

 **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, March 15, 2019

A handwritten signature in black ink, consisting of a large, stylized 'A' followed by a series of loops and a final flourish.

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

Data

The data is provided by SEAS-NVE in two data sets. The house data consists of 69 .csv-files containing 8 attributes for each house which is 499,499 data points in all. The second data set includes weather data containing 11,845 observations with 11 attributes. **Noget med hvordan data er blevet målt - hvilket udstyr, af hvilken virksomhed osv.** The main focus of this section will be how data is prepared for the further analysis.

1.1 Original data

The original house and weather data include hourly observations from the period 31-12-2017 to 29-01-2019. 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. Hourly values.
EndDateTime	End time and date for measurements.
Energy	Electricity consumption in <i>kWh</i> .
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 11: Attributes from the original house data..

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	
WindDirection	
SunHour	
Condition	
UltravioletIndex	
MeanSeaLevelPressure	
DewPoint	
Humidity	
PrecipitationProbability	
IsHistoricalEstimated	

Table 12: Attributes from the original weather data..

1.2 Cleaning and preparation

In this section, it is described how the raw data is cleaned and prepared for the statistical analysis.

Data er aggregated for at lave timeværdierne om til dagsværdier

Loader en temporary data ind, som vi modificerer indtil vi putter den ind i vores endelige data. Vi sætter navnet på den første attribute til StartDateTime. Vi ændrer formatet på de to første attributes til posix, som er `%d - %m - %Y%H : %M : %S`.

Så fjerner vi data fra 2017, fordi vi ikke har noget weather data der. 21 observationer.

For nogle huse er der nogle hourly measurements der ikke er der. Der er huller i målingerne. Disse udfyldes med null, hvilket er bedre/lettere at arbejde med.

enddays og startdays sættes for hvert hus - hvornår starter målingerne og hvornår slutter målinger. Tidspunkterne for aller første og aller sidste måling.

StartDateTime i weather formateres til rette format, så det passer med house data.

Attributen IsHistoricalEstimated ændres til logical, så vi kan compute med den.

Vi laver så temp. weather data så vi kan merge det med house data. Vi merger ikke al data, da mængden vil være en del større. Vi merger tmp weather data på house data i model processen.