DTU Library

Estimation of temperature setpoints and heat transfer coefficients among residential buildings in Denmark based on smart meter data

Gianniou, Panagiota; Reinhart, Christoph; Hsu, David; Heller, Alfred; Rode, Carsten

Published in:
Building and Environment

Link to article, DOI:
10.1016/j.buildenv.2018.05.016

Publication date:
2018

Document Version
Peer reviewed version

Estimation of Temperature Setpoints and Heat Transfer Coefficients Among Residential Buildings in Denmark Based on Smart Meter Data

Panagiota Gianniou[a, *], Christoph Reinhart[b], David Hsu[c], Alfred Heller[a], Carsten Rode[a]

[a] Department of Civil Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark

[b] Department of Architecture, Massachusetts Institute of Technology, Cambridge MA 20139, USA

[c] Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge MA 20139, USA

* Corresponding author, e-mail: pgianniou@gmail.com

Abstract

Thermal comfort preferences of occupants and their interactions with building systems are top influential factors of residential space heating demand. Consequently, housing stock models are sensitive to assumptions made on heating temperatures. This study proposes a heat balance approach, inspired by the classical degree-day method, applied to an extensive urban dataset. The goal of this analysis is to determine heterogeneous characteristics, such as temperature setpoints of heating systems and thermal envelope characteristics from an overall population of residential buildings. Measured energy data are utilized for the purpose of the study from the city of Aarhus, Denmark, where the energy usage for heating of circa 14,000 households was monitored over time via smart meters. These data are combined with actual weather data as well as data extracted by a national building database. Using linear regression and heat balance models, temperature setpoints for the whole dataset are determined with a median and average of 19ºC and 19.1ºC, respectively. Furthermore, building related characteristics such as thermal and ventilation losses per building and overall heat transfer coefficients are extracted at urban scale. The reliability of the method over its complexity is discussed with regards to the big sample that has been applied to. In general, the overall performance of the approach is satisfactory achieving a coefficient of determination with an average of 0.8, and is found to be in line with previous findings, considering

also the high uncertainty associated with building-related input parameters. The extracted setpoint distribution should be transferrable across Scandinavia.

Keywords: smart meter data, temperature setpoints, housing stock model, thermal comfort preferences, U-values, urban scale

1 Introduction

A number of modeling methodologies have been developed to obtain information on physical resource flows through the building stock [1]. These are mainly used to characterize and predict energy demand of residential building stocks and to estimate energy savings after energy retrofitting strategies. Housing stock models can thus play an important role in supporting energy policy-making. In order to be useful, they should be reliable, efficient and interpretable [2]. Housing stock models can be broadly classified into two categories: top-down and bottom-up approaches. Top-down models rely on historical energy data and cannot model in detail individual end-uses [3]. Bottom-up models consist of engineering-based and statistical models. Statistical methods usually include macroeconomic and socio-economic effects, enable the determination of end-use energy consumption and are easy to develop and be used [4]. However, they cannot model the impact of specific technologies implemented and are less flexible. Engineering-based housing stock models use actual building physics and overcome some of the limitations induced by statistical models [2]. However, the majority of them are developed at national scale to support policy making and disregard heterogeneity within a country. They are also usually time intensive and are fully dependent on input data, hence inducing a high degree of uncertainty. Therefore, there is a need to focus on regional housing stock models that handle heterogeneity.

According to the International Energy Agency in the Energy Buildings and Communities Program (IEA EBC) Annex 53: Total Energy Use in Buildings, the six driving factors of energy use in building stock are : i) climate, ii) building envelope, iii) building energy and services systems, iv) indoor design criteria, v) building operation and maintenance, and vi) occupant behavior. Even though significant progress has been made in quantifying these primary drivers, more emphasis on energy related occupant behavior in buildings is needed to develop reliable and standardized methods [5, 6]. Neglecting this aspect can lead to severe miscalculations and inaccurate conclusions about the energy performance of the building stock [7]. Occupants' interaction with building systems affects significantly the total energy use of buildings. The occupants' gratification with their thermal environment defines thermal comfort [8]. Therefore, the occupants' perception of comfort or satisfaction in the built environment drives them to perform various controls (e.g. on HVAC systems and window operations) [9]. The adjustment of thermostat setpoints and indoor thermal environment are the most influential factors of heating loads along with heated areas [10]. Some studies have even classified occupants as active, medium and passive users based on their heating setpoint preferences which impact the indoor thermal environment and energy consumption [11, 12]. Therefore, thermostat setpoints are crucial input parameters to building energy models due to their big influence on residential energy use [9]. Currently, the understanding of occupant behavior is still insufficient both in building design, operation and retrofit, leading to incorrect simplifications in modeling and analysis [5]. In the past, information about occupants' interactions with systems was based on sporadic visits to households and rough estimates of thermal preferences of occupants.

The increasing deployment of intelligent metering systems in buildings and district systems creates a vast amount of building energy use and occupant-related information. Following the Third Energy Package in the Electricity [13] and Gas Directive [14] issued by European Commission in 2006, European countries plan to convert part of their legacy meter stock to smart by 2020 with a focus on electricity. According to the projections, by 2020, it is expected that almost 72% of European consumers will have a smart meter for electricity and about 40% will have one for gas [15]. The enormous amount of information and data opens up endless opportunities for researchers and engineers to study building dynamics and performance at a large scale. In combination with weather data and cross-sectional data, they can be utilized to develop more accurate prediction models and detailed analyses on the drivers of building energy consumption [16]. Smart meter data can also help developing and applying control strategies to improve building energy performance and efficiency [17]. Therefore, they can be utilized to decrease uncertainty related to building energy performance and occupant behavior and provide detailed information on energy monitoring.

National building databases and registers can support housing stock energy analysis, by providing information about building typologies and construction characteristics. These databases are usually created with regards to building regulations and schemes. In some cases, information from building owners via questionnaires has also been collected. Building information can be updated by local authorities and by citizens. However, occupants' interventions on the building fabric (i.e. energy renovation measures) are not regularly reported to building databases. Therefore, there is a significant gap between the data that has been registered and the real energy performance of the building.

This study aims at utilizing a big urban dataset, consisting of smart meter data from more than 14,000 households in a Danish city, to estimate temperature setpoints and thermal transmittances on building level. In addition, actual weather data, as well as data collected from a national building register and a geographic information system (GIS) have been utilized. A heat balance approach is implemented to the measured energy data of one year applying linear regression analysis to extract parameters that represent the whole heating season and the total building envelope. This approach has been inspired by the degree-day theory and aims at providing a new useful tool for utilities and researchers to extract building and thermal comfort-related characteristics at urban scale based on smart meter data. The data used allows us to capture the full range of heterogeneous behavior among people, through their temperature preferences. The estimation of people's variation enables the development of customized solutions and messages for them. The estimated thermal transmittance of the building envelope indicates the refurbishment state of the building and thus, provides more accurate insights into the building stock. The generated results -in the form of distributions- can be used to improve urban building energy models for the Scandinavian housing stock.

The rest of the article is organized as follows. In section 2, related works and methods to predict room temperature setpoints are summarized. In Section 3, we present and apply the heat balance model to the smart meter dataset. In Section 4, the dataset is presented and basic information about the examined housing stock is described. In Section 5, the results are compared and validated with previous findings and relevant literature. The applicability of the methods with regards to the considerations made and the data used is discussed in Section 6. Section 7 summarizes the research findings.

2 Background

To evaluate the potential impact of different energy retrofitting scenarios in urban areas, bottom-up urban building energy models (UBEM) have been introduced over the past years. UBEMs have the potential to become key planning tools for utilities, municipalities and urban planners [18]. A key input of UBEM models are building characteristics of a given building stock from thermal envelope properties to usage patterns including the number of occupants, equipment loads and schedules as well as thermostat settings. Some of those information may be derived from census data. However, there is generally a surprising lack of data available related to the thermal performance of buildings. A useful source of information can be derived from individual building energy audits [19]. In [19], the authors used the Monte Carlo method and created a physics-based housing stock model for energy performance prediction, where inputs were probability distributions based on an Energy Performance Certification national database. Another source are nationwide building databases which include information such as floor areas, construction materials, age of construction, etc. Nevertheless, these databases may have flaws or may not be updated frequently enough. Therefore, there is high uncertainty related to input parameters of UBEMs.

Several studies have been conducted on district or urban scale making use of statistical models and data mining techniques in order to extract hidden useful knowledge from building-related data, as well as to forecast energy consumption. The authors of [20] presented a data-driven approach to modeling end user consumption based on data from 6,500 buildings in Cambridge, Massachusetts, using linear regression analysis and Gaussian process regression. In [21], electric energy data of thousands of buildings were investigated to extract specific features based on socio-economic information. In [17], a data mining method was proposed to analyze building-related data in order to establish building energy demand predictive models and examine the influence of occupant behavior on energy consumption. Older studies had also made use of regression analysis based on billing data to determine household energy. A study by [22] used monthly energy billing data to decompose energy use to weather and non-weather dependent elements, as well as explain anomalies in energy use of some households.

Determining the internal temperatures or temperature setpoints has been of particular interest, especially in residential buildings in mostly heating dominated climate since the main source of building energy demand is driven by heating which in term directly depends on the temperature difference between inside and outside.

Temperature setpoints and heating duration may differentiate across dwellings based on their preferred thermal comfort range, which affect the resulting internal temperatures. Nevertheless, many top-down urban scale models assume the same constant temperature setpoints for the whole building stock, while the rest calculate internal temperatures as a function of building envelope, occupancy and systems [23].

Most of the existing literature puts emphasis on predicting internal temperatures or measuring household room temperatures at district scale based on temperature recordings [24]. For example, the authors in [25] collected temperature and humidity data from 1,604 study dwellings in order to determine the effect of dwelling and household characteristics on indoor temperature variation. The median standardized daytime living room temperature was calculated to be 19.1^{o}C, while the night time bedroom temperature was found to be 17.1^{o}C. Temperatures were affected by the building envelope characteristics, thermal efficiency, number and age of occupants. The socio-economic status was not strongly related to them. In [26], temperature recordings from 821 English dwellings were collected and analyzed, which were monitored in different zones within the dwellings. The standardized internal temperatures were estimated by regressing the mean hourly indoor temperatures on outdoor temperature. The results showed that more efficient buildings have higher indoor temperatures at all outdoor conditions, as well as that households with children were the warmest. In another study [23], internal temperatures were predicted at high temporal resolution using panel methods based on data from 280 households. Temperature recordings were taken every 45 minutes for 6 months, while the model was generated using mean daily temperature data. It was concluded that thermostat settings play an important role in reducing total energy consumption. Moreover, centrally heated dwellings and detached homes had lower internal temperatures, while for each additional person living in a household the mean internal temperature increased by ~0.25^{o}C. In another study [27], internal temperature recordings were gathered by 54 households in China, which were found to be on average 13.5^{o}C for living rooms and 12.7^{o}C for bedrooms. These are far outside the ASHRAE steady state thermal comfort zone, highlighting the differences among climates in terms of construction and cultural behavior. Furthermore, the effect of the climate is pronounced on these findings, as this study referred to the 'hot summer-cold winter' climatic zone of China being characterized by relatively low ambient temperatures during winter season -which is quite limited- and very high temperatures during the rest of the year. However, these winter climate conditions are comparable to the ones in UK. The very low internal temperatures noticed in this study are a result of the fact that domestic heating is operated part-time-part-space in that region of China, meaning that only specific rooms are heated up instead of the whole house and only for a shorter time than a usual heating season duration.

Apart from the measurement of internal temperatures, some studies have tried to predict them. Most of them refer to smaller cases and make use of simple heat balance models. The authors in [28] used the Domestic Energy and Carbon Model [DECM] to estimate among others the annual internal temperatures in different types of English dwellings, which were on average 18.4°C. The mean internal temperature was also found to be highly correlated with the CO_2 emissions of each dwelling and outweighed the effect of climate and building fabric construction. The authors of [29] proposed a method to explore future transformations in the UK housing stock based on the English House Condition Survey data. Besides energy demand prediction, they used the heat-balance method to infer the appropriate modeling temperatures for a base year in UK. In particular, the profiles for the internal temperatures varied from 14.1 to 18.7°C for one of the two modeled zones, while these temperatures were 3°C lower for the other zone. The study by [30] modeled 37 dwellings employing the BREDEM algorithms [31] and predicted the annual internal temperatures and the heating demand temperatures which equals the temperature setpoint in most cases. It should be noted that most of the afore-mentioned studies refer to the UK housing stock, while there is a lack of similar literature for Scandinavian or Danish building stock.

In cases where internal temperatures are not available, degree-days have been used as a tool to assess and analyze weather related energy consumption in buildings for over seven decades. The concept originates from agricultural research, where variation in outdoor air temperature is also important which is also the case for building energy use [32]. Degree-days can be defined as the summation of differences between the outdoor temperature and a building reference temperature (base) over a specified time period for both heating and cooling systems. The simplicity of the concept of heating degree-days (HDD) has led to a plethora of studies in the area of building energy use analysis. The influence of HDD and heating degree hour (HDH) on hourly electricity consumption of hundreds of Norwegian households was investigated and concluded that HDD achieve a slightly higher goodness of fit compared to HDH [33]. The authors in [34] proposed a city-scale degree-day method, according to which they extracted the average building heat loss rate and a city-scale base temperature for the area of Strasbourg in order to estimate the aggregate heating energy demand, while accounting for the urban heat island effect. In addition, residential heating energy requirements and fuel consumption for the city of Istanbul were estimated making use of HDD method and air temperature records, while studying different construction types [35]. A study by [36] utilized energy data to estimate the base temperature of a single building along with the heat loss coefficient and subsequently heating degree-days using Bayesian inference. The performance line method and energy signature method were presented in [37] to estimate the building's base temperature based on daily energy data and outdoor temperatures and to extract degree-days.

3 Method

The method applied in the current experiment takes an existing heat balance approach a step further. The temperature setpoints and total heat loss coefficients are estimated from smart meter data for thousands of buildings. The granularity of energy data is hourly and the sample of buildings covers a large share of a Danish urban building stock, as opposed to previous studies that have been found in literature and cover much smaller building samples. This allows to draw conclusions about the applicability and accuracy of the method applied.

3.1 Main considerations

To treat this urban-scale sample of buildings, the following considerations were made. Firstly, only the coldest days of the year were considered. In particular, only the data that correspond to days when average daily ambient temperature was lower than 15°C were taken into account. In this way, a normal operation of the heating system would be ensured. Moreover, solar gains are reduced in Denmark especially during winter time. This results in transient phenomena being less dominating. Thus, steady-state conditions would be applicable. The investigated buildings are small enough that uniform air mixing was assumed. All households were treated as single-zone models in order to reduce complexity and computation times. In addition, the investigated housing stock is homogeneous with regards to building type (consisting only of single-family houses), allowing a somewhat accurate estimate of occupant density and construction standard. As no information about specified occupancy schedules was available either, internal heat gains were defined according to the Danish Building Research Institute guidelines [38] as 5 W/m^2 (corresponding to external floor areas), which is a sum for people loads and equipment loads for a residential building. This value was used as an average for all different spaces in households (kitchen, living room, bedrooms etc.), so that it is in line with the single-zone model approach. Furthermore, the 'equivalent' temperature setpoint, which represents the volume-weighted mean temperature across all conditioned and unconditioned spaces in every household, was assumed to be constant throughout the heating season that we investigated.

3.2 Calculation of equivalent thermostat heating temperatures and U-values

Inspired by the classical degree-day method and based on the steady-state heat balance for a room, the following formula (equation 1) can be derived. According to that, total energy loads for building space heating over a specified period are directly proportional to the building heat losses that vary with the change in the current indoor-outdoor air temperature gradient (i.e. natural ventilation and air infiltration, wall heat conduction). Similar equations to the following heat balance model, solved for Q_H, have been traditionally used to calculate the heat output of well controlled heating systems [37].

$$T_o = T_i - \frac{Q_H + Q_{SG} + Q_{IG}}{UA + c_p\, \rho\, n} \tag{1}$$

where T_i is the equivalent internal temperature in ^oC, T_o is the ambient temperature in ^oC, Q_H is the heating demand in W, Q_{SG} stands for solar gains in W, Q_{IG} represents the internal heat gains from occupants and equipment in W/m^2, U is the overall heat transfer coefficient across the building envelope in W/m^{2}K, A is the total envelope area in m^2, n is the ventilation or infiltration rate in m^3/s, c_P is the specific heat capacity for air under constant pressure which equals to 1000 J/kg K and ρ is the density of air which equals to 1.2 kg/m^3. The heat loss factor due to the air change can be also simplified to equal $0.33\, N\, V$, where N is the ventilation rate in air changes per hour (ACH) and V is the volume of the building in m^3.

Due to the extensive data quantity both in terms of number of monitored buildings and data granularity, complicated approaches would be less suitable. So, the methodology applied in this case was simple linear regression analysis. The ordinary least squares (OLS) method was used and run for each household's data. According to [39], the OLS method is the most efficient for urban scale models having the lowest deviation from measured data.

The four main statistical assumptions regarding OLS method were tested to prove whether they were satisfied on building level: the regression is linear in parameters, the sampling of observations is random, the conditional mean is zero and there is no multi-collinearity.

Thus, in the current model, the ambient temperature T_o was modeled as an affine function of the total energy load, so that $T_o = T_i - \frac{load}{losses}$, where $load = Q_H + Q_{IG} + Q_{SG}$ and $losses = UA + 0.33\ N\ V$. The internal temperature or temperature setpoint and the losses are expressed by the intercept and the slope of the linear model, respectively. The losses represent the envelope and ventilation losses. The energy load quantity will always be positive. Q_H represents the space heating consumption of the households which was given on hourly resolution. However, in order to account for latent thermal mass effects in the buildings, it was decided to aggregate the hourly energy data on daily basis and report the aggregate value of energy consumption (heating). Thus, internal temperatures in equation 1 refer to the mean of internal temperatures recorded, representing the relative heating profile of a dwelling. The internal gains were also aggregated on a daily resolution based on the indicated hourly values.

In the afore-described model, the heat losses are mainly attributed to conduction through the building envelope and to ventilation losses. For that reason, we are focusing only on the winter time or heating season when the ambient temperatures are low; furthermore, the solar gains are significantly lower. Thus, as mentioned before, the regression model was restricted to periods when the mean ambient temperature was less than 15^{o}C. Information concerning building characteristics and occupancy were retrieved by the national building database and by national building regulations, where no detailed information was available. Actual weather data on hourly resolution were acquired for the area of Aarhus covering the examined period (2014-2015) and included ambient temperature, solar radiation and relative humidity.

The calculation of solar gains was based on the actual weather data available for the area consisting of direct and diffuse radiation. An exemplary building was made in DIVA software [40], which is a daylighting and energy modeling plug-in for Rhinoceros, according to a Radiance calculation, where time-integrated solar irradiance on each surface was calculated as presented in [41, 42]. Radiance is a backward raytracer that was originally developed at Lawrence Berkeley National Laboratory [43]. An average value of integrated solar irradiance for all four surfaces of an exemplary building was taken, which was then inserted into the model. According to the following equation (2), the solar gains are calculated as the product of overall solar heat gain coefficient (*SHGC*), projected area of fenestration A_f in m^2 and incident total irradiance E_t in W/m^2.

$$Q_{SG} = SHGC_t \, A_f \, E_t \qquad (2)$$

To estimate the equivalent overall heat transfer coefficient of the building envelope for each household, the losses factor in equation 1 had to be determined. The geometry of the houses (i.e. envelope areas) was available through an additional dataset consisting of GIS data at city scale. Therefore, the ventilation and infiltration rate had to be determined taking into account the type of mechanical systems installed in each house, possible insulation state of building envelope and air leakages.

4 Data

4.1 Typical dwelling characteristics in Denmark

Approximately 45% of Danish dwellings are detached houses [44], although their share is significantly less in larger cities (e.g. Copenhagen or Aarhus). A large share of heating (namely 63%) in private Danish houses is provided by district heating, which covers both space heating and domestic hot water demand [45]. Hydronic systems are the most common heat emission systems found in residential buildings. Mechanical cooling and ventilation is mostly available in commercial buildings. The national average energy consumption for Danish single-family houses constructed in the period before 1980 is 151 kWh/m^2 per year and 102 kWh/m^2 per year for houses constructed between 1980 and 2000 [46]. That gives an indication of the energy efficiency of the investigated building stock. Moreover, according to the Danish building regulations, there should be a sufficient amount of thermal mass and insulation layer in newer buildings, which results in more airtight building envelopes. More information and typical building construction examples of Danish single-family houses can be found in [47].

4.2 Description of smart meter data

Our approach is evaluated on basis of smart meter data collected initially from 15,063 households in Aarhus, Denmark at a 60-minute granularity of one full year. The data is provided by Aarhus AffaldVarme (AVA), which supplies district heating to the inhabitants of Aarhus. The data ranges from August 2014 to August 2015. The data contains the heating consumption of residential AVA customers. The households are all single-family houses (SFH). The periods that the data cover range depending on when each of the smart meters was installed. There is a slightly smaller number of smart meters that have recordings for the whole investigated period. Thus, energy data from 14,182 households are utilized in the experiment, which correspond to the whole examined period. The data has been cleansed and aggregated at a daily time scale and the results of this analysis are presented in the results' section.

4.3 Building characterization

The dataset covers a big spectrum of single family houses in Aarhus, ranging from buildings constructed in 1800 to 2015. The floor areas range from $25m^2$ to $504m^2$, with a mean of $134m^2$. The number of floors varies between 1 and 3, with the majority of the buildings being single-floor houses. These information are presented in Figure 1.

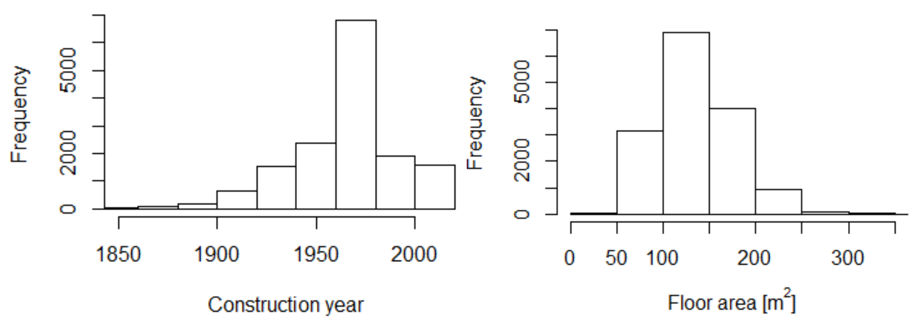

Figure 1. Building information of dataset

5 Results

The afore-mentioned method was applied to smart meter data of 14,182 households over one full year and the results are presented hereafter.

5.1 Variables' distributions

Figure 2 presents the main results of this study, which are the temperature setpoint distribution and the distribution of the total envelope and ventilation loss factor for all examined households, as estimated according to equation 1. As it can be seen, the mean equivalent temperature setpoint that was calculated for the whole sample of buildings was 19.1ºC, while the median value was 19ºC. The standard deviation was 1.54ºC. This temperature is a bit lower than the expected indoor temperature value, which would be close to 20^{o}C according to estimations of the Danish Building Research Institute [48] included in Danish Standards [49]. It can be attributed to two reasons. First, every household is modelled as a single-zone, which means that any unconditioned spaces are also included and represented with this specific equivalent parameter. Second, this figure refers to equivalent temperature setpoint, which is inferred over a whole season. So, unoccupied periods and hours with night-setback are also included in this estimation. The internal heat gains that were assumed in the model proved to have a decisive impact on the temperature results, so higher internal gains than the ones defined in the standard would lead to higher temperatures. Nevertheless, the uncertainty associated with the internal and solar gains did not result in high inaccuracy of the model. Daily internal heat gains ranged from 5 kWh to 110 kWh among the different households. The temperature setpoint distribution seems to approach normality, while the building envelope and ventilation losses graph approximates a log-normal distribution curve. The mean building envelope and ventilation losses were calculated to be 0.1 kW/K with a standard deviation of 0.03 kW/K. Additional conclusions can be drawn on how close to expectation the current used model has performed and the goodness of the accuracy that can be achieved at such a large scale. Despite the steady-state heat balance model that was applied on the building sample and the considerations that were made about input parameters, the results indicate a reasonable range of temperature setpoints and loss factor across the dwellings. In order to interpret the thermal environment and thermal comfort that occupants perceive based on the estimated equivalent internal temperature, the Predicted Mean Vote (PMV) and Predicted Percentage Dissatisfied (PPD) indices proposed by Fanger [50] were used. The average PMV was calculated to be -0.6, which corresponds to PPD equal to 12.9%. Based on these estimations, the thermal environment falls into Category C out of the three desired thermal environment categories as described in [51].

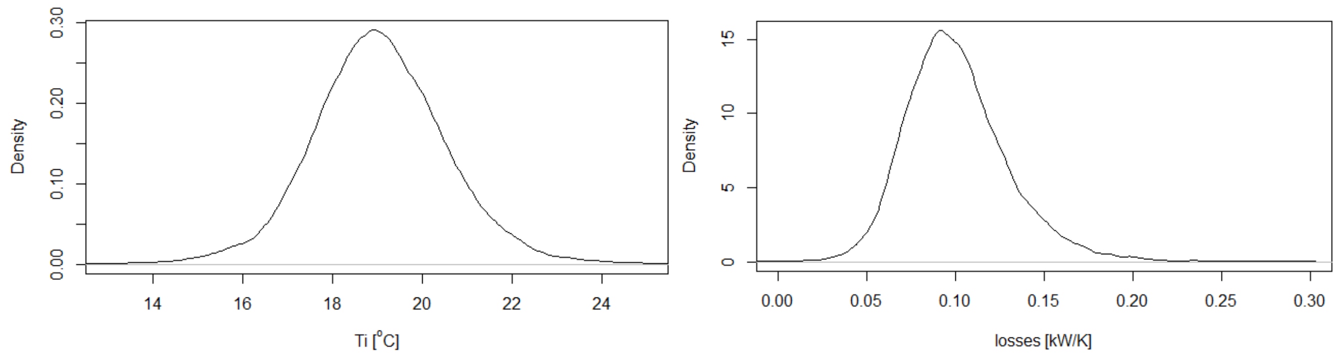

Figure 2. Distribution of temperature setpoint (left) and total building envelope and ventilation losses (right)

To assess the goodness of the fit for the simple linear regression models, the coefficient of determination (R^2) was utilized to estimate the variance of the predictable variable from the independent variable. The results are presented in Figure 3. It has to be noted that a relatively high R^2 is achieved with respect to the nature of the experiment, which regards end users energy use. The mean and median value of the coefficient of determination were found to be 0.8 and 0.83, respectively, with a standard deviation of 0.11. That confirms the good fit of the linear regression models to the smart meter data, as well as a quite consistent model, considering the large amount of data that has been used.

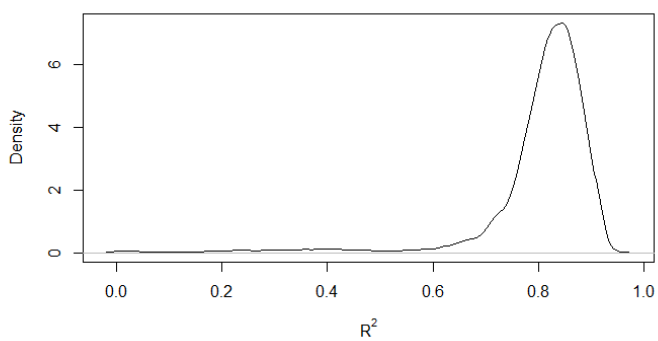

Figure 3. Distribution and boxplot for R^2 for all houses' fits

In Figure 4, the residual standard deviation distribution is illustrated, which can indicate the variability of predictions in the regression. This shows the deviation of the errors and not the errors of the regression themselves. It can be observed that the deviation of the residuals has a mean of 1.53, which is satisfactory considering also the big building sample.

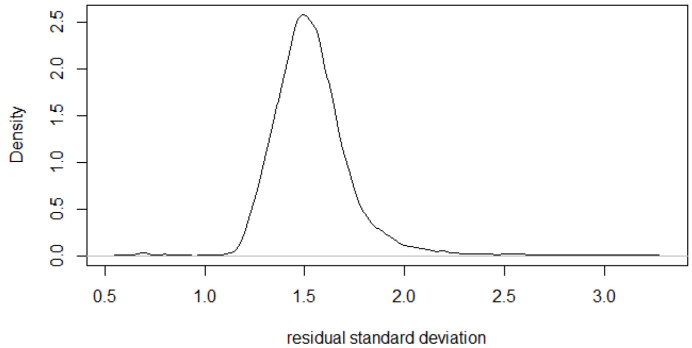

Figure 4. Residual standard deviation distribution for all houses

5.2 Exemplary households

The regression models for two randomly selected households are presented hereafter. The intercept of the models represents the estimated temperature setpoint, while the slope stands for the coefficient of envelope and ventilation losses. The coefficient of determination for these two cases were R^2=0.84 and R^2=0.8, respectively. The estimated temperature setpoints for these two models were 19.39°C and 18.91°C, respectively. Both models are quite similar, with the household presented to the right having a bit more scattered energy use data and higher number of outliers.

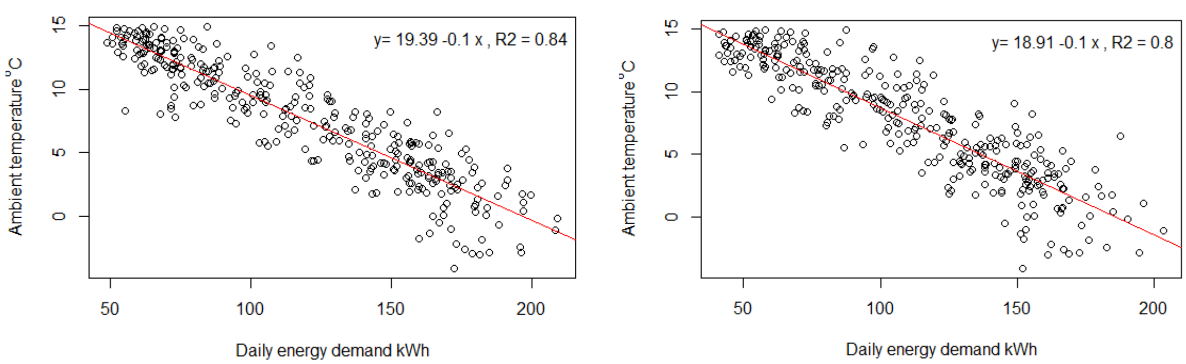

Figure 5. Regression fit for two random households

5.3 Regression diagnostics

The main assumptions of OLS were tested for numerous randomly selected models and the following conclusions were drawn. First, normality was tested by checking the probability plot of the standardized residuals against the values expected under normality, concluding that the normality assumption is satisfied. Second, the residuals did not have non-linear patterns, confirming the linearity condition. Third, the constant variance assumption was met, since the residuals were spread equally and randomly along the ranges of predictors. The errors of the regression models had zero mean as required by the Gauss-Markov theorem. The independence assumption was also satisfied. Lastly, the outliers that were influential to the regression results were determined. That means that these may be influential to the regression analysis and the results will change if we exclude these cases.

5.4 Estimation of equivalent total U-value of building envelope

After the total loss factor for each household was calculated, the equivalent overall heat transfer coefficient, U-value, for the building envelope was calculated. This would provide additional valuable information for the insulation state of the buildings and indicate any possible energy refurbishments. Thus, the ventilation and infiltration rate on building level had to be determined. According to [52], the infiltration rates for Danish single-family houses vary between 0.1 and 1 ACH. These were estimated for each household in accordance with available information about the age of the building and construction characteristics extracted from the national building register. Specifically, an algorithm was used according to which the infiltration rate took values in *[0.1, 1]* based on the construction year of the building and a binary variable that indicated if the house had undergone any energy renovation that has been registered in the national building database. These two factors would collectively indicate any thermal bridges on the building envelope affecting the infiltration rate. The estimated infiltration rate for each household was assumed to be constant throughout the course of the day. The mean value of natural ventilation rate for the investigated building stock was estimated to be 0.5 ACH. Rough information for the geometry of the houses and the envelope areas on building level was available through the Danish Building Register (BBR) [53]. BBR is a nationwide register including data for the majority of Danish buildings and households. Nowadays, it contains information about 1.6 million properties, 3.8 million buildings and 2.7 million dwellings and commercial units [54]. It was originally set up in 1977 by collecting information from building owners via questionnaires. Since then, it has been updated by local authorities and by citizens [55]. Data contained in BBR -provided for every registered house- are categorized into areas, building constructions and installations. Information about areas can be summarized to the following: total building/residential/commercial area, built-up area, number of storeys, attic and total basement area. The information extracted for the current analysis was the building footprint area and the number of floors for each building. This information was then coupled and validated with an open-source GIS dataset for the city of Aarhus [56]. Thus, the total envelope area for each building was extracted, as well as the total building volume.

The results of this analysis are presented in Figure 6. The distribution approximates a log-normal curve. The mean value of the equivalent U-values for the examined residential building stock is calculated to be 0.58 W/m^{2}K with a standard deviation of 0.22 W/m^{2}K. The median value is 0.54 W/m^{2}K. This finding gives insight into the state of the building envelopes across Danish SFH, as well as their current energy refurbishment state. According to the Danish Building Research Institute [57], the average area-weighted U-values of Danish single family houses -as calculated with regards to the national building regulations- mainly vary from 0.3 to 0.65 W/m^{2}K depending on the construction age of the building. It should be noted that the majority of the investigated single-family houses were constructed in the 1960's, which have an average area-weighted U-value of 0.52 W/m^{2}K according to [57]. The overall heat transfer coefficients of the buildings do not always comply with the ones defined in the national building databases or introduced by the building regulations. Therefore, the calculated U-values can determine the level of energy refurbishment that might have been implemented to the buildings. Such updated information, although very important for energy calculations, is missing from the majority of the Danish building register [58]. The results from this study can also validate existing values in building databases. In addition, since this equivalent U-value summarizes the heat transfer coefficient for the whole building structure, no specific conclusions can be drawn for the particular building components. However, they can be assessed in combination with the rest of building characteristics and reveal valuable information about the building energy performance of the stock.

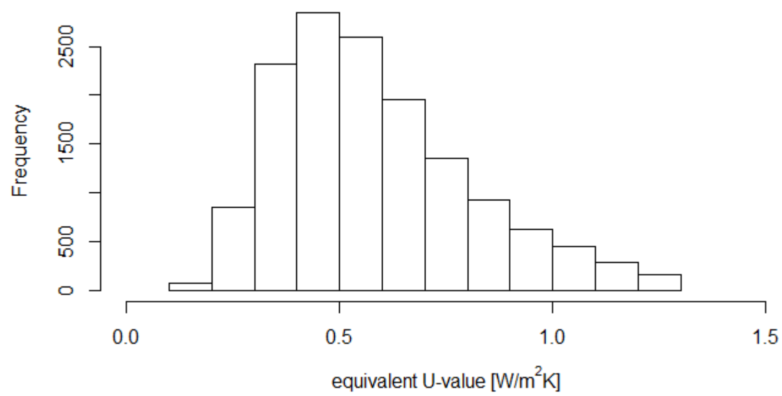

Figure 6. Distribution of equivalent U-values for the building stock

6 Discussion

The proposed method aims at investigating the applicability of a simplified heat balance approach on smart meter data to derive temperature setpoints and thermal transmittances at urban scale. Our analysis focused on fitting the overall heating energy distribution across the investigated building stock. The lack of information about the share of heated and unheated space resulted in the use of an equivalent temperature setpoint, which represented the volumetric mean temperature indoors. That means that there may be rooms that are warmer (i.e. occupied spaces) and rooms that are cooler in a building. However, the scope of our analysis was to treat the whole building as an entity. In Danish building regulations and directives, a dimensioning internal temperature of 20°C is commonly used for the majority of Danish houses. However, it is also mentioned that in older poorly insulated houses, the room temperature can be even lower than 20°C or some rooms are unheated to reduce the heating costs [48]. The thermal environment evaluation was conducted assuming that the living spaces (i.e. whole house) were heated. However, if unheated spaces were to be evaluated, the desired thermal environment follows a different classification. According to steady-state conditions of the model, changes in temperature sepoints were not taken into account. The use of a steady-state model resulted in extreme conditions and transient states not being taken into account. If a transient model was to be used instead, the temperature setpoint would no longer be represented by a single value for the whole heating season but replaced possibly with daily or even hourly ones. Moreover, the assumptions on constant infiltration rates throughout the day and over the heating season and constant internal gains throughout the day and among the different households would have to be adjusted. In particular, varying heat flow rates from occupants and appliances could be used based on hourly schedules and differentiating between weekdays and weekends. Nevertheless, the current study aimed at producing results that could be utilized subsequently by urban building energy models or housing stock models, which in their majority require a single heating temperature setpoint per building or zone. Thus, it would be outside the scope of this work to calculate dynamic temperature setpoints. In addition, this would increase the computational complexity enormously and subsequently the computation time for urban-scale or district-scale building stocks.

The biggest uncertainty on the total energy load factor comes from the internal gains and solar gains factor. Internal gains were assumed to differentiate according to floor area based on Danish standards. As a result, the residents of the investigated area were assumed to have the same occupancy patterns. Internal gains depend on occupancy behavior and schedules, as well as cultural patterns that can be estimated but not be predicted accurately at a large scale. Thus, a certain amount of assumptions based on recommendations by standards had to be taken, which resulted in a higher uncertainty in the housing stock model predictions. Solar gains had comparatively smaller impact on the energy load and the assumption about an average orientation did not seem to affect significantly the results. This should be considered in combination with the local climate, which in this case is characterized by relatively low solar radiation during the heating season and low ambient temperatures.

Despite the considerations that were made regarding the envelope properties, the performance of the proposed approach was satisfactory compared to existing knowledge and statistical values included in literature. The analysis was run on a private scientific cloud (SciCloud) using a web-based interface to interact with the data [59]. SciCloud consists of 18 physical servers with 80 cores and 564 GB of memory, as well as 4.2 TB of node storage, plus 1.2 TB of network storage. The environment used for the data analysis was *R*. The computation time for the whole housing stock analysis did not exceed two hours, due to the low complexity of the heat balance models and the OLS method. Therefore, the proposed model is expected to balance predictive accuracy and parsimony. If more complicated heat transfer phenomena were to be included in the model, the computation time would increase significantly. The big amount of historical measured energy data led to statistically significant results and strengthened our methodology. On the other hand, such large-scale energy data are mostly available to utilities and less to the research community due to privacy issues. Anonymization techniques can be applied and facilitate significantly the data acquisition and publishing process.

Measurements of internal temperatures for a smaller sample of the examined building stock would validate this methodology. The installation of smart sensors in newly constructed buildings measuring internal temperatures in rooms would allow this. However, the share of newly constructed buildings remains quite low in the city of Aarhus and the employment of smart sensors in the existing building stock would be a relatively long and costly procedure. Moreover, the application of the proposed method on a much smaller sample, which more detailed building and occupant-related information (i.e. infiltration rates, occupancy schedules) would be available for, would be a next step to continue this study. In this direction, occupant data could be used to approximate internal heat gains more accurately. The current sample is a relatively homogeneous building stock, including only single-family houses in the same urban area. Their thermal envelope characteristics varied a lot, though, in terms of building envelope insulation, construction materials and geometry. Thus, a wide range of building construction types was covered. Our method proved to work reliably on the examined stock with regards to the number of assumptions that were made. This resulted in the estimated parameters including high second-order uncertainty. To cope with this uncertainty, probabilistic methods would be required which would be less attractive for this case of thousands of buildings. Despite the uncertainty included in the input variables and estimated parameters, heterogeneous preferences of occupants regarding thermal comfort were determined successfully. Nevertheless, further work needs to be carried out to investigate how it can be expanded upon more diverse building stocks. Finally, the Danish climate is characterized by decreased solar gains and cold winters, which make the transient phenomena being less dominating. If this work was to be reproduced to different climates with increased solar gains, the quasi-steady-state conditions would be less valid. Also, equation *(1)* would have to be adjusted accordingly so that a utilization factor for solar heat gains (e.g. function of time constant) is included that accounts for the dynamic heat flows within the building.

7 Conclusion

This study has utilized an urban dataset of more than 14,000 households in Aarhus, Denmark to derive temperature setpoints and overall heat transfer coefficients at house level. This dataset comprised of measured daily heating energy data, actual weather data, basic building typological data and geometry information extracted from a building register and GIS data. A heat balance model –inspired by the degree-day theory- was proposed and applied in combination with linear regression analysis. The results showed that a good fit was achieved overall in the majority of the examined households. The results provided distributions of i) equivalent temperature setpoints, ii) a factor for building envelope and ventilation losses, as well as iii) equivalent U-values for the building envelopes. The average equivalent temperature setpoint was calculated to be 19.1ºC across the investigated Danish dwellings, considering both heated and unheated spaces in the buildings. This value represented the mean volumetric temperature indoors. The mean overall heat transfer coefficient for the total building envelope of all houses was estimated to be 0.58 W/m^{2}K. The mean value of the coefficient of determination (R^2) was 0.8, indicating a good fit of the linear regression models. It was found that Danish homes differed in heating setpoint temperatures and envelope insulation state. The energy data was proven to be highly correlated with the weather data (consisting of ambient temperature and solar radiation), as expected. Statistically significant results have been reported due to the big sample size and the consistent and granular energy measurements. Therefore, the proposed steady-state approach is applicable and recommended for urban-scale building samples when a uniform setpoint temperature is adequate to be extracted for the whole heating season. Furthermore, overall heat transfer coefficients for the whole building envelope can be used to determine any possible energy retrofit measures that have been applied to the envelope. Moreover, this method enabled the capturing of the full range of heterogeneous behavior among people, as reflected on their temperature preferences. These findings are important to characterize the thermal comfort preferences of occupants and their interactions with the building systems, which are top influential factors of residential heating energy loads.

The interest of using this method goes beyond the results that are presented here. The study provided insights that will help direct future research in identifying ways to estimate temperature setpoints, assess indoor thermal comfort and consequently improve urban building energy models. Thus, customized messages and solutions for occupants can be developed. Furthermore, it provided better understanding of the Danish building stock and its occupants, which is crucial in building energy demand calculations. Thus, the given methodology can be applied to large scale smart meter datasets to acquire building-related and thermal comfort-related characteristics. In addition, it can provide valuable information to utilities to further optimize their network and apply advanced technologies (i.e. virtual storage on district heating network) based on the temperature setpoint distribution of the district. The proposed simplified approach opens up new possibilities of building performance analysis at urban scale. Next step would be to couple this energy dataset with internal temperature recordings, which would enable the validation of the estimated variables at urban scale. In that way, the current findings could be utilized to challenge assumptions used in Scandinavian housing stock models regarding heating patterns and mostly, temperature setpoints throughout the heating period.

8 Acknowledgements

This work was undertaken as a part of the CITIES (Centre for IT-Intelligent Energy Systems in cities) project number DSF1305-00027B funded by the Danish Strategic Research Council. Special thanks should be given to Aarhus AffaldVarme, who enabled the distribution of the smart meter data to the author.

9 References

1. International Energy Agency (2001). Annex 31 - Stock aggregation: methods for evaluation of the environmental performance. International Initiative for a Sustainable Built Environment (iiSBE).
2. A.T. Booth, R. Choudhary, D.J. Spiegelhalter (2012). Handling uncertainty in housing stock models. Building and Environment 48, pp. 35-47.
3. L. Swan, V. Ugursal (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renewable and Sustainable Energy Reviews 13, pp. 1819-1835.
4. M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic (2010). A review of bottom-up building stock models for energy consumption in the residential sector. Building and Environment 45, pp. 1683-1697.

5. T. Hong, S.C. Taylor-Lange, S. D'Oca, D. Yang, S. P. Corgnati (2016). Advances in research and applications of energy-related occupant behavior in buildings. Energy and Buildings 116, 694–702.
6. K. Sun, T. Hong (2017). A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. Energy and Buildings 146, pp. 383-396.
7. W. O'Brien, I. Gaetani, S. Carlucci, P-J. Hoes, J. L.M. Hensen (2017). On occupant-centric building performance metrics. Building and Environment 122, pp. 373-385.
8. ANSI/ASHRAE, ANSI/ASHRAE Standard 55: Thermal Environmental Conditions for Human Occupancy, ANSI/ASHRAE, 2004.
9. Z. O'Neill, F. Niu (2017). Uncertainty and sensitivity analysis of spatio-temporal occupant behaviors on residential building energy usage utilizing Karhunen-Loeve expansion. Building and Environment 115, 157-172.
10. M. Bonte, F. Thellier, B. Lartigue (2014). Impact of occupant's actions on energy building performance and thermal sensation. Energy Build 76, 219–227.
11. V. Fabi, R.V. Andersen, S.P. Corgnati (2013). Influence of occupant's heating set-point preferences on indoor environmental quality and heating demand in residential buildings. HVAC&R Research 19 (5), 635–645.
12. S. D'Oca, V. Fabi, R.K. Andersen, S.P. Corgnati (2014). Effect of thermostat and window opening occupant behavior models on energy use in homes. Building Simulation 7, 683–694.
13. The European Parliament and the Council of European Union (2009). Directive 2009/72/EC concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC. Official Journal of the European Union.
14. The European Parliament and the Council of European Union (2009). Directive 2009/73/EC concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC. Official Journal of the European Union.
15. European Commission (2014). Benchmarking smart metering deployment in the EU-27 with a focus on electricity. Report from the Commission, Brussels.

16. A. Kipping, E. Trømborg (2016). Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data. Energy and Buildings 118, pp. 350-369.
17. Z. Yu (2012). Mining Hidden Knowledge from Measured Data for Improving Building Energy Performance. PhD Thesis, Concordia University.
18. C.F. Reinhart, C. Cerezo Davila (2016). Urban building energy modeling – A review of a nascent field. Building and Environment 97, pp. 196-202.
19. J.N.B. Fonseca, M.J.N. Oliveira Panão (2017). Monte Carlo housing stock model to predict the energy performance indicators. Energy and Buildings 152, pp. 503-515.
20. J. Zico Kolter, J. Ferreira (2011). A Large-scale Study on Predicting and Contextualizing Building Energy Usage. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, USA.
21. C. Becker, L. Sadamori, T. Staake, S. Santin (2014). Revealing Household Characteristics from Smart Meter Data. Energy, Vol 78, pp. 397-410.
22. E. Hirst, R. Goeltz, D. White (1986). Determination of household energy using 'fingerprints' from energy billing data. Energy Research 10, pp. 393–405.
23. S. Kelly, M. Shipworth, D. Shipworth, M. Gentry, A. Wright, M. Pollitt, D. Crawford-Brown, K. Lomas (2013). Predicting the diversity of internal temperatures from the English residential sector using panel methods. Applied Energy 102, pp. 601-621.
24. K. Vadodaria, D.L. Loveday, V. Haines (2014). Measured winter and spring-time indoor temperatures in UK homes over the period 1969-2010: A review and synthesis. Energy Policy 64, pp. 252-262.
25. T. Oreszczyn, S.H. Hong, I. Ridley, P. Wilkinson, The Warm Front Study Group (2006). Determinants of winter indoor temperatures in low income households in England. Energy and Buildings 38, pp. 245-252.
26. I.G. Hamilton, A. O'Sullivan, G. Huebner, T. Oreszczyn, D. Shipworth, A. Summerfield (2017). Old and cold? Findings on the determinants of indoor temperatures in English dwellings during cold conditions. Energy and Buildings 141, pp. 142-157.

27. B. Lin, Z. Wang, Y. Liu, Y. Zhu, Q. Quyang (2016). Investigation of winter indoor thermal environment and heating demand of urban residential buildings in China's hot summer – Cold winter climate region. Building and Environment 101, pp. 9-18.

28. V. Cheng, K. Steemers (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. Environmental Modelling & Software 26, pp. 1186-1198.

29. S. Natarajan, G. J. Levermore (2007). Predicting future UK housing stock and carbon emissions. Energy Policy 35, pp. 5719-5727.

30. S. K. Firth, K. J. Lomas, A. J. Wright (2010). Targeting household energy-efficiency measures using sensitivity analysis. Building Research & Information 38, pp. 25-41.

31. B. R. Anderson, P. F. Chapman, N. G. Cutland, C. M. Dickson, G. Henderson, J. H. Henderson, P. J. Iles, L. Kosmina, L. D. Shorrock (2002). BREDEM-12. Model description, 2001 update. BRE, Department for Environment, Food and Rural Affairs.

32. The Chartered Institution of Building Services Engineers (CIBSE) (2006). Degree-days: theory and application. TM41: 2006, London.

33. A. Kipping, E. Trømborg (2016). Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data. Energy and Buildings 118, pp. 350-369.

34. M. Kohler, N. Blond, A. Clappier (2016). A city scale degree-day method to assess building space heating energy demands in Strasbourg Eurometropolis (France). Applied Energy 184, pp. 40-54.

35. A. Durmayaz, M. Kadioglu, Z. Sen (2000). An application of the degree-hours method to estimate the residential heating energy requirement and fuel consumption in Istanbul. Energy 25, pp. 1245-1256.

36. D. Lindelöf (2017). Bayesian estimation of a building's base temperature for the calculation of heating degree-days. Energy and Buildings 134, pp.154-161.

37. A.R. Day, I. Knight, G. Dunn, R. Gaddas (2003). Improved methods for evaluating base temperature for use in building energy performance lines. Building Services Engineering Research and Technology 24, pp. 221-228.

38. S. Aggerholm, K. Grau (2016). Bygningers energibehov. SBi-Anvisning 213. Danish Building Research Institute.

39. C.E. Kontokosta, C. Tull (2017). A data-driven predictive model of city-scale energy use in buildings. Applied Energy 197, pp. 303-317.
40. DIVA for Rhino plug-in. [Online] Available at: http://www.solemma.net/Diva.html (accessed September 5, 2017).
41. C. F. Reinhart, J. Wienold (2011). The daylighting dashboard. A simulation-based design analysis for daylit spaces. Building and Environment 46, pp. 386-396.
42. C. F. Reinhart, O. Walkenhorst (2001). Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds. Energy and Buildings 33, pp. 683-697.
43. G. Ward, R. Shakespeare (1998). Rendering with RADIANCE. The Art and Science of Lighting Visualization, Morgan Kaufmann Publishers.
44. Statistics Denmark. [Online] Available at: http://www.dst.dk/en/Statistik/emner/levevilkaar/boligforhold/boliger (accessed May 15, 2017).
45. Danish Energy Agency (2015). Regulation and planning of district heating in Denmark. [Online] Available at:https://ens.dk/sites/ens.dk/files/Globalcooperation/regulation_and_planning_of_district_heating_in_denmark.pdf
46. K. Wittchen, J. Kragh (2016). Danish building typologies and building stock analyses. Danish Building Research Institute, Aalborg University, SBi 2016:18.
47. European Union (2014). Inclusion of New Buildings in Residential Building Typologies. Steps Towards NZEBs Exemplified for Different European Countries. EPISCOPE Synthesis Report No. 1. [Online]. Available at: http://episcope.eu/fileadmin/episcope/public/docs/reports/EPISCOPE_SR1_NewBuildingsInTypologies.pdf [Accessed October 5, 2017).
48. S. Aggerholm (2013). Cost-optimal levels of minimum energy performance requirements in the Danish Building Regulations. SBi 2013:25. Danish Building Research Institute, Aalborg University.
49. DS418 : 2011. Calculation of heat loss from buildings. Danish Standard.
50. P.O. Fanger (1970). Thermal Comfort-Analysis and Applications in Environmental Engineering, in: C.D.T. Press. (Ed.).

51. ISO 7730 (2005). Ergonomics of the thermal environment - Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. International Standard.
52. P. Gianniou, A. Heller, P. Sieverts Nielsen, C. Rode (2016). Identification of Parameters Affecting the Variability of Energy Use in Residential Buildings. Proceedings of CLIMA Conference, Aalborg, Denmark.
53. Bygnings- og Boligregistret (BBR). Danish Building Register. [Online] Available at: https://ois.dk/ (accessed May 20, 2017).
54. K. B. Wittchen, J. Kragh (2012). Danish building typologies. Participation in the TABULA project, Hørsholm, Denmark: SBi, Danish Building Research Institute, Aalborg University.
55. The Danish Ministry of Housing, U. a. R. A., 2006. Boligejer.dk. [Online] Available at: http://boligejer.dk/ [Accessed July 1, 2016].
56. Styrelsen for Dataforsyning og Effektivisering. Kortforsyningen. [Online] Available at: https://kortforsyningen.dk/ (accessed July 15, 2017).
57. K.B. Wittchen, L. Mortensen, S.B. Holøs, N. F. Björk, S. Vares, T. Malmqvist (2012). Building typologies in the Nordic countries, Identification of potential energy saving measures. Danish Building Research Institute, SBi2012:04.
58. P. Gianniou, A. Heller, P.S. Nielsen, K. Negendahl, C. Rode (2015). Aggregation of building energy demands for city-scale models. In Proceedings of Building Simulation 2015, Hyderabad, India.
59. X. Liu, P.S. Nielsen, A. Heller, P. Gianniou (2017). SciCloud: A Scientific Cloud and Management Platform for Smart City Data. In 2017 28th International Workshop on Database and Expert Systems Applications (DEXA), pp. 27-31. IEEE. DOI: 10.1109/DEXA.2017.22