

SCIENTIFIC REPORTS

OPEN

FISSA: A neuropil decontamination toolbox for calcium imaging signals

Sander W. Keemink^{1,2}, Scott C. Lowe^{1,3}, Janelle M. P. Pakan ³, Evelyn Dylida ³, Mark C. W. van Rossum ¹ & Nathalie L. Rochefort ³

Received: 11 October 2017

Accepted: 6 February 2018

Published online: 22 February 2018

In vivo calcium imaging has become a method of choice to image neuronal population activity throughout the nervous system. These experiments generate large sequences of images. Their analysis is computationally intensive and typically involves motion correction, image segmentation into regions of interest (ROIs), and extraction of fluorescence traces from each ROI. Out of focus fluorescence from surrounding neuropil and other cells can strongly contaminate the signal assigned to a given ROI. In this study, we introduce the FISSA toolbox (Fast Image Signal Separation Analysis) for neuropil decontamination. Given pre-defined ROIs, the FISSA toolbox automatically extracts the surrounding local neuropil and performs blind-source separation with non-negative matrix factorization. Using both simulated and *in vivo* data, we show that this toolbox performs similarly or better than existing published methods. FISSA requires only little RAM, and allows for fast processing of large datasets even on a standard laptop. The FISSA toolbox is available in Python, with an option for MATLAB format outputs, and can easily be integrated into existing workflows. It is available from Github and the standard Python repositories.

Recent developments in *in vivo* imaging and genetically-encoded calcium indicators enable monitoring the activity of hundreds to thousands of neurons in the brains of awake behaving rodents. The activity of sub-types of neurons can be directly related to the animal's behaviour, over temporal scales from hundreds of milliseconds to several weeks^{1–5}. Such imaging experiments produce large sequences of images, the analysis of which typically involves the following steps (Fig. 1A):

- (1) Correction of brain motion artefacts that lead to the misalignment of imaging frames from one time point to the next. For this step, open source software packages are available^{6–13}.
- (2) Segmentation of the imaged field-of-view into regions of interest (ROI), typically containing individual neuronal soma or sub-cellular components (e.g. dendrites and spines). Images can be segmented either manually⁴, semi-automatically^{3,14}, or automatically using either morphological criteria⁸ or activity based algorithms^{8,9,15–18}.
- (3) Extraction of the fluorescence changes across time within each ROI. Because two-photon microscopes have an elongated point spread function along the Z-axis, the signals imaged in a given focal plane are contaminated by signals from above and below this plane. The fluorescence signal from a given region of interest is thus usually contaminated by signals from surrounding neurites (axons and dendrites) and sometimes nearby neuronal somata. Correcting for such out-of-focus contamination is particularly critical for experiments in which neuropil activity itself is modulated by the experimental protocol. Decontamination is the focus of the current paper.

Currently, two main approaches are used to correct for neuropil contamination, either by subtracting a neuropil signal from the somatic signal, or by using blind source separation methods. For subtraction, a neuropil region is defined around the region of interest (e.g. a soma), and the spatially averaged neuropil signal is subtracted from the somatic signal^{3,14}. Since the neuropil signal is not spatially uniform^{11,19}, using a local neuropil region is preferable over using a global neuropil signal. The subtraction method has the advantage of being fast and intuitive.

¹Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, UK. ²Bernstein Center Freiburg, Faculty of Biology, University of Freiburg, 79104, Freiburg, Germany. ³Centre for Discovery Brain Sciences, Biomedical Sciences, University of Edinburgh, Edinburgh, EH8 9XD, UK. Sander W. Keemink and Scott C. Lowe contributed equally to this work. Mark C. W. van Rossum and Nathalie L. Rochefort jointly supervised this work. Correspondence and requests for materials should be addressed to S.W.K. (email: swkeemink@scimail.eu) or N.L.R. (email: n.rochefort@ed.ac.uk)

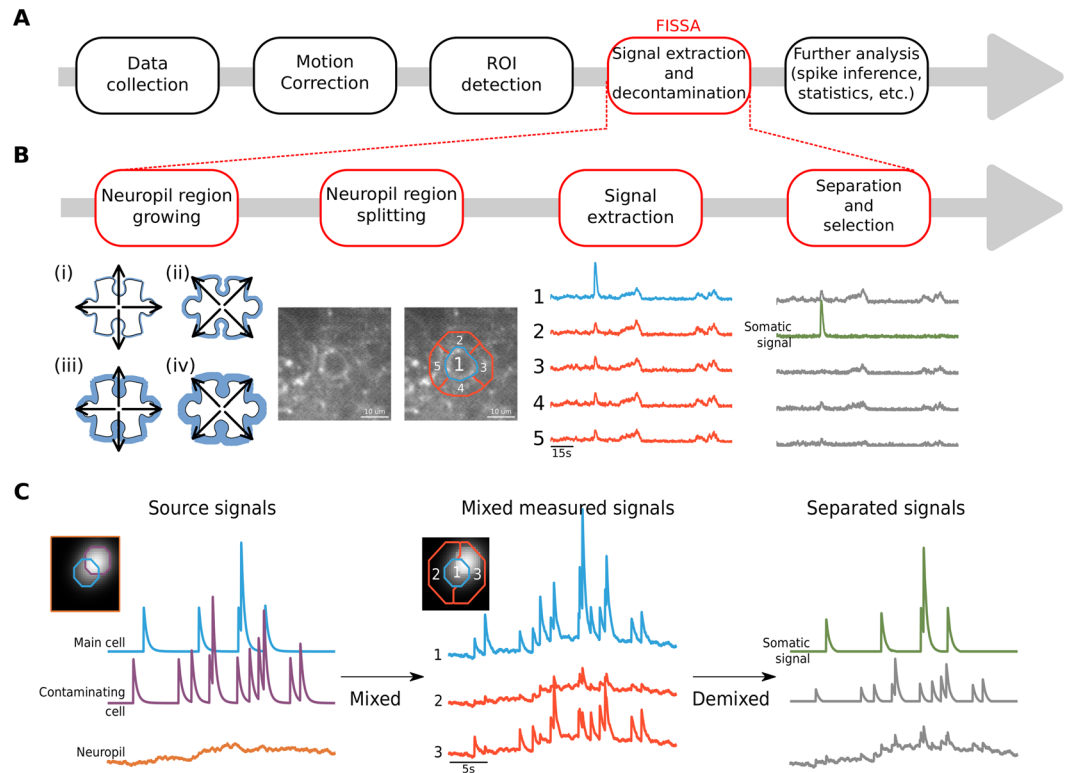


Figure 1. FISSA toolbox overview. **(A)** Schematic of a generic calcium imaging data analysis workflow. **(B)** Schematic of the four main steps of the FISSA toolbox workflow. Given a predefined region of interest (ROI), the local neuropil region is defined by expanding the ROI (white shape) alternately in the cardinal and diagonal directions (steps i to iv), until the surrounding area (blue area) is four times the ROI area. The resulting neuropil region is split into four subregions (regions 2, 3, 4, and 5). The signals from each of these regions are separated using non-negative matrix factorization (NMF), and the somatic signal is identified. **(C)** Schematic representation of blind source separation in FISSA. Left: Model of a somatic signal (blue) contaminated by two types of sources: the surrounding neuropil (orange), and an overlapping soma (purple). Middle: The measured signals in the somatic ROI (region 1) and in the surrounding neuropil (regions 2 and 3) will each be a mix of the three underlying source signals shown in the left panel. Right: NMF separates the mixed signals, recovering the original source signals. From these demixed components the one that best matches the measured ROI signal, relative to the neuropil regions, is identified as the somatic signal.

However, subtraction can lead to negative signals when the neuropil signal is larger than the somatic signal, while in other cases it does not remove all contamination. As a consequence, the subtraction parameters (such as how much weight to give to the subtracted neuropil signal) have to be adjusted for each dataset or even for each cell type (sparse vs densely firing cells¹⁴), which makes standardization of this approach challenging.

A more recent class of decontamination methods is based on blind source separation. These methods aim at finding underlying signal sources from image sequences, typically using either Independent Component Analysis (ICA)^{18,20}, Non-Negative Matrix Factorization (NMF)¹⁵, or model-based NMF^{9,11}. These approaches are standardized and user-independent. They simultaneously estimate ROIs and their associated fluorescence signals, while also accounting for neuropil contamination. However, automated segmentation into somatic ROIs is not always reliable and hand labelling is often still necessary. In addition, in early versions of blind source separation methods, the neuropil was modelled as a one-dimensional signal shared by all pixels with different weights, which can lead to an artificial decrease of correlated somatic signals⁹. Finally, cell-detection methods are computationally intensive for large datasets. Thus, the ideal neuropil decontamination method would be standardized, fast, and work with both manually and automatically drawn ROIs.

For these reasons, we have developed the Fast Image Signal Separation Analysis (FISSA) toolbox for decontaminating calcium signals. FISSA defines a set of neuropil regions around pre-defined somatic ROIs (either from hand-labelling or from automatic detection algorithms), and uses NMF to separate the signals from these regions (Fig. 1B,C). We have tested this toolbox on both simulated and *in vivo* data, and our results show that FISSA performs either similarly or better than existing published methods. Additionally, since only a few signals need to be separated, FISSA is fast and requires only little RAM, so that it is usable on a standard computer or laptop.

Results

FISSA toolbox workflow. The goal of the FISSA toolbox is to remove contamination from the ROI signals. As a consequence of the limited resolution of *in vivo* imaging methods, especially axially²¹, the signal measured from a given region of interest in a single focal plane is in fact a mixture of signals originating from this ROI as

well as from a surrounding volume (Fig. 1C). This volume includes mostly neurites (axons and dendrites) of other cells, and sometimes other somata. To demix these signals, we use the fact that the signals (photon counts) are always positive, and assume that the mixing of the different signals is linear and additive. A method of choice for demixing signals under these assumptions is non-negative matrix factorization^{22,23}, which separates signals by estimating a set of positive signals that best explains the observed mixed signals.

The FISSA toolbox relies on user-defined ROIs that can be imported either from ImageJ or defined as standard Python arrays. For each ROI, FISSA first defines a neuropil region by expanding the shape of each ROI by a fixed amount (Fig. 1B and Methods). The neuropil area is defined as the expanded shape, excluding the original ROI. Next, the neuropil region is divided into subregions of equal area. By default, FISSA defines four subregions each with the same area as the ROI; performance did not improve with more subregions (Fig. 2E).

The signals from each region (the four subregions of the neuropil and the somatic ROI) are then separated using NMF. The NMF method returns how strongly each separated signal was present in each subregion and ROI. Using these estimates, the separated signals are scaled and sorted by relative presence in the ROI compared to the surrounding subregions (see Methods). The signal with the strongest relative presence is taken as the extracted somatic signal (Fig. 1C).

Note that FISSA separates the raw fluorescence signals. The calculation of the relative change in fluorescence ($\Delta f/f_0$) can be done afterwards (see Methods). The extracted and decontaminated signals can be accessed in Python or saved in MATLAB format.

FISSA performance on simulated calcium imaging data. We first used simulated data to evaluate our approach. We modelled a set of neurons with Poisson firing statistics and calcium indicator dynamics based on GCaMP6f (see Methods). In the model, each cell had a well defined spatial structure, but its signal bled into the surrounding region and cells could overlap. A smoothly fluctuating global neuropil signal was included to model background fluctuations. A given pixel might thus contain the global neuropil contamination, and one or more cell signals. Finally, to model photon emission, the calcium indicator signal at each pixel was simulated by Poisson shot noise.

We evaluated the performance of different decontamination methods on three cases with increasing contamination (Fig. 2). We compared the performance of three decontamination methods: 1) subtraction of the local surrounding neuropil signal^{3,14}, 2) a cell detection and signal separation method including neuropil decontamination called constrained NMF (cNMF⁹), and 3) FISSA extraction. To quantify the performance of each method we calculated the Pearson correlation between the extracted signals and the ROI source signal, with the extracted signals low-pass filtered at 5 Hz to minimize differences due to high frequency noise.

We first considered a single cell whose somatic signal was only contaminated by neuropil fluctuations (Fig. 2A). All three methods successfully decontaminated the ROI signal, resulting in a high correlation between the corrected signal and the ROI source signal (Fig. 2Aiii).

Next, we modelled the same soma but added a partially overlapping neuron (Fig. 2B). The additional contaminating calcium transients reduced the correlation between the measured and the ROI source signal (Fig. 2Biii). Whilst subtracting the neuropil signal still removed slow fluctuations, it did not fully remove the extra transients, resulting in a lower correlation than the other two methods. cNMF and FISSA removed both the background fluctuations and the contaminating transients equally well, while preserving the true transients.

Finally, we added a second non-overlapping signal source with a strong, localized calcium response (Fig. 2C), leading to a neuropil signal with additional large calcium transients (Fig. 2Cii). The subtraction method led to negative transients, and the correlation between its signal and the source signal was lower than for the other two methods. FISSA and cNMF both resulted in very high correlations, with FISSA's being slightly but significantly higher ($p = 0.0051$, Fig. 2Ciii).

We then tested whether the results were consistent across a broader range of simulation parameters (Fig. 3A). While keeping the firing rates of the other cells the same, we varied the firing rates of the cell of interest (panel Ai), its spike transient amplitude (panel Aii), and the imaging frame-rate (panel Aiii). The correlations generally decreased for all methods as the signal-to-noise ratio decreases (through lower firing rates or calcium transient amplitudes). FISSA maintained the highest correlation across all parameter changes, and in particular at low signal-to-noise ratios performed better than other methods. In some cases, the difference between cNMF and FISSA results does not reflect the performance of the signal separation per se but rather shows the limit of the automatic detection method of cNMF. In cases of low signal amplitude or firing rate, cNMF may simply not detect the cell of interest. As a consequence, since there is no segmentation and thus no signal, the correlation with the source signal is very low. However, for the same reasons, a cell with very low firing rates might also not be detected manually. We then tested whether FISSA performed equally well when taking only a subset of the data. Our results show similar performance when downsampling a 120 s data set from 100 Hz to lower frame rates (Fig. 3Aiii). Finally, the shape of the cell of interest, such as more elongated shapes, also did not substantially affect the performance (Fig. 3Aiv).

FISSA has a number of user-adjustable parameters: the number of neuropil regions, the area of the neuropil subregions, and the NMF parameter α which promotes sparseness of the source separation. Performance is robust across parameter values (Fig. 3B), as long as the number of neuropil regions is larger than 3 (the default is 4), for a neuropil subregion area at least half of the original ROI's area (default is the same size as the ROI), and an α between 0.1 and 0.5 (the default is 0.1). Finally, we tested the influence of the ROI's size relative to the cell of interest by varying the threshold used to define the ROI, while keeping the simulated cell shape constant (Fig. 3C). The results show that FISSA performance remains stable for a broad range of ROI sizes, either larger or smaller than the actual cell's shape (Fig. 3C). This robustness of FISSA is a useful property when using hand-labelled ROIs that may be larger or smaller than the actual cell body.

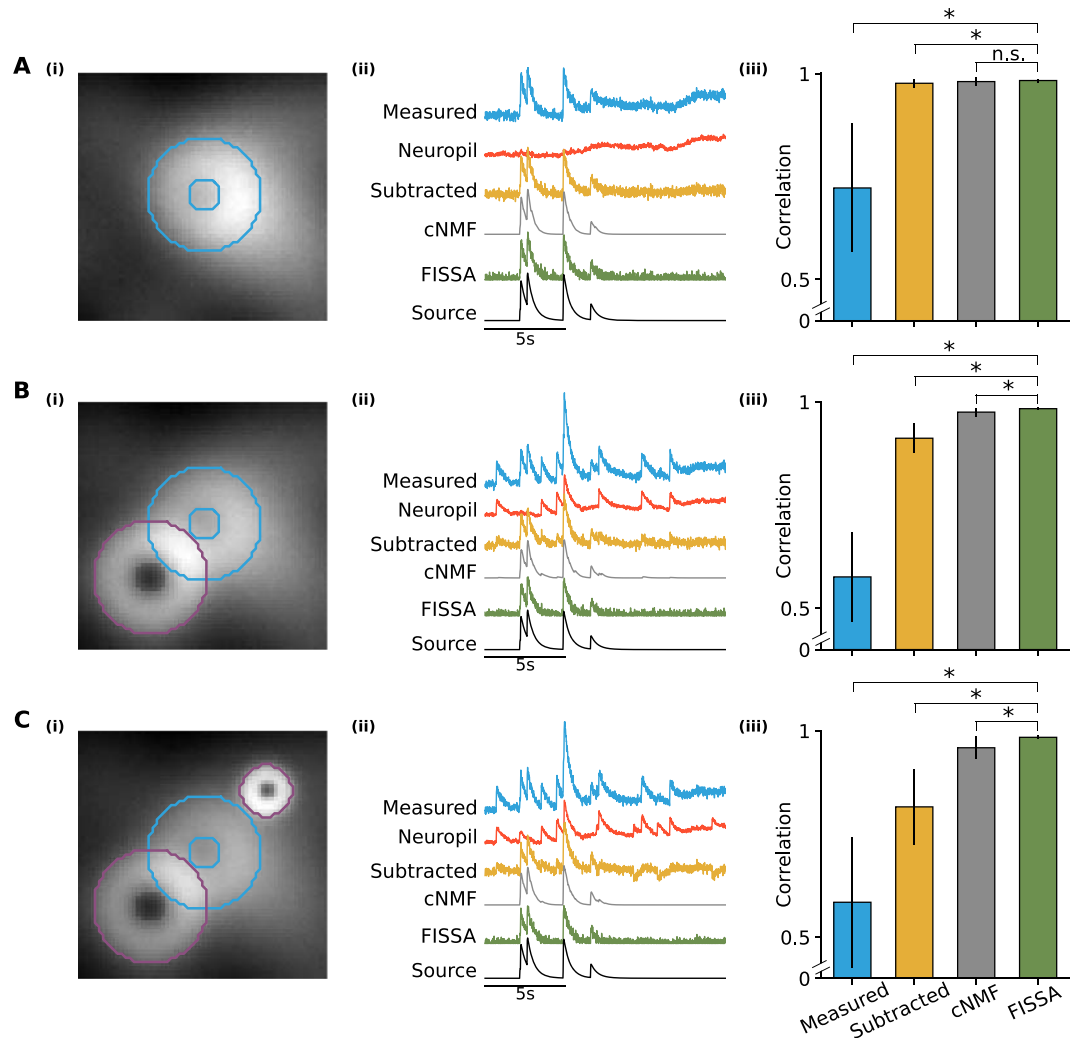


Figure 2. Comparison of FISSA performance with other decontamination methods on simulated calcium imaging data. Rows **A**, **B**, and **C** illustrate three cases with increasing levels of contamination. (**A**) (i) Image of a region containing a single doughnut-shaped cell firing at 0.5 Hz and a fluctuating neuropil background. The image is an average across all 12000 frames (120 s). The ROI is indicated in blue. (ii) Example fluorescence traces for the blue ROI indicated in panel **A**(i): average measured signal of the ROI (blue trace, ‘Measured’), the average surrounding neuropil signal (red trace), and the signals after neuropil decontamination by three different methods (subtraction in yellow, cNMF in grey, and FISSA in green). The uncontaminated source signal (as defined by the simulation) is shown in black. The cNMF trace has a lower noise level as the cNMF algorithm also includes smoothing. (iii) Average Pearson correlations between the ROI source signal and the extracted ROI signal: before decontamination (first column, ‘Measured’; 0.723), and after each decontamination method (neuropil subtraction, cNMF, and FISSA; 0.977, 0.981, and 0.984 respectively). Error bars indicate standard deviation. * $p < 0.05$; ** $p < 0.005$; n.s.: not significant ($p > 0.05$); Wilcoxon signed-rank test, $n = 10$ simulations of 120 s each (with different background neuropil signals, spike times, and photon noise). FISSA vs: measured $p = 0.0033$, subtraction $p = 0.0409$, cNMF $p = 0.0911$. (**B**) As panel **A**, but the blue ROI is additionally contaminated by an overlapping cell firing at 0.3 Hz (indicated by the purple outline). Average values for the four methods are 0.576, 0.912, 0.975, and 0.984 respectively. FISSA vs: measured $p = 0.0033$, subtraction $p = 0.0033$, cNMF $p = 0.0033$. (**C**) As panel **B**, but an additional bright localized signal firing at 0.3 Hz was added (smaller purple outline). Average values for the four methods are 0.585, 0.816, 0.959, and 0.984 respectively. FISSA vs: measured $p = 0.0051$, subtraction $p = 0.0051$, cNMF $p = 0.0051$.

FISSA performance on *in vivo* calcium imaging data. We next tested FISSA on *in vivo* two-photon calcium imaging data of GCaMP6-labelled layer 2/3 neurons in the mouse primary visual cortex. We used a publicly available dataset which contains simultaneous calcium imaging of GCaMP6-labelled neurons and simultaneous cell-attached electrophysiological recordings^{3,24}, allowing for a direct comparison between the extracted calcium signals and the recorded spikes. To quantify performance we calculated the Pearson correlation between the calcium signals estimated by each method, and the predicted calcium transients inferred from the recorded spikes.

FISSA successfully decontaminated the somatic signal of 20 cells tested: the results showed significantly improved correlation values after FISSA compared to raw data ($p = 0.0006$, Fig. 4B, ‘Measured’ vs ‘FISSA’). On

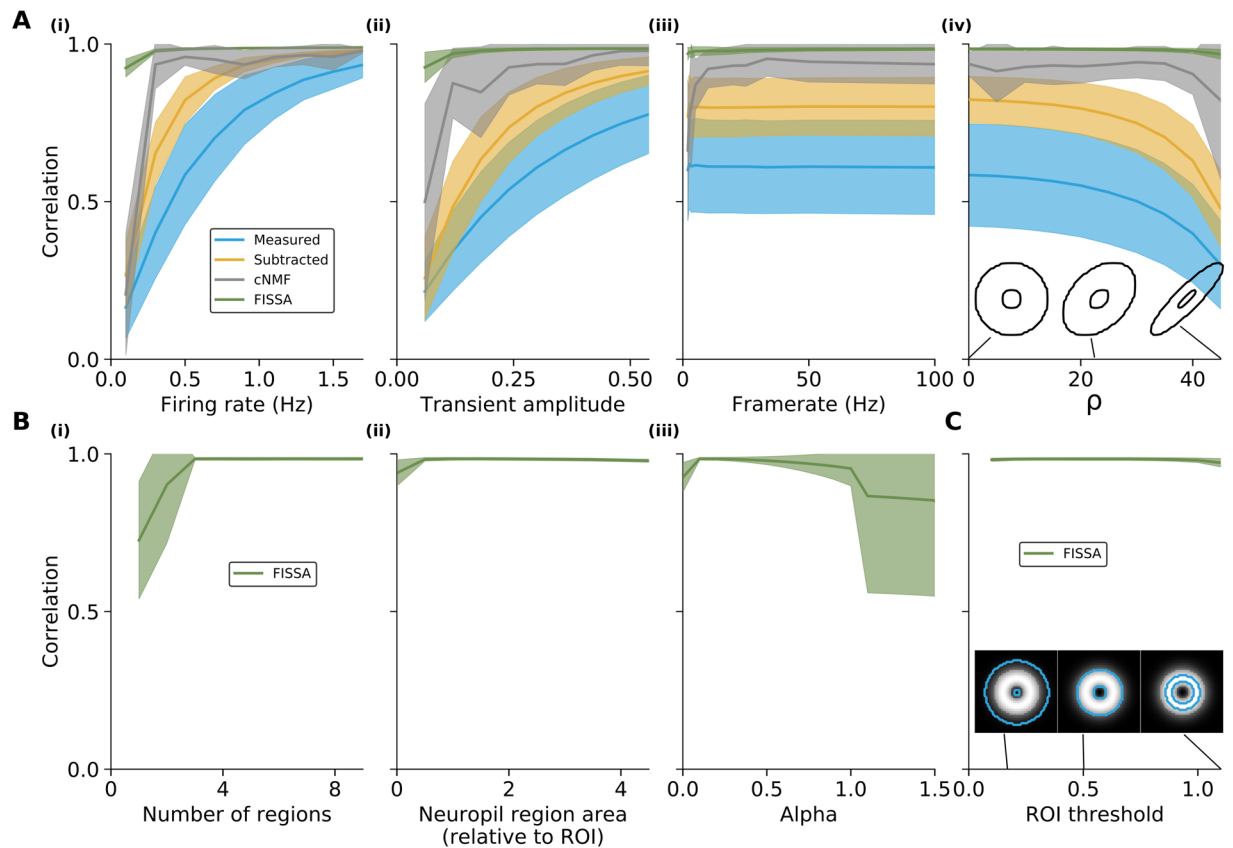


Figure 3. FISSA performance for different simulated data parameters and user-adjustable FISSA parameters, for the signal of the cell of interest from the case in Fig. 2C. **(A)** Correlations between the source and extracted signals before neuropil decontamination ('Measured') and after each decontamination method (neuropil subtraction, cNMF, and FISSA) for different simulation parameters. (i) Correlations for changes in the firing rate of the cell of interest (0.1 to 1.7 Hz, with steps of 0.2 Hz, default is 0.3 Hz). (ii) Correlations for changes in calcium transient magnitude (parameter A , see Eq. 9. 0.06 to 0.54 with steps of 0.06, default is 0.3). (iii) Correlations for changes in imaging framerate (downsampling of 100 Hz initial data to lower frame rates: by 50, 40, 30, 20, 10, 5, 4, 3, 2, and 1). (iv) Correlations for changes in cell shape, by changing the parameter ρ , see Eq. 10 (steps of 5 from 0 to 45, default is 0). The insets show example cell shapes for different ρ values. **(B)** Correlations between the source and extracted signals against different user-adjustable FISSA parameters (x-axes). (i) Number of neuropil subregions while keeping the total area constant, at four times the ROI area. (ii) Area of the neuropil subregions relative to the central ROI (0.025 for the smallest area, steps of 0.5 from 0.5 to 4), for four subregions. (iii) NMF parameter α (with steps of 0.1). **(C)** The effect of suboptimal ROI selection, by varying the threshold at which a cell mask is defined (Parameter T_{mask} in Eq. 14. 0.1 to 1.1 with steps of 0.1, default is 0.5). The insets show the same central cell with the outlines of the mask used for three example thresholds. All panels show the average values over 10 simulations; shaded areas indicate standard error.

this dataset, the results obtained with the cNMF and subtraction methods were not significantly different from those obtained with FISSA (FISSA vs subtraction $p = 0.2959$, FISSA vs cNMF $p = 0.4330$, Fig. 4B). This difference in results, compared to the simulated data in Fig. 2, is due to the relatively sparse labelling leading to a low level of contamination with few overlapping labelled structures in the field of view. As such, the main contamination source consists of background fluctuations, for which all three decontamination methods work well. However, cNMF did result in low correlation values for a small subset of cells. This can be partly explained by the high optical magnification used for the dataset (roughly one to three cells per $30 \mu\text{m}$ by $30 \mu\text{m}$ field of view, at 256 by 256 pixels), while cNMF was designed to be applied to a large field of view with hundreds to thousands of cells. For some cells, the ROIs that cNMF extracted did not fully match the outline of the soma and not all contamination was successfully removed (e.g. grey trace, Fig. 4A).

Finally, we compared the different methods for neuropil decontamination in a dataset of *in vivo* two-photon calcium imaging of layer 2/3 neurons in the primary visual cortex (V1) of awake behaving mice. All neurons were labelled through local injection of adeno-associated viruses (AAV1.Syn.GCaMP6f.WPRE.SV40) in V1⁴. After 2–3 weeks of expression, running speed of the animal and GCaMP6f signals were recorded simultaneously during the presentation of drifting gratings. It is known that visual responses of layer 2/3 neurons in V1 are modulated by locomotion^{4,25–29}. We compared the effect of locomotion on single neuron activity in this dataset before and after neuropil decontamination. The effect of locomotion was quantified for each neuron by the locomotion

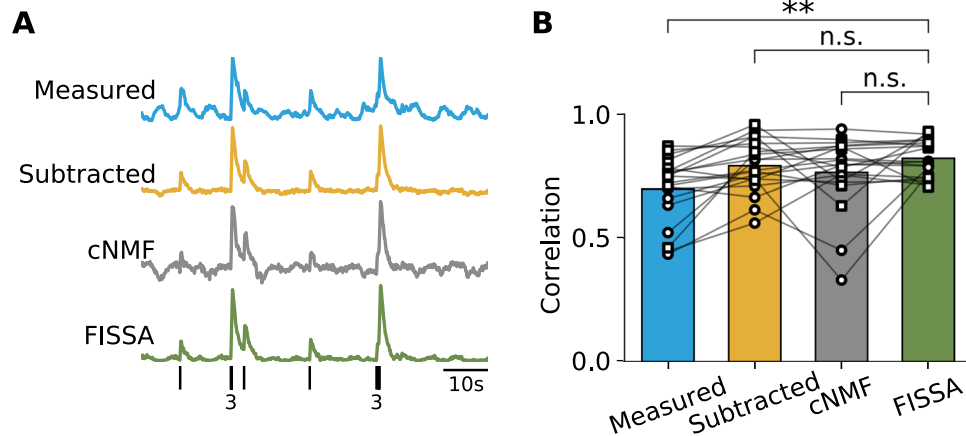


Figure 4. Comparison of neuropil decontamination methods on *in vivo* two-photon calcium imaging data from layer 2/3 neurons in mouse primary visual cortex (data from^{3,24}). **(A)** Example measured and decontaminated traces of fluorescence changes for cell 20120627_cell4 (frames 14400 to 19200). All traces were low-pass filtered at 5 Hz. The black ticks on the bottom row indicate the spikes measured electrophysiologically, with the number of overlapping spikes indicated below. **(B)** Correlations between the calcium transients inferred from the recorded spikes (see Methods), and the calcium signals estimated by each method. The data points correspond to GCaMP6s (circles) and GCaMP6f (squares) cells from the dataset^{3,24}. ** $p < 0.005$; n.s.: not significant ($p > 0.05$); Wilcoxon signed-rank test, $n = 20$ cells, FISSA vs measured: $p = 0.0006$, FISSA vs subtraction $p = 0.2959$, FISSA vs cNMF $p = 0.4330$.

modulation index (LMI)⁴, which calculates the normalized difference between the mean change in fluorescence ($\Delta f/f_0$) during locomotion (R_L) and stationary (R_S) periods, as $LMI = (R_L - R_S)/(R_L + R_S)$.

Our results show that before neuropil decontamination almost all cells in this dataset displayed positive LMI (Fig. 5, median LMI 0.29), indicating an increase of activity during locomotion. However, after neuropil decontamination by any of the three methods, LMI values strongly decreased (median LMI was 0.16, 0.04, and 0.09 using neuropil subtraction, cNMF, and FISSA, respectively). These results are in agreement with previous electrophysiological experiments which reported that 20–50% of neurons with visual responses are positively modulated by locomotion in mouse V1^{28,30}. The LMI values obtained with FISSA were not significantly different from those obtained with cNMF ($p = 0.0687$). However, in this dataset, the subtraction method led to significantly higher LMI values than those obtained after FISSA and cNMF ($p = 0.0117$).

Altogether, these results show that correction for neuropil contamination is critical for the analysis of two-photon calcium imaging data, especially for datasets in which the surrounding GCaMP6-labelled neuropil is itself modulated by the experimental conditions (e.g. by sensory stimuli or by the animal's behaviour). In addition, the results obtained with both simulated data and *in vivo* two-photon calcium imaging datasets indicate that FISSA performs either similarly or better than other published methods for neuropil decontamination.

FISSA computational resources and integration into existing workflows. FISSA is freely available at <https://github.com/rochefort-lab/fissa>. The toolbox can be applied to an existing dataset in just a few lines of code:

```

1 # FISSA toolbox import
2 import fissa
3
4 # Specify the location of the ROIs and data files
5 images_location = "../exampleData/20150629"
6 rois_location = "../exampleData/20150429.zip"
7
8 # Specify the folder where data will be stored
9 output_folder = "fissa_example"
10
11 # Signal extraction and decontamination
12 experiment = fissa.Experiment(images_location, rois_location, output_folder)
13 experiment.separate()
14
15 # Export results in MATLAB format
16 experiment.save_to_matlab()

```

First, the user defines:

- The imaging data, either as a directory containing tiff files of the acquired images or standard Numpy arrays (Python format that can be generated from other data formats).
- The regions of interest (zip files of ROIs defined in ImageJ or Numpy arrays).
- The results folder where extracted and processed data will be stored.

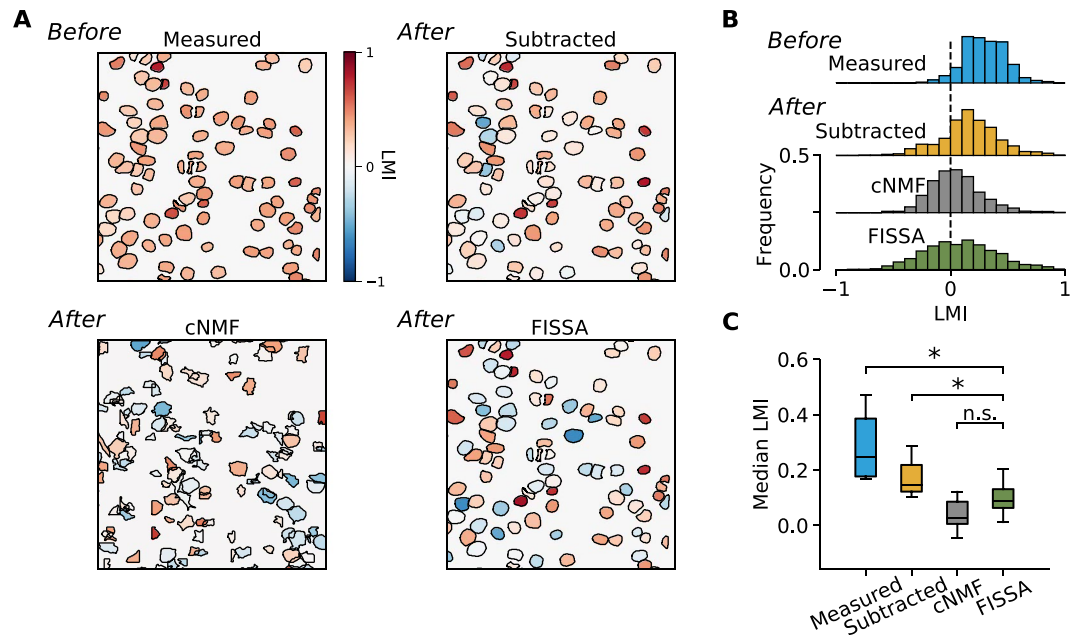


Figure 5. Impact of neuropil decontamination on neuronal responses measured with two-photon calcium imaging, in V1 layer 2/3 of awake behaving mice. Neurons were labelled with GCaMP6^{fl}. The effect of locomotion was quantified for each neuron by the locomotion modulation index (LMI). **(A)** An example field of view with somatic ROIs coloured by the LMI value obtained either before neuropil decontamination ('Measured') or after decontamination. The ROIs for the measured, subtraction, and FISSA methods are identical and were defined by hand. For the cNMF method the ROIs are presented as detected by the algorithm. LMI values were calculated for periods of visual stimulation with oriented gratings (10 to 20 trials per field of view, 60 s/trial). **(B)** The distributions of LMI values across all layer 2/3 cells in eight mice, before ('Measured') and after each decontamination method. **(C)** The distributions of median LMI values for the eight mice, before and after decontamination. * $p < 0.05$; n.s.: not significant ($p > 0.05$); Wilcoxon signed-rank test, $n = 8$ mice, FISSA vs measured: $p = 0.0117$, FISSA vs subtraction $p = 0.0117$, FISSA vs cNMF $p = 0.0687$, subtraction vs cNMF $p = 0.0687$.

After this step, only two lines of code are necessary to run the full FISSA analysis pipeline: from neuropil region definition to signal separation and selection. The results can then be accessed within Python or saved in MATLAB format. In the GitHub repository we provide example scripts which demonstrate how to integrate FISSA with existing published workflows such as SIMA⁸ and cNMF⁹.

By default, tiff files are fully loaded into memory before signal extraction. For large tiff files, there is an option to load them frame-by-frame to reduce memory usage. For other formats, there is also the option to define a custom data-loading script.

Since FISSA only has to separate a small number of signals, it can quickly separate signals across large datasets. For example, to extract isolated signals from 40 cells within a 600×600 pixel field of view over 2400 frames, FISSA only takes 40 seconds. This scales sub-linearly with the number of frames, such that for 30000 frames FISSA takes 60 seconds (for these tests FISSA was run on a workstation running Ubuntu 17.04 with a six core Intel Core i7-6800 K CPU@3.40 GHz). FISSA is thus well suited for processing large datasets with long imaging periods. However, as opposed to other methods (cNMF⁹, suite2P¹¹), FISSA does not automatically segment the images into defined ROIs; therefore, the total time for data processing also depends on the method used for defining ROIs as well as on the data loading time.

Discussion

We have developed a fast and easy to use toolbox (FISSA) for neuropil decontamination in calcium imaging datasets. The results obtained with both simulated and *in vivo* two-photon calcium imaging datasets indicate that FISSA performs either similarly or better than previously published methods for neuropil decontamination of calcium signals. In addition, FISSA presents a number of advantages. First, unlike subtraction methods, it provides a standardized, user-independent approach that removes multiple sources of contamination without leading to negative signal artefacts. Other methods based on blind-source separation do provide a standardized approach but require more computational resources and can be very slow to run on large datasets (although recent developments have improved the running time of this type of method^{11,31}).

A further advantage of FISSA is that it uses minimal computational resources and can thus be used on a standard laptop even for large datasets. However, the total analysis time will also depend on the ROI detection method. FISSA does not include an automatic cell detection algorithm: ROIs must be defined beforehand, either manually or through a third-party algorithm. Separating ROI detection and signal extraction can be an advantage for experimental data in which automatic cell detection methods are not sufficiently accurate.

Finally, FISSA's main assumptions are that measured calcium signals are positive and mix both linearly and additively. This is in contrast to other methods, that make specific assumptions about the calcium dynamics in terms of noise level and time scales^{9,11}. Thus, FISSA is more generally applicable across various experimental conditions which may include different fluorescent indicators (such as synthetic dyes or other types of protein-based sensors) as well as other imaging methods (both *in vitro* and *in vivo*). Because the FISSA toolbox can be adapted to different formats of imaging data and regions of interest, it can easily be integrated into existing data analysis workflows.

Methods

FISSA algorithm. *Neuropil subregions definition.* FISSA uses predefined ROIs, obtained from either manual or automatic segmentation. The surrounding region is automatically defined by expanding the shape of each ROI alternately in the cardinal and diagonal directions (Fig. 1B). Next, the neuropil region is divided into N equal area subregions, by taking the polar coordinates relative to the ROI centre and taking the $\frac{1}{N}$ th fraction for each subregion. By default, the total surrounding region is expanded until its area is N times the area of the central ROI (such that each subregion has the same area as the ROI). For the examples in this paper we set $N=4$, as performance saturated at this N (Fig. 3Bi).

Non-Negative Matrix Factorization implementation. We assume that the spatially averaged signal in a given ROI, $\mathbf{f}_{\text{measured}}(t)$, is given by a linear mixing of a set of underlying source signals

$$\mathbf{f}_{\text{measured}}(t) = W \mathbf{f}_{\text{source}}(t), \quad (1)$$

where $\mathbf{f}_{\text{source}}(t)$ is the set of source signals, and W is the mixing matrix. Blind source separation techniques estimate the mixing matrix V and separated sources $\mathbf{f}_{\text{sep}}(t)$ such that

$$\mathbf{f}_{\text{measured}}(t) \approx V \mathbf{f}_{\text{sep}}(t). \quad (2)$$

For FISSA, we define $\mathbf{f}_{\text{measured}}(t)$ as the $N+1$ signals from the central ROI and the neuropil subregions, V is the $N+1$ by $N+1$ mixing matrix, and $\mathbf{f}_{\text{sep}}(t)$ are the $N+1$ extracted signals. The separation thus assumes the number of output signals is the same as the number of input signals (measured $N+1$ signals). If there are fewer than $N+1$ source signals being mixed (as, for example, in Fig. 2), we did not find this to have a negative impact on performance. The main point is that the number of output signals should not be lower than the number of signal sources. However, in physiological data, the number of signal sources is unknown. For the data sets we have used (*in vivo* data from local labelling of cortical neurons in V1), the number of $N+1$ (corresponding to 5 regions) gave robust results. However, the option of changing the number of separated signals can be useful in case alternative data sets are likely to include more signal sources.

FISSA performs the blind source separation with Non-Negative Matrix Factorization (NMF), using the implementation in the scikit-learn toolbox³². There is also the option in the FISSA toolbox to use Independent Component Analysis (ICA) instead of NMF. ICA relies on the fact that the distribution of the sum of random variables will be more Gaussian than the individual components. Thus, by finding the most non-Gaussian projections, the sources can be identified. ICA is faster than NMF, but the resulting source signals or mixing coefficients can be negative, which can sometimes lead to negative extracted signals similar to the subtraction method (see Supplementary Fig. 1). NMF makes the stronger assumption that all signals and mixing coefficients are strictly non-negative, a property which is true for calcium imaging data. We therefore recommend using the NMF method, which was used for all results presented in this paper.

To perform NMF the temporal signals over time $\mathbf{f}_{\text{measured}}(t)$ and $\mathbf{f}_{\text{sep}}(t)$ are written as matrices with components

$$F_{i,t} = f_i(t). \quad (3)$$

The NMF algorithm then minimizes an objective E in alternating steps with respect to V and F_{sep} , until E reaches a target threshold^{22,33}. The objective is the total squared difference between the measured signals and the estimated signals F_{measured} , plus additional norms that encourage a sparse solution

$$\begin{aligned} E = & \frac{1}{2} \left\| F_{\text{measured}} - V F_{\text{sep}} \right\|_{\text{Fro}^2}^2 \\ & + \alpha l_{1\text{ratio}} \left\| V \right\|_1 + \alpha l_{1\text{ratio}} \left\| F_{\text{sep}} \right\|_1 \\ & + \alpha (1 - l_{1\text{ratio}}) \left\| V \right\|_{\text{Fro}^2} + \alpha (1 - l_{1\text{ratio}}) \left\| F_{\text{sep}} \right\|_{\text{Fro}^2} \end{aligned} \quad (4)$$

where the Frobenius norm is given by $\left\| A \right\|_{\text{Fro}^2} = \frac{1}{2} \sum_{i,j} A_{ij}^2$ and the element-wise L1 norm is given by $\left\| A \right\|_1 = \sum_{i,j} |A_{ij}|$. $l_{1\text{ratio}}$ determines the ratio between the Frobenius norm and the element-wise norm. α is the sparseness regularizer on both the mixing matrix V and the separated signals F_{sep} . We set $\alpha=0.1$ and $l_{1\text{ratio}}=0.5$. Finally, both the separated signals and the mixing matrix are constrained to be non-negative. Initialization of both the estimated mixing and separated matrix is done by non-negative double singular value decomposition³⁴.

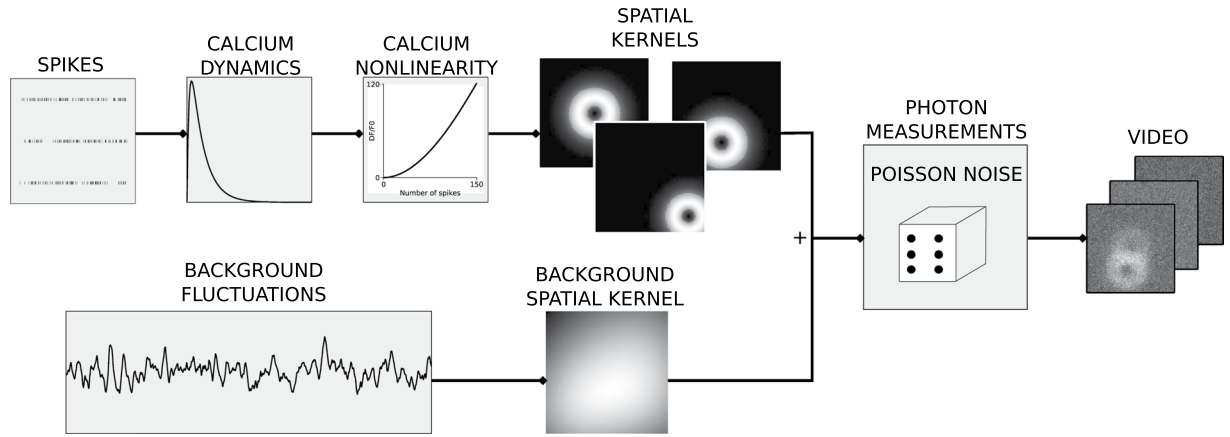


Figure 6. Simulated data generation. For each simulated neuron, a Poisson spike train is generated. The corresponding calcium indicator dynamics are simulated using GCaMP6 rise and decay times, and a nonlinearity. Next, the calcium traces are associated with a spatial kernel consisting of a doughnut shape mask (that models the cellular soma's structure) and a Gaussian that simulates signal spread. Additionally, noisy background fluctuations are generated which vary spatially through a spatial kernel. All resulting signals are summed, and passed through a Poisson generation process to simulate photon emission, resulting in the final image sequence.

Signal selection. Blind source separation returns a set of separated signals, but it does not indicate which one corresponds to the somatic signal and which ones are the contaminating signals. The estimated mixing matrix V provides the weight with which each identified source signal contributes to the mixed signals. Each row in the mixing matrix represents how strongly each of the estimated underlying signals is present in the measured signal. For each signal, we rate its relative presence in the central ROI by normalising the weights across each column of the mixing matrix

$$v'_{ij} = \frac{v_{ij}}{\sum_{i'} v_{i'j}}. \quad (5)$$

The values v'_{ij} represent how strongly each signal is present in each region, relative to its average across all regions. The estimated somatic signal is then given by the signal for which v'_{0j} is the highest, multiplied by its contribution to the measured ROI signal

$$f_{\text{est}}(t) = v_{0j_{\text{max}}} f_{\text{sep}}^{j_{\text{max}}}(t), \quad (6)$$

where $f_{\text{sep}}^j(t)$ is the j -th signal as separated by blind source separation, and

$$j_{\text{max}} = \arg \max_j v'_{0j}. \quad (7)$$

Thus, in order to find the signal that corresponds to the somatic signal, we assume that the somatic signal is more strongly represented in the central ROI, compared to the neuropil subregions (otherwise it is unlikely that it would have been chosen as the ROI). For example, the amplitude of the neuropil signal in the somatic ROI might be higher than the somatic signal itself, but this high neuropil signal will also be strongly present in the surrounding subregions. The somatic signal however, will be the one that is mostly present in the somatic ROI, but not in the other subregions.

Baseline. For Fig. 5, to calculate $\Delta f/f_0$, f_0 was estimated as the 5th percentile of the 1Hz low-pass filtered trace. For all extracted traces the f_0 of the non-corrected trace was used.

Multiple trials. In the case of several discrete recording sessions of the same cells, by default FISSA concatenates the trials together. This is done to ensure that the signal extracted for a given trial does not suddenly change across trials (e.g. if a cell is silent during one trial).

Software. The code is implemented in Python 2.7, using the Numpy 1.7, SciPy 0.12, Matplotlib 1.2 and HoloViews 1.6 toolboxes. We implemented NMF and ICA with the Python scikit-learn NMF and fastica functions respectively³².

In addition to the separation algorithm, the FISSA package has two utilities, which may be used independently. First, FISSA has fast TIFF reading scripts, using the open source tiff file package. Second, FISSA has a baseline estimator which estimates f_0 as the 5th percentile of the signal after applying a 1 Hz low-pass filter.

Simulated data. The simulated data generation is illustrated in Fig. 6. Each neuron spike train, $s(t)$, is generated by a Poisson process at a given rate. The calcium dynamics for a given cell are modelled as a difference of exponentials

$$\begin{aligned}\frac{dc_d}{dt}(t) &= -\frac{1}{\tau_d}c_d(t) + s(t) \\ \frac{dc_r}{dt}(t) &= -\frac{1}{\tau_r}c_r(t) + s(t) \\ c(t) &= c_d(t) - c_r(t)\end{aligned}\quad (8)$$

where $c_d(t)$ and $c_r(t)$ model the decay and rise dynamics respectively, $c(t)$ the overall calcium dynamics, and τ_r and τ_d the rise and decay time constants respectively. Using the published model of GCaMP6 dynamics^{35,36}, a polynomial non-linearity to model calcium indicator is applied to obtain the measured calcium indicator signal

$$\begin{aligned}d(t) &= \min [c_{\max}, c(t)] \\ f(t) &= A\{d(t) + p_2[d(t)^2 - d(t)] + p_3[d(t)^3 - d(t)]\}\end{aligned}\quad (9)$$

where A sets the signal change corresponding to a single spike, and p_2 and p_3 are polynomial parameters. The saturation is applied for $c > c_{\max}$, with $c_{\max} = \frac{-2p_2 - \sqrt{4p_2^2 + 12p_3(p_2 + p_3 - 1)}}{6p_3}$; otherwise the model starts reducing the output fluorescence beyond c_{\max} . Using the average values from³⁶ we set $p_2 = 0.85$, $p_3 = -0.006$, $\tau_r = 0.0156$ s, and $\tau_d = 0.76$ s for model GCaMP6f dynamics, and $p_2 = 0.81$, $p_3 = -0.056$, $\tau_r = 0.0702$ s, and $\tau_d = 1.87$ s for GCaMP6s dynamics. A was set to 0.3%, 2%, and 4% for the central cell, the overlapping cell (Fig. 2C), and the bright localised signal (Fig. 2C) respectively. For Fig. 2 the simulations ran for 120 seconds at 100 Hz. The firing rates were set at 0.5 Hz, and 0.3 Hz for the central cell and the neighbouring cells, respectively (unless otherwise noted). In order to model the effects of stimulus presentation, and to induce correlations between neurons, all firing rates were periodically doubled for a duration of 15 seconds. For predicting calcium traces from the recorded spikes for Fig. 4, the simulations ran at 60 Hz for as long as each original neuron was recorded electrophysiologically, with the electrophysiologically measured spike-times binned at 60 Hz.

This signal $f(t)$ is convolved with a two-dimensional doughnut-shaped spatial kernel given by the difference between two 2D Gaussians plus a step function

$$K_i(x, y) = \begin{cases} c + S(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - S(\boldsymbol{\mu}, \boldsymbol{\Sigma}/2), & \text{if } S(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - S(\boldsymbol{\mu}, \boldsymbol{\Sigma}/2) > T \\ S(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - S(\boldsymbol{\mu}, \boldsymbol{\Sigma}/2), & \text{otherwise} \end{cases}\quad (10)$$

where $S(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$, $\boldsymbol{\mu} = [\mu_x, \mu_y]$ indicates the x and y positions, and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{bmatrix}$ sets the spatial spread. The offset $c = 0.2$ models the physical structure of a cell soma, the threshold $T = 0.5$ determines its extent, and the Gaussian models the spread of a cell's calcium signal beyond its structure to model cross contamination between nearby structures. $K_i(x, y)$ was additionally normalized so that $\max(K) = 1$. For Fig. 2 the central cell had $\sigma^2 = 50$ and $\mu_x = \mu_y = 0$ (indicating the middle of the field of view), the overlapping cell had $\sigma^2 = 50$ and $\mu_x = \mu_y = 13$, and the small cell $\sigma^2 = 10$ and $\mu_x = \mu_y = -15$. Unless otherwise noted ρ is always set to 0.

We simulate the background neuropil contamination $B(t)$ as

$$\frac{dB(t)}{dt} = \eta dW(t),\quad (11)$$

where η is a constant (set to 0.05) and $W(t)$ is a Wiener process. A square wave signal of magnitude 0.1 and a period of 15 seconds was also added to simulate stimulus presentation. To vary the background signal spatially, the background signal is convolved with a spatial kernel given by the sum of $M = 10$ Gaussians

$$K_{\text{bg}}(x, y) = \sum_{i=1}^M \exp\left(-\frac{(x - \mu_{xi})^2 + (y - \mu_{yi})^2}{2\sigma_i^2}\right).\quad (12)$$

with $\sigma_i^2 \in [100, 200]$ and μ_{xi}, μ_{yi} give the mean x and y positions which were randomly drawn within the image limits (80 by 80 pixels). Finally, we obtained the calcium response at a given pixel at position x, y and time t as

$$F(x, y, t) = \sum_i K_i(x, y) f_{\text{true},i}(t) + K_{\text{bg}}(x, y) B(t),\quad (13)$$

where the sum over i is across structures. To simulate photon emission, F is used as the rate in a Poisson noise process to generate the final pixel responses. Signals for each ROI are estimated based on the mask

$$M_i(x, y) = \begin{cases} 1, & \text{if } K_i(x, y) > T_{\text{mask}} \\ 0, & \text{otherwise} \end{cases},\quad (14)$$

where T_{mask} determines its extent (set to 0.5 unless otherwise noted).

Other decontamination methods used to compare FISSA performance. *Neuropil subtraction.* Given a ROI's spatially averaged signal $f_{\text{ROI}}(t)$ and a surrounding neuropil region, the neuropil's spatially averaged signal $f_{\text{npil}}(t)$ was subtracted to give the estimated signal

$$f_{\text{est}}(t) = f_{\text{ROI}}(t) - kf_{\text{npil}}(t), \quad (15)$$

where k is a constant. The value of this constant has been determined manually in previous publications^{3,14}. When analysing our simulated data, we set $k = 1$, since this value removed the background fluctuations (Fig. 2A). For the experimental *in vivo* data, we set $k = 0.7$, as previously published³. For the neuropil region we used the total surrounding region (all subregions) as defined by the FISSA algorithm.

Constrained non-negative matrix factorization. For details of the cNMF algorithm see the original publication⁹. We applied cNMF to the simulated (Fig. 2) dataset using the demo_script.m script from the cNMF toolbox. For the *in vivo* data presented in Figs 4 and 5 the memory usage was much higher than for the simulated dataset due to higher resolution field of views and more frames. We thus performed the analysis in patches, rather than analysing the whole field of view. For this, we used the run_pipeline.m script from the cNMF toolbox. cNMF often detected more regions than just the soma of interest. In these cases we chose the region that resulted in the highest performance (highest Pearson correlation values for all subpanels (iii) in Figs 2 and 4B).

***In vivo* calcium imaging data.** All procedures were performed in accordance with the animal care and handling guidelines of the University of Edinburgh animal welfare committee, and were performed under a UK Home Office project license.

For the data in Fig. 5, all surgical and imaging procedures are detailed in our previous publication⁴. Briefly, adeno-associated viruses (AAV1.Syn.GCaMP6f.WPRE.SV40, University of Pennsylvania Vector Core, PA, USA) were locally injected in V1 at three different depths (-50 , -400 , and $-600 \mu\text{m}$) in 8 mice (8- to 10-week-old). The mice were obtained by crossing Cre-driver transgenic mice lines (Sst<tm2.1(cre)Zjh> [RRID:IMSR_JAX:013044] $n = 3$, Pvalb<tm1(cre)Arbr> (PV-Cre) [RRID:IMSR_JAX:008069] $n = 1$, or Vip<tm1(cre)Zjh> [RRID:IMSR_JAX:010908] $n = 4$) with Rosa-CAG-LSL-tdTomato [RRID:IMSR_JAX:007914] mice. All mice were originally obtained from Jackson Laboratory, ME, USA. Mice were group housed (typically 2–4 mice) and both male and female mice were used for the experiments. After 2–3 weeks of expression, the running speed and GCaMP6f signals were simultaneously recorded, during the presentation of drifting gratings (the total visual stimulation times were 720–1200 seconds per imaged field of view⁴). For the results presented in Fig. 5, we did not exclude tdTomato-positive neurons such that all GCaMP6f labelled neurons were included in the analysis.

Motion correction. To perform motion correction for the *in vivo* data presented in Figs 4 and 5 we used the discrete Fourier method from the SIMA toolbox⁸.

Data availability statement. The toolbox described in this paper is available at <https://github.com/rochefort-lab/fissa>.

References

- Huber, D. *et al.* Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**, 473–478 (2012).
- Margolis, D. J. *et al.* Reorganization of cortical population activity imaged throughout long-term sensory deprivation. *Nat Neurosci* **15**, 1539–1546 (2012).
- Chen, T. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
- Pakan, J. *et al.* Behavioural state modulation of inhibition is context-dependent and cell-type specific in mouse V1. *Elife* **5**, e14985 (2016).
- Attinger, A., Wang, B. & Keller, G. Visuomotor coupling shapes the functional development of mouse visual cortex. *Cell* **169**, 1291–1302 (2017).
- Dombeck, D., Khabaz, A., Collman, F., Adelman, T. & Tank, D. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
- Greenberg, D. & Kerr, J. Automated correction of fast motion artifacts for two-photon imaging of awake animals. *Journal of Neuroscience Methods* **176**, 1–15 (2009).
- Kaifosh, P., Zaremba, J. D., Danielson, N. B. & Losonczy, A. SIMA: Python software for analysis of dynamic fluorescence imaging data. *Frontiers in neuroinformatics* **8** (2014).
- Pnevmatikakis, E. A. *et al.* Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* **89**, 285–299 (2015).
- Muir, D., Roth, M., Helmchen, F. & Kampa, B. Model-based analysis of pattern motion processing in mouse primary visual cortex. *Frontiers in neural circuits* **9** (2015).
- Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, <https://doi.org/10.1101/061507> (2016).
- Dubbs, A., Guevara, J. & Yuste, R. moco: Fast motion correction for calcium imaging. *Frontiers in Neuroinformatics* **10** (2016).
- Pnevmatikakis, E. & Giovannucci, A. Normcorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *bioRxiv*, <https://doi.org/10.1101/108514> (2017).
- Peron, S., Freeman, J., Iyer, V., Guo, C. & Svoboda, K. A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron* **86**, 783–799 (2015).
- Maruyama, R. *et al.* Detecting cells using non-negative matrix factorization on calcium imaging data. *Neural Networks* **55**, 11–19 (2014).
- Diego, F. & Hamprecht, F. Sparse space-time deconvolution for calcium image analysis. *NIPS* **27**, 64–72 (2014).
- Apthorpe, N. *et al.* Automatic neuron detection in calcium imaging data using convolutional networks. *NIPS* **29** (2016).
- Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data. *Neuron* **63**, 747–760 (2009).
- Peron, S., Chen, T. & Svoboda, K. Comprehensive imaging of cortical networks. *Current opinion in neurobiology* **32**, 115–123 (2015).

20. Stetter, M. *et al.* Principal component analysis and blind separation of sources for optical imaging of intrinsic signals. *NeuroImage* **11**, 482–490 (2000).
21. Ji, N., Sato, T. & Betzig, E. Characterization and adaptive optical correction of aberrations during *in vivo* imaging in the mouse cortex. *PNAS* **109**, 22–27 (2012).
22. Cichocki, A. & Anh-Huy, P. H. A. N. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences* **92**, 708–721 (2009).
23. Langville, A. N., Meyer, C. D., Albright, R., Cox, J. & Duling, D. Algorithms, initializations, and convergence for the nonnegative matrix factorization. *arXiv* 1407.7299 (2014).
24. Svoboda, H. K. Simultaneous imaging and loose-seal cell-attached electrical recordings from neurons expressing a variety of genetically encoded calcium indicators. *GENIE project, Janelia Farm Campus, CRCNS.org* (2015).
25. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
26. Keller, G., Bonhoeffer, T. & Hübener, M. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* **74**, 809–815 (2012).
27. Ayaz, A., Saleem, A., Schölvink, M. & Carandini, M. Locomotion controls spatial integration in mouse visual cortex. *Current Biology* **23**, 890–894 (2013).
28. Erisken, S. *et al.* Effects of locomotion extend throughout the mouse early visual system. *Current Biology* **24**, 2899–2907 (2014).
29. Fu, Y. *et al.* A cortical circuit for gain control by behavioral state. *Cell* **156**, 1139–1152 (2014).
30. Dadarlat, M. & Stryker, M. Locomotion enhances neural encoding of visual stimuli in mouse v1. *Journal of Neuroscience* **37**, 3764–3775 (2017).
31. Friedrich, J. *et al.* Multi-scale approaches for high-speed imaging and analysis of large neural populations. *PLoS Comput Biol* **13** (2017).
32. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
33. Lin, C. Projected gradient methods for non-negative matrix factorization. *Neural Computation* **19**, 2756–2779 (2007).
34. Boutsidis, C. & Gallopoulos, E. Svd based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition* **41**, 1350–1362 (2008).
35. Akerboom, J. *et al.* Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *The Journal of Neuroscience* **32**, 13819–13840 (2012).
36. Deneux, T. *et al.* Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations *in vivo*. *Nature Communications* **7** (2016).

Acknowledgements

This work was funded by the BBSRC grant BB/N023161/1 to M.v.R. and N.R., by the Wellcome Trust and the Royal Society (Sir Henry Dale fellowship to N.R.), the Marie Curie Actions of the European Union's FP7 program (MC-CIG 631770 to N.R. and IEF 624461 to J.P.), the Shirley Foundation, the Patrick Wild Center, the RS MacDonald Charitable Trust Seedcorn Grant, the Simons Initiative for the Developing Brain (to N.R.), and the Graduate School of Life Sciences, University of Edinburgh (to E.D.). S.K. was supported by the EuroSpin Erasmus Mundus program and S.K. and S.L. were funded by the EPSRC Doctoral Training Centre in Neuroinformatics (EP/F500386/1 and BB/F529254/1).

Author Contributions

N.R. and S.K. designed the project. S.K. and S.L. were jointly responsible for coding the toolbox. J.P., E.D., and N.R. acquired and analysed the *in vivo* data. S.K., J.P., N.R. and E.D. tested the different versions of the toolbox and provided feedback on the results. S.K., M.v.R., and N.R. wrote the paper, with input from all authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-21640-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018