

International Journal of Development and Sustainability
ISSN: 2186-8662 – www.isdsnet.com/ijds
Volume 14 Number 11 (2025): Pages 935-976
https://doi.org/10.63212/IJDS25102301

ISDS JOURNALS

Understanding time-series building energy data: A review of challenges and forecasting approaches

Mohammed Mahmood Katun [1*], Olawale Muhammed Popoola [1], Rembuluwani Phillip Tshubwana [2]

[1] *Department of Electrical Engineering & Centre for Energy and Electric Power (CEEP), Tshwane University of Technology, Pretoria, South Africa*
[2] *Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa*

Abstract

This article provides an in-depth analysis of building energy consumption, focusing on understanding time series energy data and various predictive techniques. Amidst rapid urbanization and growing energy needs, precisely predicting energy consumption in buildings is critical for improving energy efficiency, fostering sustainability, and optimizing building management systems. The review adopts a narrative review approach, providing a conceptual synthesis of recent advances in building energy prediction by explores the fundamental characteristics of time series electricity consumption data, such as seasonality, trends, non-linearity, and temporal dependencies, spotlighting the challenges these present for forecasting. It covers diverse data acquisition technologies and methods, stressing the common problems of data quality, and effective mitigation strategies through exploratory data analysis and data cleaning techniques. Additionally, the article studies a broad range of time series energy forecasting models, classifying them into statistical, computational intelligence, and hybrid/ensemble approaches. Key considerations for model selection are also explored. The comprehensive forecasting process, from data collection and preparation to model evaluation, is detailed. Ultimately, this study emphasizes that effective data handling and precise model refinement are inherent to accurate prediction, enhancing energy efficiency and effective operations in the built environment.

Keywords: Energy Management; Time Series Energy Data; Building Energy Consumption; Building Energy Forecasting; Predictive Models

Published by ISDS LLC, Japan | Copyright © 2025 by the Author(s) | This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Cite this article as: Katun, M.M., Popoola, M.M. and Tshubwana, R.P. (2021), "Understanding time-series building energy data: A review of challenges and forecasting approaches", *International Journal of Development and Sustainability*, Vol. 14 No. 11, pp. 935-976.

[*] Corresponding author. *E-mail address:* mahmoodmk@tut.ac.za

1. Introduction

Over the previous years, occupant well-being and energy management in buildings have garnered massive academic interest. In the face of increasing urbanization, energy consumption in buildings is projected to rise (Bezzar et al., 2022; El Assri et al., 2025; Maleki et al., 2024; Ramokone et al., 2021). However, the intention of any building is to offer the occupant a pleasant indoor atmosphere, characterized by stable temperature, healthy ventilation and adequate illumination. Consequently, it is critical that buildings are designed and constructed to meet the needs of occupants (Jin et al., 2021).

Globally, buildings are regarded as significant energy consumers (Bezzar et al., 2022; Chen et al., 2025a; Mahmood and Popoola, 2025; Mora et al., 2019; Zhao et al., 2020; Zhou et al., 2023). Their energy needs substantially affect the global energy landscape with a notable portion of energy loss attributed to inadequate occupant management, inefficient appliances and subpar strategy implementation, resulting in energy wastage ranging from 2% to 20%. Global energy consumption is currently growing at a rate of approximately 2.9%, and in the years to come, it is expected to rise much faster (Kadir et al., 2023; Zhao et al., 2020). Buildings emit about 33.33% of all greenhouse gas emissions and consume about 40% of the world's energy (Hussien et al., 2025; Ngo et al., 2022; Ramokone et al., 2021; Zhou et al., 2023). Consequently, the core challenge for Building Management System (BMS) is to optimize energy efficiency without sacrificing occupant comfort (Kadir et al., 2023).

One strategy for ensuring the long-term decarbonization of the energy industry is energy efficiency in buildings; this is in accordance with the International Energy Agency. Building energy efficiency offers significant financial advantages in addition to environmental benefits. Buildings equipped with effective energy systems and management practices tend to have lower operational expenses (Deb et al., 2017). Smart and effective control systems are emphasized as among the best strategies to reduce energy usage throughout a building's life cycle (Maleki et al., 2024). In addition to cutting energy costs, reducing energy reliance, and decarbonization, energy efficiency can help countries sustain or enhance economic activity (Choi et al., 2020; Fan et al., 2019). Subsequently, nations are making energy saving and the construction of sustainable buildings their top priorities (Maleki et al., 2024; Ngo et al., 2022). Meanwhile, in research, academics have proposed different strategies to reduce building energy consumption, including renewable energy sources integration, energy forecasting for buildings, and implementation of high-efficiency building envelopes (El Assri et al., 2025; He et al., 2022; Maleki et al., 2024).

Predicting energy usage in buildings is a crucial step in increasing energy efficiency in present-day technologies (Chen et al., 2025a; He et al., 2022). Precise forecasting serves as the basis for effective energy network scheduling, facilitating the alignment and control of energy supply and demand. This also ensures economic and efficient functioning of building energy systems. Energy sustainability and waste reduction are enhanced by precise demand predictions (Choi et al., 2020; Jogunola et al., 2022; G. Li et al., 2022a; Maleki et al., 2024; Ngo et al., 2022; Popoola et al., 2024; Ramokone et al., 2021). In 2020, Kurte et al reported that through simulation a 30% cost reduction was observed and 21% reduction was recorded on real houses when the ability of pre-trained deep reinforcement learning was evaluated on HVAC-based (Heating, Ventilation, and Air Conditioning) control on a residence (Kurte et al., 2020). Through forecasting the energy consumption of a heat pump system having two tanks was reported to have reduced by 11.5% and its operating cost dropped

by 7.5% (Xie et al., 2020). Also, in a similar effort, a new approach based on energy forecasting and flexibility was presented for optimizing the energy consumption of HVAC, and it was reported that an energy saving of about 10% was experienced (W. Li et al., 2022b). Consequently, according to Wang et al. and Olu-Ayaji et al., accurate forecasting is important as it could lead to a 10% to 30% decrease in building energy usage (Olu-Ajayi et al., 2022; J. Wang et al., 2021). Therefore, to guarantee a sufficient energy supply, cost effective utilization techniques are strongly linked to energy predictions (He et al., 2022; Ngo et al., 2022).

1.1. Energy consumption forecast in building

Energy consumption forecasting is a crucial part of predictive analytics; using historical data to forecast future outcomes. It aids in building energy forecasting, assessing energy-saving potential in a building's design, operation, and retrofitting phases. The building's lifecycle, from design to demolition, significantly influenced the energy consumption. Predicting energy consumption requires understanding the architectural structure, material choices, and user behavior at different stages, including design and operational phases (Kolambe, 2024; Yin et al., 2024). Also, it requires identifying load types for precise analysis. Categorization based on time horizon, building types, and load types helps develop accurate and targeted strategies for load forecasting (Chen et al., 2025a). Researchers have focused on enhancing building energy efficiency through forecasting for demand response and management applications (Runge and Zmeureanu, 2019). Building energy consumption is typically predicted using two primary approaches; constructing realistic physical models for engineering simulations and using data-driven models (Chen et al., 2025a; Gao et al., 2025).

Over the past 25 years, building energy predictive research has made extensive use of techniques that leverage historical data, which are commonly referred to as “computational intelligence models” (Deb et al., 2017). Nevertheless, once a building becomes operational, various factors influence its energy performance, including weather conditions, occupancy patterns, the thermal characteristics of building envelope, and the intricate interactions among energy systems such as heating, ventilation, air conditioning (HVAC), and lighting (Deb et al., 2017; Popoola, 2018; J. Wang et al., 2021). According to the study done by (Popoola et al., 2018), lighting load forecasting is affected by various unpredictable (random) factors. The challenges of developing lighting energy models stem from their non-linear nature, variability, uncertainty, and the capacity to learn from past data. Additionally, research has indicated that predicting how occupants interact with buildings is complex. According to literature, the difference between predicted and actual building energy use is mainly due to occupant behavior and weather conditions. The volatility and non-linearity of energy use have driven research into models that can forecast load profiles by considering multiple variables and past data (El Assri et al., 2025; Maleki et al., 2024; Popoola, 2018; Popoola et al., 2024).

In building energy management domain, attempts are being made for precise prediction using different computational intelligence models or their hybrids however, the stochastic nature of the determinants that affect the energy records coupled with the temporal dependencies of the times series data has made it difficult for most of the AI modeling techniques to smoothly forecast building energy usage (Anan et al., 2024; El Assri et al., 2025; Ngoc-Tri et al., 2022). To this effect, this article presents an understanding of time series electricity consumption data.

The significance of time series electricity consumption data has increased owing to the rising concern on real-time data acquisition. With the introduction of additional technologies in building monitoring, there is a corresponding rise in the volume of time series data gathered and their erratic nature. Factors that influence energy consumption data, its characteristics, nature and patterns were reviewed and discussed in this paper.

The rest of this paper covers the method used to identify the need for this study in section 2.0. Section 3.0 presents time series energy data including its characteristics (3.1), data acquisition and technology (3.2), challenges with sensor technologies (3.2.1), and its mitigation (3.2.2). Times series energy forecasting is discussed in section 4.0 alongside classification forecasting models (4.1), different forecasting horizons (4.1.1), model selection approach (4.1.2), and steps involved in forecasting (4.2). Finally, the overall findings during the course of this study are presented in section 5.0 which forms the conclusion and recommended future work.

2. Current review articles on building energy forecast as related to the study

In other to identify recently published review articles on predicting building energy consumption, a keyword-based search was employed to select related publications in line with the title of this study. To the importance of this research activity, publications from 2015-2025 were considered to understudy areas covered as relates to the subject of this research, and also present an objective and detailed overview of available review literature on forecasting building energy consumption. More so, building energy forecasting has advanced rapidly in recent years, this review concentrates on post-2014 studies to represent the most current scholarly directions and techniques in time-series prediction. The database that was considered for search includes Google Scholar, Scopus and Web of Science using the keyword such as "review on forecasting building energy consumption" (or in parts; "building energy prediction," "time-series data," "machine learning," "hybrid models," and "energy forecasting"). Unlike systematic reviews, this paper does not apply rigid inclusion/exclusion criteria or PRISMA screening. Instead, it organizes insights thematically focusing on data characteristics, forecasting challenges, and integrative modeling perspectives to advance conceptual understanding and identify key research gaps. A total of 28 review studies (as in Table 1) were used to establish the need to embark on this present study.

Table 1. Summary of Reviewed Articles on Building and Broader Energy-Forecasting Literature (2015–2025)

Reference	Primary Focus / Contribution	Main Limitation or Scope Boundary
(Abu Bakar et al., 2015)	Introduced Energy Efficiency Index (EEI) as a benchmarking approach for building consumption analysis.	Focused mainly on quantification metrics, not predictive modeling.
(Mat Daut et al., 2017)	Compared conventional vs. AI-based electricity-consumption prediction in buildings.	Limited to AI models on building electricity; lacks hybrid scope.

Reference	Primary Focus / Contribution	Main Limitation or Scope Boundary
(Deb et al., 2017)	Comparative review of nine major time-series forecasting techniques for buildings.	Narrow model coverage; limited cross-domain generalization.
(Y. Wei et al., 2018)	Data-driven approaches for prediction and classification under diverse building archetypes.	Focused only on data-driven methods.
(Seyedzadeh et al., 2018)	ML (Machine Learning) methods (ANN (Artificial Neural Networks), SVM (Support Vector Machines), Gaussian regression, clustering) for building energy prediction.	Restricted to ML models; excludes DL (Deep Learning) and hybrid frameworks.
(T. Ahmad et al., 2018)	Identified research gaps in urban/rural energy-performance analytics using data-driven tools.	Summarized techniques but lacked methodological depth.
(Runge and Zmeureanu, 2019)	Reviewed ANN applications for building-energy demand.	Focused solely on ANN; excluded comparative modeling.
(Bourdeau et al., 2019)	Data-driven models for building energy consumption forecasting.	Lacked physical and hybrid model coverage.
(Fathi et al., 2020)	Systematic review on urban-scale ML forecasting (2015–2018).	Limited to ML; excludes emerging hybrid approaches.
(Mariano-Hernández et al., 2020)	Forecasting methods in intelligent buildings (residential & non-residential).	Restricted to physical & ML methods.
(T. Ahmad et al., 2020)	Reviewed eco-energy sources and electricity-consumption models (ANN-based).	ANN-only focus; omits other AI types.
(Qiao et al., 2021)	Systematic 30-year review of building electricity-use prediction.	Concentrated solely on electricity demand models.
(Lu et al., 2022)	Bibliometric survey of ANN use for electricity forecasting.	Bibliometric only; lacks technical synthesis.
(Chen et al., 2022)	Compared white-, grey-, and black-box modeling strategies.	Descriptive; limited application evidence.
(Ardabili et al., 2022)	Comprehensive analysis of ML & DL for building-energy management.	Excludes physical and statistical baselines.
(Khalil et al., 2022)	Highlighted computational-intelligence models for electricity consumption.	Predictive models only; lacks validation analysis.
(Olu-Ajayi et al., 2023)	Assessed data-driven tools under different data and building conditions.	Framework guidance only; no empirical performance.
(H. Liu et al., 2023)	Review of 116 studies on data + ML algorithms across time scales.	Focus on ML; limited coverage of hybrid models.
(Chen et al., 2023)	Examined interpretable ML for building-energy prediction.	Narrow focus on model interpretability.

Reference	Primary Focus / Contribution	Main Limitation or Scope Boundary
(Yin et al., 2024)	Comprehensive review of ANN models (2015–2023).	ANN and hybrid/ensemble only; lacks statistical models.
(Mathumitha et al., 2024)	DL forecasting for intelligent-building energy use.	Limited to DL approaches.
(Boukaf et al., 2024)	Digital-twin applications for electricity-consumption forecasting.	Focused mainly on digital-twin architecture.
(Mystakidis et al., 2024)	Survey of statistical, ML, DL & hybrid forecasting techniques.	Broad but high-level; lacks building-specific metrics.
(Eddaoudi et al., 2024)	Concise summary of recent ML forecasting techniques.	Brief overview; no deep comparative analysis.
(Chen et al., 2025a)	Compared data-acquisition methods & prediction models for building loads.	Focused on model contexts; lacks cross-validation.
(Chen et al., 2025b)	ML & DL forecasting models with feature-selection discussion.	Theoretical; limited empirical detail.
(D. Li et al., 2025)	ML methods for fault detection & energy scheduling.	ML-centric; omits other model classes.
(Z. Wang et al., 2025)	Ensemble learning models for building electricity forecasting.	Concentrated on ensemble modeling; limited broader context.

The table summarized the objectives of each selected article and their limitation. From the review, recent efforts are mostly centered and limited to types of models used for forecasting building electricity consumption, what informs the selection of a particular model and the area of application of a model. Some reviews looked at types of building and their influence on energy data and model selection. However, a comprehensive study on the characteristics, nature and patterns of building electricity consumption data was not identified amongst the review therefore are rarely studied in recent times. The characteristics, patterns and nature of data generated on electricity consumption are important information that a researcher needs to understand before venturing into BEMS and predicting building energy consumption. Electricity consumption dataset of building and its forecasting is essential for enhancing electricity management in the building. To this effect, this review focuses specifically on the time series of building energy data and forecasting aspects associated with energy consumption in buildings.

3. Time series energy data

Sustainable energy supply and efficient use of energy are closely connected to accurate forecasting (Choi et al., 2020; Deb et al., 2017; Popoola et al., 2024). Buildings are thus being observed in real-time in order to forecast their energy consumption. This monitoring not only facilitates critical assessments but also offers historical data essential for analyzing and predicting the energy usage of the building (Deb et al., 2017). A time series is a quantitative sequence of data points that capture the dynamics in a system over time. Time series energy consumption data is therefore data points that are indexed or arranged in chronological order. Because it has

a time component, a time-series dataset differs from other datasets (Ghosh, 2021). Table 2 presents various times series data generated in different application domain.

Table 2. Cross-domain context for time-series data generation.

Application Area	Exemplary Data Types and Sources
Medical sciences	Public health data, patient vital signs, medication adherence data, medical imaging data, ECG (Electrocardiogram) and EEG (Electroencephalogram) data, disease progression data, blood glucose data and physiological data (Batoure Bamana et al., 2024; Soyiri and Reidpath, 2013; "Time Series Analysis and Some Applications in Medical Research," 2022).
Energy management	Energy consumption data, demand response data, renewable energy data, industrial energy data, grid energy data, energy storage data, energy production data, building energy data, occupancy data, occupant activity data etc.
Business analytic and insights	Marketing data, operational data, financial data, sales data, customer behavior data, economic indicators data and web analytic data (Fatima & Rahimi, 2024).
Weather forecast	Temperature data, humidity data, atmospheric pressure data, weather patterns data, precipitation data, solar radiation data, wind data and climate data (Zhang et al., 2025).
Value chain management	Demand data, supply data, production data, seasonal data, lead time data, shipping and transportation data, lead data, order data and inventory data (Babai et al., 2019; Fatima and Rahimi, 2024).
Financial markets	Tick data, volatility data, intraday data, week/monthly data, daily data, volume data and open-high-low-close data (Jenčová et al., 2025).
Traffic management	Pedestrian data, cyclist data, signal control data, incident data, travel time data, traffic congestion data and traffic speed data (Runpeng and Seong-Yoon, 2025; L. Wang et al., 2022).

3.1. Characteristics of time series energy data of a building

The energy consumption pattern of a building occupant, whether residential or commercial (offices, shopping malls, or educational institutions), typically display seasonality; showing different usage patterns in different times of the year (long-term scale), months, or days (short-term scale), hours of the day, including holidays. Also, electricity consumption may reveal long-term trends, such as an increase or decrease in consumption over time. Time series energy consumption data possess a distinct feature of linking observed values to time lag from prior records, while the statistical properties of the data evolve over time, thus demonstrating temporal dependence. The relationship between past and present energy consumption data is referred to as temporal dependencies. Due to the presence of unpredictable factors (such as occupancy and occupant

behavior, weather, income etc), time series energy data frequently exhibit nonlinearity. The seasonality, trends, nonlinear relationship and other anomalies of the energy-related variables exhibit complex patterns in times series energy data (Bouktif et al., 2020; Deb et al., 2017; Isah et al., 2023). These attributes make modeling and forecasting of sequential data a challenging task. Consequently, adept data preprocessing is essential, along with specialized models capable of capturing its unique characteristics and intricacies. Figure 1 illustrates the time series energy consumption graph plus the characteristics and features of the dataset.

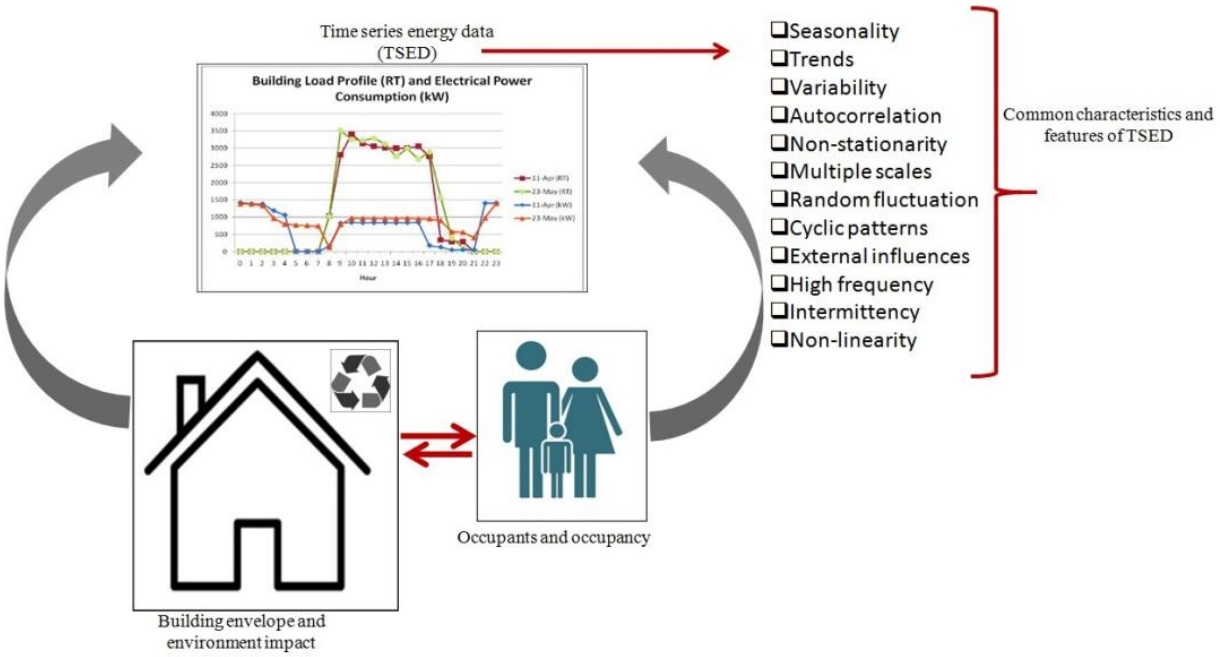


Figure 1. Time series energy consumption graph, characteristics and features of the dataset.

Time series electricity consumption data, especially in the context of buildings, demonstrates distinct characteristics and patterns. Energy consumption in buildings is recorded as a time-ordered sequence of discrete measurements, and analyzing this data usually involves breaking down the dataset (refer to section 4.0) (Deb et al., 2017). In empirical studies, time-series characteristics such as seasonality and temporal dependence are often detected using diagnostic tools like autocorrelation or seasonal-trend decomposition, which guide appropriate model selection and parameter tuning (refer to section 4.1 for linking time series data characteristics to forecasting models).

3.2. Time series data acquisition and technologies for buildings

In many ways, customer information enables energy managers to make data-driven decisions tailored to improve energy efficiency. Obtaining this information necessitates collecting specific data on individual customers, typically for research or energy efficiency programs (Beckel et al., 2014). Occupant behavior towards energy is usually influenced by its environment, activities, income, building envelope and demography (as in Figure 2) thereby making it aperiodic and stochastic. Consequently, monitoring plays a vital role, firstly

to identify energy patterns as relates to consumer's interactions induced by the determinants, and secondly, to verify the accuracy of the stochastic energy models that have been established with actual measurements (Mylonas et al., 2023). Researchers have suggested and used two common monitoring approaches; physical monitoring and occupant investigation. Sensors and energy meters are categorized as physical monitoring devices. They take records of energy consumption of the building, the building's indoor and outdoor environment, and occupancy. The information about the occupant's activities, habits, income, demography and interaction with the building envelope is the occupant investigation approach. These data are obtained through on-site visits, diaries entries, online voting, questionnaires, surveys and interviews (Alanne and Sierla, 2022; Guerra-Santin and Tweed, 2015; Mylonas et al., 2023). The primary objective of monitoring is to collect precise time series data that reveals insights into occupant activities, habits, energy consumption and interactions with building envelope and spaces for future prediction (Mylonas et al., 2023).

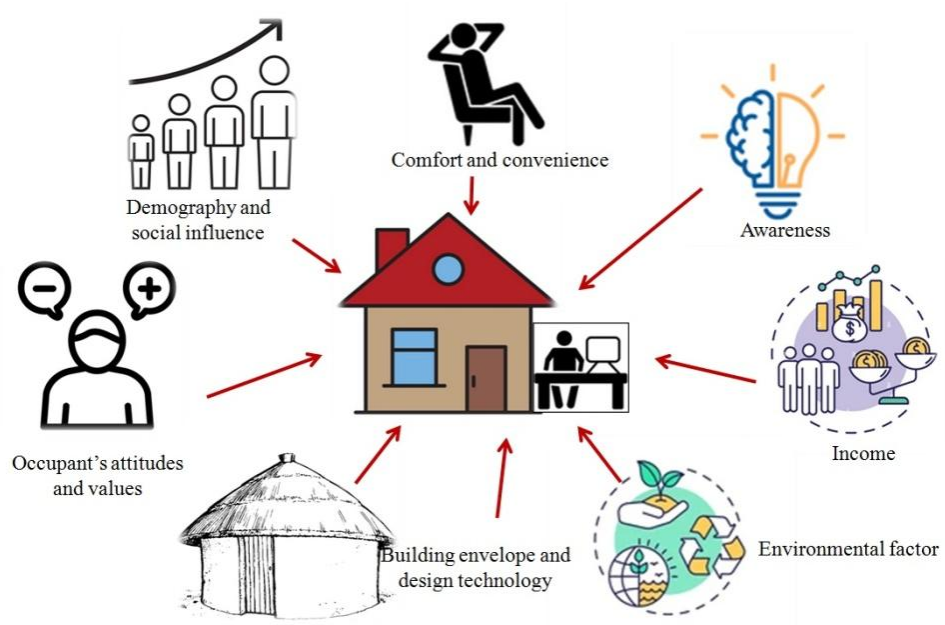


Figure 2. Variables that influence building user patterns towards energy utilization.

However, occupant investigations can be challenging due to their time-consuming and costly nature, as well as occupants' potential hesitation to share personal data. A researcher or an energy manager is left to struggle with self-reporting bias of the occupant, participant fatigue and data analysis challenge that comes with time-use data. To improve data quality and accurate prediction, technological methods which include IoT, sensors and smart energy meters are stressed in the literature for both physical monitoring and occupant investigation approaches (Alanne and Sierla, 2022; Kadir et al., 2023; Kim, Jung, et al., 2019; Mitra et al., 2020; Ramokone et al., 2021). Connecting to the Internet of Things (IoT) facilitates online data exchange between the building automation system and possibly any component within the building, such as lighting fixtures, air condition unit, heater system, and occupancy sensors. Additionally, it provides access to big datasets, particularly those related to weather, sensor data, location-specific information, street view photographs, or building codes

(Daissaoui et al., 2020; Martín-Lopo et al., 2020). In 2019, Wang et al. (Z. Wang et al., 2019) utilized video-embedded remote sensors to gather data regarding the number of occupants and created a model for predicting electricity consumption based on this data. In a similar vein, Wei et al. (Y. Wei et al., 2019a) employed CO2 concentration data to ascertain the presence of occupants and incorporated the detected occupancy as input for their electricity load prediction model. To demonstrate the importance of integrating Internet of Things (IoT) devices and technologies for the prediction of electric energy consumption, (Bedi et al., 2020) used a laboratory system featuring integrated, intelligent monitoring and control functions. The laboratory is equipped with various IoT devices such as smart power meters, occupancy sensors, smart thermostats, smart luminaire controllers, smart switches, a gateway, and a network management system for efficient monitoring and control of electrical loads. To predict electric load in a building, (Das et al., 2020) used three different kinds of sensors: plug load sensors, PIR (Passive Infra-Red) sensors for the occupant presence, and environmental sensors to record the illuminance (lighting status), temperature, and humidity. Additionally, Yi et al. introduced an innovative method for detecting occupant numbers through a CNN model, which relies on data obtained from PIR sensors and smart meters (Wu et al., 2024).

The development of smart devices has made it easier to monitor and record high-precision, long-term building data, enabling more accurate analysis and informed decision-making. (W. Li et al., 2021). Table 3 presents data collection methods for times series data for building energy consumption forecast.

Table 3. Practical building-specific data-collection instruments and their implications.

Technique/instrument	Application
Energy meters Plug-in meters Appliance audit	Energy metering serves to quantify the energy supplied, whereas energy sub-metering is utilized to analyze the distribution of energy usage within a building. This differentiation is crucial due to the contrast between regulated (building-related) and non-regulated (user-related) energy utilization as specified in building regulations and certifications (Bedi et al., 2020; Tekler and Chong, 2022). Plug-in meters have become increasingly popular and are now easily accessible. However, a significant drawback is that multiple meters are often necessary within a single building, which may disrupt the activities of the occupants. Conducting an appliance audit is generally simpler and more cost-effective than using plug-in meters. This process essentially involves calculating energy consumption based on the number and type of electrical devices, their voltage, and the duration of their use. The period of use must be assumed, observed by an examiner, or estimated by the building's occupants (Das et al., 2020; Tekler and Chong, 2022).
Questionnaire Real-time questionnaire, Retrospective questionnaire	Surveys regarding occupant comfort, activities, income levels, and energy usage can be conducted retrospectively. This approach involves asking users to evaluate their comfort and energy consumption patterns over the

Technique/instrument	Application
	past week, month, season, or year. A significant drawback of this method is the potential lack of objectivity from the user. Additionally, surveys on comfort, activities, and energy usage can be performed in real-time. In this case, users provide ratings of their comfort and energy consumption at this moment' within a specific room where indoor measurements are being collected, including the number of occupants present. Such surveys are often utilized for detailed studies related to comfort theory, adaptation, and similar topics. Real-time responses can also be integrated with long-term measurements, provided that the precise timing of the questionnaire's administration is documented (Guerra-Santin and Tweed, 2015).
Heat and water flow meters	Heat and water flow meters are employed to measure the amount of heat or water distributed to a specific area of the building. This is particularly beneficial in large structures, apartment complexes, or scenarios where the positioning of the monitored areas is crucial for energy usage (Guerra-Santin and Tweed, 2015; Kadir et al., 2023; H. Liu et al., 2023).
Sensors in windows and doors	Sensors mounted on windows and doors are important for evaluating indoor air quality and ventilation. This time series data can also be used to explore the correlation between electricity consumption and comfort (Guerra-Santin and Tweed, 2015; Tekler and Chong, 2022).
Presence, light and sound sensors	Movement sensors serve as valuable tools for examining how spaces within buildings are utilized, offering an objective alternative to relying on an occupant presence diary. Light sensors can assess the illumination levels within the building to ascertain if natural light is adequate or to identify when artificial lighting is employed. Measurements of indoor lighting, in conjunction with outdoor lighting conditions, can yield insights into the use of curtains and blinds within a building. Sound sensors are capable of measuring noise levels originating from both indoor and outdoor sources. All the data from these sensors facilitate the analysis of the energy utilization (Guerra-Santin and Tweed, 2015; Kadir et al., 2023; H. Liu et al., 2023).
Daily logs/diaries	Daily logs or diaries serve the purpose of examining how occupants manage the systems within a building, such as the operation of the cookers and water heaters, the opening of windows and doors for ventilation, among other activities. The process involves each occupant completing a form for all systems that need to be investigated. Occupants are requested to do this for a short time (Guerra-Santin and Tweed, 2015).

Multiple data collection technologies are employed, such as surveillance cameras, passive infrared (PIR) sensors, Wi-Fi, reed switches, Radio-Frequency Identification (RFID), ultra-wideband (UWB), and anemometers, alongside other online network technologies(Guerra-Santin and Tweed, 2015; H. Liu et al., 2023; Mitra et al., 2020; Ramokone et al., 2021). The data obtained are either used as a single data (univariate) or multiple data (multivariate) for prediction. It is popular knowledge that the more historical data collected plus related variant data, the better chance to achieve accurate prediction. Datasets with adequate information is always required for most predicting models (He et al., 2022; Khalil et al., 2022). Occupancy-related data including its environment and energy consumption data exhibit temporal patterns and trends, which can be analyzed using time series analysis techniques. The application of time series analysis to occupancy and energy data has a long history in predictive studies. When multiple variables or features are used to predict a target outcome, it is referred to as Multivariate prediction. This approach can provide more accurate predictions than univariate models, as it considers the relationships and interactions between multiple variables (Wu et al., 2024). From a time-series data analysis point of view, the multivariate data patterns are either fed to analytical model or a machine learning model for electricity forecasting. Numerous predicting models are used to estimate the future electricity demand and occupancy (Ullah et al., 2021).

3.2.1. Challenges with sensor technology in raw data collection

In the wake of Internet of Things (IoT) and wireless networks, sensors have grown to be essential in the development of the digital society. Its applications are felt across diverse sectors, such as engineering, energy management, smart cities, transportation systems, sciences, and security (Djedouboum et al., 2018; Felipe and Emilio L., 2022). Data from a variety of IoT-based sensors is being integrated into building management systems. Data gathered from Internet of Things sensors on building equipment allows us to determine environmental changes and how they affect occupant behavior and energy use. Precise data collection from specific targets using a variety of sensors and devices is the main goal of smart building monitoring (Lavrinovica et al., 2024). For example, Energy forecasting in buildings is conducted using different kinds of sensors to monitor and collect data. These sensors include CO_2 sensors, motion sensors, air flow sensors, heat sensors, and energy-related data obtained from smart meters as illustrated in Figure 3 (Ramos et al., 2020).

Consequently, the quality of sensor data is crucial in time series forecasting, as poor data quality renders the process ineffective. A sensor is a device designed to identify and quantify physical parameters (like temperature, humidity, and light levels) or changes in the environment, transforming this data into a form that machines can interpret and analyze. This transformation enables efficient calculations and predictions based on the real-time data. Data analysis encompasses the application of statistical tools, AI methods, signal processing theory, data classification, and various other techniques (Djedouboum et al., 2018; Mboweni et al., 2023; Teh et al., 2020). Nevertheless, gathering raw data from installed sensors for predictive models encounter various challenges, such as data quality problems like incompleteness, discrepancy, and anomalies; potentially resulting to distorted estimation (Alexiou et al., 2023; Nugroho, 2023).

Time series data is utilized to predict future electricity demand and consumption. Various traditional and AI methods are employed to forecast time series data (Maleki et al., 2024). However, managing data quality

(DQ) presents a significant obstacle in sensor technology. The quality of data is always compromised due to environmental impact, network disruption, and sensor failure, leading to measurement errors (Q. Fu et al., 2024; Teh et al., 2020). Sensor inaccuracies can considerably affect both time series data and the precision of energy forecasts. Unreliable sensor readings or missing data can skew the fundamental patterns in time series, resulting in forecasting models that are unreliable (Teh et al., 2020; Yu et al., 2025). Grasping the distinct categories of sensor errors and their impacts is essential for creating reliable forecasting systems (Teh et al., 2020).

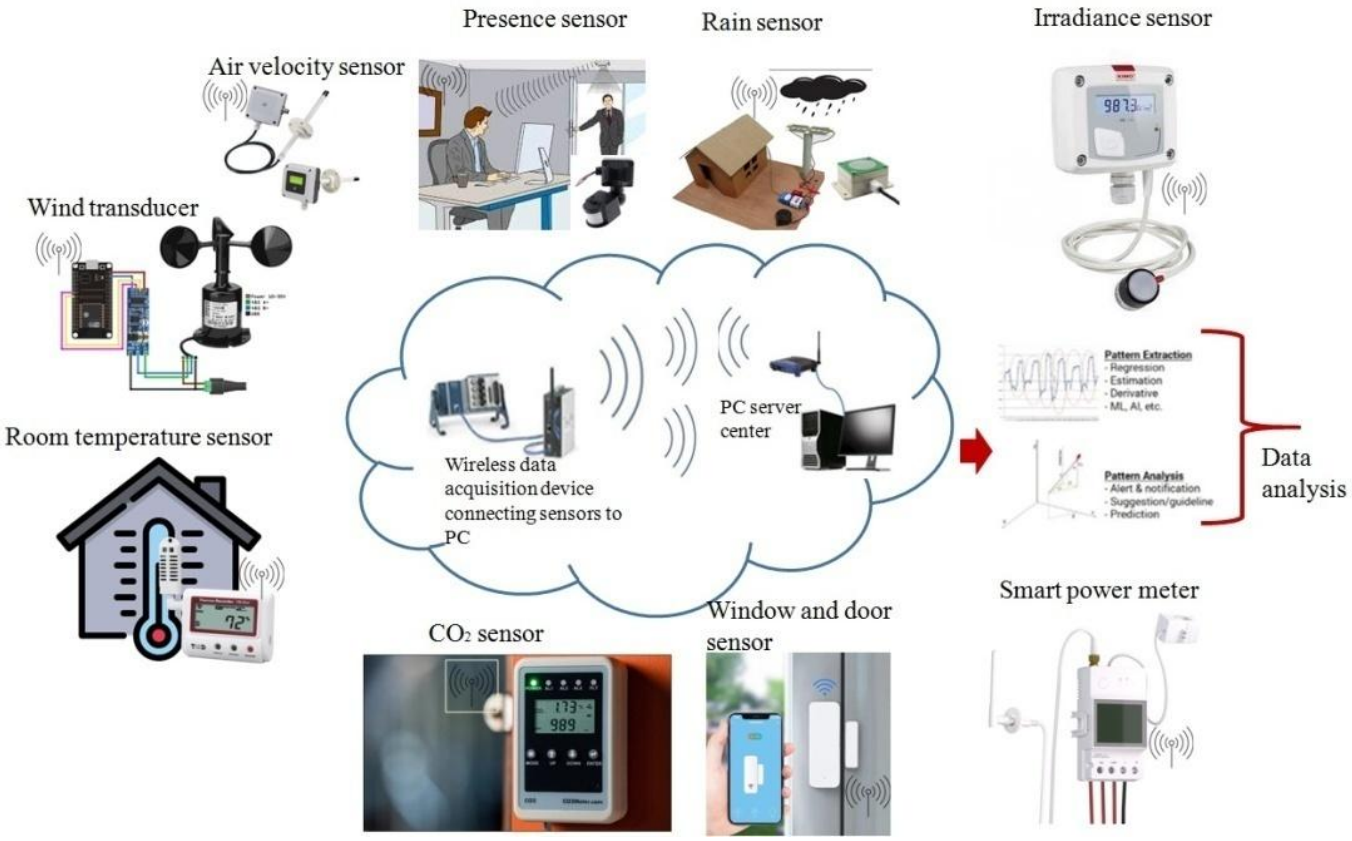


Figure 3. Illustration of IoT sensor technology and network for data collection.

As illustrated in Figure 4, there exist multiple types of errors associated with sensor data quality. These irregularities (drift, missing data, noise, bias, and others) in datasets are referred to as erratic data pattern in time series data. In times series energy forecasting, these unpredictable and irregular fluctuation in energy consumption data are influenced by different factors such as change in energy demand, data quality issue, weather variability and senor malfunctions. Erratic time series data patterns make it difficult to develop accurate forecasting models especially the traditional forecasting techniques. This eventually results to increased forecast errors, inefficient energy resource allocation and higher operational cost (Ahn et al., 2021; AlSalehy and Bailey, 2025; Lee et al., 2020; Romanuke, 2021).

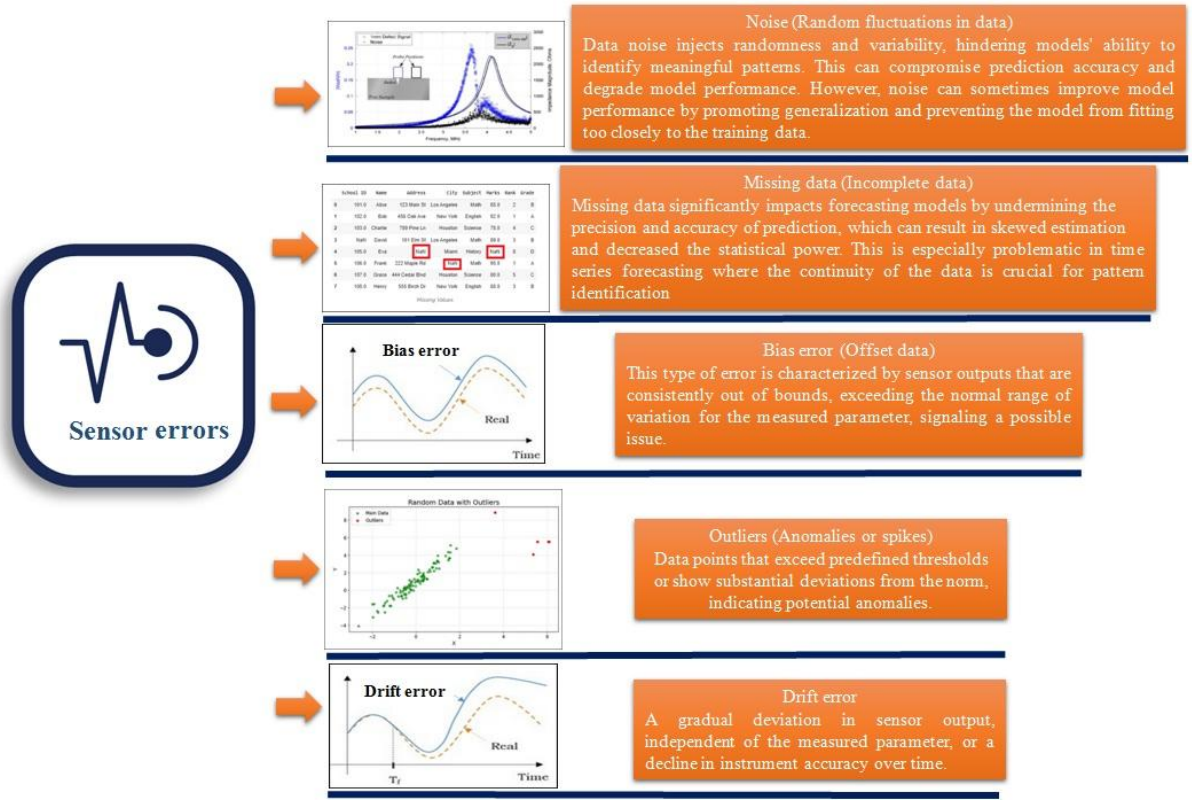


Figure 4. Sensor errors that shape the pattern of energy consumption data; other errors include constant value, uncertainty, stuck-at-zero and freeze value.

3.2.2. Mitigation strategies of erratic patterns in time series dataset

When encountering erratic data, it's important to verify the data accuracy through exploration techniques/data cleaning and adapt models or algorithms to better capture the new trends. Exploration techniques for time series energy data involve methods and tools used to analyze, visualize, and understand patterns, trends, and anomalies in energy-related data. The techniques help identify relationships (correlations between variables), seasonality, and irregularities, informing data-driven decisions (Shin et al., 2024; X. Wang and Wang, 2019). Table 4 presents key exploration techniques (or Exploratory Data Analysis) for time series energy data.

Table 4. Exploratory data-analysis techniques for time series energy data (Jason, 2018; Thangjam et al., 2023)

Technique	Approach
Visual inspection	Plotting data over time to identify trends and patterns.
Summary statistics	Calculating mean, median, and standard deviation to understand data behavior.
Seasonal (time series) decomposition	Separating trend, seasonality, and residuals.

Technique	Approach
Autocorrelation analysis	Measuring correlation between lagged values.
Spectral analysis	Decomposing time series into frequency components.
Anomaly detection	Identifying unusual patterns or outliers.
Feature engineering	Creating new variables to capture temporal relationships.

3.2.3. Data cleaning

When analyzing erratic time series energy data, effective data cleaning is paramount. It not only establishes a foundation for accurate load forecasting but also enhances data reliability and overall efficiency. Therefore, before applying any algorithm, thorough data cleaning and exploration techniques are essential to mitigate the impact of outliers, missing values, and noise. For example, a study forecasting power demand for an office building with an ARIMAX model demonstrated that simply imputing missing occupancy data (from network logins and motion sensors) using the mean of non-missing values improved forecasting accuracy, reducing the Mean Absolute Percentage Error (MAPE) from 1.244% to 1.217% (Newsham and Birt, 2010). Also in another development, a Support Vector Regression (SVR) model demonstrated drastic improvement when cleaned via input filtering and instance selection. This data cleaning lowered the MAPE from 14.32% (using all non-filtered inputs) to 0.06% (Massana et al., 2015). Each chosen algorithm's effectiveness depends on the nature of the erratic patterns and the presence of any significant events or seasonality. With the aim of enhancing predictive accuracy and computational efficiency in time series energy forecasting, El Assri et al., combined Kalman Filter and Continuous Wavelet Transform with Long Short-Term Memory neural networks to form KF-CWT-LSTM model (El Assri et al., 2025). The combination leveraged on the capacity of each algorithm to forecast energy consumption. The main objective of KF and CWT for time series data is to reduce noise and rectify uncertainty in the data to improve the signal and eliminate interference, facilitating the identification and separation of relevant patterns. The proposed KF-CWT-LSTM model was reported to outperform basic LSTM and other models as compared. A technique derived from the ARIMA (Autoregressive Integrated Moving Average) model known as the ARIMA Additive Outlier Method was employed by (Laome et al., 2021) to predict time series data that includes outliers. ARIMA AO has the capability to diminish the occurrence of outliers through iterative methods. In their study, the count of foreign tourists visiting the Port of Tanjung Priok was employed as data. The findings indicate that the model exhibits a residual with a normal distribution and a forecasting accuracy value. In (X. Li, 2024), multiple ARIMA models integrated with neural networks was developed to address the issues related to missing data. The methodology involves the synergistic combination of multiple ARIMA models, employing both model integration and inter-model compensation to achieve a significant enhancement in predictive accuracy. Another time series forecasting model capable of managing erratic data is the TBATS (Trigonometric Seasonality Box–Cox transformation, ARMA errors, Trend, and Seasonal components) model. It is integrated with Box-Cox transformation. This allows it to manage non-normality and stabilizes variance in time series data, leading to more precise forecasting. Its adaptability to manage different seasonal variations and non-linear trends renders it appropriate for identifying complex patterns in energy consumption. Without compromising prediction accuracy, TBATS is capable of managing missing values and

outliers typically found in energy consumption datasets (Serrano et al., 2024). Neural network-based Nonlinear Autoregression (NAR) is a predictive model used for forecasting nonlinear time series. It employs a recurrent autoencoder, such as a GRU autoencoder, to capture intricate non-linear relationships, temporal dependencies, and underlying trends, even in the presence of noise and non-stationarity. In the case of the TPUFIG model, this specific autoencoder is used to enhance robustness by reducing the impact of outliers and noise within the data segments (Y. Jiang et al., 2025).

In energy forecasting, models such as Random Forest, SVMs, and XGBoost demonstrate a marked superiority over linear regression by effectively capturing non-linear relationships and mitigating the impact of noise and outliers. For handling unevenly spaced time series data, however, more specialized techniques such as autoregressive models, moving averages, and RNNs are particularly well-suited for identifying underlying temporal patterns. Specifically, LSTM models excel at detecting the long-term dependencies characteristic of energy consumption's unpredictable nature. Furthermore, hybrid models that integrate traditional time-series analysis with advanced machine learning and deep learning algorithms have proven effective, yielding accurate predictions by capturing both linear and non-linear data relationships while maintaining robustness against noise and outliers (Benti et al., 2023; Deng et al., 2020). Table 5 explores some of the key data-cleaning techniques applied to handle erratic time series energy data.

Table 5. Data-cleaning and quality-control methods mitigating sensor errors.

Data cleaning techniques	**Application on erratic energy data**	**Related study**
Data Imputation Adding Constant Value Forward-fill **(Missing value)**	Missing values disrupt the continuity of time series energy data and has negative impact on forecasting accuracy. **Data Imputation:** Missing values in time-series energy data are addressed through data imputation, which replaces absent points with estimated values to maintain continuity for analysis and modeling. Common techniques include linear interpolation, multiple or regression imputation, and machine learning–based methods. For sporadic gaps, forward or backward filling preserves overall trends, while linear interpolation assumes proportional changes between observations. K-Nearest Neighbors (KNN) imputation estimates missing points using similar data patterns. Additionally, small constant values may be added to avoid division errors or undefined logarithmic operations, ensuring numerical stability and preserving data proportions	(Ahn et al., 2021; Chen et al., 2024; Lee et al., 2020)

Data cleaning techniques	Application on erratic energy data	Related study
Z-score or Standard Score **(Outliers)**	Outliers in time-series energy data are detected and managed using the Z-score method and its variants such as the Modified Z-score (mZ-score) and Median Absolute Deviation (MAD). These techniques identify data points that deviate significantly—typically beyond ±2 or ±3 standard deviations—from the mean, allowing them to be removed or adjusted to prevent distortion of forecasts. Alternative approaches like Winsorization (replacing extreme values with percentile limits), trimming**,** smoothing, data transformation, and robust regression also help reduce the undue influence of outliers on model accuracy.	(AlSalehy and Bailey, 2025; Yaro et al., 2023, 2024)
Smoothing Techniques **(Noise)**	Erratic time-series data often contain noise that obscures underlying patterns. Smoothing techniques reduce this noise and highlight key trends. Common methods include moving averages, which average nearby data points; exponential smoothing, which assigns greater weight to recent observations; and LOWESS smoothing, which fits a local curve to subsets of data for refined trend visualization.	(Hadjout et al., 2023; Panja et al., 2024; Romanuke, 2021; Singh, 1999)
Filtering Techniques **(Noise and outliers)**	Filtering techniques help detect and correct outliers while smoothing data. The Hampel filter identifies anomalies by comparing each data point with its local neighborhood, whereas the Kalman filter recursively estimates system states from observations, effectively separating true signals from noise in time-series data.	(El Assri et al., 2025; Zhu et al., 2023)
Bayesian methods	Erratic data often involves inherent uncertainty that weakens single-point predictions. Bayesian methods address this by incorporating uncertainty into forecasting, assigning probability distributions to outcomes and generating a range of possible future scenarios with their associated likelihoods.	(Mahajan et al., 2024; Shrestha et al., 2021)
Robust Statistics	Traditional statistical measures like the mean and standard deviation are highly sensitive to outliers, which can distort results. Robust statistics, such as the median and Median Absolute Deviation (MAD), provide more reliable alternatives by minimizing the influence of extreme values and offering a consistent measure of data variability.	(Hadjout et al., 2023)

4. Time series energy forecasting

Time series data analysis usually involves two key steps: understanding the underlying patterns and structure of the data, and then using this insight to build models for forecasting future values. The decomposition of time series is important in time series analysis. It establishes a theoretical framework which helps in the process of forecasting. This step will ascertain if the analyzed time series includes consistent or recurring components that can be characterized and modeled directly. These components may then suggest the necessity for data preparation, model selection, or feature engineering activities. Time series analysis conventionally proceeds by decomposing the series to identify the trend(this represents the overall long-term movement or direction of the electricity consumption variable, disregarding seasonal or irregular fluctuations), seasonality (these are predictable, periodic fluctuations in energy consumption that remain consistent in terms of timing, intensity, and direction) and residual (This is the unexplainable portion of the time series, which can sometimes be substantial enough to obscure the underlying trend and seasonality). The goal of time series forecasting is to leverage historical data to make accurate predictions about future trends (Bezzar et al., 2022; Deb et al., 2017, 2017; Mojtahedi et al., 2025). Improving forecasting accuracy has significant benefits for operational efficiency across various sectors. Recently, time series methods have been effectively utilized in the power sector across various applications that involve a temporal aspect, including energy consumption forecasting, non-intrusive activity recognition, and demand-side management and control (Deb et al., 2017; Qolomany et al., 2017). However, time series data of an energy profile are influenced by many determinants such as building envelope, occupancy and occupant activities, income, demography, season and environment (as in Figure 5(Qiao et al., 2021)). The relationship between the energy consumption and the determinants are often recorded to be non-linear. To this effect, energy consumption patterns often exhibit temporal dependencies (El Assri et al., 2025). The inherent time-series nature of building energy data and its temporal correlation with other variables necessitate a more extensive comparison and analysis of different time-series forecasting architectures (Tekler and Chong, 2022).

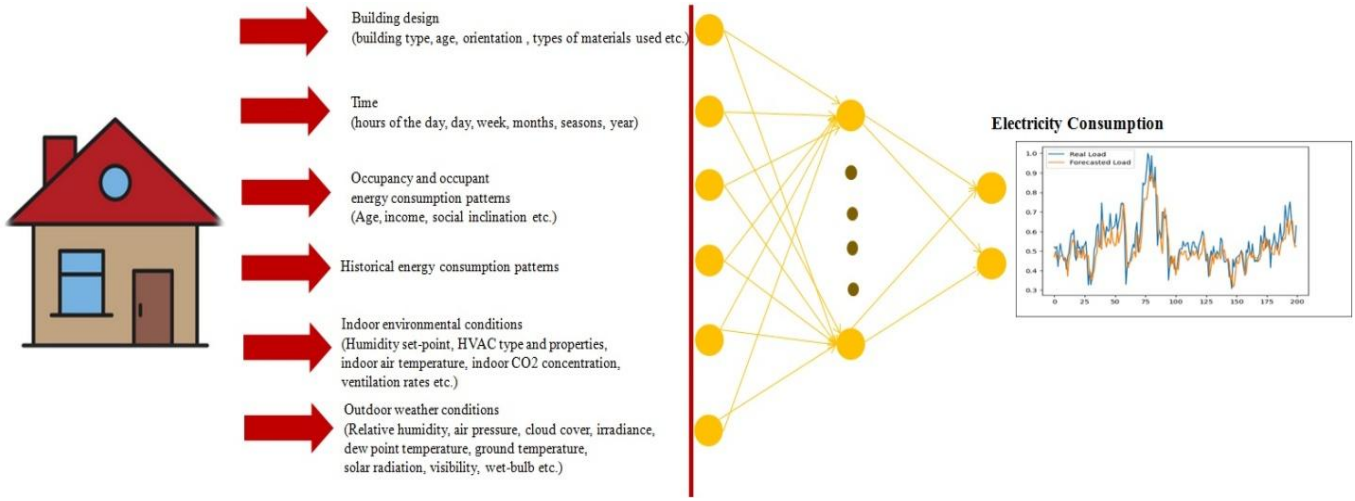


Figure 5. Commonly used prediction parameters for building energy forecast.

4.1. Models for time series building energy forecasting

Accurate forecasting of power consumption is critical for enhancing electricity optimization, securing grid stability, informing economic planning, and advancing environmental sustainability. The intricate nature of energy consumption and its generation data, exhibiting by various underlying trends and patterns, necessitates the use of various modeling architectures. Different models, including traditional statistical methods, computational intelligence approaches, and hybrid models, have been developed to address the complexities of predicting electricity usage (as illustrated in Figure 6). These algorithms seek to capture the complex interplay of determinants such as calendar effects, occupancy, demography, weather conditions, occupant behavior and socio-economic changes. The choice of a suitable forecasting model depends on the precise characteristics of the load data, the forecasting horizon, and the available computational resources. Recent trends show a strong inclination towards advanced computational intelligence techniques, ensembles and hybrid models due to their ability to handle dynamic non-linear patterns and improve prediction accuracy. Addressing issues like data missingness, outliers, and noise is also critical for maintaining forecasting accuracy (A.S. Ahmad et al., 2014; M.W. Ahmad et al., 2016; Alanne and Sierla, 2022; Anan et al., 2024; Deb et al., 2017, 2017; Mystakidis et al., 2024; Z. Wang et al., 2025).

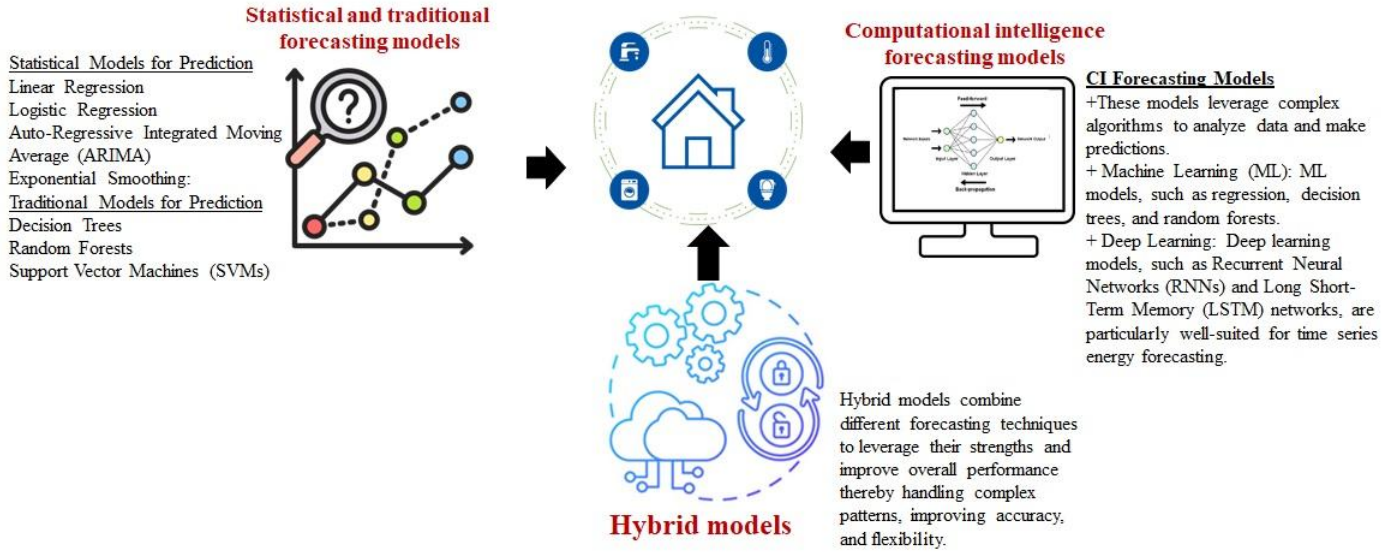


Figure 6. Illustration of different predicting techniques.

Owing to non-linear characteristics of electricity consumption data, conventional models for forecasting energy usage in buildings, such as first-principles models and statistical analysis approaches (ARIMA, SARIMA(Seasonal Autoregressive Integrated Moving Average), and ETS (Error, Trend, Seasonality)) frequently struggle (Chen et al., 2025a; El Assri et al., 2025; Sha et al., 2022). First-principles models oversimplify building features, real-world intricacies and random fluctuation in energy usage. On the other hand, statistical models often rely on linear relationships, rendering them insufficient for depicting the complex relations among factors such as temperature, income, and occupancy etc. (El Assri et al., 2025).

Consequently, data-driven techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting Regression Trees (GBRT), Multiple Linear Regression (MLR), K-Nearest Neighbors (KNN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and XGBoost have been utilized in the examination of power consumption in existing buildings. Based on previous usage trends, these methods leverage historical data to develop predictive models of energy consumption (Deb et al., 2017; Hussien et al., 2025; Le et al., 2019).

The stochastic nature of variants of building energy data exhibits complex temporal patterns, thereby complicating prediction for ML techniques (El Assri et al., 2025; Ngo et al., 2022; J. Wang et al., 2021). To capture long-term temporal patterns (trends, seasonality or cycles that occur over a long period) and dependencies accurately, deep learning (DL) modeling techniques are particularly well-suited (Tekler and Chong, 2022). Using multi-layered neural networks, deep learning effectively models complex patterns and non-linear relationships in energy data. Its techniques demonstrate exceptional performance in time series analysis, anomaly detection, and energy optimization problems for building energy management (Truong et al., 2021). Deep learning models are the Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), Convolutional Neural Networks (CNNs) and Temporal Convolutional Networks (TCNs). However, to deduce a long-term forecasting for time series data, RNN was upgraded to Long-short term memory neural network (LSTM). These LSTMs are tailored for sequential data, such as time series, and are adept at learning long-term dependencies (Anan et al., 2024; Fan et al., 2019; Hitimana et al., 2021; Jogunola et al., 2022; Mojtahedi et al., 2025; Tekler and Chong, 2022).The LSTM neural network has the ability to retain past information, enabling the model to make more accurate predictions about current and future data, outperforming networks that rely solely on present data (Leeraksakiat and Pora, 2020; Qolomany et al., 2017). However, deep learning models have their limitation, which includes the need for vast historical data, which may not be available, and being computationally intensive and time-consuming to train. They are often "black boxes," hindering interpretability important for decision-making. Achieving optimal performance with deep learning often requires complex hyperparameter tuning. Therefore, to address these shortfalls, models are either ensemble or hybridized.

In contrast with single models, ensemble models offer improved precision and robustness by merging predictions to reduce variance and smooth errors. They help in reducing overfitting and achieving better generalization to new data or pattern shifts. Ensembles also aid in handling model instability, often seen in deep learning, by combining multiple model instances or architectures. Furthermore, they leverage diverse model strengths by integrating different model types. For tasks like building energy forecasting, ensembles are often preferred for their reliability and stability in real-world applications. The combination could form heterogeneous ensemble models (these models incorporate multiple algorithms, such as combining SVR, RF, ANN, MLR, and LSTM) or homogeneous ensemble models (these models utilize multiple data subsets, often employing techniques like bagging for example Random Forest or boosting such as Multi-LSTM DANN (Buyo et al., 2025; Dannecker, 2015; Z. Wang et al., 2025). Meanwhile, hybrid models are often preferred over pure deep learning models, which heavily depend on large amounts of high-quality data. Hybrids are more robust to data limitations (like sparse or noisy data) by incorporating domain knowledge or combining different techniques. They offer greater transparency and interpretability, aligning with physical laws, which contrasts with the "black box" nature sometimes associated with deep learning. Hybrid models effectively combine strengths for handling both linear (e.g., statistical) and complex non-linear (e.g., deep learning) patterns (Farea et al., 2024; Guo et al., 2024; Z. Jiang et al., 2025; Ngoc-Tri et al., 2022; Ruan et al., 2022). This allows combining

different model types, ARIMA-ANN (Artificial Neural Network) Hybrid Models (L. Wang et al., 2013), Deep Learning (e.g., LSTM, CNN) with Physics-Informed Neural Networks (PINNs) (Z. Jiang et al., 2025), Grey-Box Models Integrating Machine Learning/Neural Networks (Guo et al., 2024), Ensemble of Hybrid Models / Homogeneous Assembly Models (Buyo et al., 2025) etc., enabling each component to handle what it is best at.

Table 6 categorized the models in types and performance adapted from Deb et al., 2017; Tekler and Chong, 2022; Z. Wang et al., 2025, and typical ranges of indicative accuracy from (A. S. Ahmad et al., 2014; Bourdeau et al., 2019; Deb et al., 2017; Kim, Kang, et al., 2019; Runge and Zmeureanu, 2019).

Table 6. Categories of models, applications and performance.

Model Category	Typical Algorithms	Typical Forecasting Context	Advantages	Indicative Accuracy (Typical Ranges)	Limitations
Statistical	ARIMA, SARIMA, Prophet	Short-term; seasonal, linear loads	Transparent, interpretable	MAE ≈ 2.01 % to 3.74; RMSE Values are highly unit-dependent; MAPE ≈ 0.33 to 9.0%	Struggles with nonlinear data
Machine Learning	SVR, RF, XGBoost	Medium-term; moderately nonlinear data	Captures complex relations	MAE ≈ 1.94 % to 9.68; RMSE ≈ 5.56 to 44.17%; MAPE ≈ 0.001 to 42.31%	Sensitive to feature scaling
Deep Learning	LSTM, GRU, CNN	Long-term; high-frequency or nonlinear data	Learns temporal dependencies	MAE ≈ 0.31 % to 9.68; RMSE ≈ 0.35 to 12.56%; MAPE ≈ 0.78 to 21.43 %	Data- and compute-intensive
Hybrid / Ensemble	ARIMA–ANN, VMD–LSTM, GA–LSTM	Multi-horizon; heterogeneous datasets	Combines linear + nonlinear strengths	MAE ≈ 0.0245 to 0.0364 %; RMSE ≈ 5.62 %; MAPE ≈ 0.132 to 14.4%	Complex tuning, low interpretability

4.1.1. Explanation of Accuracy Metrics and Model Performance

The sources indicate that hybrid models and highly optimized deep learning models often yield the lowest error rates, demonstrating MAPE values below 1% in optimal conditions. The general consensus is that hybrid techniques improve accuracy by compensating for the weaknesses of individual models, such as ARIMA handling linear components and ANN managing nonlinear residuals. Statistical models generally offer transparency and ease of implementation while competitive in certain simple time series (e.g., ARIMA/SARIMA), they are limited by assuming a linear description of the phenomenon. Machine learning

models, particularly Support Vector Machines (SVM), are recognized for dealing effectively with nonlinear problems. SVMs, specifically Support Vector Regression (SVR) and Least Squares SVM (LSSVM), frequently report high accuracy, sometimes outperforming ANN models, especially when using limited training data. ANNs themselves demonstrate wide performance variability, ranging from extremely accurate (0.001% MAPE for single-step forecasts) to significantly higher errors (42.31% MAPE for multi-step forecasts), illustrating that performance is sensitive to forecast horizon and data granularity. Meanwhile, Deep Learning (DL) models like LSTM and complex hybrid CNN architectures are designed to handle large datasets and complex non-linear relationships, particularly long-term dependencies in time-series data. The most optimized DL approaches show competitive accuracy (as low as 0.78% MAPE). The hybrid / ensemble models strategically combine multiple techniques to harness complementary strengths. The combination of LS-SVM with kNN achieved some of the lowest reported MAE percentage errors (0.0245%), reinforcing the principle that complementary models enhance accuracy and robustness (A. S. Ahmad et al., 2014; Bourdeau et al., 2019; Deb et al., 2017; Kim, Kang, et al., 2019; Runge and Zmeureanu, 2019).

At this point, where the categories of models are presented (Table 6), time series energy data characteristics can be linked to forecasting models (refer to section 3.1)**.** The inherent properties of time-series energy data directly influence model selection and forecasting performance. For instance, non-linearity in consumption patterns (as in section 4.1) driven by occupant behavior, weather variability, and appliance interactions renders linear models (e.g., ARIMA, ETS) insufficient, thereby motivating the adoption of deep learning and hybrid architectures that can learn complex, non-linear dependencies. Seasonality, often evident in daily and annual cycles, aligns well with statistical models such as SARIMA**,** TBATS, and Prophet (as discussed in sections 3.1 and 4.1), which explicitly capture periodic trends through decomposition or Fourier-based representations. Finally, temporal dependence the influence of past observations on future values justifies the use of recurrent neural networks (RNNs) and their variants (e.g., LSTM, GRU), which are designed to model sequential memory and long-term dependencies (as presented in 3.2 and 4.1). Thus, an understanding of these intrinsic data characteristics provides a theoretical basis for matching forecasting techniques to the specific dynamics of building energy behavior.

4.1.2. Time series forecasting horizons

The selection of various time frameworks influences the prediction techniques and the development of the model (Chen et al., 2025a). The forecasting of building electricity consumption typically employs a time series data horizon that is classified into three categories: short-term, medium-term, and long-term load forecasting. Forecasting horizons refer to the time period for which prediction are made. Each of these horizons has different time duration. The purpose of selecting any of these horizons is to achieve precision in a particular area. The areas could be residential, commercial, industrial, larger community or a power generation plant (A. S. Ahmad et al., 2014; Mariano-Hernández et al., 2020; Reis et al., 2025). A short-term forecasting limited to 24 hours is referred to as very short-term load forecasting; the time scale can be in the range of seconds, minutes and hourly intervals. However, if the time span exceeds 24 hours and runs into days or weeks, then the horizon is short-term forecasting. For monthly duration or quarterly forecasting, the time set is referred to as medium-term forecasting. Categories of prediction that cover several years are long-term forecasting. (A.S. Ahmad et al., 2014; Chen et al., 2025a; Wei et al., 2019b). According to the literature, recently, the long-term energy forecasting has gained attention to a range of 43-58.3%, 23-23.3% on short-term, 13.4-23% attention on a very

short-term forecasting, while medium term recorded the least, between 5-17%. The growth in application of very short-term energy forecasting is as a result of designing renewable energy solution for buildings (Mariano-Hernández et al., 2020; Wei et al., 2019b). Meanwhile, the category of horizon, data volume and its features determine the model type for prediction. For example, In contrast to short-term energy forecasting, the long- and medium-term prediction of energy consumption typically requires a less extensive volume of historical data and a less granular analysis of the intricate relationships between energy usage and various influential factors(Wei et al., 2019b).

Various models and techniques are employed for short-term, medium-term and long-term load forecasting in building energy systems. These methods range from traditional statistical approaches to advanced machine learning and computational intelligence techniques, including hybrid and ensemble approaches (as in Table 6). The models were developed to capture the complex interplay of factors such as calendar effects, weather conditions, and socio-economic changes.

- *Statistical and traditional models:* Short-term load forecasting (STLF) often utilizes regression techniques (Kiartzis et al., 1995). For example, a multiple linear regression (MLR) is a viable method for estimating the thermal energy demand for heating or cooling within buildings. This statistical technique is also used for medium-term load forecasting (MTLF). They are also fundamental in long-term energy forecasting, often incorporating the effect of variables like temperature on energy demand. For instance, a probabilistic method that includes temperature effects can be used for long-term energy demand forecasting (C.W. Fu and Nguyen, 2003; Mariano-Hernández et al., 2020). Exponential Smoothing model has been used for STLF and it performs well for short-term forecasting but encounters difficulties when forecasting long-term trends (Son and Shin, 2023). Quantile Regression shows satisfactory performance when used for MTLF (Shaikh et al., 2022). ARIMA excels at short-term forecasts but can struggle with seasonality and non-stationarity over longer periods (Son and Shin, 2023). Seasonal Autoregressive Integrated Moving Average (SARIMA) models extend ARIMA by incorporating additional parameters to capture seasonal variations in load data, such as daily, weekly, or monthly cycles. This makes them suitable for power systems with strong seasonal patterns (Akhtar et al., 2023; Mariano-Hernández et al., 2020). Generalized Additive Models (GAMs) are a prominent statistical tool for MTLF due to their ability to capture non-linear relationships while maintaining interpretability (Abu-Shikhah et al., 2011; Zimmermann and Ziel, 2025).
- *Computational intelligence models:* Artificial Neural Networks (ANNs) are widely used in STLF because they do not rely on human experience but learn the functional relationship between inputs and outputs through training; they are adaptable for various temporal forecasting ranges, including long-term (Kiartzis et al., 1995). SVR is notable for its robust generalization capabilities and its ability to address non-linear regression by determining the global optimal solution. Illustrating this, Chen et al. developed an SVR prediction model for office building power load forecasting that integrated ambient temperature data, which yielded enhanced accuracy and stability compared to alternative methods(Qiu et al., 2025). A weighted multiple SVR model was also developed for STFL energy utilization for non-residential buildings (Mariano-Hernández et al., 2020). Functional-Link Net (FLN) and Wavelet Networks were proposed and evaluated for long-term forecasting of energy demand, with wavelet functions showing good performance in forecasting accuracy. Spatiotemporal Graph Convolutional Network (ST-GCN) and KNN (Phyo et al., 2022) were applied for STLF. While primarily

for hourly predictions, the concept of graph-based methods for inter-building dependencies (that is, ST-GCN) could be extended to long-term urban energy modeling. Prophet Model is an open-source time-series forecasting library developed by Facebook, suitable for medium-term forecasting (Son and Shin, 2023). While useful, conventional RNNs suffer from vanishing gradient phenomena, limiting their ability to model long-range temporal dependencies (Qiu et al., 2025). LSTMs address the vanishing gradient problem and are effective in extracting dynamic temporal information by enhancing long-term memory capacity through forget, input, and output gates (F. Liu and Liang, 2024; Qiu et al., 2025; Son and Shin, 2023). Torres et al. used LSTM to forecast STLF due to its capability of dealing with sequences of time series data. Some studies suggest that prominent deep learning-based MTLF techniques overlap with STLF methods (Shaikh et al., 2022). LSTM and its variants like BiLSTM (Bidirectional Long Short-Term Memory) are capable of learning bidirectional temporal features and maintaining long-term memory, which could be beneficial for MTLF (Gou et al., 2024; F. Liu and Liang, 2024). GRU is a type of RNN that provides similar functionality to LSTM but with a simpler structure, effectively handling long-term dependencies and reducing computational costs compared to LSTM (Son and Shin, 2023).

- *Hybrid and ensemble models:* GA-SVR and GA-WD-SVR these models combine Genetic Algorithm (GA) with Support Vector Regression (SVR) and Wavelet Decomposition (WD), are established for short-term and ultra-short-term cooling load predictions for office buildings. Prophet and GRU Combination: this combines Prophet Model for seasonality and event handling with the GRU model for long-term and medium-term forecasting. It processes preprocessed data with Prophet first, then uses GRU with multivariate data (meteorological data and Prophet residuals) to predict electricity consumption at short- and medium-term intervals. ARIMA with XGBoost were combined for short-term forecasting (Son and Shin, 2023). GAM with Autoregressive Post-processing, GAMs are enhanced with AR post-processing to account for short-term temporal dependencies (Zimmermann and Ziel, 2025). Deep Belief Network (DBN) and Empirical Mode Decomposition (EMD): One proposed hybrid model combines DBN with EMD to address medium to long-term forecasts, aiming for improved prediction performance (Abdulrahman et al., 2021). Ensemble Voting Regressor: This approach combines the predictions from multiple machine learning regressors (e.g., Extra Trees Regressor, Random Forest Regressor, Light Gradient Boosting Machine, Gradient Boosting Regressor, and K-Nearest Neighbors) to achieve better forecasting accuracy (Phyo et al., 2022). These diverse applications of models highlight the evolving landscape of STLF, MTLF and long-term load forecasting, with a growing attention on hybrid and computational intelligence approaches to address the intricacies and non-linearities in electricity energy consumption.

In summary, the reviewed literature demonstrates a clear performance hierarchy across model families tied to data complexity and forecasting horizons. Statistical models, such as ARIMA and SARIMA, which are computationally simple and model linear temporal patterns, are predominantly effective for short-term forecasting (up to a week), particularly when data are stationary and linear. However, their inability to capture non-linear relationships often leads to higher errors compared to intelligent methods; for example, ARIMA yielded a high overall MAPE of over 12% (Phyo et al., 2022) in one short-term energy forecasting comparison. Traditional Machine Learning (ML) techniques, like Artificial Neural Networks (ANN) and Support Vector Machines (SVM), excel in handling complex non-linear data, making them highly effective for accurate short-

term load forecasting (STLF). In one STLF study comparing ARIMA and ANN, the ANN model showed a lower error rate, having a better regression factor (R) closer to 1(Tarmanini et al., 2023) than ARIMA. Deep Learning (DL) architectures, such as Long Short-Term Memory (LSTM) and Deep Recurrent Neural Networks (DRNN), are specialized in modeling prolonged dependencies, making them suitable for reliable medium-term (weeks to a year) and long-term (a year and above) predictions. An experiment demonstrated that Bi-LSTM, a DL variant, achieved superior predictive performance over statistical models (ARIMA and SARIMA) by capturing long-term dependencies, evidenced by a notably lower Root Mean Square Error (RMSE) of 0.124 (Alizadegan et al., 2025). Finally, Hybrid and Ensemble models consistently deliver the highest precision, primarily excelling in short-term and ultra-short-term forecasting; for instance, the ensemble Voting Regressor (VR) achieved the best overall short-term performance with an average Mean Absolute Percentage Error (MAPE) of 4.28% over (Phyo et al., 2022) a year-long test set, significantly outperforming the standalone ARIMA model.

4.1.3. Key considerations for model selection and application

When selecting and applying a predictive model, several crucial factors must be considered to ensure optimal performance and utility (Bezzar et al., 2022; Bourdeau et al., 2019; Chen et al., 2025a; Deb et al., 2017).

- *Forecast Horizon:* This refers to the duration of time into the future for which a time series energy forecast (short-term, medium-term, long-term) is made to significantly impacts the model accuracy. Short-term forecast (minutes to days) generally results in accurate predictions due to less uncertainty, whereas longer horizons (spanning months to years) introduce more variability and make precise prediction more difficult. Model performance varies depending on the forecasting horizon; for instance, deep learning is often better suited for very short-term predictions, while statistical models can be more effective for medium-term forecasts with distinct seasonal patterns.
- *Data Granularity and Volume:* The data granularity (the lowest level of detail contained in a building energy data collection) and volume also influence the selection of the model. High-frequency time series building energy data can gain from models capable of identifying intricate patterns, such as LSTMs and Transformers. Furthermore, larger electricity datasets often favor machine learning and deep learning approaches due to their ability to learn from huge amounts of information.
- *Exogenous Variables:* The availability and quality of exogenous variables (that is, external data not directly influenced by the model, such as weather, income, or social event) can considerably improve predicting accuracy. These variables offer additional background and information, which is helpful in domains like energy-related predictions (e.g., using weather data for electricity load forecasts). It is crucial to accurately categorize and apply these variables, whether static, prospective, or historic during both model training and prediction phases.
- *Interpretability vs. Accuracy:* An essential trade-off exists among model interpretability and accuracy. Simpler models, such as linear regression or traditional decision trees, are more interpretable, allowing for a clearer understanding of how they arrive at predictions. However, they might give up accuracy when dealing with complex and high-dimensional data. On the other hand, complex models like deep neural networks and ensemble methods often provide higher accuracy but are considered "black

boxes" due to their complex architectures, making it complicated to understand their internal workings.

- *Computational Resources:* Computational resources are a practical consideration, particularly for deep learning models. For these models, particularly those with deep architectures (e.g., CNNs or Transformers), training effectively necessitates considerable computational resources and specialized hardware, such as Graphical Processing Units (GPUs). While they can achieve high precision, their computational requirements can make them not viable for real-time applications if resources are limited.
- *Robustness to Anomalies:* Although anomaly detection is regarded as a distinct process, robustness to anomalies is a good attribute for predictive models. Robust models reduce vulnerability to outliers or strange data points contained in the training data, which can otherwise result in biased models and erroneous predictions. Strategies to mitigate (as in section 3.2.2) the impact of data anomalies include data cleaning, preprocessing techniques (like outlier removal or imputation), handling imbalanced datasets, and continuous monitoring and updating of models.
- *Probabilistic forecasting:* Finally, for risk management within energy markets, probabilistic forecasting, which provides prediction intervals rather than just point forecasts, is frequently more valuable. Numerous models can be adapted to deliver this type of forecast through techniques like Monte Carlo simulation or by integrating particular objective functions.

4.2. How time series building energy forecasting is performed

Time series energy forecasting models utilize large amount of training data to learn and recognize patterns that might not be clear to conventional statistical forecasting models. The methodology for energy forecasting encompasses a series of sequential stages including data acquisition, pre-processing, feature engineering, model selection and parameterization, and performance evaluation all of which contribute to enhancing predictive accuracy(Bezzar et al., 2022; Chen et al., 2025a; Deb et al., 2017; Fallah et al., 2018; Mystakidis et al., 2024).

- *Data Collection and Preparation:* Accurate energy forecasting relies on comprehensive data collection and meticulous preparation. This involves data aggregation, such as past energy consumption records, weather data etc. Following this, the gathered data is subjected to cleaning, imputation, normalization (commonly referred to as scaling, which seeks to convert data into a uniform range), and feature selection to guarantee its accuracy and appropriateness for model training.
- *Data Splitting:* This process entails partitioning the available data into various subsets. A widely adopted approach is to allocate a segment of the data for training the forecasting model, another segment for assessing the model's performance throughout the training phase, and the final segment for evaluating the completed, trained model.
- *Time Correlation (Autocorrelation in time series analysis):* Time correlation analysis reveals how past values in a time series influence future ones, enabling models to identify patterns and dependencies that drive accurate predictions, particularly in energy forecasting, where consumption patterns often follow temporal trends.

4.2.1. Incorporating Time Correlations into Forecasting Models

Techniques like seasonal trend decomposition using loess (STL) can decompose time series data into trend, seasonality, and residual components, allowing for better analysis of temporal patterns; this is referred to as Time-Series Decomposition. Meanwhile, feature engineering can enhance model performance by extracting time-based features like hour, day, or month, but understanding your time series data's underlying relationships is key to selecting and training an effective model.

- *Model selection, testing and training:* With a thorough understanding of temporal relationships in your data, you can choose an appropriate forecasting model and train it on historical data to make accurate predictions, leveraging patterns and trends to forecast future events. The time series forecasting algorithms undergo continuous refinement and optimization to improve performance. The algorithms are evaluated on a separate testing dataset that was not used during the training. The goal is to validate the model's performance in practical applications, thereby ensuring its reliability. By identifying and addressing prediction errors, the model is refined to enhance accuracy, underscoring the importance of iterative training and testing. This process ultimately boosts the model's predictive power and minimizes errors, with final performance evaluation determining the algorithm's effectiveness
- *Evaluation metrics of time series energy forecasting model:* Evaluation metrics for time series energy forecasting models serve the vital purpose of determining model reliability, which is important for efficient energy management, planning, and resource allocation. By validating predicted energy values against actual energy consumption data, these metrics such as MAE, RMSE and MAPE, reveal the model's effectiveness, identify potential biases (e.g., consistent over- or under-forecasting), and help in selecting the most suitable model for specific business objectives like cost minimization or grid stability. The selection of evaluation metrics usually depends on the application, as some penalize larger errors more heavily (RMSE), while others gives a percentage-based error for simple comparison across various scales (MAPE). These metrics help stakeholders to refine models, make informed decisions, and optimize energy operations. From these articles (Bezzar et al., 2022; Chen et al., 2025a; Fallah et al., 2018; Mystakidis et al., 2024), an overview of the commonly used evaluation metrics for performance assessment of regression task (time series energy data forecast) were studied. It important to note that none of these metrics is universally "most suitable" for time series energy forecasting. The best choice or, more likely, combination of metrics depends heavily on the specific business goals, data characteristics, and tolerance for different types of errors (Piotrowski et al., 2022; Serrano et al., 2024; Svoboda et al., 2024).

4.2.1.1. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) measures the average absolute percentage difference between predicted and actual energy values. A MAPE of 0% indicates perfect accuracy, and a value below 10% is considered excellent. Its primary advantage is that it's easy for non-technical users to understand because the error is expressed as a percentage. It's also independent of data scale, allowing for comparisons across different datasets or time frames. However, a significant limitation is its unreliability when actual values are zero or near-zero, which can lead to infinite or undefined results. MAPE is also sensitive to outliers and is asymmetric, meaning identical forecast errors can result in varying MAPE values.

4.2.1.2. Mean Squared Error and Root Mean Square Error

Mean Squared Error (MSE) calculates the average of the squared differences between predicted and actual values. A lower MSE signifies better accuracy, with zero being the ideal value. MSE is useful when large errors are particularly important because the squaring of errors gives them more weight. It also has strong statistical ties to variance and bias. However, it doesn't account for systematic bias and can be inflated by irrelevant variables. Root Mean Square Error (RMSE) is the square root of the MSE and evaluates the average magnitude of forecasting errors in the same units as the data. An RMSE value between 0.2 and 0.5 suggests a relatively accurate prediction, while an ideal value is 0. Like MSE, RMSE penalizes bigger errors more heavily, which is beneficial when high accuracy is critical. A major limitation is its sensitivity to outliers, as a few extreme errors can greatly increase the RMSE value and distort the overall performance assessment. Its unit dependence also complicates comparing models with different units or scales

4.2.1.3. Mean Absolute Error

Mean Absolute Error (MAE) evaluates the average magnitude of errors between predicted and actual values. A lower MAE indicates better accuracy, with a value of zero representing no error. MAE provides a simple and intuitive measure of prediction error in the same units as the target variable. It treats all errors equally, providing a clear understanding of the average error magnitude. However, because it treats all errors equally, it doesn't distinguish between overestimation and underestimation, which may be crucial in certain applications. It is also not a normalized metric and can be heavily influenced by outliers.

4.2.1.4. Coefficient of Determination

The Coefficient of Determination (R2) is a statistical measure that shows how closely a model's predictions align with the actual data. It explains the proportion of total variation in the data that the model captures. A value of 1 indicates a perfect model fit, while 0 shows no relationship between variables. R2 is helpful for assessing a model's predictive power and can aid in identifying which independent variables are most significant. It does not, however, account for systematic bias and can be inflated by adding irrelevant variables. It is also responsive to outliers and does not directly measure predictive error.

4.3. Integrative insights and research gaps

Unlike previous reviews that primarily catalog forecasting models, this study provides a unifying synthesis of how data characteristics, sensor quality, and modeling approaches interact to shape forecasting accuracy in buildings. Specifically, it (i) elucidates the influence of erratic and stochastic time-series energy data on prediction reliability, (ii) highlights data-quality mitigation frameworks that integrate EDA, imputation, and hybrid filtering, and (iii) advances an integrative perspective that connects statistical, machine learning, and hybrid paradigms within a coherent forecasting pipeline. Future research should focus on developing standardized benchmarks for evaluating hybrid models on heterogeneous and imperfect building datasets.

5. Conclusion and future work

In the face of rising urbanization and energy needs, building energy consumption forecasting is a crucial factor for advancing energy efficiency, maintaining grid stability, guiding economic decisions, and promoting

environmental sustainability. This article has presented an understanding of time series building energy consumption data and reviewed various approaches to forecasting it. The review highlights the complex characteristics of building energy time series data, including seasonality, trends, temporal dependencies, non-linearity, volatility, and stochasticity. These characteristics pose substantial hurdles for modeling and accurate forecasting. The article pointed out the increasing dependence on IoT and sensor technologies for collecting the necessary time series data. However, a notable challenge pinpointed is the poor quality of sensor data, complicated by challenges like incompleteness, discrepancy, anomalies, noise, bias, and missing values. These "erratic time series data patterns" significantly obstruct the development of accurate forecasting models, especially traditional techniques, resulting in increased prediction errors and inadequate energy resource allocation. Various data cleaning and preprocessing techniques, including imputation, smoothing, and filtering, are important mitigation strategies to address these data quality challenges before model application.

The review examined different forecasting models, ranging from traditional statistical methods like ARIMA and ETS to advanced computational intelligence approaches like ANN, SVM, RF, and Deep Learning models such as RNNs and LSTMs. Traditional models are often insufficient for capturing the complex non-linear relationships and stochastic nature of energy data. Data-driven techniques, especially deep learning models like LSTMs, have shown promise due to their ability to model complex patterns and temporal dependencies, although they often require vast data and computational resources and can lack interpretability.

Hybrid and ensemble models are presented as advantageous approaches, offering improved accuracy, robustness, and interpretability compared to single models by combining the strengths of different techniques. The selection of an appropriate forecasting model is influenced by several key factors, including the forecast horizon (short, medium, long), data granularity and volume, the availability and quality of exogenous variables (like weather or occupant behavior), the trade-off between interpretability and accuracy, computational resources, and the model's robustness to anomalies. The process of time series energy forecasting involves meticulous data collection, preprocessing, feature engineering, model selection, training, and rigorous evaluation using metrics like MAE, RMSE, and MAPE. Despite the advancements reviewed, several challenges remain that point towards necessary future work:

- Deeper understanding and characterization of energy data: The article points out that a comprehensive study on the characteristics, nature, and patterns of building electricity consumption data is rarely conducted. This strongly suggests future work should focus on developing a deeper insight into the complex characteristics of time series energy data to enhance model precision.
- Handling data quality issues: Managing data quality is a major challenge in sensor technology, giving rise to problems like incompleteness, discrepancy, anomalies, noise, bias, and missing values. These "erratic time series data patterns" complicate precise predictions, particularly with the conventional approach. Further research is required to focus on robust and potentially automated data cleaning, preprocessing, and imputation techniques.
- Improving the integration of complex determinants: Factors like occupant behavior, income, weather conditions, demography, and building envelope substantially impact energy usage, characterized by non-linear dynamics and temporal correlations that add uncertainty to prediction. Occupant interaction is particularly noted as complex and contributing to forecasting errors. Future studies could explore developing more effective methods for collecting and incorporating complex determinants,

such as occupant behavior, into forecasting models to bridge the gap between predicted and actual energy consumption.

- Developing more robust and advanced forecasting models: The study highlighted that the stochastic nature and the time-dependent patterns in building energy data can make it hard for some AI models to generate precise forecasts. Future work could involve developing cutting-edge prediction models or improving existing forecasting methods to better account for intricate relationships. While deep learning models like LSTMs are notable for capturing long-term dependencies and complex non-linear relationships, this study identified their limitations, such as requiring vast data, being computationally intensive, and often acting as "black boxes". This suggests future work should aim to develop more data-efficient, computationally lighter, or more interpretable deep learning architectures. It could be on exploring new combinations of models (hybrid) or techniques (ensemble) to further enhance forecasting performance, potentially focusing on integrating domain knowledge or addressing specific data challenges.

In summary, building on the insights and gaps highlighted in the study, future research directions would likely focus on improving the understanding and handling of complex and erratic energy data characteristics, developing more robust, accurate, and potentially more interpretable forecasting models (with a focus on leveraging hybrid and ensemble approaches), and better integrating the intricate factors that shape building energy consumption.

References

Abdulrahman, M.L., Ibrahim, K.M., Gital, A.Y., Zambuk, F.U., Ja'afaru, B., Yakubu, Z.I. and Ibrahim, A. (2021), "A Review on Deep Learning with Focus on Deep Recurrent Neural Network for Electricity Forecasting in Residential Building", *Procedia Computer Science*, Vol. 193, pp. 141–154. https://doi.org/10.1016/j.procs.2021.10.014

Abu Bakar, N.N., Hassan, M.Y., Abdullah, H., Rahman, H.A., Abdullah, M.P., Hussin, F. and Bandi, M. (2015),"Energy Efficiency Index as an Indicator for Measuring Building Energy Performance: A Review", *Renewable and Sustainable Energy Reviews*, Vol. 44, pp. 1–11. https://doi.org/10.1016/j.rser.2014.12.018

Abu-Shikhah, N., Elkarmi, F. and Aloquili, O.M. (2011), "Medium-Term Electric Load Forecasting Using Multivariable Linear and Non-Linear Regression", *Smart Grid and Renewable Energy*, Vol. 2 No. 2, Article 2. https://doi.org/10.4236/sgre.2011.22015

Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H. and Saidur, R. (2014), "A Review on Applications of ANN and SVM for Building Electrical Energy Consumption Forecasting", *Renewable and Sustainable Energy Reviews*, Vol. 33, pp. 102–109. https://doi.org/10.1016/j.rser.2014.01.069

Ahmad, M.W., Mourshed, M., Mundow, D., Sisinni, M. and Rezgui, Y. (2016), "Building Energy Metering and Environmental Monitoring – A State-of-the-Art Review and Directions for Future Research", *Energy and Buildings*, Vol. 120, pp. 85–102. https://doi.org/10.1016/j.enbuild.2016.03.059

Ahmad, T., Chen, H., Guo, Y. and Wang, J. (2018), "A Comprehensive Overview on the Data Driven and Large Scale Based Approaches for Forecasting of Building Energy Demand: A Review", *Energy and Buildings*, Vol. 165, pp. 301–320. https://doi.org/10.1016/j.enbuild.2018.01.017

Ahmad, T., Zhang, H. and Yan, B. (2020), "A Review on Renewable Energy and Electricity Requiremen Forecasting Models for Smart Grid and Buildings", *Sustainable Cities and Society*, Vol. 55, Article 102052. https://doi.org/10.1016/j.scs.2020.102052

Ahn, H., Sun, K. and Kim, K. (2021), "Comparison of Missing Data Imputation Methods in Time Series Forecasting", *Computers, Materials and Continua*, Vol. 70, pp. 767–779. https://doi.org/10.32604/cmc.2022.019369

Akhtar, S., Shahzad, S., Zaheer, A., Ullah, H.S., Kilic, H., Gono, R., Jasiński, M. and Leonowicz, Z. (2023), "Short-Term Load Forecasting Models: A Review of Challenges, Progress, and the Road Ahead", *Energies*, Vol. 16 No. 10, Article 4060. https://doi.org/10.3390/en16104060

Alanne, K. and Sierla, S. (2022), "An Overview of Machine Learning Applications for Smart Buildings", *Sustainable Cities and Society*, Vol. 76, Article 103445. https://doi.org/10.1016/j.scs.2021.103445

Alexiou, K., Pariotis, E.G. and Leligou, H.C. (2023), "Sensor Data Quality in Ships: A Time Series Forecasting Approach to Compensate for Missing Data and Drift in Measurements of Speed through Water Sensors", *Designs*, Vol. 7 No. 2, Article 46. https://doi.org/10.3390/designs7020046

Alizadegan, H., Rashidi Malki, B., Radmehr, A., Karimi, H. and Ilani, M.A. (2025), "Comparative Study of Long Short-Term Memory (LSTM), Bidirectional LSTM, and Traditional Machine Learning Approaches for Energy Consumption Prediction", *Energy Exploration & Exploitation*, Vol. 43 No. 1, pp. 281–301. https://doi.org/10.1177/01445987241269496

AlSalehy, A.S. and Bailey, M. (2025), "Improving Time Series Data Quality: Identifying Outliers and Handling Missing Values in a Multilocation Gas and Weather Dataset", *Smart Cities*, Vol. 8 No. 3, Article 82. https://doi.org/10.3390/smartcities8030082

Anan, M., Kanaan, K., Benhaddou, D., Nasser, N., Qolomany, B., Talei, H. and Sawalmeh, A. (2024), "Occupant-Aware Energy Consumption Prediction in Smart Buildings Using a LSTM Model and Time Series Data", *Energies*, Vol. 17 No. 24. https://doi.org/10.3390/en17246451

Ardabili, S., Abdolalizadeh, L., Mako, C., Torok, B. and Mosavi, A. (2022), "Systematic Review of Deep Learning and Machine Learning for Building Energy", *Frontiers in Energy Research*, Vol. 10. https://doi.org/10.3389/fenrg.2022.786027

Aydin, S. (2022), "Time Series Analysis and Some Applications in Medical Research", Journal of Mathematics and Statistics Studies, Vol. 3, pp. 31–36. https://doi.org/10.32996/jmss.2022.3.2.3

Babai, M.Z., Dallery, Y., Boubaker, S. and Kalai, R. (2019), "A New Method to Forecast Intermittent Demand in the Presence of Inventory Obsolescence", *International Journal of Production Economics*, Vol. 209, pp. 30–41. https://doi.org/10.1016/j.ijpe.2018.01.026

Batoure Bamana, A., Shafiee Kamalabad, M. and Oberski, D.L. (2024), "A Systematic Literature Review of Time Series Methods Applied to Epidemic Prediction", *Informatics in Medicine Unlocked*, Vol. 50, Article 101571. https://doi.org/10.1016/j.imu.2024.101571

Beckel, C., Sadamori, L., Staake, T. and Santini, S. (2014), "Revealing Household Characteristics from Smart Meter Data", *Energy*, Vol. 78, pp. 397–410. https://doi.org/10.1016/j.energy.2014.10.025

Bedi, G., Venayagamoorthy, G. and Singh, R. (2020), "Development of an IoT Driven Building Environment for Prediction of Electric Energy Consumption", *IEEE Internet of Things Journal*, pp. 1–1. https://doi.org/10.1109/JIOT.2020.2975847

Benti, N.E., Chaka, M.D. and Semie, A.G. (2023), "Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects", *Sustainability*, Vol. 15 No. 9, Article 9. https://doi.org/10.3390/su15097087

Bezzar, N.E.-H., Laimeche, L., Meraoumia, A. and Houam, L. (2022), "Data analysis-based time series forecast for managing household electricity consumption", *Demonstratio Mathematica*, Vol. 55 No. 1, pp. 900–921. https://doi.org/10.1515/dema-2022-0176

Boukaf, M., Fadli, F. and Meskin, N. (2024), "A comprehensive review of digital twin technology in building energy consumption forecasting", *IEEE Access*, Vol. 12, pp. 187778–187799. https://doi.org/10.1109/ACCESS.2024.3498107

Bouktif, S., Fiaz, A., Ouni, A. and Serhani, M.A. (2020), "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting", *Energies*, Vol. 13 No. 2, Article 2. https://doi.org/10.3390/en13020391

Bourdeau, M., Zhai, X.Q., Nefzaoui, E., Guo, X. and Chatellier, P. (2019), "Modeling and forecasting building energy consumption: A review of data-driven techniques", *Sustainable Cities and Society*, Vol. 48, Article 101533. https://doi.org/10.1016/j.scs.2019.101533

Buyo, N., Sheikh-Akbari, A. and Saleem, F. (2025), "An ensemble approach to predict a sustainable energy plan for London households", *Sustainability*, Vol. 17 No. 2, Article 2. https://doi.org/10.3390/su17020500

Chen, G., Lu, S., Zhou, S., Tian, Z., Kim, M.K., Liu, J. and Liu, X. (2025a), "A systematic review of building energy consumption prediction: From perspectives of load classification, data-driven frameworks, and future directions", *Applied Sciences*, Vol. 15 No. 6, Article 6. https://doi.org/10.3390/app15063086

Chen, Y., Gong, W., Obrecht, C. and Kuznik, F. (2025b), "A review of machine learning techniques for building electrical energy consumption prediction", *Energy and AI*, Vol. 21, Article 100518. https://doi.org/10.1016/j.egyai.2025.100518

Chen, Y., Guo, M., Chen, Z., Chen, Z. and Ji, Y. (2022), "Physical energy and data-driven models in building energy prediction: A review", *Energy Reports*, Vol. 8, pp. 2656–2671. https://doi.org/10.1016/j.egyr.2022.01.162

Chen, Y., Jia, W. and Wu, Q. (2024), "Fine-scale deep learning model for time series forecasting", *Applied Intelligence*, Vol. 54 No. 20, pp. 10072–10083. https://doi.org/10.1007/s10489-024-05701-w

Chen, Z., Xiao, F., Guo, F. and Yan, J. (2023), "Interpretable machine learning for building energy management: A state-of-the-art review", *Advances in Applied Energy*, Vol. 9, Article 100123. https://doi.org/10.1016/j.adapen.2023.100123

Choi, E., Cho, S. and Kim, D.K. (2020), "Power demand forecasting using long short-term memory (LSTM) deep-learning model for monitoring energy sustainability", *Sustainability*, Vol. 12 No. 3, Article 3. https://doi.org/10.3390/su12031109

Daissaoui, A., Boulmakoul, A., Karim, L. and Lbath, A. (2020), "IoT and big data analytics for smart buildings: A survey", *Procedia Computer Science*, Vol. 170, pp. 161–168. https://doi.org/10.1016/j.procs.2020.03.021

Dannecker, L. (2015), *Energy Time Series Forecasting: Efficient and Accurate Forecasting of Evolving Time Series from the Energy Domain*, Springer Fachmedien, Wiesbaden.

Das, A., Annaqeeb, M.K., Azar, E., Novakovic, V. and Kjærgaard, M.B. (2020), "Occupant-centric miscellaneous electric loads prediction in buildings using state-of-the-art deep learning methods", *Applied Energy*, Vol. 269, Article 115135. https://doi.org/10.1016/j.apenergy.2020.115135

Deb, C., Zhang, F., Yang, J., Lee, S.E. and Shah, K.W. (2017), "A review on time series forecasting techniques for building energy consumption", *Renewable and Sustainable Energy Reviews*, Vol. 74, pp. 902–924. https://doi.org/10.1016/j.rser.2017.02.085

Deng, D., Li, J., Zhao, J., Zhang, Z., Hu, W. and Huang, Q. (2020), "A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network", *IEEE Transactions on Industrial Informatics,* vol. 17, pp. 2443-2452. https://doi.org/10.1109/TII.2020.3000184

Djedouboum, A.C., Abba Ari, A.A., Gueroui, A.M., Mohamadou, A. and Aliouat, Z. (2018), "Big data collection in large-scale wireless sensor networks", *Sensors*, Vol. 18 No. 12, Article 12. https://doi.org/10.3390/s18124474

Eddaoudi, Z., Aarab, Z., Boudmen, K., Elghazi, A. and Rahmani, M.D. (2024), "A brief review of energy consumption forecasting using machine learning models", *Procedia Computer Science*, Vol. 236, pp. 33–40. https://doi.org/10.1016/j.procs.2024.05.001

El Assri, N., Jallal, M.A., Chabaa, S. and Zeroual, A. (2025), "Enhancing building energy consumption prediction using LSTM, Kalman filter, and continuous wavelet transform", *Scientific African*, Vol. 27, Article e02560. https://doi.org/10.1016/j.sciaf.2025.e02560

Fallah, S.N., Deo, R.C., Shojafar, M., Conti, M. and Shamshirband, S. (2018), "Computational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions", *Energies*, Vol. 11 No. 3, Article 3. https://doi.org/10.3390/en11030596

Fan, C., Zhang, Y., Pan, Y., Li, X., Zhang, C., Yuan, R., Wu, D., Wang, W., Pei, J. and Huang, H. (2019), "Multi-horizon time series forecasting with temporal attention learning", in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, Anchorage, Alaska, pp. 2527–2535. https://doi.org/10.1145/3292500.3330662

Farea, A., Yli-Harja, O. and Emmert-Streib, F. (2024), "Understanding physics-informed neural networks: Techniques, applications, trends, and challenges", *AI*, Vol. 5 No. 3, Article 3. https://doi.org/10.3390/ai5030074

Fathi, S., Srinivasan, R., Fenner, A. and Fathi, S. (2020), "Machine learning applications in urban building energy performance forecasting: A systematic review", *Renewable and Sustainable Energy Reviews*, Vol. 133, Article 110287. https://doi.org/10.1016/j.rser.2020.110287

Fatima, S.S.W. and Rahimi, A. (2024), "A review of time-series forecasting algorithms for industrial manufacturing systems", *Machines*, Vol. 12 No. 6, Article 6. https://doi.org/10.3390/machines12060380

Felipe, O. and Emilio L., C. (2022), "Sensor data analytics: Challenges and methods for data-intensive applications", *Entropy*, Vol. 24, Article 850. https://doi.org/10.3390/e24070850

Fu, C.W. and Nguyen, T.T. (2003), "Models for long-term energy forecasting", in: *2003 IEEE Power Engineering Society General Meeting, Conference Proceedings*, IEEE, Toronto, pp. 235–239. https://doi.org/10.1109/PES.2003.1267174

Fu, Q., Nicholson, G.L. and Easton, J.M. (2024), "Understanding data quality in a data-driven industry context: Insights from the fundamentals", *Journal of Industrial Information Integration*, Vol. 42, Article 100729. https://doi.org/10.1016/j.jii.2024.100729

Gao, Z., Zhang, X., Wang, C., Xing, J., Deng, Z. and Chen, T. (2025), "Research on building energy consumption prediction based on hybrid GRU neural network", *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-025-10151-8

Ghosh, K. (2021), "Time series analysis: A brief history and its future challenges", *Indian Science Cruiser*, Vol. 34, pp. 22. https://doi.org/10.24906/isc/2020/v34/i5/206994

Gou, Y., Guo, C. and Qin, R. (2024), "Ultra short-term power load forecasting based on the fusion of Seq2Seq BiLSTM and multi-head attention mechanism", *PLOS ONE*, Vol. 19 No. 3, Article e0299632. https://doi.org/10.1371/journal.pone.0299632

Guerra-Santin, O. and Tweed, C.A. (2015), "In-use monitoring of buildings: An overview of data collection methods", *Energy and Buildings*, Vol. 93, pp. 189–207. https://doi.org/10.1016/j.enbuild.2015.02.042

Guo, Y., Shi, J., Guo, T., Guo, F., Lu, F. and Su, L. (2024), "Grey-box method for urban building energy modelling: Advancements and potentials", *Energies*, Vol. 17 No. 21, Article 21. https://doi.org/10.3390/en17215463

Hadjout, D., Sebaa, A., Torres, J.F. and Martínez-Álvarez, F. (2023), "Electricity consumption forecasting with outliers handling based on clustering and deep learning with application to the Algerian market", *Expert Systems with Applications*, Vol. 227, Article 120123. https://doi.org/10.1016/j.eswa.2023.120123

He, N., Liu, L., Qian, C., Zhang, L., Yang, Z. and Li, S. (2022), "A closed-loop data-fusion framework for air conditioning load prediction based on LBF", *Energy Reports*, Vol. 8, pp. 7724–7734. https://doi.org/10.1016/j.egyr.2022.05.289

Hitimana, E., Bajpai, G., Musabe, R., Sibomana, L. and Kayalvizhi, J. (2021), "Implementation of IoT framework with data analysis using deep learning methods for occupancy prediction in a building", *Future Internet*, Vol. 13 No. 3, Article 3. https://doi.org/10.3390/fi13030067

Hussien, A., Maksoud, A., Al-Dahhan, A., Abdeen, A. and Baker, T. (2025), "Machine learning model for predicting long-term energy consumption in buildings", *Discover Internet of Things*, Vol. 5 No. 1, Article 18. https://doi.org/10.1007/s43926-025-00115-7

Isah, A., Shin, H., Oh, S., Oh, S., Aliyu, I., Um, T. and Kim, J. (2023), "Digital twins temporal dependencies-based on time series using multivariate long short-term memory", *Electronics*, Vol. 12 No. 19, Article 19. https://doi.org/10.3390/electronics12194187

Jason, B. (2018), *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*, Machine Learning Mastery, Queensland.

Jenčová, S., Vašaničová, P., Košíková, M. and Miškufová, M. (2025), "A time series approach to forecasting financial indicators in the wholesale and retail trade", *World*, Vol. 6 No. 1, Article 1. https://doi.org/10.3390/world6010005

Jiang, Y., Yu, F., Tang, Y., Ouyang, C. and Li, F. (2025), "Trend-pattern unlimited fuzzy information granule-based LSTM model for long-term time-series forecasting", *International Journal of Approximate Reasoning*, Vol. 180, Article 109381. https://doi.org/10.1016/j.ijar.2025.109381

Jiang, Z., Wang, X., Li, H., Hong, T., You, F., Drgoňa, J., Vrabie, D. and Dong, B. (2025), "Physics-informed machine learning for building performance simulation – A review of a nascent field", *Advances in Applied Energy*, Vol. 18, Article 100223. https://doi.org/10.1016/j.adapen.2025.100223

Jin, Y., Yan, D., Chong, A., Dong, B. and An, J. (2021), "Building occupancy forecasting: A systematical and critical review", *Energy and Buildings*, Vol. 251, Article 111345. https://doi.org/10.1016/j.enbuild.2021.111345

Jogunola, O., Adebisi, B., Hoang, K.V., Tsado, Y., Popoola, S.I., Hammoudeh, M. and Nawaz, R. (2022), "CBLSTM-AE: A hybrid deep learning framework for predicting energy consumption", *Energies*, Vol. 15 No. 3, Article 810. https://doi.org/10.3390/en15030810

Kadir, J., Hanafi, A., Shareef, A., Khatib, H., Baharin, T., Sulaima, K., Hasan, A., Rehman, H., Hossain, J., Kadir, A., Hanafi, A., Shareef, H., Khatib, T., Baharin, K. and Sulaima, M.F. (2023), "A review on optimal energy management in commercial buildings", *Energies*, Vol. 16, Article 1609. https://doi.org/10.3390/en16041609

Khalil, M., McGough, A.S., Pourmirza, Z., Pazhoohesh, M. and Walker, S. (2022), "Machine learning, deep learning and statistical analysis for forecasting building energy consumption – A systematic review", *Engineering Applications of Artificial Intelligence*, Vol. 115, Article 105287. https://doi.org/10.1016/j.engappai.2022.105287

Kiartzis, S.J., Bakirtzis, A.G. and Petridis, V. (1995), "Short-term load forecasting using neural networks", *Electric Power Systems Research*, Vol. 33 No. 1, pp. 1–6. https://doi.org/10.1016/0378-7796(95)00920-D

Kim, S., Jung, S. and Baek, S.-M. (2019), "A model for predicting energy usage pattern types with energy consumption information according to the behaviors of single-person households in South Korea", *Sustainability*, Vol. 11 No. 1, Article 1. https://doi.org/10.3390/su11010245

Kim, S., Kang, S., Ryu, K.R. and Song, G. (2019), "Real-time occupancy prediction in a large exhibition hall using deep learning approach", *Energy and Buildings*, Vol. 199, pp. 216–222. https://doi.org/10.1016/j.enbuild.2019.06.043

Kolambe, M. (2024), "Forecasting the future: A comprehensive review of time series prediction techniques", *Journal of Electrical Systems*, Vol. 20, pp. 575–586. https://doi.org/10.52783/jes.1478

Kurte, K., Munk, J., Kotevska, O., Amasyali, K., Smith, R., McKee, E., Du, Y., Cui, B., Kuruganti, T. and Zandi, H. (2020), "Evaluating the adaptability of reinforcement learning-based HVAC control for residential houses", *Sustainability*, Vol. 12 No. 18, Article 18. https://doi.org/10.3390/su12187727

Laome, L., Wibawa, G., Raya, R., Makkulau and Asbahuna, A. (2021), "Forecasting time series data containing outliers with the ARIMA additive outlier method", in: *Journal of Physics: Conference Series*, IOP Publishing, Vol. 1899, Article 012106. https://doi.org/10.1088/1742-6596/1899/1/012106

Lavrinovica, I., Judvaitis, J., Laksis, D., Skromule, M. and Ozols, K. (2024), "A comprehensive review of sensor-based smart building monitoring and data gathering techniques", *Applied Sciences*, Vol. 14 No. 21, Article 21. https://doi.org/10.3390/app142110057

Le, X.-H., Ho, H.V., Lee, G. and Jung, S. (2019), "Application of long short-term memory (LSTM) neural network for flood forecasting", *Water*, Vol. 11 No. 7, Article 7. https://doi.org/10.3390/w11071387

Lee, B., Lee, H. and Ahn, H. (2020), "Improving load forecasting of electric vehicle charging stations through missing data imputation", *Energies*, Vol. 13 No. 18, Article 18. https://doi.org/10.3390/en13184893

Leeraksakiat, P. and Pora, W. (2020), "Occupancy forecasting using LSTM neural network and transfer learning", in: *Proceedings of the 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, IEEE, Chiang Mai, pp. 470–473. https://doi.org/10.1109/ECTI-CON49241.2020.9158103

Li, D., Qi, Z., Zhou, Y. and Elchalakani, M. (2025), "Machine learning applications in building energy systems: Review and prospects", *Buildings*, Vol. 15 No. 4, Article 4. https://doi.org/10.3390/buildings15040648

Li, G., Li, F., Xu, C. and Fang, X. (2022a), "A spatial-temporal layer-wise relevance propagation method for improving interpretability and prediction accuracy of LSTM building energy prediction", *Energy and Buildings*, Vol. 271, Article 112317. https://doi.org/10.1016/j.enbuild.2022.112317

Li, W., Gong, G., Fan, H., Peng, P., Chun, L. and Fang, X. (2021), "A clustering-based approach for 'cross-scale' load prediction on building level in HVAC systems", *Applied Energy*, Vol. 282, Article 116223. https://doi.org/10.1016/j.apenergy.2020.116223

Li, W., Gong, G., Ren, Z., Ouyang, Q., Peng, P., Chun, L. and Fang, X. (2022b), "A method for energy consumption optimization of air conditioning systems based on load prediction and energy flexibility", *Energy*, Vol. 243, 123111. https://doi.org/10.1016/j.energy.2022.123111

Li, X. (2024), "Time series forecasting with missing values using multiple models: An application to wind energy systems", *Preprints*, No. 2024021547. https://doi.org/10.20944/preprints202402.1547.v1

Liu, F. and Liang, C. (2024), "Short-term power load forecasting based on AC-BiLSTM model", *Energy Reports*, Vol. 11, pp. 1570–1579. https://doi.org/10.1016/j.egyr.2024.01.026

Liu, H., Liang, J., Liu, Y. and Wu, H. (2023), "A review of data-driven building energy prediction", *Buildings*, Vol. 13 No. 2, Article 2. https://doi.org/10.3390/buildings13020532

Lu, C., Li, S. and Lu, Z. (2022), "Building energy prediction using artificial neural networks: A literature survey", *Energy and Buildings*, Vol. 262, 111718. https://doi.org/10.1016/j.enbuild.2021.111718

Mahajan, A., Das, S., Su, W. and Bui, V.-H. (2024), "Bayesian-neural-network-based approach for probabilistic prediction of building-energy demands", *Sustainability*, Vol. 16 No. 22, Article 22. https://doi.org/10.3390/su16229943

Mahmood, M. and Popoola, O. (2025), "Occupant behavior and energy efficiency in higher education buildings: A systematic review", *International Journal of Development and Sustainability*, pp. 476–497. https://doi.org/10.63212/IJDS25071601

Maleki, N., Lundström, O., Musaddiq, A., Jeansson, J., Olsson, T. and Ahlgren, F. (2024), "Future energy insights: Time-series and deep learning models for city load forecasting", *Applied Energy*, Vol. 374, 124067. https://doi.org/10.1016/j.apenergy.2024.124067

Mariano-Hernández, D., Hernández-Callejo, L., García, F.S., Duque-Perez, O. and Zorita-Lamadrid, A.L. (2020), "A review of energy consumption forecasting in smart buildings: Methods, input variables, forecasting horizon and metrics", *Applied Sciences*, Vol. 10 No. 23, Article 23. https://doi.org/10.3390/app10238323

Martín-Lopo, M.M., Boal, J. and Sánchez-Miralles, Á. (2020), "A literature review of IoT energy platforms aimed at end users", *Computer Networks*, Vol. 171, 107101. https://doi.org/10.1016/j.comnet.2020.107101

Massana, J., Pous, C., Burgas, L., Melendez, J. and Colomer, J. (2015), "Short-term load forecasting in a non-residential building contrasting models and attributes", *Energy and Buildings*, Vol. 92, pp. 322–330. https://doi.org/10.1016/j.enbuild.2015.02.007

Mat Daut, M.A., Hassan, M.Y., Abdullah, H., Rahman, H.A., Abdullah, M.P. and Hussin, F. (2017), "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review", *Renewable and Sustainable Energy Reviews*, Vol. 70, pp. 1108–1118. https://doi.org/10.1016/j.rser.2016.12.015

Mathumitha, R., Rathika, P. and Manimala, K. (2024), "Intelligent deep learning techniques for energy consumption forecasting in smart buildings: A review", *Artificial Intelligence Review*, Vol. 57 No. 2, p. 35. https://doi.org/10.1007/s10462-023-10660-8

Mboweni, I.V., Ramotsoela, D.T. and Abu-Mahfouz, A.M. (2023), "Hydraulic data preprocessing for machine learning-based intrusion detection in cyber-physical systems", *Mathematics*, Vol. 11 No. 8, Article 8. https://doi.org/10.3390/math11081846

Mitra, D., Steinmetz, N., Chu, Y. and Cetin, K.S. (2020), "Typical occupancy profiles and behaviors in residential buildings in the United States", *Energy and Buildings*, Vol. 210, 109713. https://doi.org/10.1016/j.enbuild.2019.109713

Mojtahedi, F.F., Yousefpour, N., Chow, S.H. and Cassidy, M. (2025), "Deep learning for time series forecasting: Review and applications in geotechnics and geosciences", *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-025-10244-5

Mora, D., Fajilla, G., Austin, M.C. and De Simone, M. (2019), "Occupancy patterns obtained by heuristic approaches: Cluster analysis and logical flowcharts. A case study in a university office", *Energy and Buildings*, Vol. 186, pp. 147–168. https://doi.org/10.1016/j.enbuild.2019.01.023

Mylonas, A., Tsangrassoulis, A. and Pascual, J. (2023), "A systematic review of time user surveys-questionnaires and monitoring variables used to track occupant behaviour in residential buildings", *IOP*

Conference Series: Earth and Environmental Science, Vol. 1196 No. 1, 012102. https://doi.org/10.1088/1755-1315/1196/1/012102

Mystakidis, A., Koukaras, P., Tsalikidis, N., Ioannidis, D. and Tjortjis, C. (2024), "Energy forecasting: A comprehensive review of techniques and technologies", *Energies*, Vol. 17 No. 7, Article 7. https://doi.org/10.3390/en17071662

Newsham, G.R. and Birt, B.J. (2010), "Building-level occupancy data to improve ARIMA-based electricity use forecasts", in: *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ACM, Zürich, pp. 13–18. https://doi.org/10.1145/1878431.1878435

Ngo, N.-T., Pham, A.-D., Truong, T.T.H., Truong, N.-S. and Huynh, N.-T. (2022), "Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings", *Scientific Reports*, Vol. 12 No. 1, 15775. https://doi.org/10.1038/s41598-022-19935-6

Nugroho, H. (2023), "A review: Data quality problem in predictive analytics", *International Journal of Applied Information Technology (IJAIT)*, Vol. 7, p. 79. https://doi.org/10.25124/ijait.v7i02.5980

Olu-Ajayi, R., Alaka, H., Owolabi, H., Akanbi, L. and Ganiyu, S. (2023), "Data-driven tools for building energy consumption prediction: A review", *Energies*, Vol. 16 No. 6, Article 6. https://doi.org/10.3390/en16062574

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F. and Ajayi, S. (2022), "Building energy consumption prediction or residential buildings using deep learning and other machine learning techniques", *Journal of Building Engineering*, Vol. 45, 103406. https://doi.org/10.1016/j.jobe.2021.103406

Panja, M., Chakraborty, T., Kumar, U. and Hadid, A. (2024), "Probabilistic autoregressive neural networks for accurate long-range forecasting", in: Luo, B., Cheng, L., Wu, Z.-G., Li, H. and Li, C. (eds.), *Neural Information Processing*, Springer Nature, pp. 457–477. https://doi.org/10.1007/978-981-99-8178-6_35

Phyo, P.-P., Byun, Y.-C. and Park, N. (2022), "Short-term energy forecasting using machine-learning-based ensemble voting regression", *Symmetry*, Vol. 14 No. 1, Article 1. https://doi.org/10.3390/sym14010160

Piotrowski, P., Rutyna, I., Baczyński, D. and Kopyt, M. (2022), "Evaluation metrics for wind power forecasts: A comprehensive review and statistical analysis of errors", *Energies*, Vol. 15 No. 24, 9657. https://doi.org/10.3390/en15249657

Popoola, O. (2018), "Computational intelligence modelling based on variables interlinked with behavioral tendencies for energy usage profile – A necessity", *Renewable and Sustainable Energy Reviews*, Vol. 82, pp. 60–72. https://doi.org/10.1016/j.rser.2017.09.020

Popoola, O., Munda, J. and Mpanda, A. (2018), "Residential lighting load profile modelling: ANFIS approach using weighted and non-weighted data", *Energy Efficiency*, Vol. 11, pp. 1–? (online first). https://doi.org/10.1007/s12053-017-9557-9

Popoola, O., Ramokone, A. and Awelewa, A. (2024), "Adept domestic energy load profile development using computational intelligence-based modelling", *International Transactions on Electrical Energy Systems*, 2024(1), 6656970. https://doi.org/10.1155/2024/6656970

Qiao, Q., Yunusa-Kaltungo, A. and Edwards, R.E. (2021), "Towards developing a systematic knowledge trend for building energy consumption prediction", *Journal of Building Engineering*, Vol. 35, 101967. https://doi.org/10.1016/j.jobe.2020.101967

Qiu, H., Hu, R., Chen, J. and Yuan, Z. (2025), "Short-term electricity load forecasting based on CEEMDAN and improved SSA–CNN–BiLSTM", *Mathematics*, Vol. 13 No. 5, Article 5. https://doi.org/10.3390/math13050813

Qolomany, B., Al-Fuqaha, A. and Gupta, A. (2017), "Role of deep LSTM neural networks and WiFi networks in support of occupancy prediction in smart buildings", *arXiv preprint*. https://doi.org/10.48550/arXiv.1711.10355

Ramokone, A., Popoola, O., Awelewa, A. and Temitope, A. (2021), "A review on behavioural propensity for building load and energy profile development – Model inadequacy and improved approach", *Sustainable Energy Technologies and Assessments*, Vol. 45, 101235. https://doi.org/10.1016/j.seta.2021.101235

Ramos, D., Teixeira, B., Faria, P., Gomes, L., Abrishambaf, O. and Vale, Z. (2020), "Use of sensors and analyzers data for load forecasting: A two stage approach", *Sensors*, Vol. 20 No. 12, Article 12. https://doi.org/10.3390/s20123524

Reis, J., Muñoz Tabora, J., Lima, M., Monteiro, F., Monteiro, S., Bezerra, U. and Tostes, M. (2025), "Medium and long term energy forecasting methods: A literature review", *IEEE Access*, Vol. PP, pp. 1–1. https://doi.org/10.1109/ACCESS.2025.3540999

Romanuke, V. (2021), "Time series smoothing improving forecasting", *Applied Computer Systems*, Vol. 26, pp. 60–70. https://doi.org/10.2478/acss-2021-0008

Ruan, Y., Wang, G., Meng, H. and Qian, F. (2022), "A hybrid model for power consumption forecasting using VMD-based the long short-term memory neural network", *Frontiers in Energy Research*, Vol. 9, Article 772508. https://doi.org/10.3389/fenrg.2021.772508

Runge, J. and Zmeureanu, R. (2019), "Forecasting energy use in buildings using artificial neural networks: A review", *Energies*, Vol. 12 No. 17, Article 17. https://doi.org/10.3390/en12173254

Runpeng, L. and Seong-Yoon, S. (2025), "A review of traffic flow prediction methods in intelligent transportation system construction", *Applied Sciences*, Vol. 15, pp. 3866. https://doi.org/10.3390/app15073866

Serrano, A.L.M., Rodrigues, G.A.P., Martins, P.H. dos S., Saiki, G.M., Filho, G.P.R., Gonçalves, V.P. and Albuquerque, R. de O. (2024), "Statistical comparison of time series models for forecasting Brazilian monthly energy demand using economic, industrial, and climatic exogenous variables", *Applied Sciences*, Vol. 14 No. 13, Article 13. https://doi.org/10.3390/app14135846

Seyedzadeh, S., Pour Rahimian, F., Glesk, I. and Roper, M. (2018), "Machine learning for estimation of building energy consumption and performance: A review", *Visualization in Engineering*, Vol. 6, Article 5. https://doi.org/10.1186/s40327-018-0064-7

Sha, H., Xu, P., Yan, C., Ji, Y., Zhou, K. and Chen, F. (2022), "Development of a key-variable-based parallel HVAC energy predictive model", *Building Simulation*, Vol. 15 No. 7, pp. 1193–1208. https://doi.org/10.1007/s12273-021-0885-0

Shaikh, A.K., Nazir, A., Khan, I. and Shah, A.S. (2022), "Short term energy consumption forecasting using neural basis expansion analysis for interpretable time series", *Scientific Reports*, Vol. 12 No. 1, 22562. https://doi.org/10.1038/s41598-022-26499-y

Shin, J., Moon, H., Chun, C.-J., Sim, T., Kim, E. and Lee, S. (2024), "Enhanced data processing and machine learning techniques for energy consumption forecasting", *Electronics*, Vol. 13 No. 19, Article 19. https://doi.org/10.3390/electronics13193885

Shrestha, A., Ghimire, B. and Gonzalez-Longatt, F. (2021), "A Bayesian model to forecast the time series kinetic energy data for a power system", *Energies*, Vol. 14 No. 11, Article 11. https://doi.org/10.3390/en14113299

Singh, S. (1999), "Noise impact on time-series forecasting using an intelligent pattern matching technique", *Pattern Recognition*, Vol. 32 No. 8, pp. 1389–1398. https://doi.org/10.1016/S0031-3203(98)00174-5

Son, N. and Shin, Y. (2023), "Short- and medium-term electricity consumption forecasting using Prophet and GRU", *Sustainability*, Vol. 15 No. 22, Article 22. https://doi.org/10.3390/su152215860

Soyiri, I.N. and Reidpath, D.D. (2013), "An overview of health forecasting", *Environmental Health and Preventive Medicine*, Vol. 18 No. 1, pp. 1–9. https://doi.org/10.1007/s12199-012-0294-6

Svoboda, R., Kotik, V. and Platos, J. (2024), "Data-driven multi-step energy consumption forecasting with complex seasonality patterns and exogenous variables: Model accuracy assessment in change point neighborhoods", *Applied Soft Computing*, Vol. 150, 111099. https://doi.org/10.1016/j.asoc.2023.111099

Tarmanini, C., Sarma, N., Gezegin, C. and Ozgonenel, O. (2023), "Short term load forecasting based on ARIMA and ANN approaches", *Energy Reports*, Vol. 9, pp. 550–557. https://doi.org/10.1016/j.egyr.2023.01.060

Teh, H.Y., Kempa-Liehr, A.W. and Wang, K.I.-K. (2020), "Sensor data quality: A systematic review", *Journal of Big Data*, Vol. 7 No. 1, 11. https://doi.org/10.1186/s40537-020-0285-1

Tekler, Z. and Chong, A. (2022), "Occupancy prediction using deep learning approaches across multiple space types: A minimum sensing strategy", *Building and Environment*, Vol. 226, 109689. https://doi.org/10.1016/j.buildenv.2022.109689

Thangjam, A., Jaipuria, S. and Dadabada, P.K. (2023), "Time-varying approaches for long-term electric load forecasting under economic shocks", *Applied Energy*, Vol. 333, 120602. https://doi.org/10.1016/j.apenergy.2022.120602

Truong, N.-S., Ngo, N.-T. and Pham, A.-D. (2021), "Forecasting time-series energy data in buildings using an additive artificial intelligence model for improving energy efficiency", *Computational Intelligence and Neuroscience*, 2021(1), 6028573. https://doi.org/10.1155/2021/6028573

Ullah, F.U.M., Khan, N., Hussain, T., Lee, M.Y. and Baik, S.W. (2021), "Diving deep into short-term electricity load forecasting: Comparative analysis and a novel framework", *Mathematics*, Vol. 9 No. 6, Article 6. https://doi.org/10.3390/math9060611

Wang, J., Chen, X., Zhang, F., Chen, F. and Xin, Y. (2021), "Building load forecasting using deep neural network with efficient feature fusion", *Journal of Modern Power Systems and Clean Energy*, Vol. 9, pp. 160–169. https://doi.org/10.35833/MPCE.2020.000321

Wang, L., Chen, J., Wang, W., Song, R., Zhang, Z. and Yang, G. (2022), "Review of time series traffic forecasting methods", in: *Proceedings of ICCR 2022*, IEEE, p. 5. https://doi.org/10.1109/ICCR55715.2022.10053870

Wang, L., Zou, H., Su, J., Li, L. and Chaudhry, S. (2013), "An ARIMA-ANN hybrid model for time series forecasting", *Systems Research and Behavioral Science*, Vol. 30 No. 3, pp. 244–259. https://doi.org/10.1002/sres.2179

Wang, X. and Wang, C. (2019), "Time series data cleaning: A survey", *IEEE Access*, Vol. PP, pp. 1–1. https://doi.org/10.1109/ACCESS.2019.2962152

Wang, Z., Hong, T. and Piette, M.A. (2019), "Predicting plug loads with occupant count data through a deep learning approach", *Energy*, Vol. 181, pp. 29–42. https://doi.org/10.1016/j.energy.2019.05.138

Wang, Z., Hong, Y., Huang, L., Zheng, M., Yuan, H. and Zeng, R. (2025), "A comprehensive review and future research directions of ensemble learning models for predicting building energy consumption", *Energy and Buildings*, Vol. 335, 115589. https://doi.org/10.1016/j.enbuild.2025.115589

Wei, N., Li, C., Peng, X., Zeng, F. and Lu, X. (2019b), "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review", *Journal of Petroleum Science and Engineering*, Vol. 181, 106187. https://doi.org/10.1016/j.petrol.2019.106187

Wei, Y., Xia, L., Pan, S., Wu, J., Zhang, X., Han, M., Zhang, W., Xie, J. and Li, Q. (2019a), "Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks", *Applied Energy*, Vol. 240, pp. 276–294. https://doi.org/10.1016/j.apenergy.2019.02.056

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M. and Zhao, X. (2018), "A review of data-driven approaches for prediction and classification of building energy consumption", *Renewable and Sustainable Energy Reviews*, Vol. 82, pp. 1027–1047. https://doi.org/10.1016/j.rser.2017.09.108

Wu, Y., Chen, S., Jin, Y., Xu, H., Zhou, X., Wang, X., Chong, A., Li, J. and Yan, D. (2024), "Novel occupancy detection method based on convolutional neural network model using PIR sensor and smart meter data", *Advanced Engineering Informatics*, Vol. 62, 102589. https://doi.org/10.1016/j.aei.2024.102589

Xie, Y., Hu, P., Zhu, N., Lei, F., Xing, L., Xu, L. and Sun, Q. (2020), "A hybrid short-term load forecasting model and its application in ground source heat pump with cooling storage system", *Renewable Energy*, Vol. 161, pp. 1244–1259. https://doi.org/10.1016/j.renene.2020.07.142

Yaro, A.S., Maly, F. and Prazak, P. (2023), "Outlier detection in time-series receive signal strength observation using Z-score method with Sn scale estimator for indoor localization", *Applied Sciences*, Vol. 13 No. 6, Article 6. https://doi.org/10.3390/app13063900

Yaro, A.S., Maly, F., Prazak, P. and Malý, K. (2024), "Outlier detection performance of a modified Z-score method in time-series RSS observation with hybrid scale estimators", *IEEE Access*, Vol. 12, pp. 12785–12796. https://doi.org/10.1109/ACCESS.2024.3356731

Yin, Q., Han, C., Li, A., Liu, X. and Liu, Y. (2024), "A review of research on building energy consumption prediction models based on artificial neural networks", *Sustainability*, Vol. 16 No. 17, Article 17. https://doi.org/10.3390/su16177805

Yu, Z., Chen, H., Qi, L. and Zhou, H. (2025), "Influence of sensor positional uncertainties on prediction accuracy of inverse heat conduction problems", *Journal of Thermophysics and Heat Transfer*, pp. 1–12. https://doi.org/10.2514/1.T7141

Zhang, H., Liu, Y., Zhang, C. and Li, N. (2025), "Machine learning methods for weather forecasting: A survey", Atmosphere, Vol. 16 No. 1, Article 1. https://doi.org/10.3390/atmos16010082

Zhao, Y., Zhang, C., Zhang, Y., Wang, Z. and Li, J. (2020), "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis", *Energy and Built Environment*, Vol. 1 No. 2, pp. 149–164. https://doi.org/10.1016/j.enbenv.2019.11.003

Zhou, X., Lu, Y., Hu, S., Yang, Z. and Yan, D. (2023), "New perspectives on temporal changes in occupancy characteristics of residential buildings", *Journal of Building Engineering*, Vol. 64, 105590. https://doi.org/10.1016/j.jobe.2022.105590

Zhu, J., Zhao, Z., Zheng, X., An, Z., Guo, Q., Li, Z., Sun, J. and Guo, Y. (2023), "Time-series power forecasting for wind and solar energy based on the SL-Transformer", *Energies*, Vol. 16 No. 22, Article 22. https://doi.org/10.3390/en16227610

Zimmermann, M. and Ziel, F. (2025), "Efficient mid-term forecasting of hourly electricity load using generalized additive models", *Applied Energy*, Vol. 388, 125444. https://doi.org/10.1016/j.apenergy.2025.125444