

Parameter and sensitivity analysis for large system of ODEs

Summary With the recent rise of systems biology, computational models play an increasingly important role in the study of cellular signaling networks in molecular biology. This has led to a number of novel, interesting problems concerning the inference of network topologies and parameters. In these two projects we will focus on challenges in estimating parameters and in analysing models' sensitivity to parameter variation. In particular, we will be concerned with 1) the quality and applicability of different methods for fitting differential equation models to biological data, and 2) techniques to estimate the impact of parameter deviations from this best fit estimate. The first part is challenging because experimental data is, in general, incomplete, inaccurate, and in part erroneous. Estimating the impact of deviations from the estimates is therefore crucial to gauge the predictive quality of the model. The second part is interesting also because it hints at how the dynamical system might be regulated or evolve. We will consider methods both for a local and a global stability analysis in one and more dimensions.

The system These methods will be investigated with focus on two related models [1, 2] that describe two different stress responses in *B. subtilis*, one of which involves cell differentiation (formation of spores). The key step is the activation of the transcription factors σ^F and σ^B . These are regulated by a network of three homologous proteins, a kinase, a phosphatase and their common substrate (Fig. 1A). The system is interesting for a number of reason. Firstly, it provides a simple, well-studied example of cell differentiation. Furthermore, similar networks exist in other pathological bacteria, including *M. tuberculosis*, and an understanding of the networks in *B. subtilis* can be hoped to help combating infection. Moreover, the availability of mathematical models for homologous networks provides an opportunity to study network evolution based on a concrete example.

The models are based on detailed experimental information (kinetic binding studies and enzymatic assays) and successfully predict essentially all mutants for which biochemical information is available. Given their detailed nature the models are large (150 reactions and 50 variables in case of σ^F and more than doubly as many in case of σ^B); the 150 reactions are governed by 21 different kinetic parameters. The system of equations is derived from Figure 2 in the Supplementary Material in [1] using the law of mass action. Accordingly the ODEs are coupled, non-linear and of degree 2; a general form would be

$$\frac{dx_i}{dt} = \mu_i + \sum_{l,m=1;l \neq m;l,m \neq i}^{l,m=50} \alpha_{l,m} x_l x_m - \sum_{l=1}^{l=50} \beta_l x_l x_i - \gamma_i x_i \quad (1)$$

with $i = [1; 50]$. Most decay terms, γ_i , and all but four source terms, μ_i , are zero; for those that are not zero $\mu_i = [10^{-3}; 10^{-2}]$, $\gamma_i = 4.1 \times 10^{-4}$. Most parameters were estimated independently but standard deviations are not available for all of these, and should be obtained in this project.

Project 1 A number of algorithms exist to estimate parameters [3], and software packages (including Matlab, Mathematica, Gepasi, Madonna) have implemented these. PottersWheel (<http://www.fdm.uni-freiburg.de/maiwald/PottersWheel>), the software package that we will use in conjunction with Matlab, uses the weighted χ^2 value for parameter optimization,

$$\chi^2 = \sum_{i=1}^N \frac{(y_{Model}(i) - y_{Meas}(i))^2}{\sigma_{Meas}^2(i)}. \quad (2)$$

N is the number of all data points and $\sigma_{Meas}^2(i)$ the standard deviation of measurement i , $y_{Meas}(i)$. The objective function, χ , is minimized using either of various *optimization strategies*, including line search (steepest-descent, Newton's method) and trust regions, or *Simulated Annealing*, or *Genetic Algorithms*.

The scope of this project is to learn about these algorithms in detail by using them on two existing, large systems of coupled ODEs that describe the stress response in the bacterium *B. subtilis*. Subsequently, the methods shall be used to efficiently estimate the parameter values for a new, but related signaling network (RsbRST) that exhibits a similar topology as the two networks studied initially but which has a different biological function (and thus exhibits different dynamics), and accordingly can be expected to employ different parameter values. An interesting aspect of the network (which most likely will provide some challenges) is that the network organises into larger complexes. The significance of these complexes is not understood and this question is currently receiving ample experimental attention [4]. Mathematical modeling can be expected to make important contributions as the likely consequence of these complexes is further cooperativity (non-linear effects) which are best understood by a combination of experimental and theoretical studies. However, this will then require some important changes to the equations. The last part would thus touch on ODE inference which in itself constitutes an independent project.

Workflow

- 1st-2nd week** read up on parameter estimation algorithms (a good starting point is the PottersWheel manual and references therein), get acquainted with the two available computational models for the regulation of σ^F and σ^B
- 3rd-4th week** transfer models for the σ^F and σ^B networks from Matlab code (available on request) into PottersWheel format, use PottersWheel to refit the data, obtain confidence intervals; compare to previous parameter values
- 5th week** simulate the model with the newly estimated parameters - do key predictions in [1] (e.g. sigmaF release in response to a 3-fold increase in SpoIIE activity) still hold?
- 6th-7th week** generate a model for a third, related network (RbsRST): start by reading up on the regulation of σ^B by the RST module, get data from publications, liaise with experimental collaborator (R. Lewis, Newcastle)
- 8th week-end** generate parameter estimates, check consistency of model with available literature (also on mutants) - What role do the protein complexes play?

Project 2 The second project is concerned with a sensitivity analysis of the models with respect to changes in the parameter values. This is important to understand how both inaccuracies in our parameter estimates as well as physiological and evolutionary changes affect the model's behaviour (dynamics). Given the inherent uncertainty in our parameter predictions gauging the reliability of the models' predictions is crucial. Equally understanding the regulatory impact of single parameters as well as of parameter combinations is important to understand the logic of signaling networks, and in order to be able to affect signaling networks effectively, for instance in treating diseases. Such insights are interesting also to obtain a better insight into the evolvability of networks. Finally, it could also guide experimentation if we knew in what fraction F of the (physiological) parameter domain the dynamics remain essentially the same, and shrinking of which parameter range (by further experiments) increases F the most.

Comprehensive sensitivity analyses of large computational models are challenging. Already a local perturbation analysis that involves perturbations in all possible parameter pairs is computationally costly for large models. A global sensitivity analysis is unequally harder. A number of algorithms have been proposed of how to explore the physiological parameter space efficiently [5], and all of them involve some kind of Monte Carlo approach. The basic idea of these algorithms is to remove correlation between parameters (to save computational time), and to perturb each parameter by drawing parameter values randomly from a distribution centered at the estimated parameter value. The variance of the distribution should be set based on the standard deviation in each parameter estimation. In order to explore the evolvability of these networks the variance should be set based on physiological mutation rates. For such a Monte Carlo approach to be efficient the dynamical system must be solved using C++ rather than Matlab. In order to make the task easier we will consider a simplified model for σ^F regulation that considers only eight key components and which is described in detail in the appendix.

Workflow

- 1st week read up on algorithms for parameter estimation (a good starting point is the PottersWheel manual (<http://www.fdm.uni-freiburg.de/maiwald/PottersWheel>) and references therein or for more details [3]) and sensitivity analyses (local and global) [5]; get acquainted with the (simplified) σ^F model described above
- 2nd week import the (simplified) σ^F model into PottersWheel (Matlab), use PottersWheel to fit the data, obtain confidence intervals
- 3rd-4th week solve the (simplified) σ^F model using C++ for the new parameter values, compare model predictions (sigmaF release in response to an increase in SpoIIE activity, irreversibility of sigmaF activation, energy efficiency to name the most important features) to those of the detailed model - conclusion on suitability of methods and models?
- 5th week-end code algorithms for sensitivity analysis, start with local perturbation method, then turn to global sensitivity algorithms; use previously determined confidence intervals to adjust distributions. Measure the effects of parameter perturbations by monitoring the change in the speed and level of σ^F that is released in response to an increase in SpoIIE activity. Compare results from different algorithms, comment on speed and quality of different algorithms.

The simplified model for σ^F regulation : σ^F (s), AB-ATP (b), AB-ADP (b_d), $\sigma^F \cdot$ AB-ATP (c), $\sigma^F \cdot$ AB-ATP-AA (c_1), AB-ADP-AA (c_2), AA (a), AA-P (a_p). The scheme in Figure 1B can be translated into the following set of eight differential equations

$$\dot{c} = k_{11}c_1 + k_5bs - l_1ac \quad (3)$$

$$\dot{b} = k_2b_d - k_5bs \quad (4)$$

$$\dot{b}_d = k_4c_1 + k_{16}c_2 - (k_6a + k_2)b_d \quad (5)$$

$$\dot{c}_1 = l_1ac - (k_{11} + k_4)c_1 \quad (6)$$

$$\dot{c}_2 = k_6ab_d - k_{16}c_2 \quad (7)$$

$$\dot{a}_p = k_4c_1 - k_3a_p \quad (8)$$

$$\dot{s} = k_4c_1 - k_5bs \quad (9)$$

$$\dot{a} = k_{11}c_1 + k_{16}c_2 + k_3a_p - (k_6b_d + l_1c)a \quad (10)$$

The dot refers to the time derivative (e.g. $\dot{c} = \frac{dc}{dt}$) such that equations (3-10) model the dynamics of c , b , b_d , c_1 , c_2 , a_p , s , a respectively. Thus, binding of AA (a) to $\sigma^F \cdot$ AB-ATP (c) at rate l_1 leads to the formation of $\sigma^F \cdot$ AB-ATP-AA (c_1) which either unbinds a at rate k_{11} or binds s at rate k_4 . In the cell, unbinding of σ^F permits the binding of a second AA. This step is neglected in this model, and in order to keep the model simple it is instead assumed that unbinding of s coincides with the phosphorylation of the bound a ; it has been checked that this simplification does not substantially alter the overall dynamics. Phosphorylated a can be reactivated at rate k_3 ; k_3 reflects the product of the turn-over rate and the concentration of the phosphatase SpoIIE. AB-ADP (b_d) either binds a at rate k_6 to form AB-ADP-AA (c_2) or exchanges ADP for ATP at rate k_2 to form b . b and s can rebind at rate k_5 to form c .

As compared to the detailed model described in [1] this model makes several important simplifica-

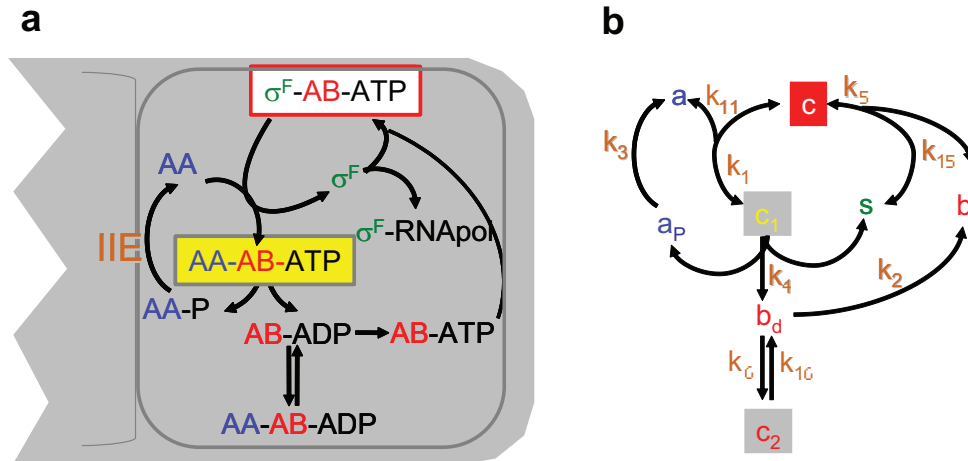


Figure 1: A simple model of the network that controls σ^F activity in *Bacillus subtilis*. (A) An overview of the interactions in the network that controls σ^F activity. For details see text. (B) The simplified model that is analysed mathematically. s refers to σ^F , a to AA, a_p to AA-P, b to AB-ATP, b_d to AB-ADP, c to $\sigma^F \cdot$ AB-ATP, and c_1 to $\sigma^F \cdot$ AB-ATP-AA. The rate constants are summarized in Table 1.

Table 1: Kinetic rate constants and protein concentrations employed in the model.

parameter	functional significance	simulation value
k_1	AB·ATP-AA on-rate	$8 \times 10^5 \text{ M s}^{-1}$
k_{11}	AB·ATP-AA off-rate	0.001
k_2	ADP-ATP exchange rate	0.02 s^{-1}
k_3	effective rate of AA-P dephosphorylation	0.003 s^{-1}
k_4	σ^F -AB·AA off-rate	0.4 s^{-1}
k_5	σ^F -AB on-rate	$4 \times 10^5 \text{ M s}^{-1}$
k_6	AB·ADP-AA on-rate	$8 \times 10^5 \text{ M s}^{-1}$
k_{16}	AB·ADP-AA off-rate	0.0074
a_T	total AA concentration	$30 \mu\text{M}$
b_T	total AB concentration	$25 \mu\text{M}$
s_T	total σ^F concentration	$15 \mu\text{M}$

tions. The most important one is that only binding of one AA per AB is considered. The impact of allostery can nonetheless be explored in this model since its main effect and importance in the detailed model was to facilitate AA-dependent σ^F release. This effect can be captured by the choice of the parameter value for k_4 . Thus without allostery k_4 would correspond to the normal σ^F ·AB offrate which is so small ($5.5 \times 10^{-3} \text{ s}^{-1}$) that the back-reaction from c to b and s is not included in the model; the allosteric conformational change increases the rate to 0.4 s^{-1} and the impact of such difference in k_4 can be studied here. Direct rebinding of σ^F after its release from σ^F ·AB·ATP·AA is ignored in the simple model since we know from the detailed model that in the presence of sufficient AA to release σ^F the neglected binding of the second AA and phosphorylation of AA are more efficient than the rebinding reaction. Equally we neglect binding of σ^F to AB·ADP because the quantitative model demonstrates that this interaction is disfavoured, presumably because of the inefficient and slow σ^F -dependent closure of the nucleotide-”lid” in the presence of ADP and because of the efficient formation of AB·ADP·AA when sufficient AA is around to bind AB (which leads to lid closure) [1]. Another important simplification concerns the phosphatase SpoIIE which is not modeled explicitly, but which is assumed to be sufficiently scarce compared to SpoIIAA-P that the phosphatase turn-over rate rather than the SpoIIE-SpoIIAA-P binding reaction is rate-limiting. In fact the IIE concentration is about $15 \mu\text{M}$ at the time of septation [6] and the IIE-AA-P on-rate has been estimated as $10^5 \text{ M}^{-1} \text{ s}^{-1}$ [1] such that the effective rate at which AA-P becomes bound by IIE is $10^5 \text{ M}^{-1} \text{ s}^{-1} \times 15 \times 10^{-6} \text{ M} = 1.5$ while the *in vivo* phosphatase rate has been estimated as 0.0087 s^{-1} [1], which is two orders of magnitude lower. Accordingly, in the detailed quantitative model about 60% of all SpoIIE is bound by SpoIIAA-P at the time of septation such that this is a reasonable assumption. Asymmetric septation results in a sudden increase of the IIE concentration in the prespore by at least 2.5-fold [1]. Since we are interested only in the effect of this sudden septation-dependent increase in the IIE activity in the prespore we can neglect protein expression and degradation which do not play a role on this timescale. Accordingly the total concentrations of σ^F , AB and AA are conserved, and there are only five independent variables in the model since we have

$$s_T = s + c + c_1 = \text{const}, \quad b_T = b + b_d + c + c_1 + c_2 = \text{const}, \quad a_T = a + a_p + c_1 + c_2 = \text{const}. \quad (11)$$

When solving the model we can use the parameter values that we estimated before for the detailed model [1]. The value for k_4 , the maximal rate of AA-dependent σ^F ·AB·ATP dissociation, has been discussed above. We further have $k_5 = 4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ for the σ^F -AB·ATP on-rate. k_2 represents the rate limiting steps in the AB ADP-ATP exchange reaction which we have estimated as 0.02 s^{-1} . As discussed

above the phosphatase reaction rather than IIE-AA-P binding is rate limiting in AA phosphorylation, and this rate has been estimated as 0.0087 s^{-1} . However, since only 60% of all IIE is bound to its substrate this rate would be an overestimate for k_3 , and $k_3 = 0.003 \text{ s}^{-1}$ would be a more realistic estimate for the pre-septation situation. Any increase in IIE and AA-P is assumed to translate into a direct increase in k_3 , the rate at which AA-P is turned over, since enzyme is the limiting factor and the more enzyme is available to bind and process the substrate the higher the turn-over rate. The concentration of SpoIIAB and σ^F at the time of septation have been measured as $b_T \sim 25 \times 10^{-6} \text{ M}$ and $s_T \sim 15 \times 10^{-6} \text{ M}$ [7, 6, 8]. The SpoIIAA concentration is about twice the concentration of SpoIIAB at the time of septation [9]. However, given that in the model binding of only one AA per AB is considered we are using $a_T = 30 \times 10^{-6} \text{ M}$.

References

- [1] Iber D, Clarkson J, Yudkin MD, Campbell ID (2006) The mechanism of cell differentiation in *Bacillus subtilis*. *Nature* 441:371–374.
- [2] Iber D Stress-dependent regulation of σ^B .
- [3] Raol J, Girija G, Singh J (2004) Modelling and Parameter Estimation of Dynamical Systems. IEE.
- [4] Hecker M, Pane-Farre J, Voelker U (2007) SigB-Dependent General Stress Response in *Bacillus subtilis* and Related Gram-Positive Bacteria. *Annu Rev Microbiol* 61:215–36.
- [5] Zheng Y, Rundell A (2006) Comparative study of parameter sensitivity analyses of the TCR-activated Erk-MAPK signalling pathway. *IEE Proc Syst Biol* 153:201–211.
- [6] Lucet I, Borriss R, Yudkin MD (1999) Purification, kinetic properties, and intracellular concentration of SpoIIE, an integral membrane protein that regulates sporulation in *Bacillus subtilis*. *J Bacteriol* 181:3242–3245.
- [7] Ho MS, Carniol K, Losick R (2003) Evidence in support of a docking model for the release of the transcription factor σ^F from the antisigma factor SpoIIAB in *Bacillus subtilis*. *J Biol Chem* 278:20898–20905.
- [8] Lord M, Barillà D, Yudkin MD (1999) Replacement of vegetative σ^A by sporulation-specific σ^F as a component of the RNA polymerase holoenzyme in sporulating *Bacillus subtilis*. *J Bacteriol* 181:2346–2350.
- [9] Magnin T, Lord M, Yudkin MD (1997) Contribution of partner switching and SpoIIAA cycling to regulation of σ^F activity in sporulating *Bacillus subtilis*. *J Bacteriol* 179:3922–3927.