# Dealing with Stochasticity in Biological ODE Models

**Hamda bint e Ajmal, Michael Madden and Catherine Enright**

Discipline of Information Technology, National University of Ireland Galway

Email:h.ajmal1@nuigalway.ie, michael.madden@nuigalway.ie, cathenright@gmail.com

### Abstract

Mathematical modeling with Ordinary Differential Equations (ODEs) has proven to be extremely successful in a variety of fields, including biology. However, these models are completely deterministic given a certain set of initial conditions. We convert mathematical ODE models of three benchmark biological systems to Dynamic Bayesian Networks (DBNs). The DBN model can handle model uncertainty and data uncertainty in a principled manner. They can be used for temporal data mining for noisy and missing variables. We apply Particle Filtering algorithm to infer the model variables by re-estimating the models parameters of various biological ODE models. The model parameters are automatically re-estimated using temporal evidence in the form of data streams. The results show that DBNs are capable of inferring the model variables of the ODE model with high accuracy in situations where data is missing, incomplete, sparse and irregular and true values of model parameters are not known.

## Introduction

Mathematical models of physical systems are widely available in many domains (Ottesen et al., 2004). These models embody existing expert knowledge and can be considered sufficient statistics of all prior experimentation in the domain.

Mathematical modeling with ODEs has a very long tradition in biology. The earliest biological ODE models date back to 18th century (Malthus, 1798; Verhulst, 1845) . ODE models assume that the observed dynamics of a system are exclusively driven by internal, deterministic mechanisms and there is no uncertainty in the process. However, in reality, biological systems are always subject to 'unexplainable' influences which are neither easily understood nor is it feasible to model them explicitly. All biological dynamical systems evolve under stochastic forces. Models of biological systems are concerned with subsystems of the real world, therefore ideally, they must include the effect of external random influences as they cannot be completely isolated from model (Ditlevsen and Samson, 2013). There are many factors in biological processes that are difficult to model explicitly with ODEs, for example, harmonic oscillations, enzymatic processes, or individual characteristics like body mass index, genes, lifestyle, habits, age etc. Such factors can contribute to the erratic behavior in a model. Other factors may include measurement noise or data uncertainty. These multiple sources of noise need to be modeled explicitly, however, simple ODE models tend to ignore them, which effects the analysis of the studied biological system and makes them impractical to be applied to the real world. There is also increasing evidence of stochastic behavior in critical biological processes, such as gene regulation and cellular behavior. All these factors point to the possible need to account for stochasticity in mathematical models (Lipniacki et al., 2006a,b).

ODE models generally describe population level behaviors. For example, ODE models already exist to describe the pharmacokinetics and pharmacodynamics of all drugs (see Shargel et al., 2007). They are developed during clinical trials, published for all approved drugs and used to

determine safe dosage regimes for the average patient. Therefore, parameterization of these models is generic, typically done in a theoretical manner or based on laboratory or experimental data. However, the critically ill patient's unique parameters can vary considerably from those of the average patient, such population based models often fail to capture specific clinical scenarios.

To describe individuals, model parameters must be re-calibrated using observations of the individual. However, the observed data may be missing or noisy, or it could sparse or infrequent relative to the dynamics of the underlying system thus, making individualisation a challenging task.

Previous research work (Enright and Madden, 2015) in our research group has shown that we can overcome the limitations of ODE models by mapping them to Dynamic Bayesian Networks (DBNs) incorporating a first order Euler solver. DBNs are well suited to handle uncertainty and deal with noisy and missing data. Learning the structure of the DBNs is a challenging task, especially in situations where data is sparse or incomplete. The expert knowledge available in form of ODEs can be utilized as ODEs are encapsulated into the DBN structure. DBNs can reason efficiently with this powerful combination of domain knowledge and real time data. They explicitly model measurement uncertainty and parameter uncertainty, allowing model parameters to be adjusted from initial approximate values to their correct values using real-time evidence. By doing this, the knowledge elicitation bottleneck is bypassed. The technique was previously applied to the problem of modeling glycaemia in patients in an Intensive Care Unit (Enright et al., 2010), producing promising results.

We apply the methodology of encapsulating ODE models into DBNs on three small to medium sized benchmark biological ODE models using the software package PROFET [1] and evaluate the results of ODE model variable inference and model parameter re-estimation.

## 1 DBNs for Biological ODE Models

In order to *individualize* a general biological ODE model to a specific case, we convert it into a DBN using the software application PROFET. The DBN framework explicitly models noise as measurement and parameter uncertainty and then reduces the uncertainty over time by individualizing model parameters using temporal evidence. We apply this approach on three benchmark biological ODE models which were used by Dondelinger et al. (2013) to apply their ODE model parameter inference methods. We evaluate the results of ODE model variable inference and model parameter re-estimation on these ODE models; however our experimental setup is different from that of Dondelinger et al. (2013). The essence of their work is to infer ODE model parameters from multiple noisy time series data. Our aim is to *individualise* the model parameters on a single time series data stream. Therefore, for each ODE model, benchmark data is obtained using the R package deSolve. The values of the model parameters and the initial state of model variables to generate the benchmark data are taken from Dondelinger et al. (2013). Instead of adding noise to the data, we assume that the true values of the ODE model parameters are unknown and must be inferred from the population values. To discover the correct model parameters, evidence (which in these cases are the true values taken from the benchmark solutions) is incorporated. Complete details of methodology to convert ODE models to DBNs incorporating first order Euler solver can be found in Enright (2012). Below we describe each model and explain the results.

---

[1]http://profet.it.nuigalway.ie/

## 1.1  The PIF4/5 Model

As described in Dondelinger et al. (2013), this is a model of gene regulation of genes *PIF4* and *PIF5* by *TOC1* in the ciracadian clock gene regulatory network of *Arabidopsis thaliana*. The complete network is described by a Locke 2-loop model (Locke et al., 2005). Following Dondelinger et al. (2013), the model was simplified, such that, the genes *PIF4* and *PIF5* are combined as *PIF4/5*. The DBN created from the $PIF4/5$ ODE model is shown in Figure 1a. Model parameter $s$ is the promoter strength, $K_d$ is the rate constant, $h$ represents the Hill coefficient of the regulation by $TOC1$ and $d$ is the degradation rate of $PIF4/5$. $PIF4/5$ and $TOC1$ represent the concentration of the genes. The model parameters are modeled as continuous nodes with linear Gaussian distribution. The initial state model distribution of the DBN can be viewed as the distribution of the population. We assume that we do not know the real values of the model parameters but we only have a rough guess of the population parameters. The true values of the model parameters must be inferred from the data. Model parameters are allowed to vary in each time step. Evidence for the observed data is sampled from the benchmark data at different time points. The evidence is deliberately sampled at sparse and irregular intervals. However, it does not contain any noise.

We run standard fixed time step Particle Filtering algorithm (Gordon et al., 1993) to infer the values of model parameters and model variables. Figure 2a shows a comparison of benchmark solution and the predicted solution. It can be seen that even though incorrect values of model parameters were chosen at the outset, the accuracy of the predicted values of $PIF4/5$ concentration begins to improve as evidence is incorporated into the system. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of the predicted data is 0.070 and 0.035 respectively.

## 1.2  The Lotka-Volterra Model Predator Prey Model

The Lotka-Volterra model, also known as the predator-prey model is described by two first order, non linear ODEs. These were first proposed by Edward J. Lotka in 1910 and since then have been frequently used to describe the dynamics of biological systems in which two species interact, one as the predator and the other as the prey. The DBN structure corresponding to the ODE model is shown in Figure 1b. The model variable node $X$ represent the number of prey and $Y$ represents the number of some predator. Model parameters $a, b, c$ and $d$ are the positive real numbers that describe the interaction of the two species. This ODE model exhibits periodicity and the interactions between model variables are non-linear.



**(a)**               **(b)**               **(c)**
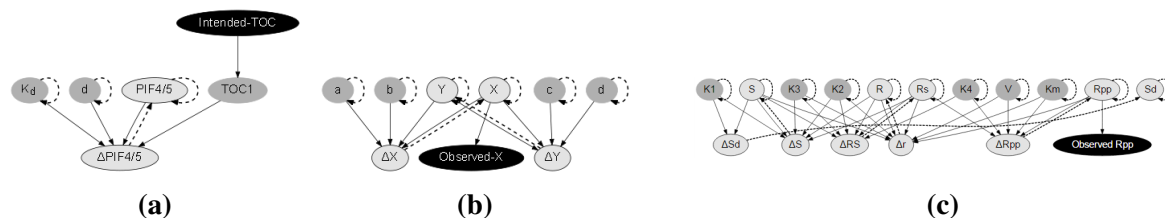
**Figure 1:** (a) DBN structure of the (a) The PIF4/5 Model (b) The Lotka-Volterra model (c) The Signal Ttransduction Cascade Model

We follow the similar steps described in the previous section and run the DBN inference in time interval [0,2]. Sparse evidence is sampled from the benchmark data of the model variable $X$. Graph in Figure 2b shows predicted values of prey population over time. As before, can see that even though incorrect values of model parameter were chosen at the start, the predicted results are

very close to those of benchmark solution. RMSE and MAE of predicted values of $X$ is 0.287 and 0.137 respectively.

## 1.3 The Signal Transduction Cascade Model

The model of signal transduction cascade was described in Vyshemirsky and Girolami (2008). The DBN constructed by PROFET for this ODE model is shown in Figure 1c. The input signal is represented by the concentration of protein $S$ which can bind to protein $R$ to form a complex $RS$ which activates protein $R$ into its phosphorylated form $Rpp$. Protein $Rpp$ can then be de-activated back to protein $R$. The model also defines the degradation of input signal, that is, the conversion of protein $S$ into its degraded form $dS$. This system represents a realistic formulation of signal transduction as a mathematical model using mass action and Michaelis-Menten kinetics.

We run the DBN inference in time interval [0,100]. Following Dondelinger et al. (2013), we sample the evidence of $Rpp$ concentration at more time points during the earlier part of the time series where the dynamics tend to be faster. In Figure 2c, we plot the value of $Rpp$ concentration predicted along with the benchmark solution. We can see that the DBN was able to predict the values of model variable $Rpp$ with high accuracy, even though incorrect model parameters were chosen in the outset. RMSE and MAE of the predicted data is 0.0085 and 0.0053 respectively.



|   (a)   |   (b)   |   (c)   |
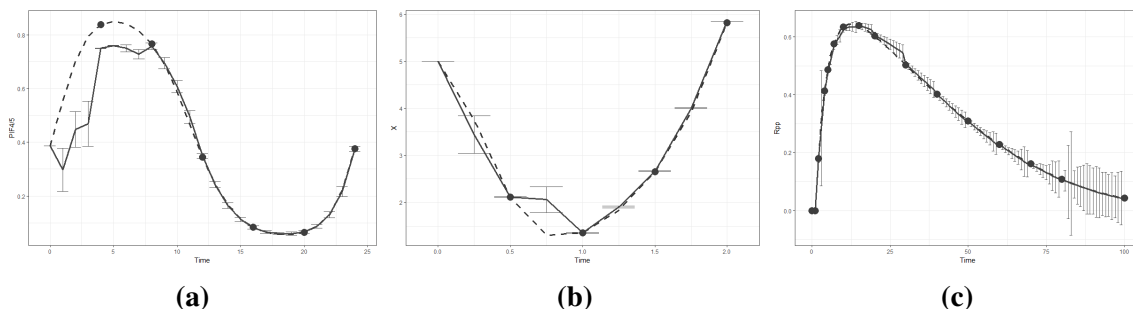
**Figure 2:** Predicted values of model variables in the ODE system. Dashed lines represent the benchmark solution. Solid lines are the predicted trajectories. Error bars show one standard deviation. Gray dots are the points where evidence is received. (a) The PIF4/5 model (b) Lotka-Volterra Model (c) The Signal Transduction Cascade model

## Conclusions

This paper has discussed the challenges of applying ODE modeling approaches to real world biological systems. ODE models are deterministic and they can not account for the uncertainty in the real world. We proposed to tackle this problem by converting the ODE models to DBNs that incorporate a first order Euler solver as proposed by Enright et al. (2013). We tested the methodology of various biological ODE models. We simulated a real-life situation by assuming that the true values of the ODE model parameters are not known and data is sparse, incomplete and collected at irregular intervals. Our DBN framework uses Particle Filtering inference algorithm to infer the values of model variables of the ODE model by re-estimating the values of model parameters at each time step. We have shown that our DBN modeling and inference works well for this task.

# References

Susanne Ditlevsen and Adeline Samson. Introduction to stochastic models in biology. In *Stochastic biomathematical models*, pages 3–35. Springer, 2013.

Frank Dondelinger, Dirk Husmeier, Simon Rogers, and Maurizio Filippone. ODE parameter inference using adaptive gradient matching with Gaussian processes. In *AISTATS*, 2013.

Catherine Enright. *A probabilistic framework based on mathematical models with application to medical data streams*. PhD thesis, 2012.

Catherine Enright, Michael G Madden, Norm Aleks Stuart Russell, Geoffrey Manley, John Laffey, Brian Harte, Anne Mulvey, and Niall Madden. Modelling glycaemia in ICU patients a dynamic Bayesian network approach. In *Proceedings of BIOSIGNALS-2010, 3rd International Joint Conference on Biomedical Engineering Systems and Technologies*, Valencia, 2010.

Catherine G Enright and Michael G Madden. Modelling and monitoring the individual patient in real time. In P.F. Lucas (eds.) A. Hommersom, editor, *Foundations of Biomedical Knowledge Representation*, pages 107–136. 2015.

Catherine G Enright, Michael G Madden, and Niall Madden. Bayesian networks for mathematical models: techniques for automatic construction and efficient inference. *International Journal of Approximate Reasoning*, 54(2), 2013.

Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F-Radar and Signal Processing*, volume 140. IET, 1993.

Tomasz Lipniacki, Pawel Paszek, Allan R Brasier, Bruce A Luxon, and Marek Kimmel. Stochastic regulation in early immune response. *Biophysical journal*, 90:725–742, 2006a.

Tomasz Lipniacki, Pawel Paszek, Anna Marciniak-Czochra, Allan R Brasier, and Marek Kimmel. Transcriptional stochasticity in gene expression. *Journal of theoretical biology*, 238(2):348–367, 2006b.

James CW Locke, Megan M Southern, László Kozma-Bognár, Victoria Hibberd, Paul E Brown, Matthew S Turner, and Andrew J Millar. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular systems biology*, 1(1), 2005.

Alfred J Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14(3): 271–274, 1910.

Thomas Malthus. An essay on the principle of population. 1798.

Johnny T Ottesen, Mette S Olufsen, and Jesper K Larsen. *Applied mathematical models in human physiology*. Siam, 2004.

Leon Shargel, Susanna Wu-Pong, and Andrew BC Yu. *Applied biopharmaceutics & pharmacokinetics*. McGraw-Hill, 2007.

PF Verhulst. Mathematical researches into the law of population growth increase. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18:1–42, 1845.

Vladislav Vyshemirsky and Mark A Girolami. Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6), 2008.