Modelling RNA Oscillatory Circuits

J. Binysh[1]

[1] Centre for Complexity Science
University of Warwick

September 22, 2015

Outline

- My project consisted of trying to estimate unknown parameters in an ODE model, representing a recently designed genetic regulatory circuit, using some recent experimental data.
- Begin by giving biological background, describing the system we are interested in. Then discuss recent experimental data, and the ODE model proposed to represent it.
- Then move on to talk about how to estimate unknown parameters in this model - first discussing methodology, then results.
- I will show that the majority of the parameters in the model cannot be uniquely estimated given the available data. I will discuss why this might be the case, and suggest the use of a simplified model, which can describe available data equally well.
- then conclude and suggest further work.

Description of RNA Oscillatory system

- In the cell, DNA is copied into an RNA molecule, which then gets made into a protein.
- This process is naturally regulated in many ways. One stage it can be regulated at is when the RNA - called mRNA - is on its way to get made into a protein
- In order for the mRNA to be made into a protein, a piece of molecular machinerey called the Ribosome must attach itself to it. This occurs at a place on the mRNA molecule called the Ribsome Binding site (RBS). If the RBS is blocked, the mRNA cannot be made into a protein
- mRNA can ‘self repress’ - its shape is such that the molecules tail is looped over the RBS
- This self repression can be undone by introducing another RNA - a small RNA which binds to the mRNA, and uncovers the RBS
- Such a setup can be synthetically engineered - mRNA-sRNA pairs that act in this fashion can be designed. In addition, we can choose that the protein made be GFP, which fluoresces, so its concentration in the cell can be monitored.
- Further, the whole setup can be placed inside a cell, and indirectly controlled by two chemcials, aTC and IPTG, whose concentrations affect the rates of mRNA and sRNA production.
- Whole system forms a logical ‘and’ gate.

Recent Experimental data

- In recent experiments, cells with the above set-up in them have been forced with a varying aTc concentration - what we see here are individual cell fluoresence time series, with aTC forcing schematically shown.

ODE Model for system (1)

- We can model this system with a set of ODE's with mass action kinetics.
- The first set represent the production of the two RNA's discussed above, and their hybridization. This is represented as a two step process - first into an unstable, then a stable, complex.
- aTc forcing is modelled via the y term. The various delta's are just degradation rates of the complexes involved, and mu represents dilutions of the chemicals due to cell growth.

ODE Model for system (2)

- The next set of equations explicitly models protein production and observed fluorescence. beta m + fs beta c term describes the translation of the mRNA into an immature GFP (p). This then matures into mature GFP (g), which fluoresces (z)

Parameter Estimation in ODE Model

- So we have our ODE model, and some recent time series data. Our goal is to estimate some of the unknown parameters in the described model, by fitting it to the data.
- To do this, we numerically simulate our ODE model, giving a predicted fluorescence time series for some set of parameters. We then define a least squares error between this simulation and our observed data, and seek to find the set of parameters which minimise this error.
- The error landscape may have many minima - to perform the minimisation we use an Evolutionary algorithm, the CMA-ES.

Parameters to be estimated in our model

- In our case, we have 9 unknown parameters, highlighted in red. 5 (name(only appear in the model before the translation step, in the initial hybridization process. Two (name) are associated with the translation process. Mu appears throughout the equations, and theta just represents machine calibration.

Initial Estimation Results

- Prediction from a typical set of parameters found shows model capable of quantitatively capturing data

Initial Estimation Results

- However, many parameters are poorly estimated. These histograms were made by randomly starting the CMA-ES at a point within an initial bounding box, and letting it run 200 times. We see little consistency in estimates for many parameters. The x axis on the histograms are the initial bounding box given, and we see some estimates spread right across the range.
- However, some parmaters (mu, theta) are relatively consistently estimated.
- Why is this the case?

Sensitivity Analysis

- Taking one of the estimated parameter sets, and performing a local sensitivity analysis about it, suggests a possible explanation.
- We see that many parameters, when slightly perturbed, have similar effects on model output - thus making their effects hard to distinguish , and so hard to simultaneously estimate
- Curves for Mu and theta relatively distinct.
- This result suggests the minimisation algorithm may be working with many fewer degrees of freedom than we gave it - only 3, rather than 9.

Model fixed point for estimated parameters

- Although individual parameter estimates are very loose, if we look at all estimated parameter sets, there is some consistency to be found. Recall that the initial hybridization steps etc were communicated to the equations modelling translation by the $\beta m + f_s \beta c$ term (flick back). If we solve for the models fixed point, and look at the fixed point values of $\beta m + f_s \beta c$, we see similarity between all parameter sets. This suggests that the algorithm is really only altering the values of $\beta m + f_s \beta c$, mu and theta, rather than the full 9 degrees of freedom.
- a biological interpretation of this fact may be that translation is a rate limiting step.

Simplified Model (1)

- This suggests we can simplify by assuming the first set of equations occur instantly, and instead just force and model the translation step. We can then re-estimate our 3 parameters as before.

Simplified Model Results (2)

- As before, running the CMA- ES provides us with parameter estimates. Unlike before, they are clear of the initial bounding box Further, we achieve error values as low as full model. Two peak structure likely a consequence of local minima.

Simplified Model Results (3)

- The model is also now simple enough we can plot the error landscape, the contours of which offer a better indiction of uncertainity in F and theta than the histograms do.

Future Work