A regression approach for assessment of building energy performance

Jimmy Vesterberg
# Table of Contents

Table of Contents i  
Abstract iii  
Sammanfattning v  
Preface vii  
List of original publications vii  
Abbreviations ix  
1. General introduction 1  
   1.1 Scope of the thesis 2  
2. Data-driven and forward modeling 3  
   2.1 IDA-ICE and validation 3  
3. Sustainable Ålidhem project 5  
   3.1 Studied buildings 5  
4. Data 7  
5. Analysis of measured data 9  
   5.1 Derivation and justification of used regression model 9  
   5.2 Potential uses of the extracted parameters 12  
6 Measurement uncertainty 13  
   6.1 Data smoothing techniques 14  
   6.2 Discussion of uncertainties in the intercept 16  
7. Concluding remarks and summary of papers 17  
   7.1 Paper I 17  
   7.2 Paper II 18  
   7.3 Paper III 18  
   7.4 Future research 19  
References 21
Abstract

Reliable evaluation methods is needed to ensure that investments in energy conservation measures (ECMs) and the construction of new energy efficient buildings lives up to the promised and expected performance.

This thesis presents and evaluates a regression method for estimation of influential building parameters: transmission losses above ground (including air leakage), ground heat loss, and overall heat loss coefficient.

The analysis is conducted with separately metered electricity, heating and weather data using linear regression models based on the simplified steady-state power balance for a whole building.

The evaluation consists of analyzing the robustness of the extracted parameters, their suitability to be used as input values to building energy simulations (BES) tools. In addition, differences between uncalibrated and calibrated BES models are analyzed when they are used to calculate energy savings. Finally the suitability of using a buildings overall heat loss coefficient as a performance verification tool is studied.

The presented regression method exhibits high robustness and good agreement with theory. Knowledge of these parameters also proved beneficial in BES calibration procedures as well as in performance verifications. Thus, the presented method shows promising features for reliable energy performance assessments of buildings.
Sammanfattning

Tillförlitliga utvärderingsmetoder behövs för att säkerställa att investeringar i energibesparande åtgärder och uppförandet av nya energieffektiva byggnader håller den utlovade och förväntade prestandan. Denna avhandling presenterar och utvärderar en regressions metod för att skatta en byggnads termiska prestanda parametrar: transmissions förluster mot uteluft (inklusive luftläckage), värmeförlust till mark och en byggnads totala förlustfaktor ($K_{tot}$) innehållande transmission (mot uteluft) och ventilations-förluster.

Analysen genomförs med separat uppmätt energianvändning av el, värme och väderdata med hjälp av linjära regressions-modeller baserade på förenklade effektbalanser för en byggnad.

Utvärderingen består av att analysera robustheten i de extraherade parametrarna, deras lämplighet att användas som ingångsvärden i byggnads energi-simuleringer (BES) program. Vidare analyseras fördelen med att använda kalibrerade BES modeller. Slutligen analyseras lämpligheten att använda $K_{tot}$ som ett komplement till den allmänt vedertagna svenska energiprestanda indikatorn (kWh/m²,år).

Den presenterade regressions-metoden uppvisar hög robusthet och god överensstämmelse med teori. Kunskap om dessa parametrar visade sig också vara fördelaktigt i kalibrering av BES modeller samt i energi-prestanda utvärderingar. Således visar den presenterade metoden lovande egenskaper att användas för tillförlitliga energi-prestanda bedömningar av byggnader.
Preface

This thesis is based on research work carried out at the department of Applied Physics and Electronics, Umeå University during the years 2010-2014. The research work was funded by the Industrial Doctoral School at Umeå University and Bostaden Umeå AB. I would like to gratefully acknowledge the financial support as well as to thank my supervisors Professor Thomas Olofsson and Docent Staffan Andersson for their invaluable advice and support throughout the study. I would also like to express my sincere thanks to colleagues at Applied Physics and Electronics. Finally, I offer my thanks to all the talented people at Bostaden AB, especially Berndt Elstig, Fredrik Westerlund, Folke Parkle, Niklas Broddeskog, Christer Nyhlen and Ann-Sofi Tapani for the fruitful cooperation.

List of original publications

The thesis consists of a brief introduction to the field, a summary and a short extension of the following appended papers, referred to by their Roman numerals.


III. J. Vesterberg, Utilizing a regression approach for troubleshooting energy performance of Swedish buildings, Manuscript.
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_o$</td>
<td>Outdoor air temperature</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Indoor air temperature</td>
</tr>
<tr>
<td>$T_s$</td>
<td>Air handling unit supply temperature</td>
</tr>
<tr>
<td>$T_x$</td>
<td>Air temperature directly after the heat exchanger</td>
</tr>
<tr>
<td>$P_G$</td>
<td>Ground heat loss</td>
</tr>
<tr>
<td>$P_{dyn}$</td>
<td>Released or stored power by the mass of a building</td>
</tr>
<tr>
<td>$P_{rad}$</td>
<td>Supplied heating power to radiator system</td>
</tr>
<tr>
<td>$P_{vent}$</td>
<td>Ventilation losses, controlled and uncontrolled</td>
</tr>
<tr>
<td>$P_{airh}$</td>
<td>Supplied heating power to ventilation system</td>
</tr>
<tr>
<td>$P_{elec}$</td>
<td>Total supplied electrical power</td>
</tr>
<tr>
<td>$P_p$</td>
<td>Internal heat gained from persons</td>
</tr>
<tr>
<td>$P_{sun}$</td>
<td>Heat gain from the sun</td>
</tr>
<tr>
<td>$P_{dhw}$</td>
<td>Heat gain due to domestic hot water usage</td>
</tr>
<tr>
<td>$Q_s$</td>
<td>Ventilation supply airflow</td>
</tr>
<tr>
<td>$Q_e$</td>
<td>Ventilation exhaust airflow</td>
</tr>
<tr>
<td>$Q_L$</td>
<td>Infiltration airflow</td>
</tr>
<tr>
<td>$C_p$</td>
<td>Specific heat capacity of air</td>
</tr>
<tr>
<td>$A$</td>
<td>Total surface envelope area above ground</td>
</tr>
<tr>
<td>$U_t$</td>
<td>Total heat transmission coefficient above ground</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Part of electricity use that contributes to heating (gain factor)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Density of air</td>
</tr>
<tr>
<td>$K_{tot}$</td>
<td>Mean thermal transmittance through building envelope to the outdoor air by transmission and ventilation</td>
</tr>
<tr>
<td>$S_b$</td>
<td>Standard error of the slope, regression coefficient</td>
</tr>
<tr>
<td>$S_a$</td>
<td>Standard error of the intercept, regression coefficient</td>
</tr>
<tr>
<td>$S_{YX}$</td>
<td>Standard error of the regression model</td>
</tr>
</tbody>
</table>
1. General introduction

Quantification of a building's energy use is the basis to make any decision for enhancing energy efficiency. According to ASHRAE Handbook [1] modeling of buildings energy use can be classified into two different approaches: forward and data-driven methods. The forward modeling approach includes Building Energy Simulation (BES) models where the objective is to predict the output based on detailed knowledge of the physical system to be modeled.

BES programs have also been accepted as powerful tools for analyzing building energy performance [2]. Its role in energy performance assessment has increased significantly in the last two decades [3] to the extent that it has now become a mainstream approach for quantifying expected energy savings due to various proposed energy conservation measures (ECMs) [4]. However, it is still often a challenge to perform accurate simulations for existing buildings due to the extensive amount of input parameters required and site inspections needed to confirm those inputs.

In Sweden, simulation competitions have been conducted based on blueprints, standardized user parameters and spot measurements. The participants used large variations in calculation methods and the models showed poor agreement with measured data even after the models was updated with audit stage parameters [5].

It is generally accepted that calibration of BES models is needed for more reliable predictions. BES models calibrated with measured data is classified into the data-driven approach according to [1] but can also be classified as a hybrid approach according to Wang et al [6]. Although there are standard criteria's given in the guidelines [8] and [9] for determining when a BES model can be considered calibrated, no consensus guidelines currently exists on how to perform a BES calibration [10] making this an open research question for the future.

In data-driven approaches (e.g. regression models) the output and input have been measured and the data is used to define a physical description of the system. Assessments that are based on data-driven models are often done by identifying the parameters of the used model. Of the many data driven models used and developed in the past, traditional energy signature (ES) models are perhaps the most widely used due to its simplicity. The disadvantages include insensitivity to dynamic effects and to variables other than temperature (e.g. humidity and solar gains). Hence, the ES method have been considered to be a good tool for rough estimations but less accurate then more complex data driven methods such as neural network or multivariate models [1].
Earlier ES based studies generally use a subset of the supplied energy for heating as a response variable plotted against the outdoor temperature as an independent variable in the regression models e.g. [11] [12] [13] [14] [15] [16] [17] [18] and [19]. The heat from the household and property electricity, occupants, sun etc. is usually not considered. Moreover, the used data is often of monthly resolution due to the easy accessibility of energy bill data. Therefore, studies which utilize data of high time-resolution (daily or hourly) and/or sub-metered are scarce. However, in the near future it is believed that detailed data will be more commonly available which presents new opportunities and challenges on how to best utilize this data.

1.1 Scope of the thesis

The aim of this thesis is to investigate the practical applications of a refined ES approach. The regression method presented in this thesis is based on detailed data and is able to quantify building thermal performance parameters which have significant influence on a building's heating load. That is, transmission losses above ground (including air leakage), ground heat loss, and overall heat loss coefficient. Parameters which are difficult to directly measure but is expected to be very useful in building performance analysis, which is discussed in section 5.2. The specific objectives of this thesis are:

- To analyze the robustness of regressed building thermal performance parameters.
- To investigate the suitability of using these thermal performance parameters as input values to BES tools.
- To analyze if calibrated BES models improves the prediction accuracy when implementing ECMs.
- To analyze the suitability of using the overall heat loss coefficient as a tool for performance verification of buildings.
2. Data-driven and forward modeling

A difficulty when using data-driven models for parameter identification is that the true values of the extracted parameters are unknown. One indirect way to validate the results is to use the parameters as feedback to a commercial BES tool such as IDA Indoor Climate and Energy (IDA-ICE) and compare predicted versus measured data when the BES tool is adjusted to coincide with the regressed estimations. In Paper I and II, the results of the simulation accuracy is given when the parameters were used to define IDA-ICE models of two multifamily buildings, building no.1 and 2 described in section 3.

2.1 IDA-ICE and validation

In the study by Crawley et al [20] the common features and capabilities of twenty major building energy simulation programs such as DOE-2 [21], ESP-r [22], EnergyPlus [23], IDA ICE [24], and TRNSYS [25] are described. As one of the widely used BES tools, IDA Indoor Climate and Energy, (IDA-ICE) has been under development since the 80s. IDA-ICE was originally developed by the Division of Building Services Engineering at the Royal Institute of Technology and the Swedish Institute of Applied Mathematics [26]. Today it is maintained and supported commercially by EQUA [24] in Stockholm. The accuracy of IDA-ICE has been analyzed in several studies over the years. For instance in the studies [27] and [28] it is showed that IDA-ICE perform well in comparison with the other previous mentioned BES programs.

Several empirical validation studies have also been conducted. In [29] Travesi et al compared several building energy simulation programs output against measured data, including IDA-ICE. It was concluded that “the computer models were successful in calculating the measured values of several parameters of interest in building energy simulation. The calculated values generally fall within the range of experimental uncertainty of the measured data”. Loutzenhiser et al [30] conducted a similar study where it was concluded that “All simulation results corresponded well with the experimental data”. A more extensive review of IDA-ICE validation studies can be found in the doctoral dissertation thesis [26]. Due to the proven accuracy of IDA-ICE, its user friendly interface as well as the fact that it is under continuously development, IDA-ICE was chosen as an analysis tool in this thesis.
3. Sustainable Ålidhem project

The research in this thesis is connected to the refurbishment project “Sustainable Ålidhem” located in Ålidhem, Umeå, Sweden. The Sustainable Ålidhem project is a refurbishment effort of 21 multifamily buildings constructed during the late 60s and early 70s. These buildings had before the refurbishment a relative high specific energy demand of 200 kWh/m² and year (as defined by the Swedish building regulations [31]).

In addition, within the project, four new multifamily buildings (Building A-E) have been constructed with a specific energy performance goal, of 65 kWh/m²/yr. The municipal housing company, AB Bostaden is conducting the project supported by the Swedish governmental body “The Delegation for Sustainable Cities” [32].

The overall energy efficiency goal, within the project is a 40-50% reduction of supplied energy for domestic hot water, building electricity and space heating. In order to evaluate if these goals were feasible, a measurement system was installed in a pilot building (Building no.1) and in a neighboring building used as a reference (Building no.2). The evaluation was conducted by comparing the post-retrofit performance of Building no.1 with the performance of Building no.2 when it was kept in its initial state (a comparison possible to very high building similarities initially).

The measurement setup and experiences from the evaluation of Building no.1 and 2 have given valuable insights within the project as well as made a number of awards possible e.g. Sustainable Energy Europe Awards 2013 and National Energy Globe Award 2014. A more complete description of the project is given at the property holder AB Bostadens webpage [33].

3.1 Studied buildings

The first building to be refurbished within the project, Building no.1 and Building no.2 have been used as case studies. Together with the four newly built multifamily buildings (Buildings A-E) constructed during 2011/2012. Paper I and II are based on monitored data from Building no.1 and 2, whereas paper III is based on data collected from Buildings A-E. These case studied buildings are described more in the appended papers.
4. Data

The supplied energy to the studied buildings was measured at hourly intervals and retrieved from the local energy company (Paper I and II) and from the property holder (paper III). The buildings use District Heating (DH) for space heating and domestic hot water preparation. Electricity is used for household equipment’s (tenants) and to operate building electrical systems. For the analysis in paper I and II, the supplied DH for space heating was monitored in three parts: supplied DH to the radiators and air heaters, sub metering of supplied DH to the air handling units and DH for domestic hot water circulation. The use of household and property electricity was monitored at whole building level and the indoor temperatures were measured in half of the apartments as well as air handling unit air supply temperatures.

The analysis conducted in paper III (the newly constructed buildings) was based on data from a similar measurement setup with the difference of that the supplied DH to the air handling units and domestic hot water circulation was lacking. All measurements that were made by the author (outdoor, indoor, air handling unit temperatures and outdoor relative humidity) were conducted with the previous models in the RTR-500 tiny-loggers serie, described at the manufacturer’s webpage [34]. The used loggers were manufacturer-calibrated with a specified average error of ±0.3°C and ±5% RH. The data measured by the author was logged and stored at a time step of 15 minutes.
5. Analysis of measured data

In this chapter an extension of the physical derivation of the model used for estimation of building parameters in paper I and III is presented, together with a discussion of potential use of the regressed parameters.

5.1 Derivation and justification of used regression model

Assessments that are based on data-driven models are either done by parameter identification or with energy predictions e.g. [35]. To obtain parameter estimates that allow direct physical interpretation, one needs to formulate models based on physical principles such as the buildings power balance.

This can be done by defining a system (a control volume) that includes the entire building. Then the change in the total energy of the control volume $\Delta P_{CV}$ is equal to the difference between the total energy entering and leaving the system, that is,

$$\Delta P_{CV} = P_{\text{entering}} - P_{\text{leaving}} = P_h + P_p + \alpha P_{\text{elec}} + P_{\text{sun}} - P_{\text{tr}} - P_{\text{vent}} - P_G$$ (1)

Where, $P_h$ and $\alpha P_{\text{elec}}$ are the separately metered whole building use of DH and electricity for space heating. $P_p$ and $P_{\text{sun}}$ are the free heatgains due to occupancy and solar irradiation. Lastly, $P_{tr}$, $P_G$ and $P_{\text{vent}}$, are heat losses due to transmission through the building envelope to the outside air, to the ground and heat losses due to air exchange respectively.

When the time scale under study is long enough to diminish the thermal lag effect the system can be considered in a steady state and the left hand side of eq. 1 ($\Delta P_{CV}$ denoted $P_{\text{dyn}}$ in paper I) yields zero.

The approach used to minimize the contribution from the dynamic effects ($P_{\text{dyn}}$) have been to preprocess the data by averaging over a time period longer than the estimated time constant of the analysed buildings. The time constant for these types of buildings in the area were experimentally estimated in the thesis [36]. The contributions from the sun, $P_{\text{sun}}$, have been assumed to be negligible as data was collected during a time period when the global solar radiation was small.

The internal heat gain due to occupancy, $P_p$, is very difficult to measure and have been estimated from public records and guideline [37]. Based on this an average of 80 W per occupant was assumed for emitted body heat during an assumed attendance time of 14 hours per day. With known apartment sizes the number of tenants was calculated in paper III with standardized values in [37] and received from public records in paper I. With the above assumptions and given that supplied power for space heating of a
building is related to the overall heat loss coefficient, $K_{tot}$, which is the mean thermal transmittance through building envelope to the outdoor air by transmission and ventilation eq. 1 may be written as

$$P_\text{h} + P_\text{p} + \alpha P_{\text{elec}} = P_\text{tr} + P_{\text{vent}} + P_G = K_{\text{tot}}(T_1 - T_o) + P_G \quad (2)$$

Since all parameters on the left side of eq.2 were known in paper III, linear regression gave estimates of $K_{tot}$, and $P_G$, under the assumption that $P_G$ was fairly constant during the analysed time period.

This approach has been previously explored by Sjögren et al [38] [39] and Andersson et al [40] and formed the basis of the analysis in paper III.

In paper I, the gained heat due to constant circulation of domestic hot water usage, $P_{\text{dhw}}$ was metered separately and included in the supplied DH to the radiators, $P_{\text{rad}}$. $P_{\text{dhw}}$ is however typically small in new buildings and therefore assumed negligible in paper III. For the older buildings studied in paper I and II, $P_{\text{dhw}}$ was significant (in the order of 3 kW, before the refurbishment) and therefore it was continuously monitored and included in $P_{\text{rad}}$ in the power balance. In addition, submetering was done of the air handling units, enabling $P_\text{h}$ in eq.1 to be divided as:

$$P_\text{h} = P_{\text{rad}} + P_{\text{airh}} \quad (3)$$

In addition, the air flows and associated temperatures were monitored in paper I, making it possible to separate $K_{tot}$ into its transmission and ventilation parts. Focusing on the exhaust air stream $Q_e$, $K_{tot}$ may be expressed as,

$$K_{\text{tot}}(T_1 - T_o) = AU_t(T_1 - T_o) + Q_e(T_1 - T_o)\rho C_p \quad (4)$$

For a building, with a balanced mechanical ventilation system, the exhaust ventilation airflow $Q_e$ is balanced by the supply and infiltration airflow $Q_s$ and $Q_L$ respectively. Rewriting the ventilation losses, $Q_e(T_1 - T_o)\rho C_p$ in eq.4 with $Q_e=Q_s+Q_L$ gives

$$Q_e(T_1 - T_o)\rho C_p = Q_s(T_1 - T_o)\rho C_p + Q_L(T_1 - T_o)\rho C_p \quad (5)$$

The right side of eq.5 may be divided for a building with balanced mechanical ventilation system, without heat recovery (building no. 2) as

$$Q_e(T_1 - T_o)\rho C_p = Q_s(T_s - T_o)\rho C_p + Q_s(T_1 - T_s)\rho C_p + Q_L(T_1 - T_o)\rho C_p \quad (6)$$

Where, the first term accounts for the supplied heat to the ventilation air heater ($P_{\text{airh}}$). The second term accounts the heat transfer associated with the
convergence of the air supply temperature \( T_s \) and indoor temperature, \( T_i \) and the last term accounts for the heat loss due to air leakage.

For a building, with a mechanical ventilation system with heat recovery (building no. 1), the right side of eq.5 may be divided in a similar way as

\[
Q_s \left( T_s - T_o \right) \rho C_p = Q_s \left( T_s - T_o \right) \rho C_p + Q_s \left( T_i - T_o \right) \rho C_p + Q_s \left( T_i - T_s \right) \rho C_p
\]

where the first term accounts for the supplied heat to the ventilation air heater \( (P_{airh}) \), which is used if the inlet air temperature after the heat exchanger, \( T_x \) is lower than the set air supply temperature, \( T_s \).

The second and third term describes, as in eq.6, the heat transfer associated with the convergence of \( T_s \) and \( T_i \) and heating of air leakage, combining eq. 2, 3, 4 and 6 for the case of building no. 2 yields,

\[
P_{rad} + P_{airh} + P_{p} + \alpha P_{elec} = \left[ A U_t + Q_L \rho C_p \right] (T_i - T_o) + P_G = [A U_t + Q_L \rho C_p] \left[ T_i - T_o \right] + Q_s \left[ T_s - T_o \right] + (T_i - T_s) + P_G
\]

In a similar way, rewriting the power balance for the case of Building no. 1, now utilizing eq. 2, 3, 4 and 7 we get

\[
P_{rad} + P_{airh} + P_{p} + \alpha P_{elec} = \left[ A U_t + Q_L \rho C_p \right] (T_i - T_o) + P_G = [A U_t + Q_L \rho C_p] \left[ T_i - T_o \right] + Q_s \left[ T_s - T_o \right] + (T_i - T_s) + P_G
\]

In order to estimate the transmission losses above ground including air leakage \( (A U_t + Q_L \rho C_p) \) and ground heat loss \( (P_G) \) from a regression analysis of measured data, eq.8 and 9 have to be rewritten.

Utilizing the fact that \( P_{airh} = Q_s \left( T_s - T_o \right) \rho C_p \) for Building no. 2 and \( P_{airh} = Q_s \left( T_s - T_o \right) \rho C_p \) for Building no. 1, we finally obtain

\[
P_{rad} + P_{p} + \alpha P_{elec} - Q_s \left( T_i - T_s \right) \rho C_p = \left( A U_t + Q_L \rho C_p \right) (T_i - T_o) + P_G
\]

A linear regression of eq. 10 yields estimates of \( (A U_t + Q_L \rho C_p) \) and \( P_G \), under the assumption that \( P_G \) is fairly constant during the analysed time period. The reason for rewriting eq.2 to the form of eq.10 is the possibility to estimate the impact from ECMs which aims to improve a buildings envelope. A condition for this however, is that the method is robust between different years. The robustness was studied in paper I. In addition, with good knowledge of the transmission losses, a significant part of the process of defining a satisfactory BES model for a building has been reached. Therefore, the extracted parameters from eq.10 were used as feedback to IDA-ICE models to analyze if the regression method also could be used as a relative simple BES calibration method.
5.2 Potential uses of the extracted parameters

Expected uses of the estimated parameters include: 1) input values to BES tools for existing buildings 2) building performance verification 3) evaluation of savings due to ECMs 4) detection of operational changes in energy management systems with continuous monitoring in large building stocks.

The evaluation of point 1 and 2 are described in detail in the appended papers. Point 3 is not explicitly investigated but the improvement of transmission losses above ground (including air leakage) due to implementation of ECMs in building no 1 can easily be assessed by comparing the difference between the slopes presented in paper I. The comparison is possible due to identical building designs as well as heating and ventilation systems before the refurbishment of building no 1. These similarities were also confirmed through a comparison of the heat demand of building no 1 and 2 a couple of months prior to the refurbishment.

The change in slopes, \((U_t+Q_i\rho C_p)\) between building no. 1 (retrofitted) and building no. 2 used as a reference can be estimated to \(0.67-0.51 \pm \sqrt{0.04^2 + 0.02^2} = 0.16 \pm 0.04 \text{ kW/K} \) (standard errors added in quadrature) during the first analyzed year and \(0.18 \pm 0.04 \text{ kW/K} \) the following year. As seen in the results, the uncertainty is fairly large (on average ca. 26%). This is due that the implemented ECMs affecting \((U_t+Q_i\rho C_p)\) in building no 1 were fairly moderate, the uncertainties in both building estimations of \((U_t+Q_i\rho C_p)\) is considered (since a difference is evaluated) and lastly the choice of pre-processing method of measured data.

In section 6.1 the influence of pre-processing methods on the regression output is briefly discussed.
6 Measurement uncertainty

As described in section 4 the used data in this thesis originates from electricity, DH and temperatures measurements. Since, the errors from electricity meters and air temperature loggers are small (in the order of ±0.2% [41] and ±0.3°C [34] respectively. The largest measurement error is believed to be associated with the DH energy measurements where commercial flow meters of accuracy class 2 can have a maximum error of 5% in confinement with the European Norm for energy metering (EN1434). This is relative large error compared to the accompanying platinum temperature sensors which is allowed to have a temperature difference error of maximum ±1.5°C [42]. However, both the flow and temperature errors decrease with increased flow and temperature difference. This results in smaller DH errors during the heating season, during which the regression analysis is conducted. As the total measurement error is the sum of all the above described uncertainties a stringent error analysis is not trivial.

One indirect way to estimate the impact of the measurement error is to analyze its influence on the regression analysis. All measurement errors consist of two components: systematic error and random error [43] and the amount of random error are inversely related to the degree of reliability (consistency) of the measurement equipment [44]. In other words, the more random error the less consistency will be seen in the measured data. Thus, if linear regression analysis is performed on data with a significant amount of random error, the results (slope and intercept) would be expected to be inconsistent between different analysed periods. Systematic or constant errors on the other hand would mainly be limited to influencing the intercept of the regression line.

In paper I the robustness of the slopes and intercepts were investigated for Building no. 1 and 2. It was found that the linear models showed a small year-to-year variation in the slopes, less than 2.0 %. In addition, for the same periods, the physical interpretations of the intercepts (ground heat loss) agreed well with calculations in accordance with the European standard [45], indicating that measurement errors had small impact on the analysis.

In paper I and III the uncertainties was expressed with the standard error of the regression coefficients, for the slope ($S_b$) and the intercept ($S_a$), which is a measure of the precision of the estimate of the coefficients [46]. The smaller the standard error, the more precise is the estimate assumed to be. However it should be noticed that this statistical measure is dependent on the number of data points (N) used in the regression analysis and hence dependent on the choice of pre-processing method for reducing the thermal dynamic effects. This dependency is demonstrated in the following section for the buildings studied in paper III.
6.1 Data smoothing techniques

The reason for preprocessing the measured data before the regression analysis is to smooth the noise in the data due to occupants, stored thermal energy and random errors. A typical smoothing approach is to average the data over time-steps longer than the estimated time constant for the analyzed building. This approach is used for instance in [40] [47] and was also used in paper I and III. Many commercial buildings operate however in occupied (weekdays) and unoccupied modes (weekends) in such cases, even if the regression is based on all the supplied energy from the grid as a response variable, the swings in internal heat gains from occupants can motivate weekly time aggregation as was advocated in [48].

For the multifamily buildings, studied in this thesis, the swings in internal loads due to body heat between weekends and weekdays are assumed to be negligible. Andersson et al [40] suggested that pairing of data from days with similar magnitude of decrease and increase of outdoor temperatures might work to reduce the thermal dynamics. One of the benefits with this approach is the reduction of data is less compared to the ordinary averaging interval approach. Another alternative to keep a significant part of the data to the subsequent regression analysis is to use moving average intervals.

To investigate the influence of the chosen data smoothing method on the regression output, K_{tot} is extracted for buildings A-E analysed in paper III based on data pre-processed by moving average intervals and the ordinary average approach. The used data was collected during two months symmetrically around the winter solstice 2013/2014 and the results are shown in table 1 for four days and weekly average intervals.
Table 1. Overall heat loss coefficient and associated standard errors for buildings A-E. The lowercase numbers denotes cases of four and seven days averaging.

<table>
<thead>
<tr>
<th>B.no.</th>
<th>Ordinary average (kW/K)</th>
<th>Moving average (kW/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_{tot4}$</td>
<td>$K_{tot7}$</td>
</tr>
<tr>
<td>A</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>B</td>
<td>1.31</td>
<td>1.32</td>
</tr>
<tr>
<td>CD</td>
<td>1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>E</td>
<td>1.38</td>
<td>1.40</td>
</tr>
<tr>
<td>Mean</td>
<td>1.17</td>
<td>1.18</td>
</tr>
</tbody>
</table>

When comparing moving and traditional averaging the difference in $K_{tot}$ is only of the order of one percent. When comparing the averaging methods sensitivity to the chosen averaging window, from one day to seven days (which due to space constraints couldn’t be shown in table 1) the maximum difference in $K_{tot}$ is about 4% for both methods. A significant difference however, between the methods can be seen in the standard error of the slopes ($S_b$) in table 1 which can be explained from the definition of $S_b$ (taken from [49]).

$$S_b = \frac{1}{\sqrt{N}} \frac{S_{YX}}{S_X} \tag{11}$$

Where, $S_X$ is the standard deviation of the used x-variable, $S_{YX}$ is the standard error of regression model also termed the root-mean-square-error (RMSE) which is a measure of how well the regression model represents the measured data. In eq. 11 it can be seen that low values of N markedly increase the uncertainty in the estimation of the slope ($S_b$). This is a well-known fact in statistics but not found to be discussed in these applications often. As to the knowledge of the author, few regression studies mention the used pre-processing method in detail, (only the study by Andersson et al [40] has been found). In addition, no consensus guideline exists on the optimal choice of data aggregation approach (interval or method) indicating an additional need for research.
6.2 Discussion of uncertainties in the intercept

Linear regression analysis should ideally not be used to predict variables outside the range of the monitored data (extrapolation) as the uncertainties tend to be too large. Thus, to obtain good results of the interpretation of intercept (ground heat loss, $P_G$) it is beneficial with significant temperature variations during the monitored period. $P_G$ is also dependent on the assumptions made on the electrical heat gain factor and body heat from the tenants which adds to the challenge of quantifying $P_G$ with reasonable accuracy.

In paper I the regressed estimations of $P_G$ was however in good agreement with calculations made in accordance with the European Standard [45]. Andersson et al [40] also found good estimations of $P_G$ with the same method as presented here but implemented on buildings without heat recovery.

Nevertheless, during winters with small temperature variations the regressed estimations of $P_G$ should be taken with caution and further research is needed to establish the expected consistency. With that said $P_G$ is often not the primary interest to analyze. In the expected use of the estimated parameters it is only when they are used as input values to BES tools as a supplement to calibration procedures it is advantageous that $P_G$ is reasonable accurate.
7. Concluding remarks and summary of papers

This work presents and evaluates linear regression models based on the simplified steady-state power balance for a whole building. The choice of modeling method was due to its practical advantage before other data-driven methods. Such as: computationally fast, anchored in established physical principles, small influence from the user behavior on the results, easy to use and hence believed to be a good candidate as analytical tool for measurement and verification practitioners.

7.1 Paper I

In this paper the robustness of extracted mean estimates of \((AU_t+Q_L\rho C_p)\) and \(P_G\) based on regression analysis of eq.10 was investigated for Building no. 1 and 2. It was found that the linear models showed a high goodness of fit against the measured data \((R^2\text{-values } > 0.96)\) and small year-to-year variation in \((AU_t+Q_L\rho C_p) < 2.0\%\). The regressed estimates of \(P_G\) were also in good agreement with calculations in accordance with the European Standard [45] during the analysed period.

In addition, the suitability of these parameters to serve as feedback to BES tools as a supplement to calibration procedures was also analysed with (IDA-ICE ver. 4.5) simulation software. It was found that the improvement in simulation accuracy was significant when design stage BES models based on inputs from standardized templet and as-built drawings were compared with BES models calibrated with the regressed parameters and measured audit data.

Mean bias error between monthly simulated and measured energy use (electricity and DH) for the analyzed year decreased from +27.6\% and +15.8\% to −3.5\% and −3.4\% and the cumulative variation of root mean squared error improved from 34.9\% and 25.2\% to 6.0\% and 5.6\% for Buildings no. 1 and 2, respectively. Due to the high quality output from the regression method it was concluded that the method could advantageously be used in a BES calibration process.
7.2 Paper II

In this paper design and calibrated BES models were used to predict energy savings due to different planned ECMs in Building no.2. The overall aim with the study was to analyze prediction differences between the BES models and thereby implicitly, investigating the necessity of using calibrated BES models. In addition to model prediction differences due to individual ECMs the calculations were also compared with the actual outcome in a Building no 1 where the analyzed ECMs have been implemented. The result indicated that a calibrated BES model should be used in order to accurately predict the post retrofit energy demand and energy savings due to heat recovery of the exhaust air. The results suggested further that BES calibration is of minor importance if only investigation of ECMs which targets a buildings transmission loss is of interest.

7.3 Paper III

In this paper $K_{tot}$ (per envelope surface area above ground) was used to interpret differences in measured energy performance (EUI) of four newly built multifamily buildings with the same design performance. In addition, to $K_{tot}$ other simple methods were investigated based on different area normalization techniques.

Since many factors are influencing the EUIs such as: user behavior, shape-factor solar heat gain and the exclusion of household electricity (Swedish regulations), different used area normalization approaches did not significantly reduce the variation. However, through comparison of the buildings $K_{tot}$ (per building envelope surface area above ground) it could be verified, (with 95% confidence), that the differences in EUIs were not due to variations in the buildings thermal performance and HVAC systems.

With the implication that, in order to improve the possibility to meet requirements of equal actual EUIs the developer must either design the buildings to be of the same size or adjust the insulation levels to the selected building geometry (envelope surface area).

Due to the used regression methods simplicity and ability to minimize the users influence on the results it was concluded $K_{tot}$ could be used to refine the discussions in the verification process and help to establish if the property holder has received the performance he/she paid for. The study thus strengthens the findings in Sjögren et al [38] that $K_{tot}$ can be used as a complementary indicator to the traditional EUI key figure.
7.4 Future research

Future studies could advantageously include data from additional buildings to further establish the consistency and accuracy of the presented regression method. In the context of using the method for evaluation of ECMs, further research is needed for quantifying the associated uncertainty e.g. how to handle the influence from the choice of data smoothing approach. Questions relating to the uncertainty are important as it provides the stakeholders the information necessary to assess the risks of a financial investment [50].

Danov et al [51] investigated the sensitivity of $K_{tot}$ to solar heat gain using data from nine public buildings in Spain. In this thesis, the influence from the sun on the extracted parameters was assumed negligible. This approach makes the analysis easier but also limits the method to locations with similar solar conditions. In a similar way as the study by Danov et al, the sensitivity of the extracted parameters in this thesis could be analyzed for conditions when the solar heat gain cannot be assumed negligible.

Using the regression method in conjunction with BES models should be further analysed since typically, regression models are not used to serve as feedback to BES models for calibration purposes. Most studies focus on the inverse situation i.e. fitting regression models to simulated data e.g. [52] [53] [54] and [55]. Moreover, rather complex iterative procedures is the common feature of the recent proposed calibration procedures e.g. [56] [57] [58] and [59]. It would be desirable to simply these as much as possible for practical use. The results look promising so-far that the extracted parameters can help to reduce these iterative efforts. But it would be beneficial with a comparison between BES-calibration that target regressed estimations as proposed here with other similar calibration methods such as developed by Liu et al [58].
References


[14] S. Hammarsten, A critical appraisal of energy-signature models,


[43] American Society of Mechanical Engineers, Test Uncertainty-Instruments and Apparatus, American Society of Mechanical Engineers


[57] T. A. Reddy et al, Calibrating Detailed Building Energy Simulation Programs with Measured Data—Part II: Application to Three Case
