CosMIC: A Consistent Metric for Spike Inference from Calcium Imaging

Stephanie Reynolds[1,2], **Simon R Schultz**[2,3] **and Pier Luigi Dragotti**[1]
[1]Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
[2]Centre for Neurotechnology, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
[3]Department of Bioengineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

Keywords: Two-photon calcium imaging, Spike detection, Spike train similarity.

Abstract

In recent years, the development of algorithms to detect neuronal spiking activity from two-photon calcium imaging data has received much attention. Meanwhile, few researchers have examined the metrics used to assess the similarity of detected spike trains with the ground truth. We highlight the limitations of the two most commonly used metrics, the spike train correlation and success rate, and propose an alternative, which we refer to as CosMIC. Rather than operating on the true and estimated spike trains directly, the proposed metric assesses the similarity of the pulse trains obtained from convolution of the spike trains with a smoothing pulse. The pulse width, which is derived from the statistics of the imaging data, reflects the temporal tolerance of the metric. The final metric score is the size of the commonalities of the pulse trains as a fraction of their average size. Viewed through the lens of set theory, CosMIC resembles a continuous Sorensen-Dice coefficient — an index commonly used to assess the similarity of discrete, presence/absence data. We demonstrate the ability of the proposed metric to discriminate the precision and recall of spike train estimates. Unlike the spike train correlation, which appears to reward overfitting, the proposed metric score is maximised when the correct number of spikes have been detected. Furthermore, we show that CosMIC discriminates the temporal precision of estimates more consistently than either the success rate or the spike train correlation.

1 Introduction

Two-photon calcium imaging has enabled neuronal population activity to be monitored in vivo in behaving animals (Dombeck et al., 2010; Peron et al., 2015). Modern microscope design allows neurons to be imaged at sub-cellular resolution in volumes spanning multiple brain areas (Sofroniew et al., 2016). Coupled with the current generation

of fluorescent indicators (Chen et al., 2013), which have sufficient sensitivity to read out single spikes, this imaging technology has great potential to further our understanding of information processing in the brain.

The fluorescent probe, however, does not directly report spiking activity. Rather, it reads out a relatively reliable indicator of spiking activity — a cell's intracellular calcium concentration — from which spike times must be inferred. A diverse array of techniques have been proposed for this task, including deconvolution approaches (Vogelstein et al., 2010; Friedrich et al., 2017; Pachitariu et al., 2017), methods that identify the most likely spike train given a signal model (Vogelstein et al., 2009; Deneux et al., 2016) and approaches that exploit the sparsity of the underlying spike train (Oñativia et al., 2013). To enable the investigation of neural coding hypotheses, reconstructed spike trains must have sufficient temporal precision for analysis of synchrony between neurons and behavioural variables (Huber et al., 2012), whilst accurately inferring the rate of spiking activity.

Although the development of spike detection algorithms has received a lot of recent attention, few researchers have examined the metrics used to assess an algorithm's performance. At present, there is no consensus in the best choice of metric. In fact, 44% of papers presenting a new method assess its performance using a metric unique to that paper (Fig. 1). This inconsistency impedes progress in the field — algorithms are not directly comparable and, consequently, data collectors cannot easily select the optimal algorithm for a new dataset.

The two most commonly used metrics, the spike train correlation (STC) and the success rate, are not well-suited to the task. The STC, which is invariant under linear transformations of the inputs, is not able to discriminate the similarity of the spike rates of two spike trains (Paiva et al., 2010). Moreover, the temporal binning that occurs prior to spike train comparison, impairs the STC's ability to compare spike train synchrony (Paiva et al., 2010). These limitations suggest that the STC, whilst a quick and intuitive method, is not appropriate for assessing an algorithm's spike detection performance. The success rate, which does accurately compare spike rates, does not reward increasing temporal precision above a given threshold. Consequently, it is not an appropriate metric for evaluating an algorithm's performance when the end goal is, for example, to investigate the synchrony of activations within a network.

In this paper, we present a metric that can discriminate both the temporal and rate precision of an estimated spike train with respect to the ground truth spike train. Unlike the STC, we do not bin the spike trains. Rather, spike trains are convolved with a smoothing pulse that allows comparison of spike timing with an implicit tolerance. The similarity between the resulting pulse trains is subsequently assessed. This type of continuous approach is also preferred by metrics assessing the relationship between spike trains from different neurons (van Rossum, 2001; Schreiber et al., 2003). We set the pulse width to reflect the temporal precision that an estimate is able to achieve given the statistics of the dataset. As such, the metric is straightforward to implement since there are no parameters to tune. For convenience, we refer to the proposed metric as CosMIC — a Consistent Metric for spike Inference from Calcium imaging. In the following, we demonstrate CosMIC's ability to discriminate spike train similarity on simulated data. We include comparisons against the two most commonly used metrics, the spike train correlation and the success rate.

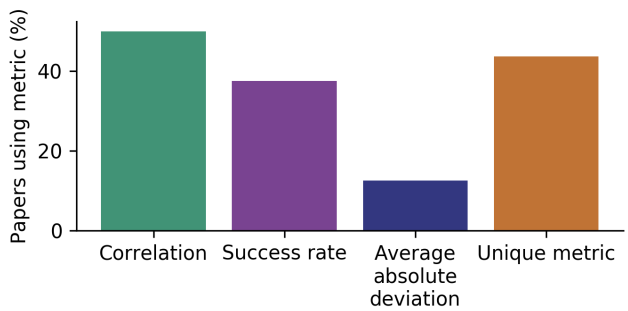

Figure 1: There is no consensus in the best metric for scoring spike train estimates. This chart illustrates the number of papers, of the 16 presenting a spike detection algorithm, that use each metric to assess their algorithm. Of the surveyed papers, 44% use a metric unique to that paper and half use more than one metric. Under the umbrella of the success rate, we include the true and false positive rates, from which the success rate is defined, and the error rate — the complement of the success rate.

2 Constructing the metric

In this paper, we present a metric for comparing the similarity of two sets of spikes: a ground truth set, $\mathrm{S} = \{t_k\}_{k=1}^{K}$, and a set of estimates, $\hat{\mathrm{S}} = \{\hat{t}_k\}_{k=1}^{\hat{K}}$. Due to limiting factors, such as noise and model mismatch, it is improbable that an estimate will match a true spike with infinite temporal precision. As such, we do not expect that $\hat{t}_j = t_k$ for any j or k. Rather, we wish to reward estimates within a reasonable range of accuracy given the limitations of the data. We achieve this by leveraging results from fuzzy set theory (Zimmermann, 2010).

In contrast to classical sets, to which an element either belongs or does not belong, fuzzy sets contain elements with a level of certainty represented by a membership function — the higher the value of the membership function, the more certain the membership. In the following, we define two fuzzy sets, S_ϵ and $\hat{\mathrm{S}}_\epsilon$, which represent the original sets of spikes, S and $\hat{\mathrm{S}}$, with a level of temporal tolerance defined by a parameter ϵ. We set ϵ to reflect the temporal precision that an estimate is able to achieve given the statistics of a dataset (see Section 3). The corresponding membership functions $y(t)$ and $\hat{y}(t)$, which are defined for $t \in \mathbb{R}$, are calculated through convolution of the spike trains,

$$x(t) = \sum_{k=1}^{K} \delta(t - t_k) \quad \text{and} \quad \hat{x}(t) = \sum_{k=1}^{\hat{K}} \delta(t - \hat{t}_k), \tag{1}$$

with a triangular pulse. The resulting functions have local maxima at the locations of the respective sets of spikes (Fig. 2A). As $x(t)$ and $\hat{x}(t)$ are analogous to the membership functions of the classical sets of spikes, we can think of the convolution as a temporal smoothing of the membership. The pulse that we employ is a triangular B-spline (Fig. 2B),

$$p_\epsilon(t) = \begin{cases} \frac{\epsilon - |t|}{\epsilon} & |t| \leq \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

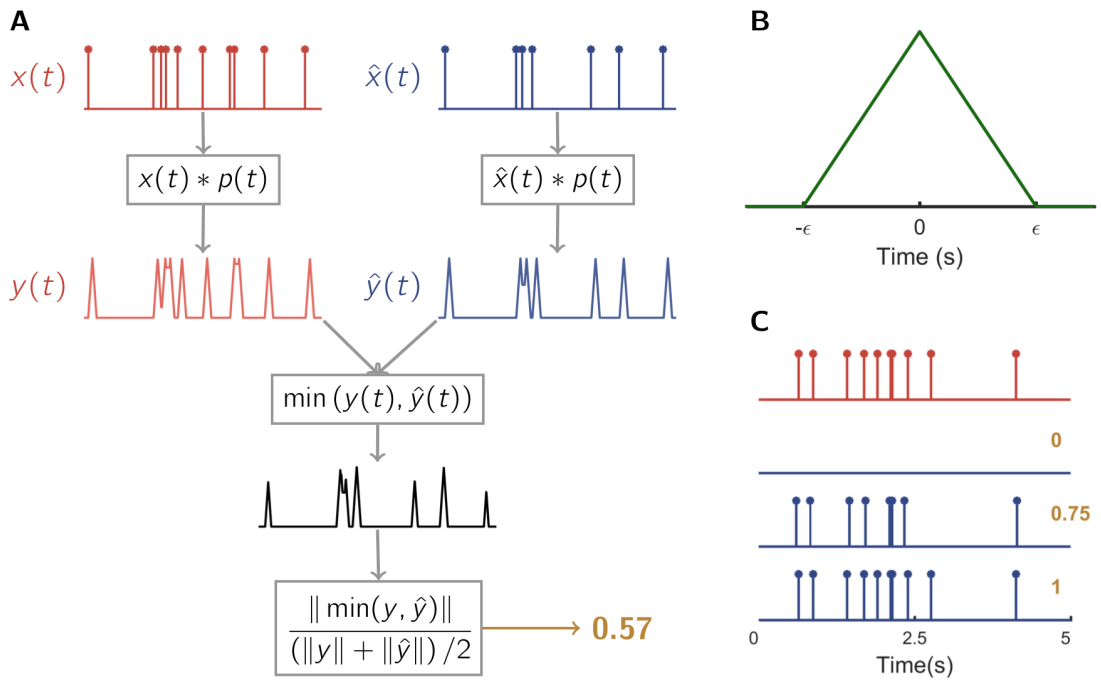

Figure 2: A flow diagram of the proposed metric. The ground truth spike train and estimated spike train are convolved with a triangular pulse (**B**), whose width is determined by the statistics of the data. The metric compares the difference between the resulting pulse trains (**A**). Metric scores are in the range [0,1] — a perfect estimate achieves score 1 and an empty spike train is scored 0 (**C**).

Using this triangular pulse means that, the further a time point, t, is from a spike, the less weight the membership function receives at that point. Past a certain distance (ϵ), the membership function receives no weight. Many pulse shapes could be chosen to introduce this grading of temporal precision, we select a triangular pulse as it is straightforward to examine analytically and implement computationally.

We design the proposed metric to quantify the size of the intersection of the fuzzy sets of true and estimated spikes with respect to the average size of the sets, such that

$$M\left(\mathrm{S}, \hat{\mathrm{S}}\right) = \frac{\mu(\mathrm{S}_\epsilon \cap \hat{\mathrm{S}}_\epsilon)}{\left(\mu\left(\mathrm{S}_\epsilon\right) + \mu(\hat{\mathrm{S}}_\epsilon)\right)/2}, \tag{3}$$

where μ is the L1-norm: $\mu(\mathrm{S}_\epsilon) = \|y\| = \int_{\mathbb{R}} |y(t)| \, \mathrm{d}t$. An analogous formula was presented for discrete fuzzy sets by Pappis and Karacapilidis (1993). Our formula can be interpreted as the continuous version of the Sørensen-Dice coefficient (Dice, 1945; Sørensen, 1948) — a score which is commonly used to assess the similarity of discrete, presence/absence data. In the context of spike detection, the Sørensen-Dice coefficient is referred to as the success rate, see Section 4.1.

The membership function of an intersection of sets is the minimum of their respective membership functions. It follows that

$$\mu(\mathrm{S}_\epsilon \cap \hat{\mathrm{S}}_\epsilon) = \|\min(y, \hat{y})\| = \int_{\mathbb{R}} |\min(y(t), \hat{y}(t))| \, \mathrm{d}t. \tag{4}$$

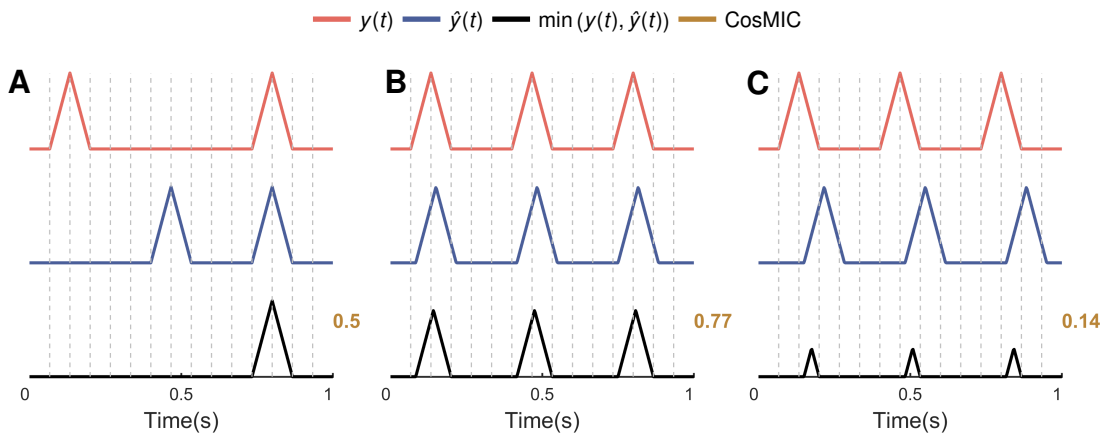

Figure 3: The proposed metric quantifies the commonalities of the sets of true and estimated spikes as a proportion of the average size of those sets. Commonalities are found by taking the minimum of the pulse trains — as such, spikes that appear in only one pulse train are excluded (**A**) and estimates with lower temporal precision receive a lower score (**B** and **C**).

Taking the minimum of the membership functions produces a conservative representation of the intersection of two sets; in our context, spikes that appear in one spike train and not in the other are removed (Fig. 3A) and spikes that are detected with poor temporal precision are assigned less weight (Fig. 3B and 3C).

The metric can also be written in alternative form

$$M\left(\mathrm{S}, \hat{\mathrm{S}}\right) = 1 - \frac{\|y - \hat{y}\|}{\|y\| + \|\hat{y}\|}, \tag{5}$$

the derivation of which is shown in Appendix A.1. From Eq. (5), it is clear that the maximal score of 1 is achieved when the membership functions, and therefore the sets of true and estimated spikes, are equivalent. The minimal score of 0 is achieved when the support of the membership functions do not overlap, i.e. no estimates are within the tolerance of the metric (Fig. 2C).

3 Temporal error tolerance

The width of the triangular pulse with which the spike trains are convolved reflects the accepted tolerance of an estimated spike's position with respect to the ground truth (Fig. 2). To set this width, we calculate a lower bound on the temporal precision of the estimate of one spike — the Cramér-Rao bound (CRB) — from the statistics of the data. The CRB reports the lower bound on the mean square error of any unbiased estimator (Kay, 1993). It is therefore useful as a benchmark; an estimator that achieves the CRB should be awarded a relatively high metric score. In Section 3.1, we detail the calculation of the CRB. In Section 3.2, we outline how we use this bound to determine the pulse width.

3.1 Cramér-Rao bound for spike detection

We consider the problem of estimating the location of one spike, t_0, from noisy calcium imaging data. The fluorescence signal is modelled as

$$f(t) = A\left(e^{-\alpha(t-t_0)} - e^{-\gamma(t-t_0)}\right) 1_{t>t_0}, \tag{6}$$

where α, γ and A are assumed to be known parameters that determine the shape and amplitude of the calcium transient. We assume that we have access to N noisy samples, such that

$$y[n] = f[n] + \xi[n], \quad n \in \{0, 1, ..., N-1\}, \tag{7}$$

where $\xi[n]$ are independent samples of a zero-mean Gaussian process with standard deviation σ and $f[n] = f(nT)$ are samples of the fluorescence signal with time resolution T. The Cramér-Rao lower bound on the uncertainty in the estimated position of t_0 is

$$\text{CRB}(t_0) = \left[\frac{A^2}{\sigma^2}\sum_{n=0}^{N-1}\left(\alpha e^{-\alpha(nT-t_0)} - \gamma e^{-\gamma(nT-t_0)}\right)^2 1_{nT>t_0}\right]^{-1}. \tag{8}$$

This bound was first presented by Schuck et al. (2017). The bound is derived by calculating the inverse of the Fisher Information, which, in the case of samples corrupted by independent, zero-mean, Gaussian noise, is

$$I(t_0) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial f}{\partial t_0}(nT)\right)^2,$$

where $\partial f/\partial t_0$ is the derivative of the fluorescence signal with respect to the spike time, t_0:

$$\frac{\partial f}{\partial t_0}(nT) = A\left(\alpha e^{-\alpha(t-t_0)} - \gamma e^{-\gamma(t-t_0)}\right) 1_{nT>t_0}.$$

In this work, we use the CRB to set the temporal tolerance of the metric. In order that the CRB holds for an arbitrarily placed spike, we remove the dependency on the true spike time by averaging the result over several values of t_0. We compute $\sigma^2_{\text{CRB}} = \frac{1}{M}\sum_{m=1}^{M}\text{CRB}(t_0^m)$, where t_0^m are evenly placed in the interval $(nT, (n+1)T)$ for a fixed n. In Fig. 4, we plot σ_{CRB} as the sampling rate and peak signal-to-noise ratio (PSNR) of the data vary. The PSNR is computed as A^2/σ^2, where A is the amplitude of the fluorescence signal and σ is the standard deviation of the noise. For this example, we use $\alpha = 3.18$ and $\gamma = 34.49$; the parameters for a Cal-520 AM pulse (Tada et al., 2014). We see that the CRB decreases as either the scan rate or the PSNR of the data increases.

3.2 Pulse width

The CRB can be used as a benchmark for temporal precision of any unbiased estimator. As such, we set the pulse width to ensure that, on average, an estimate at the precision

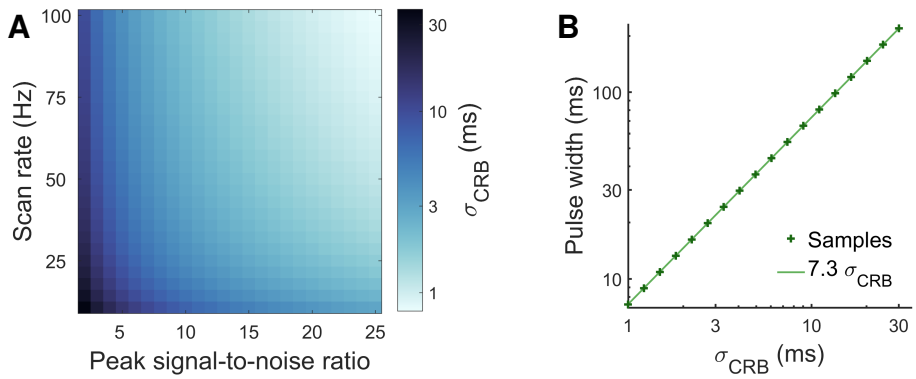

Figure 4: The pulse width is set to reflect the temporal precision achievable given the statistics of the dataset. We calculate the Cramér-Rao bound (CRB), σ^2_{CRB}, a lower bound on the mean square error of the estimated location of one spike from calcium imaging data (**A**). This bound decreases as the scan rate (Hz) and peak signal-to-noise ratio (squared calcium transient amplitude/noise variance) increase. We set the pulse width to ensure that an estimate of one spike at the temporal precision of the CRB achieves, on average, a score of 0.8. This results in a pulse width of approximately 7.3 σ_{CRB} (**B**).

of the CRB achieves a relatively high score. We set the benchmark metric score at 0.8, as this represents a relatively high value in the range of the metric, which is between 0 and 1.

We consider a true spike at t_0 and an estimate, U, normally distributed around it at the precision of the CRB, such that $U \sim \mathcal{N}(t_0, \sigma^2_{\text{CRB}})$. Then, we fix the pulse width so that, on average, $\mathbb{E}\left[M(t_0, U)\right] = 0.8$. In Appendix A.3, we show that this condition is satisfied when

$$0.4 = (\Phi(1/\beta) - 0.5)\left(\beta^2 + 1\right) + \frac{\beta}{\sqrt{2\pi}}\left(\exp(-1/2\beta^2) - 2\right), \tag{9}$$

where $\beta = \sigma_{\text{CRB}}/w$, w is the pulse width and Φ denotes the cumulative distribution function of the standard normal distribution. We see that, the pulse width that solves this equation is approximately equal to $7\sigma_{\text{CRB}}$ (Fig. 4B).

3.3 Implementation

In the following, we provide some practical advice on the implementation of CosMIC. Code to implement the metric can be found at github.com/stephanierey/metric along with a demonstration. In order to use the metric, one must have estimates of the fluorescence signal parameters, $\{\alpha, \gamma, A, \sigma\}$, see Eq. (6). The parameters that define the speed of the rise and decay of the pulse — α and γ — are defined by the fluorescent indicator that generated the imaging data. In Table 1, we provide the values of α and γ for four commonly used fluorescent indicators, extracted from the corresponding references: Cal-520 AM (Tada et al., 2014), OGB-1 AM (Lütcke et al., 2013), GCaMP6f and GCaMP6s (Chen et al., 2013). We note that the time taken for a calcium transient to rise to its peak and the decay time are functions of both α and γ; the values presented

Fluorescent indicator	α	γ
GCaMP6f	4.88	60.97
GCaMP6s	1.26	15.16
OGB-1 AM	1.5	101.5
Cal-520 AM	3.18	34.39

Table 1: To calculate CosMIC's pulse width, the parameters that define the speed of rise and decay of the calcium transient, α and γ, are required. Here, we provide the values of these parameters for four commonly used fluorescent indicators.

in Table 1 are thus not easily interpretable in terms of the shape of a calcium transient pulse.

It is typically necessary for a spike detection algorithm to know the value of the amplitude parameter, A, in order to detect spikes. If, however, this is not the case, we recommend that the parameter is fit from the data samples and the signal model:

$$y(t) = b(t) + A \sum_{k=1}^{K} \left(e^{-\alpha(t-t_k)} - e^{-\gamma(t-t_k)} \right) 1_{t>t_k}, \tag{10}$$

where $b(t)$ is a baseline component and α, γ and $\{t_k\}_{k=1}^{K}$ are known. When the baseline component is constant and there is no indicator saturation, this is a linear problem: $\mathbf{y} = A\mathbf{g} + b$, where we write $\mathbf{y}$ and $\mathbf{g}$ as vectors of data and signal model samples, respectively. The standard deviation of the noise, σ, can be computed as the sample standard deviation of a portion of the data in which there were no calcium transients.

4 Numerical experiments

To assess the discriminative ability of CosMIC, we simulate true and estimated spike trains in various informative scenarios. We compare CosMIC with the two most commonly used metrics, which we define in Sections 4.1 and 4.2 for completeness. In the following, we fix the metric parameters so that their scores are comparable (Fig. 5).

4.1 Success rate

The success rate, which is defined as a function of the true and false positive rates or, alternatively, as a function of precision and recall, appears in various forms in the literature. Spike inference performance has been assessed using true and false positive rates (Rahmati et al., 2016), precision and recall analysis (Reynolds et al., 2017) and using the complement of the success rate, the error rate (Deneux et al., 2016). We study this class of metrics under the umbrella of the success rate, which we define here.

A ground truth spike is deemed to have been 'detected' if there is an estimate within $\delta_1/2$ (s) of that spike, where δ_1 is a free parameter. Only one estimate can be deemed to detect one ground truth spike. The recall is the percentage of ground truth spikes that were detected. The precision is the percentage of estimates that detect a ground truth

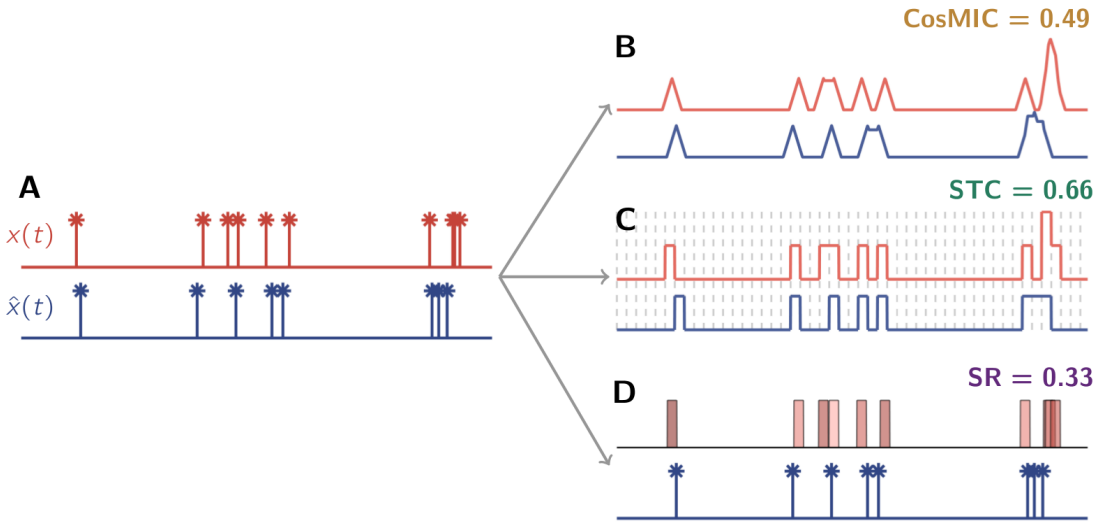

Figure 5: In this paper, we compare the scores of three metrics: CosMIC, the spike train correlation (STC) and the success rate (SR). None of the metrics compute scores directly from the true and estimated spike trains, $x(t)$ and $\hat{x}(t)$, shown in **A**. Rather, CosMIC initially convolves the spike trains with a triangular pulse (**B**). The STC first discretizes the temporal interval and utilises the counts of spikes in each time bin, the bin edges and counts are plotted in **C**. The SR uses a bin centered around each true spike — an estimate in that bin is deemed a true detection (**D**). In order that the metric scores are comparable, we fix the STC and SR bin widths to be half the width of CosMIC's pulse.

spike. Then, the success rate is the harmonic mean of the precision and recall, such that

$$\text{success rate} = 2\,\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \tag{11}$$

A binary true detection region centred around each ground truth spike is analogous to an implementation of CosMIC with a box function pulse. To ensure that the success rate 'pulse' has the same norm as CosMIC's pulse, we set $\delta_1 = \epsilon$, where ϵ is half the pulse width, see Eq. (2).

4.2 Spike train correlation

The first step in the calculation of the spike train correlation (STC) is the discretization of the temporal interval into bins of width δ_2. Two vectors of spike counts, $\mathbf{c}$ and $\hat{\mathbf{c}}$, are subsequently produced, whose i^{th} elements equal the number of spikes in the ith time bin for the true and estimated spike trains, respectively. The STC is the Pearson product-moment correlation coefficient of the resulting vectors, i.e.

$$\text{STC} = \frac{\langle \mathbf{c} - m(\mathbf{c}), \hat{\mathbf{c}} - m(\hat{\mathbf{c}}) \rangle}{\sqrt{v(\mathbf{c})}\sqrt{v(\hat{\mathbf{c}})}}, \tag{12}$$

where $\langle \cdot, \cdot \rangle$, $m(\cdot)$ and $v(\cdot)$, represent the inner product, sample mean and sample variance, respectively. To remain consistent with the success rate, in all numerical experi-

ments, we define $\delta_2 = \delta_1 = \epsilon$.

The STC takes values in the range [-1, 1]. In practise, however, it is rare for a spike detection algorithm to produce an estimate that is negatively correlated with the ground truth (Berens et al., 2017). Moreover, an estimate with maximal negative correlation is equally as informative as one with maximal positive correlation. In this paper, we utilise the normalised spike train correlation, the absolute value of the STC. This ensures that the range of each metric that we analyse is equivalent (and equal to [0,1]) and that, as a consequence, the distribution of metric values are comparable.

5 Results

To investigate the properties of CosMIC and the two most commonly used metrics, the spike train correlation and success rate, we simulate estimated and ground truth spike trains and analyse the metric scores. To mimic the temporal error in spike time estimates, unless otherwise stated, the estimates are normally distributed about the true spike times. We refer to the standard deviation of the normal distribution as the 'jitter' of the estimates. We use the standard deviation at the Cramér-Rao bound, σ_{CRB}, as a benchmark for the minimum attainable jitter a spike train estimate can achieve (see Section 3.1).

5.1 CosMIC penalises overfitting

As opposed to the spike train correlation, CosMIC penalises overfitting (Fig. 6). We simulate spike train estimates that are normally distributed about the true spike times, which are chosen with replacement when the number of detected spikes ($\hat{K}$) exceeds the number of true spikes (K) . We observe that, rather than penalising overfitting, the spike train correlation increases with the overfitting ratio ($\hat{K}/K$). In contrast, CosMIC peaks when the correct number of spikes are detected. This behaviour is consistent as the jitter of the estimated spikes is varied; in this example, the jitter is σ_{CRB} (Fig. 6A), $3\ \sigma_{\text{CRB}}$ (Fig. 6B) and $5\ \sigma_{\text{CRB}}$ (Fig. 6C), respectively.

In Fig. 6, at each level of jitter and overfitting, we plot the mean and standard deviation of metric scores over 100 realizations of spike train estimates. It can be seen that CosMIC is more consistent than the spike train correlation; it exhibits much smaller variance in the scores of spike train estimates of the same jitter and overfitting ratio.

5.2 CosMIC evaluates temporal precision consistently

CosMIC scores spike train estimates of the same temporal precision more consistently than the spike train correlation or success rate (Fig. 7). First, we illustrate this at the level of estimates of a single spike, t_0. CosMIC depends only on the absolute difference between the estimate, $\hat{t}_0$, and the true spike — the further the distance, the smaller the score. The relationship between CosMIC and the temporal error, $\delta = |t_0 - \hat{t}_0|$, is

$$M(t_0, \hat{t}_0) = \begin{cases} \left(\frac{|\delta|}{w} - 1\right)^2 & \text{if} \quad |\delta| < w \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

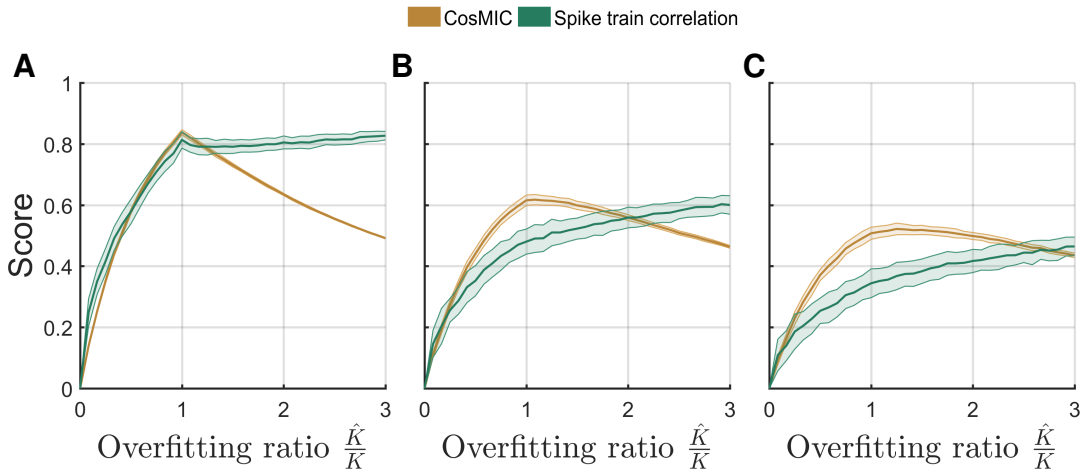

Figure 6: In contrast to the spike train correlation, CosMIC decreases as the level of overfitting increases. In this simulated example, estimated spikes are normally distributed about the true spikes, with jitter σ_{CRB} (**A**), 3 σ_{CRB} (**B**) and 5 σ_{CRB} (**C**), respectively. Here, σ_{CRB} is the root mean square error at the Cramér-Rao bound. When the number of estimated spikes ($\hat{K}$) is greater than the number of true spikes (K), estimates are distributed around a set of locations chosen with replacement from the set of true spikes. For each metric we plot the mean (darker central line) and standard deviation (edges of shaded region) of metric scores on a set of 100 spike train estimates generated at each overfitting and temporal precision combination.

where w is the width of the pulse. The derivation of this result is given in Appendix A.2. The success rate, on the other hand, does not reward precision above the bin width. Moreover, the spike train correlation is asymmetric in the temporal error: estimates the same distance from the true spike are not guaranteed to be awarded the same score. This asymmetry stems from the correlation's temporal discretization, its extent depends on the location of the true spike with respect to the bin edges.

The inconsistent scoring of temporal precision that is characteristic of both the correlation and success rate at the level of a single spike results in inconsistent scoring of spike train estimates. In Fig. 7B, we display the mean and standard deviation of metric scores awarded to spike train estimates that detect the correct number of spikes with varying jitter. Over 100 realizations of spike train estimates at each temporal precision, we observe significantly smaller standard deviation of CosMIC when compared to the spike train correlation and success rate. From this, we infer that spike trains of the same temporal precision are being scored most consistently by CosMIC.

5.3 CosMIC evaluates recall and precision consistently

By construction, CosMIC bears a strong resemblance to the Sørensen-Dice coefficient, which, in the context of spike detection, is referred to as the success rate. The success rate is the harmonic mean of the precision and recall, two intuitive metrics which represent the proportion of true spikes detected and the proportion of estimates not in the set of true spikes. In this section, we demonstrate that CosMIC can accurately discriminate

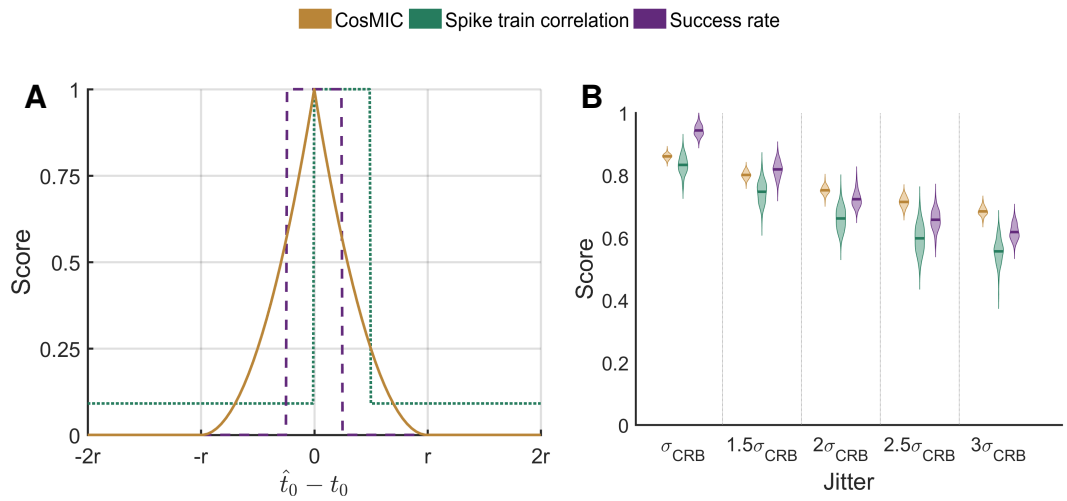

Figure 7: CosMIC scores spike train estimates of the same temporal precision more consistently than the spike train correlation (STC) or success rate (SR). Unlike the STC, the proposed metric awards estimated spikes ($\hat{t}_0$) with the same proximity from the true spike (t_0) the same score (**A**). In contrast to both the STC and SR, CosMIC rewards increasing precision above the pulse radius (r) with strictly increasing scores. When scoring spike train estimates of the correct number of spikes but varying root mean square error ('jitter'), these characteristics mean that CosMIC scores temporal precision more consistently (**B**). In **B**, the distribution of metric scores are shown for 100 realizations of spike train estimates per jitter level, where estimates are normally distributed around the true spike trains with standard deviation a multiple of σ_{CRB}, the standard deviation at the Cramér-Rao bound.

both the precision and recall of spike train estimates.

When a spike train estimate detects exactly a subset of the true spikes, plus no remainders, CosMIC and the success rate depend only on the percentage of true spikes detected (the recall) and not the location of that subset, see Fig. 8A and 8D. Denoting the size of the subset of true detections as $K - R$, with K the number of true spikes and $0 \leq R \leq K$, we have

$$M\left(\mathrm{S}, \hat{\mathrm{S}}\right) = 1 - \frac{1}{2K/R - 1}, \tag{14}$$

see Appendix A.4 for a proof. Thus, CosMIC depends only on the proportion of 'missing' spikes, R/K, not their location. In contrast, the spike train correlation exhibits significant variation at each level of recall. This is illustrated in Fig. 8A, in which we plot the distribution of CosMIC, success rate and correlation scores over 100 realizations of spike train estimates at each level of recall. It can be seen that, in this setting, CosMIC and the success rate are fixed with the recall of the spike train estimates.

When all the true spikes are exactly detected plus $R \geq 0$ surplus spikes, CosMIC and the success rate depend only on the level of precision not the location of the surplus spikes, see Fig. 8B and 8E. We have

$$M(\mathrm{S}, \hat{\mathrm{S}}) = \frac{1}{1 + R/2K}, \tag{15}$$

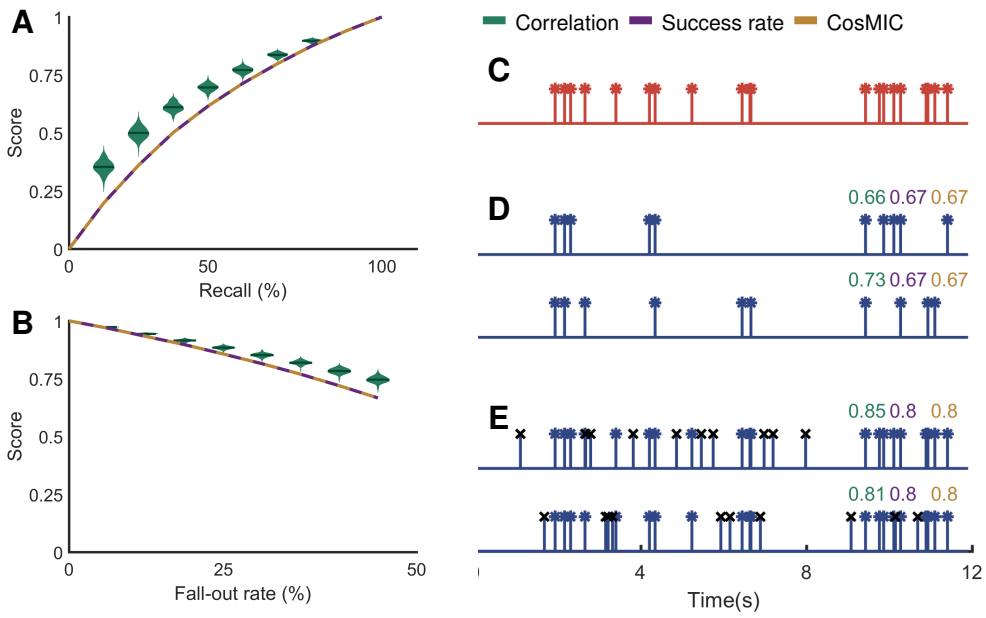

Figure 8: CosMIC scores estimated spike trains of the same recall and fall-out rate consistently, unlike the spike train correlation (STC). When a spike train estimate detects precisely the location of a subset of spikes from a true spike train, the scores of CosMIC and the success rate depend only on the percentage of spikes detected (the recall), not the location of the detected spikes (**A, D**). In contrast, the STC varies with the subset of spikes that were detected. When a spike train estimate detects all the true spikes precisely plus a number of uniformly distributed surplus spikes, the STC varies with the placement of the surplus spikes (**B, E**). In contrast, the success rate and CosMIC depend only on the percentage of over-detection (the fall-out rate). The distribution of correlation scores plotted in **D** and **E** stem from 100 realizations of estimated spike trains at each recall and fall-out rate. In **C**, we plot an example of a true spike train. In **D** and **E**, we plot estimated spike trains, with a recall and fall-out rate of 50% and 33%, respectively, along with the corresponding metric scores. The spikes with a black 'x' marker in **E** indicate the surplus spikes.

where K is the number of true spikes, see Appendix A.5 for a proof. The correlation, on the other hand, varies with the location of the surplus spikes. In Fig. 8E, we plot the distribution of the correlation scores for 100 realizations of spike train estimates at each level of precision. CosMIC and the success rate, which are constant (and identical) at a given precision, in this scenario, are also shown.

6 Discussion

Much recent attention has been focused on the development of algorithms to detect spikes from calcium imaging data, while the suitability of the metrics that assess those algorithms have been predominantly overlooked. In this paper, we presented a novel metric ('CosMIC') to assess the similarity of a spike train estimate to the ground truth spike train. Our results demonstrate that CosMIC accurately and consistently discrimi-

nates both the temporal and rate precision of estimates compared to the ground truth.

Using two-photon calcium imaging, the activity of neuronal populations can be monitored in vivo in behaving animals. Inferred spike trains can be used to investigate neural coding hypotheses, by analysing the rate and synchrony of neuronal activity with respect to behavioural variables. To justify such analysis, the ability of spike detection algorithms to generate accurate spike train estimates must be verified. In this paper, we have found the current generation of spike inference metrics to be lacking and have proposed an alternative. Unlike the spike train correlation, which appears to reward overfitting, CosMIC is maximised when the correct number of spikes are detected (Fig. 6). Moreover, CosMIC rewards increasing temporal precision more consistently than the spike train correlation and success rate (Fig. 7).

The current inconsistency in the metrics used to assess spike detection algorithms, hinders both experimentalists, aiming to select an algorithm for data analysis, and developers. In light of this problem, a recent benchmarking study tested a range of algorithms on a wide array of imaging data (Berens et al., 2017). Although informative, the study, which relied heavily on the spike train correlation to assess algorithm performance, may not provide the full picture. By introducing a new metric, we hope to complement such efforts in the pursuit of a thorough, quantitative evaluation of spike inference algorithms.

By construction, CosMIC bears a resemblance to the Sørensen-Dice coefficient, which is commonly used to compare discrete, presence/absence data (Dice, 1945; Sørensen, 1948). This metric is widely used in many fields, including ecology (Bray and Curtis, 1957) and image segmentation (Zou et al., 2004). When applied to spike inference, this coefficient is referred to as the success rate and is one of the two most commonly used metrics. We have demonstrated that this construction confers some of the advantages of the success rate to CosMIC. In particular, CosMIC is able to accurately discriminate the precision and recall of estimated spike trains (Fig. 8). We have also shown the advantages of CosMIC over the success rate; most importantly, it evaluates a spike train estimate's temporal precision more consistently than the success rate (Fig. 7). Furthermore, the CosMIC's parameter is defined with respect to the statistics of the dataset; unlike the success rate's bin size, it does not need to be selected by a user.

The width of the pulse is computed from a lower bound on temporal precision (Section 3), which, in turn, is derived from the statistics of the dataset. As a result, the metric will be more lenient for spike inference algorithms on noisier or lower sampling rate data. This is due to our assumption that a metric score should reflect the difficulty of the spike inference problem. To calculate the bound, knowledge of the calcium transient pulse parameters and the standard deviation of the noise are required. These parameters are typically used by algorithms in the spike detection process (Vogelstein et al., 2010; Deneux et al., 2016).

The problem of comparing a ground truth and estimated spike train is analogous to that of comparing spike trains from different neurons. In the spike train metric literature, binless measures have been found to outperform their discrete counterparts (Paiva et al., 2010). It is also common to convolve spike trains with a smoothing pulse prior to analysis (van Rossum, 2001; Schreiber et al., 2003). In that context, the width and shape of the pulse reflect hypotheses about the relationship between neuronal spike trains. In future work, CosMIC's pulse could be adapted to a one-sided or binary kernel

in order to explore its application to neuronal spike train comparison.

Finally, we note that the developed metric is able to accurately assess an estimate's temporal and rate precision. This information is unified in a single score that summarises the overall performance of an algorithm. We consider a 1D summary score to be practical for users who do not have the time or desire to analyse multi-dimensional trade-offs. Those that prefer to analyse the source of errors may, however, prefer to use CosMIC alongside other metrics, such as the mean square error and the true and false positive rates.

Acknowledgments

This work was supported by European Research Council starting investigator award [grant number 277800] (Pier Luigi Dragotti); Biotechnology and Biological Sciences Research Council [grant number BB/K001817/1] (Simon R. Schultz); EU Marie Curie FP7 Initial Training Network [grant number 289146] (Simon R. Schultz).

A Appendices

In the appendices, we provide derivations of some results presented in the main text. The following notation is consistent throughout. We denote with $x(t)$ and $\hat{x}(t)$ the true and estimated spike trains, see Eq. (1). We denote the triangular smoothing pulse with $p_\epsilon(t)$, see Eq. (2). The true and estimated pulse trains are denoted $y(t) = x(t) * p_\epsilon(t)$ and $\hat{y}(t) = \hat{x}(t) * p_\epsilon(t)$, respectively. The proposed metric score, when comparing the similarity between a ground truth set of spikes, S = $\{t_k\}_{k=1}^{K}$, with a set of estimates, $\hat{\text{S}}$ = $\{\hat{t}_k\}_{k=1}^{\hat{K}}$, is

$$M(\text{S}, \hat{\text{S}}) = 2\frac{\|\min(y, \hat{y})\|}{\|y\| + \|\hat{y}\|}, \tag{A.16}$$

where $\|\cdot\|$ is the L1-norm.

A.1 Alternative metric form

In the following, we derive an alternative equation for CosMIC; we show that

$$M(\text{S}, \hat{\text{S}}) = 1 - \frac{\|y - \hat{y}\|}{\|y\| + \|\hat{y}\|}. \tag{A.17}$$

We have

$$\begin{aligned} 1 - \frac{\|y - \hat{y}\|}{\|y\| + \|\hat{y}\|} &= \frac{\|y\| + \|\hat{y}\| - \|y - \hat{y}\|}{\|y\| + \|\hat{y}\|} \\ &= \frac{\int_{\mathbb{R}} y(t) + \hat{y}(t)\,\mathrm{d}t - \int_{\mathbb{R}} |y(t) - \hat{y}(t)|\,\mathrm{d}t}{\|y\| + \|\hat{y}\|}, \end{aligned}$$

where we have used the fact that $y(t)$ and $\hat{y}(t)$ are non-negative for all $t \in \mathbb{R}$. Decomposing both integrals over $\mathbb{R}$ into their counterparts over the disjoint sets $\{t \in \mathbb{R} : y(t) > \hat{y}(t)\}$ and $\{t \in \mathbb{R} : y(t) \leq \hat{y}(t)\}$ and subsequently combining them, we have

$$1 - \frac{\|y - \hat{y}\|}{\|y\| + \|\hat{y}\|} = \frac{2\int_{y>\hat{y}} \hat{y}(t)\mathrm{d}t + 2\int_{y\leq\hat{y}} y(t)\mathrm{d}t}{\|y\| + \|\hat{y}\|} = \frac{2\|\min(y, \hat{y})\|}{\|y\| + \|\hat{y}\|} = M(\mathrm{S}, \hat{\mathrm{S}}).$$

Eq. (A.17) then follows.

A.2 Score for estimate of one spike

We now derive an expression for the metric score of the estimate of the location of one spike in terms of the temporal error of the estimate, $|u|$. We see that, as the temporal precision increases above the threshold precision (ϵ), the metric score increases monotonically.

Proposition 1. *The score given to an estimate of the location of a single spike, t_0, with temporal error $u \in \mathbb{R}$ is*

$$M(t_0, t_0 + u) = \begin{cases} \left(\frac{|u|}{2\epsilon} - 1\right)^2 & \textit{if} \quad |u| < 2\epsilon \\ 0 & \textit{otherwise}, \end{cases} \tag{A.18}$$

where ϵ is half the width of the pulse, $p_\epsilon(t)$, as in Eq. (2)*.*

Proof. Without loss of generality, we let the true spike location be at $t_0 = 0$, as the metric score depends on the relative rather than absolute locations of the estimated and ground truth spikes. From Eq. (A.16), we have

$$M(0, u) = 2\frac{\|\min(p_\epsilon(t), p_\epsilon(t-u))\|}{\|p_\epsilon(t)\| + \|p_\epsilon(t-u)\|}.$$

When $|u| > 2\epsilon$, the pulses do not overlap and, consequently, the numerator is equal to 0. Therefore, the metric score is zero for all $|u| > 2\epsilon$. For $|u| \leq 2\epsilon$, we write

$$M(0, u) = \frac{1}{\epsilon}\left(\int_A p_\epsilon(t)\,\mathrm{d}t + \int_B p_\epsilon(t-u)\,\mathrm{d}t\right), \tag{A.19}$$

which follows from $\|p_\epsilon\| = \epsilon$, $A = \{t \in \mathbb{R} : p_\epsilon(t) < p_\epsilon(t-u)\}$ and $B = \{t \in \mathbb{R} : p_\epsilon(t) \geq p_\epsilon(t-u)\}$. From the change of variables $v = t + u$, we see that $M(0, u) = M(0, -u)$. As M is even in the second argument, we must only calculate $M(0, u)$ for $0 < u < 2\epsilon$. To identify the support of A and B, we must identify the point at which $p_\epsilon(t) = p_\epsilon(t-u)$. We have

$$p_\epsilon(t) = p_\epsilon(t-u) \Leftrightarrow 1 - \frac{|t|}{\epsilon} = 1 - \frac{|t-u|}{\epsilon} \Leftrightarrow |t| = |t-u|.$$

For $0 < u < 2\epsilon$, the intersection point occurs in the right half of $p_\epsilon(t)$ and the left half of $p_\epsilon(t-u)$, it follows that $t = u/2$. Eq. (A.19) becomes

$$\begin{aligned} M(0,u) &= \frac{1}{\epsilon}\left(\int_{u/2}^{\epsilon} p_\epsilon(t)\,\mathrm{d}t + \int_{u-\epsilon}^{u/2} p_\epsilon(t-u)\,\mathrm{d}t\right) \\ &= \frac{1}{\epsilon}\left(\int_{u/2}^{\epsilon} p_\epsilon(t)\,\mathrm{d}t + \int_{-\epsilon}^{-u/2} p_\epsilon(v)\,\mathrm{d}v\right) \\ &= \frac{2}{\epsilon}\int_{u/2}^{\epsilon} p_\epsilon(t)\,\mathrm{d}t, \end{aligned}$$

which follows from the change of variables $v = t + u$ and the symmetry of $p_\epsilon(t)$ about 0. Evaluating the integral, we obtain $M(0,u) = (|u|/2\epsilon - 1)^2$, for $|u| < 2\epsilon$.

□

A.3 Metric score at precision of CRB

The Cramér-Rao bound (CRB) is commonly used as a benchmark for algorithm performance in parameter estimation problems. In the context of calcium imaging, it has been previously used to evaluate detectability of spikes under different imaging modalities (Reynolds et al., 2015; Schuck et al., ress). In this case, the CRB reports the minimum uncertainty achievable by any unbiased estimator when estimating the location of one spike. We thus set the width of the pulse to ensure that, on average, an estimate of the location of one spike at the precision of the CRB achieves a metric score of 0.8. This benchmark score is relatively high in the range of the metric, which is between 0 and 1, whilst allowing leeway to be exceeded.

Proposition 2. *Let t_0 denote the location of the true spike. The estimate is normally distributed about the true spike at the precision of the CRB, it is modelled with the random variable $U \sim \mathcal{N}(t_0, \sigma^2_{CRB})$. We denote $\beta = \sigma_{CRB}/w$, where w is the pulse width. Then, we have $\mathbb{E}[M(t_0, U)] = 0.8$ if β satisfies the following equation,*

$$0.4 = \Phi(1/\beta)\left(\beta^2 + 1\right) + \frac{\beta}{\sqrt{2\pi}}\left(\exp(-\beta^2/2) - 2\right), \tag{A.20}$$

where Φ denotes the cumulative distribution function of the standard normal distribution.

Proof. We want to identify the pulse width at which $0.8 = \mathbb{E}[M(t_0, U)]$. Without loss of generality, we consider the case where $t_0 = 0$. Due to the fact that $M(0,\cdot)$ is even and the results of Appendix A.2, we have

$$\begin{aligned} \mathbb{E}[M(0,U)] &= \int_{\mathbb{R}} M(0,u) f(u)\mathrm{d}u \\ &= 2\int_0^w \left(\frac{u}{w} - 1\right)^2 f(u)\mathrm{d}u \\ &= \frac{2}{w^2}\int_0^w u^2 f(u)\mathrm{d}u - \frac{4}{w}\int_0^w u f(u)\mathrm{d}u + 2\int_0^w f(u)\mathrm{d}u \\ &= \frac{2}{w^2} I_1 - \frac{4}{w} I_2 + 2 I_3, \end{aligned}$$

where $f(\cdot)$ is the probability density function of U. Applying integration by parts to I_1, we obtain

$$I_1 = \sigma_{\text{CRB}}^2 \Pr\left(U \in [0, w)\right) - \sigma_{\text{CRB}}^2 f(w) w.$$

The remaining integrals are $I_2 = -\sigma_{\text{CRB}}^2 (f(w) - f(0))$ and $I_3 = \Pr\left(U \in [0, w)\right)$, respectively. Putting the integrals together:

$$\mathbb{E}\left[M(0, U)\right] = 2\left[\Pr\left(U \in [0, w)\right)\left(\frac{\sigma_{\text{CRB}}^2}{w^2} + 1\right) + \frac{\sigma_{\text{CRB}}^2}{w}\left(f(w) - 2f(0)\right)\right].$$

Writing $\beta = \sigma_{\text{CRB}}/w$, we have

$$\mathbb{E}\left[M(0, U)\right] = 2\left(\left(\Phi(1/\beta) - 0.5\right)\left(\beta^2 + 1\right) + \frac{\beta}{\sqrt{2\pi}}\left(\exp(-1/2\beta^2) - 2\right)\right).$$

□

A.4 Exact detection of subset of true spikes

We have a set of K true spikes, S, and $\hat{K}$ estimates, $\hat{\text{S}}$. The set of estimates contains a subset of the ground truth spike times with the exception of R missing spikes and no extras, such that $\hat{K} = K - R$ with $0 \leq R \leq K$. Due to the distributivity of the convolution operation

$$\hat{y}(t) = \hat{x}(t) * p_\epsilon(t) = (x(t) - x_r(t)) * p_\epsilon(t) = y(t) - r(t),$$

where $x_r(t)$ and $r(t)$ are the spike train and pulse train, respectively, of the spikes missing from $\hat{\text{S}}$. From the form in Eq. (A.17), the metric score becomes

$$\begin{aligned} M(y, \hat{y}) &= 1 - \frac{\|y - \hat{y}\|}{\|y\| + \|\hat{y}\|} \\ &= 1 - \frac{\|y - (y - r)\|}{\|y\| + \|y - r\|} \\ &= 1 - \frac{R\|p_\epsilon\|}{K\|p_\epsilon\| + (K - R)\|p_\epsilon\|} \\ &= 1 - \frac{1}{2K/R - 1}. \end{aligned}$$

A.5 Exact detection of all true spikes with overfitting

We have a set of K true spikes, S, and $\hat{K}$ estimates, $\hat{\text{S}}$. The set of estimates contains all the ground truth spike times plus $R \geq 0$ extra spikes, such that $\hat{K} = K + R$. Due to the distributivity of the convolution operator, the estimated pulse train can be written

$$\hat{y}(t) = \hat{x}(t) * p_\epsilon(t) = (x(t) + x_r(t)) * p_\epsilon(t) = y(t) + r(t),$$

where $x_r(t)$ and $r(t)$ are the spike train and pulse train, respectively, of the surplus spikes. From the form in Eq. (A.17), the metric score becomes

$$\begin{aligned} M(\mathrm{S}, \hat{\mathrm{S}}) &= \frac{2\|\min(y, \hat{y})\|}{\|y\| + \|\hat{y}\|} \\ &= \frac{2\|\min(y, y + r)\|}{\|y\| + \|y + r\|} \\ &= \frac{2\|y\|}{2\|y\| + \|r\|} \\ &= \frac{1}{1 + \|r\|/(2\|y\|)}, \end{aligned}$$

where the penultimate line follows from the non-negativity of y and r. As $\|y\| = K\|p_\epsilon\|$ and $\|r\| = R\|p_\epsilon\|$, it follows that

$$M(\mathrm{S}, \hat{\mathrm{S}}) = \frac{1}{1 + R/2K}.$$

References

Berens, P., Freeman, J., Deneux, T., Chenkov, N., McColgan, T., Speiser, A., Macke, J. H., Turaga, S., Mineault, P., Rupprecht, P., Gerhard, S., Friedrich, R. W., Friedrich, J., Paninski, L., Pachitariu, M., Harris, K. D., Bolte, B., Machado, T. A., Ringach, D., Reimer, J., Froudarakis, E., Euler, T., Roman-Roson, M., Theis, L., Tolias, A. S., and Bethge, M. (2017). Community-based benchmarking improves spike inference from two-photon calcium imaging data. *bioRxiv*.

Bray, J. R. and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4):325–349.

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K., and Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300.

Deneux, T., Kaszas, A., Szalay, G., Katona, K., Lakner, T., Grinvald, A., Rózsa, B., and Vanzetta, I. (2016). Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nature Communications*, 12190.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L., and Tank, D. W. (2010). Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience*, 13:1433–1440.

Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. *PLOS Computational Biology*, 13(3):1–26.

Huber, D., Gutnisky, D. A., Peron, S., O'Connor, D. H., Wiegert, J. S., Tian, L., Oertner, T. G., Looger, L. L., and Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature*, 484:473–478.

Kay, S. M. (1993). *Fundamentals of statistical signal processing*. Prentice Hall signal processing series. Prentice Hall PTR, Upper Saddle River, NJ.

Lütcke, H., Gerhard, F., Zenke, F., Gerstner, W., and Helmchen, F. (2013). Inference of neuronal network spike dynamics and topology from calcium imaging data. *Frontiers in Neural Circuits*, 7:201.

Oñativia, J., Schultz, S. R., and Dragotti, P. L. (2013). A finite rate of innovation algorithm for fast and accurate spike detection from two-photon calcium imaging. *Journal of Neural Engineering*, 10(4):046017.

Pachitariu, M., Stringer, C., and Harris, K. D. (2017). Robustness of spike deconvolution for calcium imaging of neural spiking. *bioRxiv*.

Paiva, A. R. C., Park, I., and Príncipe, J. C. (2010). A comparison of binless spike train measures. *Neural Computing and Applications*, 19(3):405–419.

Pappis, C. P. and Karacapilidis, N. I. (1993). A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems*, 56(2):171 – 174.

Peron, S. P., Freeman, J., Iyer, V., Guo, C., and Svoboda, K. (2015). A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783 – 799.

Rahmati, V., Kirmse, K., Marković, D., Holthoff, K., and Kiebel, S. J. (2016). Inferring neuronal dynamics from calcium imaging data using biophysical models and bayesian inference. *PLOS Computational Biology*, 12(2):1–42.

Reynolds, S., Abrahamsson, T., Schuck, R., Sjöström, P. J., Schultz, S. R., and Dragotti, P. L. (2017). ABLE: An activity-based level set segmentation algorithm for two-photon calcium imaging data. *eNeuro*, 4(5).

Reynolds, S., Oñativia, J., Copeland, C. S., Schultz, S. R., and Dragotti, P. L. (2015). Spike detection using FRI methods and protein calcium sensors: performance analysis and comparisons. In *11th international conference on Sampling Theory and Applications (SampTA 2015)*, Washington, DC, USA.

Schreiber, S., Fellous, J., Whitmer, D., Tiesinga, P., and Sejnowski, T. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54(Supplement C):925 – 931.

Schuck, R., Go, M. A., Garasto, S., Reynolds, S., Dragotti, P. L., and Schultz, S. (in press). Multiphoton minimal inertia scanning for fast acquisition of neural activity signals. *Journal of Neural Engineering*.

Sofroniew, N. J., Flickinger, D., King, J., and Svoboda, K. (2016). A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *eLife*, 5:e14472.

Sørensen, T. J. (1948). *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. København: I kommission hos E. Munksgaard.

Tada, M., Takeuchi, A., Hashizume, M., Kitamura, K., and Kano, M. (2014). A highly sensitive fluorescent indicator dye for calcium imaging of neural activity in vitro and in vivo. *European Journal of Neuroscience*, 39(11):1720–1728.

van Rossum, M. C. W. (2001). A novel spike distance. *Neural Computation*, 13(4):751–763.

Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691–3704.

Vogelstein, J. T., Watson, B. O., Packer, A. M., Yuste, R., Jedynak, B., and Paninski, L. (2009). Spike inference from calcium imaging using sequential monte carlo methods. *Biophysical Journal*, 97(2):636–655.

Zimmermann, H.-J. (2010). Fuzzy set theory. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):317–332.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., and Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index: Scientific Reports. *Academic Radiology*, 11(2):178 – 189.