mathematics

MDPI

Article

Inverse Coefficient Problem for Epidemiological Mean-Field Formulation

Viktoriya Petrakova [1,2]

[1] Institute of Computational Modeling, Siberian Branch of the Russian Academy of Sciences, Akademgorodok, 50/44, 660036 Krasnoyarsk, Russia; rikka@icm.krasn.ru
[2] Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, Koptyuga Ave., 4, 630090 Novosibirsk, Russia

Abstract: The paper proposes an approach to solving the inverse epidemiological problem, written in terms of the "mean-field" theory. Finding the coefficients of an epidemiological SIR mean-field model is reduced to solving an optimization problem, for the solution of which only zero-order methods can be used. An algorithm for the solution of the inverse coefficient problem is proposed. Computational experiments were carried out to compare the obtained solutions with respect to synthetic and real data. The results of computational experiments have shown the efficiency of this approach. Ways to further improve the approach have also been determined.

Keywords: SIR mean-field problem; inverse coefficient epidemiological problems; optimization problem; Nelder–Mead method

MSC: 92B99

check for updates

Citation: Petrakova, V. Inverse Coefficient Problem for Epidemiological Mean-Field Formulation. *Mathematics* **2024**, *12*, 3581. https://doi.org/10.3390/math12223581

Academic Editor: Cheng-Hung Huang

Received: 3 October 2024
Revised: 4 November 2024
Accepted: 14 November 2024
Published: 15 November 2024

1. Introduction

The coronavirus epidemic, which spread most in 2020–2022, showed the need to develop measures to contain the spread of disease to reduce mortality among the population and burdens on healthcare systems. At the same time, the measures introduced can vary quite significantly: from the development of a hand-washing culture to complete long-term isolation. These measures are perceived by society differently [1–4] and have different effects on the economy and other areas of public life. Methods of mathematical modeling help assess the consequences of decisions taken at different levels. At present, models of optimal control of various kinds are being significantly developed, since this approach is more flexible and allows accounting for the slightest changes in the choice of population and external planner behavior strategies. Thus, using the basic epidemiological model of the SIS, in [5], an optimal control model is proposed in which a planner uses tax revenues as direct funds for preventive measures for the population or for treatment of the infected. By analyzing the social costs of prevention and treatment, the authors determine the policy that is most cost-effective in different situations. In [6], a "mean-field model" is proposed, based on the premise that the selfish actions of each individual maximize personal utility, in contradiction to a socially optimal strategy that maximizes the total utility for the entire population. The authors assume that the reduction of contact with infected people should continue long after an epidemic has subsided.

One approach to modeling the dynamics of epidemic spread is optimal control models written in terms of "mean-field" (MF) theory. Here, it is assumed that the population is large enough and the influence of strategy adopted by an individual is negligible and does not affect the overall behavior of the population. This assumption allows one to pass the limit in explicit agent-based modeling problems and describe the behavior of the population with a small number of equations. One of the epidemiological mean-field models will be discussed in more detail in Section 2.1. We also note the paper with a review

of the use of various mean-field approaches for modeling the dynamics of epidemics [7]. Previously, studies [8,9] demonstrated the advantage of mean-field predictive models on real data over commonly used compartmental SIR-type models.

The need to solve inverse problems for various mathematical epidemiological formulations is caused by the presence of statistical data (usually noisy) describing the dynamics of disease development and a large set of parameters in a mathematical model used for forecasting. Parameters being reconstructed are usually physical, but difficult to determine, since they depend on a large number of factors, e.g., the contagiousness parameter (rate of virus spreading), used in an overwhelming majority of epidemiological models, depends on infectiousness of virus, frequency of contacts in the population, incubation period, and compliance with anti-virus restrictions.

The problems of reconstructing the coefficients of different epidemiological models from real data are the subject of interest to many researchers (see, e.g., studies [10–12]). But there are only a few papers on solving inverse mean-field problems [13–15]. The mentioned studies focus on the reconstruction of the Hamiltonian of the system (not a coefficient problem) and on non-epidemiological models. These papers consider classical MF formulations in the form of a system of two partial differential equations, used mainly to solve economic problems. The difficulties in finding solutions to inverse MF problems arise due to the strict restrictions on the choice of cost functional describing the behavior of the system to ensure the existence and uniqueness of the solution to the direct problem. Such conditions for epidemiological MF formulation are discussed in more detail in Section 2.2.1. To overcome them, in [13], e.g., the continuity equation is considered instead of the Fokker-Planck one, which is used in mean-field systems to describe the evolution of the population. This assumption leads to the fact that with zero current costs and a certain type of terminal conditions, the formulation is similar to a transfer optimization problem, for which the inverse problem has already been solved. The papers [14,15] consider the reconstruction of the Hamiltonian based on a limited set of noisy partial observations of population dynamics with a limited aperture. To achieve this goal, the authors formalize the inverse problem as an optimization problem with constraints on the residual function using the least-squares method with appropriate norms. The Chambolle–Pock method (a variation of the gradient method) is used to solve the inverse problem. The authors note that due to the strong incorrectness of the formulation, it is very difficult to obtain a high-quality reconstruction.

In this paper, the coefficient inverse MF problem is proposed for epidemiology. The formulation is based on the intersection of two ideas. First, with the correct choice of parameters and functionality, epidemiological mean-field models can provide a more accurate forecast of the virus spread dynamics due to additional assumptions about population behavior than simple compartmental SIR-type models. Second, researchers in the field of solving inverse mean-field problems (not epidemiological) note that due to restrictions on the existence and uniqueness of the solution of direct problems, inverse problems are incorrect and difficult to implement. Note that for epidemiological mean-field models, these restrictions are even stricter (see Section 2.2.1). However, the sensitivity analysis of the epidemiological MF model with respect to epidemiological parameters and the type of functional (see Section 2.2.2 for details), shows that the model is highly sensitive to the determination of epidemiological parameters. This allows us to consider the inverse mean-field epidemiological problem as a coefficient problem and to choose the function in such a way that it satisfies the specified uniqueness restrictions. This allows one to consider a simpler formulation in comparison to those proposed in [13–15]. Thus, the aim of this paper is to propose a formulation of the inverse mean-field epidemiological problem and to study it on real and synthetic data. In this sense, this paper follows the same ideas proposed in works [10,11,16] for SIR-based models described in the form of ordinary differential equation (ODE) systems. Inverse coefficient problems proposed for more complicated epidemiological models are discussed in [12,17,18] and others.

This work is also aimed at solving practical problems arising from epidemiological modeling. The proposed model is applied to real data on COVID-19 incidence in Novosibirsk in 2021. The inverse problem is formulated as a multicriterial optimization problem with constraints, which can only be solved by zero-order methods. Sections 3–5 present formulations, results, and comments for computational experiments on reconstruction of model parameters on synthetic and real data.

2. Mathematical Formulation

2.1. Direct Problem

As a direct mean-field problem, the formulation used was first introduced in [19] and called the *SIR Mean-Field Model with Total Control for All Epidemiological Groups (SIR TGC MF)*. This formulation differs from the one proposed, e.g., in [20], by the generality of the chosen strategy for the entire population. Their comparison is given in the work [19]. Here, the entire population is divided into three groups: susceptible people (S), infected (I), and recovered (R), who receive immunity after illness or dying. The formulation of a direct problem is the following: find the minima of the cost functional

$$J(m_{SIR}, \alpha) = \int_0^T \int_0^1 \left(\sum_{i \in \{S,I,R\}} G_i(m_{SIR}(t,x), \alpha(t,x)) m_i(t,x) + g(t,x,m_{SIR}(t,x)) \right) \mathrm{d}x\mathrm{d}t + \int_0^1 \Phi(m_{SIR}(T,x))\mathrm{d}x \tag{1}$$

with restrictions in the form of the system of convection-diffusion equations

$$\begin{cases} \partial_t m_S + \nabla(m_S \alpha) + \beta m_S m_I - \sigma_S^2 \Delta m_S/2 = 0, \\ \partial_t m_I + \nabla(m_I \alpha) - \beta m_S m_I + \gamma m_I - \sigma_I^2 \Delta m_I/2 = 0, \\ \partial_t m_R + \nabla(m_R \alpha) - \gamma m_I - \sigma_R^2 \Delta m_R/2 = 0 \end{cases} \tag{2}$$

with initial

$$m_i(0,x) = m_{0i}(x) \text{ on } \Omega \tag{3}$$

and Neumann boundary conditions

$$\partial m_i/\partial x = 0 \ \forall t \text{ and } x \in \Gamma_\Omega. \tag{4}$$

Here, the stochastic processes within the population are described using non-negative parameters σ_i, $i \in \{S,I,R\}$; $m_i(t,x) : [0,T] \times \Omega \to \mathrm{R}$ are the functions presenting the distribution of individuals in each epidemiological group $i \in \{S,I,R\}$ over the state space Ω at each time moment $t \in [0,T]$. State variable x indicates the population's loyalty to quarantine measures: $x = 0$ is the agent's dedication to imposed restriction measures, and $x = 1$ is the opposite. Function $\alpha(t,x) : [0,T] \times [0,1] \to \mathrm{R}$ denotes the representative agent's strategy for avoiding infection. Parameters β and γ are epidemiological and describe the process of virus spreading. Parameter β determines the contagiousness of the virus and γ is the individual's transition rate from the infected group to the recovered (recovery rate).

Functional (1) determined by three components. Here, G_i, $i = \{S,I,R\}$ is called *running cost* and denotes the cost for strategy realization for each epidemiological group of the population. The function g is called *current cost* and denotes the cost that is not explicitly dependent on strategy, and Φ is the *terminal cost*. The subscript $\cdot_{SIR}$ denotes any linear combination of functions m_S, m_I, m_R that is appropriate for the description of the process under research.

The using of the Lagrange multiplier method for Optimization Problem (1)–(4) leads to the conjugate system of partial differential equations

$$\begin{cases} \partial\psi_S/\partial t \ + {\sigma_S}^2\Delta\psi_S/2 + \alpha\cdot\partial\psi_S/\partial x + \beta m_I(\psi_I-\psi_S) = \\ \qquad\qquad -\sum\limits_{i\in\{S,I,R\}} m_i\partial G_i/\partial m_S - G_S \ - \partial g/\partial m_S, \\ \partial\psi_I/\partial t \ + {\sigma_I}^2\Delta\psi_I/2 + \alpha\cdot\partial\psi_I/\partial x + \beta m_S(\psi_I-\psi_S) + \gamma(\psi_R-\psi_I) = \\ \qquad\qquad -\sum\limits_{i\in\{S,I,R\}} m_i\partial G_i/\partial m_I - G_I \ - \partial g/\partial m_I, \\ \partial\psi_R/\partial t \ + {\sigma_R}^2\Delta\psi_R/2 + \alpha\cdot\partial\psi_R/\partial x = \\ \qquad\qquad -\sum\limits_{i\in\{S,I,R\}} m_i\partial G_i/\partial m_R - G_R \ - \partial g/\partial m_R \end{cases} \tag{5}$$

with conditions on the time horizon

$$\psi_i(T,\,x) = \frac{\partial\Phi}{\partial m_i}(T,x) \quad \forall\, x\in[0,1],\ \forall i\in\{S,I,R\} \tag{6}$$

and the boundary conditions

$$\partial\psi_i/\partial x = 0\ \forall t \text{ and } x\in\Gamma_\Omega. \tag{7}$$

In addition to the conjugate system, optimal conditions on α function should be satisfied for all $\bar{\alpha}\in \mathrm{R}\ \forall(t,x)\in[0,T]\times[0,1]$

$$\sum_{j\in\{S,I,R\}} m_j\left(\frac{\partial G_j}{\partial\bar{\alpha}} + \frac{\partial\psi_j}{\partial x}\right) = 0. \tag{8}$$

Note that Systems (5)–(8) are valid when the following conditions are performed:

$$\alpha(t,0) = \alpha(t,1) = 0 \ \ \forall t\in[0,T]. \tag{9}$$

Thus, the set of systems of Equations (2)–(9) describes the optimum of a dynamic system, the evolution of which is determined by Cost Function (1).

For the described computational experiments below, an iterative algorithm is used for the solution of the epidemiological mean-field problem (2)–(9). This algorithm is described in [19] in detail. The monotone finite-difference approximation for Systems (2)–(7) is proposed. Stability and convergence estimates are shown. The optimal value of strategy α is found iteratively by successive solutions of finite-difference analogs of Equations (2)–(7). The process usually converges from 4 to 5 to about 20 iterations.

2.2. Motivation for Form of Inverse Problem Formulation Choice

2.2.1. Restriction on Existence and Uniqueness of Direct Problem

The main difficulties of the solution of the direct MF problem used for real-life applications are based on strict restrictions on the domain of the existence and uniqueness of the solution. The most general condition for satisfying them is the convexity of the functional with respect to control variables. In [19], these restrictions are formulated for the SIR MF problem with individual control for each epidemiological group (see Property 4 in [19]). Below, a reformulation of these restrictions is proposed for the SIR TGC MF model (2)–(9) used in this study.

Property 1 (Reformulation of Property 4 [19]). *Conditions for the existence and uniqueness of a weak solution of the SIR TGC MF model. Assume that running cost function G_i grows quadratically with respect to control function α, and the following conditions are satisfied:*

$$\begin{aligned} &\exists M > 0: \quad \left(\frac{\partial G_i}{\partial \alpha}(m_{SIR}, \alpha) - \frac{\partial G_i}{\partial \tilde{\alpha}}(m_{SIR}, \tilde{\alpha})\right) \cdot (\alpha - \tilde{\alpha}) > 0 \\ &\forall \alpha, \tilde{\alpha}: \alpha \neq \tilde{\alpha}, \text{ whenever } \left|\frac{\partial G_i}{\partial \alpha}\right|, \left|\frac{\partial G_i}{\partial \tilde{\alpha}}\right| > M. \end{aligned} \tag{10}$$

Assume that g, Φ are non-decreasing with respect to m, bounded below and satisfy

$$\begin{aligned} &\forall L > 0 \; g_L := \sup_{m \in [0,L]} |g(t,x,m)| \in L^1((0,T) \times \Omega), \\ &\forall L > 0 \; \Phi_L := \sup_{m \in [0,L]} |\Phi(T,x)| \in L^1(\Omega), \\ &f_i \in L_1((0,T) \times \Omega), \; \forall i \in \{S, I, R\}, \end{aligned} \tag{11}$$

where f_i, $i \in \{S, I, R\}$ are the right parts of (2):

$$f_S = -\beta m_S m_I; \; f_I = \beta m_S m_I - \gamma m_I; \; f_R = \gamma m_I. \tag{12}$$

Then, for any $m_{0i} \in L^\infty(\Sigma)_+$ such as $\log m_{0i} \in L^1(\Omega)$, there exists a unique weak solution (m_{SIR}, ψ_{SIR}) to the system of (2)–(9).

2.2.2. Sensitivity Analysis

Thus, we are quite strongly limited by the area in which a unique solution to a direct problem exists. At the same time, the study of the difference in sensitivity of the modeling results to the parameters of the model for both formulation [19] in the form of the MF SIR problem and for the compartmental SIR model shows that the main contribution to the dispersion of output is made by change in the epidemiological parameters of the model (parameters of contagiousness β and recovery γ). In Figure 1, the result of the sensitivity analysis for the MF SIR model is proposed. This result was first presented in the author's previous work [19], but will be repeated here because it is important. Here, the modeled numbers of susceptible (S group) and infected people (I group) were used as outputs. As input parameters, the following vector was used:

$$\vec{q} = (\beta, \gamma, I_0, \sigma_S, \sigma_I, \sigma_R, x_S^c, x_I^c, x_R^c, \sigma_S^c, \sigma_I^c, \sigma_R^c). \tag{13}$$

Here, $x_S^c, x_I^c, x_R^c, \sigma_S^c, \sigma_I^c, \sigma_R^c$ characterize the initial distribution of each epidemiological part of the population by the following way:

$$m_{0i} = \frac{A_i}{B_i}\left(\exp\left(-\frac{(x - x_i^c)^2}{2(\sigma_i^c)^2}\right)/\sigma_i^c\sqrt{2\pi} + a_i x^2 + b_i(1-x)^2\right) \; \forall i \in \{S, I, R\}, \tag{14}$$

where A_i is the proportion of the current group in relation to the total population at the initial time; B_i is the normalization coefficient equal to the integral over Ω of the expression in brackets; $a_i = \exp\left(-(1-x_i^c)^2 \Big/ 2(\sigma_i^c)^2\right)(1 - x_i^c)/(2(\sigma_i^c)^3\sqrt{2\pi})$ and $b_i = \exp\left(-(x_i^c)^2 \Big/ 2(\sigma_i^c)^2\right) \cdot (x_i^c)/(2(\sigma_i^c)^3\sqrt{2\pi})$ are expressions that ensure boundary conditions (4) for $m(0,x)$. Physically, this means that we define initial distributions as Gaussian (though with some small corrections to comply with the boundary conditions). Parameter I_0 is the proportion of the current group in relation to the total population at the initial time for the infected group. Here, A_I coincides with I_0. For the estimation of sensitivity, the Extended Fourier Amplitude Sensitivity Test (eFAST) [21] was used. The eFAST method allows dividing the total variance of model output into components, corresponding to the

model's input parameters. Variance caused by any given parameter and their interaction is quantified by sensitivity indices being measurable indicators of the model's sensitivity to parameter identification. The amplitudes of sensitivity indices are presented in Figure 1. The inscriptions in the figure correspond to the order in which parameters are written, defined in (13), so *'sig_S'*, *'sig_I'*, *'sig_R'* correspond to model parameters σ_S, σ_I, σ_R; *'E_S'*, *'E_I'*, *'E_R'* for x_S^c, x_I^c, x_R^c and *'disp_S'*, *'disp_I'*, *'disp_R'* for $\sigma_S^c, \sigma_I^c, \sigma_R^c$.

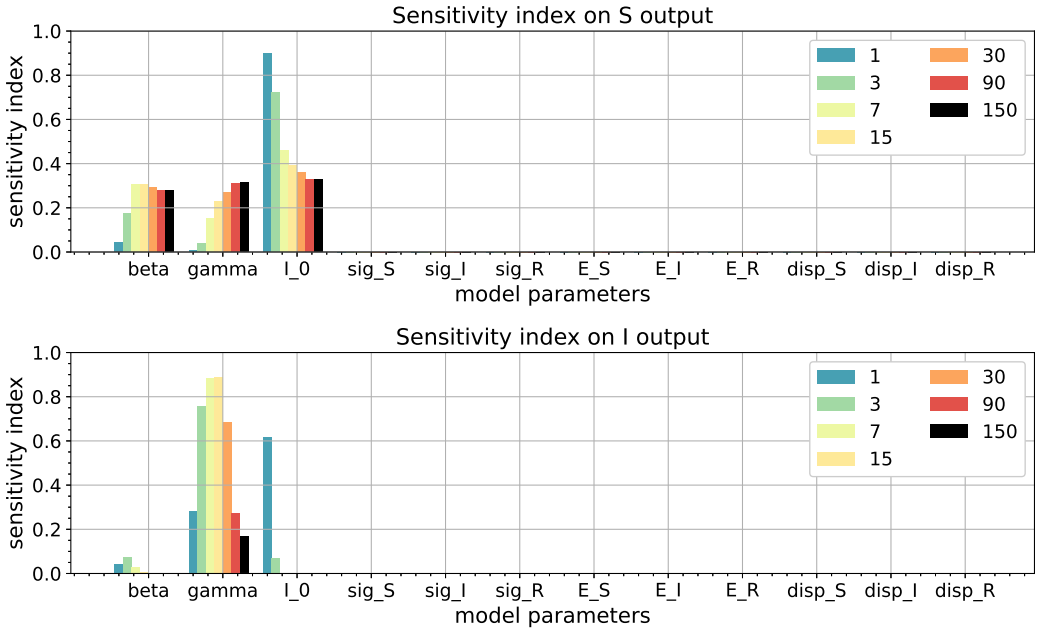

Figure 1. Sensitivity indices for SIR MF model (2)–(9) with set of input parameters (13) for different simulation times: $T = 1, 3, 7, 15, 30, 90, 150$ days for S and I outputs.

Let us comment on the results of the sensitivity analysis. The presented model with constant parameters describes only one outbreak of an epidemic. In the case averaged over all possible scenarios, the peak of epidemic development occurs at two weeks, which is reflected by sensitivity indexes on Ith output. After that, the number of infected people falls, and the variation of the parameter γ responsible for the recovery rate γ, accordingly plays a smaller role. This explains the changes in sensitivity indices for γ, Ith output. In contrast, the number of people in the S group is greater than in others and close to the population size. It implies that the values of sensitivity indices remain at the same level over time. For a small time period, the initial number of infected people (I_0) is a key parameter.

The influence of functional components in formulation (1) on the type of optimal control and modeling result is presented in the work [19] (Section 5). The main conclusion is that even for a modeling period of 10 days, the relative difference between the predicted number of infected people for the MF SIR model and the simple compartmental SIR one with the same epidemiological parameters can reach 0.1%. For a population size of 1 million people, this difference is about 1000 people.

2.3. Inverse Problem Formulation

Thus, based on the assumption that the epidemiological parameters are key for model (2)–(9), for real applications, the inverse coefficient problem is considered, instead of recovering the functions, as was carried out in [13–15]. The idea here is to put

$$G_i(m_{SIR}(t,x),\alpha) = a_c\frac{\alpha^2}{2}m_i;\;\; g(t,x,m_{SIR}(t,x)) = dm_I^2(t,x);\;\; \Phi(m_{SIR}(T,x)) = b_c m_I^2(T,x)$$

in direct problem functional (1). Here, the $a_c, b_c \in R$ are unknown parameters, $d \in R$ is known. This choice is made for several reasons. First, assuming that the functions G_i, g, Φ

are quadratic in control and distribution densities m_i, we remain in the domain of existence and uniqueness of the solution to the direct problem. Second, a very general assumption is made here about the development of an epidemic, which is that the population acts in such a way (or is subject to such actions on the part of the government) to reduce the number of those infected (choice of function g).

Then, the inverse problem is to find the unknown coefficients $\vec{\theta} = \{\beta, \gamma, a_c, b_c\}$ that minimize the following function:

$$\min_{\vec{\theta}} L = \sum_q \left(\int_0^1 m_I(q,x)dx - I_q \right)^2 + \sum_p \left(\int_0^1 m_R(p,x)dx - R_p \right)^2. \quad (15)$$

Here, I_q and R_p are known measures of infected and recovered parts of the population at some time moments, determined by p, q, the number of which is known in advance. Thus, in the considered formulation, we optimize both the epidemiological component (through parameters β, γ) and the description of the population (parameters a_c, b_c).

3. Materials and Methods

3.1. Nelder–Mead Optimization Method

Here, the Nelder–Mead method is used to solve the optimization problem. In 1965 [22], the Nelder–Mead (N-M) sequential optimization method was proposed as a direct method for local optimization of unconstrained problems and was a modification of the simplex method. The idea of the method is to compare values of the target function at $(n+1)$ vertices of the simplex and move the simplex towards the optimal point using an iterative procedure, where n is a number of model parameters to be reconstructed. In the original simplex method, a regular simplex was used at each stage. Nelder and Mead proposed several modifications of this method that allow simplices to be irregular. The result was a very reliable direct search method, which is one of the most effective if $n \leq 6$. Its effectiveness in practical applications has been demonstrated by a rapid initial decrease in function values [23,24]. Recently, this well-known algorithm has been used in combination with global search algorithms such as random search [24] and the genetic algorithm [25].

The implementation of the algorithm involves calculating the values of the target function (defined here as (15)) at vertices of the simplex. Depending on the obtained values, the simplex changes the worst point from the set to a new one, which is closer to the local minimum. In a sense, the simplex creeps to the minimum value in the region. The condition for termination of the iterative process is the smallness of sides or area of the simplex obtained at the next iteration.

Here, the choice of the Nelder–Mead method is determined by the fact that (a) this method does not require calculating the derivatives of the target function with respect to the reconstructed parameters; (b) it is effective at a low speed of objective function calculation. Recall that to calculate one value of the target function here, to solve the MF direct problem is required, and the solution to the MF direct problem is found by an iterative algorithm. Thus, for future works, the N-M method can be replaced by any suitable one for the considered formulation and constraints. More information about zero-order methods can be found in the monograph [26].

3.2. The Algorithm for the Solution of the Coefficient Inverse SIR MF Problem

Thus, for the constructed coefficient inverse problem, the following Algorithm 1 can be used.

Algorithm 1 Algorithm for the solution of the coefficient inverse SIR MF problem

1. Choose the initial value of vector $\vec{\theta}$ being reconstructed and restrictions on the search area of each component of $\vec{\theta}$. Denote the obtained initial vector as $\vec{\theta}^0$;
2. Put number of iteration k equaled zero ($k = 0$);
3. Solve direct SIR MF problem (2)–(9), (15) with known parameters $\vec{\theta}^k$;
4. Put $k = k + 1$;
5. Compute $2 \cdot |\vec{\theta}^k| + 1$ values of target function (15) on kth iteration at a small deviation from each vertex of the simplex under study. Do the iteration of N-M algorithm and find new vector $\vec{\theta}^k$;
6. Check the stopping criterion of optimization method. If it is satisfied, then set θ^k as the desired solution to the optimization problem; otherwise, return to step 3.

The solution of the direct problem of SIR MF was carried out using finite-difference approximation proposed in [19].

3.3. The Description of Computational Experiments

This paper proposes two types of computational experiments on synthetic and real data. Below is a description of each computational experiment.

Experiment №1

For the first experiment, the recovery possibility of only the β and γ parameters will be studied on synthetic data. As a set of data, the solution of a simple compartmental SIR model of the form

$$\begin{cases} dm_S/dt = -\beta m_S m_I, \\ dm_I/dt = \beta m_S m_I - \gamma m_I, \\ dm_R/dt = \gamma m_I \end{cases} \tag{16}$$

will be considered with initial values

$$m_S(0) = S_0;\ m_I(0) = I_0;\ m_R(0) = R_0$$

and parameters describing the population and epidemiological process by the following way:

$$N = 1000,\ S_0 = 800/N,\ I_0 = 200/N,\ R_0 = 0,\ \beta = 0.7,\ \gamma = 0.3,\ d = 0.001. \tag{17}$$

Here, N is the population size. These parameters describe a single outbreak of an epidemic over a period of $T = 10$ days with dynamics visually displayed in Figure 2.

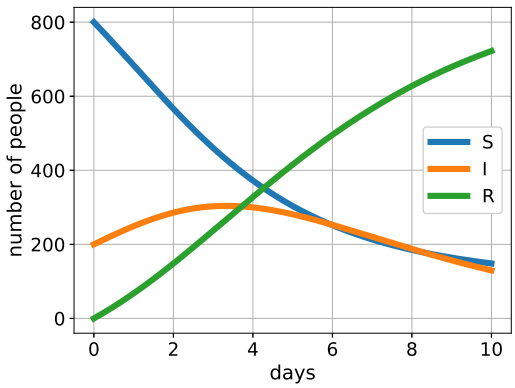

Figure 2. Visual representation of epidemic dynamics described by SIR model (16) with parameters (17).

For the inverse SIR MF problem, values of β and γ parameters were reconstructed only based on 10 daily measurements of the infected part of the population, that is, in target function (15) we have the measurements $\vec{I} = (I_0, \ldots, I_9)$ and do not have any of R_p. The parameters a_c and b_c are known here and $a_c = 1$, $b_c = 0$.

Experiment №2

For the second experiment, the same data as in Experiment №1 will be used, but a "single noise" into the known solutions $(I_0, \ldots, I_9)$ that does not exceed 1 or 5% is introduced. The noise value is generated the same for each daily measurement and displays the systematic error of the observer. The term "single noise" means that for an experiment, a single value was chosen from a uniform distribution to capture the noise level. This single value was distributed over the entire set $(I_0, \ldots, I_9)$.

Experiments №3 and №4

Experiments №3 and №4 repeat Experiments №1 and №2, differing only in that all four parameters will be reconstructed: $\vec{\theta} = (\beta, \gamma, a_c, b_c)$.

Experiment №5

The fifth experiment is aimed at testing the performance of the algorithm on real data. As real data, the dynamic of the spread of COVID-19 in Novosibirsk is chosen for the 150 days from 13 April 2021. The data collected in this city and some others can be found at the link https://covid19-modeling.ru/data (acessed on 13 November 2024). As in the previous experiment, the full set of parameters $\vec{\theta} = (\beta, \gamma, a_c, b_c)$ will be reconstructed. For the target function (15) computation, the daily measurements of the number of infected people $\vec{I} = (I_0, \ldots, I_{149})$ will be used. The modeling will be carried out for different time periods, which is named the *window of modeling* (w). This means that the parameters are reconstructed using only data of w days and for the specified period. For the full 150-day period, gluing of the corresponding number of periods on w days will be made. The comparison of errors of modeling will be provided for several values of w.

Experiment №6

Experiment №6 repeats Experiment №5, but here, for the parameter reconstruction the daily measurements of both the infected and recovered parts of the population will be used. That is, in target function (15), we have the full set of measurements $\vec{I} = (I_0, \ldots, I_{149})$ and $\vec{R} = (R_0, \ldots, R_{149})$.

Experiment №7 and №8

Experiments №7 and №8 estimate the quality of parameter reconstruction with a limited amount of data. For Experiment №7, it is proposed that the entire set of daily measurements of infected people is known $\vec{I} = (I_0, \ldots, I_{149})$, but daily measurements of recovered people are not full. So, for $\vec{R}$, only part of the sample is known. Information on what days the measurements are known is chosen randomly. The number of known measurements was chosen as 90, 70, 50, and 30 % of the number of whole measurements. The solving of the inverse problem was conducted 10 times for $w = 10$ days. The averaged parameters β, γ, a_c, b_c, and their standard deviations will be obtained.

For Experiment №8, the same will be carried out, except that here, daily measurements of infected people are known partially too.

4. Results

This section presents the results of the computational experiments described above.

Experiment №1 results

For Experiment №1, which was aimed at recovering the main epidemiological parameters β, γ from synthetic data without any noise, the recovered parameters are $\{0.69179788, 0.30081996\}$ versus the true parameter, which equaled $\{0.7, 0.3\}$. The initial values of β, γ for the optimization process were chosen equaled $\{0.5, 0.5\}$. Relative errors of recover are 1.2% and 0.3%. The 44 iterations of the algorithm, described in Section 3.2, were required.

Experiment №2 results

For the second experiment, the noise in measurements was put on a thousand similar optimization problems, where different generated values of noise were solved. The result is presented in Figures 3 and 4 for noises, the values of which do not exceed 1% and 5%.

Figure 3 should be interpreted as follows. Experiment №2 was conducted 1000 times for different noise levels within the specified limits (does not exceed 1 or 5%). The resulting samples of 1000 reconstructed β and γ values were divided into two parts of 500 elements. Scatter diagrams were constructed based on the resulting subsamples. Ideally, the points should be concentrated around the intersection of the red lines, which denotes the true solution. The location of the "cloud" of points shows the bias of the estimates obtained relative to the true value. The same description is true for Figures 4–6.

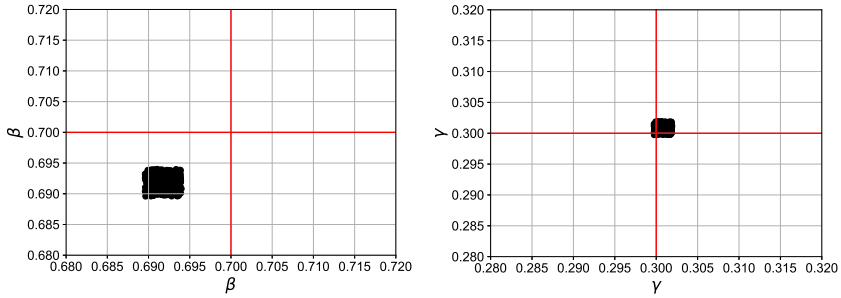

Figure 3. Values of reconstructed parameters β (**left**) and γ (**right**) for 1000 carried out Experiments №2 with uniform noise in data under 1%.

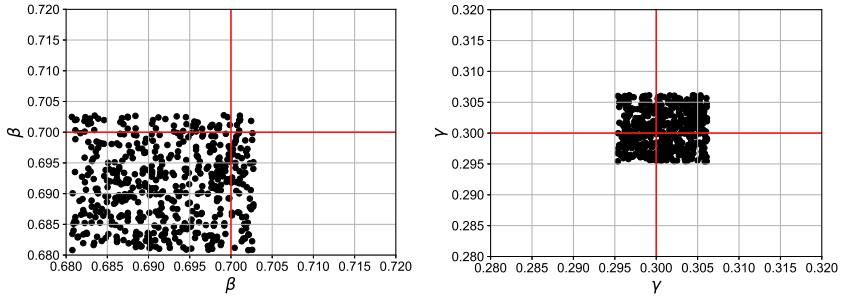

Figure 4. Values of reconstructed parameters β (**left**) and γ (**right**) for 1000 carried out Experiments №2 with uniform noise in data under 5%.

Here and after, metrics *rmse* and *rae* are used for estimating the results of computational experiments. The metrics are explained below.

$$rmse = \sqrt{\frac{1}{M}\sum_{k}(y - y_k)^2},$$

$$rae = \frac{100}{M}\sum_{k}\frac{|y - y_k|}{|y|},$$

where y is the true observation; y_k is the obtained value from the kth experiment realization; M is the number of realizations. Thus, *rmse* represents the root of the mean square error and *rae* is relative absolute error.

So, for Experiment №2, when introducing noise does not exceed 1%:

rmse for β: 0.00830; *rae* for β: 1.17%;

rmse for γ: 0.00103; *rae* for γ: 0.29%.

When introducing noise does not exceed 5%:

rmse for β is 0.01036; *rae* for β is 1.22%;

rmse for γ is 0.00320; *rae* for γ is 0.91%.

Note that the reconstructed β parameters are generally lower than the true value. It can be explained in the following way. First, the solution to MF problem is also obtained

with some approximation error due to use of finite-difference approach to solve the system of partial differential equations. This introduces an additional error into the target function. Second, due to the choice of the initial value $\beta = 0.5$, the simplex comes near the true solution $\beta = 0.7$ from below. Thus, at some value of β falling within the region shown in Figure 3, the target function reaches its minimum. This can be corrected by increasing the grid dimension when solving the mean-field problem. However, this leads to a huge increase in the calculation time, since the algorithm for solving the mean-field problem is iterative.

Experiment №3 results

The same pattern of computational results can be found for Experiments №3 and №4. The reconstructed parameters for Experiment №3 are $\{0.69366844, 0.30132925, 0.58164058, 0.00000000\}$ for β, γ, a_c, b_c. Thus, the relative absolute errors for β and γ amount to 0.90% and 0.44% correspondingly. The reconstructed parameter b_c equals zero because here as the true measurements the solution of compartmental SIR model (16) was used. This means that the optimal strategy α used in SIR MF model should be zero, and this is achieved when $b_c = 0$ and any value of a_c.

Experiment №4 results

For the fourth experiment, where the noise in measurements was added, the three hundred similar optimization problems with different generated values of noise were solved. The number of similar computations was reduced because of the long computation time of recovering four parameters. The visual representation of distribution of reconstructed parameters β and γ is similar to that obtained from Experiment №2 (see Figures 3 and 4). The *rmse* and *rae* values for β and γ parameters are presented below.

When introducing noise does not exceed 1%:

rmse for β: 0.00830; *rae* for β: 1.17%;

rmse for γ: 0.00103; *rae* for γ: 0.29%.

When introducing noise does not exceed 5%:

rmse for β is 0.01036; *rae* for β is 1.22%;

rmse for γ is 0.00320; *rae* for γ is 0.91%.

The reconstructed parameters a_c and b_c are presented in Figures 5 and 6.

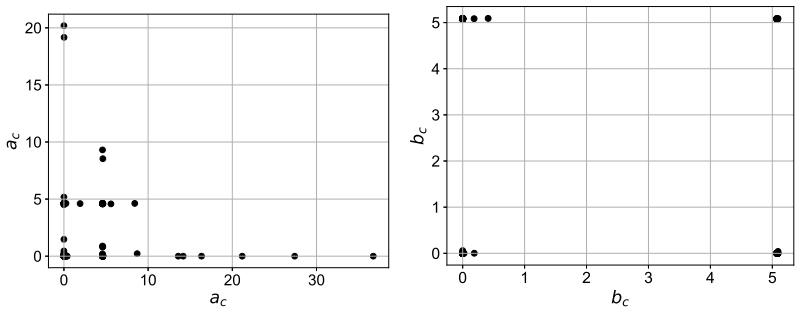

Figure 5. Values of reconstructed parameters a_c (**left**) and b_c (**right**) for 300 carried out Experiments №4 with uniform noise in data under 1%.

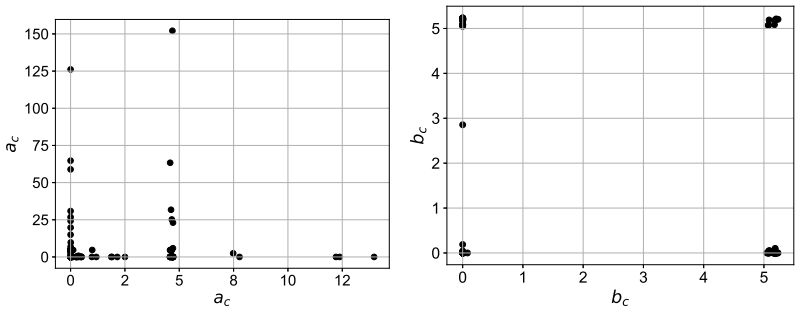

Figure 6. Values of reconstructed parameters a_c (**left**) and b_c (**right**) for the 300 carried out Experiments №4 with uniform noise in data under 5%.

Experiments №5 and №6 results

Figures 7 and 8 show the result of computational experiments carried out on real data. In the case of Figure 7, parameters were reconstructed using only measurements of the infected part (I group) of the population. After recovering, the obtained parameters were used for direct modeling to compare the quality of the approximation. For the result depicted in Figure 8, the same was carried out, but for parameter reconstruction, the measurements of infected and recovered parts (I and R groups) of the population were used. Figures 7 and 8 show that the result for modeling window length (w) equaled 10 days.

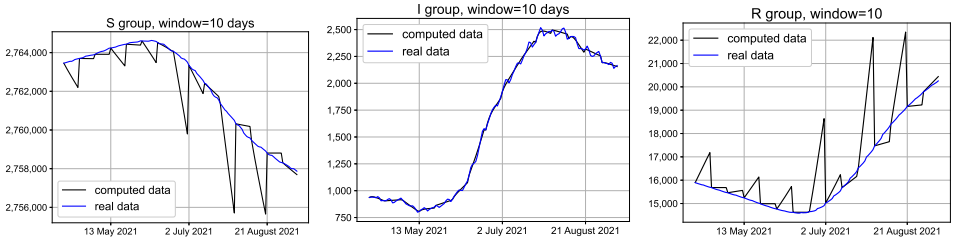

Figure 7. Result of direct modeling using SIR MF model in comparison with real data when model parameters were reconstructed using daily measurements of I group.

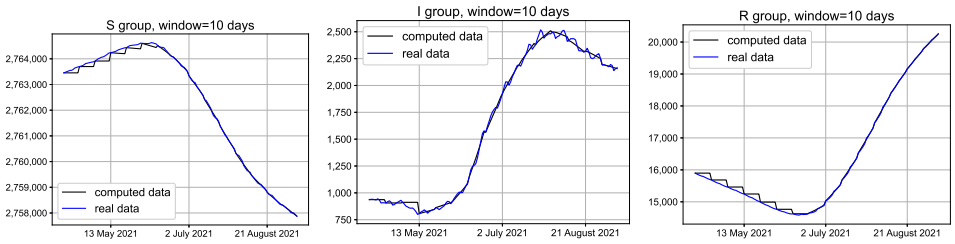

Figure 8. Result of direct modeling using SIR MF model in comparison with real data when model parameters were reconstructed using daily measurements of I and R groups.

Table 1 includes the values of *rmse* and *rae* for several values of window length w. Note that as the window length increases, the modeling error increases. This happens because the spread of disease in reality is a non-stationary process, but the modeling parameters are kept constant for the entire modeling period. The longer modeling period, the worse reconstructed parameters approximate the real epidemiological process.

Table 1. Error values of SIR MF model in comparison with real data depending on chosen window length (w) for reconstruction of model parameters.

Window Length (w) (in Days)	***rmse* (in People)**			***rae* (in Percent)**		
	***S* Group**	***I* Group**	***R* Group**	***S* Group**	***I* Group**	***R* Group**
When using only measurements of infected (I) part of population						
3	779.1	5.2	779.0	0.01	0.14	2.01
5	1052.0	9.0	1052.1	0.02	0.31	3.69
10	1124.1	19.5	1124.0	0.02	0.93	3.77
15	1198.5	21.6	1198.2	0.03	1.00	5.15
When using only measurements of infected and recovered (I, R) part of population						
3	72.9	36.8	54.9	0.00	1.11	0.21
5	110.7	50.7	79.8	0.00	2.07	0.33
10	89.8	28.3	79.5	0.00	1.49	0.33
15	336.4	147.1	298.8	0.01	5.51	1.15

Figure 9 shows the difference in the determination of parameters β, γ, a_c, b_c depending on whether they were reconstructed using only the measurements of the infected part of the population or both infected and recovered. Note, that values change non-smooth with changing time intervals, especially for the a_c and b_c. The black dots for parameters β, γ behave quite smoothly. The blue ones change non-smoothly due to the non-uniqueness of the minimum target function, which takes into account data only for the infected group. Thus, additional data on the recovered group level out the situation. Parameters a_c and b_c determine the growth of components of the cost functional (1). Their values depend on the chosen optimal population's strategy in the considered time period of modeling and the current incidence rate. These values are described by the behavior of both real data and population on each individual interval independently and should not change smoothly.

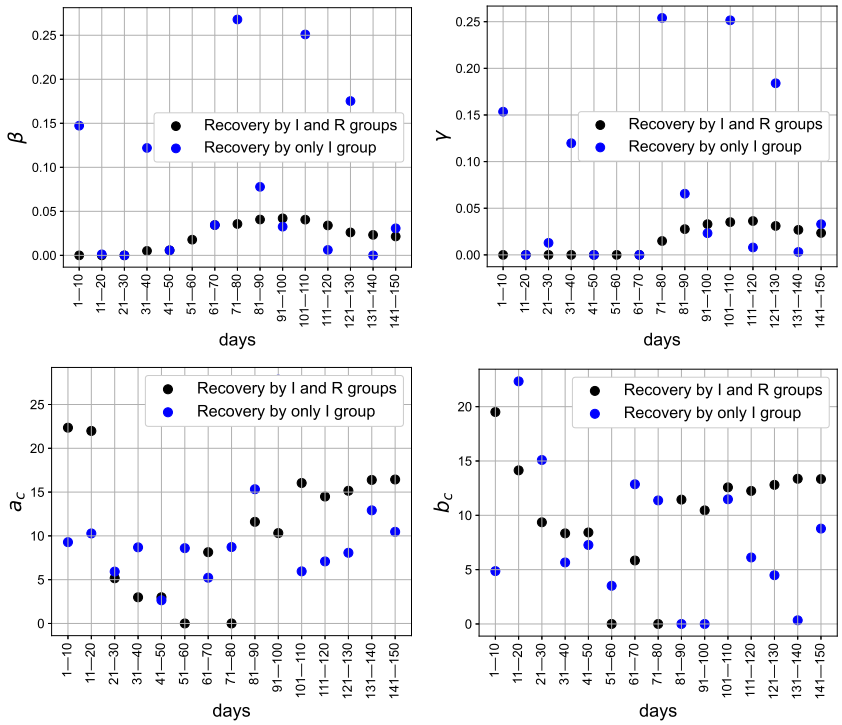

Figure 9. Value of the reconstructed parameters in each 10-day simulation window relative to real data.

Experiment №7 results

Figure 10 shows the modeling result when the parameters were reconstructed using a full set of measurements of the infected part (I group) of the population and a limited number of measurements of recovered one (R group). The length of the simulation window w was 10 days. The size of the available sample for the R group for reconstruction was chosen to be 30% rounded down. Thus, for computational experiment, the parameters $\vec{\theta}$ were recovered over a 10-day period using known daily measurements of number of infected $\vec{I} = (I_0, \ldots, I_9)$ and only three measurements of recovered group $\vec{R} = (R_{p_1}, R_{p_2}, R_{p_3})$. The values $p_1, p_2, p_3 \in \{0, 1, \ldots, 9\}$ were chosen randomly. Ten identical runs were conducted with different choices of p_1, p_2, p_3. The reconstructed parameters were averaged and a forecast built for them is shown in Figure 10. The *rmse* and *rae* for the obtained graph are 32.5 (in people) and 1.6% correspondingly for the I group. Thus, the use of additional data (in addition to infected measurements), even in small quantities, allows us to significantly improve the forecast. Table 2 contains the obtained *rmse* and *rae* values and their standard deviations over similar runs for the described above experiment in dependence on the size of the available sample for R group. The higher the percentage indicated, the more data from group R were used in simulations.

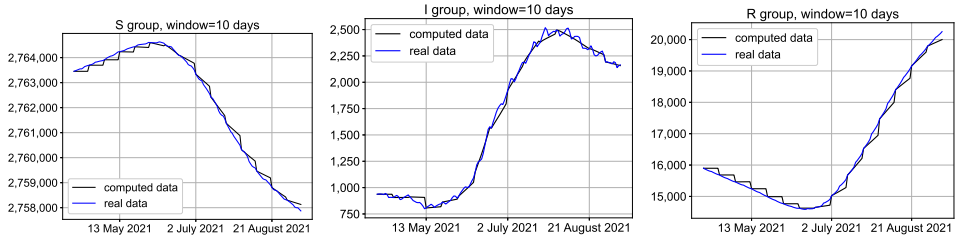

Figure 10. Result of direct modeling using SIR MF model in comparison with real data when model parameters were reconstructed using daily measurements of I group and only 30% of known measurements of R group.

Table 2. Error values (*rmse* (in people), *rae* (in percent)) of SIR MF model in comparison with real data for $w = 10$ days depending on percentage of measurements in R group used for modeling of total number of available ones. Here, daily data of measurements in I group were used. Designation *std* denotes standard deviation of depicted error for 10 similar runs.

Percentage	*S* Group				*I*				*R*			
	rmse	***std, rmse***	***rae***	***std, rae***	***rmse***	***std, rmse***	***rae***	***std, rae***	***rmse***	***std, rmse***	***rae***	***std, rae***
90	89.7	0.8	0.0021	0.0000	28.4	0.3	1.4916	0.0223	89.7	0.8	0.3319	0.0007
70	92.2	10.5	0.0022	0.0002	31.6	8.7	1.5903	0.1567	92.2	10.5	0.3371	0.0114
50	109.1	39.7	0.0025	0.0008	32.5	10.4	1.6104	0.2386	109.1	39.7	0.3913	0.1008
30	199.8	33.5	0.0045	0.0008	44.1	10.6	1.9892	0.3371	199.8	33.5	0.7151	0.1343

Experiment №8 results

The same was carried out when the limited set of measurements of I and R groups were used for parameter reconstruction. Figure 11 shows the modeling result when parameters were reconstructed using a limited set of measurements of the infected part (I group) of the population and a limited set of measurements of recovered one (R group). The length of the simulation window w was 10 days. The size of the available sample for I and R groups for reconstruction was chosen to be 30% rounded down. Thus, for the computational experiment, parameters $\vec{\theta}$ were recovered over a 10-day period using known three measurements of infected $\vec{I} = (I_{q_1}, I_{q_2}, I_{q_3})$ and three measurements of the R group $\vec{R} = (R_{p_1}, R_{p_2}, R_{p_3})$ of recovered group. The values $q_1, q_2, q_3 \in \{0, 1, \ldots, 9\}$ and $p_1, p_2, p_3 \in \{0, 1, \ldots, 9\}$ were chosen randomly. Ten identical runs were conducted with different choices of p_1, p_2, p_3. The reconstructed parameters were averaged, and a forecast built for them is shown in Figure 11. The *rmse* and *rae* for the obtained graph are 59.8 (in people) and 2.5% correspondingly for the I group. Thus, even for a limited set of available data on different groups, reconstruction is performed better than when using complete knowledge about one group. Table 3 contains the obtained *rmse* and *rae* values and their standard deviations over similar runs for the described above experiment in dependence on the size of the available sample for I and R groups and modeling period w.

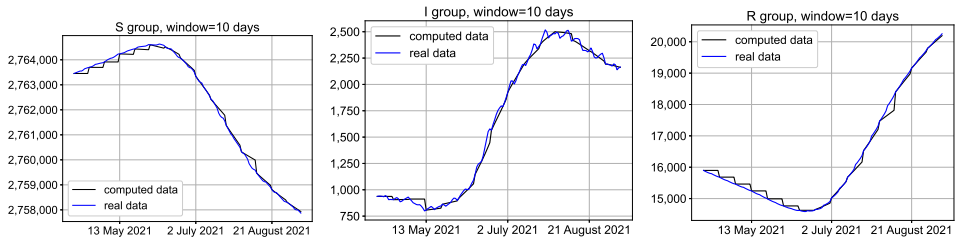

Figure 11. Result of direct modeling using SIR MF model in comparison with real data when model parameters were reconstructed using only 30% of known measurements of I and R groups.

Table 3. Error values (*rmse* (in people) and *rae* (in percent)) of SIR MF model in comparison with real data for several w days depending on percentage of measurements in I and R group used for modeling of total number of available ones.

Percentage	*S*				*I*				*R*			
	rmse	***std, rmse***	***rae***	***std, rae***	***rmse***	***std, rmse***	***rae***	***std, rae***	***rmse***	***std, rmse***	***rae***	***std, rae***
w = 5 days												
90	110.3	8.8	0.0026	0.0002	46.8	4.7	1.8558	0.1300	110.3	8.82	0.3502	0.0407
70	103.8	9.3	0.0024	0.0003	45.7	7.8	1.9018	0.2730	103.8	9.3	0.3467	0.0337
50	109.5	6.9	0.0026	0.0001	47.4	6.9	1.8844	0.2189	109.5	6.9	0.3766	0.0291
30	1014.9	101.9	0.0141	0.0023	74.4	8.9	2.5751	0.2320	1014.9	101.9	2.2034	0.3550
w = 10 days												
90	89.2	1.1	0.0021	0.0000	28.7	0.4	1.5121	0.0345	89.2	1.1	0.3319	0.0007
70	92.1	12.1	0.0022	0.0002	34.9	10.2	1.7122	0.2827	92.1	12.1	0.3391	0.0123
50	115.3	32.5	0.0026	0.0005	42.4	13.3	1.8706	0.2765	115.3	32.5	0.3853	0.0632
30	171.7	41.4	0.0037	0.0009	59.8	16.5	2.4699	0.4784	171.7	41.4	0.5309	0.0977
w = 15 days												
90	349.9	115.8	0.0085	0.0032	130.3	33.5	4.9078	1.0619	349.9	115.8	1.2234	0.4341
70	361.1	80.5	0.0083	0.0025	136.4	21.5	4.9777	0.8437	361.1	80.5	1.1672	0.3440
50	368.6	63.1	0.0084	0.0022	124.4	30.2	4.2193	0.9106	368.6	63.1	1.1692	0.2966
30	423.6	55.4	0.0101	0.0018	125.2	33.5	4.5973	1.3023	423.6	55.4	1.4555	0.2190
w = 30 days												
90	1016.1	1.1	0.0271	0.0000	529.0	1.8	19.1277	0.0402	1016.1	1.1	2.4698	0.0057
70	1016.1	1.1	0.0272	0.0000	528.7	4.5	19.1206	0.1014	1016.1	0.0	0.0272	0.0000
50	1016.1	1.1	0.0272	0.0000	527.3	4.3	19.0929	0.0950	1016.1	0.0	2.4750	0.0136
30	917.4	156.1	0.0243	0.0046	458.7	74.7	16.5436	2.5045	917.4	156.1	2.4173	0.5132

5. Discussion

This section is devoted to the analysis of the obtained results of computational experiments, discussion of current and future problems, and ways of their possible solution.

As Section 2.2.1 of this paper shows, the main difficulty in solving the inverse MF problem lies in the restricted domain of existence and uniqueness of the solution of the direct problem. This makes the problem of reconstruction of the system's Hamiltonian strongly ill-posed. However, the inverse problem for the epidemiological MF model can be reduced to a coefficient one. Sensitivity analysis of the model [19] shows the greatest influence on the modeling result is exerted by the choice of epidemiological parameters (for the SIR model, these are β and γ), and the form of the cost functional can be chosen that satisfies the restriction, but this growth can be operated by the set of parameters determined by the inverse problem.

The assertion that it is more important to restore the epidemiological parameters of the model relative to the general form of the Hamiltonian of the system is indirectly confirmed by the results of computational Experiments №1–4. Here, the solution of the differential SIR model with known parameters was chosen as "exact measurements", and the obtained restored epidemiological parameters for the mean-field model were similar to those used. At the same time, parameters a_c and b_c, describing the population, converged to values at which the solution of the mean-field model would coincide with the solution of the differential SIR model. Describe Figures 5 and 6 more detailed. In most cases, the value of the b_c parameter was determined to be close to zero. This is an expected value because when b_c equals zero and a_c is any, the solution of the SIR MF system coincides with the solution of the SIR differential model. This also approves the variation in the determination

of the a_c parameter. But in several cases, the value of the b_c parameter is close to 5. This value was chosen as the initial value of parameters for the optimization problem. Thus, the simplex method used in the optimization algorithm determines that the initial value of the parameter is close to the optimal one.

Note that here, for all experiments only, the parameters determining the growth of costs of implementing the strategy (a_c) and the terminal cost (b_c) were reconstructed while the growth rate of current costs remained unchanged, i.e., the function g in the cost functional was chosen to be reliably known. Such simplification is due to the fact that there are some relations on the growth rate of all three parts of the cost functional (cost of implementing the strategy, current costs, and terminal one), which determine the area of existence of the solution to the direct problem. At present, these relations are unknown and difficult to evaluate. This is the future direction of studies for our research group.

Another problem related to the reconstruction of the MF model parameters is the limited availability of measurements in each epidemiological group and the degree of confidence in such data. Statistical data of this kind are difficult to measure since there is a large number of asymptotic patients and/or deaths from the consequences of infection, expressed in the form of exacerbation of chronic diseases. Thus, to recover the model parameters, it would be desirable to use only the part of the data in which there is strict confidence. Attempts to implement this in computational experiments with synthetic data, when the reconstruction was made only by measuring the number of infected people, gave an encouraging result. However, in real data (Experiments №5 and №6), the use of such an approach leads to the non-uniqueness of the solution to the problem. An attempt to reduce the period for which the reconstruction is made also did not give results (see Table 1). Thus, to reconstruct the model parameters on real data, additional measurements in other epidemiological groups should be used. Experiments №7 and №8 show that the reconstruction on even a small number of data, but chosen in different groups, gives an improvement to the forecast.

Also, it is important to make a remark that any epidemiological model approximates the existing data set with some physically justified assumptions about the nature of the virus's behavior. The main problems associated with the reconstruction of the model parameters from real data seem to be the following.

– *Problem of statistical measurement consistency*. For example, incorrect data collection or interpretation can lead to the inability of a mathematical model to explain the ongoing epidemiological process. Inconsistent measurements mathematically lead to the presence of a minimum of the target function that differs from the real observed value. Reliability of the assessment of such parameters can be increased by using only those measurements that the researcher trusts (e.g., data on the number of hospitalized or tested individuals are known with enough reliability). However, as shown in the paper, this may not be enough. In this case, it is possible to include less reliable data in the target function, but with a certain weight factor. More details on taking into account poorly measured indicators, for example, asymptotic patients, are given in the work [27]. Here, just a part of the epidemiological groups is used to reconstruct the parameters of the compartmental SEIR-HCD model. Experiments using a limited set of real data to reconstruct the coefficients are carried out in this work (Experiments №7 and №8). Comparing the forecast errors obtained for the model with reconstructed parameters, it can be noted that the use of partially known data is possible.
– *Using regularization methods* can presumably provide an improvement in reconstruction. Regularization methods are used in many studies on forecasting the dynamics of epidemic spread for models in the form of a system of differential equations. For example, Tikhonov regularization is used to reconstruct parameters of compartmental SIR-type models, see, e.g., works [10,28]. The point here is to discretize the model and reduce it to a system of linear algebraic equations, and then use regularization. The same was carried out in [29]. The authors of [30] propose to use the total variation regularization method, which is popular in image denoising. Here, for a model with

available data, the likelihood function is constructed, which is optimized using a modification of the N-M algorithm. In the work [31], the total variation regularization method is used directly for the compartmental SIRQ model, where a regularizing integral term is added to the target function in the form penalty for irregularity. For the MF model studied here, the issue seems to be more complex, because for one calculation of the target function value, the iteration procedure for solving the discrete analog of the mean-field system (2)–(9) is performed. There is an idea of how to use assumptions about already recovered parameters for regularization. Let the parameters of the model for some relatively short period of modeling, e.g., several days be known. Then, the *epidemiological* parameters for the next consecutive period of time of the same duration will not differ much from those already known. This idea is indirectly confirmed by the smooth change of black dots in Figure 9 for parameters β, γ. This assumption can be used as a regularization term. Unfortunately, it has not been formalized in theory and tested in practice yet. This is a topic for future research.

– Another way to assess the reliability of the parameter's value being reconstructed is *sensitivity analysis*. If the model is not sensitive to the parameter being studied, then its reconstructed value may differ significantly from that observed in reality. For the mean-field model studied in this paper, the sensitivity analysis with respect to the epidemiological parameters is shown in Figure 1. The analysis of the dependence of control behavior on the choice of the type of cost functional is given in the paper [19]. Its result is described in Section 2.2.2. Note that the sensitivity indices with respect to the entire set of reconstructed parameters are not estimated here, and this is quite difficult to do for two reasons. First, the upper limit of the change in parameters a_c and b_c is unknown. Note that for a significant change in the "output" of the model (let's say, the number of infected people on the time horizon), values a_c and b_c can change by several orders. Second, there is some dependence between the values β, γ, a_c, b_c, which is difficult to estimate. For example, if the parameters $\beta > \gamma > 0.5$, then the number of infected people first begins to grow rapidly, and then sharply falls. If the parameters a_c and b_c are also large, this will lead to sharp jumps in the right-hand side of the Hamilton-Jacobi-Bellman equation. This leads to the inoperability of most numerical methods for solving mean-field problems. Moreover, there is a dependence between the components of the functional, determined by the epidemiological situation described in reality. Thus, terminal costs (the third integral component of the functional) should not be much greater than current and running costs (the first and second components of functional), but can be significantly less. We can try to estimate the boundaries of change in parts of the functional, based on the analysis of real epidemiological data for several outbreaks of the epidemic. Thus, assuming the isolation function α is known (based on real data), the value of each component of the functional can be estimated, and identify the ranges of change in parameters a_c and b_c. But now this remains an issue for future research.

6. Conclusions

This paper is devoted to a possible formulation of the mean-field epidemiological problem. Section 2 proposes a formulation of the coefficient inverse problem, and provides its motivation in the form of sensitivity analysis results. Section 3 describes the algorithm for solving such a problem and describes the computational experiments. Section 4 is devoted to the modeling results, and Section 5 is devoted to their discussion and future plans.

Research has shown that the presented formulation can be used to solve real epidemiological problems, but can be refined in two directions: regularization of the model to overcome the non-uniqueness of the solution to the problem on real data, and evaluation of the growth rates of cost functional components of the problem.

Funding: This work was performed according to the government research assignment for the Sobolev Institute of Mathematics SB RAS, project FWNF-2024-0002.

Data Availability Statement: Statistical data on COVID-19 incidence in Novosibirsk used for modeling are available at the link: https://covid19-modeling.ru/data (accessed on 13 November 2024). Data aggregated by epidemiological groups (S, I, R), used directly for the software implementation of the model are available at the link: https://disk.yandex.ru/d/jDVK3xsPwnNvPA (accessed on 13 November 2024). Jupyter notebooks with software implementation of computational experiments are available at the link: https://drive.google.com/drive/folders/18wLiEpQCo9VrCuOXMy7Q4vloD6Kl6tDI?usp=sharing (accessed on 13 November 2024).

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

SIS	Susceptible–Infected–Susceptible (differential model)
MF	mean field
ODE	ordinary differential equations
SIR	Susceptible-Infected-Recovered (differential model)
SIR TGC MF	SIR Mean-Field Model with Total Control for All Epidemiological Groups
eFAST	Extended Fourier Amplitude Sensitivity Test
SIR MF	Susceptible-Infected-Recovered mean field (differential model)
N-M	Nelder—Mead (method)
SEIR-HCD	Susceptible-Exposed-Infected-Recovered-Hospitalized-Critical-Dead
SIRQ	Susceptible-Infected-Recovered-Quarantined (differential model)

References

1. Zanini, D.S.; Peixoto, E.M.; Andrade, J.M.D.; Tramonte, L. Practicing social isolation during a pandemic in Brazil: A description of psychosocial characteristics and traits of personality during covid-19 lockout. *Front. Sociol.* **2021**, *6*, 615232. [CrossRef] [PubMed]
2. Chen, J.; Vullikanti, A.; Santos, J.; Venkatramanan, S.; Hoops, S.; Mortveit, H.; Marathe, A. Epidemiological and economic impact of COVID-19 in the US. *J. Abbr.* **2021**, *11*, 20451. [CrossRef] [PubMed]
3. Gao, X.; Yu, J. Public governance mechanism in the prevention and control of the COVID-19: Information, decision- making and execution. *J. Chin. Gov.* **2020**, *2*, 178–197. [CrossRef]
4. Capano, G. Policy design and state capacity in the COVID-19 emergency in Italy: If you are not prepared for the (un) expected, you can be only what you already are. *Policy Soc.* **2020**, *3*, 326–344. [CrossRef] [PubMed]
5. La Torre, D.; Malik, T.; Marsiglio, S. Optimal control of prevention and treatment in a basic macroeco-nomic–epidemiological model. *Math. Soc. Sci.* **2020**, *108*, 100–108. [CrossRef]
6. Cho, S. Mean-Field Game Analysis of SIR Model with Social Distancing. Available online: https://arxiv.org/abs/2005.06758 (accessed on 2 October 2024).
7. Roy, A.; Singh, C.; Narahari, Y. Recent advances in modeling and control of epidemics using a mean field approach. *Sādhanā* **2023**, *48*, 207. [CrossRef]
8. Petrakova, V.; Krivorotko, O. Mean field game for modeling of COVID-19 spread. *J. Math. Anal. Appl.* **2022**, *514*, 126271. [CrossRef]
9. Petrakova, V.; Krivorotko, O. Mean Field Optimal Control Problem for Predicting the Spread of Viral Infections. In Proceedings of the 19th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS), Novosibirsk, Russia, 14–22 August 2023.
10. Leonov, A.; Nagornov, O.; Tyuflin, S. Inverse problem for coefficients of equations describing propagation of COVID-19 epidemic. *J. Phys. Conf. Ser.* **2021**, *2036*, 012028. [CrossRef]
11. Comunian, A.; Gaburro, R.; Giudici, M. Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic. *Phys. D Nonlinear Phenom.* **2020**, *413*, 132674. [CrossRef]
12. Georgiev, S.; Lubin, G. Coefficient identification in a SIS fractional-order modelling of economic losses in the propagation of COVID-19. *J. Comput. Sci.* **2023**, *69*, 102007. [CrossRef]
13. Ding, L.; Li, W.; Osher, S.; Yin, W. A mean field game inverse problem. *J. Sci. Comput.* **2022**, *92*, 7. [CrossRef]
14. Chow, Y.T.; Fung, S.W.; Liu, S.; Nurbekyan, L.; Osher, S. A numerical algorithm for inverse problem from partial boundary measurement arising from mean field game problem. *Inverse Probl.* **2022**, *39*, 014001. [CrossRef]
15. Liu, H.; Mou, C.; Zhang, S. Inverse problems for mean field games. *Inverse Probl.* **2023**, *39*, 085003. [CrossRef]
16. Kabanikhin, S.I.; Krivorotko, O.I.; Ermolenko, D.V.; Kashtanova, V.N.; Latyshenko, V.A. Inverse problems of immunology and epidemiology. *Eurasian J. Math. Comput. Appl* **2017**, *5*, 14–35.

17. Coronel, A.; Huancas, F.; Sepúlveda, M. Identification of space distributed coefficients in an indirectly transmitted diseases model. *Inverse Probl.* **2019**, *35*, 115001. [CrossRef]
18. Georgiev, S.; Lubin G. Numerical coefficient reconstruction of time-depending integer-and fractional-order SIR models for economic analysis of COVID-19. *Mathematics* **2022**, *10*, 4247. [CrossRef]
19. Petrakova, V.; Krivorotko, O. Comparison of Two Mean-Field Approaches to Modeling An Epidemic Spread. *JOTA* 2024, *submitted*. Available online: https://arxiv.org/abs/2411.02800 (accessed on 2 October 2024).
20. Lee, W.; Liu, S.; Tembine, H.; Li, W.; Osher, S. Controlling propagation of epidemics via mean-field control. *SIAM J. Appl. Math.* **2021**, *81*, 190–207. [CrossRef]
21. Saltelli, A.; Tarantola, S.; Chan, K.S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **1999**, *41*, 39–56. [CrossRef]
22. Nelder, J.A.; Mead, R. Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [CrossRef]
23. Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [CrossRef]
24. Kolundzija, B.M.; Olcan, D.I. Antenna Optimization using Combination of Random and Nelder-Mead Simplex Algorithms. *IEEE Antennas Propag. Soc.* **2003**, *1*, 185–188.
25. Chelouah, R.; Siarry, P. Genetic and Nelder-Mead Algorithms Hybridized for a More Accurate Global Optimization of Continuous Multiminima Functions. *Eur. J. Oper. Res.* **2003**, *148*, 335–348. [CrossRef]
26. Larson, J.; Menickelly, M.; Wild, S. Derivative-free optimization methods. *Acta Numer.* **2019**, *28*, 287–404. [CrossRef]
27. Krivorot'ko, O.I.; Kabanikhin, S.I.; Zyat'kov, N.Y.; Prikhod'ko, A.Y.; Prokhoshin, N.M.; Shishlenin, M.A. Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region. *Numer. Anal. Appl.* **2020**, *13*, 332–348. [CrossRef]
28. Pacheco, C.; de Lacerda, C. Function estimation and regularization in the SIRD model applied to the COVID-19 pandemics. *Inverse Probl. Sci. Eng.* **2021**, *29*, 1613–1628. [CrossRef]
29. Nuraini, N.; Syukriah, Y.; Indratno, S. Estimating parameter of influenza transmission using regularized least square. *AIP Conf. Proc.* **2014**, *1587*, 74–77.
30. Zheng, W. Total Variation Regularization for Compartmental Epidemic Models with Time-Varying Dynamics. Available online: https://arxiv.org/abs/2004.00412 (accessed on 2 October 2024).
31. Hamelin, F.; Iggidr, A.; Rapaport, A.; Sallet, G.; Souza, M. About the identifiability and observability of the SIR epidemic model with quarantine. *IFAC-Pap. OnLine* **2023**, *56*, 4025–4030. [CrossRef]