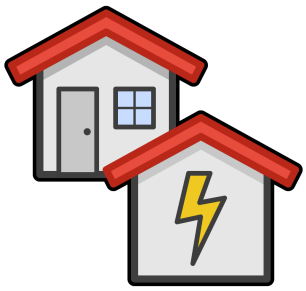
Energy Twin

Energy Twin model description



Jan Široký, PhD
Martin Zázvorka

04/2021

1 Introduction

Energy Twin (ET) is a machine learning SkySpark extension for energy consumption analysis designed for efficient multiple building monitoring using artificial intelligence. ET aims to identify problems and reveal the potential for future energy consumption savings and optimization.
The ET model is a model based on the TOWT model[1]. The dependent variable is typically measured power. However, it could be, for example, also water consumption or temperature of domestic hot water etc. The independent variables are always outdoor temperature, occupancy, time of the week. No other independent variables can be included in the ET model.
ET prediction is given as a composition of two loads: weather dependent load and time dependent load. Both loads are visualized separately in various views in ET (see Figures 1.1 and 1.2).

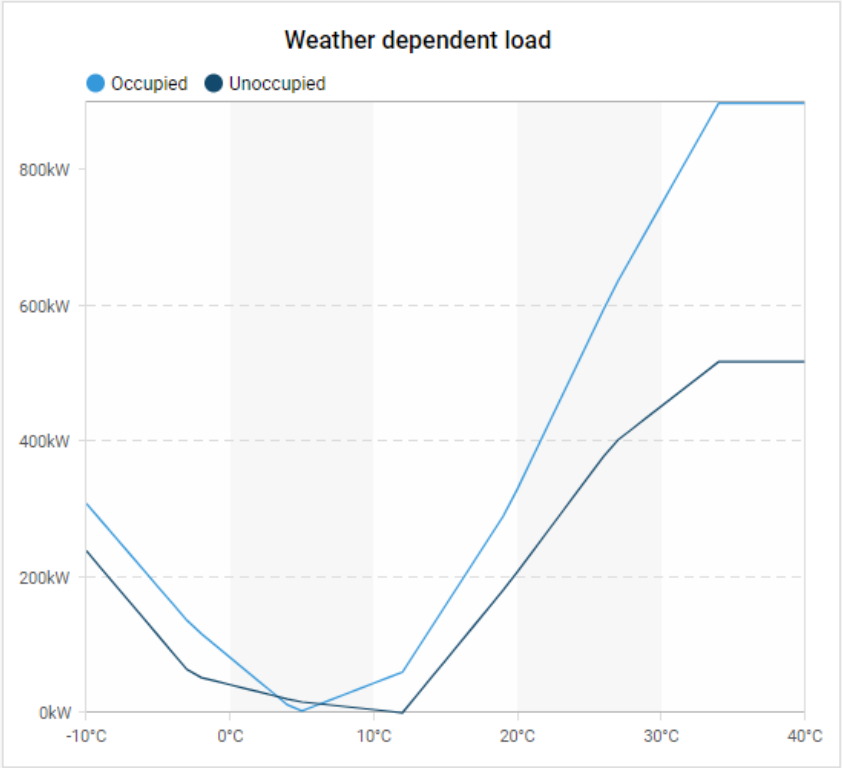


Figure 1.1: Weather dependent load

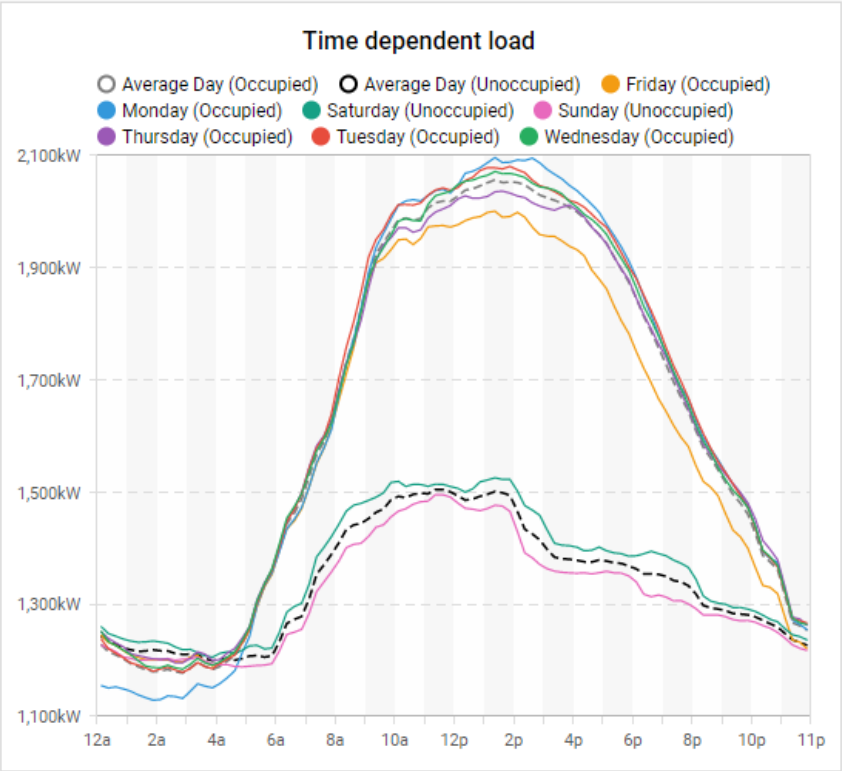


Figure 1.2: Time dependent load

2 ET model structure

One ET model consists of eight linear submodels in total. The set of submodels is determined by all combinations of the three following categories (e.g., there is a submodel representing combination: unoccupied days - full model - deterministic):

- **occupancy**
  - occupied days (e.g., workdays in case of an office building)
  - unoccupied days (e.g., the weekend in case of an office building)
- **weekday aggregation**
  - full model (each weekday is modeled separately)
  - simple model (days are aggregated)
- **model components**
  - deterministic (predicts mean value)
  - stochastic (predicts standard deviation)

Each of the above-mentioned categories will be described in detail in the following sections.

2.1 Occupancy

The ET model focuses on the energy consumption of buildings. The consumption is closely related to the building regime. ET, therefore, introduces two distinct regimes of building use. A set of linear submodels captures each regime. The prediction for a particular time step is provided by one of the models according to the specific day's occupancy. Note that the SkySpark schedule captures occupancy in a minute's detail. The ET aggregates occupancy data on a daily basis. If there is no occupancy, then the particular day is unoccupied; otherwise, the day is occupied.
Tip: Occupancy can be used for distinguishing any two distinct regimes, not necessarily only occupied and unoccupied regimes. Using this trick, one can handle a non-linearity in the data - for example, completely different regimes given by the use of energy-intensive appliances. Occupancy, in such a case, can represent information whether the particular device was on or off.

2.2 Weekday aggregation

The full model has a set of parameters for each weekday. As a result, one can distinguish, e.g. difference between an occupied Monday and Friday or an unoccupied Saturday and Sunday. ET also handles holidays and other special events using SkySpark schedule.
Such as, e.g., one Wednesday in the training period that was not occupied (i.e., there was precisely one holiday on Wednesday in the training period). In this case, ET will model unoccupied Wednesday according to this particular Wednesday in the training period. However, one would expect that any future unoccupied Wednesday would look more like an "average" unoccupied day than the exact Wednesday from a training period. For this purpose, ET uses the simple model. It aggregates all occupied days (or unoccupied days) and identifies an average occupied (occupied) day (see Figure 2.1). Weather dependent model parameters for the simple model are the same as for the full model.

ET prediction, by default, takes into account the amount of data used for identification. When the identification data set is small (e.g., just one unoccupied Wednesday in the training period), ET uses the simple model instead of the full model for that particular day (e.g., unoccupied Wednesday in the future).

Occupied regime parameters	Unoccupied regime parameters
Weather dependency	Weather dependency
Intercept	Intercept
Monday	Monday
Tuesday	Tuesday
Wednesday	Wednesday
Thursday	Thursday
Friday	Friday
Saturday	Saturday
Sunday	Sunday
Average occupied day	Average unoccupied day

Figure 2.1: ET model parameters for an office building with one unoccupied Wednesday in the training period. Green cells indicate a sufficient amount of data for estimating parameters (e.g., 52 occupied Mondays per year). Red cells indicate no data for estimating parameters. For example, there were no data for an occupied Sunday or an unoccupied Monday. Lastly, yellow cells indicate that the amount of data was insufficient for reliable parameters estimate (the simple model will be used for prediction instead of the full model).

2.3 Model components

ET's primary goal is to predict the expected value of the dependent variable given all independent variables. ET also takes the uncertainty in input data into account. Therefore, it can also provide information about data variability. See confidence intervals in Figure 2.2. One can observe a small variance in the Sunday early morning and a high variance in the Monday midday.
It is expected that residuals (difference between the measured and predicted value) are Gaussian. The stochastic component is then captured by the standard deviation for each time interval [1].

[1] The ET team has spent a lot of effort on modeling uncertainty as a function of both time of the week and weather. However, no feasible solution that could be implemented in SkySpark has not been found yet. If you are interested in this research topic and you want to contribute or share a common interest, please do not hesitate to contact the ET team.

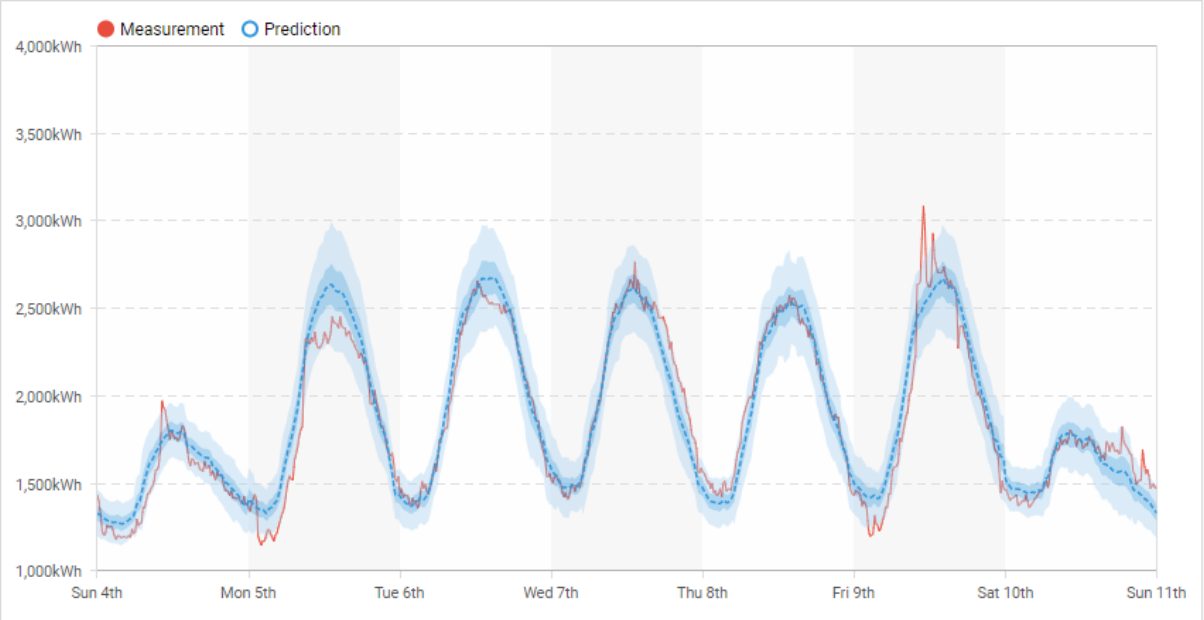


Figure 2.2: Confidence intervals, the dark blue interval is defined by $mean \pm 1 \cdot std$, while the light blue is defined by $mean \pm 3 \cdot std$. In other words, 68% of all measured values shall lay in the dark blue area, and 99.7% of all measured values shall lay in the light blue area. The graph captures one week's worth of data using a model with a sampling period of 15 minutes. One could expect two measured values to lay outside the light blue in average in such a case.

3 Implementation

The model identification is performed by a custom Fantom code that uses SMILE[1] library that is distributed as a standard SkySpark library. ET does not share any data and is not dependent on any external libraries or cloud services.

[1] https://haifengl.github.io/

Bibliography

[1] Price P, *Methods for Analyzing Electric Load Shape and its Variability*. Lawrence Berkeley National Laboratory Report LBNL-3713E, May 2010.