Renewable and Sustainable Energy Reviews 74 (2017) 902–924

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

A review on time series forecasting techniques for building energy consumption

Chirag Deb[a], Fan Zhang[b], Junjing Yang[a,*], Siew Eang Lee[a], Kwok Wei Shah[a]

[a] Department of Building, School of Design and Environment, National University of Singapore, Singapore 117566, Singapore
[b] Institute of Systems Science, National University of Singapore, Singapore 119615, Singapore

ARTICLE INFO

Keywords:
Building energy
Time series forecasting
Energy forecasting
Machine learning
Hybrid models

ABSTRACT

Energy consumption forecasting for buildings has immense value in energy efficiency and sustainability research. Accurate energy forecasting models have numerous implications in planning and energy optimization of buildings and campuses. For new buildings, where past recorded data is unavailable, computer simulation methods are used for energy analysis and forecasting future scenarios. However, for existing buildings with historically recorded time series energy data, statistical and machine learning techniques have proved to be more accurate and quick. This study presents a comprehensive review of the existing machine learning techniques for forecasting time series energy consumption. Although the emphasis is given to a single time series data analysis, the review is not just limited to it since energy data is often co-analyzed with other time series variables like outdoor weather and indoor environmental conditions. The nine most popular forecasting techniques that are based on the machine learning platform are analyzed. An in-depth review and analysis of the 'hybrid model', that combines two or more forecasting techniques is also presented. The various combinations of the hybrid model are found to be the most effective in time series energy forecasting for building.

1. Introduction

The International Energy Agency has identified energy efficiency in buildings as one of the five measures to secure long-term decarbonisation of the energy sector[1] [1]. Along with environmental benefits, building energy efficiency also presents vast economic benefits. Buildings with efficient energy systems and management strategies have much lower operating costs. Many countries have now accelerated the implementation of energy codes and regulations for various building types. These regulations outline basic requirements to achieve an energy efficient design for new buildings with a view to reduce the final energy consumption and related CO_2 emissions. In addition, many computer softwares have also been developed and widely implemented for energy efficient design of new buildings. Some of the most popular ones are EnergyPlus, DOE-2, eQUEST, IES, ECOTECT etc. A detailed study on the available computer-aided building energy analysis techniques and software tools is available in [2,3]. These regulations and computer-aided tools pertain to new buildings and are indeed very effective. However, once the building is functional, many factors govern the energy behavior of a building such as weather conditions, occupancy schedule, thermal properties of building materials, complex interactions of the energy systems like HVAC and lighting etc. Due to these complex interactions, accurate computation of energy consumption through computer simulation modeling is very difficult. For these reasons, data driven techniques for building energy analysis of existing buildings are very crucial. These techniques rely on past recorded data and attempt to model the energy consumption based on previous energy use patterns. Other factors influencing energy consumption can be used to improve the accuracy of such time series models. These techniques that make use of past data often fall under the of 'machine learning' and have been actively applied to building energy forecasting studies in the last two decades. The advantages and disadvantages of such data-driven techniques are presented in Table 1. This paper presents a comprehensive review of nine of the most popular machine learning techniques. The details of the techniques are described in later sections.

1.1. Significance to building performance optimization

To achieve an optimum level of energy performance in buildings, installation of efficient energy systems should be followed by appropriate operational and management strategies. This requires contin-

* Corresponding author.
E-mail address: bdgyj@nus.edu.sg (J. Yang).
[1] © OECD/IEA 2015 *World Energy Outlook Special Report*, IEA Publishing.
License: [http://www.iea.org/t & c/termsandconditions/].

http://dx.doi.org/10.1016/j.rser.2017.02.085
Received 4 July 2016; Received in revised form 25 January 2017; Accepted 27 February 2017

Table 1
Comparison between data-driven and deterministic techniques for forecasting.

Technique	Advantages	Disadvantages
Data-driven	1. Very fast in computation with real-time data	1. Requires past recorded data
	1. Suitable for non-linear modeling	1. Non-transparent and confined
	1. Often more accurate than deterministic models	1. Difficult to generalize
Deterministic	1. Based on the science of building physics	1. Difficult to model real scenarios
	1. Transparent and no training data needed	1. Data unavailability of building properties
	1. Easy to generalize	1. Not very accurate

uous monitoring and management of the time series energy data along with other factors influencing energy performance of buildings. In relation to continuous monitoring and management of energy consumption in existing buildings, forecasting plays a significant role. It can provide a set of boundary conditions and targets for the building facility managers and owners within which the building's energy consumption should ideally fall (daily, weekly, monthly, and annual targets). As the time series forecasting model learns from previous energy consumption usage patterns, a gradual increase in the forecasted energy consumption values over a period of time may also notify the facility managers on the maintenance aspects of the building and energy systems. Besides the time series forecasting approach, other non-time series approaches can be adopted for building optimization purposes and they can also be combined with other computer simulation models like EnergyPlus to derive occupancy and other operational factors [4]. Yang et al. applied simulation based energy optimization for a test building in Spain. A web-based parallel genetic algorithm (GA) optimization framework making use of distributed computing resources was used to reduce the computation time [5]. Petri et al. presented a modular based optimization system that combines energy simulation and optimization using artificial neural network [6]. This helped in generating optimal set-points for a large scale building facility. The application showed significant energy reduction (kWh) in a real scenario. However, this required the building to be equipped with sensors and actuators for monitoring, control, and optimization. This may not be the case with most of the existing building infrastructure. Such challenges are discussed in detail by Aste et al. [7]. However, it is also noted that a relevant energy saving potential in buildings is related to commissioning, performance tracking and advanced control strategies. This being dependent on many factors including financial resources, policy support, green awareness, green material and technology etc. [8]. Zong et al. discussed on the challenges of implementing an economic model predictive control strategy (EMPC) for smart buildings [9]. It was observed that there are still challenges in application of model predictive control including the compromise between simplification and complication of building thermal dynamic modeling and balance between multi-energy systems. Realizing the challenges in integration of building performance data with other building related data, Hu et al. presented a novel method of linking traditionally disconnected data to construct data sources to enable in-depth and insightful building performance assessment [10].

Time series forecasting is integral to building performance optimization. Any optimization technique requires information either on future scenarios or in finding the best solutions against a test criterion. Machine learning techniques are useful in this regard and are often made use in solving these two problems. However, this review focuses on the time series forecasting aspects of building optimization rather than looking at the optimization problem in total (Fig. 1). An integration of these two shall be dealt with in a separate review.

1.2. Objectives of the review

Recent review studies on energy forecasting provide detailed accounts of the existing forecasting models and their classification. Zhao and Magoules reviewed and classified the existing methods for building energy consumption prediction into five categories [11]. Hippert et al. presented a review on short-term load forecasting [12]. Suganthi and Samuel presented a review on energy demand models for demand forecasting [13]. Fumo presented a review on building energy estimation and also studied the way the estimation models are classified [14]. Martinez-Alvarez et al. presented a survey on data mining techniques for time series forecasting of electricity [15]. The survey focused on the characteristics of the models and their configurations. Raza and Khosravi presented a review on short-term load forecasting techniques based on Artificial Intelligence (AI) techniques [16]. A recent study by Mat Daut et al. presented a review on building electrical energy consumption forecasting analysis using conventional and AI methods [17]. It was observed that the hybrid combination of SVM and swarn intelligence (SI) method has superior performance compared to other methods. The superior performance of the hybrid and ensemble models was also highlighted in recent review studies [18,19].

All these reviews provide vital information on energy forecasting models on different scales and emphasize on the superior performance of the hybrid models. A forecasting model can either be based on static data that usually fits a dependent variable to a set of independent variable, or it can make use of a single or parallel time series data. This study emphasizes on the forecasting techniques using time series data, which is also reflected in the title of this review. The significance to time series analysis is due to the increased awareness in real-time data collection and monitoring. Time series energy consumption can also be clubbed with time series data of indoor environmental conditions. With more sensors being deployed in buildings and more time series data being gathered, a suitable framework to analyze and to identify the forecasting capabilities is important. This review aims to understand the existing time series forecasting techniques and present their advantages and challenges. A detailed assessment of the hybrid model is also presented due to it being increasingly used in the literature. Since the hybrid model combinations are many, these are critically reviewed in a later section following the critical review of major, established techniques like ANN and ARIMA. This review paper shall also provide a basis of qualitative and quantitative comparison for all the 9 techniques mentioned here. It is to be noted that the hybrid model is considered as one technique among the 9 techniques presented. Within the hybrid model, there are a total of 29 combinations that are covered in this review. The objectives of this review paper are:

- To provide a collective and exhaustive review of the 9 major time series forecasting techniques with respect to building energy consumption.
- To perform a comparative analysis that includes both qualitative and quantitative aspects of these techniques.
- To elaborate on the various combinations of the hybrid model while assessing their performance and novelty.

1.3. Summary of papers reviewed

This section describes the structure of this paper and summarizes information on the papers that have been reviewed (Table 2). The structure of the reviews is as follows - first, an overview of each machine learning technique is provided, followed by a review of a set of studies exemplifying the application of this technique. For each paper reviewed, emphasis is given to the objective of the study, details of the time series data, and the accuracy of the model which is generally measured in Mean Absolute Percentage Error (MAPE). In the next half