TU Dortmund
Faculty of Statistics

technische universität
dortmund

Master Thesis

Prediction of heat pump power consumption using a state space model

Author: Ortrud Wartlick
(Student ID: 192323)

Supervisor: Prof. Dr. Christine Müller

November 12, 2023

Eidesstattliche Versicherung

(Affidavit)

Wartlick, Ortrud Berta
Name, Vorname
(surname, first name)

192323
Matrikelnummer
(student ID number)

☐ Bachelorarbeit
(Bachelor's thesis)

■ Masterarbeit
(Master's thesis)

Titel
(Title)

Prediction of heat pump power consumption using a state space model

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

Dortmund, den 12.11.2023
Ort, Datum
(place, date)

O. Wartlick
Unterschrift
(signature)

Belehrung:
Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin") zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:
Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:*

Dortmund, den 12.11.2023
Ort, Datum
(place, date)

O. Wartlick
Unterschrift
(signature)

***Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.**

Contents

1 Introduction

In a world becoming more aware of climate change, heat pumps are an obvious sustainable heating solution when powered by electricity from renewable energy sources (Die Zeit 2022). In this master thesis, we analyse the electrical power consumption of heat pumps in 38 single-family households in Hamelin, Germany, in the years 2019 and 2020 (Schlemminger et al. 2002). The aim of this project is to predict electrical power consumption based on external variables, such as time of year, time of day, temperature and daylight hours, as well as internal variables, such as household-specific features and typical heat pump usage patterns.

In the first part of this thesis, the dataset is discussed (chapter 3) and potential predictor variables for heat pump power consumption are created (chapter 5). A descriptive analysis of heat pump power consumption data shows that like temperature, the seasonal profile of power consumption is dominated by seasonalities with periods of 1 year and 1 day. Furthermore, pumps can be sorted into distinct clusters based on their mean daily load profiles (usage patterns). Finally, when yearly and daily seasonalities are removed from the data, autoregressive components are visible in the noise. Based on this descriptive analysis, 145 potential predictor variables and interactions are then created that may affect heat pump power consumption (chapter 5).

The second part of this thesis is concerned with modeling. First, static linear models are used to answer some questions about fitting the data: How many pumps should be fit simultaneously? Should pumps be fit in the clusters defined by their usage patterns? Which of the 145 potential predictors are most relevant for modeling? Chapter 6 addresses these questions using linear model "experiments" in combination with forward selection. As a result of this analysis, we decide that 17 pumps can be fit simultaneously and fitting these pumps in smaller, separate clusters is not necessary. Of the 145 predictor variables and interactions constructed in chapter 5, only 48 are relevant - the others are never selected during forward selection. Furthermore, only 4 selection rounds are needed for a model to describe power consumption data comprehensively (chapter 6).

In the final chapter (chapter 7), four different state space model are formulated and estimated using the Kalman filter (Petris et al. 2009). The first state space model discussed is the "V-model", which assumes that the true feature effects are constant and is therefore similar to the static linear model in chapter 6. The second model is the "W-model", which extends the V-model by assuming that the true feature effects have a variance. The third model is the "P-model", which uses a particle filter for estimation instead of the Kalman filter. Finally, the "A-model" extends the W-model by adding AR(2) noise. The four models are optimised and the mean prediction errors of the 17 included pumps are compared. The results show that the W-model performs best both for short-term (1 hour) and longer-term (24-hour) predictions of heat pump power consumption data. However, interesting points are raised in the analysis which suggest that W- and A-models could be improved by further modification of the model structure and / or the estimation conditions (chapter 7).

2 Aims

This project aims to address the following questions:

(1) Which variables may influence the electrical power consumption of heat pumps?
(2) Can heat pump power consumption be predicted using a state space model?

Question (1) is discussed in chapters 5 and 6 and question (2) is discussed in chapter 7.

3 Dataset

The dataset is described by Schlemminger et al. 2002. Here some information is summarised.

3.1 Description

The dataset consists of electrical power consumption data (in Watt), measured in 38 separate single-family households (SFH) near Hamelin in Lower Saxony, Germany. The buildings included in the dataset are numbered from 3 to 40. All buildings in the dataset are equipped with a water-water-heat pump for underfloor space heating, a 6 kW heating rod as a backup heater to deliver additional space heating in case the heat pump is not sufficient, and solar thermal panels to heat water for domestic use (Schlemminger et al. 2002). If not enough solar energy is available (mainly in winter), the heat pumps are also used to heat domestic water. Four buildings (13, 15, 26 and 35) are also equipped with a photovoltaic system (PV) to produce electricity. Three households have a ventilation system installed (5, 9 and 16; Schlemminger et al. 2002).

Power consumption data is measured with direct-connected electricity meters in each household. Separate electricity meters are used to measure the household and the heat pump power consumption. For the households with a ventilation system, its power consumption is included in the heat pump measurement, as is the use of the backup heater (Schlemminger et al. 2002). Measurements are read in intervals of 10 seconds and transmitted to a database server using the building's internet connection. The raw power consumption data is then validated for internal consistence between power data and also measured energy data and aggregated from 10 second to 1 minute, 15 minute and 60 minute timeseries data, using the arithmetic mean to aggregate power consumption data. The data is stored in HDF5 format on the Zenodo platform (Schlemminger et al. 2002; Zenodo Platform 2022). Some of the data is also available in CSV format on the OpenMeter platform but this is not used here (Schlemminger et al. 2002; Logarithmo GmbH und Co. KG 2022).

Table 1 shows the names of the files downloaded from the Zenodo platform for the present thesis. Because of technical limitations (RAM), only 60 minute timeseries data is used. Power consumption data is available from May 2018 to December 2020, and Schlemminger et al. 2002 also provide weather data from WetterOnline for the same time period (WetterOnline 2021).

File name
2018_data_60min.hdf5
2019_data_60min.hdf5
2020_data_60min.hdf5
2018_weather.hdf5
2019_weather.hdf5
2020_weather.hdf5

Table 1: Files downloaded from Zenodo.

HDF5 files are hierarchical in structure, ordering datasets in groups, similar to directories. The structure of the HDF5 files used here is shown in Table 2 (reproduced from Schlemminger et al. 2002, Table 3).

File type	Top-level nodes	Middle-level nodes	Low-level nodes
Power HDF5-files (e.g. 2018_data_60min.hdf5)	MISC	ES1	TRANSFORMER
		PV1	EAST
			SOUTH
			WEST
	NO_PV	SFH3	HOUSEHOLD
			heat pump
		...	HOUSEHOLD
			heat pump
		SFH40	HOUSEHOLD
			heat pump
	WITH_PV	...	...
Weather HDF5-files (e.g. 2018_weather.hdf5)	WEATHER_SERVICE	IN	..TEMPERATURE..
			..WIND_SPEED..
			...

Table 2: HDF5 file structure of power files and weather files.

Variable	**Type**	**Based on variable in file**
PV	binary (0: "NO_PV", 1: "WITH_PV")	power_file.keys() ($\in$ {"NO_PV","WITH_PV","MISC"})
id	string ("SFH{n}"; $n \in \mathbb{N}^+$)	power_file[{PV}].keys()
timestamp	datetime	power_file[{PV}/{id}/HOUSEHOLD].index (in seconds since 1970)
	(YYYY-mm-dd HH:MM:SS)	power_file[{PV}/{id}/heat pump].index (in seconds since 1970)
household	numerical (in Watt)	power_file[{PV}/{id}/HOUSEHOLD].P_TOT
pump	numerical (in Watt)	power_file[{PV}/{id}/heat pump].P_TOT
number_of_persons	non-negative integer (absolute count)	power_file[{PV}/{id}/HOUSEHOLD].attrs["n_inhabitants"]
square_metres	numerical (in m^2)	power_file[{PV}/{id}/HOUSEHOLD].attrs["living_space"]
temperature	numerical (in °C)	weather_file[WEATHER_SERVICE/IN/ WEATHER_TEMPERATURE_TOTAL].TEMPERATURE_TOTAL

Table 3: Variables extracted from HDF5 files.

The variables extracted from the HDF5 files are shown in Table 3. Extraction of variables is based on a Python script provided by Jawana Gabrielski. The extracted variables include the household id (SFH3 to SFH40), whether or not the household produces electricity using a photovoltaic system (PV; the "MISC"-property is not used), the floor size of the building in square metres (ranging from 87 to 230m^2) and the number of persons living in the building (ranging from 1 to 4 persons). The household and pump power consumptions are given in Watt in 60 minute intervals, and a timestamp is provided in seconds since 1970 (UNIX timestamp). The timestamp variable is converted to a datetime variable using Python's `pandas` package

(The pandas development team 2020).

Schlemminger et al. 2002 provide a rough validation of household power consumption data, excluding households which are missing more than one month of data in 2019 (households 6, 13, 17, 24, 25, 31, 37 and 40; see also section 3.2). The remaining households have an average of 2.38 inhabitants. The German Federal Statistics Office provides a benchmark mean total power consumption of 3221 kWh for 2 person-households and 4978 kWh for (on average) 3.65 persons per year (German Federal Statistics Office 2021). Linear interpolation suggests that the expected annual household power consumption for the households in this dataset should therefore be 3625 kWh for 2.38 persons per year (Schlemminger et al. 2002). Excluding the electricity commonly used for space heating and hot water (17.9%; Frondel et al. 2020), which are here provided instead by heat pumps and solar thermal panels, results in an expected annual household power consumption of 2976 kWh. The observed mean household power consumption in the dataset is 2829 kWh in 2019, which is consistent with the expectation and thus broadly validates the measurements. The median of the annual household consumption is 2996 kWh, the minimum is 884 kWh (SFH15; a household with PV) and the maximum is 5489 kWh (SFH10).

The mean heat pump power consumption in the dataset is 4993 kWh in 2019, with a median of 4012 kWh, minimum of 1431 kWh (SFH35; one of the smallest with $100m^2$) and maximum of 14840 kWh (SFH20; one of the largest with $220m^2$), which apparently is a bit less than expected, probably because domestic hot water is heated by solar energy when available, and not by the heat pump (Schlemminger et al. 2002). Schlemminger et al. 2002 warn that the households with the highest heat pump power consumption (4, 20, 22, 34, 40) use the backup heating rod extensively (from 50% to 88% of power consumption measured as heat pump power consumption is actually due to the backup heater), whereas 12 households only use it rarely (15% of pump power consumption; Schlemminger et al. 2002). Unfortunately, use of the backup heater is not reported as a variable. However, this *caveat* should be kept in mind because use of the backup heater may well affect the heat pump power consumption profiles which are studied later in this thesis (chapter 5ff).

For the present thesis, many new variables are created from the existing variables described in Table 3. For coherence, existing variables are discussed in more detail later, together with new variables in chapter 5.

3.2 Cleaning

There is some missing data in the dataset: Schlemminger et al. 2002 explain that they achieve a data coverage of more than 90% for 23 of the 38 households. Small data gaps can be due to technical failures of the data logger or a long disconnection from the internet so that measured data cannot be transferred and is eventually lost. For two households (24 and 25), more than 50% of the data is missing because measurement was halted by the building owners. Six further households are missing more than one month of data (6, 13, 17, 31, 37 and 40), with no reason given. Schlemminger et al. 2002 correct small data gaps of up to one day by linear interpolation. Unfortunately, there is no information about which timestamps have

been corrected this way. However, larger data gaps are clearly marked as missing (NA) for each affected timestamp.

As will be discussed later, small data gaps are not a huge problem for the analyses in this thesis because the affected time points can be excluded from the analysis (chapter 6) or "bridged" using the mean power consumption of other pumps at the same time points (chapter 7). However, outliers - unusually large peaks and troughs in power consumption - can be a problem when modeling the data, because these outliers may not accurately reflect the dependence of power consumption on external factors. For this reason, some data cleaning procedures are performed on the dataset before moving on to the analysis. The cleaning procedures are illustrated in Fig. 1, using household 3 as an example. This is a household with very little missing data (Fig. 1A), which makes it easier to illustrate outliers.

We consider two types of outliers: The first are unusually large peaks in the heat pump data in some households, which we think reflect a programmed, temporary burst of heating power to kill *Legionella* bacteria in the system (mostly relevant in wintertime, when the pumps are also used to heat water; Rosenkranz, A. 2021). These types of outliers can be very regular (Fig. 1B), but for some pumps they only appear in the winter months or not at all. This irregularity necessitates an automatic approach to their detection and removal. The second type of outliers are unusal troughs in the household power consumption data (Fig. 1C). We assume that these are due to building owners being on holiday, so that the regular household power consumption is reduced to standby levels. Because the timing of holidays is unknown, we also use an automatic approach to detect and remove these.

The procedure to detect unusual peaks in heat pump power consumption data was already developed in a Fallstudien 2 seminar together with Bianca Strauss and Michael Schweitzer. The procedure is as follows: First, using a centered, rolling window of 13 days width, a rolling standard deviation of power consumption is calculated for each time point (R Core Team 2021). At the beginning and end of the timeseries, a window of decreasing bandwidth is used. If the window contains missing values, these are ignored; if the window contains *only* missing values, then the value of the standard deviation is set to 0 for the time point. The rolling standard deviation is usually larger overall in the winter months than in the summer months, when the pump is mostly on standby. To "detrend" the rolling standard deviation, a rolling median of the rolling standard deviation is calculated using a centered window of 7 days width. The rolling median is removed from the rolling standard deviation to yield a "detrended rolling standard deviation". This detrended standard deviation is then thresholded to detect outliers. The (empirical) threshold used is 300 Watt, which has been found to work well for most pumps (Fig. 1B). time points with a detrended standard deviation greater than this threshold are marked as "outliers" (Fig. 1B).

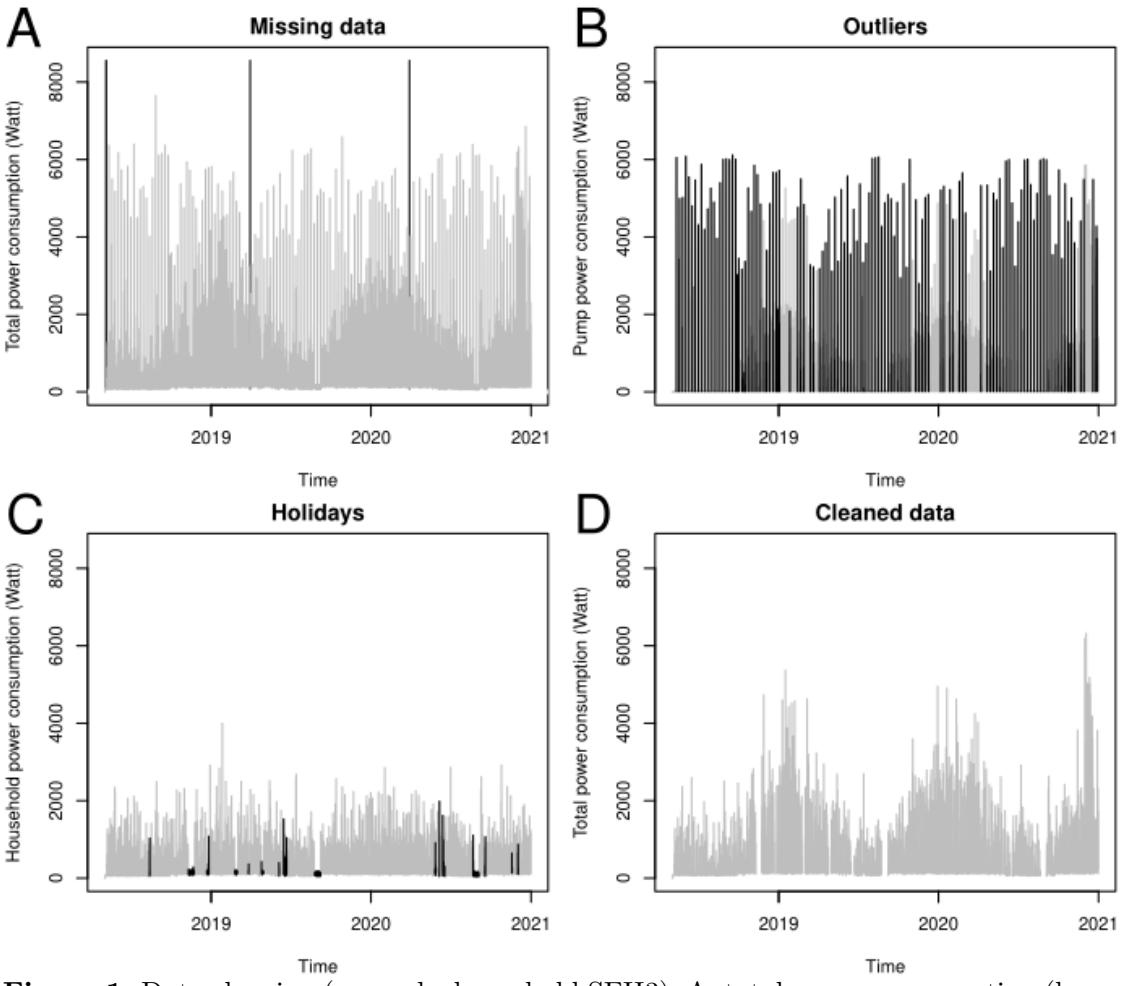


Figure 1: Data cleaning (example: household SFH3). A, total power consumption (household + pump) with missing values in black; B, pump power consumption with detected outliers in black; C, household power consumption with detected holidays in black; D, cleaned total power consumption data after removal of outliers and holidays.

The procedure to detect holidays in household pump power consumption data is as follows: First, the day-night difference in household power consumption is calculated. For this purpose, day-time is defined as between 8:00 and 19:59 hours of a given day, and night-time is defined as between 20:00 on the same day until 7:59 on the next day. Thus the data is subdivided into alternating day-time and night-time blocks, and the mean household power consumption is calculated for each block (R Core Team 2021). Missing data is ignored for this calculation. If the difference between subsequent mean day-time and night-time power consumptions is smaller than 30 Watt and the absolute household power consumption for the affected time points is smaller than 300 Watt, then these time points are marked as a potential holiday, using a new, temporary variable X, which is 1 for potential holidays and 0 otherwise. For time points with missing data, X is set to 0 by default. This method detects a lot of isolated potential holidays, which are unlikely to be real holidays. We assume that a real holiday is at least three days long. We therefore calculate a moving median of X using a centered rolling window of five days width. The value of the moving median will be 1 if at least three in five consecutive days are marked

as potential holidays. For example, a sequence of 000100011110101110000010 will become 000000011111111111000000, detecting a single holiday of ten days duration. This method is not perfect but it seems to detect holidays fairly reliably (Fig. 1C).

Before further analysis, the power consumption of time points with outliers or holidays are set to NA to ensure that they do not distort the analysis in chapters 5ff. Fig. 1D shows an example of a cleaned dataset (household 3), with outliers and holidays removed. Of course, this introduces more missing data into the dataset. Fig. 2 shows the percentage of missing data for each household in the dataset, after cleaning. The large data gaps in 2019 and / or 2020 for households 6, 8, 10, 11, 13, 15, 24, 25, 31, 35, 37 and 40 are due to missing data reported already in the original dataset (Schlemminger et al. 2002). Smaller data gaps of usually one week or less have been introduced by removing holidays. The removal of other outliers does not have a big effect because these only isolated time points.

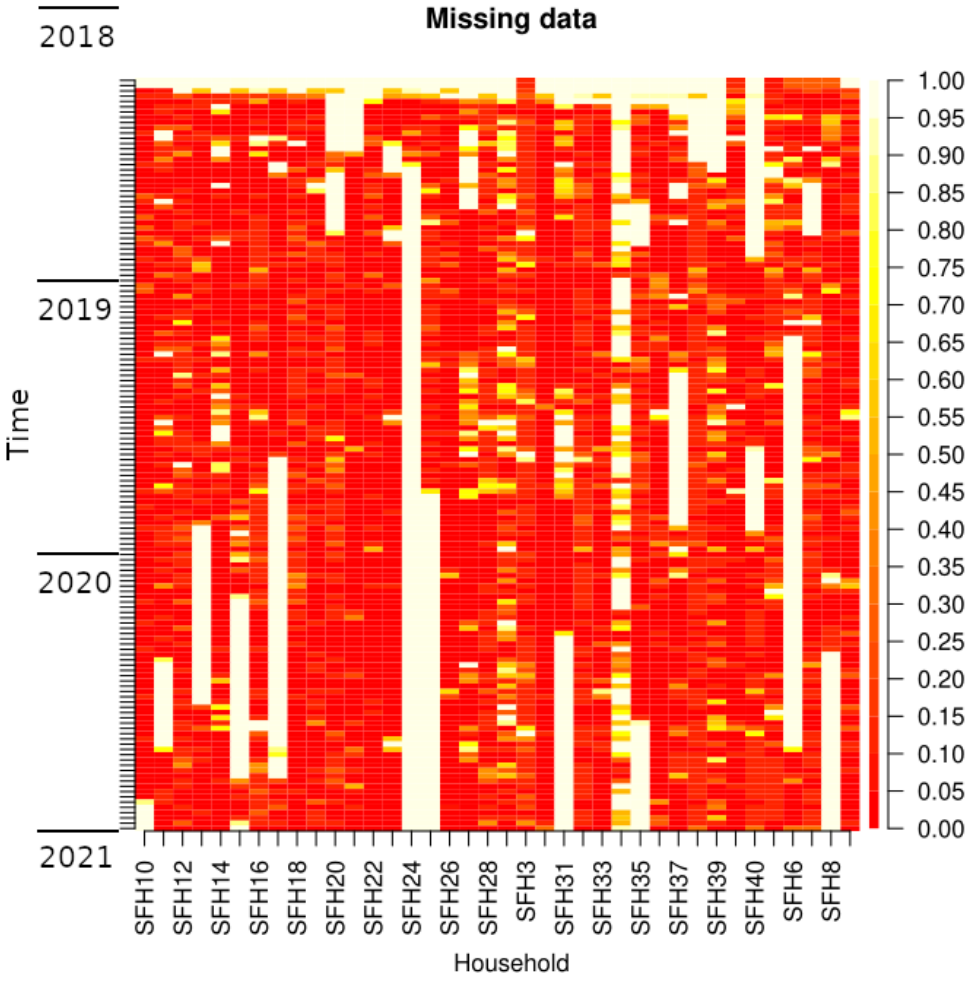


Figure 2: Fraction of missing data in pump power consumption for each household over time (aggregated per week). Light areas indicate large fractions of missing data. Note that the households are sorted alphabetically: SFH10-SFH30, SFH3, SFH31-SFH40, SFH4-SFH9.

Gaps in the data are taken into account in later chapters: for ordinary least squares estimation (OLS; chapter 6), the time points with missing power consumption values are simply discarded. This is possible because OLS does not take into account the order of a timeseries. For Kalman filter and particle filter estimations (chapter 7), which do progress through time points in an ordered fashion, data gaps in the timeseries are "bridged" for each individual pump using the mean heat pump power consumption of all the other pumps at the affected time points to impute missing data. Even then, the imputation is only temporary: it is only used so that the Kalman and particle filters can proceed across the time points with missing data, but the affected time points are not included in the analysis of predictions and residuals for the affected pumps.

In principle, it is therefore possible to include all households in the analyses used in this work. However, we do exclude some households (Table 4), either because of very large blocks of missing data - then the household does not contribute much - or for other reasons. For example, households 13, 15, 26 and 33 use a photovoltaic system to produce electricity. This naturally affects their power consumption profiles - the power consumption can even become negative when more electricity is produced than is consumed. These households are therefore excluded from all further analysis. In addition, household 24 has no data at all in 2019 and 2020 and is excluded. This leaves 33 households which can in principle be included in the analysis. Of these, households 6, 17, 25, 31, 34, 35 and 37 have very large data gaps. Households 5, 8, 11, 14, 15, 20, 22, 28, 34, 39 and 40 have unusually large or oddly shaped heat pump power consumption profiles. In some cases (20, 22, 34 and 40), these atypical profiles may be due to extensive use of the backup heater which can distort the heat pump measurements (see section 3.1), but in other cases there is no known explanation. Therefore, we do not exclude these households outright. In chapter 6 we experiment with including or excluding "atypical" pumps to test the effect these may have on the analysis.

Households	Excluded / Included?
13, 15, 26 and 33	Excluded from most further analysis because of PV
24	Excluded from further analysis because of no data in 2019-2020
6, 17, 25, 31, 34, 35 and 37	Sometimes excluded because of missing data in 2019-2020
20, 22, 34 and 40	Sometimes excluded because of extensive use of the backup heater
5, 8, 11, 14, 15, 28 and 39	Sometimes excluded because of atypical pump profiles
3-12, 14, 16-25, 27-32, 34-40	Included in some analysis (chapter 6; m = 33)
3, 4, 7, 9, 10, 12, 16, 18, 19, 21, 23, 27, 29, 30, 32, 36, 38	Included in all further analysis (chapter 6f; m=17)

Table 4: Excluded and included households; m: number of households in the group.

Finally, note that later in this thesis, only data from 2019 and 2020 is used - data from 2018 is discarded because it is not complete (Fig. 2). Data from 2019 is used for feature generation (chapter 5) and the year 2019 is also used as a training or adaptation period for different modeling approaches (chapter 6f). Data from 2020 is used for the analysis of predictions and residuals (chapter 6f).

4 Methods

4.1 Spectral analysis

Every time series $(x_t)_{t\in\mathbb{N}}$ can be represented by a Fourier series: a sum of sine and cosine waves with varying amplitudes and frequencies (or periods). In exponential form, this can be expressed as (Shumway and Stoffer 2017, p.180, Eq. 4.27):

$$x_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j)\ \exp(2\pi i w_j t). \tag{4.1}$$

Here, n is the number of time points, $\omega_j = j/n$ are the Fourier frequencies and $d(\omega_j)$ is related to the amplitude of the oscillation with frequency ω_j (Shumway and Stoffer 2017, p.180, Eq. 4.27; other definitions are available but this is the one used here). The Fourier periods are given by $l_j = 1/\omega_j$ for $j > 0$, with $l_0 := 0$.

Spectral analysis makes use of the discrete Fourier transform to determine the dominant periods of seasonalities in the data, i.e. those periods with the largest amplitude. In Shumway and Stoffer 2017, the discrete Fourier transform (DFT) is defined as:

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t \exp(-2\pi i w_j t). \tag{4.2}$$

In `R`, the DFT can be calculated using `stats::fft(x)/sqrt(length(x))` (R Core Team 2021).

In Shumway and Stoffer 2017, the "periodogram" or "sample spectral density" is defined as:

$$I(\omega_j) = |d(\omega_j)|^2 \tag{4.3}$$

for each ω_j (Shumway and Stoffer 2017, p.180, Eq. 4.28). The dominant periods are those periods with the largest $I(\omega_j)$. These correspond to the Fourier oscillations with the largest amplitudes which explain most of the variability in the data. The plots in Fig. 3 are created by plotting $I(\omega_j)$ vs. l_j. Periods are plotted instead of (the more usual) frequencies so that the dominant periods can be directly visually identified.

4.2 Sample autocorrelation function (ACF)

The autocovariance of a stationary time series $(X_t)_{t\in\mathbb{N}}$ is defined as

$$\gamma(h) = \mathrm{Cov}(X_{t+h}, X_t), \tag{4.4}$$

where $h \geq 0$ is the "lag" between the two time points (Shumway and Stoffer 2017, p. 21, Eq. 1.23). The autocorrelation function is then given by

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \tag{4.5}$$

(Shumway and Stoffer 2017, p. 21, Eq. 1.24). The autocovariance can be estimated using observations $x_t, t = 1, \ldots, n$, and

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \overline{x})(x_t - \overline{x}), \tag{4.6}$$

with $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$ (Shumway and Stoffer 2017, p. 27, Eq. 1.36). The sample autocorrelation function (ACF) can then be calculated using

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} \tag{4.7}$$

(Shumway and Stoffer 2017, p. 28, Eq. 1.37). Note that $\widehat{\rho}(h = 0) = 1$ by definition.

If the time series $(X_t)_t$ is white noise, then for large n, the sample ACF is approximately normally distributed with mean zero and standard deviation $\sigma_{\widehat{\rho}(h)} = 1/\sqrt{n}$ (Shumway and Stoffer 2017, p. 28. Eq. 1.38). Thus, if $\widehat{\rho}(h) \notin [\pm 2/\sqrt{n}]$ for $h > 1$ then $\widehat{\rho}(h)$ is significantly different from 0 and $(X_t)_t$ is not white noise (Shumway and Stoffer 2017, p.29). In the ACF plots in chapters 6 and 7, the interval $[\pm 2/\sqrt{n}]$ is indicated by dashed blue lines. We use `stats::acf` to calculate and plot the sample ACF (R Core Team 2021).

Since time series models are not the major focus of this thesis they are not discussed here in detail. However, note that the ACF of an autoregressive stationary process of any order decays exponentially as $h \to \infty$ (Shumway and Stoffer 2017, p. 95), whereas for an invertible moving average process of order q, $\rho(h) = 0$ for $h > q$ (Shumway and Stoffer 2017, p. 94, Eq. 3.43). For an ARMA process, $\rho(h)$ decays with increasing h (Shumway and Stoffer 2017, p. 99).

4.3 Sample partial autocorrelation function (PACF)

The partial autocorrelation function of a stationary time series $(X_t)_{t \in \mathbb{N}}$ is defined as

$$\phi(h) = \text{Corr}(X_{t+h} - \widehat{X}_{t+h}, X_t - \widehat{X}_t) \tag{4.8}$$

where $h \geq 1$ is the "lag" between the two time points (Shumway and Stoffer 2017, p. 97, Eq. 1.23). The partial autocorrelation measures the correlation between random variables X_{t+h} and X_t with the effects of $X_{t+1}, \ldots, X_{t+h-1}$ removed by linear regression of X_{t+h} and X_t on $X_{t+1}, \ldots, X_{t+h-1}$ yielding $\widehat{X}_{t+h}$ and $\widehat{X}_t$ (Shumway and Stoffer 2017, p. 97).

For $h = 1$, there are no elements between X_{t+h} and X_t and therefore the partial autocorrelation is just the autocorrelation: $\phi(h = 1) = \rho(h = 1)$ (see also section 4.2). For $h \geq 2$, the sample partial autocorrelation function (PACF) can be estimated using observations $x_t, t = 1, \ldots, n$, for linear regressions,

$$\begin{aligned} \widehat{x}_{t+h} &= \textstyle\sum_{k=1}^{h-1} \beta_k x_{t+h-k} \\ \widehat{x}_t &= \textstyle\sum_{k=1}^{h-1} \beta_k x_{t+k} \end{aligned} \tag{4.9}$$

and then calculating $\widehat{\phi}(h) = \text{Corr}(x_{t+h} - \widehat{x}_{t+h}, x_t - \widehat{x}_t)$ (Shumway and Stoffer 2017, p. 97). Note that, because of stationarity, $\mathbb{E}(X_t) = 0$, so no intercepts are needed in Eq. 4.9. Also because of stationarity, the coefficients $\beta_1, \ldots, \beta_{h-1}$ are the same in both regressions (Shumway and Stoffer 2017, p. 97). In practice, we use `stats::pacf` to calculate and plot the sample PACF (R Core Team 2021).

Again, time series models are not discussed here since they are not the major focus of this thesis. However, note that for an autoregressive stationary process of order p (AR(p)), $\phi(h) = 0$ for $h > p$ (Shumway and Stoffer 2017, p. 98). For an invertible moving average process of any order, the PACF decays as $h \to \infty$ (Shumway and Stoffer 2017, p. 99). For an ARMA process, $\phi(h)$ also decays as h increases (Shumway and Stoffer 2017, p. 99).

4.4 Agglomerative hierarchical clustering

Given a finite set of objects $X = \{x_1, \ldots, x_n\}$ with k dimensions ($x_i = (x_{i1}, \ldots, x_{ik})^T$), a clustering method partitions this set into disjoint subsets (or clusters) based on a similarity or dissimilarity measure, so that the elements within a cluster are similar to each other, and elements in different clusters are dissimilar (Miyamoto 2022, p. 1f). In this thesis, agglomerative hierarchical clustering is used to cluster households by the mean centered load profiles of their pumps (see chapter 5). In this case, $n = 38$ households, x_i is the mean centered daily load profile of pump i, and $k = 24$ hours.

For agglomerative hierarchical clustering, each observation i is initially assigned to its own cluster $C_i = \{x_i\}$ (Miyamoto 2022, p. 3). Then the pair-wise similarity S or dissimilarity D between clusters is calculated. The clusters with the largest similarity or smallest dissimilarity are merged into one cluster. This process is repeated until there is one large cluster that contains all objects (Miyamoto 2022, p. 3). The method is agglomerative because a bottom-up approach is used, starting with clusters of one object each and forming increasingly bigger clusters, and hierarchical because the order in which clusters are merged is recorded and yields a hierarchy of clusters that can (for example) be displayed by a dendrogram (such as Fig. 5 in chapter 5).

Here, the Ward method is used to determine the dissimilarity between clusters (Miyamoto 2022, p. 24). For the Ward method, we first define the squared error within cluster C as

$$E(C) = \sum_{x_i \in C} ||x_i - \overline{x}_C||_2^2 \tag{4.10}$$

where $\overline{x}_C$ is the mean of the objects in cluster C (Miyamoto 2022, p. 24, Eq. 2.22). The increase in the error when clusters C and C' are merged is

$$\Delta E(C, C') = E(C \cup C') - E(C) - E(C') \tag{4.11}$$

(Miyamoto 2022, p. 24, Eq. 2.23). The increase in the error is used as a dissimilarity measure, $D(C, C') = \Delta E(C, C')$. It is large when dissimilar clusters are merged, and small when similar clusters are merged. For the initial clusters consisting of one

object, the dissimilarity between two clusters $C = \{x\}$ and $C' = \{y\}$ is given by

$$D(C, C') = \Delta E(\{x\}, \{y\}) = \left|\left|x - \frac{x+y}{2}\right|\right|_2^2 + \left|\left|y - \frac{x+y}{2}\right|\right|_2^2 = ||x-y||_2^2, \quad (4.12)$$

i.e. the Euclidean distance (Miyamoto 2022, p. 24, Eq. 2.25). As clusters are merged during subsequent clustering steps, the dissimilarity calculation between clusters is updated using an algorithm based on Eq. 4.11 (Miyamoto 2022, p. 24-26, Eq. 2.26). In R, the Ward method is implemented by the `stats::hclust(..., method="ward.D")` function (R Core Team 2021).

Dendrogram

A dendrogram is used to visualise the hierarchical cluster (e.g. Fig. 5). A dendrogram displays the cluster as a tree whose endpoints (or leaves) are the individual objects. The merging of two clusters is displayed as a merge of the cluster branches (Miyamoto 2022, p. 11). Smaller clusters can be extracted from the big cluster by cutting branches at a certain distance from the leaves.

4.5 Paired t-test

In chapter 6, we compare a number of measures M in different conditions C. For example, the mean squared error (MSE) or median absolute error (MAE) of a linear model fit are calculated for each of $m = 17$ pumps, first under the condition that these pumps are all fit simultaneously ($C_{\text{"cluster"}} = 0$) and then under the condition that they are fit in smaller clusters ($C_{\text{"cluster"}} = 1$). In general, the values of a measure M measured for the same objects under conditions $C = 0$ and $C = 1$ can be compared using a paired t-Test (Heumann et al. 2016, p. 225). Here, the one-sided hypothesis for these tests is given by:

$$H_0 : \mu_M \leq 0 \text{ vs. } H_1 : \mu_M > 0, \quad (4.13)$$

where $\mu_M = \mathbb{E}(M(C = 0) - M(C = 1)) = \mathbb{E}(D_M)$, i.e. the expected difference between measures in pre- and post-conditions. If condition $C = 1$ improves the fit and therefore decreases M, then $M(C = 0) > M(C = 1)$ and $\mu_M > 0$. We assume that the difference $D_M = M(C = 0) - M(C = 1)$ is a random variable that is approximately normally distributed at the boundary of the null hypothesis, i.e. $D_M \sim N(0, \sigma_D^2)$ (Heumann et al. 2016, p. 225). Then for a random sample of n i.i.d. differences $D_{M,1}, \ldots, D_{M,n}$, the test statistic

$$T(D_M) = \sqrt{n}\frac{\overline{D}_M}{\sqrt{\text{Var}(D_M)}} \quad (4.14)$$

is t-distributed with $n-1$ degrees of freedom (Heumann et al. 2016, p. 225). The null hypothesis is rejected if the p-value $p = P_{H_0}(T \geq t(d_M))$ of the realised test statistic $t(d_M)$ for observations $d_{M,1}, \ldots, d_{M,n}$ is < 0.05. In `R`, the one-sided paired t-test is performed using `stats::t.test(`$M \sim C$`, data, paired = TRUE, alternative = 'greater')` (R Core Team 2021).

A note of caution: In the tests performed in chapter 6, a condition C comprises four experiments, with measurements for the same 17 pumps in each experiment, so $n = 4m = 68$ (see Table 9). The four experiments per condition differ in other fit conditions and so the values of M are not expected to be the same for the same pump across experiments, but strictly speaking, the measurements are not independent since each pump appears four times. This affects the estimation of the variance and the test statistic. These effects are neglected here.

4.6 Classical linear model

The classical linear model is used in chapter 5 to fit mean daily power consumption as a function of mean daily temperature (see Eq. 5.3) and to fit the mean daily temperature over time (see Eq. 5.4). It is also used in chapter 6 to fit heat pump power consumption simultaneously for m pumps (see Eq. 6.1). Here, we discuss only the simple case with one intercept. The more complex case of simultaneous fits with several intercepts is described in detail in chapter 6.

The linear model is described by the following equation (Fahrmeir et al. 2013, p. 74):

$$Y = \alpha + x_1\beta_1 + \ldots + x_p\beta_p + \epsilon, \qquad \epsilon \sim N(0, \sigma^2), \tag{4.15}$$

where Y is a random variable and $x_1, \ldots, x_p$ are p different regressors or predictor variables. The error term ϵ is a random variable that is assumed to be i.i.d. normally distributed with $\mathbb{E}[\epsilon] = 0$ and homoscedastic variance σ^2. Under these conditions, Y is also normally distributed with $\mathbb{E}[Y] = \alpha + x^T\beta$ and variance σ^2. The parameter α describes an intercept while $\beta_1, \ldots, \beta_p$ describe the effects of the predictor variables $x_1, \ldots, x_p$.

Given n observations $(y_i, x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$, the unknown parameters α, $\beta_1, \ldots, \beta_p$ can be estimated by minimising the squared error, $\Sigma_{i=1}^n (y_i - \widehat{y}_i)^2$, of predictions $\widehat{y}_i$. This is also referred to as the "ordinary least squares"-method (OLS). To simplify the estimation, we first rewrite the model equation in matrix form (Fahrmeir et al. 2013, p. 75):

$$\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{4.16}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}. \tag{4.17}$$

If $\boldsymbol{F}$ has full column rank, the OLS estimator is given by (Fahrmeir et al. 2013, p. 109):

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\theta}\|_2^2 = (\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{y}. \tag{4.18}$$

Given the assumptions for ϵ (i.i.d normally distributed with expectation 0 and homoscedastic variance), the OLS estimator is also normally distributed, with

$$\mathbb{E}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta} \text{ and } \text{Cov}(\widehat{\boldsymbol{\theta}}) = \sigma^2(\boldsymbol{F}^T\boldsymbol{F})^{-1} \tag{4.19}$$

(Fahrmeir et al. 2013, p. 119). The variance σ^2 can be estimated using

$$\widehat{\sigma}^2 = \frac{1}{n-(p+1)} \left|\left| \boldsymbol{y} - \boldsymbol{F}\widehat{\boldsymbol{\theta}} \right|\right|_2^2 \tag{4.20}$$

(Fahrmeir et al. 2013, p. 109). For the applications of the linear model in chapter 5, the `R stats::lm`-function is used for OLS estimation (R Core Team 2021). For the application with multiple intercepts and an expanding training window described in chapter 6, custom functions are used based on Eq. 6.3. Time points with missing data are excluded in either case. The `R`-code is submitted with this thesis.

Note that, for convenience, bold face notation for vectors and matrices is dropped in the remainder of the thesis.

4.7 State space model

$$\begin{array}{ccccccccc} \theta_0 \to & \theta_1 \to & \ldots \to & \theta_{t-1} \to & \theta_t \to & \theta_{t+1} \to & \ldots \\ & \downarrow & & \downarrow & \downarrow & \downarrow & \\ & Y_1 & & Y_{t-1} & Y_t & Y_{t+1} & \end{array}$$

Formally, a state space model consists of an $\mathbb{R}^q$-valued latent time series $(\theta_t)_{t\in\mathbb{N}_0}$ and an $\mathbb{R}^m$-valued observable time series $(Y_t)_{t\in\mathbb{N}}$, where

(1) $(\theta_t)_t$ is a Markov chain and
(2) Y_t depends only on θ_t and
(3) the Y_t are independent given (θ_t)

(Petris et al. 2009, p. 40). Assumptions (1-3) imply that a state space model is completely specified by the initial distribution of θ_0 and the conditional distributions of $\theta_t|\theta_{t-1}$ and $Y_t|\theta_t, t \geq 1$ (Petris et al. 2009, p. 40).

Dynamic linear models are a subgroup of state space models that assume linear relationships between Y_t and θ_t and between θ_t and θ_{t-1}, and normal distributions for θ_0 and any error terms (Petris et al. 2009, p. 41). A dynamic linear model is thus given by

$$\theta_0 \sim N_q(c_0, C_0) \tag{4.21}$$

and two linear equations for $t \geq 1$:

$$\begin{array}{lll} Y_t &= F_t\theta_t + v_t & v_t \sim N_m(0_m, V_t) \\ \theta_t &= G_t\theta_{t-1} + w_t & w_t \sim N_q(0_q, W_t) \end{array} \tag{4.22}$$

(Petris et al. 2009, p. 41). The known matrices $F_t \in \mathbb{R}^{mxq}$ and $G_t \in \mathbb{R}^{qxq}$ describe the relationship between Y_t and θ_t and θ_t and θ_{t-1}, respectively. The sequences $(v_t)_{t\geq 1}$ and $(w_t)_{t\geq 1}$ are two independent sequences of independent Gaussian random vectors with expectation 0 and known variance matrices $V_t \in \mathbb{R}_+^{mxm}$ and $W_t \in \mathbb{R}_+^{qxq}$ (Petris

et al. 2009, p. 41). It is assumed that θ_0 is independent of v_t and w_t (Petris et al. 2009, p. 41).

Note that the linearity and normality assumptions imply that $\theta_t|\theta_{t-1} \sim N_q(G_t\theta_{t-1}, W_t)$ and $Y_t|\theta_t \sim N_m(F_t\theta_t, V_t)$ (Petris et al. 2009, p. 42). In fact, all joint, marginal and conditional distributions involved are Gaussian (Petris et al. 2009, p. 53). This later simplifies the estimation and forecasting process.

All state space models described in chapter 7 can be reduced to a form of Eq. 4.22, with time-invariant $G_t = G$, $V_t = V$ and $W_t = W$. The concrete specifications of F_t, G, V and W are described in chapter 7 for each model discussed. In the following sections, two approaches to estimating $(\theta_t)_t$ and forecasting $\mathbb{E}(Y_{t+k}|\theta_t), k \geq 1$ are described: the Kalman filter and the particle filter.

4.7.1 Kalman filter

The Kalman filter uses a Bayesian approach to estimate θ_t sequentially for $t \geq 1$ given observations $y_{1:t} = y_1, \ldots, y_t$, an initial distribution for θ_0, the linear state equations 4.22 and the normality assumption. Let $\theta_0 \sim N_q(c_0, C_0)$. Then for $t \geq 1$,

(1) $\theta_t|y_{1:t-1} \sim N_q(a_t, R_t)$
with $a_t = G_t c_{t-1}$ and $R_t = G_t C_{t-1} G_t^T + W_t$
(2) $Y_t|y_{1:t-1} \sim N_m(f_t, Q_t)$
with $f_t = F_t a_t$ and $Q_t = F_t R_t F_t^T + V_t$
(3) $\theta_t|y_{1:t} \sim N_q(c_t, C_t)$
with $c_t = a_t + R_t F_t^T Q_t^{-1}(y_t - f_t)$ and $C_t = R_t - R_t F_t^T Q_t^{-1} F_t R_t$

where the third step represents the filtering step (Petris et al. 2009, p. 53).

In practice, we choose $c_0 = 0_q$ and $C_0 = \text{diag}_q(10^7, \ldots, 10^7)$. If all elements of y_t are missing at time point t, then y_t is set to 0_m for that time point. If fewer than m elements are missing in y_t at time point t, then the missing elements are imputed by setting the missing values to the mean $\overline{y}_t$ of the measured values at that time point. Alternatively, one could set the missing value to the preceding value of the same pump, however, this is problematic when there are data gaps spanning a large time period. It is actually not strictly necessary to replace the missing values at all, see Petris et al. 2009, p. 59, but this is the approach taken here.

The distributions estimated by the Kalman filter can be used for forecasting as follows. Let $a_t := c_t$ and $R_t := C_t$ for $k = 0$. Then, for $k \geq 1$,

(1) $\theta_{t+k}|y_{1:t} \sim N_q(a_{t+k}, R_{t+k})$
with $a_{t+k} = G_{t+k} a_{t+k-1}$ and $R_{t+k} = G_{t+k} R_{t+k-1} G_{t+k}^T + W_{t+k}$
(2) $Y_{t+k}|y_{1:t} \sim N_m(f_{t+k}, Q_{t+k})$
with $f_{t+k} = F_{t+k} a_{t+k}$ and $Q_{t+k} = F_{t+k} R_{t+k} F_{t+k}^T + V_{t+k}$

(Petris et al. 2009, p. 70). So first, the next value of the state vector, θ_{t+k}, is estimated and then Y_{t+k} is predicted based on the θ_{t+k} estimate.

Note a special case: If the variances in $V_t = V$ are zero or close to zero, then the Kalman filter equations yield $Q_t \approx F_t R_t F_t^T$ and $c_t = a_t + R_t F_t^T Q_t^{-1}(y_t - f_t) \approx a_t +$

$R_t F_t^T (F_t^T)^{-1} R_t^{-1} F_t^{-1}(y_t - F_t a_t) = F_t^{-1} y_t$. This implies that $f_t = F_t a_t = F_t G_t c_{t-1} \approx F_t G_t F_{t-1}^{-1} y_{t-1} \approx y_{t-1}$ if $G_t \approx I_{qxq}$. Similarly, for k-step ahead predictions, $f_{t+k} \approx y_t$. In other words, the forecasts are equal to the last observed value. Thus, careful thought has to be given to specification of the variances in the model. For most models discussed in chapter 7, the variances in V are large (specifically, much larger than the variances in W), but for the final "A-model" discussed in chapter 7, V is subsumed in $\tilde{W}$ and $\tilde{V}$ is zero (see Eqs. 7.11f). This may cause a "shift" of the predictions with respect to the observations. This problem is also discussed in chapter 7.

Petris 2010 provide an `R`-package for kalman filtering (Petris 2010, `dlm`). However, to have more flexibility for model definitions and variance optimisation, custom `R`-functions are used to implement the Kalman filter in this thesis. The `R`-code is submitted with this thesis.

4.7.2 Particle filter

"Sequential Monte Carlo"-methods or "particle filters" approximate $\theta_t|y_{1:t}$ by a set of support points, or particles, $x_t^{(i)} \in \mathbb{R}^q, i = 1, \ldots, N$:

$$\widehat{\mathbb{E}}(\theta_t|y_{1:t}) = \sum_{i=1}^{N} x_t^{(i)} w_t^{(i)} \tag{4.23}$$

where $w_t^{(i)}$ is the so-called importance or weight of $x_t^{(i)}$ (Petris et al. 2009, p. 209). The particles $x^{(1)}, \ldots, x^{(N)}$ with the associated weights $w^{(1)}, \ldots, w^{(N)}$ can be thought of as a discrete approximation of the probability distribution of $\theta_t|y_{1:t}$.

The question then simply becomes how to determine the particles and the weights. Any distribution assumption can be made, but for comparison with the Kalman filter, we will assume that the normality assumptions for Eq. 4.22 still hold and that $\theta_t|\theta_{t-1} \sim N_q(G_t\theta_{t-1}, W_t)$ and $Y_t|\theta_t \sim N_m(F_t\theta_t, V_t)$. Let the initial N particles $x_0^{(1)}, \ldots, x_0^{(N)}$ be drawn independently from a distribution $\theta_0 \sim N_q(c_0, C_0)$, and let the initial weights be $w_0^{(i)} = 1/N, i = 1, \ldots, N$. Then for $t \geq 0$,

(1) Draw new particles $x_t^{(i)}, i = 1 \ldots, N$ from $N_q(G_t\overline{x}_{t-1}, W_t)$,
(2) Calculate weights $\tilde{w}_t^{(i)} = f_{Y_t}(Y_t = y_t|F_t x_t^{(i)}, V_t)$ and normalise them: $w_t^{(i)} = \tilde{w}_t^{(i)} / \sum_{i=1}^{N} \tilde{w}_t^{(i)}$ for $i = 1, \ldots, N$,
(3) Perform multinomial resampling: Draw a random sample of N particles $x_t^{(i)}, i = 1 \ldots, N$ from the discrete distribution $P(\theta_t = x_t^{(i)}) = w_t^{(i)}, i = 1, \ldots, N$,
(4) Estimate $\widehat{\mathbb{E}}(\theta_t|y_{1:t}) = \overline{x}_t$ and $\widehat{\mathrm{Cov}}(\theta_t|y_{1:t}) = \mathrm{Cov}(x_t)$

(Petris et al. 2009, p. 211; Cappe et al. 2007; Dahlin and Schön 2019).

In practice, we set c_0 and C_0 to the average of the estimates for $\mathbb{E}[\theta_t|y_{1:t}]$ and $\mathrm{Cov}[\theta_t|y_{1:t}]$ found by the Kalman filter for the full W-model for the time points in year 2020 (2019 is used as adaptation period for the kalman filter; Table 14). Note that W_t here is also set to C_0 and is not updated during the algorithm. Updating $W_t = \widehat{\mathrm{Cov}}(\theta_t|y_{1:t})$ was also tried, but this leads to worse estimates: the variances in W_t can become very small, which decreases the spread of particles drawn in step

(1), and that makes the particle filter less flexible and results in bad adaptation in subsequent steps. Therefore, $W_t := C_0$ is used. This is even preferred to using the optimised W of the W-model, because C_0 contains larger variances, making the particle filter more flexible. Note that no optimisation is necessary because all variances are assumed as known. Missing data is treated the same way as for the Kalman filter. Custom `R`-functions are used to implement the particle filter. The `R`-code is submitted with this thesis.

Calculation of the k-step ahead predictions $f_{t+k} := \mathbb{E}(Y_{t+k}|y_{1:t})$ with covariance $Q_{t+k} := \mathrm{Cov}(Y_{t+k}|y_{1:t})$ is based on the linear state equations (Eq. 4.22) and therefore completely analogous to forecasting with the kalman filter. Using the particle filter estimates $a_t := \widehat{\mathbb{E}}(\theta_t|y_{1:t}) = \overline{x}_t$ and $R_t := \widehat{\mathrm{Cov}}(\theta_t|y_{1:t}) = \mathrm{Cov}(x_t)$, the following procedure is used (Petris et al. 2009, p. 70): For $k \geq 1$,

(1) $a_{t+k} = G_{t+k}a_{t+k-1}$ and
$R_{t+k} = G_{t+k}R_{t+k-1}G_{t+k}^T + W_{t+k}$
(2) $f_{t+k} = F_{t+k}a_{t+k}$ and
$Q_{t+k} = F_{t+k}R_{t+k}F_{t+k}^T + V_{t+k}$.

4.7.3 Optimisation of unknown parameters

Optimisation of unknown parameters is performed using the `stats::optim`-function in `R` (R Core Team 2021). The optimisation method used is "Nelder-Mead". The `optim`-function takes as input a vector of initial values for the parameters to be optimised, upper and lower bounds for the parameters to be optimised, an objective function whose value is to be minimised and any additional parameters required for the objective function. Here, the parameters to be optimised are the variances contained in V and W and also AR-coefficients for the A-model. Initial values for variances in V are set to 300 in the V-model and to 100282 (the optimised value) in the W- and A-models. The bounds are set to $(0, 10^6)$. Initial values for variances in W are set to 1, with bounds of $(0, 25000)$. Initial values for AR-parameters of the A-model are set to 0.1, with bounds of (-1,1) for the A-model and (0,1) for the A (+) model. The objective function to be minimised is a custom function that returns an optimisation measure for a model fit with the Kalman filter (V-model, W-model and A-model; no optimisation is necessary for LM- and P-models). Different optimisation measures are described and tested in section 7.1. The `R`-code for optimisation is submitted with this thesis.

5 Variables

In this chapter, we will first have a closer look at the target variable to better understand what features may influence it and how the dependence may be modeled. Based on these insights, features (covariates or predictors) are created that will be used in the subsequent modeling chapters.

5.1 Target variable

The main target variable in this thesis is the electrical power consumption (in Watt) of the heat pump ($Y = Y_{\text{pump}}$). Sometimes, the electrical power consumption of the household without the pump is used for comparison ($Y_{\text{household}}$). In the following, observations of the target variable are denoted by y.

Heat pump data shows strong seasonal behaviour (Fig. 3, 4; household 3 is used as a typical example). A spectral analysis is performed to determine the dominant seasonalities in the data, i.e. the periods of the Fourier basis functions that contribute most to the variation in the data (see section 4.1). For heat pump data, the dominant periods are 365.46 days (about 1 year) and 1.00 days (Fig. 3A). The yearly seasonality may not hit 365 days exactly because 2020 is a leap year. The seasonality with period of 1 day reflects the day-night cycle. Other seasonalities are also observed (e.g. 0.5 days), but are less powerful (Fig. 3A). In contrast, for household power consumption, the dominant periods are 1 day (the day-night cycle), some periods smaller than 1 day, and 7 days (the working week) and while there is a yearly seasonality, it is much less pronounced (Fig. 3B). In fact, the heat pump spectrum is more similar to temperature than household data. Temperature data also has dominant periods of 365.46 days and 1.00 day (Fig. 3C). The spectrum of temperature is "cleaner", however, lacking many of the small, less powerful periods apparent in pump data (Fig. 3A, C). Thus, the heat pump data probably reflects changes in temperature during the seasons, but on top of that there are other, unknown influences on heat pump power consumption.

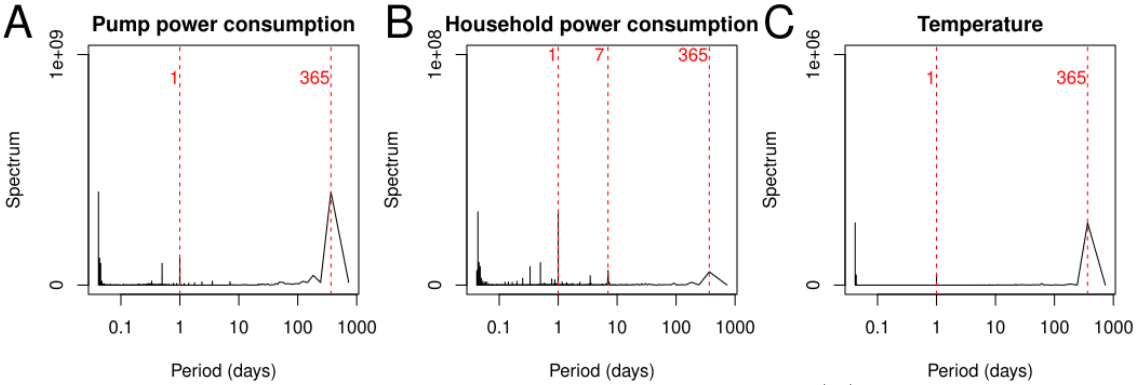


Figure 3: Spectral analysis for pump power consumption data (A) and household power consumption data (B) of household 3, as well as temperature data (C).

Knowing that the data is mainly composed of yearly and daily seasonalities, we can now separate y into different components to try to understand the structure of the data even better. We do not use the Fourier transform to decompose and recompose the data into separate parts but instead use arithmetic operations to stay as close to the data as possible:

$$y(d,h) = y_{\text{daily\_mean}}(d) + y_{\text{profile\_mean}}(d,h) + y_{\text{remainder}}(d,h) \tag{5.1}$$

where d is the date (day), which runs from the first of January 2019 to the 31st of December in 2020 and h is the daytime in full hours; $h = 0, ..., 23$. The means and the remainder in equation 5.1 are defined as follows:

$$\begin{aligned} y_{\text{daily\_mean}}(d) &= \tfrac{1}{24}\textstyle\sum_{h=0}^{23} y(d,h) \\ y_{\text{profile\_mean}}(d,h) &= \tfrac{1}{14}\textstyle\sum_{t=d-14}^{d}(y(t,h) - y_{\text{daily\_mean}}(t)) \\ y_{\text{remainder}}(d,h) &= y(d,h) - y_{\text{daily\_mean}}(d) - y_{\text{profile\_mean}}(d,h) \end{aligned} \tag{5.2}$$

Because there are only two years of measurements, we use the daily mean as an approximate for the yearly seasonality instead of a mean across years. The profile mean is the daily usage pattern. This can vary during the year so we calculate it using a left-aligned rolling window with only 14 days width. Fig. 4 shows the y-separation of heat pump data. An analogous analysis of household data is shown in Appendix A, Fig. 44.

Fig. 4 shows that $y_{\text{daily\_mean}}(d)$ captures the large scale variation in electrical power consumption over the course of a year (Fig. 4A; red line). The remaining data $(y(d,h) - y_{\text{daily\_mean}}(d))$ is mostly captured by the variation during the day, (Fig. 4B, C; blue line). Fig. 4D confirms that $y_{\text{daily\_mean}}(d) + y_{\text{profile\_mean}}(d)$ captures the essence of the target variable. The remainder, $y_{\text{remainder}}(d,h)$, contains outliers that were not detected during cleaning (section 3.2) and noise (Fig. 4E). Note that the noise is not Gaussian (Fig. 4F-H). The noise may be generated by an ARMA process: although no obvious parametrisation suggests itself, AR-components are clearly visible in the ACF and PACF (Fig. 4G-H). There also appears to be a small 1-day seasonal component still present in the noise, indicating that the profile mean may not capture they daily seasonality completely.

The y-separation analysis suggests that a simple state space model may not be sufficient to explain the data unless it allows for structured, non-white noise. On the other hand, a simpler model could be used to explain the main seasonal data, $y_{\text{daily\_mean}}(d) + y_{\text{profile\_mean}}(d,h)$.

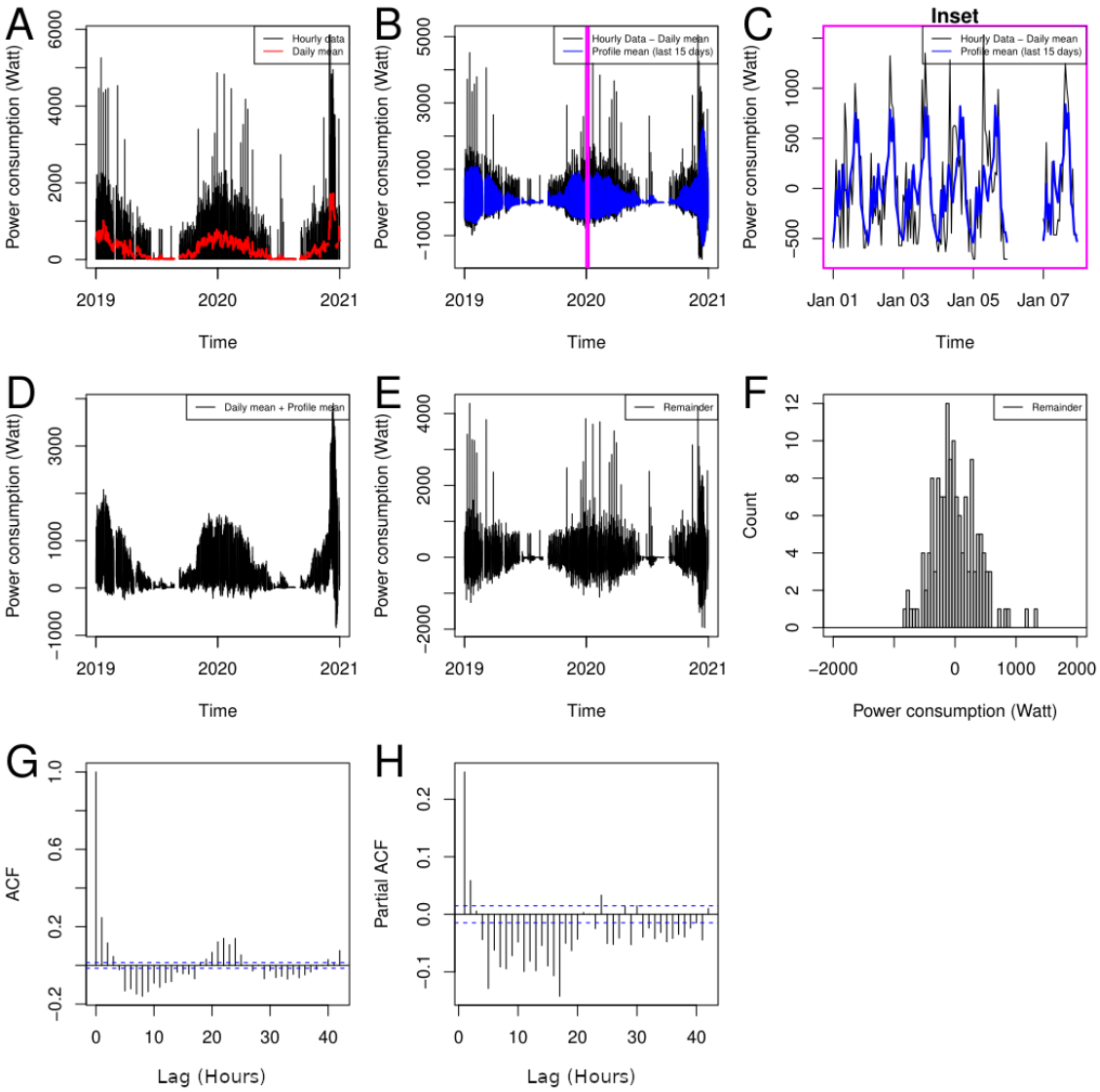


Figure 4: Typical electrical power consumption of a heat pump (household 3). A: cleaned data in black, $y_{\text{daily\_mean}}$ in red; B: $y - y_{\text{daily\_mean}}$ in black, $y_{\text{profile\_mean}}$ in blue; C: close-up of the region marked in magenta in B; D: $y_{\text{daily\_mean}} + y_{\text{profile\_mean}}$; E: $y_{\text{remainder}}$, note large peaks; F: histogram of $y_{\text{remainder}}$, note that the histogram is cut off at 1200 Watt, so does not show the largest peaks visible in E; G: ACF of $y_{\text{remainder}}$; H: PACF of $y_{\text{remainder}}$.

Load profiles

Fig. 4C shows that daily usage patterns contribute significantly to the structure of the data. These usage patterns can be different in different households, but there could also be similarities, for example if two households have pumps of the same type or are inhabited by persons with a similar lifestyle. To analyse usage patterns more closely, we therefore cluster households by their mean daily usage patterns (Fig. 5). For the purpose of clustering, each pump is characterised by a 24-dimensial vector - the mean power consumption of the pump during the course of a day after subtracting the daily mean, or "centered load profile", calculated based on 2019-2020 data. The load profiles are centered (but not scaled), so that the baseline is

zero for all pumps and does not interfere with the mean profiles and clustering. The centered load profiles are then clustered using Ward's minimum variance method (an agglomerative hierarchical clustering method that minimises the total within-cluster variance; see section 4.4). This cluster analysis is done separately for pump load profiles (Fig. 5, Fig. 6) and household load profiles (excluding the pumps; Appendix A: Fig. 45, Fig. 46), because the usage of pumps is expected to be independent of the usage of other electrical equipment.

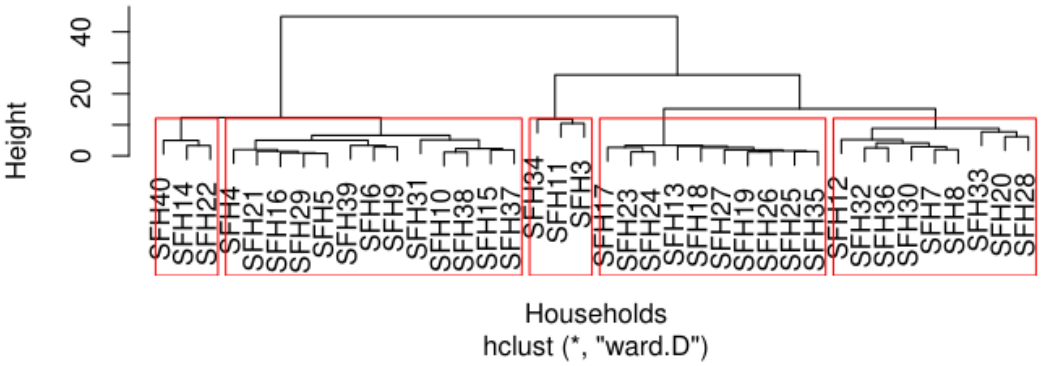


Figure 5: Pumps clustered based on mean daily electrical power consumption profiles in 2019-2020.

Cluster	Households	Size
1	4, 5, 6, 9, 10, 15, 16, 21, 29, 31, 37, 38 and 39	$m_1 = 13$
2	3, 11 and 34	$m_2 = 3$
3	7, 8, 12, 20, 28, 30, 32, 33 and 36	$m_3 = 9$
4	13, 17, 18, 19, 23, 24, 25, 26, 27 and 35	$m_4 = 10$
5	14, 22 and 40	$m_5 = 3$

Table 5: Pump clusters.

Fig. 5 shows five clearly separable pump clusters (Table 5). Note that these include all households, also households that are later excluded from further analysis (see chapter 3). Since the clusters are very distinct, it is worth looking at the in-cluster load profiles in more detail to see how they differ. For this purpose, we recalculate the centered load profiles for each cluster, but this time we also separate the profiles by season and weekday / weekend. Note that we use six seasons to take daylight saving time into account (Appendic B: Table 32; Grolemund and Wickham 2011, `R: lubridate::dst`), but for clarity, only summer and winter are shown in Fig. 6.

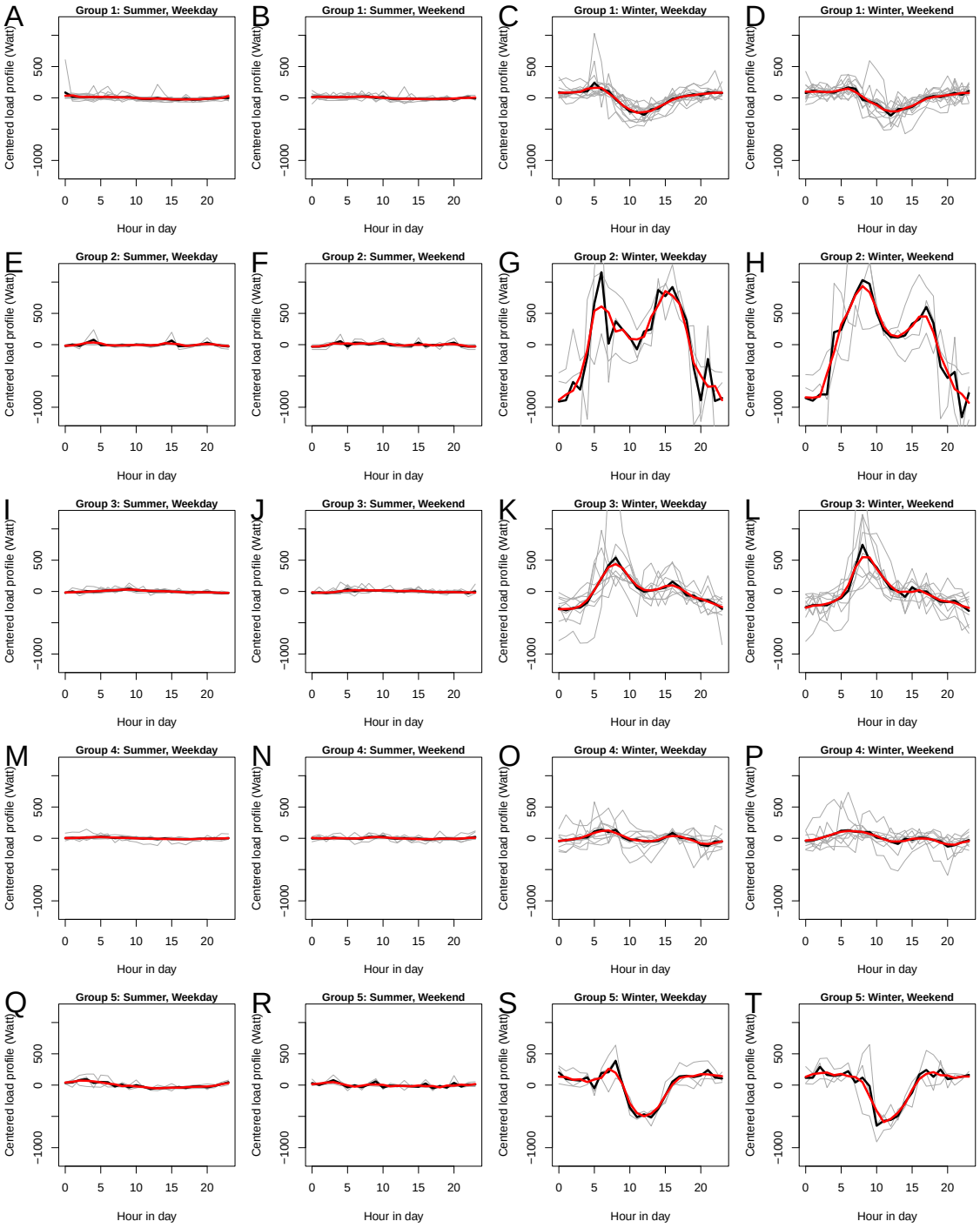


Figure 6: Mean centered pump load profiles by cluster, season and weekend. Rows represent pump clusters. The first two columns are summer weekdays and summer weekend profiles; the last two columns represent winter weekday and winter weekend profiles. Gray: load profiles of the individual pumps included in each cluster; black: mean profile; red: moving average of the mean (using a circular rolling window of three hours width).

In summer, all centered pump cluster load profiles are close to zero (Fig. 6A-R). This is expected because no or little space heating is required during the summer months. However, in winter, there are large differences between clusters. For example cluster 1 has a small amplitude, indicating that the pump is used almost equally throughout

the day, with only a small dip around noon (Fig. 6C). By contrast, cluster 2 has a very large amplitude with heating mostly in the morning and late afternoon, a small dip at noon and a large dip in heating during the night. Cluster 3 heats more in the morning than in the afternoon. In conclusion, clusters really reflect different usage patterns of the pumps. It is also interesting to note that pump clusters and household clusters are not identical, confirming that pump usage is independent of other electrical equipment (Fig. 5, Appendix A: Fig. 45).

5.2 Predictor variables

The previous section shows that there are slow (yearly) and fast (daily) variations in pump power consumption (Fig. 3, 4). Slow variations could be explained by seasonal changes in temperature or daylight hours, which both increase in the summer and decrease during winter time. Fast variations could be explained by temperature- or daylight-changes during the day as well as house-specific usage patterns. In this section, we look at temperature, seasons, load profiles and house-specific features as potential predictors. We mainly use 2019 data to create new features for the dataset. All features are then normalised to their maximum value in 2019 to make them comparable in scale. Note that "predictor variables", "variables" and "features" are used synonymously in the following sections.

Temperature

A heat pump heats space by transferring heat from a cool place (outside) to a warm place (inside). Therefore, the outside temperature (T) is the most obvious predictor for heat pump power consumption. For the households in this dataset, the desired inside temperature is unknown, but measurements for the outside temperature are available (Table 3). The relationship between power consumption and (outside) temperature is shown in Fig. 7. Because the variance in the hourly data is huge, Fig. 7A, D also show the relationship between the daily means of electrical power consumption ($y_{\text{daily\_mean}}$; equation 5.2) and temperature ($T_{\text{daily\_mean}}(d) = \frac{1}{24}\sum_{h=0}^{23} T(d,h)$).

For the heat pump, power consumption decreases with increasing temperature (Fig. 7A). This relationship appears to be linear, with a "kink" between 15° C and 18° C, after which power consumption no longer decreases (Fig. 7A). This could correspond to the outside temperature at which pumps stop heating. The "kink" appears to be present in the pump data of all households (Fig. 7B).

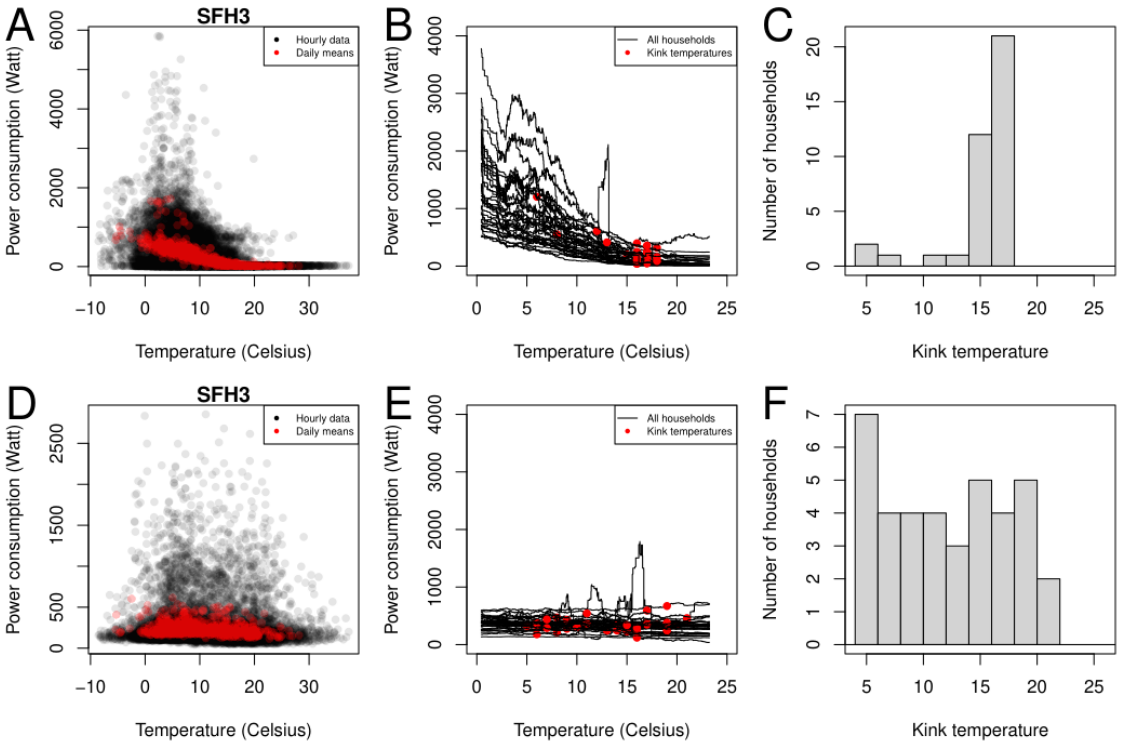


Figure 7: Typical dependence of electrical power consumption on temperature (house SFH3) for the heat pump (A-C) and the remaining household (D-F). A, D: cleaned data in black, $y_{\text{daily\_mean}}$ vs. $T_{\text{daily\_mean}}$ in red; B, E: moving averages of $y_{\text{daily\_mean}}$ vs. $T_{\text{daily\_mean}}$ using a rolling window of 41 days width (each black line represents one household), with the "kink" temperatures of the different households marked in red; C, F: histogram of the "kink" temperatures.

Because the "kink" temperature is relevant for modeling, we determine it for each pump individually, using the following linear model:

$$\begin{aligned} Y_{\text{daily\_mean}}(\text{j}) = \alpha(j, T_{\text{kink}}) \quad & + T_{\text{daily\_mean}} \cdot \mathbb{I}_{[T \leq T_{\text{kink}}]} \cdot \beta_1(j, T_{\text{kink}}) \\ & + T_{\text{daily\_mean}} \cdot \mathbb{I}_{[T > T_{\text{kink}}]} \cdot \beta_2(j, T_{\text{kink}}) + \epsilon \end{aligned} \tag{5.3}$$

for $j = 1, \ldots, 38$ pumps and $T_{\text{kink}} = 5°C, 6°C, ..., 25°C$. Here, ϵ is assumed to be i.i.d normally distributed with expectation zero and a variance > 0. Note that the intercepts α and the slopes β_1 and β_2 are of no particular interest here and therefore no estimates are given. Instead, the question is which value for T_{kink} yields the best fit for each individual pump. To determine the best-fitting "kink" temperature, T_{kink} is varied from 5° C to 25° C in steps of 1° C. A fit is performed for each value of T_{kink} and the R^2 of the fit is recorded. The value of T_{kink} which yields the largest R^2 is used as an estimate for the true kink temperature for the given pump. For the heat pumps in this dataset, the kink temperatures range mainly from 15° C to 18° C (Fig. 7B, C). For comparison, the same analysis is performed for the household power consumption, excluding heat pumps. In this case, no strong dependence on temperature is observed (Fig. 7D) and there is no well-defined "kink" temperature (Fig. 7E, F). This is consistent with the previous observation that unlike pump power consumption, household power consumption does not strongly mirror seasonalities in temperature (Fig. 3).

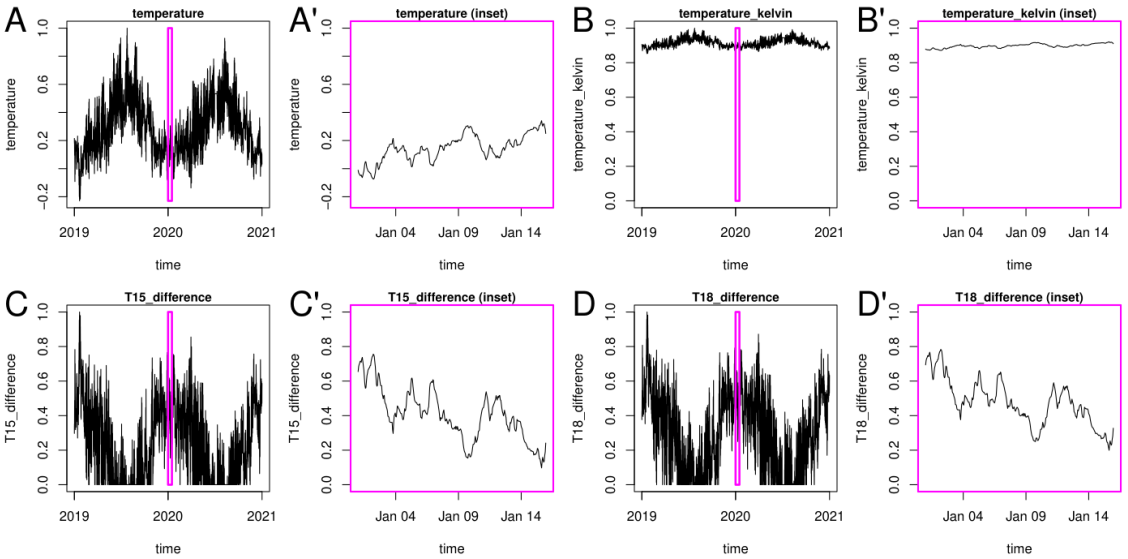


Figure 8: Predictor variables based on temperature. Features are shown on the left and a close-up of a two-week period in January 2020 is shown on the right (magenta). All variables are normalised to the maximum value in year 2019. A, A': temperature based on ° C; B, B': absolute temperature based on ° K; C, C': "T15_difference"; D, D': "T18_difference".

Based on this analysis, the following predictor variables are created (Fig. 8): the temperature (in ° Celsius), the absolute temperature (in ° Kelvin), "T15_difference" $= (15°C - T) \cdot \mathbb{I}_{[T \leq 15°C]}$, "T18_difference" $= (18°C - T) \cdot \mathbb{I}_{[T \leq 18°C]}$, "Tse15" $= \mathbb{I}_{[T \leq 15]}$ (se: "smaller equal"), "Tg15" $= \mathbb{I}_{[T > 15]}$ (g: 'greater'), "Tse18" $= \mathbb{I}_{[T \leq 18]}$ and "Tg18" $= \mathbb{I}_{[T > 18]}$. The absolute temperature is included for comparison and because its values cannot become negative (or zero in the present data). Because the "kink" can occur at different temperatures for different pumps, two cutoff temperatures are featurised (15° C and 18° C). "T15_difference" and "T18_difference" cut off values above the respective "kink" temperatures and then flip the temperature profile upside down. The other variables are binary variables describing the "kink". Some of the predictor variables are complementary (for example: Tse15 and Tg15), however, this is not a problem; it just means that later, at most one of them will be selected during forward feature selection. All variables are normalised to the maximum value in 2019 to make them comparable in scale (Table 6).

Variable	Maximum in 2019
number_of_persons	4.00
square_metres	230.00
pump_power_at_zero	4008.72
household_power_at_zero	613.81
temperature	37.25
T15_difference	23.54
T18_difference	26.54
temperature_kelvin	310.40
day_profile_pump	937.27
day_profile_household	651.96
altitude	1.06
maxaltitude	2.12
daylight_hours	16.81

Table 6: Maximum values of new metric variables in 2019, used for normalisation.

Season profiles

Fig. 7A-C suggest that the electrical power consumption of heat pumps depends on temperature. The temperature data contains a lot of noise, which may or may not be directly reflected in the heat pump profile. We provide a smooth alternative: the season profile (Fig. 9, 10). The season profile is determined based on the mean daily temperature, using the following model (Fig. 9):

$$\begin{aligned} &T_{\text{daily\_mean}}(d) = \alpha + \sin(\tfrac{2\pi d}{l})\beta_1 + \cos(\tfrac{2\pi d}{l})\beta_2 + \sin(\tfrac{4\pi d}{l})\beta_3 + \cos(\tfrac{4\pi d}{l})\beta_4 + \epsilon, \\ &\widehat{\alpha} = 10.6796; \widehat{\beta}_1 = -2.0194; \widehat{\beta}_2 = -8.1666; \widehat{\beta}_3 = 0.8902; \widehat{\beta}_4 = 0.4472 \\ &R^2 = 0.7654 \end{aligned} \tag{5.4}$$

where l is the period of the maximum spectral density of the temperature data (365 days: Fig. 3C), and $d = 1, ..., 365$ are consecutive days in the year 2019. Including second harmonics effectively also includes the period of 182.5 days ($\frac{4d}{l} = \frac{2d}{l/2}$), which improves the fit. The error term ϵ is assumed to be i.i.d normally distributed with expectation zero and variance greater zero. The OLS estimator for the intercept, $\widehat{\alpha}$, the amplitudes of the periodic terms, $\widehat{\beta}$, and the R^2 are given.

The OLS estimator is used to calculate a season profile for 2019 and 2020 (for simplicity, the value of 29.2.2020 is set to the same value as 28.2.2020). The season profile obtained this way is min-max-normalised and flipped upside-down to obtain a "season_profile" variable with values between 0 and 1 that peaks in winter (Fig. 10A). Separate winter and summer profiles are then obtained from the "season_profile" variable by subtracting 0.5, discarding values above zero (for the summer profile) or below zero (for the winter profile) and min-max scaling the resulting profiles (Fig. 10B, C). The season, summer and winter profiles allow a future model to select different kinds of profile shapes with different steepness.

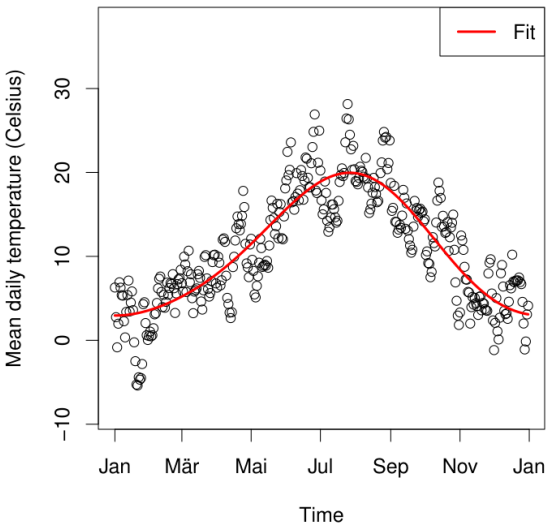


Figure 9: Seasonal profile based on the mean daily temperature in 2019.

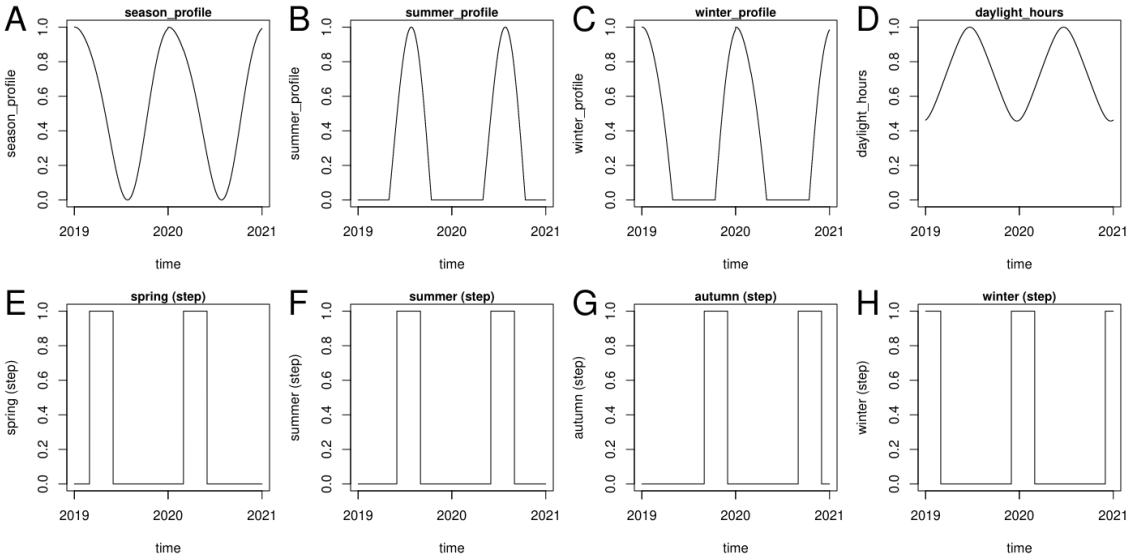


Figure 10: Predictor variables based on seasons. All variables are normalised to the maximum value in year 2019.

Household power consumption (excluding heat pumps) may not depend directly on temperature (Fig. 7D-F) but there may still be a small yearly seasonal dependence (Fig. 3B), for example on the variation in daylight hours: people tend to use more electric light during the winter months, when the days are shorter, than during the summer. For this reason, we also include "daylight_hours" = time of sunset−time of sunrise as a feature (Fig. 10D; Thieurmel and Elmarhraoui 2022, `R: sunCalc::getSunlightTimes`). Finally, we also include binary variables for the traditional four seasons (spring (month = March, April or May), summer (month = June, July or August), autumn (month = September, October or November) and winter (month = December, January or February); Fig. 10E-H), although it is unlikely that these features will be selected by any model because they change much

more abruptly than pump or household data (Fig. 4, Appendix A: Fig. 44).

In summary, in addition to temperature, the following predictor variables can capture long-term seasonal dependencies (Fig. 10): season_profile, summer_profile, winter_profile, daylight_hours, spring, summer, autumn and winter (binary variables). All variables are normalised to the maximum value in year 2019 to make them comparable in scale (Table 6).

Day profiles

Fig. 4 shows that short-term (daily) seasonalities also contribute to electrical power consumption. For pumps, temperature could explain yearly as well as daily seasonalities. However, household power consumption is more likely to depend on usage patterns (for example: times of cooking, watching television, using the internet, using the washing machine, and so on). These cannot be predicted by external factors. For this reason, two new features are created: "day_profile_pump" and "day_profile_household" (Fig. 11A, B). These features are based on the mean centered load profiles of year 2019 (2020 data is not included). The centered load profiles are available by cluster, season (Appendix B, Table 32) and weekday / weekend (Fig. 6 and Appendix A, Fig. 46). The values of the features are determined as follows: For every time point in the dataset and for every household in the dataset, the "day_profile_pump" feature is assigned the value of the centered load profile value for the cluster of the given household's pump, and for the season, weekday / weekend and hour of day of the given time point. The "day_profile_household" feature is constructed analogously (Fig. 11B). Note that excluding 2020 data is necessary when constructing the features because later, 2020 data is used to analyse the goodness of fits. Unfortunately, 2019 was a pre-Covid year and 2020 was not, so the load profile features may not be able to capture 2020 data perfectly.

In addition, temperature and daylight both depend on the sun altitude in the sky, and sun altitude also varies both yearly and daily, so could contribute to both seasonal variations in pump and household data. For this reason, the sun altitude at the location of Hamelin (latitude 52.3759°N and longitude 9.7320°E) is also added to the predictor variables (Thieurmel and Elmarhraoui 2022, `R: sunCalc::getSunlightTimes`). The sun altitude is measured in radians (for example it is 0 when the sun is at the horizon and $\pi/2$ at the zenith). Importantly, the sun altitude is negative if the sun is located below the horizon, i.e. at night. To avoid negative numbers in the variable, we shift and flip the variable: "sun_altitude" = (max(altitude) − altitude) (Fig. 11C). All variables are normalised to the maximum value in year 2019 (Table 6).

Finally, the weekend (day = Saturday or Sunday) is added to the predictor variables (Fig. 11D; hours are not shown), mainly because in the spectral analysis, a period of 7 days is seen in the spectral analysis for both pump consumption data (Fig. 3A; not very powerful) and household power consumption (Fig. 3B; quite powerful). To capture periods smaller than one day, the hours of the day (00:00-03:59, 04:00-07:59, 08:00-11:59, 12:00-15:59, 16:00-19:59, 20:00-23:59) are also added as binary variables (not shown).

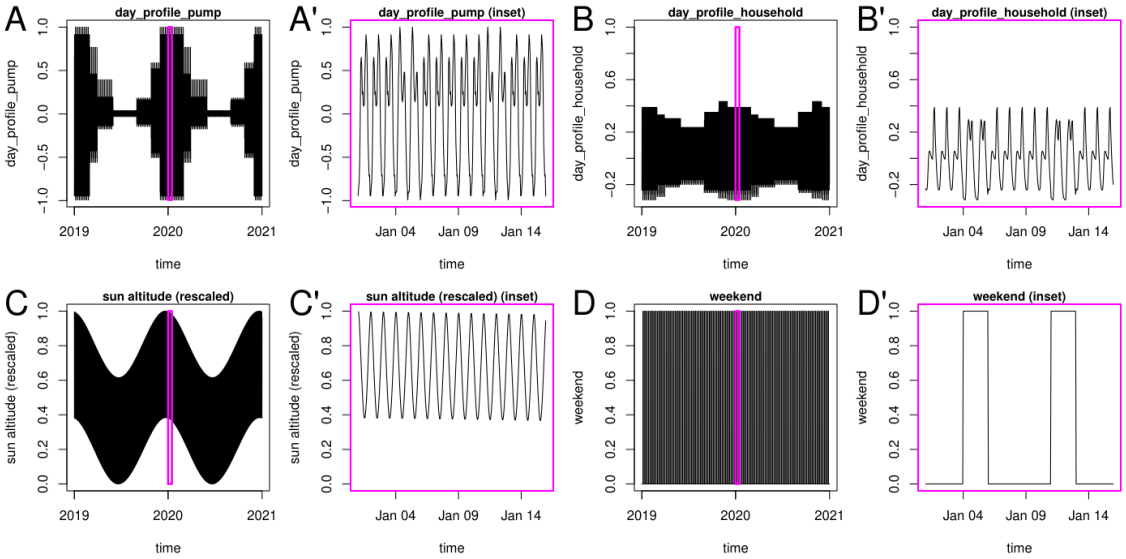


Figure 11: Predictor variables based on daily patterns. For readability, in A and B, only the profile for one cluster is shown (A: pump cluster 2, B: cluster 5; both clusters include household 3). C, C': sun altitude; D, D': weekend factor. All variables are normalised to the maximum value in year 2019.

Household-specific features

Households do not only differ by their usage patterns. Larger houses take more power to heat, and older, less efficient pumps may require more electricity to heat a house of the same size. Similarly, household power consumption is likely to be affected by the number of persons living in the house (German Federal Statistics Office 2021). For these reasons, we also consider the number of persons and the building floor size in m^2 as predictor variables (Table 7, Fig. 12). For these features, missing values for a variable are imputed using the mean of the variable values across households. The efficiency of pumps is unknown, as is the desired inside temperature. To capture these effects, we calculate the mean pump power consumption and mean household power consumption for each household for temperatures $|T| < 1°C$ in 2019, yielding new variables "pump_power_at_zero" and "household_power_at_zero" (Table 7, Fig. 12). The magnitude of the "pump_power_at_zero" variable is affected by (unknown) pump efficiency but also building size and desired indoor temperature, so it is not a clean measure of any one of the three, but it does give some indication of efficiency. All household specific features are normalised to the maximum value of the feature across all households (Fig. 12, Table 6).

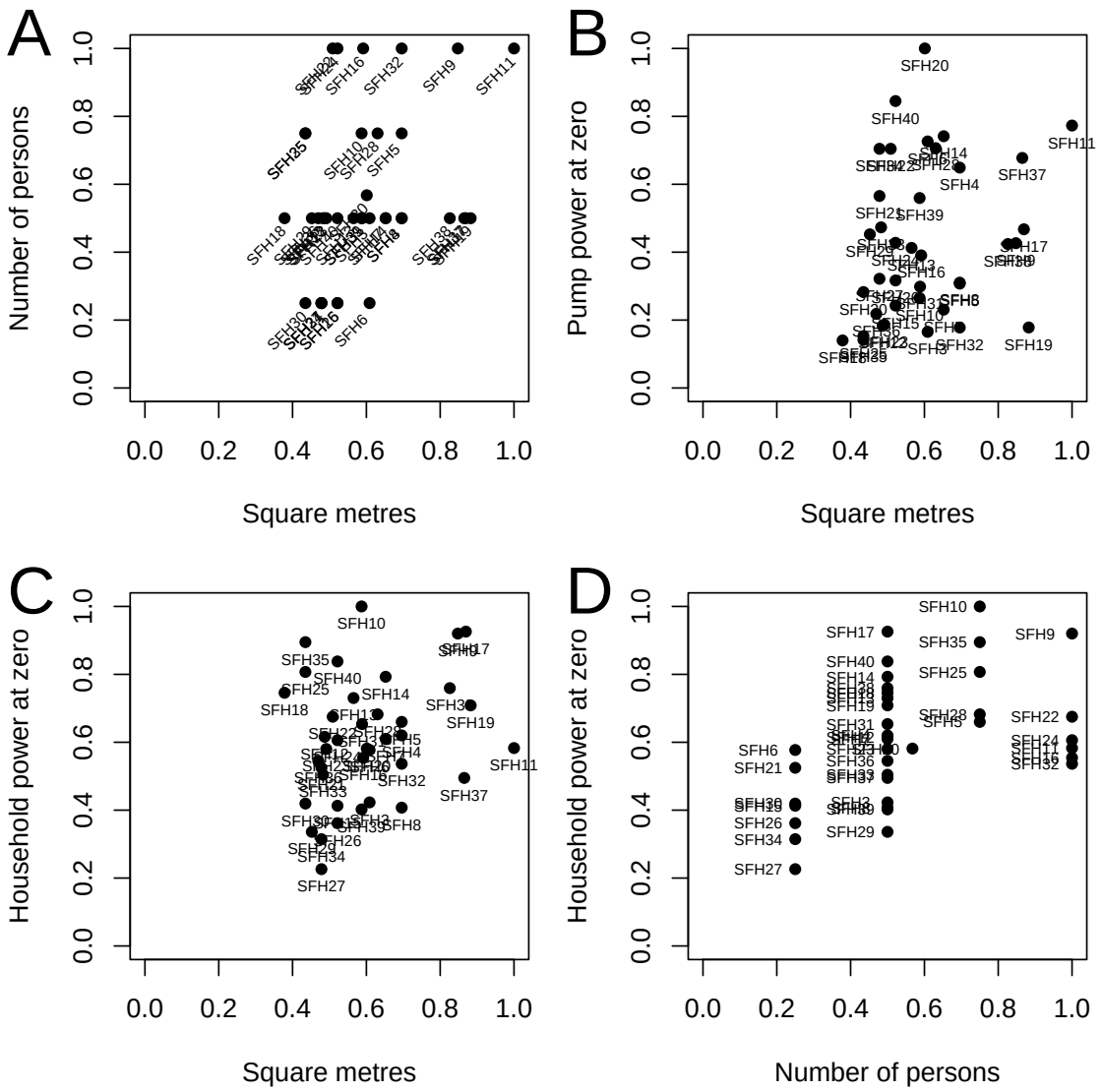


Figure 12: Household-specific features. Since these variables are constant in time, they are plotted against each other rather than over time. All variables are normalised to the maximum value of all houses.

Interactions

In addition to the new predictor variables introduced in the previous sections, interactions between these predictors are also included up to the fourth order. Especially plausible are interactions of household-specific features with seasons and temperature, but interactions between season profiles and temperature are also included. This corresponds roughly to quadratic effects of temperature (since the season profile is determined based on the mean daily temperature in year 2019), but is less noisy than quadratic temperature. Since $y_{\text{daily\_mean}}$ and $y_{\text{profile\_mean}}$ are additive, one would not expect interactions between seasons or temperature and load profiles. However, we do include some interactions between season profiles and load profiles just in case. It is better to have too many interactions rather than too few, because later, forward selection is performed, so irrelevant predictors will not be selected.

All existing, new and interaction variables are shown in Table 8. The naming of interaction variables is agglunative. For example, temperature_T15_season_profile is the interaction between temperature, Tse15 and season_profile variables. The feature season_profile2_pump_power_at_zero is the interaction between season_profile squared with the pump_power_at_zero variable. In total, there are 145 features and interactions in the dataset.

id	number_of_persons	square_metres	pump_power_at_zero	household_power_at_zero
SFH3	2.00	140.00	662.83	259.60
SFH4	2.00	160.00	2602.23	380.69
SFH5	3.00	160.00	1232.18	405.09
SFH6	1.00	140.00	2910.90	354.30
SFH7	2.00	150.00	925.33	374.36
SFH8	2.00	160.00	1243.39	250.04
SFH9	4.00	195.00	1709.75	564.79
SFH10	3.00	135.00	1062.00	613.81
SFH11	4.00	230.00	3098.70	357.82
SFH12	2.00	112.00	735.63	378.18
SFH13	2.00	130.00	1653.92	448.21
SFH14	2.00	150.00	2972.34	486.67
SFH15	1.00	120.00	973.10	253.52
SFH16	4.00	136.00	1565.96	340.29
SFH17	2.00	200.00	1873.57	568.35
SFH18	2.00	87.00	561.42	457.74
SFH19	2.00	203.00	714.28	434.94
SFH20	*2.27	*138.25	4008.72	356.86
SFH21	1.00	110.00	2267.07	322.10
SFH22	4.00	117.00	2824.33	414.46
SFH23	2.00	113.00	747.87	356.12
SFH24	4.00	120.00	*1713.96	*372.02
SFH25	3.00	100.00	613.07	495.42
SFH26	1.00	120.00	1269.63	222.01
SFH27	1.00	110.00	1289.65	138.84
SFH28	3.00	145.00	2831.70	418.81
SFH29	2.00	104.00	1813.90	206.31
SFH30	1.00	100.00	1132.64	257.44
SFH31	2.00	135.28	1196.56	401.20
SFH32	4.00	160.00	713.39	329.36
SFH33	2.00	111.00	1896.61	309.52
SFH34	1.00	110.00	2823.94	193.01
SFH35	3.00	100.00	568.94	549.24
SFH36	2.00	108.00	874.73	334.63
SFH37	2.00	199.00	2715.97	303.77
SFH38	2.00	190.00	1699.21	466.14
SFH39	2.00	135.00	2243.06	246.88
SFH40	2.00	120.00	3387.94	514.38

Table 7: Values of constant household-specific features. Starred values indicate imputation of a missing value by the mean of the variable across all households.

Variable
number_of_persons, square_metres, pump_power_at_zero, household_power_at_zero,
Hour0, Hour4, Hour8, Hour12, Hour16, Hour20, weekend,
Spring, Summer, Autumn, Winter, season_profile, summer_profile, winter_profile,
altitude, maxaltitude, daylight_hours,
day_profile_pump, day_profile_household,
temperature, Tse15, Tg15, Tse18, Tg18, T15_difference, T18_difference, temperature_T15, temperature_T18,
temperature_persons, temperature_sqmetres,
temperature_T15_persons, temperature_T15_sqmetres, temperature_T18_persons, temperature_T18_sqmetres,
temperature_kelvin, temperature_kelvin_T15, temperature_kelvin_T18,
temperature_kelvin_T15_sqmetres, temperature_kelvin_T18_sqmetres,
T15_difference_persons, T18_difference_persons, T15_difference_sqmetres, T18_difference_sqmetres,
temperature_T15_season_profile, temperature_T18_season_profile,
season_profile_sqmetres, winter_profile_sqmetres,
T15_difference_season_profile, T18_difference_season_profile,
T15_difference_season_profile_sqmetres, T18_difference_season_profile_sqmetres,
temperature_T15_winter_profile, temperature_T18_winter_profile,
T15_difference_winter_profile, T18_difference_winter_profile,
T15_difference_winter_profile_sqmetres, T18_difference_winter_profile_sqmetres,
season_profile_sqmetres_T15, season_profile_sqmetres_T18,
winter_profile_sqmetres_T15, winter_profile_sqmetres_T18,
day_profile_pump_sqmetres, day_profile_pump_sqmetres_daylight, day_profile_pump_sqmetres_sunaltitude,
day_profile_pump_sqmetres_season_profile, day_profile_pump_sqmetres_winter_profile,
day_profile_pump_sqmetres_season_profile_T15, day_profile_pump_sqmetres_winter_profile_T15,
day_profile_pump_T15, day_profile_pump_T18, day_profile_pump_T15_difference, day_profile_pump_T18_difference,
day_profile_pump_sqmetres_T15, day_profile_pump_sqmetres_T18,
day_profile_pump_sqmetres_T15_difference, day_profile_pump_sqmetres_T18_difference,
day_profile_pump_sqmetres_daylight_T15, day_profile_pump_sqmetres_daylight_T18,
day_profile_pump_sqmetres_sunaltitude_T15, day_profile_pump_sqmetres_sunaltitude_T18,
day_profile_household_persons, day_profile_household_sqmetres,
day_profile_household_daylight, day_profile_household_sunaltitude,
day_profile_household_sqmetres_daylight, day_profile_household_sqmetres_sunaltitude,
temperature_T15_pump_power_at_zero, temperature_T18_pump_power_at_zero,
T15_difference_pump_power_at_zero, T18_difference_pump_power_at_zero,
season_profile_pump_power_at_zero, winter_profile_pump_power_at_zero,
winter_profile_pump_power_at_zero_T15, winter_profile_pump_power_at_zero_T18,
season_profile_pump_power_at_zero_T15, season_profile_pump_power_at_zero_T18,
T15_difference_season_profile_pump_power_at_zero, T15_difference_winter_profile_pump_power_at_zero,
T18_difference_season_profile_pump_power_at_zero, T18_difference_winter_profile_pump_power_at_zero,
T15_difference_season_profile_sqmetres_pump_power_at_zero, T15_difference_winter_profile_sqmetres_pump_power_at_zero,
T18_difference_season_profile_sqmetres_pump_power_at_zero, T18_difference_winter_profile_sqmetres_pump_power_at_zero,
day_profile_pump_power_at_zero, day_profile_pump_power_at_zero_daylight, day_profile_pump_power_at_zero_sunaltitude,
day_profile_pump_power_at_zero_season_profile, day_profile_pump_power_at_zero_winter_profile,
day_profile_pump_power_at_zero_season_profile_T15, day_profile_pump_power_at_zero_winter_profile_T15,
day_profile_pump_power_at_zero_T15, day_profile_pump_power_at_zero_T18,
day_profile_pump_power_at_zero_T15_difference, day_profile_pump_power_at_zero_T18_difference,
day_profile_pump_power_at_zero_daylight_T15, day_profile_pump_power_at_zero_daylight_T18,
day_profile_pump_power_at_zero_sunaltitude_T15, day_profile_pump_power_at_zero_sunaltitude_T18,
day_profile_household_power_at_zer,o
day_profile_household_power_at_zero_daylight, day_profile_household_power_at_zero_sunaltitude,
temperature_T18_2, temperature_T15_2,
T15_difference2, T18_difference2,
T15_difference2_sqmetres, T18_difference2_sqmetres,
T15_difference2_pump_power_at_zero, T18_difference2_pump_power_at_zero,
T15_difference2_sqmetres_pump_power_at_zero, T18_difference2_sqmetres_pump_power_at_zero,
season_profile2, winter_profile2,
season_profile2_sqmetres, winter_profile2_sqmetres,
season_profile2_pump_power_at_zero, winter_profile2_pump_power_at_zero,
season_profile2_sqmetres_pump_power_at_zero, winter_profile2_sqmetres_pump_power_at_zero

Table 8: All 145 predictor variables and interactions included in the dataset. For interactions, the names of the interacting features are concatenated. 2:= squared; T15 := Tse15; T18 := Tse18.

6 Exploration using a linear model

There are 145 predictors, and Kalman optimisation is slow and therefore unsuited to exploration. For this reason, we use linear modeling with ordinary least squares (OLS) first to narrow down predictors and answer questions such as: What is the necessary number of forward selection rounds? Should one fit households individually, in clusters, or all together? Is a manual pre-selection of households based on missing data and other criteria sensible (Table 4)? Should one fit the full data or the reduced seasonal data $y_{\text{daily\_mean}} + y_{\text{profile\_mean}}$ (Eq. 5.2; Fig. 4)? Since the real θ is assumed to be constant in the linear model, the linear model is simpler than a state space model and does not require optimisation of variances, which makes it faster. Furthermore, since the linear model is not dynamic, i.e. does not depend on time, missing data can just be excluded from the estimation.

The linear model used in this chapter is:

$$Y = I_{mxm}\alpha + X_{mxp}\beta + \epsilon, \tag{6.1}$$

where

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} \in \mathbb{R}^m,\ \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} \in \mathbb{R}^m,\ \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \in \mathbb{R}^p \text{ and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{bmatrix} \in \mathbb{R}^m. \tag{6.2}$$

Here, Y is the power consumption of m heat pumps observed at the same time. In this chapter, m can range from $m = 1$ to $m = 33$, depending on how many households are fit simultaneously. The vector α contains the intercepts for the m pumps, and the vector β contains the effects of the p predictor variables, which form the columns of X_{mxp}. These effects are assumed to be the same for all pumps. The random vector ϵ is assumed to have expectation zero and variance $\sigma^2 I_{mxm}$, but for the moment it is not assumed to follow a multivariate normal distribution, since that is unnecessary for the work in this chapter. This means that the distribution of Y is unknown. However, because the expectation of ϵ is zero, the expectation of Y equals $I_{mxm}\alpha + X_{mxp}\beta$ and its variance is $\sigma^2 I_{mxm}$.

For a training dataset that comprises time points t_0 up to t, the OLS estimator is given by:

$$\widehat{\theta}_t = \begin{bmatrix} \widehat{\alpha}_t \\ \widehat{\beta}_t \end{bmatrix} = \left(\tilde{X}_t^T \tilde{X}_t\right)^{-1} \tilde{X}_t^T y_t, \text{ where } \tilde{X}_t := \begin{bmatrix} I_{mxm} & X_{mxp}(t_0) \\ I_{mxm} & X_{mxp}(t_1) \\ \vdots & \vdots \\ I_{mxm} & X_{mxp}(t) \end{bmatrix} \text{ and } y_t := \begin{bmatrix} y(t_0) \\ y(t_1) \\ \vdots \\ y(t) \end{bmatrix}. \tag{6.3}$$

Because we do not assume a multivariate normal distribution for ϵ, the distribution of $\widehat{\theta}_t$ is unknown. This means that we cannot calculate confidence intervals and significance for parameter estimates. However, we can still calculate the variance:

$$\text{Cov}\left(\widehat{\theta}_t\right) = \sigma^2 (\tilde{X}_t^T \tilde{X}_t)^{-1}. \tag{6.4}$$

An estimate of this covariance matrix can be calculated using the following estimate for σ^2: $\widehat{\sigma}^2 = \frac{1}{mT-(m+p)} \left|\left| y_t - \tilde{X}_t \widehat{\theta}_t \right|\right|_2^2$. Here, T is the total number of time points used in Eq. 6.3. The diagonal elements of the covariance estimate are used as estimators for the variances of the individual θ elements.

Finally, the 1-step ahead prediction for all included pumps at time point $t+1$ with data $X_{mxp}(t+1)$ and design matrix $\tilde{Z}_{t+1} = \begin{bmatrix} I_{mxm} & X_{mxp}(t+1) \end{bmatrix}$ is given by

$$\widehat{y}(t+1) = \tilde{Z}_{t+1}\widehat{\theta}_t, \quad \text{Cov}\left(\widehat{y}(t+1)\right) = \sigma^2 \tilde{Z}(\tilde{X}_t^T \tilde{X}_t)^{-1} \tilde{Z}^T + \sigma^2 I_{mxm}, \tag{6.5}$$

and the prediction residuals are $r(t+1) = y(t+1) - \widehat{y}(t+1)$. Again, we can estimate the covariance in Eq. 6.5 by replacing σ^2 with $\widehat{\sigma}^2$. The diagonal elements of the estimated covariance matrix are used to estimate the variances of the 1-step ahead predictions for the pumps included in the fit. Note that no distribution assumptions are necessary to obtain these variances.

In the remainder of this chapter, we proceed the following way: First, we define a number of fit "conditions". These are simply binary decisions that could be made when fitting the data and should provide insights into how these decisions affect the goodness of the fit. The purpose here is to determine the best fit conditions using the fast linear model procedure so that these conditions can then be carried forward when fitting the state space model in chapter 7. Conditions include:

- Pump selection: Using only pre-selected households with "typical" pump profiles (m = 17; "selected"=1) vs. using all households without PV (m = 33; "selected"=0); see also Table 4.
- Clusters: Fitting pumps in clusters ("cluster"=1) vs. all together ("cluster"=0).
- Data reduction: Fitting only the seasonal components ($y_{\text{daily\_mean}} + y_{\text{profile\_mean}}$; "reduced"=1) or the full data ("reduced"=0); see also Eq. 5.2 and Fig. 4.

A combination of conditions (for example: "selected" = 0 and "cluster"=1 and "reduced"=0, or 1|0|0 for short) is here referred to as an "experiment". The experiments performed in this chapter are shown in table 9. For household power consumption, linear model experiments are summarised in Appendix A, Table 47.

The following metrics are used in this chapter to evaluate the goodness of fits:

- MAE(j) = median($|r_j|$) for $j = 1, \ldots, m$
- MSE(j) = $\frac{1}{8783} ||r_j||_2^2$ for $j = 1, \ldots, m$
- $|\overline{r}|$(j) = $|\overline{r}_j|$ (absolute mean of residuals) for $j = 1, \ldots, m$
- s(j) = $\frac{\sum_{i=1}^{8783}(r_{j,i} - \overline{r}_j)^3}{8782 \cdot \widehat{\sigma}^3}$ (skewness of residuals) for $j = 1, \ldots, m$

where r_j are the residuals for the predictions made for pump j. The MAE (median absolute error) and MSE (mean squared error) are used to evaluate prediction errors; the absolute mean and skewness of the residuals are used to evaluate the shape of the residuals. Note that, even though there are always m or m_c (c: clustered) pumps fit simultaneously in an experiment, each metric is calculated individually for each pump. Sometimes, we then use the mean of only those 17 pumps which are included in all experiments, e.g. $\overline{\text{MAE}}_{17} := \frac{1}{17} \sum_{j=1}^{17} \text{MAE}(j)$, so that the outcomes of different conditions can be easily compared.

selected	cluster	reduced		m	$\overline{\text{MAE}}_{17}$	$\overline{\text{MSE}}_{17}$	$\overline{\lvert\overline{r}\rvert}_{17}$	$\overline{\text{s}}_{17}$
0	0	0		33	122.96	111679.95	40.19	1.20
0	0	1		33	114.02	60995.63	40.94	0.51
0	1	0	$\sum_{c=1}^{5} m_c = 33$		107.92	101691.17	43.61	1.39
0	1	1	$\sum_{c=1}^{5} m_c = 33$		98.73	51173.68	45.07	0.46
1	0	0		17	108.72	100975.00	45.79	1.39
1	0	1		17	89.50	51022.66	45.64	0.31
1	1	0	$\sum_{c=1}^{4} m_c = 17$		98.91	98481.67	44.13	1.44
1	1	1	$\sum_{c=1}^{4} m_c = 17$		85.32	46612.71	45.27	0.36

Table 9: Linear model experiments. Each row corresponds to a combination of experimental conditions: "selected": all pumps are included (0) vs. only selected pumps are included (1); "cluster": all included pumps are fit simultaneously (0) vs. only pumps of the same cluster are fit simultaneously (1); "reduced": all data is used for fitting (0) vs. only reduced (seasonal) data is used (1). m: the number of pumps included in the experiment; m_c: the number of pumps in cluster c. There are five clusters, but if only selected pumps are included, then the fifth cluster is empty. The MAE (median absolute error), MSE (mean squared error), $|\overline{r}|$ (absolute mean of residuals) and s (skewness of residuals) are calculated separately for each pump, but then for each experiment, the mean outcomes of those 17 pumps which are included in all experiments is calculated, so that the outcomes of different experiments can be compared directly.

We first perform forward selection for each "experiment". Forward selection starts with an intercept model, which is fit on 2019 data and used to predict 2020 data (for all time points at once). The mean median absolute error ($\overline{\text{MAE}}$; mean across all included pumps) on 2020 predictions is used to assess the goodness of the fit. (The MAE rather than the MSE is used because there are many outliers in the data and the MAE is more robust.) In each round of forward selection, we then iterate through all predictor variables in the dataset, perform a fit on 2019 data including the variable, use the fit to predict 2020 data and calculate the $\overline{\text{MAE}}$ for those predictions (mean across all included pumps). At the end of each selection round, the feature which most reduces the $\overline{\text{MAE}}$ is selected permanently. This process usually continues until no feature can be found which reduces the $\overline{\text{MAE}}$ even further. For the dataset used here, only four selection rounds are necessary; further rounds only reduce the MAE minimally (Fig. 13). However, it is clear that some conditions (for example: "selected"=0, including all pumps) have a significantly worse $\overline{\text{MAE}}$, which may well affect which features are selected. This is a first indication that it may be better to exclude "atypical" pumps.

Tables 10 and 11 show which features are commonly selected across experiments. The same information is also available for household power consumption (Appendix A, Tables 25 and 26). Unsurprisingly, most selected features are based on temperature profiles or load profiles. Table 11 is useful because features which are not in this table are never selected and therefore likely irrelevant. This information can be used to speed up the Kalman filter analysis in chapter 7. Note that the effects of selected features is discussed later in this chapter, when a preferred model has been chosen.

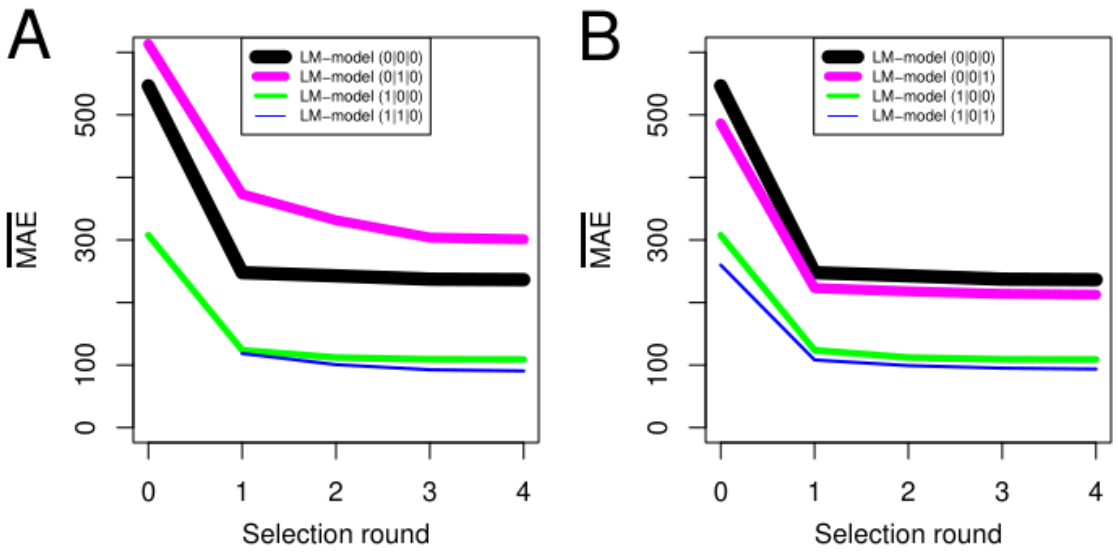


Figure 13: Model selection for different linear model experiments. Only the most relevant experiments are shown. The cluster example is based on cluster 2 (Table 5). $\overline{\text{MAE}}$ is the mean of the MAEs of all pumps included in an experiment.

Variable	Frequency
T15_difference_season_profile_pump_power_at_zero	6
T18_difference_season_profile_sqmetres_pump_power_at_zero	3
T15_difference_season_profile_sqmetres_pump_power_at_zero	2
T18_difference_winter_profile_sqmetres_pump_power_at_zero	2
T15_difference_season_profile	1
T15_difference_season_profile_sqmetres	1
T15_difference_winter_profile_pump_power_at_zero	1
T18_difference_season_profile	1
winter_profile_pump_power_at_zero_T15	1
winter_profile_sqmetres_T15	1
winter_profile2_sqmetres_pump_power_at_zero	1

Table 10: Features selected during the first round of forward selection in at least one experiment, together with the frequency with which they are selected across experiments. Note that there are 4 un-clustered experiments and 4 clustered experiments (Table 9; =2 clustered fits of five clusters for all pumps + 2 clustered fits of 4 clusters for selected pumps; one cluster is empty when "atypical" pumps are excluded), so there are 20 fits with model selection in total.

Variable	Frequency
T15_difference_season_profile_pump_power_at_zero	6
T18_difference_winter_profile_sqmetres_pump_power_at_zero	5
day_profile_pump_power_at_zero_winter_profile_T15	4
T18_difference_season_profile_sqmetres_pump_power_at_zero	4
T15_difference_season_profile_sqmetres_pump_power_at_zero	3
T15_difference_winter_profile_sqmetres_pump_power_at_zero	3
T18_difference_season_profile	3
day_profile_pump_sqmetres_T15_difference	2
T15_difference_persons	2
T15_difference_pump_power_at_zero	2
T15_difference_season_profile	2
T15_difference_sqmetres	2
T15_difference_winter_profile	2
T15_difference2_sqmetres_pump_power_at_zero	2
temperature_T15_winter_profile	2
winter_profile_sqmetres_T15	2
winter_profile2_pump_power_at_zero	2
Autumn	1
day_profile_pump_power_at_zero_season_profile	1
day_profile_pump_power_at_zero_season_profile_T15	1
day_profile_pump_power_at_zero_sunaltitude	1
day_profile_pump_power_at_zero_T15_difference	1
day_profile_pump_power_at_zero_T18_difference	1
day_profile_pump_power_at_zero_winter_profile	1
day_profile_pump_sqmetres_sunaltitude_T15	1
day_profile_pump_sqmetres_T18_difference	1
day_profile_pump_sqmetres_winter_profile	1
day_profile_pump_sqmetres_winter_profile_T15	1
day_profile_pump_T15_difference	1
season_profile2_sqmetres	1
season_profile2_sqmetres_pump_power_at_zero	1
summer_profile	1
T15_difference	1
T15_difference_season_profile_sqmetres	1
T15_difference_winter_profile_pump_power_at_zero	1
T15_difference2_pump_power_at_zero	1
T18_difference_winter_profile	1
T18_difference_winter_profile_pump_power_at_zero	1
T18_difference_winter_profile_sqmetres	1
T18_difference2_sqmetres	1
temperature_kelvin_T15_sqmetres	1
temperature_sqmetres	1
Winter	1
winter_profile_pump_power_at_zero_T15	1
winter_profile_sqmetres	1
winter_profile_sqmetres_T18	1
winter_profile2	1
winter_profile2_sqmetres_pump_power_at_zero	1

Table 11: Features selected at least once in four rounds of forward selection in any experiment, together with the frequency with which they are selected across experiments.

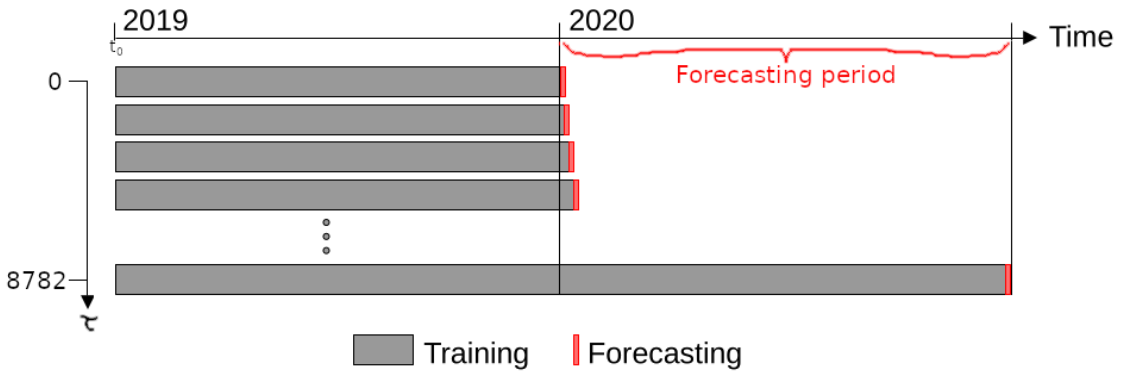


Figure 14: For linear model experiments, fits are performed on an expanding training window, t_0 = 2019-01-01 00:00:00 to t = 2020-01-01 00:00:00+τ, for $\tau = 0, ..., 8782$ hours (8783 fits in total). Predictions and evaluations are performed on the time period given by the sequence (2020-01-01 00:01:00+τ)$_\tau$.

Forward selection is only used to determine the features to be used in each experiment. To make comparison with a Kalman filter approach possible later (see chapter 7), for each experiment we then use the features selected for that experiment to fit the data using an expanding training window, including data from t_0 = 2019-01-01 00:00:00 to t = 2020-01-01 00:00:00+τ, for $\tau = 0, ..., 8782$ hours (Fig. 14; $366 \cdot 24 - 1 = 8783$ hours in 2020 are included in the analysis). For each τ, we obtain an estimate for θ and then calculate the 1-step-ahead prediction, i.e the prediction for the time point 2020-01-01 00:00:00 $+\tau + 1$ hour. This means that for each experiment, there are 8783 OLS estimates and 1-step-ahead predictions for all the time points in 2020, which can be used to evaluate the experiment. To simplify the evaluation, we define the prediction residuals for each pump included in an experiment as a vector given by the sequence $r_{pump} := (y_{pump}(t+1) - \widehat{y}_{pump}(t+1))_t$, where t = 2020-01-01 00:00:00+τ, for $\tau = 0, ..., 8782$ hours.

For the 17 pumps included in all experiments, one-sided, paired t-tests of the evaluation metrics are performed with null hypothesis $H_0 : \mu_M \leq 0$, where $\mu_M = \mathbb{E}(M(C = 0) - M(C = 1))$. The null hypothesis is rejected if $M(C = 1)$ is significantly smaller than $M(C = 0)$, i.e. if the fit improves under $C = 1$ compared to $C = 0$ as measured by M (for test details see section 4.5). The test results are shown in Fig. 15.

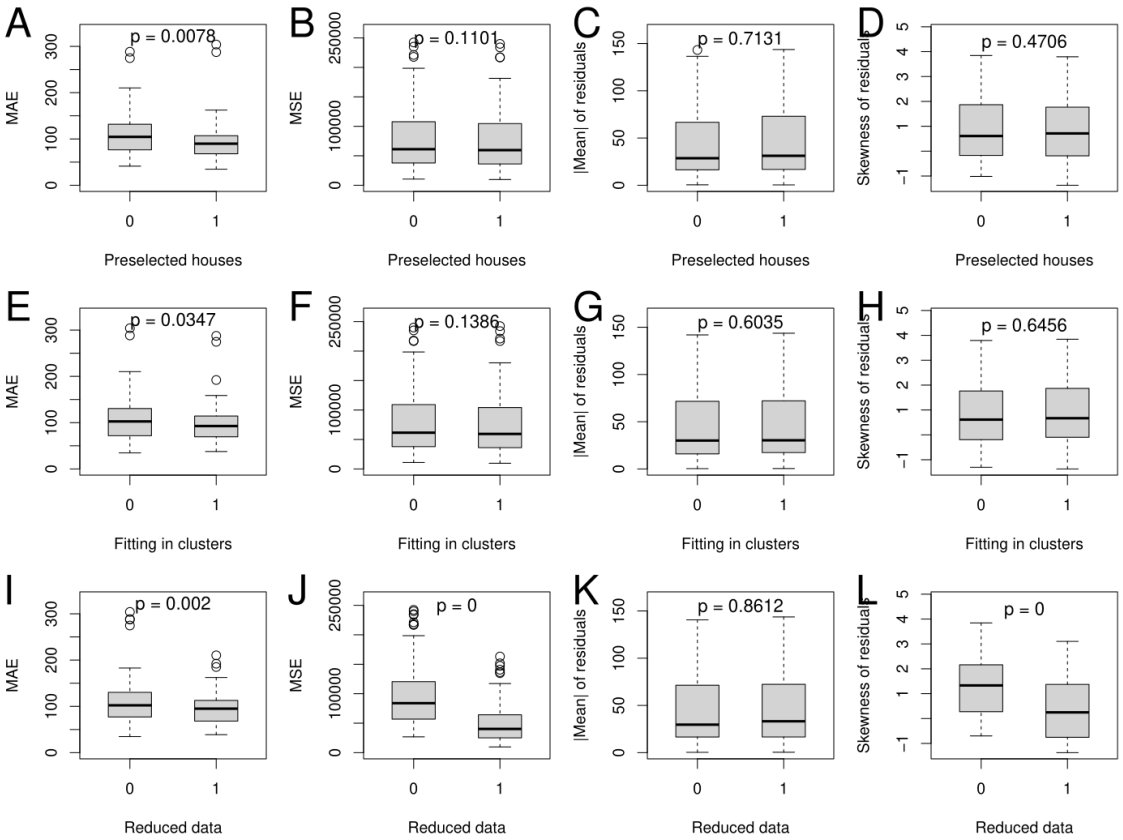


Figure 15: Effects of different fit conditions on MAE (median absolute error), MSE (mean squared error), $|\overline{r}|$ (absolute mean of residuals) and s (skewness). A-D (first row): effect of pump selection; E-H (second row): effect of fitting in clusters; I-L (third row): effect of using only reduced (seasonal) data. Paired t-tests are performed for each outcome, only including those 17 pumps which are included in all experiments. Since there are four experiments for each condition (Table 9), $n = 68$ for each test (see section 4.5).

Selecting only "typical" pumps to be fit simultaneously improves the MAE for those 17 pumps compared to when all pumps are fit simultaneously (Fig. 15A; $p = 0.0078$). In other words, including "atypical" pumps in a simultaneous fit of all pumps worsens predictions for the "typical" pumps. That could be an indication that "typical" and "atypical" pumps are best described by different features and should therefore not be fit simultaneously. However, MSE, $|\overline{r}|$ and s of the "typical" pumps do not improve significantly when "atypical" pumps are excluded (Fig. 15B-D). Similar effects are observed when fitting in clusters, though the reduction in MAE is less strong (Fig. 15E-H; Table 9). Using only seasonal data for the fits reduces prediction errors (MAE: $p = 0.002$; MSE: $p = 0$) and reduces the skewness of residuals (s: $p = 0$). However, in this case, reduction of MAE and MSE is expected because much of the noise in the data is removed before the fit (Fig. 4). Similarly, because asymmetric peaks are removed from the profiles before the fit, more symmetric (less skewed) residuals are also an expected result. Nevertheless, this results suggests that reduction of the data to the seasonal component is a good idea - **if** one is only interested in the seasonal component because, of course, while outcomes improve, all other information that may be contained in the noise is lost.

The t-test results are interesting, but an in-depth analysis of predictions and residuals paints a more nuanced picture. Fig. 16 shows examples for predictions in different linear model experiments. Household 3 is used as an example because it has little missing data and a "typical" pump profile (Table 4), so its predictions should only be affected by the fit conditions. There is no big difference between predictions in different experiments (Fig 16A-J), except when reduced data is used (Fig 16K-L). In that case, no "day_profile"-feature is selected and the fit cannot reproduce the daily variation in the data, indicating that perhaps, in this case more than four rounds of feature selection are required. Note that this behaviour is not captured by the MAE because it is robust with respect to peaks and troughs. In this case, the MSE is actually a better measure, although the MSE is also smaller for reduced data (Table 9) because much of the noise has been removed beforehand.

Fig. 17 shows the residual analysis for the pump of household 3 in the 1|0|0 experiment ("selected"=1, "cluster"=0, "reduced"=0). This example is used here to provide an overview of the general residual structure before different experiments are compared. It is clear that residuals are not homoscedastic (Fig. 17A, B), simply because pumps are much less active in the summer than in the winter. Accordingly, summer residuals are much smaller, more symmetric and more "normal" than winter residuals (Fig. 17B, C). Winter residuals still have a daily pattern (Fig. 17D), i.e. the fit does not capture the daily variation in the data very well. For the same reason, a seasonality of 24 hours is still observed in the ACF and PACF of residuals (Fig. 17E, F). The ACF and PACF also show that there is other structure in the noise, like at least two AR components (Fig. 17E, F). This is not unexpected, because even though load profiles are included as a feature, these are only the mean cluster load profiles of 2019 by season and weekday/weekend, whereas predictions and residuals are analysed in the 2020 period. Individual pumps may deviate a little bit from the mean cluster profile. Furthermore, 2019 was a pre-Covid year, while in 2020 there were many lockdown periods, which may have led to changes in pump usage patterns.

Fig. 18A shows that including only "typical" pump profiles or fitting in clusters does not make residuals more "normal". There is actually no reason why it should, since "typical" pumps also have asymmetric peaks. What is more interesting is that even in "reduced" conditions, where noise is removed before fitting, the residuals are still not "normal" (Fig. 18B). Thus, fitting only the seasonal components of the data does not actually improve residual structure. However, this may be partly due to the fact that in this experiment, no "day_profile" feature is selected during forward feature selection (Fig. 16K, L). Similarly, residuals still have daily pattern in all experiments (Fig. 19A-F). These results suggest that selecting, clustering or reducing data does not improve residual structure.

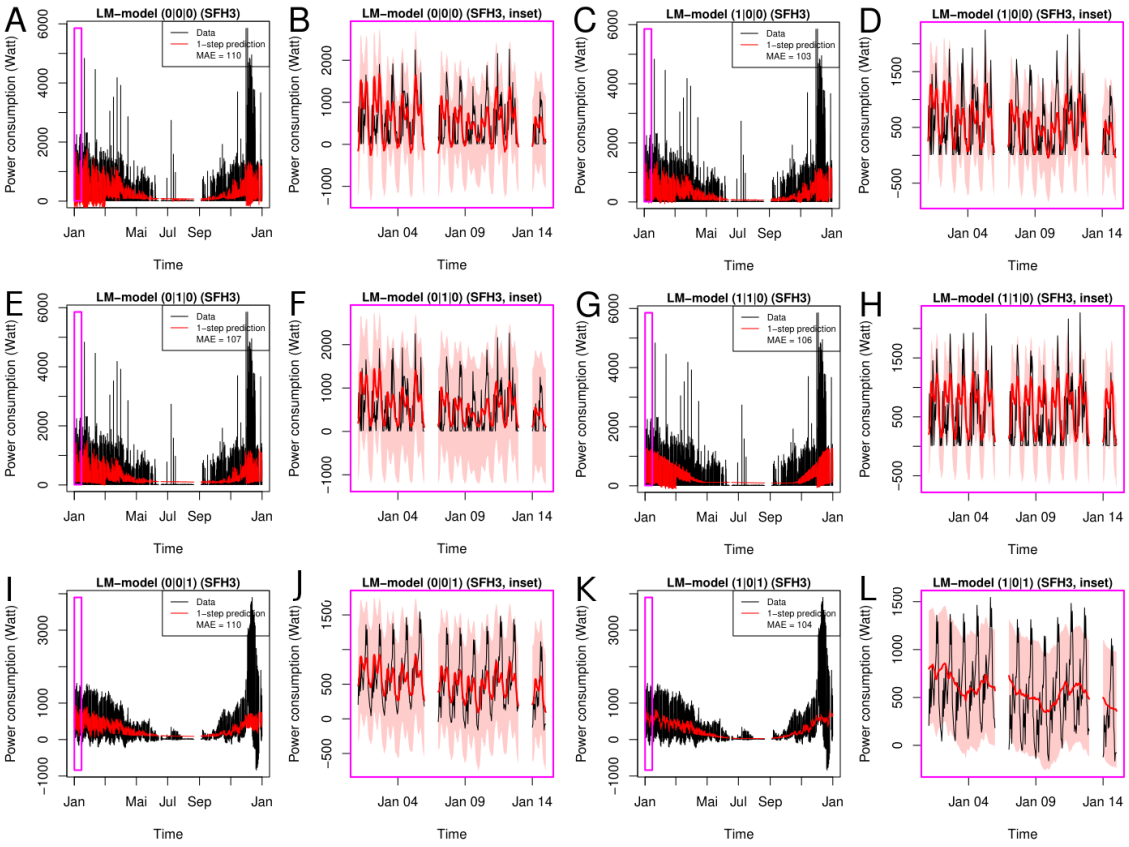


Figure 16: Comparison of 1-step-ahead predictions for the pump of household SFH3 for different linear model experiments with a simultaneous fit of m pumps. Only the most relevant experiments are shown. The prediction intervals given are based on ± two standard deviations, based on the diagonal elements of the covariance matrix of the 1-step ahead prediction described in 6.5. No distribution assumption is made, so the interval shown does not (necessarily) correspond to a 95%-prediction interval.

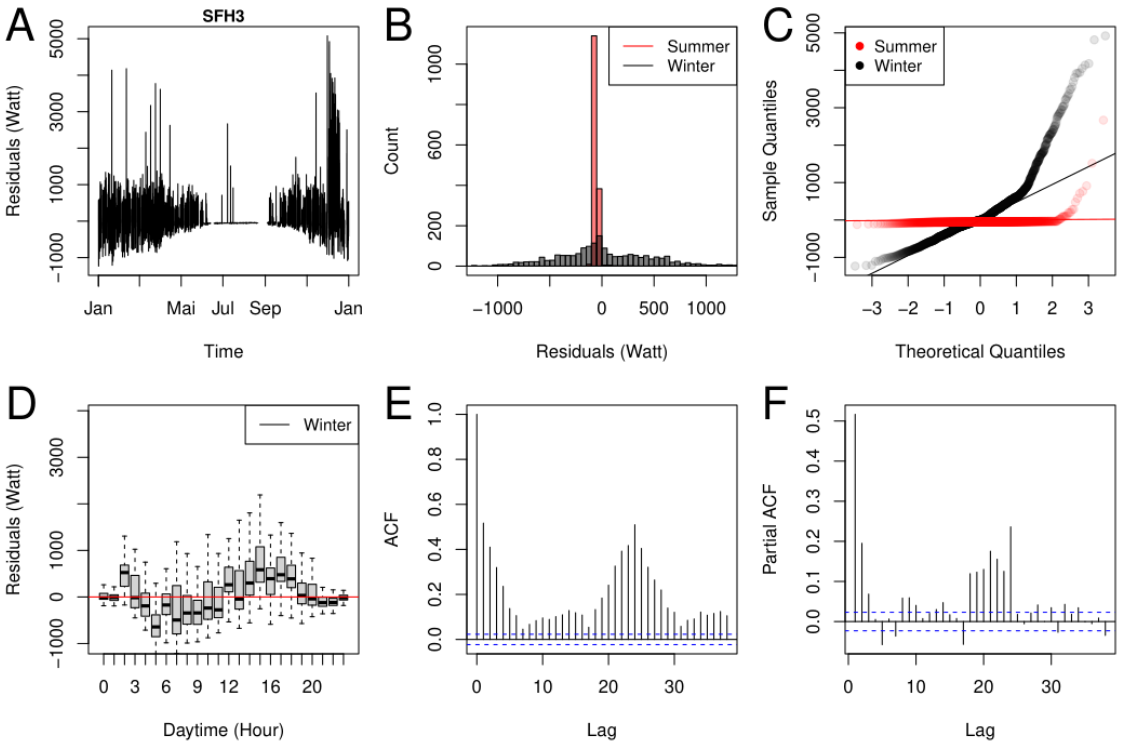


Figure 17: Residual analysis in the 1|0|0 linear model experiment. Note that in this experiment, 17 pumps are fit simultaneously, but for clarity, the residual analysis of only one of these pumps (SFH3) is shown.

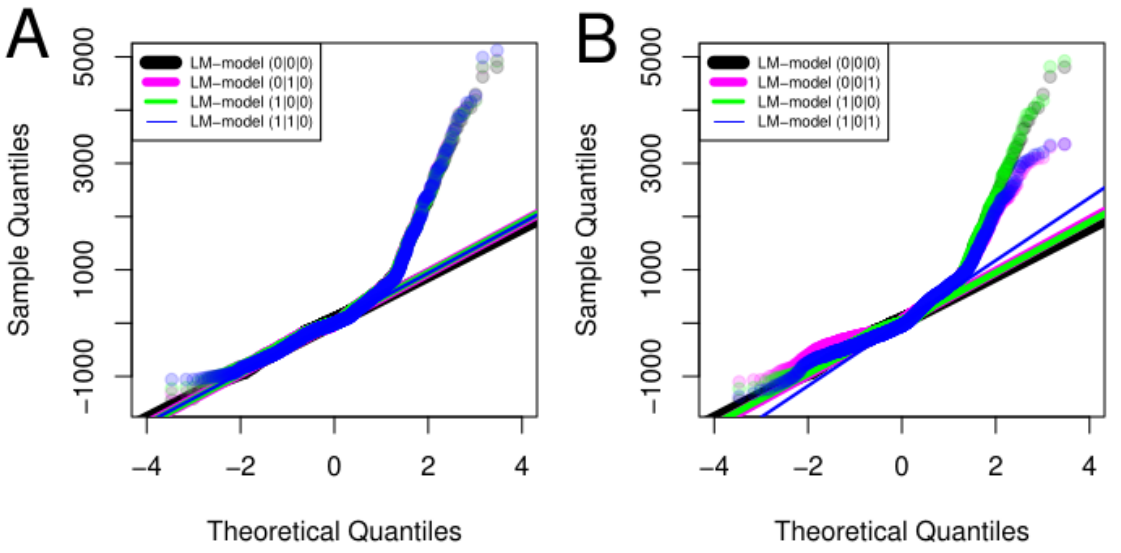


Figure 18: Comparison of the normality of winter residuals in different experiments. Only the most relevant experiments are shown.

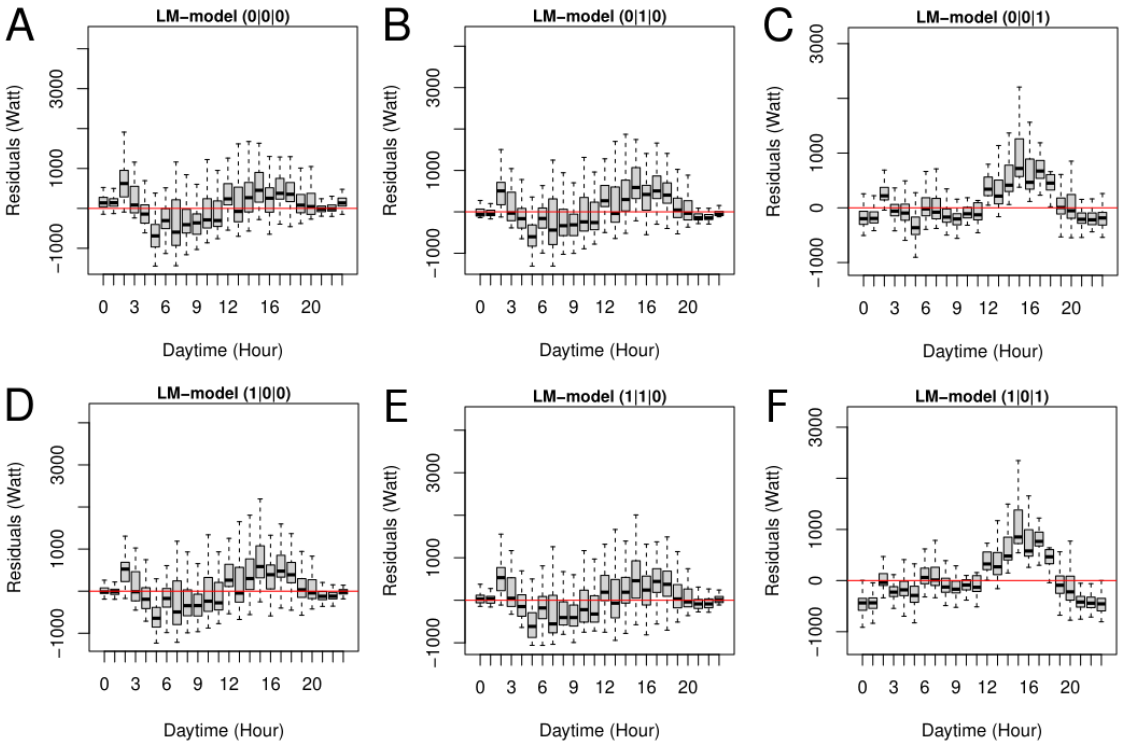


Figure 19: Comparison of patterns of winter residuals in different fit conditions.

In summary, there is a case for including only "selected", "typical" pump profiles in fits: "atypical" pumps have larger MAEs and may therefore distort the mean MAE that is used as a selection criterion during forward selection (Fig. 13), so excluding them may lead to better / more targeted feature selection. The same is not true for clustering, which has only a small effect on feature selection. A cluster may even have a larger MAE than a non-clustered fit (Fig. 13A), perhaps because fitting fewer pumps simultaneously can lead to less generalisable models. However, clustering does improve the $\overline{\text{MAE}}_1 7$ of "typical" pumps overall, so it is also a valid choice (Table 9). Reducing the data to seasonal components is a valid choice if one is only interested in the seasonal components, however, reductions in MAE and MSE are probably only down to the removal of noise, and residual structure is not improved (Fig. 18, 19).

Even though the approach taken in this chapter may in hindsight seem unnecessarily complicated, the process actually helped make the following decisions clear. Taking all the above considerations into account, we will proceed with "typical" pumps ("selected"=1), no clustering ("selected"=0), and no data reduction ("reduced"=0). We decide not to cluster because it is faster to fit all pumps simultaneously than to fit five clusters separately, and clustering brings no huge improvement. We decide not to reduce the data to seasonal components because in the next chapter, we will try to also explain some of the structure in the noise (chapter 7). Other decisions are definitely possible and valid.

In the last section of this chapter, the OLS estimator and the effects of selected features are discussed for the 1|0|0 linear model ("selected"=1, "cluster"=0, "re-

duced"=0). The following features are selected in this experiment:

- T15_difference_season_profile_pump_power_at_zero,
- T18_difference_winter_profile_sqmetres,
- day_profile_pump_sqmetres_T18_difference and
- day_profile_pump_sqmetres_sunaltitude_T15

Note that the first two features describe the long-term seasonality with an interaction of temperature and season profiles (see also chapter 5). Recall that season and winter profile are derived from the mean daily temperature with slightly different profile shapes. So the interaction is similar to a quadratic effect. The features are modified by an interaction with household-specific features (pump efficiency and building size). This allows the profiles to adapt a little bit to each individual pump in the simultaneous fit. Using two features of this type allows the model to form a weighted sum of the two which may adapt better to the real shape.

The last two features are load profile features that describe variation during the day. The interactions with household-specific features and, even more strangely, temperature or sun altitude are odd in this case because the load profiles are based on real data from 2019. However, they only include mean cluster profiles and 2020 may have different usage patterns for Covid-related reasons, so these adaptations may be due to that.

Fig. 20A shows the OLS estimators for the intercepts of the 17 pumps included in the experiment. The estimators for the four selected features, are shown in Fig. 20B-E - for these features, the estimator values are the same for all pumps. Note that there are 8783 estimates, one for every time point in 2020 (see Fig. 14), so the estimators are plotted over time. If the selected model is close to the true, unknown data-generating process, then the $\widehat{\theta}$ estimates should be constant over time. They are not. Perhaps this is not too surprising: the data is very noisy, so adding a single data point to the training dataset in the expanding training window can change the fit parameters substantially. They show definite patterns, however, which suggests that there is one or more unknown features missing in the dataset that are still needed to explain the data. Note that constant stretches during summer time may simply be due to the fact that data points added during that period have feature values of zero due to temperatures larger than 15 or 18°C. However, the estimator values are at least very far from zero (Fig. 20B-E), suggesting that the selected features have significant effects. (Even though significance cannot be shown here because no distribution assumption has been made.)

Because the estimators are not constant in time, it is difficult to table feature effects in the traditional way. Instead, Fig. 21 visualises the effect of each feature by plotting the contribution of each feature to the model over time in a cumulative fashion, using the pump of household 3 as an example. The contribution of a feature is defined as $x_{\mathrm{t,feature}} \cdot \widehat{\theta}_{\mathrm{t,feature}}$. Fig. 21A shows the effect of the intercept and the first selected feature. Fig. 21B shows the effect of the intercept and the first two selected features. This combination already captures much of the long-term seasonal profile. Adding the last two features (Fig. 21C and D) adds much "noise" in the form of daily pump usage patterns.

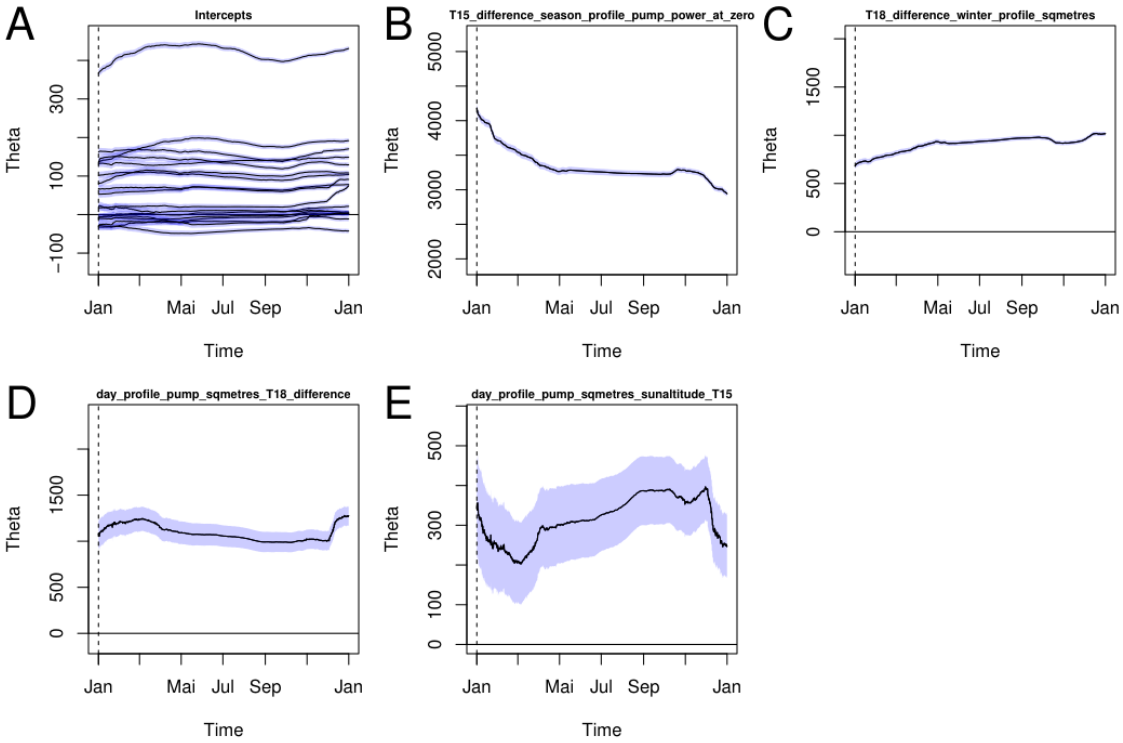


Figure 20: OLS estimators for the 1|0|0 linear model experiment. A, intercepts. Note that there are 17 intercepts because 17 pumps are fit simultaneously (Eq. 6.3). B, OLS estimator for T15_difference_season_profile_pump_power_at_zero; C, OLS estimator for T18_difference_winter_profile_sqmetres; D, OLS estimator for day_profile_pump_sqmetres_T18_difference; E, OLS estimator for day_profile_pump_sqmetres_sunaltitude_T15. An interval of ± two standard deviations is given. This is based on the diagonal elements of the covariance matrix in Eq. 6.3. No distribution assumption is made, so these intervals are not equivalent to 95%-confidence intervals.

Note that the first two features already produce quite a "noisy" profile because they are linked to temperature and thus reflect hourly variation in the temperature. We also tried excluding features with hourly temperature values from the fits completely (not shown), thinking that the temperature variation may interfere with the load profile features, in the sense that the temperature variation already covers some of the pump power consumption variation that should actually be covered by a load profile feature instead. However excluding features with hourly changes in temperature did not improve the fit (not shown).

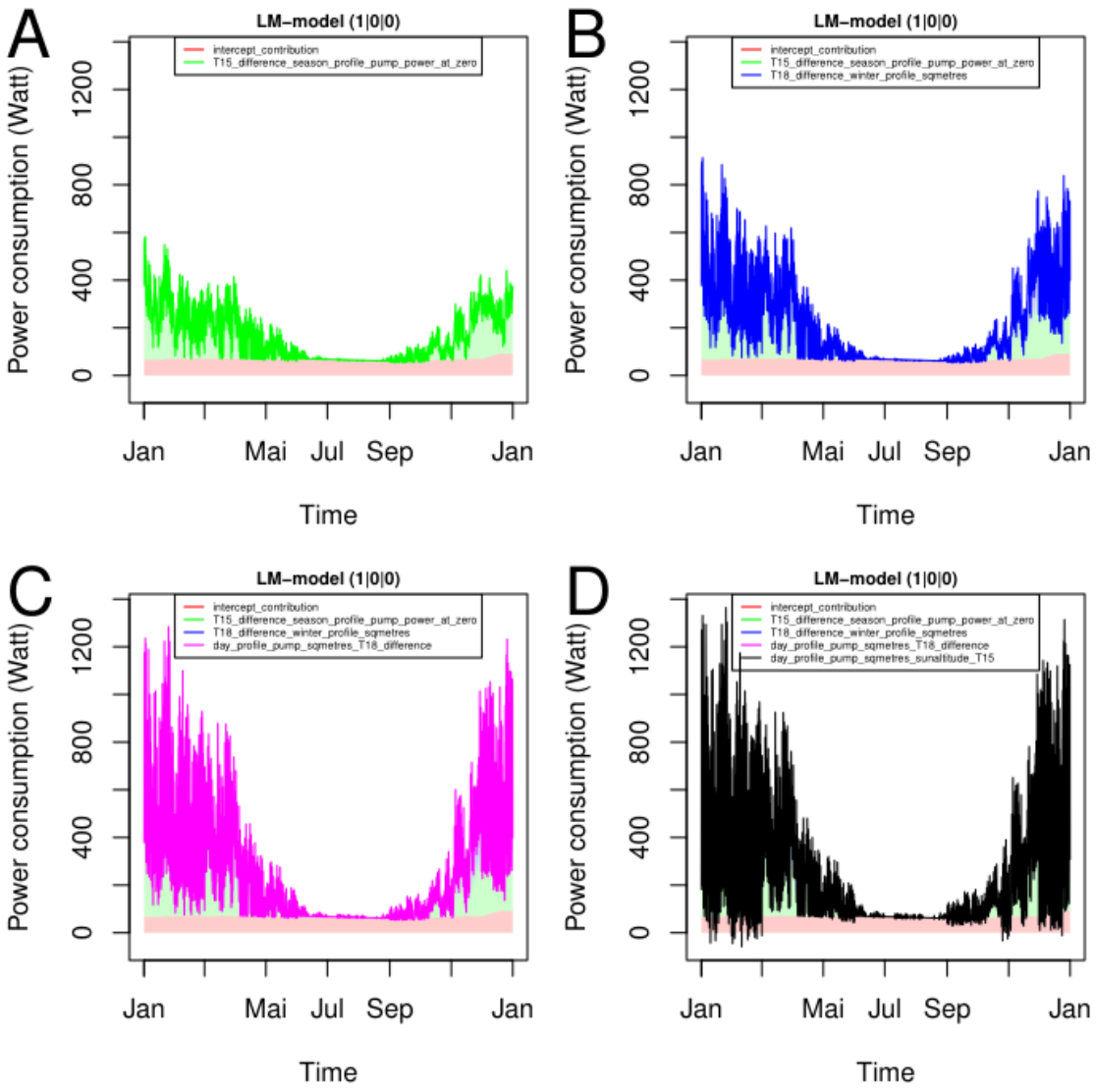


Figure 21: Contribution of selected features to the model in the 1|0|0 linear model experiment. A, contribution of the intercept (red) and the T15_difference_season_profile_pump_power_at_zero feature (green); B, adding the T18_difference_winter_profile_sqmetres feature (blue); C, adding the day_profile_pump_sqmetres_T18_difference feature (magenta); D, adding the day_profile_pump_sqmetres_sunaltitude_T15 feature (black).

7 State space model

We now treat heat pump power consumption, $(Y_t)_{t\in\mathbb{N}}$, as an observable time series in $\mathbb{R}^m$ which only depends on latent states described by $(\theta_t)_t$ and $(\mu_t)_t$. In this case, heat pump power consumption can be described using the following state space model (Petris et al. 2009):

$$\begin{aligned} Y_t &= F_t\theta_t + \mu_t \\ \theta_t &= G_t\theta_{t-1} + w_t \\ \mu_t &= \phi_1\mu_{t-1} + \phi_2\mu_{t-2} + v_t \end{aligned} \tag{7.1}$$

where t runs from 2019-01-01 00:00:00 to 2020-12-31 23:00:00 in steps of 1 hour and

$$Y_t = \begin{bmatrix} Y_{t,1} \\ \vdots \\ Y_{t,m} \end{bmatrix} \in \mathbb{R}^m,\ F_t = \begin{bmatrix} I_{mxm} & X_{mxp}(t) \end{bmatrix} \in \mathbb{R}^{mx(m+p)}, G_t = I_{(m+p)x(m+p)},$$

$$\phi_1 \in (-1,1),\ \phi_2 \in (-1,1),$$

$$\theta_t = \begin{bmatrix} \alpha_{t,1} \\ \vdots \\ \alpha_{t,m} \\ \beta_{t,1} \\ \vdots \\ \beta_{t,p} \end{bmatrix} \in \mathbb{R}^{m+p} \text{ and } \theta_0 \sim N_{m+p}(0_{m+p}, C_0), \tag{7.2}$$

$$\mu_t = \begin{bmatrix} \mu_{t,1} \\ \vdots \\ \mu_{t,m} \end{bmatrix} \in \mathbb{R}^m \text{ and } \mu_0 \sim N_m(0_m, Z_0),$$

$$\text{with } C_0 = \text{diag}_{m+p}(10^7, \ldots, 10^7) \text{ and } Z_0 = \text{diag}_m(10^7, \ldots, 10^7).$$

The sequence $(\theta_t)_{t\in\mathbb{N}_0}$ is assumed to be a Markov series in $\mathbb{R}^p$, where α contains the intercepts of m pumps and β the feature effects for p features contained in the columns of X (same notation as in chapter 6). In chapter 6, we show that $m = 17$ pumps can be fit simultaneously and only $p = 4$ features are required to describe the data well.

The sequence $(\mu_t)_{t\in\mathbb{N}_0}$ with $\mu_t = Y_t - F_t\theta_t$ describes the (latent) stationary time series remaining when the trends and seasonalities described by $F_t\theta_t$ are removed from Y_t (see also Fig. 4). $\mathbb{E}[\mu_t]$ is assumed to be 0, so that $\mathbb{E}[Y_t|\theta_t] = F_t\theta_t$. Note that the equation for μ_t (7.1) describes an AR(2) process, but this is just a first, simple assumption.

Finally, the error terms are assumed to follow multivariate normal distributions:

$$\begin{aligned} &v_t \sim N_m(0_m, V_t) \text{ with } V_t = V = \sigma^2_{\text{pump}} I_{mxm} \text{ and} \\ &w_t \sim N_{m+p}(0_{m+p}, W_t) \text{ with } W_t = W = \text{diag}_{m+p}(\sigma^2_{\alpha,1}, \ldots, \sigma^2_{\alpha,m}, \sigma^2_{\beta,1}, \ldots, \sigma^2_{\beta,p}). \end{aligned} \tag{7.3}$$

The state space model in Eq. 7.1 has been purposefully constructed in such a way that all the models discussed in this chapter build on each other (Table 12). To start

with, the V-model assumes that $\phi_1 = \phi_2 = 0$ and that $\sigma^2_{\alpha,1} = \ldots = \sigma^2_{\alpha,m} = \sigma^2_{\beta,1} = \ldots = \sigma^2_{\beta,p} = 0$ (therefore $Y_t = F_t\theta + v_t$). Thus, the V-model is a static linear model and the main difference to chapter 6 is the use of the Kalman filter for estimation (Petris et al. 2009; see section 4.7.1). Only one variance has to be optimised for this model type (σ^2_{pump}), which is still comparatively fast. Therefore, for the V-model we also perform four rounds of forward selection on the features of Table 11 to determine the best 4 predictor variables to use in this chapter. Next, the W-model extends the V-model by allowing θ to have a distribution, with $\sigma^2_{\alpha,1}, \ldots, \sigma^2_{\alpha,m}, \sigma^2_{\beta,1}, \ldots, \sigma^2_{\beta,p} > 0$ (Table 12). This corresponds to a dynamic linear model (Petris et al. 2009). No forward selection is performed here because now 22 variances need to be optimised, which takes a long time (see section 7.2). The P-model is the same as the W-model but uses particle filter estimation instead of Kalman filter estimation (Chen et al. 2004; Dahlin and Schön 2019; see section 4.7.2). Finally, the A-model uses the full model described in Eq. 7.1, including autoregressive noise. Different approaches are of course possible (e.g. SARIMAX-models), but these are not discussed here.

Model name	Simplifications	Estimation method	Effects
V-model	$\phi_1 = \phi_2 = 0$ $W = 0_{(m+p)x(m+p)}$	Kalman filter	Assuming Gaussian white noise and constant θ (similar to LM).
W-model	$\phi_1 = \phi_2 = 0$	Kalman filter	Assuming only Gaussian white noise. θ now has a distribution and can vary more.
P-model	Same as W-model	Particle filter	Effect of different estimation method.
A-model	$\phi_1 \in (-1, 1)$ and $\phi_2 \in (-1, 1)$	Kalman filter	Effect of autoregressive noise.

Table 12: Simplifications of the state space model described in Eq. 7.1.

In the following sections, V-, W-, P- and A-models are discussed separately, followed by a direct comparison of the best models from each category in section 7.5.

	V (MAE)	V (MAE_Q90)	V (MAE_I2SD_Q90)	W	W (I0)	W (F0)	A	A (+)
σ^2_{pump}	1000000	0.000394	100282.6	100281.999421	100281.999845	100281.999991	100281.999983	100281.999974
$\sigma^2_{\alpha,1}$				1.015163	0	2.577791	1.045823	1.075869
$\sigma^2_{\alpha,2}$				1.011326	0	1.213115	0.934970	0.893348
$\sigma^2_{\alpha,3}$				1.029908	0	1.467405	1.182054	1.286944
$\sigma^2_{\alpha,4}$				0.999150	0	1.081836	0.793606	0.685980
$\sigma^2_{\alpha,5}$				0.994385	0	1.013149	0.827437	0.718173
$\sigma^2_{\alpha,6}$				1.005898	0	1.685200	1.038117	1.062203
$\sigma^2_{\alpha,7}$				0.973658	0	0.000000	0.455367	0.130865
$\sigma^2_{\alpha,8}$				1.005757	0	1.018199	0.941601	0.908465
$\sigma^2_{\alpha,9}$				1.019647	0	0.415401	0.875419	0.810874
$\sigma^2_{\alpha,10}$				1.023965	0	1.870225	1.744559	2.197428
$\sigma^2_{\alpha,11}$				1.005329	0	0.915891	1.183200	1.296882
$\sigma^2_{\alpha,12}$				1.016609	0	1.371562	0.670738	0.507946
$\sigma^2_{\alpha,13}$				0.999994	0	0.858268	1.050405	1.083880
$\sigma^2_{\alpha,14}$				1.000780	0	0.912704	0.684084	0.515346
$\sigma^2_{\alpha,15}$				1.010383	0	0.001154	1.092611	1.158760
$\sigma^2_{\alpha,16}$				0.992137	0	3.413601	1.130336	1.209185
$\sigma^2_{\alpha,17}$				1.018701	0	0.785319	0.889871	0.806919
$\sigma^2_{\beta,1}$				1.035839	7.587610	0	0.893560	0.817043
$\sigma^2_{\beta,2}$				1.017378	3.549609	0	1.184449	1.281179
$\sigma^2_{\beta,3}$				0.997226	2.091091	0	0.684916	0.456805
$\sigma^2_{\beta,4}$				0.987317	1.150874	0	0.569645	0.297151
ϕ_1							-0.507544	0.012122
ϕ_2							0.539784	0.890863

Table 13: Optimised model parameters for different models.

$\widehat{\overline{\theta}}$\Model	LM	V (MAE)	V (MAE_Q90)	V (MAE_I2SD_Q90)	W	W (I0)	W (F0)	P (10^3)	P (10^4)	A	A (+)
Intercept SFH10	178.65±6.82	178.59±18.22	178.6±0	178.6±5.72	234.17±36.81	172.9±5.82	177.41±45.19	220.66±36.71	204.81±36.74	233.67±37.52	256.12±116.62
Intercept SFH12	116.53±7.31	120.71±18.37	120.72±0	120.72±5.77	194.14±37.02	109.23±5.89	137.32±37.59	170.26±36.92	158.46±36.95	190.75±36.74	210.22±111.66
Intercept SFH16	46.71±7.18	49.14±18.17	49.14±0	49.14±5.7	98.97±37.03	36.98±5.85	38.38±39.29	94.01±36.92	75.04±36.96	101.12±38.74	117.15±121.96
Intercept SFH18	79.38±7.28	91.79±18.61	91.8±0	91.8±5.84	199.84±37.3	71.09±6.02	141.37±36.65	181.66±37.19	161.28±37.22	194.22±35.68	187.09±105.02
Intercept SFH19	-35.46±7.56	-31.33±19.82	-31.31±0	-31.31±6.22	46.03±37.14	-2.85±6.38	-0.58±36.02	24.4±37.05	18.11±37.08	47.95±35.98	109.76±105.94
Intercept SFH21	251.69±6.9	251.92±18.34	251.9±0	251.91±5.75	217.27±37.23	222.93±6.04	150.24±40.72	221.07±37.14	195.4±37.15	219.66±37.95	244.18±116.62
Intercept SFH23	6±6.95	4.36±18.35	4.37±0	4.37±5.76	115.69±36.67	-6.75±5.89	43.31±6.36	92.72±36.56	80.12±36.61	110.57±30.96	106.17±73.9
Intercept SFH27	55.41±7.47	60.74±18.09	60.74±0.01	60.74±5.68	101.89±36.86	42.38±5.85	41.28±35.93	90.98±36.74	73.29±36.79	100.47±36.69	121.1±111.94
Intercept SFH29	109.14±8	103.01±18.12	103.01±0	103.01±5.68	93.75±37.11	77.06±5.93	29.99±28.81	91.43±37.01	68.8±37.05	93.84±36.2	122.4±108.96
Intercept SFH3	65.33±7.64	70.32±18.58	70.34±0	70.33±5.84	182.45±37.03	71.8±5.92	132.46±41.8	161.9±36.93	149.88±36.96	182.77±42.58	188.97±139.26
Intercept SFH30	9.96±6.79	5.19±18.19	5.2±0	5.2±5.7	26.2±36.97	-15.47±5.89	-34.14±35.05	16.06±36.89	-4.09±36.91	23.38±38.85	47.77±122.26
Intercept SFH32	99.75±7.41	98.92±18.79	98.94±0	98.93±5.9	202.91±36.95	108.59±5.99	151.62±38.7	184.08±36.86	170.71±36.87	202.15±33.85	230.1±97.4
Intercept SFH36	50.41±7.08	66.87±18.28	66.88±0	66.88±5.74	184.53±36.9	52.14±5.88	124.8±34.5	165.22±36.8	148.47±36.83	181.26±37.75	149.3±117.01
Intercept SFH38	158.42±7.62	136.77±19.37	136.76±0	136.76±6.08	221.24±37.56	140.98±6.19	162.06±35.08	224.43±37.44	206.57±37.49	229.32±34.8	275.92±98.42
Intercept SFH4	12.71±7.73	29.99±20.08	29.96±0	29.96±6.31	10.38±38.72	9.62±6.57	13.96±8.18	27.49±38.6	7.77±38.65	20.42±39.9	56.61±120.74
Intercept SFH7	139.97±7.38	132.92±18.44	132.94±0	132.93±5.79	211.62±36.61	135.05±5.87	148.34±48.45	190.79±36.51	180.89±36.55	210.88±38.25	263.32±120.05
Intercept SFH9	389.05±7.58	364.28±19.55	364.28±0	364.28±6.13	451.7±37.85	369.97±6.25	392.82±33.83	454.32±37.75	439.09±37.77	460.03±37.13	508.36±109.6
T15_difference _season_profile _sqmetres	579.79±39.93	592.07±103.31	591.83±0	591.87±32.43	976.58±109.33	1474.78±160.32	677.6±45.1	953.92±108.58	908.82±108.73	913.45±103.3	97.08±163.04
T18_difference _winter_profile _sqmetres _pump_power_at_zero	4659.69±75.26	4276.59±195.62	4277.3±0	4277.21±61.43	4431.47±205.33	2960.57±203.74	4708.32±87.16	3869.55±204.69	4383.31±204.78	4245.32±210.13	4459.04±461.41
T18_difference _winter_profile	285.09±22.51	356.69±57.87	356.69±0	356.68±18.16	-359.82±107.06	24.82±119.27	53.85±32.07	-394.72±105.43	-193.13±105.57	-290.75±95.15	-18.95±144.03
day_profile_pump _T15_difference	1104.73±27.88	1019.3±71.86	1019.32±0	1019.32±22.55	1008.42±97.2	973.1±102.48	1000.57±22.82	1026.03±97.07	1123.66±97.11	996.91±77.62	990.26±70.84

Table 14: $\widehat{\overline{\theta}}_i \pm 2\sqrt{\widehat{\overline{C}}(i,i)}$ for $i = 1, \ldots, 21$ in 2020 for different models.

7.1 V-model

The V-model is given by

$$\begin{aligned} Y_t &= F_t\theta_t + v_t \\ \theta_t &= \theta_{t-1} = \ldots = \theta_0 = \text{const.}, \end{aligned} \tag{7.4}$$

with the definitions of Y_t, F_t, θ_t and v_t given in Eqs. 7.2ff. The V-model is a linear model (Eq. 7.4 reduces to Eq. 6.1 in chapter 6), but now the Kalman filter is used for estimation (see section 4.7.1). The Kalman filter uses a Bayesian approach. This means that even though the true θ is constant in the V-model, for estimation, we start with a non-informative prior, $\widehat{\theta_0} \sim N_{m+p}(0_{m+p}, C_0)$ (with $C_0 = \text{diag}_{m+p}(10^7, \ldots, 10^7)$, see Eq. 7.2), which adapts to the data as estimation proceeds sequentially through the time points, from 2019-01-01 00:00:00 to 2020-12-31 23:00:00. In the end, the Kalman filter yields an estimate for θ for each time point (for more detail, see section 4.7.1):

$$\begin{aligned} c_t &:= \widehat{\mathbb{E}}[\theta_t|y_{1:t}] \\ C_t &:= \widehat{\text{Cov}}[\theta_t|y_{1:t}]. \end{aligned} \tag{7.5}$$

Note that C_t is not zero even though the true θ is constant because the estimator has a variance. The estimation procedure also yields k-step-ahead predictions:

$$\begin{aligned} f_{t+k} &:= \widehat{\mathbb{E}}[Y_{t+k}|y_{1:t}] \\ Q_{t+k} &:= \widehat{\text{Cov}}[Y_{t+k}|y_{1:t}] \end{aligned} \tag{7.6}$$

(unless noted otherwise, here, k = 1 hour). The diagonal of $Q_{t+1|t}$ yields m prediction variances for time point $t+1$, one for each of the m pumps fit simultaneously. Note that, like in chapter 6, estimation is performed on the whole data period, but we only analyse estimators, predictions and residuals in the year 2020. Data from 2019 is used as an adaptation period.

Another difference to the linear model of chapter 6 is that one now has to optimise the variance in the model specification. The V-model uses $v_t \sim N_m(0_m, V)$ with $V = \sigma^2_{\text{pump}} I_{mxm}$, so there is only one variance to optimise: σ^2. We use Nelder-Mead optimisation (R Core Team 2021, `R: optim(..., method="Nelder-Mead")`). The parameter σ^2_{pump} is systematically varied between 0 and 10^6 and its optimum value is determined using a combination of the following measures:

- $\overline{MAE}_{17}$: This is the mean MAE of the 1-step ahead predictions in 2020, across all pumps included in the fit (as defined in chapter 6). Note that, if $\overline{MAE}_{17}$ is minimal, it means that predictions are good, but Q is not constrained, so prediction intervals can be unexpectedly large or small.
- Q90_{17}: "Q90" is the 90%-quantile of the prediction variances estimated by the Kalman filter for the timepoints in 2020. The quantile is calculated for each individual pump included in the fit, and then the 90%-quantile of these quantiles is calculated to obtain Q90_{17}. Note that if Q90_{17} is minimal, Q is very small for all pumps, so prediction intervals are very small.

- $\overline{I2SD}_{17}$: "I2SD"(j) is a measure for the mean fraction of data points $y_{t+1}(j)$ contained in the intervals $[f_{t+1|t}(j) \pm 2\sqrt{Q_{t+1|t}(j,j)}]$ across time points in the year 2020 for the jth pump ($j = 1, \ldots, m$). $\overline{I2SD}_{17}$ is the mean of this measure across all 17 pumps included in the fit. Note that, if $\overline{I2SD}_{17}$ is close to 0.95, the prediction intervals on average contain 95% of the data and are neither too large nor too small.

The V-model is the first time in this thesis that optimisation becomes necessary. Therefore, for the V-model, we test different optimisation measures based on the above criteria. Three measures are tested: "MAE" (minimising $\overline{MAE}_{17}$), "MAE_Q90" (minimising $\overline{MAE}_{17} + \text{Q90}_{17}$) and "MAE_I2SD_Q90", an empirically devised function of the three:

$$\text{MAE\_I2SD\_Q90} = \begin{cases} \overline{MAE}_{17} + 100 \cdot (1 - \frac{5}{6}\overline{I2SD}_{17}) & \text{if } \overline{I2SD}_{17} < 0.9 \\ \overline{MAE}_{17} + 25 + \text{Q90}_{17} & \text{if } \overline{I2SD}_{17} = 1 \\ \overline{MAE}_{17} + (100 \cdot \overline{I2SD}_{17} - 95)^2 & \text{otherwise.} \end{cases} \tag{7.7}$$

The coefficients in Eq. 7.7 have been chosen in such a way that MAE_I2SD_Q90 is continuous over $\overline{I2SD}_{17}$ and that the penalty due to $\overline{I2SD}_{17}$ is around the same order of magnitude as $\overline{\text{MAE}}_{17}$. The effect of $\overline{I2SD}_{17}$ is linearly reduced as $\overline{I2SD}_{17} \to 0.9$. For $0.9 \leq \overline{I2SD}_{17} < 1$ there is a quadratic effect with a minimum for $\overline{I2SD}_{17} = 0.95$. For $\overline{I2SD}_{17} = 1$, the Q90_{17}-measure is added to penalise larger variances.

Fig. 22, Fig. 23, Table 13 and Table 15 show the effects of using different optimisation measures. Forward feature selection is still based on the $\overline{MAE}_{17}$ (see chapter 6), but now, in each feature trial, σ^2_{pump} is optimised based on one of the above three measures. Forward selection is almost the same for each of the three measures (Fig. 22A), and indeed, the same features are selected in all cases (Table 15). Note that the features selected here are different from the features selected for the 1|0|0 linear model (see chapter 6). This is probably due to the fact that in chapter 6, whole-scale 2020 predictions are used to speed up forward selection, whereas here, only 1-step ahead predictions are used. In addition, the Kalman filter is more flexible than the OLS estimator.

Importantly, the prediction variance as estimated by Q90_{17} is very different for the three optimisation measures (Fig. 22B, Table 13). At the end of the fourth round of forward selection, optimisation by "MAE" yields the model with the largest variance ($\sigma^2_{\text{pump,opt}} = 10^6$), "MAE_Q90" yields the smallest variance ($\sigma^2_{\text{pump,opt}} = 0.000394$) and for "MAE_I2SD_Q90", the variance lies in between ($\sigma^2_{\text{pump,opt}} = 100282.6$). This effect can also be seen when comparing predictions (using pump of household 3 as an example): "MAE_I2SD_Q90" is the only optimisation measure that yields plausible prediction intervals (Fig. 23, compare prediction intervals in B, D and F).

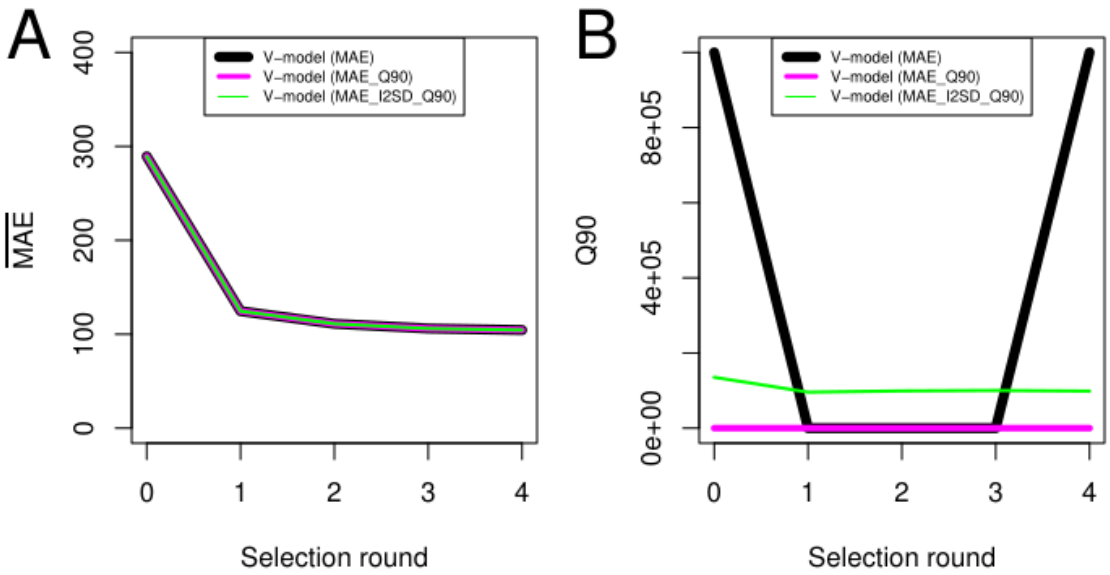


Figure 22: Forward selections for different V-models.

Model name	Forward-selected features	m	$\overline{\text{MAE}}_{17}$ (k=1)	$\overline{\text{MSE}}_{17}$ (k=1)	$\overline{\text{I2SD}}_{17}$ (k=1)
V-model (MAE)	T15_difference_season_profile_sqmetres T18_difference_winter_profile_sqmetres_pump_power_at_zero T18_difference_winter_profile day_profile_pump_T15_difference	17	104.43	98643.51	1
V-model (MAE_Q90)	T15_difference_season_profile_sqmetres T18_difference_winter_profile_sqmetres_pump_power_at_zero T18_difference_winter_profile day_profile_pump_T15_difference	17	104.44	98645.74	0
V-model (MAE_I2SD_Q90)	T15_difference_season_profile_sqmetres T18_difference_winter_profile_sqmetres_pump_power_at_zero T18_difference_winter_profile day_profile_pump_T15_difference	17	104.44	98645.33	0.95

Table 15: Comparison of different V-models.

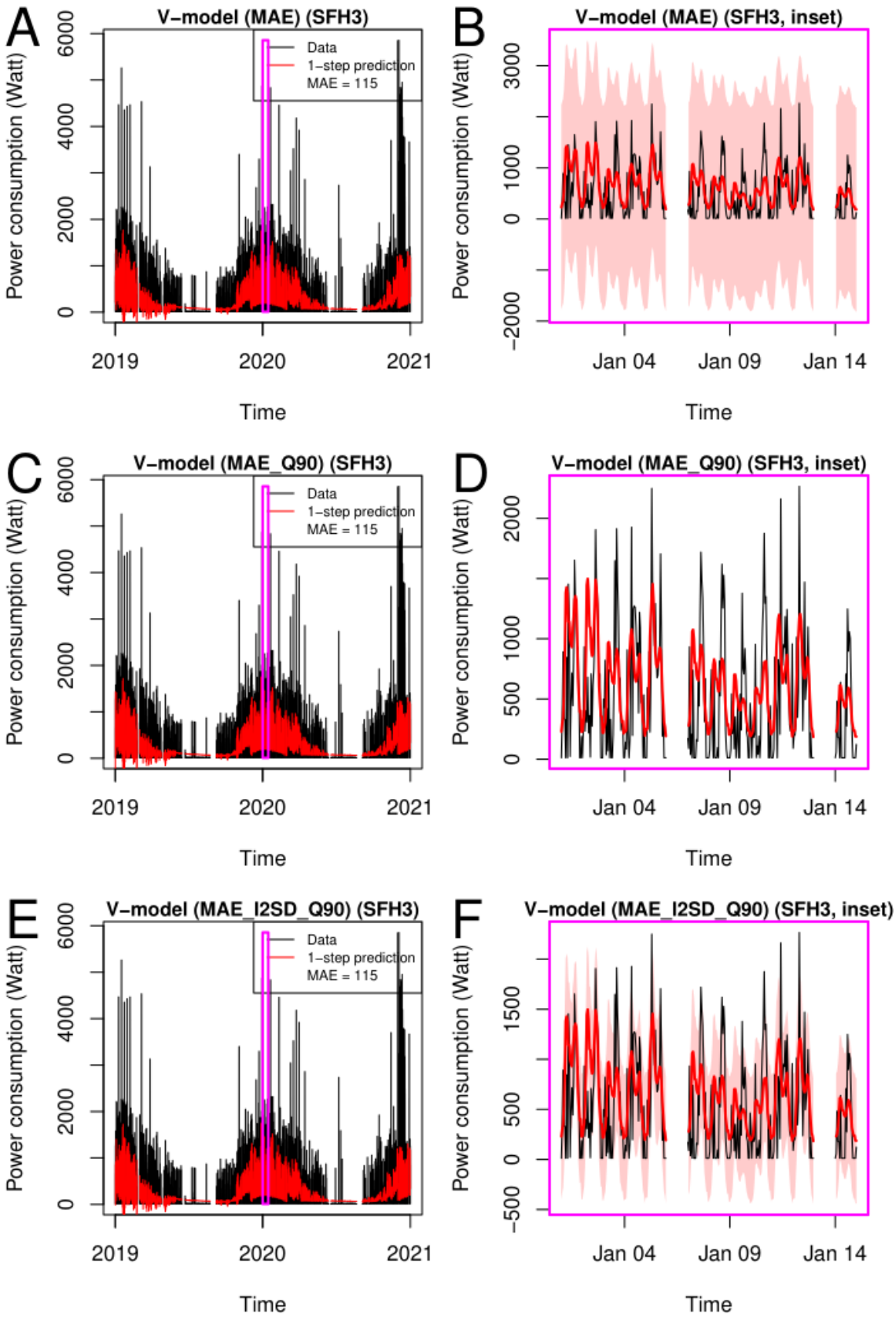


Figure 23: 1-step predictions for the pump of household 3 for different V-models with a simultaneous fit of m=17 pumps. A, B: V-model (MAE); C, D: V-model (MAE_Q90); E, F: V-model (MAE_I2SD_Q90).

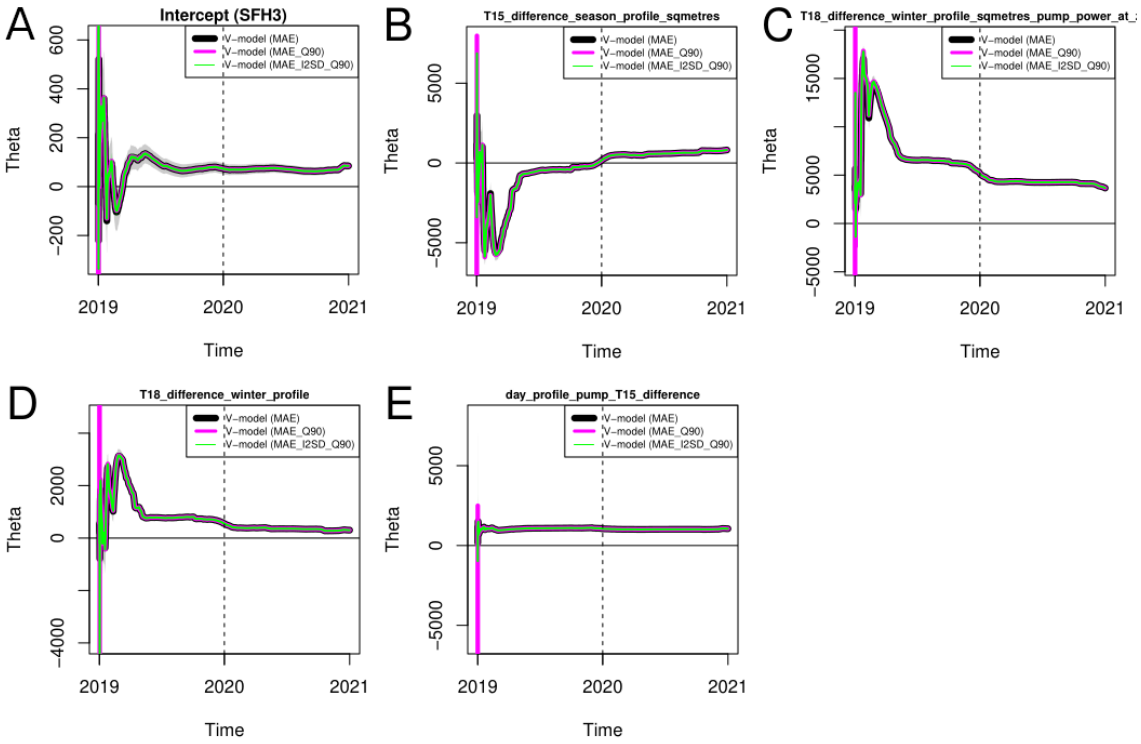


Figure 24: Thetas for different V-models. Note that for better readability, in A, only one of the 17 intercepts is shown (that for pump of household 3) for all three models. B, theta for feature T15_difference_season_profile_sqmetres; C, theta for feature T18_difference_winter_profile_sqmetres_pump_power_at_zero; D, theta for feature T18_difference_winter_profile; E, theta for feature day_profile_pump_T15_difference. The dashed vertical line indicates the beginning of the analysis period (2020).

Not only the selected features are the same, but their estimated θ values are also (almost) the same, since the estimation of the expected value is not affected by σ^2_{pump} (Table 14, Fig. 24; section 4.7.1). Note that there is a somewhat tumultuous adaptation period where the estimation of θ fluctuates wildly, but by 2020, θ values are much more constant (Fig. 24). The fact that the estimators are essentially the same for the three different V-models means that feature effects (Fig. 25), predictions (of the expected value, not the variance; Fig. 23) and residuals (Fig. 26) are also essentially the same. Therefore, only the MAE_I2SD_Q90 V-model will be discussed further.

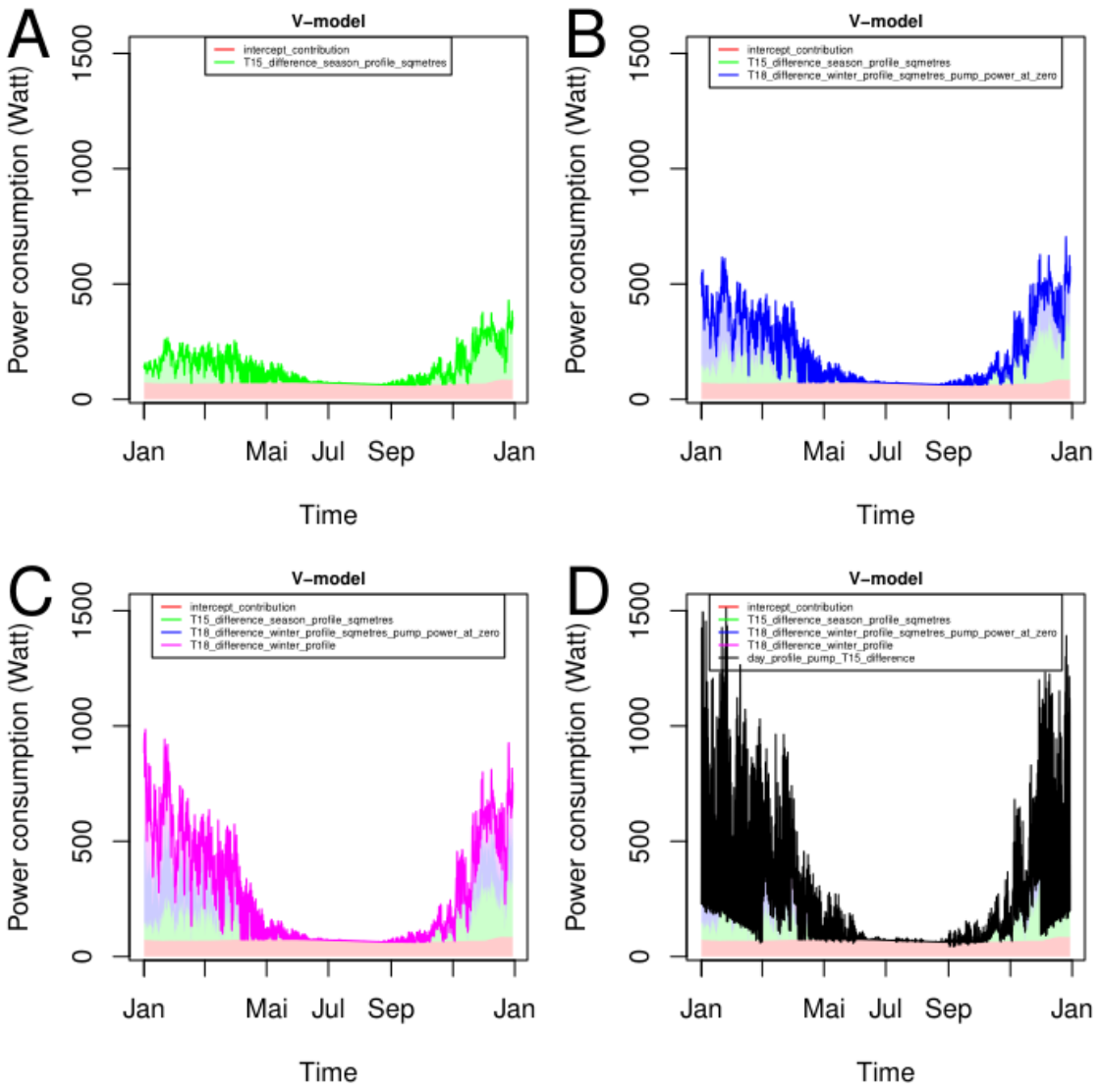


Figure 25: Feature effects for V-model (MAE_I2SD_Q90).

Feature effects for the V-model (MAE_I2SD_Q90) are shown in Fig. 25, using the pump of household 3 as an example. Because the estimators are not constant in time, the effect of each feature is visualised by plotting the contribution of each feature to the model over time. As in chapter 6, the contribution of a feature at time point t is defined as $x_{t,\text{feature}} \cdot \widehat{\theta}_{t,\text{feature}}$. Fig. 25A-C show the effects of the intercept and the first three selected features:

- T15_difference_season_profile_sqmetres,
- T18_difference_winter_profile_sqmetres_pump_power_at_zero and
- T18_difference_winter_profile.

This combination captures the long-term seasonal profile. The last selected feature (day_profile_pump_T15_difference) adds daily variation in the form of daily pump usage patterns (Fig. 25D).

Note that the first three features describe the long-term seasonality with an interaction of temperature (T15_difference or T18_difference) and season profiles (sea-

son_profile or winter_profile). The season and winter profile are derived from the mean daily temperature, allowing for slightly different profile shapes (chapter 6, Fig. 10). So the interaction between temperature and season profiles is similar to a quadratic temperature effect. In addition, the features are modified by an interaction with household-specific features (building size and pump efficiency). This allows the long-term seasonal profile to adapt a little bit to each individual pump in the simultaneous fit. Using three features of this type allows the model to form a weighted sum to adapt better to the real shape than a single such feature might. Although the selected features are slightly different than for the 1|0|0 linear model (chapter 6), the overall effects are very similar.

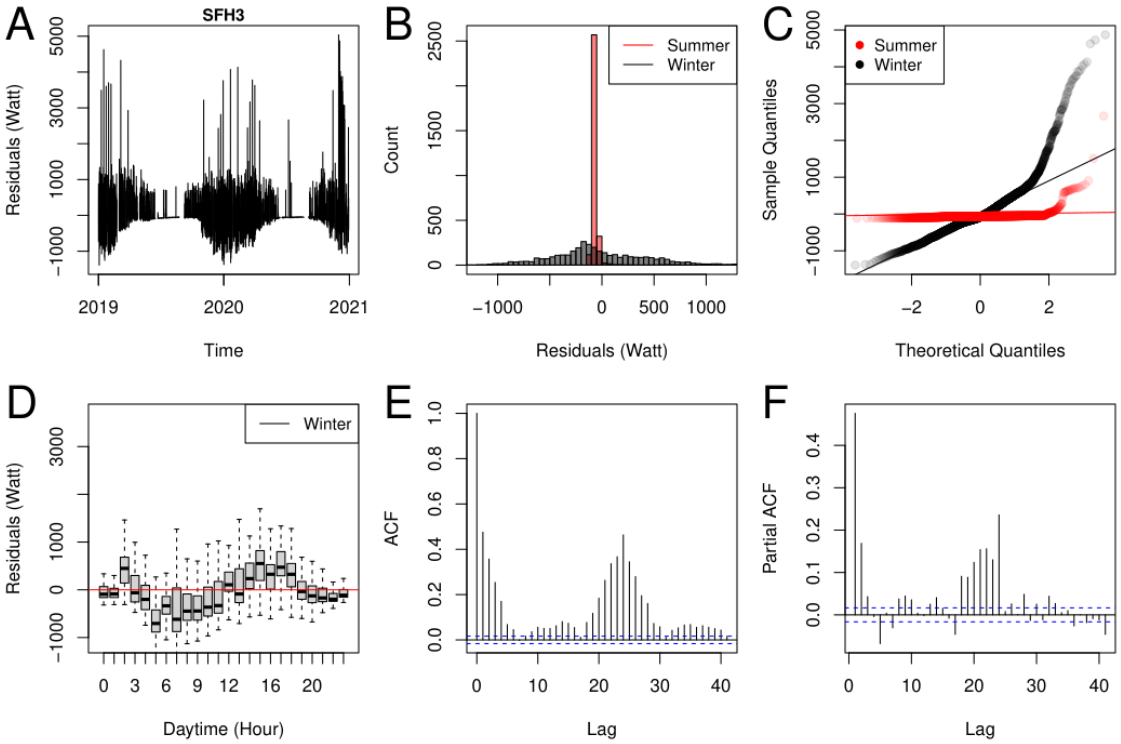


Figure 26: Residual analysis for the V-model (MAE_I2SD_Q90). Note that in this condition, 17 pumps are fit simultaneously, but for clarity, the residual analysis of only one of these pumps (SFH3) is shown here.

The residual analysis in Fig. 26 is also very similar to what was observed for the linear model (Fig. 17). This is expected because the structure of the data has not changed. The Kalman filter can adapt better to data than the OLS estimator, but as we set θ-variances to zero in the V-model ($W = 0_{(m+p)x(m+p)}$; Table 12), this adaptability is still constrained. For this reason, the mean prediction error is only slightly smaller in the V-model than in the 1|0|0 linear model ($\overline{MAE}_{17}$=104.44 Watt vs. 108.72 Watt in the linear model; $\overline{MSE}_{17}$=98645.33 vs. 100975.00 in the linear model; Table 9 and 15). In the next section, we will observe the effect of removing the restriction on W.

7.2 W-model

The W-model is given by:

$$\begin{aligned} Y_t &= F_t\theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t \end{aligned} \tag{7.8}$$

with the definitions of Y_t, F_t, θ_t, v_t, w_t and θ_0 given in Eqs. 7.2ff. Note that the expectation of θ_t is still constant, but θ_t now has a distribution and so can vary. There are now 22 parameters to be optimised: $\sigma^2_{\text{pump}}, \sigma^2_{\alpha,1}, \ldots, \sigma^2_{\alpha,m}, \sigma^2_{\beta,1}, \ldots, \sigma^2_{\beta,p}$. For optimisation, the "Nelder-Mead"-method and the "MAE_I2SD_Q90"-measure described in the previous section are used. Because there are now 22 parameters, optimisation takes a very long time (three days on a computer with 4 CPUs and 20 GB RAM), and for this reason no forward selection is performed for the W-model. Instead, the features selected for the V-model are carried forward (Table 15). These features are:

- 15_difference_season_profile_sqmetres,
- T18_difference_winter_profile_sqmetres_pump_power_at_zero,
- T18_difference_winter_profile and
- day_profile_pump_T15_difference.

Three models are compared in this section: The full W-model, with 22 parameters, the W (I0) model, where $\sigma^2_{\alpha,1} = \ldots = \sigma^2_{\alpha,m} = 0$ (the variances of the intercept thetas), so there are only 5 parameters, and W (F0), where $\sigma^2_{\beta,1} = \ldots = \sigma^2_{\beta,p} = 0$ (the variances of the feature thetas), so that there are 18 parameters. The values of the optimised parameters are shown in Table 13. In general, $\sigma^2_{\text{pump}} \approx 100282$ while the θ variances have values between 0 and 10.

Model name	m	$\overline{\text{MAE}}_{17}$ (k=1)	$\overline{\text{MSE}}_{17}$ (k=1)	$\overline{\text{I2SD}}_{17}$ (k=1)
W-model	17	89.27	77112.73	0.96
W-model (I0)	17	100.92	87742.34	0.96
W-model (F0)	17	88.59	81277.84	0.96

Table 16: Comparison of different W-models.

Predictions for the different W-models are compared in Fig. 27, using the pump of household 3 as an example.

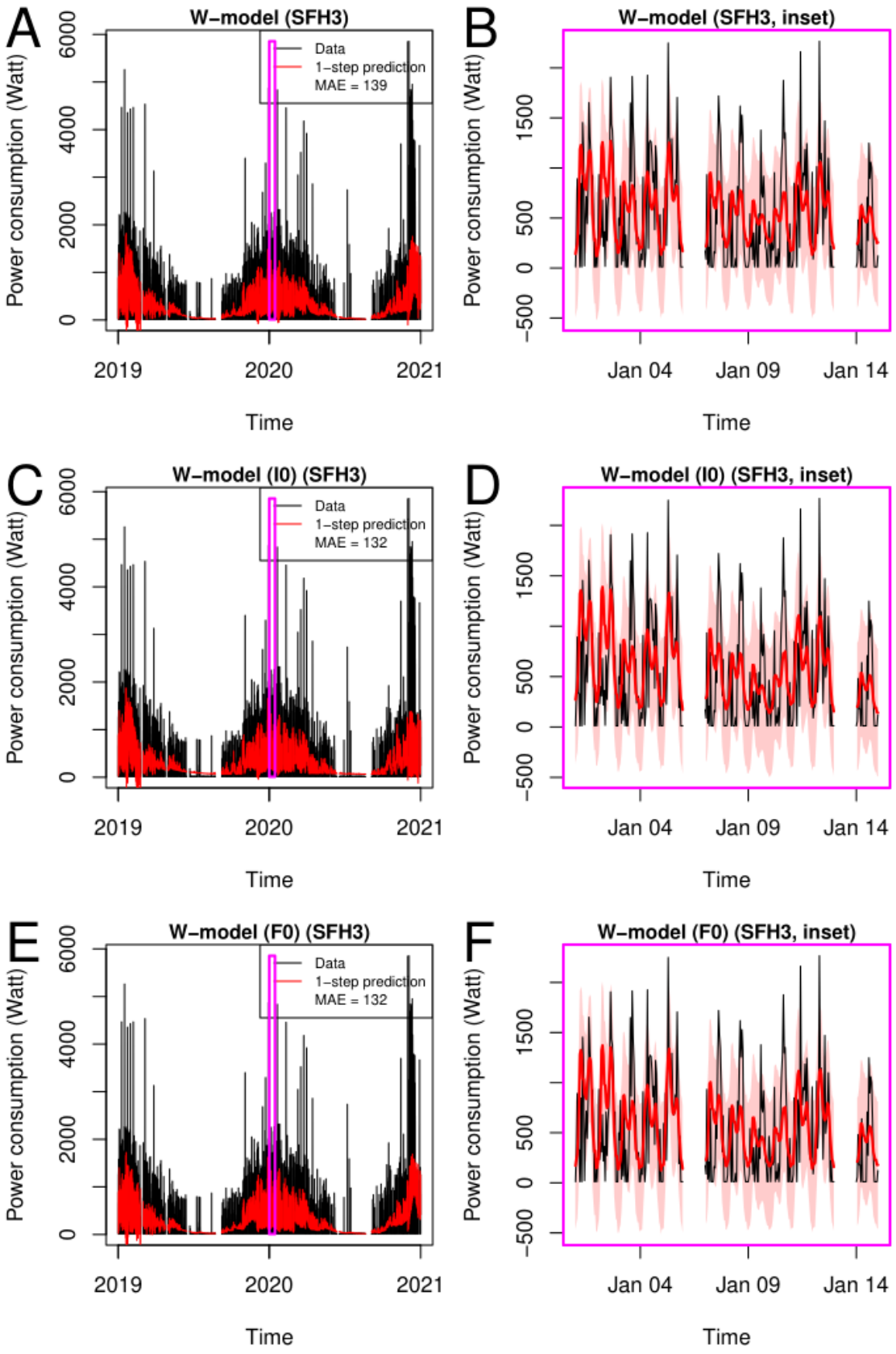


Figure 27: 1-step predictions for the pump of household 3 for different W-models with a simultaneous fit of m=17 pumps. A, B: W-model; C, D: W-model (I0); E, F: W-model (F0).

Table 16 summarises the three W-models. Overall the W (I0) model performs worst. For that model, the intercept variances are 0, and the θ values of the features are constrained by the fact that 17 pumps are fit simultaneously - these pumps have individual intercepts, but there is only one θ for each feature which has to work for all the pumps (Eq. 7.1f). So the variance of the feature thetas does not have much scope to work in. Even so, the W (I0) model still performs better than the V-model ($\overline{MAE}_{17}$=100.92 Watt vs. 104.44 Watt in the V-model; $\overline{MSE}_{17}$=87742.34 vs. 98645.33 in the V-model; Table 15 and 16). The W (F0) model performs even better ($\overline{MAE}_{17}$=88.59 Watt) because now the intercepts can vary freely. Finally, the full W-model performs similar to the W (F0) model with respect to the $\overline{MAE}_{17}$, but much better for the $\overline{MSE}_{17}$ (77112.73 vs. 81277.84; Table 16).

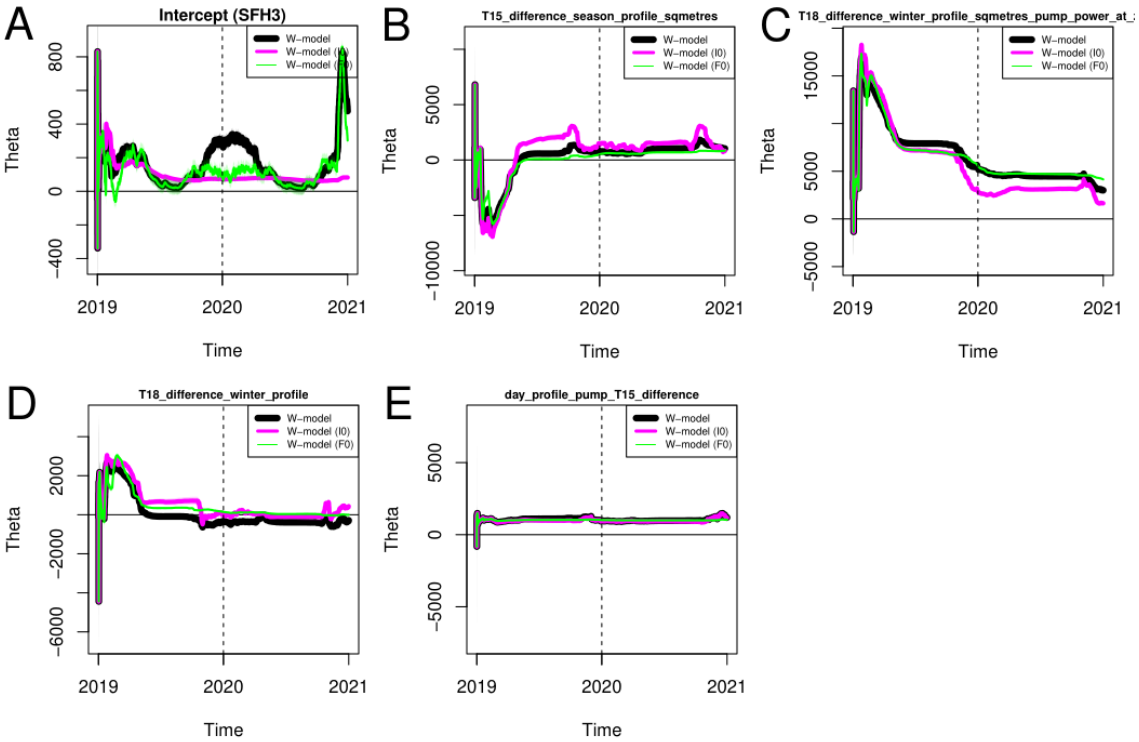


Figure 28: Thetas for different W-models. Note that for better readability, in A, only one of the 17 intercepts is shown (that for pump SFH3) for all three models. B, theta for feature T15_difference_season_profile_sqmetres; C, theta for feature T18_difference_winter_profile_sqmetres_pump_power_at_zero; D, theta for feature T18_difference_winter_profile; E, theta for feature day_profile_pump_T15_difference. The dashed vertical line indicates the beginning of the analysis period (2020).

Fig. 28 shows the Kalman estimator for the different W-models. Unlike for the V-models, θ is quite different for the three W-models. For example, the intercept values vary a lot in the full W-model and the W (F0) model, but not as much in the W (I0) model, which has zero intercept variances and fairly constant intercept values in 2020 (Fig. 28A). The other models seem to make use of the intercept variance to adapt more closely to the data and in this way achieve better predictions. For

example, there are peaks in the intercept value during winter time including a particularly large peak at the end of 2020, reflecting increased power consumption (e.g. see Fig. 27A). This adaptation by the intercepts is not a good thing: one would prefer the θ estimator to be fairly constant over time because that would be a sign that the model reflects the true data-generating process. This is not the case here. However for short-term predictions, it does not matter if there are large changes in θ, as long as these changes are happening on a much larger timescale than the prediction (see also section 7.5).

Residuals for the full W-model and the pump of household 3 are shown in Fig. 29. The residual analysis of the W-model is not notably different from the V-model (compare Fig. 26 and Fig. 29). Furthermore, there is no big difference in the normality and patterns of winter residuals in different W-models (Fig. 30 and Fig. 31), indicating that although the W-model is more adaptable than the V-model (compare Table 15 and Table 16), it still does not capture all the variability in the data. In the next section, we test the particle filter, an alternative estimation method which is better able to capture data variability that happens on a very short timescale.

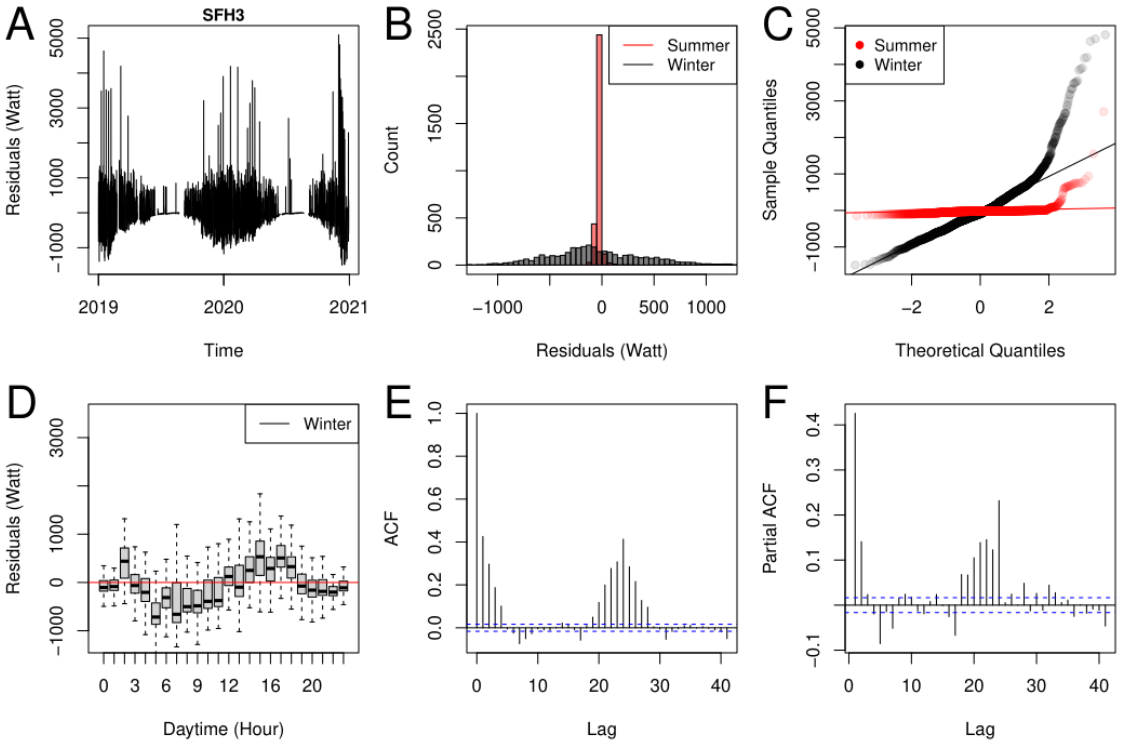


Figure 29: Residual analysis for the W-model. Note that in this condition, 17 pumps are fit simultaneously, but for clarity, the residual analysis of only one of these pumps (SFH3) is shown here.

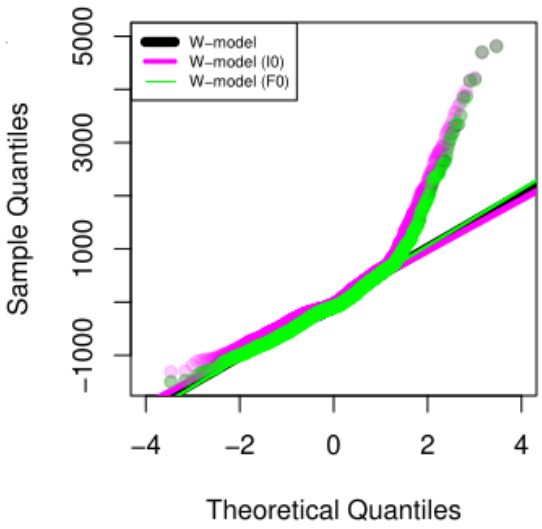


Figure 30: Normality analysis of winter residuals for different W-models.

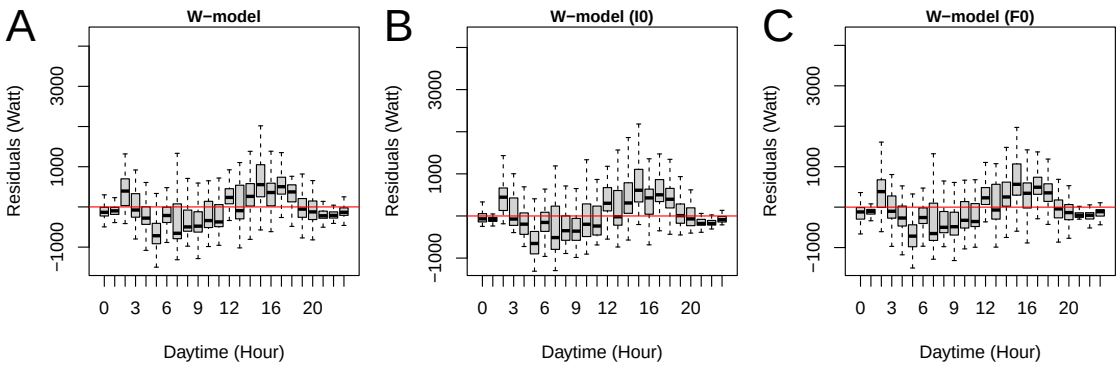


Figure 31: Residual patterns for different W-models.

7.3 P-model

The P-model is based on the same description as the W-model.

$$\begin{aligned} Y_t &= F_t\theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t \end{aligned} \tag{7.9}$$

with the definitions of Y_t, F_t, θ_t, v_t and w_t given in Eqs. 7.2ff. Note that like for the W-model, the features of the V-model are used. However, now a "Sequential Monte Carlo"-method or "particle filter" is used for estimation instead of the Kalman filter (see section 4.7.2).

The particle filter users N particles to approximate the distribution of $\theta_t|y_{1:t}$. Any distribution assumption can be made. We use $\theta_0 \sim N_{m+p}(c, C)$, where c and C are the average of the estimates for $\mathbb{E}[\theta_t|y_{1:t}]$ and $\mathrm{Cov}[\theta_t|y_{1:t}]$ found by the Kalman filter for the full W-model in 2020 (Table 14). N initial particles, $x_0^{(i)}, i = 1, \ldots, N$,

are drawn from this distribution. Each particle represents a potential value for θ_0. Proceeding sequentially through the time points, θ_t is estimated as follows (see section 4.7.2 for more detail): First, N new particles $x_t^{(i)}, i = 1, \ldots, N$ are drawn from $N_{m+p}(\overline{x}_{t-1}, W)$, where $\overline{x}_{t-1}$ is the mean of the N particles of the previous time point and $W := C$. Then, for each new particle $x_t^{(i)}, i = 1, \ldots, N$, the value of the probability density at $Y_t = y_t$ under $Y_t \sim N_m(F_t x_t^{(i)}, V)$ is calculated, using the optimised V of the full W-model. The probability density values are normalised to their sum over all particles and the normalised values are used as weights in the next step, when resampling the particles with replacement. Weighted resampling ensures that particles which describe the data at this time point well are enriched in the sample. Then the estimates are calculated for the time point: $\widehat{\mathbb{E}}[\theta_t|y_{1:t}] = \overline{x}_t$ and $\widehat{\mathrm{Cov}}[\theta_t|y_{1:t}] = \mathrm{Cov}[x_t]$ (see section 4.7.2).

Here we test two versions of the particle filter: $N = 10^3$ and $N = 10^4$ (Table 17, Fig. 32). Given that θ_t is 21-dimensional, even 10^4 particles are probably too few particles to cover the parameter space adequately. Unfortunately, increasing the number of particles further slows the estimation to an unfeasible degree (for 10^6 particles, the computation did not complete within five days on a computer with 4 CPUs and 20 GB RAM). However, even with comparably few particles, the particle filter works suprisingly well (Table 17, Fig. 32). In fact, with respect to the $\overline{MAE}_{17}$ measure, the best P-model (10^4) is as good as the best W-model ($\overline{MAE}_{17} = 88.76$ vs. 88.59 for the W (F0) model), and the $\overline{MSE}_{17}$ is even better ($\overline{MSE}_{17} = 75512.31$ vs. 77112.73), indicating that perhaps now some peaks and troughs, which do not influence the $\overline{MAE}_{17}$ but which do influence the $\overline{MSE}_{17}$, are predicted slightly better.

There is not a big difference between 10^3 and 10^4 particle filters when it comes to predictions, feature effects and residuals (Fig. 32-36). Like the Kalman filter estimator for the W-model, the intercepts of P-models vary considerably, absorbing much of the data variation that cannot be explained by the features, whose θ values are more constrained by fitting 17 pumps simultaneously. Recall that, during summer time, most of the features are zero because temperatures are higher than 15°C or 18°C, meaning that data points in this period do not affect the value of the θ estimator much. However, in the winter periods, the particle filter estimator is much "noisier" than the Kalman filter estimator (Fig. 33; for a direct comparison, see also Fig. 43 in section 7.5). That reflects the fact that particles can adapt better to the data and explains the smaller $\overline{MSE}_{17}$. Despite the greater flexibility, the particle filter is not a huge improvement on the Kalman filter for the W-model, suggesting that structural changes in the model, rather than changes in the estimation method are required to improve the model further.

Model name	m	$\overline{\mathrm{MAE}}_{17}$ (k=1)	$\overline{\mathrm{MSE}}_{17}$ (k=1)	$\overline{\mathrm{I2SD}}_{17}$ (k=1)
P-model (10^3)	17	90.84	76475.07	0.96
P-model (10^4)	17	88.76	75512.31	0.96

Table 17: Comparison of different P-models.

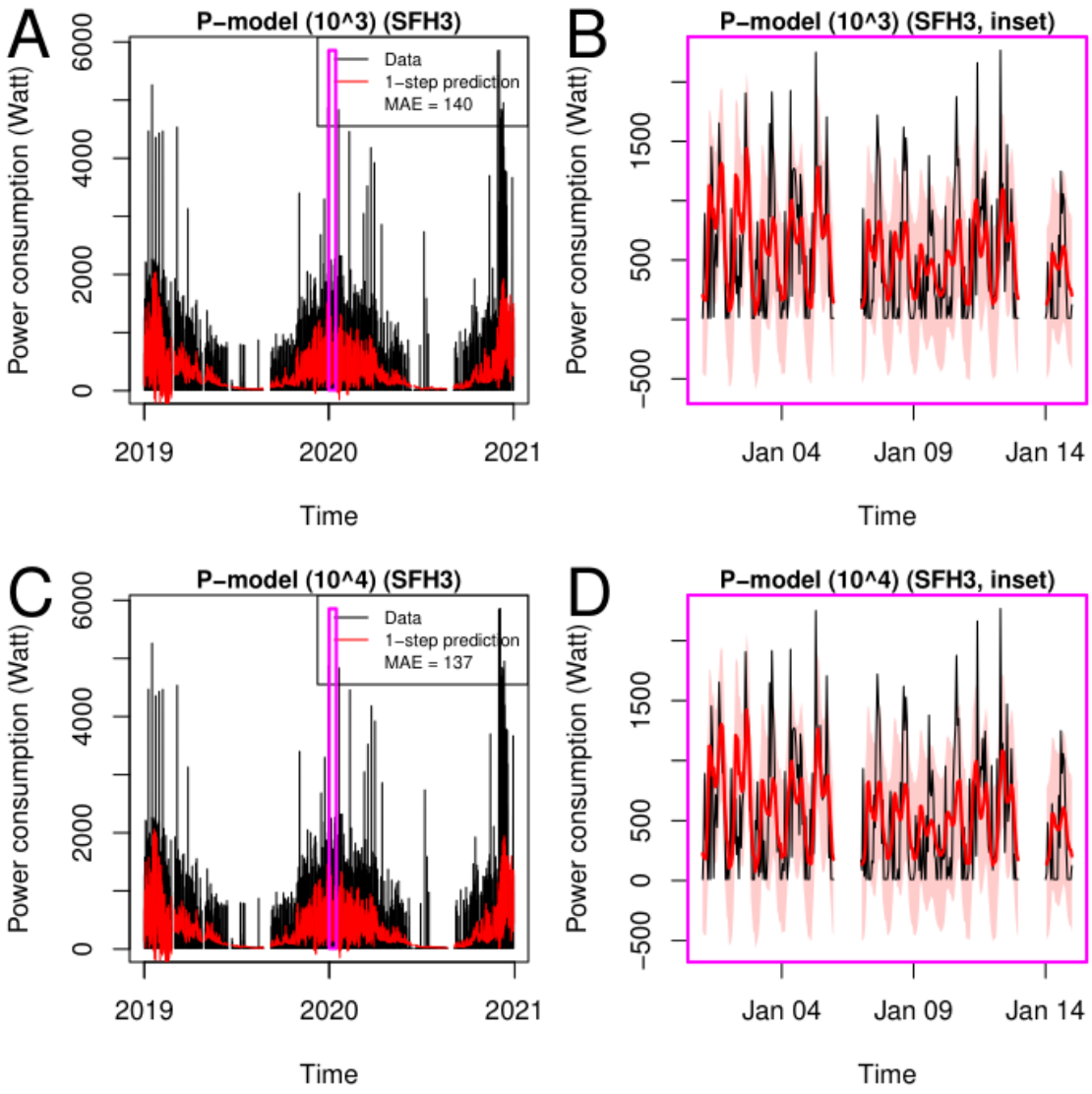


Figure 32: 1-step predictions for the pump of household 3 for different P-models with a simultaneous fit of m=17 pumps. A, B: P-model (10^3); C, D: P-model (10^4).

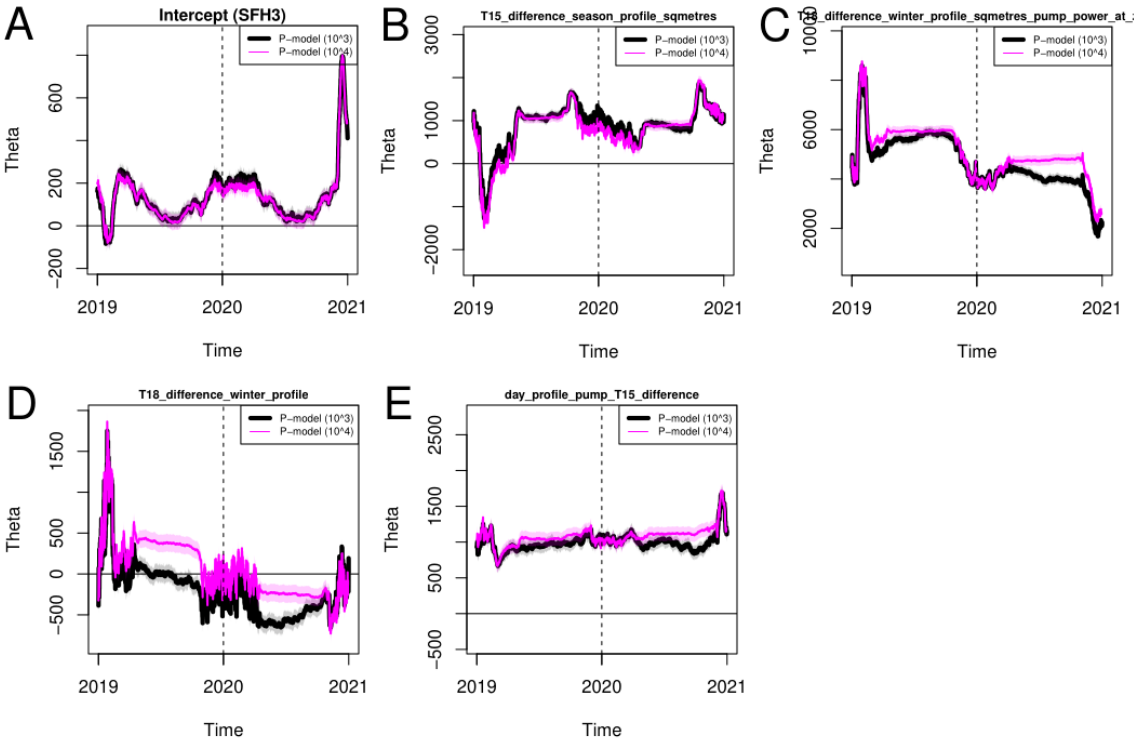


Figure 33: Thetas for different P-models. Note that for better readability, in A, only one of the 17 intercepts is shown (that for pump of household 3). B, theta for feature T15_difference_season_profile_sqmetres; C, theta for feature T18_difference_winter_profile_sqmetres_pump_power_at_zero; D, theta for feature T18_difference_winter_profile; E, theta for feature day_profile_pump_T15_difference. The dashed vertical line indicates the beginning of the analysis period (2020).

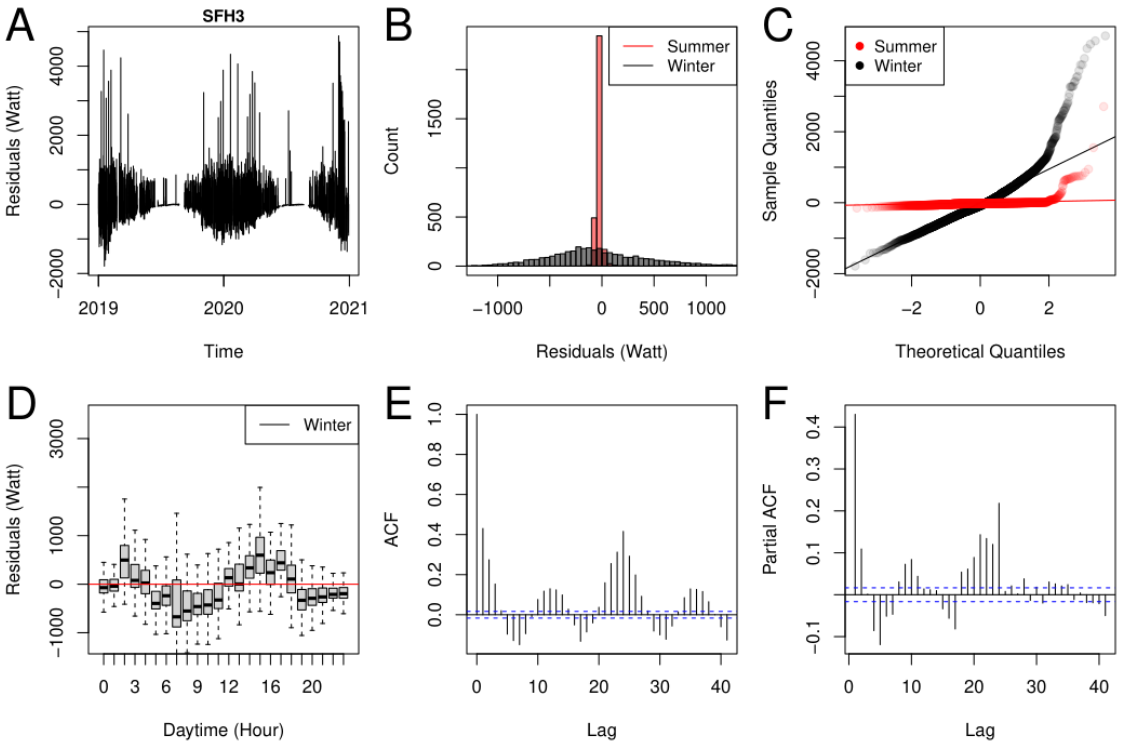


Figure 34: Residual analysis for the P-model with 10^4 particles. Note that in this condition, 17 pumps are fit simultaneously, but for clarity, the residual analysis of only one of these pumps (SFH3) is shown here.

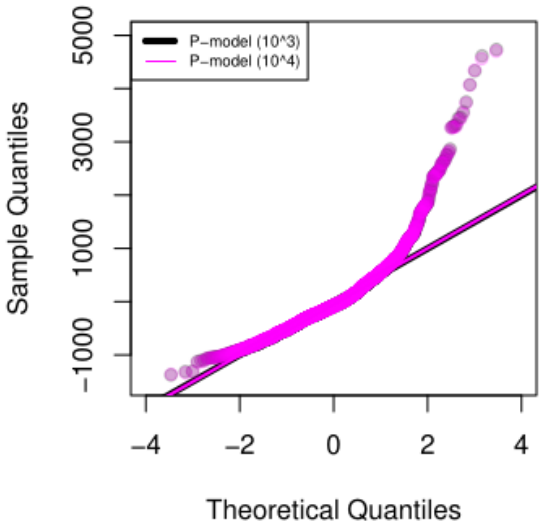


Figure 35: Normality analysis of winter residuals for different P-models.

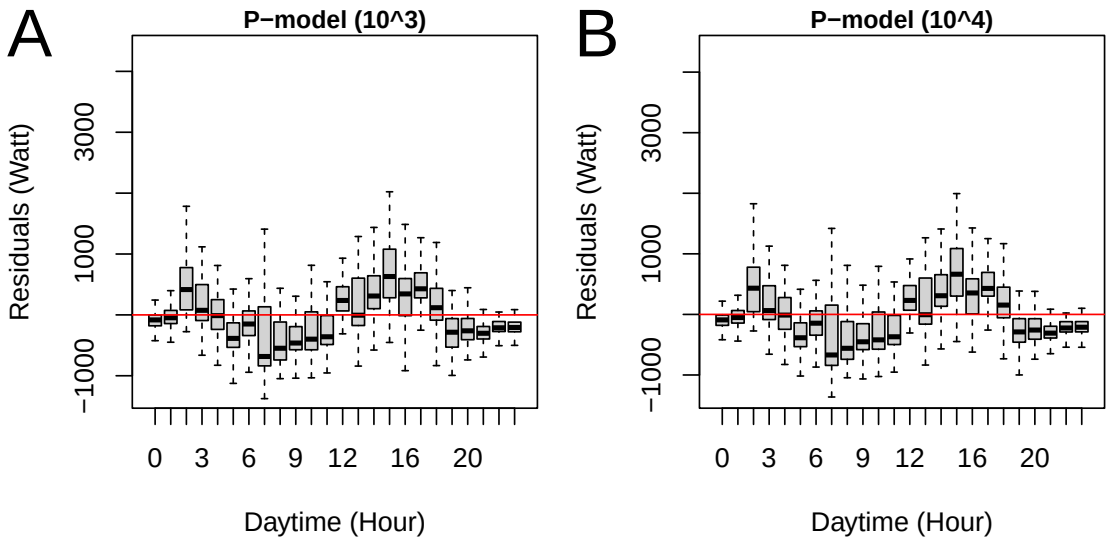


Figure 36: Residual patterns for different P-models.

7.4 A-model

The next step in this state space model evolution is the A-model:

$$\begin{aligned} Y_t &= F_t\theta_t + \mu_t \\ \theta_t &= \theta_{t-1} + w_t \\ \mu_t &= \phi_1\mu_{t-1} + \phi_2\mu_{t-2} + v_t \end{aligned} \tag{7.10}$$

with the definitions of Y_t, F_t, θ_t, μ_t, θ_0, μ_0, v_t, w_t, ϕ_1 and ϕ_2 given in Eqs. 7.2ff. Here, we assume that $\mathbb{E}[\mu_t] = 0$, so that $\mathbb{E}[Y_t|\theta_t] = F_t\theta_t$ like in all previous models. The sequence $(\mu_t)_t$ is used to model $Y_t - F_t\theta_t$ as a stationary AR(2) process. The A-model is a very simple extension of the W-model that tries to take into account autoregressive noise (see also Fig. 29). This is only a first approach to non-white noise in heat pump power consumption data that could be extended in the future.

For the A-model, Kalman filter estimation is used, because its parameter space is even higher-dimensional than the W-model, so the particle filter is (probably) less suited. The higher dimensionality becomes apparent when Eq. 7.10 is rewritten for use with the Kalman filter:

$$\begin{aligned} Y_t &= \begin{bmatrix} F_t & I_{mxm} & 0_{mxm} \end{bmatrix} \begin{bmatrix} \theta_t \\ \mu_t \\ \mu_{t-1} \end{bmatrix} \\ \begin{bmatrix} \theta_t \\ \mu_t \\ \mu_{t-1} \end{bmatrix} &= \begin{bmatrix} I_{(m+p)x(m+p)} & 0_{(m+p)xm} & 0_{(m+p)xm} \\ 0_{mx(m+p)} & \phi_1 I_{mxm} & \phi_2 I_{mxm} \\ 0_{mx(m+p)} & I_{mxm} & 0_{mxm} \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \mu_{t-1} \\ \mu_{t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ v_t \\ 0_m \end{bmatrix}. \end{aligned} \tag{7.11}$$

With $\tilde{F}_t = \begin{bmatrix} F_t & I_{mxm} & 0_{mxm} \end{bmatrix}$, $\tilde{\theta}_t = (\theta_t^T, \mu_t^T, \mu_{t-1}^T)^T$, $\tilde{w}_t = (w_t^T, v_t^T, 0_m^T)^T$ and

$$\tilde{G}_t := \begin{bmatrix} I_{(m+p)x(m+p)} & 0_{(m+p)xm} & 0_{(m+p)xm} \\ 0_{mx(m+p)} & \phi_1 I_{mxm} & \phi_2 I_{mxm} \\ 0_{mx(m+p)} & I_{mxm} & 0_{mxm} \end{bmatrix},$$

Eq. 7.11 reduces to:

$$\begin{aligned} Y_t &= \tilde{F}_t \tilde{\theta}_t \\ \tilde{\theta}_t &= \tilde{G}_t \tilde{\theta}_{t-1} + \tilde{w}_t. \end{aligned} \tag{7.12}$$

This reformulation of the A-model can be used for estimation with the Kalman filter (see section 4.7.1). Note that v_t is subsumed in $\tilde{w}_t$ and that $\tilde{v}_t = 0_m$. One could think about changing this in a future version; remember that this is just a first extension of the W-model. Also note that the "new" matrix $\tilde{G}_t$ describes the relationship between $\tilde{\theta}_t$ and $\tilde{\theta}_{t-1}$. In the case of the W-model, the relationship is described by the identity ($G_t = I_{(m+p)x(m+p)}$). Finally, note that $\tilde{\theta}_t$ has dimension $3m + p = 55$. This is why we do not use the particle filter for estimation.

In the A-model, there are 24 parameters to be optimised: σ^2_{pump}, $\sigma^2_{\alpha,1}, \ldots, \sigma^2_{\alpha,m}$, $\sigma^2_{\beta,1}, \ldots, \sigma^2_{\beta,p}, \phi_1$ and ϕ_2. For optimisation, the "Nelder-Mead"-method and the "MAE _I2SD_Q90"-measure described in section 7.1 are used. Like for the W-model, optimisation takes a very long time (three days on a computer with 4 CPUs and 20 GB RAM), and for this reason no forward selection is performed. Instead, the features previously selected for the V-model are used (Table 15). These features are:

- 15_difference_season_profile_sqmetres,
- T18_difference_winter_profile_sqmetres_pump_power_at_zero,
- T18_difference_winter_profile and
- day_profile_pump_T15_difference.

Two models are tested in this chapter. For the standard A-model, ϕ_1 and ϕ_2 are constrained to be in $(-1, 1)$ (optimised values are: $\phi_1 = -0.5075$ and $\phi_2 = 0.5398$; Table 13). For the A (+) model, ϕ_1 and ϕ_2 are coerced to be in $(0, 1)$ (optimised values are: $\phi_1 = 0.0121$ and $\phi_2 = 0.8909$; Table 13). The reason for the constraint is the empirical finding that the $\overline{MAE}_{17}$ is much better if ϕ_1 and ϕ_2 are optimised on $(0, 1)$ instead of $(-1, 1)$ (Table 18).

Model name	m	$\overline{\text{MAE}}_{17}$ (k=1)	$\overline{\text{MSE}}_{17}$ (k=1)	$\overline{\text{I2SD}}_{17}$ (k=1)
A-model	17	73.23	100024.86	0.95
A-model (+)	17	53.69	102130.69	0.93

Table 18: Comparison of different A-models.

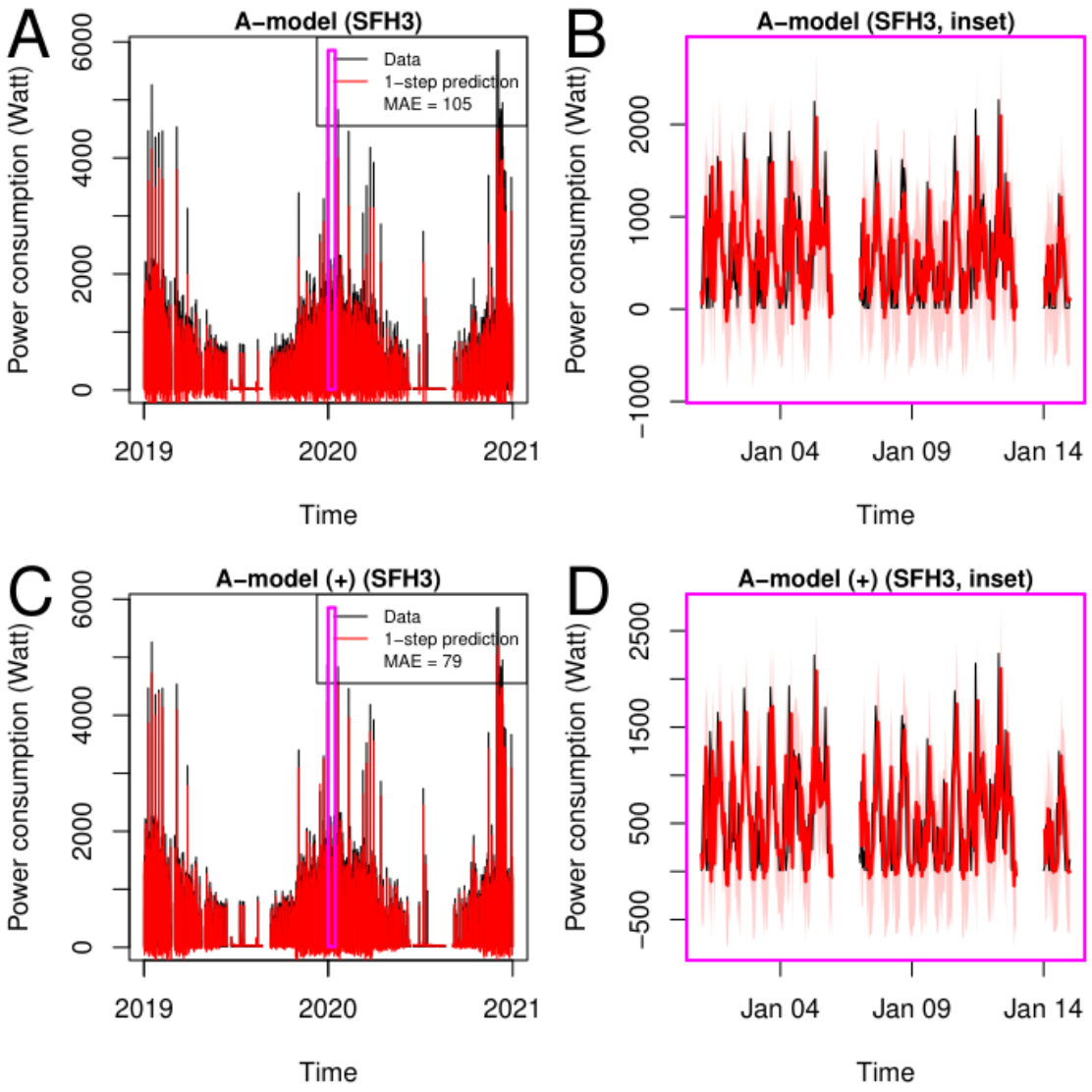


Figure 37: 1-step predictions for the pump of household 3 for different A-models, with a simultaneous fit of m=17 pumps. A, B: A-model; C, D: A-model (+).

Compared to the W-model, the $\overline{MAE}_{17}$ of both A-models is much better ($\overline{MAE}_{17} = 53.69$ for the A (+) model compared to 88.59 for the best W-model; Table 16 and 18). However, the $\overline{MSE}_{17}$ is much worse (100024.86 for the best A-model compared to 77112.73 for the best W-model). To understand this discrepancy, it is useful to look closely at the predictions and residuals (Fig. 37 and 38). The predictions follow the data much more closely than with previous models, although they appear to be slightly shifted to the right (Fig. 37B and D). This may be due to the fact that $\tilde{v}_t$ is zero in Eq. 7.12. In that case, $f_{t+1} \approx y_t$ (for details, see section 4.7.1). This problem could be fixed by introducing $\tilde{v}_t \neq 0_m$ in Eq. 7.12) in a future extension of the model. In any case, the residuals are more symmetric (Fig. 38B, C and Fig. 39) and less patterned (Fig. 38D and Fig. 40), which explains the smaller $\overline{MAE}_{17}$, but because of the shift the magnitude of the residuals is still large (Fig. 38D), and the now large negative residuals contribute more to the $\overline{MSE}_{17}$. Interestingly, the ACF and PACF show that while the AR components have been successfully incorporated

in the model, there still are two MA components visible in the residuals (Fig. 38E, F). Thus, an ARMA(2,2) model may be more appropriate to describe the noise.

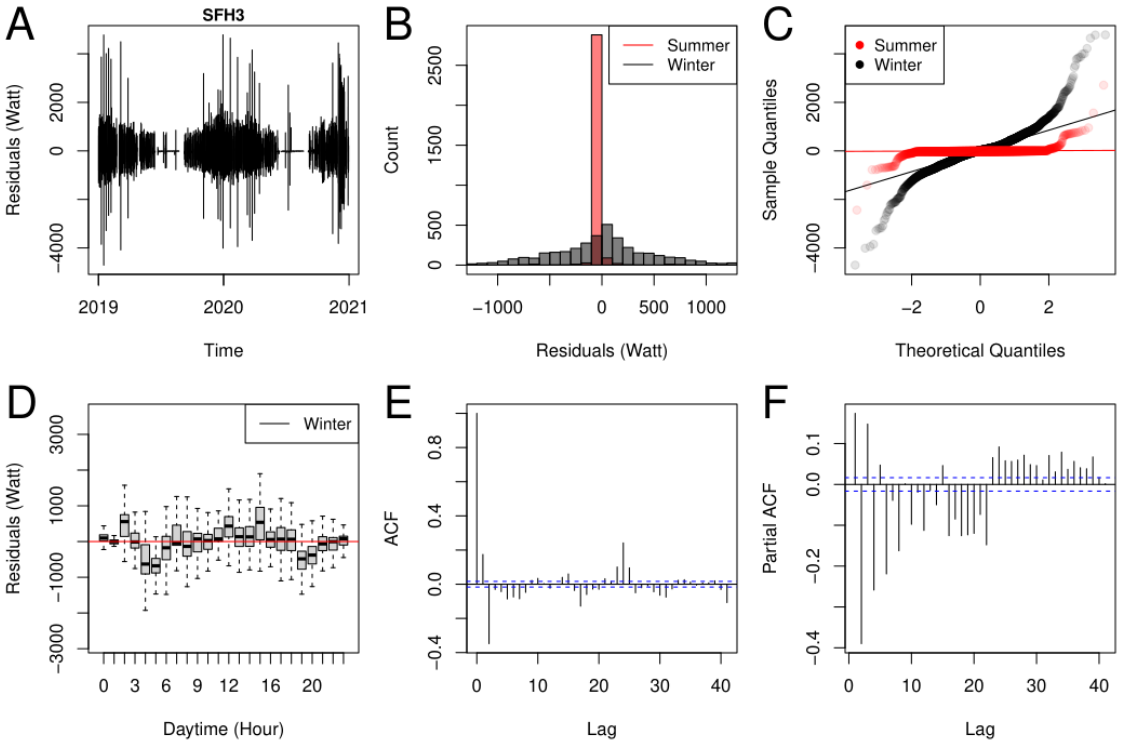


Figure 38: Residual analysis for the A-model with positive AR coefficients. Note that 17 pumps are fit simultaneously, but for clarity, the residual analysis of only one of these pumps (SFH3) is shown here.

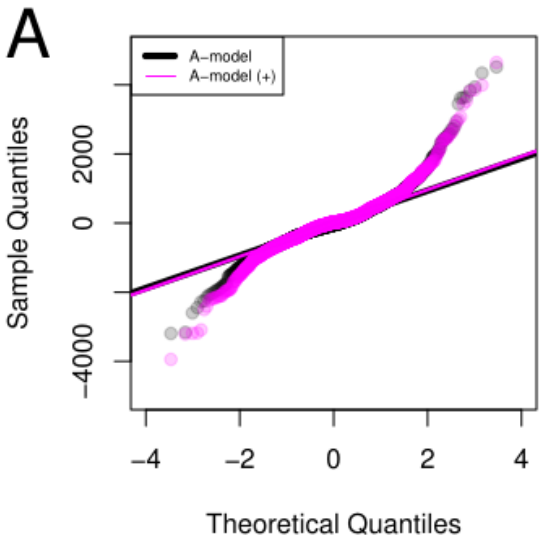


Figure 39: Normality analysis of winter residuals for different A-models.

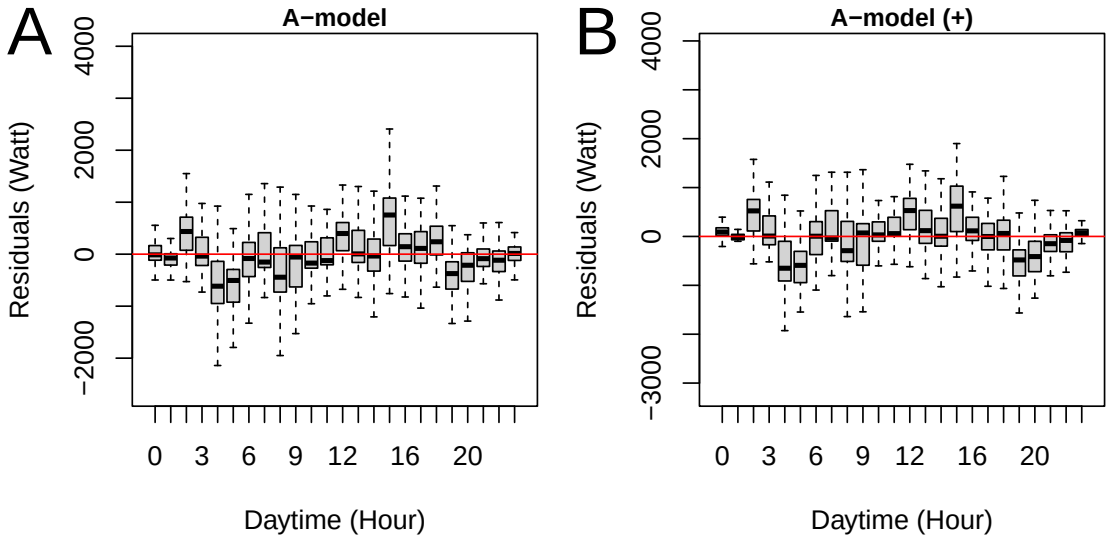


Figure 40: Patterns of winter residuals for different A-models.

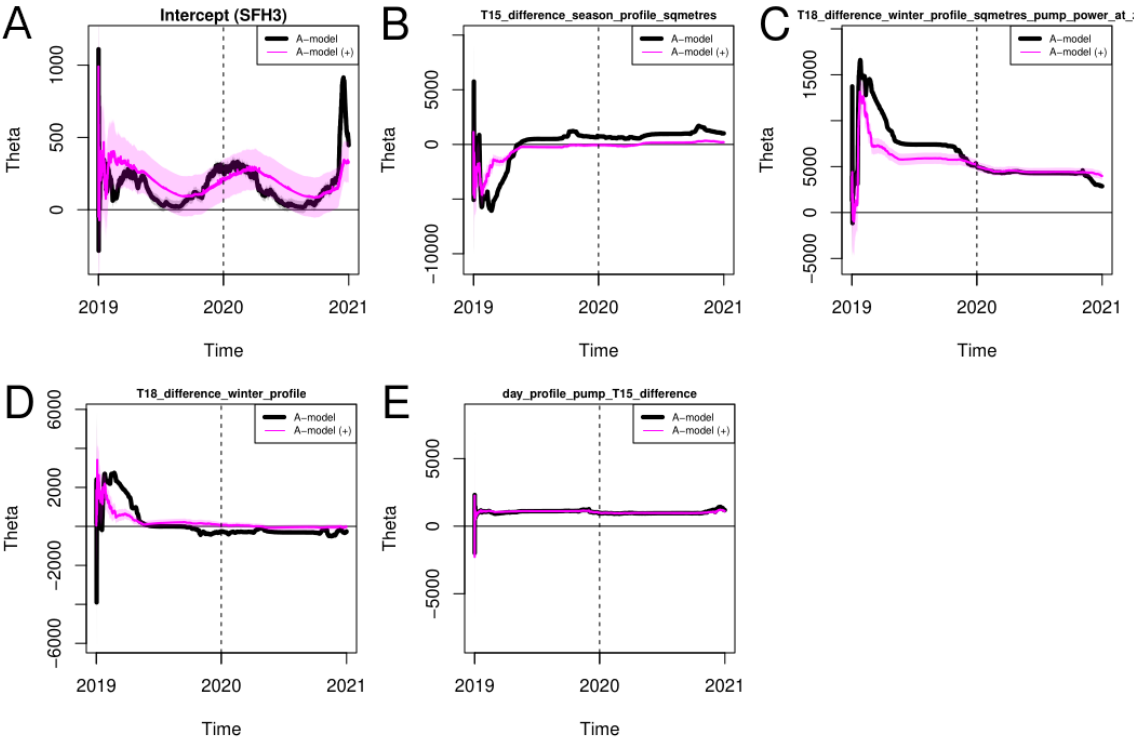


Figure 41: Thetas for different A-models. Note that for better readability, in A, only one of the 17 intercepts is shown (that for the pump of household 3). B, theta for feature T15_difference_season_profile_sqmetres; C, theta for feature T18_difference_winter_profile_sqmetres_pump_power_at_zero; D, theta for feature T18_difference_winter_profile; E, theta for feature day_profile_pump_T15_difference. The dashed vertical line indicates the beginning of the analysis period (2020).

Except for the difference in $\overline{MAE}_{17}$ (Table 18), there is no big difference between the two A-models: in both cases, residuals are more symmetric but still not normal (Fig. 39 and 40). The estimates for θ are quite different for the intercepts and feature 1 and 3 (Fig. 41B and (Fig. 41D; Table 14). The intercepts also are shifted with respect to each other, which must be due to the different optimisation constraints on ϕ_1 and ϕ_2, since that is the only difference between the two models. Note that the averages of the estimates for μ_t lie within two standard deviations of 0 (Table 19), consistent with the assumption that $\mathbb{E}[\mu_t] = 0$ (Eq. 7.10).

	estimate	lower_bound	upper_bound	contains_zero
$\overline{\widehat{\mu}}_1$	-1.44	-119.24	116.37	TRUE
$\overline{\widehat{\mu}}_2$	-9.59	-122.45	103.28	TRUE
$\overline{\widehat{\mu}}_3$	3.40	-119.68	126.47	TRUE
$\overline{\widehat{\mu}}_4$	5.24	-101.05	111.53	TRUE
$\overline{\widehat{\mu}}_5$	-0.71	-108.54	107.12	TRUE
$\overline{\widehat{\mu}}_6$	-24.39	-142.21	93.42	TRUE
$\overline{\widehat{\mu}}_7$	21.08	-55.25	97.41	TRUE
$\overline{\widehat{\mu}}_8$	-14.61	-127.73	98.51	TRUE
$\overline{\widehat{\mu}}_9$	-32.96	-143.20	77.29	TRUE
$\overline{\widehat{\mu}}_{10}$	25.94	-114.23	166.11	TRUE
$\overline{\widehat{\mu}}_{11}$	-22.73	-145.91	100.45	TRUE
$\overline{\widehat{\mu}}_{12}$	8.82	-90.49	108.13	TRUE
$\overline{\widehat{\mu}}_{13}$	66.56	-51.52	184.64	TRUE
$\overline{\widehat{\mu}}_{14}$	-11.12	-112.14	89.89	TRUE
$\overline{\widehat{\mu}}_{15}$	-15.46	-137.82	106.91	TRUE
$\overline{\widehat{\mu}}_{16}$	-47.37	-168.61	73.86	TRUE
$\overline{\widehat{\mu}}_{17}$	-9.29	-121.02	102.43	TRUE

Table 19: $\overline{\widehat{\mu}}(j) \pm 2\sqrt{\overline{\widehat{C}}(j,j)}$ for $j = 1, \ldots, 17$ in 2020 for the A (+) model.

7.5 Comparison

In this section, the linear model (LM), V-model (MAE_I2SD_Q90), the full W-model, the P-model (10^4) and the A-model (+) are compared (Table 20ff). The same analysis for household data is shown in Appendix A, Tables 27-31 and Fig. 52. Note that for this section, the estimations have been re-run with slightly different starting values for θ_0; instead of 0_{m+p}, $\overline{\widehat{\theta}}$ of 2020 has been used (Table 20). For this reason, the $\overline{\text{MAE}}_{17}$ (k=1) and $\overline{\text{MSE}}_{17}$ (k=1) may be slightly different in Table 20 than in previous tables. Furthermore, the LM-model listed in these tables does not correspond directly to the 1|0|0-model described in chapter 6, which used different features; for better comparability, that model was re-run with the features selected by the V-model (Table 15), so that all models in this section are using the same features. These features are:

- 15_difference_season_profile_sqmetres,
- T18_difference_winter_profile_sqmetres_pump_power_at_zero,

- T18_difference_winter_profile and
- day_profile_pump_T15_difference.

The linear model with these features actually performs slightly better than the model from chapter 6 (Table 20 vs. Table 9), indicating that forward selection with the V-model has identified better features. (Recall that model selection was performed slightly differently for the linear model, not using 1-step-ahead predictions to minimise CPU time, see chapter 6).

Model name	Method	m	$\overline{\text{MAE}}_{17}$ (k=1)	$\overline{\text{MSE}}_{17}$ (k=1)	$\overline{\text{I2SD}}_{17}$ (k=1)
LM-model	OLS	17	104.24	98498.98	0.97
V-model (MAE_I2SD_Q90)	Kalman	17	103.96	97492.82	0.95
W-model	Kalman	17	88.93	76402.41	0.96
P-model (10^4)	Particle	17	88.76	75512.31	0.96
A-model (+)	Kalman	17	53.49	101795.13	0.95

Table 20: Comparison of different model types.

Model name	m	$\overline{\text{MAE}}_{17}$ (k=1)	$\overline{\text{MAE}}_{17}$ (k=6)	$\overline{\text{MAE}}_{17}$ (k=12)	$\overline{\text{MAE}}_{17}$ (k=18)	$\overline{\text{MAE}}_{17}$ (k=24)
LM-model	17	104.24	104.21	104.19	104.17	104.14
V-model (MAE_I2SD_Q90)	17	103.96	103.73	103.76	103.97	104.08
W-model	17	88.93	89.24	89.35	89.54	90.09
P-model (10^4)	17	88.76	97.68	102.18	99.74	93.04
A-model (+)	17	53.49	92.72	102.91		

Table 21: $\overline{\text{MAE}}_{17}$ of the best models from each model type for different k.

Model name	m	$\overline{\text{MSE}}_{17}$ (k=1)	$\overline{\text{MSE}}_{17}$ (k=6)	$\overline{\text{MSE}}_{17}$ (k=12)	$\overline{\text{MSE}}_{17}$ (k=18)	$\overline{\text{MSE}}_{17}$ (k=24)
LM-model	17	98498.98	98369.63	98080.22	97942.95	97892.23
V-model (MAE_I2SD_Q90)	17	97492.82	97492.1	96565.66	96371.02	96247.94
W-model	17	76402.41	77018.04	76484.83	76523.74	77009.31
P-model (10^4)	17	75512.31	90529.88	95117.3	94270.08	81213.1
A-model (+)	17	101795.13	103619.02	92469.09		

Table 22: $\overline{\text{MSE}}_{17}$ for the best models from each model type for different k.

Model name	m	$\overline{\text{I2SD}}_{17}$ (k=1)	$\overline{\text{I2SD}}_{17}$ (k=6)	$\overline{\text{I2SD}}_{17}$ (k=12)	$\overline{\text{I2SD}}_{17}$ (k=18)	$\overline{\text{I2SD}}_{17}$ (k=24)
LM-model	17	0.97	0.97	0.97	0.97	0.97
V-model (MAE_I2SD_Q90)	17	0.95	0.95	0.95	0.95	0.95
W-model	17	0.96	0.96	0.96	0.96	0.96
P-model (10^4)	17	0.96	0.95	0.95	0.95	0.96
A-model (+)	17	0.93	0.99	0.99		

Table 23: $\overline{\text{I2SD}}_{17}$ for the best models from each model type for different k.

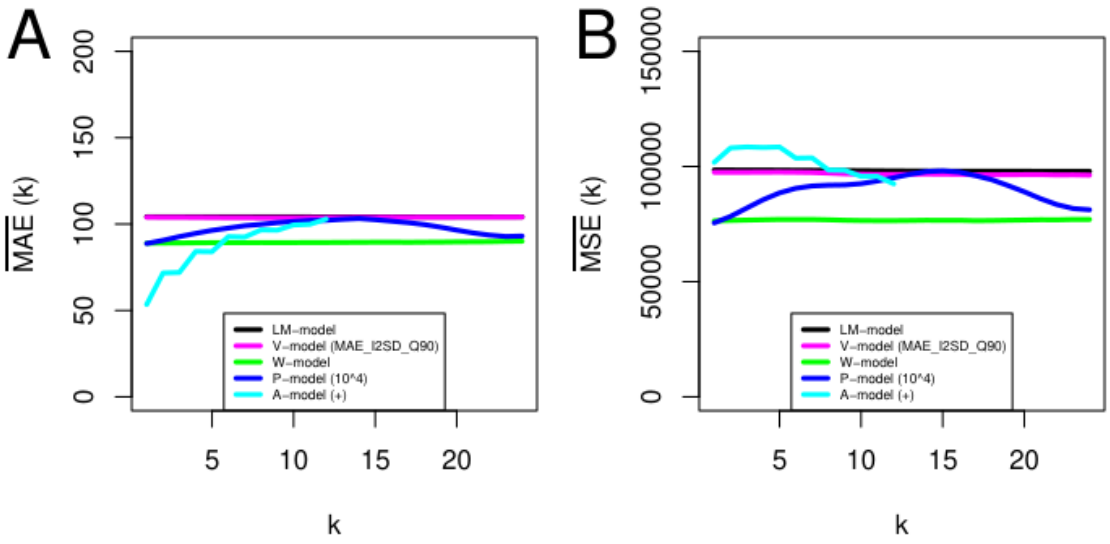


Figure 42: $\overline{\text{MAE}}_{17}$ (A) and $\overline{\text{MSE}}_{17}$ (B) for k-step ahead predictions for different k and different model types (mean over all 17 included pumps). Note that the lines for the LM-model and the V-model overlap.

Tables 21-22 and Fig. 42 show how the $\overline{\text{MAE}}_{17}$(k) and $\overline{\text{MSE}}_{17}$ for a k-step-ahead prediction range from k=1 to k=24 hours. Note that, for the A-model, because of computational (RAM) limitations only values for up to k=12 could be calculated. With respect to the $\overline{\text{MAE}}_{17}$, the A-model performs best up for to 5-step ahead predictions (Fig. 42A). After that, predictions get worse as the AR-effect on the prediction reduces to zero the farther ahead one moves from currently known data. The P-model performs as well as the W-model for 1-step-ahead predictions, but interestingly, not after that (Fig. 42A). The reason may be that the P-model θ estimates are actually too closely adapted to the data. This means they capture more data variation but are less suited for prediction. Note that $\overline{\text{MAE}}_{17}$(k) of the P-model even oscillates. This suggests that the particle estimator perhaps even captures some daily seasonality, so that paradoxically, the estimator could be used to predict 1 and 24 hours ahead, but should not be used for predictions in between. Of course, this makes the P-model less attractive overall. Finally, the V-model is very similar to the LM-model, which is expected because they are very similar in structure. As expected, the V-model performs worse than the more flexible W-model (Fig. 42A).

The results for the $\overline{\text{MSE}}_{17}$ are similar: the W-model performs best (Fig. 42B). However, here the A-model shows a very strange behaviour: its $\overline{\text{MSE}}_{17}$ is larger for small k and decreasing with increasing k (Fig. 42A). It appears that, with respect to the $\overline{\text{MSE}}_{17}$, the autoregressive component actually worsens predictions, and losing the AR-effect for increasing k improves predictions again. This suggests two things: (1) The MAE, while robust, is perhaps not a good measure for the prediction error, at least not on its own, and (2) the A-model as estimated here is not suitable for prediction. This is partly due to the model structure as discussed in the previous section, as well as estimation conditions.

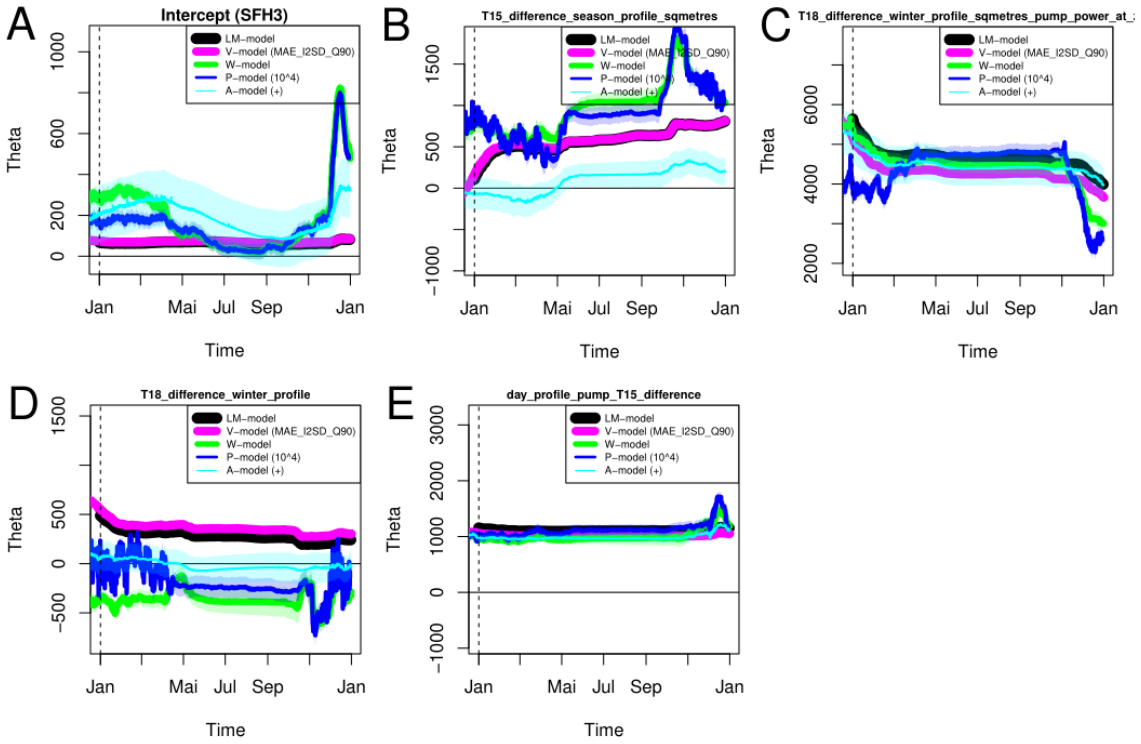


Figure 43: Thetas for different model types. Note that for better readability, in A, only one of the 17 intercepts is shown (that for pump SFH3). B, theta for feature T15_difference_season_profile_sqmetres; C, theta for feature T18_difference_winter_profile_sqmetres_pump_power_at_zero; D, theta for feature T18_difference_winter_profile; E, theta for feature day_profile_pump_T15_difference. The dashed vertical line indicates the beginning of the analysis period (2020).

Fig. 43 shows the estimators for different model types in 2020. Recall that all models use the same features. Yet the estimated θ have very different values (see also Table 14). Those for LM- and V-models are similar, and those for W- and P-models are similar (though P is much noisier), but W-, P- and A-models are very different to the V-model. No forward selection was performed for W-, P- and A-models for computational reasons, and Fig. 43 suggests that this may be a mistake. For example, $\overline{\widehat{\theta}} = 356.68$ in 2020 on average for the V-model but $\overline{\widehat{\theta}} = -359.82$ for the W-model and $\overline{\widehat{\theta}} = -18.95$ for the A-model (Table 14). The models are very similar in structure and the matrix F_t is the same for all, so this huge difference in the estimates could support the assumption that the "wrong" features are being used in the W- and A-models. Another problem may be that the W- and A-models have many more parameters to optimise than the V-model (1 for the V-model, 22 for the W-model and 24 for the A-model). That makes the W- and A-models more vulnerable to optimisation problems such as getting stuck in local minima.

To summarise, the W-, P- and A-models described in this thesis may not yet be the best of their type for structural and estimation reasons. I think that these models could be substantially improved by making the following changes:

- W-model:

  - Try setting $\sigma^2 := \sigma^2_{\alpha,1} = \ldots = \sigma^2_{\alpha,m} = \sigma^2_{\beta,1} = \ldots = \sigma^2_{\beta,p} > 0$. This would probably not be a big problem since all these variances are small compared to σ^2_{pump} and it would reduce the number of parameters to optimise to from 22 to 2 (σ^2_{pump} and σ^2).
  - Use forward selection, possibly based on $\overline{\text{MSE}}_{17}$ instead of $\overline{\text{MAE}}_{17}$, to identify optimal features for the W-model, instead of using the features selected by the V-model. If only two parameters are optimised during each feature trial, optimisation should be fast enough to make forward selection feasible.
- P-model:
  - Use forward selection, or at least features selected for the W-model instead of the V-model. (Since the P-model has the same structure as the W-model, using features selected for the W-model would be justified.)
  - Try different distribution assumptions, because the normal distribution assumption does not appear to be justified (e.g. Fig. 35). (This is easiest to test with the particle filter, since the particle filter does not require specific distribution assumptions.)
- A-model:
  - Add a $\tilde{v}_t \sim N_m(0_m, \tilde{V})$-term to Eq. 7.12 to avoid a shift in predictions (see also section 4.7.2). Use $\tilde{V} = \text{diag}_m(\sigma^2_v)$ to add only one new optimisation parameter.
  - Try setting $\sigma^2 := \sigma^2_{\alpha,1} = \ldots = \sigma^2_{\alpha,m} = \sigma^2_{\beta,1} = \ldots = \sigma^2_{\beta,p} > 0$. For the A-model, this would reduce the number of parameters to optimise to from 24 to 5 (σ^2_{pump}, σ^2_v, σ^2, ϕ_1 and ϕ_2).
  - Use forward selection to identify the best features for the A-model.
  - Use ARMA(2,2) instead of AR(2) noise (Fig. 38).

In conclusion, we have presented five types of models that can be used to model and predict heat pump power consumption data. Of these, the W-model performs best in terms of $\overline{\text{MAE}}_{17}$ and $\overline{\text{MSE}}_{17}$ for predictions up to 24 hours ahead. However, there are still many improvements that could be made, both in the structure of the models and in their estimation, that should improve predictions even further.

8 Summary

Heat pumps play an important role in renewable energy policy because instead of fossil fuels they use up to 75% environmental heat for space heating (Robert Bosch GmbH 2022). However, heat pumps are still operated using electrical power (Robert Bosch GmbH 2022). In this project, hourly electrical power consumption data of 38 heat pumps from households in Hamelin, Germany (Schlemminger et al. 2002), is used to study, model and predict heat pump power consumption in the years 2019 and 2020.

Analysis of external and internal variables, such as temperature, seasons, pump load profiles and household-specific features demonstrates that (1) heat pump power consumption, like temperature, shows seasonal behaviour which is dominated by periods of 1 year and 1 day; (2) the mean centered daily load profiles of heat pumps can be sorted into distinct clusters, which may reflect pump usage patterns or technical specifications; and (3) after removal of yearly and daily seasonalities, structure is visible in the remaining noise, including autoregressive effects (chapter 5). Based on this analysis, 145 potential predictor variables and interactions are created that may affect heat pump power consumption.

A preliminary study using linear models (LM) and ordinary least squares (OLS) estimation shows that (1) 17 of 38 pumps can be fit simultaneously; (2) these 17 pumps do not need to be fit in separate clusters; (3) only 48 of 145 available features are relevant for heat pump power consumption; and (4) only four rounds of forward selection are necessary to determine the most relevant features for a model (chapter 6). The goodness of model predictions can be described using the mean MAE (median absolute error) and MSE (mean squared error) of the 17 pumps fit simultaneously in a given model.

Based on the above findings, different variations of a state space model are developed and tested (chapter 7). The "V-model" has the same structure as a linear model and the true feature effects are assumed to be constant, but are now estimated using a Kalman filter. Because the Kalman filter is more flexible than OLS estimation, the V-model outperforms a simple linear model, but not by much (V: $\overline{\text{MAE}}_{17} = 103.96$, $\overline{\text{MSE}}_{17} = 97492.82$ vs. LM: $\overline{\text{MAE}}_{17} = 104.21$, $\overline{\text{MSE}}_{17} = 98498.98$ for 1-hour-ahead predictions). Forward feature selection with the V-model yields three temperature- and season-profile features whose weighted sum predicts the long-term seasonal behaviour apparent in heat pump power consumption data, while the fourth selected feature is based on mean cluster load profiles and adds daily variation.

An extension of the V-model is the "W-model": a dynamic linear model where feature effects have a variance. For computational reasons, no forward selection is performed; instead, the features discovered for the V-model are used. Because of the W-model's greater flexibility, the W-model has much smaller prediction errors than the V-model (W: $\overline{\text{MAE}}_{17} = 88.93$, $\overline{\text{MSE}}_{17} = 76402.41$). The "P-model" is the same as the "W-model", but a particle filter is used for estimation instead of a Kalman filter. However, this does not substantially reduce the prediction error as compared to the W- model (P: $\overline{\text{MAE}}_{17} = 88.76$, $\overline{\text{MSE}}_{17} = 75512.31$).

Finally, the "A-model" is an extension of the "W-model" that incorporates AR(2) noise. The A-model paradoxically performs better than the W-model with respect to the MAE but worse with respect to the MSE (A: $\overline{\text{MAE}}_{17} = 53.49$, $\overline{\text{MSE}}_{17} = 101795.13$). This suggests that the model specification and estimation conditions for the A-model are not yet optimal. Possibly, an ARMA-model should be used to describe the noise, feature selection should be performed per model type instead of relying on features found by the V-model, and optimisation should be more reliant on the MSE than the MAE. These are avenues that can be explored further in the future. For the moment, the W-model can be used to predict heat pump power consumption for at least up to 24 hours ahead.

References

Cappe, O., S. Godsill, and E. Moulines (2007). "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo". In: *Proceedings of the IEEE* 95, pp. 899–924. DOI: 10.1109/JPROC.2007.893250.

Chen, T., J. Morris, and E. Martin (2004). "Particle filters for the estimation of a state space model". In: *European Symposium on Computer-Aided Process Engineering-14*. Ed. by A. Barbosa-Póvoa and H. Matos. Vol. 18. Computer Aided Chemical Engineering. Elsevier, pp. 613–618. DOI: https://doi.org/10.1016/S1570-7946(04)80168-8.

Dahlin, J. and T. B. Schön (2019). "Getting Started with Particle Metropolis-Hastings for Inference in Nonlinear Dynamical Models". In: *Journal of Statistical Software, Code Snippets* 88.2, pp. 1–41. DOI: 10.18637/jss.v088.c02.

Die Zeit (2022). *Bundesregierung plant Offensive bei Wärmepumpen*. https://www.zeit.de/politik/2022-06/waermepumpen-offensive-energie-bundesregierung-robert-habeck. Retrieved on 26.10.2023.

Fahrmeir, L., T. Kneib, and S. Lang (2013). "The Classical Linear Model". In: *Regression: Models, Methods and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 73–168. ISBN: 978-3-642-34333-9.

Frondel, M., R. Janßen-Timmen, and S. Sommer (2020). *Erstellung der Anwendungsbilanzen 2018 für den Sektor der Privaten Haushalte und den Verkehrssektor in Deutschland*. ger. RWI Projektberichte. Additional information: Forschungsprojekt im Auftrag der Arbeitsgemeinschaft Energiebilanzen e.V. - Endbericht - Mai 2020.

German Federal Statistics Office (2021). *Stromverbrauch der privaten Haushalte nach Haushaltsgrößenklasse [Electricity consumption of private households by household size class]*.

Grolemund, G. and H. Wickham (2011). "Dates and Times Made Easy with lubridate". In: *Journal of Statistical Software* 40.3, pp. 1–25.

Heumann, C., M. Schomaker, and Shalabh (2016). *Introduction to Statistics and Data Analysis*. 1. Edition. Springer.

Logarithmo GmbH und Co. KG (2022). *OpenMeter*. https://appstore.logarithmo.de/app/openmeterplatform/v1/demo/page-datenuebersicht?lang=DE. Retrieved on 22.11.2022.

Miyamoto, S. (2022). *Theory of Agglomerative Hierarchical Clustering (Behaviormetrics: Quantitative Approaches to Human Behavior, 15)*. 1. Edition. Springer.

Petris, G. (2010). "An R Package for Dynamic Linear Models". In: *Journal of Statistical Software* 36.12, pp. 1–16.

Petris, G., S. Petrone, and P. Campagnoli (2009). *Dynamic linear models with R*. Springer New York. ISBN: 978-0-387-77238-7.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Robert Bosch GmbH (2022). *Stromverbrauch von Wärmepumpen*. https://www.bosch-thermotechnology.com/de/de/wohngebaeude/wissen/heizungsratgeber/waermepumpe/stromverbrauch-waermepumpe/. Retrieved on 12.12.2022.

Rosenkranz, A. (2021). *Legionellen im Leitungswasser verhindern*. https://www.heizung.de/ratgeber/diverses/legionellen-im-leitungswasser-verhindern.html. Retrieved on 12.12.2022.

Schlemminger, M., T. Ohrdes, E. Schneider, and M. Knoop (2002). "Dataset on electrical single-family house and heat pump load profiles in Germany". In: *Sci Data* 9.56.

Shumway, R. and D. Stoffer (2017). *Time Series Analysis and Its Applications*. 4. Edition. Springer.

The pandas development team (2020). *pandas-dev/pandas: Pandas*. Version latest. DOI: 10.5281/zenodo.3509134.

Thieurmel, B. and A. Elmarhraoui (2022). *suncalc: Compute Sun Position, Sunlight Phases, Moon Position and Lunar Phase*. R package version 0.5.1.

WetterOnline (2021). *Die WetterOnline Wetter-API - weltweite Wetterdaten in JSON und XML*. https://wo.wetteronline.de/produktwelt/website/. Retrieved on 21.10.2023.

Zenodo Platform (2022). *WPuQ*. https://zenodo.org/record/5642902\#.ZBljl62ZOUm. Retrieved on 21.10.2023.

Appendix

A Household power consumption

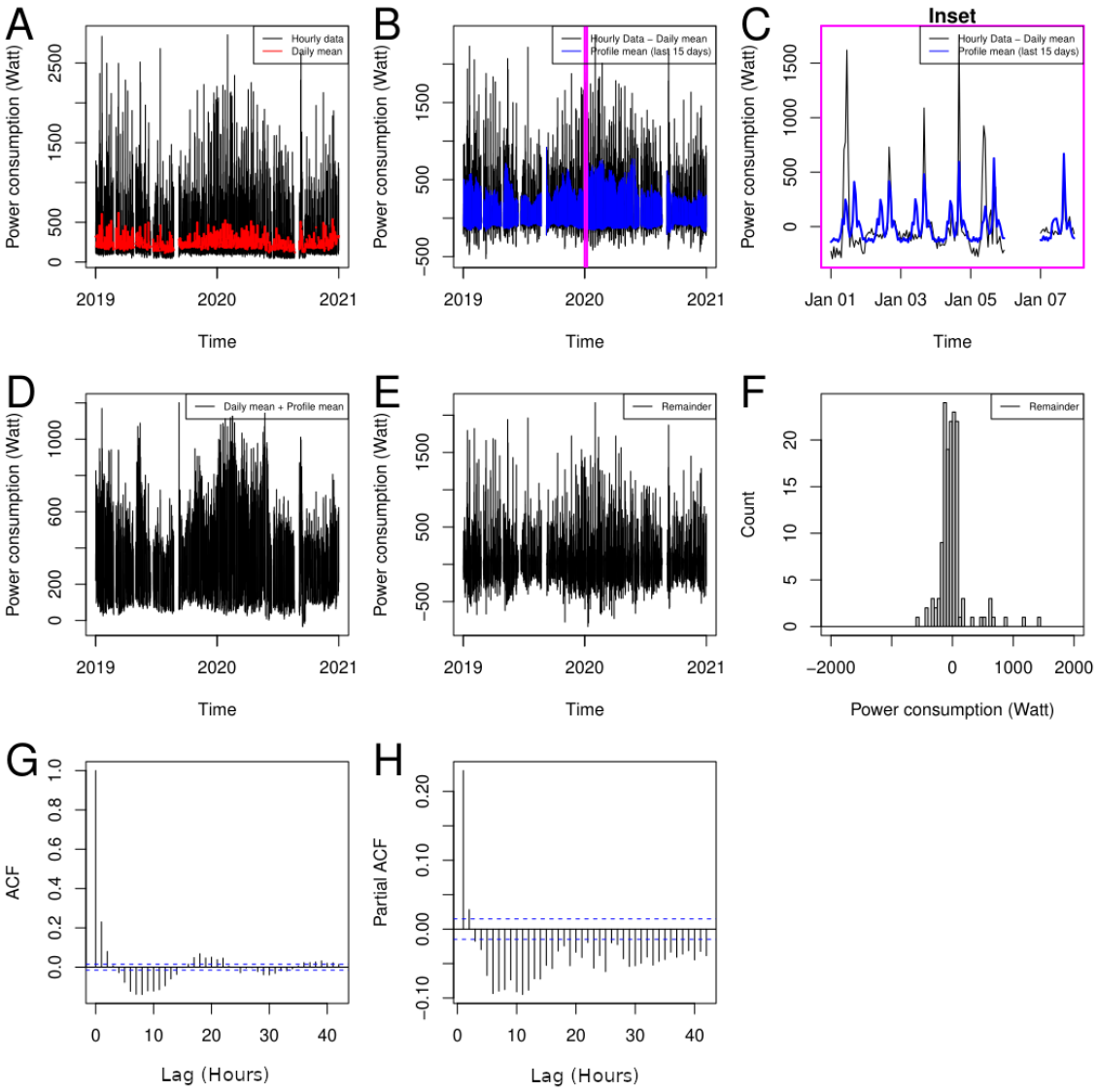


Figure 44: Typical electrical power consumption of a house without its heat pump ($y_{\text{household}}$; house SFH3). A: cleaned data in black, $y_{\text{daily\_mean}}$ in red; B: $y - y_{\text{daily\_mean}}$ in black, $y_{\text{profile\_mean}}$ in blue; C: close-up of the region marked in magenta in B; D: $y_{\text{daily\_mean}} + y_{\text{profile\_mean}}$; E: $y_{\text{remainder}}$, note large peaks; F: histogram of $y_{\text{remainder}}$, note that the histogram is cut off at 1200 Watt, so does not show the largest peaks visible in E; G: ACF of $y_{\text{remainder}}$; H: PACF of $y_{\text{remainder}}$.

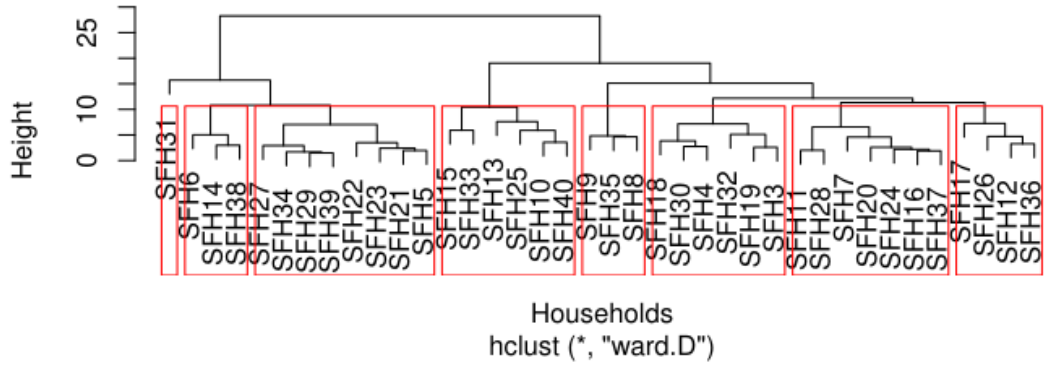
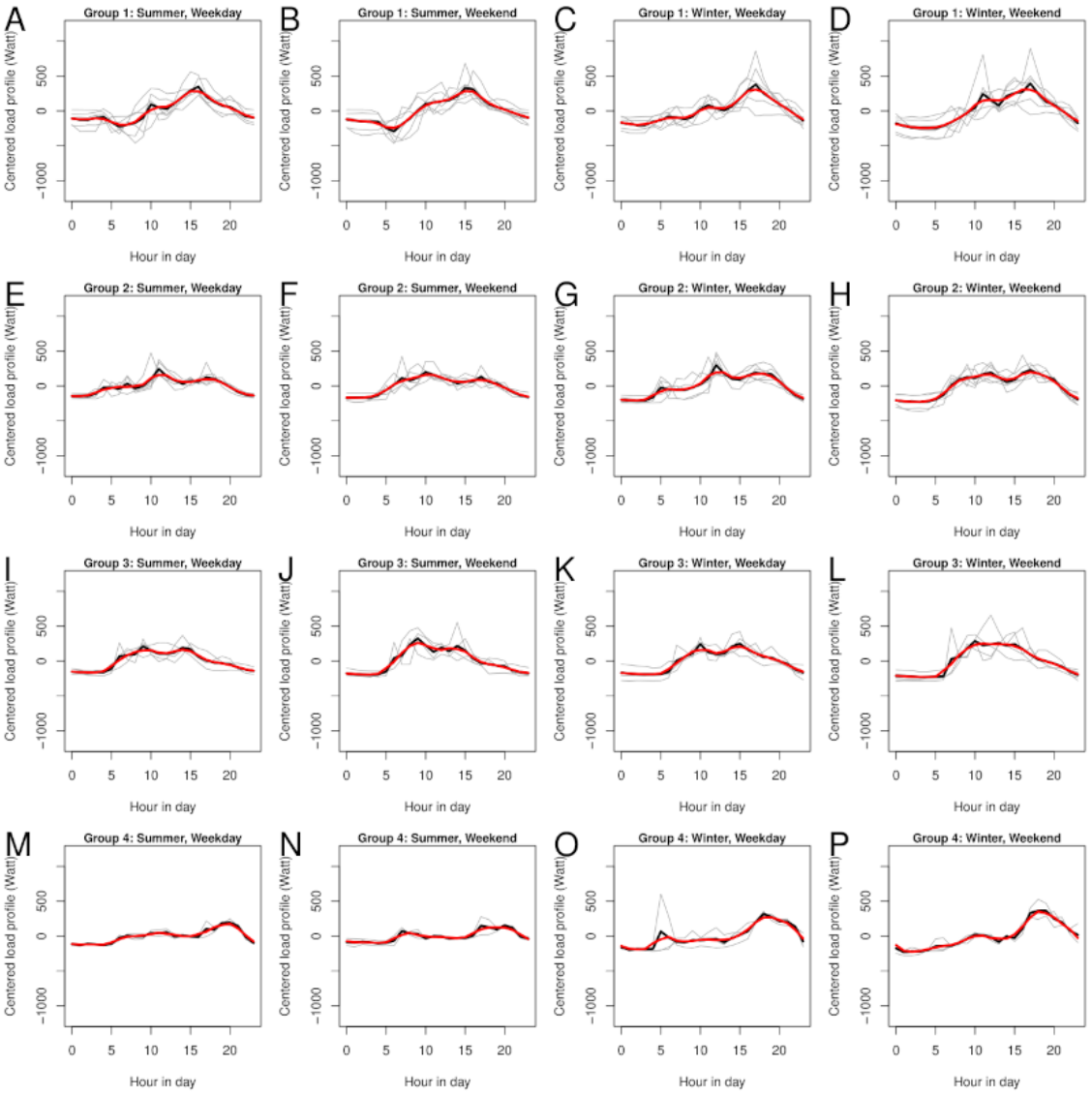


Figure 45: Houses clustered based on mean daily electrical power consumption profiles of their households excluding the pumps.



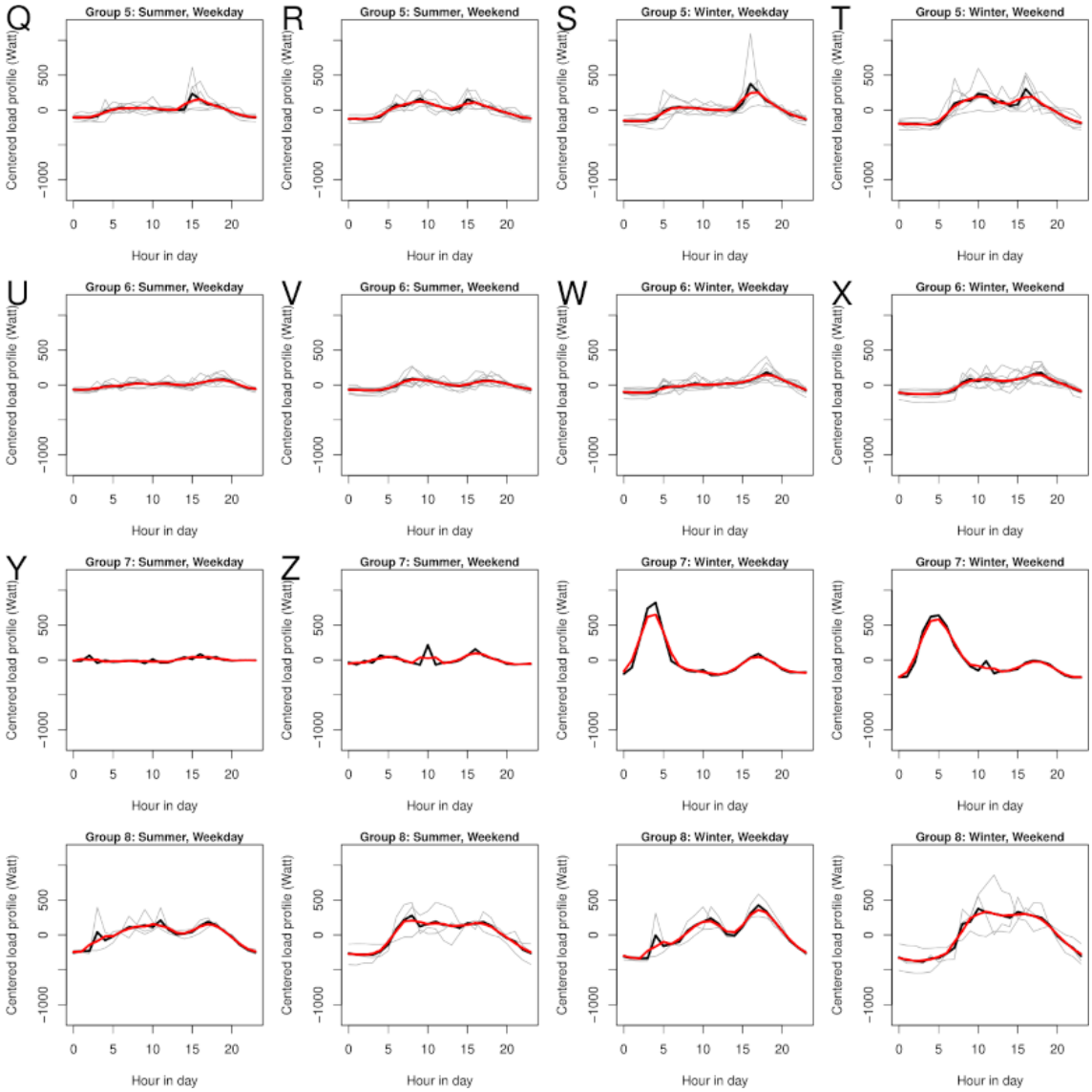


Figure 46: Mean load profiles of household power consumption during the course of a day in summer (first two columns), winter (last two columns), on weekdays (first and third column) and weekends (second and last column) for eight clusters (rows).

selected	cluster	reduced	m	$\overline{\text{MAE}}_{17}$	$\overline{\text{MSE}}_{17}$	r_mean	r_skewness
0	0	0	33	94.36	71822.13	11.40	2.86
0	0	1	33	76.97	18612.25	11.44	0.83
0	1	0	33	92.65	71129.07	11.26	2.80
0	1	1	33	76.04	18461.51	11.26	0.86
1	0	0	33	93.32	71241.57	11.39	2.84
1	0	1	33	76.96	18324.82	11.09	0.81
1	1	0	33	91.70	70838.47	11.43	2.80
1	1	1	33	73.69	17598.83	11.24	0.87

Table 24: Linear model experiments. Selected: all households are included (0) vs. only selected households are included in the fit (1); cluster: all households are fit simultaneously (0) vs. only households of the same cluster are fit simultaneously (1); reduced: all data is used for fitting (0) vs. only reduced data is used (1); m: the number of households included in the experiment. Outcomes MAE (median absolute error), MSE (mean squared error), r_mean, r_skewness), given as the mean of only those 17 households which are included in all experiments, so that the outcomes of different experiments can be easily compared.

Variable	Frequency
day_profile_household	9
day_profile_household_daylight	8
day_profile_household_power_at_zero	6
day_profile_household_sqmetres	5
day_profile_household_power_at_zero_daylight	3
day_profile_household_persons	1

Table 25: First selected features during forward selection in any experiment, together with the frequency with which they are selected across experiments, for household power consumption data.

Variable	Frequency
weekend	19
day_profile_household	11
Hour8	11
day_profile_household_daylight	9
daylight_hours	9
day_profile_household_power_at_zero	8
Hour4	7
day_profile_household_power_at_zero_daylight	6
Autumn	5
day_profile_household_sqmetres	5
day_profile_household_power_at_zero_sunaltitude	4
Hour12	4
Hour16	4
altitude	3
Hour20	3
Spring	3
maxaltitude	2
season_profile_sqmetres	2
winter_profile2	2
day_profile_household_persons	1
day_profile_household_sqmetres_daylight	1
day_profile_household_sqmetres_sunaltitude	1
day_profile_household_sunaltitude	1
Hour0	1
Summer	1
summer_profile	1
temperature_kelvin	1
Winter	1
winter_profile_sqmetres	1
winter_profile2_sqmetres	1

Table 26: Features which are selected at least once in four rounds of forward selection in any experiments, together with the frequency with which they are selected across experiments, for household power consumption data.

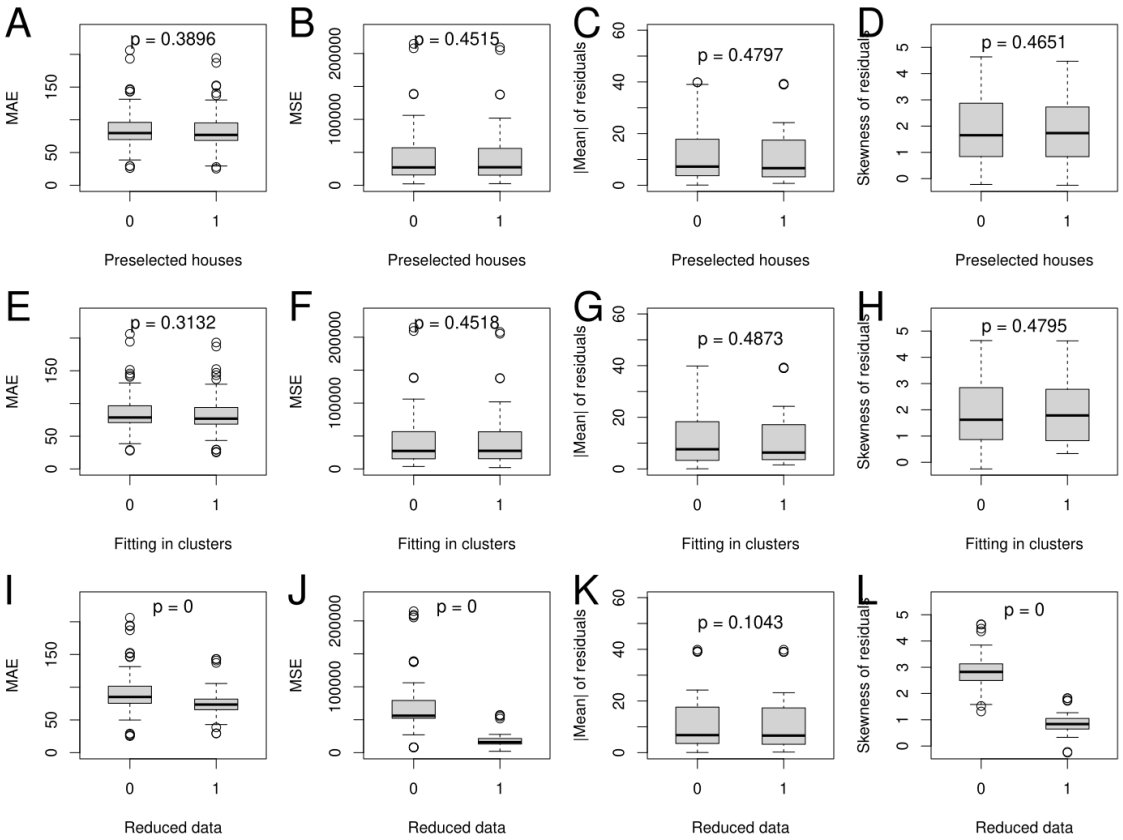


Figure 47: Effects of different fit conditions on MAE, MSE, r and s.

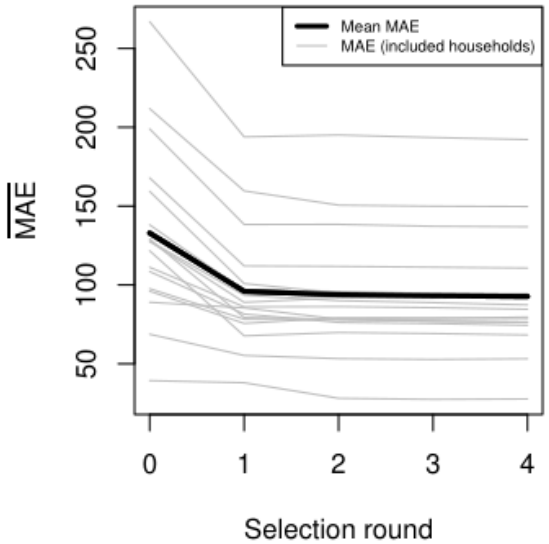


Figure 48: Model selection for the 1|0|0 linear model for household power consumption data.

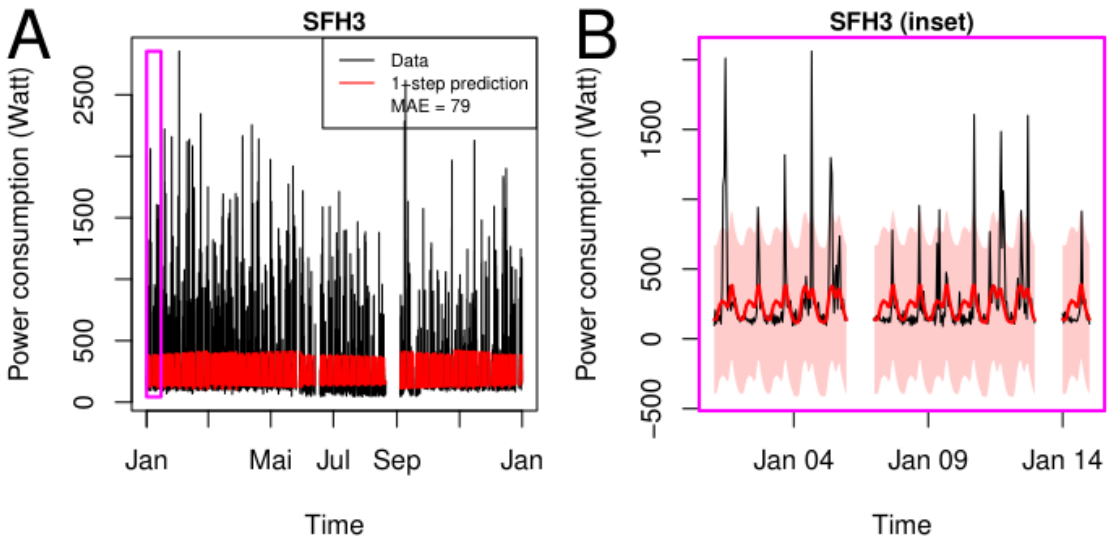


Figure 49: 1-step prediction for 1|0|0 linear model for household power consumption data (example: household 3).

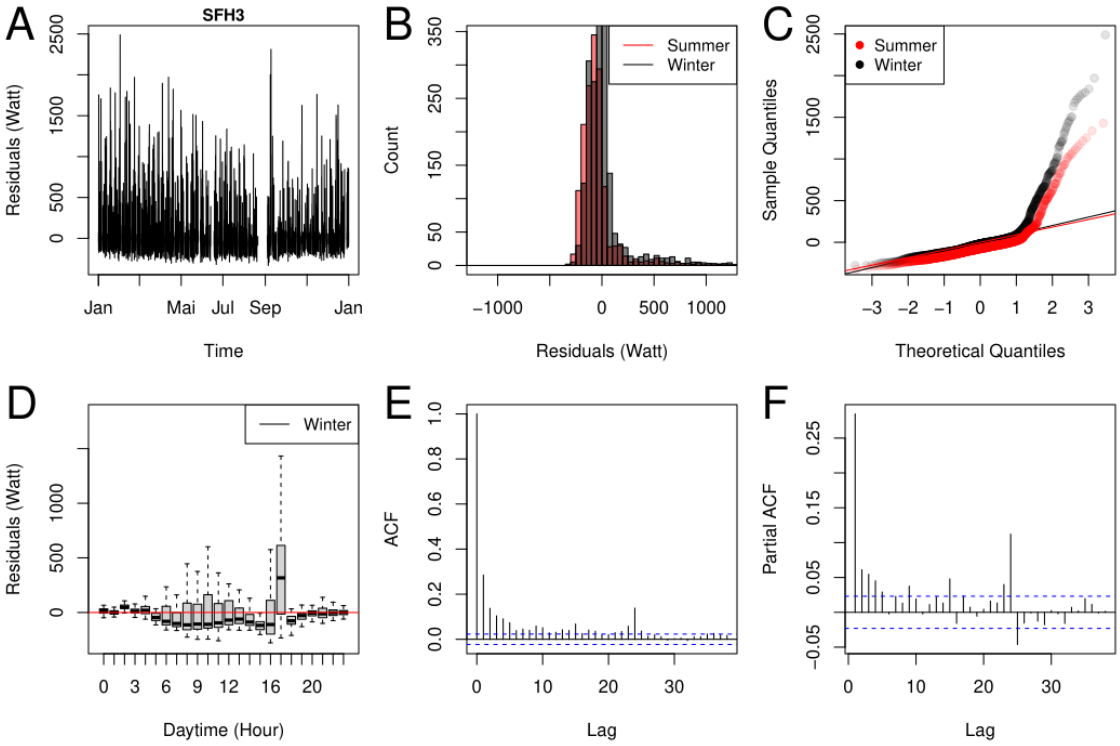


Figure 50: Residual analysis for the 1|0|0 linear model for household power consumption data (example: household 3).

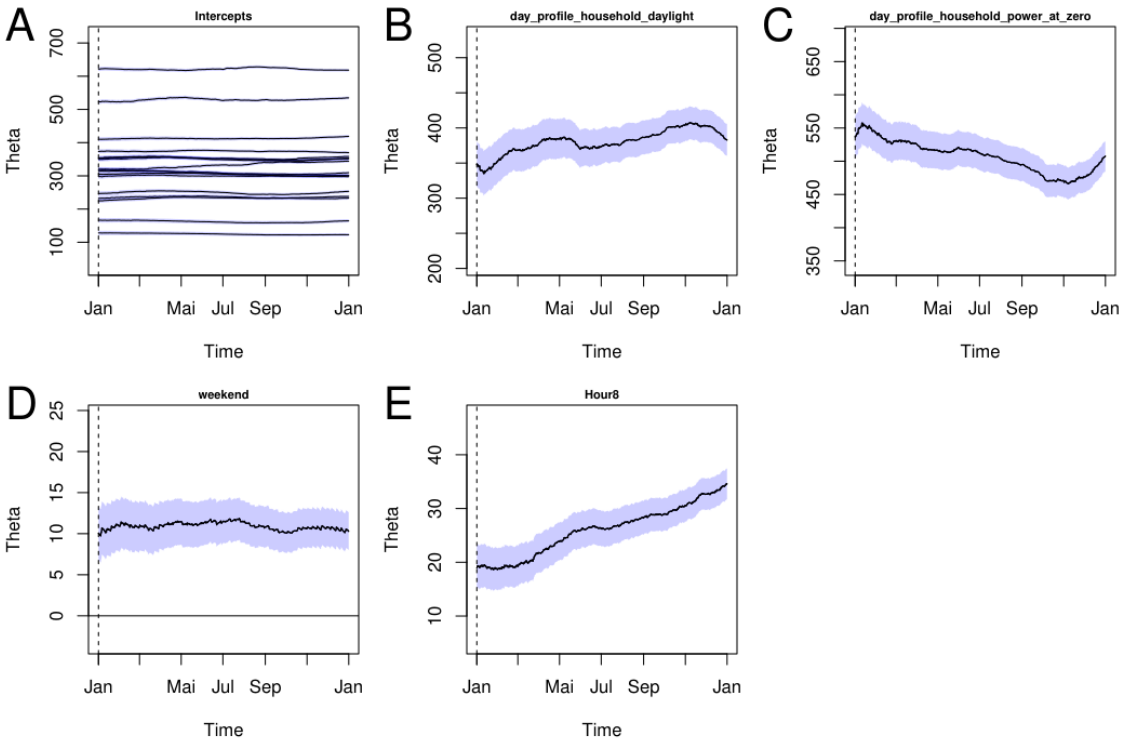


Figure 51: Theta estimates for the 1|0|0 linear model for household power consumption data.

Feature
day_profile_household_daylight
Hour4
weekend
day_profile_household_power_at_zero

Table 27: Features selected by the V-model for household power consumption data.

Model name	m	$\overline{\text{MAE}}_{17}$ (k=1)	$\overline{\text{MAE}}_{17}$ (k=6)	$\overline{\text{MAE}}_{17}$ (k=12)	$\overline{\text{MAE}}_{17}$ (k=18)	$\overline{\text{MAE}}_{17}$ (k=24)
LM-model	17	93.56	93.59	93.59	93.57	93.6
V-model (MAE_I2SD_Q90)	17	98.27	99.28	98.43	97.29	97.99
W-model	17	98.05	99.31	98.32	97	97.81
P-model (10^4)	17	103.27	148.77	176.49	161.05	98.53
A-model (+)	17	83.08	107.6	104.46		

Table 28: MAEs of the best models from each model type for household power consumption data.

Model name	m	$\overline{\text{MSE}}_{17}$ (k=1)	$\overline{\text{MSE}}_{17}$ (k=6)	$\overline{\text{MSE}}_{17}$ (k=12)	$\overline{\text{MSE}}_{17}$ (k=18)	$\overline{\text{MSE}}_{17}$ (k=24)
LM-model	17	71474.73	71499.19	71468.94	71458.54	71482.71
V-model (MAE_I2SD_Q90)	17	71995.41	72757.71	72105.18	71234.38	71830.78
W-model	17	70745.13	71611.56	70943.32	70081.17	70750.33
P-model (10^4)	17	72903.33	103857.03	119006.98	98058.97	71071.29
A-model (+)	17	91812.01	78309.35	72407.74		

Table 29: MSEs of the best models from each model type for household power consumption data.

	V (MAE_I2SD_Q90)	W	A (+)
σ^2_{pump}	72923.66	72922.999967	72923.000016
σ^2_1		1.145805	1.333099
σ^2_2		0	0.941278
σ^2_3		1.515356	0.925855
σ^2_4		0.155041	0.968894
σ^2_5		1.875026	1.129895
σ^2_6		1.128112	1.023865
σ^2_7		0.938157	1.070511
σ^2_8		0.256259	0.862536
σ^2_9		2.249404	1.137614
σ^2_{10}		1.224707	1.051620
σ^2_{11}		0.603326	1.135784
σ^2_{12}		0	0.968070
σ^2_{13}		0.776388	1.097203
σ^2_{14}		1.147777	1.019182
σ^2_{15}		1.629266	0.739525
σ^2_{16}		1.328594	1.298165
σ^2_{17}		1.803883	0.631629
σ^2_{18}		0.963884	0.829717
σ^2_{19}		0.966094	0.451437
σ^2_{20}		0.538686	1.085892
σ^2_{21}		0.324318	0.388784
ϕ_1			0.792010
ϕ_2			0.051122

Table 30: Optimised parameters for different model types for household power consumption data.

Theta	LM	V (MAE_I2SD_Q90)	W	P (10^4)	A (+)
SFH10	627.29±4.86	621.21±4.94	601.28±34.19	602.69±34.11	617.32±89.15
SFH12	337.68±5.15	339.35±4.94	338.89±4.86	343.44±4.86	363.85±81.75
SFH16	255.12±5.12	266.14±4.94	278.06±36.65	277.89±36.56	270.54±81.42
SFH18	315.23±5.09	319.14±4.94	308.07±20.78	306.78±20.76	305.39±82.34
SFH19	379.79±4.91	378.59±4.94	375.12±38.64	376.1±38.53	377.22±85.56
SFH21	314.31±4.87	315.37±4.95	299.94±34.06	299.49±33.98	306.98±83.48
SFH23	358.02±4.9	357.54±4.95	357.01±32.54	356.57±32.47	354.87±84.41
SFH27	130.56±5.3	167.41±4.94	149.3±23.55	150.69±23.53	151.25±79.99
SFH29	167.95±5.7	221.62±4.95	223.06±40.43	223.29±40.32	219.71±85.7
SFH3	240.95±5.3	258.27±4.94	259.19±34.76	258.79±34.68	256.79±84.04
SFH30	239.01±4.83	240.5±4.95	260.28±29.15	260.31±29.1	257.68±85.66
SFH32	305.55±5.11	309.98±4.95	309.51±4.87	319.05±4.87	312.92±82.33
SFH36	309.57±5.04	312.93±4.95	302.49±31.03	304.13±30.99	304.16±84.93
SFH38	418.64±5.14	409.3±4.94	424.11±34.21	424.33±34.13	418.55±83.39
SFH4	358.65±4.93	357.68±4.94	373.39±37.32	373.26±37.22	366.55±76.98
SFH7	356.77±5.18	355.31±4.94	358.39±35.47	358.42±35.39	353.95±88.56
SFH9	535.27±5	518.8±4.95	540.81±38.27	540.14±38.17	527.54±74.02
day_profile_household_daylight	410.49±26.1	430.61±25.41	785.63±69.8	829.6±69.62	588.42±108.25
Hour4	-8.76±3.29	-10.55±3.21	-3.88±25.89	-3.95±25.8	-4.46±25.35
weekend	10.91±2.68	9.27±2.61	11.47±20.39	10.59±20.33	14.77±46.2
day_profile_household_power_at_zero	490.26±26.21	452.65±25.72	138.48±61.25	89.23±61.14	339.65±100.32

Table 31: Thetas obtained for different model types for household power consumption data. The mean is calculated for the year 2020 and two standard deviations are given.

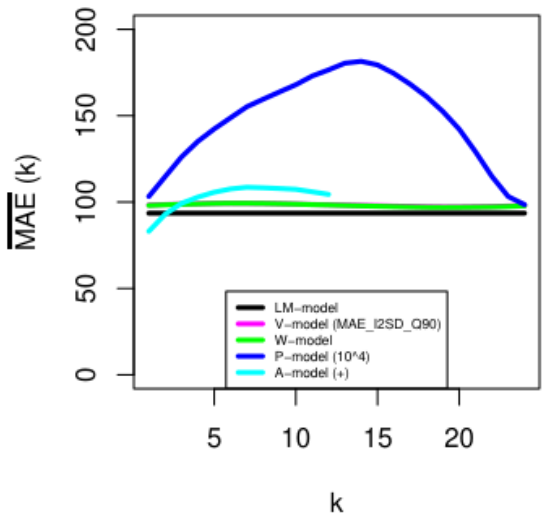


Figure 52: $\overline{\text{MAE}}_{17}$ for k-step ahead predictions for different k and different model types for household power consumption data.

B Miscellaneous

Season	2019	2020
Spring 1	01.03.2019, 00:00 - 31.03.2019, 02:59	01.03.2020, 00:00 - 30.03.2020, 02:59
Spring 2	31.03.2019, 03:00 - 30.05.2019, 23:59	30.03.2020, 03:00 - 30.05.2020, 23:59
Summer	01.06.2019, 00:00 - 31.08.2019, 23:59	01.06.2020, 00:00 - 31.08.2020, 23:59
Autumn 1	01.09.2019, 00:00 - 27.10.2019, 01:59	01.09.2020, 00:00 - 25.10.2020, 01:59
Autumn 2	27.10.2019, 02:00 - 30.11.2019, 23:59	25.10.2020, 02:00 - 30.11.2020, 23:59
Winter	01.12.2019, 00:00 - 31.12.2019, 23:59	01.12.2020, 00:00 - 31.12.2020, 23:59
	01.01.2019, 00:00 - 28.02.2019, 23:59	01.01.2020, 00:00 - 29.02.2020, 23:59

Table 32: Seasons for cluster profile analysis.