

Article

Not peer-reviewed version

Inverse Coefficient Problem for Epidemiological Mean Field Formulation

[Viktoriya Sergeevna Petrakova](#) *

Posted Date: 8 October 2024

doi: 10.20944/preprints202410.0592.v1

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



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Inverse Coefficient Problem for Epidemiological Mean Field Formulation

Viktoriya Petrakova^{1,2} ¹ Institute of computational Modelling SB RAS, 660036, Krasnoyarsk, Akademgorodok, 50/44; vika-svetlakova@yandex.rut² Sobolev Institute of Mathematics SB RAS, 630090, Novosibirsk, Koptyuga Ave., 4

Abstract: The paper proposes an approach to solving the inverse epidemiological problem, written in terms of the “mean field” theory. Finding the coefficients of the epidemiological SIR mean field model is reduced to solving an optimization problem, for the solution of which only zero-order methods can be used. The Nelder-Mead method was chosen as such a method. The algorithm for 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

1. Introduction

The coronavirus epidemic, which was spreading most in 2020-2022, has shown the need to develop measures to contain the spread of the disease due to reduce mortality among the population and the burden on the healthcare system. At the same time, the measures introduced can vary quite significantly: from the development of a hand-washing culture to complete long-term isolation, be perceived by society differently [1–4] and have different effects on the economy and other areas of public life. Methods of mathematical modeling help to 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 us to assess the sensitivity of the modeled system to the slightest changes in the choice of the population behavior strategy and the external planner. Thus, using the basic epidemiological model of the SIS, in [5] an optimal management model is proposed in which the planner uses tax revenues to direct funds either to preventive measures for the population or to 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 are likely to maximize personal utility, as opposed to a socially optimal strategy that maximizes the total utility for the entire population. The authors assume that the reduction of contacts with infected people should continue long after the epidemic has subsided.

One approach to modeling the dynamics of epidemic spread is optimal control models written in terms of “mean field” theory. Here it is assumed that the population is large enough and the influence of the strategy adopted by an individual is negligible and does not affect the overall behavior of the population. This assumption allows one to pass to 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, the studies [8,9] demonstrated the advantage of mean-field predictive models 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 the mathematical model used for forecasting. The parameters

being reconstructed are usually physical, but difficult to determine, since they depend on a large number of factors. For example, the contagiousness parameter (the rate of virus spread), used in the overwhelming majority of epidemiological models, depends on the infectiousness of the virus, the frequency of contacts in the population, the incubation period of the virus, and compliance with anti-virus restrictions.

There are only a few papers on solving inverse mean field problems [10–12]. These papers consider classical formulations in the form of a system of two partial differential equations, used mainly to solve economic problems. The problems of finding solutions of inverse problems for mean-field models arise due to the strict restrictions on the choice of the cost functional describing the behavior of the system to ensure the existence and uniqueness of the solution of direct problem. Such conditions for epidemiological mean field formulation are discussed in more detail in Section 2.2. To overcome them, in [10], for example, 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 the transfer optimization problem, for which the inverse problem has been already solved. The papers [11,12] consider the reconstruction of the Hamiltonian of the classical mean-field model 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 functional 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.

This work is aimed at solving practical problems arising from epidemiological modeling. Here, to remain within the framework of the uniqueness of the existence of a solution to the direct problem, an inverse coefficient problem is formulated. The inverse problem is formulated as a multicriterial optimization problem with constraints, which can only be solved by zero-order methods. Sections “Materials and Methods”, “Results” and “Discussion” present 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 it is used the formulation which was first introduced in [13] and called the *SIR Mean Field model with Total Control for all epidemiological Groups (SIR TGC MF model)*. This formulation differs from the one proposed, e.g. in [14], by the generality of the chosen strategy for the entire population. Their comparison is given in the work [13]. The formulation of a direct problem is the following:

find the minima of the cost functional

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

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

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

with initial

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

and Neumann boundary conditions

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

Here the stochastic processes within the population are described using non-negative parameters σ_i , $i \in \{S, I, R\}$; $m_i(t, x) : [0, T] \times \Omega \rightarrow \mathbb{R}$ 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 the imposed restriction measures and $x = 1$ is the opposite. Functions $\alpha_i(t, x) : [0, T] \times [0, 1] \rightarrow \mathbb{R}$, $i \in \{S, I, R\}$ denote the compliance strategy of the representative agent of each group in the population.

The using of Lagrange multiplayer method for optimization problem (1)–(4) leads to the conjugate system of partial differential equations

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

with the conditions on time horizon

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

and the boundary conditions

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

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

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

Note, that systems (4)–(6), (8) are valid when the following conditions are performed:

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

Thus, the set of systems of equations (2)–(9) describes the optimum of a dynamic system, the evolution of which is determined by the cost function (1).

2.2. Inverse Problem Formulation

The main difficulties of the solution of the mean field problem used for real-life application is based on the strict restrictions on the domain of the existence and uniqueness of the solution. The most general condition for meeting these needs is the convexity of the functional with respect to the control variables. In paper [13] these restrictions are formulated for the SIR mean field problem with individual control for each epidemiological group (see Property 4 in [13]). Below reformulation of these restrictions is proposed for the SIR TGC MF model (2)–(9) used in this study.

Property 1. [Reformulation of Property 4 [13]] 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:

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

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

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

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

$$f_S = -\beta m_S m_I; \quad f_I = \beta m_S m_I - \gamma m_I; \quad f_R = \gamma m_I. \quad (12)$$

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

Thus, we are quite strongly limited by the area in which a unique solution to the direct problem exists. At the same time, the study of the difference in the sensitivity of the modeling results to the parameters of the model for formulation of the mean field problem and in the form of a system of ordinary differential equations of the SIR type for describing the epidemiological process shows that the main contribution to the dispersion of the output is made by the change in the epidemiological parameters of the model (parameters of contagiousness β and recovery γ). In Figure 1 the result of the sensitivity analysis is proposed. This result was first presented in the author's previous work [13], but will be repeated here because it is important. Here the modeled number of infected people (I group) was used as output. For the input parameters, the following vector was used:

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

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:

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

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; $a_i = \exp \left(-(1 - x_i^c)^2 / 2(\sigma_i^c)^2 \right) (1 - x_i^c) / (2(\sigma_i^c)^3 \sqrt{2\pi})$ and $b_i = \exp \left(-(x_i^c)^2 / 2(\sigma_i^c)^2 \right) (x_i^c) / (2(\sigma_i^c)^3 \sqrt{2\pi})$ are the expressions which ensure boundary conditions (4) for $m(0, x)$. Physically, this means that we define the initial distributions as Gaussian (though with some small correction to comply with the boundary conditions). The 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)[15] was used. Method eFAST allows to divide the total variance of model output into components, corresponding to the model's input parameters. The 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 amplitude of sensitivity indices is 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 $\sigma_S, \sigma_I, \sigma_R$; 'E_S', 'E_I', 'E_R' for x_S^c, x_I^c, x_R^c and 'disp_S', 'disp_I', 'disp_R' for $\sigma_S^c, \sigma_I^c, \sigma_R^c$.

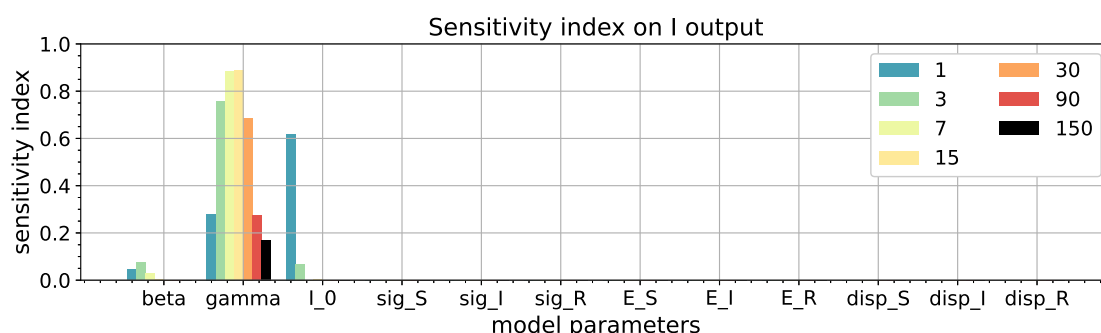


Figure 1. Sensitivity indices for the SIR MF model (2)–(9) with a set of input parameters (13) for different simulation times: $T = 1, 3, 7, 15, 30, 90, 150$ days

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 done in [10–12]. The idea here is to put

$$G_i(m_{SIR}, \alpha) = a_c \frac{\alpha^2}{2} m_i; \quad g(t, x, m_{SIR}) = m_I^2(t, x); \quad \Phi(m_{SIR}(T, x)) = b_c m_I^2(T, x) \quad (15)$$

in direct problem functional (1). Here the a_c, b_c are unknown parameters. This choice is due to several reasons. First, assuming that the functions G_i, g, Φ are quadratic in control and the distribution densities, 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 the epidemic - the population acts in such a way (or is subject to such actions on the part of the government) to reduce the number of infected (choice of function g).

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

$$\min_{\vec{\theta}} L = \sum_q \left(\int_0^1 m_I(q, x) dx - I_q \right)^2 + \sum_p \left(\int_0^1 m_R(p, x) dx - R_p \right)^2. \quad (16)$$

Here I_q and R_p are known measures of infected and recovered parts of the population in some moments of time, 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 Optimiaztion Method

Here, the Nelder-Mead method is used to solve the optimization problem. In 1965 [16], 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 the values of the objective function at $(n + 1)$ vertices of the simplex and move the simplex towards the optimal point using an iterative procedure, where n is the 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 the rapid initial decrease in the function values [17,18].

Recently, this well-known algorithm has been used in combination with global search algorithms such as random search [18] and genetic algorithm [19].

The implementation of the algorithm involves calculating the values of the objective function (defined here as (16)) at the vertices of 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 the termination of the iterative process is the smallness of the sides or area of the simplex obtained at the next iteration.

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

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

Algorithm

1. Choose the initial value of vector $\vec{\theta}$ being restored and the 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 equal to zero ($k = 0$);
3. Solve the direct SIR MF problem (2)–(9),(15) with known parameters $\vec{\theta}^k$;
4. Put $k = k + 1$;
5. Compute the $2 \cdot |\vec{\theta}^k| + 1$ value of target function (16) on k th iteration at a small deviation from each vertex of the simplex under study. Do the iteration of Nelder-Mead algorithm and find new vector $\vec{\theta}^k$;
6. Check the stopping criterion of the 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 done using the finite difference approximation proposed in [13].

3.3. The description of computational experiments

The 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, we will compare the recovery possibility of only the β and γ parameters. As a set of synthetic data, we will consider the solution of a simple differential SIR model of the form

$$\begin{cases} dm_S/dt = -\beta m_S m_I, \\ dm_I/dt = \beta m_S m_I - \gamma m_I, \\ dm_R/dt = \gamma m_I \end{cases} \quad (17)$$

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:

$$N = 1000, S_0 = 800/N, I_0 = 200/N, R_0 = 0, \beta = 0.7, \gamma = 0.3. \quad (18)$$

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.

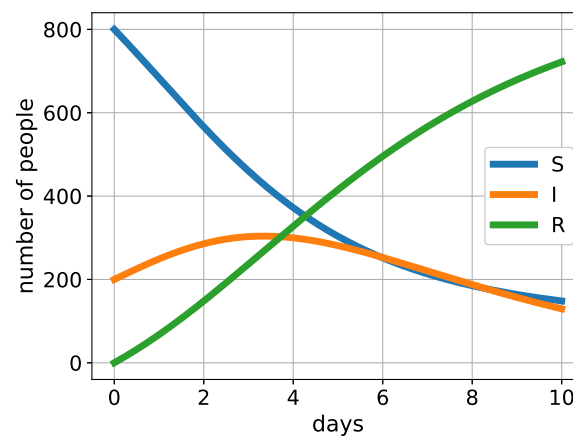


Figure 2. Visual representation of the epidemic dynamics described by the SIR model (17) with parameters (18)

For the inverse SIR MF problem, the values of the β and γ parameters were reconstructed only based on 10 daily measurements of the infected part of the population, that is, in the target function (16) we have the measurements $\vec{I} = (I_0, \dots, I_9)$ and don't have any of R_p .

Experiment №2

For the second experiment, we use the same data as in Experiment 1, but introduce a single noise into the known solutions $(I_0, \dots, I_9)$ that does not exceed 1 or 5%. The noise value is generated the same for each day measurement and displays the systematic error of the observer.

Experiments №3,4

Experiments 3 and 4 repeat Experiments 1 and 2, differing only in that all 4 parameters are 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 spread of Covid-19 in Novosibirsk was chosen for the 150 days from 2021-04-13. The data collected in this city and some others can be found at the link <https://covid19-modeling.ru/data>. As in the previous experiment, the full set of parameters $\vec{\theta} = (\beta, \gamma, a_c, b_c)$ was restored. For the target function (16) computation, the daily measurements of the number of infected people $\vec{I} = (I_0, \dots, I_{149})$ were used. The modeling was carried out for different time periods, which was called 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-days period, gluing of the corresponding number of periods for w days was done. The comparison of errors of modeling is provided for several values of w .

Experiment №6

And the last experiment repeats Experiment 5, but here for the parameters reconstruction the daily measurements of the infected and recovered parts of the population were used. That is, in the target function (16) we have the measurements $\vec{I} = (I_0, \dots, I_{149})$ and $\vec{R} = (R_0, \dots, R_{149})$.

4. Results

This section presents the results of computational experiments described above.

Experiment №1

For the simple 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 one 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 3.2, were required.

Experiment №2

For the second experiment where the noise in measurements was put on the thousand of similar optimization problems with different generated value of noise were solved. The result of this is presented in Figures 3 and 4 for noises the values of which don't exceed 1% and 5%.

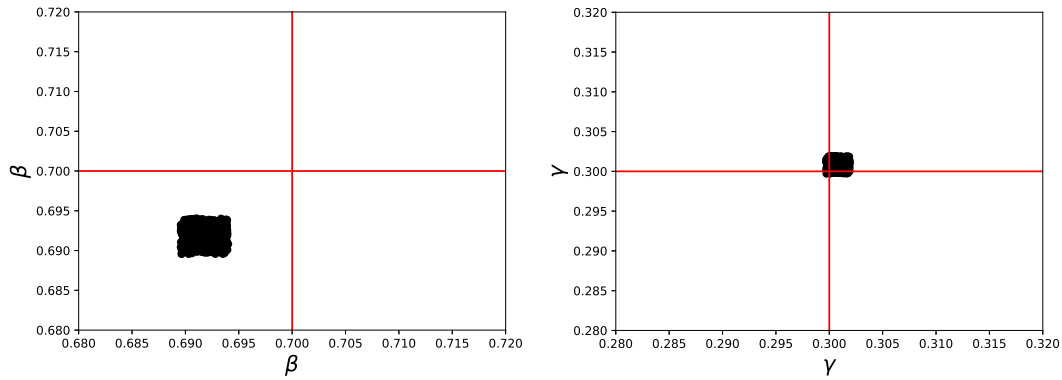


Figure 3. The values of recovered parameters β (left) and γ (right) for the 1000 carried out in similar Experiments №2 with uniform noise in data under 1%

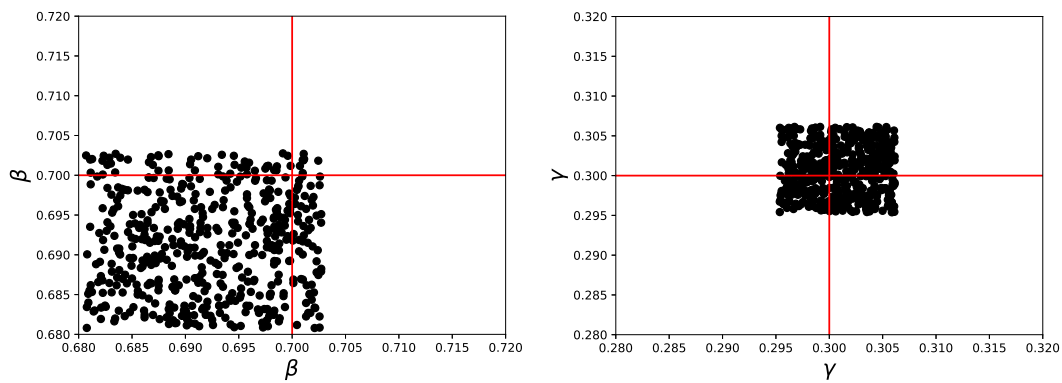


Figure 4. The values of recovered parameters β (left) and γ (right) for the 1000 carried out in similar Experiments №2 with uniform noise in data under 5%

Here and after, metrics *rmse* and *mae* 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 true observation; y_k obtained value from k th experiment realization; M is number of realization. Thus, *rmse* represents the root of the mean square error and *rae* is relative absolute error. So, for the Experiment №2, when introducing noise doesn't 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 doesn't 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%.

Experiment №3

The same pattern of computational results can be found for Experiments 3,4. The recovered 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 SIR epidemiological problem (17) 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

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

When introducing noise doesn't 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 doesn't 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.

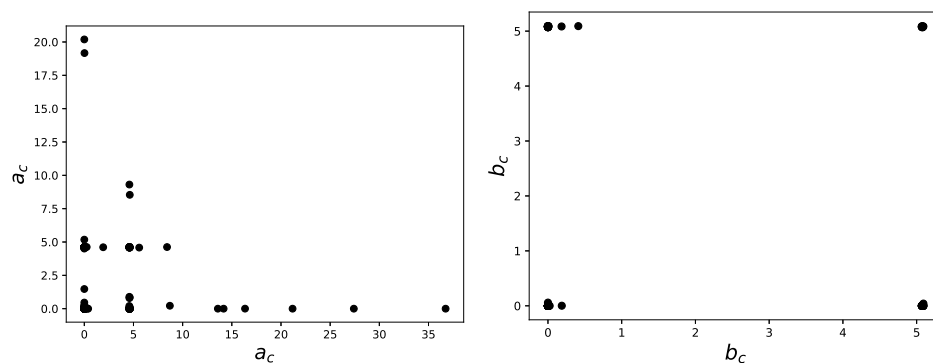


Figure 5. The values of recovered parameters a_c (left) and b_c (right) for the 300 carried out in similar Experiments №4 with uniform noise in data under 1%

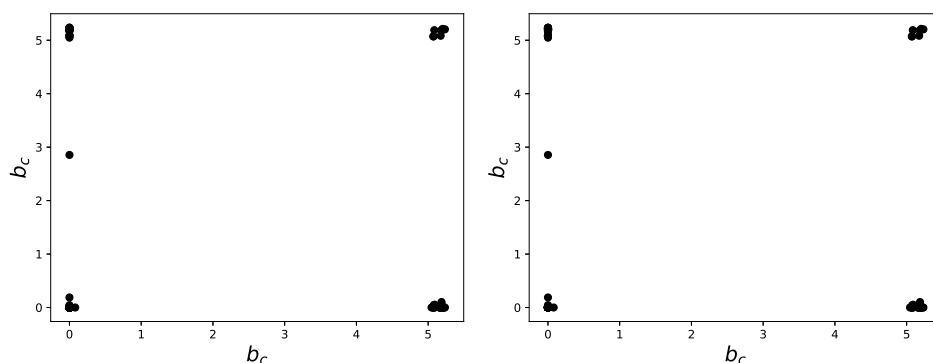


Figure 6. The values of recovered parameters a_c (left) and b_c (right) for the 300 carried out in similar Experiments №4 with uniform noise in data under 5%

Experiments №5,6

The Figures 7 and 8 show the result of computational experiments carried out on real data. In the case of Figure 7 the parameters were recovered using only measurements of the infected part (I group)

of the population and after recovering they were used for direct modeling to compare the quality of approximation. For the result depicted in Figure 8 the same was done, but fore the recovering the measurements of infected and recovered parts (I and R groups) of the population were used. The figures depict the result for modeling window length (w) equaled 10 days.

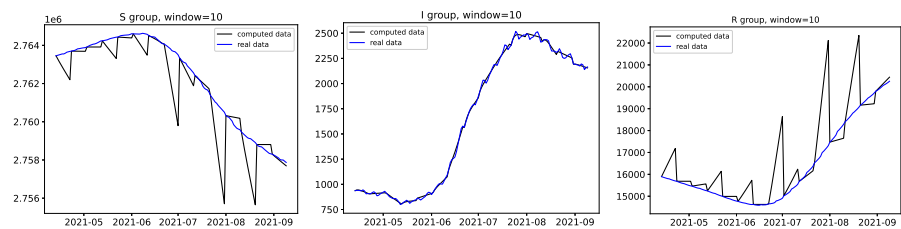


Figure 7. The result of direct modeling using SIR MF model in comparison with real data when the parameters of the model were reconstructed using the measurements of I group

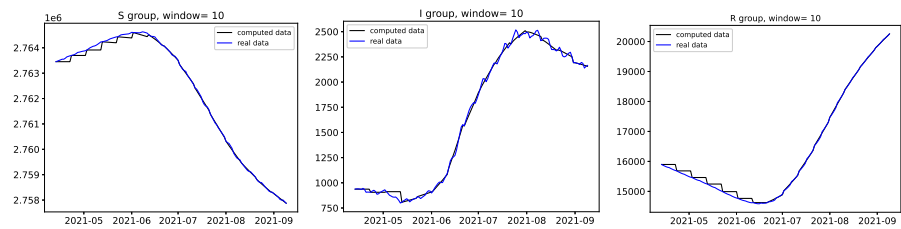


Figure 8. The result of direct modeling using SIR MF model in comparison with real data when the parameters of the model were reconstructed using the measurements of I and R groups

The Table 1 includes the values of $rmse$ and rae for several values of windows length w .

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

window length (w) (in days)	mae (in people)			rae (in percent)		
	S group	I group	R group	S group	I group	R group
Recovery by infected group (I) only						
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
Recovery by infected and removed groups (I, R)						
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

And the last picture 9 shows the difference in determination of parameters β, γ, a_c, b_c when they were recovered using only the measurements of the infected part of population or both infected and recovered.

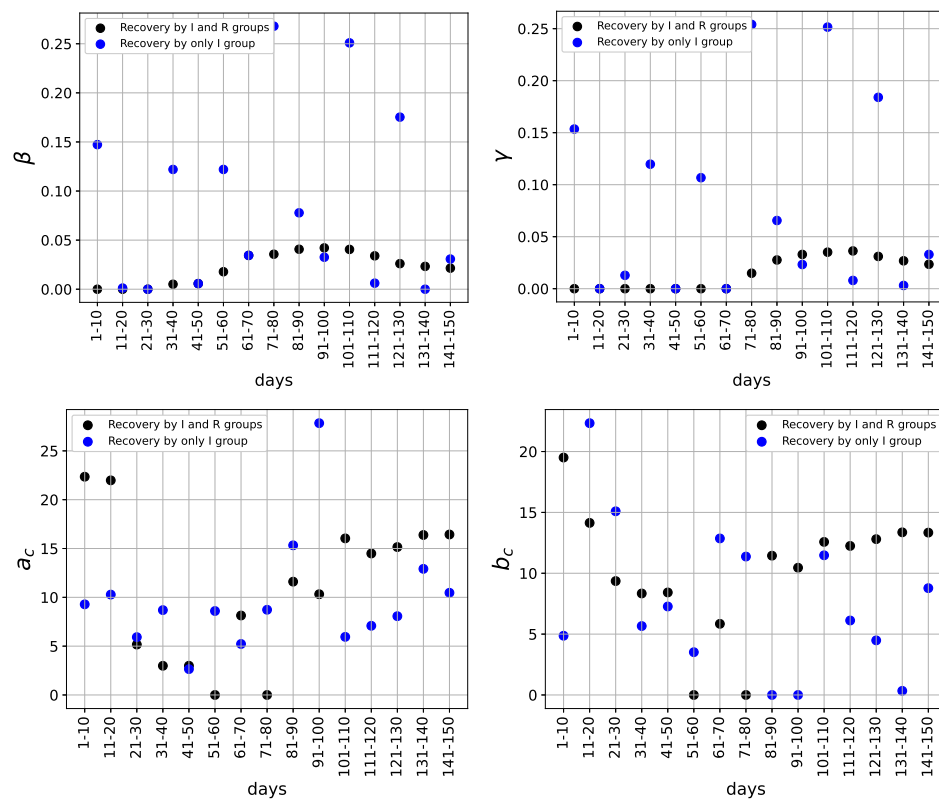


Figure 9. The value of the reconstructed parameters in each 10-day simulation window relative to the real data

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 solution.

As Section 2 of the work shows, the main difficulty in solving the inverse mean field 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 Hamiltonian of the system strongly ill-posed. However, the inverse problem for the mean mean field model describing the epidemiological process can be reduced to a coefficient problem. As the sensitivity analysis of the model, methodically carried out in [13], 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 growth rate of agent costs over time, that is, the choice of coefficients in the cost functional for a given behavior of the system.

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, the 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 5,6 more detailed. In most cases, the value of b_c parameter was determined to be close to zero. This is an expected value because when b_c is equaled zero and a_c is any, the solution of SIR MF system coincides with the solution of SIR differential model. This also approves the variation in determination of a_c parameter. But in several cases, the value of 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. This indirectly confirms the hypothesis of a lower sensitivity of the model to the parameters describing the population, since when introducing

an error into the "exact data", the greatest contribution to the system is made by the definition of the epidemiological parameters β and γ .

Note that here only the parameters before the costs of implementing the strategy (a_c) and the terminal conditions (b_c) were restored, 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 (the cost of implementing the strategy, current costs and terminal costs), which determine the area of existence of the solution to the direct problem. At present, these relations are unknown, but there is a hypothesis for their assessment based on the description of phase behavior of more complicated differential SIR models when describing a single outbreak. This is the future direction of studies of our research group.

Another problem related to the recovery of the mean field 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 asymptomatic 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 recovery was made only by measuring the number of infected people, gave an encouraging result. However, on real data (Experiments 5-6), the use of such an approach leads to non-uniqueness of the solution to the problem. An attempt to reduce the period for which the recovery is made also did not give results (see Table 1). Thus, to recover the model parameters on real data, there is a need for additional use of measurements in other epidemiological groups, at least partially, for example, the number of recovered people). This problem can be solved by using regularization, based on the work [20], where the regularization component is based on an assumption about the known values of the parameters. For example, the parameters being recovered over a certain period of time are not very different from those recovered over a previous period of time. This hypothesis also needs to be tested in future studies.

6. Conclusion

This work was devoted to a possible formulation of the mean field epidemiological problem. Section 2 proposes a formulation of the inverse problem coefficient, and provides its justification 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.

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 ratios of the components of the cost functional of the problem.

Funding: The work was performed according to the Government research assignment for 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>. 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>.

Abbreviations

The following abbreviations are used in this manuscript:

SIS	Susceptible - Infected - Susceptible (differential model)
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)

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. *Frontiers in Sociology* **2021**, *6*, 615232.
2. Chen, J.; et al. pidemiological and economic impact of COVID-19 in the US. *Journal Abbreviation* **2021**, *11*, 20451.
3. Gao, X.; Yu, J. Public governance mechanism in the prevention and control of the COVID-19: information, decision- making and execution. *Journal of Chinese Governance* **2020**, *2*, 178–197.
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 and Society* **2020**, *3*, 326–344.
5. La Torre, D.; Malik T.; Marsiglio S. Optimal control of prevention and treatment in a basic macroeconomic–epidemiological model. *Mathematical Social Sciences* **2020**, *108*, 100–108.
6. Cho, S. Mean-field game analysis of SIR model with social distancing. Available online: <https://arxiv.org/abs/2005.06758>.
7. Roy, A.; Singh, Ch.; Narahari, Y. Recent advances in modeling and control of epidemics using a mean field approach **2023**, *48*, 207.
8. Petrakova, V.; Krivorotko, O. Mean field game for modeling of COVID-19 spread. *Journal of Mathematical Analysis and Applications* **2022**, *514*, 126271.
9. Petrakova, V.; Krivorotko, O. Mean Field Optimal Control Problem for Predicting the Spread of Viral Infections. In Proceedings of 19th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS), Novosibirsk, Russia, 14-22 August 2023.
10. Ding, L.; Li, W.; Osher, S.; Yin, W. A mean field game inverse problem. *Journal of Scientific Computing* **2022**, *92*, 7.
11. 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 Problems* **2022**, *39*, 014001.
12. Liu, H.; Mou, C.; Zhang, S. Inverse problems for mean field games. *Inverse Problems* **2023**, *39*, 085003.
13. Petrakova, V.; Krivorotko, O. Comparison of Two Mean-Field Approaches to Modeling An Epidemic Spread. *JOTA* **2024**, *submitted*
14. Lee, W.; Liu, S.; Tembine, H.; Li, W.; Osher, S. Controlling propagation of epidemics via mean-field control. *SIAM Journal on Applied Mathematics* **2021**, *81*, 190-207.
15. Saltelli, A.; Tarantola, S.; Chan, K.S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **1999**, *41*, 39–56.
16. Nelder, J.A.; Mead, R. Simplex Method for Function Minimization. *Computer Journal* **1965**, *7*, 308–313.
17. Lagarias, J.C.; Reeds, J.A.; Wright, M.H; Wright, P.E. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization* **1998**, *9*, 112–147.
18. Kolundzija, B.M.; Olcan, D.I. Antenna Optimization using Combination of Random and Nelder-Mead Simplex Algorithms. *IEEE Antennas and Propagation Society* **2003**, *1*, 185–188.
19. Chelouah, R.; Siarry, P. Genetic and Nelder-Mead Algorithms Hybridized for a More Accurate Global Optimization of Continuous Multim minima Functions. *European Journal of Operational Research* **2003**, *148*, 335–348.
20. Shaydurov, V.; Petrakova, V. Three-parameter Regularization Algorithm for Pseudo-solution of Non-compatible Systems. *Lobachevskii Journal of Mathematics* **2024**, *1*, 328–335.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.