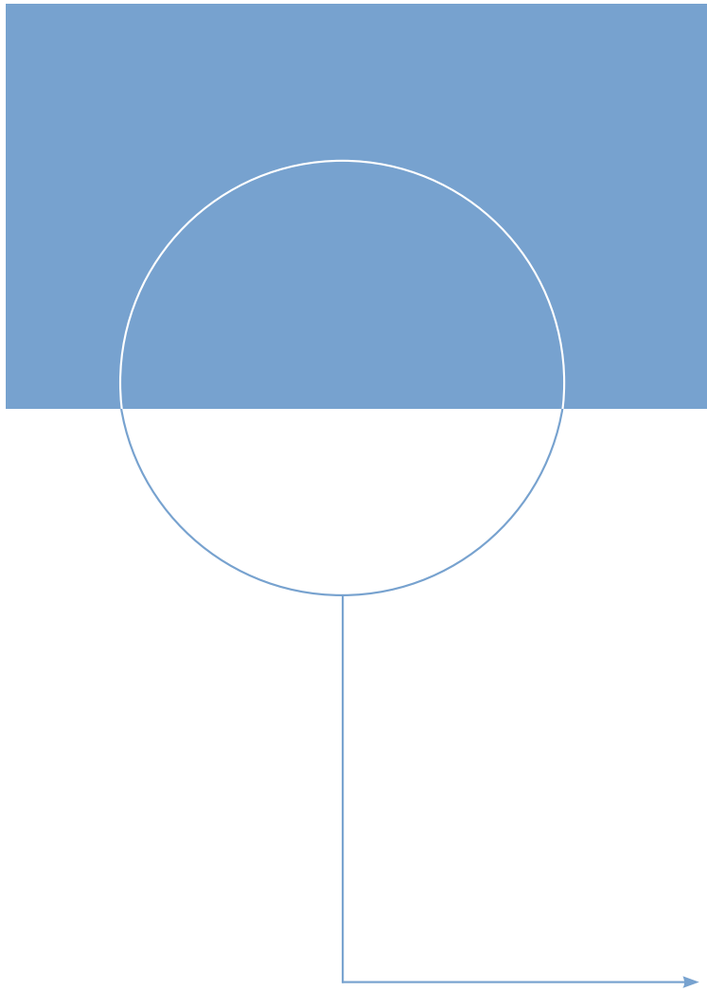


Master's thesis

2019

Anne Siri Fardal

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology and Electrical
Engineering
Department of Mathematical Sciences

Anne Siri Fardal

A Bayesian Model for Prediction of Heat Consumption

June 2019

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

A Bayesian Model for Prediction of Heat Consumption

Anne Siri Fardal

Master of Science in Applied Physics and Mathematics
Submission date: June 2019
Supervisor: Sara Martino

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Accurate prediction of heat consumption is important when developing new neighborhoods to ensure a suitable power grid. In this work, we present latent Gaussian models, estimated in R-INLA, for long-term prediction of hourly heat consumption per square meter. They include, among other things, weather conditions and terms for seasonality, both daily, weekly and annually. An additive regression model is utilized, set up as a latent Gaussian model, to support non-linear terms. Different combinations of covariates and how they are modeled are tested on collections of office buildings in Oslo and Trondheim. It was found that there was little difference in predictive power between the models explored. The two areas require different models, where the model for Oslo that was most successful was the one containing the cycles mentioned above, in addition to the weather conditions temperature and wind speed. For Trondheim, a model similar to the one in Oslo, but where the effect of wind speed is omitted, and a long-term linear trend is included, offered the greatest fit. Both models were able to catch the underlying process well, but correlation in the residuals signifies that more work needs to be done.

Sammendrag

Når man planlegger utbygging av nye områder, er det viktig med korrekte prognoser for varmeforbruk for å kunne sikre et passende kraftnett. I denne oppgaven presenterer vi latente gaussiske modeller, implementert i R-INLA, for langsiktig prediksjon av varmeforbruk per time per kvadratmeter. Modellene inkluderer blant annet ulike værforhold, og effekter for sykluser, både daglige, ukentlige og årlige. Vi anvender en additiv regresjonsmodell, satt opp som en latent gaussisk modell, for å støtte ikke-lineære effekter. Ulike kombinasjoner av kovariater og hvordan de blir modellert er testet på grupper av kontorbygg i Oslo og Trondheim. Resultatene viste at de to områdene krevde ulike modeller. Modellen for Oslo som ga best resultat var den som inkluderte syklusene nevnt ovenfor, i tillegg til værforholdene temperatur og vindhastighet. I Trondheim var det en lignende modell som i Oslo som ga best resultat, men effekten av vind er utelatt, og en langsiktig lineær trend er inkludert. Begge modellene fanget opp de underliggende prosessene som forårsaker observasjonene, men vi hadde korrelasjon i residualene som betyr at videre arbeid med modellene er nødvendig.

Preface

This master's thesis completes my Master of Science in Applied Physics and Mathematics at the Norwegian University of Science and Technology. The work has been carried out during the spring of 2019 at the Department of Mathematical Sciences.

I would like to express my deep gratitude for the guidance and motivation provided by my supervisor Sara Martino. Her tips and comments have been valuable, both in this work and during the specialization project in the fall of 2018. I would also like to thank Igor Sartori at Sintef byggforsk for providing the data on heat consumption and for his interest in the work and helpfully answering all of our questions about the data set.

Finally, I have to thank Bjørnar Keiseraas for his continuous support and encouragement during my work with this thesis, especially during the final weeks.

Contents

1 Introduction **1**

2 Data **5**
2.1 Aggregated Data Set 5
2.2 Weather Data 12

3 Background **17**
3.1 Latent Gaussian Models 17
3.2 Random Effects 18
3.3 Integrated Nested Laplace Approximation (INLA) 20
3.4 Deviance Information Criterion 22
3.5 Evaluation of the Predictive Performance 22

4 Statistical Models for Heat Consumption **25**
4.1 Basic Model 25
4.2 Results from the Basic Model 27
4.3 Additional Models 32
4.4 Separating the Counties 37
4.5 Comparison of Goodness of Fit 41
4.6 Predictive Performance 43

5 Results and Discussion **51**
5.1 Main Conclusions From the Experimental Results 51
5.2 Comparison with the Sintef Model 53
5.3 Further Work 57
5.4 Discussion 57

Appendices **61**

A Additional Results for Trondheim **61**
A.1 Basic Model 61
A.2 Comparison with Sintef 64

B Implementation in R-INLA **67**
B.1 Data Input 67
B.2 Model Specification and Prediction 68

Chapter 1

Introduction

When planning new neighborhoods, knowledge about the estimated future annual heat consumption is of utter importance. In order to create a suitable power grid that can handle the demand, but at the same time is not unnecessary extensive, a thorough understanding of the consumption is crucial in order to avoid financial distress. A prediction that results in underestimation of actual heat consumption will lead to costly supplementary services. In the reverse case, when the demand is overestimated, there will be an excess of supply and complexity of the power system, which is not desirable.

In this study, we will discuss relevant aspects of some statistical models suited for predicting time series of heat consumption, and we will apply the theory to a data set provided by Sintef byggforsk to test the performance of the suggested models. Our goal is to create a model that is able to make good predictions for aggregated heat consumption per square meter far into the future, as far as one year ahead.

In most of the literature available on power load forecasting, short-term load forecasts for electricity have been the area of interest. The interest in short-term is highly due to the fact that electric utilities[1] have to report every day to an energy exchange market and account for the quantities of electricity that they wish to trade the next day. Inaccurate predictions can cause severe financial penalties[2]. For short time-frames, various univariate forecasting models for electricity have successfully been applied, including exponential smoothing methods (Taylor et al., 2006) and (S)ARIMA models (El-Hawary, 2017). When predicting further ahead than one day, it is critical to directly allow for weather-induced variations (Taylor and Buizza, 2003). Methods allowing this include multiple regression (Kaytez et al., 2015) and Kalman filters (Takeda et al., 2016), to mention some. We can also include here some methods that recently have become popular to consider, which are artificial intelligence methods such as recurrent neural networks (Bianchi et al., 2017) and support vector regression (Kaytez et al., 2015), to name a few. In Bianchi et al. (2017) a comparative study on the use of different classes of neural networks in the short-term load forecasting is found. Furthermore, Srivastava et al. (2016) provide an extensive review of even more methods, classical as well as modern approaches.

When it comes to models that have been explored for energy prediction, as the ones mentioned above, forecasting of electricity consumption rather than heat consumption

[1] An example of a large electric utility in Norway is NTE (Nord Trøndelag Elektrisitetsverk).

[2] The prices for trading a surplus or a deficit of electricity the next day, if the utilities experience a significant difference between predicted consumption and the actual need, are typically unfavorable compared to the market where trading for future demand is handled. See more about Nord Pool, one of the largest European energy exchange markets, at `http://www.nordpoolgroup.com`.

has been the main concern. Since both the daily and annual power load, or electricity demand, in the relevant literature have the same form as in the heat consumption data explored in this study, in addition to the dependency on outdoor temperature mentioned in several of the papers, there is reason to believe that the electricity consumption in general includes heat consumption. Thus it is of highly relevance to treat models for prediction of electricity in general as comparable to those for heat prediction. Furthermore, the abundance of work on electricity prediction offers an incentive to investigate heat consumption separately.

As described above, developing forecasting models for heat consumption is of high interest for the building industry. The researchers at Sintef byggforsk therefore have come up with a suggestion for a model (Lindberg et al., 2019) that can be used for prediction of heat consumption. They estimate a linear model for each hour of the day, for each day type. Further, they assume a piecewise linear relationship between temperature and heat consumption, as seen in Figure 1.1. The *change point temperature* (CPT) defines this linear relationship in the sense that when you find the CPT, you say that the consumption at temperatures below the CPT has a negative linear relationship with temperature and is temperature dependent, and for consumption at temperatures above the CPT, the consumption is temperature independent. This is justified by the

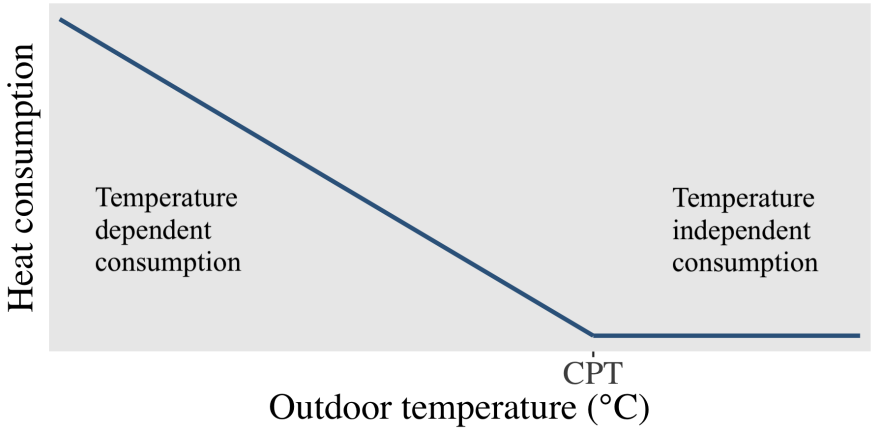


Figure 1.1: Illustration of the idea behind the change point temperature (CPT).

idea that above a certain temperature the heating system is switched off and the consumption consists of just a small basic consumption that cannot be avoided. However, the CPT is unknown, and Lindberg et al. (2019) use an ad hoc iterative procedure to find the CPT.

For the work in this thesis we will employ a Bayesian hierarchical model, specifically a latent Gaussian model (LGM), and R-INLA will be used for model estimation. We propose to model the dependence between temperature and heat consumption with a smooth effect so that it is not necessary to define the CPT. By exploiting the usefulness that LGMs impose by being able to model complicated processes through a hierarchy of simpler models (Hue and Steinsland, 2016), we can hopefully improve the current

model from Sintef. LGMs are particularly useful because of the freedom to include a range of different effects. As a consequence, we are able to include typical effects used to model time series. Evaluation in R-INLA ensures fast and accurate forecasts. No papers have been found by the authors on the subject of models for heat prediction that have been estimated using R-INLA. A related problem, about short-term wind power forecasting, have been considered in Lenzi et al. (2018). Another motivating factor is accordingly to contribute with a paper on this topic.

The rest of the report will continue with an introduction of the data set in Chapter 2 before we go through some important background theory and describe how we evaluate the models in Chapter 3. In Chapter 4 we will present our family of models. Finally, an analysis and discussion of the models and results are given in Chapter 5.

Remark: In this thesis, the terms "load", "demand" and "consumption" will be used interchangeably as there seems to be no distinction between them in the relevant literature. In addition, "Sintef" will for simplicity sometimes be used instead of "Sintef byggforsk".

Chapter 2

Data

The data set with hourly observations of heat consumption was provided by Sintef byggforsk. The time series were recorded over a period of three years, from January 1st 2009 to December 31st 2011, in 27 office buildings in Norway. The buildings are all located in either Oslo or Trondheim. In Table 2.1 mean annual heat consumption per square meter for each building is displayed, along with the size of each building and some quantitative information about each time series. The size of the buildings ranges from 2 570 m^2 to 50 576 m^2. In addition to time series of heat for each building, and its floor area, two time series of outdoor temperature in the same period of time were provided, one for each county. The same goes for wind speed, and for Trondheim an additional time series of solar radiation was given. The stations where the weather measurements are recorded are known: Voll in Trondheim and Blindern in Oslo, controlled by The Norwegian Meteorological Institute. The exact location of each building within a county is unknown.

2.1 Aggregated Data Set

Although we have data for each building, we are actually interested in a neighborhood that is a collection of buildings. In addition, we only have two time series at the most for each weather measurement. We therefore create two time series of aggregated heat consumption, one for each county. Further, the aggregated consumption is divided by the sum of the floor area of the buildings in the corresponding county to get measurements per square meter.

2.1.1 Quality of the Data

Inspection of the data is necessary in order to establish whether there are any outliers or meaningless values present that should be disregarded and not included in the aggregated data set. Table 2.1 displays the percentage of missing values and zeros in the heat consumption data. Looking at the information in this table we first of all notice that three buildings have more than 33% of missing values for heat. The issue with missing values is not actually an issue. When we create an aggregated data set of heat per square meter, we simply ignore the observations that are missing and do not add the area of the associated building to the aggregated area for that observation hour. There would have been problems with the accuracy of a model devoted to make predictions

for one single building when more than 30% of the observations are missing, but since we operate on an aggregated level, this is not a concern.

	Amount of NAs for heat (%)	**Amount of zeros for heat** (%)	**Mean annual heat consumption** (kWh/m^{2}year)	**Size** (m^2)
Trondheim	10.96	0.28	71.90	4 923
	0.00	4.64	108.38	6 334
	2.50	4.20	91.08	6 767
	21.65	0.32	46.13	7 065
	33.33	0.49	39.96	7 360
	0.00	1.63	95.62	9 905
	0.12	0.00	103.98	12 427
	0.00	0.01	113.61	20 311
	3.14	1.91	54.46	20 659
	0.39	0.09	74.85	22 957
Oslo	0.00	12.86	109.32	2 570
	11.97	5.47	106.67	8 129
	0.05	7.41	81.97	9 149
	0.00	11.17	83.63	10 750
	0.03	3.18	95.50	12 920
	3.11	6.03	73.70	13 359
	12.03	1.21	69.21	16 600
	0.37	0.50	90.88	19 320
	44.47	0.00	45.33	21 362
	0.00	0.25	95.72	21 723
	0.02	3.33	145.29	22 000
	4.55	3.96	32.47	50 576
	0.00	31.20	59.67	3 528
	0.00	18.82	88.93	5 153
	0.00	43.72	69.05	5 763
	6.18	33.58	29.66	9 018
	38.18	23.11	29.16	34 500

Table 2.1: Number of missing values and zeros for each building, in addition to mean annual heat consumption per square meter observed from 2009 to 2011 and the size of each building. The buildings marked in red are removed from the aggregated data set.

Next we want to investigate the amount of zeros in each time series to be able to decide whether any of the time series have an unreasonable amount of zeros and will give a wrong attribution on an aggregated level. Approximately 8% of the observed values of heat are zero, and as we can see in Table 2.1 the percentages in 20 of the

buildings are below that. For about one fourth of the buildings the ratio is higher, in fact as high as 44% for one of them. It is unreasonable to assume that all these values are the correct amount of heat consumption considering the low amount in the majority of the buildings.

An explanation for this could be found in the step sizes for recorded heat consumption. Many of the buildings have values of hourly heat consumption in a range where the step sizes are rather large; 15, 20 and even 100 kWh/h, which implies that the recorded values are either rounded to a multiple of the step size, or a value is recorded as zero unless it exceeds a multiple of the step size. It could also be due to a fault in the measuring device. Since our data consist of measurements made in the buildings, in contrast to measurements made at the electric utility providing the power[3], the quality of the data is not expected to be perfect. In any case, the value zero most probably does not reflect the actual heat consumption in some cases. Five buildings, all in Oslo, exhibited such a high amount of zeros that it is hard to tell if they are true or not. Since we want to create a model that is appropriate for all buildings in a county as a whole, we choose to leave these five buildings out. They are marked in red in Table 2.1.

A new data set is created using observations from 22 of the buildings, excluding five buildings in Oslo. We reduce the number of locations from 27 to two, and the number of observations from 709 560 to 52 560, by aggregating the observations in the chosen buildings and placing them in either Oslo or Trondheim. Each county has time series of heat per square meter, temperature and wind, and columns with information about day type (working day, weekend, holiday), day of the year and time of the day. In addition, Trondheim has a time series of solar radiation. Since we will be working on an aggregated level, heat consumption will have the unit of measurement kWh/m^{2}h for the rest of this thesis, and all heat consumption will come from the aggregated data set, with the exception of in Figure 2.2.

2.1.2 Daylight Saving Time

When we have time series of hourly observations covering a whole year, considerations must be taken around the transition points regarding Daylight Saving Time. The transition points appear as time shifts either one hour forward in the spring, or one hour back in the fall. In the data set that we received the time shifts are apparent at slightly different hours than when the actual time shifts appear: the hour 00:00 on the transition day in the spring is missing, and the hour 00:00 the day after the transition day in the fall is counted twice. In Trondheim we notice even more deviating data: the observations of heat in the extra hour we get in the transition in the fall are added to the observations at 04:00 on the transition day. To fix the issue with two observations added together and marked as one, we divide these six measurements (two measurements for each of the three years) by two.

[3]The electric utilities have a stronger need for financial reasons, and therefore better systems, for accurately measuring the power they are providing.

At hour 03:00 on the transition day in the spring all observations of heat in Trondheim are 0. This is due to the missing hour caused by the time shift forward. To avoid dealing with obvious false observations, we copy the observations from hour 04:00 to hour 03:00.

2.1.3 Distribution of the Data

Figure 2.1 displays in the first row the distribution of heat consumption in Oslo and Trondheim. Both the distributions are skewed which suggest that they might belong

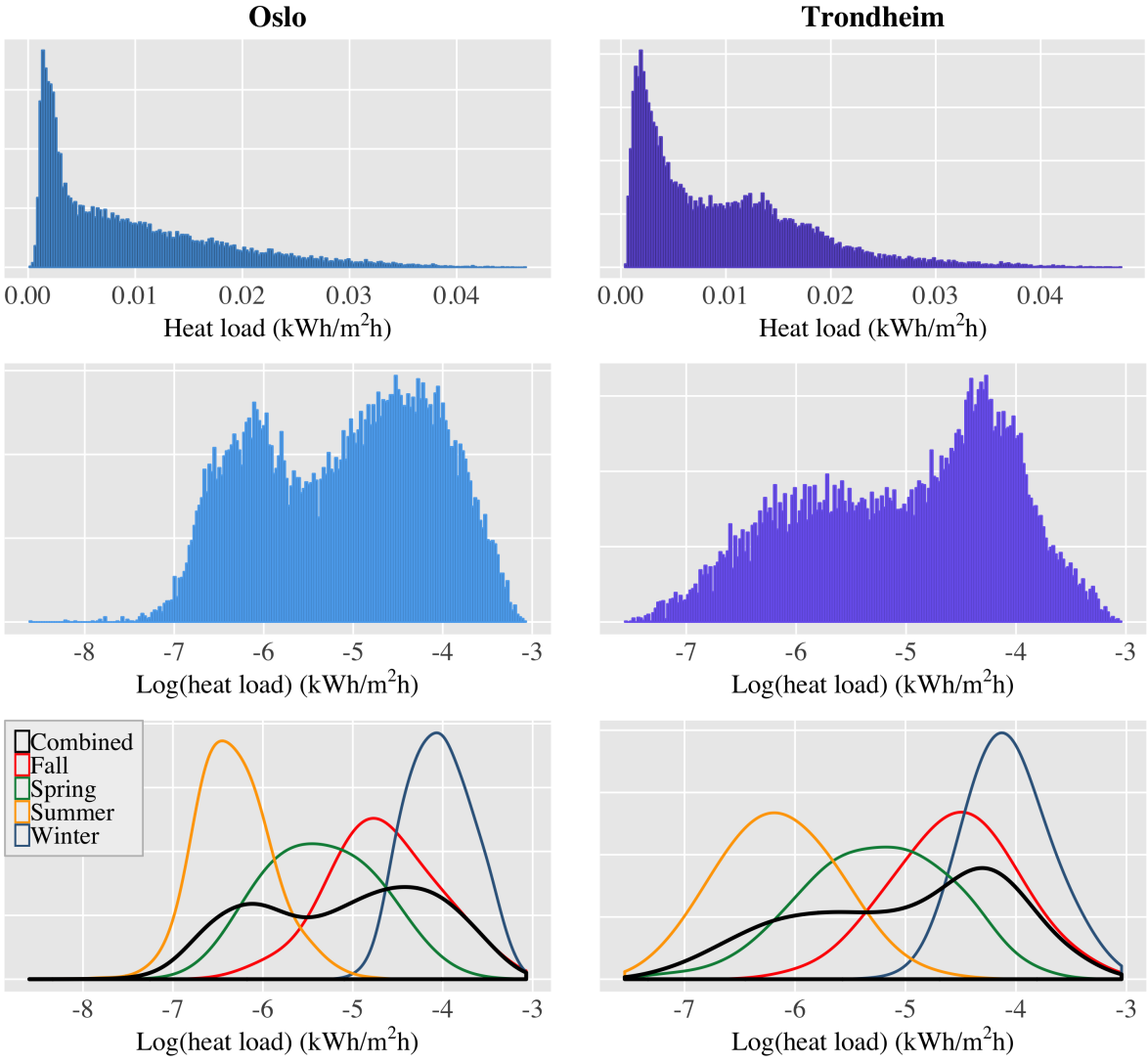


Figure 2.1: Distribution of heat consumption in Oslo (first column) and Trondheim (second column). First row: before log transformation. Second row: after log transformation. Third row: kernel density estimate of the log transformed distributions after separating the observations into seasons.

to a lognormal distribution. In the second row of the same figure the distributions of the log transformed heat consumption are displayed. The resulting shapes are no longer skewed, but they are not exactly as of the familiar shape of the Gaussian, as one would expect when taking the logarithm of a lognormal distribution. It does, however, look like the transformed distributions could be a combination of several distributions. We take a closer look by dividing the transformed observations into seasons. The four new distributions, in addition to the combined distributions from the second row, are shown in the bottom row of Figure 2.1. By looking at each season separately, it becomes clear that the log transformed heat consumption does, in fact, follow the normal distribution. The season separated distributions are not shown here for the untransformed data, but they reveal four skewed distributions.

Figure 2.2 reveals the skewness and variability of the original distribution in seven of the buildings by using box plots and it also, as Figure 2.1, demonstrates how the log transformation transforms the distribution more close to normal. Since this transformation removes the skewness and variability of the data we will, from now on, work with the log transformed time series of heat consumption per square meter.

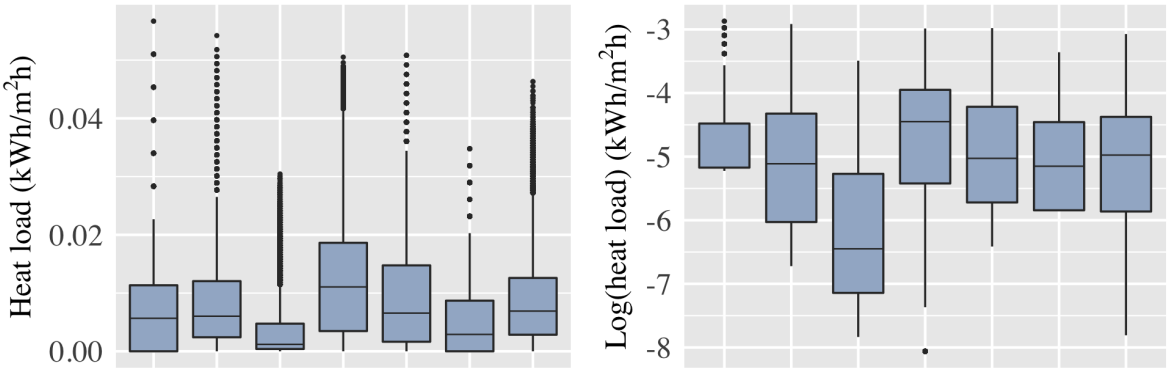


Figure 2.2: Box plot of heat consumption per square meter in seven of the buildings in the original scale (left), and after taking the log transformation (right).

2.1.4 Seasonality

In Figure 2.3 the heat consumption for the log transformed time series from the aggregated data set is displayed. The annual pattern is apparent, and it is similar both across years, but also across the counties. The difference between Oslo and Trondheim is the larger variability that is present in Trondheim. To be able to see the seasonalities within a year, the log transformed heat load in January 2011 in Oslo is displayed in Figure 2.4, and the daily periodic patterns are shown in Figure 2.5. The five days following January 10th in Figure 2.4 are all working days, but they exhibit notably different amounts of heat consumption. The reason is outdoor temperature. In the first three out of these five days the daily mean outdoor temperature was considerably warmer than during the two next days.

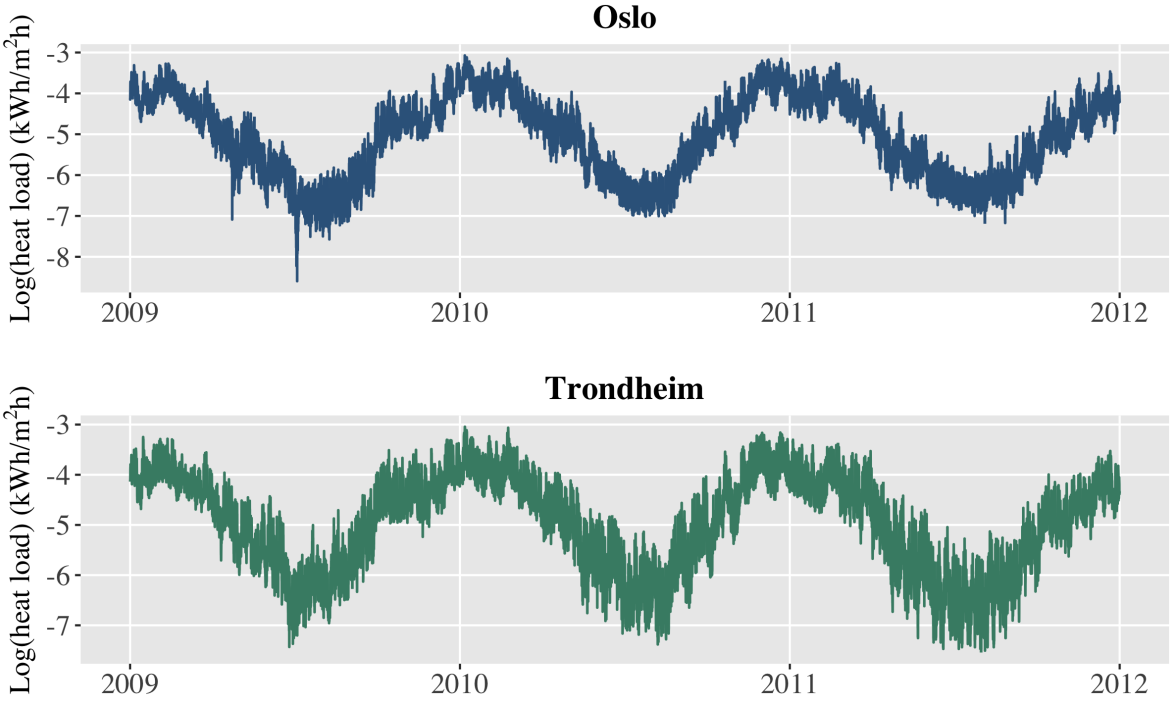


Figure 2.3: Heat consumption for the three years of observations. The labels on the horizontal axes are placed on January 1st in the given years.

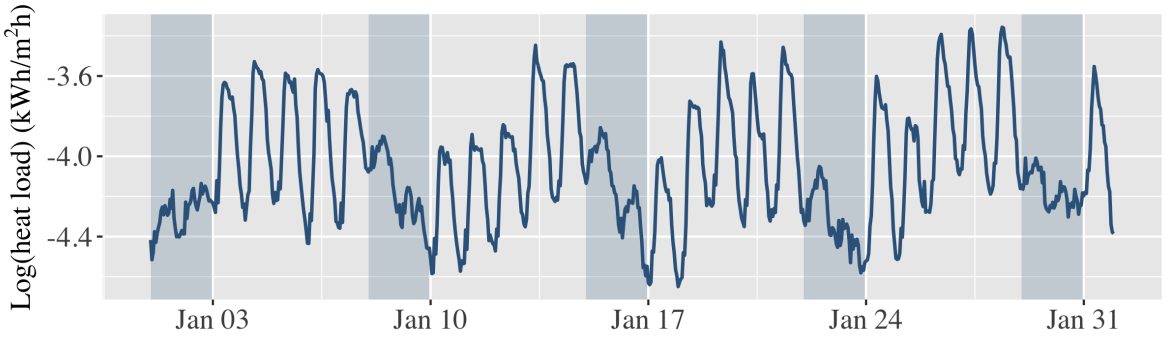


Figure 2.4: Hourly heat load in Oslo in January 2011. Weekends are shaded with a darker background than workdays.

The daily patterns of heat consumption change depending on different factors. In Figure 2.5, the data set is divided into the seven days of the week, in addition to separating holidays, and then averaged over each hour of the day. The main distinction is between working days and non-working days, but there also is some variation between the different non-working days, more so in Oslo. Further there exists some disparity between the two counties. The second half of the day for working days in Trondheim experiences a steeper decrease in consumption. This decrease also happens a bit earlier

here than in Oslo. The differences between the distinct working days are subtle, but in both counties the peak load appears on Wednesdays, while Mondays show a trend of having generally lower consumption throughout the day.

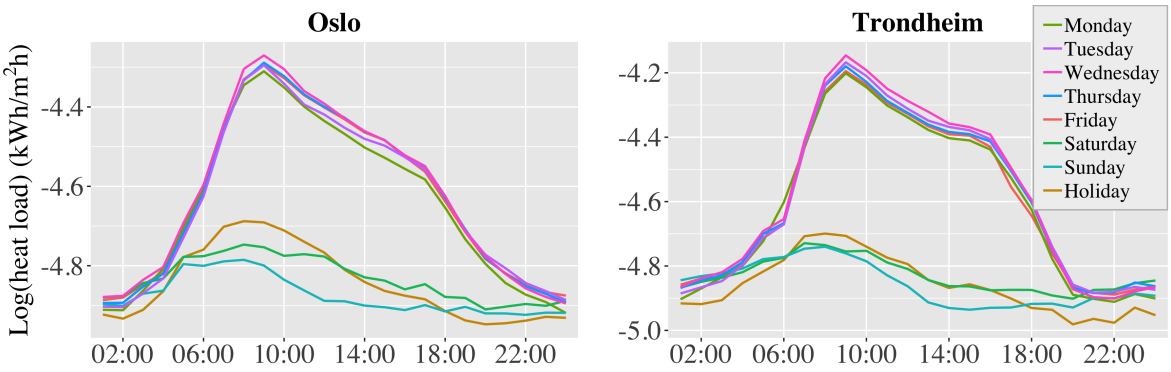


Figure 2.5: Mean daily heat load for all weekdays, included holidays, in Oslo (left) and Trondheim (right).

2.1.5 Time Trends

One thing that is often included when modeling time series is a linear trend. We want to examine if that should be included in our model for heat consumption, and for that reason we have plotted mean monthly values of heat consumption from 2009 to 2011 in Figure 2.6 and 2.7, together with least-squares regression lines.

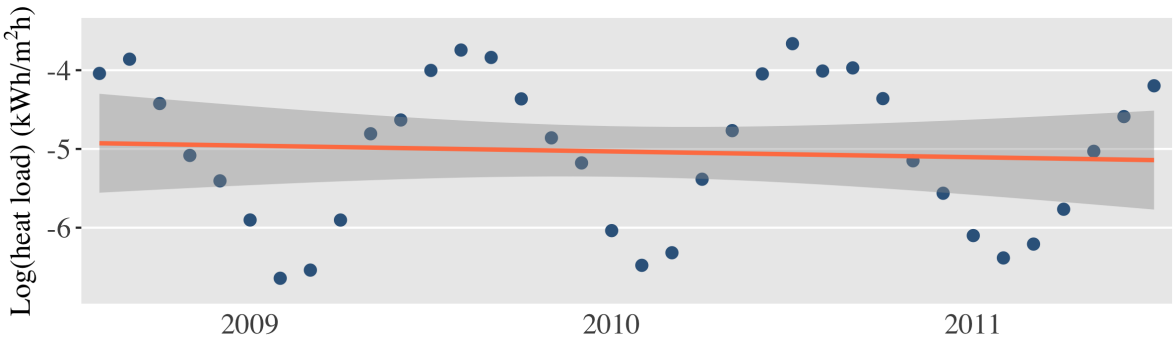


Figure 2.6: Mean monthly values of heat consumption (blue dots) in Oslo for each month. The red line shows the best linear model for the consumption, and shaded area denotes the 95% confidence interval.

The slope of the fitted lines in Figure 2.6 and 2.7 indicates, in Oslo, a monthly average decrease in heat consumption of 0.006 kWh/m^{2}h, and a monthly average decrease of 0.015 kWh/m^{2}h in Trondheim.

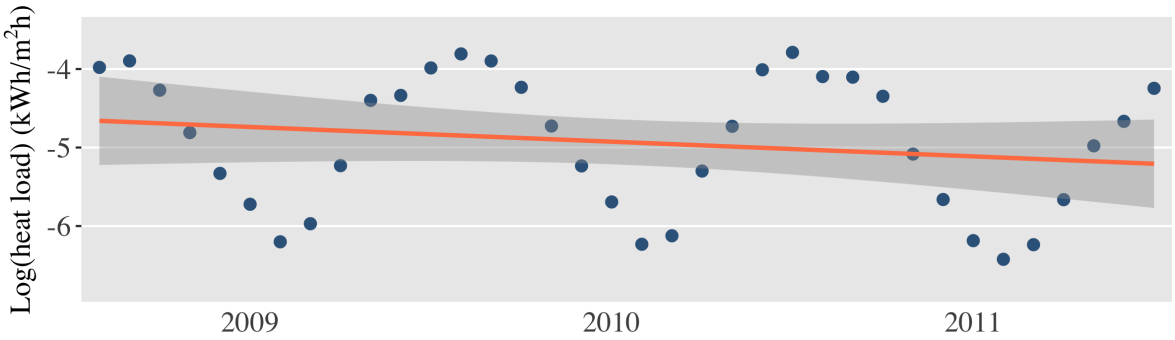


Figure 2.7: Mean monthly values of heat consumption (blue dots) in Trondheim for each month. The red line shows the best linear model for the consumption, and shaded area denotes the 95% confidence interval.

2.2 Weather Data

The observed time series of outdoor temperature are taken at one weather station in each county. Time series of wind speed and solar radiation are also provided. The measurements of solar radiation are only given for Trondheim, and they do not cover all three years of observations of heat consumption. The solar observations stop at May 4th 2011, and there is a large amount of missing values in 2011 until this date. For the wind and temperature time series we have little missing data.

It is known that there is a clear relationship between heat consumption and outdoor temperature. This is because we need to use more heating when it is cold outside to achieve a comfortable indoor temperature. In Figure 2.8 we have scaled the temperature in Oslo to the same range as the log transformed heat consumption and plotted the

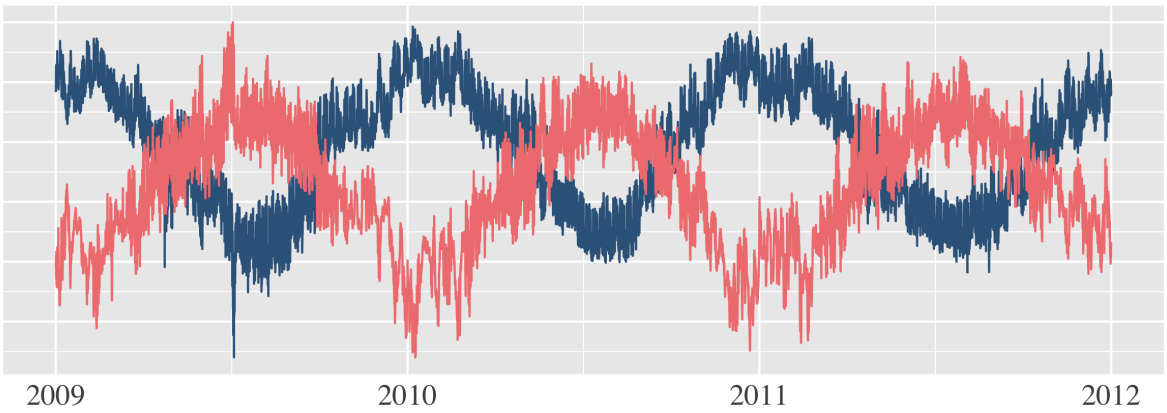


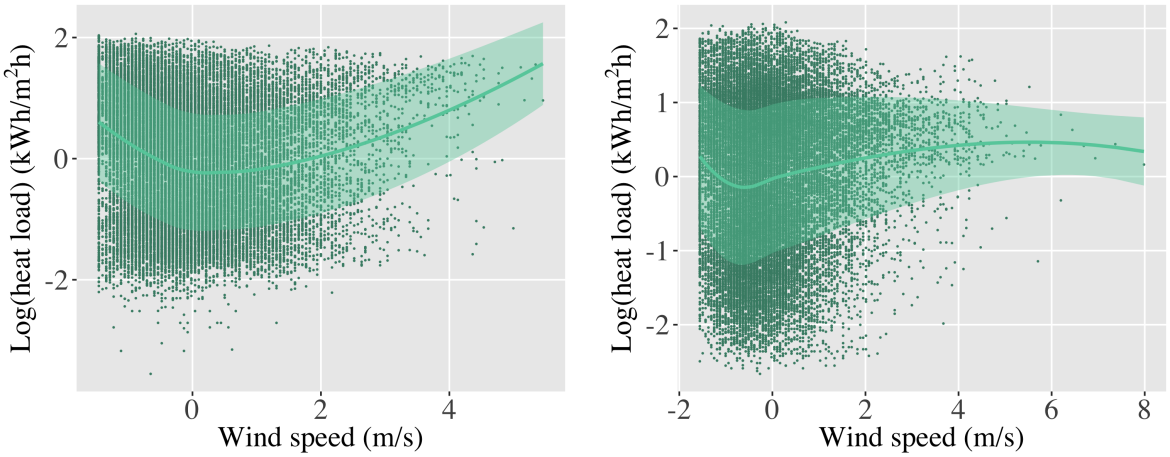
Figure 2.8: Log transformed aggregated heat consumption per square meter (blue) in Oslo from 2009 to 2011 together with outdoor temperature (coral). The temperature is scaled to the same range as the heat consumption.

resulting time series together with observations of heat in Oslo. The correlation between

them is evident.

In order to further detect any relationship between our time series, heat load values from all three years of observations are plotted against the various weather conditions in Figure 2.9, 2.10 and 2.11. Each time series is standardized to zero mean and unit variance so that values can more easily be compared across data types. A local polynomial regression is fitted in each panel in Figure 2.9, 2.10 and 2.11. We keep Oslo and Trondheim separated due to different weather time series, and also to be able to reveal possible differences between the counties.

Starting with wind speed, the relationship is weak in both counties. It switches between negative and positive for increasing wind speed. In Trondheim the regression curve is not that different from a horizontal line, and with the wide confidence interval it is not significant. A reason for the divergence between the counties could be that the buildings in Oslo are more exposed to wind than those in Trondheim. If a building is surrounded by other buildings or is sheltered in some way, it will not experience the



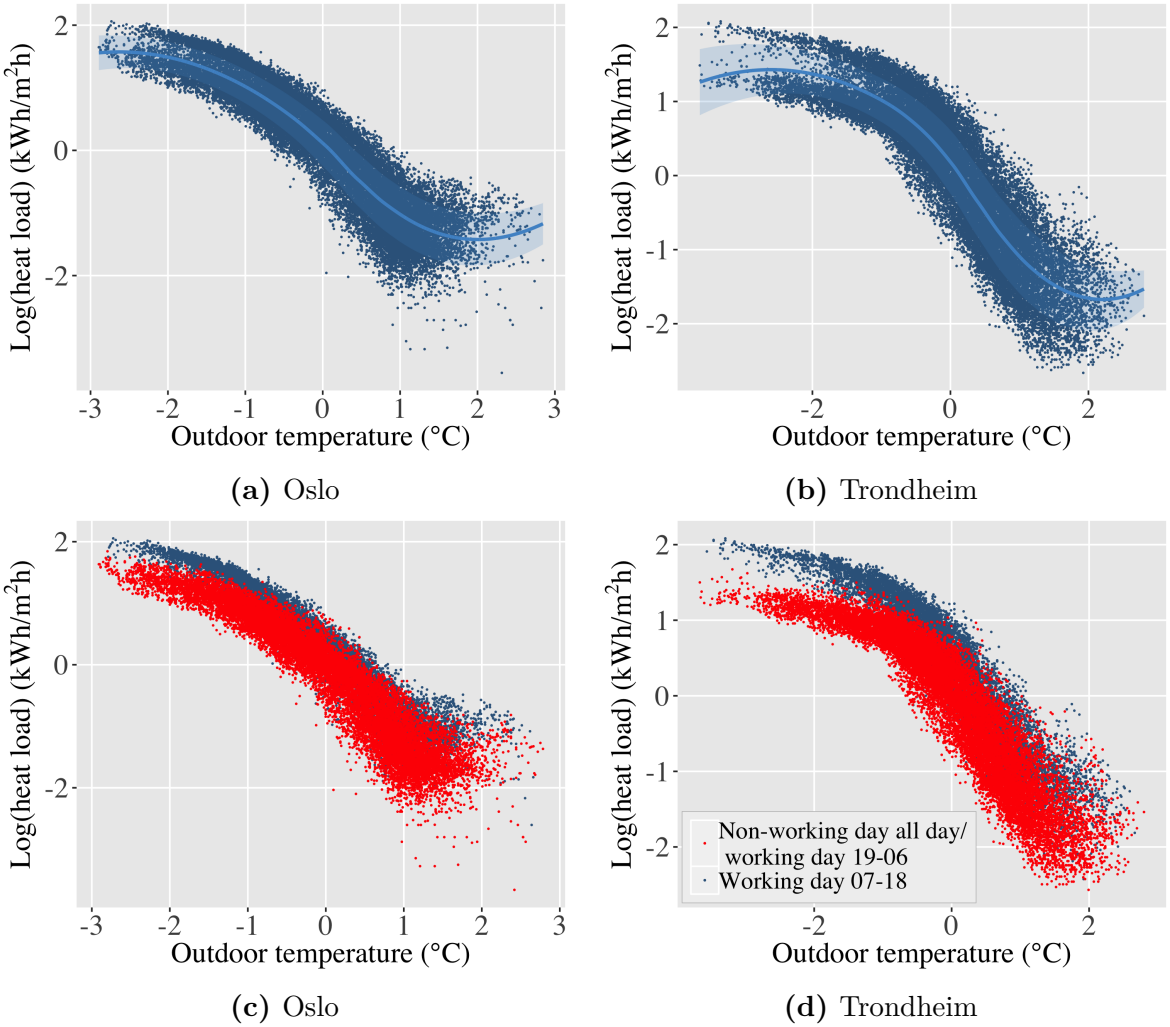
(a) Oslo **(b)** Trondheim

Figure 2.9: Heat load values from 2009 to 2011 are plotted with respect to wind speed. All time series are standardised. A local polynomial regression is fitted to the data in each panel. Translucent bands denote the 95% confidence intervals.

observed wind exposure that is recorded in a location elsewhere in the same county. Since we do not know the location of each building other than the county it is located in, we have no information about the wind exposure of the buildings. Differences in the insulation of the buildings could also be a reason for different impact on heat consumption with increasing wind speed.

The regression curves in Figure 2.10a and 2.10b show that we have a clear, mostly negative, relationship between outdoor temperature and heat consumption that is not linear. The confidence intervals are more narrow than for wind in Figure 2.9 indicating we have a stronger relationship between temperature and heat. To explain the division of the data in the upper left corner in Figure 2.10b, which is also slightly visible in

Figure 2.10a albeit more subtle, we divide the observations of heat and temperature into two parts in Figure 2.10c and 2.10d. The red color denotes all observations that are made either at night time on a working day, defined here as the hours from 19:00 to 06:00



(a) Oslo **(b)** Trondheim

(c) Oslo **(d)** Trondheim

Figure 2.10: Heat load values from 2009 to 2011 are plotted with respect to temperature. All time series are standardised. A local polynomial regression is fitted to the data in the top row. Translucent bands denote the 95% confidence intervals.

the next day, or during all hours on non-working days. For the rest of the observations, which are at day time (07:00 to 18:00) on working days, we have the color blue. The result is that it is this division that causes the data to separate in the upper left corner of the temperature-heat plots. The reasoning behind dividing the day into two parts like this is because of what we saw in Figure 2.5 in Section 2.1.4, namely that the consumption on non-working days is similar to that at night time on working days.

For the solar radiation in Figure 2.11 there is a negative, close to linear relationship

with heat consumption, with a wide 95% confidence interval.

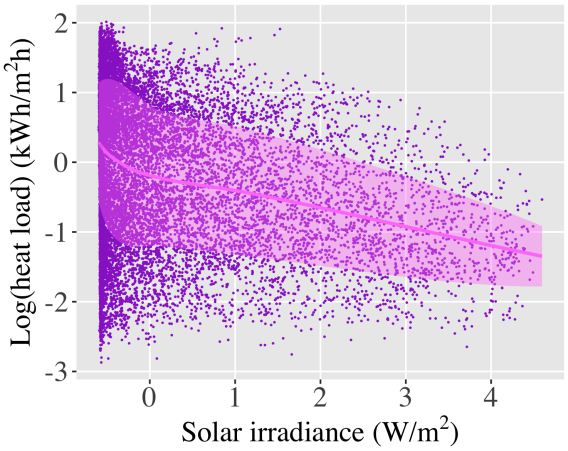


Figure 2.11: Heat load values from 2009, 2010 and parts of 2011 are plotted with respect to solar radiation in Trondheim. The time series are standardised, and a local polynomial regression is fitted to the data. Translucent band denotes the 95% confidence interval.

In conclusion, temperature is by far the most important predictor of the three. The plots of wind speed also implied a possible relationship with heat consumption that is worth considering. Temperature and wind speed will therefore be included in our model. We are missing observations of solar radiation for relatively large parts of our time series of heat consumption, and the relationship with heat consumption in Figure 2.11 is weak. Because of this, solar radiation will not be included in our model.

A note on the solar radiation is that measurements thereof have for a long time been sparse in time, but in recent years more observations are made and are available. For newer data sets it would therefore be of interest to look more into the relationship between solar radiation and heat consumption. This would be of particular interest when predicting for one building rather than an aggregated set as solar exposure varies largely based on the surroundings of each building.

Chapter 3

Background

To be able to reach the goal for this thesis, which is to construct a model for prediction of heat consumption, we need some theory and concepts. In this chapter we will therefore present some important background theory.

3.1 Latent Gaussian Models

A way to model complicated processes is through a hierarchical model consisting of three stages with relatively simple statistical models (Hue and Steinsland, 2016). An often used and quite wide class of models of this structure is that of latent Gaussian models (LGMs) (Rue et al., 2009). They are a subset of structured additive regression models with the property that the latent field is Gaussian.

At the first level of an LGM we have a model for the observed data $\boldsymbol{y}$ called the observation model or observation likelihood. We assume the data to be conditionally independent given some unobserved (latent) stochastic variables. The likelihood, and first stage, is then

$$\pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_1) = \prod_{t=1}^{n} \pi(y_t \mid \eta_t(\boldsymbol{x}), \boldsymbol{\theta}_1),$$

where $\boldsymbol{y} = (y_1, ..., y_n)^\top$ is the response vector of n observations and $\boldsymbol{x} = (x_1, ..., x_n)^\top$ is the vector of latent variables called the latent field. $(\cdot)^\top$ denotes the transpose of a vector. $\boldsymbol{\theta}_1$ contains some parameters that we call hyperparameters and η_t is the t-th linear predictor that connects the data to the latent field. $\boldsymbol{x}$ includes all the parameters in η_t, including η_t. For example, if we choose a Gaussian likelihood $y_t \sim \mathcal{N}(\eta_t, \sigma_\epsilon^2)$ we have that $\theta_1 = \sigma_\epsilon^2$ while η_t would be the mean μ_t. The linear predictor η_t can then be modeled to include e.g. spatial dependence or the effect of covariates as

$$\eta_t = \beta_0 + \boldsymbol{v}^\top \boldsymbol{\beta} + \sum_{j=1}^{p} f_j(z_{jt}), \tag{3.1}$$

where β_0 is a scalar representing the intercept, and the coefficients $\boldsymbol{\beta} = (\beta_1, ..., \beta_m)^\top$ quantify the linear effect of the covariates $\boldsymbol{v} = (v_{1t}, ..., v_{mt})^\top$ on the response. $\boldsymbol{f} = \{f_1(\cdot), ..., f_p(\cdot)\}$ is a collection of functions, defined in terms of the set of non-linear covariates $\boldsymbol{z}_t$, that models the random or nonlinear effects of $\boldsymbol{z}_t$. They can also model interactions. $\boldsymbol{x}$ now includes $(\beta_0, \beta_1, ..., \beta_m, f_1, ..., f_p, \boldsymbol{\eta})$, and Equation (3.1) defines an

additive regression model.

In the second stage we assign a multivariate Gaussian prior on $\boldsymbol{x}$ with mean $\mathbf{0}$ and precision matrix $\boldsymbol{Q}(\boldsymbol{\theta}_2)$ (Blangiardo and Cameletti, 2015) so that

$$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}_2) \sim \mathcal{N}\Big(\mathbf{0}, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2)\Big).$$

θ_2 typically governs the smoothness of the latent field. The third and final stage is formed by the prior distributions assigned to the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$,

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}).$$

LGMs include as a special case many of the models commonly used in statistical science like mixed-effects models and smoothing time series of binomial data. The terms $f_j(\cdot)$ in equation (3.1) can take on many different forms, and for this reason LGMs are very flexible and can accommodate a wide range of models (Blangiardo and Cameletti, 2015).

The matrix $\boldsymbol{Q}$ can be very large, and if it is dense, computations can be demanding. But many of the models that are commonly used as prior for $f_j(\cdot)$ in Equation 3.1 belong to the class of Gaussian Markov random fields (GMRF). Examples are autoregressive models used in time series analysis and random walk models used to model smooth effects of covariates. Many LGMs in the literature satisfy the property that the latent field in an LGM is a GMRF with sparse precision matrix $\boldsymbol{Q}$ (Rue et al., 2009). The Markov property is linked to the sparse structure of $\boldsymbol{Q}$ in the sense that if two elements in $\boldsymbol{x}$ are conditionally independent given the rest, then the corresponding entry in $\boldsymbol{Q}$ is zero.

3.2 Random Effects

Some Markov models that can be used to model the smooth effects $\boldsymbol{f}$ of Equation (3.1) are introduced next. Further we will talk more about the additive regression model we mentioned in Section 3.1 before we will offer a framework for estimating these models and its parameters.

3.2.1 Random Walk Models

The general definition of a random walk (RW) is that it is a random process where we assume that the increments are iid (independent and identically distributed) with a Gaussian distribution. The second-order random walk, RW2, is commonly used for smoothing data and for modeling response functions. They are especially useful when analyzing time series data (Fahrmeir et al., 2013). The RW2 is a Markov model where the following holds:

$$\Delta^2 z_i = z_i - 2z_{i+1} + z_{i+2} \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau_z^{-1}), \qquad i = 1, ..., n-2,$$

where τ_z is the precision, and the density of the RW2 model is

$$\pi(\boldsymbol{z}) \propto \exp\Big(-\frac{\tau_z}{2} \sum_{i=2}^{n-1} (z_{i-1} - 2z_i + z_{i+1})^2 \Big).$$

The Markov property ensures that

$$\pi(z_i \mid \boldsymbol{z}_{-i}) = \pi(z_i \mid z_{i-2}, z_{i-1}, z_{i+1}, z_{i+2}).$$

In turn this means that we have a sparse precision matrix.

Since RW are intrinsic models, identification problems can arise when they are used as prior for smooth effects in LGMs. This is because the RW1 is invariant under the addition of a constant, and the RW2 is invariant under the addition of a constant or a line. Hence, the overall level of the RW model is arbitrary unless we impose further restrictions. The fixing of the level of the model is usually obtained by centering the functions around zero such that you get a sum-to-zero constraint:

$$\sum_{t=1}^{T} f_j(x_{jt}) = 0 \qquad \text{for all } j \in (1, ..., n),$$

where T is the length of the vector $\boldsymbol{y}$.

3.2.2 Autoregressive Models

Another form that $f_j(z_{jt})$ in Equation (3.1) can take on for the element z_{jt} is that of the autoregressive (AR) model. In this case a value from a time series is regressed on p previous values of the same time series, contrary to on some additional covariates. The value p refers to the order of the model, and an AR(p) model is defined as

$$z_{jt} = \sum_{i=1}^{p} \varphi_i z_{j,t-1} + \epsilon_t, \qquad t = p+1, ..., T, \qquad |\varphi_i| < 1,$$

where $\varphi_1, ..., \varphi_p$ are the parameters of the model, and ϵ_t is white noise. When $p = 1$, we get an AR1 model defined by

$$z_{jt} = \varphi z_{j,t-1} + \epsilon_t, \qquad t = 2, ..., T,$$

with

$$\epsilon_t \sim \mathcal{N}(0, \tau_\epsilon^{-1}).$$

In the time series vocabulary, you often work with *lagged* values. A value at lag k is an observation made k time stamps before observation z_{jt}. $z_{j,t-k}$ hence denotes the observation at lag k, compared to time t. A way to measure the linear relationship between an observation z_{jt} and a lagged value is by the *autocorrelation function* (ACF).

The ACF measures the correlation between lagged values of a time series, and for a given k it is given by (Fahrmeir et al., 2013)

$$\text{Corr}(z_{jt}, z_{j,t-k}) = \frac{\text{Cov}(z_{jt}, z_{j,t-k})}{\text{Var}(z_{jk})}.$$

At lag 0 you will have perfect correlation because you are comparing the time series to itself, so you get an autocorrelation equal to 1. For lags larger than 0, you compare the time series to delayed versions of itself.

When you have estimated a model for a time series, looking at the ACF of your residuals is a useful tool to determine whether your model is able to catch the time dependencies in your data. Because the error term e_t in any regression model is assumed to be iid by $e_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau_e^{-1})$ for all t, a good model would reveal no significant autocorrelation after lag 0 in a plot of the ACF of the residuals. If you do, however, have significant spikes in this ACF, your model may benefit from including an AR term.

One important thing to keep in mind when including such a term in your regression model, is the impact it has on flexibility. The variance τ_ϵ^{-1} from the AR term for each z_{jt} can lead to overfitting because the degree of the curve oscillation of the predictions grows large, and the model is able to fit the data perfectly (Zheng and Bakka, 2018). To amend this, we can fix the variance of the response y_t to a small value.

3.3 Integrated Nested Laplace Approximation (INLA)

In Bayesian analysis the goal is often to find the posterior marginal distributions $\pi(x_j|\boldsymbol{y})$ and $\pi(\theta_k|\boldsymbol{y})$ in order to gain knowledge about the unknown distribution of $\boldsymbol{x}$ and the unknown parameters $\boldsymbol{\theta}$. This is where integrated nested Laplace approximation (INLA) offers a fast and computationally cheap method as an alternative to more traditional MCMC methods. It is a method for estimating latent Gaussian models and is therefore suitable to use for estimation of our model.

INLA utilizes numerical integration to approximate

$$\tilde{\pi}(x_j|\boldsymbol{y}) = \int \tilde{\pi}(x_j, \boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} = \int \tilde{\pi}(x_j|\boldsymbol{\theta}, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{3.2}$$

and

$$\tilde{\pi}(\theta_k|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-k} \tag{3.3}$$

where $\boldsymbol{\theta}_{-k}$ denotes the vector $\boldsymbol{\theta}$ without element k. What we need in order to solve the integrals in equation (3.2) and (3.3) is to

1. approximate $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ by $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$

2. approximate $\pi(x_j|\boldsymbol{\theta}, \boldsymbol{y})$ by $\widetilde{\pi}(x_j|\boldsymbol{\theta}, \boldsymbol{y})$.

To approximate (1) INLA utilizes that, in general,

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} = \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{y})} \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}$$

$$\approx \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\widetilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})} = \widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}),$$

where $\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation, given by the Laplace approximation of $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ at its mode $\boldsymbol{x}^*(\boldsymbol{\theta})$ for a given $\boldsymbol{\theta}$ (see Blangiardo and Cameletti (2015) for details). This approximation is usually very precise as the prior of $\boldsymbol{x}$ is Gaussian and the full posterior $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ is close to Gaussian.

Approximating (2) is more complex with more expensive computation as there tends to be a larger number of components in $\boldsymbol{x}$ than in $\boldsymbol{\theta}$. One of the options proposed by Rue et al. (2009) is to use Laplace approximation again:

$$\pi(x_j|\boldsymbol{\theta}, \boldsymbol{y}) = \frac{\pi((x_j, \boldsymbol{x}_{-j})|\boldsymbol{\theta}, \boldsymbol{y})}{\pi(\boldsymbol{x}_{-j}|x_j, \boldsymbol{\theta}, \boldsymbol{y})} = \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{\theta}|\boldsymbol{y})\pi(\boldsymbol{x}_{-j}|x_j, \boldsymbol{\theta}\boldsymbol{y})} \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{x}_{-j}|x_j, \boldsymbol{\theta}, \boldsymbol{y})}$$

$$\approx \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})}{\widetilde{\pi}(\boldsymbol{x}_{-j}|x_j, \boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-j}(\boldsymbol{x}_j, \boldsymbol{\theta})} \coloneqq \widetilde{\pi}(x_j|\boldsymbol{\theta}, \boldsymbol{y}),$$

where $\tilde{\pi}(\boldsymbol{x}_{-j}|x_j, \boldsymbol{\theta}, \boldsymbol{y})$ is the Laplace approximation of $\pi(\boldsymbol{x}_{-j}|x_j, \boldsymbol{\theta}, \boldsymbol{y})$ at its mode $\boldsymbol{x}^*_{-j}(\boldsymbol{x}_j, \boldsymbol{\theta})$ for a given $\boldsymbol{\theta}$ and x_j.

INLA exploits the fact that $\boldsymbol{Q}$ in the second stage of the LGM is Gauss-Markov, and thus sparse, to attain computational efficiency. Approximate results are obtained very quickly and are usually also very precise. INLA has been used on a large number of applied projects: disease mapping, evolution of the Ebola virus, search for evidence of gene expression heterosis, effects of measurement errors and so on. More examples are found in Rue et al. (2016).

3.3.1 Implementation in R-INLA

The R-INLA package is used to implement the models in this thesis. There are two steps:

1) Define the linear predictor through a `formula` object.

2) Fit the model using the function `inla()`. The fitted model is returned as an `inla` object.

The `formula` can include fixed effects and random effects defined through the `f()`

function. The `inla` object includes the posterior marginal distributions of the latent effects and hyperparameters, as well as summary statistics. Model choice criteria such as DIC are also available. In R-INLA predictions must be done as a part of the model fitting itself. As prediction can be regarded as fitting a model with missing data, we can simply set `y[i]=NA` for those i's we want to predict. `inla()` does not return a predictive distribution, it only returns the posterior marginals of the linear predictor at the missing locations. To obtain a predictive distribution one has to add the observational noise to the fitted values.

An example on making predictions in R-INLA is provided in Appendix B.

3.4 Deviance Information Criterion

The deviance information criterion (DIC) is a tool for Bayesian model selection. Spiegelhalter et al. (2002) proposed this criterion based on a trade-off between goodness of fit of the model, and the corresponding complexity of the model. DIC is based on the fit to observed data given the posterior mean (Tsai, 2015) and is defined by

$$\text{DIC} = -2\log\Big(p(\boldsymbol{y}|\hat{\boldsymbol{\theta}})\Big) + 2p_{\text{DIC}},$$

where $\boldsymbol{y}$ is the data, $p(\boldsymbol{y}|\hat{\boldsymbol{\theta}})$ is the likelihood and $\widehat{\boldsymbol{\theta}}$ is the posterior mean $\mathbb{E}(\boldsymbol{\theta}|\boldsymbol{y})$ of the parameters $\boldsymbol{\theta}$. p_{DIC} is an estimate of the so-called effective number of parameters in the model, and it penalizes the complexity of a model by

$$p_{\text{DIC}} = 2\Big(\log(p(\boldsymbol{y}|\hat{\boldsymbol{\theta}})) - \mathbb{E}[\log(p(\boldsymbol{y}|\boldsymbol{\theta}))]\Big). \tag{3.4}$$

The expectation in the second term is an average of $\boldsymbol{\theta}$ over its posterior distribution (Gelman et al., 2013).

The first term in Equation (3.4) favors a good fit, and the second term penalizes model complexity. A lower value of DIC of one model compared to another indicates a better fit.

3.5 Evaluation of the Predictive Performance

In order to evaluate and compare the predictive performance of the different models to be presented, we will employ two criterions: the root-mean-square error (RMSE) and the continuous ranked probability score (CRPS). They both measure how close the predicted values are to the observations, but with different approaches. We will in this paper work with averaged annual RMSE and CRPS.

3.5.1 RMSE

The definition of the RMSE in year j is

$$\text{RMSE}_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\Big(\hat{y}_t - y_t\Big)^2},$$

where T is the number of predictions made in a year (one for each hour), and $\hat{y}_t$ and y_t are the predicted value and the observation, respectively, at time t in year j. As mentioned, the mean annual value of RMSE, averaged over all years r, will be used as measurement of the predictive performance, and it is found by

$$\overline{\text{RMSE}} = \frac{1}{r}\sum_{j=1}^{r}\text{RMSE}_j.$$

3.5.2 CRPS

The second criterion that will be used for assessing the predictive power of our models is the CRPS. This is a probabilistic forecast score that evaluates the performance of forecast densitites, and is defined, for one predicted value, as

$$\text{CRPS}(F, y_t) = \int_{-\infty}^{\infty}\Big(F(u) - 1\{y_t \leq u\}\Big)^2 du.$$

Here F is the predictive cumulative distribution function and y_t is the observed value of heat consumption. Computation of the CRPS is carried out in R using the function `crps()` found in the `verification` package. Here it is assumed that the posterior distribution F is Gaussian with mean and standard deviation equal to the posterior mean and standard deviation of heat consumption. The annual mean is then found by

$$\overline{\text{CRPS}} = \frac{1}{r}\sum_{j=1}^{r}\frac{1}{T}\sum_{t=1}^{T}\text{CRPS}\Big(F_{tj}, y_j(t)\Big).$$

When comparing two values of $\overline{\text{RMSE}}$ or $\overline{\text{CRPS}}$ for two different models, a lower value indicates more accurate predictions.

Chapter 4

Statistical Models for Heat Consumption

Theory and knowledge from the preceding chapters will now be used to construct a number of latent Gaussian models for aggregated heat consumption per hour. A basic model common for both Oslo and Trondheim will be introduced first, before we test new covariates and effects in addition to estimate separate models for the two counties. The computations will be carried out using the R-INLA methodology to ensure fast calculations.

4.1 Basic Model

We will set up the model for the log transformed time series of heat consumption as an additive regression model with Gaussian likelihood. We choose GMRF priors for the effects of covariates and we are therefore in the family of LGMs. Furthermore, with priors on the hyperparameters, this is a Bayesian LGM which allows us to use R-INLA for the model estimation.

With an essentially unlimited number of models, made up by various different components, we restrict ourselves to first consider a basic model we believe will capture most of the patterns that are present in the underlying process. In our basic model for the linear predictor η_t we incorporate the dependencies and structures we discovered in Chapter 2. As mentioned above, the basic model is common for both counties, so we will train the model on data from both Oslo and Trondheim. For two effects we will make a division between the counties by including an interaction with county in the smooth effect, and by adding a fixed effect of county. This way we allow for differences between the areas that after all are located at two rather different locations. The observations of heat consumption is assumed to follow $y_t \sim \mathcal{N}(\eta_t, \sigma_y^2)$, and the basic model looks like this:

$$\begin{aligned} \eta_t &= \boldsymbol{x}^\top \boldsymbol{\beta} + f(\text{temp}_t) + f(\text{day}_t) + f(\text{hour}_t, \text{daytype}_t) + f(t, \text{county}) \\ y_t &= \eta_t + e_t. \end{aligned} \tag{4.1}$$

Here $f(\text{temp}_t)$ represents a smooth effect of the observed temperature while $f(\text{day}_t)$ is a smooth effect of the day of the year: $\text{day}_t \in (1, ..., 365)$. For these two effects we use an RW2 model with unknown precision. For the annual cycle we assume the the RW2 model is circular, meaning that the last day of the year is assumed to be neighbor with

the first one. This is because it makes sense to assume that these two time stamps have a similar effect.

The $f(\text{hour}_t, \text{daytype}_t)$ term indicates an interaction between the hour of the day, $\text{hour}_t \in (0, ..., 23)$, and the day type (working day and non-working day). We assume that each day type has its own daily profile but that all daily profiles have the same smoothness. We achieve this by using the `replicate` option in R-INLA when specifying the corresponding `f()` function. We assume a circular RW2 model for the daily cycles. We include this interaction because of what we saw in Figure 2.10c and 2.10d, namely that the level of heat consumption shift depending on hour_t and daytype_t.

$f(t, \text{county})$ is an AR1 model that is used in order to capture some of the residual autocorrelation in the time series. We assume that each of the two counties has its own autoregressive effect. Again we model this using the `replicate` option.

Finally, the $\boldsymbol{x}^\top \boldsymbol{\beta}$ part of the model includes linear effects of the covariates county and daytype_t. We include these linear effects due to the fact that replicates of the daily cycle in the implementation are only allowed different shapes, but they all sum to zero. Because we know that the heat consumption is larger during working days than non-working days, we need to allow for different amounts as well as different shapes. In addition, by including the linear effect of county we allow for differences between Oslo and Trondheim. The reference levels, which are included in the intercept, are county = Oslo and daytype_t = working day.

The basic model has five hyperparameters for which we employ uninformative priors. The hyperparameters consist of one precision parameter for each of the random effects and one autocorrelation parameter for the AR1 model. We have chosen to fix the precision parameter of the likelihood to a high value in order to avoid overfitting. If we were to not fix it we would have two unstructured random noises at each step of the time series: one coming from the innovation of the AR1 model and one from the likelihood error (Zheng and Bakka, 2018). Fixing the likelihood also has the advantage that the posterior distribution for the linear predictor returned by `inla()` coincides with the predictive distribution for the unobserved data.

e_t is the measurement error that is assumed to be iid by $e_t \sim \mathcal{N}(0, \tau_e^{-1})$, where τ_e is fixed.

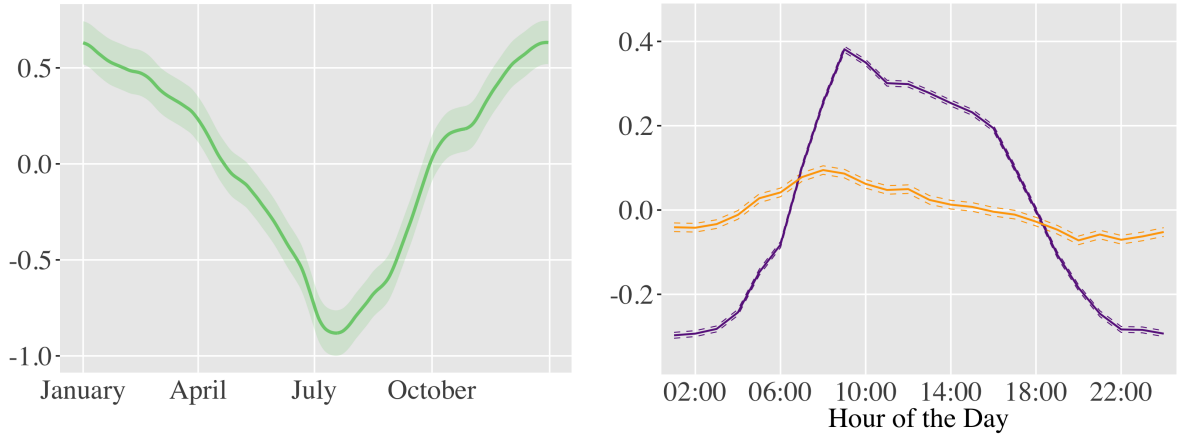
A problem that can arise with the model is stability issues. In Section 2.2 we talked about the correlation between heat consumption and outdoor temperature. So when including the annual cycle in the basic model, which is highly correlated to temperature, stability issues arise (Fahrmeir et al., 2013). Since it is the temperature that causes the annual cycle, it would seem enough to just include temp_t (temperature) in the model, and not day_t (annual cycle). However, when excluding day_t, the annual cycle appears in the residuals. Apparently temp_t is not able to capture the entire annual cycle, and day_t is required in the model. Like we said, because of multicollinearity, this could cause stability issues when estimating the model.

4.2 Results from the Basic Model

In this section, and the following one, we present the results of training the basic model, and more models to be introduced, on *all* observations from both Oslo and Trondheim. We therefore talk about fitted values and not predictions.

The annual effect is shown in Figure 4.1a. Remembering how one year of consumption looked like in Figure 2.3, we see that the annual effect and the observed annual cycle have, roughly, the same shape. In Figure 4.1b the daily profiles are shown, and there is a clear distinction between the day types. The shapes are similar to those in Figure 2.5. For working day there is a sharp increase in the morning with a peak around 09:00. The consumption then decreases when the working day is over. In the non-working days the consumption is more or less flat.

It may seem more natural to separate between, perhaps, eight day types, which represent each of the seven days in the week and holidays. Figure 2.5 suggested that there was little difference between the working days (Monday-Friday), and that the weekend and holidays share the same characteristics. A basic model with these eight day types was also tested on the data, and it actually produced a slightly higher value of DIC and certainly did not improve the predictions. This is why we only work with two day types in our models.



(a) Annual effect.

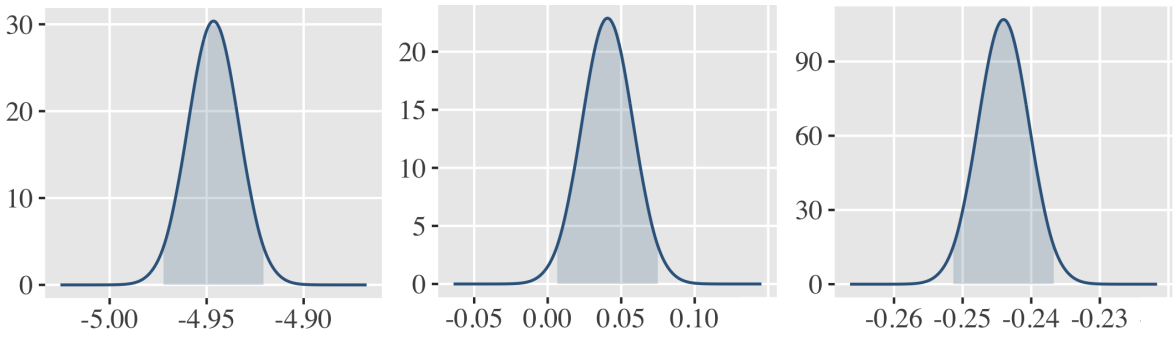
(b) Daily profiles for working days (purple) and non-working days (orange).

Figure 4.1: Annual effect and daily profiles. Translucent band and dashed lines indicate the 95% credible intervals.

The linear effects in the basic model are shown in terms of their posterior marginal distributions in Figure 4.2. The intercept, which includes the reference levels county = Oslo and daytype_t = working day, is shown in Figure 4.2a. In Figure 4.2b we have the linear effect that is added to the linear predictor if county = Trondheim is true. The mean is small, but the 95% credible interval does not cover zero, so it is significantly different from zero, and we get a small increase in heat consumption in Trondheim.

From what we have seen about the consumption in the two counties, the amounts are similar (see for example Figure 2.3) and thus a small value for the additional effect of Trondheim is to be expected.

When we have a non-working day, the consumption is reduced by about 0.245, as seen in Figure 4.2c. This is in line with what we saw in Figure 2.5, namely that the level of consumption on non-working days is lower than on working days.



(a) Oslo and working day **(b)** Trondheim **(c)** Non-working day

Figure 4.2: The posterior marginal distributions of the linear effects. Shaded areas denote the 95% credible intervals.

Finally, the effect of temperature is shown in Figure 4.3. It is similar to the shape of the distributions of heat load with respect to temperature in Figure 2.10a and 2.10b in Section 2.2. As expected, the consumption decreases with higher temperatures. The curve is relatively flat for very low temperatures and decreases steeply between -10°C and 15°C and shows a tendency to stabilize, with some increase, for temperatures higher than 15-20°C. This is expected as for high temperatures the heating system is shut down. The slight increase might be due to the fact that, for some buildings, energy

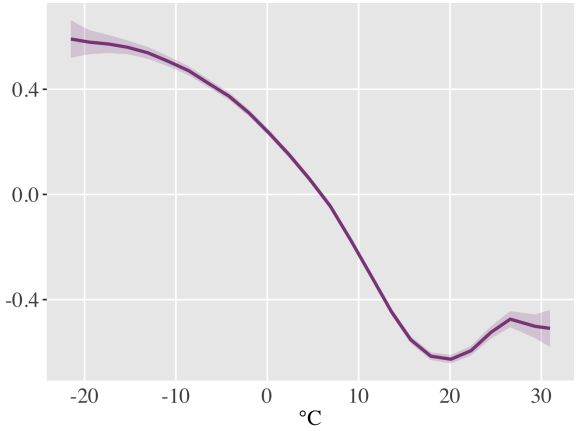


Figure 4.3: Effect of temperature. Translucent band indicate the 95% credible interval.

is used for cooling[4]. The intervals are wider at the ends of the temperature scale due to few observations of heat consumption at these temperatures.

4.2.1 Fitted Values Compared to Observations

The shapes and values of the effects above tell us little about how well the observations are estimated. We will in this section compare the fitted values in Oslo from the basic model with the observed heat consumption. Corresponding plots and results are found in Appendix A for Trondheim. The estimated value of the linear predictor that is used for comparison with observations throughout this thesis is the median, or 0.5 quantile. We use the median instead of the mean because if we want to look at the predictions in the original scale, i.e. not on the log scale, we can transform the median, but we cannot do the same for the mean.

In Figure 4.4 we have the fitted values (orange) from the basic model, together with the observations (blue) for 2011 in Oslo. The observations are hard to see here since the fitted values overlap with them for large parts of the year. In Figure 4.5 two single weeks are shown, making it possible to also see the 95% prediction interval.

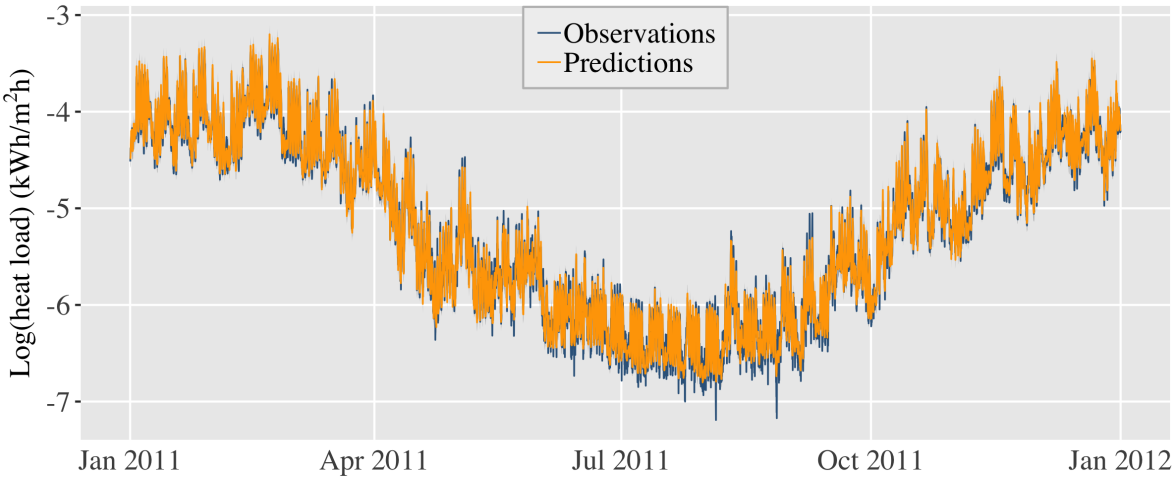


Figure 4.4: The median of the predictive distribution and observations for a year in Oslo.

The curves in both Figure 4.4 and 4.5, where orange is the curve for the fitted values, tell us that the basic model captures the annual cycle well, in addition to the weekly cycle and the daily cycles. Moreover, it seems as though the effect of temperature works well in the sense that the last five days in the top plot in Figure 4.5, which are working days, are not all at the same level.

The fitted values follow a smoother curve than the curve for the observations which is preferable since we do not want to overfit. We notice there seems to be a shift in the daily peak load in the top plot in Figure 4.5. This could indicate that lagged

[4]This information was obtained through communication with researchers at Sintef.

temperatures would be more appropriate to include as input to the model instead of the actual temperatures. Figure 2.10a and 2.10b suggested that there is a strong correlation between outdoor temperature and heat consumption, but it is not obvious how long it takes for changes in the temperature to appear in the amount of heat that is consumed. Because buildings in Norway are generally well insulated, it is to be anticipated that fluctuations in temperature do not appear instantaneously in the demand. We will test this idea later in Section 4.3.4.

Note that it does not seem like the whole day is shifted, only the peak. Furthermore, in the bottom plot in Figure 4.5 this shift does not occur as clearly. If the shift is caused by the effect of temperature, it is expected that we do not see this characteristic in this plot as it shows a week in the summer. Remembering the effect of temperature in Figure 4.3, we know that temperature has less effect on the consumption for high temperatures.

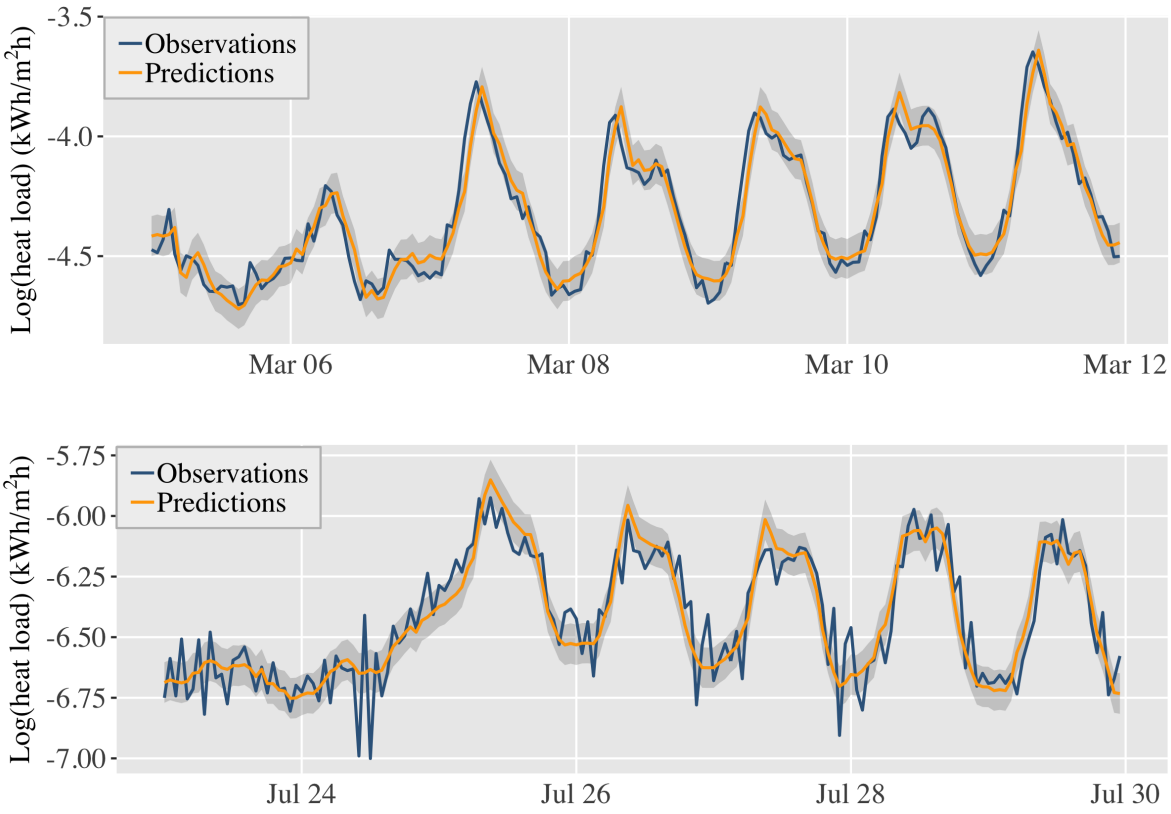


Figure 4.5: The median of the predictive distribution and observations for two weeks in 2011 in Oslo along with the 95% prediction intervals.

In Figure 4.4 it is apparent that it is in the summer that our model performs least well. We see a tendency of what appears to be overestimation of the consumption. In the bottom plot in Figure 4.5 a week from this period is shown. The small scale variation of the observations that occur here is not captured by the model. However,

the larger scale pattern seems to be well represented by the fitted values. Furthermore, based on the same plot it seems like the assumption we made of overestimation in the summer is not the case for all consumption here, but rather a lack of hitting the extreme values, meaning the lowest daily demands.

The increased difference between observations and fitted values at summer time is reflected in the residuals for Oslo in Figure 4.6, where there is a higher variance in the residuals in the summer. We note that the summer time is the period where heat consumption is at its lowest. Moderately under- or overestimation of the demand here will therefore not have a big impact on the calculation of the annual demand.

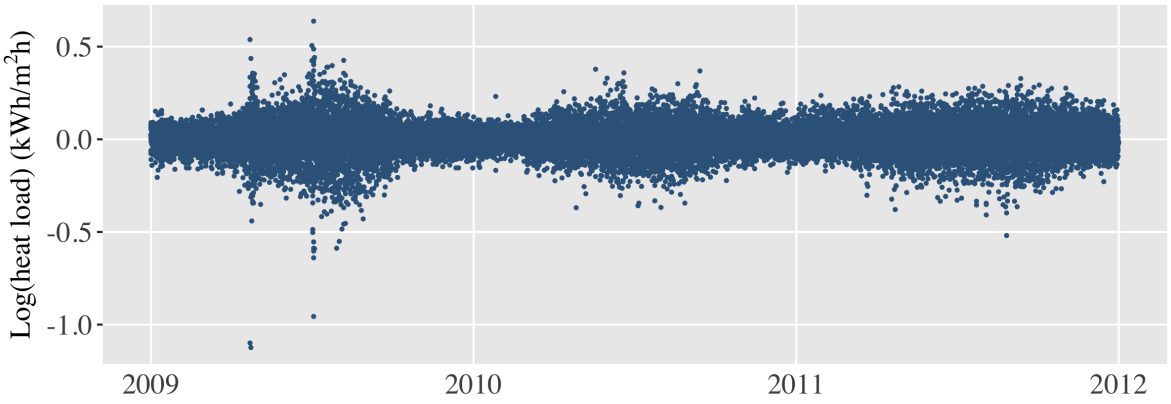


Figure 4.6: Residuals in Oslo.

The higher variance in the residuals in the summer is the most notable feature of the residuals. In order to detect whether there is more correlation between them, we have plotted the ACF of the residuals in Figure 4.7. We see a 24 hour cycle in the ACF

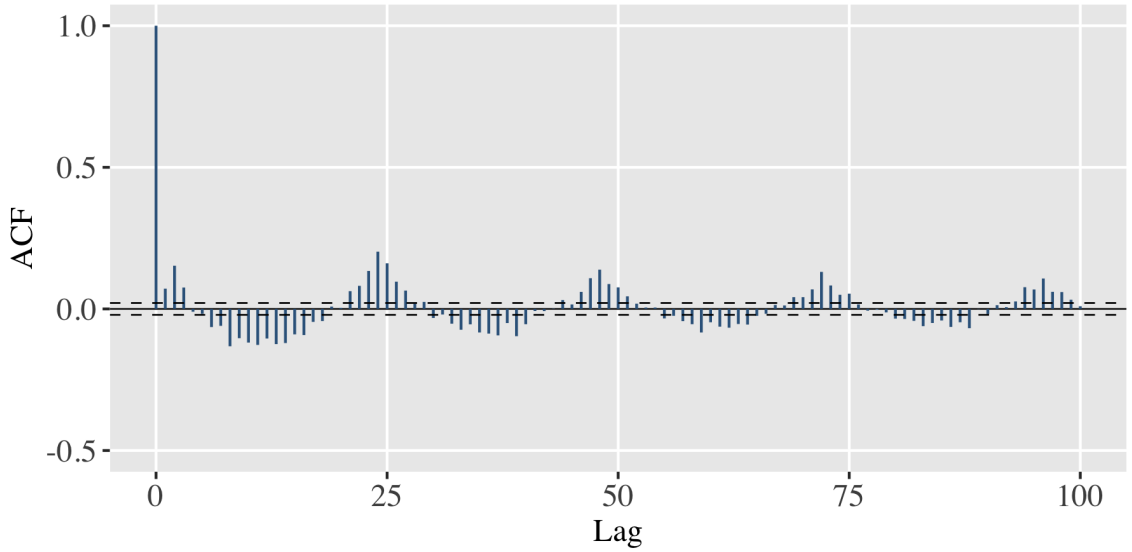


Figure 4.7: ACF of residuals in 2009 in Oslo.

that the model has not been able to catch. In Figure A.4 in Appendix A the ACF of the same year is found for Trondheim. The result is similar as in Oslo.

The estimated value of φ of the AR1 model $f(t, \text{county})$ has a value very close to 1, making the inputs to $f(t)$ highly correlated to the outputs. This, along with the correlation in Figure 4.7, seems to suggest that the deterministic part of the model does not capture all the variability in the data.

4.3 Additional Models

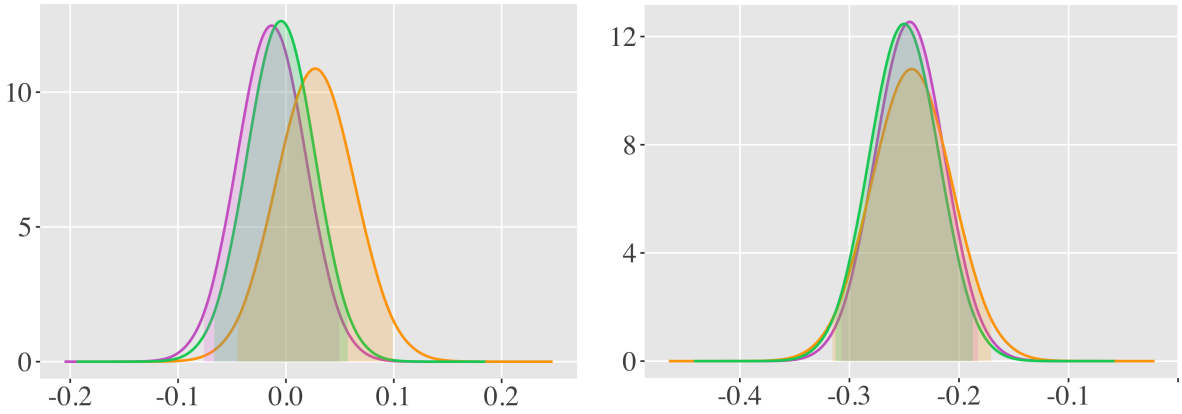
We now expand our set of proposal models by altering the basic model. We start by testing whether adding interactions between covariates improves the model. Next we introduce new covariates and effects. The models are, as in the previous section, trained on all observations from Oslo and Trondheim. We keep the new effects common for both counties, and in Section 4.4 we will include unique effects for each county, allowing us to discover potential differences between them.

4.3.1 Interaction Between Day Type and Season

An interesting question could be whether the daily cycle changes not only depending on day type, but also from season to season. This could be due to the fact that working habits are different in winter and summer. To investigate this we add to our model an interaction effect between daily cycles and season. We add the interaction both in the fixed effect and in the random effect. This will be called Model S4. We have defined the four seasons here as winter from December 21st to March 20th, spring is the next three months, then three months of summer and finally three months of fall, ending December 20th. This means that now instead of two daily profiles (working day/non-working day) we have eight ($\text{daytype}_t \times \text{season}_t$), with two and four levels of daytype_t and season_t, respectively.

The fixed effect season_t = winter is included in the intercept together with the reference levels Oslo and working day. The posterior marginal distributions of the fixed effects of spring, summer and fall are shown in Figure 4.8. The interaction between these seasons and working day is found in Figure 4.8a, and the interaction with non-working day is found in Figure 4.8b.

In Figure 4.8a we see the outcome of giving the linear effect of working day the opportunity to change through the seasons, and the result is that it does not contribute much to the response. The 95% credible intervals for all interactions cover zero, thus the effects are not significantly different from zero. The mean of the effect of summer is slightly positive, but since zero is included in the credible interval, the effect is not significant.



(a) Working day **(b)** Non-working day

Figure 4.8: Posterior marginal distributions of the linear effects for interaction between day type and the seasons spring (purple), summer (orange) and fall (green). Shaded areas denote the 95% credible intervals.

Figure 4.8b shows the effects of season interacting with non-working day. They are all significantly different from zero, and the similarity between them indicates that it is non-working day that is the reason for this, and not the interaction with season.

The different daily profiles for $f(\text{hour}_t, \text{daytype}_t, \text{season}_t)$ are shown in Figure 4.9. We have separated the seasons by different colors, and the day types by solid and dashed lines.

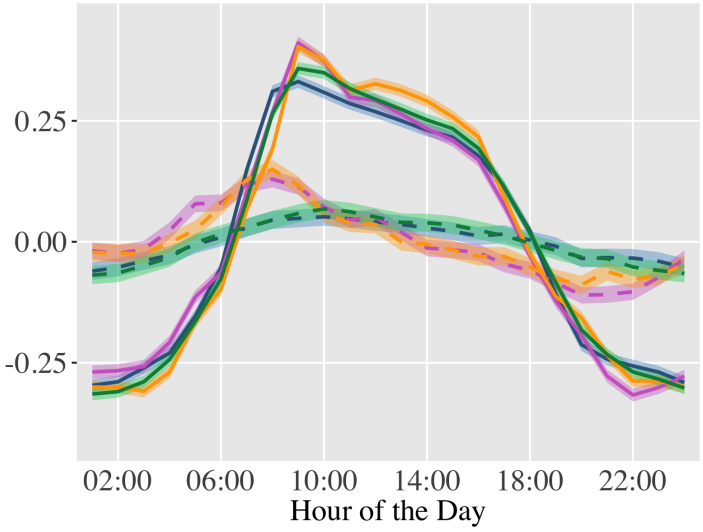


Figure 4.9: The four seasons are represented in blue (winter), purple (spring), orange (summer) and green (fall). Dashed lines are non-working days, and solid lines are working days. The shaded bands denote the 95% credible intervals.

All four effects for each day type in Figure 4.9 are similar, but for non-working days

there is more distinction between the seasons. Or rather, we see groupings of two and two seasons, where spring and summer are very similar, and so are fall and winter, and there is a clear difference between these two groups. For working days, this grouping does not appear.

Based on the result of the effects for non-working days in Figure 4.9, we replace the four seasons in season_t by two six-month seasons in Model S2. Spring/summer contains spring and summer, and fall/winter contains fall and winter. This way we reduce the model complexity in terms of the number of effects we have to estimate, which is favored by the DIC. The new effects for the daily cycles are plotted in Figure 4.10. The resulting profiles resemble the groupings we saw in Figure 4.9. We notice that the peaks occur at different times of the day in the spring/summer and in the fall/winter: for non-working days, the peak occurs about two hours before in the spring/summer than in the fall/winter, and the peak reaches a higher level here than in the fall/winter. The peak is higher in the spring/summer for working days as well, but it occurs at the same time as in the fall/winter. We also see a second, local peak at about 14:00 on working days in the spring/summer.

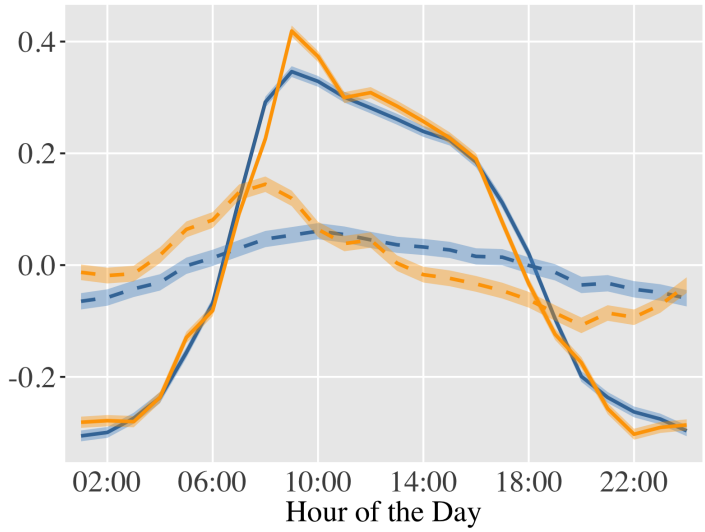


Figure 4.10: The daily profiles for the new six-month seasons are represented in blue ("winter") and orange ("summer"). Dashed lines are non-working days, and solid lines are working days. The shaded bands denote the 95% credible intervals.

4.3.2 Adding an Effect of Wind

Together with temperature and solar radiation, wind speed is considered one of the climatic factors that can influence the heat load of buildings. Since we saw a possible relationship between wind and heat consumption in Figure 2.9, we include a smooth effect of wind in Model W by adding

$$f(\text{wind}_t) \sim \text{RW2}$$

to the basic model. The resulting effect of wind is shown in Figure 4.11. There is a small increase in heat consumption for low temperatures, but for higher temperatures the relationship is weak. We will see later in Section 4.4.2 that there is a difference between the estimated effect of wind in Oslo and Trondheim, where the common effect in Figure 4.11 resembles the effect in Trondheim. A reason for the effect in Trondheim to dominate for high wind speeds in the common effect could be that the highest observations of wind speeds are recorded in Trondheim: the highest value recorded in Oslo was about 12.5 m/s, while in Trondheim it was just above 15 m/s.

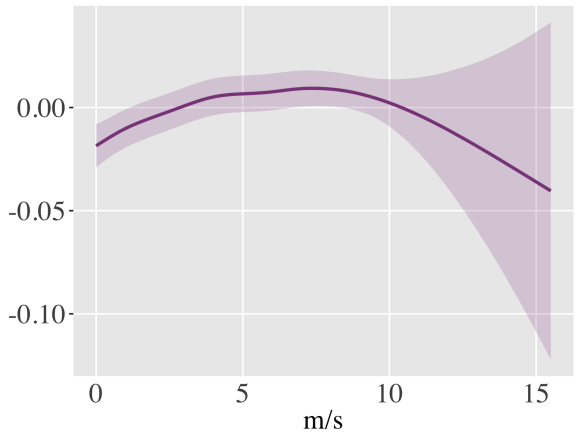


Figure 4.11: Effect of wind. Translucent band indicates the 95% credible interval.

4.3.3 Adding a Linear Trend

We include in the model a global time trend in the form of a linear regression with respect to the number of months from the beginning of the observation time. This is called Model M. Such a global trend could represent for example an improvement of the infrastructures that leads to decreasing energy usage. The exploratory analysis in Section 2.1.5 suggests the presence of a small negative trend in heat consumption over time that appears more prominent in Trondheim.

The resulting slope of the time trend is shown in Figure 4.12 in terms of its posterior marginal distribution. The effect has a small value, but it is significant. Moreover, the sign of the effect corresponds to the sign of the slopes in Figure 2.6 and 2.7.

Also by looking at the effect of time in Model M compared to the basic model, we can get a hint on how well a linear trend is suited for the data. This is because $f(t, \text{county})$ picks up the effects that are not captured by the deterministic part of the model. It was intended to catch possible correlation between the fitted values, and if we have a, say, monthly decrease that is not modeled by any of the other elements of the model, this will appear in $f(t, \text{county})$.

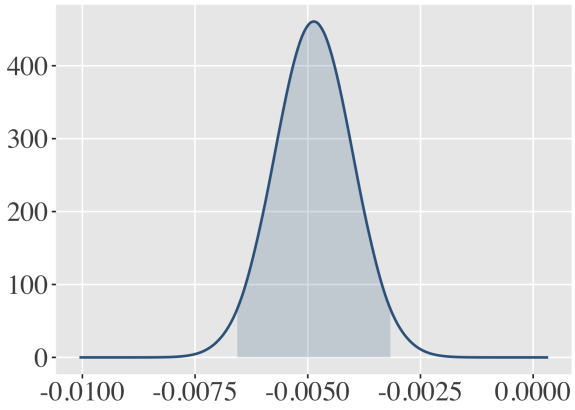


Figure 4.12: The posterior marginal distribution of the effect of month. Shaded area denotes the 95% credible interval.

In Figure 4.13 we have compared the effect of time from the basic model and Model M. In both Oslo and Trondheim the effects mostly overlap for the two models, but we see a slight monthly decrease in the effect from the basic model compared to the effect from Model M, indicating that a linear trend is suitable, although it is very small. There is no notable difference between the effects in Oslo and Trondheim.

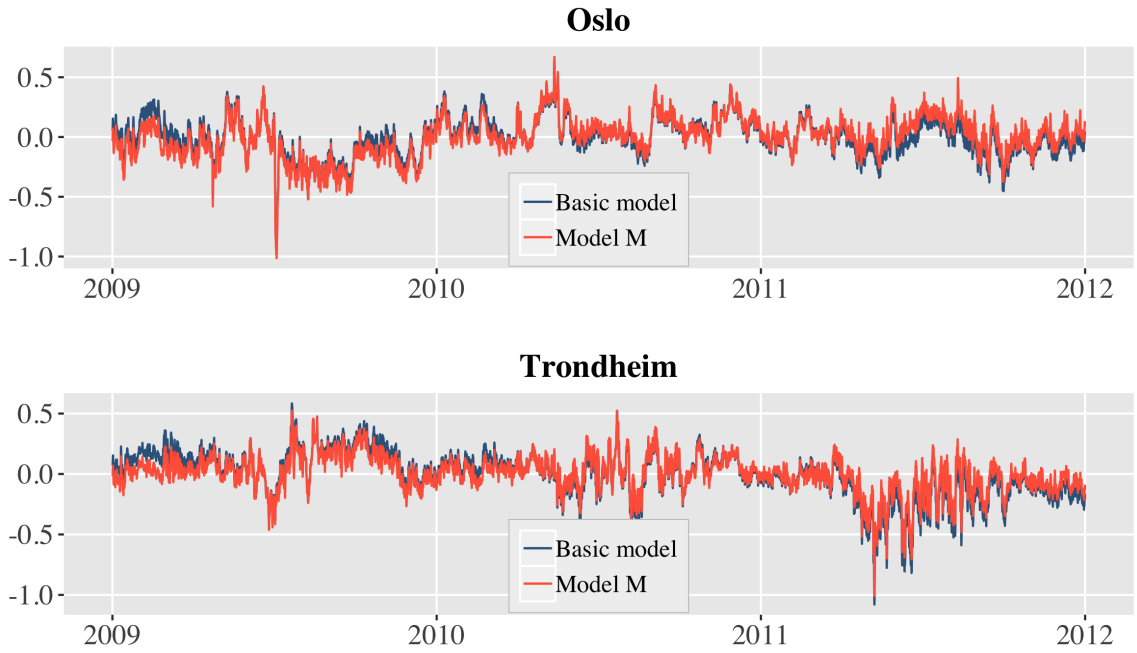
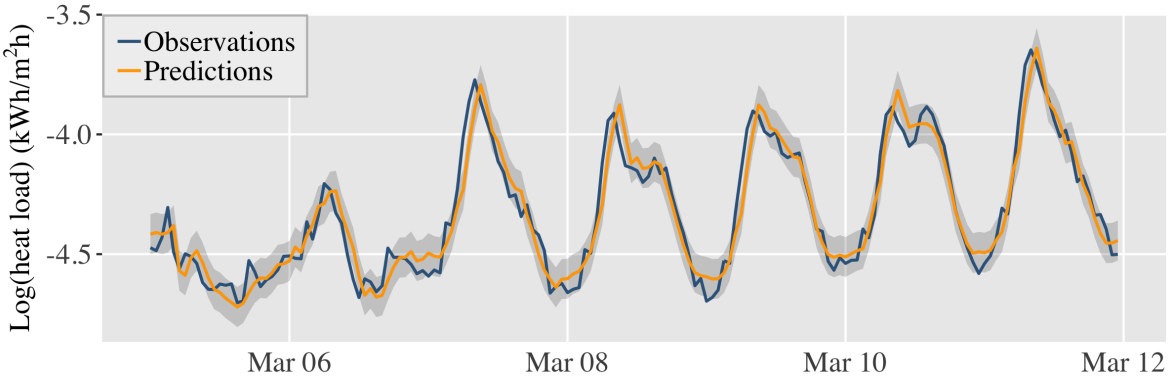


Figure 4.13: Effect of time in the basic model (blue) and in Model M (red).

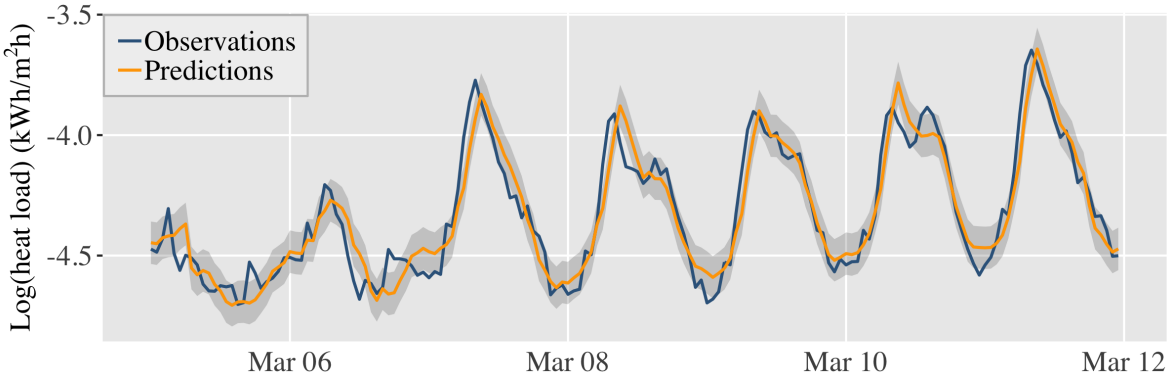
4.3.4 Lagged Covariates

We talked about in Section 4.2.1 that we are not sure about how long it takes for changes in the temperature to appear as an increase or decrease in heat consumption.

The difference in when the peak load for one day appears for the observations and for the fitted values from the basic model in Figure 4.5 suggested that lagged values of the temperature covariate could be a better choice for input to the model than the actual temperature. Based on the findings in Skagestad (2018), where likewise buildings in Norway were studied, we choose a lag of $k = 2$. The model with a lagged temperature covariate is called Model LT, and the fitted values from the same week as in the top of Figure 4.5 is shown in Figure 4.14, together with the fitted values from the basic model. No change or improvement of the placement of the daily peak loads from the fitted values can be seen from this figure.



(a) Basic model



(b) Model LT

Figure 4.14: The median of the predictive distribution and observations for one week in 2011 in Oslo along with the 95% prediction intervals.

4.4 Separating the Counties

The idea behind a common model for both counties is to invoke the possibility of learning from each other. For instance if the daily profiles are similar in Oslo and Trondheim, then estimating them on observations from both places renders more observations and thus a larger training set. However, if the effects in our model differ between the loca-

tions, the common effects would, in the worst case scenario, not be suitable for neither of the locations. In this section we will first fit the basic model introduced earlier on one county at a time and compare the new effects for each county. Next, more models from the previous section are estimated on one county at a time. When we only have observations from one county in the training set, and we make predictions for a year in the same county, we say that we use a *one-county* model, in contrast to the *common* models presented above.

When we estimate the basic model on each county separately, the interaction in Equation (4.1) with county is removed from $f(t)$, and the fixed effect of county is omitted. By estimating models on one county at a time, we essentially let all the effects in the model have an interaction with county, creating the possibility of differences between Oslo and Trondheim in all effects.

4.4.1 Effect of Temperature and Cycles

To get an impression of how similar (or dissimilar) these two one-county basic models are, we look at the smooth effects $f(\text{temp}_t)$ in Figure 4.15, and at $f(\text{hour}_t)$ and $f(\text{day}_t)$ in Figure 4.16.

For temperatures between 0°C and about 18°C the effect of temperature is more or less identical in Oslo and Trondheim, as seen in Figure 4.15. Furthermore, the relationship here is close to linear. For temperatures outside this interval, we have a larger effect in Oslo. The decrease in effect for the highest temperatures is only seen in Oslo. The difference in effect for the extreme temperatures is probably caused by the uncertainty here due to few observations at said temperatures. The credible intervals are not shown in Figure 4.15 and 4.16 for readability reasons, but they would have revealed the same characteristic as seen in Figure 4.3, namely wider intervals at the ends of the temperature scale.

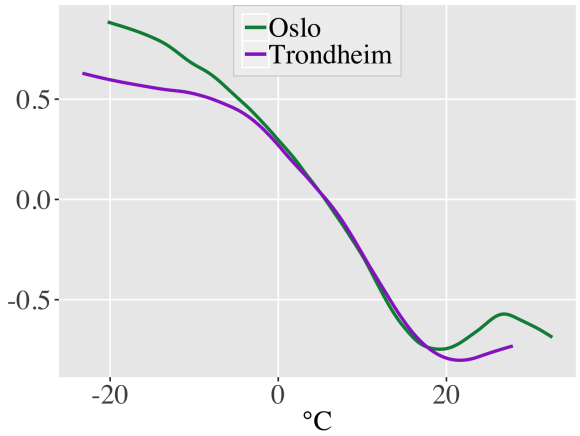
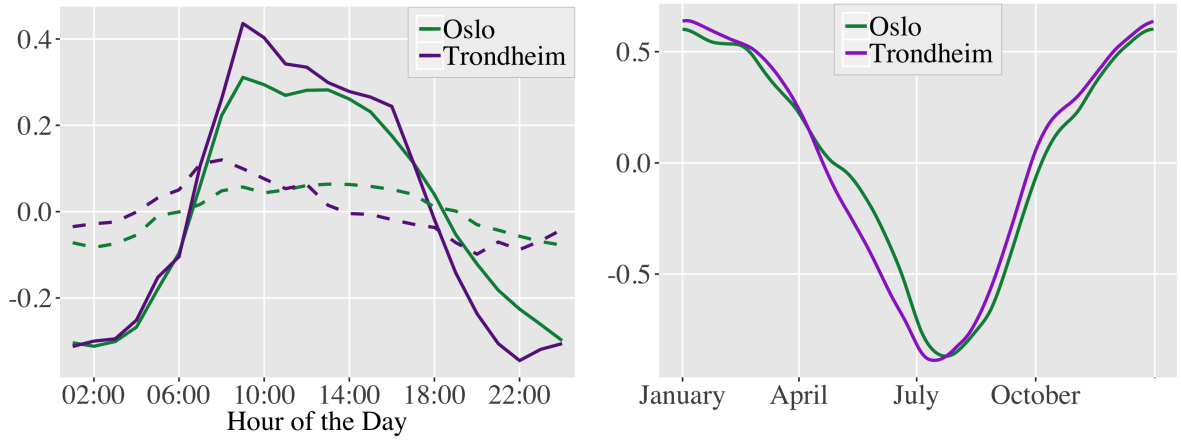


Figure 4.15: Effect of temperature.

The daily cycles for working days in Figure 4.16a resemble what we saw in Figure 2.5: the peak load in Trondheim is higher than in Oslo, and the heat consumption

experiences a steeper decrease at the end of the day in Trondheim. For non-working days, which are visualized as dashed lines, the shape during a day is similar in both counties, but is shifted some hours ahead in time for Oslo such that the peak load happens later in the day. For the annual cycle in Figure 4.16b, we also have similar shapes in Oslo and Trondheim where the one for Oslo is shifted slightly to the right compared to Trondheim.

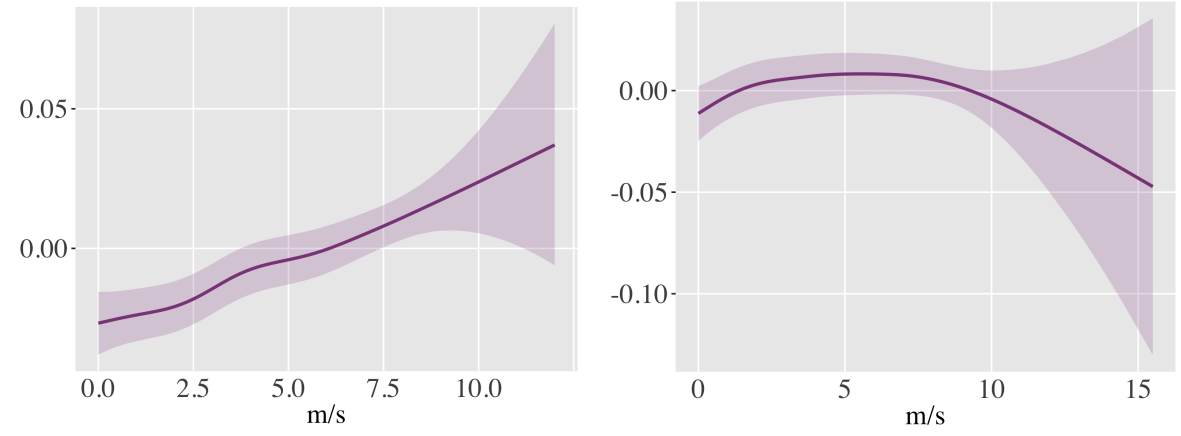


(a) Daily profiles. **(b)** Annual effects.

Figure 4.16: Daily profiles and annual effects in Oslo (green) and Trondheim (purple).

4.4.2 Effect of Wind

The effects of wind from the one-county Model W are displayed in Figure 4.17. We first of all notice how the effect of wind in the common Model W is more or less identical to the effect for Trondheim in Figure 4.17b. We also see that the positive linear effect



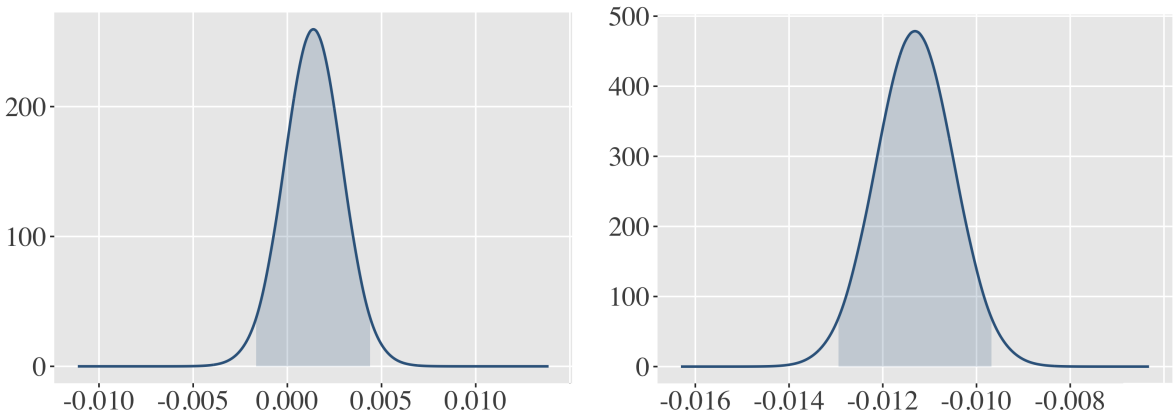
(a) Oslo **(b)** Trondheim

Figure 4.17: Effect of wind. Translucent bands indicate the 95% credible intervals.

with increasing wind speed, which we saw in Figure 2.9a, is present in the effect for Oslo in Figure 4.17a.

4.4.3 Effect of a Linear Trend

The effect of the linear trend in the common Model M was small, but significant. As with the effect of wind, we have from Chapter 2 that the mean monthly decrease is different between the counties. Thus we expect to see significant differences between Oslo and Trondheim in the new estimated linear effects. Our assumption is confirmed: in Figure 4.18 we see that the effect of month in Oslo is not significantly different from zero, and the effect in Trondheim is larger than in the common model. Moreover, the effects of time in Figure 4.19 from the one-county basic models and the one-county Model M show that, for Trondheim, the inclusion of a linear trend removes the decreasing mean from the AR term $f(t)$.



(a) Oslo **(b)** Trondheim

Figure 4.18: The posterior marginal distributions of the effect of month. Shaded areas denote the 95% credible intervals.

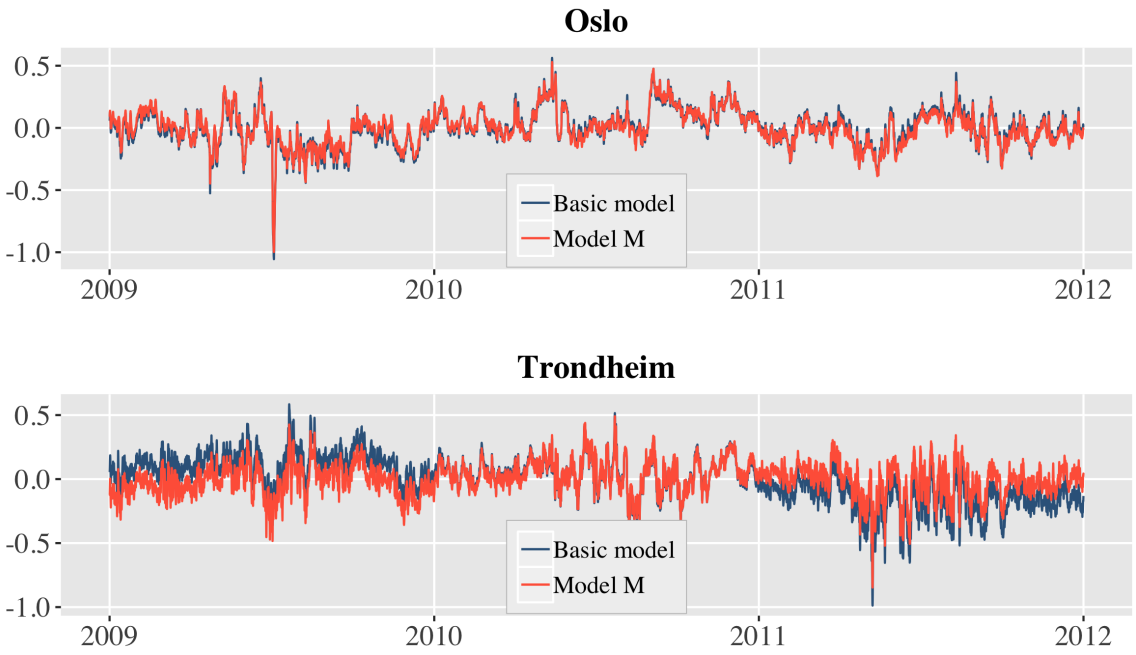


Figure 4.19: Effect of time in the basic model (blue) and in Model M (red).

4.5 Comparison of Goodness of Fit

We will in this section look at the values of DIC for the different models explored in the previous sections. DIC will not be our main criteria for evaluating which model offers the best fit to our data. This is because it is a measurement of *expected* predictive power for out-of-sample prediction. We are more interested in how our models perform when tasked with actually predicting for unknown observations. However, we still include the resulting values of DIC from estimating the various models as it is interesting to discover whether any of the models are strongly preferred compared to another, with respect to the DIC.

The DIC of all common models that are explored and estimated on all observations in this chapter are listed in Table 4.1. There is also included a column where the difference between each model and the basic model is listed. The first row in this column reads the effects in the basic model. We have in addition estimated all models separately for each county. The resulting DICs for the one-county models are shown in Table 4.2. We have omitted the column with model specification here as it is identical to the column in Table 4.1, with the exception of the county covariate in $f(t)$, and the linear effect of each county that is removed from $\boldsymbol{x}^\top\boldsymbol{\beta}$.

Model	Formula	DIC
Basic model	$\boldsymbol{x}^\top\boldsymbol{\beta} + f(\text{temp}_t) + f(\text{day}_t)$ $+f(\text{hour}_t, \text{daytype}_t) + f(t, \text{county})$	-79 563
Model S4	Added interaction with four seasons and daytype_t	**-80 256**
Model S2	Added interaction with two seasons and daytype_t	-80 141
Model W	Added a smooth effect of wind	-79 576
Model M	Added a long-term time effect	-79 634
Model LT	Used lagged temperatures	-78 098

Table 4.1: A list of the different models and their DIC. The items in the Formula column as of the second row refer to the alterations made to the basic model. All observations are included in the model estimations.

Model	DIC_O	DIC_T
Basic model	-41 905	-39 393
Model S4	**-42 388**	**-40 382**
Model S2	-42 099	-40 212
Model W	-41 898	-39 403
Model M	-41 898	-39 405
Model LT	-40 056	-39 584

Table 4.2: A list of the different models and their DIC. Subscripts O and T denote Oslo and Trondheim, respectively. All observations in one county were included in the model estimation of the same county.

In Table 4.1 Model LT has the highest DIC and Model S4 has the lowest. Considering the high absolute value of all the DICs in Table 4.1, the differences between them is not that large, and there is not one model that stands out with a significant low value

of DIC. Likewise, the highest value of DIC found in Table 4.1 is not that different from the rest.

In both Table 4.1 and 4.2 Model S4 gives the lowest DIC, and Model LT has the highest value for the common model and for the one-county model in Oslo. In Trondheim, we have that the one-county basic model yields the highest value of DIC.

4.6 Predictive Performance

Until now we have included all observations from the data set when estimating the models. This way we learn whether the models are able to catch the underlying process causing the observations, and we are also able to decide which effects are significant. Additionally we have compared the DIC of the different models. But as we said in Section 4.5, we are more interested in how the models perform in terms of the RMSE and CRPS when predicting for new observations.

4.6.1 RMSE and CRPS

As in Section 4.5 we start by considering the models that are common for both counties. We create a training set where we remove one year from one county at a time. For example, when predicting for 2011 in Oslo, we use the observations from 2009 and 2010 in Oslo, and from 2009-2011 in Trondheim. We then compare the predicted values to the observations. The values of RMSE and CRPS that we will consider are the mean annual ones. They are found in Table 4.3.

	$\overline{\text{RMSE}}_O$	$\overline{\text{RMSE}}_T$	$\overline{\text{CRPS}}_O$	$\overline{\text{CRPS}}_T$
Basic model	0.213	0.249	0.119	0.146
Model S4	0.215	0.248	0.121	0.144
Model S2	0.212	0.252	**0.118**	0.146
Model W	**0.211**	0.250	**0.118**	0.147
Model M	0.257	**0.239**	0.157	**0.138**
Model LT	0.221	0.249	0.124	0.147

Table 4.3: A list of the different models and their $\overline{\text{RMSE}}$ and $\overline{\text{CRPS}}$. Subscripts O and T denote Oslo and Trondheim, respectively. Observations from both counties, excluding the year of prediction in the given county, were included in the training sets.

For the one-county models, the training set for predicting 2011 in Oslo is only the observations from 2009 and 2010 in Oslo, and none from Trondheim. Likewise for the training sets for prediction in Trondheim. The resulting mean annual RMSE and CRPS are listed in Table 4.4.

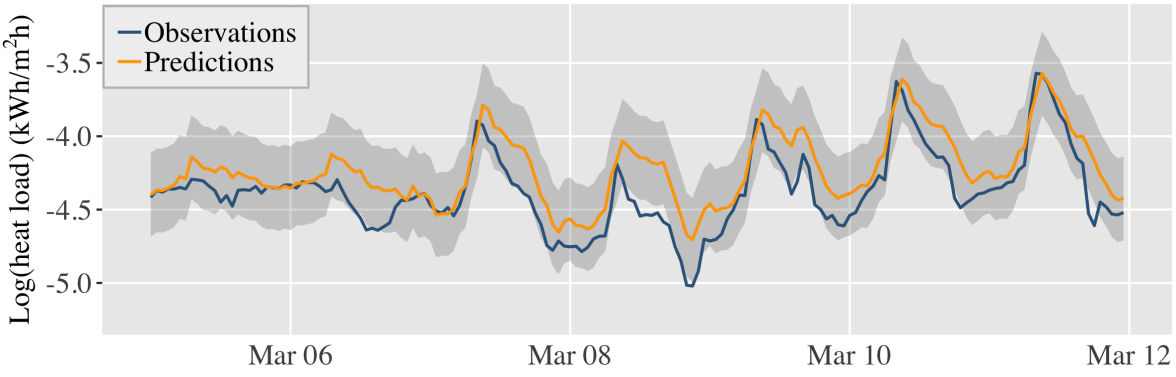
For all models except for one, Model M, we have lower values of RMSE and CRPS in Oslo. Moreover, here the RMSE and CRPS for all models in Table 4.4 are lower than in Table 4.3, where observations from Trondheim were also included in the training set. This is to be expected when the effects are trained on data from Oslo to predict for data from Oslo. We do not see the same result for Trondheim: for all but two models, the RMSE and CRPS in Table 4.4 are *higher* than in Table 4.3.

	$\overline{\text{RMSE}}_O$	$\overline{\text{RMSE}}_T$	$\overline{\text{CRPS}}_O$	$\overline{\text{CRPS}}_T$
Basic model	0.208	0.256	**0.115**	0.152
Model S4	0.210	0.252	0.118	0.146
Model S2	0.207	0.264	**0.115**	0.151
Model W	**0.206**	0.258	**0.115**	0.153
Model M	0.249	**0.209**	0.146	**0.115**
Model LT	0.215	0.255	0.121	0.152

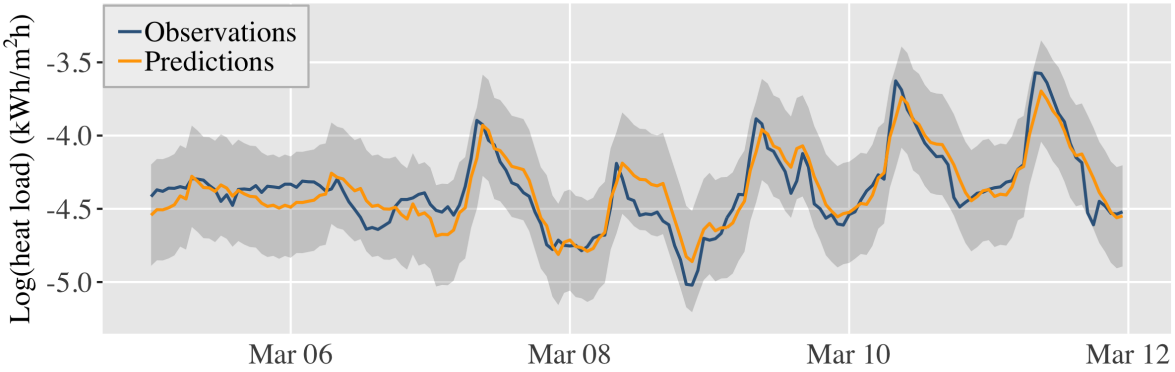
Table 4.4: A list of the different models and their DIC, $\overline{\text{RMSE}}$ and $\overline{\text{CRPS}}$. Subscripts O and T denote Oslo and Trondheim, respectively. Two years of observations from the county that is predicted for are included in the training set.

4.6.2 Predictions

In Trondheim, the model that gave the lowest value of RMSE and CRPS was the one-county Model M. The findings in Section 2.1.5 suggested that a linear trend with a monthly decrease was suitable for Trondheim, and the resulting RMSE and CRPS endorse this assumption. In Figure 4.20 we show predicted values using both the one-county basic model and the one-county Model M, along with observations, for one week in 2011 in Trondheim. The predictions in Figure 4.20a follow the same shape as the observations, and they mostly stay inside the 95% prediction interval. We do however see an overestimation of the observed heat consumption throughout the week. In Figure 4.20b, where the predictions from Model M are shown, this overestimation is no longer present, and we also have that the observations stay inside the prediction interval for all predictions.



(a) Basic model



(b) Model M

Figure 4.20: The median of the predictive distribution and observations from one week in 2011 in Trondheim using the one-county basic model (top) and the one-county Model M (bottom). Shaded bands denote the 95% prediction intervals.

A year from the one-county Model M is shown in Figure 4.21. What we saw in Figure 4.20b, which was a good fit in a month in the spring, is also seen in Figure 4.21. In fact, the model performs remarkably better during the first and last months of the year than in the summer, where we have an overestimation of the demand. Since Model S4 is designed to allow for both different levels and daily cycles during the year, and since this model reported the lowest value of DIC, we look at the predictions from the one-county Model S4 in Trondheim in the same year as in Figure 4.21. The result is shown in Figure 4.22, and we see that the overestimation in the summer has not been reduced with Model S4 compared to Model M. We have a slightly higher level of all predicted values here, because the linear trend is not included in Model S4.

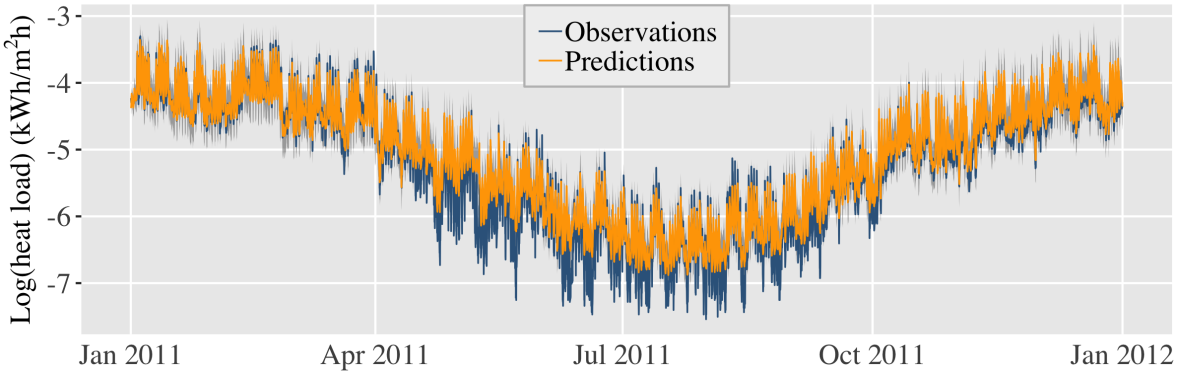


Figure 4.21: The median of the predictive distribution and observations for 2011 in Trondheim with one-county Model M. Shaded band denotes the 95% prediction interval.

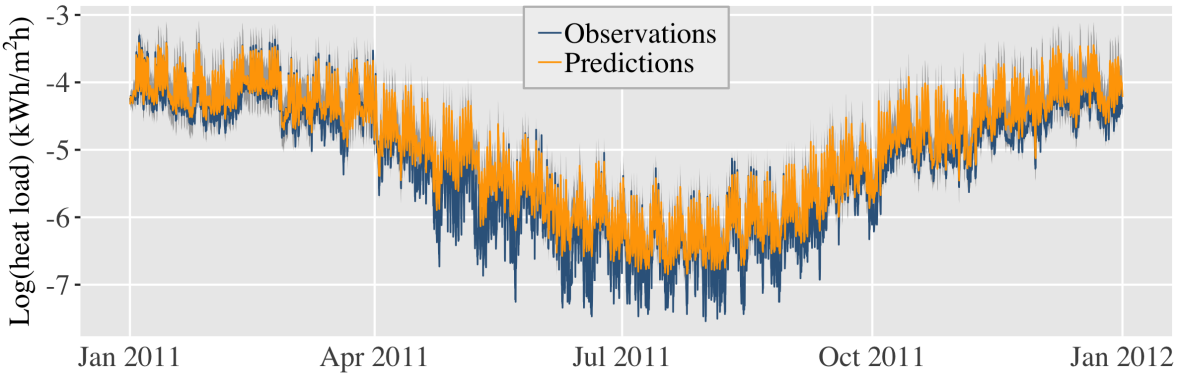


Figure 4.22: Median of the predictive distribution and observations for 2011 in Trondheim with one-county Model S4. Shaded band denotes the 95% prediction interval.

We saw in Figure 2.3 that there is more variability in the heat consumption in Trondheim, and in Figure 4.21 and 4.22 we see that the models do not capture this variability in the summer.

For completeness, we include a plot of the residuals from predicting for 2011 in Trondheim with the one-county Model M. This is shown in Figure 4.23. As is expected from both Figure 4.6 and 4.21, we have higher variance in the residuals in the summer. We also have correlation.

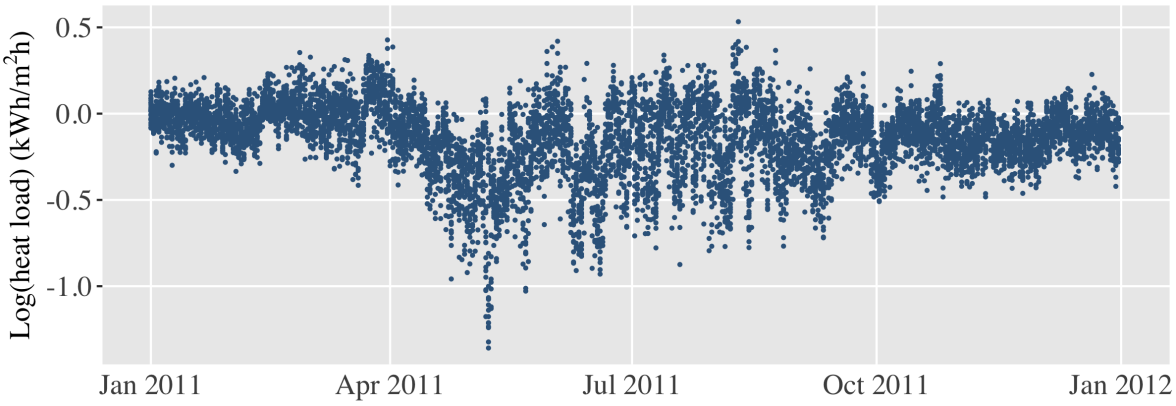
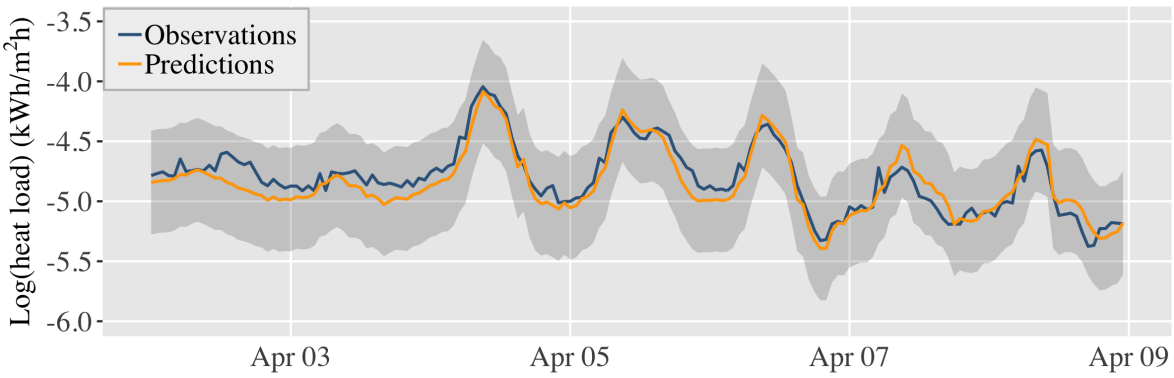
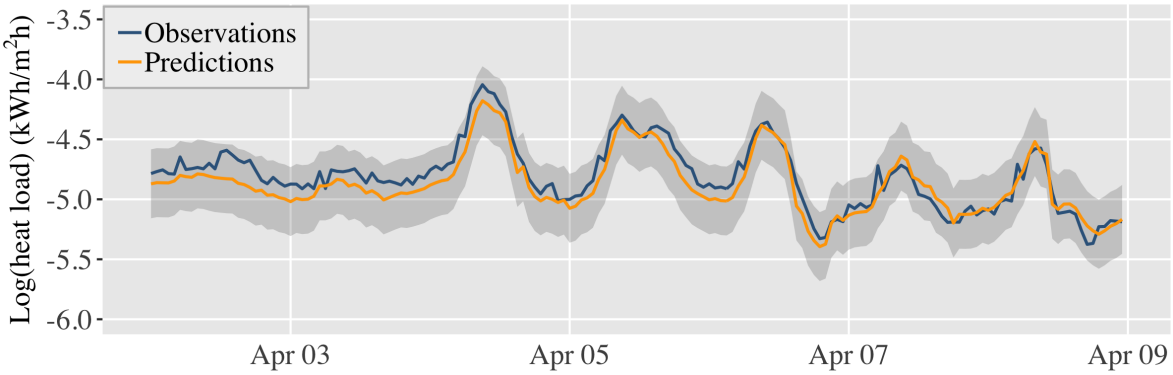


Figure 4.23: Residuals in Trondheim.

The models with the lowest values of RMSE and CRPS for Oslo was the one-county Model W. As the difference in RMSE between the one-county Model W and the basic model was minor, we choose to compare predicted values made with the common Model W and the one-county Model W. The difference in RMSE between these two models was, for Oslo, slightly higher. These predictions are given in Figure 4.24. The most notable difference between these two models, is the width of the prediction intervals. When all observations from both counties are included in the training set, except the year in Oslo we make predictions for, it is wider than when we only include two years of observations from Oslo. The predicted values in Figure 4.24a and 4.24b are more or less the same, due to the small impact the effect of wind yields, as seen in Figure 4.17.



(a) Common Model W



(b) One-county Model W

Figure 4.24: Median of the predictive distribution and observations from one week in 2011 in Trondheim using the common Model W (top) and the one-county Model W (bottom). Shaded bands denote the 95% prediction intervals.

A reason for this decrease in width of the prediction interval may be found in the effect of wind. What is special about this effect, compared to the effects included in our other models that rather yield wider intervals for the one-county models, is the large difference in the effect of wind between the counties. The effect of wind in the common Model W had a very large credible interval, which we saw in Figure 4.17 was caused by the effect of wind in Trondheim. Thus, when having a common effect of wind that is mostly influenced by the uncertain effect in Trondheim, more noise is added to the predictions and, in turn, the prediction interval is wider.

We also include the predictions for the whole year of 2011 in Oslo using the same models as in Figure 4.24. They are shown in Figure 4.25 for the common Model W and in Figure 4.26 for the one-county Model W. The predictions for the common and one-county Model W are similar. The difference is in the width of the prediction interval, as seen also in Figure 4.24, and that the one-county model underestimates the consumption more in the summer than the common model does.

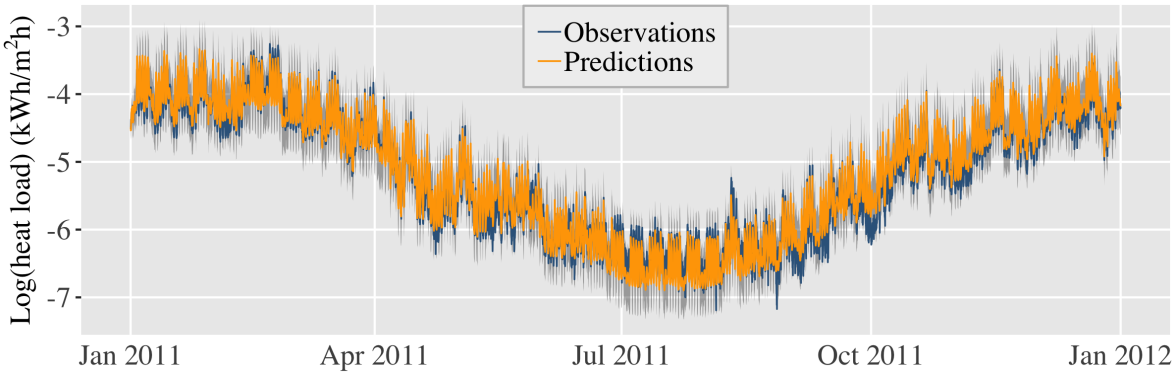


Figure 4.25: Median of the predictive distribution and observations for 2011 in Oslo with the common Model W. Shaded band denotes the 95% prediction interval.

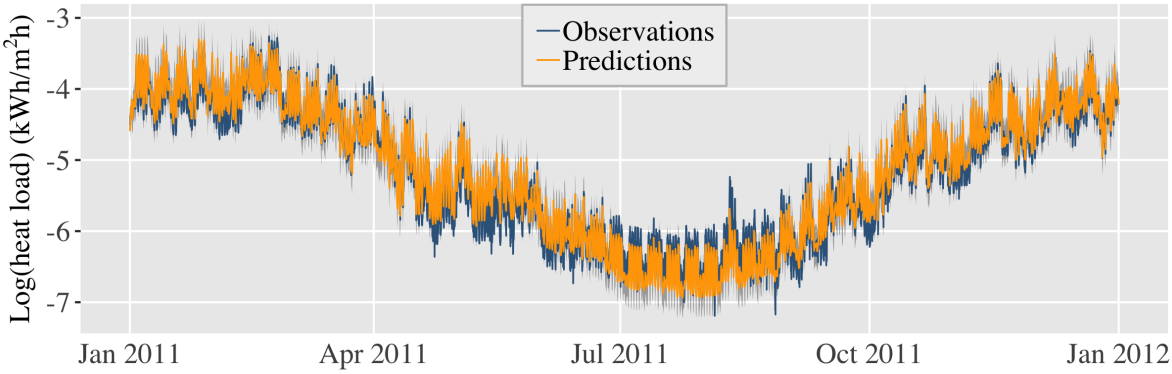


Figure 4.26: Median of the predictive distribution and observations for 2011 in Oslo with one-county Model W. Shaded band denotes the 95% prediction interval.

Chapter 5

Results and Discussion

In this chapter we will analyse how the inclusion of different covariates and interactions affects the predictive performance of the models in Chapter 4. Next, we will study how the predictive performance of our models compare to that of the model suggested by Sintef, before we discuss the findings in this work.

5.1 Main Conclusions From the Experimental Results

We will here summarize and draw conclusions from the experimental results in Chapter 4, with focus on the smooth effects.

5.1.1 Temperature

In Figure 4.20 and 4.24 we saw that the predictions follow both the shape and level of the observations when we have a varying level of the observed day-to-day heat consumption. This indicates that the effect of temperature is strong. This is in line with common knowledge that we use more heat when the weather is cold.

The shape of the estimated effect of temperature is similar to that of the piecewise linear relationship assumed by Sintef, shown in Figure 1.1. We saw that for a certain temperature range, the effect of temperature is close to linear, and for another range, namely higher temperatures, the effect stabilizes.

Figure 4.5 suggested that a lagged series of temperatures could be appropriate for our models. A lag of $k = 2$ was implemented and tested, showing no improvement in neither the DIC nor the RMSE or CRPS.

5.1.2 Daily Cycle

The estimated daily cycles from our models correspond well to the observed ones in Figure 2.5. We saw some small changes when we included an interaction with both four and two seasons, but this had little impact on the predictive performance in neither of the counties.

5.1.3 Wind Speed

The effect of wind appears to be significant in Oslo but not in Trondheim. When the effect of wind is included in the common Model W it appears to increase the uncertainty, while it improves the predictive performance in Oslo with the one-county Model W. The reason for this is not known.

5.1.4 Monthly Linear Trend

We revealed a monthly decrease in heat consumption in Figure 2.7. When a long-term time trend was included in the basic model, which created Model M, we saw improvements in the predictive power in Trondheim. Both in the RMSE and CRPS, and also by comparing the one-county basic model and the one-county Model M in Figure 4.20. Overall, Model M showed the largest improvement, with respect to RMSE and CRPS, compared to the basic model, among the models considered. Model M was also the model that produced values of RMSE and CRPS for Trondheim that were closest to the best results in Oslo.

5.1.5 Learning Between the Counties

A reason to estimate the models on Oslo and Trondheim together is that some of the effects are assumed to be alike. For example the effect of temperature and the annual cycle, which we confirmed in Figure 4.15 and 4.16b. However, we saw that the effect of wind is different, and in Trondheim there is a long-term trend that we do not have in Oslo.

In Oslo there was little difference in RMSE and CRPS between the one-county models and the corresponding common county models. The one-county models performed slightly better for each model. In Trondheim we saw lower values of RMSE and CRPS in general when we used the common models. However, for the model with the lowest value of RMSE and CRPS, Model M, the one-county model performed better. Model M stood out in Oslo by yielding a high value of RMSE and CRPS compared to the others.

We have data from two more buildings in Oslo than in Trondheim in our aggregated data set. A reason for that the common models yield better results in Trondheim than the one-county models could be that we have less data here, and if the data in addition is noisy, the predictions from estimating a model only on observations from Trondheim will contain noise.

5.1.6 Interaction with Season

In Table 4.1 and 4.2 we saw that the values of DIC of the models that included an interaction between daily cycle and season with four levels, Model S4, were the lowest ones recorded. However, the resulting effects from these interactions, shown for the common Model S4 in Figure 4.8, 4.9 and 4.10 showed that the fixed effects were not

significant, and the smooth effects showed little change from the basic model, where only the interaction between daily cycle and day type was included. Furthermore, in Figure 4.22, Model S4 did not improve the fit of the model in the summer compared to the fit of the one-county Model M in Figure 4.21. It appears that there is not an additional effect of season that is not already recognized by our models.

5.2 Comparison with the Sintef Model

The model presented by the researchers at Sintef in Lindberg et al. (2019) is actually a number of linear models, estimated for each hour of the day for different day types. Unlike the models presented in Chapter 4, no considerations regarding the skewness of the data is taken. The relationship between heat and temperature is defined differently at each side of the CPT - for heat consumption below the CPT there is a negative linear relationship, and for temperatures above, the consumption is temperature independent such that the relationship is modeled as a horizontal line. An ad hoc method is used to find the CPT as this temperature is unknown. The effects they include are the ones of temperature and day type mentioned above, in addition to daylight and monthly effects. The daylight effect is a dummy variable that is equal to one if the time t occurs during daylight, and zero otherwise, and the monthly effect is a factor for each month of the year. The effect of temperature is modeled in two terms: One direct temperature effect, and one 24-hour moving average temperature effect. A building specific fixed effect is also included. This is calculated for a number of buildings, and when predicting for an unknown building, an average of these effects for the buildings in the same category, e.g. office buildings, is employed.

This model was presented to us as a sort of calculator. All we need to (and can) do is to plug in the floor area of the buildings we want to make predictions for, along with a time series of temperature for one year. In return we receive the predicted values of heat consumption per hour for the selected year. No uncertainty measures are provided. Furthermore, we do not have access to any of the model parameters from the model by Sintef and can therefore not make any comparison of the effects in their model and ours. The exception is the effect of temperature, which we know is modeled as in Figure 1.1.

Since we want to compare their results to our predictions, which are heat per square meter, we enter 1 m^2 as the floor area. We compare the predictions we receive to transformed quantiles of our predictions. We transform them back to the original scale so we are in the same scale as Sintef. To be clear, we still work with heat consumption per square meter in this section, but no longer on the log scale.

5.2.1 Important Differences Between the Models Presented in This Work and by Sintef

Some alterations to our models were necessary to make sure the comparison with Sintef is reasonable. In the edition of their model that was used to make predictions for our aggregated data set, there was no distinction between regular days and holidyas. Thus, the days in our data set that were holidays and accordingly marked as non-working days, got a new label. Either working day if the holiday appeared on a Monday to Friday, or they stayed a non-working day if they appeared in the weekend. Further, their model is designed to predict for an arbitrary year, so by default January 1st is set to be a Monday. None of the years in our data set had this property. To ensure that the time series of temperature provided by us starts on a Monday, we move the start of the series to the first Monday of the year. In turn we get a little less than a year of observations.

The only drawback about the method of starting the year on the first Monday is that Sintef has separate model parameters for each month. So when we present a time series of temperature that we make their model believe starts on Monday January 1st, but actually starts at, say, Monday January 3rd, the last predictions in their January will be marked as January even though the observations actually were recorded in February. We believe that this will not cause major disadvantages in the predictive performance of their model. Firstly, because we assume two adjacent months do not have very different model parameters. And secondly, this problem only occurs for a minor part of the days of the year.

In 2009, the first Monday appeared on January 5th, while in 2010 and 2011 it appeared on January 4th and 3rd, respectively. That means that, at the most, we lose four days in one year to make predictions for.

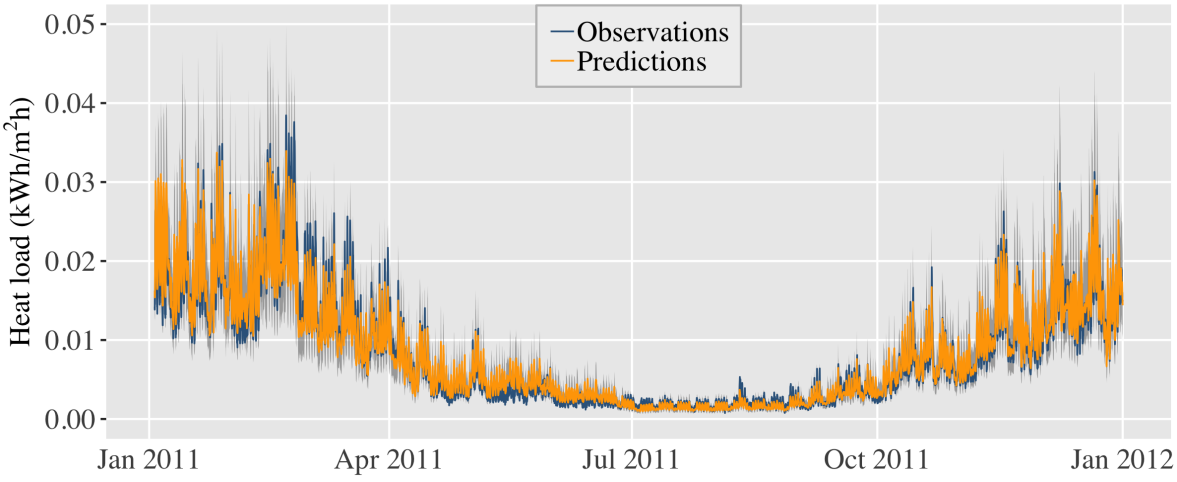
5.2.2 Models Used for Comparison with Sintef

For Oslo, the one-county Model W is used for making predictions. One year at a time, the model is trained on the two years from Oslo to predict for the year that is left out. In Trondheim, where the inclusion of wind showed little effect on the predictive performance, the one-county Model M is applied. This model is trained only on observations from Trondheim, and we make predictions in the same way as in Oslo. Plots of predictions from Model M together with predictions from the Sintef model are available in Appendix A.

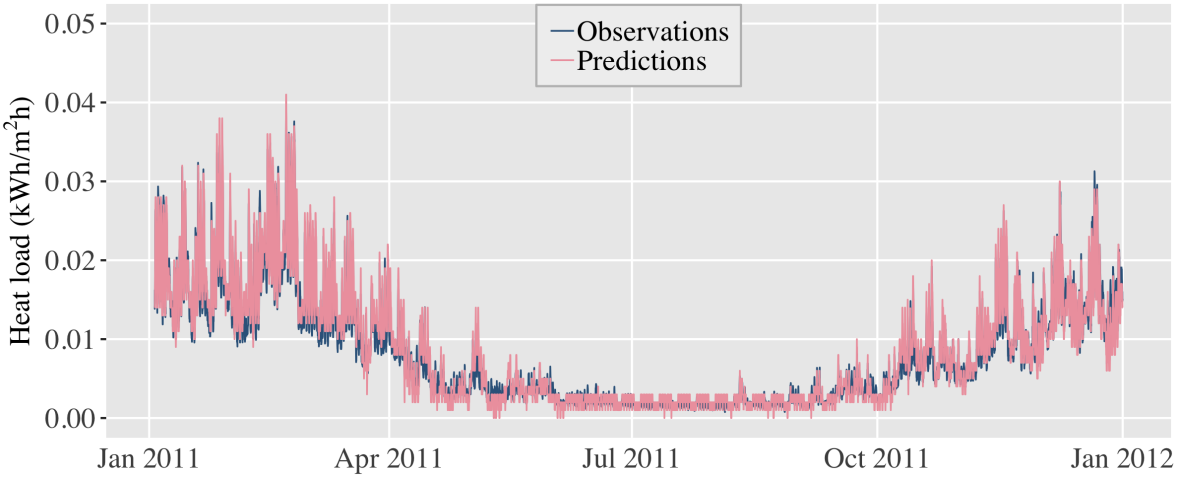
We note that we will not use any formal procedure to assess the model fits. This is because our models and the Sintef model are so different in how they are defined. The CRPS, for instance, is not possible to calculate for the predictions from Sintef as they do not come with uncertainty measures.

5.2.3 Results

In Figure 5.1 the result of predicting for 2011 in Oslo with the one-county Model W and the Sintef model is shown. The grey peaks that are present in Figure 5.1a come from the prediction interval. In general we see that both models resemble the form of the observations. We do see a more rigid pattern in the predictions from the Sintef model in the summer. This is because they assume temperature has no effect here. Their predictions in the summer are the estimates of the basic consumption that cannot be avoided.



(a) Model W



(b) Sintef

Figure 5.1: Top: Median of the predictive distribution in Oslo in 2011 estimated with the one-county Model W along with the 95% prediction interval and observations. Bottom: Predictions in the same year and county as above, made with the Sintef model, along with observations.

We zoom in on Figure 5.1 in Figure 5.2, where we have plotted the predictions from both models together for three weeks in 2011 in Oslo. The time series in this figure that are labeled "Predictions" refer to the predictions made by our model, and the time

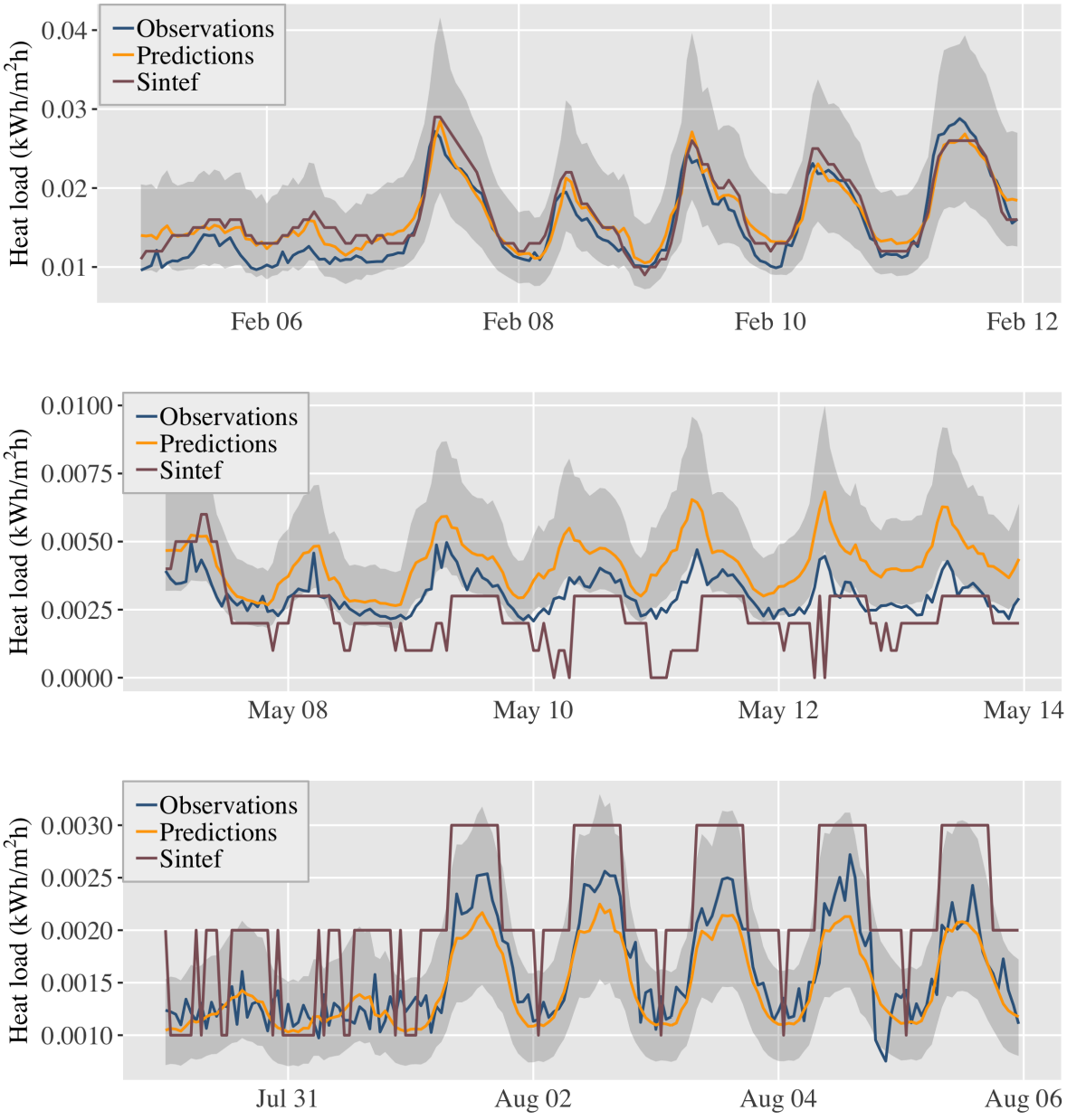


Figure 5.2: Median of the predictive distribution using the one-county Model W, predictions from the Sintef model and observations from three weeks in 2011 in Oslo. Shaded bands denote the 95% prediction intervals of the predictive distribution from using Model W.

series labeled "Sintef" mark the predictions made by the Sintef model. In February both models perform well, following the shape of the observations. The temperatures in this week were in the range where the effect of temperature in our model is linear. Thus the effect of temperature, for this week, is modeled in the same way in our model as in the Sintef model.

For the other two weeks, which are in the spring and summer, there are large differences between the shape of the observations and the predictions from Sintef, as we also saw in Figure 5.1 and commented above.

5.3 Further Work

As discussed in Section 4.2.1 we have correlation and structure in the residuals which could be investigated further. The inclusion of informative priors on the hyperparameters, in addition to an analysis of the prior sensitivity of these parameters, would also be interesting to consider as we did not put much effort into choosing the priors.

Finally, for newer observations of heat prediction than the ones considered here, observations of solar radiation is available. A thorough study on the relationship between solar radiation and heat consumption would, too, be interesting to consider.

5.4 Discussion

In this work we have presented several latent Gaussian models for prediction of heat consumption in an aggregated set of office buildings. They include the weather conditions temperature and wind speed, cycles for the daily, weekly and annual load, in addition to a linear trend, fixed effects and interactions with the seasons.

The motivation for modeling heat consumption was to come up with a model that could compete with the one presented by Sintef in terms of predictive power. A comparison between this model and the models presented in this work was made in Section 5.2.3 in Chapter 5. The result is that our models follow the patterns of the observations more closely than the Sintef model. In addition, we provide uncertainty measures which Sintef could not provide. By modeling the effect of temperature with a smooth effect, we did not need to go through the trouble of finding the CPT. The Sintef model provides more variability in the predictions which results in overestimation of the peak loads for various parts of the year. Our models tend to both over- and underestimate the peak loads. If one is interested in modeling the peaks rather than the expected values, one should use something like quantile regression (Fahrmeir et al., 2013).

We estimated all the models that were explored both on observations from Oslo and Trondheim together, and on observations from one county at a time by separate one-county models. For Oslo, the lowest values of RMSE and CRPS, although only marginally, for all proposed models, were achieved for the one-county models. The model that offered the best fit for this county was the one-county model that included a smooth effect of wind, Model W, although the effect of wind was weak. For Trondheim

the difference in RMSE and CRPS between the best performing model and the second to best was bigger than in Oslo. The model that offered the best fit here was the one-county Model M which included a long-term time trend. For all models except for this, Oslo performed better in terms of RMSE and CRPS.

When the one-county models are used for prediction, only two years of observations are included in the training set. A training set of this size compared to the number of predicted values, which is one year, is considered small.

In conclusion, the models proposed in this thesis present satisfactory results in terms of predictive power and the ability to model the heat consumption in Oslo and Trondheim. Correlation and lack of constant variance in the residuals suggest that the models will benefit from further work in terms of informative priors on the hyperparameters, defining new effects for the covariates and possibly the inclusion of new covariates such as solar radiation, for time series of heat consumption where this is available.

Bibliography

F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen. *Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis*. Springer, 2017.

M. Blangiardo and M. Cameletti. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley, 2015.

M. E. El-Hawary. *Advances in Electric Power and Energy Systems: Load and Price Forecasting*. John Wiley & Sons, 2017.

L. Fahrmeir, T. Kneib, S. Lang, and B. Marx. *Regression: Models, Methods and Applications*. Springer, 2013.

A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 2013.

X. Hue and I. Steinsland. Spatial modeling with system of stochastic partial differential equations. *WIREs Comput Stat*, 2016.

F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac. Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*, 2015.

A. Lenzi, I. Steinsland, and P. Pinson. Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels. *Environmetrics*, 2018.

K. B. Lindberg, S. J. Baker, and I. Sartori. Modelling electric and heat load profiles of non-residential buildings for use in long-term aggregate load forecasts. *Utilities Policy*, 2019.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian Inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2009.

H. Rue, A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren. Bayesian Computing with INLA: A Review. *Annual Review of Statistics and Its Application*, 2016.

R. Skagestad. Electricity Demand Forecasting with Gaussian Process Regression. Master's thesis, Norwegian University of Science and Technology, Trondheim, 2018.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 2002.

A. K. Srivastava, A. S. Pandey, and D. Singh. Short-term load forecasting methods: A review. *2016 International Conference on Emerging Trends in Electrical Electronics Sustainable Energy Systems (ICETEESES)*, 2016.

H. Takeda, Y. Tamura, and S. Sato. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy*, 2016.

J. W. Taylor and R. Buizza. Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 2003.

J. W. Taylor, L. M. de Menezes, and P. E. McSharry. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 2006.

R.-C. Tsai. Lecture notes in Introduction to Bayesian Statistics. Department of Mathematics, National Taiwan Normal University, May 2015.

R. Zheng and H. Bakka. Analysing unemployment data with the AR1 Model. `https://haakonbakka.bitbucket.io/btopic115.html`, 2018. Accessed: 2019-05-19.

Appendix A

Additional Results for Trondheim

Some results for Trondheim, which were showed for Oslo in Chapter 4 and 5, are included here. We first look at the results from estimating the common basic model in Trondheim, and next we will look how our predictions for Trondheim compare to those of the Sintef model. This is done in the same way as in Section 5.2.3, namely that we transform the median of our predictions to the original scale and compare them to the heat consumption that the Sintef model has predicted for 1 m^2.

A.1 Basic Model

The fitted values for Trondheim from the common basic model is shown in Figure A.1 for one year and in Figure A.2 for two weeks. The residuals from 2011 are shown in Figure A.3, and the ACF of these residuals are plotted in Figure A.4. We see the same results as for Oslo in Section 4.2.1.

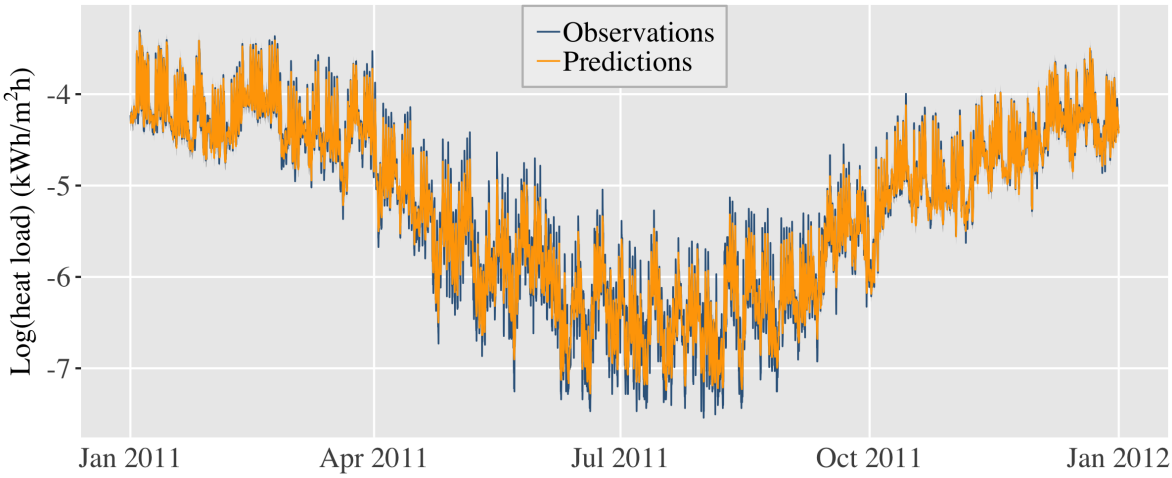


Figure A.1: Median of the predictive distribution and observations for a year in Trondheim.

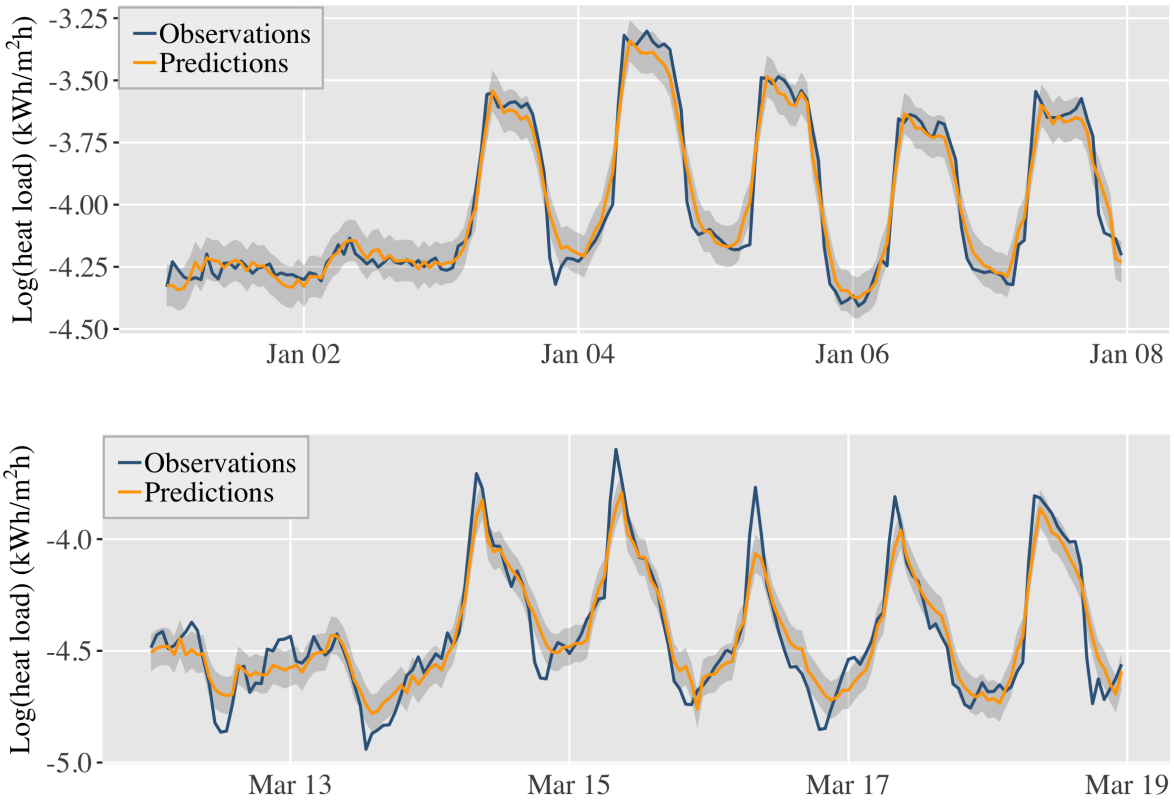


Figure A.2: Median of the predictive distribution from the basic model and observations for two weeks in 2011 in Trondheim along with the 95% prediction intervals.

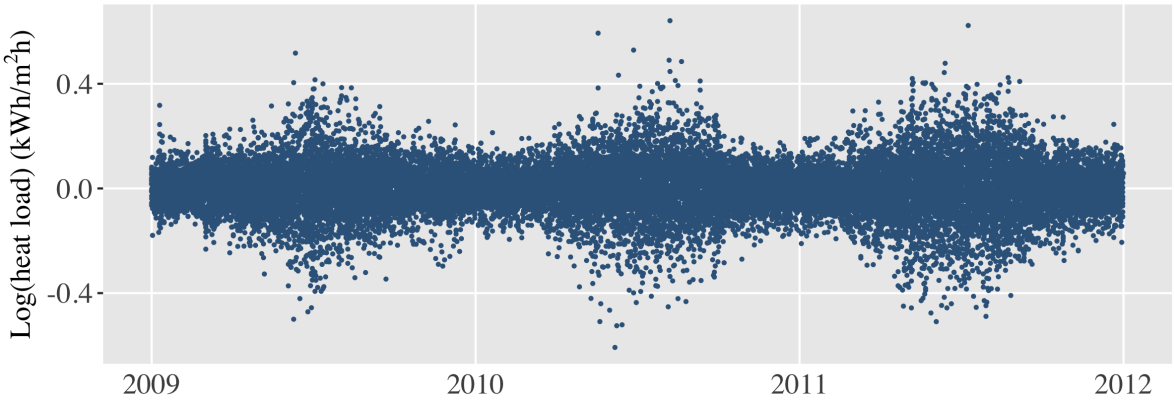


Figure A.3: Residuals in Trondheim.

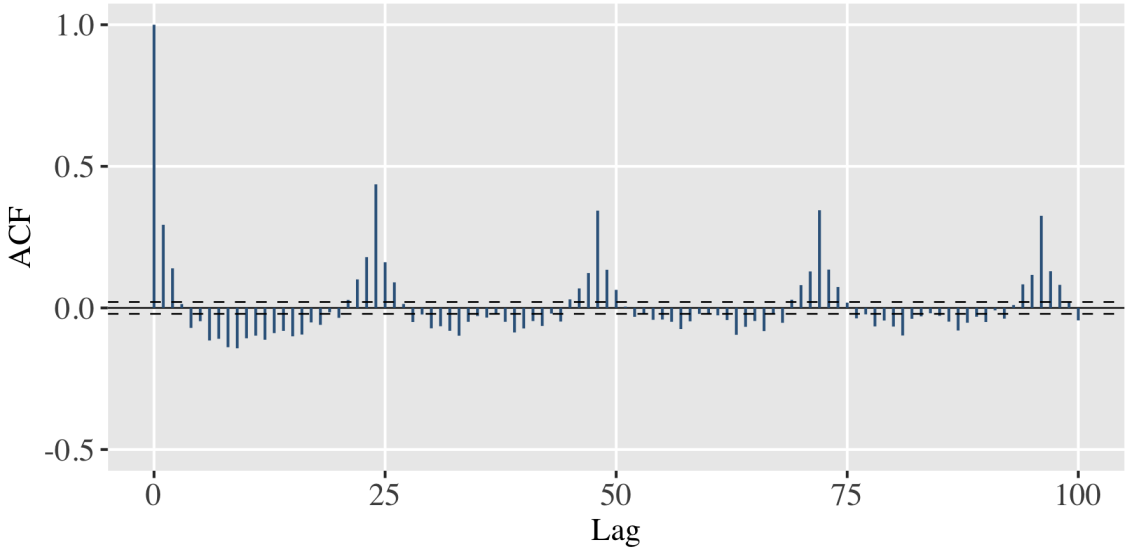
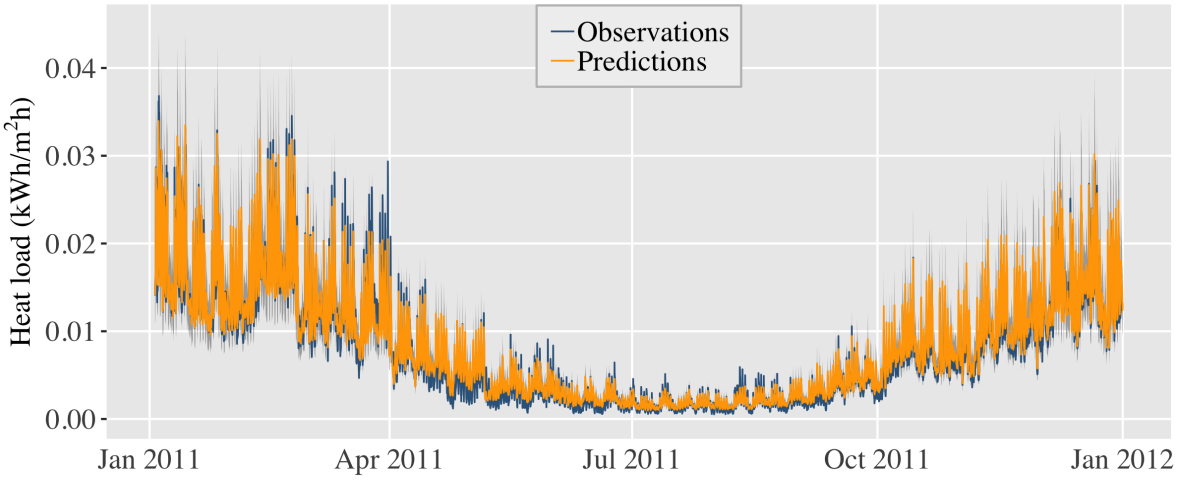


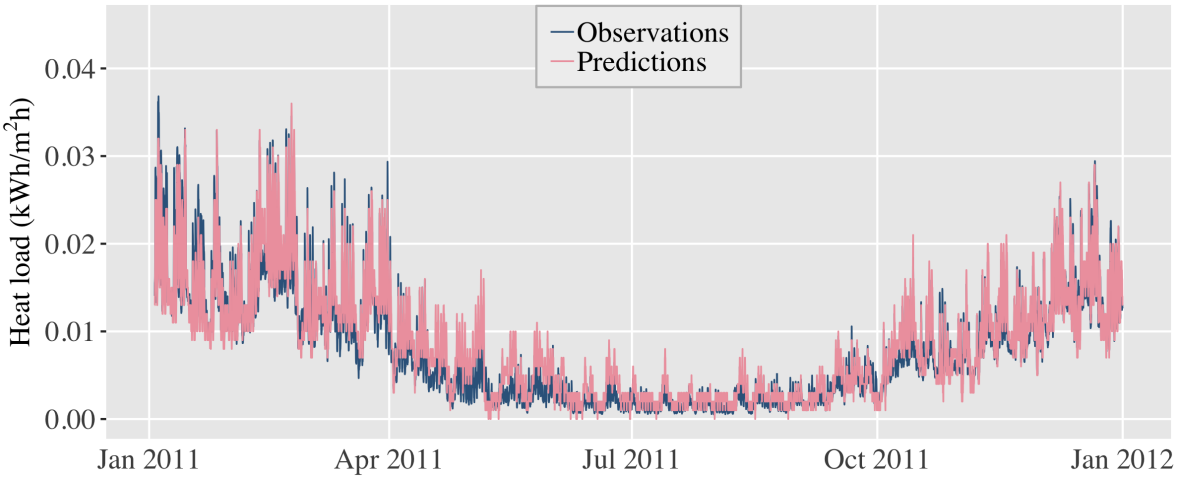
Figure A.4: ACF of residuals in 2009 in Trondheim.

A.2 Comparison with Sintef

In Section 5.2.3 we showed predictions made for Oslo for one year and for three weeks. Corresponding results for Trondheim are shown in Figure A.5 and A.6, comparing the predicted values to those obtained by the Sintef model.



(a) Model M



(b) Sintef

Figure A.5: Top: Median of the predictive distribution in Trondheim in 2011 estimated with the one-county Model M along with the 95% prediction interval and observations. Bottom: Predictions in the same year and county as above, made with the Sintef model, along with observations.

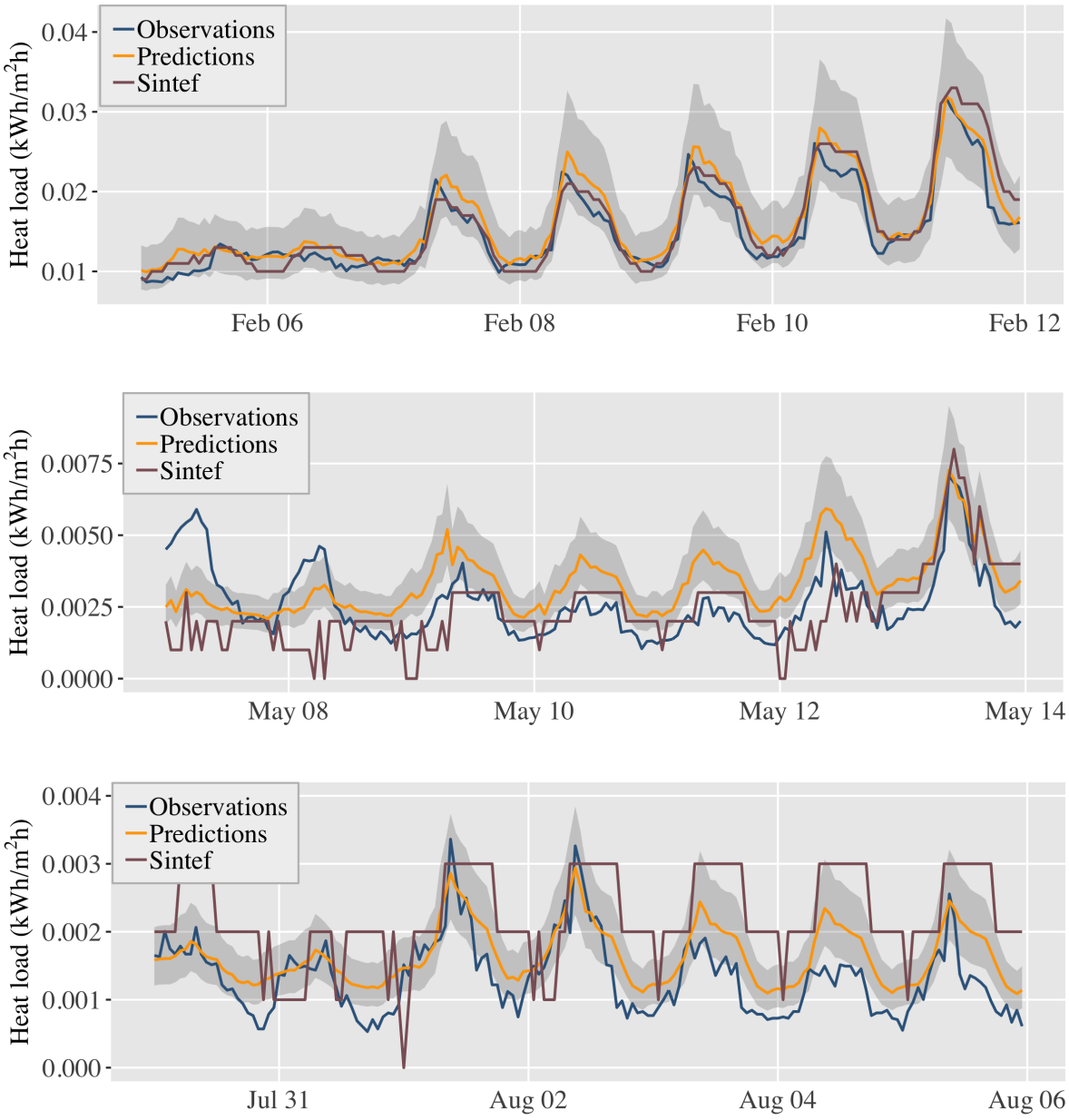


Figure A.6: Median of the predictive distribution using the one-county Model M, predictions from the Sintef model and observations from three weeks in 2011 in Trondheim. Shaded bands denote the 95% prediction intervals of the predictive distribution from using Model W.

Appendix B

Implementation in R-INLA

An example on how to make predictions for log transformed heat consumption per square meter in R-INLA is presented here. We will illustrate how to make predictions for 2011 in Oslo by using the one-county Model W.

B.1 Data Input

A data frame called `data` with nine columns and 26 280 rows with information about Oslo from the aggregated data set is created to use as input to the model. The first six rows are displayed here:

```
head(data)

##           y temp wind hour day daytype daytype2 t year
## 1 -4.119143 -5.7  0.5    1   1       2        2 1 2009
## 2 -4.151404 -5.4  0.3    2   1       2        2 2 2009
## 3 -4.131119 -5.7  0.6    3   1       2        2 3 2009
## 4 -4.056907 -5.6  0.8    4   1       2        2 4 2009
## 5 -3.989406 -6.4  0.2    5   1       2        2 5 2009
## 6 -3.908487 -5.6  2.6    6   1       2        2 6 2009
```

`y` is the time series of the log transformed heat consumption in Oslo per square meter, and `temp` and `wind` are the time series of outdoor temperature and wind speed. `hour` is a column that repeats the sequence `1,...,24`, once for every day, to be input to the smooth effect that models the daily cycles. `day` is similar to `hour` in that matter, only that it repeats the numbers `1,...,365`, 24 times for each number, for the three years of observations. Thus we are able to mark every observation by the day of the year it belongs to. `daytype` and `daytype2` are identical, but as we use them as input in two effects in our model, we need to make copies of them. They refer to what day type we have. `1` denotes a working day, and `2` denotes a weekend or a holiday. Since our time series starts on January 1st, and this is a public holiday, the first elements in these two columns are 2s. `t` is a column with `1,...,26 280`, such that all observations in the three years are marked as one time series, to be input to $f(t, \text{county})$. Finally, `year` marks what year the observations are made, and we use it to specify what year we want to predict.

When we want to make predictions for 2011, we replace the observations `y` in 2011 by `NA`s. This way, the model learns from 2009 and 2010 and uses the information available about 2011 in the data frame, which are all columns except `y`.

```
data$y[which(data$year==2011)] = NA
```

B.2 Model Specification and Prediction

Next, we specify the formula. Recall that Model W for one county is defined as

$$\eta_t = \boldsymbol{x}^\top \boldsymbol{\beta} + f(\text{temp}_t) + f(\text{day}_t) + f(\text{hour}_t, \text{daytype}_t) + f(\text{wind}_t) + f(t).$$

Transferring this to the R-INLA environment yields

```
formula = y ~ f(temp, model="rw2", constr=T) +
    f(day, model="rw2", cyclic=T, constr=T) +
    f(t, model="ar1") +
    f(hour, model="rw2", cyclic=T, constr=T, replicate=daytype) +
    f(wind, model="rw2", constr=T) +
    as.factor(daytype2)
```

where `constr=T` in the RW2 models specifies that we impose the constraint of sum-to-zero, discussed in Section 3.2.1. `cyclic=T` denotes a circular RW2, which we explained in Section 4.1.

When `formula` is specified, we can make inference and predictions using the `inla()` function to estimate the model parameters:

```
model = inla(formula, data=data, verbose=TRUE,
    control.predictor=list(compute=T),
    control.family=list(hyper=list(theta=list(fixed=TRUE))),
    control.compute=list(dic=TRUE), inla.call="remote")
```

The first line in the preceding code chunk specifies that we want to estimate our model using `formula` defined above, and that the covariates and response in `formula` are found in the data frame `data`. Adding `verbose=TRUE` allows us to see the computations in the console as `inla()`. Further, the argument in the second line specifies that we want to compute the posterior distribution of the linear predictor. In the third line we fix the precision parameter of the likelihood to a high value, as discussed in Section 4.1. In the bottom line we specify that we want the DIC returned, and the last argument tells R-INLA to run `inla()` on a remote server that we have a connection to.

Retrieving the predictions is simple. They are stored in the object `model`, and the fitted values are retrieved by typing `model$summary.fitted.values`:

```
head(model$summary.fitted.values)

##                              mean         sd 0.025quant  0.5quant
## fitted.Predictor.00001 -4.036594 0.05311429  -4.140715 -4.036591
## fitted.Predictor.00002 -4.057386 0.04984726  -4.155099 -4.057385
## fitted.Predictor.00003 -4.034754 0.04734837  -4.127568 -4.034753
## fitted.Predictor.00004 -4.006229 0.04557057  -4.095555 -4.006230
## fitted.Predictor.00005 -3.912899 0.04440196  -3.999940 -3.912897
## fitted.Predictor.00006 -3.928386 0.04345217  -4.013557 -3.928388
##                        0.975quant      mode
## fitted.Predictor.00001  -3.932455 -4.036585
## fitted.Predictor.00002  -3.959646 -4.057383
## fitted.Predictor.00003  -3.941914 -4.034752
## fitted.Predictor.00004  -3.916869 -4.006232
## fitted.Predictor.00005  -3.825841 -3.912892
## fitted.Predictor.00006  -3.843176 -3.928392
```

In our case, as we model the log transform of the heat consumption, we will use the median, `0.5quant`, as the predicted value instead of the mean. The reason for this was explained in Section 4.2.1.

We observe that `mean` and `0.5quant` are very similar. This is because of the assumption of Gaussian distributed log transformed heat consumption. The uncertainty measures of each prediction are located in the columns `0.025quant` and `0.975quant`, constituting the 95% prediction interval.

To get the predictions for the year we left out, 2011, we use the column `year` in `data`:

```
head(model$summary.fitted.values[which(data$year==2011),])

##                              mean         sd 0.025quant  0.5quant
## fitted.Predictor.17521 -4.529538 0.05854770  -4.644297 -4.529541
## fitted.Predictor.17522 -4.583514 0.06302053  -4.707038 -4.583520
## fitted.Predictor.17523 -4.557791 0.06889907  -4.692842 -4.557794
## fitted.Predictor.17524 -4.511654 0.06970234  -4.648268 -4.511662
## fitted.Predictor.17525 -4.422401 0.07254563  -4.564601 -4.422402
## fitted.Predictor.17526 -4.402019 0.07572685  -4.550445 -4.402027
##                        0.975quant      mode
## fitted.Predictor.17521  -4.414723 -4.529548
## fitted.Predictor.17522  -4.459924 -4.583529
## fitted.Predictor.17523  -4.422681 -4.557798
## fitted.Predictor.17524  -4.374948 -4.511679
## fitted.Predictor.17525  -4.280142 -4.422405
## fitted.Predictor.17526  -4.253503 -4.402041
```

The smooth effects are located in `model$summary.random`. We show the first six rows of the effect of wind as example, where the column `ID` refers to the wind speed:

```
head(model$summary.random$wind)

##    ID        mean          sd  0.025quant    0.5quant   0.975quant
## 1 0.0 -0.02438979 0.007858127 -0.03982639 -0.02438760 -0.008978529
## 2 0.1 -0.02363391 0.007542211 -0.03844557 -0.02363323 -0.008838965
## 3 0.2 -0.02288779 0.007307877 -0.03723603 -0.02288822 -0.008550036
## 4 0.3 -0.02216240 0.007138756 -0.03617560 -0.02216384 -0.008154043
## 5 0.4 -0.02146658 0.007018899 -0.03524132 -0.02146910 -0.007690727
## 6 0.5 -0.02081973 0.006934298 -0.03442508 -0.02082341 -0.007207018
##          mode          kld
## 1 -0.02438210 4.089109e-06
## 2 -0.02363096 2.783882e-06
## 3 -0.02288824 1.671624e-06
## 4 -0.02216591 8.530127e-07
## 5 -0.02147335 3.723921e-07
## 6 -0.02082993 2.216462e-07
```

The fixed effects are located in `model$summary.fixed`:

```
model$summary.fixed

##                             mean         sd 0.025quant   0.5quant
## (Intercept)          -4.9177180 0.01433494 -4.9459685 -4.9176881
## as.factor(daytype2)2 -0.2226348 0.00557459 -0.2335793 -0.2226351
##                      0.975quant       mode          kld
## (Intercept)          -4.8896655 -4.9176365 1.777096e-06
## as.factor(daytype2)2 -0.2116982 -0.2226352 1.056984e-07
```

And the marginal distribution for each of the fixed effects is available in `model$marginals.fixed`:

```
head(model$marginals.fixed$`(Intercept)`)

##              x            y
## [1,] -5.061258 5.186572e-17
## [2,] -5.032551 9.078260e-11
## [3,] -5.003845 3.693528e-06
## [4,] -4.989492 2.787650e-04
## [5,] -4.975139 1.279472e-02
## [6,] -4.960786 3.263960e-01
```

```
head(model$marginals.fixed$`as.factor(daytype2)2`)

##                x            y
## [1,] -0.2783979 1.675234e-20
## [2,] -0.2672451 9.699726e-13
## [3,] -0.2560924 1.101437e-06
## [4,] -0.2505161 2.669355e-04
## [5,] -0.2449397 2.395037e-02
## [6,] -0.2393633 7.931067e-01
```

If we wish to plot these marginal distributions, for example for the intercept, we can use the function `inla.smarginal()` to achieve more coordinates for a smoother curve:

```
x = inla.smarginal(model$marginals.fixed$`(Intercept)`)$x
y = inla.smarginal(model$marginals.fixed$`(Intercept)`)$y
```

where `x` is the vector of x-coordinates of the distribution, and `y` is the vector of y-coordinates.

Since we added the argument `control.compute=list(dic=TRUE)` in `inla()`, we can get the value of DIC from the model estimation:

```
model$dic$dic

## [1] -27982.93
```

Note that the DIC is more interesting to consider when we estimate a model on data where no observations are left out. We still include it in this example in order to demonstrate how the value is retrieved.