www.icbdc.org

# ICBDC 2021

## 2021 6TH INTERNATIONAL CONFERENCE ON BIG DATA AND COMPUTING

**Shenzhen, China | May 22-24, 2021**

**2021 6th International Conference on Big Data and Computing will be held during May 22-24, 2021 in Shenzhen, China. It will provide a forum for researchers from both academia and industry to exchange the latest innovations and research advancements in big data, soft computing, cloud computing, networking and computer security. The aim of ICBDC 2021 is to present the latest research and results of scientists related to Big Data and Computing topics.**

## PUBLICATION

Accepted papers will be included into ICBDC 2021 conference proceedings. The Proceedings will be published in the **International Conference Proceedings Series** by **ACM** (**ISBN: 978-1-4503-8980-8**), which will be archived in the **ACM Digital Library**, and indexed by **Ei Compendex** and **Scopus** and submitted to be reviewed by Thomson Reuters Conference Proceedings Citation Index (ISI Web of Science).

## CONFERENCE PROCEEDINGS

- ICBDC 2020 Proceedings has been archived into ACM Digital Library and indexed by EI Compendex http://dl.acm.org/citation.cfm?id=3404687
- ICBDC 2019 Proceedings has been archived into ACM Digital Library and indexed by EI Compendex and SCOPUS https://dl.acm.org/citation.cfm?id=3335484
- ICBDC 2018 Proceedings has been archived into ACM Digital Library and indexed by EI Compendex and SCOPOS four months after the conference: https://dl.acm.org/citation.cfm?id=3220199

## CALL FOR PAPERS

**Topics of interest include, but are not limited to:**

- Algorithmic, experimental, prototyping and implementation
- Big Data Algorithms, Applications and Services
- Government and Industrial Experiences for Cloud and Big Data
- Computing, scheduling and resource management for sustainability
- Data-driven innovation, computational modelling and data integration
- Green Computing and Networking Technologies for Cloud and Big Data
- Software, hardware and algorithm co-design, high-performance computing
- Cloud Computing Solutions and Platforms
- Big Data Processing and Querying
- Big Data Visualization
- Big Data Education
- Data intensive computing theorems

**For details about topics, please visit at http://www.icbdc.org/cfp.html**

## SUBMISSION

Online submission system: **http://www.easychair.org/conferences/?conf=icbdc2021**
More detail about submission, please visit at **http://www.icbdc.org/author.html#sub**

官方微信号
一对一咨询服务

## IMPORTANT DATES

Submission Deadline: January 15, 2021 ♦ Notification Date: February 4, 2021
♦Registration Deadline: February 25, 2021

## Contact US

Conference Secretary: Ms. Echo Yang
E-mail: icbdc@iact.net
Tel: +86-18081079313

Please note that if you do nothing (not even click on the menu) for more than two hours, your session will expire and you will have to log in again.

Some of your submissions were either withdrawn by authors or deleted by chairs, they are shown using a grey background.

| # | Authors | Title | View | paper | Program |
|---|---|---|---|---|---|
| 4 | Meili Liu, Liwei Wang, Chun-Te Lee and Jeng-Eng Lin | Data analysis on the precipitation and temperature of city of Wenzhou | | | |
| 5 | Meili Liu, Yichao Dai, Chun-Te Lee and Jeng-Eng Lin | Regression Analysis on Classification and Quality of Wine | | | |
| 6 | Meili Liu, Kewei Zhang, Chun-Te Lee and Jeng-Eng Lin | A Comparative Study of SARIMA and Holt-Winters Models in Time Series Forecasting | | | |
| 7 | Meili Liu, Bingxu Han, Chun-Te Lee and Jeng-Eng Lin | Data analysis on water resources and characteristics of East China city | | | |
| 8 | Chun-Te Lee, Jie Shen, Meili Liu and Jeng-Eng Lin | On the study of meteorological data with applications to time series analysis | | | |
| 10 | Meili Liu, Jianxiong Zhao, Chun-Te Lee and Jeng-Eng Lin | Data analysis on the housing prices in Taipei using Machine Learning method | | | |
| 11 | Meili Liu, Tang Jinghuan, Chun-Te Lee and Jeng-Eng Lin | Visualization and Analysis of the Air Quality Data in Certain Provinces of China | | | |
| 12 | Chun-Te Lee, Tao Zheng, Meili Liu and Jeng-Eng Lin | Data Analysis on Relationship between Expenditure of Local Government and Total Output in China | | | |

# A Comparative Study of SARIMA and Holt-Winters Models in Time Series Forecasting

Meili Liu

*Institute of Artificial Intelligence and Research Center, Midea Group, Shenzhen, China*

Kewei Zhang

*Department of Mathematics, Wenzhou-Kean University, Wenzhou, China*

Chun-Te Lee*

*School of Mathematical Sciences College of Science and Technology Wenzhou-Kean University, Wenzhou, China*

Jeng-Eng Lin

*Department of Mathematical Sciences George Mason University, USA.*

## Abstract

This paper is aimed to provide periodic change of meteorology of temperature and precipitation in Wenzhou area, East of China. We propose two models, namely, SARIMA and Holt-Winter to predict the future temperature and amount of rainfall. The data set of weather conditions that we used are downloaded from URL:https://geographic.org/global_weather/china/wenzhou_659.html, which involves two characteristics precipitation and temperature. Time series models are applied on dataset recorded from 1955 to 1997, and the forecasting results are compared and models are evaluated. The forecasting results show that the Holt-Winter model was the best model to make prediction on meteorological data. In comparison with with SARIMA model, the performance of Holt-Winter model has lower estimation error for weather featuring variables. In conclusion, Holt-Winter model is effective

*Corresponding author

*Email addresses:* 12mlliu@gmail.com (Meili Liu), zhangke@kean.edu (Kewei Zhang), scat1440@yahoo.com (Chun-Te Lee), jelin@gmu.edu (Jeng-Eng Lin)

and produces better forecasting results on climate data, which only provides effective algorithms for predicting the weather, but also helps prevent climate change disasters caused by the greenhouse effect.

*Keywords:* Time Series, forcasting, exponential smoothing, Holt-Winters, ARIMA

## 1. 1.Introduction

Time series is widely known a a type of data that is defined as a set of values observed sequentially through time, and forecasting on time series has become one of the important issues in machine learning because of the complexity and components involved in prediction problems. Time series forecasting involves taking models fit on historical data and using them to predict future observations. The skill of a time series forecasting model is determined by its performance at predicting the future. The primary objective of this article is to provide effective time series models and predict future values based on source data. To obtain a better prediction result, it is essential to provide a reliable and effective data analysis forecasting algorithm that allows to develop forecasting models on source data. In this report the seasonal autoregressive integrated moving average (SARIMA) and Holt-Winter model will be implemented to obtain the best prediction on time series data. In section 2 the basic mathematics of the models are introduced and described. In section 3 the evaluation methods used for comparing the performance of each forecasting methods are introduced, and the forecasting procedure and experimental results for test data of each model are discussed. Finally, conclude the time series analysis on meteorological data in section 4.

## 2. 2. Methodology

The process of our methodological on time series forecasting model is shown in Fig. 1

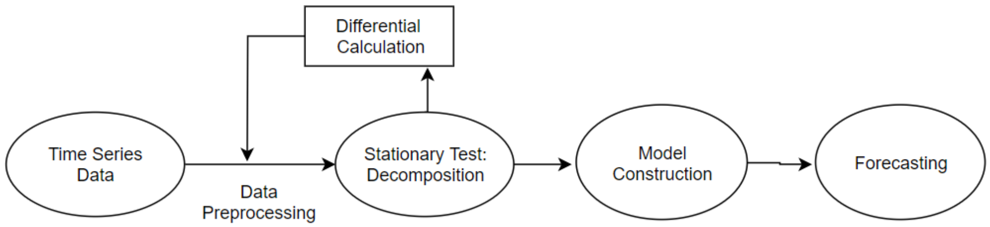


Figure 1: Flow Chart

Time series has wide applications in prediction and forecasting the data against time. There are several different time series forecasting methods, e.g., the moving average, exponential smoothing and ARIMA, which are linear statistical model and able to provide relatively satisfactory prediction results. To better under stand these essential traditional statistical forecasting model, the seasonal ARIMA (SARIMA) and triple exponential smoothing also known as Holt-Winters model are proposed in this article for predicting data.

### *2.1. 2.1 ARIMA and SARIMA*

Auto regressive integrated moving average(ARIMA) is one of the most commonly used linear method in time series forecasting, which used three integers p, d and q to parametrize the ARIMA model. the parameters p, q, d account for components of time series, namely seasonality, trend, and noise. Here p is the auto-regressive parameter which provides the number of autoregressive terms, it is defined by the partial autocorrelations of the appropriately differenced series. The parameter d is the integrated parameter which indicates number of nonseasonal differences needed for stationarity, and q is moving average parameter which defines the number of lagged forecast errors in the prediction equation. The parameter is determined from the autocorrelations of the appropriately differenced series

Notice that in the model identification step, the stationarity and seasonality of the series need to be determined because stationarity is one necessary condition for building one ARIMA model. Differencing and power transformation are common used method to stabilize the data series. To capture the salient dynamic features as parameters of the model, Box and Jenkins

In general, the three-step iterative cycles will be repeated several times to find the best fitted model. Furthermore, for seasonal data series, Box and Jenkins recommend adding three more parameters P, D and Q which are defined as seasonal parameters corresponding to p, d and q. The complete SARIMA model is defined as "ARIMA(p,d,q)x(P,D,Q). Indicated by Box and Jenkins

Notice that the SARIMA model generally involves the following four-step iterative cycles:

(a) Identification of SARIMA model structure

(b) Estimation of unknown parameters

(c) Goodness-of-fit tests on the estimated residuals

(d) Forecasting

Notice that p, d, q, P, D, Q are model parameters, s is the periodicity,B is the backward shift operator, $\phi(B), (B), \theta(B), \Theta(B)$ are polynomials in B respectively of degree p, d and q, d is the level of difference, D is the level of seasonal difference, $Z_t$ is the actual value at time period t, $a_t$ is the estimated residual error at time period t.

### *2.2. 2.2 Holt-Winters Model*

Exponential smoothing is another forecasting method for time series. Similar to Box-Jenkins ARIMA method whose prediction is one weighted linear sum of past observations of lags

$$
\begin{aligned}
&L_t = \alpha(y_t - St - s) + (1 - \alpha)(Lt - 1 + b_{t-1}) \\
&b_t = \beta(L_t - Lt - 1) + (1 - \beta)bt - 1 \\
&S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s} \\
&F_{t+k} = L_t + kb_t + St + k - s,
\end{aligned}
\tag{1}
$$

where $L_t$ is the level equation, defined the weighted average between the seasonally adjusted observation and the non-seasonal forecasting at time period t. The parameter $b_t$ is the trend equation which present the weighted average value based on difference between $L_t$ and $L_{t-1}$ and $b_{t-1}$. The variable $S_t$ is the seasonal function which presents the weighted average between seasonal in two years. $\alpha$, $\beta$ and $\gamma$ are smoothing parameters which all ranged in 0 and 1, and s is the length of seasonal period

## 3. 3. Experimental Result

We perform the numerical methods on the data set containing precipitation and temperature of Wenzhou from 1951 to 1997. The source data downloaded from URL:https://geographic.org/global_weather/china/wenzhou_659.html will be analyzed as the source data. The data set contains four variables, namely precipitation, max temperature, min temperature and average temperature. Here the Precipitation and temperature variables are essential for forecasting the time series data, which are very practical approach in the analysis. Our modeling approach will start from a simple univariate time series analysis to forecast the precipitation with different functionality in python, then the complex multivariate time series with different temperature variables are introduced and demonstrated. The characteristics of data sets will be demonstrated presented and estimation results obtained from each model will be compared.

### *3.1. 3.1 Time Series analysis on Precipitation*

Precipitation is one of the major parts of global water cycle, analysis on precipitation can contribute to the understanding of the weather in specific area and it is one essential factor that has a dramatic effect on agriculture. The precipitation time series from January 1951 to December 1997 showed seasonality trend is shown in Fig. 2. The peaks generally fall on summer, drops usually occurs in winter. After decomposition on the data, the seasonality trend is more clarity which is shown in Fig. 3. The data from 1951 January to 1987 December will be used as the training data to forecast the data from 1988 January to 1997 December. It is assumed that the seasonality and trend pattern existed in the past will still extent to the future.

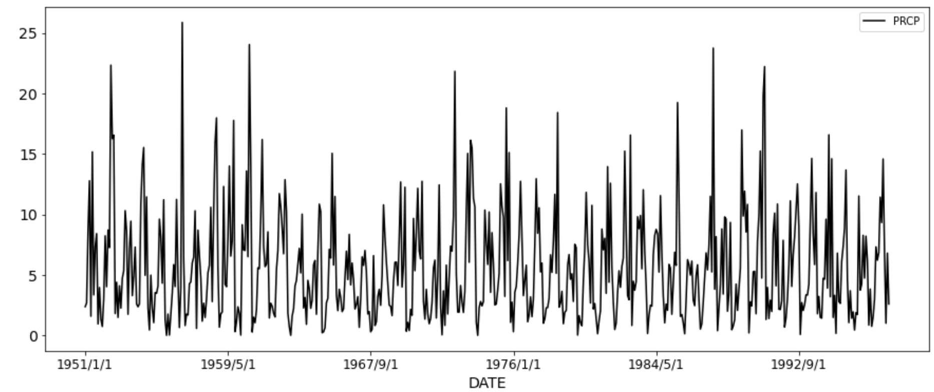


Figure 2: Precipitation data

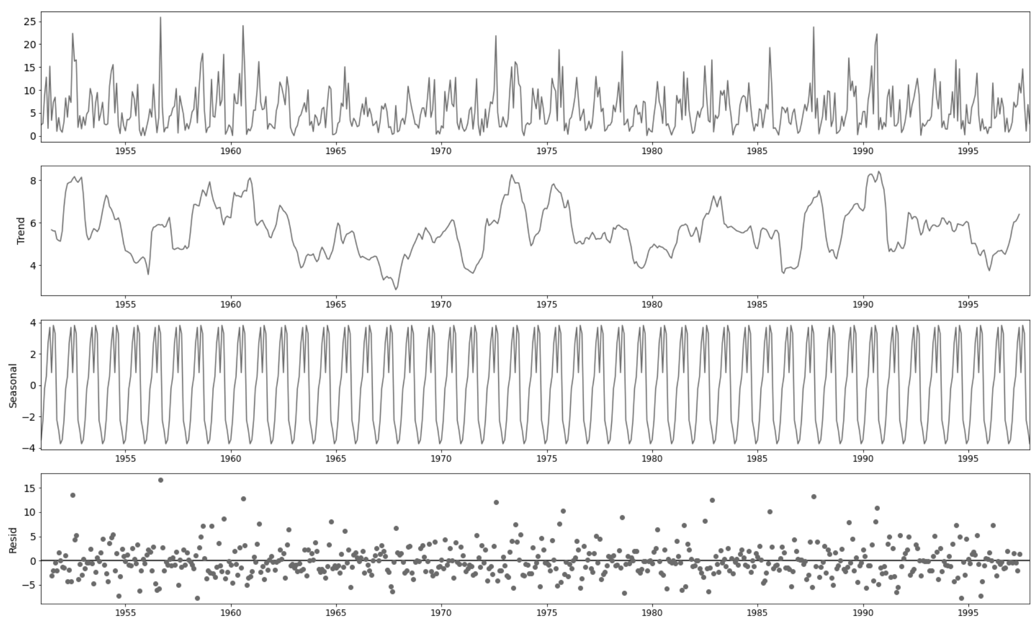


Figure 3: Precipitation data after Decomposition

### *3.1.1. 3.1.1 SARIMA Model on Precipitation Data*

We resort to the use of auto arima function to formulate the SARIMA model in python. The function automatically computes the best model through comparing the Akaike Information Criterion (AIC). The model generated from training data set is $ARIMA(5,1,0)(5,1,0)_12$. The visualization of the forecasting results and observed results from 1988 January to 1997 December are shown in Fig 4.

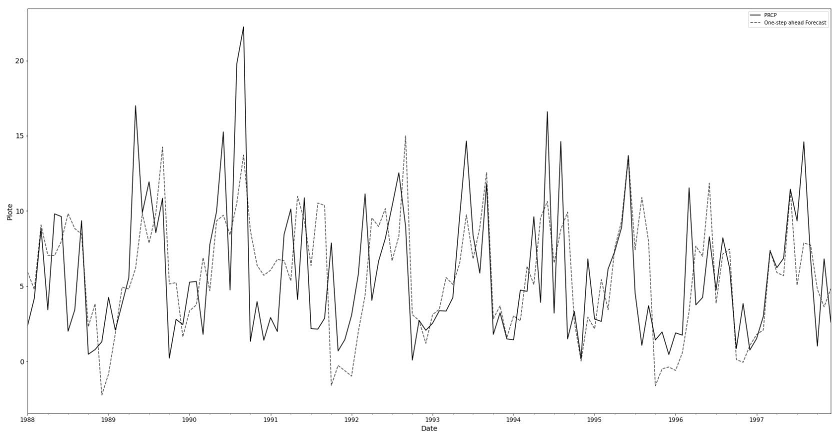


Figure 4: Forecasting Result of SARIMA Model

*3.1.2. 3.1.2 Holt-Winters Model on Precipitation Data*

In this subsection, the additive Holt-Winters method is utilized because it has been observed from Fig. 2 that the seasonal variations are roughly constant. To ensure the best model is obtained, Both the additive and multiplicative Holt-Winters methods will be applied, the best estimation will be selected based based on their performance. In python code, the library will automatically optimizes the smoothing parameters. Seasonal period is defined as 12. The visualization of the forecasting results from 1988 January to 1997 December is shown in Fig. 5

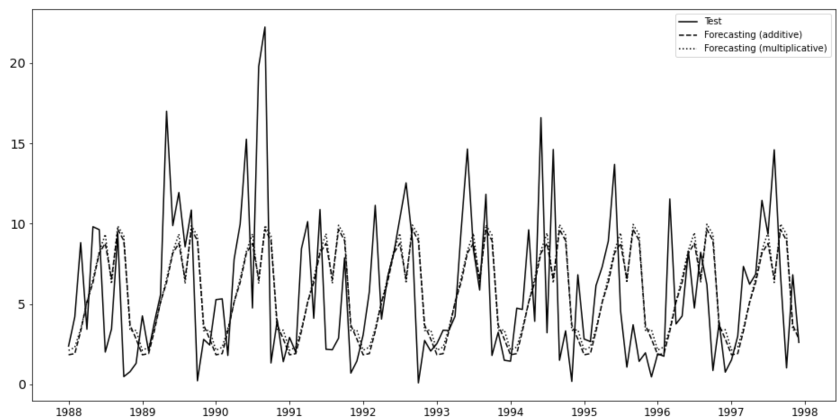


Figure 5: Forecasting Result of Holt-Winters Model

### *3.2. 3.2 Temperature Time Series*

Temperature is one essential physical quantity which directly reflects the climate and atmosphere condition of one location. Temperature forecasting can help people predict the nature disaster in advance, and it is also very important for agriculture development. The temperature time series from January 1951 to December 1997 showed strong seasonality trend, shown in Fig. 6. The decomposition plot of trend and seasonality is shown in Fig. 7. The data from 1951 January to 1987 December will be used as the training data to forecast the data from 1988 January to 1997 December. It is believed that the seasonality and trend patterns are existed and the phenomenon will still be extended to the future.

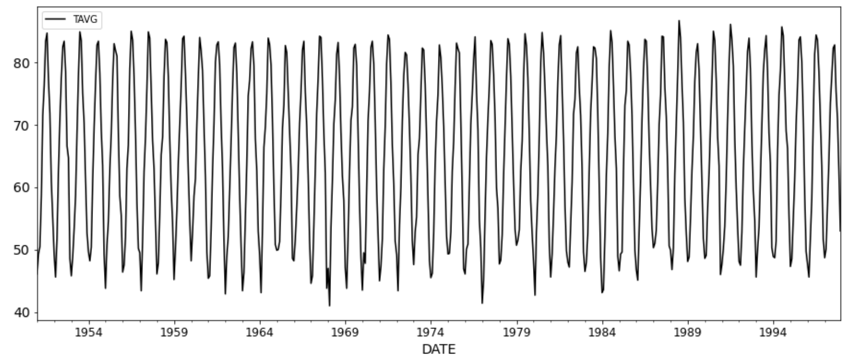


Figure 6: Temperature Data

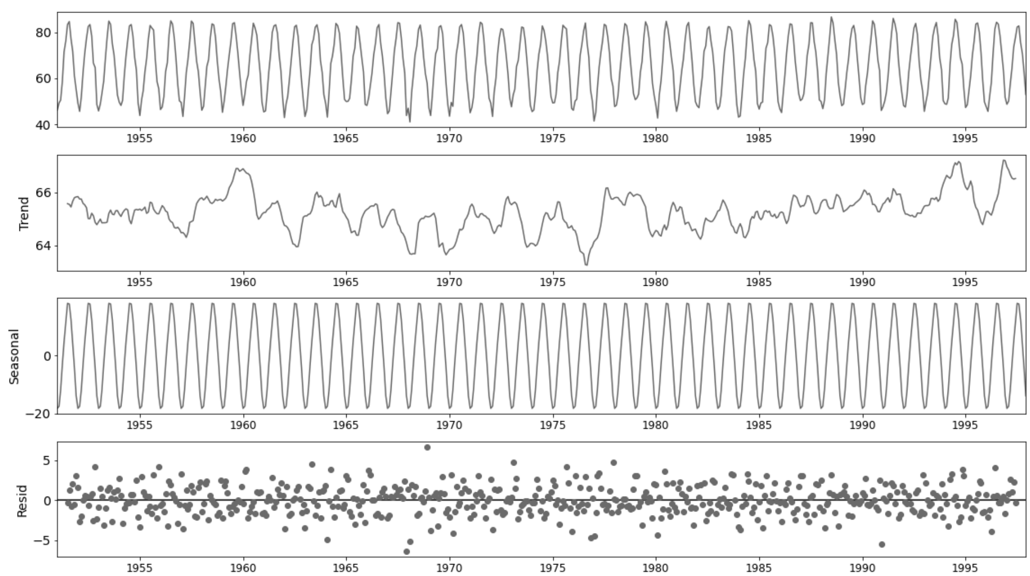


Figure 7: Temperature Data After Decomposition

### *3.2.1. 3.2.1 SARIMA Model on Temperature Data*

The auto arima function is used to apply the SARIMA model in python. Automatically computed best model generated from training data set is $ARIMA(5,1,0)(5,1,0)_12$. The visualization of the forecasting results and observation results from 1988 January to 1997 December are shown in Fig 8.

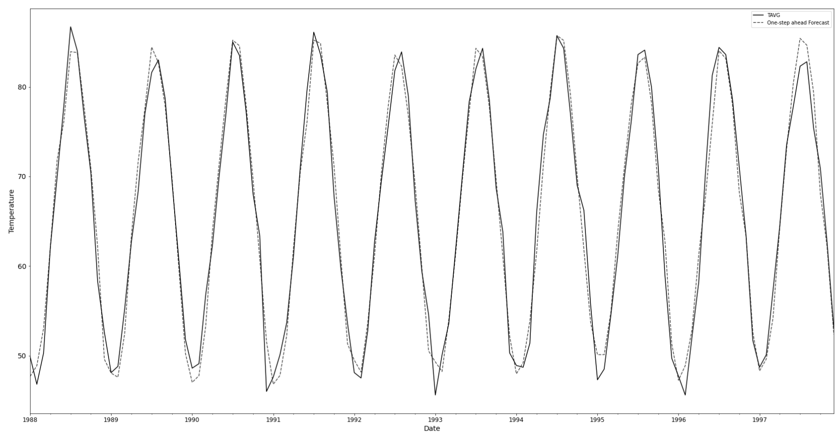


Figure 8: Forecasting Result of SARIMA Model

*3.2.2. 3.2.1 Holt-Winters Model on Temperature Data*

The Holt-Winters additive and multiplicative methods will be applied to guarantee the best fitted model is found. The best estimation will be selected based on their performance. In python the library will automatically optimizes the smoothing parameters. Seasonal period is defined as 12. The visualization of the forecasting results from 1988 January to 1997 December is shown in Fig. 9.

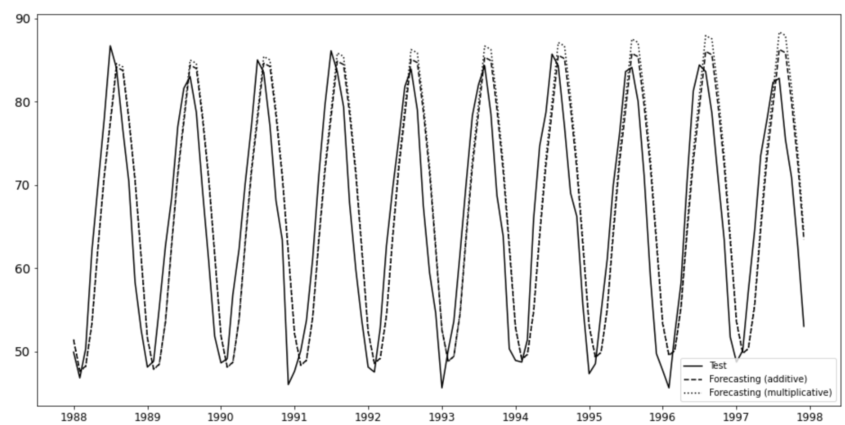


Figure 9: Forecasting Result of Holt-Winters Model

## 4. 4. Evaluation of Models and Comparison of Performance

The performances between each method will be compared in this section. The measurements methods of accuracy involve MSE, RMSE and MAE.The mathematical expressions of each methods are:

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (e_t)^2 \tag{2}$$

$$RMSE = sqrt \frac{1}{n} \sum_{t=1}^{n} (e_t)^2 \tag{3}$$

$$MAE = \frac{1}{n} sum_{t=1}^{n} |e_t| \tag{4}$$

Table 1: Errors(training set) of the precipitation time series

| METHOD | SARIMA | Holt-Winters(add) | Holt-Winters(mul) |
|---|---|---|---|
| MSE | 30.27 | 19.0876 | 19.6607 |
| RMSE | 5.5 | 4.3689 | 4.4340 |
| MAE | 4.6776 | 3.2520 | 3.2887 |

Based on the comparison of forecasting error of the precipitation time series data in Table 1, the results indicate that the additive Holt-Winters model is superior to other methods. The MSE of Holt-Winters additive method is 19.0876, which is lower than multiplicative Holt-Winters of 19.6607 and 30.27 in SARIMA model. In addition, it has the lowest MAE as 3.2520, comparing with the SARIMA which has 4.6776 and 3.2887 of the multiplicative Holt-Winters model. It is noted that the Holt-Winters models are both superior to the SARIMA model, which means Holt-Winters model remained in a stable performance on the precipitation time series data.

Table 2: Errors(training set) of the temperature time series

| METHOD | SARIMA | Holt-Winters(add) | Holt-Winters(mul) |
|---|---|---|---|
| MSE | 3.45 | 50.3374 | 54.4503 |
| RMSE | 1.86 | 7.0949 | 7.3790 |
| MAE | 1.4656 | 6.2317 | 6.4451 |

Table 2 indicates the SARIMA showed outstanding performance than two Holt-Winters models for the temperature data set.

## 5. 4 Conclusion

In this article, two essential forecasting methods of ARIMA and exponential smoothing model have been implemented. Both of the models have been applied to the prediction of the precipitation and temperature in time series data in Wenzhou area, East-China, from 1955 to 1997. The evaluation method including MAE, RMSE and MSE have been calculated and results are fairly compared. The results showed that both methods are possible to obtain the best forecasting model. For time series data in precipitation the Holt-Winters models have better performance, in which the ARIAM model shows much lower error on the temperature than the Holt-Winters model.

## References

[1] G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970

[2] G.Peter Zhang,Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, Volume 50 (2003) 159-175

[3] Fang-Mei Tseng, Hsiao-Cheng Yu, Gwo-Hsiung Tzeng, Combining neural network model with seasonal time series ARIMA model, Technological Forecasting and Social Change Volume 69 (1) (2002) 71-87

[4] Brownlee Jason, How to Grid Search Triple Exponential Smoothing for Time Series Forecasting in Python, Machine Learning Mastery, 2018

[5] George Aryee, Raymond Essuman, Robert Djagbletey, Ebenezer Owusu Darkwa, Comparing the Forecasting Performance of Seasonal Arima and Holt -Winters Methods of Births at a Tertiary Healthcare Facility in Ghana, JBiostat Epidemiol. 5(1) (2019) 18-27