomas Deneux, Nicolay Chenkov, Thomas McColgan,  
520 Artur Speiser, Jakob H Macke, Srinivas C Turaga, Patrick Mineault, Peter Rupprecht,  
521 Stephan Gerhard, Rainer W Friedrich, Johannes Friedrich, Liam Paninski, Marius Pa-  
522 chitariu, Kenneth D Harris, Ben Bolte, Timothy A Machado, Dario Ringach, Jasmine  
523 Stone, Luke E Rogerson, Nicolas J Sofroniew, Jacob Reimer, Emmanouil Froudarakis,  
524 Thomas Euler, Miroslav Román Rosón, Lucas Theis, Andreas S Tolias, and Matthias  
525 Bethge. Community-based benchmarking improves spike rate inference from two-photon  
526 calcium imaging data. *PLoS Comput. Biol.*, 14(5):e1006157, May 2018.

- 527 Emery N Brown, Robert E Kass, and Partha P Mitra. Multiple neural spike train data  
528 analysis: state-of-the-art and future challenges. *Nat. Neurosci.*, 7(5):456–461, May 2004.
- 529 Tsai-Wen Chen, Trevor J Wardill, Yi Sun, Stefan R Pulver, Sabine L Renninger, Amy  
530 Baohan, Eric R Schreiter, Rex A Kerr, Michael B Orger, Vivek Jayaraman, Loren L  
531 Looger, Karel Svoboda, and Douglas S Kim. Ultrasensitive fluorescent proteins for  
532 imaging neuronal activity. *Nature*, 499(7458):295–300, July 2013.
- 533 John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural  
534 recordings. *Nat. Neurosci.*, 17(11):1500–1509, November 2014.
- 535 Thomas Deneux, Attila Kaszas, Gergely Szalay, Gergely Katona, Tamás Lakner, Amiram  
536 Grinvald, Balázs Rózsa, and Ivo Vanzetta. Accurate spike estimation from noisy calcium  
537 signals for ultrafast three-dimensional imaging of large neuronal populations in vivo.  
538 *Nat. Commun.*, 7:12190, July 2016.
- 539 Samuel Andrew Hires, Diego A Gutnisky, Jianing Yu, Daniel H O’Connor, and Karel  
540 Svoboda. Low-noise encoding of active touch by layer 4 in the somatosensory cortex.  
541 *Elife*, 4, August 2015.
- 542 Sean Jewell and Daniela Witten. Exact spike train inference via  $\ell_0$  optimization. March  
543 2017.
- 544 Andreas Klaus, Gabriela J Martins, Vitor B Paixao, Pengcheng Zhou, Liam Paninski, and  
545 Rui M Costa. The spatiotemporal organization of the striatum encodes action space.  
546 *Neuron*, 95:1171–1180.e7, Aug 2017. ISSN 1097-4199. doi: 10.1016/j.neuron.2017.08.  
547 015.
- 548 Henry Lütcke, Felipe Gerhard, Friedemann Zenke, Wulfram Gerstner, and Fritjof Helm-  
549 chen. Inference of neuronal network spike dynamics and topology from calcium imaging  
550 data. *Front. Neural Circuits*, 7:201, December 2013.
- 551 Marcus R Munaf and George Davey Smith. Robust research needs many lines of evidence.  
552 *Nature*, 553:399–401, January 2018. ISSN 1476-4687. doi: 10.1038/d41586-018-01023-3.
- 553 Daniel H O’Connor, Simon P Peron, Daniel Huber, and Karel Svoboda. Neural activity  
554 in barrel cortex underlying vibrissa-based object localization in mice. *Neuron*, 67(6):  
555 1048–1061, September 2010.
- 556 Marius Pachitariu, Carsen Stringer, Sylvia Schröder, Mario Dipoppa, L Federico Rossi,  
557 Matteo Carandini, and Kenneth D Harris. Suite2p: beyond 10, 000 neurons with stan-  
558 dard two-photon microscopy.
- 559 António R C Paiva, Il Park, and José C Príncipe. A comparison of binless spike train  
560 measures. *Neural Comput. Appl.*, 19(3):405–419, April 2010.
- 561 Simon Peron, Tsai-Wen Chen, and Karel Svoboda. Comprehensive imaging of cortical  
562 networks. *Curr. Opin. Neurobiol.*, 32:115–123, June 2015a.
- 563 Simon P Peron, Jeremy Freeman, Vijay Iyer, Caiying Guo, and Karel Svoboda. A cellular  
564 resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783–799,  
565 May 2015b.
- 566 Stephanie Reynolds, Simon R Schultz, and Pier Luigi Dragotti. CosMIC: A consistent  
567 metric for spike inference from calcium imaging. December 2017.

- 568 K Svoboda. Simultaneous imaging and loose-seal cell-attached electrical recordings from  
569 neurons expressing a variety of genetically encoded calcium indicators. *GENIE project,*  
570 *Janelia Farm Campus, HHMI; CRCNS.org*, 2015.
- 571 Lucas Theis, Philipp Berens, Emmanouil Froudarakis, Jacob Reimer, Miroslav  
572 Román Rosón, Tom Baden, Thomas Euler, Andreas S Tolias, and Matthias Bethge.  
573 Benchmarking spike rate inference in population calcium imaging. *Neuron*, 90(3):471–  
574 482, May 2016.
- 575 J D Victor and K P Purpura. Nature and precision of temporal coding in visual cortex:  
576 a metric-space analysis. *J. Neurophysiol.*, 76(2):1310–1326, August 1996.
- 577 Adrien Wohrer, Mark D Humphries, and Christian K Machens. Population-wide distri-  
578 butions of neural activity during perceptual decision-making. *Prog. Neurobiol.*, 103:  
579 156–193, April 2013.
- 580 Emre Yaksi and Rainer W Friedrich. Reconstruction of firing rate changes across neuronal  
581 populations by temporally deconvolved ca<sub>2+</sub> imaging. *Nat. Methods*, 3(5):377–383, May  
582 2006.