

Вторая конференция математических центров

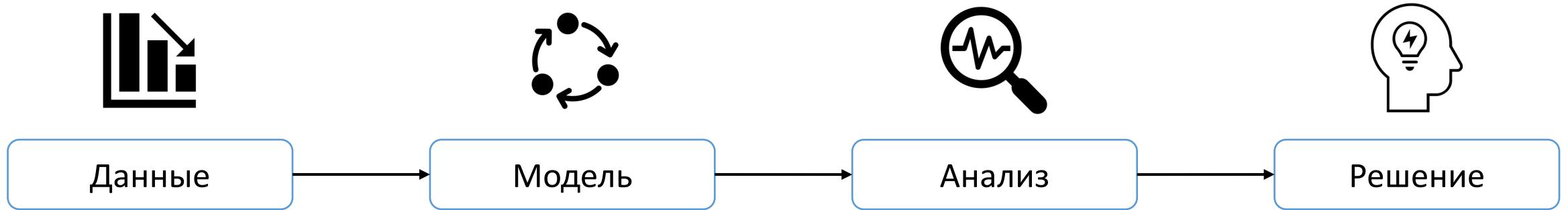
Москва, МГУ-МИАН, 07 – 11 ноября, 2022

*Математическое моделирование социально-
значимых заболеваний в Российской Федерации*

к.ф.-м.н. О.И. Криворотко

Москва, 11 ноября 2022

Построение моделей распространения инфекционного заболевания



Эпидемиологические
Демографические
Социальные
Экономические

...

Дифференциальная
Стохастическая
Основанная на графах
Игра среднего поля

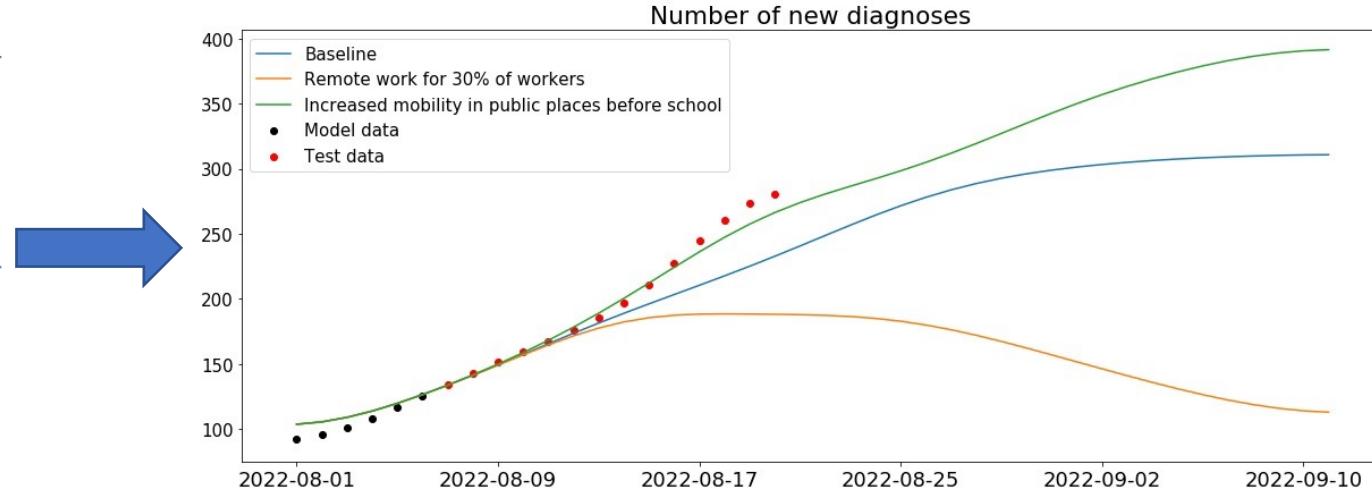
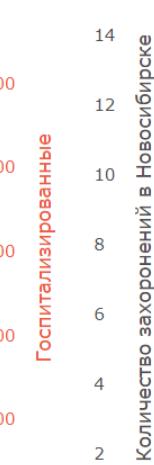
...

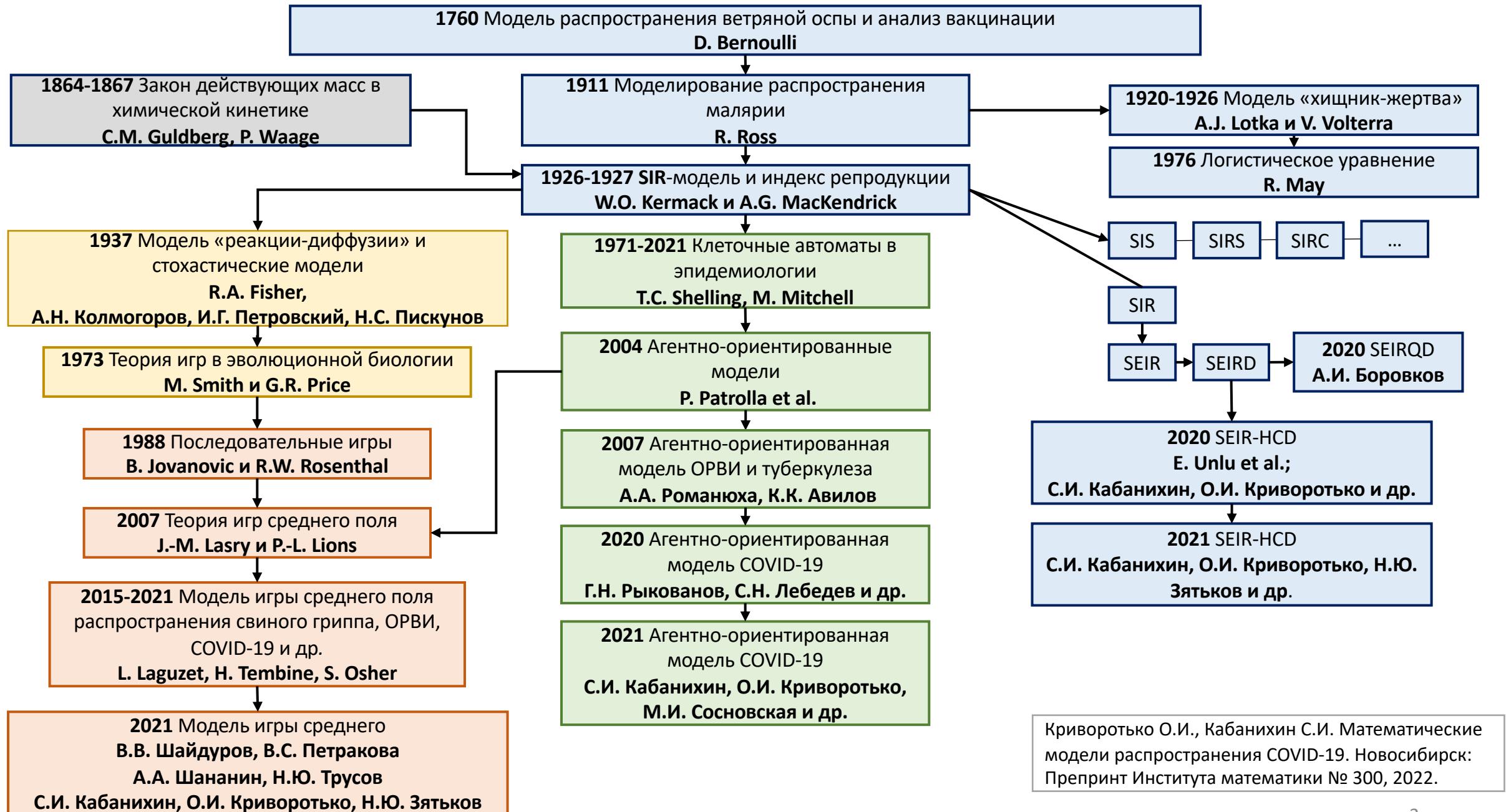
Идентифицируемости
Устойчивости
Сходимости
Регуляризации

...

Прямой задачи
Обратной задачи
(оптимизация)
Прогнозирование

...





Математические модели эпидемиологии: два основных подхода

Камерные модели

Моделирование сверху-вниз

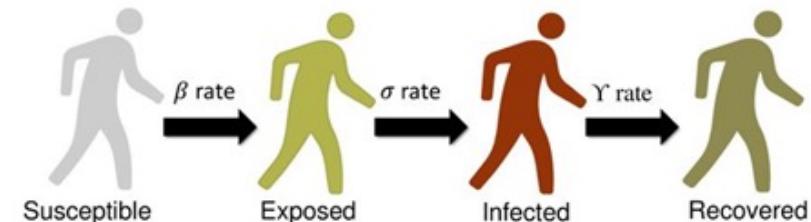
Детерминистские

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).

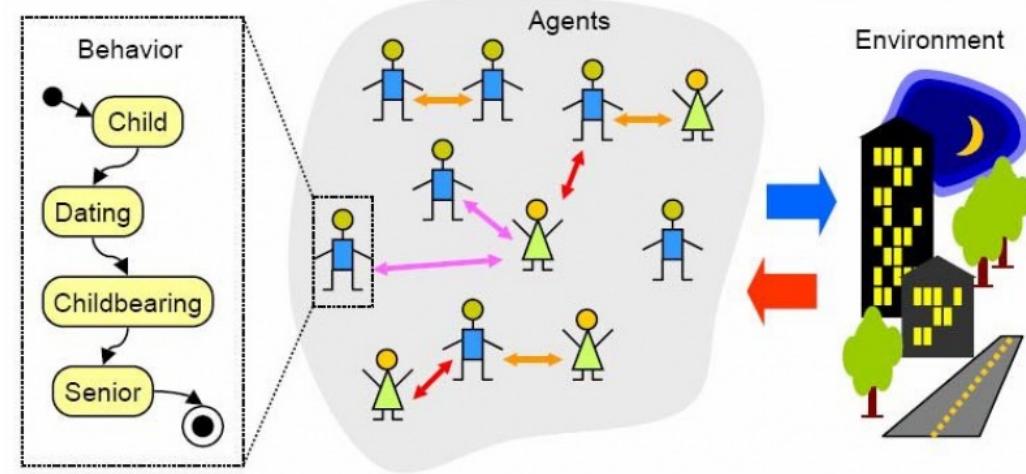


Закон действующих масс:

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

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

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

Прямая задача: найти вектор (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 $q(t) = (\alpha_E(t), \alpha_I(t), \varepsilon_{HC}(t), \mu(t), E_0, I_0)$ по данным $I(t_k) = (1 - b_k)f_k, C(t_k) = C_k, D(t_k) = g_k$, $\beta(t)$ – концентрация антител IgG с учетом доп.ограничений.



Свойства SEIR-HCD модели

Базовый индекс репродукции вируса (начало эпидемии):

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

При $\mathcal{R}_0 < 1$ состояние равновесия без инфекции.

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.

| 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) |
| 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 |

Агентно-ориентированная модель распространения COVID-19: 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):

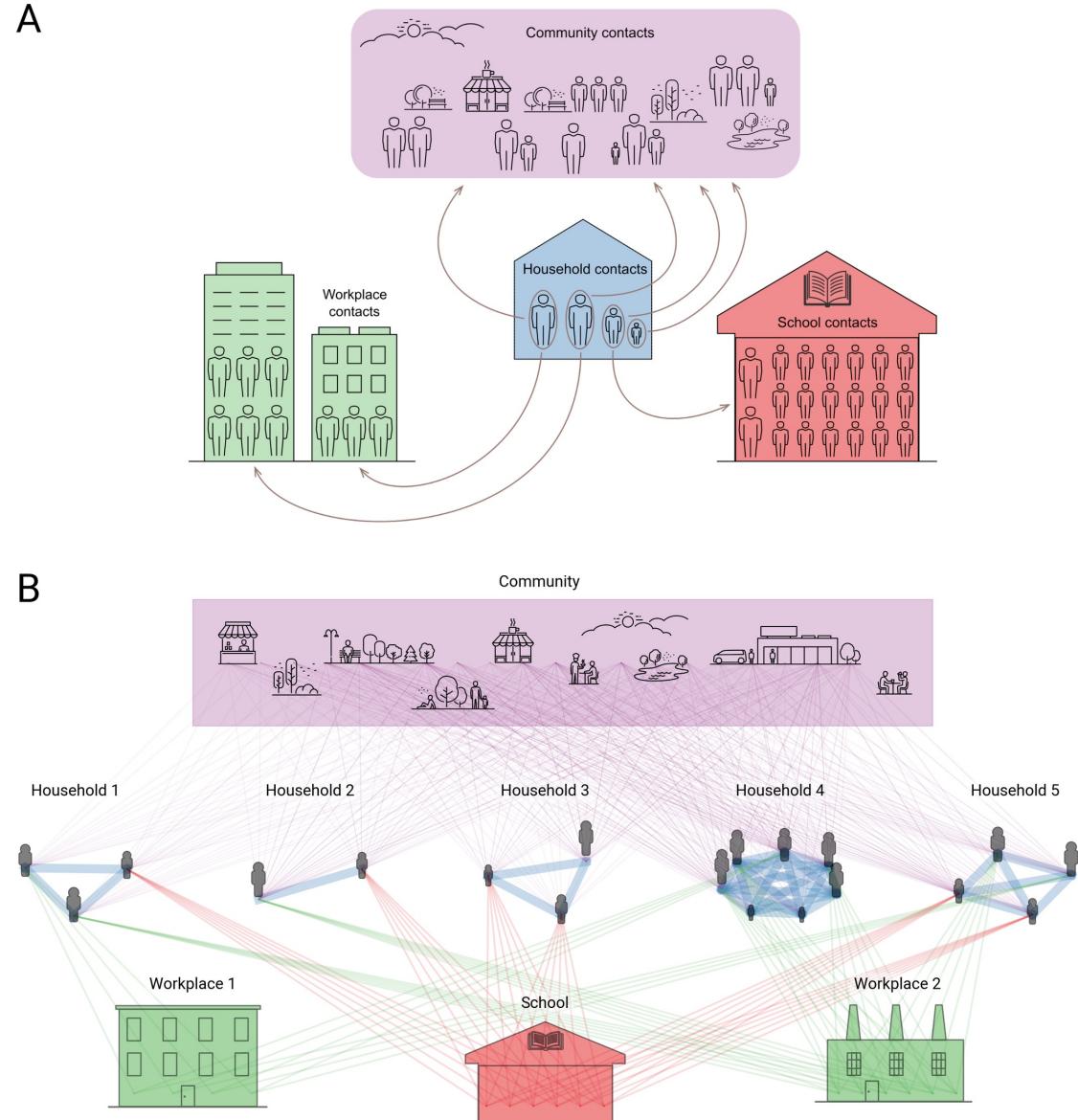
- для домохозяйства – *размер семьи*,
- для общественных мест и образовательных учреждений – 20,
- для работы – 8.

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>

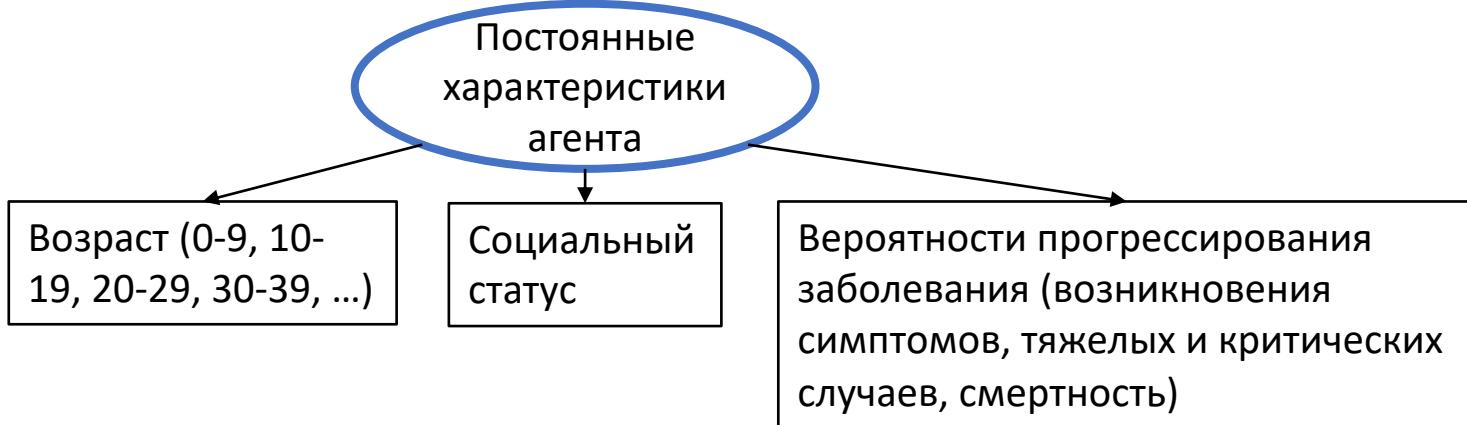
[3] Household Size <https://population.un.org/Household/\#/countries/840>



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) | $\beta \cdot 0.6$ |
| Публичные 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.

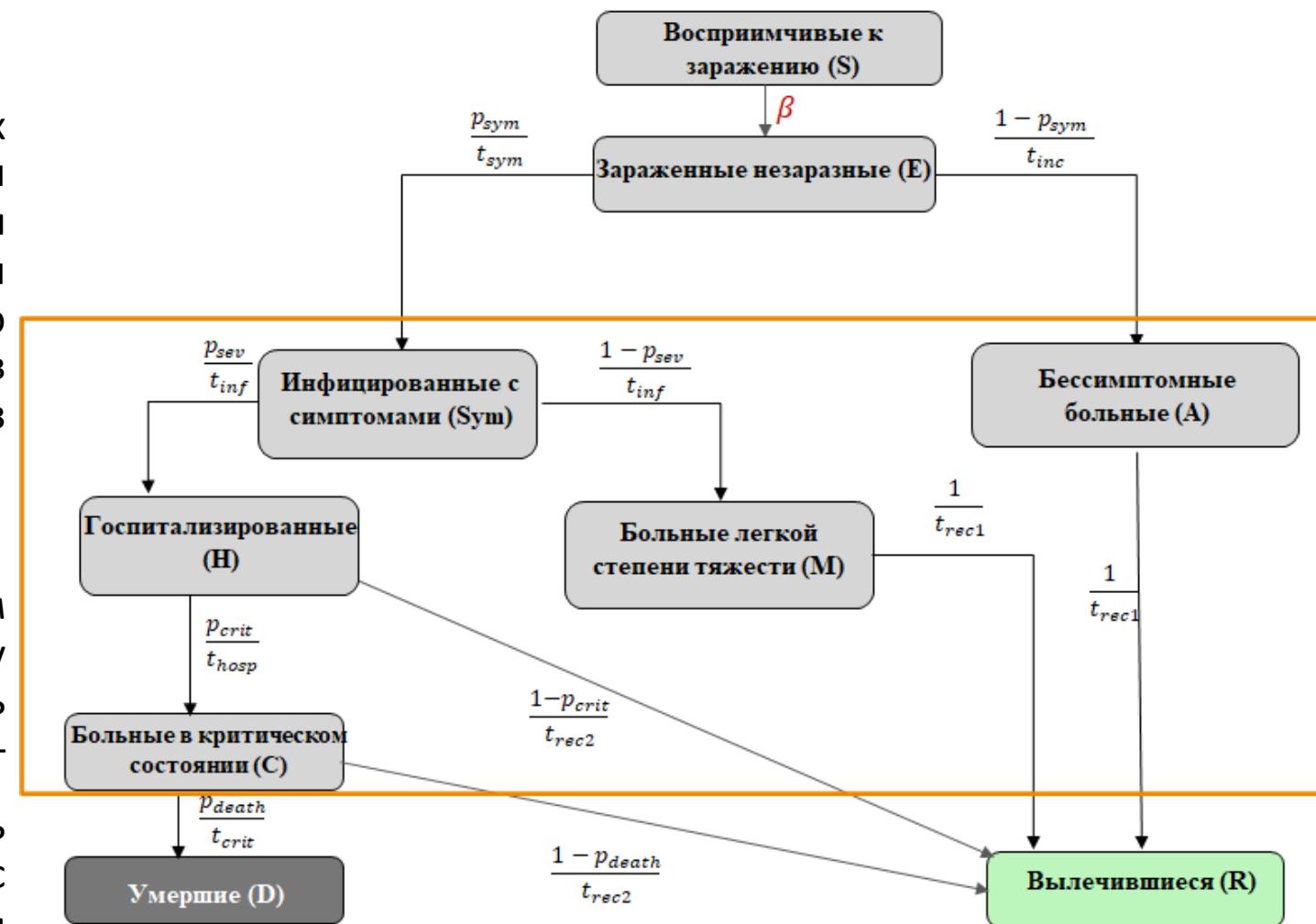
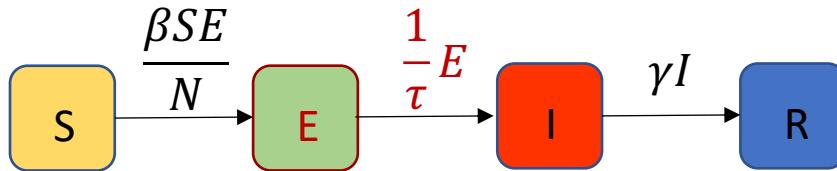


Диаграмма состояний агентов в модели Covasim. Оранжевой рамкой обозначены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19.

Длительность стадий заболевания

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

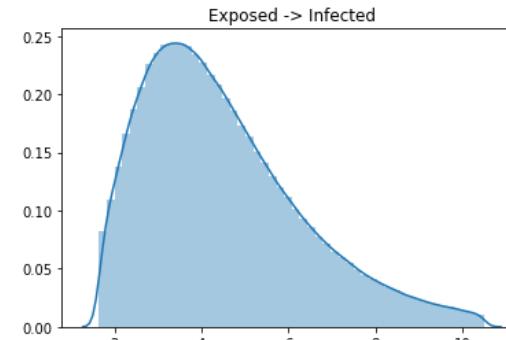
В простейшем случае для SEIR модели



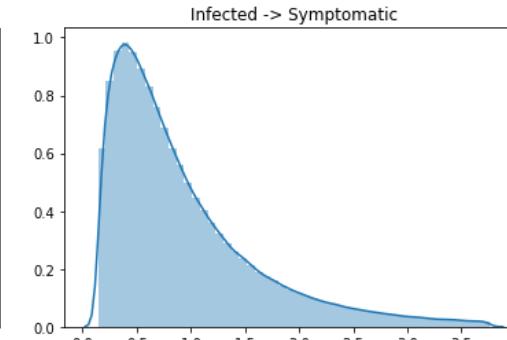
$$\frac{dE}{dt} = \frac{\beta SE}{N} - \frac{1}{\tau} E.$$

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

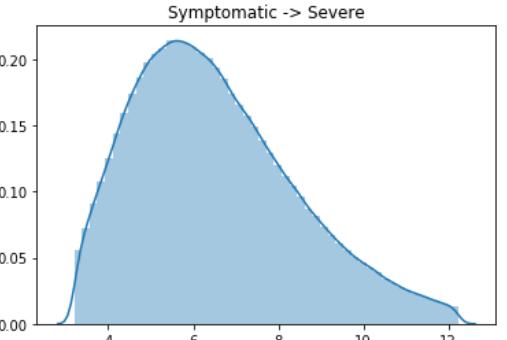
Вероятность $\frac{1}{\tau}$ перехода из группы E в группу I характеризует обратную величину к среднему времени τ пребывания индивидуума в группе E (в линейном приближении).



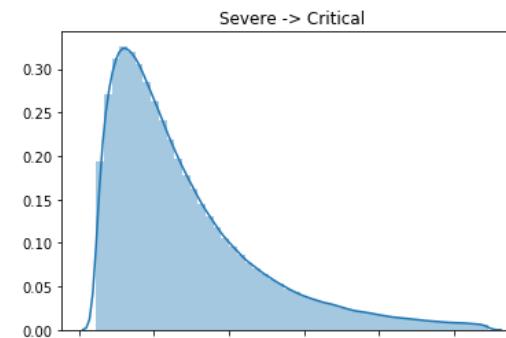
$$t_{inc} \sim \text{LogN}(4.6, 4.8)$$



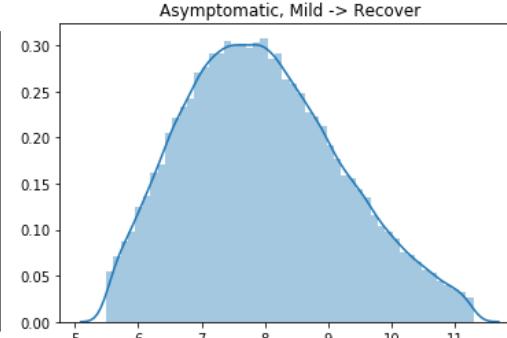
$$t_{sym} \sim \text{LogN}(1.0, 0.9)$$



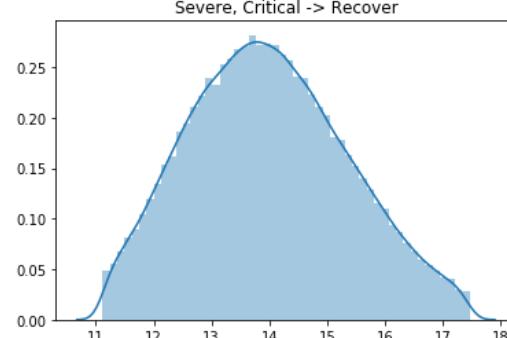
$$t_{inf} \sim \text{LogN}(6.6, 4.9)$$



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



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



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

Течение заболевания: пример заражения восприимчивого агента

Пусть контакт между двумя индивидуумами происходит в домохозяйстве.

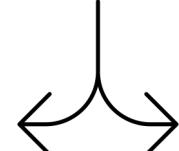
Восприимчивый к заражению (S)



Инфицированный с симптомами (Sym)

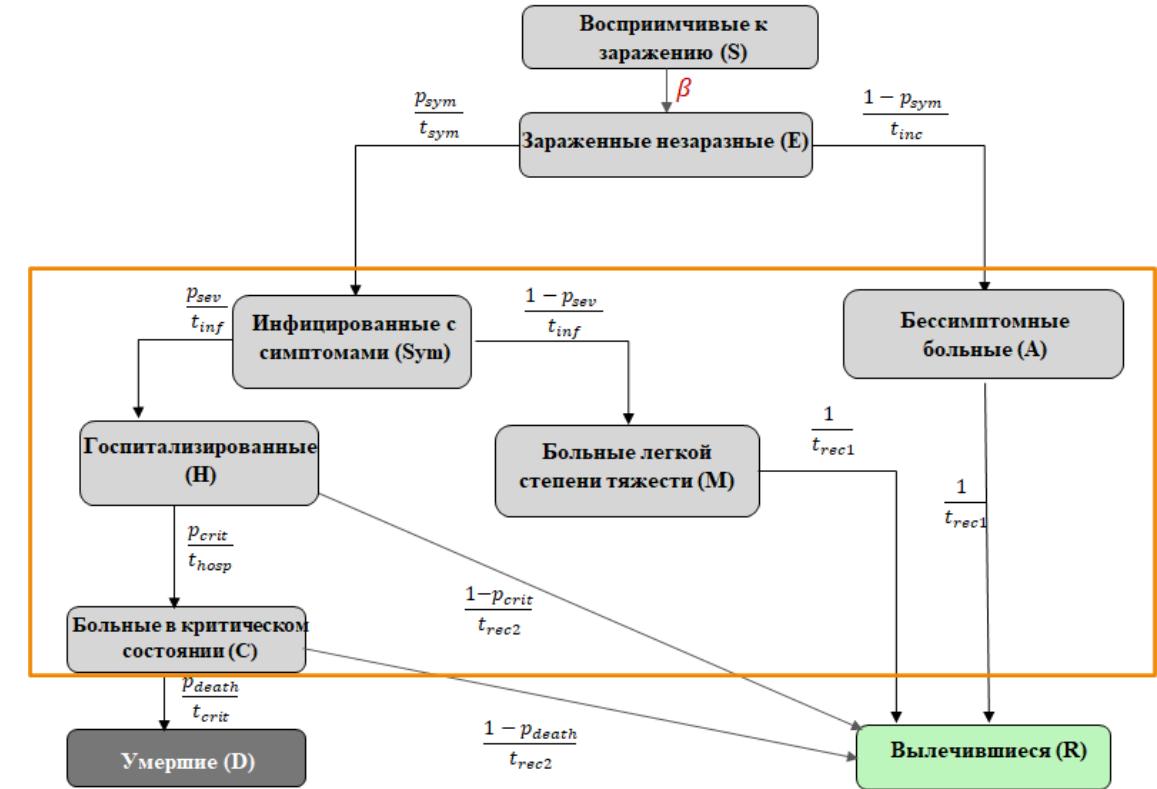


$$3 \cdot \beta$$



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

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



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

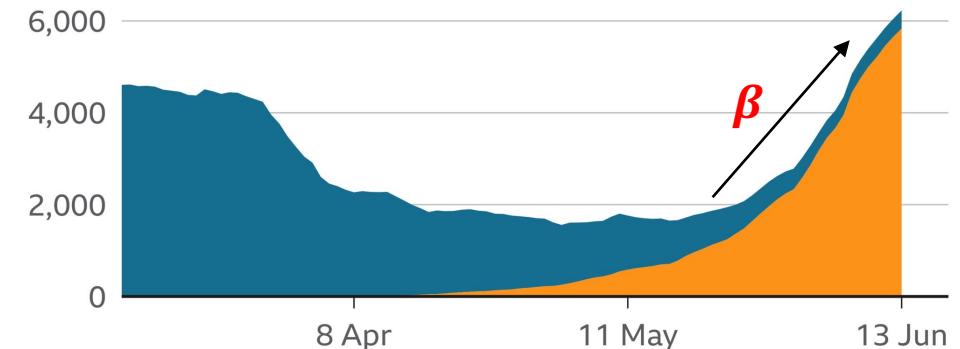
Введение карантинных мер

- изменение значения параметра β (характеризующее уменьшение передачи вируса например при соблюдении масочного режима, соблюдения социальной дистанции);
- удаление ребер в графах (характеризующее самоизоляцию, переход на дистанционное обучение или работу).

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

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).

Обратная задача для агентной модели

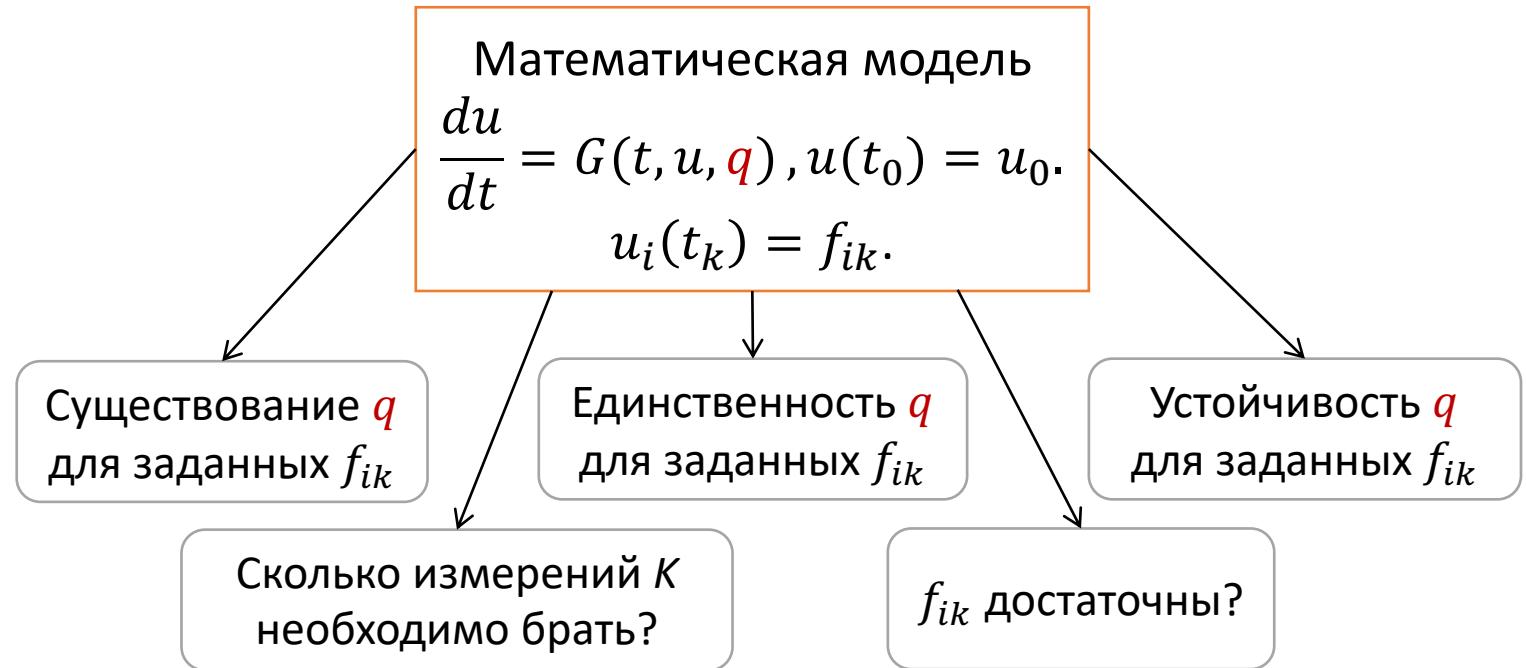
Определить вектор параметров $q(t) = (E(0), \beta, \beta_d, \beta_c, p)$ по дополнительной информации о количестве **ежедневно проведенных ПЦР-тестов, выявленных и умерших случаев от COVID-19** в конкретном регионе, а также усредненных статистических показателей для региона (размер семьи, средняя продолжительность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

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

Сложность применения АОМ возрастает с увеличением моделируемой популяции и учетом внешних воздействий на систему. Усреднение в течение 1 дня моделирования в пределе приводит к SIR модели.

При усложнении SIR-модели, вводя возрастные разграничения в популяции и пространственные перемещения, мы получим первое приближение АОМ.

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

Структурная

- Зависит от структуры модели
- Измерения гладкие

Методы

- Передаточной функции
- Дифференциальной алгебры
- Разложение в ряд Тейлора

Какая комбинация параметров q является идентифицируемой по заданным измерениям f_{ik} ?
Единственность решения обратной задачи.

Практическая

- Зависит от уровня шума в данных
- Необходима аппроксимация параметров

Методы

- Монте-Карло
- Корреляционной матрицы

Какое минимальное число измерений K необходимо для устойчивого решения обратной задачи?

Анализ чувствительности

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

Методы

- Собственных значений
- Ортогональный
- Соболя



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

1. Криворотко 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.
2. Криворотко 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.

Анализ чувствительности для SEIR-HCD и агентной модели COVID-19

$$\frac{du}{dt} = G(t, u, \mathbf{q}), u(t_0) = \mathbf{u}_0$$

$$\mathbf{y} = (f_k, C_k, g_k)^T$$

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

Данные аппроксимируются регрессионной моделью

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

Где параметры регрессии определяются из задачи максимизации функции правдоподобия.

SEIR-HCD

$$\mathbf{q}_1 = (\alpha_E, \alpha_I, \beta_{IgG}, \varepsilon_{HC}, \mu, E_0)^T$$
$$\text{data} = (f_k, C_k, g_k)^T, k = 1, \dots, K$$

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 процентах людей с антителами IgG

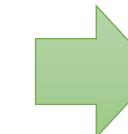
АОМ

$$\mathbf{q}_2 = (\beta, E(0), \tilde{p})^T$$
$$\text{data} = (f_k, C_k, g_k)^T, k = 1, \dots, K$$

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

Алгоритмы решения обратных задач



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

Методы глобальной оптимизации $+$ Градиентные (локальные) методы $=$ $q^* = \arg \min_{\mathbf{q} \in Q} J(\mathbf{q})$

- Природоподобные алгоритмы
- Тензорная оптимизация
- Байесовские методы
- Машинное обучение

Выражение для градиента целевого функционала

- Регуляризация А.Н. Тихонова
- Численная регуляризация
- Гиперболическая регуляризация (для параболических уравнений и систем)

Тензорная оптимизация в случае пространства параметров большой размерности

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

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

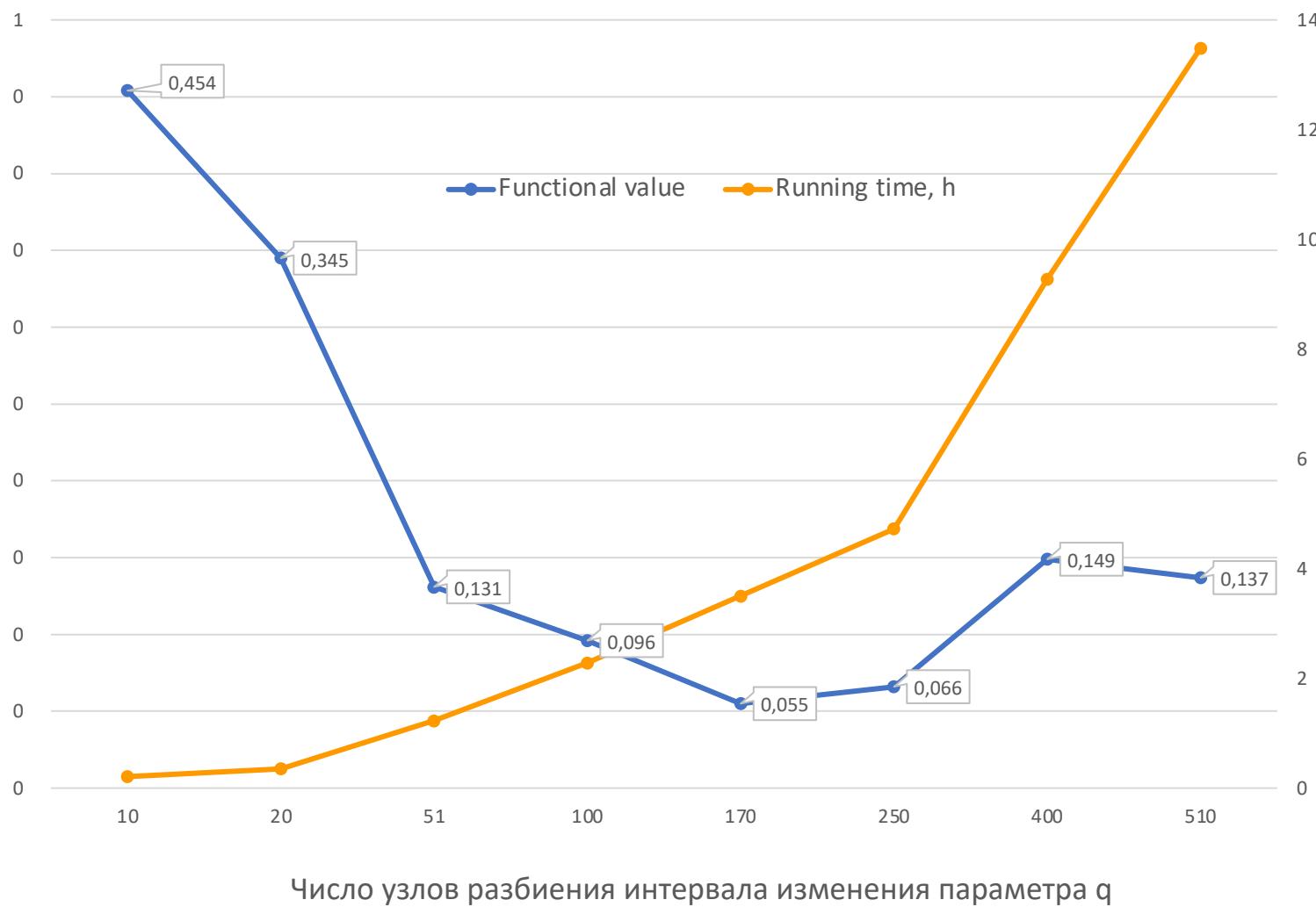
$h(a)$ – непрерывная монотонно убывающая функция, α – некоторое начальное приближение к минимуму функционала $J(q)$.

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 $d=6$ 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.

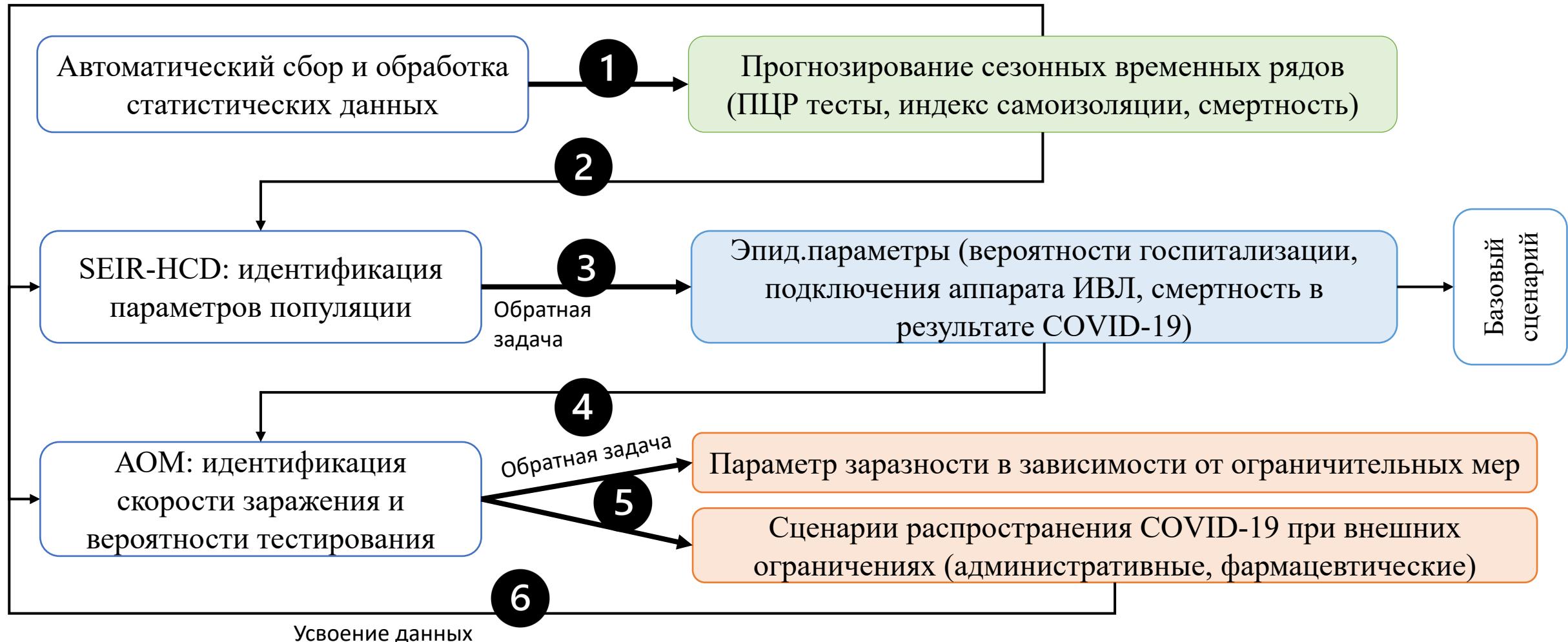
4.3. Вычислительная сложность тензорной оптимизации



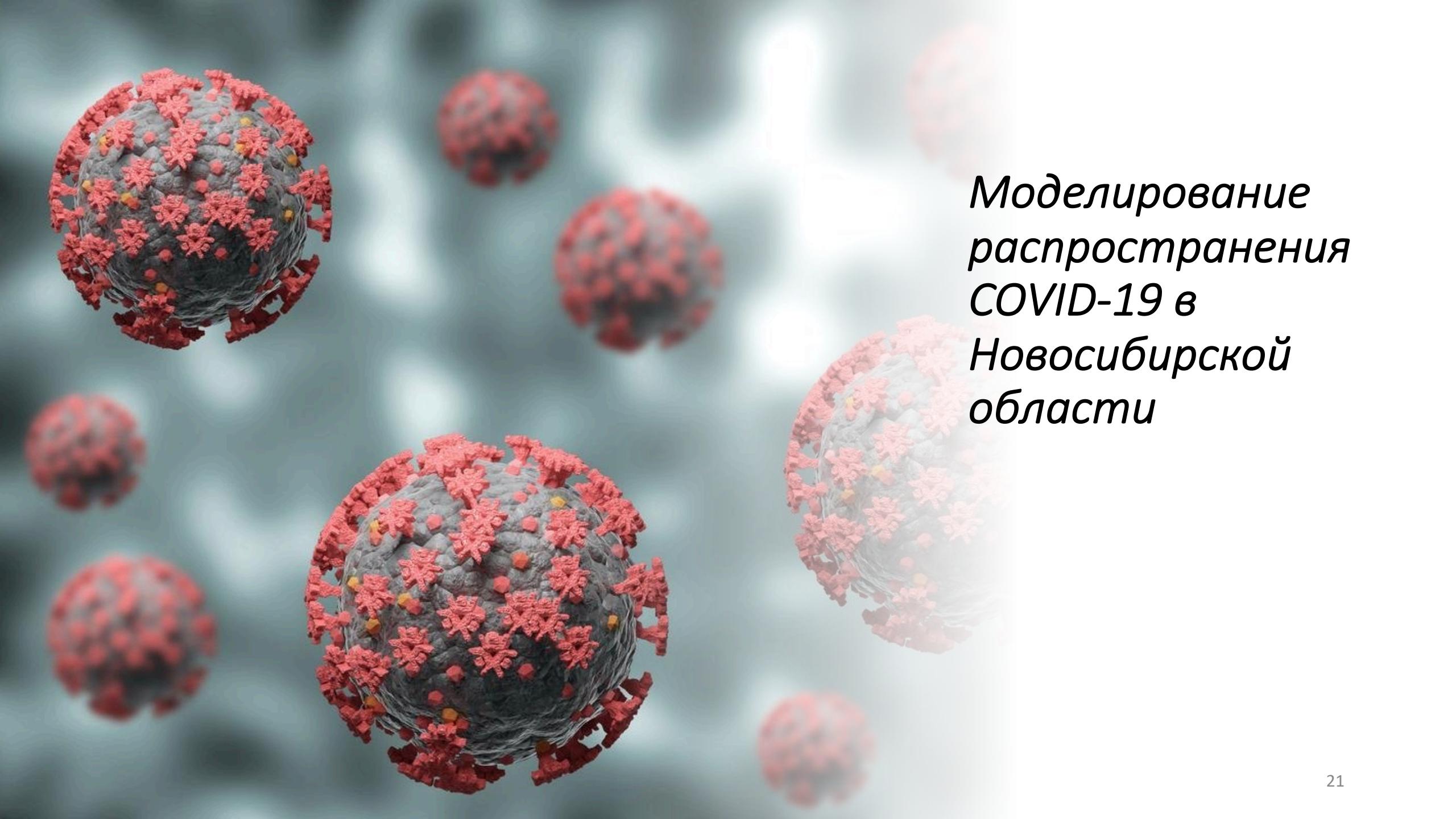
Характеристики вычислительного узла в ССКЦ СО РАН:

2 процессора Intel Xeon Gold 6248R (3 GHz, 24 ядра)
384 GB оперативной памяти

При увеличении размерности тензора вычислительное время растет экспоненциально.



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

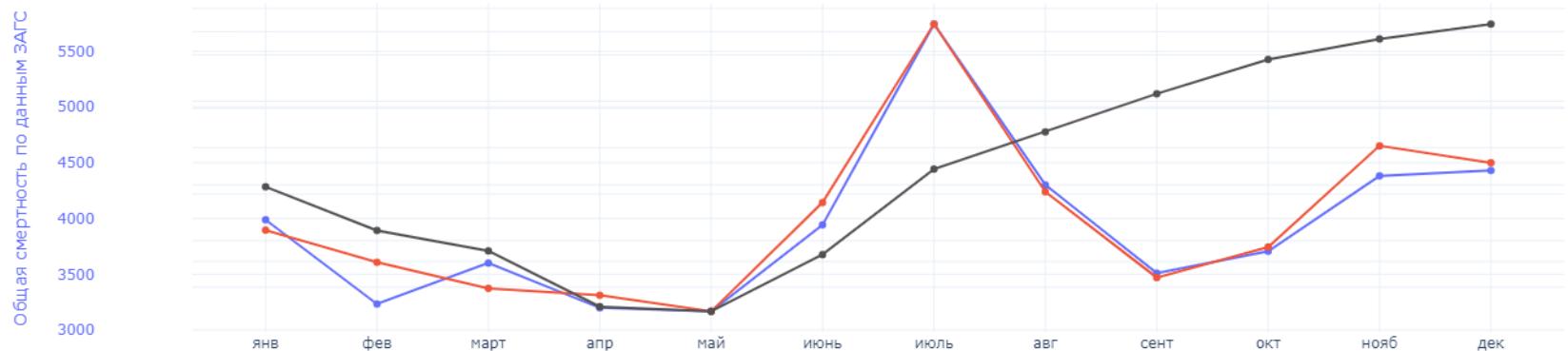


*Моделирование
распространения
COVID-19 в
Новосибирской
области*

Эпидемиологические показатели в Новосибирской области



Смертность в 2021



Черная линия – смертность от COVID-19, **красная линия** – захоронения с диагнозом COVID-19, **синяя линия** – общая смертность

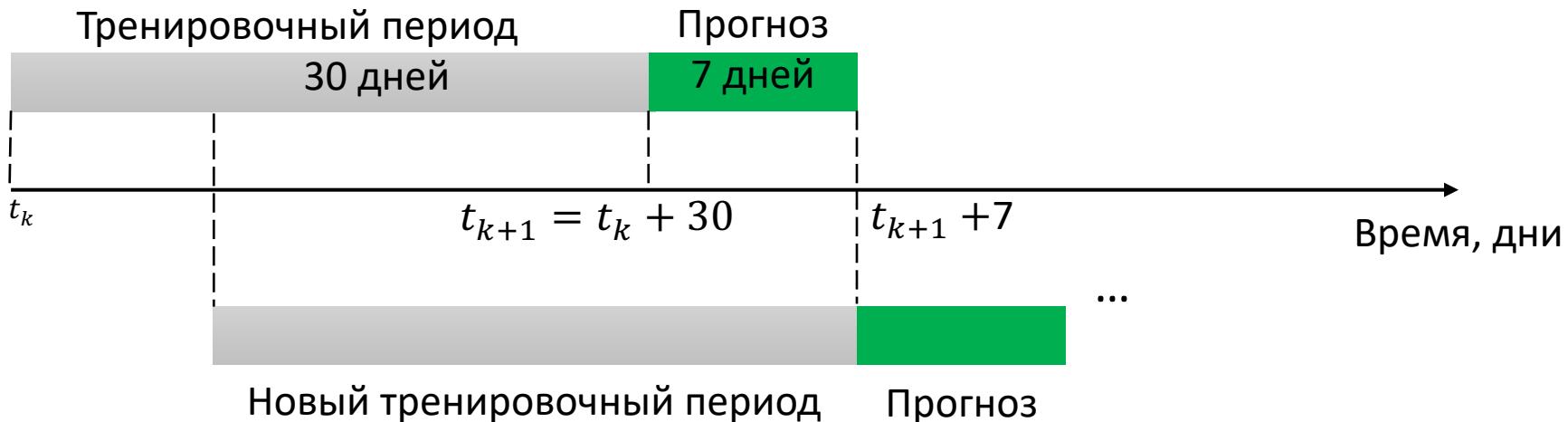
<http://covid19-modeling.ru/>

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**).

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

Алгоритм усвоения данных

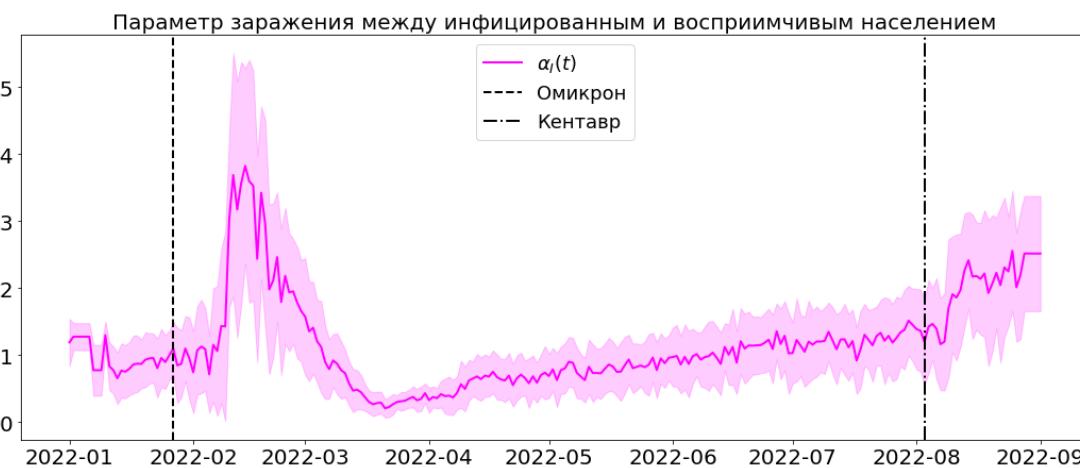
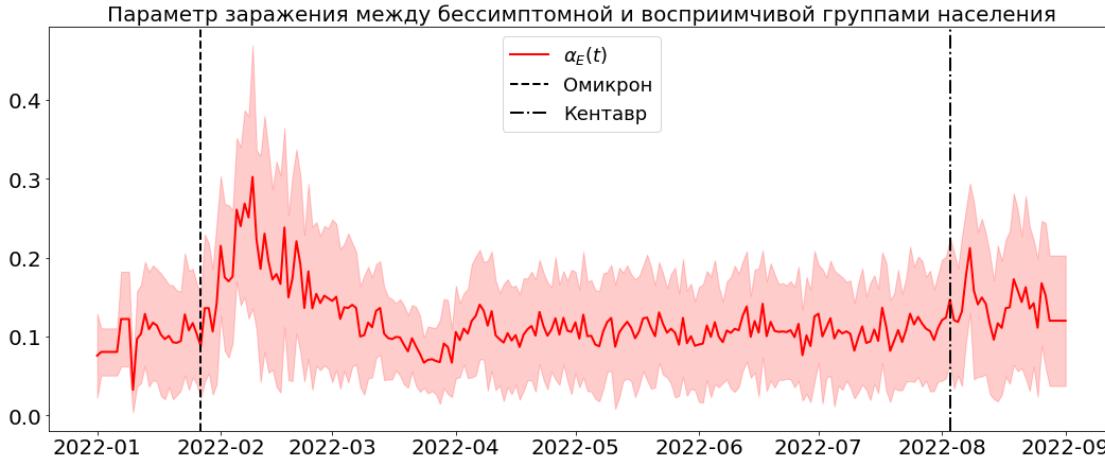


Тренировочный период (30 дней) – уточнение параметров q .

Прогноз (7 дней) рассчитывается при найденных q .

Новый период сдвигается на 7 дней и снова решается обратная задача (уточнение параметров q).

Динамика эпидемиологических параметров SEIR-HCD модели в Новосибирской области с 01.2022 по 09.2022.



Скорость распространения вируса SARS-CoV-2 при штамме «Омикрон» увеличивается как для симптомных, так и для бессимптомных инфицированных. Сохраняется уровень смертности в регионе. Его естественная мутация, штамм «Кентавр» не обладает высокой заразностью (только от симптомной группы), смертность уменьшается.

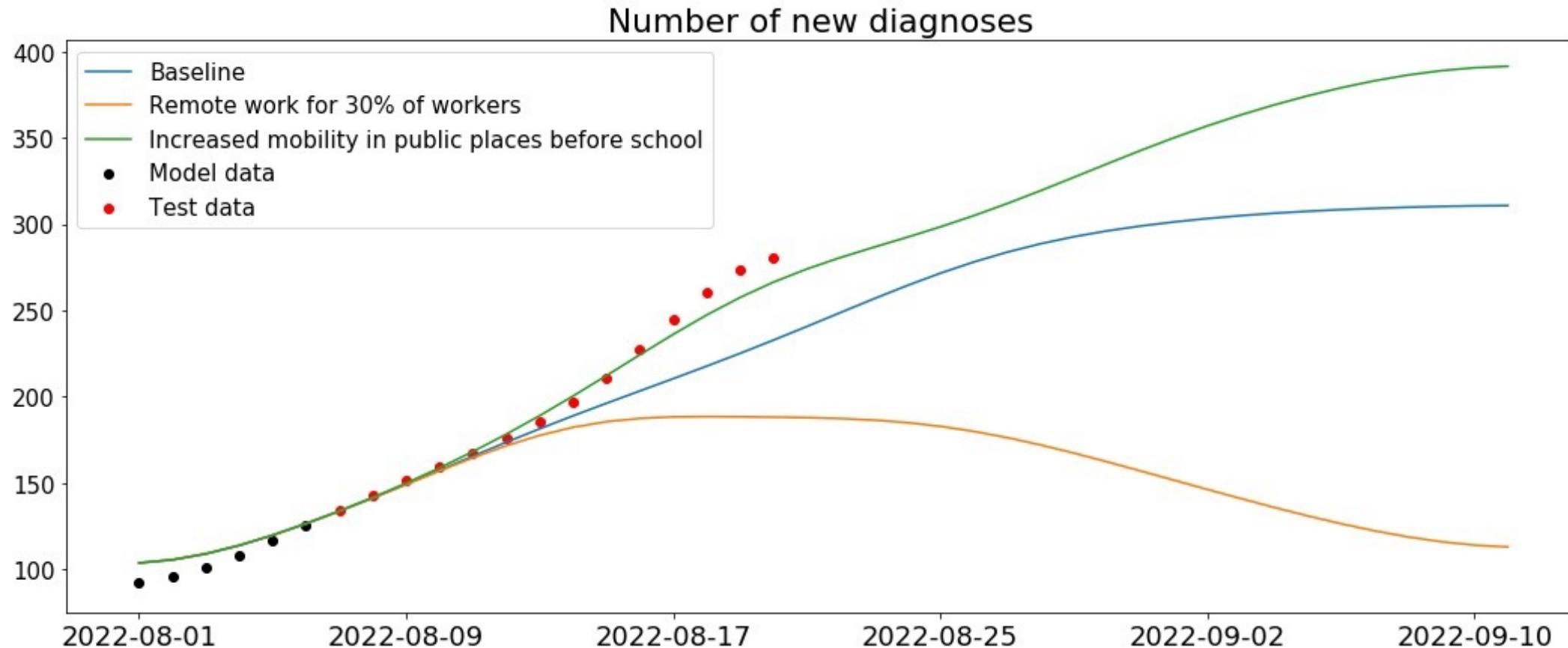
Влияние ограничительных мер и появления новых штаммов на изменение параметра заразности в НСО с 04.2020 по 08.2022.

Изменение параметра заразности в Новосибирской области



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.

Черные точки - реальные данные по 06.08.2022, участвующие в решении обратной задачи (<https://covid19-modeling.ru/>).

Красные точки – тестовые данные с 07.08.2022 по 22.08.2022 (для верификации модели).



Russian Science
Foundation



RUSSIAN
FOUNDATION
FOR BASIC
RESEARCH

Спасибо за внимание!



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



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