Estimation of UA-values for single-family houses

January 16, 2008

ENFOR

Forecasting and Optimization for the Energy Sector

ENFOR A/S
Lyngsø Allé 3
DK-2970 Hørsholm
www.enfor.dk

Document ID: 03EKS0001A003-A

Preface

The ideas, methods, and results that are presented in this report have been discussed during several meetings between ENFOR, and Søren Østergaard Jensen, Danish Technological Institute, and Henrik Madsen, Informatics and Mathematical Modeling, Technical University of Denmark. The results have also been discussed with the project's advisory panel, consisting of Hans Skyum Larsen, IT Energy, Bjarne Olesen, Department of Mechanical Engineering, Technical University of Denmark, Poul Erik Pedersen, Elsparefonden, Carsten Rode, Department of Civil Engineering, Technical University of Denmark, and Kim Wittchen, Danish Building Research Institute.

ENFOR would like to acknowledge the mentioned participants for their contributing comments.

Contents

1 Introduction

This report considers estimation of UA-values for single-family houses, using measurements of energy consumption, ambient air temperature, and possibly indoor temperature. The data originates from "det jysk-fynske forbrugs måleprogram" and dates back to the early 1990s. The full data set consists of measurements from 25 houses plus a single wind-speed measurement for part of the period. All houses were supplied with electrical heating.

2 Data

The houses without a wood burning stove and not using night-time drop are numbered internally 1, 9, 20, and 24. Data are collected every 15 minutes over the period from Feb. 01, 1991 until May. 01, 1993. The measurements are:

`total`: Total electricity consumption in kWh over the previous 15 minutes.

`heat`: Electricity consumption of the heating equipment in kWh over the previous 15 minutes.

`water`: Electricity consumption for producing hot tap water in kWh over the previous 15 minutes.

`Ti`: Indoor temperature in ^oC representing the previous 15 minutes.

`Ta`: Ambient temperature in ^oC representing the previous 15 minutes.

`ws`: Wind speed in m/s measured 2m above the roof top of house number 11. These data are only available for part of the period and are not used here.

The data were investigated with the aim of identifying erroneous measurements. Initially, simple range checks were used, but due to the seasonality this type of validation will not catch all possible errors. For this resson the individual series were decomposed into an overall trend, a diurnal/weekly variation, and a remainder. The method used is described in [7, Sec. 3.3]. Following this, the remainder was inspected and a few extreme values were identified. Furthermore, a few occasions where the measurements freeze at a constant level were identified. Time series plots of data, excluding the outliers, can be found in Appendix A. Plots of daily averages can be found in Appendix A.

In all the cases mentioned, the outliers identified as described above were set to be missing and therefore excluded from the analysis. However, for the temperatures "filled", series without missing values were produced. The procedure used to produce these filled series, called `Tif` and `Taf`, is based on decomposition of the original series (with outliers set to be missing) as described above, followed by filling the remainder using an approximate conditional mean robust smoother based on a robustly estimated autoregressive model[1].

During the period, power supply for heating equipment (room and water) was disconnected for short periods of time. Furthermore, the price paid for electricity varied according to the time of day.

3 Analysis of daily averages

Due to the occasional disconnection of the power supply as described above and due to the changing price over the course of the day, the analysis starts by considering the daily averages of energy consumption (`total`, `heat`, and `water`) and temperatures (`Ta` and `Ti`).

3.1 Initial investigation of data

Figure 1 depicts the daily averages of heat consumption against the daily averages of the temperature differences for each of the four houses. Only data from the months Nov. to Feb. are used in the plots. Both plots based on the variables `heat` and the difference between variables `total` and `water` are supplied, the latter taking into account both the heat used by the heating equipment and the heat supplied by energy use not directly related to heating. Furthermore, two lines are overlayed on each plot; one line is the best linear fit to the points on the plot and the other is the best linear fit through the origin. In all but one case the slope of the best linear fit is lower than the slope of the best line through the origin. This implies that *extrapolation* of the best linear fit to temperature differences of 0^oC will result in positive heat consumption.

Figure 2 shows the same plots based on the remaining months. In this case the order of the slopes are reversed, implying negative heat consumption for temperature differences of 0^oC. From Figure 2 it is seen that the linear regression line intersects the horisontal line of 0 $kWh/15min$ at approximately 5^oC.

The negative heat consumption may originate from heat input not accounted for in the data, i.e. heat from persons living in the house and from solar radiation. Note however that the relation is not strictly linear during these months.

[1]Functions `acm.smo` and `ar.gm` in S-PLUS verdion 7 `www.insightful.com` were used for this purpose.

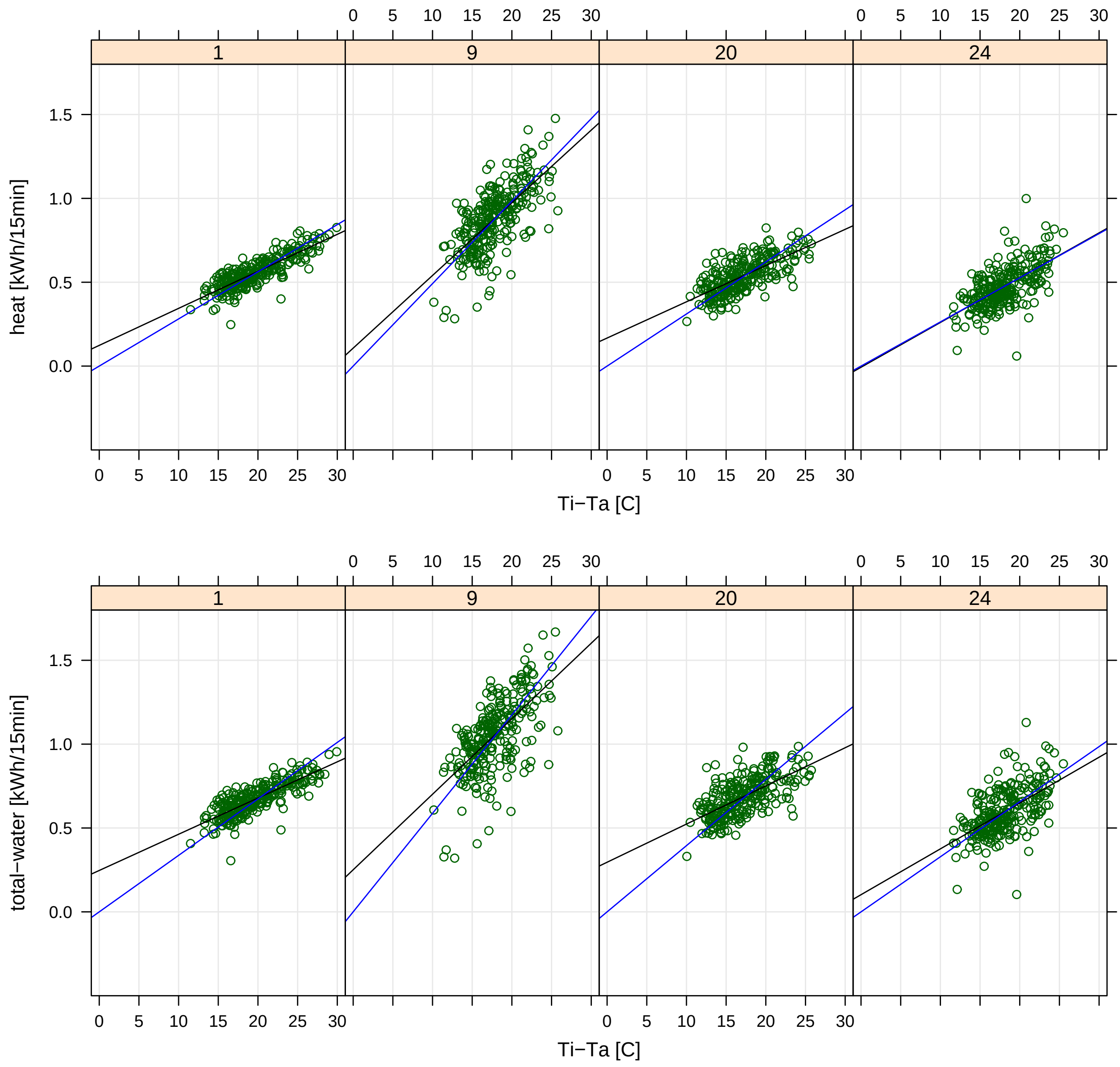


Figure 1: Daily averages of heat consumption ($kWh/15min$) versus temperature differences (^oC) for the four houses for months November, December, January, and February. The black line indicates the best linear fit and the blue line is the best linear fit through the origin.

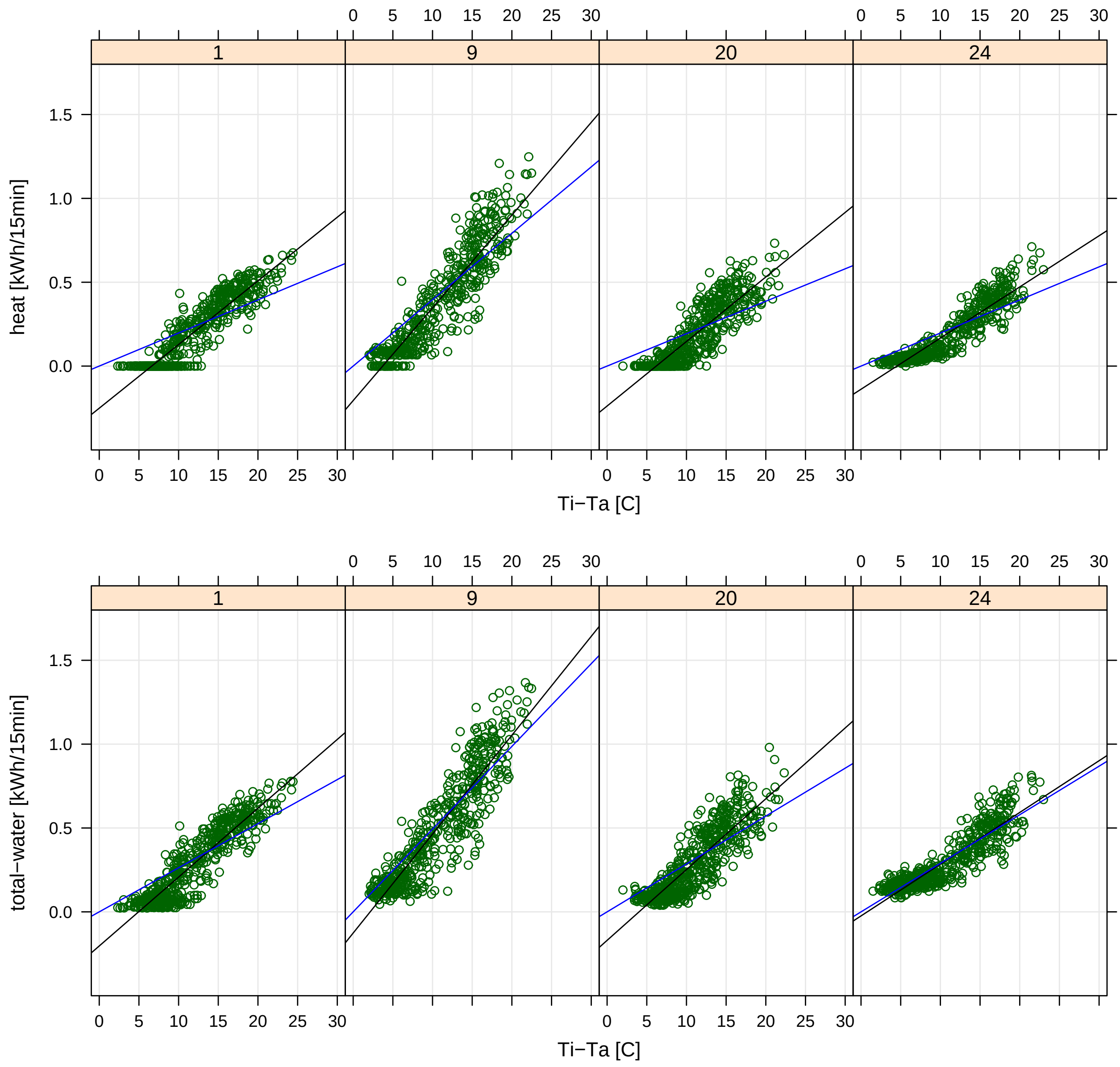


Figure 2: Daily averages of heat consumption ($kWh/15min$) versus temperature differences (^oC) for the four houses for months March to October. The black line indicates the best linear fit and the blue line is the best linear fit through the origin.

3.2 Estimating UA-values separately for each month.

With appropriately defined units[2], assuming that effects other than the ambient air temperature and the indoor temperature can be neglected when determining the heating requirement, and furthermore assuming that dynamic effects are neglible when considering daily averages, the following linear regression through the origin states the relation between heat consumption and temperatures

$$Q_t = UA(T_{i,t} - T_{a,t}) + e_t \,. \tag{1}$$

Where t is the time index indicating the day, Q is the heat consumption, T_i is the indoor air temperature, T_a is the ambient air temperature, UA is the UA-value, and e is the error term.

As indicated by the results presented in the previous sub-section the data indicates that when $T_{i,t} = T_{a,t}$ heat consumption is not zero. Especially not when only considering winter periods where the results corresponding to $T_{i,t} = T_{a,t}$ amounts to extrapolation, cf. Figure 1. For this reason we also consider an extension of model (1) to allow for a non-zero intercept Q_0:

$$Q_t = Q_0 + UA(T_{i,t} - T_{a,t}) + e_t \,. \tag{2}$$

The models (1) and (2) are fitted separately to each month of the data and the results are displayed for both the case where Q_t is defined as `heat` and the case where Q_t is defined as `total - water`. The resulting estimates of UA are shown in Figure 3. It is seen that the estimates are somewhat noisy and, generally, as expected the estimates for summer periods are unrealistically low. For model (1) the estimates based on `total-water` are generally higher than for the estimates based on `heat`. Comparing the estimates based on models (1) and (2) are consistent with the observations stated in Section 3.1. Actually, UA in model (2) is the marginal change in heat consumption when the difference between indoor and ambient air temperature change 1^oC. Generally, there is a tendency for this value to be relatively high during autumn and spring, while stabilising at a lower level during winter.

The estimates presented in Figure 3 and used until this point are so-called Least Squares (LS) estimates, which is the usual estimation method for linear regression. LS-estimates are defined as the set of estimates minimising the sum of the the squared errors on the dataset and therefore LS-estimates place rather significant emphasis on large errors. If the data contain relatively many large errors compared to the normal distribution, the estimates may be too sensitive to these errors. Instead a so-called robust estimation method may be used. Here the estimates obtained when minimising the sum of the absolute errors, i.e. $\sum |e_t|$, are considered. This method is sometimes called L1-regression or median regression, since it is a special case of quantile regression [4]. Results obtained using this method are displayed in Figure 4. Based on comparison of Figures 3 and 4, it is concluded that in most cases the differences between the two estimation methods are marginal. Consequently, throughout the rest of this report only LS-type estimates will be considered since these are simpler to calculate than robust estimates.

[2] If the heat consumption (power) is measured $kWh/15min$ and the temperatures in ^oC the slope of the linear regression as e.g. in Figure 2 must be multiplied by 4000 in order to be given in the unit $W/^oC$.

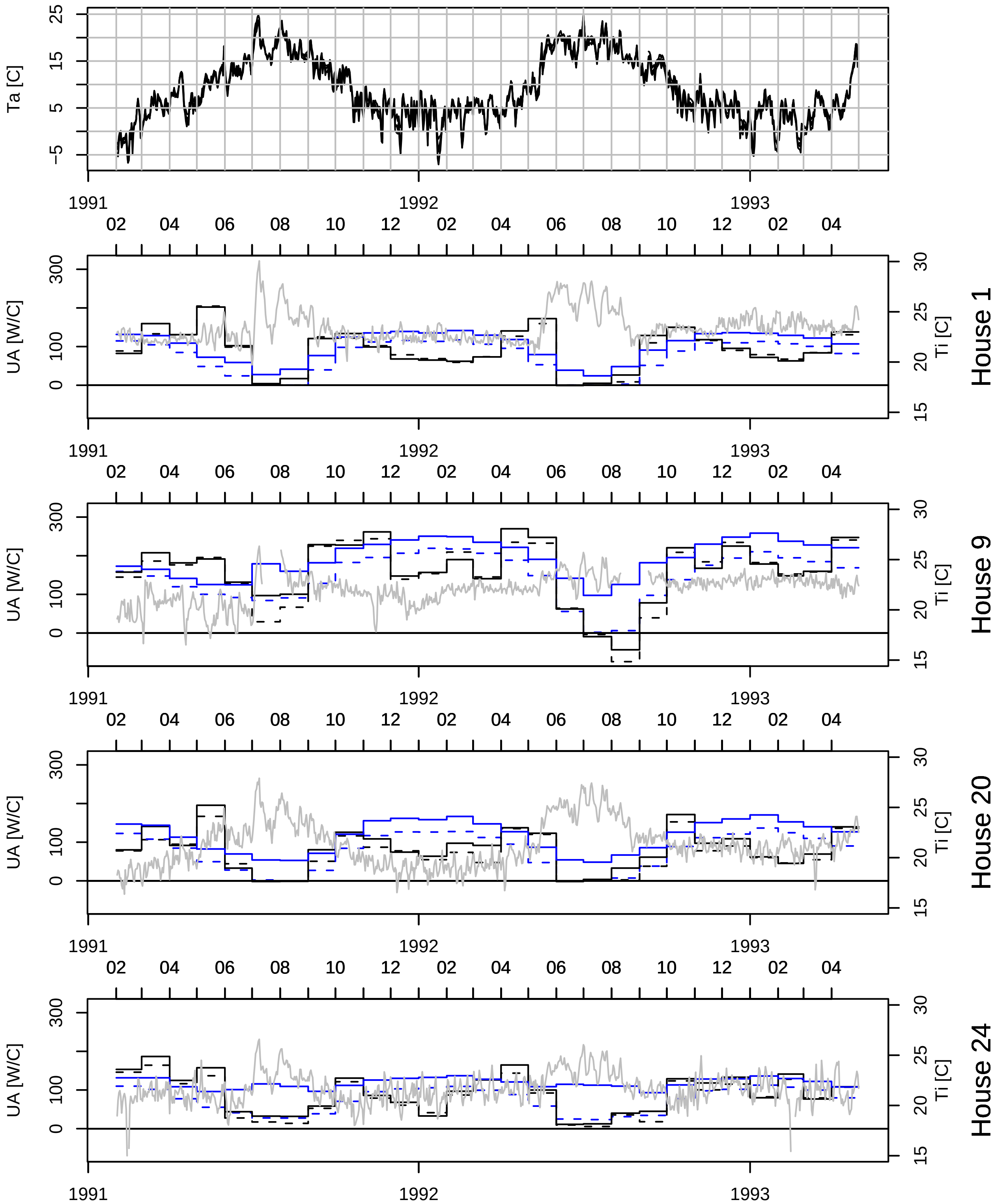


Figure 3: Monthly estimates based on daily averages (black: linear regression, blue: linear regression through the origin, dotted: 'heat' as the response, full: 'total-water' as the response).

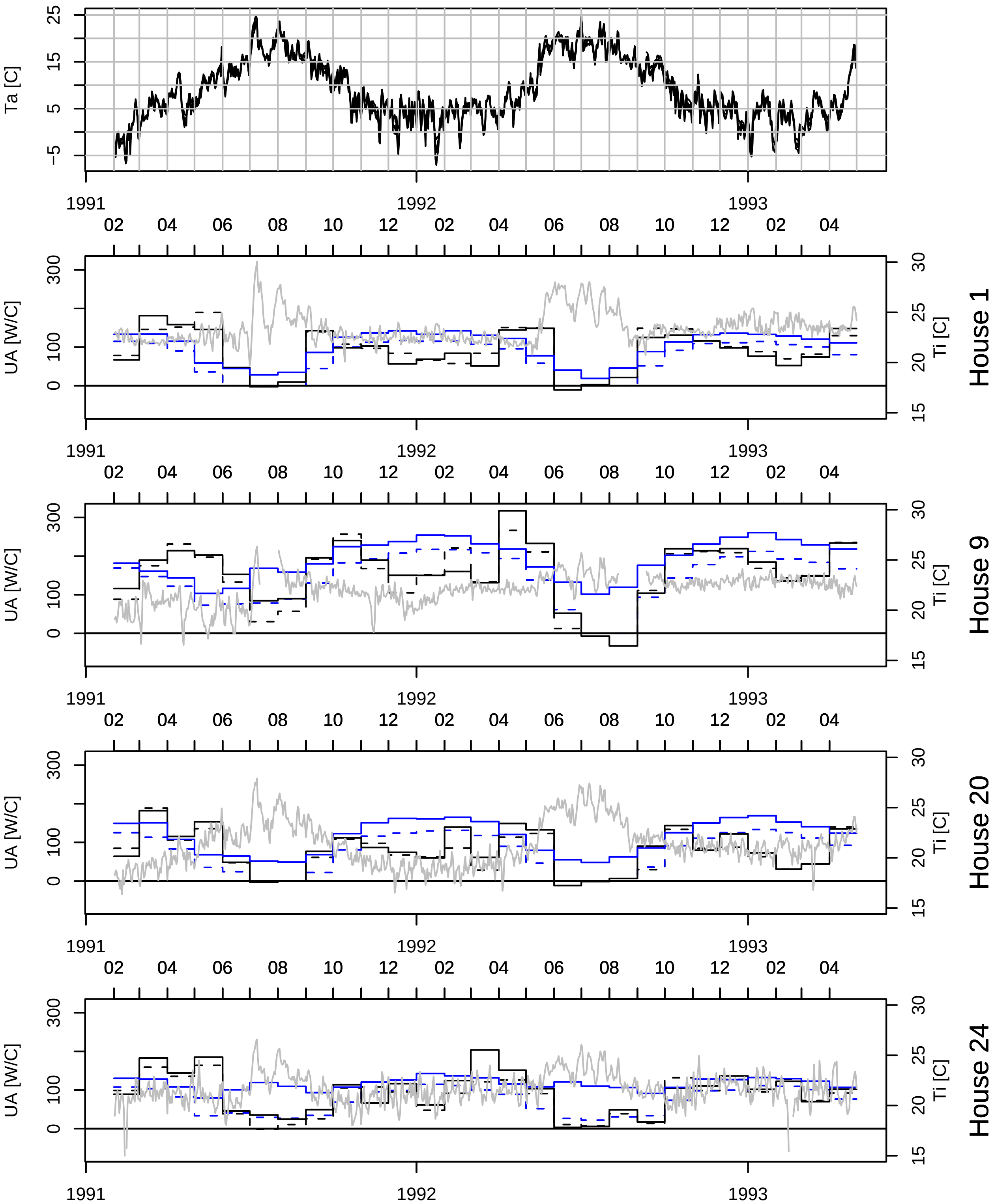


Figure 4: MEDIAN REGRESSION: Monthly estimates based on daily averages (black: linear regression, blue: linear regression through the origin, dotted: 'heat' as the response, full: 'total-water' as the response).

In the above it has been assumed that, when considering daily averages, the dynamic effect of temperature differences can be neglected. Applying the usual method of pre-whitening and estimation of the cross correlation function [5] to part of the period results in the estimates of impulse response as depicted in Figure 5. As described in the caption of the figure, the horizontal lines indicate approximate 95% confidence intervals for uncorrelated series. In order to judge the significance of a given lag, the estimated values are compared to the bounds of the confidence interval. As seen from the plots there is a weak tendency of dependence on the temperature levels on the previous day. However, this is not true for house 24.

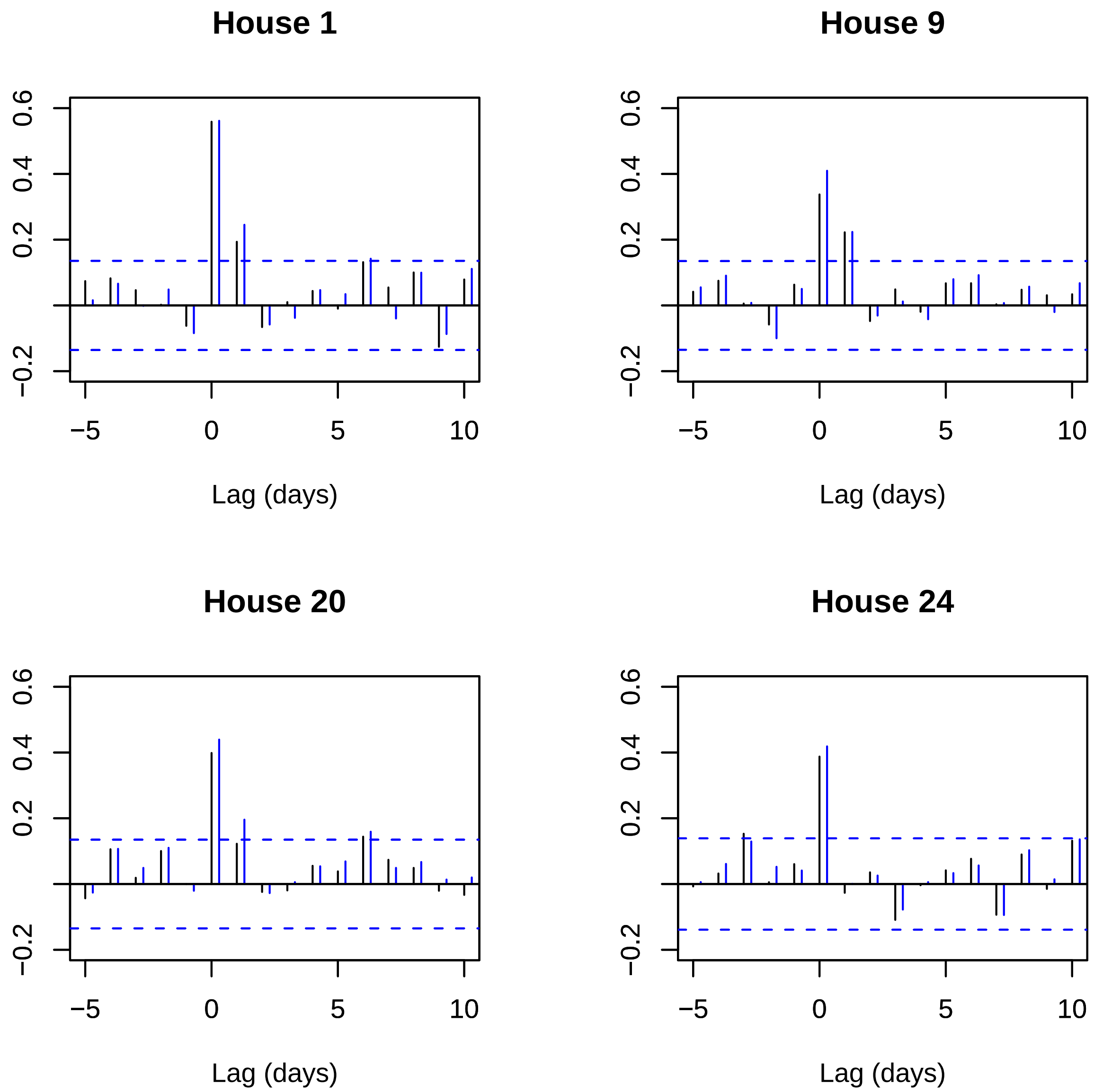


Figure 5: Dynamic response of 'total-water' (black) or 'heat' (blue) on $T_i - T_a$ when using data covering the period from Oct. 01, 1992 to Apr. 30, 1993. The estimates are found as the sample cross correlation of pre-whited series and thus the estimates are reported as numbers proportional to the impulse response. The horizontal lines indicate approximate 95% confidence intervals for uncorrelated series. Pre-whitening is performed using the best AR-model (in AIC-sense) for the input series $(T_i - T_a)$.

Extending the model to include the average temperature difference on the previous day also results in the following model including a constant term

$$Q_t = Q_0 + c_0(T_{i,t} - T_{a,t}) + c_1(T_{i,t-1} - T_{a,t-1}) + e_t \, , \tag{3}$$

where c_0 and c_1 are regression coefficients. Since the stationary response to a change in temperature is $c_0 + c_1$ this quantity is interpreted as the UA-value.

Figures 6 and 7 show the estimates of $UA = c_0 + c_1$ and Q_0, respectively, for both the case described and when c_1 is fixed at 0. It is seen that for houses 1 and 9, which based on Figure 5 have the most clear dynamic dependence, the estimates of UA obtained from model (3) are generally larger than those obtained when using the non-dynamic model. For houses 20 and 24 the estimates obtained using the two approaches are similar.

Figure 8 shows the estimates of UA obtained when excluding the constant term from the model. It is seen that the estimates are very similar in this case.

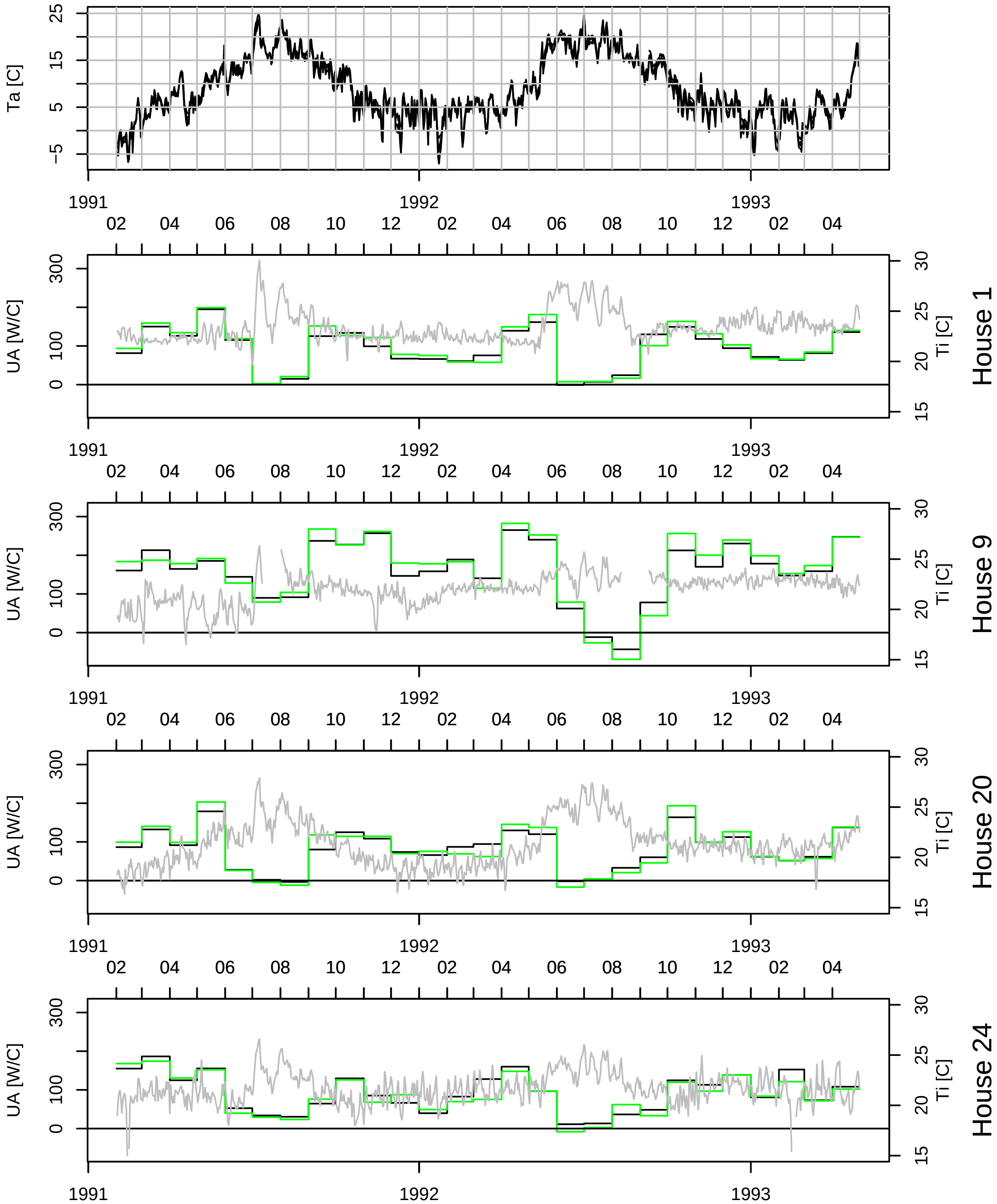


Figure 6: Monthly estimates of UA based on daily averages and using 'total-water' as the dependent variable (black: linear regression, green: linear regression including previous day also).

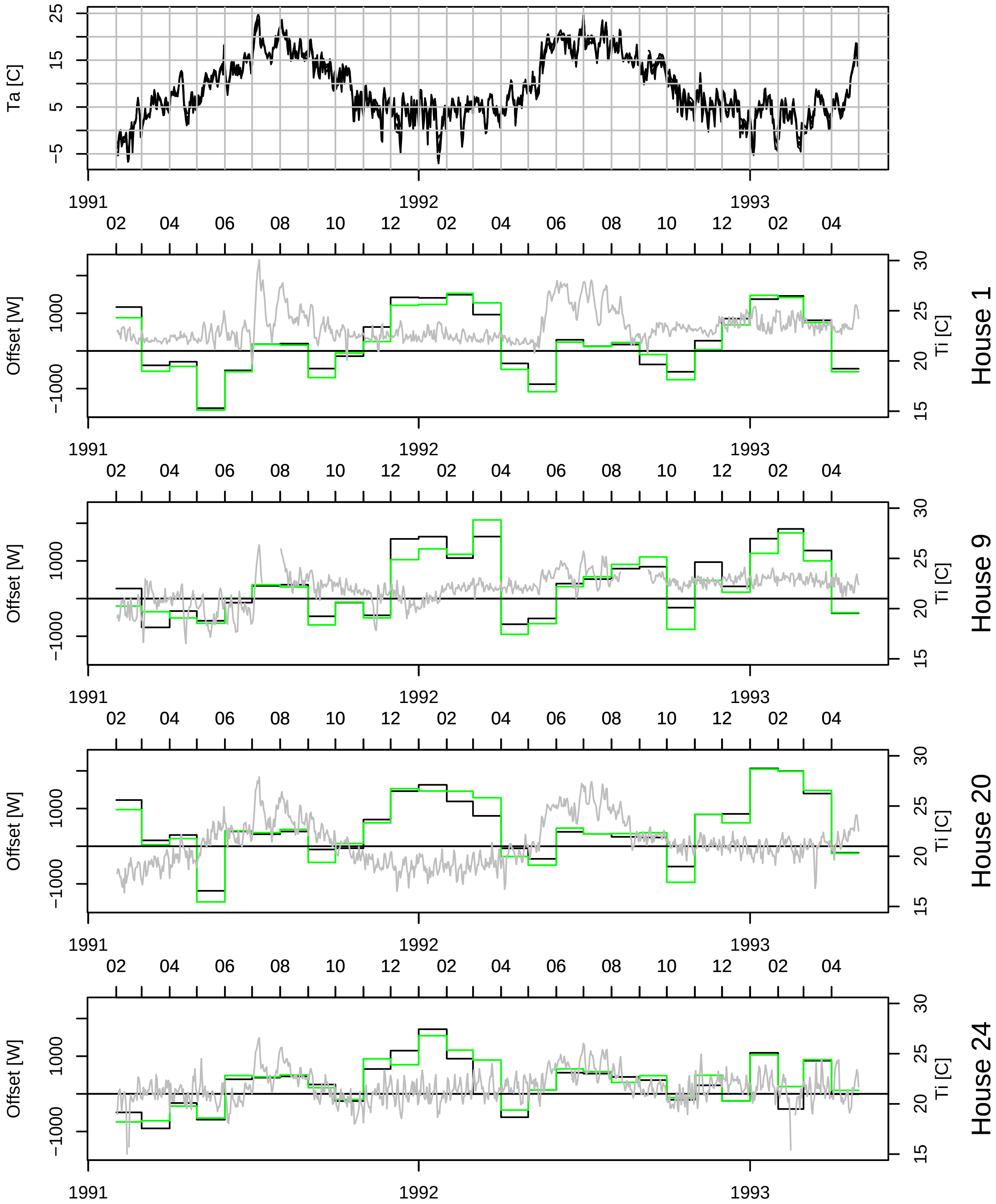


Figure 7: Monthly estimates of the constant term Q_0 based on daily averages and using 'total-water' as the dependent variable (black: linear regression, green: linear regression including previous day also).

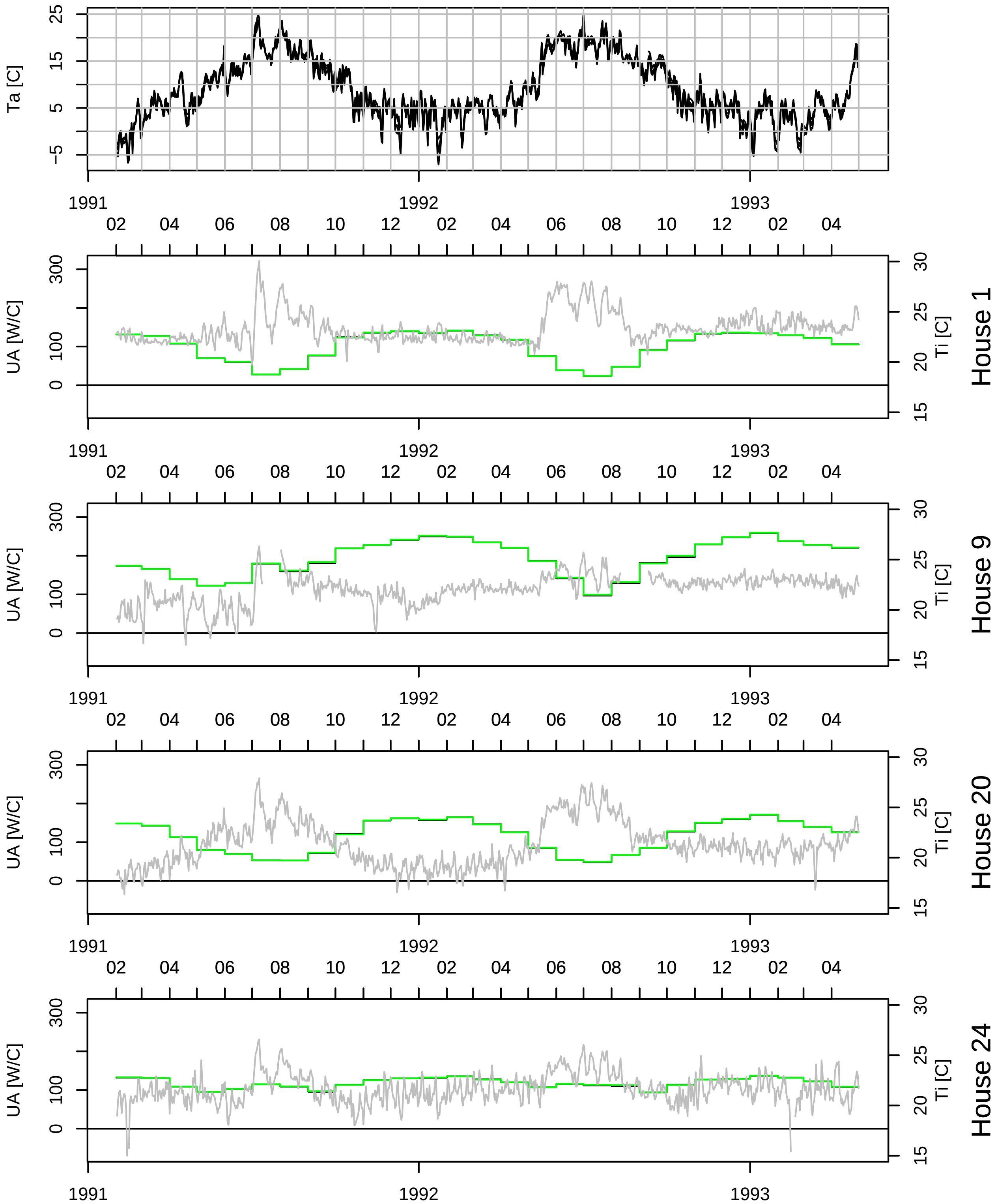


Figure 8: Monthly estimates of UA based on daily averages, but excluding the constant term of the models (black: linear regression, green: linear regression including previous day also).

3.3 Smooth time-varying estimates of UA-values

In Section 3.2 estimates were calculated separately for each month. The estimates obtained appear somewhat noisy, which in turn makes it difficult to interpret the general patterns. In order to obtain smooth time-varying estimates of UA, we therefore explicitly estimate the time-varying parameters in the model

$$Q_t = Q_0(t) + c_0(t)(T_{i,t} - T_{a,t}) + c_1(t)(T_{i,t-1} - T_{a,t-1}) + e_t \, , \tag{4}$$

where $Q_0(t)$, $c_0(t)$, and $c_1(t)$ are smooth functions of time. The estimate of the UA-value at time t is denoted $\widehat{UA}(t) = \hat{c}_0(t) + \hat{c}_1(t)$. Estimates in the model (4) can be obtained by a procedure related to locally weighted regression [2]. The extended procedure is indicated in [1], but described in more detail in [6].

The procedure is based on approximating the functions using low-order polynomials, followed by fitting the resulting linear model locally to a fixed value of the argument of the functions. Repeating the procedure for a large number of fixed values of the argument produces tables of the estimate function values. In this case local linear approximations are used together with a fixed bandwidth of 60 days. With this setup, and due to the down-weighting as depicted in Figure 9, it is implicitly assumed that a linear approximation fits well within approximately one month. With the aim of estimating smoothly varying functions over the heating season, this seems to be a reasonable choice. In general, the method has increased variance in boundary regions. Due to the size of the bandwidth, estimates for the first and last 60 days of the data period will be more unreliable than estimates for the remaining period.

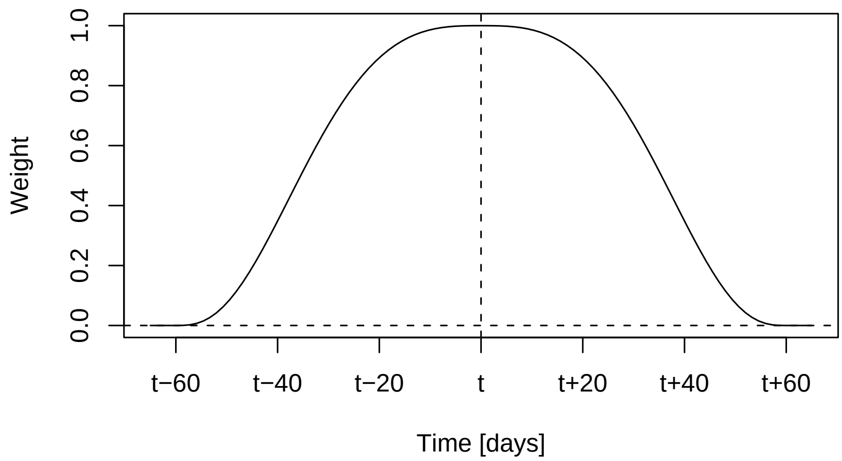


Figure 9: Weights used when estimating the function values at time t when a bandwidth of 60 days is used. The underlying weight function used is the so-called tricube [2].

Estimates are calculated for the following cases:

- `heat` used as the response and both indoor and ambient air temperature used in forming the regressors.
- `heat` used as the response but only ambient air temperature used in forming the regressors, i.e. removing T_i from model (4).
- `total-water` used as the response and both indoor and ambient air temperature used in forming the regressors.
- `total-water` used as the response but only ambient air temperature used in forming the regressors.
- `total` used as the response and both indoor and ambient air temperature used in forming the regressors.
- `total` used as the response but only ambient air temperature used in forming the regressors.

In the case where T_i is removed from the model, the Q_0 will increase drastically since the term will also contain a contribution originating from the indoor temperature.

Figure 10 shows the estimated UA-values as functions of time for the six different cases mentioned above. Except for boundary regions, the estimates are fairly similar for all six cases and the estimates obtained during winter months are consistent between heating seasons.

Figure 11 shows the constant term Q_0 for the three cases where the models are based on the difference between indoor and ambient air temperature. These plots show the amount of energy consumption which cannot be explained by the temperature difference. Negative values indicate energy entering from other sources than accounted for in the data. It is noted that for e.g. May there is a drop in all curves and when `heat` is used as the response variable the values obtained during May are negative in all cases. This could possibly be explained by significant heating of the houses by solar radiatio, while the ambient air temperature is still so low enough to allow this energy to be utilised. However, note that generally the response to temperature differences (i.e. the estimated UA-values) is larger in May than e.g. in the months just before May.

Considering model (4), if Q_0 is assumed to be zero and if $T_{i,t}$ is assumed to be approximately constant over time so that $T_{i,t} \approx T_{i,t-1}$, the model can be written as $Q_t = (c_0(t) + c_1(t))T_{i,t} - c_0(t)T_{a,t} - c_1(t)T_{a,t-1} + e_t$. If $(c_0(t) + c_1(t))T_{i,t}$ is estimated as a constant term varying smoothly with time, the ratio between the constant term and the UA-value $(c_0(t) + c_1(t))$ is an estimate of the indoor temperature.

Figure 12 shows the ratio between the constant term and the estimated UA-value for the three cases where the models are based on ambient air temperature alone. Assuming Q_0 in model (4) to be zero, the ratio should indicate the level of the indoor temperature. For houses 1 and 9, when using `heat` as the response the overall levels of the indoor temperature agree reasonably well with measured indoor temperatures. For houses 20 and 24 the agreement is poor. This seems to be caused by a low response to temperature differences during winter months, resulting in low estimates of UA.

Figure 13 shows the three sets of estimates based on temperature differences from Figure 10, together with three corresponding sets of estimates where the Q_0 is fixed at zero. Again it is seen that when fixing Q_0 at zero the method is sensitive as to which of the response variables is used (`heat`, `total-water`, or `total`). This sensitivity is very low when Q_0 is included in the model.

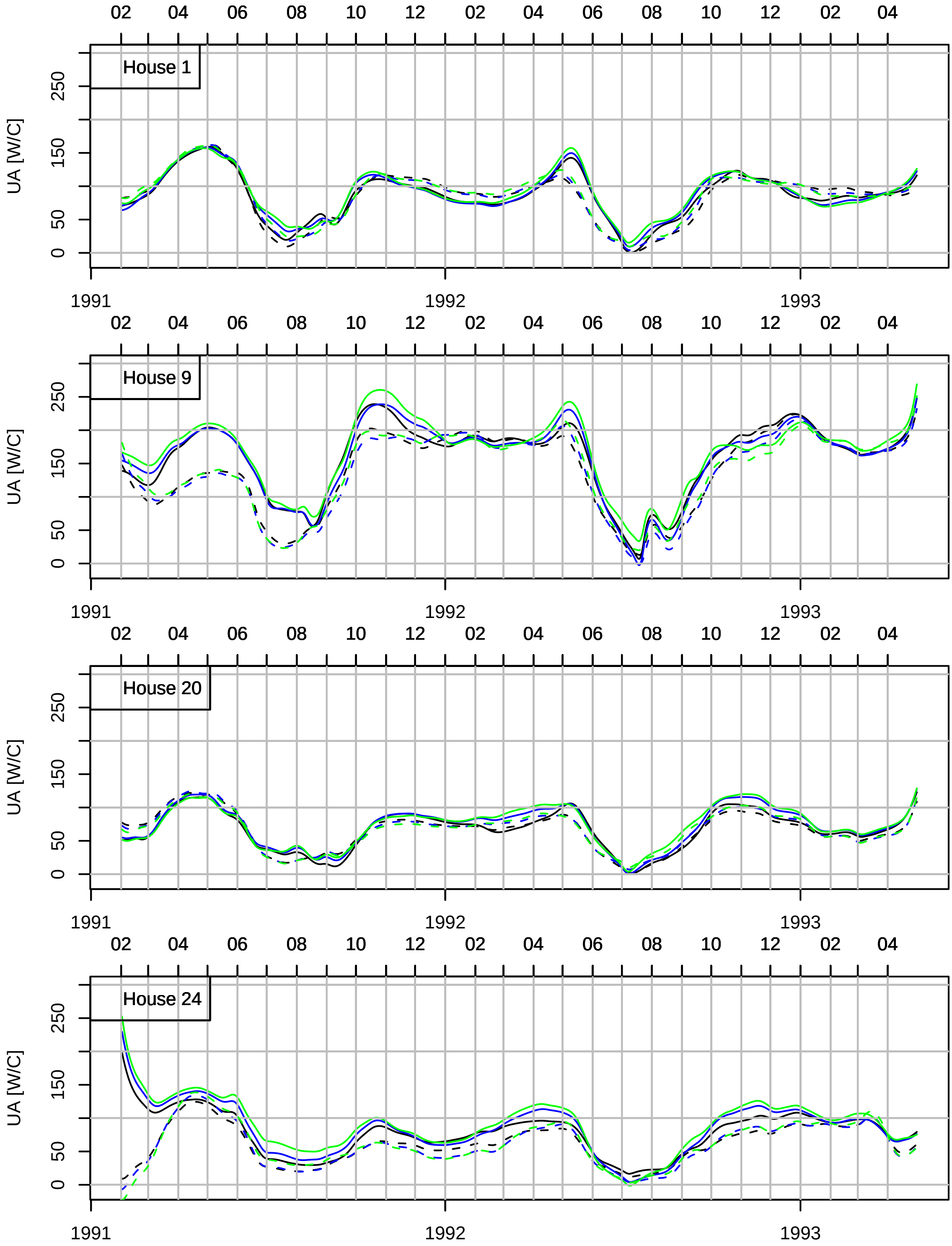


Figure 10: Estimates of UA obtained using local linear regression on daily averages including current and previous day in the model (full; based on $T_i - T_a$, dotted; based on T_a, black; 'heat' as the response, blue; 'total − water' as the response, green; 'total' as the response). A fixed bandwidth of 60 days is used.

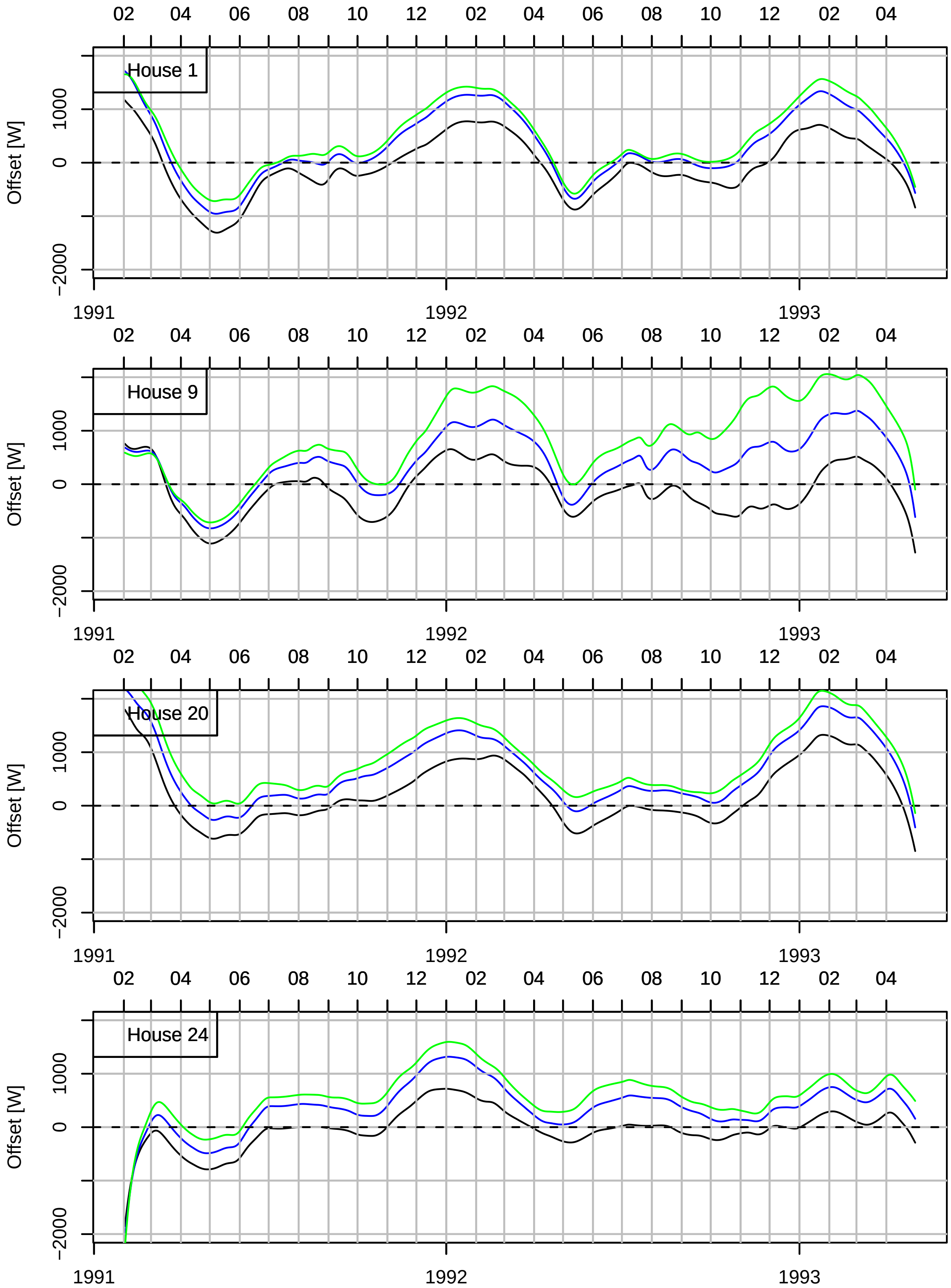


Figure 11: The constant term for the models based on $T_i - T_a$ for which the estimates of UA are shown in Figure 10.

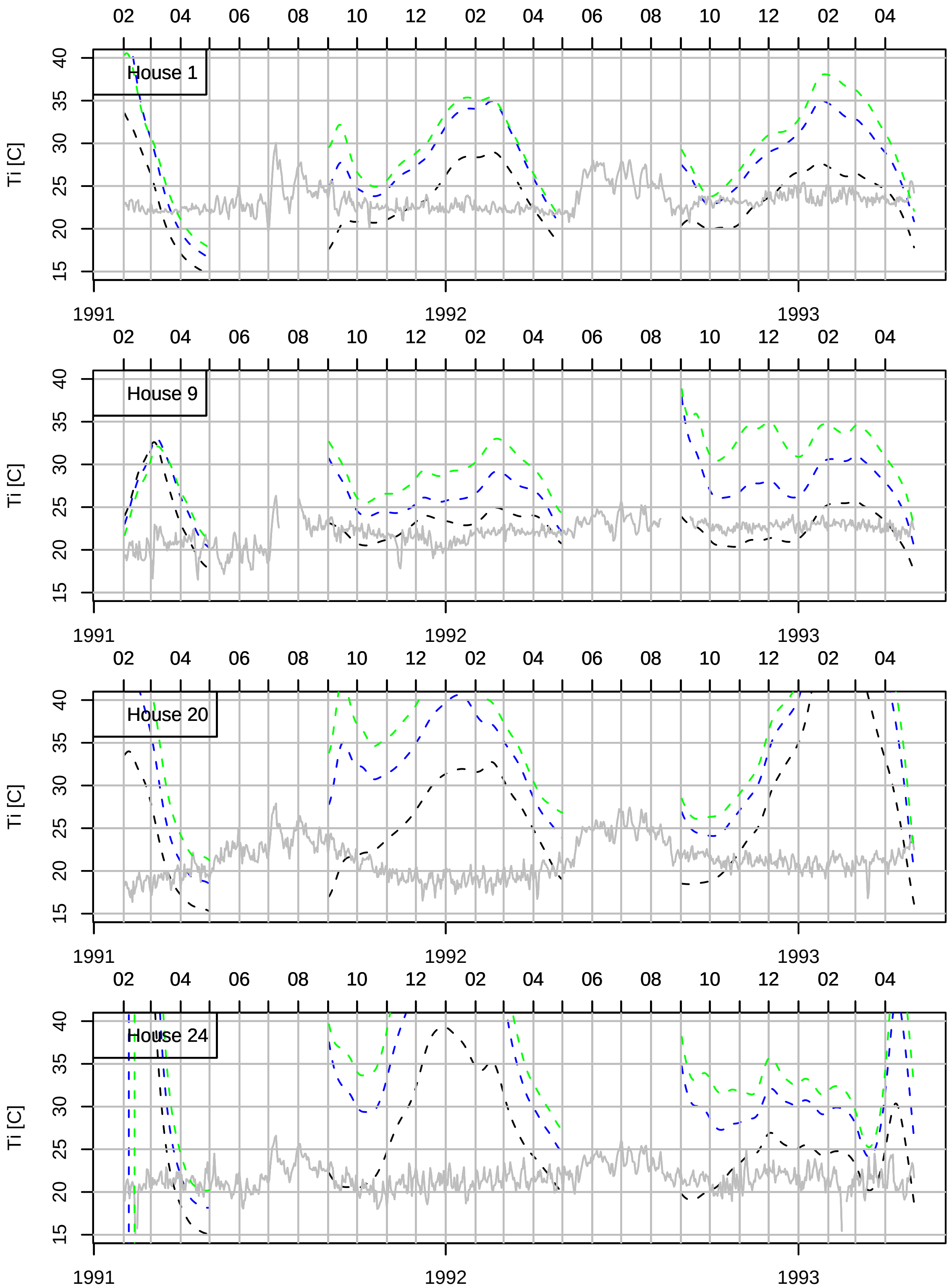


Figure 12: Estimates of the indoor temperature based on three different definitions of heat consumption (black; 'heat' as the response, blue; 'total − water' as the response, green; 'total' as the response). The estimates obtained for the months May to August are not shown. The measured indoor temperatures are shown in grey.

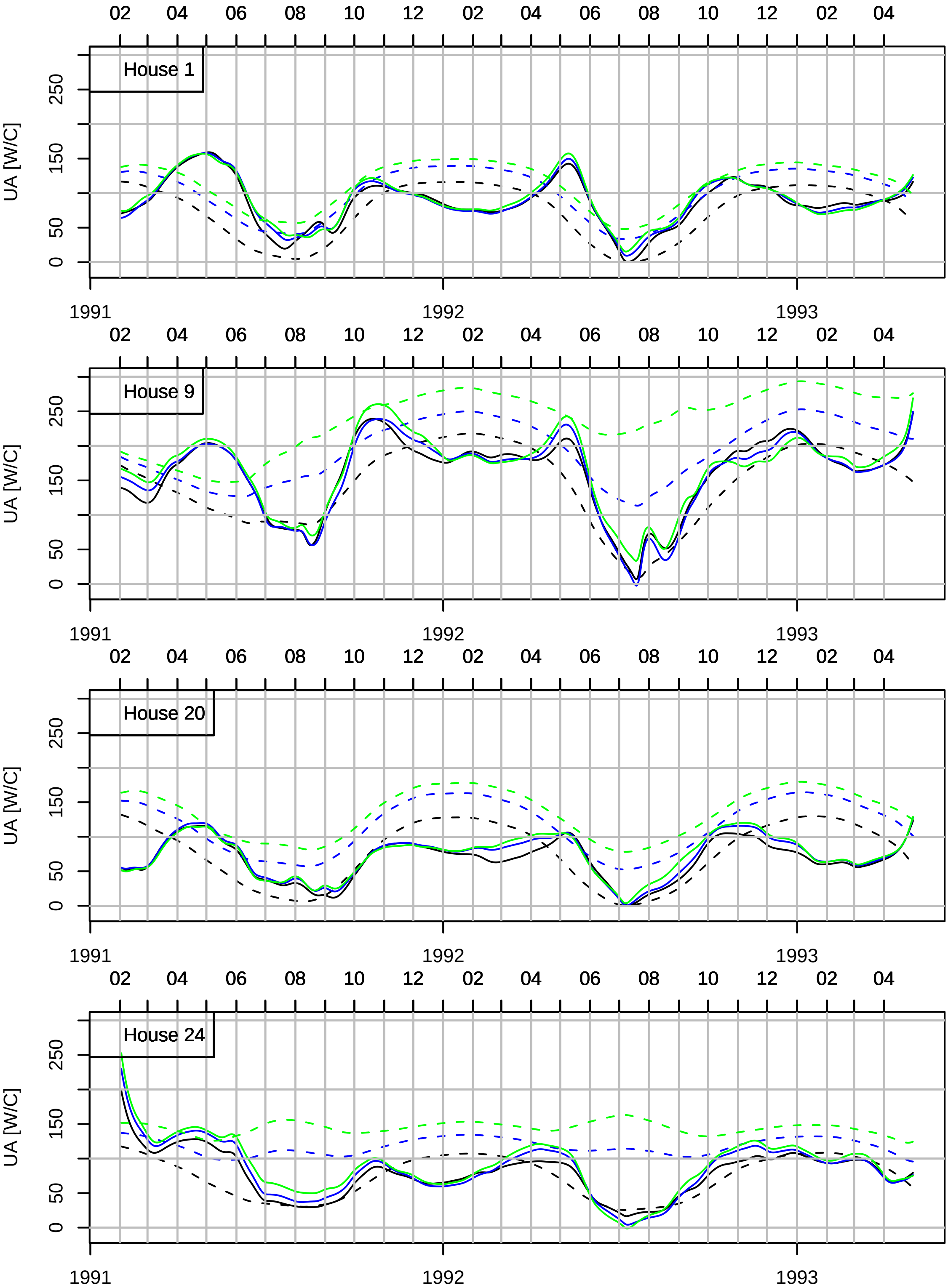


Figure 13: Comparing the estimates of UA based on $T_i - T_a$ (full) which are shown in Figure 10 with models where the intercept is excluded (dotted).

4 Analysis of the original samples using dynamic models

With the aim of investigating if the data can be modelled on the original time scale of 15 minutes, dynamical models are applied.

Let:

- t be a normalised time index corresponding to the 15 min averages, i.e. $t \in \mathcal{Z}$
- $Q_{H,t}$ be the electrical heat supplied to the heating equipment or similarly, if "free" heat is included in the analysis, i.e. if `total-water` is analysed.
- Q_t be heat loss from the building.
- $T_{i,t}$ be the indoor temperature.
- $T_{a,t}$ be the ambient (outdoor) air temperature.
- q^{-1} be the backward shift operator so that given the signal x_t, the operator yields $q^{-1}x_t = x_{t-1}$.

The most simple discrete-time dynamic model of the relation between the heat input and loss is obtained using a infinite impulse response filter with one pole ϕ_h and a stationary gain of one:

$$Q_t = \frac{1-\phi_h}{1-\phi_h q^{-1}} Q_{H,t} \tag{5}$$

Similarly, assuming that the main part of the thermal mass is placed on the outside of the walls yields the following relation:

$$Q_t = UA\left(T_{i,t} - \frac{1-\phi_w}{1-\phi_w q^{-1}} T_{a,t}\right) \tag{6}$$

Which is further supported by the fact that the indoor temperature is fairly constant. Combining (5) and (6) and isolating the indoor temperature yields:

$$T_{i,t} = (UA)^{-1} \frac{1-\phi_h}{1-\phi_h q^{-1}} Q_{H,t} + \frac{1-\phi_w}{1-\phi_w q^{-1}} T_{a,t} \tag{7}$$

Using a logistic-type relation to constrain the parameters of the model (UA, ϕ_h, ϕ_w), the estimates can be found using the non-linear least squares method.

Initially, ϕ_h is set at 0.8. The step response of the filter is shown in Figure 14. It is seen that almost full response is achieved within approximately 2 hours, which from a physical point of view seems reasonable.

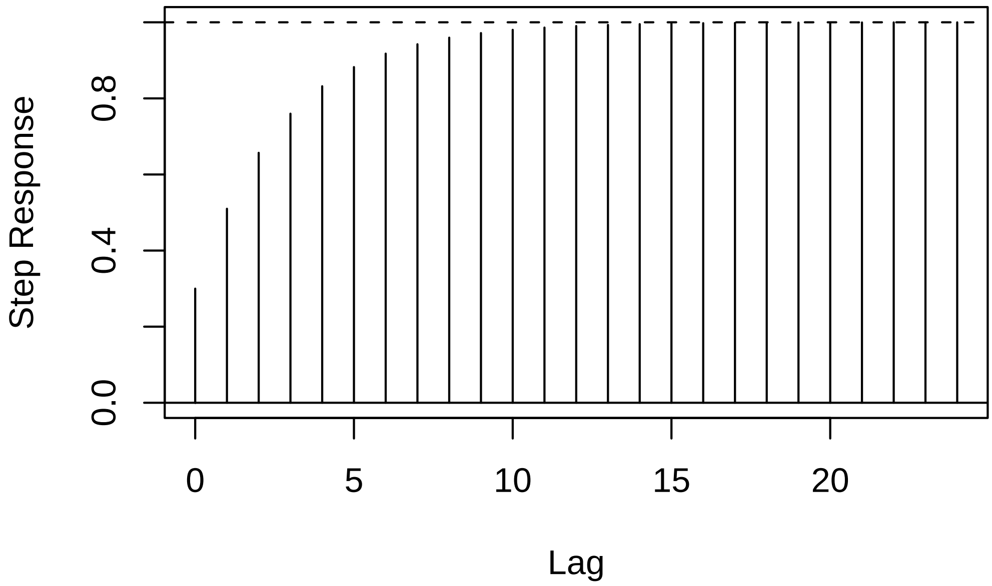


Figure 14: Step response of the filter $0.2/(1-0.8q^{-1})$. The time unit is lags, i.e. intervals of 15-mins length and the total length of the 1st axis is 6 hours.

Data from house no. 9, covering the period from Sep. 01, 1991 to Jun. 01, 1992 is used in the analysis presented here. The main reason for selecting this house is that based on the analysis of daily averages, this house shows the best agreement between measured and estimated indoor temperature, cf. Figure 12, when using `heat` to represent the heat consumption. Furthermore, there are no missing measurements of energy consumption for the period considered (a few missing temperature measurements are replaced by appropriate values).

The UA-value is estimated at 228 $W/^oC$ and ϕ_w is estimated at 0.952, corresponding to the step response function shown in Figure 15. Figure 16 shows a summary of the fit.

The fitted values do not catch the observations; too many high frequencies from the heating system seem to be entering the model. In an attempt to circumvent this a 4th order low-pass Butterworth filter is used (coner frequency 0.125 samples^{-1}, i.e. $1/0.125 = 8$ quarters or 2 hours), `http://www-users.cs.york.ac.uk/~fisher/mkfilter/trad.html`. The spectrum of the filter is shown in Figure 17 and the spectrum of the filtered and unfiltered series is shown in Figure 18.

The results obtained when using the filtered series are shown in Figure 19. Again, it is seen that, despite the low-pass filter applied, too many high frequencies from the heating system enter into the response (T_i). This may be caused by the particular heating equipment used. Electrical heating equipment is often controlled with an on/off switch which receives signals from a temperature sensor[3]. Such a controller may cause large differences in the heat used

[3] Together with some hysteresis which keeps the on/off frequency within reasonable limits.

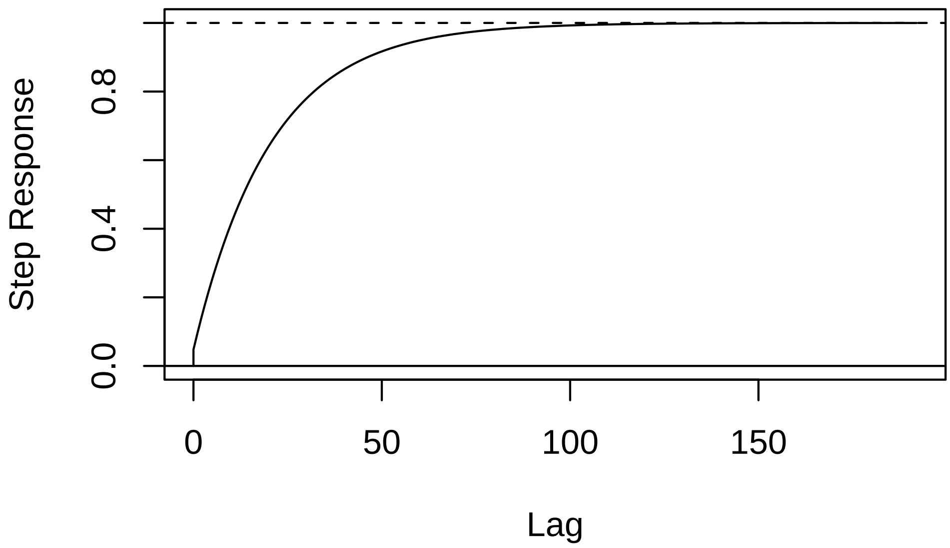


Figure 15: Step response over two daya of the filter $(1 - \hat{\phi_w})/(1 - \hat{\phi_w} q^{-1})$. The time unit is lags, i.e. intervals of 15 mins length.

during neighbouring time intervals. Note also that during the beginning and end of the period, where the heating equipment is not used very much, the high frequencies are significantly less predominant.

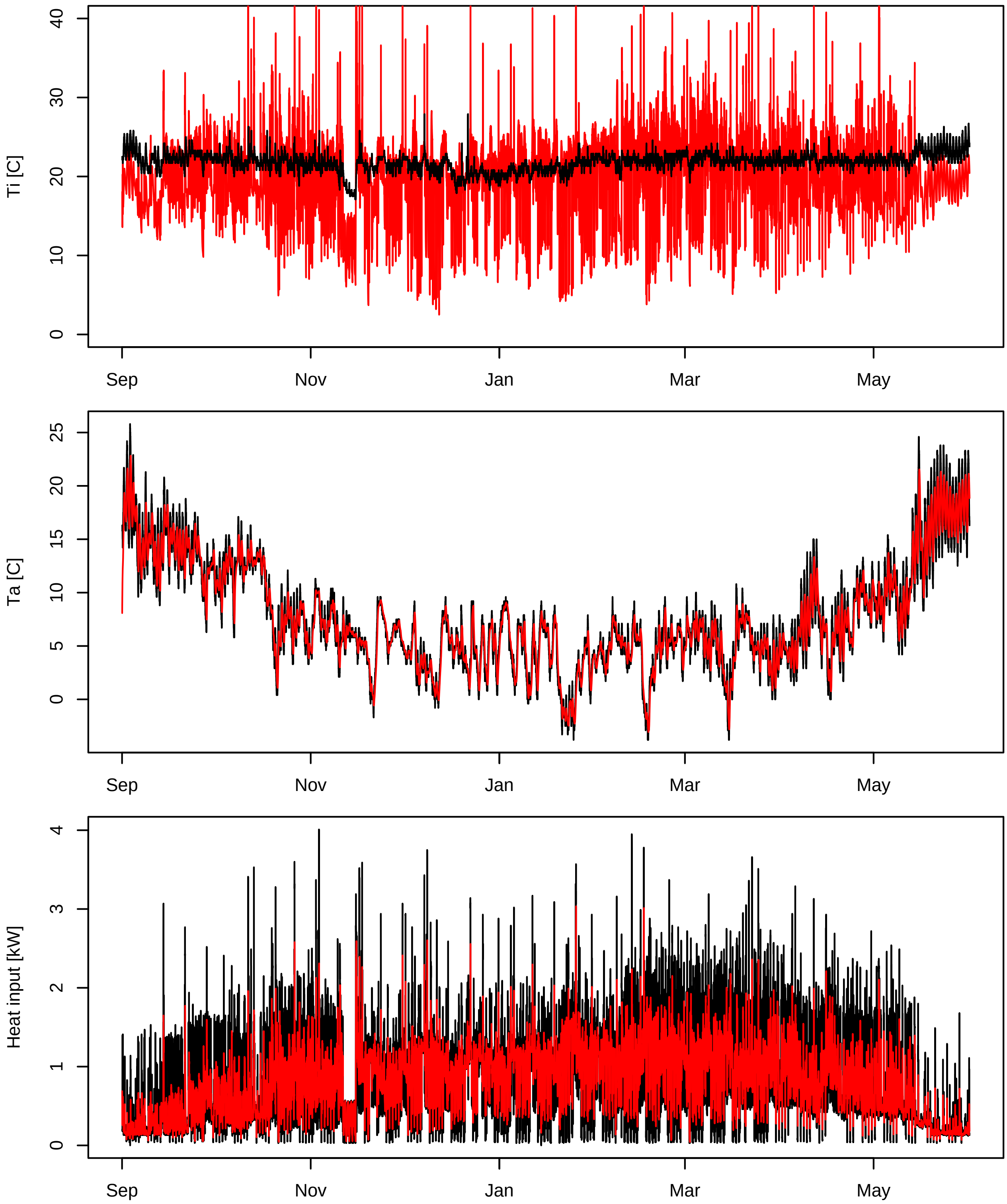


Figure 16: Summary of the fit of the dynamic model (7). Top: Observations of T_i (black) and fitted values (red). Middle: Observations of T_a (black) and low-pass filtered values using $\hat{\phi}_w$ (red). Bottom: Observations of heat input $Q_{H,t}$ ('total' − 'water') and low-pass filtered values using $\hat{\phi}_h = 0.8$ (red).

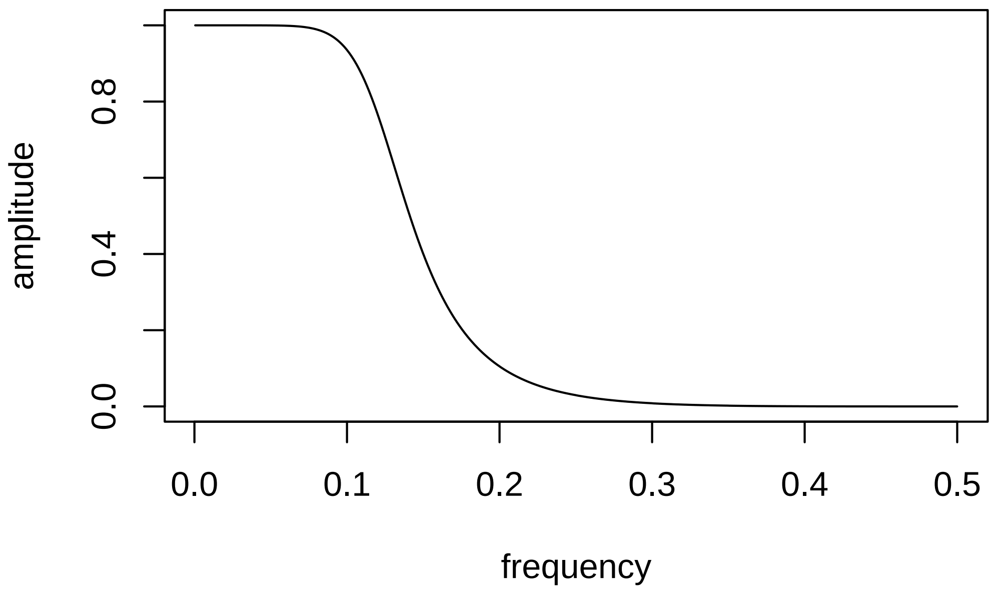


Figure 17: Spectrum of the low-pass filter applied to $Q_{H,t}$.

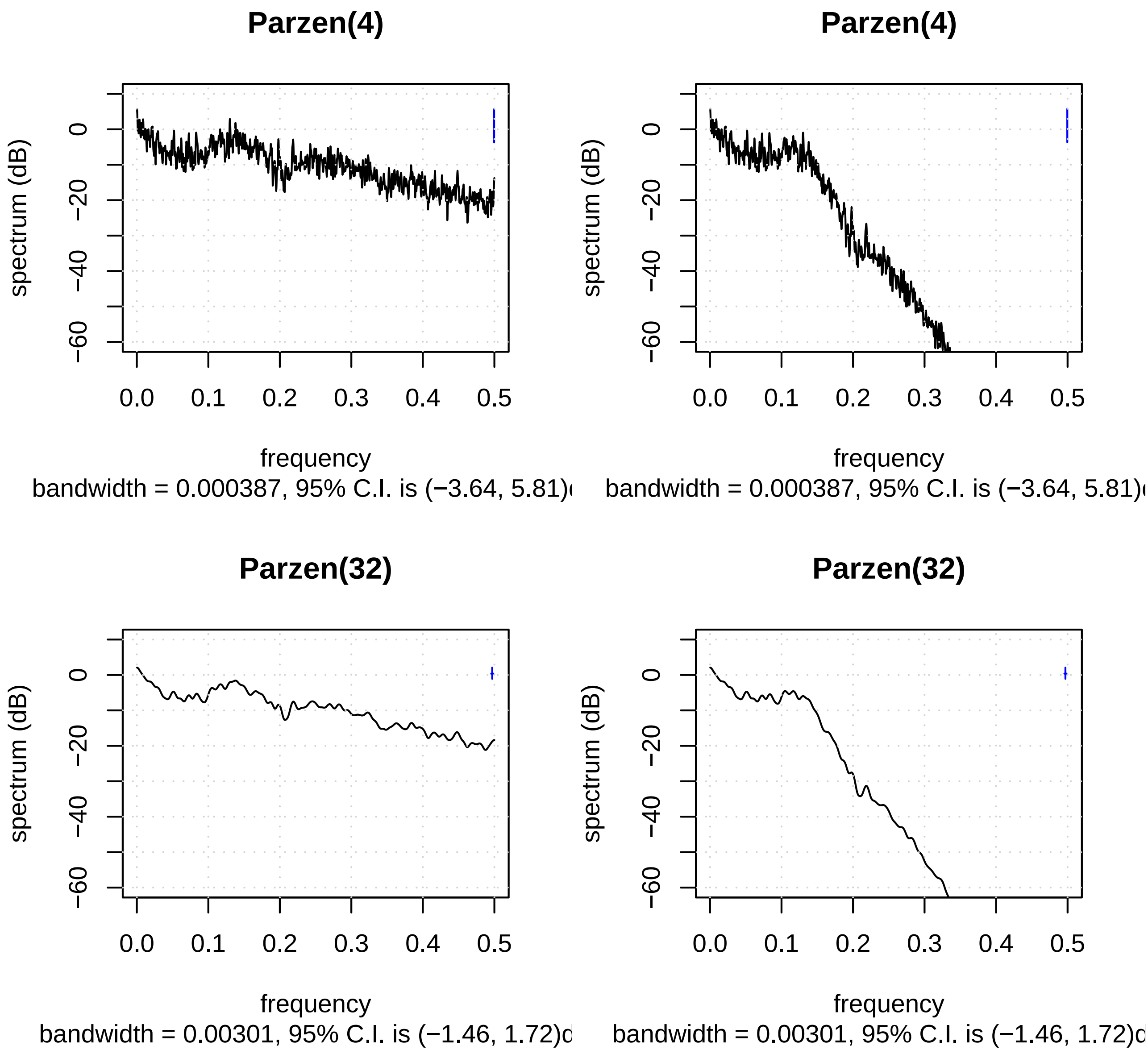


Figure 18: Spectrum of unfiltered $Q_{H,t}$ (left column) and filtered $Q_{H,t}$ (right column) after removing the smooth trend with a local linear smoother estimated using a bandwidth of 30 days. For spectral smoothening a Parzen window is used, with the window sizes as indicated. The spectra is only estimated for par of the series, representing a clear winter situation.

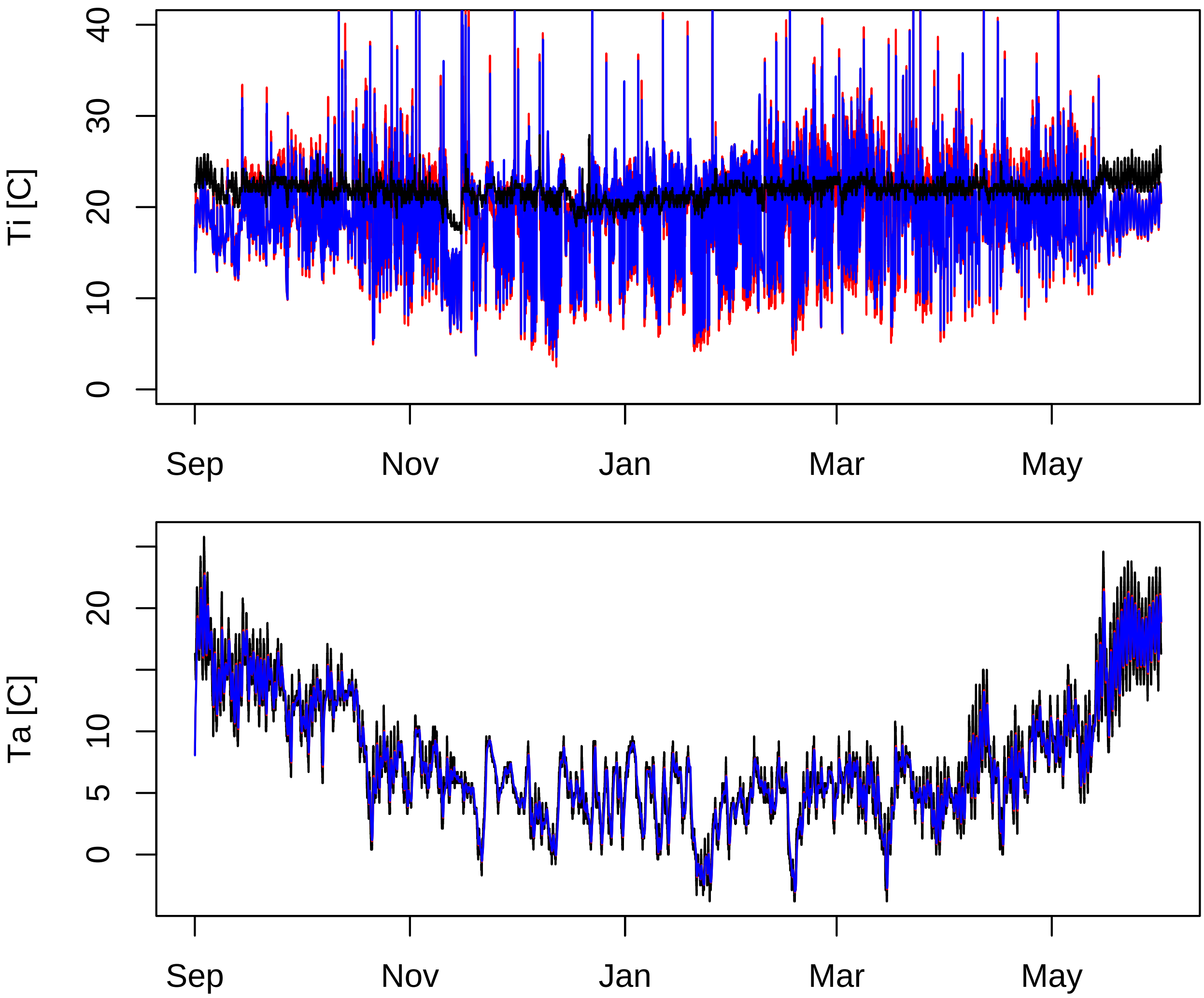


Figure 19: Summary of the fit of the dynamic model (7). Top: Observations of T_i (black), fitted values (red), and fitted values when using low-pass filtered $Q_{H,t}$ (blue). Bottom: Observations of T_a (black), low-pass filtered values using $\hat{\phi}_w$ (red), and low-pass filtered values using $\hat{\phi}_w$ obtained when using low-pass filtered $Q_{H,t}$ (blue).

5 UA-values estimated for all 25 houses

In this section the UA-values estimated for each house and for each of the heating seasons 91/92 and 92/93, are considered using model (3) on page 15[4]. Furthermore, the UA-values have been estimated in two ways (i) using both the temperature difference between indoor and ambient air temperature, and (ii) using the ambient air temperature alone. In all cases daily averages are used. Generally, the estimated UA-value divided by the ground area of the house is considered. Besides these variables, information is used in the analysis about the number of persons living in the house, whether the house is equipped with a wood burning stove, and whether night-time drop is used. The analysis also considers the heating season and the type of model used for estimating the UA-value, i.e. case (i) and (ii) above.

5.1 Graphical analysis

An overview of the normalised estimates is given in Figures 20 and 21.

Figures 22 and 23 show the estimated UA-values and offsets (Q_0) for each of the 25 houses. Generally, for the UA-values, it is seen that there is good agreement between methods, i.e. whether the indoor temperature is used or not. This is further investigated in Figure 24 and it is seen that in general it is rather unimportant whether or not the indoor temperature is used in the model. However, 2 out of 50 cases show a difference of approximately 0.5 $W/^oC/m^2$ in the value of UA divided by ground area of the house. This is seen for the group with wood burning stove and night-time drop.

Both with respect to UA and Q_0 some houses show good agreement between heating seasons, whereas for other houses the agreement between seasons is poor.

Figure 25 investigates this aspect further. Generally, for houses not equipped with a wood burning stove, there is good agreement between seasons. However, for houses equipped with a wood burning stove, there can be some difference between seasons.

[4]Only months November to March (both included) are used. For 91/92 the average ambient temperature was 4.2^oC and 3.7^oC for 92/93.

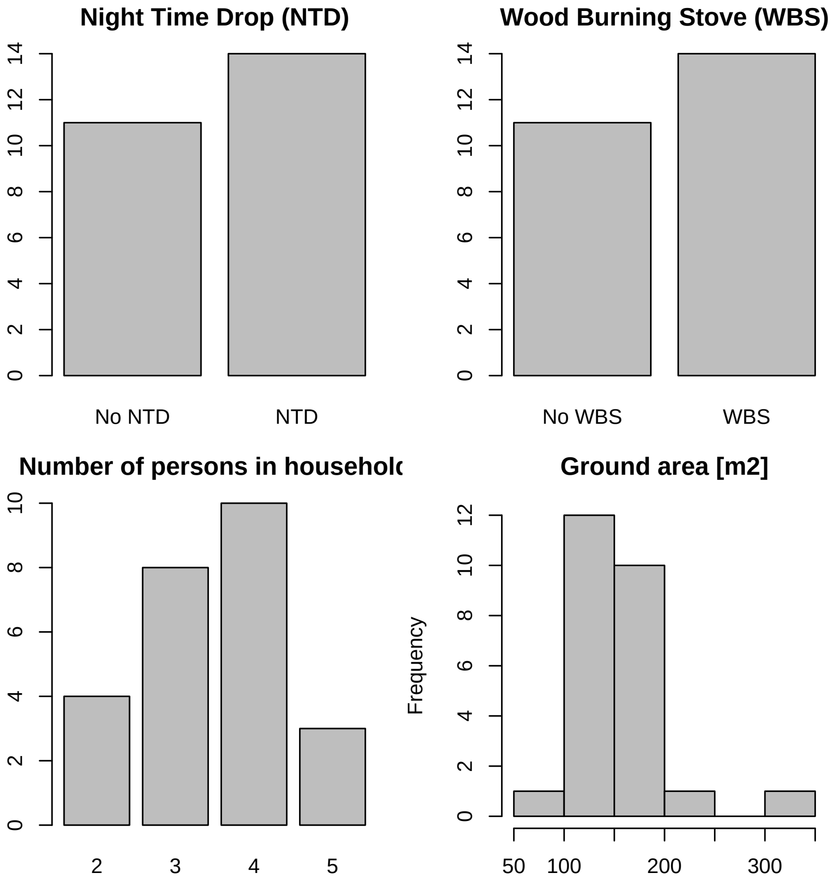


Figure 20: Distribution of variables for the 25 houses/households.

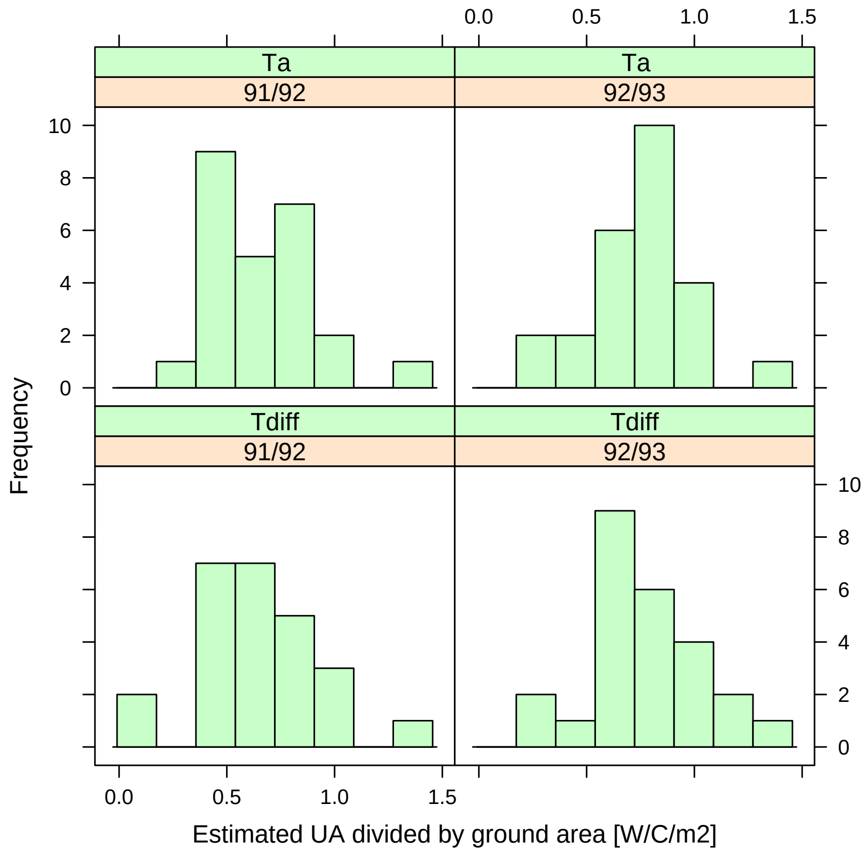


Figure 21: Distribution of estimated UA-values normalised by the ground area of the house for the two heating seasons and for the model where only the ambient air temperature is used (`Ta`) and the model where the temperature difference is used (`Tdiff`).

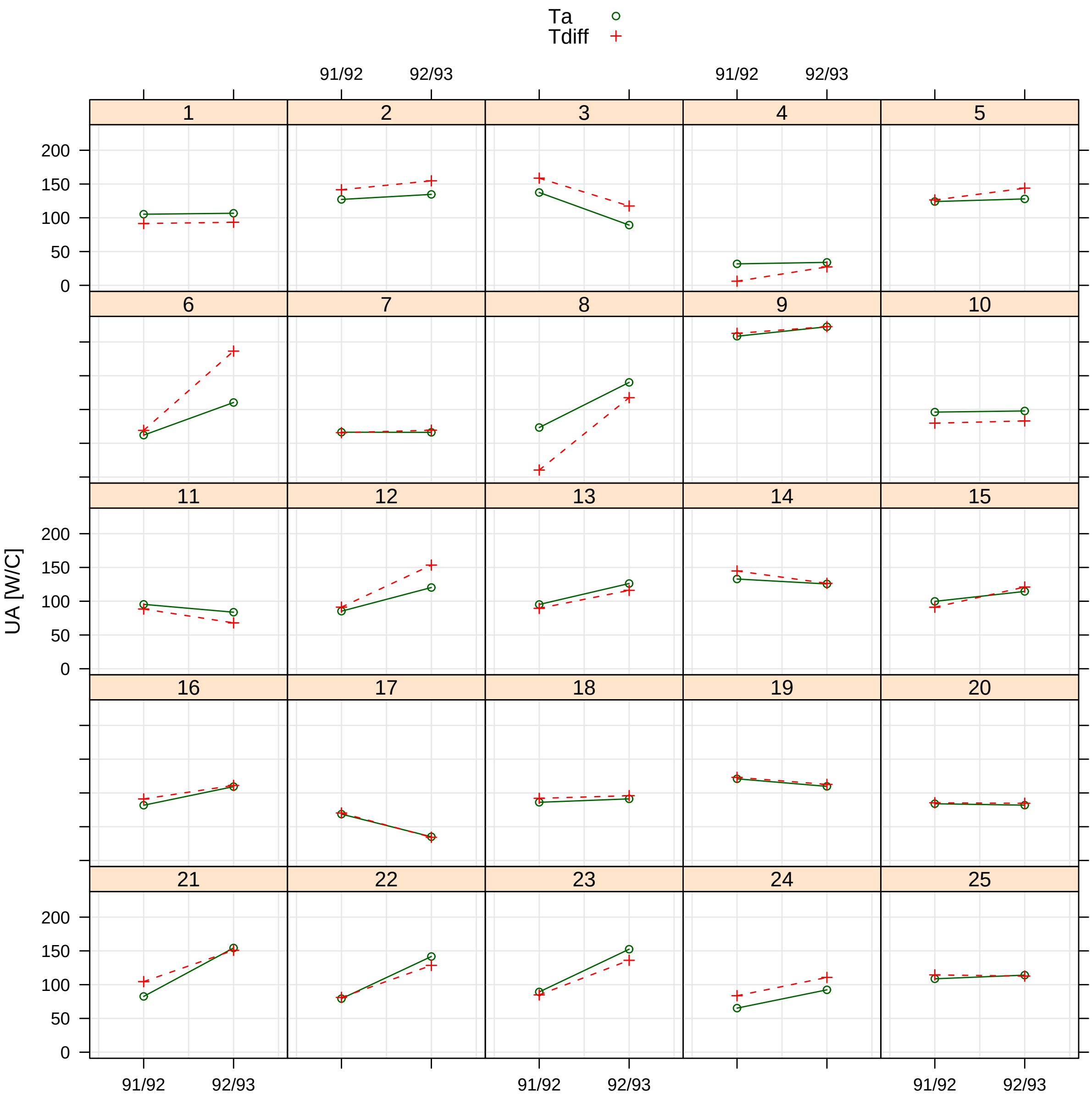


Figure 22: Estimates of UA for each of the 25 houses and for each heating season, Both estimates obtained using the temperature difference (`Tdiff`) as the regressor and just the ambient air temperature (`Ta`) as the regressor are shown.

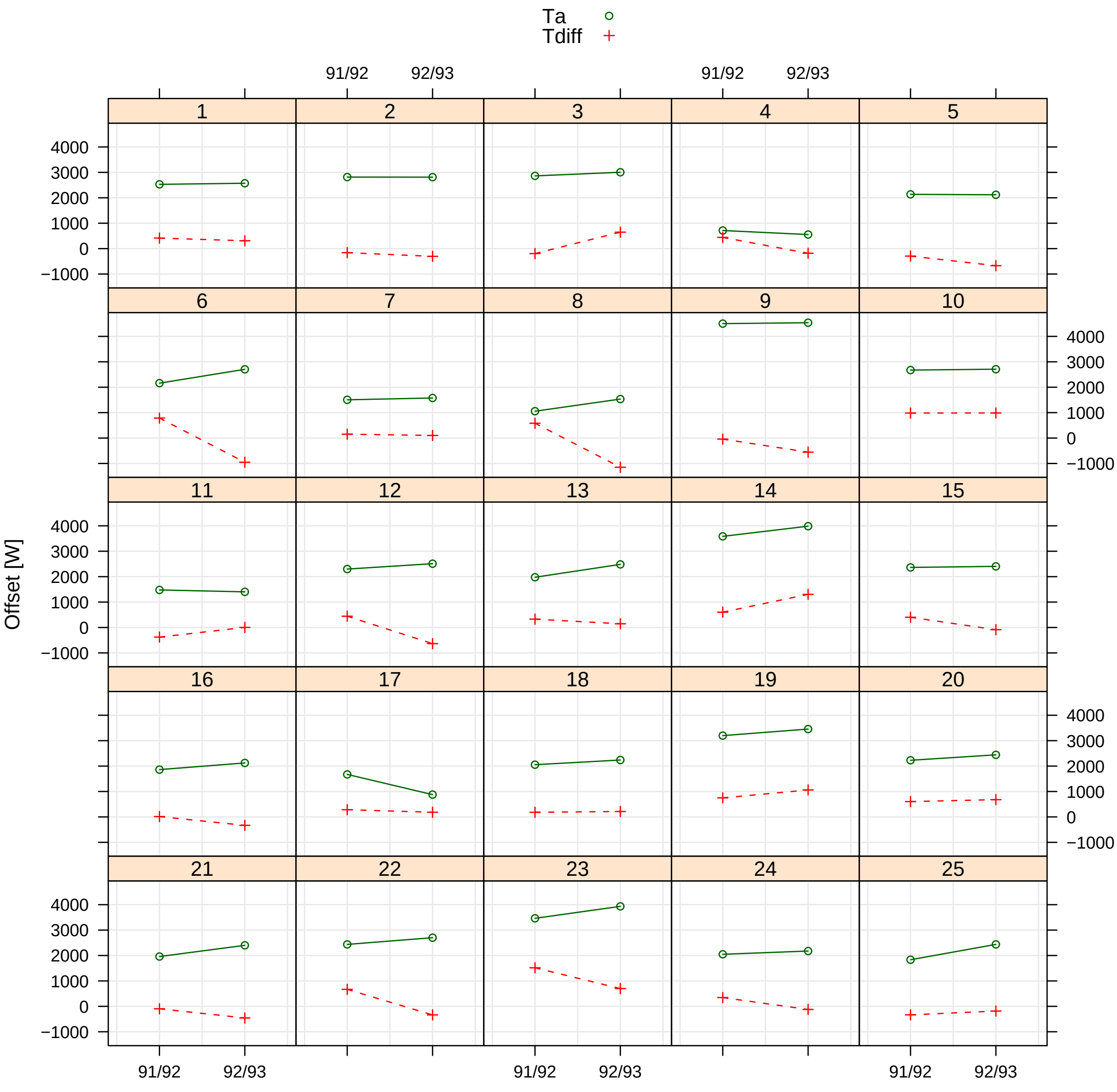


Figure 23: Estimates of the offset (Q_0) for each of the 25 houses and for each heating season, Both estimates obtained using the temperature difference (`Tdiff`) as the regressor and just the ambient air temperature (`Ta`) as the regressor are shown.

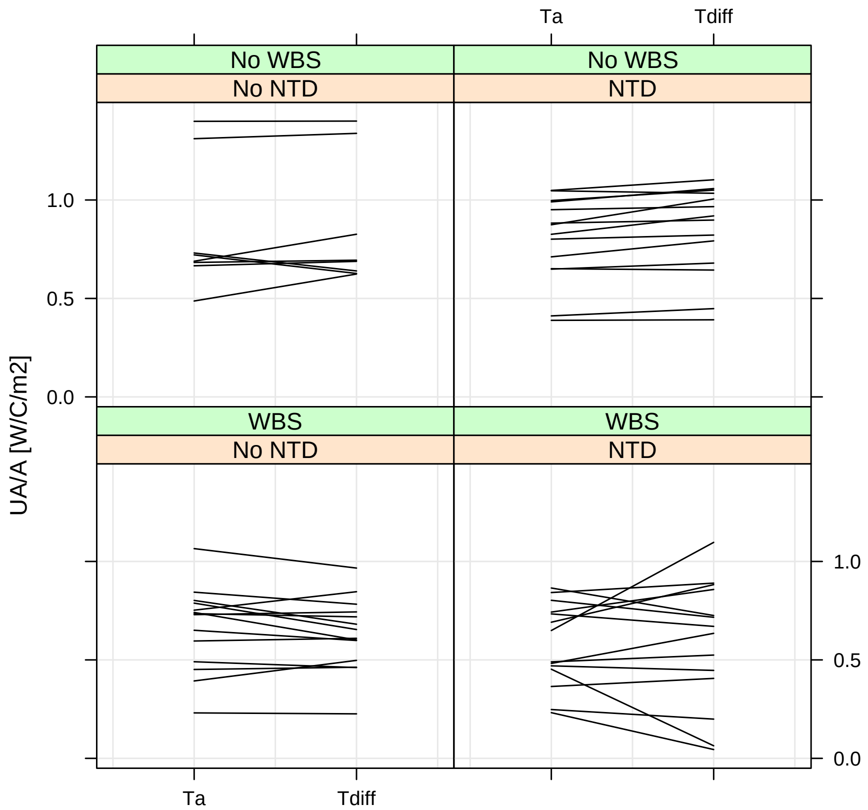


Figure 24: Estimates of UA depending on the type of regressor used (`Ta` or `Tdiff`) for the groups of houses defined by their use of night time drop (NTP) and whether the houses are equipped with a wood burning stove (WBS).

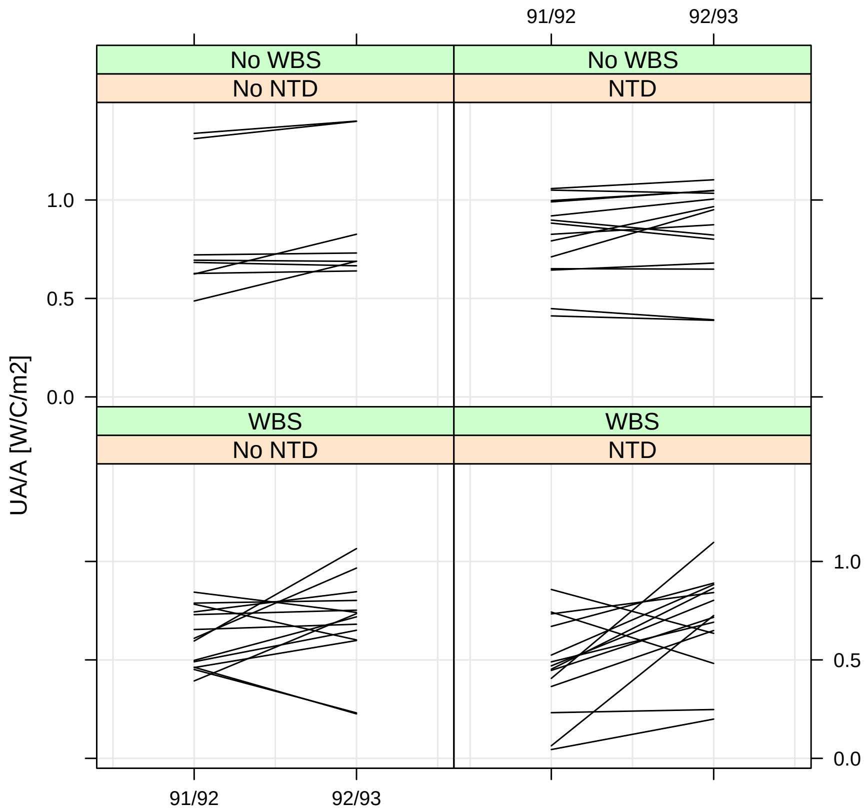


Figure 25: Estimates of UA depending on the heating season for the groups of houses defined by their use of night time drop (NTP) and whether the houses are equipped with a wood burning stove (WBS)

5.2 Analysis of variance

The analysis is performed in parallel for each of the cases defined by whether the indoor temperature is used in the regression or not. The main reason for this choice is that the pairs of estimates are based on the same data. For this reason inference (p-values etc.) may be incorrect if the type of model is included in the analysis. In this section, standard statictical terms such as p-value, interaction, and confidence interval are used. A description of these terms is beyond the scope of this report, but can be found in most books about introductory statistics, see e.g. [3].

The response variable analysed is the estimated UA-value normalised by the ground area of the house. Initially, the data are analysed using analysis of variance with effects defined by

- the number of persons living in the house (variable `persons`),
- whether night time drop is used (variable `ntd`),
- whether the house is equipped with a wood burning stove (variable `bstove`), and
- the heating season, i.e. 91/92 or 92/93 (variable `season`).

Main effects and second order interactions are included in the model. However, for the factor defined by the number of persons living in the house, only the main effect is included in the model.

Figure 26 on page 44 shows diagnostic plots of the two models. Generally the standard diagnostics, i.e. the first two rows of subplots, are acceptable indicating no obvious violations of the assumptions of homogenerity of variance and normality of errors. However, the last row of subplots indicate a rather clear correlation of the two residuals for each house. This observation indicates that inference based on the models used here could be misleading.

The observed correlation of the two residuals for each house is positive indicating that, if the residual is e.g. positive in one heating season, it tends to be positive in the other heating season as well. Consequently, the data show that there are differences in normalised UA-values between houses, which can not be accounted for by the factors in the model. This is hardly surprising, but it needs to be handled appropriately when analysing the data.

The phenomena can be modelled using random effects models. Here the error term is split in to one contribution for the house which is common for each of the two heating seasons and a residual contribution which differs between heating seasons. Thus, with this model the individual characteristics of the house are modelled (as a random effect) and it is then possible to compare effects of `persons`, `ntd`, `bstove`, and `season`.

The analysis is performed using the linear mixed effects models as implemented in the package `nlme`, version 3.1 for R (see `www.r-project.org`). In the analysis of variance tables shown here the result of marginal tests are presented, i.e. tests where the full model is compared with the model where the term under consideration is deleted. Generally, tests for main effects are not considered when the main effect under consideration appears in any interaction term.

Table 1 shows the analysis of variance tables. It is seen that all 2nd order interactions are non-significant. However, there seems to be a weak indication of an interaction between "wood burning stove" and "season". Table 2 shows the analysis of variance tables when only the main effects are included in the models. From these tables it is seen that the use of night time drop is non-significant. Also, the number of persons in the house is non-significant (taking the 5% level as the limit). However, when not using the indoor temperature when estimating UA, the number of persons in the house is near-significant. The clear significant effects are "wood burning stove" and "season".

	DF1	DF2	F-value (Tdiff)	p-value (Tdiff)	F-value (Ta)	p-value (Ta)
persons	3	18	1.6748	0.2080	2.3159	0.1101
ntd	1	18	0.4272	0.5217	0.1317	0.7209
bstove	1	18	3.6941	0.0706	2.1383	0.1609
season	1	22	0.1802	0.6753	0.2652	0.6117
ntd:bstove	1	18	0.0229	0.8815	0.0547	0.8177
ntd:season	1	22	1.9474	0.1768	0.0809	0.7788
bstove:season	1	22	3.1513	0.0897	1.0529	0.3160

Table 1: Analysis of variance for the case where the indoor temperature is used when estimating UA (`Tdiff`) and the case where only the ambient air temperature is used when estimating UA (`Ta`). DF1 and DF2 indicate the numerator and denominator degrees of freedom, respectively. Interactions are indicated by ":".

	DF1	DF2	F-value (Tdiff)	p-value (Tdiff)	F-value (Ta)	p-value (Ta)
persons	3	19	1.7723	0.1866	2.4523	0.0948
ntd	1	19	0.3054	0.5869	0.4846	0.4948
bstove	1	19	6.1369	0.0228	4.8136	0.0409
season	1	24	6.9335	0.0146	6.5815	0.0170

Table 2: Analysis of variance when excluding interactions from the models for the case where the indoor temperature is used when estimating UA (`Tdiff`) and the case where only the ambient air temperature is used when estimating UA (`Ta`). DF1 and DF2 indicate the numerator and denominator degrees of freedom, respectively.

Significant effects of "wood burning stove" and "season" have been identified. These observations are somewhat in contrast to the graphical analysis where the season effect mainly seemed to originate from the houses equipped with a wood burning stove (Figure 25). However, as part of the analysis it was noted that the interaction between "wood burning stove" and "season" is not clearly insignificant. For this reason Tables 3–4 and 5–6 analyse the data separately for houses without and with a wood burning stove, respectively.

From Tables 3 and 4 it is seen that when only houses without a wood burning stove are used in the analysis, no effects are significant. In principle, this could be because only part of the data is used and thus the observation alone is inconclusive. However, as is seen from Tables 5 and 6, the season is significant, at least when the indoor temperature is used when calculating UA. These observations are in agreement with the graphical analysis

	DF1	DF2	F-value (Tdiff)	p-value (Tdiff)	F-value (Ta)	p-value (Ta)
persons	3	6	1.2722	0.3655	0.7872	0.5434
ntd	1	6	0.0959	0.7673	0.0820	0.7843
season	1	9	2.3174	0.1623	2.0059	0.1904
ntd:season	1	9	0.5278	0.4860	0.2210	0.6495

Table 3: Analysis of variance when only considering houses **without** a wood burning stove.

	DF1	DF2	F-value (Tdiff)	p-value (Tdiff)	F-value (Ta)	p-value (Ta)
persons	3	6	1.2722	0.3655	0.7872	0.5434
ntd	1	6	0.1841	0.6828	0.1363	0.7247
season	1	10	2.5652	0.1403	3.2338	0.1023

Table 4: Analysis of variance when only considering houses **without** a wood burning stove and excluding interactions.

	DF1	DF2	F-value (Tdiff)	p-value (Tdiff)	F-value (Ta)	p-value (Ta)
persons	3	9	1.0446	0.4189	3.4024	0.0669
ntd	1	9	0.9310	0.3598	0.3631	0.5617
season	1	12	0.3640	0.5575	1.2012	0.2946
ntd:season	1	12	2.9033	0.1141	0.2171	0.6496

Table 5: Analysis of variance when only considering houses **with** a wood burning stove.

	DF1	DF2	F-value (Tdiff)	p-value (Tdiff)	F-value (Ta)	p-value (Ta)
persons	3	9	1.0446	0.4189	3.4024	0.0669
ntd	1	9	0.0135	0.9100	0.1614	0.6973
season	1	13	5.7037	0.0328	4.3245	0.0579

Table 6: Analysis of variance when only considering houses **with** a wood burning stove and excluding interactions.

Table 7 shows approximate 95% confidence intervals corresponding to the effects `season` and `ntd` for the case where the analysis is only based on houses without a wood burning stove, i.e. corresponding to Table 4 above. The intervals are obtained as the estimated effects ± two times the calculated standard error of the estimate. It is seen that compared to the range of normalised UA-estimates (Figure 21 on page 35) the confidence intervals are small[5]. Two times the residual variance is estimated to 0.123 $W/C/m^2$ when the indoor temperature is used and 0.136 $W/C/m^2$ when the only the ambient air temperature is used.

	Lower	Upper
Season difference: '92/93' - '92/91' (Ta)	-0.01	0.11
Season difference: '92/93' - '92/91' (Tdiff)	-0.01	0.09
NTD difference: 'NTD' - 'No NTD' (Ta)	-0.44	0.30
NTD difference: 'NTD' - 'No NTD' (Tdiff)	-0.43	0.28

Table 7: Approximate 95% confidence intervals [$W/C/m^2$] for effects `season` and `ntd`.

[5] A range of 0.1 $W/C/m^2$ corresponds to an uncertainty of 150 W for a 10^oC temperature difference for a house with a 150 m^2 ground area.

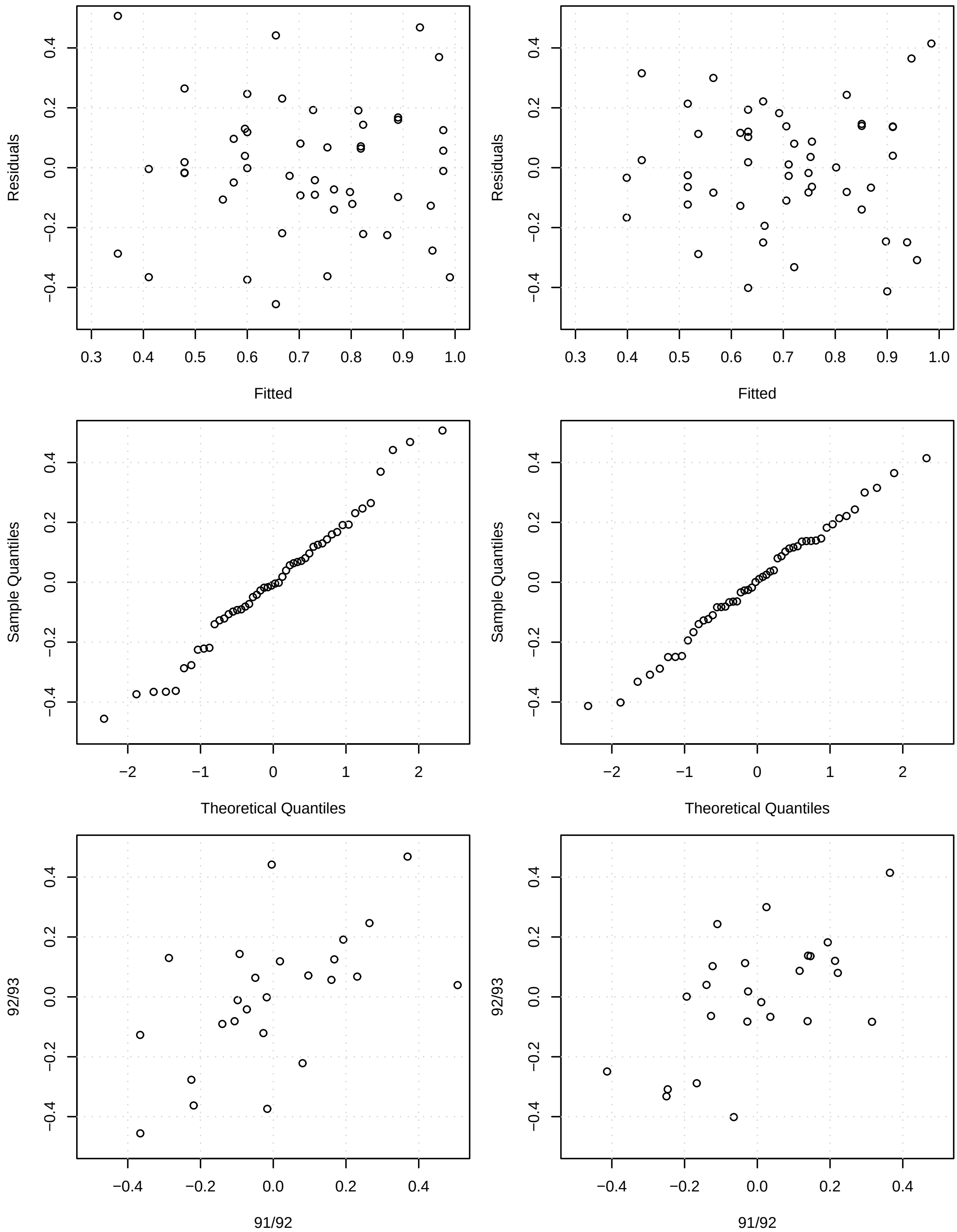


Figure 26: Model diagnostics when the indoor temperature is used when estimating UA (left column) and when only the ambient air temperature is used when estimating UA (right column). The top row show the residuals versus the fitted values, normal Q-Q plots of the residuals are shown in the middle row, and the bottom row show the two residuals for each house plotted against each other.

6 Conclusion

Measurements of heat consumption, indoor temperature, and outdoor temperature are analysed for four single-family houses not equipped with a wood burning stove and not using night-time drop.

The primary aim of the analysis is to devise methods of estimating the UA-value from data. The houses are heated via electrical heating equipment and measurements of (i) the total electricity consumption, (ii) the electricity consumption used by the heating equipment, and (iii) the electricity consumption used for hot tap water are available. Generally, this results in three different definitions of heat consumption which can be used for estimating the UA-values; a) item (i) above, b) item (ii) above, or c) the difference between items (i) and (iii) above. Definition c) aims at including the "free" heat in the measurement of energy supplied to the house. The data are available as 15-minute values.

Primarily the data are analysed as daily averages. It is concluded that:

- If an offset term is not included in the model, the estimated UA-value is highly sensitive to the definition of heat consumption.
- If an offset term is included in the model, the estimated UA-value is not sensitive to the definition of heat consumption.
- Also, at least during winter months, when an offset term is included in the model, the result is not very sensitive to whether the indoor temperature is known.
- The offset term is generally positive and largest during winter months. Naturally, the offset is smallest when using the measured electricity consumption of the heating devices (i.e. item (ii) above) in the models. It is also noted that in this case the time-varying estimates of the offset vary around zero. This indicates that if the offset is excluded from the model, the true underlying UA-values of the houses are most appropriately represented by the results obtained using item (ii) above as representing the actual heat supplied to the house. A consequence of this conclusion is that, generally, the "free" heat is not utilized very efficiently. However, this is also the estimate in best agreement with the estimates obtained when including the offset in the model and possibly only using the ambient air temperature in the model, cf. Figures 10, 11, and 13.

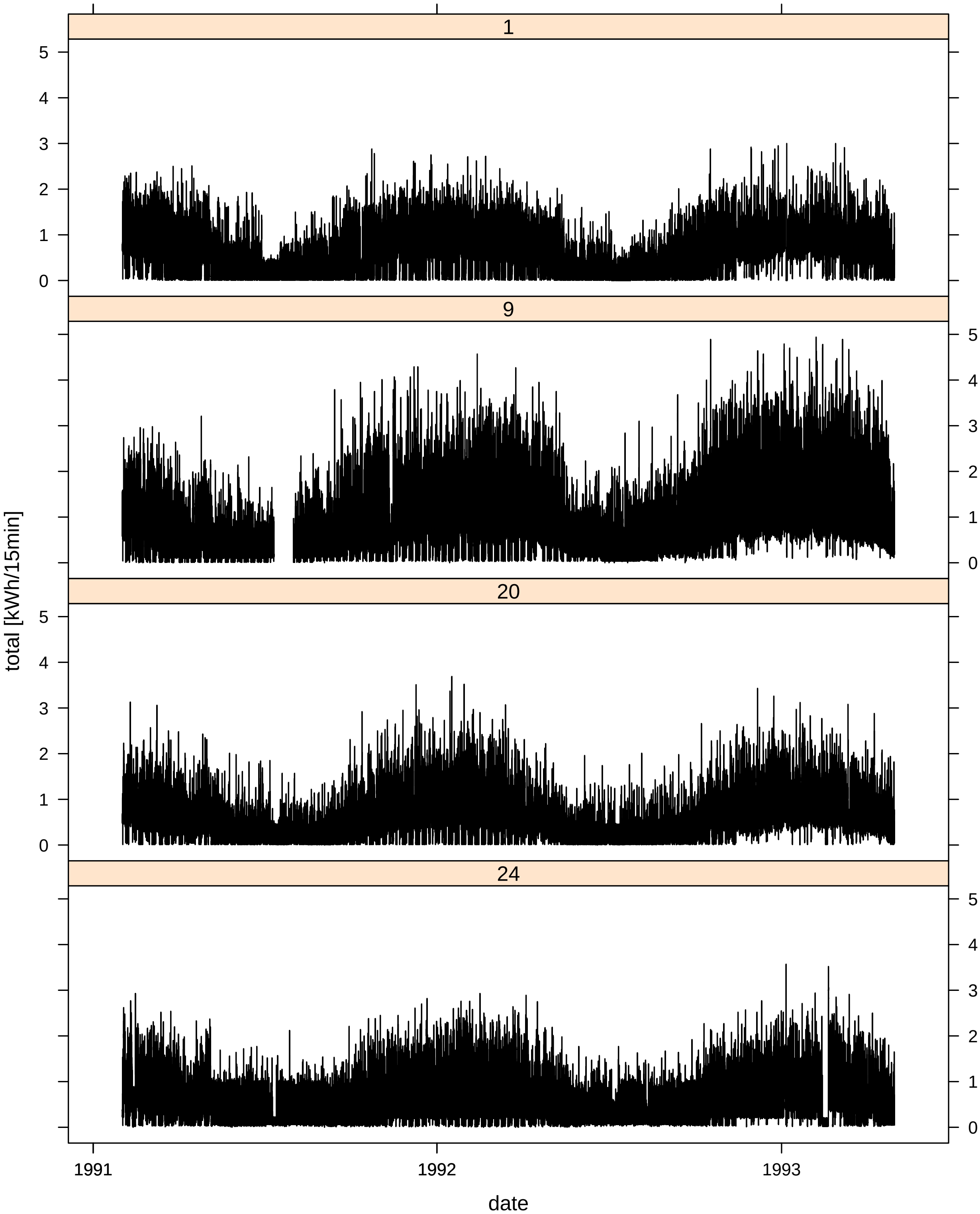
In conclusion it seems reasonable to base an estimate of the UA-value of an house on measurements of ambient air temperature and measurements of the energy consumption of the house. It is not crucial whether the latter contains a base-load not related to climate. The estimates can be obtained using linear regression with a constant term, the average temperature for the current day, and the average temperature for the previous day. The estimation should be based on a winter period, e.g. November to March.

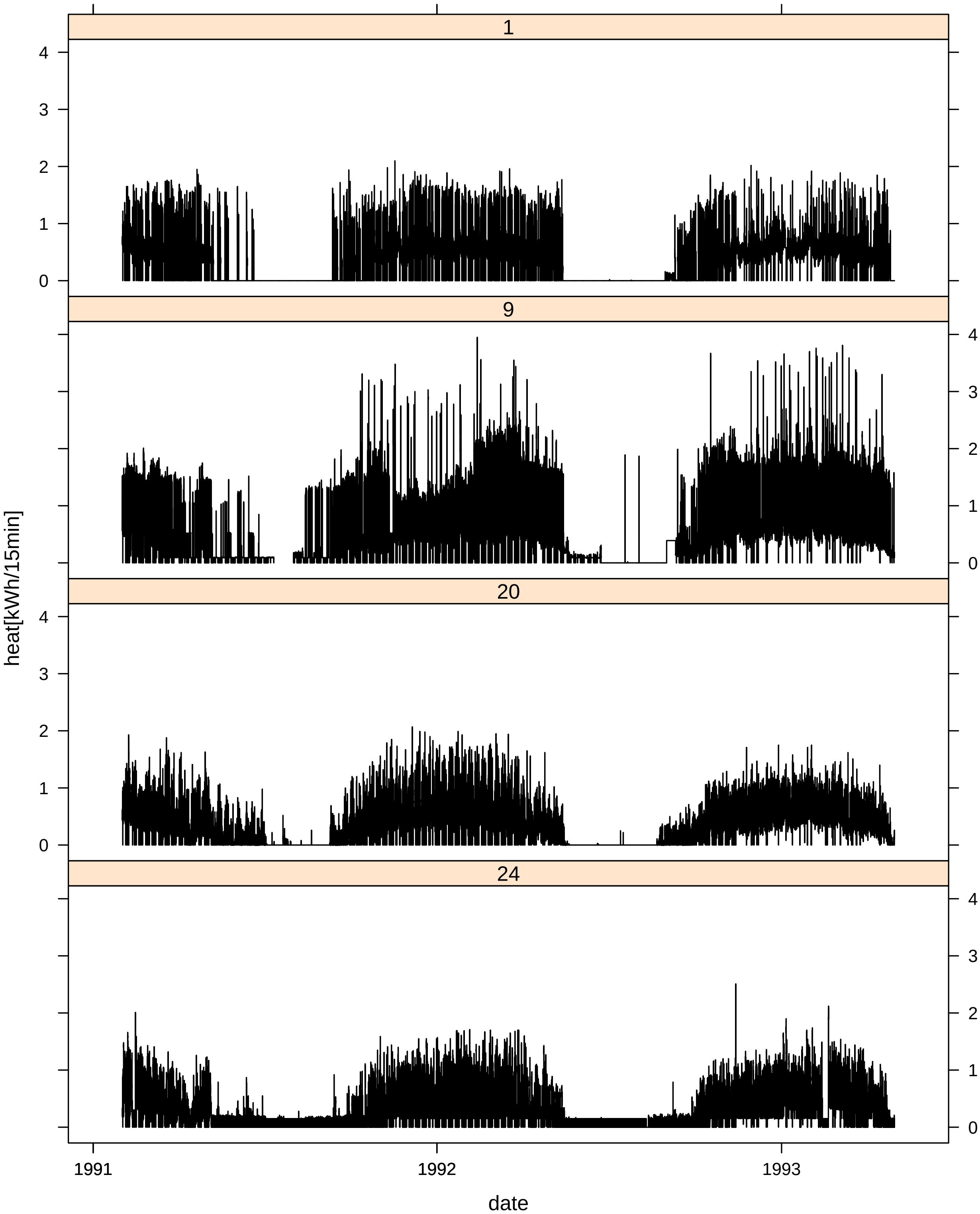
An attempt to model the data on the original time scale was also made. Although the estimated UA-values are comparable in size to the estimates obtained using daily averages the indoor temperature predicted by the model contain too many high frequencies originating from the measurements of heat consumption. This is also true despite low-pass filtering of the heat consumption.

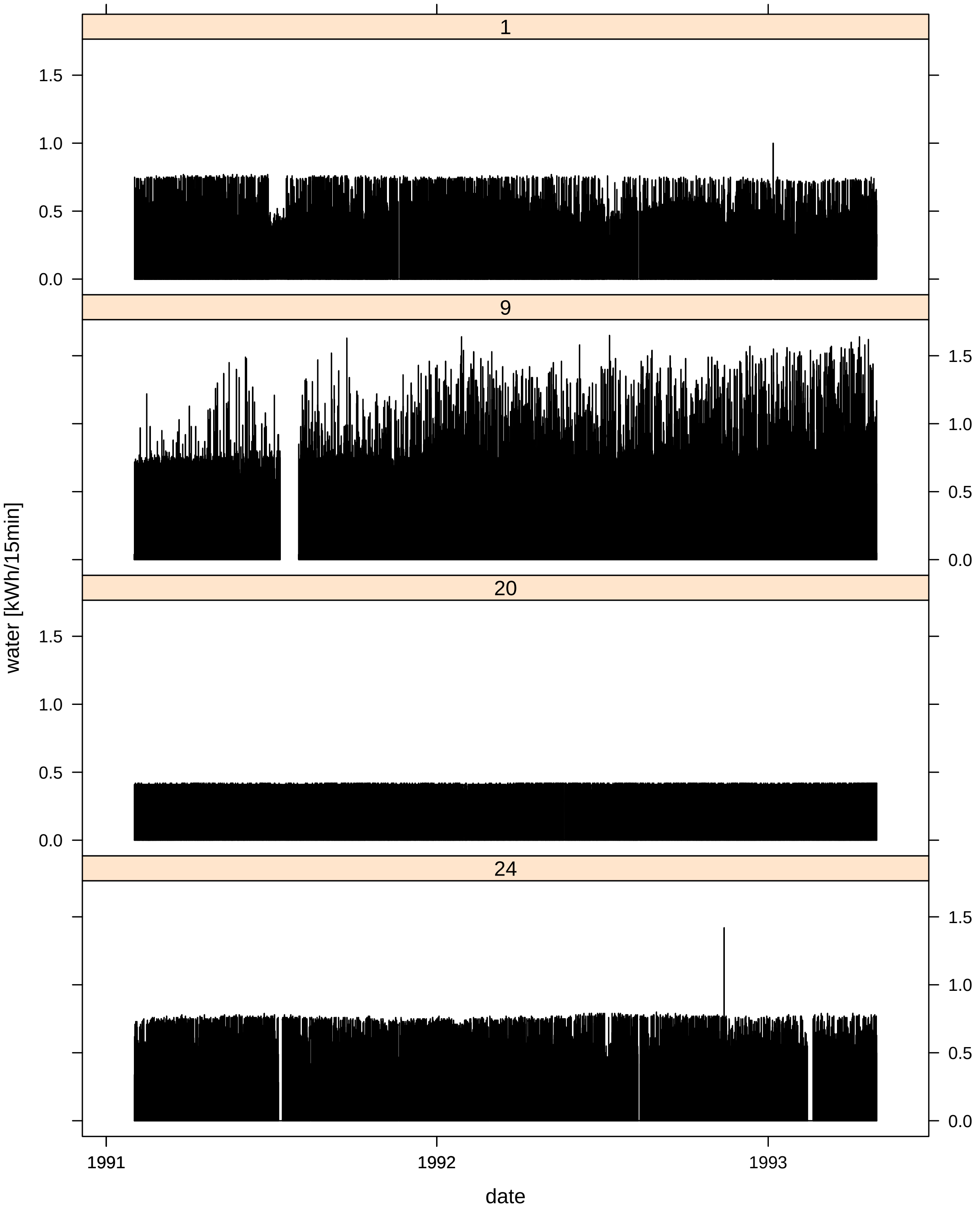
Following the analysis of the daily averages for the four houses that do not use night-time drop and without a wood burning stove, UA-values were estimated for each of the heating seasons 91/92 and 92/93 (months November to March) for all 25 houses for which data are available. It is concluded that the method performs well when a wood burning stove is not used since in this case litte or no effect of whether the indoor temperature is used or not, season, and the use of night time drop can be observed in data.

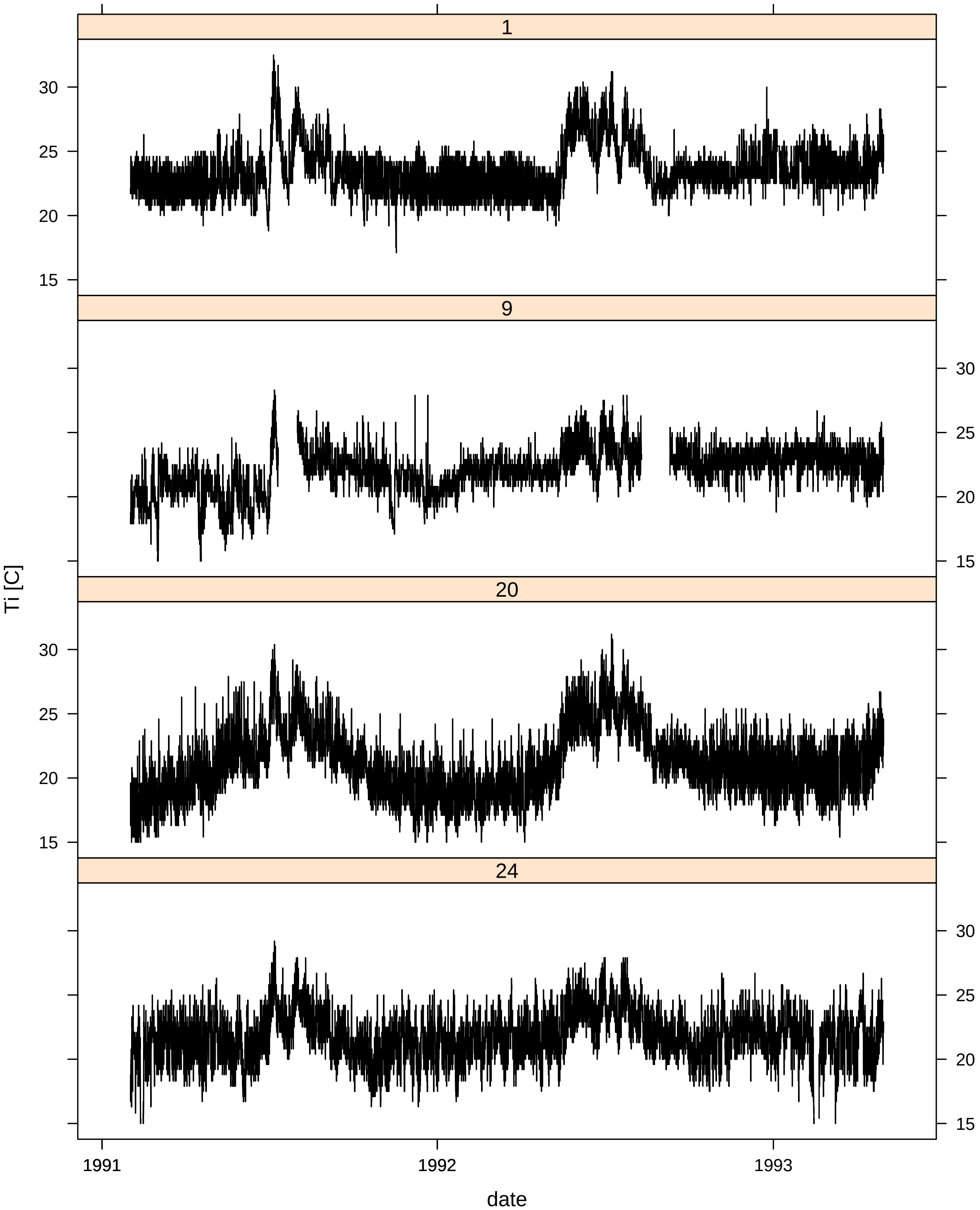
At this stage, however it is an open question how to deal with houses where a wood burning stove is used extensively. One possibility may be to model the heat supplied from the wood burning stove as an unobserved state in a state space model [5]. However, such an appoach is rather advanced and may require some sensor-input which can indicate if the wood burning stove is in use or not.

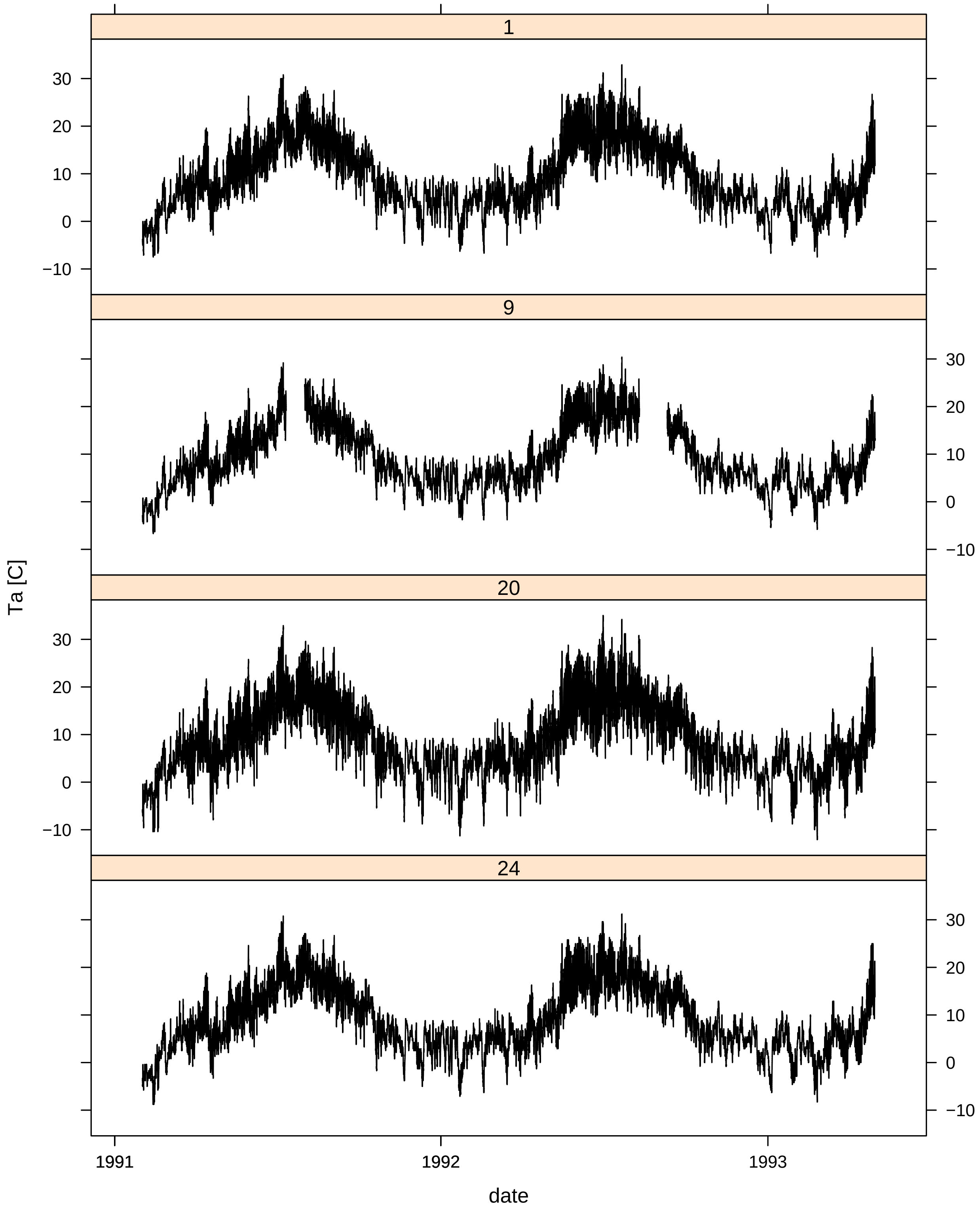
A Plots of data

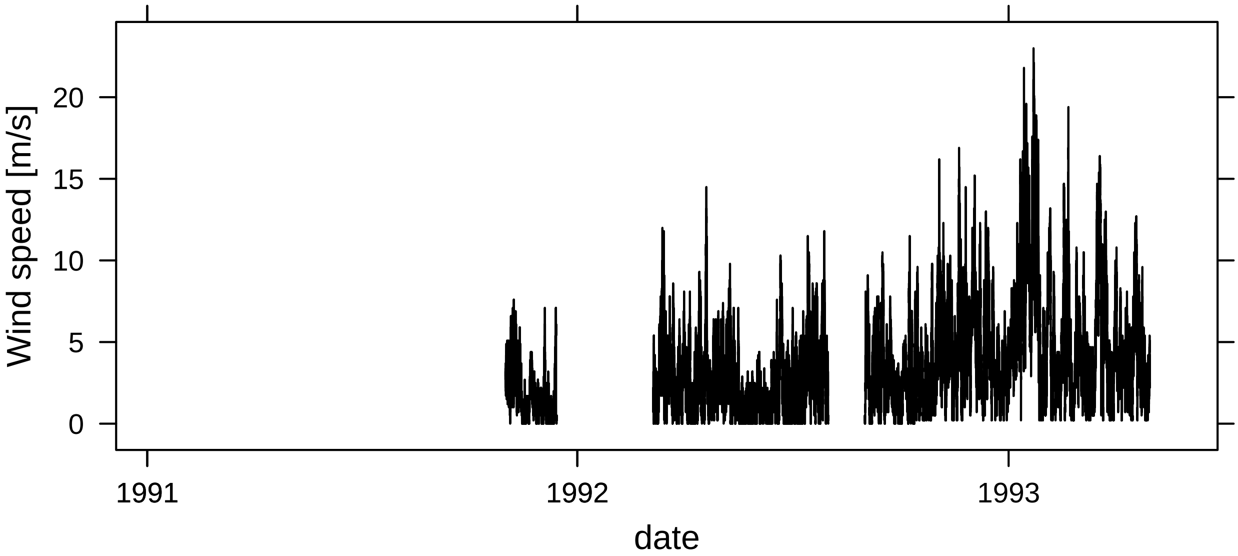



1
9
20
24
0
1
2
3
4
heat[kWh/15min]
1991
1992
1993
date

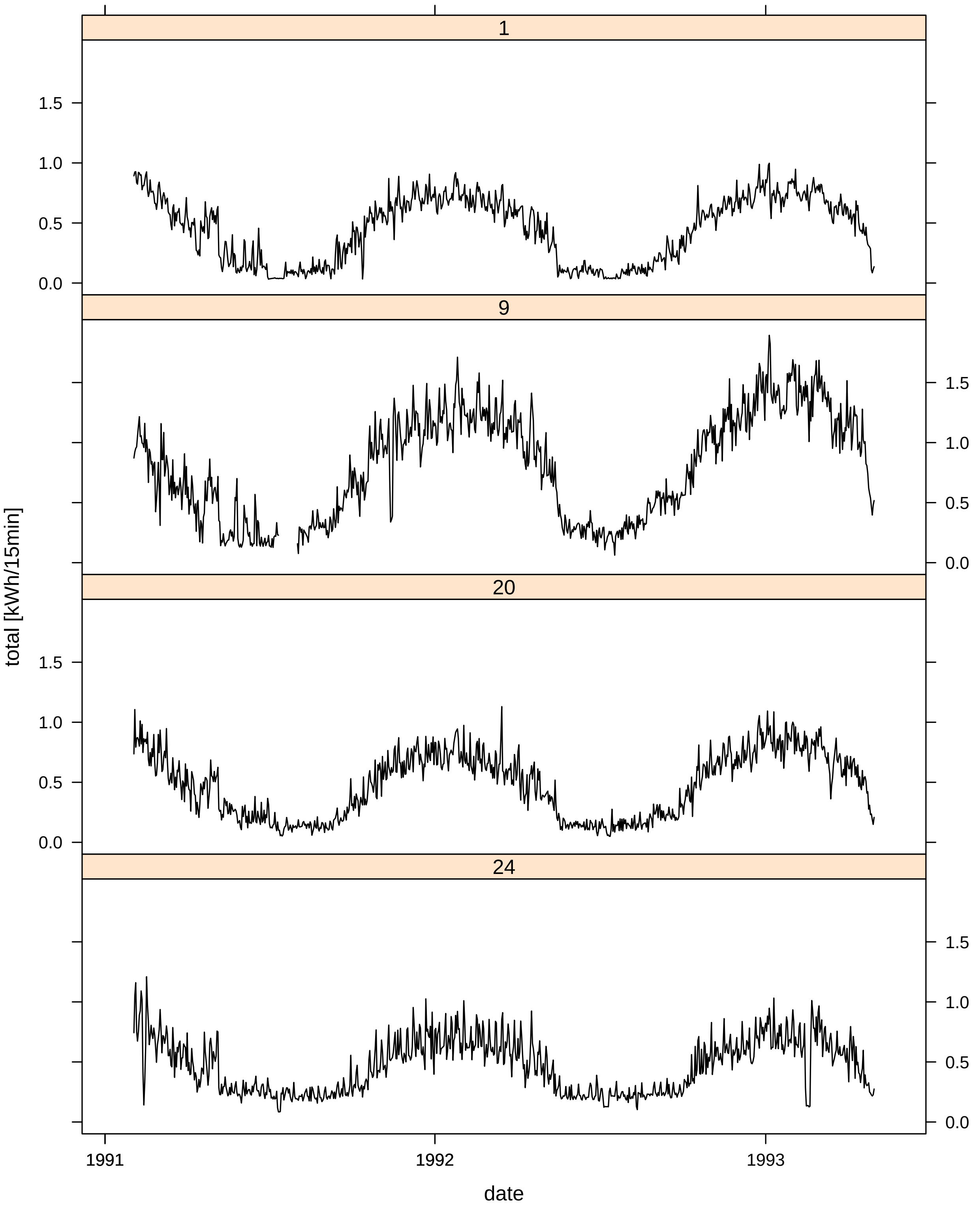

1
9
20
24
water [kWh/15min]
0.0
0.5
1.0
1.5
1991
1992
1993
date

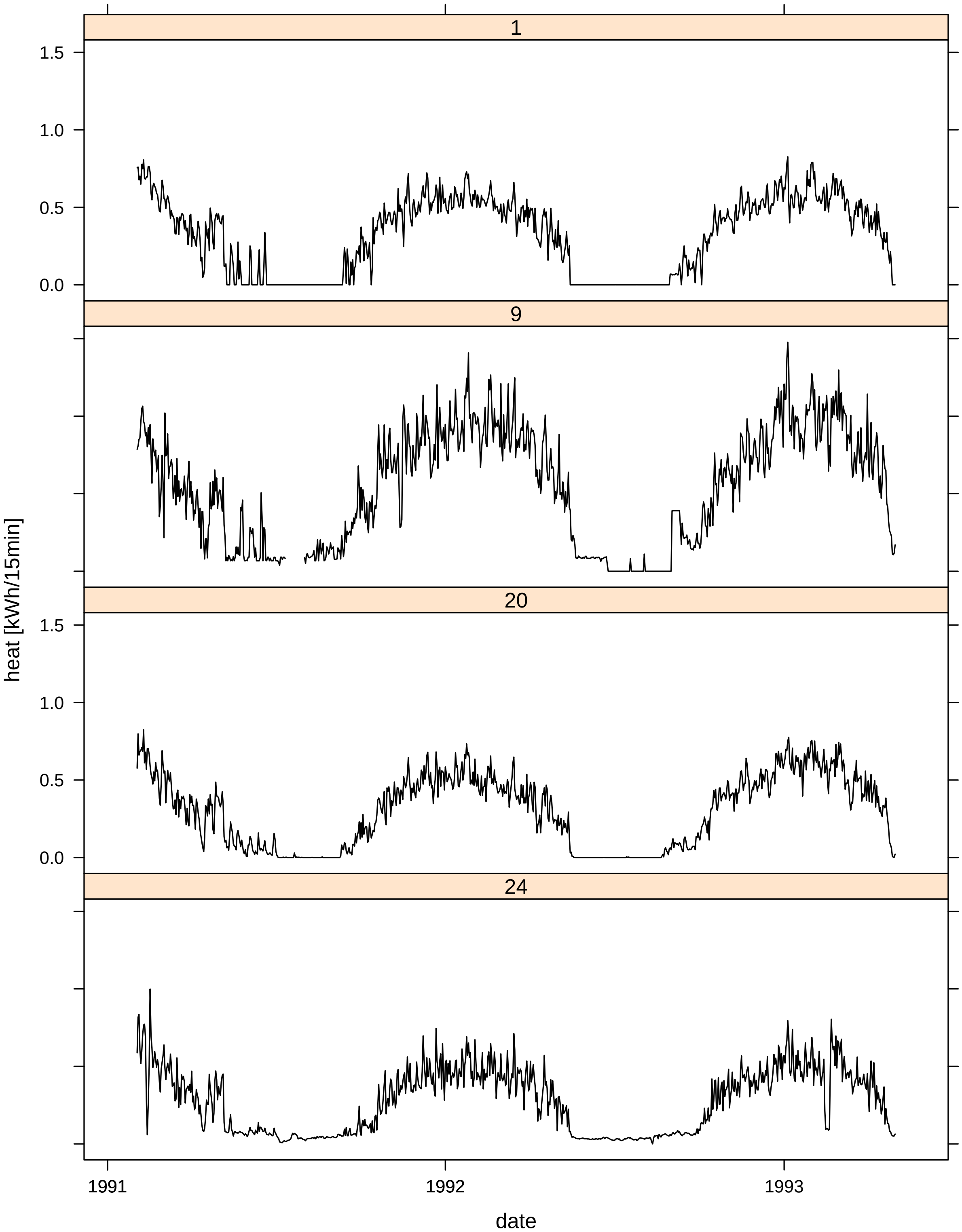

1
9
20
24
30
25
20
15
Ti [C]
1991
1992
1993
date

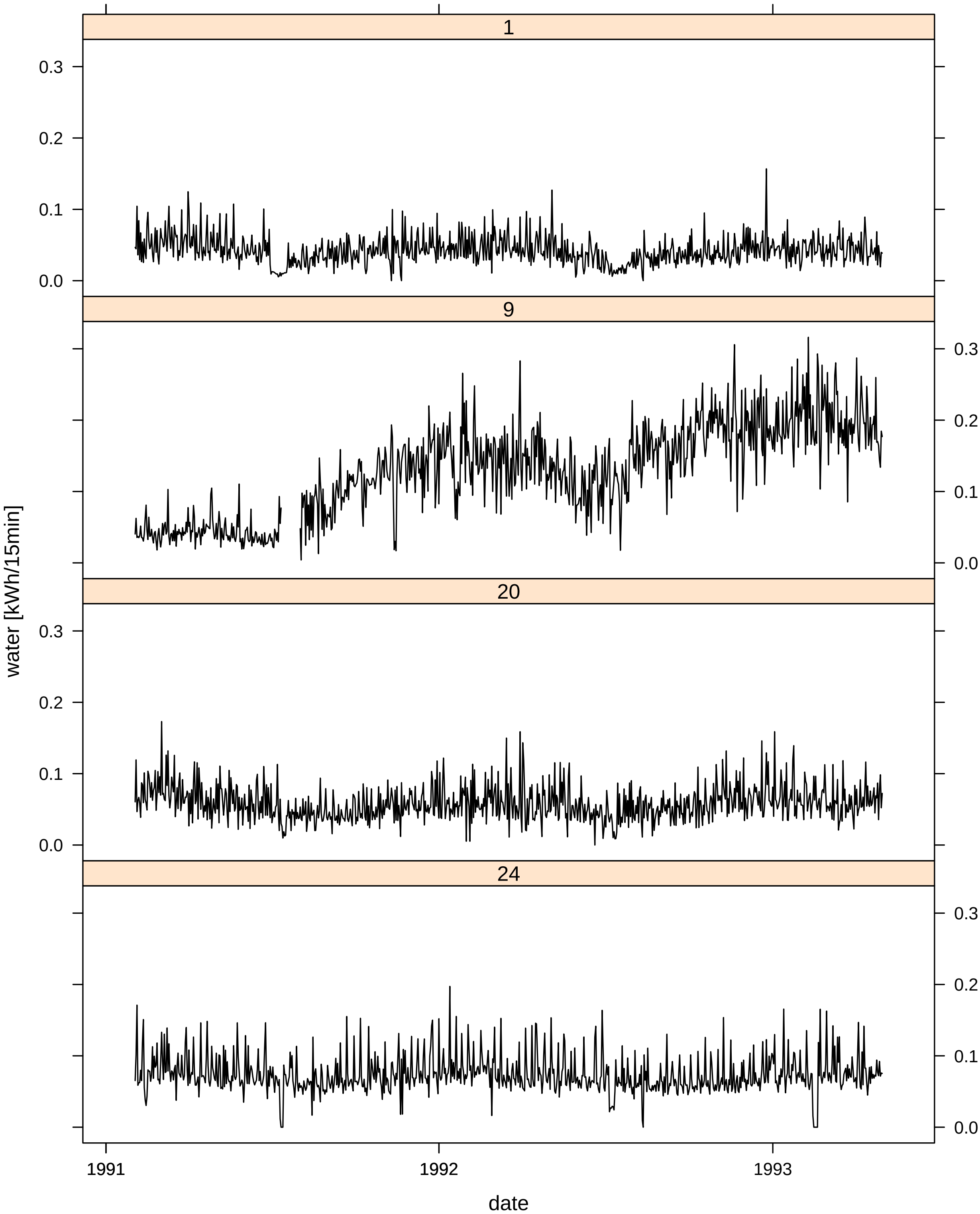

1
9
20
24
30
20
10
0
−10
Ta [C]
1991
1992
1993
date

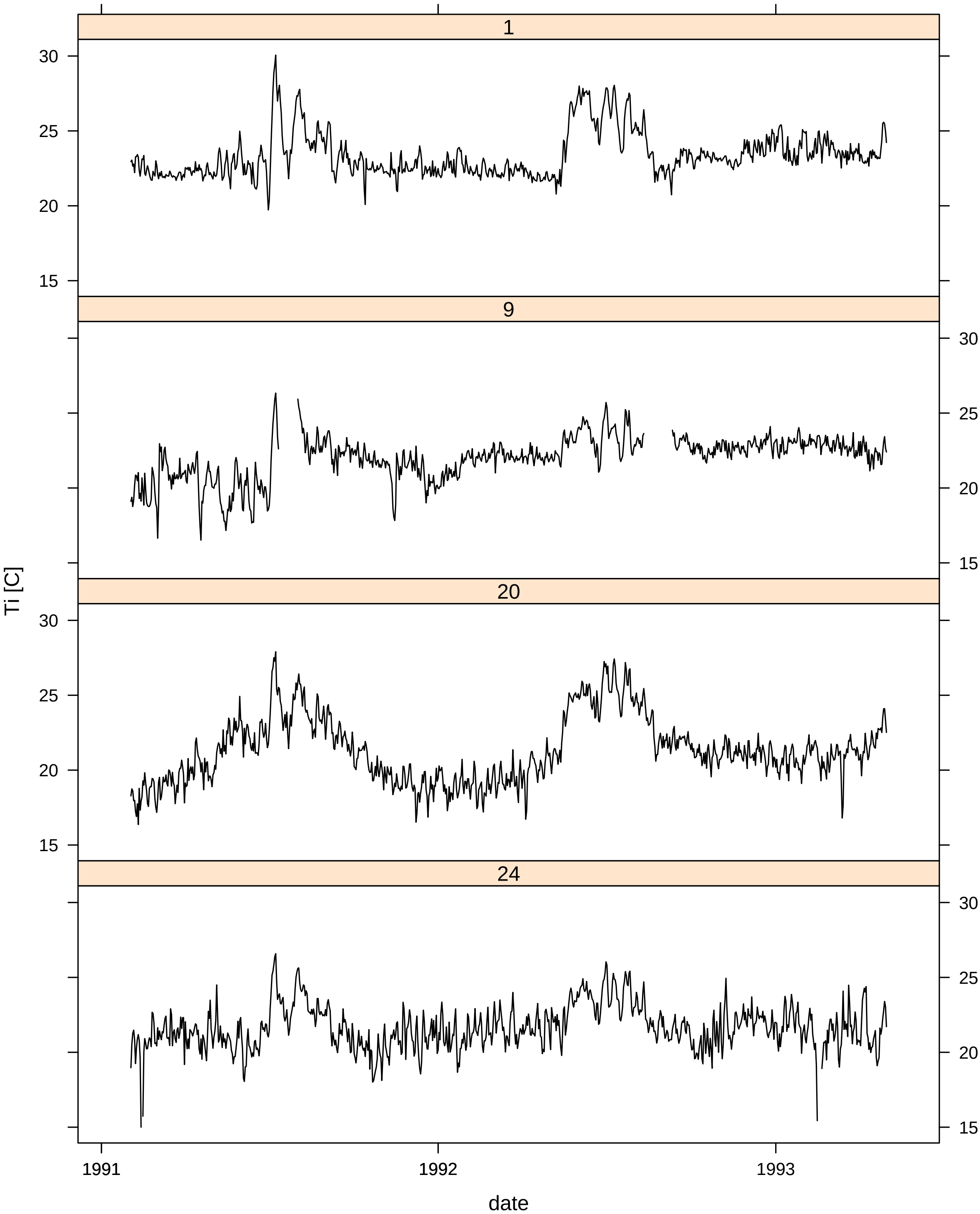

Wind speed [m/s]
0
5
10
15
20
1991
1992
1993
date

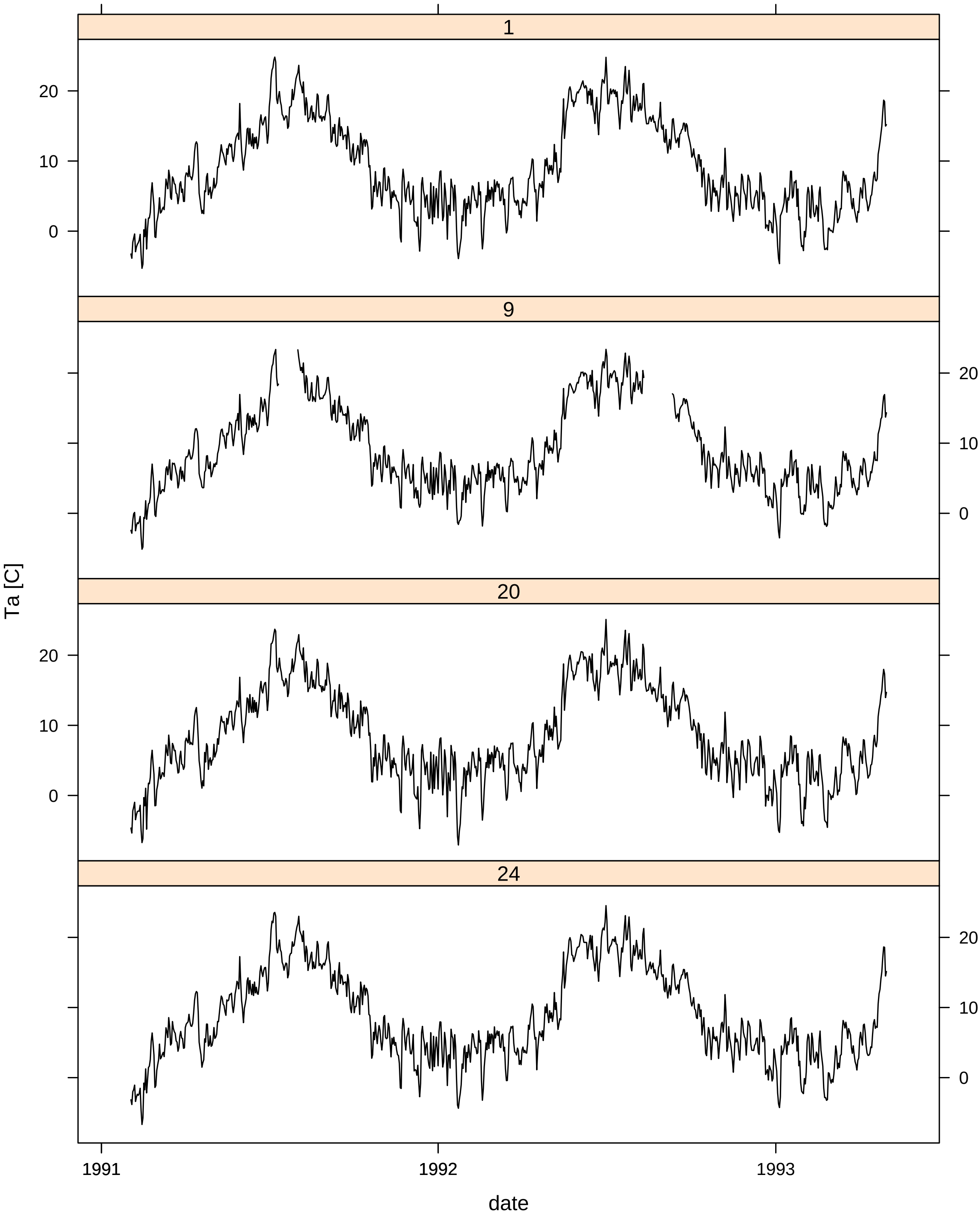
A Plots of daily averages




1
9
20
24
heat [kWh/15min]
0.0
0.5
1.0
1.5
1991
1992
1993
date


1
9
20
24
0.0
0.1
0.2
0.3
water [kWh/15min]
1991
1992
1993
date


1
9
20
24
30
25
20
15
Ti [C]
1991
1992
1993
date


1
9
20
24
20
10
0
Ta [C]
1991
1992
1993
date

References

[1] William S. Cleveland. In T. W. Anderson, K. T. Fang, and I. Olkin, editors, *Multivariate Analysis and Its Applications*, chapter Coplots, Nonparametric Regression, and conditionally Parametric Fits, pages 21–36. Institute of Mathematical Statistics, Hayward, 1994.

[2] William S. Cleveland and Susan J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83: 596–610, 1988.

[3] Richard A. Johnson. *Miller & Freund's Probability and Statistics for Engineers*. Pearson Prentice Hall, 7 edition, 2005.

[4] Roger Koenker. *Quantile Regression*. Cambridge University Press, 2005.

[5] Henrik Madsen. Time series analysis. Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, 2001. 316 pages.

[6] Henrik Aalborg Nielsen. LFLM version 1.0, an S-PLUS / R library for locally weighted fitting of linear models. Technical Report 22, Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark, 1997.

[7] Henrik Aalborg Nielsen and Henrik Madsen. *Predicting the Heat Consumption in District Heating Systems using Meteorological Forecasts*. Department of Mathematical Modelling, Technical University of Denmark, DK-2800 Lyngby, Denmark, 2000. URL `http://www.imm.dtu.dk/~han/pub/efp98.pdf`. ENS. J. Nr. 1323/98-0025.