

# Modelling the fluorescence produced by fluorescent calcium indicators in response to arbitrary sequences of action potentials

Thomas J. Delaney  
Cian O'Donnell, Michael C. Ashby

### Abstract

The use of fluorescent calcium indicators, such as GCaMP6s, to monitor neuronal activity is widespread. But the relationship between the fluorescence signal and action potential firing is poorly understood. Furthermore, the effects of the indicator characteristics on this fluorescence signal are unknown. For example, it is known that genetically encoded indicators accumulate within neurons over weeks and months. This makes comparison of activity levels at different time points difficult. As a result, whether or not spike train inference is always possible using fluorescent calcium indicators remains unknown.

The aim of this project was to model the fluorescence traces produced by a fluorescent calcium indicator in a neuron soma, given parameters such as binding rate, dissociation rate, and molecular concentration from a specific spike train. The ultimate goal of the model is to allow benchmarking of the various spike inference algorithms that have been developed, and to understand how indicator characteristics affect the quality of spike train inference.

The modelled cell contents consisted of free calcium, fluorescent indicator molecules, and mobile and immobile endogenous calcium buffers. The indicator molecules which were bound to a calcium molecule could be either excited, i.e. able to release a photon, or relaxed. In order to reproduce the noise in the system dynamics, we modelled the release of photons from the excited indicator bound calcium as a stochastic process.

The fluorescence traces produced by the simulation were calibrated to reproduce the signal-to-noise ratio observed in experimental data. Spike inference algorithms were used to infer spike trains from the experimental fluorescence traces and the modelled fluorescence traces. The parameters of the model were then varied in order to determine the effect on the system dynamics and the effects on spike inference.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Fluorescent calcium indicators . . . . .	4
1.2	Spike Inference Algorithms . . . . .	5
<b>2</b>	<b>Methods &amp; Data Sources</b>	<b>7</b>
2.1	Modelling . . . . .	7
2.1.1	Calcium Dynamics . . . . .	7
2.1.2	The Model . . . . .	7
2.1.3	Photon release & capture . . . . .	8
2.1.4	Parameter optimisation . . . . .	9
2.1.5	Julia . . . . .	11
2.2	Spike Inference . . . . .	11
2.2.1	Perturbation analysis . . . . .	13
2.3	Data sources . . . . .	13
<b>3</b>	<b>Results</b>	<b>14</b>
3.1	Concentration & Fluorescence Modelling . . . . .	14
3.1.1	Buffered Calcium Concentrations . . . . .	14
3.1.2	Fluorescence Modelling . . . . .	15
3.2	Spike Inference . . . . .	16
3.2.1	Observed fluorescence vs Modelled fluorescence . . . . .	16
3.2.2	Perturbation analysis . . . . .	18
<b>4</b>	<b>Discussion</b>	<b>19</b>

# 1 Introduction

## 1.1 Fluorescent calcium indicators

Since the invention of two-photon calcium imaging, the use of fluorescent calcium indicators to study the activity of populations of neurons has grown in popularity. In this process, a calcium sensitive dye is injected into the body of each cell to be monitored. Or the cells are altered by a virus or genetically modified to make the cells express a calcium sensitive protein. The dye or the proteins fluoresce when they come into contact with and bind to calcium molecules. This fluorescence allows us to locate the cells containing the fluorescent indicator. Furthermore, since the firing of an action potential involves an influx of calcium molecules, the indicator makes the firing of action potentials observable by an increase and subsequent decay in fluorescence inside the cell.

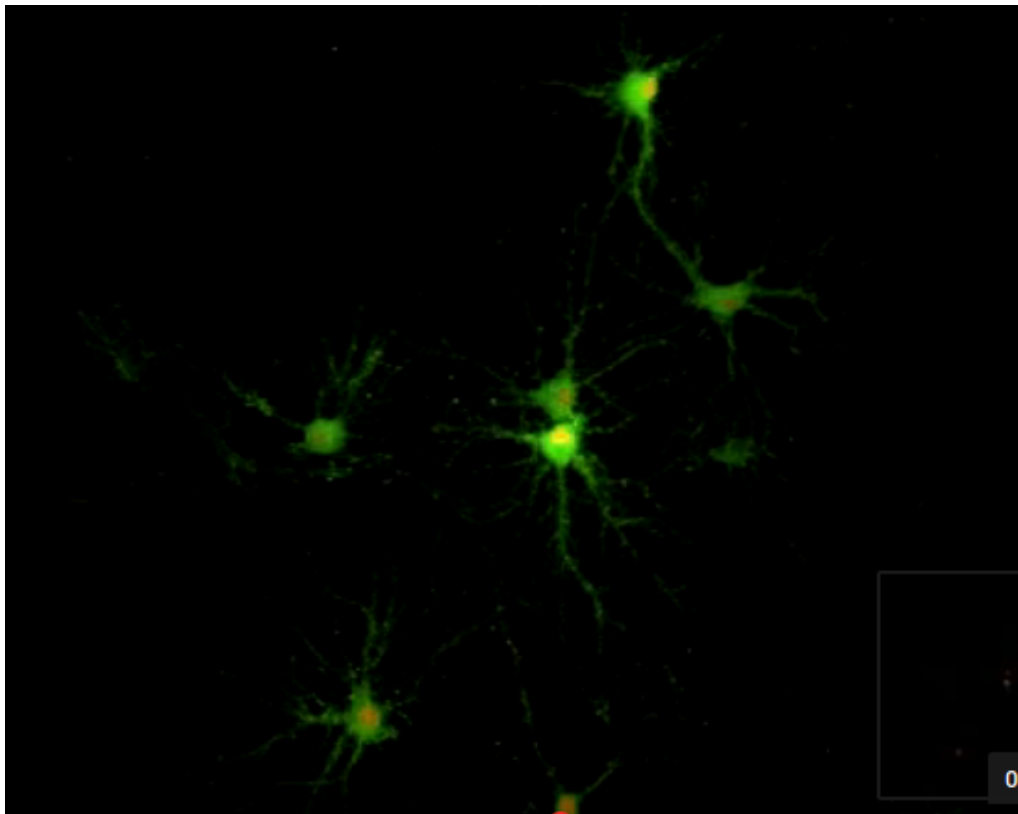


Figure 1: Single frame from video of spontaneous activity recorded from cultured mouse hippocampal neurons using GCaMP6s fluorescent indicator. Image courtesy of [brainslicemethods.com](http://brainslicemethods.com)

Calcium imaging has a number of advantages over the more traditional electrophysiological methods. Firstly, electrophysiological methods involve inserting an electrode or a number of electrodes into the brain to measure extra-cellular voltages. The electrodes need to be attached by wires to a device that collects their data. This inhibits the movement of the experimental subject, and makes longitudinal experiments impossible. The calcium imaging can be performed *in vivo* [2], and the apparatus can be removed from the subject when the subject is not under observation.

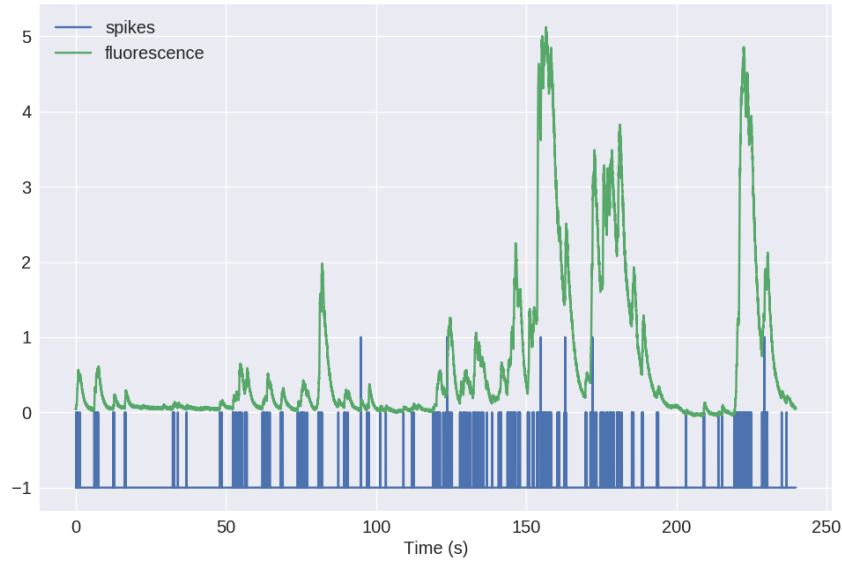
This enables longitudinal experiments as the subject can survive and move around normally when not being examined. Furthermore, inserting these electrodes can cause cells, or the connections between cells to be damaged. Using calcium imaging avoids this kind of damage. Also, in electrophysiology, the strength of the spiking signal, and the position of each electrode are used by purpose designed algorithms to identify the position of individual neurons and ‘sort’ the spikes, i.e. attribute the spikes to a cell. This method is very good for spike detection and attribution. But these methods are biased toward readings from the most active cells. In fact, an inactive cell will be completely ignored by electrophysiological methods. With calcium imaging, inactive cells are still detected since each cell contains a baseline level of calcium [8], and therefore fluoresces a small amount. Finally, electrophysiological methods can only read from the cells nearby the electrodes. Genetically encoded fluorescent calcium indicators such as GCaMP6s can be targeted at neurons with a specific genetic make-up.

fMRI methods can also be used to monitor brain activity. But the resolution is too low to monitor individual cells.

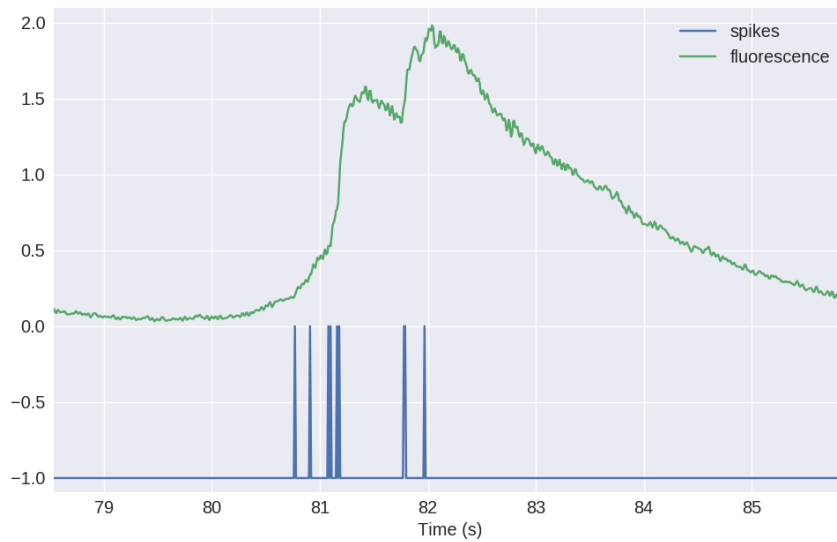
There are some drawbacks to fluorescent calcium indicators. For example, the dynamics of the fluorescent indicators are much slower than the dynamics of action potentials. After excitation, the cross-membrane potential of a cell will return to baseline in 2 – 3ms. After a single action potential, the fluorescence of GCaMP6s will return to baseline after approximately 3s. At least a thousand times longer. An example of this can be seen in a figure 2, showing a fluorescence trace and the corresponding spike train. The relaxation time is extended when more than one action potential is fired in quick succession [3]. Furthermore, readings are usually taken from cells after a ‘loading period’, so that the indicator is saturated within the cells. At this point the increase in fluorescence in response to an action potential will be greatest. So another issue with the genetically encoded indicators is that it is impossible to know whether or not the indicator is saturated within the cell. This is especially a problem with longitudinal studies, as the amplitude of fluorescence change in response to an action potential will change as the experiment continues. This means that the same amount of activity in cells will cause different amounts of fluorescence changes at different times, which makes analysis difficult. It is claimed that despite this, it is still possible to detect individual spikes even in spike bursts [3].

## 1.2 Spike Inference Algorithms

There are existing methods for converting the fluorescence traces read during these experiments into spike trains [4, 5, 6, 7], but these methods have not been tested rigorously. Also, these methods do not take experimental variations into account, such as a variation in the concentration of the indicator. Furthermore, the process of collecting the photons emitted by the indicator introduces some noise. Whether or not this noise is enough to obfuscate the spiking signal has never been investigated. A model of the fluorescence induced by a spike train is required to assess whether or not it is actually possible to infer an accurate spike train from a fluorescence trace. A model of this kind could also be used to test the methods that infer a spike train from a fluorescence trace.



(a) Plot of a spike train and the corresponding GCaMP6s fluorescence trace. Data courtesy of [spikefinder.codeneuro.org](https://spikefinder.codeneuro.org)



(b) The same image as (a) but zoomed into the period from 80s to 86s. A group of six action potentials around the 81s point followed by a group of three action potentials just before the 82s point are shown. The decay of the fluorescence trace is much slower than the dynamics of an action potential.

Figure 2

## 2 Methods & Data Sources

### 2.1 Modelling

#### 2.1.1 Calcium Dynamics

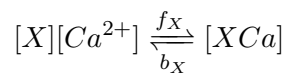
When a neuron fires an action potential, voltage-dependent calcium ion-channels open up that allow a current of  $\text{Ca}^{2+}$  to flow into the neuron [9]. The increase in the free calcium ion concentration inside of the cell along with changes in the concentration of potassium and sodium causes the change in cell membrane potential, which must be depolarised. The depolarising process consists of free calcium ions leaving the cell through open ion channels, binding to molecules within the cell called *buffers*, and some calcium storage by organelles such as the endoplasmic reticulum. There are several different types of calcium buffer, each with different dynamics and different concentrations within different types of excitable cell. The fluorescent calcium indicator is another calcium buffer, with the useful property that when it is bound to a calcium ion, the bound molecule may become excited by a photon and release a photon in return. This is what creates the fluorescence. After the action potential has taken place, the free calcium concentration within the cell will return to a baseline level [8].

#### 2.1.2 The Model

We modelled the the dynamics of five molecular concentrations within the body of a neuron,

- Free calcium ion concentration,  $[\text{Ca}^{2+}]$ ,
- Fluorescent indicator bound calcium,  $[\text{BCa}]$ ,
- Endogenous mobile buffer bound calcium,  $[\text{ECa}]$ ,
- Endogenous immobile buffer bound calcium,  $[\text{ImCa}]$ ,
- Excited buffered calcium,  $[\text{BCa}^*]$ .

The term ‘buffering’ refers to free calcium ions coming into contact with buffer molecules and binding together. Diagrammatically this is:

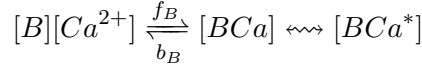


where  $[X]$  represents any buffer molecule, and  $f_X$  and  $b_X$  represent the binding and unbinding (dissociation) rates in units of per molar concentration per second ( $M^{-1}s^{-1}$ ) and per second ( $s^{-1}$ ) respectively. The speed of this chemical reaction is determined by the binding and unbinding rates.

There are a number different endogenous buffers in any neuron. The buffers present, and the buffers’ concentrations vary from cell to cell. In order to capture the effects of mobile and immobile endogenous buffers without introducing several parameters they were modelled as two buffers. One representing mobile buffers and the other representing immobile buffers. Each with their own binding and unbinding rates.

The fluorescent calcium indicator behaves similarly to the other calcium buffers. The calcium is buffered by the indicator in the same way. But an indicator bound calcium molecule can become excited by absorbing the energy from a photon. An

excited indicator bound calcium molecule can then release a photon to go back to its ‘relaxed’ state.



The released photons are captured by a photon collector. This gives us the fluorescence trace.

Ignoring the baseline level of free calcium in a neuron, the system of equations we used to model all of these interactions is as follows:

$$\begin{aligned} \frac{d[Ca^{2+}]}{dt} &= b_i[BCa] + b_E[ECa] + b_{Im}[ImCa] \\ &\quad - f_i[B][Ca^{2+}] - f_E[E][Ca^{2+}] - f_{Im}[Im][Ca^{2+}] \\ \frac{d[BCa]}{dt} &= f_i[B][Ca^{2+}] - b_i[BCa] + r[BCa^*] - e[BCa] \\ \frac{d[ECa]}{dt} &= f_E[E][Ca^{2+}] - b_E[ECa] \\ \frac{d[ImCa]}{dt} &= f_{Im}[Im][Ca^{2+}] - b_{Im}[ImCa] \\ \frac{d[BCa^*]}{dt} &= e[BCa] - r[BCa^*] \end{aligned} \tag{1}$$

where  $e$  is the excitation rate for indicator bound calcium,  $r$  is the photon release rate for the excited indicator bound calcium, and  $f$  and  $b$  are used to indicate the forward and backward rates for chemical reactions respectively. The excitation rate defines the proportion of indicator bound calcium that becomes excited at each time step. The photon release rate defines the proportion of excited indicator bound calcium that releases a photon and returns to its relaxed state at each time step. An action potential is modelled as a discontinuous increase in the free calcium concentration to an appropriate value [8].

The model also includes influx and outflux forces pushing the level of free calcium within the cell back to the baseline level of free calcium as measured by Maravall et al. (2000). Only free calcium, and the calcium bound to endogenous mobile buffer is subject to both the influx and outflux forces. Only the calcium bound to endogenous immobile buffer is subject to the influx force.

### 2.1.3 Photon release & capture

We used a simple model for the photon release. The number of photons released was proportional to the number of excited indicator bound calcium molecules in the cell. The size of that proportion was defined by the ‘release rate’. The release rate is an optimised parameter of the model.

In two-photon excitation microscopy, the photons scattered by the fluorescent indicator get scattered in all directions. Therefore the number of photons detected is stochastic. This made the process for capturing photons the natural source of noise in the system. The number of photons captured, and therefore the intensity of the fluorescence, is modelled using a binomial distribution.

A binomial distribution is the discrete probability distribution of the number of successes in a sequence of independent trials with a binary outcome. The parameters defining the distribution are:

- $n$ , The number of trials



- $p$ , The probability of success

The probability mass function of the distribution is

$$Pr(\text{number of successes} = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (2)$$

The number of photons released was used as number of trials. This was a natural choice as each photon is either captured, or not captured. The probability of success is set equal to the ‘capture rate’. The capture rate is a free parameter of the model to be optimised.

#### 2.1.4 Parameter optimisation

The free parameters of the model are as follows:

**Capture rate** The average proportion of photons captured by the photon detector.

**Influx** When the concentration of calcium within the cell is greater than the baseline, the influx defines what proportion of the extra calcium leaves the cell at each time step.

**Outflux** When the concentration of calcium within the cell is less than the baseline, the outflux defines what proportion of the calcium deficit enters the cell at each time step.

**Excitation rate** The proportion of indicator bound calcium that is excited by photon bombardment at each time step.

**Release rate** The proportion of excited indicator bound calcium that releases a photon at each time step.

To optimise the free parameters given a fluorescence trace, we applied the following procedure:

1. The frequency power spectrum of the trace was measured.
2. The power spectrum was smoothed using a boxcar smoother (aka. sliding average, box smoother).
3. The log of the smoothed power spectrum was measured.
4. Use the model to create a modelled fluorescence trace
5. Apply steps 1, 2, and 3 to the modelled fluorescence trace.
6. Calculate the root mean squared difference between the log power of the actual fluorescence trace, and the log power of the modelled fluorescence trace.
7. Use an optimisation algorithm to reapply this process, attempting to minimize the root mean squared difference at each iteration.

The Nelder-Mead method (aka downhill simplex method) was used for the optimisation. This method is useful for maximising or minimising an objective function in a multidimensional space. Therefore it was ideal for minimising the root mean squared difference that we calculated here. The objective function used was the root mean squared difference between the log power spectra. The procedure was applied ten times, using the optimised parameters from the previous iteration as the starting parameters for the next. The free parameters were optimised for each individual fluorescence trace.

An example of an observed trace, a modelled trace created with optimised parameters, and the smoothed power spectrum of each of these traces can be seen in figure 8.

The trajectory of the mean and standard deviation of the root mean squared difference across all of the twenty-one traces over each of the iterations is shown in figure 3. The root mean squared difference is smaller on average after each iteration, but it never reaches zero for a number of reasons. Firstly, the fluorescence traces carry low frequency fluctuations that cannot be captured by the model (see figure 4). Secondly, the model assumes that the process of calcium binding to the fluorescent indicator is linear in time (see equation 1), but there are more complicated dynamics involved here. Fluorescent calcium indicators are often built upon the calcium binding protein called ‘calmodulin’. This protein has four calcium binding sites. These sites are locally split into two pairs. Each pair has a different affinity for calcium, and the affinity of the binding sites is affected by the occupancy of the other binding sites [10]. So the calcium to calcium indicator binding process is non-linear, but the model does not take this into account.

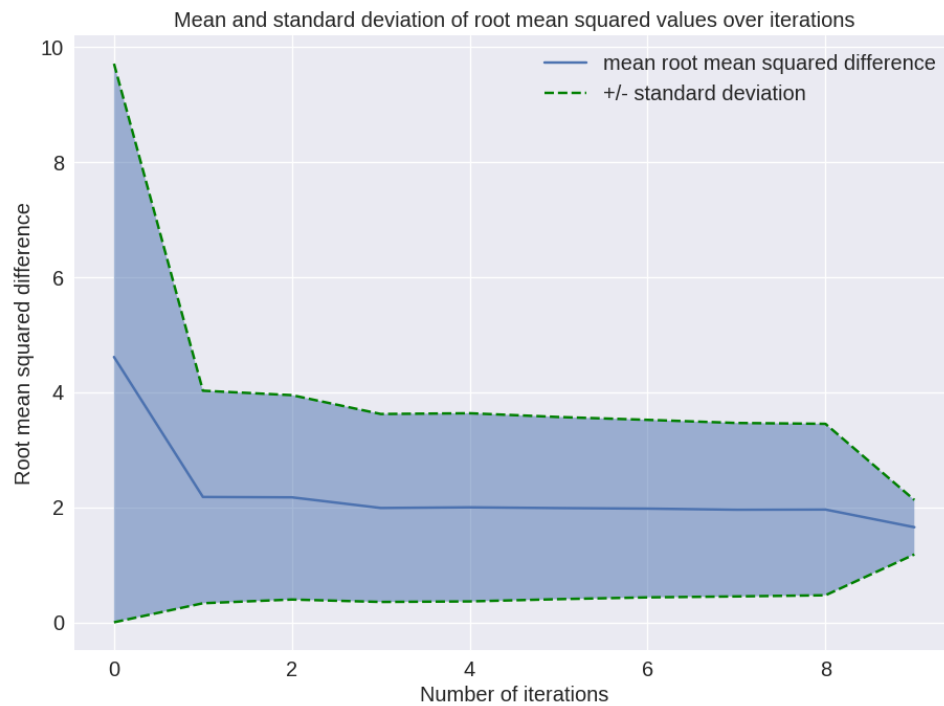


Figure 3: The optimisation procedure was applied ten times, using the optimised parameters from the previous iteration as the starting parameters for the next. We had twenty one traces for which to optimise the parameters. Shown here is the mean and standard deviation (coloured region) of the root mean squared difference over the number of iterations executed. As the number of iterations increases, the means root mean squared difference reduces, and the standard deviation reduces.

### 2.1.5 Julia

The programming language used to write and execute the model was ‘Julia’. Julia is a dynamic programming language designed for technical computing. Julia was designed specifically to provide a convenient high-level dynamic language similar to MATLAB<sup>®</sup>, or Python, with improved performance. Julia’s type system and Julia’s direct interfaces with C and Fortran allow this aim to be achieved [11]. The Julia version of the ‘Sundials’ package for ODE solving was used to solve equations 1.

## 2.2 Spike Inference

We used spike inference algorithms to compare the quality of spike inference using the modelled traces to the quality of spike inference using the observed traces. We also used the spike inference algorithms to assess the effect of parameter perturbation on the spike inference. Three algorithms were used:

**Constrained non-negative deconvolution algorithm (aka Paninski algorithm)** This algorithm uses a constrained version of non-negative Weiner deconvolution to infer a calcium signal and a ‘spiking activity signal’ from the fluorescence trace [5, 4]. The spiking activity signal is a non-negative vector of real numbers reflecting the cell’s activity rather than an actual spike train. We inferred a spike train by choosing an optimised threshold for the spiking activity signal. Whenever the spiking activity signal exceeded that threshold, an action potential was inferred. The threshold was optimised by minimising the difference between the number of spikes observed and the number of spikes predicted.

**$\ell_0$  optimisation algorithm** This algorithm is based on a first order auto-regressive model model of the fluorescence trace, with sparse spiking [6]. By removing the restriction that the contribution from a spike must be non-negative, Jewell et al. (2010) reduce the problem of optimising this model using an  $\ell_0$  norm to a *changepoint detection problem*, for which efficient algorithms already exist [6]. Given a fluorescence trace, this algorithm returns the optimised parameter of the  $\ell_0$  problem, the time constant of the calcium concentration, and an estimated spike train.

**Online Active Set method to Infer Spikes (OASIS)** This algorithm is once again based on an auto-regressive model of the fluorescence trace, but can be generalised to any order. The algorithm itself is a generalisation of the *pool adjacent violators algorithm* (PAVA) that is used in isotonic regression. The OASIS algorithm works through the fluorescence trace from beginning to end, this combined with speed of the algorithm means that it could be used for real-time online spike inference [7]. Given a fluorescence trace, the algorithm will return the most likely spike train and an inferred denoised fluorescence signal.

In order to quantify the quality of spike inference for a given algorithm, we ran that algorithm on all of the fluorescence traces in dataset number eight of the spike finder datasets. Then we measured some binary classification measures on the results. These measures included

- Accuracy
- True positive rate, (aka recall, sensitivity, hit rate)
- True negative rate, (aka specificity)

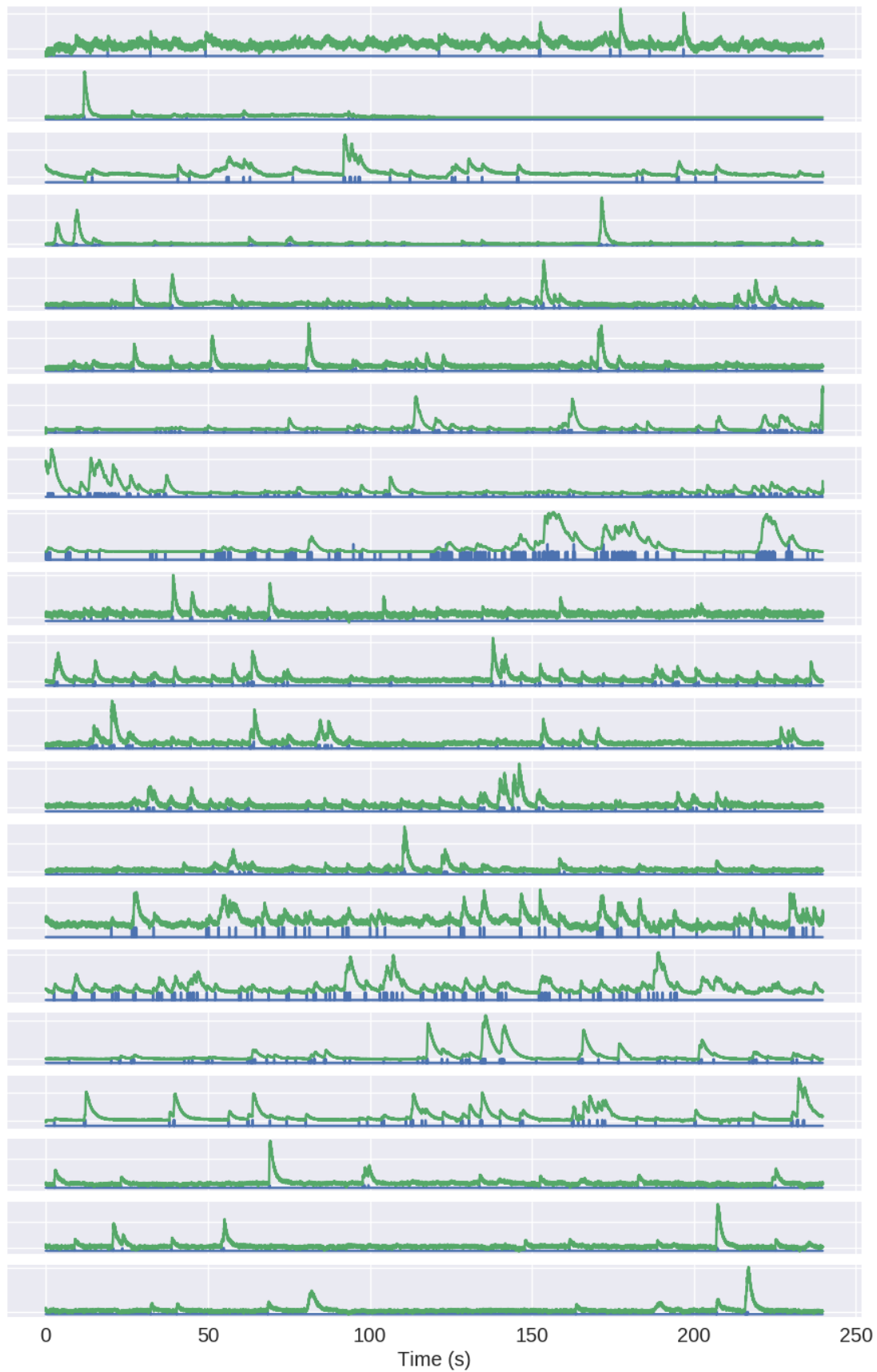


Figure 4: All the spikefinder data. Fluorescence traces in green, spike trains in blue.

- Precision, (aka positive predicted value)
- Negative predicted value,
- False negative rate, (aka miss rate)
- False positive rate, (aka fall-out)
- False discovery rate,
- False omission rate

In making these measurements, we allowed a tolerance of two subsequent time bins for spike prediction. For example, the spike train data is a vector of 0s and 1s, with one element for each time bin. A '0' denotes inactivity, a '1' denotes the presence of at least one action potential. The inferred spike train produced by the spike inference algorithms take the same form. In our analysis, if a spike appeared in the inferred spike train up to two time frames after a spike in the observed spike train, that spike was considered correctly inferred i.e. a true positive. However, once a spike in the inferred spike train was matched to a spike from the observed spike train, the inferred spike could not be matched to another observed spike. To illustrate, if two spikes were inferred in the two time bins following an isolated observed spike, the first inferred spike was considered correctly inferred, but the second inferred spike was considered incorrectly inferred, i.e. a false positive.

The most useful measure was the true positive rate. This is because the spiking is sparse and this measurement is sensitive to the number of spikes observed and inferred. After optimising the parameters for each fluorescence trace we measured the spike inference quality for the observed fluorescence traces, and compared this to the spike inference quality for the modelled traces.

### 2.2.1 Perturbation analysis

In order to measure the sensitivity of spike inference to changes in a given model parameter, we perturbed the parameter and compared the quality of spike inference with the perturbed parameters to the quality of spike inference with the experimental or optimised parameters. In order to maximise the possibility of observing a difference due to the perturbation, we perturbed the chosen parameter by a relatively large amount. For example, the experimental value for the molar concentration of the fluorescent indicator within the cell was  $10^{-4}\text{M}$  [8]. The perturbed values used for this parameter were  $10^{-2}\text{M}$ ,  $10^{-3}\text{M}$ ,  $10^{-5}\text{M}$ , and  $10^{-6}\text{M}$ . The quality of the inference was compared by measuring the true positive rate for each perturbed value and using a *t*-test to compare the distributions of the results.

This analysis was performed firstly without any optimisation of the free parameters for use with the perturbed parameters. Then the analysis was performed after the optimised parameters for each perturbed value were calculated.

## 2.3 Data sources

All of the data used in this project was sourced from the 'Spike Finder' project ([spikefinder.codeneuro.org](http://spikefinder.codeneuro.org)). The data consisted of a collection of datasets with simultaneously measured fluorescence traces and action potentials. All of the fluorescence traces and action potential trains from the dataset that we used most heavily can be seen in figure 4.

## 3 Results

### 3.1 Concentration & Fluorescence Modelling

The spike trains from the Spike Finder project were used as input to the fluorescence model.

#### 3.1.1 Buffered Calcium Concentrations

In order to produce a fluorescence trace, the model first simulates the traces of five chemical concentrations within the cell (detailed in section 2.1.2)  $[Ca^{2+}]$ ,  $[BCa]$ ,  $[ECa]$ ,  $[ImCa]$ ,  $[BCa^*]$ . The trajectory of each of the concentrations over a sample spike train can be seen in figure 5. Each of the action potentials caused a discontinuous increase in the concentration of  $[Ca^{2+}]$  in the cell.

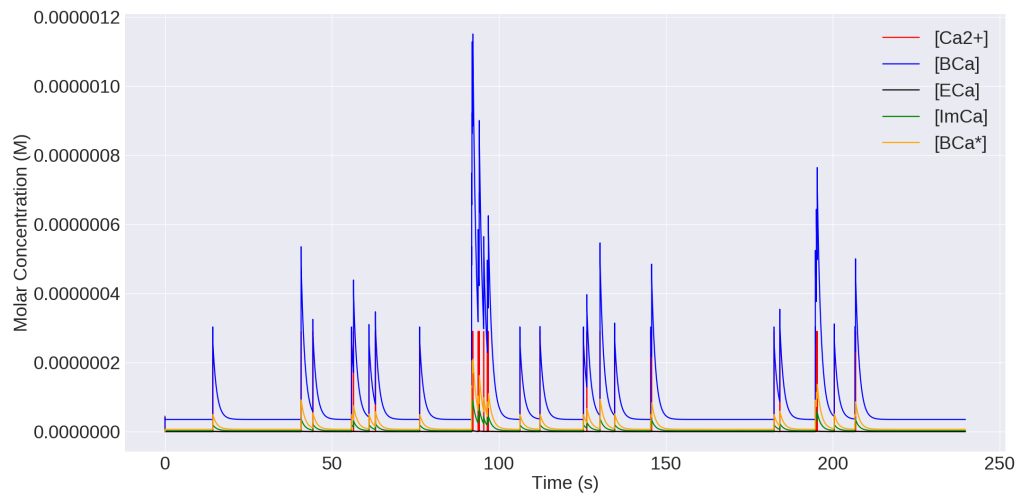


Figure 5: The concentrations of free calcium ions,  $[Ca^{2+}]$ , fluorescent indicator bound calcium,  $[BCa]$ , endogenous buffer bound calcium,  $[ECa]$ , immobile buffer bound calcium,  $[ImCa]$ , and excited fluorescent indicator bound calcium,  $[BCa^*]$ , over the course of a sample trial from the spike finder data. At each action potential there is a discontinuous increase in free calcium concentration, causing a chain reaction in the other concentrations.

The trajectories of  $[BCa]$ ,  $[ECa]$ , and  $[ImCa]$  are all similar. Each of the proteins buffer some of the baseline calcium, so the buffered concentrations settle to a stable baseline in the absence of action potentials. At the onset of an action potential, the jump in free calcium causes a sudden jump in each of these three concentrations. The concentrations then begin to decay very quickly. This very quick decay is not visible in figure 5, but can be observed in the zoomed in version figure 6. Then the decay slows down to a slower rate. The trace of each of these concentrations can be accurately modelled by the sum of two exponential decays, one of which decays much more quickly than the other.

The trajectory of the excited indicator bound calcium,  $[BCa^*]$ , is slightly different. There is a baseline level of this concentration in the absence of an action

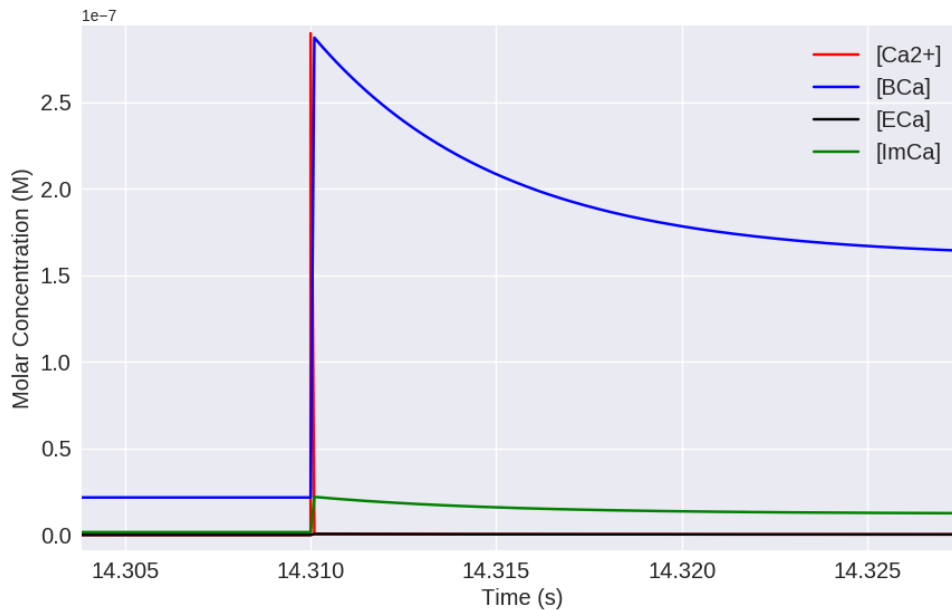


Figure 6: A zoomed in version of figure 5 showing only the  $[Ca^{2+}]$ ,  $[BCa]$ ,  $[ECa]$ , and  $[ImCa]$  concentrations during the period 14.30 – 14.33s. The fast decay dynamics of the buffered concentrations can be seen here. The excess of free calcium that enters the cell at the time of an action potential is immediately buffered by a combination of the fluorescent indicator and the endogenous mobile and immobile buffers.

potential. This is in agreement with experimental observations of a baseline level of fluorescence. The increase in this concentration does not mirror the instantaneous jump in free calcium. Instead, the increase resembles a logarithmic increase up to some limit. After reaching a peak the concentration then begins an exponential decay back to a baseline level. These effects can be observed in figures 5 and 7. The trace of the  $[BCa^*]$  concentration can be accurately modelled by the difference of two exponentials.

### 3.1.2 Fluorescence Modelling

The fluorescence trace is defined by the number of photons released and subsequently captured. The number of photons released was proportional to the number of excited indicator bound calcium molecules in the cell. The size of that proportion was defined by another optimised parameter of the model called the ‘release rate’. The actual fluorescence trace was sampled at each time step from a binomial distribution. For details on the binomial distribution see section 2.1.3. The number of trials in the distribution was equal to the number of photons released from the excited indicator bound calcium molecules,  $[BCa^*]$ , in the cell. The probability of success was an optimised parameter of the model called the ‘capture rate’. The model caused the fluorescence trace to follow the excited indicator bound calcium trace, with some noise associated with the binomial distribution.

An example observed fluorescence trace along side its modelled counterpart can be seen in figure 8

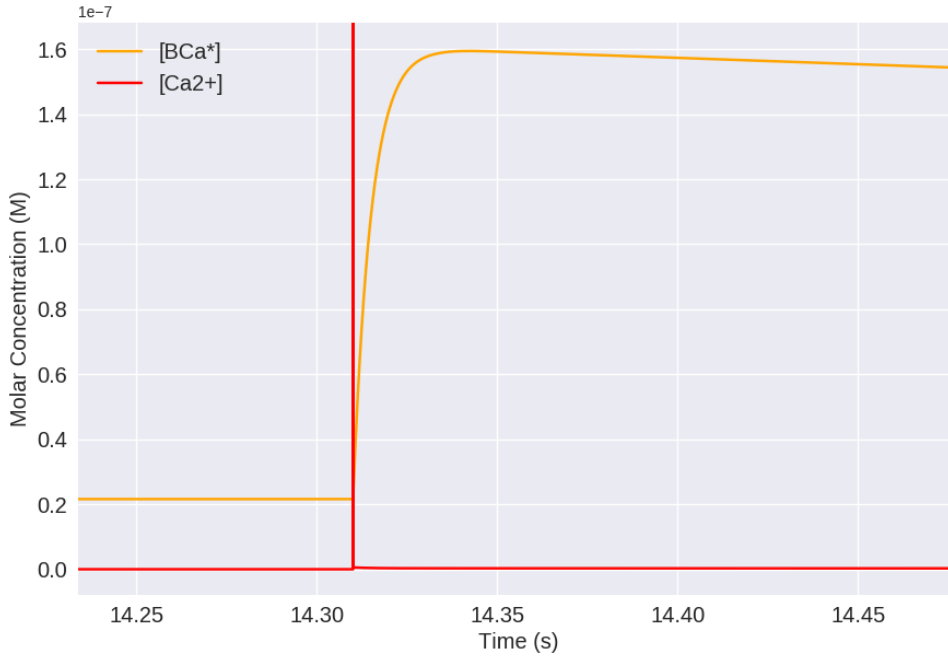


Figure 7: A zoomed in version of figure 5 showing only the  $[Ca^{2+}]$ , and  $[BCa^*]$  concentrations during the period 14.30 – 14.33s. The  $[Ca^{2+}]$  is only included to show the presence of an action potential. There is a logarithmic increase in the  $[BCa^*]$  concentration immediately after the action potential, followed by a much slower exponential decay.

## 3.2 Spike Inference

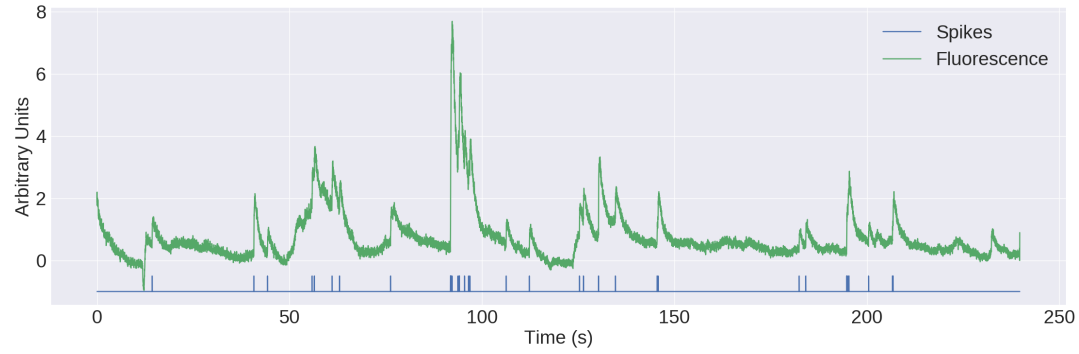
### 3.2.1 Observed fluorescence vs Modelled fluorescence

We used the three spike inference algorithms mentioned in section 2.2 to infer spike trains from each of the observed fluorescence traces, and each of the modelled fluorescence traces. We compared the performance of the inference algorithms on the observed fluorescence traces to their performance on the modelled fluorescence traces by measuring the true positive rates and true negative rates of inferred spike trains. The results of these measurements can be seen figures 9 and 10.

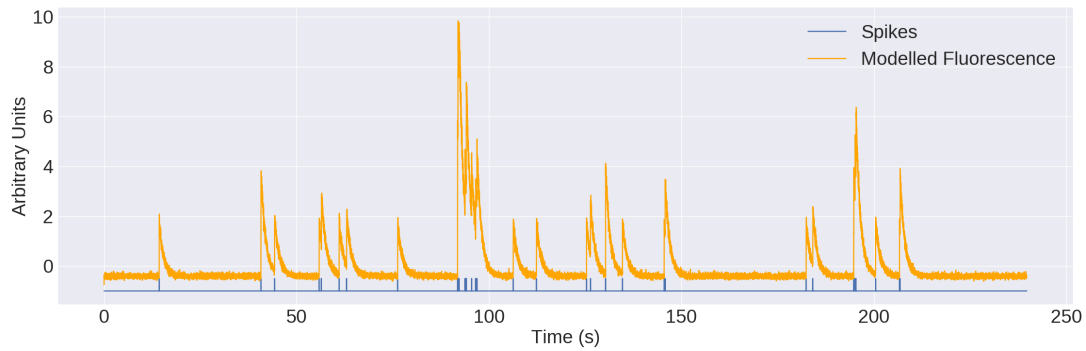
Two out of three of the algorithms performed better on the modelled spike trains. The Paninski algorithm didn't perform significantly better or worse on modelled traces in comparison to the observed traces.

For the algorithms that performed better on the modelled traces, the spike trains that the algorithms inferred from the modelled fluorescence traces were more similar to the observed spike trains than the spike trains inferred from the observed fluorescence traces. Since the observed fluorescence traces contain low frequency fluctuations that cannot be captured by the model, the spikes should be easier to infer from the modelled fluorescence traces. So the improved performance is expected.

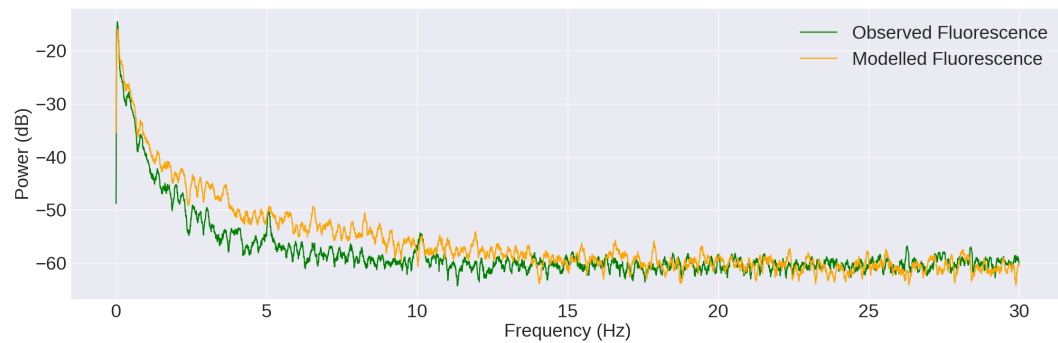




(a) An experimentally observed fluorescence trace and the spike train that induced the trace. Note that is the same example trace as we used in figures 5, 6, and 7.



(b) The modelled fluorescence trace created from the same spike train as shown in (a). Note that the experimental fluorescence trace contains some low frequency oscillations that are not captured by the model.



(c) The smoothed log power spectra for both the observed fluorescence trace and the modelled fluorescence trace shown in (a) and (b) respectively. The aim of our optimisation process was to make these two spectra as alike as possible.

Figure 8

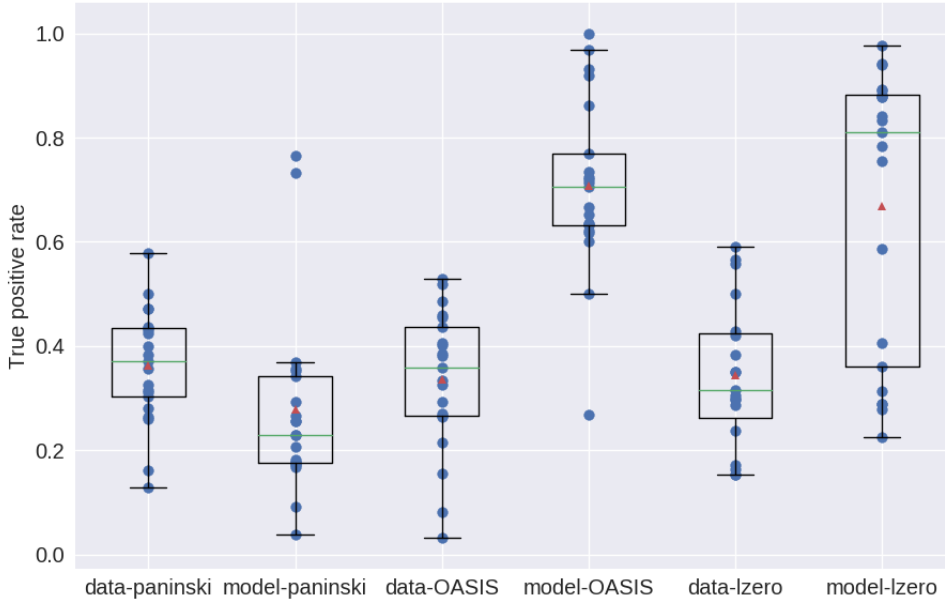


Figure 9: Box plot showing the distribution of the true positive rate measurements for spike inference on each of the observed and modelled fluorescence traces. Each of the box plots are labelled with either ‘data’ for the observed fluorescence traces, or ‘model’ for the modelled fluorescence traces. One box plot for observed fluorescences, and one box plot for modelled fluorescences for each of the three algorithms mentioned in section 2.2 are shown. So six plots in total. The box extends from the lower to the upper quartile. The median is shown as a green line, and the mean is shown as a red triangle. The whiskers show the range of the data. The individual true positive rates are represented by the blue dots. The OASIS and  $\ell_0$  algorithms performed better on the modelled traces than the observed traces ( $p < 10^{-8}$  and  $p < 10^{-4}$  respectively). The Paninski algorithm did not perform significantly differently on the two datasets ( $p > 0.05$ ).

### 3.2.2 Perturbation analysis

In order to measure the sensitivity of the spike inference algorithms to changes in the concentration of fluorescent calcium indicator within the cell, the spike inference analysis outlined in section 2.2 was carried out, using different values for the concentration of fluorescent indicator. The experimental, and therefore unperturbed, value for the indicator concentration was  $10^{-4}\text{M}$ . The perturbed values used were different from this value by orders of magnitude in order ensure that any effects of the perturbation would be observed. These values were  $10^{-2}\text{M}$ ,  $10^{-3}\text{M}$ ,  $10^{-5}\text{M}$ , and  $10^{-6}\text{M}$ .

The dynamics of the five free calcium/calcium molecule concentrations simulated by the model can be seen in figure 11. There is one figure for each perturbed value as well as one for the unperturbed value. There is not much difference between the models using the three highest values for the indicator concentration. In each of these cases, all of the free calcium that enters the cell upon the firing of an action

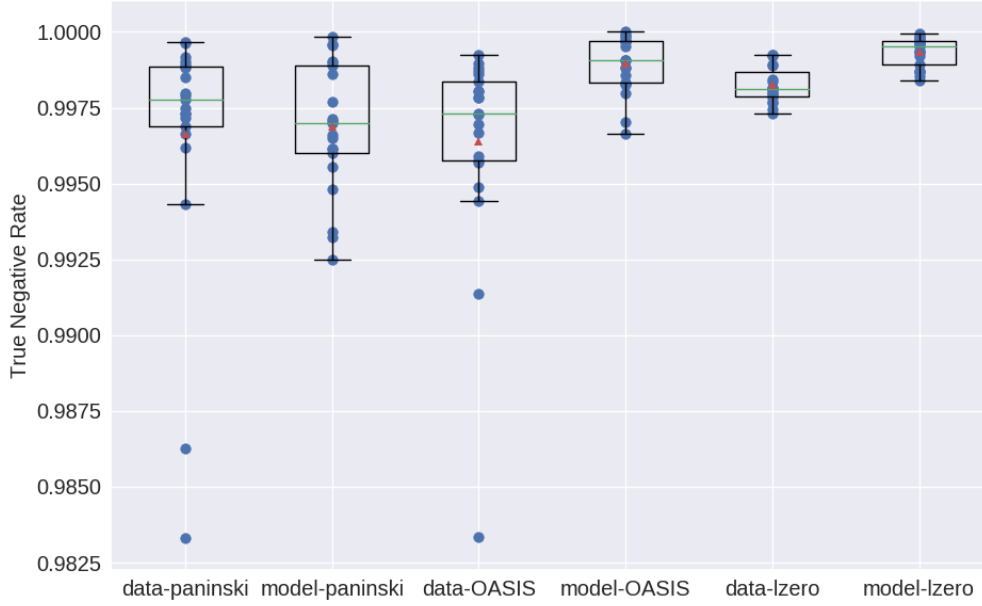


Figure 10: This figure is the same as figure 9 with true positive rates replaced with true negative rates. Once again, the OASIS and  $\ell_0$  algorithms performed better on the modelled traces than the observed traces ( $p < 0.005$  and  $p < 10^{-6}$  respectively). The Paninski algorithm did not perform significantly differently on the two datasets ( $p > 0.05$ )

potential is buffered by a combination of the fluorescent indicator and the mobile and immobile endogenous buffers. For the models using the lowest two values of the indicator concentration, the relative difference between the indicator bound calcium and immobile buffer bound calcium traces is either much smaller, or the sign of the difference is reversed. This is a result of the indicator becoming saturated by the free calcium, and the immobile buffer binding to the remaining free calcium. A consequence of this is a smaller change in the amplitude of the fluoresce in response to an action potential.

## 4 Discussion

We wrote a program to model the fluorescence traces read from neuron somata containing a fluorescent calcium indicator during spiking activity. The model simulated five molecular concentrations within the soma. The photon collection process was modelled as the noisy part of the system, with photons being captured stochastically according to a binomial distribution. We used the trained model to create synthetic fluorescence traces using the spike trains in the *Spike Finder* dataset. We ran three spike inference algorithms on the modelled fluorescence traces as well as on the observed ground truth fluorescence traces and compared both the spike trains inferred from the modelled traces and from the observed traces with the ground truth spike trains.

We found that the observed fluorescence traces contained low frequency fluctua-

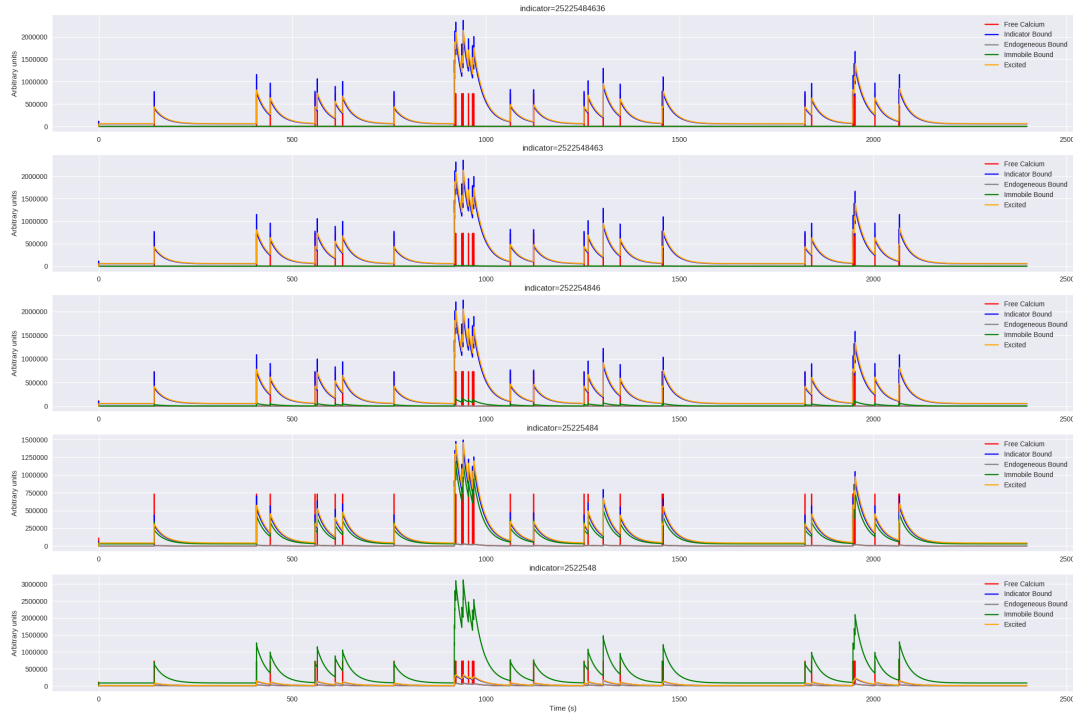


Figure 11: The dynamics of all the concentrations modelled by the model for five different concentrations of fluorescent indicator. From top to bottom, the concentration of the fluorescent indicator is  $10^{-2}\text{M}$ ,  $10^{-3}\text{M}$ ,  $10^{-5}\text{M}$ , and  $10^{-6}\text{M}$ . The top three values all look very similar. Having more fluorescent indicator doesn't make any difference when all of the free calcium is buffered by the indicator. For the two lowest values of fluorescent indicator, we can see the concentration of excited buffered calcium is much smaller, and therefore the fluorescence will be reduced. We can also see the other calcium buffers bonding to more free calcium as the fluorescent indicator becomes saturated.

tions that were not captured by the model consisting of first order ordinary differential equations.

In comparing the quality of spike inference from observed fluorescence traces to the quality of spike inference from modelled fluorescence traces, we found that two of the three algorithms used performed better on the modelled fluorescence traces, with no significant difference for the other algorithm used. We expected an improvement in performance due to the lack of low frequency oscillations in the modelled fluorescence traces.

## References

- [1] György Buzsáki, *Large-scale recording of neuronal ensembles*. Nature Neuroscience 7, 446-451, (2004)
- [2] Christoph Stosiek, Olga Garaschuk, Knut Holthoff, Arthur Konnerth, *In vivo two-photon calcium imaging of neuronal networks*. Proceedings of the National Academy of Sciences 100, 7319-7324, (2003)
- [3] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, Vivek Jayaraman, Loren L. Looger, Karel Svoboda, Douglas S. Kim, *Ultrasensitive fluorescent proteins for imaging neuronal activity*. Nature 499, 295-300, (2013)
- [4] Joshua T. Vogelstein, Adam M. Packer, Timothy A. Machado, Tanya Sippy, Bakdash Babadi, Rafael Yuste, Liam Paninski, *Fast non-negative deconvolution for spike train inference from population calcium imaging*. Journal of Neurophysiology 104, 3691-3704, (2010)
- [5] E.A. Pnevmatikakis, D. Soudry, Y. Gao, T.A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T.M. Jessell, D.S. Peterka, R. Yuste, L. Paninski *Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data*. Neuron 89, 285-299, (2016)
- [6] Sean Jewell, Daniela Witten, *Exact Spike Train Inference Via  $\ell_0$  Optimization*. arXiv:1703.08644v2, (2017)
- [7] Johannes Friedrich, Liam Paninski, *Fast active set methods for online spike inference from calcium imaging*. 30th Conference on Neural Information Processing Systems, (2017)
- [8] M. Maravall, Z. F. Mainen, B. L. Sabatini, K. Svoboda, *Estimating intracellular calcium concentrations and buffering without wavelength ratioing*. Biophysical Journal 78, 2655-2667, (2000)
- [9] Christoff Koch, *Biophysics of Computation: Information processing in single neurons*. Oxford University Press, (1999)
- [10] Marie-Claude Kilhoffer, Daniel M. Roberts, Abiodun O. Adibi, D. Martin Watersons, Jacques Haiech, *Investigation of the Mechanism of Calcium Binding to Calmodulin*. Journal of Biological Chemistry 32, 17023-17029, (1988)
- [11] Jeff Bezanson, Stefan Karpinski, Viral B. Shah, Alan Edelman, *Julia: A fast dynamic language for the technical computing*. arXiv:1209.5145, (2012)