Data Assimilation for Systems and Mathematical Biology

Benjamin Engelhardt [1,2], Dominik Kahl[3], Andreas Weber[4], Maik Kschischo[3,*]

March 27, 2018

Abstract

Mathematical models are increasingly used as a tool to deal with the tremendous complexity of biological systems. Data Assimilation, defined as the process of combining models with experimental observations, is a key step in order to better align the model outputs with reality. Despite new and improved experimental techniques, it is usually impossible to directly observe all the states of a biological system. This renders Data Assimilation an inverse problem requiring sophisticated mathematical and statistical techniques and the systematic integration of prior knowledge or assumptions.

In this article, we will highlight how Data Assimilation is key to dealing with incomplete and uncertain biological information. We will survey the basic concepts and methodological approaches relevant to systems biology and indicate similarities and differences with Data Assimilation problems faced in the environmental and geosciences. In addition, we will illustrate the Dynamic Elastic-Net, a recent data driven approach to dealing with incomplete models and structural model errors.

1 Rheinische Friedrich-Wilhelms-Universität Bonn, Algorithmic Bioinformatics, Bonn, Germany
2 Current Address: AbbVie Deutschland GmbH & Co. KG, Ludwigshafen, Germany
3 University of Applied Sciences Koblenz, RheinAhrCampus, Department of Mathematics and Technology, Remagen, Germany
4 Rheinische Friedrich-Wilhelms-Universität Bonn, Institute for Computer Science, Bonn, Germany
*Corresponding author: kschischo@rheinahrcampus.de

1 Introduction

A central goal of the life sciences is to understand, predict or manipulate the dynamics of biological systems. For example, ecologists record the number of individuals of a group or species in a certain area to reveal the inner workings of dynamical ecosystems. Molecular and cell biologists measure the abundance of proteins and other gene products as a function of time in order to better characterise molecular pathways. The central theme of mathematical and computational systems biology is to combine these data with mathematical models and to use computer simulations to better understand the system, to predict its behaviour after certain interventions and to design controls to manipulate the system in a desired way.

One of the greatest challenges of these combined modelling and experimentation endeavours is Data Assimilation (DA), defined here as the process of combining the mathematical model with the observed data. An important DA problem is the estimation of dynamic state variable in the rather typical situation that not all the variables of a state space model can directly be experimentally accessed. This state observerer problem is well known in control engineering and a great part of research has been devoted to deriving conditions under which the state can be reconstructed from the outputs (observability) and to design state observers, i.e. artificial dynamical systems synchronising their state to the dynamics of the true system using only the available output data [TOCITE].

Measurement noise makes the state estimation problem even harder. As in other areas of science and engineering, measurement noise originates from shortcomings of the measurement

device or the experimental procedures. If the system states itself can be considered to be deterministic, the state estimation problem reduces to the estimation of parameters and initial conditions. The challenges of parameter estimation have extensively been studied in the context of systems biology [TOCITE].

In Fig.1 we illustrate the state estimation problem using a deterministic dynamical model for the regulation of the Epo receptor [1]. At the moment, we will refrain from discussing the model in detail and just focus on the principled state estimation problem. The model has six state variables $\boldsymbol{x} = (x_1, \ldots, x_6)^T$ and their dynamics $\boldsymbol{x}(t)$ is described by a system of coupled ordinary differential equations (ODEs). These interactions can be displayed graphically, see section 2.1 for details. The state variables represent the concentrations of different biomolecules, which can not directly be measured. Instead, the measurable outputs $\boldsymbol{y} = (y_1,\ y_2, y_3)^T$ of the system are scaled linear combinations of the states. The DA task involves the estimation of the parameters and the initial state from noisy time course observations $\boldsymbol{y}(t_k)$ at certain measurement times t_k. Given that the system is deterministic and has a unique solution, the parameters and the initial condition provide full information about the dynamic state $\boldsymbol{x}(t)$.

A fundamentally different source of noise in living systems is related to biological diversity. No two apparently alike cells in an organ or no two individuals of the same species are in fact completely identical or react in exactly the same way to an environmental stimulus. This variation is the basis for natural selection, but it also poses additional challenges for modelling and DA. In state space models, the diversity can be represented by stochastic dynamical sytems, where the state variables are targeted by noise. This is sometimes called systems noise. Other sources of systems noise can be environmental fluctuations not encompassed by the model. Thus, both intrinsic and extrinsic noise processes lead to stochastic dynamics and the task of state estimation is then expressed as the problem of inferring the probability distribution of the states given the observed data.

The majority of DA algorithms is derived from the assumption that the governing equations of the system are known. However, in biology we have only partial information about all the interations and processes in a real living system and the decision of which terms in a model can be neglected is often not easy to make. In addition, biological systems are open and exchange matter, energy and information with their environment. Thus, the perfect model assumption is usually unrealistic and structural model errors can not be represented by zero mean noise processes. DA algorithms dealing with structural model error are still in its infancy, but progress to deal with model errors and incomplete information is urgently required.

In our view, DA is a key technique in all areas of quantitative biology, although the most recent methodological progress has been largely driven by applications in other fields, in particular the atmospheric and oceanographic sciences. Our motivation is to review the most important concepts of DA in the context of quantitative biology and to introduce researchers in the life sciences to recent developments and open research questions relevant to their modelling work.

2 The Data Assimilation problem

2.1 Continous time state space models

We assume that the state of a biological system can be characterised by a vector of state variables $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))^T$ at time t. One example is the state of a biochemical reaction network, which is is given by the concentrations x_i of all the reacting substances i. Other examples are the state of a cellular population, which can be characterised by the abundance x_i of each cell type i or the number of individuals x_i of a species i in a population dynamics model. For continuous time $t \in \mathbb{R}_{\geq 0}$, the dynamical model is formulated as a system pf

ordinary differential equations

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \boldsymbol{f}\left(\boldsymbol{x}(t), \boldsymbol{u}(t)\right) && \text{(1a)} \\ \boldsymbol{y}(t) &= \boldsymbol{h}\left(\boldsymbol{x}(t)\right). && \text{(1b)} \\ \boldsymbol{x}(t_0) &= \boldsymbol{x}_0, && \text{(1c)} \end{aligned}$$

where the function $\boldsymbol{f} : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^n$ with $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{U} \subset \mathbb{R}^q$ encodes the effect of the interacting state variables and the known input $\boldsymbol{u}(t) \in \mathcal{U}$ on the time derivative $\dot{\boldsymbol{x}}(t)$. Note, that one can also arrive at such an ordinary differential equation model (1) from spatio-temporal models by discretising in space. Typically, it is not possible to obtain direct measurements for all the state variables. Instead, one observes the output variables $\boldsymbol{y}(t) = (y_1(t), \ldots, y_m(t))^T$, which are related to the states in (1) via the measurement function $\boldsymbol{h}$. Usually, we will set $t_0 = 0$ in the inital condition (1c).

To illustrate different aspects of DA we will use a model for the information processing at the erythropoietin (Epo) receptor (EpoR) as a running example [1]. This model is illustrated in Fig. 1 together with its influence graph, which has a node for each state variable x_i and an directed edge from x_i to x_j whenever $\frac{\partial f_j}{\partial x_i} \neq 0$. In addition, the effect of the state to the outputs is visualised by drawing a dashed arrow from x_i to y_l, if $\frac{\partial g_l}{\partial x_i} \neq 0$.

2.2 The Filtering, smoothing and prediction tasks

In reality, one observes the output $\boldsymbol{y}$ of a state space model at discrete measurement time points $t_1 < t_2 < \ldots < t_T$. The observations $\boldsymbol{y}^o(t_k)$ are in addition corrupted by measurement noise $\boldsymbol{v}(t_k)$ and thus

$$\boldsymbol{y}^o(t_k) = \boldsymbol{h}\left(\boldsymbol{x}(t_k)\right) + \boldsymbol{v}(t_k) \qquad (2)$$

describes the sequence of measurements, cf. (1b). We assume that the measurements devices are unbiased and thus the noise $\boldsymbol{v}(t_k)$ is a stochastic process with expectation zero: $E\left(\boldsymbol{v}(t_k)\right) = \boldsymbol{0}_{\mathbb{R}^m}$ for all times $k = 1, 2, \ldots, n$.

Assume that we have observed the output of our dynamical system and collected an $m \times T$ data matrix $\boldsymbol{y}^o_{1:T} = (\boldsymbol{y}^o(t_1), \boldsymbol{y}^o(t_2), \ldots, \boldsymbol{y}^o(t_T)) \in \mathbb{R}^{m \times T}$. The goal of state estimation is to combine the data with the model (1) in order to infer the state $\boldsymbol{x}(t)$ at time t. Depending on t, the task of estimating the state $\boldsymbol{x}(t)$ given the data $\boldsymbol{y}^o_{1:T}$ is called

(a) *smoothing* if $t \in [t_1, t_T]$

(b) *filtering* if $t = t_T$

(c) *prediction* if $t > t_T$

Thus, smoothing is an offline task, where we use all the observations to infer the state, whereas flitering refers to the online task of estimating the state immediately after observing $\boldsymbol{y}(t_T)$. Predictions are estimates of the future behaviour beyond the data assimilation window $[t_1, t_T]$.

3 Discrete time state space model

The continous-time dynamical system (1a) can can also be considered at discrete time points via the map

$$\begin{aligned} \phi_s(\boldsymbol{x}(t_0)) =& \boldsymbol{x}(t_0 + s) \qquad (3) \\ =& \boldsymbol{x}(t_0) + \int_{t_0}^{t_0+s} \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t))dt. \end{aligned}$$

The resulting discrete-time dynamic system

$$\boldsymbol{x}_{k+1} = \boldsymbol{\phi}_t\left(\boldsymbol{x}_k\right), \qquad (4)$$

describes the dynamics of $\boldsymbol{x}_k = \boldsymbol{x}(ks)$. Here, we have suppresssed the known input $\boldsymbol{u}$ in the notation. An advantage of the discrete-time description is that the technicalities of stochastic differential equations can be avoided, when we consider systems with state noise. Thus, in the following derivations we will consider the general model

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{\phi}_t\left(\boldsymbol{x}_k\right) + \boldsymbol{w}_k && \text{(5a)} \\ \boldsymbol{y}_k &= \boldsymbol{h}\left(\boldsymbol{x}_k(t)\right) + \boldsymbol{v}_k && \text{(5b)} \end{aligned}$$

where $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ describe the state and observation noise, respectively. These stochastic processes are assumed to be independent and

identically distributed (i.i.d) sequences of random variables densities p_w and p_v. We use the notation

$$\boldsymbol{w}_k \sim p_w \quad \text{and} \quad \boldsymbol{v}_k \sim p_v \quad \forall k \in \mathbb{N}_0. \tag{5c}$$

We assume zero expectation $E(w_k) = 0$ and $E(v_k) = 0$. In addition, the initial state $\boldsymbol{x}_0$ is often unknown and thus considered a random variable $\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0)$.
The stochastic dynamics (5) can be reformulated as a probabilistic state space model. The probability density $p(\boldsymbol{x}_{0:T})$ of a state sequence $\boldsymbol{x}_{0:T} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ and the conditional density of an observation sequence $\boldsymbol{y}_{1:T} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$ given the state sequence can be recursively computed as

$$p(\boldsymbol{x}_{0:T}) = p(\boldsymbol{x}_0) \prod_{k=1}^{T} p(\boldsymbol{x}_k \,|\boldsymbol{x}_{k-1}) \tag{6}$$

$$p(\boldsymbol{y}_{1:T} \,|\boldsymbol{x}_{0:t}) = \prod_{k=1}^{T} p(\boldsymbol{y}_k \,|\boldsymbol{x}_k). \tag{7}$$

Here, (6) indicates that the state space model (5a) defines a Markov process of first order, where the state $\boldsymbol{x}_k$ at time step k depends only on the previous state $\boldsymbol{x}_{k-1}$. The state transition probabilities are given as

$$p(\boldsymbol{x}_{k+1} \,|\boldsymbol{x}_k) = p_w\left(\boldsymbol{x}_{k+1} - \boldsymbol{\phi}_t(\boldsymbol{x}_k)\right). \tag{8}$$

The product in (7) expresses the conditional independence of the outputs $\boldsymbol{y}_k$ at different time points, given the state $\boldsymbol{x}_k$. By comparison with (5b) we have

$$p(\boldsymbol{y}_k \,|\boldsymbol{x}_k) = p_v\left(\boldsymbol{y}_k - \boldsymbol{h}(\boldsymbol{x}_k)\right). \tag{9}$$

In summary, we see that the discrete time state space model (5) together with specifications of the noise distributions (5c) is equivalent to specyfying a state transition probability $p(\boldsymbol{x}_{k+1} \,|\boldsymbol{x}_k)$ and an output distribution $p(\boldsymbol{y}_k \,|\boldsymbol{x}_k)$. Readers familiar with sequence analysis (TOCITE-DURBIN) recognise the structure of a Hidden Markov Model (HMM) with hidden state $\boldsymbol{x}_k$ and observations $\boldsymbol{y}_k$. However, here we consider continous state spaces $\mathcal{X}$ instead of the discrete state spaces used for protein or nucleotide sequence analysis.

4 The Smoothing and the filtering distributions

For DA, we are want to infer the states given the observations. The posterior distribution of the states follows from Bayes' theorem as

$$p(\boldsymbol{x}_{0:t} \,|\boldsymbol{y}_{1:T}) = \frac{p(\boldsymbol{y}_{1:T} \,|\boldsymbol{x}_{0:t})\, p(\boldsymbol{x}_{0:T})}{p(\boldsymbol{y}_{1:T})} \tag{10}$$

$$= \frac{p(\boldsymbol{x}_0) \prod_{k=1}^{T} p(\boldsymbol{x}_k \,|\boldsymbol{x}_{k-1}) p(\boldsymbol{y}_k \,|\boldsymbol{x}_k)}{p(\boldsymbol{y}_{1:T})} \tag{11}$$

with the marginal output distribution

$$p(\boldsymbol{y}_{1:T}) = \int p(\boldsymbol{y}_{1:T} \,|\boldsymbol{x}_{0:t})\, p(\boldsymbol{x}_{0:T})\, d\boldsymbol{x}_0 \ldots d\boldsymbol{x}_T. \tag{12}$$

If we insert the observed data $\boldsymbol{y}_{1:T} = \boldsymbol{y}^o_{1:T}$ into the posterior smoothing distribution, then the posterior smoothing distribution (10) provides maximum information about the states. Inferring this distribution is called the *smoothing problem*. As we will see, the integral in (12) is often difficult or impossible to calculate exactly. Smoothing algorithms approximate the smoothing distribution or certain of its statistical moments (10).
Filtering refers to the online task of estimating $\boldsymbol{x}_k$ from data $\boldsymbol{y}^o_{1:k}$ accumulated up to time k. Thus, we are interested in the filtering distribution $p(\boldsymbol{x}_k \,|\boldsymbol{y}_{1:k})$ for time steps $k = 1, 2, \ldots, T$. This filtering distribution is updated from $p(\boldsymbol{x}_k \,|\boldsymbol{y}_{1:k})$ to $p\left(\boldsymbol{x}_{k+1} \,\middle|\boldsymbol{y}_{1:k+1}\right)$ via two steps, *prediction* and *analysis*. Prediction refers to the the mapping from $p(\boldsymbol{x}_k \,|\boldsymbol{y}_{1:k})$ to $p(\boldsymbol{x}_{k+1} \,|\boldsymbol{y}_{1:k})$, i.e. prediction of the next state given the output data up to time point k. The corresponding distribution $p(\boldsymbol{x}_{k+1} \,|\boldsymbol{y}_{1:k})$ is often called the *background* distribution. Analysis considers the mapping from the background $p(\boldsymbol{x}_{k+1} \,|\boldsymbol{y}_{1:k})$ to the *analysis* distribution $p\left(\boldsymbol{x}_{k+1} \,\middle|\boldsymbol{y}_{1:k+1}\right)$ to incorporate the next observation $\boldsymbol{y}_{k+1}$ into the state estimate.
Prediction: $p(\boldsymbol{x}_k \,|\boldsymbol{y}_{1:k}) \to p(\boldsymbol{x}_{k+1} \,|\boldsymbol{y}_{1:k})$
The prediction step is based on the Chapman-Kolmogorov equation

$$p(\boldsymbol{x}_{k+1} \,|\boldsymbol{y}_{1:k}) = \int_{\mathcal{X}} p(\boldsymbol{x}_{k+1} \,|\boldsymbol{x}_k)\, p(\boldsymbol{x}_k \,|\boldsymbol{y}_{1:k})\, d\boldsymbol{x}_k\,. \tag{13}$$

Here, we have used $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{y}_k) = p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$, which means that the information about the perfect state $\boldsymbol{x}_k$ cannot be improved by the noisy measurement $\boldsymbol{y}_k$. In the special case of a deterministic state space model, i.e. in the absence of process noise ($\boldsymbol{w}_k = 0$ for all $k \in \mathbb{N}_0$ in (5a)), the state transition probabilty density degenerates to a delta distribution $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) = \delta(\boldsymbol{x}_{k+1} - \boldsymbol{\phi}_t(\boldsymbol{x}_k))$ and the prediction step correponds to a single iteration of the deterministic discrete time state space model.

Analysis: $p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k}) \rightarrow p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1})$ The analysis step is based on an application of Bayes' theorem

$$\begin{aligned} &p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k+1}) \\ &= p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{k+1}\,\boldsymbol{y}_{1:k}) \\ &= \frac{p(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}\,\boldsymbol{y}_{1:k})\; p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})}{p(\boldsymbol{y}_{k+1}|\boldsymbol{y}_{1:k})} \\ &= \frac{p(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1})\; p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{1:k})}{p(\boldsymbol{y}_{k+1}|\boldsymbol{y}_{1:k})}, \end{aligned} \tag{14}$$

where in the last line we have used once again the conditional indepence of the observation from previous observations given the current state. Together, the two equations (13, 14) are approximated by different filtering algorithms to recursively estimate the filtering distribution $p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. This recursive approach is often called *sequential data assimilation*.

5 Variational smoothing and filtering

Inferring the smoothing and filtering posterior distributions can be challenging and computationally expensive. In some cases, it is sufficient to summarise these distributions by point estimates. Variational methods for smoothing and filtering seek a maximum a posteriory estimator (MAP estimator), i.e. they locate peaks of the smoothing or filtering distribution, respectively, by solving an optimisation problem.

The posterior smoothing distribution (10) can be rewritten as

$$\begin{aligned} p(\boldsymbol{x}_{0:t}|\boldsymbol{y}_{1:T}) &= e^{-\Psi(\boldsymbol{x}_{0:t}|\boldsymbol{y}_{1:T})} \\ &\propto e^{-E_0(\boldsymbol{x}_0) + \sum_{k=1}^{T}\left[E(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) + U(\boldsymbol{y}_k|\boldsymbol{x}_k)\right]}, \end{aligned} \tag{15}$$

where we have assumed that the state transition and the output probabilities are nonzero everywhere on their domain and

$$\begin{aligned} p(\boldsymbol{x}_0) &\propto \exp(-E_0(\boldsymbol{x}_0)) \\ p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) &\propto \exp(-E(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})) \\ p(\boldsymbol{y}_k|\boldsymbol{x}_k) &\propto \exp(-U(\boldsymbol{y}_k|\boldsymbol{x}_k)). \end{aligned} \tag{16}$$

The constant term originating from the denominator in (10) does not depend on the state sequence and is omitted in the proportionality (15). The MAP estimator for the smoothing problem is obtained from minimising $\Psi(\boldsymbol{x}_{0:t}|\boldsymbol{y}^o_{1:T})$ over all possible state sequences $\boldsymbol{x}_{0:t}$ for given data $\boldsymbol{y}^o_{1:T}$. The constants of proportionality in (16) are the normalisation constants independent of the arguments of the three "energies" E_0, E and U.

We illustrate this minimisation for the important case that the process noise $\boldsymbol{w}_k$, the measurement noise $\boldsymbol{v}_k$ and the initial state $\boldsymbol{x}_0$ are iid Gaussian random variables, i.e. $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}_{\mathbb{R}^n}, Q)$, $\boldsymbol{v}_k \sim \mathcal{N}(\boldsymbol{0}_{\mathbb{R}^m}, R)$ and $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{0}_{\mathbb{R}^n}, S)$ with zero mean and coviance matrix Q, R and S respectively. Then*,

$$E_0(\boldsymbol{x}_0) = \frac{1}{2}\|\boldsymbol{x}_0\|^2_S = \tag{17}$$

$$E(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{\phi}_t(\boldsymbol{x}_k)\|^2_Q \tag{18}$$

$$U(\boldsymbol{y}_k|\boldsymbol{x}_k) = \frac{1}{2}\|\boldsymbol{y}_k - \boldsymbol{h}(\boldsymbol{x}_k)\|^2_R \tag{19}$$

6 Linear models: The Kalman filter and smoother

BIS HIER

*Here we have used the notation $\boldsymbol{a}^T B \boldsymbol{a} = \|\boldsymbol{a}\|^2_B$ for a column vector $\boldsymbol{a} \in \mathbb{R}^n$ and a square matrix $B \in \mathbb{R}^{n\times n}$.

The variances and covariances between the different output measurements are collected in the covariance matrix $R_s = Cov(\boldsymbol{v}_s) = cov\left(v_k(t_s), v_l(t_s)\right)_{k,l=1,\ldots,m}$ at time t_s. In many cases the measurement noise can be assumed to be Gaussian $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^m}, R)$ and is thus completely characterised by R. Non-gaussian measurements can often be transformed to a Gaussian distribution. In this review we assume Gaussian measurement noise, although other distributions of the measurement noise can also be considered, but often at the price of more complicated DA equations.

- State observation
- discrete time and noise

7 Monte Carlo algorithms for DA

8 Structural model error

9 Outlook

A Appendix name

References

[1] Becker V, Schilling M, Bachmann J, Baumann U, Raue A, Maiwald T, et al. Covering a Broad Dynamic Range: Information Processing at the Erythropoietin Receptor. Science. 2010 Jun;328(5984):1404–1408. Available from: `http://www.sciencemag.org/cgi/doi/10.1126/science.1184913`.

bla

$$\dot{\boldsymbol{x}} = \underbrace{\begin{pmatrix} k_t B_{max} - k_t x_1 - k_{on} x_1 x_2 + k_{off} x_3 + k_{ex} x_4 \\ -k_{on} x_1 x_2 + k_{off} x_3 + k_{ex} x_4 \\ k_{on} x_1 x_2 - k_{off} x_3 - k_e x_3 \\ k_e x_3 - k_{ex} x_4 - k_{di} x_4 - k_{de} x_4 \\ k_{di} x_4 \\ k_{de} x_4 \end{pmatrix}}_{\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})} \begin{matrix} \text{EpoR} \\ \text{Epo} \\ \text{Epo-EpoR} \\ \text{Epo-EpoR}_i \\ \text{dEpo}_i \\ \text{dEpo}_e \end{matrix}$$

$$\boldsymbol{y} = \underbrace{\begin{pmatrix} \kappa_1 (x_2 + 2x_6) \\ \kappa_2 (x_3) \\ \kappa_3 (x_4 + x_5) \end{pmatrix}}_{\boldsymbol{h}(\boldsymbol{x})} \begin{matrix} \text{Epo} + \text{dEpo}_i \\ \text{Epo-EpoR} \\ \text{Epo-EpoR}_i + \text{dEpo}_i \end{matrix}$$

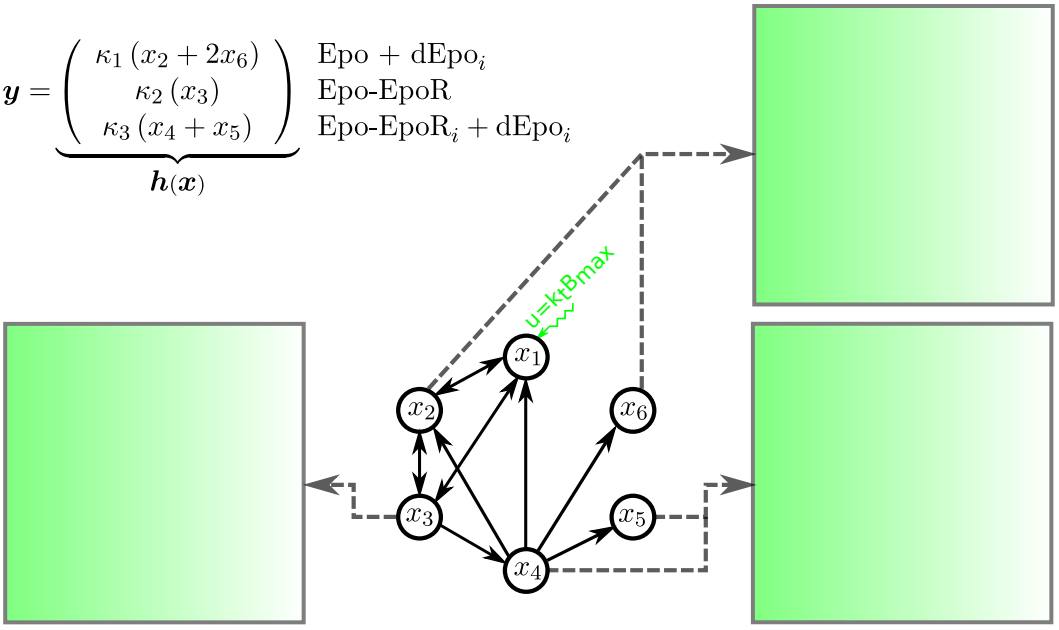

Figure 1: The Epo receptor regulation model [1] used as a running example. The state $\boldsymbol{x} = (x_1, \ldots, x_6)^T$ of this model is given by the concentrations of the Epo receptor (x_1) on the cell surface which can bind to Epo (x_2) and build the ligand-receptor complex (x_3). This complex is able to activate subsequent signaling cascades, e.g. the JAK-STAT signaling pathway. In addition the ligand-receptor complex can be internalized (x_4) and dissociate from Epo which is then degraded (x_5) and transported to the extracellular space (x_5). The complex regulation of this receptor is characterized by receptor mobilization, turnover and recycling. The rate constants correspond to (i) receptor turnover (k_t), (ii) ligand-receptor binding (k_{on}) or dissociation (k_{off}), (iii) ligand-induced endocytosis (k_e), (iv) recycling (k_{ex}) and (v) internal (k_{di}) or external (k_{de}) degradation of Epo. Only the the Epo concentration in medium (y_1), on surface (y_2) and in cells (y_3) can be measured up to some scaling paramters κ_j, $j \in \{1, 2, 3\}$. The influence graph illustrates the interactions between the state variables (black arrows) and the measured outputs y_k, $k \in \{1, 2, 3\}$ (dashed arrows).