Nonlinear identification of a tumor growth model for validating cancer treatments

György Eigner*, Gábor Szögi [†], Péter Pausits [†], Imre Rudas [†], Levente Kovács *
*Physiological Controls Research Center and Research, Innovation, and Service Center,
Óbuda University, Budapest, Hungary,
Email: {eigner.gyorgy,kovacs.levente}@nik.uni-obuda.hu
[†]Antal Bejczy Center for Intelligent Robotics and Research, Innovation, and Service Center,
Óbuda University, Budapest, Hungary
Email: {gabor.szogi,peter.pausits}@irob.uni-obuda.hu, rudas@uni-obuda.hu
.

Abstract—In case of physiological related researches the appropriate adjustment of the parameters of the mathematical models describing biological phenomenons is a crucial issue. These models are essential in many research field such as the personalized health care or the control of physiological processes. Despite the available identification techniques there is no general solution in those cases where the mathematical model is given, but highly nonlinear in order to capture the main dynamical attitude of the physiological processes to be described. One of our aims was to develop such a general nonlinear identification framework which is flexible, can be easily used and supports the identification of these kind of models. We defined different metrics to measure the performance of the developed system. From the other hand, our goal was to successfully realize the identification framework in case of tumor growth beside anti-angiogenic treatment which is essential in our future work in order to validate the performance of advanced control algorithms. The results show that the nonlinear identification framework performed well in this case, since the predefined requirements from the applied metrics points of view were satisfied in all cases.

Index Terms—Identification, Nonlinear Least Mean Square Method, Tumor Growth, Tumor Model Identification

I. INTRODUCTION

Targeted Molecular Therapies (TMT) represent an advanced treatment opportunity beside the well known classical cancer therapies as radiotherapy, chemotherapy and surgical intervention. The TMTs – based on different drugs and/or other substances – are used to eliminate the tumor growth and spread in the body. TMTs interfering with specific molecules which play important roles in the physiological processes concerning to the growth, progression and spread of the cancer. Thus the main aim of TMTs is the inhibition of certain physiological processes and not to – directly – kill the tumor itself. Although, in many cases the inhibition of a specific process leads to the devastation of cancer cells. The main advantages of using TMTs are the limited side effects and the more focused therapy compared to the conservative treatments. Several TMTs exist,

albeit, the most important ones are the apoptosis inducers, signal transmission inhibitors, gene expression modulators and anti-angiogenic therapies [1].

One of the frequently applied TMT is the anti-angiogenic therapy [1]. The core concept in this case is to inhibit the growing of the supplying vasculature of the cancer concourse. It is known, that the supplying vasculature is needed in order to provide the oxygen and nutrients after the tumor grows beyond a certain volume due to the limited efficacy of local diffusion. Diffusion itself – after a given volume – is not able to satisfy the requirements of further growth from nutrition point of view and to maintain the reached size. By blocking the formation of new blood vessels the size of the cancer concourse can be decreased or kept below a given level which is tolerable from the human body point of view. Normally, the angiogenic processes are infrequent in healthy adults, but frequent concerning to tumor growth. By inhibiting this process only limited side effects occur compared to the conservative therapies.

The angiogenic processes are regulated by anti- and pro-angiogen factors. The most important pro-angiogen factor is the vascular endothelial growth factor (VEGF). The VEGF controls the endothelial proliferation – the reproduction of endothelial cells from which blood vessels are formed – which makes it excellent target for such inhibitor therapy [1]–[3].

In clinical practice, three approaches are used for drug delivery in case of anti-angiogenic therapy: bolus doses therapy (BDT), metronomic low dose therapy (MLDT) and continuous infusion therapy (CIT) [4]. In case of BDT the amount of the injected drug is the maximum dose which can be tolerated by the human body. Between the boluses there is no anti-angiogenic kind drug injection. The main disadvantages are the frequently occurring side effects and the remaining tumor cells may become resistant to the therapy due to their fast evolution and proliferation [5]. MLDT provides a better alternative at which the anti-tumor drugs are delivered in minimal dosage based on predefined schedule over longer periods [6]. The in silico- and animal-experiments have shown that application of CIT protocol can be the most effective treatment among the current anti-cancer therapies [7]–[9]. Albeit, to realize the CIT

protocol in practice as a personalized therapy precise models are indispensable. These models are also needed in order to validate the performance of advanced control algorithm in a personalized manner which promise a better outcome for the individuals compared to today's medical protocols.

To bridge this problem the most straightforward solution from control engineering point of view is to develop an identification process via the parameters of the existing mathematical models – which focus to the given phenomena – can be adjusted. In this way instead the expensive clinical trials a large population of virtual patients can be realized and the treatments coming from new medical protocols or the performance of the advanced control algorithms in this regard can be validated in silico before their validation in clinical environment. Accordingly, one of our goals was to develop such an identification framework which is able to deal with the afore mentioned issues, but also it is general and can be used regardless the physiological processes to be described and the mathematical model to be identified.

The paper is structured as follows: first, the applied tumor growth model is introduced. After, the developed identification framework is presented. In Sec. IV., we introduced the applicability of the identification framework in case of the introduced tumor growth model. Finally, we present our findings.

II. Applied Tumor Growth Model

We intended to use a well-known model in this regard which became a standard on the research field in the recent years. Due to we decided to apply the Hahnfeldt model [10] which describes the tumor growth beside angiogenic inhibition. The model is able to describe the dynamics of the growth of the tumor and the supporting vasculature as follows:

$$\dot{x}_1(t) = -\lambda_1 x_1(t) \log\left(\frac{x_1(t)}{x_2(t)}\right) \ , \tag{1}$$

$$\dot{x}_2(t) = bx_1(t) - dx_1^{2/3}(t)x_2(t) - \eta x_2(t)g(t) \ , \tag{2}$$

$$\dot{g}(t) = -\lambda_3 g(t) + u(t) \ , \tag{3}$$

where the $x_1(t)$ mm^3 is the tumor volume and $x_2(t)$ mm^3 is the volume of the supporting vasculature. The $g(t)$ mg/kg describes the antiangiogenic inhibitor serum level in time. The model's output is the $x_1(t)$ which is the only measurable state. The original model parameters were adjusted based on mice experiments (tumor: Lewis lung carcinoma, inhibitor: endostatin). These parameters are the following: $\lambda_1 = 0.1921$ 1/day, $b = 5.851$ 1/day, $d = 0.00871$ 1/(mm^2 day), $\eta = 0.66$ kg/(mg day) and $\lambda_3 = 1.31$ 1/day [10] .

The tumor growth model contains unfavorable mathematical properties due to the phenomena which is described by it. The model contains nonlinearities and nonlinear coupling between the states. Moreover, strong limitations make the handling of the model challenging, namely $x_1(t), x_2(t) > \forall t \ (t > 0)$ and feasibility problems can be occurred due to the $\log\left(\frac{x_1(t)}{x_2(t)}\right)$ term in (1) – which may causes $(0/0)$ kind singularity, when

$x_1(t)$ and $x_2(t)$ are equal to zero. Further, if $x_1(t)$ and $x_2(t)$ are close to zero numerical stability problems can be appeared.

It has to be mentioned that the Hahnfeldt model is a general model. Namely, it is appropriate to simulate tumor growth beside angiogenic inhibition regardless the type of the tumor and inhibitor. Furthermore, despite it was developed by using the considerations based on mice experiments, it can be used to describe the state of the human being as well after the appropriate adjustment of parameters of the model.

III. Identification Framework

We have built a convenient identification framework in the Mathworks' MATLAB system based on the so-called Nonlinear Least Squares Optimization (NLSO) method [11]–[13]. Naturally, this framework is general, thus it can be used in case of numerous nonlinear model identification problem.

During the realization we applied the MATLAB embedded NLSO algorithm. This algorithm – via $lqsnonlin$ function – is able to solve nonlinear least-squares curve fitting problems in the following form:

$$\min_{\mathbf{\Omega}} \|f(\mathbf{\Omega})\|_2^2 = \min_{\mathbf{\Omega}} \left(f_1(\mathbf{\Omega})^2 + f_2(\mathbf{\Omega})^2 + ... + f_n(\mathbf{\Omega})^2\right) \ , \tag{4}$$

where $f(\mathbf{\Omega})$ is the objective function to be minimized and $\mathbf{\Omega}$ is the parameter vector to be identified. Albeit, instead of computing the sum of squares the implemented $lsqnonlin$ function requires a user-defined vector valued function [13].

We considered a difference based objective function for the NLSO process in the following form:

$$f_1(\mathbf{\Omega}) = \sqrt{(\mathbf{z} - \hat{\mathbf{z}}(\mathbf{\Omega}))^2} \ , \tag{5}$$

where $\mathbf{z}$ is the measured data and $\hat{\mathbf{z}}(\mathbf{\Omega})$ is the simulated data based on the in silico experiments. We have tried other objective functions as well (absolute error function, norm-based functions, and others, but we experienced that (5) provided the best achievements according to the predefined metrics).

Consequently, (4) has to be modified accordingly (5) as follows:

$$\min_{\mathbf{\Omega}} \|f(\mathbf{\Omega})\|_2^2 = f_1(\mathbf{\Omega})^2 \ . \tag{6}$$

Which makes the developed framework easy-to-use is the implementation of the modeling environment. Each model to be identified has to be realized graphically in SIMULINK, however, the identification process (via the $lsqnonlin$ algorithm) is running from the MATLAB framework. In order to make these sub-systems compatible we developed different intermediate layers which doing the data synchronization and matching. The structure of the identification framework can be seen on Fig. 1.

In accordance with Fig. 1. the operation of the system can be decomposed into three sub-systems: *(i)* initialization process, *(ii)* operation of the NLSO framework and *(iii)* adjusting the model parameters.
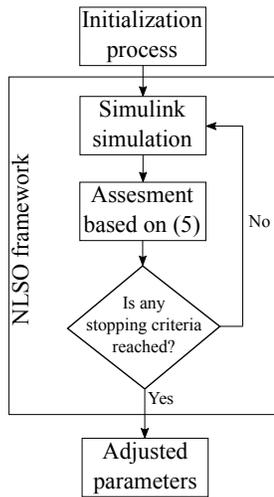
Figure 1. The core structure of the identification framework

During the initialization process all initial data have to be loaded as follows:

- NLSO related initialization: stopping criteria, type of the *lsqnonlin* algorithm and boundaries ($\mathbf{\Omega}_{ub}$, $\mathbf{\Omega}_{lb}$) of the model parameters. We predefined the following stopping criteria: MaxIter – maximum number of allowed iterations, FunTol – termination tolerance on the function value, StepTol – termination tolerance on the maximum step distance. All of the others were considered with their default values. We applied the following options: $\mathbf{O} = [\text{MaxIter,FunTol,StepTol,defaults}] = [600, 1e-12, 1e-12, \text{defaults}]$, where defaults mean the not predefined options. We selected the trust-region algorithm [14] to execute the NLSO process. $\mathbf{\Omega}_{ub}$ upper and $\mathbf{\Omega}_{lb}$ lower parameter boundaries (vectors) also have to be predefined.
- $\mathbf{\Omega}_{init}$ initial model parameters.
- Parameters of SIMULINK environment. These are the $\mathbf{x}(0)$ initial values of the model, T sampling time (has to be the same as the sampling of the $\mathbf{z}$ measurements), MaxStepSize (we applied MaxStepSize=1), parameters of the input (piecewise or continuous depend on circumstances of the measurements).

The operation of the iterative NLSO framework can be summarized as follows:

- Based on the initialized data the $\hat{\mathbf{z}}(\mathbf{\Omega})$ occurring via the realized simulation environment.
- Assessment is done according to the predefined metrics based on (5) to determine the goodness of the occurred $\hat{\mathbf{z}}(\mathbf{\Omega})$.
- Examination of the stopping criteria. If one of the stopping criteria is satisfied than the iterative operation of the NLSO framework terminates. Else, a new cycle is started.

The result of the NLSO based nonlinear identification process is the adjusted $\mathbf{\Omega}$ parameter set. The adjusted $\mathbf{\Omega}$ is that parameter set with which the $\hat{\mathbf{z}}(\mathbf{\Omega})$ simulated values are the closest to the $\mathbf{z}$ measurements.

IV. IDENTIFICATION OF THE MODEL

The adjustment of the parameters of the model was done by using measured data. The measurements coming from mice experiments (species: C57BL/6) at which C38 colon adenocarcinoma was implanted subcutaneously. The applied angiogenic inhibitor was bevazicumab [15]. The mice were divided into control group and case group accordingly the original goal of the investigation. The case group received inhibitor intake (injection) at each day during the treatment (the amount was $1/18$ mg/kg). The control group received one injection at the beginning of the treatment (the amount was 10 mg/kg) concerning to the existing medical protocols. We had $3-3$ measurements in the control and the case groups, respectively. In both cases, the experiments were 10 days long, thus, each measurement contains 10 data points [15], [16].

The original – literary – model parameters were adjusted based on measurement data by using Lewis lung carcinoma and endostatin inhibitor [10]. If the goal is to use arbitrary data source all model parameters have to be identified. Our goal was to adjust the $\mathbf{\Omega}_i = [\lambda_{1,i}, b_i, d_i, \eta_i, \lambda_{3,i}]^\top$ model parameter set from (1)-(3) by using the above detailed method.

During the identification we considered the followings regard to the applied initial parameter set, initial state values, initial parameters of the simulation framework and applied boundaries in both (control and case groups) cases:

- $\mathbf{\Omega}_{init} = [\lambda_{init}, b_{init}, d_i, \eta_i, \lambda_{3,i}]^\top = [0.1921, 5.851, 0.00871, 0.66, 1.31]^\top$. Namely, the $\mathbf{\Omega}_{init}$ initial parameter set was the same as the literary parameters accordingly [10]. This was a reasonable choice, thus these are validated and the numerical values are almost in the middle of the target range (bounded by $\mathbf{\Omega}_{lb}$ and $\mathbf{\Omega}_{ub}$).
- The initial values of $x_1(t_1)$ and $x_2(t_1)$ in all cases were the same as the initial values of the belonging $z(t_1)$, namely, $x_1(t_1) = x_2(t_1) = z(t_1)$. Although, the third initial state $(g(t_1))$ was considered as zero. As it was detailed above, the $u(t_1)$ initial inhibitor intake was different in the case group ($1/18$ mg/kg at each day) and control group (10 mg/kg at the beginning of the therapy).
- The applied lower and upper boundaries were $\mathbf{\Omega}_{lb} = [0.05, 2.5, 0.001, 0.1, 0.01]^\top$ and $\mathbf{\Omega}_{ub} = [1.1, 10, 0.04, 2.1, 0.9]^\top$, respectively – according to our previous researches these are appropriate ranges in case of this tumor type and mice model [15].

The effectiveness of the identification process was measured by three quality metrics. Each metric represents an other type of quality and by applying both of them provides a full picture about the goodness of all identified model. These metrics are usually used regard to the assessment of identification of physiological models [17]–[19].

The first one was the mean percentage variance accounted for (VAF) [18], [19] in the form of

$$\text{VAF}_{\text{mean}} = \text{mean}\left(\left(1 - \frac{\text{var}(\mathbf{z} - \hat{\mathbf{z}})}{\text{var}(\mathbf{z})}\right)\right) \times 100\% \ , \quad (7)$$

where $\mathbf{z}$ is the measured data and $\hat{\mathbf{z}}$ is the identified model output at given appropriate time instance. If $\mathbf{z}$ and $\hat{\mathbf{z}}$ are equal then the VAF parameter is 100% – in case of differences between them the VAF is lower. The second applied metric was the well-known root mean square error (RMSE) [20] in the following form:

$$\text{RMSE} = \sqrt{\frac{(\mathbf{z} - \hat{\mathbf{z}})^\top (\mathbf{z} - \hat{\mathbf{z}})}{N}} \ , \quad (8)$$

where N is the number of samples of the measured data. The last metric measures the fitness of the model by using the simple fit metric (SFM) [18] in percentage. The SFM can be calculated in the following way:

$$\text{SFM} = \left(1 - \frac{\sqrt{(\mathbf{z} - \hat{\mathbf{z}})^\top (\mathbf{z} - \hat{\mathbf{z}})}}{\sqrt{(\mathbf{z} - \bar{z})^\top (\mathbf{z} - \bar{z})}}\right) \times 100 \ , \quad (9)$$

where $\bar{z} = \text{mean}(\mathbf{z})$. If $\mathbf{z} = \hat{\mathbf{z}}$ then the SFM parameter is 100% – else, the SFM is lower.

The gray box based nonlinear identification process beside tight lower and upper bounds can be considered successful if the VAF_{mean} is not lower than 90%, the difference between $\mathbf{z}$ and $\hat{\mathbf{z}}$ is not higher than 10% at each compared data point and this reflects in the RMSE value and the SFM is higher than 75% [17], [19].

As aforesaid, we had $3-3$ measurements (both with $10-10$ data points) regard to the control and case groups, each data point represented 1 day. Due to in the simulation environment T sampling time was set to $T = 1$ and all simulations provided a 10 virtual days long simulated $\hat{\mathbf{z}}(\mathbf{\Omega})$ data vector.

The results of identification concerning to the control group can be seen on Fig. 2 whereon blue stars represent the $\mathbf{z}$ measurement vector and the red line is the linear interpolation between the $\hat{\mathbf{z}}$ simulated data points produced by the identified model beside the applied $\mathbf{\Omega}$ adjusted parameter set. As the results shown that the identification process was successful and provided acceptable results. In the second case a higher deviation appeared between $\mathbf{z}$ and $\hat{\mathbf{z}}$ mostly due to the different distribution of the data points. The upper part of Table I. shows the results of the numerical assessment based on the predefined VAF_{mean}, RMSE and SFM metrics. According to the VAF_{mean} low variance occurred in the difference between the $\mathbf{z}$ and $\hat{\mathbf{z}}$. Although, the numerical values regard to the RMSE metric seem high, but by taking into consideration the magnitudes of the measurements the results were satisfying. The SFM metric strengthen the previous results, namely, the identification framework provides the best fitting in the first and third cases.

Figure 3. shows the results of identification regarding to the case group (blue stars represent the $\mathbf{z}$ measurement vector and the red line is a linear interpolation between the $\hat{\mathbf{z}}$ simulated data points produced by the identified model beside the applied
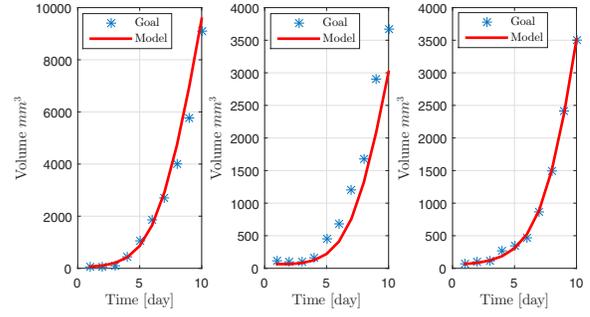

Figure 2. Results of identification based on the control group data

$\mathbf{\Omega}$ set). According to the diagram, the identification framework performed with lower accuracy, mostly because of the daily injections (1/18 mg/kg on each day) which was a higher external load than the previous case. Although, in all cases the results were acceptable and the identified models provided appropriate outcomes. The numerical values of the applied metrics can be seen in the bottom part of Table I. According to the VAF_{mean} low variance appeared in the difference between the $\mathbf{z}$ measurements and $\hat{\mathbf{z}}$ simulated values. The RMSE and SFM values shown that the value of the error was higher and the fitting of the identified model was less accurate than the previous case. Although, both values were acceptable respect to the magnitude of $\mathbf{z}$.
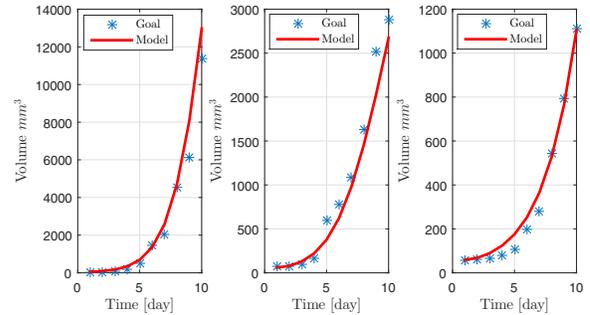

Figure 3. Results of identification based on the case group data

Table I
RESULTS OF THE IDENTIFICATION: CONTROL AND CASE GROUPS. EACH COLUMN REPRESENTS A GIVEN CASE

Control Group		
VAF$_{mean}$	98.6810%	95.2147%	99.9494%
RMSE	486.9614	394.7545	37.0242
SFM	82.8796%	67.3902%	96.6476%
Case Group		
VAF$_{mean}$	97.5985%	98.1631%	99.3287%
RMSE	815.7506	198.7766	42.6905
SFM	76.8989%	79.7959%	87.7860%

V. CONCLUSION

In this paper we described a convenient nonlinear identification framework which can be used concerning to physiological model identification. We systematically introduced the operation of the developed system and the applied metrics for the assessment. The other aim of the paper was to adjust the parameters of a well-known tumor growth model beside angiogenic inhibition via the developed process. From the results it is clear that the nonlinear identification framework performed well in this case, since the predefined requirements from the applied metrics points of view were satisfied in all cases. One of the main challenges was the low number of samples in the individual measurements. This limitation is a consequence of the tolerance of the mice respect to given higher tumor volume and the strict rules of the animal experiments due to ethical restrictions. In our future work we will improve the introduced framework in order to increase the accuracy of the identified model. Moreover, we are going to investigate the performance of the developed tool by using measurements in which more data points are available.

REFERENCES

[1] N. Vasudev and A. Reynolds, "Anti-angiogenic therapy for cancer: current progress, unresolved questions and future directions," *Angiogenesis*, vol. 17, no. 3, pp. 471–494, 2014.

[2] B. Al-Husein, M. Abdalla, M. Trepte, D. DeRemer, and P. Somanat, "Anti-angiogenic therapy for cancer: An update," *Pharmacotherapy*, vol. 32, no. 12, pp. 1095–1111, 2012.

[3] Y. Kubota, "Tumor angiogenesis and antiangiogenic therapy," *Keio J Med*, vol. 61, pp. 47 – 56, 2012.

[4] D. Drexler, J. Sápi, and L. Kovács, "Potential Benefits of Discrete-Time Controller-based Treatments over Protocol-based Cancer Therapies," *ACTA Pol Hung*, vol. 14, no. 1, pp. 11–23, 2017.

[5] O. Scharovsky, L. Mainetti, and V. Rozados, "Metronomic chemotherapy: changing the paradigm that more is better," *Curr Oncol*, vol. 16, no. 2, pp. 7–15, 2009.

[6] T. Browder, C. Butterfield, B. Kraling, B. Shi, B. Marshall, M. O'Reilly, and J. Folkman, "Antiangiogenic scheduling of chemotherapy improves efficacy against experimental drug-resistant cancer," *Cancer Res*, vol. 60, no. 7, pp. 1878–1886, 2000.

[7] J. Sápi, L. Kovács, D. Drexler, P. Kocsis, D. Gaári, and Z. Sápi, "Tumor volume estimation and quasi-continuous administration for most effective bevacizumab therapy," *PLoS ONE*, vol. 10, no. 11, 2015.

[8] J. Sápi, D. Drexler, I. Harmati, Z. Sápi, and L. Kovács, "Qualitative analysis of tumor growth model under antiangiogenic therapy: choosing the operating point and design parameters for controller design," *Optim Contr Appl Met*, vol. 37, p. 848–866, 2015.

[9] J. Sápi, D. Drexler, and L. Kovács, "Comparison of Protocol-based Cancer Therapies and Discrete Controller-based Treatments in the Case of Endostatin Administration," in *SMC 2016 - IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, pp. 3830–3835.

[10] P. Hahnfeldt, D. Panigrahy, J. Folkman, and L. Hlatky, "Tumor development under angiogenic signaling: A dynamical theory of tumor growth, treatment response, and postvascular dormancy," *Cancer Res*, vol. 59, pp. 4770–4775, 1999.

[11] C. Kelley, in *Iterative Methods for Optimization*, ser. SIAM Frontiers in Applied Mathematics. SIAM, 1999, no. No. 18.

[12] T. Strutz, *Data Fitting and Uncertainty (A practical introduction to weighted least squares and beyond)*, 2nd ed. Heidelberg, Germany: Springer Verlag, 2016.

[13] Mathworks Inc., *Optimization Toolbox. Revised for Version 7.6 (Release 2017a)*, 2017. [Online]. Available: https://www.mathworks.com/help/pdf_doc/optim/optim_tb.pdf

[14] T. Coleman and Y. Li, "Trust Region Approach for Nonlinear Minimization Subject to Bounds," *SIAM J Optim*, vol. 6, pp. 418 – 445, 1996.

[15] J. Sápi, L. Kovács, D. Drexler, P. Kocsis, D. Gajári, and Z. Sápi, "Tumor volume estimation and quasi-continuous administration for most effective bevacizumab therapy," *PLoS ONE*, vol. 10, no. 11, pp. 1 – 20, 2015.

[16] D. Drexler, J. Sápi, and L. Kovács, "A minimal model of tumor growth with angiogenic inhibition using bevacizumab," in *SAMI 2017 - IEEE 15th International Symposium on Applied Machine Intelligence and Informatics*. IEEE, p. 185–190.

[17] J. Bronzino and D. Peterson, *The Biomedical Engineering Handbook*, 4th ed. Boca Raton, Florida, USA: CRC Press, 2015.

[18] M. Tárnik, "T1DM model parameters adjustment based on PK, PD and CGM data," in *INES 2015 - IEEE 19th International Conference on Intelligent Engineering Systems*. IEEE, pp. 313 – 318.

[19] O. Nelles, *Nonlinear System Identification*, 1st ed. Heidelberg, Germany: Springer Verlag, 2001.

[20] W. Levine, Ed., *The Control Engineering Handbook*, 2nd ed. Boca Raton, USA: CRC Press, Taylor and Francis Group, 2011.