Parameter estimation of complex mathematical models of human physiology using remote simulation distributed in scientific cloud

Tomáš Kulhánek*,[1], Marek Mateják[1], Jan Šilar[1]
and Jiří Kofránek[1]

***Abstract*— A generic system for estimation of model parameters —calibrate models— is introduced. The proposed system architecture is built of several loosely coupled modules behaving as RESTful web services and allowing to integrate other parts of the system via HTTP protocol and data exchanged in JSON format. The system was designed in such a way that the most demanding computational part is computed in parallel and computation may be distributed to remote computational resources. A test deployment was done in scientific cloud provided by czech NGI CESNET. Parameter identification of complex models got significant speedup on cloud computing resources.**

I. INTRODUCTION

There are several methodologies and technologies how to model a complex reality in biological domain. The large-scale mathematical description of physiological systems was introduced by Guyton et al. in 1972[1] and continues today by Hester et al. who introduced HumMod - a derivative of the Guyton's model and in-house modeling language and simulation tool[2][3].

Kofranek et al. implemented Guyton's 1972 model in MATLAB®Simulink[4]. However, the complexity of the model increased from Guyton's model to HumMod and it becomes too complicated keeping the model up-to-date using block oriented tools like MATLAB®Simulink. Therefore an implementation of the HumMod model was introduced in object oriented Modelica language[5][6]. Modelica is an acausal object oriented language introduced by Fritzson et al.[7] and it is curently mantained by the international Modelica association and implemented by several vendors[8].

The parameter identification is a task to estimate the unknown model parameters in such a way that the model simulation fits the experimental data[9]. The objective of this task could be for example to minimize the following function (least squares):

$$f(\vec{p}) = \sum_{i=1}^{n} (M(t_i, \vec{p}) - d(t_i))^2 \rightarrow min \qquad (1)$$

where $\vec{p}$ is vector of values of parameters, $M(t_i, \vec{p})$ is model simulated at time t_i with the given parameter values $\vec{p}$ and $d(t_i)$ is the measured experimental value at time t_i.

The models of human physiology are in general set of linear and non-linear algebraic and differential equations, some of them may change it's behavior based on discrete conditions, thus output of such model can be non-differentiable and non-continuous. Thus global optimization methods that work without derivatives has to be used in general to find minimum of the objective function. There were tested several global optimization methods to identify parameters of models in Modelica language[10][11]. The simulation must be performed many times using this methods and it may take extremely long for complex models or larger space of parameters. The computation time can be however reduced using distributed computing techniques. Maffioletti et al. introduced GC3Pie framework and shown workflow to identify parameters using grid computing[12]. Humphrey et al. calibrated hydrology models utilizing cloud computing [13]. We proposed a system which should support the process of parameter identification mentioned in above scientific publication in more generic way, so the researcher may focus on experimental data, selecting the parameters from a model to identify and interpreting the estimates during computation and hide the technical details of configuring the computational modules in distributed systems.

The proposed system integrates visualization, identification algorithm and simulation into loosely coupled modules opened to any modeling and numerical technology. We implemented the system and tested it with the models of human physiology in Modelica language, we selected genetic algorithm as a global optimization method for parameter identification and we distributed simulation into scientific cloud.

II. SYSTEM DESCRIPTION

The proposed system as seen on Figure 1 consists of several loosely coupled distinct pieces of software modules which can be replaced by another technology or implementation with no or minimal intervention into other related modules. The communication among modules is done via HTTP protocol and endpoints follows REST architectural style[14].

The simulation module consist of Modelica model exported as a Functional Mockup Unit (FMU) conforming the standardized Functional Mockup Interface[16]. The FMU is in fact DLL library for MS Windows platform. We wrapped this FMU by the ServiceStack [17] framework to provide web interface and control the simulation via HTTP protocol and JSON format. The simulation module can be deployed in multiple instances and each one can be executed in parallel.

*corresponding author tomas.kulhanek@lf1.cuni.cz
[1]Institute of Pathological Physiology, First Faculty of Medicine, Charles University in Prague, Czech Republic

978-1-4799-2131-7/14/$31.00

TABLE I

TIME AND SPEEDUP OF ESTIMATING THE PARAMETERS OF MODELS.

CPUs involved	1	2	3	4	10	20	30	40	100
model HumMod[6]									
time	71d21h	38d1h	25d11h	20d8h	7d16h	3d21h	2d10h	1d20h	18h
speedup	–	1.8x	2.8x	3.5x	9.3x	18.5x	29.7x	39.2x	95.8x
model Rossi-Bernardi[15]									
time	50 min	27 min	20 min	19 min					
speedup	–	1.8x	2.5x	2.6x					

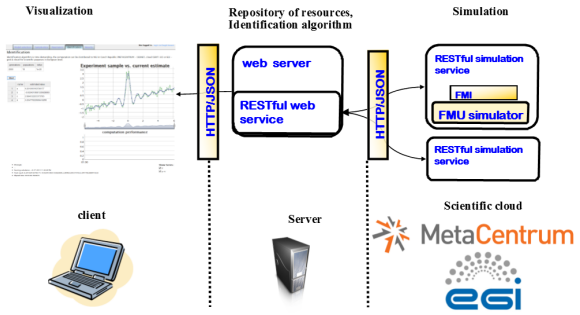

Fig. 1. Architecture of the system for parameter estimation

The simulation module instance registers it's unique endpoint URL to the repository module.

The repository module is a RESTful web service which provides storage of several data entities and provides access to the data via HTTP protocol and JSON format. The parameters are presented in web application as a table of names and values, the experiment and simulation data are visualized in a graph.

The identification algorithm is executed within this module and is used for model parameter estimation. We selected genetic algorithm from MATLAB-Optimization toolbox exported into a DLL library able to be executed with Matlab Common Runtime (MCR) environment. Each step of the algorithm produces a vector of parameter values which needs to be simulated and can be computed parallely by the instances of simulation modules.

The visualization module is implemented in HTML (version 5) and Javascript. This module gives user a list of available models. The table for experimental data allows to insert manualy or to copy&paste data from a desktop application. The table for parameters defines names, initial value and estimated maximal and minimal value, which will be taken into account by identification algorithm. After starting the identification process, the current best estimation is visualized periodically in a graph together with experimental data. From this perspective the identification algorithm behaves as a curve fitting process.

The whole system can be deployed in single computer, however it was designed to be deployed in several different computing elements. We deployed simulation modules in a local cluster and in a virtual infrastructure within scientific cloud provided by the Czech grid infrastructure provider CESNET[1], member of the European Grid Infrastructure foundation (EGI[2]). The repository module and identification algorithm controls the connected simulation modules.

III. RESULTS

For testing purposes we selected one known parameter from the HumMod model[5] and identify it again. We also tested to identify 4 parameters of model of hemoglobin saturation curve in variable condition of acidity and concentration of carbon dioxid based on the model of Rossi-Bernardi1967[15] implemented in Modelica language. We set the genetic algorithm to finish after 200 thousands single simulations for both models giving the best result found during the computation. The simulation was distributed into local cluster (up to 4 parallel simulation processes, CPU Intel XEON 2.7GHz)) and in scientific cloud (up to 100 parallel simulation processes, CPU Intel E5-2620 2GHz).

Experimental data for HumMod model were generated from single simulation of the model with specified parameter. In the case of the Rossi-Bernardi1967 model, we took the experimental data from the publication[15]. The values of parameters identified during the computation were comparable with the known values. However, we focus on the time of computation and possible speedup when the simulation was distributed into more parallel processes. The measured computation time and speedup is in the table I.

The single simulation of HumMod model takes about 30s of computation time. And we estimate the whole process of computation to 71 days. We didn't wait more than 2 months for this results, rather we estimate this after couple of hours from the number of the simulation done. When distributed into the local cluster up to the 4 CPU, we got the speedup about 3.5 times. When the computation was distributed into virtual infrastructure of 10 computers each contributing by 10 CPU with computation (totally 100CPU) we got the speedup about 96x and the estimation of 1 parameter was done in 18 hours.

The single simulation of the Rossi-Bernardini1967 model takes about 15miliseconds. When computation was distributed into 3 parallel nodes (3 CPU) the utilization of the service module was high and adding another computation node (totally 4 CPU) we did not get any other significant speedup. Distributing the computation to the scientific cloud we got even worse results influenced mainly by the network latency and increased communication overhead.

[1]http://www.cesnet.cz
[2]http://www.egi.eu

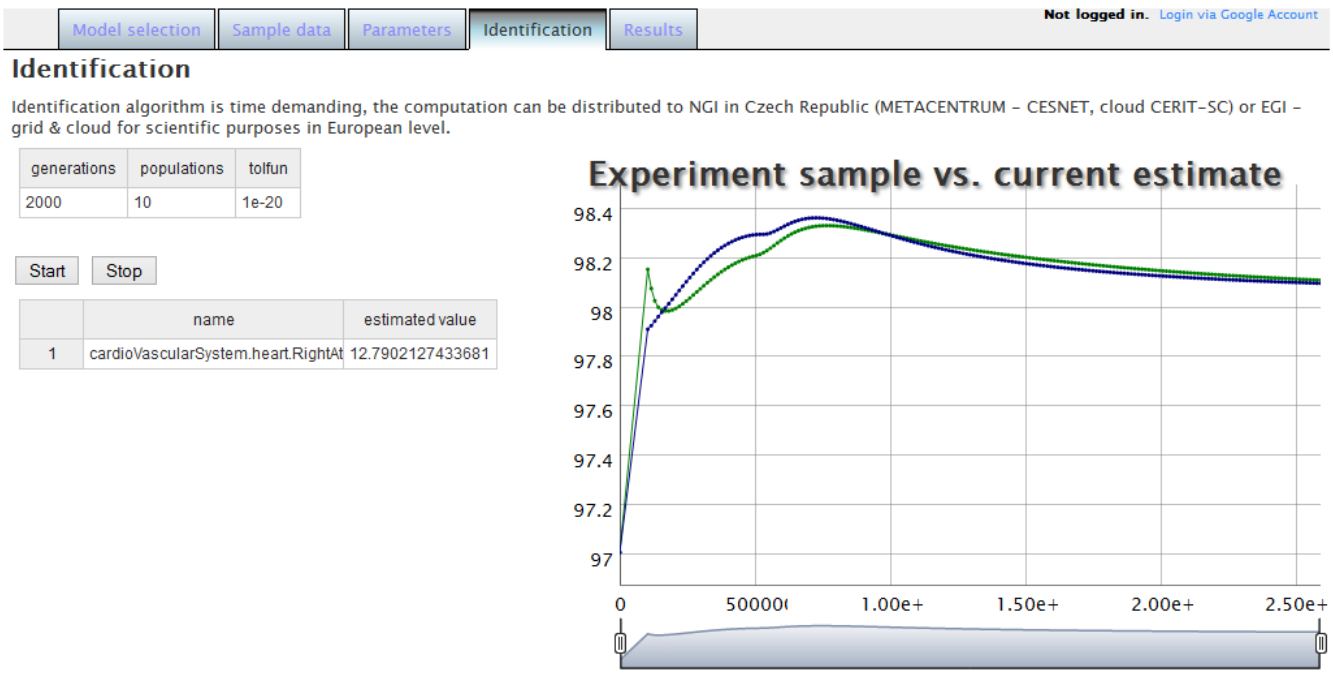

Fig. 2. Screenshot showing running estimation. Current best estimation of parameters values on the left table and curve of estimated model data (blue) vs. experimental data (green).

IV. DISCUSSION

The parameter identification of the complex model spends majority of it's time in parallel simulation and minority in communication and synchronization procedures. This type of tasks can achieve significant speedup if the computation is distributed into remote capacity e.g. within a cloud or grid. On the other hand, the parameter identification of simpler models converges to a highly parallel computation system where time spent in parallel simulation is moreless same as in communication and synchronization procedures.

We estimate that if the single simulation takes more than a second, then the identification task is worth to deploy into cloud computing environment and gain a speedup from it. However if the single simulation takes less for simpler models, than the identification task should stay on local cluster or should be computed in some supercomputer. Distributing them into cloud or grid using this system we do not get any significant speedup. However more exact distinction should be done in further studies.

The parameter estimation within this paper was provided by genetic algorithm, however, there are other identification algorithm (e.g. other evolutionary algorithms) which can gain significant speedup utilizing distributed computing environment.

The system was tested with models implemented in Modelica language, however, significant contribution to the knowledge of human physiology were done by other projects, e.g. VPH[18] or IUPS Physiome[19], which has a so called Physiome model repository[20] and the majority of models are in CellML modeling language or JSIM modeling language. There is an effort to develop translation tool among the modeling technologies to give researchers freedom of choosing the modeling technology e.g.[21]. The further development of the system for parameter estimation can be enhanced to support and simulate the models in the above mentioned modeling technologies.

The web application provides minimal set of functionality for system analysis. Further development of the web application and introduced system needs to do an usability survey and incorporate most useful functionality. For other methods and tasks related to system analysis in physiology use other specialized tools e.g. Design.Calibrate library available in Dymola[22], Optimization toolbox from MATLAB®etc.

V. CONCLUSION

We presented a system to support identification of physiological system in the phase of parameter identification. The loosely coupled part of the system might be deployed into remote distributed computational capacity and significant speedup was shown in the case of the large complex physiological model computed in cloud computing infrastructure.

The described system is accessible via a web application and allows user to focus on input experimental data, names of parameters, visual control of calibration process and hide unnecessary complexity of configuration of the remote computation.

The continued work is oriented to enhance the complex model of human physiology - HumMod and to integrate other modeling technologies.

ACKNOWLEDGMENT

This work was supported by project MPO FR-TI3/869 and by project FR CESNET number 431. The computation was

done on virtual resources of scientific cloud provided by the czech national grid infrastructure, the association CESNET.

REFERENCES

[1] A. C. Guyton, T. G. Coleman, and H. J. Granger, "Circulation: overall regulation." *Annual review of physiology*, vol. 34, pp. 13–46, Jan. 1972. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/4334846

[2] R. L. Hester, A. J. Brown, L. Husband, R. Iliescu, D. Pruett, R. Summers, and T. G. Coleman, "Hummod: a modeling environment for the simulation of integrative human physiology," *Frontiers in physiology*, vol. 2, 2011.

[3] R. L. Hester, R. Iliescu, R. Summers, and T. G. Coleman, "Systems biology and integrative physiological modelling," *The Journal of physiology*, vol. 589, no. 5, pp. 1053–1060, 2011.

[4] J. Kofránek and J. Rusz, "Restoration of Guyton's diagram for regulation of the circulation as a basis for quantitative physiological model development." *Physiological research / Academia Scientiarum Bohemoslovaca*, vol. 59, no. 6, pp. 897–908, Jan. 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20533860

[5] J. Kofranek, M. Matejak, and P. Privitzer, "Hummod-large scale physiological models in modelica," in *Proceedings 8th Modelica Conference, Dresden, Germany*, 2011, pp. 713–724.

[6] J. Kofránek, M. Mateják, P. Privitzer, M. Tribula, T. Kulhánek, J. Šilar, and R. Pecinovský, "Hummod-golem edition: large scale model of integrative physiology for virtual patient simulators," in *Proceedings of World Congress in Computer Science 2013 (WORLDCOMP'13), International Conference on Modeling, Simulation and Visualisation Methods (MSV'13)*, 2013, pp. 182–188.

[7] P. Fritzson and P. Bunus, "Modelica-a general object-oriented language for continuous and discrete-event system modeling and simulation," in *Simulation Symposium, 2002. Proceedings. 35th Annual*. IEEE, 2002, pp. 365–380.

[8] "Modelica and the Modelica Association," https://modelica.org, accessed: 2013-12-01.

[9] M. C. Khoo, *Physiological control systems*. IEEE press, 2000, ch. 7: Identification of Physiological Control Systems.

[10] S. Proßand B. Bachmann, "An Advanced Environment for Hybrid Modeling and Parameter Identification of Biological Systems," *Proceedings 8th Modelica Conference, Dresden, Germany*, pp. 557–571, Jun. 2011.

[11] T. Kulhanek, J. Silar, M. Matejak, P. Privitzer, J. Kofranek, and M. Tribula, "Distributed computation and parameter estimation in identification of physiological systems," *Proceedings of VPH conference*, 2010.

[12] S. Maffioletti, R. Murri, B. Jonen, and S. Scheuring, "Computational workflows with gc3pie," in *Proceedings of the EGI Community Forum*, 2012.

[13] M. Humphrey and N. Beekwilder, "Calibration of watershed models using cloud computing," *Proceedings of IEEE 8th International Conference on E-Science*, 2012.

[14] R. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, University of California, 2000.

[15] L. Rossi-Bernardi and F. Roughton, "The specific influence of carbon dioxide and carbamate compounds on the buffer power and Bohr effects in human haemoglobin solutions," *The Journal of Physiology*, pp. 1–29, 1967.

[16] T. Blochwitz, M. Otter, M. Arnold, C. Bausch, C. Clauß, H. Elmqvist, A. Junghanns, J. Mauss, M. Monteiro, T. Neidhold, D. Neumerkel, H. Olsson, J. Peetz, S. Wolf, G. I. T. I. Gmbh, and D. L. R. Oberpfaffenhofen, "The Functional Mockup Interface for Tool independent Exchange of Simulation Models The Functional Mock-Up Interface," *Modelica*, pp. 105–114, 2011.

[17] "Servicestack," https://github.com/ServiceStack/ServiceStack, accessed: 2014-01-06.

[18] P. Kohl and D. Noble, "Systems biology and the virtual physiological human," *Molecular Systems Biology*, vol. 5, no. 1, 2009.

[19] P. Hunter, P. Robbins, and D. Noble, "The IUPS human physiome project," *European journal of physiology*, vol. 445, no. 1, pp. 1–9, 2002.

[20] T. Yu, C. M. Lloyd, D. P. Nickerson, M. T. Cooling, A. K. Miller, A. Garny, J. R. Terkildsen, J. Lawson, R. D. Britten, P. J. Hunter, and P. M. F. Nielsen, "The Physiome Model Repository 2." *Bioinformatics (Oxford, England)*, vol. 27, no. 5, pp. 743–4, Mar. 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21216774

[21] L. P. Smith, E. Butterworth, J. B. Bassingthwaighte, and H. M. Sauro, "SBML and CellML translation in Antimony and JSim." *Bioinformatics (Oxford, England)*, Nov. 2013.

[22] H. Elmqvist, H. Olsson, and S. Mattsson, "Optimization for design and parameter estimation," *Proceedings of the 4th International Modelica Conference, Hamburg*, 2005. [Online]. Available: http://elib.dlr.de/12296/