THE ENERGY PERFORMANCE CERTIFICATE PREDICTIVE MODEL

NEEM HUB

Nordic Energy Efficient Mortgage Hub

17 May 2023

PUBLICATION INFORMATION

SOURCE ACTIVITY: WP2, PUB
AUTHORS: P. VISETTI, J. AALTO, G. SINDERMANN, J. SMERTINAS, P. BACHER (GDFA, DTU)
FINAL DATE: 28.04.23
CONTRACTUAL DELIVERY DATE: MONTH 12

The Nordic Energy Efficient Mortgage Hub aims to scale-up lending to energy renovations in the Nordics and will publish a blueprint on how to accomplish this which will be implementable in other regions of Europe and, indeed, the world. In striving to increase energy renovations, the NEEM Hub will help achieve the targets of the European Green Deal and contribute to addressing ambitious national climate targets.

The NEEM Hub will be comprised of a long list of institutions from the financial sector, behavioural scientists, mortgage specialists and authorities, and digital technologies communities from across the Nordics, all guided by leading European Economics Consultancy, Copenhagen Economics.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101032653.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

One of the NEEM project's goals was to facilitate and bring to the market a modular software platform for the estimation of EPC labels for buildings supplied with district heating. The project has succeeded in this goal by creating a software environment based on existing energy signature models, expanding them, and applying the software to Scandinavian buildings in pilot studies.

The models developed were validated on Norwegian school buildings, where a high level of control and detailed measurements of the heating system were available. After validation, the software was applied to EPC label estimation in Denmark on buildings supplied with district heating. Additionally, two pilot studies were performed in Norway and Sweden, where the software was applied to buildings whose main energy carrier was electricity.

This report highlights substantial challenges faced in the implementation and testing of the software on buildings in the NEEM Hub. As seen by testing it on many buildings, the main issues appear to be the requirements for nonlinear optimisation methods for the models. While the model specification is simple, due to real-world application, the numerical stability of model optimisation must be ensured so that false results are not generated. The necessity exists for C++-based automatic differentiation for the scalability of the application, which also presented numerical issues that were overcome.

In conclusion, the project produced scalable modular software that can be applied to more than district-heating-only applications. In the NEEM projects, this software was used to estimate the EPC labels of buildings based on their simulated yearly energy consumption. Additionally, the software performs highly accurate long-term forecasts of energy demand. It also provides a diagnostic overview of building characteristics that were used to validate the EPC label and a quick overview of the building’s condition. We assess these to be critical conditions for a model to perform the tasks required in NEEM.

Chapter 1
INTRODUCTION

1.1 NEEM HUB AND NEEM CORE SOLUTION

As part of the Energy Efficient Mortgages Initiative supported by Horizon 2020, the Nordic Energy Efficient Mortgage (NEEM) Hub aims to scale up lending to energy renovations in the Nordics. To facilitate that, the project has developed a NEEM core solution: a three-step guide for banks to deliver customer-specific energy renovation recommendations to their customers. The solution is designed to simplify and automate the process of finding profitable energy renovations for residential houses.

The NEEM core solution required building an estimation framework for the energy performance of buildings in the Nordic region heated by the common district-heating system. The goal of the model development was to adapt existing research into a largely automated energy efficiency modelling framework. This project produced an R-based software environment that leverages the modularity of the energy signature models to forecast the yearly energy consumption of buildings and consequently provide them with an EPC label.

1.2 ABOUT THIS REPORT

This report describes the foundational models and development of the modular software for automated EPC estimation, which is based on hourly energy consumption and weather data. It includes a validation case based on a Norwegian school, but the main results from the application of the software are presented in the Deliverable 5.2 report. It also provides a brief overview of complications with EPC label estimation in Denmark, Norway, and Sweden and the general intricacies of automated nonlinear model estimation.

Chapter 2

CONCEPTUAL UNDERSTANDING AND EXPLANATION OF THE MODEL

2.1 CONCEPTUAL UNDERSTANDING

The current EPC labelling procedure takes place partly through a combination of different measurements of the building taken on site (e.g., materials, kind of building, heating devices) with physical knowledge in an equation that weighs and accounts for all the factors. In the end, this formula produces primary energy demand for the building. Each label in the current EPC labelling system consists of a range of energy usage per square meter and spans the entire real line, also negative in case the building produces more energy than it consumes.

This manual method of classifying buildings with an EPC label is time-consuming, expensive, and most importantly uncertain. Such a manual classification cannot look at how the building performs as a whole and may significantly misclassify.

As stated by Haldi and Robinson, several studies have shown that the discrepancy between actual and estimated energy consumption is significant. One study showed that the difference between estimated and actual energy consumption can exceed 100% due to occupants' behaviour (Brohus et al.), and a difference of 300% was observed between identical buildings in a study by R.H. Socolow.

The IEA EBC Annex 53 report (Yoshino et al.) states that the energy consumption of a building is influenced by six main factors: climate, building envelope characteristics, building services and energy systems characteristics, building operation and maintenance, occupant activities and behaviour, and indoor environmental quality provided. A similar categorisation is found in Yu et al.

As stated above, a major factor in energy performance is occupant behaviour. However, only minor focus has been put on data-driven methods where human interaction with the building is explicitly taken into account when estimating buildings' thermophysical properties. A building's heat consumption and dynamics can be heavily influenced by occupants' changing preferences for the indoor environment, and operational staff's adjustments of the energy systems might lead to significant disturbances when modelling a building's heat consumption. The research done in this project addresses this.

The discrepancy between the prior construction's anticipated energy consumption and the actual, however, is not limited to the occupants' behaviour. Several studies have shown that the thermal properties of buildings prescribed in the design and reality can vary significantly. In a study from 2011, it was found that 18 out of 18 (100%) newly built British dwellings had a significantly higher heat-loss coefficient than anticipated in the design when it was assessed by co-heating methods on the finished building (Wingfield et al.). The Danish Energy Agency also found that 23% of the energy performance certificate labels issued in 2018 were misclassified, and 21% and 31% were misclassified in 2017 and 2016, respectively (Energistyrelsen).

A pertinent example of the physical factors influencing energy consumption in the Scandinavian climate is leaky windows. A manual examination of the building may have a difficult time identifying a single heavily leaking window. If identified, it is nearly impossible to determine the additional heat usage of the leak, i.e., how much should this affect the energy performance of the building, and how should it affect the EPC label of the building? A building may perform worse than the EPC manually established label says.

Consequently, the discrepancies between the intended building energy performance and the actual performance are hard to quantify as the effects of building characteristics are practically difficult to separate from the occupant-related effects.

In the NEEM project, to estimate energy efficiency, we have employed a type of data-driven model that can separate the energy performance of the building from the usage to the extent possible using simple scalable meter readings combined with weather data.

This report presents the basic concepts and developments on how the steady-state energy signature model presented by Rasmussen et al. (2020) was altered and expanded to produce test results for the concrete pilot tests in the NEEM Hub. As these data-driven models are proprietary technology of the Technical University of Denmark (DTU) pending publication, they will not be explicitly detailed in this report. However, their general idea and advantages are outlined below, with a validation case.

Data-driven models

Data-driven models are mathematical or algorithmic procedures that process data to produce a desired result. They are used to create models that describe physical, static, or dynamic phenomena, such as the temperature evolution in buildings, to predict the electricity load at a substation, and to describe social patterns. These models can calibrate using data and predicting features of the phenomenon. Grey-box models are a combination of white- and black-box models, aiming to bridge the gap between the strengths and weaknesses of each. They are simple, low-dimensional models that are calibrated to data and capture the most important dynamics, leaving out small and often impossible-to-capture dynamics that have little influence on the system as a whole.

White-box models are not data-driven and are based entirely on physical knowledge of the system. They tend to be detailed and become large due to the equations needed. A direct consequence of these large and coupled models is the significant number of parameters appearing. Typically, parameters need to be calibrated to the system at hand. However, white-box models tend to be difficult to calibrate and match to the given system.

Black-box models are purely based on data and carry no explicit information about physical properties, dynamics, or constraints. This can be an advantage in situations where the system is too complex to model by physical knowledge alone. However, due to the complete lack of physical knowledge, this kind of model typically requires large amounts of data to sufficiently learn the dynamics of the phenomenon. Another problem arising from the lack of physical knowledge is black-box models' lack of ability to deal with unobserved situations.

Small, uncapturable dynamics that are left out by grey-box models are not just lost in the modelling. Often, these small "disturbances" in the system cause the data to become noisy, and these models then characterise the noise and include the description. The noise description may be useful for determining the uncertainty of the predictions made by the model, which in turn is useful for decision-making to know how close to constraints the system is allowed to be.

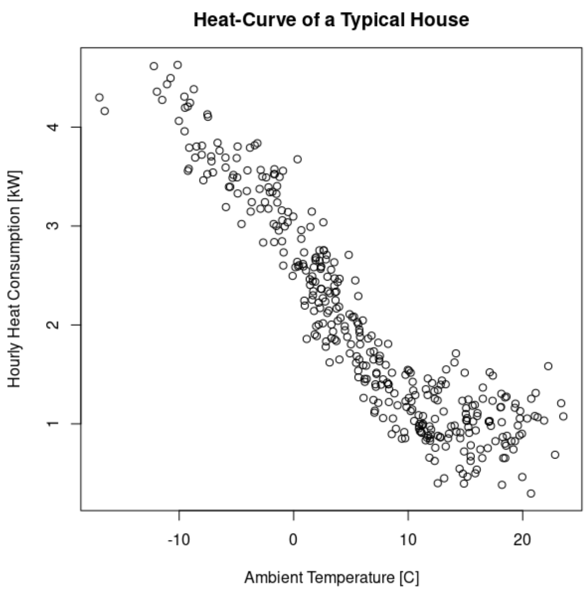
The model described in this report and used in the NEEM project is of the grey class; it includes interpretable and physical parameters and states and is calibrated using data.

2.2 Building energy signature and the digital twin

1.1.1

The most intuitive and reliable way to visualise and explain the heat consumption of a building is the energy signature approach. The energy signature is the relationship between heat consumption described as a function of the outdoor air temperature.

Figure 1
Hourly heat consumption against ambient air temperature for a district-heated building in Denmark



Source: DTU

Figure 1 above illustrates some typical observations of a Danish building's heat consumption concerning ambient temperature. The relationship appears rather linear, however, at a temperature of 10 degrees Celsius, the relationship transitions to a constant.

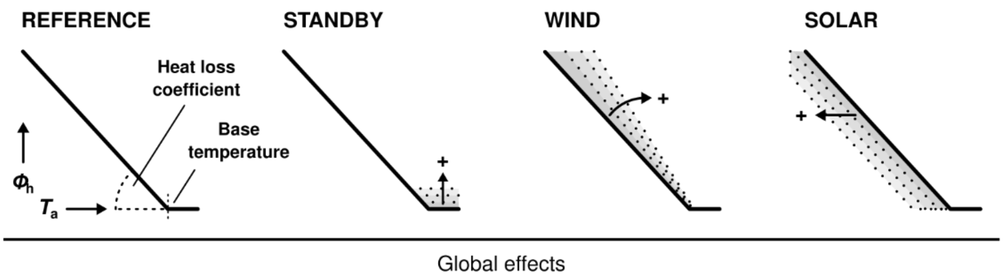
As such, the energy signature is often approximated as a linear function concerning temperature. However, it can be more aptly described as a sigmoid function as the heating capacity of the energy system becomes saturated in extreme conditions. For the sake of simplicity, we assume that the heat consumption relationship is linear concerning ambient air temperature.

The slope and intercept of even the simple linear approximation carry meaningful information about a building's energy performance. For instance, the slope represents the heat-loss coefficient (HLC) of the building, which describes the rate of heat flow through the building's envelope when a temperature difference exists between the indoor air and the outdoor air under steady-state conditions.

During weather-independent periods such as summer, heat consumption is modelled as a constant for buildings without cooling and heat recovery. The change point from the weather-dependent or heating period to the weather-independent period is described as the base temperature (T_b). The best temperature of the building is the temperature at which it is in thermal equilibrium with the outside, which can be thought of as the operating temperature of the building. Higher T_b often means higher heat loss at lower ambient temperatures.

While the relationship between heat consumption and ambient temperature can often be described as linear, its relation to other weather phenomena such as wind and solar is more complicated. Nevertheless, energy signatures can be used to visualise their influence and gain information on the building's condition.

Figure 2
Heat curve and its responses to common effect



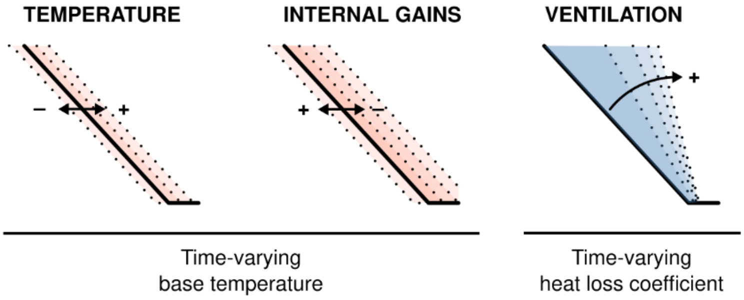
Source: DTU

Common effects in Figure 2 include increasing standby energy needs of the building, the effect of wind on heat loss, and solar radiation heating the building. The HLC (the slope) and the base temperature (transition point) are highlighted.

Figure 2 above illustrates how the HLC of the building is influenced by the wind and how the entire energy signature is shifted by solar radiation. Cold, windy conditions increase the HLC, i.e., the building bleeds the heat faster, and solar radiation warms the building by entering the house through windowed surfaces and direct induction into the material of the building. In both cases, the effects can be modelled as a function of the wind speed and solar irradiation if one has data on the two variables.

Figure 3

Effect of changing indoor temperature and internal heat gains on base temperature

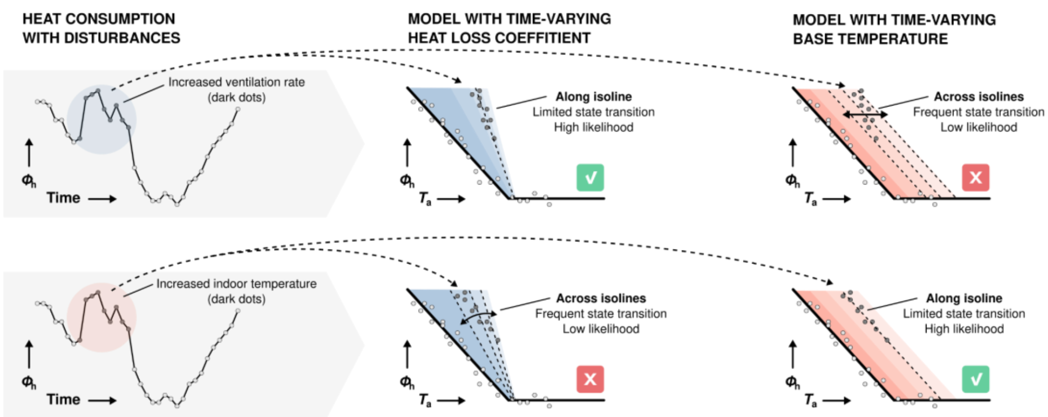


Source: DTU

The two red sub-plots in Figure 3 show how the effect of changing indoor temperature and internal heat gains alter the base temperature, and the blue sub-plot shows how the ventilation rate alters the apparent HLC. For other effects driving heat consumption, such as changing indoor temperatures, internal heat gains, and ventilation rates, measurements of relevant variables are often unobtainable. These are visualised in Figure 4.

Figure 4

Disturbances and the effect on energy signature



Source: DTU

Figure 4 above shows a conceptual illustration of how two fundamental types of disturbances (changing ventilation rate and indoor temperature) affect the energy signature and consequently how they dictate which model to use. It demonstrates one of the difficulties in attributing disturbances in heat consumption to appropriate weather effects. The figure displays two identical time series of observed heat consumption. In both cases, nine of the observations are assigned to an unknown disturbance.

In the first row of Figure 4, the disturbance is caused by an increased ventilation rate and hence increased heat consumption (dark dots). In the second row of the figure, the disturbance is caused by increased indoor temperature and hence increased heat consumption (dark dots). In the first scenario, the data can be modelled by including a time-varying HLC or a time-varying base temperature as shown in the first row, second and third columns. When choosing the model with a time-varying HLC, the disturbance can be explained as a single increment in the ventilation rate (i.e., a one-time change in the HLC). On the other hand, when choosing the model with a time-varying base temperature, the disturbance can only be explained equally well by several subsequent changes in the base temperature.

Regardless of which approach is correct, assigning the most accurate model for a variety of buildings autonomously poses a substantial challenge. This project undertook and developed several approaches to tackle this, and it was paramount to find an accurate model describing the most common disturbances experienced by each building.

If the energy signature is well estimated, i.e., the effects of weather phenomena on the energy signature are well captured, then the simulation of the building's total energy expenditure over a year becomes trivial. With models that can accurately estimate the yearly energy expenditure of a building, the assignment of an EPC label becomes an exercise in knowing the correct legislation for each country.

2.3 The model developed

The previous sections covered the modelling approach heuristically. Here, a more rigorous introduction to the base model for heat signature is provided. This is followed by an introduction to its extensions, where the variations in ventilation rates and base temperature shifts are accounted for.

The static model introduced in this section was used as the base model in all pilot studies carried out by the NEEM project. Additional testing and validation of extended models were performed during the Danish pilot studies.

Generally, the static energy signature model consists of two regimes, one when the energy demand is weather dependent (when building heating is active) and the other when energy demands are constant (when building heating is unnecessary, i.e., in summer). The weather-dependent regime (denoted $f(\cdot)$) is built to capture a building's response to weather changes. The weather-independent regime (denoted $g(\cdot)$) provides the baseline energy needs of the building irrespective of weather dependence, such as hot water use, and some electricity draw.

The two regimes allow us to capture useful information about the building, for example, exactly how it responds to different weather phenomena and how the underlying energy demand of the building behaves. This can be used for two obvious tasks: building diagnostic overviews and building energy demand forecasts on large time scales, which is of great interest for this project task.

As it is not immediately obvious when building heating is activated, the model is an autonomous mixture of the two regimes and transitions between these.

2.3.1 The static model

The static model is the foundation for the model extensions. It is the simplest model used as a baseline for estimating the EPC labels in the NEEM project. Also, for the Norwegian and Swedish cases, the simple model was exclusively used without comparisons to the extended models.

In simple terms, the mathematical model of the heat consumption Φ_{heat} can be expressed as

$$\phi_{heat} = max[f(\cdot), g(\cdot)]$$

where $f(\cdot)$ is a function describing the heat consumption during periods with heat demand and $g(\cdot)$ is a function describing the heat consumption during periods without heat demand. The maximum value is then between the two regimes. The dot notation is a placeholder for any explanatory variable used in the functions. This is typically the outdoor temperature, wind speed, solar irradiation, or other driving forces. Together, the functions form the energy signature introduced in the previous chapter.

For buildings without cooling or heat recovery systems, a constant heat consumption model is used during periods without heat demand. The heat consumption model for periods without heat demand, $g(\cdot)$, is therefore defined as

$$g = \phi_0 + e$$

where ϕ is the constant daily base heat load related to heat losses from regular building use (e.g., hot water) and e is i.i.d white noise.

For periods with heating demand, i.e., weather-dependent heat consumption, the heat consumption can be derived from the heat balance:

$$\phi_{heat} - \phi_{tr} + \phi_{sol} + \phi_{int} - \phi_{vent} + \phi_{mass} + \phi_{latent} = 0$$

where ϕ_{heat} is the heat consumption, ϕ_{tr} is the transmission loss, ϕ_{sol} is the solar gain, ϕ_{int} is the internal heat gains, ϕ_{vent} is the ventilation loss, ϕ_{mass} is the release of thermal energy stored in the thermal mass and ϕ_{latent} is the energy absorption and release due to evaporation and condensation in the thermal zone.

For daily averaged heat consumption and weather data, as used in this study, the heat exchange with the internal thermal mass (e.g., building structures and furniture) and the latent heat exchange can be ignored.

The following form is a robust, simple, accurate energy signature approximation:

$$\phi_{heat} = LSE(f(\cdot), g(\cdot))$$

where the function $g(\cdot)$ is defined as above, the function $f(\cdot)$ is defined as

$$f \ = (UA_0 + W_s UA_w)(T_b - T_a) \ + gA\, I_g + e$$

the parameter UA_0 is the base HLC, UA_w is the influence of the wind speed W_s on the HLC, T_a is the ambient temperature observation, and T_b is a parameter for base temperature. gA is the solar gain factor for the global solar irradiation observations I_g.

$LSE(\cdot)$ is the log-sum exponential, which acts as a soft-maximum function describing the smooth transition between the two heating regimes. It is controlled by some hyperparameters that are omitted here for the sake of simplicity.

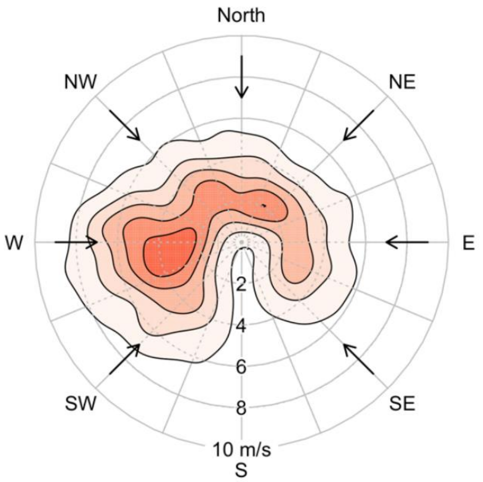
Weather parameters used in this model are chosen from previous research (Rasmussen et al., 2020) that has shown them to be by far the most significant influences on buildings' energy consumption.

It is noteworthy that with the base model, the energy signature of a building is well defined, meaning it is possible to accurately forecast long-term energy consumption, which is necessary for EPC label estimation.

Additionally, as will be demonstrated later in this report, the model already possesses many parameters that quickly provide valuable information about a building's potential shortcomings. It is obvious that a high UA_w value indicates a large influence of wind on a building's energy performance.

The most important feature of the static model is that it is modular, so each interaction parameter can be easily extended depending on the needed model complexity and data availability.

Figure 5
Wind directions and speed (10 m) at the building site obtained from August 2019 to June 2020



Source: Rasmussen et al., 2021

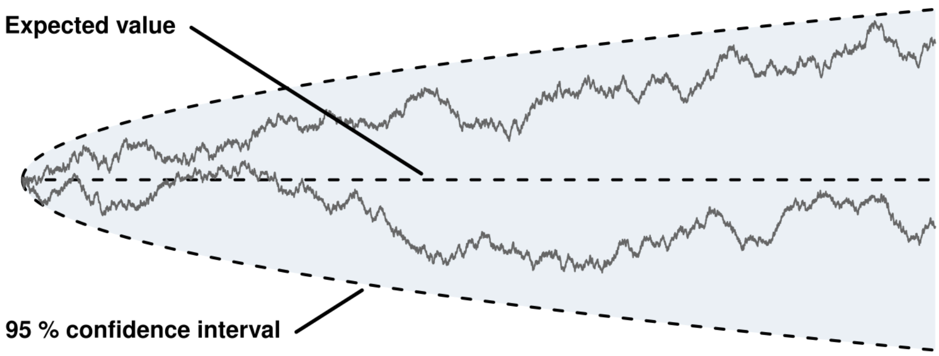
Darker colours in the density plot in Figure 5 mean more frequent observation, (Rasmussen et al., 2021). For example, if the data for wind direction are available, then the parameter W_s can be easily replaced with a function describing a building's wind susceptibility in a given direction. This function estimate can later be reconstructed to show which windows facing a given direction would need further insulation.

2.3.2 Model extensions

As mentioned in the introduction, the NEEM project enabled the testing of novel methods for estimating a building's physical thermal properties while synchronously estimating time-varying effects caused by human interactions with the building. This is done by combining an advanced smooth and nonlinear formulation of the steady-state energy signature model known from the literature with a hidden state formulated as a random walk to describe the human interaction with the building. For the sake of brevity, we omit some mathematical rigour.

Presuming that the indoor temperature, internal heat gains, and ventilation rate are not necessarily constant, they can be treated as hidden states evolving.

Figure 6
Two realisations of a random walk of 10,000 steps each, their expected mean and 95% confidence interval



Source: DTU

For the sake of simplicity, we form two models where only a single hidden state may be described by a random walk, namely:

1.1 Time-Varying ${UA_0}^t$ - allowing the HLC to vary provides insight into the building's use and estimates the specific ventilation losses.
2.1 Time-Varying ${T_b}^t$ - allowing the base temperature to vary provides insight into the building's internal temperature change, internal heat gains and user-driven use of the building.

One approach to estimate the model parameters and hidden state is to use maximum likelihood estimation (MLE). The advantage of the MLE method is that it allows for estimating parameters related to the noise term, and in this setting the hidden state. The outline of the MLE method is omitted here, but details can be found in Rasmussen et al. (2020). In practice, the model parameters are found by maximising the log-likelihood function, and the hidden state is found by Laplace approximation using R (R core team) and the Template Model Builder (TMB) developed by DTU (Kristensen & Nielsen; Thygesen et al.).

Models with hidden and unconstrained states such as those presented here are prone to overfitting. To what extent they overfit depends on the data, the estimation approach, and the objective function, which is subject to optimisation to find a suitable set of model parameters. To quantify the bias-variance trade-off, the models must be cross-validated. These extended models were fitted on buildings in the Danish pilot studies.

Chapter 3
VALIDATING THE MODEL FOR NEEM

This chapter presents a model validation of the static model and the two extended models. They were compared and validated on a highly monitored test case to verify the models' fit for the NEEM pilot testing before applying them to the NEEM project's Scandinavian building data. The model selection process we used for the buildings in the NEEM project is described in Section 3.2. Both statistical and physical validations are included, and how the extended model with time-varying parameters can be used for deep building energy performance diagnostics is shown.

3.1 TEST CASE: NORWEGIAN SCHOOLS

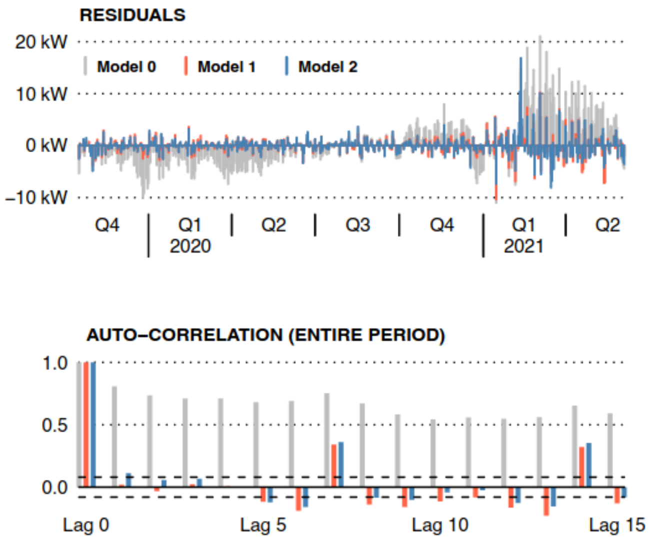
The test case is the newly built Montessori school in Drøbak, Norway. The building is a lower secondary school for 60 pupils with two floors and a heated area of approximately 900 m^2. The school was built with the vision to become Norway's most environmentally friendly school. The basis for the energy concept design is a well-insulated building envelope with minimal heat loss, a very efficient lighting system, a high-performance ventilation system, and a ground-source heat pump system that provides low-temperature heating in winter and free cooling in summer.

3.2 MODEL VALIDATION

Three sets of models were tested on the school buildings, namely the static model (Model 0), the time-varying ${UA_0}^t$ model (Model 1) and the time-varying ${T_b}^t$ model (Model 2), These models were later tested in the Danish pilot study in the NEEM project. The prediction errors of the models (Models 0, 1, and 2) and their estimated autocorrelation functions are displayed in Figure 7. The autocorrelation function describes the time dependence of the data, and in the ideal case, it should be equal to 1 lag 0 and close to 0 for the other lags. The residuals are rather small for the first year and begin to increase in 2021. This may indicate a systematic or behavioural switch in the usage of the building.

Regardless of this change in prediction accuracy, the autocorrelation functions for Models 1 and 2 (with time-varying parameters) look significantly better compared to the base model (Model 0). The autocorrelation function for time series data says something about how much time dependence (or time-related correlation) is present in the data. If the amounts of time dependence in the data are significant, this means there are dynamics or contributions that we (the modellers) do not capture, meaning the model can be improved. Model 0 seems to have systematic errors for all time lags, whereas Models 1 and 2 have significant lag dependencies in lag 7 and 14 only. This could indicate some weekly dynamics these models do not describe.

Figure 7
Upper: The residuals (prediction errors) of Models 0, 1, and 2

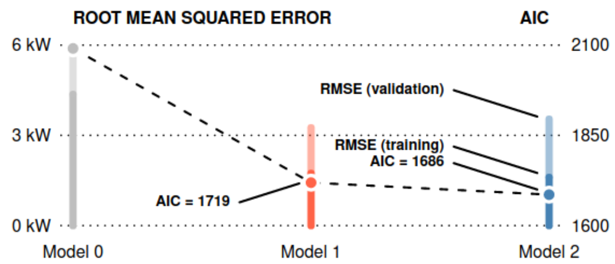


Source: DTU

In Figure 7, in Q4 2019 and Q1 and Q2 2020, Model 0 tends to consistently estimate lower energy consumption. Models 1 and 2, in contrast, seem to have smaller residuals and are more evenly spread around 0. These things could indicate that Model 1 or 2 better describes the building's thermal dynamics.

Looking at the performance in terms of accuracy, Figure 8 below shows the estimated Root-Mean-Squared-Error (RMSE). A smaller RMSE means better accuracy of the model, which is used to choose the best model among Models 0, 1, and 2. In the NEEM project, we fit all three models to a building and choose the one that has the best performance.

Figure 8
The performance of the model in terms of RMSE and AIC



Source: DTU

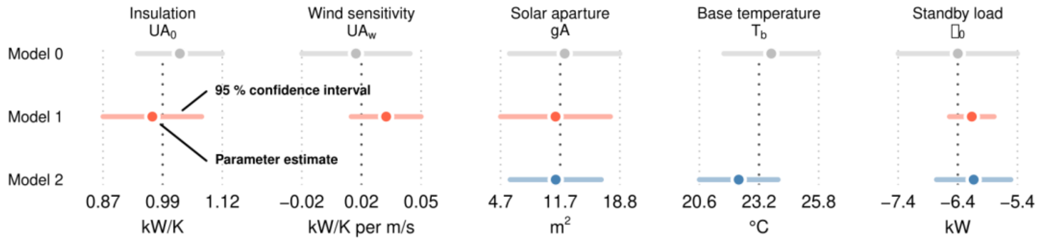
Based on the model evaluation of this real-life case study, the model captures the most important drivers and influences that contribute to the heat demand and heat usage of the Norwegian building (e.g., wind speed and whether it changes). This is indicated by analysing the performance of the model and looking at the prediction errors.

While the model behaviour appears to be consistent in this particular school building, as further validation, the models were tested on Danish buildings heated via district heating as part of the NEEM project.

3.3 DATA-DRIVEN BUILDING DIAGNOSTICS

This section outlines some of the results achieved for the validation case. This should illustrate how the models are different. In Figure 9, the estimated values of the physical parameters are shown for the three models. The dots indicate the estimate, and the horizontal bars indicate the 95% confidence intervals. For Model 1, the base temperature (Tb) is modelled as a random walk and is therefore not shown here. For Model 2, the insulation and wind sensitivity (UA0 and UAw) are omitted for the same reason. Although there are differences between the estimates, they are minor and thus all three models provide equivalent estimates, which validates the reliability between the models. We thus believe that the estimated physical parameters of the buildings in the NEEM project are reliable.

Figure 9
Estimated physical building parameters by Models 0, 1, and 2



Source: DTU

Additional insights can be achieved using the extended with time-varying ${T_b}^t$ – allowing for the base temperature to vary provides insight into the building's internal temperature change, internal heat gains and user-driven use of the building and time-varying ${UA_0}^t$ – allowing for the HLC to vary provides insight into the building's use and estimates the specific ventilation losses. We have omitted this, and we refer the reader to future publications about this model.

Chapter 4

ESTIMATING ACTUAL ENERGY EFFICIENCY PERFORMANCE

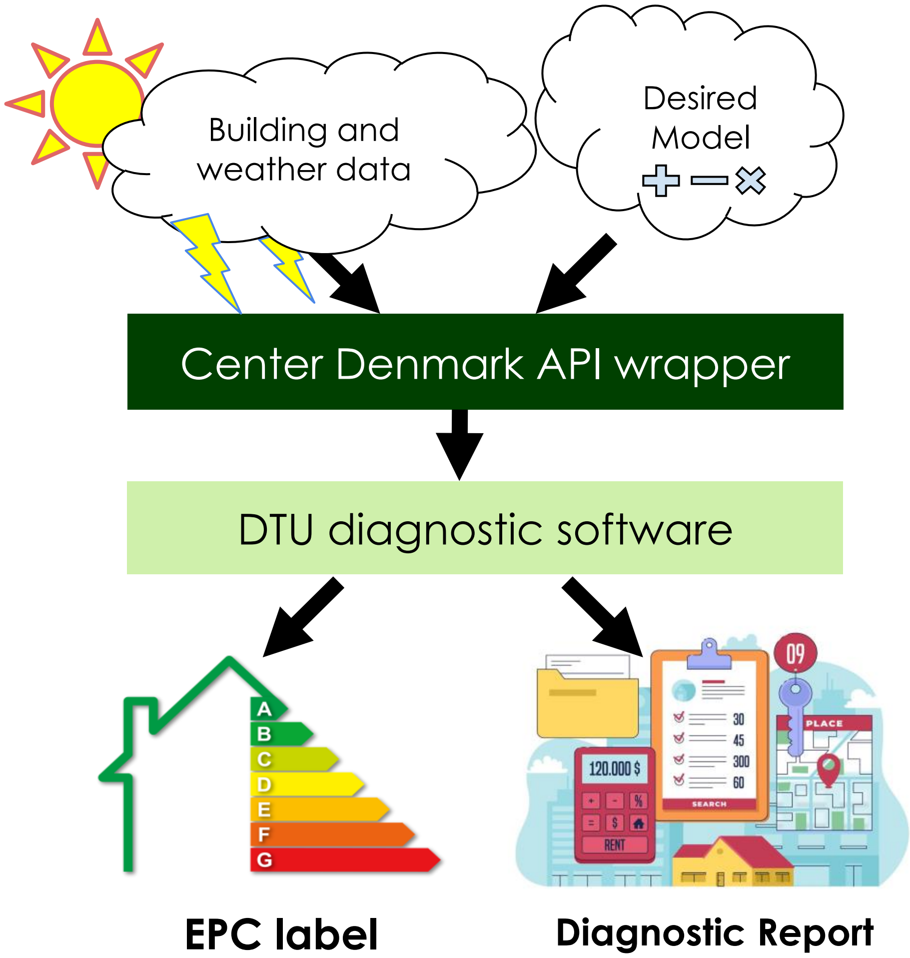
This section describes the modular software for automated EPC estimation developed by DTU and how it interacts with Center Denmark's data. It also provides a brief overview of complications with EPC label estimation in each Nordic country, as this information is non-trivial to obtain and must be reflected on when assessing buildings energy saving potential in each country and the testing results from the pilot tests. Lastly, challenges with automated nonlinear model estimation are discussed.

This project focused on building an estimation framework for the energy performance of buildings in the Nordic region heated by the common district-heating system. The main pilot studies on district-heated buildings were carried out in Denmark. The project carried out additional pilot studies in Norway and Sweden where the main heat carrier for buildings was electricity.

4.1 METHOD FOR AUTOMATED EPC ESTIMATION

The goal was to adapt the previously showcased research into a largely automated energy efficiency modelling framework. This project produced an R-based software environment that leverages the modularity of the energy signature models to forecast the yearly energy consumption of buildings and consequently provide them with an EPC label. The NEEM core solution's EPC labelling software functionality used in the project, described below in Figure 11, was developed by DTU and Center Denmark.

Figure 10
Visualisation of NEEM core solution's EPC labelling software functionality



Source: NEEM Hub

In the end, the project produced API-compatible software that can be remotely requested to estimate an energy signature model for a given household. The energy signature model provides an EPC label and a diagnostic report. Automated software was not available through most of the project's runtime, and the software had to be executed manually, together with much oversight. While the software was intended for use with district-heated buildings only, it was also successfully applied to buildings with electricity as the energy carrier.

4.2 THE FOUNDATIONAL CONCEPT FOR EPC LABELS

This project focused on EPC label estimation in Denmark, Norway, and Sweden. Based on a building's EPC label and other parameters, an evaluation of renovation potential was produced. While the EPC label in each country is based on a similar foundational concept, the specifics differ greatly and are worth exploring.

The foundational concept for EPC labels is the following:

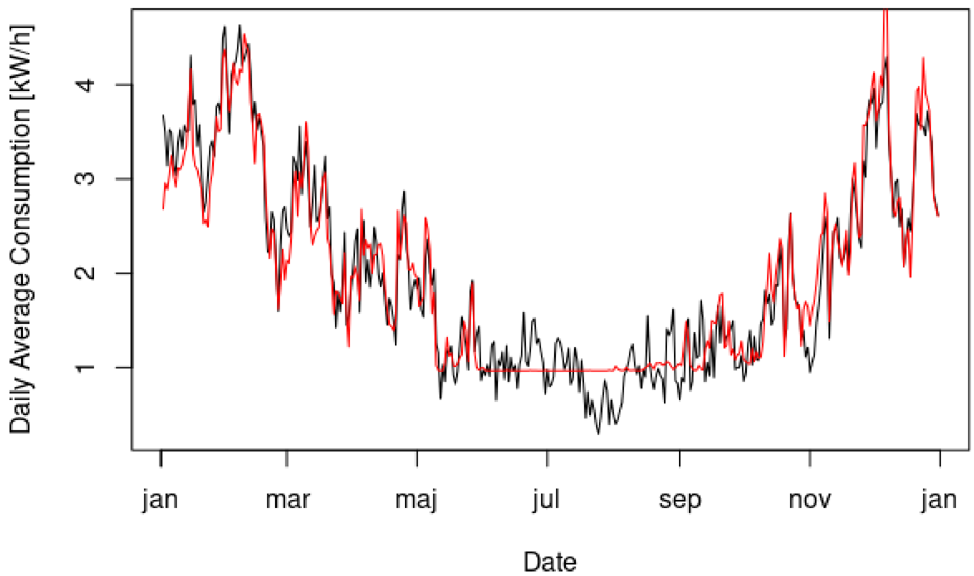
$$EP\ measure = Energy\ Scaling\ Factor \times \frac{Yearly\ Energy\ Consumption}{Building\ Area}$$

The main component is pertinent to model correctly the yearly energy consumption of a given building. Luckily, this measure is not particularly complicated to model. This is because yearly energy consumption can be expressed as a sum of smaller daily energy consumption estimates:

$$Yearly\ Energy\ Consumption\ = \sum_{i=1}^{365}\ Daily\ Energy\ Use$$

As long as a given model gives a mean-zero estimate for daily energy consumption, due to the statistical properties, the error in the cumulative yearly measurement will usually be low. Energy signature models developed by DTU are suitable for this task.

Figure 11
Daily average building energy consumption and simulated estimate



Source: NEEM Hub

Figure 12 depicts a representative example of the observed energy consumption of a building (black) and the simulated estimate (red) by the static energy signature model. This data is from a Danish building used in the NEEM Hub. Note the regime switch between the two parts of the model in mid-May, where the mode switches from being weather dependent to a constant estimate. As is clear, even the static model describes the energy consumption of a given building adequately well for the task of estimating yearly energy consumption.

It is worth noting that modelling a building's energy consumption is a much easier task with district-heating data, as it is less prone to bias from utility use from the residents. Nevertheless, this project also attempted to apply the DTU modelling framework in Sweden and Norway on buildings heated via electricity as the mean heat carrier. This required more care, which will be covered later in this report.

Lastly, while all countries' guidelines agree on what yearly energy consumption is, they differ in every component in estimating the energy performance (EP) measure. The following sections provide a brief yet important overview for estimating a building's yearly energy consumption and then specifics of EPC labels in each country. This knowledge is important when contextualising the pilot study. The main takeaway is that the EPC labels from the three countries are in no way inter-compatible, with the Swedish EPC labelling being by far the harshest.

4.3 DANISH EPC LABELS

The Danish and Norwegian EPC labels are awarded according to nearly identical methodologies. However, the values to which the $EP\ measure$ is compared vary. The following section provides a necessary overview and comparison of the label's definitions in the three countries, which in turn influences how recommendations for the sale or renovation of the building are given.

4.3.1 EPC label formulas for Denmark

Generally, an EPC label is awarded according to which threshold value the $EP\ measure$ falls under, subject to area correction. More explicitly,

$$EP\ measure \leq (Maximum\ EP + Area\ Correction).$$

The $Maximum\ EP$ threshold values for Danish buildings can be found in the following table (where *Area* is the useful area of the building in $[\boldsymbol{m^2}]$).

Table 1
EPC label thresholds for Denmark

EPC	Maximum EP $[kWh/(m^2 year)]$	Area Correction
A2020	27	
A2015	30	$+1000/Area$
A2010	52.5	$+1650/Area$
B	70	$+2200/Area$

C	110	$+3200/Area$
D	150	$+4200/Area$
E	190	$+5200/Area$
F	240	$+6500/Area$
G	∞	

Source: DTU

Additionally, the primary energy factors used to calculate the $EP\ measure$ for Danish buildings are as follows:

Table 2
Primary energy factors for Danish buildings

Energy Carrier	**Primary Energy Factor**
Electricity	1.9
District Heating	0.85

Source: DTU

4.4 NORWEGIAN EPC LABELS

The Norwegian approach is the simplest one considered in this project. It is largely similar to the Danish EPC labelling system. However, there is only a single A-standard, primary energy factors are not used (Brekke et al.) and the values for EP categories are different.

Table 3
EPC label thresholds for Norway

EPC	**Maximum EP** $[kWh/(m^2 year)]$	**Area Correction**
A	95	$+800/Area$
B	120	$+1600/Area$
C	145	$+2500/Area$
D	175	$+4100/Area$
E	205	$+5800/Area$
F	250	$+8000/Area$
G	∞	

Source: DTU

4.5 SWEDISH EPC LABELS

As with the others, the Swedish EPC labels depend on the EP measure. What sets them apart is that the EP is compared to that of newly built buildings. For example, for newly-built single-family homes as of 26/01/2023, the EP is 90.

EPC label categories are defined as a fraction of the required energy performance. Concrete thresholds are shown in Table 4.

Table 4
EPC label thresholds for Sweden

EPC	Fraction	Minimum EP	Maximum EP
A	0-50%	0	45
B	50-75%	45	67.5
C	75-100%	67.5	**90**
D	90-135%	90	121.5
E	135-180%	121.5	162
F	180-235%	162	211.5
G	>235%	211.5	∞

Source: DTU

Some common energy scaling factors are shown in Table 5 below.

Table 5
Energy scaling factors for Sweden

Energy Carrier	Primary Energy Factor
Electricity	1.6
District Heating	1
Gas	1

Source: DTU

As Sweden is a rather large country, across its length the climate shifts toward harsher in the north. As such, the yearly energy consumption score must be offset by a geographical adjustment factor in the following way:

$$Adjusted\ Yearly\ Energy\ Consumption = \frac{Yearly\ Energy\ Consumption}{Geographical\ Adjustment\ Factor}$$

Some examples of geographical adjustment factors are shown in Table 6.

Table 6
Geographical adjustment factors for Sweden

County	Municipality	Geographical Adjustment Factor
Norrbotten	Kiruna	1.9
Stockholm	All Municipalities	1.0
Skåne	Lund	0.9

Source: DTU

During January 2023, NEEM Hub carried out a pilot study on 12 Swedish private family homes. Notable differences between the Swedish system and the other two countries are twofold. First, there is no building-area adjustment for the EPC label, which generally makes the criteria more stringent. Second, there is a geographical adjustment, which this project believes to be a valuable addition. If a building is located in a more northern part of the country, given the same modern insulation it will still produce a larger power draw. Geographical adjustment facilitated easy building comparison between regions.

While the Swedish EPC label system adjustments are welcome, they unfortunately make the calculations for savings potential in a building's renovation more complicated, which was an issue that Copenhagen Economics encountered in the last pilot study.

Chapter 5
MODELLING AND SOFTWARE CHALLENGES

This chapter presents some of the particular challenges encountered in the process of estimating data-driven EPC labels for the NEEM project.

5.1 NONLINEAR ESTIMATION PROBLEM

At first glance, the static energy signature model appears to be simple in appearance: a mixture of two semi-linear regimes. However, that is very much not the case. Every weather dependency apart from temperature is nonlinear. Coupled with hidden-state estimation for the extended models, the problem becomes nonlinear hidden-state estimation that requires nonlinear optimisation and other machine-learning techniques.

This, of course, was a substantial challenge to overcome as if the models do not converge towards true building characteristics, the resulting buildings assessment by the NEEM project risks being false. As the goal was to apply the methods in the real world using actual customers of Scandinavian banks, the model robustness was paramount. The wide-ranging Danish pilot study served to highlight the existence of numerical edge cases that led to improvements in the optimisation procedure.

The main issue was unforeseen numerical stability issues when using automatic differentiation methods for accelerated gradient descent. The current set of solutions for these problems is a clever choice of optimisation algorithm, an educated initial guess for all parameters coupled with statistical validation of the simulation results on training data. However, this still fails in some cases, resulting in the model not converging. In those cases, the results are not provided. This challenge was present in all three pilot studies.

5.2 C++ AND SCALABILITY

For the software to be widely applicable and scalable, it is paramount that the model computation be short. In the context of the Danish pilot study, where the NEEM project evaluated over 20,000 buildings, the difference in computing time for a simple building changing from a minute to a few seconds becomes a difference from five hours to two weeks when computing the pilot study results.

For the further large-scale application of the software, it is unfeasible to have a model estimation for single buildings last more than a second, so accelerating the computation is necessary. However, this comes with numerical stability issues that are difficult to account for and catch autonomously. As the software is tested on an increasing number of buildings and validated by comparison to other models, more errors are caught. This means that the software in its current state requires continuous development. However, most major issues have been removed and the software is ready for deployment.

The current implementation of model estimation relies on the open-source R software package TMB developed at DTU Aqua. This package allows the user to implement random effect models through simple C++ templates that can take advantage of the automatic differentiation of the model loss function. This allows for highly sped-up optimisation, taking time spent on a single model estimation from minutes to seconds.

5.3 HOUSE APPLIANCES, HEAT PUMPS AND ELECTRIC VEHICLES

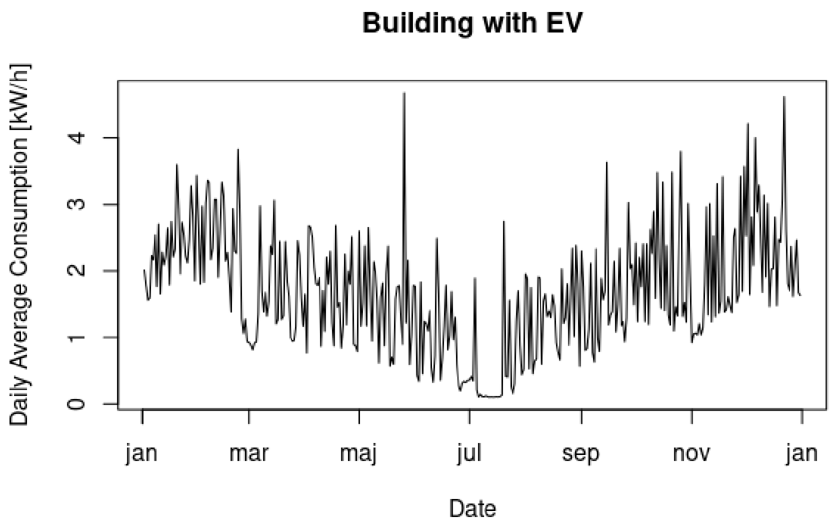
While the main objective of the project was to develop software to evaluate buildings supplied by district heating, the project carried out additional pilot studies on buildings where the main energy carrier is electricity, the cases from Norway and Sweden. The biggest hindrances in estimating the accurate EP of buildings heated by electricity are accounting for home appliances, heat-pump efficiency, and electric vehicles.

Generally, the combined energy draw of typical home appliances such as fridges, dishwashers, washing machines and televisions is negligible compared to the energy needed to heat the building through most of the year. This is the case in Nordic countries but would be a more substantial consideration in central and southern parts of Europe. In the opinion of this project, it is reasonable to include typical home appliances in the heating envelope of the building.

What is more difficult to consider is the efficiency of air-to-water or geothermal heat pumps that use electricity to heat the building. Every heat pump has a coefficient of performance. This coefficient is not known and is subject to change with the weather. In cases when the buildings use electricity through heat pumps for heating, the energy signature model says more about the weather's influence on the heat pump and not the building. While it is generally acceptable to assess buildings with heat pumps as a singular entity and assign an EPC label, the model becomes less useful in its diagnostic capabilities.

The most disruptive and challenging home appliance to account for when processing electricity data is an electric vehicle (EV).

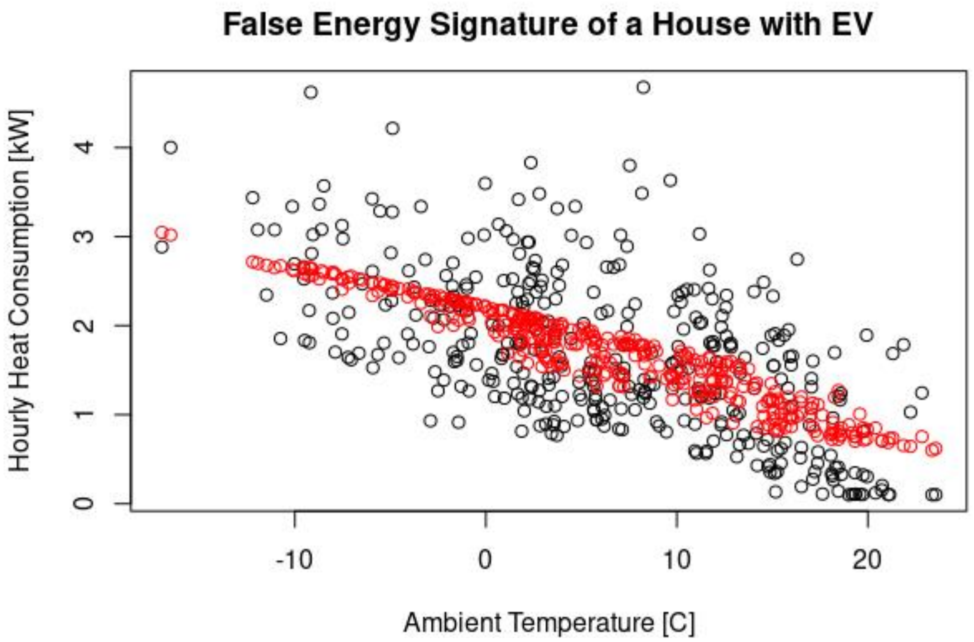
Figure 12
Daily average electricity consumption of a building with an EV over a year



Source: NEEM Hub

Figure 12 depicts observations obtained from the Swedish pilot study carried out by the NEEM Hub. The figure shows the daily average energy consumption of a building heated by electricity paired with an EV charger. With the addition of an EV, the daily energy use readings gain a large but consistent variability. If the data are kept as is, the energy signature model will be ill-fitted.

Figure 13
False fit of the static energy signature model on the data of a building with an EV



Source: NEEM Hub

Figure 13 depicts data from the same Swedish pilot study as in Figure 13, showing a false fit of the energy signature due to the inclusion of an EV in the data. The main impact of an EV is in artificially inflating the baseline energy draw of the building. This mostly presents itself in a large upward bias in baseline energy consumption, removing significance in less significant interactions with the weather (solar, wind) and a small bias in response to temperature changes (HLC).

The project produced several ways of dealing with these issues. However, all of them are difficult to automate reliably. The best option is to have data available directly from the heat source, e.g., district heating or the energy gauge on the private heat pump. In that way, the EV and other utilities are not included in the data.

If direct measurements are not an option, energy consumption data with higher time resolution (hourly, as opposed to daily) could be manually processed. As EV power draw is usually highly pronounced, constant, and can generally be spotted by looking at hourly data, with some effort, these observations can be removed from the dataset manually.

However, this might not be an option if a country has fixed grid fees to encourage people to spread their electricity consumption as evenly as possible throughout the day, i.e., hourly electricity overdraw penalties in Norway. In these cases, the smart EV charger will spread the draw through many hours, often for 12-hour periods at a time. This means that nearly half of the data must be disregarded.

In this case, it is possible to resort to the fact that EV charging will have a limited biasing effect on a building's response to temperature changes. Few countries (such as Sweden) have reference tables for EPC labels based on the HLC of the buildings, and thus a crude estimate of EPC labels can still be issued.

5.4 DATA SCARCITY

Building energy consumption consists of two regimes, one describing heat consumption during periods of heat demand and one describing heat consumption during periods of constant heat consumption (e.g., summer). Therefore, to fit an accurate energy signature model, a dataset overlaps periods where both regimes are active. However, it is more important that some observations of the weather-independent regime are present, as their lack can make the model underestimate the overall heat consumption.

The section on the performance of extended models highlighted that the models require validation datasets to confirm that the models were not over-fitted on the training data set and provide an unbiased insight into building heat dynamics. This naturally requires more data, which is not always available. Regardless of this, the project found that the base static model often describes energy demand sufficiently well with comparatively little data and is always an available option.

Chapter 6
CONCLUSION

One of the NEEM project's goals was to develop a scalable and modular software platform for the estimation of EPC labels for buildings supplied with district heating. The project succeeded in this goal by creating a software environment based on energy signature models and expanding them. The models that were developed as part of this project extract buildings' dependence on the three main weather influences, namely temperature, solar radiation, and wind. The extracted information on buildings' heat-demand characteristics can then be used as an initial diagnostic of different components of the building and for accurate long-term energy demand forecasts. The yearly energy demand forecasts are used in the estimation of the EPC label for the building.

Before the models were applied in testing within the NEEM Hub, they were validated on Norwegian school buildings where great control and detailed measurements of the heating system were available.

After validation, the software was applied to EPC label estimation in Denmark on buildings supplied with district heating. Additionally, two pilot studies were performed in Norway and Sweden, where the software was applied to buildings where the main energy carrier was electricity. This report highlights the different approaches to EPC labelling in each country and additional considerations necessary when working with electricity data. The results from these cases are presented in Deliverable 5.2.

Lastly, the report highlighted substantial challenges it faced in the implementation and testing of the software on a large number of buildings. The main issues are the requirements for nonlinear optimisation methods for the models, the necessity for C++-based automatic differentiation for scalability, the numerical issues present in that, and finally, data scarcity. These issues were all tackled and solved during the project's development; however, some are still benefitting from active development.

The developed software can solve the basic EPC labelling problem through a data-driven scalable implementation. Furthermore, it shows how building performance characteristics can be estimated in various ways, from simple to detailed building diagnostics with time-varying parameters tracking changes over time. This research can lead to deep diagnostic applications in future projects.

REFERENCES

Rasmussen, C.; Bacher, P.; Calì, D.; Nielsen, H.A.; Madsen, H. Method for Scalable and Automatised Thermal Building Performance Documentation and Screening. Energies 2020, 13, 3866. https://doi.org/10.3390/en13153866

Frédéric Haldi and Darren Robinson. The impact of occupants' behaviour on building energy demand. Journal of Building Performance Simulation, 4(4):323–338, 2011.

Henrik Brohus, Per Heiselberg, Allan Simonsen, and Kim C Sørensen. Influence of occupants' behaviour on the energy consumption of domestic buildings. Aalborg University, Denmark, 2010.

Robert H. Socolow. The Twin Rivers Program on Energy Conservation in Housing: Highlights and Conclusions*. Energy and Buildings, 1(3): 207–242, 1977.

Hiroshi Yoshino, Tianzhen Hong, and Natasa Nord. IEA EBC Annex 53: Total energy use in buildings – Analysis and evaluation methods. Energy and Buildings, 152:124–136, 2017.

Zhun Yu, Benjamin CM Fung, Fariborz Haghighat, Hiroshi Yoshino, and Edward Morofsky. A systematic procedure to study the influence of occupant behaviour on building energy consumption. Energy and buildings, 43(6):1409–1417, 2011.

J Wingfield, D Miles-Shenton, and M Bell. Comparison of Measured Versus Predicted Heat Loss for New Build UK Dwellings. Unpublished Data, Leeds Metropolitan University, Leeds, UK, 2011.

Energistyrelsen (The Danish Energy Agency). Status for Energimærkningsordning for Bygninger. Technical report, Energistyrelsen (The Danish Energy Agency), 2018.
URL {https://ens.dk/sites/ens.dk/files/Energimaerke/status_for_energimaerkningsordningen_for_bygninger.pdf}.

Christoffer Rasmussen, Peder Bacher, Davide Calì, Henrik Aalborg Nielsen, and Henrik Madsen. Method for scalable and automatised thermal building performance documentation and screening. Energies, 2020.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

Kasper Kristensen, Anders Nielsen, Casper W. Berg, Hans Skaug, and Bradley M. Bell. TMB: Automatic differentiation and Laplace approximation. Journal of Statistical Software, 70(5):1–21, 2016. doi:10.18637/jss.v070.i05.

Uffe H. Thygesen, Christoffer M. Albertsen, Casper W. Berg, Kasper Kristensen, and Anders Nielsen. Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics, pages 1–23, 2017. doi: 10.1007/s10651-017-0372-4.

Copernicus Climate Data Store. ERA5-Land hourly data from 1950 to present. https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview, 2022. Accessed:2023-03-14.

Brekke, Tor, et al. "EPBD implementation in Norway." 2016.