第 47 卷 第 9 期
2020 年 9 月

湖南大学学报（自然科学版）
Journal of Hunan University (Natural Sciences）

Vol. 47. No. 9.
September 2020

Prediction case notification rates for tuberculosis in eight countries

Seng Hansun

Universitas Multimedia Nusantara, Tangerang, Indonesia

Abstract: This article describes a novel approach to predicting tuberculosis case notification rates based on historical data for the eight countries with the highest tuberculosis rates. Using a relatively new prediction method known as Brown's Weighted Exponential Moving Average, we produced excellent prediction results for the selected countries. In one example, we describe the implementation of the proposed tuberculosis case notification rates prediction model in China, which led to the most accurate prediction result in the study. The effectiveness of this new method was confirmed by our calculation of the prediction error using the mean absolute percentage error, showing that China had the lowest mean absolute percentage error value at 2.3606444%. New research results confirm that there is an increasing trend in tuberculosis case notification rates for most countries included in this research. These results can be used to support the decision-making process for all related stakeholders, including the governments of these countries, when managing the spread of tuberculosis.

Keywords: Brown's weighted exponential moving average, case notification rate, high burden countries, prediction, tuberculosis.

八個國家的結核病預測病例通報率

摘要：本文介绍了一种基于结核病发病率最高的八个国家的历史数据预测结核病病例报告率的新颖方法。使用称为布朗的加权指数移动平均值的相对较新的预测方法，我们为选定的国家/地区提供了出色的预测结果。在一个示例中，我们描述了在中国实施所建议的结核病病例报告率预测模型的方法，从而使该研究的预测结果最准确。通过我们使用平均绝对百分比误差计算预测误差，证实了该新方法的有效性，表明中国的平均绝对百分比误差值最低，为 2.3606444％。新的研究结果证实，这项研究包括的大多数国家/地区的结核病报告率呈上升趋势。这些结果可用于管理所有相关利益攸关方（包括这些国家的政府）在管理结核病蔓延时的决策过程。

关键词：布朗的加權指數移動平均線，病例通報率，高負擔國家，預測，結核病.

Introduction

Tuberculosis (TB) is a communicable disease that originates from bacteria called *Mycobacterium tuberculosis* [1]. It is easily spread through the air and typically infects the lungs (Pulmonary TB) or other sites (Extrapulmonary TB), such as the kidneys, bones, and spine [2]. In the past decade, tuberculosis was listed as one of the top ten causes of death around the world [3], which should concern stakeholders everywhere, especially governments and public health experts.

In particular, TB has become a public health threat in low- and middle-income countries (LMICs) [4]. Out of the 202 countries and territories mapped by the World Health Organization (WHO), the 48 high TB-burden countries are predominantly LMICs [1]. According to the same report, in 2018, two-thirds of the global burden of TB was shared by only eight countries: namely, "India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (6%), Nigeria (4%), Bangladesh (4%), and South Africa (3%)" (Figure 1).

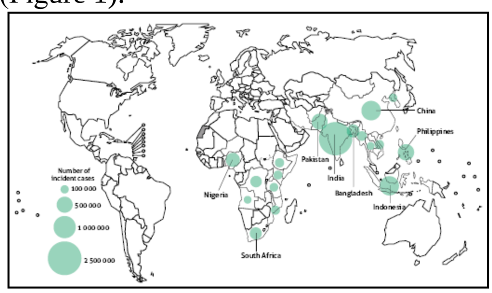


Fig. 1 TB incidence estimation in 2018 [1]

Received (date): August 1, 2020
About the author: Seng Hansun, Universitas Multimedia Nusantara, Tangerang, Indonesia
Corresponding author Seng Hansun, seng.hansun@lecturer.umn.ac.id

In this study, the TB case notification rates (CNR) for these eight countries will be predicted using a time series forecasting method. TB notification is defined as "the reporting process of diagnosed TB cases to relevant health authorities" [5]. TB notification is important for local contact tracing and assessing a country's TB incidence [6]. Therefore, the aim of this study is to find patterns or trends in TB CNR to help the decision-makers in these countries take proper action and form policies to handle the global burden of TB. Of note, two previous studies have discussed TB case prediction. Halim et al. [7] used the fuzzy linear regression method to predict TB CNR in the province of Surabaya and obtained good prediction results. In addition, Li et al. [8] used the autoregressive integrated moving average and generalized regression neural network (ARIMA-GRNN) hybrid model to predict TB incidence rates in the city of Lianyungang, Jiangsu Province, China. Both publications used soft computing methods, which can produce good results but generally are more complex and computationally expensive.

In this study, TB case notifications for all the eight countries will be predicted by using a time series forecasting method. TB notification is defined as "the reporting process of diagnosed TB cases to relevant health authorities" [5]. It is important for the local contact tracing and assessing the country's TB incidence case [6]. Therefore, the aim of this study is to find the pattern or trend of TB case notifications in those countries that can help the decision-makers to take proper action and policy in handling TB global burden.

There are two previous pieces of research that had discussed the TB case prediction. Halim et al. [7] use the fuzzy linear regression method to predict Surabaya province's TB case notification. They found that the model built by using the prediction method could give a good prediction result. Another publication is from Li et al. [8], where they used the hybrid model from ARIMA-GRNN to predict TB incidence rate in Lianyungang city, Jiangsu province, China. Both publications used soft computing methods that could give a good result, but generally are more complex and computationally expensive.

As one of the time series forecasting methods, Brown's Weighted Exponential Moving Average (B-WEMA) was first introduced in 2016 [9]. Since then, the method has been widely accepted and applied to different kinds of problems, such as Forex forecasting [10], stock rate of return [11; 12], and air quality index prediction [13]. In [10], B-WEMA was compared with WEMA and got a smaller Mean Square Error (MSE) value at 6.97030E-5. Similarly, in [11], B-WEMA also was compared with WEMA and got smaller MSE at 6124.222 and smaller Mean Absolute Percentage Error (MAPE) at 1.831685%. In [12], B-WEMA was compared with its building block methods, i.e., Weighted Moving Average (WMA) and Brown's Double Exponential Smoothing (B-DES). From the experiments undertaken, they found that B-WEMA has the lowest MSE and MAPE values at 2100.488 and 1.4959%. Lastly, in [13], B-WEMA excels WMA, B-DES, and Exponential Moving Average (EMA) with MSE at 114.326 and MAPE at 20.062%. Based on those research, B-WEMA is proven to be a robust and accurate forecasting method that can be used in different kinds of scenarios. Hence, in this study, B-WEMA will be used to predict the TB case notifications for the top eight TB-burden countries.

The next section will briefly discuss B-WEMA as the forecasting method used in this study. Section 3 will present the forecasting results of TB incidence for the top eight high burden countries, and some concluding remarks will be given in Section 4.

1 B-WEMA method

B-WEMA is a hybrid forecasting method that combines the procedure of Weighted Moving Average with Brown's Double Exponential Smoothing (B-DES) [9]. It can be used to predict the future values of a given time series data with a trend pattern [13]. Since it follows the B-DES procedure, it has similar steps with a little variation, as described below [10].

Step 1. Find a base value, B_t, using Eq. (1)

$$B_t = \frac{nP_m + (n-1)P_{m-1} + \cdots + 2P_{(m-n+2)} + P_{(m-n+1)}}{n + (n-1) + \cdots + 2 + 1} \quad (1),$$

where n is the total span period as weighting criteria, m is the total historical data in the dataset, and P_m shows the actual value at index m.

Step 2. Find the forecasting value using the B-DES procedures as follows

First, as shown in Eq.(2), we need to set the initial values for S' and S'', i.e., the single-smoothed series and the double-smoothed series, respectively.

$$S'_{t-1} = S''_{t-1} = B_t \quad (2)$$

Then, we calculate the following values for both smoothed series using the same constant smoothing factor, α $(0 \leq \alpha \leq 1)$, as shown in Eq.(3) and Eq.(4).

$$S'_t = \alpha\, Y_t + (1-\alpha)S'_{t-1} \quad (3)$$

$$S''_t = \alpha\, S'_t + (1-\alpha)S''_{t-1} \quad (4)$$

Next, we move to calculate the estimated level (L) and estimated trend (T) at time t using Eq.(5) and Eq.(6).

$$L_t = 2S'_t - S''_{t-1} \quad (5)$$

$$T_t = \frac{\alpha}{1-\alpha}\left(S'_t - S''_{t-1}\right) \quad (6)$$

Lastly, we could calculate the forecast value for Y_{t+k}, for any $k \geq 1$,as shown in Eq.(7).

$$F_{t+k} = L_t + kT_t. \quad (7)$$

Step 3. Repeat Step 1 and 2 until each data point has been gone through.

Although it was just recently introduced, B-WEMA has been widely accepted in the community. It had been tested and applied to different scenarios, such as foreign exchange transaction data prediction [10], air quality index prediction [13], and foreign tourist arrival prediction [14].

2 Results and discussion

We start this section with the description of the dataset being used in this study. Next, the forecasting results of TB incidence for the top eight countries will be given, followed by the analysis of the forecasting results.

2.1 The Dataset

In this study, we used data from WHO's Global Tuberculosis Database [15], specifically the Case Notifications CSV file with around 2Mb size in total. It contains 8,286 records and 163 attributes for more than 200 countries and territories. Eight countries form the main focus of this study: India, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa, hence the data collected are only for those countries. Moreover, since there are some missing values for several countries prior to the year 2000, we limited the data collection from 2000 up to 2018. Table 1 shows the data recapitulation for the top eight countries that experience a high burden of the tuberculosis (TB) disease.

Table 1 TB case notifications

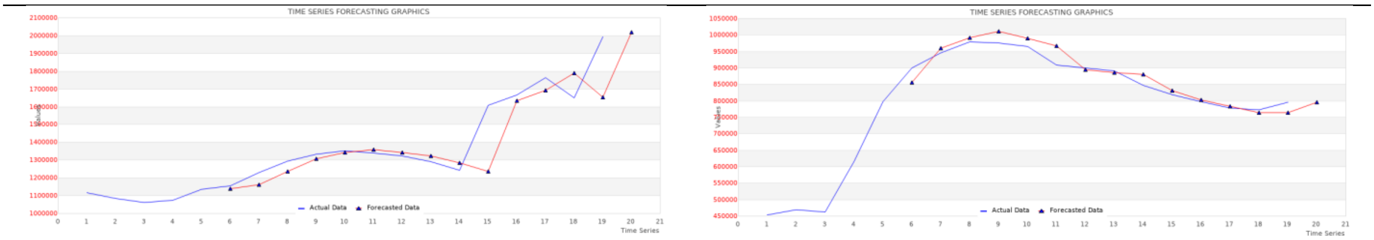
Year	**"India"**	**"China"**	**"Indonesia"**	**"Philippines"**	**"Pakistan"**	**"Nigeria"**	**"Bangladesh"**	**"South Africa"**
2000	1115718	454372	84591	119914	11050	25821	75557	151239
2001	1085075	470221	92792	107133	20707	45842	76302	148257
2002	1060951	462609	155188	118408	48344	38628	83485	215120
2003	1073282	615868	174174	132759	69632	44445	88156	227320
2004	1136182	798433	210229	130530	94327	57553	98336	267290
2005	1156248	899729	254601	137100	142017	63990	123118	270178
2006	1228827	945175	277589	147305	177376	70734	145186	303114
2007	1295943	979502	275193	140588	230468	82417	147342	315315
2008	1332267	975821	296514	139603	245635	85674	151062	348241
2009	1351913	965257	292754	146565	264248	88589	160875	360183
2010	1339866	908399	300659	166323	264235	84121	153892	354786
2011	1323949	899669	318949	195560	264934	86778	154358	362453
2012	1289836	890645	328824	216627	267475	92818	168683	323664
2013	1243905	847176	325582	229918	288910	94825	184506	312380
2014	1609547	819283	322806	243379	308417	86464	191166	306166
2015	1667136	798439	331703	276672	323856	87211	206915	287224
2016	1763876	778493	364671	332941	356390	97279	222248	237045
2017	1649694	773150	442172	317266	359224	102387	242639	243297
2018	1994000	795245	563879	371668	360472	103921	267143	227999

2.2 Prediction Results

Figure 2 shows the graphs of forecasting results for each country starting from "India, China, Indonesia, Philippines, Pakistan, Nigeria, Bangladesh, and South Africa". The actual values are shown as blue line, while the predicted values using the B-WEMA forecasting method are shown as red line. In this study, the constant smoothing factor, α, will be increased iteratively on each looping, started from 0 to 1 with two decimal places to get the smallest forecast error. The initial data being used is five, and the span data is also five. Table 2 shows the best α values for each country in this study.

Table 2 Best α for each country

Country	**Best α**
"India"	0.41
"China"	0.46
"Indonesia"	0.57
"Philippines"	0.5
"Pakistan"	0.49
"Nigeria"	0.46
"Bangladesh"	0.52
"South Africa"	0.46



(a) India *(b) China*

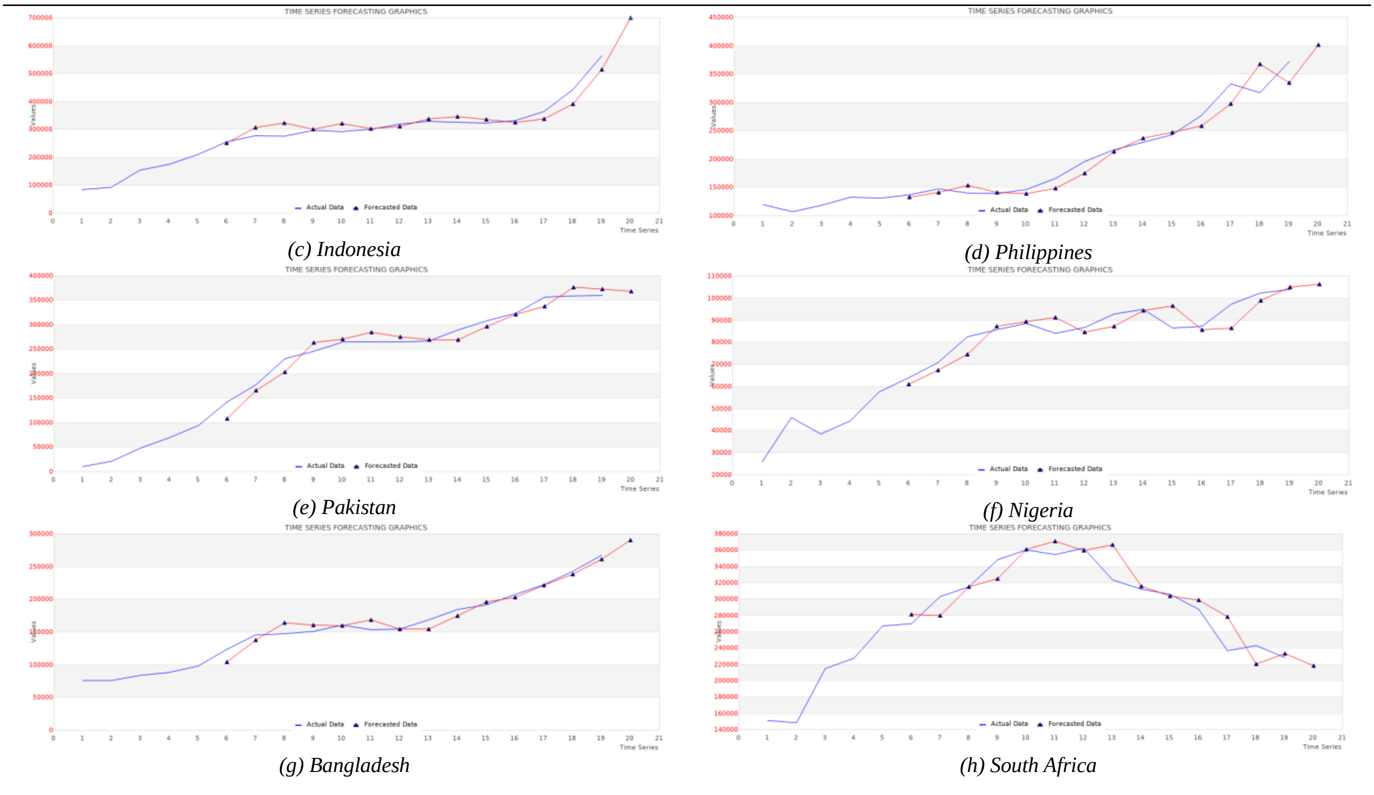


Fig.2 TB case notifications prediction

2.3 Analysis

Most of the countries studied in this research have a rising linear trend of TB case notifications since 2000, except for China and South Africa. However, China has an increased value of TB notification in 2018 and is predicted to have an increasing trend again in the future. Meanwhile, South Africa tends to have a declining trend in the future, which can be drawn from the graphic results in Figure 2.

Moreover, to calculate the accuracy level of the forecasting method applied in this study, the Mean Absolute Percentage Error (MAPE) was used. It is one of the forecast error measurement that has the property of being scale-independent. It has a formula, as shown in Eq.(8) [16].

$$MAPE = mean(|100\ e_t/Y_t|) \quad (8),$$

where Y_t is the actual value at time t, F_t is the forecasted value of Y_t, and lastly $e_t = Y_t - F_t$. Table 3 shows the MAPE values for each country incorporated in this study.

Table 3 MAPE for each country

Country	MAPE Value
"India"	5.5381966
"China"	2.3606444
"Indonesia"	6.1865528
"Philippines"	6.6085564
"Pakistan"	6.3175929
"Nigeria"	4.8628874
"Bangladesh"	5.0844156
"South Africa"	5.1957879

Table 3 demonstrates that China has the smallest MAPE value; hence the forecasting results for TB case notifications in China will be highly regarded than those of the other countries. Since China also has been predicted to have an increase in TB case notification next year, it should become the main focus of all related stakeholders, especially the Government.

3 Conclusion

Tuberculosis (TB) case notifications for the top eight high burden countries of TB disease have been successfully predicted by using the B-WEMA forecasting method. As can be seen in Table 3, B-WEMA produces small forecast error results; in this case, MAPE values under 7% for each country. Out of all eight countries, China has the smallest MAPE value of 2.36%, which means that the forecasting results should be more accurate than those of the other countries. However, there is an increasing trend in the forecasting results for China in recent years; hence it should become the main focus of all the stakeholders, especially the Government, to handle the situation. One of the possible immediate solutions is by utilizing the primary health care facilities in latent TB infection screening and treatment [17].

Acknowledgments

This study is partially supported by the research grant from Kemenristekdikti (Contract No. 1581/HD/LPPM-UMN/IV/2019). The author also would like to say` thank you for the support given by Universitas Multimedia Nusantara for this study.

References

[1] WORLD HEALTH ORGANIZATION. Global Tuberculosis Report 2019. Geneva, 2019. https://www.who.int/tb/publications/global_report/en/

[2] DEPARTMENT OF HEALTH & HUMAN SERVICES. Tuberculosis - The Facts. https://www2.health.vic.gov.au/Api/downloadmedia/%7BB9E6FB0F-D3B6-47C2-A163-6E150A5A628C%7D

[3] KURNIAWATI A., PADMAWATI R. S., and MAHENDRADHATA Y. Acceptability of mandatory tuberculosis notification among private practitioners in Yogyakarta, Indonesia. *BMC Research Notes*, 2019, 12(1): 543. https://doi.org/10.1186/s13104-019-4581-9

[4] DEPARTMENT OF HEALTH & HUMAN SERVICES. Management, Control and Prevention of Tuberculosis: Guidelines for Health Care Providers, 2016. https://www2.health.vic.gov.au/about/publications/policiesandguidelines/tuberculosis-guidelines-2015

[5] UPLEKAR M., ATRE S., WELLS W. A. et al. Mandatory tuberculosis case notification in high tuberculosis-incidence countries: policy and practice. *European Respiratory Journal*, 2016, 48: 1571–1581. https://doi.org/10.1183/13993003.00956-2016

[6] SHELDON C.D., KING K., COCK H. et al. Notification of tuberculosis: how many cases are never reported?. *Thorax*, 1992, 47(12): 1015–1018. http://dx.doi.org/ 10.1136/thx.47.12.1015

[7] HALIM S., INTAN R., and DEWI L.P. Fuzzy linear regression for tuberculosis case notification rate prediction in Surabaya. Proceedings of the International Conference on Advanced Information Science and System, New York, 2019, pp. 1–5. https://doi.org/10.1145/3373477.3373492

[8] LI Z., WANG Z., SONG H. et al. Application of a hybrid model in predicting the incidence of tuberculosis in a Chinese population. *Infection and Drug Resistance*, 2019, 12: 1011–1020. https://doi.org/10.2147/IDR.S190418

[9] HANSUN S. A New Approach of Brown's Double Exponential Smoothing Method in Time Series Analysis. *Balkan Journal of Electrical and Computer Engineering*, 2016, 4(2): 75–78. https://doi.org/10.17694/bajece.14351

[10] HANSUN S. WEMA versus B-WEMA Methods in Forex Forecasting. Proceedings of the 9th International Conference on Machine Learning and Computing, New York, 2017, pp. 268–271. https://doi.org/10.1145/3055635.3056565

[11] ABDULLAH N.H., JUNAIDI, and HANDAYANI L. Peramalan Rate of Return Saham Menggunakan Metode Brown's Weighted Exponential Moving Average dengan Optimasi Levenberg-Marquardt. *Natural Sciences: Journal of Science and Technology,* 2019, 8(3): 171–176. https://doi.org/10.22487/25411969.2019.v8.i3.14955

[12] MUKHLASHIN P.A.R. and NUGRAHA J. Brown's Weighted Exponential Moving Average (B-WEMA) with Levenberg-Marquardt Optimization to Forecasting Rate of Return. *The Turkish Online Journal of Design, Art and Communication,* 2018, 8: 1744–1749. https://doi.org/10.7456/1080SSE/232

[13] HANSUN S. and KRISTANDA M.B. AQI Measurement and Prediction using B-WEMA Method. *The International Journal of Engineering Research and Technology,* 2019, 12(10): 1621–1625. http://irphouse.com/ijert19/ijertv12n10_02.pdf

[14] HANSUN S. and KRISTANDA M.B. Forecasting Foreign Tourist Arrivals to Bali: Hybrid Double Exponential Smoothing Approach. *The International Journal of Engineering Research and Technology,* 2019, 12(11): 1864–1868. http://irphouse.com/ijert19/ijertv12n11_05.pdf

[15] WORLD HEALTH ORGANIZATION. WHO's Global Tuberculosis Database. https://www.who.int/tb/country/data/download/en/

[16] HYNDMAN R.J. and KOEHLER A.B. Another look at measures of forecast accuracy. International Journal of Forecasting., 2006, 22: 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[17] CHATTERJI M., MARKS G., LIAW S.-T., and VAN DRIEL M. Challenges in latent TB screening and treatment in primary care: A proposal to improve latent TB screening and treatment in Australia. Communicable Diseases Control Conference, Canmberra, 2019, pp. 105-108.

参考文:

[1] 世界卫生组织。2019年全球结核病报告. 2019年，日内瓦. https://www.who.int/tb/publications/global_report/en/

[2] 卫生与公共服务部。结核病-事实。https://www2.health.vic.gov.au/Api/downloadmedia/%7BB9E6FB0F-D3B6-47C2-A163-6E150A5A628C%7D

[3] KURNIAWATI A.，PADMAWATI R. S.和MAHENDRADHATA Y. 印度尼西亚日惹的私人从业者对强制性肺结核通报的接受程度。BMC研究笔记，2019，12（1）：543. https://doi.org/10.1186/s13104-019-4581-9

[4] 卫生与公共服务部。结核病的管理，控制和预防：医疗服务提供者指南，2016年。https://www2.health.vic.gov.au/about/publications/policiesandguidelines/tuberculosis-guidelines-2015

[5] UPLEKAR M.，ATRE S.，WELLS W. A.等。结核高发国家的强制性结核病例通报：政策和实践。欧洲呼吸杂志，2016，48：1571-1581。https://doi.org/10.1183/13993003.00956-2016

[6] SHELDON C.D.，KING K.，COCK H. 等。结核病的通知：从未报告过多少病例？胸部，1992，47（12）：1015-1018。http://dx.doi.org/ 10.1136 / thx.47.12.1015

[7] HALIM S.，INTAN R. 和 DEWI L.P. 在泗水的结核病

例通报率预测的模糊线性回归。国际先进信息科学和系统会议会议录，纽约，2019年， 第1-5页。https://doi.org/10.1145/3373477.3373492
[8] 李中，王中，宋红等。混合模型在预测中国人群结核病发病率中的应用。感染与耐药性，2019，12：1011-1020。https://doi.org/10.2147/IDR.S190418
[9] HANSUN S.时间序列分析中的布朗双指数平滑法的一种新方法。巴尔干电机与计算机工程学报，2016，4（2）：75-78。https://doi.org/10.17694/bajece.14351
[10] 外汇预测中的HANSUN S. WEMA与B-WEMA方法。第9届国际机器学习与计算国际会议论文集，纽约，2017年，第268–271页。https://doi.org/10.1145/3055635.3056565
[11] ABDULLAH N.H.，JUNAIDI 和HANDAYANI L. 巴拉马兰的回报率使用布朗方法的股票的加权指数移动平均线丹甘·奥帕马蒂斯·莱文贝格·马夸特。自然科学：科学技术学报，2019，8（3）：171–176。https://doi.org/10.22487/25411969.2019.v8.i3.14955
[12] MUKHLASHIN P.A.R.以及NUGRAHA J. 棕色的加权指数移动平均值（乙-WEMA）和莱文贝格-马夸特优化来预测回报率。土耳其在线设计，艺术与传播在线杂志，2018年，8：1744-1749。https://doi.org/10.7456/1080SSE/232
[13] HANSUN S. 和 KRISTANDA M.B. 使用B-WEMA方法进行空气质量指数测量和预测。国际工程研究与技术杂志，2019，12（10）：1621-1625。http://irphouse.com/ijert19/ijertv12n10_02.pdf
[14] HANSUN S. 和 KRISTANDA M.B. 预测到巴厘岛的外国游客到来：混合双指数平滑法。国际工程研究与技术杂志，2019，12（11）：1864-1868。http://irphouse.com/ijert19/ijertv12n11_05.pdf
[15] 世界卫生组织。世卫组织的全球结核病数据库。https://www.who.int/tb/country/data/download/zh/
[16] HYNDMAN R.J. 和 科勒A.B. 再看一下预测准确性的度量。国际预测杂志。，2006，22：679–688。https://doi.org/10.1016/j.ijforecast.2006.03.001
[17] CHATTERJI M.，MARKS G.，LIAW S.-T. 和 VAN DRIEL M. 初级保健中潜在结核病筛查和治疗的挑战：在澳大利亚改善潜在结核病筛查和治疗的建议。传染病控制会议，坎培拉，2019年，第105-108页。