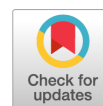


TIME SERIES ANALYSIS FOR MODELING AND PREDICTING CONFIRMED CASES OF INFLUENZA A IN ALGERIA



D. Seba^a, N. Benakle^b, K. Belaïde^b

^a Higher School of Informatics, Sidi Bel Abbès, Algeria

^b University of Bejaïa, Bejaïa, Algeria

Abstract. Influenza A is a subtype of the influenza virus that primarily infects birds and mammals, causing respiratory illness. It is characterized by its ability to mutate rapidly, leading to various strains and occasional pandemics. Objective. This paper is dedicated to studying the distribution behavior and predicting confirmed cases of Influenza A within the Algerian context, a highly infectious disease that causes widespread illness and deaths both in Algeria and globally. *Materials and methods.* To predict confirmed cases of Influenza A, we implemented several statistical models, including ARIMA, Seasonal ARIMA (SARIMA), ETS, BATS, and the machine learning technique RNN, which is widely recognized in the literature. We then conducted a comparative study using performance measures to evaluate these models. *Results.* We used RMSE to determine the best-performing model. Our findings indicate that RNN outperformed the others due to its ability to handle complex patterns, including seasonal components and memory. SARIMA and BATS also performed well, thanks to their capacity to manage seasonal patterns. In contrast, ARIMA and ETS showed the poorest performance. *Conclusion.* This study employed a comprehensive approach to develop a model for predicting confirmed cases of Influenza A in Algeria. The results enhance our understanding of the potential future behavior of this disease and contribute to effective risk management strategies.

Key words: influenza A, prediction, risk management, time series, BATS, Algeria.

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ ДЛЯ МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ ПОДТВЕРЖДЕННЫХ СЛУЧАЕВ ГРИППА А В АЛЖИРЕ

Себа Д.¹, Бенаклеф Н.², Белаиде К.²

¹ Высшая школа информатики, г. Сиди-Бель-Аббес, Алжир

² Университет Беджаи, г. Беджая, Алжир

Резюме. Грипп А является подтипом вируса гриппа, который в первую очередь поражает птиц и млекопитающих, вызывая респираторные заболевания, и характеризуется способностью быстро мутировать, что приводит к появлению разнообразия штаммов и периодическим пандемиям. Настоящая статья посвящена изучению в Алжире распространения и прогнозированию подтвержденных случаев гриппа А, высокоинфекционного заболевания, которое вызывает широко распространенные заболевания и смертность как в Алжире, так и во всем мире. *Материалы и методы.* Для прогнозирования подтвержденных случаев гриппа А были применены несколько статистических моделей, включая ARIMA, Seasonal ARIMA (SARIMA), ETS, BATS и широко признанный метод машинного обучения RNN. Далее мы провели сравнительное исследование с использованием показателей производительности для оценки указанных моделей. *Результаты.* Для определения наиболее

Адрес для переписки:

Джиллали Себа
22016, Алжир, г. Сиди-бель-Аббес, ул. 8 мая 1945 года,
Высшая школа информатики.
Тел.: +213-7-82-01-74-34. E-mail: d.seba@esi-sba.dz

Contacts:

Djillali Seba
22016, Sidi Bel Abbès, Algérie, Rue du 8 Mai 1945,
École Supérieure en Informatique.
Phone: +213-7-82-01-74-34. E-mail: d.seba@esi-sba.dz

Для цитирования:

Себа Д., Бенаклеф Н., Белаиде К. Анализ временных рядов для моделирования и прогнозирования подтвержденных случаев гриппа А в Алжире // Инфекция и иммунитет. 2025. Т. 15, № 1. С. 168–172.
doi: 10.15789/2220-7619-TSA-17693

Citation:

Seba D., Benakle N., Belaïde K. Time series analysis for modeling and predicting confirmed cases of influenza A in Algeria // Russian Journal of Infection and Immunity = Infektsiya i immunitet, 2025, vol. 15, no. 1, pp. 168–172. doi: 10.15789/2220-7619-TSA-17693

эффективной модели проводилась оценка среднеквадратической ошибки. Наши результаты показывают, что RNN превзошел другие модели благодаря своей способности обрабатывать сложные шаблоны, включая сезонные компоненты и наличие памяти. SARIMA и BATS также показали хорошие результаты благодаря своей способности управлять сезонными закономерностями. Напротив, ARIMA и ETS показали самые плохие результаты. **Вывод.** В приводимом исследовании использовался комплексный подход для разработки модели прогнозирования подтвержденных случаев гриппа А в Алжире. Полученные результаты расширяют наше понимание потенциального будущего распространения данного заболевания и способствуют эффективным стратегиям управления рисками.

Ключевые слова: грипп А, прогнозирование, управление рисками, временные ряды, BATS, Алжир.

Introduction

Influenza A is a viral infection that affects the respiratory system. It is one of the four types of influenza viruses and can cause symptoms such as cough, body aches, and sore throat. Highly contagious, Influenza A spreads through tiny droplets of bodily fluid released during coughing, sneezing, or talking. Symptoms often include fever, chills, fatigue, and other related discomforts.

In the early 20th century, scientific knowledge was advanced enough to predict the recurrence of influenza, which had twice reached pandemic levels in the late 19th century. However, it was largely ineffective in mitigating the devastating impact of the 1918 pandemic. Since then, humanity has made significant strides against the disease, developing the capability to design and produce vaccines and antiviral drugs to prevent or lessen infections.

The World Health Organization (WHO) estimates that globally there are 3–5 million cases of severe illness and 290 000–650 000 deaths annually due to influenza-related respiratory conditions.

Nowadays, predicting Influenza A helps minimize the health, economic, and social impacts of the virus by enabling proactive and well-coordinated responses. The debate on forecasting Influenza A involves researchers from various disciplines who use a range of methodologies, including statistical, machine learning, and deep learning techniques, referencing various studies such as Goldstein et al. [7], Xu et al. [14], Zheng et al. [16], Kandula et al. [8], Khan et al. [9], Cheng et al. [4], Wolk et al. [13], Xue et al. [15], Boostani et al. [3], Al-Qaness et al. [2], Seba et al. [12].

This topic has been treated in the Algerian context by several works such as [6] and [11]. The objective of our work is to predict the behavior of new cases in Algeria using time series analysis. We employ two widely recognized approaches from the literature: statistical methods and machine learning techniques for time series analysis.

Methods and Materials

Descriptive Data

The epidemiology of seasonal influenza is well defined in many parts of the world, especially in de-

veloped countries. However, in other regions, much less is known about the epidemiology of Influenza A, notably in Algeria (Fig. 1).

We collect data from: Our world in data-Influenza. We observed that Influenza A exhibits a seasonal winter pattern because the cold, dry air of winter provides ideal conditions for the virus's prolonged survival. The reduced humidity during this season enhances the likelihood of infection. The significant decrease in Influenza cases from 2020 to 2022 can be attributed to several factors related to the COVID-19 pandemic: Implementing public health strategies, including mask wearing, hand hygiene, social distancing, and lockdowns on a large scale, substantially decreased the spread of respiratory infections, including Influenza A.

Changes in the individual's awareness of safety amid the COVID-19 outbreak [10]. Analyzing monthly confirmed cases involves treating the data as a time series. Throughout the literature, various approaches such as statistical, machine learning methods have commonly been employed.

Preprocessing data analysis

Our data does not contain any missing values but contains seasonal patterns.

We employ the Augmented Dickey-Fuller (ADF) test to assess the stationarity of the present time series. A p-value of 0.01, which is smaller than the significance level of 0.05, indicates that the time series is stationary.

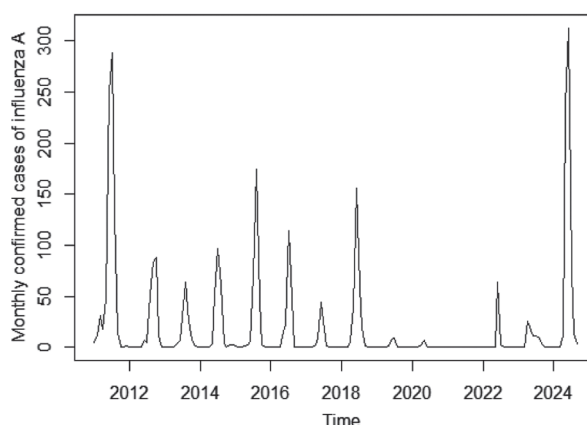


Figure 1. Monthly confirmed cases of influenza A

For data to be considered stationary, the statistical characteristics of the system must remain constant over time. This does not mean that the values of each data point must be identical; rather, the overall behavior of the data should remain consistent.

We introduce the Partial Autocorrelation Function (PACF) and the Autocorrelation Function (ACF) to measure the memory of the most effective model from the ACF and PACF plots, we observe that the autocorrelation decays exponentially, indicating that the data has short memory. Additionally, we note the presence of seasonal components (Fig. 2).

Methodology

Various time series analysis models and techniques are employed to determine the most efficient method for handling validated cases of influenza A, utilizing both statistical and machine learning models.

We take our data as a time series and split it into two separate sets: the training set and the test set. The test set, comprising 15% of the data, is used to validate the best model, while the training set consists of the remaining 85%. We evaluate a model's performance using the Root Mean Square Error (RMSE).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - z_i)^2} \quad (1)$$

Results

Predictive models

ARIMA and ARFIMA model

Autoregressive integrated moving average (ARIMA) models predict future values based on past values, it gauges the strength of one dependent variable relative to other changing variables.

A stochastic process $(X_t)_{t \geq 0}$ is said to be an ARIMA(p, d, q) an integrated mixture autoregressive moving average model if it satisfies the following equation:

$$\phi(L)(1-L)^d X_t = \theta(L)\varepsilon_t \quad \forall t \geq 0 \quad (2)$$

where $d \in \mathbb{N}$, L is lag operator, $\varepsilon_t \sim N(0, \sigma^2)$ i.i.d. errors, with $\sigma^2 < \infty$.

$$\phi(L) = (1 - \phi_1 L - \dots - \phi_p L^p) \text{ with } \phi_p \neq 0$$

$$\theta(L) = (1 - \theta_1 L - \dots - \theta_q L^q) \text{ with } \theta_q \neq 0$$

Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.

There are four seasonal elements that are not part of ARIMA that must be configured; they are: P: Seasonal autoregressive order, D: Seasonal difference order, Q: Seasonal moving average order, m: The number of time steps for a single seasonal period SARIMA(p,d,q)(P,D,Q)m

$$\phi'(L)\phi(L)(1-L^m)^d(1-L)^d X_t = \theta'(L)\theta(L)\varepsilon_t \quad \forall t \geq 0 \quad (3)$$

ETS model

The ETS models are time series models with an underlying state space model consisting of a level component, a trend component (T), a seasonal component (S), and an error term (E). This method produces forecasts that are weighted averages of past observations where the weights of older observations exponentially decrease.

BATS model

The BATS (Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components) model is a time series forecasting model that was proposed by De Livera et al. [5].

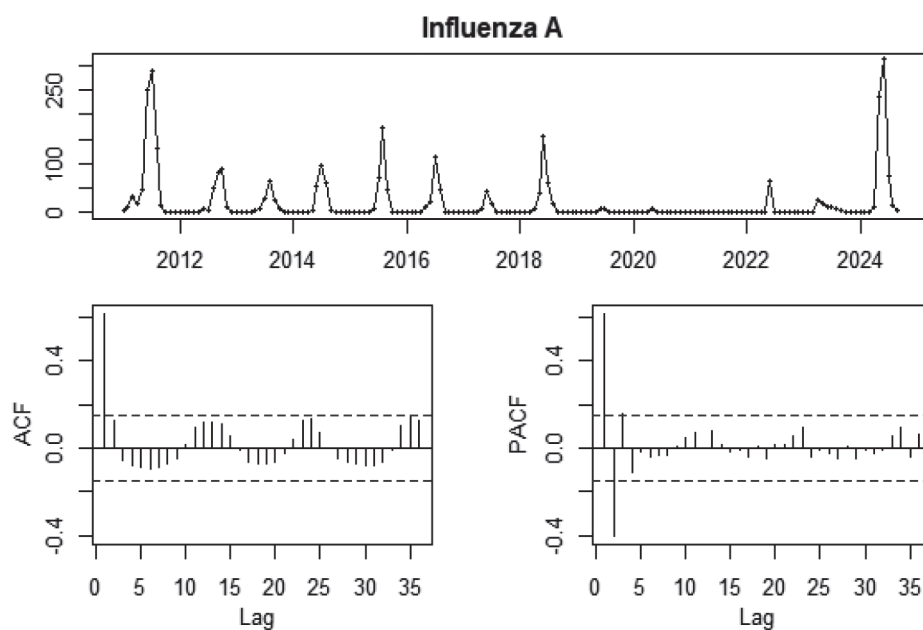


Figure 2. ACF and PACF plots

Box-Cox Transformation component is used to transform the data to achieve normality and stabilize the variance. The ARMA (Autoregressive Moving Average) Errors component is used to model the residuals of the time series data, which are assumed to be independent and identically distributed. Finally, the Seasonal component is used to model the seasonal patterns in the data.

Recurrent Neural Network RNN model

A Recurrent Neural Network (RNN) model for regression is a type of neural network designed to process sequential data by maintaining a memory of previous inputs.

Sequential Data Handling: RNNs are ideal for tasks where data points are dependent on previous ones, due to their ability to maintain information over sequences.

Memory: RNNs have internal memory (hidden states) that captures information from previous time steps, allowing them to learn patterns and dependencies over time.

Structure: An RNN consists of layers of neurons where each neuron receives inputs not only from the current time step but also from its own previous output.

Backpropagation Through Time (BPTT): The training of RNNs involves a variation of backpropagation (BPTT), which updates weights by considering the entire sequence of data.

Empirical Results

Here, we illustrate the predicted results graphically using several models “Statistical and machine learning models” (Fig. 3).

We show numerical results using this measure such as RMSE, a smaller RMSE and MAE indicates better performance (Tabl.).

Discussion

We have applied statistical models: ARIMA, SARIMA, BATS, and ETS, as well as a machine learning model, RNN.

ARIMA is not suitable for modeling and predicting this data due to the lack of seasonal patterns. Therefore, we opted for SARIMA, which can handle seasonal patterns. SARIMA performed well in forecasting the confirmed cases in winter 2023 but exhibited poor performance for the winter of 2024 due to its short persistence.

ETS had the worst performance. We applied simple exponential smoothing but could not use multiplicative errors and seasonal components due to negative values. Similarly, the additive case also resulted in negative values.

BATS performed well due to its ability to handle seasonal components and its treatment of errors as ARMA, meaning they are autocorrelated (dependent), which is more realistic compared to the independent errors assumed by ARIMA and SARIMA.

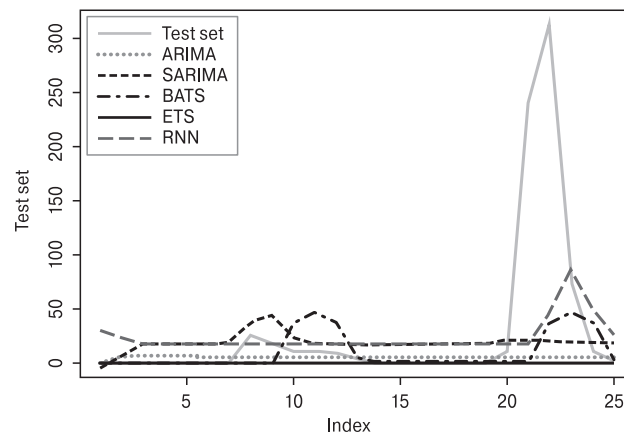


Figure 3. Predicting confirmed cases of Influenza A in Algeria

Table. Measure of performance

Models	ARIMA	SARIMA	ETS	BATS	RNN
RMSE	78.319	75.176	80.353	74.088	71.335

To improve the results, we replaced negative values with zeroes.

The RNN model performed the best due to its capacity to handle complex patterns, including non-linear and periodic trends, and its ability to overcome the problem of memory. However, it had problems modeling the rest of the year except for the wintertime.

Suggestions for reducing the number of confirmed cases of Influenza A

A combination of public health initiatives, personal efforts, and preventative measures are needed to decrease the number of confirmed cases of influenza A. The following are some crucial procedures:

Vaccination: Boost the use of yearly influenza vaccinations, which aim to protect people against the most prevalent strains that emerge each season.

Public Health Campaigns: Organize awareness programs to inform people about the value of immunizations, good hand cleanliness, and proper respiratory protocol.

Social Distancing: Take steps to avoid close proximity in crowded regions, particularly during the prime time of the flu season. This can involve advising people to stay away from crowded places, work from home, and keep a safe distance from other people.

Surveillance and Early Detection: Implement robust surveillance systems to: Monitor Influenza Activity: Track the spread and evolution of influenza strains in real-time. Identify Outbreaks: Quickly detect and respond to outbreaks to contain their spread.

Data Sharing: Collaborate with international health organizations for data sharing and coordinated response efforts.

Conclusion

To summarize, this work examined the behavior of confirmed Influenza A cases in Algeria and applied various statistical and machine learning models to predict the future behavior of this phenomenon. This approach enhances our understanding of the disease's future trends. Based on our find-

ings, we have suggested recommendations to help reduce the number of confirmed Influenza A cases.

Acknowledgement

We acknowledge the support of "Direction Générale de la Recherche Scientifique et du Développement Technologique DGRSDT". MESRS ALGERIA.

References

1. Ali S.T., Cowling B.J. Influenza virus: tracking, predicting, and forecasting. *Annu Rev. Public Health*, 2021, vol. 42, pp. 43–57. doi: 10.1146/annurev-publhealth-010720-021049
2. Al-Qaness M.A.A., Ewees A.A., Fan H., Abd Elaziz M. Optimized forecasting method for weekly influenza confirmed cases. *Int. J. Environ. Res. Public Health*, 2020, vol. 17, no. 10: 3510. doi: 10.3390/ijerph17103510
3. Boostani R., Rismanchi M., Khosravani A., Rashidi L., Kouchaki S. Presenting a hybrid method in order to predict the 2009 pandemic influenza A (H1N1). *J. Health. Med. Inform.*, 2012, vol. 3, no. 1, pp. 31–43. doi: 10.4172/2157-7420.1000112
4. Cheng H.Y., Wu Y.C., Lin M.H., Liu Y.L., Tsai Y.Y., Wu J.H., Pan K.H., Ke C.J., Chen C.M., Liu D.P., Lin I.F., Chuang J.H. Applying machine learning models with an ensemble approach for accurate real-time influenza forecasting in Taiwan: development and validation study. *J. Med. Internet Res.*, 2020, vol. 22, no. 8: e15394. doi: 10.2196/15394
5. De Livera A.M., Hyndman R.J., Snyder R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.*, 2011, vol. 106, no. 496, pp. 1513–1527. doi: 10.1198/jasa.2011.tm09771
6. Feradi F., Bouhata R., Kalla M.I., Kalla M. Assessing avian influenza vulnerability using geographically weighted regression, Batna Algeria. *The Arab. World. Geographer*, 2023, vol. 26, no. 1, pp. 76–87. doi: 10.5555/1480-6800-26.1.76
7. Goldstein E., Cobey S., Takahashi S., Miller J.C., Lipsitch M. Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. *PLoS Med.*, 2011, vol. 8, no. 7: e1001051. doi: 10.1371/journal.pmed.1001051
8. Kandula S., Yamana T., Pei S., Yang W., Morita H., Shaman J. Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *J R Soc. Interface*, 2018, vol. 15, no. 144: 20180174. doi: 10.1098/rsif.2018.0174
9. Khan M.A., Abidi W.U.H., Ghamdi M.A.A., Almotiri S.H., Saqib S., Alyas T., Khan K.M., Mahmood N. Forecast the influenza pandemic using machine learning. *Comput. Mater. Contin.*, 2021, vol. 66, no. 1, pp. 331–340. doi: 10.32604/cmc.2020.012148
10. Lu Y., Wang Y., Shen C., Luo J., Yu W. Decreased incidence of influenza during the COVID-19 pandemic. *Int. J. Gen. Med.*, 2022, vol. 15, pp. 2957–2962. doi: 10.2147/IJGM.S343940
11. Mejia K., Viboud C., Santillana M. Leveraging Google search data to track influenza outbreaks in Africa. *Gates Open Research*, 2019, vol. 3, no. 1653: 1653. doi: 10.12688/gatesopenres.13072.1
12. Seba D., Belaïde K. Forecasting infection fatality rate of COVID-19: measuring the efficiency of several hybrid models. *Russian Journal of Infection and Immunity*, 2024, vol. 14, no. 2, pp. 313–319. doi: 10.15789/2220-7619-FIF-17548
13. Wolk D.M., Lanyado A., Tice A.M., Shermohammed M., Kinar Y., Goren A., Chabris C.F., Meyer M.N., Shoshan A., Abedi V. Prediction of influenza complications: development and validation of a machine learning prediction model to improve and expand the identification of vaccine-hesitant patients at risk of severe influenza complications. *J. Clin. Med.*, 2022, vol. 11, no. 15: 4342. doi: 10.3390/jcm11154342
14. Xu Q., Gel Y.R., Ramirez Ramirez L.L., Nezafati K., Zhang Q., Tsui K.L. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. *PLoS One*, 2017, vol. 12, no. 5: e0176690. doi: 10.1371/journal.pone.0176690
15. Xue H., Bai Y., Hu H., Liang H. Regional level influenza study based on Twitter and machine learning method. *PLoS One*, 2019, vol. 14, no. 4: e0215600. doi: 10.1371/journal.pone.0215600
16. Zheng Y., Wang K., Zhang L., Wang L. Study on the relationship between the incidence of influenza and climate indicators and the prediction of influenza incidence. *Environ. Sci. Pollut. Res. Int.*, 2021, vol. 28, no. 1, pp. 473–481. doi: 10.1007/s11356-020-10523-7

Авторы:

Себа Д., д.мат.н., доцент, лаборатория прикладной математики, факультет математики, Высшая школа информатики, г. Сиди-бель-Аббес, Алжир;
Бенаклеф Н., аспирант по математике, специальность «Вероятность и статистика» в университете Беджая, лаборатория прикладной математики в Университете Беджая, г. Беджая, Алжир;
Белаиде К., д.мат.н., профессор, лаборатория прикладной математики, факультет математики, Университет Беджая, г. Беджая, Алжир.

Authors:

Seba D., Doctor in Mathematics, Assistant Professor, Laboratory of Applied Mathematics, Department of Mathematics, Higher School of Computer Science (Ecole Supérieure en Informatique), Sidi Bel Abbes, Algeria;
Benaklef N., PhD Student in Mathematics, Speciality "Probability and Statistics" at the University of Bejaia, Member of Applied Mathematics Laboratory, University of Bejaia, Bejaia, Algeria;
Belaïde K., Doctor in Mathematics, Full Professor, Laboratory of Applied Mathematics, Department of Mathematics, University of Bejaia, Bejaia, Algeria.

Поступила в редакцию 14.06.2024
 Отправлена на доработку 28.07.2024
 Принята к печати 29.07.2024

Received 14.06.2024
 Revision received 28.07.2024
 Accepted 29.07.2024