

## Daily Activities

- 19/09/2016 - 15/11/2016** During this time I read most of the papers mentioned in Cian's model paper describing different models of neuronal networks. I also attempted to model some of Mike Ashby's data, but we only had data for one point in time. So analysing changes in internal representation was not possible. I also worked on a Kaggle challenge to predict epileptic seizures. I also started working on a project to model the fluorescence created by calcium indicators.
- 16/11/2016** Read about piecewise-deterministic Markov processes (PDMPs), mostly mathematical definitions and concepts. The main point to take away is: In order to solve the system, we get the jump times for poisson process, and we solve the ODE for the periods in between the jump times. We change the values of the ODE variables according to the jumps of the poisson process. So, solve the ODE for the interval  $t = [0, T_{1^{st}jump})$ , then apply the variable changes, then solve the ODE for the interval  $t = [T_{1^{st}jump}, T_{2^{nd}jump})$ , and so on.
- 17/11/2016** Worked on understanding the `julia` implementation of a PDMP solver.
- 18/11/2016** Attempted to write a functioning PDMP in `julia`. Having problems with concentrations going negative.
- 21/11/2016** Meeting with Cian. Recalculated the number of molecules of buffer and free calcium. Use of these numbers in `julia` results in an error from the Sundials ODE solver. Learning what a 'stiff' ODE is may help. Trying other solvers shouldn't be too difficult. Need to read about this 'true jump method'.
- 22/11/2016** Read about Ramain Veltz's method for solving the PDMPs. Didn't learn much, the problem is with the ODE solver. The `julia` API to `cvoid` is not working. Need to try out different ones, or find one that works.
- 23/11/2016** Increased the relative tolerance to  $10^{-6}$  from  $10^{-7}$ . The `Sundials.cvoid` ODE solver works with those values. Need to make the model stable with a calcium influx and outflux around a baseline. Introducing spikes and bursts of spikes seems very possible.
- 24/11/2016** Good progress today. Managed to model a single spike and a burst of spikes using Julia. Emailed Cian. Need to hear from him. Maybe do some epilepsy tomorrow.
- 25/11/2016** Tidied up the `calcium_dynamics.jl` script. Implemented command line parameters. Need to find a way to keep the previous number of  $BCa^*$  and use it to calculate the jump.
- 26/11/2016** Ensured that the jump in the continuous variable associated with the change in the discrete variable was being implemented properly.
- 27/11/2016** No work. Raspberry Pi instead. Can't ssh between the pi and my laptop. I'm blaming eduroam.

**28/11/2016** Meeting with Cian and Mike today, need a list of questions for Mike. Mike lent me a book. It contains useful information, and formulas for analysis. Implementing this info could be useful if possible. The main takeaway from the meeting was that I should introduce the endogenous buffer, as this will change the dynamics of the system quite significantly. Read more tomorrow, and start thinking about the implementation of the endogenous buffer.

**29/11/2016** Used Thomas Griffith's thesis to get numbers for the endogeneous buffer. Started writing code to implement endogeneous buffer. Guessed at the forward/backward rates for Oregon Green BAPTA.

**30/11/2016** Added the code for the endogeneous buffer. Causes the rates to go negative immediately. All the figures need to be in the same dimensions. Currently putting all the figures in molecules per litre units. This causes concentrations to be massive, and rates to be tiny.

**01/12/2016** Rates going negative was caused by a typo. I am concerned that my numbers are not all in the same dimensions. Try to per litre, and uM dimensions again without the typos, and if they don't work, get in touch with Cian about this issue. Apart from that, including the endogeneous buffer doesn't make much difference.

**02/12/2016** Using per litre rates causes the time steps used in solving the ODE to become very small. This means that ccode gives up on solving altogether. I still think my dimensions are incorrect though. I have emailed Cian about this matter. I should read the Sejnowski paper in the meantime. Also, Thom Griffith modeled fixed buffer, I may need to model free endogenous buffer.

**05/12/2016** My dimensions are definitely incorrect. I need to find a way to convert from molar concentration to molecules reliably. Then use that to convert all rates to per molecule rates. They hopefully the numbers will be runnable.

**06/12/2016** Converted all parameters to molecules, or per molecule per second. After some messing, the simulation still runs, and looks reasonable. Need to read the Sejnowski paper, and read other papers, and prepare for whatever it is that I'm suppose to write for early next year. I've emailed Cian, but he is very busy right now with his new daughter. May be a few days before he gets back to me.

**07/12/2016** Read some of Sejnowski, discovered the forward and backward rates for OGB1. After some struggling, implemented the new numbers.

**08/12/2016** Tomorrow I need to continue to read Sejnowski. The actions of the various substances within the cell should be explained and this should be useful.

**09/12/2016** Read some of the Sejnowski paper. Contains some useful information about: ATPases, smooth endoplasmic reticula, ion exchangers, mobile, immobile, and fluorescent buffers, voltage dependent calcium channel types, etc., etc.

**12/12/2016** Read all the best bits of the Sejnowski paper on simulating  $\text{Ca}^{2+}$  transients in post-synaptic dendrites. Has information on immobile  $\text{Ca}^{2+}$  buffer molecules. Will probably need to introduce this, but should ask Mike about how these buffers work.

Will read the Ganmor paper on reliable interaction model for neuronal ensemble responses.

**13/12/2016** Read some Ganmor paper sounds interesting. Will read more tomorrow. Need to email Mike.

**14/12/2016** Read the rest of the Ganmor (reliable interactions model) paper. It sounded like quite a good model that could be useful for some situations. Emailed Mike about immobile  $\text{Ca}^{2+}$  buffer. Need to add the immobile  $\text{Ca}^{2+}$  to my own model.

**15/12/2016** Added code for the immobile buffer. Perhaps may not be correct. Need to consider the requirement that calcium buffered by the immobile buffer cannot be pumped out of the cell. Also, remember that Mike said that calcium pumps act more like immobile buffers. The buffer the molecules first, and then pump them out at a slower rate.

**16/12/2016** Started reading the generalised linear model for population response analysis. Need to email Cian on Monday to see if we have a meeting.

**19/12/2016** Finished reading about the generalised linear model (Pillow et al 2008). Easy to understand and implement, but overall not the most useful model that exists. Started reading about the population coupling model (Okun et al 2015).

**20/12/2016** Continue reading about the population coupling model. Meeting with Cian, must ask about the Calcium deconvolution project. What should the next step be? Took a lot from Cian's meeting. A choice of things to do. Printed out some papers to read.

**09/01/2017** Didn't do anything at all over the break. Invigilation this morning. Today I will find out what is required for the BNA 'Festival of Neuroscience' poster competition, and email Cian with any questions on the matter. Other things to do include altering the order of variable in the Calcium Deconvolution julia script, and looking back on what Cian and I discussed before Christmas.

Tomorrow I need to look into adding noise to the calcium deconvolution script, that can be varied in amplitude. And respond to whatever Cian thinks about the email I sent to him today.

**10/01/2017** Funding form. Couldn't finish the funding form, need input from Gina Stuart. Wrote a first draft of the abstract and sent it to Cian.

Tomorrow need to check that the order of parameters doesn't matter, and need to add noise in some way.

**11/01/2017** Wrote an updated draft of the poster abstract. Wrote a draft of the abstract for my overall project, requested by the university. Sent both to Cian. Tomorrow, need to begin writing code again. Possibly need to register for the conference tomorrow, submit the abstract the next day.

**12/01/2017 - 17/01/2017** Wrote and submitted the poster abstract. Was sick all weekend. Changed the calcium deconvolution script to accept parameters in normal

units (Molar concentration, and seconds). Included the functionality to introduce noise. Introducing noise messes everything up, causes  $[Ca^{2+}]$  to go negative. Had a look at the spike finder data, the GCaMP6s looks pretty rubbish as an indicator. Currently writing a deterministic version of the model, will introduce noise as part of a model of the photon collection process.

**18/01/2017** Continued with developing the deterministic model. Meeting with Cian cancelled again. Simulation appears to be working ok. Need to introduce a model of the photon collection process, this would be the noisy part.

**19/01/2017** Refactored the calcium dynamics code, looks nice now. Need to decide how to model the noise in the collection process. The number of photons released in each time step will be some proportion of the  $[BCa^*]$  concentration. The number of photons collected by the theoretical collector will be some fraction of that. How do I characterise the stochasticity of this process? Poisson process, with the rate equal to some proportion of the  $[BCa^*]$ .

**20/01/2017** Didn't have a meeting with Cian. Adding the stochasticity using a Gaussian distribution looks OK. Need to compare to actual data. Need to do Okinawa. Need to claim payment for invigilation.

**23/01/2017** Is the spike finder data always from somata, or always from dendrites, or both? Made some progress on taking in files to define spike trains.

**24/01/2017** After meeting with Cian, I need to write a draft of the Okinawa application. I also need to analyse the power spectrum of the spike train, noise, and the overall fluorescence trace. From that it should be possible to subtract the spike train and noise power spectra from the overall spectra to get a spectrum for the GCaMP6s component of the overall spectrum. This could be used to tune input parameter. But, the model is not working well right now anyway, so I need to fix it.

**25/01/2017** Have written a draft of some of the Okinawa school application. Need to think of a project to do, and expand the section on personal motivation. Also need to find out how to measure the power spectrum in julia.

**26/01/2017** Finished writing the application to the OCNC. Sent the draft off to Cian.

**27/01/2017** Wrote a `getPowerSpectrum` function for julia. On Monday, must start doing the analysis that Cian suggested, preferably get something calculated before our meeting in the evening.

**30/01/2017** Meeting with Cian, rewrote some of the OCNC application. Need to again rewrite the OCNC application. Need to pay attention to the claims I make about the RI model. In particular, is it maximum entropy, and will my plan actually create a normalisable probability distribution?

**31/01/2017** Updated the personal motivations part of the application. The Ganmor paper is not maximum entropy after all. I need to think of a new project to do in the OCNC.

- 01/02/2017** Ganmor paper may or may not be maximum entropy, it's not clear.
- 02/02/2017** Emailed Elad Schneidman about the Reliable interaction model. It's NOT a maximum entropy model, but could form the basis of one. Also read Okun's paper on population coupling. Population coupling could be a very useful property in any model.
- 03/02/2017** Sent edited OCNC application to Cian. Cian returned with comments. Made some progress on the power spectrum analysis.
- 04/02/2017** Edited the OCNC application again, and submitted.
- 06/02/2017** Created some power spectrum images. Made a start on the presentation.
- 07/02/2017** Continuing with the presentation. Made some progress at least.
- 08/02/2017** Presentation. Still.
- 09/02/2017** Actually did the presentation. Went OK. Tomorrow, I need to find out what a spectrogram is, I also need to read the population model equivalent over the week-end.
- 10/02/2017** Need to read Gardella over the weekend.
- 13/02/2017** Read Gardella over the weekend and today. Fairly clear about the contents of the paper itself, but need to look at whether or not this model and Cian's model are equivalent before the meeting on Thursday.
- 14/02/2017** Summarised Gardella's paper, compared to Cian's paper. They said that there is a one-to-one correspondence between their parameters  $h_{iK}$  and Cian's parameters, which is correct. There are differences in the training however.
- 15/02/2017** Went to the Enterprising Research Talks. One was very good, another was good, two were ok, one was very bad.
- 16/02/2017** Made some spectrograms, they look quite good now, although I'm skeptical about how useful they are tbh. The meeting with Nick was very good. Cian mentioned that lower dimensional models of the brain makes sense because the brain itself is unlikely to see all possible patterns a number of times and therefore is not using all possible correlations. Improving one of the models by introducing a temporal variable sounds promising to me, because the data is highly temporally correlated anyway, but the efforts of making a maximum entropy model that includes temporal correlations ran into the curse of dimensionality. The following things were mentioned: restricted Boltzmann machines, semi-restricted Boltzmann machines, HMM. All that.
- 17/02/2017** Sent the spectrograms to Cian, he thought they weren't very informative. I agree.

- 20/02/2017** Printed out some docs in order to get reimbursed for the BNA registration. Had meeting with Cian. Need to concentrate on writing the report for the review. Needs to be done before next Monday. Need a draft to be finished by Thursday night ideally.
- 21/02/2017 - 26/02/2017** Doing the six month review report.
- 27/02/2017** Read Köster et al (2014), it is the *Restricted Boltzmann Machine* paper. Nothing special in there. The hidden units introduce loads of parameters potentially.
- 28/02/2017** Finished reading Köster et al (2014), had a meeting with Nick Whiteley about a paper that looks pretty good. It describes a model that could be used to model population tracking and temporal correlations. Need to read the paper and discuss it with Cian.
- 01/03/2017** Started reading the Shimazaki paper. Made some progress, need an example of the EM algorithm to figure out how it works in practise. Continue reading tomorrow.
- 02/03/2017** Continued reading the Shimazaki paper. Need to finish before the next meeting with Cian. Seems like a very good paper at the moment.
- 03/03/2017** Still reading the paper, it's a long one.
- 04/03/2017 - 12/03/2017** Finished reading most of the paper. Discussed it with Cian. He hadn't read most of it. Fixed the calcium fluorescence model, and fiddled around with the parameters until it looked correct. At the moment, it looks quite good for one trace at least. Need to check more traces really. Some of the parameter settings may not be correct, and the simulation may look nice by coincidence.
- 13/03/2017** Emailed Shimazaki, got code from him. Had long meeting with Cian about finishing the calcium fluorescence project. I have some kind of road map towards the end now. Need to follow it and get as much done as possible before the poster. Tomorrow, I need to find a smoothing algorithm, and implement it in my simulation.
- 14/03/2017** Implemented the smoothing function. Re-factored the Spike-finder and Calcium Deconvolution projects.
- 15/03/2017** Added some of the spike-finder functions to the calcium project. Used them to analyse the power spectrum of the modelled fluorescence.
- 16/03/2017** Applied a zscore to the real and modelled fluorescence traces before calculating and comparing the power. More comparable. Emailed Philipp Berens about the drop off  $> 30\text{Hz}$ . Need to learn how to implement and optimize an objective function in `julia`.
- 17/03/2017** Looked into the JuMP modelling package for Julia. Seems to be just what I'm looking for.
- 20/03/2017** Had a meeting with Cian. Need to get JuMP working for modelling the Calcium Deconvolution project. Also need to do the Cajal summer school application.

- 21/03/2017** Modelling with JuMP is not simple. It seems like all the model variables need to be in the one scope, which is not helpful at all. Look further into it tomorrow, and start on the Cajal application.
- 22/03/2017** Struggled with JuMP all day, didn't make an progress. Don't think it will work.
- 23/03/2017** JuMP won't work, but PyCall combined with Scipy.optimize might work. Probably should have used python from the start tbh.
- 24/03/2017 - 26/03/2017** Trying to optimize the parameters of the model.
- 27/03/2017** Had a meeting with Cian. Need to smoothen the power spectra, start making the poster. Make a proper to do list. Also submitted application for Cajal Lisbon summer school today.
- 28/03/2017** Found a way to train two of the parameters. Keep getting a CNode failure when trying to train any of the other parameters. Don't know what to do about that.
- 29/03/2017 - 02/04/2017** Working on the BNA poster. Some strange results from varying the fluorescent indicator concentration. Basically, varying the concentration doesn't do anything. Also started making the poster itself. Using beamerposter, it's quite good.
- 03/04/2017** Made some progress with the perturbation experiments. Started to implement the deconvolution algorithm. Gives out probabilities, need to decide threshold somehow.
- 04/04/2017 - 24/04/2017** Used a kNN classifier to convert probabilities into a spike train. It worked reasonably well. Made the poster. Presented it. The conference was good. There were good talks. Spent last week on the Jentropy package, hoping to add documentation and release today. Meeting with Cian tomorrow.
- 25/04/2017 - 01/05/2017** Changed the Calcium deconvolution code, now using Molar instead of molecules. Still trying to optimise the free parameters. Will then test the effect of varying the concentration. Also got the state-space analysis model running in MATLAB
- 02/05/2017** Got the state-space analysis model up and running well. Also refreshed the model theory. Will try to get the model running in Python, then find some other data and try to get it working on that.
- 03/05/2017** Got the state-space log linear model up and running in Python. Tomorrow will run through it line by line to see what is being calculated and understand how everything is working.
- 04/05/2017 - 09/05/2017** Ran through Shimazaki line by line. Learned a lot about how the whole thing worked, but still not really clear. Got the  $l_0$  spike inference model running. Wrote some code to run the model.

- 10/05/2017** Wrote a script to run the  $l_0$  spike inference model on the spike-finder data. It's killing my laptop. Waiting for IT to install R on my desktop.
- 11/05/2017** Got the  $l_0$  model to run on my laptop using a smaller file. The  $\lambda$  parameter makes a big difference in the number of spikes predicted. For dataset number 8  $\lambda = 10$  works well. This may need to be learned for each dataset though, which is a limiting factor for applications of the spike inference model.
- 12/05/2017** Made some good progress reading Shimazaki's model again. I'm getting to grips with the optimization process quite well, although some parts are still a bit fuzzy. Need to explain the whole thing to Cian on Monday. Tomorrow I need to look up the *fixed-interval smoothing algorithm* for a Gaussian state and observation equation.
- 15/05/2017** Had a meeting with Cian. Need to separate the release and capture rate in to noise and capture rate. Use binomial distribution for the distribution of collected/uncollected photons. Profile the Shimazaki code. There's some problem with the baseline dynamics, need to investigate and prioritise.
- 16/05/2017 - 17/05/2017** Model is really really not working now. I might consider going back to molecules instead of molar.
- 17/05/2017** Went back to molecules. Without the bug that I have fixed, the model runs much much better, and it optimises well.
- 18/05/2017 - 25/05/2017** Edited the photon capture modelling process to use a binomial distribution. The variance is much lower than it needs to be, but the whole process makes more sense. Created simulated traces for the 8<sup>th</sup> dataset.
- 26/05/2017** Need to run the  $l_0$  spike inference on the simulated traces, and compare the results to the inference from the actual traces (which was really bad). For the future I need to find out why the Paninski algorithm works so badly on simulated traces. I also have to prepare for the journal club next week.
- 27/05/2017** Currently running the  $l_0$  inference on the simulated traces. The bad precision of the other inference algorithm may have been down to comparing different traces with one another. Need to look back at how I built the files involved.
- 29/05/2017** Compared the performance of the  $l_0$  spike inference algorithm when applied to spike-finder data and modelled fluorescence traces. Again, the algorithm performs better on actual data than on modelled data. But, this algorithm performs much worse than the Paninski algorithm.
- 30/05/2017** Had a meeting with Cian. To do:
- Create figures showing the fluorescence trace with spikes and predicted spikes, and the simulated fluorescence trace with spikes and predicted spikes. Do this for both the Paninski algorithm and the  $l_0$  algorithm for dataset number 8



- The monkey data is sampled at a frequency of 1Hz. Therefore the time bins are 1ms. This makes observing correlations unlikely. So process the data so that it uses larger time bins. Do for 5, 10, and 20ms bins and see if there's any difference.

**31/05/2017** Did the Journal club. Only three people showed, other than myself and Cian. Also ran the sub-state log linear model using different time bins. Using a bigger time bin gives more correlations, which is nice.

**01/06/2017 - 02/06/2017** Came in very late. TO DO:

- Organise the trip to Cardiff. Early Career neuroscientist.
- Calcium Deconvolution: Check the conversion functions still make sense.
- Email Nick asking to meet up and discuss the model and how it could be improved. Have a suggestion to hand.
- Calcium Deconvolution: Make the figures comparing spike inference methods. Refactor code if necessary.

**05/06/2017** At the Early Career Neuroscientist Day in Cardiff.

**06/06/2017** Checked the conversion functions. There was a scaling issue, and a question about the definition of a Molar was raised. Eliminating the scaling issue caused the variance to be increased relative to the mean. That's a good thing.

**07/06/2017** Had a meeting with Nick, we discussed what our aims are and where we should look into the implementation of the model.

**08/06/2017** Got a script running for simulating a fluorescence trace for each spike train in a given spike finder data set. Investigated the implementation of the Newton Raphson method in the State Space Log Linear model. It looks like they calculate  $\eta_t$  using  $\theta_t$  every time. That means that they need to calculate products of matrices, the size of which scale with the number of neurons, and the order or correlations. This is what disables the scaling of the model.

**09/06/2017** Had a meeting with Nick. He is looking into some improvements to the model. Wrote a summary email. Profiled the State Space log linear model and sent a summary to Nick and Cian.

**12/06/2017** Came in very late. TODO:

- Email Nick to organise meeting.
- Write script to run inference algos on fluorescence traces.
- Make the figures comparing inferred spikes with actual.

**13/06/2017** Had the meeting with Nick. Shimazaki has already been on a paper that solves the problem. Need to run the two codes side by side and have a look at the differences.

- 14/06/2017** Made some progress implementing the approximate state space log linear model. Found some apparent bugs.
- 15/06/2017** Went to the 'Data Intensive research workshop' in the morning. In the afternoon, found out that the code was only designed for pairwise models. Using higher orders causes the code to break down.
- 16/06/2017** Ran the exact and approximate models side by side, using  $N = 2$ , and  $O = 2$ . Definitely found differences between the two. The exact model contains more information entropy and the time series of the entropy seems to reflect the make-up of the stimulus.
- 17/06/2017** Prepared all the necessary files for comparing traces and spike inference.
- 18/06/2017** Made all the figures.
- 19/06/2017** Sent the figures off to Cian. Had a meeting with Cian. Cian will try to catch up on the work that Nick and I did while he was on holidays. I need to remake the figures under a few different scenarios. First of all, I need to fix the scaling issue with converting from molar concentration to molecules.
- 20/06/2017** Prepared the files for more comparisons, with the scaling issue fixed. Also started writing the Hourly paid teaching application.
- 21/06/2017** Sent some more comparisons to Cian. Started doing yet more comparisons. Also applied for hourly paid teaching.
- 22/06/2017 - 30/06/2017** Applied for hourly paid teaching, applied for Janelia. Scaled spikefinder data and sent the results to Paninski. The changes in performance are at least partly down to the KNN algorithm.
- 03/07/2017** Had a meeting with Cian. Need to sort out what the story is with the effect of scaling on performance of the Paninski algorithm. Cian suggests calculating the false positive rate, and false negative rate at each increment of an incrementally increasing threshold of spike or no spike. Needs to be performed in matlab. Could be annoying.
- 04/07/2017** Prepared the figures for Cian showing the performance of the algorithm with different thresholds. See what he has to say about that.
- 05/07/2017** Prepared another figure, aggregating over all the false positive/false negative rates, showing the mean and standard error. I think I have forgotten what the point of this is.
- 06/07/2017** Used a threshold to predict spikes instead of knn. It performs much worse, but it's more realistic on the other hand. Anyway I suggested to Cian that this madness must end.
- 07/07/2017** Need to have a read of the original paninski paper to find out what the last output field is actually supposed to be. I'll see if I can think of some reasonable way to convert this vector into a spike train after that.

**09/07/2017** Emailed Pnevmatikakis about the Paninski algorithm. I want to know what the 'sp' return variable is actually supposed to be. Hopefully he will respond.

**10/07/2017** Printed out a paper by Pnevmatikakis. Hasn't replied so far.

**17/07/2017** Arrived back from holiday. Actually need to read the paper I mentioned in my last entry.

**18/07/2017 - 26/07/2017** Got a reply from Pnevmatikakis. He said that the algorithm wasn't really designed for inferring spike trains. So I gave up on that algorithm. Since then I have look implementing other algorithms and seeing how the optimisation is going. Tomorrow I need to read a bit of a paper about a new algorithm, and play around with the examples for that algorithm. Hopefully, I can actually implement this one properly.

**27/07/2017** Make a program that optimises the five free parameters of the model before running the model. Run that on all of the traces in dataset eight. Make figures showing the power spectrum for each trace. See what Cian thinks of that.

**28/07/2017** Finish making the program described above. Run it. Send figures to Cian. Practise with the OASIS spike inference algorithm.

**31/07/2017** Had a meeting with Cian. It was good. Almost finished with the optimisation. Need to put the optimisation figures on the Google drive. Need to make a figure of the rms path as the number of iterations increases. Need to make notes on what I have done in this project. Hopefully a paper can be written on it. Need to email the Deneux guy.

**01/08/2017** Tomorrow I need to continue with learning about the OASIS algorithm. Hopefully I can implement it soon. Its seems somewhat simple. I should also redo the optimisation with slightly different starting points.

**02/08/2017 - 03/08/2017** I had the flu for these days.

**04/08/2017** Tomorrow, I have to attempt to get the OASIS algorithm running on the spike-finder data. The examples in the github repo are not very helpful. The quality of the coding is poor in the examples, but good in the actual code.

**05/08/2017** Had another look at the OASIS algorithm. I should apply it the same way as it is applied in example number 5. Vary the minimum amount of non-zero activity within each time-bin until we get a reasonable number of spikes. In order to infer the actual spikes, do the same thresholding as in example number 5.

**06/08/2017** Wrote a good script for the OASIS algorithm. Made figures and saved traces and spike trains. Need to move the figures into the google drive. Tell Cian they're there. Tell Cian to have a look at them, and the other figures before our meeting tomorrow. Also need to do some classification analysis on the results of this algorithm.

- 07/08/2017** Made the classification analysis for the OASIS algorithm. Need to implement some kind of window in the classification analysis. Need to finish of the work on applying a positive window to the classification analysis.
- 08/08/2017** Finished implementing the window. The OASIS algorithm performs better on the modelled data than on the actual data now. Apply the window to the Lzero data next.
- 09/08/2017** The Lzero data was also improved by using the window. The algorithm performs better on the modelled traces than on the actual fluorescence traces, when evaluated using the window. Need to apply the methodology that I applied to the OASIS algorithm to the paninski algorithm and see how that goes. This will require writing some matlab.
- 10/08/2017** The Paninski algorithm now behaves as expected after performing analysis using an 'optimised threshold' and the two frame window. Tomorrow, I should investigate to verify that everything is actually working the way I think it is (too good to be true). Next week I will have to move on to perturbation, or something. Will also be meeting Nick, and maybe Mike as well.
- 11/08/2017** I investigated the results. I'm happy enough that they are genuine. This is good. I will now write a new project for doing the perturbation analysis.
- 14/08/2017** Made some progress with the perturbation analysis code. Doing it in a slightly different way. Should work well, I hope.
- 15/08/2017** The only problem with the code so far is saving down the results. I'm trying to save them all down in the one hdf5 file, but it's not cooperating. Saving as a mat file may be a better option.
- 16/08/2017** Got the hdf5 save down working well. The problem was with over writing the data that I had already written. Just need to write the python to make the figures now.
- 17/08/2017 - 22/08/2017** Had a meeting with Cian. Had a look at the perturbation figures. They're quite good. Need to make more figures for perturbing different parameters. Currently working on taking the perturbed modelled fluorescence traces, inferring spike trains from them, and evaluating the quality of the inference.
- 23/08/2017 - 27/08/2017** Managed to evaluate the difference in spike inference quality with perturbed fluorescence indicator concentration. Need to make the labels a bit clearer. But also need to get the lzero algorithm working better.
- 28/08/2017 - 30/08/2017** Got the Lzero algorithm working well. It goes slowly, but it goes well. Still need clearer labels, although I can control that really. Need to figure out a way of perturbing the optimised parameters. Should graph the values of the optimised params.
- 31/08/2017 - 01/09/2017** Re-registered. Wrote a lot of code for making boxplots of the fluorescence model parameters' optimised values.

**02/09/2017 - 10/09/2017** On holiday. When I come back, I need to look at the discrepancy between lzero performance. The perturbed values are worse than the values from the model, even though they should be similar.

**11/09/2017 - 12/09/2017** Made the figure comparing the optimised parameters pairwise. Not sure what to make of it. Need to investigate the lzero performance tomorrow.

**13/09/2017 - 15/09/2017** Investigating the Lzero performance currently. It still doesn't behave like the other algorithms. It's difficult to tell why. Need to make more graphs, take more measurements.

**16/09/2017 - 19/09/2017** Had a meeting with Cian. He's ok with Lzero performing differently to the other algorithms. It's just not a very good algorithm. Need to do the following:

- Make a graph showing the difference in number of spikes predicted, actual vs predicted, with different colours for different perturbation values.
- Calculate the spike triggered average concentrations of each of the buffers in the model
- Measure the time constants of all the concentrations.
- Change the optimised parameter pairwise comparison graph, so that all of the axes start at zero.

**20/09/2017** Changed the optimised parameter graphs. Made a good start on the number of spikes difference graphs.

**21/09/2017 - 22/09/2017** Made the spike count comparison graphs. Also managed to delete all of my h5 files. I will have to remake those overnight.

**23/09/2017 - 27/09/2017** Remade all the h5 files quite quickly. Also created some figures using PyPlot and Seaborn in julia. Installed julia v0.6. Measured all the spike induced averages of the traces. Now need to measure the time constants.

**28/09/2017 - 30/09/2017** Wrote a script for measuring spike induced changes, and time constants, each of the buffer traces has two time constants, one fast and one slow. The excited indicator bound calcium has an inverse logarithmic increase followed by an exponential decay. Need to ask Cian if the spike induced change measurement is correct, and what he wants from the time constants. I also need to start writing up this work as a chapter for my thesis.

**02/10/2017** Did not have meeting with Cian. Need to convert time constants and spike induced changes into more sensible units. Need to start writing. Writing this the top priority for tomorrow.

**03/10/2017 - 15/10/2017** Started writing the thesis chapter on the fluorescence model. It's going well so far. Also wrote a function for converting from molecules to molar concentration. Need to get started on running the perturbation analysis with optimisation.

**16/10/2017** Edited the Perturbation analysis process to run the analysis with optimisation. The optimisation is the slow part, and it is running. The analysis is not running, there is some bug. But that can be fixed in the future. Need to keep going with writing the chapter. Start with including the smoothed power spectrum. Then move on to the spike inference methods.

**17/10/2017 - 20/10/2017** Continued to write up the thesis chapter. Currently running the optimised perturbation analysis. Need to start writing about the perturbation analysis next week. Good talk today from a man from Google Deepmind. Should have spoken further.

**23/10/2017** Need to write up the methods and analysis on the perturbation analysis. The start on the presentation. The perturbation analysis with optimisation mostly worked.

**24/10/2017 - 03/11/2017** Prepared and presented the presentation. Had a useful meeting with both Mike and Cian, with follow up work to do. Started work on further perturbation, but the results are strange. Need to investigate the calcium dynamics/fluorescence traces produced by the perturbation process. The most likely explanation is a coding error on my part.

**04/11/2017 - 07/11/2017** Found a bug in the perturbation with optimisation bash script. Fixed the bug, now re-running all of the analysis. Need to see if I can write some of the discussion. Should revise the introduction part also. The analysis will take an age to run.

**08/11/2017** Continued with redoing the analysis. It takes a while. Re-wrote some of the introduction to the thesis chapter. Meeting with Cian tomorrow.

**09/11/2017 - 16/11/2017** Meeting went very badly. Cian suggested changing the model to remove some of the parameters. Got started on that. Presented a paper to the O'Donnell group. Wrote a draft application for COSYNE. Tomorrow I need to create some traces with the new model using a very large value for the calcium rate. See if that makes any difference.

**17/11/2017 - 20/11/2017** Making the calcium rate very large did make some difference. Need to change the objective function, and make more traces. Wrote a submission for COSYNE, need to submit that this evening.

**21/11/2017 - 11/12/2017** Managed to the new model working well. I have changed the objective function to include the difference between the two traces. I have re-made almost all analysis. Still running the Lzero spike inference on the perturbation. Need to make new dynamics comparisons and trace comparisons for the perturbation data. Need to have a go at making an RNN to infer spikes.

**12/12/2017 - 13/12/2017** Many other people have made combinations of RNNs and CNNs to solve the spike inference problem. All of these solutions require tensorflow and a GPU however. After much messing around I have managed to get the 'MLspike' spike inference algorithm working. Although it doesn't perform very well. It's better than Lzero.

- 14/12/2017 - 17/12/2017** Managed to get MLSpikes working well enough to include in the results. Had to update my version of Julia due to Sundials (I think). Took all day on the 15th but got it working. Need to analyse the perturbation without optimisation results. On the spike inference performance figures make sure the Y-axis is between 0 and 1. Read about the false discovery rate and Tukey's test. These may be good ways to compare the inference performance at perturbed values.
- 18/12/2017** Managed to do the perturbation without optimisation. Results are not that different from the optimised version. Also changed the spike inference y-axis range.
- 18/12/2017 - 16/01/2018** Need to put the perturbation spike inference quality section in the document. Also need to Add the table of sources and values for fixed parameters, seriously. Need to register for COSYNE.
- 17/01/2018 - 22/01/2018** Registered for COSYNE, booked flights and most of the accommodation. Changed the perturbed fluorescence figure to see all of the traces individually. Problem with the lower perturbed values. Still need to add table of sources and values for parameters.
- 23/01/2018** Meeting with Cian. Need to measure the signal to noise ratio for each perturbed value. Must measure the change induced by a spike averaged over 100 samples. Must also make the fluorescence model into a single repository. Need to move unit conversion from modelling to model.
- 24/01/2018 - 26/01/2018** Wrote a script for measuring SNR. Put it in the chapter. Made the fluorescence model into a single repository. Moved unit conversion as well. Need to put in a table of sources and whatnot. Also need to perturb another parameter.
- 26/01/2018 - 29/01/2018** Started perturbing the immobile buffer. Still running. Need to edit the perturbation code to accept two parameters to be perturbed at the same time.
- 30/01/2018** Finished off putting in the table of fixed parameters. Chapter is almost finished now. Still need to edit the fluorescence\_training script to accept two parameters.
- 31/01/2018** Need to edit the fluorescence\_training script still. But the priority is the Okinawa application.
- 01/02/2018** Have an idea of the Okinawa application. Preparing for the workshop tomorrow at the moment.
- 02/02/2018** Sent a draft of the application to Cian. Waiting for a reply now. Need to get back to the parameter perturbation over the weekend. Still need to rewrite fluorescence\_training.jl.
- 05/02/2018 - 09/02/2018** Did the analysis on the immobile buffer concentration perturbation. Nice result. Also edited the training/perturbation process to accept more than one parameter at a time, currently running. Need to ask Cian about an alternative to reporting p-values.

- 12/02/2018** Did the analysis on the binding/unbinding rates, results were fine. Need to start editing the poster.
- 13/02/2018** Edited the poster. Sent to Cian. Some major changes to the layout may be necessary in order to fit everything in there.
- 14/02/2018** Changing the poster quite a bit now. Five columns, much wider. Should be able to fit more things. Try not to put in too much text, mostly results.
- 15/02/2018** Edited the title of the poster. Mostly did HPT work.
- 16/02/2018** Made some small progress with the poster. Subcaption environment breaks the compilation.
- 17/02/2018** Did a little bit on the cosyne poster.
- 18/02/2018 - 21/03/2018** Made the COSYNE poster and went to COSYNE. The conference was very good. The talks were interesting on the whole. Some were very particularly good. The poster session went well. People were interested in the model and how it could be useful for them. Some people took my email address etc. Since I got back, I added more to the calcium fluorescence thesis chapter. I also had a review meeting with Conor that went quite well. I had a meeting with both Cian and Mike about the thesis chapter and turning it into a paper. They both think that what I have written is fine as a thesis chapter. But it needs a little more to work as a paper. In particular something that makes it more interesting to the field. For example, comparing two different types of cells, or comparing two different types of indicator.
- 22/03/2018** Need to do some HPT. Also need to think about what to add to the chapter to make it into a paper. Also need to see what I can download from the Allen Institute's brain data. Looking for V4 data in the mouse brain.
- 23/03/2018** Did the HPT. Tried to install the allensdk, but can't manage it. Asked IT to do it instead. Need to have a look at the references sent by Mike.
- 26/03/2018** Read some of the references sent over by Mike. Emailed Mike and Cian. Going to make a new project for simulating spike trains.
- 27/03/2018 - 17/04/2018** Simulated spike trains with different mean firing rates. Work on finishing then fluorescence modelling paper is ongoing with Mike and Cian. Had a meeting with Nick and Cian. Re-read the Kolazyck paper on the statistical modelling for hierarchical systems. Sent my thoughts on the paper to Cian and Nick. Started looking for a placement, attended a meeting, have more meetings in the upcoming days.
- 18/04/2018** Start on the paper as outlined by Cian. Start with figures. Attend meeting in Adarga.
- 19/04/2018 - 26/04/2018** Attending meetings at Adarga and CheckRisk. Will be applying with both of these companies. Need to email CheckRisk soon, and Adarga



also. Made the figures for Cian. He suggested changes. Start on those tomorrow evening. Also received data from Mike/Anna Simpson, started work on modelling fluorescence traces for that.

**27/04/2018 - 16/05/2018** Spent a lot of time on the applications for CheckRisk and Adarga. Now submitted. Remade the figures for Cian, they look more like what would be seen in a paper now. Need to ask Cian about what the next step is on the paper. Did the analysis for Mike/Anna Simpson. Only acknowledging one spike per bin may be a problem. Currently working on loading in Neuropixels data, and applying the hierarchical model to this data.

**17/05/2018 - 06/06/2018** Re-did the analysis for Mike, worked out quite well. Mike was happy with it anyway. I redid the figures for the paper. Cian was happy with most of them, but a few still need to be redone. The application with CheckRisk was accepted, I will start on the 25th of June. Must tell Cian about that. Wrote a Julia script to do most of the same analysis as is in the MATLAB script that goes with the Neuropixels data.

**07/06/2018 - 08/06/2018** Wrote a Julia script to count the spikes for each cell, and created a csv holding all the necessary data. Need to fit the hierarchical model to the data. Sent the figures to Mike. He said he would discuss the paper with Cian. Also need to update the figures.

**09/06/2018 - 14/06/2018** Used statistical tests to see if the cells are more Poisson or more Gaussian. More of them are Gaussian, but most of them are neither. Need to write some functions to do the parameter transformation. Also still need to edit the paper figures. But Cian and Mike have not looked at them yet. Starting placement the week after next.

**15/06/2018 - 22/06/2018** Wrote the functions for the parameter transformations. Also wrote some functions to get the model distributions. There may be a problem with the model. Need to re-read the paper and make sure I understand. Placement is supposed to start soon.

**23/06/2018 - 08/11/2018** Attempted to derive all the results given in the Kolaczyk paper. Ran into trouble with the posterior estimation of  $\omega_{j,k}$ . Went on three month placement, learned about time series analysis, forecasting, and recurrent neural networks. Suggested time series related topic for chapter 3 of thesis. Caught up again with the Kolaczyk paper. Added to the notes on the Kolaczyk paper. Emailed Cian to ask him about my problem. Learned about empirical Bayes.

**09/11/2018 - 23/11/2018** Implemented some toy examples of Bayesian parameter estimation. Learned quite a bit about it, definitely have a better understanding now. Also did a lot of teaching assistance for the Machine Learning course. Need to infer the parameters of the synthetic hierarchical data.

**24/11/2018 - 28/11/2018** Implemented the Gaussian version of the hierarchical model to synthetic data. Seems to work ok. Told Cian and Nick. Need to do more on the Machine Learning course before next week.

- 29/11/2019 - 22/02/2019** Implemented Poisson version of the multiscale model, fixed bugs in the Gaussian version. Sampled from Gaussian version. Did progress meeting with Conor. Started on regional correlations project. Wrote code to get the cell info and save it down. Wrote code to select random neurons and calculate the regional correlations between them.
- 23/02/2019** Started working on calculating the marginal likelihood for the multiscale model. Need to calculate the likelihood, then find a way of comparing the likelihoods for two different tree structures. One correct, and one incorrect.
- 24/02/2019** Added code to the multiscale project to calculate the marginal likelihood of a model/tree. Added code to allow different trees to be tried out as models. Must compare the marginal likelihoods of the true tree and a false tree. Must email Christian Konrad about the drop-in session on Friday.
- 25/02/2019** Added a script for plotting marginal likelihoods over number of samples. Results not as expected. Emailed Christian. Need to read correlations review. Need to read about Barcelona conference. Need to read about Entrepreneur competition.
- 26/02/2019 - 01/03/2019** Read most of the correlations review. Some new measurements to take, and filters to write. Did applied stats lab. Met with Cian. Need to write the abstract for Barcelona. Also need to do the Entrepreneurship competition. No harm in entering either way.
- 02/03/2019 - 05/03/2019** Started on the abstract for Barcelona. Added a little bit to the correlations matrix plot. Wrote a blog post for Aby. Need to do more on the correlations. Need to experiments with different trees and correlation regimes in Multi-scale project.
- 06/03/2019 - 11/03/2019** Had a meeting with Cian. Reviewed the OCNS abstract. Read about ‘incremental mutual information’. Got permission from Steinmetz to use the dataset. Entered Entrepreneurship competition. Read some algorithms. Submitted abstract for OCNS. Must catch up on both algorithms and Applied statistics. Must implement time bin size in regional correlations.
- 12/03/2019 - 13/03/2019** Caught up with algorithms, did some teaching. Tried to write a script to replicate results from the IMI paper. Having difficulty. Still need to implement time bins.
- 14/03/2019 - 15/03/2019** Did even more teaching. Managed to replicate the correlation results. Still need to implement time bins.
- 16/03/2019** Started on calculation of IMIs. Don’t forget time bins.
- 17/03/2019 - 19/03/2019** Calculated some IMIs. Also added code to calculate correlation strength vs bin width. Need to measure lots of results now.
- 20/03/2019** Measured lots of correlation coefficients for different bin widths. Looks like I can get a good result. But I need to refactor the code first.

- 21/03/2019** Refactored the code. Wrote a script to get 30 pairs of strongly responding pairs from each region, and calculate their correlation coefficients across different time bins. Also different stimuli. Need to run it tomorrow, and hope it works.
- 22/03/2019 - 23/03/2019** Ran the code, created a big csv. Used the csv to create figures. Wrote a document outlining what I did. Will present to Cian on Monday, hopefully.
- 24/03/2019 - 26/03/2019** Had a meeting with Cian. The correlations vs bin width figures and results are good. Trying to make some cross-correlograms. Not working so far.
- 27/03/2019** Attended CS Dept Seminar given by Cian. Still trying to make cross-correlograms.
- 28/03/2019 - 30/03/2019** Made the cross corellograms, but they don't look as expected. I'm not worried. Wrote a script to calculate the signal correlation. Started plotting the signal correlation.
- 31/03/2019** Made the signal correlation plots. Need to start on the histograms, and bifurcated spike count correlations vs time bins.
- 01/04/2019** Made the bifurcated spike count correlation plots. Need to add to the pdf and start on the histograms.
- 02/04/2019** Added to the pdf. Made strong pairs correlation histograms. Need to make all pair versions.
- 03/04/2019 - 11/04/2019** Did some teaching. Started measuring the mutual information. Added all pairs histograms and strong histograms for pairwise correlations and mutual information. Got accepted to the CNS conference. Had a good meeting with Cian. Need to get on with tasks arising from meeting.
- 12/04/2019** Added firing rates measurements and images. Started on geomtric means vs pairwise measurements. Probably need to remeasure the mutual information using bias correction, at least to examine the difference.
- 15/04/2019** Added the correlation and mutual information histograms for strongly responding neurons and their corresponding statistical tests. Updated the README.md. Need to get on with measuring the MI with bias correction and learning how to do community detection.
- 16/04/2019** Implemented bias correction techniques in measuring the MI. Need to rebuild the big csvs on both machines and decide on an MI measure to use for figures etc. Also need to read about community detection. Also need to read emails about other stuff.
- 17/04/2019** Rebuilt the big csvs. Started reading about community detection. Meeting with Cian tomorrow.
- 18/04/2019** Had a meeting with Cian. Started measuring the 'symmetric uncertainty'.
- 19/04/2019** Started rebuilding csvs. Made symmetric uncertainty histograms. Made correction comparison figures.

- 20/04/2019 - 22/04/2019** Edited the correlation histograms script. Now choses the best stimulus for the strong responses. Makes strong vs all comparisons valid. Added the firing rate histograms to the pdf. Should also add symmetric uncertainty histograms. Need to create mutual information vs symmetric uncertainty scatter plots. Need to email the lads who sent code to me, and take a quick look at the code as well.
- 23/04/2019 - 26/04/2019** Added to the pdf. Had a meeting with Cian. Started to implement the community detection.
- 27/04/2019 - 30/04/2019** Updated the 'useful\_lines.py' script to have a look at the community detection. There are a number of functions to use from a few different sources. The Brain Connectivity Toolbox should be useful.
- 01/05/2019** Reading about consensus clustering. Writin some code to do it.
- 02/05/2019 - 09/05/2019** Read more about consensus clustering. Wrote some functions that should be useful. Read Humphries paper from this year. Need to implement that paper's methods on my data. Need to also get started with the other smaller project.
- 10/05/2019 - 13/05/2019** Haven't started the smaller project yet. Have a meeting with Cian tomoroow. Will outline Humphries paper for him.
- 14/05/2019 - 16/05/2019** Added to the useful lines script. Implemented full and sparse WCM now. Need to do something on the smaller project really.
- 17/05/2019** Getting the Low D space now works. Need to do the network rejection.
- 18/05/2019 - 21/05/2019** Implemented the network rejection. There are some differences to the matlab results, but they are mostly the same. Need to implement the synthetic networks code.
- 22/05/2019** Added more to network rejection. Implemented kmeans.
- 23/05/2019 - 25/05/2019** Added more to network rejection. Trying to implement consensus convergence.
- 23/05/2019 - 04/06/2019** Continued implementing the Network Noise Rejection project. Also went on holiday. Have to do a lot of marking over the next few days.
- 05/06/2019 - 09/06/2019** Did the computational neuroscience marking. Did more on the Network Noise Rejection. Just need to implement plotting function, then put the project in it's own repo. Then I can get back to finding communities using mutual information.
- 10/06/2019** Added plotClusterMap. Not finished yet.
- 11/06/2019** Finished plotClusterMap. Started moving code into new repo.
- 12/06/2019 - 13/06/2019** Moved code into new repo. Appears to work well. Wrote example scripts for the new repo. Had a meeting with Cian. Need to bring in my laptop next Tuesday.

- 14/06/2019** Tested the example scripts on the StarWars data. There are differences. It should be useful to find the source of these differences.
- 15/06/2019 - 18/06/2019** Implemented Otsu's method to match Humphries's implementation. It was causing a problem with all 0 clusters after exploration. Need to start 'sparsifying' the info and correlation matrices, and applying the clustering there.
- 19/06/2019** Wrote a function for sparsifying. Without sparsifying, the clustering still works. Should sparsify though. Need to save some files to speed things up a bit.
- 20/06/2019** Started making the poster. Need to sparsify to different levels and save some results from each community detection. Region label, and cell\_id label would be good. Need to continue making the poster too.
- 21/06/2019** Did a little bit of work on the poster.
- 22/06/2019** Did some more work on the poster.
- 23/06/2019** Did a bit more work on the poster. Added a raster plotting function to the regional correlations project.
- 24/06/2019** Added more to the poster. Need to rebuild `strong_pairs.csv`. Need to make some custom figures. Might need to make some correlation networks. Emailed Humphries.
- 25/06/2019** Added more to the poster. Sent to Cian for review. May need to change tomorrow. Hoping for more community detection in correlations over night.
- 26/06/2019** Added more to the poster. Need to include the Bristol CNU pdf.
- 27/06/2019 - 05/07/2019** Made the poster. Got it printed. Went to the Neural Computation conference. Presented the poster. Spoke to Mark Humphries, Matt Evans, Matthias Hennig. All went well. Downloaded the new 8 probe data. Started a repo for analysing that. Need to look into SpikeInterface to data. Need to measure shuffled correlations and information when analysing.
- 06/07/2019 - 07/07/2019** Started some work on the Eight Probe project. Added cell info csv.
- 08/07/2019 - 10/07/2019** Looked into Steinmetz's eight probe data. Got Matt Jones's data from Cian, also looked at that. Started analysing Eight Probe data.
- 11/07/2019** Did more work on the eight probe data. Almost in a position to start on the results (I think).
- 12/07/2019 - 18/07/2019** At CNS 2019, in Barcelona. Presented the same poster as in Nottingham. Less interest because there were more posters, and my poster was away in a corner. Some good talks, especially on the last day. Talks about transfer entropy were interesting. Direct information could be a useful measure.
- 19/07/2019 - 22/07/2019** Had some expenses to query about. Started working on the Tennant data. Added a script to show spike count vs position. Emailed Cian.

**23/07/2019** Read about decoding. Looked at Tennant data. Need more data. Added a bit to the Eight Probe project. Need to get back on Matt Jones data as well.

**24/07/2019 - 29/07/2019** Had a meeting with Cian. Would be nice to do some quick analysis on cells with high info, but low correlation. Also started reading about multiprocessing in Python. Have implemented some code. But it doesn't help.

**30/07/2019** Tested multiprocessing on BlueCrystal, it helps. Pyentropy is only compatible with python 2.7, but multiprocessing is only doable in python3. I'm moving on from pyentropy and using pyitlib instead. I won't be calculating the bias corrected info anymore, unless I calculate it myself.

**31/07/2019 - 01/08/2019** Still adding parallelisation. Trying to parallelise across nodes. Meeting with Cian cancelled. Must respond to email.

**02/08/2019 - 09/08/2019** Parallelised across nodes and processes. Still had to use a subset of 50 neurons in order to do the bin width analysis. Some very surprising results comparing mutual information to shuffled mutual information.

**10/08/2019 - 22/08/2019** Spent some time wondering about the mutual information. Read Panzeri et al 1996, mentioned that shuffled info may overestimate the bias if a lot of bins are empty. Need to write a function to measure the Panzeri-Treves bias. Also need to hear from Sarah Tennant, and do something with the Matt Jones data. Also need to extract a dataframe with 5ms bins to analyse spike trains, and distribution of the number of active cells.

**23/08/2019** Wrote function to measure PT-bias. Ran new analysis on BlueCrystal. Need to get files copied across. Started on analysis for spike trains. Need to parallelise some of the code.

**24/08/2019 - 26/08/2019** Did analysis on the number of active cells in each region across bin widths. Some things to think about.

- Changes in the distribution across time bins. There is a big change from 1ms to 5ms. But there is a small change from 3.5s to 4s.
- Correlations between the number of active cells across brain regions. Including time delayed correlations.
- Time delay correlations between the number of active cells in a region, and the individual cells in that region.
- Need to read Okun, and Cian's paper too.

**27/08/2019** Read Cian's draft for the Fluorescence modelling project. Hard to see what the point was. I could have written the same thing years ago. In fact, I wrote a longer version in 2017. Need to familiarise myself with running the code again. It almost certainly won't work properly.

**28/08/2019 - 05/09/2019** Started upgrading to Julia 1.2. Got it running to some level, but doesn't work correctly. Need to look into it more. Did presentation for TRENDS (for Amelia). It was ok. Need to try to do further analysis on the Eight Probe

data, using 1s, 2s, and 3s bin widths. Need to implement network rejection and community detection also.

**06/09/2019** Added a script for making firing rate histograms. Added some firing rate histogram figures also. Need to move this code over to BlueCrystal. Then build huge spike count frame files. Then scp these files to the university machine. Then put those files on the one drive, or a USB.

**07/09/2019 - 12/09/2019** Added even more histograms. BlueCrystal is failing to run the analysis. Need to get that running. Also need to learn how to fit the Conway-Maxwell Binomial distribution. Still need to get the Fluorescence Modelling back working as before.

**13/09/2019 - 16/09/2019** Got the Fluorescence Modelling working as before (I think). Blue Crystal still failing. Need to choose a handful of cells from each region for each mouse, and go with that. Worked on maximum likelihood estimation for the parameters of the Conway-Maxwell Binomial distribution. Flying to France on the 21st.

**17/09/2019** Worked on the maximum likelihood estimation of the parameters of the Conway-Maxwell distribution. Wrote a decent document about it, and have a good idea of what is going on.

**18/09/2019** Did more on maximum likelihood. Looks like standard approach will not work. Also added a function to Eight\_Probe to distribute the cells used over the regions. Trying to analyse the data.

**19/09/2019 - 20/09/2019** Read more about the exponential family, will use the maximum likelihood estimation methods for exponential families to estimate the values of  $p$  and  $\nu$  in the Conway-Maxwell binomial distribution. No progress on Eight\_Probe data so far.

**21/09/2019 - 26/09/2019** Attended NCCD 2019 in Capbreton, France. Learned more about exponential families and estimating the parameters of them.

**27/09/2019 - 28/09/2019** Implemented exponential family style bayesian estimation for binomial distribution. Needs some tweaks, but works in principal. Also want to implement animations in python.

**29/09/2019 - 30/09/2019** Did a better job with the binomial bayesian estimation. Added chunking to Eight\_Probe data collection. Currently having a problem with greenlet. Emailed HPC help.

**01/10/2019** Must check to see if any other packages are using greenlet. Added more to Conway-Maxwell binomial document. Forgot the log in the code.

**02/10/2019** Fixed problem with greenlet by using more smaller chunks. Had a meeting with Cian. Conway-Maxwell plan is approved. Started measurement distribution script for Eight\_Probe data. Worked on the Conway-Maxwell posterior kernel function. Needs logs.

- 03/10/2019 - 05/10/2019** Worked on fitting the Conway-Maxwell binomial distribution. Using MAP. Posterior kernel function appears to be working now.
- 06/10/2019 - 08/10/2019** Prior kernel now working. Worked on plotting functions. Need to move back to Eight\_Probe analysis. Must do COSYNE abstract and present to O'Donnell group with this analysis.
- 09/10/2019** Changed the 3-d prior and posterior plots to contour plots. They look nicer. Need to look at python maximising functions. Then get back to Eight\_Probe.
- 10/10/2019 - 14/10/2019** Changed the Conway-Maxwell estimation to Maximum likelihood. works much better. Started harder work on Eight\_Probe data. Made histograms. Making maxtrix plots.
- 15/10/2019 - 16/10/2019** Made firing rate histograms, correlation histograms, and regional correlation matrix plots for bin widths 0.2, 1, 2, and 3s. Had a meeting with Cian. Emailed Steinmetz.
- 17/10/2019** Wrote the community detection script for Eight\_Probe data. Need to make the other plots, and imitate that other plot too.
- 18/10/2019** Made regional community plots. Need to imitate the other plot, and to start the presentation.
- 19/10/2019 - 23/10/2019** Managed to imitate the figure plot. Found a similar result. Need to show the result with a statistical test. (Mann-Whitney U test). Prepared and delivered presentation. Went reasonably well. Had meeting with Cian. Need to get some draft of application finished by Monday.
- 24/10/2019** Wrote 'summary' part of COSYNE application. Need to write additional detail part. Need to keep an eye on BlueCrystal. Need to search for homogeneity measure.
- 25/10/2019 - 27/10/2019** Wrote more of the COSYNE application. Added 0.05 to the selected bin widths.
- 28/10/2019** Did some HPT, poorly. Reviewed other people's COSYNE applications. Edited my own. Emailed Steinmetz, Robbins is the wild type?
- 29/10/2019 - 30/10/2019** Caught up on Data Structures and Algorithms, a bit. Had a meeting about my COSYNE abstract. Made some changes. Submitting tomorrow.
- 31/10/2019 - 05/11/2019** Completed and submitted COSYNE 2020 abstract. Caught up with Data Structures and Algorithms.
- 06/11/2019** Attended Fortran seminar. Started Eight\_Probe latex directory.
- 07/11/2019** Attended GIT seminar, it was very good. Added to the Eight\_Probe latex.
- 08/11/2019** Added more to the Eight\_Probe latex. Started work on the fluorescence modelling paper draft. Added methods section and subsections. Looking at references and so on at the moment. Probably going to go badly.



- 09/11/2019 - 15/11/2019** Added a methods section to the fluorescence modelling paper draft. Tried cleaning up some of the references. Also added to the Eight\_Probe document. Also put together a list of conferences to go to potentially. Need to do another TODO list.
- 16/11/2019 - 18/11/2019** Thought about implementing the active population model. Added to the Eight Probe readme.
- 19/11/2019 - 22/11/2019** Read about comparing clusterings. Treated the anatomical classification as a cluster. Compared the anatomical clustering with the community detection clustering using a range of measurements. Found nice results. Had a meeting with Cian. Only very small things left to do with the Eight Probe project.
- 23/11/2019 - 24/11/2019** Added a summary frame for measuring the modularity of consensus clusterings etc. Added a pair counting metric for comparing anatomical 'clusterings' to community detection 'clusterings'. Need to run some scripts in the morning.
- 25/11/2019 - 26/11/2019** Tried to run the scripts. Crashed my machine. Will try again. Added more to the Eight\_Probe latex.
- 27/11/2019** Added more to the Eight\_Probe latex. Methods might be finished now. Had a meeting with Cian. Some options for travelling to Denver. Need to check them out.
- 28/11/2019 - 29/11/2019** Moved the calcium fluorescence paper word document over to latex and put it on overleaf. Waste of time. Need to fix the `detect_communities.py` script, and look for money to go to COSYNE.
- 30/11/2019 - 02/12/2019** Applied for travel funding on the COSYNE webpage. Unlikely to get that. Had a meeting with Mike. Need to do yet more work on the Eight Probe project. Need to work the behaviour SVD into analysis somehow. Did some marking for Data Structures & Algorithms.
- 03/12/2019** Fixed the bug with the `detect_communities.py` script. The problem was with the `bctpy` package. Fixed on desktop machine. Currently trying to process 500 neurons at 5ms time bins.
- 04/12/2019** Editing the Eight\_probe functions to avoid having to build a spike count frame. This will avoid a memory bottleneck, and hopefully will not create another one.
- 05/12/2019 - 10/12/2019** Avoiding the memory bottleneck introduced a time problem. The process takes a very long time to run now. Need to do some profiling to find out where the problem lies. Read the Gerkin paper. Don't yet know if it was useful. Need to learn about the SVD and how it was used in the Stringer paper.
- 11/12/2019 - 17/12/2019** Played around with different settings for the BlueCrystal submission script. Eventually got `bin_width_variation.py` script running. Currently running `detect_communities.py`, hopefully it'll go ok. Will have to

get a quote from Ian Allen before I get confirmation from COSYNE. Need to read more about controlling for external variables.

**18/12/2019 - 29/12/2019** On Christmas holidays. Looked for something useful in Stringer related to controlling for mouse facial movement measures. Looked again Gerkin. Implementing the procedure described in the supplemental material in Gerkin. Wrote a good function for getting the required ingredients. Need to start working on the part that involves pairs. Don't fully understand what I am doing. But, hopefully it will work out.

**30/12/2019 - 16/01/2020** Working on calculating the expected value of the conditional covariance between spike counts using the law of total covariance. It's a pain. Also invigilating at the moment. Will start on teaching assistance next week.

**17/01/2020 - 20/01/2020** Still working on calculating the expected value of the conditional variance. Have some small changes to make tomorrow, but I have no confidence in this method. Still invigilating. Meeting about applied statistics tomorrow.

**21/01/2020 - 27/01/2020** Method is not working. Over-estimation of conditional expectations may be the problem. Need to find another way of modelling. Prepared for teaching tomorrow.

**28/01/2020 - 31/01/2020** Changed the method of calculating the conditional covariance. Now rank the components by individual conditional covariance, keep the top 4, and work with those only. 4 is small enough to avoid memory issues. Need to speak to Cian about his suggestion of linear regression, could be better. Also started teaching on Tuesday. Should prepare for next Tuesday.

**01/02/2020 - 05/02/2020** Spoke to Cian about his linear regression idea. Definitely worth a shot. Did some more teaching. Created a Conway-Maxwell-binomial distribution github repo. Edited the code to make it faster also.

**06/02/2020 - 07/02/2020** Made some more edits to the repo. Started work on implementing linear regression model.

**08/02/2020 - 10/02/2020** Prepared for teaching. Need to continue with linear regression. Need to choose separated samples to break up autocorrelation in SVD components.

**11/02/2020** Added an autocorrelation function to the Eight Probe project. Need to note the autocorrelation whenever I am doing the downsampling. Maybe choose some time difference between retained samples that gives a decent number of samples remaining?

**12/02/2020 - 14/02/2020** Chose the downsampling based on number of samples required rather than autocorrelation. Need to compare linear regression using principal components to linear regression using raw data.

**15/02/2020 - 18/02/2020** Compared the performance of linear models to generalised linear models using the `statsmodels` package. The GLMs are no better than

the linear models from sklearn. Currently comparing performance using different amounts of downsampling. At some point the autocorrelation in the PCs should make the models lose performance.

**19/02/2020** Finished model comparison. Emailed Cian with details. Added figures to illustrate the linear model performance. Started on a function to use the model without doing a lot of comparison.

**20/02/2020** Model comparison was slightly inaccurate, but could be worse. Wrote full infrastructure for conditional analysis. Started something running.

**21/02/2020 - 24/02/2020** Went to Saint Helens for long weekend. Process that I started was killed because of memory issues. Prepared for teaching Applied stats.