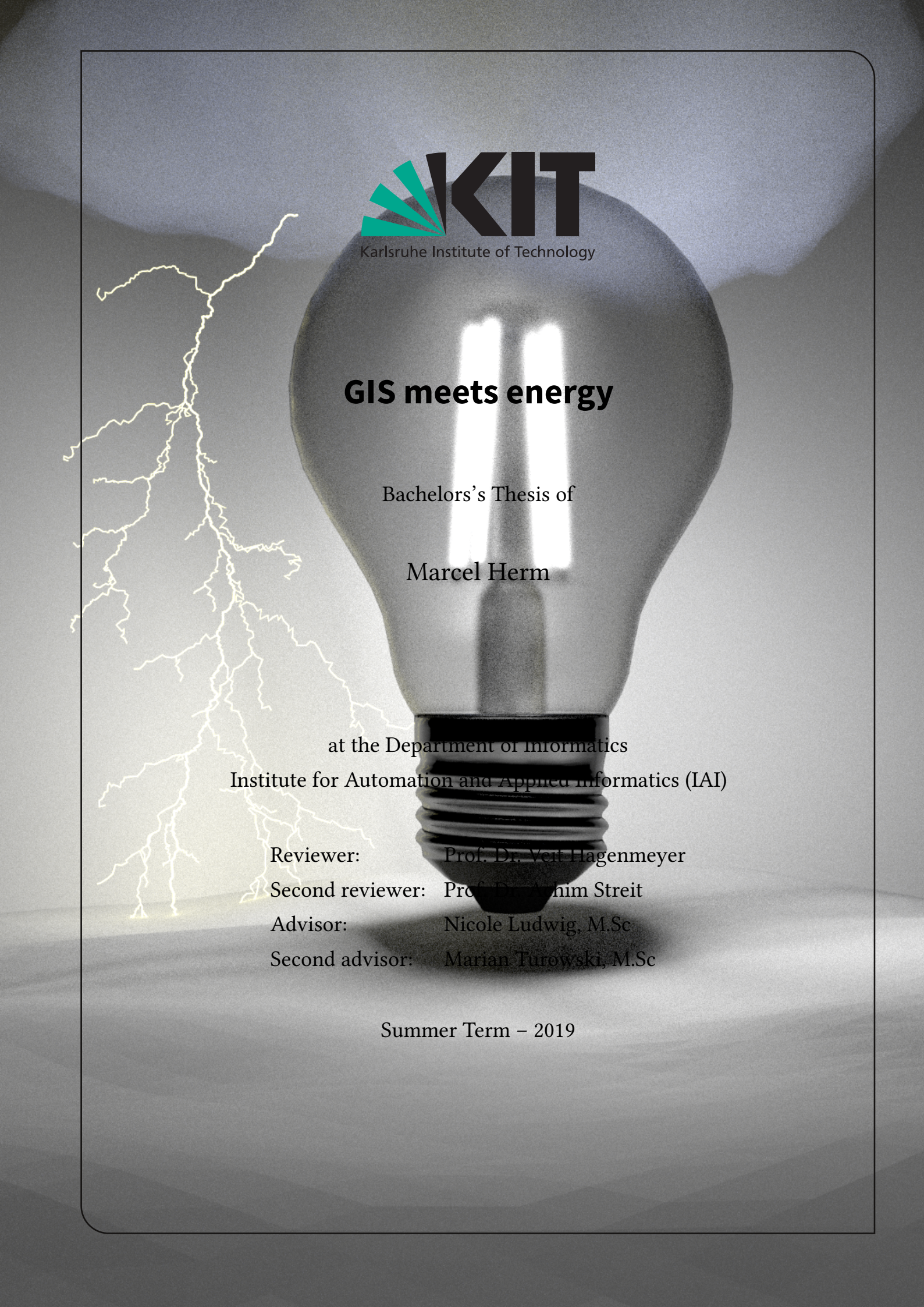

Karlsruhe Institute of Technology

# GIS meets energy

Bachelors's Thesis of

Marcel Herm

at the Department of Informatics
Institute for Automation and Applied Informatics (IAI)

| | |
|---|---|
| Reviewer: | Prof. Dr. Veit Hagenmeyer |
| Second reviewer: | Prof. Dr. Achim Streit |
| Advisor: | Nicole Ludwig, M.Sc |
| Second advisor: | Marian Turowski, M.Sc |

Summer Term – 2019

Karlsruher Institut für Technologie
Fakultät für Informatik
Postfach 6980
76128 Karlsruhe

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.
**PLACE, DATE**

. . . . . .

(Marcel Herm)

# Abstract

As the share of electricity from regenerative sources is growing constantly, the weather becomes an increasingly important factor in the analysis of electricity markets. Hence, this thesis uses local weather data to predict electricity spot prices. More precisely, we include wind speed and temperature from individual German weather stations into time series and statistical learning models. However, as the available weather information is vast and renewable power is not generated everywhere, we use random forests and Bayesian structural time series to perform a feature selection. Overall, we manage to improve our forecasting accuracy of the EPEX electricity prices by up to 7.69 % in terms of root mean squared error and up to 8.19 % in terms of mean absolute error.

# Contents

# List of Figures

# List of Tables

# 1. Introduction

Your thesis should start with an introduction. The introduction is supposed to motivate your thesis. Discuss the relevance of your topic, why are you looking into it, why is it relevant in the field? Cite important research related to your motivation. Briefly state the problem as in the abstract and repeat the contribution, for example in the form of research questions.
Give an outline of your thesis.

# 2. Related Work

Regarding grid based research in the field of energy, there has not been published v

# 3. Methodology

Using weather data from ECMWF Copernicus Climate Change Service (C3S).
Using load data from `https://data.open-power-system-data.org/`.
First downloaded whole Datasets from 2006-2019, but as the load for germany is properly available since 2015, now reduced dataset to 2015-2019.
Also checked for non-existing values, only 2 last timestamps values for the load are missing.

## 3.1. Method 1

Maybe use Random Forests for variable selection as in Nicoles paper? (Ludwig et al. 2015)

You can also use equation numbering if you need to refer to an equation later e. g. Equation (3.1).

$$a^2 + b^2 = c^2 \tag{3.1}$$

Additionally, simple equations can be put inline with the text, for example, $x \in X$. Remember to set all variables in math font i. e. all $x$, $i$ and so on.

## 3.2. Method 2

…

# 4. Evaluation

## 4.1. Research

Searching information always is a key element in research. Therefor, arXiv and Google Scholar were used in order to find suitable reading.
It proofed to be difficult to find such papers using keywords such as "grid-based" or "geographic", because most of the results referred to either other grids, such as in Smart Grid, or completely different geographic research subjects.
As weather data from the European Centre of Medium-Range Weather Forecasts (ECMWF) is used in this thesis, which is grid based, one solution was to search for "ECMWF", as this type of data is widely used within the research field of energy.

also used BASE which proposed "energy network geographic data" after searching a bit, I came to the term "energy network ecmwf" which gave some hits.
One criteria is the title which tells much about the subject of the work. If the title implies both, working with geographic or grid-based data and also has a connection to the field of energy networks, it might be suitable for this work.
Another one is the abstract and/or introduction which tells some more details about the paper. If there are some points about applying geographic data in algorithms to forecast some power production or load, it is most likely that this paper contains some valuable information for this thesis.

## 4.2. Data

### 4.2.1. ECMWF

The data used in this thesis originates from ECMWF, which is a research institute that produces global numerical weather predictions and other data.
It is time series based and for each timestamp there is a 2-dimensional array referred to by

longitude and latitude respectively.

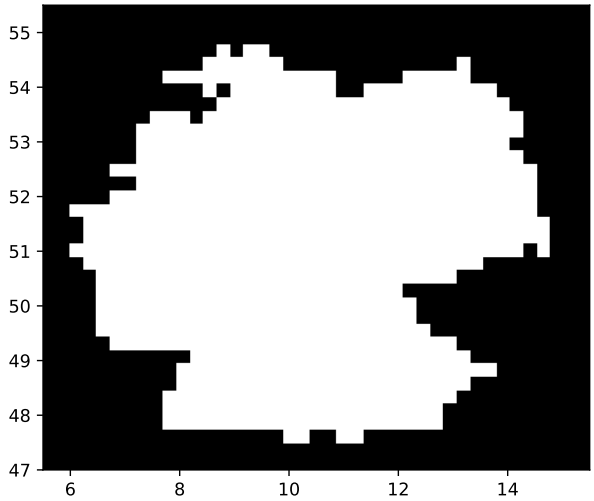


**Figure 4.1.** This figure shows the 2D boolean numpy.ndarray used to filter grid squares that are within germany. It was created by using a shapefile of germany (TODO insert source `https://ec.europa.eu/eurostat/cache/GISCO/distribution/v2/nuts/nuts-2016-files.html`) and checking for each point of the grid if it is within the shapefile. (TODO shorter explanation, put explanation in text)

| variable name | units | min | max |
|---|---|---|---|
| 10 metre U wind component | m s**-1 | -18.56 | 21.92 |
| 10 metre V wind component | m s**-1 | -21.51 | 20.00 |
| 2 metre temperature | K | 240.97 | 313.26 |
| Leaf area index, high vegetation | m**2 m**-2 | 0.00 | 4.90 |
| Leaf area index, low vegetation | m**2 m**-2 | 0.00 | 3.84 |
| Low cloud cover | (0 - 1) | 0.00 | 1.00 |
| Soil temperature level 1 | K | 257.91 | 313.64 |
| Surface latent heat flux | J m**-2 | -2203977.00 | 359411.00 |
| Surface net thermal radiation | J m**-2 | -663417.00 | 142945.02 |
| Surface sensible heat flux | J m**-2 | -1703159.00 | 801354.00 |
| Total cloud cover | (0 - 1) | 0.00 | 1.00 |
| Total column rain water | kg m**-2 | 0.00 | 2.73 |
| Total sky direct solar radiation at surface | J m**-2 | -0.12 | 3088320.00 |

**Table 4.1.** List of exogenous weather variables used to forecast the load including min, max values.

### 4.2.2. Load data

Table 4.1 is an example table. Remember to use full sentences in your caption and explain everything one can see in the table there as well. You can of course also use a simpler format for your table.

## 4.3. Programming part

### 4.3.1. Programming Language

For the programming part, Python3.6+ has been chosen, as there is a variety of libraries to process all used file formats and because it tends to be a time saving language, also for visualization.

### 4.3.2. Documentation

In regard to coding styles, especially when it comes to docstrings, the numpy conventions were used. The three major points for this were first, that it is a popular and often used style, then it is also a visually oriented style which means, that it is easy to read and last it is supported by several (TODO check which, sphinx?!) autodoc tools that create a HTML based documentation from existing source code with docstrings.

## 4.4. Results

Describe the results you have obtained using your methods described above. Again use proper visualization methods.

### 4.4.1. Experiment 1

...

### 4.4.2. Experiment 2

...

# 5. Discussion

This chapter is supposed to discuss your results. Point out what your results mean. What are the limitations of your approach, managerial implications or future impact?
Explain the broader picture but be critical with your methods.

# 6. Conclusion

Repeat the problem and its relevance, as well as the contribution (plus quantitative results).
Look back at what you have written in the introduction.
Provide an outlook for further research steps.

# List of Terms

**ECMWF** European Centre of Medium-Range Weather Forecasts. 7

# Bibliography

Ludwig, N., S. Feuerriegel, and D. Neumann (2015). *Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests*. In: *Journal of Decision Systems*, Vol. 24, No. 1, pp. 19–36.

# A. Appendix

## A.1. First Section

…