Sensitivity and identifiability analysis of a third-order tumor growth model

Máté Siket, György Eigner, Levente Kovács

Abstract— The growing cancer cases attract more and more scientific research and introductions of new models, applied control algorithms and methods. The models are fundamental in the area of computer generated low-dose metronomic (LDM) chemotherapy, when the administration of the drug is ought to be optimized. Generally the in-silico tests and investigations are based on a model, which is hypothesized to describe the given process reliably and accurately. The analysis of the models and its parameters is crucial for therapy generation. We performed an analysis of a third-order tumor growth model based on sensitivity analysis and identifiability tests. The results show that a subset of parameters can be fixed as population values and the rest of the parameter sets results in an identifiable system with minor loss of accuracy.

I. INTRODUCTION

Cancer treatment and its related fields are intensively studied subjects, since it is the second leading cause of death globally [1], and it is predicted that there will be 21.4 million cases of cancer by 2030 [2]. The application of modern control algorithms to adjust the dosage of drugs has many unexploited potential [3]. With models accurately describing the underlying physiological processes control engineers have the possibility to define optimal therapies. System engineers thrive to provide a proper description of the patients, however the inter- and intrapatient variability and the effects of different drugs are cumbersome to model. Several models were introduced over the years [4]–[7]. One of them is a third-order model describing living and dead tumor cells and drug injections using pegylated liposomal doxorubicin [7].

In the case of physiological systems, it is of great importance to get an overall view and understanding of the model structure and its parameters. Identification is the key process when the dynamic model is ought to be tailored to a specific individual. However, other approaches are also present, e.g., online parameter estimation, a priori analysis of readily available data [8]. What is common in all these techniques is that they require knowledge about the identifiability of the underlying model and the parameter sensitivities.

Structural identifiability is crucial because it provides information about whether during ideal conditions certain parameters can be uniquely determined given the time series of the inputs and outputs of the system. Although in realistic scenarios, structural identifiability does not inherently guarantee the identifiability. If the system is not fully identifiable, meaning that only a subset of the parameters can be uniqely determined, the sensitivity analysis provides useful information about which parameters should be under focus, i.e., which have the largest impact on the output.

The paper is structured as follows: Section II describes the investigated tumor growth model. In Section III and IV we present the toolboxes and applied methods for the model analysis. In Section V the results are summarized. Finally, in Section VI the paper is concluded.

II. THE APPLIED TUMOR GROWTH MODEL

The investigated tumor growth model is a third-order nonlinear system. The model differentiates between the living and dead tumor cells by introducing separate state variables: x_1 [mm^3] and x_2 [mm^3], respectively. The third variable x_3 [mg/kg] describes the drug concentration. The dynamics is described by the equations given in [6] at time t:

$$\dot{x}_1(t) = (a - n)x_1(t) - b\frac{x_1(t)x_3(t)}{ED_{50} + x_3(t)}, \quad (1)$$

$$\dot{x}_2(t) = nx_1(t) + b\frac{x_1(t)x_3(t)}{ED_{50} + x_3(t)} - wx_2(t), \quad (2)$$

$$\dot{x}_3(t) = -c\frac{x_3(t)}{K_B + x_3(t)} - b_k\frac{x_1(t)x_3(t)}{ED_{50} + x_3(t)} + u(t), \quad (3)$$

where a describes the proliferation rate of the living tumor cells, n is the tumor cell necrosis rate, w is the dead tumor cell washout rate, c is the drug depletion rate. The pharmacodynamics is affected by the b and ED_{50} (median effective dose) parameters. The effect of the drug and the depletion of the drug is characterized by equations following the dynamics of Michaelis-Menten kinetics, thus ED_{50} and K_B are Michelis-Menten parameters.

The therapy can be only effective if it can decrease the tumor volume, which is determined by the value of the parameters. In order to get an insight, suppose that we have infinite dose of drug, thus the effect of the drug is described by

$$\lim_{x_3 \to \infty} \frac{x_3}{ED_{50} + x_3} = 1, \quad (4)$$

which is maximum. In this case, (1) reduces to

$$\dot{x}_1 = (a - n - b)x_1, \tag{5}$$

and due to the positivity of x_1 [7], the drug can decrease the tumor volume if and only if

$$0 > (a - n - b) \tag{6}$$

which results in the condition $b > a - n$.

It is beneficial to keep the term in (4) close to 1, since it maximizes the effect of the drug. On the other hand due to the harmful side effects, it is also beneficial to keep the drug level as low as possible. The optimal scheduling of injections is a distinct focus of the scientific field [9], [10].

III. SENSITIVITY ANALYSIS

Sensitivity provides information about the effects of parameter value variations on the output. We utilized the AMIGO2 Matlab toolbox [11] for the sensitivity analysis. This information can be used to reduce the number of parameters which need to be estimated online, while we can use population mean values for the other parameters. The analysis is usually performed using various measures on the difference between a nominal trajectory and a perturbed trajectory caused by changes in parameter values. The most typical measure is the root-mean-square error (RMSE), but the direction of change can be extracted using a normalized mean error as follows.

If we define a general nonlinear system in the form of (7), where p are the parameters, then the sensitivities s_p can be given with the Jacobians in (9).

$$0 = f(\dot{x}, x, p, u, t), \tag{7}$$

$$s_p = \frac{\partial x}{\partial p} \tag{8}$$

$$\frac{\partial s_p}{\partial t} = \frac{\partial f}{\partial x} s_p + \frac{\partial f}{\partial p}, \tag{9}$$

$$s_p(0) = 0, \tag{10}$$

$$S_{p,k} = \frac{p_i}{y(k)} s_{p,k}, \tag{11}$$

where $S_{p,k}$ is the normalized or relative sensitivity of parameter p in the $k-th$ discrete step, $y(k)$ is the output and p_i is the $i-th$ sample from the LHS. In (11) the toolbox normalizes with the value of the $i-th$ parameter sample and with the value of the output in the given $k-th$ discrete step:

$$\delta_p^{msqr} = \frac{1}{n_e n_{lhs} n_k} \sqrt{\sum_{e=1}^{n_e} \sum_{i=1}^{n_{lhs}} \sum_{k=1}^{n_k} (S_{p,k})^2} \tag{12}$$

where n_e is the number of different scenarios for experiments (initial conditions or input schemes), n_{lhs} is the number of sampled parameter values, n_k is the number of discrete points at which the system is evaluated.

The sampling of the parameter values in a given range are done by the Latin Hypercube Sampling (LHS), resulting in

equal distribution in the range and in the individual parameters. Because of the nature of the tumor growth, the system is unstable if untreated, and there are parameter combinations which result in an unstable system even in the case of high drug injections. To avoid the analysis of these combinations, after the sampling, the unstable patients are removed from the cohort. This results in a non-symmetric distribution of the a, b, n parameters, as it is otherwise expected with the LHS.

The resulting trajectories are dependent on the input, thus the sensitivities also. In order to investigate the effect of different kind of inputs, we applied two different drug administration scheme. The first is similar to a typical protocol, where the patients get the maximum tolerable dose every week [12]. The other is a metronomic therapy using much smaller doses. The doses for the second therapy are from the investigation performed by [13]. The results indicate that the optimal frequency of injections is between two and three days, with an injection of around $4.5 \cdot 10^{-3}$ [mg/kg/day].

The nominal values of the parameters and standard deviations (SDs) are calculated from an identification done on eight mice [6] these values are given in the Table I. The upper and lower bounds for the parameters used during the sensitivity analysis are defined as ± 2 SD.
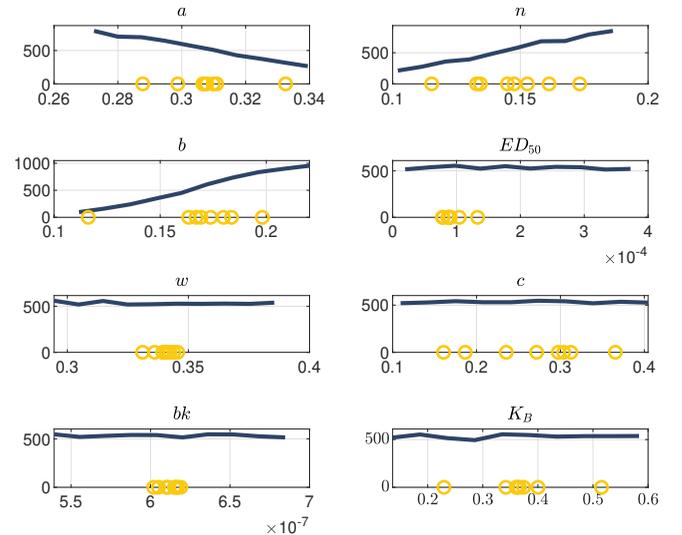

Fig. 1. Histograms of the constrained LHS parameter sets, showing the number of simulations with the given parameter values.

In Fig. 1, the histograms show the number of simulations with the corresponding parameter values. In the case of the a, b, n parameters the asymetric distribution can be seen and in the case of ED_{50} the +- 2 SD would have allowed negative values, thus it is constrained to be positive. The yellow circles show the parameters of the previously identified mice.

IV. IDENTIFIABILTY ANALYSIS

We use the definitions of structurally globally identifiable (s.g.i.) and structurally locally identifiable (s.l.i.) for the

Parameter	PLD1	PLD2	PLD3	PLD4	PLD5	PLD6	PLD8	PLD9	Nominal	SD
a [1/day]	0.333	0.307	0.307	0.310	0.289	0.299	0.308	0.311	0.306	0.0186
b [1/day]	0.116	0.169	0.198	0.180	0.163	0.184	0.174	0.167	0.166	0.0302
c [1/day]	0.235	0.297	0.304	0.272	0.312	0.365	0.187	0.161	0.257	0.0820
n [1/day]	0.115	0.148	0.153	0.173	0.134	0.161	0.133	0.145	0.144	0.0235
$b_k \left[\frac{10^{-7}mg}{kg \cdot day \cdot mm^3}\right]$	6.15	6.05	6.02	6.10	6.19	6.16	6.17	6.11	6.12	0.404
K_B [mg/kg]	0.367	0.361	0.342	0.230	0.362	0.374	0.515	0.400	0.36	0.1242
ED_{50} [10^-5mg/kg]	8.89	9.03	10.4	13.3	8.64	7.91	7.79	8.94	9.71	1.48
w [1/day]	0.346	0.344	0.331	0.341	0.341	0.339	0.336	0.342	0.34	0.0253

TABLE I

THE MEAN VALUES AND STANDARD DEVIATIONS OF THE IDENTIFIED PARAMETERS [6] BASED ON MICE EXPERIMENTS [12].

classification of systems [14] [15]. The structurally adverb indicates that the result is only valid from theoretical point of view. In practice due to noise and errors the identifiability is not guaranteed using this method. Globally identifiable means that there is only one unique parameter set that describes a given trajectory with a given input. On the other hand, in the case of local identifiability, there is a finite amount of parameter sets. If the system is non-identifiable, it means that there are infinite number of combinations. For the identifiability analysis the GenSSI toolbox [16] has been utilized.

If a dynamic system given in the form:

$$\dot{x} = f(x, p) + g(x)u, \tag{13}$$
$$y = h(x, p), \tag{14}$$

where x is the n-dimensional state vector, $\mathbf{f}$ and $\mathbf{g}$ are analytic vector fields. The output scalar field is given by $\mathbf{h}$ and u is the scalar input. This example is given with one input, however the method can be extended to multiple input systems.

Using the Lie-derivatives $\mathcal{L}$ of $\mathbf{x}$ along $\mathbf{f}$ and $\mathbf{g}$ the system can be written in the form [17]:

$$\dot{\mathbf{x}}(t) = \mathcal{L}_{\mathbf{f}}(\mathbf{x}(t), \mathbf{p})[\mathbf{x}(t)] + \mathcal{L}_{\mathbf{g}}(\mathbf{x}(t), \mathbf{p})[\mathbf{x}(t)]u(t). \tag{15}$$

Based on this representation the Nth order Peano-Baker series become:

$$\begin{aligned} \mathbf{y}(t) = \; & \mathbf{h}(\mathbf{x}(0), \mathbf{p}) + \\ & + \sum_{n=0}^{N} \sum_{k(0),...,k(n) \in \{\mathbf{f},\mathbf{g}\}} \left(\mathcal{L}_{\mathbf{k(0)}}...\mathcal{L}_{\mathbf{k(n)}} \; \mathbf{h}(\mathbf{x}, \mathbf{p})|_{\mathbf{x}(0)} \right. \\ & \left. \cdot \; \mathbf{s_{k(0)}}...\mathbf{s_{k(n)}} \right), \end{aligned} \tag{16}$$

where $\mathbf{s_f}$ represents $\int_0^t d\sigma$ and $\mathbf{s_g}$ represents $\int_0^t u(\sigma)d\sigma$. The task from identifiability point of view is to find kernels $\mathcal{L}_{\mathbf{k(0)}}(\tau)...\mathcal{L}_{\mathbf{k(n)}}(0)\mathbf{h}(\mathbf{x}(0), \mathbf{p})|_{\mathbf{x}(0)}$ which can be uniqely determined. The difficulty is that we do not know a priori the necessary order N of Lie derivatives. Moreover, the required high number of Lie derivatives can turn into computionally unfeasible situations. For example if $N = 2$, the combinations are limited: $\mathcal{L}_{\mathbf{f}}h$, $\mathcal{L}_{\mathbf{g}}h$ $\mathcal{L}_{\mathbf{f}}\mathcal{L}_{\mathbf{f}}h$, $\mathcal{L}_{\mathbf{f}}\mathcal{L}_{\mathbf{g}}h$, $\mathcal{L}_{\mathbf{g}}\mathcal{L}_{\mathbf{g}}h$, $\mathcal{L}_{\mathbf{g}}\mathcal{L}_{\mathbf{f}}h$. However it can be seen that with higher order of derivatives the combinations quickly become voluminous.

V. RESULTS

A. Sensitivity

The results of the sensitivity analysis are investigated with two different drug administration scheme as discussed earlier in Section III. However it is important to be certain about the consistency and reliability of the results, since usually based on the findings of the sensitivity analysis one continues with the selection of key parameter either during parameter estimation or identification. To this end consistency tests are done with different number of parameter samples, amount of injected drug and initial tumor volume. In each case the "nominal" values (on which the results were based) are the initial values and they are modified in increasing and decreasing directions.

Figure 2 showcases the global relative RMSDs of the model parameters with varying test parameters. It can be seen that the results slightly differ with only 100 samples, and above it remains constant, so the finally selected 1000 samples can be trusted from grid resolution point of view. The multiples of the optimized drug injections are also investigated. The ED_{50} and K_B parameters show a decreasing tendency and b an increasing tendency. This behavior can be explained if one observes the system equations, in particular the fraction in (4). If the input is large enough, it increases the drug concentration x_3 to an extent, that the fraction converges to one. This means that above a certain drug concentration, increasing the injections makes less difference. On the other hand, when the fraction saturated, the b parameter becomes the only variable as a scaling factor, thus its relative importance grows. The initial tumor volume is also modified with multiples of the nominal 40 [mm^3], the quasi horizontal lines indicate that the analysis is very robust with respect to the different initial conditions.

In Fig. 3 the normalized measures can be seen. The normalization is done in order to make each measure visible in a single plot. We are interested in the relative values given a measure, not the absolute ones. When ranking parameters usually the RMSD or MAD is preferred. The two measures in this case provide almost identical results. The most sensitive parameter is a, followed by b and n (with sensitivity half of a) which are very similar. The less sensitive ones are the w, c, K_B, ED_{50} in this order, bk becoming the least sensitive one. The mean difference gives insight about the sign of the change in the output.
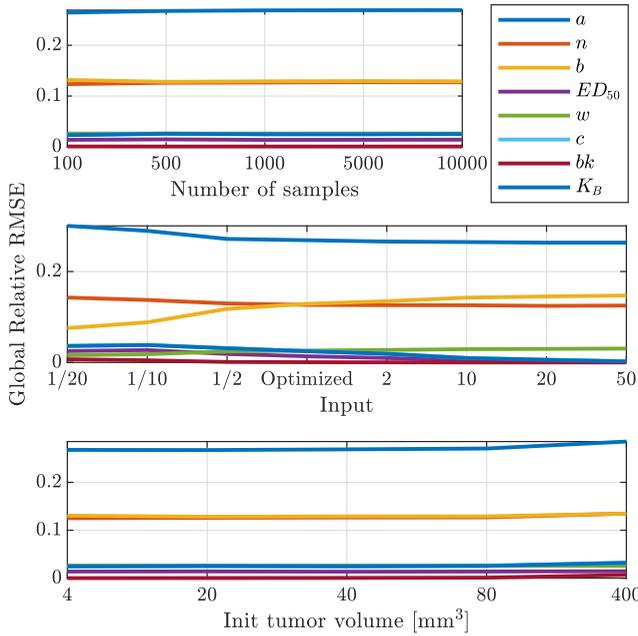
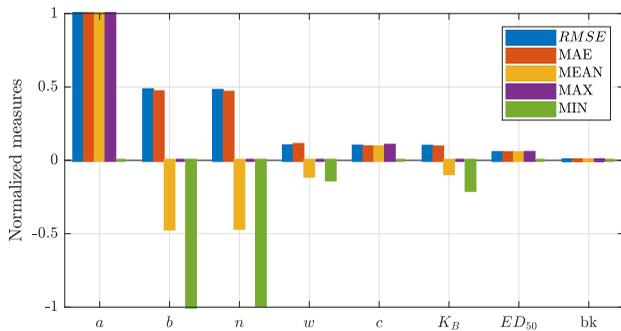Fig. 2. Consistency of the different simulation setups.

Fig. 3. Normalized, global relative sensitivities based on the optimized therapy.

Figure 4 depicts various drug concentration trajectories, when the model parameters are varied in the given range. The spikes occur every third day, when the patient receives the chemotherapeutic drug. The red lines show the drug concentration value when the fraction given in (4) equals to 0.95, for the highest ED_{50} value (upper line), for the lowest ED_{50} value (lower line). The two overlapping blue lines show the concentration when the fraction equals to the ED_{50} value.

Figure 5 depicts the results with the MTD therapy. Although the ranking of the most sensitive parameters, and their relative sensitivity to each other remained very similiar, there is a minor difference compared to the optimized therapy. The relative effect of the less sensitive parameters become even less, this can be explained due to the fact that the high input
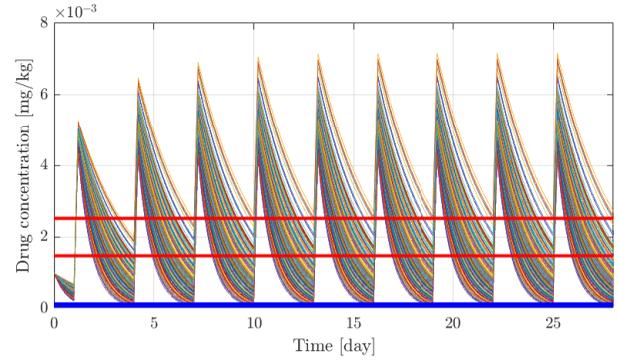

Fig. 4. Drug concentration trajectories in the population range.
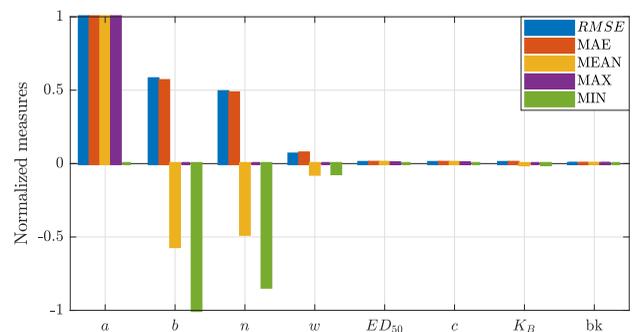

Fig. 5. Normalized, global relative sensitivities based on the MTD therapy.

values make the Michaelis–Menten kinetics saturate and their parameters become less relevant. This implies the importance of the reasonable input-parameter ratios, otherwise the model goes out of the physiologically relevant domain.

B. Identifiability

The assessment of identifiability requires high order Lie derivatives, due to the high number of model parameters. In order to be feasible computationally in each analysis only a subset of parameters are tested. The rest of the parameters are hypothesized to be a constant population value. The important parameters were prioritized based on the result of the sensitivity analysis.

Parameters	s.g.i.	s.l.i.
a, b, n, w	-	all
a, b, n, c	all	-
a, b, n, K_B	all	-
a, b, n, ED_{50}	all	-
a, b, n, b_k	all	-
a, b, n, w, ED_{50}	-	all
a, b, n, c, ED_{50}	a	b, n, c, ED_{50}

VI. CONCLUSIONS

In this paper we presented the sensitivity and identifiability analysis of a nonlinear third-order tumor growth model,

with ensuring the consistency and reliability of the results. The results clearly show that the system is significantly more sensitive to a subset of parameters. By selecting those parameters to be identified and the rest as nominal population values, the system is identifiable. Further identifications, parameter estimations and design of control algorithms can be enhanced based on these findings.

REFERENCES

[1] "WHO: Cancer," https://www.who.int/news-room/fact-sheets/detail/cancer, accessed: 2020-03-11.

[2] H.-P. Ren, Y. Yang, M. S. Baptista, and C. Grebogi, "Tumour chemotherapy strategy based on impulse control theory," *Philosophical Transactions Mathematical Physical & Engineering Sciences*, vol. 375, no. 2088, 2017.

[3] D. A. Drexler, T. Ferenci, and L. Kovács, "Extended tumor growth model for combined therapy," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Oct 2019, pp. 886–891.

[4] P. Hahnfeldt, D. Panigrahy, J. Folkman, and L. Hlatky, "Tumor Development under Angiogenic Signaling," *Cancer Research*, vol. 59, no. 19, p. 4770, Oct. 1999. [Online]. Available: http://cancerres.aacrjournals.org/content/59/19/4770.abstract

[5] A. d'Onofrio and A. Gandolfi, "Tumour eradication by antiangiogenic therapy: analysis and extensions of the model by Hahnfeldt et al. (1999)," *Mathematical Biosciences*, vol. 191, no. 2, pp. 159–184, Oct. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002555640400121X

[6] D. A. Drexler, T. Ferenci, A. Lovrics, and L. Kovács, "Modeling of tumor growth incorporating the effect of pegylated liposomal doxorubicin," in *in Proceedings of the 2019 IEEE 23nd International Conference on Intelligent Engineering Systems*. IEEE, 2019, pp. 369–373.

[7] D. A. Drexler, T. Ferenci, A. Lovrics, and L. Kovács, "Tumor dynamics modeling based on formal reaction kinetics," *Acta Polytechnica Hungarica*, vol. 16, no. 10, pp. 31–44, 2019.

[8] I. Hajizadeh, M. Rashid, S. Samadi, J. Feng, M. Sevil, N. Hobbs, C. Lazaro, Z. Maloney, R. Brandt, X. Yu, K. Turksoy, E. Littlejohn, E. Cengiz, and A. Cinar, "Adaptive and Personalized Plasma Insulin Concentration Estimation for Artificial Pancreas Systems," *Journal of Diabetes Science and Technology*, vol. 12, no. 3, pp. 639–649, May 2018. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1932296818763959

[9] F. Cacace, V. Cusimano, and P. Palumbo, "Optimal impulsive control with application to antiangiogenic tumor therapy," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 1, pp. 106–117, Jan 2020.

[10] D. A. Drexler and L. Kovács, "Optimization of impulsive discrete-time tumor chemotherapy," in *Proceedings of the 2019 IEEE 1st International Conference on Societal Automation.*, 2019.

[11] E. Balsa-Canto, D. Henriques, A. Gábor, and J. R. Banga, "AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology," *Bioinformatics (Oxford, England)*, vol. 32, no. 21, pp. 3357–3359, Nov. 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27378288

[12] A. Füredi, K. Szebényi, S. Tóth, M. Cserepes, L. Hámori, V. Nagy, E. Karai, P. Vajdovich, T. Imre, P. Szabó, D. Szüts, J. Tóvári, and G. Szakács, "Pegylated liposomal formulation of doxorubicin overcomes drug resistance in a genetically engineered mouse model of breast cancer," *Journal of Controlled Release*, vol. 261, pp. 287 – 296, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168365917307083

[13] B. Péceli, D. A. Drexler, and L. Kovács, "Optimal scheduling of low-dose metronomic chemotherapy: an in-silico analysis," in *Proceedings of the IEEE 15th International Conference on System of Systems Engineering*, 2020, accepted for publication.

[14] C. Cobelli and G. Romanin-Jacur, "Controllability, Observability and Structural Identifiability of Multi Input and Multi Output Biological Compartmental Systems," *IEEE Transactions on Biomedical Engineering*, vol. BME-23, no. 2, pp. 93–100, Mar. 1976. [Online]. Available: http://ieeexplore.ieee.org/document/4121014/

[15] A. F. Villaverde and A. Barreiro, "Identifiability of Large Nonlinear Biochemical Networks," p. 38.

[16] O. Chiş, J. R. Banga, and E. Balsa-Canto, "GenSSI: a software toolbox for structural identifiability analysis of biological models," *Bioinformatics*, vol. 27, no. 18, pp. 2610–2611, Sep. 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3167050/

[17] E. Walter and L. Pronzato, *Identification of Parametric Models: from Experimental Data*, ser. Communications and Control Engineering. London: Springer-Verlag, 1997. [Online]. Available: https://www.springer.com/gp/book/9783540761198