Aggregated Power Consumption Models for the Eastern Part of Denmark

Henrik Aalborg Nielsen & Henrik Madsen

14 January 1997

IMM

Time Series Group
Department of Mathematical Modelling
Technical University of Denmark

This document was prepared with LaTeX and printed by NESA A/S, Hellerup.

Acknowledgements

First of all the authors whish 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 whish to thank NESA A/S, Hellerup, Elkraft A.m.b.A., Ballerup and Department of Agricultural Sciences, The Royal Vet. and Agric. Univ., Copenhagen for suppling data.

A special thank to Conny Thorsted at NESA for collecting data, providing information, answering all our questions about different aspects of power production, comments and suggestions about this report, and for managing this project. We would also like to thank the other members of the project team Willy Bergstrøm (NESA), Johan Hakmann (KB), Jan Møller (DEFU), Preben Munter (SEAS), Henning Parbo (ELSAM), and Jens Rubæk (NVE) for comments and fruitful discussions. Also, thank to Keld Oksbjerg (Elkraft) for discussions regarding the realtion between economic variables and power consumption.

Summary

In this report models for the hourly power consumption in the Eastern part of Denmark are developed. The development is based on empirical methods. Data covering the years from 1974 to 1994 are used. The models are intended for long term predictions and scenario studies.

In order to make the models well suited for long term predictions no models containing autoregressive terms in the power consumption are used. Instead the variations are modelled using the time in the diurnal and annual cycle, the type of day and the climate. The climate is expressed as air temperature, wind speed and global radiation measured at Taastrup. The models are estimated separately for each year in order to account for slow changes in the system.

Local regression techniques have been used to identify a functional relationship between the climate and the power consumption in the setting of static models. These results has then been used to formulate static parametric models. Furthermore models describing the dynamic response on the global radiation and the air temperature have been established. For all parametric models the diurnal fluctuation varies over the year. The residuals of the models are not white noise and hence it should be possible to refine them further.

The parametric models have values of the squared degree of determination (R^2 values) of approximately 0.98 or 0.99. The largest residuals are observed during Christmas and the modelling of this period is not satisfactory. The dynamic effect of the most important climate variables is investigated, and it is shown that the dynamic response from the global radiation is much faster that the dynamic response from the air temperature. In particular the step responses from global radiation stabilize within five hours, whereas the step responses from air temperature seems to stabilize within three days. The range of the diurnal fluctuation is approximately 1000 MWh/h and for the annual fluctuation the corresponding number is approximately 400 MWh/h. However, the influence of temperature adds approximately further 500 MWh/h to the annual fluctuation.

An approach to scenario analysis is suggested. For this purpose the estimates obtained for the individual years are investigated for certain internal relationships and temporal development. Furthermore relationships to external variables measured on annual basis are investigated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This report forms a 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.

In this report models for hourly observations of the power consumption in the Eastern part of Denmark are developed. The development is based on data covering the period from 1974 until 1994, and is largely empirical. However, prior knowledge and reasonable assumptions are used in the model building.

The models are intended for long term predictions and scenario studies, where the climate is assumed known. For this reason the models contains no autoregressive part, since the temporal correlation will yield good approximations only for short term predictions. However, some models includes a dynamic response on the climate variables. Due to slow changes during the period (see Chapter 2) it has not been possible to find reasonable estimates for the entire period. For this reason most of the investigations are performed on a yearly basis.

Besides the Summary and Introduction all readers should read Chapter 2, in which the data are described, and probably also the Conclusion and the Discussion. Readers who are mainly interested in the results obtained should further read Section 7.7 in which the final model is presented and Chapter 8 in which the results of the model are investigated. Furthermore, Chapter 9, in which an approach to scenario studies is presented, may be of interest. Readers interested in understanding the rationale of the final model should read the entire report.

Chapter 2

Data

2.1 Hourly Observations of Power and Climate

The electricity power data consists of hourly observations of the power consumption for the Eastern part of Denmark (Sjælland, Møn, Falster and Lolland) for the years 1974–1994. The data is provided by Elkraft A.m.b.A., Ballerup. The unit of measurement is MWh.

More precisely the power consumption is defined as the production of electricity at public primary power stations at the injection point, i.e. including distribution loss but not including the production at public secondary power stations[1] (windmills and small local production units). The power consumption is representing national usage only, i.e. the values are corrected for import/export. Few, hourly data on the private power production are available. It is however known that there has been a growth of the private production during the period (wind mills, gas engines, etc.). According to the Association of Danish Electric Utilities (Danske Elværkers Forening) the sum of the production at public secondary power stations and of the private production was 1234.2 kWh in 1994. The power consumption as defined above was 12313 MWh. Hence, the non-registered production is a rather small fraction of the consumption.

To stress that the power consumption may be interpreted either as energy consumed or the average power (in the sense of physical science) over the preceding hour the unit MWh/h is used throughout this report.

The climate data consists of observations of hourly averages of ambient air temperature ($^{\circ}C$), wind speed (m/s) and global radiation (W/m^2) for the years 1974–1994. The data is measured at Taastrup by the Department of Agricultural Sciences, The Royal Vet. and Agric. Univ., Copenhagen. Observations of the global radiation below 3 W/m^2 are set equal to zero.

[1] This number will be called the "power consumption" throughout this report.

The data was corrected for some outliers and other errors (the value for the particular observation is set to missing). Only the climate data contains missing values and nearly all occurs after 1985. The pattern of missing values are nearly identical for all three climate variables. Although up to 80% of data are missing for a single month the most typical fraction of missing values is below 20% and for the majority of months no missing values are found, see Figure 2.1 below. The fraction of missing values per month were compared with the mean and standard deviation of the non-missing data by means of graphical methods. No dependencies were revealed. Plots the corrected data are shown in Appendix C, together with plots only covering part of the period.

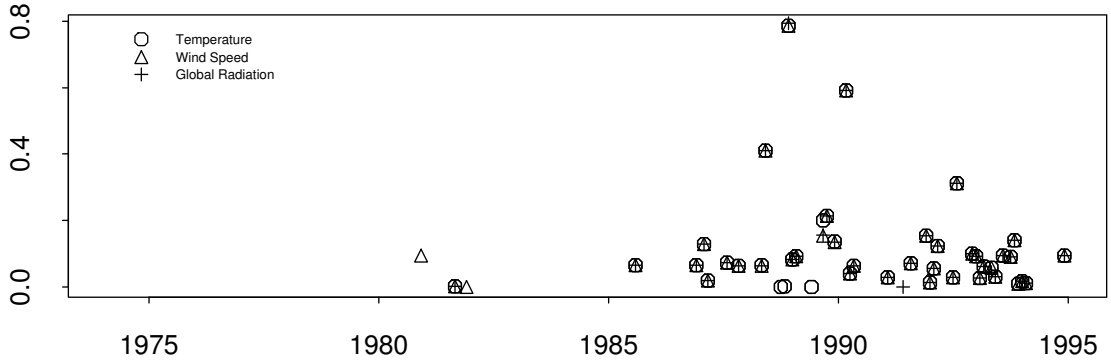


Figure 2.1. *Fraction of missing values per month.*

In this report the following notation will be used:

- $P(t)$: Power consumption during the hour $t-1$ to t.
- $T_a(t)$: Average ambient air temperature during the hour $t-1$ to t.
- $W(t)$: Average wind speed during the hour $t-1$ to t.
- $R_g(t)$: Average global radiation during the hour $t-1$ to t.

For each observation the day of week and the type of day are registered. The type of day is one of "normal working day", "Saturday or Half-Holy day", or "Sunday or Holy day".

2.2 Yearly Observations

The trend and other main characteristics of the power consumption data may be related to economic and/or population data. Since the private production of power is not accounted for in the data, the wind dependence will probably be related to the installed capacity on private wind mills. Furthermore the production on a few

non-private wind mills is not accounted for in the power consumption data. These potential explanatory variables are described in this section.

In Table 2.1 and 2.2 the installed wind power capacity and some economic and population variables are displayed, respectively. All variables are related to the Eastern part of Denmark only. The Gross Domestic Product at factor cost (GPD_{fc}) is measured by use of two sightly different methods. For a period of four years both methods were used. These years have been used to calculate approximate values of method 1 for 1992-1994. This was done by fitting a strait line with series 1 as the dependent variable.

The index of campaign intensity is based on figures from NESA's annual account. Direct costs of wages plus an addition of 70 per cent for employees who have a permanent tenure with the company and costs of materials for the departments in question is included. Exact information from NESA's annual account exist back to 1980. For the period 1974 – 1979 the index have been based on data stored by employees who have been with the company for many years. The index have been reflated using an implied price index based on the gross domestic product at factor cost.

Year	≤ 82	83	84	85	86	87	88	89	90	91	92	93	94
Private	0.0	4.9	6.0	8.6	13.6	19.3	32.9	45.1	54.4	56.3	60.4	64.8	70.3
Other	0.0	0.0	0.0	0.0	3.8	3.0	14.3	18.9	31.3	41.7	49.5	57.0	68.4
Total	0.0	4.9	6.0	8.6	17.4	22.3	47.2	64.0	85.7	98.0	109.9	121.8	138.7

Table 2.1. *Installed capacity (MW) of wind mills not accounted for in the power consumption data. Source of data: (DEF, 1995).*

2.3 Some Political and Economical Events

In this section some political and economical events from 1974 to 1994, which may have influenced the power consumption, is listed. The source of data is NESA A/S, Hellerup.

General political and economical events:

- 21 December 1979: Passage of a number of laws to strengthen the balance of payments for Denmark.
- 16 October 1982: New government and changes of the economic politics.
- 17 October 1986: Passage of a number of laws which tighten the private economy.
- 11 December 1991: EEC summit meeting in Maastricht.

Year	GPD_{fc} 1 (DKr)	GPD_{fc} 2 (DKr)	GPD_{fc} (DKr)	TFP (-)	Price (DKr/kWh)	C. idx. (-)	Population (No.)	Households (No.)
1974	138189	–	138189	95.9	0.4196	0.93	2288274	–
1975	135711	–	135711	96.7	0.4258	0.94	2290761	–
1976	141547	–	141547	98.2	0.3658	0.95	2288551	–
1977	143537	–	143537	99.3	0.3326	0.96	2286424	–
1978	145410	–	145410	100.1	0.3434	0.98	2286586	–
1979	150722	–	150722	101.6	0.3294	0.99	2265694	–
1980	152874	–	152874	100.0	0.3976	1.00	2283454	973862
1981	152066	–	152066	100.9	0.4824	1.01	2278612	980299
1982	156068	–	156068	102.0	0.4992	1.03	2272669	984066
1983	158460	–	158460	102.7	0.4351	1.04	2268036	991821
1984	163455	–	163455	104.3	0.3655	1.05	2262596	999403
1985	168897	–	168897	105.3	0.3670	1.06	2258644	1007790
1986	175256	–	175256	106.0	0.3581	1.08	2256788	1015645
1987	177595	–	177595	106.4	0.3244	1.09	2256514	1022869
1988	181895	181626	181895	107.3	0.3539	1.10	2254659	1029005
1989	182439	181428	182439	108.0	0.3768	1.11	2252007	1036436
1990	185682	184003	185682	108.9	0.3766	1.13	2251807	1042610
1991	186223	184422	186223	109.1	0.3565	1.14	2255054	1049943
1992	–	182872	184063	107.9	0.3441	1.15	2260890	1058232
1993	–	183774	185323	108.6	0.3278	1.16	2268081	1064032
1994	–	188951	192556	111.0	0.3312	1.18	2275090	1069143

Table 2.2. *Economic, power and population related variables for the years 1974–1994. Gross Domestic Product at factor cost (GPD_{fc}), Index of Total Factor Productivity – real production (TFP), a weighted overall price on electricity in fixed prices (Price), index of campaign intensity (C. idx.), Population size, No. of households. Source of data: NESA A/S, Hellerup.*

Political events related to energy consumption:

- 6 November 1973: A number of restrictions in the oil consumption were introduced due to the oil crisis. The restrictions were abolished on the 12 February 1974.
- In 1976 it was possible for the first time to obtain economic support from the government to implement energy conserving arrangements. Also in 1976 the second energy crisis took place, but it mainly affected the oil and transport sectors.
- In 1977 a separate tax on power was implemented.
- 20 January 1989: The tax on a number of power consuming goods were abolished or decreased significantly.
- 30 May 1989: The tax on low energy bulbs were abolished.

Chapter 3

Statistical Methods

3.1 Local Regression

In in this report local regression is used to estimate non-linear relationships for which the exact parametric form are unknown. For instance, since heating by power is used to some extent in Denmark, it is clear that the air temperature affects the power consumption. However, it is not known whether the relationship can be well approximated by e.g. a straight line.

For these kind of analyses local regression has been applied. Classical local regression is not directly applicable since one would typically like to estimate the response on air temperature while adjusting for the time of day at the same time. For this reason conditional parametric models have been used. A program for estimation in such models has been developed. Since classical local regression forms the basis of the method, it is briefly described below.

Classical Local Regression

The model considered in classical local regression is

$$y_i = f(x_i) + e_i; \quad i = 1, \ldots, n, \tag{3.1}$$

where x_i is the independent variables, y_i is the dependent variable, n is the No. of observations and f is an unknown, but smooth, function. Furthermore it is assumed that e_i is independently identical distributed (i.i.d.).

The purpose of local regression is to estimate the value of $f(x)$ for any x within the range defined by the observations. In practice $\hat{f}(x)$ is calculated in a number of points and interpolation formulas are applied when other points are considered. Often these interpolation formulas are quite advanced. In order to use local regression

it is important that locally $f(x)$ can be well approximated by a parametric function. Motivated by Taylors formula polynomials are often used.

The degree of locality is determined by a coefficient called the bandwidth (h). If an estimate of the unknown function f is wanted in the point u_j, and a 2nd order polynomial g_j is used as a local approximation, then

$$g_j(x_i) = a_{j0} + a_{j1}(x_i - u_j) + a_{j2}(x_i - u_j)^2; \quad i = 1, \ldots, n \tag{3.2}$$

is fitted with larger weight on observations close to u_j than to those further away. The weight, w_{ji}, on the observation x_i when estimating the function at the point u_j is

$$w_{ji} = W(||x_i - u_j||_2/h_j), \tag{3.3}$$

where $W : \mathcal{R}_0 \to \mathcal{R}_0$ is a nowhere increasing function, h_j is the (scalar) bandwidth used for u_j and $||x_i - u_j||_2$ is the Euclidean distance between x_i and u_j.

In this study

$$W(u) = \begin{cases} (1 - u^3)^3, & u \in [0; 1) \\ 0, & u \in [1; \infty) \end{cases}, \tag{3.4}$$

is used – note that $W(u)$ is positive only on a bounded interval. The estimate is then $\hat{f}(u_j) = \hat{g}_j(u_j) = \hat{a}_{j0}$. If the bandwidth is constant the method is denoted "local regression with fixed bandwidth". Often the bandwidth at the point u_j is chosen such that a certain fraction of the observations fulfill the criterion $||x_i - u_j||_2 \leq h_j$. This is called a nearest neighbour bandwidth.

Together the bandwidth and the degree of the local polynomial form a model of the observations. If a nearest neighbour bandwidth is used the location of the x_i's is also involved in the resulting model.

Classical local regression is described in e.g. (Cleveland and Devlin, 1988), which contains many other references. Also in (Cleveland and Devlin, 1988) it is shown how to test hypothesis and calculate pointwise confidence intervals.

Conditional Parametric Models

Conditional parametric models are just ordinary regression models which fit locally to some other regressors. If these last mentioned regressors are fixed, the model is globally parametric, hereof the name. In this case models of the form

$$y_i = f(x_i) + z_i^T \theta(x_i) + e_i; \quad i = 1, \ldots, n \tag{3.5}$$

are considered. f, y_i, x_i, e_i and n is defined as above, but now the independent variables z_i and the multidimensional function θ are added. Equation (6.1) in Section 6.1 is an example of such a model.

By regarding all the independent variables in (3.5) as x_i in (3.1) it is seen that (3.1) actually contains (3.5). The latest model is more restrictive and hence assumes more prior knowledge about the system. The benefit is that if a good approximation is used, then the model is able to give more precise results, especially when the combined dimension of x_i and z_i is large.

In order to estimate the functions f and θ at the point u_j the parameters of

$$Y = Z\beta_j, \tag{3.6}$$

are estimated by Weighted Least Squares (WLS). β_j is the parameters, Y contains the observations of the dependent variables and the rows of Z is formed by the independent variables x_j and z_j. The weights on the observations are defined by x_i and found as described in the section above.

In this case θ is always estimated by local constants (kernel estimates), whereas the degree of the polynomial locally approximating f has to be chosen. If x has dimension one and a local 2nd order polynomial is used for f, then row i of Z becomes $[1\ x_i\ x_i^2\ z_i^T]$. With $\beta_j = [\beta_{j0}\ \beta_{j1}\ \beta_{j2}\ \ldots]$ the estimate of $f(u_j)$ is $\hat{\beta}_{j0} + \hat{\beta}_{j1}u_j + \hat{\beta}_{j2}u_j^2$.

The WLS problem is solved by the algorithm described in (Miller, 1992). The algorithm was originally written in Fortran but a port to C by A. Shah is used in this study. This may be obtained as `pub/C-numanal/as274_fc.tar.z` by anonymous ftp from `usc.edu`.

3.2 Non-linear Regression Models

Based on the methods described in Section 3.1 structured parametric models of the investigated relationships are identified. These parametric models are often non-linear in the parameters. The parametric models used in this report belongs to the class of non-linear regression models. Some fundamental theory of these models is outlined below. For further reading see e.g. (Gallant, 1987). The estimation is performed using `proc model` of SAS Version 6, see (SAS Institute Inc., 1993).

In the non-linear regression model the expected value of the dependent variable is a function of the independent variables. Except for the value of the parameters the function is known. This may be represented as

$$y_i = f(x_i, \theta) + e_i; \quad i = 1, \ldots, n, \tag{3.7}$$

where x_i is the independent variables, y_i is the dependent variable, n is the No. of observations, θ is a p dimensional vector containing the unknown parameters, f is a known function, and e_i is i.i.d. with zero mean and variance σ^2.

Least Squares Estimates

If the parameters θ are estimated by Ordinary Least Squares (OLS or just LS) the estimates $\hat{\theta}$ are chosen to minimize

$$V(\hat{\theta}) = \sum_{i=1}^{n}(y_i - f(x_i, \hat{\theta}))^2. \tag{3.8}$$

Let $F = F(\theta)$ be a matrix with rows $(\partial/\partial\theta^T)f(x_i, \theta)$. According to (Gallant, 1987) then for large n

$$\hat{\theta} \sim N(\theta, \sigma^2(F^T F)^{-1}), \tag{3.9}$$

where $N(\mu, \Sigma)$ denotes the p-dimensional normal distribution with mean μ and covariance Σ. In practice the elements of the covariance matrix must be estmated. An estimate of σ^2 is

$$s^2 = \frac{V(\hat{\theta})}{n-p}, \tag{3.10}$$

and an estimate of $(F^T F)^{-1}$ is

$$\hat{C} = [F^T(\hat{\theta})F(\hat{\theta})]^{-1}. \tag{3.11}$$

When $f(x_i, \theta)$ is linear in θ then $(F^T F)^{-1}$ is independent of θ and only σ^2 needs to be estimated. An approximate $(1-\alpha) \times 100\%$ confidence interval for the unknown parameter θ_j (the j'th element of θ) is thus

$$\hat{\theta}_j \pm t(n-p)_{\alpha/2}\sqrt{s^2\hat{c}_{jj}}, \tag{3.12}$$

where $\hat{c}_{jj}$ is the j'th diagonal element of $\hat{C}$ and $t(n-p)_{\alpha/2}$ is the $\alpha/2$ quantile of the central t-distribution with $n-p$ degrees of freedom. The expression (3.12) also form the basis of the t-test, see (Gallant, 1987).

For the above stated distributional properties to hold the sequence of independent variables $\{x_i\}$ must behave properly as n tends to infinity, see (Gallant, 1987). One way this proper behaviour may be obtained is by random sampling. The function $f(x_i, \theta)$ must be twice differentiable with respect to the argument (x_i, θ) and the second partial derivatives must be continuous. Furthermore two estimability conditions must be fulfilled for $n \to \infty$. First $\frac{1}{n}\sum_{i=1}^{n}[f(x_i, \tilde{\theta}) - f(x_i, \theta)]^2$ must have a unique minimum for $\tilde{\theta}$ at the true value θ, and second $\frac{1}{n}F^T(\theta)F(\theta)$ must be non-singular. Generally these two conditions can not be verified, except for situations where it is possible to prove that they hold for any value the true θ may have.

The squared degree of determination

The squared degree of determination R^2, in linear regression also called the squared multiple correlation coefficient, is a single number which describes the model error.

The value express how large a part of the variation in the dependent variable is explained by the model. For a particular model R^2 is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{e}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{3.13}$$

where y_i is the observation of the dependent variable, $\bar{y}$ is the average of $y_i, i = 1, \ldots, n$, and $\hat{e}_i, i = 1, \ldots, n$ is the residuals or model errors.

Note that R^2 is calculated based on the residuals from the data set for which the parameters of the model are estimated. Hence, a model for which a large R^2 is obtained is not necessarily desirable, since it might be due to an over fitting of the data.

Chapter 4

Main Characteristics of the Power Consumption Data

In this chapter the power consumption is split into trend, annual fluctuations, diurnal fluctuations, and the days are grouped. The analysis is performed using time as the only explanatory variable, hence the analysis is very preliminary. Primarily local regression methods are used. If not stated otherwise the degree of the locally fitted polynomial is two.

4.1 Trend

In Figure 4.1 the power consumption data for the whole period is shown. The trend of these data is estimated by local regression with fixed bandwidths of 1/8, 1/4, 1/3, and 1/2 of the total time span. These estimates are also displayed on the figure. The estimates are calculated in 20 points equally spaced over the total time span.

It is seen that the lowest bandwidth results in inappropriate estimates at the low time points. Hence a fixed bandwidth of approximately 1/4 of the total time span seems appropriate. However, at the low time points the estimate still seems to be biased upwards. This is probably due to the winter peak at the beginning of the data period.

In order to have each point estimate based on the same No. of observations a 50% nearest neighbour estimate of the trend is calculated. At the mid time points the bandwidth of the nearest neighbour estimate equals the fixed bandwidth of 1/4 of the total time span. At the end points it equals the fixed bandwidth of 1/2 of the total time span.

In Figure 4.2 the two estimates of the trend are displayed. It is seen that the nearest neighbour estimate avoids the upward bias at the beginning of the period.

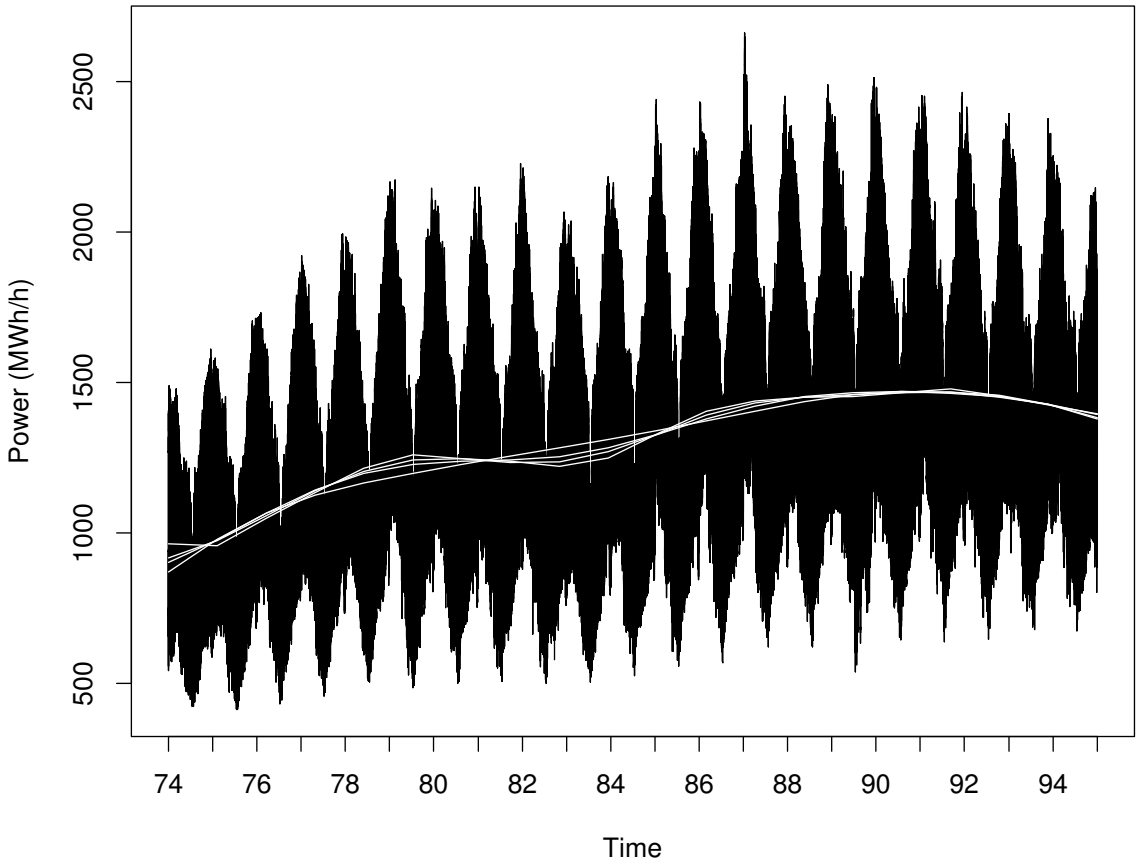


Figure 4.1. *Power consumption with estimates of the trend superimposed. The estimates corresponds to fixed bandwidths of 1/8, 1/4, 1/3, and 1/2 of the total time span.*

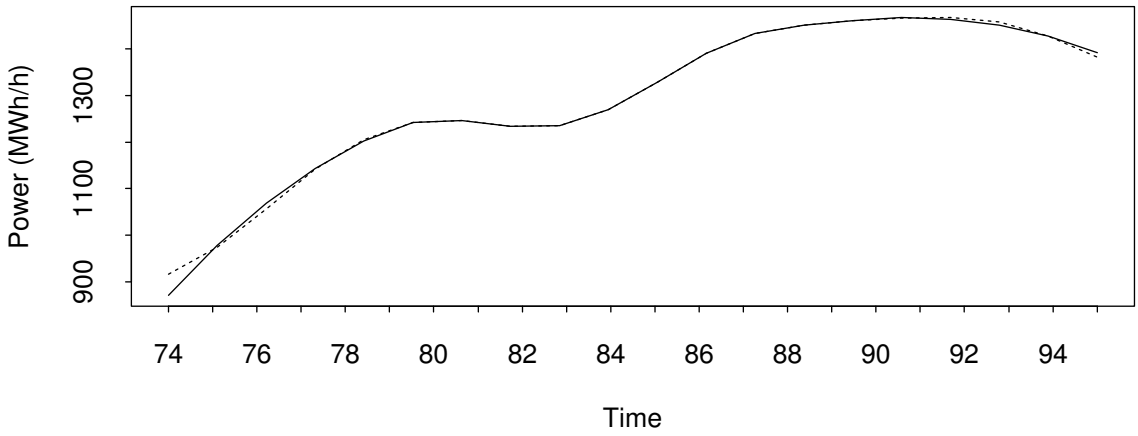


Figure 4.2. *Estimated trend of power consumption. "—" 50% nearest neighbour, "- -" fixed bandwidth of 1/4 of the total time span.*

4.2 Annual Fluctuations

The annual fluctuations are estimated like the trend but with a bandwidth low enough to catch fluctuations during the year. More precisely a fixed bandwidth corresponding to 1.5 month is chosen. This bandwidth should be able to catch fairly quick changes when a 2nd order polynomial is used. The estimate is calculated in 3000 points equally spaced over the total time span. The estimated fluctuations are displayed on Figure 4.3 together with the trend.

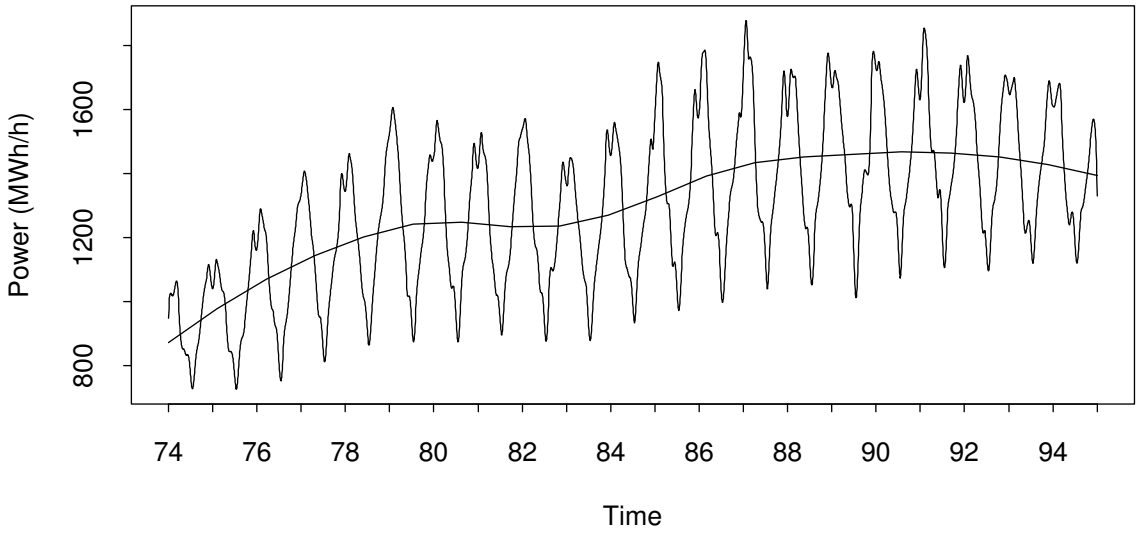


Figure 4.3. *Estimated annual fluctuations and trend of power consumption. The trend is estimated with a 50% nearest neighbour bandwidth. The annual fluctuations are estimated with a fixed bandwidth corresponding to 1/8 of a year.*

Figure 4.4 shows the annual fluctuations from Figure 4.3 with the trend displayed on the same figure subtracted. The annual fluctuations are plotted against the normalized time in year. For a specific hour this number is defined as the No. of hours since the beginning of the year divided by the total No. of hours in the year.

4.3 Diurnal Fluctuations and Grouping of Days

The trend and the annual fluctuations are subtracted from the observations of power consumption by means of linear interpolation between the points in which the annual fluctuations are calculated. Hereafter an approximate 95% confidence interval of the median is calculated for each of the times 01:00, 02:00, ..., 24:00 (see the description of `boxplot()` in (Statistical Sciences, 1993)). The calculation is based on the assumption that the observations are independent. Since the observations probably are positive correlated these confidence intervals form a lower bound of the true confidence intervals.

The calculations are performed for each of the seven groups (i)-(v) Normal working days split in the day of week, (vi) Saturdays or Half-Holy days, and (vii) Sundays or Half-Holy days. The result is shown on Figure 4.5, and it is readily seen that it is reasonable to pool the normal working days Tuesday, Wednesday, and Thursday into one group which will be called "Midweek".

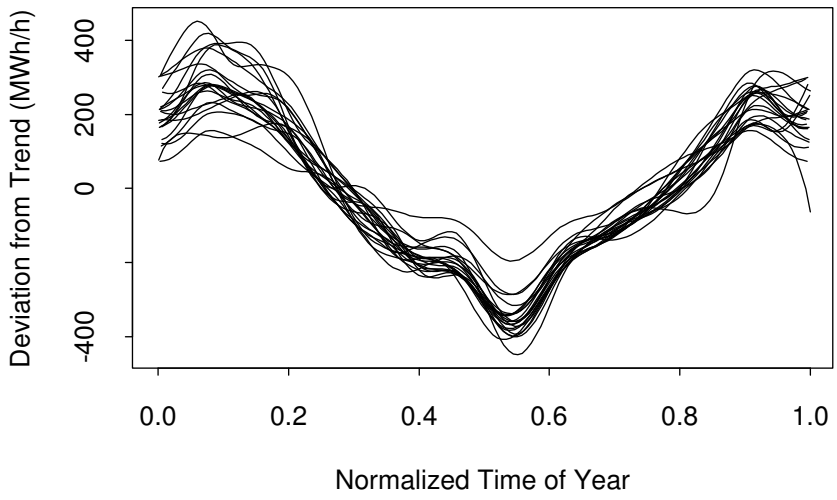


Figure 4.4. *Annual fluctuations when the trend is subtracted. Based on estimates shown in Figure 4.3.*

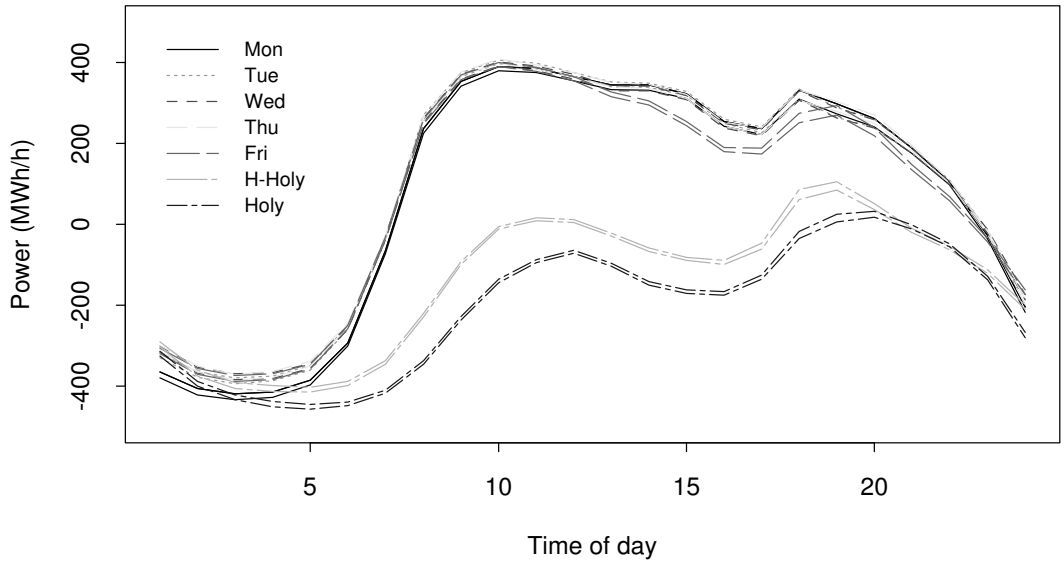


Figure 4.5. *Approximate 95% confidence intervals for the median under the assumption that the observations are independent.*

Chapter 5

Investigation Regarding Transformation of Data

Various transformations of the power consumption data are investigated. Only the results for the logarithmic transformation are included in this chapter. The other transformations investigated are all of the form $\sqrt{P-\lambda}$. Due to the ratio between level and amplitude for the power consumption data range-mean plots (group size: 24) for $\sqrt{P}$ resemble those for P. When λ is around 400 MWh/h the range-mean plots of $\sqrt{P-\lambda}$ resemble those for $\log P$.

A time plot of the logarithm transformed series (not shown) has large negative peaks, which is somewhat undesirable. In Figure 5.1 range-mean plots of untransformed and logarithm transformed power consumption data are shown. The group size is 24 and hence the plots address the dependence of the diurnal amplitude on the daily level. Two groups are clearly seen on the plots, since Weekends, Half-holy and Holy days from April to September form the group with low range.

The range-mean plots indicate that when using logarithm transformed data the daily amplitude is independent of the level, and hence this transformation seems to be applicable. This is also indicated by the range-mean plots for each year shown on Figures 5.2 and 5.3.

It is obvious that the power consumption depends on the time of the day and year and on the climate variables. For this reason it is not clear whether the logarithmic transformation should be applied[1]. For this reason the following parametric model

[1]If the power consumption data could be modelled as a pure stochastic ARIMA(p,d,q) process the range-mean plots clearly indicate that the logarithmic transformation should be applied.

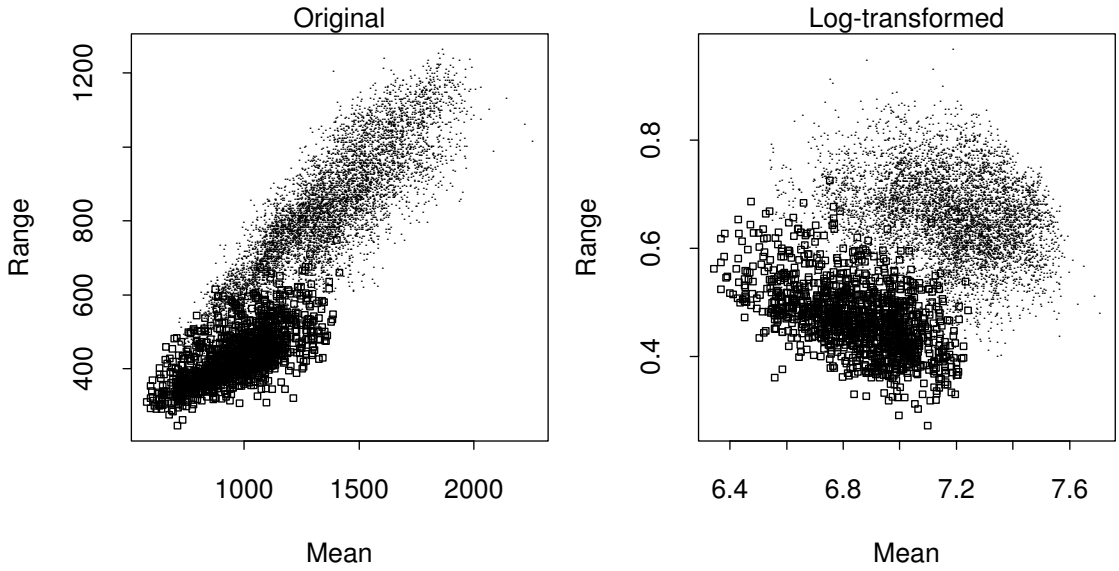


Figure 5.1. *Range-mean plots of power consumption data using a group size of 24. Weekends, half-holy and holy days from April to September are shown as "□".*

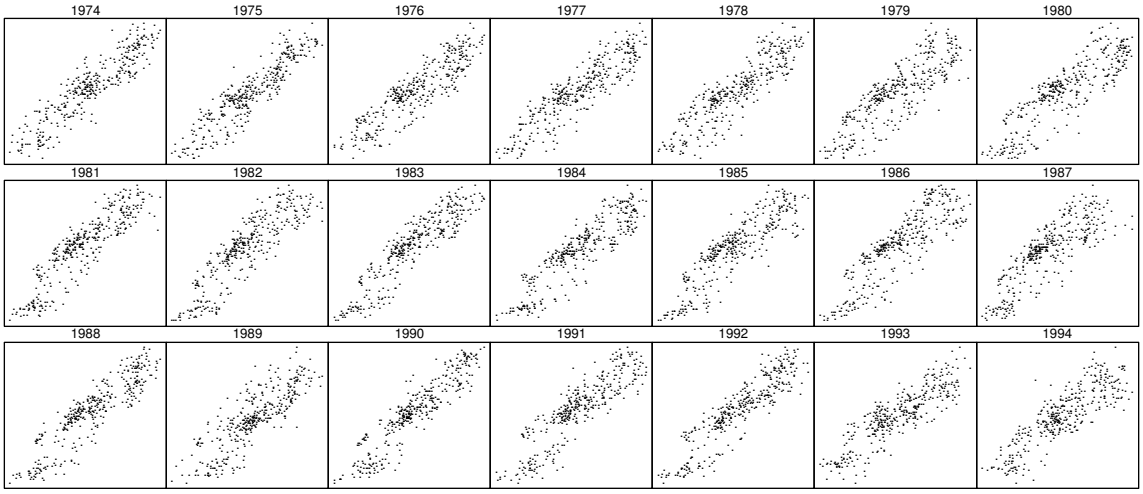


Figure 5.2. *Range-mean plots of power consumption data for each year. Group size: 24.*

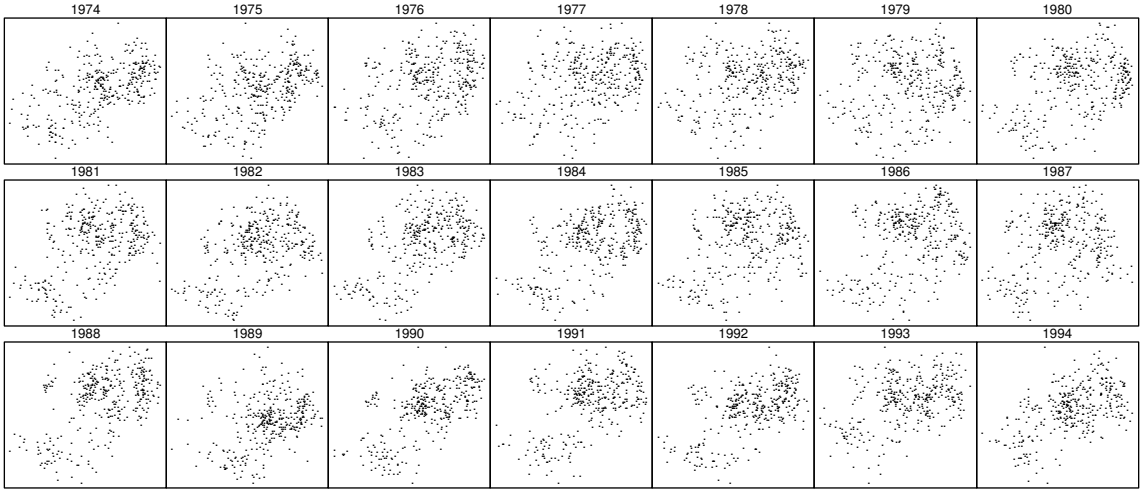


Figure 5.3. *Range-mean plots of logarithm transformed power consumption data for each year. Group size: 24.*

is fit to both original and transformed data on a monthly basis.

$$\begin{aligned} E[Y(t)] &= \mu + \sum_{i=1}^{2} c_i h_m^i(t) + \sum_{i=1}^{2}[a_{ti}T_a^i(t) + a_{wi}W^i(t) + a_{ri}R_g^i(t)] \\ &+ a_{tw}T_a(t)W(t) + a_{tr}T_a(t)R_g(t) + a_{wr}W(t)R_g(t) \\ &+ \sum_{j=1}^{5}[I_j(t)\sum_{i=1}^{5}[\mu_i + d_{1ji}\cos\frac{2\pi i h_d(t)}{24} + d_{2ji}\sin\frac{2\pi i h_d(t)}{24}] \qquad (5.1) \\ \mu_5 &= -\sum_{i=1}^{4}\mu_i \end{aligned}$$

where $Y(t)$ is either the power consumption or the logarithmic transformed power consumption at the time t. The No. of hours since the beginning of the month is denoted $h_m(t)$ and $h_d(t)$ is the time of day. $T_a(t)$, $W(t)$ and $R_g(t)$ is the temperature, wind speed and global radiation, respectively. According to the results found in Chapter 4 the days are split in five groups indexed by j; (1) Mondays, (2) Midweek, (3) Fridays, (4) Saturdays and Half-holy days, and (5) Sundays and Holy days. The function $I_j(t)$ is defined as

$$I_j(t) = \begin{cases} 1, & \text{if } t \text{ belongs to group } j \\ 0, & \text{otherwise} \end{cases} \qquad (5.2)$$

The trend during the month is accounted for by the 2nd order polynomial in the No. of hours since the beginning of the month. The deviation from the trend for days of type i is accounted for by the parameters $\mu_1, \ldots, \mu_5$. The 2nd order polynomial in the climate variables adjusts for the influence of climate. The parameters are estimated using least squares techniques. The estimates of the parameters in the Fourier expansion for Midweeks ($j = 2$) are plotted against the calendar time (the end of each month) in Figures 5.4 and 5.5.

The most clear difference is observed for the estimate of the parameter corresponding to $\cos(2\pi h_d(t)/24)$, i.e. d_{121} (`cos 1` on the plots). This is probably because the logarithmic transformation stabilize the amplitude of the daily fluctuations with respect to the level. For both the transformed and the untransformed case annual fluctuations and trends in the estimates are observed, i.e. the logarithmic transformation does not stabilize the shape of the daily fluctuations. From the plots it is judged that it will be more easy to model the annual fluctuations of the estimates as Fourier expansions with a main period of one year, for the untransformed power consumption than for the logarithmic transformed values. For this reason further studies are performed on the untransformed power consumption data.

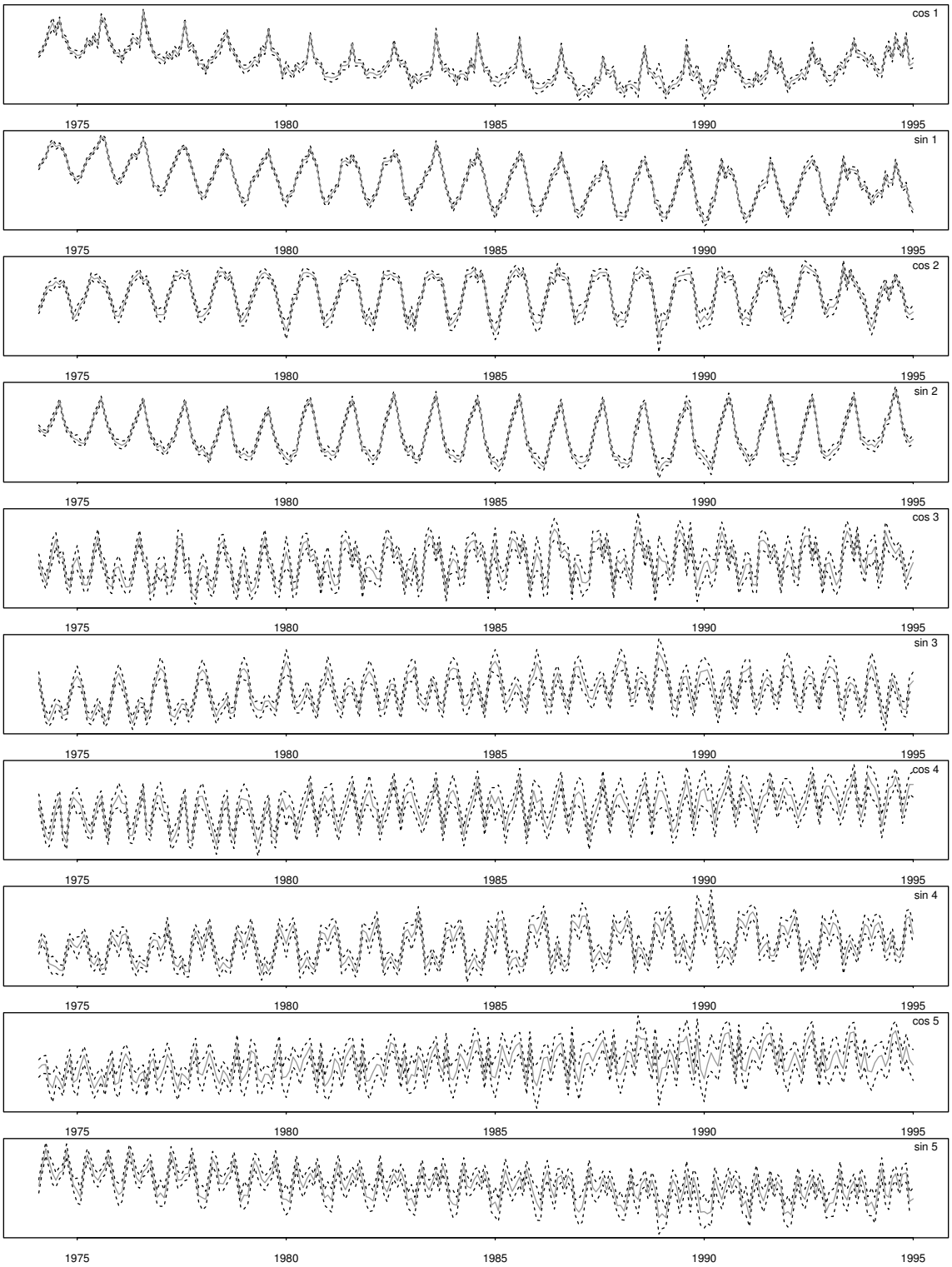


Figure 5.4. *Parameter estimates in the Fourier expansion of Midweek fluctuations, untransformed power consumption. Parameters indicated by "cos" or "sin" and the No. of periods per day. Approximate 95% confidence intervals indicated by "- -".*

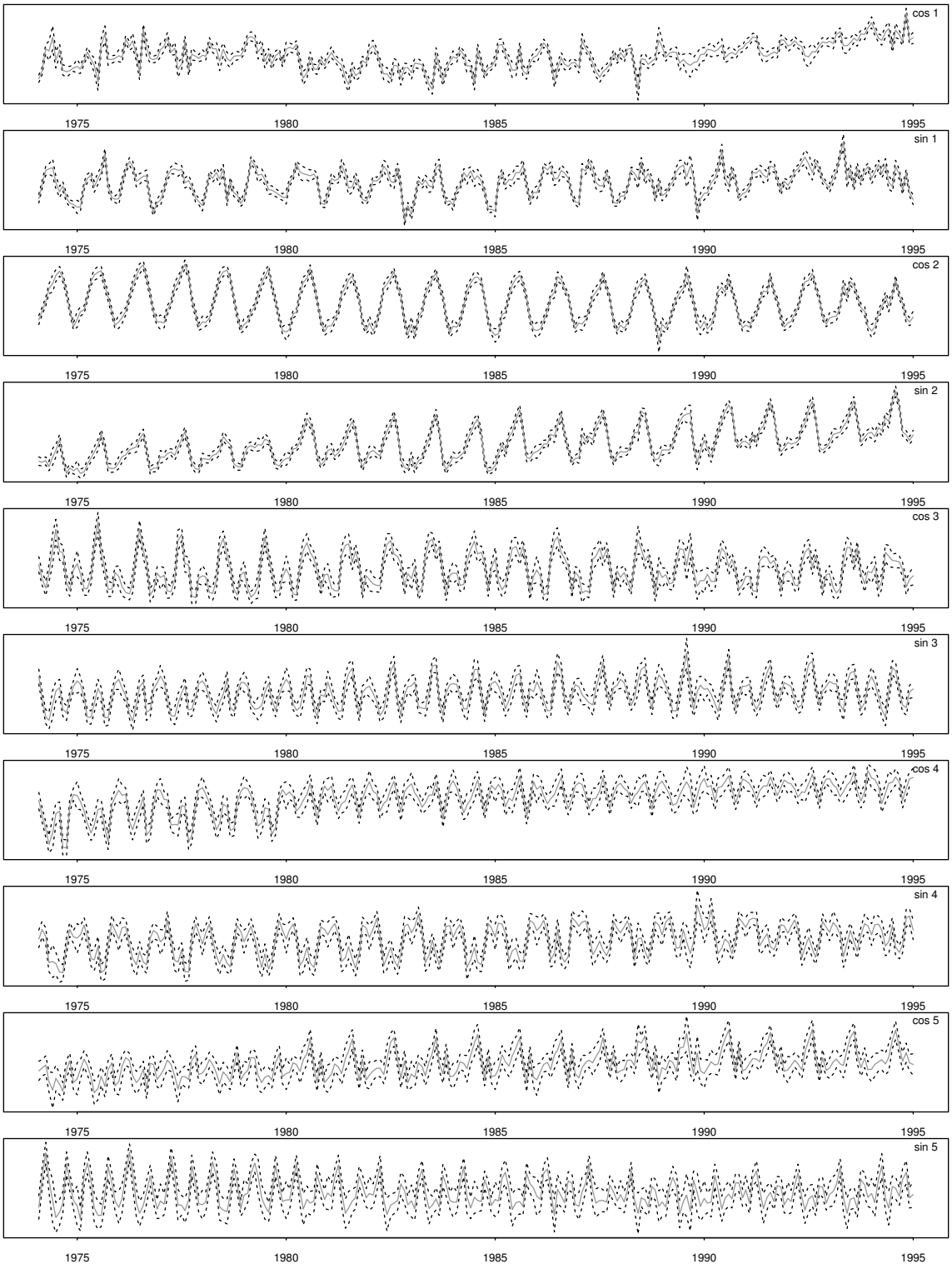


Figure 5.5. *Parameter estimates in the Fourier expansion of Midweek fluctuations, logarithmic transformed power consumption. Parameters indicated by "cos" or "sin" and the No. of periods per day. Approximate 95% confidence intervals indicated by "- -".*

Chapter 6

Preliminary Investigation of the Influence of Climate

The influence of climate on power consumption is investigated by use of local regression. Using hourly observations it is however necessary to adjust for the time of the day. This is done by modelling the daily fluctuations using a Fourier expansion.

In Figure 6.1 the median of the power consumption adjusted for trend and annual fluctuations for Midweek data is plotted. Also the result of fitting Fourier expansions of orders 4, 5 and 6 to the medians is shown. It is seen that the order must be at least five and that the sixth order fit is almost identical to the fifth order fit. Investigations show that in order to fully catch the raise from 17:00 to 18:00 one must use a Fourier expansion of order nine. However, in order to keep the number of parameters down the fifth order expansion is chosen.

Only data from one year is used, since then (i) it is not necessary to include a trend in the model, and (ii) it makes the computations feasible and relatively fast. Since the temperature is assumed to be the most important climate variable, the year is chosen based on this variable. I 1982 both low and high temperatures occured (-21 to 28.5 ^oC for all days, -15.6 to 27.4 ^oC for Midweek). Hence data from this year is used.

6.1 Temperature and Wind Speed

The following model is assumed

$$E[P] = f(T_a, W) + \sum_{i=1}^{5}[c_{1i}(T_a, W)\cos(\frac{2\pi ih}{24}) + c_{2i}(T_a, W)\sin(\frac{2\pi ih}{24})], \qquad (6.1)$$

where P, T_a, W, and h are simultaneous values of power consumption, ambient air temperature, wind speed, and time of day. The functions $f(T_a, W)$ and

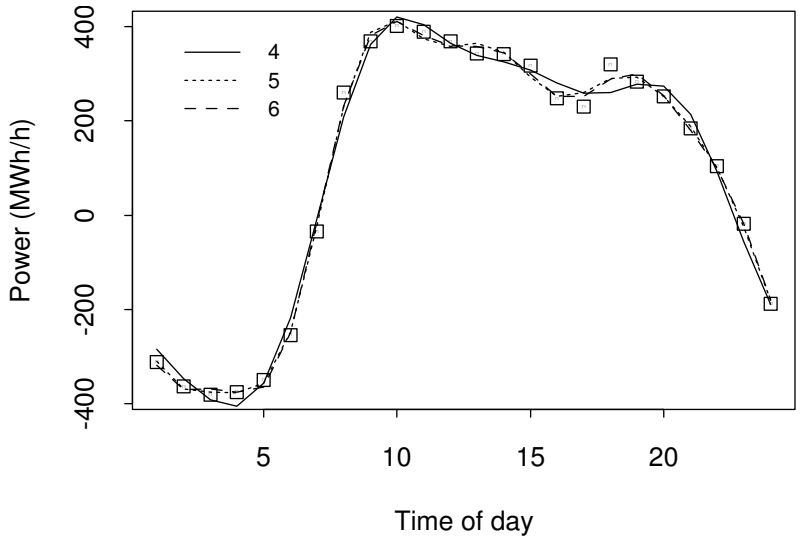


Figure 6.1. *Median of power consumption adjusted for trend and annual fluctuations for Midweek (Tuesday to Thursday) (□). Fitted fitted curves using Fourier expansions of orders 4, 5 and 6 are also shown.*

$c_{ji}(T_a, W); i = 1, \ldots, 5; j = 1, 2$ are estimated by local regression as described in Section 3.1. $f(T_a, W)$ is approximated by a local 2nd order polynomial, while the other functions are approximated by local constants. Since the temperature varies over the year model (6.1) allows for slow changes in the diurnal profile.

The functions in model (6.1) are estimated using nearest neighbour bandwidths ranging from 0.2 to 0.7 and with no lower limit on the resulting bandwidths. The scaling factors are chosen so that the scaled variables have approximately same range. This is accomplished by a scaling factor of 1 ^oC for T_a and $1/3$ m/s for W. The points in which the estimation is performed are shown in Figure 6.2.

Contour plots of estimates of $f(T_a, W)$ in (6.1) are shown in Figure 6.3. From the plots it seems reasonable to select a bandwidth equal to 0.4 or larger (the estimates are too noisy for bandwidths 0.2 and 0.3). An upper limit on the bandwidth is not easily deducted, but the following observations holds for bandwidths equal to 0.4 or larger.

Figures 6.4 and 6.5 show the estimates for bandwidths 0.4 to 0.7 for fixed T_a and W, respectively. From these plots it is seen that the influence of T_a is much stronger than the influence of W.

When the ambient air temperature is above 10 ^oC wind speeds below 3 m/s do not affect the power consumption (nearly vertical contour lines). For wind speeds above 3 m/s the contour lines are parallel, which indicates that $f(T_a, W)$ is additive in T_a and W when the wind speed is above 3 m/s. From these findings it follows that for the whole range of T_a and W, the function $f(T_a, W)$ is not additive in the arguments. From the plots it is, however, seen that the non-additivity is weak, and

hence possibly an artifact.

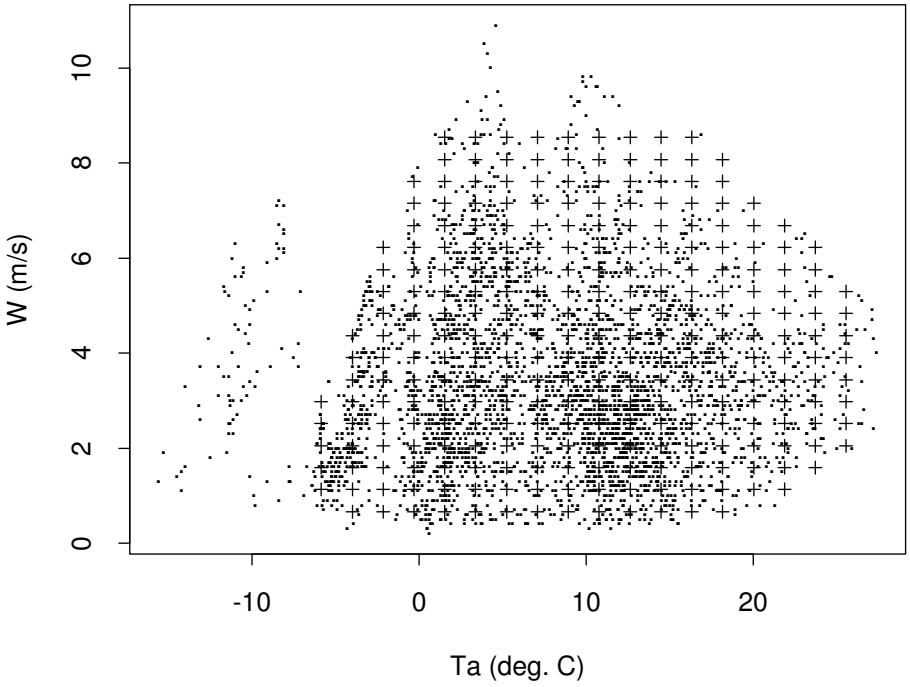


Figure 6.2. *Points (+) in which model (6.1) is estimated. Data shown as small dots. Climate variables: (Jensen, 1995).*

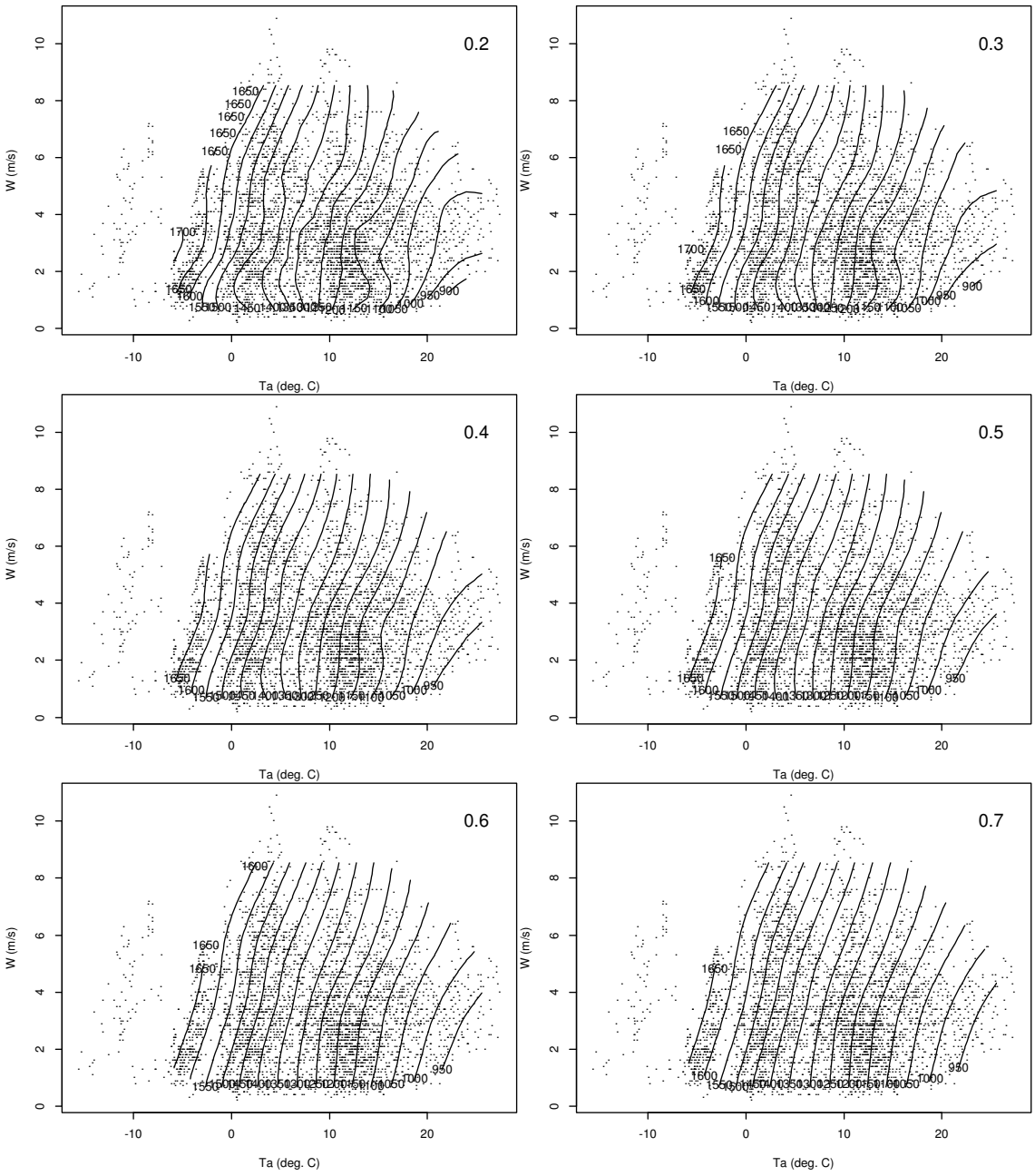


Figure 6.3. *Contour plots of estimates of $f(T_a, W)$ in (6.1) for nearest neighbour bandwidths in the interval 0.2 to 0.7 (no lower limit on the bandwidth). Scale factors: 1 °C for T_a, 1/3 m/s for W. Data shown as small dots. Climate variables: (Jensen, 1995)*

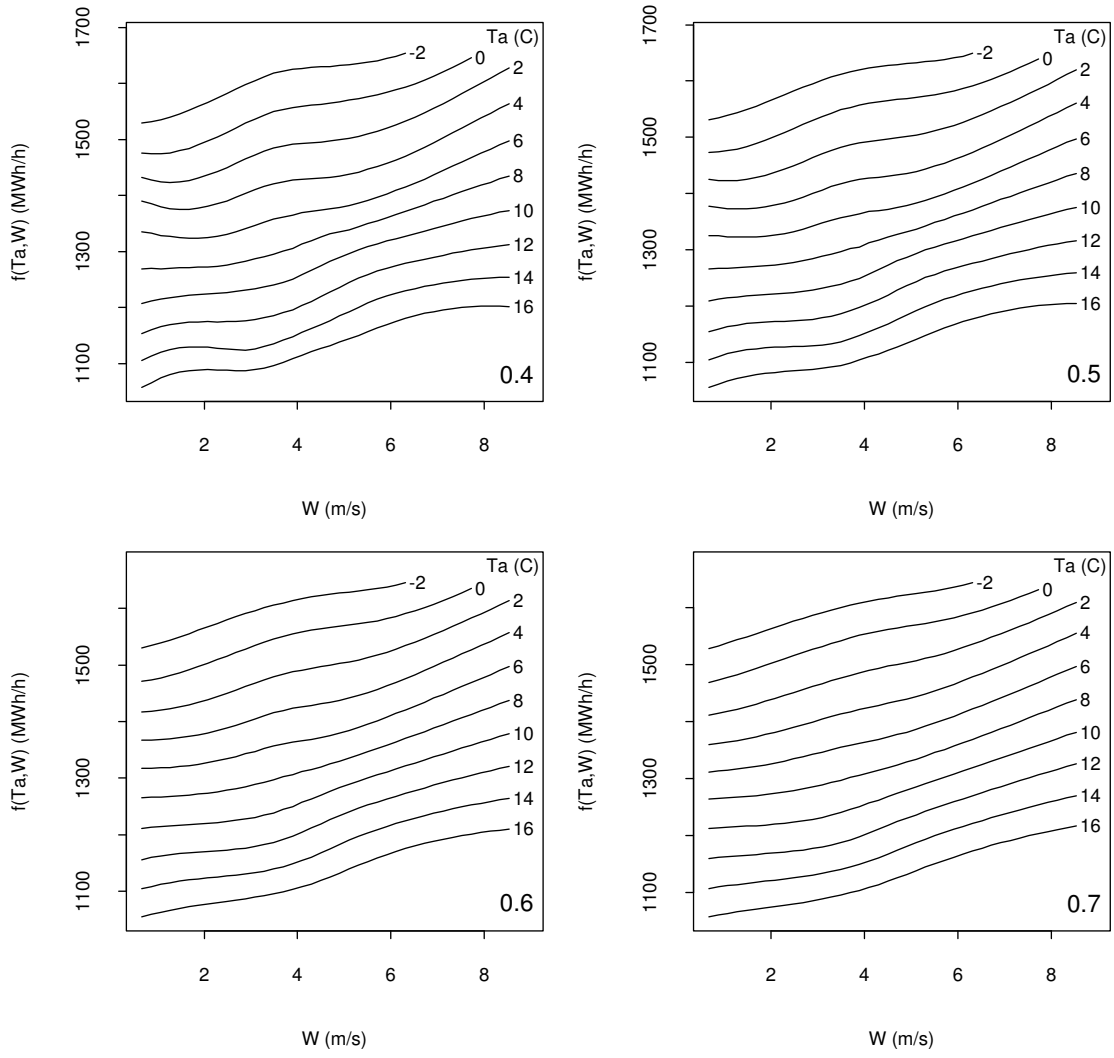


Figure 6.4. *The estimates of $f(T_a, W)$ shown in Figure 6.3 for selected values of T_a. Only estimates for nearest neighbour bandwidths in the interval 0.4 to 0.7 are shown. Climate variables: (Jensen, 1995)*

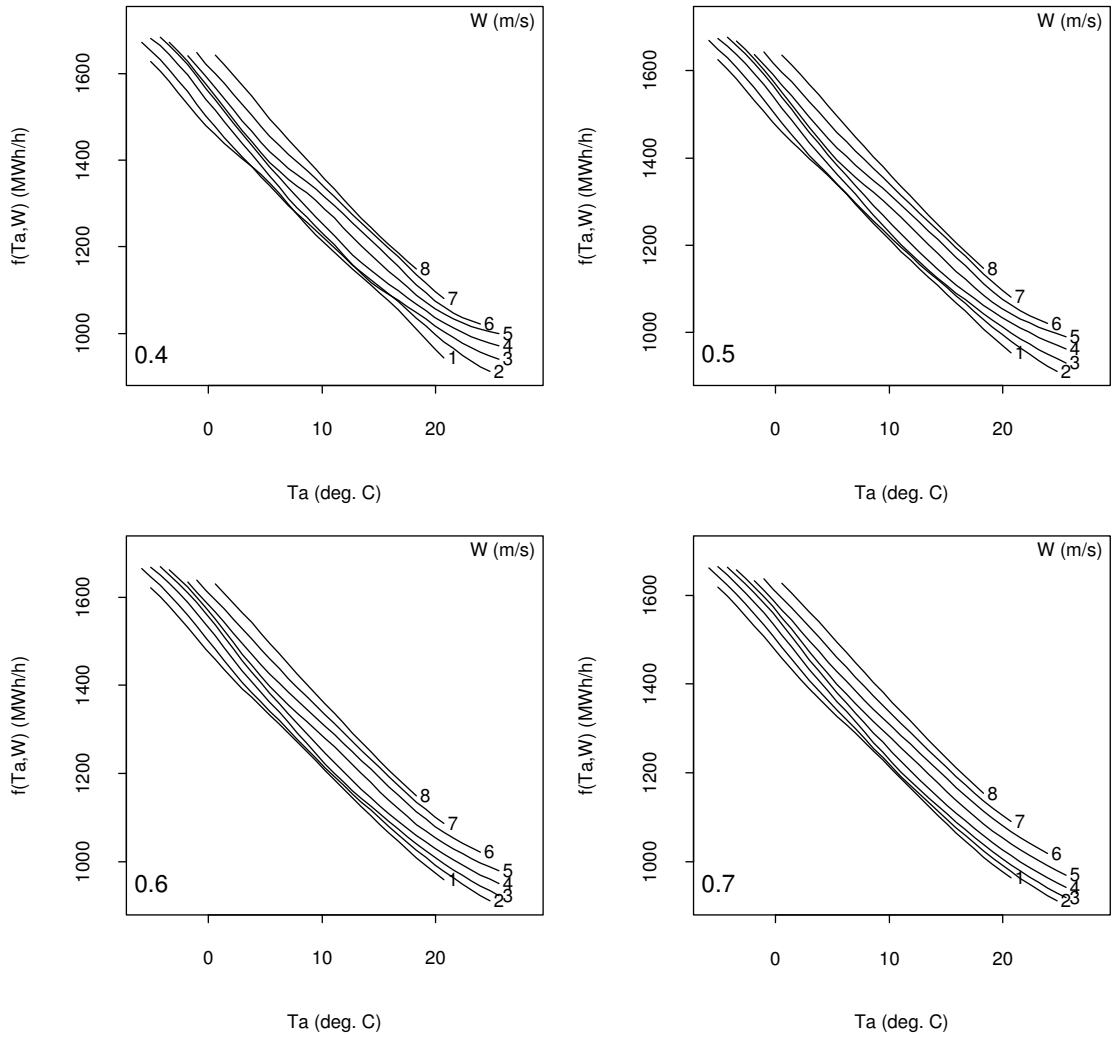


Figure 6.5. *The estimates of $f(T_a, W)$ shown in Figure 6.3 for selected values of W. Only estimates for nearest neighbour bandwidths in the interval 0.4 to 0.7 are shown. Climate variables: (Jensen, 1995)*

Investigations have shown that the wind speed is nearly uncorrelated with the ambient air temperature, the time of day, and the time of year. Furthermore the wind speed is of less importance than the ambient air temperature, as concluded above. For these reasons it is reasonable to exclude the wind speed from model (6.1), and hence the model

$$E[P] = f(T_a) + \sum_{i=1}^{5}[c_{1i}(T_a)\cos(\frac{2\pi ih}{24}) + c_{2i}(T_a)\sin(\frac{2\pi ih}{24})], \tag{6.2}$$

is obtained. This model is valid for wind speeds around the average of the observed wind speeds (3.4 m/s). The main function of interest ($f(T_a)$) is approximated locally using a 2nd polynomial. The remaining functions are approximated by local constants.

Figure 6.6 shows estimates of $f(T_a)$ in (6.2) using nearest neighbour bandwidths ranging from 0.1 to 0.7 (no lower limit). Note that for the small bandwidths $\hat{f}(T_a)$ is increasing for the high temperatures. Investigations have shown that the relevant temperatures are observed from 10:00 until 21:00. The reason why the estimates are not increasing when higher bandwidths are used is thus that the diurnal profile adjusts for the time of day.

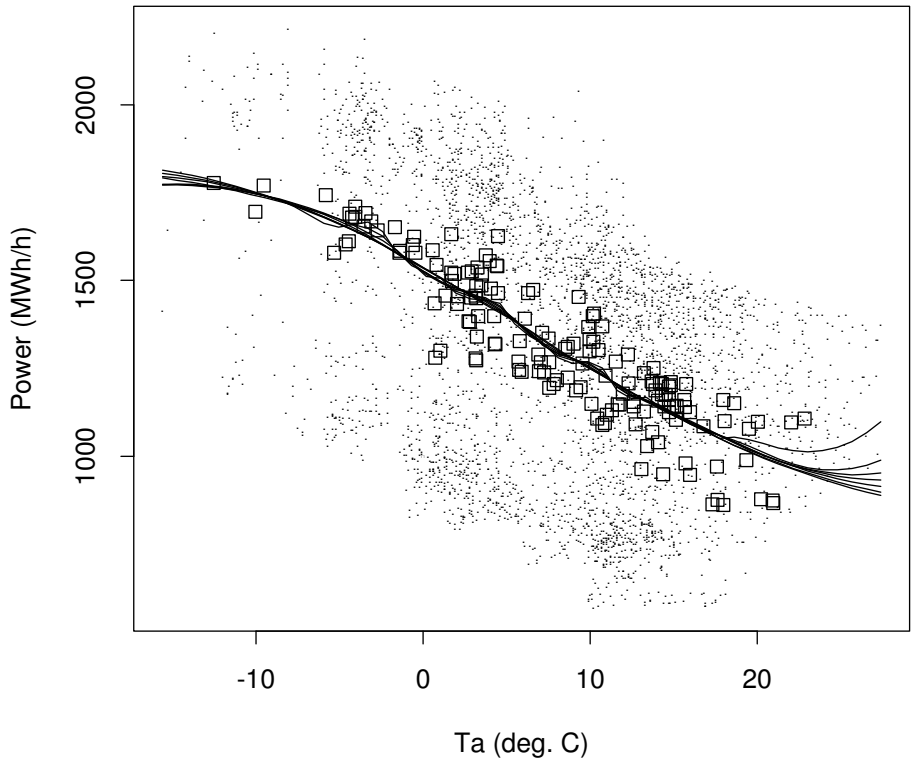


Figure 6.6. *Estimates of $f(T_a)$ in (6.2) for nearest neighbour bandwidths in the interval 0.1 to 0.7 (no lower limit on the bandwidth). Data shown as small dots, 24h averages shown as "□". Climate variables: (Jensen, 1995)*

It is noted that the 24h averages lies nicely around the estimated relation between

temperature and power consumption. This indicates that the method has been successful in splitting the variation between temperature and time.

The non-parametric estimates in Figure 6.5 and 6.6 can be used for suggesting a parametric relation between the power consumption and the temperature. It seems appropriate to use a logistic function to describe the relation. From Figure 6.4 it is seen that the relation between power consumption and wind speed can be fairly well described by a linear function.

6.2 Temperature and Global Radiation

The global radiation (R_g) is handled like the wind speed, hence the functions in the model

$$E[P] = f(T_a, R_g) + \sum_{i=1}^{5}[c_{1i}(T_a, R_g)\cos(\frac{2\pi ih}{24}) + c_{2i}(T_a, R_g)\sin(\frac{2\pi ih}{24})], \qquad (6.3)$$

is estimated by local regression. The main function of interest $f(T_a, R_g)$ is approximated locally by a 2nd order polynomial. The remaining functions are approximated by local constants. In order to obtain approximately the same range for the scaled values a scaling factor of 1 $^\circ C$ is used for T_a and a scaling factor of 30 W/m^2 for R_g.

The estimation is difficult due to the following (obvious) facts:

- 48% of the observations of R_g is 0 W/m^2,
- R_g is strongly dependent of the time of day and the time of year, and
- R_g is correlated with T_a.

To avoid too small bandwidths at the low values of R_g a lower limit of 5.0 is used. Nearest neighbour bandwidths of 0.2, 0.4 and 0.7 are used. Only 0.7 yields a stable estimate, which is due to the sparseness of data at the high values of R_g, see Figure 6.7.

In Figure 6.7 the data, the fitting points, and contour lines of $\hat{f}(T_a, R_g)$ are shown. The fit is highlighted in Figure 6.8, which shows $\hat{f}(T_a, R_g)$ for fixed T_a and R_g, respectively.

From the plots it is concluded that the relation between global radiation and power consumption may be approximated by an exponential decreasing function.

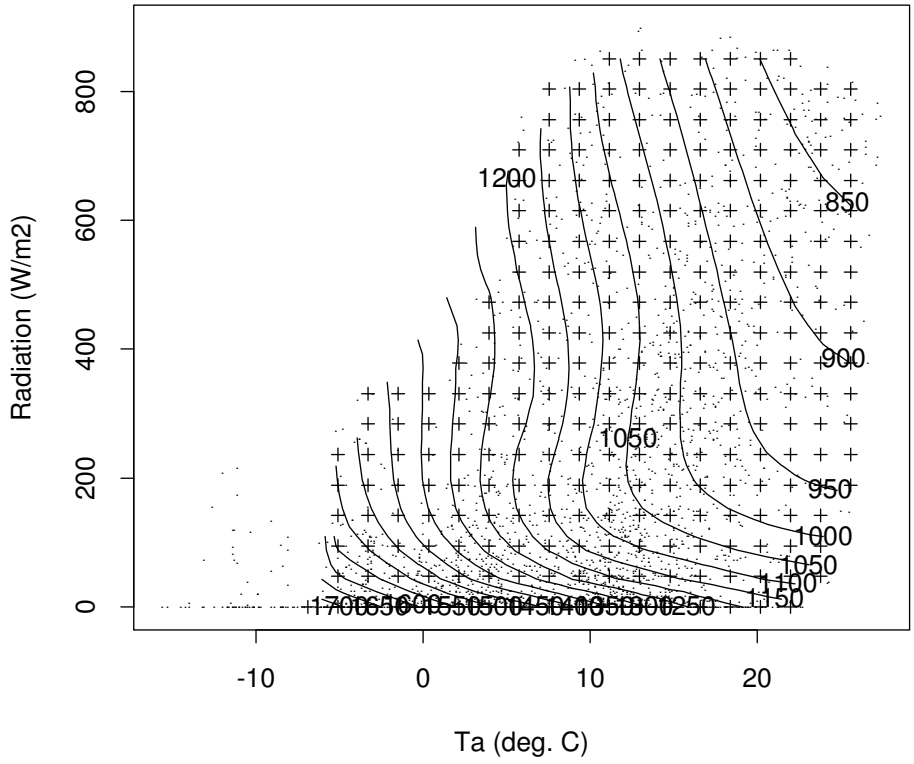


Figure 6.7. *Contour plots of the estimate of $f(T_a, R_g)$ in (6.3) for nearest neighbour bandwidth 0.7 and a lower limit on the bandwidth of 5.0. Scale factors: 1 °C for T_a, 30 W/m^2 for R_g. Estimation points shown as "+". Data shown as small dots. Climate variables: (Jensen, 1995)*

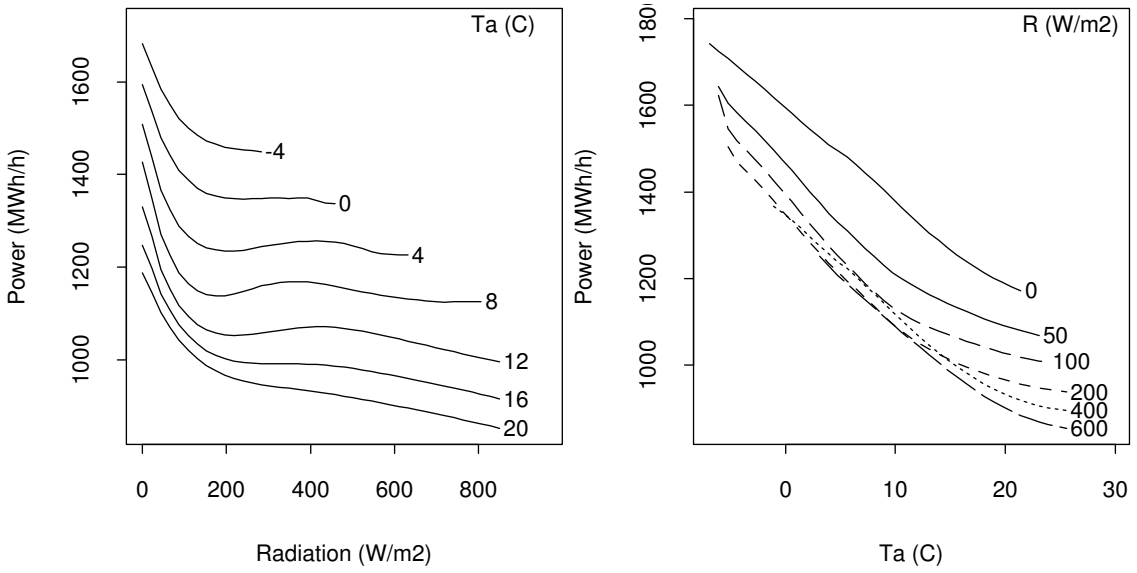


Figure 6.8. *The estimate of $f(T_a, W)$ shown in Figure 6.8 for selected values of T_a and R_g. Climate variables: (Jensen, 1995)*

Chapter 7

Parametric Models

This chapter describes globally parametric models, which are used for simultaneous estimation of periodic fluctuations and climate response. Both static models and models with a dynamic response on one or more climate variables are considered.

In order to stabilize the estimation procedure numerically the power consumption is divided by 1000 and each of the climate variables by 10. In general estimates of parameters correspond to this scaling, whereas the scaling is reversed on plots related to response, diurnal fluctuation, etc.

7.1 Identification of Annual Variations in the Diurnal Fluctuation

As described in Chapter 5 a model was estimated on monthly basis. In this section the results using model (5.1), with $Y(t)$ being the power consumption at time t, are described. The model uses simple adjustments for trend and influence from climate and hence only results related to the diurnal fluctuations will be presented.

In Figure 7.1 the estimated diurnal fluctuations are shown for each month and year. The irregular estimate for November, 1989 is caused by the fact that 566 out of 720 observations of climate variables are missing.

As expected the amplitude depends on the time of year. Minor variations over the period 1974 – 1995 are observed, the most clearly being the absence of the evening peak for the months April – September from 1980 and onwards. This corresponds exactly to the years and months in which daylight savings has been used.

The fluctuations in the parameters are shown in Figure 5.4. From this figure it is concluded that the annual fluctuation in the estimates approximated by Fourier expansions of order five, represents a reasonable compromise between the need for

adequate approximations, and the need to limit the number of parameters in models to be used on annual basis.

7.2 Identification of the Order of the Fourier Expansions

In order to formulate parametric models, for which the parameters may be estimated on a yearly basis, appropriate orders of the Fourier expansions describing

1. the diurnal fluctuation,
2. the annual variation of the diurnal fluctuation, and
3. the annual fluctuation

must be identified. Items 1 and 2 has been discussed in Chapter 6 and Section 7.1, respectively. In these cases Fourier expansions of order five seem appropriate.

Preliminary identification of the order of the Fourier expansion of the annual fluctuation is based on the estimated annual fluctuations displayed in Figure 4.4. These estimates are averaged so that one curve only describes the average estimated annual fluctuation. Kernel smoothing with triangular weight function and bandwidth 4/365, which corresponds to 4 days, is used in order to obtain averages which changes smoothly over the year. The measure of distance is angular so that the distance from 1 to 0 is zero. This ensures that the averages form a periodic function. Fourier expansions of order six and seven are fitted to these averages, and the results are displayed in Figure 7.2.

It is seen that a Fourier expansion of order seven approximates the average estimated annual fluctuation quite well. However, the drop from the end to the beginning of the year covers a period of 8 weeks. Since the drop probably is caused by the Christmas holidays the period is far to long, indicating over-smoothing at that particular time of year. For this reason the order of the Fourier expansion must be larger than seven in order to describe the drop related to the Christmas holidays, which lasts for about $1\frac{1}{2}$ week.

Using a Fourier expansion of order 16 yields a minimum period length of 52/16 = 3.25 weeks. This should be able to describe a drop of half the period length, i.e. approximately 1.6 weeks, see Figure 7.3. This number is much closer to the expected length of the Christmas holidays. Unfortunately the number decreases very slowly with the order of the Fourier expansion when the order is high, see Table 7.1. For these reasons the annual fluctuation is initially approximated with a Fourier expansion of order 16.

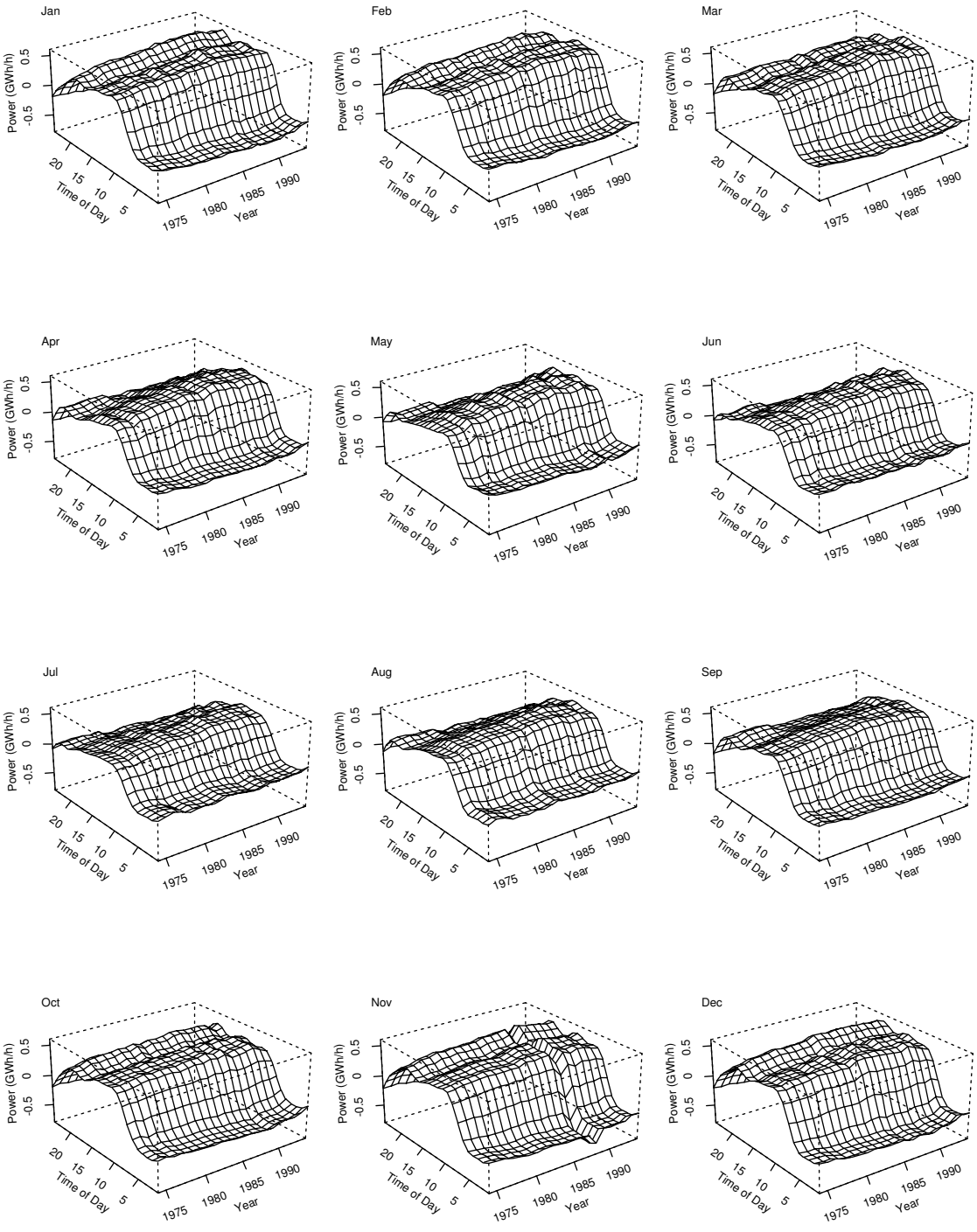


Figure 7.1. *Estimates of diurnal fluctuations for Midweek, based on model (5.1) with* $Y(t)$ *being the power consumption. The parameters are estimated on monthly basis.*

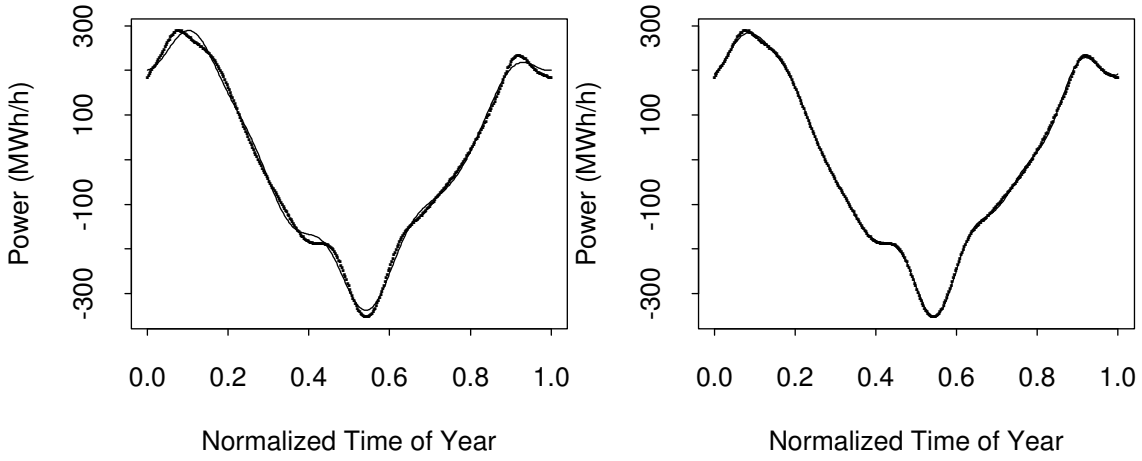


Figure 7.2. *Kernel smoothing of the estimates shown in Figure 4.4 with triangular weight function and bandwidth 4/365. Fourier expansions of order 6 (left) and 7 (right) fitted to the kernel values are superimposed.*

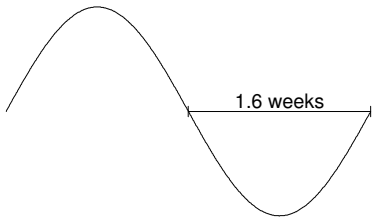


Figure 7.3. *One period of the term corresponding to the highest frequency in a 16th order Fourier expansion of the annual fluctuation.*

Order	Half Period	Decrease	Order	Half Period	Decrease
7	3.71	–	14	1.86	0.1429
8	3.25	0.4643	15	1.73	0.1238
9	2.89	0.3611	16	1.62	0.1083
10	2.60	0.2889	17	1.53	0.0956
11	2.36	0.2364	18	1.44	0.0850
12	2.17	0.1970	19	1.37	0.0760
13	2.00	0.1667	20	1.30	0.0684

Table 7.1. *Length in weeks of half the period length of the term corresponding to the highest frequency of Fourier expansions with orders 7 to 20. Also, the decrease from the previous order is displayed.*

7.3 A compact description of Fourier expansions

Fourier expansions of high order are used extensively in the following. For this reason it is reasonable to define some notation which is short but identifies the Fourier expansion uniquely. The basic notation for a Fourier expansion of order n with main period τ is $F(t;\tau,n)$, i.e.

$$F(t;\tau,n) = \sum_{i=1}^{n}[c_{1i}\cos\frac{2\pi i h_\tau(t)}{\tau} + c_{2i}\sin\frac{2\pi i h_\tau(t)}{\tau}], \tag{7.1}$$

where c_{ji} are the parameters and $h_\tau(t)$ is the time in the cycle with main period τ at the running time t, e.g. if $\tau = 24h$ and $t-(t-1) = 1h$ then $h_\tau(t)$ may be defined as the time of day.

The basic notation has some extensions as described below.

Parameters which is known functions: $F(t;\tau,n;\theta(x))$. This notation is used to emphasize that the parameters (c_{ji} in (7.1)) are functions describing how the parameters vary between pairs of i and j. If, for instance, the parameters are linear in the variable x the notation is $F(t;\tau,n;bx+a)$. In this case c_{ji} in (7.1) equals $b_{ji}x + a_{ji}$.

If the diurnal fluctuation varies during the year, and if both these variations are expressed as Fourier expansions, the notation would be $F(t;24h,n_d;F(t;1\ year,n_y))$. Note that if the main period is one year, it corresponds to 8760 hours in normal years and 8784 hours in leap years.

7.4 Summary on Static Models

Based on the investigations so far the most simple linear model describing the variations of the power consumption $P(t)$ is

$$\begin{aligned}
E[P(t)] &= \mu_0 + F_y(t;\text{1year},16) \\
&+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t;\text{24h},5;F_{iy}(t;\text{1year},5))] \\
&+ a_T T_a(t) + a_W W(t) + a_R R_g(t), \qquad (7.2)\\
\mu_5(t) &= -\sum_{i=1}^{4}\mu_i(t), \\
\mu_i(t) &= \gamma_i + F_{i0}(t;\text{1year},5); \quad i = 1,\ldots,4
\end{aligned}$$

where $E[\cdot]$ denotes expectation. Air temperature, wind speed and global radiation are denoted $T_a(t)$, $W(t)$ and $R_g(t)$, respectively, t is the time. The day groups

$i = 1, \ldots, 5$ corresponds to Monday, Midweek, Friday, Saturday / Half-holy day and Sunday / Holy day. $I_i(t)$ is the indicator function for day i defined as (5.2).

The level at a particular time of the year is $\mu_0 + F_y(t; 1\text{year}, 16)$. $\mu_i(t)$ describes the deviation from this level for days of type i at a particular time of year. The diurnal fluctuation, for days of type i, at a particular time of year is $F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))$.

As shown in Chapter 6 the air temperature is the most important climate variable. For this reason the following simple non-linear extensions are considered first:

Logistic dependence on temperature:

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; 1\text{year}, 16) \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(a_{T1}(T_a(t) - a_{T0}))} + a_W W(t) + a_R R_g(t). \end{aligned} \tag{7.3}$$

Logistic dependence on temperature (reparametrization):

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; 1\text{year}, 16) \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp((a_{T1}/a_{T2})(T_a(t) - a_{T0}))} + a_W W(t) + a_R R_g(t). \end{aligned} \tag{7.4}$$

Logistic dependence on temperature and annual fluctuation driven solely by the temperature:

$$\begin{aligned} E[P(t)] &= \mu_0 \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(a_{T1}(T_a(t) - a_{T0}))} + a_W W(t) + a_R R_g(t). \end{aligned} \tag{7.5}$$

The reason why the reparametrization (7.4) of model (7.3) is considered is explained below.

Let $f_1(T_a)$ and $f_2(T_a)$ be the non-linear functions describing the influence of T_a on $P(t)$ in (7.3) and (7.4), respectively. For both $f_1(\cdot)$ and $f_2(\cdot)$ it holds that $f(-\infty) = a_{T2}$, $f(\infty) = 0$ and $f(T_a) = a_{T2}/2 \Leftrightarrow T_a = a_{T0}$. Furthermore it holds that $f_1'(a_{T0}) = -a_{T1}a_{T2}/4$ and $f_2'(a_{T0}) = -a_{T1}/4$.

From this it is seen that for (7.4) each of the parameters has a unique interpretation. The horizontal placement is determined by a_{T0}, the height is determined by a_{T2}, and

the steepness is determined by a_{T1}. For (7.3) the steepness is determined by both a_{T1} and a_{T2}. For this reason it is expected that the correlation between the estimates of the parameters a_{T1} and a_{T2} is numerically lower for (7.4) than for (7.3).

The parameters of the four models mentioned above are estimated by Ordinary Least Squares (OLS). The estimation is based on data from 1982 only. In Table 7.2 R^2 values for the models are shown. As seen from the table the R^2 values of the models are all quite high. However, model (7.5) in which the annual fluctuation are solely described by the temperature has a somewhat lower R^2.

Model	No. of Parameters	R^2
(7.2)	630	0.9853
(7.3)	632	0.9856
(7.4)	632	0.9856
(7.5)	600	0.9595

Table 7.2. *No. of parameters and R^2 for simple models when fitted to data from 1982.*

For model (7.3) the correlation between estimates of a_{T1} and a_{T2} is -0.9359, whereas it is 0.3892 for model (7.4). This confirms the theoretical argument stated above.

The models do not describe all systematic variation in the data. This is clearly seen from Figure 7.4 in which a time-plot of the residuals are shown. All parameters describing the annual fluctuation, except c_{15}, are highly significant (t-test: $p = 0.0012$ for c_{210}, the remaining have $p < 0.0001$). Although the residuals are not white noise and statistical tests of the significance of parameters therefore are not strictly valid, it was decided to increase the order of the Fourier expansion of the annual fluctuation from 16 to 18. For the diurnal fluctuations the t-tests of many of the parameters show insignificant p-values. For this reason the parametrization of the diurnal fluctuations are not changed.

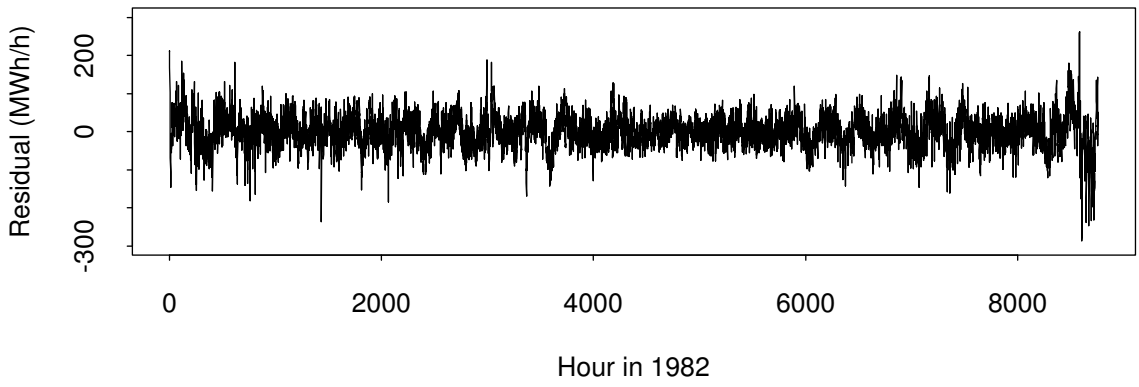


Figure 7.4. *Residuals when model (7.3) is fitted to data from 1982.*

Now the model will be extended by including a parametric description of the dependency on the global radiation. The above findings and the presumed exponential

dependence on global radiation result in the following model[1].

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; 1\text{year}, 18) \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(a_{T1}(T_a(t) - a_{T0}))} + a_W W(t) + a_{R2} \exp(-a_{R1} R_g(t)). \quad (7.6) \end{aligned}$$

The parameters of this model are fitted for each year using OLS. In Table 7.3 the R^2 values are shown. It is seen that the increase in R^2 from Table 7.2 to 7.3 is large compared with the increase observed when the linear model (7.2) is compared with the initial non-linear model (7.3) in Table 7.2.

Year	R^2	Year	R^2	Year	R^2
1974	0.9830	1981	0.9872	1988	0.9896
1975	0.9865	1982	0.9870	1989	0.9809
1976	0.9864	1983	0.9881	1990	0.9869
1977	0.9874	1984	0.9882	1991	0.9869
1978	0.9869	1985	0.9869	1992	0.9882
1979	0.9881	1986	0.9886	1993	0.9821
1980	0.9875	1987	0.9883	1994	0.9800

Table 7.3. *R^2 values for model (7.6). The model has 637 parameters.*

In Figure 7.5 the residuals when model (7.6) is fitted to data from 1982 are shown. Compared with model (7.3) the maximum size of the residuals at the end of the year has been reduced with approximately 100 MWh/h to approximately 200 MWh/h. This is probably due to the increase in the order of the Fourier expansion of the annual fluctuation. The time-plot shows that the residuals are not white noise.

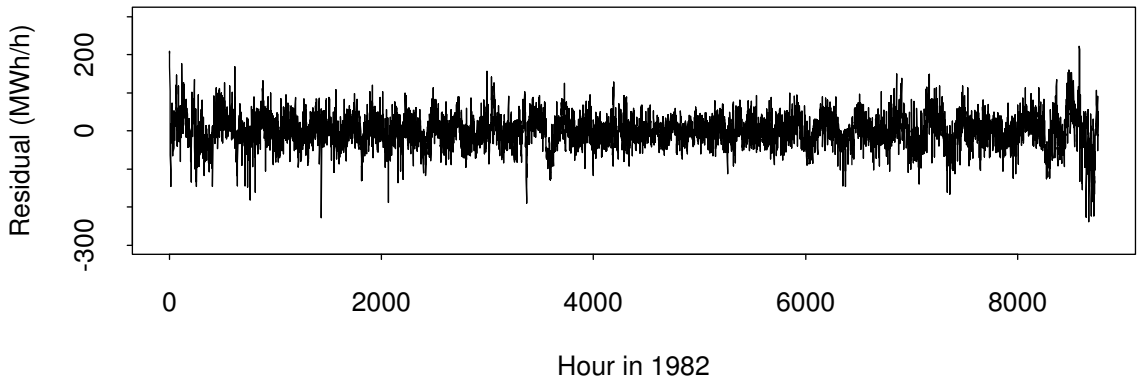


Figure 7.5. *Residuals when model (7.6) is fitted to data from 1982.*

When the residuals are compared with the climate variables no dependencies are observed. Hence the static climate response seems to be fairly well described. When

[1]The computer time used for estimation was longer for model (7.4) than for model (7.3).

the residuals are compared with the time of day, a small peak is observed around 08:00 and 18:00. As mentioned in Section 6 at least order nine of the Fourier expansion is needed to catch the peak at 18:00. This leads to too many parameters and hence the order of the Fourier expansion of the diurnal fluctuation is not changed. For all years plots like the on in Figure 7.5 show serial correlation.

Figures 7.6, 7.7 and 7.8 show the estimated annual fluctuation and climate response for all years. Note that within each year the y-axes are identical and the minima are set to zero. Furthermore the response on the global radiation is shown only for values below 800 W/m^2, although values as high as 1005 W/m^2 occur. The level stated in the figure is the level obtained when the four functions are at their minimum. The functions are plotted only in the range for which observations exists. It is seen that the wind speed and global radiation are of minor importance. The slope of the response on the wind speed changes from positive to negative over the period. The changes in the response on the wind speed are seen in the period from 1987 to 1994, with the first negative slope in 1990. This phenomena is probably due to the increase in the installed capacity on wind mills. A large part of the variation is described by the time of year. For many years the air temperature is important.

Figures 7.9 and 7.10 show the diurnal fluctuations for Midweeks (Tuesdays, Wednesdays and Thursdays) for each year. The time of the year and the day are indicated in normalized time, which is the original time interval scaled to $[0, 1]$. It is seen that the diurnal fluctuations are approximately twice the size of the annual fluctuations. In the period 1974 to 1979 the evening peak changes smoothly, whereas for 1980 to 1994 the evening peak disappears quickly in the spring and reappears quickly in the autumn. It is noted that in Denmark summer time from last Sunday in March until last Sunday in September was reinforced in 1980.

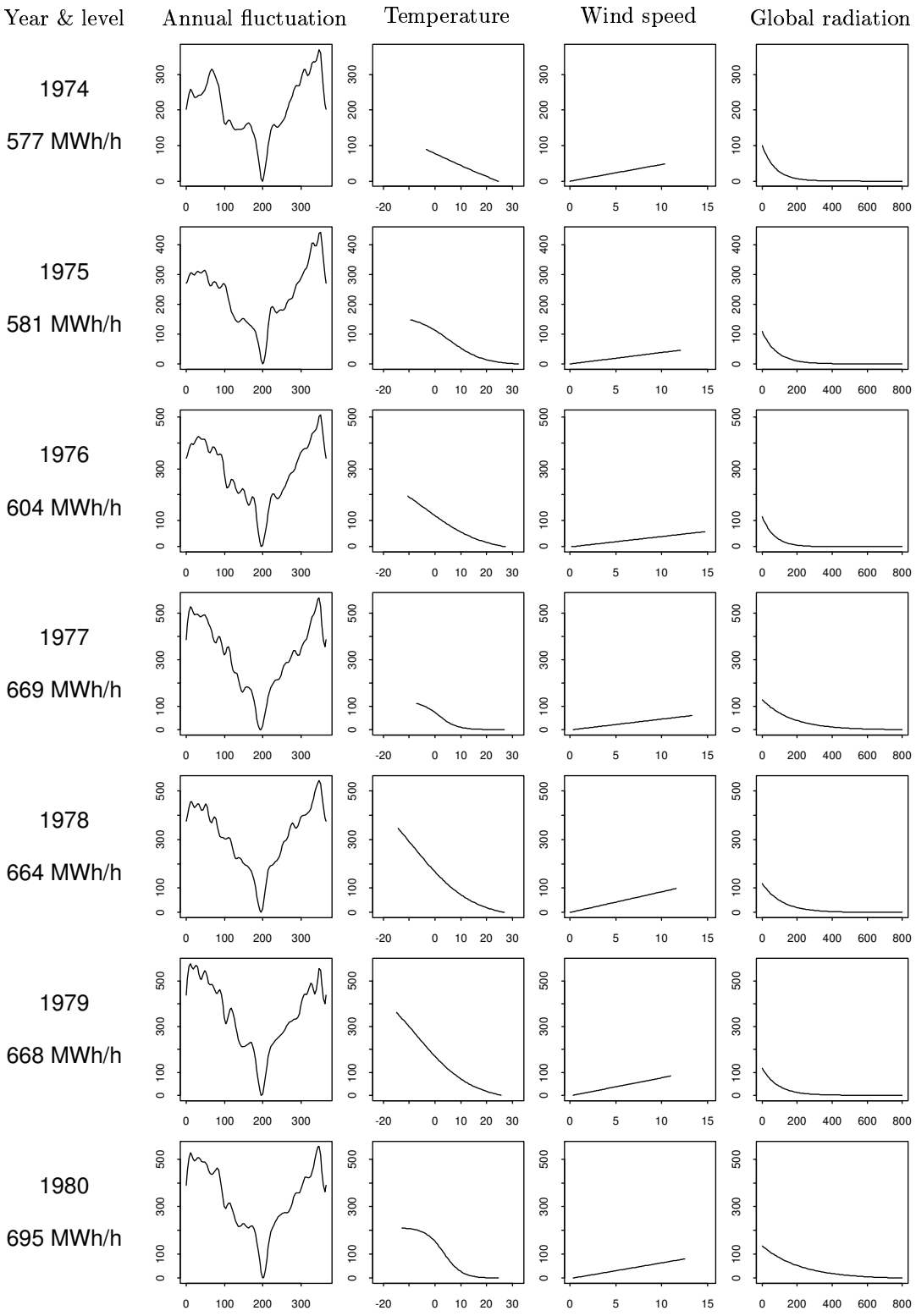


Figure 7.6. *Estimated annual fluctuation and climate response for the years 1974 – 1980 (model (7.6)). From left to right: Year and level, annual fluctuation vs. day in year, response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

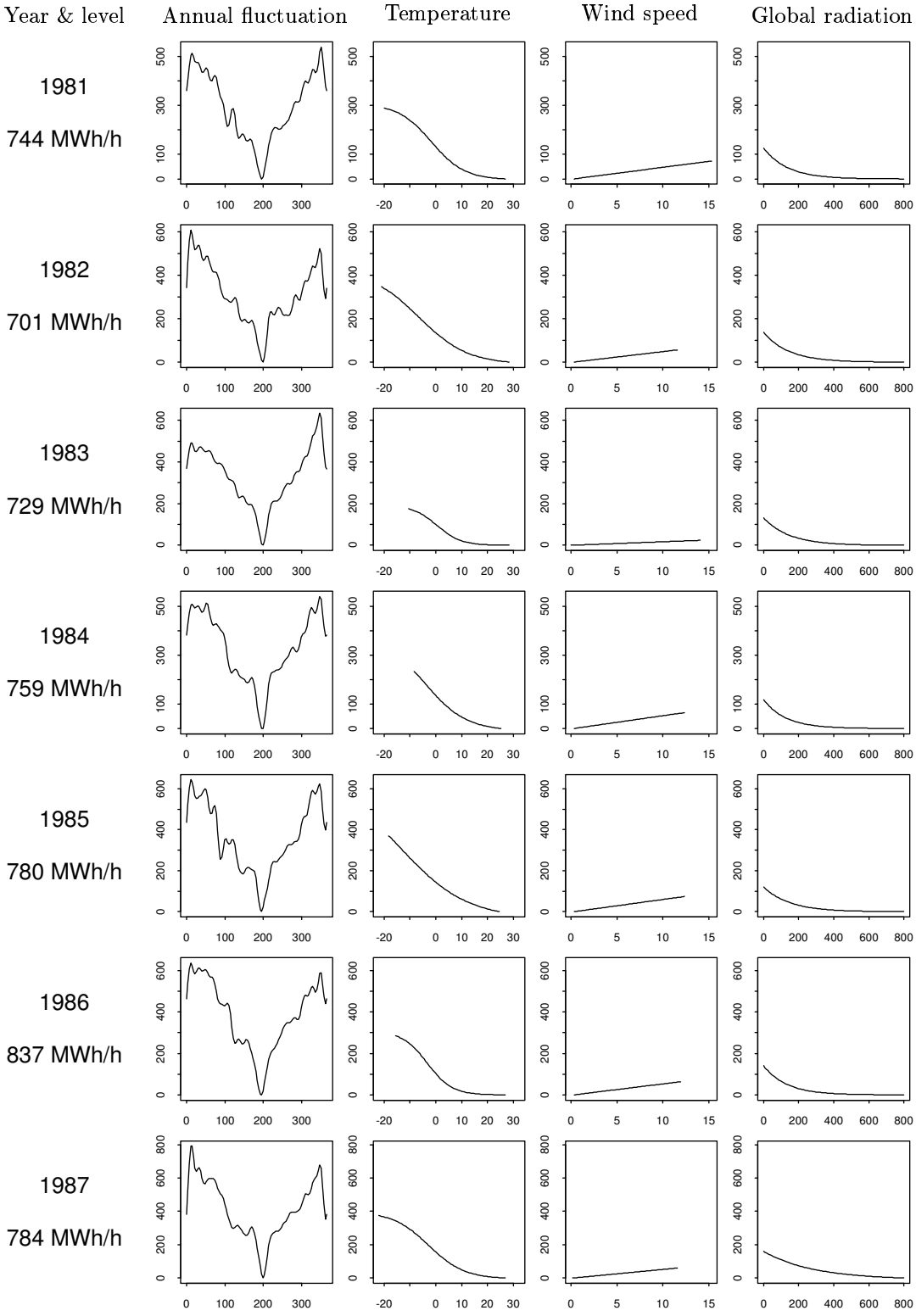


Figure 7.7. *Estimated annual fluctuation and climate response for the years 1981 – 1987 (model (7.6)). From left to right: Year and level, annual fluctuation vs. day in year, response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

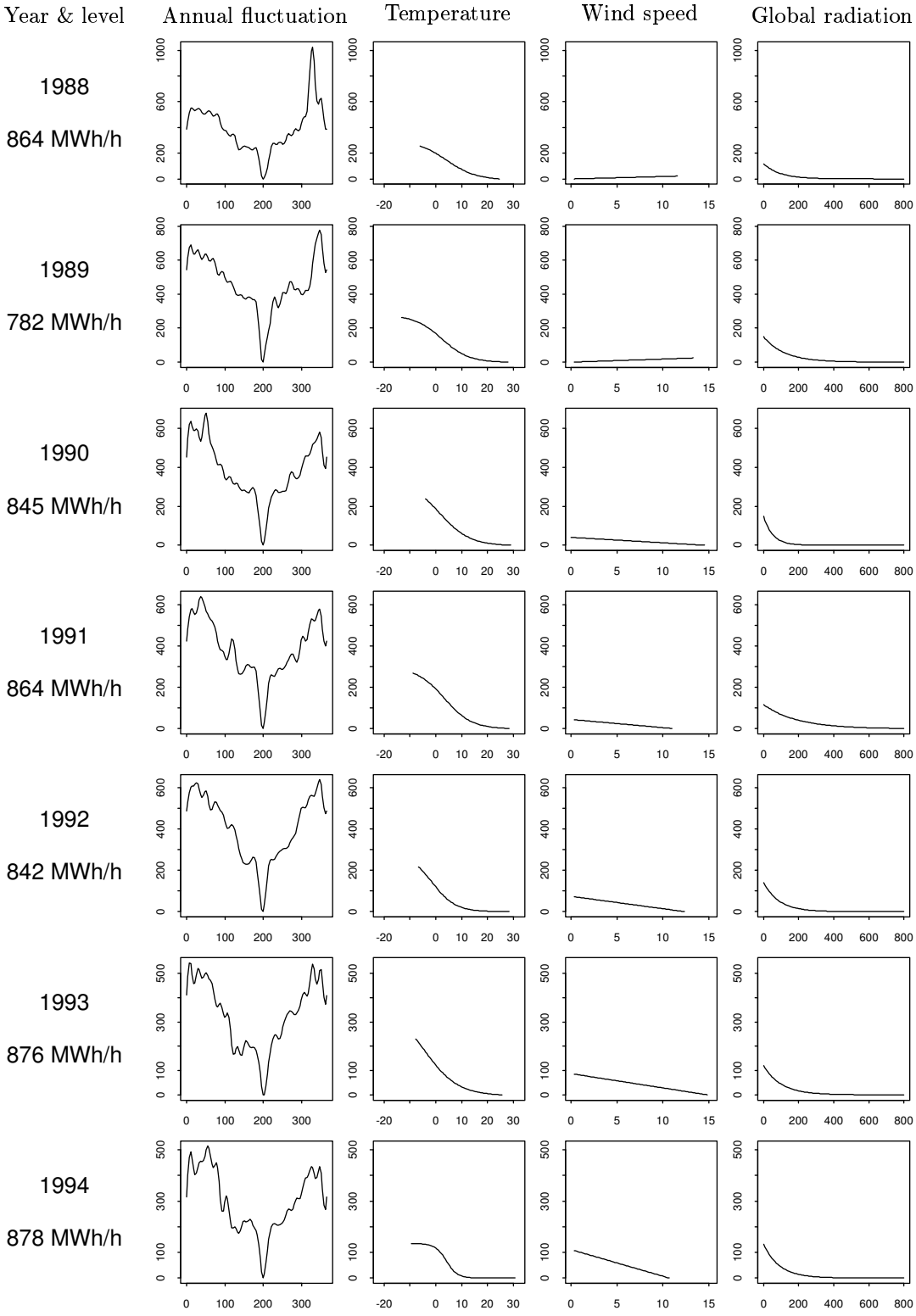


Figure 7.8. *Estimated annual fluctuation and climate response for the years 1988 – 1994 (model (7.6)). From left to right: Year and level, annual fluctuation vs. day in year, response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

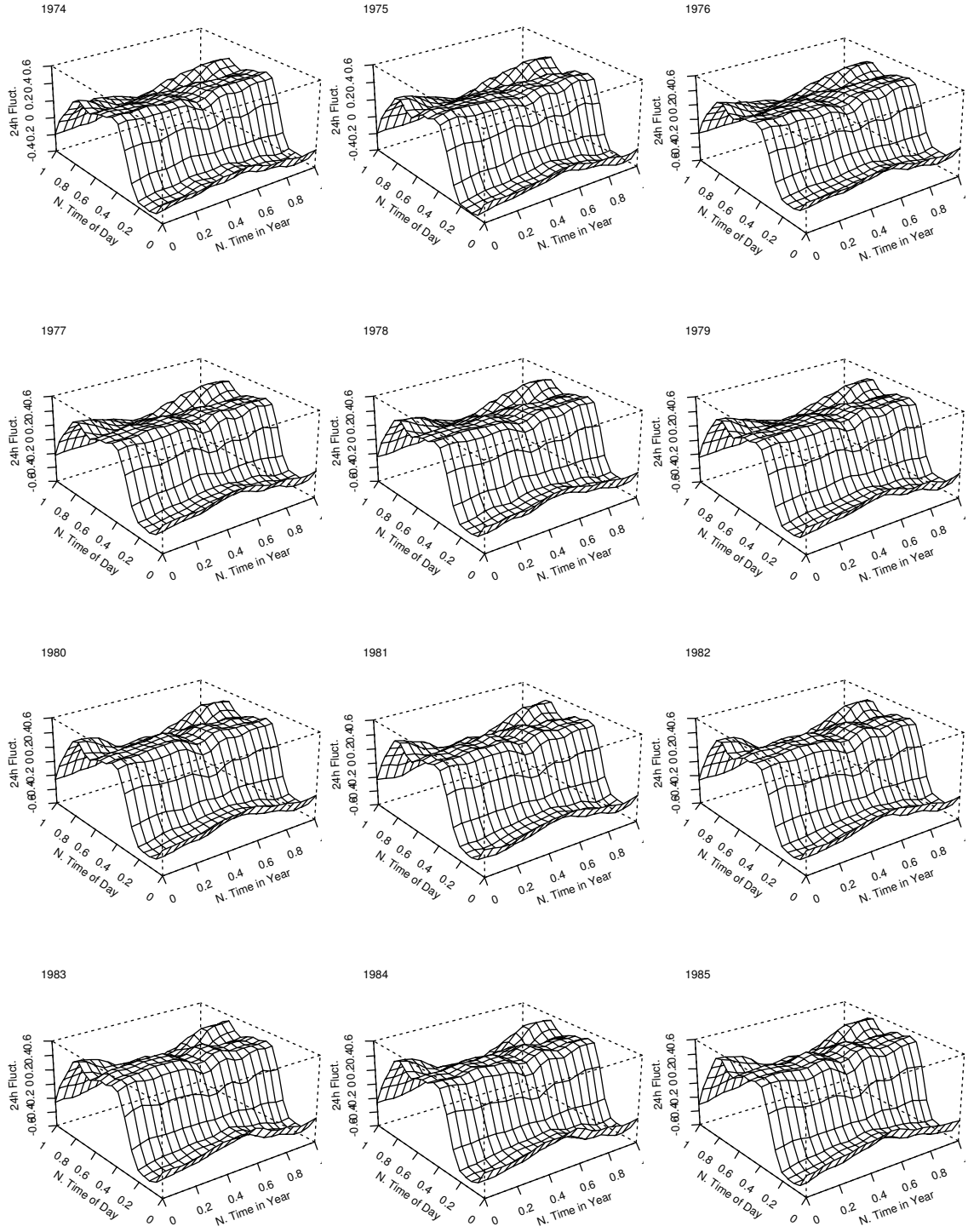


Figure 7.9. *Estimated diurnal fluctuation for Midweeks for the years 1974 – 1985 (model (7.6)). The unit is GWh/h.*

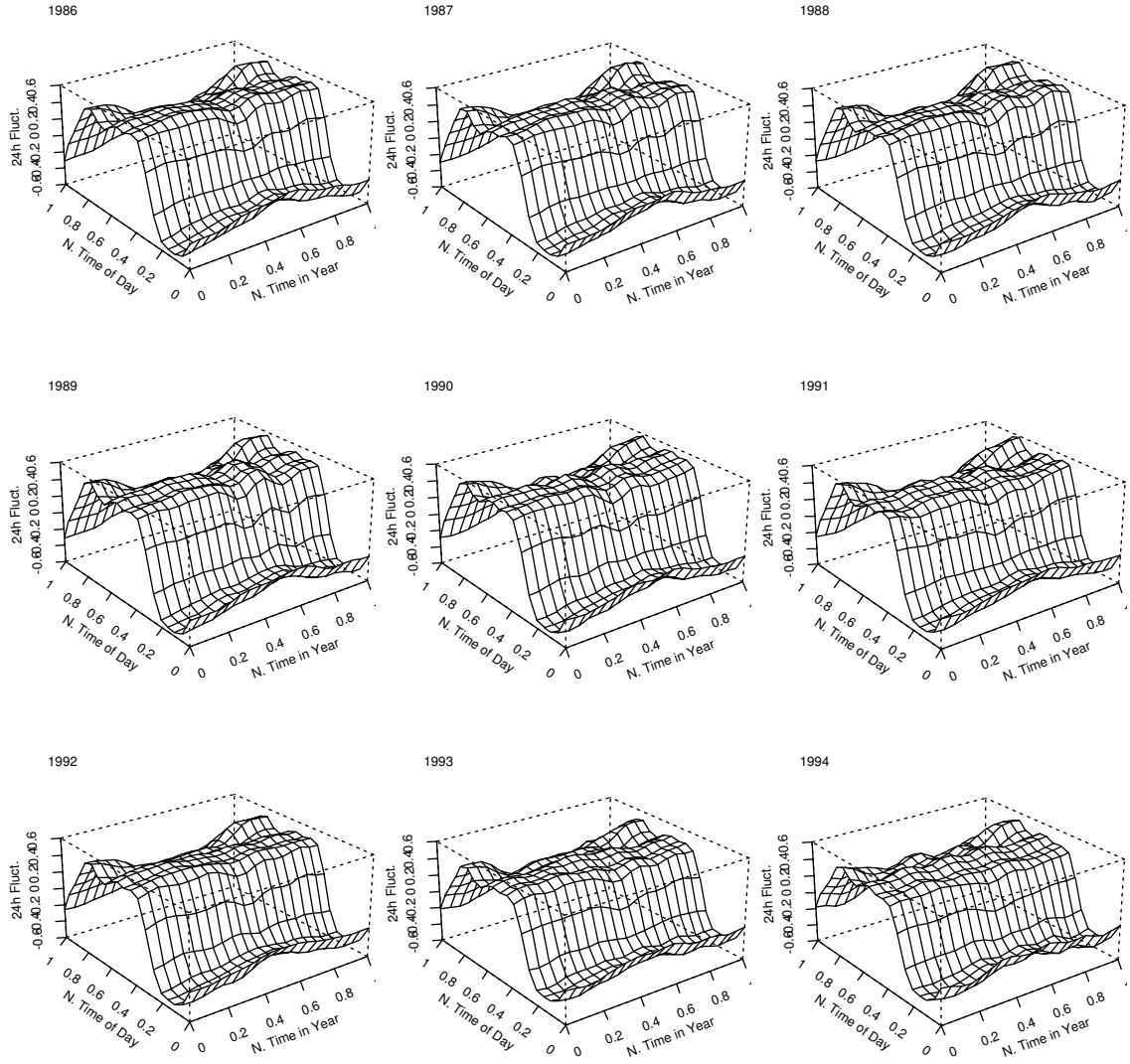


Figure 7.10. *Estimated diurnal fluctuation for Midweeks for the years 1986 – 1994 (model (7.6)). The unit is GWh/h.*

7.5 Dynamic Response from Air Temperature

Among the three climate variables the air temperature is the most important, see Section 7.4. The air temperature is expected to exhibit a dynamic response on the power consumption. This is due to the heat capacity of the buildings. For this reason the focus in this section is on dynamic response on the air temperature.

The most simple extension of the best model of Section 7.4, i.e. model (7.6), is to filter the time series of air temperature.

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; \text{1year}, 18) \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; \text{24h}, 5; F_{iy}(t; \text{1year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(a_{T1}(T_f(t) - a_{T0}))} + a_W W(t) + a_{R2} \exp(-a_{R1} R_g(t)), \quad (7.7) \\ T_f(t) &= \sum_{i=0}^{26} b_i T_a(t-i), \\ b_0 &= 1 - \sum_{i=1}^{26} b_i, \end{aligned}$$

where b_i is the filter parameters. The restriction implies that the stationary gain of the filter is one and hence the filtered value may be interpreted as an air temperature. The order of the filter has been set to 26 in order to be able to test the 24 hour lag, which has been found in (Thorsted, 1993).

The parameters of the model is estimated using OLS based on data for 1982 only. The model has 663 parameters and a R^2 value of 0.9881 is obtained. The t-test statistics of the b_i parameters were investigated although the residuals do not indicate white noise. In Figure 7.11 the b_i parameters are shown together with approximate 95% confidence intervals.

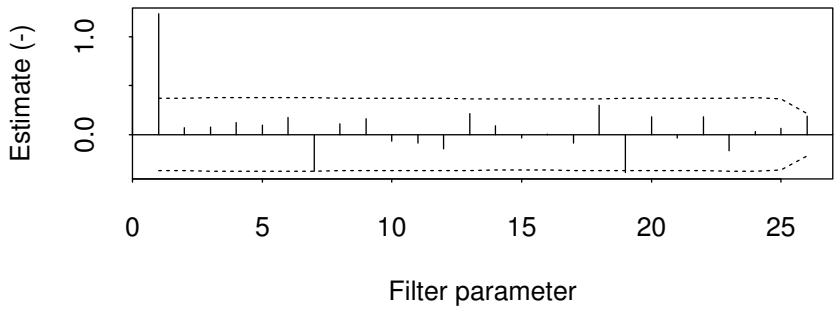


Figure 7.11. *Estimated filter parameters and approximate 95% confidence intervals for model (7.7), under the assumption that the true values are zero and the model is correct. Data from 1982.*

Only the parameter b_1 is significant. The last parameter (b_{26}) are almost significant. Since the intermediate parameters are non-significant, the significance of b_{26} does

not have any physical interpretation. In order to investigate if the significance is an artifact the order of the filter was increased to 32. Furthermore, the restriction on the filter parameters was removed since the computer time required to estimate the parameters of model (7.7) was rather high. This is only a technical aspect, which however inhibits the direct interpretation of the filtered temperature. Since b_0 is estimated directly the standard error of this estimate is directly accessible.

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; 1\text{year}, 18) \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(T_f(t) - a_{T0})} + a_W W(t) + a_{R2} \exp(-a_{R1} R_g(t)), \qquad (7.8) \\ T_f(t) &= \sum_{i=0}^{32} b_i T_a(t-i). \end{aligned}$$

R^2 values for this model are displayed in Table 7.4.

Year	R^2	Year	R^2	Year	R^2
1974	0.9837	1981	0.9889	1988	0.9898
1975	0.9875	1982	0.9881	1989	0.9805
1976	0.9874	1983	0.9893	1990	0.9871
1977	0.9884	1984	0.9894	1991	0.9875
1978	0.9886	1985	0.9887	1992	0.9877
1979	0.9892	1986	0.9898	1993	0.9829
1980	0.9890	1987	0.9900	1994	0.9794

Table 7.4. *R^2 for model (7.8). The model has 669 parameters.*

The estimates of the b_i parameters are shown together with approximate 95% confidence intervals in Figure 7.12. It is seen that the previously observed significance of b_{26} was an artifact. The reason why the confidence interval of the last parameter estimate is smaller than the remaining is probably that the parameter estimate acts as an overall adjustment of the stationary gain.

The estimates of b_0 and b_1 are of opposite sign. Since $\hat{a}_{T2} > 0$, $\hat{b}_0 < 0$ and $\hat{b}_1 > 0$ the following holds under the model. If the air temperature decreases at time t so does the power consumption at time t. However, at time $t+1$ the power consumption increases. This is seen in Figure 7.13 which shows the response on a step from 0 to -1 ^oC at $t = 0$. The stationary increase in power consumption is approximately 15 MWh/h, whereas the static model (7.6) predicts approximately 10 MWh/h. The dynamic response at $t = 0$ is rather large and of same sign as the step. For $t \geq 1$ the response is more in correspondence with what is expected from physical reasons. Possible explanations for the response at $t = 1$ includes:

- The temperature controllers in the buildings acts on a smaller time scale than the sampling time of one hour.

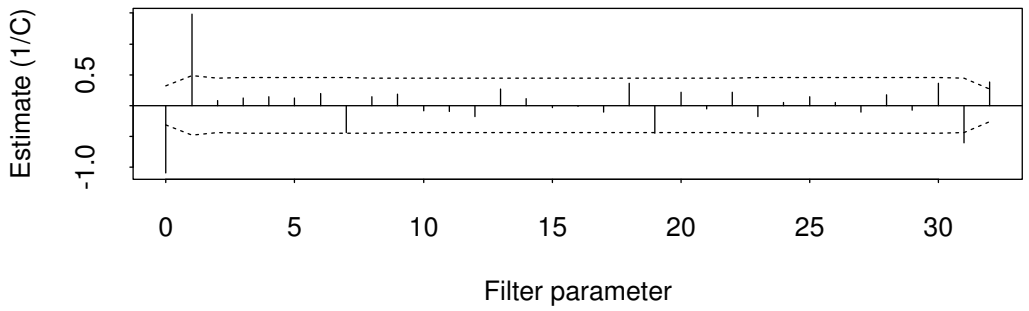


Figure 7.12. *Estimated filter parameters and approximate 95% confidence intervals for model (7.8), under the assumption that the true values are zero and the model is correct. Data from 1982.*

- The air temperature time series does not resemble steps and hence using the model to predict step responses for the system corresponds to extrapolating beyond the area in which the model has been calibrated.

To investigate the second possibility further the response was calculated in case of a more realistic input. Figure 7.14 shows the response on a ramp from 0 to -8 $°C$ over eight hours. It is seen that the ramp response is much more credible than the step response in that the decrease in the beginning is small and may actually correspond to no change. If this is true the system just waits four or five hours before it starts to respond to the ramp. On Figure 7.15 the response on a single harmonic with period 24h is displayed. Generally the response is shaped, as expected, and lacking four hours behind the input.

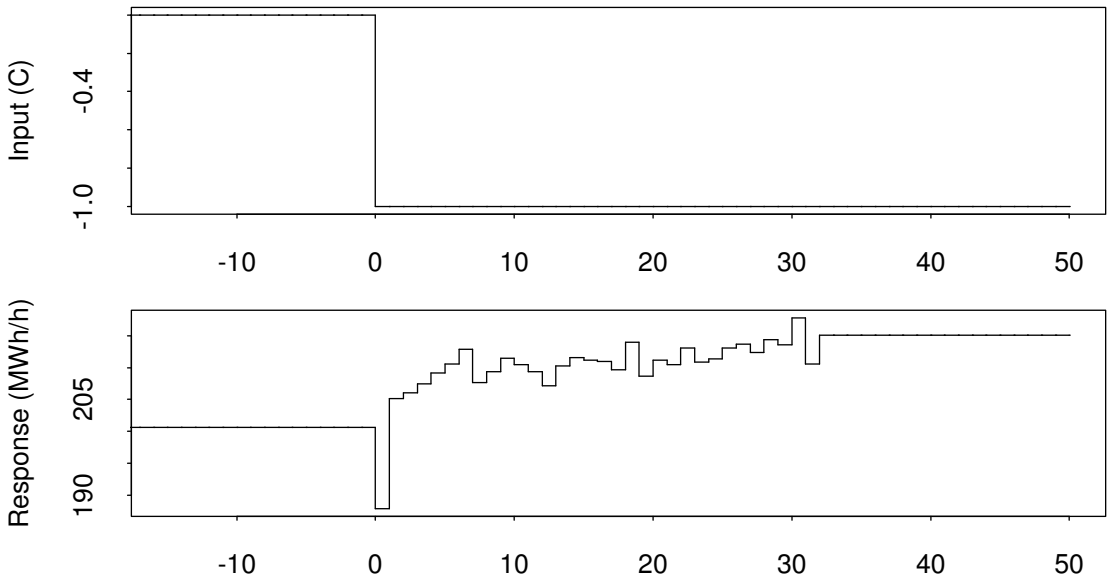


Figure 7.13. *Step response on air temperature under model (7.8) fitted to data from 1982.*

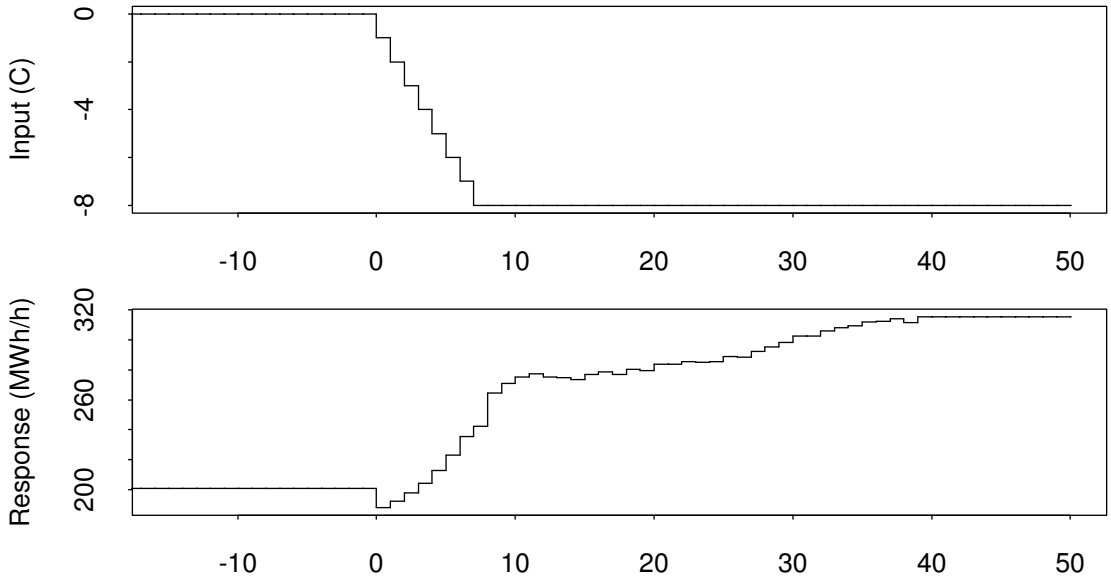


Figure 7.14. *Ramp response on air temperature under model (7.8) fitted to data from 1982.*

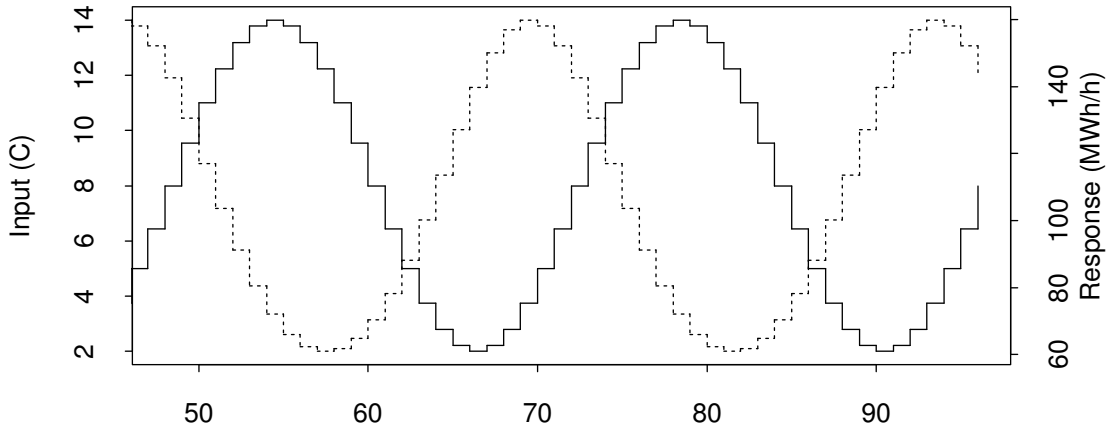


Figure 7.15. *Response on single harmonic air temperature with period 24h under model (7.8) fitted to data from 1982. Input "—", response "- - -".*

From Figure 7.12 it is tempting to conclude that only lags 0 and 1 should be included in the filter of $T_a(t)$. Although the ramp and sine responses indicate that at least part of the remaining lags are important a model like (7.8), but with the $T_a(t)$ filter replaced by $b_0T_a(t) + b_1T_a(t-1)$, is fitted to the data with the same procedure as previously. In Table 7.5 the No. of iterations used are displayed. Generally the No. of iterations used is increased when the restricted filter is used. In some cases the minimization procedure even fails to converge. This further indicates that more lags than 0 and 1 should be included in the filter.

Year	(a)	(b)	Year	(a)	(b)	Year	(a)	(b)
1974	10	34	1981	8	8	1988	8	10
1975	17	8	1982	9	11	1989	10	(50)
1976	9	9	1983	8	21	1990	9	(23)
1977	7	15	1984	7	19	1991	9	8
1978	8	9	1985	12	25	1992	12	43
1979	9	11	1986	9	14	1993	7	26
1980	9	8	1987	9	8	1994	8	(4)

Table 7.5. *Number of iterations used to estimate the parameters of model (7.8) with filter (a)* $b_0T_a(t) + \ldots + b_{32}T_a(t-32)$ *and (b)* $b_0T_a(t) + b_1T_a(t-1)$. *() indicates that the itertations failed to converge. The limit on the number of iterations is 50.*

Figures 7.16, 7.17 and 7.18 show the estimated annual fluctuation and climate response for all years. Note that within each year the y-axes are identical and the minima are set to zero. Furthermore the response on the global radiation is only displayed for values below 800 W/m^2, although values as high as 1005 W/m^2 occur. The level stated below the year is the level obtained when the four functions are at their minimum. The functions are only plotted in the range for which observations exists. Comparing these plots to the corresponding plots for model (7.6) shows that for model (7.8) the stationary response on the temperature compared to the annual fluctuation is generally larger for (7.8). This observation verifies that the dynamic description of the temperature response is important.

In Appendix A.1 time-plots of the residuals from model (7.8) are shown. Superimposed on the plots are a 2nd order local regression smoothing with a nearest neighbour bandwidth of 0.05 calculated using the `loess()` function of S-PLUS, see (Statistical Sciences, 1995). Serial correlation is seen. Large residuals are observed at the end of each year. Furthermore the loess smooth indicate some fluctuations especially in the spring and autumn. The period of these fluctuations is rather constant. The fluctuations may be an artifact caused by the high order Fourier expansion in that the high frequencies adapts to the low consumption during a part of the summer. Since the model is fitted globally this will affect the Fourier expansion for the rest of the year.

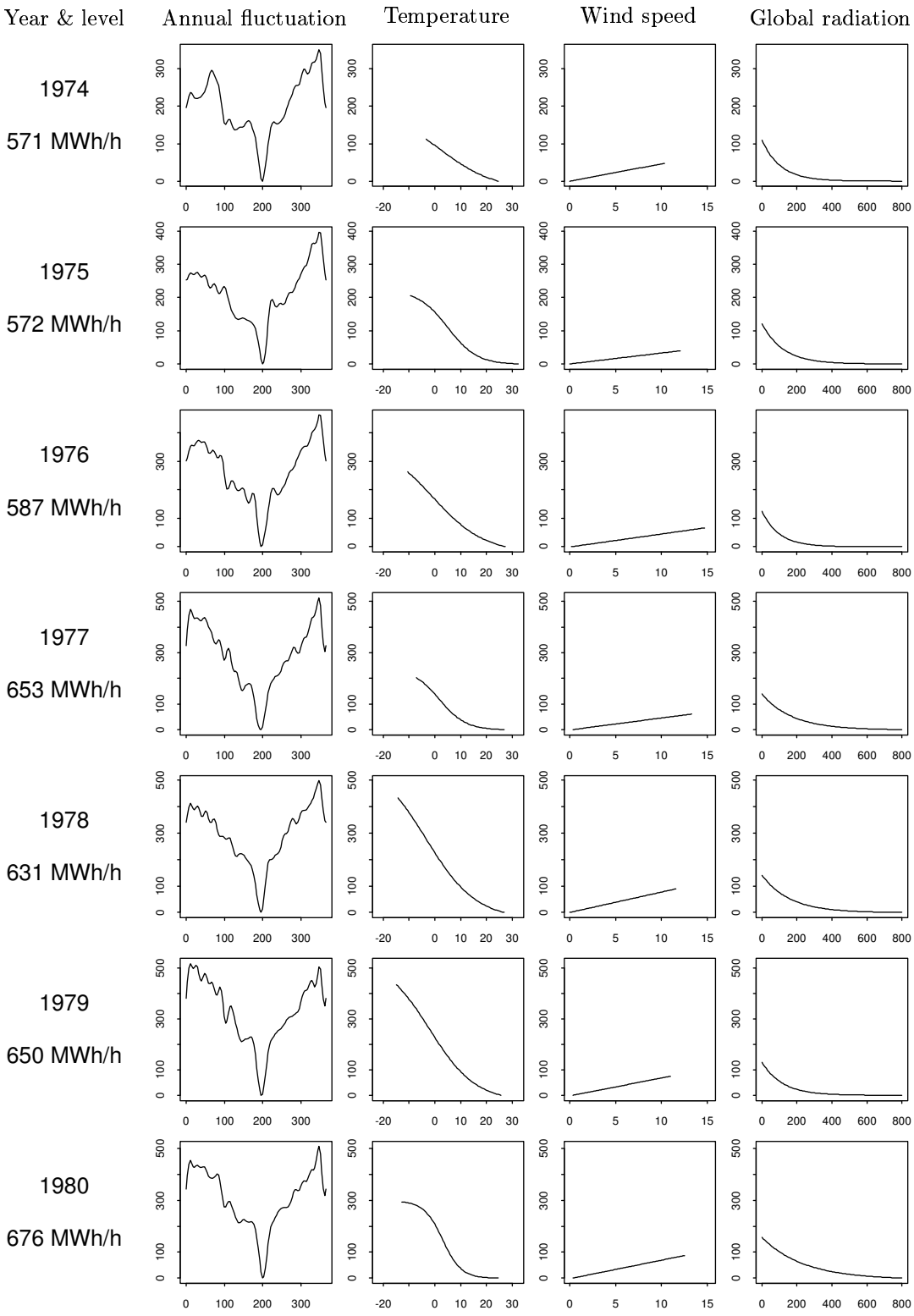


Figure 7.16. *Estimated annual fluctuation and climate response for the years 1974 – 1980 (model (7.8)). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

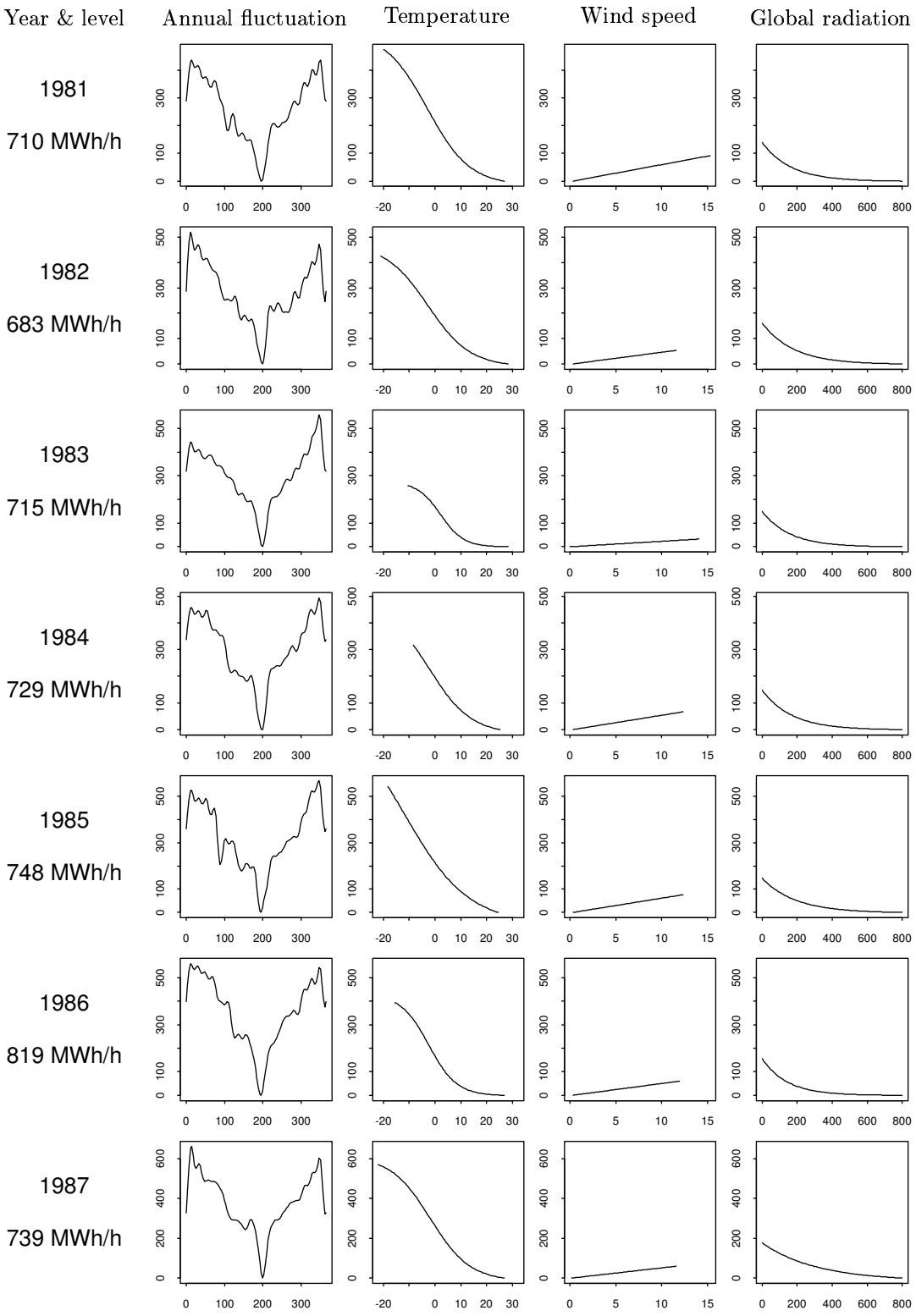


Figure 7.17. *Estimated annual fluctuation and climate response for the years 1981 – 1987 (model (7.8)). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

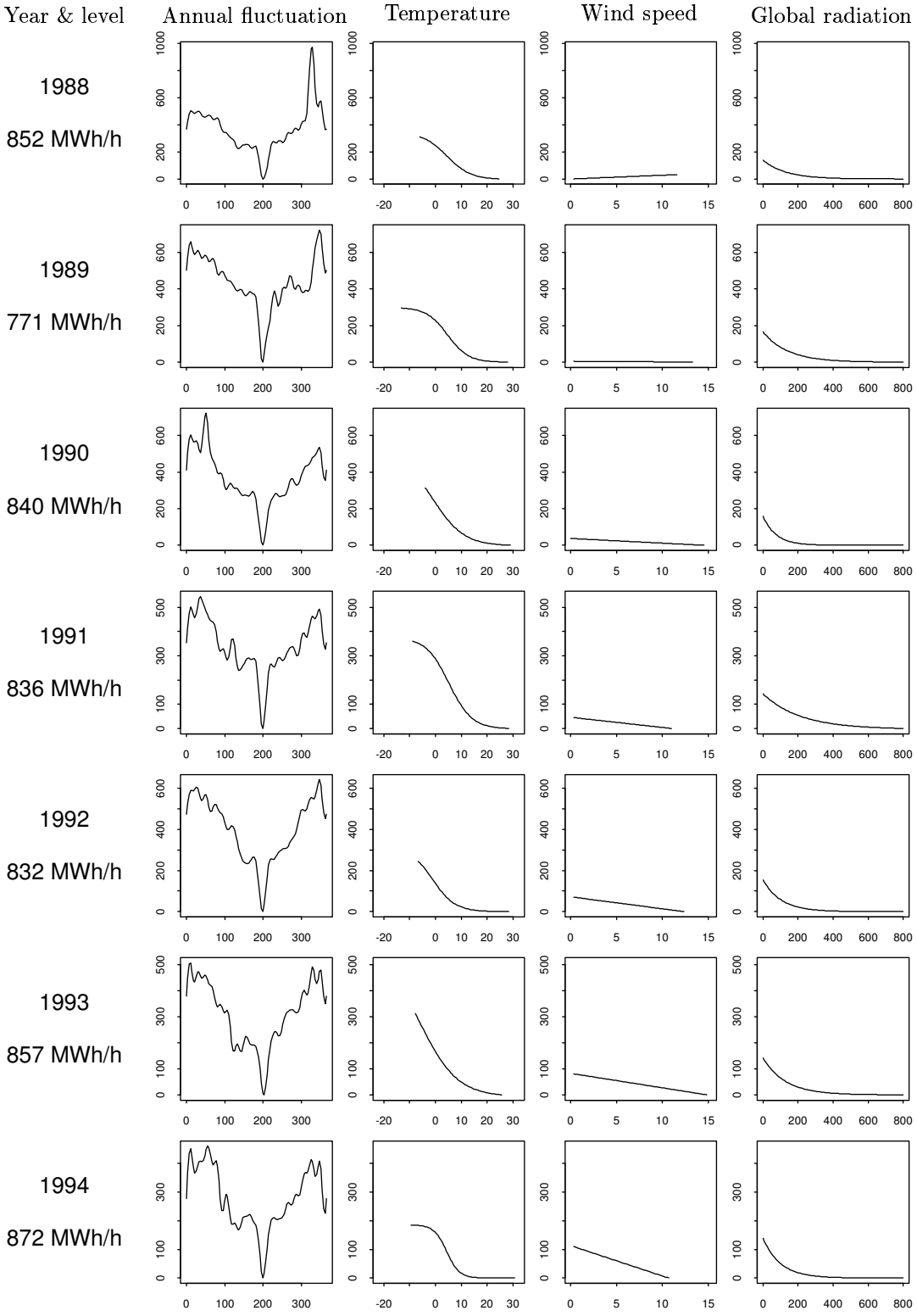


Figure 7.18. *Estimated annual fluctuation and climate response for the years 1988 – 1994 (model (7.8)). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

7.6 A Simplification of the Annual Fluctuation

The annual fluctuation estimated under model (7.6) and (7.8) may vary rather fast due to the high order of the Fourier expansion used. Figure 7.19 shows all the estimated annual fluctuations under model (7.8). It is seen that, except for short periods around Christmas and during the summer, the fluctuations on average follows a low order harmonic. In Figure 7.20 the result when fitting Fourier expansions of order 1, 2 and 3 to the estimates are shown. It is seen that a single harmonic describes the average quite well.

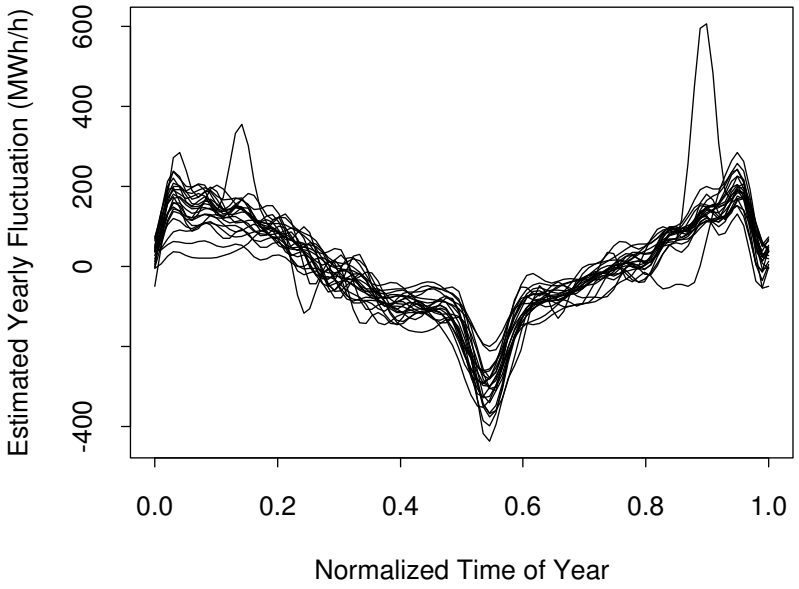


Figure 7.19. *Estimated annual fluctuations under model (7.8) for all years. The normalized time of year is the time of the year scaled to the interval* $[0, 1]$.

In order to force the estimated annual fluctuations into the shape of the average shown in Figure 7.20 the annual fluctuation is parameterized as the sum of a single harmonic, a sine correction for the summer period ranging from 0.49 to 0.6 in units of the normalized time of year, and a shift during the Christmas period. Different Christmas periods were investigated, but based on time-plots of the residuals, the period 24th December to 1st January was chosen. Furthermore days in the Christmas period classified as working days are reclassified as Saturdays / Half-Holy days:

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; 1\text{year}, 1) + I_{\{summer\}}(t) c_s \sin \frac{\pi h_y^n(t)}{0.6 - 0.49} + I_{\{christmas\}}(t)\mu_c \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(T_f(t) - a_{T0})} + a_W W(t) + a_{R2} \exp(-a_{R1} R_g(t)), \qquad (7.9) \\ T_f(t) &= \sum_{i=0}^{32} b_i T_a(t - i), \end{aligned}$$

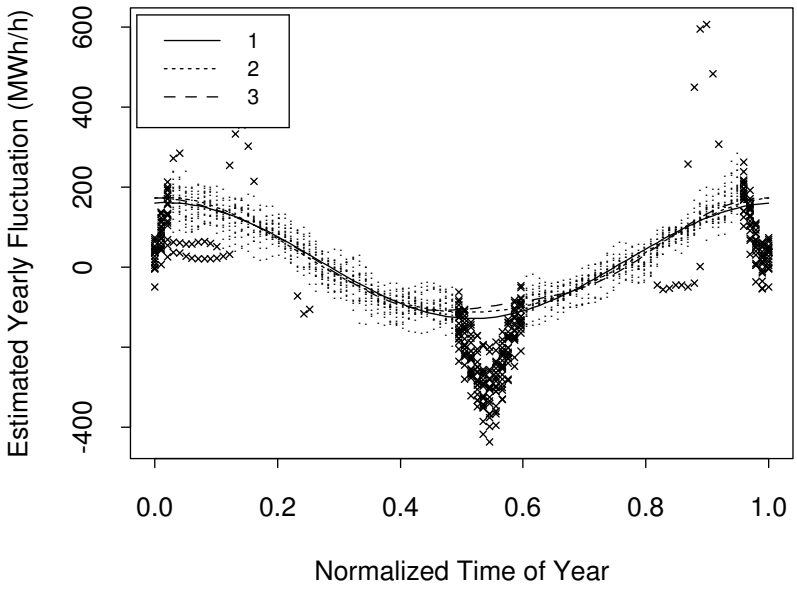


Figure 7.20. *Estimated annual profiles from Figure 7.19 approximated by Fourier expansions of order 1, 2 and 3. The points marked by "×" are excluded. The normalized time of year is the time of the year scaled to the interval* $[0, 1]$.

Where $h_y^n(t)$ is the normalized time of year, $I_{\{summer\}}(t)$ is the indicator function for the summer period, i.e. the function is zero unless $0.49 \leq h_y^n(t) \leq 0.6$, and $I_{\{christmas\}}(t)$ is the indicator function for the Christmas period (24th December – 1st January).

The parameters of model (7.9) are estimated using OLS separately for each year. The estimation failed for 1989 in that the model was found to be singular (note that 1989 also proved to be problematic in Section 7.2). This is probably due to a local minimum reached during the iterations. The R^2 values for this model are displayed in Table 7.6. Compared to the models (7.6) and (7.8) the R^2 values are lower for model (7.9).

Year	R^2	Year	R^2	Year	R^2
1974	0.9736	1981	0.9857	1988	0.9870
1975	0.9816	1982	0.9849	1989	(0.9582)
1976	0.9815	1983	0.9857	1990	0.9846
1977	0.9853	1984	0.9877	1991	0.9840
1978	0.9852	1985	0.9824	1992	0.9857
1979	0.9872	1986	0.9864	1993	0.9801
1980	0.9861	1987	0.9859	1994	0.9736

Table 7.6. R^2 *values for model (7.9). The model has 637 parameters. For 1989 the model is (locally) singular and the estimates are biased.*

Figures 7.21, 7.22 and 7.23 show the estimated annual fluctuation and climate response for all years. Note that within each year the y-axes are identical and the

minima are set to zero. Furthermore, the response on the global radiation is only displayed for values below 800 W/m^2, although values as high as 1005 W/m^2 occur. The level stated below the year is the level obtained when the four functions are at their minimum. The functions are only plotted in the range for which observations exists. Comparing these plots to the corresponding plots for the models (7.6) and (7.8) shows that for model (7.9) the climate response and especially the stationary response on the temperature is more stable between years. This is probably due the 18th order Fourier expansion of the annual fluctuation in the models (7.6) and (7.8) which adapts to fluctuations actually caused by the climate.

In Appendix A.2 time-plots of the residuals from model (7.9) are shown. Superimposed on the plots are a 2nd order local regression smoothing with a nearest neighbour bandwidth of 0.05 calculated using the `loess()` function of S-PLUS, see (Statistical Sciences, 1995). It is clear from the plots that the residuals are not white noise. From the plots it is seen that the parametrization of the annual fluctuation is probably too restrictive or that other explanatory variables are important. For long term predictions, however, the annual fluctuation of model (7.9) is quite appropiate since the fluctuations of the residuals in Appendix A.2 are unexplainable in the context of this report, see the comment on residuals and climate variables below. However, it would be desirable to estimate the annual fluctuation of model (7.9) by local regression with a user-defined bandwidth varying over the year (low at Christmas and during the summer). At pressent software which can perform this task is not accessible.

When the residuals are plotted against the time of day it is seen that some systematic variation is still present. This is not surprising when Figure 6.1 is considered. However, as mentioned in Chapter 6, a Fourier expansion of order nine must be used to explain all systematic diurnal variation. Since the coefficients of the Fourier expansion of the diurnal variation are modelled as Fourier expansions with a period of one year, it is crucial to keep the order down and hence the order is kept at five.

When the residuals are plotted against each of the climate variables lagged from zero to five hours (six plots per year) almost no systematic variation are identified. On the raw scatter-plots no systematic variation can be seen. When a 2nd order local regression smoothing with a nearest neighbour bandwidth of 0.3 calculated using the `loess()` function of S-PLUS, see (Statistical Sciences, 1995), is superimposed on the plots (not possible for the global radiation due to the high No. of observations for which this number is 0 W/m^2) only very week dependencies are seen. The logistic function describing the dependence of the power consumption on the air temperature may not be quite appropriate, but within the family of globally parametric models its probably a good choice. However, some artifacts may be introduced by the logistic function, in that it is symmetric around the point $(a_{T0}/\sum b_i, a_{T2}/2)$. The similar double exponential function $a\exp(-b\exp(cx))$ does not share this symmetry property, but, the value of the maximum and minimum curvature (the extremes of the second order derivative) are functions of each other. For this reason the double exponential function does not solve the problem. One solution would be

to estimate the static relation between air temperature and power consumption by local regression techniques, see (Hastie and Tibshirani, 1990). However, the filtering of the air temperature poses a theoretical problem.

Also a model like (7.9) but with $c_s \sin \frac{\pi h_y^n(t)}{0.6-0.49}$ replaced by $\mu_s + F(t; (0.6-0.49)year, 1)$ and with μ_c replaced by $\mu_c + F(t, 1.2week, 1)$ is fitted separately for each year. Furthermore days of type "normal working days" during the Christmas are not reclassified. When time-plots of the residuals from this model is compared with those for model (7.9) very few differences are observed. The differences observed are especially found around Christmas. Since for some years one model is sightly superior and vice versa there is no need to change model (7.9) in the direction just mentioned.

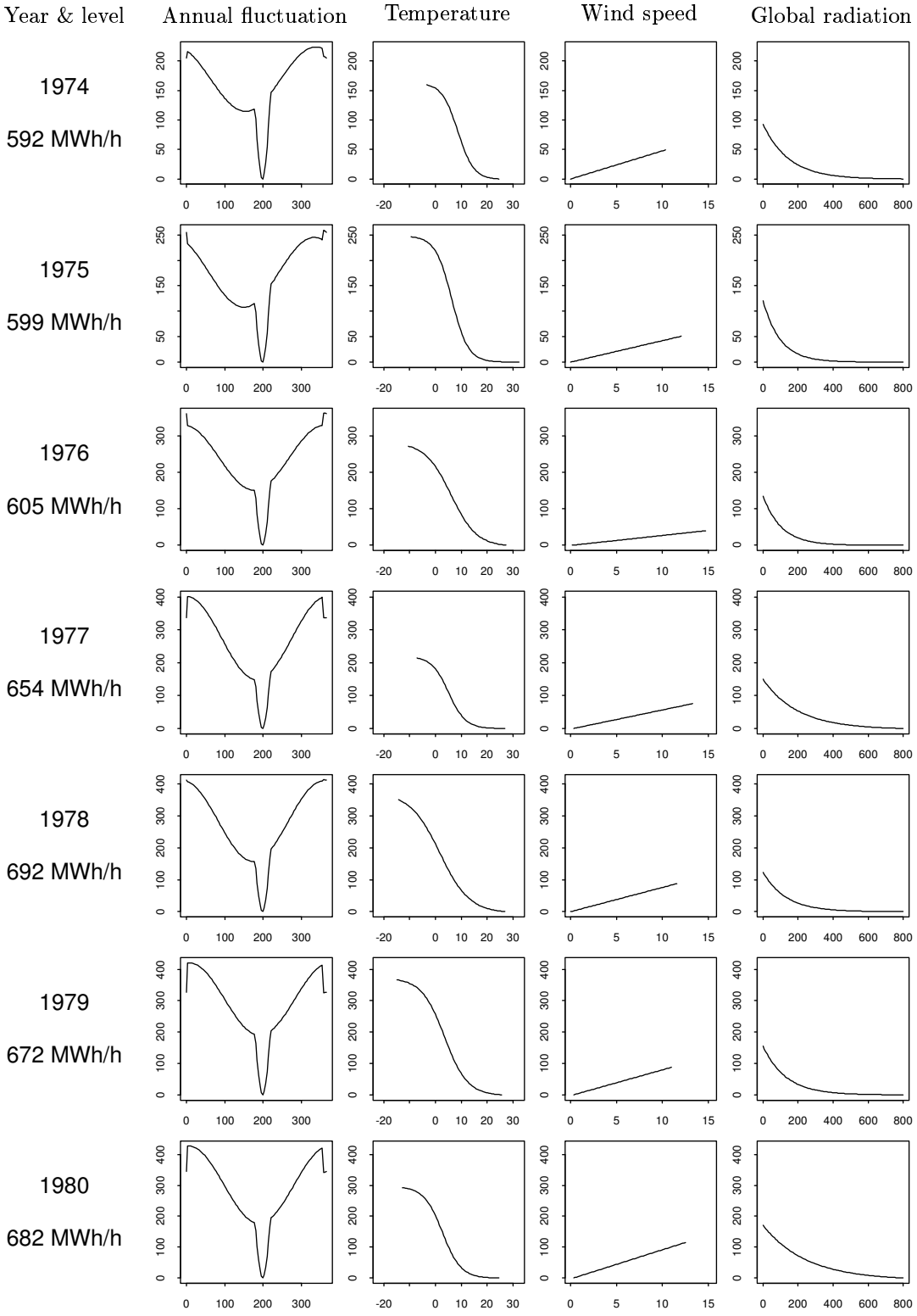


Figure 7.21. *Estimated annual fluctuation and climate response for the years 1974 – 1980 (model (7.9)). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

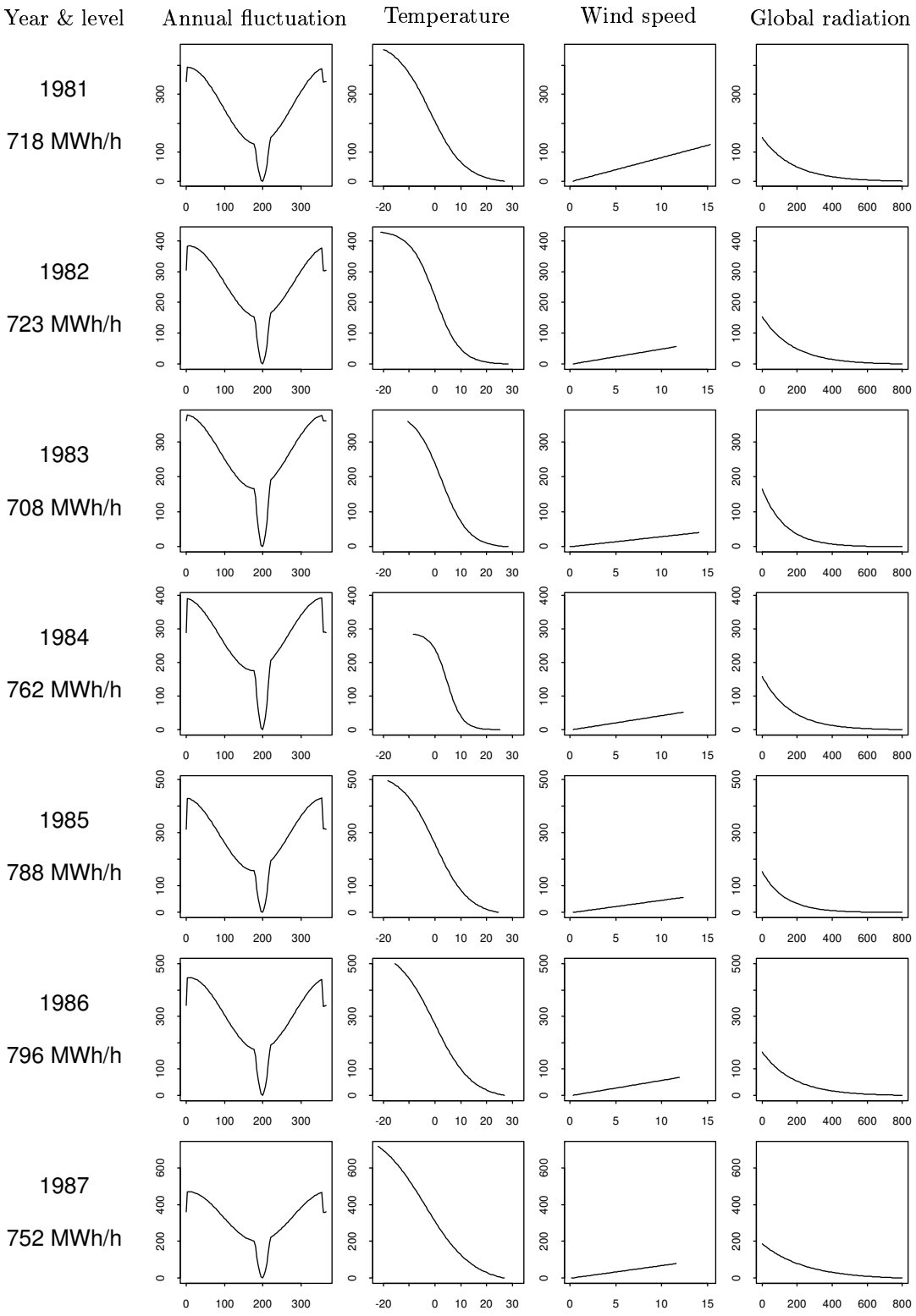


Figure 7.22. *Estimated annual fluctuation and climate response for the years 1981 – 1987 (model (7.9)). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

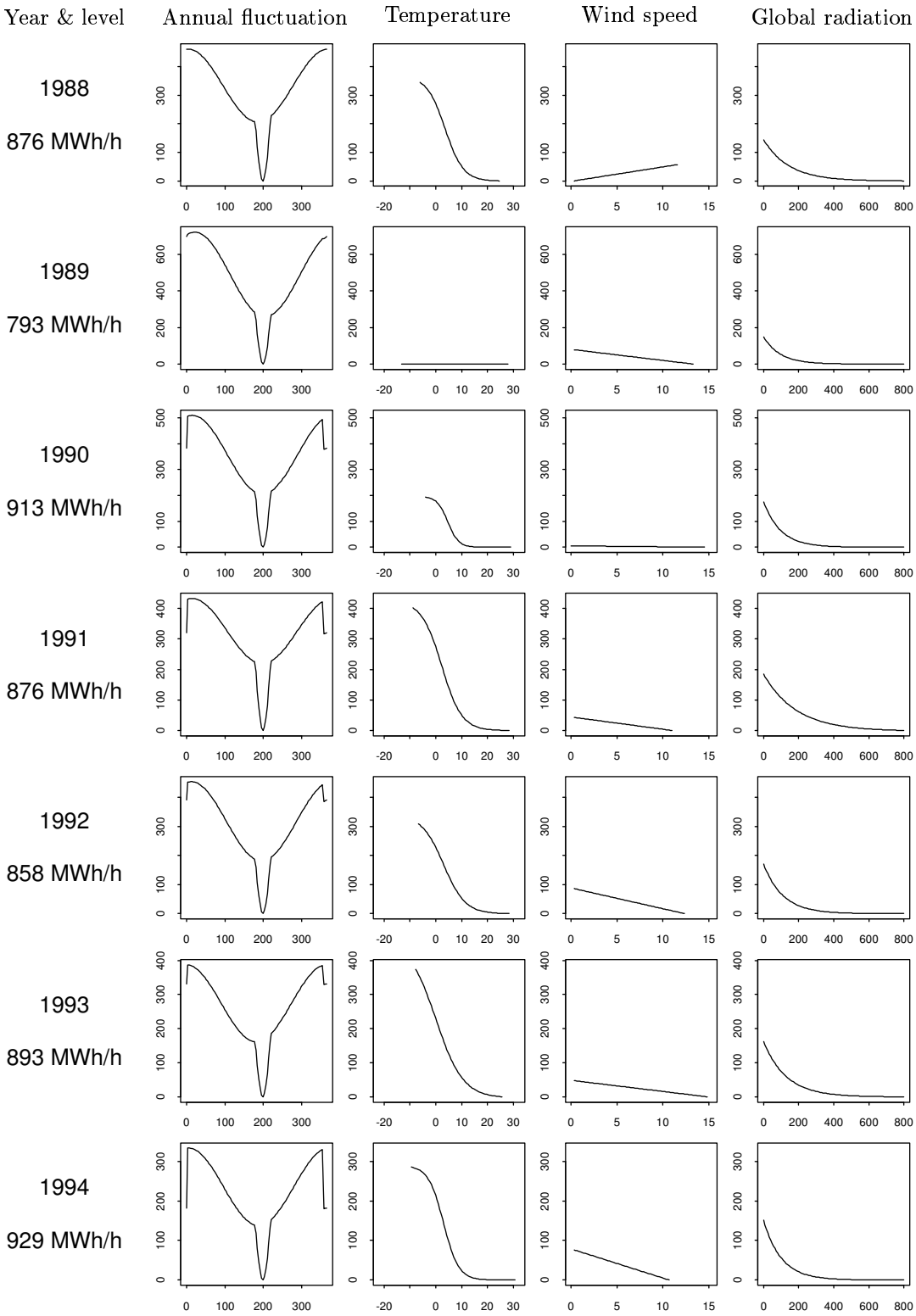


Figure 7.23. *Estimated annual fluctuation and climate response for the years 1988 – 1994 (model (7.9)). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h. For 1989 the model is (locally) singular and the estimates biased.*

7.7 Dynamic Response from Global Radiation and Air Temperature

Although residual plots do not show any dependence on lagged values of the global radiation lagged values may lead to a more adequate description of the system. This is due to the fact that the diurnal fluctuation may explain some of the variation actually due to the global radiation. For this reason the parameters of the following model is estimated with OLS separately for each year.

$$\begin{aligned} E[P(t)] &= \mu_0 + F_y(t; 1\text{year}, 1) + I_{\{summer\}}(t) c_s \sin \frac{\pi h_y^n(t)}{0.6 - 0.49} + I_{\{christmas\}}(t)\mu_c \\ &+ \sum_{i=1}^{5} I_i(t)[\mu_i(t) + F_i(t; 24\text{h}, 5; F_{iy}(t; 1\text{year}, 5))] \\ &+ \frac{a_{T2}}{1 + \exp(T_f(t) - a_{T0})} + a_W W(t) + a_{R2} \exp(-R_f(t)), \qquad (7.10) \\ T_f(t) &= \sum_{i=0}^{32} b_i T_a(t-i), \\ R_f(t) &= \sum_{i=0}^{10} g_i R_g(t-i). \end{aligned}$$

The model is almost exactly like (7.9) including the reclassification of working days during Christmas. The only difference is the filtered value of the global radiation $R_f(t)$, with the filter parameters $g_0, \ldots, g_{10}$. The R^2 values for this model are shown in Table 7.7. Note that also this model is found to be singular in 1989.

Year	R^2	Year	R^2	Year	R^2
1974	0.9743	1981	0.9868	1988	0.9877
1975	0.9826	1982	0.9859	1989	(0.9601)
1976	0.9827	1983	0.9869	1990	0.9856
1977	0.9866	1984	0.9889	1991	0.9855
1978	0.9860	1985	0.9834	1992	0.9865
1979	0.9883	1986	0.9875	1993	0.9811
1980	0.9872	1987	0.9870	1994	0.9749

Table 7.7. *The R^2 values for model (7.10). The model has 647 parameters. For 1989 the model is (locally) singular and the estimates biased.*

Figure 7.24 shows the estimated filter parameters for year 1982. Compared with model (7.8) (Figure 7.12) it is seen that the significance of the air temperature at lags 0 and 1 has disappeared. This must be caused by the introduction of the filtered global radiation. In Appendix B.1 plots like the ones in Figure 7.24 are shown for all years (except for 1989). In almost all cases only the filter parameter related to the largest lag (No. 32) is significant. The parameters of the filter corresponding to the global radiation are strongly significant for the first 3 or 4 lags. Due to the

fact that the global radiation is measured in Taastrup, negative lags might also be considered, this is however omitted from this report.

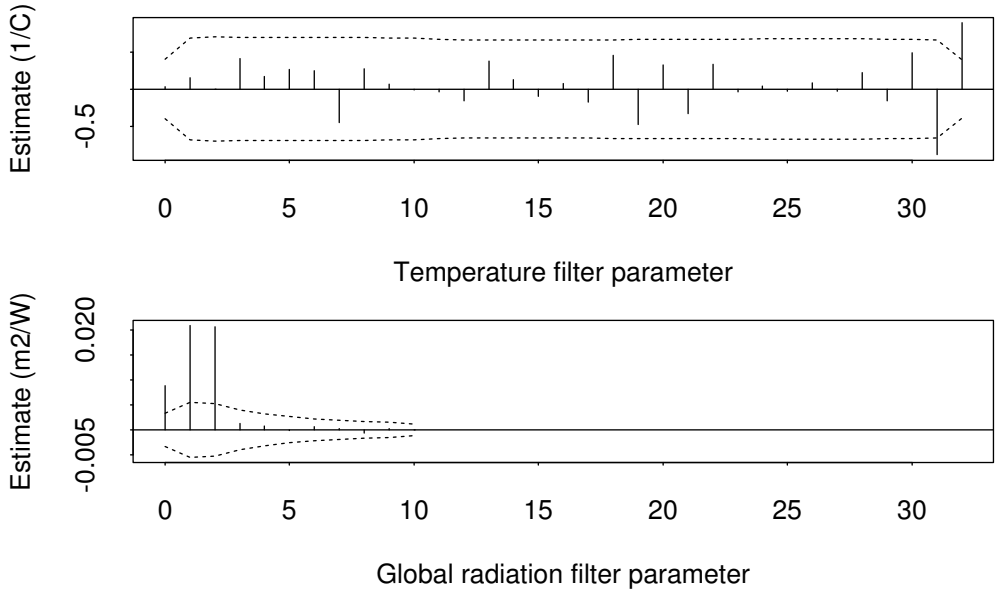


Figure 7.24. *Estimated filter parameters and approximate 95% confidence intervals for model (7.10), under the assumption that the true values are zero and the model is correct. Data from 1982.*

Figures 7.25 and 7.26 show examples of step and ramp responses for the model fitted to data from 1982. In opposition to model (7.9) the step responses are quite plausible from a physical point of view. For the temperature the significance of the last (in this case last two) lags looks strange on the step response[2], whereas the step response on the radiation is more plausible from a physical point of view. The plots in Appendix B.1 indicate that the step responses are similar for the remaining years.

Due to the results stated above it was decided to increase the order of the air temperature filter in model (7.10) from 32 to 75, since linear extrapolation on Figure 7.25 results in an approximate maximum lag of 50. For 1989 the estimation failed to converge. Table 7.8 shows the values of R^2. Compared with Table 7.7 the values have increased slightly.

In Figure 7.27 the step responses estimated based on data from 1982 are shown. Even for the temperature the shape of the response is quite like the one shown in Figure 7.25 as long as the lag is below 32. In Appendix B.2 step responses for all years except 1989 are shown. The response on the global radiation in general resembles that of a 1st order system. The temperature response is more difficult to interpret. For some years it looks like a first order system and for some it looks like a second order system. In general the plots in Appendix B.2 indicates that lags up

[2]As explained previously this phenomena is most likely due to the fact that the last lags explain some of the influence from higher lags, which are not included in the model.

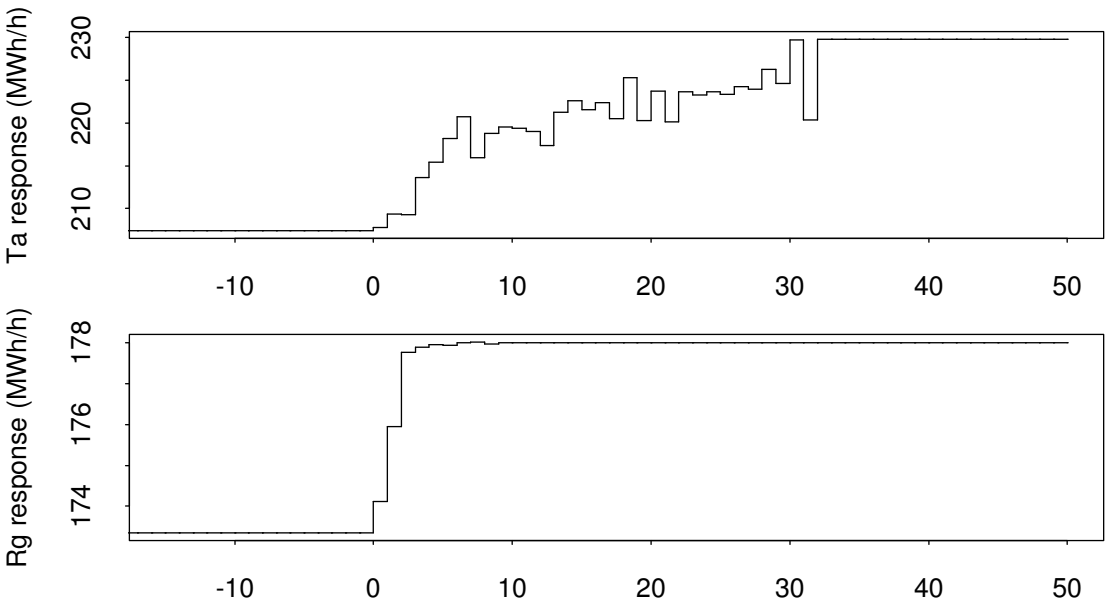


Figure 7.25. *Step response on air temperature and global radiation under model (7.10) fitted to data from 1982. The air temperature drops from 0 to -1 °C at time* $t = 0$. *The global radiation drops from 25 to 20* W/m^2 *at time* $t = 0$.

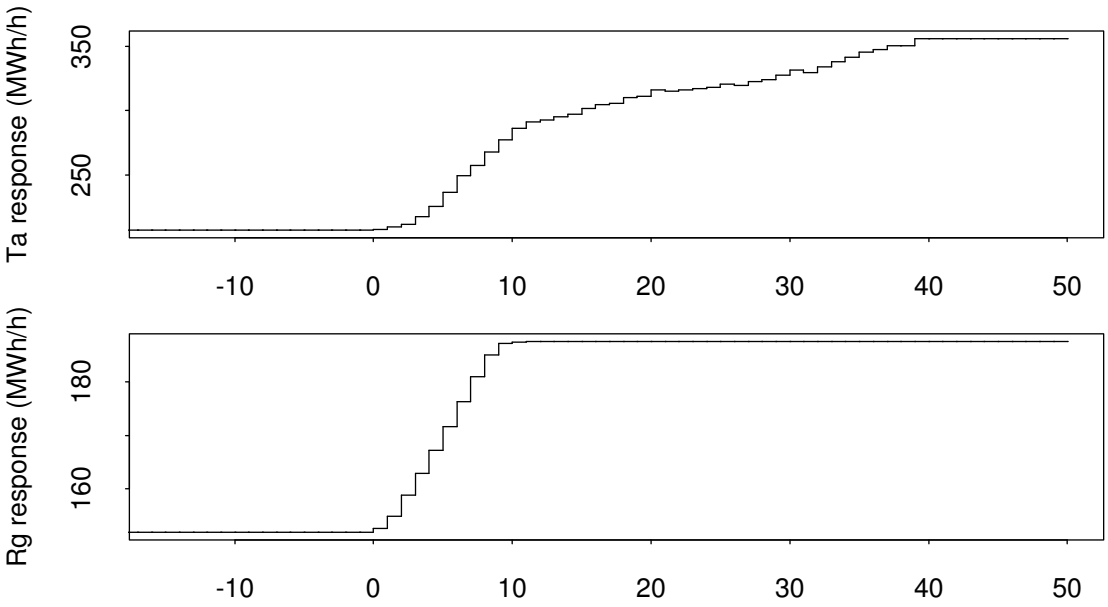


Figure 7.26. *Ramp response on air temperature and global radiation under model (7.10) fitted to data from 1982. The air temperature drops from 0 to -8 °C over 8 hours starting at* $t = 0$. *The global radiation drops from 50 to 10* W/m^2 *over 8 hours starting at* $t = 0$.

Year	R^2	Year	R^2	Year	R^2
1974	0.9759	1981	0.9880	1988	0.9891
1975	0.9838	1982	0.9868	1989	(0.9265)
1976	0.9846	1983	0.9875	1990	0.9857
1977	0.9871	1984	0.9893	1991	0.9861
1978	0.9865	1985	0.9843	1992	0.9865
1979	0.9888	1986	0.9878	1993	0.9828
1980	0.9876	1987	0.9888	1994	0.9761

Table 7.8. *The R^2 values for model (7.10), with the order of the air temperature filter increased to 75. The model has 690 parameters. For 1989 the estimation failed to converge.*

to 75 should be included in the temperature response. The high lag values in the case of the air temperature indicate that a persistent change of the air temperature causes a response in the power consumption which is not fully effective until 3 to 4 days after the change occured.

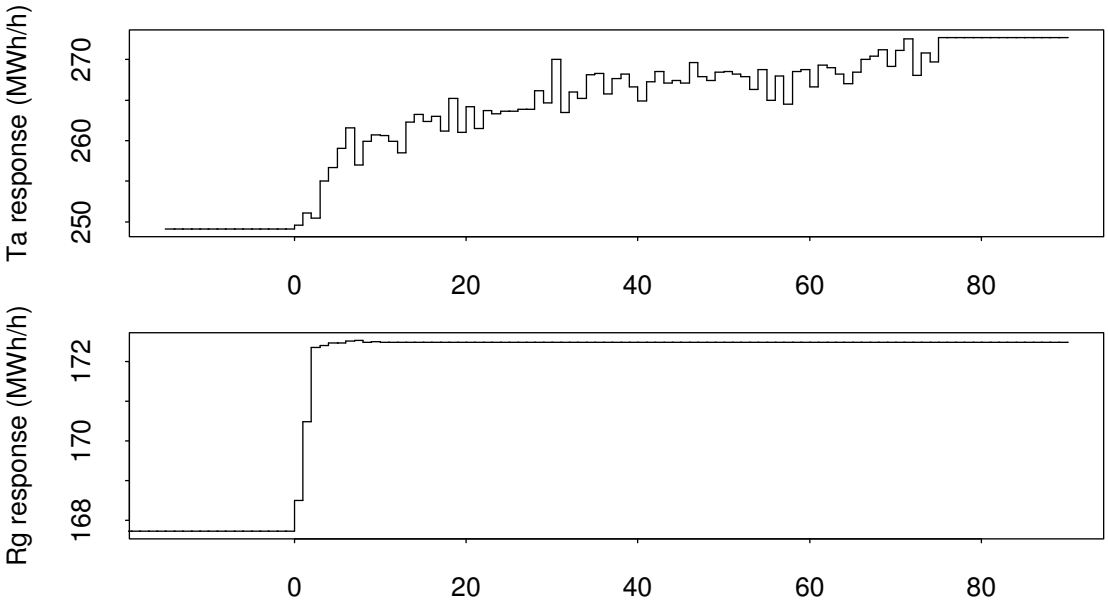


Figure 7.27. *Step response on air temperature and global radiation under model (7.10), with the order of the air temperature filter increased to 75. Data from 1982. The air temperature drops from 0 to -1 °C at time $t = 0$. The global radiation drops from 25 to 20 W/m^2 at time $t = 0$.*

Figures 7.28, 7.29 and 7.30 show the estimated annual fluctuation and stationary climate response for all years. Note that within each year the y-axes are identical and the minima are set to zero. Furthermore the response on the global radiation is only displayed for values below 800 W/m^2, although values as high as 1005 occur. The level stated below the year is the level obtained when the four functions are at their minimum. The functions are plotted only in the range for which observations exists.

Compared with model (7.9) (Figures 7.21, 7.22 and 7.23) the stationary response on the global radiation has a larger range. The remaining curves are similar.

Figures 7.31 and 7.32 show the diurnal fluctuations for Midweeks (Tuesdays, Wednesdays and Thursdays) for each year. The time of year and the time of day are indicated as a normalized time, which is the original time interval scaled to $[0, 1]$. The estimated diurnal fluctuations are very similar to those obtained with model (7.6), see Figures 7.9 and 7.10. The diurnal fluctuations are in the range of two to three times the size of the annual fluctuations. Again the reinforcement of summer time in 1980 are clearly seen in the plots.

The estimated deviation from the annual fluctuation ($\mu_i(t)$ in e.g. model (7.10)) is very constant between years and during the individual years when compared with the range of the diurnal fluctuation. For normal working days the level is approximately 100 MWh/h above the annual fluctuation, while for the remaining days the level is approximately 200 MWh/h below.

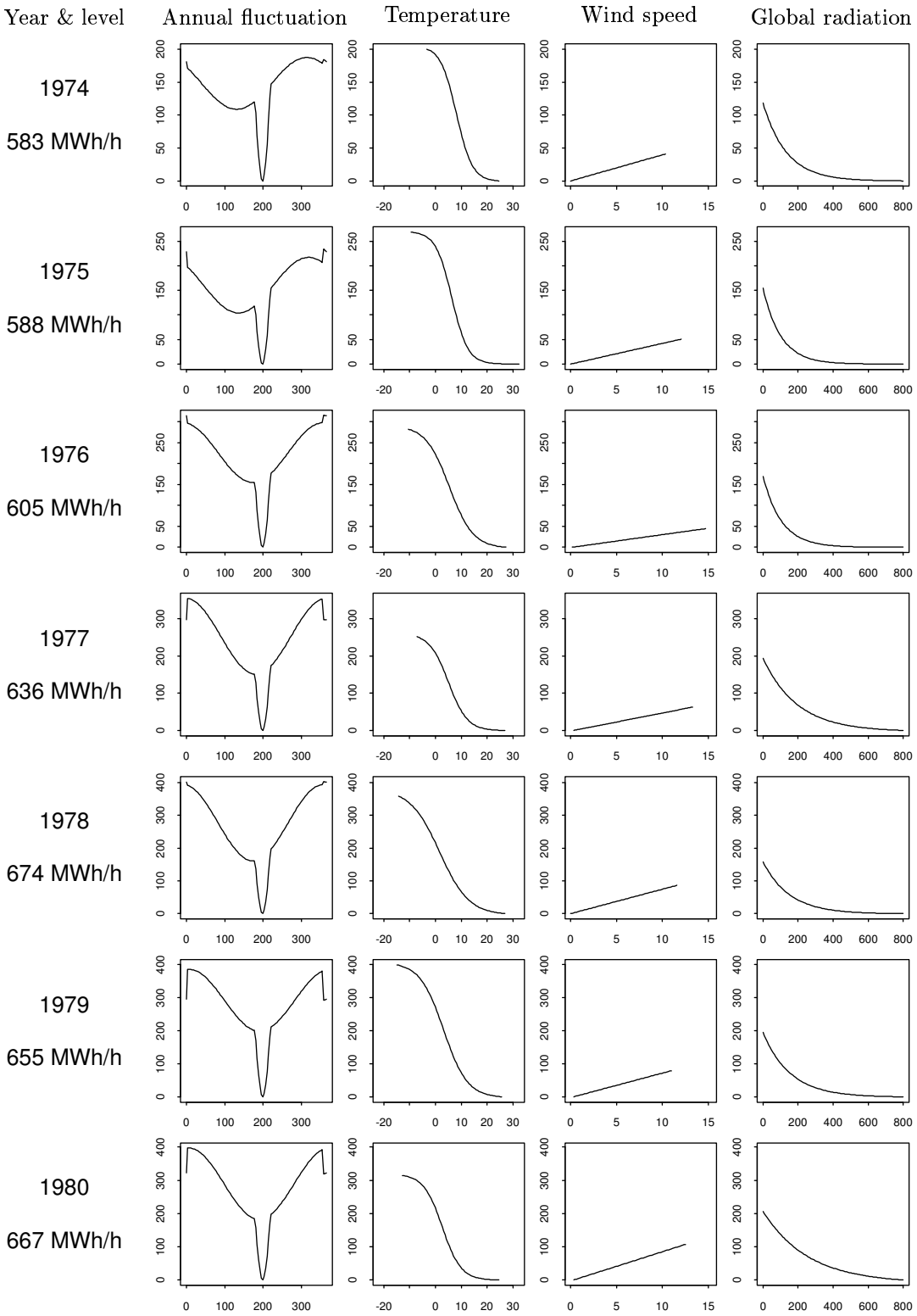


Figure 7.28. *Estimated annual fluctuation and climate response for the years 1974 – 1980 (model (7.10), with temperatures up to lag 75). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

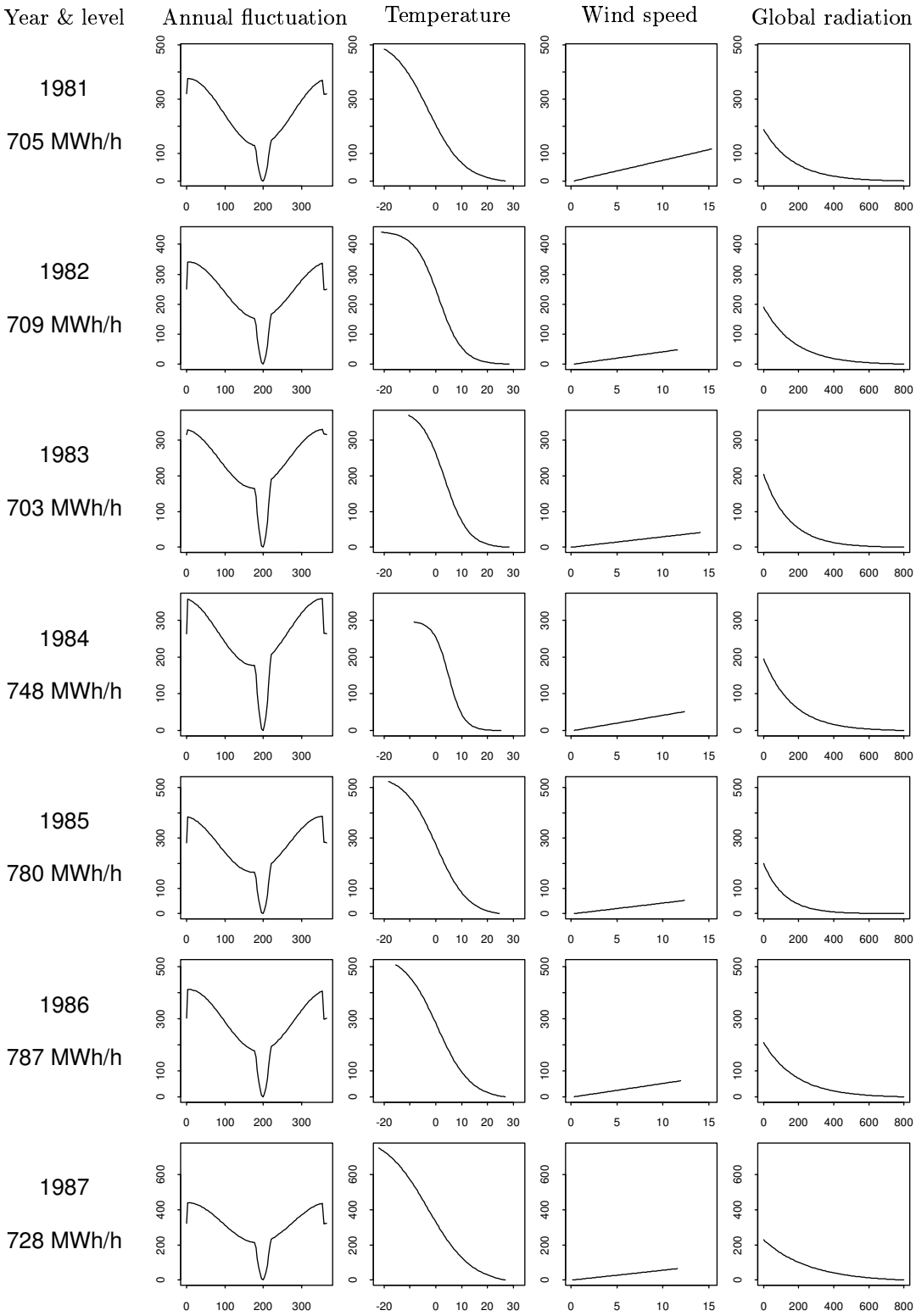


Figure 7.29. *Estimated annual fluctuation and climate response for the years 1981 – 1987 (model (7.10), with temperatures up to lag 75). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h.*

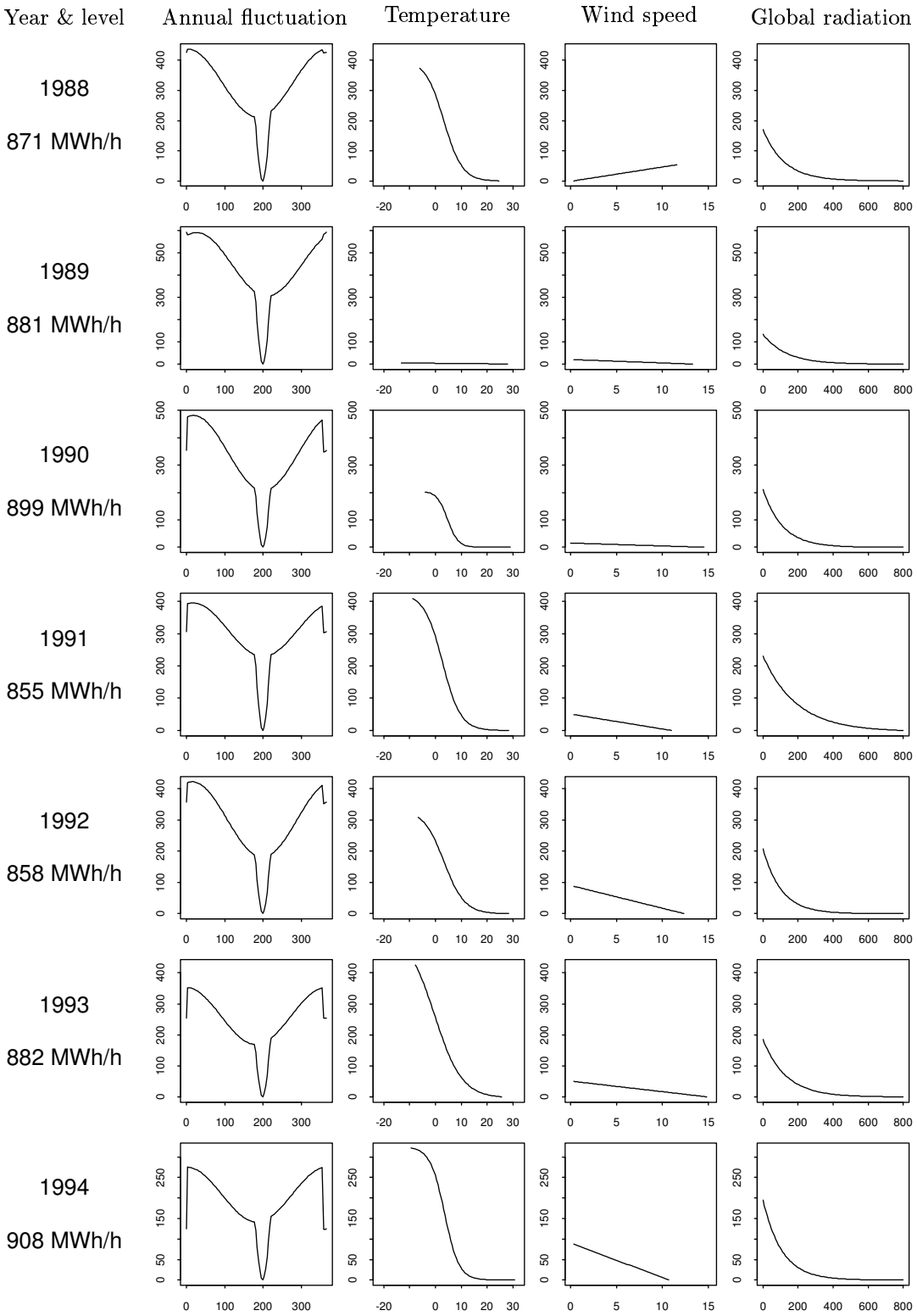


Figure 7.30. *Estimated annual fluctuation and climate response for the years 1988 – 1994 (model (7.10), with temperatures up to lag 75). From left to right: Year and level, annual fluctuation vs. day in year, stationary response on temperature (°C), response on wind speed (m/s), and response on global radiation (W/m^2). The unit on the y-axis is MWh/h. For 1989 the estimation did not converge.*

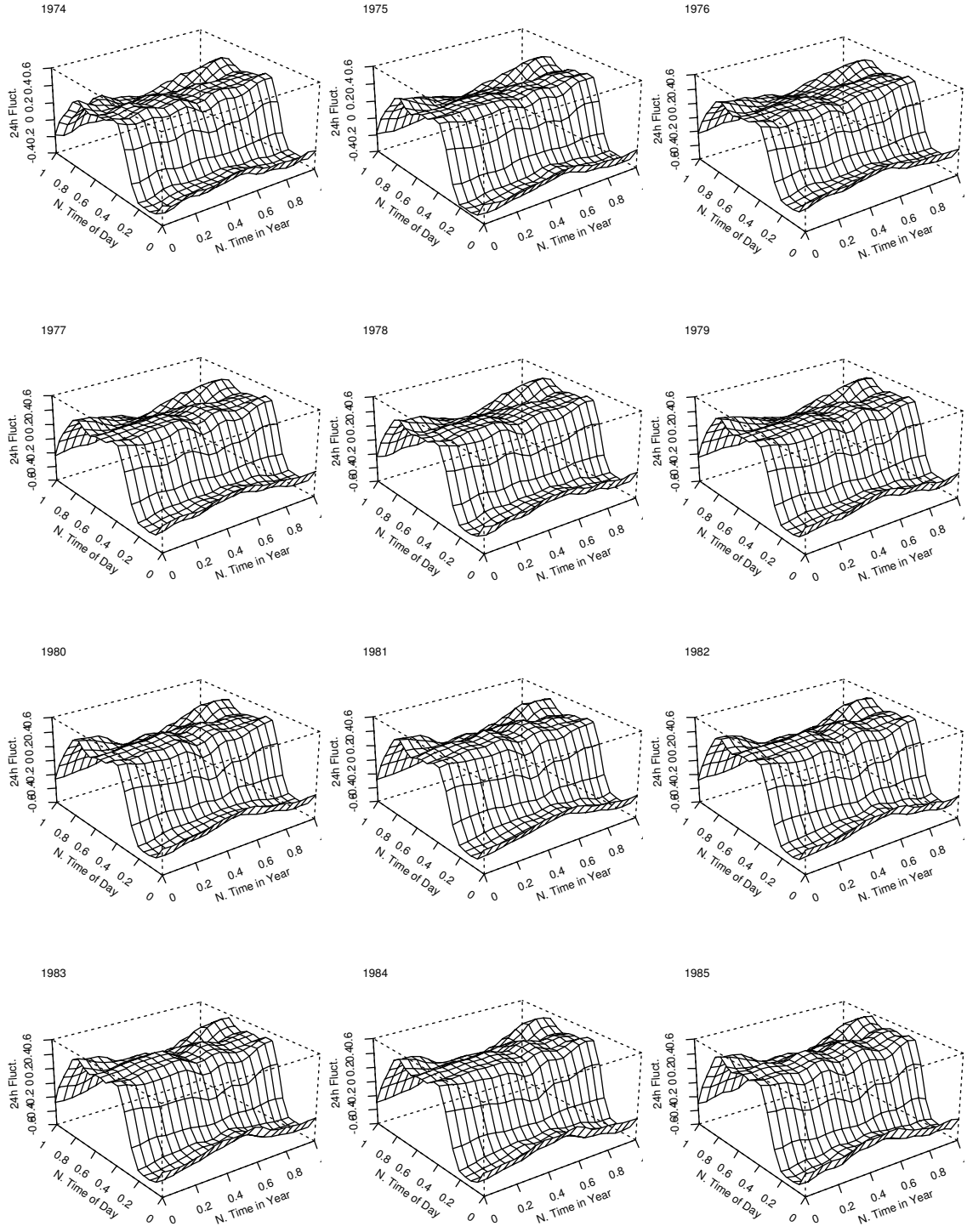


Figure 7.31. *Estimated diurnal fluctuation for Midweeks for the years 1974 – 1985 (model (7.10), with temperatures up to lag 75). The unit is GWh/h.*

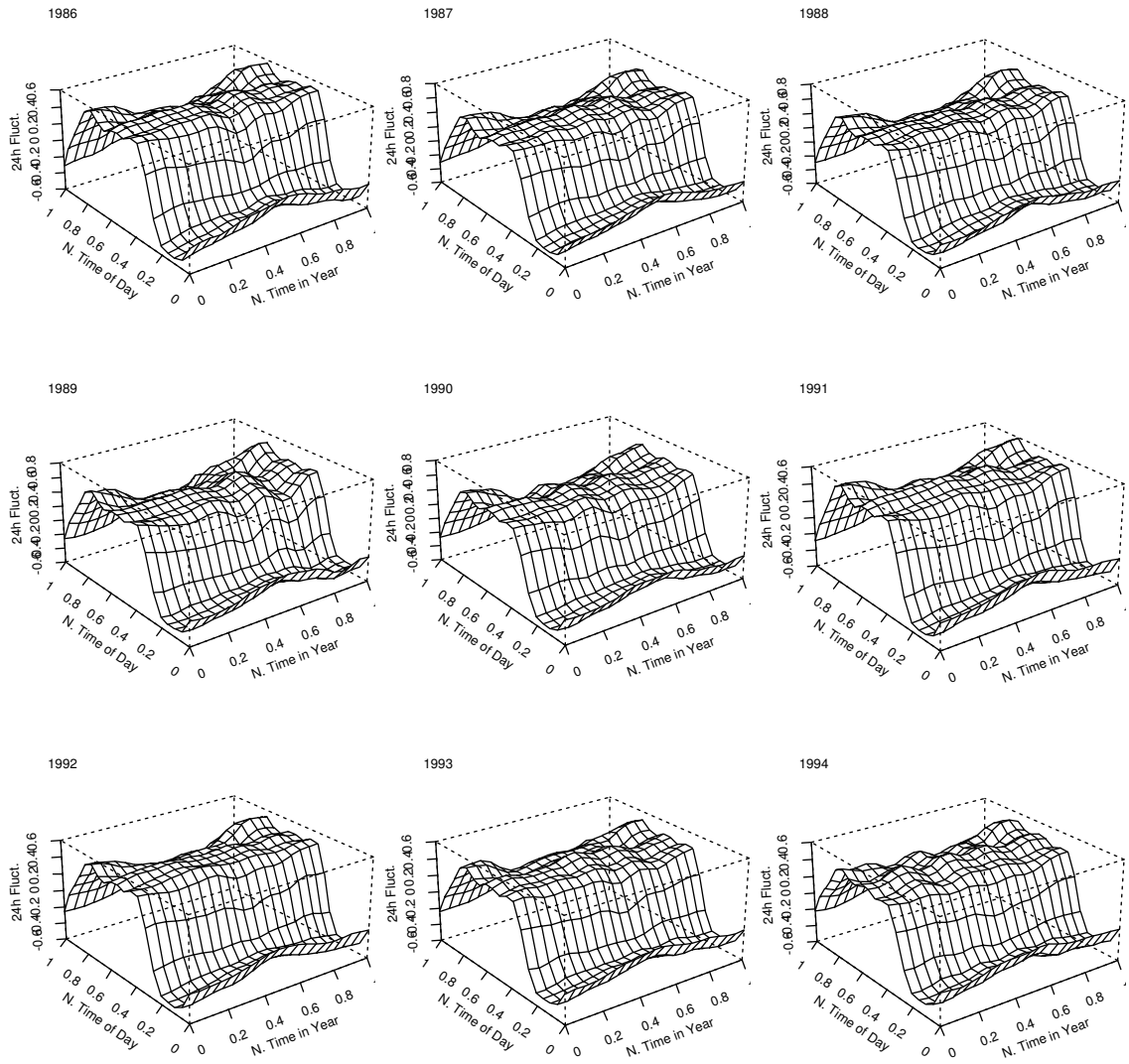


Figure 7.32. *Estimated diurnal fluctuation for Midweeks for the years 1986 – 1994 (model (7.10), with temperatures up to lag 75). The unit is GWh/h.*

Chapter 8

Analysis of Estimates

In this chapter estimates obtained under model (7.10), with the order of the air temperature filter increased to 75, are examined and discussed. The analyses are performed with the purpose of extracting/obtaining information which may be useful in relation to performing a scenario analysis, see Chapter 9. For this reason potential explanatory variables and the relation between estimates are considered. In the analysis 1989 is excluded since the estimation failed to converge for this year.

8.1 Overall Level of Power Consumption

With the purpose of trying to explain the variation from year to year, corrected for differences in climate, some economical variables are considered.

The overall level of the power consumption is probably related to all the variables shown in Table 2.1 and 2.2 (Section 2.2), but, the level used in the analysis reported below is corrected for the wind speed. However, since the level is between 1000 and 1800 MWh/h and the maximum installed wind power capacity is 120 MW. For these reasons the installed wind power capacity will have only marginal influence on the level of the power consumption.

The level is calculated by excluding the periodic components and finding the stationary response on 0 $^{\circ}C$, 0 m/s and 0 W/m^2, i.e.

$$\hat{\mu}_0 + \frac{\hat{a}_{T2}}{1 + \exp(-\hat{a}_{T0})} + \hat{a}_{R2}, \tag{8.1}$$

since the filtered air temperature and global radiation equals zero. The level determined by (8.1) is thus the level corrected for differences in the climate from year to year. In Table 8.1 the estimated level is shown for each year.

In Figure 8.2 a scatter plot matrix of the year and the regressors is shown. It is seen that many of the regressors are almost linear dependent. For this reason principal

Year	Level (MWh/h)	Year	Level (MWh/h)	Year	Level (MWh/h)
1974	1040.9	1981	1348.7	1988	1649.8
1975	1141.4	1982	1392.7	1989	–
1976	1222.6	1983	1416.7	1990	1655.5
1977	1288.3	1984	1463.4	1991	1737.9
1978	1324.7	1985	1526.7	1992	1687.9
1979	1408.8	1986	1567.4	1993	1633.6
1980	1375.4	1987	1613.2	1994	1655.6

Table 8.1. *Estimated level for model (7.10), with the order of the air temperature filter increased to 75. For 1989 the estimation failed to converge.*

components of the regressors are used in the analysis, see e.g. (Mardia et al., 1979). The number of households is known only for 1980-1994. Furthermore since this variable is almost linear dependent of many other variables (e.g. the campaign index) it is excluded from the analysis.

The principal components are based on values of GPD_{fc}, TFP, electricity price, campaign index, and popupation size scaled to zero mean and unit variance. The first two principal components explain 95% of the total generalized variance, and for the first three components the corresponding number is 99.6 %. For this reason only the first three principal components are used as regressors on the level.

In Table 8.2 the center, scale, and coefficients defining the linear combinations of the scaled variables forming the principal components are shown. The first component is a mixture of all five variables, but with a lower weight on the price of electricity. The second component is mainly the price, whereas the third is mainly the population size.

	Center	Scale	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
GPD_{fc}	1.65×10^5	1.81×10^4	-0.502	0.0633	-0.190	-0.3789	0.7513
TFP	1.04×10^2	4.36	-0.501	0.0338	-0.217	0.8365	0.0294
Price	3.77×10^{-1}	4.85×10^{-2}	0.244	0.9436	-0.210	0.0516	0.0564
Campaign	1.05	7.55×10^{-2}	-0.492	0.0904	-0.432	-0.3883	-0.6416
Population	2.27×10^6	1.37×10^4	0.441	-0.3102	-0.828	0.0575	0.1410

Table 8.2. *Center, scale, and coefficients defining the five principal components.*

Figure 8.3 is a scatter plot matrix for the year, level, and the five principal components. It is seen that the level is strongly related to the first component, whereas the remaining components seems to be of minor importance for explaining the level.

The level shown in Table 8.1 is modelled as a second order polynomial in the first three principal components (c_1, c_2 and c_3).

$$\zeta_i = \alpha_{00} + \sum_{j=1}^{3} \alpha_{j0} c_{ji} + \sum_{j=1}^{3} \sum_{k=j}^{3} \alpha_{jk} c_{ji} c_{ki} + \epsilon_i; \quad i = 1974, \ldots, 1988, 1990, \ldots, 1994, \quad (8.2)$$

where ζ_i is the level in year i, $\alpha_{..}$ is the parameters of the model and ϵ_i is the noise which is assumed to be independently identically normally distributed random variables with zero mean. For estimation and testing in models of this kind see e.g. (Myers, 1986).

Under the assumption that the model is correct the null hypothesis that the second principal component c_2 can be excluded was tested. The corresponding F-statistic is 2.9706, with 4 numerator and 10 denominator degrees of freedom, a p-value of 0.074 is obtained. Although not clearly non-significant the second component is excluded from the model.

In the reduced model all the parameters are significant, except α_{33} for which the t-test yields a p-value of 0.118. If α_{33} is fixed at zero and the resulting model is tested against model (8.2) a p-value of 0.067 is obtained. The estimates and standard errors are displayed in Table 8.3. The residual standard error is 30.32 MWh/h and $R^2 = 0.9809$. Figure 8.1 shows the results obtained for all three models. The fitted values of the final model do not differ much from those of the other models.

	Estimate	Std. error
α_{00}	1496.2	12.3
α_{10}	-95.5	3.9
α_{30}	-83.2	23.0
α_{11}	-7.5	2.7
α_{13}	-41.0	11.7

Table 8.3. *Estimates and standard errors of estimates of the parameters in the final model of the level.*

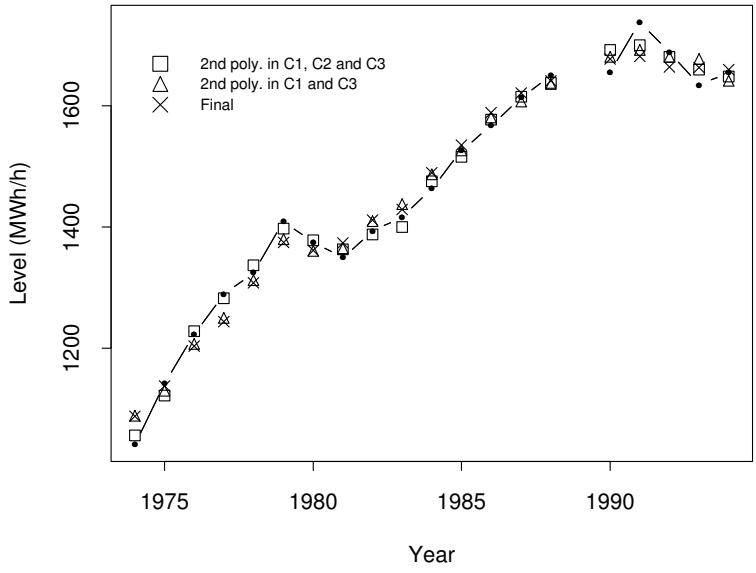


Figure 8.1. *The values of the corrected level (line-points) and the values fitted using the three models investigated.*

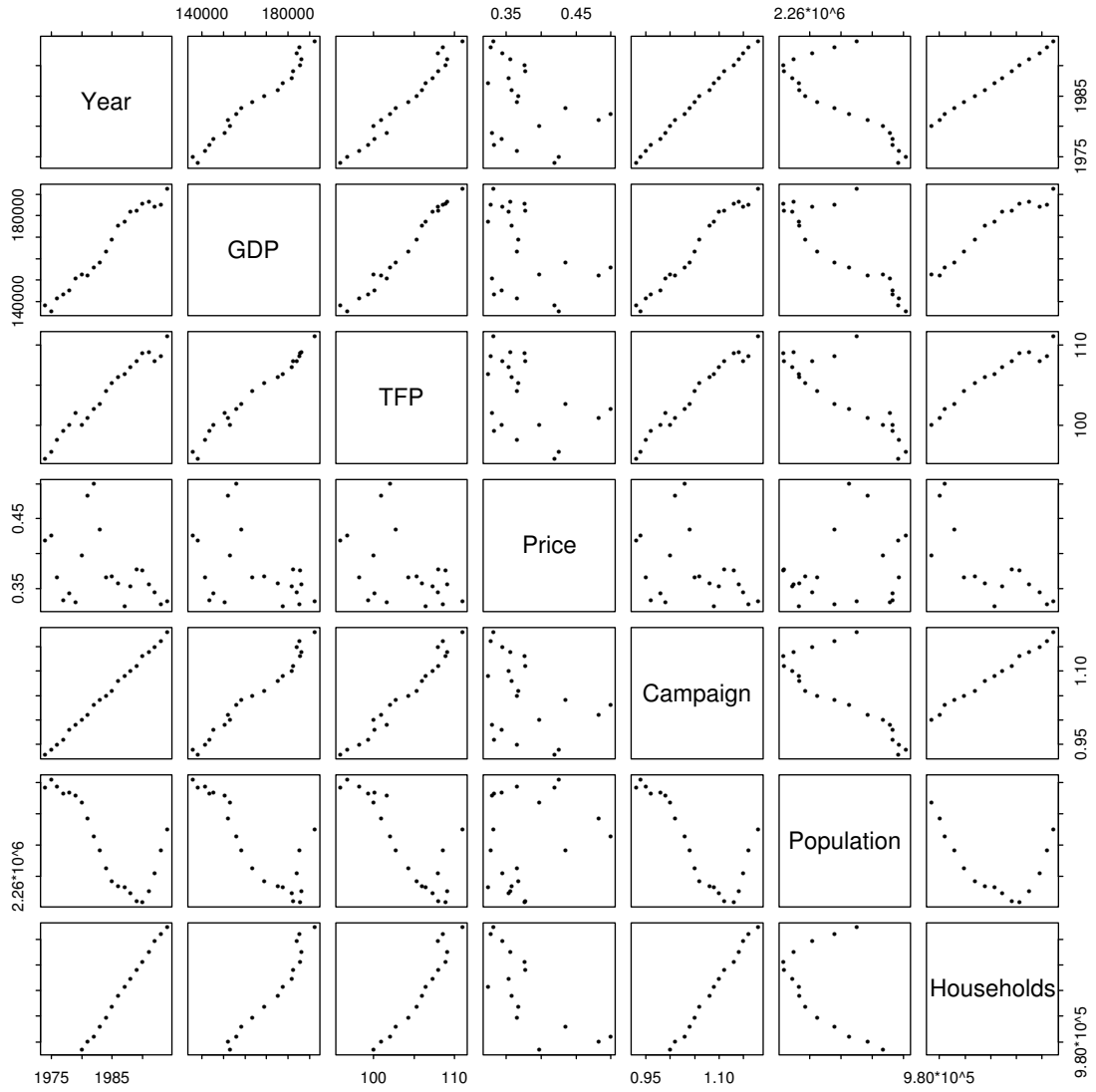


Figure 8.2. *Scatter plot matrix of year and the variables in Table 2.2. The units are identical to those used in Table 2.2. For each plot the variable on the x-axis is indicated by the column diagonal element, whereas the variable on the y-axis is indicated by the row diagonal element.*

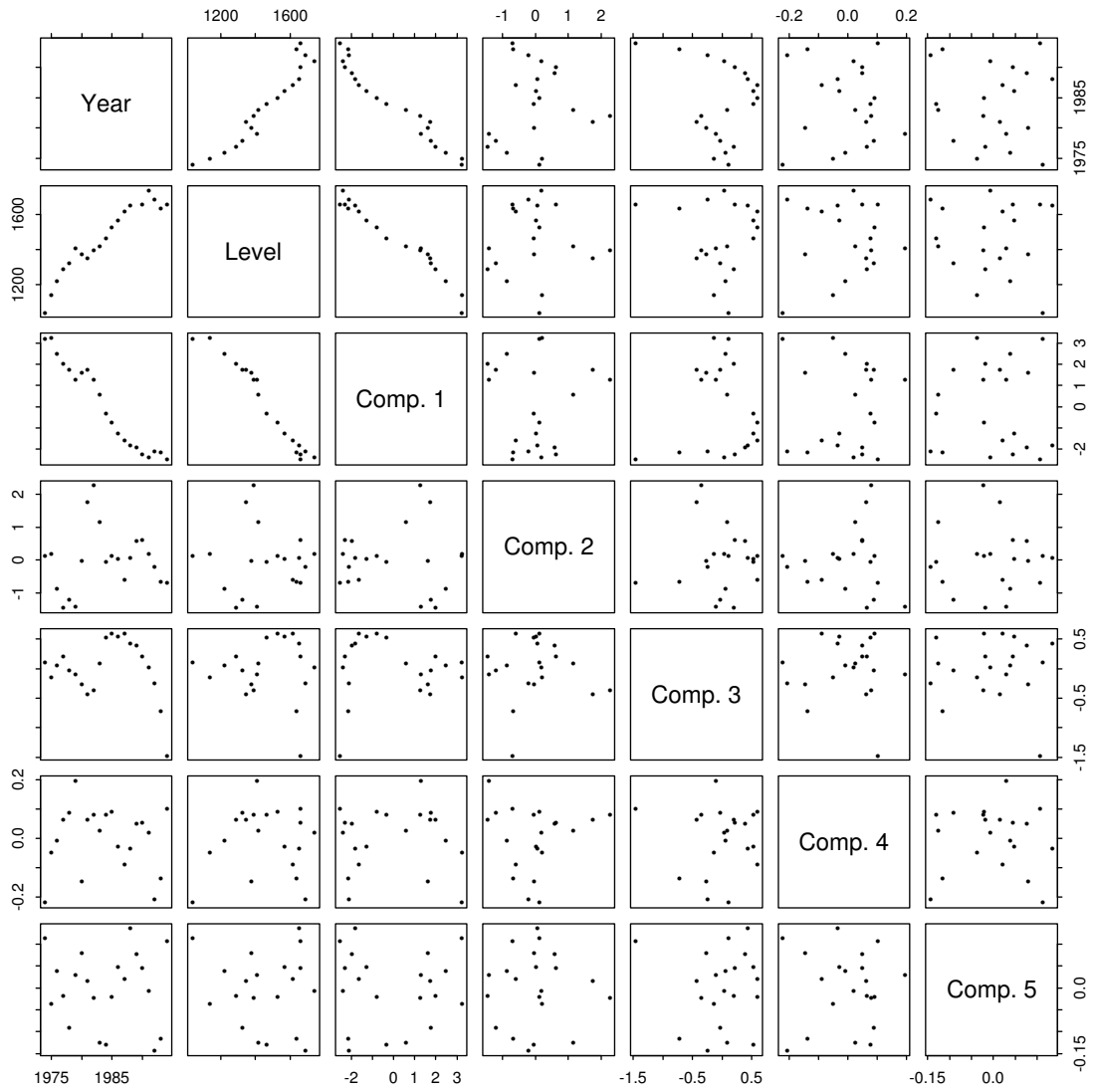


Figure 8.3. *Scatter plot matrix of year, level and the principal components.*

8.2 Periodic Variations

In this section the amplitudes of the periodic variations are compared with an appropriately defined level, e.g. the amplitude of the annual fluctuation for the individual years are compared with the level defined as (8.1). Especially it is investigated if these amplitudes are proportional to the levels.

8.2.1 Annual Fluctuation

This section addresses components of the estimated annual fluctuation based on model (7.10), with the order of the air temperature filter increased to 75. In Figure 8.4 the total amplitude and the amplitude of the first order harmonic are shown, whereas Figure 8.5 shows the size of the drop during Christmas and the size of the drop during summer. These quantities are plotted against time and the annual level defined as (8.1).

On the plots against level a least squares fits of a strait line and of a line trough the origin are shown. The years 1974–1976 and 1994 seem atypical and have been excluded from the fits displayed on Figure 8.4, whereas only 1994 is excluded from the fits displayed on Figure 8.5. In Table 8.4 the estimated coefficients are shown.

In the case of traditional least squares the coefficients of the linear fits are obtained as those who minimize the sum of the squared vertical distances between the line and the points. This is not necessarily the most appropriate fit when both the independent and dependent variable are affected by noise. For this reason also orthogonal regression is applied, see (Kendall and Stuart, 1961)[1]. When using this method the coefficients are chosen to minimize the sum of the squared orthogonal distances. However, in this case the fitted line obtained by orthogonal regression are similar to the ones obtained by traditional least squares. Consequently, orthogonal regression lines are not shown.

From the plots it is concluded that the drop during summer is proportional to the level. As an approximation this also holds for the total amplitude. On the other hand the amplitude of the first order harmonic is nearby constant, whereas the drop during Christmas is dependent of the level, but the dependence is non-proportional.

[1] (Kendall and Stuart, 1961) use the term "linear functional relationship".

	Linear		Proportional
	Intercept (MWh/h)	Slope (-)	(-)
Total amplitude	157.61	0.1557	0.2594
Amplitude of 1st harmonic	170.54	0.0279	0.1402
Drop during summer	9.06	0.1157	0.1218
Drop during Christmas	-189.82	0.1720	0.0430

Table 8.4. *Intercepts and slopes of lines shown in Figures 8.4 and 8.5.*

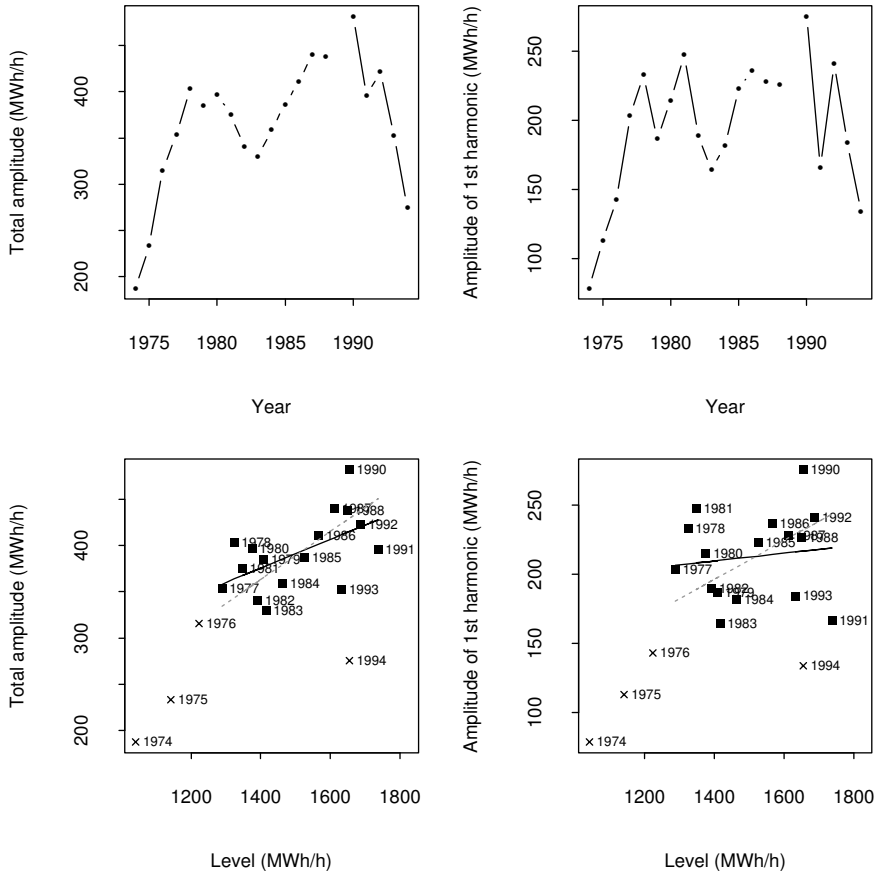


Figure 8.4. *Total amplitude (left) and amplitude of first order harmonic (right) versus time (top row) and annual level (bottom row). On the bottom row the year is indicated at the right of the data point. The linear fits are indicated by solid lines, whereas the proportional fits are indicated by dotted lines. The fits are based on data points shown as filled boxes.*

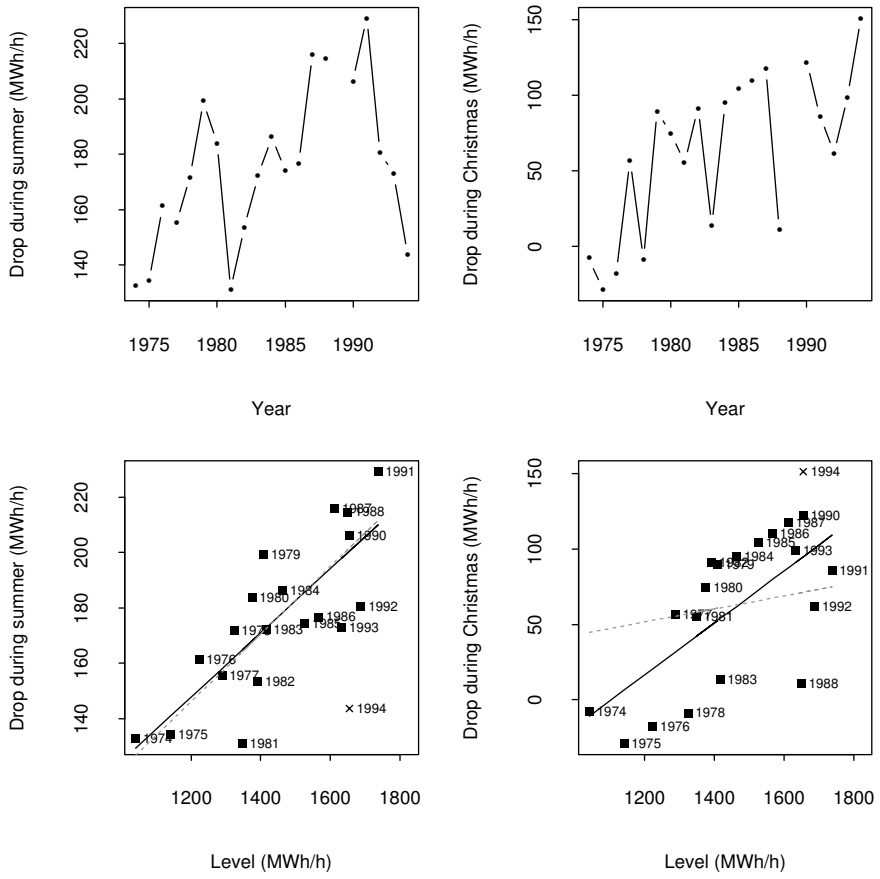


Figure 8.5. *Drop during summer (left) and Christmas (right) versus time (top row) and annual level (bottom row). On the bottom row the year is indicated at the right of the data point. The linear fits are indicated by solid lines, whereas the proportional fits are indicated by dotted lines. The fits are based on data points shown as filled boxes.*

8.2.2 Diurnal Fluctuation

This section deals with the relation between the estimated amplitude in the diurnal fluctuations for the five types of days and the estimated level of the day. The level of the day is defined as the sum of the annual fluctuation and the annual level, where the annual level is defined by (8.1). For each day the amplitude is found by calculating the value of the diurnal profile at 25 evenly spaced points starting at 00:00 and ending at 24:00. For each of the five types of days the amplitude and the level is calculated at 365 evenly spaced points covering one year. The procedure is repeated for each of the 21 sets of estimates, i.e. for each year.

Figure 8.6 shows the involved quantities against time, whereas Figure 8.7 shows the amplitudes against the level. On Figure 8.7 orthogonal, normal and proportional regression lines have been added. The coefficients of the fits are given in Table 8.5.

From the figures it is clearly seen that the drop during the summer period is not reflected in the amplitudes of the diurnal profiles. However, this is a consequence of the model and not necessarily implied by the data. The plots clearly indicate that two groups are present in the data. Investigations show that these groups are quite precisely reproduced by splitting the data in the groups (i) before 1988 and (ii) 1988 or later, this might be due to the tax reform in 1986, see Section 2.3. For these reasons the lines shown on the plots are fitted separately for the two groups defined by the year and excluding the summer period.

Before 1988 the amplitude for Mondays, Midweeks, and Fridays seems to have been proportional to the level. Which is also approximately true for half-holy and holy days. For 1988 or later the amplitudes do not seem to be proportional to the level. For these years the orthogonal fit seems more appropriate.

	Type of day	Orthogonal		Linear		Proportional
		Intercept (MWh/h)	Slope (-)	Intercept (MWh/h)	Slope (-)	(-)
1974–1987	Mondays	-388.00	1.1537	182.44	0.6892	0.8373
	Midweeks	-359.55	1.0914	97.30	0.7195	0.7984
	Fridays	-469.92	1.1475	101.42	0.6823	0.7646
	Half-holy	-1342.59	1.6803	-869.60	1.2952	0.5897
	Holy	-1060.24	1.4247	-592.84	1.0442	0.5632
1988–1994	Mondays	101.30	0.8957	180.04	0.8118	1.0007
	Midweeks	75.09	0.8704	140.85	0.8003	0.9481
	Fridays	103.12	0.8219	147.73	0.7743	0.9293
	Half-holy	-429.60	1.1377	-232.99	0.9281	0.6836
	Holy	-253.22	0.9067	-157.77	0.8050	0.6394

Table 8.5. *Intercepts and slopes of lines shown in Figure 8.7.*

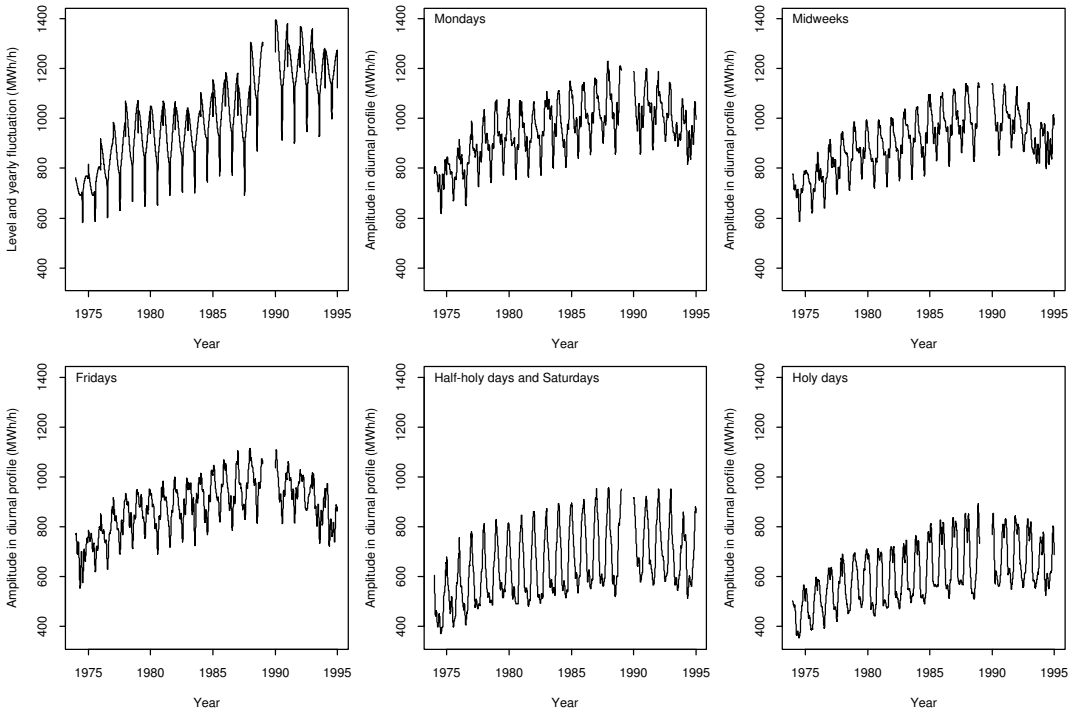


Figure 8.6. *The sum of the level and the annual fluctuation (upper left), and also the amplitude in the diurnal profile for the five different types of days versus time.*

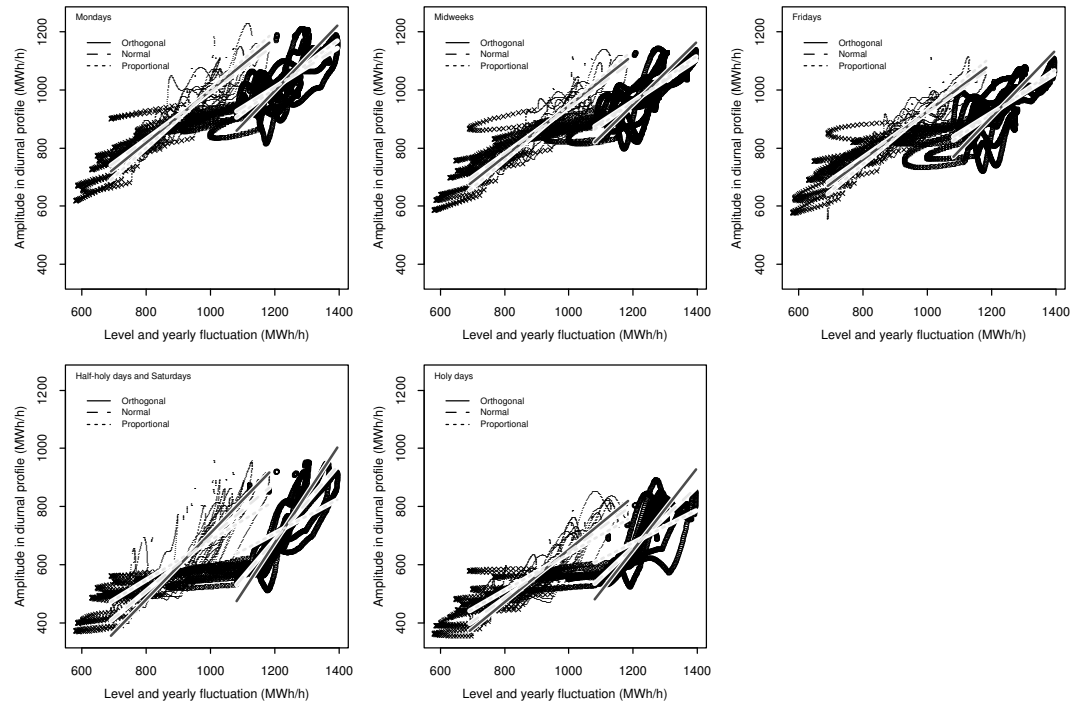


Figure 8.7. *The amplitude in the diurnal profile for the five different types of days versus the sum of the level and the annual fluctuation. Years 1988–1994 indicated by circles, summer periods indicated by crosses.*

8.2.3 Deviation from Level due to Type of Day

The main characteristics of the diurnal fluctuation, i.e. the amplitude, were analysed in the Section 8.2.2. This section addresses the deviation from the level caused by the type of day.

Again the data is split in two groups (i) before 1988 and (ii) 1988 or later, see e.g. Section 8.2.2. In Figure 8.8 the deviations for the individual type of days are plotted against the level. Furthermore fitted regression lines are shown. The coefficients of these lines are shown in Table 8.6. In this case the orthogonal regression lines are not included because they are almost identical to the normal regression lines.

From the plots it is seen that in all but one case the deviation is approximately proportional to the level. The case for which this does not hold is half-holy days in 1988 or later. However, in all cases the slope is very small and the dependence may be unimportant for practical purposes.

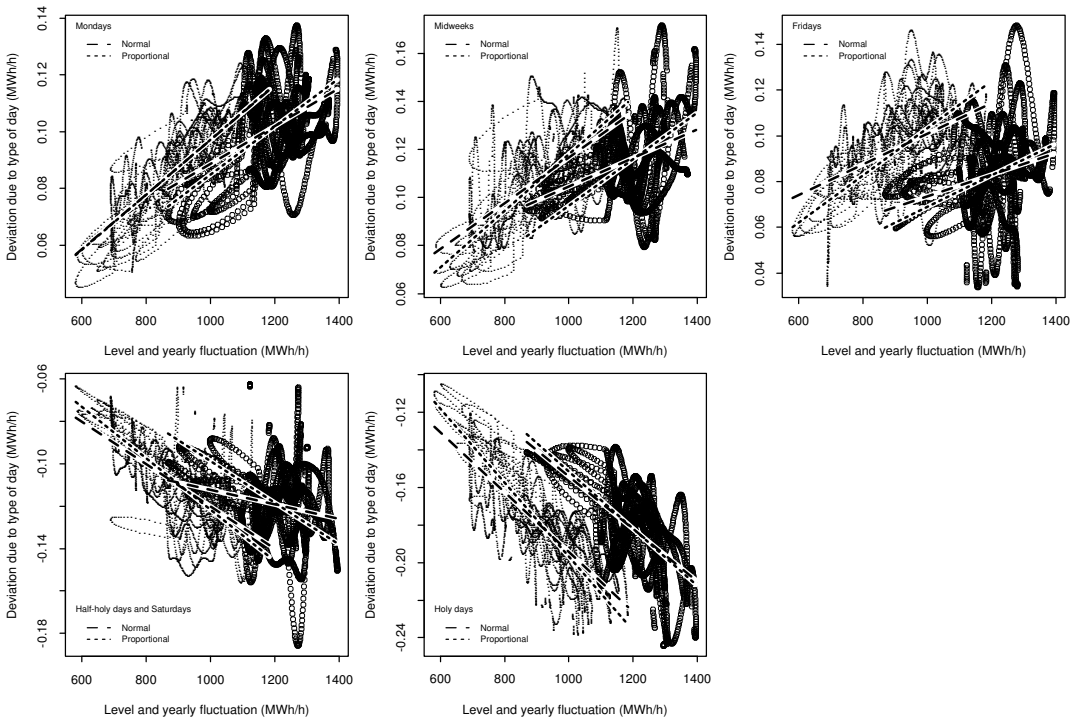


Figure 8.8. *Deviation from the level defined by the sum of the annual level and the annual fluctuation, due to the type of day. Years 1988–1994 indicated by circles.*

	Type of day	Linear		Proportional
		Intercept (MWh/h)	Slope (-)	(-)
1974–1987	Mondays	0.0005	9.66×10^{-5}	9.71×10^{-5}
	Midweeks	0.0215	9.51×10^{-5}	11.81×10^{-5}
	Fridays	0.0338	6.70×10^{-5}	10.32×10^{-5}
	Half-holy	-0.0202	-9.99×10^{-5}	-12.15×10^{-5}
	Holy	-0.0355	-15.89×10^{-5}	-19.69×10^{-5}
1988–1994	Mondays	0.0220	6.67×10^{-5}	8.48×10^{-5}
	Midweeks	0.0448	5.97×10^{-5}	9.67×10^{-5}
	Fridays	0.0257	4.77×10^{-5}	6.89×10^{-5}
	Half-holy	-0.0780	-3.43×10^{-5}	-9.87×10^{-5}
	Holy	-0.0146	-13.97×10^{-5}	-15.17×10^{-5}

Table 8.6. *Intercepts and slopes of lines shown in Figure 8.8.*

8.3 Influence of Climate

8.3.1 Wind Speed

As seen from e.g. Figures 7.28, 7.29, and 7.30 the influence of wind speed changes sign over the period. In this section the estimated slope on the wind speed dependence is compared with the total amount of installed wind power capacity (Table 2.1).

In Figure 8.9 the wind power capacity and the estimated wind coefficient are plotted. It is seen that the data is very sparse. A linear relationship for wind power capacities larger than 80 MW is not contradicted.

8.3.2 Temperature

This section focuses on the stationary and dynamic parts of the response on temperature. The stationary part is analysed by investigating the position, slope and span of the logistic function describing the stationary response on temperature. Reference is made to the notation used in (7.10).

The horizontal position of the logistic function is defined as $a_{T0}/\sum_{i=0}^{75} b_i$, the span is a_{T2}, and the slope is $-a_{T2}\sum_{i=0}^{75} b_i/4$. The estimated values of these quantities plotted against time are shown in Figure 8.10, together with pairwise scatter plots.

The negative slope increases over the period, whereas the position and span seem to fluctuate. However, as shown on one of the scatter plots the span and the position are closely related.

The dynamic response is investigated by analysing the step response function of the

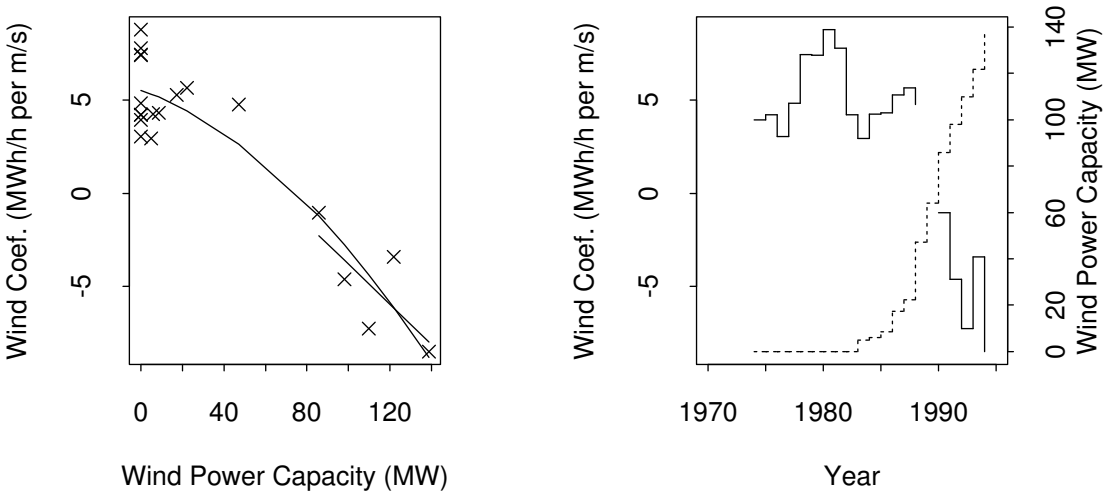


Figure 8.9. *Estimated wind coefficient against wind power capacity (left), with line fitted for 1990-1994 and second order polynomial fitted for all years. Right is shown a time-plot of the estimated wind coefficient (—) and wind power capacity (- -).*

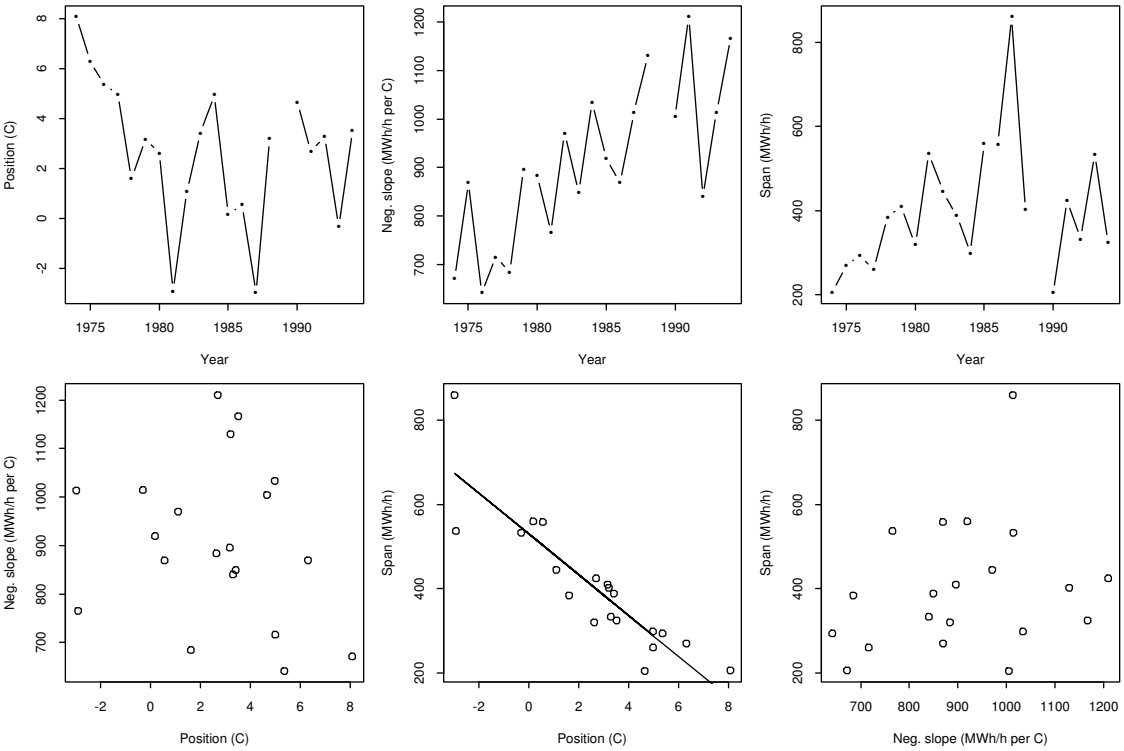


Figure 8.10. *Characteristics of stationary response on air temperature. Top row: Position of the logistic function versus time (left), negative slope versus time (middle), and span versus time (right). Bottom row: Negative slope versus position (left), span versus position with regression line* $y = 530.5 - 48.42x$ *overlayed (middle), and span versus negative slope (right).*

filter $T_f(t) = \sum_{i=0}^{75} b_i T_a(t-i)$. The step response for the individual years is displayed in Figure 8.11. Also, on the figure the least squares approximation corresponding to the rational transfer function $(c_0 + c_1 q^{-1})/(1 + a_1 q^{-1} + \ldots + a_5 q^{-5})$ is shown. The coefficients of the approximating filter are found by simulating 15000 values[2] from the original filter with uniformly distributed white noise input and fitting the rational transfer function to the obtained series. Different (low) orders of the polynomials involved were investigated. A visual comparison revealed that the stated approximation is the most appropriate.

Although the approximation is not very precise, it seems that the rational transfer function may be appropriate for scenario analyses. However, the stationary response should be adjusted, see also Chapter 9. From the shape of the approximating step response it is concluded, that only the largest pole is important. Summary statistics for this pole, c_0, and c_1 is given in Table 8.7. Because the slope of the stationary response depends on the filter parameters, see the beginning of this section, the values of c_0 and c_1 used in the table are divided by the stationary gain of the approximating filter.

	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
c_0	-0.08	0.04	0.11	0.11	0.14	0.34
c_1	0.06	0.14	0.16	0.19	0.24	0.37
Pole	0.90	0.92	0.93	0.93	0.95	0.96

Table 8.7. *Approximate low pass filter on air temperature with unit stationary gain: Summary statistics for c_0, c_1 and the largest pole.*

[2]With this No. of simulated values the difference between replicated simulations is small.

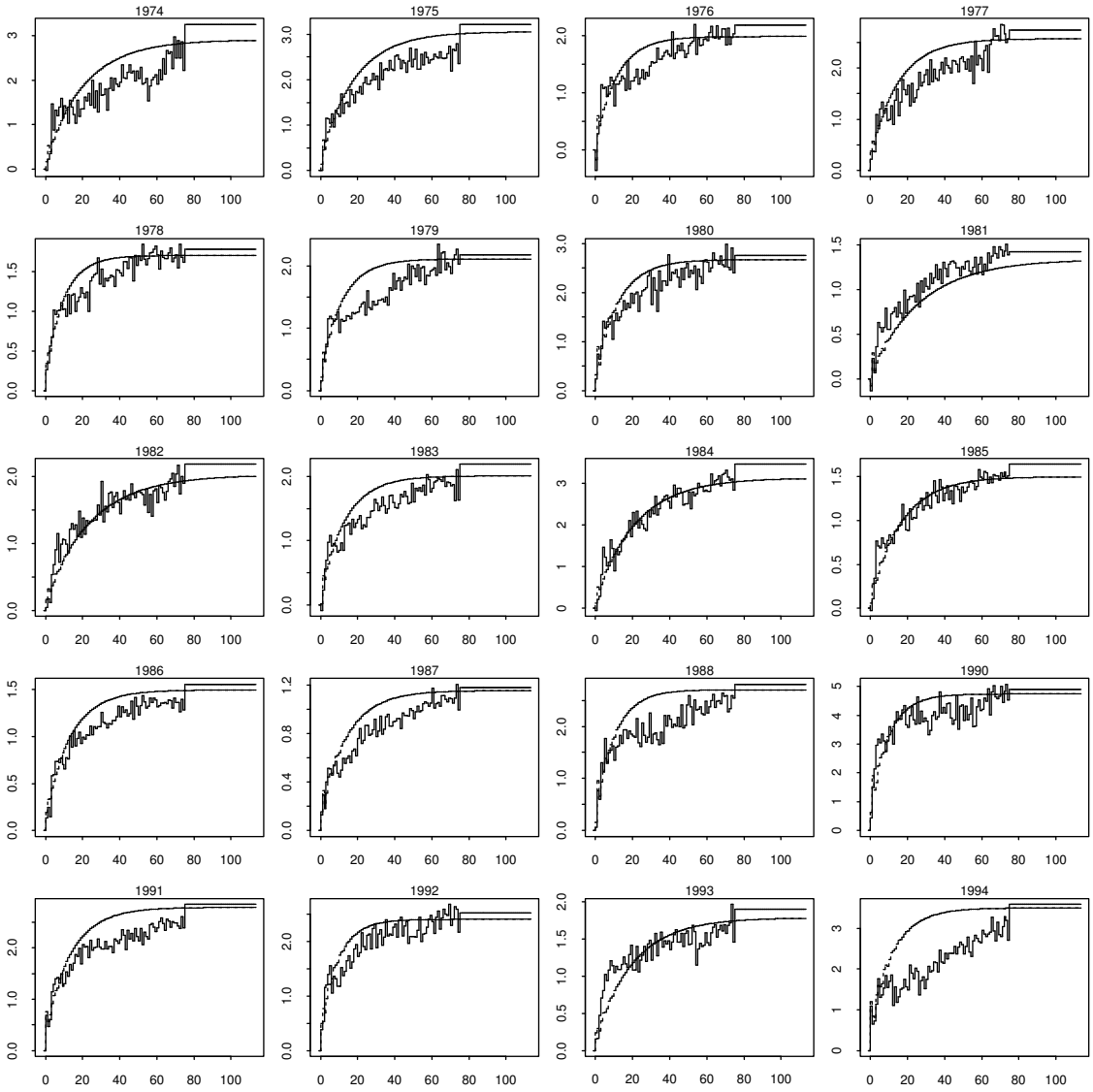


Figure 8.11. *Step response of the low pass filter on air temperature approximated by the rational transfer function* $(c_0 + c_1 q^{-1})/(1 + a_1 q^{-1} + \ldots + a_5 q^{-5})$.

8.3.3 Global Radiation

This section consideres the stationary and dynamic parts of the response on global radiation. The stationary part is analysed by investigating the slope and span of the exponential function describing the stationary response on global radiation. Reference is made to the notation used in (7.10).

The slope of the exponential function is defined as $-a_{R2}\sum_{i=0}^{10} g_i$, i.e. the slope at 0 W/m^2, whereas the span is a_{R2}. The estimated values of these quantaties plotted against time are shown in Figure 8.12, together with a scatterplot of the two quantaties. Both the span and the slope fluctuate over time.

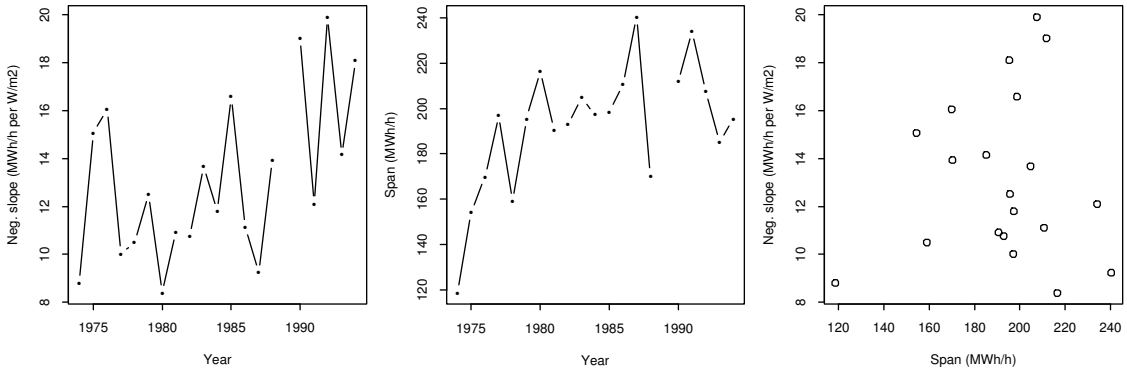


Figure 8.12. *Characteristics of stationary response on global radiation. From left to right: Negative slope at 0 W/m^2 versus time, span of the response versus time, and negative slope at 0 W/m^2 versus span of the response.*

The dynamic response is investigated by analysing the step response function of the filter $R_f(t) = \sum_{i=0}^{10} b_i R_g(t-i)$. The step response for the individual years is displayed in Figure 8.13. Also, on the figure the least square approximations corresponding to the rational transfer function $(c_0 + c_1 q^{-1})/(1 + a_1 q^{-1})$ are shown. The coefficients of the approximating filter is found by simulating 15000 values from the original filter with uniformly distributed white noise input and fitting the rational transfer function to the series obtained. Different (low) orders of the polynomials involved was investigated. A visual comparison revealed that the stated approximation is the most appropiate.

The approximation is quite precise. Summary statistics for the pole $(-a_1)$, c_0, and c_1 is given in Table 8.8. Because the slope of the stationary response depends on the filter parameters, see the beginning of this section, the values of c_0 and c_1 used in the table are divided by the stationary gain of the approximating filter.

	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
c_0	0.15	0.18	0.21	0.21	0.25	0.30
c_1	0.11	0.28	0.33	0.34	0.41	0.64
Pole	0.18	0.34	0.47	0.44	0.54	0.66

Table 8.8. *Approximate low pass filter on global radiation with unit stationary gain: Summary statistics for c_0, c_1 and the pole.*

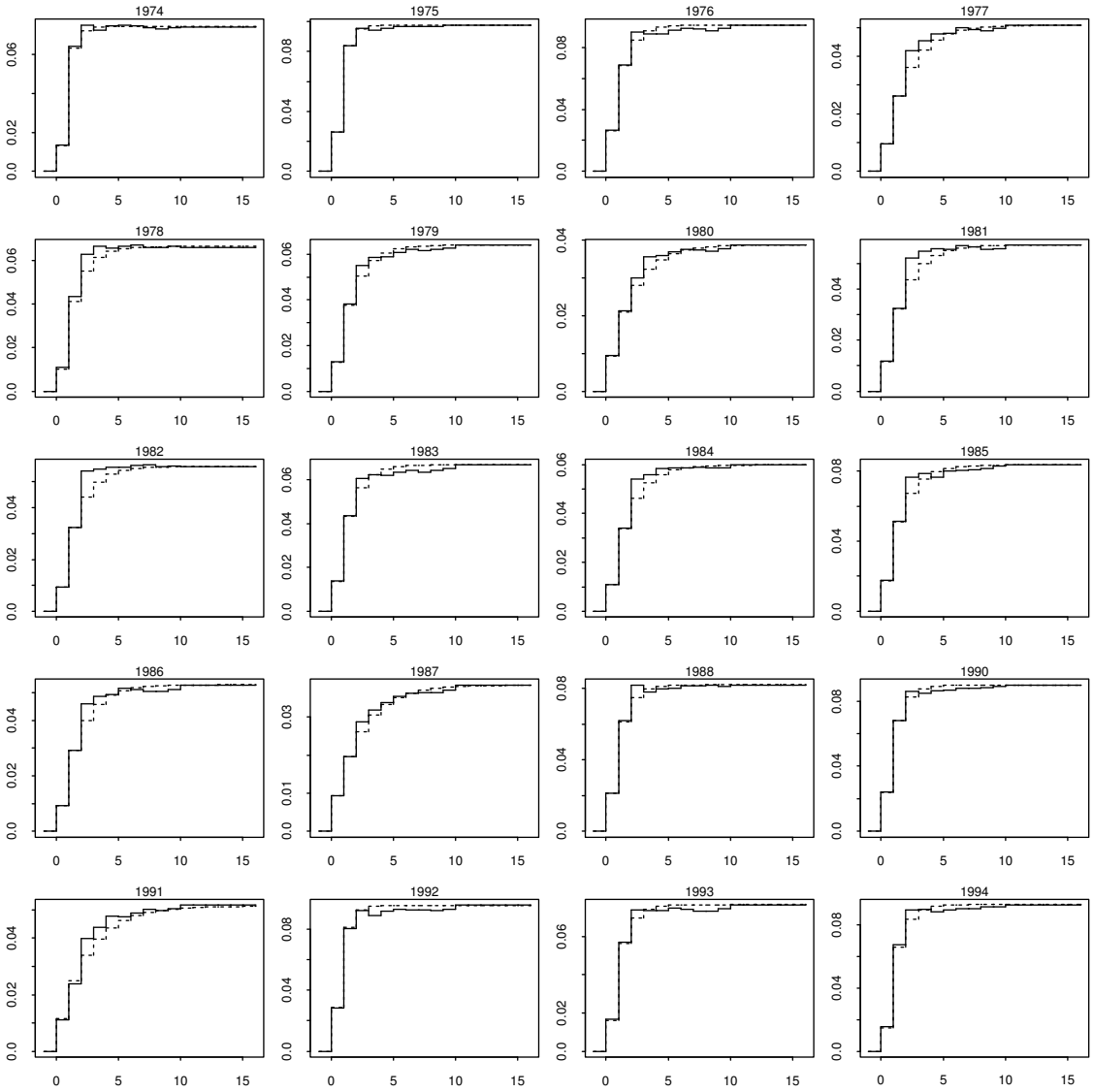


Figure 8.13. *Step response of the low pass filter on global radiation approximated by the rational transfer function $(c_0 + c_1 q^{-1})/(1 + a_1 q^{-1})$.*

8.4 Residual Variance

To facilitate the construction of e.g. 95% prediction intervals the Root Mean Square Error (RMSE) is analysed. This quantity is an estimate of the square root of the residual variance. It is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N-p}\sum_{i=1}^{N} r_i^2}, \tag{8.3}$$

where r_i is the residuals, p is the No. of parameters in the model and N is the No. of observations used in the estimation.

In Figure 8.14 the RMSE is plotted against time and against the annual level defined by (8.1). A strait line is fitted for the relation between RMSE and level, excluding years 1974, 1985 and 1994. The linear relationship between RMSE and the level is quite convincing.

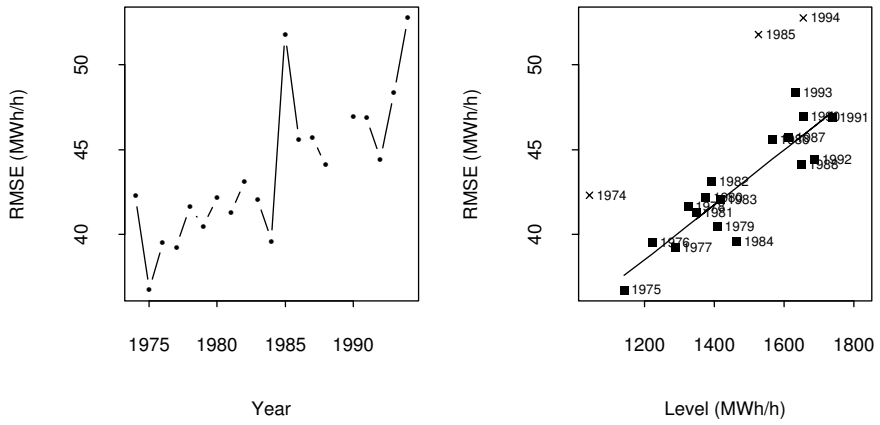


Figure 8.14. *Root mean square error versus time (left) and versus level (right). On the rightmost plot the line* $y = 19.08 + 0.01618x$ *is overlayed.*

8.5 Events

An analysis of the relation between the events mentioned in Section 2.3 and the power consumption is complicated by the fact that none of the models investigated yields residuals that resembles white noise.

For the events in 1976 an 1977 (economic support for energy conservation and tax on power) no specific dates are stated. It is however interesting to note that the speed of which the observed level increases from one year to the other is lower for 1977 to 1978 than for the preceding years, see Figure 8.1.

For the years in which the remaining events occur the residuals of model (7.10), with the order of the air temperature filter increased to 75, are shown in Figure 8.15. Note that the estimation failed for 1989. The plot of the residuals is included for completeness. Also, note that the scale on the y-axis differs between plots.

No immediate effects are observed in conjunction with the abolishment of the restrictions, related to the first oil crisis on 12 February 1974. However, an increasing trend is observed for the first part of 1974 (approximately until the middle of March). The event of 21 December 1979 is close to Christmas during which large residuals are seen every year. This also holds for the EEC summit meeting of 11 December 1991. The events of 16 October 1982 (new government) and 17 October 1986 (tightening of private economy) are both followed by a period of negative residuals. However, systematic deviations of the same size are observed periodically during the years.

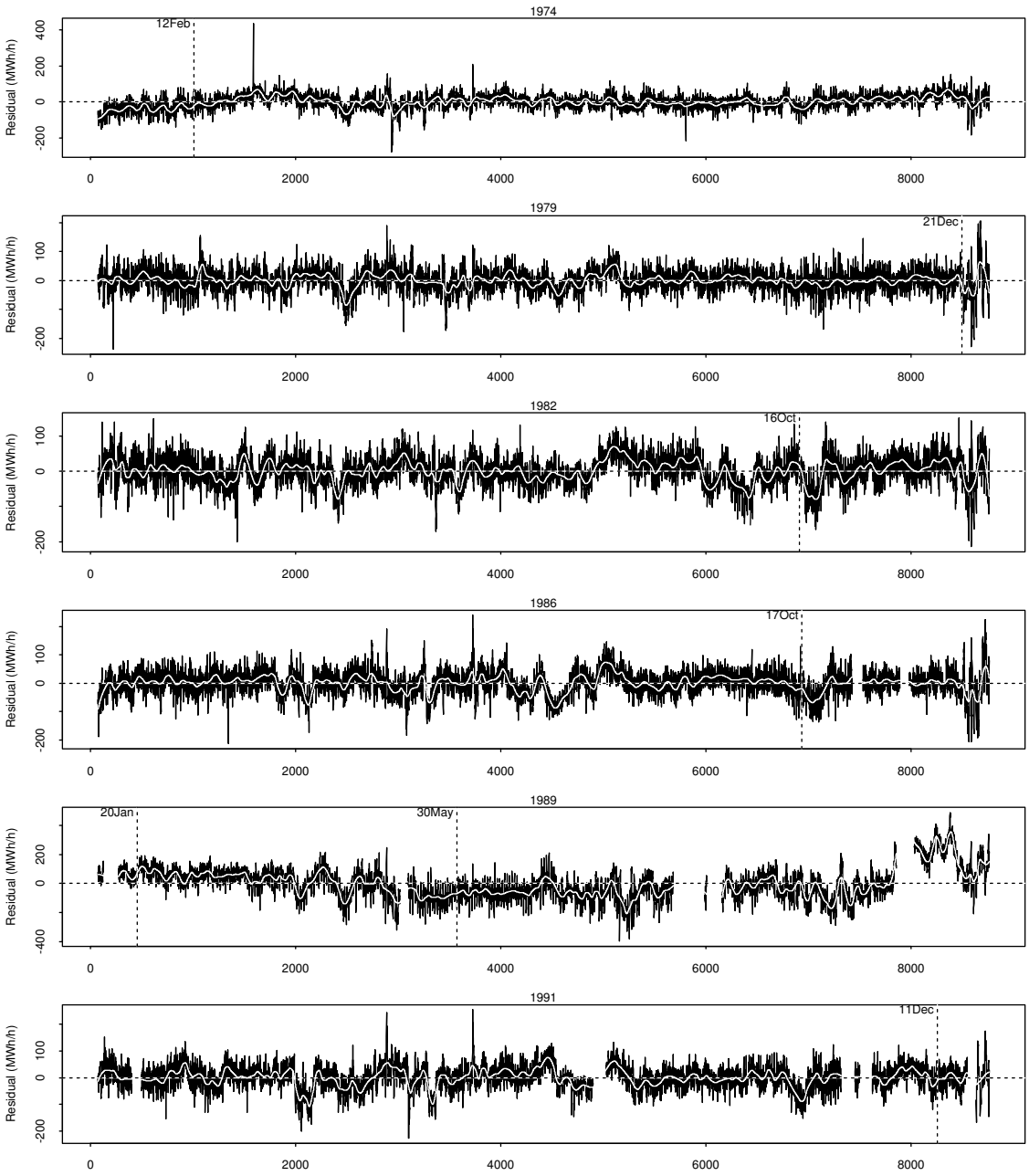


Figure 8.15. *Residuals of model (7.10), with the order of the air temperature filter increased to 75. Only years for which separate events occur are included. Superimposed on the plots are a 2nd order local regression smoothing with a nearest neighbour bandwidth of 0.02 calculated using the* `loess()` *function of S-PLUS. The estimation failed to converge for 1989.*

Chapter 9

Strategy for Long Term Predictions

This chapter deals with the basic procedure required to make long term predictions based on the results presented in this report. Since the prediction horizon is in the range of 10 to 20 years the procedure is quite involved in that the predictions must be based on an appropriate set of assumptions. Several sets of assumptions may have to be investigated in order cover the range of possible directions which the future evolution might take. In the following one such set of assumptions are called a *scenario.*

In general terms the prediction problem is the following: *Given the climate predict the hourly power consumption as defined in Section 2.1 for one full year 10 to 20 years ahead.* Furthermore, the influence of climate may be investigated by using the total available record of climate measurements, performed in accordance with (Jensen, 1995), to obtain several predictions for each hour in the year.

Due to the length of the prediction horizon the functional relationships identified on basis of data covering the years 1974–1994 should not be directly adopted. Instead they should merely be taken as a statement on how things worked in the past. The user should then use these relationships, together with the information of Chapter 8 and own assumptions, to generate a model, which is then used to calculate predictions. In the following we will assume that the general structure of this model is the same as the structure of the models described in Chapter 7, i.e.

$$E[P(t)] = \mu_0 + \mathcal{Y}(t) + \sum_{i=1}^{N_g} I_i(t)[\mu_i(t) + \mathcal{D}_i(t)] + \mathcal{T}(T_f(t)) + \mathcal{W}(W(t)) + \mathcal{G}(R_f(t)). \quad (9.1)$$

The terms of this equation are defined in Table 9.1 below.

Term	Definition
$E[P(t)]$	Expected power consumption at time t
μ_0	Overall level of power consumption
$\mathcal{Y}(t)$	Annual fluctuation about the overall level
$i = 1, \ldots, N_g$	Day groups
$I_i(t)$	Indicator function for the day corresponding to time t belonging to group i
$\mu_i(t)$	Deviation from $\mu_0 + \mathcal{Y}(t)$
$\mathcal{D}_i(t)$	Deviation from $\mu_0 + \mathcal{Y}(t) + \mu_i(t)$
$\mathcal{T}(T_f(t))$	Influence of low-pass filtered air temperature
$\mathcal{W}(W(t))$	Influence of wind speed
$\mathcal{G}(R_f(t))$	Influence of low-pass filtered global radiation

Table 9.1. Terms used in model (9.1).

9.1 Selecting a Scenario Model

The following considerations should be taken into account when fixing the terms on the right hand side of (9.1).

Overall level (μ_0). The overall level should be related to a specific setting of climate variables, e.g. in Chapter 8 the level is calculated by excluding the periodic components and finding the stationary response on 0 $^\circ C$, 0 m/s and 0 W/m^2.

If predictions of economic variables are available the determination of μ_0 may be facilitated by the analysis described in Section 8.1. However, the relation observed in the past between economic variables and the level of power consumption need not be the same in the future. In fact, due to possible structural changes in the economy, the observed relation will probably not hold in the future. For this reason it is recommended to perform a thorough economic analysis of possible future developments before μ_0 is fixed.

Annual fluctuation ($\mathcal{Y}(t)$). This variable describes the deviation from the overall level μ_0 not directly caused by the climate. A typical value for the range of this variable is 400 MWh/h, based on 1974 – 1994. Except for the Christmas and summer holidays the deviation seems to be fairly well described by a first order harmonic. The drop during Christmas is probably caused by a large part of the industry being closed between Christmas and New-year. Similarly for the summer period, but, due to non-simultaneous summer holidays, the drop is less sharp.

Grouping of days. The grouping of days used in this report will probably be applicable for long term predictions. However, possible future distributions of workdays over the week should be considered.

Diurnal variation ($\mu_i(t) + \mathcal{D}_i(t)$). The diurnal variation consists of a mean shift $\mu_i(t)$ for days of type i from the overall level at the particular time of year $\mu_0 + \mathcal{Y}(t)$, and a periodic variation specific for the day type $\mathcal{D}_i(t)$ around the resulting mean of the particular day. Note that $\sum_{i=1}^{N_g} \mu_i(t) = 0$, and that the shape of the diurnal variation changes over the year. In particular the part of the year, for which daylight savings are in effect, is important. In the future a change in the tariff structure may imply a change of the diurnal variation.

Influence of climate. The influence of climate will depend on many aspects which may change over a period of 10 to 20 years, e.g. the fraction of houses with electricity heating and the amount of installed capacity on wind mills. All such aspects must be taken into account when constructing the scenario model.

The expected amount of installed capacity on wind mills may be used to select an appropriate wind coefficient, see Section 8.3.1. The handling of the air temperature and global radiation is more difficult since both the dynamic and stationary part must be modelled. In the following it is assumed that the response on each of these climate variables can be expressed as a possible non-linear transformation of a low-pass filtered version of the variable. The following procedure is suggested:

1. Select a filter with unit stationary gain and an appropriate step response, using information from Section 8.3.2 or 8.3.3.
2. Select the function describing the stationary response, also using information from Section 8.3.2 or 8.3.3.

Since the chosen filter has unit stationary gain, the temperature or radiation in step 2 may now just be replaced by the low-pass filtered variable from 1.

9.2 Producing Predictions

After a scenario model has been selected, and the calendar for the year in question has been retrieved, it only remains to decide on the handling of the climate.

One possibility is to assume the climate to be known. In this case it is appropriate to use the Danish test reference year, see (Andersen, 1986). If climate measurements collected according to (Jensen, 1995) where in mind during the construction of the scenario model, the reference year may have to be scaled.

If the randomness of the climate is to be reflected in the predictions at least two different approaches may be applied. One approach is to identify and estimate a stochastic time-series model for the climate, see e.g. (Madsen, 1985), and use the marginal distribution of the predictions of climate to obtain confidence intervals

for the mean power consumption. If the model is non-linear in the climate some approximations will have to be used.

Another approach which requires many years of data, but is much easier to apply, is to produce several sets of predictions. In each prediction the climate is assumed to be known. The prediction is then repeated over the total record of climate observations and hence several predictions are obtained for each hour. The range of these predictions indicates the uncertainty due to the unknown climate.

Under the assumption that the model is true and all values of the regressors are known approximate prediction intervals may be constructed using of the results of Section 8.4. These intervals can of course be included in approaches where the randomness of the climate is addressed.

Chapter 10

Conclusion

The work described in this report documents that the power consumption on a hourly basis can be described by models which take into account (i) the time of year, (ii) the time and type of day, and (iii) the climate expressed as air temperature, wind speed and global radiation. It is important that the diurnal fluctuation is allowed to change over the year. Furthermore it is important to model the dynamic response on the climate. Step responses on global radiation stabilize within five hours, whereas step responses on air temperature seems to stabilize within three days. A possible explanation is that both the air temperature and global radiation affects the indoor temperature, by heat conduction through the walls of the buildings, whereas the global radiation mainly affects the indoor temperature by radiation trough the windows, see also (Madsen and Holst, 1995). Furthermore, the global radiation influences the need for electric lighting.

The range of the diurnal fluctuation is approximately 1000 MWh/h and for the annual fluctuation the corresponding number is approximately 400 MWh/h. However, the influence of temperature adds approximately 500 MWh/h to the annual fluctuation.

The models are estimated separately for each of the years and R^2 values of approximately 0.98 or 0.99 are obtained. The level of power consumption on a yearly basis, corrected for climate fluctuations, is fairly well described by economic variables. The amplitude of the annual fluctuation seems to be related to the level of the power consumption. Likewise the amplitude of the diurnal fluctuation is related to the daily level of power consumption. The slope of the wind speed dependence changes sign over the period, and a linear relationship with the amount of installed wind power capacity is indicated. The change of sign is most likely caused by the increasing amount of installed wind mills as it occurs around the years 1989 – 1991.

It should be possible to refine the models further since the residuals are not white noise. For this reason it is also difficult to address the influence of separate political events during the period.

Chapter 11

Discussion

The models of the power consumption considered in this report result in a quite adequate fit, i.e. the R^2 value is high, and the components of the models have a reasonable interpretation. Since the models are intended for long term prediction they need to describe as much as possible of the variation in the power consumption by the periodic fluctuations and the dependence on climate. For this reason the autocorrelation of the residuals have not been modelled, as the slow change in power consumption from one hour to the next will account a large part of the variation. If the autocorrelation is of interest it may be better to model it based on the residuals of the models obtained in this report. This approach would allow for e.g. stochastic simulation.

Comparison of estimates obtained separately for each year with other estimates or external variables measured only once each year seems to be valuable for the construction of models to be used for scenario analysis and long term prediction. It is not possible to validate procedures for long term prediction, since the society most certainly will change over a period of 10 or 20 years. For this reason any procedure used will rely heavily on subjective judgement. Consequently, the scientific credibility of long term predictions will always be low and they will be open for discussion. Never the less, if such predictions are important for planing purposes, efforts should be invested in obtaining appropriate predictions and the uncertainty of these. By applying the structured approach to long term prediction presented in this report it is the hope that the discussion can be moved from the actual values to the more well defined areas (i) the basic structure of the model, (ii) comparisons with the past, and (iii) the set of assumptions to investigate in order produce a range of possible future values instead of just a single value.

The remaining part of the discussion focus on possible extensions to the methodology and may be skipped by readers not interested in these aspects. If a model for which the residuals are resembling white noise is obtained it would then be appropriate to investigate components of the resulting model by statistical tests. In general unimportant parts of the model should then be identified and excluded, whereby

the degrees of freedom used by the model is reduced. There are several directions in which the models could be further developed:

1. Modelling the influence from unknown discrete events.
2. Non-parametric estimation of the functions involved.
3. Alternative parametrizations of low-pass filters.
4. Modelling the correlation of the residuals.

Re. 1. The power consumption on an hourly basis is probably affected by discrete events. Such events include political events for which the response is expected to last for weeks or months and highly popular football matches on TV. Events of these kinds are impossible to predict and hence of minor relevance to scenario analysis. However, it may be relevant to know which size and duration such events may have. Also filtering these events from the residuals could possibly improve the quality of the estimates of interest – although this could probably also be achieved by robust estimation. One possible approach to the modelling of the influence of unknown events is to model the sequence of occurrences as a sequence of independent Bernoulli variables. In the case of an event the model would then postulate a response corresponding to a transfer function filtering an impulse. This last step is related to intervention analysis, see e.g. (Box and Tiao, 1975). A rational transfer function with fixed orders of the numerator and denominator polynomials can be used to model both short and long term impulse responses.

Re. 2. For modelling power consumption non-parametric methods seems to be very attractive. For instance by only applying constraints on the smoothness of the stationary climate response difference between years could be addressed by comparing the shape of the estimated functions, and e.g. relationships applicable for high air temperatures would not affect the estimated relationship for low temperatures and vice versa. Furthermore non-parametric estimation of the periodic variations could probably replace the Fourier expansions used in this report. In Section 7.6 (Figure 7.19) it is shown how the sharp drop in power consumption during the summer holidays and the Christmas results in quick fluctuations during the remaining part of the year. This is a natural consequence of fitting a high order Fourier expansion globally. Approximating the annual fluctuation locally by a low order polynomial, as done in local regression, is attractive since for instance the drop during the summer holidays will only affect the estimated annual fluctuation near this period.

The kind of model just outlined is called a Generalized Additive Model (GAM) as described in (Hastie and Tibshirani, 1990) and implemented in the S-PLUS software. However, for this application certain non-standard extensions are needed. To model the periodic variations the bandwidth must vary depending on the time, e.g. for the annual fluctuation the bandwidth should be low during the summer holidays and the Christmas. Furthermore for the annual fluctuation the distance between time

points should be measured on a circle, whereas for the diurnal fluctuation varying over the year the distance should be measured on a sphere. The other non-standard extension needed is the low-pass filtering of the climate variables. Extensions in this direction are less strait forward in that the model will not be additive in the sense of (Hastie and Tibshirani, 1990) and hence the backfitting algorithm can not be directly applied.

Re. 3. The type of low-pass filter used in this report is very simple since it only consists of a finite order polynomial in the back-shift operator. It has been shown that the estimated filter of the global radiation can be well approximated by a rational transfer function with first order polynomials in both the numerator and the denominator. Hence is seems natural to include this in the model. The estimated filter of the air temperature seems to be difficult to approximate by a rational transfer function. However, it should be possible to apply other restrictions to the impulse response $\{h_i\}_{i=0}^{\infty}$. It seems promising to apply only a smoothness constraint on the impulse or step response.

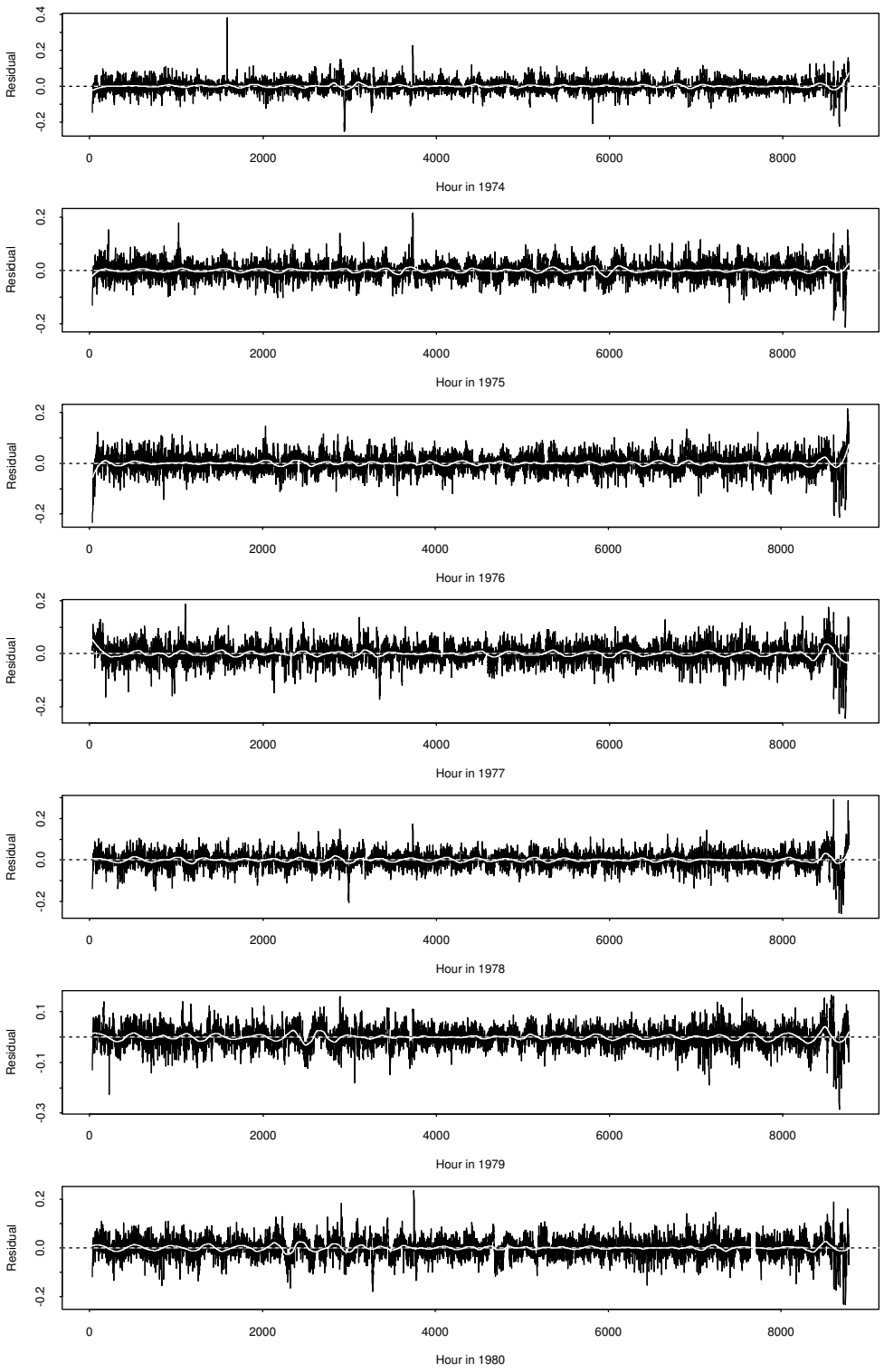
Re. 4. Modelling the correlation of the residuals will certainly improve the fit and will probably produce residuals resembling white noise. The danger of this approach is that, since the power consumption does not change very much from one hour to the other, a simple autoregressive model will result in an approximation highly appropriate for short term predictions, but the relevance of the model for long term predictions (the marginal mean) is very questionable. Furthermore, the dependencies of the power consumption on the external variables are not very important for short term predictions, and hence these dependencies will be poorly estimated. For these reasons modelling the correlation of the residuals should only be added if the results of the suggestions described above indicate that the model should be developed further. In all cases modelling the correlation of the residuals should only act as a minor refinement. This may achived by modelling the residuals, from models of the kind used in this report, using an ARMA process. That is the power consumption is modelled as $P_t = \mu(t, x_t) + N_t$, where $\mu(t, x_t)$ is a time-varying mean which depends on the external variables x_t, i.e. models of the kind considered in this report, and N_t is an ARMA process.

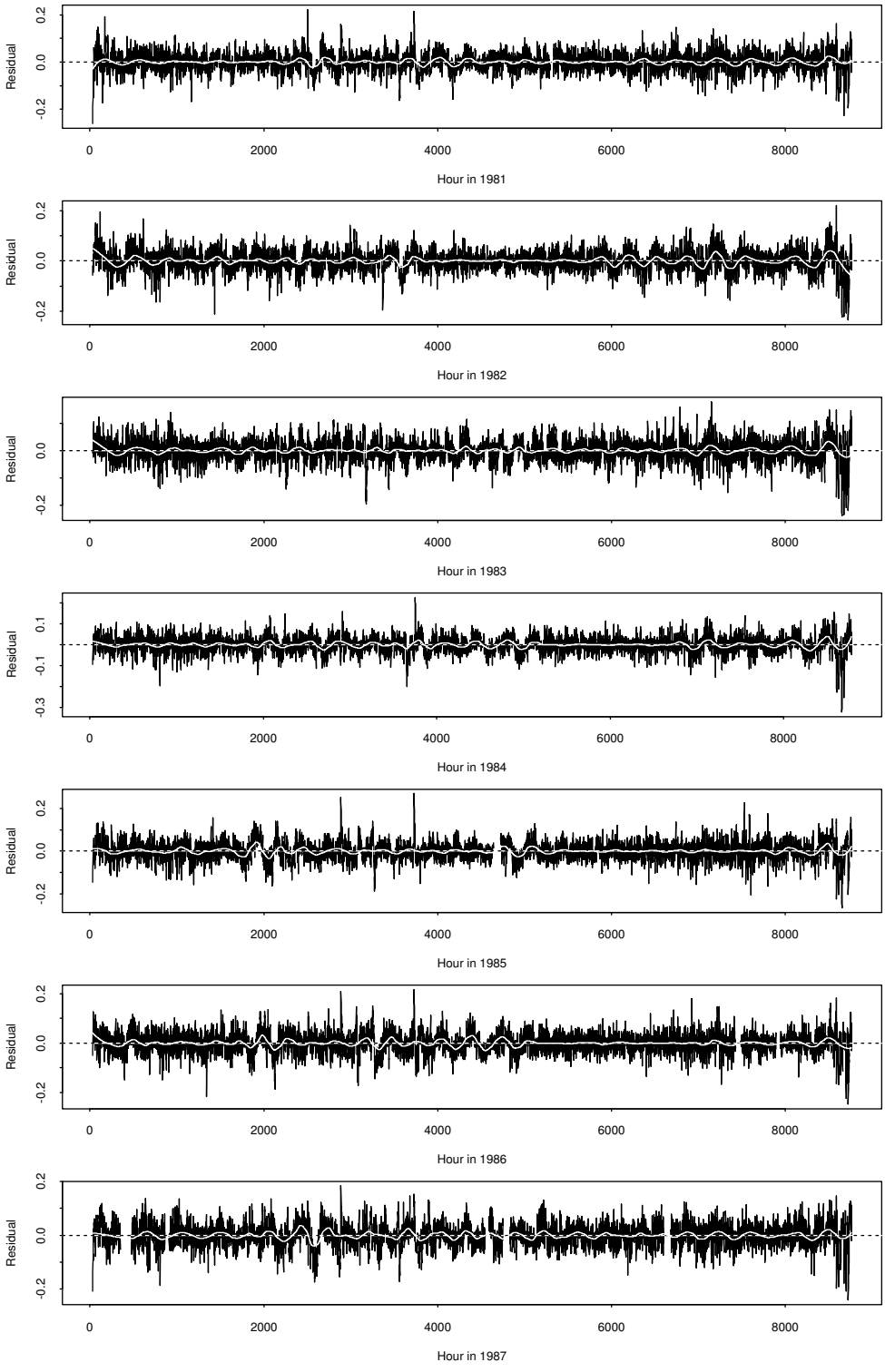
Appendix A

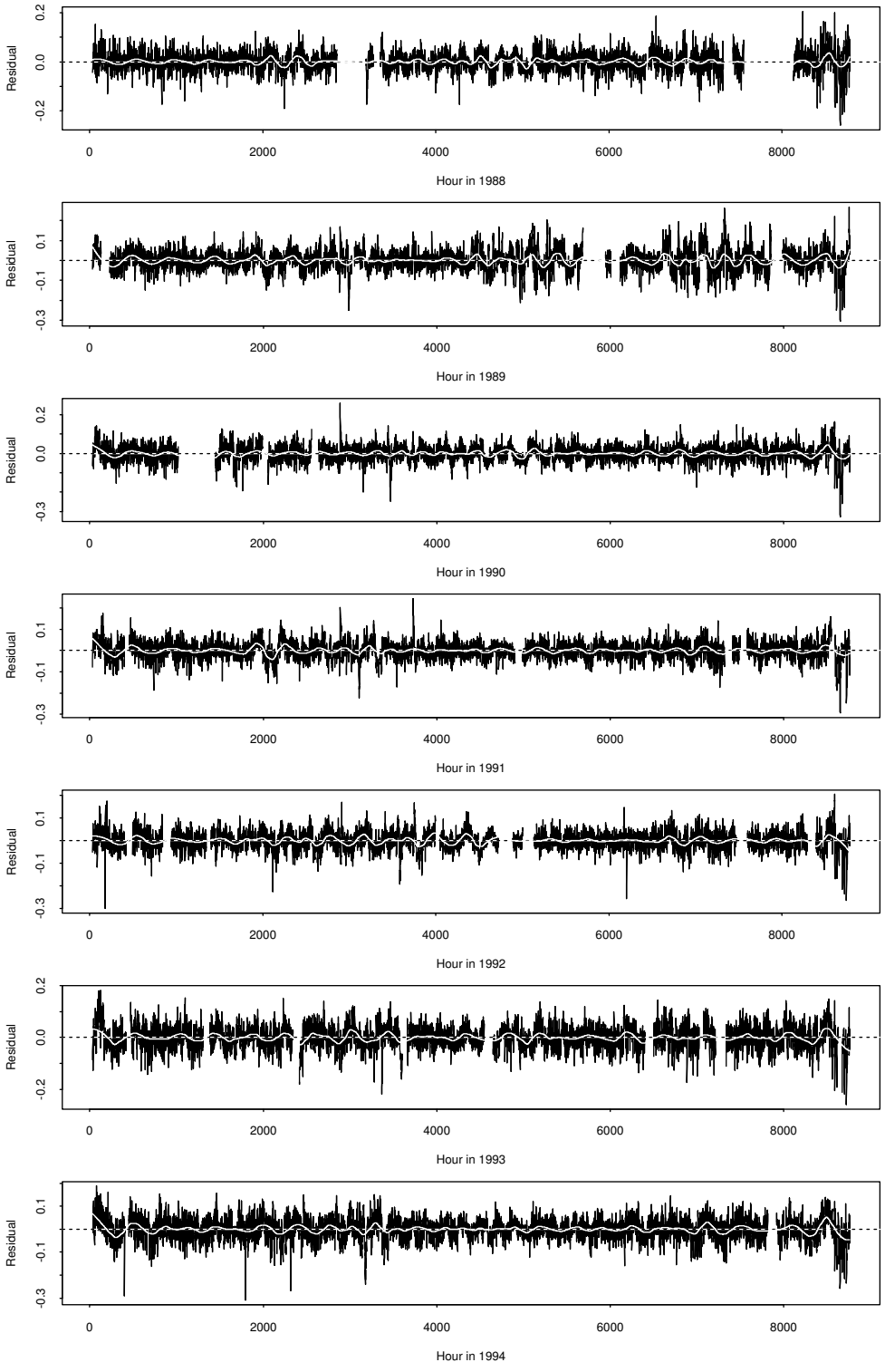
Plots of Residuals

A.1 Dynamic Temperature Response

On the following pages time-plots of the residuals from model (7.8) fitted separately for each year are shown. The unit of the residuals are GWh/h. Superimposed on the plots are a 2nd order local regression smoothing with a nearest neighbour bandwidth of 0.05 calculated by the `loess()` function of S-PLUS, see (Statistical Sciences, 1995). In order to be able to display occasional large residuals, the scale on the y-axes are not identical.

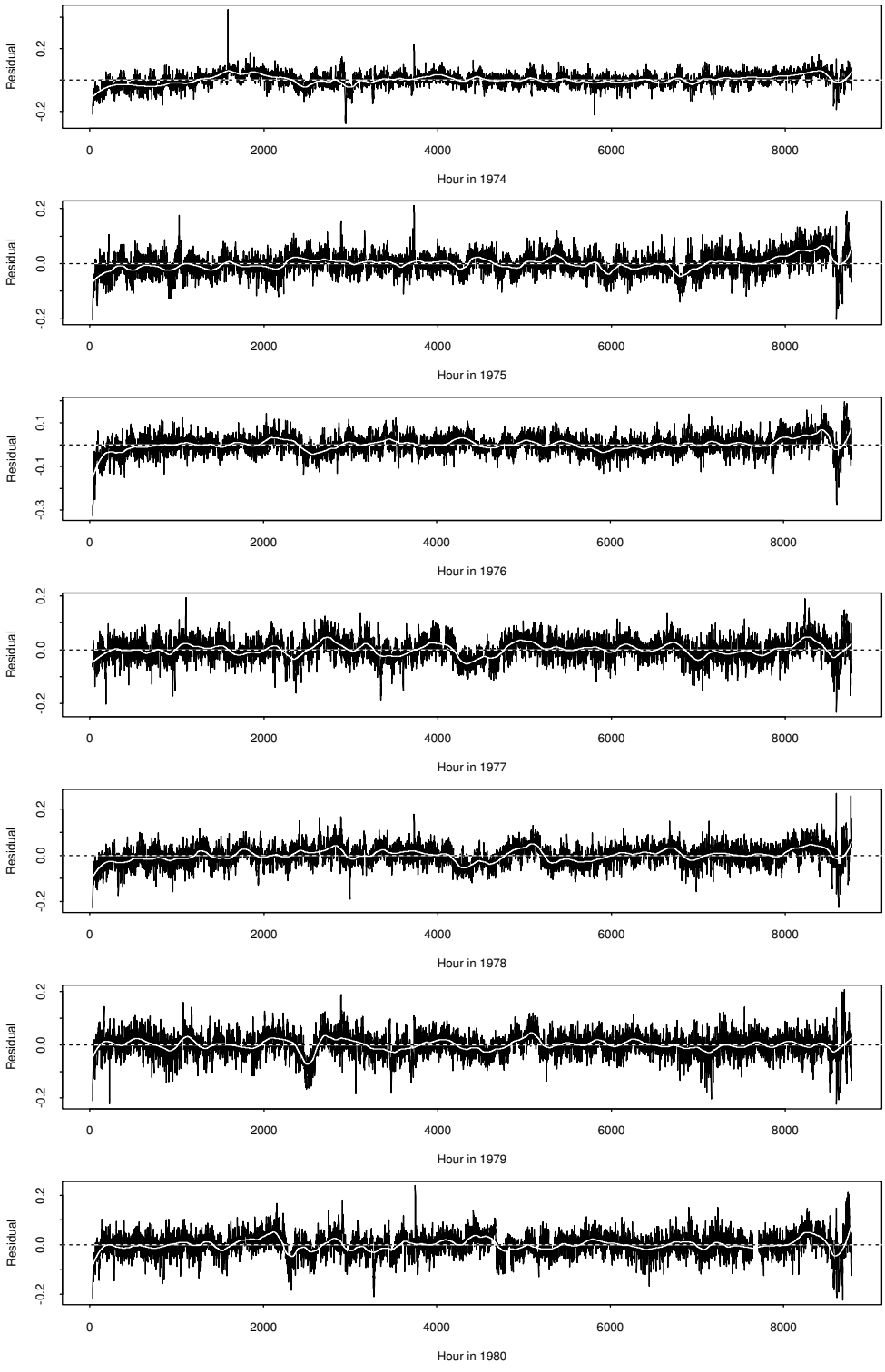

Residual
Hour in 1974
Hour in 1975
Hour in 1976
Hour in 1977
Hour in 1978
Hour in 1979
Hour in 1980
0
2000
4000
6000
8000

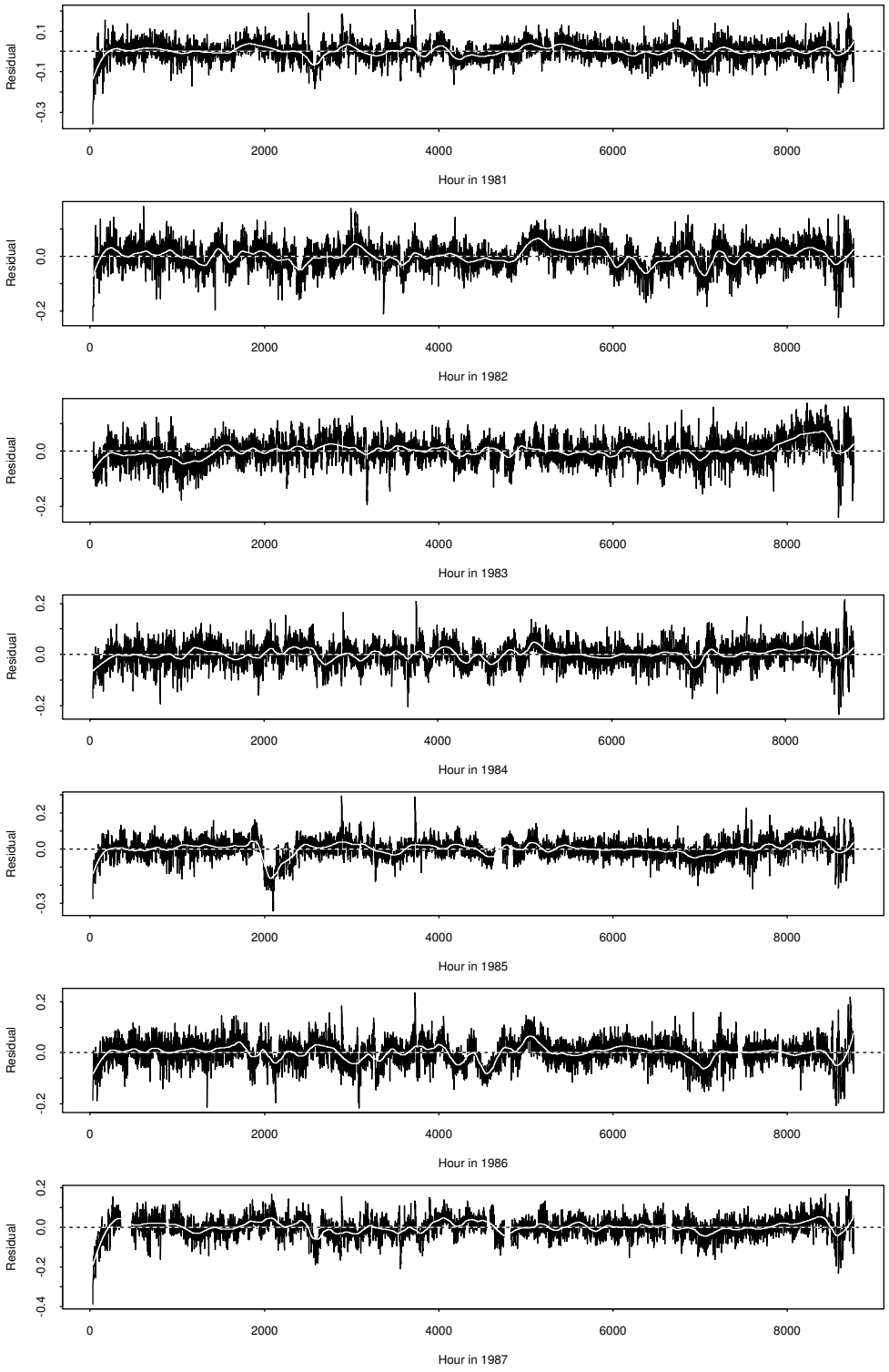

Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1981
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1982
Residual
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1983
Residual
0.1
-0.1
-0.3
0
2000
4000
6000
8000
Hour in 1984
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1985
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1986
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1987

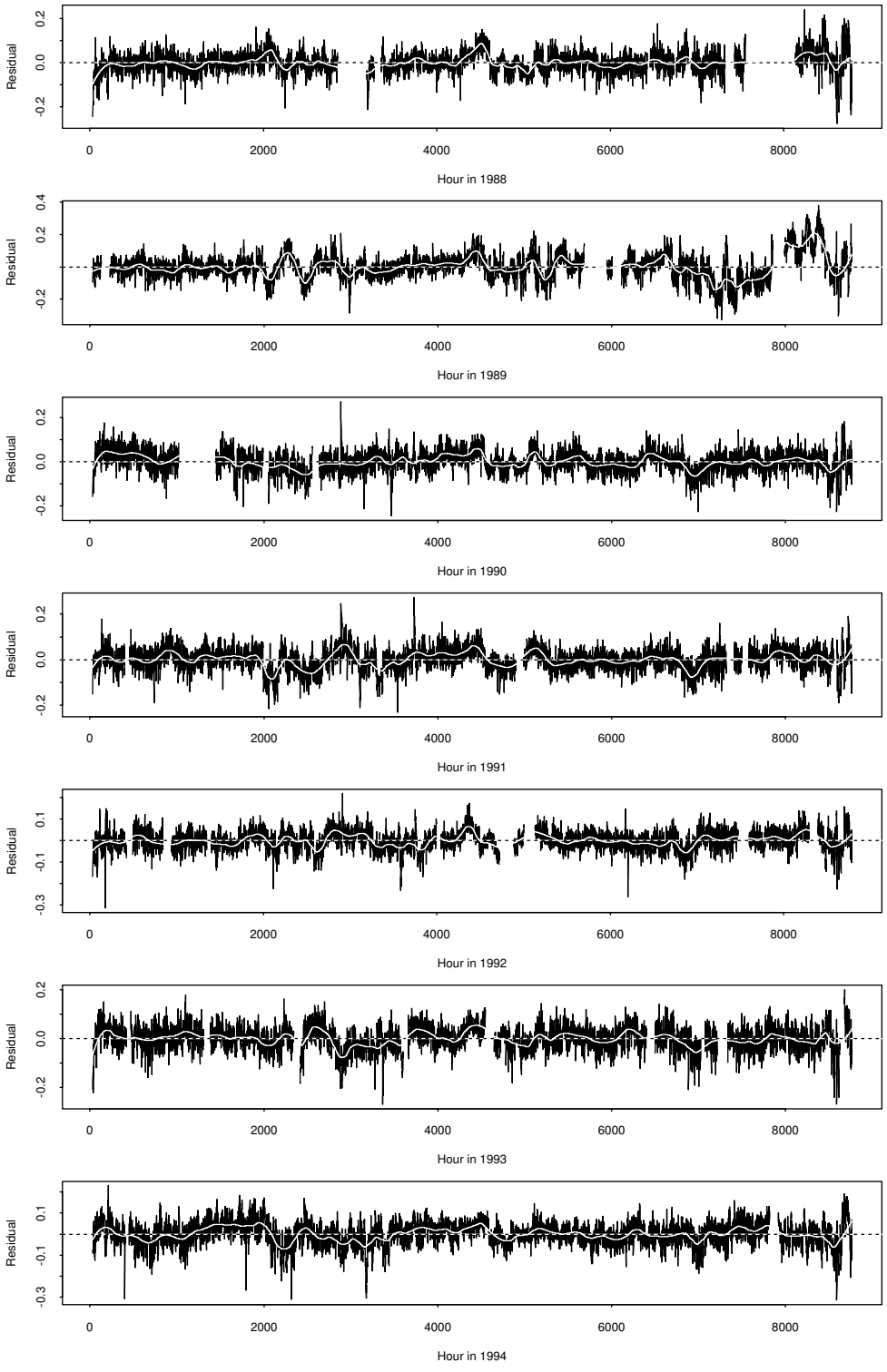

Residual
Hour in 1988
Hour in 1989
Hour in 1990
Hour in 1991
Hour in 1992
Hour in 1993
Hour in 1994
0
2000
4000
6000
8000

A.2 Model with Simplified Yearly Profile

On the following pages time-plots of the residuals from model (7.9) fitted separately for each year are shown. The unit of the residuals are GWh/h. Superimposed on the plots are a 2nd order local regression smoothing with a nearest neighbour bandwidth of 0.05 calculated by the `loess()` function of S-PLUS, see (Statistical Sciences, 1995). In order to be able to display occasional large residuals, the scale on the y-axes are not identical. Note that for 1989 the model was found to be singular.


Residual
0.2
-0.2
0
2000
4000
6000
8000
Hour in 1974
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1975
Residual
0.1
-0.1
-0.3
0
2000
4000
6000
8000
Hour in 1976
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1977
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1978
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1979
Residual
0.2
0.0
-0.2
0
2000
4000
6000
8000
Hour in 1980


Residual
-0.3
-0.1
0.1
0
2000
4000
6000
8000
Hour in 1981
Residual
-0.2
0.0
0
2000
4000
6000
8000
Hour in 1982
Residual
-0.2
0.0
0
2000
4000
6000
8000
Hour in 1983
Residual
-0.2
0.0
0.2
0
2000
4000
6000
8000
Hour in 1984
Residual
-0.3
0.0
0.2
0
2000
4000
6000
8000
Hour in 1985
Residual
-0.2
0.0
0.2
0
2000
4000
6000
8000
Hour in 1986
Residual
-0.4
-0.2
0.0
0.2
0
2000
4000
6000
8000
Hour in 1987


Residual
Hour in 1988
Residual
Hour in 1989
Residual
Hour in 1990
Residual
Hour in 1991
Residual
Hour in 1992
Residual
Hour in 1993
Residual
Hour in 1994

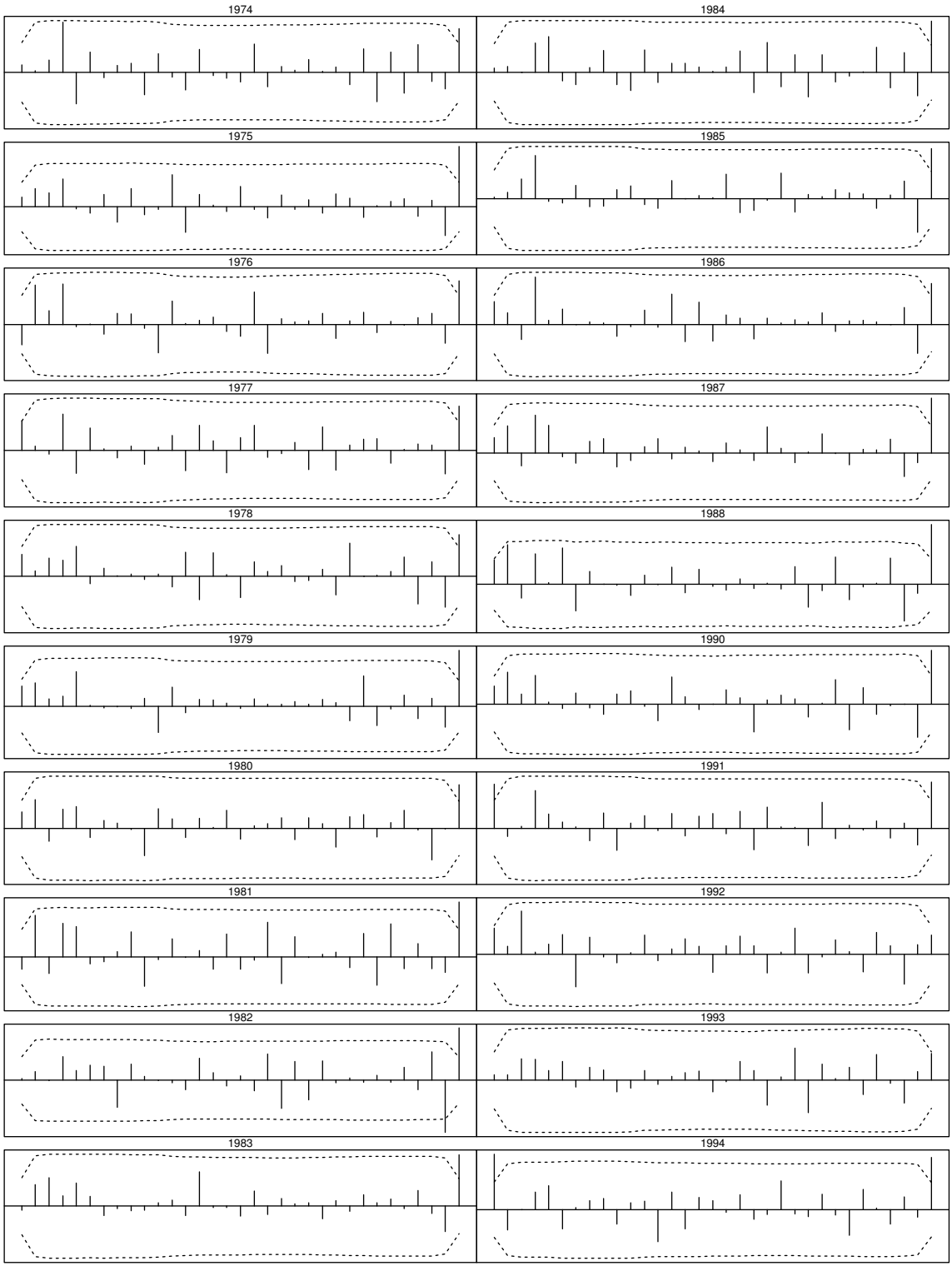
Appendix B

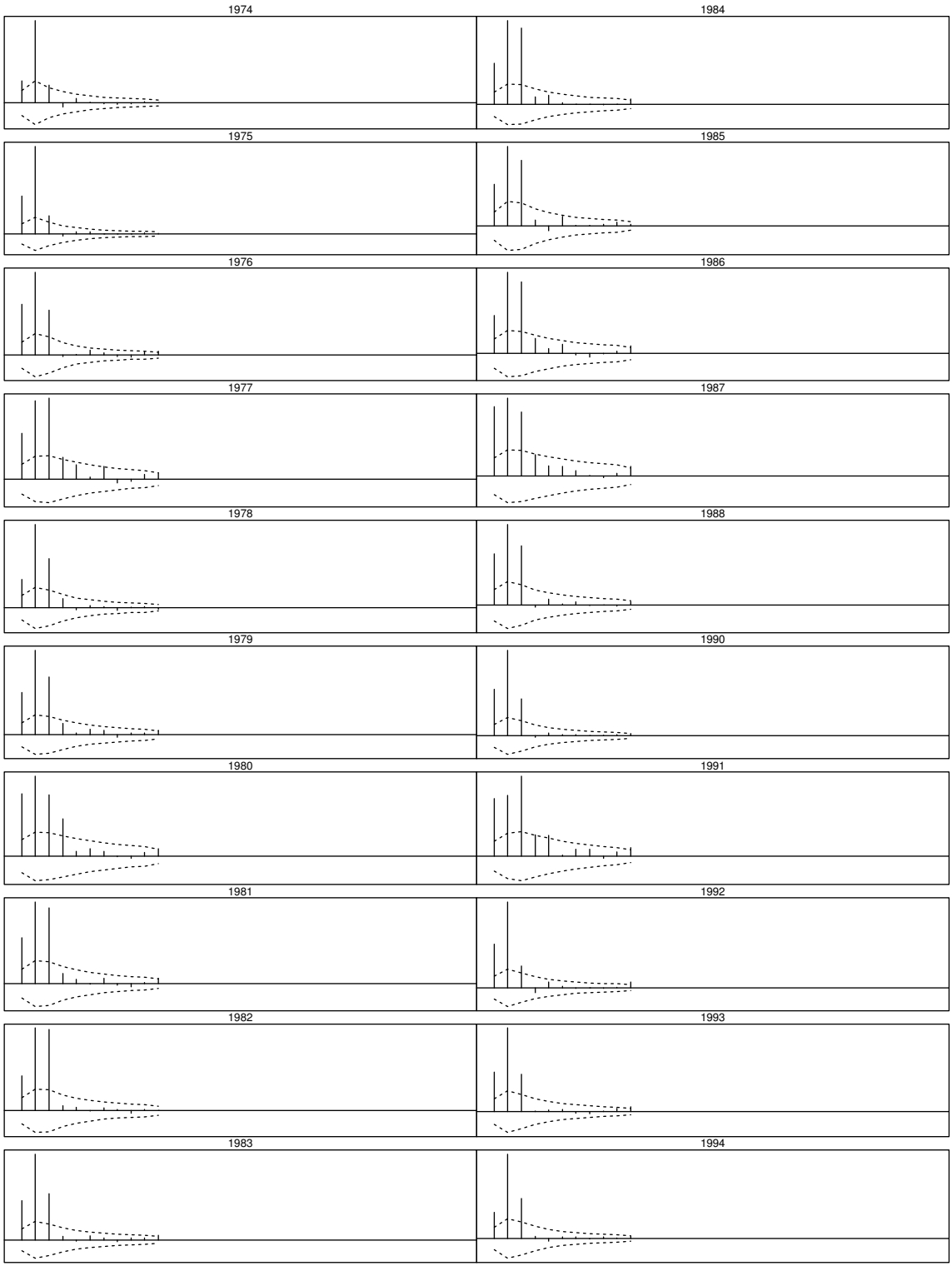
Dynamic Response on Temperature and Radiation

B.1 Filter Parameters

On the following two pages plots of the estimated filter parameters of model (7.10) are shown. Furthermore approximate 95% confidence intervals, calculated under the assumption that the model is correct and the true values zero, are indicated by dotted lines. The first page shows the estimates corresponding to the air temperature. The second page shows the estimates corresponding to the global radiation. Sizes and units on the axes are not shown.

The estimates for 1989 has been excluded since the model becomes singular this year.

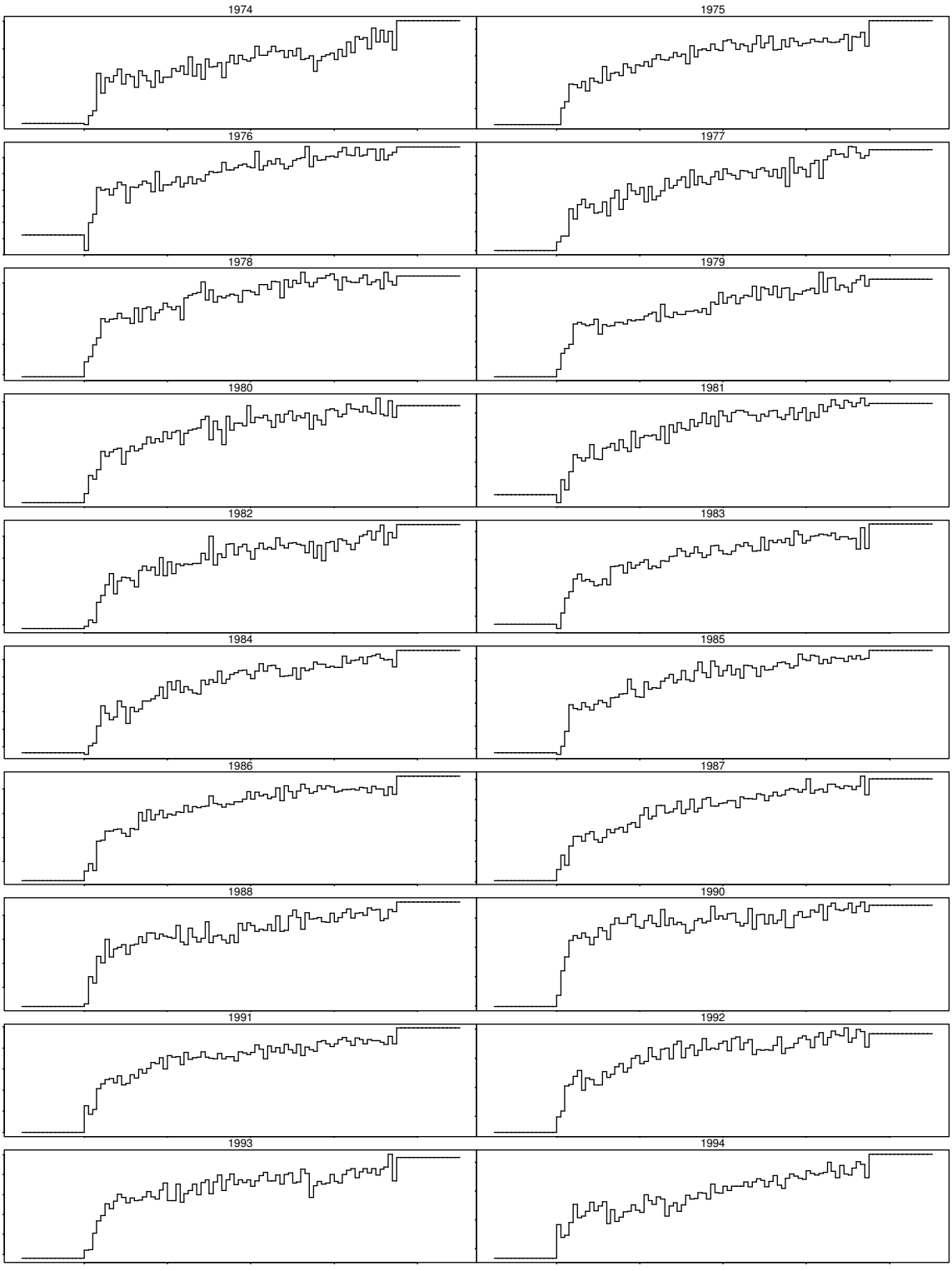

1974
1984
1975
1985
1976
1986
1977
1987
1978
1988
1979
1990
1980
1991
1981
1992
1982
1993
1983
1994

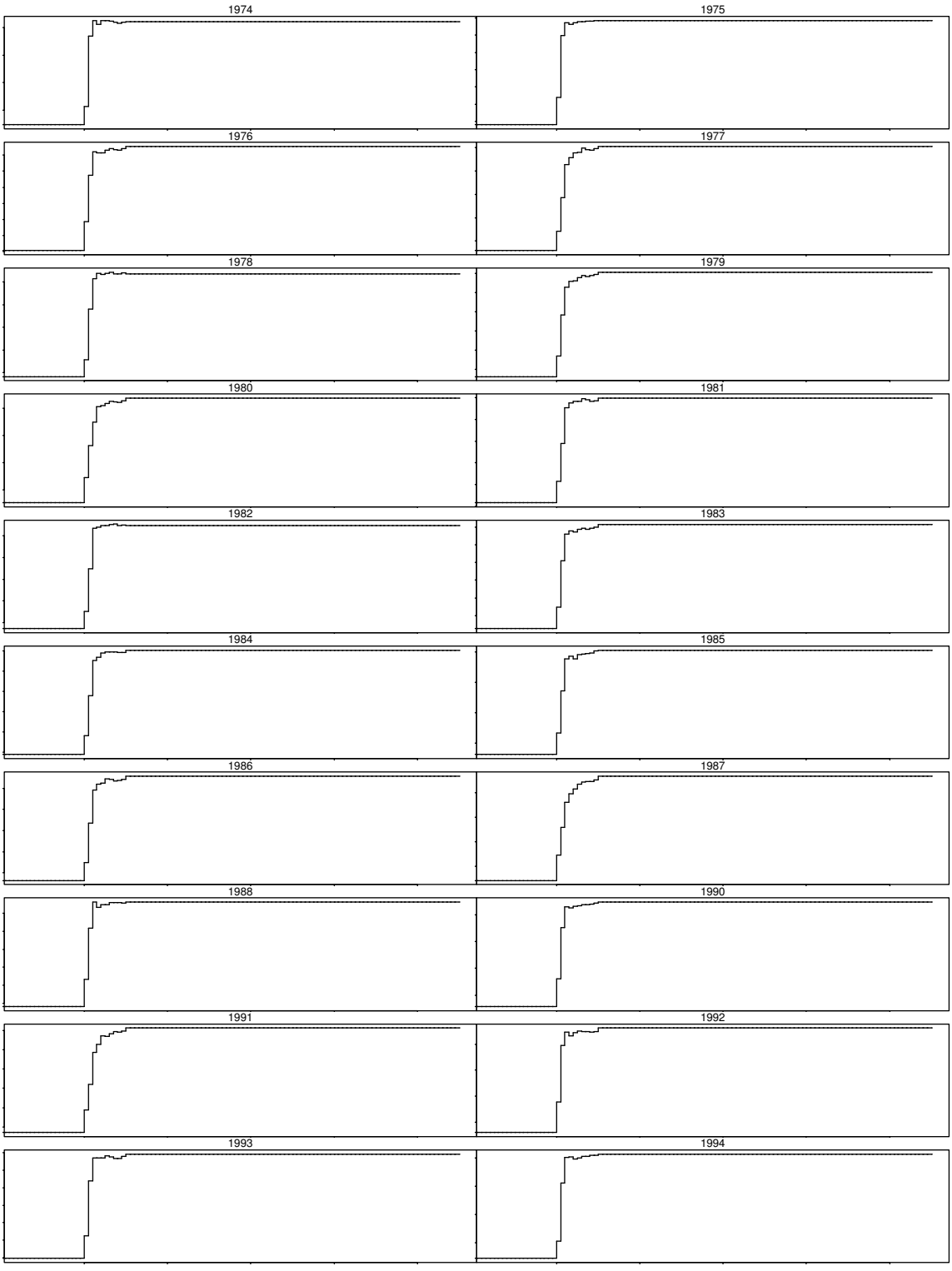

1974
1984
1975
1985
1976
1986
1977
1987
1978
1988
1979
1990
1980
1991
1981
1992
1982
1993
1983
1994

B.2 Step Response

On the following two pages plots of the estimated step response of model (7.10), with the order of the air temperature filter increased to 75, are shown. The air temparature drops from 0 to -1 $^{\circ}C$ at time $t = 0$. The global radiation drops from 25 to 20 W/m^2 at time $t = 0$. The first page shows the estimates corresponding to the air temperature. The second page shows the estimates corresponding to the global radiation. Sizes and units on the axes are not shown. The units on the y-axes are *not* identical between plots since these plots are included to highlight the dynamic characteristics.

The estimates for 1989 has been excluded since the model becomes singular this year.


1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1990
1991
1992
1993
1994

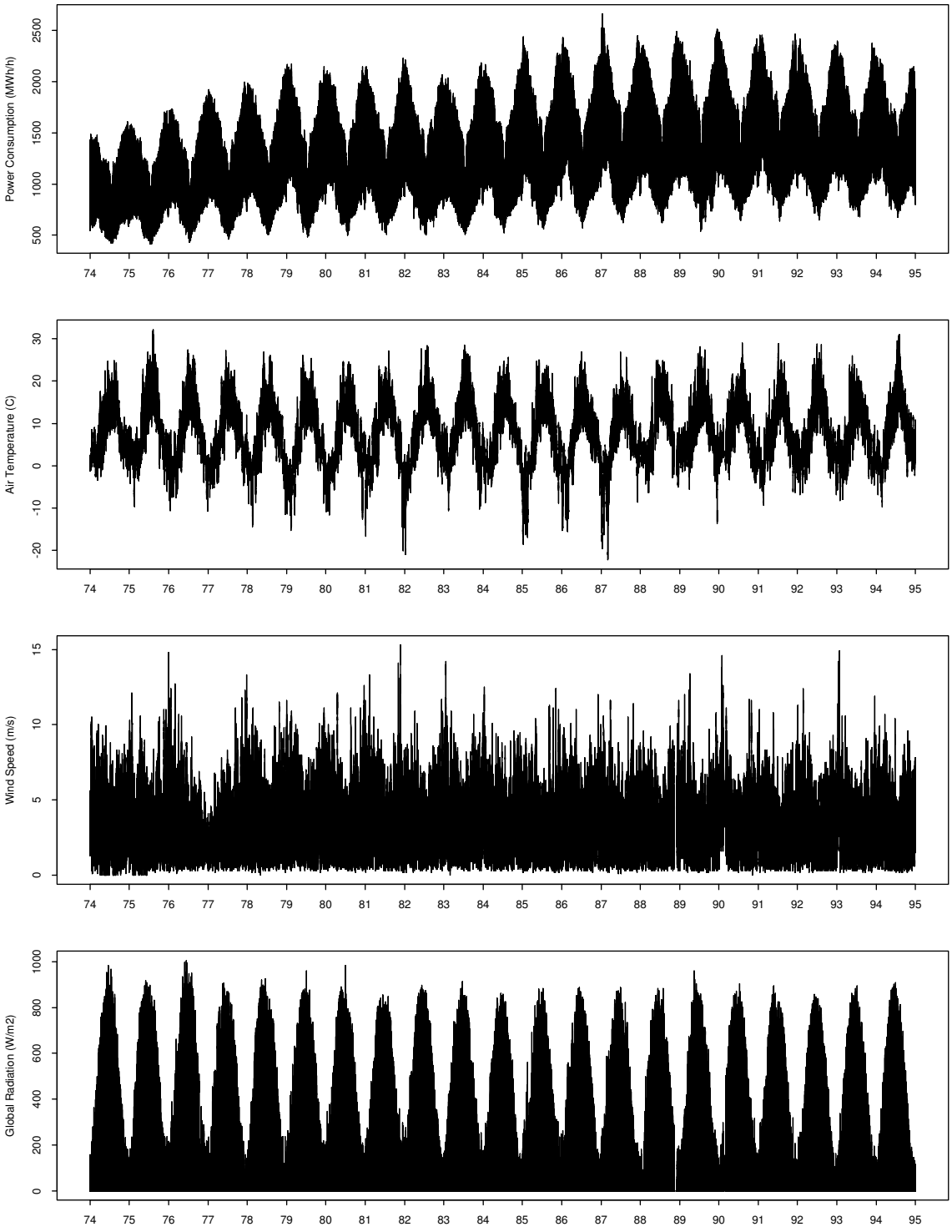

1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1990
1991
1992
1993
1994

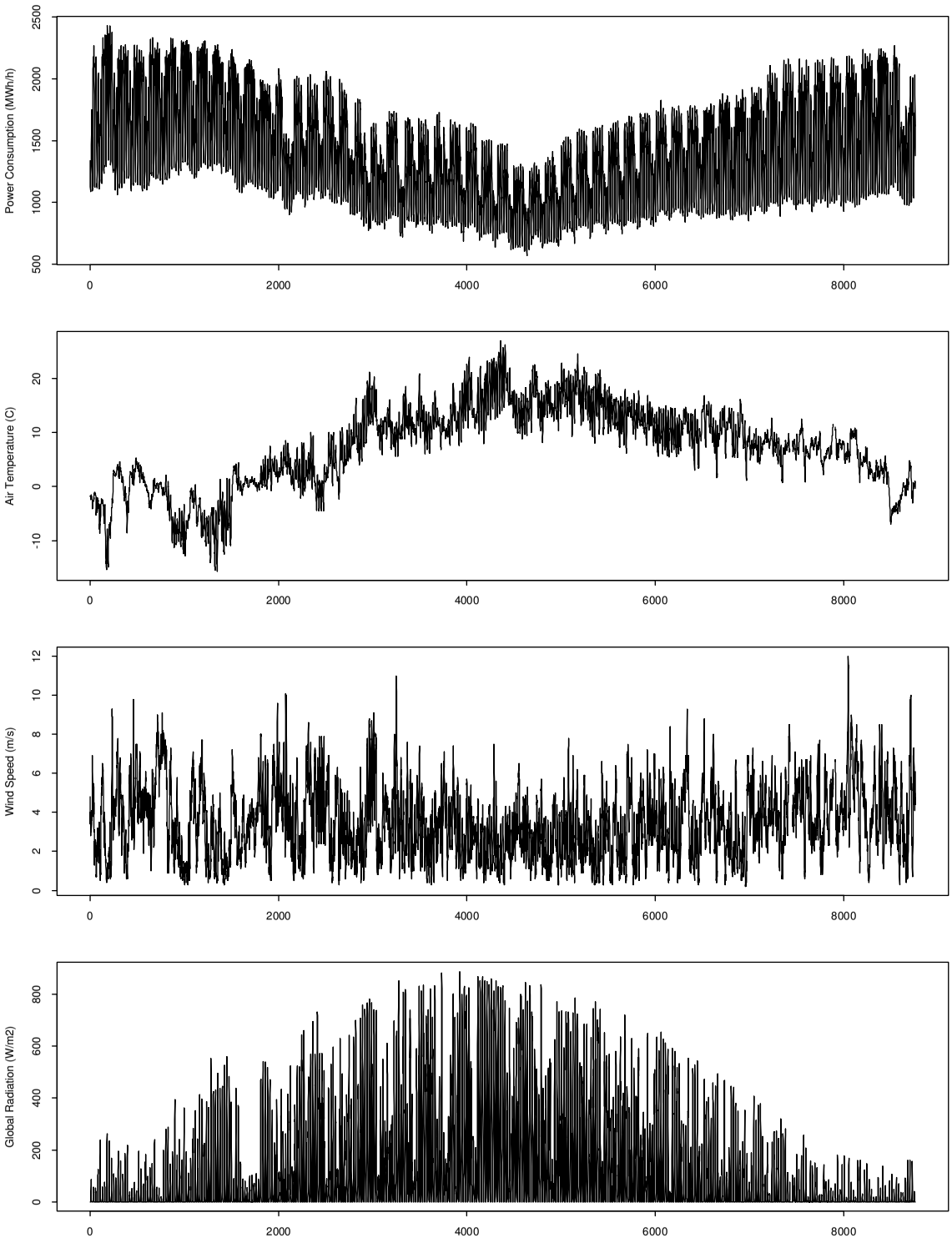
Appendix C

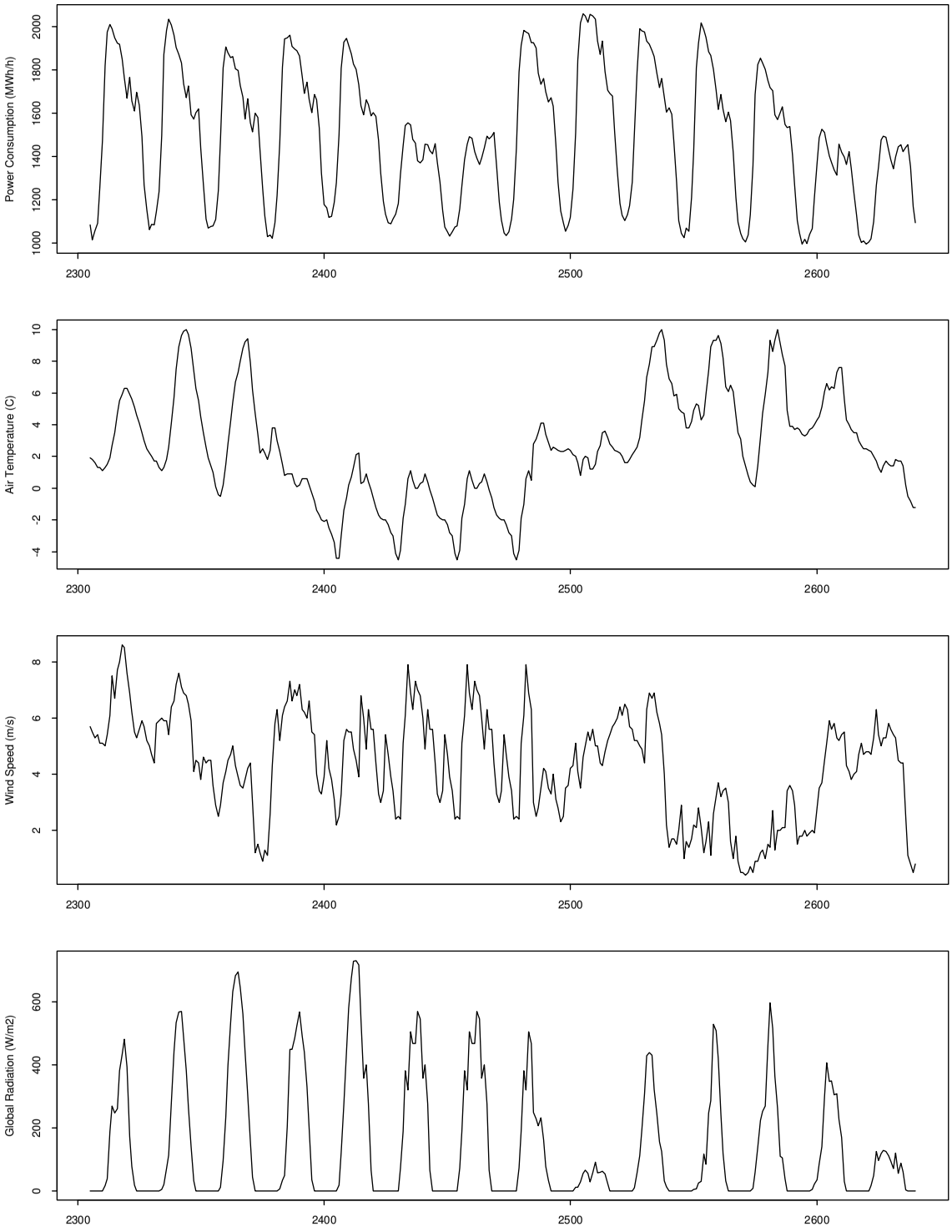
Plots of Power Consumption and Climate Data

On the following page time plots of the power consumption and climate data for the years 1974 to 1994 are shown. The unit on the x-axis is the the year. The power consumption plotted at time t is the power consumption over the time interval $t-1$ to t. The climate data plotted at time t is the average over the interval $t-1$ to t. The climate data were recorded by KVL at Højbakkegaard, see (Jensen, 1995). The plots were produced after the corrections described in Chapter 2 were implemented.

To illustrate details of the data two more pages have been included. The first shows data from 1986, and the second shows data from Monday 7 April to Sunday 20 April, 1986 (two weeks). On these pages the unix on the x-axis is hours since the start of 1986.


Power Consumption (MWh/h)
500
1000
1500
2000
2500
Air Temperature (C)
-20
-10
0
10
20
30
Wind Speed (m/s)
0
5
10
15
Global Radiation (W/m2)
0
200
400
600
800
1000
74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95


Power Consumption (MWh/h)
500
1000
1500
2000
2500
0
2000
4000
6000
8000
Air Temperature (C)
-10
0
10
20
0
2000
4000
6000
8000
Wind Speed (m/s)
0
2
4
6
8
10
12
0
2000
4000
6000
8000
Global Radiation (W/m2)
0
200
400
600
800
0
2000
4000
6000
8000


Power Consumption (MWh/h)
1000
1200
1400
1600
1800
2000
2300
2400
2500
2600
Air Temperature (C)
-4
-2
0
2
4
6
8
10
2300
2400
2500
2600
Wind Speed (m/s)
2
4
6
8
2300
2400
2500
2600
Global Radiation (W/m2)
0
200
400
600
2300
2400
2500
2600

Bibliography

Andersen, B. (1986). Danish test reference year, TRY Meteorological data for HVAC and energy. Note 174. Thermal Insulation Laboratory, Department of Buildings and Energy, Technical University of Denmark, Lyngby.

Box, G. E. P. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70:70–79.

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

Gallant, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York/Chichester.

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

Jensen, S. E. (1995). Agroclimate at Taastrup 1995. Agrohydrology and Bioclimatology. Department of Agricultural Sciences. The Royal Vet. and Agric. Univ., Copenhagen.

Kendall, M. G. and Stuart, A. (1961). *The advanced theory of statistics*, volume 2: Inference and relationship. Charles Griffin, London.

Madsen, H. (1985). *Statistically determined dynamical models for climate processes*. PhD thesis, Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby.

Madsen, H. and Holst, J. (1995). Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings*, 22:67–79.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic, New York/London.

DEF (1995). 10 year survey of the electricity supply (in danish: Elforsyningens tiårsoversigt). Association of Danish Electric Utilities (Danske Elværkers Forening), Copenhagen.

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

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.

Miller, A. J. (1992). [Algorithm AS 274] Least squares routines to supplement those of Gentleman. *Applied Statistics*, 41:458–478. (Correction: 94V43 p678).

Myers, R. H. (1986). *Classical and Modern Regression With Applications.* Duxbury Press, North Scituate, MA.

Thorsted, C. (1993). Long term forecast of hourly power consumption (in danish: Langtidsprognose for elbelastning på timebasis). Department of Mathematical Modelling, Technical University of Denmark, Lyngby.