Statistical analysis of 200 digital twins for thermal load of Swiss buildings created from smart grid monitoring data

Andreas Melillo, Esther Linder, Braulio Barahona, Philipp Schuetz

esther.linder@hslu.ch, braulio.barahonagarzon@hslu.ch, philipp.schuetz@hslu.ch

Corresponding author: andreas.melillo@hslu.ch

Abstract. Exploiting thermal inertia of buildings as flexibility for the electric grid requires information about thermal dynamics in the system. The presented paper proposes a new method to create digital twins based on smart meter monitoring data and has been applied to 200 Swiss residential buildings. The statistical analysis of these heat pump based heating systems shows expected distributions of key parameters such as heat losses, solar gain factors and thermal capacities. Regional comparisons for crucial building parameters have been carried out and the remote identification of potential renovation candidates has shown promising results.

1. Introduction
Over 60% of the energy demand in the residential sector is used for space heating [1]. On the other hand, detailed energy consumption data will be more and more collected via the Smart Meter roll outs promoted on a national and even European level. A comprehensive overview of building properties for locations using a heat pump as their primary heating source can help to identify high consumption and thereby optimization potential to increase the renovation rate most efficiently.

Standard digital twin models to estimate building parameters encompass on one side so called data driven black box model. An example is given by [2] who present an electricity demand forecasting methodology based on historical data, which can be used in decentralized demand-side planning. A similar approach was chosen by [3] to estimate the heating behaviour in residential buildings with a data mining methodology. For black box models, the model parameters have often only a weak physical interpretability, whereas on the other side, white box models employ detailed information demand and are computationally expensive. As they are simulation methods based on physics, the model parameters offer an interpretation and elucidate the dynamics of the system. An interesting comparison of three models of this type to simulate building thermal behaviour (EnergyPlus, IDA ICE and TRNSYS) was performed by [4]. In [5] the gap between actual and simulated building energy performance has been studied and the most profound factors that influence the energy building performance were identified.

The method here proposed to create digital twins of heating systems driven by heat pumps is based on the ideas from [6], [7] and combines the data driven approach using the smart meter consumption data with a fast simulation tool to a so called grey box model. With the low demands of computation resources, compared to other simulation frameworks, and the possibility to mitigate the necessity of detailed information of the buildings, we were able to extend a simulation based approach to a large size study with 200 test sites distributed all over Switzerland.
2. Material & Methods

2.1. Data Sources

For the process of creating the digital twins, two major data sources were used. The first one was the heat pump power consumption data set consisting of a total of over 200 residential buildings, with a resolution of one second. The data was recorded via custom smart meter devices by tiko energy solutions AG. As a first part of the validation, simulated consumption profiles were compared to the actual consumption profiles. Additional to the power consumption time series, meta information about the buildings were provided and used to further validate the resulting distributions of crucial, heating related building parameters. These plausibility assessments are discussed in Section 3.2. Key information used were zip-code location, type of heating system, MINERGIE® label, and information about the size of the buildings as living space and number of floors. The data set was provided by tiko energy solutions AG. The overview of the spatial distribution in Figure 1, created from the zip-code location, shows that the considered systems are broadly distributed over Switzerland.

![Figure 1: Distribution of zip-code locations in Switzerland. Data for at least one building was available close to these locations.](image)

The second necessary input data source contains weather information used for simulations. This data set contained hourly logged solar irradiance and outside temperature data and has been taken entirely from Meteoswiss, collected via its IDAWEB access to archive data of ground-level monitoring networks. In order to assign to each location a suitable weather data source, the spatially closest, according to the zip-codes of the buildings, weather station providing necessary content has been selected.

2.2. Modelling Procedure

To create the digital twins the procedure suggested in [6], [7] has been applied. Based on the iterative comparison of the real world and simulated heat consumption profiles and the continuous adjustment of the building properties, the building properties eventually converge to a set of properties reproducing accurately the yearly heat pump consumption profile. As an optimization procedure the Nelder-Mead implementation of SciPy was used. This approach allows to reproduce suitable building properties like its lumped heat capacity $C_l$ [J/Km²], its integral heat loss through the building envelope $H$ [W/Km²] and its solar gain parameter $g$ [-], which is related to the conversion of energy stemming from solar irradiation. The underlying differential equation used is of the form

$$C_l \frac{dT_{inside}}{dt} = -H \ (T_{inside} - T_{ambient}) + gI + L_{emitter} + L_{internal}$$

(1)

where $T$ [K] refers to the temperature (index inside: building modelled as one room, index ambient: outside temperature from nearest weather station), $I$ [W/m²] to the intensity of solar irradiation and $L$ [W/m²] refers to the heat contributions of the emitter system and the inhabitants and appliances (index internal), respectively.
3. Results
The results section is divided in two parts. In the first part a statistical overview of the modelling process and outcome is given, in the second part, test applications of the presented method are discussed.

3.1. Modelling Results
To assess the quality of the building models we chose the L2 norm of the difference between the actual consumption profile and the reproduction of the model. To be able to compare the different models among each other, these values are normalized by the total yearly demand. For a further comparison of the used metric with RMSE and NRMSE (normalized by total real yearly demand) see Table 1. The distribution of the normalized L2 norm (NM) is shown in Figure 2A. Figures 2B-D show examples for characteristic regions of the metric. The largest part of the buildings investigated show a very good reproduction of the demand profile by the proposed model, as depicted in Figure 2B. The region beginning after the drop of the Gaussian part in Figure 2A is characterized by peculiarities in the original demand profile, where e.g. the slope at the beginning of the year is much steeper than after summer, as shown in Figure 2C. Since the approach tries to fit the model parameters for the whole year, it cannot account for such changes, presumably introduced by a device change or renovation activities during summer, within the observation period. The building models with values of NM bigger than 8 show strange consumption patterns and are therefore considered as outliers. An example is given in Figure 2D, where in the second half of the year no demand is measured at all. A possible explanation for such appearances could be the replacement of the entire heating system or a cancellation of the monitoring. However, the results show that the approach as suggested only works if the input consumption profile follows the expected s-shaped profile. However, the problematic examples in Figure 2C and D have been kept in this study on purpose to gain a comprehensive overview of the situation.

Figure 2: Subplot A (top left) shows the distribution of the relative norm NM. Subplots B,C and D (top right, bottom left, bottom right) are sample results for particularly interesting buildings of the fitting quality (NM < 4, 4< NM<8, NM>8) and compare the reproduced yearly consumption profiles of the models (red) with the ground truth data (black)
Table 1: Overview of metrics based on the cumulative consumption profiles used to compare the quality of the models created for the sample buildings 2B, 2C, 2D: RMSE, NRMSE (normalized by total consumption) and NM

|    | RMSE [kWh] | NRMSE | NM   |
|----|------------|-------|------|
| B  | 260        | 0.009 | 0.82 |
| C  | 801        | 0.046 | 4.30 |
| D  | 985        | 0.153 | 14.30|

In Figure 3, the distributions of the crucial building parameters are summarized. The reconstructed heat transmission coefficient $H$ shows an expected right tailed distribution, which can be used to identify buildings with the highest heat losses according to their model representations and label them as possible renovation candidates. For the solar gain distribution, the peak at very low values and the group with highest values are most probably artifacts of a mismatch between the actual window area, which is unknown, and the assumptions of the simulation. Apart from that, it follows a bell shaped profile with 0.6 as a mean, which is comparable to typical window values in the last three decades. The lumped capacities $C_l$ also show an expected bell shape and due to the sensitivity analysis in [7] have not been investigated any further so far.

In order to understand how our modelling approach would transfer to other buildings outside our 200 samples, we look at the correlations between the model normalized metric (NM) and the meta data information. From a visual inspection of the scatter plots in Figure 5, it is not possible to identify clear and strong tendencies. The sample Pearson correlation coefficient $r$ and Spearman’s rank correlation coefficient $r_s$ are very low. To this point, the accuracy of our models is not strongly linked to building
size or number of inhabitants. This is an indication that the approach might be robust and that we can expect it to perform similarly in a larger sample of buildings. Other important meta data is categorical, such as type of window glazing, MINERGIE® label, and type of heating system. Thus, we resort to box plots as a visual tool to identify gross differences of the model performance metric for different groups of buildings. As shown in Figure 6, NM median and standard deviation is not very different for buildings with triple glazing (THR) as it is for those with only double glazing (TWO). Although the distribution of NM in THR buildings is more skewed than that of TWO buildings. Overall, such distributions of NM over subgroups of buildings indicate robustness under different building and heating system types.

3.2. Applications

One application of the suggested building modelling procedure is to estimate the insulation status of the building. Together with the meta data such as building year or location of the building, it is possible to verify the suggested insulation status from the building modelling procedure.

Based on the distribution of the estimated heat transmission coefficient ($H$) shown in Figure 3, the insulation status is defined. Buildings are considered as not well insulated if they show a heat transmission coefficient that is greater than 0.7 W/Km², which corresponds to the dip in the distribution shown in Figure 3. Buildings with a heat transmission coefficient below that value are considered as well insulated.

As newer or recently renovated buildings consist of modern and therefore generally better insulation material, it is expected that the fraction of not well insulated buildings relative to the well-insulated buildings decreases with time. In total there are 22 (11 % of all) buildings in the dataset identified as not well insulated (high heat transmission coefficient) and 178 (89 % of all the) buildings which are well insulated (low heat transmission coefficient). For the 200 buildings analyzed in this study, it is found that indeed over the last 50 years, the fraction of not well-insulated to well-isolated buildings decreases with a slope of −1 %/year. The newest well insulated building was built in 2001 and renovated in 2010, whereas the oldest well insulated building was built in 1800 and renovated in 1979.

Another check for the recovered insulation status is given by the MINERGIE® label of a building, a Swiss standard for sustainable building. This information is available in the meta data as a binary flag. The resulting statistics is shown in Table 2. All houses with a MINERGIE® label are recognized as houses with good insulation by our model. On the other hand, out of the buildings, which do not have the MINERGIE® label, only 9 % are poorly insulated.

| MINERGIE® | Not well insulated | Well insulated |
|-----------|-------------------|----------------|
| Not MINERGIE® | 22 | 162 |

The climate in Ticino differs from the climate in the rest of Switzerland. A crucial driver for the required insulation for a building is given by the (average/lowest) outside temperature, as the reduction of heat losses due to low ambient temperatures necessitates better isolation. Considering the distribution of the 200 buildings in this analysis as shown in Figure 1, the buildings in the region of Ticino distinguish themselves from the other buildings in this study by laying in a region of tendentially higher mean yearly temperatures. The yearly mean temperature in Ticino for the buildings in this study lies around 14 °C, whereas the yearly mean temperature for the buildings for this study in the north and northwest of Switzerland lies between 5 and 10 °C.

Hence, another plausible explanation of the suggested building insulation is given by the geographic location of the building. Table 3 gives the overview of the averaged heat transmission coefficients and solar gain factor depending on the location of the building. The mean of the heat transmission coefficient $H$ for buildings that are in Ticino is 1.4 times higher than for the buildings which are not in Ticino. This
means that buildings not in Ticino are generally better insulated, which is corresponding to the expectation given from the different yearly mean temperatures for the geographic regions.

Following Meteoswiss, in 2020 the yearly global radiation (solar irradiation on a horizontal area) is around 10 W/m² higher in Ticino than in the regions in northern Switzerland. The amount of solar irradiation translates into the solar gain that is modelled for the windows. The mean of the solar gain factor \( g \) for the buildings in this analysis is 13 % higher for buildings in Ticino relative to those not in Ticino. As there is more solar irradiation in Ticino compared to the rest of Switzerland, the solar gain factor is higher in average in Ticino as more thermal energy in the building can be gained from the solar irradiation, assuming the same window area for the buildings. The assumption of the same window area for every building is another limitation from the procedure, as a variation in boundary condition such as a change in solar irradiation must translate into a change of the solar gain factor \( g \).

| Region | Mean of H [W/Km²] | Mean of g |
|--------|-----------------|----------|
| Not Ticino | 0.36 | 0.79 |
| Ticino | 0.51 | 0.89 |

### 4. Discussion and Conclusion

The analysis of 200 digital twins of heating systems indicates that the proposed approach enables the reproduction of thermal performance related building parameters. The validation procedures, such as the recognition of MINERGIE® labelled buildings, the analysis over climatic different regions in Switzerland, as well as the change in fraction of poorly insulated buildings built over time, show that the reconstructed building parameters are an accurate representation of the real-world situation.

Confronted with this large amount of building sites, we were able to identify new key problems in the proposed method. Two of the major examples are the necessity of an automated filtering for suitable s-shaped profiles before the application of the method and the urgency of a reliable estimation of the window area in order to produce trustworthy results for the solar gain factors.

Future steps are directed to widen the application range of the proposed procedure including prediction and district heating applications, as well as carrying out a deeper analysis of the extensive data set and enlarge the building number in order to give a more comprehensive summary on the condition of Swiss residential buildings in terms of heating related parameters.

### 5. Acknowledgements

This work has been carried out in the framework of an Innosuisse funded research project (contract number 25824.1 PFES-ES). We would like to thank tiko Energy Solutions for their collaboration.

### References

[1] Eurostat, “Energy Consumption in Households,” *Stat. Explain.*, no. March 2017, p. 126, 2018.
[2] S. Williams and M. Short, “Electricity demand forecasting for decentralised energy management,” *Energy Built Environ.*, vol. 1, no. 2, pp. 178–186, 2020.
[3] M. Laskari, S. Karatasou, M. Santamouris, and M. N. Assimakopoulos, “Using pattern recognition to characterise heating behaviour in residential buildings,” *Adv. Build. Energy Rex.*, vol. 0, no. 0, pp. 1–25, 2020.
[4] D. Mazzeo, N. Matera, C. Cornaro, G. Oliveti, P. Romagnoni, and L. De Santoli, “EnergyPlus, IDA ICE and TRNSYS predictive simulation accuracy for building thermal behaviour evaluation by using an experimental campaign in solar test boxes with and without a PCM module,” *Energy Build.*, vol. 212, 2020.
[5] E. Cuerda, O. Guerra-Santin, J. J. Sendra, and F. J. Neila, “Understanding the performance gap in energy retrofitting: Measured input data for adjusting building simulation models,” *Energy Build.*, vol. 209, p. 109688, 2020.
[6] A. Melillo, R. Durrer, J. Worlitschek, and P. Schütz, “First results of remote building characterisation based on smart meter measurement data,” *Energy*, vol. 53, no. 9, p. 117525, Apr. 2020.
[7] P. Schuetz et al., “Automated modelling of residential buildings and heating systems based on smart grid monitoring data,” *Energy Build.*, vol. 229, p. 110453, 2020.