

Конференция «Женщины в математике»
27-28 мая 2023

*Обратные задачи в эпидемиологии и
идентифицируемость*

к.ф.-м.н. Криворотько О.И.

Новосибирск, 2023

Содержание

1. Подходы к построению моделей в эпидемиологии.
2. Математические модели распространения COVID-19.
 - SEIR-HCD модель распространения COVID-19 в регионах РФ.
 - Агентная модель.
3. Идентифицируемость математических моделей.
 - Дифференциальная модель.
 - Агентная модель.
4. Методы оптимизации.
5. Численные результаты моделирования COVID-19.
 - Данные и их обработка.
 - Верификация моделей для Новосибирской области.
 - Сценарии распространения COVID-19 в Новосибирской области.
 - Сравнение с моделью нейронных сетей.
 - Великобритания и Новосибирская области – сравнение результатов.



Математические модели эпидемиологии: два основных подхода

Камерные модели

Моделирование сверху-вниз

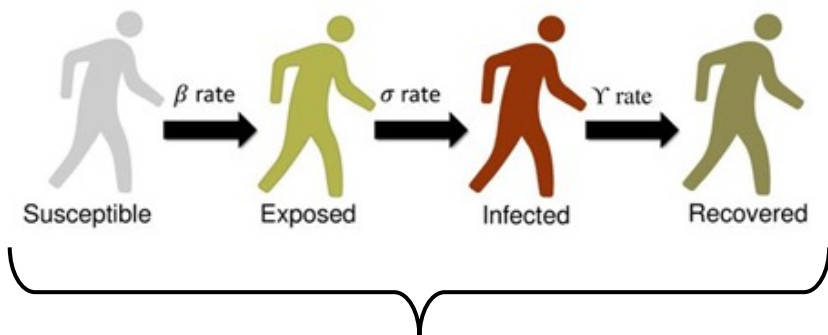
Детерминистские

Стохастические

Популяция разделена на группы со схожими признаками

SEIR Model

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

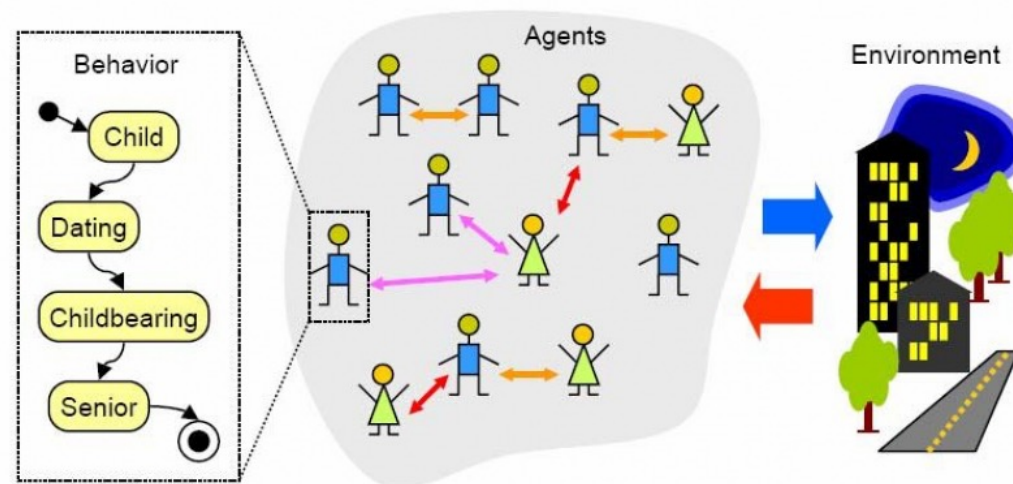


Закон действующих масс:

$$N = S(t) + E(t) + I(t) + R(t)$$
$$\frac{dN}{dt} = 0 \quad \forall t > 0.$$

Агентные модели

Моделирование снизу-вверх



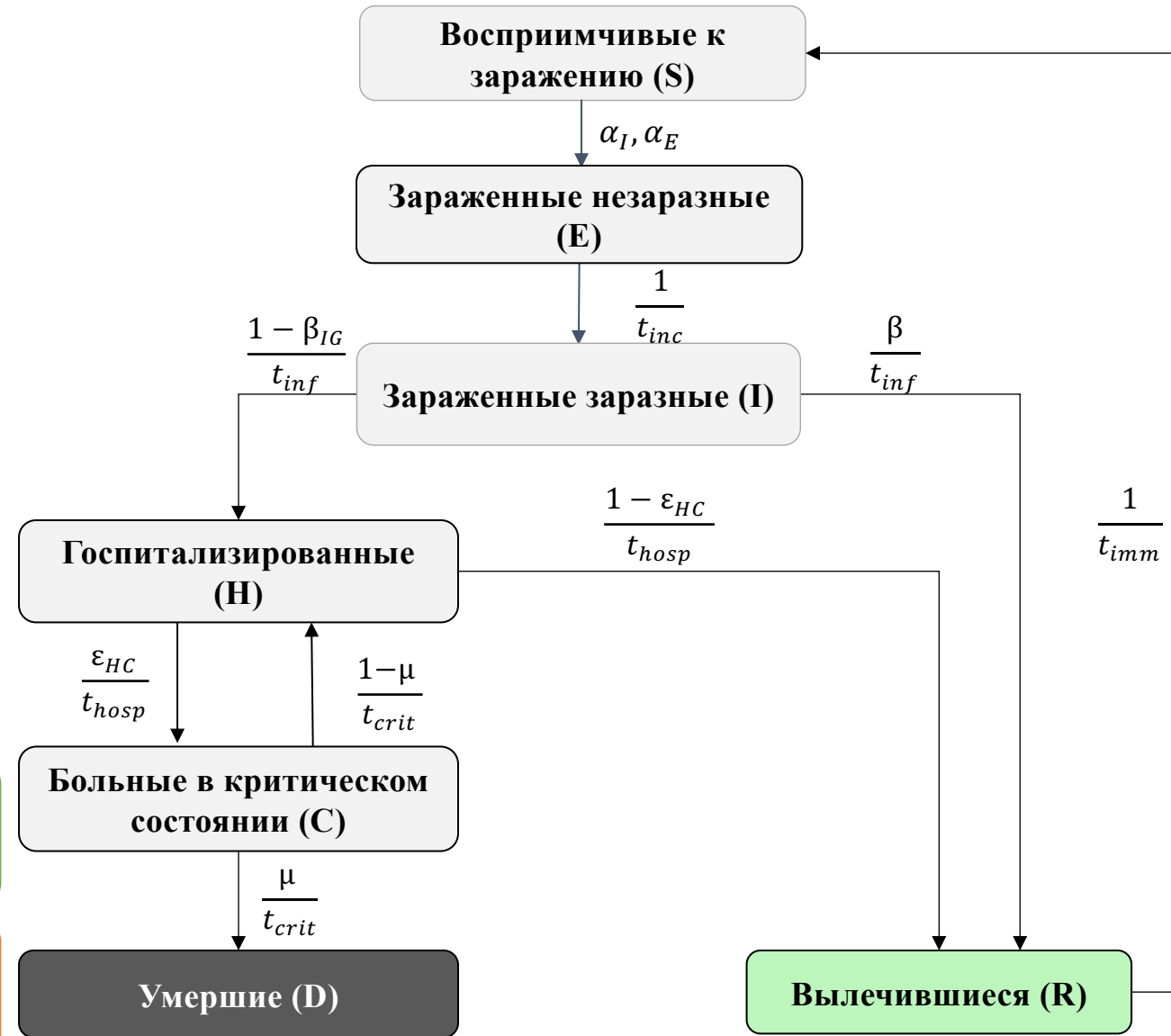
Метод имитационного моделирования, исследующий поведение децентрализованных агентов и то, как такое поведение определяет поведение всей системы в целом. Агентное моделирование включает в себя клеточные автоматы, элементы теории игр, сложных систем, мультиагентных систем и эволюционного программирования, методы Монте-Карло. Каждый агент обладает набором характеристик, а способ взаимодействия агентов из случайных графов определяется особенностями распространения заболевания.

SEIR-HCD моделирование вспышки COVID-19

$$\left\{ \begin{aligned} \frac{dS}{dt} &= -\frac{5-a(t-\tau)}{5} \left(\alpha_E(t) \frac{S(t)E(t)}{N} + \alpha_I(t) \frac{S(t)I(t)}{N} \right) + \frac{1}{t_{imm}} R(t), \\ \frac{dE}{dt} &= \frac{5-a(t-\tau)}{5} \left(\alpha_E(t) \frac{S(t)E(t)}{N} + \alpha_I(t) \frac{S(t)I(t)}{N} \right) - \frac{1}{t_{inc}} E(t), \\ \frac{dI}{dt} &= \frac{1}{t_{inc}} E(t) - \frac{1}{t_{inf}} I(t), \\ \frac{dR}{dt} &= \frac{\beta_{IG}(t)}{t_{inf}} I(t) + \frac{1-\varepsilon_{HC}(t)}{t_{hosp}} H(t) - \frac{1}{t_{imm}} R(t), \\ \frac{dH}{dt} &= \frac{1-\beta_{IG}(t)}{t_{inf}} I(t) + \frac{1-\mu(t)}{t_{crit}} C(t) - \frac{1}{t_{hosp}} H(t), \\ \frac{dC}{dt} &= \frac{\varepsilon_{HC}(t)}{t_{hosp}} H(t) - \frac{1}{t_{crit}} C(t), \\ \frac{dD}{dt} &= \frac{\mu(t)}{t_{crit}} C(t). \end{aligned} \right.$$

Прямая задача: найти вектор $u(t) = (S, E, I, R, H, C, D)$ по начальным данным $S(0) = S_0, E(0) = E_0, I(0) = I_0, R(0) = R_0, H(0) = H_0, C(0) = C_0, D(0) = D_0$.

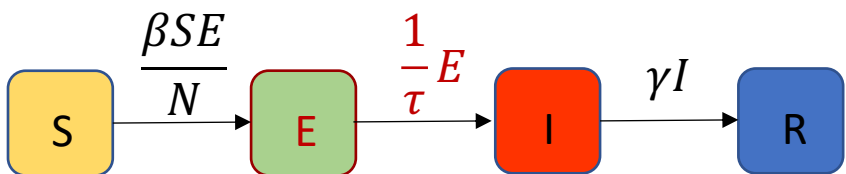
Обратная задача: найти $q(t) = (\alpha_E(t), \alpha_I(t), \beta_{IG}(t), \varepsilon_{HC}(t), \mu(t), u_0)$ по данным $I(t_k) = (1 - b_k)f_k, C(t_k) = C_k, D(t_k) = g_k$ с учетом дополнительных ограничений.



Длительность стадий заболевания

Случайная логнормальная величина с фиксированными параметрами среднего и дисперсии: $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2 / 2\sigma^2}$

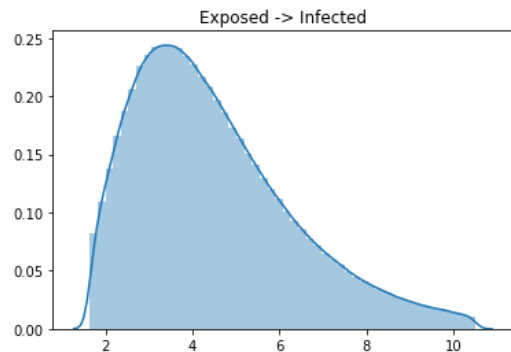
В простейшем случае для SEIR модели



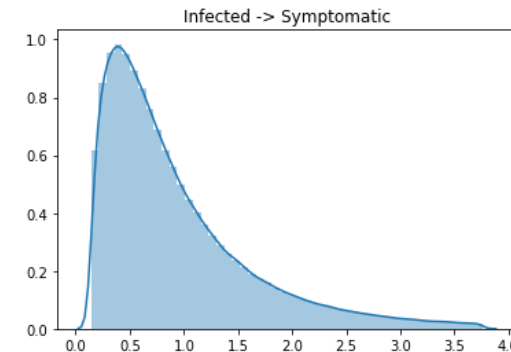
$$\frac{dE}{dt} = \frac{\beta SE}{N} - \frac{1}{\tau}E.$$

$$E(t) + \tau \frac{dE}{dt} = \frac{\beta \tau SE}{N} \approx E(t + \tau).$$

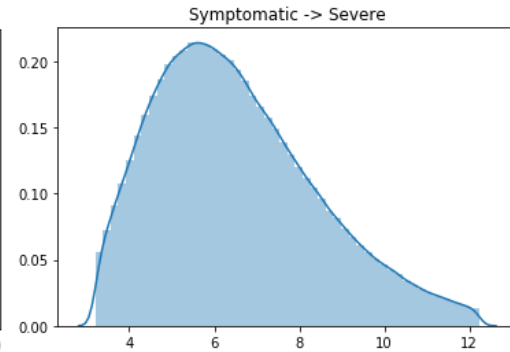
Вероятность $\frac{1}{\tau}$ перехода из группы E в группу I характеризует обратную величину к среднему времени τ пребывания индивидуума в группе E (в линейном приближении).



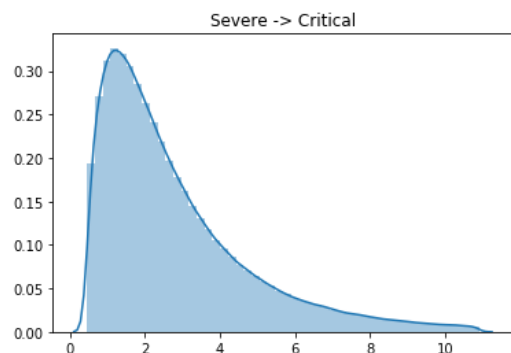
$$t_{inc} \sim \text{LogN}(4.6, 4.8)$$



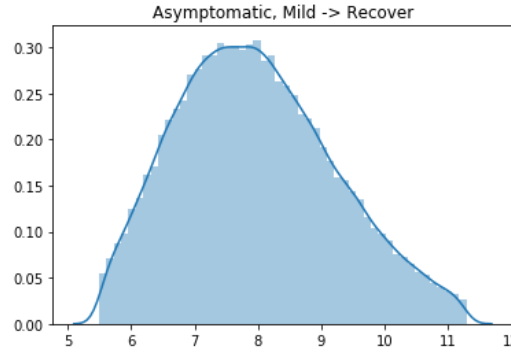
$$t_{sym} \sim \text{LogN}(1.0, 0.9)$$



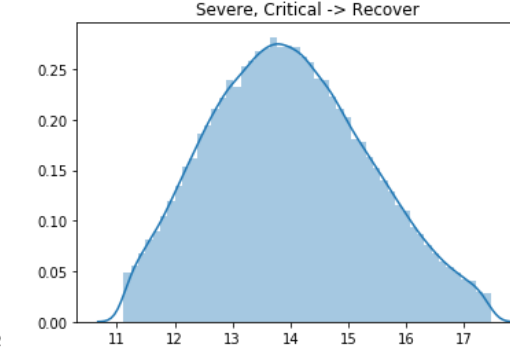
$$t_{inf} \sim \text{LogN}(6.6, 4.9)$$



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



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



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

Свойства SEIR-HCD модели

Базовый индекс репродукции вируса (начало эпидемии):

$$\mathcal{R}_0(t) = \frac{5 - a(t - \tau)}{5} (\alpha_E t_{inc} + \alpha_I t_{inf})$$

При $\mathcal{R}_0 < 1$ состояние равновесия без заболевания устойчиво по Ляпунову.

Преимущество SIR-моделей

- Простота реализации.
- Основа: закон сохранения масс с установленными взаимосвязями с различными состояниями системы.
- Описывает большие популяции (регионы, страны).
- Фундаментальность (для моделирования распространения другой инфекции в другом регионе достаточно уточнить параметры).

Недостатки SIR-моделей

- Описывает только 1 пик.
- Изменение параметров (мутации вируса, ограничительные меры, вакцинация) приводит к необходимости заново решать обратную задачу и рассчитывать сценарии с новыми параметрами.

Криворотько О.И., Кабанихин С.И. Математические модели распространения COVID-19. Новосибирск: Препринт Института математики № 300, 2022.

Параметр	Описание	Границы
$a(t)$	Индекс самоизоляции (данные Яндекса)	(0, 5)
$\alpha_E(t)$	Параметр заражения между бессимптомной $E(t)$ и восприимчивой $S(t)$ группами населения ($\alpha_E \gg \alpha_I$)	(0, 1)
$\alpha_I(t)$	Параметр заражения между инфицированным $I(t)$ и восприимчивым $S(t)$ населением	(0, 1)
$\beta(t)$	Доля инфицированных, имеющая антитела IgG к SARS-CoV-2	(0, 1)
$\varepsilon_{HC}(t)$	Доля госпитализированных случаев $H(t)$, которым требуется подключение ИВЛ	(0, 1)
$\mu(t)$	Доля смертельных случаев	(0, 0.5)
b_k	Доля симптомных случаев из выявленных больных методов ПЦР-тестирования в день k	(0, 1)
τ	Латентный период	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 ИВЛ	10-20 дней
t_{imm}	Средняя продолжительность гуморального иммунитета после выздоровления	180 дней

Анализ устойчивости решения линеаризованной обратной задачи для SEIR-HCD модели

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

$$u_i(t_k) = f_{ik}$$

В зависимости от количества и качества (уровень погрешности) измерений $f_{ik} = (f_{1k}, f_{2k}, f_{3k})$ решение обратной задачи $\hat{q} = (q, u_0)$ может не существовать, быть неединственным и/или неустойчивым.

Линеаризация в окрестности референтных значений \tilde{q}

$$\frac{du}{dt} = \frac{\partial \Phi}{\partial u}(t, \tilde{u}, \tilde{q}) u + \frac{\partial \Phi}{\partial q}(t, \tilde{u}, \tilde{q}) q,$$

$$u(t_0) = u_0.$$

$$u_i(t_k) = f_{ik}$$

$$A \hat{q} = f$$

A – матрица $3K \times 12$

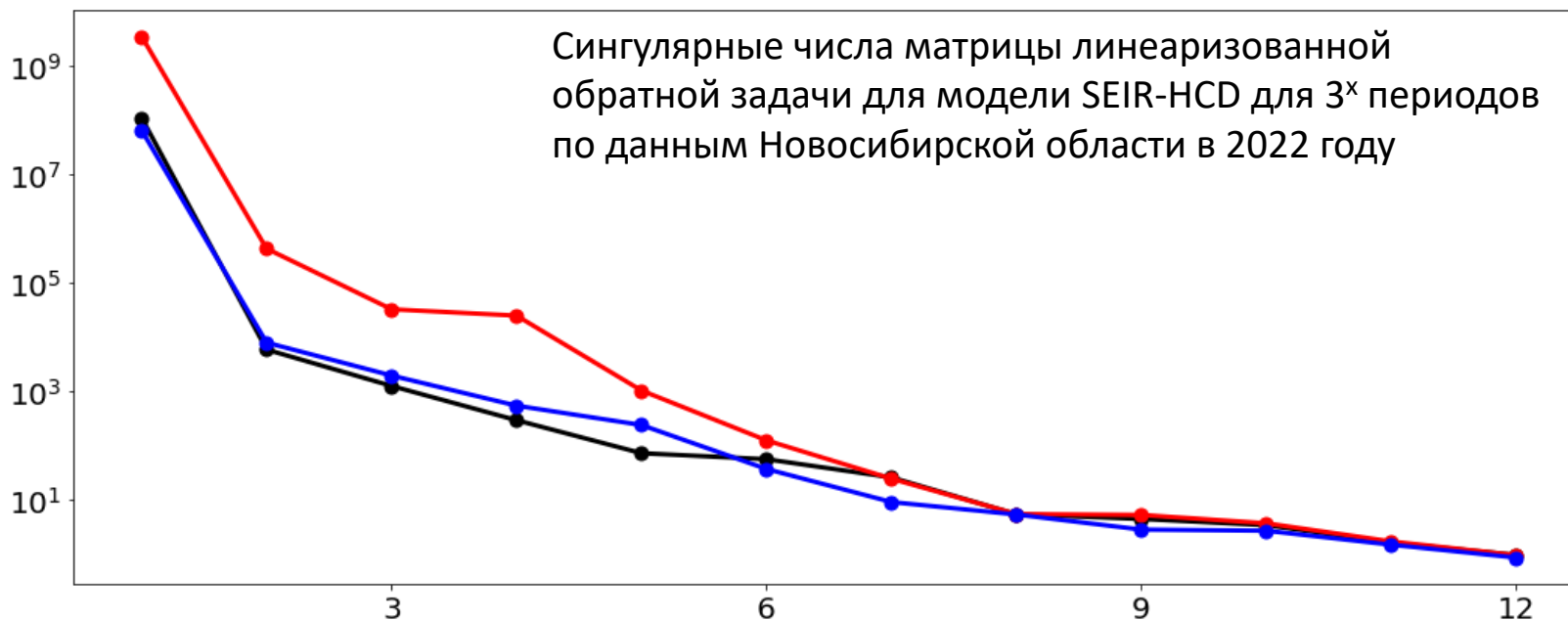
SVD $A = U \Sigma V^T$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$$

$$p = \min(3K, 12)$$

$$\frac{\|\delta q\|}{\|\hat{q}\|} \leq \text{cond}(A) \frac{\|\delta f\|}{\|f\|}$$

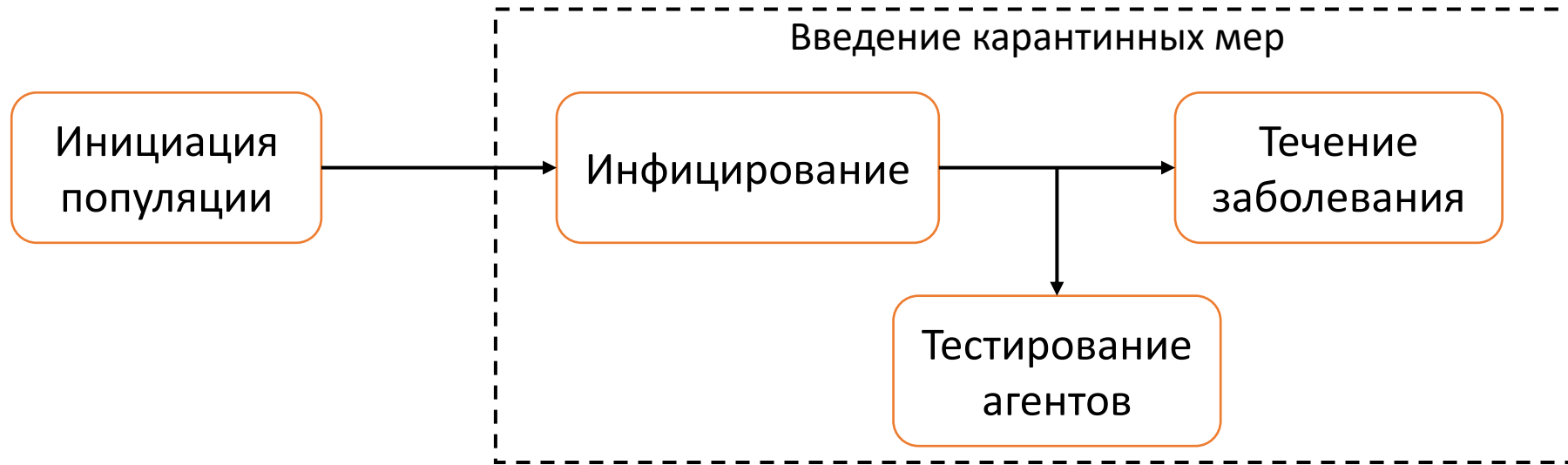
$$\text{cond}(A) = \frac{\sigma_1(A)}{\sigma_p(A)}$$



Красная линия – 02.02-03.03 (штамм Омикрон);
Синяя линия – 08.04-07.05 (стабилизация);
Черная линия – 29.07-27.08 (штамм Кентавр)

Степень убывания сингулярных чисел совпадает для всех 3-х рассматриваемых периодов моделирования. Однако $\text{cond}(A)$ в случае вспышки в результате появления штамма Омикрон на 1 порядок больше, чем остальные 2 периода моделирования.

Агентно-ориентированная 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 [1]

- Агенты разбиваются на возрастные группы по 10 лет [2];
- Популяция разделяется на 4 слоя контактов (домохозяйства, образовательные учреждения, рабочие и общественные места), которые заполняются агентами (рис. А). Среднее количество человек в семье в РФ по данным ООН [3] – 2,7.

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 (рис. В):

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

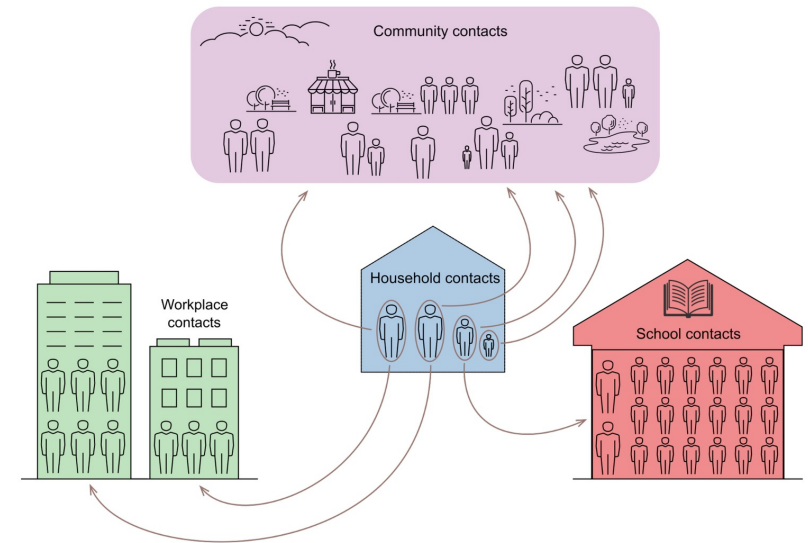
Агенты в возрасте 6-21 лет могут контактировать в образовательных учреждениях с агентами своего возраста, агенты в возрасте 22-65 лет – на работе.

[1] Kerr C. et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. Medverix. 2020. DOI: 10.1101/2020.05.10.20097469.

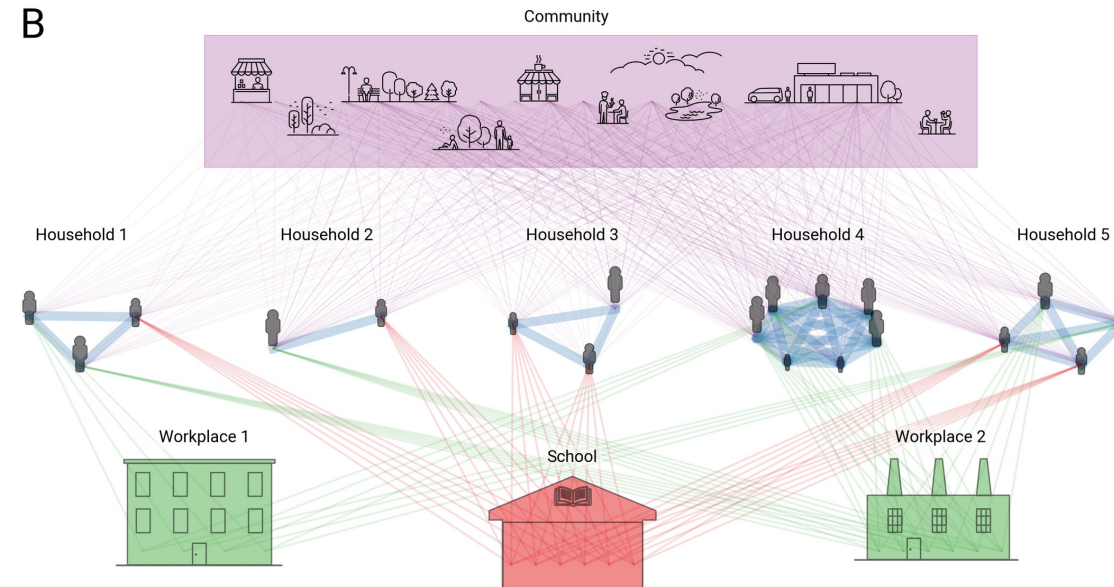
[2] Федеральная служба государственной статистики, Новосибирская обл. <https://novosibstat.gks.ru/folder/31729>

[3] Household Size <https://population.un.org/Household/#/countries/840>

A



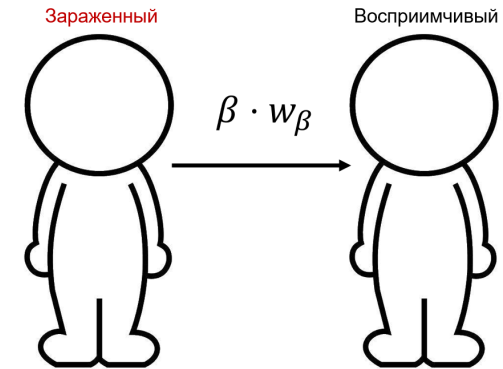
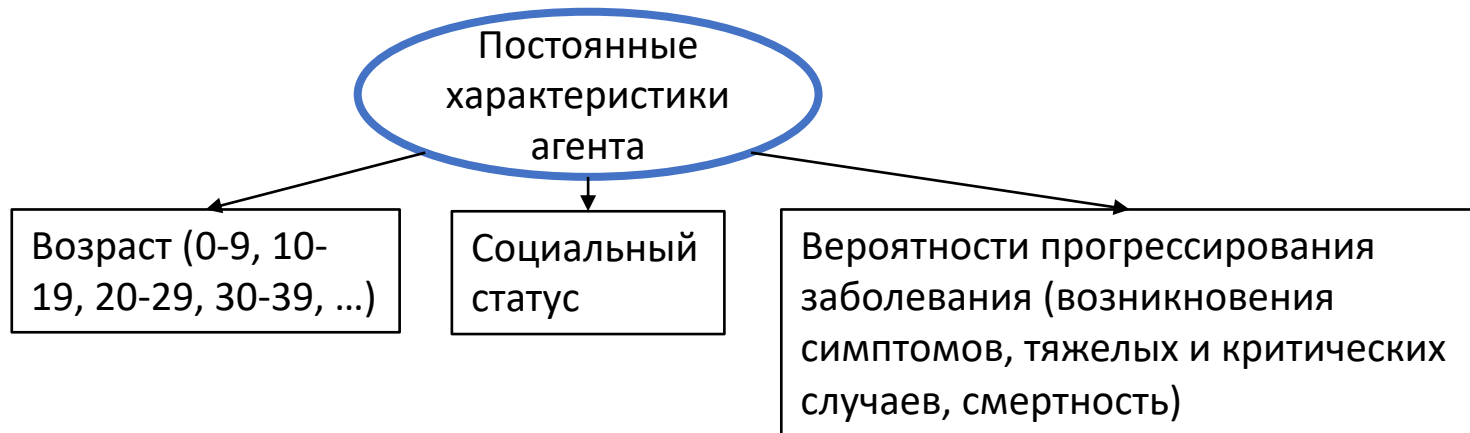
B



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	Скорость заражения
Домохозяйства	$\beta \cdot 3.0$
Школы и университеты (6-22 года)	$\beta \cdot 0.6$
Работа (22-65 лет)	$\beta \cdot 0.6$
Общественные места	$\beta \cdot 0.3$

Течение заболевания и тестирование

Каждый агент может находиться в 9 состояниях заболевания. Переход из одной стадии заболевания в другую контролируется параметрами $p_{sym}, p_{sev}, p_{crit}, p_{death}$, зависящими от возрастной группы агента. Агенты старшего возраста более уязвимы для ранних штаммов вируса (данные параметры оцениваются из статистических данных).

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

Параметр p восстанавливается в ходе решения обратной задачи.

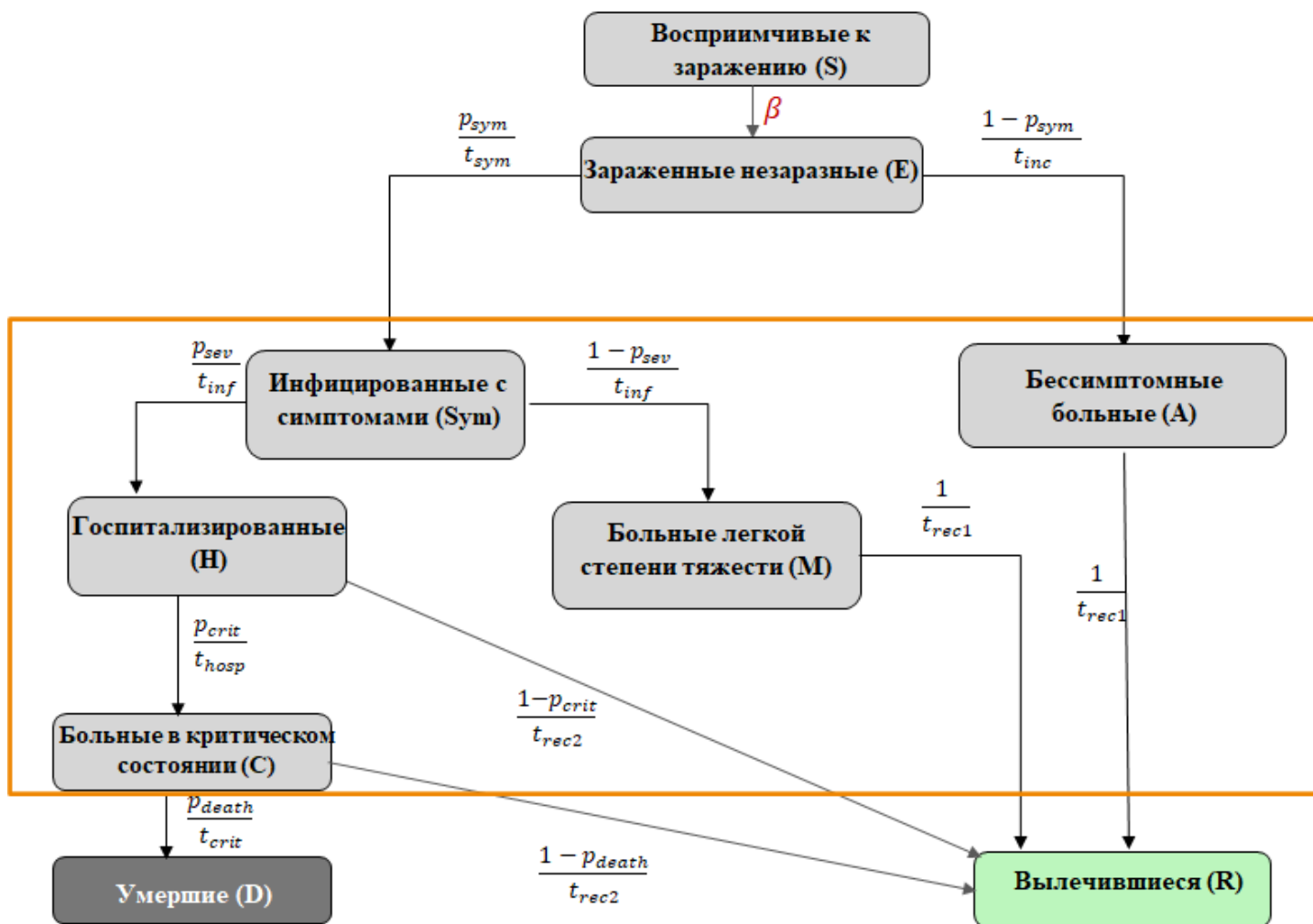


Диаграмма состояний агентов в модели Covasim. Оранжевой рамкой обозначены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19.

Течение заболевания: пример заражения восприимчивого агента

Пусть контакт между двумя индивидуумами происходит в домохозяйстве.

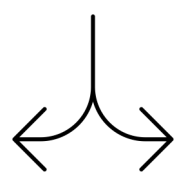
Восприимчивый к заражению (S)



Инфицированный с симптомами (Sym)

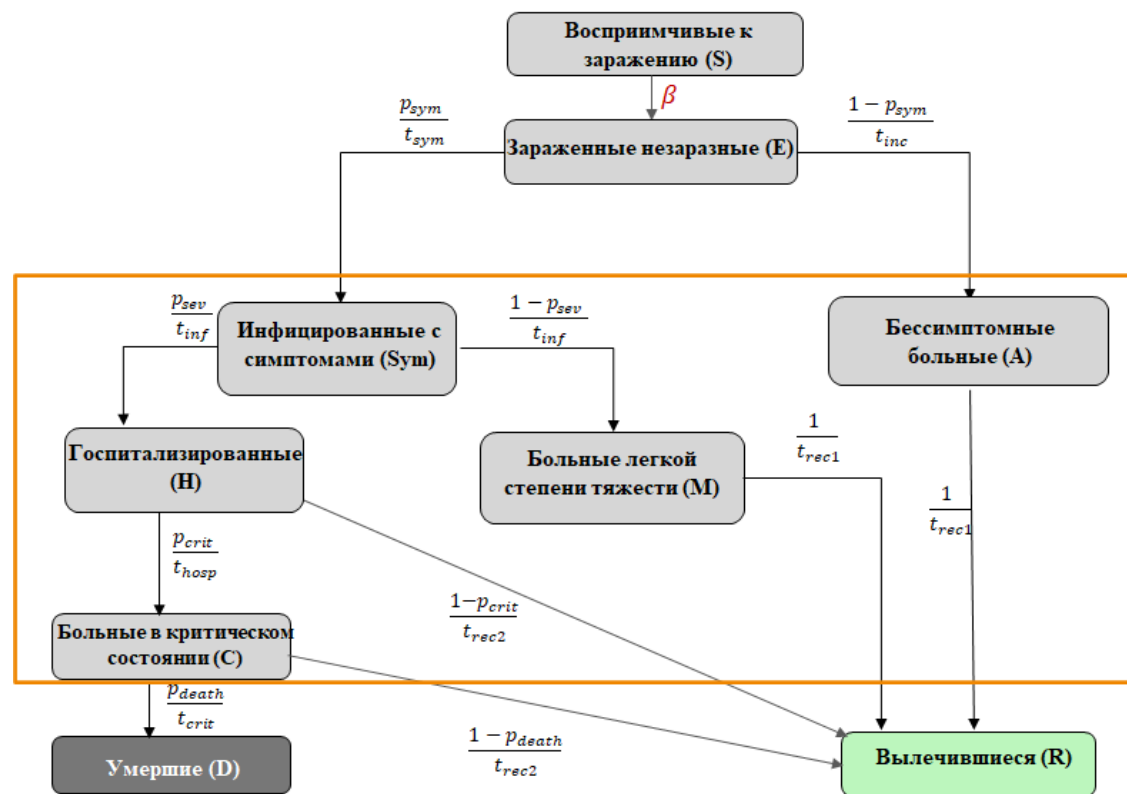


$3 \cdot \beta$



$$p(\text{Sym}) = 3 \cdot \beta \cdot \frac{p_{\text{sym}}}{t_{\text{sym}}}$$

$$p(A) = 3 \cdot \beta \cdot \frac{1 - p_{\text{sym}}}{t_{\text{inc}}}$$



Разыгрывается случайная величина γ . Если $\gamma > p_{\text{sym}}$, то инфицированный становится симптомным больным через t_{sym} дней после контакта, иначе переходит в бессимптомного носителя инфекции через t_{inc} дней.

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.

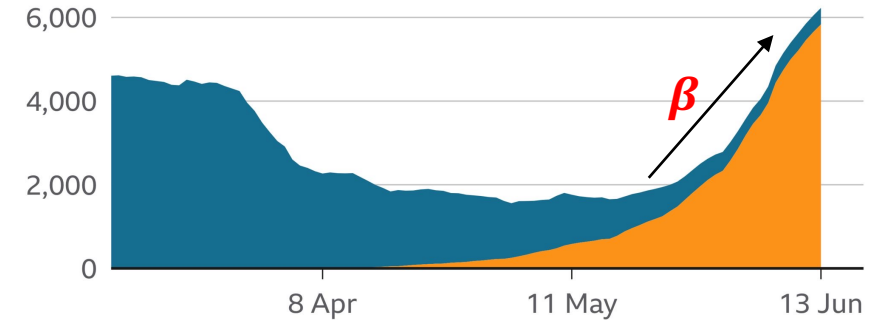
Предположения и ограничения модели

- 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;
- Состояние агента обновляется в конце дня с учетом усреднения возможного пребывания в разных структурах контактов (например, 8 часов на работе, 4 часа в общественном месте и 12 часов дома за весь моделируемый день).

Delta variant now dominant in England

Rolling 7 day average of daily cases in England

■ Other ■ Delta variant (B.1.617.2 - first detected in India)



Variant cases estimated using proportion found in sequences analysed by COG UK

Source: BBC analysis of COG-UK and gov.uk data

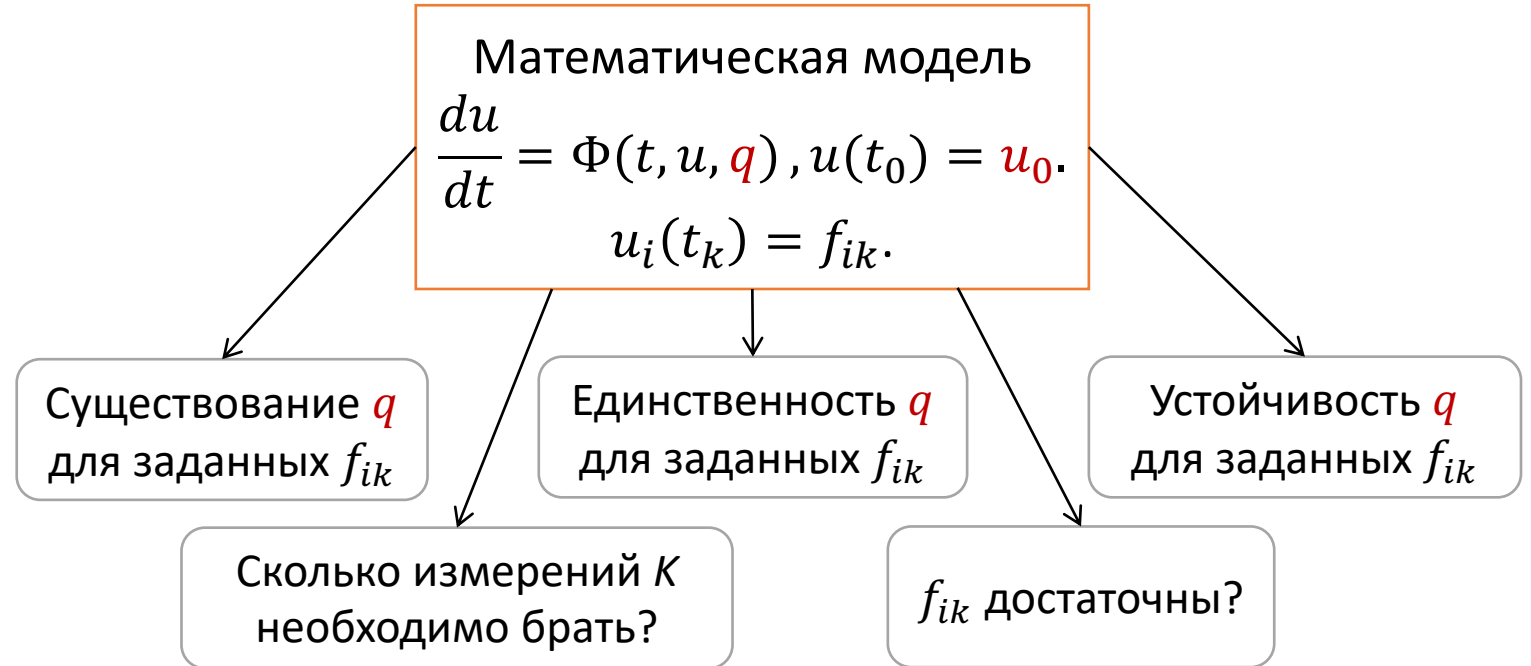
BBC

Обратная задача для агентной модели

Определить вектор параметров $q(t) = (E(0), \beta, \beta_d, \beta_c, p)$ по дополнительной информации о количестве **ежедневно проведенных ПЦР-тестов, выявленных и умерших случаев от COVID-19** в конкретном регионе, а также усредненных статистических показателей для региона (размер семьи, средняя продолжительность состояния стадии заболевания и т.п.).

β_d - день измерения параметра β , β_c - значение, на которое изменится параметр β ,
 p - параметр уровня тестирования.

Идентифицируемость дифференциальных моделей



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

Априорная идентифицируемость

Проводится без привлечения конкретных экспериментальных данных на основе качественного исследования структуры модели

Структурная идентифицируемость

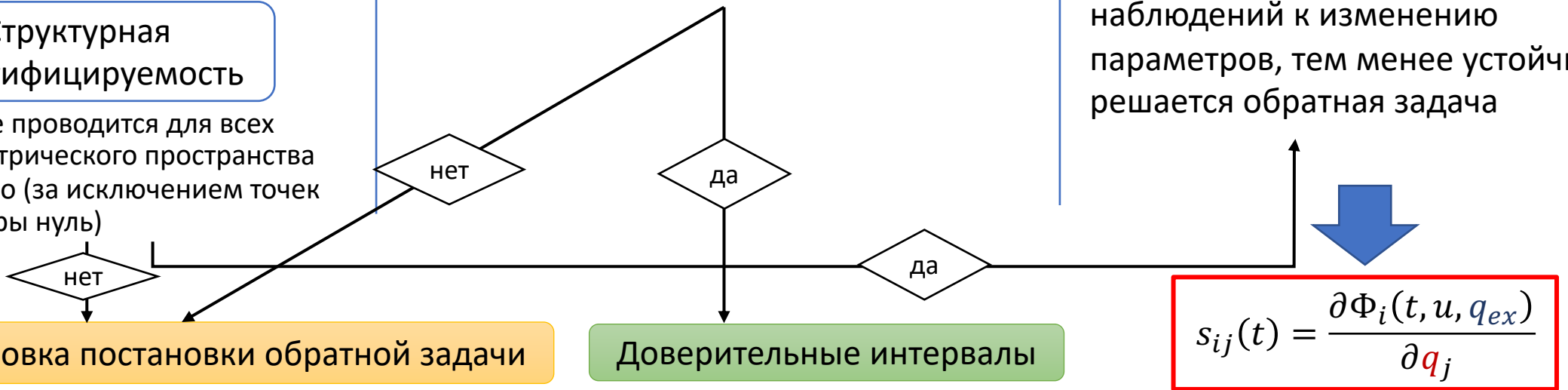
исследование проводится для всех точек параметрического пространства одновременно (за исключением точек множеств меры нуль)

Апостериорная (практическая) идентифицируемость

Проводится при наличии конкретных экспериментальных данных для оценки параметров в отдельных точках параметрического пространства

Анализ чувствительности

Необходимо знать значения параметров, относительно которых исследуется чувствительность. Чем меньше чувствительность вектора наблюдений к изменению параметров, тем менее устойчиво решается обратная задача



Авдеенко Т.В., Горский В.Г. Построение динамических моделей в пространстве состояний: Анализ структурной идентифицируемости: монография - Новосибирск: Изд-во НГТУ (Серия Монографии НГТУ), 2007. - 292 с.

Криворотько О.И., Андорная Д.В., Кабанихин С.И. Анализ чувствительности и практическая идентифицируемость математических моделей биологии. Сибирский журнал индустриальной математики, 2020, т. 23, № 1, стр. 107-125.

Криворотько О.И., Кабанихин С.И., Сосновская М.И., Андорная Д.В. Анализ чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19. Вавиловский журнал генетики и селекции, 2021, т. 25(1), стр. 82-91.


Пример 1. Анализ чувствительности SEIR-HCD модели (параметры постоянны)

Модель ОДУ:

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

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



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.

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

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

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:
 $\alpha_E, E_0, \alpha_I, \varepsilon_{HC}, \mu, \beta$

Пример 2. Анализ чувствительности пространственной SEIR-HCD модели

Модель в частных производных:

$$\frac{\partial u}{\partial t} = \Phi(t, u, q) + \nabla(Nv\nabla u),$$

$$u(x, t_0) = u_0(x), \quad u_x(0, t) = u_x(1, t) = 0.$$

Здесь $u = (S, E, I, R, H, C, D)^T$,

$$q = (\alpha_E, \alpha_I, \beta, \varepsilon_{HC}, \mu, t_{inc}, t_{inf}, t_{hosp}, t_{imm}, t_{crit})^T$$

$$v = (v_s, v_e, v_i, v_r)^T$$

Метод Соболя

$Y = f(q_1, \dots, q_m)$ - модель. Здесь $y = (f_k, C_k, g_k)^T$ - данные о ежедневно выявленных, критических и умерших случаях.

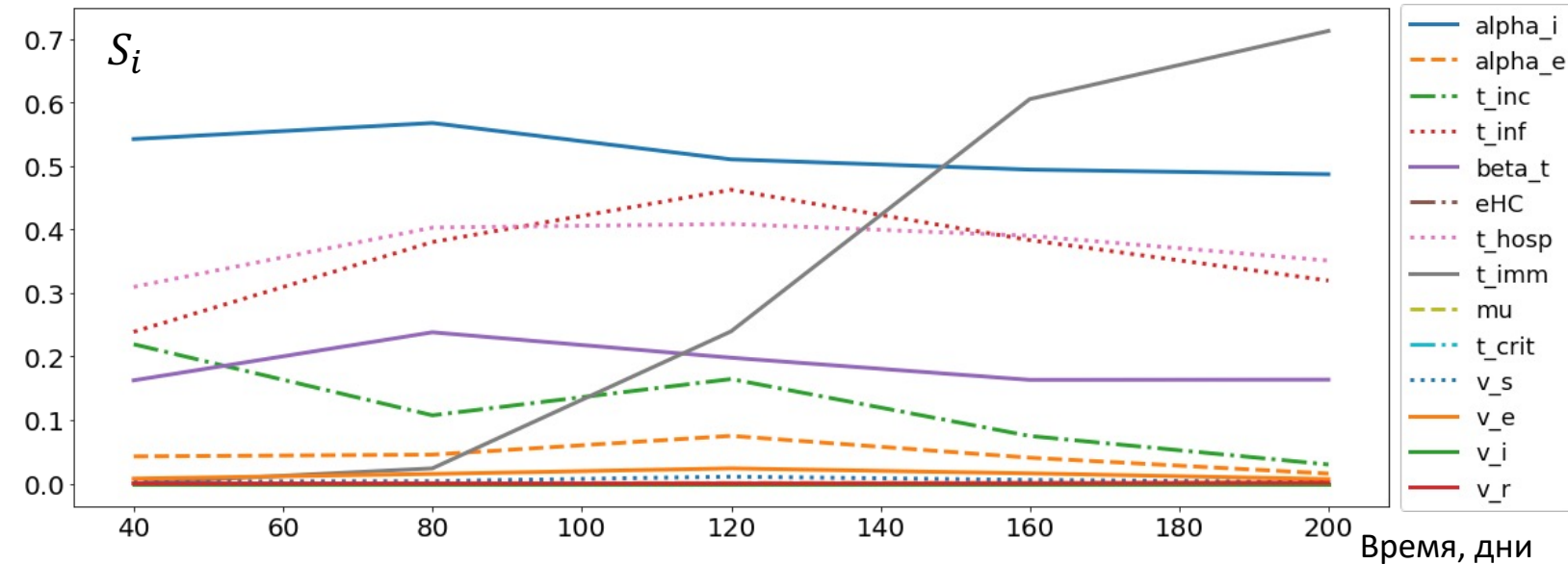
$$\int_{\Omega} i(t_k, x) dx = (1 - b_k) f_k, \quad \int_{\Omega} c(t_k, x) dx = C_k, \quad \int_{\Omega} d(t_k, x) dx = g_k$$

Мера чувствительности параметра q_i : $S_i = \frac{V_{q_i}(\mathbb{E}_{Q_i}(Y|q_i))}{V(Y)}$.

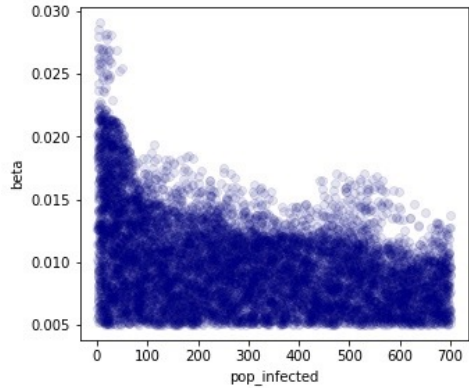
Дисперсия по всем строкам Y

Среднее значение для Y берется по всем возможным значениям пространства параметров при фиксированном $q_i, i = 1, \dots, 14$

Параметр t_{imm} со временем становится более чувствительным. Чувствительность параметров $\alpha_I, t_{inc}, t_{inf}, t_{hosp}$ уменьшается со временем. Параметры диффузионного процесса не чувствительны.



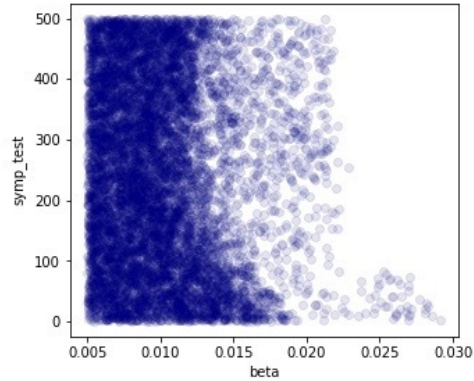
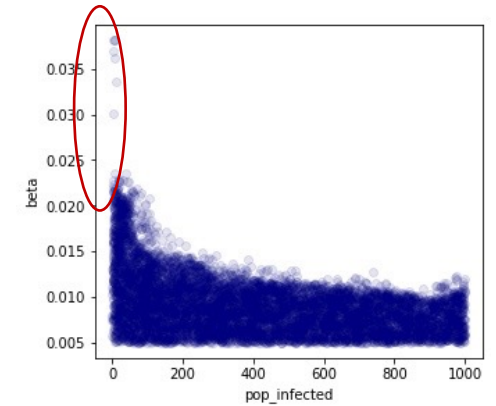
Пример 3. Анализ чувствительности для агентной модели распространения COVID-19



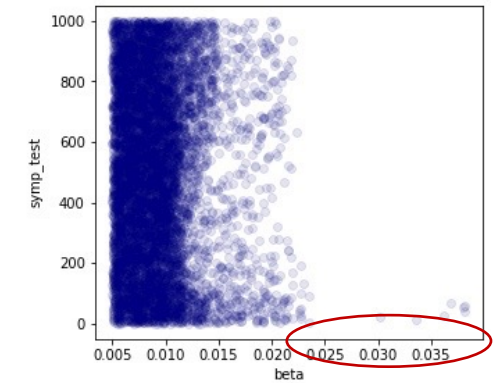
Данные аппроксимируются регрессионной моделью

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

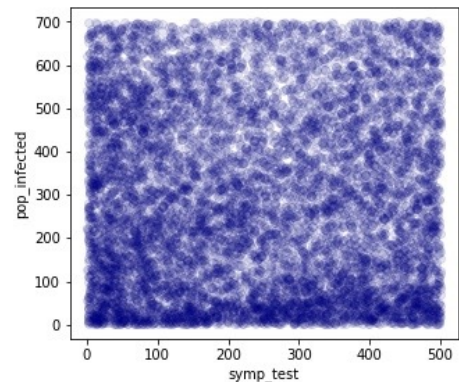
Где параметры регрессии определяются из задачи максимизации функции правдоподобия.



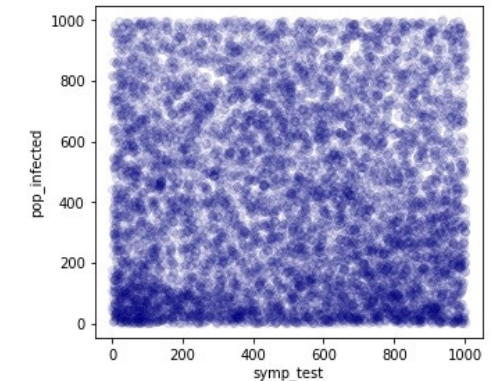
← $\mathbf{q} = (\beta, E(0), p), f$ – выявленные и умершие случаи
 Параметр contagiозности β (скорость передачи инфекции) чувствителен к дополнительным измерениям f .



\bar{f} – выявленные, критические и умершие случаи →



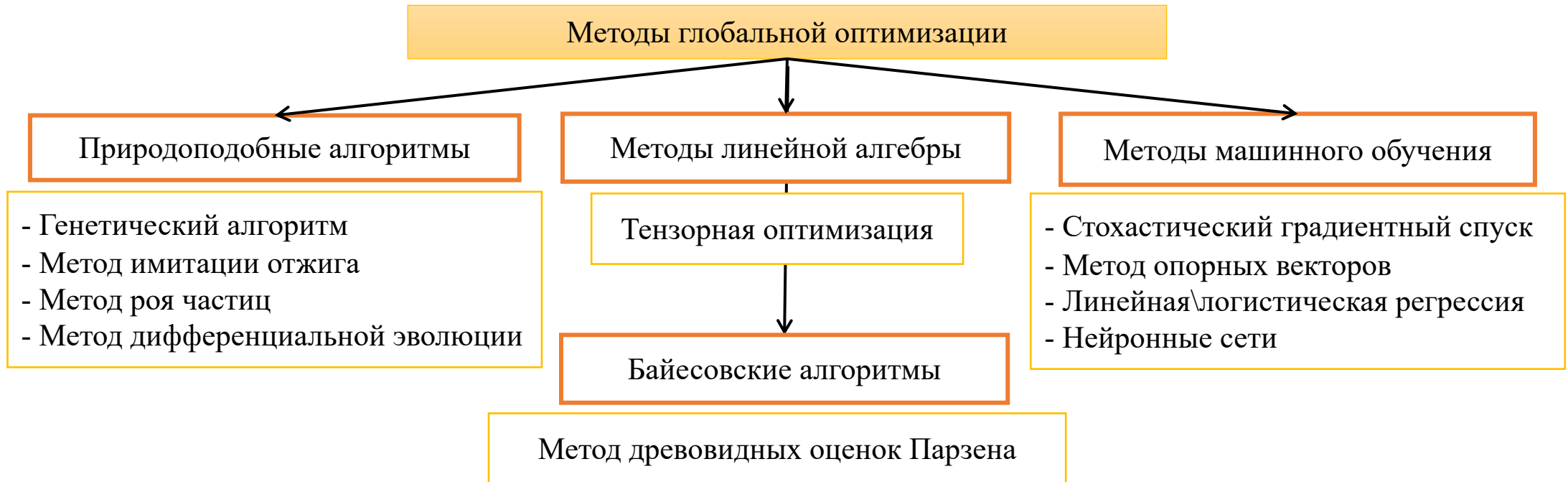
Добавление информации о критических случаях позволяет уточнить границы изменения параметра β . Количество исходных инфекционных случаев $E(0)$ и шанс быть протестированным p не зависели от используемых данных.



Алгоритмы решения обратных задач

$$J(q) = \sum_{k=1}^K \sum_s w_s \frac{|c_d^{i,s} - c_m^{i,s}(q)|^2}{(c_d^{s,t})^2} \rightarrow \min_q$$

Методы глобальной оптимизации + Градиентные (локальные) методы = $q^* = \underset{q \in Q}{\operatorname{argmin}} J(q)$



Градиентные методы: выражение градиента целевого функционала

ОДУ

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

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

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

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

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

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.

УЧПР

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

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

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

$$\begin{cases} \frac{\partial \Psi}{\partial t} = -\mu_u^T(u(x, t), q) - d \frac{\partial^2 \Psi}{\partial x^2}, t > 0, x \in \Omega, \\ \Psi(T) = 0, \quad \frac{\partial \Psi}{\partial x} \Big|_{\partial \Omega_i} = 0, \\ [\Psi_i]_{t=t_k} = 2\omega_i \left(\sum_i u(x_i, t_k; q) - f_k \right), k = 1, \dots, K. \end{cases}$$

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.

Природоподобные алгоритмы

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:

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

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.

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

Метод древовидных оценок Парзена (пакет OPTUNA)

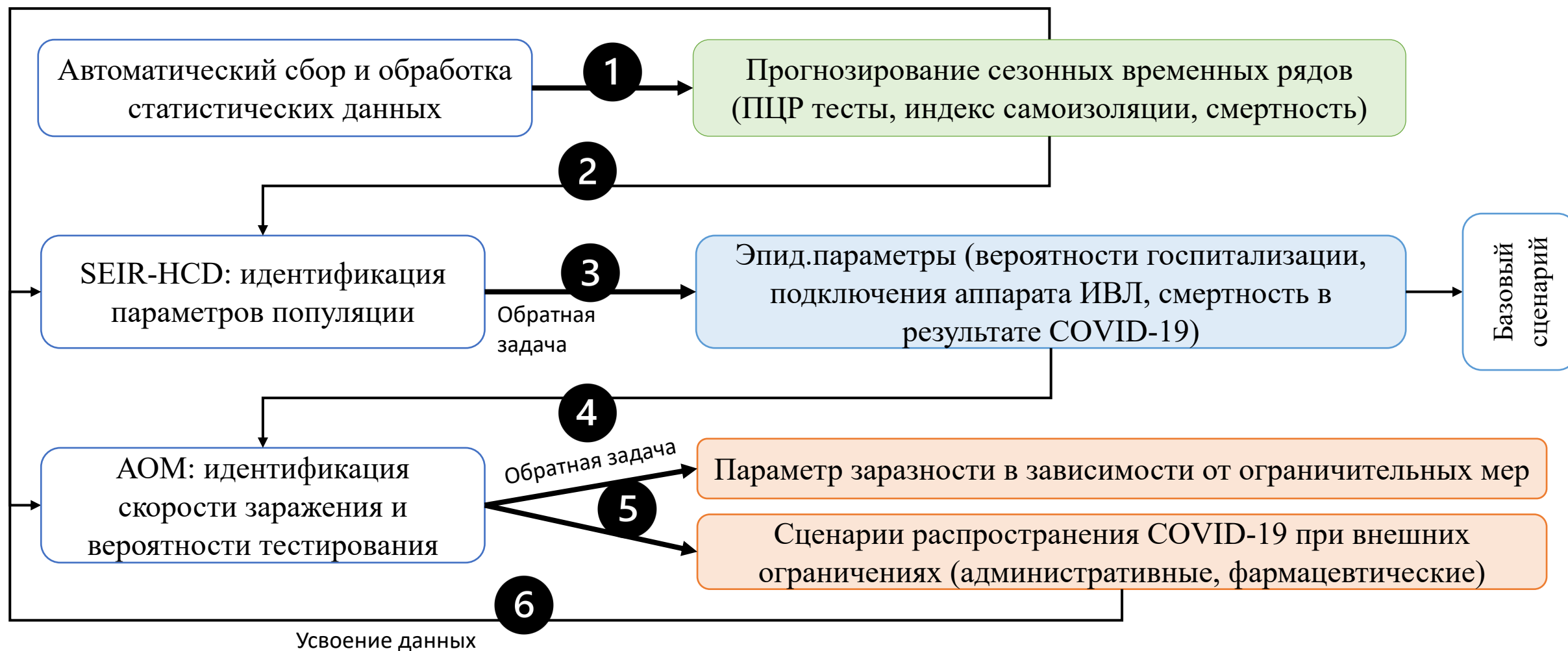
1. Выбираем J^* как квантиль γ , т.е. $p(J < J^*) = \gamma$ и моделируем $p(J)$.
2. Все остальные значения разделяются на 2 группы: те, которые доставляют значение функционала меньше $J^*(q^i)$ и остальные.
Т.о. строятся плотности вероятности значений q^i – это $l(q)$ и плотности остальных значений – $g(q)$. Получаем вероятность

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

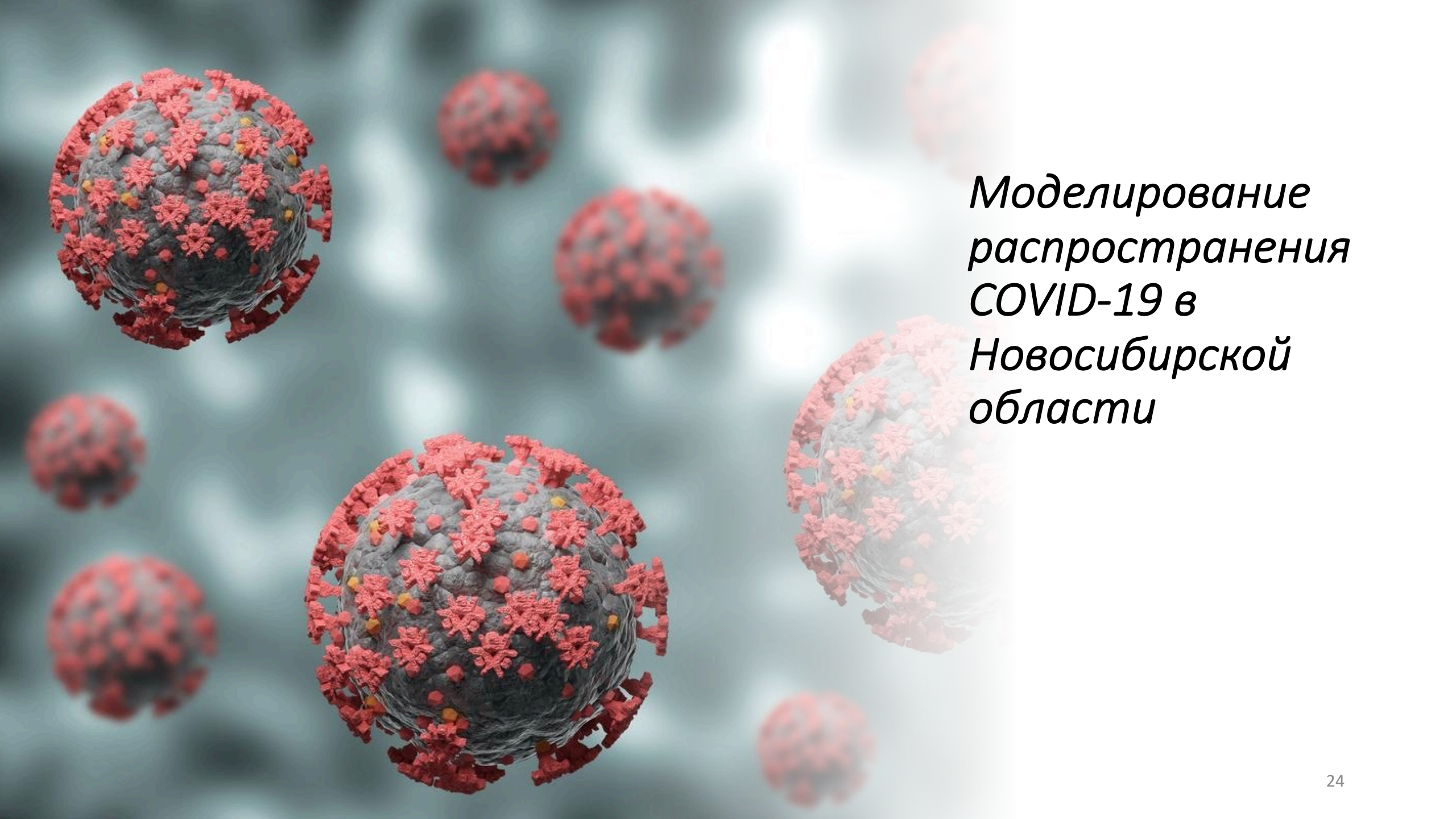
3. Используя формулу полной вероятности, задача оптимизации сводится к задаче максимизации $l(q)/g(q)$.

Сходимость: $E \ln \frac{|q_k - q^*|}{|q_{k-1} - q^*|} \rightarrow -c < 0, \quad k \rightarrow \infty.$

Комбинированная модель распространения COVID-19: цикл 1 дня (<https://covid19-modeling.ru/>)



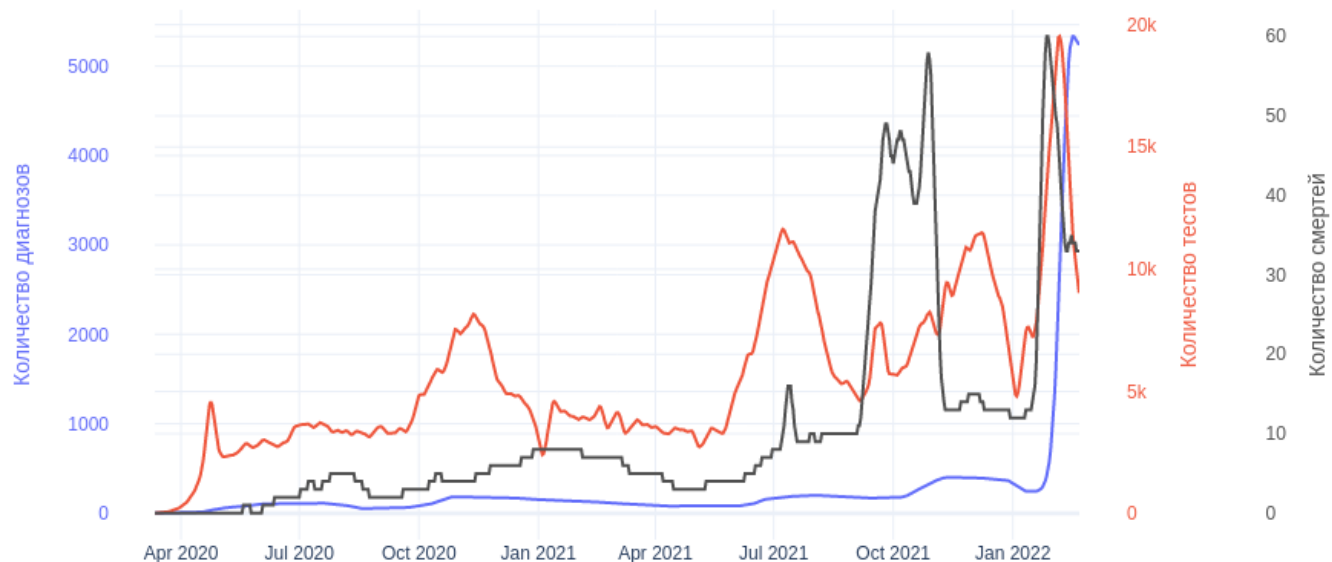
Ежедневный сценарий распространения COVID-19 обновляется на основе комбинации SIR и агентного подходов.

The background of the slide features several 3D rendered COVID-19 virus particles. These particles are spherical with a greyish-blue textured surface and are covered with numerous red, crown-shaped spike proteins. Some particles are in sharp focus, while others are blurred in the background, creating a sense of depth. The overall color palette is cool, with blues and greys, contrasted by the bright red of the spikes.

*Моделирование
распространения
COVID-19 в
Новосибирской
области*

Эпидемиологические показатели в Новосибирской области

Данные по Новосибирской области

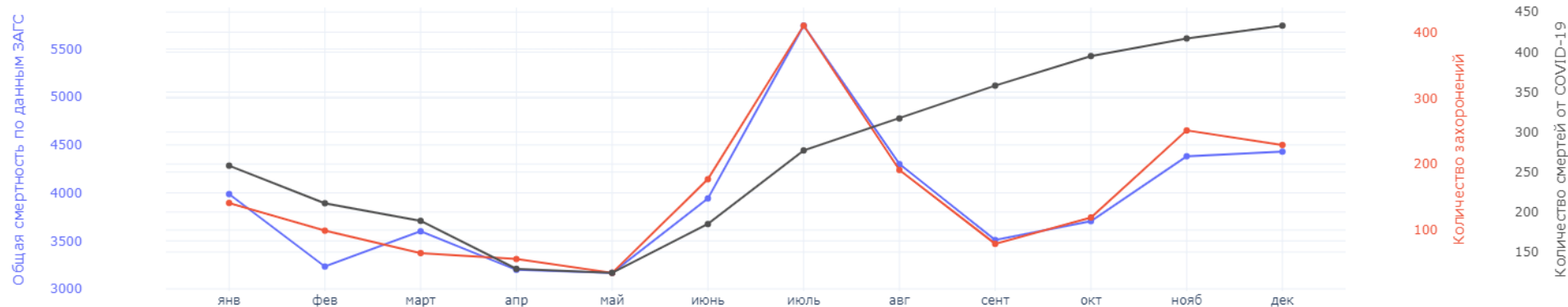


<http://covid19-modeling.ru/>

Сглаженные данные с 12.03.2020 по 12.02.2022 (стопкоронавирус.рф)

- Ежедневное количество проведенных ПЦР-тестов (красная линия),
- Ежедневно выявленные случаи (синяя линия),
- Ежедневное количество смертей от COVID-19 (черная линия).

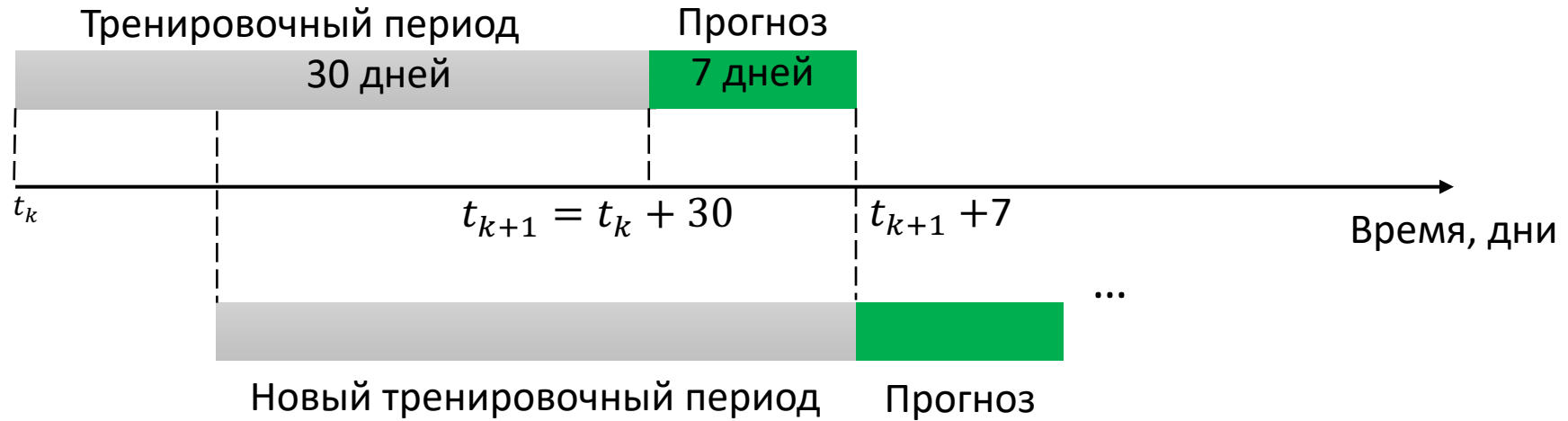
Смертность в 2021



Количество захоронений от COVID-19, полученные муниципальной информационной системой «Ритуал» (мэрия города Новосибирска) коррелирует с данными по общей смертности (ЗАГС).

Черная линия – смертность от COVID-19, красная линия – захоронения больных, синяя линия – общая смертность

Алгоритм усвоения данных



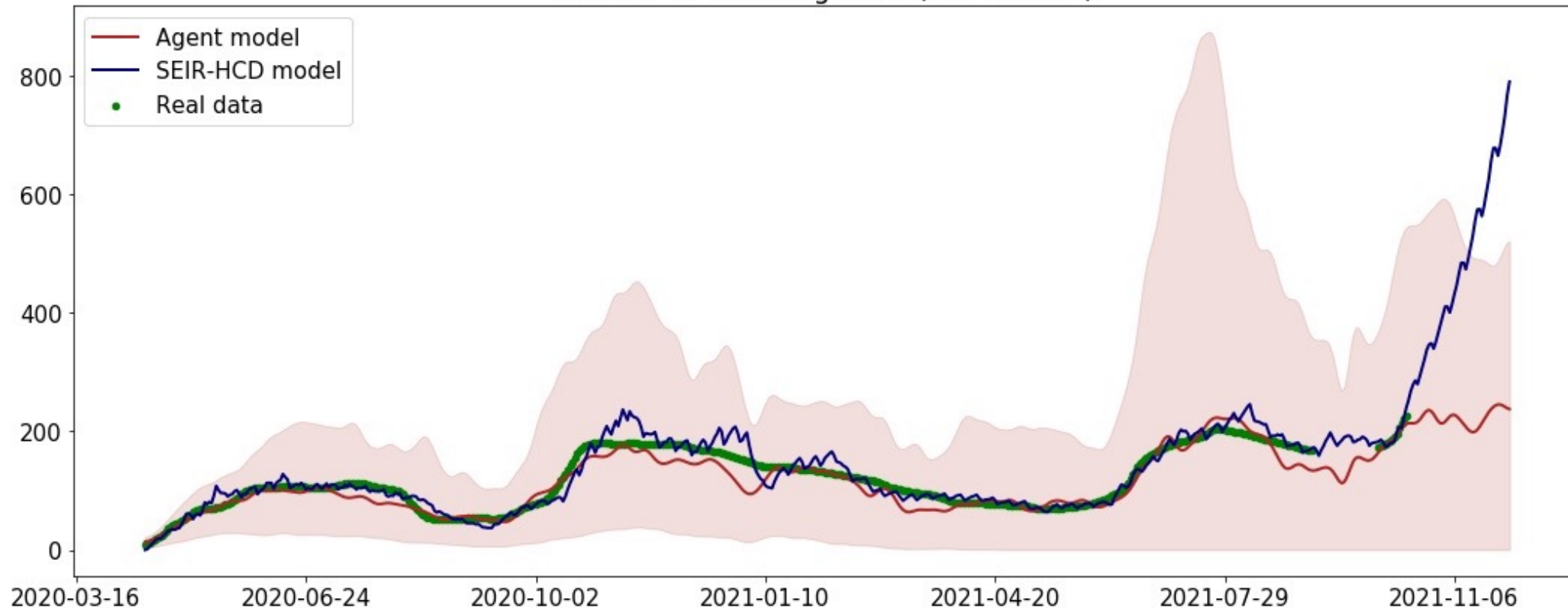
Тренировочный период (30 дней) – уточнение параметров q .
Прогноз (7 дней) рассчитывается при найденных q .
Новый период сдвигается на 7 дней и снова решается обратная задача (уточнение параметров q).

Алгоритм построения доверительных интервалов (для агентной модели)

- Решаем прямую задачу с идентифицированными параметрами 1000 раз (вычислительное время решения прямой задачи для агентной модели на кластере = 10 сек);
- Для каждого дня считаем квантили уровня 0.1, 0.5, 0.9 (строим функцию распределения случайной величины выявленных случаев и выбираем ее значения в точках 0.1, 0.5 и 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).

Моделирование распространения 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)



Красная линия – агентная модель, **синяя линия** - SEIR-HCD модель, **жирная зеленая линия** - реальные данные с 12.03.2020 по 16.10.2021. Прогноз сделан на 45 дней с 17.10 по 30.11. Закрашенная красная область - 80% доверительный интервал.

Сравнение прогноза на 24.10.2021:

- 250 человек – агентная модель
- 392 человека – SEIR-HCD,
- 281 человек – реальные данные

Глубокое обучение для задачи моделирования COVID-19

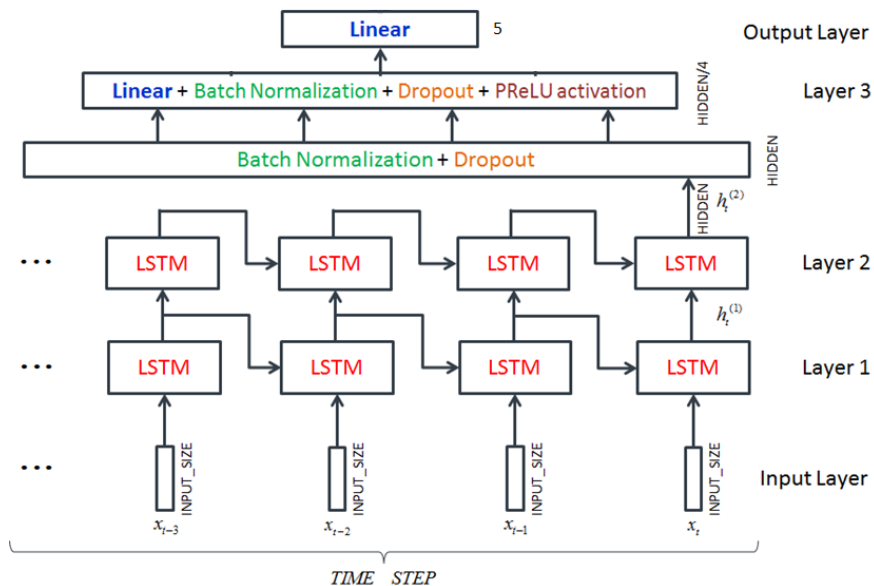
Обучающие признаки:

- 1) новые выявленные случаи COVID-19,
- 2) госпитализированные,
- 3) подключённые к ИВЛ,
- 4) ежедневно захороненные,
- 5) логарифм каждого признака 1)-4),
- 6) индекс самоизоляции Яндекс,
- 7) процент бессимптомных случаев,
- 8) процент населения с антителами IgG,
- 9) логарифмические lag-приращения признаков 1)-4), где lag = 3, 7, 14 дней.
- 10) текущий день недели,
- 11) ввело ли правительство ужесточение в текущий день (1 или 0),
- 12) ввело ли правительство послабление в текущий день (1 или 0).

Целевые значения:

Будем строить прогноз **новых выявленных случаев** COVID-19 на следующие **7 дней** в Новосибирской области

Архитектура нейронной сети:



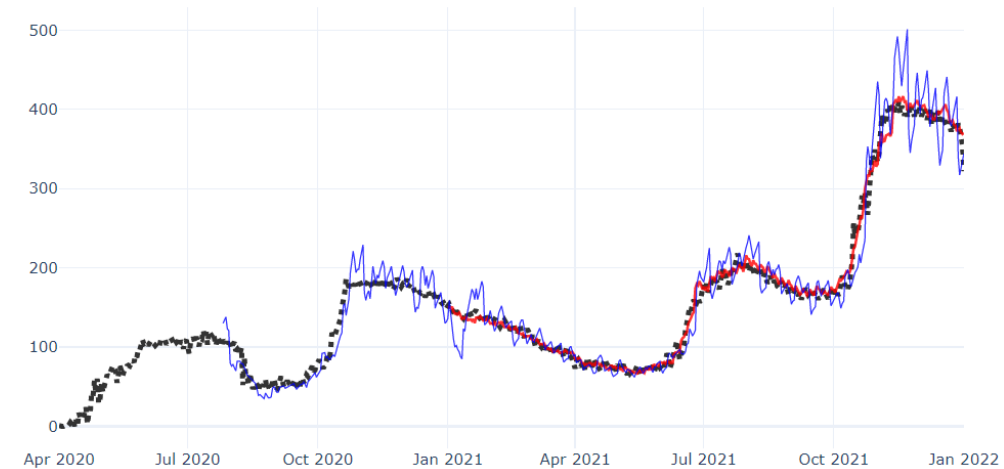
Сравнение с комбинированной моделью:

Синяя линия – прогноз новых выявленных случаев с помощью комбинированной модели на следующие 7 дней.

Красная линия – прогноз новых выявленных случаев с помощью нейронной сети на следующие 7 дней.

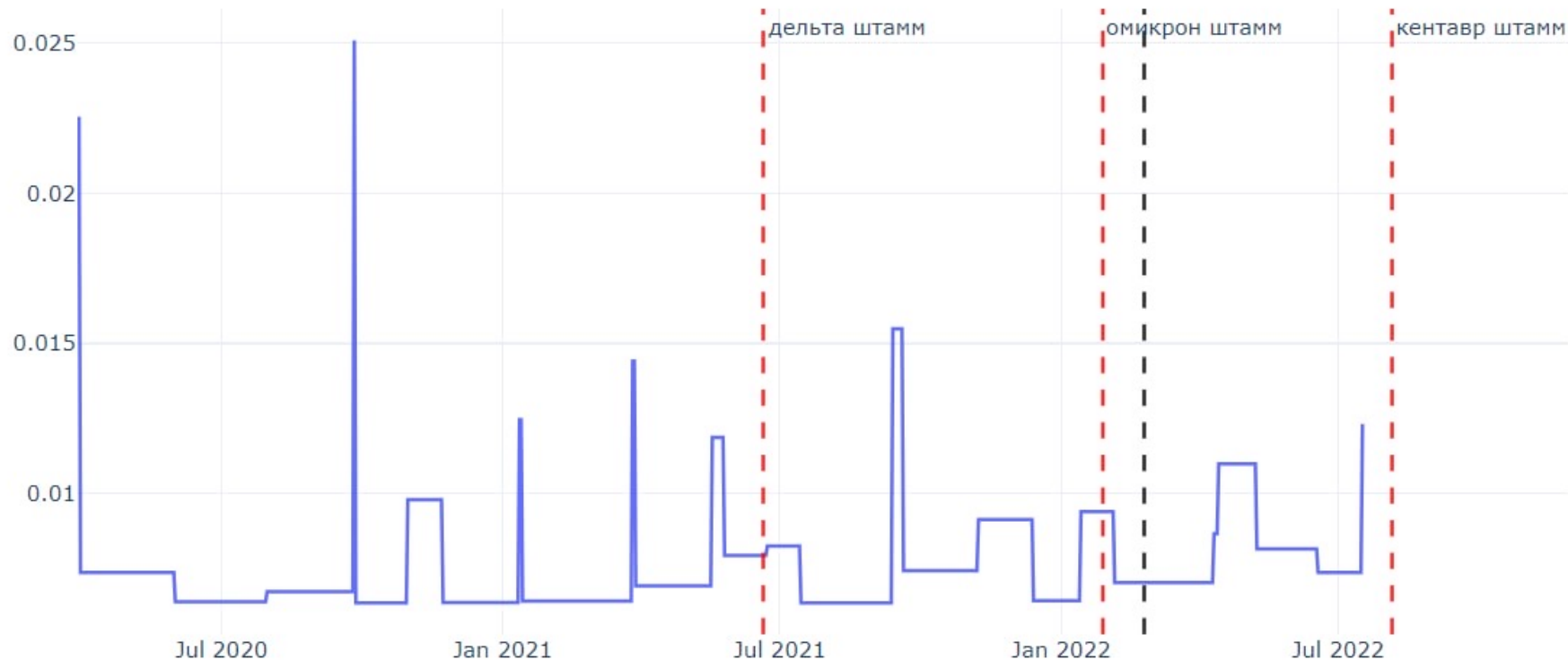
Черные точки – реальные данные новых выявленных случаев COVID-19 (<https://стопкоронавирус.рф>).

Новые случаи COVID-19 в Новосибирской области, 2020-04-01 - 2022-01-01



Влияние ограничительных мер и появления новых штаммов на изменение параметра заразности в НСО с 04.2020 по 08.2022.

Изменение параметра заразности в Новосибирской области



Синяя линия – параметр заразности β ,

Пунктирные **красные линии** – появление штамма в РФ,

Черная линия – отмена масочного режима в регионе 24.02.2022.

Отмена масочного режима в весенний период 2022 года повлекло за собой увеличение параметра заражения, после чего в летний период его значение стабилизировалось. Увеличение параметра заразности в конце июля свидетельствует о более раннем функционировании нового штамма «Кентавр» в регионе.

Сценарии распространения COVID-19 для новых выявленных случаев в Новосибирской области с 07.08 по 12.09.2022



Голубая линия – базовый сценарий с 07.08.2022 по 12.09.2022.

Оранжевая линия – сценарий, когда 30% рабочих мест переходят в удаленный режим с 21.08.2022.

Зеленая линия – сценарий увеличения мобильности граждан в общественных местах с 21.08.2022.

Красная линия – базовый сценарий SEIR-HCD модели

Красные точки – тестовые данные с 07.08.2022 по 19.08.2022 (для верификации модели).

Изучена идентифицируемость агентных, дифференциальных и стохастических моделей эпидемиологии с учетом экономических и социальных процессов

SEIR-HCD

$$q_1 = (\alpha_E, \alpha_I, \beta_{IgG}, \varepsilon_{HC}, \mu, E_0)^T$$
$$\text{data} = (f_k, C_k, g_k)^T, k = 1, \dots, K$$

Последовательность чувствительных параметров от наибольшего к наименьшему $\alpha_E, E_0, \alpha_I, \varepsilon_{HC}, \mu, \beta_{IgG}$

Доля инфицированных, у которых заболевание протекает в легкой форме β_{IgG} , определяется из информации о проценте людей с антителами IgG

ABM

$$q_2 = (\beta, E(0), \tilde{p})^T$$
$$\text{data} = (f_k, C_k, g_k)^T, k = 1, \dots, K$$

На основе байесовского подхода и метода Монте Карло уточнены границы изменения параметров

$$\beta \in [0.005, 0.05] \rightarrow \hat{\beta} \in [0.005, 0.022]$$
$$E_0 \in [1, 1000] \rightarrow \hat{E}_0 \in [1, 1000]$$
$$\tilde{p} \in [1, 1000] \rightarrow \hat{\tilde{p}} \in [1, 1000]$$

Удалось уменьшить границы изменения параметра заразности β более чем в два раза



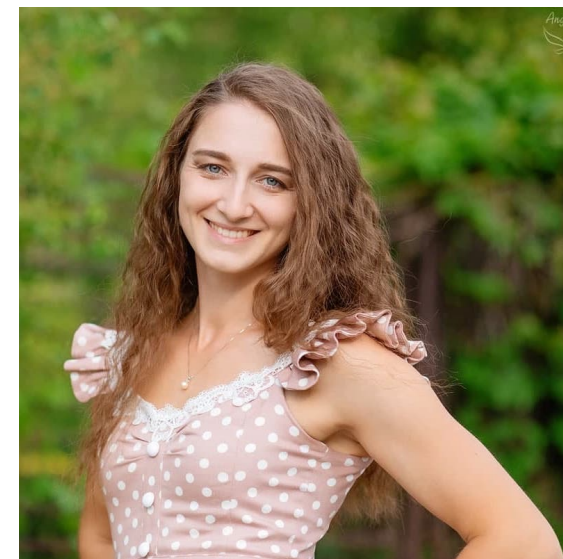
На этой основе построены новые градиентные методы решения обратных задач с использованием априорной информации



Спасибо за внимание!



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



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