Bachelor Project – Exploring the link between data and model using the Profile Likelihood method

Dynamical models are often used to gain a more comprehensive understanding of a biological system. Such models consist of state variables (or states), which are quantities that change over time and parameters (which are fixed with respect to time). These are then embedded in a system of equations which relate the different states of the model to time. To simulate the model and make predictions, these model parameter values are required.
In signaling applications, the parameters are typically not directly accessible by experiment and are obtained indirectly by calibrating the observable model outputs to their measured counterparts by iteratively adjusting model parameters.

Precise model parameterization requires a large amount of data, requiring simulation of various experimental conditions. Additionally, measurements on the system are typically performed using different experimental modalities. There are two issues with this.

1. Sometimes different experimental results can be in disagreement with each other.
2. A large number of conditions can hamper model analysis due to computational complexity.

It is therefore interesting to explore how the model and data are tied together. In this project, we will explore the relation between a model and the individual data sets which were used to parameterize it. Preliminary work revealed that many of the data sets correlate, and typically only a few datasets provide actual parameter constraints.

In this preliminary work, we made use of the profile likelihood. The profile likelihood is the result of continually re-calibrating parameters, while forcing one parameter to change over a range of values. Hereby a path is traced through parameter space which provides an optimal fit under the constraint that the profiled parameter takes on specific values. For identifiable parameters, forcing the system to deviate from the optimal parameter set results in loss of data fidelity (the distance between model and data increases). If we now plot the individual contributions to the distance between model and data for each dataset, we obtain Figures 1 and 2, which show both correlations and anti-correlations.

The aims of this project are threefold.

1. Explore constraints provided by data for various (toy) model and data configurations. Do all datasets imply the same model constraints, or are there inconsistencies? Can inconsistencies reliably be detected and what would they look like?
2. Can we reduce the amount of data to a representative set based on information we obtain from a profile likelihood? Can we translate this to predictions?
3. When a model doesn't fit, can we explore the contradictory constraints using a similar approach (for example by performing a profile likelihood for the estimated model errors)? Can we relate this to the model parameters? Does it provide insight on which parts of the model are wrong?

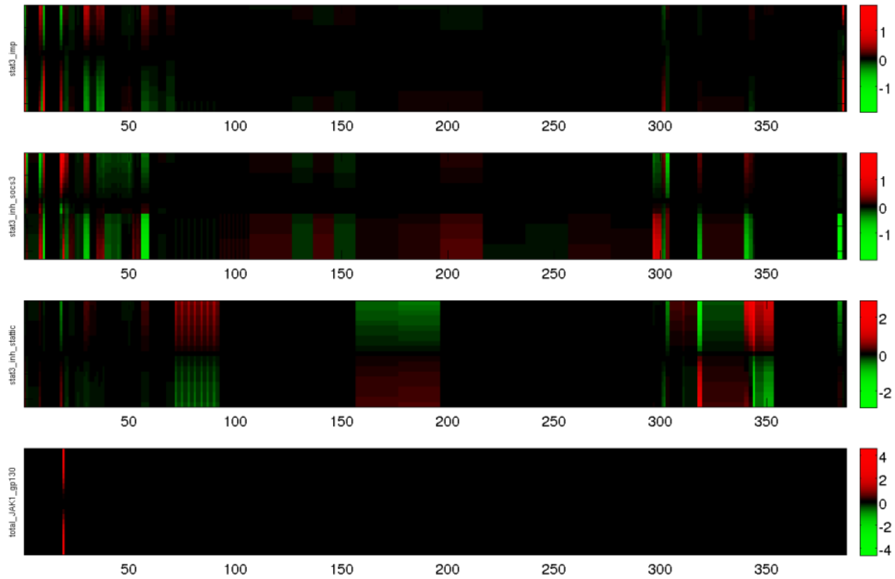

Figure 1: Contributions of different datasets to the profile likelihood. Each column corresponds to a single dataset. Red indicates that the data has a constraining effect (prevents the parameter from being at this value), while green indicates the opposite. Note that certain datasets correlate.

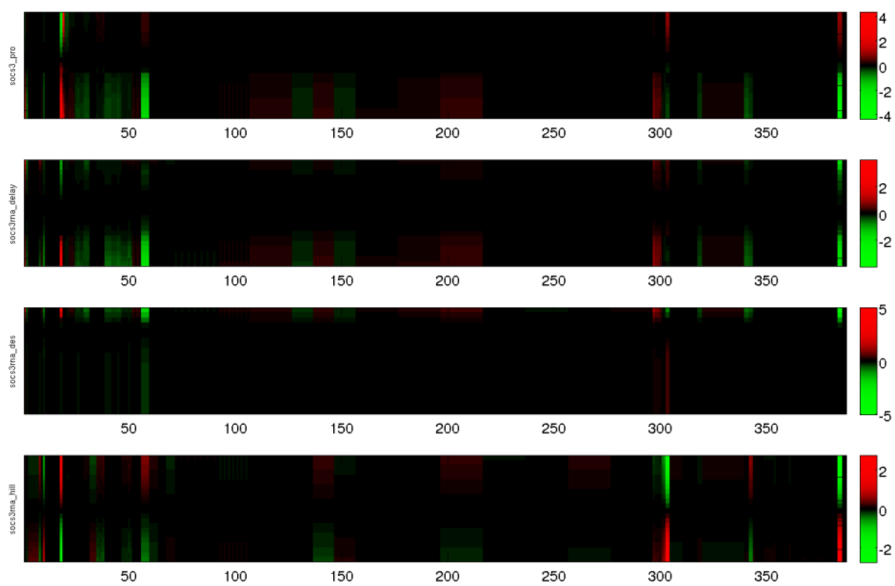

Figure 2: Contributions of different datasets to the profile likelihood. Each column corresponds to a single dataset. Red indicates that the data has a constraining effect (prevents the parameter from being at this value), while green indicates the opposite. Note that certain datasets are contradictory.