Thailand Statistician
October 2023; 21(4): 887-909
http://statassoc.or.th
Contributed paper

A Novel Approach for Combined Forecasting Model Systems Based on the Correlation Coefficient Ranking of the Individual Forecasting Models

Lily Ingsrisawang and Chantha Wongoutong*
Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, Thailand.
*Corresponding author; e-mail: fscictw@ku.ac.th

Received: 3 July 2022
Revised: 24 November 2022
Accepted: 7 June 2023

Abstract

This study investigates the efficiency of forecasting by two combined forecasting model methods (simple-average and Bates-Granger) comprising up to five individual forecasting models. The novel approach of ranking the correlation coefficients between the predicted values produced by the individual forecasting model and the actual values is used to rank the individual forecasting models used in the combined forecasting model methods. The results of a simulation study and using real datasets with a stationary pattern time series demonstrate that the simple-average and Bates-Granger combined approaches based on the two highest-ranked individual forecasting models improved the forecasting performance better than with five, four, or three individual forecasting models, especially for a short time-series dataset. Hence, the novel approach using two vital individual models in their correlation coefficients for combined forecasting produces fewer errors and improves the forecasting accuracy than 3-5 individual models in the combined forecasting approach.

Keywords: Combined forecasting method, correlation coefficient, model weighting, stationary.

1. Introduction

Time-series forecasting is used across many fields of study, such as weather forecasting, earthquake prediction, astronomy, statistics, econometrics, signal processing (Kim 2003; Cortez et al. 2004; Ahmed et al. 2010; Yolcu et al. 2013). A time-series forecasting model is used by extracting knowledge from past observations to identify the evolution of a phenomenon in the present and facilitate future projections (Box et al. 1994). The challenges of time-series modeling are the high complexity of time-series data and the low accuracy and poor generalization ability of the prediction model (Hyndman and Athanasopoulos 2018). In practical situations, when applying a statistical modeling approach for time-series data analysis and forecasting, one faces the important issue of choosing the best model among various candidates. Since a single forecasting model has both advantages and disadvantages, using it alone is only effective to a certain extent under certain circumstances (Taylor and Bunn 1999). Furthermore, setting the conditions and other factors will also affect a single model. Therefore, predictions are often based on a range of insufficient information sources, leading to difficulties in meeting the requirements of the forecasting scenario (Xiao and Wu 2008).

Forecasting accuracy can be improved by using several methods on the same time series and averaging the resulting forecasts by using a combined forecasting model method. Over half a century ago, Bates and Granger (1969) systematically studied a combined forecasting model for the first time and showed that combining forecasts from several methods often leads to better forecasting accuracy. Likewise, Clement and Hendry (1989) and Armstrong (2001) claimed that combining forecasts by averaging them can reduce errors, which is particularly useful when uncertainty exists about which forecasting model is the best for a certain prediction. Moreover, studies on combined forecasts with univariate time-series models show that combining forecasting models can reduce errors (Makridakis and Winkler 1983, Sanders and Ritzman 1989, Lobo 1992, Thaithanan and Wongoutong 2020). Consequently, the point of using a combined forecasting model method is to integrate the prediction results from various forecasting models via various statistical approaches to obtain more accurate forecasting.

Combining forecasts is associated with the performance consistency of the individual forecasting models, and combining at least of them results in more accurate forecasting (Armstrong, 2001; Jose and Winkler 2008; Khandelwal et al. 2015). Thus, it is critical to select appropriate but diverse individual forecasting models from many candidates. Indeed, the number of individual forecasting models is a significant factor affecting the accuracy of the combined forecasting model, albeit the lack of theory makes it difficult to determine this. According to Yang (Yang 2004) the two main directions for combining forecasts are adaptation and improvement: the first targets the best individual performance among a pool of forecasting candidates while the second uses a new forecasting candidate to attempt to outperform the previous one.

As mentioned above, it is not easy to obtain the perfect forecasting model, and thus, eliminating or reducing the errors of a forecasting model is proving to be an eternal topic. Actually, instead of designing a more sophisticated forecasting model, combining existing methods to improve the preciseness of a forecasting model is feasible in statistics. However, combining three or more individual forecasting models may not always produce satisfactory performance, while selecting more appropriate individual forecasting models for the combination may result in better forecasting. Clarifying and developing this area is the motivation behind the present study.

Thus, selecting the individual forecasting models from many candidates and how many individual forecasting models should be used in the combined forecasting model are of particular interest. This study, the novel to build a combined forecasting model method based on ranking the individual forecasting models via correlation coefficients is proposed for two well-known combined forecasting models with the simple-average and Bates-Granger.

The rest of this paper is organized as follows. In Section 2, the data used in the study are presented, while Section 3 provides the individual forecasting models, combination forecasting methods, and the proposed method used in the study. The experimental study to compare the methods is described in Section 4. The results of the experimental study and a discussion are given in Section 5. Finally, conclusions on the study are included in Section 6.

2. The Forecasting Methods and Models Used in the Study

Time-series forecasting is when past observations of a variable are collected in a sequence of data points indexed in temporal order and analyzed to develop a method or model that describes the underlying relationship. Finding a model that represents reality and predicts efficaciously is the main objective of forecasters. Many researchers have developed and improved time-series modeling over the past several decades. Assembling several individual forecasting models to improve prediction is known as combined forecasting, which has been shown to outperform forecasting by using the individual methods in most cases (De Menezes et al. 2000). Furthermore, Armstrong (2001) claimed that the number of individual forecasting models to optimize combinatorial efficacy is five. Hence, a novel approach for selecting the individual forecasting models for the combined forecasting model method from many candidates and how many individual forecasting models should be used are presented herein. The individual forecasting models and combined forecasting model methods briefly described in this section are available in the R statistical package.

2.1. The individual forecasting models

2.1.1 Simple moving average

The simple moving average procedure is a simple and common type of smoothing used in time-series analysis and forecasting that is effective when the time-series data are assumed to be stable over time with no trend or seasonality (Svetunkov and Petropoulos 2018). Thus, the simple moving average method can forecast the next value(s) in a time series depending on the average of the previous values over specified k periods for which each point is assigned an equal weight $(1/k)$. The formula for this is

$$\hat{y}_t = \frac{1}{k}\sum_{i=1}^{k} y_{t-i}, \tag{1}$$

where y_t is the actual value, $\hat{y}_t$ is the forecasted value for period t, and k is the length of the simple moving average.

2.1.2 Single exponential smoothing

Exponential smoothing is a very popular scheme for smoothing a time series in which exponentially decreasing weights are assigned as the observations get older (Brown 1956; and Holt 2004). The simplest approach is single exponential smoothing for stable data over time without seasonality or trend. The forecasting model in period t is formulated as

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t, \tag{2}$$

where $\hat{y}_{t+1}$ is the forecast in period $t+1$, $\hat{y}_t$ is the forecast in period t, y_t is the actual value in period t, and α is a smoothing parameter between 0 and 1.

2.1.3 Box-Jenkins

The autoregressive moving average (ARMA) or Box-Jenkins model is commonly used in time-series modeling (Box and Jenkins 1970). It can be used to describe a stationary time series in terms of polynomials of the two components: autoregression and moving average. We begin with the assumption that the process that generated the time series can be approximated by using an ARMA model if it is stationary (Box et al. 2015). The general ARMA model is expressed as

$$(1-\phi_1 \mathrm{B} - \ldots - \phi_\mathrm{p} \mathrm{B}^\mathrm{p}) y_t = (1-\theta_1 \mathrm{B} - \ldots - \theta_\mathrm{q} \mathrm{B}^\mathrm{q})\varepsilon_t, \tag{3}$$

where $\varepsilon_t \sim N(0,\sigma^2)$, real parameters $\phi_1, \phi_2, \phi_3, \ldots, \phi_p$ are autoregressive coefficients, and $\theta_1, \theta_2, \theta_3, \ldots, \theta_q$ are moving average coefficients. Consequently, the model's parameters are estimated by applying the Box-Jenkins methodology in a four-step iterative procedure as follows:

1. Model identification: this involves identifying the order of the model (p and q) required to capture the data's salient dynamic features. Subsequently, the time series, the autocorrelation function (ACF), and the partial ACF are graphically determined.

2. Parameter estimation: this is based on the maximum likelihood or the minimum least-squares methods.

3. Diagnostic checking: plots and statistical tests of the residual errors are used to determine the adequacy of the model fitting, and if need be, consider alternative models.

4. Forecasting: use the appropriate model for forecasting.

2.1.4 Artificial neural network (ANN)

ANN is a computer system that simulates the human brain's learning process that has become popular due to its nonlinear modeling capability for data time-series forecasting (Shi et al. 2013). An ANN usually consists of an input layer, an output layer, and a hidden layer(s). The input layer introduces the data to the network, the hidden layer(s) processes the data, and the output layer produces the results. The most popular ANN architecture in the forecasting domain is the multilayer perceptron (MLP), a class of feedforward ANN that is a nonlinear autoregressive model. The structure of a typical ANN with MLP architecture is shown in Figure 1, including a feedforward structure for the input layer, one or more hidden layers, and an output layer. The nnetar function from the R statistical package was used for fitting the ANN model for the time-series data in this study. This function creates feedforward neural networks with a single hidden layer using lagged inputs for forecasting a univariate time series.

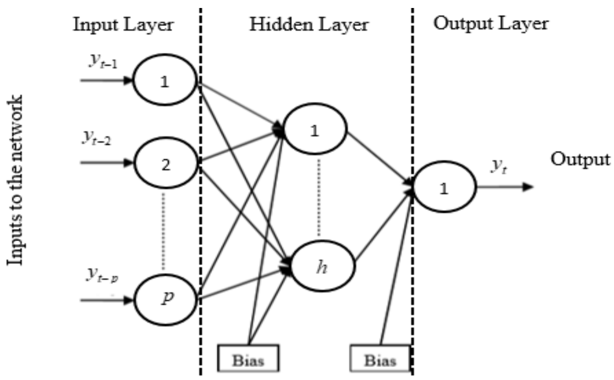


Figure 1 A typical ANN structure

In a fully connected ANN model with p input, h hidden, and a single output node, the relationship between the inputs y_{t-i} $(i = 1, 2, ..., p)$ and the output y_t is given by

$$y_t = G\left(\alpha_0 + \sum_{j=1}^{h} \alpha_j F\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-1}\right)\right), \tag{4}$$

where α_j and β_{ij} ($i = 1, 2, ..., p$; $j = 1, 2, ..., h$) are the connection weights, α_0 and β_{0j} are the bias terms, and F and G are the network activation functions.

2.1.5 Support vector machine (SVM)

The SVM method first suggested by Cortes and Vapnik (Cortes and Vapnik 1995) has been used in many fields, such as in classification, data mining, regression, and time-series forecasting (Zhao et al. 2006; Okasha 2014; Guo et al. 2011). The ability of SVM to solve nonlinear regression estimation problems has led to it being successfully applied to time-series forecasting. In the SVM process, the input space is mapped nonlinearly into a higher dimensional feature space, after which an optimally separating hyperplane is extracted. A set of mathematical functions defined as the kernel function involves taking data as input and transforming them into the required form. Different SVM algorithms use different kernel functions (e.g., linear, nonlinear, polynomial, sigmoid, and radial basis function (RBF)). RBF is the most appropriate type of kernel function for time series (Karatzoglou et al. 2004), and so it was used in the present study. The SVM regression algorithm can be applied in time-series forecasting by adopting a sliding time window defined by the set of time lags $\{k_1, k_2, ..., k_I\}$ used to build the forecast. Subsequently, the SVM algorithm determines the best linear separating hyperplane in the feature space while tolerating a few errors when fitting the data as follows:

$$y_t = w_0 + \sum_{i=1}^{m} w_i \phi_i(\mathbf{y}). \tag{5}$$

For given time period t, the model inputs are $\mathbf{y} = (y_{t-k_I}, ..., y_{t-k_2}, y_{t-k_1})$ and the desired output is y_t. In SVM regression, the input $\mathbf{y} = (y_{t-k_I}, ..., y_{t-k_2}, y_{t-k_1})$ is transformed into a high m-dimensional feature space by using nonlinear mapping (ϕ) that depends on the kernel function. Figure 2 shows SVM regression with the ε-insensitive loss function adapted from (Karatzoglou et al. 2004).

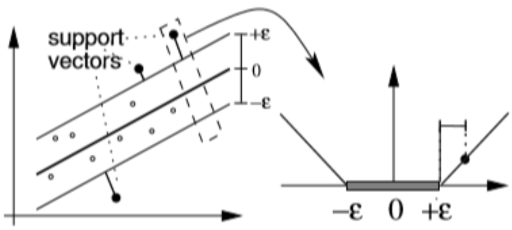


Figure 2 Linear SVM regression and the ε-insensitive loss function.

2.2. The combined forecasting model methods

Combining individual forecasting models is a well-established procedure for improving forecasting accuracy (Armstrong 2001; Makridakis and Hibon 2000; Thaithanan and Wongoutong 2020). Armstrong (2001) claimed that combined forecasting should be derived from methods that differ substantially and draw from different sources of information and when feasible, by using five or more of them.

A combined forecasting model method is associated with the performance consistency of each individual forecasting model and assigning combinatory weights. Consider $Y = [y_1, y_2, ..., y_N]^T$ as the actual time series to be forecasted using n different individual forecasting models, $\hat{Y}^{(i)} = [\hat{y}_1^{(i)}, \hat{y}_2^{(i)}, ..., \hat{y}_N^{(i)}]^T$ as the forecast obtained from the i^{th} model $(i = 1, 2, \ldots, n)$, and $\hat{Y}^{(c)} = [\hat{y}_1^{(c)}, \hat{y}_2^{(c)}, ..., \hat{y}_N^{(c)}]^T$ as the combined forecasted series of the original time series. In this study, two combined forecasting model methods were used to improve forecasting accuracy: simple-average and Bates-Granger.

2.2.1 The simple-average combined forecasting model method

This is used to assign equal weights to all of the individual forecasting models. Although it may appear to be a naïve approach for combining forecasts, more complex methods do not often improve its accuracy (Clemen and Hendry 1989). It is well-documented that using the simple-average approach provides a robust combination method that is difficult to beat (Stock and Watson 2004; Timmermann 2006). Assigning equal weights $(w_i = 1/n)$ to each of the individual forecasting models can be achieved as follows

$$\hat{y}_k^{(SA)} = w_1\hat{y}_k^{(1)} + w_2\hat{y}_k^{(2)} + \ldots + w_n\hat{y}_k^{(n)} = \sum_{i=1}^{n} w_i\hat{y}_k^{(i)} : w_i = 1/n, \tag{6}$$

where $\hat{y}_k^{(i)}(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, N)$ and $\hat{y}_k^{(c)}(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, N)$ denote the individual forecasts and the simple-average combination forecast, respectively.

2.2.2 The Bates-Granger combination model

Bates and Granger (Bates and Granger 1969) first showed that combining forecasts often leads to better forecasting accuracy. They combined forecasts by using several methods on the same time series and combining the resulting forecasts. They used the diagonal elements of the estimated mean-squared prediction error matrix to compute combination weights as follows

$w_i = \dfrac{\hat{\sigma}_{(i)}^{-2}}{\sum_{j=1}^{n}\hat{\sigma}_{(j)}^{-2}}$, where $\hat{\sigma}_{(i)}^{-2}$ is the estimated mean-squared prediction error of the i^{th} model. The combined forecast can then be obtained as

$$\hat{y}_k^{(BG)} = w_1\hat{y}_k^{(1)} + w_2\hat{y}_k^{(2)} + \ldots + w_n\hat{y}_k^{(n)} = \sum_{i=1}^{n} w_i\hat{y}_k^{(i)} : w_i = \hat{\sigma}_{(i)}^{-2} / \sum_{j=1}^{n}\hat{\sigma}_{(j)}^{-2}, \tag{7}$$

where $\hat{y}_k^{(i)}(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, N)$ and $\hat{y}_k^{(c)}(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, N)$ denote the individual forecasting models and the Bates-Granger forecasting combination method, respectively.

3. Methodology

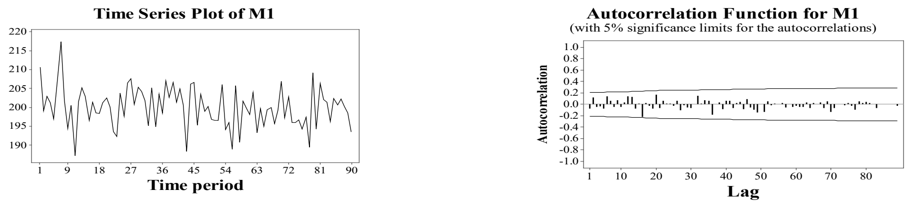
3.1. The data used in the study

The simulation of data with short, medium, and long stationary time series was used to clarify the performances of the combined forecasting model methods. The details of these datasets are reported in Table 1, while Figure 3 shows time-series and autocorrelation plots for some of them.

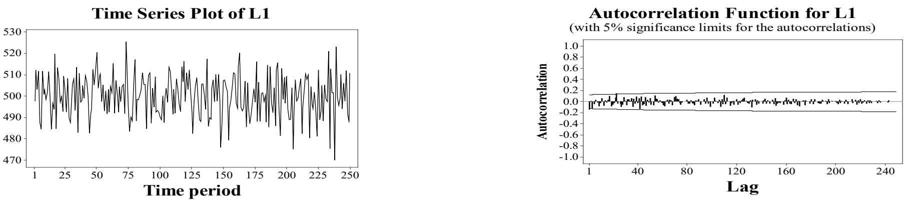
Table 1 The simulation of data with short, medium, and long stationary time series

Type	Dataset	Simulation Model	ε_t (the error term is normally distributed) μ	σ^2	Size
Short	S1	$Y_t = 100 + \varepsilon_t$	0	25	40
	S2	$Y_t = 20 + \varepsilon_t$	0	1	50
	S3	$Y_t = 2500 + \varepsilon_t$	0	100	50
	S4	$Y_t = 300 + \varepsilon_t$	0	4	40
	S5	$Y_t = 1500 + \varepsilon_t$	0	100	50
Medium	M1	$Y_t = 200 + \varepsilon_t$	0	25	90
	M2	$Y_t = 3000 + \varepsilon_t$	0	100	70
	M3	$Y_t = 300 + \varepsilon_t$	0	25	85
	M4	$Y_t = 250 + \varepsilon_t$	0	4	60
	M5	$Y_t = 30 + \varepsilon_t$	0	1	75
Long	L1	$Y_t = 500 + \varepsilon_t$	0	100	250
	L2	$Y_t = 200 + \varepsilon_t$	0	25	150
	L3	$Y_t = 1000 + \varepsilon_t$	0	100	250
	L4	$Y_t = 3000 + \varepsilon_t$	0	100	150
	L5	$Y_t = 30 + \varepsilon_t$	0	1	300

(a)



(b)



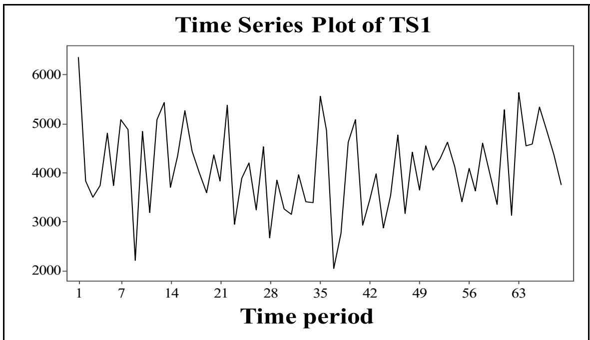
(c)

Figure 3 Time-series plots (above) and autocorrelation plots (below) of some of the simulated datasets: (a) S1, (b) M1, and (c) L1 (short, medium, and long stationary time series, respectively)

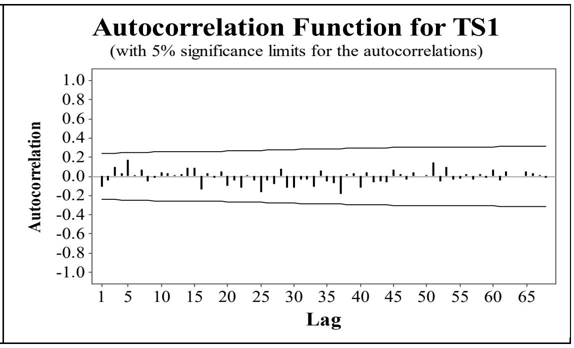
Ten real datasets comprising time series with a stationary pattern were used in this study to evaluate the performances of the combined forecasting models. Moreover, the augmented Dickey-Fuller test was used to check whether the process was stationary. All of the datasets were obtained from the M3 competition conducted by Makridakis and Hibon (2000). Brief details of these datasets are reported in Table 2, and time-series plots and autocorrelation plots for some of them are presented in Figure 4.

Table 2 Time-series data from the M3 Competition used to compare the combined forecasting model methods

Data	M3 Competition Code	Time Period	Size	Mean	SD
TS1	N1449 (Q)	1990Q1-2007Q1	69	4083.84	864.65
TS2	N1442 (Q)	1990Q1-2007Q1	69	5671.01	1466.48
TS3	N1472 (Q)	1990Q1-2007Q1	69	6340.70	657.65
TS4	N234 (Y)	1950-1993	44	6661.50	1070.99
TS5	N1447 (Q)	1990Q1-2007Q1	69	4581.45	849.19
TS6	N1987 (M)	1979M1-1990M12	144	5093.10	488.67
TS7	N217 (Y)	1947-1993	47	6418.40	1251.64
TS8	N2521(M)	1962M1-1973M12	144	5807.64	1700.30
TS9	N2172(M)	1982M1-1993M12	144	4993.23	479.80
TS10	N2125(M)	1978M1-1989M12	144	5677.08	1252.80



(a)



(b)

Figure 4 Real dataset of TS1: (a) Time-series plots, (b) the autocorrelation plots

3.2. The measures for forecasting accuracy

Two of the most frequently used error indices, root-mean-squared error (RMSE) and mean-absolute-percentage error (MAPE) (Clement and Hendry 1998; Wongoutong 2021) were used in the present study. The MAPE is a relative error measure using absolute values that can be used to compare the forecasting accuracy when using differently scaled time-series data. Meanwhile, RMSE is an absolute error measure by using the square of the deviation that can prevent positive and negative deviation values from canceling each other out. MAPE and RMSE are respectively defined as follows:

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left(\left|\frac{y_t - \hat{y}_t}{y_t}\right|\right), \quad \text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}},$$

where y_t and $\hat{y}_t$ are the true and predicted values at time t and n is the number of data points.

3.3. The steps in the experimental study

The steps used in this study illustrated as a flow chart in Figure 5 were as follows:

Step 1. Create the forecasting values with the five individual forecasting models for each time-series dataset (15 simulated and 10 real datasets).

Step 2. Compute the accuracy as the RMSE and MAPE values for each time-series dataset using the following methods: simple moving average, single exponential smoothing, Box-Jenkins, ANN, and SVM.

Step 3. Compute the correlation coefficient between the actual and the forecasted values for each model: $r(i) = \frac{\sum_{k=1}^{n}(y_{ik} - \overline{y})(\hat{y}_{ik} - \overline{\hat{y}})}{\sqrt{\sum_{k=1}^{n}(y_{ik} - \overline{y})^2 \sum_{k=1}^{n}(\hat{y}_{ik} - \overline{\hat{y}})^2}}$; $i = 1, 2, 3, 4, 5$ and $k = 1, 2, 3, \ldots, n$, where $r(i)$ is the correlation coefficient value between the actual and forecasted values for the i^{th} model and y_{ik} and $\hat{y}_{ik}$ refer to the actual and forecasted values for the i^{th} model, respectively.

Step 4. Rank the $r(i)$ values from smallest to largest and order them as $r(1) < r(2) < r(3) < r(4) < r(5)$.

Step 5. Create four groups by ranking order of their correlation coefficient values.

Group1 (5M)	Group2 (4M)	Group3 (3M)	Group4 (2M)
Forecasted values from five individual models corresponding to $r(1) < r(2) < r(3) < r(4) < r(5)$	Forecasted values from four individual models corresponding to $r(2) < r(3) < r(4) < r(5)$	Forecasted values from three individual models corresponding to $r(3) < r(4) < r(5)$	Forecasted values from two individual models corresponding to $r(4) < r(5)$

Step 6. Apply the simple-average and the Bates-Granger combined forecasting models to the forecasted values from groups 1-4.

Step 7. Compute the performances of the two combined forecasting model methods as MAPE and RMSE values for each time-series dataset and each group.

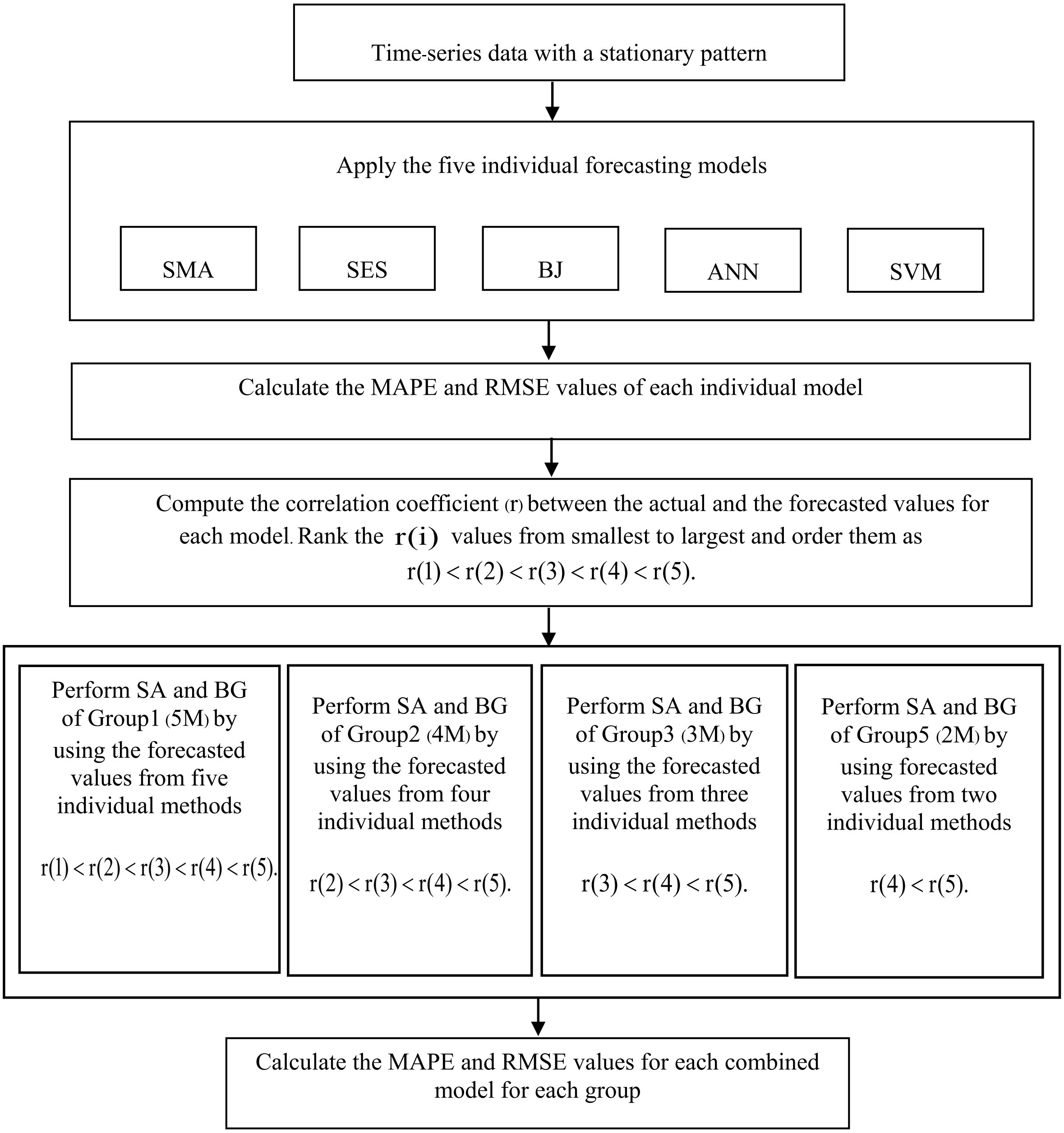


Figure 5 A flow chart of the experimental study with simulated and real time-series datasets with stationary patterns. SMA, simple moving average; DMA, double moving average; SES, single exponential smoothing; BJ, Box-Jenkins; SVM, support vector machine; ANN, artificial neural network; SA, simple-average; BG, Bates-Granger

4. Results and Discussion

4.1. Comparison of the combined forecasting model performances with the simulated datasets

The simulation time-series datasets with short (S1-S5), medium (M1-M5), and long (L1-L5) time series were treated with five (5M), four (4M), three (3M), and two (2M) individual forecasting models, which was followed by combining them with the simple-average and Bates-Granger models. The

MAPE and RMSE values for forecasting with the two combined forecasting models are reported in Tables 3 and 4, respectively.

For example, for the S1 short stationary time-series dataset when using 5M, 4M, 3M, and 2M, the combined forecasting simple-average model produced RMSE values of 3.365, 3.096, 2.707, and 2.098, respectively, whereas the combined forecasting Bates-Granger model produced RMSE values of 1.744, 1.581, 1.477, and 1.472, respectively (Table 3). Similarly, under the same experimental conditions, the combined forecasting simple-average model produced MAPE values of 2.823, 2.542, 2.175, and 1.608, respectively, whereas the combined forecasting Bates-Granger model produced MAPE values of 1.332, 1.167, 1.133, and 1.037, respectively (Table 4). These results indicate that the simple-average and Bates-Granger combined forecasting model methods performed the best with 2M, and the latter outperformed the former. This was the case for all of the short, medium, and long stationary time-series datasets shown in Figures. 6, 7, and 8, respectively. These figures offer a clear visual demonstration of the superiority of the simple-average and Bates-Granger combined forecasting models when selecting two vital individual models over five, four, or three of them for all three time-series scenarios.

Table 3 The RMSE results for the simple-average and Bates-Granger combined forecasting models with the simulated datasets

Dataset	Simple-Average				Bates-Granger			
	5M	4M	3M	2M	5M	4M	3M	2M
S1	3.365	3.096	2.707	2.098	1.744	1.581	1.477	<u>1.472</u>
S2	0.780	0.731	0.683	0.601	0.709	0.674	0.635	<u>0.596</u>
S3	8.290	7.854	7.417	6.768	7.439	7.100	6.812	<u>6.416</u>
S4	1.408	1.318	1.143	0.870	0.608	0.561	0.531	<u>0.507</u>
S5	3.873	3.679	3.450	2.964	3.243	3.066	2.918	<u>2.665</u>
M1	4.819	4.685	4.566	4.522	4.762	4.657	4.546	<u>4.514</u>
M2	8.541	8.295	7.728	7.280	8.046	7.786	7.458	<u>7.118</u>
M3	4.103	3.965	3.675	3.384	3.798	3.657	3.480	<u>3.286</u>
M4	1.612	1.557	1.489	1.404	1.537	1.479	1.415	<u>1.379</u>
M5	0.791	0.757	0.705	0.651	0.730	0.697	0.666	<u>0.638</u>
L1	9.598	9.467	9.405	<u>9.360</u>	9.566	9.449	9.381	9.364
L2	4.509	4.449	4.346	4.223	4.464	4.400	4.321	<u>4.213</u>
L3	9.732	9.675	9.533	9.519	9.701	9.637	9.528	<u>9.519</u>
L4	9.521	9.415	9.190	9.031	9.438	9.322	9.143	<u>9.004</u>
L5	0.976	0.969	0.952	0.932	0.969	0.960	0.946	<u>0.928</u>

Table 4 The MAPE results for the simple-average and Bates-Granger combined forecasting models with the simulated datasets

Dataset	Simple-Average				Bates-Granger			
	5M	4M	3M	2M	5M	4M	3M	2M
S1	2.823	2.542	2.175	1.608	1.332	1.167	1.133	1.037
S2	3.204	2.985	2.743	2.462	2.885	2.727	2.569	2.430
S3	0.258	0.246	0.227	0.205	0.232	0.222	0.210	0.197
S4	0.375	0.350	0.299	0.219	0.156	0.146	0.135	0.127
S5	0.399	0.379	0.334	0.303	0.355	0.333	0.310	0.303
M1	1.855	1.795	1.738	1.715	1.831	1.783	1.730	1.715
M2	0.225	0.219	0.204	0.195	0.211	0.204	0.197	0.188
M3	1.111	1.073	0.968	0.905	1.007	0.971	0.910	0.859
M4	0.524	0.506	0.481	0.452	0.496	0.476	0.453	0.443
M5	2.134	2.018	1.846	1.725	1.961	1.856	1.760	1.688
L1	1.563	1.537	1.516	1.512	1.557	1.534	1.517	1.510
L2	1.847	1.819	1.749	1.684	1.822	1.790	1.736	1.678
L3	0.782	0.776	0.760	0.753	0.779	0.772	0.759	0.753
L4	0.250	0.247	0.243	0.237	0.248	0.245	0.241	0.237
L5	2.651	2.634	2.567	2.528	2.631	2.608	2.552	2.516

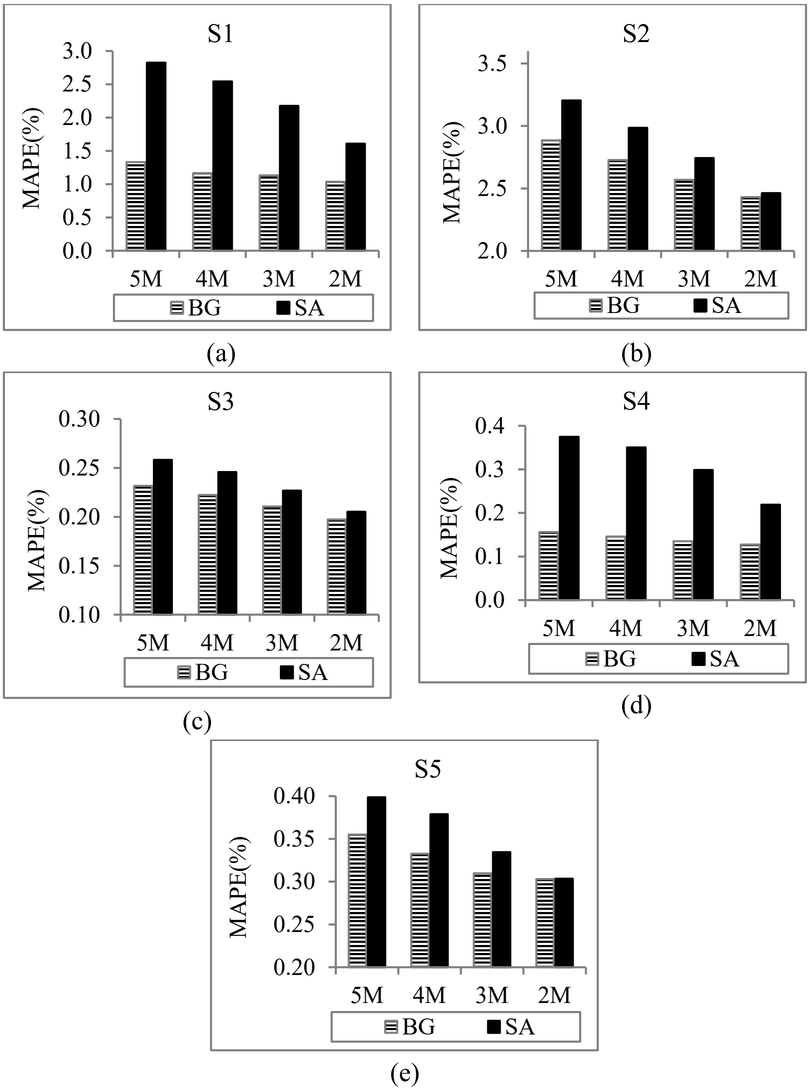


Figure 6 The MAPE results for the simple-average and Bates-Granger combined forecasting model methods when using five, four, three, or two individual forecasting models for simulated data with a short stationary time series (S1-S5); SA, simple-average; BG, Bates-Granger

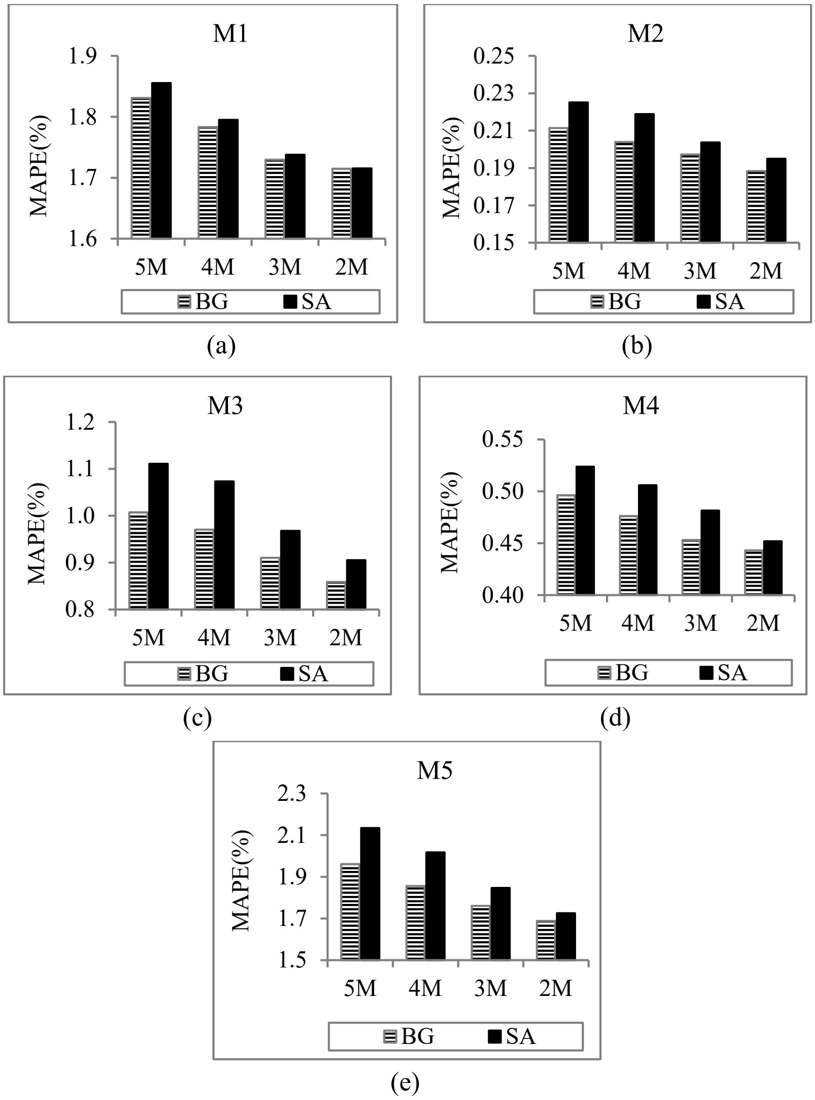


Figure 7 The MAPE results for the simple-average and Bates-Granger combined forecasting model methods when using five, four, three, or two individual forecasting models for simulated data with a medium stationary time series (M1-M5); SA, simple-average; BG, Bates-Granger

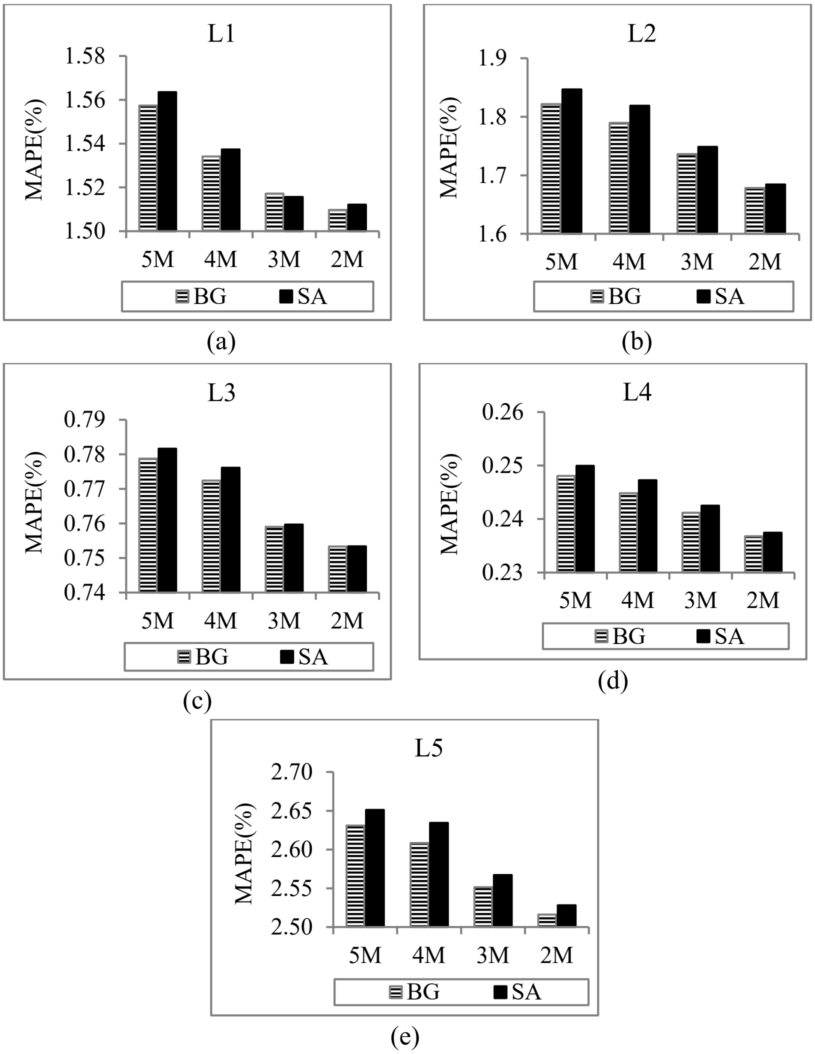


Figure 8 The MAPE results for the simple-average and Bates-Granger combined forecasting model methods when using five, four, three, or two individual forecasting models for simulated data with a long stationary time series (L1-L5); SA, simple-average; BG, Bates-Granger

4.2. Comparison of the combined forecasting model method performances with the real datasets

The MAPE and RMSE values for forecasting using real time-series datasets with a stationary pattern (TS1-TS10) are reported in Tables 5 and 6, respectively. When using the simple-average and Bates-Granger combined models with 5M, 4M, 3M, and 2M individual methods, their RMSE values followed a similar pattern to when using the simulated datasets; i.e., the RMSE and MAPE values decreased in the same direction.

For example, with the TS1 time-series dataset, the simple-average combined forecasting model when using 5M, 4M, 3M, and 2M produced RMSE values of 738.057, 725.913, 662.264, and 606.772, respectively, whereas the Bates-Granger combined forecasting model produced RMSE values of 686.591, 667.651, 633.861, and 600.742, respectively (Table 5). Similarly, the simple-average combined forecasting model produced MAPE values of 16.064, 15.446, 13.836, and 11.956, respectively, whereas the Bates-Granger combined forecasting model produced MAPE values of 14.683, 13.859, 12.903, and 11.326, respectively (Table 6). These results indicate the same trend as with the simulated datasets in that the combined forecasting models with 2M performed the best and the Bates-Granger combined forecasting model outperformed the simple-average combined forecasting model. Visualization of these results is shown in Figure. 9, which clearly demonstrates the superiority of the combined forecasting model methods using two vital individual models compared to five, four, or three. Thus, the results of using the novel method of selecting two vital individual models for the simple-average and Bates-Granger combined forecasting model methods were similar when using simulated and real time-series datasets.

Table 5 The RMSE results for the simple-average and Bates-Granger combined forecasting model methods with real data

Dataset	Simple-Average				Bates-Granger			
	5M	4M	3M	2M	5M	4M	3M	2M
TS1	738.057	725.913	662.264	606.772	686.591	667.651	633.861	600.742
TS2	1229.68	1202.11	1168.01	1107.85	1201.67	1175.67	1140.44	1103.50
TS3	523.900	486.565	438.405	360.711	358.725	335.912	309.345	288.002
TS4	677.394	645.164	600.152	552.037	608.177	584.737	552.251	530.328
TS5	658.128	620.953	577.746	512.872	585.829	553.599	528.313	497.379
TS6	375.814	361.641	343.239	324.513	358.863	347.200	333.871	321.203
TS7	889.666	825.437	738.908	606.021	634.590	585.590	536.197	498.657
TS8	1505.30	1454.77	1392.42	1325.21	1449.65	1404.55	1359.84	1305.35
TS9	370.731	360.216	347.709	332.523	353.990	345.126	333.749	322.539
TS10	1091.44	1074.50	1031.49	994.486	1054.34	1034.68	1003.57	972.973

Table 6 The MAPE results for the simple-average and Bates-Granger combined forecasting models of real data

Dataset	Simple-Average				Bates-Granger			
	5M	4M	3M	2M	5M	4M	3M	2M
TS1	16.064	15.446	13.836	11.956	14.683	13.859	12.903	11.326
TS2	18.771	18.083	17.583	16.054	18.160	17.501	16.801	15.975
TS3	6.533	5.992	5.284	4.232	4.305	3.940	3.525	3.437
TS4	8.024	7.568	7.029	6.688	7.138	6.877	6.606	6.594
TS5	12.080	11.116	9.805	8.521	10.341	9.606	8.916	8.639
TS6	5.883	5.599	5.229	5.071	5.591	5.353	5.094	5.053
TS7	12.940	12.202	11.007	8.602	9.127	8.429	7.537	6.748
TS8	26.468	25.373	23.705	22.205	25.182	24.189	22.982	21.555
TS9	5.974	5.756	5.510	5.165	5.683	5.499	5.272	5.033
TS10	16.524	16.169	15.564	14.962	15.934	15.566	15.172	14.705

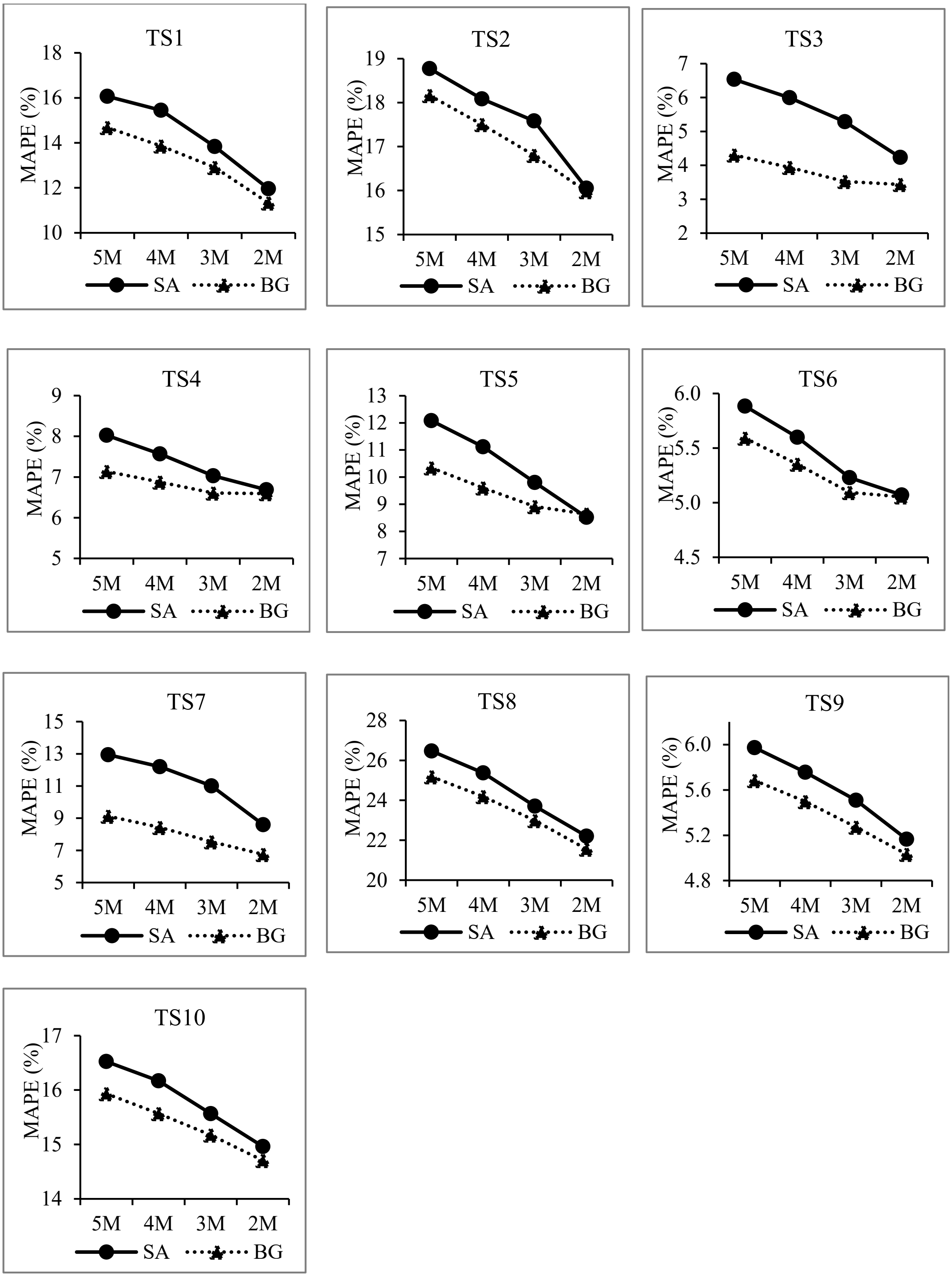


Figure 9 The MAPE results for the simple-average and Bates-Granger combined forecasting model methods when using five, four, three, or two individual forecasting models for real datasets TS1 to TS10; SA, simple-average; BG, Bates-Granger

The percentage improvements in terms of RMSE and MAPE when selecting two vital individual forecasting models compared to five, four, or three for the simple-average and Bates-Granger combined forecasting model methods with both simulated and real time-series datasets with a stationary pattern are reported in Tables 7 and 8, respectively. For simulated datasets S1–S5, using two vital individual forecasting models instead of five, four, or three improved the RMSE by 28.13%, 23.45%, and 16.25% with the simple-average combined method and 15.94%, 10.16%, and 5.10% with the Bates-Granger combined method, respectively. Similarly, for simulated datasets M1–M5, the improvements were 13.81%, 10.84%, and 5.61% with the simple-average combined method and 10.62%, 7.40%, and 3.51% with the Bates-Granger combined method, respectively, and for simulated datasets L1–L5, the improvements were 4.04%, 3.05%, and 1.46% with the simple-average combined method and 3.65%, 2.59%, and 1.24% with the Bates-Granger combined method, respectively. Finally, for real stationary time-series datasets (TS1-TS10), the improvements were 17.61%, 14.29%, and 8.67% with the simple-average combined model and 12.68%, 9.18%, and 4.64% with the Bates-Granger combined forecasting model method, respectively. Meanwhile, the results of the percentage improvements in MAPE followed the same trend as for RMSE.

It is once again evident that the novel approach of selecting two vital individual methods for the simple-average and Bates-Granger combined forecasting model methods was far superior to selecting five, four, or three, as shown in Figures 6, 7 and 8, respectively. Moreover, selecting two vital individual methods for the simple-average combined forecasting model for the short stationary time-series pattern improved the MAPE considerably more than for the medium and long size time-series patterns.

When comparing the improvements in MAPE by both combined forecasting methods with two vital individual methods, the plots in Figures. 10 and 11 for the Bates-Granger method are always under those of the simple-average method. These results clearly show that the latter considerably outperformed the former. Moreover, these results also support that the Bates-Granger combined forecasting model method when selecting two vital individual models is more efficient than selecting five, four, or three models.

Table 7 The percentage improvement in RMSE by the simple-average and Bates-Granger combined forecasting model methods with two individual methods compared with three, four, or five methods

Source	Dataset	Simple-Average with 2M			Bates-Granger with 2M		
		5M	4M	3M	5M	4M	3M
Simulation data	S1	37.65	32.24	22.50	15.60	6.89	0.34
	S2	22.95	17.78	12.01	15.94	11.57	6.14
	S3	18.36	13.83	8.75	13.75	9.63	5.81
	S4	38.21	33.99	23.88	16.61	9.63	4.52
	S5	23.47	19.43	14.09	17.82	13.08	8.67
	Average	**28.13**	**23.45**	**16.25**	**15.94**	**10.16**	**5.10**
	M1	6.16	3.48	0.96	5.21	3.07	0.70
	M2	14.76	12.24	5.80	11.53	8.58	4.56
	M3	17.52	14.65	7.92	13.48	10.14	5.57
	M4	12.90	9.83	5.71	10.28	6.76	2.54
	M5	17.70	14.00	7.66	12.60	8.46	4.20
	Average	**13.81**	**10.84**	**5.61**	**10.62**	**7.40**	**3.51**
	L1	2.01	0.65	0.48	1.93	0.72	0.18
	L2	6.34	5.08	2.83	5.62	4.25	2.50
	L3	2.19	1.61	0.15	1.88	1.22	0.09
	L4	5.15	4.08	1.73	4.60	3.41	1.52
	L5	4.51	3.82	2.10	4.23	3.33	1.90
	Average	**4.04**	**3.05**	**1.46**	**3.65**	**2.59**	**1.24**
Real data	TS1	17.79	16.41	8.38	12.50	10.02	5.22
	TS2	9.91	7.84	5.15	8.17	6.14	3.24
	TS3	31.15	25.87	17.72	19.72	14.26	6.90
	TS4	18.51	14.43	8.02	12.80	9.30	3.97
	TS5	22.07	17.41	11.23	15.10	10.16	5.86
	TS6	13.65	10.27	5.46	10.49	7.49	3.79
	TS7	31.88	26.58	17.98	21.42	14.85	7.00
	TS8	11.96	8.91	4.83	9.95	7.06	4.01
	TS9	10.31	7.69	4.37	8.88	6.54	3.36
	TS10	8.88	7.45	3.59	7.72	5.96	3.05
	Average	**17.61**	**14.29**	**8.67**	**12.68**	**9.18**	**4.64**

Table 8 The percentage improvement in MAPE by the simple-average and Bates-Granger combined forecasting model methods with two individual methods compared with three, four, or five methods

Source	Dataset	Simple-Average with 2M			Bates-Granger with 2M		
		5M	4M	3M	5M	4M	3M
Simulation data	S1	43.04	36.74	26.07	22.15	11.14	8.47
	S2	23.16	17.52	10.24	15.77	10.89	5.41
	S3	20.54	16.67	9.69	15.09	11.26	6.19
	S4	41.60	37.43	26.76	18.59	13.01	5.93
	S5	24.06	20.05	9.28	14.65	9.01	2.26
	Average	**30.48**	**25.68**	**16.41**	**17.25**	**11.06**	**5.65**
	M1	7.55	4.46	1.32	6.34	3.81	0.87
	M2	13.33	10.96	4.41	10.90	7.84	4.57
	M3	18.54	15.66	6.51	14.70	11.53	5.60
	M4	13.74	10.67	6.03	10.69	6.93	2.21
	M5	19.17	14.52	6.55	13.92	9.05	4.09
	Average	**14.47**	**11.25**	**4.96**	**11.31**	**7.83**	**3.47**
	L1	3.26	1.63	0.26	3.02	1.56	0.46
	L2	8.83	7.42	3.72	7.90	6.26	3.34
	L3	3.71	2.96	0.92	3.34	2.46	0.79
	L4	5.20	4.05	2.47	4.44	3.27	1.66
	L5	4.64	4.02	1.52	4.37	3.53	1.41
	Average	**5.13**	**4.02**	**1.78**	**4.61**	**3.42**	**1.53**
Real data	TS1	25.57	22.59	13.59	22.86	18.28	12.22
	TS2	14.47	11.22	8.70	12.03	8.72	4.92
	TS3	35.22	29.37	19.91	20.16	12.77	2.50
	TS4	16.65	11.63	4.85	7.62	4.12	0.18
	TS5	29.46	23.34	13.10	16.46	10.07	3.11
	TS6	13.80	9.43	3.02	9.62	5.60	0.80
	TS7	33.52	29.50	21.85	26.07	19.94	10.47
	TS8	16.11	12.49	6.33	14.40	10.89	6.21
	TS9	13.54	10.27	6.26	11.44	8.47	4.53
	TS10	9.45	7.46	3.87	7.71	5.53	3.08
	Average	**20.779**	**16.73**	**10.148**	**14.837**	**10.439**	**4.802**

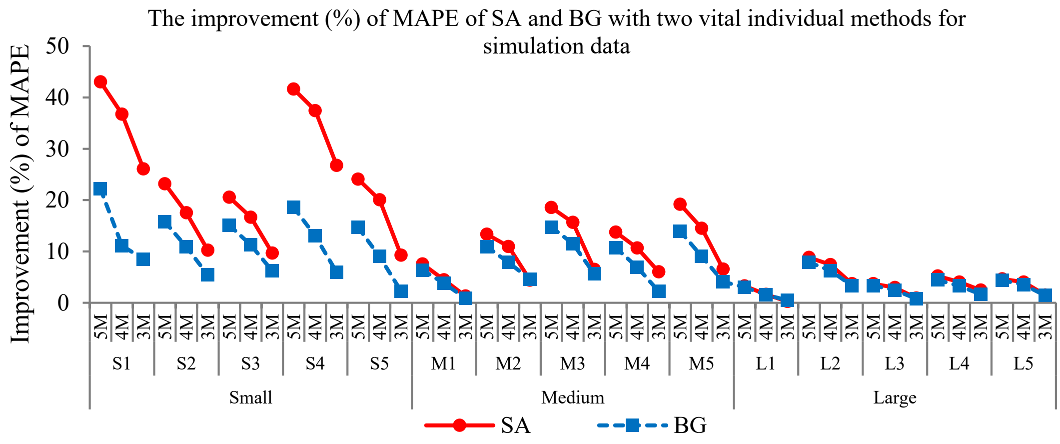


Figure 10 The percentage improvement in MAPE by the simple-average and Bates-Granger combined forecasting model methods with two individual methods compared with three, four, or five for the simulated data

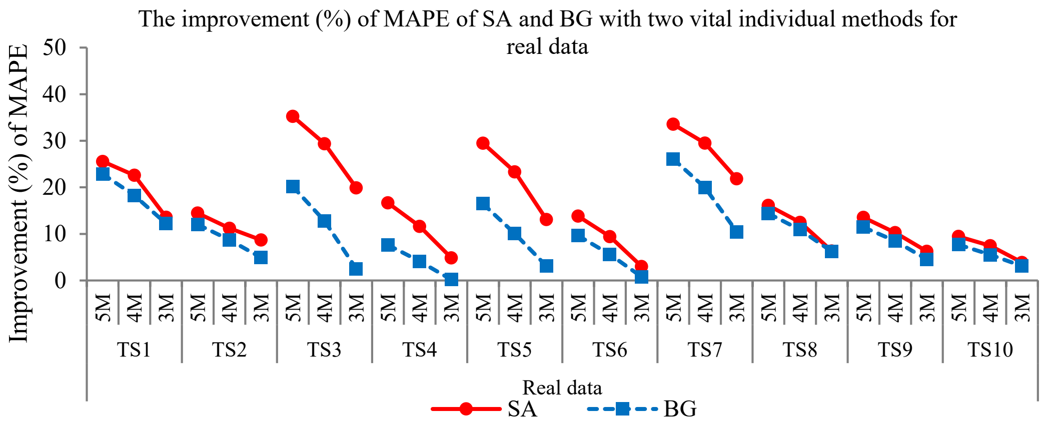


Figure 11. The percentage improvement in MAPE by the simple-average and Bates-Granger combined forecasting model methods with two individual methods compared with three, four, or five for real data

5. Conclusion

The significant discovery of this study is the approach to building a combined forecasting model method based on ranking of the individual forecasting models via correlation coefficients. The performances of the simple-average and Bates-Granger combined forecasting model methods with five, four, three, and two individual forecasting models were analyzed. For stationary time-series datasets of any length, the simple-average and Bates-Granger combined forecasting model methods using five, four, three, and two individual forecasting models performed equally well in terms of RMSE and MAPE. Moreover, their best performances were when using the novel approach of selecting the two most vital individual forecasting models (especially the Bates-Granger approach), thereby indicating that the approach in this study can effectively integrate the information from two individual forecasting models and produce accurate prediction results.

Acknowledgments

The authors are grateful for funding from the Thailand Institute of Scientific and Technological Research Institute (TISTR). In addition, the authors also would like to thank Kasetsart University and International SciKU Branding (ISB) for providing the facilities to conduct the research.

References

Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H. An empirical comparison of machine learning models for time series forecasting. Econom Rev. 2010; 29(5-6): 594-621.

Armstrong JS. Principles of forecasting: a handbook for researchers and practitioners. New York: Kluwer Academic Publishers; 2001.

Bates JM, Granger CW. The combination of forecasts. J Oper Res Soc.1969; 20(4): 451-468.

Box GE, Jenkins GM, Reinsel G. Time series analysis: forecasting and control. New Jersey: Prentice-Hall; 1994.

Box GE, Jenkins GM, Reinsel G. Forecasting and control. J Time Ser Anal. 1970; 3(75): 1970.

Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. New York: John Wiley & Sons; 2015.

Brown RG. Exponential smoothing for predicting demand. Cambridge: Arthur D. Little; 1956.

Clements M, Hendry D. Forecasting economic time series. Cambridge University Press; 1998.

Cortez P, Rocha M, Neves J. Evolving time series forecasting ARMA models. J Heuristics. 2004; 10: 415-429.

Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995; 20: 273-297.

De Menezes LM, Bunn DW, Taylor JW. Review of guidelines for the use of combined forecasts. Eur J Oper Res. 2000; 120(1): 190-204.

Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecast. 2004; 20(1): 5-10.

Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. Melbourne: OTexts; 2018.

Jose VR, Winkler RL. Simple robust averages of forecasts: Some empirical results. Int J Forecast. 2008; 24(1): 163-169.

Karatzoglou A, Hornik K, Smola A, Zeileis A. kernlab-an S4 package for kernel methods in R. J Stat Softw. 2004; 11(9): 1-20, https://doi.org/10.18637/jss.v011.i09.

Khandelwal I, Adhikari R, Verma G. Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. Procedia Comput Sci. 2015; 48: 173-179.

Kim KJ. Financial time series forecasting using support vector machines. Neurocomputing. 2003; 55(1-2): 307-319.

Lobo GJ. Analysis and comparison of financial analysts', time series, and combined forecasts of annual earnings. J Bus Res. 1992; 24(3): 269-280.

Makridakis S, Winkler RL. Averages of forecasts: Some empirical results. J Manag Sci. 1983; 29(9): 987-996.

Makridakis S, Hibon M. The M3-Competition: results, conclusions and implications. Int J Forecast. 2000; 16(4): 451-476.

Okasha MK. Using support vector machines in financial time series forecasting. Int J Stat Appl. 2014; 4(1): 28-39.

Sanders NR, Ritzman LP. Some empirical findings on short-term forecasting: technique complexity and combinations. Decis Sci. 1989; 20(3): 635-640.

Shi HY, Hwang SL, Lee KT, Lin CL. In-hospital mortality after traumatic brain injury surgery: a nationwide population-based comparison of mortality predictors used in artificial neural network and logistic regression models. J Neurosurg. 2013; 118(4): 746-752.

Svetunkov I, Petropoulos F. Old dog, new tricks: a modelling view of simple moving averages. Int J Prod Res. 2018; 56(18): 6034-6047.

Taylor JW, Bunn DW. Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: a simulation study. Int J Forecast 1999; 15(3): 325-339.

Thaithanan, J., Wongoutong, C. A combined forecasting model for predicting the number of road traffic accident deaths in Thailand. ADAS. 2020; 64(2): 143-163.

The R Foundation. The R project for statistical computing. 2020 [cited 2022 Feb 14]. Available from: http://www.r-project.org/

Timmermann A. Forecast combinations. Handb Econ Forecast. 2006; 1: 135-196.

Stock JH, Watson MW. Combination forecasts of output growth in a seven-country data set. J Forecast. 2004; 23(6): 405-430.

Wongoutong C. The Effect on forecasting accuracy of the holt-winters method when using the incorrect model on a non-stationary time series. Thail Stat. 2021; 19(3): 565-582.

Yang Y. Combining forecasting procedures: some theoretical results. Econ Theory. 2004; 20(1): 176-222.

Yolcu U, Egrioglu E, Aladag CH. A new linear & nonlinear artificial neural network model for time series forecasting. Decis Support Syst. 2013; 54(3): 1340-1347.

Xiao Z, Wu W. The application of combining forecasting based on PSO-PLS to GDP. J Manag Sci. 2008; 21(3): 115-122.

Zhao CY, Zhang HX, Zhang XY, Liu MC, Hu ZD, Fan BT. Application of support vector machine (SVM) for prediction toxic activity of different data sets. J Toxicol. 2006; 217(2-3): 105-119.