Josef Martínek; Václav Kučera
Numerical optimization of parameters in systems of differential equations

In: Jan Chleboun and Pavel Kůs and Jan Papež and Miroslav Rozložník and Karel Segeth and Jakub Šístek (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Jablonec nad Nisou, June 19-24, 2022. Institute of Mathematics CAS, Prague, 2023. pp. 123–132.

Persistent URL: http://dml.cz/dmlcz/703194

Programs and Algorithms of Numerical Mathematics 21
J. Chleboun, P. Kůs, J. Papež, M. Rozložník, K. Segeth, J. Šístek (Eds.)
Institute of Mathematics CAS, Prague 2023

NUMERICAL OPTIMIZATION OF PARAMETERS IN SYSTEMS OF DIFFERENTIAL EQUATIONS

Josef Martínek, Václav Kučera

Faculty of Mathematics and Physics, Charles University
Sokolovská 83, Praha 8, 186 75, Czech Republic
martinek@karlin.mff.cuni.cz, kucera@karlin.mff.cuni.cz

Abstract: We present results on the estimation of unknown parameters in systems of ordinary differential equations in order to fit the output of models to real data. The numerical method is based on the nonlinear least squares problem along with the solution of sensitivity equations corresponding to the differential equations. We will present the performance of the method on the problem of fitting the output of basic compartmental epidemic models to data from the Covid-19 epidemic. This allows us to draw several conclusions on the natural limitations of these models and their validity.

Keywords: Ordinary differential equations, parameter estimation, nonlinear least squares, mathematical epidemiology

MSC: 65L05, 92D30, 65K10

1. Introduction

Ordinary differential equations (ODEs) are one of the most common mathematical tools to describe natural phenomena. Extensive literature exists on how to build more or less sophisticated mathematical models leading to ODEs. Typically the resulting equations contain unknown parameters (constants) which must be tailored to the specific application. These can be obtained by measurement, theoretical considerations, etc., but in certain situations it is difficult to come up even with a rough estimate of the real-life parameters of the model. One possibility then is to tune the parameters of the model so that its output agrees best with measured data. There are many approaches to solve such a data-fitting problem, cf. [6]. Here we build on the approach of [2] which uses so-called sensitivity equations to obtain the dependence of the solution of the ODEs on the considered parameters (Section 2). This is then used in a gradient-based Levenberg-Marquardt optimization algorithm which solves a nonlinear least-squares problem of fitting the output to the data (Section 3). We test the data-fitting algorithm on compartmental models from epidemiology (Section 4), specifically we take data from the COVID-19 epidemic in the Czech Republic (Section 5) and discuss the validity of such simple models.

DOI: 10.21136/panm.2022.12

2. Ordinary differential equations and sensitivity equations

We use the notion of a system of ODEs in the following way:

Definition 1. *Let* $n \in \mathbb{N}$, *and* $f_i \colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ *for* $i \in \{1, \dots, n\}$. *By a* system of differential equations *we mean a system of the form*

$$\begin{aligned} y_1' &= f_1(y_1, \dots, y_n, t), \\ y_2' &= f_2(y_1, \dots, y_n, t), \\ &\vdots \\ y_n' &= f_n(y_1, \dots, y_n, t). \end{aligned} \tag{1}$$

We use vector notation $y' = f(t, y(t))$ *for brevity. By an* initial value problem *we mean the system* (1) *along with a point* $(t_0, y^0) \in \mathbb{R} \times \mathbb{R}^n$ *called the* initial condition. *We seek a solution of the system of differential equations such that* $y(t_0) = y^0$.

Throughout this contribution, we consider the case when the system of ODEs (1) contains some known or unknown parameters, in which case the resulting solution also depends on the choice of the parameter. Specifically, instead of y being only a function of t, i.e. $y(t)$, we will have also the dependence on some parameter(s) c: hence we write $y(t, c)$. To simplify the notation, for some fixed value of the parameter c we will sometimes omit the second argument and write $y(t, c) = y(t)$. Similarly, we write $y'(t, c) = y'(t) = \frac{\partial y}{\partial t}(t, c)$ if the right-hand side is defined. This will simplify the notation for ODEs, where t is the relevant variable and c is only a parameter.

Now we follow the paper of Dickinson and Gelinas [2] and the monograph [6] by Schittkowski. Let us consider an initial value problem

$$y'(t, c) = f\big(y(t, c), t, c\big), \quad y(0, c) = y^0, \tag{2}$$

which depends on a real parameter c. For now we assume that the initial condition y^0 does not depend on c. In order to optimize the parameters in our models we need to determine the so called *sensitivity* of the system with respect to c.

Definition 2. *Let* $i \in \{1, \dots, n\}$. *We define the* sensitivity of the i-th variable with respect to the parameter c *by*

$$z_i(t, c) = \frac{\partial y_i}{\partial c}(t, c).$$

The sensitivities defined above can be obtained as a solution of a system of ODEs called the sensitivity equations which we derive now. Let $i \in \{1, \dots, n\}$. We assume that all functions involved are sufficiently smooth. Then we obtain by Definition 2 and the rule for interchanging the order of differentiation

$$\frac{\partial z_i}{\partial t}(t, c) = \frac{\partial}{\partial t}\left(\frac{\partial y_i}{\partial c}(t, c)\right) = \frac{\partial}{\partial c}\left(\frac{\partial y_i}{\partial t}(t, c)\right). \tag{3}$$

By using (2), the chain rule for differentiation and Definition 2, we have

$$\begin{aligned}\frac{\partial z_i}{\partial t}(t,c) &= \frac{\partial}{\partial c}\big[f_i\big(y_1(t,c),\dots,y_n(t,c),t,c\big)\big] \\ &= \frac{\partial f_i}{\partial c}\big(y_1,\dots,y_n,t,c\big) + \sum_{j=1}^{n}\frac{\partial f_i}{\partial y_j}\big(y_1,\dots,y_n,t,c\big)\frac{\partial y_j}{\partial c}(t,c) \\ &= \frac{\partial f_i}{\partial c}\big(y_1,\dots,y_n,t,c\big) + \sum_{j=1}^{n}\frac{\partial f_i}{\partial y_j}\big(y_1,\dots,y_n,t,c\big)\, z_j(t,c).\end{aligned} \tag{4}$$

We have obtained the so-called sensitivity equations. These are a system of n ODEs which can be solved simultaneously with the original system (2). We now determine the initial condition of the sensitivity equations. Since the initial condition of the original system (2) does not depend on the parameter c, we have by Definition 2:

$$z_i(0,c) = \frac{\partial y_i}{\partial c}(0,c) = \frac{\partial y_i^0}{\partial c} = 0. \tag{5}$$

Definition 3. *Let $y'(t,c) = f(y(t,c),t,c)$, $y(0,c) = y^0$ be an initial value problem of the form (2) and suppose that the initial condition does not depend on the parameter $c \in \mathbb{R}$. We define the corresponding* sensitivity equations *by*

$$z_i'(t,c) = \frac{\partial f_i}{\partial c}\big(y_1,\dots,y_n,t,c\big) + \sum_{j=1}^{n}\frac{\partial f_i}{\partial y_j}\big(y_1,\dots,y_n,t,c\big)\, z_j(t,c), \quad z_i(0,c) = 0,$$

for $i \in \{1,\dots,n\}$.

2.1. Multiple parameters and parameter in initial condition

The previous derivation generalizes straightforwardly to the case of multiple parameters (in the equation only), where we use the vector form $c = (c_1,\dots,c_m)^T \in \mathbb{R}^m$. We define the *sensitivity of the i-th variable with respect to the parameter c_j* by

$$z_i^j(t,c) = \frac{\partial y_i}{\partial c_j}(t,c).$$

Proceeding similarly as in the derivation in the previous case (4) we obtain the sensitivity equation for the sensitivity z_i^j in the form

$$(z_i^j)' = \frac{\partial f_i}{\partial c_j} + \sum_{k=1}^{n}\frac{\partial f_i}{\partial y_k} z_k^j,$$

along with the initial conditions $z_i^j(0,c) = 0$.

Until now we have discussed the case when the initial condition does not depend on c. However, a parameter may appear both in the equation and in the initial condition. Consider for example the following initial value problem:

$$y'(t,c) = f\big(y(t,c),t,c\big), \quad y(0,c) = (c, y_2^0,\dots,y_n^0)^T, \tag{6}$$

The sensitivity equations themselves are identical to those in Definition 3. As for the initial condition, for the first variable z_1 we have by Definition 2

$$z_1(0,c) = \frac{\partial y_1}{\partial c}(0,c) = \frac{\partial}{\partial c}c = 1$$

and for $i \in \{2,\dots,n\}$ we get $z_i(0,c) = 0$ as in the previous section.

3. Algorithms for parameter optimization

We now address the problem of optimizing the parameters in ODEs, i.e. finding the set of parameters for which the solution of the ODE has the best agreement with given data obtained e.g. from measurement or observation. There are many possibilities how to approach this problem, see [6]. Our approach is the following: The resulting function obtained as a solution to the considered model fits the measured data in the least squares sense. More precisely, consider the initial value problem (2) which depends on m parameters $c = (c_1,\dots,c_m)^T \in \mathbb{R}^m$. Suppose we have a set of data points $\{(t_j, Y^j) \in \mathbb{R}^{n+1},\ j = 0,\dots,M\}$. We want to find a vector of parameters $c_{\min} \in \mathbb{R}^m$ such that it satisfies the condition

$$c_{min} = \underset{c\in\mathbb{R}^m}{\operatorname{argmin}} \sum_{j=0}^{M} \|y(t_j,c) - Y^j\|^2 = \underset{c\in\mathbb{R}^m}{\operatorname{argmin}} \sum_{j=0}^{M}\sum_{i=1}^{n} \big(r_{ij}(c)\big)^2, \tag{7}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^n$ and the *residuals* are defined by

$$r_{ij}(c) = y_i(t_j,c) - Y_i^j. \tag{8}$$

A minimization problem of the form (7) is called a *nonlinear least squares* problem. In the case when $r_{ij}(c)$ depend linearly on c, the problem reduces to (linear) least squares. Since we are typically unable to find analytic solutions to our ODEs, we cannot write the explicit formulae for $r_{ij}(c)$. However, one can solve the equations numerically, in our case by a fourth order Runge-Kutta method. Moreover, we can also compute the partial derivatives of the residuals w.r.t. the parameters:

$$\frac{\partial r_{ij}}{\partial c_k}(c) = \frac{\partial y_i}{\partial c_k}(t_j,c) = z_i^k(t_j,c), \quad k \in \{1,\dots,m\}.$$

We can therefore evaluate the partial derivatives of r_{ij} by solving the sensitivity equations (also using Runge-Kutta) in parallel with the original ODEs. This allows us to apply a gradient-based optimization algorithm for the numerical solution of problem (7). Specifically, we use the Levenberg-Marquardt method, which produces the a sequence of approximations to c_{min} using the iterative process

$$c^{(l+1)} = c^{(l)} - (J_l^T J_l + \lambda_l I)^{-1} J_l^T r(c^{(l)}), \quad l = 0,\dots,$$

where J_l is the Jacobi matrix of the residuals r_{ij} w.r.t. the parameter vector c at the l-th iteration. The constant λ_l is a 'damping' parameter which interpolates

between Gauss-Newton method ($\lambda_l = 0$) and steepest descent ($\lambda_l \to \infty$). There are various choices of λ_l, we adopted the simple strategy from the original paper of Marquardt [4], which proved sufficient in our case. The more basic method, Gauss-Newton's method, did not converge in several of our test cases or converged very locally probably due to the near-singularity of the Jacobi matrices. The Levenberg-Marquardt method can be viewed as Gauss-Newton using a trust region approach.

4. Compartmental epidemiological models

We will test the performance of the parameter optimization algorithm on systems of ODEs coming from mathematical biology, namely models for the spreading of infections diseases in a population. Mathematical models in epidemiology may be sorted into various categories according to different criteria – discretization of time (models with discrete intervals and continuous time models), allowing for randomness (stochastic and deterministic models), structure of the population etc. Here we take into account exclusively *deterministic*, *continuous time* models where the population is assumed to be a *homogeneous continuum*. Presumably the most widely known representatives of this kind of models are the standard compartmental models. These models are based on the principle of dividing the population into several labeled *compartments* (eg. Infectious, Recovered etc.) under certain simplifying assumptions. The development of the epidemic is then determined by relations describing the flow between compartments, namely the rate of flow between a pair of compartments. The model is formulated mathematically as a system of ODEs.

4.1. SIR model

The *SIR model* is the most basic compartmental model, cf. [5]. The population is divided into three groups, each group a function of time:

- *Susceptible* (S) — those who have not come across the disease and can fall ill if they come into contact with an infectious person, thus becoming infectious.
- *Infectious* (I) — those who spread the disease among the susceptible population. After recovery they move to the compartment R:
- *Recovered* (R) — those who are removed from the compartment I either due to recovery or due to death.

The relations between the compartments are based on four fundamental assumptions:

1. The vital dynamics is neglected and the size of the population is supposed to be constant, we denote it by $N > 0$.
2. The population is assumed to be a homogeneous continuum, i.e. all people have an equal number of contacts, the probability of the transmission of the disease between a susceptible and an infectious person during their contact remains constant and the infectious are equally distributed among the population.

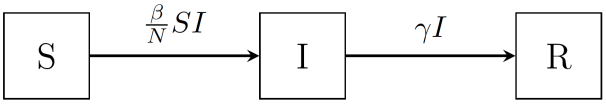

Figure 1: SIR model.

3. The rate of flow between the compartments I and R is directly proportional to the size of the compartment I.

4. The recovered acquire immunity and cannot spread the infection. Those who fall victims to the disease are treated as recovered.

Let r be the number of contacts of a person per unit time and let $p \in (0, 1)$ be the probability of the transmission between an infectious and a susceptible person when they meet. It is desired to find the number of people an infectious person infects per unit time. The fraction of susceptible population within the total population is $\frac{S}{N}$. Therefore, the infectious person meets a total of $r\frac{S}{N}$ susceptible people per unit time. It follows that the number of infected susceptible people per infectious person per unit time is $pr\frac{S}{N}$. It proves convenient to define a new constant $\beta = pr$. Because the total number of infectious people is equal to I, it can be concluded that the total number of people an infectious person infects per unit time is $\beta I\frac{S}{N}$.

We now determine the relation between compartments I and R. As stated in the assumption 2, the rate of flow between the compartments I and R is directly proportional to the size of the compartment I. Denote by γ the coefficient of proportionality. The rate of flow is then equal to γI. The value $\frac{1}{\gamma}$ can be interpreted as the expected time spent in the compartment I, cf. [5].

The resulting model is described mathematically by the system of ODEs

$$S' = -\frac{\beta}{N}SI, \quad I' = \frac{\beta}{N}SI - \gamma I, \quad R' = \gamma I. \tag{9}$$

The model is shown schematically in Figure 1. The system (9) is equipped with the following initial conditions. Let $I_0 > 0$ and $R_0 \geq 0$. We set

$$S(0) = N - R_0 - I_0, \quad I(0) = I_0, \quad R(0) = R_0. \tag{10}$$

There are many generalizations of the SIR model, usually based on the introduction of various other compartments. For example, the *SIQR* model is based on the additional assumption that every infectious subject is quarantined after the infection is detected. In addition to S, I, and R, we define a new compartment called *Quarantined* denoted Q. The infectious move from the compartment I to the compartment Q with a rate of flow directly proportional to the size of I. Analogously, the quarantined leave the compartment Q and move on to the compartment R with

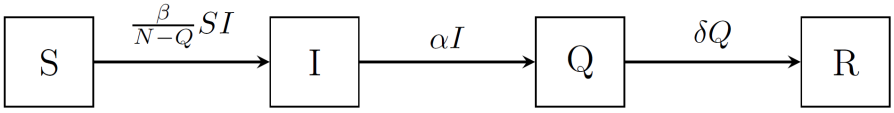

Figure 2: SIQR model.

a rate of flow directly proportional to the size of Q. The coefficients of proportionality are denoted by α and δ, respectively. Finally, we assume that the quarantined are unable to interact with the rest of the population. Thus the flow rate between S and I in the SIR model has to be modified appropriately. The model is shown schematically in Figure 2. The resulting system of ODEs reads

$$S' = -\frac{\beta}{N-Q}SI, \quad I' = \frac{\beta}{N-Q}SI - \alpha I, \quad Q' = \alpha I - \delta Q, \quad R' = \delta Q. \tag{11}$$

Apart from SIR and SIQR, we considered several other variants, such as the SEIR and SEIQR models and a different version of the SIQR model. Here E stands for *Exposed*, this compartment contains infected people who are not infectious yet, effectively adding a latency period to the standard model. We only mention these models in passing, since they gave us results almost identical with the basic SIR model on the considered data and thus present no added value in our case.

5. Numerical results

The approach to parameter optimization described in Sections 2 and 3 was implemented in MATLAB and tested on COVID-19 epidemiological data from the Czech Republic using the models from Section 4. However first we have tested the algorithms on artificially generated data and, more interestingly, on a standard test-case of data from a well studied and documented local influenza epidemic.

5.1. Influenza epidemic in a boarding school

The SIR model is derived under certain assumptions on the population and the disease. This may significantly affect the accuracy of the model in practice. We present here one case, which is as close as possible to satisfying the assumptions, the case of an influenza outbreak in an English boarding school from 1978, cf. [5].

In total, 763 boys were present, one boy had an influenza-like illness from the A/USSR/90/77(H1N1) virus. Over the next two weeks, a total of 512 boys developed similar symptoms spending between three and seven days in the college infirmary. We want to estimate the values of parameters β and γ from the SIR model (9) corresponding to this epidemic. The population remains constant over the whole period, i.e. $N = 763$. Contacts of the pupils were limited to the people in school, thus forming a closed community – it seems that the population is as homogeneous as possible. The presymptomatic period is short, no deaths occurred and the recovered

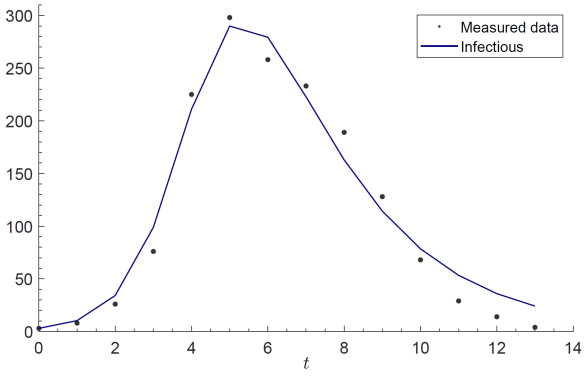

Figure 3: Measured data of the flu epidemic and the estimate of compartment I.

acquired sufficient immunity. One problem concerning the available data may occur, since in practical cases we do not possess the data which fit into the structure of the SIR model precisely. The data consists of the number of students confined to bed each day. Following [5], we assume the data to be from the compartment I.

Since we have data only from compartment I, we define the optimization problem of the form: Find β_m, γ_m satisfying

$$(\beta_m, \gamma_m)^T = \underset{(\tilde{\beta},\tilde{\gamma})^T \in \mathbb{R}^2}{\operatorname{argmin}} \sum_{j=0}^{D} |\tilde{I}(j\tau, \tilde{\beta}, \tilde{\gamma}) - I^j|^2, \tag{12}$$

where $\tau = 1$ corresponds to one day, which is the period with which we know the number of infected, I^j on the j-th day, $j = 0, \ldots, D$ with $D = 13$. The initial estimate is given by $(\beta^{(0)}, \gamma^{(0)})^T = (1, \frac{1}{7})^T$ and the stopping criterion

$$\|(\beta_m, \gamma_m)^T - (\beta_{m+1}, \gamma_{m+1})^T\|_\infty < 10^{-5}, \tag{13}$$

was satisfied after six iterations of the Levenberg-Marquardt algorithm. The resulting estimate of the parameters is $(\beta_m, \gamma_m)^T \approx (1.6998, 0.4469)^T$. Figure 3 shows that the estimated values of the compartment I are in good agreement with the data. However, after closer examination we find that the results do not quite correspond to the available data. Namely, the SIR model with the optimized parameters shows that the total number of people who suffered from the illness is 744, whereas the true number was 512. In addition, the value $\frac{1}{\gamma} \approx 2.24$ represents the expected time (in days) one spends in the Infectious compartment. This value is less then the observed value, which was three to seven days. This suggests that even in this simple case some unexpected issues limiting the accuracy of the model occur. This is a consequence of several facts. As stated above, the available data do not fit the model precisely – a person diagnosed with the illness has limited possibilities of spreading the disease because their contacts with the susceptible population are restricted. In addition, the pattern of the SIR model may not be entirely convenient for this particular disease. In order to adjust the model in accordance with the disease we need additional medical information which is not available.

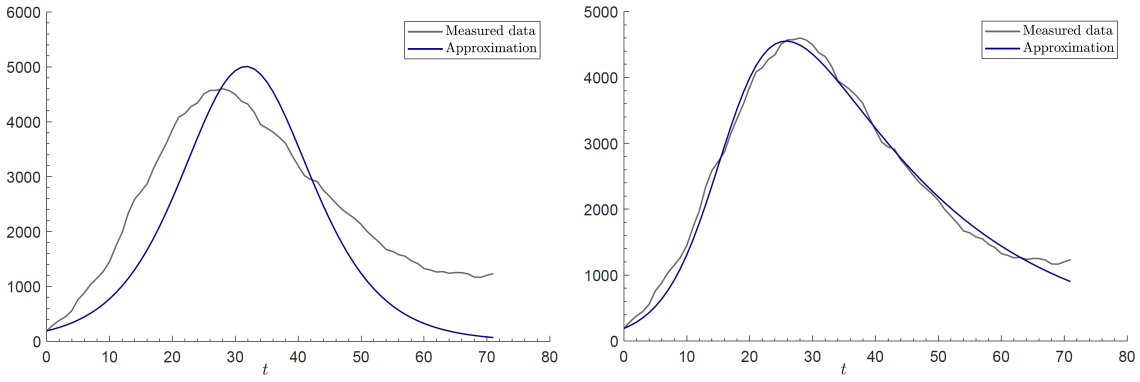

Figure 4: COVID-19, Infectious: Full population (left), effective population (right)

5.2. COVID-19 epidemic in the Czech Republic

Finally, we apply the presented numerical methods to the COVID-19 epidemiological data from the Czech Republic provided by the Ministry of Health of the Czech republic [3]. We chose the period from March 13, 2020, to May 24, 2020. The reasons are the following: On 13 March, the key measure forbidding retail sales and the sales of services in business premises came into effect and on 25 May the crucial part of the restrictive measures ended. It is therefore reasonable to assume that β and γ remain constant within this period, since adopting some restrictive measures against the spread of the disease decreases the value of parameter β, because the number of contacts of a person is reduced. The chosen period was the longest during the epidemic, where external conditions remained the same.

We optimized the parameters β and γ using the data from the compartment I only, i.e. the function to minimize is of the form (12) with $D = 71$ and $N = 1.065{\cdot}10^7$. The initial guess of the parameters is again given by $(\beta^{(0)}, \gamma^{(0)})^T = (1, 1)^T$. The stopping criterion (13) was achieved after 10 iterations. The computed estimate is $(\beta_m, \gamma_m)^T \approx (4.6687, 4.5244)^T$. The results from compartment I can be seen in Figure 4 (left). We note that the computed estimate gives the expected time a person remains infectious $\frac{1}{\gamma_m} \approx 0.22$ days, which is clearly unrealistic. Moreover, the model shows that the total number of recovered people at the and of the considered time interval is $6.15 \cdot 10^5$, while the actual value was 7750.

The reason why the SIR model gives such unrealistic results for the presented data is that the number of infected was very small in proportion to the total population of the Czech Republic and the population was not homogeneous, since the epidemic consisted of small local outbreaks, thus violating one of the basic assumptions of the SIR model. This consideration leads us to the introduction of an *effective population size*. The idea is to use a reduced population size which reflects the assumption of homogeneity within that smaller sub-population. The question is how to determine the size of the effective population. Our approach is to consider N not as a fixed constant (as it has been until now), but to treat it as an unknown parameter. For-

mally, the change is that instead of the parameter vector $(\beta, \gamma)^T$ for the SIR model, we now have the extended parameter vector $(\beta, \gamma, N)^T$. We note that N is present not only in the equations (9), but also in the initial condition (10), thus we use the approach from Section 2.1.

The initial guess was $(\beta^{(0)}, \gamma^{(0)}, N^{(0)})^T = (1, 1, 10^6)^T$. The computed results are $(\beta_m, \gamma_m, N_m)^T \approx (0.2587, 0.0444, 8593)$ and were reached after 50 iterations. Agreement with measured data has improved, cf. Figure 4 (right). The estimated total number of recovered is 7636, which is a good approximation of the true value 7750. The expected length of the infectious period is approximately 22 days. This is close to the length of the potential maximal infectious period of 15 to 21 days estimated in meta-analysis [1]. The estimate of the basic reproduction number $R_0 = \frac{\beta_m}{\gamma_m} \approx 5.8$ exceeds the values in the interval 2.4 to 3.4 estimated by meta-analysis.

To conclude, the presented method of the effective population considerably increased the accuracy of the basic SIR model in the situation when the SIR model itself failed due to high inconsistency of the measured data with the assumptions of the model. We have also tried other compartmental models such as the SIQR model, however not much improvement was observed over the basic SIR model with optimization of the effective population size.

Acknowledgements

This work was supported by grant No. 20-01074S of the Czech Science Foundation.

References

[1] Byrne, A.W. et al.: Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. BMJ Open, **10**(8) (2020).

[2] Dickinson, R.P. and Gelinas, R.J.: Sensitivity analysis of ordinary differential equation systems - a direct method. J. Comput. Phys., **21** (1976), 123–143.

[3] Komenda, M. et al.: COVID-19: Přehled aktuální situace v ČR, onemocnění aktuálně, https://onemocneni-aktualne.mzcr.cz/covid-19, accessed 30-11-2022.

[4] Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math., **11**(2) (1963): 431–441.

[5] Murray, J.D.: *Mathematical Biology I. An Introduction*, Springer, New York, 2002.

[6] Schittkowski, K.: *Numerical Data Fitting in Dynamical Systems: A Practical Introduction with Applications and Software*. Kluwer Academic Publishers, 2002.