Manual parts of spike-sorting

1. Are these waveforms from a neuron or noise?
2. Are the selected waveforms sufficiently homogenous/isolated?
   - Isolated from waveforms from other units
   - Isolated from noise

Automating the Manual parts of spike-sorting

1. Are these waveforms from a neuron or noise?
2. Are the selected waveforms sufficiently homogenous/isolated?
    - Isolated from waveforms from other units
    - Isolated from noise

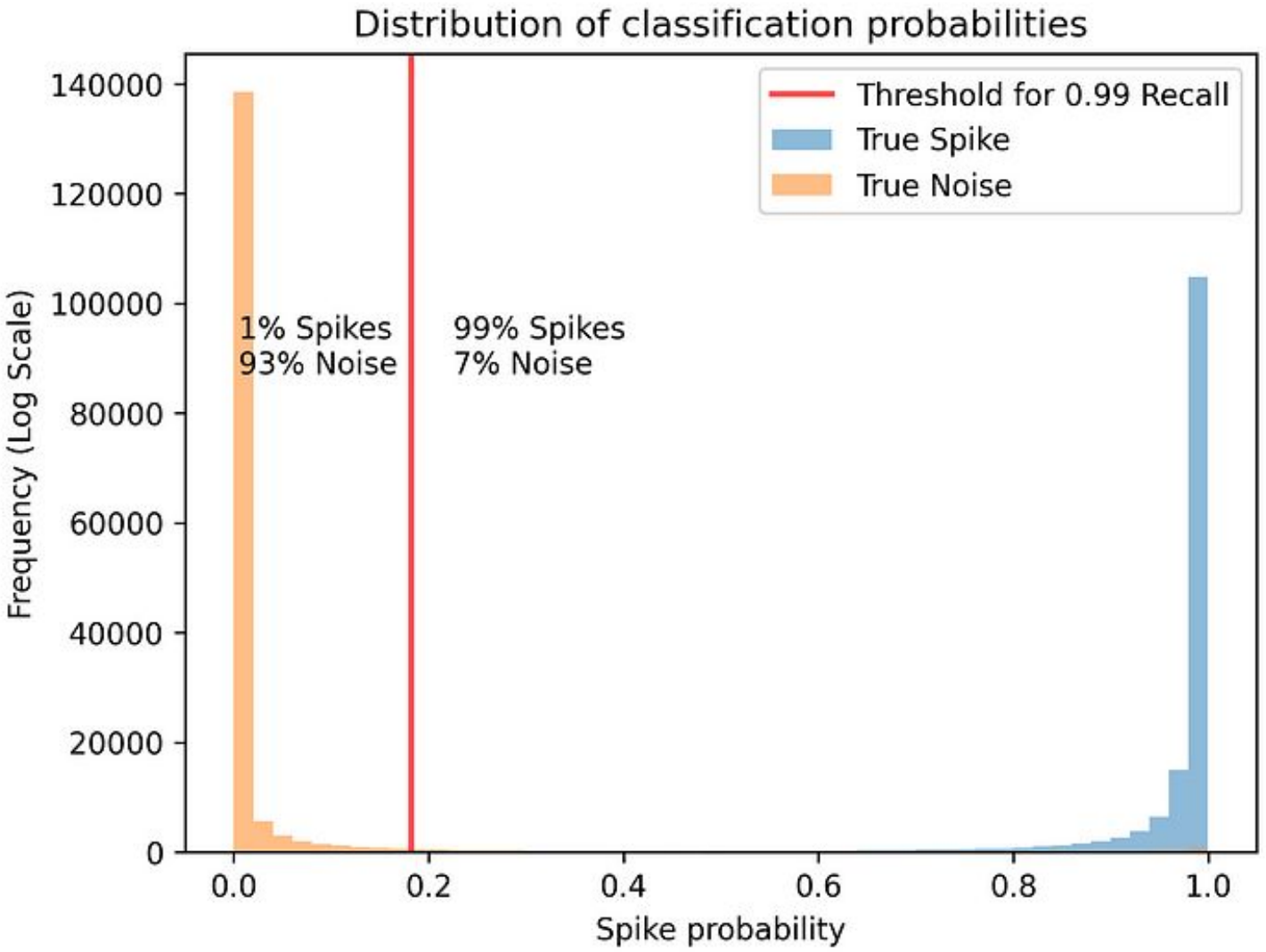
Potential Solutions

1. Train a classifier to say neuron or noise
2. Use a Bayesian Gaussian Mixture model to determine "best" number of clusters
    - Implicitly, this will suggest a clustering configuration with the best isolation given the data
3. Select the right clusters

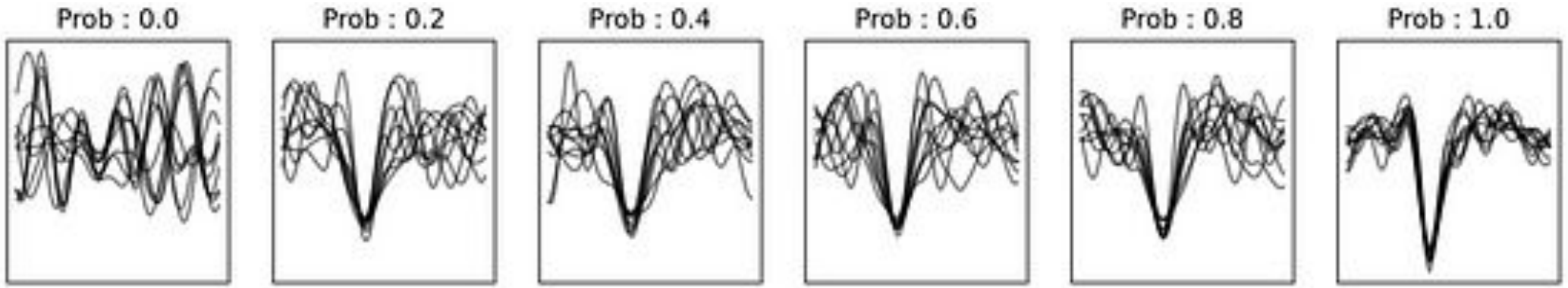
1.Making and testing the classifier

- XGBoost classifier
  - Fast
  - Effective
- ~2.5 million waveforms
  - ~48% true spikes
  - ~52% noise
  - manually labelled

Maximizing Recall (True spikes kept)

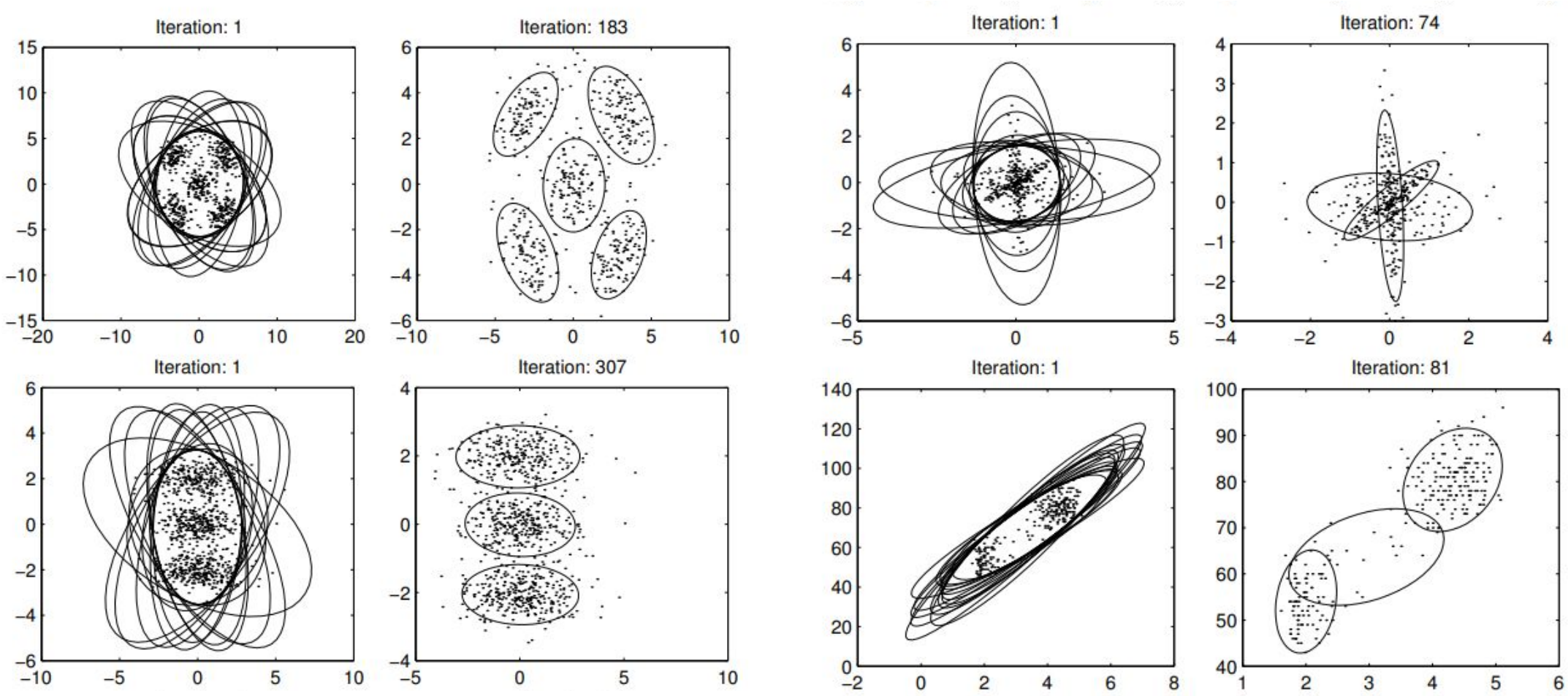


1.Making and testing the classifier



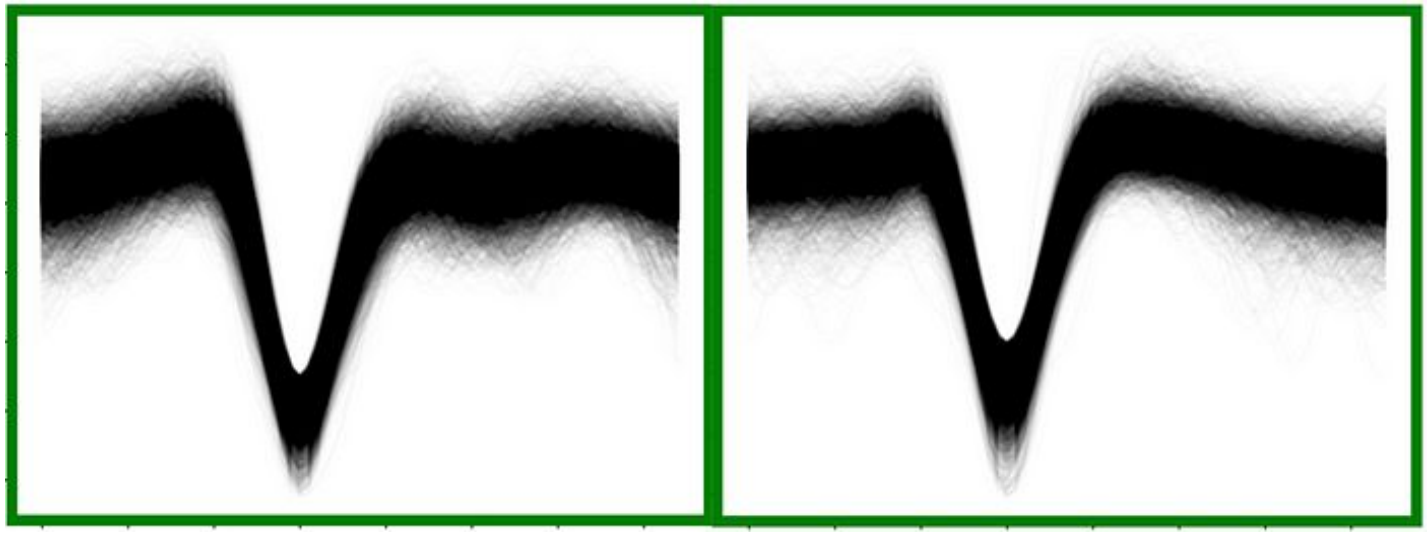
2. Automating clustering of waveforms

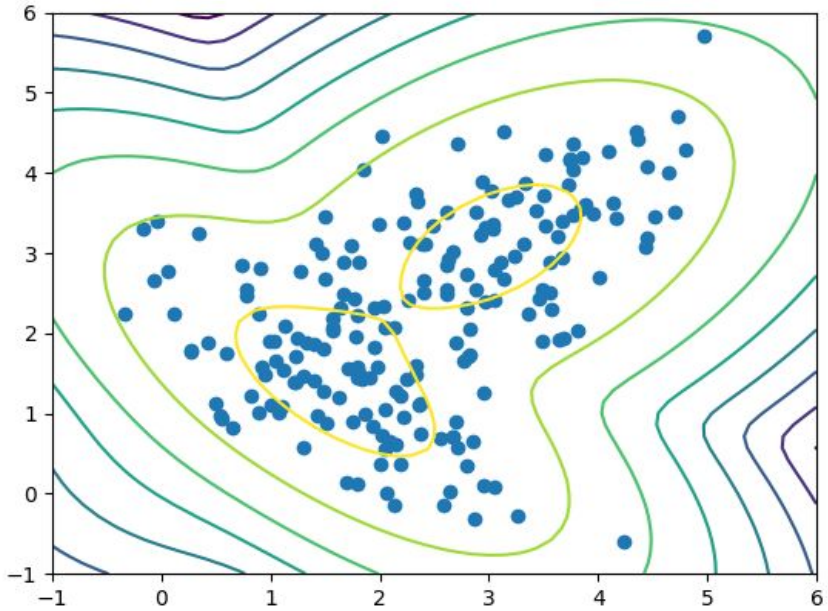
- Bayesian Gaussian Mixture Model
  - Only have to specify maximum allowed number of clusters
  - Sklearn.mixture.BayesianGaussianMixture



2. Automating clustering of waveforms

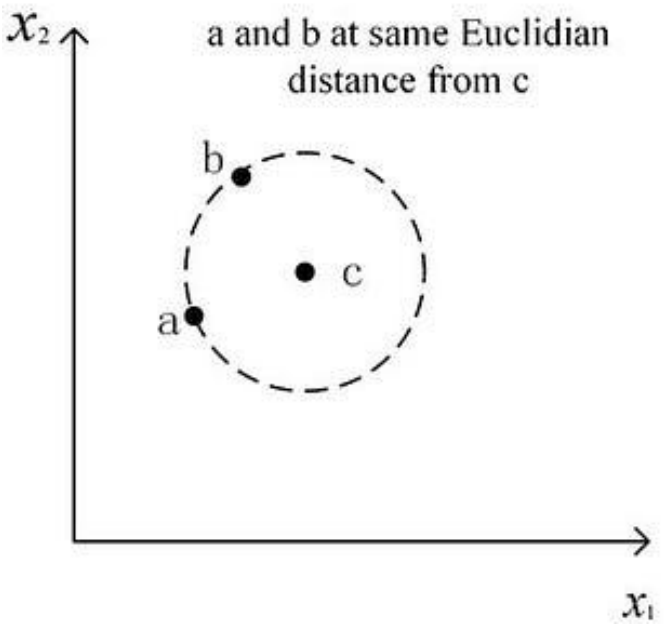
- BGM tends to over-split
- Have to figure out how to automatically merge clusters

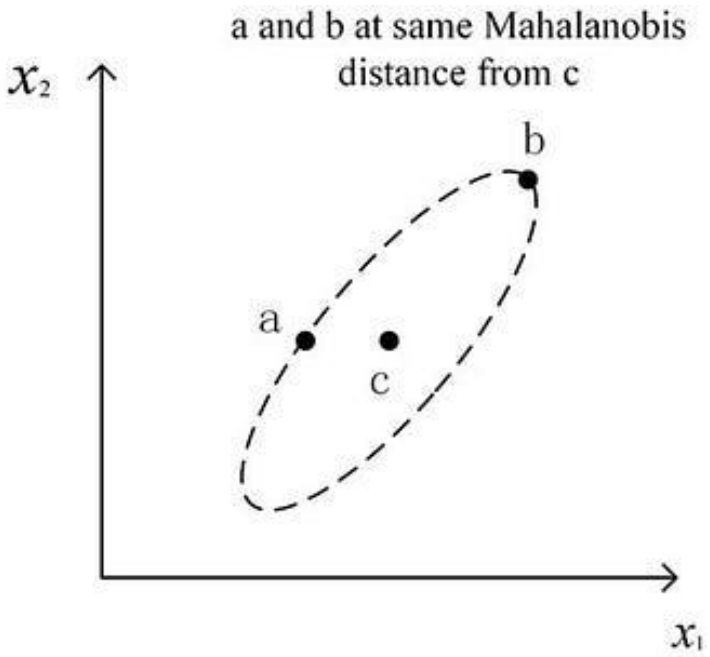


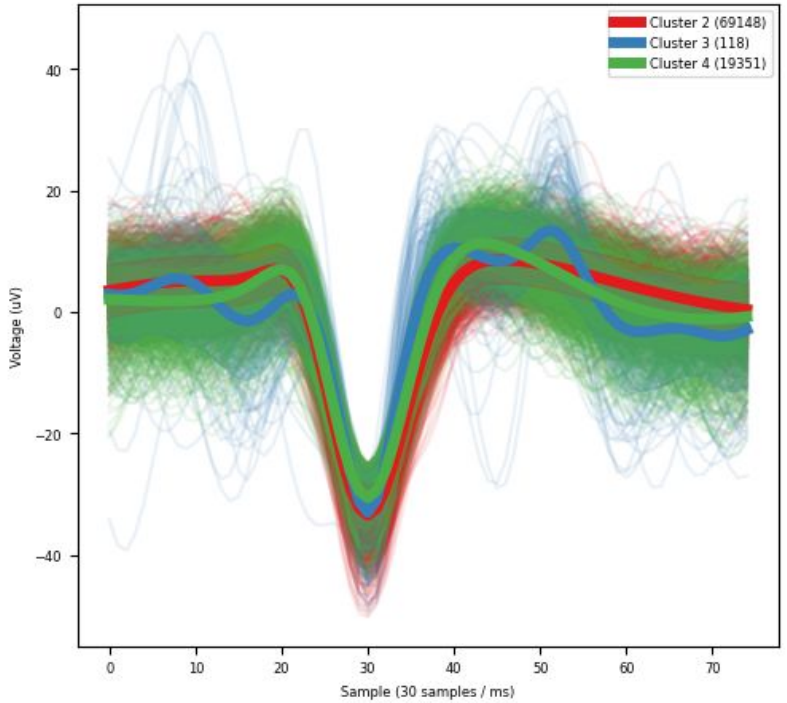


2. Automating clustering of waveforms

- BGM tends to over-split
- Have to figure out how to automatically merge clusters







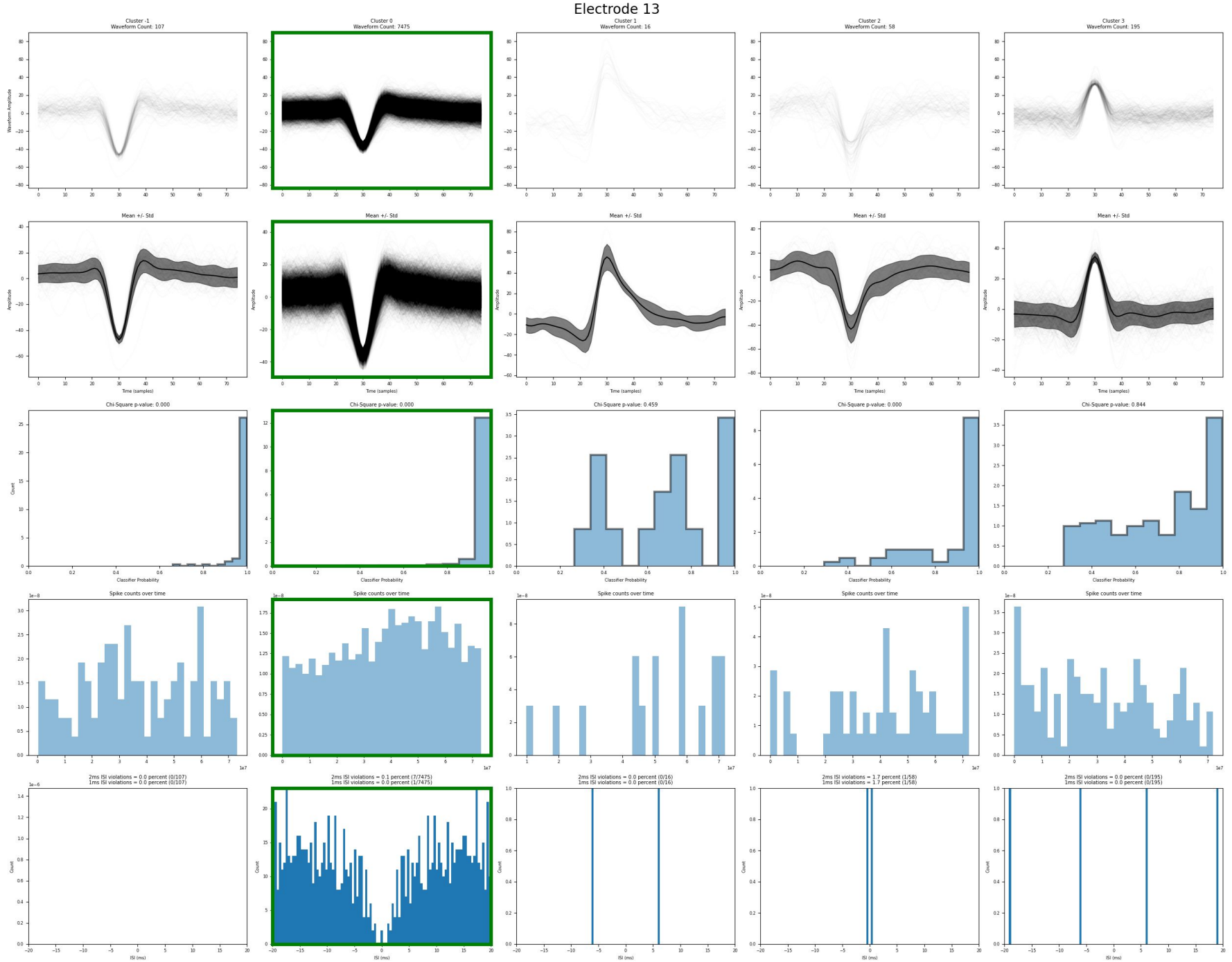
3. Selecting the right clusters

1. We know whether a single waveform is noise or spike
2. How do we select whether a cluster is noise or spiking

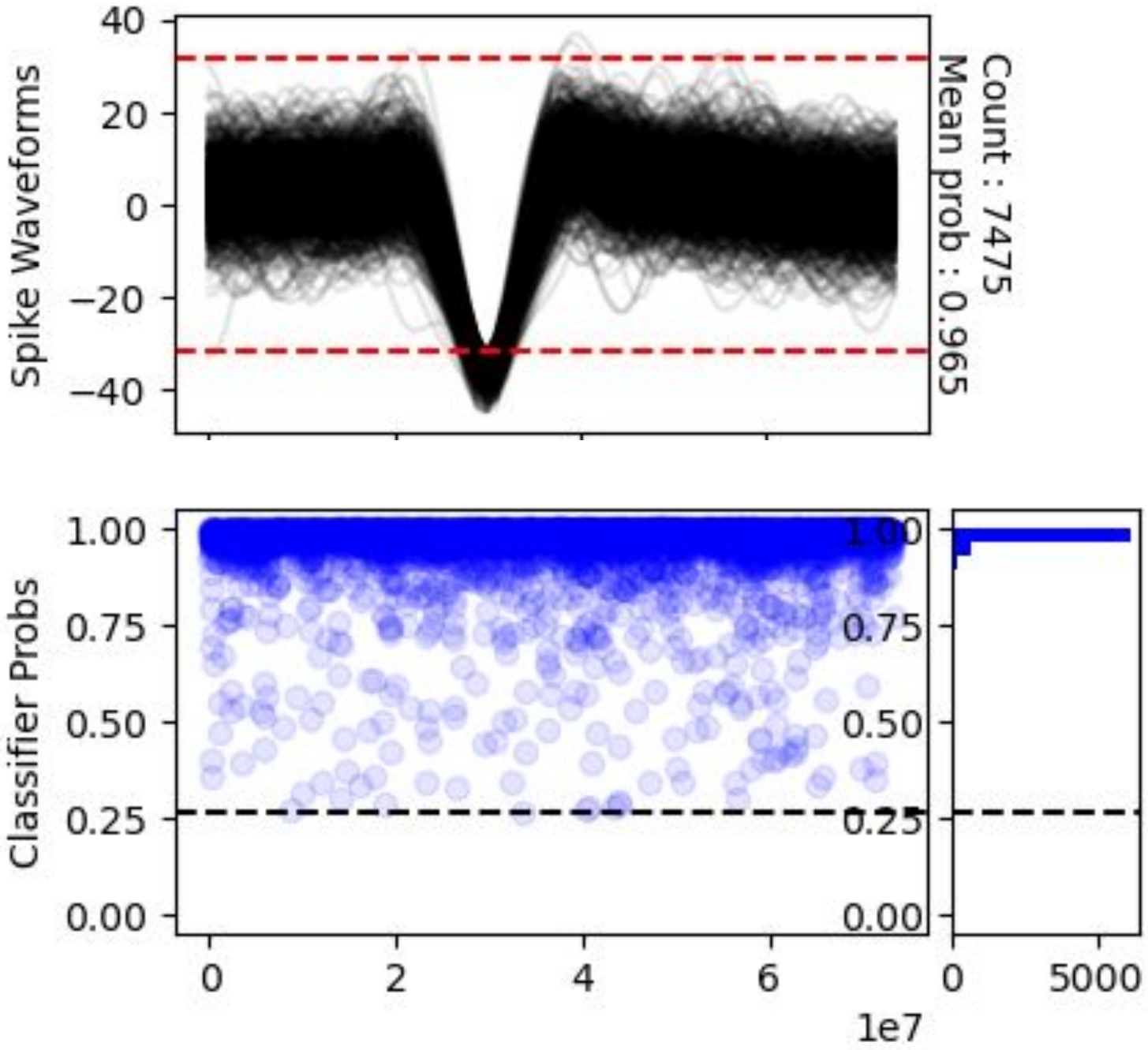
Criteria

1. Minimum number of waveforms (~2000)
2. Inter-spike interval violations
3. Distribution of classification probability (chi-square test for non-uniformity)

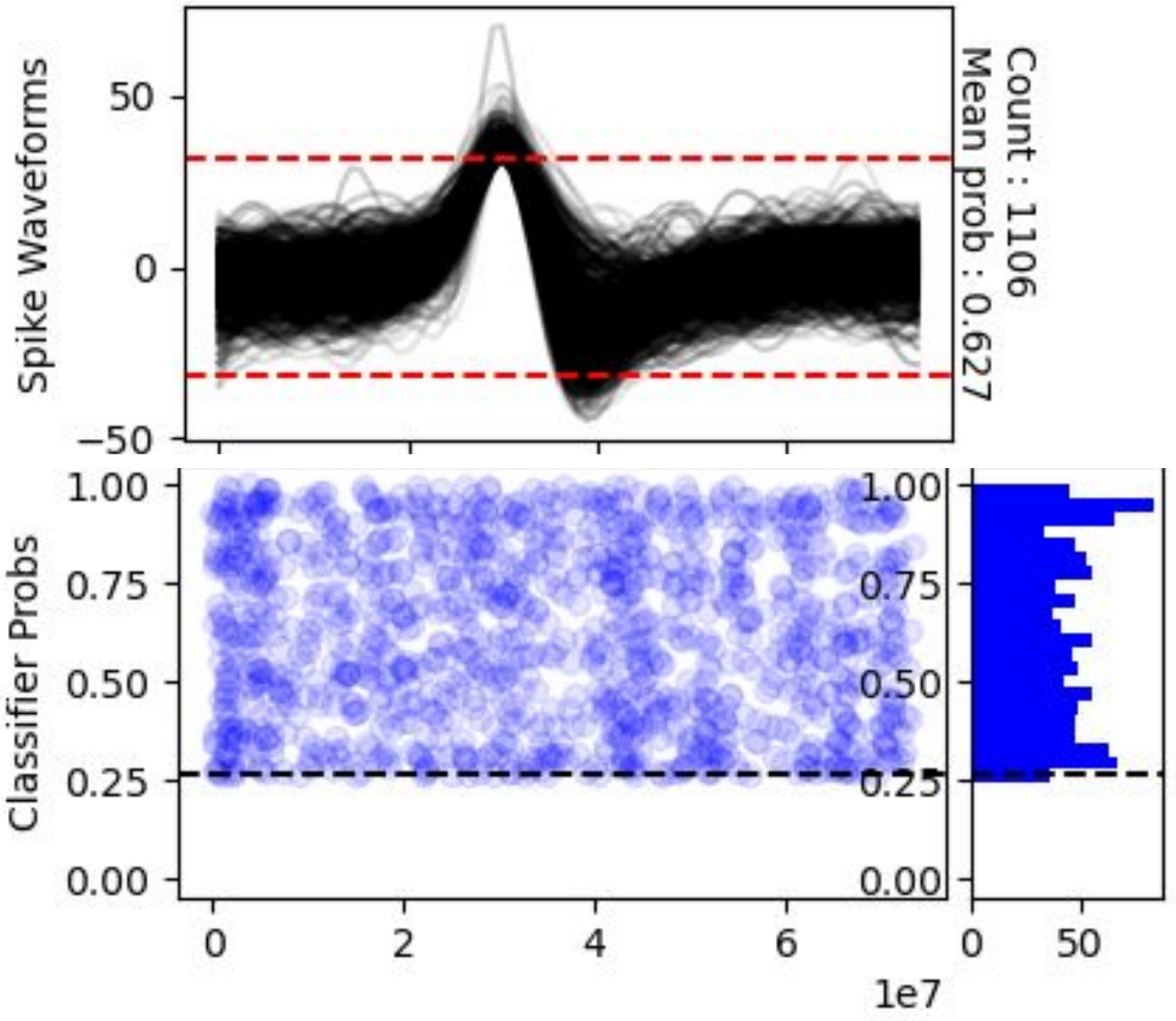
3. Selecting the right clusters



3. Selecting the right clusters



3. Selecting the right clusters

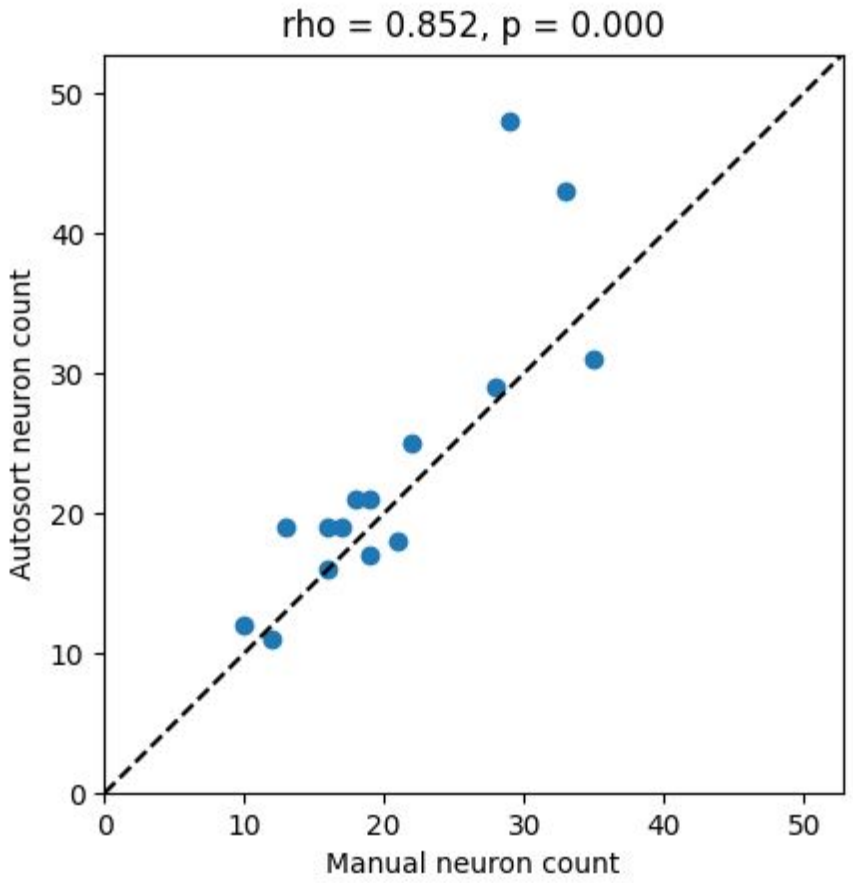


How well does it work?

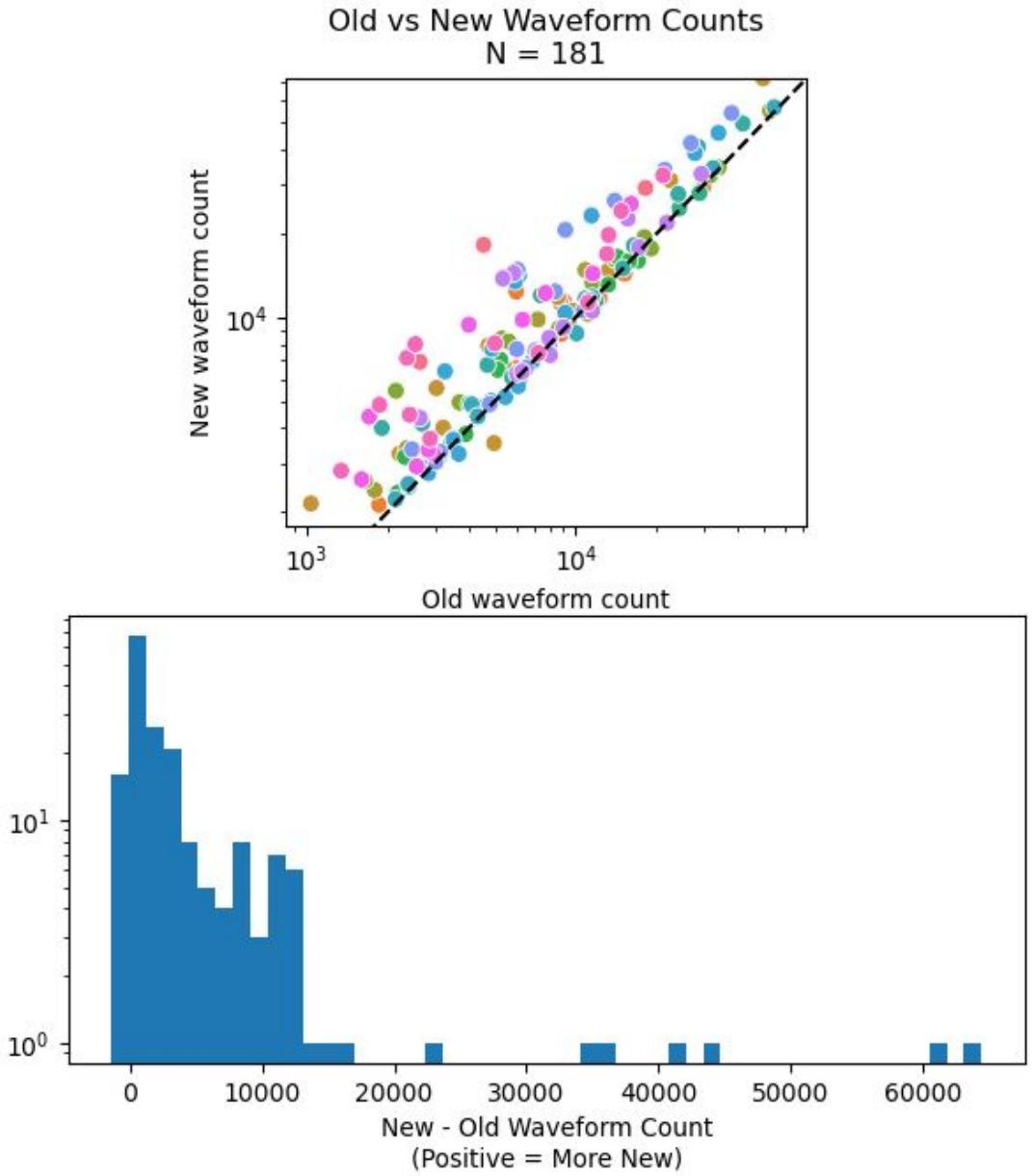
1. At the unit level
   - Comparing number of detected neurons
   - Comparing waveforms detected per unit
2. At the activity level
   - Single-trial activity
   - Mean-trial activity
   - Changes in similarity of activity with neuron counts

How well does it work?

Comparing selected neuron counts

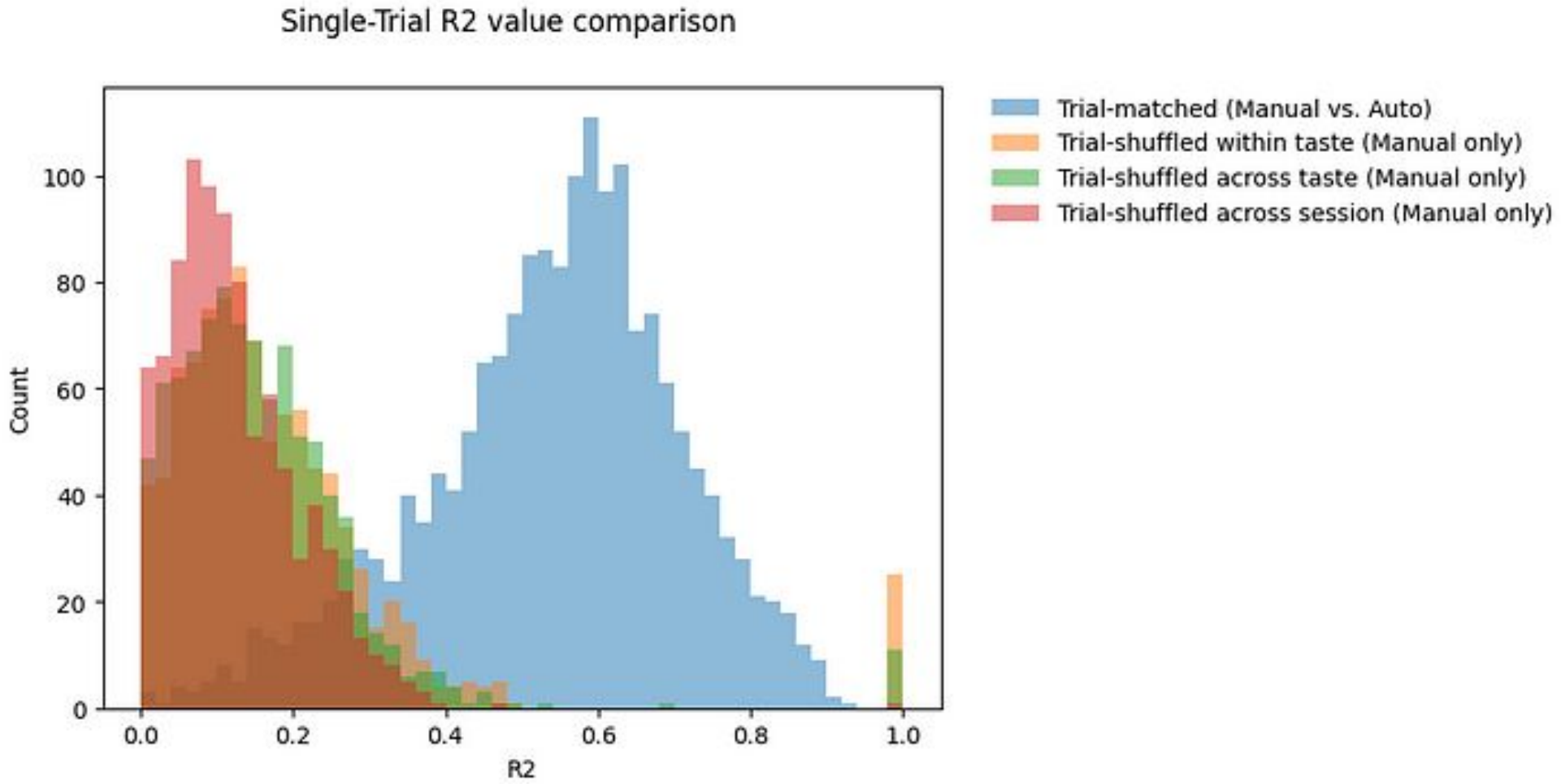


Comparing waveform counts



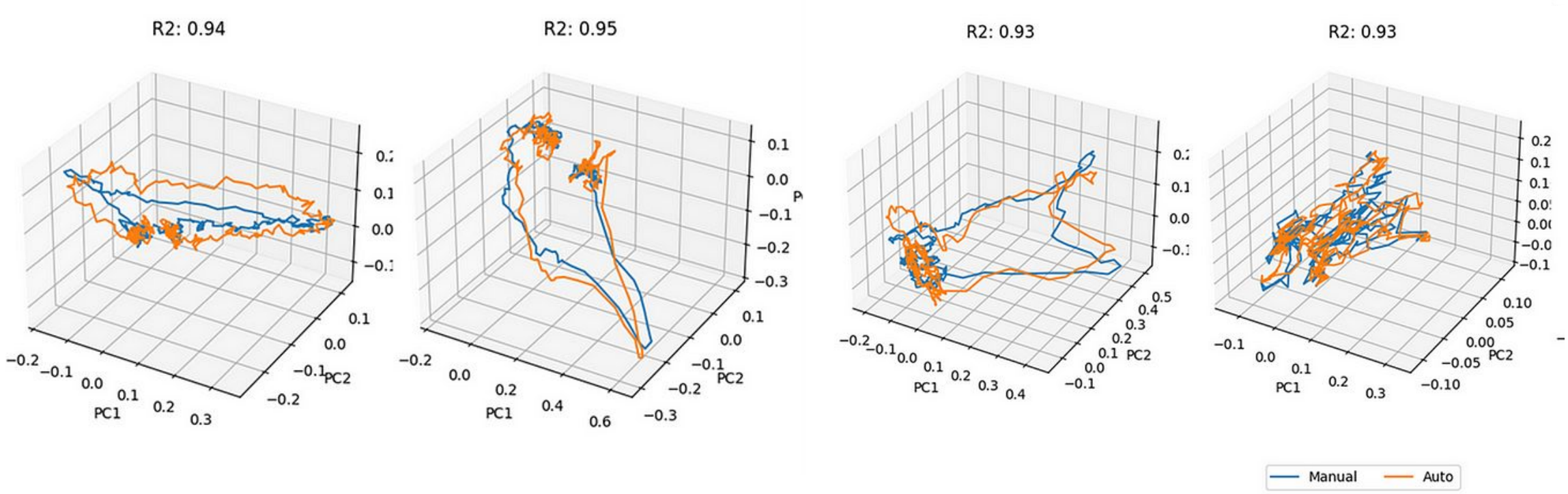
How well does it work?

Comparison of single-trial firing rates (population PCA)



How well does it work?

Trial-averaged population PCA



How well does it work?

Changes in similarity of mean neural activity vs. Neuron Counts