# Review on Sensitivity Analysis in Biochemical Models

Agata Charzyńska

Institute of Computer Science, Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warsaw, Poland

**Abstract.** In this work we focus on sensitivity analysis of biochemical models. The research includes both formal modelling of signalling and metabolic pathways as well as developing theoretical methods for model assessment. The aim of this study is to better understand a natural phenomenon in quantitative an qualitative manner by mathematical modelling. To validate any model it is crucial to determine which factors are most influential for a modelled system behaviour. Part of my study is to develop a new sensitivity analysis method based on mutual information that provides an efficient identification of parameters and group of parameters that are crucial for a modelled system, providing additionally information about interactions between parameters in accordance to the model output.

In the first section of this paper we briefly present the motivation behind formal modelling and its necessity in any experimental design.

The second section contains review of classical sensitivity analysis methods based on literature and in the second part of this section we recall two recently invented SA methods: Stochastic Noise Decomposition (SND) and Sensitivity Analysis (SA) based on Mutual Information (MI). We tested and implemented the SND method in direct cooperation with authors of this method (cf. Komorowski et al., 2013) the results of our work were presented in application note - StochDecomp Matlab package (Jetka et al., 2014). The second method SA based on MI was deeply studied and developed by us with the application to continuous random variables. We introduce a novel correction to the classical $k$-nn entropy estimator to reduce the bias of estimation in finite sample size for highly dimensional data.

The third section is devoted to a brief summary of biochemical models of our interest. Some models were adopted from literature and used as a test example for application of theoretical SA methods e.g. p53-Mdm2 negative feedback loop model and other models were fully developed and implemented by us e.g. sphingolipid metabolism model (Wronowska et al., 2015). To all presented in this section models we applied several SA methods.

## 1 Motivation

Mathematical modelling of biological phenomena described e.g. by dynamical systems, complements experimental technologies used to identify and comprehend a role of system components. The process in which a model is formulated and refined helps to articulate hypotheses and thereby supports the design of experiments to validate these hypotheses and the model itself. Once the model is validated it is used to speculate about mechanisms underlying cell functions.

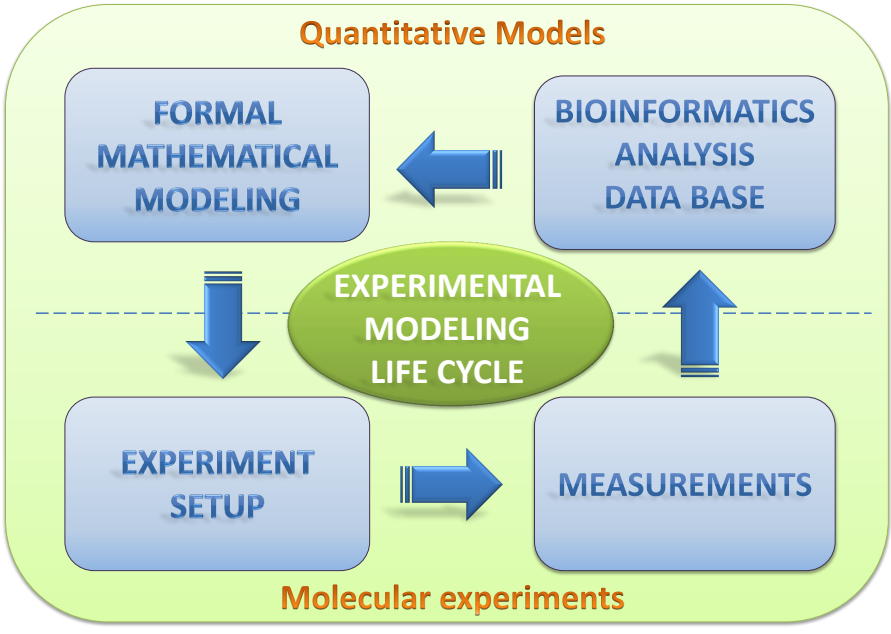


**Fig. 1.** Scheme of experimental modelling life cycle

The standard approach to understand dynamics of biological system is to observe the behaviour of as many as possible system components. An important element of the model design is analysis and identification of most informative and responsive to perturbations model elements (sensitivity analysis) to reveal the spectrum of available dynamical regimes (validated by model checking). Verification of the model design together with parameters estimates is carried out experimentally by comparing model predictions with experimental results for stimulation profiles. Any necessary corrections in model structure and parameter estimates must be made. The experimental modelling life cycle scheme is depicted in Fig. 1.

The construction and analysis of mechanistic models of biological systems is a part of recently established, highly interdisciplinary fields of systems and computational biology. Computational modelling of signal transduction integrates available knowledge about pathway regulation, and the general chemical and physical principles with experimental data from different biotechnology platforms. Such approach constitutes a powerful solution for formalizing and extending traditional molecular and cellular biology.

## 2 Sensitivity Analysis

Mathematical modelling of biological phenomena can be carried out in a deterministic, stochastic or hybrid manner. The first approach is based on the Ordinary Differential Equations (ODEs), while the second one is based on the stochastic processes or stochastic differential equations (SDEs) theory. Both types of models are usually based on some simplifying assumptions i.e. that the temperature of a chemical environment is constant, and that the diffusion process occurs immediately, which ensures an even distribution of a substance over a limited volume. Deterministic models describe changes in mean concentrations of reagents (species) over time, and they do not include the effect of fluctuations which occur in reality. This means that for given initial conditions, a deterministic model will always provide the same results. While stochastic models describe the evolution of the probability distribution of all possible system states with respect to time. Both types of modelling requires proper verification and analysis.

A biochemical model described by ODEs can be expressed in the matrix form:

$$\frac{d\mathbf{S}(t)}{dt} = M\mathbf{v}(\mathbf{S}(t)),$$

where the system state is represented by the time dependent state vector $\mathbf{S}(t)$ of species concentration, $M$ denotes the stoichiometry matrix and $\mathbf{v}(\mathbf{S}(t))$ denotes a vector of reaction fluxes (in simplest standard modelling according to Mass Action Law (MAL) or Michaelis Menten (MM) kinetics possibly including inhibition rates).

The most popular approach to describe discrete stochastic model of biochemical pathway is Chemical Master Equation (Chapman-Kolmogorov equation of Markov chain modelling the evolution of the system):

$$\frac{pP(\mathbf{x},t)}{dt} = \sum_j a_j(\mathbf{x}-\mathbf{m}_j)P(\mathbf{x}-\mathbf{m}_j,t) - \sum_j a_j(\mathbf{x})P(\mathbf{x},t),$$

where the system state is denoted by the vector $\mathbf{X}(t) \in \mathbb{N}^N$ of numbers of molecules each row for one of $N$ reacting species, $\mathbf{m}_j$ denotes the $j$-th column of stoichiometry matrix $M = (\mathbf{m}_1, \ldots, \mathbf{m}_R)$ and $P(\mathbf{x},t)$ denotes the time- and state-dependent distribution of system being in state $\mathbf{X}(t) = \mathbf{x}$ and finally $a_j(\mathbf{X}(t))$ denotes the propensity function associated with the $j$-th reaction (Charzyńska et al., 2012).

### 2.1 Classification of SA Methods

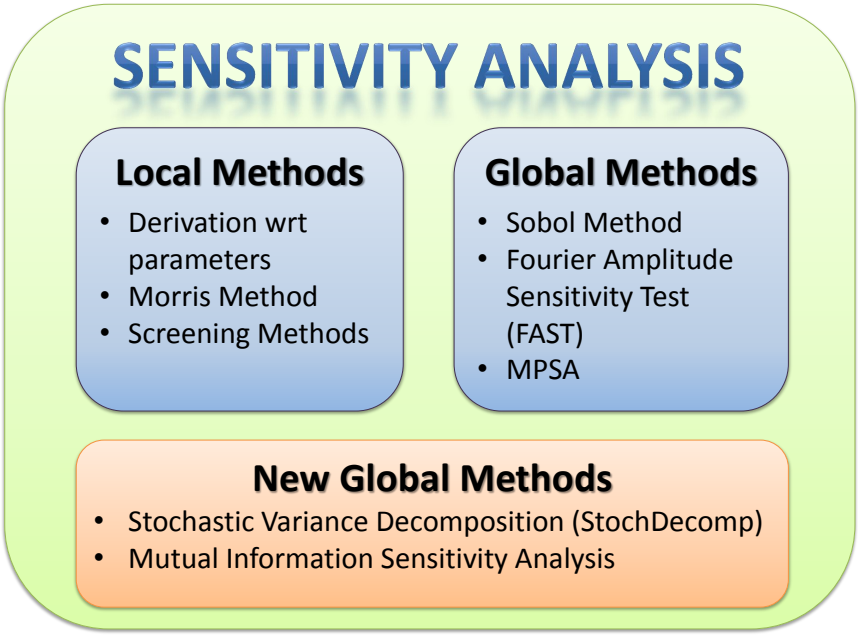


**Fig. 2.** Local and Global Sensitivity Analysis Methods

Sensitivity analysis is used to determine dependencies between input parameters and the results of the model. One can chose as input parameters for example initial concentrations of modelled species or reaction rates. Result of the biochemical model is most commonly defined as the density of species as a function of time. SA is very useful in mathematical modelling, as it describes dependencies between different elements of the model, it is also applicable to empirical experiments planning and enables verification of theoretical model results together with numerical and empirical results. SA also enables recognition of model's conceptual and implementational omissions.

Sensitivity analysis investigates the relations between uncertain parameters of a model, and a property of the observable outcome, which represents some prototypic features of the modelled system (Saltelli et al., 2008) . SA has been used in various parametrization tasks for models of biological systems, such as finding essential and insignificant parameters for the prioritization (Yue et al., 2008) , identifying parameters interactions or or parameters clustering (Mahdavi et al., 2007).

Classically, sensitivity of the model to the parameters is determined by the partial derivation of the outcome variables with respect to parameters. SA methods based on such quantities are called local (LSA), as the derivative is taken at a fixed point in the state space. Sensitivity indices are defined as partial derivatives of system states with respect to parameters integrated by time:

$$s_{n,i} = \int_0^T \left| \frac{\partial S_n(t)}{\partial \theta_i} \right|_{\theta=\theta_0} dt$$

where $S_n$ are different species concentrations, $\theta$ is the vector of parameters and $\theta_0$ is some fixed point in parameters space. One of disadvantage of this method is the high dependence of sensitivity indices to arbitrary choice of time horizon $T$ that can influence the SA results.

Moreover, these methods belong to the class of one-factor-at-time (OAT) methods, because the net effect of a parameter to the model outcome is taken while assuming that all other factors are fixed. However, most of the biochemical reactions networks yield models of a non-linear nature and for these models, OAT methods can be of limited use if not outright misleading (Saltelli et al., 2005) . Possible solution is to ingestive of the influence of simultaneous changes in parameters values by assessing higher order partial derivatives (Mahdavi et al., 2007), where the order depends on the non-linearity level of the model. Nevertheless, it is still a local method, highly dependent on the given values of parameters.

On the other hand, there are so-called global sensitivity analysis (GSA) methods, that simultaneously examine a whole range of input parameters values. Exemplary implementations of the GSA indices are the model-free, global sensitivity measures such as the variance decomposition (Saltelli et al., 2008), or the parameters space mapping method of Monte Carlo filtering (MCF) such as the multi-parameter sensitivity analysis (MPSA) (Hornberger and Spear, 1981).

In between, there are screening techniques which approximate the GSA indices. Screening techniques, such as the weighted average of local sensitivities (Bentele et al., 2004) or the elementary effects of Morris (1991), are global in the sense that they scan a whole range of parameters values, but they use local OAT methods for each analysed set of parameter values.

For a sake of clarity, if not explicitly stated otherwise, we will use a term local method meaning the local and OAT method, as well as a term global method meaning GSA method (Global and simultaneous).

Finally, there are SA methods tailored specifically to the stochastic models (Gunawan et al., 2005). These methods recognize that the response is in form of distribution rather than a single value corresponding, for instance, to the mean value. Consequently, for systems where a parameter disruption does not significantly influence the mean but significantly influences the distribution itself, the model-free SA indices can incorrectly indicate a lack of sensitivity of the model (cf. Degasperi and Gilmore, 2008).

To extend the range of available global sensitivity analysis methods we recall here new approaches: method based on information theoretic measure and

stochastic noise decomposition. Both methods can be applied to dynamical systems whether formulated in deterministic or stochastic manner. Each method has its specificity: noise decomposition method allows to track how the stochastic noise distribute within the biochemical system in division into single reactions noise compartments, whereas mutual information method based on entropy estimation provides sensitivity indices and interaction indices for any group of parameters that represent model input.

### 2.2 Stochastic Noise Decomposition Method

The question which molecular species or parts of a network contribute most to the variability of a system or are responsible for most of the information loss has attracted much attention in recent years. Stochasticity is an indispensable aspect of biochemical processes not only but especially at the cellular level. Studies on how the noise enters and propagates in biochemical systems provides a non-trivial insights into the origins of noise in a model. Numerous studies focus on analysis of noise in signalling networks in detail and decomposition of the noise into contributions attributable to fluctuations in species concentration.

Recently developed StochDecomp (Jetka et al., 2014) is a flexible and widely applicable noise decomposition tool that allows to calculate contributions of individual reactions to the total variability of a system output. The method allows to quantify how the noise enters and propagates in biochemical systems. It is based on recently developed method (Komorowski et al., 2013) that allows to analyse how the structure of biochemical networks gives rise to noise in its outputs. In principle, this allows to efficiently calculate the contribution each reaction makes to the variability in all concentrations for any network, which can be modelled within the Linear Noise Approximation (LNA) framework. LNA is one of the possible simplification of the Chemical Master Equation , with the system dynamic modelled as Poisson process:

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{j=1}^{R} \mathbf{m}_j N_j \left( \int_0^t f_j(\mathbf{X}(\tau), \tau) d\tau \right)$$

where $N_j(\mathbf{X}(t), t)$ denotes Poisson process dependent on time and a system state $\mathbf{X}(t)$, corresponding to occurrence of $j$-th reaction. The probability that $j$-th reaction occur during the time interval $[t; t + dt)$ equals $f_j(x, t)dt$, where the $f_j(x, t)$ is called the transition rate.

It is more efficient to transit from discrete to continuous process, as accurate discrete models describe the exact evolution of probability distribution of the system state counted in molecules number. Discrete biochemical models are computationally not efficient, as simulations require significant resources. Consequently by use of deterministic approximation:

$$\Phi(t) = \Phi(0) + \sum_{j=1}^{R} m_j \int_0^t f_j(\Phi(s), s) ds$$

where $\Phi(t)$ is the mean system state being the solution of the ODEs, one can describe the system state evolution by dividing it into deterministic and stochastic part:

$$x(t) = \xi(t) + \Phi(t)$$

where $\Phi(t)$ is the deterministic part and $\xi(t)$ is the Weiner process describing stochastic noise of a system state (Komorowski et al. 2009). The next step of stochastic noise decomposition is to divided noise linearly into noise steaming from separate reactions. The total variance:

$$\Sigma(t) = \langle (x(t) - \langle x(t) \rangle)(x(t) - \langle x(t) \rangle)^T \rangle$$

is described by the differential equation

$$\frac{d\Sigma}{dt} = A(t)\Sigma + \Sigma A(t)^T + D(t), \tag{1}$$

where

$$\{A(\Phi, t)\}_{ik} = \sum_{j=1}^{r} m_{ij} \frac{\partial f_j(\Phi, t)}{\partial \Phi_k}$$

and $D(t)$ denotes diffusion matrix. The fact, that the variance can be represented as the sum of individual contributions,

$$\Sigma(t) = \Sigma^{(1)}(t) + \ \ldots \ + \Sigma^{(r)}(t). \tag{2}$$

results directly from the decomposition of the diffusion matrix $D(t) = \sum_{j=1}^{r} D^{(j)}(t)$ and the linearity of the equation for $\Sigma(t)$.

The Stochastic Noise Decomposition method is based on the LNA, which is assumed to provide a reasonable representation of analysed systems even in case of priori deterministic formulation. Origins of variability can be therefore assigned to individual reactions and arbitrarily defined network components.

Contrary to most available methods Stochastic Noise Decomposition is tailored for biochemical dynamical systems and provides an insight in time evolution of noise decomposition into reaction network. The tool is computationally effective even for vast biochemical models (compare Fig. 11) and can successfully provide a required information, see Fig. 3.

### 2.3 Sensitivity Analysis Based on Mutual Information

Another recently developed method for sensitivity analysis of multi-variables system has been oryginaly proposed by Lüdtke (Lüdtke et al. 2008). One of the biggest advantage of this method is its applicability to investigation of a model sensitivity to groups of parameters, and not only to single parameters, so it is not OAT method. Moreover this method provides an insight into interactions between parameters.

However, the approach proposed by Lüdtke is based on discrete variables and discrete entropy estimator. Consequently it requires computationally inefficient

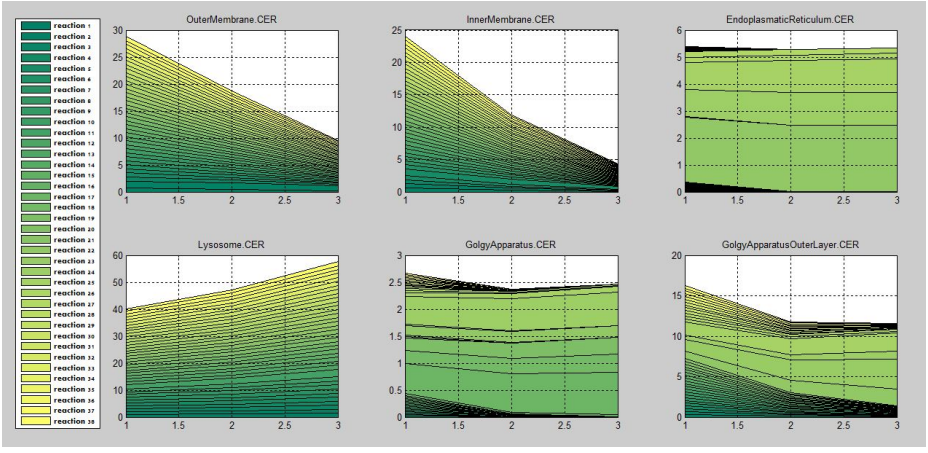


**Fig. 3.** Stochastic Noise Decomposition into single reactions vs. time for the Ceramide Metabolism Model - Scheme in Fig. 11

variable discretization procedure, that is highly biased and inefficient in high-dimensional space (i.e. in many parameters case). Having this concern in mind and due to fact that biochemical models deal with continuous measurements we propose to amend the method to continuous variables case.

The fundamental concept for sensitivity indices is mutual information Eq. (3) between random variables defined by parameters and random variables being model output. Let us denote by variable $X \sim g(x)$ model parameters and by variable $Y \sim f(y)$ model output, then the mutual information between this continuous random variables is defined by

$$I(X;Y) := \int_X \int_Y \log \frac{h(x,y)}{g(x)f(y)} h(x,y)dydx \tag{3}$$
$$= \mathbb{E}\left[\log \frac{h(x,y)}{g(x)f(y)}\right] = H(Y) + H(X) - H(X,Y),$$

where $g(x)$ and $f(y)$ denotes probabilities densities functions and $h(x,y)$ is the joint probability density function of joint random variable $(X,Y)$.

Measurements of MI is based on entropy estimation. In our approach to SA based on MI we use differential entropy Eq. (4), so there is no need of variables discretization.

$$H(X) := \int_X -\log g(x)dx = \mathbb{E}\left[-\log g(x)\right] \tag{4}$$

As a starting point for differential entropy estimation we used the k-th nearest neighbour entropy estimator. In order to achieve more reliable results we introduced more efficient $k$-nn differential entropy estimator for multivariate random variables. We have noticed the biased behaviour of the $k$-nn entropy estimator

Eq. in higher dimension and propose bias correction, which yields more accurate $k$-nn entropy estimates especially in higher dimensions. Our improved $k$-nn entropy estimator explore the idea of correcting the density function evaluation near to the boundary of random variable support.

**Definition 1** *Assume that $X_i$ are the parameters of the model and $Y$ is the model output, then* ***sensitivity indices*** *are defined as:*

$$I(X_i;Y) = H(Y) + H(X_i) - H(X_i,Y) = H(Y) - H(Y|X_i).$$

*Analogously,* ***sensitivity indices for pairs of parameters*** *are defined as:*

$$I(X_i,X_j;Y) = H(Y) + H(X_i,X_j) - H(Y,X_i,X_j) = H(Y) - H(Y|X_i,X_j).$$

The sensitivity indices reflect the impact of parameters on the model output, in other words this definition indicates correlations between parameters and the output. Definition 1 can be extended for any subset of parameters.

The group sensitivity index for a pair of parameters may have high value indicating the significant influence of these parameters to the model output, while two sensitivity indices for these two single parameters may in the same time have low value. We interpret such case as opposite -negative interaction between this pair of parameters, compare Fig. 4.

**Definition 2** *Let $X_i$ denote parameters of a model and $Y$ denote model output, then* ***interactions indices within pair of parameters*** *are defined by:*

$$I(X_i;X_j;Y) = \mathbb{E}_{X_i,X_j,Y}\left[-\log\frac{p(x_i)p(x_j)p(y)p(x_i,x_j,y)}{p(x_i,x_j)p(x_i,y)p(x_j,y)}\right]$$
$$= H(X_i) + H(X_j) + H(Y) - H(X_i,Y) - H(X_j,Y) - H(X_i,X_j) + H(X_i,X_j,Y)$$
$$= I(X_i;Y) + I(X_j;Y) - I(X_i,X_j;Y).$$

## 3 Biochemical Models

Within our research we concentrate on investigation and development of formal sensitivity analysis methods and also we implement and test the methods on various dynamical biochemical models. The complexity of a model depends on the number of variables and parameters and the kinetics defined in ODEs or SDEs, compare Fig. 5. In order to better understand and capture model features we test different standard and novel approaches.

### 3.1 Ligand-induced receptor model

In paper (Charzyńska et al., 2012) we focus on recently available methods of sensitivity analysis for dynamic biochemical models, such as local sensitivity analysis based on derivatives with regards to single parameters, and global sensitivity

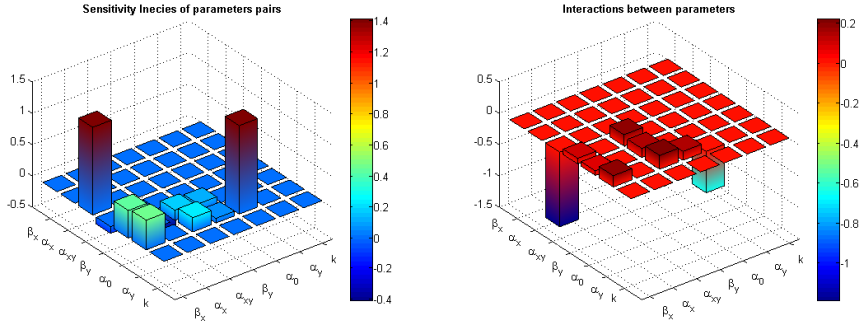


**Fig. 4.** Sensitivity analysis based on MI for the p53-Mdm2 negative feedback loop model - scheme of the model in Fig. 8. (Submitted to Entropy)

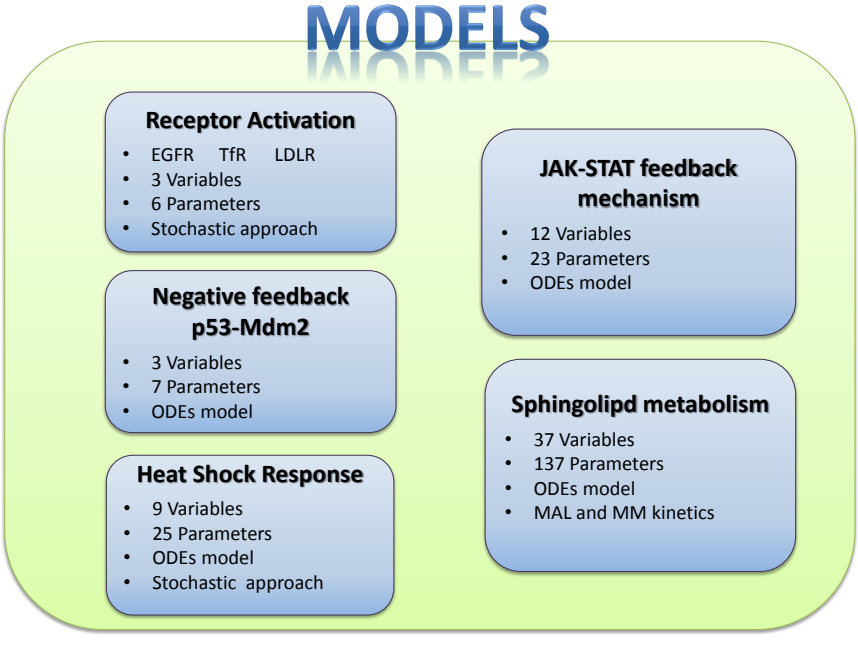


**Fig. 5.** Examined biochemical models

analysis methods i.e. variance decomposition, Fourier Amplitude Sensitivity Test (FAST), screening methods or stochastic methods. As an illustrative example of the presented ideas we consider the mathematical model of ligand-induced receptor system (Shankaran et al. 2007), see Fig. 6.

We transformed the classical deterministic version into a stochastic model. For both approaches, appropriate SA were applied. The model reflects a system of cell surface receptors in a single cell and describes the time evolution of three different species: ligands in the inter-cellular space, free receptors on a cell membrane and ligand-receptor complexes, see Fig. 6. The set of the model parameters contain also the volume of the inter-cellular space that falls for a single cell $V$ and the level of receptors concentration in the steady state $R_T$. We investigate four types of receptors:

- epidermal growth factor receptor, (EGFR), which stimulates cell division and plays an important role in the process of tumour formation,
- transferrin receptor (TfR), responsible for the transport of iron into cells,
- low-density lipoprotein receptor (LDLR), transporting cholesterol into cells,
- vitellogenin receptor (VtgR), which mediates the uptake of vitellogenin (Vtg) in oocyte development.

The results for sensitivity analysis based on Morris method were presented in Fig. 7. In three of four analysed receptors types the crucial parameters were $k_{off}$ and $k_{on}$ corresponding respectively to rate of complexes disintegration and complexes binding, as well $R_T$ corresponding to concentration of receptor in stationary state.

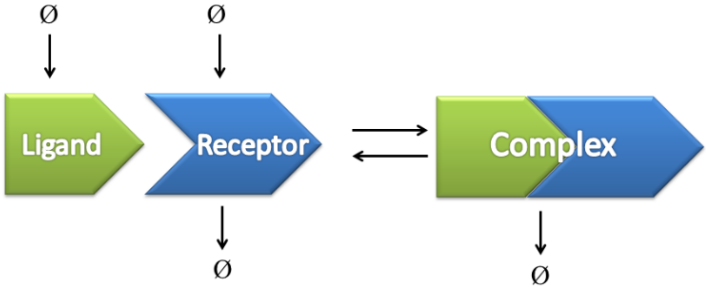


**Fig. 6.** Ligand-induced Receptor Activation Model

### 3.2 Negative feedback model of p53-Mdm2

To validate new approach of the global sensitivity analysis based on mutual information measure we tested the method on a well known and widely studied

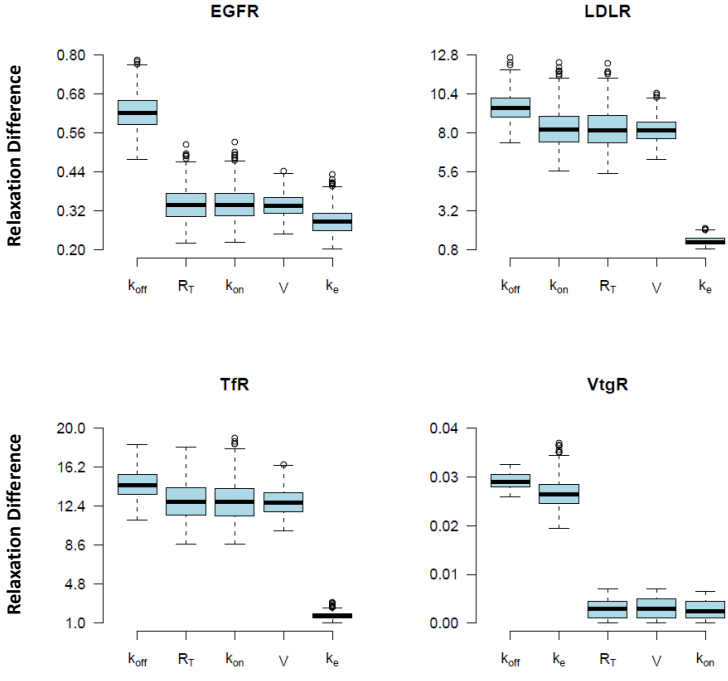


**Fig. 7.** SA based on Morris method for Ligand-induced Receptor Activation Model (Figure previously published in Biotechnologia by Charzyńska at al., 2012)

example of negative feedback loop of p53 protein and Mdm2 ligase (Zatorsky et al. 2006) for the model scheme see Fig. 8.

Tumour suppressor p53 protein also known as TP53 transcription protein 53 is a transcription factor determining the fate of a cell in case of DNA damage; p53 indirectly, via activation of transcription of the p21 gene encoding, can block cell cycle to repair DNA or activate a process of programmed cell death called apoptosis. The main regulator of the concentration of p53 protein is ligase Mdm2 / Hdm2 (double minute 2 mouse / human double minute 2), which through ubiquitination leads to degradation of p53 in the proteasome. In more than half of the cases of human cancers p53 is inactivated or absent, which allows the mutated tumor cells to replicate and determines their immortality. Consequently, this protein is under investigation due to its property to lead to self-destruction of cancer cells, which could be successfully used as therapy in many types of cancer.

By use of SA method based on MI we were able to capture the negative interactions between parameters $\beta_x$ and $\alpha_{xy}$ corresponding respectively to p53 inflow and Mdm2 negative loop, for the results of SA see Fig. 4.

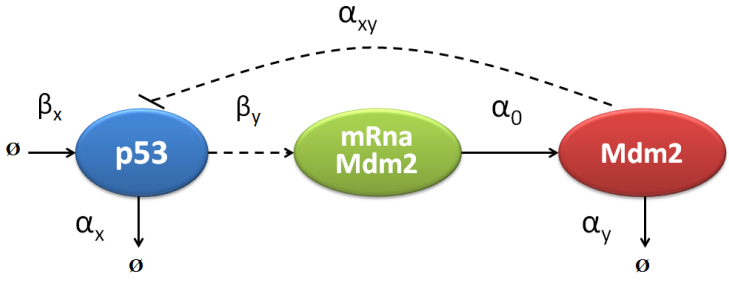


**Fig. 8.** Negative Feedback p53-Mdm2 Model

### 3.3 Heat shock response model

One of the most important questions in cell biology is how cells cope with rapid changes in their environment. The range of molecular responses includes a dramatic change in gene expression pattern and higher synthesis of so-called heat shock (or stress) proteins (HSPs). Induction of HSPs increases cell survival under stress conditions (Morimoto 1993). To test hypothesis about heat shock treatment we implemented and verify a mathematical model of heat shock protein synthesis induced by an external temperature stimulus (see Fig. 9), both in deterministic and stochastic meaner. The deterministic model consists of a system of nine non-linear ordinary differential equations describing the temporal evolution of the key variables involved in the regulation of HSP synthesis.

Computational simulations of the model were carried out for different external temperature stimuli. Stochastic version of the model was implemented by use of Chemical Master Equation. To validate the stochastic model output we used the indices of variance to mean ratio for all modelled species. The greatest variance to mean ratio both in homoeostasis and in heat shock case was scored by HSP indicating highest sensitivity to stimuli of rapid substrate concentration raise.

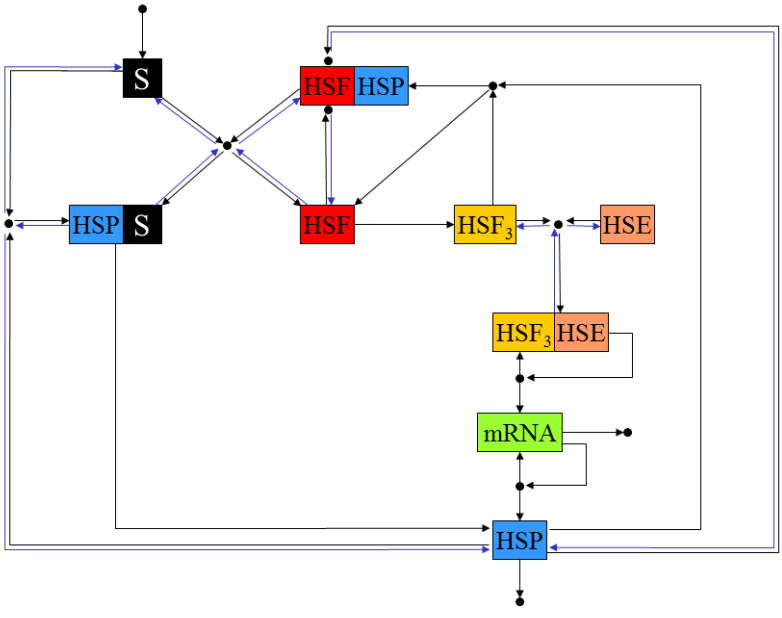


**Fig. 9.** Heat Shock Response Model

### 3.4 JAK-STAT feedback model

The paper (Gambin et al., 2013) is a review of computational models of JAK1/2-STAT1 signalling pathway. Despite conceptually simple mechanism of *JAK-STAT* signalling pathway it has highly complex behaviour. This model describes a control mechanism and factors influencing kinetics of the JAK-STAT pathway with increased IFN-$\gamma$ activity. The model is relatively complex as it captures all essential elements in the JAK1/2-STAT1 signaling. Scheme of the model is depicted in Fig. 10. The model can be informally divided into three modules: receptor module, transcription factor module (the STAT life-cycle) and post-translational feedback module.

Computational modelling is a tool to investigate complex molecular signalling pathways and formalize the description of the dynamics of the system. Computational models integrate experimental data with formal description of a modelled system and consequently they allow to test new hypotheses about interactions between modelled species. In paper (Gambin et al., 2013) we compared three different approaches to the modelling of JAK1/2-STAT1 phenomenon. The sensitivity analysis was useful not only to find the crucial parameters of any analysed JAK-STAT models, but we used it also as a tool to compare different modelling approaches.

steaming

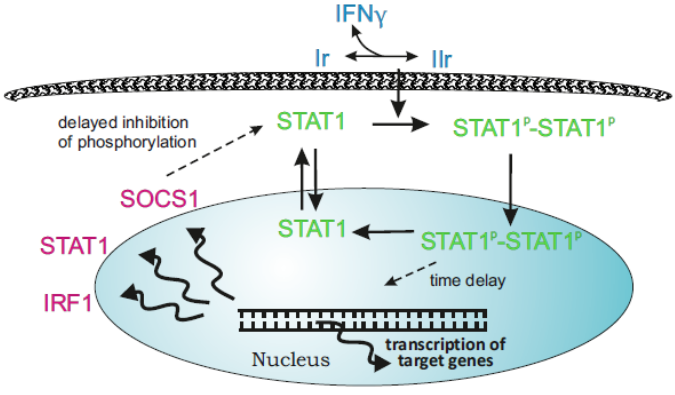


**Fig. 10.** JAK-STAT Feedback Model

### 3.5 Sphingolipid metabolism model

In paper (Wronowska et al., 2015) we propose the first comprehensive computational model of sphingolipid metabolism in human tissue. Contrary to the previous attempts, we use a model that reflects cell compartmentalization thereby highlighting the differences among individual organelles, see Fig. 11.

It has been proven that a significant role in the cell apoptosis pathway can be played the ceramides - bioactive lipids, members of sfingolipid family. The exact role of ceramides in signals transduction within nerve cells is still not fully explained. Our motivation was to formally describe an empirically observed correlation between ceramides concentration and cell viability response in human neuroblastoma SH-SY-5Y. Ceramides in low concentrations increase cell viability and may stimulate proliferation but in high concentrations ceramides induce cell apoptosis. One of the hypothesis which may explain the pro-survival role of ceramides in low concentrations is connection with sphingosine 1-phosphate

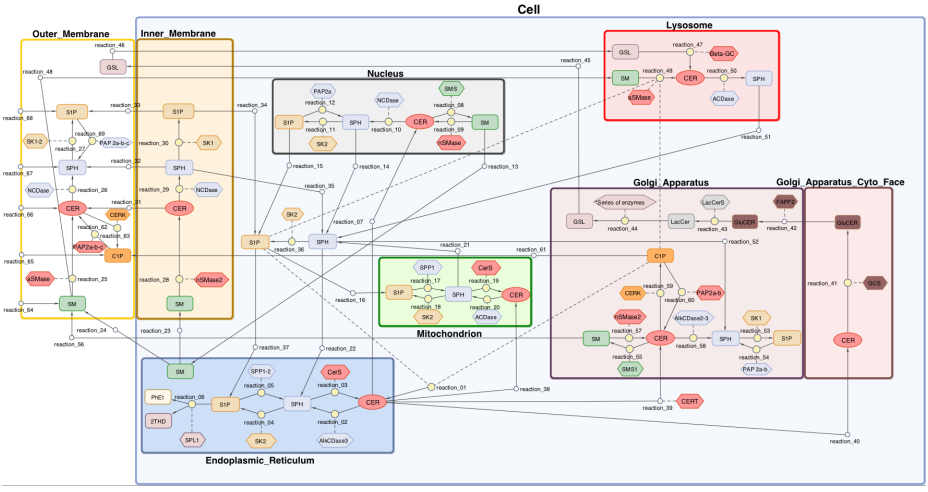


**Fig. 11.** Sphingolipid Metabolism Model (Figure previously published in BMC Systems Biology by Wronowska et al., 2015)

(S1P) synthesis involving sphingosine ceramide kinase activity. We aim to examine molecular mechanism of cell death evoked by ceramides within nerve cells for both pathological states: cancer and neurodegeneration.

The model that we had build was validated using recently proposed methods of model analysis, allowing to detect the most sensitive and experimentally non-identifiable parameters and determine the main sources of model variance. Moreover, we demonstrate the usefulness of our model in the study of molecular processes underlying Alzheimer's disease, which are associated with sphingolipid metabolism. This model allows to study sphingolipid metabolism disorders that have been observed in various pathological conditions such as cancer and neurodegeneration.

We performed local sensitivity analysis for this ODE model, but unfortunately due to the model complexity the method was of limited use and must have been complemented by the other SA method. The reason for limited applicability of LSA was due to its sensitiveness to the arbitrary choice of time horizon. In case of sphingolipid model we found useful the stochastic noise decomposition method based on LNA described in Section 2.2, for the results see Fig. 3. The StochDecomp method allowed to detect parameters of highest variance components and it complemented the LSA method.

## 4 Summary

There are plenty recently available SA methods that were developed over decades. We briefly recalled most popular SA methods in Section 2.1. Nevertheless each method has some limitations in its applicability to model assessment. Our in-

terest in development of new SA methods resulted from the need for efficient analytical tools to assess computational models that deeply explore natural complex phenomena. Section 3 contain description of some larger models examples of transcriptional signalling and metabolic pathways that were under our investigation. For this models classical methods such as local SA was of limited use duet to its applicability only to one at the time factor. Consequently it was not enough to perform simple LSA to understand all model dependencies. In case of larger networks (cf. model of sphingolipids metabolism Fig. 11) to understand the complex relations within modelled species and parameters we prefer to investigate all parameters at once, as they may interact one with another and they can have common impact to the model. In case of sphingolipid metabolism model we found useful to compare the results of LSA with the StochDecop output that let us identify the species with the greatest variance component resulting from different reactions.

We also focused on development of a new method based on MI in lieu of its discrete equivalent. This method seems to be promising as it can be applied to any subset of parameters and can provide the information about interactions within parameters groups with respect to the model output. We used this method to compare with the LSA results of the p53-Mdm2 negative feedback loop model. Contrary to LSA that can only provide the sensitivities of a single species to a single parameter separately, SA based on MI provide us with the information of sensitivity of global output to any subset o parameters and moreover with the information of parameters interactions.

To conclude there are many SA methods with different applications. LSA is most popular for biochemical dynamical models but it is a OAT method. The alternative GSA methods are usually computationally inefficient especially for stochastic version of biochemical models. The StochDecomp tool is a solution that by LNA can provide the variance decomposition steaming from separate reactions and can be easily applied to any biochemical model. The SA method based on MI can by applied to samples of data from continuous random variables and can provide the sensitivity indexing for any subset of parameters, consequently we found it useful for biochemical models.

## Acknowledgements

The paper is co-founded by the European Union from resources of the European Social Fund. Project PO KL "Information technologies: Research and their interdisciplinary applications", Agreement UDA-POKL.04.01.01-00-051/10-00.

## References

1. Andrews S.S. et al., *Detailed simulations of cell biology with smoldyn 2.1*, PLoS Comput. Biol., 6, e1000705, 2010
2. Bentele M., Lavrik I., Ulrich M., Stösser S., Heermann D.W., Kalthoff H., Krammer P.H., Eils R, *Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis*, J. Cell Biol. 166: 839-851, 2004

3. Butcher J.C. , *Numerical Methods for Ordinary Differential Equations*, Wiley, 2003
4. Campolongo F., Saltelli A., Tarantola S. , *Sensitivity analysis as an ingredient of modelin*, Statistic. Sci. 15: 377-395, 2000
5. Charzyńska A., Nałęcz A., Rybiński M., Gambin A., *Sensitivity Analysis of Mathematical Models of Signaling Pathways*, BioTechnologia 93(3), 291–308, 2012.
6. Charzyńska A., Gambin A., *Improvement of k-NN entropy estimator with applications in systems biology*, Entropy - submited.
7. Cukier R.I., Levine H.B., Shuler K.E., *Nonlinear sensitivity analysis of multiparameter model systems*, J. Comput. Phys. 26: 1-42, 1978
8. Degasperi A., Gilmore S., *Sensitivity analysis of stochastic models of bistable biochemical reactions*, Lect. Notes Comput. Sci. 5016: 1-20, 2008
9. Dorval A., *Probability distributions of the logarithm of inter-spike intervals yield accurate entropy estimates from small datasets*, Journal of Neuroscience Methods 173, 129-139, 2008
10. Gambin A., Charzyńska A., Ellert-Miklaszewska A., Rybiński M., *Computational models of JAK1/2-STAT1 signaling*, JAK-STAT 2:3, e24672, Landes Bioscience, 2013.
11. Gillespie D.T. , *A rigorous derivation of the chemical master equation*, Physica A: Stat. Mechan. Applic. 188: 404-425, 1992
12. Gillespie D.T. , *The chemical Langevin equation*, J. Chem. Phys. 113: 297-306, 2000
13. Goutsias J., *Classical versus stochastic kinetics modeling of biochemical reaction systems*, Biophys. J. 92: 2350-2365, 2007
14. Gunawan R., Cao Y., Petzold L., Doyle F.J., *Sensitivity analysis of discrete stochastic systems*, Biophys. J. 88: 2530-2540, 2005
15. Hornberger G.M., Spear R.C., *An approach to the preliminary analysis of environmental systems*, J. Environ. Manag. 12: 7-18, 1981
16. Jetka T., Charzyńska A., Gambin A., Stumpf M.P.H., Komorowski M., *StochDecomp - Matlab package for noise decomposition in stochastic biochemical systems*, Bioinformatics 30(1), 137–138, 2014.
17. Kampen N.V., *Stochastic Processes in Physics and Chemistry*, North Holland, Third Edition, 2007
18. Kazachenko L.F., Leonenko, N.N., *Sample Estimate of the Entropy of a Random Vector*, Probl. Peredachi Inf. 23(2), 9–16 1987.
19. Komorowski M. et al., *Decomposing noise in biochemical signalling systems highlights the role of protein degradation*, Biophys. J., 104, 1783-1793, 2013
20. Komorowski M. et al., *Bayesian inference of biochemical kinetic parameters using the linear noise approximation*, BMC Bioinformatics, 10, 343, 2009
21. Kraskov A., Stögbauer H., Grassberger P., *Estimating mutual information*, Phys. Rev. E 69, 066138, 2004; Erratum Phys. Rev. E 83, 019903, 2011
22. Lüdke N., Panzeri S., Brown M., Broomhead D.S., Knowles J., Montemurro M.A., Kell D.B., *Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks*, J. R. Soc. Interface 5(19), 223–235, 2008.
23. Lazo A., Rathie, P., *On the entropy of continuous probability distributions. Information Theory*, IEEE Transactions on 24, 120-122, 1978
24. Ma J., Sun Z., *Mutual Information is Copula Entropy* , http://arxiv.org/abs/0808.0845v1, 2008
25. Mahdavi A., Davey R.E., Bhola P., Yin T., Zandstra P.W., *Sensitivity analysis of intracellular signaling pathway kinetics predicts targets for stem cell fate control*, PLoS Comput. Biol. 3: e130, 2007

26. Morimoto R.I., *Cells in stress: transcriptional activation of heat shock genes*, Science 259, 1409-1410, 1993
27. Morris M.D., *Factorial sampling plans for preliminary computational experiments*, Technometr. 33: 161-174, 1991
28. Pérez-Cruz F., *Estimation of Information Theoretic Measures for Continuous Random Variables*, In: Advances in Neural Information Processing Systems 21 (NIPS), Vancouver, 2008
29. Ramaswamy R. et al., *Discreteness-induced concentration inversion in mesoscopic chemical systems*, Nat. Commun., 3, 779, 2012
30. Rateitschak K., Karger A., Fitzner B., Lange F., Wolkenhauer O., Jaster R., *Mathematical modelling of interferon-gamma signalling in pancreatic stellate cells reflects and predicts the dynamics of STAT1 pathway activity*, Cell Signal, 22:97-105, 2010
31. Raykar V.C., *Probability Density Function Estimation by different Methods*, ENEE 739Q SPRING 2002, course assignment 1 report, 1-8, 2002
32. Saltelli A., Ratto M., Andres T., Campolongo F., Cariboni J., Gatelli D., Saisana M., Tarantola S., *Global Sensitivity Analysis*, The Primer, Wiley-Interscience, 2008
33. Saltelli A., Ratto M., Tarantola S., Campolongo F., *Sensitivity analysis for chemical models*, Chem. Rev. 105: 2811-2828, 2005
34. Saltelli A., Tarantola S., Chan K.P.S., *A quantitative model-independent method for global sensitivity analysis of model output*, Technometrics 41: 39-56, 1999
35. Shankaran H. Resat H., Wiley H.S., *Cell surface receptors for signal transduction and ligand transport: a design principles study*, PLoS Comput. Biol. 3: 0986-0999, 2007
36. Shankaran H., Wiley H.S., Resat H., *Modeling the effects of HER/ErbB1-3 co-expression on receptor dimerization and biological response*, Biophys. J. 90: 3993-4009, 2006
37. Sjöberg P., Lötstedt P., Elf J., *Fokker-Planck approximation of the master equation in molecular biology*, Comput. Visual. Sci. 12: 37-50, 2009
38. Sobol I.M., Kucherenko S., *Global sensitivity indices for nonlinear mathematical models*, Review. Wilmott 1: 2-7, 2005
39. Sobol I.M., *Sensitivity analysis for nonlinear mathematical models*, Mathem. Model. Comput. Exper. 1: 407-414, 1993
40. Sricharan K., Raich R. III A.O.H., *Boundary Compensated k-NN Graphs*, Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2010
41. Sricharan K., Raich R., Hero A, *Optimized intrinsic dimension estimation*, Proc. IEEE Intl. Conf on Acoustics, Speech and Signal Processing (ICASSP), 5418-5421, 2010
42. Swameye I. et al., *Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling*, Proc. Natl Acad. Sci. USA, 100, 1028, 2003
43. Szymańska Z., Zylicz M., *Mathematical modeling of heat shock protein synthesis in response to temperature change*, Journal of Theoretical Biology 259(3), 562-569, 2009
44. Thomas P. et al., *Intrinsic noise analyzer: a software package for the exploration of stochastic biochemical kinetics using the system size expansion*, PLoS One, 7, e38518, 2012
45. Turanyi T. , *Sensitivity analysis of complex kinetic systems. tools and applications*, J. Mathem. Chem. 5: 203-248, 1990
46. van Kampen N.G., *Stochastic Processes in Physics and Chemistry*, North-Holland Personal Library, Elsevier Science, 2007

47. Vanlier J. et al., *An integrated strategy for prediction uncertainty analysis*, Bioinformatics, 28, 1130-1135, 2012
48. Wallace E. et al., *Linear noise approximation is valid over limited times for any chemical system that is sufficiently large*, IET Syst. Biol., 6, 102-115, 2012
49. Westerhoff H.V., Palsson B.O., *The evolution of molecular biology into systems biology*, Nature Biotechnology, 22, 1249-1252, 2004
50. Wolkenhauer O., Ullah M., Kolch W., Cho K.H., *Modeling and simulation of intracellular dynamics: Choosing an appropriate framework*, IEEE Trans. Nanobiosci. 3: 200-207, 2004
51. Wronowska w., Charzyńska A., Nienałtowski K., Gambin A., *Computational modeling of sphingolipid metabolism*, BMC Systems Biology 9:47, 2015
52. Yamada S., Shiono S., Joo A., Yoshimura A., *Control mechanism of JAK/STAT signal transduction pathway*, FEBS Lett 534:190-6, 2003
53. Yue H., Brown M., He F., Jia J., Kell D., *Sensitivity analysis and robust experimental design of a signal transduction pathway system*, Intern. J. Chem. Kinet. 40: 730-741, 2008
54. Zatorsky N.G., Rosenfeld N., Itzkovitz S., Milo R., Sigal A., Dekel E., Yarnitzky T., Liron Y., Polak P., Lahav G., Alon U, *Oscillations and variability in the p53 system*, Molecular Systems Biology 2: 2006.0033, 2006