Forecasting of Electromagnetic Radiation Time Series: An Empirical Comparative Approach

Zeydin Pala [1], İbrahim Halil Ünlük [1], and Erkan Yaldız [2]

[1] Department of Computer Engineering/Computer Programming
University of Mus Alparslan, Mus, 49250, Turkey
z.pala@alparslan.edu.tr, ih.unluk@alparslan.edu.tr

[2] Municipality Information Processing Office, Mus, Turkey
erkan1yaldiz@gmail.com

Abstract **─** This study compares the performance of time series models for forecasting electromagnetic radiation levels at Yesilce neighborhood in Mus, Turkey.

To make successful predictions using EMF time series, which is obtained in the 36-month measurement process using the calibrated Wavecontrol SMP2 device, nine different models were used. In addition to Mean, Naive, Seasonal Naïve, Drift, STLF and TBATS standard models, more advanced ANN models such as NNETAR, MLP and ELM used in the R software environment for forecasting. In order to determine the accuracy of the models used in the EMF time series used in the study, mean absolute error (MAE), relative mean absolute error (RMAE) metrics were used. The best results obtained with NNETAR, Seasonal Naïve, MLP, STLF, TBATS, and ELM models, respectively.

Index Terms **─** Electromagnetic radiation, ELM, forecasting models, MLP, NNETAR, time series.

I. INTRODUCTION

Based on past and current data, predicting the future and making plans according to it has a very great meaning for practical life. Predicting the future is of great importance in health as well as in many areas. Time series analysis for forecasting of time depending variables, which is the collection of data at specific intervals over a period is not new [1], but is an important area of machine learning. Usually, the measurement of time series is made at regular time intervals that are a list of observation where the ordering matters. Due to dependency, ordering is very important in time series. In other words, changing time series of measurements means changing means of the data. Time series analysis is frequently used to estimate a relationship between adverse health outcomes and short-term exposes to ambient electromagnetic radiation (EMR) [2]. Modeling some forecasting models for low-frequency high voltage EMR characteristics based on these historical data may be useful. In order to model the time series correctly and efficiently and predict the future, it is possible to find many important models in the literature.

Therefore, time series used extensively in the fields of industry, engineering, and science, especially in the fields of economics, business, physical sciences [3], and finance.

The research question in this paper is that in living environments weather EMR effects on humans and forecasting EMR time series data using popular algorithms.

The main objectives of this study are: To act an empirical study and compare analysis and highly extensible forecasting methods of R, which is a language and environment for statistical computing and graphics. To compare the performance of statistical and ANN based models. To investigate the influence of electromagnetic radiation on humans.

II. MATHEMATICAL BACKGROUND

Time series are sequential measurements data points of the same variable over successive times. Our EMF time series can be mathematical defined as a set of vectors $\{x_1, x_2, x_3, \ldots, x_T\}$, $t=0,1,2,3,\ldots, T$, where T is the set of times at which the process was, will or can be observed, t is an index denoting the period in time in which x occurs. Here x_t is treated as a random variable. Time series are called univariate or multivariate according to the time-dependent variable they contain. The EMF time series used in this study is called univariate because it is based on one time-dependent variable [4].

For a time series modeling a general approach is to plot the series and examine the main features of the graph, checking in particular whether there is (i) a trend, (ii) a seasonal component, (iii) any apparent sharp changes in behavior and any some observations [5].

In time series, forecasting predictor variables are used usefully. In this study, we wish to forecast the

Submitted On: December 5, 2018
Accepted On: May 12, 2019

monthly low-frequency electromagnetic fields (EMF) of a high voltage electric lines. A model with predictor variables might be of the form:

$$EMF = f(\text{device sensitivity}, \textit{distance}, \textit{time}, \grave{\alpha}). \quad (1)$$

The ϵ_t term on the right allows effects of relevant variables and for random variation that is not included in the model. Since the model helps explain, what causes the variation in EMF it can be calls, explanatory model. In this situation, a convenient time series forecasting equation is of the form:

$$EMF_{t+1} = f(EMF_t, EMF_{t-1}, EMF_{t-2}, \ldots, \grave{\alpha}), \quad (2)$$

where t is the present month, t+1 is the next month, t−1 is the previous month, t−2 is two months ago, and so on [6].

A. Forecasting methods in R

The statistical functions of the R language are very strong. In this study, all functions such as Mean, Naive, Seasonal Naive, Drift, STLF, and TBATS, which are used for prediction purposes in the R environment, are defined within the forecast class. All of the above-mentioned functions can be used for prediction of the desired time series after importing the forecast class. The following paragraphs give a brief description of these functions:

The Main method in the forecast class, which returns estimates and prediction ranges, for an independent and identically distributed (iid) model that is applied to the time series. If we use the data of EMF time series y_1, y_2, y_3, ..., y_T, then we can write the forecasts as:

$$\hat{y}_{T+\Delta|T} = \frac{y_1+y_2+y_3+\cdots+y_T}{T}. \quad (3)$$

Where, the left side of the equation expresses the short-term forecast, in other words, it means the forecast $y_{T+\Delta}$ taking account of y_1, y_2, y_3, ..., y_T and Δ is the forecast horizon.

We used the Naïve method for forecasts and we simply set all forecasts to be the value of the last observation of the EMF dataset. That is,

$$\hat{y}_{T+\Delta|T} = y_T. \quad (4)$$

The Seasonal Naïve method, which is a similar and useful method for highly seasonal data. Using this method, we foresee that each estimate will be equal to the last observed monthly value starting from the same month of the year.

Estimating the monthly EMF, which is normally much more than T + Δ, can be written as follows:

$$\hat{y}_{T+\Delta|T} = y_{T+\Delta-\alpha(\beta+1)}, \quad (5)$$

where α is the seasonal period and β is the integer part of (Δ−1)/α. For example, with monthly data, the forecast for all future January values is equal to the last observed January value.

One of the simplest method to predict using the general trend of the time series is the Drift method, a variation on the naïve method. Thus, the forecast for time T+Δ is given by [6]:

$$\hat{y}_{T+\Delta|T} = y_T + \frac{\Delta}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_t + \Delta\left(\frac{y_T-y_1}{T-1}\right). \quad (6)$$

The Drift method is equivalent to drawing a line between the first and last observations and calculating it for the future. TBATS model uses a combination of Fourier terms with an exponential smoothing state space model and a Box-Cox transformation. One drawback of TBATS models, however, is that they can be slow to estimate, especially with long time series. The STLF function, which is expressed as seasonal and trend decomposition using loess, is a fairly simple yet powerful estimation function. The STLF model assumes that time series can be divided into trend, seasonality, and error components.

B. Artificial Neural Networks (ANN) in R

Artificial neural networks (ANNs), which is accepted as one of the applications of artificial intelligence, is the information processing technology that analyzes the existing data by mimicking the working structure of the human brain and creates new information with different learning algorithms. In this study, along with the NNETAR algorithm, which is based on the artificial neural network in the Forecast class, multilayer perceptron (MLP) and Extreme learning machines (ELM) algorithms in the NNFOR class were used.

The neural network autoregression (NNAR) model consists of three layers: input, hidden and output. This model is defined in the R forecast library and the NNAR model becomes available after the forecast library is imported. It generally performs better than conventional algorithms such as Main, Naive, Seasonal Naive, and STLF. The NNAR model is a parametric and non-linear estimation model [7]. The MLP is a fully connected feed-forward networks, supervised learning network with up to one or more hidden layers, and probably the most common network architecture in use. Training of the model is usually performed by error backpropagation or a related procedure [8, 9]. The ELMs are also feed-forward neural networks, primarily used in classification, regression and clustering with a single layer or multiple hidden node layers. They are especially preferred in terms of presenting models for a large class of natural and artificial phenomena, which are difficult to handle using classical parametric techniques [10].

C. Predictive results comparative analysis

The estimation error of a model can be defined as the difference between the actual value of the model and its estimated value. It can be written as [6]:

$$e_{T+\Delta} = y_{T+\Delta} - \hat{y}_{T+\Delta|T}, \quad (7)$$

where the training data is given by $\{y_1, y_2, y_3, \ldots, y_T\}$ and the test data is given by $\{y_{T+1}, y_{T+2}, y_{T+3}, \ldots\}$.

The performance of the models used in this study was evaluated using statistical metrics such as the mean absolute error (MAE) and relative mean absolute error

(RMAE). The MAE, which is measures the average absolute deviation of forecasted values from original ones can be defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_{T+\Delta}|. \quad (8)$$

Where n is the number of observation. The RMAE is used to evaluate models [11]:

$$RMAE = \frac{1}{|\hat{y}_{T+\Delta|T}|}\sum_{i=1}^{n}|e_{T+\Delta}|. \quad (9)$$

MAE is the average absolute deviation of forecasted values from original ones.

III. IMPLEMANTATION AND DISCUSSION

The dataset used in this study was obtained as a result of 36 months of regular measurements. Measurements were made at a height of 2.9 m from the high voltage line and 2.5 m from the ground. In the measurements, the highest EMF value was measured as E_{max}=6792 V/m in January 2018 and the lowest value was E_{min} 5388 V/m in July 2017. All measurements made at the points where high voltage lines are located in Yeşilce neighborhood of Muş province have been carried out according to the standards determined by ICNRP. The measurement time was made as the average of 6-minute measurements as determined by these standards. These values measured monthly using the Wavecontrol SMP2 device exceed ICNIRP (1998) and Council of European Union (1999) public exposure limits. Using the R programming language version 3.5.1, all analyzes were performed in RStudio.

In addition, the time series is divided into multiple components using the decomposition method to make predictions more realistic. Equipment used for the periodic measurement of the electric field of high voltage lines in outdoor environment and decomposition graphs of the data obtained in the measurements are shown in Fig. 1 and Fig. 2, respectively. As seen in the time series, it is observed that electricity consumption decreases in summer and increases in winter months. Accordingly, EMF values also varied. By using decomposition method in the Studio environment, EMR time series is divided into four components as data, seasonal, trend and remainder. These components are shown in Fig. 2.

Fig. 1. Periodic measurement image under high voltage line.

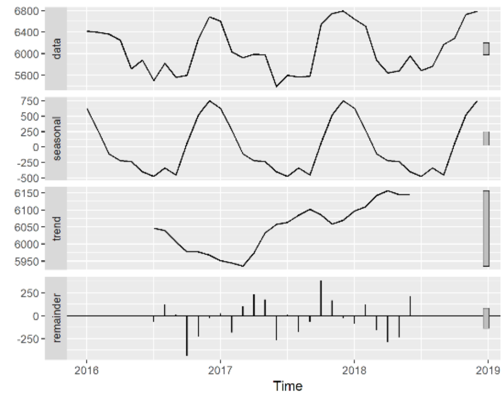

Fig. 2. The EMF data series (top) and its three additive components.

In the first analysis, the first 30-month portion of the EMF time series, which was generated as a result of the 36-month measurement, was used for training (84%) purposes, and the remaining were used for testing (16%) purposes.

In the second analysis, the dataset was divided into 70% and 30% for training and testing, respectively. It is also necessary to consider the forecasting horizon (Δ) when using forecasting methods. For all functions used in the analysis, the value of the Δ parameter was taken as 12 for the prediction of the next 12 months.

In order to have a better understanding on the performance of the selected methods, the performance metrics of the forecasted methods are shown in Table 1. When both MAE and RMAE performance metrics were evaluated together, Seasonal Naive from the standard functions showed the best performance, whereas, for neural network functions, the NNETAR function showed the best performance. In general, the RMAE average value of ANN algorithms such as NNETAR, MLP, and ELM is smaller than the average value of the other classical algorithms. This result shows that ANN algorithms perform better than classical algorithms [7]. As shown in Table 1, the division of the dataset (84%, 16%) for training and testing operations resulted in a better performance than dividing by (70%, 30%).

Table 1: Comparison metrics of forecasting methods

Model Strategy	**Training (84%) Test (16%)**		**Training (70%) Test (30%)**	
Method/Metric	**MAE**	**RMAE**	**MAE**	**RMAE**
Seasonal Naïve	190.66	0.184	262.36	0.475
STLF	205.59	0.198	262.47	0.476
TBATS	209.86	0.202	239.80	0.434
MEAN	402.90	0.389	358.98	0.650
NAIVE	434.00	0.419	585.18	1.060
DRIFT	473.65	0.457	606.68	1.098
MLP	205.55	0.197	504.76	0.880
NNETAR	183.71	0.177	486.75	0.875
ELM	303.04	0.292	612.13	1.108

The visual predictions of the classical algorithms defined in the Forecast class and the ANN algorithms defined in the NNFOR class are shown in Fig. 3 and Fig. 4, respectively.

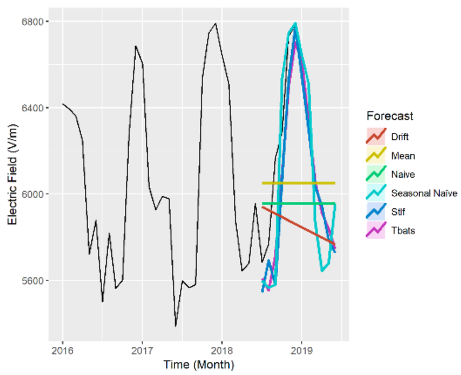

Fig. 3. Comparative forecasts for EMF monthly time series.

As shown in Fig. 3, considering the seasonal changes in the EMF time series, it is possible to say that the 12-month estimates of the Seasonal Naive, STLF and TBATS functions are quite satisfactory.

The MLP and ELM models used in the estimation process consisted of three layers. The numbers of neurons in the inputs, hidden and output layers of each model were respectively (3, 5, 1) and (3, 17, 1). As shown in Table 1, ANN based algorithms are NNETAR, MLP and ELM, respectively, based on metric values.

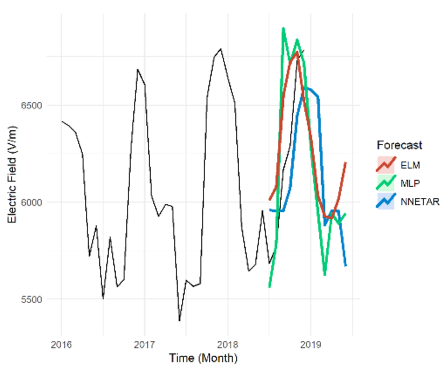

Fig. 4. Future predictions of ANN-based algorithms.

IV. CONCLUSION

In this study, a dataset based on 36-month EMF measurements was used. The monthly values obtained in the measurement environment generally exceed the ICNIRP public exposure limits. In addition, we compared the accuracy of Mean, Drift, Naïve, Seasonal Naïve, STL, and TBATS model as representative methods when forecasts EMF time series data. The comparative study of presented work is also performed with the ANN models for forecasting of the EMF time series. In the next 12-month forecasting of the EMF time series, the best performance values were obtained using NNETAR, Seasonal Naïve, MLP, STLF, TBATS and ELM models, respectively.

REFERENCES

[1] C. Chatfield and A. S. Weigend, "Time series prediction: Forecasting the future and understanding the past," *International Journal of Forecasting*, vol. 10, p. 161-163, 1996.

[2] G. Redlarski, B. Lewczuk, A. Zak, et al., "The influence of electromagnetic pollution on living organisms: Historical trends and forecasting changes," *BioMed Research International*, vol. 2015, pp. 1-18, 2015.

[3] Z. Pala and R. Atici, "Forecasting sunspot time series using deep learning methods," *Solar Physics*, vol. 294, no. 50, pp. 1-14, 2019.

[4] K. W. Hipel and A. I. Mcleod, *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam, Elsevier 1994.

[5] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. 3rd ed., Springer, Switzerland, 2016.

[6] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. 2nd ed., Monash University, Australia, 2018.

[7] D. Sena and N. K. Nagwani, "A neural network autoregression model to forecast per capita disposable income," vol. 11, pp. 13123-13128, 2016.

[8] K. S. R. Krishna, J. L. Narayana, and L. P. Reddy, "ANN models for microstrip line synthesis and analysis," *World Acad. Sci. Eng. Technol.*, vol. 2, no. 10, pp. 724-8, 2008.

[9] Z. D. Marinković and V. V. Marković, "Temperature-dependent models of low-noise microwave transistors based on neural networks," *Int. J. RF Microw. Comput. Eng.*, vol. 15, no. 6, pp. 567-77, 2005.

[10] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006.

[11] F. Güneş, P. Mahouti, S. Demirel, M. A. Belen, and A. Uluslu, "Cost-effective GRNN-based modeling of microwave transistors with a reduced number of measurements," *Int. J. Numer. Model. Electron. Networks, Devices Fields*, vol. 30, no. 3-4, pp. 1-12, 2017.