IDENTIFICATION OF MAK2 MODEL VIA VARIOUS OPTIMIZATION METHODS

Michal Rost, Vladimír Španihel

Czech Technical University
Department of Software Engineering in Economics
Břehová 7, 115 19, Praha 1
Czech Republic
rostmich@fjfi.cvut.cz, spanivla@fjfi.cvut.cz

Abstract: *This paper deals with application of optimization heuristics to identification of the Mass Action Kinetic model (MAK2) which is used for Polymerase Chain Reaction (PCR) quantification. In order to choose a proper heuristic for the future implementation in a PCR diagnostic software, four heuristic were tested on the same dilution series data. Standard errors of regression models and numbers of evaluations of the objective function were accepted as a quality indicators for the test.*

Keywords: *PCR, mass action kinetic model, model identification, global optimization, differential evolution, fast simulated annealing.*

1 Introduction

Our team is concerned with the development of a control application for thermocycler device [3]. This kind of device is used for performing Polymerase chain reaction (PCR) [2, 8], during which the patient's DNA is being replicated. The replicated DNA is exposed to a beam of laser light, and, finally, the fluorescent light emmited from DNA is collected by photomultiplier and sent to the application for a subsequent analysis.

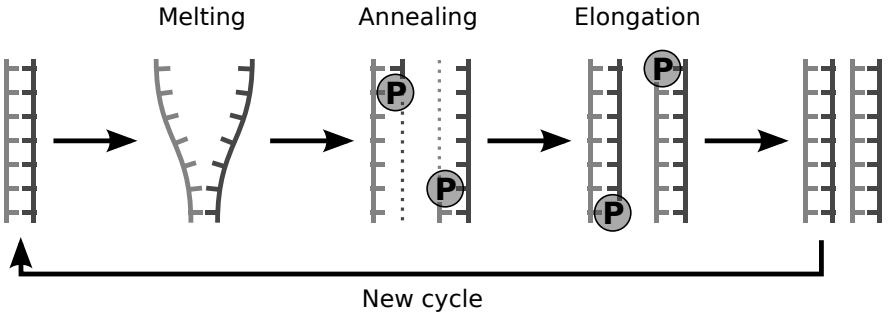

Figure 1: Polymerase chain reaction (PCR).

Standard curve method [6] is commonly used for *quantitative analysis*[1] [1, 6] of fluorescence data. In spite of the fact that this method is very accurate, each time the analysis is performed for a new data set, the callibration curve has to be constructed or re-calibrated using *standard samples* of a known concentration. In 2010 Boggy and Woolf developed a new quantitative method, referred to as MAK2 [1]. It has been expertimentally proved [1] that quantification by MAK2 is as reliable as standard curve method. Moreover, MAK2 does not require any pre-built standard curves and expects only one standard sample in each data set.

1.1 MAK2 method

Mass Action Kinetic model with 2 parameters (MAK2) is based on the chemical kinetics that take place during the annealing and elongation steps of PCR (Fig. 1). As described in [1], the original model has been simplified into the form of two simple reactions (1, 2) where S represents a single DNA strand, while D represents double stranded DNA; parameters k_a and k_b are the reaction rates.

[1] estimates initial concentration of given DNA segment

$$S \xrightarrow{k_a} D \tag{1}$$

$$S + S \xrightarrow{k_b} D \tag{2}$$

$$\begin{aligned} S' = \frac{dS}{dt} &= -k_a S - k_b S^2, \quad S(0) = S_0 \\ D' = \frac{dD}{dt} &= k_a S + \frac{1}{2} k_b S^2, \quad D(0) = 0 \end{aligned} \tag{3}$$

The mathematical representation of equations (1) and (2) contains system of two differential equations (3) from that a final MAK2 recursive formula (4) has been derived [1]. In the (4), variable D_n corresponds to the fluorescence intesity after n PCR cycles, constant k characterizes the dynamics of PCR and F_b represents a background fluorescence as a noise.

$$D_n = D_{n-1} + k \cdot \ln\left(1 + \frac{D_{n-1}}{k}\right) + F_b \tag{4}$$

2 Problem Formulation

Regression analysis has to be carried out for (4) to estimate initial DNA concentration D_0 as well as the other parameters: k and F_b. Boggy and Woolf used the sum of squared residuals as a regression cost function and suggested *Constrained Nelder-Mead* [9] heuristic for its minimization.

In order to choose a proper optimization heuristic for the developed application, it has been decided to reproduce the results from [1], and compare them with the results of the other heuristics. *Fast simulated annealing*, *Differential evolution* and *Competetive differential evolution* algorithms has been chosen for comparison.

2.1 Used heuristics

Constrained Nelder-Mead (CNM) heuristic is a non-linear optimization technique; it is also called *Simplex method*. In our case, the simplex is a set of four vertices in consequence of the three-dimensional parameter space of the cost function. The simplex adapts itself to the local landscape of the cost function on the given domain.

Fast simulated annealing (FSA) [11] is based on the standard Simulated annealing (SA) [7] method; SA employs a global parameter T referred to as temperature which starts at the initial value T_0, and in each step it is cooled due to the cooling factor n_0. In addition, during each step SA attempts to replace the current solution by a new point which is generated at random, following a given probability distribution; then the new point can be accepted with a probability based on the temperature T. FSA method utilizes a general D-dimensional Cauchy distribution as the probability distribution for the generation of the new points.

Differential evolution uses populations while finding the minimum of the cost function. New populations are created by three evolutionary operators: selection, crossover and mutation. There are many modifications of differential evolution. The variant referred to as DER in [12] was selected as a third tested heuristic.

Competitive differential evolution (DEBR18) [13] joins two widely used implementations of differential evolution DER and DEBEST, each with 9 different parameter settings. In each step of the algorithm, the settings compete among themselves with respect to their previous successfulness, so that no parameter tuning is needed.

2.2 Boundaries of search space

Original boundaries of search space and coordinates of starting point that were proposed by [1] are shown in Table 1 where $\vec{F} = (F_1, F_2, \ldots, F_n)$ represents fluorescence data of one sample collected during n cycles of PCR. However, in this case the search interval is limited in the directions D_0 and k and quite wide in the direction of F_b. That is why it has been decided to use decadic logarithm of parameters k, D_0. Consequently, the boundaries and starting point coordinates were changed to values that are shown in Table 2.

Table 1: Original boundaries of search space and coordinates of starting point.

Parameter	Minimal value	Maximal value	Starting value
k	10^{-3}	$\max_i F_i$	1.0
D_0	10^{-12}	10^{-2}	10^{-10}
F_b	$-1.1 \cdot (\lvert F_1 \rvert + 10^{-4})$	$1.1 \cdot (\lvert F_1 \rvert + 10^{-4})$	$\lvert F_1 \rvert + 10^{-4}$

Table 2: Used boundaries of search space and coordinates of starting point.

Parameter	Minimal value	Maximal value	Starting value
$\log k$	-3	$\log \max_i F_i$	0
$\log D_0$	-12	-2	0
F_b	$-1.1 \cdot (\lvert F_1 \rvert + 10^{-4})$	$1.1 \cdot (\lvert F_1 \rvert + 10^{-4})$	$\lvert F_1 \rvert + 10^{-4}$

2.3 Experimental data

All following computations are based on qPCR dilution series data taken from [1]. The dilution series contains 6 samples; for each sample a fluorescence intensity was collected during 40 PCR cycles. If the fluorescence of one sample were plotted against the number of PCR cycle and resulting points were interpolated, they would form a curve referred to as the *amplification curve* [6, 14].

The amplification curves are considered to be sigmoids (S-shape), because of the limited number of PCR components in the samples [14]. In many methods, including the standard curve method and MAK2, only the exponential phase of the amplification curve is being fitted [1, 6, 10, 14]; that is why only the flurescence that correspond with this phase had to be extracted from the dilution series data.

The extraction was achieved by the utilization of second differences and moving averages. First, the second difference series $\vec{F}'$ were calculated from the original fluorescence series $\vec{F}$ of each sample using (5). Second, moving averages were used to identify the trend of $\vec{F}'$. Next, an index i of cycle corresponding to the inflection point of interpolation of $\vec{F}$ was found for each sample. Finally, all sample's fluorescence F_j with $j \leq i$ were extracted from its original data $\vec{F}$.

$$F'_{i-2} = (F_i - F_{i-1}) - (F_{i-1} - F_{i-2}) = F_i - 2 \cdot F_{i-1} + F_{i-2} \quad \forall i \in \{3, \ldots, n\} \tag{5}$$

3 Problem Solution

The previously mentioned heuristics: CNM, FSA, DER and DEBR18 were used for regression analysis of fluorescence extracted from dilution data. Since the FSA algorithm requires an initial temperature (T_0), and cooling rate (n_0) on its input, they were set to $T_0 = 100$, and $n_0 = 0.0001$, respectively. Both differential evolution algorithms use size of population (N) and a tolerance[2] (f_{tol}) parameters; these were set to $N = 30$ and $f_{tol} = 10^{-8}$. Parameters F and CR of DER heuristic were set to $F = 0.8$ and $CR = 0.5$. The results are shown in Table 3 where *ne* stands for the number of evaluations of the objective function and *se* represents the standard error of regression model.

Table 3: Application of heuristics on dilution series data.

	CNM		FSA		DER		DEBR18	
Sample	*ne*	*se* $\times 10^6$	*ne*	*se* $\times 10^6$	*ne*	*se* $\times 10^6$	*ne*	*se* $\times 10^6$
1	222	3.0380	5000	3.2058	5490	3.0380	3540	3.0380
2	214	3.5559	5000	3.7085	5940	3.5559	2550	3.5559
3	300	3.0887	5000	3.5593	4800	3.0887	2820	3.0887
4	239	4.5212	5000	4.7597	5400	4.5212	2790	4.5212
5	202	2.9087	5000	3.9244	5160	2.9087	2790	2.9087
6	194	3.0207	5000	3.6347	4380	3.0207	2610	3.0207

After all six samples had been tested, second sample was chosen at random for further analysis. Next, each heuristic was executed on its data one hundred times. Finally, estimates E of expected values and standard deviations S of objective function's evaluation count (ne) and standard error (se) were calculated and recorded

[2] difference between the best and worst objective function value in the population

in Table 4. As shown in Table 4 Constrained Nelder-Mead method is the fastest and has same smallest standard error as both DER and DEBR18. Moreover, CNM has the smallest standard deviation of error.

Table 4: Estimates of expected values and standard deviations of objective function's evaluation count and standard error, for 100 observations.

Heuristic	CNM	FSA	DER	DEBR18
E_{ne}	214.0000	5000.0000	5.4417×10^{3}	2.8461×10^{3}
S_{ne}	0.0000	0.0000	352.1853	230.2826
E_{se}	3.5559×10^{-6}	4.3112×10^{-6}	3.5559×10^{-6}	3.5559×10^{-6}
S_{se}	5.9591×10^{-21}	7.1586×10^{-7}	5.5048×10^{-12}	7.4963×10^{-12}

The changes of population size from $N = 30$ to $N = 120, 100, 80, 60, 40, 30, 20$ in DEBR18 did not affected the size of standard errors significantly (Table 5); however, in the most cases smaller population size led to smaller number of evaluations of the objective function (Fig 2, Table 5).

Table 5: Complexity and model error of DEBR18 in relation to population size.

	DEBR18			
N	E_{ne}	S_{ne}	$E_{se} \times 10^{10}$	$S_{se} \times 10^{12}$
120	11161.2000	698.7771	35559.0853	2.0977
100	9282.0000	525.3916	35559.0903	1.8335
80	7520.8000	551.9766	35559.0972	2.7803
60	5652.0000	410.6757	35559.1045	3.2344
40	3811.6000	286.6622	35559.1250	4.9053
30	2837.1000	245.7736	35559.1696	9.3156
20	1951.2000	203.6499	35559.2110	12.4392

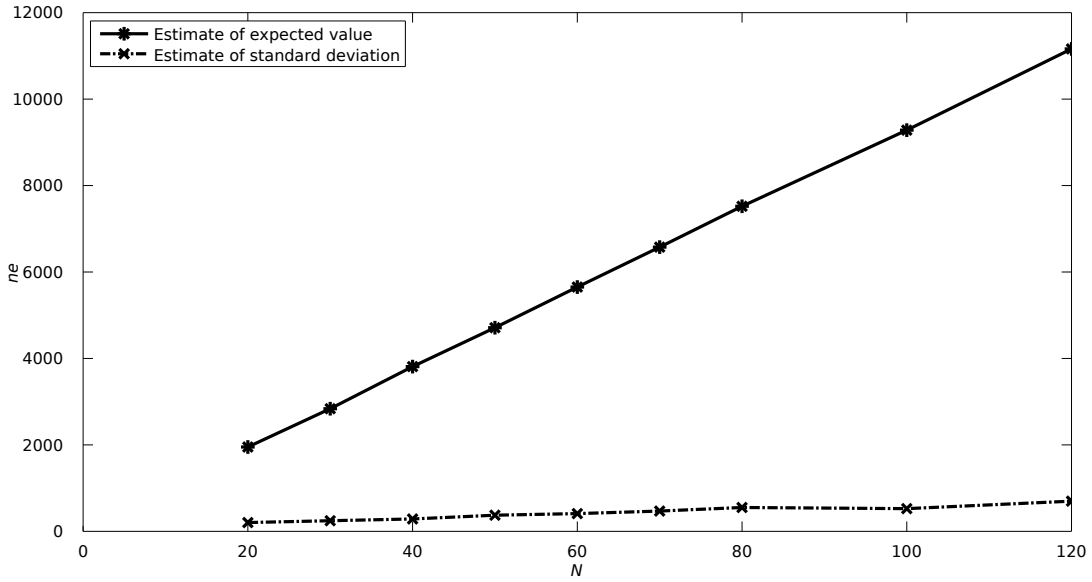

Figure 2: Estimates of expected values and standard deviations of objective function's evaluation count plotted against DEBR18 population size.

4 Conclusion

As shown in tables 3, 4 and 5 the tested data set has been best fitted by CNM, DEBR and DEBR18 methods. While there is no significant difference among standard errors of these three methods, the CNM method has unambiguously the lowest number of evaluations of the objective function. However, CNM method is strongly

dependent on the coordinates of starting point, while the DEBR and DEBR18 do not require any starting point. Moreover, the quantification analysis will be performed expost in developed diagnostic software, so complexity is not so important in this case. We are considering choosing DEBR18 method for the implementation; nevertheless, other heuristics like Harmony Search [4, 5] and Cuckoo Search [15] will be tested before the implementation.

Acknowledgement: The research reported in this paper has been supported by grant: MPO FR-TI1/591; creation of this paper was partially supported by grant: SGS 11/167.

References

[1] Boggy, G. J., Woolf, P. J., A Mechanistic Model of PCR for Accurate Quantification of Quantitative PCR Data, Michigan, *Plos One*, Vol.5(8), 2010, doi: 10.1371/journal.pone.0012355

[2] Erlich, H. A., Gelfand, D. H., Higuchi, R., Horn, G. T., Mullis, K. B., Saiky, R. S., Scharf, S. J., Stoffel, S., Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase, *Science*, Vol.239, 1988, pp. 487-491

[3] Fiala, P., Rost, M., Španihel, V., Virius, M., Development of Modern Application for In-Vitro diagnostics. In *Proceedings of 37th Conference on Information Technology SOFTWARE DEVELOPMENT 2011*, Ostrava, 2011, pp. 19-26

[4] Forbes, F., Qin, A. K., Harmony Search with Differential Mutation based Pitch Adjustment. In *In Proceedings of 13th Annual Conference on Genetic and Evolutionary Computation GECCO 2011*, Dublin, 2011, pp. 545–552

[5] Geem, Z. W., Kim, J. H., Loganathan, G. V., A New Heuristic Optimization Algorithm: Harmony Search, *Simulation*, Vol.76(2), 2001, pp. 60-68

[6] Hodek, J., Kučera, L., Ovesná J., Interferences of PCR Effectivity: Importance for Quantitative Analyses, Praha, *Czech Journal of Food Sciences*, Vol.27, 2009, pp. 42-49

[7] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., Optimization by Simulated Annealing, *Science*, Vol.220, 1983, pp. 671-680

[8] Mullis, K. B., The Unusual Origin of the Polymerase Chain Reaction, *Scientific American*, Vol.262(4), 1990, pp. 56-65

[9] Nelder, J.A., Mead, R., A Simplex Method for Function Minimization, *Computer Journal*, Vol.7(1), 1964, pp. 308–313

[10] Pfaffl, M. W., Quantification Strategies in Real-time PCR. In *A-Z of quantitative PCR*, IUL, 2004, pp. 87-112

[11] Szu, H., Hartley, R., Fast Simulated Annealing, *Physics Letters A*, Vol.122(3), 1987, pp. 157-162

[12] Tvrdík, J., *Evoluční algoritmy*, Ostravská univerzita, 2004

[13] Tvrdík, J., Competitive Differential Evolution. In *Proceedings of 16th International Conference on Soft-Computing MENDEL 2010*, Brno, 2010, pp. 7-12.

[14] Van Pelt-Verkuil, E., Van Belkum, A., Hays, J. P., *Principles and Technical Aspects of PCR Amplification*, Springer, 2008

[15] Yang, X.-S., Deb, S., Cuckoo search via Lévy flights. In *Proceedings of World Congress on Nature & Biologically Inspired Computing NaBIC 2009*, Coimbatore, 2009, pp. 210-214