Contents

1 Motivation

In their review entitled 'Measuring and interpreting neural correlations', Cohen and Kohn studied the effects of response strength and time bin width on neural correlation measurement. Using simulated data, they found that correlation measurements taken between weakly responding neurons (i.e. neurons with firing rates of < 10 spikes per s) will be less than the true pairwise correlation value. They also found that binning the neural spiking data into small time bins can cause the measured correlation to be less than the true value. How small is 'small' in this case is related to width of the cross-correlogram of the pair of neurons. But, it was shown in simulation that time bins should be at least 0.1 seconds wide in order for the measured correlation value to be close to the true correlation value [1].

Our aim here is to measure a big enough sample of pairwise correlations from actual neural data across a selection of time bin widths and compare our results to those of Cohen and Kohn. In order to do this we used data collected using 'Neuropixels' probes [2]. These data are made available publicly online by Dr. Nick Steinmetz [1].

2 Data

The recent development of Neuropixels probes has allowed extracellular voltage measurements to be collected from multiple brain regions simultaneously routinely, and in much larger numbers than traditional methods.

Using two probes, spiking activity was simultaneously collected from over 800 neurons in an awake mouse brain for a period of 84 minutes. During this period, the mouse was shown various visual stimuli. The 800 neurons were distributed across 5 different brain regions: V1, hippocampus, thalamus, motor cortex, and striatum.

The data consist of spike timings and cell/cluster identifiers associating spikes with a certain cell/cluster and a certain time. Each cell/cluster is classified as 'good', 'multi-unit activity', or 'unsorted', referring to the quality of spike sorting. This classification was performed by those who analysed the data. Only the cell/clusters classified as 'good' are used in this project.

2.1 Stimulus

The stimulus was a full field moving bar grating. There were 17 stimulus conditions corresponding to 16 drift directions (0 degrees to 337.5 degrees in 22.5 degree increments) with 2Hz temporal frequency and 0.08 cycles/degree spatial frequency (conditions $1-16$) plus a blank condition (17). Each condition was presented 10 times for 2 seconds each time, with 1.5 seconds between trials.

[1] `http://data.cortexlab.net/dualPhase3/`

3 Methods

3.1 Binning data

The data were divided into time bins or various widths ranging from 0.01 to 2 seconds. If the bin width was not an integer divisor of the trial period (2 seconds), only the bins that lay totally within the trial period were included. For example, when dividing the trials into bins of 0.3 seconds, the final bin of 0.2 seconds was excluded. We measured the number of spikes in each time bin.

When performing calculations on the binned data, each bin was treated as an individual measurement. For example, when calculating the spike count correlation coefficient for a given pair of neurons, if the time bin used was 2.0 seconds, then we had 10 measurements for each neuron with which to calculate the coefficient. But, if we were using a bin width of 1.0 second, then we would have 20 measurements for each neuron.

3.2 Pairing strongly responding neurons

A weak response, or low firing rate, has a diminishing effect on measured correlation [1]. In order to avoid this effect, we filtered out any neurons with a mean firing rate of less than 10 spikes per second measured across the 10 trials. Once these neurons were filtered out, we randomly chose 30 pairs from all the possible pairs of the remaining neurons. We used these 30 pairs to calculate 30 correlation coefficients. We repeated this process for each brain region, and each stimulus condition.

If less than 9 strongly responding neurons were found it was not possible to make 30 pairs of strongly responding neurons. In that case, we just used all the strongly responding pairs.

3.3 Correlation coefficient

We calculated Pearson's correlation coefficient for pairs of neurons. For jointly distributed random variables X and Y, Pearson's correlation coefficient is defined as:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{1}$$

$$= \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{2}$$

where E denotes the expected value, μ denotes the mean, and σ denotes the standard deviation. The correlation coefficient is a normalised measure of the covariance. It can take values between 1 (completely correlated) and -1 (completely anti-correlated). Two independent variables will have a correlation coefficient of 0. But, having 0 correlation does not imply independence.

If we do not know the means and standard deviations required for equation 1, but we have samples from X and Y, Pearson's sample correlation coefficient is defined as:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3}$$

where $\{(x_i, y_i)\}$ for $i \in \{1, \ldots, n\}$ are the paired samples from X and Y, and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the sample means.

In practice we used the python function `scipy.stats.pearsonr` to calculate the correlation coefficients.

3.3.1 Spike Count Correlation, r_{SC}

The spike count correlation (r_{SC}) of two cells is the correlation between the spike counts of those cells in response to a given stimulus condition.

3.3.2 Signal Correlation, r_{signal}

The signal correlation of two cells (r_{signal}) is the correlation between the mean responses of those cells to each stimulus condition.

3.4 Separating Correlations & Anti-correlations

In order to compare the effect of bin width on measures of negative r_{SC} (anti-correlation) and positive r_{SC} separately, we had to separate correlated and anti-correlated pairs. To do this, we simply measured the mean r_{SC}, taking the mean across all the bin widths. If this quantity was positive or zero we regarded the pair as positively correlated. If this quantity was negative we regarded the pair as anti-correlated.

4 Results

4.1 Spike Count Correlations

For each stimulus condition and each region, we randomly chose 30 pairs of strongly responding neurons (or chose as many pairs as we could find), measured their absolute spike count correlations ($|r_{SC}|$) and absolute signal correlation ($|r_{signal}|$) using various values for the time bin width and examined the relationship between the coefficients and the bin width. We found that $|r_{SC}|$ increased approximately log-linearly up to a width of 1s and appeared to level off thereafter, see figure 1, or figure 2 for a linear version. This is in agreement with the findings of the simulated experiment in [1]. For these figures, we chose the stimulus condition that evoked the strongest response from a given region. This guaranteed us 30 strongly responding pairs for each region.

4.1.1 Correlations & Anti-correlations

We separated the 30 randomly chosen pairs into correlated and anti-correlated pairs as per section 3.4. We then measured r_{SC} across various bin width values. We found that the positive r_{SC} got more positive as the bin width increased. We also found that the negative r_{SC} got more negative as the bin width increased. Both positive and negative coefficients showed a log-linear increase similar to the absolute spike count correlation coefficients. See figure 3.

4.2 Signal Correlations

We found that $|r_{signal}|$ maintained a constant value across different bin widths. This was the case in all regions, with some variations in V1. See figure 4.

Things to discuss with Cian

1. Is all of the above clear?
2. Are my methods sound?
3. Could any of my methods be improved?
4. How can this document be improved? (The figures with the linear x-axis might be better.)

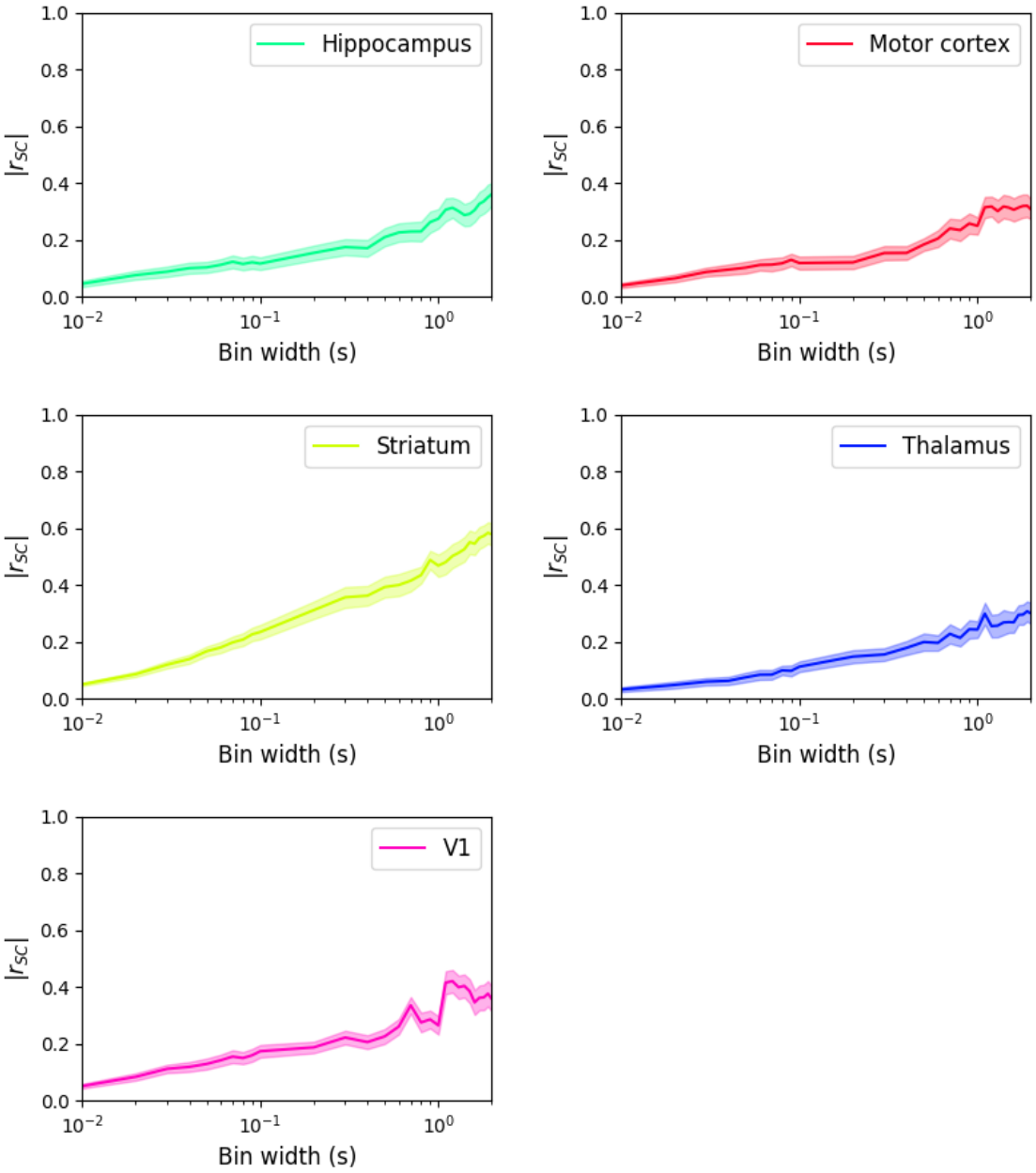

Figure 1: **Spike Count Correlations:** Absolute values of correlation coefficients measured using different time bin widths. Shaded areas indicate standard errors. One figure for each brain region from which data were available. The absolute correlation increases approximately log-linearly with the bin width up to a bin width of 1 second and levels out somewhat thereafter.

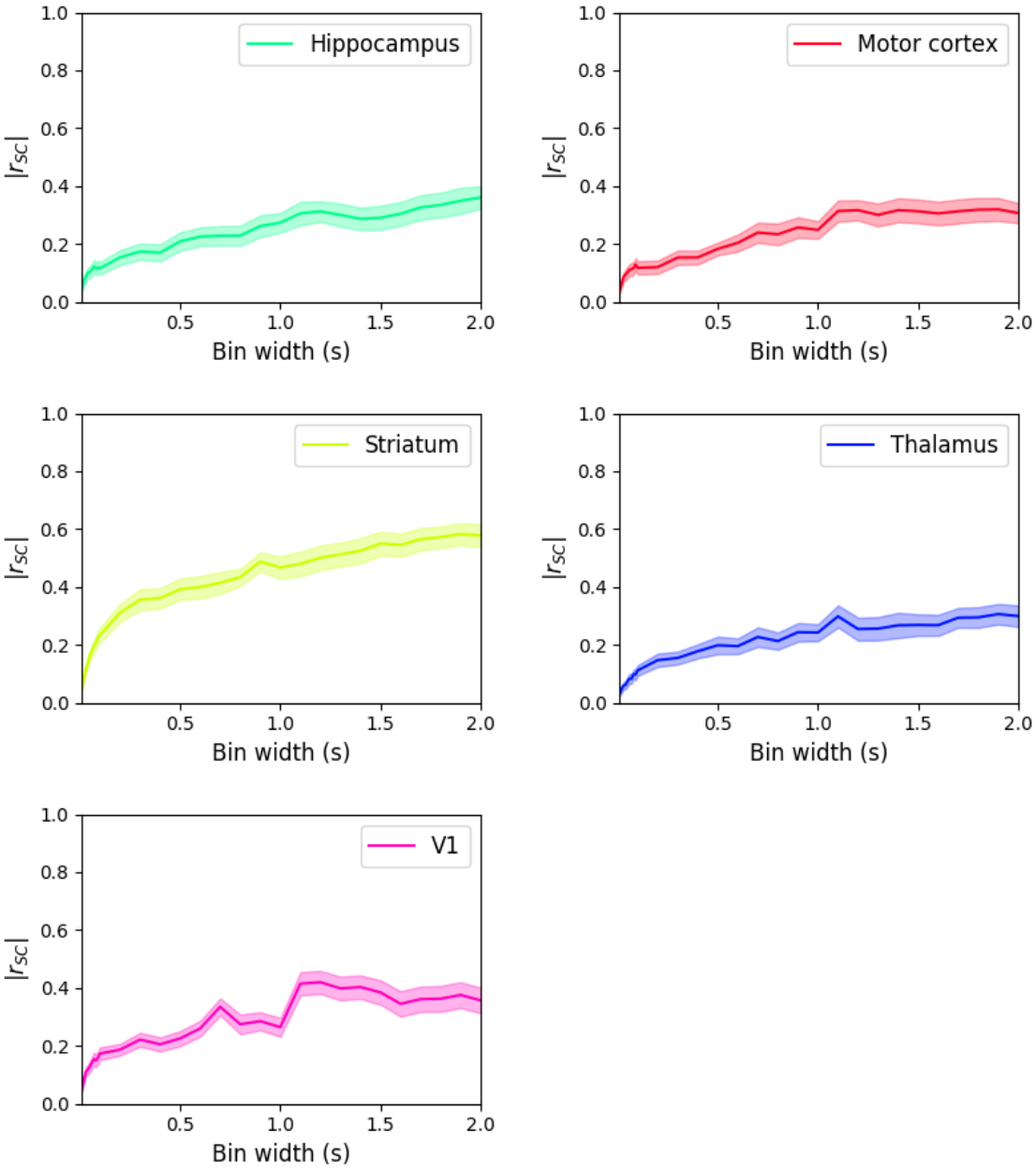

Figure 2: **Spike Count Correlations:** Absolute values of correlation coefficients measured using different time bin widths. Shaded areas indicate standard errors. One figure for each brain region from which data were available. The absolute correlation increases approximately log-linearly with the bin width up to a bin width of 1 second and levels out somewhat thereafter.

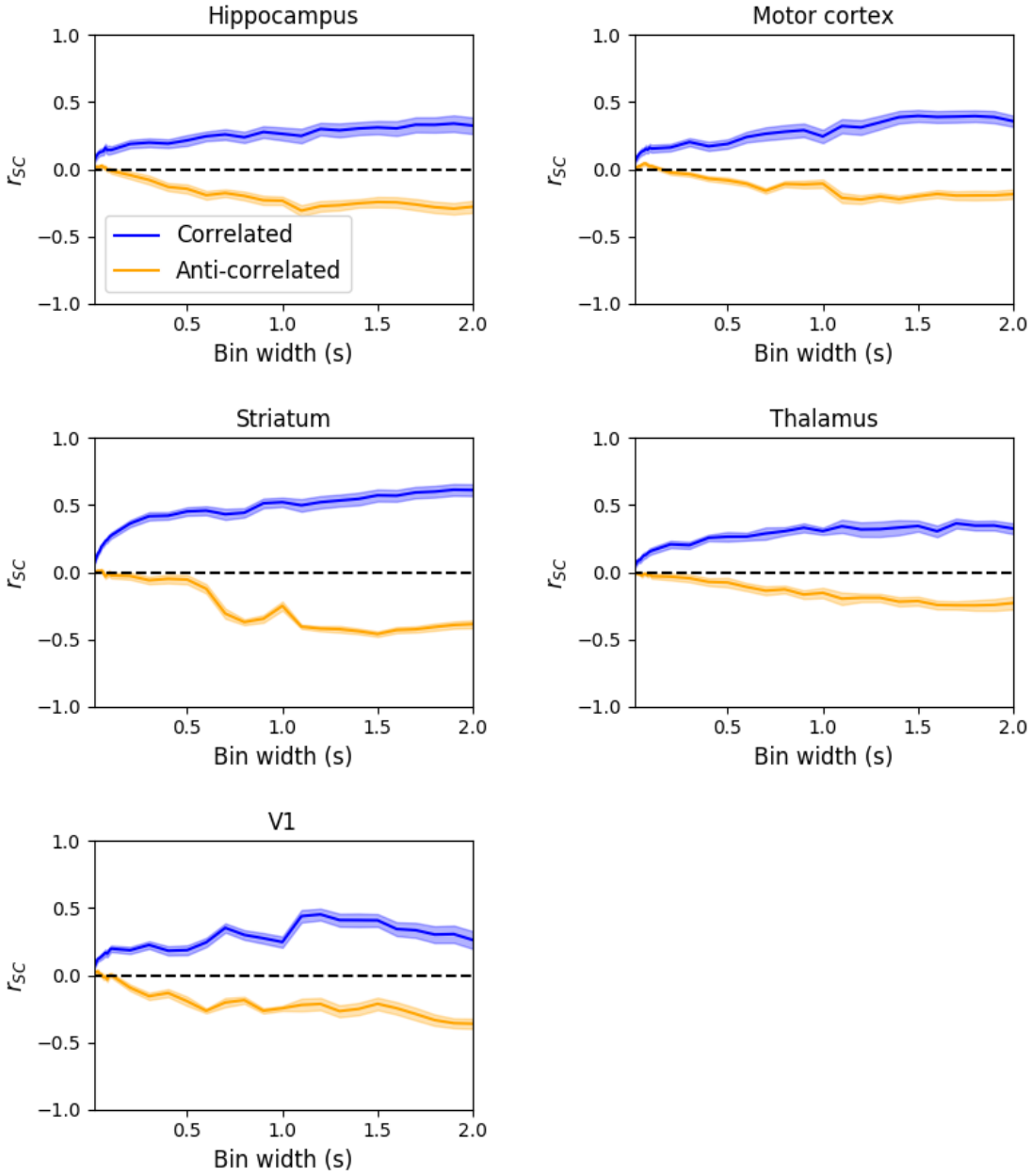

Figure 3: **Spike Count Correlations and Anti-correlations:** Correlation coefficients measured using different time bin widths and separated into positive and negative sets. Shaded areas indicate standard errors. One figure for each brain region from which data were available. The correlations or anti-correlations get stronger as the bin width increases. We see the log linear increase similar to figures 1 and 2.

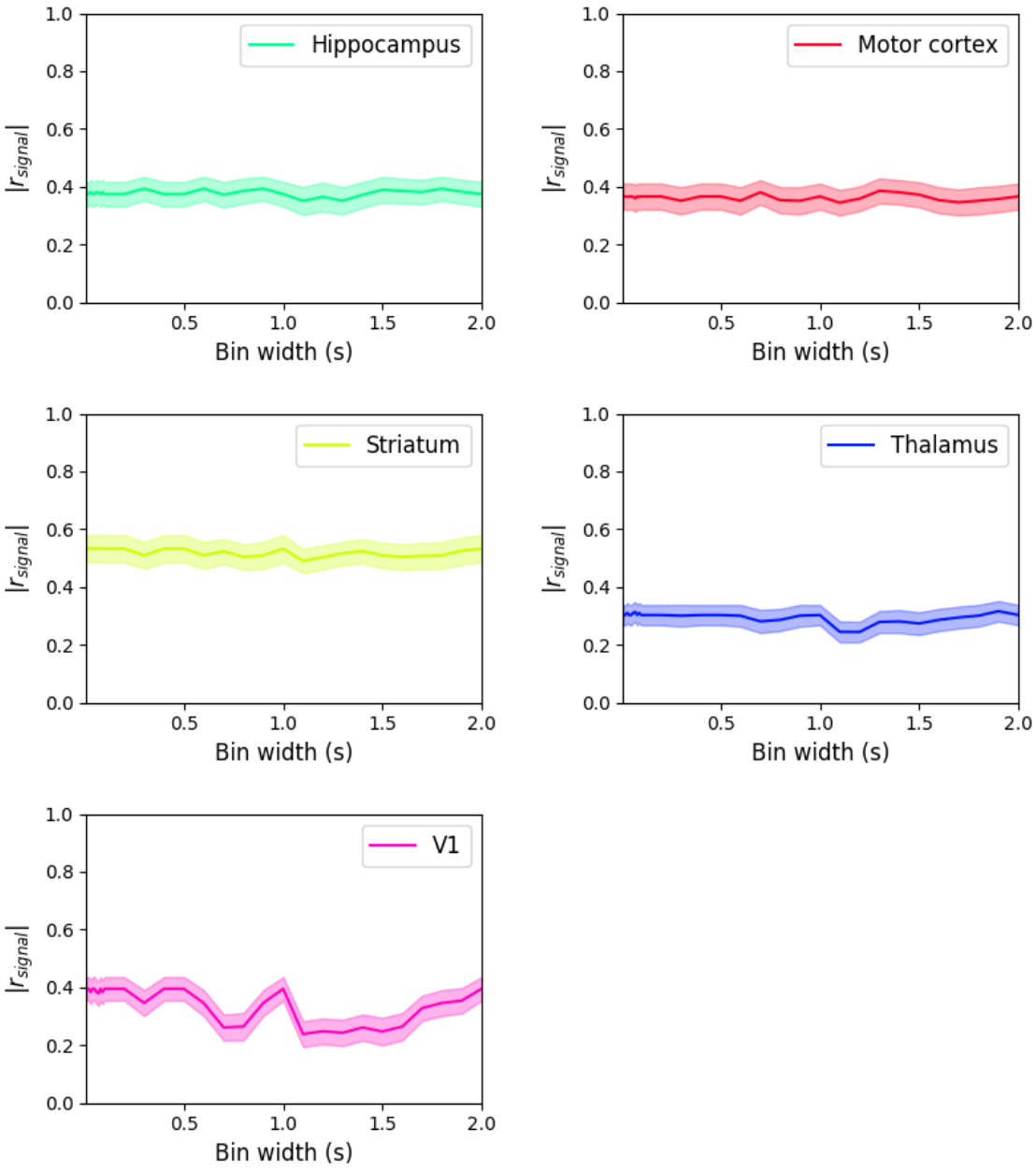

Figure 4: **Signal Correlations:** Absolute values of signal correlation coefficients measured using different time bin widths. Shaded areas indicate standard errors. One figure for each brain region from which data were available. The signal correlation maintains a fairly constant value across bin widths for most brain regions, with some variations in V1.

References

[1] Marlene R Cohen, Adam Kohn, *Measuring and interpreting neural correlations*. Nature Neuroscience 14, 811-819, (2011)

[2] James J. Jun, Nicholas A. Steinmetz, Timothy D. Harris, *Fully integrated silicon probes for high-density recording of neural activity*. Nature 551, 232–236, (2017)