ASM Sc. J., 12, Special Issue 5, 2019 for ICoAIMS2019, 160-166

The Combination of Forecasts with Different Time Aggregation

Nur Haizum Abd Rahman[1*] and Muhammad Hisyam Lee[2]

[1]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia*

[2]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia*

In forecasting, it is important to improve forecast accuracy. Thus, the forecast combination have been proposed in the literature. Usually, the classical approach in forecast combination obtain from the composite of two (or more) available forecasts with identical timings. However, forecast horizon, short and long term do affect the forecast performance. Therefore, unlike previous combinations, this paper combined the forecasts with different time aggregations in order to capture the unique information of the data set. We had considered the problems in forecasting daily air pollutant index (API) as well as the monthly aggregate, by using the Box-Jenkins method and fuzzy time series method as the time series approach. Then, the monthly aggregate forecasts were interpolated to obtain the forecasts on a daily basis. Each of the original forecasts was used to determine the weights in forming the combined forecast. The error magnitude measurements were used to measure the accuracy. The result showed that the forecast combinations with different timings outperformed the individual forecasts and traditional forecast combinations with identical timing. Hence, the combination of different timing data sets produced better forecasting accuracy, which can be a good practice in many types of data with different time horizon.

Keywords: forecast combination; forecasting; aggregate data; box-jenkins; fuzzy

I. INTRODUCTION

Time series forecasting has been used in various fields to predict the future events particularly in decision making (Hyndman & Athanasopoulos, 2018). Forecasting method can be classifiend into classical and modern methods. Although recent studies show that the newer and more advanced forecasting techniques tend to result in improved forecast accuracy under certain circumstances, no clear-cut evidence shows that any one model can consistently outperform other models in the forecasting competition (Song & Li, 2008). Hence, improvements in forecasting performance remain as the main issues search by forecasters in many organizations (Makridakis, Spiliotis, & Assimakopoulos, 2018).

Forecast combination is one of the developed methods used to improve the forecast accuracy. It was firstly introduced by Bates and Granger (Bates & Granger, 1969). Usually, forecaster tends to discover many time series methods to get the best forecast result. Then, it will be chosen as the best method while others were being discarded (Prybutok, Yi, & Mitchell, 2000). However, the discarded method could contain some useful independent information. This is because, different method considered different information with dissimilar assumption in relating between the time series behaviour (Bates & Granger, 1969). For this reason, it is important to applied forecast combination in forecasting.

In the past, the forecast combination has been well presented in many studies and successful as alternative techniques of the individual forecasts (Trabelsi & Hillmer, 1989). Traditionally, the forecast combination can be achieved by combining at least two or more forecast results by assigning weights to the individual forecasts (Winkler & Makridakis, 1983). Some ways of combining forecasts are based on minimum variance method and the arithmetic mean. Both of these weighting methods were considered

*Corresponding author's e-mail: nurhaizum_ar@upm.edu.my

by most of forecasters to perform the forecast combination (Andrawis, Atiya, & El-Shishiny, 2011; Martins & Werner, 2012).

Rapidly studies were published in order to present new ways of combining. Especially in different forecasting methods with identical timings and comparing the simpler existing methods with more complex and sophisticated weights assigning to the individual forecasts (Bates & Granger, 1969; Wong, Song, Witt, & Wu, 2007). However, combining forecasts using different time aggregation instead of combining with the identical timings data set also important. This different time scales forecast combination beneficial in capturing the different dynamics in time series data. However, limited number of studies documented in conducting forecast combination with different time aggregation

Andrawis *et al.* (2011) considered tourism data in Egypt with monthly series for short term data set and by using summation they aggregate the monthly series to yearly series to get long term data set. Then, compared with different weight assigning in performing the combination forecast. For this paper, the forecast combination between short term and long term data set will be explored where the Box-Jenkins methods as the main time series approach. To obtain the aggregate data at monthly frequencies, the summation and the average of daily data will be used. Following by comparison between the combination forecast with different time aggregation, single forecast and traditional composite forecast.

This paper was organized as follows. Section 2 contains brief explanation on methodology, application procedure of forecast combination and statistics to evaluate the accuracy of forecasting performance. Followed by the results and discussion in Section 3. Finally, the conclusions contained in Section 4.

II. MATERIALS AND METHOD

In this section, all the details of proposed procedures, seasonal ARIMA method and forecast combinations methods are discussed.

A. Combination different time aggregation

Normally, the data have been examined in forecast combination are all being identical timing. However, in this paper we interested in different timing time series data consist of short term and long term dataset to obtain single forecast by using forecast combination approach. The long term time series data is a time aggregation step from short term time series data. For this study, two aggregations step have been perform which are the summation and average of short term time series dataset.

In our approach, we used specific time series data generating model, Box-Jenkins due to the capability of this method in determine many types of dataset for forecasting (Zhang *et al.*, 2015). Then, to combine the two different timing forecasts, the interpolation step is needed in order to interpolate the long term series into the forecasts at short term frequency. Therefore, this study will have the single forecast based on the combination of common practise and short term forecast derived from the long term time series. This combination step can be written as:

$$\hat{y}_{t+h} = w_1\hat{x}^{(1)}_{t+h} + w_2\hat{x}^{(2)}_{t+h} \qquad (1)$$

where w_1 and w_2 are the combination weights, $\hat{x}^{(1)}_{t+h}$ forecast at short term series, $\hat{x}^{(2)}_{t+h}$ forecast at short term series derived from long term series and $\hat{y}_{t+h}$ is the final forecast for h step ahead forecasts.

B. ARIMA Forecasting

The Box-Jenkins method or ARIMA, is classified as a linear model that capable in presenting both stationary and non-stationary time series (Box & Jenkins, 1976). The three main models in ARIMA are autoregressive (AR), the integrated (I) and the moving average (MA). The AR and MA models suitable for stationary time series pattern and the mixture of AR and MA could obtain the ARMA models whilst for nonstationary dataset, *I* models could make the dataset stationary to obtain ARIMA models.

Tentative identification, parameter estimation and diagnostic checking are the important procedures when determine the best ARIMA model for time series data (Suhartono, 2011). For some cases, where the seasonal components are included, the model is referred to as a seasonal ARIMA model or SARIMA model. The model can be abbreviated as SARIMA (p,d,q)(P,D,Q)^S. The lowercase represents the non-seasonal part whilst the uppercase shows the seasonal part. The generalized form of the SARIMA model can be written as:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (2)$$

where

$$\phi_p(B) = 1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p$$

$$\Phi_P(B^S) = 1-\Phi_1 B^S-\Phi_2 B^{2S}-\cdots-\Phi_P B^{PS}$$

$$\theta_q(B) = 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$$

$$\Theta_Q(B) = 1-\Theta_1 B^S-\Theta_2 B^{2S}-\cdots-\Theta_Q B^{QS}$$

The forecaster usually used multiplicative SARIMA model directly when the autocorrelation function (ACF) and partial autocorrelation function (PACF) indicated the data contained both non-seasonal and seasonal orders. However, there are more precise model identification step particularly in lags implication (Suhartono, 2011). It can be either, multiplicative, subset or additive SARIMA model (see Suhartono 2011 for model identification explanation).

C. Forecast Combination Weights Assigning Methods

Different weight assigning methods have been proposed (Armstrong, 2001). However, complex methods do not always provide more accuracy than simpler methods (Makridakis & Winkler, 1983). The combination illustrated by arithmetic mean of the two individual forecasts is considered the most in composite forecasts (Martins & Werner, 2012). In this scheme, the forecast is the simple average of the two individual forecasts, where (1) becomes:

$$\hat{y}_{t+h} = \frac{1}{2}\hat{x}_{t+h}^{(1)} + \frac{1}{2}\hat{x}_{t+h}^{(2)} \quad (3)$$

Another combinations weights is minimum variance method (VAR). This approach give greater weight to the set of forecasts which seemed to contain the lower mean-square errors. The weights can be assign to the individual forecasts as below:

$$w_1 = \frac{\sigma_2^2-\sigma_{12}}{\sigma_1^2+\sigma_2^2-2\sigma_{12}};\ w_2 = 1-w_1 \quad (4)$$

where σ_1^2 and σ_2^2 are the variance for individual forecasts respectively and σ_{12} is the covariance of the two forecasts. In the case where the individual forecasts are independent, the covariance between the forecasts is null. These weights abbreviated as VAR-NO-CORR. The optimal weights can be written as:

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2};\ w_2 = 1-w_1 \quad (5)$$

C. Performance evaluations

It is important to check the forecasting performance to identify the best model with the smallest error. There are several statistical tests can be considered to measure the model validation. This paper use forecasting accuracy based on the root mean squared error (RMSE) and mean absolute error (MAE). Then,

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y}_t| \quad (7)$$

where y_t is an actual value, $\hat{y}_t$ is a predicted value, and n is the number of predicted value. All of the statistical computations measure the residual errors with the smallest values will be chosen as the best model to be used in forecasting.

The statistical indicators RMSE were utilized to estimate the obtained result. The RMSE describe the difference between the observed and forecast values with high sensitivity in extreme values due to the power term and magnitude of the error that would be relatively more useful to decision makers (Armstrong & Collopy, 1992).

III. RESULTS AND DISCUSSION

The approach of this study contemplated the use of real series. Daily air pollutant index (API) was used as the case study. These data were obtained from the Department of the Environment (DoE) Malaysia through Alam Sekitar Malaysia Sdn. Bhd., a private company which is responsible for monitoring and managing ambient air quality in Malaysia.

The API in Malaysia was developed based on the API introduced by the United State Environmental Protection Agency (USEPA). It is determined by the calculation of five main pollutants sub-indexes, namely PM_{10}, O_3, CO_2, SO_2 and NO_2. Government agencies used API to characterize the status of air quality since it provide an easy assessment and harmonization at the given location through out the countries (Sansuddin *et al.*, 2011). API also helps general public to understand easily the air quality status for their own health precaution (Kumar & Jain, 2010). This information with different ranges of human health reflects as "Good (0-50), Moderate (51-100), Unhealthy (101-200), Very Unhealthy (201-300) and Hazardous (301 and above)." In addition, API can be as the benchmark of air quality management or data interpretation of decision making processes.

Figure 1 shows the time series plot of the daily mean of API. There are remarkable seasonal variations in API with very high concentrations recorded during the dry season of the summer monsoon. Days when API readings exceeded the "Unhealthy" zone suggested by Recommended Malaysian Air Quality Guidelines (RMAQG) are infrequent. However, on most days during the summer monsoon, the API readings for all of the stations exceeded "Moderate" zone, a limit value suggested by the World Health Organization (WHO).

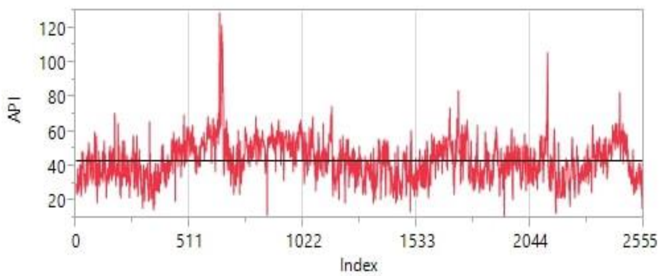

Figure 1. Air pollutant index data

The analysis given here includes the forecast combination methods and individual forecast. The daily API data was initially modelled based on Box-Jenkins procedure and analysis by using SAS software. To begin the analysis, the API plots show that the data contain seasonality trend which normally influence by seasonal wind pattern and variety of sources. Thus, the data is not satisfy the stationary condition. Therefore, before ascertaining the tentative model identification, the differencing process is necessary to produce a new series that is compatible with the assumption of stationary. By using difference both non-seasonal (d=1) and seasonal (D=1, S=365) then the data become stationary series.

Based on stationary series, the tentative model for ACF and PACF were examined to determine the best combination order of seasonal ARIMA model for each data set. Model identification is the most crucial stage in building ARIMA models. Differ from most of previous paper used directly the multiplicative model, more details identification and testing in choosing the most appropriate order of SARIMA are obtained. By using the procedure suggested by Suhartono (2011), the best SARIMA for this data set is SARIMA $([2,4,5],1,2)(0,1,1)^{365}$ which known as subset SARIMA model.

Traditionally, the forecast combination is based on two individual forecast models. Though, this study focused on SARIMA models and proposed to combine the forecast with different time aggregation but as the comparison purposes the forecast combination between two models with identical timing using SARIMA and modern methods, fuzzy time series also conducted. The details of fuzzy method can be found in (Rahman, Lee, Suhartono, & Latif, 2015).

For the first forecast, $\hat{x}_1$ the best SARIMA model denoted as $([2,4,5],1,2)(0,1,1)^{365}$. For the second forecast, $\hat{x}_2$, two aggregate monthly datasets based on summation and average of daily API with the SARIMA model are SARIMA $(1,0,0)(0,1,1)^{12}$ and SARIMA $(0,1,1)(0,1,1)^{12}$ respectively. The second forecasts from monthly series are interpolated to create daily forecast as described previously. Table 1 shows the out-sample performance for API dataset for both single models based on daily data and monthly data. Forecast accuracies for both of the aggregate daily data (summation and average) are computed after the forecasts converted to daily data using interpolation. From Table 1, the average aggregate outperform others approach with the lowest value of 6.205 and 7.765 in MAE and RMSE respectively. The second best forecast performance is summation aggregate following by fuzzy and SARIMA model.

Forecast performance of three different forecast combinations with three forecast combination methods approach limited to simple average (AVG), variance with correlation (VAR) and variance without correlation (VAR-NO-CORR) are shown in Table 2. For Model 1, the forecast combination is obtained from the individual forecasts, SARIMA and fuzzy time series with identical timing. Meanwhile for Model 2 and Model 3 are the combinations of SARIMA forecast from different timing (short term and long term dataset) by using summation and average approach correspondingly.

Table 1. The out-sample performance of the single models

Models	MAE	RMSE
SARIMA	7.860	9.651
Fuzzy	7.734	9.626
Summation aggregate	7.047	8.712
Average aggregate	6.205	7.769

Table 2. The out-sample forecast combinations performance

Forecast Evaluation	MAE	RMSE
Model 1		
AVG	7.338	9.081
VAR	7.341	9.081
VAR NO-CORR	7.340	9.080
Model 2		
AVG	6.976	8.705
VAR	6.903	8.587
VAR NO-CORR	6.905	8.613
Model 3		
AVG	6.767	8.396
VAR	6.580	8.196
VAR NO-CORR	6.672	8.289

Based on Table 2, as overall in Model 1, forecast combination achieved 6% improvement compared to the individual forecasts. As in the table, the RMSE by using three different forecast combinations methods are the same, 9.08. This is because the fuzzy taken into consideration using the lags obtained from SARIMA model. This result suggest that, the capability of individual forecast (SARIMA and fuzzy) applied are the same. Thus, the weighting of both then can be denoted as equal.

For Model 2, all the combination methods can outperform the individual SARIMA model and summation aggregation forecast. Differ from Model 2, Model 3 unsuccessful to outperform the individual average aggregation forecast but success in outperform the SARIMA model. According to Andrawis *et al.* (2011), the capability in achieving the best forecast combination compared to individual forecast depends on degree of variation in the accuracies of the underlying models. Such that, if they are methods can perform the good forecasting models, it is harder for forecast combination to beat the individual forecasts. Although Model 3 only can beat one of the individual forecast (SARIMA), Model 3 still achieved the best forecasting performance compared to Model 1 and Model 2.

Additional result from forecast horizon using single forecast, SARIMA and fuzzy including the forecast combination from Model 1, Model 2 and Model 3 were performed and shown in Table 3. Based on each forecast horizon, fuzzy give good result in forecast horizon 1, 2 and 3 but worsen after the third forecast horizon. Therefore, in average the fuzzy become the last preferred model with MAE 4.65. The third best method is SARIMA where it outperform the Model 1 due to worst forecast performance in fuzzy. These results strengthen our study where the combinations of forecasts outperform the single methods with Model 3 provide higher accuracy compared to the second best model Model 2 where both are based on forecast combination approach. Model 3 also give consistent result through out the forecast horizon.

Table 3. Forecast accuracy measures for monthly aggregate forecast based on MAE

Forecast Horizon	Methods				
	SARIMA	Fuzzy	Model 1	Model 2	Model 3
1	3.54	0.99	2.21	4.61	2.83
2	5.77	0.94	2.28	6.00	2.44
3	6.01	1.70	3.77	8.81	3.58
4	2.17	0.77	0.64	5.56	0.91
5	3.88	9.65	6.88	2.81	5.03
6	6.76	9.25	8.05	4.66	6.34
7	1.66	3.04	2.38	1.81	0.70
8	4.03	3.44	3.72	0.17	2.69

9	9.44	6.99	8.17	5.17	7.33
10	3.11	0.24	1.36	0.77	1.83
11	0.04	6.45	3.37	3.61	0.57
12	1.49	12.36	7.14	3.45	1.48
Average	3.99 (3)	4.65 (5)	4.16 (4)	3.95 (2)	2.98 (1)

IV. CONCLUSION

This paper has discussed the idea of forecast combination by using different timing of dataset namely as short term and long term including the traditional composite forecast from two different time series forecasting methods as comparison purpose. For the realization of this study, the real time series data set from air quality term, API were used. The data provided in short term daily data. By using summation and average, the daily data were aggregate into long term data set at monthly frequencies. The individual forecasts for different timing were obtained by Box-Jenkins methods then the long term forecasts were interpolate to get the forecasts in short term frequency. The combinations weights were carried out using the arithmetic mean and minimum variance. As the comparison purpose, the traditional composite forecast between two methods, Box-Jenkins and fuzzy time series also taken into consideration. The MAE and RMSE were used to measure which combinations have the best accuracy.

In general, these empirical results show that the forecast combinations must be consider in forecasting since most of the combined forecast can be better than both single forecasts. The best combination result was obtained by using our proposed different time aggregation, average as shown in Model 3. This study prove that forecast combinations provide improvement in forecast performance. Therefore, this study suggests that forecast combination with different timing must take into consideration by forecasters in order to make used all the variability obtained by using individual forecasts to perform better forecast accuracy.

V. ACKNOWLEDGEMENT

We would like to thanks Department of Environment Malaysia in providing the air quality data set. This study was supported by Universiti Putra Malaysia, Malaysia under Putra-IPM grant, 9587700.

VI. REFEREENCES

Andrawis, R. R., Atiya, A. F., & El-Shishiny, H. (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, *27*(3), 870–886.

Armstrong, J. S. (2001). *Combining Forecasts*. 417–439.

Armstrong, J. S., & Collopy, F. (1992). Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons By J. Scott Armstrong and Fred Collopy Reprinted with permission form. *International Journal of Forecasting*, *8*(1), 69–80.

Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *Or*, *20*(4), 451.

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control San Francisco. *Calif: Holden-Day*.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting : principles and practice. In *OTexts* (2nd ed.). Retrieved from OTexts.com/fpp2

Kumar, U., & Jain, V. K. (2010). ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment*, *24*(5), 751–760.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, *29*(9), 987–996.

Makridakis, Spyros, Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*(4), 802–808.

Martins, V. L. M., & Werner, L. (2012). Forecast combination in industrial series: A comparison between individual forecasts and its combinations with and without correlated errors. *Expert Systems with Applications*, *39*(13), 11479–11486.

Prybutok, V. R., Yi, J., & Mitchell, D. (2000). Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research*, *122*(1), 31–40.

Rahman, N. H. A., Lee, M. H., Suhartono, & Latif, M. T. (2015). Artificial neural networks and fuzzy time series forecasting: an application to air quality. *Quality & Quantity*, *49*(6), 2633–2647.

Sansuddin, N., Ramli, N. A., Yahaya, A. S., Yusof, N. F. F. M., Ghazali, N. A., & Madhoun, W. A. Al. (2011). Statistical analysis of PM10 concentrations at different locations in Malaysia. *Environmental Monitoring and Assessment*, *180*(1–4), 573–588.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting-A review of recent research. *Tourism Management*, *29*(2), 203–220.

Suhartono. (2011). Time Series Forecasting by using Seasonal Autoregressive Integrated Moving Average: Subset, Multiplicative or Additive Model. *Journal of Mathematics and Statistics*, *7*(1), 20–27.

Trabelsi, A., & Hillmer, S. C. (1989). A benchmarking approach to forecast combination. *Journal of Business and Economic Statistics*.

Winkler, R. L., & Makridakis, S. (1983). The Combination of Forecasts. *Journal of the Royal Statistical Society. Series A (General)*, *146*(2), 150–157.

Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, *28*(4), 1068–1078.

Zhang, X., Rui, X., Xia, X., Bai, X., Yin, W., & Dong, J. (2015). A hybrid model for short-term air pollutant concentration forecasting. *10th IEEE Int. Conf. on Service Operations and Logistics, and Informatics, SOLI 2015 - In Conjunction with ICT4ALL 2015*, 171–175.