UMYU Scientifica, Vol. 3 NO. 3, September 2024, Pp 082 – 088
https://doi.org/10.56919/usci.2433.010
ISSN: 2955 – 1145 (print); 2955 – 1153 (online)
A periodical of the Faculty of Natural and Applied Sciences, UMYU, Katsina

ORIGINAL RESEARCH ARTICLE

Modeling Wind Speed Data Using Dynamic Linear Model and Innovation State Space ETS Model

Ibrahim Lawal Kane[1]* and Ibrahim Mu'awiyya Idris[2]
[1,2]Department of Mathematics and Statistics, Umaru Musa Yar'adua University, 2218, Katsina State, Nigeria

ABSTRACT

For many years, experts have acknowledged the intricate nature of wind speed, considering it a prime example of complex systems. In recent times, wind speed patterns have become increasingly erratic. This research paper focuses on a state space approach that utilizes dynamic linear models (DLM) and an Innovation state space model based on Exponential smoothing methods to effectively model wind speed data in the Katsina metropolis from January 2013 to December 2022. By incorporating trend, seasonal, and regressive components, these methods allow for a natural interpretation of the data. The model's parameters were estimated, and their validity was assessed through residual analysis. The validated models were then used for one-step-ahead forecasts. Comparing the predicted values with the observed wind speed series, the DLM demonstrates a closer match than the Innovation state space ETS (M, N, A), indicating that the model accurately replicates the actual data. Consequently, the model is considered suitable for representing the wind speed of Katsina.

ARTICLE HISTORY

Received April 07, 2024
Accepted July 11, 2024
Published July 19, 2024

KEYWORDS

Wind speed, Dynamic Linear Model, ETS state space model, Kalman filter Ljung Box test.

INTRODUCTION

Among the meteorological factors, wind speed stands out as a crucial atmospheric element with significant implications for the energy production industry. Its impact extends to weather forecasting, aircraft and maritime operations, and other sectors. With the increasing importance of wind energy, power utilities must strategically integrate wind power. Therefore, precise prediction of wind speed is essential. It provides insights into historical and projected wind speed patterns, enabling accurate estimation of power energy generation. Numerous techniques and methodologies for developing forecasting models can be found in the literature. For example, Numerical weather prediction (NWP) models are commonly used for modeling and forecasting wind speed. These models simulate the dynamics of the atmosphere using complex mathematical equations and input data from weather stations, satellites, and other sources. While NWP models can provide valuable insights into wind speed patterns, they often struggle to accurately capture local wind phenomena and require significant computational resources (Zhang (2021).

Statistical methods, such as machine learning algorithms and time series analysis, have also been employed for wind speed forecasting (Wang *et al.,* 2022; Chatterjee & Mondal, 2021; Bessa *et al.,* 2020). These methods analyze historical wind speed data to identify patterns and trends that can be used to predict future wind speed. While statistical methods are computationally less expensive than NWP models, they may not capture the complex interactions of the atmosphere.

Recently, there has been a trend toward hybrid models that merge the advantages of statistical and physical methods. These hybrids exploit the abundance of high-quality observational data and improvements in computational capabilities to blend data-driven techniques with a physical understanding, thereby improving the precision and dependability of forecasts. This trend has sparked a growing interest in creating statistical models capable of grasping the intricate dynamics of time series data and delivering precise future predictions, e.g. (Wang *et al.,* 2019; Chai *et al.*, 2019).

State-of-the-art methods and recent probabilistic developments in wind speed prediction have been described (Zhang *et al.*, 2020). The study suggests that wind speed is a nonlinear process with varying degrees of variability. This variability leads to uncertainty in wind

Correspondence: Ibrahim Lawal Kane. Department of Mathematics and Statistics, Umaru Musa Yar'adua University, 2218, Katsina State, Nigeria. ✉ ibrahim.lawal@umyu.edu.ng.
How to cite: Kane, I. L., & Idris, I. M. (2024). Modeling Wind Speed Data Using Dynamic Linear Model and Innovation State Space ETS Model. *UMYU Scientifica*, 3(3), 82 – 88. https://doi.org/10.56919/usci.2433.010

power generation forecasts and operating system decisions. Numerous proposals have been put forward to comprehend the stochastic characteristics of wind speed (Jiang & Zhao, 2013; Azorin et al., 2018; Agada *et al.*, 2023).

However, this study focuses solely on time series forecasting models such as the dynamic linear model (DLM) framework and Innovation state space model based on Exponential smoothing methods to model the wind speed time series. The DLM is a versatile and robust statistical framework capable of handling intricate time series data. It achieves this by integrating systematic and random components into the model. Utilizing the DLM enables us to address temporal dependence, non-stationarity, and non-linearity in time series data characteristics that traditional time series models often struggle to capture. Exponential smoothing is a simple but very useful technique of adaptive time series forecasting. Standard seasonal methods of exponential smoothing include the Holt-Winters additive trend, multiplicative trend, damped additive trend, and damped multiplicative trend. The ETS (Error, Trend, Seasonality) models are a class of Exponential smoothing models used for forecasting time series data. These models are implemented in the R programming language using the ets() function, which is part of the forecast package developed by (Hyndman et al., 2008). The ets() function can be used to automatically select and fit the most appropriate ETS model to a given time series dataset, based on its error, trend, and seasonality components.

Dynamic Linear Modeling (DLM) is a potent statistical framework for analyzing time series data, finding broad applications across diverse fields such as economics, finance, and meteorology. Chen et al. (2019) employed a Dynamic Linear Model (DLM) to analyze monthly rainfall data from the Yangtze River basin in China. Their model integrated deterministic and stochastic components and utilized Bayesian inference for parameter estimation. The study concluded that the DLM effectively captured both the seasonal patterns and interannual variability of rainfall, enabling accurate predictions of future rainfall. Marangon et al. (2014) presented a Bayesian DLM for forecasting the water inflow at the Brazilian hydropower reservoirs. Nobre et al., (2001) compared the performances of a seasonal autoregressive integrated moving average (SARIMA) and a dynamic linear model (DLM) for estimating case occurrence of two diseases such as reported cases of malaria and hepatitis A from January 1980 to June 1995 for the United States and concluded that the DLM approach has some advantages over SARIMA model.

This paper tends to contribute to the field of wind speed modeling by innovative state space and dynamic linear modeling approaches that effectively captures the complexities and uncertainties inherent in wind speed patterns. By employing state-of-the-art modeling techniques, this study seeks to compare the performance of the two models.

DYNAMIC LINEAR MODEL

DLMs are a broad class of models that support time-varying parameters among their many features. Because of this, they can embed the temporal dependence within the functional relationship and eliminate the need for multiple lags of a covariate to account for the temporal nature of the data, leading to a cleaner interpretation of the parameters (Osthus et al. 2014).

Consider the wind speed data Y_t. Presumably, the series can be expressed as the total of its independent parts, as:

$$Y_t = Y_{1,t} + \cdots +, Y_{j,t} \tag{1}$$

where $Y_{1,t}$ represent a trend component, $Y_{2,t}$ a seasonal component, and so on. The $i-th$ component in $Y_{i,t}$, $i = 1,\ldots,j$, be defined by a DLM as $Y_{i,t} = F_{i,t}\theta_{i,t} + v_{i,t}$ with $v_{i,t} \sim N(0, V_{i,t})$ and $\theta_t = G_{i,t}\theta_{i,t-1} + w_{i,t}$ with $w_{i,t} \sim N(0, W_{i,t})$

The DLM for (Y_t) is then obtained by combining the DLM's components. Assuming the components are independent, it is simple to demonstrate that $Y_t = \sum_{i=1}^{h} Y_{i,t}$ is what the DLM describes as;

$$Y_t = F_t\theta_t + v_t, \quad v_t \sim N(0, V_t) \tag{2}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \tag{3}$$

where F_t and G_t are the *evolution matrices* which pre-multiplies the previous period's state vector and the v_t and w_t are two independent white noise sequences with mean zero and known covariance matrices V_t and W_t respectively.

To specify a DLM, the parameters; G_t, F_t, V_t and W_t must be specified for each period *t* (Petris, 2010). The filtering distribution of θ_t is the distribution of $\theta_t / y_1, y_2, \ldots, y_t$, while the smoothing distribution of θ_t at time s is the conditional distribution of $\theta_t / y_1, y_2, \ldots, y_s$, for s ≥ t., under the assumptions that the distributions are Gaussian.

The Kalman filter algorithm follows:

$$\alpha_t = c_t + T_t\alpha_{t-1} + R_t\eta_t \quad (4)$$

$$y_t = d_t + Z_t\alpha_t + \varepsilon_t \quad (5)$$

where $\eta_t \sim N(0, Q_t)$ and $R_t \sim N(0, H_t)$. The recurring reference is made in the outcome:

$$\alpha_{t-1} = E[\alpha_{t-1}|y_0, \dots, y_{t-1}] \quad (6)$$

$$P_{t-1} = E[(\alpha_{t-1} - a_{t-1})(\alpha_{t-1} - a_{t-1})^T] \quad (7)$$

The estimates of the state vector and its covariance matrix at time t with information available at time $t-1$, $a_{t|t-1}$ and $P_{t|t-1}$ respectively, are given by the time appraise equations:

$$a_{t|t-1} = T_t a_{t-1} + c_t \quad (8)$$

$$P_{t|t-1} = T_t P_{t-1} T_t^T + R_t Q_t R_t^T \quad (9)$$

Let $F_t = Z_t P_{t|t-1} Z_t^T + H_t$. If a new observation is available at time t, then $a_{t|t-1}$ and $P_{t|t-1}$ can be updated with the measurement update equations:

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t^T F_t^{-1}(y_t - Z_t a_{t|t-1} - d_t) \quad (10)$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t F_t^{-1} Z_t^T P_{t|t-1} \quad (11)$$

For convenience the DLM 's can be written by the joint density of the observations in the form:

$$p(y_1, \dots, y_n; \psi) = \prod_{t=1}^{n} p(y_t|D_{t-1}; \psi) \quad (12)$$

where $p(y_t|D_{t-1}; \psi)$ is the conditional density of y_t given the data up to time t-1, assuming that ψ is the value of the unknown parameter. Therefore the likelihood equation can be written as;

$$\ell(\psi) = -\frac{1}{2}\sum_{t=1}^{n} log|Q_t| - \frac{1}{2}\sum_{t=1}^{n}(y_t - f_t)' Q_t^{-1} (y_t - f_t) \quad (13)$$

where the f_t and the Q_t depend implicitly on ψ. The expression (13) can be numerically maximized to obtain the maximum likelihood estimator (MLE) of ψ:

$$\hat{\psi} = argmax\ \ell(\psi) \quad (14)$$

Innovations state space models for exponential smoothing

The exponential smoothing methods are algorithms which generate point forecasts. For each method there exist two models: one with additive errors and one with multiplicative errors. The point forecasts produced by the models are identical if they use the same smoothing parameter values. They will, however, generate different prediction intervals. A statistical model is a stochastic (or random) data generating process that can produce an entire forecast distribution. The model in this consists of a measurement equation that describes the observed data, and some state equations that describe how the unobserved components or states (level, trend, seasonal) change over time. Hence, these are referred to as state space models (Hyndman, & Khandakar, 2008).

Specifying one-step-ahead training errors as relative errors;

$$\varepsilon_t = \frac{y_t - (l_{t-1} + b_{t-1})}{(l_{t-1} + b_{t-1})} \quad (15)$$

we can write an innovations state space model underlying Holt's linear method with multiplicative errors as

$$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t) \quad (16)$$

$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$$

$$b_t = b_{t-1} + \beta(l_{t-1} + b_{t-1})\varepsilon_t$$

where for simplicity we have set $\beta = \alpha\beta^*$ and $\varepsilon_t \sim NID(0, \sigma^2)$.

Model Adequacy Checking and Comparison

To ensure that the model is correctly specified and can be used to model the wind speed data, the model is exposed for diagnostic testing. Diagnostic checks were performed to determine whether the models fit the data very well. If the model fits well, the residuals of the model should be uncorrelated. The Ljung-Box test (Dare *et al.,* 2022) is commonly used for this purpose given as

$$Q = N(N+2)\sum_{j=1}^{L} \frac{\hat{p}_j^2}{(N-j)} \quad (17)$$

Where N is sample size, L is the number of autocorrelation lags included in the statistic, and $\hat{p}_j^2$ is the squared sample autocorrelation at lag j. Under the null hypothesis of no serial correlation, the Q test statistic is asymptotically chi-square distributed. The p value above 0.05 indicate the acceptance of the null hypothesis of model adequacy at significance level 0.05.

Forecasting competency of the models can be obtained by the root mean square error (RMSE) and mean absolute percentage error (MAPE) between observed and forecasted wind speed values. RMSE and MAPE are calculated with equations 18 &19 respectively;

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \quad (18)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{Y_t} \quad (19)$$

Large MSE and MAPE value imply low accuracy and vice-versa.

RESULTS AND DISCUSSION

The Study Area

Katsina State, located in northern Nigeria, experiences a typical Sahelian climate characterized by distinct wet and dry seasons. The dry season typically lasts from November to March, characterized by hot and dry weather with temperatures often exceeding 40°C (104°F). During the wet season, which usually starts around May and lasts until September, Katsina State receives the majority of its annual rainfall. The peak of the rainy season occurs between July and August, with the rainfall supporting agricultural activities in the region, including the cultivation of crops such as millet, sorghum, and cowpeas.

Wind speeds in Katsina State can vary depending on the season. During the dry season, winds are generally light, contributing to the dry and dusty conditions. In contrast, the wet season may experience stronger winds, especially during thunderstorms and squalls associated with the rainy season. Generally, Katsina State's weather is characterized by its distinct wet and dry seasons, with temperature and wind patterns varying throughout the year.

The data set analyzed in this study is the Katsina wind speed record from January 2013 to December 2022. The statistical description of the data shows the average highest speed in January with 7.0m/s and the lowest speed of 4.3 in September. The average speed over the year is 5.8m/s. Figure 1 depicts the Katsina wind speed data with speed variations over time.

The theoretical models, which work as the basic structure of the analysis, are the dynamic linear model and ETS(M,N,A) Model. The smart possession of DLM is the recursive nature of its filtering methods which extrapolate expertly as new data developed. The parameter estimates that specify the wind speed DLM model and wind speed ETS(M,N,A) model are represented in Tables 1 & 2, respectively.

The parameters M_0 and C_0 are the mean and the variance of the prior distribution, respectively, FF is the covariates, V is observational variance, GG is evolution, and W is the evolution variance. So, based on the parameters in Table 1, the DLM wind speed forecasting model is given by;

$$Y_t = \mu_t + v_t; \quad v_t \sim N(5.466 \times 10^{-8}) \quad (20)$$

where $\mu_t = \mu_{t-1} + w_t$; $w_t \sim N(8.346 \times 10^{-1})$ with the Prior distribution, $(\mu_0|D_0) \sim N(0,\ 1 \times 10^7)$.

Model Validation and Forecast Ability

For both the DLM and ETS models, the Ljung-Box test yields p-values above 0.05, which suggests that both models adequately capture the autocorrelation in the data, and their residuals are independent, indicating that both models are suitable for forecasting the wind speed time series data.

The accuracy between both models; DLM and ETS (M, N, A) has been compared. The performance of the models was evaluated using the RMSE values and MAPE values; the lowest the better. The comparison of the RMSE value between in-sample forecasts is shown in Table 3.

Figure 2 displays the one-step ahead forecast plot from the DLM fitted and ETS (M,N,A) fitted to the wind speed data. The fitted values from the DLM model trail closely to the observed data than the ETS(M,N,A) model. Therefore, the model notably represents the wind speed time series data.

The DLM model performs better than the ETS (M, N, A) model based on RMSE and MAPE.

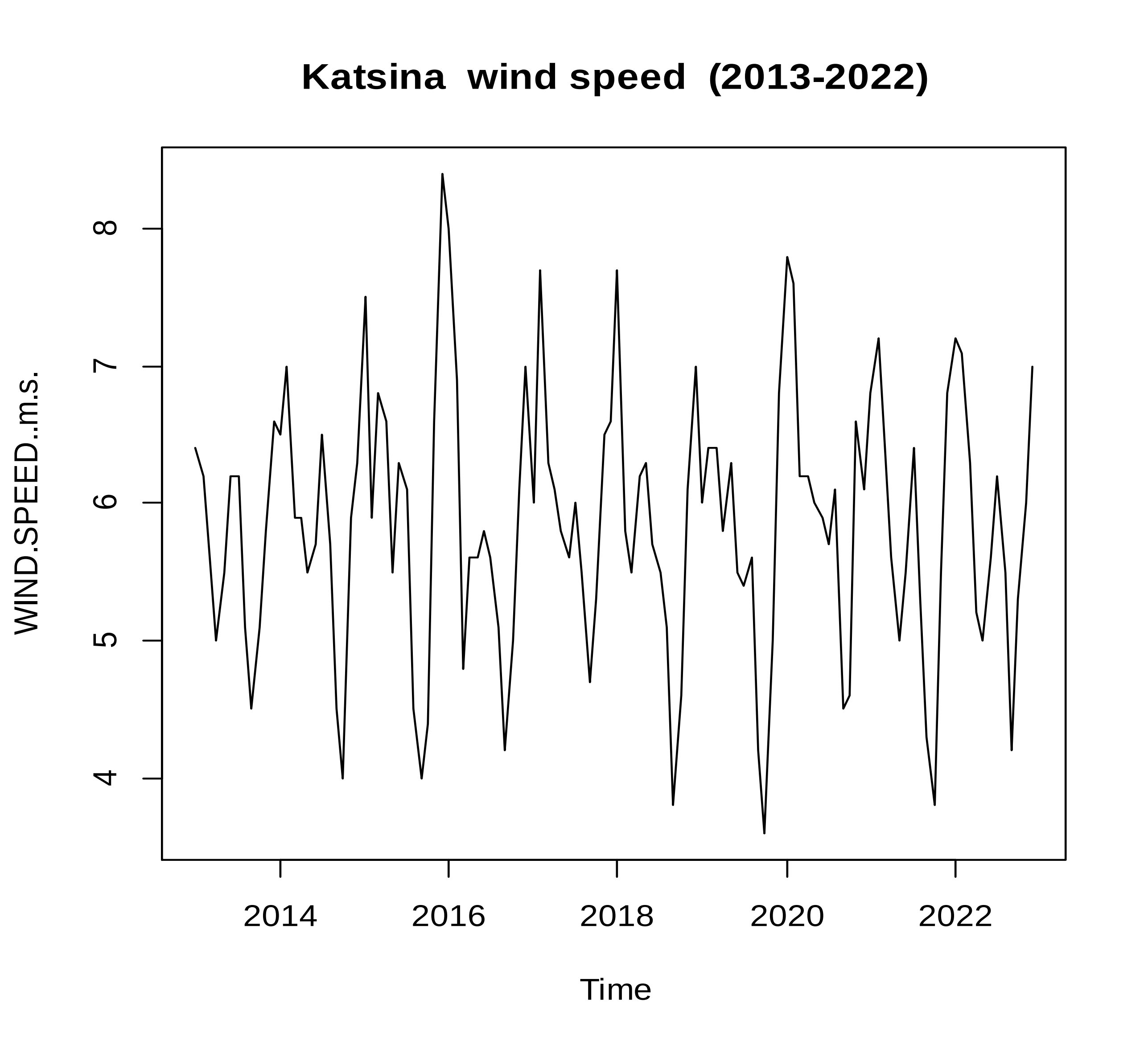

Figure 1 Katsina wind speed time plot

Table 1 Wind Speed DLM Parameters

M_0	C_0	FF	V	GG	W
0	1.0×10^7	1.0	5.466×10^{-8}	1.0	8.346×10^{-1}

Table 2: ETS(M,N,A) Model Parameters

Smoothing parameters:	α	β
	1e-04	1e-04
Initial states:	l	s
	5.8416	1.0249, 0.1724, -1.2289, -1.5814, -0.5131, 0.1024, -0.0813, -0.1926, -0.0186, 0.1965, 0.9625, 1.1573

Table 3: Models Comparison

Model	In-Sample	
	RMSE	MAPE
DLM	0.0023	1.0029
ETS (M, N, A)	0.4987	6.8797

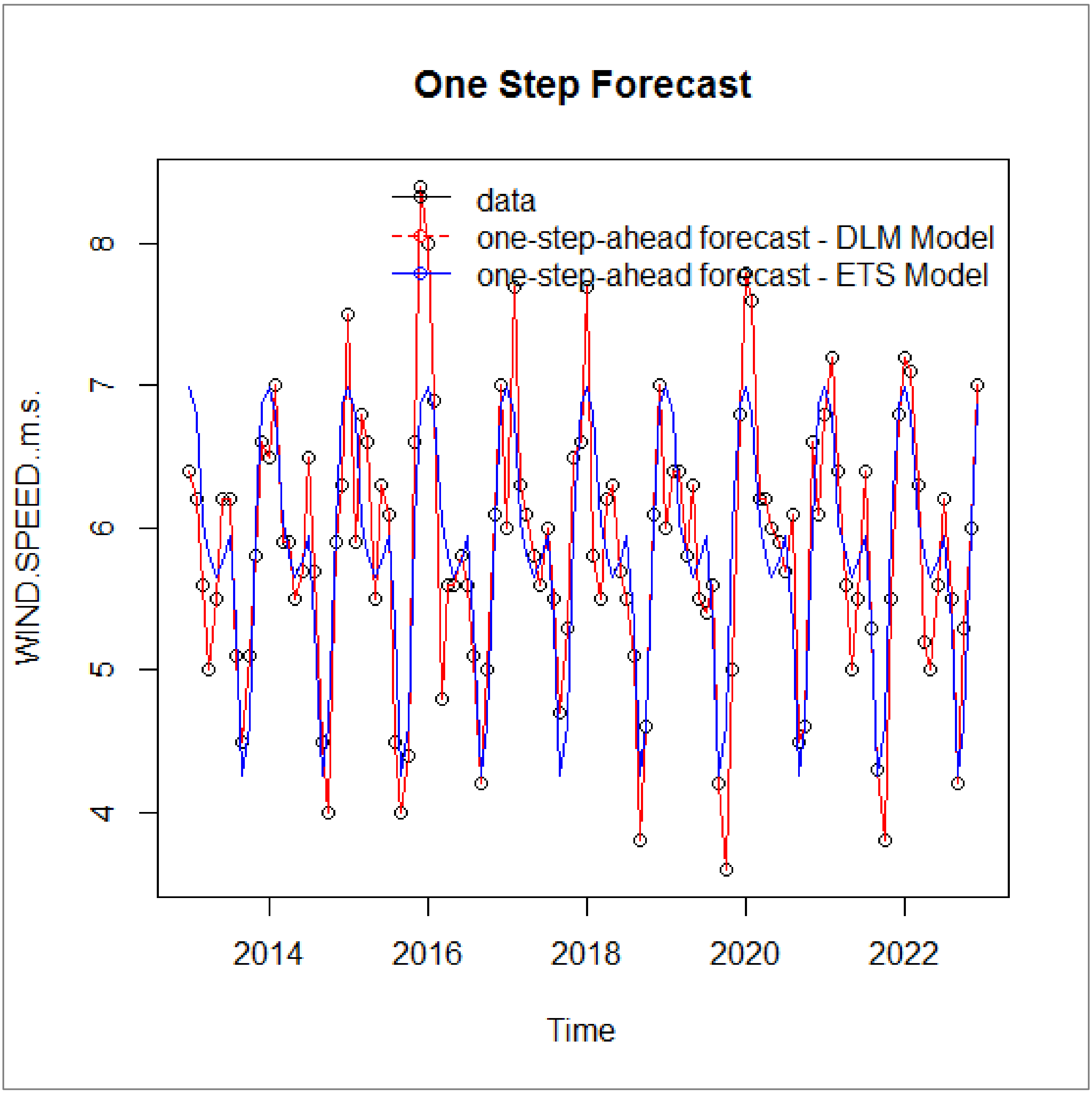

Figure 2 the one-step ahead forecasts plot

CONCLUSION

Based on the forecast plot, Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), it can be concluded that the Dynamic Linear Model (DLM) outperforms the Exponential Smoothing ETS(M, N, A) models in terms of forecasting performance. The DLM provides more accurate and reliable forecasts compared to the ETS models, as evidenced by lower MAPE and RMSE values and a closer fit of the forecasted values to the actual data points on the forecast plot. Therefore, the DLM is the preferred choice for this dataset and forecasting task for the Katsina wind speed generating forecasts.

ACKNOWLEDGMENT

The authors thank Umaru Musa Yar'adua University, Katsina, and the TETFund IBR grant for funding the research.

REFERENCES

Agada I.O., Peter O. and Agada P.O, (2023). Stochastic Modelling of Wind Speed over Northern States in Nigeria Science World Journal Vol. 18(No 3). [Crossref]

Azorin-Molina, C., S. Rehman, J. A. Guijarro, T. R. McVicar, L. Minola, D. Chen, and S. M. Vicente-Serrano, (2018) Recent trends in wind speed across Saudi Arabia, 1978–2013: A break in the stilling. *Int. J. Climatol.*, **38**, e966–e984, [Crossref].

Bessa, R. J., Miranda, V., Botterud, A., Wang, J., & Zhang, J. (2020). Wind power forecasting: A review of the state of the art. IEEE Transactions on Power Systems, 35(1), 425-440.

Chai, Tianyou, et al. (2019)."A review on hybrid forecasting approaches: Principles, applications, and challenges." Renewable and Sustainable Energy Reviews 113: 109300.

Chatterjee, S., & Mondal, S. (2021). Wind speed and power forecasting using hybrid machine learning models: a comprehensive review. Renewable Energy, 176, 624-641.

Chen, X., Wang, Y., Liu, P., & Liu, S. (2019). A Dynamic Linear Model for Monthly Rainfall Prediction Based on the Bayesian Inference Method. Journal of Hydrology, 577, 123998.

Dare, J., Patrick, A. O., & Oyewola, D. O. (2022). Comparison of Stationarity on Ljung Box Test Statistics for Forecasting. *Earthline Journal of Mathematical Sciences*, *8*(2), 325-336. [Crossref]

Hyndman, R.J. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of statistical software. [Crossref]

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. (2008). Forecasting with exponential smoothing: the state space approach. Springer-Verlag, Berlin. [Crossref]

Jiang, Y., Y. Luo, and Z. Zhao (2013). Maximum wind speed changes over China. *Acta Meteor. Sin.*, **27**, 63–74, [Crossref].

Marangon L.M. Lima, E. Popova, P. Damien (2014). Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models, International Journal of Forecasting, Volume 30, Issue3, Pages 464-476, ISSN 0169-2070, [Crossref].

Nobre, F. F., Monteiro, A. B. S., Telles, P. R., & Williamson, G. D. (2001). Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Statistics in medicine*, *20*(20), 3051-3069. [Crossref]

Osthus, D., Caragea, P. C., Higdon, D., Morley, S. K., Reeves, G. D., & Weaver, B. P. (2014). Dynamic linear models for forecasting of radiation belt electrons and limitations on physical interpretation of predictive models. *Space Weather*, *12*(6), 426–446. [Crossref]

Petris, G. (2010). An R Package for Dynamic Linear Models. *Journal of Statistical Software*, *36*(12), 1–16. [Crossref]

Wang, Xuan, et al. (2019). "A hybrid model for wind speed forecasting based on LSTM and ARIMA." Applied Energy 233-234: 124-135.

Wang, Y., Huang, S., Zhou, L., Wang, H., & Li, Z. (2022). A deep learning approach for wind speed forecasting based on long short-term memory neural networks. Energy Reports, 8, 333-342.

ZhangG, C. AzorinMolina, D. Chen, J.A. Guijarro, F. K ong, L. Minola, T.R. McVicar, S.-W. Son, P. Shi (2020). Variability of daily maximum wind speed across China, 1975–2016: an examination of likely causes J. Clim., 33, pp. 2793-2816. [Crossref]

Zhang, J., Cheng, W., Tang, M., Jiang, X., & Wang, G. (2021). A review of wind speed forecasting methods based on machine learning. Renewable and Sustainable Energy Reviews, 135, 110159.