

An Epitomized Approach to Possess Promising Predictions by using Time-Series Analysis and Forecasting in R language

^{*1}Ramachandra Rao Kurada, ²Dr. Karteeka Pavan Kanadam

¹ Asst. Prof., Department of CSE, Shri Vishnu Engineering College for Women, Bhimavaram

² Professor, Department of Computer Applications, R.V.R. & J.C. College of Engineering, Guntur

Email: ¹ ramachandraraokurada@gmail.com, ² karteeka@yahoo.com

Received: 11th February 2018, Accepted 17th March 18, Published: 30th April 2018

Abstract

The aim of this work is to exertion a plug-in, formerly named as Time Series Analysis and Forecasting (TSAF) and incorporates this plug-in into R language. The intent behind materializing this plug-in is to establish a first-rated approach to forecast in-advance extrapolations in time series data and to make accurate decisions methodically. The plug-in provides a computationally intelligent environment by accepting a preprocessed time series datasets as input and sense the direction of outputs that will transpire over the coming ages. The internal code structure and implementation details in between the input and output precincts are factorized with the general machine learning, statistical calculation, and visualization packages. The preeminence of this incarnated viewpoint is scientifically verified over time-series datasets archived in UCI repositories. The results enabled from these datasets pertain to revive qualitative nature of forecasting, which helps the users to predict or foresee changing domain trends and thereby make strategic decisions and hopefully gain lifelong encroachments in this process.

Keywords: Time series analysis, Forecasting's, Predictions, Machine Learning, Statistical study, Computational Intelligence.

1. Introduction:

A time series data is a sequence of numerical data points in successive order, measured at regular intervals, with no minimum or maximum amount of time [1]. Parenthetically, such data is assorted into one of the three sets as a) Time series: Holds a set of annotations on the data for an attribute to changes over time, b) Cross-sectional: Data of more than one or more characteristic is composed at a similar point in time and c) Pooled: Hybrid data with the combination of both time series and cross-sectional data [2].

A dependability study to comprehend the factors of trend, seasonality, cyclic effect and noise in time series data inclining statistical technique is called time series analysis. In general, Trend is a pattern in time series data, which is linear or non-linear, indicates the change in data

either ascending or descending. Seasonal is a significant pattern in time series data, always applicable to known period and is influenced by an annual factor of data and time. Seasonality is effected systematically in a fixed interval period. Cyclic pattern effects the time series data in a changing interval time as fluctuations concerning hours, day, weeks, months, quarters and years [3].

In this technology accelerated age, dialing endorsements and piloting trends in time series data is apparently resisted for decades without a demystifying procedure. Hence, the work confronts a holistic approach to provide a truly global experience by meritoriously presenting R language packages for machine learning, statistical calculation, and visualization in a customary plug-in called Time Series Analysis and Forecasting (TSAF). According to the prefatory unveiled in this work, the TSAF plug-in is configured with R language and configuration, scanning, parsing, formatting, and integration of code are expended with intelligent computing. The anticipated approach acts as a centerpiece to find the best time of time series data by pursuing a credible time series analysis and forecasting on critical applications such as business planning, economics, sociology, meteorology, medicine, and environment, etc.

2. Approach to conduct Time Series Analysis and Forecasting:

The approach presented in this section goalposts the undermined objective stated in the novel theme of the paper. This leitmotif precious inclines the statistical procedures with computational intelligence techniques and comprehends the connectivity among them. This engagement demeanor predictions, forecasting and time series analysis to find patterns in time series data. The workflow sequencing towards conducting the time series analysis and forecasting is shown in Figure 1.

Holistically Figure 1 embarks the complete procedure with staging to gain familiarity and represent knowledge to make cognizant verdicts. The time series data is accepted as an input in the form of flat files or from databases. This data is thoroughly preprocessed by finding and fixing null or missing values. Data

transformation, normalization, outlier detection and elimination is made to refine the time series and concoct time series analysis and forecasting [4].

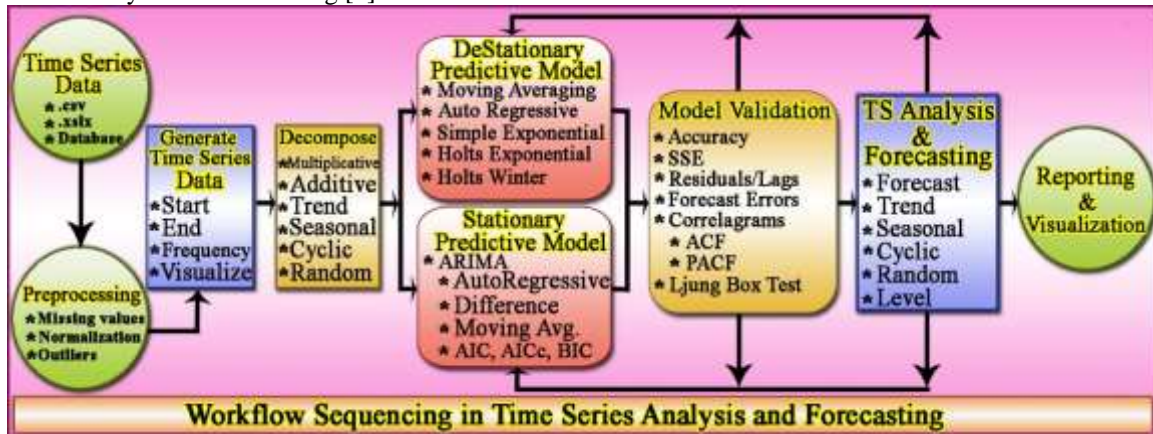


Figure 1. Workflow sequencing in Time Series Analysis and Forecasting

Such, data is intertwined basically by setting time series start and end time periodically, with a relative frequency to distinguish time interval such as an hour, day, week, month, quarter or yearly. Time series model has erected either additive or multiplicative sequence, and thereby actual data is decomposed into time series patterns such as trend, seasonal, cyclic and random [5, 30]. This putrefied is used as an input to the predictive model. The predictive models are fictional either by equally weighted observations via Moving Averaging (MV) [17] or scenery with a different set of weights via Exponential Smoothing [18].

In exponential smoothing methods, weights decay exponentially from the most recent to the most distant data points. Simple exponential smoothing [20, 21], Holt exponential smoothing [18] and Holts Winter exponential smoothing methods [22, 23] are tagged into this category [6, 12]. These methods used α , β , and γ to represent smoothing constant, trend and seasonality respectively [24]. The autoregressive integrated moving average (ARIMA) [25, 26] model is used to make time series stationary by computing the difference two successive time intervals, thereby safeguards to see both trend and seasonal composed are eliminated [7]. Once the models are constructed, the accuracy and SSE of the model are verified and validated, by forecasting the errors in the form of residuals or lags.

One type of validation is by visualizing the lags inform Autocorrelation Function (ACF) [27] and Partial autocorrelation function (PACF) [27] correlograms [10, 11]. ACF and PACF measures adhere association between current and past series values and indicate which past series values are most useful in predicting

future values. With this knowledge, you can determine the order of processes in an ARIMA model [31]. The respective lag values which exceed the significant boundaries are considered as p and q , i.e., AR and MA values to fit the order in a predictive model. The other type of validity measure is to impose Ljung box test [28, 29] to check for the existence of null hypothesis [8, 9]. In this way, the stationary or destationary models efficiency is reiterated and refitted until an optimal model is fixed. Once the final version of time series model is ratified, forecasting is instigated to extract time series patterns and predict next trends [30].

3. Methodology:

- Input: Accept the time series data as input.
 - Output: Discover the knowledge and patterns by conducting estimates, time series analysis, and forecasting's.
1. Clean the time series data by eliminating missing values, removing the outliers and normalizing the data.
 2. Generate the time series data by setting starting and ending interval of time along with frequency.
 3. Choose basic decompose models additive or multiplicative to find time series patterns such as seasonal, cyclic, trend, random.
 4. Verify the acquired patterns as stationarity or trend-stationarity and forecast the future time series. Data patterns using AR, MV or Holts Exponential or Holts Winter model or ARIMA models.
 5. Make predictive validation by estimating the accuracy and sum of squared error of the model.
 - a. Calculate the correlogram of forecast errors with ACF and PACF.

- b. Find optimal parameters to make the time series stationary using ACF and PACF.
- c. Check for significant evidence of non-zero correlations at lags using Ljung-Box test.
- d. Verify whether the forecast errors are normally distributed with mean zero and constant variance.
- e. Build the suitable ARIMA model with lowest BIC and AIC to make forecasts on the future time points.
6. If forecast errors still exist, refit the model by reiterating Step 4 until all the lag values are within significant boundaries of ACF and PACF.
7. Finally, build an optimal predictive model with low-slung of forecasting residuals and start forecasting time series patterns.

This structured approach is compiled as a TSAF plug-in, to acts as a scientific model to forecast the future process in time series data. The proposed method is used to accomplish time series analysis to conceptually diverse the entwined internal structure of time series into trend, seasonality, cyclic and random. As forecasting is crucial in time series data, the adopted ARIMA models in this plug-in assure to obtain high forecasting accuracy with low computational time.

4. Results and Discussion:

This section verifies the proficient methodology stated in the previous section by giving greater accountability to obtain trend, seasonality, cyclic, random patterns in time series data, and thereby experience a better way of forecasting in an efficient manner. To appraise the consummate work, R programming language is used to emanate time series analysis and forecasting on actual time series data archived in UCI datasets. Also, to apprehend the environment towards building the TSAF plug-in, Windows 8.0 operating system with Intel Core i3 processor, 4GB RAM and R Studio 1.0 is used.

a. Dataset 1 (Currency). To conduct time series analysis and forecasting, the currency dataset is extracted from UCI repositories [13]. This data set archives the changes in currency value of 54 different countries to the value of US dollar. Ranging from the year 1950 to 2016, a total of 3084 samples with 3 attributes each are stored in this dataset. Towards beckoning versatility and transparency in time series data, subsampling technique is chosen in this work by picking 67 samples and confining the discussion towards Indian currency and imprint results in this way. The objective is set to predict the value of Indian rupee to US Dollar for the next four years.

By applying the technique of subsampling 67 samples with 3 attributes each, are chosen as input file in .csv file format to forecast the time series data patterns. Preprocessing of data is performed by applying data cleaning primitive functions such as outlier detection and elimination, replacing the null and missing values using `tsclean()` command. Data frame is recognized with minimum value and a maximum value ranging from 4.76 to 76.195 from the year 1950 to 2016. The data set is now documented in a sequential data across an equally spaced unit of time and decomposed into time series patterns as a trend, seasonal, cyclic and random. The underlying direction in the rate of change of Indian Rupee as a long-term movement in the form of time series as a trend, slope, and level is shown in Figure 2.

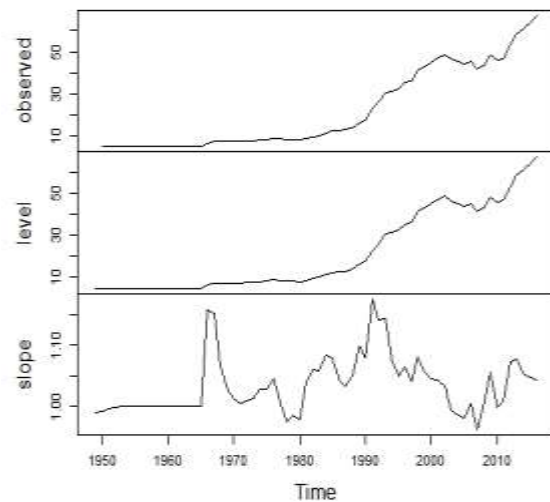
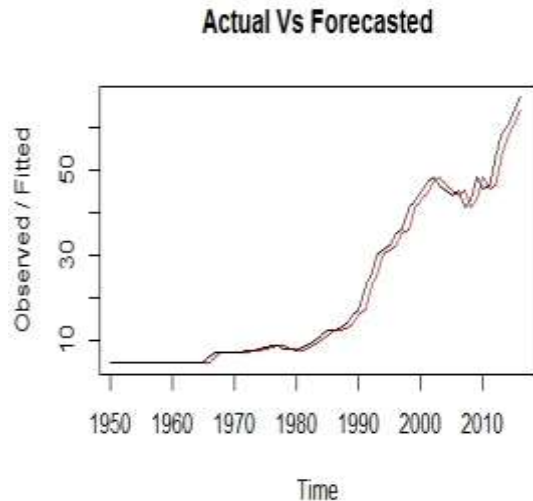


Figure 2. Decomposition of currency dataset in level, slope, and random time intervals

To verify the acquired patterns, Holt Winter exponential smoothing technique is applied on the decomposed time series data by setting the smoothing parameters alpha, beta, and gamma to determining the weight of recent data on the level, trend, and seasonality, respectively. These constant values always vary between 0 and 1. As the constant value gets closer to one, more and more weight is given to recent observations. Either additive or multiplicative do Holt Winter exponential smoothing adjustment of seasonality. In additive series, the amount of change of time series values is always constant for all the levels and for multiplicative the amount of correction of time series values varies from level to level. Figure 3 depicts the output of Holt Winter exponential smoothing algorithm in the form of establishing the relationship between actual and predicted. The value of smoothing parameter alpha is rendered as 0.9999, which is almost upper threshold value as 1. This indicates that the forecast is based on the most recent observations.

In the plot shown in Figure 3, the values of the actual trend are shown in black color, and forecasted trend in the dataset is shown in red color. It is observed from Figure 3 there exists a close approximation between actual and forecasted values of the time series data since both the actual and forecasted are moving in the same

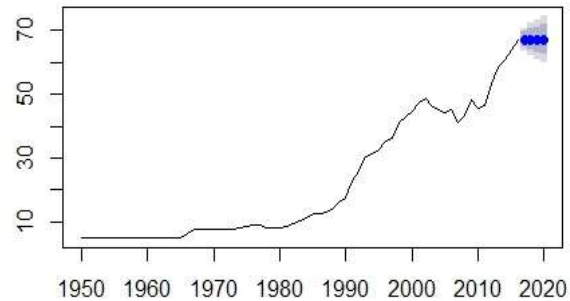


direction with the same momentum. To gain this swiftly potent in a more statistically way additive smoothing adjustment is highlighted in this work. The sum of squared error (SSE) is computed as 284.0939, and this signals a relatively low percentage of error shown in the model. After applying the smoothing function on the decomposed time series data, forecasting is scheduled with intangible confidence on the constructed predictive model.

Figure 3. Values of actual vs. forecasted time intervals in currency dataset

Table 1 shows the forecasted value of Indian rupee to US dollar with 80% and 90% prediction values. The same was exhibited as a plot with prediction intervals in Figure 4 shaded with two different colors. The accuracy of the forecasted model and training error measures are shown in Table 2. These errors are denoted as a mean absolute scaled error (MASE), mean error (ME), mean absolute percentage error (MAPE), root mean squared error (RMSE), mean percentage error (MPE) and mean absolute error (MAE). It is observed from Table 2 that maximum an error of 5% deviation happened from the actual to the forecasted and interestingness from this table is all the values are close to zero and signaling high accurateness in predicting how closely the forecasted components track the actual data.

Table 1. Forecasted value of Indian Rupee to US Dollar
Forecast with 80% and 90% prediction interval



for the year's 2017 to 2020 using Holt winter forecasting model.

Year	80% prediction interval		90% prediction interval	
	Low	High	Low	High
2017	64.81	69.57	63.54	70.84
2018	63.82	70.56	62.03	72.35
2019	63.06	71.32	60.87	73.51
2020	62.42	71.96	59.91	74.48

Figure 4. Forecasted value of Indian Rupee to US Dollar for the years 2017 to 2020 about 80% and 90% prediction intervals using Holt winter forecasted model

To verify and validate the prediction model the forecast errors are estimated as residuals. Such confirmations help the model to improve upon correlations between forecast errors for successive predictions. To happen this into reality a correlogram with 20 residuals as lags are used in ACF function. The function `acf()` is used to identify the possible autocorrelated structure of time series data. After applying this `acf()` function, the results are interpreted at the lag values. Lag 1, 2 and 4 exceeds the significant boundaries, which indicates these delays are correlated with current series, and rest of the lags lie within the important region representing the correlation of the values of time series with itself and are finally used to determine the value of q for MA model.

Table 2. Values of training errors produced when applied over Holt winter forecasting model on currency dataset

Error	Value
ME	0.9460307
RMSE	2.074718
MAE	1.317349
MPE	4.341275
MAPE	5.454227
MASE	0.0786751

This strategy was expressed in Figure 5. Consequently, `pacf()` function is also induced with 20 lags, by removing the linear dependence between the lags and establishing a correlation to itself. The same is interpreted from the plot depicted in Figure 6. An important observation was only the lags 1, and 3 exceeds the boundaries and are statistically significant to compute the partial autocorrelation between the lags and thereby determine the value of p for AR model.

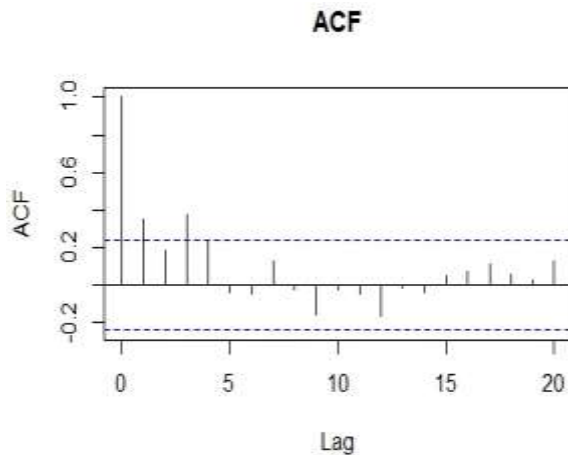


Figure 5. ACF in currency data set

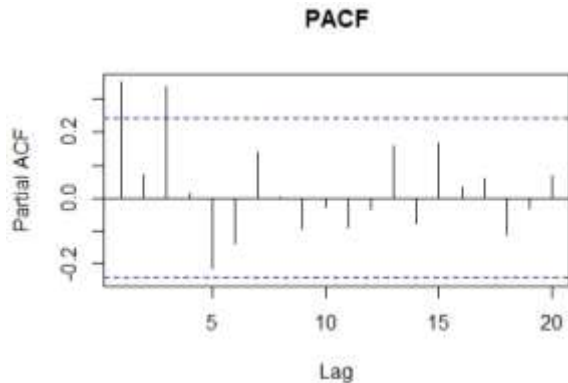
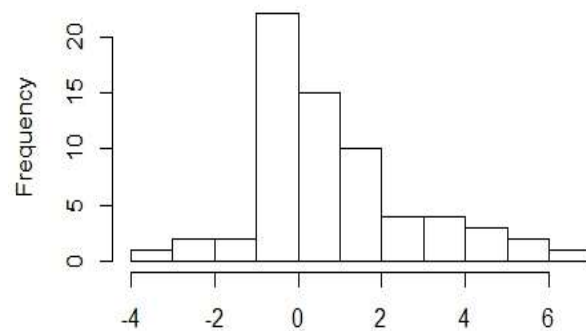


Figure 6. PACF in currency data set

After verifying statistical significance with autocorrelation and partial correlation, a verification towards null hypothesis of independence in time series data is checked using Ljung-Box test. This test results χ^2 value as 35.19 with a more significant sample auto-correlation of the residuals and p -value as 0.01911, mentioning null hypothesis cannot be rejected as p -value > 0.05 . Both these inferences give an impression that the

time series values cannot be decided either as dependent or independent. To make sure that the predictive model can be improved one more check is induced on forecasted errors by verifying in a way to state whether they are typically distributed with continual variance and zero means.

The results of this test with forecast errors and its frequency are shown as a histogram in Figure 7, having an average mean of all the residuals as 0.9460307. In Holt winter forecasting the assumptions related to forecasting between successive values of time series are not made. So, to create a better forecasting model and make the time series as stationary ARIMA models are used. ARIMA uses MA model to make short-term dependencies between successive intervals (p), and to make the forecast errors in the prediction intervals uncorrelated, i.e., to convert the time series into stationary time series with (d) differencing times, and AR model to make long-term dependencies between



successive intervals (q).

Figure 7. Forecast errors in currency dataset

The appropriate forecast model is comprised by using the value lag zero in ACF as shown in Figure 5 and thereby initializes p -value as 0. Similarly, the model uses PACF as shown in Figure 6 and thus initialize q value as 1. This strategy is used in ARIMA forecasting model as both of these lags values are the first ones to cross the significant boundaries, marked in blue color in their respective correlogram. The ultimate objective is to choose a model which claims low errors for predictions of the future, and this is done by refitting the model by comparing the errors, AIC and BIC values in the models. The diagnosis report of ARIMA (0, 2, 1) model over currency dataset is shown in Table 3. It is observed from Table 3 that the ARIMA model results in a low δ^2 value which is close to zero and the value of ME shown in Table 3 are very much closer to zero. After computing a value of 0.20 as a difference between the quality fit criteria parameters AIC and BIC, substantial evidence is communicated that the model is fit for forecasting the patterns in time series

data. The best fit of ARIMA model after making residual fit on currency dataset stationary and residuals retention within significant boundaries of ACF and PACF are shown Figure 8.

Table 3. Outcome of ARIMA forecasting model with order (0, 2, 1) on Currency dataset

Parameters	Value	Parameters	Value
δ^2	3.027	RMSE	1.7135
AIC	261.12	MAE	1.0020
AICc	261.32	MPE	0.8234
BIC	265.47	MAPE	4.0116
ME	0.1500	MASE	0.0578

Figure 8. Output of ARIMA predictive model after making currency time series dataset as stationary

Finally, to meet the problem objective in a more statistically significant way, forecasting is conducted on the stationary time series data to get a persistent end outcome. Table 4 and Figure 9 exhibits the fashionable upshots on the currency time series dataset. By comparing the low and high values in prediction intervals of Table 1 and Table 4, it unveiled a fact that ARIMA results better forecasting and the values stated in prediction intervals are more realistic to actuals. Hence statistical and computational intelligent evidence justifies that the approach claimed in this work can be recommended to conduct time series analysis and forecasting efficiently.

Table 4. Forecasted value of Indian Rupee to US Dollar for the year's 2017 to 2020 using ARIMA model.

Year	Forecast	80% prediction interval		90% prediction interval	
		Low	High	Low	High
2017	70.20	67.97	72.43	66.79	73.61
2018	73.22	69.56	76.87	67.62	78.81
2019	76.23	71.12	81.34	68.42	84.04
2020	79.24	72.61	85.87	69.10	89.39

b. Dataset 2 (Beijing city temperature). This dataset is an hourly dataset containing meteorological data of Beijing city temperature archived from Jan 1st, 2010 to Dec 31st, 2014. This dataset is extracted from UCI repositories [14] to conduct time series analysis and forecasting.

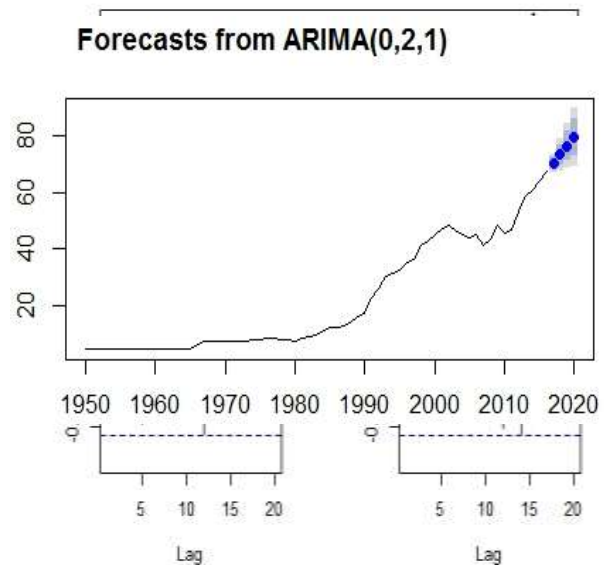


Figure 9. Forecasted value of Indian Rupee to US Dollar for the years 2017 to 2020 about 80% and 90% prediction intervals using ARIMA forecasting model

A total of 43824 samples with 13 attributes are recorded in this dataset. The characteristics include row number, year, month, day, hour, PM2.5 concentration, dew point, temperature, pressure, combined wind direction, cumulated wind speed, cumulated hours of snow and cumulated hours of rain. Here, the objective of this work is set to predict the temperature of Beijing city for the next five years using the proposed TSAF plug-in.

The Beijing city temperature dataset is read as an input from UCI archives in a .csv formatted file with 43824 rows and 13 columns each. Preprocessing of data is performed by applying data cleaning primitive functions such as outlier detection and elimination, replacing the null and missing values using tsclean() command. Data frame is recognized with minimum value and a maximum value ranging from -19 to 42 from January 2010 to December 2014. The data set is now documented in a sequential data across an equally spaced unit of time and decomposed into time series patterns as a trend, seasonal, cyclic and random. The underlying direction of decomposed time series data with a long-term movement in the form of trend, seasonal, cyclic and random are shown in Figure 10.

Holt winter exponential smoothing is used over decomposed time series data with smoothing parameters alpha, beta and gamma to build a predictive model for forecasting. Figure 11 portrays the output of Holt Winter exponential smoothing function with a relationship between actual and predicted time intervals. The value

of smoothing parameter alpha is observed as 0.9999, which indicates the forecast is based on the most recent observations. It is observed from Figure 11 there exists a close guesstimate amongst actual and forecasted values of the time series data since both of them trending in the same way.

pacf() function is also induced with 20 lags, by removing the linear dependence between the lags and establishing a correlation to itself. The same is inferred from the plot depicted in Figure 14.

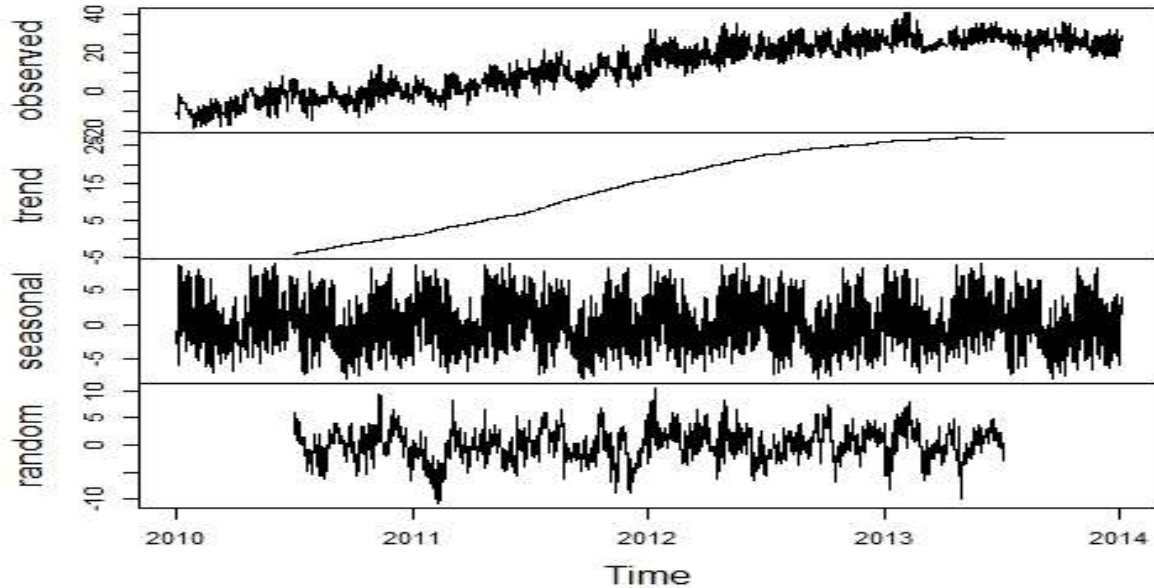


Figure 10. Decomposition of Beijing city temperature dataset into trend, seasonal, random and cyclicAlso, the insinuation at SSE indicates with a value 10630.48, signals comparatively a high percentage of error shown in the model and acclaims to refit the model as motionless. Subsequently, after refitting the model, forecasting is programmed with impalpable buoyancy, by a forecasted value on the temperature of Beijing city with 80% and 95% prediction intervals as shown in Figure 12. The accuracy of this model is tested, and training error raised during the implementation of the model are shown in Table 6 as ME, RMSE, MAE, MPE, MAPE, MASE.

It is observed from Table 6 that maximum error of 1% deviation happened from the actual to the forecasted time series intervals. The interestingness of all the error measures is close to zero, signaling high accurateness in predicting how diligently the forecasted values track with the actuals. A check to find existence to forecast errors and improvement in the predicted model is done by using correlograms. The function acf() is used to identify the possible autocorrelated structure of time series data. It is observed from Figure 13, that almost all the lags including lag 1 exceed significant boundaries, which means that these lags are correlated with current series and thereby representing the correlation of the values of time series with itself and are useful in determining the value of q for MA model. Accordingly,

A significant observation in PACF is that the lags 1 and 3 exceeds the positive side substantial boundaries and rest of the lags, negatively exceeds the significant edge. Such, observations in PACF functions are used to determine the value of p for AR model.

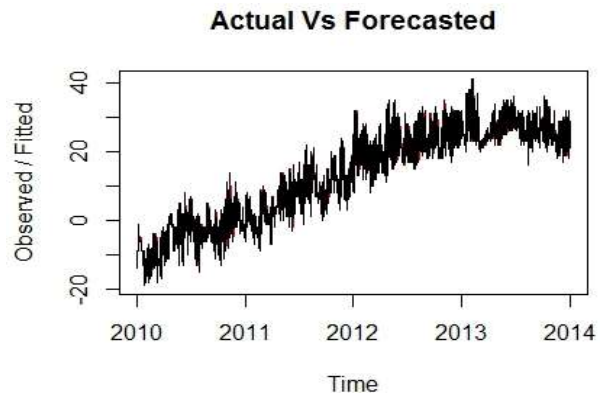


Figure 11. Values of actual vs. forecasted in Beijing city temperature dataset

After verifying statistical significance with autocorrelation and partial correlation, a verification towards null hypothesis of independence in time series data is checked using Ljung-Box test. This test results x-squared value as 6300.1 with a more substantial sample auto-correlation of the residuals and p-value as $2.2e-16$,

mentioning null hypothesis cannot be rejected as $p\text{-value} > 0.05$. Both these inferences give an impression that the time series values cannot be decided either as dependent or independent.

RMSE	1.3572
MAE	1.3173
MPE	0.9482
MAPE	NaN
MASE	Inf 0.07979

Forecasting temperature with Holt Winter

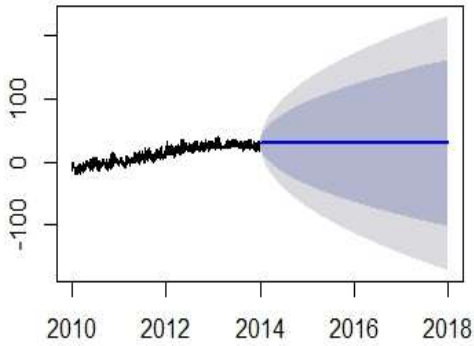


Figure 12. Forecasted temperature of Beijing city for the years 2015 to 2018, pertaining with 80% and 95% prediction intervals using Holt winter forecast model

To make sure that the predictive model can be improved one more check is induced on forecasted errors by verifying in a way to state whether they are typically distributed with continual variance and zero means. The results of this test with forecast errors and its frequency are shown as a histogram in Figure 15, having an average mean of all the residuals as 0.0069. In Holt winter forecasting the assumptions related to forecasting between successive values of time series are not made

So, to create a better forecasting model and make the time series as stationary ARIMA models are used. ARIMA uses MA model to make short-term dependencies between successive intervals (p), and to make the forecast errors in the prediction intervals uncorrelated, i.e., to convert the time series into stationary time series with (d) differencing times, and AR model to make long-term dependencies between successive intervals (q). The appropriate forecast model is comprised by using the value lag zero in ACF as shown in Figure 13 and thereby initializes p-value as 5. Similarly, the model uses PACF as shown in Figure 14 and thereby initialize q value as 4.

Table 6. Values of training errors produced by Holt winter forecasting model on Beijing city temperature dataset

Error	Value
ME	0.0069

ACF

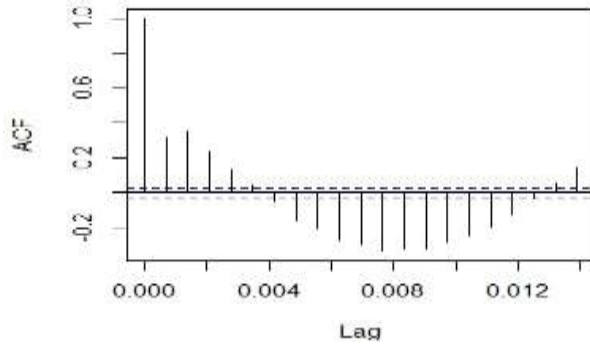


Figure 13. Lag value of ACF in Beijing city temperature dataset

PACF

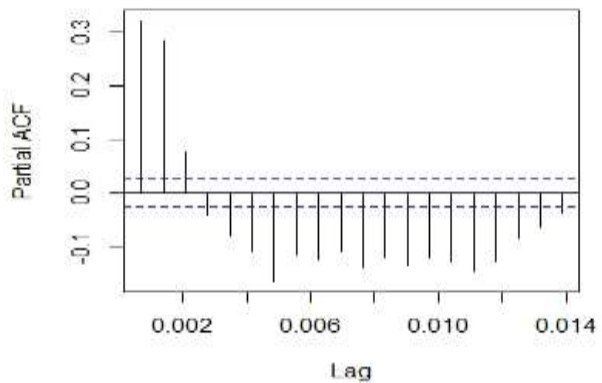


Figure 14. Lag values of partial autocorrelation frequency in Beijing city temperature dataset

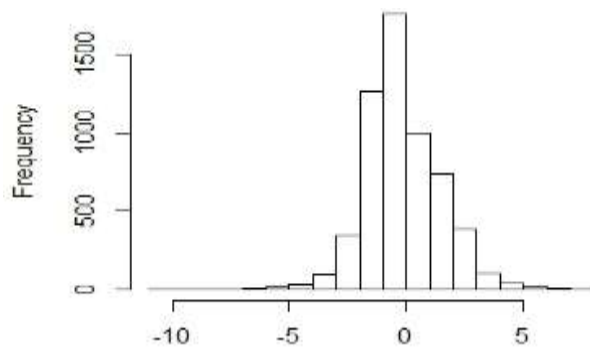


Figure 15. Forecast errors in Beijing city temperature dataset

This strategy is used in ARIMA forecasting model as both of these lags values are the first ones to cross the significant boundaries, marked in blue color in their respective correlogram. The ultimate objective is to choose a model which claims low errors for predictions of the future, and this is done by refitting the model by comparing the errors, AIC and BIC values in the models. The diagnosis report of ARIMA (5, 1, 4) model over Beijing city temperature dataset is shown in Table 7. ARIMA model results are shown in Table 7 a low δ^2 value which is close to zero, and with the value of ME shown in Table 7 are very much closer to zero. After computing a value of -66.61 as a difference between the quality fit criteria parameters AIC and BIC, substantial evidence is communicated that the model is fit for forecasting the patterns in time series data. The best fit of ARIMA model after making residual fit on Beijing city temperature dataset stationary and residuals retention within significant boundaries of ACF and PACF are shown Figure 16.

Table 7. Outcome of ARIMA forecast model with order (5, 1, 4) on currency dataset

Parameters	Value	Parameters	Value
δ^2	1.219	RMSE	1.1039
AIC	17543.1	MAE	0.7951
AICc	17543.14	MPE	NaN
BIC	17609.71	MAPE	Infinite
ME	0.0067	MASE	0.0669

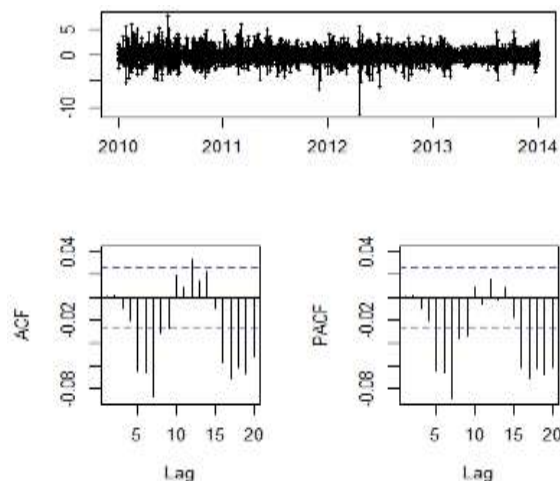


Figure 16. Outputs of ARIMA predictive model after making Beijing city temperature dataset stationary

Lastly, to encounter the problem objective in the narrated workflow sequencing of Figure 1, forecasting is steered on the stationary time series data to acquire tenacious

ending aftermaths. Table 8 and Figure 17 parades the sophisticated consequences on the currency time series dataset. Table 8 discloses the point that results of ARIMA model can be used for better forecasting with convincing prediction. Hence the paramount approach given in TSAF plug-in with computational intelligent and statistical substantiation can be commended and efficiently rationalized for time series analysis and forecasting. The R language code compiled to make time series analysis and forecasting on Beijing city temperature dataset is shown in Table 9.

Table 8. Value of Temperature in centigrade for January 2018 in Beijing city temperature using ARIMA forecast model

Year	Month	Day	Hour	Forecast	80% prediction		90% prediction	
					Low	High	Low	High
2018	1	1	0	30.13	28.7	31.5	27.9	32.3
2018	1	1	1	30.98	28.9	32.9	27.9	34.0
2018	1	1	2	31.52	28.9	34.0	27.6	35.4
2018	1	1	3	31.71	28.7	34.7	27.1	36.3

Forecasts from ARIMA(5,1,4)

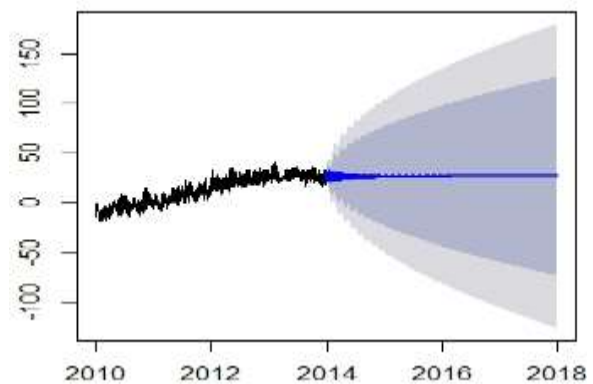


Figure 17. Forecasted value of temperature in centigrade for the years 2017 to 2020 in Beijing city temperature dataset pertaining 80% and 90% prediction intervals using ARIMA model

5. Conclusion

This paper deliberated a premeditated approach to make predictions and analysis over time series data with a TSAF plug-in. All the essential elements such as computational logic with machine intelligence primitives, statistically significant methods are incorporated into this plug-in. The user supplies the time series data as an input, and abstractly after computation of the internal logic without any intervention, the final outputs are generated in the form of time series patterns such as trend, seasonal, cyclic components. This plug-in not only figures the time series patterns but does optimized forecasting more precisely. The usage of this plug-in is thoroughly tested by imposing real-time datasets as inputs and clear-cut results, and predictions

are exhibited in the previous section more invariably by revitalizing the unique abilities of each method. For novel investigates, this chaperone can be used as a platform to practice and demeanor research in this direction by having a quick insight into the projected methodology.

Table 9. R language code on Beijing city temperature dataset to conduct time series analysis and forecasting

```
setwd("E:/Time Series analysis/Final Datasets")
library("forecast", lib.loc=~R/win-library/3.3")
library("tseries", lib.loc=~R/win-library/3.3")
#-----
mytemp<-
read.csv("beijingcitytemperature.csv",stringsAsFac
tors=FALSE)
tsclean(mytemp)
summary(mytemp)
str(mytemp)
mytemptimeseries <- ts(mytemp, start=c(2010,1),
end=c(2014,12), frequency=1440)
plot(mytemptimeseries)
mytemptimeseriescomponents <-
decompose(mytemptimeseries)
plot(mytemptimeseriescomponents)
#-----
mytemptimeseries_HW_beta_gamma <-
HoltWinters(mytemptimeseries, beta=FALSE,
gamma=FALSE)
plot(mytemptimeseries_HW_beta_gamma,
main="Actual Vs Forecasted")
mytemptimeseries_HW_beta_gamma
mytemptimeseries_HW_beta_gamma_startvalue<-
HoltWinters(mytemptimeseries, beta=FALSE,
gamma=FALSE, l.start=-11.00)
mytemptimeseriesforecasts1<-
forecast.HoltWinters(mytemptimeseries_HW_beta
_gamma_startvalue, h=5760)
head(mytemptimeseriesforecasts1)
plot.forecast(mytemptimeseriesforecasts1,
main="Forecastig temperature with Holt Winter")
accuracy(mytemptimeseriesforecasts1)
#-----
acf(mytemptimeseriesforecasts1$residuals,
lag.max=20, main="ACF")
pacf(mytemptimeseriesforecasts1$residuals,
lag.max=20, main="PACF")
Box.test(mytemptimeseriesforecasts1$residuals,
lag=20, type="Ljung-Box")
plot.ts(mytemptimeseriesforecasts1$residuals)
hist(as.vector(mytemptimeseriesforecasts1$residual
s))
mean(mytemptimeseriesforecasts1$residuals)
#-----
```

```
fit1 = arima(mytemptimeseries, order=c(5,1,4))
accuracy(fit1)
tsdisplay(residuals(fit1), lag.max=20)
#-----
fcast <- forecast(fit1, h=5760)
write.csv(fcast, file = "E:/Time Series
analysis/TSAF Datasets/btemp.csv")
plot(fcast)
```

References

- [1] Zhang, J., Zhao, Z., Xue, Y., Chen, Z., Ma, X., & Zhou, Q. (2017). Time series analysis. Handbook of Medical Statistics, 269.
- [2] Chatfield, C. (2016). The analysis of time series: an introduction. CRC press.
- [3] Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting. springer.
- [4] Fanaee-T, Hadi, and Gama, Joao, 'Event labeling combining ensemble detectors and background knowledge', Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg.
- [5] Avril Coghlan. (2017). A Little Book of R For Time Series.
- [6] Osabuohien-Irabor Osarumwense. Time Series Forecasting Models: A Comparative Study of some Models with Application to Inflation Data. Open Science Journal of Statistics and Application. Vol. 2, No. 2, 2014, pp. 24-29.
- [7] Etebong P. Clement. (2014), Using Normalized Bayesian Information Criterion (Bic) to Improve Box - Jenkins Model Building, American Journal of Mathematics and Statistics, 4(5): 214-221, DOI: 10.5923/j.ajms.20140405.02.
- [8] Box, G. E. P., and Pierce, D. A. (1970), Distribution of residual correlations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association, 65, 1509–1526.
- [9] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [10] Makridakis, S. and Hibon, M. (1979). Accuracy of Forecasting: An Empirical Investigation (with discussion). Journal of the Royal Statistical Society (A), 142, 97 – 145.
- [11] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R. and Winkler, R. (1982), "The Accuracy of Extrapolation (Time 23 Series) Methods: Results of a Forecasting Competition," Journal of Forecasting, 1, 111 – 153.
- [12] Holt, C.C. et al., (1960). Planning Production, Inventories and work force, Prentice-Hall: Englewood cliffs, Chapter 14.
- [13] Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine,

- CA: University of California, School of Information and Computer Science.
- [14] Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H. and Chen, S. X. (2015). Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society A*, 471, 20150257.
- [15] Booth et al., *Hydrologic Variability of the Cosumnes River Floodplain, San Francisco Estuary and Watershed Science*, Volume 4, Issue 2, 2006.
- [16] "NIST/SEMATECH e-Handbook of Statistical Methods". NIST. Retrieved 2010-05-23.
- [17] Charles C. Holt (1957), *Forecasting Trends and Seasonals by Exponentially Weighted Averages*, Carnegie Institute of Technology, Pittsburgh Office of Naval Research memorandum no. 52.
- [18] Brown, Robert G. (1956). *Exponential Smoothing for Predicting Demand*. Cambridge, Massachusetts: Arthur D. Little Inc. p. 15.
- [19] Oppenheim, Alan V.; Schafer, Ronald W. (1975). *Digital Signal Processing*. Prentice Hall. p. 5. ISBN 0-13-214635-5.
- [20] Goodwin, P. (2010). The holt-winters approach to exponential smoothing: 50 years old and going strong. *Foresight*, 19, 30-33.
- [21] Kalekar, Prajakta S. "Time series forecasting using holt-winters exponential smoothing." *Kanwal Rekhi School of Information Technology* 4329008 (2004): 1-13.
- [30] De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3), 443-473.
- [31] Hyndman, R. J., & Khandakar, Y. (2007). *Automatic time series for forecasting: the forecast package for R* (No. 6/07). Monash University, Department of Econometrics and Business Statistics.
- [22] Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1), 1-28.
- [23] "Notation for ARIMA Models". *Time Series Forecasting System*. SAS Institute. Retrieved 19 May 2015.
- [24] Hyndman, Rob J; Athanasopoulos, George. "Seasonal ARIMA models". *Forecasting: principles and practice*.
- [25] Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C. (2008). *Time Series Analysis, Forecasting and Control* (4th ed.). Hoboken, NJ: Wiley. ISBN 9780470272848.
- [26] Box, G. E. P.; Pierce, D. A. (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models". *Journal of the American Statistical Association*. 65: 1509–1526. JSTOR 2284333. doi:10.1080/01621459.1970.10481180.
- [27] G. M. Ljung; G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". *Biometrika*. 65 (2): 297–303. doi:10.1093/biomet/65.2.297.
- [28] Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John Wiley & sons.
- [29] De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496), 1513-1527.