Experimental Assessment of the Effect of a Power Conservation Campaign

Henrik Aalborg Nielsen & Henrik Madsen

15 December 1997

IMM

Time Series Group
Department of Mathematical Modelling
Technical University of Denmark

This document was prepared with LaTeX and printed by IMM, DTU, Lyngby.

Acknowledgements

First of all the authors wish to thank the Danish Ministry of Energy for financial support of this project under contract EFP95:1753/95-0001, Danish Energy Research Program. We also wish to thank NESA A/S, Hellerup for supplying data and for executing the large field trial in the area of power savings obtained through information of consumers, which form the basic of the work described in this report.

Special thanks to Conny Thorsted, previously at NESA and now at Ernst & Young, for collecting data and for fruitful discussions, to Per Madsen and Maria Storm Sørensen, both at NESA, for providing valuable information and for helping in the final phase of the project, and to Willy Bergstrøm (NESA), especially, for helping us keeping the overall purpose of the project in mind. We would also like to express our thanks to the members of the project support team Peter Ebdrup and Flemming Grønskov (ELFOR), Johan Hakmann (KB), Preben Munter (SEAS), Jan Møller (DEFU), Henning Parbo (ELSAM), and Jens Rubæk (NVE) for comments and fruitful discussions. Finally, we thank Steffen Torvits (NESA) for guidance during the initial planning of the trial and the household consultants at NESA for the actual execution of the trial by visiting more that two hundred households.

Summary

The purpose of the project described in this report is to develop and to investigate methods for estimating effects of power conservation campaigns, and to estimate the effect of a particular campaign. The statistical approach of experimental design, followed by analysis of measurements is applied.

Using a method known as cluster analysis prior to the initiation of the campaign, a number of substations for which measurements equipment was installed were combined into groups of highest possible similarity with respect to a number of features. It was decided to perform a campaign on a total of two substations. For this purpose the two groups of highest similarity were selected and one substation was chosen from each group to receive the campaign (the active substations). Since each of the two groups contained two substations this procedure left two substations for comparisons (the control substations).

The overall results indicate a 10-12% reduction of the power consumption when carrying out a campaign. However, some difference is observed for the two pairs of active and control substations. Consequently, the extrapolation to other substations of similar kind will be difficult. When measuring the diurnal profile as the relative deviation from the daily level during the cycle, practically no effect of the campaign can be detected. However, the amplitude of the profile measured in units of power consumption depends strongly on the daily level.

It is demonstrated that non-parametric and semi non-parametric methods, combined with traditional time series analysis and bootstrapping, are well suited as statistical tools for the analysis of data from these kinds of experiments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This report forms part of the documentation of the results obtained in a project carried out by NESA A/S, Hellerup and Department of Mathematical Modelling, Technical University of Denmark, Lyngby. The project is partly supported by the Danish Energy Research Program under contract EFP95:1753/95-0001.

The purpose of the work described in this report is to develop and investigate methods for estimating effects of power conservation campaigns. Hereafter the methods are used in order to assess the influence of a particular campaign on the actual power consumption, including diurnal variation. The campaign is aimed at power conservation in households. Due to practical and economic constraints the type of measurements obtainable is the power consumption over 15 minutes intervals for particular groups of households, corresponding to substations for which measurement equipment already has been installed. Furthermore, the campaign can only be applied to a very limited number of groups (two).

It is obvious that prior to the start of the trial control groups, corresponding to the groups for which the campaign is to be applied, should be assigned. These control groups should in ways, which are described in Chapter 2, resemble the campaign groups.

All of the above lead to a quite limited number of candidate groups for the campaign, and a very limited number of groups for which the campaign is actual applied. Consequently, inferential problems arise if the results are to be extrapolated to Danish households in general. Therefore the approach of this report will be to limit the conclusions to the type of households for which the campaign is actual applied, and leave it to the experienced power system professional to make extrapolations to other types of households.

In Chapter 2 the basis of the trial is described. This includes a description of the design of the trial, demography, climate data, and measurements of power consumption at substation level. Furthermore, corrections applied to the measurements of power consumption are considered. Chapter 3 describes the main statistical meth-

ods used. In Chapter 4 the actual analysis of the data is described and the results presented. Readers not interested in details about the methods used should skip Chapter 3. Probably, besides the conclusion and discussion, some readers may want only to read Section 4.5.2, and maybe skip the subsection on bandwidth selection. For readers not familiar with statistical terms we mention that the 95% confidence intervals used in Section 4.5.2 is an expression of the uncertainty of the estimates, e.g. of the trend component, and that bootstrapping is the method we apply to obtain these.

For the sake of cross-referencing figures, tables, and equations are numbered, although not all equations have numbers. The numbers are separated by a decimal point. The number to the left of this point is the number of the chapter in which the figure, table, or equation is placed and the number to the right is a running number, within the chapter and category. For appendices capital letters are used instead of numbers.

Chapter 2

Trial Setup and Data

2.1 Setup

Prior to the startup of the trial the power consumption, during the period September 1995 until January 1996, of 24 substations was analyzed in order to identify substations with similar patterns and sizes of consumption, cf. Appendix D. Hereafter it was decided to conduct the campaign at two substations. The NESA ID numbers for these substations and their corresponding control substation are displayed in Table 2.1. The corresponding locations are listed in Table 2.2. Table 2.3 displays the number of households and the expected yearly power consumption for the four substations. The expected yearly power consumption is calculated for every household supplied by the electrical grid, as the observed consumption during the past year adjusted for climate variations.

Type	Active	Control
0	4284	1325
1	1667	2588

Table 2.1. *ID numbers for active and control substations.*

ID No.	Location	Year
1325	Furesøvej 84-86, Virum	1954
1667	Krogholmgårdsvej 51, Trørød	1960
2588	Havlykkevej 36, Herlev	1965
4284	Mørbjergvænget near # 64, Vindinge	1971

Table 2.2. *Locations and year of establishment of the substations.*

A campaign aimed at reducing the power consumption in the households was executed in the late summer and in the fall of 1996 for the substations termed "active" in Table 2.1. The campaign and its execution is described in details in (Sørensen 1997).

	Households			Other	
		Electrical heating			
ID No.	Total	Absolute	Relative	Absolute	Relative
1325	130	6	5%	1	0.8%
1667	158	15	9%	6	3.8%
2588	219	23	11%	1	0.5%
4284	124	11	9%	4	3.2%
1325	785	141	18%	1	0.1%
1667	1025	219	21%	6	0.6%
2588	1184	321	27%	20	1.7%
4284	683	144	21%	3	0.4%

Table 2.3. *Number of households (top) and expected yearly power consumption in MWh (bottom).*

In general the campaign consisted of intensive information on how to reduce the power consumption. Furthermore, the households were given the opportunity of switching to another tariff structure in which the cost per kWh would change depending on the consumption. The number of participating households and their yearly consumption are displayed in Table 2.4. In total 120 households were visited for substation ID 1667 and 105 for ID 4284, and the dates of the individual visits are displayed in Table A.2. In Figure 2.1 the fraction of the total number of households visited is displayed.

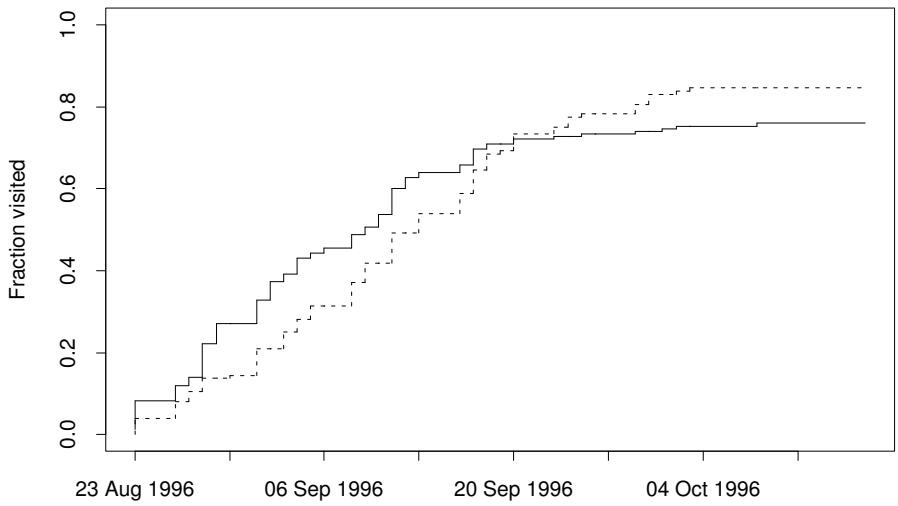

Figure 2.1. *Fraction of total number of households visited for substation ID 1667 (solid) and ID 4284 (dotted).*

	Number		Consumption	
ID No.	Total	Electrical heating	Total	Electrical heating
1667	120	12	865	205
4284	105	9	630	135

Table 2.4. *Participating households: Number and expected yearly consumption in MWh.*

2.2 Climate

The climate is measured as degree days per month. The degree day values are listed in Table A.1. In Figure 2.2 the monthly degree day values are plotted for the period in which measurements of power consumption are available (cf. Section 2.3). For the months July and August the degree day values are assumed to be zero.

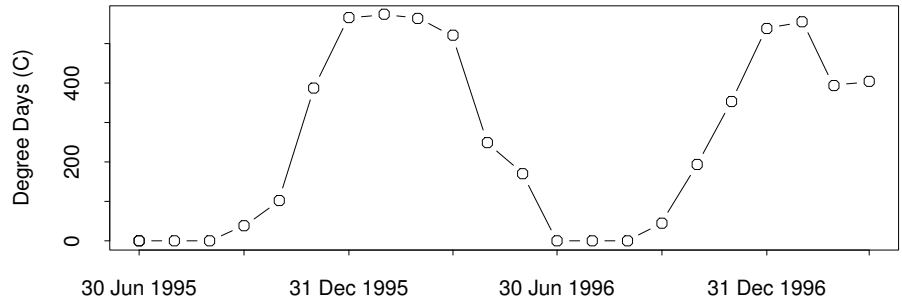

Figure 2.2. *Monthly degree days during the period for which measurements of power consumption are available.*

2.3 Measurements

The power consumption for the four substation shown in Tables 2.1 and 2.2 was recorded over 15 minutes intervals during the months July 1995 until March 1997 (including both), i.e. measurements obtained prior to the execution of the trial is used. The unit of measurement is (KWh/h).

2.4 Preprocessing of Power Data

2.4.1 Erroneous Observations

The 15 minutes values were investigated with the aim of revealing erroneous observations, termed outliers in the following. Two methods were applied

- Logarithm transformed simultaneous values for the four series were plotted pairwise, yielding six scatter plots. Possible outliers were identified by inspecting these plots. Transformed values were used to obtain ellipsis shaped plots.
- The residuals from an ARMA(2,1) model fitted to the individual series were investigated. Standardized residuals with an absolute value greater than or equal to six were taken as possible outliers. The untransformed series were used since this resulted in residuals with more constant variance, than if the logarithm transformed series were used.

In all cases of possible outliers time series plots of the neighbor observations were inspected in order to verify the problem. In many cases the actual problem did occur before the corresponding possible outlier. The decisions following from this analysis are listed in Table 2.5.

Date			Substation			
			1325	1667	2588	4284
17	Aug.	95				13:15
26	Jul.	96				09:15
3	Oct.	96	09:30, 10:30	10:30	09:45, 10:45	11:00, 12:00
1	Nov.	96				11:00, 11:15
13	Jan.	97	12:45	09:45	11:15	10:30
21	Feb.	97		07:15		

Table 2.5. *Decisions related to outliers. Values corresponding to the given periods and substations are all set to missing.*

During the process of outlier, detection a special problem was noted just prior to or after periods with missing values. The first, respectively the last, non-missing value was often small compared to their neighbor values. It is believed that this is caused by the data collection system failing *during* a 15 minute interval. For this reason all neighbor values to periods with missing values were investigated, but no other problems than the ones already identified were revealed.

2.4.2 Final Data Set

To reduce the size of the data set which must be handled during the analysis the data were summed to hourly values. This seems reasonable since this will not corrupt a diurnal variation with a morning and evening peak. Time series plots of the data are shown in Figure 2.3, where, on the right hand side of the plots bars indicate the range of the series with the smallest span (1325). To indicate that the data is given in units of kWh over the past hour, but also may be regraded as an average power (in the sense of physical science), the unit kWh/h is used throughout this report.

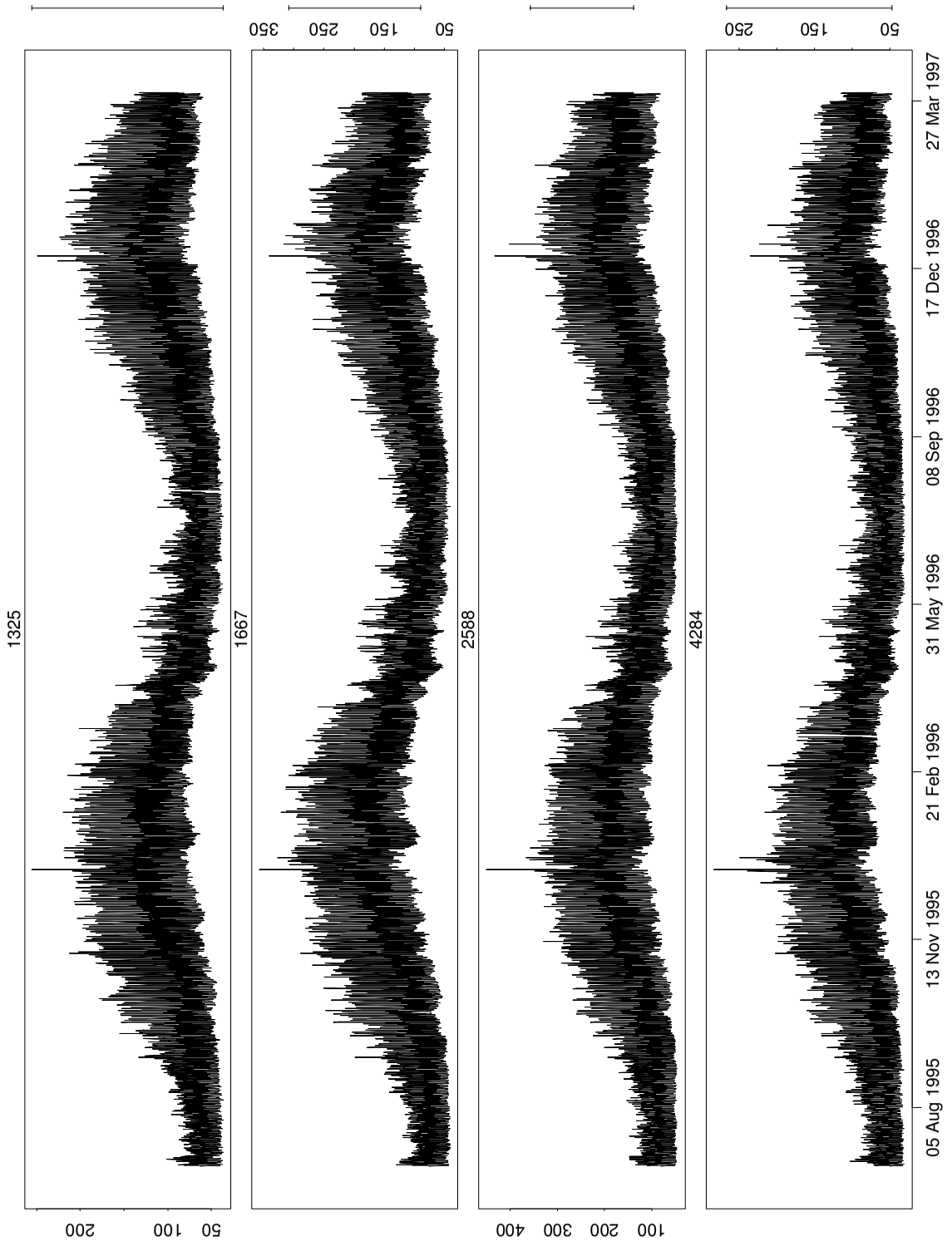

Figure 2.3. *Final data set used in analyses (kWh/h).*

2.5 Demography

In this section some demographic data are tabulated. Table 2.6 shows the distribution of income among persons, whereas Table 2.7 shows the distribution of area among houses. In Table 2.9 cross tabulations of the number of households with specific combinations of the number of adults and number of children are shown. The information in the three tables mentioned so far, were produced by the Danish statistical bureau on 18 Mar. 1996.

Furthermore, the age of the houses supplied by the substations was assessed. This information is displayed in Table 2.8.

ID No.	0 – 74999	75000 – 149999	150000 – 224999	225000 –	No. of persons
1325	51 (19%)	32 (12%)	44 (16%)	142 (53%)	269 (100%)
1667	73 (22%)	36 (11%)	55 (16%)	170 (51%)	334 (100%)
2588	102 (20%)	89 (18%)	130 (26%)	186 (37%)	507 (100%)
4284	74 (23%)	41 (13%)	78 (24%)	126 (39%)	319 (100%)

Table 2.6. *Number of persons (including children) with income 0 – 74999 Dkr., 75000 – 149999 Dkr., 150000 – 224999 Dkr, or above 225000 Dkr.*

ID No.	0 – 59	60 – 79	80 – 99	100 – 119	120 – 159	160 –	No. of houses
1325	1 (1%)	0 (0%)	4 (3%)	27 (21%)	47 (36%)	50 (39%)	129 (100%)
1667	0 (0%)	0 (0%)	4 (3%)	25 (16%)	75 (48%)	51 (33%)	155 (100%)
2588	1 (0%)	5 (2%)	20 (9%)	82 (37%)	85 (39%)	26 (12%)	219 (100%)
4284	0 (0%)	0 (0%)	0 (0%)	0 (0%)	104 (87%)	15 (13%)	119 (100%)

Table 2.7. *Number of houses with area 0 – 59 m^2, 60 – 79 m^2, 80 – 99 m^2, 100 – 119 m^2, 120 – 159 m^2, or above 160 m^2.*

ID No.	
1325	Around 1960–65, but also both older and newer, large and small
1667	Around 1965, a few older and newer, inhomogeneous
2588	Around 1965–70, terrace houses on Havlykkevej
4284	Around 1965–70, quite homogeneous

Table 2.8. *Age and other characteristics of houses supplied by the substations.*

Substation 4284

Adults	**Children**				
Frequency Percent Row Pct Col Pct	0	1	2	≥ 3	Total
1	3	0	0	0	3
	3	0	0	0	3
	100	0	0	0	
	5	0	0	0	
2	38	22	18	6	84
	32	19	15	5	71
	45	26	21	7	
	58	81	95	75	
≥ 3	24	5	1	2	32
	20	4	1	2	27
	75	16	3	6	
	37	19	5	25	
Total	65	27	19	8	119
	55	23	16	7	100

Substation 1325

Adults	**Children**				
Frequency Percent Row Pct Col Pct	0	1	2	≥ 3	Total
1	25	0	0	0	25
	20	0	0	0	20
	100	0	0	0	
	30	0	0	0	
2	45	15	24	3	87
	35	12	19	2	68
	52	17	28	3	
	55	88	96	75	
≥ 3	12	2	1	1	16
	9	2	1	1	13
	75	13	6	6	
	15	12	4	25	
Total	82	17	25	4	128
	64	13	20	3	100

Substation 1667

Adults	**Children**				
Frequency Percent Row Pct Col Pct	0	1	2	≥ 3	Total
1	19	0	2	0	21
	12	0	1	0	14
	90	0	10	0	
	20	0	8	0	
2	64	23	21	8	116
	42	15	14	5	75
	55	20	18	7	
	66	96	88	89	
≥ 3	14	1	1	1	17
	9	1	1	1	11
	82	6	6	6	
	14	4	4	11	
Total	97	24	24	9	154
	63	16	16	6	100

Substation 2588

Adults	**Children**				
Frequency Percent Row Pct Col Pct	0	1	2	≥ 3	Total
1	25	1	0	0	26
	12	0	0	0	12
	96	4	0	0	
	17	3	0	0	
2	101	17	30	3	151
	46	8	14	1	69
	67	11	20	2	
	68	53	88	100	
≥ 3	23	14	4	0	41
	11	6	2	0	19
	56	34	10	0	
	15	44	12	0	
Total	149	32	34	3	218
	68	15	16	1	100

Table 2.9. *Cross-tabulations of the number of adults and children for the four substations. The numbers are absolute and relative numbers of households, as indicated in the top left corner of each cross-tabulation.*

Chapter 3

Statistical Methods

In this chapter the main methods used in the analysis are described. In Section 3.1 a method for decomposing a time series into trend, seasonal, and irregular components is described. The method is very similar to the one described in (Cleveland, Cleveland, McRae & Terpenning 1990). In Sections 3.2, 3.3, and 3.4 the fact that the four time series of power measurements are classified according to a 2-by-2 table is used. In Section 3.2 a simple splitting in type and campaign effects for each point in time is described. In Section 3.3 a method in which the neighbor time points are used in order to split the four time series in type, campaign, and interaction effects, is presented. The method described in Section 3.4 aims at decomposing the time series and splitting e.g. the trend component in main and interaction effects, i.e. the method is a combination of the principles of Section 3.1 and 3.3. The decomposition methods are both closely related to additive models (Hastie & Tibshirani 1990), which are briefly described in Section 3.5. The basic principles of a simulation method, known as bootstrapping (Efron & Tibshirani 1993), is described in Section 3.6 for the case of confidence interval approximation. Finally, in Section 3.7 methods for estimating the autocorrelation function and inverse/partial autocorrelation function for time series with missing observations are described.

We follow Hastie & Tibshirani (1990) and call a non-parametric regression function estimator a *smoother* and the estimate produced by the smoother we call a *smooth*.

3.1 Decomposing a Time Series

This section describes a method for decomposition of a time series. A method was chosen that results in smooth estimates of the trend component and allows for non-stationary, but smooth, changes in the seasonal component. Furthermore, the method allows for different seasonal components for different types of days.

The method is very similar to the STL procedure (Seasonal-Trend decomposition

procedure based on Loess) described in (Cleveland et al. 1990). The STL procedure uses locally weighted polynomial regression smoothers iteratively to decompose a time series. For readers familiar with the STL procedure we mention that the differences are that we apply a convergence criterion in both the inner and outer loop. Furthermore, we do not perform the low-pass filtering of the smoothed cycle-subseries and the post smoothing of the seasonal component. Finally, due to the type of data, we have to allow for different seasonal components for different type of days (weekends, working days).

The algorithm consists of an outer and an inner loop. At startup an estimate of the trend component is calculated by smoothing the time series plot of the data, using a large bandwidth and ignoring the seasonal component. The outer loop down weights observations with large residuals, exactly as described in section 2.4 of (Cleveland et al. 1990). The inner loop is essentially the back-fitting algorithm (Hastie & Tibshirani 1990). In this case it is implemented as follows

Step 1 Subtract the trend component and do seasonal smoothing (see below) on the result.

Step 2 Detrend the seasonal component by smoothing the time series plot of the seasonal with the same smoother as used to estimate the trend component and subtracting the smooth from the seasonal obtained in step 1.

Step 3 Subtract the seasonal component obtained in step 2 from the data and smooth the resulting time series as for the initial estimate of the trend component.

Step 4 Check for convergence and go to step 1 if convergence has not been obtained.

Let x_t denote the original time series with the estimate of the trend component subtracted, let $h(t)$ denote the time of day at the running time t, and let $d(t)$ denote the type of day corresponding to t. The seasonal smoothing is accomplished by smoothing (t, x_t) for each individual element of the pair $(h(t), d(t))$, i.e. no smoothing is performed over neighborhood values of $h(t)$. Consequently, the seasonal component is allowed to change slowly.

For smoothing locally weighted polynomial regression (Cleveland & Devlin 1988) (with fixed bandwidth, tricube window, and linear interpolation between equally spaced points) is used. Thus, the estimates are mainly determined by (i) the span in days of the windows used for the trend and seasonal smoother, (ii) degrees of the polynomials fitted locally, and (iii) a user defined grouping of days. Furthermore, the user might decide not to distinguish between days for some of the 24 hours of the diurnal cycle. As long as the number of equally spaced points in which to calculate the smooth is not too low it is of minor importance. We have found that 50 are appropriate in most situations. However, small window spans require more points.

Let $\widehat{C}_{new}(t_j)$ be the most resent estimate of the component (seasonal or trend) evaluated in the fitting point t_j and let $\widehat{C}_{old}(t_j)$ be the corresponding value at the second most resent iteration. For both seasonal and trend components the following value is calculated

$$\frac{\max_j\{|\widehat{C}_{new}(t_j) - \widehat{C}_{old}(t_j)|\}}{\max_j\{\widehat{C}_{new}(t_j)\} - \min_j\{\widehat{C}_{new}(t_j)\}} \tag{3.1}$$

when these both are below a certain value the iterations in the inner loop are stopped. Unless otherwise stated a value of 0.001, corresponding to a maximum change of 0.1%, is used. For the outer loop the outermost max in (3.1) is replaced by a quantile calculation, where by default the 99% quantile is used.

3.2 Additive Splitting

Suppose that four time series, covering the same period of time, are available and that these are classified according to two factors, each with two levels. In the following these factors will be called type and campaign. For each point in time this corresponds to a two-way analysis of variance (Devore 1991) or (Conradsen 1984), but since only one measurement is available for each combination of type and campaign only the main effects and not the interaction can be estimated, hereof the term "additive splitting". The time series will be denoted by $y_{ijt};\ i = 0, 1;\ j = 0, 1$, with i and j corresponding to type (0, 1) and campaign (0=control, 1=active), respectively.

The factor levels must be coded. The strait forward approach would be to use 0 (no campaign / type 0) and 1 (campaign / type 1). With this setup the design matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \tag{3.2}$$

where the first column corresponds to the level, the second to the type and the third to the campaign effect. However, with (3.2) the estimate of the level is correlated with the estimates of type and campaign effect. This is seen by calculating $(\mathbf{X}^T\mathbf{X})^{-1}$. If instead

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}, \tag{3.3}$$

is used then $(\mathbf{X}^T\mathbf{X})^{-1} = \mathrm{diag}\{1, 1, 1\}/4$. Consequently, the estimates are uncorrelated. Using (3.3) estimates of the overall mean (level) μ_t, the type effect α_t, and

the campaign effect β_t are found as

$$\begin{aligned} \hat{\mu}_t &= (\ \ y_{00t} + y_{01t} + y_{10t} + y_{11t})/4, \\ \hat{\alpha}_t &= (-y_{00t} - y_{01t} + y_{10t} + y_{11t})/2, \\ \hat{\beta}_t &= (-y_{00t} + y_{01t} - y_{10t} + y_{11t})/2, \end{aligned} \tag{3.4}$$

which is seen by calculating $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}_t$, where $\mathbf{Y}_t = [y_{00t}\ y_{01t}\ y_{10t}\ y_{11t}]^T$, and observing that with (3.3) the underlying parameters corresponding to type and campaign effects are half the actual difference, e.g. for the campaign / no campiagn comparison. Let $\tilde{y}_{ijt} = y_{ijt} - \hat{y}_{ijt}$. It is well known that for fixed i then $\tilde{y}_{ijt}$ sums to zero over j, and vice versa. This leaves one degree of freedom to the error, and the model may be written as

$$y_{ijt} = \mu_t + \frac{(-1)^{i+1}}{2}\alpha_t + \frac{(-1)^{j+1}}{2}\beta_t + (-1)^i(-1)^j e_t, \tag{3.5}$$

where $\tilde{y}_{00t}$ is an estimate of e_t. The estimates described above are the maximum likelihood estimates if (i) model (3.5) is true and (ii) for fixed t the observations y_{ijt} are normally distributed with the same variance.

If one of the effects are required to be constant for all t, assuming it is known, it can be subtracted from the data. For instance if α_t is required to be constant α for all t then if the second column in (3.3) is dropped and $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widetilde{\mathbf{Y}}_t$, $\widetilde{\mathbf{Y}}_t = \mathbf{Y}_t - [-\alpha\ -\alpha\ \alpha\ \alpha\ ]^T/2$. It is then seen that μ_t and β_t are unaffected. Consequently, α does only affect the residuals.

3.3 Small Bandwidth Splitting Allowing for Interaction

If more than one observation per time series per point in time were available, then the interaction between the effects could have been estimated. Of course, this is impossible to obtain. Instead the neighbor points in time are used in the hope that the changes are slow. In this case a design matrix for the whole data set are needed, i.e. the number of rows will be four times the number of observations. Each point in time is repeated four times and the effects are coded as described in Table 3.1. The linear model corresponding to this design matrix is then fitted locally in time, with local polynomial approximations of the mean, main effects, and interaction, see (Nielsen 1997). A fixed bandwidth and tricube weight function is used.

With a design matrix as implied by Table 3.1 the estimates will be uncorrelated, if there is no missing data values, or, if some data point is missing then data for all four time series are missing. If the time series has length 100 and one observation is missing, then the off-diagonal elements of $(\mathbf{X}^T\mathbf{X})^{-1}$ (see Section 3.2) are approximately 400 times smaller than the diagonal elements, indicating almost uncorrelated

Overall Mean	Type Effect	Campaign Effect	Interaction Effect	Time Series
1	−1	−1	1	$\{y_{00t}\}$
1	−1	1	−1	$\{y_{01t}\}$
1	1	−1	−1	$\{y_{10t}\}$
1	1	1	1	$\{y_{11t}\}$

Table 3.1. *Coding, when including interaction.*

estimates. However, if a small bandwidth is used the estimates may locally be very correlated, or even undefined if too few observations are available.

In order to minimize the amount of smoothing, whereby the estimates (3.4) are closely reproduced together with the interaction, the bandwidth should be small. For small bandwidths it will be necessary to fit the linear model locally to all points in time for which measurements are available. If the fixed bandwidth is chosen to be e.g. two hours the weight will become zero exactly two hours before and two hours after the fitting point. Consequently, 3×4 observations will be used to estimate the mean and the three effects. If these are approximated by first order polynomials then the weighted least squares problem amounts to 4×2 estimates. For second order polynomials the bandwidth must be increased in order to avoid a perfect fit.

3.4 ANOVA Decomposition

It is natural to combine the concepts of Sections 3.1 and 3.3 in order to split the trend, seasonal, and remainder components into mean, type effect, campaign effect, and interaction.

The same design matrix as described in Section 3.3 is used. Trend smoothing is performed as described in Section 3.3, but here we use a much larger bandwidth and thereby excluding the high frequency variations from the estimate. The seasonal smoothing also uses the design matrix of Section 3.3 and groups the data as described in Section 3.1. However, detrending of the seasonal components is done for the mean, main effects, and interaction individually. Besides these differences the method is identical to the decomposition of one time series, described in Section 3.1.

3.5 Additive Models

Additive models are regression type models in which the expected value of the response is expressed as a sum of functions of explanatory variables, i.e.

$$Y_i = f_1(\mathbf{x}_{1i}) + f_2(\mathbf{x}_{2i}) + \ldots + f_p(\mathbf{x}_{pi}) + e_i, \tag{3.6}$$

where $i = 1, \ldots, N$ is the observation number, Y_i is a stochastic variable denoting the response, $f_\cdot(\cdot)$ is the functions, $\mathbf{x}_{\cdot i}$ is the explanatory variables, each of which may have dimension larger than one, and the errors e_i are independently identical distributed stochastic variables with zero mean. Hastie & Tibshirani (1990) describe these models in detail, also in the case of non-continuous response. Some of the functions may be of known parametric form. The parametric form must be linear in the parameters. If some or all of the functions are unknown some assumptions must be applied to allow estimation. As is common praxis (Hastie & Tibshirani 1990) we will assume that the unknown functions are smooth in the sense that, locally, they can be approximated by low-order polynomials.

Some of the explanatory variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ may have common elements. The important aspect with respect to estimation is that each of the summands on the right hand side of (3.6) relates to markedly different aspects of the expected value of the response. Actually, the decomposition methods described in Sections 3.1 and 3.4 use the same kind of estimation method as is used when estimating in (3.6).

Estimation

Estimation in (3.6) is accomplished using the backfitting algorithm (Hastie & Tibshirani 1990). This method requires initial estimates of all but one of the functions on the right hand side of (3.6). In practice zero is often used as an initial estimate. To update the estimate of a particular function adjusted observations of the response are generated by subtracting the latest estimates of the remaining functions from the original observations. Hereafter the function estimate is updated using the adjusted observations and the explanatory variables relevant for the function under consideration. The method by which this function estimate is updated may be linear regression, if it is a linear model, or locally weighted polynomial regression (Cleveland & Devlin 1988) if it is just assumed to be smooth. As described in (Hastie & Tibshirani 1990) also smoothing splines, or other smoothers, could be applied. In this way each of the function estimates are updated, and the steps are repeated until convergence.

A necessary condition for the uniqueness of the estimates is that each of the functions, except one, sums to a fixed constant (zero) over the observations. This is accomplished by subtracting the mean of the fitted values of the individual functions from the estimates each time the functions are updated. Hastie & Tibshirani (1990) requires all functions to sum to zero, but adds a constant on the right hand side of (3.6).

According to Hastie & Tibshirani (1990) convergence can be guaranteed when using smoothing splines. For locally weighted regression convergence is expected, when the additive model is reasonable as compared to the design. Results presented in this report indicate that the estimation in model (4.11), which is considered in Section 4.5.2, does not converge for the data set used. Different convergence criteria

may be applied. It is natural to use the maximum absolute change in each of the function estimates from one update to the next, and requiring all these to be below a certain value. However, we prefer to measure these changes relative to the total range of the fitted values of the full model. The criterion is not evaluated until all function estimates have been updated equally many times.

Since Least Squares or Weighted Least Squares are used to update the individual functions the noise must be close to normally distributed for the estimation method to be appropriate. However, as in linear regression, the assumption of indepent errors is not strictly necessary, but, generally, the distributional properties described in (Hastie & Tibshirani 1990) will not be valid if the errors are not independently identical distributed.

Equivalent Number of Parameters

The equivalent number of parameters, or degrees of freedom, of model (3.6) is complicated to calculate (Chambers & Hastie 1991, pp. 303-4). Here the same approximation as suggested in the reference just mentioned is used for the individual functions and the sum of these individual degrees of freedom is added to form the equivalent number of parameters.

If the function is linear in the parameters the degrees of freedom are just the number of parameters, possibly minus one, if it is required to sum to a fixed constant over the observations and if this assumption is not build into the structure of the linear model. If the summand is estimated by locally weighted regression the equivalent number of parameters is calculated as $\mathrm{tr}(\mathbf{L})$, where $\mathbf{L}$ is a $N \times N$ matrix denoted the smoother matrix (Cleveland & Devlin 1988), or $\mathrm{tr}(\mathbf{L}) - 1$ if the sum of the function values over the observations is required to be a fixed constant.

Information Criteria

When the number of unknown functions in (3.6) is small it is feasible to calculate some information criteria values for different values of the smoothing constants; degree of local polynomial, bandwidth, and window type (one set for each of the unknown functions). In this report the additive model is applied in a case in which only one unknown function is present. Furthermore, the degree of local polynomial and window type are selected based on general considerations. Hence, only one bandwidth has to be chosen and consequently the information criteria approach is applicable. The following criteria, which all should be minimized, are considered.

Generalized Cross Validation (GCV), originally introduced for splines by Craven & Wahba (1979):

$$\mathrm{GCV} = \frac{N^2 \hat{\sigma}_r^2}{(N - n_{pe})^2}, \tag{3.7}$$

where n_{pe} is the equivalent number of parameters of (3.6), $\hat{\sigma}_r^2 = \sum_{i=1}^{N} r_i^2/N$ and r_i are the residuals of the fit.

Akaike's Information Criterion (AIC), see e.g. (Tong 1990, p. 285):

$$\text{AIC} = N \log \hat{\sigma}_r^2 + 2n_{pe}. \tag{3.8}$$

Note that for $N \gg n_{pe}$ AIC approximately equals $N \log(\text{GCV})$.

Bayesian Information Criterion (BIC) (Schwarz 1978):

$$\text{BIC} = N \log \hat{\sigma}_r^2 + n_{pe} \log N. \tag{3.9}$$

Originally AIC and BIC are likelihood inspired. For this reason it seems appropriate to use $\hat{\sigma}_r^2$ as defined above. Also C_p (Mallows 1973) given by $C_p = \frac{1}{\sigma^2} \sum_{i=1}^{N} r_i^2 - N + 2n_{pe}$, which actually is the criteria considered in (Cleveland & Devlin 1988), could be used. However, it is not considered here as it is not directly applicable, because it, in its basic form, requires knowledge of σ^2, the true but unknown variance of the error.

3.6 Bootstrapping

Bootstrapping (Efron & Tibshirani 1993) is a general method mainly used to calculate confidence intervals of estimates or performing hypothesis tests in cases where other approaches are unavailable. One such example could be an additive model (3.6) in which the errors comes from an autoregressive process, cf. Section 4.5.2. A number of varieties of the method exists. Here we will consider bootstrapping of the residuals with the purpose of calculating confidence intervals.

Assume that a regression type model, possibly an additive model like (3.6), has been fitted using a particular data set. If the residuals of the fit indicate that the errors of the model are independently identically distributed, it will be possible to generate simulated sequences of model errors. Traditionally in bootstrapping this is done by drawing random samples of size N, with replacement, from the residuals of the original fit. This is equivalent to drawing samples from the empirical distribution of the residuals.

Using the original estimates and the simulated sequences of model errors bootstrap replicates of the dependent variable(s) can easily be generated. Hereafter the estimation procedure is carried out for each of the replicates, yielding a number of bootstrap replicates of the estimates. These may be regarded as possible values of the estimates if the model error follow the distribution function used for random sampling, and if the true system is described by the model and the estimates. Consequently, it is possible to use the bootstrap replicates to obtain confidence intervals of the estimates. Precisely how to do this depends on the distributional properties.

If the estimates are identically normally distributed, the most efficient way will be to use the mean and standard deviation of the bootstrap replicates of the estimates, Efron & Tibshirani (1993) denotes this standard normal intervals. If the assumption is doubtful the BC_a or ABC method of Efron & Tibshirani (1993) is often applied. However, these methods are infeasible for the work described in this report, cf. Section 4.5.2 and (Efron & Tibshirani 1993). Consequently, the percentile intervals is the only remaining standard method mentioned in (Efron & Tibshirani 1993) which can be applied in case of non-normality of the estimates. Since this method uses the empirical quantiles only it is uncertain when the desired coverage probability of the confidence interval is close to one, e.g. if 1000 replicates are generated a 95% confidence interval based on the percentile method will be determined by the 25 smallest and 25 largest values.

Generally, the precision of the bootstrap results increases with the number of bootstrap replicates. However, the limiting factor is time, including CPU-time. Also, the appropriate number of bootstrap replicates depend strongly on the normality assumption, and hence it should be considered separately for each application.

3.7 Sample ACF, PACF, and IACF When Some Observations Are Missing.

This section describes how missing observations are treated when estimating the autocovariance/correlation function (ACF), the partial autocorrelation function, and the inverse autocorrelation function. The approach described in (Madsen 1995) is used, but the actual formulas are changed slightly to make the estimators produce the usual estimates in the case of no missing observations.

The autocorrelation function

Let $x_1, x_2, \ldots, x_N$ be the observations from the process $\{X_t\}$, where some observations may be missing. Define

$$a_t = \begin{cases} 0, & \text{if } x_t \text{ is missing} \\ 1, & \text{otherwise} \end{cases} . \tag{3.10}$$

Using the temporary quantities

$$C_a(k) = \frac{1}{N-|k|} \sum_{t=1}^{N-|k|} a_t a_{t+|k|},$$

and

$$C^{\square}(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} a_t a_{t+|k|} (x_t - \bar{x})(x_{t+|k|} - \bar{x}),$$

where

$$\bar{x} = \sum_{t=1}^{N} a_t x_t / \sum_{t=1}^{N} a_t.$$

The estimate of the autocovariance in lag k is

$$C(k) = \frac{C^{\square}(k)}{C_a(k)}. \tag{3.11}$$

It is seen that the method just skips the summands which cannot be calculated due to missing observations, and adjusts the number of observations accordingly. When no observations are missing $C_a(k) = 1$ and

$$C(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} (x_t - \bar{x})(x_{t+|k|} - \bar{x}),$$

which is the usual estimator, cf. (Madsen 1995, p131). If the definition of $C_a(k)$ in (Madsen 1995, p242) is used, then $C_a(k) = (N - |k|)/N$ when no observations are missing. Consequently, $1/N$ in the usual estimator of autocovariance is replaced by $1/(N-|k|)$, and the estimator leads no longer to a non-negative definite autocovariance function, cf. (Madsen 1995, p131).

Based on the estimates of autocovariance (3.11) the sample autocorrelation function is calculated as

$$\hat{\rho}(k) = \frac{C(k)}{C(0)}. \tag{3.12}$$

The inverse autocorrelation function

The inverse autocorrelation function is calculated as in `proc arima` of SAS/ETS (SAS Institute Inc. 1993), except that here missing values are allowed. A high order autoregressive process is fitted to the series using the Yule-Walker equations and the estimates of the autocorrelation function (3.12). Hereafter the autocorrelation function is calculated for the dual process, i.e. a moving average process of the same order as the autoregressive process.

Let k_{max} be the maximum lag for which the sample inverse autocorrelation function is to be calculated. The order of the process is then

$$p = \min\left(\left[\frac{N}{2} \right], k_{max} \right),$$

where $[\cdot]$ denotes the integer value. If $k_{max} > N/2$, the estimates for the lags after $N/2$ is zero.

The partial autocorrelation function

The partial autocorrelation function is estimated based on (3.12) and calculated as described in (Madsen 1995, Appendix B).

Variance of estimates

It is well known (Madsen 1995) that if the observations come from a white noise process the variance of the estimates of all three types of functions is approximately $1/N$. It is thus natural to adjust N to account for the number of observations which cannot be used due to missing values, i.e.

$$V[\hat{\rho}(k)] \approx \left[N \frac{\sum_{t=1}^{N-|k|} a_t a_{t+|k|}}{N-|k|} \right]^{-1}, \tag{3.13}$$

which equals $1/N$ if no observations are missing.

For the SIACF and SPACF this approach is not directly applicable since it is not possible directly to relate the sum in (3.13) to a specific lag. The maximum or minimum over the lags for which the estimates are calculated could be used, depending on the particular application. However, in this report the usual $1/N$ is used, since the method will newer be used anyway if a substantial fraction of the observations is missing.

Chapter 4

Results

In this chapter the results of the analysis are presented. The analysis consists of a number of different approaches. In Section 4.1 the individual power measurement series are decomposed and the results are compared qualitatively. Section 4.2 addresses aspects of standardization of the series. In Sections 4.3 and 4.4 the series are compared quantitatively. Finally, in Section 4.5 the influence of climate is considered.

4.1 Decomposition of the Individual Time Series

The four time series of power measurements are decomposed into trend, seasonal, and remainder components as described in Section 3.1. For the seasonal smoother the data are divided into the three type of days (i) Monday to Friday which are normal working days, (ii) Half-holy days, including Saturdays, and (iii) Holy days, including Sundays. For both trend and seasonal smoothers, the weighting window spans 270 days and locally a second order polynomial in time is fitted. All smooths are calculated at 50 equally spaced points in time, a convergence criterion (3.1) of 0.001 is used in the inner loop and no robustness iterations are performed.

The results are shown in Figures B.1, B.2, B.3, and B.4, placed in Appendix B. On each figure the data, trend component, seasonal component, and remainder are plotted. To the right of each plot a vertical bar indicates the range of the component with the smallest span. The horizontal line at the bottom of each plot indicates the period in which the households were visited, cf. Section 2.1, the same period one year before is indicated by a dotted line.

Also in Appendix B seasonal diagnostic plots, i.e. the data with the trend component subtracted, grouped by time of day and type of day, and plotted against the date, are shown for working days only. The plots indicate that the smooths are appropriate,

c.f. (Cleveland et al. 1990). Plots corresponding to the remaining day groups are not included, but these are very similar to those in Appendix B.

Even though the remainder of the decomposition, clearly, is not white noise some interesting aspects are noted based on the trend and seasonal components. For two of the substations (1667 and 4284) the winter peak is clearly lower in 1996/97 than in 1995/96 and for substation 1667 the seasonal seems to "flatten" for working days during the winter 1996/97. It is certainly interesting to compare these observations with Table 2.1, from which it is revealed that exactly 1667 and 4284 are the substations to which the campaign were applied.

From the plots of the remainder component it is revealed that some irregular small scale variation is present. For this reason the trend component is estimated using a window span of 50 and 100 days also, but still keeping the remaining settings. The seasonal components are largely unaffected by this, although slightly affected around Christmas. All estimates of the trend components are displayed in Figure 4.1. The plots are arranged so that control substations are at the right of the corresponding active substation, see also Table 2.1. On the plots vertical lines indicate the period in which the households were visited, together with the corresponding period one year before. Especially substation 4284 shows low values for the second peak, indicating that the campaign has had the largest effect for this substation.

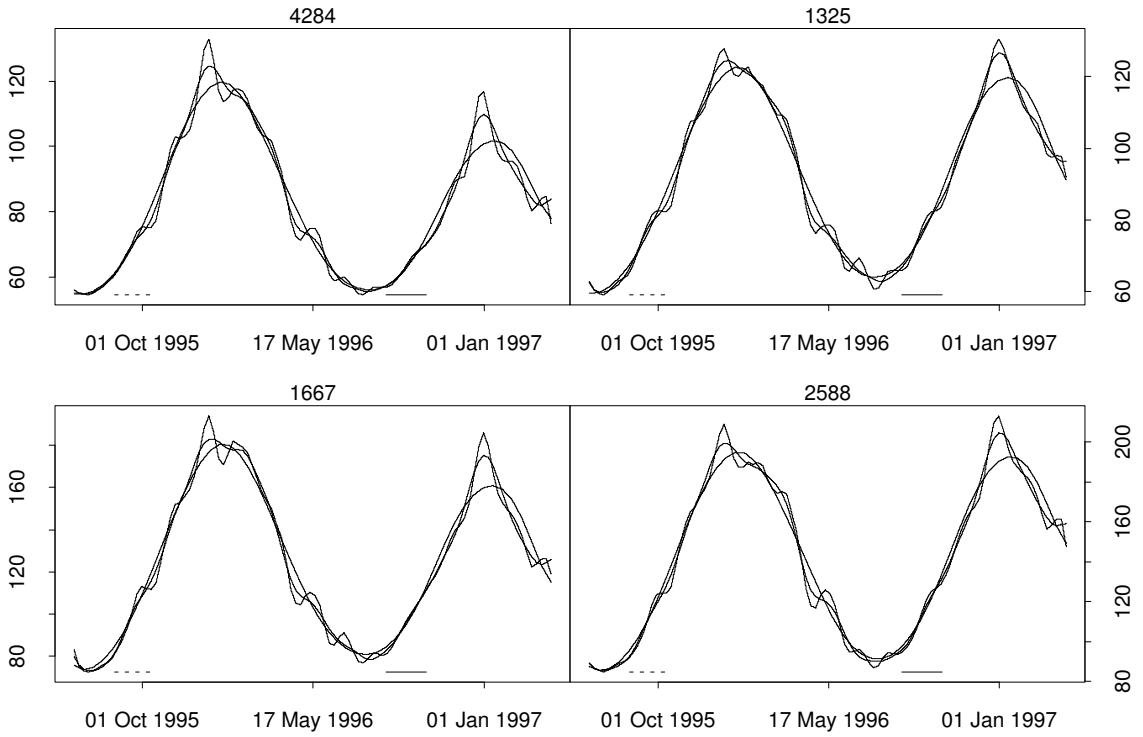

Figure 4.1. *Estimates of the trend component (kWh/h) using a window span of 50, 100, and 270 days.*

4.2 Standardization

In the previous section the four time series were compared qualitatively, e.g. in Figure 4.1 the scale on the y-axes is different. In the following sections the four series will be compared quantitatively. For this reason it is appropriate to discuss if the data should be standardized. Considering Table 2.3 it is natural to consider standardizing each of the four series by either the number of households or by the total expected power consumption of both households and other consumers. However, since the pairs active/control have been found by cluster analysis with emphasis on, size, amplitude of hourly consumption, etc. it seems natural *not* to standardize the series.

In the following section it will be indicated that for both purposes of interpretation and appropriateness of models the data should be logarithm transformed. Furthermore, since the procedures used corresponds to a series of, possibly weighted, least squares problems a standardization will only add a constant to the estimates. This is seen from the following. Let $\mathbf{Y}_t$ be a vector containing values from the four time series for which the weight when considering time point t is positive. Let $\mathbf{X}_t$ be the design matrix corresponding to the model under consideration, and let $\mathbf{W}_t$ be a diagonal weight matrix corresponding to the observations $\mathbf{Y}_t$. If the four series are standardized the logarithm transformed observations may be written $\mathbf{Y}_t - \mathbf{Y}_{std}$, note that $\mathbf{Y}_{std}$ do not depend on t. Estimates corresponding to the standardized series may then be written as

$$\hat{\boldsymbol{\theta}}_{std}(t) = \left(\mathbf{X}_t^T \mathbf{W}_t \mathbf{X}_t\right)^{-1} \mathbf{X}_t^T \mathbf{W}_t \left(\mathbf{Y}_t - \mathbf{Y}_{std}\right), \tag{4.1}$$

this equals

$$\hat{\boldsymbol{\theta}}_{std}(t) = \hat{\boldsymbol{\theta}}(t) + \left(\mathbf{X}_t^T \mathbf{W}_t \mathbf{X}_t\right)^{-1} \mathbf{X}_t^T \mathbf{W}_t \mathbf{Y}_{std}, \tag{4.2}$$

where $\hat{\boldsymbol{\theta}}(t)$ are the estimates corresponding to the original series. From this it is seen that if the matrices $\mathbf{X}_t$ and $\mathbf{W}_t$ do not change for the values of t considered the last part of (4.2) is constant. This will be true for additive splitting (cf. Section 3.2) and small bandwidth splitting with fixed bandwidth (cf. Section 3.3).

Consequently, if inference is based on changes in effects, then it will be independent of standardization.

4.3 Additive Splitting According to 2-by-2 Design

To make a simple assessment of the type and campaign effects the hourly values are split in mean, type effect, campaign effect, and remainder or residual as described in Section 3.2. However, transformation of the data should be considered. If in (3.5) y_{ijt} is the logarithm transformed power consumption $\log(P_{ijt})$ then the fitted value

of the control substations will be

$$\widehat{P}_{i0t} = \exp\left(\hat{\mu}_t + \frac{(-1)^{i+1}}{2}\hat{\alpha}_t\right)\exp(-\hat{\beta}/2) \tag{4.3}$$

while the fitted value of the active substations will be

$$\widehat{P}_{i1t} = \exp\left(\hat{\mu}_t + \frac{(-1)^{i+1}}{2}\hat{\alpha}_t\right)\exp(\hat{\beta}/2) \tag{4.4}$$

Consequently, $\widehat{P}_{i1t}/\widehat{P}_{i0t} = \exp(\hat{\beta})$ and therefore, if $\hat{\beta}$ is negative, $1 - \exp(\hat{\beta})$ is an estimate of the fraction by which the original (no campaign) power consumption may be reduced if a campaign is carried out.

Similar calculations for untransformed data reveal that $\hat{\beta}$ is an estimate of the reduction (when $\hat{\beta}$ is negative) in kWh/h a campaign will have on the power consumption, i.e., under the model, the actual reduction in kWh/h does not depend on the size of the substation. Hence, the result from an analysis using logarithm transformed data has much nicer interpretation. However, in the following both untransformed and logarithm transformed data will be analyzed with the aim of investigating which are the most appropriate from a statistical point of view.

Figure 4.2 shows for $i = 0$ and $j = 0$, i.e. substation 1325, and both untransformed and logarithmic transformed data (i) the residuals plotted against the fitted values, (ii) a normal quantile-quantile plot of the residuals (Statistical Sciences 1995), and (iii) a histogram of the residuals. From these plots it is revealed that an additive splitting of the untransformed data is highly inappropriate. For logarithm transformed data the additive splitting is more appropriate, although it is not perfect. In Figure 4.3 the sample autocorrelation function, of the residuals mentioned above when using logarithmic transformed data, is displayed. As expected, a quite high correlation as well as a seasonal correlation, corresponding to a diurnal variation, is observed.

The actual additive split on the logarithmic scale is displayed in Figure 4.6, together with the residuals (e_t in (3.5)). To the right of each plot a vertical bar indicates the range of the component with the smallest span. From the figure it is seen that, on an average, the campaign effect moves away from zero in the late summer of 1996, which is the time at which the campaign was initiated, cf. Section 2.1. The effect seems to be even more pronounced when comparing the winter 1996/97 with the winter 1995/96. The type effect seems to exhibit annual variation, and possibly on top of this it is increasing during the late summer and early fall of 1996. However, this cannot account for the campaign effect, cf. the comment on constant type effect in Section 3.2.

To assess the significance of the behavior of the estimates of campaign effect initially, the usual F-statistic for the hypothesis of no campaign effect (Devore 1991) has been calculated. These values are shown in Figure 4.4 together with quantiles of the F-distribution on $(1, 1)$ degrees of freedom corresponding to a level of significance

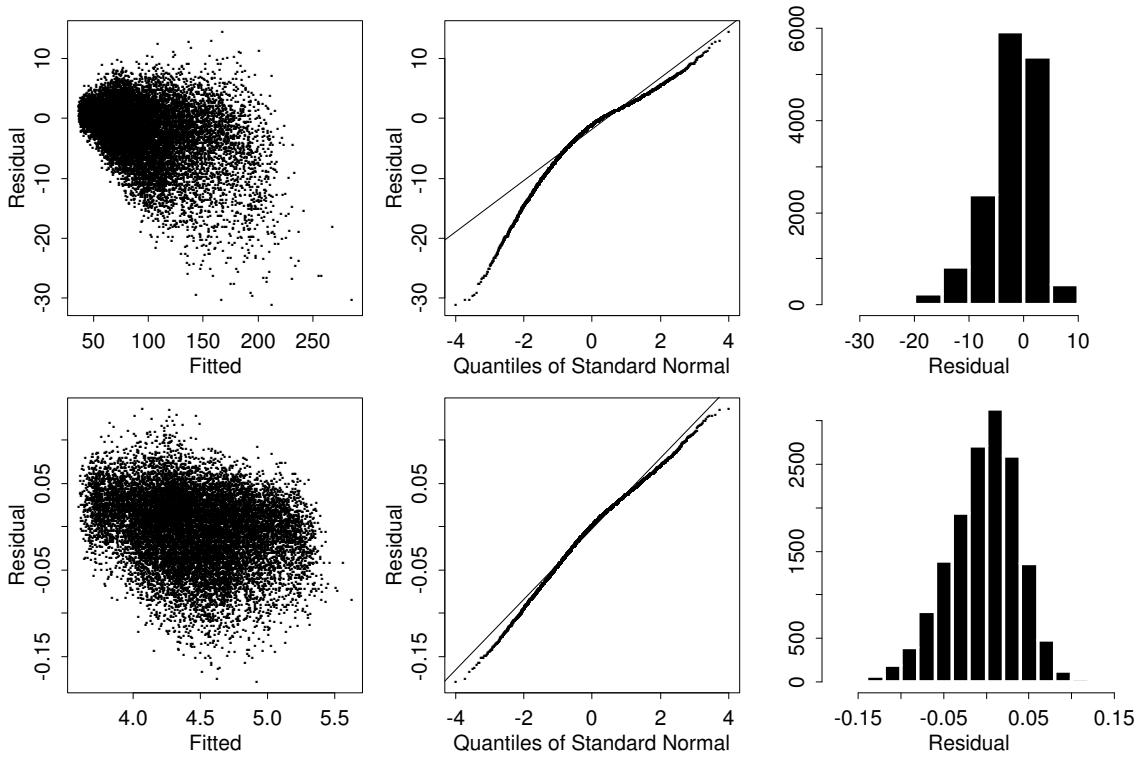

Figure 4.2. *Diagnostics when using untransformed (top row) and logarithm transformed (bottom row) data.*

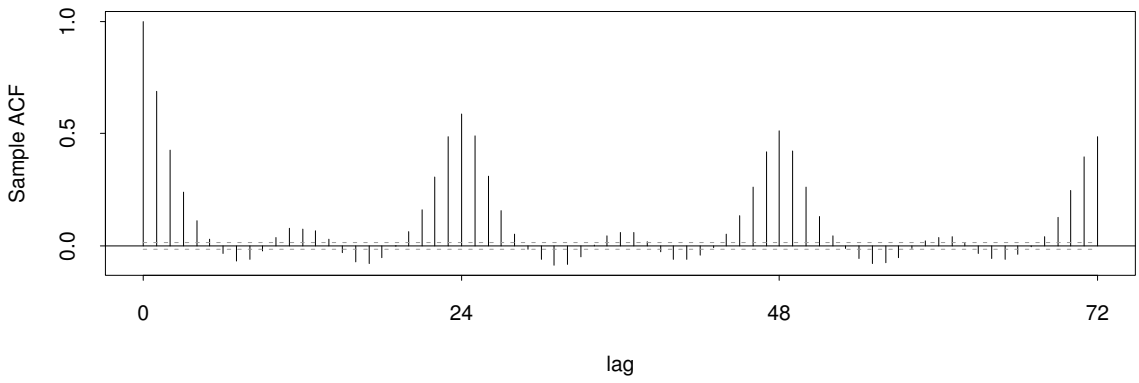

Figure 4.3. *Sample autocorrelation function of the residuals for substation 1325.*

of 1%, 5%, and 50%. It is seem that the F-statistic, during the entire period, often moves above the 5% level of significance, especially at the end of the period. However, the problem of multiple testing is evident in this setting and hence it is doubtful if the usual critical values have any relevance.

If the true campaign effect is zero then the F-statistic, as calculated above, will have median $F(1,1)_{0.5} = 1$, whereas the mean does not exist. From Figure 4.4 it is evident that the F-statistic rarely is below $F(1,1)_{0.5}$ from September 1996 and onwards, indicating a significant campaign effect. This period agrees with the fact that the main campaign efforts have been spent during August 1996, cf. Section 2.1. As in Section 4.1 the difference is most obvious when comparing the winter 1995/96 with the winter 1996/97.

The results presented in this section indicate a significant effect of the campaign, under the assumption of additivity on the logarithmic scale. This assumption seems to be an reasonable approximation, cf. Figure 4.2. Furthermore, the residuals are in general smaller than the remaining effects, cf. Figure 4.6, indicating that the deviation from additivity is small. In Figure 4.5 the campaign component from Figure 4.6 is plotted on the original scale, together with local line smooth using a tricube window and a 2% nearest neighbor bandwidth. Overall, Figure 4.5 indicates a 10% reduction in the campaign effect when comparing the winter 1995/96 with 1996/97.

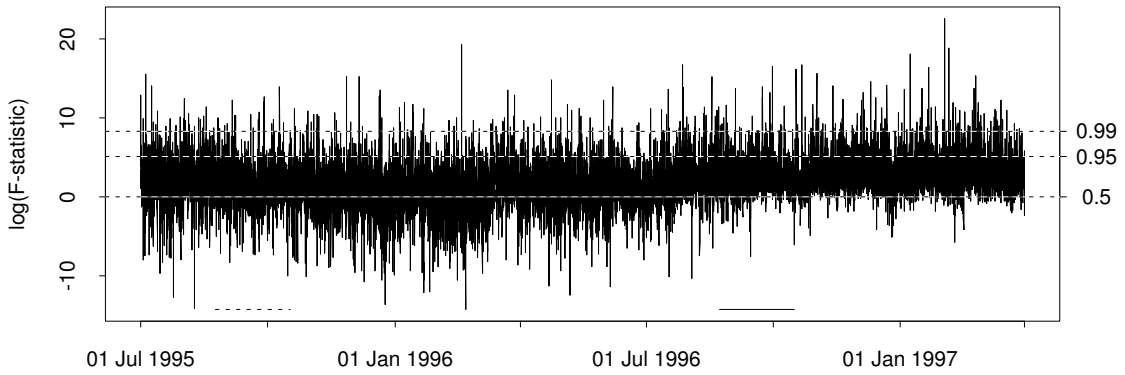

Figure 4.4. *F-statistic for the hypothesis that the campaign effect is nil, together with 50%, 95% and 99% quantiles of* $F(1,1)$.

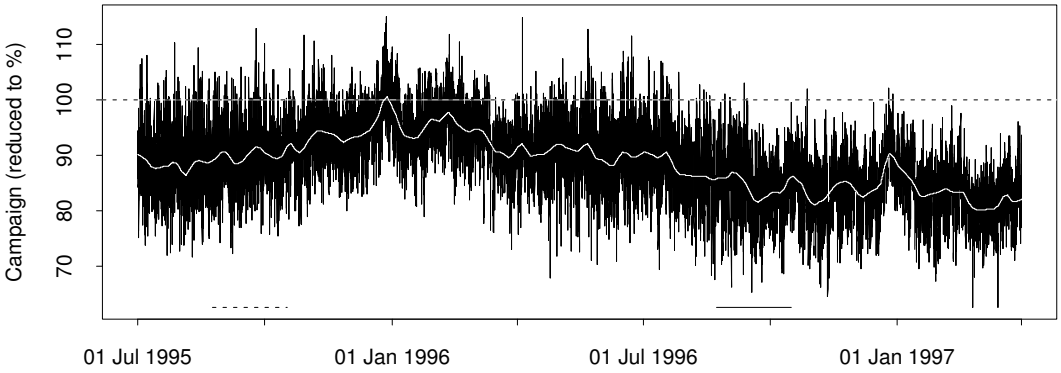

Figure 4.5. *Campaign component on the original scale.*

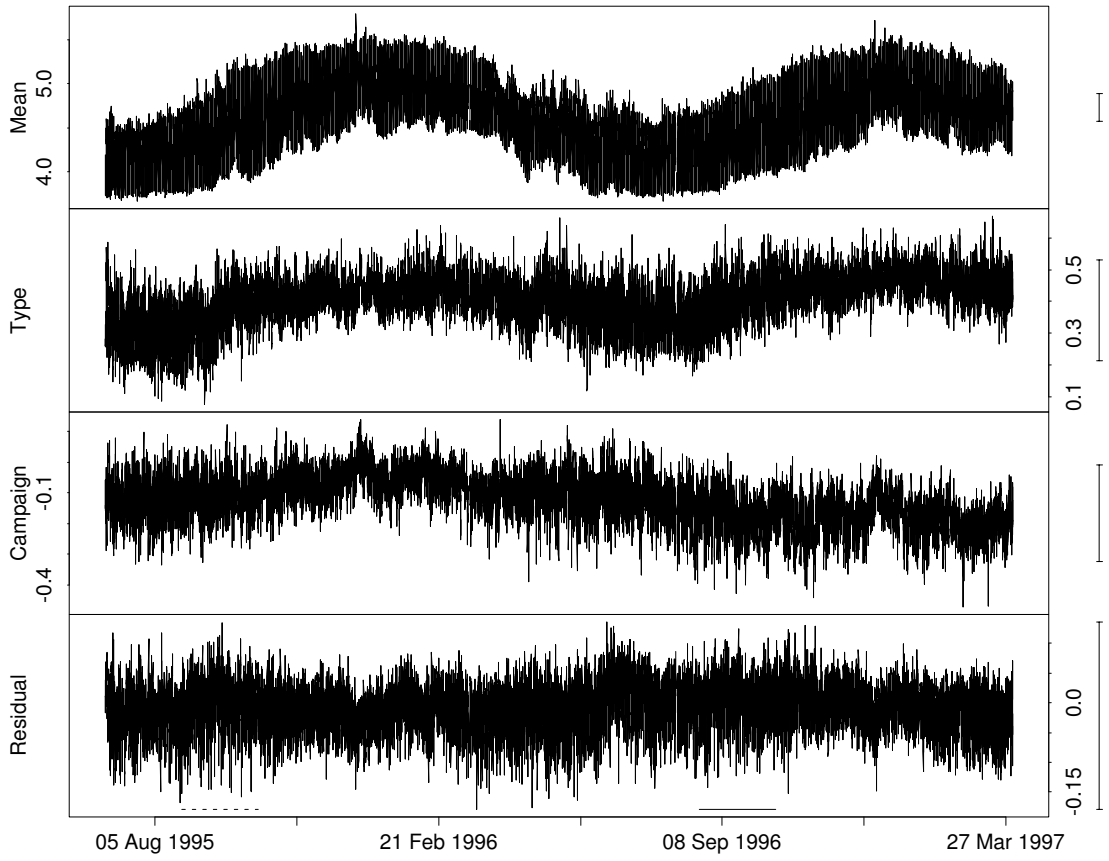

Figure 4.6. *Components of the additive split for logarithm transformed data, together with the residuals for substation 1325.*

4.4 Non-Additive Splitting

In Section 4.3 the type and campaign effects were estimated for each point in time, under the assumption that the effect of the campaign is the same for both types (0 and 1) of substations, i.e. it is assumed that the interaction between the type and campaign effects is negligible. In this section the interaction is estimated by using neighborhood data. This approach requires an assumption, namely that the effects and the mean change smoothly over time. The method is described in Section 3.3. A very small bandwidth, compared to the degree of the approximating polynomial, is used and consequently the estimates must be calculated exactly for every time point in the data set. As in Section 4.3 logarithm transformed data are used. At the end of the section the results are compared with results obtained by ANOVA decomposition, cf. Section 3.4.

Instead of using the consumption during one-hour periods and down weighting neighborhood data it may be argued that the original 15 minute data could be used and within the individual hours the data could be treated as four repetitions under the same experimental circumstances. However, since variations on a smaller time scale are now present this will not be true. For this reason the method briefly described above is applied.

Due to the diurnal variation in the data the curvature is an important property. For this reason second order polynomials are used. If the span of the tricube window is set to seven hours, then seven hours (28 observations) have non-zero weights, see Figure 4.7. Some observations have very little weight, if this is taken into account it is seen that the equivalent number of observations is $4\times(1+2\times(0.932+0.538+0.051)) \simeq 16.2$. Using second order approximations of the evolution of the mean and the three effects amounts to 12 local constants. Since the number of observations per point in time is four these numbers seem appropriate. Note that a two-sided window is used, another possibility is to use a left-sided window, but since this will introduce a phase shift we do not consider this further.

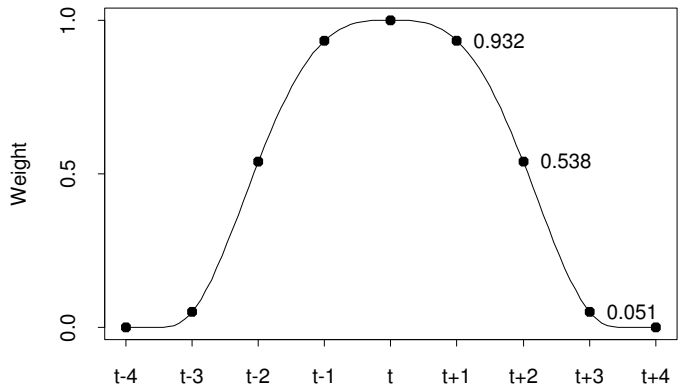

Figure 4.7. *Weights on observations when the tricube window spans seven hours.*

The bandwidth chosen corresponds to a model where the error has approximately

50% of the total degrees of freedom (number of observations). This is seen by calculating the degrees of freedom, using the smoother matrix as described in (Nielsen 1997), for a description of the concept of a smoother matrix see (Hastie & Tibshirani 1990). However, due to computational limitations, the degrees of freedom cannot be calculated for the solid data set. Instead the degrees of freedom are calculated using six consecutive days[1], yielding an error degrees of freedom of 281.4, with 576 observations this amounts to 49% of the total degrees of freedom. Varying the number of days used indicated that the degrees of freedom per observation are stable when observations from six days are used.

The result of the splitting of the four time series is displayed in Figure 4.8. With respect to the main effects (type and campaign) the result resembles that obtained for additive splitting, cf. Figure 4.3. This is no surprise since, except in case of missing values, the weighted least squares problems solved to obtain the results in Figure 4.8 all correspond to balanced designs, i.e. requiring the interaction to be zero will not influence the other estimates.

The residuals of the split are investigated. The standard deviation of the four residual series is in the range from 0.032 to 0.037. The correlation between series is in the range form 0.40 to 0.48 and the autocorrelation in lag 1 is approximately -0.4. Furthermore, a weak diurnal variation is present. Both the sample autocorrelation function, the inverse autocorrelation function, and the raw periodogram indicates that a (high frequency) MA(1) model will be able to explain the main part of the remaining variation of the individual series. To account for the cross correlation a vector MA(1) model may be applicable. However, we do not consider these aspects further.

In principle, in case of an significant interaction the campaign effect cannot be interpreted as a difference independent of type, and vice versa. However, if the interaction effect is significant the data differ with respect to both type and campaign and we may infer that there is some evidence of an campaign effect, but it may have opposite sign for the two types of substations, or it may be nil for one of them. Although, a formal test for the interaction effect has not been developed there is some evidence that this effect is in fact significant. The sum of the squared differences between fitted values, when including the interaction effect and when excluding this is 87.7 (log-scale) and the sum of the squared residuals when including the interaction effect is 70.8. As mentioned above the small bandwidth splitting used in this section has degrees of freedom approximately equal to half the number of observations ($\approx 4 \times 15000$). Furthermore it seems reasonable to assume that the four components corresponds to the same number of degrees of freedom. The usual ANOVA F-statistic is then

$$F = \frac{87.7/(60000/2 - 3/4 \times 60000/2)}{70.8/(60000 - 60000/2)} \simeq 5.0,$$

on $(7500, 30000)$ degrees of freedom. The 95% quantile of the corresponding F-

[1]The results are independent of which six days are used, provided that no values are missing.

distribution equals approximately 1.0 (which is true for all F-distributions with a large number of degrees of freedom). Consequently, using the normal F-test suggests a strongly significant interaction effect.

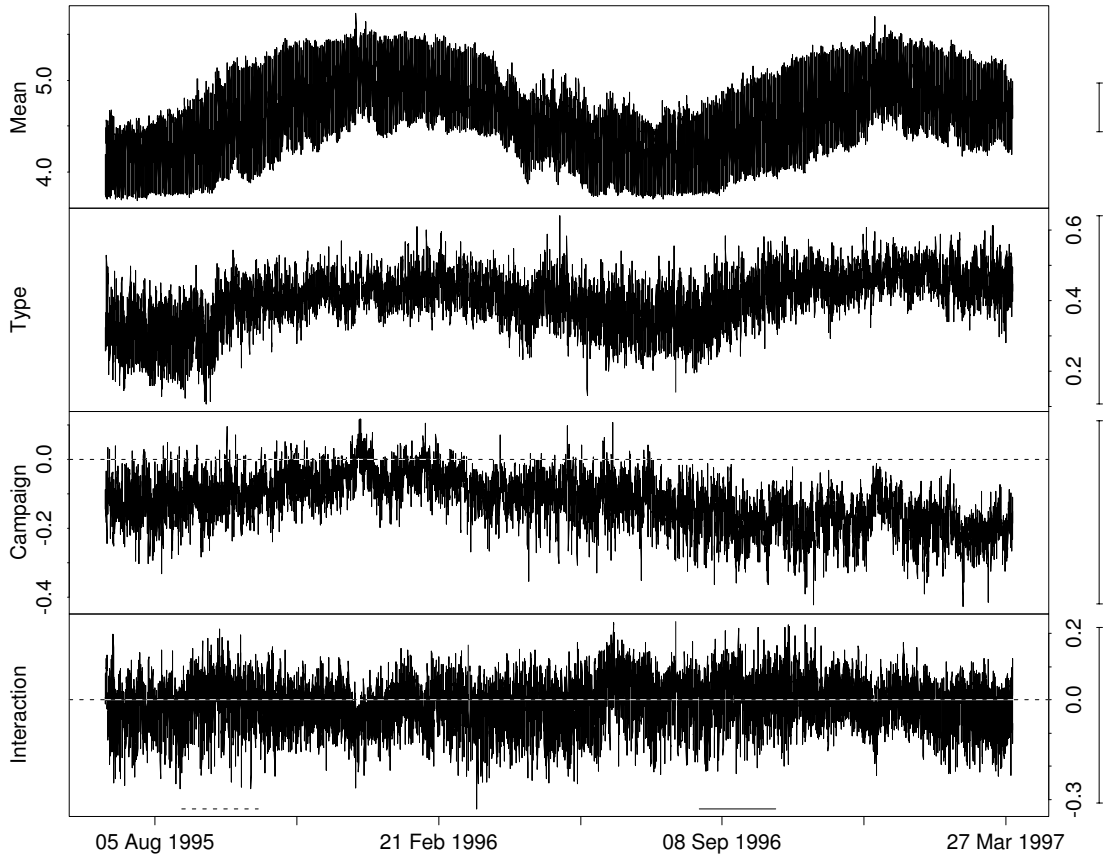

Figure 4.8. *Result of small bandwidth splitting.*

Following the procedure of small bandwidth splitting described above the fitted values on the original scale may be expressed as

$$\widehat{P}_{ijt} = e^{\hat{\mu}_t} e^{(-1)^{i+1}\hat{\alpha}_t/2} e^{(-1)^{j+1}\hat{\beta}_t/2} e^{(-1)^{i}(-1)^{j}\hat{\gamma}_t/2}; \quad i,j = 0,1, \tag{4.5}$$

where μ_t is the mean, α_t is the type (i), β_t is the campaign (j), and γ_t is the interaction effect ($j = 0$ corresponds to "control" and $j = 1$ corresponds to "active").

For type 0 (substations 1325 (control) and 4284 (active)) the ratio between the fitted values for active and control is

$$\frac{\widehat{P}_{01t}}{\widehat{P}_{00t}} = e^{\hat{\beta}_t} e^{-\hat{\gamma}_t}, \tag{4.6}$$

and for type 1 (substations 2588 (control) and 1667 (active)) the ratio between the fitted values for active and control is

$$\frac{\widehat{P}_{11t}}{\widehat{P}_{10t}} = e^{\hat{\beta}_t} e^{\hat{\gamma}_t} \tag{4.7}$$

From (4.6) and (4.7) it is seen that if the interaction effect is small compared to the campaign effect, this can be interpreted without considering the type. If the interaction (γ) is positive on the log-scale then the ratio for type 0 is smaller than for type 1, indicating that the campaign has a larger effect on type 0 than on type 1. This is illustrated by the following example.

Example 4.1 *Suppose* $\hat{\beta}_t = -0.1$ *and* $\hat{\gamma}_t = 0.05$. *From (4.6) and (4.7) the ratios between active and control is calculated for each type.*

$$\frac{\widehat{P}_{01t}}{\widehat{P}_{00t}} = 0.905 \times 0.951 = 0.86, \qquad \frac{\widehat{P}_{11t}}{\widehat{P}_{10t}} = 0.905 \times 1.051 = 0.95$$

□

If the mean over t of the interaction effect is close to zero or if it exhibits a diurnal variation but averages to a value close to zero over a period of some days a trend component of the campaign effect can be interpreted without considering the interaction. For this reason the estimates of the interaction effect are analyzed in the following.

Figure 4.9 shows box plots (Statistical Sciences 1993) of the diurnal variation of the interaction effect for the three types of days. It is seen that half-holy and holy days are quite similar, for which reason these groups will be pooled in the following. Furthermore, the box plot corresponding to working days reveals that part of the interaction originates from a phase shift of the four series.

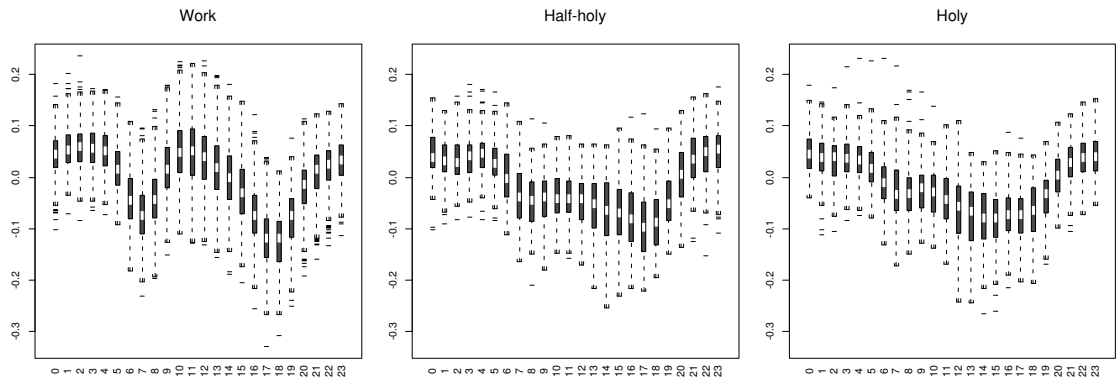

Figure 4.9. *Box plots of the interaction effect from Figure 4.8 versus the time of day for working days (left), half-holy and Saturdays (middle), and holy and Sundays (right).*

Based on the observations regarding the diurnal variation all models considered for the interaction effect contain a variation in the (marginal) mean. The variation is

modeled with a free parameter for each hour of the day. In the following d_t will denote this diurnal profile, which equals

$$d_t = I_w(t) \sum_{i=0}^{23} I_i(t)\mu_{1i} + (1 - I_w(t)) \sum_{i=0}^{23} I_i(t)\mu_{2i}, \tag{4.8}$$

where $I_w(t)$ equals 1 if t corresponds to a working day and 0 otherwise, $I_i(t)$ equals 1 if t corresponds to the time of day denoted by i, and μ_{1i} and μ_{2i} are the parameters.

A model consisting of a diurnal profile only is fitted to the data. The estimates of the autocorrelation function and the inverse autocorrelation function are displayed in Figure 4.10. The estimates indicate that a autoregressive model with a seasonal component at 24 hours is needed to account for the correlation. Many different models are fit to the series and the estimates of the autocorrelation function compared. During the process it is revealed that a seasonal component at lag 168 is also needed. For computational reasons the models are not fit directly as seasonal models, instead the corresponding model being linear in the parameters is used. Finally, an AR(48) model with a seasonal component at lag 168 is selected and the following model fit to the series by least squares

$$y_t = \sum_{i=1}^{48} a_i y_{t-i} + \sum_{i=168}^{216} a_i y_{t-i} + d_t + e_t, \tag{4.9}$$

where y_t is the interaction at time t, d_t is the diurnal profile (4.8), and $\{e_t\}$ is a zero mean white noise process with constant variance. The estimate of the autocorrelation function of the residuals from this model is displayed in Figure 4.11. It is seen that the model fits the data well. From this it is concluded that the interaction can be modeled by a stationary stochastic process, with the exception of a varying but constrained marginal mean. The pole with the largest modulus is real and lie at 0.996, this is consistent with the long memory indicated by the plot of the series (Figure 4.8, bottom) and the estimate of the autocorrelation function in Figure 4.10.

Under model (4.9) the marginal mean m_t must obey the recursive equation

$$m_t = \sum_{i=1}^{48} a_i m_{t-i} + \sum_{i=0}^{168} a_i m_{t-i} + d_t, \tag{4.10}$$

but since d_t is not strictly periodic (e.g. around Christmas) m_t will also not be periodic, not even on a weekly basis. The solution of the recursions, when fixing the initial values needed at zero, is displayed in Figure 4.12. The figure also shows a locally weighted mean using a tricube window spanning one week. When comparing with Figure 4.8, it is seen that with respect to the trend the deterministic part of the interaction is very small. This ensures that the decreasing trend in the campaign effect observed for the additive splitting (Figure 4.5) is approximately independent of the type. Comparing the period October – March 1996/97 with the same period one year earlier a reduction of approximately 10% from 95% to 85% is observed.

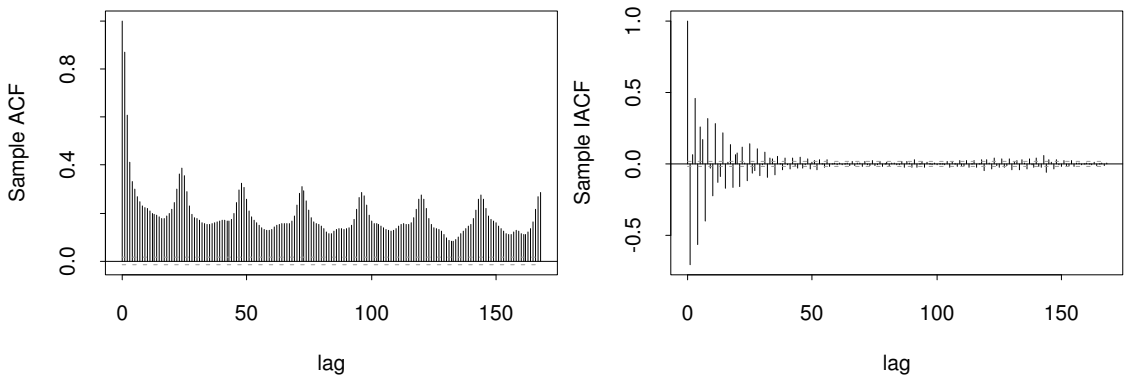

Figure 4.10. *Sample autocorrelation function (left) and inverse autocorrelation function (right) of the residuals of a model consisting of a diurnal profile only. The maximum lag is 168 hours. An approximate 95% confidence interval for white noise is indicated by dotted lines.*

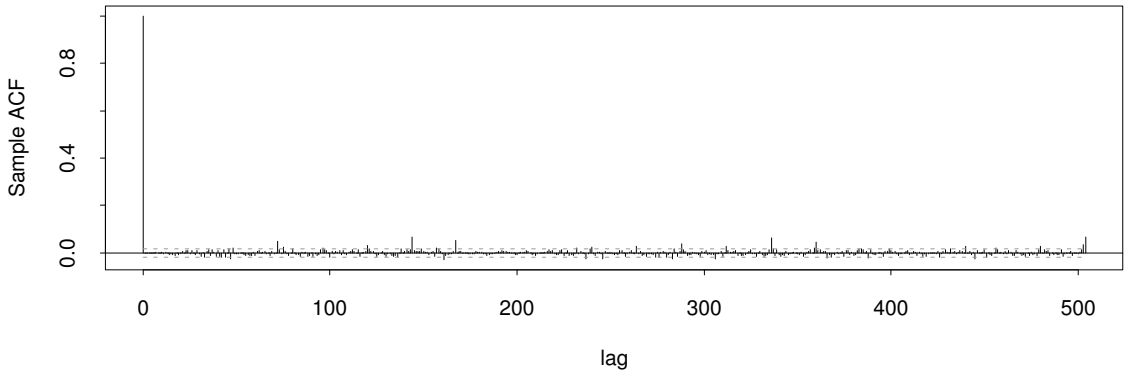

Figure 4.11. *Sample autocorrelation function of the residuals from (4.9). The maximum lag is 504 hours (three weeks). An approximate 95% confidence interval for white noise is indicated by dotted lines.*

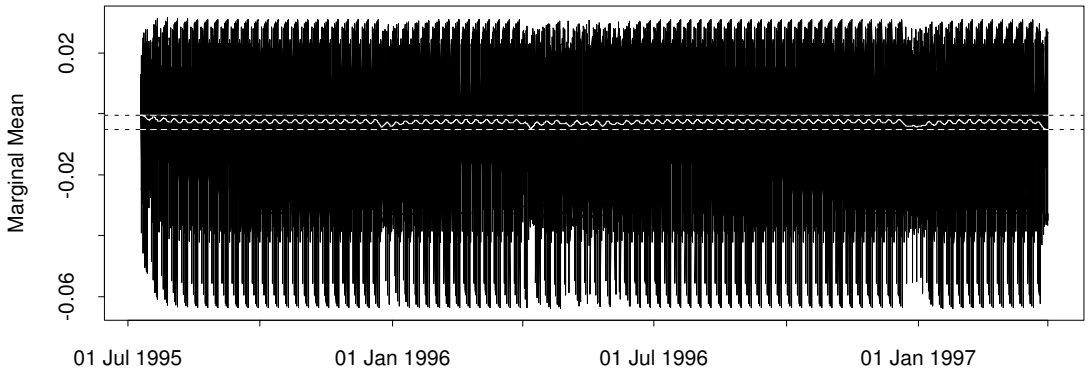

Figure 4.12. *Solution of the recursions (4.10) when fixing the initial values needed at zero, together with a locally weighted mean, and the range of this (-0.0051 to -0.0005).*

In Figure 4.13 the campaign and interaction effects from Figure 4.8 are displayed on the same plot. It is seen that for the first part of the period the interaction almost hides the campaign effect, whereas the effects seem to diverge starting in the late summer of 1996. The smoothed series overlayed on the plot clearly show that the interaction moves around zero, while the annual variation of the campaign effect observed for the period 1 July 1995 until 1 July 1996 is not repeated in the following period. The smoothed values of the campaign and interaction effects from the figure just considered are used to estimate the ratio between the power consumption of active and control substations for type 0 and 1 as described in (4.6) and (4.7). The results are shown in Figure 4.14. Also the pooled estimate, $e^{\hat{\beta}_t}$ in the equations just referred, is shown in the figure when using a smoother with a window span of 64 days. The ratios for the individual types do not differ markedly over time, and overall the pooled estimate seems appropriate.

Very similar results are obtained using ANOVA decomposition, cf. Section 3.4, of the logarithmic transformed series. Two decompositions are performed. For both series the seasonal smoothing is performed separately for working days and non-working days, using local lines and weights based on a tricube window spanning 270 days. The trend smoothers both use local second order polynomials; but one uses a tricube window spanning 50 days and the other uses a window spanning 270 days. In both cases a convergence criterion (3.1), corresponding to a maximum relative change of any of the components, of 0.001 is used and no robustness iterations are performed. The settings are chosen as described since they resemble the ones used in Section 4.1, in which seasonal diagnostic plots are used to access the appropriateness of the settings. The results obtained for the trend component are used in Figure 4.15 to obtain estimates of the ratios between active and control for the individual types. Also the pooled estimate based on the trend smoother with a window span of 270 days is shown on the plot. Compared with Figure 4.14, the results are very similar.

Plots of the remainder of the ANOVA decompositions exhibit some small scale irregular variation. This could be modelled adding another trend component varying on a smaller scale. Although this seems to be an obvious way of identifying at which time scale the results can be interpreted without considering the type, the concept is not considered further.

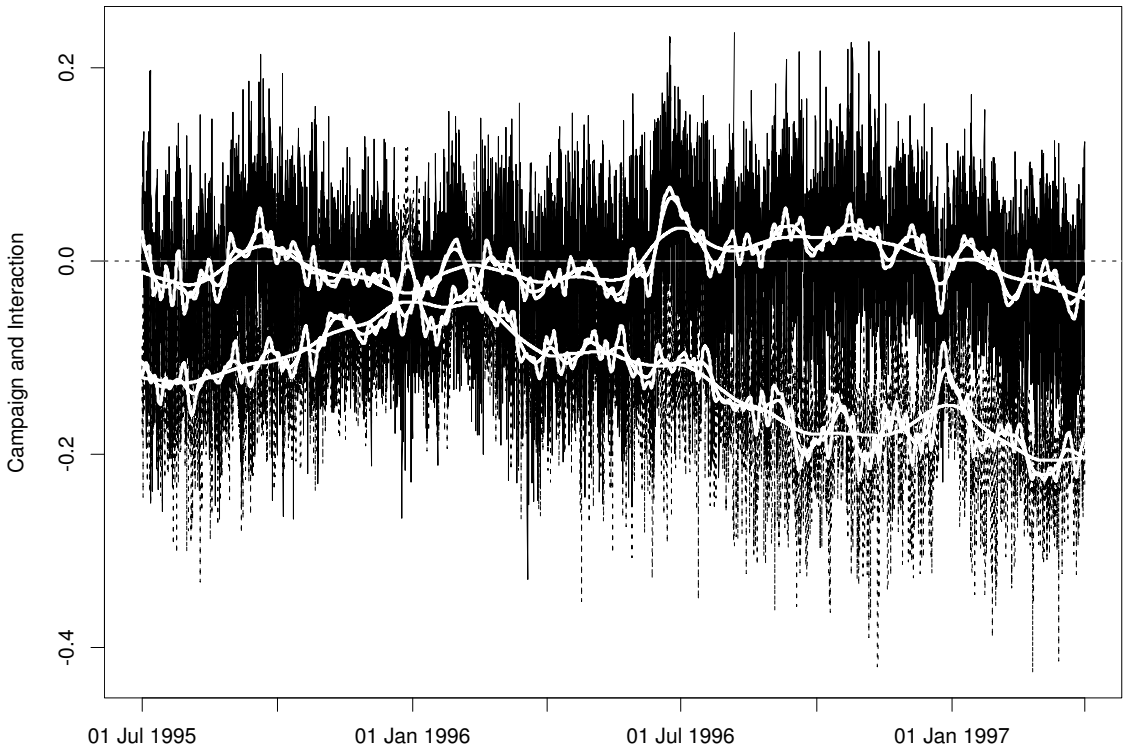

Figure 4.13. *Campaign (dotted) and interaction (solid) effects from Figure 4.8, together with robust locally weighted linear smooths for (tricube) window spans of 6.4, 19.2, and 64 days.*

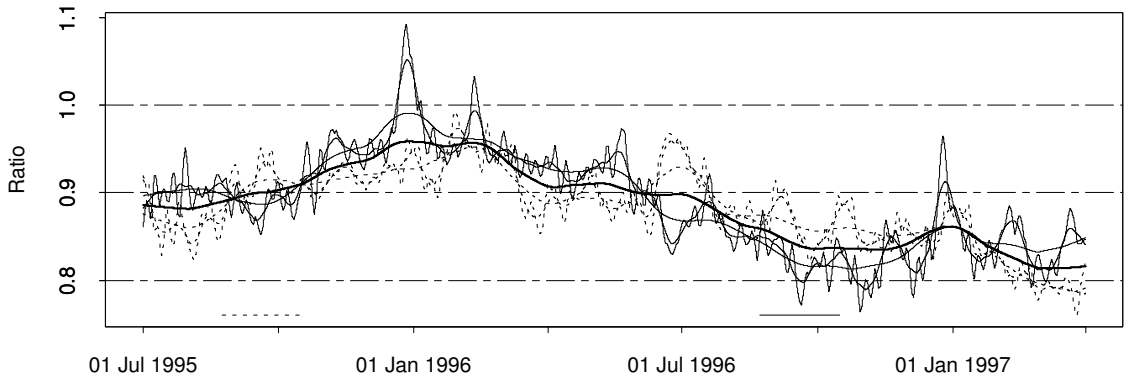

Figure 4.14. *Ratio of 4284 to 1325 (solid) and 1667 to 2588 (dotted) based on the smooths from Figure 4.13. The bold line shows the pooled estimate for the smoother with the widest window (64 days).*

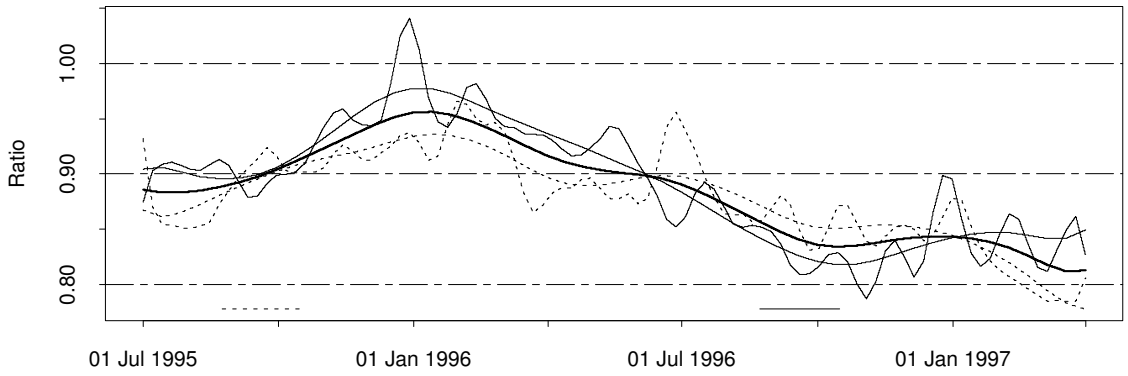

Figure 4.15. *Ratio of 4284 to 1325 (solid), 1667 to 2588 (dotted), and a pooled estimate (bold) based on ANOVA decompositions where only the span of the trend smoother window is varied (50 and 270 days).*

4.5 Influence of Climate

4.5.1 Comparing Trend Components and Climate

In the previous sections it is shown that over time the development of the power consumption indicate an effect of the campaign. However, an annual variation is present and consequently some part of the variation may be explained by the climate. For this reason the results are compared with the monthly degree days as displayed in Figure 2.2. However, these are first used to generate hourly values using linear interpolation followed by scaling to obtain values for which the sums for the individual months equal the original data.

Figure 4.16 shows the trend components from the decomposition of the individual series, shown in Figure 4.1, plotted against the degree days. It is seen that for time points later than 1 Oct. 1996 the trend component as a function of the degree days has a smaller slope for the active (left) than for the control (right).

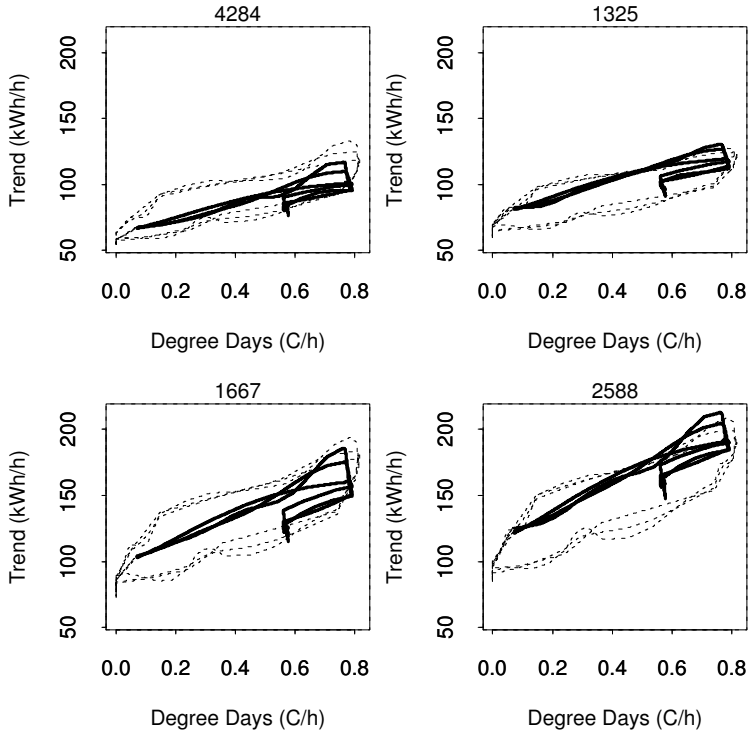

Figure 4.16. *Estimates of the trend component of the individual time series for three bandwidths (also shown in Figure 4.1) versus the degree days. Time points later than 1 Oct. 1996 are indicated by bold lines.*

Similar plots for the mean, type, campaign, and interaction effect of the trend component from the ANOVA decompositions considered in the last part of Section 4.4

are displayed in Figure 4.17. From these plots it is seen that the campaign effect is the one which most clearly separates the time points before and after 1 Oct. 1996. Furthermore, the annual variation of the mean and type effects is to a large extend explainable by the degree days. The interaction effect seems to differ slightly for the first part of the period after 1 Oct. 1996, but ends up in the same region as the values during the winter 1995/96.

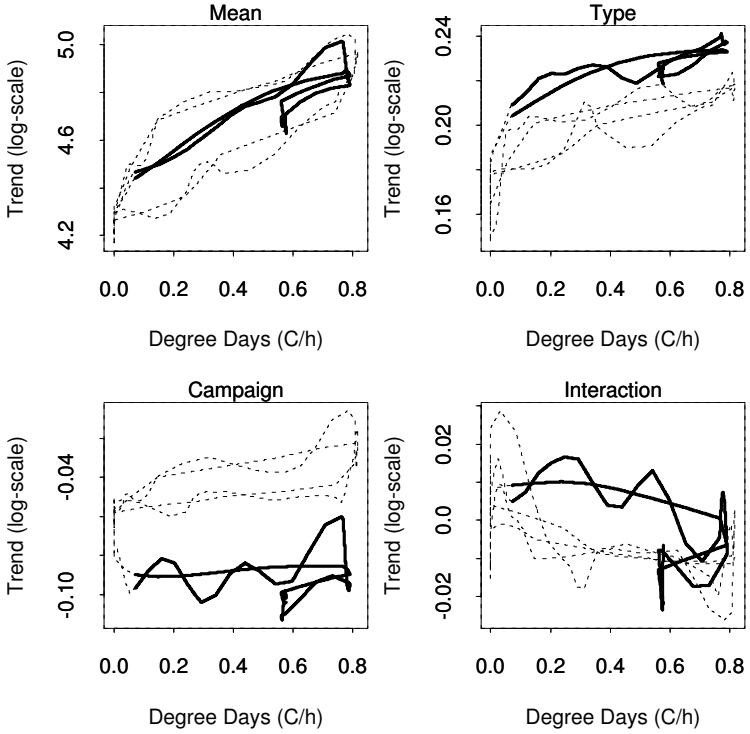

Figure 4.17. *ANOVA decomposition: Mean, type, campaign, and interaction of the trend component for logarithmic transformed data, using a window span of 50 and 270 days versus the degree days. Time points later 1 Oct. 1996 are indicated by bold lines.*

4.5.2 Modelling the Dependence on Climate

In this section the dependence on the degree days will be modelled. It is clear that besides the dependence on the degree days the data contain a diurnal variation, and from Figure 2.3 it is clear that the amplitude of the diurnal variation varies over time. Plots of the logarithmic transformed data indicate that the diurnal variation on this scale has a constant amplitude. In the following it will be assumed that this is true, but the appropriateness of the assumption will be addressed by analyzing how well the model(s) fit the data.

Figure 4.18 shows the logarithmic transformed data versus the degree days for the

four substations. Apart from observations for which the degree days are below approximately $0.1°C/h$ the relationship seems to well described by a strait line. Observations with low values of degree days are only present in the data set before the campaign was initiated. These observations will be accounted for by estimating a trend component (see below).

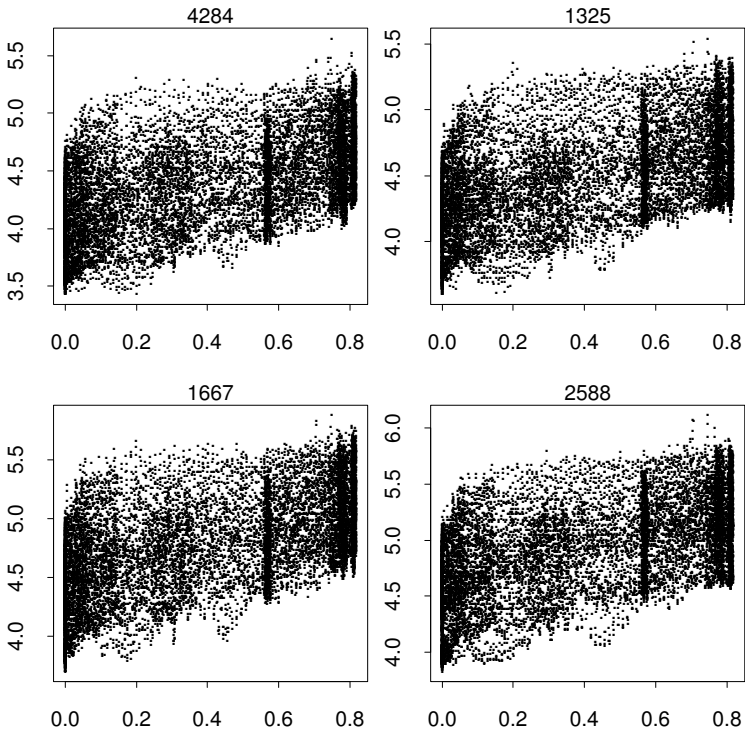

Figure 4.18. *Logarithmic transformed power consumption (kWh/h) versus the degree days ($°C/h$) for the four series.*

Since almost all visits are performed before 1 Oct. 1996, cf. Figure 2.1, the following model is fit to the individual series.

$$\log(P_t) = T_t + a + bG_t + d_t + I_{1Oct96}(t)\left[\Delta a + \Delta bG_t + \Delta d_t\right] + e_t, \tag{4.11}$$

where T_t is the trend component at time t; G_t is the degree days at time t; $I_{1Oct96}(t)$ equals -1 before and 1 after 1 Oct. 1996; a, b, Δa, and Δb are constants; d_t and Δd_t are diurnal profiles similar to (4.8); i.e.

$$d_t = I_w(t)\left[\mu + \sum_{i=0}^{23} I_i(t)\mu_{1i}\right] + (1 - I_w(t))\left[-\mu + \sum_{i=0}^{23} I_i(t)\mu_{2i}\right], \tag{4.12}$$

where $\mu_{..}$ denotes the parameters, and $\sum_{i=0}^{23}\mu_{1i} = \sum_{i=0}^{23}\mu_{2i} = 0$, i.e. for working days the mean is μ and for non-working days the mean is $-\mu$. Internally in the software used for estimation a Helmet parametrization is used (Chambers & Hastie 1991), (3.3) is an example of a coding corresponding to a Helmet parametrization. A

necessary condition for the uniqueness of the estimates is that T_t sums to zero over the period considered.

The trend component T_t of (4.11) is modelled by a smoother, and hence no specific parametric form is assumed. The smoother used is based on locally weighted polynomial regression, using a tricube window and a nearest neighbour bandwidth. In order to be able to appropriately model peaks second order polynomials are used locally. Model (4.11) is fitted by back-fitting (Hastie & Tibshirani 1990), altering between smoothing and linear regression, and initiating the procedure by fitting the linear part of the model and disregarding the trend component. In order to obtain an estimate of the trend component that sums to zero over the period considered, the smooth is adjusted by subtracting the overall mean of the smooth. The iterative procedure is stopped when both the fitted values of the linear part and the trend component do not change more than 0.1%, relative to the total range of the fitted values, from one iteration to the next.

For some bandwidths the estimation in (4.11) does not converge after 50 iterations, and in general the convergence is slow. Investigations show that this is due to the fact that for some bandwidths (windows spanning 25% of the data on the time axis) the rate of change in the estimate of the trend component is of the same order as the rate of change in the degree days. Furthermore, starting approximately at 1 June 1996 the curves are in phase. Hastie & Tibshirani (1990) call these type of problems for concurvity. Based on these observations it is concluded that Δa and Δb in (4.11) cannot be estimated unambiguously. As a consequence the reduced model

$$\log(P_t) = T_t + a + bG_t + d_t + I_{1Oct96}(t)\Delta d_t + e_t, \tag{4.13}$$

will be used. Doing this, as much as possible of the long term variation in P_t will be accounted for by a global linear model in G_t, whereas the remaining is accounted for by the trend-smoother. The linear part of the model, i.e. all except T_t, contains 96 parameters, most of which are describing the diurnal profiles.

Bandwidth Selection

Model (4.13) is fitted for nearest neighbour bandwidths of 1, 3, 5, 10, 15, 20, ..., 60, and 65%. The sample autocorrelation function (SACF) and sample inverse autocorrelation function (SIACF) of the residuals for the bandwidths 1% and 3%, when using observations from substation ID 4284 is displayed in Figure 4.19. For a bandwidth of 1% it is seen that the SIACF has a rather strange appearance, and it is concluded that bandwidths lower than 3% are inappropriate. For bandwidths higher than 3% the SIACF is nearly identical to the SIACF in the bottom row of Figure 4.19. This indicates that an AR(1) model with a seasonal component at lag 24 will be able to account for most of the temporal correlation in the residuals.

To assist the selection of bandwidths different information criteria are considered. In order to do that the equivalent number of parameters of (4.13) must be known, cf.

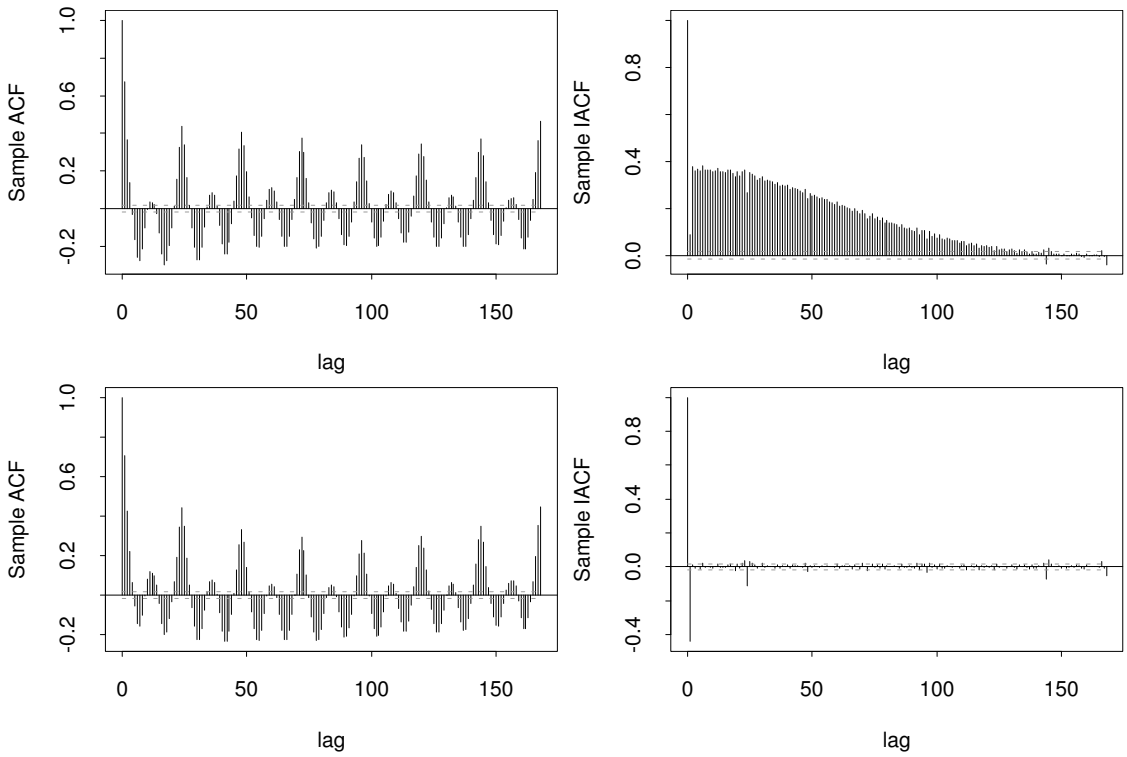

Figure 4.19. *Sample autocorrelation function (left) and inverse autocorrelation function (right) of the residuals from model (4.13), using a nearest neighbour bandwidth of 1% (top) and 3% (bottom). The maximum lag is 168. An approximate 95% confidence interval for white noise is indicated by dotted lines.*

Section 3.5. For the implementation of the backfitting algorithm the `loess` function of S-PLUS (Statistical Sciences 1995) is used for smoothing, this function provides an approximation to the equivalent number of parameters, see also (Cleveland, Devlin & Grosse 1988). In the following the equivalent number of parameters of (4.13) will be taken to be the sum of the number of parameters in the linear part of the model and the equivalent number of parameters of the trend component estimator. Based on this, together with the residuals of the fit, the information criteria Generalized Cross Validation (GCV), Akaike's Information Criterion (AIC), and Bayesian Information Criterion (BIC) are calculated for each of the bandwidths considered above, cf. Section 3.5.

Since AIC and BIC are likelihood inspired, at least one of the models considered should yield residuals resembling white noise. According to Figure 4.19 (and similar plots for other bandwidths) this is clearly not the case for model (4.13). However, the residuals can be appropriately modelled by a zero mean autoregressive model with non-zero parameters at lags 1, 24, and 25. Figure 4.20 shows the SACF after fitting the above mentioned autoregressive model to the residuals from (4.13), when using a bandwidth of 3% and data from substation 4284. Compared with Figure 4.19

it is seen that most of the temporal correlation is accounted for. The information criteria are therefore also calculated after fitting the autoregressive model to the residual series. In this case the number of observations N is decreased due to extra missing values being generated by the autoregression.

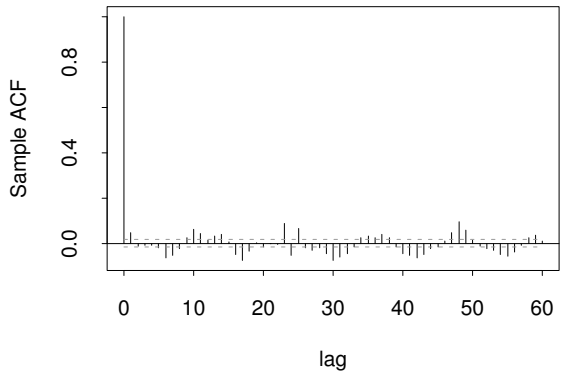

Figure 4.20. *Sample autocorrelation function after fitting an autoregressive model to the residuals from (4.13), when using a bandwidth of 3%. An approximate 95% confidence interval for white noise is indicated by dotted lines.*

Figures 4.21 and 4.22 displays the values of GCV, AIC, and BIC for different bandwidths. In Figure 4.21 the values are calculated based on the residuals after fitting (4.13), whereas in Figure 4.22 the values are calculated after fitting the autoregressive model. In this case the three parameters of the autoregressive model are added to the total equivalent number of parameters.

When not adjusting for the temporal correlation of the residual series it is seen that GCV and AIC point towards a bandwidth of 1%, whereas BIC leads to 3%, but above it is argued that 1% is inappropriate for other reasons. If the adjustment is performed as described, although inappropriate for the low bandwidth of 1%, GCV and AIC points towards a bandwidth of 3–10%, whereas BIC leads to an very high bandwidth (the linear part of the model contain 96 parameters). Overall 5% seems to be an appropriate choice, or perhaps 10% if a more smooth estimate of the trend component is desired. For these two bandwidths the estimates of the trend component are displayed in Figure 4.23. Based on this figure a bandwidth of 10% is selected. Residual diagnostic plots are included in Appendix C, and these are all fairly well behaved, although the normal quantile-quantile plots indicate that the distribution of the residuals has longer tails than the normal distribution.

In the bandwidth selection the BIC criterion is neclected since we believe that it is inappropiate for the kind of models considered, see the discussion in Chapter 6. For completeness the values of BIC are, however, displayed.

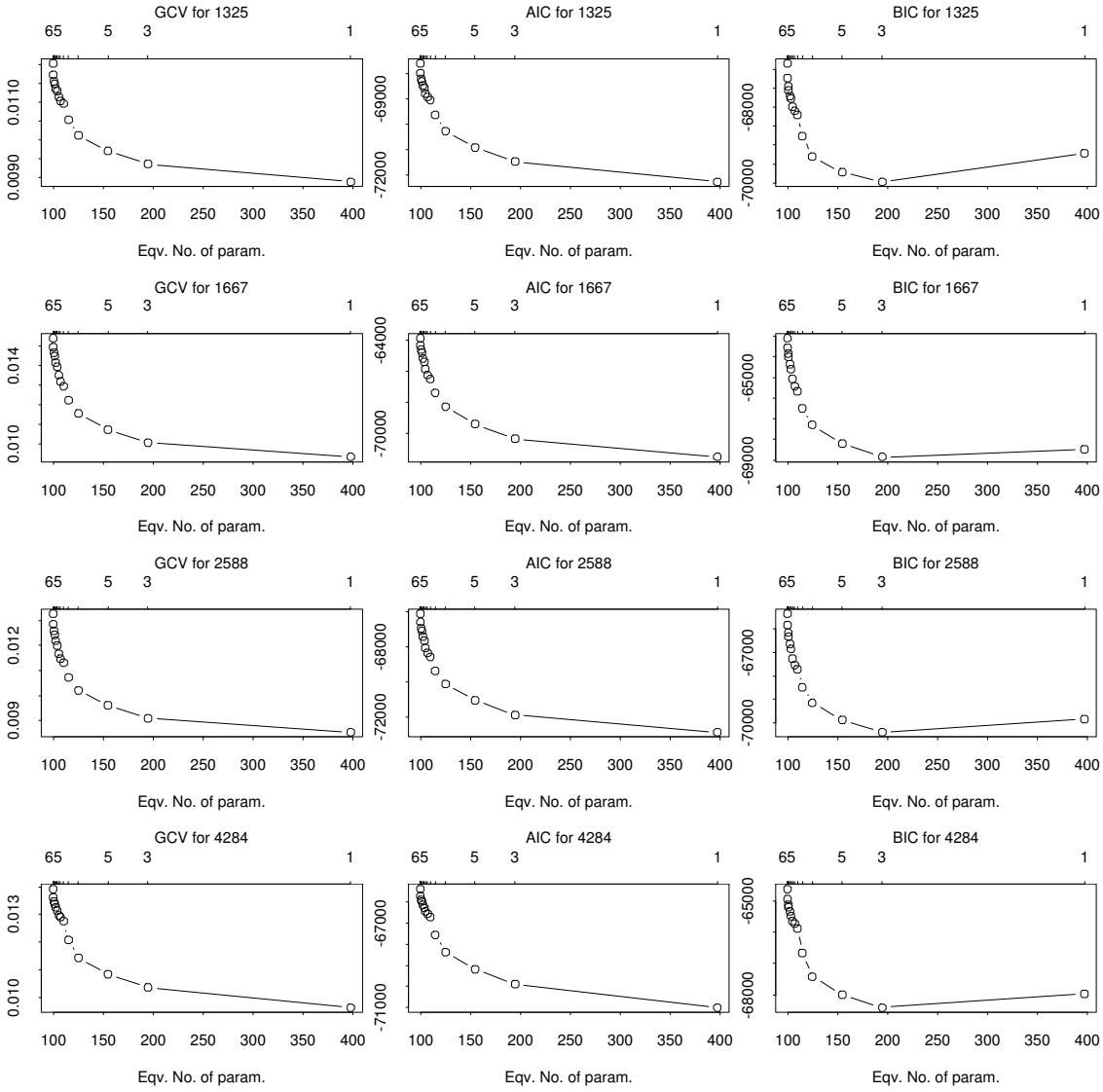

Figure 4.21. *Bandwidth selection criteria GCV, AIC, and BIC when different bandwidths (shown in percent on top of each plot) are used for the local quadratic trend smoother in (4.13).*

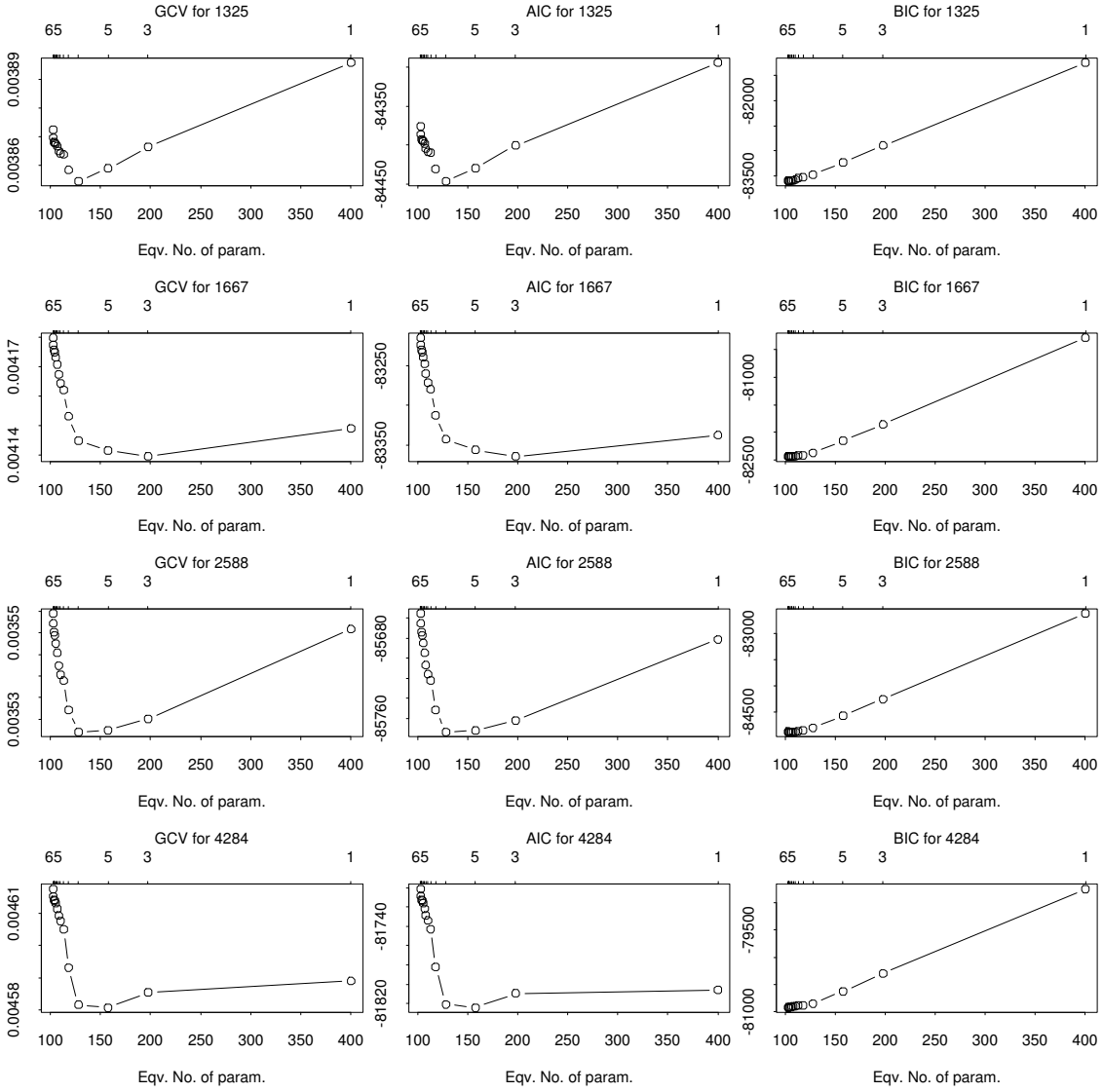

Figure 4.22. *Bandwidth selection criteria GCV, AIC, and BIC as in Figure 4.21, but calculated after modelling the temporal correlation of the residual series.*

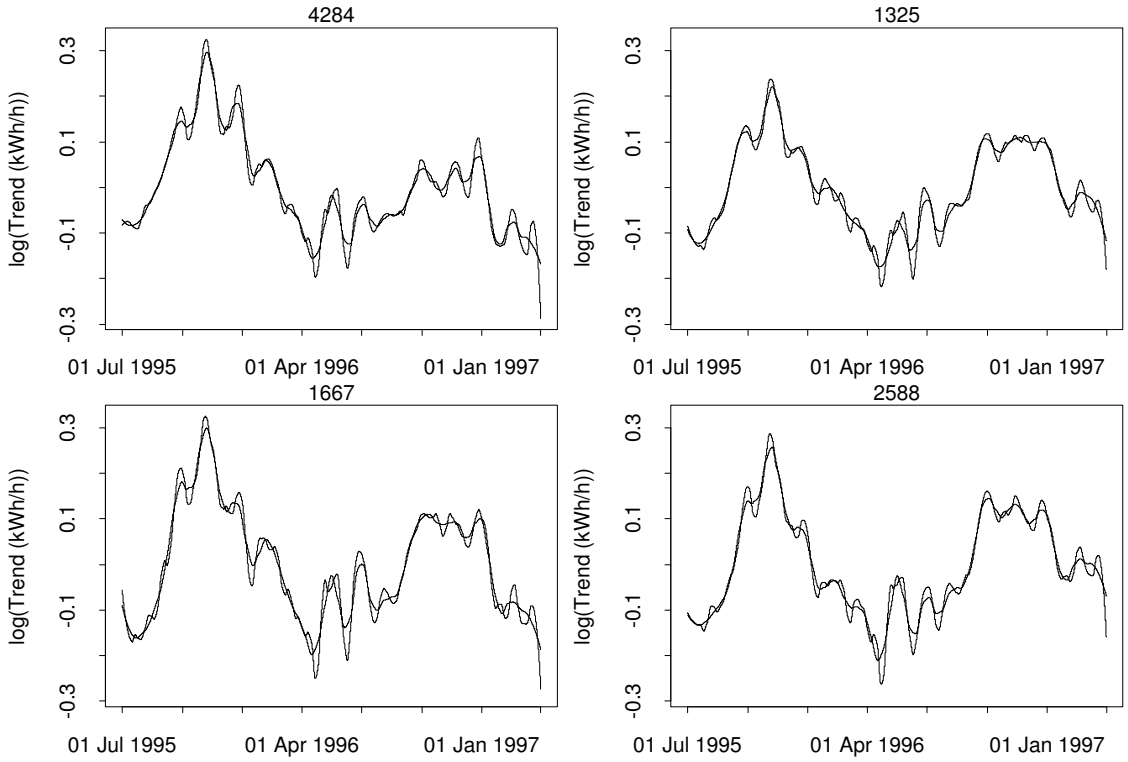

Figure 4.23. *Estimates of the trend component using bandwidths 5 and 10% for all four substation.*

Final Estimates

In the remaining part of this section the results using (4.13) together with a 10% nearest neighbour smoother, as described above, will be presented. First it is noted that the fitted values under (4.13) may be written as

$$\widehat{P}_t = e^{\hat{a}} e^{\hat{b}G_t} e^{\widehat{T}_t} e^{\hat{d}_t + I_{1Oct96}(t)\widehat{\Delta d}_t}. \tag{4.14}$$

Here $e^{\hat{a}}$ is the overall level of power consumption in kWh/h, $e^{\hat{b}G_t}$ is the factor by which the overall level is changed due to the degree days G_t, $e^{\widehat{T}_t}$ is the factor by which the overall level is changed on the long run due to other factors than G_t, and $e^{\hat{d}_t + I_{1Oct96}(t)\widehat{\Delta d}_t}$ is the factor by which the daily level $e^{\hat{a}} e^{\hat{b}G_t} e^{\widehat{T}_t}$ is changed due to the time of day and week.

The estimated level and dependence on degree days, expressed as $e^{\hat{a}}$ and $\hat{b}$, are displayed in Table 4.1. A similarity is seen within types of substations, i.e. the groups 4284/1325 and 1667/2588. For $\hat{b} = 0.85\ h/^{\circ}C$ the value of $\hat{b}G_t$ will range from 0 to 0.7, i.e. the range is very similar in size to the range of the trend component on the logarithmic scale, cf. Figure 4.23.

Figure 4.24 displays the percent-wise increase in the long run power consumption

Substation	$e^{\hat{a}}$ (kWh/h)	$\hat{b}$ $(h/°C)$
4284	59	0.76
1325	67	0.70
1667	84	0.85
2588	96	0.85

Table 4.1. *Estimated level ($e^{\hat{a}}$) and dependence on degree days ($\hat{b}$) for all four substations.*

due to other factors than the degree day values, i.e. $(e^{\widehat{T}_t} - 1) \times 100\%$, for each pair of active and control, cf. Table 2.1. On the figure the period in which the households were visited and the corresponding period one year before are indicated by horizontal line segments. It is seen for both pairs that the control starts below the active and crosses over just before the initiation of the trial. The effect is more pronounced for 4284/1325 than for 1667/2588. To partly eliminate differences before the initiation of the trial the average difference between active and control up to 1 Aug. 1996 is added to the control. The adjustment displayed is performed on the log-scale. However, performing the adjustment on the original scale gives virtually the same results. Using this adjustment it is clearly seen that the effect is more pronounced for one pair than for the other. The reason for the crossing of the active and control curves before the initiation of the trial could be (i) that the households are contacted by mail and phone before the actual visits, and/or (ii) that the actual shift is quite pronounced, which, due to the construction of the smoother, will result in slowly changing estimates starting before the actual shift. If (ii) is the case it should be detectable by visual inspection of the residuals, however, this does not seem to be the case.

The estimates of the trend component (on the log-scale) are split in mean, type, campaign, and interaction effects. The split is performed as described in Section 3.2, but observing that in this case twice the residual corresponding to substation 1325 equals the interaction as described in Section 4.4 (model (4.5)). Figure 4.25 shows the four components from which the original estimates of the trend components can be perfectly restored. The interaction effect is the smallest of all. Figure 4.26 shows the ratio between active and control, together with the overall ratio, cf. (4.6) and (4.7) in Section 4.4. Although, there seems to be some difference between types (see also Figure 4.24) the overall estimate is not totally misleading, in that the same overall behaviour is present for both types. The overall estimate, when visually smoothed, indicates a 10-12% difference between 1 Jan 1996 and 1 Jan 1997. This is also true when comparing the period just after the visits to the households, with the same period one year before. Note that this reduction is essentially the same as found without considering the climate, cf. Section 4.4 and Figure 4.5.

In Appendix C plots of the estimated diurnal profiles are displayed. In Figure C.1 and C.2 the results are compared for active and control substations, both before and

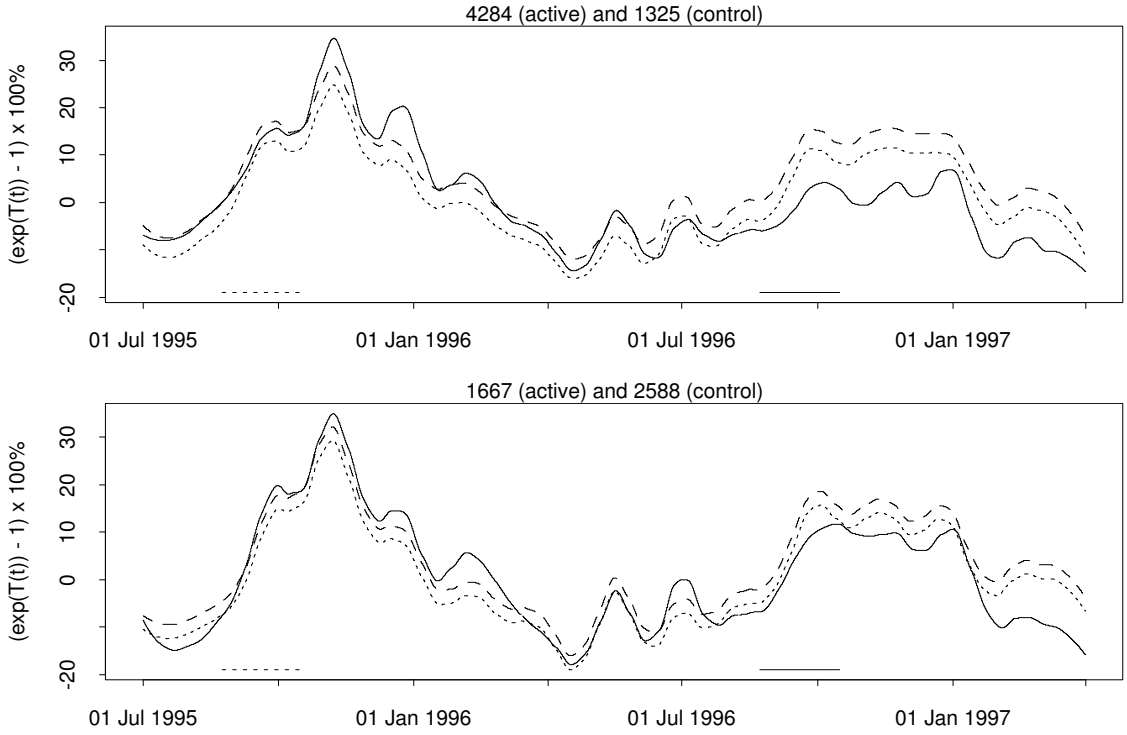

Figure 4.24. *Estimates of the trend component of (4.13) using a bandwidth of 10% for each pair of active (——) and control (- - -). An adjusted estimate of the control (– – –) is also included.*

after 1 Oct. 1996. In Figure C.3 and C.4 the individual substations are addressed, the results obtained for the first period (before 1 Oct. 1996) are compared with the results obtained for the last period (after 1 Oct. 1996). For the active/control comparisons only minor differences are observable. However, one feature seems to be consistent for both types of substations and for both working and non-working days; when comparing the last period with the first period the active substations seem to have reduced consumption in the evening when compared with the control. For the comparisons based on the individual substations the consumption after 18:00 (6 p.m.) is even larger in the last period than in the first, but this seems to be true for both active and control substations. In conclusion the diurnal profiles measured relative to the daily level do not seem to be seriously affected by the campaign. This is probably expectable since the logarithmic scale is used for the analysis. In this way the diurnal variation in kWh/h is dictated solely by the daily level and by a basic profile which express the relative change over the day and night. For instance, if the overall level for a substation is 80 kWh/h and if the campaign results in a 12% reduction and if the diurnal variation spans -40% to +75% as in Appendix C, then the amplitude of the diurnal profile will change from 92 kWh/h (48-140) to 81 kWh/h (42-123).

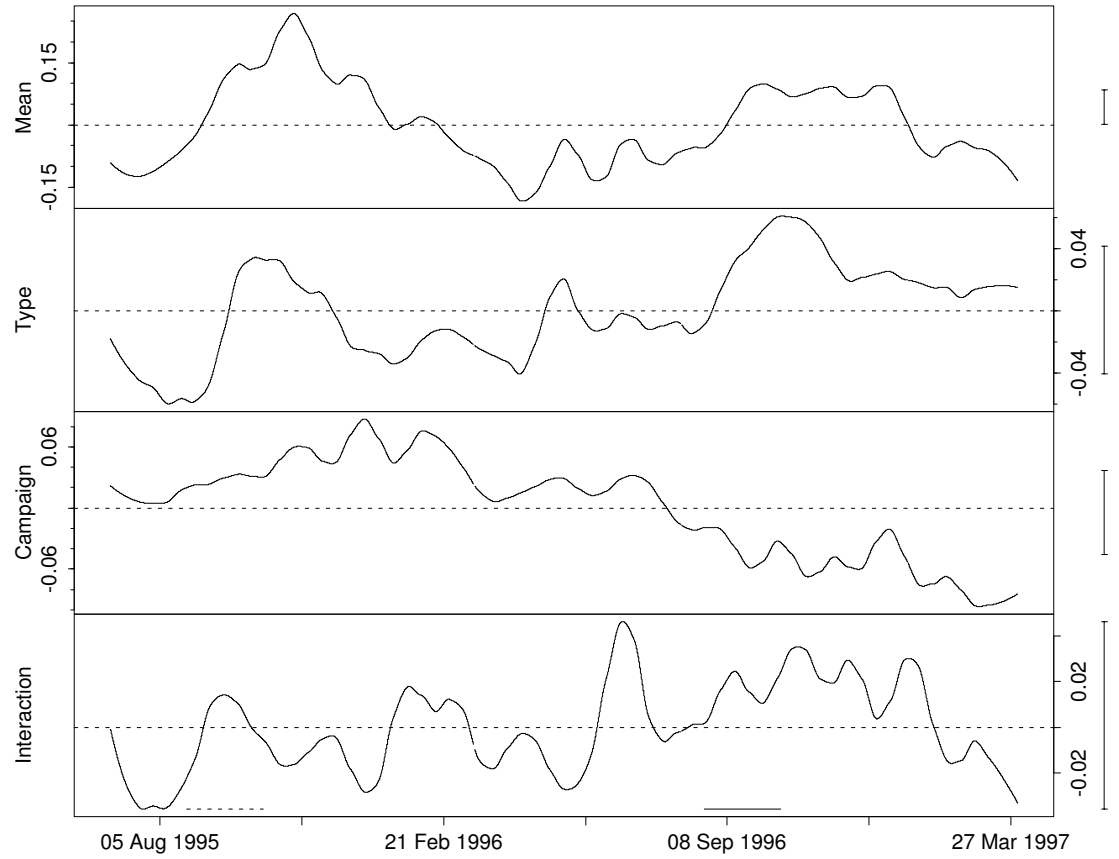

Figure 4.25. *Mean, type, campaign, and interaction of the four estimates of the trend component (log-scale) of (4.13) using a bandwidth of 10%. The bars on the right-hand side cover the same range.*

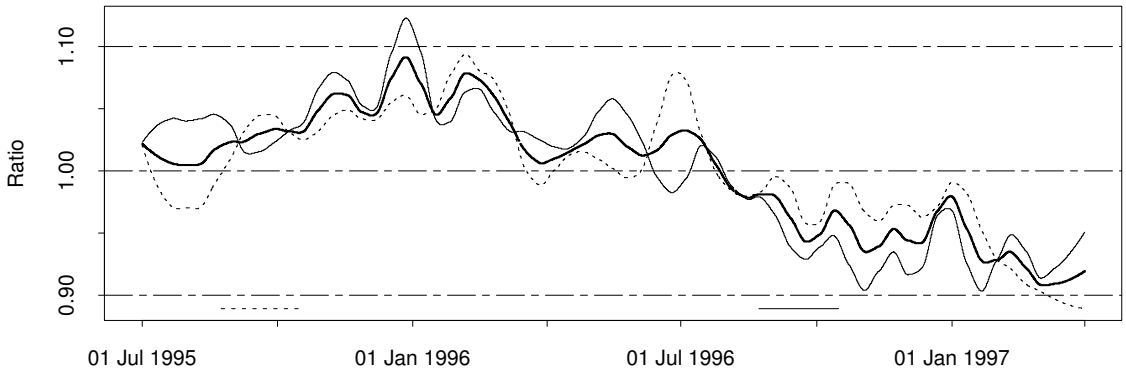

Figure 4.26. *Estimated ratio between the trend component of active and control for 4284 and 1325 (solid) and 1667 and 2588 (dotted), together with the overall ratio (bold). A 10% bandwidth is used for the estimation in (4.13).*

Pointwise Confidence Intervals

With the purpose of accessing the uncertainty of the claims stated above pointwise 95% confidence intervals are calculated for the estimates related to the trend, overall level, and dependence on degree days. Confidence intervals related to the diurnal variation are not presented since the size of the difference displayed in Figures C.1, C.2, C.3, and C.4 are judged to be unimportant anyhow.

The pointwise confidence intervals are obtained using bootstrapping, cf. Section 3.6. As shown above the residuals of model (4.13) are correlated, but this correlation is appropriately modelled by a simple autoregressive model. However, as pairwise comparisons between substations are important, see e.g. Figure 4.26, the cross-correlation between the residuals from the autoregressive models should be investigated, and possibly modelled.

Figure 4.27 shows the sample autocorrelation matrix after fitting an autoregressive model with non-zero parameters in lags 1, 24, and 25 to each of the four residual series. Since the software used for calculating the autocorrelation matrix (the `acf` function of S-PLUS (Statistical Sciences 1995)) do not allow missing values, the correlation values are calculated by just disregarding time points in which any of the observations are missing. The off-diagonal plots show a cross-correlation near 0.4 in lag 0 and 0.2 in lag ± 1.

For bootstrapping it is not necessary to remove the cross-correlation in lag zero. This correlation can be accounted for by sampling time points only and hence keeping the stochastic dependence between errors in lag zero intact. With the aim of reducing the cross-correlation, for non-zero lags and especially lag ± 1, a multivariate autoregressive model is fitted to the multivariate series consisting of the four residual series from (4.13).

The diagonal of the matrix of autoregressive operators (Jenkins & Alavi 1981) contains non-zero parameters at lag 1, 24, and 25, while the off-diagonal elements contain a non-zero parameter at lag 1. The model is fitted by observing that for each of the series the value at a particular time point is the sum of a white noise component corresponding to the time point and a linear combination of lagged values for the four series. Therefore one-step predictions are formed exactly like in autoregressive models with external input (ARX-models). This leads to an estimation procedure in which the parameters of the multivariate model are estimated as least squares estimates of the four corresponding ARX-models.

The sample autocorrelation matrix of the residuals from the multivariate model is displayed in Figure 4.28. Some improvement is seen, especially for the off-diagonal plots in lag ± 1. The improvements in lag ± 1 are highlighted in Table 4.2. Overall the cross-correlation in lag ± 1 is reduced by at least 1/3.

The residuals from the multivariate autoregressive model are judged to be close enough to multivariate white noise to allow for bootstrapping. To generate simulated

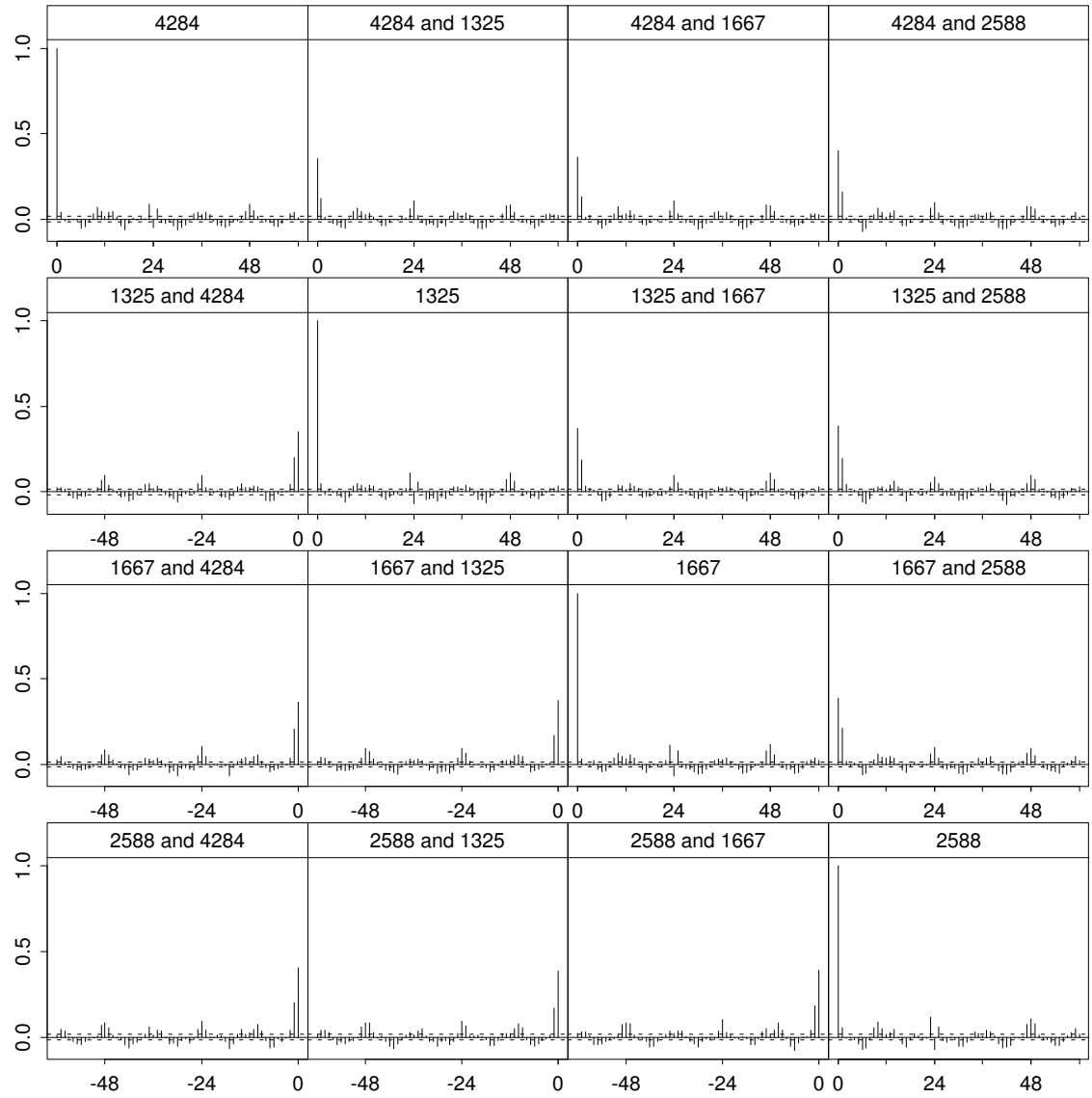

Figure 4.27. *Sample autocorrelation function matrix of the residuals of (4.13) after fitting an autoregressive model to each of the series. Approximate 95% confidence intervals are indicated by dotted lines.*

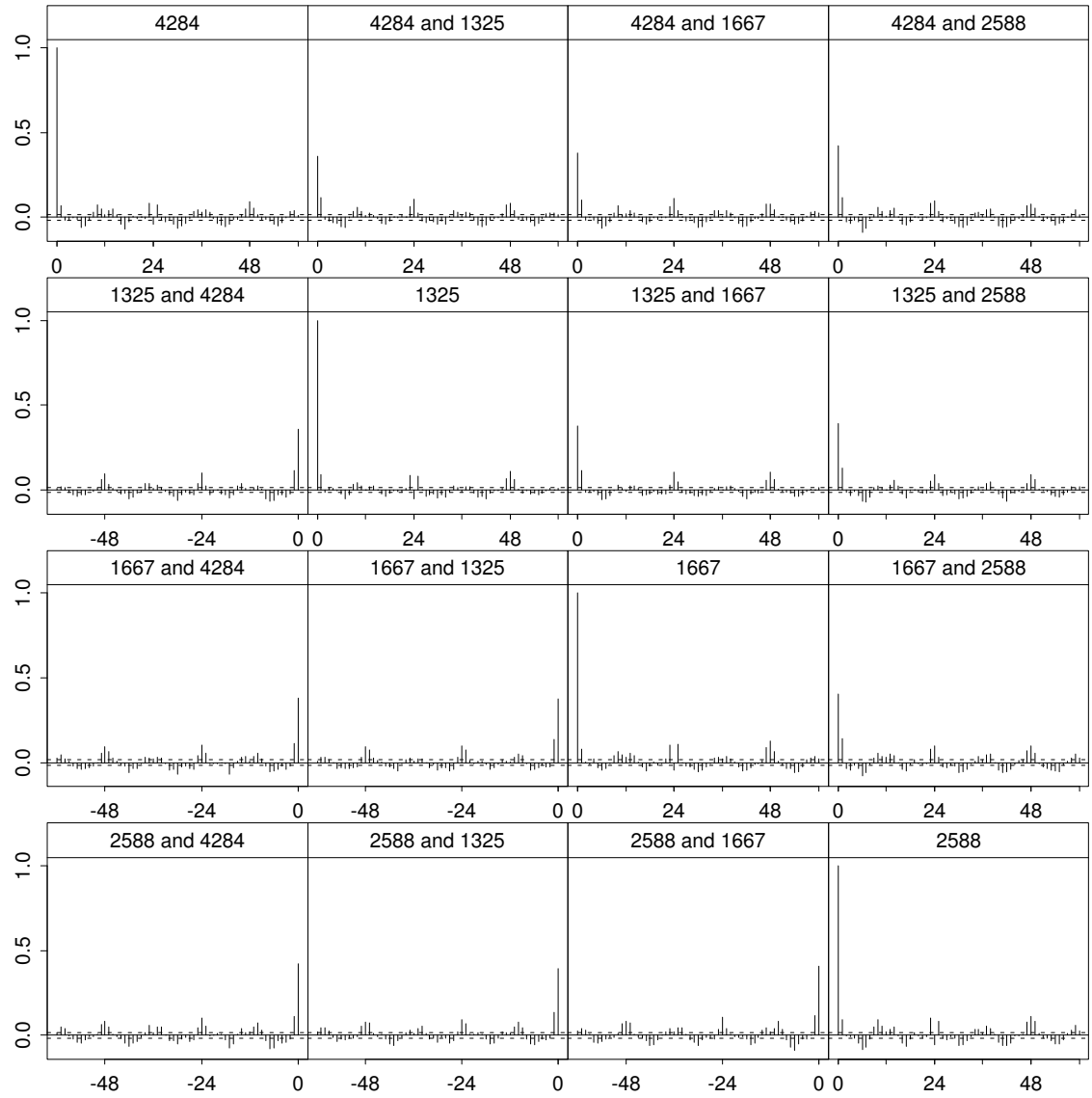

Figure 4.28. *Sample autocorrelation function matrix of the residuals of (4.13) after fitting an multivariate autoregressive model to the four series. Approximate 95% confidence intervals are indicated by dotted lines.*

	4284	1325	1667	2588
4284	0.041	0.124	0.132	0.159
1325	0.201	0.049	0.189	0.197
1667	0.206	0.171	0.035	0.211
2588	0.204	0.169	0.181	0.054

	4284	1325	1667	2588
4284	0.068	0.116	0.101	0.115
1325	0.116	0.091	0.116	0.129
1667	0.115	0.138	0.077	0.141
2588	0.110	0.134	0.114	0.094

Table 4.2. *Correlation values in lag* ± 1 *as displayed in Figures 4.27 (top) and 4.28 (bottom).*

observations the fitted values of (4.13) are calculated for each substation. Hereafter a sample of size 15358 (the number of hours in the data set) is drawn with replacement from the residuals of the multivariate autoregressive model, and by conditioning on the 25 first residuals from (4.13) a simulated error sequence corresponding to (4.13) is generated. To obtain simulated series of the dependent variable ($\log(P_t)$), for each substation, these simulated residuals are then added to the corresponding series of fitted values. The bootstrap replicate of the estimates is then generated by applying the original estimation procedure to the simulated data.

Four hundred bootstrap replicates were generated, each using approximately 5.5 CPU minute on a HP 9000/800. Since the number of observations are large and the estimates are linear combinations of the adjusted observations, cf. the backfitting-algorithm described in Section 3.5, teoretical considerations suggest that confidence intervals can be found by assuming that the estimates are normally distributed. Furthermore, this is validated by inspecting the bootstrap replicates of the estimate of the trend component in (4.13). More precisely the mean and median is plotted together with 95% percentile and standard normal intervals. Differences are only revealed when using a large plot area, cf. Figure C.9, and even on this plot the difference is hardly visible. Consequently, if the standard normal intervals are calculated based on estimates on the logarithmic scale, or based on sums/differences on this scale, it will probably result in results of high precision. This approach is applied in the following.

Table 4.3 contains 95% confidence intervals of the level and dependence on degree days. These are indeed quite narrow and this confirms that the dependence on degree days is significant.

On the following figures the period in which the households were visited and the corresponding period one year before are indicated by horizontal lines, solid and dotted, respectively. Figure 4.29 shows 95% confidence intervals of the four estimates of the trend components. For both types of substations there is a tendency of the control

Substation	$e^{\hat{a}}$ (kWh/h)	$\hat{b}$ $(h/°C)$
4284	[58.2, 59.6]	[0.731, 0.779]
1325	[66.5, 68.0]	[0.673, 0.719]
1667	[83.3, 85.4]	[0.828, 0.880]
2588	[94.9, 97.0]	[0.826, 0.875]

Table 4.3. *95% confidence intervals of the level ($e^{\hat{a}}$) and dependence on degree days ($\hat{b}$) for all four substations.*

substation being below the active substation at the beginning and above at the end of the period. However, these plots do not take the correlation between substations into account. For this reason each bootstrap replicate of the trend component (four sets of estimates) is split into mean, type, campaign, and interaction effect (which is an additive operation on the logarithmic scale).

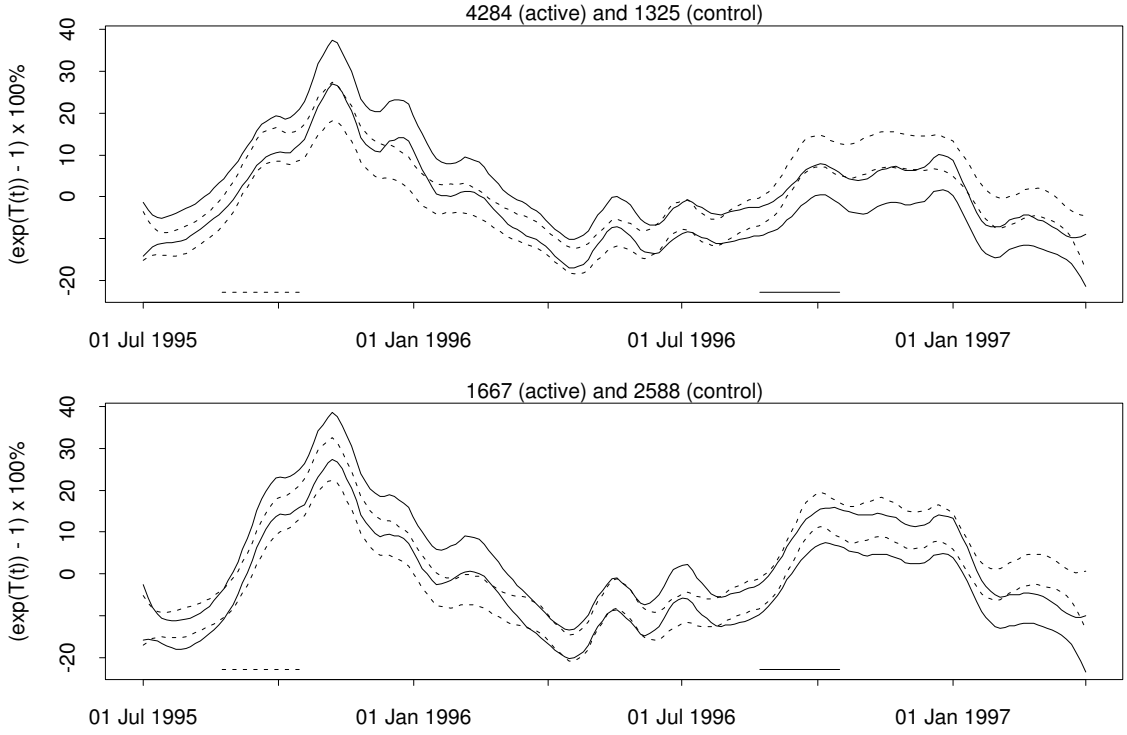

Figure 4.29. *95% confidence intervals of the estimates of the trend component. Results from control substations are indicated by dotted lines.*

Figure 4.30 shows 95% bootstrap confidence intervals of these estimates. It is seen that the interaction effect often is close to zero and compared with Figure 4.29 the effect of the campaign is far more clearly revealed. Based on the bootstrap replicates of campaign and interaction effects the 95% confidence intervals of the ratio between active and control, displayed in Figure 4.31, are found. Comparing the period after

the trial with the same period one year before (using the horizontal lines as guidance) indicate an overall reduction of 10% or more. For substations 4284 and 1325 the reduction is larger; 15% or more. For substations 1667 and 2588 the reduction is smaller; around 5%, except near the end of the measuring period were it is 10%.

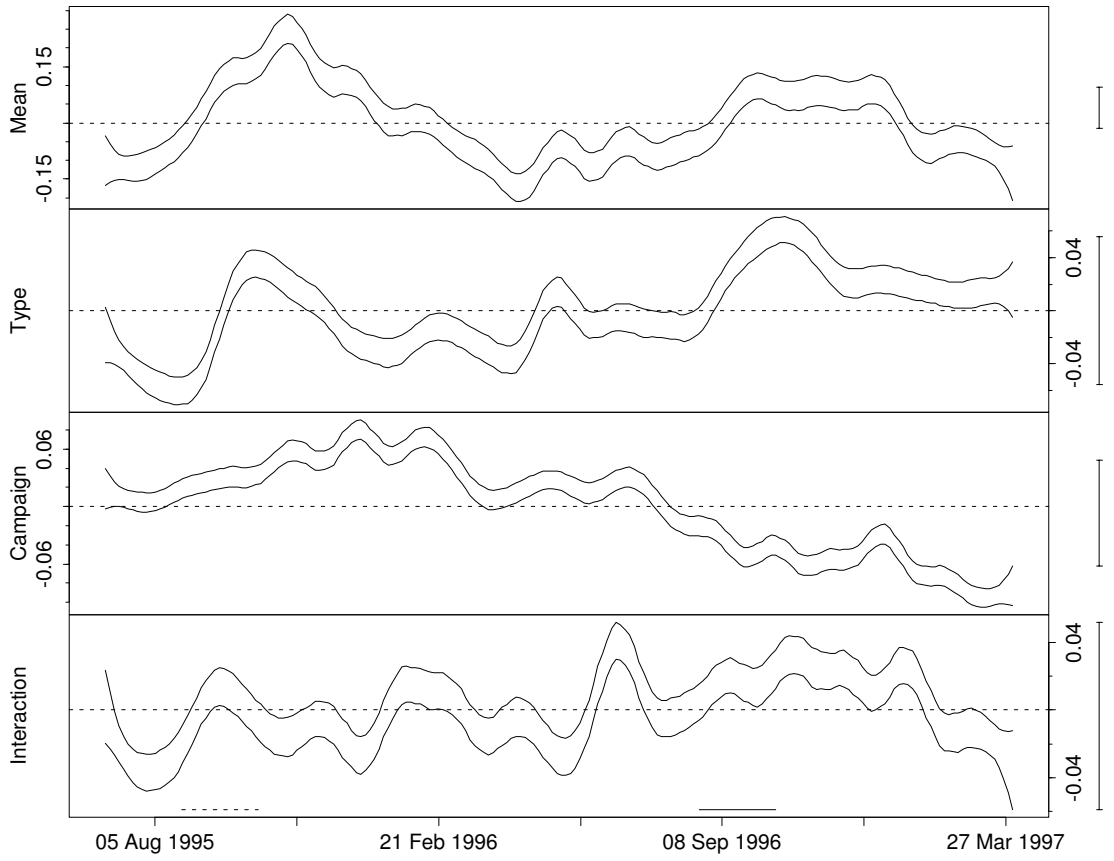

Figure 4.30. *95% confidence intervals of the mean, type, campaign, and interaction effects of the estimates of trend component on the logarithmic scale. The vertical bars cover the same range on all plots.*

Figure 4.32 further clarifies these aspects. On this figure the differences between corresponding dates in the first and last part of the confidence limits of the ratio between active and control substations are displayed. Ideally 95% confidence intervals of these differences should be calculated. However, this amounts to calculating quantiles of a stochastic variable which is the result of a difference between two, possibly dependent, log-normal stochastic variables. Such a variable will have long tails and hence the percentile intervals will be uncertain, and the standard normal intervals cannot be expected to be correct.

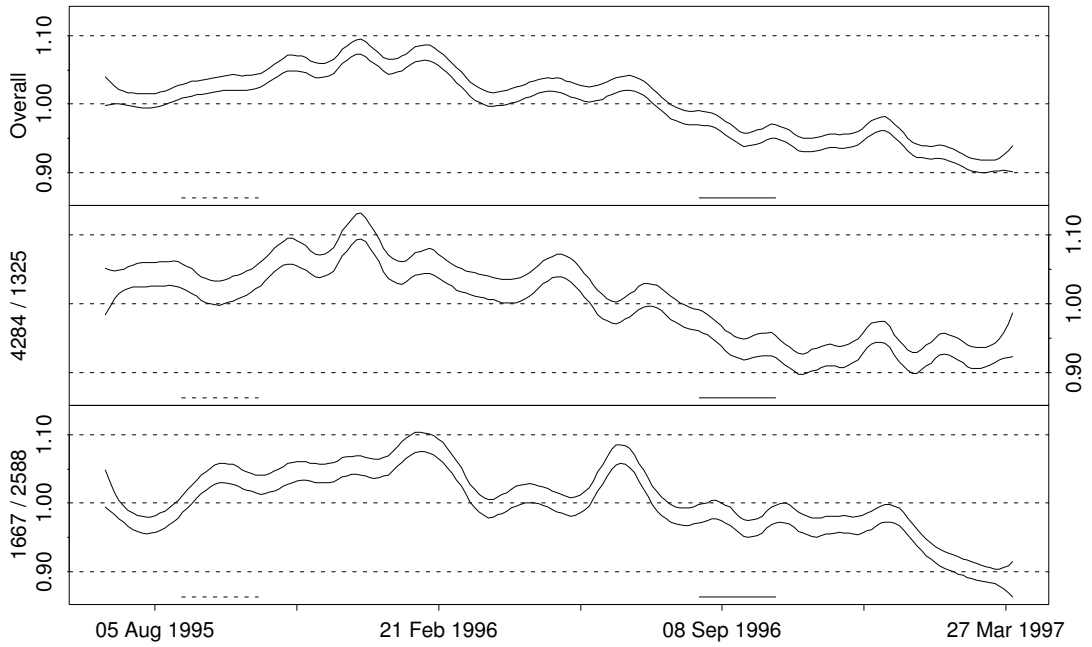

Figure 4.31. *95% confidence intervals of the overall ratio between active and control (top) and the ratio for each type (middel, bottom).*

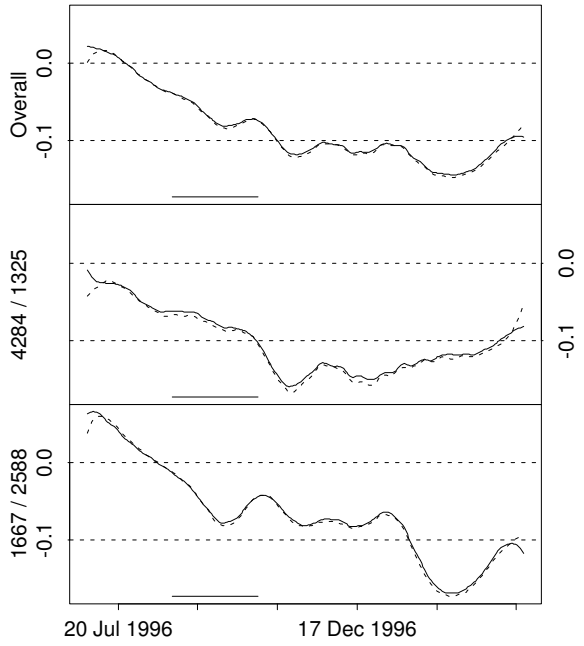

Figure 4.32. *Difference between lower (solid) and upper (dotted) confidence limits of the ratios between active and control in the first and last part of the curves displayed in Figure 4.31. Dates corresponding to the last part are used on the x-axis. Negative values correspond to a decrease over time.*

Chapter 5

Conclusion

The purpose of the project described in this report is to develop and investigate methods for estimating effects of power conservation campaigns, and to estimate the effect of a particular campaign. The main concepts used are (i) grouping of substations during the early phase of planning with the purpose of defining groups of substations which can serve as control/active in the experiment, and (ii) analysis of data taking the predefined groups into account. It is demonstrated that non-parametric and semi non-parametric methods, combined with traditional time series analysis and bootstrapping are well suited as statistical tools for these kind of data.

Measurements from a controlled trial addressing power conservations in households are analysed. The measurements are power consumptions over 15 minute intervals, measured at substation level. Before analysis these values are summed to hourly values. Two active substations, each with a control substation, are included in the trial.

Statistical analysis suggest that logarithmic transformed data should be used. Hence, the quantification of effects as percentages, instead of units of power consumption (kWh), are both statistical well founded and intuitively appealing.

For one substation the difference between years is quite large. For this reason the controls are extremely important. With respect to the number of substations it may be argued that the trial is not very large. On the other hand 225 households have been visited and the four substations supply 631 households. The homogeneity of households with respect to the tendency to respond to a campaign is probably larger within substations than between substations. For this reason the 631 households can not be regarded as random sample and the "effective number of households" is thus smaller. From Figure 4.24 it is evident that the two controls behave quite similar, also after the trial, and for this reason we are fairly confident that the findings regarding the active substations are due to the trial, i.e. a fairly large reduction for substation 4284 (Vindinge) and a smaller reduction for substation 1667 (Trørød). The overall results indicate a 10-12% reduction of the power consumption.

However, due to the observed difference between the two pairs of active and control substations, an extrapolation to other substations of similar kind will be difficult.

These results are valid both in the case where no information regarding climate is used, and when part of the long term variation in data is accounted for by the monthly degree day values. This indicates that when using controls the climate can be disregarded, at least with respect to the overall results.

When the diurnal profile is measured as the relative deviation from the daily level during the cycle, practically no effect of the campaign can be detected. This may be due to the fact that the diurnal variation is very large (-40 to +75%). Of cause, the amplitude of the profile measured in units of power consumption is strongly dependent on the daily level.

For the case where information regarding climate is used the variation over time is addressed. When comparing the time points after the campaign with the same time points one year before it is clearly seen that, compared with the control, substation 4284 (Vindinge) responds faster than substation 1667 (Trørød), and except for the last part of the period the reduction is largest for substation 4284, cf. Figure 4.32. This figure also indicates that the response is temporary, or at least larger in the months following the trial than can be expected on the long run. This should be further investigated when one or two more years of data become available.

Chapter 6

Discussion

As mentioned in the conclusion the results indicate that all hourly values of power consumption are reduced proportional to their size when carrying out a campaign, i.e. the diurnal variation seems to be dictated by the daily level of power consumption. This suggests other experimental designs in which the power consumption is measured for the individual households, but as sums over long periods (say three months or more). If measurements can be obtained inexpensively, e.g. by consumer readings of their usual equipment, this gives the possibility of experiments of which the results are more representative than the ones obtained in this trial. The ideal is random sampling of households for participation in the trial, but since not all households will be willing to participate, the representativeness of the trial must still be considered during design.

Most of the methods used in this report are of the non-parametric or semi non-parametric type. These seem to be valuable, especially for the case of trend component estimation. The estimation may be regarded as a signal extraction in which we condition on the full set of data and apply smoothness constraints to obtain the estimates. Alternatively the trend and seasonal components could, in principle, be estimated by postulating a stochastic model for the trend, maybe a random walk, and a stochastic component for the diurnal variation, maybe a seasonal walk (Madsen 1995). Estimation would then be feasible if the noise could be assumed to concentrate around zero. Although the estimates will not be smooth the approach is similar to smoothing in that the estimates have reduced variance compared with the original data. Hence this kind of smoother could probably also be used together with the backfitting algorithm. In this context also the Kalman smoother seems interesting (Shumway 1988, p. 176-7).

In some cases information criteria are used for bandwidth selection. Cross-validation is not used since it is too computational demanding. It is interesting to note that the bandwidth selected is strongly dependent on whether the temporal correlation of the residuals is modelled or not. Since the information criteria are likelihood based at least one of the models considered should have white noise residuals, i.e. the true

model should be among those considered, for the selection to work appropriately. A more intuitive explanation of this is also available, it is based on leave-one-out cross-validation (CV), for which the GVC and AIC are approximations. If the bandwidth selection is based on CV, and if the residuals are positive correlated, the CV will tend to select a very small bandwidth (Friedman 1984, p. 10). This is exactly what happens on Figure 4.21; a bandwidth of one percent is selected by GCV and AIC.

BIC leads to totally different bandwidths than AIC and GCV. A closer look on the assumptions on which BIC (Schwarz 1978) is based may explain why. The assumptions on the a priori probability of the models, i.e. bandwidths in this case, imply mutual orthogonality of the conditional a priori distribution of the model parameters, i.e. trend component and parameters in the linear part of the model. Since any bandwidth can be selected an infinity of different conditional a priori distributions of model parameters exists, and since the equivalent number of parameters is finite, these can not all be orthogonal. This implies that the assumptions on which BIC is based cannot be fulfilled in the case of bandwidth selection.

Appendix A

Data Tables

Date			Degree days ($^{\circ}C/h$)
30	Jun.	95	0
31	Jul.	95	–
31	Aug.	95	–
30	Sep.	95	38
31	Oct.	95	103
30	Nov.	95	388
31	Dec.	95	566
31	Jan.	96	573
29	Feb.	96	564
31	Mar.	96	520
30	Apr.	96	250
31	May.	96	170
30	Jun.	96	0
31	Jul.	96	–
31	Aug.	96	–
30	Sep.	96	46
31	Oct.	96	193
30	Nov.	96	352
31	Dec.	96	538
31	Jan.	97	555
28	Feb.	97	393
31	Mar.	97	405

Table A.1. *Total degree days by month, backwards in time.*

	Substation			
	1667		4284	
Date	No.	Cum.	No.	Cum.
23 Aug.	2	2	0	0
26 Aug.	11	13	5	5
27 Aug.	6	19	5	10
28 Aug.	3	22	3	13
29 Aug.	13	35	4	17
30 Aug.	8	43	0	17
1 Sep.	0	43	1	18
2 Sep.	9	52	8	26
3 Sep.	7	59	0	26
4 Sep.	3	62	5	31
5 Sep.	6	68	4	35
6 Sep.	2	70	4	39
8 Sep.	2	72	0	39
9 Sep.	5	77	7	46
10 Sep.	3	80	6	52
11 Sep.	5	85	0	52
12 Sep.	10	95	9	61
13 Sep.	4	99	0	61
16 Sep.	2	101	6	67
17 Sep.	3	104	6	73
18 Sep.	6	110	7	80
19 Sep.	2	112	5	85
20 Sep.	0	112	1	86
23 Sep.	2	114	5	91
24 Sep.	1	115	2	93
25 Sep.	0	115	3	96
26 Sep.	1	116	1	97
29 Sep.	0	116	0	97
30 Sep.	1	117	3	100
1 Oct.	0	117	3	103
2 Oct.	1	118	0	103
3 Oct.	1	119	1	104
8 Oct.	0	119	1	105
16 Oct.	1	120	0	105

Table A.2. *Number of visits on induvidual dates.*

Appendix B

Decomposition of the Individual Time Series

This appendix contains plots of the decompositions of the individual time series addressed in Section 4.1. Seasonal diagnostic plots are also included.

Figures B.1, B.2, B.3, and B.4 show the original data and the components. In these figures the vertical bar on the right of the plots covers the same range on the individual figures. The horizontal line at the bottom of each plot indicates the period in which the households were visited, cf. Section 2.1, the same period one year before is indicated by a dotted line.

Figures B.5, B.6, B.7, and B.8 show the corresponding seasonal diagnostic plots. They show, for each hour of the day (printed on top of each sub-plot), the detrended data plotted against the date, together with the seasonal smooth. On each page the range of the axes are identical. Only data corresponding to normal working days are shown, plots corresponding to the other types of days (see the beginning of Chaper 4) are simillar, but with fewer data points.

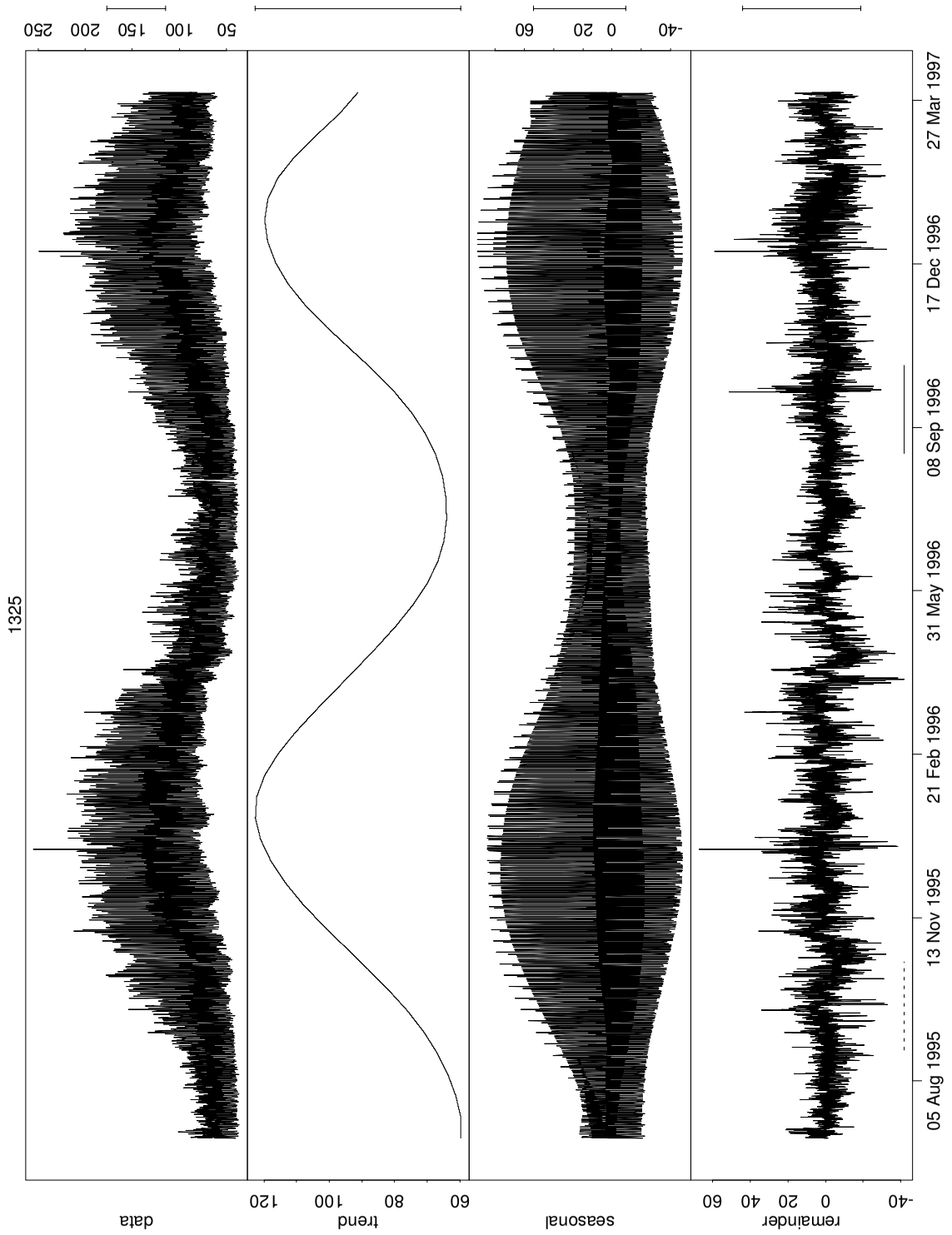

Figure B.1. *Series 1325, together with trend, seasonal, and remainder components.*

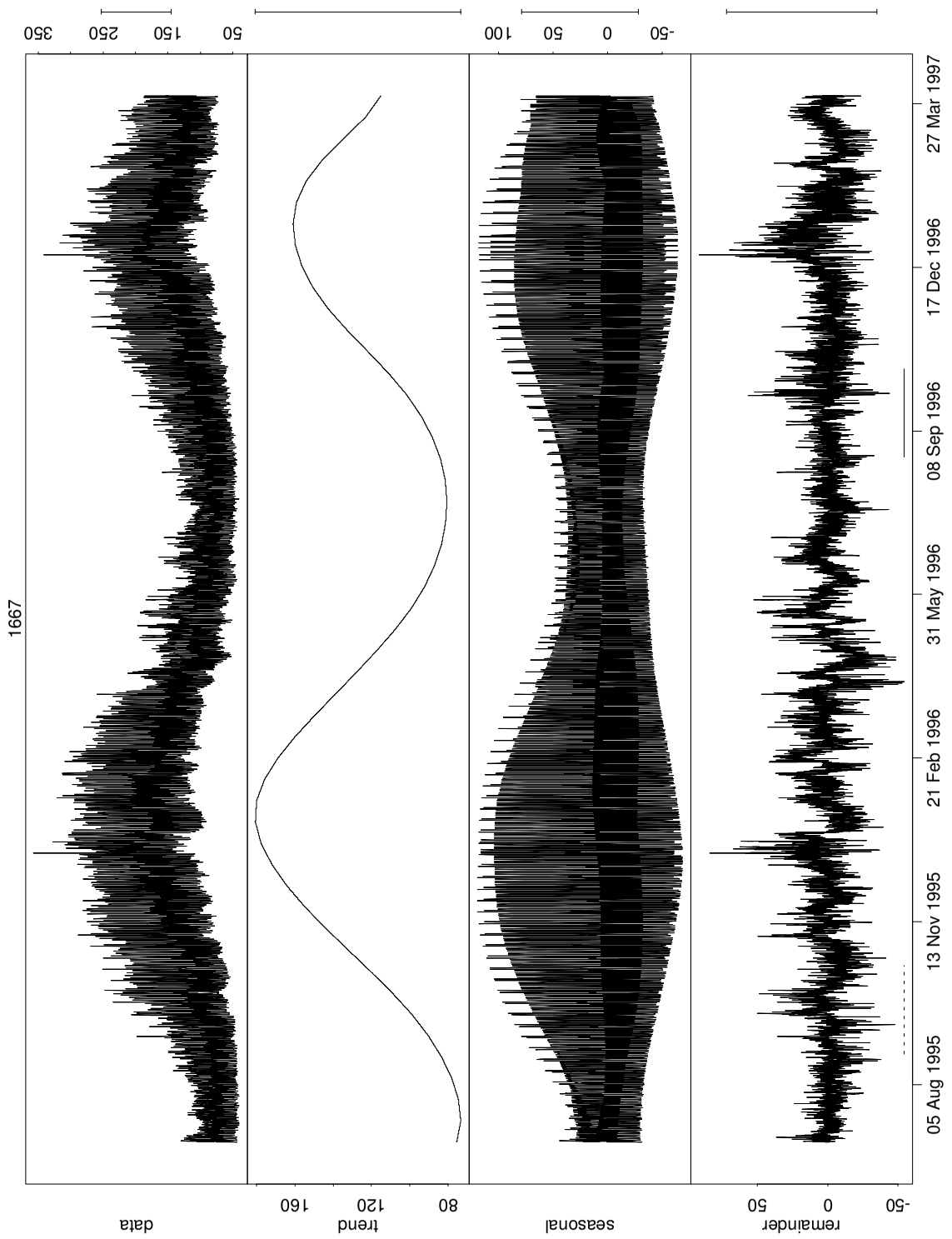

Figure B.2. *Series 1667, together with trend, seasonal, and remainder components.*

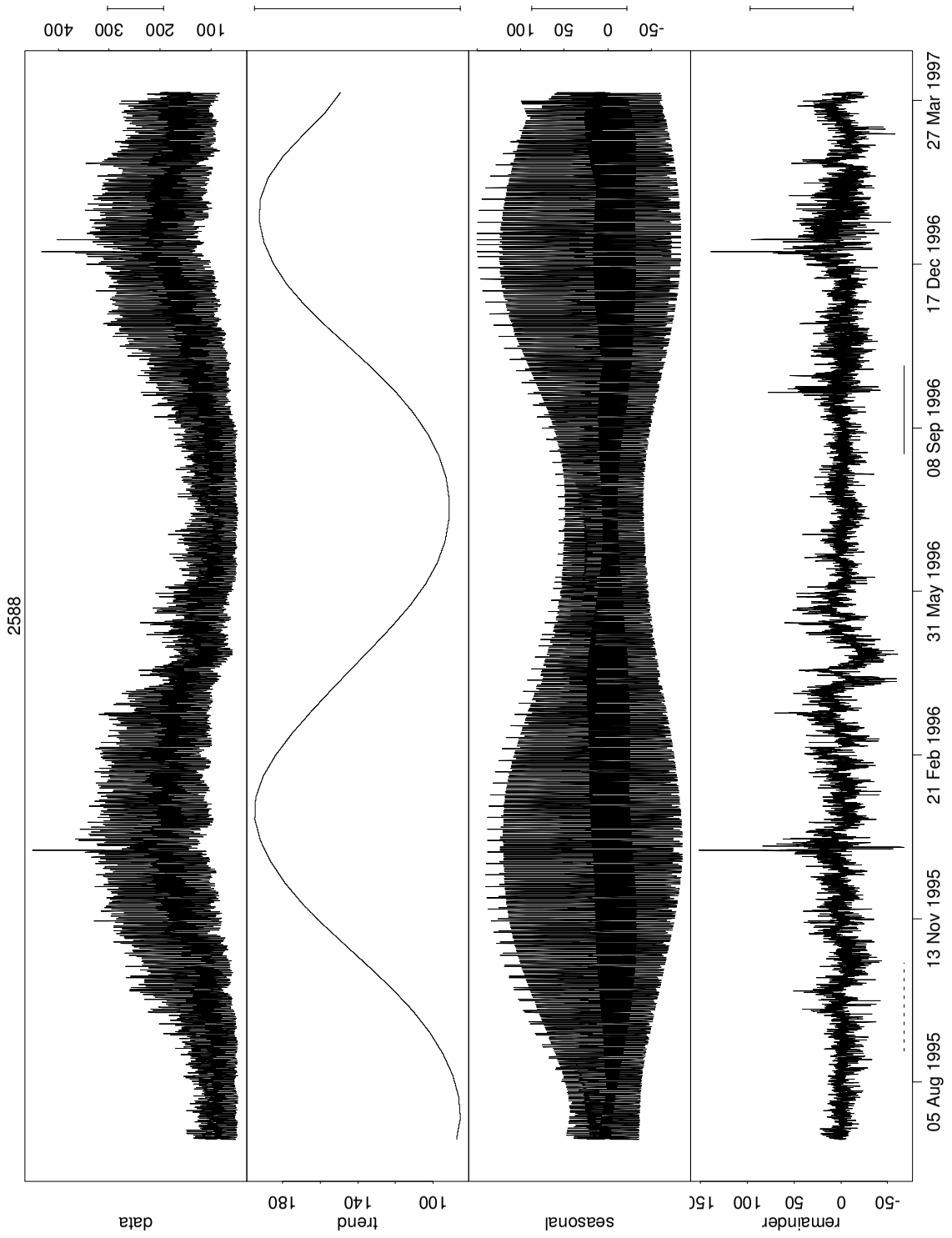

Figure B.3. *Series 2588, together with trend, seasonal, and remainder components.*

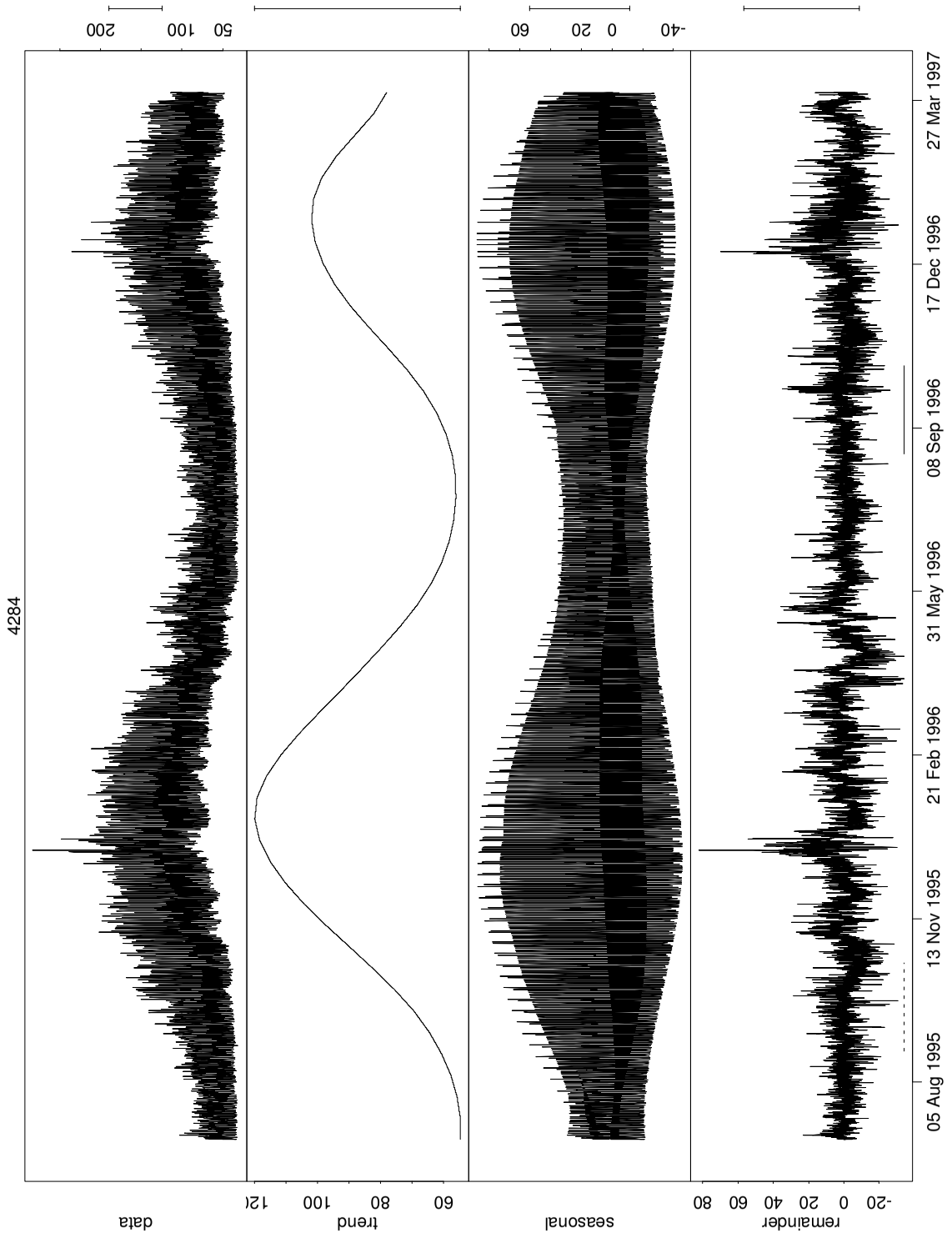

Figure B.4. *Series 4284, together with trend, seasonal, and remainder components.*

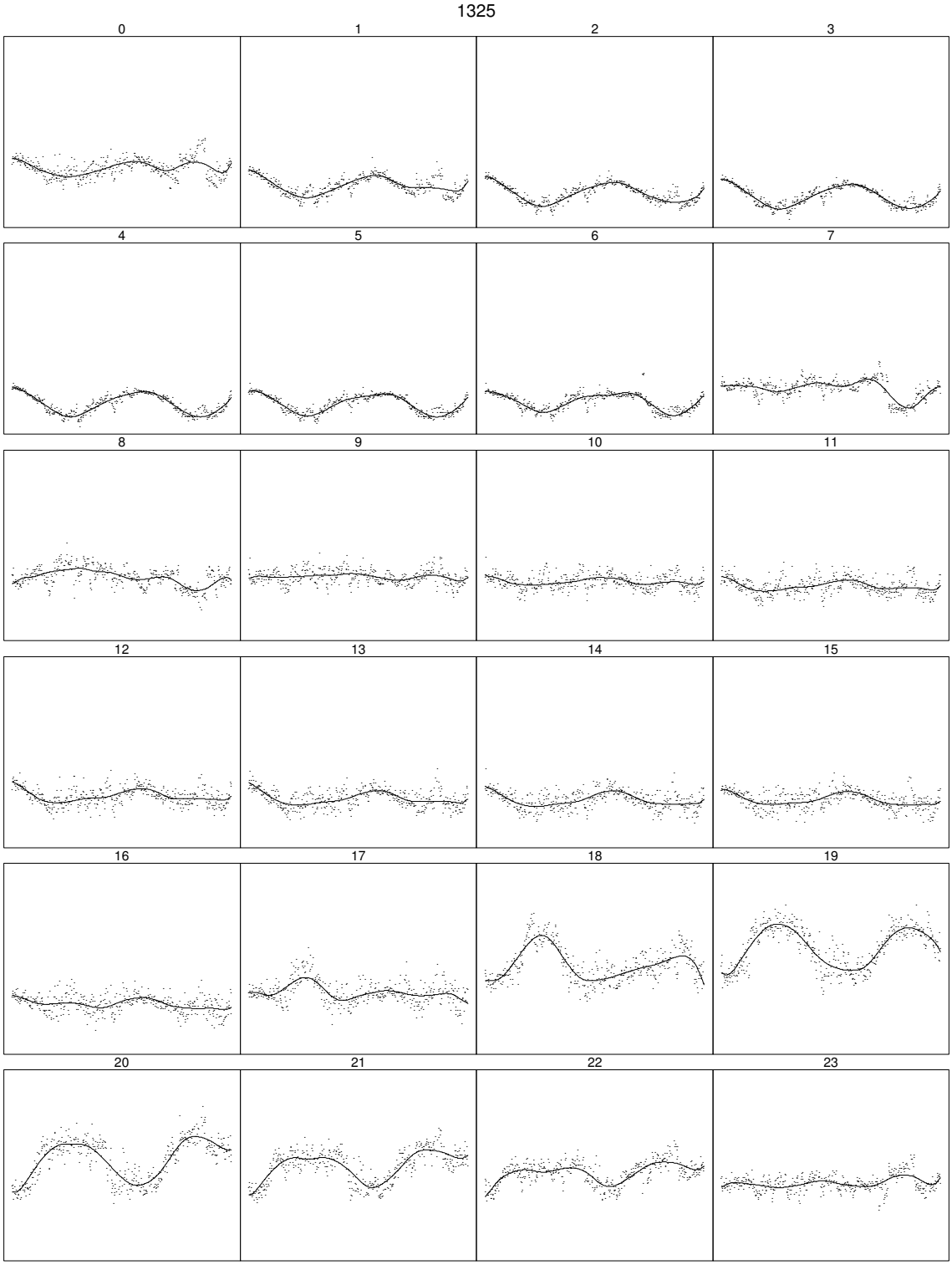

Figure B.5. *Seasonal diagnostic plot for substation 1325 (working days).*

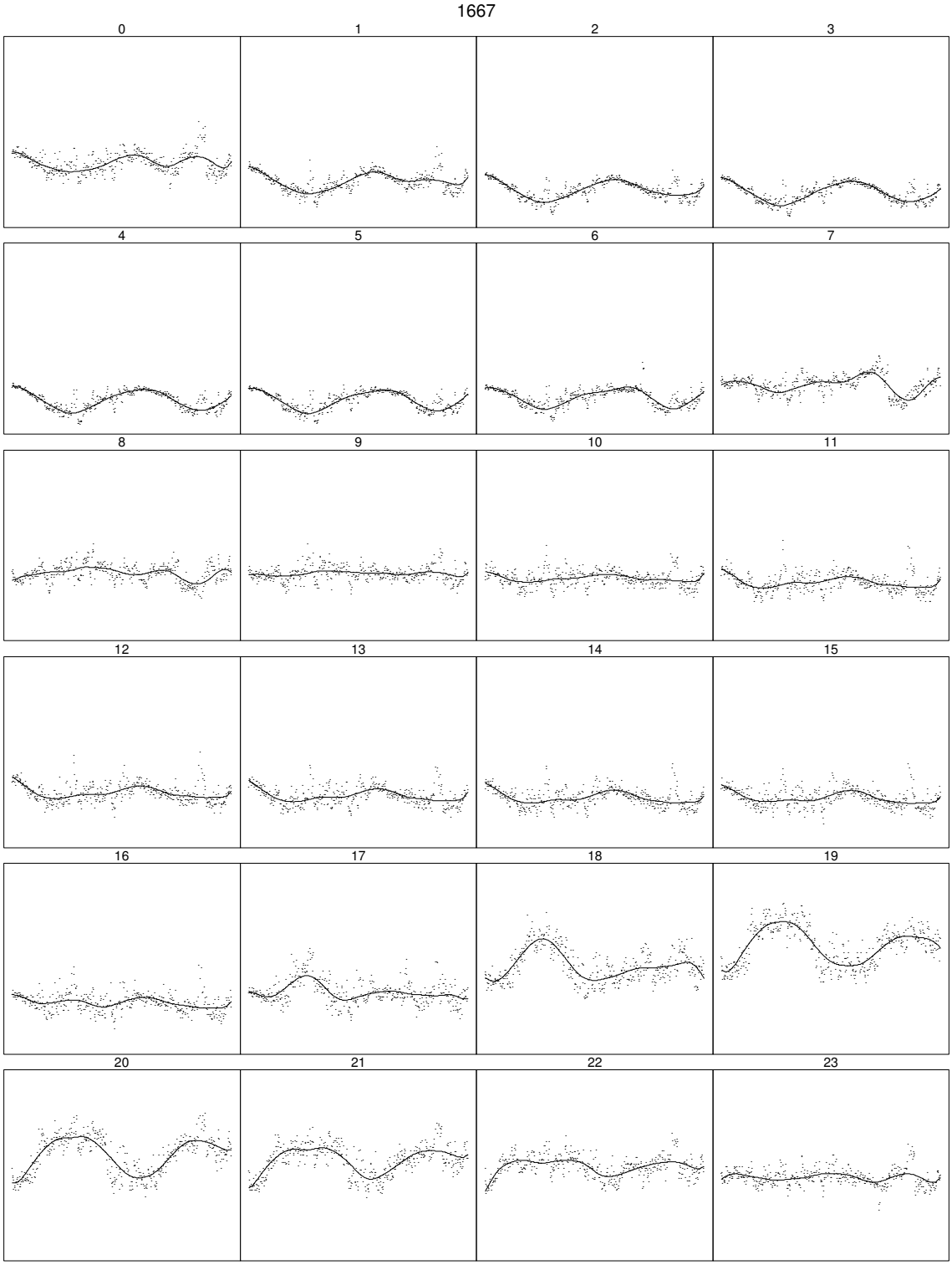

Figure B.6. *Seasonal diagnostic plot for substation 1667 (working days).*

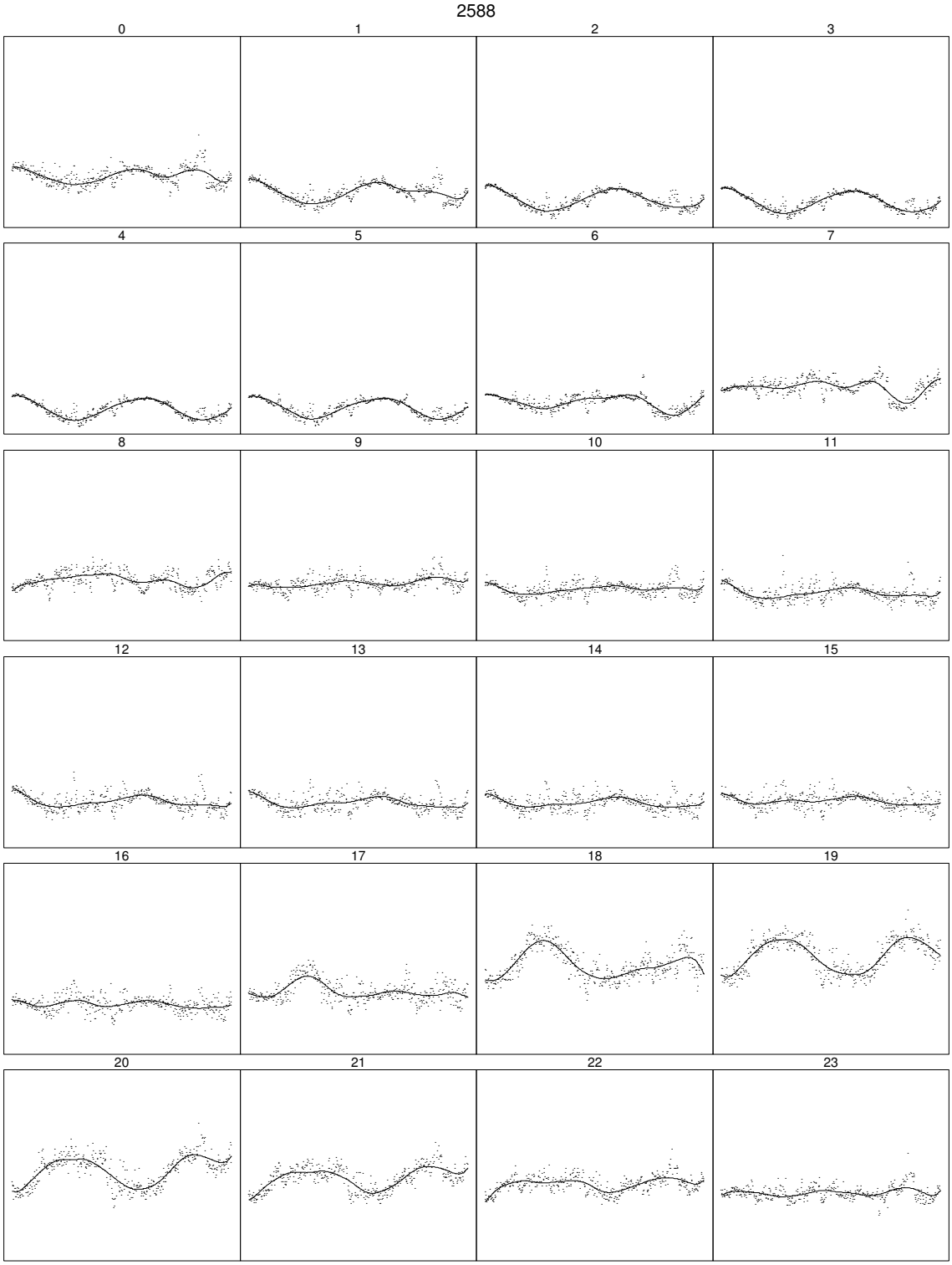

Figure B.7. *Seasonal diagnostic plot for substation 2588 (working days).*

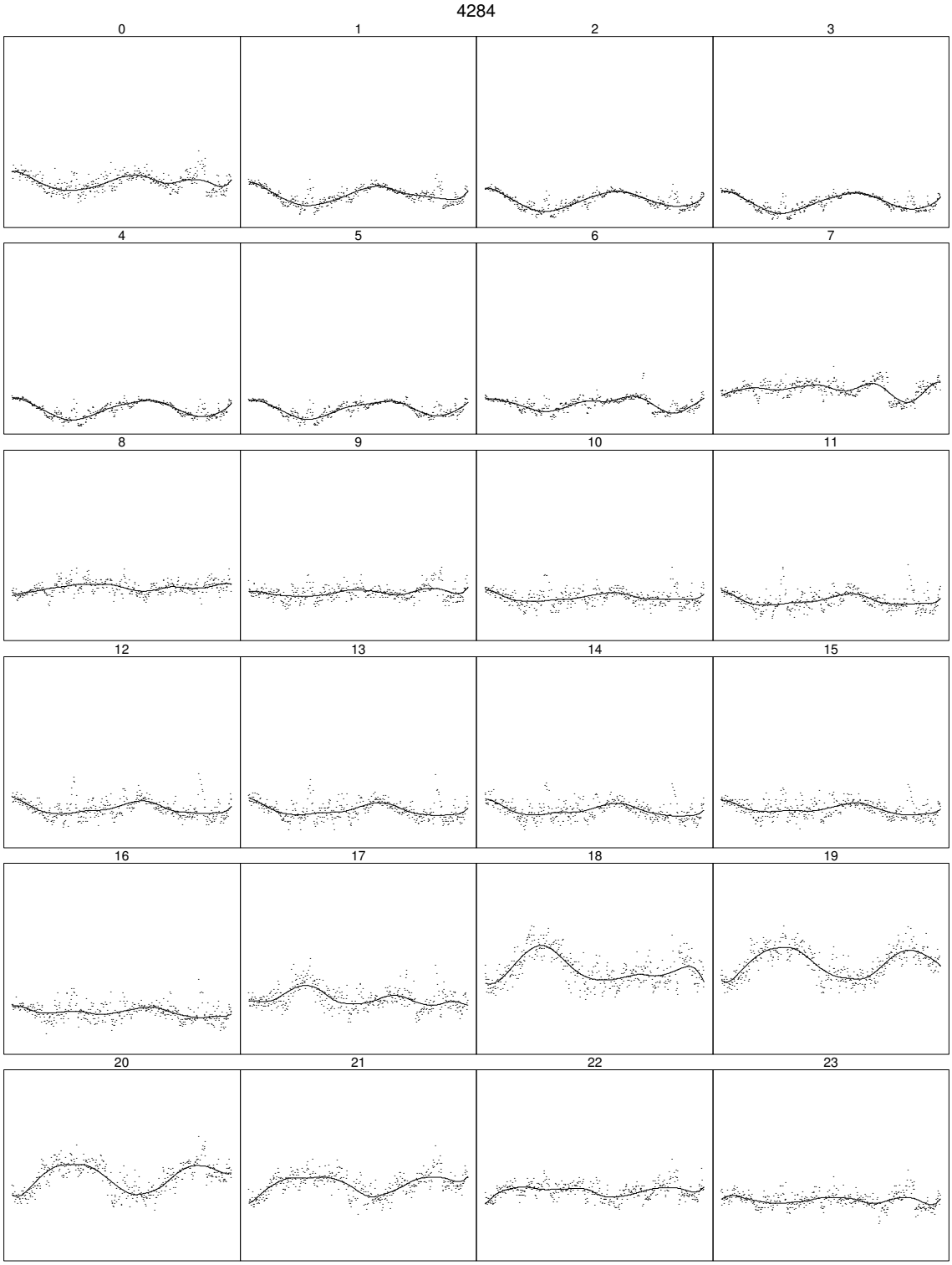

Figure B.8. *Seasonal diagnostic plot for substation 4284 (working days).*

Appendix C

Modelling the Dependence on Climate

This appendix contains plots related to Section 4.5.2 and model (4.13).

- Figures C.1, C.2, C.3, and C.4 show the estimated diurnal profiles based on (4.13) using a bandwidth of 10%. The results are presented as the percent-wise deviation from the daily level plotted against the time of day. On the plots the vertical lines not extended beyond 00 – 24 indicate the mean of the diurnal profile with which they share line type (solid or dotted). The calculation of means are performed on the log-scale.
- Figures C.5, C.6, C.7, and C.8 are diagnostic plots of the residuals, after fitting (4.13) using a bandwidth of 10%. For a description of box plots and normal quantile-quantile plots, see (Statistical Sciences 1993, Statistical Sciences 1995). The sample autocorrelation functions shown are calculated after fitting a zero mean autoregressive model with non-zero parameters at lags 1, 24, and 25.
- Figure C.9 displays the mean and median together with 95% percentile and standard normal intervals based on the bootstrap replicates of the estimate of the trend component, cf. page 54 (Section 4.5.2), for substation 4284.

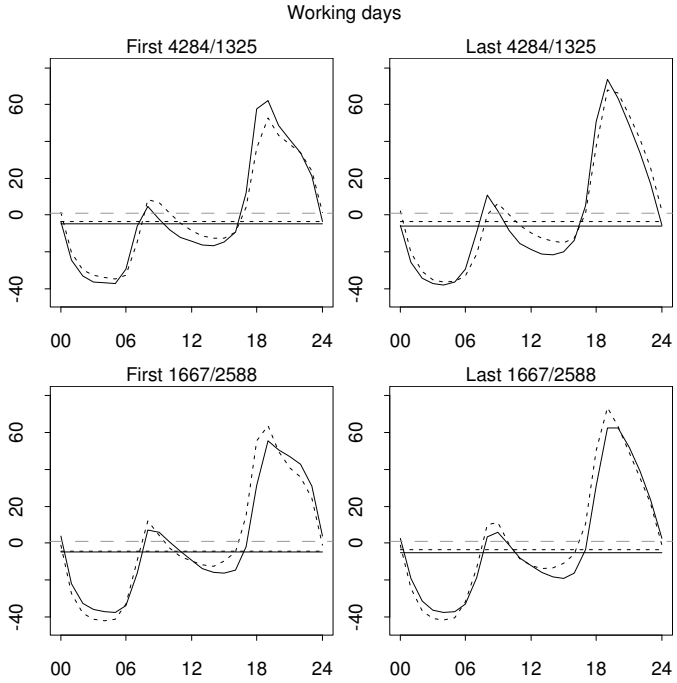

Figure C.1. *Diurnal profiles: Comparison of active (solid) and control (dotted) for working days and for both types of substations and for the first (before 1 Oct. 1996) and last (after 1 Oct. 1996) period.*

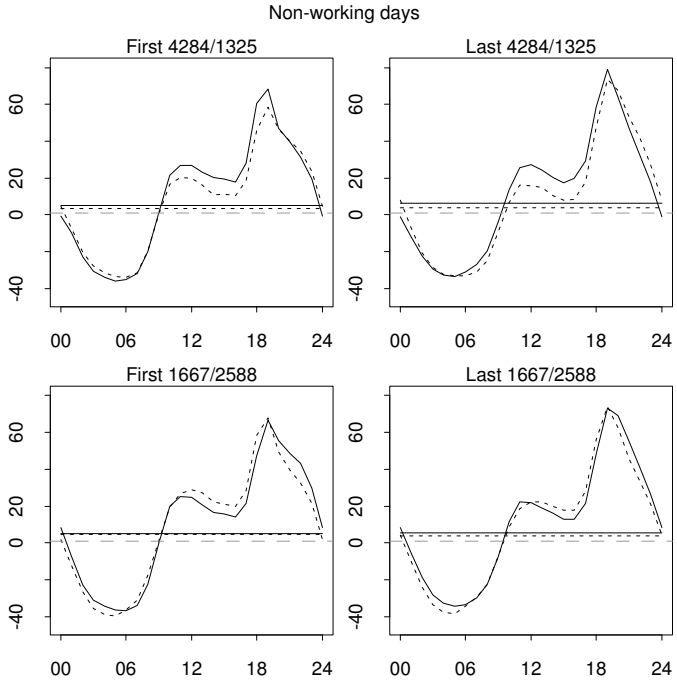

Figure C.2. *Diurnal profiles: Comparison of active (solid) and control (dotted) for non-working days and for both types of substations and for the first (before 1 Oct. 1996) and last (after 1 Oct. 1996) period.*

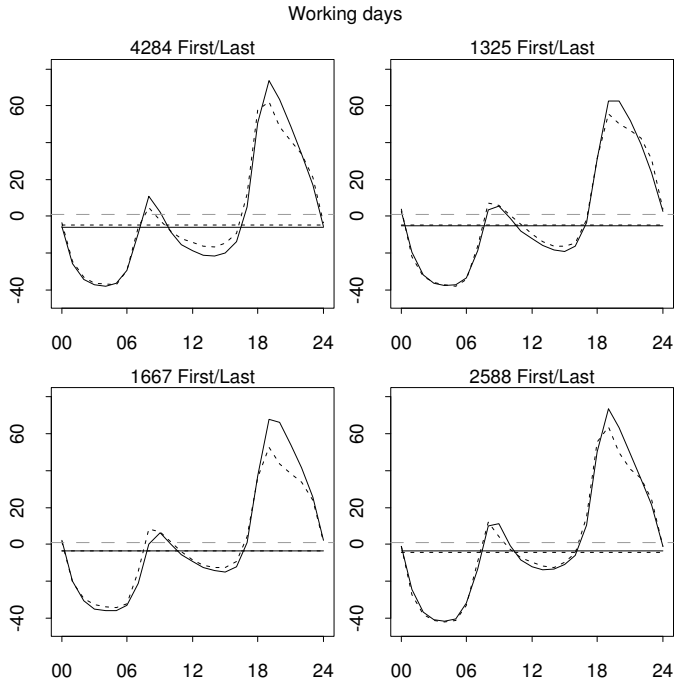

Figure C.3. *Diurnal profiles: Comparison of periods for working days for the individual substations. Before 1 Oct. 1996 (dotted) and after 1 Oct. 1996 (solid).*

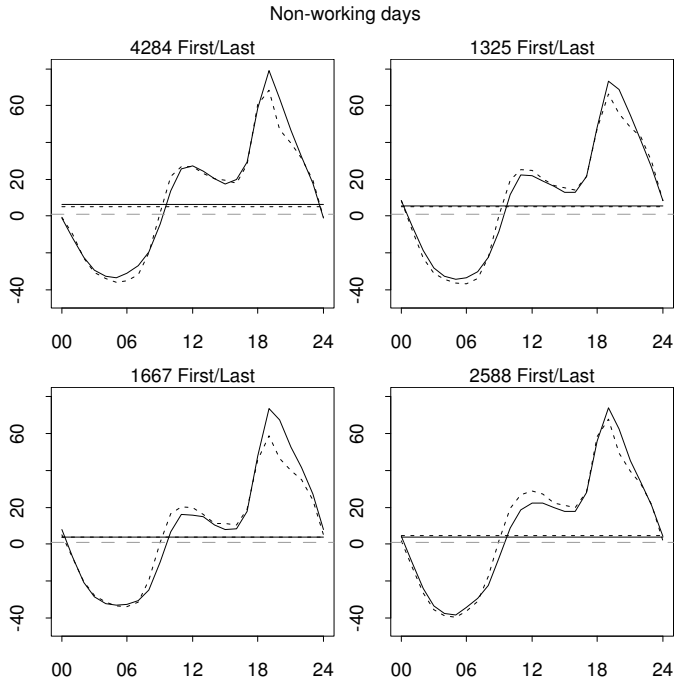

Figure C.4. *Diurnal profiles: Comparison of periods for non-working days for the individual substations. Before 1 Oct. 1996 (dotted) and after 1 Oct. 1996 (solid).*

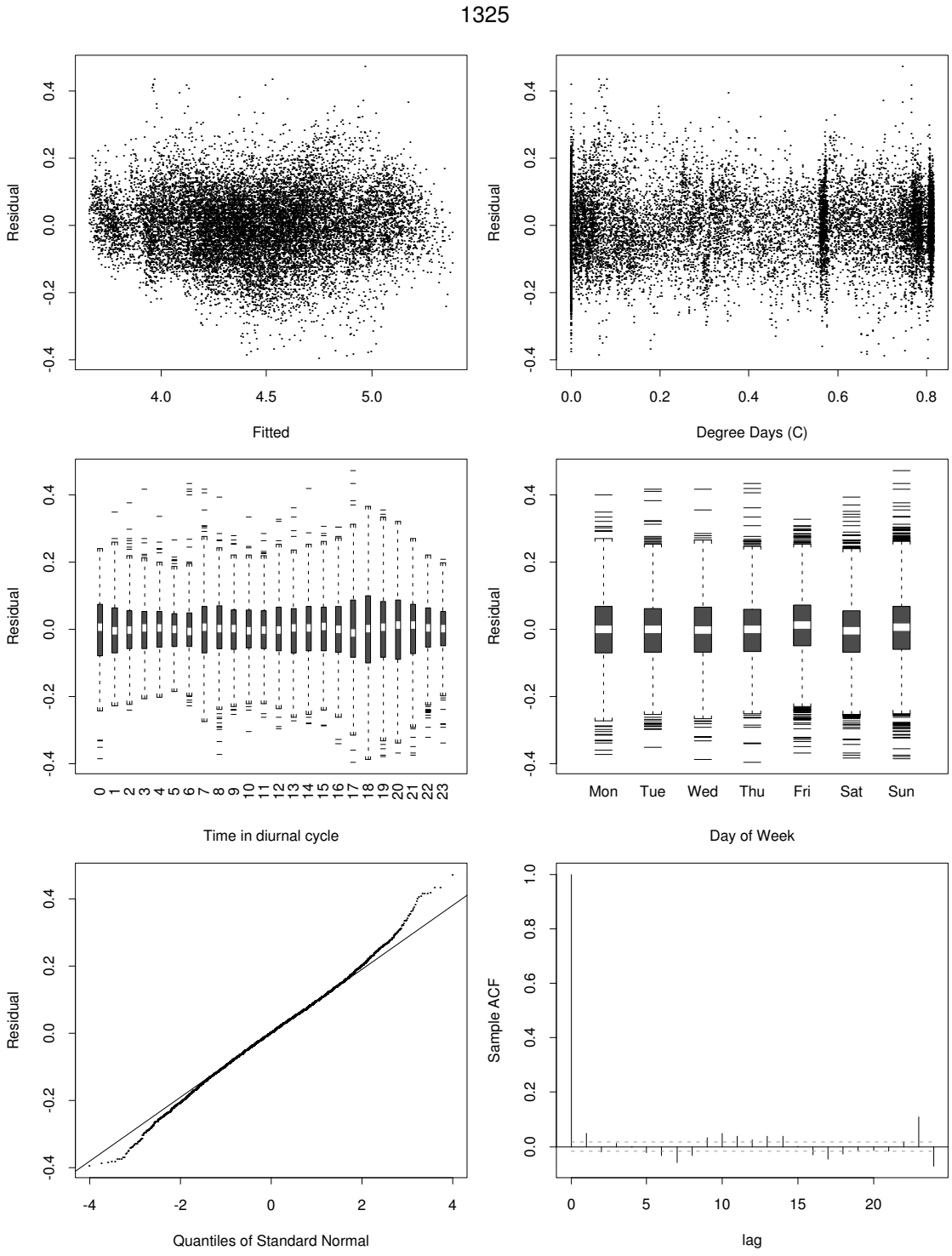

Figure C.5. *Residual diagnostic plots for substation 1325.*

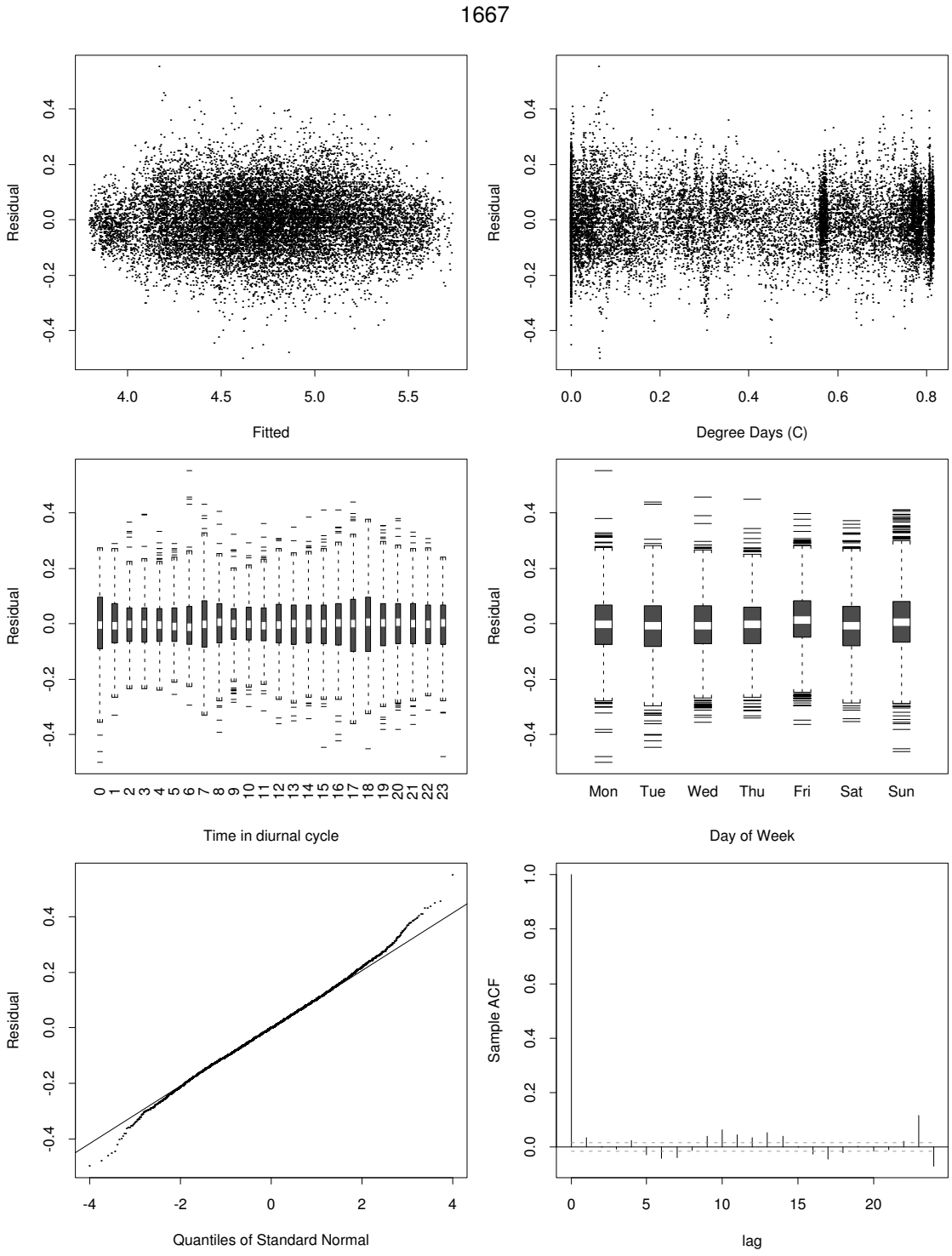

Figure C.6. *Residual diagnostic plots for substation 1667.*

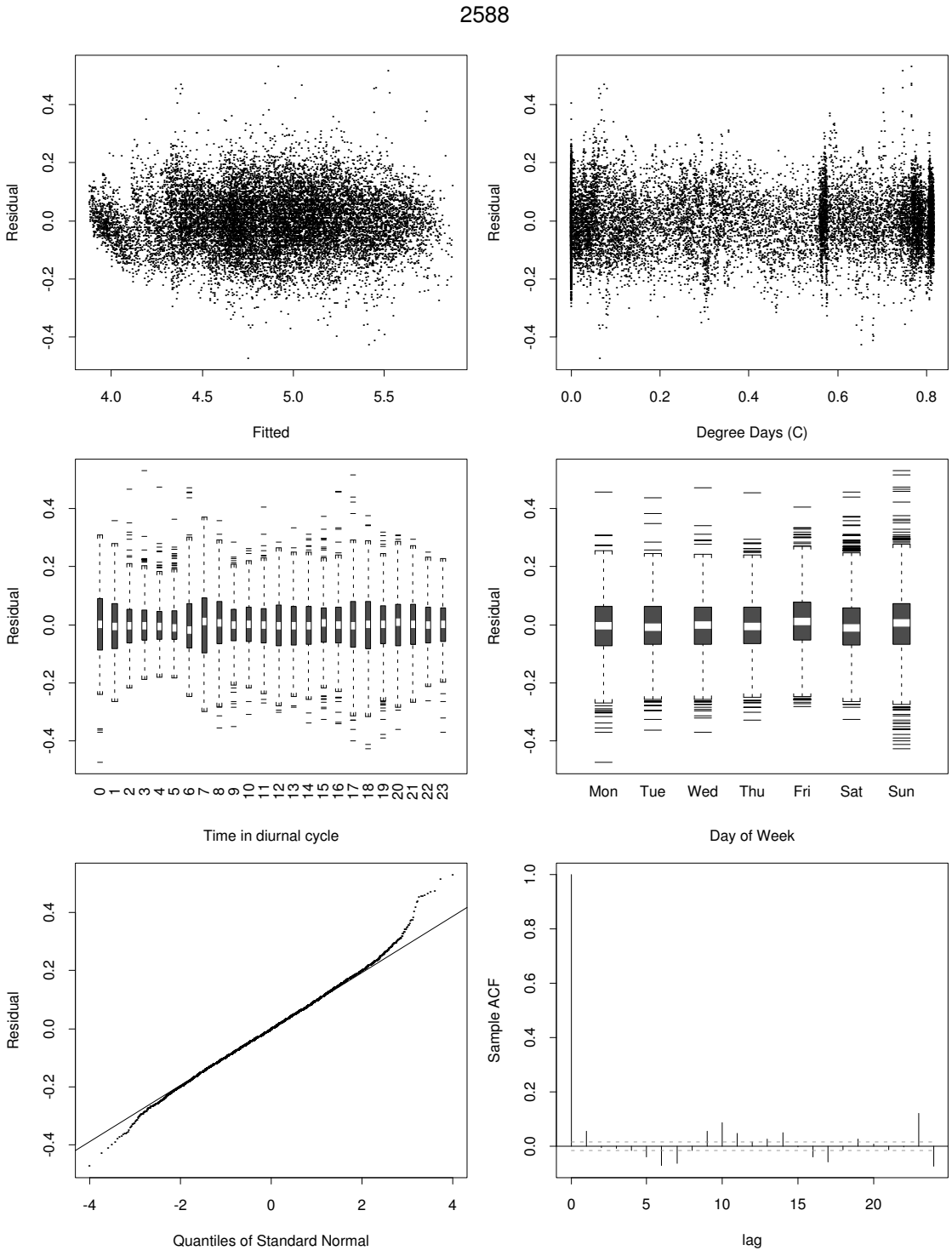

Figure C.7. *Residual diagnostic plots for substation 2588.*

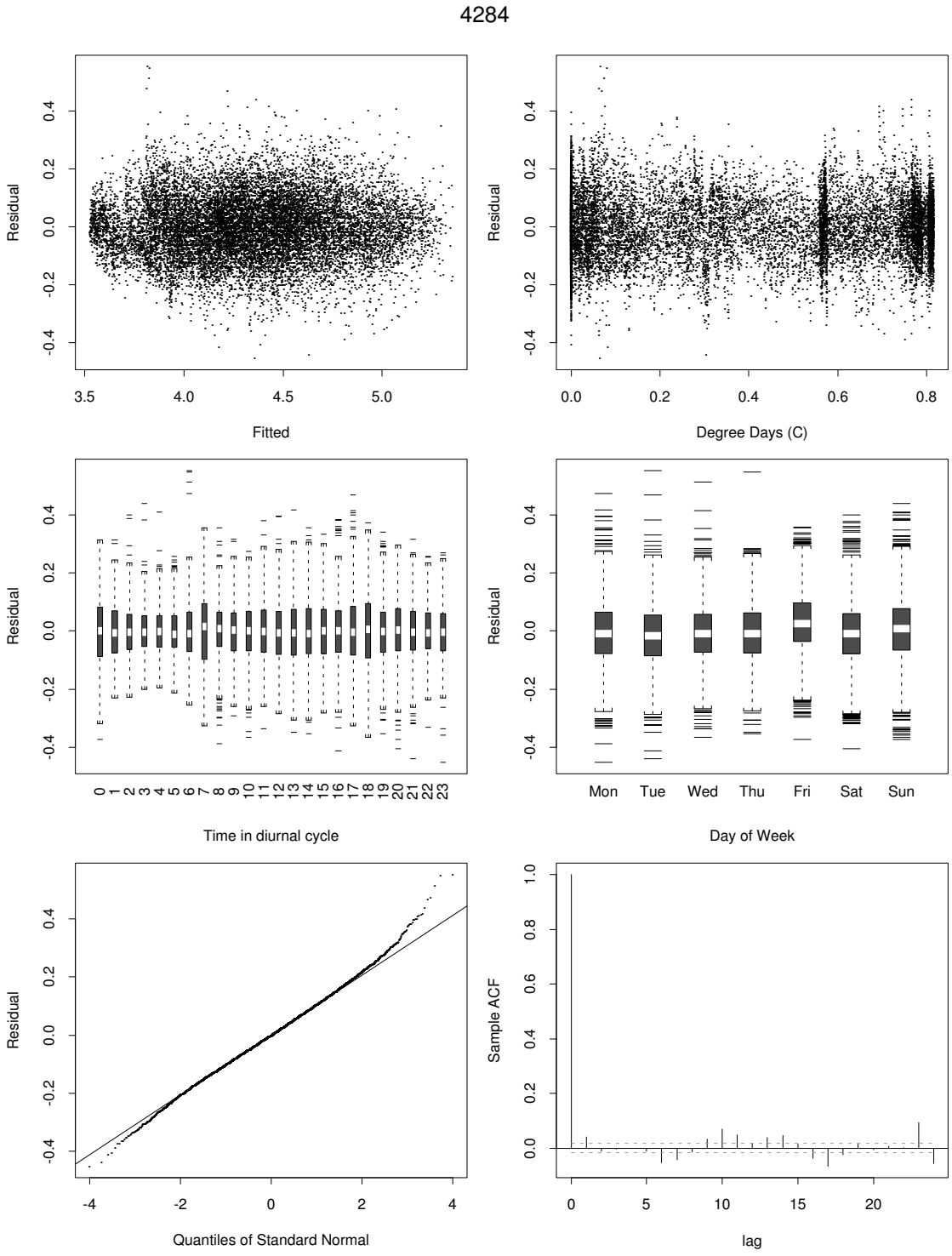

Figure C.8. *Residual diagnostic plots for substation 4284.*

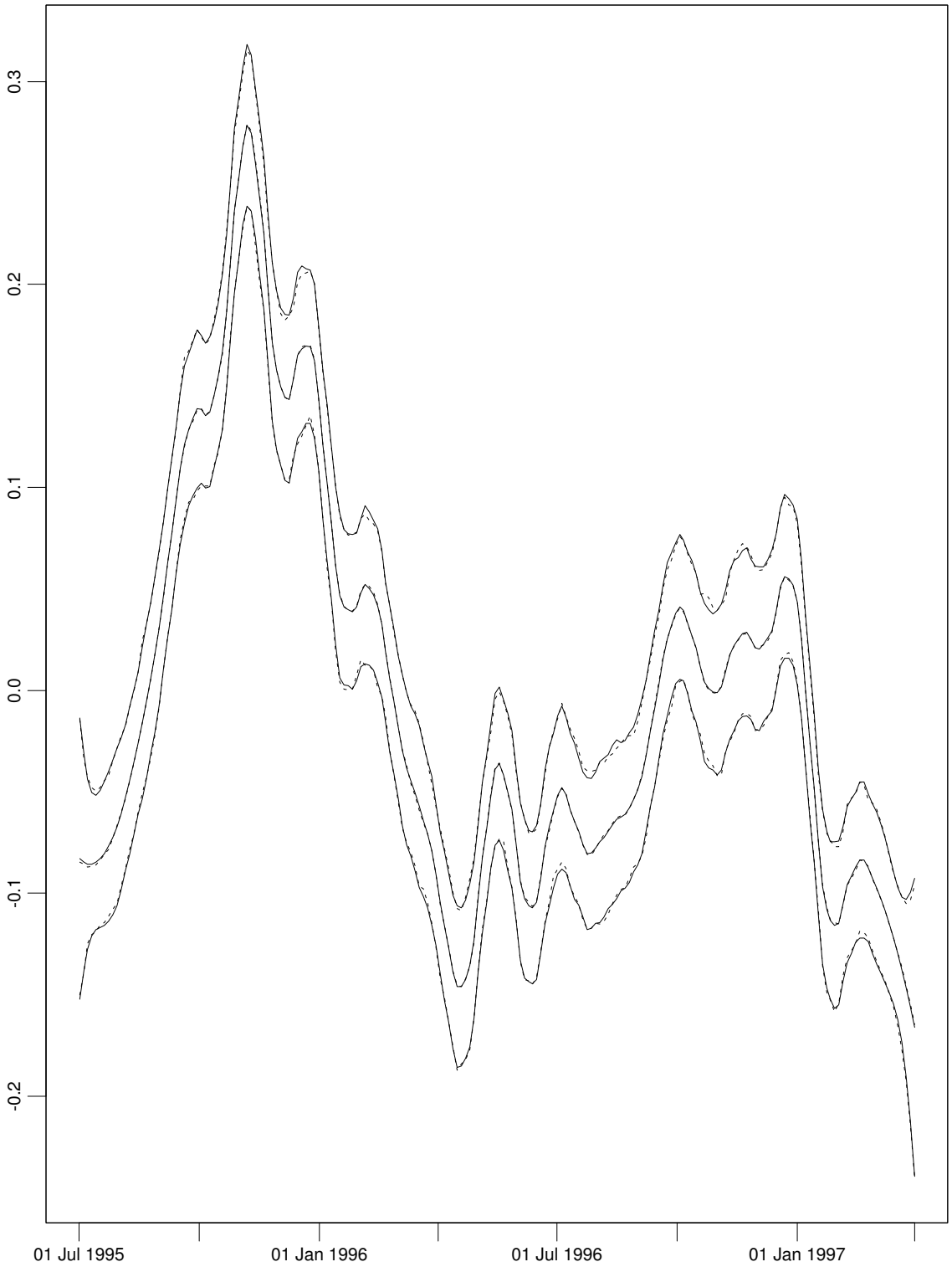

Figure C.9. *Mean (solid) and median (dotted) together with 95% percentile (dotted) and standard normal (solid) intervals of the estimate of the trend component (log-scale) of substation 4284.*

Appendix D

Grouping of Substations

Prior to the startup of the trial the power consumption, during the period September 1995 until January 1996, of 24 substations was analyzed in order to identify substations with similar patterns and sizes of consumption, cf. (Nielsen 1996). An english version of this reference is included in this appendix.

The S-PLUS functions mentioned in Section D.2.1 are described in (Statistical Sciences 1995).

Summary

The power consumption during the period September 1995 until January 1996 for 24 substations is analysed with the purpose of identifying substations with similar pattern and sizes of consumption. The fraction of apartments and the fraction of households using electrical heating are also included in the analysis. It is the purpose to identify four or five groups, and it is sufficient that each group consists of two substations.

For substations with a low fraction of households using electrical heating (below 10%), and without apartments connected, two groups are identified; (i) substations 1667 and 2588, and (ii) 1325 and 4284. For substations with a high fraction of households using electrical heating (above 96%), and without apartments connected, the group 4892 and 5128 is obtained. Finally, for substations with only apartments connected the group 2370 and 2686 is obtained.

D.1 Introduction

This note has been prepared during the planning stage of a trial in which electricity consumers connected to one or more substations are offered devices and/or advise with the purpose of reducing or rearranging their power consumption. For interpretation of the results it is desirable to be able to compare the results from substations which (actively) participate in the trial with results from similar substations where the consumers have not been offered devices and/or advise.

Before such a trial is initiated it is suitable to identify substations with similar pattern and size of consumption. In this note it is described how cluster analysis has been applied in order to identify such groups.

D.2 Approach

Hourly measurements, at the 24 substations, of the power consumption during the period September 1995 until January 1996 (including both) are used, cf. Section D.5. This period is chosen since it (i) agrees with the period during which the trial is expected to be performed, and (ii) outlying observations seem to be absent during the period.

Furthermore the fraction of households using electrical heating and the fraction of apartments are used. Note that apartments with electrical heating installed are not present in the data set.

The substations are first grouped based on the fractions. Hereafter the grouping is performed based on the measurements of power consumption. The final grouping is obtained by requiring that substations in the same group must be grouped together by both methods.

D.2.1 Cluster Analysis

In cluster analysis a number of units are grouped based on quantitative features (actually qualitative features can also be handled, but this is not considered here). First the distance, in the space defined by the features, between the units is calculated. In this case the Euclidean distance is used. The grouping is now obtained by merging the two units closest to each other. These are now regarded as one unit, called a cluster, and the procedure is repeated until all units are grouped into one cluster.

When the distance between a cluster and a unit or a cluster are calculated, one possibility is to calculate the average distance. However, in this note the maximal

distance between units in clusters is used. This method is chosen since it increases the degree of similarity within clusters.

The computer program S-PLUS version 3.2 (functions `dist()` and `hclust()`) is used.

D.3 Results

On basis of the fraction of apartments and fraction of households using electrical heating, both measured in percent, the groups 1–8 shown on Figure D.1 are identified. Note that only groups 1–4 consists of three or more substations.

A plot like the one shown in Figure D.1 is called a cluster tree. The horizontal line segments indicate levels at which units or clusters are combined. E.g. substations 5128 and 5324 are combined into one cluster at level 1% (indicated on the axis to the left of the tree).

In Section D.5 the (i) daily mean, (ii) daily 5% quantile, and (iii) difference between the daily 95% and 5% quantiles (denoted amplitude in the following) are plotted against the number of days since September 1, 1995. Based on these time series the substations have been characterized by:

- The overall mean,
- the slope of the daily mean,
- the mean of the daily 5% quantile,
- the slope of the daily 5% quantile,
- and the mean of daily amplitude.

The summary statistics are calculated based on data until 20 December 1995, since this resulted in well defined slopes. Note that the individual time series are not standardized. For this reason the analysis will tend to group substations with similar sizes of consumption, which is appropriate for the concept of the trial outlined in Section D.1.

The summary statistics (*not* the time series *but* the overall mean, etc.) are standardized to zero mean and unit standard deviation. The result of the following cluster analysis is displayed in Figure D.2. Also, in the figure the grouping obtained based on the fraction of apartments and fraction of households using electrical heating is indicated.

If a maximal distance of 2.0 is used, four groups of substations are obtained. For substations where a low fraction (10% or less) of the households use electrical heating

and with no apartments the groups (i) 1667 and 2588, and (ii) 1325 and 4284 are identified. For substations where a high fraction (96% or more) of the households use electrical heating and with no apartments the group 4892 and 5128 are identified. Finally, for substations with only apartments connected the group 2370 and 2686 are identified.

From Figure D.2 it is revealed that, if the campaign is to be executed at two substations only, the group where a high fraction of households use electrical heating (4892, 5128), and the group consisting of apartments (2370, 2686), should be omitted. The remaining groups are obtained using a maximal distance below 1.

If the groups identified are compared with the time series plots in Appendix D.5 it is revealed that the substations selected are very similar within the groups. This is especially true for the groups (1325, 4284) and (1667, 2588).

D.4 Conclusion

Using Cluster Analysis four pairs of similar substations are identified among 24 substations. If only two pairs are needed the similarity is markedly higher.

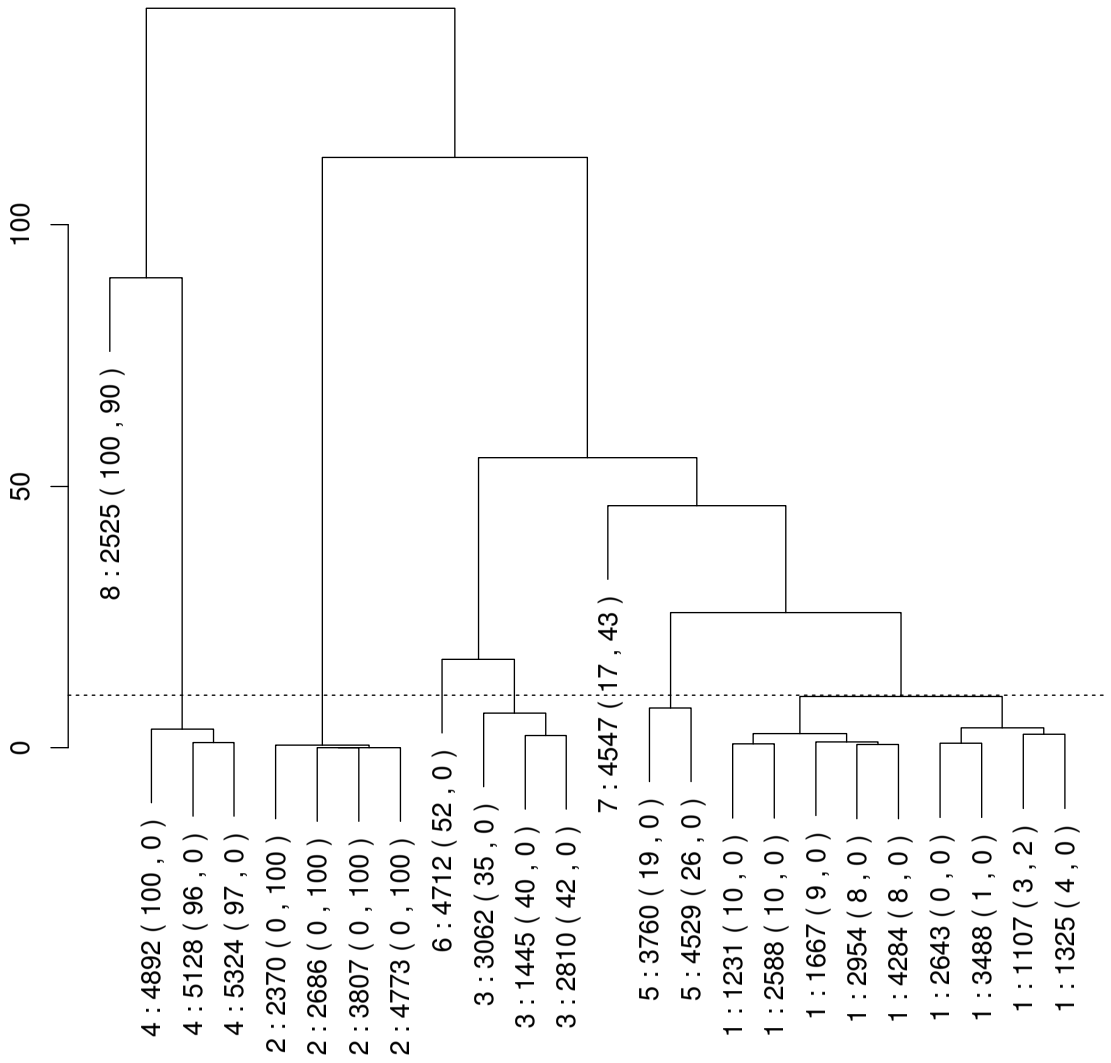

Figure D.1. *Cluster tree based on the fraction of households using electrical heating and the fraction of apartments. The distance 10% is indicated by a dotted line. The text on the tree indicates:* `Group : substation (% electrical heating, % apartments)`.

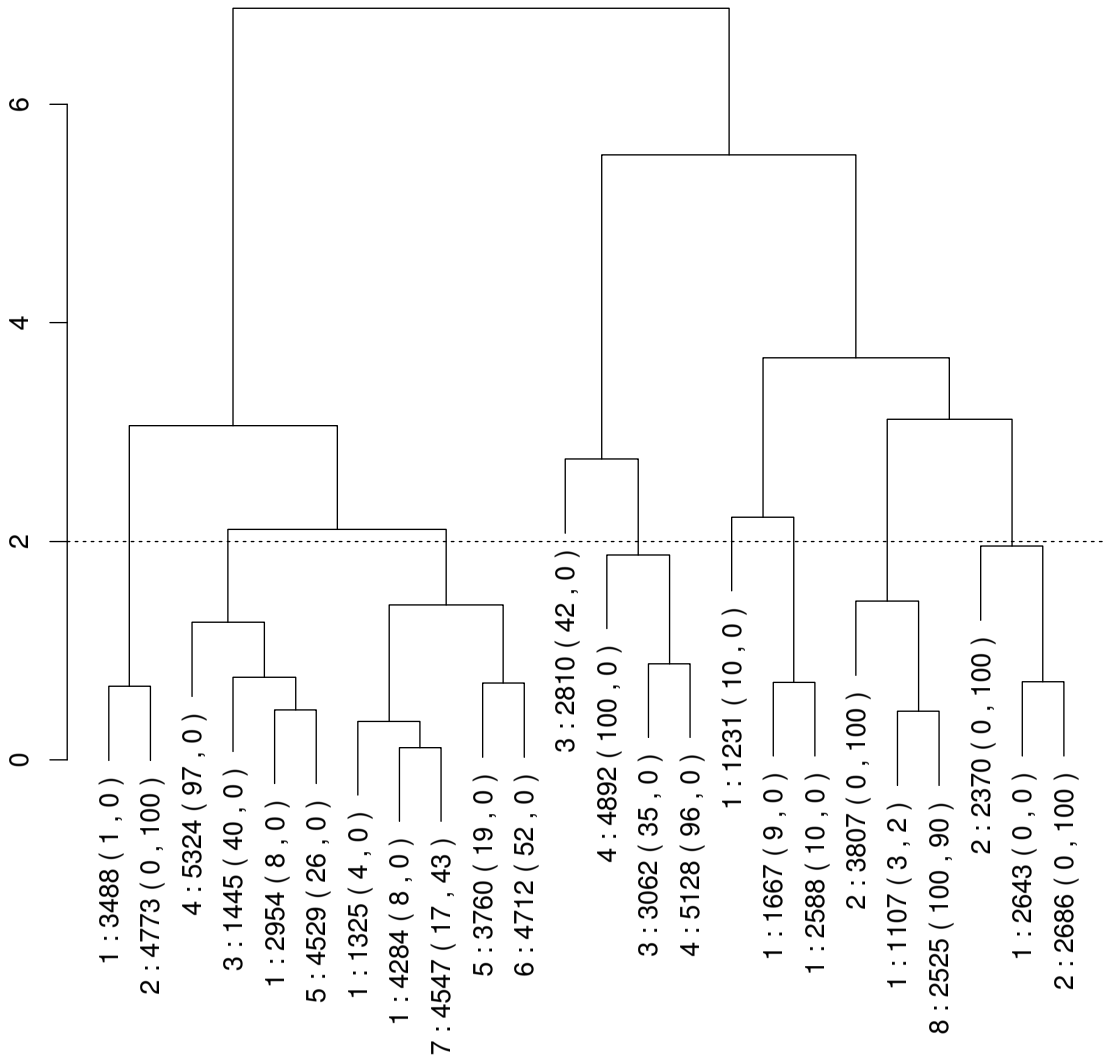

Figure D.2. *Cluster tree based on standardized summary statistics of electrical power consumption data. The distance 2 is indicated by a dotted line. The text on the tree indicates:* `Group from Figure D.1 : substation (% electrical heating, % apartments).`

D.5 Data

The primary data is hourly measurements of the power consumption (kWh) for 24 substations measured during the period September 1995 until January 1996 (including both). On the following pages

- the daily mean,
- the daily 5% quantile,
- and the difference between the daily 95% and 5% quantiles

are plotted against the number of days since 1 September 1995. The substation identification numbers are placed over each sub-plot.

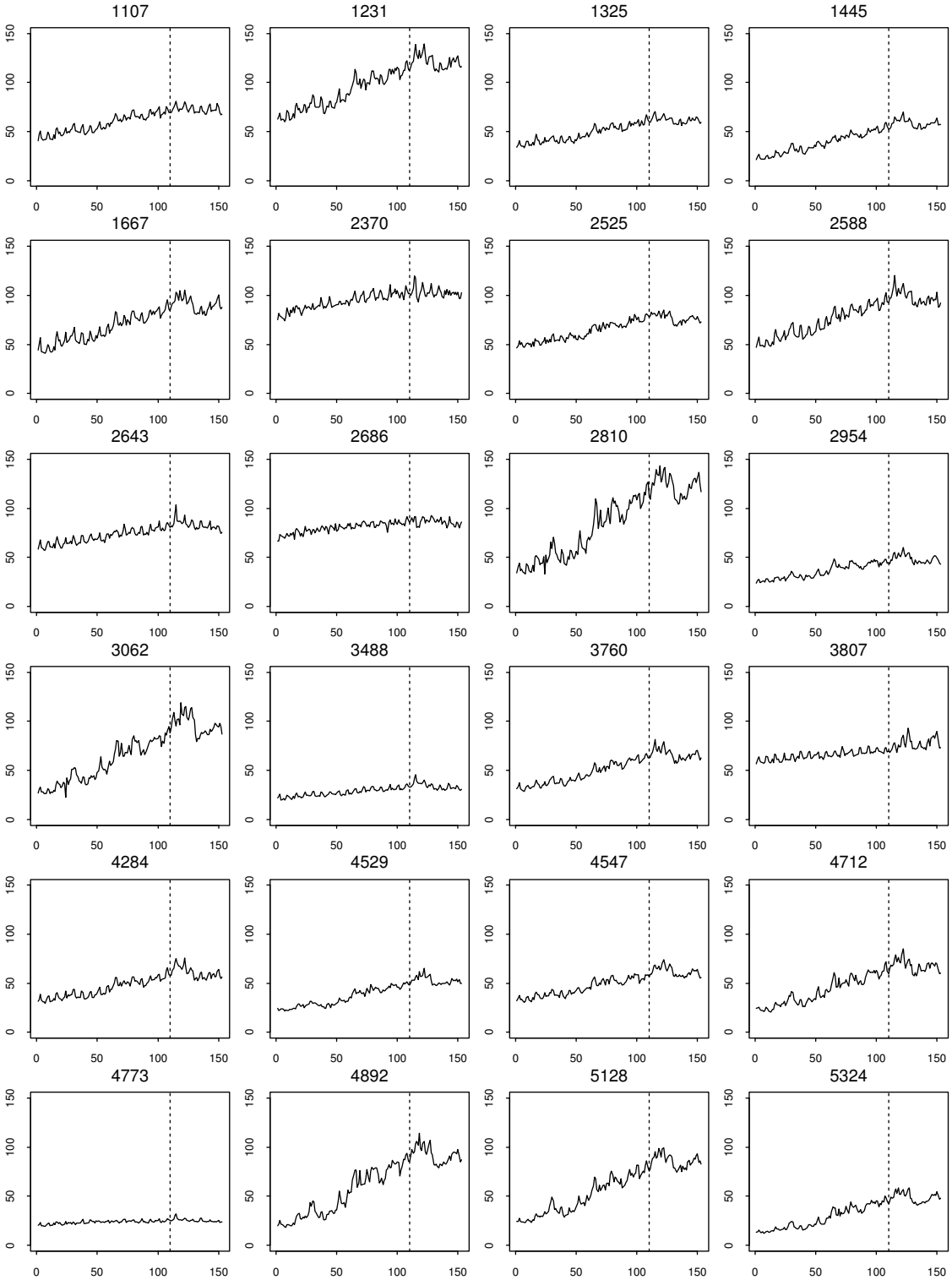

Figure D.3. *Daily means against the number of days since 1 September 1995. 20 December 1995 is indicated by a dotted line.*

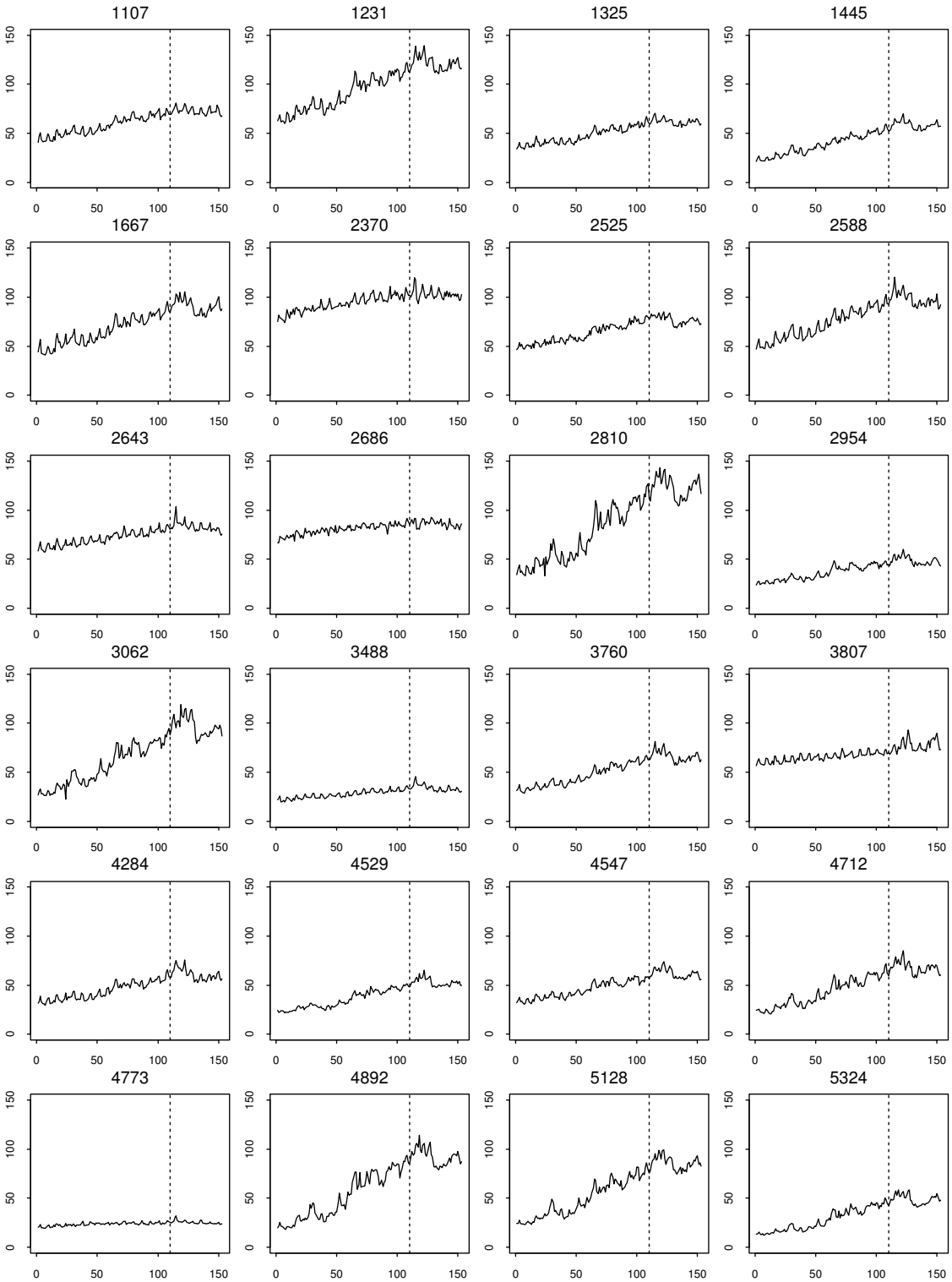

Figure D.4. *Daily 5% quantiles against the number of days since 1 September 1995. 20 December 1995 is indicated by a dotted line.*

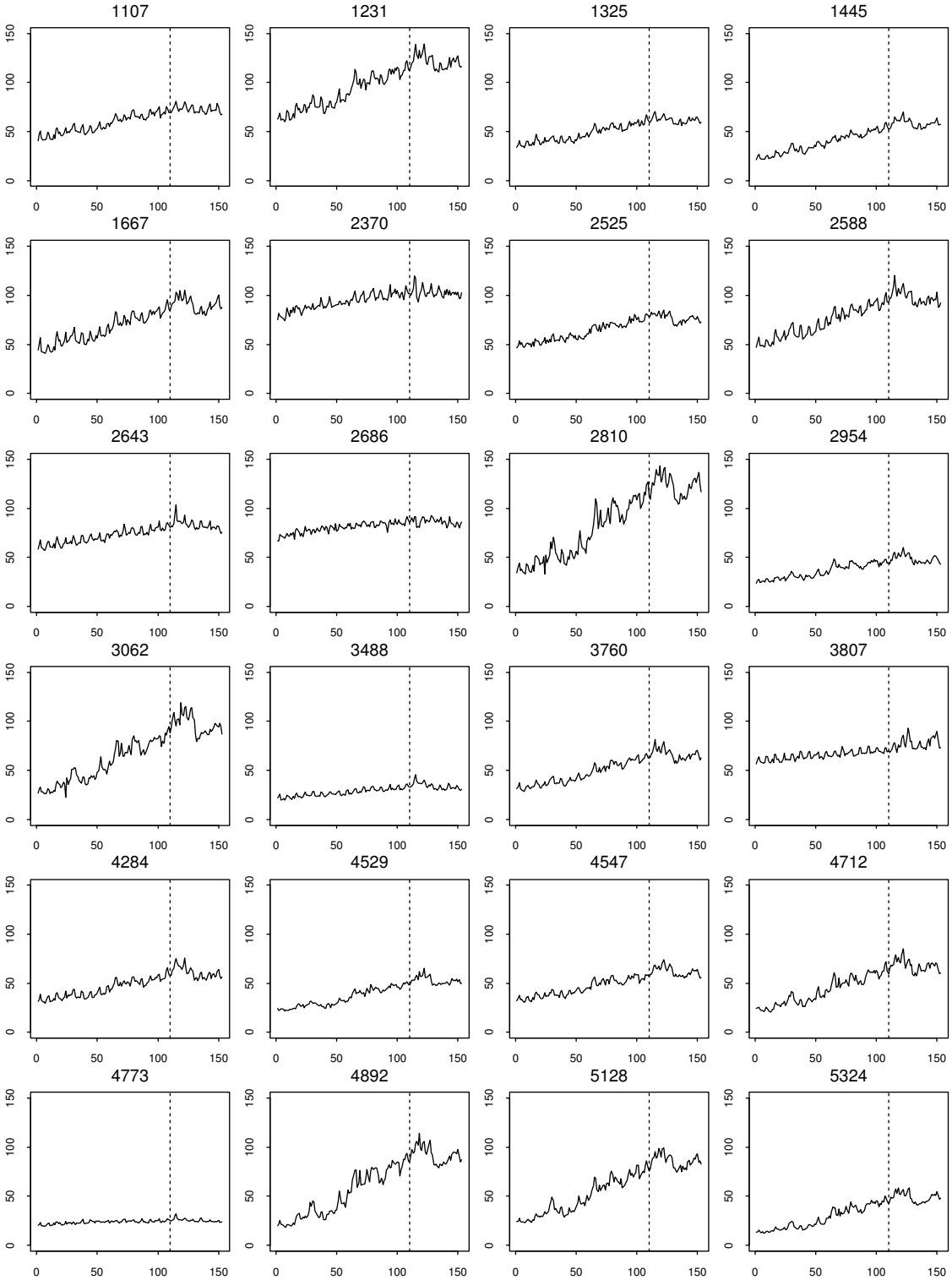

Figure D.5. *Difference between daily 95% and 5% quantiles against the number of days since 1 September 1995. 20 December 1995 is indicated by a dotted line.*

Bibliography

Chambers, J. M. & Hastie, T. J., eds (1991), *Statistical Models in S*, Wadsworth, Belmont, CA.

Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. (1990), 'STL: A seasonal-trend decomposition procedure based on loess', *Journal of Official Statistics* **6**, 3–33. (C/R: p33-73; Corr. to Comment: p343-348).

Cleveland, W. S. & Devlin, S. J. (1988), 'Locally weighted regression: An approach to regression analysis by local fitting', *Journal of the American Statistical Association* **83**, 596–610.

Cleveland, W. S., Devlin, S. J. & Grosse, E. (1988), 'Regression by local fitting: Methods, properties, and computational algorithms', *Journal of Econometrics* **37**, 87–114.

Conradsen, K. (1984), 'Introduction to statistics, 2a & 2b (in danish: En introduktion til statistik, 2a & 2b), 4th edition', Institute of Mathematical Statistics and Operations Research, Technical University of Denmark.

Craven, P. & Wahba, G. (1979), 'Smoothing noisy data with spline functions', *Numerische Mathematik* **31**, 377–403.

Devore, J. L. (1991), *Probability and Statistics for Engineering and the Sciences*, third edn, Brooks/Cole, Monterey, CA.

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, London/New York.

Friedman, J. H. (1984), A variable span smoother, Technical Report 5, Laboratory for Computational Statistics, Department of Statistics, Stanford University, California.

Hastie, T. J. & Tibshirani, R. J. (1990), *Generalized Additive Models*, Chapman & Hall, London/New York.

Jenkins, G. M. & Alavi, A. S. (1981), 'Some aspects of modelling and forecasting multivariate time series', *Journal of Time Series Analysis* **2**, 1–47.

Madsen, H. (1995), 'Time series analysis (in danish: Tidsrækkeanalyse), 2nd edition', Department of Mathematical Modelling, Technical University of Denmark.

Mallows, C. L. (1973), 'Some comments on C_p', *Technometrics* **15**, 661–675.

Nielsen, H. A. (1996), 'Note on grouping substations prior to execution of a trial (in danish: Notat om gruppering af transformerudføringer med henblik på udførelse af forsøg)', Department of Mathematical Modelling, Technical University of Denmark.

Nielsen, H. A. (1997), LFLM 1.0, an S-PLUS / R library for locally weighted fitting of linear models, Technical Report 22, Department of Mathematical Modelling, Technical University of Denmark.

SAS Institute Inc. (1993), *SAS/ETS User's Guide, Version 6*, second edn, SAS Institute Inc., Cary, NC.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**, 461–464.

Shumway, R. H. (1988), *Applied Statistical Time Series Analysis*, Prentice-Hall, Englewood Cliffs, NJ.

Sørensen, M. S. (1997), 'Development of methods for evaluation of electricity saving and load levelling measures, part 2: The planning and implementation of a power conservation campaign', NESA A/S, Hellerup, Denmark.

Statistical Sciences (1993), *S-PLUS User's Manual, Version 3.2*, StatSci, a division of MathSoft, Inc., Seattle.

Statistical Sciences (1995), *S-PLUS Guide to Statistical & Mathematical Analysis, Version 3.3*, StatSci, a division of MathSoft, Inc., Seattle.

Tong, H. (1990), *Non-linear Time Series. A Dynamical System Approach*, Oxford University Press, Oxford.