

Determination of best set of adjustable parameters with full search and limited search methods

Ivars Mozga

SIA TIBIT, Dobeles iela 10-9, LV-3001, Jelgava, Latvia
ivars.m@gmail.com

Received: 26 May 2013; accepted: 30 May 2013; published online: 31 May 2013.
This paper has no supplementary material.

Abstract: *In case of optimization of biochemical networks the best combination of adjustable parameters has to be found keeping low the costs of modification of biochemical network and reducing the risk of unpredicted side effects of processes outside the scope of the model. Dynamic models of biochemical networks usually are in form of system of differential equations and therefore can not be optimized analytically. Usually they are optimized using computationally demanding global stochastic optimization methods which may have hardly predictable duration of optimization. Optimization of all the possible combinations is necessary to find the best set of adjustable parameters per number of parameters in combination. Due to the combinatorial explosion the full search often is not feasible approach if computational and time resources are limited.*

In this study the performance of full search is compared with forward selection and three metabolic control based methods which need significantly less combinations to be examined. It is found that forward selection is a good compromise in case of larger number of adjustable parameters where the number of possible combinations exceeds the available resources needed for full search in the adjustable parameter space. It is proposed to combine full search for small number of adjustable parameters with forward selection at larger number of adjustable parameters per combination.

Keywords: ranking, adjustable parameters, optimization, dynamic model, biochemical networks.

1. Introduction

The development of models of biochemical processes becomes more common task in the developing fields of systems biology and synthetic biology which in turn help in the development of biotechnological, medical and environmental solutions. In case of dynamic models parameter estimation (Mendes et al., 2009; Moles et al., 2003; Rodriguez-Fernandez et al., 2006) is often used to find the correct values of model parameters to repeat the process of interest. Design task (Mendes and Kell, 1998) is performed to find the most effective manipulation with adjustable parameters of the process of interest. Global stochastic optimization methods are often used to solve both kind of tasks (Banga, 2008; Rodriguez-Fernandez et al., 2006). Some disadvantages of global stochastic optimization methods are 1) the hardly predictable duration of optimization (Mozga and Stalidzans, 2011a; Nikolaev, 2010) and 2) possible stagnation in local optima (Mozga and Stalidzans, 2011b; Sulins and Mednis, 2012).

Often both in parameter estimation and design tasks it is necessary to find the smallest set of adjustable parameters to satisfy the task setting: requested level of similarity between experimental measurements and the behaviour of model in case of parameter estimation task or the necessary level of improvement of objective function in case of design task. The problem appears because of the combinatorial nature of the task when the best combination of limited number of parameters has to be found (Pharkya and Maranas, 2006; Yousofshahi et al., 2013). One of approaches is to use the optimization potential to find out when further improvement becomes impossible (Mozga and Stalidzans, 2011c; Mozga,

2012). Application of parallel optimization runs is proposed to analyze (Kostromins et al., 2012) and automatically detect the moment of termination of optimization runs (Sulins and Stalidzans, 2012). Still those attempts remain computationally consuming and request analysis of large amount of combinations.

There are several approaches based on some metabolic network specific features. A method that uses all the enzymes on the way between substrate and product is proposed by Kacser and Acerenza (Kacser and Acerenza, 1993). The advantage of this approach is its simplicity. The main disadvantage is the fact that good increase of objective function can be reached also using smaller set of enzymes (Nikolaev, 2010; Rodriguez-Prados et al., 2009). Metabolic Control Analysis (MCA) approach (Fell, 1992) based solutions are proposed (Acerenza and Ortega, 2007; Hatzimanikatis, 1999; Magnus et al., 2009; Rodriguez-Prados et al., 2009; Stephanopoulos and Simpson, 1997). Still most of them need transformations of models that is hard to automate.

In this study several simple approaches including MCA based ones are tested on their efficiency and compared with full combinatorial search. The tested approaches examine small number of combinations which is close to the number of adjustable parameters. The execution of all tested methods can be performed without transformations of models and can be automated. Tests are performed using model of yeast glycolysis (Hynne et al., 2001).

2. Materials and methods

2.1. Optimization task and software

Yeast glycolysis model (Hynne et al., 2001) downloaded from Biomodels data base (Le Novère et al., 2006) is used as a

test model for optimization. The model contains 2 compartments, 24 reactions and 25 metabolites. Objective function in all optimization runs was

$$K = \frac{\text{Ethanol flow}}{\text{Glucose uptake}} + 5 * \text{Ethanol flow}$$

Concentrations of enzymes catalyzing 15 reactions were chosen as total set of adjustable parameters. COPASI (Hoops et al., 2006), build 30, is used as optimization tool. Particle swarm optimization method is applied with following method parameters: Iteration Limit: 2000; Swarm Size: 50; Std. Deviation: 1e-06; Random Number Generator: 1; Seed: 0. The values of adjustable parameters were allowed to change within a wide range from -99% up to 1000% from their initial values. "Steady state" subtask of optimization was chosen to reject solutions without steady state. ConvAn software (Kostromins et al., 2012) is used for analysis of optimization dynamics.

2.2. Ranking methods

Computationally demanding full combinatorial search of the possible space of adjustable parameter combinations is compared to several alternatives: forward selection method and three MCA based methods. The proposed MCA methods are applicable without any alterations or analysis of the biochemical network like other MCA based methods (Acerenza and Ortega, 2007; Hatzimanikatis, 1999; Magnus et al., 2009; Stephanopoulos and Simpson, 1997) bringing the advantage of possible automation of the model analysis as a single task. Flux control coefficients (FCC) or concentration control coefficients (CCC) are used depending on the expression of objective function. Full combinatorial search in this case serves as a benchmark for evaluation of other methods.

2.2.1. 1-st method

The 1st method starts with no adjustable parameters in the model and adds the next adjustable parameter with the largest module of flux control coefficient (FCC) of initial model (before optimizations) that corresponds to the objective function. That is repeated until all the adjustable parameters are involved. In each step the values of adjustable parameters reached in the previous step remain fixed. The number of tested combinations in this case corresponds to the number of adjustable parameters. In case of proposed model that is 15.

2.2.2. 2-nd method

The 2nd method is similar to the 1st method. The only difference is that the FCC coefficients of a steady state after optimization of all adjustable parameters in one set are taken into account calculating the largest module of FCC that corresponds to the objective function. This method takes into account the best possible improvement of the model instead of the initial state of the model which is not changed under influence of the objective function. The number of tested combinations in this case corresponds to the number of adjustable parameters. In case of proposed model that is 15.

2.2.3. 3-rd method

The 3rd method is reverse compared to the 2nd one: the starting point is a model which is optimized with all the adjustable parameters in one set. FCC of the optimized model is used. The main difference is that the initial state is when all the adjustable parameters are included and in each step the adjustable parameter with the smallest module of FCC that

corresponds to the objective function is removed. That is similar to the backward elimination approach. The process is completed when all the parameters are removed from the set of adjustable parameters. The number of tested combinations in this case also equals to the number of adjustable parameters (15).

2.2.4. Forward selection

Forward selection method starts with no adjustable parameters in the model and tests the addition of each adjustable parameter using objective function as criterion adding the adjustable parameter that improves the objective function the most, and repeating this process until all adjustable parameters are involved. The number of optimized combinations for n adjustable parameters is equal to their sum:

$$\sum_{i=1}^n x_i$$

That is significantly bigger number than the one for 1st, 2nd and 3rd methods. In the experiment only up to eight adjustable parameters per combination were examined because the maximal possible value of objective function (equals to the one of all 15 parameters) was already reached and further addition of adjustable parameters could not increase it.

2.2.5. Full search

Full search is the most demanding approach in terms of computational costs because of the large number of combinations. The number of combinations of m parameters out of n parameters can be calculated by formula:

$$C_n^m = \frac{n!}{m!(n-m)!}$$

The number of all possible combinations with up to n adjustable parameters out of n can be calculated as follows:

$$L_n = \sum_{m=1}^n C_n^m = \sum_{m=1}^n \frac{n!}{m!(n-m)!}$$

Therefore the number of combinations of parameters is increasing very fast and reaches 32767 possible combinations for up to 15 parameters out of the set consisting of 15 parameters.

Thus the screening of all the combinations with optimization runs is very demanding in terms of time and computational resources, complicated or almost impossible due to the huge number of combinations and unpredictable optimization time to reach the global optimum close objective function value (Mozga and Stalidzans, 2011a). The main advantage is that checking all the possible combinations it is guaranteed that the best one is found.

Due to the high number of combinations this method is performed for only up to three adjustable parameters. That means 15 combinations of one parameter, 105 combinations of two parameters and 455 combinations of three parameters (totally 475 combinations) are examined.

3. Results and discussion

The best values of objective function per number of adjustable parameters in combinations (Fig.1) demonstrate big advantage of full search method over the other ones at small number of adjustable parameters. The main drawback is the very high number of searched combinations (475) compared to 3, 3, 114 and 6 combinations for 1st, 2nd, 3rd and forward selection methods.

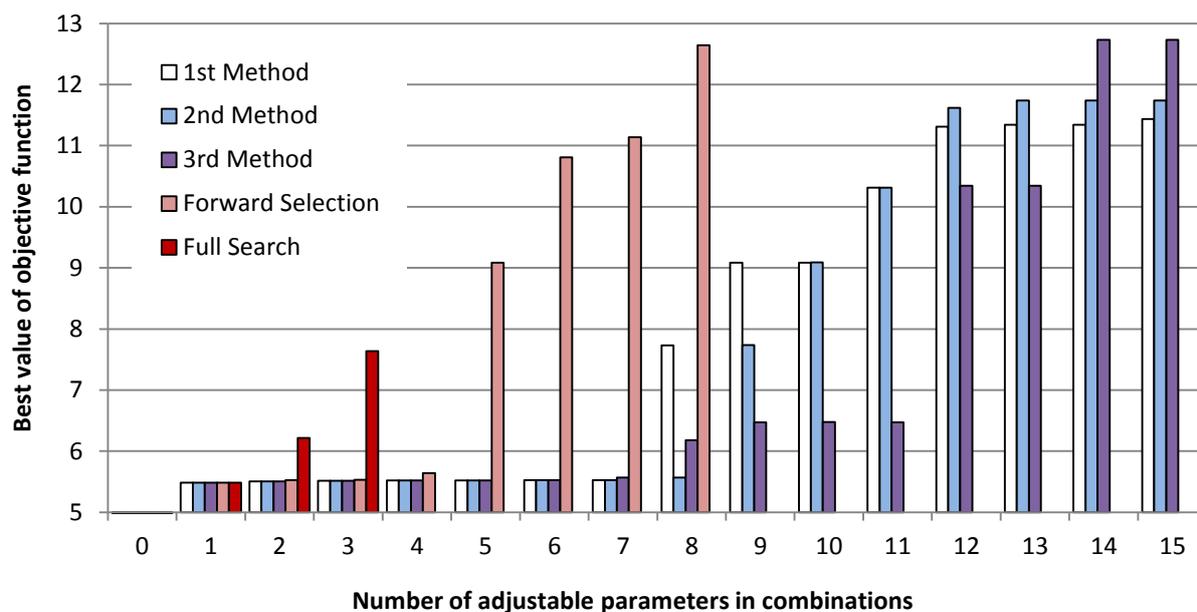


Fig. 1. The best values of objective function per number of adjustable parameters in combinations for compared methods. Forward selection method is applied for up to eight parameters. Full search is applied for up to three parameters.

Forward selection method needs 120 combinations for 15 adjustable parameters and show convincing advantage over all the MCA based methods which request almost 10 times less (15) combinations to cover all the 15 parameters. At the same time the maximal possible increase of objective function is reached already including eight parameters and correspondingly using 92 instead of 120 combinations. Thus the experiment shows clear correlation between the performance of method and the number of examined combinations.

Comparing the MCA based methods the 1st and 2nd method show very similar behavior except for 8 and 9 parameters. The 3rd method has slightly better performance number of adjustable parameters close to the maximum.

Full search for small number of adjustable parameters could be combined with forward selection taking into account the good performance of full search at small number of adjustable parameters and forward selection as the second best method.

4. Conclusion

The tested MCA based or forward selection based methods can not compete with the full search method. Therefore full search method should be used if the computational resources and time are available and high confidence about closeness of the best result to the global optima is needed.

Forward selection is a good compromise in case of larger number of adjustable parameters where the number of possible combinations exceeds the available resources needed for full search in the parameter space.

All MCA based methods demonstrate poor performance and might be outperformed even by selection of random combinations (not tested in this study).

The performed experiments can not be generalized for other optimization tasks and models but it clearly shows that poor performance of tested methods with limited search can have very poor performance. Generally the experiment shows clear correlation between the performance of method and the number of examined combinations.

Full search for small number of adjustable parameters could be combined with forward selection taking into account the good performance of full search at small number of adjustable parameters and forward selection as the second best method.

References

- Acerenza, L. and Ortega, F. (2007), "Modular metabolic control analysis of large responses," *The FEBS journal*, Vol. 274 No. 1, pp. 188–201. <http://dx.doi.org/10.1111/j.1742-4658.2006.05575.x>
- Banga, J.R. (2008), "Optimization in computational systems biology," *BMC systems biology*, Vol. 2, p. 47. <http://dx.doi.org/10.1186/1752-0509-2-47>
- Fell, D.A. (1992), "Metabolic control analysis: a survey of its theoretical and experimental development," *Biochemical Journal*, Portland Press Ltd, Vol. 286 No. Pt 2, pp. 313–330. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1132899/>
- Hatzimanikatis, V. (1999), "Nonlinear metabolic control analysis," *Metabolic engineering*, Vol. 1 No. 1, pp. 75–87. <http://dx.doi.org/10.1006/mben.1998.0108>
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., et al. (2006), "COPASI—a Complex Pathway Simulator," *Bioinformatics (Oxford, England)*, Vol. 22 No. 24, pp. 3067–74. <http://dx.doi.org/10.1093/bioinformatics/btl485>
- Hynne, F., Danø, S. and Sørensen, P.G. (2001), "Full-scale model of glycolysis in *Saccharomyces cerevisiae*," *Biophysical Chemistry*, Vol. 94 No. 1-2, pp. 121–163. [http://dx.doi.org/10.1016/S0301-4622\(01\)00229-0](http://dx.doi.org/10.1016/S0301-4622(01)00229-0)
- Kacser, H. and Acerenza, L. (1993), "A universal method for achieving increases in metabolite production," *European journal of biochemistry / FEBS*, Vol. 216 No. 2, pp. 361–7. <http://dx.doi.org/10.1111/j.1432-1033.1993.tb18153.x>
- Kostromins, A., Mozga, I. and Stalidzans, E. (2012), "ConvAn: a convergence analyzing tool for optimization of biochemical networks," *Biosystems*, Vol. 108 No. 1-3, pp. 73–77. <http://dx.doi.org/10.1016/j.biosystems.2011.12.004>
- Magnus, J.B., Oldiges, M. and Takors, R. (2009), "The identification of enzyme targets for the optimization of a valine producing *Corynebacterium glutamicum* strain using a kinetic model," *Biotechnology progress*, Wiley Online Library, Vol. 25 No. 3, pp. 754–762. doi:10.1021/bp.184
- Mendes, P., Hoops, S., Sahle, S., Gauges, R., Dada, J.O. and Kummer, U. (2009), "Computational Modeling of Biochemical Networks Using COPASI," in Maly, I. V (Ed.), *Systems Biology*, Totowa, NJ, Humana Press, Vol. 500, pp. 17–59. <http://dx.doi.org/10.1007/978-1-59745-525-1>
- Mendes, P. and Kell, D. (1998), "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation," *Bioinformatics*, Vol. 14 No. 10, pp. 869–883. <http://dx.doi.org/10.1093/bioinformatics/14.10.869>

- Moles, C.G., Mendes, P. and Banga, J.R. (2003), "Parameter estimation in biochemical pathways: a comparison of global optimization methods.," (Skjoldbrems, C. and Tryström, G., Eds.) *Genome Research*, Goetheborg, Sweden, Vol. 13 No. 11, pp. 2467–2474. <http://dx.doi.org/10.1101/gr.1262503>
- Mozga, I. (2012), *Steady state optimization procedure of biochemical networks*, Latvia University of Agriculture. Retrieved from http://lufb.llu.lv/dissertation-summary/information-technologies/Promocijas_darba_kopsavilkums_IvarsMozga_LLU_ITF_2012.pdf
- Mozga, I. and Stalidzans, E. (2011a), "Convergence dynamics of biochemical pathway steady state stochastic global optimization," *Proceedings of IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, 21-22 November 2011, Budapest, IEEE, pp. 231–235. <http://dx.doi.org/10.1109/CINTI.2011.6108504>
- Mozga, I. and Stalidzans, E. (2011b), "Convergence Dynamics of Biochemical Models to the Global Optimum," *Proceedings of the 3rd International Conference on E-Health and Bioengineering*, 24-26 November 2011, Iasi, pp. 227–230.
- Mozga, I. and Stalidzans, E. (2011c), "Optimization protocol of biochemical networks for effective collaboration between industry representatives, biologists and modellers," *Proceedings of International Industrial Simulation Conference*, 6.-8 June 2011, Venice, EUROSIS, pp. 91–96.
- Nikolaev, E.V. (2010), "The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems.," *Metabolic engineering*, Elsevier, Vol. 12 No. 1, pp. 26–38. <http://dx.doi.org/10.1016/j.ymben.2009.08.010>
- Le Novère, N., Bornstein, B.J., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., et al. (2006), "BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.," *Nucleic acids research*, Vol. 34 No. Database issue, pp. D689–91. <http://dx.doi.org/10.1093/nar/gkj092>
- Pharkya, P. and Maranas, C.D. (2006), "An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems.," *Metabolic engineering*, Vol. 8 No. 1, pp. 1–13. <http://dx.doi.org/10.1016/j.ymben.2005.08.003>
- Rodríguez-Fernandez, M., Mendes, P. and Banga, J.R. (2006), "A hybrid approach for efficient and robust parameter estimation in biochemical pathways.," *Bio Systems*, Vol. 83 No. 2-3, pp. 248–65. <http://dx.doi.org/10.1016/j.biosystems.2005.06.016>
- Rodríguez-Prados, J.C., De Atauri, P., Maury, J., Ortega, F., Portais, J.C., Chassagnole, C., Acerenza, L., et al. (2009), "In silico strategy to rationally engineer metabolite production: A case study for threonine in *Escherichia coli*.," *Biotechnology and bioengineering*, Vol. 103 No. 3, pp. 609–20. <http://dx.doi.org/10.1002/bit.22271>
- Stephanopoulos, G. and Simpson, T.W. (1997), "Flux amplification in complex metabolic networks," *Chemical Engineering Science*, Vol. 52 No. 15, pp. 2607–2627. [http://dx.doi.org/10.1016/S0009-2509\(97\)00077-8](http://dx.doi.org/10.1016/S0009-2509(97)00077-8)
- Sulins, J. and Mednis, M. (2012), "Automatic termination of parallel optimization runs of stochastic global optimization methods in consensus or stagnation cases," *Biosystems and Information technology*, Vol. 1 No. 1, pp. 1–5. <http://dx.doi.org/10.11592/bit.120501>
- Sulins, J. and Stalidzans, E. (2012), "Corunner: multiple optimization run manager for Copasi software," *Proceedings of 5th International Scientific Conference on Applied Information and Communication Technologies*, 26-27 April 2012, Jelgava, pp. 312–316.
- Yousofshahi, M., Orshansky, M., Lee, K. and Hassoun, S. (2013), "Probabilistic strain optimization under constraint uncertainty," *BMC Systems Biology*, Vol. 7 No. 1, p. 29. <http://dx.doi.org/10.1186/1752-0509-7-29>