Targeted experiment design using the posterior predictive distribution

take advantage of both Bayesian and frequentist methods. The elegance of Bayesian methodology is founded in the propagation of information content provided by experimental data and prior assumptions to the posterior probability distribution of model predictions. However, for complex applications experimental data and prior assumptions potentially constrain the posterior probability distribution insufficiently. In these situations Bayesian Markov chain Monte Carlo sampling can be infeasible. From a frequentist point of view insufficient experimental data and prior assumptions can be interpreted as non-identifiability. The profile likelihood approach offers to detect and to resolve non-identifiability by experimental design iteratively [Raue et al., 2009]. Therefore, it allows one to better constrain the posterior probability distribution until Markov chain Monte Carlo sampling can be used securely. Using an application from cell biology [Becker et al., 2010] we compare both methods and show that a successive application of both methods facilitates a realistic assessment of uncertainty in model predictions [Raue et al., 2012].

References

A Raue, C Kreutz, T Maiwald, J Bachmann, M Schilling, U Klingmüller, and J Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

V Becker, M Schilling, J Bachmann, U Baumann, A Raue, T Maiwald, J Timmer, and U Klingmueller. Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science*, 328(5984):1404–1408, 2010.

A Raue, C Kreutz, F Theis and J Timmer. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. To appear in *Phil Trans Roy Soc A*, 2012.

Joep Vanlier (TU Eindhoven), Christian Tiemann, Peter Hilbers and Natal van Riel

Targeted experimental design using the posterior predictive distribution

Introduction. Systems biology employs mathematical modeling to further our understanding of biochemical pathways. The complexity of models necessary to describe biological pathways in combination with the limited amount of quantitative data results in large parameter uncertainty which propagates into model predictions. When predictions required to test the hypothesis are insufficiently constrained more data will be required. However, it is often not immediately evident which measurement(s) at which specific time point(s) would be most informative. We focus on designing experiments specifically targeting the variance of quantities of interest that depend on model predictions.

Methods. In this work we used a Bayesian approach to infer a posterior distribution based on the model and the data. Self-normalized importance sampling of the Posterior Predictive Distribution (PPD) was used to perform Optimal Experiment Design (OED).

Results. We proposed a flexible Bayesian method for hypothesis driven experimental design that exploits relations within the posterior predictive distribution whilst considering finite measurement accuracy and model uncertainty. This approach is endowed with the ability to consider multiple measurements under multiple experimental conditions simultaneously. Moreover, the method allows great freedom in terms of quantity of interest. Experiment(s) can be optimized for any quantity that can be expressed in terms of the model and model parameters. We present our method by illustrating its use on a model of the JAK-STAT signaling pathway.

Ivan Vujacic (University of Groningen), Javier Gonzalez and Ernst Wit

A new statistical framework to infer gene regulatory networks with hidden transcription factors

Regulatory networks consist of genes encoding transcription factors (TFs) and the genes they activate or repress. Various types of systems of ordinary differential equations (ODE) have been proposed to model these networks, ranging from linear to Michaelis-Mentin approaches. In practice, a serious drawback to estimate these models is that the TFs are generally unobserved. The reason is the actual lack of high-throughput techniques to measure abundance of proteins in the cell. The challenge is to infer their activity profile together with the kinetic parameters of the ODE using level expression measurements of the genes they regulate. In this work we propose general statistical framework to infer the kinetic parameters of regulatory networks with one or more TFs using time course gene expression data. Our approach is also able to predict the activity levels of the TF. We use a penalized likelihood approach where the ODE is used as a penalty. The main advantage is that the solution of the ODE is not required explicitly as it is common in most proposed methods. This makes our approach computationally efficient and suitable for large systems with many components. We use the proposed method to study a SOS repair system in Escherichia Coli. The reconstructed TF exhibit a similar behavior to experimentally measured profiles and the genetic expression data are fitted properly.

Posters

Miguel Atencia (University of Malaga) and Gonzalo Joya

Parameter estimation for dynamical systems based upon Hopfield and Tank