Mat-2.4177 Seminar on Case Studies in Operations Research
Fortum
Project plan

Lasse Lindqvist
Juulia Happonen
Joona Karjalainen
Mika Juuti

March 13, 2014

Abstract

This project work concerns forecasting the demand of district heating for the energy company Fortum.

1 Background

District heating is the main heating form in Finland, having a market share of 46 % of the total heat production. In the largest cities, it covers over 90 % of the provided heat. The global market share of district heating is around 10 %. District heating is used for residential, commercial and industrial purposes. It is mainly used for space heating and hot-water consumption. Its applications also include space cooling, which can be used for cooling of data centers, for instance.

The demand of heat is affected by outside temperature and social behaviour, among other factors. By taking these factors into account, it is possible to predict the heat demand using appropriate mathematical models. The resulting forecasts can be used to adjust heat production.

2 Objectives

The objectives for this project work are creating a working forecast model for district heating and providing suitable instructions of how to use the model in future real forecasting cases. Additionally a working program for forecasting and handling data will be implemented in the R language.

Earlier research will be studied and possible additions to the model will be considered. A literature overview will cover general information on district heating and other methods used to estimate the demand. The report will consider possible effects of weather conditions based on literature. Some prospects of the development of the infrastructure will also be discussed.

3 Tasks

3.1 Sources of uncertainty

Temperature information is in some respect uncertain. Forecasting the demand in the future would require a forecast of the temperature too. Furthermore, a single temperature measurement station is used for all buildings, some of which might be located quite far away from the measurement station.

If the temperature forecasts are considered more uncertain than the heat demand model, then improving the quality of the temperature forecast becomes a priority.

3.2 Data

The data we have in our use contains usage data from 2013. The usage is recorded hourly giving us 8760 data points for every building. Temperature data is also hourly data collected from one measurement station. The essential information can be considered to be:

1. The quantity to be forecasted (MWh).
2. Characteristics of each building
   - Heated volume (m^3)
   - Year of construction
   - Type of building
3. Temperature is from Sepänkylä measurement station in Espoo.

3.2.1 Omitted data and comments

The original data defines the building volume, floor area and heated floor area for each building. The heated volume gives a more appropriate estimate of the building's heat consumption than the floor area, and the floor area will thus be omitted. The year of construction surely affects the energy efficiency of the house and thus also the heat consumption in some way.

The original data also contains information such as the year of restauration (if any). The year of reconstruction is not reported in some cases. Additionally the term causes some ambiguity; in some cases even repainting the building is considered restauration. Thus, the reconstruction year will be omitted.

The temperature measurements contain some blank measurements. To use the data for modeling these blanks must be filled. In the lack of greater knowledge, these values are to be interpolated. When using the model for real forecasting, forecasts for temperature are also used. Usually forecasts for every hour should be available.

The same goes for consumption data. This data does not contain blanks, but it does contain erronous values. These values too must be replaced with interpolated values. Automating consumption correction could be a hard problem, but an important one. Detecting and correcting errors has to be implemented. We will describe in the report how erroneous values can affect the resulting model and give unacceptable results.

3.3 Clustering

The data available is quite detailed. Therefore it is important to identify possible clusters which behave accordingly. Forecasting for each cluster independently provides a more powerful forecast than using the same model for all units. Statistical tests are an important tool in identifying independent clusters.

3.4 Finding suitable models

The starting point is an ARMAX model with temperature as the independent variable.

3.5 External variables

The usage of district heating is commonly thought to consist of two factors: social and weather. Of our external variables temperature deals with weather. It is also the only information available, even though wind and sunshine have been suggested to have an effect, too.

Social component is modeled with SARMA model but also possibly with external indicator variable that determines if a day is a midweek holiday or not.

3.5.1 Temperature

It is known that usage of district heating is not linearly dependent on temperature. The effect of the temperature is smaller in the cases of relatively high and relatively low temperatures. Bronislav Chramcov used a polynomial $y = aT + bT^3$. We are considering using this as our starting point to see if the fit is enough or if higher orders (5) are needed.

In Heimo Zinko's report "Marginaler i Fjärrvärmesystem" [1], the outside temperature is modeled to affect the heating of the house after some hours, due to the thermal strength of the house. Based on (cross)correlation analysis, it seems that there is no lag in the usage of central heating when considering all houses. Therefore no lagged values in the outside temperature need to be considered. The temperature forecast on the specific hour could be useful, however.

3.5.2 Midweek holidays

Midweek holidays affect social behaviour. It is unclear if a simple indicator variable is flexible and powerful enough to give the model the possibility to react to midweek holidays.

3.5.3 Lag

External variables may or may not have an immediate effect. Midweek holidays are expected to have an immediate effect, but temperature might not have.

3.5.4 Seasonality

Many things in society have multiple strong cycles. The first cycle is daily and the second weekly. These correspond to 24 and 168 hours. One important thing to do is to determine whether to use a season of 24 or 168 hours in the model.

3.6 Validating and verifying the models

The models are to be validated with different data than they were estimated with.

An important thing to do is to compare the submodels to a single model estimated for the whole grid.

3.6.1 Statistical tests

ARMA models are defined in such a way that they should produce independent normally distributed residuals. Multiple different tests have been devised for testing these assumptions. These tests will be used to validate the models.

3.6.2 Information criteria

Identifying the correct model can be done manually or automatically. R software offers a way to automatically fit different models and calculate information criteria for these.

Possible criteria include Akaike information criterion (AIC), Akaike information criterion for finite samples (AICc) and Bayesian information criterion (BIC). The AIC penalizes the number of parameters less strongly than the Bayesian information criterion (BIC) does. These criteria will be used in model selection.

3.6.3 Measures of error

The performance of the model will be evaluated using selected measures of error. These include Root Mean Squared Error (RMSE), Mean Absolute Percent Error (MAPE) and Mean Absolute Scaled Error (MASE). We will read relevant literature to find the most common ways to measure the error.

4 Schedule

17/1	It was agreed that Lasse Lindqvist was to be the project manager.
24/1	Meeting the contact persons at Fortum. The project was discussed in detail.
28/2	Presentation of the project plan. Earlier research studied and included in the report.
1/3	An example of a forecasting model ready and in the report. Considerations about corrections in data ready.
Mid-March	Analysing the usefulness of chosen groupings. Choosing the amount of past data.
End of March	Mid-report ready.
4/4	Presentation of the mid-report.
Early May	Content of the project and report ready.
16/5	Presentation of the project.

5 Resources

The project group consists of four students of systems and operations research. The group will collaborate with Teppo Luukkonen and Jaakko Luhtala from Fortum to ensure that the focus of the project is in line with the client's wishes.

6 Responsibilities

The project leader Lasse Lindqvist will be the contact person of the group. The responsibilities include contacting the client as needed, making sure the group members have the necessary information and seeing that the project follows the schedule.

Building the model with R will be mostly Lasse Lindqvist's and Mika Juuti's responsibility. Everyone will offer critique about the model, validation and the presentation of the results.

Literature review will mostly be Juulia Happonen's and Joona Karjalainen's responsibility. Juulia Happonen will write an overview of district heating. Joona Karjalainen will write about earlier research on heat demand modelling and alternative approaches to time-series modelling.

7 Risks and considerations

The project itself is very straightforward in its nature. This brings the risk of doing too little. On the other hand, we have a possibility of furthermore bringing up the methodology in applying the model in real forecasting and devising helping rules in its use. The probability of this risk is high. However, the impact may be low, as results obtained with simple models may still be valuable to the client. The main impact would be that the group would not learn enough during the course.

The overview of district heating and combined heat and power might end up being too short if sufficient earlier research and literature are not found. To avoid this, information should be searched from a versatile variety of sources. If suitable methods are still not found, we will only focus on the areas that are covered in literature. The risk is of medium significance.

The outliers and blank measurements in the omitted data can make the quality of data occasionally poor and lower the significance of the results. However, the amount of problematic data is small compared to the overall amount of data. It is unlikely that this would have a critical impact on the model selection. However, choosing a wrong model would be critical to the main results of the project.

Finding the right model for the demand can be time-consuming. In this case, the workload could become greater than expected and the project could be delayed. There is also a risk of the final model being inefficient if a suitable model does not exist. This risk is of high significance and the risk can be reduced by examining and evaluating the possible models at an early stage of the project.

In addition, there is a possibility of individual time management might fail, which can result in project being delayed. To avoid this, the tasks and actions need to be precisely planned. Online working will help to keep track of the progress and to recognize the situations where the workload between the project members is revealed to be unevenly distributed. Failures in time management are unlikely but critical to the success of the project.

References

[1] Patrik Selinder and Heimo Zinko. Marginaler i fjärrvärmesystem. *Forskning och Utveckling 2003*, 85. Svenska Fjärrvärmeföreningens Service AB, 2003.