



Математическое моделирование пандемии COVID-19: большие данные и высокопроизводительные вычисления

Криворотько О.И.
krivorotko.olya@mail.ru

Содержание

1. Введение

- Краткое определение SIR и агентных моделей.
- Индекс репродукции.

2. SIR-модели распространения COVID-19:

- Постановка прямой и обратной задач.
- Индекс репродукции для SEIR-HCD модели.
- Анализ чувствительности параметров.

3. Агентное моделирование в эпидемиологии:

- Описание агентно-ориентированной модели распространения COVID-19.
- Анализ идентифицируемости.

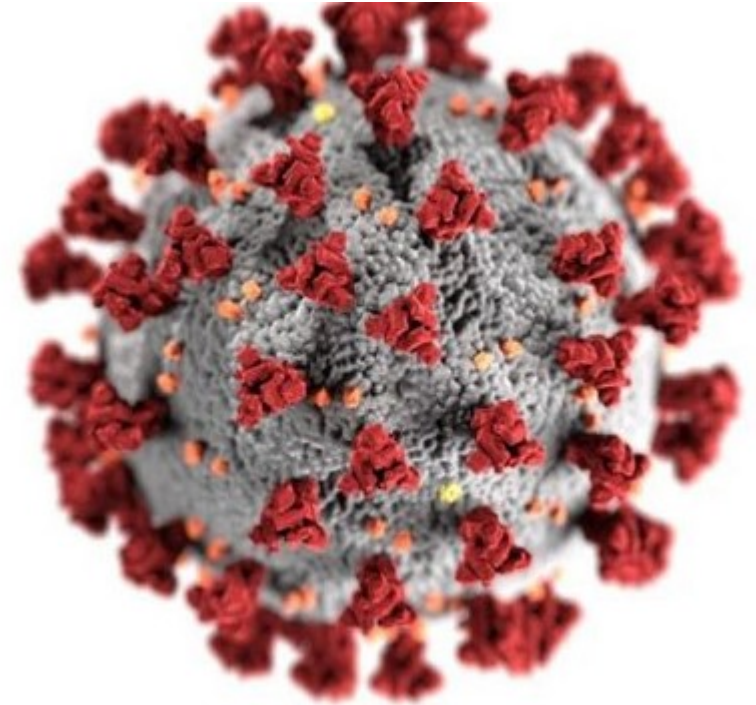
4. Постановка задач минимизации целевого функционала и методы их решения.

5. Комплекс программ:

- Алгоритм моделирования и построения сценариев развития динамики выявленных случаев COVID-19.
- Mean field games как объединение двух подходов моделирования.

6. Численные расчеты распространения COVID-19 в Новосибирской области:

- Верификация модели.
- Сценарии развития.
- Суперкомпьютерный анализ алгоритмов.

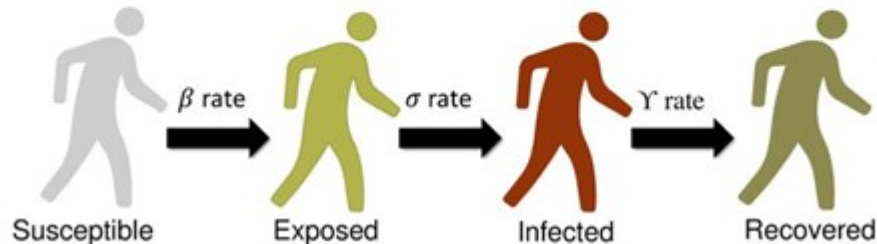


Математическое моделирование в эпидемиологии: два подхода

Моделирование «сверху-вниз»

SEIR Model

model the flows of people between four states: susceptible (S), exposed (E), infected (I), and recovered (R).



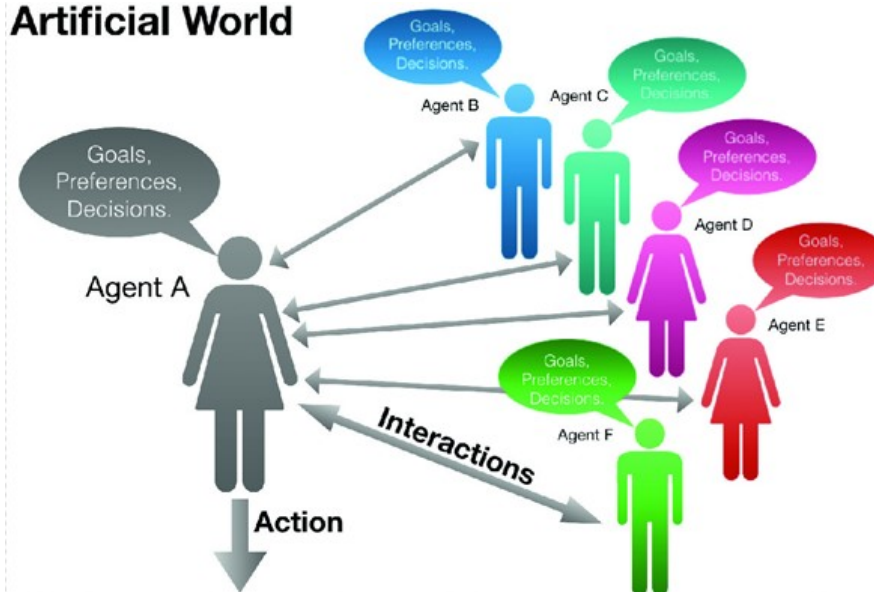
Популяция делится на группы со схожими признаками

R. Ross (1911), A.J. Lotka, V. Volterra (1920-1926), G.U. Yule (1924), **W.O. Kermack, A.G. McKendrick (1926-1927)**, R.A. Fisher (1937), А.Н. Колмогоров, И.Г. Петровский, Н.С. Пискунов (1937), J.M. Smith, G.R. Price (1973), L. Laguzet (2015), Н. Tembine (2020) и др.

SIR-модели

Моделирование «снизу-вверх»

Artificial World



Каждый агент наделен набором признаков и правилом взаимодействия с другими агентами на основе случайных графов.

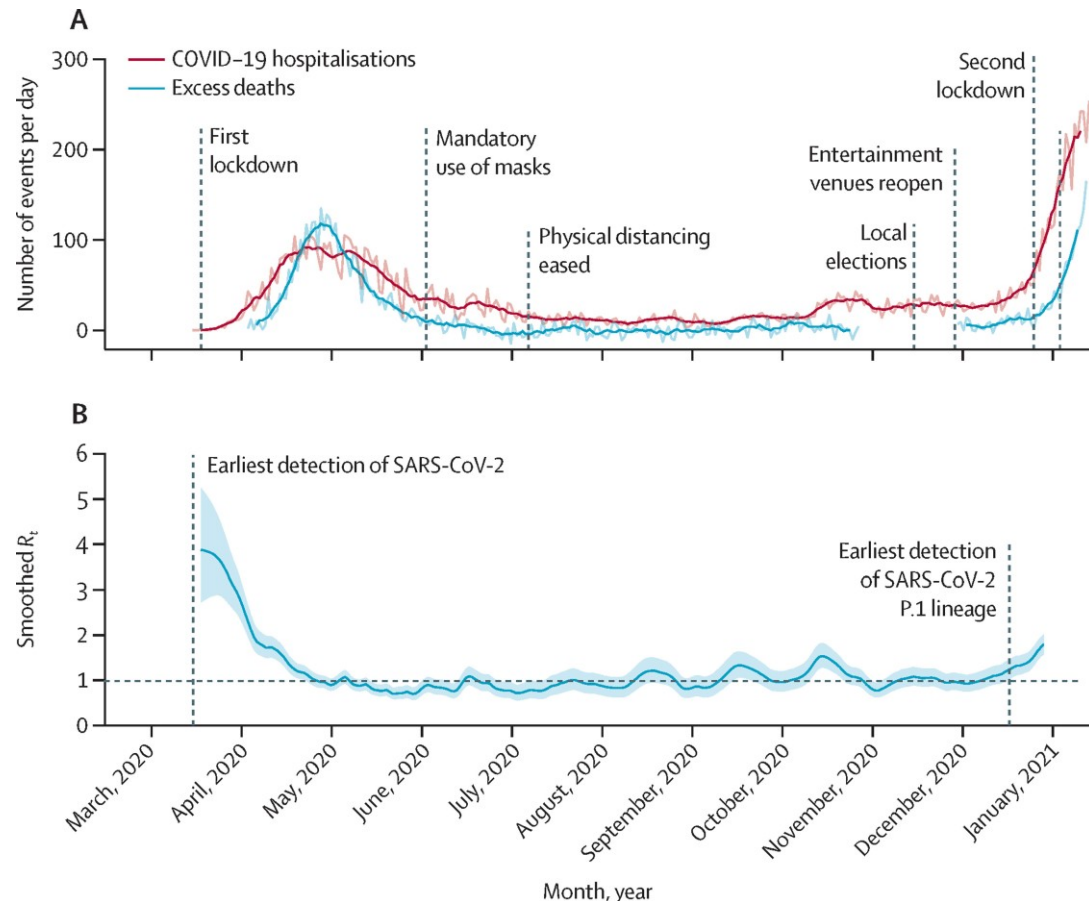
T.C. Shelling (1971) и M. Mitchel (1993), P. Patrolla (2004) и др.

Агентные модели

Индекс репродукции в моделях эпидемиологии

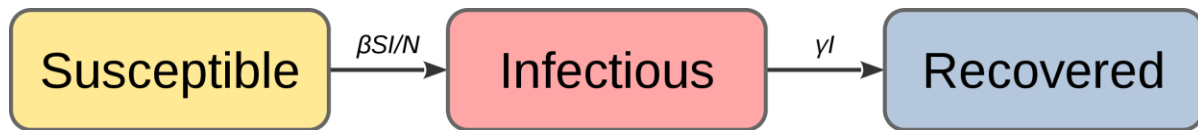
Индекс репродукции вируса \mathcal{R}_0 определяется как среднее количество людей, которых заражает активный инфицированный, попавший в полностью неиммунизированное окружение при отсутствии специальных эпидемиологических мер, направленных на предотвращение распространения заболевания.

| Заболевание | \mathcal{R}_0 | Путь передачи инфекции |
|---------------------------------|-----------------|---------------------------|
| Корь | 12-18 | Воздушный |
| Ветряная оспа | 10-12 | Воздушный |
| Краснуха | 5-7 | Воздушно-капельный |
| COVID-19 | 1.4-5.7 | Воздушно-капельный |
| СПИД | 2-5 | Жидкости тела |
| SARS-CoV (Китай, 2003) | 2-5 | Воздушно-капельный |
| Грипп «испанка» (пандемия 1918) | 1.4-2.8 | Воздушно-капельный |
| Эбола | 1.5-1.9 | Жидкости тела |
| Грипп (пандемия 2009) | 1.4-1.6 | Воздушно-капельный |
| Грипп сезонный | 0.9-2.1 | Воздушно-капельный |
| MERS (2012) | 0.3-0.8 | Воздушно-капельный |



Взаимосвязь индекса репродукции и случаев госпитализации и смертности от COVID-19 в Бразилии.

SIR-модель

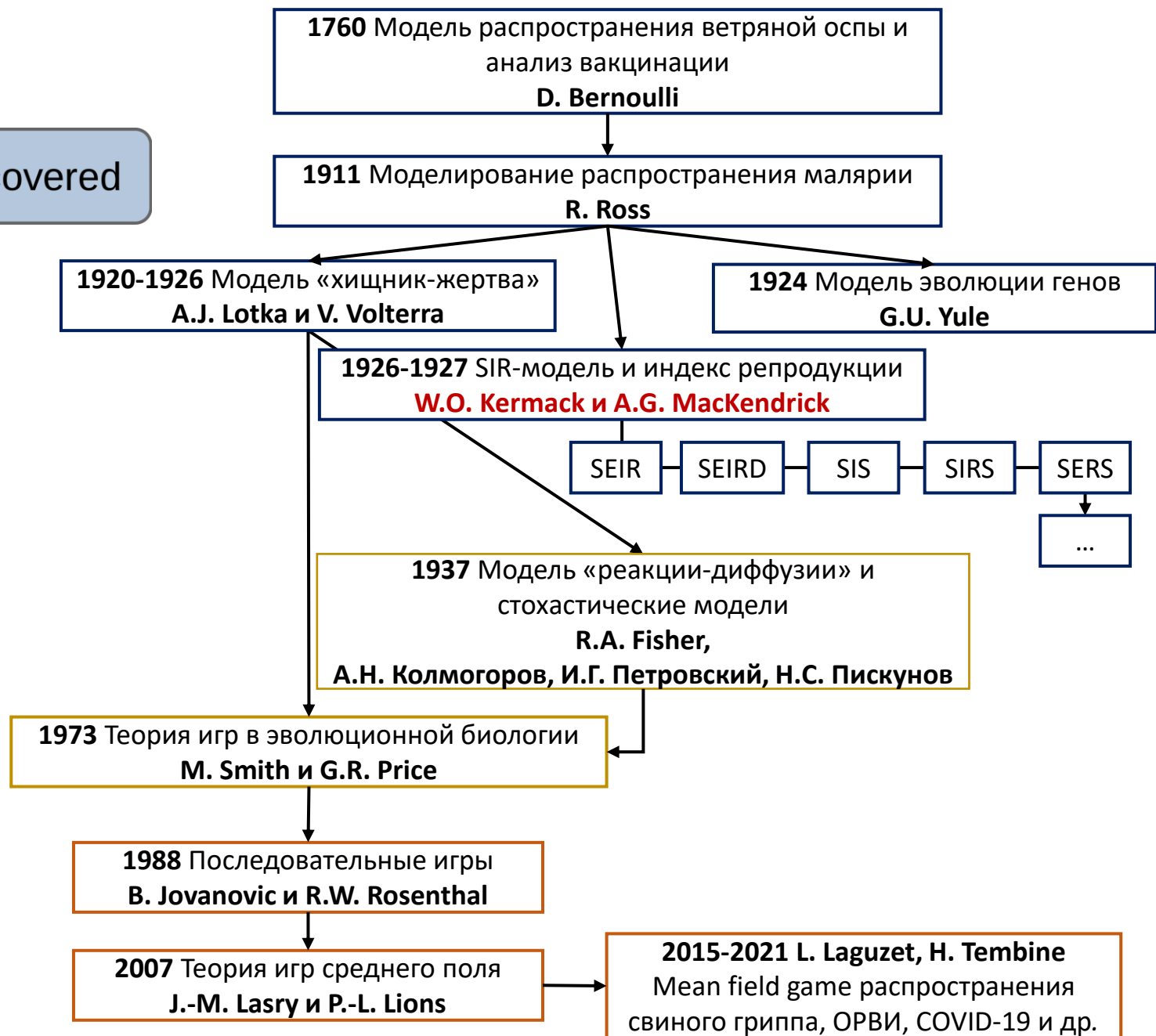


$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N}, & S(0) = N - I_0, \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, & I(0) = I_0 < N, \\ \frac{dR}{dt} = \gamma I, & R(0) = 0. \end{cases}$$

$N = S(t) + I(t) + R(t)$ - константа

Символ Описание

| Символ | Описание |
|----------|---|
| β | Уровень инфицирования, описывающий вероятность передачи болезни между восприимчивым и заразным индивидуумом |
| γ | Скорость выздоровления, определяемая средней продолжительностью заражения |
| N | Размер моделируемой популяции |



Прямая и обратная задачи для SIR-моделей

Закон баланса масс: $u(t) = \text{const}$

$$\frac{du}{dt} = \mu(t, u, q), u(t_0) = u_0.$$

Прямая задача: найти решение системы ОДУ $u(t) \geq 0$ при заданных коэффициентах q и начальных условиях $u_0 \geq 0$.

Обратная задача: найти коэффициенты q и/или начальные условия u_0 по дополнительным измерениям $u_i(t_k) = f_{ik}$, которые могут быть получены из статистических данных.

Недостатком SIR-моделей является отсутствие гибкости – невозможность учета изменения параметров (новые мутации вируса и штамма, ограничительные меры, вакцинация).
При попытке ввести в SIR-модели указанные изменения (например, сделать переменной скорость передачи инфекции $\beta = \beta(t)$), мы сталкиваемся с неединственностью и неустойчивостью решения обратной задачи идентификации этого параметра $\beta(t)$.



$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N}, & S(0) = N - I_0, \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, & I(0) = I_0 < N, \\ \frac{dR}{dt} = \gamma I, & R(0) = 0. \end{cases}$$

$N = S(t) + I(t) + R(t)$ - константа

Обратная задача состоит в уточнении коэффициентов β и γ , а также начальных данных I_0 по дополнительным измерениям количества индивидов, находящихся на лечении:
 $b_k I(t_k) = f_k, k = 1, \dots, K.$

Индекс репродукции вируса в SIR-моделях

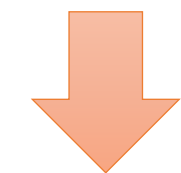
Индекс репродукции \mathcal{R}_0 является границей устойчивости состояния равновесия SIR-системы при отсутствии инфицированных.

$\mathcal{R}_0 > 1$, то на начальном этапе число заболевших будет расти экспоненциально.

$\mathcal{R}_0 \in (0,1)$, то небольшое количество инфицированных людей, попавших в полностью восприимчивую популяцию, в среднем не смогут сохранить свою группу, и эпидемии не будет.



$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N}, & S(0) = N - I_0, \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, & I(0) = I_0 < N, \\ \frac{dR}{dt} = \gamma I, & R(0) = 0. \end{cases}$$



$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

P. van den Driessche, J. Watnough. Further notes on the basic reproduction number. In: Mathematical epidemiology. Springer, 2008. P. 159-178.

Чрезвычайно важна роль как индекса репродукции, так и начальной восприимчивости. Уравнение для инфекционных индивидуумов:

$$\frac{dI}{dt} = \left(\mathcal{R}_0 \frac{S}{N} - 1 \right) \gamma I,$$

Если $\mathcal{R}_0 S(0) > N$

$$\frac{dI}{dt}(0) > 0$$

Произойдет настоящая эпидемическая вспышка с увеличением числа инфекционных (которые могут охватить значительную часть населения).

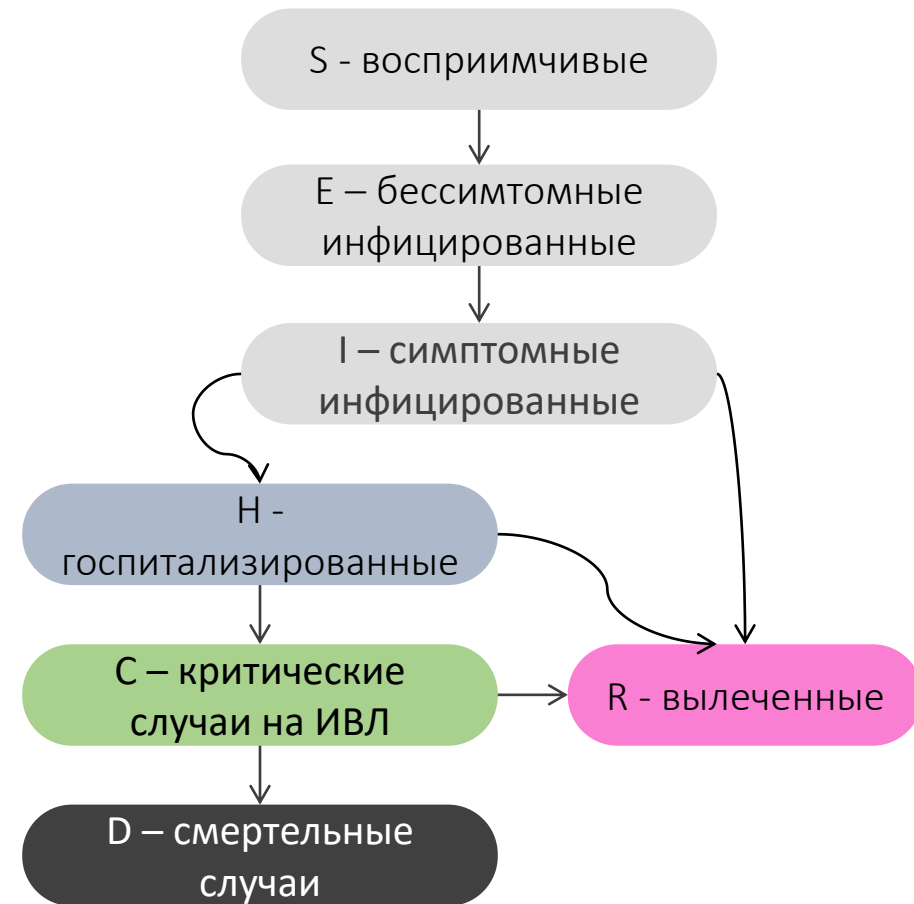
Если $\mathcal{R}_0 S(0) < N$

$$\frac{dI}{dt}(0) < 0$$

Независимо от первоначального размера восприимчивого населения болезнь никогда не может вызвать настоящую эпидемическую вспышку.

Математическая модель SEIR-HCD распространения COVID-19

$$\begin{cases} \frac{dS}{dt} = -\frac{5-a(t-\tau)}{5} \left(\alpha_E(t) \frac{S(t)E(t)}{N(t)} + \alpha_I(t) \frac{S(t)I(t)}{N(t)} \right), & S(0) = S_0, \\ \frac{dE}{dt} = \frac{5-a(t-\tau)}{5} \left(\alpha_E(t) \frac{S(t)E(t)}{N(t)} + \alpha_I(t) \frac{S(t)I(t)}{N(t)} \right) - \frac{1}{t_{inc}} E(t), & E(0) = E_0, \\ \frac{dI}{dt} = \frac{1}{t_{inc}} E(t) - \frac{1}{t_{inf}} I(t), & I(0) = I_0, \\ \frac{dR}{dt} = \frac{\beta}{t_{inf}} I(t) + \frac{1-\varepsilon_{HC}}{t_{hosp}} H(t), & R(0) = R_0, \\ \frac{dH}{dt} = \frac{1-\beta}{t_{inf}} I(t) + \frac{1-\mu}{t_{crit}} C(t) - \frac{1}{t_{hosp}} H(t), & H(0) = H_0, \\ \frac{dC}{dt} = \frac{\varepsilon_{HC}}{t_{hosp}} H(t) - \frac{1}{t_{crit}} C(t), & C(0) = C_0, \\ \frac{dD}{dt} = \frac{\mu}{t_{crit}} C(t), & D(0) = D_0. \end{cases}$$



Вектор неизвестных параметров: $\mathbf{q} = (\alpha_E(t), \alpha_I(t), \beta, \varepsilon_{HC}, \mu, E_0, I_0)$

Данные: $I(t_k) = (1 - b_k)f_k, C(t_k) = C_k, D(t_k) = g_k.$

Обратная задача: найти \mathbf{q} по данным $f_k, b_k, C_k, g_k, k = 1, \dots, K.$

SEIR-HCD модель распространения COVID-19: ограничения

В модели HE учитывается:

- Естественные рождение и смерть и некоторые другие параметры, приводящие к изменениям в популяции.
- Разница между мужчинами и женщинами (но статистика показывает, что мужчины были более уязвимы для вируса, чем женщины).
- Возрастные изменения (но известно, что пожилые люди более уязвимы для вируса).
- Возможность вакцинации.
- Грузовые перевозки и пассажиропотоки.
- Индикатор социальной дистанции.
- Индивидуальные меры против распространения вируса.
- Вариабельность распространения вируса, которую можно учитывать в параметрах модели.

Параметры модели определяются для фиксированного региона с использованием информации о ежедневных подтвержденных и критических случаях, а также случаях смерти с учетом уровня изоляции Яндекса, количества протестированных и выявленных симптомных случаев в протестированных в регионах Российской Федерации.

Параметры SEIR-HCD математической модели

| Параметр | Описание | Границы |
|--------------------|---|-------------|
| $\alpha(t)$ | Индекс самоизоляции от Яндекса | (0, 5) |
| $\alpha_E(t)$ | Параметр заражения между бессимптомной и восприимчивой группами населения ($\alpha_E \gg \alpha_I$) | (0, 1) |
| $\alpha_I(t)$ | Параметр заражения между инфицированным и восприимчивым населением, который связан с контагиозностью вируса и социальными факторами | (0, 1) |
| β | Доля инфицированных, которая переносит заболевание без осложнений | (0, 1) |
| ε_{HC} | Доля госпитализированных случаев, которым требуется подключение ИВЛ | (0, 1) |
| μ | Доля смертельных случаев в результате COVID-19 | (0, 0.5) |
| τ | Латентный период (характеризует запаздывание выделения вирионов или заразности) | 2 дня |
| t_{inc} | Длительность инкубационного периода | 2-14 дней |
| t_{inf} | Длительность периода инфицирования | 2,5-14 дней |
| t_{hosp} | Длительность периода госпитализации | 4-5 дней |
| t_{crit} | Длительность использования аппарата ИВЛ | 10-20 дней |

Анализ модели SEIR-HCD распространения COVID-19

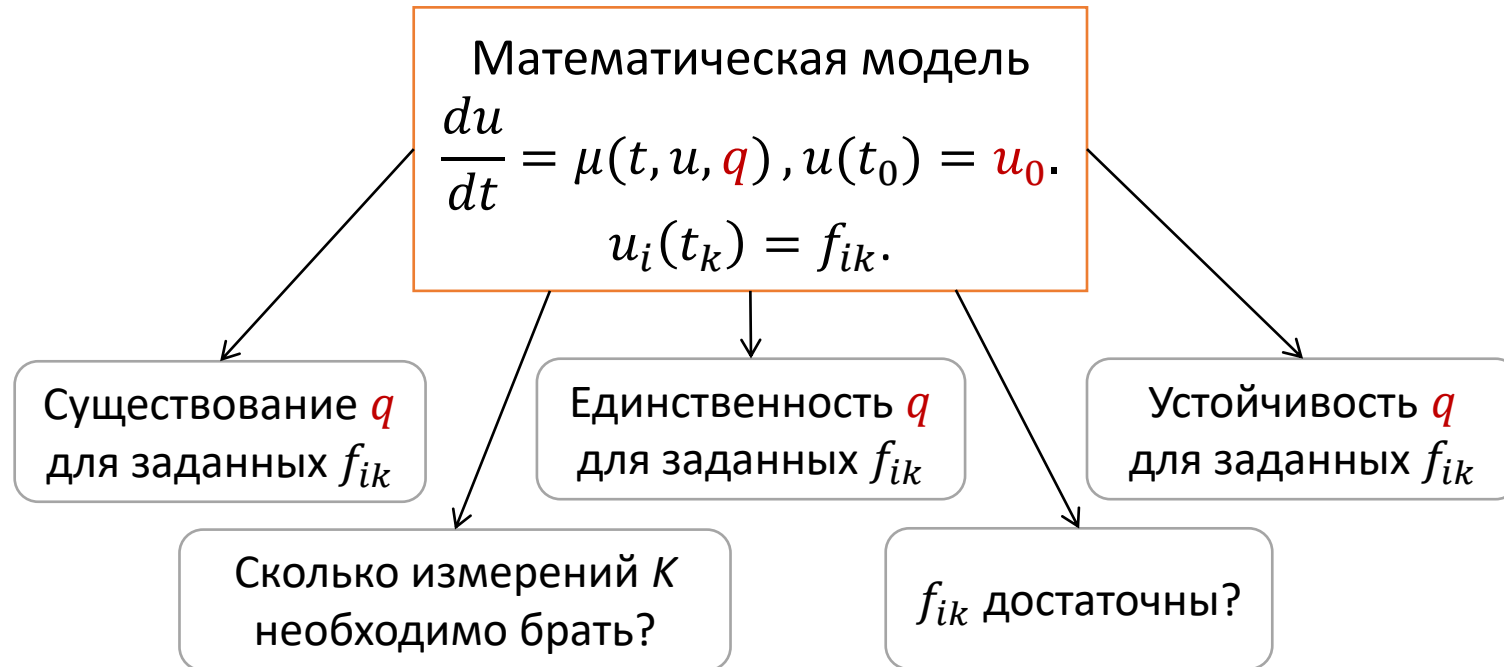
$$\begin{cases} \frac{dS}{dt} = -\frac{5 - a(t - \tau)}{5} \left(\alpha_E(t) \frac{S(t)E(t)}{N(t)} + \alpha_I(t) \frac{S(t)I(t)}{N(t)} \right), & S(0) = S_0, \\ \frac{dE}{dt} = \frac{5 - a(t - \tau)}{5} \left(\alpha_E(t) \frac{S(t)E(t)}{N(t)} + \alpha_I(t) \frac{S(t)I(t)}{N(t)} \right) - \frac{1}{t_{inc}} E(t), & E(0) = E_0, \\ \frac{dI}{dt} = \frac{1}{t_{inc}} E(t) - \frac{1}{t_{inf}} I(t), & I(0) = I_0, \\ \frac{dR}{dt} = \frac{\beta}{t_{inf}} I(t) + \frac{1 - \varepsilon_{HC}}{t_{hosp}} H(t), & R(0) = R_0, \\ \frac{dH}{dt} = \frac{1 - \beta}{t_{inf}} I(t) + \frac{1 - \mu}{t_{crit}} C(t) - \frac{1}{t_{hosp}} H(t), & H(0) = H_0, \\ \frac{dC}{dt} = \frac{\varepsilon_{HC}}{t_{hosp}} H(t) - \frac{1}{t_{crit}} C(t), & C(0) = C_0, \\ \frac{dD}{dt} = \frac{\mu}{t_{crit}} C(t), & D(0) = D_0. \end{cases}$$

Все входящие в систему уравнений константы неотрицательны; это непосредственно следует из их биологического смысла. При условии неотрицательности начальных данных решение задачи существует и единственно, а также является непрерывным и неотрицательным, т.е. $u(t) \geq 0$.

Система допускает единственное ненулевое стационарное решение:
 $S = N, E = 0, I = 0, R = 0,$
 $H = 0, C = 0, D = 0,$
 которое описывает ситуацию отсутствия инфицированных индивидуумов.

$$\mathcal{R}_0(t) = \frac{5 - a(t - \tau)}{5} \left(\alpha_E t_{inc} + \alpha_I t_{inf} \frac{1 + \varepsilon_{HC}(1 - \mu)}{1 - \varepsilon_{HC}(1 - \mu)} \right).$$

Что такое «идентифицируемость»?



Анализ идентифицируемости - это группа методов математической статистики, которые используются для определения того, насколько хорошо параметры модели оцениваются по количеству и качеству экспериментальных данных.

Идентифицируемость математических моделей для ОДУ

Структурная

- Зависит от структуры модели
- Измерения гладкие

Методы:

- Метод передаточной функции;
- Метод дифференциальной алгебры;
- Разложение в ряд Тейлора

Какой набор параметров является идентифицируемым по заданным измерениям (обратная задача имеет единственное решение)?

Анализ чувствительности

- Не использует реальные данные
- Необходима аппроксимация параметров
- Зависит от количества измерений

Методы:

- Метод собственных значений;
- Ортогональный метод



$$s_{ij}(t) = \frac{\partial \mu_i(t, u, q_{ex})}{\partial q_j}$$

Практическая

- Зависит от уровня шума в данных
- Необходима аппроксимация параметров

Методы:

- Метод Монте-Карло;
- Метод корреляционной матрицы

Какое оптимальное число измерений K необходимо для существования решения обратной задачи?

Анализ чувствительности SEIR-HCD модели

Модель ОДУ:

$$\frac{du}{dt} = \mu(t, u, q), u(t_0) = u_0.$$

Здесь $u = (S, E, I, R, H, C, D)^T$,
 $q = (\alpha_E, \alpha_I, \beta, \varepsilon_{HC}, \mu, E_0)^T$

Дифференцируем по q_j



$$\frac{d}{dt} \left(\frac{\partial u(t)}{\partial q} \right) = \frac{\partial \mu}{\partial u} \frac{\partial u(t)}{\partial q} + \frac{\partial \mu}{\partial q},$$

$$\frac{\partial u(0)}{\partial q} = \frac{\partial u_0}{\partial q}.$$

Матрица чувствительности:

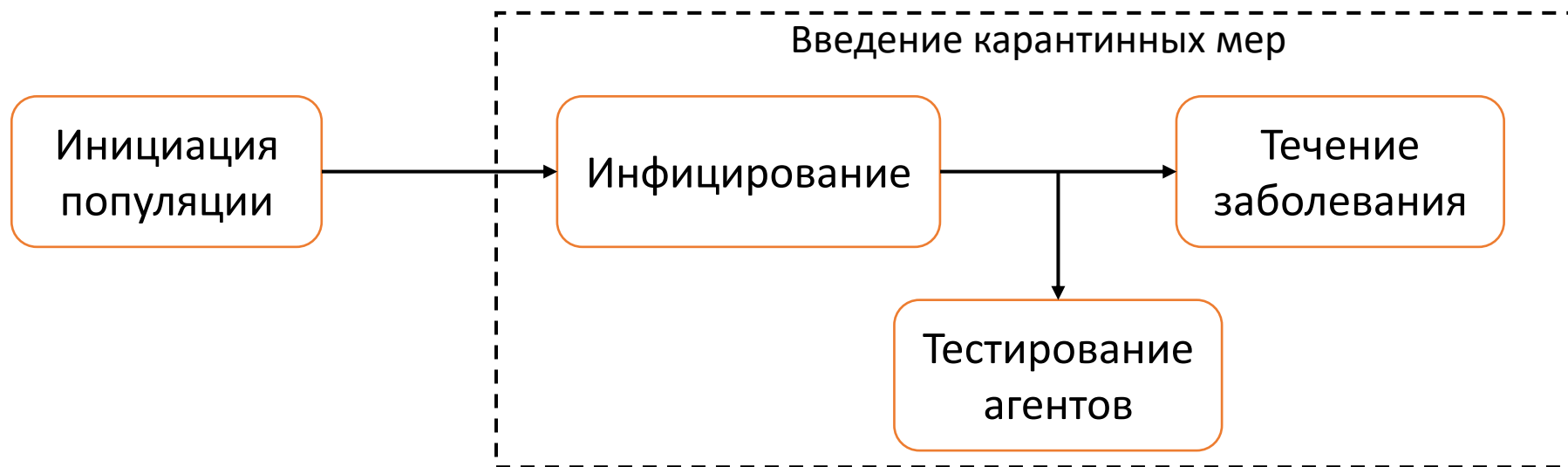
$$\frac{\partial y_l(t_k)}{\partial q_m}, l = 1, 2, 3; m = 1, \dots, 6; k = 1, \dots, K.$$

Здесь $y = (f_k, C_k, g_k)^T$ - данные о ежедневно выявленных, критических и умерших случаях.

| y_i | q_j | $\left\ \frac{\partial y_i(t)}{\partial q_j} q_j^* \right\ $ | y_i | q_j | $\left\ \frac{\partial y_i(t)}{\partial q_j} q_j^* \right\ $ | y_i | q_j | $\left\ \frac{\partial y_i(t)}{\partial q_j} q_j^* \right\ $ |
|-------|------------|---|-------|--------------------|---|-------|--------------------|---|
| f | α_E | $2.865 \cdot 10^{14}$ | g | α_E | $7.110 \cdot 10^{11}$ | g | μ | $2.982 \cdot 10^4$ |
| f | α_I | $2.396 \cdot 10^{14}$ | g | α_I | $5.947 \cdot 10^{11}$ | g | β | $2.164 \cdot 10^4$ |
| f | E_0 | $1.854 \cdot 10^{14}$ | g | E_0 | $4.601 \cdot 10^{11}$ | C | μ | $3.695 \cdot 10^2$ |
| C | α_E | $2.386 \cdot 10^{12}$ | C | ε_{HC} | $4.833 \cdot 10^4$ | f | β | $2.03 \cdot 10^{-6}$ |
| C | α_I | $1.996 \cdot 10^{12}$ | C | β | $3.428 \cdot 10^4$ | f | ε_{HC} | $1.6 \cdot 10^{-7}$ |
| C | E_0 | $1.544 \cdot 10^{12}$ | g | ε_{HC} | $3.041 \cdot 10^4$ | f | μ | $3 \cdot 10^{-8}$ |

С помощью ортогонального метода анализа матрицы чувствительности получена последовательность параметров от наиболее идентифицируемого до наименее:
 $\alpha_E, E_0, \alpha_I, \varepsilon_{HC}, \mu, \beta$

Агентная модель распространения COVID-19



| Person A | |
|----------------|-------|
| uid | 23928 |
| age | 55.1 |
| dead | 0 |
| susceptible | 1 |
| infected | 0 |
| diagnosed | 0 |
| ... | |
| date_infected | NaN |
| date_diagnosed | NaN |

...

| Person B | |
|----------------|-------|
| uid | 41135 |
| age | 13.5 |
| dead | 0 |
| susceptible | 0 |
| infected | 1 |
| diagnosed | 0 |
| ... | |
| date_infected | 44 |
| date_diagnosed | NaN |

...

| Person C | |
|----------------|-------|
| uid | 76851 |
| age | 83.2 |
| dead | 1 |
| susceptible | 0 |
| infected | 1 |
| diagnosed | 1 |
| ... | |
| date_infected | 46 |
| date_diagnosed | 53 |

Инициация популяции

Инициализируем искусственную популяцию согласно статистическим данным:

- Возрастная структура населения. Агенты разбиваются на возрастные группы по 10 лет [1];
- Структура домохозяйств (среднее количество человек в семье по данным ООН [2] – 3-5).

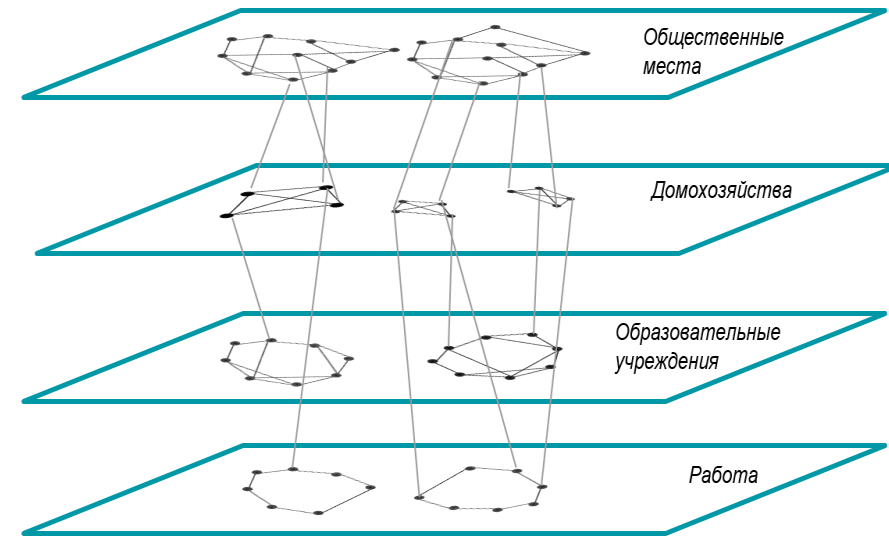
Создаем слои контактов и заполняем их агентами. В зависимости от возраста агенты контактируют друг с другом в контактных сетях, которые представлены **полносвязными** графами, количество вершин которых является пуассоновской случайной величиной с средним:

- для домохозяйства – *размер семьи*,
- для общественных мест и образовательных учреждений – 20,
- для работы – 8.

Все агенты имеют контакты в домохозяйствах и в общественных местах, агенты в возрасте 6-21 лет также могут контактировать в образовательных учреждениях с агентами своего возраста, агенты в возрасте 22-65 лет – на работе.

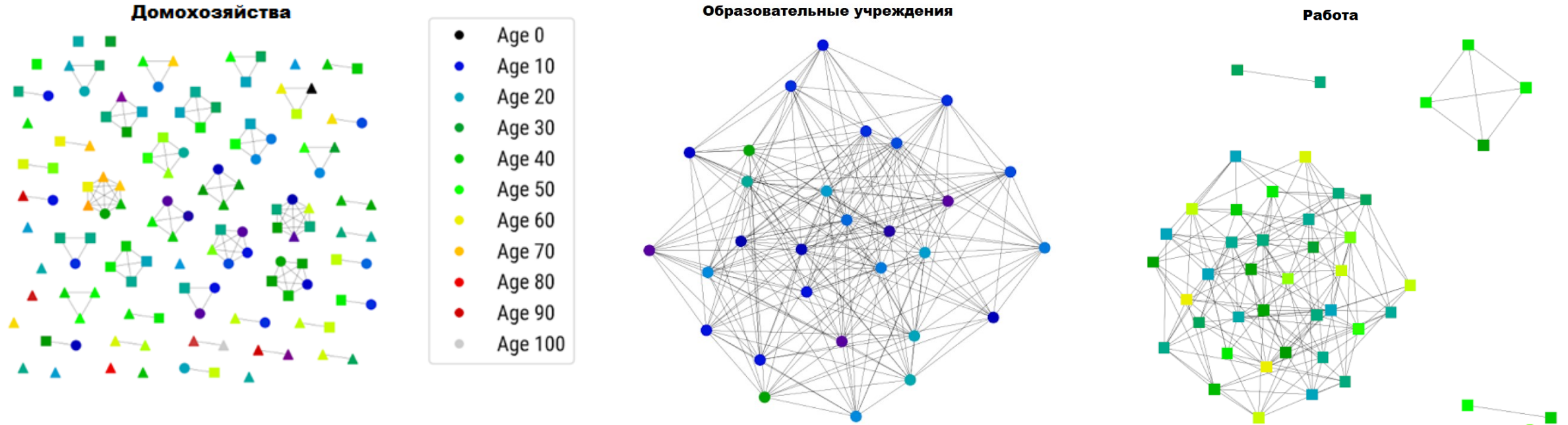
[1] Федеральная служба государственной статистики, Новосибирская обл. <https://novosibstat.gks.ru/folder/31729>

[2] Household Size, UN. <https://population.un.org/Household/\#/countries/840>



Слои контактов агентов и взаимосвязи между ними в модели распространения COVID-19.

Инициация популяции: контактные сети

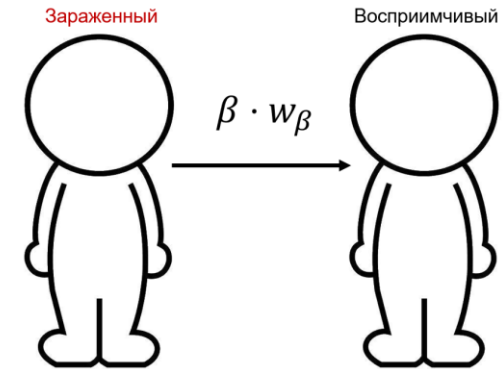
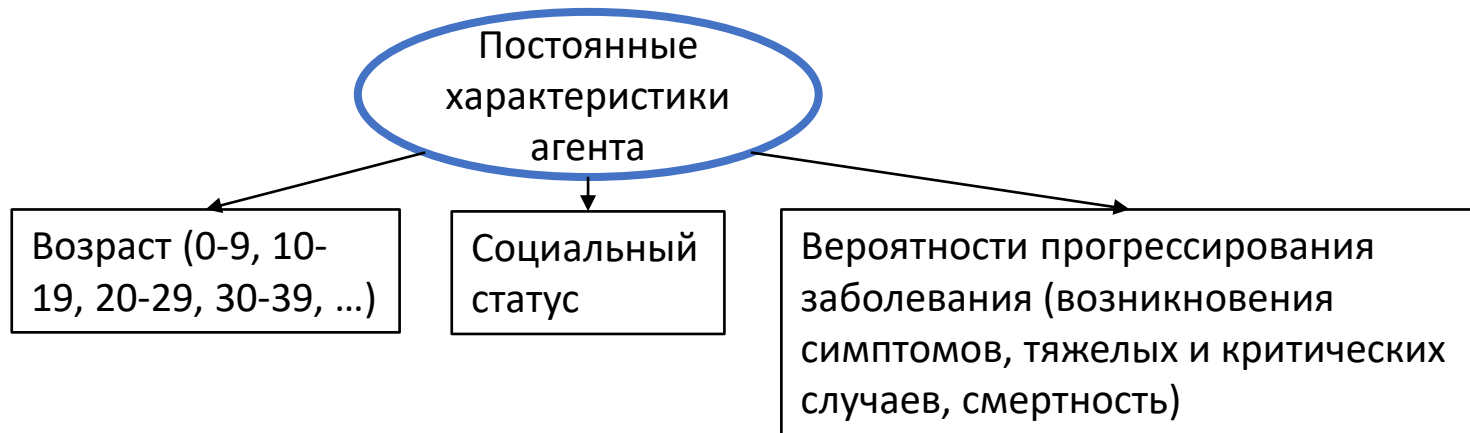


Фактические схемы связи для подвыборки 127 человек из 10 000. Все люди присутствуют в домохозяйствах, в том числе некоторые не имеют связи с домом. Как правило, эти люди, включая учителей, присутствуют в школьной сети (кружки); другое подмножество присутствует в сетях рабочих мест (квадраты); некоторые люди не входят ни в школьную, ни в рабочую сеть (треугольники).

Инфицирование

В рамках модели предполагается, что вирус передается между агентами, которые соединены между собой ребром в графе. За передачу вируса при близком контакте с зараженным отвечает параметр β и является кусочно-постоянным. В зависимости от структуры контакта, параметр β умножается на соответствующую константу (см. таблицу). Симптомные и бессимптомные агенты передают вирус одинаково.

Таким образом, вероятность передачи вируса для каждой контактной сети случайная и различная.



Вероятность заражения зависит от пребывания агента в семье, на работе, на учебе, в публичных местах.

| Структура | Вероятность заражения |
|----------------------------------|-----------------------|
| Домохозяйства | $\beta \cdot 3.0$ |
| Школы и университеты (6-22 года) | $\beta \cdot 0.6$ |
| Работа (22-65 лет) | $\beta \cdot 0.6$ |
| Публичные места | $\beta \cdot 0.3$ |

Течение заболевания

Каждый агент может находиться в 9 возможных состояниях заболевания. Переход из одной стадии заболевания в другую контролируется параметрами:

- $p_{sym} \in [0.5, 0.9]$ – вероятность проявлять симптомы после заражения: $p_{sym}[i]=0.5+0.05 \cdot i$, где i – это номер возрастной группы (0 для 0-9 лет, 1 для 10-19 лет и т.д.).
- $p_{sev} \in [0.0001, 0.184]$ – вероятность перехода больного с симптомами в тяжелое состояние (нуждается в госпитализации);
- $p_{crit} \in [0.00004, 0.10333]$ – вероятность перехода больного из тяжелого состояния в критическое (нуждается в реанимации и подключения аппарата ИВЛ);
- $p_{death} \in [0.00002, 0.093]$ – вероятность смерти для больного, находящегося в реанимации.

Данные параметры зависят от возрастной группы агента. Агенты старшего возраста более уязвимы.

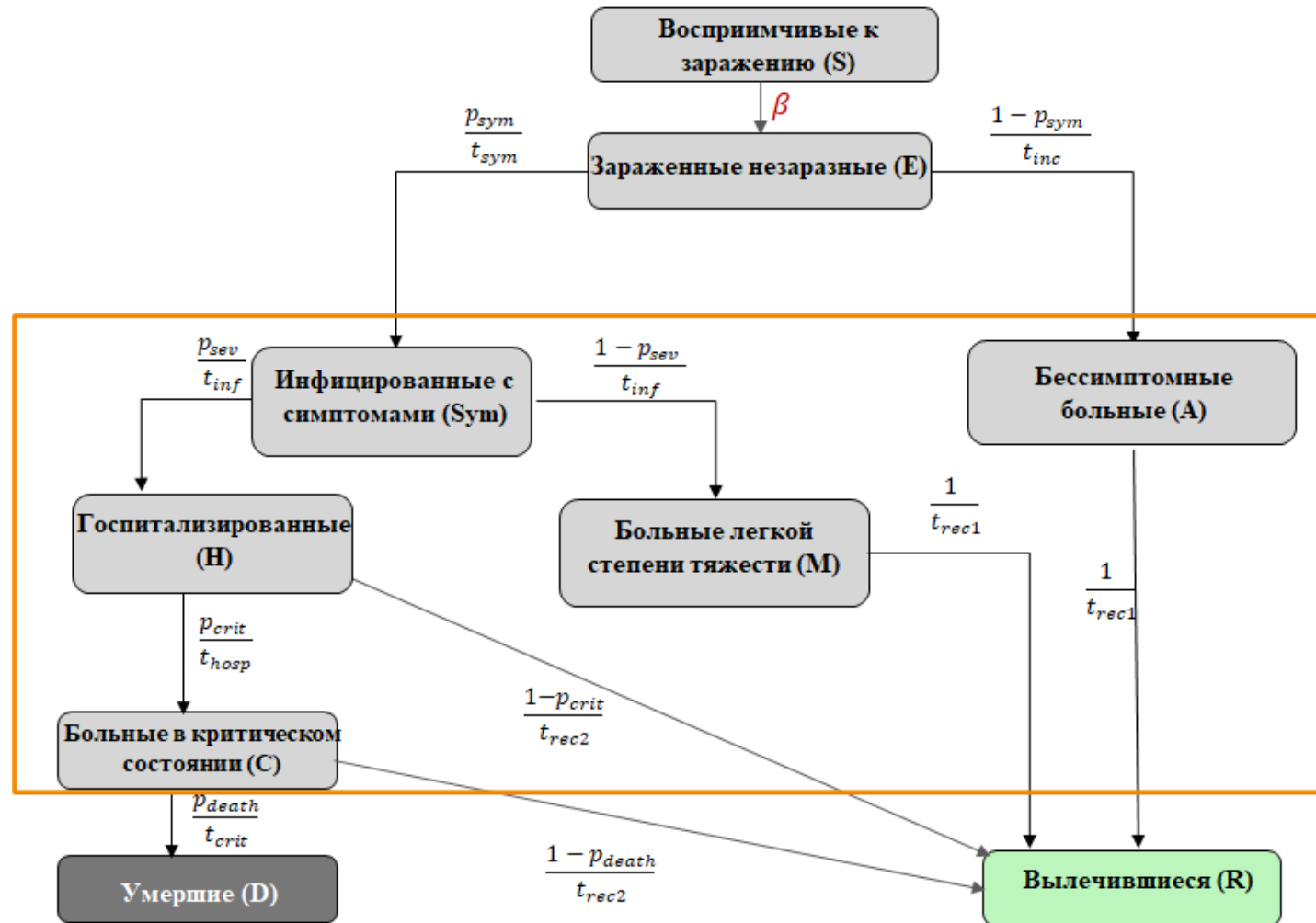
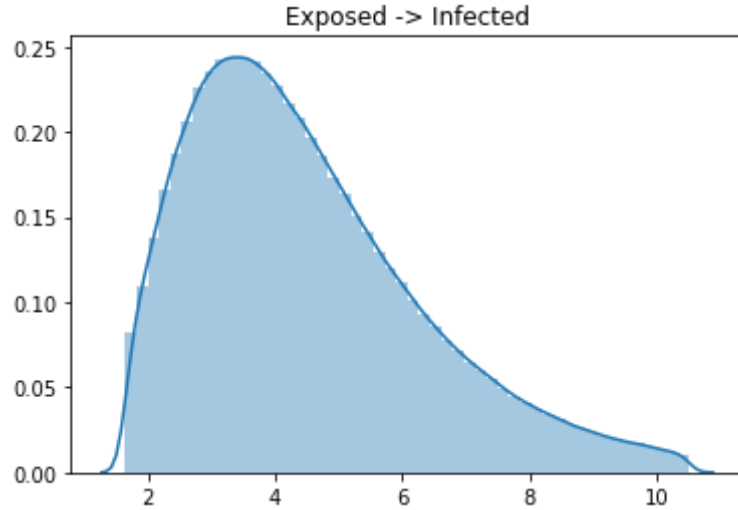


Диаграмма состояний агентов в модели Covasim. Оранжевой рамкой обозначены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19.

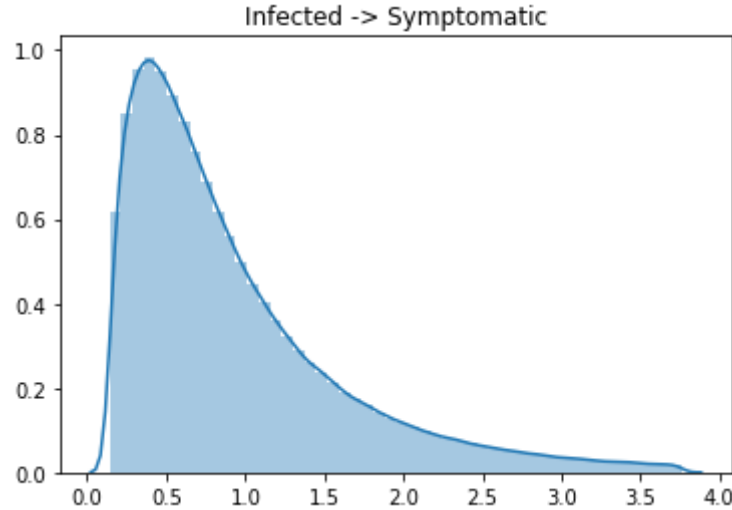
Течение заболевания

Длительность каждой стадии заболевания является случайной логнормальной величиной с фиксированными параметрами среднего и дисперсии:

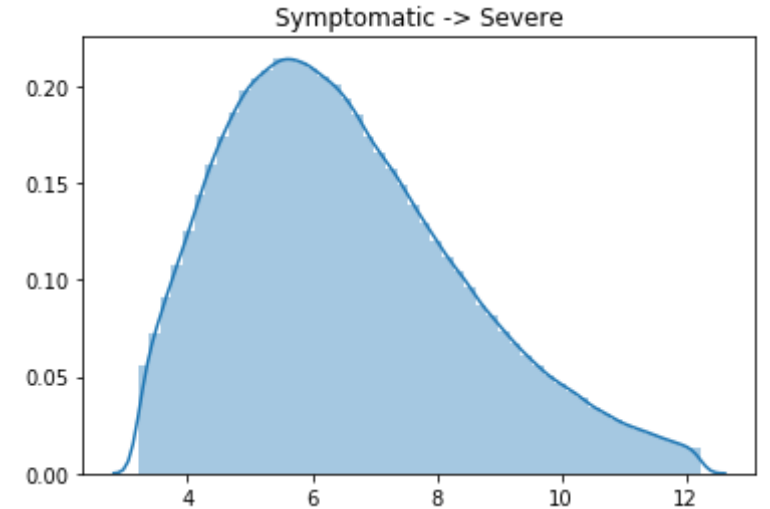
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2}$$



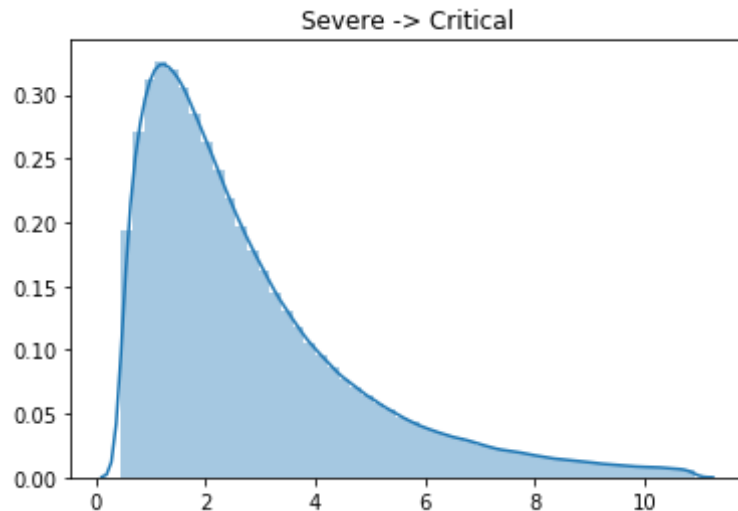
$$t_{inc} \sim \text{LogN}(4.6, 4.8)$$



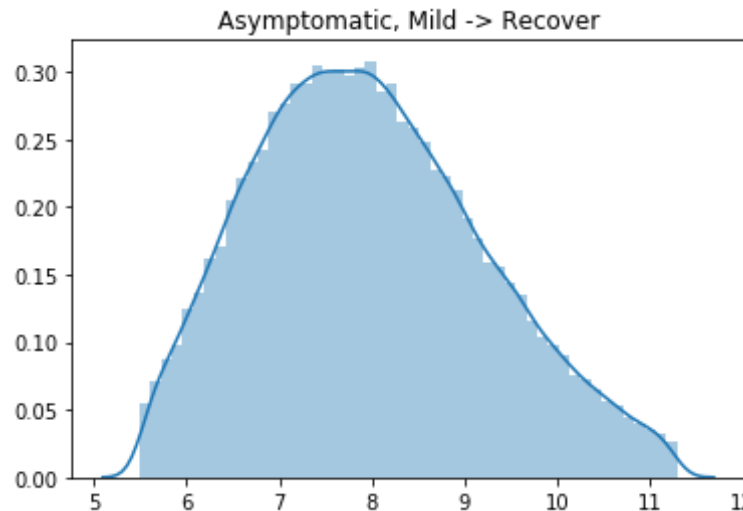
$$t_{sym} \sim \text{LogN}(1.0, 0.9)$$



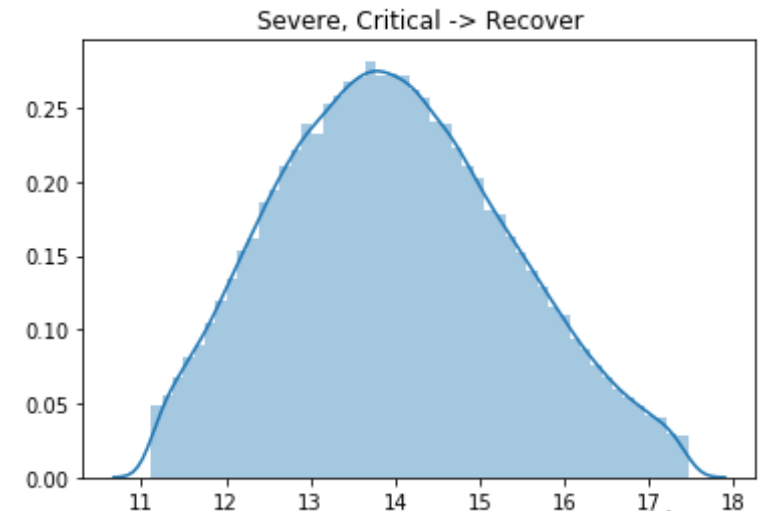
$$t_{inf} \sim \text{LogN}(6.6, 4.9)$$



$$t_{hosp} \sim \text{LogN}(3.0, 7.4)$$



$$t_{rec1} \sim \text{LogN}(8.0, 2.0)$$



$$t_{rec2} \sim \text{LogN}(14.0, 2.4)$$

Тестирование агентов

Тестирование проводится согласно ежедневным статистическим данным по количеству проведенных тестов в регионе. Шанс быть протестированным на COVID-19 $\tilde{p} = \frac{p_{sym}}{p_{asym}}$, зависит от эпидемиологического статуса агента. Здесь p_{sym} — вероятность тестирования для больных с симптомами, p_{asym} — вероятность тестирования для больных без симптомов.

Параметр \tilde{p} восстанавливается в ходе решения обратной задачи.

Положительный результат теста могут получить агенты, статус которых обведен в оранжевую рамку на схеме.

В модели предполагается, что вероятность тестирования агентов с симптомами выше, чем у бессимптомных больных. Данное соотношение шансов контролируется параметром \tilde{p} , который восстанавливается в ходе решения обратной задачи.

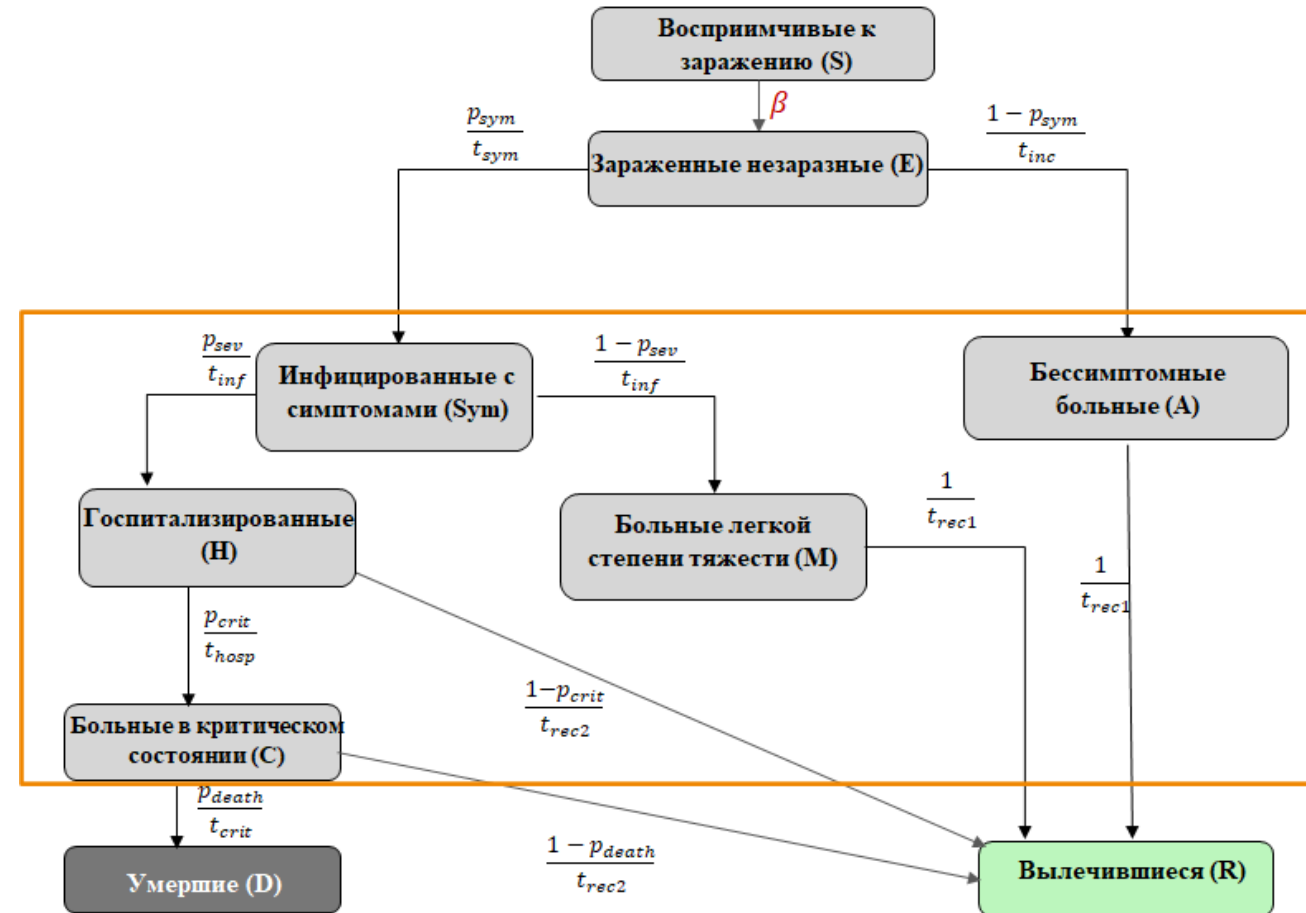


Диаграмма состояний агентов в модели Covasim. Оранжевой рамкой обозначены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19.

Введение карантинных мер

В рамках модели возможно ввести ограничительные меры глобально для всех слоев или для каждого по отдельности. Это можно сделать двумя способами:

- изменить значение параметра β (характеризующее уменьшение передачи вируса например при соблюдении масочного режима, соблюдения социальной дистанции);
- удалить ребра в графах (характеризующее самоизоляцию, переход на дистанционное обучение или работу).

В нашей модели используются оба этих подхода.

Предположения и ограничения модели

- Изначально иммунитета нет ни у одного агента;
- Есть начальная доля инфицированных в инкубационном состоянии, количество которых характеризуется параметром $E(0)$;
- Умереть могут только агенты, которые находятся в реанимации (больные в критическом состоянии);
- После выздоровления агент не может заразиться вновь в течение 4-5 месяцев.

Анализ чувствительности стохастических систем

Агентная модель является вычислительно трудоемкой

- Большое число входных параметров \mathbf{q} и выходных данных;
- Задача согласования нескольких видов данных (временных рядов) при одновременном изменении большого количества входных параметров может быть вычислительно трудоемкой.



Какие вопросы могут быть решены после проведения анализа идентифицируемости?

1. Задача идентификации параметров \mathbf{q} по экспериментальным данным \mathbf{z} может быть переопределена.
2. Если параметры вектора \mathbf{q} не идентифицируемы для некоторых измерений, то они могут быть идентифицируемы для другого набора данных.
3. Если имеется дополнительная статистическая информация о \mathbf{q} , а полученные в рамках модели оценки ведут к некорректным распределениям.

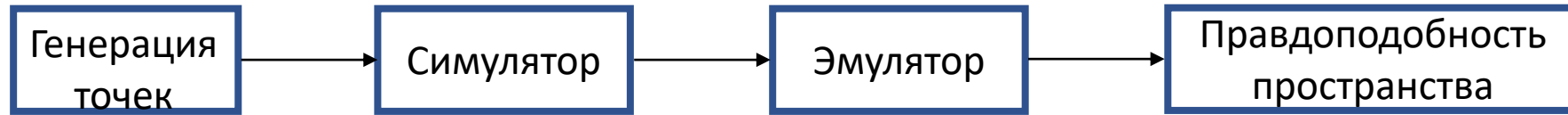
Основная идея

~~идентификация параметров агентной модели~~

Использование **простой** регрессионной модели для вариации и уточнения границ параметров \mathbf{q} .

$$g_i(\mathbf{q}) = \sum_{j=1}^k h_j(\mathbf{q})b_j + u(\mathbf{q})$$

Анализ чувствительности математических моделей стохастических систем: алгоритм



Генерация точек. Задаем $q_i \in (a_i, b_i)$, $i = 1, \dots, p$, и генерируем $n = 10p$ точек для обучения и $n_0 = p$ для валидации модели, используя Latin hypercube.

Симулятор. Решаем прямые задачи для агентной модели при n вариантах параметров и получаем выходные данные $f_i(\mathbf{q})$, $i = 1, \dots, n$.

Эмулятор. Аппроксимируем выходные данные $f_i(\mathbf{q})$ проверочной моделью (emulator):

$$g_i(\mathbf{q}) = \sum_{j=1}^k h_j(\mathbf{q}) b_j + u(\mathbf{q}).$$

Здесь $Hb = b_0 + b_1 \mathbf{q} + b_2 \mathbf{q}^2 + b_3 \mathbf{q}^3$, вектор b неизвестен, $u(\mathbf{q}) \sim \mathcal{N}(0, \sigma^2 c(\mathbf{q}, \mathbf{q}'))$ с корреляционной функцией Матерна $c(\mathbf{q}, \mathbf{q}') = (1 + \sqrt{3}W)e^{-\sqrt{3}W}$, W – взвешенное расстояние между \mathbf{q} и \mathbf{q}' .

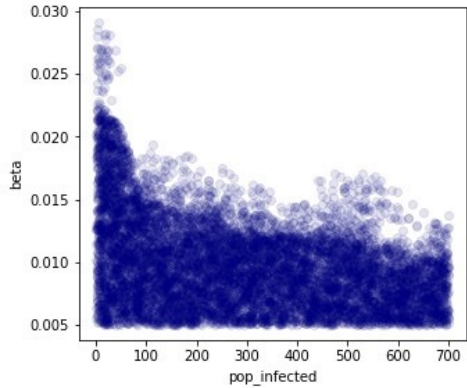
Обучение проверочной модели: максимизируем функционал правдоподобия $\mathcal{L}(g_i(\mathbf{q})|\theta)$, используя градиентный метод *L-BFGS-B*.

Правдоподобность пространства. Вычисляем меру неправдоподобности

$$I_i(\mathbf{q}) = \frac{|z_i - E(f_i(\mathbf{q}))|}{\sqrt{\text{Var}[z_i - E(f_i(\mathbf{q}))]}}.$$

Выбираем $I_i(\mathbf{q}) > c$, где параметр c выбирается по правилу «3 сигм» (Pukelsheim's 3σ rule).

Анализ чувствительности для агентной модели распространения COVID-19

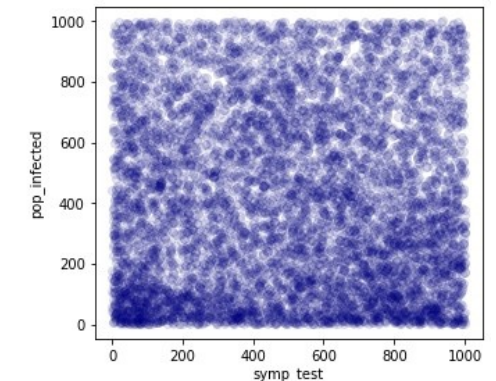
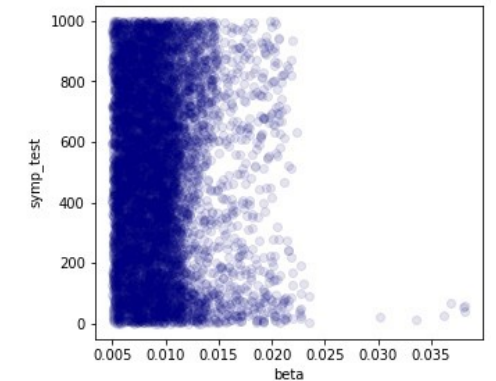
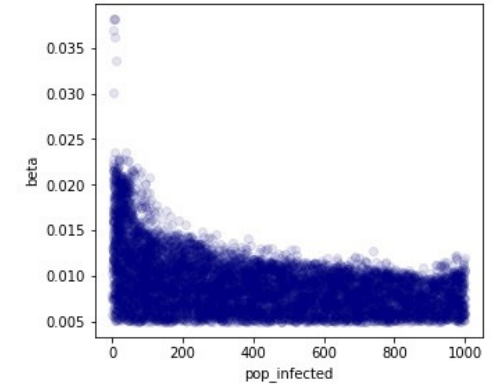
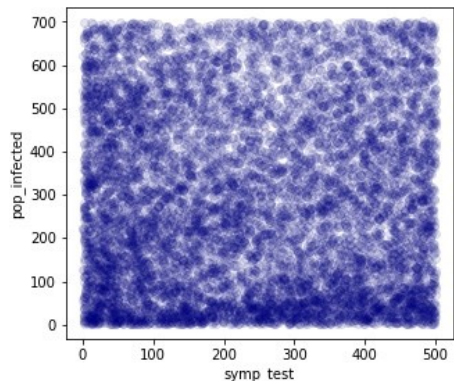
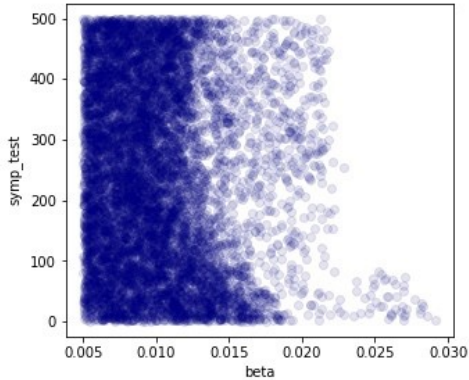


$$\mathbf{q} = (\beta, E(0), \tilde{p})$$
$$\mathbf{z} = (\text{New_diagnosed}, \text{New_deaths})$$

Параметр контагиозности β (скорость передачи инфекции) очень чувствителен к дополнительным измерениям \mathbf{z} .

Если мы добавим информацию о критических случаях, мы получим более точные границы изменения параметра β . Другие параметры исходных инфекционных случаев и количество симптоматических тестов не зависели от вариаций данных.

$$\bar{\mathbf{z}} = (\text{New_diagnosed}, \text{New_deaths}, \text{Critical_cases})$$



Идентификация неизвестных параметров путем минимизации целевого функционала

$$J(q) = \sum_{k=1}^K \sum_s w_s \frac{|c_d^{i,s} - c_m^{i,s}|}{\max_t c_d^{s,t}} \rightarrow \min_q$$

Алгоритм оптимизации: OPTUNA hyperparameter optimization framework <https://optuna.org/>

В основе лежит метод древовидных оценок Парзена*. Также в алгоритме используются:

- эволюционные методы (генетический алгоритм, дифференциальной эволюции),
- градиентные (квазиньютоновские, стохастические),
- тензорной оптимизации и
- методы случайного поиска.

$q = (E(0), \beta, \beta_d, \beta_c, t_{sym})$ – вектор неизвестных параметров для каждого участка моделирования, где β_d - день измерения параметра β , β_c - значение, на которое изменится параметр β , t_{sym} - параметр уровня тестирования.

s – статистики, по которым наблюдаются данные (количество диагностированных, протестированных, умерших),

$c_d^{i,s}, c_m^{i,s}$ - значения данных и модели,

w_s - весовой параметр для каждой статистики.

Методы глобальной оптимизации



Градиентные (локальные) методы



$$q^* = \operatorname{argmin}_{q \in Q} J(q)$$

* Parzen E. On estimation of a probability density function and mode, *Ann. Math. Stat.*, **33**, 1065-1076 (1962).

Природоподобные алгоритмы

Методы глобальной оптимизации

Природоподобные алгоритмы

- Генетический алгоритм
- Метод имитации отжига
- Метод роя частиц
- Метод дифференциальной эволюции

Полуаналитические методы

- Тензорная оптимизация

Методы машинного обучения

- Стохастический градиентный спуск
- Метод опорных векторов
- Линейная\логистическая регрессия
- Нейронные сети

Преимущества:

- ✓ Определяет область глобального экстремума
- ✓ Не чувствителен к начальному приближению
- ✓ Быстро сходится к области «точного» решения
- ✓ Простая структура (имитирует простейшую задачу физики, биологии)

Ограничения:

- Стохастическая сходимость
- Настройка параметров методов трудоемкая

Стохастическая сходимость:

- А.А. Жиглявский, А.Г. Жилинскас. Методы поиска глобального экстремума. М.: Наука, 1991:

Теорема. Пусть $J(q)$ – непрерывная функция в области A , для которой существует точка глобального минимума q^* , и пусть для любого $\varepsilon > 0$ выполнены следующие условия:

$$\sum_{k=1}^{\infty} q_k(\varepsilon) = \infty, \quad \text{где } q_k(\varepsilon) = \inf_{Q_i, 1 \leq i \leq k} P_k(B(\varepsilon)),$$

где P_k – распределение вектора q' .

Тогда последовательность случайных векторов q_1, q_2, \dots , полученная в результате применения природоподобного алгоритма (генетическим, имитацией отжига, роя частиц, дифференциальной эволюции), попадает в множество $A(\varepsilon)$ с вероятностью 1 почти наверное.

Градиентные методы: выражение градиента целевого функционала

ОДУ

$$\frac{du}{dt} = \mu(t, u, q), u(t_0) = u_0 \quad + \quad u_i(t_k) = f_{ik}$$

$$(q_{n+1}, u_{0_{n+1}}) = (q_n, u_{0_n}) - \alpha_n J'(q_n, u_{0_n}), \quad \alpha_n > 0$$

$$J(q, u_0) = \sum_{k=1}^K \sum_i \omega_i (u_i(t_k; q, u_0) - f_{ik})^2 \rightarrow \min_{q, u_0}$$

$$J'(q, u_0) = \left(- \int_{t_0}^T \Psi^T(t) \mu_q(u(t), q) dt, -\Psi(t_0) \right)$$

$$\begin{cases} \dot{\Psi} = -\mu_u^T(u(t), q)\Psi, \\ \Psi(T) = 0, \\ [\Psi_i]_{t=t_k} = 2\omega_i (u_i(t_k; q) - f_{ik}), k = 1, \dots, K, \\ i \in I \subset \{1, \dots, N\}. \end{cases}$$

УЧПР

$$\begin{cases} \frac{\partial u}{\partial t} = d \frac{\partial^2 u}{\partial x^2} + g(u, q), \\ u(x, t_0) = u_0(x) \end{cases} \quad + \quad u(x_i, t_k) = f_{ik}$$

$$J(q, u_0) = \sum_{i \in I} \sum_{k=0}^K \omega_i |u(x_i, t_k; q, u_0) - f_{ik}|^2 \rightarrow \min_{q, u_0}$$

$$J'(q, u_0) = - \left(\int_{t_0}^T \int_{\Omega_l} \Psi^T \mu_q dt dx, -\Psi(x, t_0) \right)^T$$

$$\begin{cases} \frac{\partial \Psi}{\partial t} = -\mu_u^T(u(x, t), q) - d \frac{\partial^2 \Psi}{\partial x^2}, t > 0, x \in \Omega, \\ \Psi(T) = 0, \frac{\partial \Psi}{\partial x} \Big|_{\partial \Omega_l} = 0, \\ [\Psi_i]_{t=t_k} = 2\omega_i (u_i(x, t_k; q) - f_{ik}), k = 1, \dots, K, l = 1, \dots, L. \end{cases}$$

[1] S.I. Kabanikhin, O.I. Krivorotko. Identification of biological models described by systems of nonlinear differential equations. J. Inverse Ill-Posed Probl. 2015. V. 23(5). P. 519-527.

[2] O. Krivorotko, S. Kabanikhin, Sh. Zhang, V. Kashtanova. Global and local optimization in identification of parabolic systems. J. Inverse Ill-Posed Probl. 2020. V. 28(6). P. 899-913

Тензорная оптимизация

$$J(q) \rightarrow g_\alpha(q)$$

$$g_\alpha(q) = h(J(q) - \alpha)$$

$h(a)$ – непрерывная монотонно убывающая функция, α – некоторое начальное приближение к минимуму функционала $J(q)$.

Исходная задача минимизации функционала $J(q)$, $q \in \mathbb{R}^d$ заменяется на эквивалентную задачу поиска максимума по модулю функции g .

Алгоритм:

1. Вводится сетка с n узлами по каждому из d направлений, и тензор значений функции g на этой сетке обозначается через $A(i_1, \dots, i_d)$.
2. Затем составляются подматрицы $A(R_k, C_k)$ матриц разверток $A_k(i_1 \dots i_k, i_{k+1} \dots i_d) = A(i_1, \dots, i_d)$. Они приближаются с помощью матричного крестового метода.
3. Далее вычисляются пространственные координаты точек, соответствующих элементам матриц.
4. Найденные точки локально оптимизируются и добавляются в множество локальных оптимумов M .
5. Определяются ближайшие элементы матриц к элементам этого множества и строковые и столбцовые номера этих элементов записываются в множества R_k и C_k соответственно.
6. Обновляется сдвиг функционала $\alpha = \min_{q \in M} J(q)$.
7. Повторяем эти действия, пока не выполнится критерий остановки.

Тензорная оптимизация: зависимость времени вычислений от разбиения вектора параметров

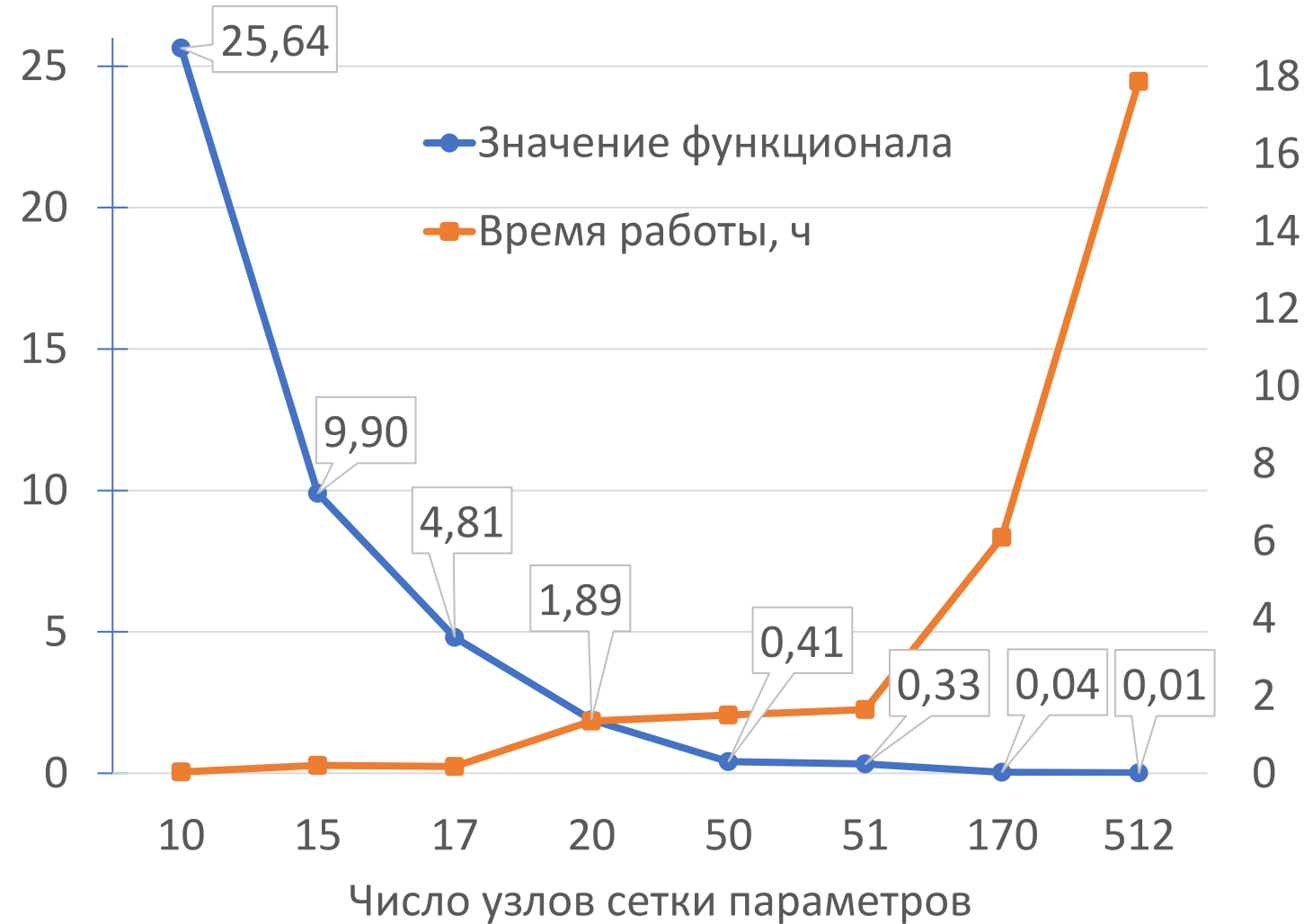
Вычислительный узел ССКЦ:

2 процессора Intel Xeon Gold 6248R (3 ГГц, 24 ядра)

384 Гбайт основной оперативной памяти

Вычисления производились на 2 процессорах при прочих равных условиях.

При увеличении числа узлов сетки параметров повышается точность восстановления параметров и увеличивается время вычислений.



Алгоритм моделирования и построения сценариев распространения COVID-19

1. Сбор и обработка данных по количеству ежедневно выявленных, протестированных, умерших, критических и госпитализированных случаев. Ежедневно обновляемые данные по Новосибирской области и близлежащим регионам расположены на сайте: <http://covid19-modeling.ru/>
2. Уточнение границ неизвестных параметров методами анализа чувствительности.
3. Решение задачи идентификации параметров комбинацией методов глобальной и локальной минимизации целевого функционала.
4. Для построения сценариев развития COVID-19:
 - Экстраполируем заданные временные ряды методами машинного обучения (индекс самоизоляции от Яндекса, количество ожидаемых тестов в регионе);
 - Для уточненных параметров и ограничительных мер решаем прямую задачу.
5. Построение доверительных интервалов (для агентных моделей)
 - Решаем прямую задачу с идентифицированными параметрами 10 раз;
 - Для каждого дня считаем квантили уровня 0.1, 0.5, 0.9;
 - Получаем 3 массива точек с шагом один день, по которым строим медианное значение (квантиль уровня 0.5, сплошная линия на графиках ниже), верхнюю и нижнюю границу (закрашенная область).

Mean field games – объединение 2х подходов моделирования

Для больших однородных и хорошо перемешанных популяций SIR-модели можно интерпретировать как приближение среднего поля вероятностной модели клеточного автомата.



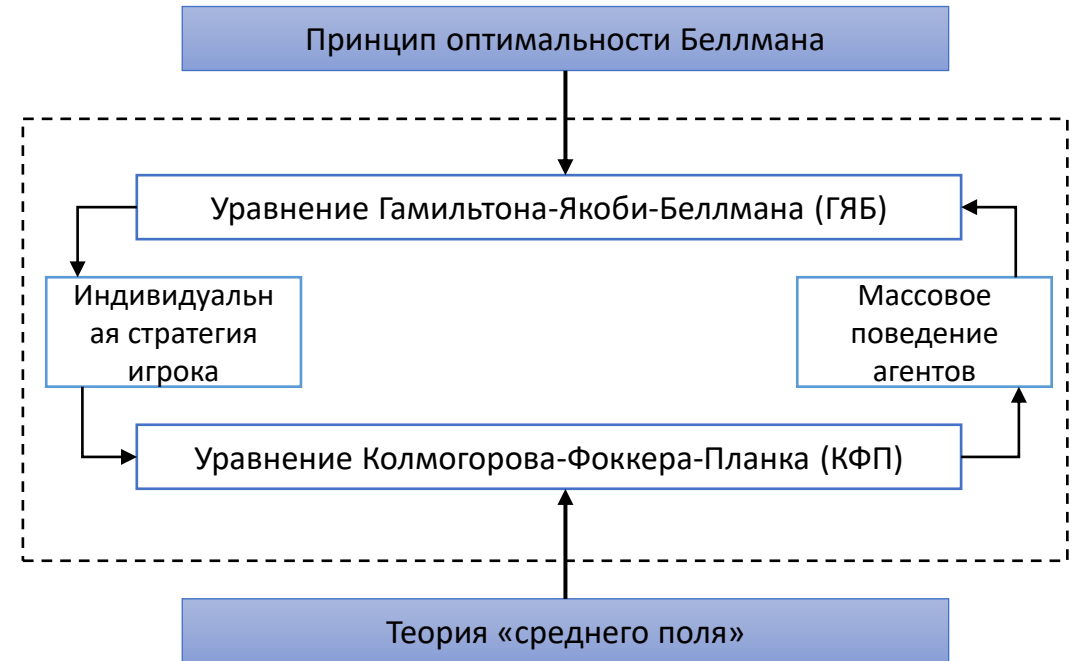
$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N}, & S(0) = N - I_0, \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, & I(0) = I_0 < N, \\ \frac{dR}{dt} = \gamma I, & R(0) = 0. \end{cases}$$

КФП

Введем плотность распределения агентов $m_i: [0, T] \times [0, 1] \rightarrow \mathbb{R}$, где $i \in \{S, I, R\}$, $x \in [0, 1]$ обозначает склонность населения соблюдать карантинные меры (0 – агент склонен соблюдать социальную дистанцию, 1 – не склонен).

$$\begin{cases} \frac{\partial m_S}{\partial t} + \nabla(m_S \alpha_S) + \frac{\beta m_I m_S}{N} - \frac{\sigma_S^2 \Delta m_S}{2} = 0, & m_S(0, x) = m_{S0}(x), \\ \frac{\partial m_I}{\partial t} + \nabla(m_I \alpha_I) - \frac{\beta m_I m_S}{N} + \gamma m_I - \frac{\sigma_I^2 \Delta m_I}{2} = 0, & m_I(0, x) = m_{I0}(x), \\ \frac{\partial m_R}{\partial t} + \nabla(m_R \alpha_R) - \gamma m_I - \frac{\sigma_R^2 \Delta m_R}{2} = 0, & m_R(0, x) = m_{R0}(x). \end{cases}$$

$$\frac{\partial m_i}{\partial x} = 0, x = 0, 1.$$



H. Tembine. COVID-19: Data-Driven Mean-Field-Type Game Perspective. Games. 11, 51 (2020).

Mean field games – объединение 2х подходов моделирования

$$\begin{cases} \frac{\partial m_S}{\partial t} + \nabla(m_S \alpha_S) + \frac{\beta m_I m_S}{N} - \frac{\sigma_S^2 \Delta m_S}{2} = 0, & m_S(0, x) = m_{S0}(x), \\ \frac{\partial m_I}{\partial t} + \nabla(m_I \alpha_I) - \frac{\beta m_I m_S}{N} + \gamma m_I - \frac{\sigma_I^2 \Delta m_I}{2} = 0, & m_I(0, x) = m_{I0}(x), \\ \frac{\partial m_R}{\partial t} + \nabla(m_R \alpha_R) - \gamma m_I - \frac{\sigma_R^2 \Delta m_R}{2} = 0, & m_R(0, x) = m_{R0}(x). \end{cases}$$

$$\frac{\partial m_i}{\partial x} = 0, x = 0, 1.$$

КФП

В предположении рациональности агентов набор стратегий $\alpha_i(t, x)$ доставляет минимум функционалу

$$J(m_{SIR}, \alpha_{SIR}) = \int_0^T \int_0^1 \sum_{i \in \{S, I, R\}} (F_i(\alpha_{SIR}, t, x) m_i + g_i(t, x, m_{SIR})) dx dt + \int_0^1 \frac{m_I^2(T, x)}{2} dx$$

$$\begin{cases} \frac{\partial v_S}{\partial t} + \frac{\sigma_S^2 \Delta v_S}{2} + \alpha_S \frac{\partial v_S}{\partial x} + \beta m_I (v_I - v_S) = -F_S - \frac{\partial g_S}{\partial m_S}, \\ \frac{\partial v_I}{\partial t} + \frac{\sigma_I^2 \Delta v_I}{2} + \alpha_I \frac{\partial v_I}{\partial x} + \beta m_S (v_I - v_S) - \gamma v_I = -F_I - \frac{\partial g_I}{\partial m_I} - \delta(T - t) m_I, \\ \frac{\partial v_R}{\partial t} + \frac{\sigma_R^2 \Delta v_R}{2} + \alpha_R \frac{\partial v_R}{\partial x} = -F_R - \frac{\partial g_R}{\partial m_R}. \end{cases}$$

ГЯБ

$$v_i(T, x) = 0,$$

$$\frac{\partial m_i}{\partial x} = 0, x = 0, 1.$$

Методы, используемые при построении алгоритма решения обратных задач

Обработка статистических данных

- Методы регрессионного анализа для обработки данных.
- Методы машинного обучения (нейронные сети, линейная регрессия) для экстраполяции данных с целью построения прогноза.



Получено свидетельство о государственной регистрации программы для ЭВМ №2021614740 от 29.03.2021

Анализ идентифицируемости

- Методы дифференциальной алгебры и разложения в ряды Тейлора анализа структурной идентифицируемости.
- Ортогональный метод и метод сингулярного разложения получения упорядоченной последовательности чувствительных параметров для систем ОДУ.
- Методы Монте-Карло и корреляционной матрицы анализа практической идентифицируемости.
- Байесовский подход к анализу чувствительности стохастических систем.



Использовано программное обеспечение DAISY (<https://daisy.dei.unipd.it/>)



Получено свидетельство о государственной регистрации программы для ЭВМ №2019661851 от 10.09.2019



Использована и адаптирована библиотека <https://github.com/alan-turing-institute/mogp-emulator>

Получены последовательности идентифицируемых параметров и оценки их областей для моделей внутриклеточной динамики ВИЧ, ко-инфекции туберкулеза и ВИЧ, распространения COVID-19.

Построение агентной модели

- Метод случайных графов с учетом особенностей COVID-19



Улучшена библиотека COVASIM <https://github.com/InstituteforDiseaseModeling/covasim>
Свидетельство №2021614740 от 29.03.2021

Методы, используемые в представленных работах

Методы решения задач минимизации

Глобальные

- Метод тензорной оптимизации (использована библиотека <https://github.com/oseledets/TT-Toolbox>).
- Генетический алгоритм, методы дифференциальной эволюции, имитации отжига, роя частиц (использована библиотека *SciPy.Optimize Python*).



Для модели эпидемии туберкулеза с возникновением лекарственно-устойчивых штаммов получено внедрение в учебный процесс ЕНУ им. Л.Н. Гумилева в 2020 году.

Модель внутриклеточной динамики ВИЧ-инфекции зарегистрирована в ФАП СО РАН 24.12.2018 (№ PR18006)

Локальные

- Градиентные методы.
- Метод Нелдера-Мида (использована библиотека *SciPy.Optimize Python*).



Получено выражение для градиента целевого функционала в общем случае.

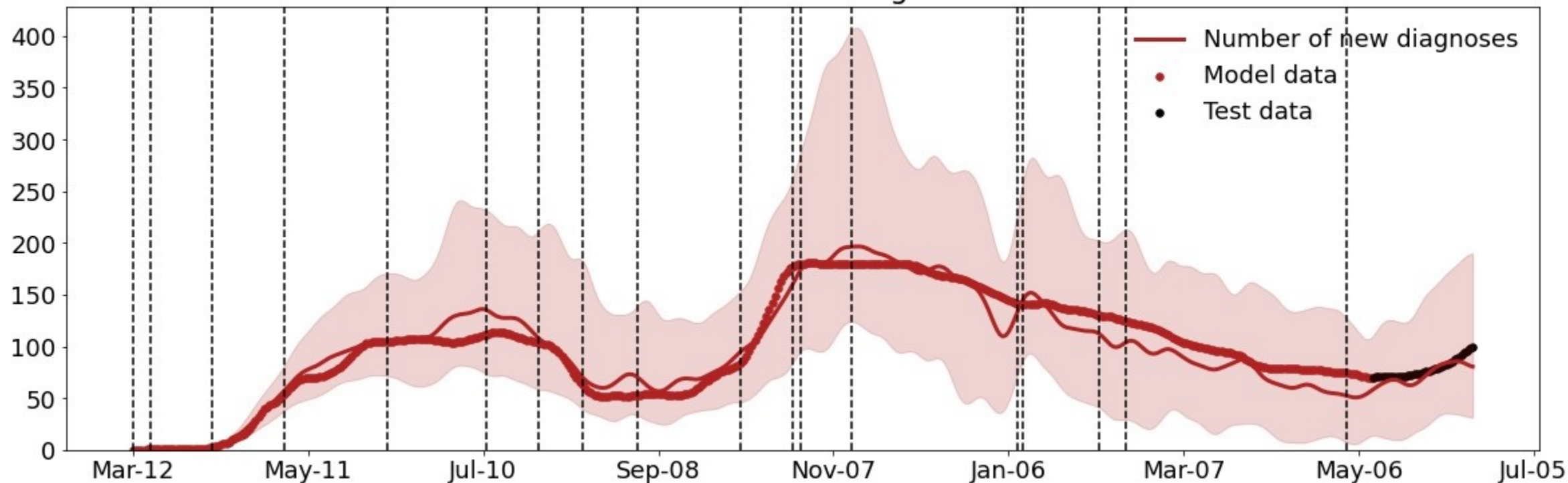
Получена оценка остаточного члена в разложении производное Фреше целевого функционала.

Карантинные меры в Новосибирской области

| Дата | Карантинные меры |
|-------------------|---|
| 18 марта 2020 | Университеты начали удаленные занятия. Каникулы в школах с 18.03 по 01.04. |
| 28 марта 2020 | Приостановлены все массовые, развлекательные, общественные мероприятия |
| 31 марта 2020 | Усилены меры против коронавируса (социальная дистанция, самоизоляция). |
| 27 апреля 2020 | Указ губернатора об обязательном ношении масок в магазинах. |
| 12 мая 2020 | Возобновление работы продовольственных рынков во всех населенных пунктах, кроме областного центра. Разрешено заниматься спортом на улице, соблюдая социальную дистанцию. Разрешены прогулки без посещения общественных мест, детских площадок группам не более двух человек или семьям, живущим вместе. |
| 8 июня 2020 | Возобновление работы некоторых непродовольственных магазинов. |
| 6 июля 2020 | Открытие летних веранд в кафе и ресторанах. |
| 15 июля 2020 | Открытие торговых центров. |
| 10 августа 2020 | Открытие бассейнов и ресторанов. |
| 1 сентября 2020 | Начало учебного года (школы, университеты). |
| 23 октября 2020 | Школьные каникулы и возобновление удаленного режима в учебных заведениях. |
| 31.12.21-10-01.21 | Новогодние праздники |
| 01.05.21-10.05.21 | Майские праздники |

Верификация агентной модели распространения COVID-19 в Новосибирской области с 12.03.2020 по 21.06.2021: ежедневно выявленные случаи

Number of new diagnoses

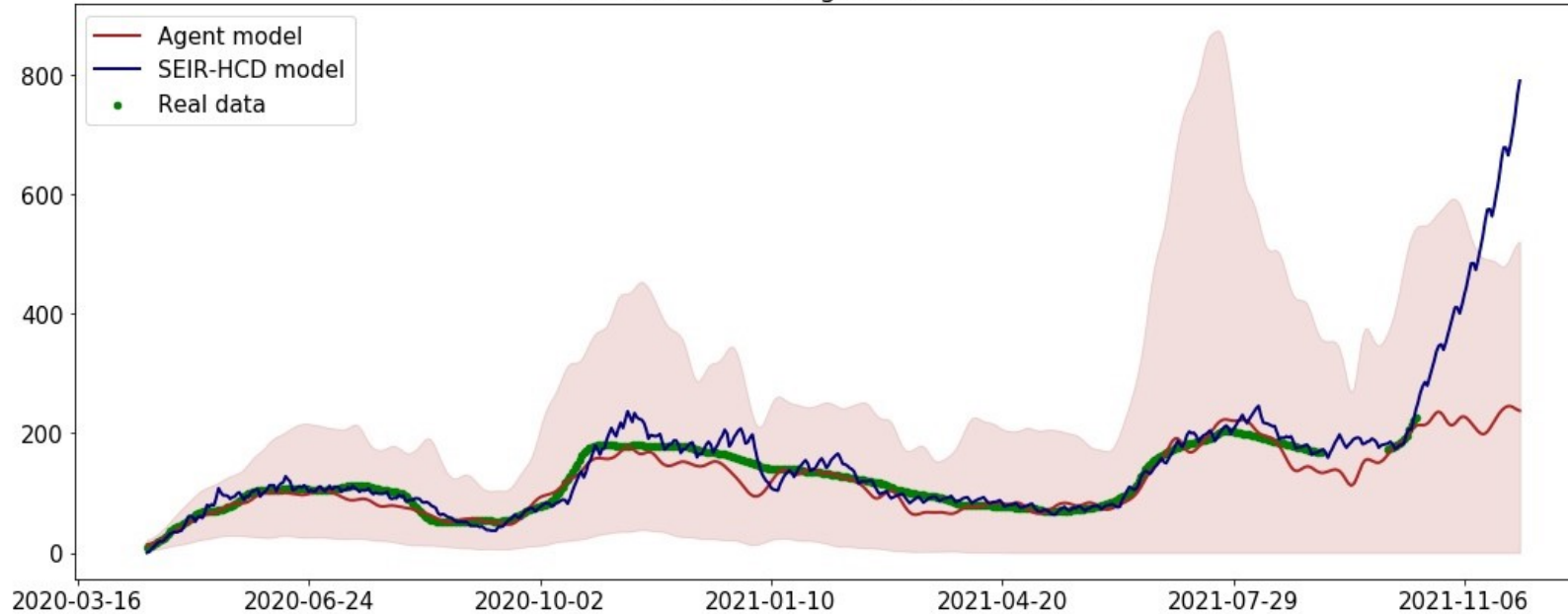


Моделирование распространения ежедневно выявленных случаев COVID-19 в Новосибирской области с 12 марта 2020 по 21 июня 2021 (**красная тонкая линия**). Реальные данные с 12.03.2020 по 10.05.2021, используемые при решении обратной задачи, обозначены **красной жирной линией**. Реальные данные с 11.05.2021 по 21.06.2021, обозначенные **черной жирной линией**, использовались для верификации модели.

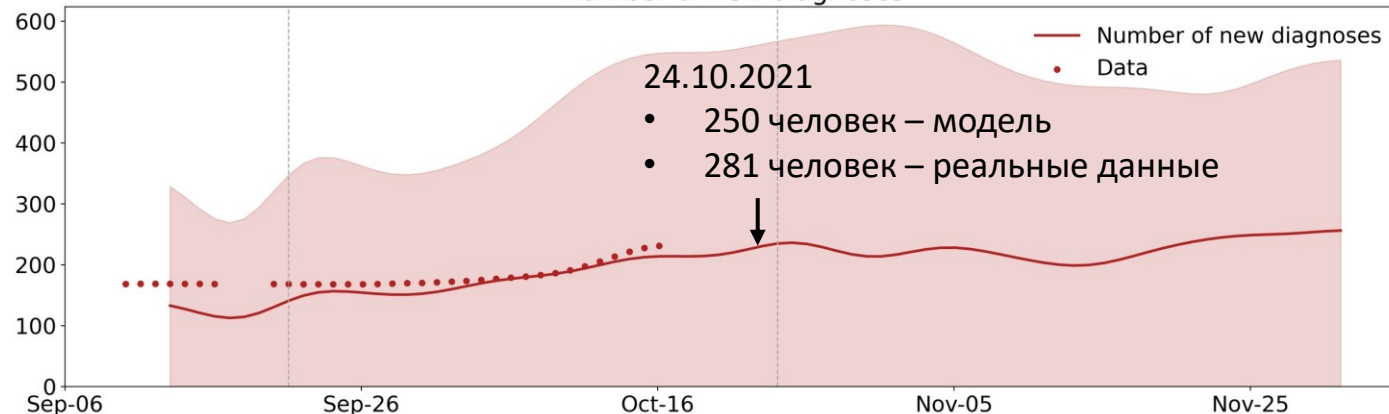
Вертикальные пунктирные линии – время изменения параметра contagiозности, закрашенная красная область - 90% доверительный интервал.

Распространение COVID-19 в Новосибирской области с 12.03.2020 по 30.11.2021: ежедневно выявленные случаи (на момент расчетов 16.10.2021)

Number of new diagnoses (Novosibirsk)



Number of new diagnoses



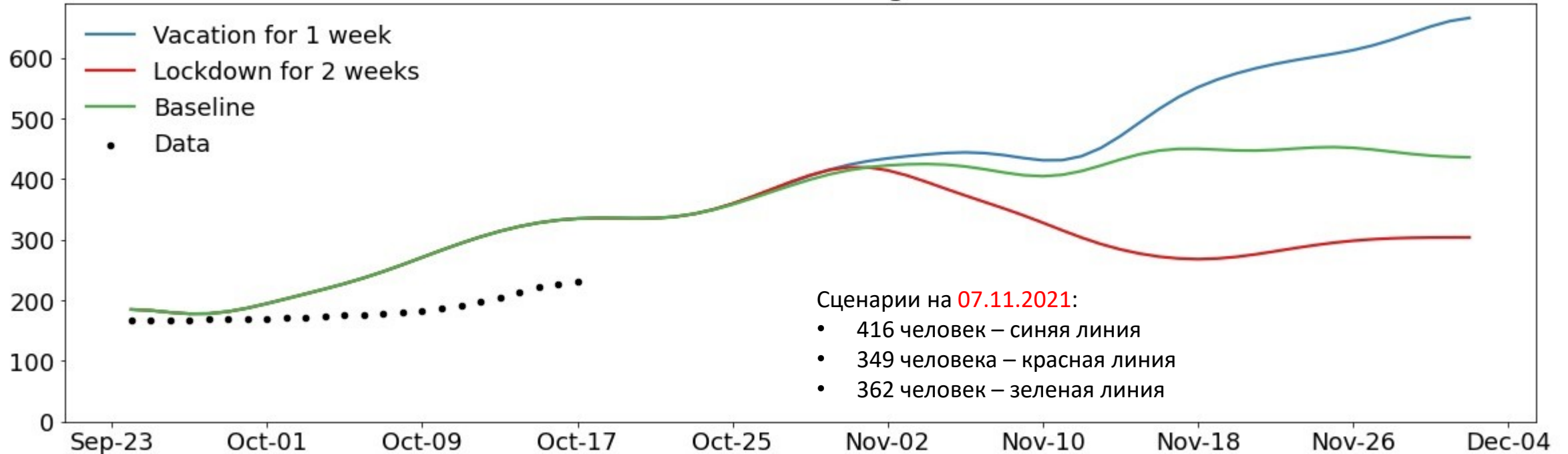
Моделирование распространения ежедневно выявленных случаев COVID-19 в Новосибирской области с 12 марта 2020 по 30 ноября 2021 в рамках агентной модели (**красная линия**) и SEIR-HCD модели (**синяя линия**). Реальные данные с 12.03.2020 по 16.10.2021, используемые при решении обратной задачи, обозначены **зеленой жирной линией**. Прогноз сделан на 45 дней с 17.10 по 30.11. Вертикальные пунктирные линии – время изменения параметра контагиозности, закрашенная красная область - 80% доверительный интервал.

Нижний график является приближением верхнего на даты 15.09.2021 – 30.11.2021.

При сохранении карантинных мер к 01.11.2021 в НСО ожидается 217 выявленных случаев: <http://covid19-modeling.ru>

Распространение COVID-19 в Новосибирской области с 17.10.2021 по 30.11.2021: сценарии распространения ежедневно выявленных случаев

Number of new diagnoses



Синяя линия – объявить нерабочие дни с 30.10 по 07.11 (в течение которых уменьшается на 40% людей на работе и учебе, потом увеличивается заболеваемость из-за привезенных случаев в 2 раза).

Красная линия – устроить локдаун на 2 недели до 07.11 (полностью закрыть школы, 50% общественных мест и 50% рабочих на перевести на удаленную работу).

Зеленая линия – базовый сценарий развития на дату 17.10.2021 с учетом уменьшения количества людей в общественных местах на 40% с 25.10.2021.

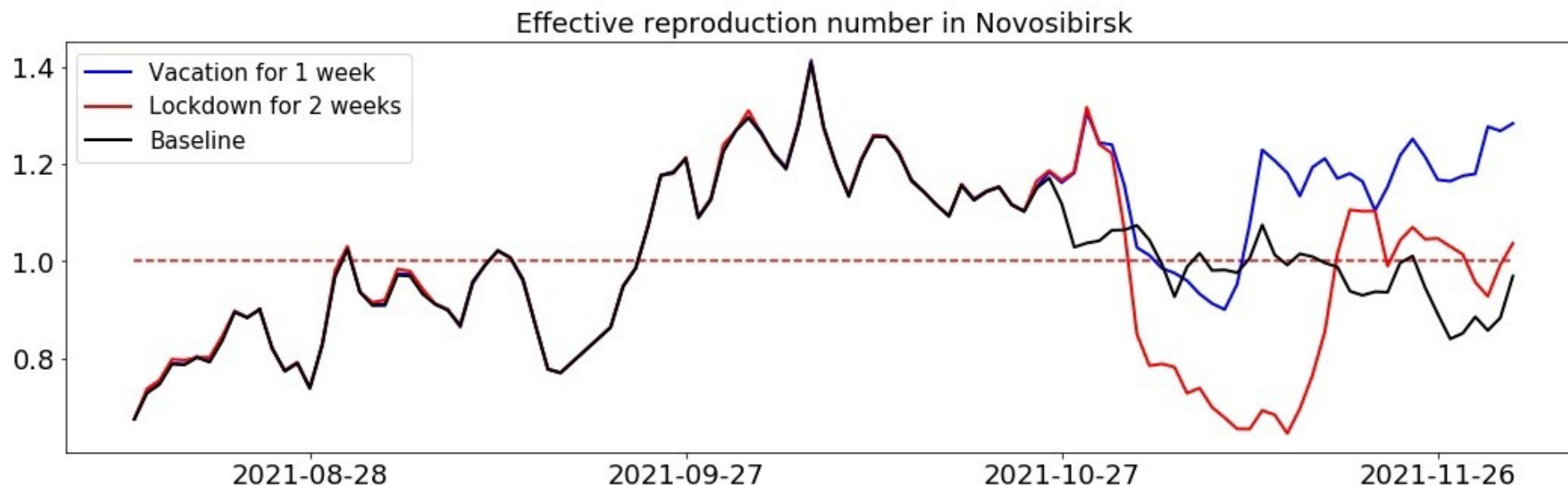
Черные точки – реальные данные выявленных случаев до 17.10.2021 (www.стопкоронавирус.рф).

Индекс репродукции вируса для Новосибирской области (от 17.10.2021)

Основным показателем распространения эпидемии является значение числа репродукции $\mathcal{R}_0(t)$, который описывает, сколько в среднем людей заражает инфицированный человек в течение своей болезни:

$$\mathcal{R}_0(t) = \frac{I_N(t) \cdot d}{I_C(t)}.$$

Здесь $I_N(t)$ - количество новых случаев инфицирования в день t , $I_C(t)$ - количество людей с активным инфекционным заболеванием в день t , d – среднее время инфицирования.



Базовый сценарий развития (**черная линия**) предполагает небольшой рост выявленных случаев до середины ноября, а затем стабилизацию ситуации. Закрытие учебных заведений и 50% общественных и рабочих мест (**красная линия**) до 07.11 приведет к существенному снижению количества выявленных случаев, рост которых начнется только в 20х числах ноября. Объявление нерабочих дней с 30.10 по 07.11 (**синяя линия**) приведет к существенному увеличению числа выявленных COVID-19 до 30.11.2021.

Используемые ресурсы ССКЦ

Решаемые задачи на языке программирования Python:

- Решение обратной задачи для агентной модели COVID-19,
- Моделирование социальных процессов.

Используемые пакеты:

- Numba (для jit и cuda)
- Optuna (для оптимизационных задач)
- Covasim (для агентного моделирования)
- Tensor-train (для тензорного разложения)

Недостающие ресурсы и возможности

1. Возможность одновременного использования ресурсов для разных вычислительных задач.
2. Использование современных графических ускорителей (GPU) для оптимизации процессов вычислений.
3. Возможность размещения сайта <http://covid19-modeling.ru/> на вычислительном узле (необходимо выделить IP адрес для подключения домена).

Эффективность работы на примере решения обратной задачи распространения COVID-19 по данным за 4 месяца:

| Подзадача | Время, с | |
|-----------------------------|-------------|-----------|
| | На кластере | На ПК |
| Прямая задача | 10.057 | 23.154 |
| Обратная задача по 1 месяцу | 770.075 | 18010.759 |

Вычислительный узел: 2 процессора Intel Xeon Gold 6248R (3 ГГц, 24 ядра)
ПК: Intel Core I5 с 4 потоками.

Достигнутое ускорение: 2.3
Эффективность ускорения на ядро: 0.19

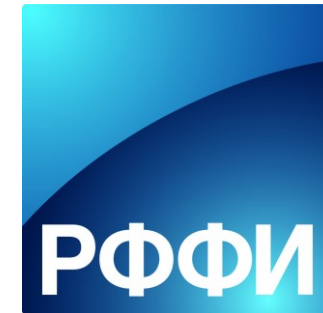
Основные результаты

1. Построены и проанализированы численные алгоритмы решения задач идентификации математических моделей распространения новой коронавирусной инфекции в Новосибирской области (и в других регионах РФ, США, Великобритании)
 - на основе обыкновенных дифференциальных уравнений (SIR-модели),
 - агентные модели,
 - на основе игры «среднего поля».
2. На основе алгоритмов идентифицируемости выявлены построены последовательности чувствительных параметров к ошибкам в измерениях, а также уточнены границы изменения чувствительных параметров.
3. Разработан комплекс программ моделирования и построения сценариев распространения COVID-19 в Новосибирской области (сайт: <https://covid19-modeling.ru/>) в зависимости от ограничительных мер, сдерживающих эпидемию. А также
 - проведена адаптация алгоритмов на параллельные технологии, включая суперкомпьютерные вычисления и графические ускорители,
 - комплекс программ можно адаптировать для моделирования вирусно-бактериальной инфекции (ВИЧ, туберкулез, ОРВИ и др.).

Комбинация математических моделей более точно описывает динамику эпидемии в зависимости от влияния ограничительных мер. Сценарии развития используются в Правительстве Новосибирской области для принятия оперативных решений по сдерживанию распространения эпидемии COVID-19.



Российский
научный фонд



Спасибо за внимание!



Ольга Криворотько
krivorotko.olya@mail.ru
+7-983-303-3083

Результаты моделирования
<http://covid19-modeling.ru/>



Метод древовидных оценок Парзена (Tree Parzen Estimator)

$$q^* = \operatorname{argmin}_{q \in Q} J(q)$$

Его идея напоминает Байесовский оптимизатор, однако в отличие от моделирования $p(J|q)$ (как в Байесовском случае), TPE вычисляет $p(q|J)$ и $p(J)$.

Алгоритм TPE:

1. Выбираем J^* как квартиль γ (например квартиль Q3), т.е. $p(J < J^*) = \gamma$. Основываясь на этом факте моделируется $p(J)$.
2. Затем все остальные значения разделяются на 2 группы: те, которые доставляют значение функционала меньше $J^*(q^i)$ и остальные. Таким образом строятся плотности вероятности значений q^i – это $l(q)$ и плотности остальных значений - $g(q)$. Тогда получаем вероятность попадания в одну из групп:

$$p(q|J) = \begin{cases} l(q), & J < J^*, \\ g(q), & J \geq J^*. \end{cases}$$

3. Используя формулу полной вероятности, задача оптимизации сводится к задаче максимизации $l(q)/g(q)$.

Сходимость: $E \ln \frac{|q_k - q^*|}{|q_{k-1} - q^*|} \rightarrow -c < 0, k \rightarrow \infty$ (в среднем расстояние до q^* уменьшается на постоянную e^{-c}).

1. Bergstra J.S, Bardenet R., Bengio Y., Kegl B. Algorithms for hyper-parameter optimization. *In Advances in Neural Information Processing Systems*. 2011. P. 2546-2554.

2. Madrigal F., Maurice C., Lerasle F. Hyper-parameter optimization tools comparison for multiple object tracking applications. *Mach. Vision Appl.* Vol. 30, no. 2, 2019. P. 269–289. <https://doi.org/10.1007/s00138-018-0984-1>.