

## Применимость обобщённого стохастического подхода к моделированию развития заболеваний на примере прогнозирования распространения гриппа

*П.Н. Подзолков*

*Тюменский государственный университет*

**Аннотация:** В статье рассматриваются методы моделирования распространения инфекционных заболеваний. Рассматриваются особенности обобщённого компартментального подхода к моделированию эпидемий, в рамках которого популяция подразделяется на непересекающиеся группы индивидов. Прогноз моделей, построенных при использовании данного подхода, подразумевает оценку численности групп индивидов во времени. Предложен метод оценки параметров модели по статистическим данным. Предложен метод оценки доверительных интервалов прогноза модели, основанный на серии стохастических запусков модели. Проведён вычислительный эксперимент, демонстрирующий работу предложенных методов на данных о распространении гриппа в странах Европы. Результаты показывают эффективность модели в прогнозировании динамики эпидемии и оценке доверительных интервалов прогноза. Обоснована возможность приложения описанных методов к моделированию хронических заболеваний.

**Ключевые слова:** моделирование эпидемий, компьютерное моделирование, компартментальные модели, стохастическое моделирование, оценка параметров, доверительный интервал, прогноз, грипп.

### Введение

Методы математического и компьютерного моделирования применяются во многих сферах деятельности человека, в том числе и для прогнозирования прогрессирования хронических и распространения инфекционных заболеваний. В ситуации развивающейся эпидемии инфекционного заболевания оценка численности индивидов в различных состояниях в заданное время в будущем может быть использована для формирования корректных мер защиты населения [1].

Одной из самых устоявшихся в применении моделей распространения эпидемий является модель SIR (В модели выделяют следующие состояния, в которых могут находиться индивиды: Susceptible (восприимчивые) – Infected (инфицированные) – Recovered (выздоровевшие); первые буквы слов составляют название модели). Данная модель основана на

компарментальном подходе, в рамках которого популяция разделяется на непересекающиеся группы индивидов – компарменты. Каждый компармент соответствует одному из выделенных состояний индивида, значимых с точки зрения развития и распространения заболевания [2].

В оригинальном виде SIR, как математическая модель, представляет собой систему из трёх дифференциальных уравнений, описывающих скорость изменения численности каждого из компарментов:

$$\begin{cases} \frac{dS}{dt} = -\frac{S \cdot \beta \cdot I}{N} \\ \frac{dI}{dt} = \frac{S \cdot \beta \cdot I}{N} - I \cdot \gamma \\ \frac{dR}{dt} = I \cdot \gamma \end{cases}$$

где  $\beta$  – коэффициент интенсивности инфицирования;  $\gamma$  – коэффициент интенсивности выздоровления.

Помимо SIR модели существует множество её модификаций, расширяющих количество компарментов, следовательно, учитывающих дополнительные состояния индивидов.

**Целью данного исследования** является разработка обобщённых методов оценки параметров и получения прогноза для моделей с различным набором компарментов, а также проведение вычислительного эксперимента, демонстрирующего работу предложенных методов, основанного на статистических данных о распространении инфекционного заболевания.

### **Обобщённая модель распространения эпидемии**

Обобщая компарментальный подход, для построения модели эпидемии инфекционного заболевания можно исходить из предположения, что каждый индивид в определённый момент времени находится в одном из состояний  $s_i$ , принадлежащих множеству  $S = \{s_i\}_{i=1}^n$  [3]. Между состояниями

---

возможны вероятностные переходы, которые составляют множество  $F = \{f_j\}_{j=1}^m$ .

В общем случае модель распространения инфекционного заболевания может быть представлена в виде графа переходов  $G(S, F)$ , в котором  $S$  – множество узлов, а  $F$  – множество рёбер.

Пусть  $q(s_i, t)$  – количество индивидов в состоянии  $s_i$  в момент времени  $t$ , а  $Q(S, t)$  – вектор, компонентами которого являются численности индивидов в каждом состоянии в момент времени  $t$ .

Пусть также  $q(f_j, t)$  – интенсивность перехода  $f_j$ , то есть количество индивидов совершивших переход  $f_j$  в момент времени  $t$ , а  $Q(F, t)$  – вектор, компонентами которого являются интенсивности каждого из переходов в момент времени  $t$ .

Тогда изменение численности индивидов в каждом из состояний во времени определяется через:

$$q(s_i, t + \Delta t) = q(s_i, t) + \Delta t \cdot \sum_{j|end(f_j)=s_i} q(f_j, t) - \Delta t \cdot \sum_{j|start(f_j)=s_i} q(f_j, t) \quad (1)$$

где  $start(f_j)$  – состояние-источник перехода  $f_j$ ;  $end(f_j)$  – состояние-цель перехода  $f_j$ ;  $\Delta t$  – изменение момента моделирования.

Определение интенсивности перехода  $f_j$  производится на основании соответствующего ему веса  $v_j$  из вектора  $V = \{v_j\}_{j=1}^m$ , количества индивидов в состоянии-источнике данного перехода и количества индивидов в состояниях, численность которых влияет на данный переход. Будем считать, что определена функциональная зависимость интенсивности каждого перехода от указанных величин:

$$q(f_j, t) = func(v_j, Q(S, t)) \quad (2)$$

**Цель моделирования:** для заданного  $t_0$  определить прогнозируемые значения  $Q(S, t)$  и  $Q(F, t)$ , где  $t > t_0$ .

Результаты моделирования представляют собой матрицу прогнозных значений  $R_{T \times n} = \{r_{ii}\}$ , где  $T$  – длительность моделирования,  $r_{ii} = q(s_i, t)$ . Следовательно, строками данной матрицы являются векторы  $Q(S, t)$ . Также в процессе моделирования может быть получена матрица прогнозных значений интенсивностей переходов во времени  $B_{T \times m} = \{b_{ij}\}$ , где  $m$  – количество переходов, а  $b_{ij} = q(f_j, t)$ . Строками данной матрицы, следовательно, являются векторы  $Q(F, t)$ .

### Стохастическое моделирование распространения эпидемии

Специфика предметной области предполагает наличие большого количества неизвестных факторов, влияющих на динамику распространения заболевания. Однако моделирование численности индивидов в различных состояниях на основе уравнений (1) и (2) подразумевает получение только детерминированного прогноза. В то же время учёт объективной неопределённости в модели может быть осуществлён за счёт включения в уравнения модели случайных величин. Таким образом может быть реализовано стохастическое моделирование, используемое для получения серии отличающихся прогнозов.

Разработанный нами подход к стохастическому моделированию использует величины, получаемые из уравнения (2) в качестве средних значений для распределения Пуассона, которое широко используется для моделирования процессов с заданной интенсивностью в том числе в эпидемиологии [4].

Для увеличения вариативности получаемых прогнозов был также применен принцип дважды стохастического моделирования, при котором изменение момента моделирования также являлся случайной величиной с распределением Пуассона и средним значением равным 1 [5].

Таким образом, при стохастическом моделировании численность индивидов в каждом состоянии в заданный момент времени определяется на основании одного из предшествующих моментов моделирования:

$$q(s_i, t_{new}) = q(s_i, t) + 1 \cdot Ps \left( \sum_{j|end(f_j)=s_i} q(f_j, t) \right) - 1 \cdot Ps \left( \sum_{j|start(f_j)=s_i} q(f_j, t) \right) \quad (3)$$

где  $Ps(x)$  – случайная величина из распределения Пуассона с математическим ожиданием  $x$ ;  $t_{new}$  – новый момент моделирования, для которого вычисляются прогнозные значения. Изменение момента моделирования определяется соотношением:

$$t_{new} = t + Ps(1) \quad (4)$$

Так, для классической SIR модели результаты детерминированного и стохастического запусков представлены на рис. 1.

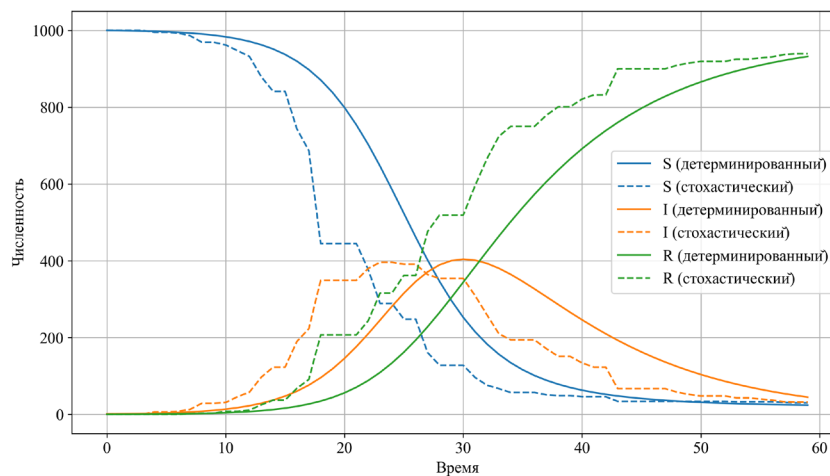


Рис. 1. – Детерминированный и стохастический результаты SIR модели

Стоит учитывать, что реализация нулевого изменения момента моделирования, возможна только в рамках компьютерного моделирования. В таком случае при вычислении вектора  $Q(S, t_{new})$  его значения могут быть пересчитаны неопределённое число раз, при совпадении  $t$  и  $t_{new}$ . Вычисление при этом на каждой итерации производится на основе предыдущих значений данного вектора.

Далее представлен псевдокод алгоритма получения стохастических результатов моделирования.

---

### Алгоритм 1 Стохастическое моделирование

---

```
1  инициализация МАССИВА [T, n] R результатов моделирования
2  заполнение строки R[0] значениями из Q(S, t0)
3  инициализация начального момента моделирования t значением t0
4  ПОКА момент моделирования t в пределах от t0 до T:
5      вычисление нового момента моделирования t_new по уравнению (4)
6      вычисление значений вектора Q(S, t_new) по уравнению (3)
7      заполнение строки R[t_new] значениями из Q(S, t_new)
8      изменение момента моделирования t на значение t_new
9  КОНЕЦ ПОКА
10 ВЕРНУТЬ R
```

---

### Оценка параметров модели

Статистические данные о распространении инфекционного заболевания будем называть *реальными данными* в противопоставлении результатам моделирования. Реальные данные могут содержать оценку количества индивидов во времени в различных состояниях: болеющие, выздоровевшие, не сталкивавшиеся с инфекцией, умершие и т.д. Такие данные легко могут быть сопоставлены с результатами моделирования в матрице  $R$ . Например, для  $SIR$  модели, в матрице  $R$  будут содержаться данные о количестве индивидов в состояниях «S», «I», «R», что сопоставимо с количеством не сталкивавшихся с инфекцией, болеющих и выздоровевших из соответствующего набора реальных данных. Тогда обозначим матрицу, содержащую реальные данные, через  $\bar{R}_{T \times m}$ .

Однако чаще реальные данные о распространении инфекционных заболеваний содержат информацию лишь о количестве новых случаев заболевания, произошедших в заданный промежуток времени. Дополнительно может быть собрана статистика о динамике выздоровлений, содержащая информацию о количестве новых выздоровлений, произошедших в заданный промежуток времени. Такие показатели могут быть сопоставлены с результатами моделирования, содержащимися в матрице  $B$ . Так для прогноза  $SIR$  модели содержимое матрицы  $B$  отражает

интенсивности переходов из «S» в «I» (инфицирование) и из «I» в «R» (выздоровление), что сопоставимо с данными о новых случаях заболевания и выздоровления. Обозначим матрицу, содержащую реальные данные об интенсивностях переходов между состояниями, через  $\bar{B}_{T \times m}$ .

Пусть  $X$  – вектор параметров, необходимых для получения прогнозов  $R$  и  $B$ . Тогда определение значений  $X$  является задачей оценки параметров модели для последующего прогнозирования.

Если задана матрица  $\bar{R}$  (Ситуация 1), вектор  $X$  должен содержать только параметры, необходимые для оценки интенсивностей переходов при заданном начальном распределении индивидов между состояниями  $Q(S, t_0) = \{\bar{r}_{0i}\}_{i=1}^n$ . В таком случае  $X$  будет соответствовать вектору  $V$  весов всех переходов. Соответственно, задача нахождения вектора  $X$  заключается в минимизации различий между матрицами  $R$  и  $\bar{R}$  [6].

Введём функцию ошибки  $E_r(X, \bar{R})$ :

$$E_r(X, \bar{R}) = \sum_t \sum_j^n (r_{ij} - \bar{r}_{ij})^2$$

Тогда найти вектор  $X$  можно через минимизацию функции ошибки:

$$X = X : E_r(X, \bar{R}) = \min E_r$$

Однако если задана только матрица  $\bar{B}$  (Ситуация 2), вектор  $X$  должен включать помимо элементов вектора  $V$ , элементы вектора  $Q(S, t_0)$ . Задача нахождения вектора  $X$  тогда заключается в минимизации различий между матрицами  $B$  и  $\bar{B}$ .

Введём соответствующую функцию ошибки  $E_b(X, \bar{B})$ :

$$E_b(X, \bar{B}) = \sum_t \sum_j^m (b_{ij} - \bar{b}_{ij})^2 \quad (5)$$

В таком случае, нахождение вектора  $X$  будет соответствовать следующему отношению:

$$X = X : E_b(X, \bar{B}) = \min E_b$$

Таким образом, задача оценки параметров модели заключается в минимизации функции ошибки, выбор которой зависит от заданной на основе реальных данных матрицы  $R$  или  $B$ . Модель, в которой используются найденные значения параметров, будем называть *обученной*.

### Доверительный интервал прогноза модели

Под доверительным интервалом прогноза  $q(s_i, t)$  или  $q(f_i, t)$  будем понимать интервал, в который соответствующее значение из реальных данных для момента времени  $t$  попадает с некоторой заданной вероятностью [7]. Для оценки доверительных интервалов детерминированных результатов моделирования может быть использована совокупность стохастических прогнозов, полученных при тех же параметрах стохастической модели. При этом предполагается, что возможное отклонение детерминированного прогноза относительно реальных данных не превышает значение соответствующей метрики относительно стохастического прогноза.

Также стоит учитывать, что вариативность стохастических прогнозов не симметрична относительно детерминированного прогноза. Вычислительные эксперименты показали, что количество стохастических прогнозов  $q(s_i, t)$ , значения которых меньше детерминированного прогноза, может преобладать над количеством превышающих его стохастических прогнозов. Справедливо и обратное утверждение. Таким образом, детерминированный прогноз в общем случае может находиться не в середине доверительного интервала.

Пусть результаты  $L$  стохастических запусков образуют трёхмерную матрицу  $H_{L \times T \times n} = \{h_{kti}\}$ , где  $h_{kti}$  – прогнозное значение для  $q(s_i, t)$ , полученное в результате  $k$ -го стохастического запуска.

Тогда границы доверительного интервала для каждого момента времени могут быть определены следующим образом:



$$r_{ii}^{(down)} = Percentile(\{h_{kti} \mid h_{kti} < r_{ii} \}_{k=1}^L, 100 - \gamma)$$

$$r_{ii}^{(up)} = Percentile(\{h_{kti} \mid h_{kti} > r_{ii} \}_{k=1}^L, \gamma)$$

где  $r_{ii}^{(down)}$ ,  $r_{ii}^{(up)}$  – нижний и верхний пределы доверительного интервала для прогноза  $q(s_i, t)$ ;  $r_{ii}$  – детерминированный прогноз  $q(s_i, t)$ ;  $Percentile(A, \gamma)$  –  $\gamma$ -й процентиль набора значений  $A$ .

Аналогичным образом может быть определён доверительный интервал для прогноза интенсивностей переходов, определённых в матрице  $B$ .

### Вычислительный эксперимент

Инструменты для построения модели, оценки её параметров, получения прогнозов и доверительных интервалов были реализованы в виде пакета `epidemmo` (<https://pypi.org/project/epidemmo/>) на языке программирования Python. Для минимизации функции ошибки использовался метод Нелдера-Мида, реализованный в пакете *SciPy* [8, 9].

Для апробации описанных методов был выбран набор данных о распространении гриппа в мире – *FluNet* (<https://www.who.int/tools/flunet>), собираемый глобальной программой по надзору за гриппом и принятию ответных мер в рамках деятельности всемирной организации здравоохранения (ВОЗ). Данные включают еженедельную сводку о новых случаях различных штаммов гриппа от участвующих в проекте стран [10].

Для построения модели из представленного набора были извлечены данные о суммарном количестве новых случаев всех штаммов гриппа для всех стран Европы за период с 1 ноября 2024 года по 20 января 2025 года, представляющий временной ряд из 12 значений.

В качестве базовой модели эпидемии была выбрана классическая SIR модель. Задача оценки параметров для выбранного набора данных соответствует *Ситуации 2*. При моделировании использовался шаг времени равный одному дню. Полученные значения численности перехода из

состояния «S» в состояние «I» (инфицирование) суммировались для каждого недельного интервала и сравнивались с реальными данными в соответствии с уравнением (5).

После минимизации функции ошибки были получены следующие значения параметров модели: начальные значения численности состояний «S», «I» и «R» – 182452, 936 и 0 соответственно; значения весов переходов инфицирования ( $\beta$ ) и выздоровления ( $\gamma$ ) – 0.157 и 0.12 соответственно. Результаты работы *обученной SIR модели* представлены на рис. 2 и рис. 3.

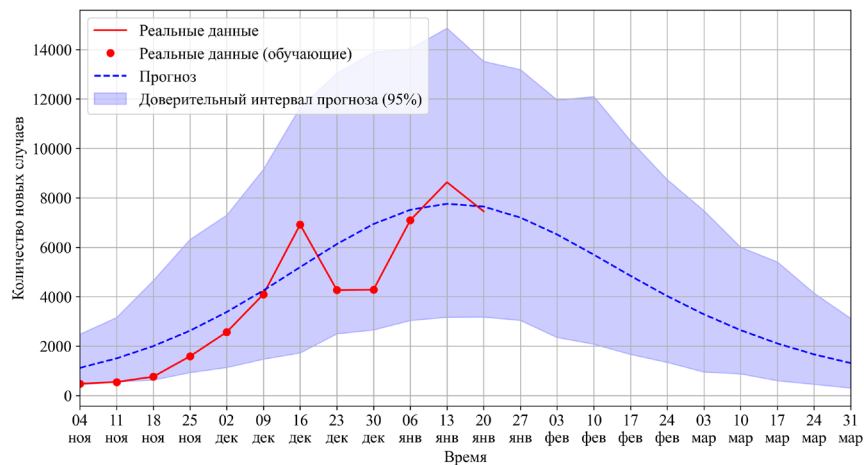


Рис. 2. – Прогноз интенсивности инфицирования (количества новых случаев), полученный в результате работы *обученной SIR модели*

Из полученного прогноза следует, что наибольшая интенсивность инфицирования приходится на середину января, что соответствует данным о распространении гриппа за предыдущий год. Доверительный интервал прогноза покрывает реальные данные, на которых производилось обучение модели и последующие значения.

Зависимости, представленные на рис. 3, показывают, что пик численности инфицированных гриппом может приходиться на промежуток времени с начала января до конца февраля. Данный интервал определён на основе оценки разброса времени наступления максимума численности инфицированных в наборе стохастических запусков модели.

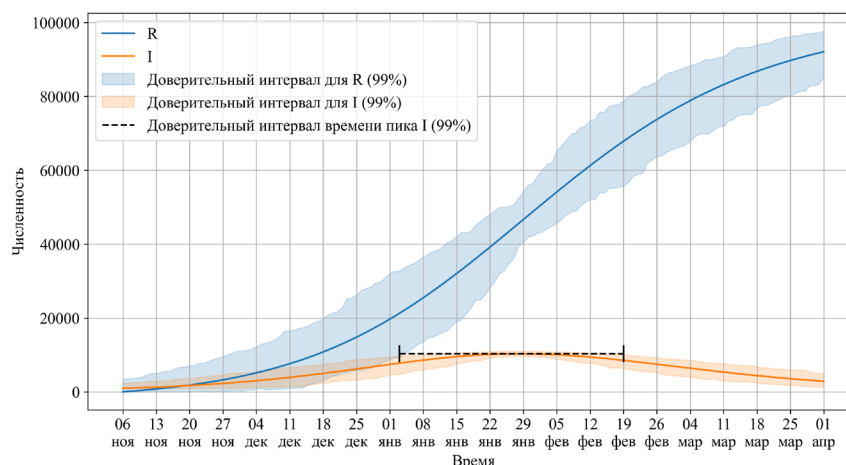


Рис. 3. – Прогноз численности инфицированных и выздоровевших, полученный в результате работы *обученной SIR модели*

На основании полученного прогноза можно сделать вывод, что в период сезонного распространения гриппа (2024 – 2025гг.) в странах Европы, согласно модели, переболеют приблизительно 100 тыс. человек.

### Заключение

Опираясь на модель эпидемии, основанную на графе переходов между состояниями, можно получить детерминированный и стохастический результаты прогнозирования. Совокупность стохастических результатов моделирования отражает вариативность получаемого прогноза, которую можно оценить и представить в виде доверительного интервала.

Предложенный метод оценки параметров модели заболевания, основанный на минимизации отличий между детерминированным прогнозом модели и статистическими (реальными) данными, позволяет оценить валидность и при необходимости скорректировать соответствующую модель.

Описанный метод оценки доверительных интервалов может быть модифицирован за счёт использования альтернативных случайных величин при стохастическом моделировании. В свою очередь дальнейшее исследование предложенного подхода к оценке параметров модели

предполагает проведение вычислительных экспериментов при иных методах минимизации ошибки.

Отметим, что представление процесса развития заболевания в виде графа переходов применимо также для моделирования прогрессирования хронических заболеваний. Целочисленные показатели стохастического прогноза в таком случае могут описывать количество индивидов, у которых заболевание развивается схожим образом. Учитывая это, предложенный метод оценки доверительных интервалов может быть использован применительно к показателю вероятности развития каждой стадии заболевания в заданный момент времени.

### Литература

1. Махмуд Б.М. Повышение качества и точности противоэпидемической ситуации с применением комбинированной имитационной модели на основе стохастической компартментной модели и клеточного автомата // Инженерный вестник Дона. 2014. № 1. URL: [ivdon.ru/ru/magazine/archive/n1y2014/2273/](http://ivdon.ru/ru/magazine/archive/n1y2014/2273/)
2. Tang L., Zhou Y., Wang L., Purkayastha S., Zhang L., He J., Wang F., Song P. X.-K. A review of multi-compartment infectious disease models // International Statistical Review. 2020. Vol. 88. No. 2. pp. 462–513.
3. Reyné B., Saby N., Sofonea M. T. Principles of mathematical epidemiology and compartmental modelling application to COVID-19 // Anaesthesia, Critical Care & Pain Medicine. 2021. Vol. 41. No. 1. URL: [doi.org/10.1016/j.accpm.2021.101017](https://doi.org/10.1016/j.accpm.2021.101017)
4. Nguyen-Van-Yen B., Del Moral P., Cazelles B. Stochastic epidemic models inference and diagnosis with Poisson random measure data augmentation // Mathematical Biosciences. 2021. Vol. 335. URL: [doi.org/10.1016/j.mbs.2021.108583](https://doi.org/10.1016/j.mbs.2021.108583)

5. Jang J., Oh R. A review on Poisson, Cox, Hawkes, shot-noise Poisson and dynamic contagion process and their compound processes // *Annals of Actuarial Science*. 2021. Vol. 15. No. 3. pp. 623-644.

6. Молев М.Д., Занина И.А., Стуженко Н.И. Синтез прогнозной информации в практике оценки эколого-экономического развития региона // *Инженерный вестник Дона*. 2013. Т. 27. № 4. URL: [ivdon.ru/ru/magazine/archive/n4y2013/1993](http://ivdon.ru/ru/magazine/archive/n4y2013/1993)

7. Cortés-Carvajal P.D., Cubilla-Montilla M., González-Cortés D.R. Estimation of the instantaneous reproduction number and its confidence interval for modeling the COVID-19 pandemic // *Mathematics*. 2022. Vol. 10. No. 2. URL: [doi.org/10.3390/math10020287](https://doi.org/10.3390/math10020287)

8. Virtanen P., Gommers R., Oliphant T.E., Haberland M., Reddy T., Cournapeau D., Burovski E., Peterson P., Weckesser W., Bright J., Walt S.J., Brett M., Wilson J., Millman K.J., Mayorov N., Nelson A.R.J., Jones E., Kern R., Larson E., Carey C.J., Polat İ., Feng Y., Moore E.W., VanderPlas J., Laxalde D., Perktold J., Cimrman R., Henriksen I., Quintero E.A., Harris C.R., Archibald A.M., Ribeiro A.H., Pedregosa F., Mulbregt P. SciPy 1.0: fundamental algorithms for scientific computing in Python // *Nature methods*. 2020. Vol. 17. No. 3. pp. 261-272. URL: [doi.org/10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)

9. Январев С.Г., Бакланов А.Н., Шепелева А.О. Оптимизационный Метод Нелдера-Мида // *Компьютерные технологии в науке, производстве, социальных и экономических процессах: Материалы 16-ой Международной научно-практической конференции, посвященной 110-летию Южно-Российского государственного политехнического университета (НПИ) имени М.И. Платова*. Новочеркасск. 2016. С. 67-71.

10. Flahault A., Dias-Ferrao V., Chaberty P., Esteves K., Valleron A., Lavanchy D. FluNet as a tool for global monitoring of influenza on the Web // *Jama*. 1998. Vol. 280. No. 15. pp. 1330-1332.

---

## References

1. Makhmud B.M. Inzhenernyj vestnik Dona. 2014. № 1. URL: [ivdon.ru/ru/magazine/archive/n1y2014/2273/](http://ivdon.ru/ru/magazine/archive/n1y2014/2273/)
2. Tang L., Zhou Y., Wang L., Purkayastha S., Zhang L., He J., Wang F., Song P. X.-K. International Statistical Review. 2020. Vol. 88. No. 2. pp. 462–513.
3. Reyné B., Saby N., Sofonea M. T. Anaesthesia, Critical Care & Pain Medicine. 2021. Vol. 41. No. 1. URL: [doi.org/10.1016/j.accpm.2021.101017](https://doi.org/10.1016/j.accpm.2021.101017)
4. Nguyen-Van-Yen B., Del Moral P., Cazelles B. Mathematical Biosciences. 2021. Vol. 335. URL: [doi.org/10.1016/j.mbs.2021.108583](https://doi.org/10.1016/j.mbs.2021.108583)
5. Jang J., Oh R. Annals of Actuarial Science. 2021. Vol. 15. No. 3. pp. 623-644.
6. Molev M.D., Zanina I.A., Stuzhenko N.I. Inzhenernyj vestnik Dona. 2013. T. 27. № 4. URL: [ivdon.ru/ru/magazine/archive/n4y2013/1993](http://ivdon.ru/ru/magazine/archive/n4y2013/1993)
7. Cortés-Carvajal P.D., Cubilla-Montilla M., González-Cortés D.R. Mathematics. 2022. Vol. 10. No. 2. URL: [doi.org/10.3390/math10020287](https://doi.org/10.3390/math10020287)
8. Virtanen P., Gommers R., Oliphant T.E., Haberland M., Reddy T., Cournapeau D., Burovski E., Peterson P., Weckesser W., Bright J., Walt S.J., Brett M., Wilson J., Millman K.J., Mayorov N., Nelson A.R.J., Jones E., Kern R., Larson E., Carey C.J., Polat İ., Feng Y., Moore E.W., VanderPlas J., Laxalde D., Perktold J., Cimrman R., Henriksen I., Quintero E.A., Harris C.R., Archibald A.M., Ribeiro A.H., Pedregosa F., Mulbregt P. Nature methods. 2020. Vol. 17. No. 3. pp. 261-272. URL: [doi.org/10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
9. Yanvarev S.G., Baklanov A.N., Shepeleva A.O. Komp'yuternye tekhnologii v nauke, proizvodstve, sotsial'nykh i ekonomicheskikh protsessakh. 2016. pp. 67-71.
10. Flahault A., Dias-Ferrao V., Chaberty P., Esteves K., Valleron A., Lavanchy D. Jama. 1998. Vol. 280. No. 15. pp. 1330-1332.

**Дата поступления: 21.02.2025**

**Дата публикации: 25.04 2025**