FORECAST OF NUMERICAL OPTIMIZATION PROGRESS OF BIOCHEMICAL NETWORKS

Ivars Mozga, Andrejs Kostromins, Egils Stalidzans
Latvia University of Agriculture
ivars.m@gmail.com

Abstract. Increasing size of biochemical models of cellular processes influence the time consumption for optimization of control systems in biotechnological plants to increase the productivity. Nonlinear systems of differential equations can be optimized by time consuming stochastic numerical methods that work for most different nonlinear models but can not guarantee finding of the global optimum. In case of a high number of possible parameter combinations early rejection of parameter combinations with low potential of criteria increase becomes important. Statistics of convergence dynamics is collected to predict the optimization potential of optimization parameter combinations and use them for early rejection of parameter combinations with low optimization potential. A prediction tool is developed to predict the distance to the global optimum depending on the number of parameters, size of the model, and the number of parameters in the model. The prediction tool returns distance to the expected optimal solution as a function of Central Processor Unit (CPU) time. The forecast tool can be used for models of different size and using different numerical optimization methods.

Keywords: biochemical network, numerical optimization, glycolysis, ethanol.

Introduction

Global optimization of nonlinear biochemical networks is a common problem in biotechnological applications of metabolic engineering to evaluate the potential of engineered strains of microorganisms before investment in their development in vivo [1]. *In silico* optimization phase reduce the time and resources needed to reach the best possible result in biotechnological production. Even in case of small models a systematic *in silico* scanning [2] of possible solution space may be a huge effort because of combinatorial explosion.

Generally two groups of methods: deterministic and stochastic [3] can be used for solution of this problem. Still the stochastic methods are more universal and less dependent on the type of problem [4; 5]. Therefore, we concentrate on analysis of stochastic methods which have also drawbacks: they can not guarantee global optimality and take unknown time to reach the best value.

Thus, universality of stochastic algorithms bring also practical disadvantages in their practical applications during optimization experiments specially when many optimization runs have to be performed. The main questions are: 1) when one can conclude that the best solution is found hoping that the global optimum is reached and 2) can one assume that a long optimization run will guarantee a global optimum close solution. These questions are not well studied so far.

It is proposed to apply statistical analysis of several optimization runs for estimation of the convergence speed of the promising combinations of parameters and make early rejection of slowly converging parameters. Possibilities to generalize statistics about the model to different optimization methods and number of optimized parameters are discussed.

Optimization runs are made using software COPASI [6] build 30 and two of built-in optimization methods: particle swarm and evolutionary programming. The model of yeast glycolysis [7] with 24 reactions and 25 metabolites is used as a test model. One, five, ten and fifteen enzyme concentrations were optimized and in each case five identical optimization runs were performed and compared.

The particle swarm method demonstrated good convergence to the best known value independent on the number of modified reactions while evolutionary programming reached the best value of criteria in case of one and fifteen reactions and did not reach it in case of five and ten reactions. The convergence speed was better in case of particle swarm compared to evolutionary programming in case of five, ten and fifteen optimized reactions while evolutionary programming was faster in case of one optimized reaction.

Thus, it was concluded that convergence properties of stochastic global optimization methods are hardly predictable even within the same model depending on the number of optimized reactions. It is necessary to collect convergence statistics for different number of optimized reactions to use the knowledge about convergence speed in similar optimization runs.

A software tool is developed for automated creation of convergence statistics and forecast of time that is needed to reach the best criteria value that might be the global optimum or to be confident that particular closeness to the best value is reached.

The aim of the paper is to find out the differences in convergence speed dynamics in case of different number of optimization parameters within one optimization method and in comparison with another. A forecast tool is developed to choose the method and forecast the convergence dynamics based on the experience of previous runs.

Materials and methods

Yeast glycolysis model [7] is used as a test model for optimization. This model is described with 2 compartments, 24 reactions and 25 metabolites. COPASI (build 30) [6] is used as optimization tool. Two optimization methods are applied: 1) evolutionary programming [8] with following default method parameters: Number of Generations: 30000 (increased to have longer optimization runs); Population Size: 20; Random Number Generator: 1 (uniform distribution); Seed: 0 and particle swarm [9] with the following default method parameters: Iteration Limit: 2000; Swarm Size: 50; Std. Deviation: 1e-06; Random Number Generator: 1; Seed: 0. The optimization parameters were allowed to change within a range from -99 % up to 1000 % from their initial values. "Steady state" subtask of optimization was chosen. Optimization criteria in all optimizations were calculated with the same formula (1).

$$K = \frac{Ethanol\ flow}{Glucose\,uptake} + 5 \cdot Ethanol\ flow, \quad (1)$$

where 5 is expert – determined constant that indicates the increased importance of ethanol flow as a criteria compared to the other part of criteria.

The sequence of modified reactions was chosen in decreasing order of the module of flux control coefficients of ethanol flow (Table 1) using Metabolic Control Analysis (MCA) [10] task of COPASI for the steady state found for the initial values of the glycolysis model as it is downloaded from the Biomodels database [11].

Table 1

Sequence of modified reactions

Sequence number accordingly MCA	Use in optimization experiment sets of 1, 5, 10 and 15 reactions	Reaction name	MCA flux control coefficient for Ethanol flow	Module of MCA flux control coefficient for Ethanol flow
1	1,5,10,15	Hexokinase	7.92e-01	7.92e-01
2	5,10,15	Alcohol dehydrogenase	1.84e-01	1.84e-01
3	5,10,15	ATP consumption	5.21e-02	5.21e-02
4	5,10,15	Glycerol synthesis	-3.13e-02	3.13e-02
5	5,10,15	Phosphofructokinase	2.81e-02	2.81e-02
6	10,15	Glyceraldehyde 3-phosphate dehydrogenase	2.01e-02	2.01e-02
7	10,15	Storage	-1.89e-02	1.89e-02
8	10,15	Triosephosphate isomerase	9.18e-03	9.18e-03
9	10,15	Pyruvate kinase	5.20e-03	5.20e-03
10	10,15	Glucose uptake	4.37e-03	4.37e-03
11	15	Phosphoglucoisomerase	2.32e-03	2.32e-03
12	15	Phosphoenolpyruvate synthesis	1.72e-03	1.72e-03
13	15	Pyruvate decarboxylase	7.82e-16	7.82e-16
14	15	Aldolase	4.04e-16	4.04e-16
15	15	Adenylate kinase	2.13e-18	2.13e-18

The progress of convergence to the global optimum was recorded as time series of CPU time and criteria values.

Five optimization experiments were performed for each number of reactions for each optimization method on a server running 64-bit Microsoft Windows Server 2008 Standard Service Pack 2 operating system. Server has 4x QuadCore Intel Xeon MP E7330 2400 MHz CPU and 32768 MB of RAM. Several optimization experiments were run in parallel. Single processor per task was used as COPASI does not support optimization with parallel task distribution.

The forecast tool was made using Microsoft Visual Studio 2008 Professional Edition and Microsoft .NET Framework version 3.5 Service Pack 1. Programming language is C#. ZedGraph dynamic-link library [12] was used to draw the result curves on the plot. Microsoft SQL Server 2008 Compact Edition version 3.5 was chosen as optimization data storage.

Results and discussion

Convergence speed experiments

Plots of normalized convergence speed of optimization in percent from global optimum with COPASI using evolutionary programming and particle swarm methods for different number of reactions are demonstrated in Figure 1. 0 % level in all the graphs is the criteria value of the original model (4.99). 100 % level corresponds to the best criteria value found in any run of any optimization method and is considered to be the global optimum. So 100 % correspond to 5.02 optimizing one reaction (Fig. 1, a), 6.38 optimizing 5 reactions (Fig. 1, b), 6.48 optimizing 10 reactions (Fig. 1, c) and 12.73 optimizing 15 reactions (Fig. 1, d).

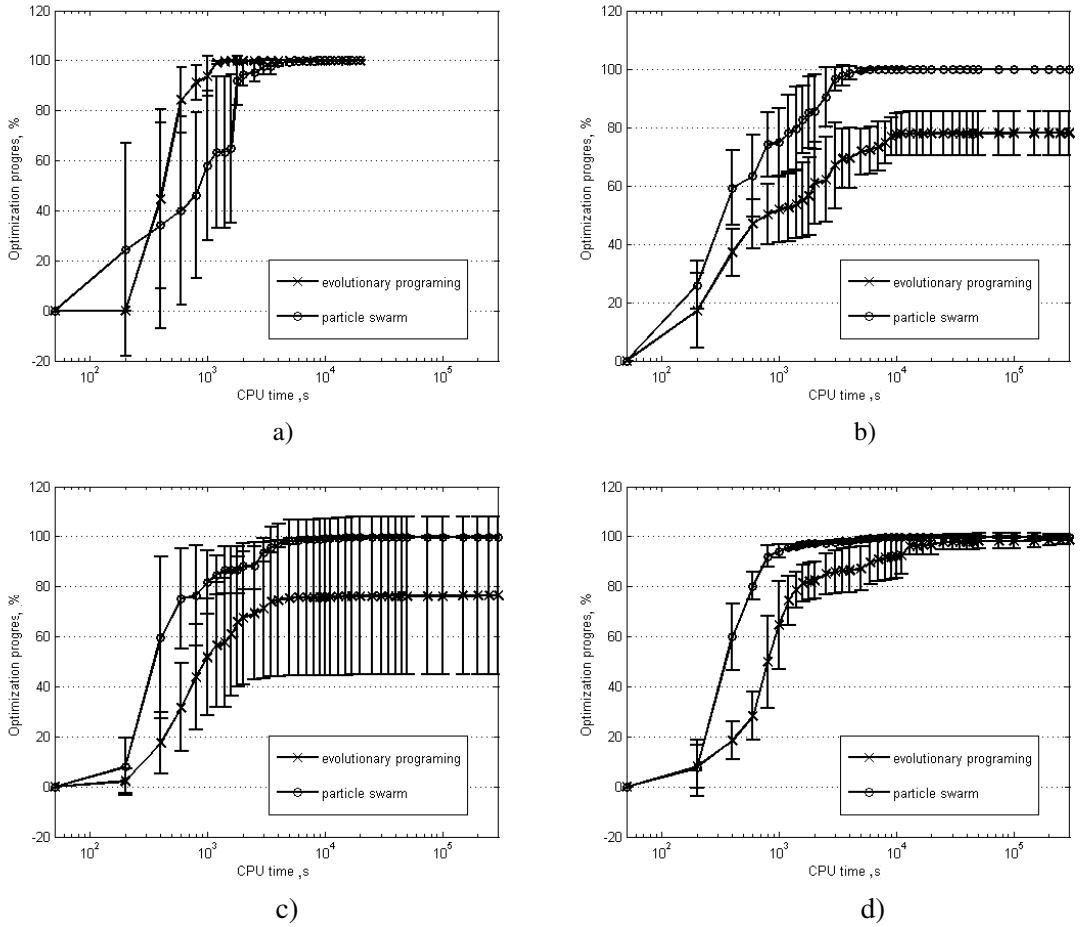

Fig. 1. **Average normalized convergence speed of evolutionary programming and particle swarm optimization methods (error bars represent standard deviation of five experiments):**
a – one reaction optimization; b – five reaction optimization; c – ten reaction optimization; d – fifteen reaction optimization

In case of the given model the optimization performance of both methods has advantages and disadvantages. Evolutionary programming has better performance in case of one optimized reaction both in terms of convergence speed and standard deviation while particle swarm has better performance in case of five, ten and fifteen optimized reactions both in convergence speed and standard deviation. In case of five and ten reactions evolutionary programming has very poor performance. In case of five reaction optimization none of optimization runs reach 100 % level in spite of long optimization runs of 500 000 seconds of CPU time that correspond to 5.8 days. In case of 10 reactions just some optimization runs come close to 100 % level while other runs stay at the level of about 40 %. That may be explained by small preset probability of mutations for evolutionary programming method.

Even comparing just two global optimization methods let us conclude that 1) within optimization of the same model one method is not always better than the other one (evolutionary programming is better in case of one reaction optimization while in all other cases the particle swarm performs better) and 2) there are global optimization methods which converge to the global optimum depending on the number of reactions to be optimized (while there is a good convergence of evolutionary programming in case of one and fifteen reactions it is poor in case of five and ten reactions). Therefore, we suggest collecting knowledge about the convergence progress learning from experience in optimizing of a particular model with a particular optimization tool and method. This approach is reasonable in case if big numbers of similar optimizations have to be performed.

Optimization time/accuracy forecast tool

Software for forecast of relationship between optimization time and distance of best solution from the global optimum has been developed.

The software is universal, so it does not matter which optimization tool is used to get global optimum. The input data consist of two columns representing CPU time for x axis and optimization values for y axis. The input file format is text file where columns are delimited by tab separator. In the forecast tool there is possible to store date in database for further analyses and calculations. The forecast tool supports profile creation according to the user's needs where the number of metabolites, number of reactions, number of parameters, number of variable parameters and optimization method must be specified as configuration parameters for every new profile. A function for method adding also is present in the developed forecast software. All analyses functions are based on previously gathered data.

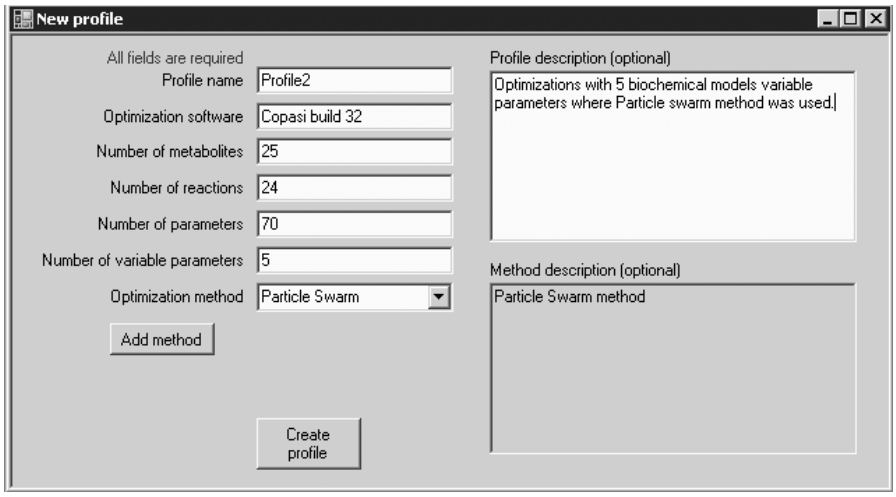

Fig. 2. **New profile creation in forecast tool**

Firstly, the output result tool can calculate and put on the single plot all curves and corresponding average curve with error bars with standard deviation for particular profile.

Secondly, software is able to analyze two methods in parallel to see differences between them. The number of metabolites, number of reactions, number of parameters, number of variable parameters, two methods names and corresponding x and y axes labels must be specified as input

parameters for this function. As a result the function returns the best value and a list of methods which have ever reached the best calculated value for such function input parameters. Moreover, the function also returns the best values for the methods analyzed and the chart is created for representation of two corresponding average curves with standard deviation error bars. It helps the user to choose appropriate methods based on the input parameters. Curve normalization by default is based on automatically calculated the best value for two methods, but it is possible to define another value manually.

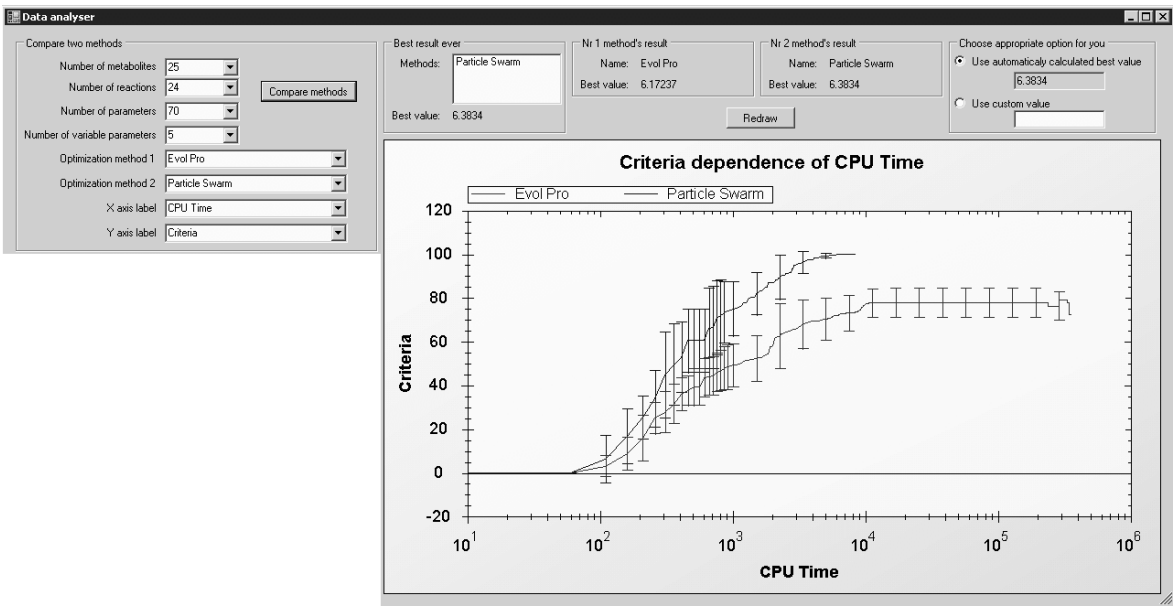

Fig. 4. **Comparison of two methods and finding best global optimum value based on chosen methods**

Thirdly, software offers the analyses of particular method that clearly shows behavior of optimization curves in case of different number of the model variable parameters. The user can freely choose how many curves to show at ones depending on his interests. In this step the global optimum value is also selectable from options. In Figure 5 an example with 5 and 15 variable parameter curves, where the global optimum value is 13 is represented.

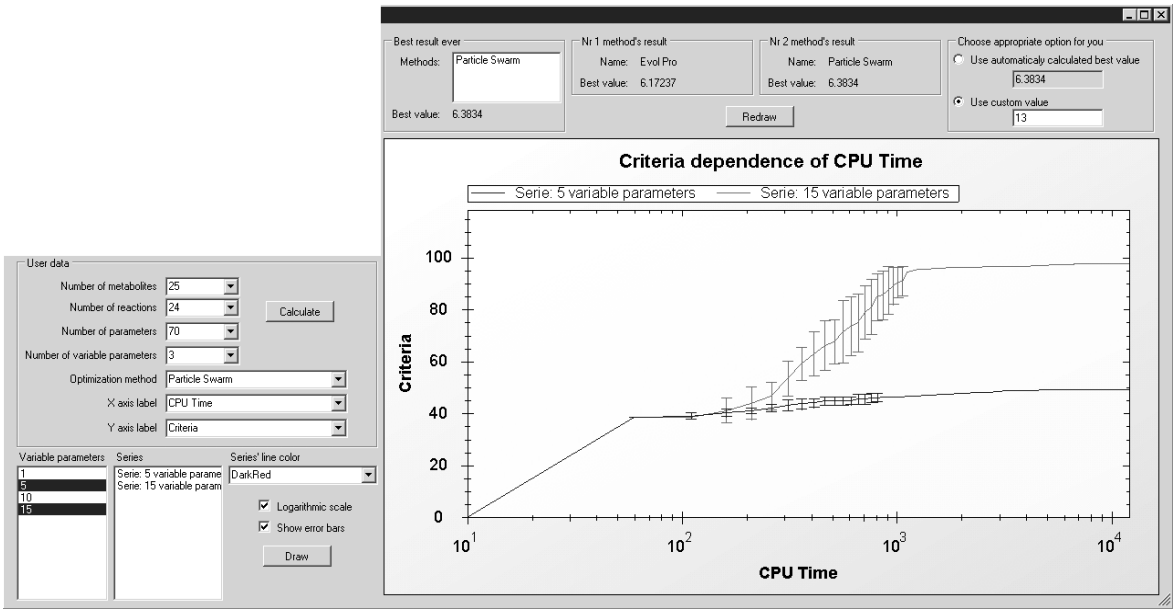

Fig. 5. **Comparison of variable parameter curves based on chosen input parameters and manually specified global optimum value**

Conclusions

Comparing the two global optimization methods for optimization of yeast glycolysis model it can be concluded that 1) within optimization of the same model one method is not always better than the other one (evolutionary programming is better in case of one reaction optimization while in all other cases the particle swarm performs better) and 2) there are global optimization methods which converge to the global optimum strongly depending on the number of reactions to be optimized (while there is a good convergence of evolutionary programming in case of one and fifteen reactions it is poor in case of five and ten reactions).

The developed optimization time/accuracy forecast tool can be used 1) to choose the best method of optimization on the same model and tool tested methods depending on the number of reactions to be optimized and 2) to predict the necessary time to reach the needed level of closeness of the best solution to the global optimum.

Acknowledgments

This work is funded by a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv

References

1. Mendes P., Kell D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics (Oxford, England), 14(10), 1998, pp. 869-883. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9927716.
2. Schulz M., Bakker B.M., Klipp E. Tide: a software for the systematic scanning of drug targets in kinetic network models. BMC bioinformatics, 10, 344. doi: 10.1186/1471-2105-10-344, 2009.
3. Ali M.M., Storey C., Torn A. Application of Stochastic Global Optimization Algorithms to Practical Problems. Journal of Optimization Theory and Applications, 95(3), 1997, pp. 545-563.
4. Balsa-Canto E., Alonso A.A., Banga J.R. Dynamic optimization of bioprocesses : deterministic and stochastic strategies. In C. Skjoldebremd (Eds.), ACoFop IV (Automatic Control of Food and Biological Processes) (Vol. 2, pp. 1-6). Goteborg, Sweden 1998.
5. Moles C.G., Mendes P., Banga J. R. Parameter estimation in biochemical pathways: a comparison of global optimization methods. (Skjoldebremd C., Trystrom G., Eds.) Genome Research, 13(11), 2003, pp. 2467-2474. Goetheborg, Sweden. doi: 10.1101/gr.1262503.
6. Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu L., Mendes P., Kummer U. COPASI — a Complex Pathway Simulator. Bioinformatics 22, 2006, pp. 3067-3074.
7. Hynne F., Danø S., Sørensen P.G. Full-scale model of glycolysis in Saccharomyces cerevisiae. Biophys Chem, 94(1-2):1, 2001, pp. 21-63.
8. Bäck T., Schwefel H.P. An overview of evolutionary algorithms for parameter optimization. Evolutionary Computation. 1. 1997, pp. 1 - 23.
9. Kennedy J. and Eberhart R. Particle Swarm Optimization. Proceedings of the Fourth IEEE International Conference on Neural Networks, Perth, Australia. 1995, pp. 1942-1948.
10. Fell D.A. Metabolic Control Analysis. In: Alberghina L., Westerhoff H.V. (eds.) Systems Biology, Definitions and Perspectives, Springer-Verlag Berlin Heidelberg, Germany, 2005, pp. 69-80.
11. Li C., Donizelli M., Rodriguez N., Dharuri H., Endler L., Chelliah V., Li L., He E., Henry A., Stefan M.I., Snoep J.L., Hucka M., Novère N.L. and Laibe C. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Systems Biology 2010, 4:92
12. Champion J. A flexible charting library for .NET. [online] [20.03.2011.]. Available at: http://www.codeproject.com/KB/graphics/zedgraph.aspx