2015 IEEE International Conference on Systems, Man, and Cybernetics

Tumor model identification and statistical analysis

Johanna Sápi*, Tamás Ferenci*, Dániel András Drexler† and Levente Kovács‡
*Óbuda University, Research and Innovation Center of Óbuda University,
Physiological Controls Group,
Bécsi út 96/B H-1034 Budapest, Hungary
Email: {sapi.johanna, ferenci.tamas, kovacs.levente}@nik.uni-obuda.hu
†Budapest University of Technology and Economics,
Department of Control Engineering and Information Technology
Magyar tudósok krt 2. H-1117 Budapest, Hungary
Email: drexler@iit.bme.hu

***Abstract*—Tumor growth model identification under antiangiogenic therapy is a very current issue since the existing models in the literature have some limitations and usually they are not clinically validated. We have carried out animal experiments to observe valid data; mice were transplanted with C38 colon adenocarcinoma and they were treated with bevacizumab. Two groups were created, control group was treated according to the protocol, while case group members receive much lower doses daily. We created fixed and mixed models for the groups. Mixed models differs from fixed ones in random effects – in the case of mixed models both the intercept and the slope are random variables. These models are appropriate when the aim is to model not the concrete subjects in the sample, but rather, to describe the imagined population from which the samples were coming.**

***Index Terms*—tumor growth, identification, mixed model, C38 colon adenocarcinoma, bevacizumab**

I. Introduction

Antiangiogenic therapy [1] is a special type of tumor treatment, which inhibits angiogenesis. Angiogenesis is the process of forming new blood vessels; normally it occurs in the human body only at specific times (e.g. in case of wound healing). Tumor cells can break through this strict control and become able to form own blood vessels, which is essential for survival after a certain tumor size ($1-2$ mm diameter). The aim of antiangiogenic therapy is to prevent tumors from forming new blood vessels, because without angiogenesis tumor growth is inhibited [2], [3]. Bevacizumab (Avastin) [4] is a drug for antiangiogenic therapy, which inhibits the biological activity of human VEGF (vascular endothelial growth factor) [5].

II. Materials and Methods

A. Experimental Settings

In our experiment eight weeks old male C57Bl/6 mice with implanted C38 colon adenocarcinoma was used applying bevacizumab treatment. A piece of tumor was transplanted subcutaneously in the recipient animal on the 1st day of the experiment. Two groups were created: control and case groups. Control group (5 mice) received bevacizumab in one dose for an 18-day treatment according to the protocol (200 μg bevacizumab with 455 μl 0.9% NaCl solution) intraperitoneally on the -1st day and on the 17th day. Case group (9 mice) received one-tenth dose of control dose intraperitoneally spread over 18 days (1.11 μg bevacizumab with 45 μl 0.9% NaCl solution) every day from the -1st day of the experiment. The treatment period was 20 days.

Tumor volume was measured in two different ways. First way is the digital caliper measurement; in that case tumor diameters (width, length) are measured with caliper. It can be carried out in vivo during the experiment because of the subcutaneous localization of the tumor. Tumor volume (and the third diameter) has to be approximated, assuming a certain shape for the tumor. Measurements with caliper were done on the 0th, 2nd, 4th, 6th, 8th, 10th, 12th, 14th, 16th, 18th and 19th days of the experiment. The other way to measure tumor volume is the usage of small animal MRI, a non-invasive in vivo technology giving the possibility of a more precise volume measurement [6]. Measurements with small animal MRI were done on the 0th, 4th, 7th, 11th, 14th and 19th days of the experiment.

B. Model Description

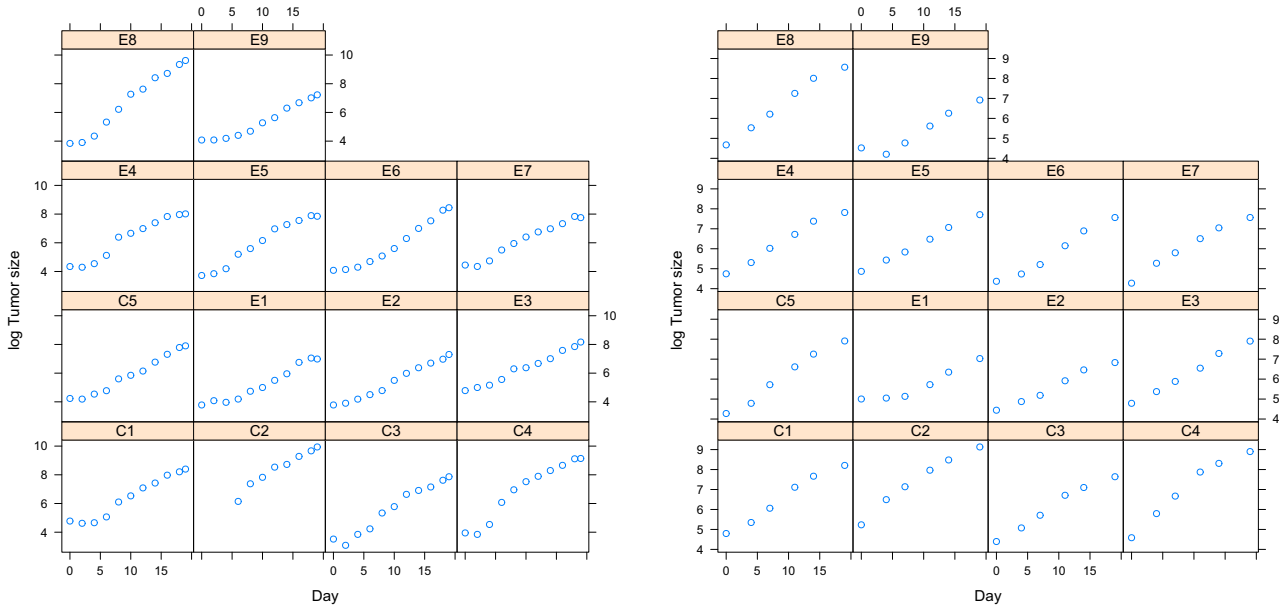
Statistical analysis was carried out in R program package [7] version 3.1.2 with a custom script developed for this purpose that is available from the corresponding author on request. For the mixed effects model, R library nlme [8] version 3.1-118 was also used.

The growth of the tumors exhibit exponential shape irrespectively of the animal and the measurement method, thus, the logarithm of the tumor size will be used in subsequent calculations to achieve linearity. The growth patterns are illustrated on Figure 1.

The tumor growth curves were modelled with mixed effects models, which is one of the most widely used tools in analyzing longitudinal data [9], [10], [11]. Such growth curves represent clustered data (with the repeated measurements within the same subject forming clusters), which can be effectively handled with mixed models [12].

More specifically, it was presumed that the logarithm of the tumor size exhibits linear growth in time, but both the intercept and the slope may possibly depend on the group (control or

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.434



(a) Caliper

(b) MRI

Fig. 1. Growth of the tumors in each animal shown with the logarithm of tumor size, both with caliper and MRI measurement method

case) of the mice. This was allowed to grab the potential impact of the treatment. Thus, the **fixed part** of the regression was

$$\text{logSize}_{i,t} = \beta_0 + \beta_{0\text{Case}}\text{Case}_i + \beta_1 t + \beta_{1\text{Case}}\text{Case}_i t + \varepsilon_{ij}, \quad (1)$$

where i is the subject (mouse), t is the time (in days, since the initiation of the study), ε_{ij} is the error term and Case_i is a dummy variable indicating if i^{th} subject belongs to the case group, viz.

$$\text{Case}_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ subject belongs to the case group} \\ 0 & \text{if the } i^{\text{th}} \text{ subject belongs to the control group} \end{cases}$$

In other words, the intercept in the control group is β_0, in the case group it is $\beta_0 + \beta_{0\text{Case}}$, the slope is β_1 in the control group and $\beta_1 + \beta_{1\text{Case}}$ in the case group (i.e. $\beta_{0\text{Case}}$ and $\beta_{1\text{Case}}$ is the difference in the intercept and slope, respectively, of the case group, compared to the controls).

Next, those coefficients were allowed to vary between subjects, by adding a random effect, thus the **mixed model** is

$$\begin{aligned}\text{logSize}_{i,t} = {} & (\beta_0 + b_{0,i}) + (\beta_{0\text{Case}} + b_{0\text{Case,i}})\,\text{Case}_i + \\ & + (\beta_1 + b_{1,i})\,t + (\beta_{1\text{Case}} + b_{1\text{Case,i}})\,\text{Case}_i t + \varepsilon_{ij}.\end{aligned} \quad (2)$$

This means a similar regression model, but with coefficients being random variables – coming from a distribution – instead of being fixed numbers. This model is appropriate when the aim is to model not the concrete subjects in the sample, but rather, to describe the imagined population from which the samples were coming. This structure also allows to account for the intra-individual correlations characteristic of longitudinal data. Here, the estimated quantity is not the individual parameters of a subject but rather the variances of the b random components – which is a single number independently of the sample size. Thus, a mixed model represents a kind of compromise between fitting a – global – fixed regression to every subject (parsimonious, but neglects individual differences) and an – individual – fixed regression to each subject itself (provides a good fit even in the presence of individual differences, but prevents inferring on the population of the subjects).

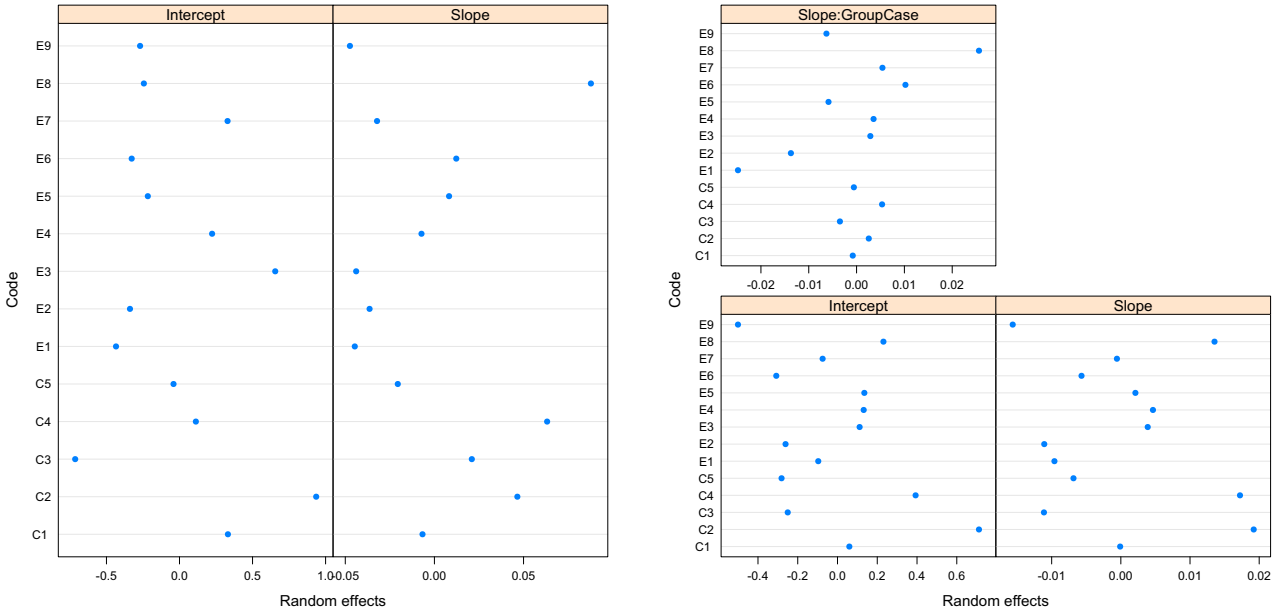
It was presumed that $\begin{pmatrix} b_{0,i} & b_{0\text{Case,i}} & b_{1,i} & b_{1\text{Case,i}} \end{pmatrix}^T \sim \mathcal{N}(0, \Psi)$ – i.e. they have normal distribution – that is, the random effects are allowed to have arbitrary correlation structure. (However, random effects of different subjects are assumed to be independent.) Also, its is assumed that $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ independent and identically distributed, and also independent of the random effects [12].

Usual model diagnostics – inspecting residual vs. fitted value plot to look for signs of heteroskedasticity and QQ-plot for residual normality, also stratified according to groups – were performed in all cases.

III. Results

A. Model Identification for Caliper Measurements

For the caliper measurements, the case and control groups did not differ significantly, neither in intercept ($p = 0.5668$), nor in slope ($p = 0.1703$). After removing these, the intercept – the initial tumor volume – was 3.85 (47.0 in the original scale) and the slope was 0.23 (that is: 26.4% increase in tumor volume each day, $p < 0.0001$). The random effects had a standard deviation of 0.48 for the intercept and 0.044 for the



(a) Caliper (b) MRI

Fig. 2. Growth of the tumors in each animal, both with caliper and MRI measurement method

slope. The random effects are shown on Figure 2(a) for each subject. The residual standard deviation was 0.29.

B. Model Identification for MRI Measurements

For the MRI measurements, the case and control groups did differ significantly, but only in terms of slope, that is, the growth rate of the tumor ($p = 0.003$), but not in the intercept, the initial tumor volume ($p = 0.343$). Therefore the difference in the intercept was removed from the model; in the resulting final model, the intercept – the initial tumor volume – was 4.62 (101.5 in the original scale) and the slope was 0.20 (22.2% increase in tumor volume each day, $p < 0.0001$) in the control group with the difference between the two groups being -0.038 ($p = 0.0047$), thus the growth rate in the case group was 17.7% each day. The random effects had a standard deviation of 0.34 for the intercept, 0.013 for the slope and 0.019 for the interaction. The random effects are shown on Figure 2(b) for each subject. The residual standard deviation was 0.34.

The overall fit of the mixed effects models – indicating the prediction both with the fixed part-only and with the whole mixed model – is shown on Figure 3.

IV. Discussion

Knowing the general equations of the created models, we have to investigate separately the resulted models and the biological meaning of the different variables. Besides these, we have to transform the logarithmic scale models into linear ones for control system identification.

$\beta_{0\mathrm{Case}}$ describes the difference in the intercept between control and case groups:

if $\beta_{0\mathrm{Case}} < 0$	then the initial tumor volume is smaller in case group than control group
if $\beta_{0\mathrm{Case}} > 0$	then the initial tumor volume is larger in case group than control group.

$\beta_{1\mathrm{Case}}$ describes the difference in the slope between control and case groups:

if $\beta_{1\mathrm{Case}} < 0$	growth rate is slower in case group than control group
if $\beta_{1\mathrm{Case}} > 0$	growth rate is faster in case group than control group.

A. Identified Models for Caliper Measurements

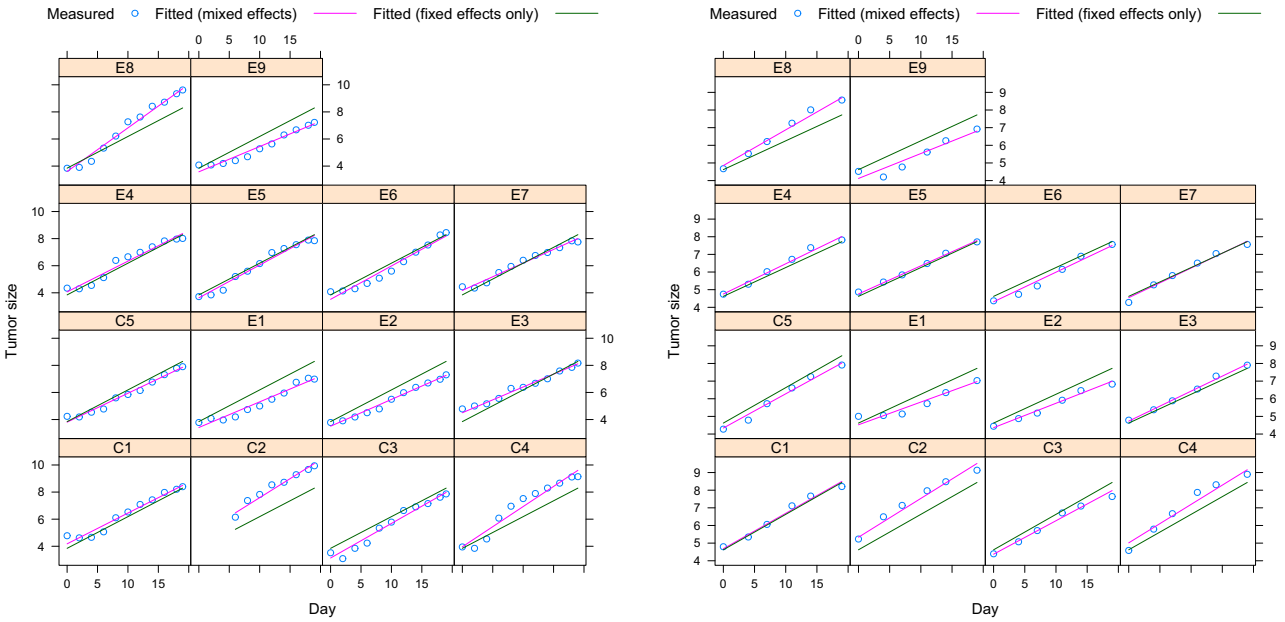
1) Fixed Model for Caliper Measurements: In the case of caliper measurements, the case and control groups did not differ significantly, neither in intercept, nor in slope, therefore $\beta_{0\mathrm{Case}} = 0$ and $\beta_{1\mathrm{Case}} = 0$.

$$\log\mathrm{Size}_{i,t}^{\mathrm{caliper,fixed}} = \beta_0 + \beta_1 t + \varepsilon_{ij} = 3.78 + 0.24 \cdot t + \varepsilon_{ij}, \quad (3)$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, 0.29)$.

$$\mathrm{linSize}_{i,t}^{\mathrm{caliper,fixed}} = 43.8 \cdot e^{0.24t} \cdot \varepsilon_{ij}^{\mathrm{lin}}, \quad (4)$$

where $\varepsilon_{ij}^{\mathrm{lin}}$ is the error term in the linear model, $\varepsilon_{ij}^{\mathrm{lin}} \sim \mathcal{LN}(0, 0.29)$ ($\mathcal{LN}$ stands for lognormal distribution).



(a) Caliper (b) MRI

Fig. 3. Predictions of the mixed models, with the fixed part shown separately

2) Mixed Model for Caliper Measurements: Since there was no difference in the intercept and slope between control and case groups, these terms cannot have random effects ($b_{0\text{Case,i}} = 0$, $b_{1\text{Case,i}} = 0$).

$$\begin{aligned}\text{logSize}_{i,t}^{\text{caliper,mixed}} &= (\beta_0 + b_{0,i}) + (\beta_1 + b_{1,i})\, t + \varepsilon_{ij} = \\ &= (3.85 + b_{0,i}) + (0.23 + b_{1,i})\, t + \varepsilon_{ij},\end{aligned} \tag{5}$$

where $b_{0,i} \sim \mathcal{N}(0, 0.48)$ and $b_{0,i} \sim \mathcal{N}(0, 0.044)$.

$$\text{linSize}_{i,t}^{\text{caliper,mixed}} = \left(47.0 \cdot b_{0,i}^{\text{lin}}\right) \cdot e^{(0.23+b_{1,i})t} \cdot \varepsilon_{ij}^{\text{lin}}, \tag{6}$$

where $b_{0,i}^{\text{lin}} \sim \mathcal{LN}(0, 0.48)$.

B. Identified Models for MRI Measurements

1) Fixed Model for MRI Measurements: Since MRI technique provides much more precise tumor volume measurement than caliper method, we have identified difference in the slope between control and case groups. As the difference have negative value, we can state that the growth rate is slower in case group than control group, i.e. our alternative administration was more effective than the protocol-based treatment. In the intercept there was no difference ($\beta_{0\text{Case}} = 0$), which is correct according to the fact that both group's members was transplanted by the same method and the first treatment happened just one day before the first measurement.

$$\begin{aligned}\text{logSize}_{i,t}^{\text{MRI,fixed}} &= \beta_0 + \beta_1 t + \beta_{1\text{Case}}\text{Case}_i t + \varepsilon_{ij} = \\ &= 4.62 + 0.21t - 0.06 \cdot \text{Case}_i t + \varepsilon_{ij},\end{aligned} \tag{7}$$

where $\varepsilon_{ij} \sim \mathcal{N}(0, 0.34)$.

$$\text{logSize}_{i,t}^{\text{MRI,fixed,control}} = 4.62 + 0.21t + \varepsilon_{ij}. \tag{8}$$

$$\text{logSize}_{i,t}^{\text{MRI,fixed,case}} = 4.62 + 0.15t + \varepsilon_{ij}. \tag{9}$$

$$\text{linSize}_{i,t}^{\text{MRI,fixed,control}} = 101.5 \cdot e^{0.21t} \cdot \varepsilon_{ij}^{\text{lin}}, \tag{10}$$

where $\varepsilon_{ij}^{\text{lin}} \sim \mathcal{LN}(0, 0.34)$.

$$\text{linSize}_{i,t}^{\text{MRI,fixed,case}} = 101.5 \cdot e^{0.15t} \cdot \varepsilon_{ij}^{\text{lin}}. \tag{11}$$

2) Mixed Model for MRI Measurements: Since there was no difference in the intercept between control and case groups, this term cannot have random effect ($b_{0\text{Case,i}} = 0$).

$$\begin{aligned}\text{logSize}_{i,t}^{\text{MRI,mixed}} &= (\beta_0 + b_{0,i}) + (\beta_1 + b_{1,i})\, t + \\ &+ (\beta_{1\text{Case}} + b_{1\text{Case,i}})\, \text{Case}_i t + \varepsilon_{ij} = \\ &= (4.62 + b_{0,i}) + (0.20 + b_{1,i})\, t + \\ &+ (-0.038 + b_{1\text{Case,i}})\, \text{Case}_i t + \varepsilon_{ij},\end{aligned} \tag{12}$$

where $b_{0,i} \sim \mathcal{N}(0, 0.34)$, $b_{1,i} \sim \mathcal{N}(0, 0.013)$ and $b_{1\text{Case,i}} \sim \mathcal{N}(0, 0.019)$.

$$\text{logSize}_{i,t}^{\text{MRI,mixed,control}} = (4.62 + b_{0,i}) + (0.20 + b_{1,i})\, t + \varepsilon_{ij}. \tag{13}$$

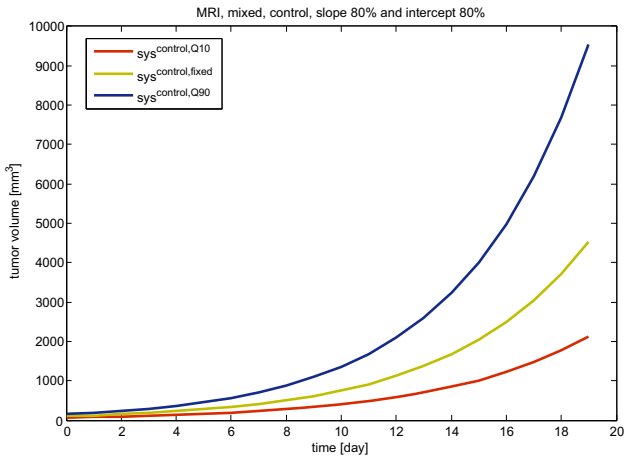


Fig. 4. Impulse response of the control group between $Q10$ and $Q90$ quantiles

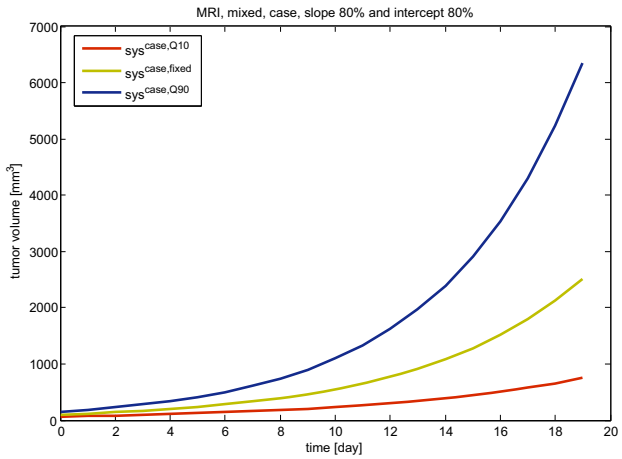


Fig. 5. Impulse response of the case group between $Q10$ and $Q90$ quantiles

$$\begin{aligned}\text{logSize}_{i,t}^{\text{MRI,mixed,case}} &= (4.62 + b_{0,i}) + \\ &+ (0.162 + b_{1,i} + b_{1\text{Case,i}})\, t + \varepsilon_{ij},\end{aligned} \tag{14}$$

where $b_{1,i} + b_{1\text{Case,i}} = b_{11\text{Case,i}} \sim \mathcal{N}(0, 0.026)$.

Finally, the linear mixed models for MRI measurements:

$$\text{linSize}_{i,t}^{\text{MRI,mixed,control}} = \left(101.5 \cdot b_{0,i}^{\text{lin}}\right) \cdot e^{(0.20+b_{1,i})t} \cdot \varepsilon_{ij}^{\text{lin}}, \tag{15}$$

where $b_{0,i}^{\text{lin}} \sim \mathcal{LN}(0, 0.34)$.

$$\text{linSize}_{i,t}^{\text{MRI,mixed,case}} = \left(101.5 \cdot b_{0,i}^{\text{lin}}\right) \cdot e^{(0.162+b_{11\text{Case,i}})t} \cdot \varepsilon_{ij}^{\text{lin}}. \tag{16}$$

C. Control System Identification

1) Control System Identification for Control and Case Groups: A linear dynamic model in time domain is written in the following form (state space representation):

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{17}$$

$$y(t) = Cx(t) + Du(t), \tag{18}$$

where (17) defines the dynamics of the system (x is the state variable and u is the input of the system), while (18) defines the output of the system.

The state space model of the control group, using the mixed model of MRI measurements (based on Equantion (15)):

$$\dot{x}(t) = (0.20 + b_{1,i})\, x(t) + u(t) \tag{19}$$

$$y(t) = \left(101.5 \cdot b_{0,i}^{\text{lin}}\right) x(t), \tag{20}$$

where $b_{1,i} \sim \mathcal{N}(0, 0.013)$ and $b_{0,i}^{\text{lin}} \sim \mathcal{LN}(0, 0.34)$.

The state space model of the case group, using the mixed model of MRI measurements (based on Equantion (16)):

$$\dot{x}(t) = (0.162 + b_{11\text{Case,i}})\, x(t) + u(t) \tag{21}$$

$$y(t) = \left(101.5 \cdot b_{0,i}^{\text{lin}}\right) x(t), \tag{22}$$

where $b_{11\text{Case,i}} \sim \mathcal{N}(0, 0.026)$ and $b_{0,i}^{\text{lin}} \sim \mathcal{LN}(0, 0.34)$

2) Investigating the Effect of the Random Variables: To simulate the effect of the random variables of the MRI mixed model, we calculated two quantiles of the distributions, hence we defined a range and we investigated the extrema of the range. To take into account the 80% of the elements, we calculated the $Q10$ and $Q90$ quantiles (which splits off the lowest 10% of data, and the highest 90% of data, respectively).

The mixed model for the control group, using the discrete value of $Q10$ quantile both for slope and intercept random variables:

$$\text{sys}^{\text{control,Q10}} = \left(101.5 \cdot Q10(b_{0,i}^{\text{lin}})\right) \cdot e^{(0.20+Q10(b_{1,i}))t}, \tag{23}$$

where $Q10(b_{0,i}^{\text{lin}}) = 0.6404$ and $Q10(b_{1,i}) = -0.0167$.

Similarly, using $Q90$ quantile both for slope and intercept random variables:

$$\text{sys}^{\text{control,Q90}} = \left(101.5 \cdot Q90(b_{0,i}^{\text{lin}})\right) \cdot e^{(0.20+Q90(b_{1,i}))t}, \tag{24}$$

where $Q90(b_{0,i}^{\text{lin}}) = 1.5314$ and $Q90(b_{1,i}) = 0.0167$.

Omiting the random variables, the "fixed" model for the case group:

$$\text{sys}^{\text{control,fixed}} = 101.5 \cdot e^{0.20t}. \tag{25}$$

The mixed model for the case group, using the discrete value of $Q10$ quantile both for slope and intercept random variables:

$$\text{sys}^{\text{case,Q10}} = \left(101.5 \cdot Q10(b_{0,i}^{\text{lin}})\right) \cdot e^{(0.162+Q10(b_{11\text{Case,i}}))t}, \tag{26}$$

where $Q10(b_{11\text{Case,i}}) = -0.0333$.

Similarly, using $Q90$ quantile both for slope and intercept random variables:

$$\text{sys}^{\text{case,Q90}} = \left(101.5 \cdot Q90(b_{0,i}^{\text{lin}})\right) \cdot e^{(0.162+Q90(b_{11\text{Case,i}}))t}, \tag{27}$$

where $Q90(b_{11\text{Case,i}}) = 0.0333$.

Omiting the random variables, the "fixed" model for the case group:

$$\text{sys}^{\text{case,fixed}} = 101.5 \cdot e^{0.162t}. \tag{28}$$

One can see the impulse responses of the systems in Figure 4 for control group, and in Figure 5 for case group. We plan further investigations on the random effect, however it is clear that this mixed model approach is a new, promising method in tumor growth identification.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support by the Óbuda University research fund. Levente Kovács was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

[1] H. C. Wu, C. T. Huang, and D. K. Chang, "Anti-angiogenic therapeutic drugs for treatment of human cancer," *J Cancer.*, vol. 4(2), pp. 37–45, 2008.

[2] D. M. McDonald, "Angiogenesis and vascular remodeling in inflammation and cancer: Biology and architecture of the vasculature," in *Angiogenesis: An Integrative Approach from Science to Medicine*, W. D. Figg and J. Folkman, Eds. Springer Science+Business Media, LLC, 2008.

[3] A. Hoeben, B. Landuyt, M. Highley, H. Wildiers, A. T. Van Oosterom, and E. A. De Bruijn, "Vascular endothelial growth factor and angiogenesis," *Pharmacol Rev.*, vol. 56, pp. 549–580, 2004.

[4] Genentech, "Prescribing information of Avastin (Bevacizumab)," http://www.gene.com/download/pdf/avastin_prescribing.pdf, 2013.

[5] European Medicines Agency, "Scientific discussion of Avastin," http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000582/WC500029262.pdf, 2005.

[6] V. Koo, P. W. Hamilton, and K. Williamson, "Non-invasive in vivo imaging in small animal research," *Cell Oncol*, vol. 28(4), pp. 127–139, 2006.

[7] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/

[8] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team, *nlme: Linear and Nonlinear Mixed Effects Models*, 2014, r package version 3.1-118. [Online]. Available: http://CRAN.R-project.org/package=nlme

[9] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, *Applied longitudinal analysis*. John Wiley & Sons, 2012.

[10] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*. Springer Science & Business Media, 2009.

[11] P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger, *Analysis of longitudinal data*. Oxford University Press, 2002.

[12] J. C. Pinheiro and D. M. Bates, *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, 2000.