A REVIEW ON DATA-DRIVEN MODELS FOR MUTI-TIMESCALE BUILDING ENERGY PREDICTION

Y.Q.Pan[1],*, M.Y.Zhu[2] and Z.Z.Huang[3]

[1]School of Mechanical Engineering, Tongji University,Shanghai 200092, China
[2]School of Mechanical Engineering,, Tongji University,Shanghai 200092, China
[3]Sino-German College of Applied Sciences,Tongji University, Shanghai 200092, China

ABSTRACT
Data-driven Models (DMs) play a significant role in the building energy prediction researches. The lack of targeted summary on DMs usually makes researchers and scientists engaged in promoting building energy performance confused of the selection and application of DMs. This paper primarily summarizes the application of DMs in the field of building energy prediction. According to the methodology of DMs development, the Regression Model (RM), Time Series Model (TSM), Genetic algorithm (GA), Artificial Neural Network (ANN)/ Support Vector Machine (SVM) and Calibrated Simulation (CS) are classified into DMs. From the view point of the model application, the function of the prediction load/energy achieved by a specific DM depends on its time scale. Under different building management goals for different levels of decision-makers and researchers, the model selection, the data requirements and applicable objects are all distinct accordingly. Different from previous reviews that mainly pay attention to the methodology and the performance comparison of building energy prediction models, we review the DMs' model development and application in the multi-timescale building energy prediction, including hourly/daily, monthly, annual and mutidecadal timescale. This paper will contribute to reasonable selection of existing DMs on the basis of the building operation stage, the available monitored data quality, the building energy prediction requirement, and the building energy efficiency goal for an object building or a building type.

KEYWORDS
Data-driven Models (DMs), building energy prediction, multiple timescales, model application

INTRODUCTION
The main two purposes of estimating building energy use are generalized by ASHRAE (2013): (1) energy assessment for building and HVAC system design and associated design optimization, and (2) analyzing energy use of existing buildings for establishing baselines, calculating retrofit savings, and implementing model predictive control. A large number of building energy prediction researches varying in magnitude from aiming at a single building to aiming at a building type have applied/proposed various kinds of building energy prediction models in recent years. By applying the science of mathematical modeling to physical systems, Rabl (1988) generalized two broad and distinct modeling approaches for building energy estimation: Forward Approach and Data-driven Approach.

* Corresponding author email: yiqunpan @tongji.edu.cn

The forward modeling approach/procedure and its application is quite unified in the related researches and a comprehensive comparison of mainstream software based on forward approach has been reviewed (Zhao et al. 2012). On the contrary, with various statistical methods and actual building performance data as the base, data-driven modeling approach contains a variety of types and has more complex application according to different data mining method or different available/required data. Here focuses on the recent work related to the Data-driven models applied to building energy prediction. This review paper is primarily targeted to summarize the DMs' model development and their feasibility in the field of muti-timescal building energy prediction for researchers and scientists engaged in promoting building energy performance.

DATA-DRIVEN MODELS (DMs) APPROACH AND DEVELOPMENT

Data-driven Modeling method

In the view of mathematical modeling, a model is a description of the behavior of a system made up of three components: input, system structure and output. For a building energy model, the input variables have two types: controllable by the experimenter (e.g., internal heat gains, thermostat settings, etc.), and uncontrollable (e.g., outdoor temperature, solar radiation, etc.). The model provides the necessary physical description of building energy system (e.g., thermal mass or mechanical properties of the elements). The outputs of the model, e.g. energy use, describe the reaction of the building energy system to the inputs variables. In this case, the procedure of a DM for building energy prediction is to determine a mathematical description of the building energy system and to estimate system parameters with known/measured input and output variables.

Classification approaches for Data-driven Models (DMs)

There mainly exist two kinds of classification for DMs. Under the first classification, DMs are divided into steady-state models and dynamic models. Steady-state models do not consider the effects such as thermal mass or capacitance that cause short-term temperature transients. Generally, these models are appropriate for monthly, weekly, or daily data and are often used for baseline model development. Dynamic models capture effects such as building warm-up or cooldown periods and peak loads, and are appropriate for building load control, FDD, and equipment control. The other classification is based on data requirements, time and effort needed to develop the DM, user skill demands, and sophistication and reliability provided. In this case, DMs can be classified into three broad categories: Black-box, Calibrated simulation, and Gray-box. With Black-box approach, a relationship between measured energy use and the various influential parameters (e.g., climatic variables, building occupancy) is built, either by purely statistical method or loosely based on some basic engineering formulation of energy use in the building. It's adequate for evaluating demand-side management programs to identify simple and conventional energy conservation measures in an actual building and for baseline model development in energy conservation measurement and verification projects. The Calibrated simulation uses an existing building simulation program and tunes or calibrates the various physical inputs to the program so that observed energy use matches closely with that predicted by the simulation program. The calibrated model has the applicability to baseline model development for M&V purposes and as a diagnostic tool for identifying potential operational problems and for estimating potential savings from optimized operating parameters. Rabl (1992) defined that Gray-box approach first formulates a physical model to represent the structure or physical configuration of the building or HVAC&R equipment or system, and then identifies important parameters representative of certain key and aggregated physical parameters and characteristics by statistical analysis. Gray-box model has great potential, especially for fault detection and diagnosis (FDD) and online control.

As shown in Table 1, the first two different classification of existing several DMs by ASHRAE (2013) overlaps each other. Recent review works basically follow the second one, adopting the Black-box/Statistical & Gray-box idea in the Category 3 (Zhao 2012) and 4 (Foucquier 2013). Under the later two classifications, some models in the former two classifications are excluded from the common DMs, e.g., Polynomial fitting, Bin, Differential equation, and Modal analysis models. Category 3 takes the ANN and SVM as the machine learning model category. The later two classifications exclude the Calibrated simulation from DMs, especially in the Category 3 taking it as Engineering/Forward model. Actually, both the procedure and the result of Calibrated simulation fit the development of a DM. The achieved model by calibrated simulation is an accurate model built by a mature building energy simulation tool, having a more complex format than other DMs. Besides, Category 4 classifies the Genetic algorithm (GA) as the Statistical model. Properly speaking, GA is a stochastic optimization method in statistics, deduced from an analogy with the evolution theory of Darwin. In the development of DMs for building energy prediction, GA is often used as a parameter estimation method for selected model structure and model variables by researchers to determine the model parameters of each input variables (Hossein et al. 2011). Some researches regard GA as one of machine learning techniques (Fan et al. 2014). As a result, we classify the Calibrated simulation and the optimization method, GA into the Data-driven Methods.

Table 1 *Existing different classifications of DMs*

Modeling method	Category 1	Category 2	Category 3	Category 4
Regression Model	Steady-state	Black-box	Statistical	Statistical
Polynomial fitting	Steady-state	Gray-box	-	-
Bin method	Steady-state	Black-box	-	-
ANN/SVM	Dynamic	Black-box	Machine learning	Statistical
Thermal network	Dynamic	Gray-box	Gray-box	Gray-box
Time Series Model	Dynamic	Gray-box	Statistical	-
Differential equation	Dynamic	Gray-box	-	-
Modal analysis	Dynamic	Gray-box	-	-
Calibrated simulation	Dynamic	Calibrated Simulation	Engineering	-
Genetic algorithm	-	-	-	Statistical

According to the methodology of DMs development, we classify the following several DMs: Regression Model (RM), Time Series Model (TSM), Genetic algorithm (GA), Artificial Neural Network (ANN)/ Support Vector Machine (SVM) and Calibrated Simulation (CS) into the Data-driven Methods/Models, with consideration of the similarity and difference of existing classifications of recently common DMs. The characters and performance of the mentioned DMs are summarized and compared, shown in Table 2, based on a hard work of reviewing related researches.

***Table 2** Performance comparison of DMs*

DMs	Complexity	Difficulty of Using	Calculation Speed	Data Requirements	Accuracy
RM	low	low	very quick	historical data	medium
TSM	medium	low	very quick	historical data	medium
GA	high	high	slow	historical data	high
ANN	high	high	quick	historical data	high
SVM	very high	high	slow	historical data	very high
CS	high	medium	quick	historical data and building information	high

LITERATURE REVIEW ON APPLICATION OF DMs FOR MUTI-TIMESCALE BUILDING ENERGY PREDICTION

Previous review researches mainly pay attention to the modeling methodology and the performance comparison of building energy prediction models (Foucqier et al. 2013, ASHRAE, 2013, Zhao et al. 2012). The lack of targeted summary on DMs usually makes researchers and scientists engaged in promoting building energy performance confused of the selection and application of DMs. From the view point of the application in the building management, the function of the prediction load/energy achieved by a specific DM depends on its time scale. Hahn (2009) broadly classified the existing research on building energy prediction into three timescales: short-term (up to one week ahead), medium-term (from one week to one year ahead) and long-term (longer than one year ahead) predictions. Under different building management goals for different levels of decision-makers and researchers, the model selection, the data requirements and applicable objects are all distinct accordingly. This section reviewed the DMs' model development and application in the multi-timescale building energy prediction for different building management goals.

DMs' application in the Hourly/Daily energy predication

The short-term (hourly/daily) building load/energy predictions have a significant role in optimizing the building energy use and achieving an energy efficient operation. The hourly building energy prediction enables an optimization of control sequences and operations of building systems by identifying periods of excessive consumption and proposing improved operation strategies (Platon et al. 2015). Currently, the daily prediction mainly focuses on the predictions of the daily peak demand, daily energy consumption and daily load profiles for secure and profitable operation of modern power utilities (Fan et al. 2014) and convenient building maintenance management. Radu et al. (2015) developed the predictive models for hourly building electricity consumption of a Canadian office building by means of two artificial intelligence techniques, ANN and case-based reasoning (CBR). By Principal Component Analysis (PCA) procedure, significant input variables were selected, including outside air temperature and humidity, AHU supply air temperatures, chiller/ boiler water outlet temperature/flow rate. The hourly measured data covers the period of March 2013 – May 2014. The results revealed that the ANN models consistently outperform the CBR models. Cheng et al. (2014) developed ensemble models for predicting next-day energy consumption and peak power demand of a high-rising commercial building in Hong Kong. After the abnormal building energy consumption profile and the data-driven input selection work, eight base DMs are developed using RM, SVM, TSM, etc. The final predictive ensemble models are constructed by combining base models by application of GA to optimize the weights of eight base models. The modeling inputs contain the building power consumption, collected from the building energy management system at 15-min intervals, and daily meteorological data for one year. By comparison, the accuracies of the ensemble models are better than those of the individual base models, and the SVM models result in the best performance among the individual base models.

DMs' application in the Monthly energy prediction

The monthly building energy prediction has close linkage to the building energy cost budget, which is crucial for building owners and managers' decision making of optimizing allocation of energy resources, protecting future energy supply/demand and promoting efficiency and conservation efforts. Williams et al.(2016) applied RM to predict future monthly energy consumption for residential buildings in USA using building characteristics and monthly climate data. In the model training datasets, the building energy data is collected from over 426,305 local homes with four years (2010-2013) of monthly energy consumption as the response variable. The considered factors contain 17 building characteristics, including building size, fuel type, built year, construction type, etc., and 2 climate variables, including average monthly temperature and humidity. With the large scale measured datasets, regression models are developed and compared given influence rank of building characteristics and uncertainty in climate forecasts. This modeling route is similar to the second research orientation for a building type in the following section 3.3. Zhou et al.(2013) proposed a monthly energy consumption prediction method for commercial buildings by adopting TSM combined with RM. The conventional TSM mostly comprise a trend term (Ht) and a periodic term (Pt). In the application of the proposed method, the historical monthly energy consumption data of the previous three years (Jan., 2007-Dec., 2009) for two commercial buildings in Shanghai is regarded as an original sequence. After the seasonal adjustment and pre-process, the Ht and Pt term of future monthly building energy model is established. Different from traditional TSM, four impact factors at monthly scale, accumulated temperature, relative humidity, workdays, and non-workdays, are added to the above two terms in the form of a regression term to achieve the final predictive TSM.

DMs' application in the Annual energy predication

In the building energy prediction field of multiple timescales, the annual energy prediction attracts the most researchers with the aim of building performance evaluation, Energy Conservation Measures (ECMs) investigation, energy saving calculation and verification. The objects in the related researches vary from a specific building to a building type, in which office and commercial building are in the majority. For a specific building, the calibrated simulation is the representative DMs applied for whole building energy performance analysis. Raftery et al. (2011) detailed a calibration of a whole building energy model to hourly measured energy consumption data using a systematic, evidence-based methodology. The case study building was a 30,000 m^2, four-floor office building located on Intel's campus in Ireland. The process of calibration simulation has high requirements for both the object building information and the measured data, including as-built drawings, details on materials and constructions, design information and operation & maintenance manuals for major HVAC equipments, and electrical power consumption of major electrical panel or motor control center. The measured data-points are logged at 1-min intervals for weather data and 15-min intervals for energy data for 5 years. The final calibrated model is used to investigate 12 selected ECMs for feasibility on the basis of the annual building energy analysis and energy saving calculation.

Another related research orientation targets the quick building energy performance assessment for a type of buildings. Statistical models (e.g., RM and ANN) are usually applied to a building performance database (BPD) (Lee et al. 2015) including large scale energy factors and the corresponding energy use data. Currently, there are two technical routes for the BPD establishment. One is achieved by collecting empirical data on actual building energy performance, physical and operational characteristics of the object type of buildings, such as the 2003 CBECS (Commercial Building Energy Consumption Survey) database. Aranda et al. (2012) implemented multiple linear RM to predict the annual energy consumption in the banking sector in Spain and to suggest energy saving strategies to increase the energy efficiency, using survey data of 55 local banks from the 12 total climatic areas across Spain. For the analysis, a two-phase energy audit was carried out from June 2010 to May 2011to obtain the 11 variables, including annual energy consumption, climatic condition, office area, number of employees, HVAC installed power, etc. The achieved regression model is appropriate for predicting the energy consumption of bank branches with good energy consumption performance and detecting inefficiencies in bank branches with poor energy consumption performance. The other route applies forward models to conduct large-scale building energy simulations. The inputs (building variables) and outputs (building energy consumption) of these massive calculation constitute the training dataset for the mentioned statistical models. Shideh et al. (2015) combined DOE-2 with a randomized approach to generate a commercial building performance dataset having 30,000 combinations of building variables and the corresponding energy use, with consideration of 17 key building design variables related to building envelope, building orientation, occupant schedule for different climate zones in the U.S. Multiple regression models for office building energy prediction is developed after the stepwise regression and sensitivity analysis for the regressor determination. Without the limitation of the practical measured data requirement, the developing models by this route can be used to estimate building energy consumption in early stages of design. Given that building energy consumption depends on many operational and design parameters, large numbers of simulations and the corresponding massive computation time are needed to cover the whole design space. To tackle this problem, some researches choose the sampling methods (Amiri et al. 2015), such as Monte Carlo method, for reducing the simulation times.

DMs' application in the Multidecadal energy prediction

Under the global concern on climate change, the long-term prediction of building energy demand under climate change is greatly significant to building energy saving, urban energy planning, urban strategy development and energy policy formulation for effective response to climate change. Zhu et al. (2016) classified the current prediction researches of multidecadal building energy demand into two group: (1) direct prediction through statistical model, and (2) indirect prediction using calibrated building model. For the direct prediction, a statistical method (mainly Regression) (Wan et al. 2011, Braun et al. 2014) is usually applied to achieve the relationship between the local meteorological parameters (degree day, dry-bulb/wet-bulb temperature) and the specific building energy demand based on the sufficient historical observed data. Wan et al. (2011) developed several RMs to correlate monthly building cooling loads and the total energy use of office building in five representative cities in China, Harbin, Beijing, Shanghai, Kunming and Hong Kong, with a new climatic index Z (a function of dry-bulb/wet-bulb temperature and global solar radiation in 1971-2008). Future trends of building heating/cooling loads and energy demand in the 21st century are predicted by the RMs based on IPCC's (Intergovernmental Panel on Climate Change) climate predictions (IPCC 2007). The indirect prediction usually needs two main parts as the research foundation: (1) Hourly future weather data. As the indispensable inputs of building simulation, the future weather data is mainly achieved based on IPCC's prediction in the recent researches by means of statistical methods, such as Morphing, TSM (Zhu et al. 2016, Chan 2011). (2) Calibrated prototypical building models. After calibrated by investigated actual energy use data, the models can mainly reflect the general situation of different building types in a specific region. In this way, Chan (2011) developed future TMY files for Hong Kong and used them in local prototypical buildings to predict office and residential building energy demand for three future periods (2011–2030, 2046–2065 and 2080–2099).

DISCUSSION

According to the methodology of DMs development, we classify the RM, TSM, GA, ANN/ SVM and CS into DMs. Each model has its own advantages and disadvantages. It is difficult to say which one is better without complete consideration of certain cases of applications. Actually the feasibility of a specific DM and the corresponding required data quality of the DM development mainly depends on its time scale of prediction results needed for different goals. In the short-term (hourly and daily) building energy prediction for optimal and secure operation of a specific building energy system, machine learning (GA, ANN/SVM), RM, and TSM are usually used. RM, TSM and CS are often applied in the medium-term (monthly and annual) prediction to assess energy cost budget and alternative ECMs for a specific building. In the researches on quick building performance evaluation and energy saving strategy suggestion for a type of building, the RM and ANN are often combined with a building performance dataset as the training data, which is built by practical survey or forward simulation. For response to climate change, RM and CS are often combined with weather data prediction study to evaluate the long-term (mutidecadal) energy demand trend of a building sector.

CONCLUSION

This paper reviews the recent work on the application of Data-driven Models for muti-timescale building energy prediction. This paper focuses related work using the following several DMs: RM, TSM, GA, ANN/ SVM and CS. Different from previous reviews that mainly pay attention to the methodology and the performance comparison of general building energy prediction models, we review the DMs' model development and application in the building energy prediction from the view point of different functions realized by multi-timescale building energy prediction results. The model selection, the data requirements and applicable objects are all distinct accordingly for different building management goals and different levels of decision-makers. This paper will contribute to reasonable selection of existing DMs on the basis of the building operation stage, the available monitored data quality, the building energy prediction requirement, and the building energy efficiency goal for an object building or a building type.

ACKNOWLEDGEMENTS

The authors would like to thank the support of China Mobile Communications Corporation, who funded this research. This work is part of the Ministry of Education (MOE) – China Mobile Research Foundation 2014 Project No. 5-2: Energy simulation model development and validation for China Mobile office buildings.

REFERENCES

Amiri S S, Mottahedi M, Asadi S. 2015. Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S.[J]. Energy and Buildings, 109: 209-216.

Aranda A, Ferreira G, Mainar-Toledo M D, Scarpellini S, Llera Sastresa E. 2012. Multiple regression models to predict the annual energy consumption in the Spanish banking sector[J]. Energy and Buildings, 49: 380-387.

Ashrae. 2013. ASHRAE Handbook of Fundamentals[M]. American Society of Heating, Refrigerating, and Air Conditioning Engineering, Inc..

Braun M R, Altan H, Beck S B M. 2014. Using regression analysis to predict the future energy consumption of a supermarket in the UK[J]. Applied Energy, 130: 305-313.

Chan A L S. 2011. Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong[J]. Energy and Buildings, 43(10): 2860-2868.

Fan C, Xiao F, Wang S. 2014. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques[J]. Applied Energy, 127: 1-10.

Foucquier A, Robert S, Suard F, Stéphan L, Jay A. 2013. State of the art in building modelling and energy performances prediction: A review[J]. Renewable and Sustainable Energy Reviews, 23: 272-288.

Hahn H, Meyer-Nieberg S, Pickl S. 2009. Electric load forecasting methods: Tools for decision making[J]. European Journal of Operational Research, 199(3): 902-907.

Hossein S, Mahdi Z, Mohamad H. 2011. Estimation of electricity demand in residential sector using genetic algorithm approach[J]. International Journal of Industrial Engineering and Production Research, 22(22): 43-50.

Ipcc. 2007. Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [M]. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Lee S H, Hong T, Piette M A, Taylor-Lange S C. 2015. Energy retrofit analysis toolkits for commercial buildings: A review[J]. Energy, 89: 1087-1100.

Platon R, Dehkordi V R, Martel J. 2015. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis[J]. Energy and Buildings, 92: 10-18.

Rabl A. 1988. Parameter estimation in buildings: Methods for dynamic analysis of measured energy use[J]. Journal of Solar Energy Engineering, 110(110): 52-66.

Rabl A, Rialhe A. 1992. Energy signature models for commercial buildings: test with measured data and interpretation[J]. Energy and Buildings, 19(2): 143-154.

Raftery P, Keane M, Costa A. 2011. Calibrating whole building energy models: Detailed case study using hourly measured data[J]. Energy and Buildings, 43(12): 3666-3679.

Wan K K W, Li D H W, Liu D, Lam J C. 2011. Future trends of building heating and cooling loads and energy consumption in different climates[J]. Building and Environment, 46(1): 223-234.

Williams K T, Gomez J D. 2016. Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach[J]. Energy and Buildings, 128: 1-11.

Zhao H, Magoulès F. 2012. A review on the prediction of building energy consumption[J]. Renewable and Sustainable Energy Reviews, 16(6): 3586-3592.

Zhou R, Pan Y, Huang Z, Wang Q. 2013. Building Energy Use Prediction Using Time Series Analysis[J]. 2013 IEEE SIXTH INTERNATIONAL CONFERENCE ON SERVICE-ORIENTED COMPUTING AND APPLICATIONS (SOCA): 309-313.

Zhu M, Pan Y, Huang Z, Xu P. 2016. An alternative method to predict future weather data for building energy demand simulation under global climate change[J]. Energy and Buildings, 113: 74-86.