Available online at www.sciencedirect.com
ScienceDirect
Energy Procedia 78 (2015) 3378 – 3383
Energy Procedia

6th International Building Physics Conference, IBPC 2015

Data driven approaches for prediction of building energy consumption at urban level

Giovanni Tardioli[a,b]*, Ruth Kerrigan[b], Mike Oates[b], James O`Donnell[a], Donal Finn[a]

[a:] School of Mechanical & Materials Engineering, University College of Dublin, Dublin, Ireland
[b:] Integrated Environmental Solutions (IES) R&D, Glasgow, UK

Abstract

The ability to predict building energy consumption in an urban environment context, using a variety of performance metrics for different building categories and granularities, across varying geographic scales, is critical for future energy scenario planning. The increased quantity and quality of data collected across urban districts facilitates the utilization of data-driven approaches, thereby realizing the potential for energy prediction as a complementary or alternative option to the more traditional physics based approaches. The majority of research to date that exploits data-driven approaches, has mainly focused on analysis at an individual building level. There are few examples in the literature of studies that utilize data-driven models for building energy prediction at an urban scale. The current paper provides a literature review of the recent applications of data-driven models at an urban scale, underlining the opportunities for further research in this context.

Peer-review under responsibility of the CENTRO CONGRESSI INTERNAZIONALE SRL

Keywords: large scale data-driven models; buildings clustering; building energy consumption estimation; energy mapping

1 Introduction

1.1 Background and context.

The issues surrounding building energy consumption in the urban context has grown steadily in the last few decades due to growth in the urban population, improvement of building services, associated comfort levels and the

* Corresponding author. Tel.: 00353 8311889645
E-mail address: giovanni.tardioli@ucdconnect.ie

Peer-review under responsibility of the CENTRO CONGRESSI INTERNAZIONALE SRL
doi:10.1016/j.egypro.2015.11.754

increased time occupants spend inside buildings [1]. Reports on the energy end-use consumption in the EU 27 for 2011 show that residential and commercial buildings account for 40% of total energy consumption of the built environment and about 14% of greenhouse gas emissions [2]. Buildings, as a large part of the overall energy system, represent one of the major contributors to energy inefficiency. Given the proposed ambitious EU, national and local energy goals, there is an increasing requirement for fast, reliable and accessible simulation tools. Such tools need to provide decision makers with information regarding individual and collective building energy and pollution footprints at different levels of granularity in urban areas [3]. The construction sector has the largest potential for delivering long term significant and cost effective greenhouse gas emission reductions, through both retrofitting of the existing building stock and the implementation of new innovative energy efficient strategies for new buildings [4]. However, to make this change possible, major efforts are required to go beyond existing technical and economic barriers. These include the ability to represent the energy consumption of the entire building portfolio, while at the same time identifying energy reduction opportunities at the building level, thereby improving urban planning strategies for the entire building portfolio life-cycle (feasibility, design and operational). As such, the reduction of cost barriers for energy efficient solutions and the improvement of reliable indicators to measure building energy performance at a city scale are important contributions for achieving urban sustainability targets [5]. By creating solutions that start at the city scale, but address both districts and buildings, the potential for the uptake of synergistic and scalable solutions is significant.

1.2 Motivation and problem identification.

The advantages of performing large scale energy prediction through simulation are many, for example, the identification of (i) energy resources (e.g., waste power or heat) in districts of a city or in different buildings of the same district [6], (ii) energy outliers [7], (iii) demand side management operations and local balancing [8], (iv) candidates for retrofit intervention [9], (v) large benchmarking analysis engaging whole communities [10, 11], (vi) peak power demand [12] and (vi) better urban planning in a specified area. To analyze energy at an urban scale, the available information and the level of granularity of the data must be evaluated. The quantity of data which is possible to collect from single buildings has increased in the last few years due to the penetration of smart metering, greater access and understanding of utility data and the implementation of Building Management Systems (BMS). However, even if data is theoretically available for analysis, privacy and protection policies may exclude them as sources of information. Aggregation or anonymization techniques are therefore required, sometimes compromising the quality of the datasets [13]. Further to this, the evaluation of large scale building energy consumption can be extremely time-consuming if performed with single building simulation approaches, due to data gathering processes, simulation and monitoring techniques and estimation of uncertainties [14]. In this context, the identification of new methods to collect and use real building data in a time efficient manner while maintaining a high level of granularity that does not compromise the final outcome is a significant contribution for tools that will aid decision support.

1.3 Current paper aim.

Using existing approaches for the modelling of single buildings as a reference baseline [14], it is possible to categorize building energy simulation methods into three high-level categories as follows: (i) physics-based approaches, (ii) data-driven approaches and (iii) hybrid approaches. At a building level, these categories are often classified as white, black-box or grey-box approaches, respectively [15, 16]. The current paper is concerned with a review of the state of the art of data driven approaches applied to building energy analysis at an urban scale. The paper is motivated by the need to identify the research opportunities and points of departure for urban level analysis of building energy performance using data-driven approaches. Cognisance will be taken of variety of issues including: varying building types and end-use, different energy performance metrics, multiple levels of granularity and urban scales.

2 Large scale buildings energy prediction methods

2.1 White-box based approaches.

The development of white-box models for an entire urban building stock would require a considerable amount of time for the simulation of each individual building and the collection of detailed information required to ensure that the simulated Building Energy Models (BEM) are of sufficient accuracy. The urban energy building portfolio is described as a white-box based approach for the simulation of representative buildings, called archetypes; these are developed after an accurate identification of the most common characteristics of different groups of similar buildings [15-17]. This method analyses a smaller part of the building stock, producing BEMs in detail, which still retain the ability to adequately characterize the energy performance of the entire building portfolio. Furthermore, the classification of the building stock with representative buildings, allows for the creation of accurate benchmarking models at a local level. Another important advantage of this approach, is the possibility to assess the potential of deep retrofit Energy Conservation Measures (ECM's) and corresponding what if scenarios for the entire city [16, 18]. This is achieved through total energy results aggregation. Enhancing the building stock classification makes it then possible to provide approximate information for the single building, boosting accurate energy mapping and profiling possibilities at individual building level also. Nevertheless, such kind of representation requires dedicated simulation engines and large amounts of data to correctly represent the building population of a given city.

2.2 Grey-box based approaches.

Grey-box based approaches [19-22] combine prior physical knowledge with information from data sources. Usually grey-box models have a hybrid structure combining first principle physics and data driven approaches. They have some advantages but also limitations of the white and black-box models. In the majority of large scale models the building stock is represented based on analogy with an electrical circuit, where a reduced order resistance- capacitance (RC) circuit is able to describe the energy behaviour of the building [23].

2.3 Black-box based approaches.

Large scale black-box based models use building level data-analysis and data-mining tools to predict and forecast energy consumption at a larger scale than the single building level. While black-box models are widely used for prediction and forecasting of energy use, based on the selection of hierarchically important inputs [24-26], there are fewer examples of the application these approaches at large scale [27, 28]. The most popular black-box approaches for prediction and forecasting at building level are [29]: simple regression model (SRM), multiple linear regression (MLR), decision trees (DT), artificial neural networks (ANN) and support vector machine (SVM). Data driven models rely on the availability of prior data to forecast energy behaviour. For this reason it is important to establish a database to train the models, however data privacy policy and economic interests make the data collection process difficult, often compromising the quality of the final results. Geographic information system (GIS) is increasingly an important resource to develop large scale building energy models, since they give the possibility to allocate and visualize contextual building energy information in a user-friendly mode [14, 30-35], however very few city GIS databases contain the information relevant for understanding the energy performance of a city. Other relevant datasets include: census [14], national resources [14, 31], normative [14, 31], national and local surveys [14, 30-35], questionnaires [30] and meteorological data. Recently, other innovative sources of information such as crowd data sourcing techniques have been used to develop and populate entire city database [36]. Depending on the energy estimation methodology, different information needs to be collected, the most relevant being: construction period, use, area, shape, perimeter, height, form factor (surface/volume), windows area, surface/glazing factor, envelope materials, number of floors, orientation, solar shading, typical efficiency system plants, energy consumption data at building level or aggregated level, energy bills information and meteorological data.

3 Data driven methods: large scale techniques

Black-box based models have been used to address different large scale building energy related problems, e.g. mapping urban energy consumption, forecasting and prediction, benchmarking, profiling, and initial clustering analysis for selection of parameters to develop archetypes. Table 1 summarizes the literature review for this topic and related techniques. To achieve forecasting capabilities, data-driven models must be trained on large detailed (hourly or sub-hourly readings) dataset, collected by smart metering systems and/or BMS. At building level when smart metering systems and BMS data are available, black-box models achieve a high level of prediction [37, 38]. At large scale, studies are typically carried out on building groups, university campuses or entire districts [39] and the training procedure is accomplished on aggregated measured data [40] which is often deficient due to the high costs of metering. This limits the scope of a data-driven framework for energy consumption prediction and forecasting.

Table 1. Literature review of data driven models for large scale building energy-related applications

Data-driven model	Forecasting prediction	Benchmarking	Energy mapping	Profiling
Statistical and regression based	[41]	[14, 15]	[7, 14, 33, 42]	
ANN	[27, 41]	[43, 44]		
SVM	[24, 27, 41]	[45]		
Clustering based	[46, 47]	[15, 48-51]	[34, 52, 53]	[25, 54-56]

There have been several attempts to use available urban data to classify energy performance for the entire building stock or part of it [14, 57, 58]. Benchmark models have been developed following the three main approaches, i.e., black box, grey-box and white-box methods [59, 60]. The choice of method is dictated by operational requirements, available inputs, available monitoring data, and modeller experience. Among data driven models used for benchmarking, it is possible to find: multiple linear regression, support vector regression, gaussian process regression, artificial neural networks and decision trees. Neural networks show good results for benchmarking purposes with high prediction capability after the selection of important model input parameters [43, 44]. Energy mapping methods, usually multi-layered GIS-based models, consider data-driven techniques both for pre- and post-processing operations. Among large scale techniques to estimate building energy consumption are: modified degrees day method [30, 32, 33], archetype based simulation [61] and multiple linear regression [7, 33].
Building energy profiling is concerned with the accurate characterization of energy consumption of a building over a period of time. The provision of information regarding the energy consumption profile is important for demand side management, energy resources estimations, renewables energy integration, energy saving scenarios and accurate energy mapping. Assimilating profile information is not an easy task at building level, due to uncertainties connected with unpredictable events (e.g. occupant behaviour, weather) and difficulties in gathering exact information connected with the building. There are few examples in the literature of profiling techniques at large scale [54] and they are usually developed by mixing data-driven approaches and detailed building simulation [62].
Clustering is an unsupervised data-analysis technique with the aim to discover hidden information structures in an unlabelled set of data [54]. Clustering algorithms divide a group of objects into sub-groups, where every element in a given group is similar to another in the same cluster but different from the elements in the other clusters [63]. Clustering techniques have been used as starting points to create benchmarking models to identify common features of representative buildings for use as a comparison baseline [50, 51]. Clustering of the building sector has been widely used in the literature to perform a first step in finding representative buildings (centroids) and develop archetypes [48]. Although clustering is very well documented, it is mostly used to perform classification on a specific category of buildings [49, 50, 64] and there are few studies conducted at urban level [65, 66]. Most common clustering algorithms include: k-means, model based clustering, hierarchical agglomerative clustering and k- medoids.

4 Conclusions

Several efforts have been made to develop city scale building energy consumption models at different level of granularity and functionality. A range of different techniques have been developed to this end. Data-driven models offer a good balance between reducing the time taken to create the model while maintaining an adequate level of accuracy [37, 38]. The described techniques are widely used at building level but there are fewer examples of their application at large scale. The review process, summarized in Table1, underlined the possible level of granularity achievable with data driven approaches at urban level which is directly connected with data availability. It has emerged that combinations of different techniques are employed at large scale for predicting the energy consumption and to overcome the lack of available data. The review process showed that some well-known data driven models such as ANN and SVM are still not employed at large scale to provide detailed information on the building stock (building profiling). This is reasonable considering the high level of detail of single building data which is required to develop these models. A city level data driven framework requires a large penetration of metering systems and possibilities to explore private data of the entire building stock; these conditions are still not easily accessible. Future research works on possibilities of using large scale data driven models is required in order to answer questions such as: (i) what is the achievable level of accuracy that data-driven based approaches can produce when widespread penetration of metering systems is absent?, (ii) how is it possible to provide detailed information (building energy profiling) for buildings characterized by different levels of available data?, and (iii) which combined techniques need to be taken into account to achieve the desired level of granularity?

5 Acknowledgement

The authors gratefully acknowledge the European Commission for providing financial support during conduct of research under FP7-PEOPLE-2013 Marie Curie Initial Training Network "CI-NERGY" project with Grant Agreement Number 606851.

6 References

[1] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. *Energ Buildings*. 2008;40(3):394-8.
[2] Eurostat: Energy, transport and environment indicators. 2013.
[3] Luıs de Sousa CE, Leopold U, Baume O, Braun C. iGUESS - A web based system integrating Urban Energy Planning and Assessment Modelling for multi-scale spatial decision making. iEMSs Leipzig, Germany.
[4] Buildings and Climate Change Summary for Decision-Makers. Sustainable Buildings & Climate Initiative, United Nation Environment Programme 2009.
[5] Chung W, Hui YV, Lam YM. Benchmarking the energy efficiency of commercial buildings. *Appl Energ*. 2006;83(1):1-14.
[6] Yik FWH, Burnett J, Prescott I. Predicting air-conditioning energy consumption of a group of buildings using different heat rejection methods. *Energ Buildings*. 2001;33(2):151-66.
[7] Howard B, Parshall L, Thompson J, Hammer S, Dickinson J, Modi V. Spatial distribution of urban building energy consumption by end use. *Energ Buildings*. 2012;45:141-51.
[8] Yang C, Létourneau S, Guo H. Developing data-driven models to predict BEMs energy consumption for demand response systems. Lecture Notes in Computer Science. 2014;8481(1):188-97.
[9] Hong T, Yang L, Hill D, Feng W. Data and analytics to inform energy retrofit of high performance buildings. *Appl Energ*. 2014;126:90-106.
[10] Yalcintas M, Aytun OU. An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. *Int J Energ Res*. 2007;31(4):412-21.
[11] Yalcintas M. An energy benchmarking model based on artificial neural network method with a case example for tropical climates. Int J *Energ Res*. 2006;30(14):1158-74.
[12] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl Energ*. 2014;127:1-10.
[13] EU directive 95/46/EC, 2012/148/EU2014/724/EU: Commission Recommendation of 10 October 2014 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems.
[14] Caputo P, Costa G, Ferrari S. A supporting method for defining energy strategies in the building sector at urban scale. *Energy Policy*. 2013;55(0):261-70.
[15] Ballarini I, Corgnati SP, Corrado V, Talà N. Improving energy modeling of large building stock through the development of archetype buildings. 2011. p. 2874-81.
[16] Ballarini I, Corgnati SP, Corrado V. Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project. *Energy Policy*. 2014;68(0):273-84.
[17] Korolija I, Marjanovic-Halburd L, Zhang Y, Hanby VI. UK office buildings archetypal model as methodological approach in development of regression models for predicting building energy consumption from heating and cooling demands. *Energ Buildings*. 2013;60(0):152-62.
[18] Mustafaraj G, Marini D, Costa A, Keane M. Model calibration for building energy efficiency simulation. *Appl Energ*. 2014;130(0):72-85.
[19] Kämpf JH, Robinson D. A simplified thermal model to support analysis of urban resource flows. Energ Buildings. 2007;39(4):445-53.
[20] Robinson DFH, Kämpf J, Leroux P, Perez D, Rasheed A, Wilke U. CITYSIM: comprehensive micro-simulation of resource flows for sustainable urban planning. Proc. Building Simulation 2009.
[21] Nielsen HA, Madsen H. Modelling the heat consumption in district heating systems using a grey-box approach. *Energ Buildings*. 2006;38(1):63-71.
[22] Lauster M, Teichmann J, Fuchs M, Streblow R, Mueller D. Low order thermal network models for dynamic simulations of buildings on city district scale. *Building Environment*. 2014;73(0):223-31.
[23] Kim E-J, Plessis G, Hubert J-L, Roux J-J. Urban energy simulation: Simplification and reduction of building envelope models. *Energ Buildings*. 2014;84(0):193-202.
[24] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. *Energ Buildings*. 2005;37(5):545-53.
[25] Zhou K-l, Yang S-l, Shen C. A review of electric load classification in smart grid environment. *Renew Sust Energ Rev*. 2013;24:103-10.
[26] Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew Sust Energ Rev*. 2014;33:102-9.
[27] Humeau S, Wijaya TK, Vasirani M, Aberer K. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households. 2013 Sustainable Internet and ICT for Sustainability - Conference Proceeding. 2013:1-6.
[28] Powell KM, Sriprasad A, Cole WJ, Edgar TF. Heating, cooling, and electrical load forecasting for a large-scale district energy system. *Energy*. 2014;74(0):877-85.

[29] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renew Sust Energ Rev*. 2009;13(8):1819-35.
[30] Ascione F, De Masi RF, de Rossi F, Fistola R, Sasso M, Vanoli GP. Analysis and diagnosis of the energy performance of buildings and districts: Methodology, validation and development of Urban Energy Maps. *Cities*. 2013;35(0):270-83.
[31] Ascione F, Canelli M, De Masi RF, Sasso M, Vanoli GP. Combined cooling, heating and power for small urban districts: An Italian case-study. *Appl Therm Eng*. 2014;71(2):705-13.
[32] Taylor S, Fan D, Rylatt M. Enabling urban-scale energy modelling: a new spatial approach. *Build Res Inf*. 2013;42(1):4-16.
[33] Mastrucci A, Baume O, Stazi F, Leopold U. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. *Energ Buildings*. 2014;75(0):358-67.
[34] Jones P, Patterson J, Lannon S. Modelling the built environment at an urban scale - Energy and health impacts in relation to housing. *Landscape Urban Plan*. 2007;83(1):39-49.
[35] Orehounig K, Mavromatidis, G., Evins, R., Dorer, V., & Carmeliet, J. (2010). Predicting energy performance consumption of a neighbourhood using building performance simulation. *Energy*; 2011.
[36] Glasgow Energy App: http://www.iesve.com/news/article/New-Energy-App-Helps-Residents-Cut-Bills/4332?modal=1.
[37] Paudel S, Elmtiri M, Kling WL, Corre OL, Lacarrière B. Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. *Energ Buildings*. 2014;70(0):81-93.
[38] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl Energ*. 2014;123:168-78.
[39] Albert A, Rajagopal R. Building dynamic thermal profiles of energy consumption for individuals and neighborhoods. Big Data, IEEE International Conference; 2013.
[40] Wojdyga K. Predicting heat demand for a district heating systems. 2007.
[41] Wijaya TK, Humeau SFRJ, Vasirani M, Aberer K. Residential Electricity Load Forecasting: Evaluation of Individual and Aggregate Forecasts. EPFL, 2014.
[42] Tian W, Song J, Li Z. Spatial regression analysis of domestic energy in urban areas. *Energy*. 2014;76(0):629-40.
[43] Yalcintas M. An energy benchmarking model based on artificial neural network method with a case example for tropical climates. *Int J Energ Res*, *30*(14), 1158-11742006.
[44] Hong S-M, Paterson G, Mumovic D, Steadman P. Improved benchmarking comparability for energy consumption in schools. *Build Res Inf*. 2013;42(1):47-61.
[45] Shabunko V, Lim CM, Brahim S, Mathew S. Developing building benchmarking for Brunei Darussalam. *Energ Buildings*. 2014;85(0):79-85.
[46] Wang M, Dongfang Wei, Jian Li, Hui Jiang, and Juliang Jin. A Novel Clustering Model Based on Set Pair Analysis for the Energy Consumption Forecast in China. *Mathematical Problems in Engineering* 2014.
[47] Alzate C, Sinn M. Improved Electricity Load Forecasting via Kernel Spectral Clustering of Smart Meters. IEEE 13th International Conference on Data Mining. 2013:943-8.
[48] Arambula L, Romagnoli P, Gasparella A. Selection of representative buildings through preliminary cluster analysis. International High Performance Buildings Conference 2014.
[49] Heidarinejad M, Dahlhausen M, McMahon S, Pyke C, Srebric J. Cluster analysis of simulated energy use for LEED certified U.S. office buildings. *Energ Buildings*. 2014;85(0):86-97.
[50] Nikolaou TG, Kolokotsa DS, Stavrakakis GS, Skias ID. On the Application of Clustering Techniques for Office Buildings' Energy and Thermal Comfort Classification. Smart Grid, IEEE Transactions on. 2012;3(4):2196-210.
[51] Gao X, Malkawi A. A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. *Energ Buildings*. 2014;84(0):607-16.
[52] Jones PJ, Lannon S, Williams J. Modelling building energy use at urban scale IBPSA conference, Rio de Janeiro, Brazil 2001.
[53] Lannon S, Georgakaki A, Macdonald S. Modelling urban scake retrofit pathways to 2050 low carbon residential building stock. *BS 2013*.
[54] Panapakidis IP, Papadopoulos TA, Christoforidis GC, Papagiannis GK. Pattern recognition algorithms for electricity load curve analysis of buildings. *Energ Buildings*. 2014;73(0):137-45.
[55] Jeremy Rawlings PC, Joe Doak and Ben Burfoot. A clustering approach to support SME carbon reduction. 2014.
[56] Wijaya TK, Ganu T, Chakraborty D, Aberer K, Seetharam DP. Consumer Segmentation and Knowledge Extraction from Smart Meter and Survey Data. SIAM International Conference on Data Mining (SDM14). 2014.
[57] Dall'O' G, Galante A, Torri M. A methodology for the energy performance classification of residential building stock on an urban scale. *Energ Buildings*. 2012;48(0):211-9.
[58] Fabbri K, Zuppiroli M, Ambrogio K. Heritage buildings and energy performance: Mapping with GIS tools. *Energ Buildings*. 2012;48(0):137-45.
[59] Li Z, Han Y, Xu P. Methods for benchmarking building energy consumption against its past or intended performance: An overview. *Appl Energ*. 2014;124(0):325-34.
[60] Chung W. Review of building energy-use performance benchmarking methodologies. *Appl Energ*. 2011;88(5):1470-9.
[61] Yoshiyuki Shimoda TF, Takao Morikawa and Minoru Mizuno. Development of residential energy end use simulation model at city scale. IBPSA Eindhoven, Netherlands 2003.
[62] Fumo N, Mago P, Luck R. Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. *Energ Buildings*. 2010;42(12):2331-7.
[63] Bishop CM. Pattern recognition and Machine Learning. Singapore: Springer; 2006.
[64] Gaitani N, Lehmann C, Santamouris M, Mihalakakou G, Patargias P. Using principal component and cluster analysis in the heating evaluation of the school building sector. *Appl Energ*. 2010;87(6):2079-86.
[65] Hsu D. Characterizing energy use in Ney York City commercial and multifamily buildings. ACEEE summer study on energy efficiency in buildings 2012.
[66] Zhao H-X, Magoulès F. A review on the prediction of building energy consumption. *Renew Sust Energ Rev*. 2012;16(6):3586-92.