B.Sc. Thesis
Bachelor of Science in Mathematics and Technology

DTU Compute
Department of Applied Mathematics and Computer Science

Statistical models for analysis of frequent readings of electricity, water and heat consumption from smart meters

In cooperation with SEAS-NVE

Anton Stockmarr (s164170)
Ida Riis Jensen (s161777)
Mikkel Laursen (s164199)

Kongens Lyngby 2019

DTU Compute
Department of Applied Mathematics and Computer Science
Technical University of Denmark

Matematiktorvet
Building 303B
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Preface

This xxx thesis was prepared at the department of Applied Mathematics and Computer Science at the Technical University of Denmark in fulfillment of the requirements for acquiring a yyy degree in zzz.

Kongens Lyngby, June 3, 2019

Anton Stockmarr (s164170)
Ida Riis Jensen (s161777)
Mikkel Laursen (s164199)

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

CHAPTER 1

Introduction

1.1 Motivation

1.2 Introduction to WATTS app

CHAPTER 2

Data

The data is provided by SEAS-NVE in three data sets. The house data consists of 71 `.csv`-files containing 8 attributes for each house which is 513877 data points in all. The second data set includes weather data containing 10,140 observations and predictions of the next 2283 data points, all with 11 attributes. Furthermore, the third data set is from Bygnings- og Boligregistret (BBR) and contain details for each of the houses e.g. total area, year of construction and type of house. The main focus of this section will be how this data is prepared for the further analysis.

2.1 Original data

The original house and weather data include hourly observations from the period 31-12-2017 23:00 to 7-02-2019 10:00. The time period varies in the house data which will be taken into account when cleaning the data.
Table **??** below shows the attributes from the house data set.

Variable	**Description**
StartDateTime	Start time and date for measurements.
EndDateTime	End time and date for measurements.
Energy	Electricity consumption in kWh.
Flow	Amount of water passed through meter in $m^3/hour$.
Volume	in m^3.
TemperatureIn	Temp. of the water flowing into a house in Degrees/C.
TemperatureOut	Temp. of the water flowing out of a house in Degrees/C.
CoolingDegree	Difference between Temp.In and Temp.Out in Degrees/C.

Table 2.1: Attributes from the original house data..

The Heat Consumption is defined as

$$Q = V \cdot \Delta T \tag{2.1}$$

The weather data set consists of the attributes seen in Table **??**.

Variable	Description
StartDateTime	Start time and date for measurements. Hourly values.
Temperature	Temperature outside in Degrees/C.
WindSpeed	Wind speed in m/s
WindDirection	Wind direction i degrees from 0 to 360, 0 being North
SunHour	The level of sunshine in the hour in a scale from 0 to 1
Condition	
UltravioletIndex	The UV index level
MeanSeaLevelPressure	
DewPoint	
Humidity	
PrecipitationProbability	
IsHistoricalEstimated	Binary variable, true if the datapoint is a prediction

Table 2.2: Attributes from the original weather data..

The BBR data set consists of the attributes seen in Table **??**.

Variable	Description
Key	The house ID key
HouseType	Type of house: Apartment, house, industrial etc.
TotalArea	The total area of the house in m^2
Floors	The number of floors in the house
Basement	How many m^2 basement there is in the house
Attic	How many m^2 attic there is in the house
ConstructionYear	The year of construction for the house
Surfaces	The material on the surface of the outdoor walls of the house
ReconstructionYear	The year of the latest reconstruction of the house
AdditionalHeating	If there are any additional heating installed in the house. Fireplace etc.

Table 2.3: Attributes from the BBR data..

2.2 Cleaning and preparation

In this section, it is described how the raw data is cleaned and prepared for the statistical analysis.
Due to the fact, that `StartDateTime` and `EndDateTime` is always one hour apart, it is redundant to use both of the attributes. The obervations of most of the attributes are made at time `EndDateTime`, and for that reason it is used as `ObsTime` for the observations. For the weather data set, the observations is made at time `StartDateTime`, and there is no `EndDateTime` for this data set. When merging these data sets, `ObsTime` is

alligned with `StartDateTime`. The format of these attributes is changed to a `Posixct` value with d-m-Y H:min:sec as the structure.
Every now and then, one or more data points in a row are missing. When this happen, a data point with NA-values for all of the attributes except `ObsTime`, is placed in the data set, which makes the data set easier to use in the modelling process. In the data sets there are no indication of whether or not it is weekend. This attribute is added as well as the school holidays.
Both weather data and the house data are aggregated with mean values for each day in order to convert hourly values into daily values since there are of interest when modelling in chapter 3, two of the attributes is aggregated in a different way, which is explained later.

In the house data there are some measurements missing and it can therefore be difficult to do modelling for the houses in question. To avoid these difficulties, a so called "Data Checking" function has been made in order to check whether several constraints for the data are fulfilled. There must be a certain number of observations and the amount of missing data should not exceed a certain fraction of the data observation period.

2.2.1 The sun and the wind

A physical factor that could possibly affect the heat consumption is the sun. In raw data, the attributes `Condition`, `SunHour`, and `UltraVioletIndex` can be seen as explanatory variables for the sun. Instead, an attribute, `Radiation`, is added to calculate the solar radiation for a given day. This attribute is determined with use of the `R` function `calcSol` from the library `solaR`. The ultraviolet index is a measurement of the strength of ultraviolet radiation and since the attribute `Radiation` is more exact, `UltravioletIndex` is removed from the weather data set.

Another physical factor that might be of importance is the wind. There are data available for both the wind direction in degrees and the wind speed. When the data is aggregated into daily values, it is important to pay special attention to the wind attributes, since it is not logical to take the average of degree values. For example, the average wind direction of 359 degrees and 0 degrees is not 179.5 degrees. Instead the wind direction and wind speed are interpreted as polar coordinates in a coordinate system. They are converted to rectangular coordinates. Then they are aggregated from hourly values into daily values, and returned to polar coordinates. When the wind is aggregated this way, wind directions with high wind speeds are weighted higher than wind directions with low wind speeds. Also the problem with the periodicity of the wind direction is solved.