Applied Energy 144 (2015) 261–275

Contents lists available at ScienceDirect

Applied Energy

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

# Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach

Xiaoshu Lü [a,*], Tao Lu [a], Charles J. Kibert [b], Martti Viljanen [a]

[a] Department of Civil and Structural Engineering, School of Engineering, Aalto University, P.O. Box 12100, FIN-02015 Espoo, Finland
[b] Powell Center for Construction & Environment, University of Florida, P.O. Box 115703, Gainesville, FL 32611-5703, USA

## HIGHLIGHTS

- This paper presents a new modeling method to forecast energy demands.
- The model is based on physical–statistical approach to improving forecast accuracy.
- A new method is proposed to address the heterogeneity challenge.
- Comparison with measurements shows accurate forecasts of the model.
- The first physical–statistical/heterogeneous building energy modeling approach is proposed and validated.

ARTICLE INFO

*Article history:*
Received 12 May 2014
Received in revised form 3 December 2014
Accepted 6 December 2014
Available online 25 February 2015

*Keywords:*
Modeling method
Energy consumption
Physical–statistical approach
Heterogeneous buildings

ABSTRACT

Energy consumption forecasting is a critical and necessary input to planning and controlling energy usage in the building sector which accounts for 40% of the world's energy use and the world's greatest fraction of greenhouse gas emissions. However, due to the diversity and complexity of buildings as well as the random nature of weather conditions, energy consumption and loads are stochastic and difficult to predict. This paper presents a new methodology for energy demand forecasting that addresses the heterogeneity challenges in energy modeling of buildings. The new method is based on a physical–statistical approach designed to account for building heterogeneity to improve forecast accuracy. The physical model provides a theoretical input to characterize the underlying physical mechanism of energy flows. Then stochastic parameters are introduced into the physical model and the statistical time series model is formulated to reflect model uncertainties and individual heterogeneity in buildings. A new method of model generalization based on a convex hull technique is further derived to parameterize the individual-level model parameters for consistent model coefficients while maintaining satisfactory modeling accuracy for heterogeneous buildings. The proposed method and its validation are presented in detail for four different sports buildings with field measurements. The results show that the proposed methodology and model can provide a considerable improvement in forecasting accuracy.



## 1. Introduction

The demand for reliable building energy forecasting models is rapidly increasing because energy and environmental issues related to the building sector have become a prominent concern of society. In addition, the ever-growing dynamics and complexity of current and future building systems, e.g. the integration of information and communication technologies into the heating, ventilation and air-conditioning (HVAC) systems, has made building energy modeling more sophisticated and forecasting more difficult. However, building energy forecasting models have always been useful in helping predict, transmit, distribute and plan building energy use to deal with capacity needs and to ensure energy efficiency and demand response. And models are especially needed when tackling difficult and complex building systems because they can improve and deepen our understanding of these systems. Therefore, modeling will continue to encompass a large and important literature in building energy research.

Although a variety of models ranging from simple to complex exist in the literature, two general types of modeling methods are widely used: physical and statistical plus their combination. The physical models consist of partial differential equations of physical laws that govern energy flows in buildings. For example,

* Corresponding author.
*E-mail address:* xiaoshu.lu@aalto.fi (X. Lü).

http://dx.doi.org/10.1016/j.apenergy.2014.12.019
0306-2619/

**Nomenclature**

| | |
|---|---|
| $a$ | model parameter referred to Eq. (16) |
| $A$ | surface area ($m^2$) |
| $b$ | model parameter referred to Eq. (16) |
| $B$ | backward shift operator referred to Eq. (16) |
| $C^{air}$ | heat capacity of indoor air ($J\,m^{-3}\,K^{-1}$) (product of density and specific heat) |
| $d$, $D$ | diagonal elements and matrix referred to Eqs. (22) and (23) |
| $E$ | expectation operator referred to Eq. (31) |
| $F$ | thermal transmission defined in Eq. (12) |
| $\dot{G}$ | heat generation rate of the occupancy (W) |
| $h$ | convective heat transfer coefficient ($W\,m^{-2}\,K^{-1}$) |
| $I$ | solar radiation ($W\,m^{-2}$) |
| $i$, $j$ | indices |
| $k$ | conductivity ($W\,m^{-1}\,K^{-1}$) |
| $l$, $L$ | layer thickness (m) |
| $l$ | time ahead forecasting referred to Eq. (31) |
| $n$ | layer number |
| $n^{air}$ | air change rate ($h^{-1}$) |
| $n^{occupancy}$ | number of occupants |
| $p$, $q$ | model parameter referred to Eq. (16) |
| $r$ | rank of approximation referred to Eq. (23) |
| $R$ | error between predicted and measured |
| $\dot{Q}$ | heat flow ($W\,m^{-3}$) |
| $t$ | time (s) |
| $T$ | temperature (°C/K) |
| $u$, $U$ | vector and matrix referred to Eqs. (22) and (23) |
| $v$, $V$ | vector and matrix referred to Eqs. (22) and (23) |
| $V$ | volume ($m^3$) |
| $w$ | weights referred to Eqs. (21) and (27) |
| $x$ | space coordinate |
| $x$, $y$ | variables or time series |

*Greek symbols*

| | |
|---|---|
| $\alpha$ | thermal diffusivity ($m^2\,s^{-1}$) |
| $\alpha^{sol}$ | fraction of the surface for the incident solar radiation |
| $\beta$ | model parameters referred to Eq. (14) and Eq. (25) |
| $\varepsilon$ | error term ($\sim N(0,\sigma)$ or zero mean and variance $\sigma$) |
| $\gamma$ | coefficient referred to Eq. (26) |
| $\sigma^{sol}$ | transmissivity |

*Superscripts and subscripts*

| | |
|---|---|
| $t$ | time of the observed datum referred to time series |
| + | indoor |
| − | outdoor |

*Abbreviations*

| | |
|---|---|
| AIC | Akaike information criterion |
| ANN | artificial neural networks |
| AR | autoregressive |
| ARIMA | autoregressive-integrated-moving-average |
| HVAC | heating, ventilation and air-conditioning |
| RBF | radial basis function |
| RMSE | root mean squared error |
| $R^2$ | coefficient of determination |
| SRM | structural risk minimization |
| SVD | singular value decomposition |
| SVM | support vector machine |
| VFD | variable frequency drive |
| $CO_2$ | carbon dioxide |
| ACF | autocorrelation function |
| PACF | partial autocorrelation function |

Ascione et al. [1] applied a physical model to predict cooling energy savings with reference to a well-insulated massive building to investigate the effect of phase change materials on the exterior building envelope during the cooling season [1]. Physical models can provide valuable insight into general physical mechanism and potential knowledge, but are often limited to simple systems. Buildings, especially large or complex ones, are inherently complex and nonlinear because of the multiple interconnections among their diverse energy systems. Simplifications of the model equations and lack of knowledge of the physical mechanisms underlying complex systems may lead to a lack of precision or incorrect results. In particular, models often approximate individual buildings which are not representative enough. As a result, many physical models are only applicable to micro-scale validations which can hardly be generalized. In contrast to physical models, statistical models are constructed based on experimental data for flexibly coping with the various complexities. As examples, statistical machine learning models have been applied to forecast energy consumption and loads of residential buildings [2,3]. Alan et al. [4] employed a statistical model for load forecasting in an application of long-term electric power transmission planning [4]. The major limitation of the statistical models is that they provide much less physical understanding and, thus, substantial amounts of data are needed for training the models due to the unknown underlying mechanisms. Model output is estimated directly from input–output data without model structures and as a result the models suffer the same type of accuracy and application limitations due to the possible non-representative data.

In a typical application, many models fall between these two extremes. Neto and Fiorelli [5] compared physically-based EnergyPlus [6] and statistically-based artificial neural network (ANN) models in forecasting building energy consumption. They found that even though the differences in accuracy were small, with the ANN model providing a slightly better prediction, occupant behavior and weather changes could significantly affect the energy consumption profile. This made forecasting more difficult or inaccurate for both models for air conditioned buildings. Xu et al. [7] extended EnergyPlus with ANNs for a holistic energy consumption model at the inter-building level which considered the influence of residents and the neighborhood context. Lee and Tong [8] combined a dynamic model with genetic programming to improve the forecasts of energy consumption for the traditional statistical approach. Focusing on the impact of different shading devises and building envelopes characteristics on the demand for air-conditioning public buildings, Ouedraogo et al. [9] adopted simplified physical energy-balance and empirical models to investigate weather trends and how the trends affected public buildings' cooling loads for 2010–2080 [9]. Using a similar forecasting approach, Gouveia et al. [10] performed energy end-use demand prediction in residential buildings for 2050 and aimed to identify the parameters governing energy services demand uncertainty [10]. A similar approach was also employed by Kwak et al. [11] to predict short-term and real-time energy demand for the effective operation and management of buildings using case studies. Roldan-Blay et al. [12] adopted a more complicated model of ANNs for predicting short-term building energy consumption. The input variables were reduced to physical variables to avoid too much variability in the model. Mavromatidis et al. [13] combined physical-thermal and fractional factorial simulation methods to obtain regression models in the form of polynomial functions of building envelope's physical characteristics for dynamic thermal performance forecasting [13].