Stochastic Models of Gene Regulatory Networks

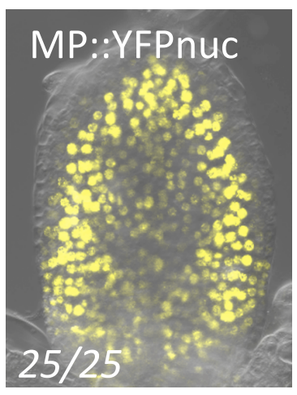
John Ng, Laura Brown and Henrik Jönsson

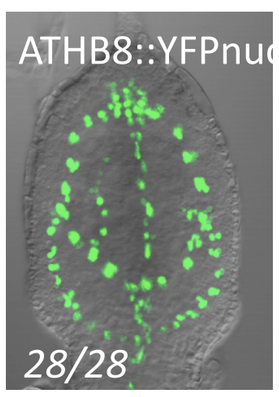
Introduction

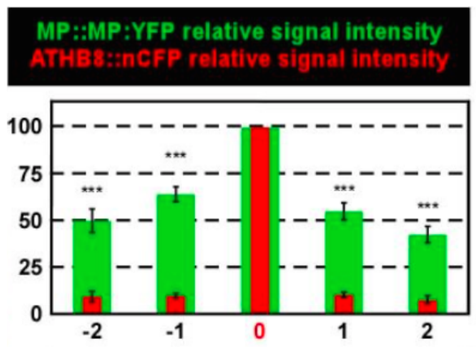
Vascular development of leaves are intricately regulated by gene expression. In order to gain insights of the processes and dynamics involved, we wish to derive computational models from the gene regulatory network and test those models against experimental data. This helps to guide further experimentation. The experimental data in this project was collect from the model plant *Arabidopsis thaliana*.

Biological Background

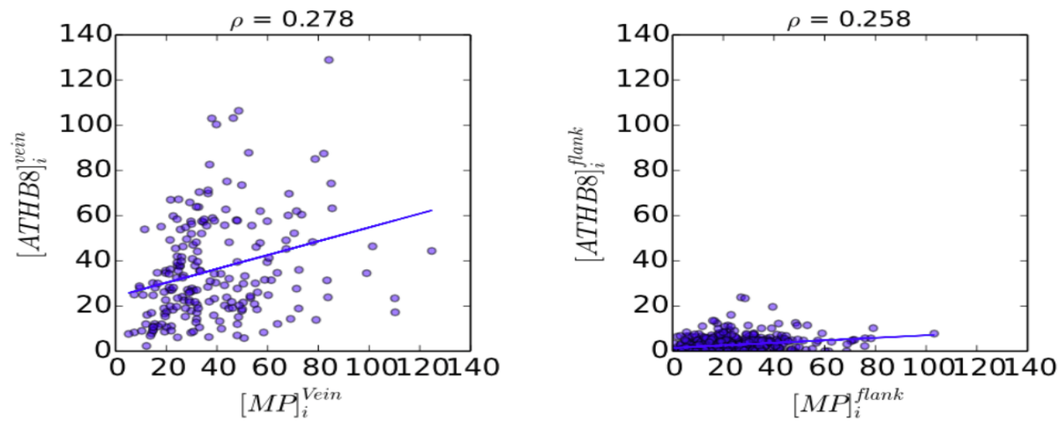
MONOPTEROS (MP), an auxin response factor, has broad expression. However, HD-ZIP III gene **ARABIDOPSIS THALIANA HOMOEBOX8 (ATHB8)**, a pre-provascular marker, has narrow domains. The response of ATHB8 to manipulation of MP levels has been quantified.







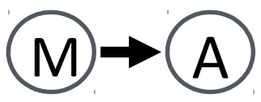
Experimental Data: We observe a lot of variation in the experimental data. Originally the data of ATHB8 and MP levels was categorized into flank and vein. However since both sets have to be described by the same equation, it is sensible to merge the data.



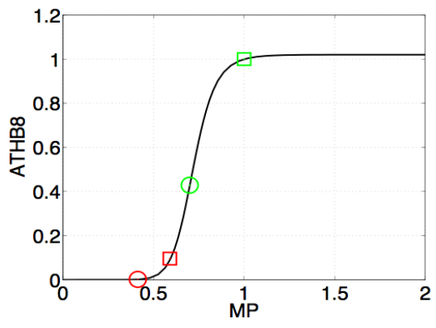
Modelling

Model 1 : A simple model of transcription

This is the simple Hill's equation (steady state) derived from the Michaelis-Menten formalism.



$$A = \frac{VM^n}{(M^n + K^n)}$$



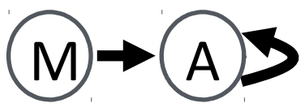
Stochastic Modelling: In this project we mostly consider models with stochasticity/noise, despite deterministic modeling being simpler. Previous studies have found that noise changes the position or even the number of stable states. Stochastic stabilisation has been demonstrated with Hill-type equations. Noise is essential in genetic regulation and cell functioning, i.e. it is essential functionally and not as a correction [1].

We utilize a useful stochastic tool known as the **Langevin equation.** Noise is inversely proportional to *E*, the effective cell volume.**ξ** is the uncorrelated Gaussian white noise whose fluctuation depends on the magnitude of components.

$$\frac{dA_i}{dt} = V\frac{M_i^n}{K^n + M_i^n} - A_i + \sqrt{\frac{V\frac{M_i^n}{K^n + M_i^n} + A_i}{2E}}.\xi_i(t)$$

Equation 1: Langevin equation describing Model 1 with stochasticity

We use Heun algorithm with Ito interpretation to numerically integrate the Langevin.



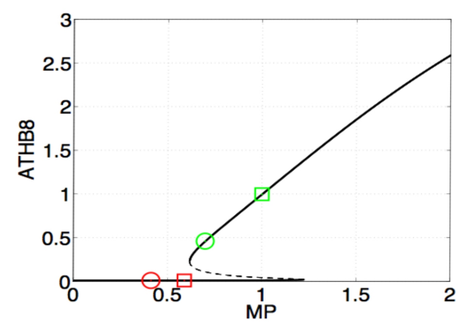
Model 2: Stochastic Hill's with self regulation (A)

On the left is the resulting model when we add self-regulation (ATHB8). It follows a more general Hill-form. The ε term is the multiplicative noise from the Langevin equation.

The bifurcation diagram of this model shows the existence of two stable and one unstable fixed points for an intermediate value of MP.

Note that we implicitly assume in model 1 and 2 that MP obeys a deterministic differential equation with a degradation term and a constant rate of change.

$$A = \epsilon + \frac{VM^nA^n}{(M^n + K_1^n)(A^n + K_2^n)}$$



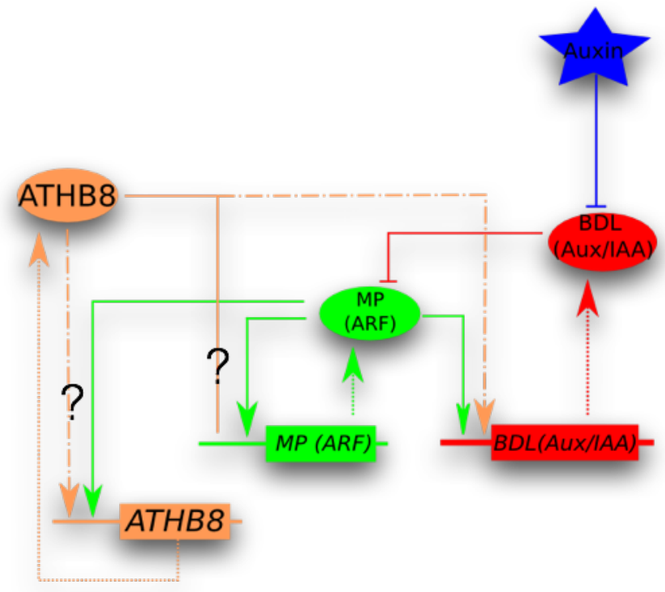
We can consider more complex models by assuming the rate equation of MP is subject to noise:

Model 3: Model 1 + stochastic MP

Model 4: Model 2 + stochastic MP

On the left is the gene regulatory network involved in vascular development. Other modelling options include:

1. Adding more feedback terms
2. Adding the role of BDL



Analysis

A. Motivation

1. Simulation: We use the software "Organism" to simulate species outcome. Varying parameters helps to understand the system and how the parameters affect the bigger picture.

2. Optimisation: Given a model, we want to find the best parameters that fits the actual data.

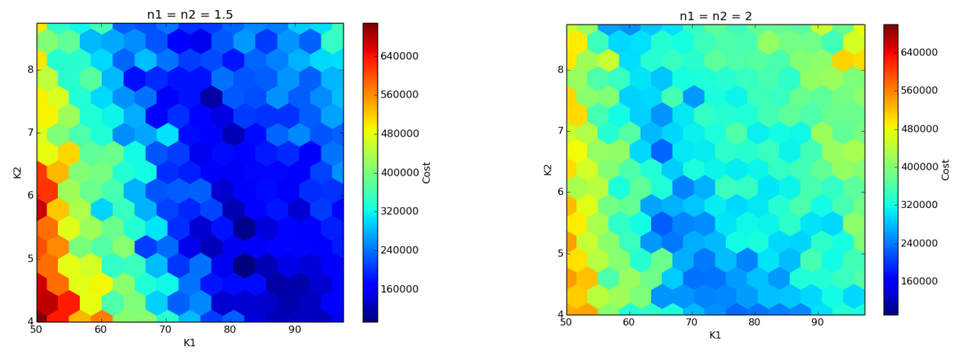
B. Stochastic Optimisation

1. Define an objective function (Input = Parameter *n*-tuple, Output = cost). This defines a distance between the actual dataset and the parameter-tuple (via the simulated datasets in Organism).

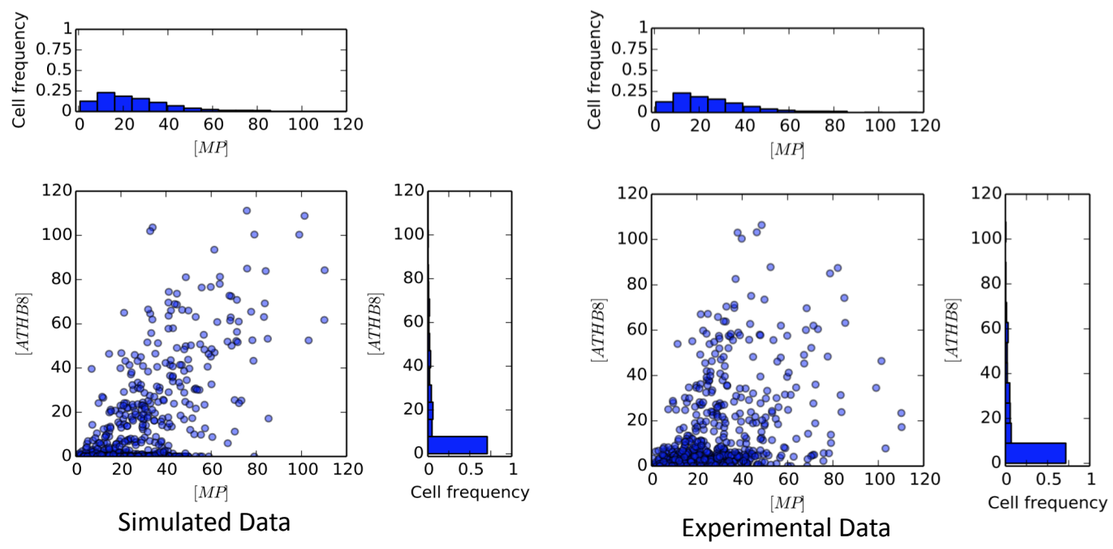
2. Perform a parameter sweep. Choose, say *k* tuples with lowest costs, then run optimisation.

C. An Example: Finding suitable parameters for Model 2

The following is an example of parameter sweep (in practice, optimisation is done over higher dimensions). n=1.5 has more extreme values than n=2. Looking for low cost, K1=80, K2=5.5, n=1.5 seems to be a good parameter candidate, we then apply local optimisation method to it.



Scatterplots/Histogram : The right side illustrates the experimental data, 540 out of 754 data points have ATHB8 level less than 8.9. The left side is a simulated dataset from (V=140, K1=80, K2=5.5, n1=n2=1.5, E=0.08). This parameter-tuple performs well, consistently giving simulations close to actual data by observing the histogram of cell frequency. This particular simulation has 541 out of 754 data points having ATHB8 level below 8.9.



D. Discussion

1. Objective functions: More elaborate versions incorporate skewness, kurtosis, mean square etc

2. Model comparison: Again by using objective function. We can include more feedback terms or modify the nature of noise. We want a good model yet not overfitting one. So next we consider:

3. Predictive power: We can use 'cross-validation' to choose the model with best predictive power: Randomly divide the data into *n* sets, train *n-1* sets and test on 1 set, and repeat.

Reference

[1] Weber M, Buceta J (2013) Stochastic stabilization of phenotypic states: The genetic bistable switch as a case study. PLoS ONE 8(9): e73487.doi:10.1371/journal.pone.0073487