

Ю.С. Попков, Ю.А. Дубнов, А.Ю. Попков  
**ПРОГНОЗИРОВАНИЕ РАЗВИТИЯ ЭПИДЕМИИ COVID-19 В  
СТРАНАХ ЕВРОПЕЙСКОГО СОЮЗА С ИСПОЛЬЗОВАНИЕМ  
ЭНТРОПИЙНО-РАНДОМИЗИРОВАННОГО ПОДХОДА**

*Попков Ю.С., Дубнов Ю.А., Попков А.Ю.* Прогнозирование развития эпидемии COVID-19 в странах Европейского союза с использованием энтропийно-рандомизированного подхода.

**Аннотация.** Работа посвящена прогнозированию развития эпидемии COVID-19 с помощью нового метода рандомизированного машинного обучения. Основу метода составляет идея оценивания распределений вероятностей параметров модели по реальным данным вместе с распределением вероятностей измерительных шумов. Энтропийно-оптимальные распределения соответствуют состоянию максимальной неопределенности, что позволяет использовать получаемые в итоге прогнозы, как прогнозы наиболее “негативного” сценария исследуемого процесса. Полученные оценки параметров и шумов, которые представляют собой распределения вероятностей, необходимо генерировать, получая таким образом ансамбль траекторий, который требуется анализировать статистическими методами. Для целей такого анализа проводится вычисление средней и медианной по ансамблю траектории, а также траектории, соответствующей средним по распределению значениям параметров модели. Предлагаемый подход используется для прогнозирования общего количества инфицированных с помощью трехпараметрической логистической модели роста. Проведенный эксперимент основан на реальных данных о распространении COVID-19 в нескольких странах Европейского союза. Основной целью эксперимента является демонстрация энтропийно-рандомизированного подхода для прогнозирования эпидемического процесса на основе реальных данных вблизи пика. Существенная неопределенность, содержащаяся в доступных реальных данных моделируется аддитивным шумом в пределах 30%, который используется как на этапе обучения модели, так и при прогнозировании. Для настройки гиперпараметров модели используется схема их настройки по тестовой выборке с последующим переобучением. Показано, что при одинаковых наборах данных, предлагаемый подход позволяет более эффективно прогнозировать развитие эпидемии по сравнению со стандартным подходом, основанным на методе наименьших квадратов.

**Ключевые слова:** моделирование эпидемий, SARS-CoV-2, COVID-19, рандомизированное машинное обучение, энтропия, энтропийное оценивание, прогнозирование, рандомизированное прогнозирование.

**1. Введение.** Эпидемия новой коронавирусной инфекции началась предположительно в ноябре-декабре 2019 года в Китае, и уже к февралю 2020 года Всемирная организация здравоохранения присвоила ей статус пандемии, тем самым подтвердив ее глобальное распространение. Быстрый рост числа инфицированных и высокая смертность от вызываемой вирусом SARS-CoV-2 болезни COVID-19 привели к перегрузке систем здравоохранения почти во всех странах, включая страны Европейского союза и США. К марту 2020 года правительствам всех этих стран, в том числе и России, пришлось прибегать к крайним противоэпидемиче-

ским мерам, включая тотальный локдаун, запрет на перемещения людей, закрытие границ между регионами и государствами.

С самого начала пандемии начали проводиться научные исследования, направленные на прогнозирование развития эпидемии, основными целями которых являлась оценка масштабов эпидемии в кратко- и среднесрочной перспективе. Понимание эпидемических процессов и их прогнозирование необходимо для оценки готовности системы здравоохранения, принимаемых мер сдерживания эпидемии, состояния экономики.

Основными подходами к моделированию эпидемического процесса на сегодняшний день являются подходы, основанные на применении статических моделей подгонки кривой под имеющиеся данные; динамические модели, основанные на системах дифференциальных уравнений, описывающих динамику основных показателей эпидемии и сетевые модели, ориентированные на моделировании динамики эпидемических процессов с учетом неоднородной структуры общества, что отражается на неравномерном развитии эпидемии в разных группах населения.

Большинство подходов первого типа базируются на идеях, восходящих к моделям популяций в биологических системах, начиная с моделей экспоненциального роста конца 19 – начала 20 веков [1]. Модели экспоненциального роста показывают свою эффективность на начальном этапе развития эпидемии, когда происходит резкий рост числа инфицированных. Со временем, когда появляются естественные (например, увеличение доли иммунных членов популяции) или искусственные ограничения этого процесса (например, введение физических барьеров для ограничения контактов между членами популяции), происходит снижение количества инфицированных. Этот переломный момент в развитии эпидемического процесса, а также его дальнейшую динамику, простые модели экспоненциального роста не могут предсказать. По этой причине, в современном мире, в котором вводятся различные меры по сдерживанию эпидемии, применяются в основном модели логистического типа, которые используются для моделирования общего количества инфицированных. В начале 2020 года появилось большое количество исследований, основанных на логистических моделях, которые описывали эпидемический процесс с хорошим уровнем качества [2–8].

В настоящее время, практически во всех странах на уровне правительств используются динамические модели прогнозирования развития эпидемии, которые более адекватно описывают эпидемический процесс. Не вдаваясь в подробности разнообразных динамических моделей, можно отметить такое их общее свойство, как чувствительность их параметров, что приводит к новой проблеме адекватной и устойчивой их оценке,

без которой эффективное функционирование динамических моделей невозможно.

На начальном этапе эпидемии, исследование которого является предметом настоящей работы, произвести качественную оценку параметров динамических моделей не представлялось возможным из-за недостатка надежных реальных данных и понимания многих аспектов поведения вируса SARS-CoV-2 и болезни COVID-19. По этой причине, на рассматриваемом этапе развития эпидемии получили широкое распространение статические модели, в том числе логистического типа. Результаты применения этих моделей использовались многими исследователями для оценки параметров динамических моделей.

В работе рассматривается применение нового подхода к обучению моделей с использованием реальных данных и прогнозированию, называемого *Рандомизированным машинным обучением* (РМО) [9]. Основным достоинством этого метода является независимость от реальных характеристик используемых данных. Для корректного применения метода не требуется подтверждения или предположения о нормальности данных (или иных их вероятностных свойствах), а полученные в результате обучения распределения получены в условиях максимальной энтропии (максимальной неопределенности), таким образом отражая наиболее “плохой” сценарий развития исследуемого процесса. Эти свойства энтропийного подхода установлены в работах Больцмана [10], Джейнса [11, 12], Шеннона [13]. Еще одной важной особенностью метода является получение, вместе с оптимальными распределениями параметров, энтропийно-оптимальных распределений шумов (стохастических компонент данных), содержащихся в данных. Это свойство существенно отличает метод от классических подходов, в которых делаются различные предположения о характеристиках шумов.

**2. Исходные данные и модели.** Одним из важных и основных индикаторов эпидемии является общее (накопленное, куммулятивное) количество инфицированных. В случае эпидемии инфекционной болезни, важность этого показателя становится еще больше, так как количество инфицированных необходимо для оценивания и прогнозирования медицинских ресурсов, а также для оценивания других показателей эпидемического процесса, в том числе и его динамических характеристик.

Для моделирования рассматриваемого показателя в настоящее время (на начало 2021 года) в большинстве случаев применяются динамические модели, основанные на дифференциальных уравнениях с параметрами, которые оцениваются на основе реальных данных или с помощью других моделей. Статические модели, основанные на нели-

нейных уравнениях, параметры которых предполагается оценивать по реальным данным, стали популярными на начальной стадии эпидемии COVID-19 главным образом из-за недостатка реальных данных, и невозможности вследствие этого качественно оценить основные параметры динамических моделей. Кроме этого, опыт прошлых эпидемий, а также идея о снижении доли инфицированных в популяции из-за выздоровления или смерти, позволяет предположить, что кривая заражений будет иметь форму логистической кривой с участками плавного роста, быстрого роста и последующего снижения количества зараженных.

**2.1. Модель.** Динамика инфицированных членов популяции  $N$  в биологической системе может быть описана следующим уравнением [1]:

$$\frac{dN}{dt} = \lambda N \left( 1 - \frac{N}{K} \right), \quad (1)$$

где  $\lambda$  — скорость роста инфицированных,  $N$  — количество инфицированных,  $K$  — объем популяции. Решением этого уравнения является кривая Ферхюльста [14, 15]

$$N(t) = \frac{K}{1 + B e^{-\lambda t}}, \quad B = \frac{K - N_0}{N_0}, \quad (2)$$

где  $N_0$  — количество инфицированных в популяции в начальный момент времени [15].

Уравнение вида (2) активно использовалось в начале 2020 г. для построения предсказательных моделей заражения (предсказания общего количества заболевших) [2–8] и показали свою эффективность на начальном этапе развития эпидемии. Эти модели относятся к так называемым *логистическим моделям роста* (Logistic Growth Model, LGM), которые применяют для описания накопленного количества зараженных членов замкнутой популяции (кривую заражения).

Рассмотрим данную модель в общем виде. Для этого будем использовать уравнение трехпараметрической логистической кривой, которая определяет преобразование скалярного входа  $x$  в выход  $\hat{y}$  с использованием логистической нелинейной функции

$$\hat{y} = \Phi(x, \mathbf{a}) = \frac{a_3}{1 + a_1 e^{-a_2 x}}, \quad (3)$$

где  $\mathbf{a} = (a_1, a_2, a_3)$  — вектор параметров модели. Входом модели является порядковый номер (или индекс) дня, а выходом — накопленное количество инфицированных.

Основной задачей для использования модели в прогнозировании, является ее обучение на реальных данных (оценивание ее параметров). После получения оценок параметров, возникает следующая задача, связанная с ее использованием для получения прогноза. Стандартные методы машинного обучения (статистического оценивания) состоят в вычислении точечных или интервальных оценок параметров модели, с последующим построением прогноза с использованием обученной модели. Этот подход является в большинстве случаев эффективным, однако при использовании нелинейных моделей он сопряжен с существенными трудностями в установлении свойств получаемых оценок, которые требуется учитывать при построении прогноза.

Подход к оцениванию параметров модели, предлагаемый в работе, основан на теории *рандомизированного машинного обучения* (РМО) [9, 16, 17], который предполагает получение не точечных оценок параметров, а их распределений вероятностей, причем вместе с распределениями вероятностей шумов, неизбежно содержащихся в данных. Важно отметить, что никаких априорных предположений о вероятностных свойствах реальных данных, которые бы учитывались при оценивании, не требуется. Для построения прогноза необходимо генерировать полученные оценки распределений параметров и шумов, что приводит к вычислению не одной прогнозной траектории, а ансамбля траекторий, который требуется анализировать статистическими методами.

Применение РМО приводит к построению *рандомизированной модели*, которая определяет специальную методику ее использования при прогнозировании — *рандомизированное прогнозирование* [18–21].

При применении методов математического моделирования эффективным является путь, который состоит в использовании непрерывных математических объектов и, соответственно, непрерывных моделей. Это обусловлено как достижениями в различных областях математики, так и точностью, которую могут обеспечить только непрерывные объекты. В то же время, работа в рамках непрерывной теории сопряжена с существенными вычислительными трудностями в случае многомерных объектов. Одним из путей решения этой проблемы является переход к дискретным объектам. Адаптация теории РМО на дискретный случай для нелинейных моделей была предложена в [22]. При применении этого подхода, можно построить *рандомизированную логистическую модель роста (РЛМР)*, которая является частным случаем *нелинейной рандомизированной дискретной статической модели (НРДСМ)*:

$$v = \hat{y} + \xi = \Phi(x, \mathbf{a}) + \xi, \quad (4)$$

где  $\Phi$  определяется (3),  $x$  определяет вход модели, которым в рассматриваемой задаче является порядковый номер дня,  $\xi$  — шум при каждом измерении входа, действующий аддитивно,  $\hat{y}$  — выход модели (общее количество инфицированных),  $v$  — выход искаженный шумом.

Параметры и шумы реализуются интервальными дискретными случайными величинами с распределениями

$$a_{k\ell} \in A_k, \quad p_{k\ell} \in [0,1], \quad k = \overline{1,d}, \quad \ell = \overline{1,M}, \quad (5)$$

$$\xi_{jh} \in \Xi_j, \quad q_{jh} \in [0,1], \quad j = \overline{1,m}, \quad h = \overline{1,L}, \quad (6)$$

где  $a_{k\ell}$ ,  $\xi_{jh}$  — значения случайных величин,  $p_{k\ell}$ ,  $q_{jh}$  — вероятности их реализации,  $A_k$ ,  $\Xi_j$  — интервалы значений случайных величин,  $m$  — количество точек данных,  $d = 3$ .

**2.2. Данные.** Для обучения и прогнозирования используются данные Университета Джонса Хопкинса [23, 24] о ежедневном количестве инфицированных. На рисунке 1 представлены данные вместе с 7-дневным скользящим средним для нескольких стран Европейского союза. Из данных видно, что зимой–весной 2020 года во всех странах наблюдался пик заражений, после которого произошел спад, продолжавшийся все лето, после которого возникла так называемая “вторая волна”. Учитывая, что исследования, проводимые в настоящей работе, направлены на моделирование и прогнозирование роста заражений вблизи пика, для обучения модели выбирались данные за 7 дней до пика “первой волны”, положение пика определялось по максимуму скользящего среднего за 7 дней.

Для обучения данные были масштабированы на отрезок  $[0,1]$  с целью уменьшения сложностей, связанных с переполнением при вычислениях. При прогнозировании полученный выход модели преобразовывался к исходному масштабу.

**2.3. Обучение.** Рандомизированное машинное обучение предполагает обучение модели (3) с использованием реальных данных для получения оптимальных распределений параметров модели. Используемая здесь техника энтропийного оценивания позволяет определять такие распределения, которые доставляют максимум их энтропии. Полученные энтропийно-оптимальные распределения будут отражать наиболее неопределенный («плохой») сценарий, что в условиях полного отсутствия информации о реальных характеристиках является единственным доступным в этих условиях решением [9, 25, 26].

Вычисление оптимальных распределений обеспечивается решением задачи условной максимизации энтропии распределений параметров и шумов измерений при условиях нормировки соответствующих рас-

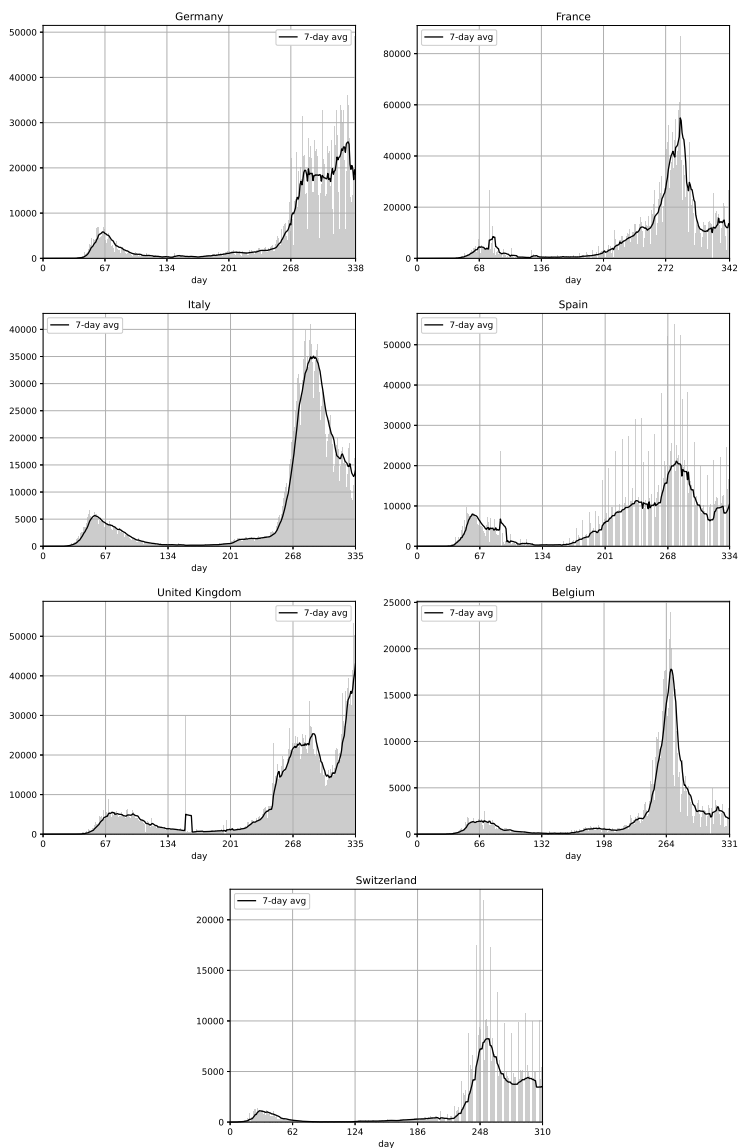


Рис. 1. Дневное количество инфицированных

пределений и выполнении условий на баланс среднего выхода модели с измерением выхода объекта, которая формулируется следующим образом [9]:

$$H(P, Q) = - \sum_{k=1}^d \sum_{\ell=1}^M p_{k\ell} \ln p_{k\ell} - \sum_{j=1}^m \sum_{h=1}^L q_{jh} \ln q_{jh} \rightarrow \max_{P, Q}, \quad (7)$$

где  $P$  и  $Q$  — распределения параметров и шумов (5) и (6), а  $d = 3$  при условиях:

$$\sum_{\ell=1}^M p_{k\ell} = 1, \quad \sum_{h=1}^L q_{jh} = 1, \quad k = \overline{1, d}, \quad j = \overline{1, m}, \quad (8)$$

$$\mathbb{E}[v_j] = \mathbb{E}[\Phi(x_j, \mathbf{a}) + \xi_j] = y_j, \quad (9)$$

где  $y_j$  — реальные данные выхода,  $\Phi(x, \mathbf{a})$  определяется (3),  $\xi_j$  — шум.

Условие (9) определяет баланс среднего выхода модели с реальными данными выхода:

$$\begin{aligned} \mathbb{E}[v_j] &= \mathbb{E}[\Phi(x_j, \mathbf{a}) + \xi_j] = \mathbb{E}[\Phi(x_j, \mathbf{a})] + \mathbb{E}[\xi_j] = \\ &= \sum_{\substack{\ell_k=1 \\ k=\overline{1, d}}}^M \Phi(x_j, a_{1\ell_1}, \dots, a_{d\ell_d}) p_{1\ell_1} \cdots p_{d\ell_d} + \sum_{h=1}^L \xi_{jh} q_{jh} = \\ &= \bar{\Phi}(x_j) + \sum_{h=1}^L \xi_{jh} q_{jh} = y_j, \quad j = \overline{1, m}. \end{aligned} \quad (10)$$

Сумма в выражении для  $\bar{\Phi}$  содержит  $M^d$  членов, суммирование осуществляется для всех комбинаций значений случайных величин  $a_{k\ell}$ .

Для решения задачи (7)-(9) используем метод множителей Лагранжа, применение которого позволяет получить выражения энтропийно-оптимальных распределений параметров и шумов измерений, зависящих



от параметра  $\lambda$  (множителей Лагранжа):

$$p_{k\ell}^*(\lambda) = \frac{\exp\left(-\sum_{j=1}^m \lambda_j \frac{\partial \bar{\Phi}_j}{\partial p_{k\ell}}\right)}{\sum_{\ell=1}^M \exp\left(-\sum_{j=1}^m \lambda_j \frac{\partial \bar{\Phi}_j}{\partial p_{k\ell}}\right)}, \quad k = \overline{1, d}, \ell = \overline{1, M}, \quad (11)$$

$$q_{jh}^*(\lambda) = \frac{\exp(-\lambda_j \xi_{jh})}{\sum_{h=1}^L \exp(-\lambda_j \xi_{jh})}, \quad j = \overline{1, m}, h = \overline{1, L}. \quad (12)$$

Множители  $\lambda$  определяются решением системы уравнений, получающихся подстановкой выражений (11)-(12) в балансовые соотношения (9):

$$\sum_{\substack{\ell_k=1 \\ k=1, d}}^M \Phi(\mathbf{x}_j, a_{1\ell_1}, \dots, a_{d\ell_d}) \prod_{\substack{\ell_s=1 \\ s=1, d}}^M p_{s\ell_s}^*(\lambda) + \sum_{h=1}^L \xi_{jh} q_{jh}^*(\lambda) = y_j, \quad j = \overline{1, m}. \quad (13)$$

Решение системы (13) позволяет определить требуемые энтропийно-оптимальные распределения параметров и шумов измерений. Для решения этой системы, очевидно, требуется привлекать численные методы, так как ее решение аналитически сопряжено с существенными трудностями. В работе для этой цели применяется усеченный метод Ньютона ТНС [27, 28].

**2.4. Прогнозирование.** В результате обучения модель оказывается снабжена энтропийно-оптимальными оценками распределений параметров и измерительных шумов, формируя таким образом *рандомизированную предсказательную модель (РПМ)*. Применение РПМ для прогнозирования основано на специальной методике *рандомизированного прогнозирования*, которая определяет технологию использования энтропийно-оптимальных распределений параметров и шумов для построения прогноза. Указанные распределения можно использовать двумя способами. Первый состоит в использовании их средних значений в качестве точечных оценок параметров и шумов модели (4). Второй способ обладает большей эффективностью и состоит в генерации указанных распределений и построением ансамбля выхода модели для каждой rea-

лизации соответствующих случайных величин. Этот способ используется в вычислениях в настоящей работе, наряду с первым способом, который участвует в сравнительном исследовании. Для выполнения сравнительного исследования также используется стандартный способ прогнозирования, основанный на подгонке кривой нелинейным методом наименьших квадратов.

Вычисляя выход РПМ для каждой реализации параметров и шумов, мы получаем ансамбль траекторий, причем, для каждой реализации параметров, соответствующий шум в каждой точке данных генерируется независимо в количестве нескольких экземпляров.

Пусть имеется выборка параметров из распределения  $P^*$  объема  $S$ , определяемого (11). Для каждой реализации параметра  $\mathbf{a}_i$ ,  $i = \overline{1, S}$  генерируется  $U$  экземпляров шума с распределением  $Q^*$  определяемого (12) для каждого входа РПМ. Используя эти параметры и шумы, ансамбль выхода модели будет определяться блочным вектором или матрицей

$$\mathcal{V} = [\mathbf{v}_1, \dots, \mathbf{v}_W] = \begin{bmatrix} v_{11} & \dots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{W1} & \dots & v_{Wm} \end{bmatrix}, \quad (14)$$

где векторы  $\mathbf{v}_k$  (строки матрицы) соответствуют траектории выхода РПМ для одной реализации параметров и шумов, а общее количество таких векторов (строк)  $W = SU$ .

По полученному ансамблю можно вычислить различные траектории, характеризующие выборочные статистические величины. В настоящей работе используются средняя траектория (обозначаемая далее  $\text{mean}$ ), медианная траектория ( $\text{med}$ ), область стандартного отклонения ( $\text{std}$ ). Кроме этого, для целей сравнения, может быть вычислена прогнозная траектория, соответствующая среднему значению параметров ( $\text{avg}$ ).

**3. Дизайн эксперимента.** Данные об общем количестве инфицированных, используемые в экспериментальном исследовании для каждой страны, были организованы следующим образом.

Каждый набор данных был разделен на 3 непересекающиеся группы (интервала): интервал обучения  $\mathcal{T}_{train}$ , интервал тестирования  $\mathcal{T}_{test}$  и интервал прогнозирования  $\mathcal{T}_{pred}$ . Следуя принятой в области анализа данных и машинного обучения методике, модель требуется обучить на интервале обучения, далее ее необходимо протестировать на интервале тестирования с целью улучшения предсказательной способности, и далее осуществить прогнозирование на интервале прогнозирования.

Таблица 1. Параметры набора данных

	$d_0$	$T_{train}$	$T_{test}$	$T_{pred}$
Германия	08-03-2020	[40, 57]	[58, 71]	[72, 101]
Франция	08-03-2020	[44, 69]	[70, 83]	[84, 113]
Италия	29-02-2020	[29, 47]	[48, 61]	[62, 91]
Испания	09-03-2020	[37, 51]	[52, 65]	[66, 95]
Великобритания	14-03-2020	[43, 66]	[67, 80]	[80, 109]
Швейцария	05-03-2020	[9, 21]	[22, 35]	[36, 65]
Бельгия	16-03-2020	[41, 63]	[64, 77]	[78, 107]

Для обучения предсказательной модели для каждой страны используются данные начиная со дня, когда впервые было зафиксировано общее количество инфицированных, превышающее 1000 человек (100 человек для Швейцарии), и заканчиваясь за 7 дней до пика весны 2020 года. Интервал тестирования составлял 14 дней, и интервал прогнозирования — 30 дней. Конфигурация всех наборов данных по каждой стране указаны в таблице 1, в колонке  $d_0$  указана дата, соответствующая первому дню интервала обучения для соответствующей страны.

Рандомизированная модель (4)-(6) содержит несколько гиперпараметров, которые настраиваются на интервале тестирования. Этими параметрами являются интервалы  $A_k$ , интервалы шума устанавливаются для всех распределений равными  $[-0,3; 0,3]$ , что соответствует шуму в пределах 30%, параметры  $L$  и  $M$  принимаются постоянными и равными 5.

Основной метрикой для сравнения и настройки гиперпараметров является коэффициент детерминации  $R^2$ , определяемый формулой:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (15)$$

где  $y$  — реальные данные,  $\hat{y}$  — выход модели,  $\bar{y}$  — среднее значение, вычисленное по реальным данным.

Настройка гиперпараметров РПМ проводится следующим образом. Формируется набор конфигураций интервалов  $A_k$  путем вычисления сетки по каждому параметру с определенным шагом. Далее проводится обучение с каждой конфигурацией, и на интервале тестирования выбирается конфигурация интервалов с максимальным значением  $R^2$  для средней по ансамблю траектории  $mean$ . Начальные интервалы устанавливаются в пределах 20% от точечной оценки, полученной методом наименьших квадратов на том же наборе данных.

После настройки гиперпараметров на интервале тестирования происходит переобучение РПМ на совмещенном наборе данных  $\mathcal{T}_{train} \cup \mathcal{T}_{test}$  с полученными при настройке оптимальными интервалами  $A_k^*$ , и дальнейшее прогнозирование на интервале  $\mathcal{T}_{pred}$ , при этом при прогнозировании используется распределение шума, полученное для последней точки на интервале  $\mathcal{T}_{train} \cup \mathcal{T}_{test}$ .

Полученные рандомизированные прогнозы сравниваются с подгонкой кривой с помощью метода наименьших квадратов на том же наборе данных.

Кроме показателя  $R^2$  также вычислялась средне-квадратичная ошибка (Mean-squared error, MSE) для истинных (реальных, true) значений  $y$  и предсказанных (модельных, predicted) значений  $\hat{y}$  по формуле

$$MSE = \frac{1}{n} \sum_{i=1}^{n-1} (y_i - \hat{y}_i)^2. \quad (16)$$

Все программное обеспечение для экспериментов реализовано на платформе Python 3.8 в среде Windows 10, метод наименьших квадратов реализован функцией `curve_fit`, метод оптимизации TNC, используемый для обучения РПМ, реализован функцией `minimize` библиотеки `scipy.optimize`.

**4. Результаты и обсуждение.** Полученные в экспериментах результаты изображены на рисунках 2-8 со следующими метками:

- `real` — реальные данные;
- `ols` — метод наименьших квадратов;
- `avg` — рандомизированное прогнозирование со средними по распределению значениями параметров модели;
- `mean` — рандомизированное прогнозирование со средним по ансамблю;
- `med` — рандомизированное прогнозирование с медианой по ансамблю;
- `7-avg` — 7-дневное скользящее среднее.

На рисунках указаны общие и ежедневные объемы инфицированных. На рисунках с общими объемами светло-серым цветом отмечены траектории, составляющие полученный ансамбль, темно-серым цветом — область стандартного отклонения по ансамблю. На рисунках с ежедневными объемами инфицированных светло-серым цветом указаны реальные данные и нанесены кривые, соответствующие 7-дневному скользящему среднему.

Таблица 2. Интервалы параметров РПМ

	$A_1$	$A_2$	$A_3$
Германия	[195231, 292847]	[0.1618, 0.2427]	[0.5902, 0.8853]
Франция	[8833, 13249]	[0.0940, 0.1411]	[1.0129, 1.5193]
Италия	[15476, 23215]	[0.1511, 0.2267]	[0.4993, 0.7489]
Испания	[207419, 311129]	[0.1753, 0.2630]	[0.5333, 0.8000]
Великобритания	[65218, 97828]	[0.1291, 0.1937]	[0.4687, 0.7030]
Швейцария	[1197, 1796]	[0.2108, 0.3162]	[0.5060, 0.7590]
Бельгия	[13241, 19862]	[0.1229, 0.1844]	[0.6245, 0.9368]

Все эксперименты производились для 1000 реализаций из распределения параметров модели, и 100 реализаций по распределениям шумов для каждого значения параметра. Генерация распределений шума проводилась для каждой точки соответствующего интервала. Таким образом, полученный ансамбль состоял из  $10^5$  траекторий. Вертикальная синяя пунктирная линия нанесена в точке начала интервала тестирования, красная пунктирная линия нанесена в точке начала интервала прогнозирования.

В таблице 2 указаны интервалы параметров  $A_k$ , полученные в результате тестирования модели и использованные при финальном обучении на итоговом интервале  $\mathcal{T}_{train} \cup \mathcal{T}_{test}$ .

В таблицах 3-4 приведены значения метрик, используемых в экспериментах для интервала обучения  $\mathcal{T}_{train} \cup \mathcal{T}_{test}$  и прогнозирования  $\mathcal{T}_{pred}$ .

Анализируя полученные результаты, можно отметить следующее. Данные, используемые в экспериментах, соответствуют периоду «первой волны» эпидемии COVID-19 (весна 2020 года). В этот период все страны столкнулись с массовой нехваткой тестов, неготовностью медицинской и лабораторной инфраструктуры, а также с неготовностью протоколов лечения и диагностики, в следствие чего данные о заболевших в тот период имеют существенные искажения. Кроме этого, в тот период совершенствование тест-систем и протоколов тестирования происходило фактически в реальном времени, что приводило к получению более точных и полных данных. Эти соображения могут объяснить рост числа заболевших в определенные периоды времени, в то время как ожидалось их снижение вследствие введенных общенациональных локдаунов практически во всех европейских странах. Отмеченные эффекты наблюдаются в полученных в экспериментах результатах, в частности, такая ситуация видна в Великобритании. Стоит также отметить, что продолжающийся, хотя и более плавный, рост числа заболевших может быть связан с недостаточными

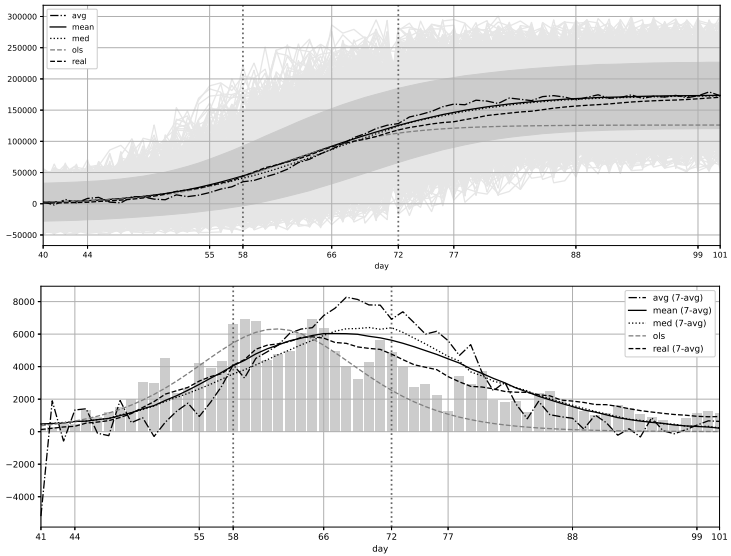


Рис. 2. Общее и ежедневное количество инфицированных в Германии

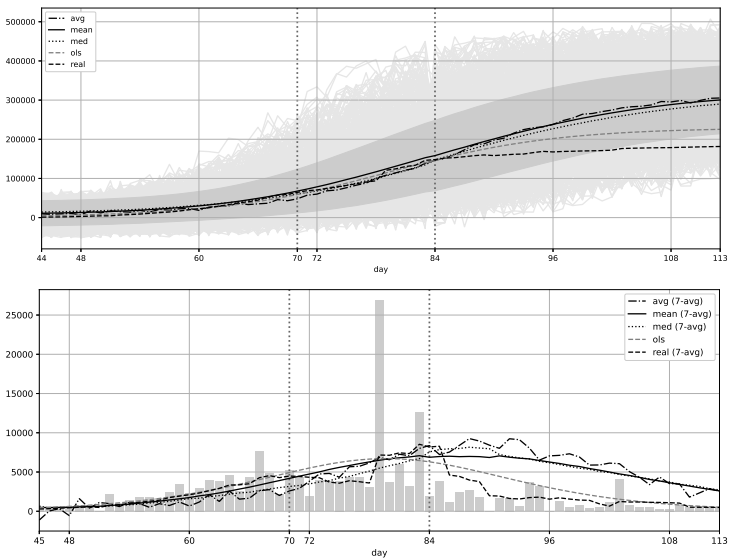


Рис. 3. Общее и ежедневное количество инфицированных во Франции

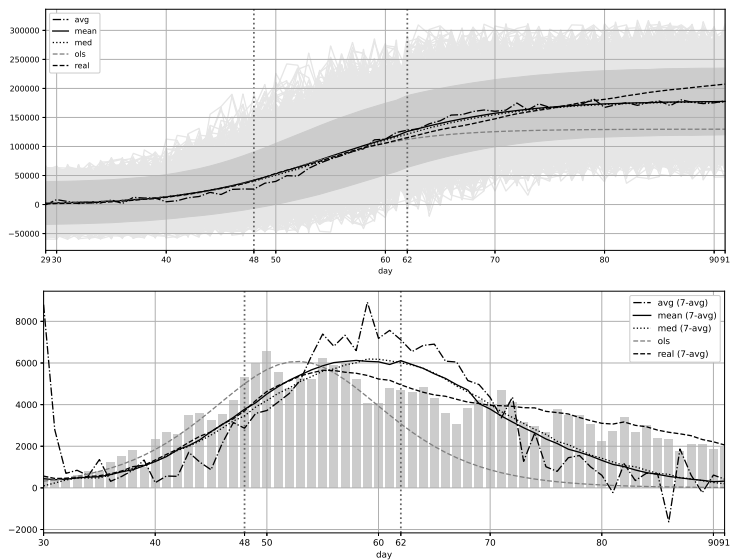


Рис. 4. Общее и ежедневное количество инфицированных в Италии

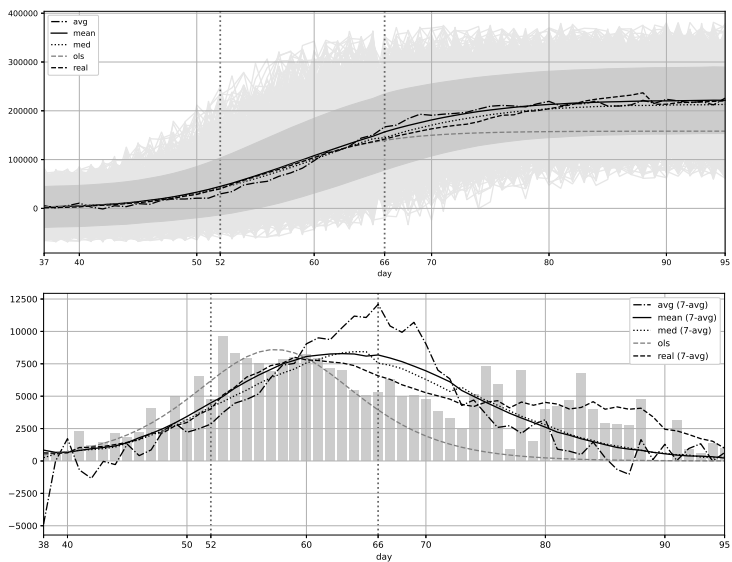


Рис. 5. Общее и ежедневное количество инфицированных в Испании

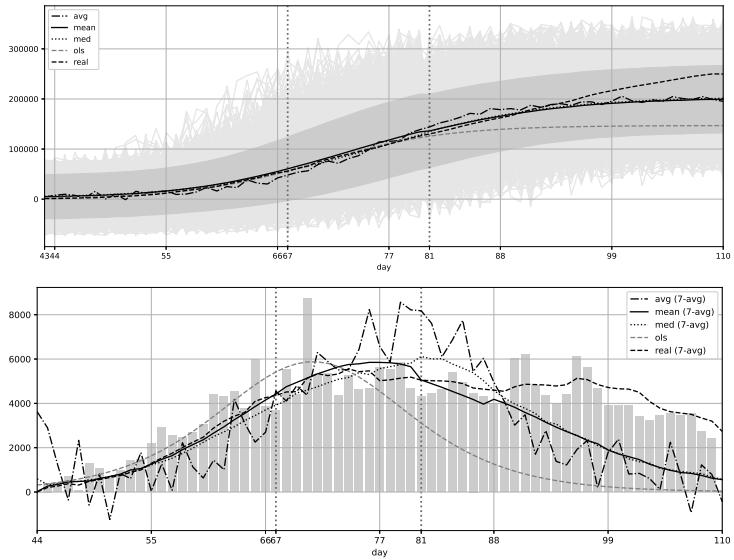


Рис. 6. Общее и ежедневное количество инфицированных в Великобритании

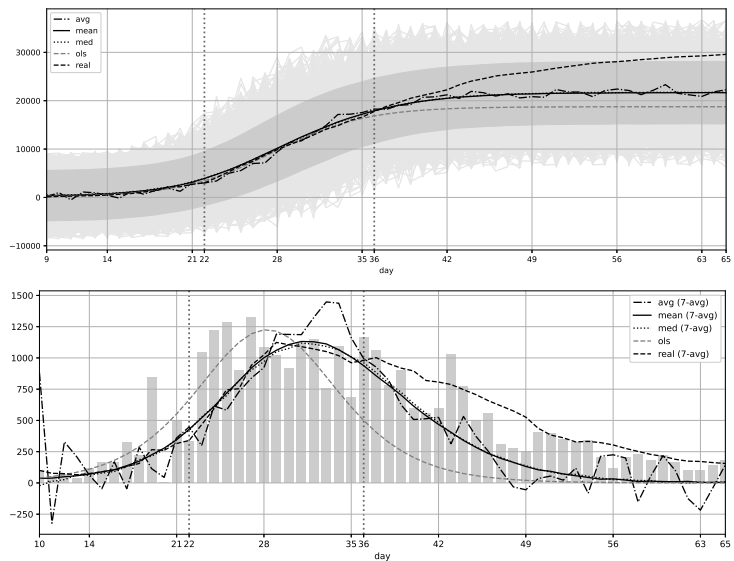


Рис. 7. Общее и ежедневное количество инфицированных в Швейцарии



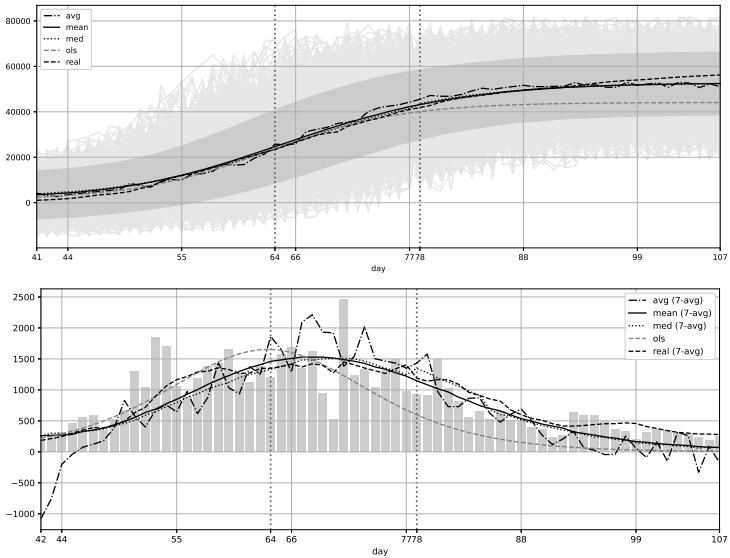


Рис. 8. Общее и ежедневное количество инфицированных в Бельгии

Таблица 3. Значение показателей качества на интервале обучения  $\mathcal{T}_{train} \cup \mathcal{T}_{test}$

	ols	avg	mean	med
	$R^2$			
Германия	0.9983	0.9483	0.9933	0.9875
Франция	0.9878	0.9639	0.9416	0.9372
Италия	0.9993	0.9583	0.9967	0.9933
Испания	0.9993	0.9533	0.9922	0.9956
Великобритания	0.9983	0.9754	0.9812	0.9691
Швейцария	0.9987	0.9846	0.9969	0.9958
Бельгия	0.9944	0.9651	0.9846	0.9758
	$MSE$			
Германия	0.0001	0.0025	0.0003	0.0006
Франция	0.0007	0.0020	0.0033	0.0035
Италия	0.0000	0.0013	0.0001	0.0002
Испания	0.0000	0.0017	0.0003	0.0002
Великобритания	0.0000	0.0006	0.0005	0.0008
Швейцария	0.0000	0.0005	0.0001	0.0001
Бельгия	0.0003	0.0019	0.0009	0.0013

Таблица 4. Значение показателей качества на интервале прогнозирования  $\mathcal{T}_{pred}$ 

	ols	avg	mean	med
$R^2$				
Германия	-2.5876	-0.1733	0.4272	<b>0.5883</b>
Франция	<b>-11.5808</b>	-55.8622	-44.6427	-47.4572
Италия	-1.9247	0.6243	0.6975	<b>0.7083</b>
Испания	-2.1955	0.5739	0.8489	<b>0.8536</b>
Великобритания	-1.8281	0.2007	0.1640	<b>0.2189</b>
Швейцария	-4.0840	<b>-1.2270</b>	-1.5965	-1.4652
Бельгия	-3.1254	0.6807	0.7199	<b>0.7511</b>
$MSE$				
Германия	0.0298	0.0097	0.0048	<b>0.0034</b>
Франция	<b>0.0363</b>	0.1640	0.1317	0.1398
Италия	0.0532	0.0068	0.0055	<b>0.0043</b>
Испания	0.0422	0.0056	0.0020	<b>0.0019</b>
Великобритания	0.0655	0.0185	0.0194	<b>0.0181</b>
Швейцария	0.0689	<b>0.0302</b>	0.0352	0.0334
Бельгия	0.0222	0.0017	0.0015	<b>0.0013</b>

противоэпидемическими мерами, а также с массовым несоблюдением этих мер населением.

Согласно показателям качества (см. табл. 3), полученным на интервале обучения, стандартных подход, основанный на методе наименьших квадратов, кажется более эффективным, качество приближения реальных данных с использованием трехпараметрической логистической модели превосходит соответствующие расчетные данные рандомизированной модели. Однако, на интервале прогнозирования ситуация существенно меняется в пользу последних (см. табл. 4).

Рандомизированный подход к прогнозированию, основная идея которого состоит в анализе ансамбля выхода модели, может быть реализован и для средних по распределению значений параметров модели. Такой способ реализации предсказательной модели соответствует стандартному подходу к прогнозированию, основанному на точечных оценках параметров. Из полученных результатов такого способа прогнозирования видно, что модельные значения выхода имеют большое отклонение от модельных значений, полученных при генерации соответствующих распределений, и иногда находятся в области отрицательных значений. Последнее наблюдение обусловлено свойством рандомизированной модели, в которой на выходе действует аддитивный шум, по причине которого модельные значения могут сильно отклоняться от средних по ансамблю траекторий. Это свойство такого прогноза может быть сглажено другим способом

учета шума при реализации модели, например, полной его фильтрацией. В целом необходимо отметить, что реализация указанного прогноза приводится в работе с целью сравнения как со стандартным методом, так и с подходом, основанным на генерации энтропийно-оптимальных распределений параметров и шумов.

Таким образом, можно констатировать преимущество энтропийно-рандомизированного подхода и его эффективность даже в условиях неполных и искаженных (зашумленных) данных. Необходимо также отметить, что при применении метода рандомизированного машинного обучения какие-либо априорные предположения о вероятностных свойствах данных не требуется.

**5. Заключение.** В работе предложен подход к прогнозированию развития эпидемии COVID-19, основанный на новом методе рандомизированного машинного обучения, состоящего в оценке распределений вероятностей параметров модели вместе с измерительными шумами. Применение данного метода продемонстрировано в эксперименте, одной из целей которого является сравнительное исследование со стандартными методами. Показаны преимущества предложенного метода в условиях малого количества зашумленных данных, каковыми являются эпидемические данные начального этапа. Важной особенностью используемого метода является его независимость от вероятностных свойств данных. Полученные результаты показывают, что даже в таких сложных условиях возможно построение адекватного прогноза, существенно превосходящего качества прогноза, полученного методом наименьших квадратов. Настоящая работа содержит результаты первого применения метода рандомизированного машинного обучения к моделированию эпидемического процесса. Успешное его применение к другим задачам интеллектуального анализа данных, таких как задачи классификации и задачи прогнозирования с помощью динамических регрессионных моделей, позволяет надеяться на его эффективность и в задачах, рассматриваемых в настоящей работе.

### Литература

1. van den Driessche P. *Mathematical Epidemiology* / ed. by Brauer F., van den Driessche P., Wu J. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Vol. 1945. pp. 147–157.
2. Kumar J., Hembram K.P.S.S. Epidemiological study of novel coronavirus (COVID-19). ArXiv. 2020. URL: <http://arxiv.org/abs/2003.11376> (accessed 02.09.2021).
3. Yang W., Zhang D., Peng L., Zhuge C., and Hong L. Rational evaluation of various epidemic models based on the COVID-19 data of China. ArXiv. 2020. URL: <http://arxiv.org/abs/2003.05666> (accessed 02.09.2021).
4. Tátrai D., Várallyay Z. COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. ArXiv. 2020. URL: <http://arxiv.org/abs/2003.14160> (accessed 02.09.2021).

5. Morais A.F. Logistic approximations used to describe new outbreaks in the 2020 COVID-19 pandemic. ArXiv. 2020. URL: <http://arxiv.org/abs/2003.11149> (accessed 02.09.2021).
6. Shen C.Y. Logistic growth modelling of COVID-19 proliferation in China and its international implications. *International Journal of Infectious Diseases*. 2020. vol. 96. pp. 582–589. URL: <https://doi.org/10.1016/j.ijid.2020.04.085> (accessed 02.09.2021).
7. Wang P., Zheng X., Li J., Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*. 2020. vol. 139. P. 110058. URL: <https://doi.org/10.1016/j.chaos.2020.110058> (accessed 02.09.2021).
8. Chen D.-G., Chen X., Chen J.K. Reconstructing and forecasting the COVID-19 epidemic in the United States using a 5-parameter logistic growth model. *Global Health Research and Policy*. 2020. vol. 5, no. 1. P. 25. URL: <https://doi.org/10.1186/s41256-020-00152-5> (accessed 02.09.2021).
9. Popkov Yu.S., Popkov A.Yu., Dubnov Yu.A. Randomizirovannoe mashinnoe obuchenie pri ogranichennyh naborah dannyh: ot jempiricheskoj verojatnosti k jentropijnoj randomizacii. [Randomized machine learning with limited datasets: from empirical probability to entropy randomization]. M. : LENAND, 2019.
10. Boltzman L. O svyazi mezhdv vtorym nachalom meh anicheskoj teorii teploty i teoriej verojatnostej v teoremah o teplovom ravnovesii. [On the connection between the second principle of the mechanical theory of heat and the theory of probabilities in theorems on thermal equilibrium]. Boltzman L.E. *Izbrannye trudy*. [Selected Works]. Ed. Slag L.S. Moscow: Nauka, 1984.
11. Jaynes E.T. Information theory and statistical mechanics. *Physical review*. 1957. vol. 106, no. 4. pp. 620–630.
12. Jaynes E.T. *Probability theory: the logic of science*. Cambridge university press. 2003.
13. Shannon C.E. Communication theory of secrecy systems. *Bell Labs Technical Journal*. 1949. vol. 28. no. 4. pp. 656–715.
14. Verhulst P.-F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* 1838. Vol. 10. pp. 113–126.
15. Singer H.M. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Physical Biology*. 2020. vol. 17. no. 5. P. 055001.
16. Popkov Yu.S., Dubnov Yu.A., Popkov A.Yu. Randomized machine learning: Statement, solution, applications. *Intelligent Systems (IS), 2016 IEEE 8th International Conference on IEEE*. 2016. pp. 27–39.
17. Popkov Y.S., Dubnov Y.A., Popkov A.Y. Introduction to the Theory of Randomized Machine Learning. *Learning Systems: From Theory to Practice / ed. by Sgurev V., Piuri V., Jotsov V. Cham: Springer International Publishing, 2018. pp. 199–220.*
18. Popkov Y.S., Dubnov Y.A., Popkov A.Y. New method of randomized forecasting using entropy-robust estimation: Application to the World population prediction. *Mathematics*. 2016. vol. 4, no. 1. pp. 1–16.
19. Popkov Y.S., Volkovich Z., Dubnov Y.A., Avros R., Ravve E. Entropy 2-Soft Classification of Objects. *Entropy*. 2017. vol. 19, no. 4. P. 178.
20. Dubnov Y.A. Entropy-Based Estimation in Classification Problems. *Automation and Remote Control*. 2019. vol. 80. pp. 502–512.
21. Popkov Y.S., Popkov A.Y., Dubnov Y.A., Solomatine D. Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models. *Mathematics*. 2020. No. 8. P. 1119.
22. Popkov A.Y. Randomized machine learning of nonlinear models with application to the prediction of the development of epidemic process. *Automation and Remote Control*. 2021.

23. Dong E., Du H., Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*. 2020. vol. 20. no. 5. pp. 533–534.
24. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. URL: <https://github.com/CSSEGISandData/COVID-19> (accessed 02.09.2021).
25. Golan A., Judge G., Miller D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York : John Wiley & Sons. 1996.
26. Golan A. Information and entropy econometrics. A review and synthesis. *Foundations and trends in Econometrics*. 2008. vol. 2, no. 1–2. pp. 1–145.
27. Nocedal J., Wright S. *Numerical optimization*. Springer Science & Business Media. 2006.
28. Nash S.G. Newton-type minimization via the Lanczos method. *SIAM Journal on Numerical Analysis*. 1984. vol. 21. no. 4. pp. 770–788.

**Попков Юрий Соломонович** — д-р техн. наук, профессор, академик РАН, главный научный сотрудник, Федеральный исследовательский центр "Информатика и управление" Российской академии наук; главный научный сотрудник, Институт проблем управления Российской академии наук. Область научных интересов: энтропийные методы, макросистемы, рандомизированное машинное обучение. Число научных публикаций — 220. popkov@isa.ru; ул. Вавилова, 44/2, 119133, Москва, Россия; р.т.: +79852111090.

**Дубнов Юрий Андреевич** — научный сотрудник, Федеральный исследовательский центр "Информатика и управление" Российской академии наук; Старший преподаватель, Национальный исследовательский университет "Высшая школа экономики". Область научных интересов: машинное обучение, байесовское оценивание. Число научных публикаций — 30. yury.dubnov@phystech.edu; ул. Вавилова, 44/2, 119133, Москва, Россия; р.т.: +79099634498.

**Попков Алексей Юрьевич** — канд. техн. наук, ведущий научный сотрудник, Федеральный исследовательский центр "Информатика и управление" Российской академии наук. Область научных интересов: энтропийные методы, рандомизированное машинное обучение, интеллектуальный анализ данных, разработка программного обеспечения. Число научных публикаций — 47. aropkov@isa.ru; ул. Вавилова, 44/2, 119133, Москва, Россия; р.т.: +7(499)135-62-60.

**Поддержка исследований.** Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 20-04-60119).

Y. POPKOV, Y. DUBNOV, A. POPKOV  
**FORECASTING DEVELOPMENT OF COVID-19 EPIDEMIC IN  
EUROPEAN UNION USING ENTROPY-RANDOMIZED APPROACH**

---

*Popkov Y., Dunbov Y., Popkov A. Forecasting Development of COVID-19 Epidemic in European Union Using Entropy-Randomized Approach.*

**Abstract.** The paper is devoted to the forecasting of the COVID-19 epidemic by the novel method of randomized machine learning. This method is based on the idea of estimation of probability distributions of model parameters and noises on real data. Entropy-optimal distributions correspond to the state of maximum uncertainty which allows the resulting forecasts to be used as forecasts of the most “negative” scenario of the process under study. The resulting estimates of parameters and noises, which are probability distributions, must be generated, thus obtaining an ensemble of trajectories that considered to be analyzed by statistical methods. In this work, for the purposes of such an analysis, the mean and median trajectories over the ensemble are calculated, as well as the trajectory corresponding to the mean over distribution values of the model parameters. The proposed approach is used to predict the total number of infected people using a three-parameter logistic growth model. The conducted experiment is based on real COVID-19 epidemic data in several countries of the European Union. The main goal of the experiment is to demonstrate an entropy-randomized approach for predicting the epidemic process based on real data near the peak. The significant uncertainty contained in the available real data is modeled by an additive noise within 30%, which is used both at the training and predicting stages. To tune the hyperparameters of the model, the scheme is used to configure them according to a testing dataset with subsequent retraining of the model. It is shown that with the same datasets, the proposed approach makes it possible to predict the development of the epidemic more efficiently in comparison with the standard approach based on the least-squares method.

**Keywords:** epidemic modelling, SARS-CoV-2, COVID-19, randomized machine learning, entropy, entropy estimation, forecasting, randomized forecasting.

---

**Popkov Yuri** — Ph.D., Dr.Sci., Professor, Academician of RAS, Chief research scientist, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences; Chief research scientist, Institute of Control Problems of Russian Academy of Sciences. Research interests: entropy, macrosystems, randomized machine learning, optimization. The number of publications — 220. popkov@isa.ru; 44/2, Vavilov Str., 119133, Moscow, Russia; office phone: +79852111090.

**Dubnov Yuri** — Research scientist, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences; Senior lecturer, National Research University Higher School of Economics. Research interests: machine learning, bayesian estimation. The number of publications — 30. yury.dubnov@phystech.edu; 44/2, Vavilov Str., 119133, Moscow, Russia; office phone: +79099634498.

**Popkov Alexey** — Ph.D., Leading research scientist, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences. Research interests: entropy methods, machine learning, data mining, software development. The number of publications — 47. apopkov@isa.ru; 44/2, Vavilova Str., 119133, Moscow, Russia; office phone: +7(499)135-62-60.

**Acknowledgements.** This work was supported by Russian Foundation for Basic Research (project no. 20-04-60119).

## References

1. van den Driessche P. *Mathematical Epidemiology* / ed. by Brauer F., van den Driessche P., Wu J. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Vol. 1945. pp. 147–157.
2. Kumar J., Hembram K.P.S.S. Epidemiological study of novel coronavirus (COVID-19). ArXiv. 2020. URL: <http://arxiv.org/abs/2003.11376> (accessed 02.09.2021).
3. Yang W., Zhang D., Peng L., Zhuge C., and Hong L. Rational evaluation of various epidemic models based on the COVID-19 data of China. ArXiv. 2020. URL: <http://arxiv.org/abs/2003.05666> (accessed 02.09.2021).
4. Tátrai D., Várallyay Z. COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. ArXiv. 2020. URL: <http://arxiv.org/abs/2003.14160> (accessed 02.09.2021).
5. Morais A.F. Logistic approximations used to describe new outbreaks in the 2020 COVID-19 pandemic. ArXiv. 2020. URL: <http://arxiv.org/abs/2003.11149> (accessed 02.09.2021).
6. Shen C.Y. Logistic growth modelling of COVID-19 proliferation in China and its international implications. *International Journal of Infectious Diseases*. 2020. vol. 96. pp. 582–589. URL: <https://doi.org/10.1016/j.ijid.2020.04.085> (accessed 02.09.2021).
7. Wang P., Zheng X., Li J., Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*. 2020. vol. 139. P. 110058. URL: <https://doi.org/10.1016/j.chaos.2020.110058> (accessed 02.09.2021).
8. Chen D.-G., Chen X., Chen J.K. Reconstructing and forecasting the COVID-19 epidemic in the United States using a 5-parameter logistic growth model. *Global Health Research and Policy*. 2020. vol. 5, no. 1. P. 25. URL: <https://doi.org/10.1186/s41256-020-00152-5> (accessed 02.09.2021).
9. Popkov Yu.S., Popkov A.Yu., Dubnov Yu.A. Randomizirovannoe mashinnoe obuchenie pri ogranichennyh naborah dannyh: ot jempiricheskoj verojatnosti k jentropijnoj randomizacii. [Randomized machine learning with limited datasets: from empirical probability to entropy randomization]. М.: LENAND, 2019.
10. Boltzman L. O svjazi mezhdu vtorym nachalom meh anicheskoj teorii teploty i teoriej verojatnostej v teoremah o teplovom ravnovesii. [On the connection between the second principle of the mechanical theory of heat and the theory of probabilities in theorems on thermal equilibrium]. Boltzman L.E. *Izbrannye trudy*. [Selected Works]. Ed. Slag L.S. Moscow: Nauka, 1984.
11. Jaynes E.T. Information theory and statistical mechanics. *Physical review*. 1957. vol. 106, no. 4. pp. 620–630.
12. Jaynes E.T. *Probability theory: the logic of science*. Cambridge university press. 2003.
13. Shannon C.E. Communication theory of secrecy systems. *Bell Labs Technical Journal*. 1949. vol. 28. no. 4. pp. 656–715.
14. Verhulst P.-F. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* 1838. Vol. 10. pp. 113–126.
15. Singer H.M. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Physical Biology*. 2020. vol. 17. no. 5. P. 055001.
16. Popkov Yu.S., Dubnov Yu.A., Popkov A.Yu. Randomized machine learning: Statement, solution, applications. *Intelligent Systems (IS)*, 2016 IEEE 8th International Conference on IEEE. 2016. pp. 27–39.

17. Popkov Y.S., Dubnov Y.A., Popkov A.Y. Introduction to the Theory of Randomized Machine Learning. Learning Systems: From Theory to Practice / ed. by Sgurev V., Piuri V., Jotsov V. Cham: Springer International Publishing, 2018. pp. 199–220.
18. Popkov Y.S., Dubnov Y.A., Popkov A.Y. New method of randomized forecasting using entropy-robust estimation: Application to the World population prediction. Mathematics. 2016. vol. 4, no. 1. pp. 1–16.
19. Popkov Y.S., Volkovich Z., Dubnov Y.A., Avros R., Ravve E. Entropy 2-Soft Classification of Objects. Entropy. 2017. vol. 19, no. 4. P. 178.
20. Dubnov Y.A. Entropy-Based Estimation in Classification Problems. Automation and Remote Control. 2019. vol. 80. pp. 502–512.
21. Popkov Y.S., Popkov A.Y., Dubnov Y.A., Solomatine D. Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models. Mathematics. 2020. No. 8. P. 1119.
22. Popkov A.Y. Randomized machine learning of nonlinear models with application to the prediction of the development of epidemic process. Automation and Remote Control. 2021.
23. Dong E., Du H., Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases. 2020. vol. 20. no. 5. pp. 533–534.
24. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. URL: <https://github.com/CSSEGISandData/COVID-19> (accessed 02.09.2021).
25. Golan A., Judge G., Miller D. Maximum Entropy Econometrics: Robust Estimation with Limited Data. New York : John Wiley & Sons. 1996.
26. Golan A. Information and entropy econometrics. A review and synthesis. Foundations and trends in Econometrics. 2008. vol. 2, no. 1–2. pp. 1–145.
27. Nocedal J., Wright S. Numerical optimization. Springer Science & Business Media. 2006.
28. Nash S.G. Newton-type minimization via the Lanczos method. SIAM Journal on Numerical Analysis. 1984. vol. 21. no. 4. pp. 770–788.