

Исследование адаптивно-компаратментной модели распространения КОВИД-19 в некоторых регионах РФ методами оптимизации

**Левашкин С.П.^{*1,2}, Агапов С.Н.¹, Захарова О.И.¹, Иванов К.Н.¹,
Кузьмина Е.С.¹, Соколовский В.А.¹, Моисова А.С.¹, Воробьев А.В.¹,
Апешин Д.Н.¹**

¹*Поволжский государственный университет телекоммуникаций и информатики,
Самара, Россия*

²*Самарский государственный технический университет, Самара, Россия*

Аннотация. Предлагается системный подход к исследованию новой многопараметрической модели распространения пандемии COVID-19, который имеет конечной целью оптимизацию ее управляющих параметров. Подход состоит из двух основных частей: 1) адаптивно-компаратментной модели распространения эпидемии, которая является обобщением классической модели SEIR и 2) модуля настройки параметров этой модели по эпидемиологическим данным методами интеллектуальной оптимизации. Данные для тестирования предлагаемого подхода на примере распространения пандемии в некоторых регионах РФ, брались на ежедневной основе из открытых источников в течении первых 130 дней эпидемии, начиная с марта 2020 г. Для этого была разработана и внедрена на локальном сервере так называемая «ферма данных» (автоматизированная система сбора, хранения и предварительной обработки данных из гетерогенных источников), которая в сочетании с методами оптимизации позволяет наиболее точно настраивать параметры нашей модели, превращая ее таким образом в интеллектуальную систему для поддержки принятия управленческих решений. Среди всех используемых параметров модели наиболее важными являются: скорость передачи инфекции, действия властей и реакция населения.

Ключевые слова: *многопараметрическое моделирование, COVID-19, модель распространения эпидемии, многопараметрическая оптимизация, функция потерь.*

ВВЕДЕНИЕ

Сегодня весь мир столкнулся с пандемией COVID-19, которая, по всей вероятности, будет иметь долгосрочные последствия для экономики государств и общества в целом. Данную ситуацию однозначно можно рассматривать как серьезную биогенную угрозу, которая наносит ощутимый ущерб социально-экономическим системам, государственным образованиям и обществу. При этом особенно актуальными научными задачами являются коррекция существующих и разработка новых математических моделей распространения эпидемий и методов цифрового мониторинга не только темпов распространения вируса, но и социально-экономической среды. И те, и другие должны отражать современные реалии и как можно точнее представлять, как саму биогенную угрозу, так и ее последствия. Поэтому своевременный мониторинг экономических, психологических и иных социальных процессов, протекающих внутри такой среды,

* ai_lab@psuti.ru

позволит принимать более оптимальные управленческие решения и минимизировать последствия пандемии за счет более точной настройки параметров прогнозной модели.

Начало применению математических методов при изучении эпидемий было положено Даниилом Бернулли в середине XVIII века [1]. Он впервые применил простейший математический аппарат для оценки эффективности профилактических прививок против натуральной оспы. Вслед за этим последовал значительный перерыв, который завершился работами английского ученого Уильяма Фара [2]. Он изучал и моделировал статистические показатели смертности населения Англии (Уэльса) от эпидемии натуральной оспы в 1837–1839 гг. Этот ученый впервые получил математические модели показателей «движения» эпидемии натуральной оспы в виде статистических закономерностей, что позволило ему в итоге составить прогностическую модель этой эпидемии. В начале XX века статистический подход У.Фарра в изучении эпидемий был переосмыслен и затем развит в работах Джона Браунли, в которых он анализировал статистические закономерности «движения» эпидемиологических показателей с помощью малоизвестных методов математической статистики. Однако этот статистический подход в изучении закономерностей развития эпидемий существенно отличается от аналитического подхода, который был предложен в конце XIX века сначала в России [3], а затем в Англии [4]. Благодаря этим исследователям, в начале XX века были сформулированы основы современной теории математического моделирования эпидемий, разработаны первые прогностические модели эпидемий (корь, ветрянка, малярия и др.), изучены их основные свойства, получены аналитические формулы для прогнозирования эпидемий.

Качественно новый этап исследований начинается с появлением электронно-вычислительных машин, когда число научных работ и публикаций по математическому и компьютерному моделированию эпидемий стало быстро увеличиваться. В работах того времени стали появляться все более сложные математические модели, в которых существенную роль играли случайные факторы эпидемического процесса, поэтому большинство моделей этого периода имели стохастический (вероятностный) характер, а рабочим аппаратом была теория вероятностей и случайных процессов. Этот этап характеризуется «натиском» на эпидемиологию «чистых» математиков, которым удалось создать множество абстрактных моделей, но с весьма ограниченным эпидемиологическим содержанием [5].

Первые исследования, которые наметили пути преодоления указанного «перекоса», были выполнены в 1960-е годы в СССР акад. О.В. Барояном и проф. Л.А. Рвачевым [6]. Ими была разработана новая методология математического моделирования эпидемий – эпиддинамика. Данная методология основана на методе научной аналогии в отображении эпидемического процесса (процесс «переноса» возбудителя инфекции от больных к здоровым) с процессом «переноса» материи (энергии, импульса и др.) в уравнениях математической физики [7]. Действительно, в ходе развития эпидемии среди населения территории, пораженной инфекционным заболеванием, формируется сложный самоподдерживающийся процесс «переноса» популяции возбудителя на сообщество восприимчивых людей. Эпидемиологическое содержание данного процесса связано с адекватным его отображением, как в календарном времени t , так и во «внутреннем времени τ , которое фиксирует развитие инфекционного заболевания у множества лиц, пораженных инфекцией. Система уравнений, которая описывает развитие эпидемического процесса, представляет собой систему нелинейных интегро-дифференциальных уравнений в частных производных с соответствующими начальными и граничными условиями, весьма «схожими» с уравнениями гидродинамики.

Современный этап в развитии моделирования эпидемий связан прежде всего с быстрым прогрессом в области компьютерных технологий (разработаны мощные

компьютеры с новейшими инструментами программирования и моделирования) [8]. На сегодняшний день известны и общепризнаны несколько основных математических моделей распространения эпидемий, которые берут свое начало от модели Барояна – Рвачева, в которой уже рассматривались основные стадии-состояния инфекционного процесса типа SEIR, где: S – восприимчивые, E – в инкубационном периоде (латентные), I – инфицированные, R – выздоровевшие. Рассмотрим эти модели более подробно. Заметим, что для каждой из них характерно последовательное добавление новых стадий-состояний инфекционного процесса для более точного описания реальной динамики распространения эпидемии. Один из пионеров в области математического моделирования эпидемий, Росс, разработал математические модели, описывающие перемещение взрослых комаров и модель распространения малярии [4]. Макдональд, основываясь на идеях Росса, более подробно изучил энтомологические и демографические аспекты распространения малярии [9]. Хоппэнстэд опубликовал одну из первых моделей эпидемии с возрастной структурой и предложил несколько различных видов дифференциальных систем, описывающих динамику передачи болезни [10]. Бейли, см. выше, расширил вторую модель Росса и представил общую модель распространения болезней в виде SIR и SI моделей. Ньютон и Рейтер опубликовали SEIR–SEI модели для лихорадки денге. В [11] подробно исследованы SEI–SEIR модели, доказаны существование и единственность неотрицательного решения, и существование стационарных точек. В [12] с помощью разделения человеческой популяции на конечное число возрастных групп автор нашел базовый репродуктивный номер болезни для задачи с SEIR системой, зависящей от возраста. SEI–SEIR модели с возрастной структурой построены на основе популяционной модели Маккендрика-фон Ферстера. В зависимости от свойств болезни, в дифференциальных системах, которые описывают динамику распространения инфекции, используются различные наборы популяционных групп. В SEI моделях общее количество индивидов в популяции N делится на три класса: S – восприимчивых к болезни, E – зараженных, которые еще не являются заразными, и I – зараженных индивидов, способных к передаче инфекции. В некоторых SEIR моделях добавляется в рассмотрение класс иммунных к болезни индивидов R. Отличие SI–SIR моделей от SEI–SEIR заключается лишь в том, что в первых убирается из рассмотрения класс людей, которые заразились, но еще не являются заразными. Это удобно в случае, когда инкубационным периодом патогена можно пренебречь. До сих пор все перечисленные модели и их многочисленные модификации являются актуальными [13–27] для входящих в них процессов, начальными условиями и большим набором параметров, зачастую определяемых весьма субъективно [30]. Таким образом, изначально постановка задачи является нелинейной и многопараметрической, и потому разработать для нее качественные методы, обладающие большей эффективностью, чем широко известные методы градиентного типа [29], а также методы, основанные, скажем, на принципе максимума Понтрягина, представляется довольно затруднительным.

Кроме того, эпидемиологические прогнозы выполняются для различных сроков и в зависимости от них служат разным целям. Так, краткосрочный прогноз на несколько недель вперед применяется в оперативном управлении и при выявлении эпидемических вспышек заболеваемости. Наиболее полезным можно считать среднесрочный прогноз сроком от двух месяцев до полугода, используемый в тактическом управлении. Конечно, он менее точен, нежели краткосрочный, но оставляет достаточно времени для подготовки к возможным чрезвычайным ситуациям и проведения превентивных мероприятий [28].

Чтобы преодолеть эти и другие трудности, в нашем исследовании, взяв за основу модель SEIR, мы ее существенно модифицируем и, добавляя ряд параметров и процессов, строим адаптивно-компарментную модель, которая характеризуется возможностью исследования проблемы при помощи современных подходов

оптимизации и интеллектуального анализа данных. Причем при выборе модели мы исходили из соображений оптимального ограничения количества регулирующих параметров, то есть выбора наиболее простой модели из множества возможных для максимально точного описания имеющихся данных и построения среднесрочного прогноза.

Дальнейшее изложение организовано следующим образом. В разделе «Материалы и методы» описаны модель, данные и основные методы, использованные для получения результатов. Раздел «Результаты» содержит результаты численных экспериментов для двух регионов РФ (Москва, Самара). Последние два раздела называются соответственно «Дискуссия» и «Заключение», где обсуждаются перспективы дальнейшего развития модели и подводятся итоги данному этапу исследования.

МАТЕРИАЛЫ И МЕТОДЫ

Модель

В нашей работе за основу взяты ряд общепринятых в эпидемиологии компартментных моделей (compartmental models) SIR, SEIR, SEIRS, SEIRD, описанных в [13–19] и [21–27].

При модификации моделей мы использовали методологию, аналогичную описанной в работах [20, 30, 31].

При моделировании мы придерживались следующих общепринятых в эпидемиологии допущений:

- Перинатальный материнский иммунитет игнорируется и новорожденные включаются в класс восприимчивых к болезни S.
- Предполагается, что шансы на выздоровление или смерть не изменяются в течение всего периода инфекции.
- Предполагается, что заразность не меняется в течение всего периода инфекции.
- Инфекционные периоды (и время пребывания во всех категориях) распределяются экспоненциально.

В нашей работе за основу взята модель SEIRD, то есть модель SEIR, расширенная для учета смертельных случаев, параметры нелинейной скорости передачи инфекции, действия властей и реакцию населения. При выборе и модификации модели мы исходили из соображений ограничения количества регулирующих параметров, то есть подбора наиболее простой модели из возможных для описания данных. Наша модель описывается следующей системой уравнений:

$$\begin{aligned} \frac{dS}{dt} &= -\beta(t) \frac{SI}{N}, \\ \frac{dE}{dt} &= \beta(t) \frac{SI}{N} - \sigma E, \\ \frac{dI}{dt} &= \sigma E - (1-\alpha)\gamma I - \alpha R_0 I, \\ \frac{dR}{dt} &= (1-\alpha)\gamma I, \\ \frac{dD}{dt} &= \alpha R_0 I, \end{aligned} \tag{1}$$

где $\beta(t) = \beta_0 * g(t) * s(t)$ – функция изменения базового коэффициента скорости передачи инфекции β_0 (transmission rate);

t – время;

N – общий размер популяции;

$$g(t) = \begin{cases} g_1, & \text{если } t < tg \\ g_2, & \text{если } t > tg \end{cases} - \text{ функция управления со стороны государства (карантины и}$$

прочие ограничения);

значения g_1 , g_2 и tg приведены в Таблице 1;

$$s(t) = \frac{1}{1 + e^{c_1(si(t)-c_2)}} - \text{ функция преобразования индекса самоизоляции Яндекса } si(t),$$

берется из данных; значения c_1 и c_2 приведены в Таблице 1;

R_0 – 1/время нахождения в критическом состоянии перед смертью;

σ – коэффициент передачи инфекции от зараженных E к инфицированным I;

γ — скорость выздоровления;

α – коэффициент смертности (от количества инфицированных).

Данные

При проектировании системы хранения и сбора данных ключевыми были следующие критерии:

1. Наличие нескольких источников данных, состав которых может корректироваться (добавляться и удаляться).

2. Вероятность возникновения новых факторов, которые могут оказать существенное влияние на модель.

3. Приоритетность официальных источников информации, несмотря на возможную неточность в данных.

4. Возможность как дублирования, так и расхождения в показателях при получении данных из разных источников.

5. Возможная избыточность данных для использования при построении нескольких видов моделей.

6. Данные могут быть представлены как числовые, временные ряды, тексты.

Функционально архитектура разбита на три группы операций:

1. Получение данных.

2. Хранение данных.

3. Выдача данных.

При сборе данных было проанализировано более 200 различных ресурсов. Данные для обработки в модели были взяты с официального российского сайта стопкоронавирус.рф (<https://стопкоронавирус.рф/>); данные по самоизоляции были взяты с сайта Яндекс Datalens (<https://cloud.yandex.ru/services/datalens>), как наиболее актуальные и обновляемые на ежедневной основе (все другие собираемые нами на ежедневной основе данные, будут использованы для уточнения и модификации модели).

Выбор основывался на определенных параметрах, благодаря которым результат будет более точным. Основными критериями выбора являлись данные о случаях заражения, выздоровления и летальных исходах, о мерах, предпринимаемых руководством страны, реакция населения, а также данные по России, с детализацией по регионам и городам.

В результате реализации разработанной архитектуры был создан файл с данными aggregation_RU. На момент формирования результатов количество документов в базе aggregation_RU составляло порядка 30000. В качестве примера приводятся фрагменты файлов с данными по Самаре и Москве (рис. 1).

C Samara_dataset (1).csv *						C Moscow_dataset (1).csv *					
A						A					
	1	2	3	4	5	1	2	3	4	5	6
	1	375	427	483	540	1	362	414	468	524	582
	2	2020-03-22	9.0	0.0	0.0	2020-03-22	191.0	0.0	0.0	2020-03-25	410.0
	3	0.0	0.0	0.0	0.0	2020-03-23	262.0	0.0	0.0	2020-03-26	546.0
	4	Самарская область, RU-SAM	3.1	1.6	1.4	2020-03-24	290.0	0.0	0.0	2020-03-27	703.0
	5	Самарская область, RU-SAM	1.3	1.4	1.3	2020-03-28	817.0	4.0	18.0	2020-03-31	1613.0
	6	Самарская область, RU-SAM	1.3	2.8	2.8	2020-04-01	1880.0	16.0	115.0	2020-04-02	2475.0
	7	Самарская область, RU-SAM	2.8	3.5	3.3	2020-04-03	2923.0	20.0	168.0	2020-04-04	3357.0
	8	Самарская область, RU-SAM	2.8	2.9	2.9	2020-04-04	194.0	0.0	0.0		
	9	Самарская область, RU-SAM	2.8	3.5	3.5						
	10	Самарская область, RU-SAM	2.8	2.9	2.9						
	11	Самарская область, RU-SAM	2.9	3.4	3.4						
	12	Самарская область, RU-SAM	3.4	3.5	3.3						
	13	Самарская область, RU-SAM	3.3	3.3	3.4						
	14	Самарская область, RU-SAM	3.3	3.4	3.4						
	15	Самарская область, RU-SAM	3.6	3.8							

Рис. 1. Выгрузка данных из aggregation_RU: а) по Самаре, б) по Москве.

Интеллектуальная оптимизация

Для подбора параметров модели мы использовали несколько пакетов оптимизации:

- Scipy.optimize <https://docs.scipy.org/doc/scipy/reference/optimize.html>;
- Pyswarm pso, алгоритм роя <https://pythonhosted.org/pyswarm/>;
- Julia BBO optimize [32], алгоритм дифференциальной эволюции.

Наилучшие результаты были получены при использовании BBO optimize. В библиотеке реализовано несколько вариантов алгоритмов, которые можно запускать в пакетном режиме одной командой:

```
compare_optimizers
```

Функция потерь

В качестве функции потерь при оптимизации использовался следующий функционал (среднеквадратичная ошибка L_2):

$$w_1(I - data_I)^2 + w_2(R - data_R)^2 + w_3(D - data_D)^2 \equiv w_1 \cdot loss_I + w_2 \cdot loss_R + w_3 \cdot loss_D,$$

где w_i , $i = 1, 2, 3$ – некоторые веса, которые по умолчанию все считаются равными единице (простая функция потерь). Ниже будет рассмотрена некая стратегия определения весов. При этом $data_I$, $data_R$, и $data_D$ преобразовывались из начальных данных следующим образом: временной ряд в начальных данных представлен нарастающим итогом, мы его сначала дифференцировали, а потом брали скользящее среднее по всем временному ряду с окном в 14 дней. Окно такой протяженности выбрано поскольку это соответствует среднему времени течения болезни средней степени тяжести [33]. В результате этой операции мы получили примерное значение показателя распространённости заболевания (prevalence).

РЕЗУЛЬТАТЫ

Для определения интервалов оптимизируемых параметров, мы проанализировали современные (2020 года) исследования моделей COVID-19 и оценки, полученные на основе отчетов ВОЗ [34]. Результаты анализа представлены в Таблице 1.

Для оптимальной настройки параметров модели были использованы три различных пакета оптимизаторов, которые предназначены для этих целей. В результате многочисленных экспериментов выяснилось, что наилучшие результаты работы модели получаются с использованием алгоритма дифференциальной эволюции Julia BBO optimize. На рисунках 2–4 приведены фрагменты кода, который реализует этот алгоритм.

Таблица 1. Диапазоны эпидемических параметров

Название параметра	Диапазоны значений параметров	Ссылка на источник
β_0	0.2–3.5	[17, 19, 24, 25, 36]
c_1	0–1.0	оценка [34]
c_2	0–5.0	оценка [34]
σ	0–1.0	[18, 22, 25, 26]
γ	0.1–0.2	[13, 18, 19, 36, 37]
α	0–1.0	оценка [34]
R_0	0–1.0	[13, 22]
tg	40–130.0	оценка [34]
g_1	0–1.0	оценка [34]
g_2	0–1.0	оценка [34]

```
1 bounds = [(0.2, 3.5), (0., 1.0), (0., 5.0), (0., 1.0), (0.1, 0.2), (0., 1.0), (0., 1.0), (40.0, 70.0), (0., 1.0), (0., 1.0)]
2 res = bboptimize(loss_multi; SearchRange = bounds);
```

Рис. 2. Настройка параметров и запуск оптимизатора.

```
Starting optimization with optimizer DiffEvoOpt{FitPopulation{Float64},RadiusLimitedSelector,BlackBoxOptim.AdaptiveDiffEvoRandB
in{3},RandomBound{ContinuousRectSearchSpace}}
0.00 secs, 0 evals, 0 steps
0.52 secs, 123 evals, 71 steps, improv/step: 0.479 (last = 0.4789), fitness=1213799577.255521774
1.03 secs, 239 evals, 159 steps, improv/step: 0.421 (last = 0.3750), fitness=1112888422.016420364
1.53 secs, 479 evals, 375 steps, improv/step: 0.347 (last = 0.2917), fitness=695073836.044374347
2.03 secs, 747 evals, 639 steps, improv/step: 0.297 (last = 0.2273), fitness=388497217.235055089
2.53 secs, 979 evals, 864 steps, improv/step: 0.266 (last = 0.1778), fitness=261997267.789818347
3.04 secs, 1267 evals, 1148 steps, improv/step: 0.226 (last = 0.1021), fitness=233700978.395002186
3.54 secs, 1530 evals, 1409 steps, improv/step: 0.200 (last = 0.0881), fitness=130509446.407769263
4.05 secs, 1781 evals, 1657 steps, improv/step: 0.184 (last = 0.0927), fitness=130509446.407769263
4.55 secs, 2045 evals, 1921 steps, improv/step: 0.171 (last = 0.0871), fitness=130509446.407769263
5.05 secs, 2300 evals, 2176 steps, improv/step: 0.157 (last = 0.0549), fitness=130509446.407769263
5.55 secs, 2522 evals, 2398 steps, improv/step: 0.148 (last = 0.0586), fitness=130509446.407769263
6.07 secs, 2766 evals, 2642 steps, improv/step: 0.139 (last = 0.0492), fitness=130509446.407769263
6.57 secs, 3016 evals, 2892 steps, improv/step: 0.132 (last = 0.0600), fitness=130509446.407769263
7.07 secs, 3264 evals, 3140 steps, improv/step: 0.124 (last = 0.0282), fitness=130509446.407769263
7.58 secs, 3518 evals, 3394 steps, improv/step: 0.119 (last = 0.0630), fitness=86818651.263051867
8.09 secs, 3764 evals, 3640 steps, improv/step: 0.115 (last = 0.0569), fitness=86818651.263051867
8.61 secs, 4003 evals, 3879 steps, improv/step: 0.110 (last = 0.0293), fitness=86818651.263051867
9.11 secs, 4275 evals, 4151 steps, improv/step: 0.105 (last = 0.0331), fitness=86818651.263051867
9.62 secs, 4517 evals, 4393 steps, improv/step: 0.100 (last = 0.0248), fitness=86818651.263051867
10.13 secs, 4744 evals, 4620 steps, improv/step: 0.098 (last = 0.0573), fitness=86818651.263051867
10.64 secs, 5005 evals, 4881 steps, improv/step: 0.096 (last = 0.0536), fitness=86818651.263051867
11.14 secs, 5251 evals, 5127 steps, improv/step: 0.093 (last = 0.0447), fitness=86818651.263051867
11.65 secs, 5451 evals, 5327 steps, improv/step: 0.091 (last = 0.0200), fitness=86818651.263051867
12.16 secs, 5698 evals, 5575 steps, improv/step: 0.088 (last = 0.0242), fitness=86818651.263051867
12.67 secs, 5944 evals, 5821 steps, improv/step: 0.085 (last = 0.0203), fitness=86818651.263051867
13.17 secs, 6159 evals, 6036 steps, improv/step: 0.083 (last = 0.0186), fitness=83363057.487712845
13.68 secs, 6440 evals, 6317 steps, improv/step: 0.082 (last = 0.0605), fitness=83363057.487712845
14.18 secs, 6667 evals, 6544 steps, improv/step: 0.080 (last = 0.0352), fitness=83363057.487712845
14.69 secs, 6896 evals, 6773 steps, improv/step: 0.078 (last = 0.0306), fitness=83363057.487712845
15.19 secs, 7166 evals, 7043 steps, improv/step: 0.077 (last = 0.0519), fitness=83363057.487712845
15.70 secs, 7425 evals, 7302 steps, improv/step: 0.076 (last = 0.0309), fitness=83363057.487712845
16.21 secs, 7624 evals, 7501 steps, improv/step: 0.075 (last = 0.0452), fitness=83363057.487712845
16.71 secs, 7898 evals, 7775 steps, improv/step: 0.073 (last = 0.0365), fitness=77284467.462728173
17.21 secs, 8145 evals, 8022 steps, improv/step: 0.073 (last = 0.0445), fitness=77284467.462728173
17.73 secs, 8379 evals, 8256 steps, improv/step: 0.072 (last = 0.0470), fitness=77284467.462728173
18.24 secs, 8641 evals, 8518 steps, improv/step: 0.070 (last = 0.0267), fitness=66056836.283918992
18.74 secs, 8882 evals, 8759 steps, improv/step: 0.070 (last = 0.0373), fitness=66056836.283918992
19.25 secs, 9092 evals, 8969 steps, improv/step: 0.068 (last = 0.0143), fitness=66056836.283918992
19.76 secs, 9358 evals, 9235 steps, improv/step: 0.068 (last = 0.0451), fitness=60460529.830031902
20.26 secs, 9610 evals, 9487 steps, improv/step: 0.067 (last = 0.0476), fitness=60460529.830031902
20.76 secs, 9840 evals, 9717 steps, improv/step: 0.067 (last = 0.0478), fitness=60460529.830031902
21.27 secs, 10072 evals, 9949 steps, improv/step: 0.066 (last = 0.0388), fitness=60460529.830031902

Optimization stopped after 10001 steps and 21.43 seconds
Termination reason: Max number of steps (10000) reached
Steps per second = 466.73
Function evals per second = 472.47
Improvements/step = 0.06600
Total function evaluations = 10124

Best candidate found: [0.655329, 0.609039, 1.01636, 0.350847, 0.110794, 0.494167, 0.00915563, 44.958, 0.671146, 0.48525]
Fitness: 55479682.906840093
```

Рис. 3. Пример работы оптимизатора.

```

1 #[ $\theta$ ,  $c_1$ ,  $c_2$ ,  $\sigma$ ,  $\gamma$ ,  $\alpha$ ,  $R_0$ ,  $tg$ ,  $g_1$ ,  $g_2$ ]
2 params = [2.93199, 0.160801, 5.55758, 0.00220122, 0.19814, 0.771531, 0.0010888, 69.0643, 0.258591, 0.771477]
3

```

Рис. 4. Вывод оптимальных параметров модели.

Результаты моделирования по Самаре

Настройки кода:

1. `data = dfm [1:130, :]` – промежуток в днях;
2. `beta_gov(t, beta_0, c_1, c_2, tg, g_1, g_2) = beta_0 * gov(t, tg, g_1, g_2) * c_sig(t, c_1, c_2)` – бета функция;
3. `loss = loss_I + loss_R + loss_D;`
4. Функция `si` берет все значения из `isolation_start`.

Промежутки для параметров: $[\beta_0, c_1, c_2, \sigma, \gamma, \alpha, R_0, tg, g_1, g_2]$.

Ранее средне-полученное значение для tg составляло от 53 до 64 дней, что хоть и является первой половиной эпидемии, исходя из пика на графике(-ах), можно сказать, что это действительно так, кривая начинает изменять свое поведение в пределах от 50 до 75 дней.

Значения параметров для моделирования по Самаре, подобранных с помощью Julia BBO optimize:

`params = [0.486821, 0.88213, 2.18778, 0.198157, 0.102617, 0.490706, 0.00134632, 55.6341, 0.844033, 0.117687]`

Ниже представлены графики (рис. 5–7) с результатами моделирования распространения коронавирусной эпидемии в г. Самара, с указанными значениями параметров.

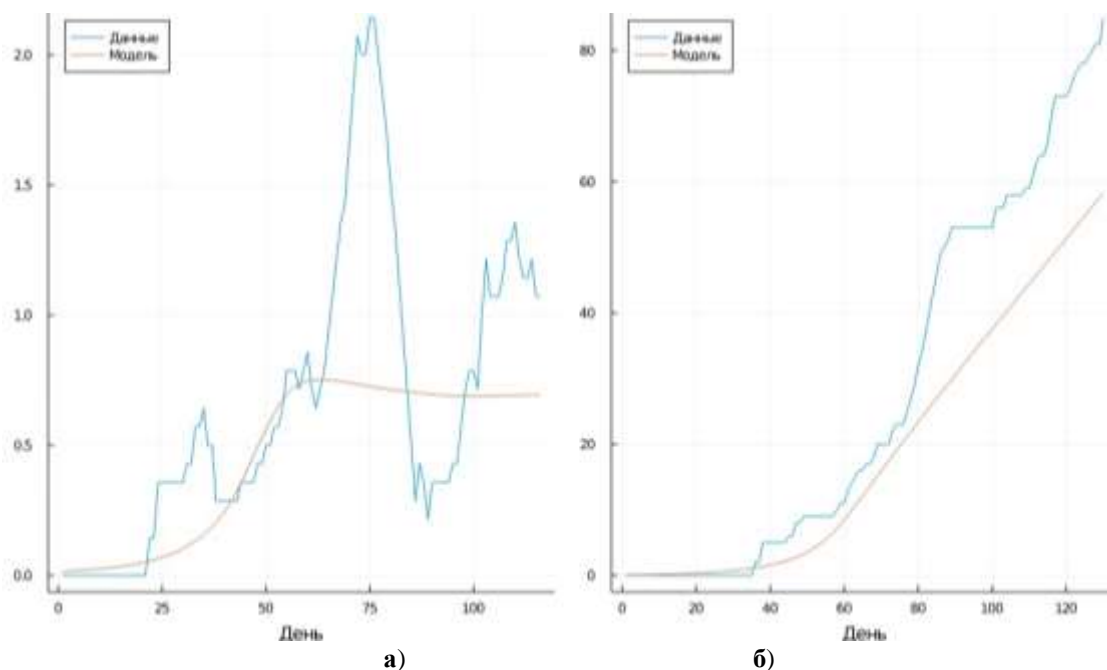


Рис. 5. Результаты моделирования по Самаре: а) Умершие D , дифференциальная сглаженная; б) Умершие D , накопительная.

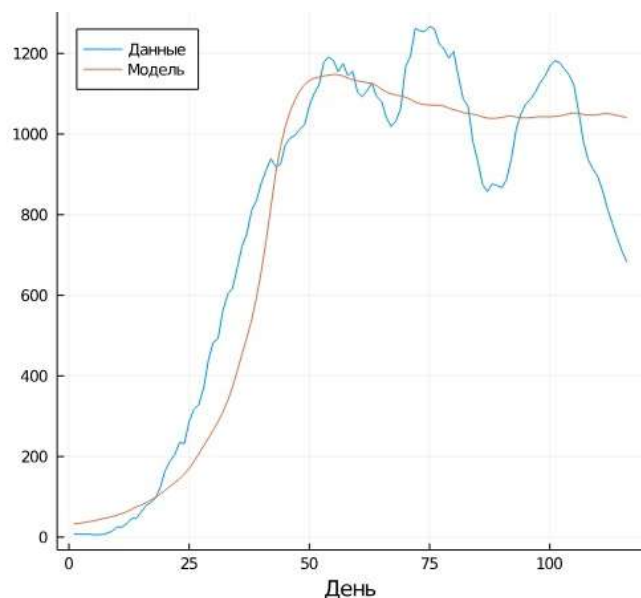


Рис. 6. Результаты моделирования по Самаре: Инфицированные I .

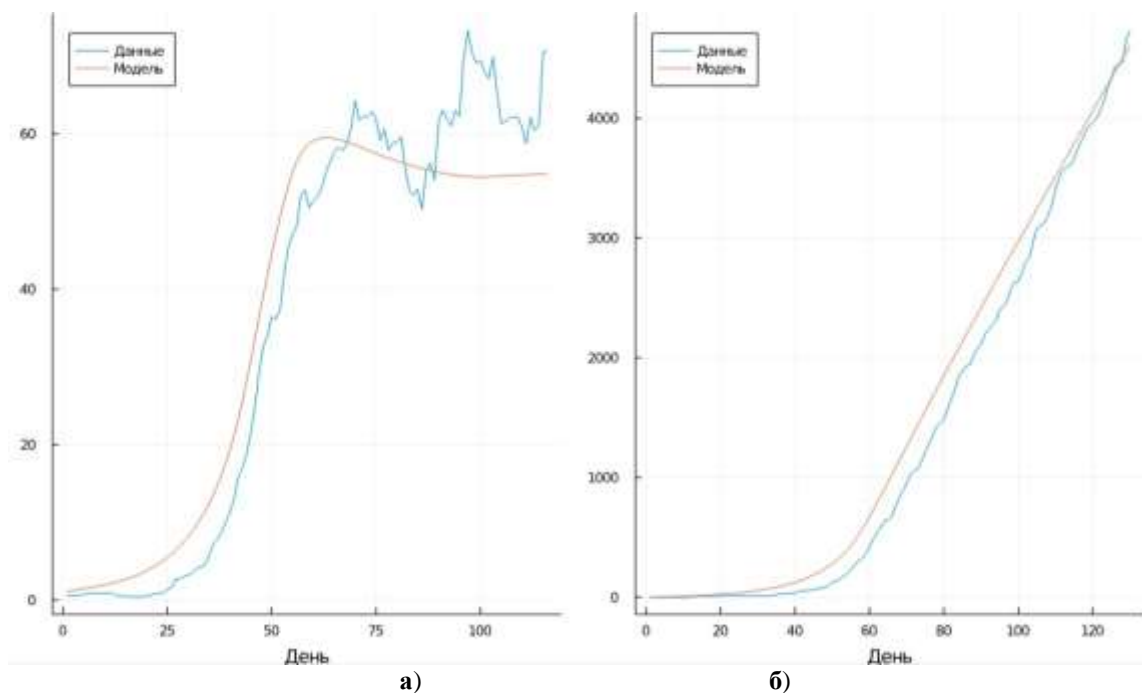


Рис. 7. Результаты моделирования по Самаре: а) Выздоровевшие R , дифференциальная сглаженная; б) Выздоровевшие R , накопительная.

Результаты моделирования по Москве

Настройки кода:

1. `data = dfm[1:130, :]` – промежуток в днях;
2. `beta_gov(t, beta_0, c_1, c_2, tg, g_1, g_2) = beta_0 * gov(t, tg, g_1, g_2) * c_sig(t, c_1, c_2)` – бета функция;
3. `loss = loss_I + loss_R + loss_D;`
4. Функция `si` берет все значения из `isolation_start`.

Промежутки для параметров: $[\beta_0, c_1, c_2, \sigma, \gamma, \alpha, R_0, tg, g_1, g_2]$.

Первая дата по данным, которые используется в модели – 23.03.2020. Обычно при подборе параметров tg равно $[40, \dots, 50]$ дней.

Промежуток этот выпадает на середину апреля – введение пропускного режима в Москве.

Значения параметров для моделирования по Москве, подобранных с помощью Julia BBO optimize:

params = [0.484185, 0.280583, 4.25395, 0.359855, 0.128215, 0.576838, 0.00190299, 49.1466, 0.924046, 0.215798]

Ниже представлены графики (рис. 8–10) с результатами моделирования распространения коронавирусной эпидемии в г. Москва, с указанными значениями параметров.

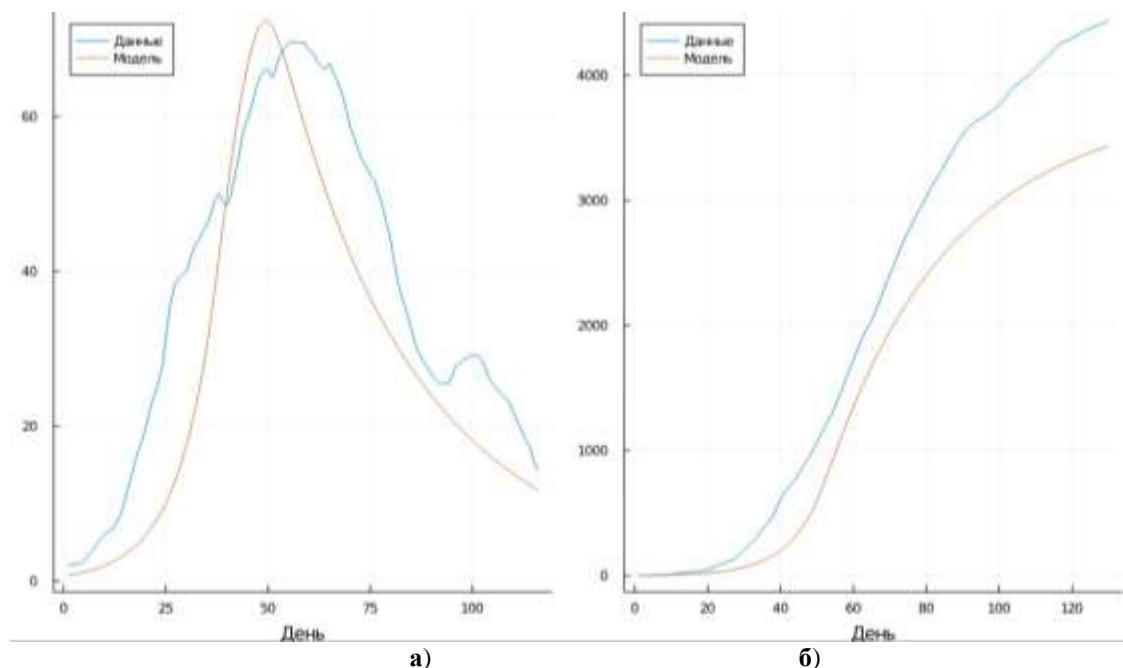


Рис. 8. Результаты моделирования по Москве: а) Умершие D , дифференциальная, сглаженная; б) Умершие D , накопительная.

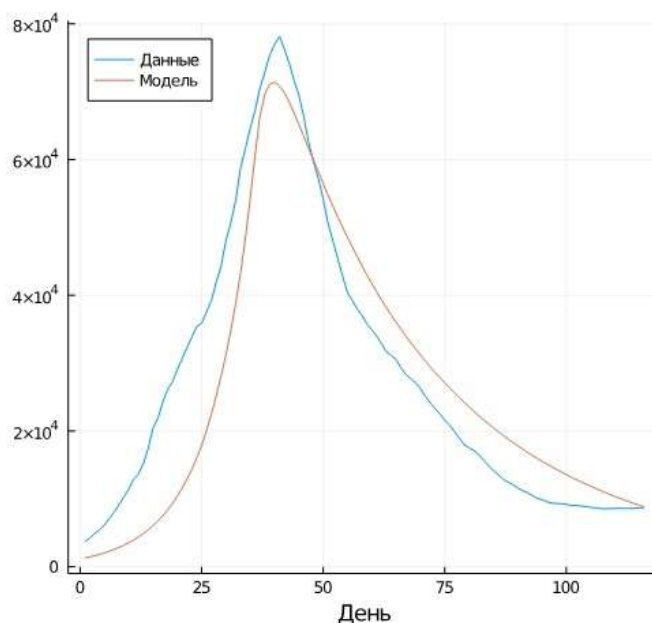


Рис. 9. Результаты моделирования по Москве: Инфицированные I .

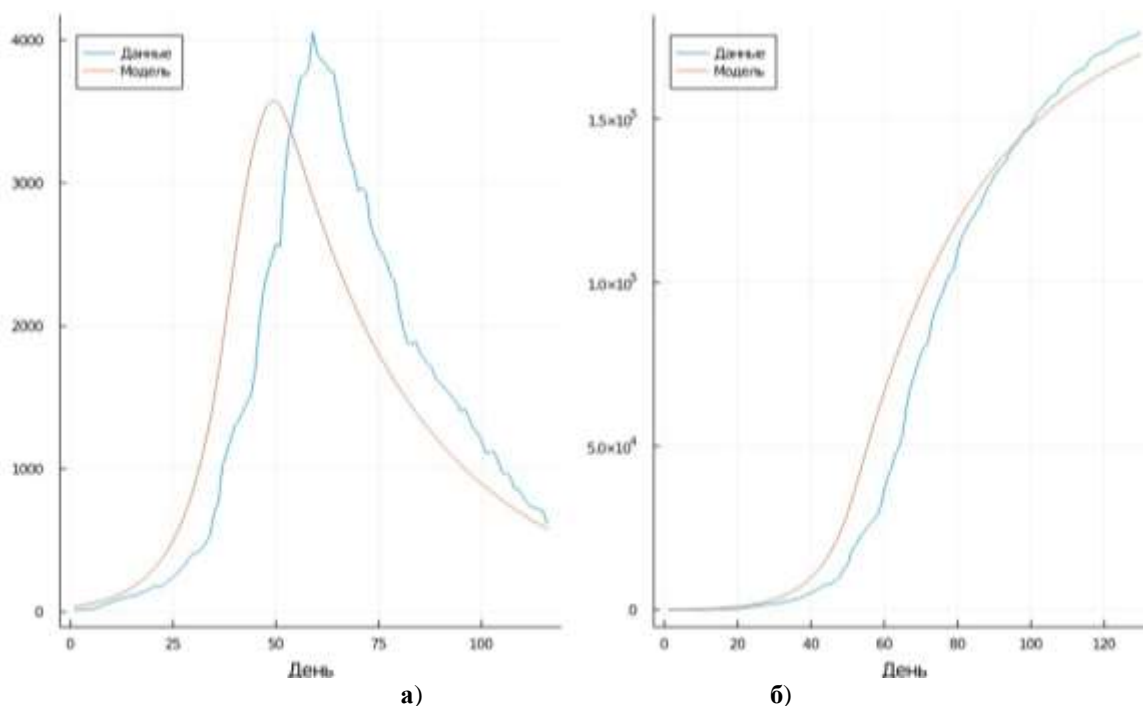


Рис. 10. Результаты моделирования по Москве: а) Выздоровевшие R , дифференциальная сглаженная; б) Выздоровевшие R , накопительная.

Сравнение результатов моделирования по Москве с простой и взвешенной функцией потерь

В качестве функции потерь мы используем сумму среднеквадратичных ошибок по трём переменным $loss_I$, $loss_R$ и $loss_D$. Так как оптимизатор стремится сделать это число как можно меньше, наименьшая ошибка при простом суммировании будет достигаться при более точном аппроксимировании функции переменной I , в то время как слагаемое $loss_D$ не будет так сильно влиять на изменение суммы, поскольку среднее число зараженных за все дни эпидемии превышает среднее число умерших почти в 65 раз.

```
mean(data.deaths)
```

```
2088.5923076923077
```

```
mean(data.confirmed)
```

```
135025.4153846154
```

```
mean(data.recovered)
```

```
73481.4076923077
```

Таким образом, при оптимизации более точное приближение переменной D к реальным значениям количества смертей почти никак не влияет на изменение функции потерь. Для того, чтобы избежать этого, можно привести слагаемые в функции потерь к одному масштабу (например, уменьшив значение переменной $loss_I$ в те же ~ 65 раз). Для того, чтобы точно определить веса для каждого слагаемого в функции потерь, разделим среднее значение числа смертей в наших данных сначала на среднее число выздоровевших, а потом на среднее число умерших.

```
mean(data.deaths)/mean(data.recovered)
```

```
0.02842341176203337
```

```
mean(data.deaths)/mean(data.confirmed)
```

```
0.015468142066017885
```

Таким образом получим взвешенную функцию потерь: $0.015468*loss_I + 0.02842*loss_R + loss_D$.

Снижение влияние среднеквадратичной ошибки на изменение переменной I , увеличивает значение функции потерь по ней. Однако, это позволяет достичь намного более высокой точности по остальным двум переменным.

Кроме того, с учетом описанной выше методики расчета среднеквадратичной ошибки, оптимизатор подбирает разные параметры в зависимости от того, каким образом мы учитываем ошибку каждой переменной. Для сравнения будем использовать три основных варианта среднеквадратичной ошибки: обычное суммирование ($loss_I + loss_R + loss_D$), взвешенная сумма ($w_1*loss_I + w_2*loss_R + w_3*loss_D$) и сумму без учета одного из значений ошибки (один из коэффициентов w_1 , w_2 или w_3 из предыдущего выражения равен нулю).

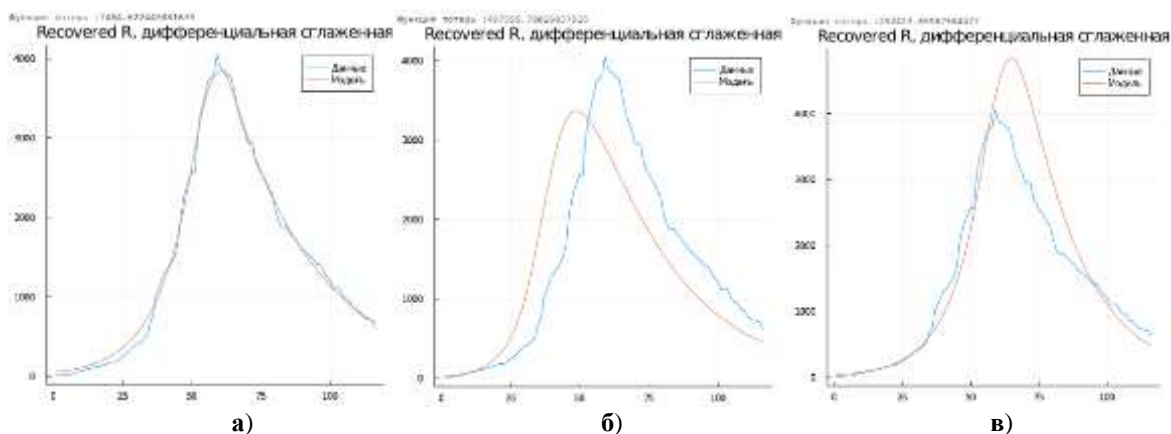


Рис. 11. Результаты моделирования по Москве: а) выздоровевшие R , дифференциальная сглаженная с взвешенной суммой; б) Выздоровевшие R , дифференциальная сглаженная с обычным суммированием; в) Выздоровевшие R , дифференциальная сглаженная с обнулением одного из значений веса.

ОБСУЖДЕНИЕ

Заметим, что данные по смертям D в целом представляют из себя сильно разряженные данные (или малые данные), поскольку их количество (особенно в случае Самары) незначительно по сравнению с общим размером популяции и количеством подверженных эпидемии. Тем не менее наша модель достаточно хорошо приближает количество D . Если сравнить результаты по Москве и Самаре, то хорошо прослеживается эффект больших данных – результаты по Москве получаются точнее, поскольку там данных значительно больше. Тем не менее, модель в целом правильно отражает нужный момент для реакции властей.

Мы также провели численные эксперименты по Нижегородской, Новосибирской областям и Санкт-Петербургу. Результаты получились примерно того же уровня качества, что для Москвы и Самары. Это свидетельствует о том, что предложенный подход, который решает сложную задачу настройки параметров модели эпидемии из реальных данных, вполне эффективен. Однако, ввиду ограниченности объема статьи, мы эти результаты здесь не приводим.

В развитие предложенного подхода, планируется включить в модель важный процесс вакцинации населения. В моделях с вакцинацией значение параметра R_0 может быть использовано для оценки уровня вакцинации [35–37]. Если поддерживать уровень вакцинации населения ниже критического уровня $P_c = 1 - 1/R_0$, то патогенное воздействие будет рассеиваться, а эпидемия затухать.

ЗАКЛЮЧЕНИЕ

Представлена гибридная адаптивно-компаратментная модель распространения короновиральной эпидемии и показана возможность ее оптимизации за счет точной настройки параметров модели из данных по эпидемии. Принятие реальных решений на предотвращение/смягчение угроз требует, чтобы проигрывалось множество возможных вариантов [38, 39] с относительно несложным аппаратом интерпретации, за небольшой отрезок времени. Поэтому мы используем сочетание современных технологий интеллектуальной оптимизации и методов системной динамики. Такой подход позволяет справиться с основными трудностями рассматриваемой задачи, а именно: 1) будучи реальным феноменом, она содержит большое число параметров, которые, в конце концов, необходимо учесть и настроить, чтобы сделать модель как можно более приближенной к реальным процессам; 2) ожидаемая размерность пространства данных, необходимая для достижения той же цели, весьма высока, а количество данных велико. Эти два обстоятельства делают практически неприменимым традиционный математический аппарат статистического анализа, регрессий, сеточных супервычислений и т.п. Заметим также, что рассматриваемая задача носит комплексный междисциплинарный характер и связана с разработкой методологии, инструментов и технологий повышения эффективности государственного управления в кризисных ситуациях.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта по гранту № 20-04-60160_Вирусы.

Вычисления проведены с помощью гибридного суперкомпьютера К-60, установленного в Центре коллективного пользования ИПМ им. М.В. Келдыша РАН.

СПИСОК ЛИТЕРАТУРЫ

1. Bernoulli D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculum pour la prévenir. *Mém. Math. Phys. Acad. Roy. Sci. Paris*. 1760. P. 1–45.
2. Farr W. Progress of epidemics. 2d Report of the regist. London: General of England and Wales, 1840.
3. Енько П.Л. О ходе эпидемий некоторых заразных болезней. В: *Врач*. СПб, 1889. № 46–48.
4. Ross R., Hudson H.P. An application of the theory of probabilities to the study of a priori pathometry. *Proc. R. Soc. Lond.* 1916-1917. A 93. P. 212-240.
5. Бейли Н. *Математика в биологии и медицине*. М.: МИР, 1970. 326 с.
6. Бароян О.В., Рвачев Л.А. *Математика и эпидемиология*. М.: Знание, 1977. 63 с.
7. Боев. Б.В. Современные этапы математического моделирования процессов развития и распространения инфекционных заболеваний. В: *Эпидемиологическая кибернетика: модели, информация, эксперименты*. 1991. С. 6–13.
8. Anderson R.M., May R.M. *Infectious Diseases of Humans. Dynamics and Control*. Oxford: Oxford University Press, 1991.
9. Macdonald G. The measurement of malaria transmission. *Proc. R. Soc. Med.* 1955. V. 48. № 4. P. 295–302.

10. Hoppensteadt F. An age dependent epidemic model. *Journal of the Franklin Institute*. 1974. P. 325–333.
11. Gupur G., Li Xue-Zhi, Zhu Guang-Tian. Threshold and Stability Results for an Age-Structured Epidemic Model. *Computers and Mathematics with Applications*. 2001. V. 42. P. 883–907.
12. Park T. *Age-dependence in epidemic models of vector-borne infections*. Huntsville: The University of Alabama, 2004.
13. Giordano G., Blanchini F., Bruno R., Colaneri P., Di Filippo A., Di Matteo A., Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*. 2020. V. 26. P. 855–860. doi: [10.1038/s41591-020-0883-7](https://doi.org/10.1038/s41591-020-0883-7)
14. Букин Ю.С., Джигоев Ю.П., Бондарюк А.Н., Ткачев С.Е., Злобин В.И. *Применение универсальной математической модели эпидемического процесса «SRID» для прогноза развития эпидемии COVID-19 в городе Москва*: препринт. 2020. 21 с. doi: [10.24108/preprints-3112045](https://doi.org/10.24108/preprints-3112045)
15. Derrode S., Gauchon R., Ponthus N., Rigotti C., Pothier C., Volpert V., Loisel S., Bertoglio J.-P., Roy P. Piecewise estimation of R_0 by a simple SEIR model. Application to COVID-19 in French regions and departments until June 30. *Université Lyon 1 - Claude Bernard*. 2020. URL: <https://www.semanticscholar.org/paper/Piecewise-estimation-of-R0-by-a-simple-SEIR-model.-Derrode-Gauchon/7394add56a191e2fed3121c2d935d7f3f278320a#paper-header> (accessed 14.05.2020).
16. Kočańczyk M., Grabowski F., Lipniacki T. Super-spreading events initiated the exponential growth phase of COVID-19 with \mathcal{R}_0 higher than initially estimated. *Biology, Medicine. Royal Society Open Science*. 2020. Article No. 200786. doi: [10.1098/rsos.200786](https://doi.org/10.1098/rsos.200786)
17. Wang X., Tang T., Cao L., Aihara K., Guo Q. Inferring key epidemiological parameters and transmission dynamics of COVID-19 based on a modified SEIR model. *Math. Model. Nat. Phenom.* 2020. V. 15. P. 74. Article No. 2020050. doi: [10.1051/mmnp/2020050](https://doi.org/10.1051/mmnp/2020050)
18. Grigorieva E.V., Khailov E. N., Korobeinikov A. Optimal quarantine strategies for COVID-19 control models. *arXiv:2004.10614 [math.OC]* 2020. 21 p. URL: <http://arxiv.org/abs/2004.10614v3> (accessed 14.05.2020).
19. Zhou X., Ma X.-d., Hong N., Su L., Ma Y., He J., Jiang H., Liu C., Shan G., Zhu W., Zhang S., Long Y. Forecasting the Worldwide Spread of COVID-19 based on Logistic Model and SEIR Model. 2020. *Medicine, Geography. medRxiv*. Article No. 20044289. doi: [10.1101/2020.03.26.20044289](https://doi.org/10.1101/2020.03.26.20044289)
20. Lin Q., Zhao S., Gao D., Lou Y., Yang S., Musa S. S., Wang M.H., Cai Y., Wang W., Yang L., He D. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases*. doi: [10.1016/j.ijid.2020.02.058](https://doi.org/10.1016/j.ijid.2020.02.058)
21. Zamir M., Abdeljawad T., Nadeem F., Wahid A., Yousef A. An optimal control analysis of a COVID-19 model Author links open overlay panel. *Alexandria Engineering Journal*. 2021. V. 60. No. 3. P. 2875-2884. doi: [10.1016/j.aej.2021.01.022](https://doi.org/10.1016/j.aej.2021.01.022)
22. Okhuese V.A., Mathematical predictions for COVID-19 as a global pandemic. *medRxiv*. 2020. doi: [10.1101/2020.03.19.20038794](https://doi.org/10.1101/2020.03.19.20038794)
23. Zeng Y., Guo X., Deng Q., Luo S., Zhang H. Forecasting of COVID-19: spread with dynamic transmission rate. *Journal of Safety Science and Resilience*. 2020. V. 1. No. 2. P. 91-96. doi: [10.1016/j.jnlssr.2020.07.003](https://doi.org/10.1016/j.jnlssr.2020.07.003)

24. Томчин Д.А., Фрадков А.Л. *Прогнозирование распространения вируса COVID-19 в России на основе простых математических моделей эпидемий*. СПб.: ИПМаш РАН. 2020. 17 с.
25. Mamo D.K. Model the transmission dynamics of COVID-19 propagation with public health intervention. *Results in Applied Mathematics*. 2020. V. 7. Article No. 100123. doi: [10.1016/j.rinam.2020.100123](https://doi.org/10.1016/j.rinam.2020.100123)
26. Garba S. M., Lubuma J. M.-S., Tsanou B. Modeling the transmission dynamics of the COVID-19 Pandemic in South Africa. *Mathematical Biosciences*. 2020. V. 328. Article No. 108441. doi: [10.1016/j.mbs.2020.108441](https://doi.org/10.1016/j.mbs.2020.108441)
27. Aghdaoui H., Tilioua M., Sooppy Nisar K., Khan I. A Fractional Epidemic Model with Mittag-Leffler Kernel for COVID-19. *Mathematical Biology and Bioinformatics*. 2021. V. 16. № 1. P. 39-56. doi: [10.17537/2021.16.39](https://doi.org/10.17537/2021.16.39)
28. Кондратьев М.А. Методы прогнозирования и модели распространения заболеваний. *Компьютерные исследования и моделирование*. 2013. Т. 5. № 5. С. 863-882.
29. Васильев Ф.П. *Методы оптимизации*. М.: Факториал Пресс, 2002.
30. Krivorotko O., Kabanikhin S., Zyatkov N., Prikhod'ko A., Prokshoshin N., Shishlenin M. Mathematical modeling and prediction of COVID-19 in Moscow and Novosibirsk region. *Numerical Analysis and Applications*. 2020. V. 13. P. 332–348. doi: <https://doi.org/10.1134/S1995423920040047>
31. He D., Dushoff J., Day T., Ma J., Earn D. J.-D. Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proceedings of the Royal Society B: Biological Sciences*. 2013. doi: [10.1098/rspb.2013.1345](https://doi.org/10.1098/rspb.2013.1345)
32. *BlackBoxOptim.jl*. URL: <https://github.com/robertfeldt/BlackBoxOptim.jl> (accessed 14.05.2020).
33. Самородская И.В. Проблемы диагностики и учета заболеваемости COVID-19. *Медвестник*. 2020. URL: <https://medvestnik.ru/content/medarticles/Problemy-diagnostiki-i-ucheta-zabolevaemosti-COVID-19.html> (accessed 14.05.2020).
34. *Клиническое ведение случаев COVID-19: временное руководство*. ВОЗ, 2020. URL: <https://apps.who.int/iris/bitstream/handle/10665/332196/WHO-2019-nCoV-clinical-2020.5-rus.pdf> (accessed 14.05.2020).
35. Anderson R.M., May R.M. Directly transmitted infectious diseases: Control by vaccination. *Science*. 1982. P. 1053-1060.
36. Roberts M., Heesterbeek H. Bluff your way in epidemic models. *Trends in Microbiology*. 1993. V. 1. No. 9. P. 343–348. doi: [10.1016/0966-842x\(93\)90075-3](https://doi.org/10.1016/0966-842x(93)90075-3)
37. Ferguson N.M., Keeling M.J., Edmunds W.J., Gani R., Grenfell B.T., Anderson R.M., Leach S. Planning for smallpox outbreaks. *Nature*. 2003. V. 425. № 6959. P. 681–685.
38. Frost I., Craig J., Osen G., Hauck S., Kalanxhi E., Schueller E., Gatalo O., Yan-y Y., Tseng K., Lin G., Klein E. Modeling COVID-19 Transmission in Africa: Country-wise Projections of Total and Severe Infections Under Different Lockdown Scenarios. *medRxiv*. 2020. Article No. 20188102. doi: [10.1101/2020.09.04.20188102](https://doi.org/10.1101/2020.09.04.20188102)
39. Bertozzi A.L., Franco E., Mohler G., Short M.B., Sledge D. The challenges of modeling and forecasting the spread of COVID-19. *PNAS*. 2020. V. 117. No. 29. P. 16732-16738. doi: [10.1073/pnas.2006520117](https://doi.org/10.1073/pnas.2006520117)

Рукопись поступила в редакцию 23.04.2021, переработанный вариант поступил 21.05.2021.
Дата опубликования 24.05.2021.

Study of SEIRD Adaptive-Compartmental Model of COVID-19 Epidemic Spread in Russian Federation Using Optimization Methods

Levashkin S.P.^{1,2}, Agapov S.N.¹, Zakharova O.I.¹, Ivanov K.N.¹,
Kuzmina E.S.¹, Sokolovsky V.A.¹, Monasova A.S.¹, Vorobiev A.V.¹,
Apeshin D.N.¹

¹*Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia*

²*Samara State Technical University, Samara, Russia*

Abstract. A systemic approach to the study of a new multi-parameter model of the COVID-19 pandemic spread is proposed, which has the ultimate goal of optimizing the manage parameters of the model. The approach consists of two main parts: 1) an adaptive-compartmental model of the epidemic spread, which is a generalization of the classical SEIR model, and 2) a module for adjusting the parameters of this model from the epidemic data using intelligent optimization methods. Data for testing the proposed approach using the pandemic spread in some regions of the Russian Federation were collected on a daily basis from open sources during the first 130 days of the epidemic, starting in March 2020. For this, a so-called "data farm" was developed and implemented on a local server (an automated system for collecting, storing and preprocessing data from heterogeneous sources, which, in combination with optimization methods, allows most accurately tune the parameters of the model, thus turning it into an intelligent system to support management decisions). Among all model parameters used, the most important are: the rate of infection transmission, the government actions and the population reaction.

Key words: *multivariable modeling, COVID-19, epidemic spread model, multivariable optimization, loss function.*

This research was partially supported by the Russian Foundation for Basic Research under grant No. 20-04-60160_Viruses.

The calculations were performed on the hybrid supercomputer K-60 installed in the Supercomputer Center of Collective Usage of KIAM RAS.