How data assimilation helps to illuminate complex biology

The dynamic elastic-net

Maik Kschischo

Department of Mathematics and Technology
University of Applied Sciences Koblenz
RheinAhrCampus Remagen
Joseph-Rovan-Allee 2
53424 Remagen
Email: kschischo@rheinahrcampus.de

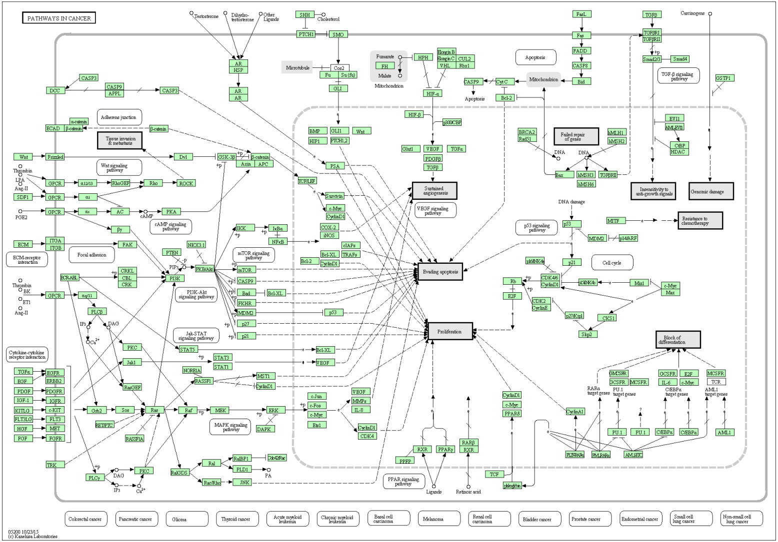
RISDA 2017

Overview

1. Some challenges for modelling in molecular systems biology
2. The dynamic elastic net: An algorithm for state and model error estimation in imperfect models
3. Test examples
4. Outlook

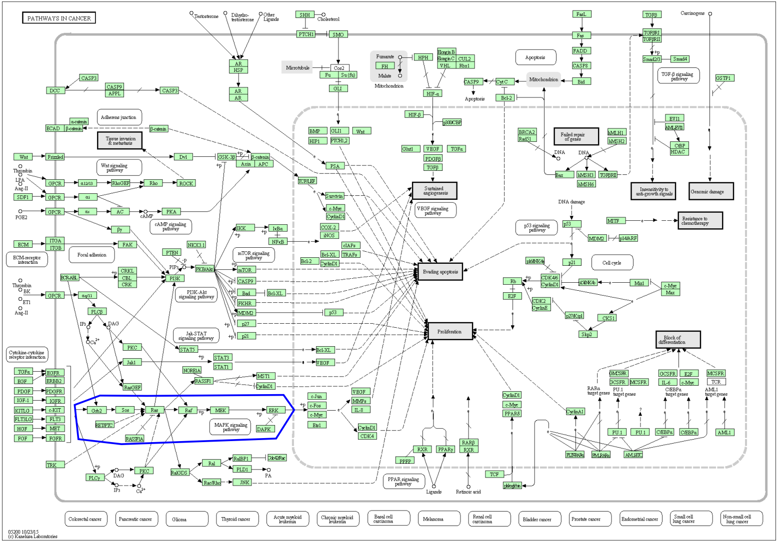
How to model a complex dynamical reaction system?

Pathways in Cancer (KEGG hsa05200, 10/23/25, Kaneshisa Lab)



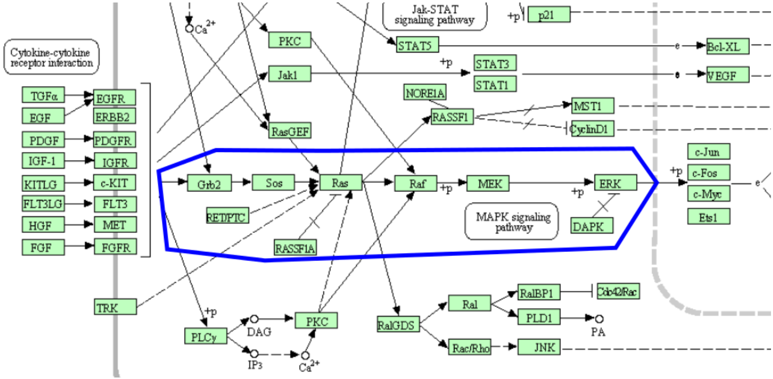
Modelling subnetworks

Pathways in Cancer (KEGG hsa05200, 10/23/25, Kaneshisa Lab)



Crosstalk and external feedback loops

Pathways in Cancer (KEGG hsa05200, 10/23/25, Kaneshisa Lab)



Challenges for Modelling

Kahm *et al.* , PLoS Comput Biol 8, 2012;
Tsigkinopoulou *et al.*, Trends in Biotechnology, 2016

We have only incomplete or uncertain information about the

- Reaction network
- Reaction rate functions (dynamic laws of the reactions)
- Parameter values

Data: Molecular concentration measurements over time (e.g. protein)

- Not all the molecular substances can directly be measured
- Measurements are often very noisy
- Some proteins can not be identified

Challenges for Modelling

Kahm *et al.* , PLoS Comput Biol 8, 2012;
Tsigkinopoulou *et al.*, Trends in Biotechnology, 2016

We have only incomplete or uncertain information about the

- Reaction network
- Reaction rate functions (dynamic laws of the reactions)
- Parameter values

Data: Molecular concentration measurements over time (e.g. protein)

- Not all the molecular substances can directly be measured
- Measurements are often very noisy
- Some proteins can not be identified

Methods for modelling under uncertain and incomplete information are needed!

Outline

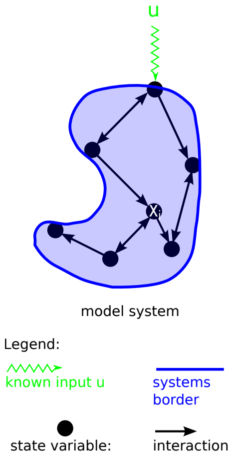
- State vector: $\boldsymbol{x} = (x_1, \ldots, x_n)^T$
- Output (measured): $\boldsymbol{y} = (y_1, \ldots, y_m)^T$

Ordinary differential equation (ODE) model

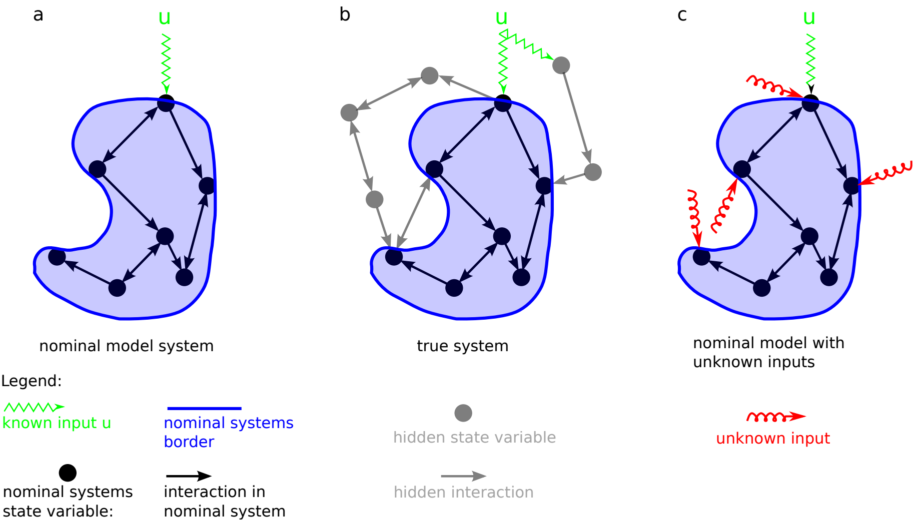
$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \\ \boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x}) \end{aligned}$$

Influence graph

- Each state variable x_i is a vertex
- Draw a directed edge $x_i \rightarrow x_j$ iff $\frac{\partial f_j}{\partial x_i} \neq 0$



The nominal model and the true system



Defining the model error

Ordinary differential equation (ODE) models

The nominal model (the draft)

$$\begin{aligned}\dot{\tilde{\boldsymbol{x}}} &= \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) \\ \tilde{\boldsymbol{y}} &= \boldsymbol{h}(\tilde{\boldsymbol{x}})\end{aligned}$$

The nominal model (the draft)

$$
\begin{aligned}
\dot{\tilde{\boldsymbol{x}}} &= \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}, \boldsymbol{u}) \\
\tilde{\boldsymbol{y}} &= \boldsymbol{h}(\tilde{\boldsymbol{x}})
\end{aligned}
$$

True model (with unkown error)

$$
\begin{aligned}
\dot{\boldsymbol{x}} &= \tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{w} \\
\boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x})
\end{aligned}
$$

The nominal model (the draft)

$$\begin{aligned}\dot{\tilde{\boldsymbol{x}}} &= \tilde{\boldsymbol{f}}(\tilde{\boldsymbol{x}}, \boldsymbol{u})\\ \tilde{\boldsymbol{y}} &= \boldsymbol{h}(\tilde{\boldsymbol{x}})\end{aligned}$$

True model (with unkown error)

$$\begin{aligned}\dot{\boldsymbol{x}} &= \tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{w}\\ \boldsymbol{y} &= \boldsymbol{h}(\boldsymbol{x})\end{aligned}$$

Model error = hidden input to nominal model

$$\boldsymbol{w}(t) = \dot{\boldsymbol{x}}(t) - \tilde{\boldsymbol{f}}(\boldsymbol{x}(t), \boldsymbol{u}(t))$$

Note: $\boldsymbol{x}(t)$ is the trajectory of the true system.

Data assimilation problem with structural model error

What we have:

- The nominal model (a model draft)
- Data: $\boldsymbol{y}(t_k)$ for some time points $t_k,\ k = 1, \dots, N$

Data assimilation problem with structural model error

What we have:

- The nominal model (a model draft)
- Data: $\boldsymbol{y}(t_k)$ for some time points t_k, $k = 1, \ldots, N$

What we want to know:

1. Where is the nominal model wrong?
   - Set of target nodes $\mathcal{W} = \{i \in \{1, \ldots, n\} | w_i \not\equiv 0\}$
2. The dynamics of the model error $\boldsymbol{w}(t)$
3. The unknown state trajectory $\boldsymbol{x}(t)$ of the true system

Data assimilation problem with structural model error

What we have:

- The nominal model (a model draft)
- Data: $\boldsymbol{y}(t_k)$ for some time points $t_k,\ k = 1, \ldots, N$

What we want to know:

1. Where is the nominal model wrong?
   - Set of target nodes $\mathcal{W} = \{i \in \{1, \ldots, n\} | w_i \not\equiv 0\}$
2. The dynamics of the model error $\boldsymbol{w}(t)$
3. The unknown state trajectory $\boldsymbol{x}(t)$ of the true system

This is an ill posed problem!

- Different models/model errors could explain the data equally well $\Longrightarrow$ (Non-)observability of the model error

The model error signal $\boldsymbol{w}(t)$

1. is a smooth function
2. a sparse vector
    - The set of target nodes $\mathcal{W} = \{i \in \{1, \ldots, n\} | w_i \not\equiv 0\}$ is small
    - Most parsimonious explanation for observed discrepancy between the nominal model output and the data

Estimating the model error using the Dynamic Elastic-Net

Which simple error signal could fit the data?

Data: Measure $\boldsymbol{y}(t_k)$ for some time points t_k, $k = 1, \ldots, N$

Idea:

- Build an observer system (copy of the true system):

$$\begin{aligned} \dot{\hat{\boldsymbol{x}}} &= \tilde{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t), \boldsymbol{u}(t)) + \hat{\boldsymbol{w}}(t) \\ \hat{\boldsymbol{y}}(t) &= \boldsymbol{h}(\hat{\boldsymbol{x}}(t)) \end{aligned}$$

- Infer the "simplest" $\hat{\boldsymbol{w}}(t)$ that explains the data:

$$\min_{\hat{\boldsymbol{w}}(t)} \left\{ \sum_{k=1}^{N} \|\boldsymbol{y}(t_k) - \hat{\boldsymbol{y}}(t_k)\|^2_{Q(t_k)} + \mathcal{R}[\hat{\boldsymbol{w}}] \right\}$$

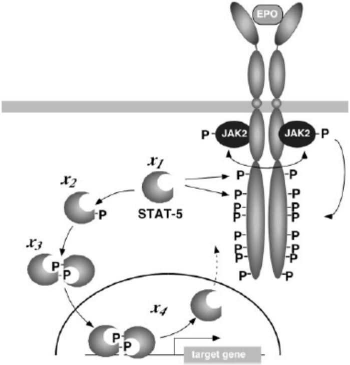
- L_1 and L_2 regularisation (dynamic elastic-net) induces **sparsity** and **smoothness**

$$\mathcal{R}[\hat{\boldsymbol{w}}] = \alpha_1 \|\hat{\boldsymbol{w}}(t)\|_1 + \frac{\alpha_2}{2} \|\hat{\boldsymbol{w}}(t)\|_2^2$$

Outline

Testing the algorithm

Example 1: The Jak-Stat-model from Swameye *et al.*, PNAS, 2001.



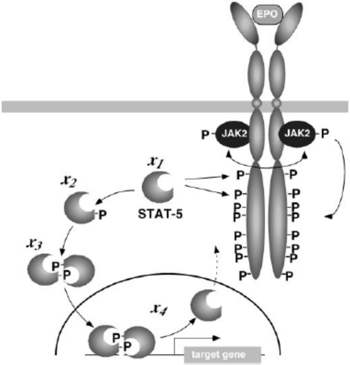
Swameye *et al.*, *PNAS*, 100, 2001, Figure 1 on page 1029.

Nominal Model of Swameye *et al.*:

- Epo binds the EpoR (receptor)
- Cytoplasmatic STAT5 ist activated
- Phosphorylated STAT5 dimerises
- Export to the nucleous

Testing the algorithm

Example 1: The Jak-Stat-model from Swameye *et al.*, PNAS, 2001.



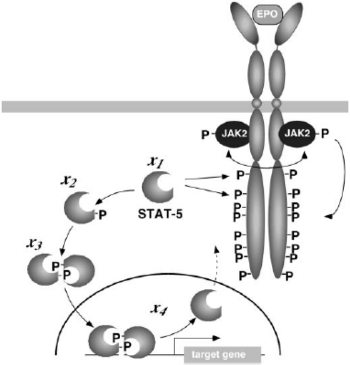
Swameye *et al.*, *PNAS*, 100, 2001, Figure 1 on page 1029.

Nominal Model of Swameye *et al.*:

- Epo binds the EpoR (receptor)
- Cytoplasmatic STAT5 ist activated
- Phosphorylated STAT5 dimerises
- Export to the nucleous

Missing reaction:

- Recycling of nuclear STAT (Raue *et al.*, *Bioinformatics*, 2009.)



Swameye *et al.*, *PNAS*, 100, 2001, Figure 1 on page 1029.

Nominal Model of Swameye *et al.*:

- Epo binds the EpoR (receptor)
- Cytoplasmatic STAT5 ist activated
- Phosphorylated STAT5 dimerises
- Export to the nucleous

Missing reaction:

- Recycling of nuclear STAT (Raue *et al.*, *Bioinformatics*, 2009.)

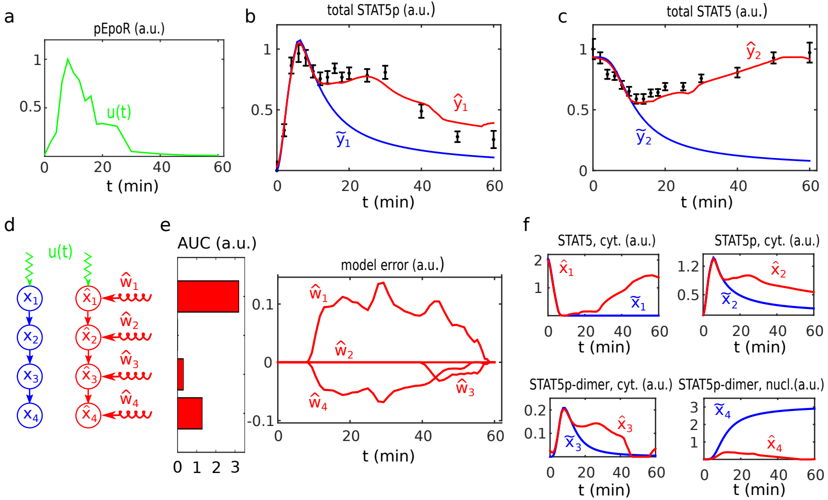
Could this have been learned from the data?

- Apply the dynamic elastic-net

The Jak-Stat-System

Detecting nucleocytoplasmatic cycling (data points from Swameye *et al.*, 2001.)

Legend: known input, data, nominal model, dynamic elastic-net

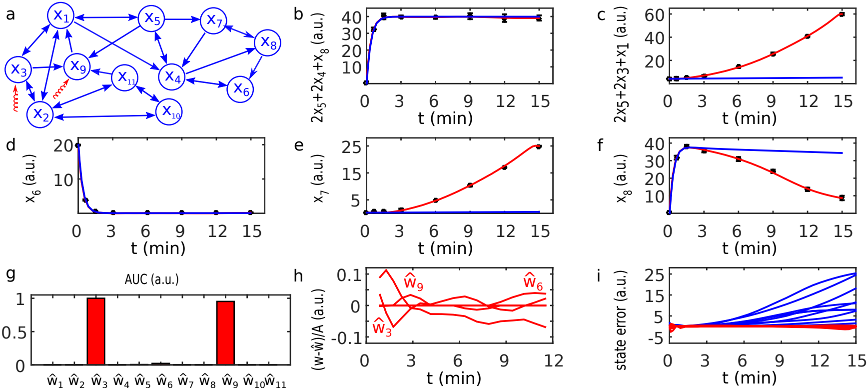


Testing the algorithm

Example 2: The UVB-system

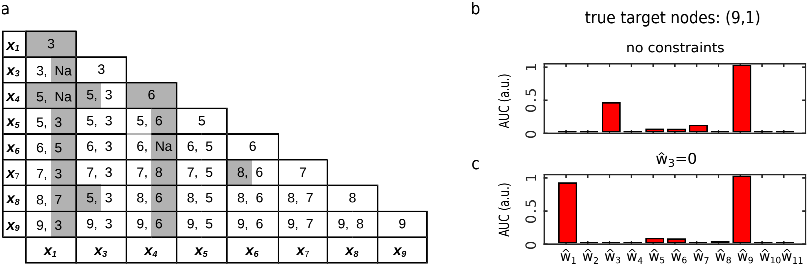
- A model for photomorphogenic UV-B signalling in plants *(Ouyang, X. et al. Proc. Natl. Acad. Sci. USA , 2014)*
- 11 protein concentrations $\boldsymbol{x}(t)$ participating in 10 reactions
- The system was perturbed by a known model error
- Can we recover this error $\boldsymbol{w}(t)$ and the true state $\boldsymbol{x}(t)$?

Legend: data, nominal model, dynamic elastic-net



Testing the limitations

All single and pairwise combinations in the UVB-system



- Some model errors are non-observable
- Simple algorithm to explore suboptimal solutions (next slide)
- The true target nodes are found amongst the highest ranking suboptimal solutions

Exploring suboptimal solutions

A heuristic algorithm

1. Init $B = I_n$ (identity matrix)
2. Solve the DEN using the observer system

$$\dot{\hat{\boldsymbol{x}}} = \tilde{\boldsymbol{f}}(\hat{\boldsymbol{x}}(t), \boldsymbol{u}(t)) + B\hat{\boldsymbol{w}}(t)$$

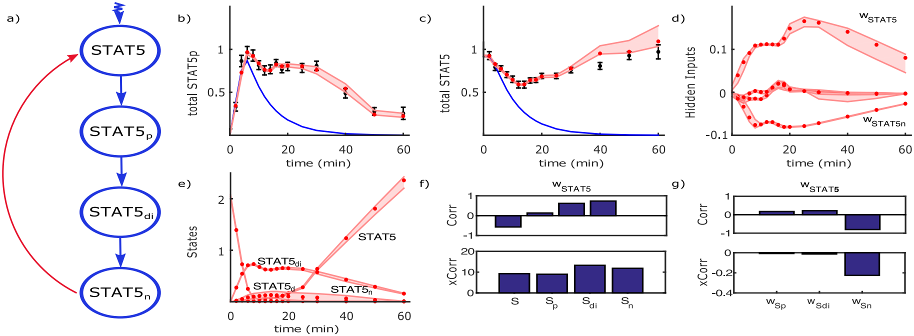
   and estimate $\hat{\boldsymbol{w}}(t)$ and $\hat{\mathcal{W}} = \{k | \hat{w}_k(t) \not\equiv 0\}$
3. If the squared error $\sum_{k=1}^{N} \|\boldsymbol{y}(t_k) - \hat{\boldsymbol{y}}(t_k)\|^2_{Q(t_k)}$ is too big, then STOP.
4. Find the index $k \in \hat{\mathcal{W}}$ of the "smallest" nonzero signal $\hat{w}_k(t)$ (as measured by L_1 norm)
5. Delete the k-th column in B and the corresponding row of $\hat{\boldsymbol{w}}$ and GOTO 2.

Result: A set of suboptimal solutions of the DEN providing candidate model errors and state estimates.

Summary and Outlook

- **Dynamic elastic-net**
  - Estimates the nodes in a network, where the model could be wrong and
  - Estimates the true system state trajectory, even for incomplete or incorrect models
- **Non-observable model error**
  - No unique solution possible
  - Greedy approach to explore suboptimal solutions
- **Further research**
  - Network properties and model error observability
  - Optimal experimental design to improve observability of model error
  - Assign uncertainty/confidence (Engelhardt *et al.*, submitted)

JAK-STAT signalling pathway (Swameye et al., 2003)



Uncertainty quantification for state and model error estimates (Engelhardt *et al.*, submitted)

Acknowledgement

Thanks for the great collaboration

Holger Fröhlich[1,2] **Benjamin Engelhardt**[1]

(1) Rheinische Friedrich-Wilhelms-Universität Bonn, Institute for Computer Science, Algorithmic Bioinformatics, c/o Bonn-Aachen International Center for IT

(2) UCB Biosciences GmbH

For some details:

Engelhardt, B., Fröhlich, H. and Kschischo, M. 2016. Learning (from) the errors of a systems biology model. Scientific Reports 6:20772.