

УДК 519.246.8+51-76
DOI: 10.15827/0236-235X.126.337-342

Дата подачи статьи: 20.12.18
2019. Т. 32. № 2. С. 337–342

Прогнозирование временного ряда инфекционной заболеваемости

С.А. Тарасова¹, к.пед.н., старший преподаватель, shedrina19@bk.ru

¹ Курский государственный медицинский университет, г. Курск, 305041, Россия

В статье отражены актуальность и степень разработанности проблемы прогнозирования инфекционной заболеваемости населения, предложен один из способов прогнозирования заболеваемости населения различными инфекциями на базе классической декомпозиции временного ряда.

Обычно в структуре временных рядов инфекционной заболеваемости выделяют тренд и сезонную составляющую с одним или двумя пиками в зависимости от типа инфекции, а также остаточную компоненту, которая должна удовлетворять условиям случайности, независимости и нормального распределения уровней с математическим ожиданием, равным нулю. При выполнении этих условий методы классической декомпозиции достаточно хорошо позволяют выявить как долгосрочную тенденцию развития процесса, так и сезонные изменения. Методика заключается в последовательной реализации процедур алгоритмического и аналитического выравнивания временного ряда и нахождении сезонной вариации в виде усредненных нормированных отклонений фактических уровней ряда от линии тренда, а также не предполагает в формировании индексов сезонности остаточной компоненты, что способствует более точным прогнозам детерминированных составляющих временного ряда.

На первом этапе алгоритма ряд выравнивается с помощью скользящих средних, что позволяет уменьшить остаточную компоненту и получить комбинацию трендовой и сезонной составляющих временного ряда. На втором этапе с помощью метода наименьших квадратов составляется уравнение тренда, отражающего долгосрочную тенденцию динамики. На третьем этапе рассчитываются индексы сезонности, которые показывают степень отклонения сезонного временного ряда от тренда. На четвертом этапе прогнозная модель проверяется на адекватность. На пятом этапе на основе экстраполяции тренда и с учетом индексов сезонности осуществляется прогноз инфекционной заболеваемости на будущие периоды.

В результате исследования с помощью описанной процедуры разработана адекватная модель прогнозирования заболеваемости населения России острыми респираторными вирусными инфекциями, верификация которой показала достаточную точность и достоверность выполненных на ее основе прогнозов.

Ключевые слова: прогноз, прогнозирование, модель, моделирование, временной ряд, тренд, сезонность, декомпозиция, инфекционная заболеваемость.

На сегодняшний день сохраняется весьма неблагоприятная эпидемиологическая обстановка по целому ряду опасных инфекционных заболеваний [1]. Выход из сложившейся ситуации заключается в своевременном и точном прогнозе заболеваемости населения различными инфекциями и принятии наиболее перспективных решений в плане подготовки системы медико-профилактических и терапевтических мероприятий. Таким образом, остается актуальной проблема теоретической разработки и практической реализации методов прогнозирования инфекционной заболеваемости для обеспечения санитарно-эпидемиологического благополучия граждан, сохранения и улучшения их здоровья.

В настоящее время математический аппарат прогнозирования инфекционной заболеваемости представлен довольно широким спектром

методик и процедур, обеспечивающих эффективность прогностических решений, принимаемых на их основе. Современными направлениями научного поиска в этой области являются как модификация классических методов анализа временных рядов применительно к заболеваемости населения, так и разработка принципиально новых, опирающихся на последние достижения науки и техники способов изучения динамики распространения инфекционных заболеваний. Так, в статье [2] сделан обзор наиболее значимых и употребительных подходов к прогнозированию инфекционной заболеваемости и развития эпидемического процесса, описаны методы исследования временных рядов на базе фильтрации, классического регрессионного анализа, байесовских и искусственных нейронных сетей, рассуждений на основе прецедентов. Автор подробно оста-

навливаются на так называемом биологическом подходе, который состоит в моделировании эпидемических показателей с помощью систем дифференциальных уравнений. Отмечается наибольшая эффективность смешанных техник прогнозирования, основанных на совместном использовании нескольких методов. В соответствии с этим в работе [3] для идентификации временных рядов инфекционной заболеваемости предложено вариативное моделирование, основанное на одновременном применении периодограммного и сингулярного подходов. В исследовании [4] для прогнозирования течения инфекционных заболеваний пациентов был апробирован целый ряд методов: параметрическая идентификация SIR-модели, анализ временных рядов с использованием показателя Херста, адаптивное прогнозирование Брауна и Хольта–Уинтерса, искусственные нейронные сети, тестирование нейро-нечеткой гибридной сети. Надо отметить, что нейросетевые модели прогнозирования временных рядов инфекционной заболеваемости обладают значительными преимуществами по сравнению с другими моделями и потому в настоящее время пользуются наибольшей популярностью [5]. В статье [6] приведены прогнозные модели, которые, являясь междисциплинарными, могут успешно использоваться и для анализа динамики распространения инфекционных заболеваний. В [7] предложен многокритериальный аппарат выбора оптимальной модели для прогнозирования временных рядов распространения инфекций. В работе [8] проведен анализ динамического ряда инфекционной заболеваемости с учетом влияния различных факторов, в частности, погодных-климатических условий, которые являются ведущими в процессах возникновения и распространения инфекционных заболеваний.

В структуре временных рядов инфекционной заболеваемости выделяют тренд и сезонную составляющую с одним или двумя пиками в зависимости от типа инфекции. Для анализа таких рядов обычно требуется провести их декомпозицию. В научной литературе описаны способы осуществления декомпозиции временного ряда [9, 10], чаще всего они основываются на спектральном анализе или процедурах сглаживания. После разделения ряда на тренд и сезонную составляющую оценивается вклад каждой компоненты в развитие процесса, делаются прогностические выводы.

В настоящей работе показан подход к прогнозированию рядов инфекционной заболевае-

мости населения на базе классической декомпозиции временного ряда, который заключается в последовательной реализации процедур алгоритмического и аналитического выравнивания ряда и нахождения сезонной вариации в виде усредненных нормированных отклонений фактических уровней ряда от линии тренда. В отличие от существующих в науке модификаций классической декомпозиции временного ряда [11, 12] представленная методика не предполагает в формировании индексов сезонности остаточной компоненты, что способствует более точным прогнозам детерминированных составляющих временного ряда. Алгоритм состоит из следующих существенных этапов. На первом этапе ряд выравнивается с помощью скользящих средних, что позволяет уменьшить остаточную компоненту и получить комбинацию трендовой и сезонной составляющих временного ряда. На втором этапе с помощью метода наименьших квадратов составляется уравнение тренда, отражающего долгосрочную тенденцию динамики. На третьем этапе рассчитываются индексы сезонности, которые показывают степень отклонения сезонного временного ряда от тренда. Заметим, что прогнозирование на базе классической декомпозиции временного ряда считается адекватным, если уровни остаточной компоненты ряда являются случайными, независимыми и нормально распределенными с математическим ожиданием, равным нулю, поэтому на четвертом этапе необходима проверка прогнозной модели на адекватность. На пятом этапе на основе экстраполяции тренда и с учетом индексов сезонности осуществляется прогноз инфекционной заболеваемости на будущие периоды.

Рассмотрим каждый этап подробнее. Используем статистические данные по заболеваемости населения ОРВИ за 2000–2017 гг. [13]. Обозначим исходный ряд:

$$y_t = y_{ij}, \quad (1)$$

где t – сквозной номер уровня ряда, $t = \overline{1, 216}$; i – номер года, $i = \overline{0, 17}$; j – номер месяца, $j = \overline{1, 12}$, причем $t = t_{ij} = 12i + j$, и представим его графически (рис. 1). Размерность уровня ряда (1) – 10^5 человек.

Анализ статистических данных показывает, что временной ряд (1) содержит тренд и сезонную составляющую с периодом, равным 12 месяцам. Для их выделения выравниваем ряд (1) скользящей средней с интервалом сглаживания, равным периоду сезонной составляющей,

с помощью формулы $\bar{y}_t = \frac{y_{t-6} + \sum_{m=t-5}^{t+5} y_m + y_{t+6}}{12}$, $t = \overline{7, 210}$.

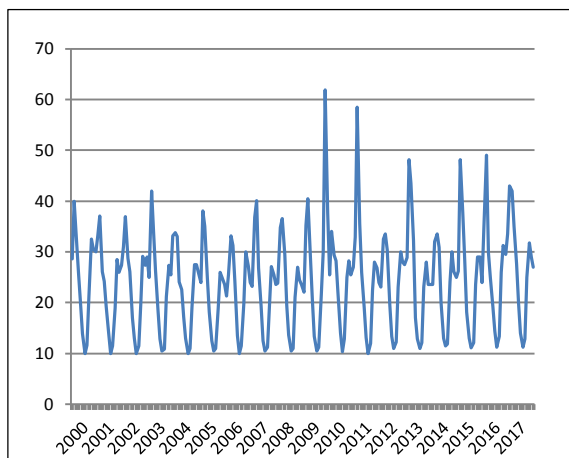


Рис. 1. Статистика заболеваемости населения ОРВИ в России

Fig. 1. Statistics of ARVI morbidity in Russia

Это необходимо для того, чтобы не исказить сезонную компоненту ряда.

Невыровненные значения в начале и в конце ряда отбрасываем. По оценке тренда \bar{y}_t методом наименьших квадратов составляем его уравнение $y = f(t)$. В рассматриваемом случае $y = 0,015t + 22,464$. (2)

Тренд представлен на рисунке 2.

Вычисляем отклонения исходного ряда от выровненного: $\Delta_{ij} = y_{ij} - y(t_{ij})$, которые будут содержать сезонную и остаточную компоненты.

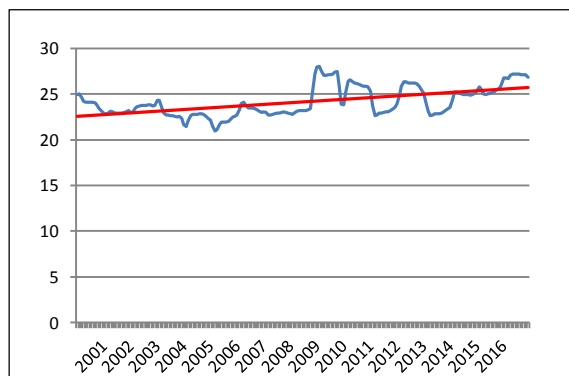


Рис. 2. Выровненный ряд заболеваемости населения ОРВИ с линией тренда

Fig. 2. Aligned ARVI morbidity time series with a trend line

Для каждого года $i = \overline{1, 16}$ рассчитываем дисперсию полученных отклонений по фор-

муле $\sigma_i^2 = \frac{\sum_{j=1}^{12} \Delta_{ij}^2 - \frac{\left(\sum_{j=1}^{12} \Delta_{ij}\right)^2}{12}}{11}$ и нормируем их:

$$\Delta'_{ij} = \frac{\Delta_{ij}}{\sigma_i}$$

Усредняя нормированные отклонения по годам, получаем сезонную волну (рис. 3):

$$v_j = \frac{\sum_{i=1}^{16} \Delta'_{ij}}{16}$$

Таким образом, действие остаточной компоненты временного ряда исключается.

Сезонную компоненту (рис. 4) ряда (1) получаем умножением среднего квадратического отклонения каждого года на сезонную волну: $v_{ij} = \sigma_i \cdot v_j$.

Рассчитываем индексы сезонности (рис. 5)

$$I_j = \frac{\sum_{i=1}^{16} \frac{y(t_{ij}) + v_{ij}}{y(t_{ij})}}{16}$$

Вычисляем остаточную компоненту ряда: $\varepsilon_{ij} = y_{ij} - y(t_{ij}) - v_{ij}$.

С помощью критериально-оценочного аппарата на уровне значимости 0,01 было выявлено, что остаточная компонента временного ряда (1) является случайной (критерий пиков), нормально распределенной (критерий Колмогорова), с математическим ожиданием, равным нулю (критерий Стьюдента), тем не менее, критерий Дарбина–Уотсона показал наличие слабой автокорреляции уровней остаточной компоненты. Следовательно, разработанная прогнозная модель является вполне адекватной и может быть использована для прогнозирования заболеваемости населения острыми респираторными вирусными инфекциями.

В соответствии с этим, экстраполируя тренд на будущие периоды времени, находим прогнозные значения для основной тенденции, а затем, умножая их на соответствующий индекс сезонности, получаем прогнозы, которые отражают как долгосрочную тенденцию, так и сезонную вариацию.

Например, рассчитаем прогноз на октябрь 2018 года ($t = 12 \cdot 18 + 10 = 226$). Сначала в уравнение тренда (2) подставляем $t = 226$: $y(226) = 0,015 \cdot 226 + 22,464 = 25,85$.

Затем, умножая получившееся значение на индекс сезонности для октября, получаем прогнозируемый уровень заболеваемости ОРВИ в

октябре 2018 года: $25,85 \cdot 1,19 = 30,76 \cdot 10^5$ человек.

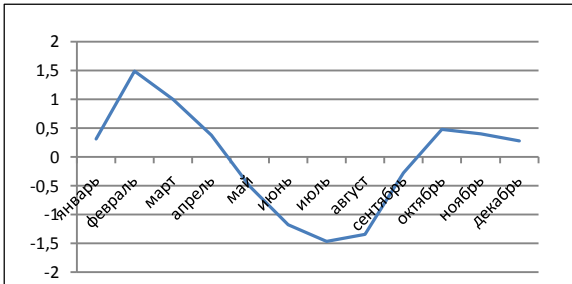


Рис. 3. Сезонная волна заболеваемости населения ОРВИ

Fig. 3. A seasonal wave of ARVI morbidity

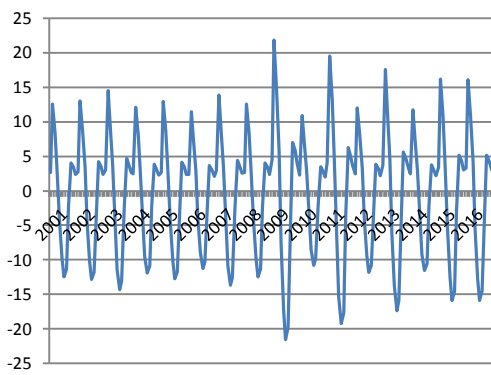


Рис. 4. Сезонная компонента временного ряда заболеваемости населения ОРВИ

Fig. 4. A seasonal component of ARVI morbidity time series

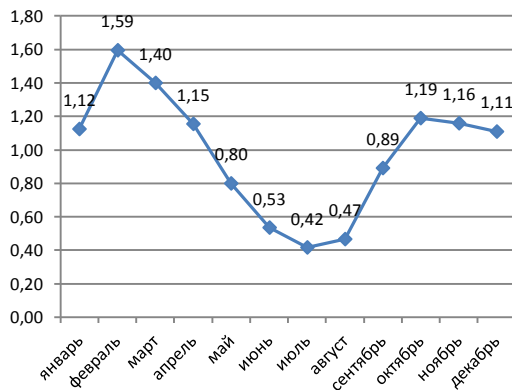


Рис. 5. Индексы сезонности заболеваемости населения ОРВИ

Fig. 5. Seasonality indices of ARVI morbidity

Фактический уровень заболеваемости в этом месяце был $29,26 \cdot 10^5$ человек, соответственно, ошибка прогноза составила 5,1 %.

Верификация модели по данным Роспотребнадзора за 2018 г. показала, что средняя относительная ошибка прогноза не превышает 7 %. На рисунке 6 представлены прогнозные и фактические значения заболеваемости населения ОРВИ в 2018 году.



Рис. 6. Ряд прогнозных значений с рядом фактических значений заболеваемости населения ОРВИ

Fig. 6. Series of forecast values and series of actual values of ARVI morbidity

Таким образом, метод прогнозирования инфекционной заболеваемости населения на основе классической декомпозиции временного ряда достаточно хорошо позволяет выявить как долгосрочную тенденцию, так и повторяющиеся сезонные колебания. Однако метод не отражает циклические изменения процесса (если таковые присутствуют), соответственно, в зависимости от необходимой точности разрабатываемых прогнозов могут потребоваться дополнительные исследования этой составляющей временного ряда.

Литература

1. Infectious diseases. World health statistics 2018: monitoring health for the SDGs, sustainable development goals. World Health Organization. URL: https://www.who.int/gho/publications/world_health_statistics/en/ (дата обращения: 30.11.2018).
2. Кондратьев М.А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование. 2013. Т. 5. № 5. С. 863–882.
3. Альсова О.К., Губарев В.В., Локтев В.Б. Использование вариативного моделирования при идентификации временных рядов инфекционной

- заболеваемости // Изв. ВолгГТУ. 2011. № 11. С. 42–47.
4. Котин В.В. Прогнозирование заболеваемости: динамические модели и временные ряды // Биотехносфера. 2014. № 3. С. 45–47.
5. Головинова В.Ю., Киреев С.Г., Котенко П.К., Минаев Ю.Л., Штамбург И.Н., Кузьмин С.Г. Нейросетевые модели прогнозирования заболеваемости в организованных коллективах // Вестн. Российской воен.-мед. акад. 2014. № 3. С. 150–154.
6. Пранов Б.М. Адекватные междисциплинарные модели в прогнозировании временных рядов статистических данных // Программные продукты и системы. 2018. Т. 31. № 3. С. 444–447. DOI: 10.15827/0236-235X.123.444-447.
7. Рыков А.С., Хорошилов В.О., Щипин К.С. Система прогнозирования инфекционной заболеваемости на основе многокритериального анализа временных рядов // Проблемы управления. 2005. № 1. С. 26–32.
8. Сташевский П.С., Яковина И.Н. Метод профилей для селекции признаков из временных рядов в задачах анализа данных // Автоматика и программная инженерия. 2015. № 4. С. 59–64.
9. Токмакова А.А. Выделение периодической компоненты из временного ряда // Машинное обучение и анализ данных. 2011. Т. 1. № 1. С. 40–50.
10. Юров В.М. Технология прогнозирования периодических экономических процессов на основе методов гармонического анализа в MS Excel // Вестн. МГОУ. 2018. № 3. С. 19–28. DOI: 10.18384/2310-6646-2018-3-19-28.
11. Dibrivny O.A. Comparative analysis of time series forecasting based on the trend model and adaptive Brown's model // Телекомунікаційні та інформаційні технології. 2018. № 1. С. 88–95 (англ.).
12. Мухаметжанова Ж.С. Прогнозирование по аддитивной и мультипликативной модели временного ряда // Вестн. Кыргызского национальн. ун-та. 2016. № 2. С. 7–15.
13. Щербакова Е.М. Заболеваемость населения России, 2016–2017 годы // Демоскоп Weekly. 2018. № 767–768. URL: <http://demoscope.ru/weekly/2018/0767/barom01.php> (дата обращения: 30.11.2018).

Software & Systems

DOI: 10.15827/0236-235X.126.337-342

Received 20.12.18

2019, vol. 32, no. 2, pp. 337–342

Forecasting time series of infectious morbidity

S.A. Tarasova¹, Ph.D. (Pedagogy), Senior Lecturer, shedrina19@bk.ru

¹ Kursk State Medical University, Kursk, 305041, Russian Federation

Abstract. The paper presents the topicality and the extent of prior investigation of the problem of forecasting infectious morbidity of the population. It also proposes one of the methods of forecasting population's infectious morbidity based on the classical time series decomposition.

Typically, the structure of infectious morbidity time series consists of a trend and a seasonal component with one or two peaks depending on the type of infection, as well as a residual component, which must satisfy the conditions of randomness, independence and normal distribution of levels with a mathematical expectation equal to zero. When these conditions are fulfilled, the classical decomposition methods identify both the long-term tendency of the process development and seasonal changes. The technique assumes algorithmic and analytical alignment of time series, finding seasonal variations as averaged normalized deviations of actual series levels from the trend line. It does not imply a residual component in seasonality indices, which provides more accurate forecasts of deterministic components of the time series.

The algorithm consists of the following stages. At the first stage, moving averages align the time series, which allows reducing a residual component and obtaining a combination of a trend and seasonal component of the time series. The second stage includes generation of a trend equation using the method of least squares. The trend equation reflects a long-term tendency of the dynamics. The third stage includes calculation of seasonality indices, which show the degree of the seasonal time series deviation from the trend. At the fourth stage, the forecasting model is checked for adequacy. At the fifth stage includes forecasting infectious morbidity for future periods based on extrapolation of the trend and taking into account seasonality indices.

The study represents an adequate model for forecasting the population's morbidity of acute respiratory viral infections in Russia; its verification has shown sufficient accuracy and reliability of further forecasts.

Keywords: forecast, forecasting, model, modeling, time series, trend, seasonality, decomposition, infectious morbidity.

References

1. Infectious diseases. World health statistics 2018: monitoring health for the SDGs, sustainable development goals. *World Health Organization*. Available at: https://www.who.int/gho/publications/world_health_statistics/en/ (accessed November 30, 2018).
2. Kondratev M.A. Forecasting methods and models of diseases spread. *Computer Research and Modeling*. 2013, vol. 5, no. 5, pp. 863–882 (in Russ.).
3. Alsova O.K., Gubarev V.V., Loktev V.B. Using variable modeling to identify infectious morbidity time series. *Izvestiya VSTU*. 2011, no. 11, pp. 42–47 (in Russ.).
4. Kotin V.V. Forecasting morbidity: dynamic models and time series. *Biotekhnosfera*. 2014, no. 3, pp. 45–47 (in Russ.).
5. Golovinova V.Yu., Kireev S.G., Kotenko P.K., Minaev Yu.L., Shtamburg I.N., Kuzmin S.G. Neural network models for forecasting morbidity in organized groups. *The Bulletin of S.M. Kirov Military Medical Academy*. 2014, no. 3, pp. 150–154 (in Russ.).
6. Pranov B.M. Adequate interdisciplinary models in forecasting time series of statistical data. *Software & Systems*. 2018, vol. 31, no. 3, pp. 444–447 (in Russ.). DOI: 10.15827/0236-235X.123.444-447.
7. Rykov A.S., Khoroshilov V.O., Shchipin K.S. System for forecasting infectious morbidity based on multi-criteria time series analysis. *Control Sciences*. 2005, no. 1, pp. 26–32 (in Russ.).
8. Stashevsky P.S., Yakovina I.N. Method of feature engineering for time series in data analysis problems. *Automatics & Software Enginery*. 2015, no. 4, pp. 59–64 (in Russ.).
9. Tokmakova A.A. Selection a periodic component from a time series. *J. of Machine Learning and Data Analysis*. 2011, vol. 1, no. 1, pp. 40–50 (in Russ.).
10. Yurov V.M. A harmonic analysis-based technique for forecasting periodic economic processes in MS Excel. *Bulletin MRSU*. 2018, no. 3, pp. 19–28 (in Russ.). DOI: 10.18384/2310-6646-2018-3-19-28.
11. Dibrivny O.A. Comparative analysis of time series forecasting based on the trend model and adaptive Brown's model. *Telecommunication and Information Technologies*. 2018, no. 1, pp. 88–95.
12. Mukhametzhanova Zh.S. Forecasting by additive and multiplicative models of time series. *Bulletin of Kyrgyz National Univ*. 2016, no. 2, pp. 7–15 (in Russ.).
13. Shcherbakova E.M. Morbidity of Russian population, 2016-2017. *Demoskop Weekly*. 2018, no. 767–768. Available at: <http://demoscope.ru/weekly/2018/0767/barom01.php> (accessed November 30, 2018).

Для цитирования

Тарасова С.А. Прогнозирование временного ряда инфекционной заболеваемости // Программные продукты и системы. 2019. Т. 32. № 2. С. 337–342. DOI: 10.15827/0236-235X.126.337-342.

For citation

Tarasova S.A. Forecasting time series of infectious morbidity. *Software & Systems*. 2019, vol. 32, no. 2, pp. 337–342 (in Russ.). DOI: 10.15827/0236-235X.126.337-342.