

A distributed approach to parameter estimation in systems biology models.



CNR - ITB

www.itb.cnr.it

Ettore Mosca¹, Ivan Merelli¹, Roberta Alfieri^{1,2}, Luciano Milanese¹

¹ Istituto di Tecnologie Biomediche - CNR, via F.lli Cervi 93, 20090 Segrate, Milano, Italy

² CILEA, Consorzio Intrauniversitario per l'Elaborazione Automatica, via R. Sanzio 4 20090, Segrate, Milano, Italy

Introduction

Nowadays a systems biology approach is essential to understand complex biological processes, such as cell cycle regulation [1]. This approach relies on mathematical models used to describe biological systems and makes useful predictions. Due to the lack of experimental measurements, experimental errors and biological variability, the value of many parameters of the models is yet unknown or uncertain [2]. A possible computational solution is the parameter estimation, which can be informally stated as the identification of the parameter values such that the model best fits the experimental data. Considering models based on a system of non-linear Ordinary Differential Equations (ODEs) the problem of parameter estimation is represented by a non-convex non-linear programming (NLP) problem. Since the global optimization (GO) is NP-hard [3], large instances of the problem cannot be solved in a reasonable time, even considering stochastic approaches. We have already developed a automated system to face the problem of parameter estimation of systems biology ODE models based on grid computing [4] and on cluster of processors [5]. Here we present a distributed approach to compute an Evolution Strategy algorithm for parameter estimation of ODEs models.

Methods

Considering a system of n non-linear ODEs $f=(dx/dt, x, p, t)$, where p are real-valued constants (parameters), and experimental data for T temporal points for each variable of the ODE system, the parameter estimation is stated as the minimization of the cost function $J=D(A, Y(p))$ by varying p , where A is the matrix of experimental data, $Y(p)$ is the matrix of the solutions of the ODE system at the same time of A , $D(X, Y) \rightarrow [0, 1]$ is the operator which measures the distance from two real-valued matrices. Here the constraints are the ODE system, the initial conditions $x(t_0)=x_0$ and boundaries on parameter values $p^L < p < p^H$. To solve the GO problem we use an Evolution Strategy (ES), a sub-class of nature-inspired optimization methods belonging to the class of Evolutionary Algorithms (EAs). EAs apply mutation, recombination, and selection operators to a population of individuals representing candidate solutions in order to evolve iteratively better and better solutions. In particular, we chose the Stochastic Ranking Evolution Strategy algorithm (SRES) [6], which has shown good performance when applied to parameter estimation in biochemical pathways [7]. Although this algorithm is very efficient, it can take long time to compute a sufficiently high number of iterations in order to find a satisfying solution. An interesting approach to reduce the execution time relies in running different instances of the algorithm simultaneously, swapping periodically the best results among the processes. Using this method, the convergence to the optimal solution speed up thanks to a wider search on the solution space. To accomplish this task, we have developed an environment to distribute each run of the evolution algorithm on a different computational resource. The core feature of the implementation is a relational database that allows the user to swap the individuals (i.e. solutions of the GO) among the working nodes during the computations.

Results

We have developed an automated system to manage the parameter estimation of ODE models which enables a distributed approach of computing. The system is made of two components: the application for parameter estimation and the infrastructure which manages the distribution over the computational resources relying on a relation database. The system accepts as input models encoded in the SBML standard. Preliminary results indicate that this approach can successfully lead to the parameter estimation of more complex ODE models, overcoming both the computational load and the difficulty of the GO problem for large instances of the parameter estimation problem.

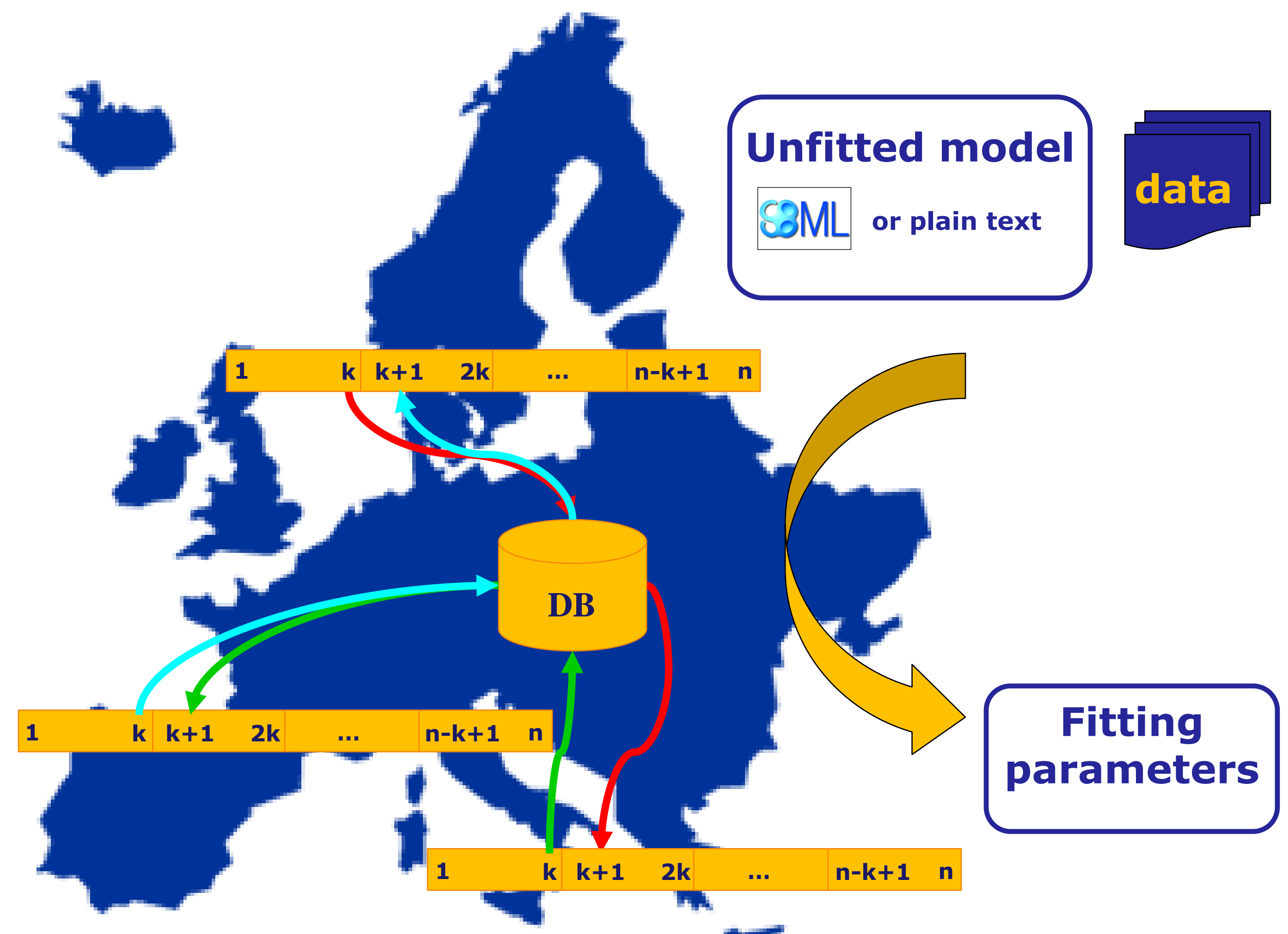


Figure 1. The framework of the parameter estimation which relies on the distributed approach. The system accepts as input a ODE based model (formatted as plain text or encoded in the SBML standard) and the experimental data and outputs the best fitting parameters. During the computation candidate solutions of separated evolution processes are swapped every k iterations thanks to a relational database.

References

1. R. Alfieri, I. Merelli, E. Mosca, L. Milanese (2008) The cell cycle DB: a systems biology approach to cell cycle analysis, *Nucleic Acids Res.* 36(Database issue): D641-D645.
2. W. Liebermeister, E. Klipp (2005) Biochemical networks with uncertain parameters, *Systems Biology, IEE Proceedings*, 152: 97-107
3. S.H Choi, J.W. Ko, V. Manousiouthakis (1999) A stochastic approach to global optimization of chemical processes. *Computers and Chemical Engineering*, 23: 1351-1356.
4. R. Alfieri, E. Mosca, I. Merelli, L. Milanese (2007). Parameter estimation for cell cycle ordinary differential equation (ODE) models using a grid approach. *Stud Health Technol Inform.* 2007;126:93-102.
5. E. Mosca, R. Alfieri, I. Merelli, L. Milanese (2007) Parameters estimation for cell cycle ODE models using high performance computing. *Sysbio-Health Symposium 2007*, Milan, Italy
6. T.P. Runarsson, X. Yao (2000) Stochastic Ranking for Constrained Evolutionary Optimization. *IEEE Transactions on Evolutionary Computation*, 4: 284-294.
7. C.G. Moles, P. Mendes, J.R. Banga (2003) Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Research*, 13: 2467-2474.