Intelligent Environments 2024: Combined Proceedings of Workshops and Demos & Videos Session
M.J. Hornos et al. (Eds.)

doi:10.3233/AISE240022

A Simplified and Sustainable Approach for Energy Prediction

Zahra ZIRAN [a,1], Massimo MECELLA [a] Francesco LEOTTA [a]
[a] *Sapienza Università di Roma, Rome, Italy*
ORCiD ID: Zahra Ziran https://orcid.org/0000-0002-3529-7380, Massimo Mecella https://orcid.org/0000-0002-9730-8882, Francesco Leotta https://orcid.org/0000-0001-9216-8502

Abstract. Nowadays, efficient management of energy consumption is crucial for the sustainability of our cities, and overall of our planet. Approaches investigated so far, mostly adopt complex approaches, often based on deep learning, which have an important footprint. This study focuses on the importance of using simpler methods to predict energy consumption in smart buildings, emphasizing a methodological approach that prioritizes simplicity, transparency, and computational efficiency, especially when data is scarce. It emphasizes that even the prediction of energy consumption at the scale of a building, which is sometimes ignored due to computational complexity, is feasible and can make a big difference. By using simple analytical models combined with outlier detection, this research contributes to the field by showing how we can still gain valuable insights with limited data. Therefore, this study provides a practical and scalable way to improve energy efficiency and sustainability in buildings, which has a significant contribution to energy management practices.

Keywords. energy management, smart buildings, statistical modeling, predictive analytics

1. Introduction

In the rapidly urbanizing global landscape, the accurate forecasting of energy consumption has emerged as a critical component for achieving sustainable growth. The building sector, a significant consumer of energy, accounts for approximately a fifth of total energy usage worldwide. With projections indicating a steady 1.3% annual increase in energy consumption through 2050, as mentioned in [1,2], the imperative for efficient energy management strategies becomes unequivocal. This context sets the stage for the development and application of advanced predictive analytics, aimed at enhancing the sustainability and operational efficiency of smart environments. Recent research efforts have extensively focused on various methodologies for energy consumption forecasting to facilitate more effective energy management. Notable studies that explore diverse approaches include [3,4,5], highlighting the breadth of investigative work in this area.

In this context, the reliance on deep learning and complex algorithmic models, while offering powerful predictive capabilities, often complicates interpretability and demands

[1]Corresponding Author: Zahra Ziran, *Email:* zahra.ziran@uniroma1.it

extensive computational resources. Addressing these challenges, this study proposes a more streamlined and efficient alternative, employing simple statistical and machine learning models coupled with outlier detection techniques. This methodological adjustment suggests a gradual move towards analytical frameworks that aim to balance simplicity and transparency with computational efficiency, all while maintaining the rigor of energy consumption prediction accuracy. One of the advantages of the methodology considered in this study is its transparency, which plays a crucial role in the adoption and effectiveness of predictive analytics for smart building energy management. In contrast to many advanced artificial intelligence (AI) methods, often criticized for their "black box" nature [6], this research utilizes statistical models known for their interpretability. The clear transparency of these models enables a deeper understanding of the variables and mechanics influencing predictions. This feature is especially valuable as it allows everyone, including those with limited technical expertise, to comprehend, trust, and effectively utilize the findings. By prioritizing methods that offer easy interpretability, this study not only meets the growing demand for accountable and explainable AI but also enhances the practical applicability of its outcomes.

This research is investigated based on the CU-BEMS dataset [7], a repository of energy consumption data derived from the Building Energy Management System (BEMS) developed at Chulalongkorn University. This dataset, meticulously compiled from a seven-story academic office building within the university's campus, encompasses approximately 11,700 square meters and is characterized by a peak energy load of about 700 kW. The granularity of the CU-BEMS dataset, which records energy consumption across various zones of the building, offers an opportunity for a detailed examination of energy usage patterns. The dataset not only includes data on significant energy consumers within commercial settings, such as air conditioning units, lighting, and plug loads, but also captures environmental conditions—temperature, humidity, and ambient light levels. These comprehensive data points provide a multifaceted lens through which the dynamics of energy demand can be explored and understood. The methodology adopted in this study is an analysis, venturing into the realm of data refinement through the strategic implementation of outlier detection techniques. By identifying and mitigating anomalies within the dataset — instances of abnormally high or low energy consumption — this research aims at enhancing the precision and reliability of subsequent energy consumption forecasts. This focus on outlier management is pivotal, given its potential to rectify distortions in predictive models caused by such anomalies, thereby yielding more accurate and actionable insights for energy management.

Opting for a straightforward modeling approach does not imply a reduction in analytical depth or complexity. Instead, employing autoregressive models [8,9] highlights the effectiveness of simplicity in capturing the inherent temporal dynamics of energy consumption data. This methodology provides a clear insight into the cyclical and trend-driven nature of energy usage, further enhanced by incorporating outlier detection to increase both the model's accuracy and its interpretability.

This study's insights, derived from the meticulous analysis of the CU-BEMS dataset and the subsequent predictive modeling exercise, transcend the limitations of a single case study. The methodologies and principles outlined are scalable and adaptable, demonstrating applicability across a wide range of building types and energy management scenarios. Notably, the success of this approach in residential buildings indicates its potential applicability in broader domains within the smart building context. The repeatability of the study is also enforced, by adopting a widespread and open dataset and providing access to all techniques, in order to enforce a truly *open science* approach[2].

The integration of statistical models with outlier detection for accurate energy prediction reinforces the idea that achieving advanced outcomes does not necessarily require complex analytical procedures, especially in the realm of smart buildings. This study promotes a move towards predictive analytics that are more accessible, transparent, and computationally efficient,not only addressing the immediate challenges of forecasting energy consumption but also establishing a foundation for future advancements in the field.

The structure of the paper is as follows: In Section 2, we state the problem and explore the historical context of energy consumption and the various techniques that have been applied within this area, providing a solid foundation for understanding the significance and evolution of forecasting methodologies. Following this, Section 3 details the methodology of the work, outlining the specific approaches and analytical techniques employed in this research to address the forecasting challenges. In Section 4, the results of the study are discussed, offering an in-depth analysis of the findings and their implications for energy management practices. Finally, the paper concludes in Section 5, by summarizing the key insights gained from the research and suggesting directions for future studies in the field of energy consumption forecasting.

2. Literature Review

The global population is increasing each year, which leads to higher electrical energy demand for residential use and national development. Meeting this growing energy demand while protecting the environment is crucial. Hence, forecasting electrical energy consumption is vital for energy conservation and minimizing environmental harm [10,11,12,13]. Research in electrical energy consumption forecasting has developed into various methods. These methods fall into two main categories: artificial intelligence (AI)-based approaches [14,15,16] and traditional methods [17,18].

AI-based techniques utilize deep learning (DL) to predict energy usage accurately, processing large datasets to identify patterns. Although various methodologies such as transfer learning, as discussed in [19,20], offer a valuable solution to compensate for the lack of datasets and have significantly addressed the issue of dataset limitations, the focus of this study remains on utilizing the existing dataset with methods that require less time and resources. Consequently, this research specifically discusses energy consumption forecasting based on the available dataset. In contrast, traditional methods rely on statistical and mathematical models, offering simplicity and reliability without extensive computational demands. Both approaches play a critical role in understanding energy consumption trends and informing sustainable energy management strategies. When deciding whether to prioritize conventional methods over AI/DL techniques or vice versa, the intended application of the method must be carefully considered. The mere fact that AI/DL methods are more modern does not automatically render them superior for all purposes. Notably, traditional approaches continue to play a significant role in predictive

[2]Cf. `https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en` and `https://www.fosteropenscience.eu/content/what-open-science-introduction`.

analytics, which is crucial for attaining precise forecast outcomes. Before AI/DL became prevalent, conventional methods like stochastic time series and regression-based models were extensively utilized. These traditional techniques are particularly adapt at addressing linear problems and offering satisfactory outcomes. The principle behind stochastic time series, for example, relies on the concept of reliable forecasting through the extrapolation of time series models into the future [21]. Furthermore, the regression-based approach offers the distinct advantage of elucidating the relationship between dependent and independent variables. This method allows for the precise determination of how these variables interact, enhancing the predictive analysis framework.

In reviewing the literature on the advantages and disadvantages of artificial neural network methods, as outlined in [22,23], it becomes evident that while neural networks possess robust capabilities in handling non-linear patterns and operational resilience even with certain failures within the network, they are not without significant drawbacks. The autonomous learning capacity of deep learning and its broad applicability across various applications are indeed commendable. However, the methodological intricacies and operational demands of deep learning prompt a reconsideration of their suitability for forecasting tasks, particularly energy consumption forecasting.

The primary challenges associated with deep learning stem from their substantial requirements for training time and the necessity to meticulously emulate the architecture of the neural network for effective operation. Moreover, the scalability of deep learning becomes a concern, as managing a large network necessitates prolonged computational efforts. These constraints underscore the potential limitations of deploying deep learning in environments where computational resources are finite or where rapid deployment is crucial.

Given these considerations, the choice to employ a statistical approach for forecasting energy consumption is driven by a strategic assessment of both the methodological fit and operational efficiency [24]. Statistical methods, by their nature, offer a more straightforward and interpretable framework for predictive analysis. Unlike deep learning, these methods do not require extensive training periods or complex network architectures, making them more agile and easier to implement within a shorter timeframe. Furthermore, the transparency and simplicity of statistical models facilitate a clearer understanding of the underlying relationships between variables, an aspect that is crucial for actionable insights and decision-making in energy management.

Therefore, while recognizing the sophisticated capabilities of deep learning, the preference for statistical methods in forecasting energy consumption arises from the desire for simplicity in methodology, operational efficiency, and clarity in interpretation. These considerations are critical in ensuring the forecasting model's accuracy and practical applicability, given the limitations of resources and the immediate nature of energy management goals.

This study adopted a detailed strategy to utilize the AutoReg model for energy consumption forecasting, focusing on the dataset's time-stamped nature as essential for capturing temporal patterns, following insights from [25]. Recognizing the data's inherent seasonality and trends, adjustments were made according to Box [26], to prepare for accurate modeling. The choice of lags, determined by PACF and the Akaike Information Criterion (AIC), aimed to precisely model the dataset's temporal structure without overfitting, a principle outlined in [27]. Upon implementing the AutoReg model, its predictions were rigorously evaluated against actual data using metrics suited for time series analysis, as recommended in [28]. This approach was not only methodologically sound but also chosen for its practicality and cost-effectiveness after analyzing the dataset's characteristics. It underscores the research's commitment to a balance between accuracy, practicality, and cost-efficiency in forecasting energy consumption.

3. Methodology

3.1. Research Design

This study employs a quantitative research approach to predict weekly energy consumption. Given the objective nature of the data and the research questions, statistical and regression modeling techniques were chosen for their ability to provide measurable and precise predictions. The focus on predictive modeling aligns with the study's goal to identify patterns in energy usage and apply these findings to enhance energy management practices.

3.2. Data Collection

The dataset central to this research is sourced from the Building Energy Management System (BEMS) at Chulalongkorn University. Energy consumption data was collected from a seven-story academic office building on the university campus, which occupies an area of approximately 11,700 square meters and experiences a peak energy load of about 700 kW. The data represents a comprehensive record of the building's energy usage, essential for conducting a detailed analysis of consumption patterns from July 1, 2018, to December 31, 2019. In this study, the overall energy consumption, measured in kilowatts (kW), has been utilized.

3.3. Dataset Preparation

The dataset underwent an extensive preparation phase to ensure its readiness for predictive modeling. Initially, data cleaning was conducted to correct any errors and remove inconsistencies, thereby enhancing the dataset's overall quality. To address missing values, statistical imputation techniques were applied, ensuring that the dataset remained comprehensive and representative of the building's energy consumption patterns. A critical component of the preparation process involved the meticulous identification and treatment of outliers. To mitigate the potential impact of these anomalies on predictive modeling, several statistical techniques were employed. The Interquartile Range (IQR) method was utilized to detect outliers by identifying data points that fell outside the 1.5 * IQR threshold above the third quartile and below the first quartile. Additionally, the Z-score method was applied to identify outliers based on the standard deviation. These techniques allowed for the careful adjustment or removal of outliers, ensuring they did not distort the modeling process. Furthermore, to address the dataset's seasonality and trends—common characteristics in energy consumption data—differencing and detrending methods were applied. This step was pivotal in achieving stationarity, a prerequisite for the effective application of the AutoReg model. By transforming the data to a stationary state, we ensured that the model's assumptions were met, thereby enhancing the accuracy of the predictions generated.

3.4. Predictive Modeling

The AutoReg model was selected for its effectiveness in leveraging temporal dependencies within the dataset to predict future energy consumption. The choice of this model was informed by its proven track record in time series forecasting, offering a balance between simplicity and accuracy. The AutoReg model operates on the premise that current observations in a time series can be predicted from past values. The AutoReg model is defined in Equation 1.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t \tag{1}$$

As described in Equation 1, Y_t represents the forecasted value at time t, α is the intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients indicating the influence of past p observations on the current value, and ε_t is the error term. The selection of an appropriate number of lags (p) is crucial for capturing the dataset's temporal dynamics accurately without overfitting. This selection was guided by examining the Partial Autocorrelation Function (PACF) and employing criteria such as the Akaike Information Criterion (AIC) to determine the best model fit.

After optimizing the model parameters, including the number of lags, the AutoReg model was applied to the prepared dataset. This approach utilized historical data points to generate a prediction, adhering to the linear relationships identified between past and present energy consumption levels. The model's forecasting capability was rigorously evaluated against actual data, ensuring the accuracy of its predictions.

3.5. Realization of the Techniques

In our research, we have used Python and its powerful libraries such as NumPy, Pandas, and Scikit-learn to implement the methods we adopted to predict energy consumption, including auto-regression, on the CU-BEMS dataset. These techniques make it possible to understand and process data. Standard-scaler normalization played an essential role in improving the accuracy of our model. Matplotlib, Seaborn, and Plotly were used for data visualization. Initially tested on the DGX device for functionality, the model is available on Google Colab, expanding research opportunities through the Smart-Home GitHub repository[3]

4. Discussion and Results

In this research, the methodology commenced with training the model on data from the first six months, aiming to predict energy consumption for the first week of the seventh month, cf. Figure 1. Following this initial stage, the training dataset was incrementally expanded month by month, allowing for the prediction of energy consumption in the first week of subsequent months, as shown in Figures 2 to 6, progressing in this manner until the end of 2019. The dataset in the year 2019 served as the foundational basis for this sequential training and forecasting approach. This predicted value was subsequently compared with the actual recorded energy consumption. Similarly, the model training was

[3] Smart-Home. https://github.com/zahraziran/Smart-Home/tree/main.

extended monthly and was performed to predict energy consumption in the first week of the next month. The accuracy of these predictions was evaluated against the actual consumption figures, which made it possible to determine the accuracy percentage of the model prediction. In this study, we have used a standard metric, the coefficient of determination (R^2), to evaluate the accuracy of our prediction model. This measure is important because of its widespread acceptance in quantifying the variance in the dependent variable that can be predicted from the independent variable. Using R^2 enables a standardized approach to measuring the predictive power of our model, and ensures that our analysis conforms to conventional practices. This enables an objective comparison with established criteria in subsequent studies. As shown in Table 1, the prediction accuracy obtained in this study was more than 70% for the five months under review. This level of accuracy can be considered very satisfactory for energy management in smart buildings, especially when forecasting with smaller data sets and considering that the developed model does not use complex deep learning calculations.

The progress of this model and the observed prediction error rate highlight the challenges and limitations inherent in predicting energy consumption with limited data sets and simplified approaches. Nevertheless, the findings of this study contribute by showing that basic models can yield useful insights for energy management in smart buildings. The ability of these models to provide reasonable predictions, despite a reasonable margin of error, emphasizes the feasibility of using such efficient approaches for cost-effective energy management. This is especially important in situations where comprehensive data may not be available, but informed decisions are essential to increasing energy efficiency and sustainability.

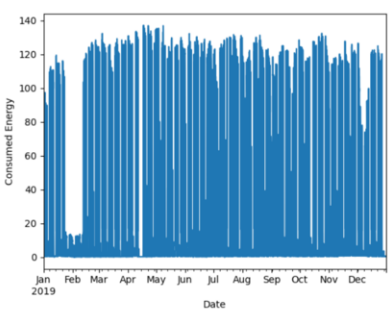

Figure 1. Total Energy consumption in 2019.

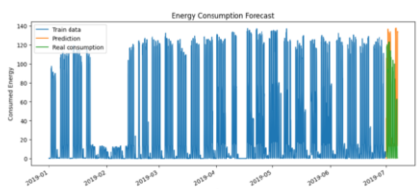

Figure 2. First week of July 2019.

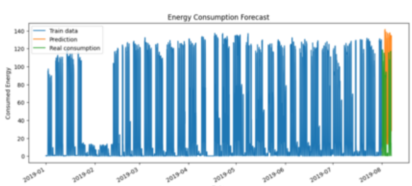

Figure 3. First week of August 2019.

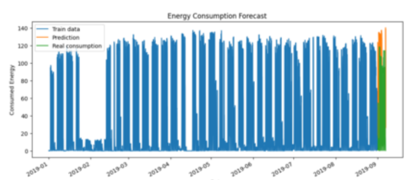

Figure 4. First week of September 2019.

The consequences obtained from these results are into two categories. On the one hand, they emphasize the necessity of developing adaptive forecasting models that can accom-

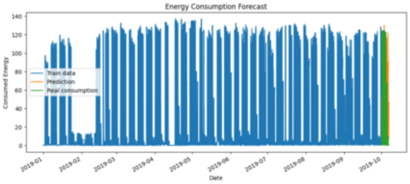

Figure 5. First week of October 2019.

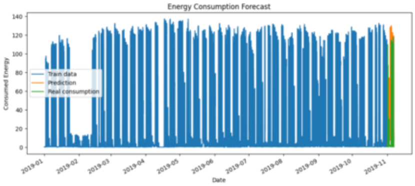

Figure 6. First week of November 2019.

Table 1. Summary of Results.

Prediction Period	Accuracy of prediction (R^2)
From 01/07/2019 to 07/07/2019	79.86%
From 01/08/2019 to 07/08/2019	76.39%
From 01/09/2019 to 07/09/2019	81.57%
From 01/10/2019 to 07/10/2019	85.42%
From 01/11/2019 to 07/11/2019	75.69%

modate different degrees of data availability and complexity. On the other hand, they highlight the ongoing need for model refinement and validation against broader datasets, which could potentially reduce the margin of error and improve prediction accuracy. Essentially, this research lays the foundation for further studies aimed at optimizing energy consumption prediction models, thereby contributing to more sustainable and efficient energy management practices in the smart building sector.

5. Conclusion

In conclusion, this study has underscored the effectiveness of statistical methods for predicting energy consumption in smart buildings, achieving a predictive accuracy of over 75%. This accomplishment highlights the substantial potential of adopting simpler, more transparent analytical models for energy management, especially when faced with limitations such as small datasets and the need to avoid complex computational methodologies like deep learning. The research herein not only enriches the academic discourse around energy consumption forecasting but also provides practical insights for enhancing energy efficiency and sustainability in smart buildings. By demonstrating that advanced outcomes can be achieved without resorting to intricate algorithms, this study advocates for a shift towards more accessible and computationally efficient predictive analytics. Reflecting on the methodology employed in this study, a significant advantage is its understandability, which starkly contrasts with the often opaque nature of complex AI systems. The statistical approaches used here illuminate the prediction processes, offering clarity on how inputs are transformed into forecasts. This transparency is crucial as it facilitates informed decisions based on clear insights rather than obscure computations. Future research will continue to explore and emphasize the value of explainability in predictive models. Furthermore, future studies will not only compare the results obtained from this approach with those derived from deep learning techniques but will also evaluate the efficacy of this method against other predictive models based on statistical and machine learning algorithms. This comparative analysis will further validate the effectiveness and applicability of our approach in various forecasting scenarios.

References

[1] IEA. World energy outlook 2022. IEA Paris, France; 2022. .

[2] Wang H, Ou X, Zhang X. Mode, technology, energy consumption, and resulting CO2 emissions in China's transport sector up to 2050. Energy Policy. 2017;109:719-33.

[3] Himeur Y, Ghanem K, Alsalemi A, Bensaali F, Amira A. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. Applied Energy. 2021;287:116601.

[4] Araya DB, Grolinger K, ElYamany HF, Capretz MA, Bitsuamlak G. An ensemble learning framework for anomaly detection in building energy consumption. Energy and Buildings. 2017;144:191-206.

[5] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy and buildings. 2008;40(3):394-8.

[6] Castelvecchi D. Can we open the black box of AI? Nature News. 2016;538(7623):20.

[7] Pipattanasomporn M, Chitalia G, Songsiri J, Aswakul C, Pora W, Suwankawin S, et al. CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets. Scientific Data. 2020;7(1):241.

[8] Akaike H. Fitting autoregreesive models for prediction. In: Selected Papers of Hirotugu Akaike. Springer; 1969. p. 131-5.

[9] Claeskens G, Croux C, Van Kerckhoven J. Prediction-focused model selection for autoregressive models. Australian & New Zealand Journal of Statistics. 2007;49(4):359-79.

[10] Papakostas K, Mavromatis T, Kyriakis N. Impact of the ambient temperature rise on the energy consumption for heating and cooling in residential buildings of Greece. Renewable Energy. 2010;35(7):1376-9.

[11] Cao X, Yao R, Ding C, Zhou N, Yu W, Yao J, et al. Energy-quota-based integrated solutions for heating and cooling of residential buildings in the Hot Summer and Cold Winter zone in China. Energy and Buildings. 2021;236:110767.

[12] Kim DD, Suh HS. Heating and cooling energy consumption prediction model for high-rise apartment buildings considering design parameters. Energy for Sustainable Development. 2021;61:1-14.

[13] Bazazzadeh H, Nadolny A, Safaei SSH. Climate change and building energy consumption: A review of the impact of weather parameters influenced by climate change on household heating and cooling demands of buildings. European Journal of Sustainable Development. 2021;10(2):1-1.

[14] Niu D, Wang Y, Wu DD. Power load forecasting using support vector machine and ant colony optimization. Expert systems with Applications. 2010;37(3):2531-9.

[15] Escrivá-Escrivá G, Álvarez-Bel C, Roldán-Blay C, Alcázar-Ortega M. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. Energy and Buildings. 2011;43(11):3112-9.

[16] Chen S, Gooi H, Wang M. Solar radiation forecast based on fuzzy logic and neural networks. Renewable energy. 2013;60:195-201.

[17] Nowicka-Zagrajek J, Weron R. Modeling electricity loads in California: ARMA models with hyperbolic noise. Signal Processing. 2002;82(12):1903-15.

[18] Khashei M, Bijari M, Ardali GAR. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). Neurocomputing. 2009;72(4-6):956-67.

[19] González-Vidal A, Mendoza-Bernal J, Niu S, Skarmeta AF, Song H. A transfer learning framework for predictive energy-related scenarios in smart buildings. IEEE Transactions on Industry Applications. 2022;59(1):26-37.

[20] Feng B, Zhou Q, Xing J, Yang Q, Chen Y, Deng Z. Attention-empowered transfer learning method for HVAC sensor fault diagnosis in dynamic building environments. Building and Environment. 2024;250:111148.

[21] Islam BU. Comparison of conventional and modern load forecasting techniques based on artificial intelligence and expert systems. International Journal of Computer Science Issues (IJCSI). 2011;8(5):504.

[22] Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews. 2014;33:102-9.

[23] Damborg M, El-Sharkawi M, Aggoune M, Marks R. Potential of artificial neural networks in power system operation. In: IEEE International Symposium on Circuits and Systems. IEEE; 1990. p. 2933-7.

[24] Daut MAM, Hassan MY, Abdullah H, Rahman HA, Abdullah MP, Hussin F. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. Renewable and Sustainable Energy Reviews. 2017;70:1108-18.

[25] Shanika L Wickramasuriya GA, Hyndman RJ. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. Journal of the American Statistical Association. 2019;114(526):804-19. Available from: https://doi.org/10.1080/01621459.2018.1448825.

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

[27] Brockwell PJ, Davis RA. Introduction to time series and forecasting. Springer; 2002.

[28] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006;22(4):679-88. Available from: https://www.sciencedirect.com/science/article/pii/S0169207006000239.