Title
To deconvolve, or not to deconvolve: Inferences of neuronal activities using calcium imaging data.

Authors
Shen, Tong
Lur, Gyorgy
Xu, Xiangmin
et al.

Publication Date
2022

DOI
10.1016/j.jneumeth.2021.109431

HHS Public Access

Author manuscript

J Neurosci Methods. Author manuscript; available in PMC 2022 September 17.

Published in final edited form as:

J Neurosci Methods. 2022 January 15; 366: 109431. doi:10.1016/j.jneumeth.2021.109431.

To deconvolve, or not to deconvolve: Inferences of neuronal activities using calcium imaging data

Tong Shen[a], **Gyorgy Lur**[b], **Xiangmin Xu**[c,d,e,f,g], **Zhaoxia Yu**[a,*]

[a] Department of Statistics, University of California, Irvine, CA 92697, USA

[b] Department of Neurobiology and Behavior, University of California, Irvine, CA 92697, USA

[c] Department of Anatomy and Neurobiology, School of Medicine, University of California, Irvine, CA 92697-1275, USA

[d] The Center for Neural Circuit Mapping, University of California, Irvine, CA 92697, USA

[e] Department of Biomedical Engineering, University of California, Irvine, CA 92697-2715, USA

[f] Department of Microbiology and Molecular Genetics, University of California, Irvine, CA 92697-4025, USA

[g] Department of Computer Science, University of California, Irvine, CA 92697-3435, USA

Abstract

Background: With the increasing popularity of calcium imaging in neuroscience research, choosing the right methods to analyze calcium imaging data is critical to address various scientific questions. Unlike spike trains measured using electrodes, fluorescence intensity traces provide an indirect and noisy measurement of the underlying neuronal activities. The observed calcium traces are either analyzed directly or deconvolved to spike trains to infer neuronal activities. When both approaches are applicable, it is unclear whether deconvolving calcium traces is a necessary step.

Methods: In this article, we compare the performance of using calcium traces or their deconvolved spike trains for three common analyses: clustering, principal component analysis (PCA), and population decoding.

Results: We found that (1) the two approaches lead to diverging results; (2) estimated spike trains, when smoothed or binned appropriately, usually lead to satisfactory performances, such as more accurate estimation of cluster membership; (3) although estimate spike train produce results more similar to true spike data than trace data, we found that the PCA results from trace data might better reflect the underlying neuronal ensembles (clusters); and (4) for both approaches, decobability can be improved by using denoising or smoothing methods.

* Corresponding author. zhaoxia@ics.uci.edu (Z. Yu).

Competing Interests Statement

The authors (Tong Shen, Gyorgy Lur, Xiangmin Xu, and Zhaoxia Yu) declare that there is no conflict of interest.

CRediT authorship contribution statement

Tong Shen: Software, Visualization, Validation, Writing – original draft, Writing – review & editing. **Xiangmin Xu:** Conceptualization, Supervision, Writing – review & editing. **Gyorgy Lur:** Conceptualization, Writing – review & editing. **Zhaoxia Yu:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Comparison with existing methods: Our simulations and applications to real data suggest that estimated spike data outperform trace data in cluster analysis and give comparable results for population decoding. In addition, the decobability of estimated spike data can be slightly better than that of calcium trace data with appropriate filtering / smoothing methods.

Conclusion: We conclude that spike detection might be a useful pre-processing step for certain problems such as clustering; however, the continuous nature of calcium imaging data provides a natural smoothness that might be helpful for problems such as dimensional reduction.

Keywords

Spike detection; Neural activity; Neural calcium imaging; Analysis; Population decoding; Neuronal ensembles

1. Introduction

With the technical advances in optical imaging devices, sensitive genetically encoded indicators, and pre-processing methods (Tian et al., 2009; Dombeck et al., 2010; Ghosh et al., 2011; Grienberger and Konnerth, 2012; Chen et al., 2013; Yang and Yuste, 2017), calcium imaging has been increasingly adopted as a supplement or substitute to the traditional electrophysiological measurements of neuronal firing activities. Compared to electrophysiological methods, imaging methods offer flexibility in a number of ways such as targeting specific neuronal subpopulations of interest, improved spatial resolutions, longer follow up time, and increased number of neurons that can be simultaneously recorded. One trade-off for the greater flexibility is the requirement of more sophisticated pre-possessing, as the measurement of neuronal activities from calcium imaging is indirect, and its estimate of calcium concentration is complicated by several factors such as measurement noise and contamination from non-neuronal cells (Johnston et al., 2020). As a consequence, each calcium trace is only a proxy of the underlying spiking activities with a reduced signal-to-noise ratio and temporal resolution. Thus, recent comparisons of electrophysiology and calcium imaging data are timely to guide us on how to interpret the results obtained from calcium imaging data. Using matched neuron populations and experimental conditions, Wei et al. (2020) compared electrophysiology and calcium data on temporal dynamics (within each trial), trial-type selectivity, sources of variances, and population decoding. As calcium imaging has been widely adopted in longitudinal studies to track the activities of large populations of neurons, an emerging follow-up question is whether the recorded calcium traces should be analyzed directly or deconvolved to estimate the underlying spike trains first.

Despite the continued improvement in the quality of calcium imaging, extracting the underlying spike activities that would otherwise be accurately measured by electrophysiology data is still one major challenge in analyzing calcium imaging data. As a result, numerous spike deconvolution methods have been proposed, from the simple thresholding with 2–3 standard deviations away from the calcium trace baseline to formal and fully Bayesian models, such as Yaksi and Friedrich (2006), Vogelstein et al. (2010), Pnevmatikakis et al. (2016), Deneux et al. (2016), Jewell and Witten (2018), Pachitariu et al. (2018), Berens et al. (2018), Stringer and Pachitariu (2019), Pnevmatikakis (2019),

Giovannucci et al. (2019), and three of our recent work on multi-trial data (Johnston et al., 2020; Shen et al., 2021; D'Angelo et al., 2021). Comparisons between estimated and true spike data from simulated or benchmark data usually showed that deconvolution leads to satisfactory results; on the other hand, the estimated spikes might be inconsistent across methods, which produces different estimates of firing rate, number of tuned neurons, and distribution of estimated firing rates (Evans et al., 2019).

Some research problems, such as investigating temporal coding (Abeles et al., 1993; Abbott, 1994; Theunissen and Miller, 1995) and estimating instantaneous firing rates, require the precise timing of action potentials. In addressing many other scientific questions, however, either calcium traces or estimated spike data can be used. For visualization purposes, both heat maps of calcium traces and raster plots of estimated spike trains are commonly presented. Cluster analysis is a routine practice to group neurons or trials with similar temporal patterns using spike train or estimated spike train data (Adler et al., 2012; Humphries, 2011; Diana et al., 2019). Due to the increased availability of calcium recordings, in recent work, calcium trace data have also been directly used for clustering (Ozden et al., 2008; Dombeck et al., 2009; Barbera et al., 2016). For example, cluster analysis of calcium trace data suggests that spatially compact neural clusters exist in awake mouse motor cortex at not only macrocircuitry but also microcircuitry levels (Dombeck et al., 2009). These clusters may represent cell assemblies that are stable over days; their dynamics often represent unique behavioral states and carry useful encoding information for behaviors (Barbera et al., 2016). When calcium imaging data are available, one can conduct cluster analysis using either calcium traces or the deconvolved spike data (Romano et al., 2017). In this situation, one natural question is should one conduct cluster analysis on calcium traces or the deconvolved data?

A related analysis is principal component analysis (PCA), which aims to extract components (linear combinations of time series from multiple neurons) that can keep as much as possible the variation in the original data. In neuroscience, PCA is frequently performed for dimension reduction and data visualization. Quantifying the dimensionality, which is defined as a function of the corresponding eigenvalues, may also shed light on the dynamics of the underlying neural circuit connectivity of various tasks and stimuli (Gao et al., 2017). For example, it is widely believed that, for a population of neurons, high dimensionality and low correlation imply coding efficiency whereas low dimensionality and high correlation produce robust/reliable coding (Stringer et al., 2019). PCA analysis can be conducted using either electrically recorded spikes trains or optically recorded calcium imaging data (Cunningham and Byron, 2014). One advantage of calcium recording over electrophysiology recording is that calcium imaging allows a large number of neurons to be recorded simultaneously over several weeks. A popular method to compactly visualize the neuronal dynamics is to plot PCA trajectories over time using the first two or three components (Churchland et al., 2012; Cunningham and Byron, 2014). Because calcium traces are continuous, it is more tempting and straightforward to conduct PCA using calcium traces than using the estimated spike data, which are binary. In PCA analysis of spike train data, filtering methods (Paiva et al., 2010) such as Gaussian kernel (Churchland et al., 2012) are often applied to spike trains because PCA is more natural for continuous data. When the components of the true spike activities are of interest, PCA based on calcium trace data

might not be desirable, as a substantial source of the variation might be from the random noise and slow decay in calcium intensity, rather than the signals of action potentials. Thus, one important question is whether deconvolution will help recover the true underlying components.

Another important analysis in neuroscience is decoding, which refers to finding the mapping from neural activities to either external stimuli such as presented images and experimental types or animal behavioral outcomes such as movements, speed, positions, and decision making. An interesting question is whether and how information is represented in an ensemble of neurons. Thus, population decoding methods have been widely adopted to investigate the joint activities of a group of neurons using spike train data (Brown et al., 2004). For example, Yang and Masmanidis (2020) analyzed simultaneously measured spike train data from two brain regions to compare their population decoding of choices in a two-alternative choice task. With the increased popularity of calcium imaging, more and more population decoding analyses are conducted using calcium traces. In many published studies, the calcium trace data were first deconvolved to spike train data before being used for population decoding. A recent study (Wei et al., 2020) found that, as expected, the decoding accuracy of the deconvolved spike data from calcium traces was much lower than the electrophysiology data; however, counter-intuitively, calcium trace data have higher predictability than both their deconvolved spike train data and the less noisy electrophysiology data. One possible explanation suggested by Wei et al. (2020) is, due to the slow decay rate of the observed calcium transients, the calcium trace data naturally perform “integrating” rather than instantaneous decoding. This is evidenced by the improved predictability of the electrophysiology data when a filter of 1 s was applied. It is unknown whether it is beneficial to conduct deconvolution when “integrating effects” are taken into consideration to make a fair comparison between calcium traces and their deconvolved spike trains.

In this paper, our goal is to assess the necessity of spike estimation for calcium trace data in three widely used methods, which are cluster analysis, PCA analysis, and population decoding.

2. Cluster analysis

K-means is perhaps the easiest to understand and the most widely used clustering method. The idea is to allocate a neuron to the cluster with the nearest centroid. Its improved versions have also been used in clustering either neurons or trials. For example, the meta k-means, which is a consensus clustering method that aggregates clustering results from multiple runs of k-means, has been developed to increase the clustering stability of neurons (Ozden et al., 2008). Another example is the fuzzy k-means, which accounts for clustering uncertainty by assigning cluster probabilities to each trial or neuron (Dunn, 1973; Bezdek, 2013). Here we use the conventional k-means, due to both its simplicity and popularity in cluster analysis of calcium traces. Because of the trial-to-trial and neuron-to-neuron variation in the magnitudes of peak fluorescence intensity, we use the correlation distance metric.

2.1. Results based on simulated data

We first compare the clustering accuracy of calcium traces and their deconvolved spike trains using a simulation study, where both the true spike trains and the underlying cluster structure of the neurons are known. The spike train data we analyze are a subset of the spike trains simulated by Fellous (2004). Each simulated data set consists of three clusters, which are treated as three neuron clusters here. The 35 neurons (spike trains) within each cluster share 4–6 spikes with the spike times uniformly distributed between the time interval (0–1 s). The following three ways are used to generate spike noises at various levels.

- 15% spike events are dropped randomly.
- X extra random spikes are added to each train.
- All the spike times are jittered by a value drawn from a normal distribution with mean 0 and standard deviation J million second (ms).

We consider seven noise levels (Table 1), from well-separated clusters to very "fuzzy" clusters. The raster plots in Fig. 1 show two sets of spike trains with the left panel representing low noise (level 2) and the right panel representing high noise (level 8). In Fellous (2004), 30 data sets were generated for each of the seven noise levels to account for variations in data simulations. Each fluorescence trace is generated for a given spike train using an AR(1) model (Vogelstein et al., 2010). In the AR(1) model, the calcium fluorescence for a neuron at a single trial, denoted by $y(t)$, $t = 1, \cdots, T$, is modeled using the following first-order auto-regressive model

$$\begin{aligned} y(t) &= c(t) + \epsilon(t), \epsilon(t) \sim N\left(0, \sigma^2\right), \\ c(t) &= \gamma c(t-1) + s(t), \end{aligned} \tag{1}$$

where $c(t)$ denotes the underlying true calcium concentration, $s(t)$ represents the non-negative change in calcium concentration between time points $t - 1$ and t with $s(t) > 0$ indicates a spike at time t, γ is the decay rate of calcium transients, and σ^2 denotes the variance of the noise in measuring calcium concentration. We choose the following parameters: the rate of decay $\gamma = 0.96$, the magnitude of each spike $s(t) = 1$ for any $s(t) > 0$, and four levels of Gaussian noise $\sigma \in \{0.1, 0.2, 0.3, 0.4\}$. These values coverage realistic levels of noise in calcium trace, from easy to challenging in spike deconvolution. In addition, to mimic realistic decay rates of genetically encoded indicators, we re-scale the data to 100 time data points with a sampling rate of 15 Hz. To estimate spikes from simulated calcium data, we use the ℓ_0 penalized approach of Jewell and Witten (2018) for each calcium trace.

We perform cluster analysis using the k-means with three clusters to all the three types of data, namely, the calcium traces, the estimated spike trains from calcium traces, and the true spike trains. Due to the sparsity of spike data, we apply a Gaussian filter with a relatively small bandwidth ($\sigma = 1/5$ s) to the spike data such that a little bit of continuity is gained without losing too much of the binary characteristic of spike data. More discussions on data smoothing using Gaussian filter or binning are provided in Section 3.2.

The clustering accuracy is quantified using two metrics – the Rand index (Rand, 1971) and the normalized mutual information $I(A, B)$ (Danon et al., 2005) between the estimated and

true cluster memberships. Both metrics measure the consistency between two categorical variables, with the maximum value 1 indicating perfect agreement. Not surprisingly, as presented in Fig. 2, the true spike data have the best clustering performance. Importantly, estimated spikes perform better than calcium trace at almost all noise levels.

In practice, the true number of clusters is unknown. To examine whether the results are sensitive to the choice of the number of clusters, we also calculate the Rand indices and mutual information for four and five clusters. The results (not shown) lead to the same conclusion, i.e., estimated spike data perform uniformly better than calcium traces on clustering.

2.2. A decision-making task

We then compare neuron clusters estimated from trace data to those from estimated spike data in two real studies. These applications fall into the situation of no ground truth, as neither the true number of clusters nor the clustering membership is unknown a priori. Similar to Pachitariu et al. (2018), who used concordance between repeated trials as a metric for estimated spikes, we use Rand index to examine the between-trial consistency for estimated clusters. This method is particularly relevant for the first study, as the calcium imaging data during a decision-making task were collected after the participating mice had mastered the task. Thus, it is reasonable to treat trials with the same behavior outcome as repeated trials. In this multi-trial experiment, mice were first trained to discriminate pole location using their whiskers and report the perceived pole position by licking the correct lickport to receive a small drop of water reward (Li et al., 2015; Wei et al., 2020). Their neuronal activities in the left anterolateral motor cortex region were then measured using two-photon calcium imaging using GCaMP6s. Each 5-s long trial consisted of three epochs: sample epoch (mice presented with a vertical pole), delay epoch (the pole was removed), and response epoch (mice cued to give a response). The behavioral outcome of each trial is “correct” (i.e., 31 trials of correctly lick the left lickport and 21 trials correctly lick the right lickport) or “incorrect” (i.e., eight trials of incorrectly lick the left lickport and 13 trials of incorrectly lick right lickport).

For a randomly chosen mouse with 67 neurons, K-means was used to cluster the neurons into four clusters. The results from three or five clusters are not omitted because they lead to similar conclusions. Presented in the left and right upper panels of Fig. 3 are the heat maps of Rand index matrices between trials using trace or estimated spike (Gaussian filtered with $\sigma = 1/3s$), suggesting that estimated spike data with Gaussian filtering lead to increased trial-to-trial agreement in neuron clusters. The histogram of the difference in Rand index between the two data types for all trial pairs is given in Fig. 4 (right lower panel, red-colored). To examine whether this difference is statistically significant, we compared the mean of the observed differences to those from shuffling the labels of the two data types 1000 times. Presented in the blue histogram of Fig. 3 is the histogram of the differences from one random permutation. The p-value is less than 0.001, suggesting the improvement in trial-to-trial agreement by estimated spike data is significant.

An interesting question is whether the estimated clusters from these two data types, i.e., calcium trace and estimated spike data, are consistent. The Rand index matrix between

clusters estimated from trace or estimated spike data (Fig. 3, lower left panel) indicates that there is only moderate concordance between clusters from trace and estimated spike data. Note that the blue bands are caused by trials with one empty cluster from the trace data.

2.3. A fear-based contextual discrimination experiment

The example provided in Section 2.2 is a representative example of repeated trials. As a comparison, in this study, calcium recording started in the first learning trial and lasted for a few weeks; therefore, neural dynamics are expected as a result of the learning process and neural plasticity. For this reason, incorporating neuronal dynamics in firing rate estimation is likely to improve spike detection and firing rate estimation.

In one of our previous studies (Johnston et al., 2020), mice were trained to recognize two contexts via fear conditioning (foot shock). Meanwhile, fluorescence miniature microscopes (Ghosh et al., 2011) were used to track cell populations in contextual discrimination experiments in mice's hippocampus using a genetically encoded calcium indicator (GCaMP6f) for up to 60 days. The whole experiment included four stages: habituation (mice freely exploring environment), learning (a stimulus context with foot shock), extinction (no foot shock), and relearning (stimulus reinstated).

We analyze a mouse whose 141 neurons were measured in 21 trials, with the first 11 trials in the learning stage and the last 10 trials in the relearning stage. The calcium data collected 10 s before and 60 s after foot shock were used for analysis. Note that several decisions need to be made before clustering, such as the number of clusters, the degree and the way that spike trains are smoothed. To check whether the comparison results between trace and estimate spike are sensitive to those choices, we examine combinations of these factors, including $K \in \{3, 4, 5\}$ clusters, Gaussian filtering with $\sigma \in \left\{\frac{1}{15}, \frac{1}{5}, \frac{1}{3}, \frac{2}{3}, 1\right\}$ second, and bin size $\in \left\{\frac{1}{15}, \frac{2}{15}, \frac{1}{5}, \frac{1}{3}, \frac{2}{3}, 1\right\}$ second.

However, the Rand index values can be very low for binned data with a size less than 1/3 s. This is presumably due to the discreteness with small bin sizes. Shown in Fig. 4 is the Rand index matrices of trace data (upper left), spike data with a Gaussian filter of 1/3 s (upper right), binned data with a size of 1/3 s (bottom left), and trace vs spike (Gaussian filtered, bottom right) when $K = 5$ clusters are assumed. Importantly, the Gaussian filtered spike data achieve higher trial-to-trial agreement with most large Rand index values close to the diagonal (i.e., neighboring trials). Using the permutation method described in Section 2.2, the p-values of significant difference are less than 0.01 for Gaussian filtered ($\sigma = 1/3s$) vs trace and less than 0.05 for binned (size = 1/3 s) vs trace. Of particular interest is the high consistency of the relearning trials, especially when Gaussian filtered spike data were used for clustering. This may suggest that a stable neuronal ensemble has been formed. For the trace vs spike plot (lower right, Fig. 4), the value in the ith row and jth column is the Rand index between the clusters of the ith trial based on trace and the clusters of the jth trial based on spike data. The high Rand index values on the diagonal suggest that trace and spike data at least show good agreement on the same trials, which is expected.

Similar conclusions are obtained when the neurons are clustered into three or four clusters. Due to the lack of ground truth in neuron clusters, caution should be taken when relating results to potential misspecifications of cluster numbers.

3. Principal component analysis (PCA)

PCA is a classical method of dimension reduction. In PCA analysis of neuron data, the aim of PCA is to project neurons into a low-dimensional subspace and the extracted components are linear combinations of individual neurons. It has been widely used in analyzing large-scale neuron data (Chapin and Nicolelis, 1999; Paiva et al., 2010; Churchland et al., 2012; Harvey et al., 2012; Bekolay et al., 2014; Cunningham and Byron, 2014; Kobak et al., 2016; Stringer et al., 2019). For example, choice-specific neuron trajectories were reported in the parietal cortex during a virtual-navigation decision task (Harvey et al., 2012). Using a delayed response task, Wei et al. (2020) compared PCA between electrophysiology and calcium imaging data, and found that the proportions of explained variance of the first three PCs are quite different between calcium trace and electrophysiological data. In this article, we will use both the simulated and real data presented in Section 2 to examine the difference between trace and spike data in PCA analysis.

3.1. PCA results based on simulated data

We first computed an interesting summary metric of the distribution of the eigenvalues – the effective dimension of an embedding (Victor and Purpura, 1997), which is also noted as the dimensionality (Gao et al., 2017). It is defined as $E = \frac{(\sum \lambda_i)^2}{\sum \lambda_i^2}$ where λ_I is the i-th eigenvalue of the covariance matrix of calcium traces/true spike trains/estimated spike trains between neurons. The result in the left panel of Fig. 5 shows that the dimensionality calculated from estimated spike data is more similar to that from the true spikes, especially when the noise levels of spike and calcium are low. At the lowest noise level (level 2, left panel of Fig. 1), there are three clear clusters. The effective dimension from the trace data with a small Gaussian noise is the closest to the true number of clusters than that from spike or estimated spike data. As the noise level increases, three clusters might not be adequate because there is more variation across neurons (right panel of Fig. 1). As a result, it is expected that the effective dimension should increase as the spike noise level increases from level 2 to level 8. All data types, except the trace with a large Gaussian noise, demonstrate this expected trend (Fig. 5, left panel). On the other hand, the effective number from the true or estimated spike data is much larger than three, even for a noise level with three noticeable clusters such as level 6 (Fig. 1, right panel); as a comparison, the estimated dimension from trace data is still small. One possible explanation is that the slow decay in calcium fluorescence might compensate for the jittering noises of spike times.

To understand whether the loadings from spike data agree with those from trace data, we plotted the cumulative proportions of explained variances by the 10 leading principal components for the data set at spike noise level 3 and calcium noise $\sigma = 0.1$. As illustrated in the right panel of Fig. 5, the curves based on the true spike trains (black) are very similar to those based on the estimated spike trains (blue) but different (marginally significant) from

those based on the calcium traces (red). Although the estimated spike data are closer to the true spike data in terms of spectral characteristics, such as effective dimension and percent of variance explained, the trace data seem to be more robust against jittering noise in spike times.

3.2. PCA results for the fear-based discrimination data

We next examine the effective dimensions computed from calcium traces or estimated spikes for the fear-based contextual discrimination experiment. As described in Section 2.3, each session is 70 s long with a sampling rate of 15 Hz. The resulting number of data points is over 1000, which allows us to examine the effect of binning spike trains, in addition to Gaussian smoothing. The following σ (for Gaussian) and bin size are considered: {1/15, 1/5, 1/3, 2/3, 1} second. Shown in Fig. 6 is the results using selected smoothing parameters. The effective dimension seems to decrease during the learning stage and vary less during the relearning, which might suggest the formation of neural ensembles after learning. However, given the small number of trials, there is not enough information to make a conclusion on the trend of effective dimensions over time.

When comparing the effective dimensions across data types and smoothing methods, we notice that the trace data provide similar results with smoothed spike data using a Gaussian filter with $\sigma = 1$ s. Similarly, binning with a size of 1 s is similar to Gaussian filtering with $\sigma = 1/3$ s, and binning with a size of 2/3 s is similar to Gaussian filtering with $\sigma = 1/5$ s. It should be kept in mind that here we focus on a single summary of PCA analysis, i.e., effective dimension. As PCA is a multi-purpose tool (such as dimension reduction and visualization), trace and spike data might differ in multiple ways. For instance, based on our own experience, neuron trajectories could vary substantially with the chosen data type and smoothing method. It is worth conducting more research on PCA with calcium imaging.

4. Population decoding analysis

4.1. Data “pre-processing”

Our goal in this section is to investigate whether deconvolution is helpful in population decoding. Wei et al. (2020) reported that electrophysiology and estimated spike data are less informative in population decoding at each time bin. This is somewhat surprising but not totally unexpected. There are several possible explanations such as the integration effect of the calcium trace data (Wei et al., 2020) and the sparsity of spike data in each time window. One support of these explanations is that the one-second filtered electrophysiology data reached higher peak decoding accuracy than the original analysis of electrophysiology data with a bin size 67 ms. In fact, temporal filtering is a common practice in population decoding, and choosing the appropriate filtering encourages data integration temporally, leading to improved decoding accuracy (Yates et al., 2020; Tu et al., 2020; Park et al., 2014). For this reason, we consider both an instantaneous approach, which uses the original calcium trace resolution (15 Hz), and the 1 s filtered analysis, which uses the sum of a current time point and its previous fourteen time points for each neuron.

In addition, it has been known that coding timescales vary across cortical areas, stimulus types, and choice contexts (Runyan et al., 2017). For example, sensory cortices were often reported to have short coding timescales at both individual and population levels (Barbera et al., 2016; Runyan et al., 2017), whereas the posterior parietal cortex showed improved population decoding accuracy for behavioral choices when a cumulative decoding, rather than instantaneous decoding, was conducted (Runyan et al., 2017), Extended Fig. 7). They reasoned that the longer timescale could be due to the coupling of neurons in the posterior parietal cortex. In their analysis, cumulative decoding refers to a decoding strategy that either uses all the data up to a given time point (Runyan et al., 2017) or uses all available time points (Berens et al., 2012). In this article, in addition to the 1 s filtered method, we examine two other approaches to take neural firing history into consideration: the cumulative approach uses the total number of spikes up to a time t for each neuron, and the history approach uses the whole time course up to time t for each neuron. Statistically, at each given time t, our cumulative approach is a simple linear decoder that can be obtained from a multiple logistic regression. Compared to both the 1 s filtered and the cumulative approach, the history approach is more data-driven in that the weights of different time points are determined by the data. The details of the history method are described in Section 4.3.

4.2. Data types

Note that the spike detection framework based on the AR(1) model in Eq. (1) not only produces estimated timestamps of spike events but also provides estimates of spike magnitudes, denoted by $s(t)$. The spike magnitude $s(t)$ was originally motivated by the number of spikes in a small time bin (Vogelstein et al., 2010). Due to pre-processing such as data normalization, the interpretation of a positive $s(t)$ is not clear; nevertheless, it might still provide useful quantitative information related to spike counts and is therefore useful for population decoding. Another quantity estimated by the AR(1) model is the calcium concentration $c(t)$. The estimated $c(t)$ can be considered as a denoised version of the observed calcium trace. Interestingly, $c(t)$ can also be viewed as a filtered version of the underlying spike train. As spike train filters are often applied in various spike train analyses in neuroscience research, it is reasonable to use the estimated $c(t)$ as a candidate for population decoding. Thus, we will focus on the following four types of data: the raw calcium traces, the estimated spike trains, the estimated changes $s(t)$, and the estimated/denoised traces $c(t)$.

4.3. Prediction methods

Taken together, the data types and ways of pre-processing create various combinations of decoding, as illustrated in Table 2. Note that in real studies, the category of spike train might be spike train data from electrophysiology measurements or estimated spike trains from calcium trace data.

When decoding using traces and cumulative counts, we apply a linear support vector machine (SVM) (Cortes and Vapnik, 1995) with 5-fold cross-validation on the training data. To assess the dependency of results on predictive models, we also used Fisher's linear discriminant analysis. Because they give similar results in the scenarios we consider, only results from SVM will be reported.

For the history decoding approach, predictors are a matrix of the time series (up to time t) from a neuron population. Traditional methods such as SVM reshape the matrices into vectors, causing a loss of structural information such as time history. Here we apply a sparse support matrix machine method (SSMM) (Zheng et al., 2018), which is a regularized binary matrix classifier that uses a ℓ_1 penalty to ensure sparsity and a nuclear norm penalty to encourage low rank of the coefficient matrix. Given a set of samples $\{\mathbf{X}_i, y_i\}_{i=1}^n, \mathbf{X}_i \in \mathbb{R}^{p \times q}$ is the ith predictor matrix and $y_i \in \{1, -1\}$ is its corresponding label. The proposed SSMM (Zheng et al., 2018) is based on support vector machine (Cortes and Vapnik, 1995) and it combines a hinge loss with a new regularization on the regression coefficient matrix $\mathbf{W}$. The objective function of SSMM is presented as:

$$\operatorname{argmin}_{\mathbf{W}, b} \left\{ \lambda \|\mathbf{W}\|_1 + \tau \|\mathbf{W}\|_* + \sum_{i=1}^{n} \left\{ 1 - y_i \left[\operatorname{tr}\left(\mathbf{W}^T \mathbf{X}_i\right) + b \right] \right\}_+ \right\}$$

where $\mathbf{X}_i, \mathbf{W} \in \mathbb{R}^{p \times q}$. It minimizes the penalized hinge loss with both an ℓ_1 norm $\|\mathbf{W}\|_1$ and a nuclear norm $\|\mathbf{W}\|*$ on $\mathbf{W}$. The ℓ_1 norm controls the sparsity of $\mathbf{W}$ and the nuclear norm encourages $\mathbf{W}$ to be low-rank. An efficient algorithm to solve the optimization problem was presented in Zheng et al. (2018), in which a generalized smooth hinge loss with Lipschitz-continuous gradient was used. The computational cost is $\mathcal{O}\left(n^2 pq\right)$.

The tuning parameters λ and τ in SSMM control the sparsity and rank of the input matrix, respectively. Zheng et al. (2018) reported that positive values of the tuning parameters τ and λ often give better results than no regularization. We used a 5-fold cross-validation to the training data to choose their optimal tuning parameters.

4.4. Decoding accuracy for the water lick data

The water lick data has been described in Section 2.2. The primary goal of the original study was to compare electrophysiology and imaging data for measuring neuronal activities in the anterior lateral motor cortex (Wei et al., 2020). Although the electrophysiology and calcium imaging recordings were separate, neuron populations of matched depth were measured from the same delayed response task. Here we analyzed the calcium imaging data for 1493 neurons from 4 mice with adeno-associated virus expressing GCaMP6s and electrophysiology data for 720 neurons from 19 mice. To match the temporal resolution of calcium trace data, the spike train data from electrophysiology recordings were temporally subsampled to 15 Hz. Note that the data we downloaded from https://doi.org/10.6084/m9.figshare.12786296.v1 have been pre-processed. We also followed several other strategies used in Wei et al. (2020), such as resampling trials to cope with the limited number of trials, resampling the same number of neurons when comparing different data types, and using cross-validation to choose tuning parameters (60% data for training and 40% data for testing). The full description of the experiments and data processing is available from Wei et al. (2020) and further backgrounds about the experimental design can be found in other articles from the same lab (Li et al., 2015; Wei et al., 2019). Fig. 7 presents the decoding accuracy, which is defined as the proportion of correct predictions in testing data, of the prediction methods in Table 2. The top panel replicated the reported difference between the

decodability of the calcium trace and the electrophysiology recordings (Wei et al., 2020). Although electrophysiology data showed earlier latency (defined as the time reaching %70 decoding accuracy), surprisingly, it had lower instantaneous decoding accuracy (the solid black curve) than the calcium trace data, which are noisy measurements of neural activities. This difference might be due to the integration effect of the calcium trace data, as a calcium transient is characterized by a rapid rise but slow decay. This hypothesis is supported by the improved decoding accuracy of the 1-second filtered electrophysiology data. However, the 1-second filtered electrophysiology data lost the early latency. As a comparison, the spike history, a new method considered here, outperforms the calcium trace data in all time bins without losing its early latency. Thus, when used appropriately, the less noisy spike train data from electrophysiology recordings do achieve both earlier latency and higher maximum decodability.

The lower panel of Fig. 7 focused on the effects of deconvolving calcium traces. Consistent with Wei et al. (2020), the estimated spike data has lower instantaneous decodability (the solid red curve) than both calcium traces and electrophysiology data. The magnitude of changes, i.e., $s(t)$ has a similar instantaneous accuracy rate (the solid blue curve). However, their cumulative, 1 s filtered, and history decoders have much higher accuracy, with the cumulative and 1 s filter decoders outperforming the calcium trace data. It is worth pointing out that the estimated calcium concentrations (“estimated calcium trace”), which is a denoised version of the calcium traces, also demonstrate higher decoding accuracy than the original calcium traces. These results indicate that the decodability of a neuron population can be improved by appropriate analysis of the deconvolved data.

One noticeable difference between the two plots in Fig. 7 is that the cumulative approach works better than the history approach for electrophysiology data but worse for the estimate spike data. There are several possibilities for this inconsistency. First, estimating spike trains is challenging due to multiple factors. For example, when firing rate is high and kinetics is slow, adjacent spikes cannot be accurately estimated due to the overlapping fluorescence responses (Ali and Kwan, 2019). In this situation, the number of spikes might be underestimated. For neurons that are active in bursts, such as pyramidal cells, we are only able to detect the starting time of a calcium event rather than individual spikes (Stefanini et al., 2020). Second, the observed difference of different decoding methods in this study need to be verified using similar experimental settings and neural populations. Additionally, it is not known whether the results depicted in Fig. 7 can be generalized to other tasks or neural populations. Nevertheless, the results indicate that decodability can be improved by smoothing discrete data (electrophysiology data and estimate spike data) and denoising the raw calcium trace data through deconvolution.

5. Discussion

In this article, we assessed the necessity of spike estimation by comparing it with calcium trace data in three analyses: cluster analysis, PCA, and population decoding. Using simulated data, we showed that deconvolution tends to improve clustering accuracy. In our analysis of real studies, the estimated spike trains produced neural clusters with higher trial-to-trial concordance. In addition, when analyzed using appropriate predictive models

and smoothing methods, they give slightly more accurate results in population decoding. However, the PCA results from trace data seem to be more reasonable, especially in the presence of high noise in spike times. In practice, we recommend analyzing both trace and estimated spike data, as they might provide supplementary information to each other.

For computational efficiency, we chose a recently published ℓ_0 penalized method (Jewell and Witten, 2018) to estimate spike trains from calcium traces. Because different deconvolution methods have varying accuracy in spike estimation, it is expected that the comparison results depend on the choice of deconvolution methods, as reported in (Evans et al., 2019). On the other hand, given the consistent observations in our study and previous studies (Evans et al., 2019; Wei et al., 2020), it is reasonable to assume that the difference between choosing calcium trace and estimated spike data is mainly due to the nature of the data, such as quantitative vs binary, continuous vs sparse, and delayed/integrated vs instantaneous.

In examining cluster analysis, we used a prefixed number of clusters. In practice, this is unknown. Further, it has been shown that the number of clusters changes during behavioral states (Paz et al., 2005). Similarly, we conducted PCA one trial at a time. It might be useful to examine PCA results in a refined temporal resolution to characterize dynamics population activity (Rossi-Pool et al., 2017). It is interesting to examine how trace or spike data characterize the evolution and dynamics of neural ensembles at different time resolutions.

The results also likely vary with the specific analysis methods chosen, such as the choice of clustering methods. There is a rich body of literature on clustering, PCA, and population decoding methods that are tailored for spike train and calcium imaging data. For example, clustering methods that account for clustering uncertainty by relaxing the strict membership (Dunn, 1973; Bezdek et al., 1984; Bezdek, 2013) or by aggregating clustering results from multiple runs of k-means (Ozden et al., 2008). For PCA, kernel-based methods have been developed to handle the sparsity of spike train data (Carnell and Richardson, 2005; Paiva et al., 2009; Schrauwen and Van Campenhout, 2007, 2010); recently, novel PCA methods have been proposed for PCA analysis of neuron responses. For example, the cross-validated PCA extracts reliable components by removing trial-to-trial variability or unrelated cognitive or behavioral variables (Stringer et al., 2019). In the demixed PCA (Berens et al., 2012; Kobak et al., 2016), the variation in data is demixed into task-related components using a decomposition method that is similar to the variance (covariance) decomposition in multivariate analysis of variance (MANOVA). We chose the most basic ones, i.e., k-means, the conventional PCA, and linear decoders as they might be researchers' first choices. While adopting more sophisticated methods might further improve the performance of both calcium traces and the deconvolved spike data, we expect that the main conclusion remains.

Acknowledgements

This research is supported by NIH/NINDS R01NS104897 (Drs. Xu and Yu) and NIH/NIMH R01MH123686 (Drs. Lur and Yu).

References

Abbott LF, 1994. Decoding neuronal firing and modelling neural networks. Q. Rev. Biophys. 27, 291–331. [PubMed: 7899551]

Abeles M, Bergman H, Margalit E, Vaadia E, 1993. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol. 70 (4), 1629–1638. [PubMed: 8283219]

Adler A, Katabi S, Finkes I, Israel Z, Prut Y, Bergman H, 2012. Temporal convergence of dynamic cell assemblies in the striato-pallidal network. J. Neurosci. 32 (7), 2473–2484. [PubMed: 22396421]

Ali F, Kwan AC, 2019. Interpreting in vivo calcium signals from neuronal cell bodies, axons, and dendrites: a review. Neurophotonics 7 (1), 011402. [PubMed: 31372367]

Barbera G, Liang B, Zhang L, Gerfen CR, Culurciello E, Chen R, Li Y, Lin DT, 2016. Spatially compact neural clusters in the dorsal striatum encode locomotion relevant information. Neuron 92 (1), 202–213. [PubMed: 27667003]

Bekolay T, Laubach M, Eliasmith C, 2014. A spiking neural integrator model of the adaptive control of action by the medial prefrontal cortex. J. Neurosci. 34 (5), 1892–1902. [PubMed: 24478368]

Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS, 2012. A fast and simple population code for orientation in primate v1. J. Neurosci. 32 (31), 10618–10626. [PubMed: 22855811]

Berens P, Freeman J, Deneux T, Chenkov N, McColgan T, Speiser A, Macke J, Turaga SC, Mineault P, Rupprecht P, 2018. Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. PLoS Comput. Biol. 14 (5), e1006157. [PubMed: 29782491]

Bezdek JC, 2013. Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business Media.

Bezdek JC, Ehrlich R, Full W, 1984. Fcm: the fuzzy c-means clustering algorithm. Comput. Geosci. 10 (2–3), 191–203.

Brown EN, Kass RE, Mitra PP, 2004. Multiple neural spike train data analysis: state-of-the-art and future challenges. Nat. Neurosci. 7 (5), 456–461. [PubMed: 15114358]

Carnell A, Richardson D, 2005. Linear algebra for time series of spikes. ESANN 363–368.

Chapin JK, Nicolelis MA, 1999. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. J. Neurosci. Methods 94 (1), 121–140. [PubMed: 10638820]

Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, et al. , 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499, 295–300. [PubMed: 23868258]

Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV, 2012. Neural population dynamics during reaching. Nature 487, 51–56. [PubMed: 22722855]

Cortes C, Vapnik V, 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Cunningham JP, Byron MY, 2014. Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. 17, 1500–1509. [PubMed: 25151264]

D'Angelo L, Canale A, Yu Z, Guindani M, 2021. Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data. arXiv:2102.09403, Under minor revision for Biometrics.

Danon L, Diaz-Guilera A, Duch J, Arenas A, 2005. Comparing community structure identification. J. Stat. Mech.: Theory Exp P09008.

Deneux T, Kaszas A, Szalay G, Katona G, Lakner T, Grinvald A, Rózsa B, Vanzetta I, 2016. Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. Nat. Commun. 7, 12190. [PubMed: 27432255]

Diana G, Sainsbury TT, Meyer MP, 2019. Bayesian inference of neuronal assemblies. PLoS Comput. Biol. 15 (10), e1007481. [PubMed: 31671090]

Dombeck DA, Graziano MS, Tank DW, 2009. Functional clustering of neurons in motor cortex determined by cellular resolution imaging in awake behaving mice. J. Neurosci. 29 (44), 13751–13760. [PubMed: 19889987]

Dombeck DA, Harvey CD, Tian L, Looger LL, Tank DW, 2010. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. Nat. Neurosci. 13 (11), 1433–1440. [PubMed: 20890294]

Dunn JC, 1973. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. J. Cybern. 3, 32–57.

Evans M, Petersen RS, Humphries MD, 2019. On the use of calcium deconvolution algorithms in practical contexts. bioRxiv 871137.

Fellous JM, Tiesinga PH, Thomas PJ, Sejnowski TJ, 2004. Discovering spike patterns in neuronal responses. J. Neurosci. 24 (12), 2989–3001. [PubMed: 15044538]

Gao P, Trautmann E, Yu B, Santhanam G, Ryu S, Shenoy K, Ganguli S, 2017. A theory of multineuronal dimensionality, dynamics and measurement. BioRxiv 214262.

Ghosh KK, Burns LD, Cocker ED, Nimmerjahn A, Ziv Y, El Gamal A, Schnitzer MJ, 2011. Miniaturized integration of a fluorescence microscope. Nat. Methods 8 (10), 871. [PubMed: 21909102]

Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, Taxidis J, Najafi F, Gauthier JL, Zhou P, 2019. CaImAn an open source tool for scalable calcium imaging data analysis. Elife 8, e38173. [PubMed: 30652683]

Grienberger C, Konnerth A, 2012. Imaging calcium in neurons. Neuron 73 (5), 862–885. [PubMed: 22405199]

Harvey CD, Coen P, Tank DW, 2012. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484, 62–68. [PubMed: 22419153]

Humphries MD, 2011. Spike-train communities: finding groups of similar spike trains. J. Neurosci. 31 (6), 2321–2336. [PubMed: 21307268]

Jewell S, Witten D, 2018. Exact spike train inference via l0 optimization. Ann. Appl. Stat. 12 (4), 2457. [PubMed: 30627301]

Johnston KG, Grieco SF, Yu Z, Jin S, Shen T, Crary R, Guzowski J, Holmes T, Nie QC, Xu X, 2020. Robust population single neuronal calcium signal extraction using SCOUT allows for longitudinal analysis of behavior-associated neural ensemble dynamics. bioRxiv.

Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi XL, Romo R, Uchida N, Machens CK, 2016. Demixed principal component analysis of neural population data. Elife 5, e10989. [PubMed: 27067378]

Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K, 2015. A motor cortex circuit for motor planning and movement. Nature 519, 51–56. [PubMed: 25731172]

Ozden I, Lee HM, Sullivan MR, Wang SSH, 2008. Identification and clustering of event patterns from in vivo multiphoton optical recordings of neuronal ensembles. J. Neurophysiol. 100, 495–503. [PubMed: 18497355]

Pachitariu M, Stringer C, Harris KD, 2018. Robustness of spike deconvolution for neuronal calcium imaging. J. Neurosci. 38 (37), 7976–7985. [PubMed: 30082416]

Paiva AR, Park I, Principe JC, 2009. A reproducing kernel hilbert space framework for spike train signal processing. Neural Comput. 21 (2), 424–449. [PubMed: 19431265]

Paiva AR, Park I, Principe JC, 2010. Inner products for representation and learning in the spike train domain. In: Statistical Signal Processing for Neuroscience and Neurotechnology. Academic Press, pp. 265–309.

Park IM, Meister ML, Huk AC, Pillow JW, 2014. Encoding and decoding in parietal cortex during sensorimotor decision-making. Nat. Neurosc. 17, 1395–1403.

Paz R, Natan C, Boraud T, Bergman H, Vaadia E, 2005. Emerging patterns of neuronal responses in supplementary and primary motor areas during sensorimotor adaptation. J. Neurosci. 25 (47), 10941–10951. [PubMed: 16306407]

Pnevmatikakis EA, 2019. Analysis pipelines for calcium imaging data. Curr. Opin. Neurobiol. 55, 15–21. [PubMed: 30529147]

Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, Reardon T, Mu Y, Lacefield C, Yang W, et al. , 2016. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. Neuron 89 (2), 285–299. [PubMed: 26774160]

Rand WM, 1971. Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. 66 (336), 846–850.

Romano SA, Pérez-Schuster V, Jouary A, Boulanger-Weill J, Candeo A, Pietri T, Sumbre G, 2017. An integrated calcium imaging processing toolbox for the analysis of neuronal population dynamics. PLoS Comput. Biol. 13 (6), e1005526. [PubMed: 28591182]

Rossi-Pool R, Zainos A, Alvarez M, Zizumbo J, Vergara J, Romo R, 2017. Decoding a decision process in the neuronal population of dorsal premotor cortex. Neuron 96 (6), 1432–1446. [PubMed: 29224726]

Runyan CA, Piasini E, Panzeri S, Harvey CD, 2017. Distinct timescales of population coding across cortex. Nature 548, 92–96. [PubMed: 28723889]

Shen Tong, Johnston Kevin, Lur Gyorgy, Guindani Michele, Ombao Hernando, Yu Zhaoxia, 2021. Time-varying l_0 optimization for spike inference from multi-trial calcium recordings. arXiv:2103.03818, (submitted for publication).

Schrauwen B, Van Campenhout J, 2007. Linking non-binned spike train kernels to several existing spike train metrics. Neurocomputing 70, 1247–1253.

Stefanini F, Kushnir L, Jimenez JC, Jennings JH, Woods NI, Stuber GD, Kheirbek MA, Hen R, Fusi S, 2020. A distributed neural code in the dentate gyrus and in ca1. Neuron 107 (4), 703–716. [PubMed: 32521223]

Stringer C, Pachitariu M, 2019. Computational processing of neural recordings from calcium imaging data. Curr. Opin. Neurobiol. 55, 22–31. [PubMed: 30530255]

Stringer C, Jian M, Steinmetz N, Carandini M, Harris KD, 2019. High-dimensional geometry of population responses in visual cortex. Nature 571, 361–365. [PubMed: 31243367]

Theunissen F, Miller JP, 1995. Temporal encoding in nervous systems: a rigorous definition. J. Comput. Neurosci. 2 (2), 149–162. [PubMed: 8521284]

Tian L, Hires SA, Mao T, Huber D, Chiappe ME, Chalasani SH, Petreanu L, Akerboom J, McKinney SA, Schreiter ER, 2009. Imaging neural activity in worms, flies and mice with improved gcamp calcium indicators. Nat. Methods 6 (12), 875–881. [PubMed: 19898485]

Tu M, Zhao R, Adler A, Gan WB, Chen ZS, 2020. Efficient position decoding methods based on fluorescence calcium imaging in the mouse hippocampus. Neural Comput. 32 (6), 1144–1167. [PubMed: 32343646]

Victor JD, Purpura KP, 1997. Metric-space analysis of spike trains: theory, algorithms and application. Netw.: Comput. Neural Syst. 8, 127–164.

Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, Paninski L, 2010. Fast nonnegative deconvolution for spike train inference from population calcium imaging. J. Neurophysiol. 104 (6), 3691–3704. [PubMed: 20554834]

Wei Z, Inagaki H, Li N, Svoboda K, Druckmann S, 2019. An orderly single-trial organization of population dynamics in premotor cortex predicts behavioral variability. Nat. Commun. 10, 216. [PubMed: 30644387]

Wei Z, Lin BJ, Chen TW, Daie K, Svoboda K, Druckmann S, 2020. A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. PLoS Comput. Biol. 16 (9), e1008198. [PubMed: 32931495]

Yaksi E, Friedrich RW, 2006. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging. Nat. Methods 3, 377–383. [PubMed: 16628208]

Yang L, Masmanidis SC, 2020. Differential encoding of action selection by orbitofrontal and striatal population dynamics. J. Neurophysiol. 124 (2), 634–644. [PubMed: 32727312]

Yang W, Yuste R, 2017. In vivo imaging of neural activity. Nat. Methods 14, 349–359. [PubMed: 28362436]

Yates JL, Katz LN, Levi AJ, Pillow JW, Huk AC, 2020. A simple linear readout of mt supports motion direction-discrimination performance. J. Neurophysiol. 123 (2), 682–694. [PubMed: 31852399]

Zheng Q, Zhu F, Qin J, Chen B, Heng PA, 2018. Sparse support matrix machine. Pattern Recognit. 76, 715–726.

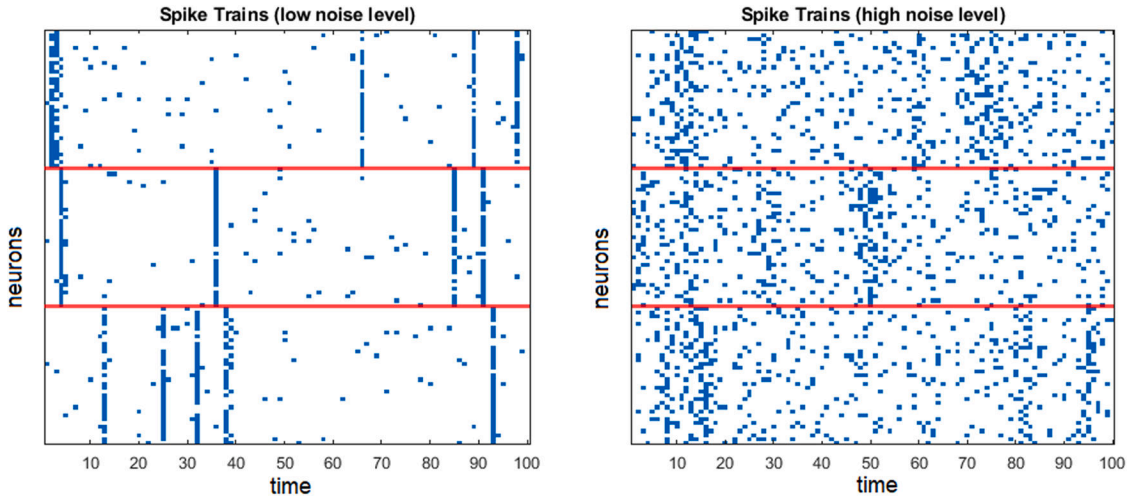

Fig. 1.

Raster plot of a simulated spike data set with $G = 3$ groups (separated by red lines) and 35 spike trains in each cluster. *Left*: with $X = 3$ random spikes and $J = 3$ ms jitter. *Right*: with $X = 11$ random spikes and $J = 15$ ms jitter.

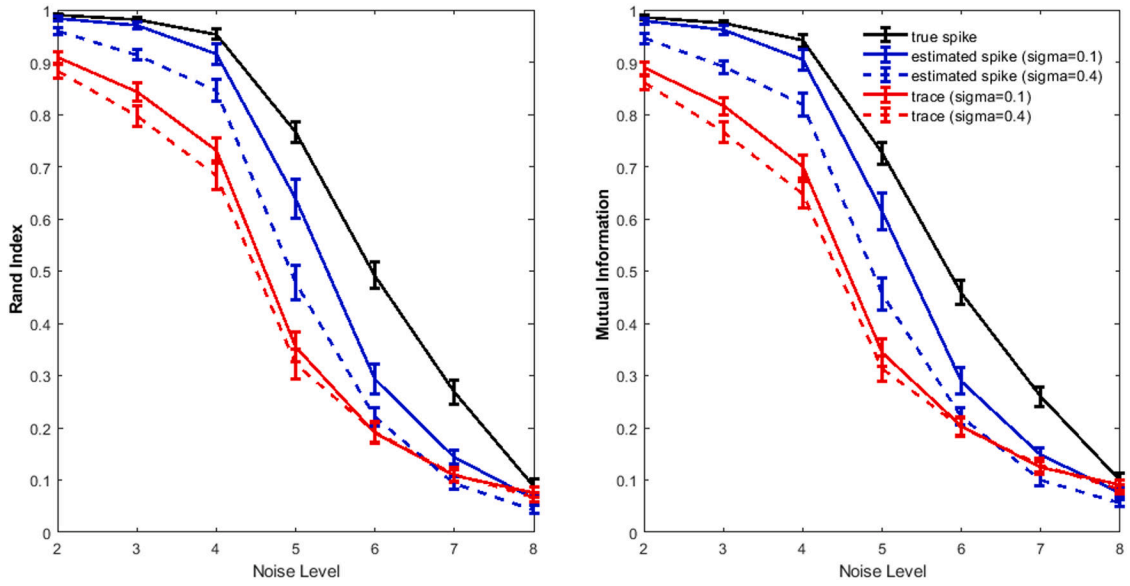

Fig. 2.

Clustering results for simulated data over seven spike noise levels. For the ease of visualization,only two calcium noise levels are presented: $\sigma = 0.1$ and $\sigma = 0.4$. Presented are means and 95% confidence intervals over 30 data sets in each simulation setting. *Left*: Rand index. *Right*: mutual index.

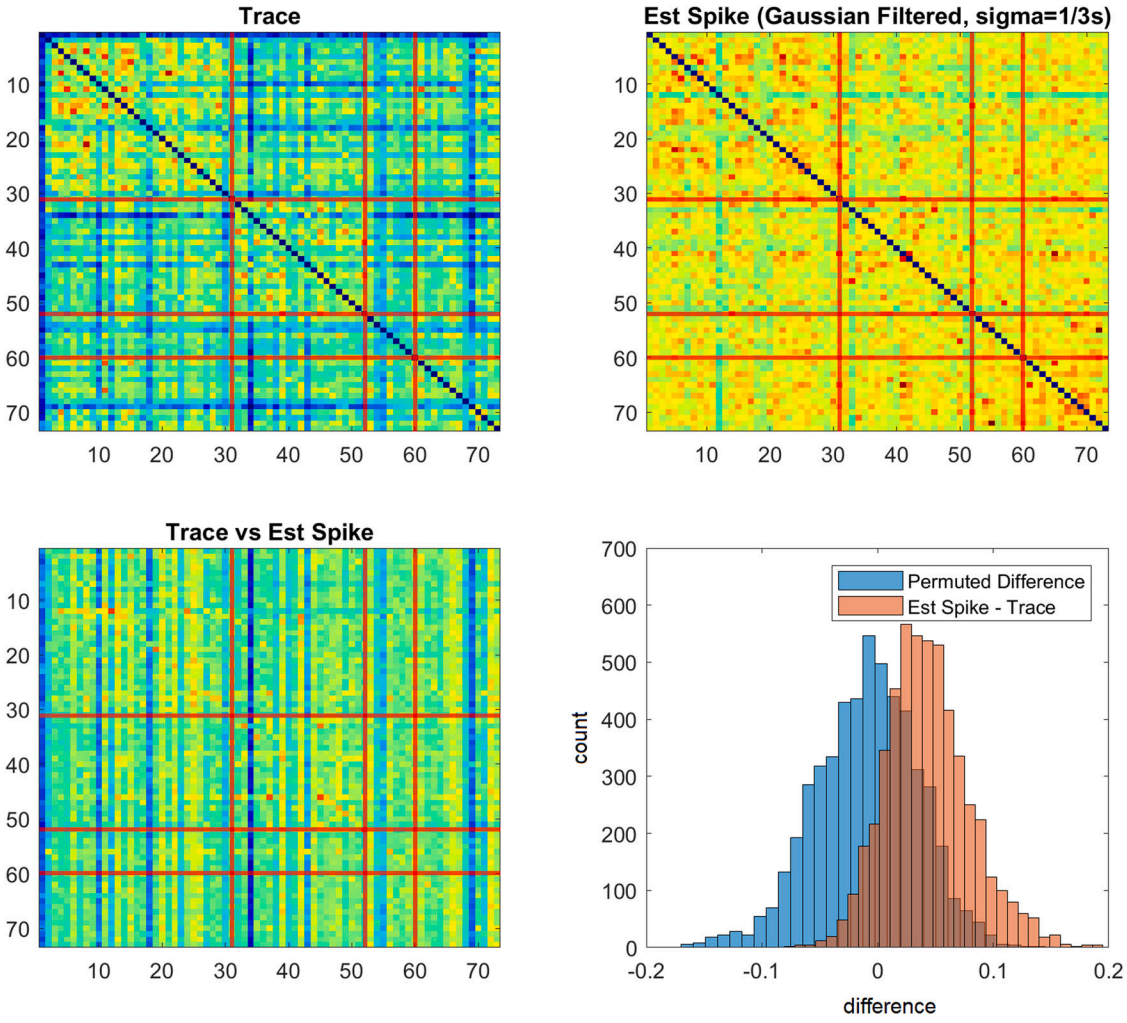

Fig. 3.

Rand index obtained from a mouse in 71 repeated sessions of a decision-making task. The results are based on the assumption of four clusters. The range of values in the heat maps is from 0.45 (dark blue) to 0.75 (dark red). Note the diagonal elements of the three above Rand index matrices are all equal to the maximum value 1. For better visualization, we manually set them as missing. Each element in the heat map matrix is the Rand index for a pair of trials. The red lines on one each heat map separate the trials according to the trial types. Specifically, from left to right and top to bottom: correctly lick the left lickport (31 trials), correctly lick the right lickport (21 trials), incorrectly lick the left (8 trials), and incorrectly lick the right (13 trials). *Upper Left*: Rand index based on calcium traces; *Upper Right*: Rand index using Gaussian filtered spike trains with $\sigma = 1/3$ s; *Lower Left*: trace vs spike, with the value in the ith row and jth column being the Rand index between the clusters of the ith trial based on trace and the clusters of the jth trial based on spike data. In several trials, the trace data produced three, rather than four clusters; as a result, the Rand index between those trials and other trials is low, creating blue-colored (low Rand index values) lines in the heat maps. *Lower Right*: histograms of the observed and permuted Rand indices.

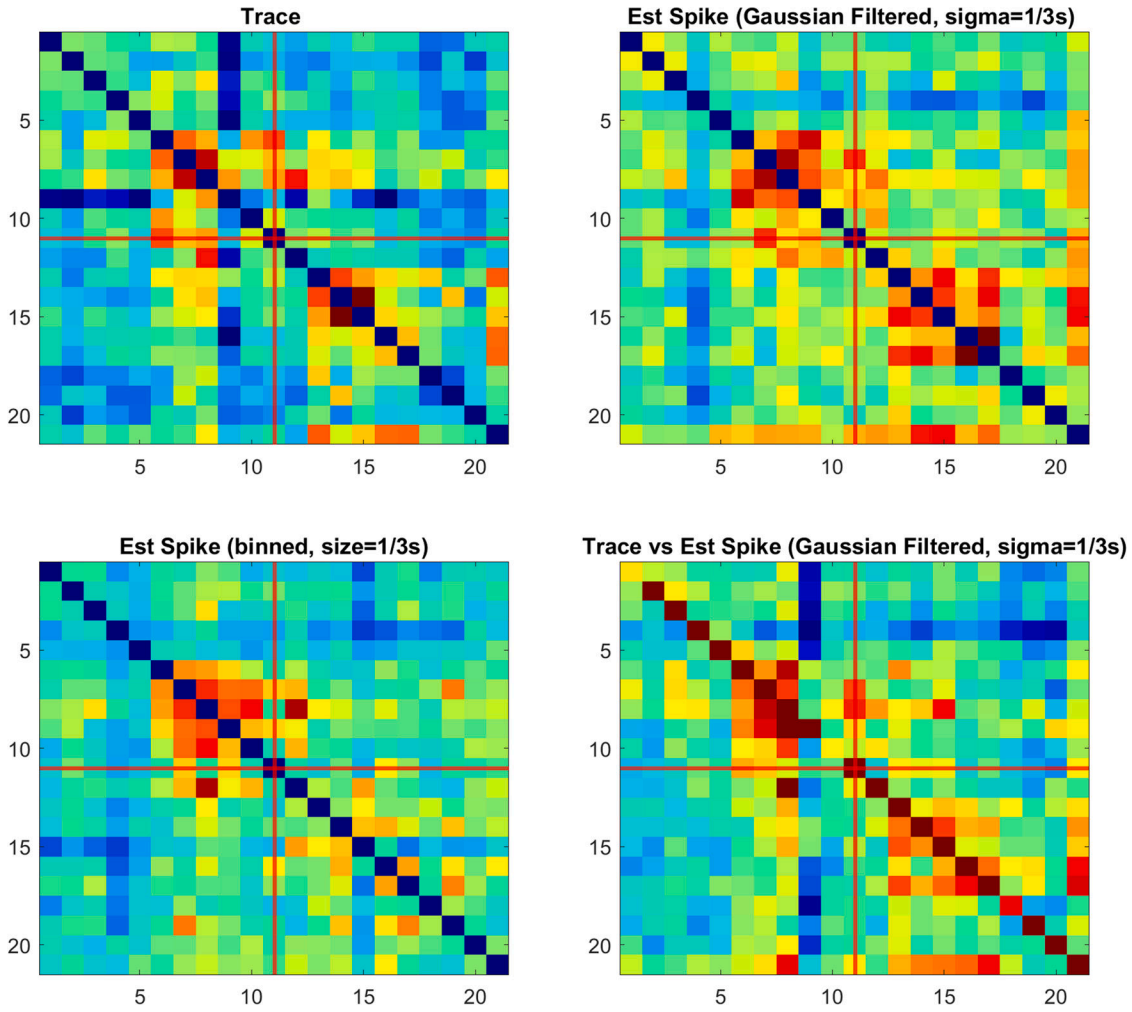

Fig. 4.

Rand index for the fear conditioning data obtained from a mouse in 31 shock sessions. The results are based on the assumption of five clusters. The range of values in the heat maps is from 0.65 (dark blue) to 0.85 (dark red). Note the diagonal elements of the three above Rand index matrices are all equal to the maximum value 1. For better visualization, we manually set them as missing. Each element in the matrix is the Rand index for a pair of trials. The first 11 trials were obtained during the learning stage and the last 10 trials were from the relearning stage. *Upper Left*: Rand index based on calcium traces; *Upper Right*: Rand index using estimated spike trains (Gaussian filtered with $\sigma = 1/3$ s); *Lower Left*: Rand index using estimated spike trains (binned with a bin size of 1/3 s). *Lower Right*: trace vs spike, with the value in the ith row and jth column being the Rand index between the clusters of the ith trial based on trace and the clusters of the jth trial based on spike data.

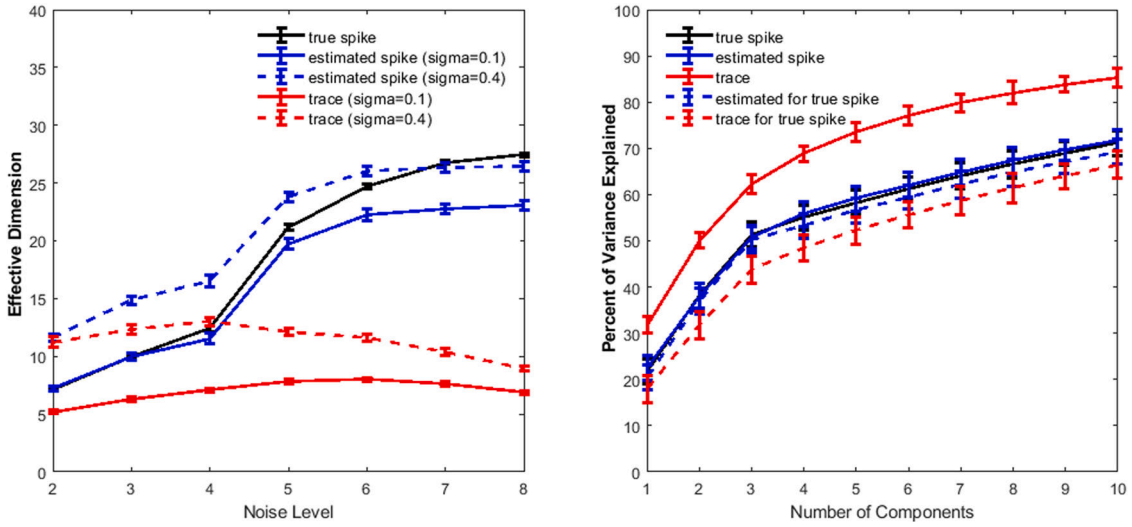

Fig. 5.

Effective dimension and cumulative proportion of variance explained. *Left*: the effective dimension computed using three data types (black color for true spike, blue color for estimated spike, and red for trace) for seven spike noise levels ({2, 3, ⋯, 8}) and two calcium noise levels (solid for $\sigma = 0.1$ and dashed for $\sigma = 0.4$). *Right*: cumulative proportion of variance explained for the data at spike noise level 3 and calcium noise level $\sigma = 0.14$. The solid lines present explained variance of a data type by itself (black for true spike, blue for estimated spike, and red for trace). The blue dashed line is for the variance in true spike that is explained by components from estimated spike data; the red dashed line is the variance in true spike that is explained by components from trace data.

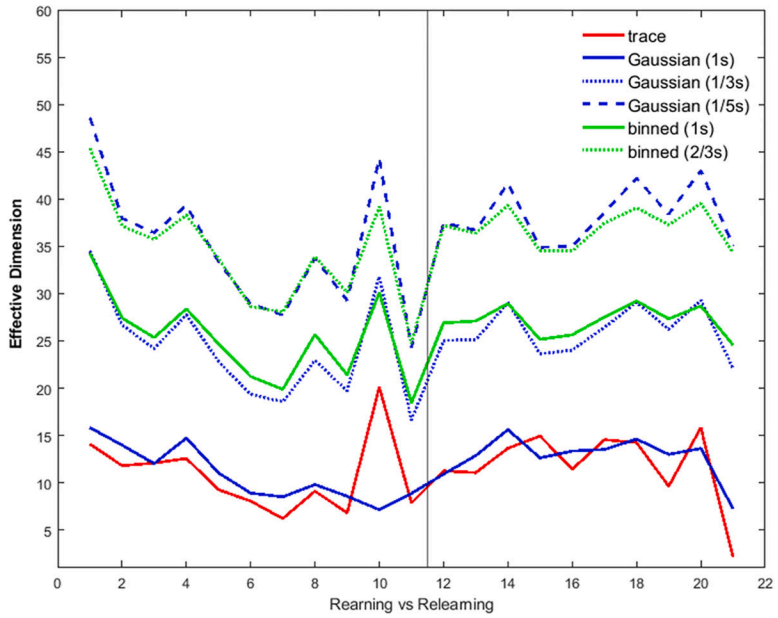

Fig. 6.

Effective dimensions estimated using the fear-based discrimination data. Blue: Gaussian filtered spike data (estimated) with $\sigma \in \{1/5, 1/3, 1\}$ second. Red: trace data. Green: binned spike data (estimated) with a size of 2/3 s or 1 s.

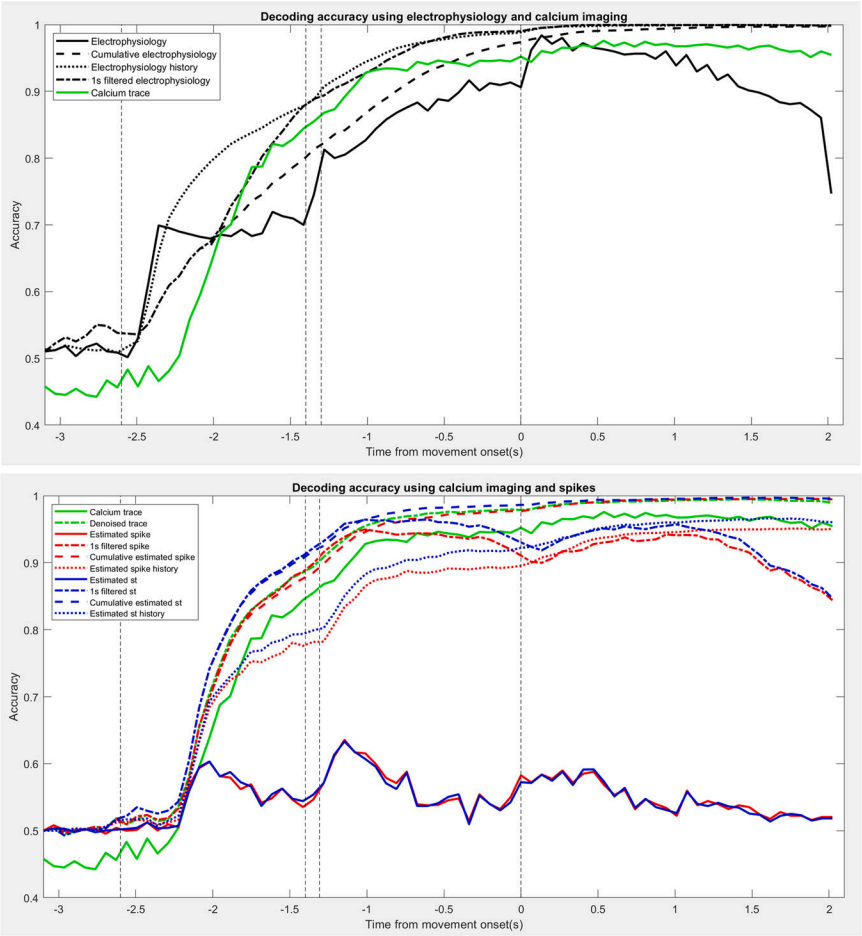

Fig. 7.

Decoding results for the water lick data. The decoding accuracy is averaged over 100 subsamples. In each subsample, 50 neurons are randomly selected and 500 trials are sampled from each neuron. Vertical dashed lines from left to right: the start time of sample epoch, delay epoch of calcium imaging data, delay epoch of electrophysiology, and response epoch. *Top*: decoding accuracy using electrophysiology and calcium imaging data. Solid black: instantaneous decoding using electrophysiology data; dashed black: cumulative decoding using electrophysiology data; dotted black: decoding based on electrophysiology history. Dash-dotted line: 1 s filtered decoding using electrophysiology data. Green line: calcium trace. *Bottom*: decoding accuracy using calcium traces and estimated spike data. Solid green: calcium traces; dot-dashed green: denoised traces. Solid red: instantaneous decoding using estimated spikes; dotted red: spike history; dashed red: cumulative spikes; dot-dashed red: 1 s filtered spike data. Solid blue: instantaneous decoding using spike magnitude; dotted blue: spike magnitude history; dashed blue: cumulative spike magnitude; dot-dashed blue: 1 s filtered spike magnitude.

Table 1

The noise levels in the simulated spike trains from Fellous (2004). X is the extra number of random spikes added to per spike train; J is the standard deviation when jittering the spike times. The labels of the noise levels from Fellous (2004) are used.

Noise level	2	3	4	5	6	7	8
X	2	3	4	8	11	15	20
J	1	3	5	10	15	20	30

Table 2

Population decoding methods. The “1 s filtered”, “cumulative”, and “history” methods are excluded for trace data due to lack of justification.

	Calcium trace	**Estimated trace *c* (*t*)**	**Spike train**	***s* (*t*)**
instantaneous	✓	✓	✓	✓
1 s filtered			✓	✓
cumulative			✓	✓
history			✓	✓