A Two-level Approach for Fusing Early Signaling Events and Long Term Cellular Responses

Nadine Rudolph ∗ Tina Meyer ∗ Kristina Franzen ∗∗ Christoph Garbers ∗∗∗ Fred Schaper ∗∗ Stefan Streif ∗∗∗∗ Anna Dittrich ∗∗ Rolf Findeisen ∗

∗ Laboratory for Systems Theory, Otto-von-Guericke University Magdeburg, Germany
∗∗ Chair for Systems Biology, Otto-von-Guericke University Magdeburg, Germany
∗∗∗ Unit for Cytokine and Metallproteinase Research, University of Kiel, Germany
∗∗∗∗ Department of Computer Science and Automation, Ilmenau University of Technology, Germany

Abstract: In biological systems, reactions on different time scales exist and need to be considered for the analysis of physiological reactions such as proliferation. In this contribution we describe a two-level approach to decode biological reactions on a short time-scale, e.g. signaling, into long term cellular responses, e.g. proliferation. First, we derive a valid and parametrized dynamic model for signaling events on the short time-scale using set-based estimation methods allowing to take uncertainties into account. Second, the derived model candidate for early signaling events is fused with a model for long term proliferation using shape-based signaling properties. This approach is realized in the specific case study of Interleukin-6-induced signaling and proliferation. Our modeling approach enables us to consider both, dynamic early signaling events and static long term events. Furthermore, it allows a deeper understanding of how cells process information from early signaling events to long term cellular responses.

Keywords: set-based approach; model invalidation; parameter estimation; convex optimization; information processing; early signaling events and long term biological responses

1. INTRODUCTION

Biological signaling processes can be divided into different time scales. Early reactions span seconds to hours and include the activation of receptors and the transmission of signals into the nucleus. Late reactions that occur within hours to days include biological responses such as differentiation or proliferation as a consequence of the initial early reactions [Weng et al., 1999]. By studying the dynamical behavior of early signaling events with mathematical models one gains insight into interactions of components of signaling pathways. For example, mathematical modeling of the Janus kinase - Signal Transducers and Activators of Transcription (Jak-STAT) signaling pathway helped to understand molecular mechanisms underlying this pathway [Quaiser et al., 2011, Blätke et al., 2013]. However, the analysis of early signaling events alone is not sufficient to understand how a biological signal is processed into late cellular responses because the underlying interconnections are often unclear. Nonetheless, understanding and predicting long term cellular responses is essential to gain insight into physiological and pathophysiological responses [Hood et al., 2004]. In addition, predictions of responses on longer time scales, i.e. from days over weeks to months enable to reflect the development of diseases and are the ultimate goal for modeling and analyses in systems medicine. Integrating early signaling events and long term cellular responses into a mathematical model is challenging due to the multiple biological factors that influence this integration. These factors are often not yet well characterized and thereby can not be modeled in detail. That is why phenomenological approaches to combine early signaling events and long term responses have been developed, e.g. [Schneider et al., 2012]. Schneider et al. postulate to combine early signaling events and long term cellular responses by linearly correlating shape properties of activated signaling molecules, e.g. the maximum peak height of a signal with the strength of long term responses. However, so far these approaches have not been established in combination with set-based estimation methods. Set-based approaches allow to take uncertainties in measurement and input data as well as initial conditions and parameters into account. Especially for guaranteed invalidation of uncertain model candidates, set-based analysis methods are useful tools.
In this contribution, we develop a valid mathematical model of Interleukin-6-induced (IL-6-induced) Jak-STAT signaling on the short time-scale and combine this model phenomenologically with a model for IL-6-induced cell proliferation. The model for early signaling events is able to explain the transient activation of Jak-STAT signaling by IL-6 and is derived by testing two competing model candidates describing different mechanisms of IL-6-induced receptor complex assembly. As a framework, we apply a set-based estimation approach as described in [Rumschinski et al., 2010] which allows to consider uncertainties in measurement data, initial conditions, parameters and inputs and to make guaranteed statements about model invalidity. We estimate the feasible parameter set of the superior model candidate and determine valid parametrizations by a sample-based approach.

Furthermore, we combine the validated model candidate with a phenomenological model for cell proliferation. Using shape properties of transient STAT phosphorylation we relate early signaling events to long term responses by computing, 1) the integral of the trajectory of STAT phosphorylation, 2) the maximal peak height of STAT phosphorylation. As long term model output, proliferation is simulated and compared with experimentally determined data for IL-6-induced proliferation of cells from a murine bone-marrow derived pre-B-cell line (Ba/F3).

This contribution is structured as follows: In Section 2 the set-based estimation framework is outlined. Section 3 introduces IL-6-induced Jak-STAT signaling and depicts the results of the set-based model invalidation and parameter estimation. In Section 4 the approaches for phenomenologically decoding early signaling events into long term cellular responses are presented and employed.

2. SET-BASED ESTIMATION

In this section, we introduce the problem of model invalidation and consistent parameter estimation using set-based feasibility methods.

2.1 Problem Setup

In the first part of this contribution, we consider a class of polynomial discrete-time systems of the form

\[ g(x(k+1), x(k), u(k), p) = 0 \] (1a)
\[ h(y(k), x(k), u(k), p) = 0, \] (1b)

where \( g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_x} \) are polynomial or rational functions, \( x(k) \in \mathbb{R}^{n_x} \) is the time-variant state vector, \( u(k) \in \mathbb{R}^{n_u} \) the time-variant input vector, and \( p \in \mathbb{R}^{n_p} \) the time-invariant parameter vector. Time is indexed by \( k \in \mathbb{N} \). The model output equations are given by \( h : \mathbb{R}^{n_y} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_y} \), which are assumed as polynomial functions, and \( y(k) \in \mathbb{R}^{n_y} \) denote the time-variant model output vector.

For parameter estimation and model invalidation, we assume that different measurements are taken at time instances indexed by \( k \in \mathcal{T} \), where \( \mathcal{T} \) describes the considered finite time horizon.

Typically, the measurements are uncertain. Hence, we assume that each data point at time instance \( k \) (long or short term), i.e. \( y(k) \) and \( u(k) \) lie in the compact sets \( \mathcal{Y}_k \) and \( \mathcal{U}_k \):

\[ y(k) \in \mathcal{Y}_k \] (2)
\[ u(k) \in \mathcal{U}_k. \] (3)

Furthermore, we assume that initial conditions \( x_0 \) (i.e. at \( k = 0 \)) and parameters \( p \) lie in the possibly large sets \( \mathcal{X}_0 \) and \( \mathcal{P} \), respectively:

\[ x_0 \in \mathcal{X}_0, \] (4)
\[ p \in \mathcal{P}. \] (5)

Typically, \( \mathcal{X}_0 \) and \( \mathcal{P} \) can be derived from initial knowledge, conservation relations or from the physical meaning of the variables (e.g. concentrations have to be positive).

2.2 Model Invalidation and Parameter Estimation Using Infeasibility Certificates

As framework, we consider a set-based approach for the invalidation of competing hypotheses and parameter estimation [Rumschinski et al., 2010] using the following notion:

Definition 1. (Model consistency). Model (1) is said to be consistent with the measurements \( \mathcal{Y} \) and \( \mathcal{U} \), if there exist \( p \in \mathcal{P} \) and \( x_0 \in \mathcal{X}_0 \), such that \( y(k) \in \mathcal{Y}_k \) and \( u(k) \in \mathcal{U}_k, \forall k \in \mathcal{T} \).

Based on Definition 1 we can state the first problem:

Problem 1. (Estimation of feasible parameter set). Estimate the feasible parameter set \( \mathcal{P}_f \) for a model candidate such that \( y(k) \in \mathcal{Y}_k \) and \( u(k) \in \mathcal{U}_k, \forall k \in \mathcal{T} \).

The general aim of the set-based approach is to rule out competing model hypotheses by checking whether the feasible parameter set \( \mathcal{P}_f \) is empty or not. If \( \mathcal{P}_f \) is found to be empty the model is an invalid candidate and therefore rejected. If \( \mathcal{P}_f \) is found to be not empty the model is a not invalid candidate and will be used for further analyses.

To check invalidity, we consider the following feasibility problem:

\[
\text{FP} : \begin{cases}
\text{find } \xi F_{\mathcal{P}} \\
\text{s.t.} \quad \text{equations (1)},
\end{cases}
\]

\[
x_0 \in \mathcal{X}_0, \quad p \in \mathcal{P}, \quad y(k) \in \mathcal{Y}_k, \quad u(k) \in \mathcal{U}_k, \forall k \in \mathcal{T},
\]

where \( \xi F_{\mathcal{P}} \in \mathbb{R}^{(n_x+n_u+n_y)n_t+n_y} \) is a vector that contains all the variables in (1), i.e. state, input, and output variables at the time-instances \( k \in \mathcal{T} \), and the parameters \( p \).

Note, when solving FP optimality does not matter rather we want to find any feasible solution. Due to the nonlinearities in (1), the set of feasible parameters \( \mathcal{P}_f \) is difficult to derive. However, as shown it is possible to relax the FP into a convex semi-definite feasibility problem (SDP) or – if larger problems are considered - into a linear program (LP) [Borchers et al., 2009]. Relaxation here refers to the fact that the feasible parameter set \( \mathcal{P}_f \) of the FP is always contained in the solution set of the SDP and the LP, i.e. \( \mathcal{P}_f \subseteq \mathcal{P}_\text{SDP} \subseteq \mathcal{P}_\text{LP} \).

Consistency can then efficiently be checked using the Lagrangian-dual of the linear relaxation. The weak-duality theorem guarantees that if the objective of the dual pro-
gram is unbounded, then the FP does not admit a solution \cite{Borchers et al., 2009}.

2.3 Outer Approximation of Feasible Parameter Sets for Validation and Estimation

Relaxation of the FP introduces conservatisms, i.e. the solution space increases. However, it allows definite statements about model invalidity checking the global parameter space. Furthermore, if a model is found to be not invalid, set-based methods allow an efficient outer approximation of the feasible parameter space. In this contribution we use the approach of bisectioning the parameter space \( \mathcal{P} \) by dividing it into partitions and checking each partition for consistency. Using a recursive algorithm, the feasible parameter set \( \mathcal{P}_f \) can be approximated systematically and up to a chosen precision threshold \cite{Borchers et al., 2009, Rumschinski et al., 2010}. Note, that only an outer-approximation of the feasible parameter set is not sufficient to show validity of a model. In a next step, feasible parameter samples have to be found (e.g. using Monte Carlo sampling or inner approximations \cite{Streif et al., 2013}) such that \( y(k) \in \mathcal{Y}_k \) and \( u(k) \in \mathcal{U}_k \), \( \forall k \) holds.

3. EARLY JAK-STAT SIGNALING

To outline and illustrate the set-based approach as described in Section 2, we consider in the following IL-6-induced receptor complex assembly and downstream signaling, i.e. activation of the Jak-STAT pathway.

3.1 Biological Principles of IL-6-induced Jak-STAT Signaling

IL-6 is a pleiotropic cytokine and a key regulator of inflammatory processes. Dysregulation of IL-6 function leads to numerous pathological states. It has been shown that high and persistent serum levels of IL-6 foster diseases like rheumatoid arthritis, Crohn’s disease and multiple sclerosis \cite{Scheller et al., 2006}. The mechanisms by which IL-6 initiates signal transduction are the following (cf. Fig. 1), \cite{Heinrich et al., 2003}; IL-6 forms a complex with the receptor subunit glycoprotein 80 (IL-6R\(\alpha \)). Subsequently, the receptor subunit glycoprotein 130 (gp130) binds to the complex of IL-6 and IL-6R\(\alpha \). Tyrosine kinases of the Jak family are constitutively bound to the intracellular domain of gp130. After receptor complex assembly, Jaks become activated and in turn phosphorylated gp130. Subsequently, STATs are recruited to phosphorylated gp130 and are phosphorylated by Jak3 leading to the formation of active STAT dimers. STAT dimers translocate into the nucleus where they activate the transcription of target genes coding, e.g. for Suppressors of Cytokine Signaling (SOCS) – a negative feedback inhibitor of Jak kinases \cite{Endo et al., 1997}. The inhibition of Jak by SOCS leads to a transient activation of Jak-STAT signaling. Note, also other genes controlling cellular responses are activated in parallel. On a long time-scale, IL-6 induces proliferation in certain cell types, such as Ba/F3.

Fig. 1. Schematic and simplified representation of IL-6-induced receptor complex assembly, Jak-STAT pathway activation and activation of target genes.

3.2 Modeling Assumptions

To keep the mathematical models simple, we do not model the mechanisms described in Section 3.1 in detail but rather assume the following:

- formation of IL-6, IL-6R\(\alpha \) and gp130 complex is considered as active signaling complex
- Jak kinases are represented by gp130 species
- STATs are phosphorylated by the active receptor complex
- STAT activation represents phosphorylation and dimerization of STAT proteins.

For model invalidation and parameter estimation we use simulated data that describe the transient course of IL-6-induced STAT phosphorylation (black bars Fig. 2(a)). In addition, a theoretical measurement error of 10\% and parameter uncertainties of \( p_i = [0.001, 0.50] \) are considered. Furthermore, we define a constant input concentration in the range of IL-6 = [0.18, 0.20] nM. The dynamic processes of IL-6-induced Jak-STAT signaling as well as inhibition via SOCS are modeled using the law of mass action and Michaelis-Menten type kinetics, respectively as described for Interferon-\( \gamma \) signaling by \cite{Yamada et al., 2003}. For illustration of the set-based approach, two competing model candidates are implemented using the freely available toolbox ADMIT \cite{Streif et al., 2012}.

3.3 Model Candidate 1

The first model candidate describes an oversimplified activation of IL-6-induced signaling. We aggregate the stepwise receptor complex formation in response to IL-6 treatment, i.e. we assume that the active receptor complex consisting of IL-6, IL-6R\(\alpha \) and gp130 forms immediately in the presence of IL-6 \((u)\) resulting in the formation of the active receptor complex \( R_{\text{complex}} \) \((x_1)\). \( R_{\text{complex}} \) is subsequently phosphorylated to \( (p)R_{\text{complex}} \) \((x_2)\). Then, STATs are phosphorylated to \( (p)\text{STATs} \) \((x_3)\) leading to SOCS mRNA expression \((x_4)\). Translation of SOCS mRNA results in the synthesis of SOCS protein \((x_5)\).

For Model Candidate 1 we consider the following conserved moieties: IL-6R\(\alpha_{\text{total}} \) = IL-6R\(\alpha \) + \( R_{\text{complex}} \) + \( (p)R_{\text{complex}} \) and STAT\(\text{total} \) = STAT + \( (p)\text{STAT} \). The reaction mechanisms can be described as follows:
\[ x_1(k+1) = x_1(k) + h \left( p_1 (1 - x_1(k) - x_2(k))u(k) - (p_2 + p_3)x_1(k) \right) \]
\[ x_2(k+1) = x_2(k) + h \left( \frac{p_3x_1(k)}{1 + p_nh x_5(k)} - p_4x_2(k) \right) \]
\[ x_3(k+1) = x_3(k) + h(p_5x_2(k)(1 - x_3(k)) - p_6x_3(k)) \]
\[ x_4(k+1) = x_4(k) + h(p_7x_3(k) - p_8x_4(k)) \]
\[ x_5(k+1) = x_5(k) + h(p_9x_4(k) - p_{10}x_5(k)) \]

where \( h \) results from time-discretization and is set to 1 min.

Using the set-based approach we could invalidate Model Candidate 1 as the feasible parameter set was found to be empty. In Fig. 2(a) 10 sample trajectories are depicted (red) illustrating that Model Candidate 1 can not explain the data (black bars).

### 3.4 Model Candidate 2

The second model candidate considers a stepwise association of the receptor subunits IL-6Rα and gp130. After binding of IL-6 to IL-6Rα, the complex IL-6/IL-6Rα \((x_1)\) associates with gp130 \((x_2)\) to the active receptor complex \(R_{\text{complex}}(x_3)\). Then, \(R_{\text{complex}}\) is phosphorylated to \((p)R_{\text{complex}}(x_4)\). Equations for \((p)\)STAT \((x_5)\), SOCS mRNA \((x_6)\), SOCS \((x_7)\) as well as inhibition through SOCS are implemented in analogy to Model Candidate 1:

\[ x_1(k+1) = x_1(k) + h(p_1 (1 - x_1(k) - x_2(k))u(k) - (p_2 + p_3)x_1(k)) \]
\[ x_2(k+1) = x_2(k) + h(p_4x_3(k) - p_3x_1(k)x_2(k)) \]
\[ x_3(k+1) = x_3(k) + h(p_5x_1(k)x_2(k) - p_4x_3(k) - p_5x_3(k) + p_6x_4(k)) \]
\[ x_4(k+1) = x_4(k) + h(p_5x_3(k) - p_7x_4(k)) \]
\[ x_5(k+1) = x_5(k) + h(p_7x_3(k)(1 - x_5(k)) - p_8x_5(k)) \]
\[ x_6(k+1) = x_6(k) + h(p_9x_5(k) - p_{10}x_6(k)) \]
\[ x_7(k+1) = x_7(k) + h(p_{11}x_6(k) - p_{12}x_7(k)) \]

Note, in addition to the conserved moieties from Model Candidate 1, we consider \(gp130_{\text{total}} = gp130 + R_{\text{complex}} + (p)R_{\text{complex}}\). The feasible parameter set for Model Candidate 2 was found to be not empty. Therefore, we could not invalidate Model Candidate 2. To show validity we first performed an outer approximation of the feasible parameter set that reduced the initial parameter space as defined in Section 3.2 (see blue boxes in Fig. 2(b)). Then we run a Monte-Carlo sampling procedure to find parameter samples within the reduced parameter space that are consistent with the data. Exemplarily derived parametrizations are depicted in Fig. 2(b) for \(p_3, p_4\) and \(p_5\) as red crosses. Trajectories for the valid parametrizations are shown in Fig. 2(a) in blue.

In summary, we developed two competing model candidates for IL-6-induced Jak-STAT signaling. Using the set-based approach, we could invalidate Model Candidate 1 as the parameter set was found to be empty. Model Candidate 2 was found to be not invalid and we could derive valid parameterizations that are able to explain the data. Hence, Model Candidate 2 will be used next for further analyses. In more detail, we extend the candidate phenomenologically using shape properties of the trajectory of \((p)\)STAT and derive a model that integrates both, early Jak-STAT signaling and long term proliferation.

### 4. COMBINING EARLY SIGNALING EVENTS AND LONG TERM CELLULAR RESPONSES

In this section, we describe the approaches and results to decode Jak-STAT signaling into cell proliferation. Cell proliferation – which serves as model output on the long time-scale – was experimentally determined. For this, Ba/F3 cells expressing IL-6Rα and gp130 on their surfaces were stimulated with different IL-6 concentrations ranging from 0.004-7.7 nM. After two days, proliferation was determined as described in [Garbers et al., 2011]. The results are shown in Fig. 4 as black bars.

### 4.1 Problem Setup

To fuse early signaling events with long term responses we transfer the discrete-time model from Section 2 (Eq. (1)) into a continuous-time model because it is not possible to compute shape properties of a signal within the presented set-based framework. The continuous-time system is described as follows:

\[ \dot{x}(t) = g(x(t), u(t), y(t), p) \]  \( (6a) \)
\[ y(t) = h(x(t), u(t), p) \]  \( (6b) \)

where \(x(t), u(t),\) and \(y(t)\) and \(p\) are as before.
After converting Model Candidate 2 into a continuous-time system, we extend the model candidate phenomenologically with a model for cell proliferation. We propose to estimate the model parameters on the early signaling and long-time scale applying least squares estimation [Mendes and Kell, 1998] and using fmincon as solver. To do so, we set the following constraints: \( p_i \geq [0.001, 0.50] \) for \( i = 1, 2, 6, ..., 12 \) and \( x_0 = [0, 1.5, 0, 0, 0, 0, 0, 0, 0, 0] \). For the parameters \( p_3, p_4 \), and \( p_5 \) we assume the estimated outer bounds based on the set-based analyses in Section 3, i.e. \( p_3 = [0.001, 10.7], p_4 = [0.001, 0.31], p_5 = [0.001, 12.5] \) (cf. Fig. 2(b)).

For the continuous-time model, we use IL-6=0.19 nM as input concentration. While on the early signaling scale several, time-dependent measurements for \((p)\)STAT are available (cf. Fig. 2(a)), on the long time-scale only one measurement point for proliferation corresponding to IL-6=0.19 nM (cf. Fig. 4) is available to infer the unknown kinetic parameters. Therefore, we first estimate the parameters by fitting the model to the measurement data on the early signaling and long time-scale and then based on these parameter estimates – predict and compare proliferation for the remaining input concentrations.

4.2 Approaches for Fusing Early and Late Signaling Events

We extend the derived Model Candidate 2 from Section 3 using the approaches described in [Schneider et al., 2012]. One possibility to fuse early and long term events is to correlate the integral of the trajectory of a signal \( S(t) \), i.e. \((p)\)STAT linearly with the strength of the long term response \( L_{int} \), i.e. proliferation (cf. Fig. 3(a)):

\[
L_{int} = \alpha \int_{t_1}^{t_2} S(t) \, dt,
\]

where \( \alpha \) describes the linear dependency of the integral and the strength of the long term response and \( t_1 \) as well as \( t_2 \) are the lower and upper time points for computing the integral.

Another possibility is to correlate the maximal peak height of a signal \( S(t) \) linearly with the strength of the response \( L_{max} \) (cf. Fig. 3(b)), i.e.

\[
L_{max} = \beta \max S(t),
\]

where \( \beta \) describes the linear dependency of maximal signal peak height and the strength of the long term response.

In the following we will employ and demonstrate the described approaches using the derived Model Candidate 2. Note, initial constraints for \( \alpha \) and \( \beta \) were set to: \( \alpha = \beta = [0.001, 100] \) and \( t_1 = 0 \) min, \( t_2 = 90 \) min.

4.3 Results for Fusing IL-6-induced Early Jak-STAT Signaling and Cell Proliferation

At first, our aim is to determine estimates for the parameters \( p, \alpha \) and \( \beta \) such that the constraints for the model output on the early signaling scale, i.e. kinetics of \((p)\)STAT and constraints for the model output on the long time-scale, i.e. cell proliferation are reproducible.

Fig. 3. Schematic presentation of decoding early signaling events \( S(t) \) into long term biological responses \( L \) using shape properties of a signal. (a) Integral of signal \( S(t) \). (b) Maximum peak height of signal \( S(t) \).

By randomly generating initial parameter combinations and comparison of simulations and model output constraints on the early signaling and long time-scale we obtained several satisfying parametrizations explaining both, early Jak-STAT signaling and long term cell proliferation. Using these parametrizations, we predicted proliferation for input concentrations higher and lower than IL-6=0.19 nM and compared the predicted proliferation with measurement data (cf. Fig. 4).

Fig. 4(a) and 4(b) depict the results for decoding the integral of the \((p)\)STAT signaling curve as well the maximum height of the \((p)\)STAT signal into cell proliferation for 30 estimates (out of 1000) for \( p, \alpha \) and \( \beta \) which yielded the best fit to the data. It can be seen, that especially for higher IL-6 concentrations the approach of correlating the maximal peak height of \((p)\)STAT with cell proliferation seems to be more suitable than the integral of the \((p)\)STAT trajectory. However, so far no clear conclusions about the validity of one of the approaches can be drawn.

5. CONCLUSIONS AND OUTLOOK

This paper proposed a two-level approach for fusing early signaling events and long term biological responses. In the first part of this contribution, we derived two competing model candidates describing IL-6-induced receptor complex assembly and Jak-STAT signaling. Applying a set-based estimation approach, we could invalidate an oversimplified early signaling model candidate as the parameter set was found to be empty. For the remaining model candidate we outer-approximated the feasible parameter set \( P_f \) and derived valid parametrizations within \( P_f \) running a Monte-Carlo sampling procedure. Outer approximations of a feasible parameter set always introduce additional invalid solutions due to the conservatism of the set-based approach. Hence, a sample-based approach as used in this contribution is indispensable to show validity. As an alternative, inner approximations of \( P_f \) could be computed which will be done in a next step for Model Candidate 2 [Streif et al., 2013].

In the second part of this contribution, we extended the validated model candidate for early Jak-STAT signaling with a model for long term cell proliferation using phe-


