www.rericjournal.ait.ac.th

Using Weather Patterns to Forecast Electricity Consumption in Sri Lanka: An ARDL Approach

Priyadarshana A.D.A.D.*, [1], Lokupitiya R.S.*,
Kuruppuarachchi D.[+], and Lokupitiya E.[#]

ARTICLE INFO

Article history:
Received 30 October 2020
Received in revised form
22 March 2021
Accepted 28 March 2021

Keywords:
Autoregressive distributed lag model
Electricity consumption forecasting
Inverse distance weighted interpolation
Missing value imputation
Weather impact

ABSTRACT

It is crucial to plan the electricity supply to match the future demand since electricity has become a dominant utility. Sri Lanka as a developing country, has over 98% of households electrified, which sometimes suffer from interruptions in supply. This study aims at forecasting monthly electricity consumption in Sri Lanka by considering the influence of weather patterns. Rainfall, humidity, and temperature are the three main weather parameters found to affect the electricity demand. We compared eight forecasting approaches including four econometric models and four algorithmic forecasting methods in forecasting monthly electricity consumption. Twenty meteorological stations were considered to spatially interpolate the weather data using the Inverse Distance Weighted (IDW) interpolation method. Results revealed that Autoregressive Distributed Lag (ARDL) model which incorporates the weather patterns as predictors outperforms in forecasting the monthly electricity consumption compared with all other forecasting approaches.

1. INTRODUCTION

Sri Lanka is a tropical island in the Indian Ocean, subject to frequent fluctuations in weather patterns which include extreme weather events such as droughts. Such volatile weather patterns influence both electricity generation and consumption. However, less attention has been paid so far on forecasting the electricity demand in the country. On the other hand, to achieve sustainable development goals, there is a national plan to increase the use of renewable energy sources to produce electricity in the country. Nevertheless, continuous increasing electricity demand is evident due to trends in household and industrial consumptions [1]. The main source of the electricity generation in Sri Lanka has become the thermal sources which are up to 80 percent of the supply under extreme drought periods [2]. In principle, the electricity supply should be planned based on both the power generation capacity and the demand. Thus, the ability to forecast electricity demand is crucial to achieve a sustainable supply-demand balance.

The Ceylon Electricity Board (CEB) being the only generation and transmission utility in Sri Lanka, is responsible for the entire supply chain. The CEB is controlled centrally by a System Control Centre (SCC) which is the nerve center of operational activities. As alternating current (AC) electricity cannot be stored in real-time, the SCC needs to match the demand for electricity in the country at any given time with the exact amount of generation simultaneously at a pre-specified frequency tolerance of 50 Hz ±1% [3]. This "operation planning" process needs to be done in steps spanning from a week to several months in advance. One of the fundamental pre-requisites of operations planning is the forecasts for short-term demand. However, such a comprehensive forecasting mechanism is not available yet for SCC. Also, region-wise demand forecasting is not manipulated by the CEB at present although there is such a requirement. This study attempts to fill this gap.

More specifically, the objectives of this study are, 1), to assess the association between electricity demand and weather parameters (*i.e.* temperature, rainfall, and humidity), 2) to formulate a suitable demand forecasting model incorporating weather parameters, and 3) to evaluate the performance of the best model (against the other seven models considered) in forecasting electricity demand in Sri Lanka. Although region-wise demand plays a major role in planning activities of CEB, any approach for such forecasting approach cannot be seen within the current system. Therefore, in addition to the above mentioned objectives, we have also aimed at proposing a suitable approach for CEB in forecasting region-wise consumption.

The availability of a valid, efficient, and effective forecasting model for electricity loads is a successful way of reducing the need for new generation capacity, scheduling resources, and it could help improve the efficiency in electricity end-use. In a broader economic

*Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka.

[+]Department of Accountancy and Finance, University of Otago, Dunedin, New Zealand.

[#]Department of Zoology and Environment Sciences, Faculty of Science, University of Colombo, Sri Lanka.

[1]Corresponding author:
Email: anupriyadarsh@gmail.com

sense, such models help in achieving a more efficient allocation of resources. Unlike long-term expansion planning where the demand growth over a 20-year time horizon is driven by external factors such as economic growth, population, tariff, etc., the short-term forecasting is primarily driven by weather and short-term social factors such as holidays and festivals [4]. Along with the impact of the demand from higher living standards, weather conditions also affect the short-term electricity demand due to the use of electric appliances such as fans and air conditioners during warm periods. At present, system operators (system control engineers) perform the short-term demand forecasting merely using time trend analysis and with their experience which is inefficient [1]. Intra-day forecasting (which is done for shorter time steps of hours) is primarily done using historical hourly demand curves whereas weekly forecasting is done considering past data (in daily time steps). However, at the SCC, monthly forecasts (in monthly time steps) are computed based only on the time trend analysis. Absence of seasonal weather patterns in monthly forecasts caused several power cuts during the last decade in the country. In fact, there were 12 power cuts imposed by the CEB during 1969-2012 due to the inadequate supply [5]. Even though weather fluctuations and extreme weather conditions could have a significant impact on electricity demand and consumption, this fact has not been accounted for the current forecasting system. Thus, this study contributes to a significant improvement in the demand-forecasting mechanism of the SCC.

Many studies on forecasting electricity consumption using the impact of weather have been conducted internationally. Such studies have brought out the importance of considering the impact of weather fluctuations on electricity consumption forecasts. For example, in addition to economic and national factors, electricity demand in Greece is found to be dependent on seasonal changes revealing peaks of the demand in winter and summer [6]. Moreover, the electricity consumption in England and Whales has shown a significant sensitivity to temperature during seasons. In addition to the weather-related factors, it has been found that economic factors such as gross domestic product and population growth also affect the electricity demand [7]. Alabbas and Nyangon also forecasted the electricity demand using weather conditions in Saudi Arabia up to 2040 [8]. This study has emphasized the importance of the awareness of climate change on future heating and cooling equipment investments. In an electricity consumption analysis done for the city of Kragujevac in Serbia by Jovanović, it has been pointed out that the mean daily air temperature is the most influential meteorological factor on the additional consumption of electrical power [9]. The same study found that geographic location and the climate zone affect electric power consumption.

Lou and Dong have observed that temperature and humidity have strong correlations with the power load using a correlation analysis [10]. There are many external factors such as climatic conditions, number of users, time/s of the year (*e.g.* holidays and weekends, etc.), and production activities affecting the electrical load fluctuation [11]. Among the various electricity load forecasting techniques used in the past decade, the best-known model is considered as the univariate models [12]. Among those models, autoregressive integrated moving average (ARIMA) models tend to give favorable results [13], while exponential smoothing models [14], Bayesian models [15], Kalman filter models [16], and regression models [17] also have been widely used in forecasting load consumption. These models do not consider any external factors such as climate conditions and they deal with historical electrical loads in forecasting [12].

In particular, artificial neural networks (ANNs) [18], [19], knowledge-based system (KBS) models [20], [21], fuzzy inference models [10], [22], and kernel-based models [23] have been employed in developing electricity load forecasting models based on artificial intelligence. Also, in short term load forecasting, hybrid models [24]-[26], and combined models [19], [27] have become more popular in recent years. Moreover, the support vector regression (SVR) model has been extensively used in non-linear modeling of electric load forecasting [26], [28].

On the other hand, many researchers have used temperature as one of the crucial factors in their forecasting models [17]. Engle *et al*. [29] analyzed the relationship between electricity usage and temperature using partially linear models. Chikobvu and Sigauke have assessed the temperature impact on daily peak demand in a study based on South Africa [30]. The results of this study have demonstrated that electricity demand is sensitive to cold weather in the country.

Lebotsa *et al*. has shown the improvement of short load forecast accuracy due to inclusion of temperature variables in addition to the impact of time factors [17]. The relationship between the load demand and the explanatory variables such as calendar effects, temperature effects, and lagged load observations has been investigated. Aggregated coastal and inland temperature variables have been considered in their paper. Furthermore, it is emphasized that selection of weather stations is an important aspect in forecasting load demand since weather effects have a major impact on load demand. Also, due to the presence of weather conditions and economic factors collectively, the electricity demand pattern becomes complex [31], [32].

Although the above mentioned weather incorporated models have considered the past electricity consumption to gain a significant forecasting accuracy, the attention has not been paid on the past weather parameters for the forecasts. In contrast, the proposed ARDL approach in this study significantly contributes by using historical weather parameters along with consumer accounts as a proxy for the demand. Moreover, the current study introduces region-wise forecasting for the country. Instead of using aggregated weather information that has been used in the past [17], the current study employs spatial interpolation techniques to derive weather parameters to cover the entire country that overcome the uneven spread of weather stations.

The rest of this paper is organized as follows. Section 2 presents the materials and methods. Section 3 demonstrates the results. Section 4 discusses findings and section 5 concludes.

2. MATERIALS AND METHODS

The following steps demonstrate the proposed methodology in this paper. These steps are summarised in Figure 1.

Step 1: Data preparation for monthly electricity consumption of CEB areas for the period of January 2006 to June 2017.

Step 2: Aggregate area-wise monthly electricity consumption to get the total monthly electricity consumption in the country. Divide it into training and test sets. We considered January 2006 to December 2015 to train the model for all forecasting methods and use January 2016 to June 2017 data as the test data set.

Step 3: Train the models employing Classical, Exponential smoothing, and Box-Jenkins methods as univariate time series.

Step 4: Forecasting for the test period using the trained models.

Step 5: Compile monthly weather data of 20 meteorological stations around the country.

Step 6: Missing value imputation of weather parameters. We used Seadec algorithm in R for this step.

Step 7: Divide the country into sub-regions such that each region would possess similar weather conditions.

Step 8: Imputation of location-specific weather parameters from IDW interpolation for the Sri Lanka grid. As representative locations map centroids of the each and every district and CEB area using ArcMap 10.3 version.

Step 9: Exploratory data analysis and explore the relationship between electricity consumption and weather factors of each district.

Step 10: Train the ARDL models for each and every sub-region predetermined in Step 7.

Step 11: Forecasting the monthly electricity consumption based on the estimated ARDL models. Aggregate the forecasts of the sub-regions and obtain the forecasted total monthly electricity consumption of the country.

Step 12: Model evaluation and verification of the superiority of the proposed ARDL approach with other estimated models.

The CEB has declared 62 distribution areas covering the entire country. Representing all these areas, monthly data were obtained over the period from January 2006 to June 2017 covering tariff-wise consumption and consumer accounts data from the Information Technology Branch of CEB. Moreover, minimum and maximum of air temperature, minimum and maximum of relative humidity, and rainfall data were retrieved monthly basis from the Meteorology Department and the Department of Census and Statistics representing 20 meteorological stations around the country along with their latitude and longitude information. This study employed a battery of econometric forecasting methods for the monthly electricity consumption data as follows:

1. Classical approach of time series analysis
2. Exponential smoothing
3. Box-Jenkins method
4. Autoregressive Distributed Lag (ARDL) modeling

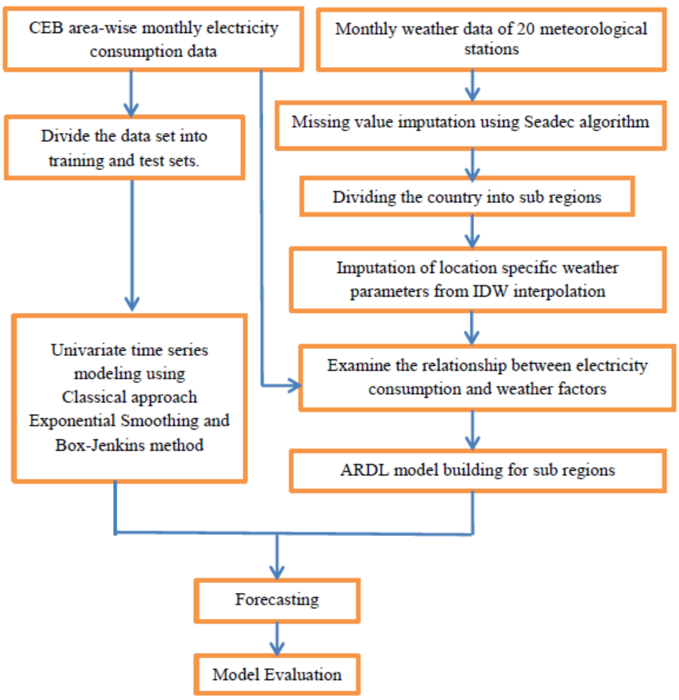

Fig. 1. Flowchart of the proposed forecasting algorithm.

2.1 Classical Approach of Time Series Analysis

In this study, additive and multiplicative models were used. In the additive model, the decomposition of time series was done based on the assumption that the effects of various components are additive in nature and are independent of each other. If the size of a seasonal effect does not change with the trend, then the additive model is appropriate. The additive model specifies the electricity consumption in the given month t (Y_t) based on the trend value (T_t) seasonal variation (S_t) and the irregular variation (I_t) in the same month as in Equation 1.

$$Y_t = T_t + S_t + I_t \tag{1}$$

Alternatively, in the multiplicative model, the decomposition of time series was done as a product of the components assuming that the effects of the components are not necessarily independent of each other. If the size of the seasonal effect is directly proportional to the trend then the multiplicative model is appropriate. It treats the time series values of electricity consumption as the product of these three components, T_t, S_t, and I_t as illustrated in Equation 2 [33];

$$Y_t = T_t \times S_t \times I_t \tag{2}$$

2.2 Exponential Smoothing Method

Exponential smoothing method is popular in producing smoothed time series forecasts. In particular, Exponential smoothing method assigns exponentially decreasing weights on the past data as the observation gets older. That is, the recent observations are given relatively more importance than the older observations for forecasting. There are three types of exponential smoothing procedures: single exponential smoothing, double exponential smoothing, and triple exponential smoothing. Triple exponential smoothing method is most useful when the seasonal component and the trend component are changing at different paces over time which we have used in this study as in Equation 3.

$$Y_{t+m} = (S_t + mT_t)I_{t-L+m} \tag{3}$$

Here, Y_{t+m} denotes forecasted future electricity consumption value m months ahead from month t. Furthermore, S_t, T_t, and I_t represent overall smoothing, trend smoothing, and seasonal smoothing coefficients respectively while L represents the lag length [33].

2.3 Box-Jenkins Method

Box-Jenkins method includes autoregressive moving average (ARMA) models under stationary conditions. Moreover, in the presence of a seasonal effect, ARIMA model can be extended into a seasonal ARIMA model (SARIMA). A $SARIMA\ (p, d, q)\ (P,D,Q)_S$, model is illustrated in Equation 4a where p, d, and q are the non-seasonal autoregressive, integration, and moving average orders respectively. Furthermore, P, D, Q are the seasonal autoregressive, integration, and moving average orders respectively [34].

$$\theta_p(B)\Theta_P(B^S)(1-B)^d(1-B^S)^D X_t = \phi_q(B)\varphi_Q(B^S)Z_t \tag{4a}$$

Here, B is the backshift operator, S is seasonal length, and $Z_t \sim WN\ (0, \sigma^2)$ while;

$$\theta_p(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p \tag{4b}$$

$$\Theta_P(B^S) = 1 - \alpha_1{}' B^{1S} - \alpha_2{}' B^{2S} - \ldots - \alpha_P{}' B^{PS} \tag{4c}$$

$$\phi_q(B) = 1 + \beta_1 B + \beta_2 B^2 + \ldots + \beta_q B^q \tag{4d}$$

$$\varphi_Q(B^S) = 1 + \beta_1{}' B^{1S} + \beta_2{}' B^{2S} + \ldots + \beta_Q{}' B^{QS} \tag{4e}$$

2.4 ARDL Modelling Approach

ARDL models consist of lags of both the dependent variable and explanatory variables as regressors. Specifically, an $ARDL\ (p,q_1,\ldots,q_k)$ can be specified such that p is the autoregressive order and q_j is the lag order of the j th explanatory variable such that $j=1,2,3\ldots,k$ for k number of explanatory variables. The corresponding ARDL model is illustrated in Equation 5.

$$Y_t = \alpha + \sum_{i=1}^{p} \gamma_i Y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q_j} X_{j,t-i} \beta_{j,i} + \varepsilon_t \tag{5}$$

Here, α is a constant term while γ_i and $\beta_{j,i}$ are the coefficients associated with autoregressive and distributed lagged terms respectively[35].

Apart from the above-mentioned econometric models, we adopt four popular algorithms commonly used in the literature namely, ANN [36], [37], SVR [28], SVR hybridized with linear regression (SVR-LR) [38], and dynamic harmonic regression (DHR) [39]. This helps us to make our comparison not only to econometric model-based forecasting but also to algorithm-based forecasting.

For all the above-discussed forecast approaches except the ARDL and the ANN model, monthly electricity consumption in Sri Lanka was used as a univariate time series whereas in ARDL framework and ANN, weather factors were used as predictor variables. Minimums, maximums, and averages of air temperature, relative humidity, and rainfall were employed as weather-related factors. In addition, the number of consumer accounts was used as a demand-related predictor. Due to frequent technical failures, there were a significant number of missing values in weather variables and hence, missing value imputation was done using Seasonally Decomposed Missing Value Imputation (Seadec) [2] approach. Seadec approach initially removes the seasonal component from the time series, then performs imputation on the trend and

[2] Seadec approach among other methods was used due to its outstanding performance which was evident from the preliminary analysis. These results are not reported but are available upon request.

irregular components and finally adds the seasonal component again [40].

2.5. Model Building for Sub Regions in the Country

Sri Lanka as a tropical country has been traditionally divided into three main climatic zones, namely wet, dry and intermediate zones. Thus, the weather variation is not uniform over the country. Hence unlike in classical, exponential smoothing, and Box-Jenkins forecasting approaches, the total monthly electricity consumption of the country cannot be taken as the response variable when the weather factors are included in ARDL approach. Therefore, the need for dividing the country into sub-regions such that each region would possess similar weather conditions was arisen [6].

Ceylon Electricity Board has divided the country into 62 zones for the convenience of its functions and these zones are called "CEB areas". The climatic zones in the country and the CEB zones mismatch each other. To overcome this limitation, corresponding districts of the CEB zones were identified first. Among these 62 CEB areas, 59 could be assigned to 20 districts by matching CEB area boundaries with district boundaries and the remaining 3 CEB areas were considered separately. Then we assigned corresponding values of weather factors with respect to each district's centroid. This resulted 20 districts and 3 CEB areas which were used for monthly electricity consumption forecasting.
Apparently, location-specific weather parameters were imputed on each and every district and CEB area. As representative locations, centroids of the districts were mapped using ArcMap 10.3 Version. The conversion of longitude and latitude from WGS84, the reference system being used by the Global Positioning System into the current coordinate system used by Survey Department of Sri Lanka (SLD99) [41], was done for the 20 meteorological stations. Due to sparseness of the meteorological stations, distance-based IDW interpolation [42] was employed in estimating the weather parameters at district centroids.

2.6 Measurement of Forecasting Performance

Mean absolute percentage error (MAPE) was used to measure the accuracy of the forecasts obtained from different methods. This measure averages the absolute distance between each pair of actual and forecast values and provides the error as a percentage. Consequently, MAPE equation has managerial appeal and became popular in forecasting.

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|\frac{Y_t-\hat{Y}_t}{Y_t}\right|\times 100}{n} \tag{6}$$

Here, Y_t is the actual value, and $\hat{Y}_t$ is the forecast value at time t while n is the number of observations. For the robustness check, alternative forecasting performance measures such as mean absolute error (MAE) and root mean squared error (RMSE) were used in this study with the first preference on MAPE due to its superiority as a percentage error and its extensive usage in the related literature [24], [43]. MAE and RMSE are expressed as shown in the following equations.

$$\text{MAE} = \frac{\sum_{t=1}^{n}\left|Y_t-\hat{Y}_t\right|}{n} \tag{7}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}\left(Y_t-\hat{Y}\right)^2}{n}} \tag{8}$$

2.7 Tests of Significance of Superiority of Forecasting Performance

The Wilcoxon signed-rank test and the Friedman test are two useful statistical significance tests that can be used to verify the enhanced forecasting performance [44], [45]. Wilcoxon signed-ranks test is a simple, but robust nonparametric test for pairwise statistical comparisons [46].

The central tendencies between two data sets of the same size can be compared by employing the Wilcoxon signed-rank test [44]. Differences between forecasts of any two forecasting models are ranked ignoring the sign. If the difference between the pair equals to zero, that pair is eliminated, and the sample size is reduced accordingly. Then each rank is labeled with the corresponding sign of the difference. Let r^+ be the sum of the ranks of the positive differences, and r^-, the sum of the ranks of the negative differences. Thus, the test statistic for the Wilcoxon signed-rank test, W is computed as follows.

$$W = \min\{r^+, r^-\} \tag{9}$$

The Friedman test is the best-known procedure for testing the differences between more than two related samples [46]. This non-parametric statistical test can be used to identify significant differences in the forecasting errors of more than two models [12]. The equality of the medians of the forecasting errors of all models of interest is the null hypothesis of this test. The test statistic of the Friedman test, F can be calculated by Equations 10 and 11 [46].

$$F = \frac{12N}{m(m+1)}\left[\sum_{j=1}^{m} R_j^2 - \frac{m(m+1)^2}{4}\right] \tag{10}$$

$$R_j = \frac{1}{N}\sum_{i=1}^{N} s_i^j \tag{11}$$

Here, N denotes the total number of forecasting errors; m is the total number of compared models; and R_j denotes the average of the number of each rank for each forecasting model where s_i^j is the number of rank from the j th compared model at the i th point forecast error. [44].

3. RESULTS

Figure 2 illustrates the interpolation for average relative humidity in January 2006 as an example. Similarly, all weather factors were interpolated over the sampled period on monthly basis. After identifying the relevant grid points, for each variable in each month, the imputed values were gathered and a summary of those values is illustrated in Table 1 for the average temperature. It indicates the variation in average temperature over the country by districts such that the minimum temperature was recorded by Nuwara Eliya district while the maximum was recorded by Colombo district.

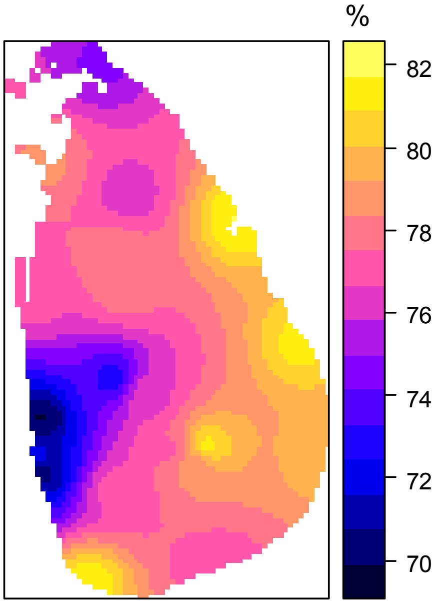

Fig. 2. IDW interpolation for average relative humidity (percentage) in January 2006.

Table 1. Descriptive statistics (in ℃) of "interpolated average temperature" by district for January over the study period (January 2006 – June 2017).

District	Minimum	Maximum	Mean	Standard Deviation
Ampara	23.262	24.992	24.036	0.394
Anuradhapura	25.009	26.456	25.792	0.419
Badulla	20.181	21.584	20.868	0.458
Batticaloa	24.410	26.456	25.448	0.549
Colombo	26.392	28.128	27.058	0.440
Galle	25.561	27.180	26.195	0.418
Gampaha	25.975	27.771	26.777	0.440
Hambantota	25.761	27.647	26.513	0.560
Jaffna	25.011	25.970	25.527	0.279
Kalutara	25.433	27.275	26.284	0.442
Kandy	22.490	24.724	23.325	0.608
Kegalle	23.518	25.496	24.411	0.458
Kilinochchi	25.149	26.067	25.585	0.271
Kurunegala	24.315	26.608	25.490	0.533
Mannar	25.629	26.877	26.215	0.349
Matale	23.214	25.221	24.119	0.473
Matara	24.866	26.534	25.545	0.414
Moneragala	20.770	22.249	21.385	0.361
Mullaittivu	25.182	26.146	25.592	0.319
Nuwara Eliya	14.486	16.010	15.150	0.439
Polonnaruwa	24.104	25.673	24.819	0.383
Puttalam	25.151	26.583	25.759	0.366
Ratnapura	23.791	25.686	24.781	0.481
Trincomalee	25.593	26.979	26.174	0.396
Vavuniya	24.869	26.293	25.404	0.411

Then the relationship between electricity consumption and weather factors in Table 2 was examined. The null hypothesis that there is no significant linear correlation between the two variables was tested using the corresponding p-value such where smaller the p-value the higher the evidence against the null hypothesis [47]. A 5% level of significance was used in concluding. Results indicated the existence of significant linear correlations justifying the relevance of the proposed ARDL framework in electricity demand forecasting.

Table 2. Correlations between weather variables and monthly electricity consumption.

CEB Area[a]	Weather Variable	Pearson Correlation Coefficient	p value
Kurunegala	RH_avg[b]	-0.271	0.007
	Temp_min[b]	0.289	0.002
	Temp_max[b]	0.218	0.021
	Temp_avg[b]	0.332	0.000
Ratmalana	Temp_min	0.201	0.035
	Temp_max	0.233	0.014
	Temp_avg	0.284	0.002
Ratnapura	Temp_min	0.201	0.034
Trinco	Temp_max	0.222	0.027
Vavuniya	RH_avg	-0.386	0.000
Jaffna	RH_max[b]	-0.609	0.000
	Temp_max	0.197	0.040
Hambantota	Temp_min	0.298	0.002
Colombo	RH_min[b]	0.208	0.039
	Temp_min	0.443	0.000
	Temp_max	0.252	0.008
	Temp_avg	0.487	0.000

[a]Reported only significant correlations at 5% level
[b]RH_avg = average relative humidity; RH_max = maximum relative humidity; RH_min = minimum relative humidity; Temp_min = minimum temperature; Temp_max = maximum temperature ; Temp_avg = average temperature.

Jaffna CEB area had a strong negative linear relationship with maximum relative humidity while electricity consumption in Colombo CEB area depends on the temperature. Overall, temperature appeared to be positively correlated with the electricity demand while the humidity affects negatively. It was evident from Table 2 that the temperature seemed to be the prominent weather factor affecting electricity consumption in the country. Moreover, Figure 3 illustrates the relationship between temperature and electricity consumption in Colombo district as an example. It was evident that the electricity consumption depends on the monsoons.

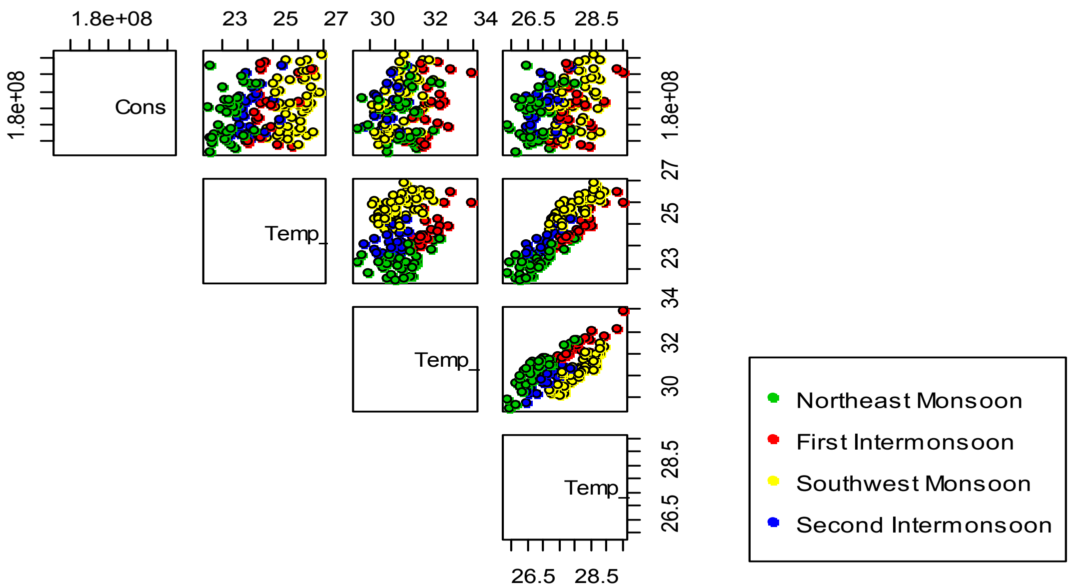

Fig. 3. Correlation matrix between electricity consumption and temperature by monsoon types for Colombo district in January 2006.

Next, location-specific ARDL models for each region were estimated by incorporating the above-mentioned location-wise weather factors. Since weather factors are seasonal, missing values were imputed using "Seadec" algorithm in R. The total forecast of consumption was obtained by aggregating those location-specific forecasts. Table 3 presents the estimated ARDL model for forecasting monthly electricity consumption in Matale district as an example.

Table 3. Estimated ARDL model for Matale district's electricity consumption forecasting.

Model: ARDL (2, 0, 0, 2)

Variable	Coefficient	Std. Error	t-Statistic	Prob
$Cons_{t-1}$ [a]	0.341	0.087	3.916	0.000
$Cons_{t-2}$	0.354	0.091	3.881	0.000
CA_t [a]	0.288	0.079	3.651	0.000
RH_max_t [a]	0.045	0.022	2.035	0.044
$Temp\_avg_t$ [a]	-0.012	0.034	-0.355	0.723
$Temp\_avg_{t-1}$	0.120	0.052	2.323	0.022
$Temp\_avg_{t-2}$	-0.088	0.038	-2.304	0.023
C	-0.001	0.019	-0.029	0.977

R-squared	0.953	F-statistic	318.736
Adjusted R-squared	0.950	Prob(F-statistic)	0
S.E. of regression	0.172	Sum squared resid	3.264
Akaike info criterion	-0.614	Schwarz criterion	-0.426

[a] $Cons_{t-i}$= electricity consumption of *i* month(s) before month t; CA_t= number of consumer accounts in month *t*; RH_max_t = maximum relative humidity of month *t*; $Temp\_avg_{t-i}$= average temperature of *i* month(s) before month *t*.

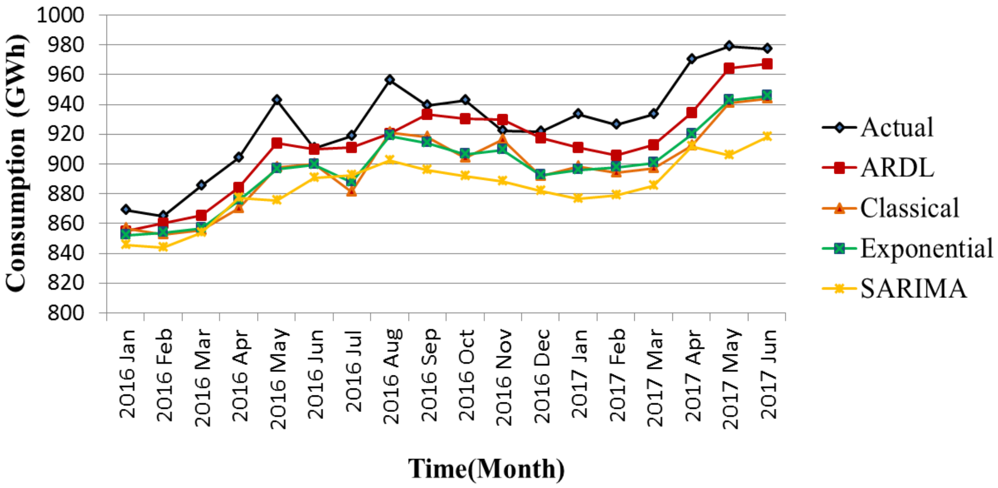

Fig. 4. Comparison of forecasts of the total consumption with the econometric models considered.

After aggregating all ARDL-based monthly forecasts, the total monthly electricity consumption in the country was computed. Figure 4 illustrates the forecasts obtained by econometric forecasting methods demonstrated in section 2 along with the actual monthly electricity consumption over the sample period of 18 months. Unreported results [3] revealed that additive models perform better than the multiplicative models in Classical and exponential smoothing approaches. Hence, additive models were used for the comparison in Figure 4. Moreover, ARIMA $(1,1,0)(2,0,0)_{12}$ model was the selected specification from several candidate models in Box-Jenkins approach due to its superior goodness-of-fit among other specifications.

Forecasts generated using ARDL model are the closest to the actual values compared with other forecasts. All the other forecasts stay below the actual values significantly indicating a positive bias in the forecasting error. Thus, among the econometric forecasting methods demonstrated in section 2, ARDL performs to the best over other models. Furthermore, Figure 5 plots actual versus predicted values generated from all econometric forecasting methods. Box-Jenkins SARIMA model appeared to be the worst-performing method among others due to the prominent deviation of predicted values from the actuals. Classical method and triple exponential smoothing models performed similar to each other, however with noticeable deviations. More importantly, the proposed ARDL model depicted the best performance with the nearest forecasts to the actual values.

[3] Available upon request

Finally, we compared forecasting accuracy among the four different methods, as shown in Table 4. We used MAPE in this comparison as demonstrated in Section 2.6. It was evident that ARDL approach reports the smallest MAPE. Alternative performance measures presented in Table 4 were also robust with this finding. Hence it can be stated that the ARDL approach outperforms all the other three methods used in this study. Moreover, exponential smoothing approach is the second-best method while the classical approach took the third place. Apparently, SARIMA model provided the weakest forecasts.

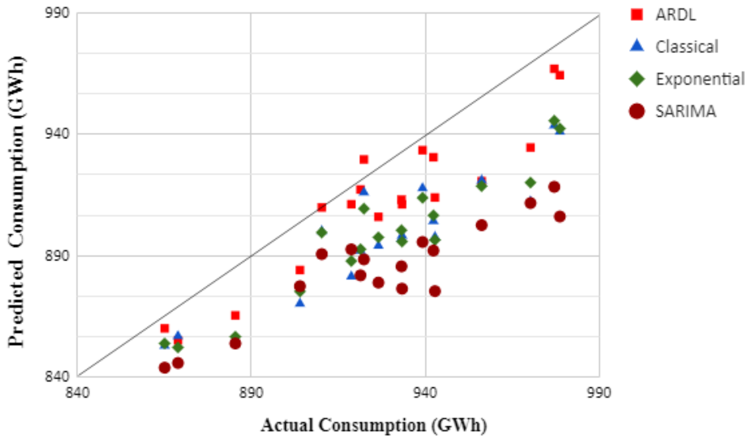

Fig. 5. Predicted (by the econometric models) vs. actual values of the total consumption.

Table 4. Measures of accuracy of the forecasts given by the econometric models.

Compared Models	Forecasting accuracy indices		
	MAE	RMSE	MAPE (%)
Classical	30,197.58	32,820.76	3.23
Exponential	29,552.30	31,446.67	3.16
SARIMA	43,316.13	46,145.52	4.62
ARDL	15,745.93	18,737.10	1.59

Finally, to ensure that the proposed ARDL model offers a significant improvement in forecasting accuracy, the two statistical tests in Section 2.7, Wilcoxon signed-rank test and Friedman test, were conducted. For this purpose, proposed ARDL model was compared with each of the alternative models (Wilcoxon signed-rank test) all together jointly (Friedman test). As reported in Table 5, the test results indicate that the proposed ARDL model significantly outperforms all the alternative methods.

Furthermore, we compared the superiority in forecasting performances of the proposed ARDL model with four alternative algorithms introduced in Section 2. Figure 6 demonstrates corresponding results over the test sample period of 18 months. In Figure 6, SVR-LR shows the enhancement of the SVR model incorporating linear regression is worth in forecasting the electricity consumption. However, among all alternatives, ARDL model stays on the top which is also evident from MAPE, RMSE, and MAE in Table 6.

Table 5. Results of Wilcoxon signed-rank test and Friedman test for the econometric models.

Compared models	Wilcoxon signed-rank test (α = 0.05)		Friedman Test (α= 0.05, two-tail test)
	W-value	p-value	
ARDL vs. Classical	168	0.0000	H_0:* $e_1 = e_2 = e_3 = e_4$
ARDL vs. Exponential	171	0.0000	F = 38.6
ARDL vs. SARIMA	171	0.0000	p value = 0.0000

*e_i = the population median of the absolute forecast error of the ith model.

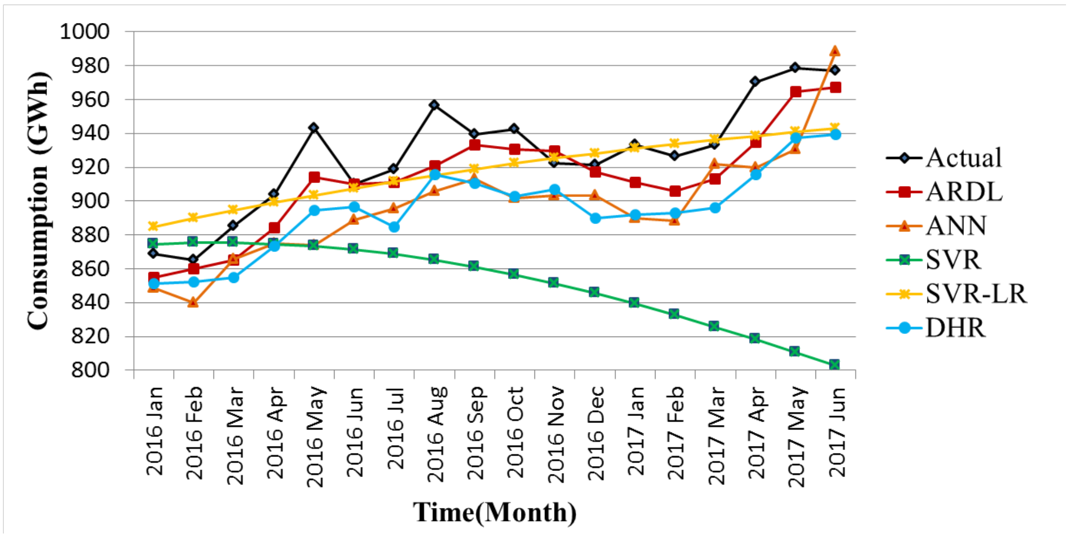

Fig. 6. Comparison of forecasts of the total consumption with the algorithmic forecasting methods and the ARDL model.

Table 6. Comparison of measures of accuracy of the algorithmic forecasting methods with the ARDL model.

Compared Models	Forecasting accuracy indices		
	MAE	RMSE	MAPE (%)
ARDL	15,745.93	18,737.10	1.59
ANN	31,335.22	34,954.11	3.36
SVR	78,112.96	92,349.84	8.26
SVR-LR	17,308.51	22,138.54	1.84
DHR	32,749.63	34,661.94	3.50

Table 7. Results of Wilcoxon signed-rank test and Friedman test for the algorithmic forecasting models and the ARDL model.

Compared models	Wilcoxon signed-rank test (α = 0.05)		Friedman Test (α= 0.05, two-tail test)
	W-value	p-value	
ARDL vs. ANN	154	0.0030	$H_0 : e_1 = e_2 = e_3 = e_4 = e_5$ F = 43.422 p value = 0.0000
ARDL vs. SVR	162	0.0010	
ARDL vs. SVR-LR	60	0.2760	
ARDL vs. DHR	171	0.0000	

In addition, the Wilcoxon signed-rank and Friedman tests were used similar to Table 5 and corresponding results are reported in Table 7. It is evident that the proposed ARDL model significantly outperforms ANN, SVR, and DHR but SVR-LR model that gives similar performance to ARDL model on average. However according to the fluctuation patterns observed in Figure 6, ARDL model outperforms the SVR-LR model by capturing the seasonality in forecasts.

4. DISCUSSION

This study assesses the performance of eight methods (four statistical models and four alternative algorithms) to forecast electricity consumption in Sri Lanka. We found the impact of weather conditions on the electricity consumption in consistent with the previous studies such as in Greece, England, and Whales [6], [7]. Also, we identified a monsoon-wise dependence of electricity demand in Sri Lanka. This study reinforces the importance of the awareness of weather impact on the electricity demand for effective planning and managing the electricity generation sources as documented in the literature [8].

Also, this study found that geographic location is a predominant factor in forecasting demand as in the study by Jovanović *et al.* [9]. Thus, we propose ARDL models that provide location-specific consumption forecasts with outperforming evidence. Thus, our proposal would be important for CEB to implement region-wise forecasting which is absent yet in their operations planning process. More importantly, our approach can be used to plan sustainable mini-electricity generation projects (*e.g.* hydro, wind, and solar) within regions with the ability to fulfill the region-wise short-term demand.

5. CONCLUSION

The current study finds the best econometric model to forecast monthly electricity consumption in Sri Lanka. This study also explored the influence of weather-related factors on electricity consumption. Among all econometric and algorithmic forecasting methods compared, ARDL model with weather-related predictors is found to be the most appropriate model in forecasting the monthly electricity consumption in Sri Lanka. The results confirm that the proposed ARDL approach can simultaneously provide forecasting with both accuracy and interpretability with its generalization ability (unlike in algorithmic forecasting). The proposed ARDL model estimated the electricity consumption district-wise which could help in planning region-wise electricity supply. Thus, we suggest using ARDL approach for CEB's regional electricity demand management.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Ceylon Electricity Board and the Department of Meteorology in Sri Lanka for providing the required data.

REFERENCES

[1] De Silva G.V. and L. Samaliarachchi. 2013. Peak electricity demand prediction model for Sri Lanka power system. *Engineer: Journal of the Institution of Engineers*, *Sri Lanka* 46(4).

[2] Ceylon Electricity Board, 2019. *Energy Generation*, Retrieved from the World Wide Web: https://ceb.lk/electricity-generated/en

[3] Bambaravanage T., Kumarawadu S. and Rodrigo A., 2016. Comparison of three under-frequency load shedding schemes referring to the power system of Sri Lanka. *Engineer: Journal of the Institution of Engineers*, *Sri Lanka* 49(1).

[4] Papalexopoulos A.D. and T.C. Hesterberg, 1990. A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems* 5(4): 1535-1547.

[5] Ceylon Electricity Board, 2017, *Sales and Generation Data Book*. Retrieved from the World Wide Web: https://ceb.lk/front_img/img_reports/1546506487Sales_and_Generation_Data_Book_2017.pdf

[6] Mirasgedis S., Sarafidis Y., Georgopoulou E., Lalas D.P., Moschovits M., Karagiannis F. and Papakonstantinou D., 2006. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy* 31(2-3): 208-227.

[7] Hor C.L., Watson S.J. and Majithia S., 2005. Analyzing the impact of weather variables on monthly electricity demand. *IEEE Transactions on Power Systems* 20(4): 2078-2085.

[8] Alabbas N. and J. Nyangon. 2016, October. Weather-based long-term electricity demand forecasting model for Saudi Arabia: a hybrid approach using end-use and econometric methods for comprehensive demand analysis. In *Implications of North American energy self-sufficiency, 34th USAEE/IAEE North American Conference*, Oct 23-26, 2016. International Association for Energy Economics.

[9] Jovanović S., Djordjević Z., Bojić M. and Stepanović S.S.B., 2012, November. Weather conditions impact on electricity consumption. In *1st International Scientific Conference*: 28-30.

[10] Lou C.W. and M.C. Dong. 2015. A novel random fuzzy neural networks for tackling uncertainties of electric load forecasting. *International Journal of Electrical Power & Energy Systems* 73: 34-44.

[11] Fan G.F., Peng L.L., Hong W.C. and Sun F., 2016. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing* 173: 958-970.

[12] Zhang Z. and W.C. Hong. 2019. Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dynamics* 98(2): 1107-1136.

[13] Sen P., Roy M. and Pal P., 2016. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy* 116: 1031-1038.

[14] Yang D., Sharma V., Ye Z., Lim L.I., Zhao L. and Aryaputera A.W., 2015. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy* 81: 111-119.

[15] Li Y., Jiang X., Zhu H., He X., Peeta S., Zheng T. and Li Y., 2016. Multiple measures-based chaotic time series for traffic flow prediction based on Bayesian theory. *Nonlinear Dynamics* 85(1): 179-194.

[16] Takeda H., Tamura Y. and Sato S., 2016. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy* 104: 184-198.

[17] Lebotsa M.E., Sigauke C., Bere A., Fildes R. and Boylan J.E., 2018. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. *Applied Energy* 222: 104-118.

[18] Kelo S. and S. Dudul. 2012. A wavelet Elman neural network for short-term electrical load prediction under the influence of temperature. *International Journal of Electrical Power & Energy Systems* 43(1): 1063-1071.

[19] Hernández L., Baladrón C., Aguiar J.M., Carro B., Sánchez-Esguevillas A. and Lloret J., 2014. Artificial neural networks for short-term load forecasting in microgrids environment. *Energy* 75: 252-264.

[20] Lusis P., Khalilpour K.R., Andrew L. and Liebman A., 2017. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Applied Energy* 205: 654-669.

[21] Duan Q., Liu J. and Zhao D., 2017. Short term electric load forecasting using an automated system of model choice. *International Journal of Electrical Power & Energy Systems* 91: 92-100.

[22] Zhang W., Zhang S. and Zhang S., 2018. Two-factor high-order fuzzy-trend FTS model based on BSO-FCM and improved KA for TAIEX stock forecasting. *Nonlinear Dynamics* 94(2): 1429-1446.

[23] Hua J.C., Noorian F., Moss D., Leong P.H. and Gunaratne G.H., 2017. High-dimensional time series prediction using kernel-based Koopman mode regression. *Nonlinear Dynamics* 90(3): 1785-1806.
[24] Fan G.F., Wei X., Li Y.T. and Hong W.C., 2020. Forecasting electricity consumption using a novel hybrid model. *Sustainable Cities and Society* 61: 102320.
[25] Fan G.F., Guo Y.H., Zheng J.M. and Hong W.C., 2020. A generalized regression model based on hybrid empirical mode decomposition and support vector regression with back-propagation neural network for mid-short-term load forecasting. *Journal of Forecasting* 39(5): 737-756.
[26] Hong W.C. and G.F. Fan. 2019. Hybrid empirical mode decomposition with support vector regression model for short term load forecasting. *Energies* 12(6): 1093.
[27] Fan G.F., Peng L.L. and Hong W.C., 2018. Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. *Applied Energy* 224: 13-33.
[28] Fan G.F., Qing S., Wang H., Hong W.C. and Li H.J., 2013. Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting. *Energies* 6(4): 1887-1901.
[29] Engle R.F., Granger C.W., Rice J. and Weiss A., 1986. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association* 81(394): 310-320.
[30] Chikobvu D. and C. Sigauke. 2013. Modelling influence of temperature on daily peak electricity demand in South Africa. *Journal of Energy in Southern Africa* 24(4): 63-70.
[31] Fahad M.U. and N. Arbab. 2014. Factor affecting short term load forecasting. *Journal of Clean Energy Technologies*, 2(4): 305-309.
[32] Fan S. and R.J. Hyndman. 2011. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems* 27(1): 134-141.
[33] Benli T.O., 2016. A comparison of various electricity tariff price forecasting techniques in turkey and identifying the impact of time series periods. *arXiv preprint arXiv*:1610.08415.
[34] Faraway J. and C. Chatfield. 1998. Time series forecasting with neural networks: a comparative study using the airline data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47(2): 231-250.
[35] Bildirici M.E., 2013. The analysis of relationship between economic growth and electricity consumption in Africa by ARDL method. *Energy Economics Letters* 1(1): 1-14.
[36] Saviozzi M., Massucco S. and Silvestro F., 2019. Implementation of advanced functionalities for Distribution Management Systems: Load forecasting and modeling through Artificial Neural Networks ensembles. *Electric Power Systems Research* 167: 230-239.
[37] Wu H., Zhou Y., Luo Q. and Basset M.A., 2016. Training feedforward neural networks using symbiotic organisms search algorithm. *Computational intelligence and neuroscience*, 2016.
[38] Lin K., Lin Q., Zhou C. and Yao J., 2007, August. Time series prediction based on linear regression and SVR. In *Third International Conference on Natural Computation (ICNC 2007) 1*: 688-691. IEEE.
[39] Young P.C., Pedregal D.J. and Tych W., 1999. Dynamic harmonic regression. *Journal of Forecasting*, 18(6): 369-394.
[40] Fung D.S., 2006. Methods for the estimation of missing values in time series. MS Thesis, Edith Cowan University Edith Cowan University, Perth, Western Australia.
[41] Abeyratne P.G.V., Featherstone W.E. and Tantrigoda D.A., 2010. On the geodetic datums in Sri Lanka. *Survey Review* 42(317): 229-239.
[42] Sluiter R., 2009. Interpolation methods for climate data: literature review. *KNMI intern rapport, Royal Netherlands Meteorological Institute*, *De Bilt.*
[43] Taylor J.W. and R. Buizza. 2003. Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting* 19(1): 57-70.
[44] Zhang Z., Ding S. and Sun Y., 2020. A support vector regression model hybridized with chaotic krill herd algorithm and empirical mode decomposition for regression task. *Neurocomputing* 410: 185-201.
[45] Diebold F.X. and R.S. Mariano. 2002. Comparing predictive accuracy. *Journal of Business & economic statistics* 20(1): 134-144.
[46] Derrac J., García S., Molina D. and Herrera F., 2011. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation* 1(1): 3-18.
[47] Montgomery D.C. and Runger G.C., 2010. *Applied statistics and probability for engineers*. John Wiley & Sons.