Excercise 3 - Grey-box models (continued)

Models for the heat dynamics of a building

Marco Hernandez Velasco

August 2018

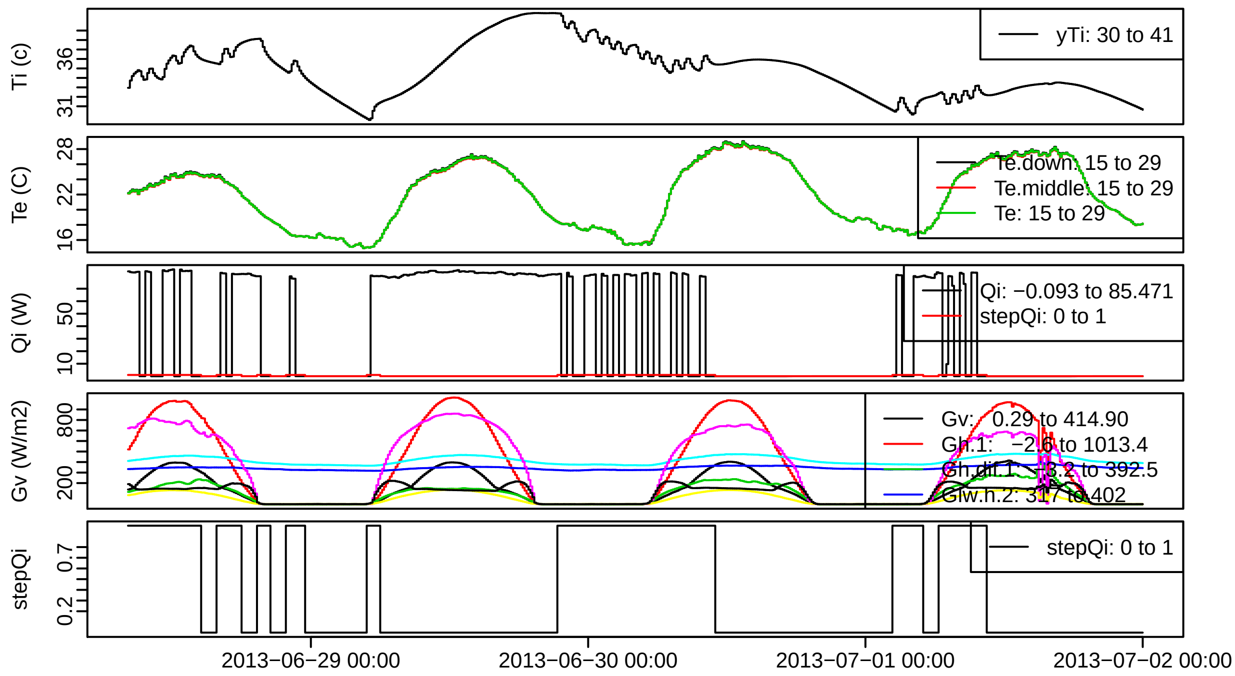
The exercise is focused on grey-box modelling of the heat dynamics of a (small) building using stochastic differential equations (SDEs). In addition to the first exercise on greybox modelling, we will in this exercise test different techniques to:

1. Alter the noise level or system uncertainty to account for e.g. non-linear phenomena.
2. Build a semi-parametric model to take into account that the solar penetration (i.e. relation between measured solar radiation and radiation entering into the building) as function of the position of the sun.
3. Balance heat gains to the air temperature and the temperature of the thermal mass.

The data consists of several measurement from a small test box with a single window. In this exercise the following signals are used:

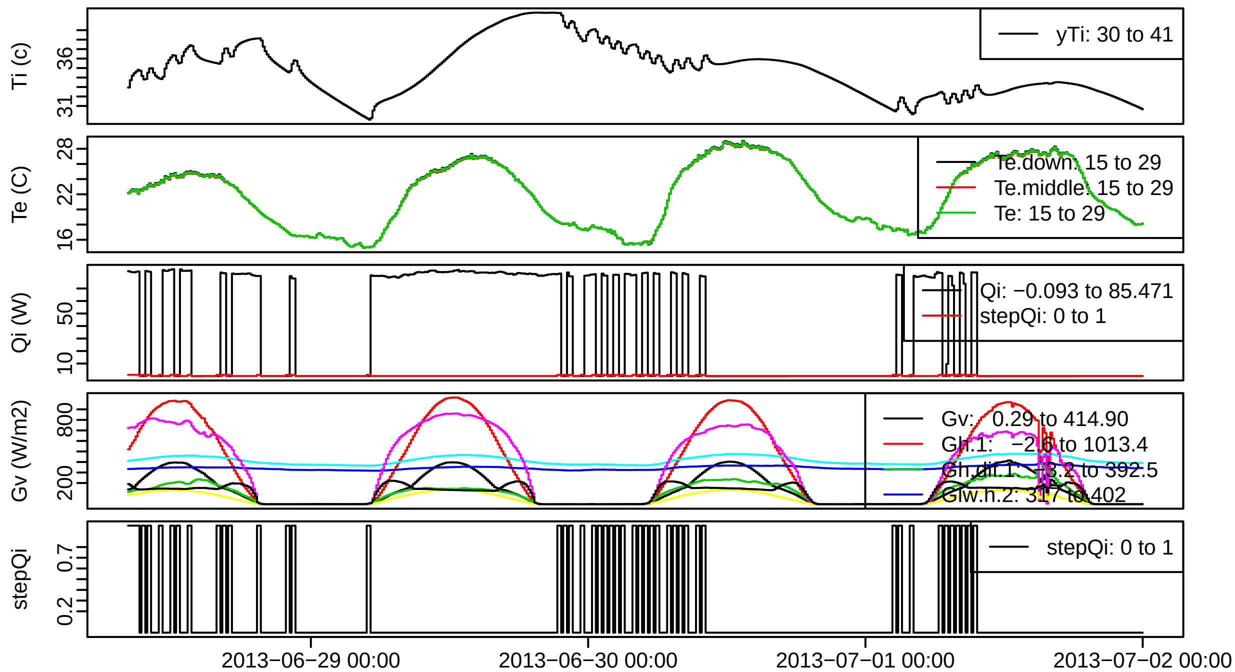
- Ti (yTi in data) the observed indoor temperatures. (C)
- Qi (Qi in data) the heat emitted by the electrical heaters in the test box (W)
- Te (Te in data) the ambient temperature (C)
- Gv (Gv in the data) the vertical south total solar radiation (W/m2)
- Gvn (Gvn in data) the vertical north total solar radiation (W/m2)

Question 1



- The lower time series plot is of stepQi, which goes from 0 to 1. Try to change the argument samples_after_Qi_step in the function preparing the data. How does it change stepQi?

Changing the stepQi makes the ON or OFF intervals longer or shorter and makes it able to correspond better to the actual heater's behavior. Like below with a step = 0.5

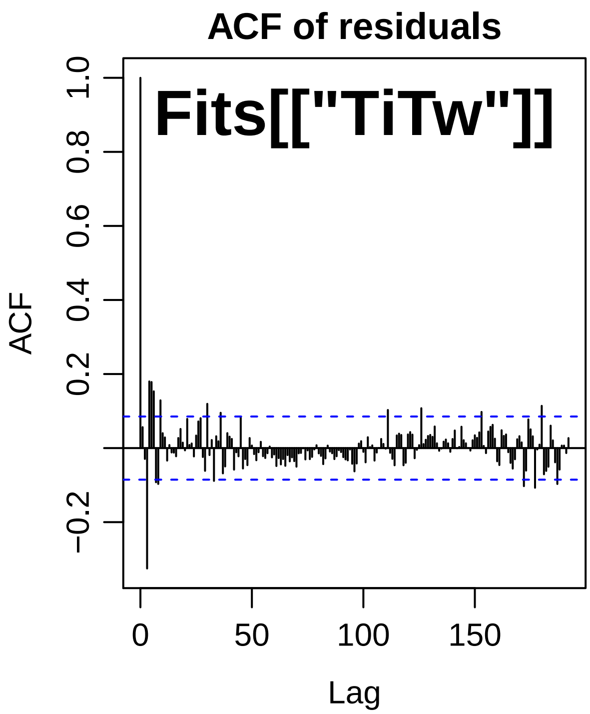


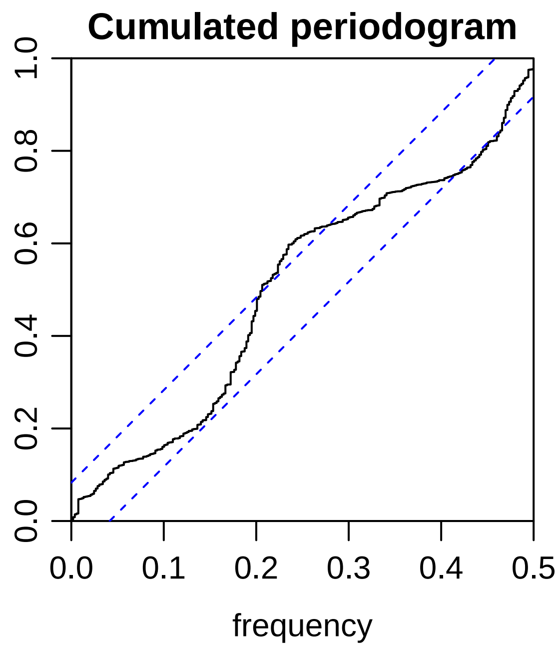
Comparing the two models

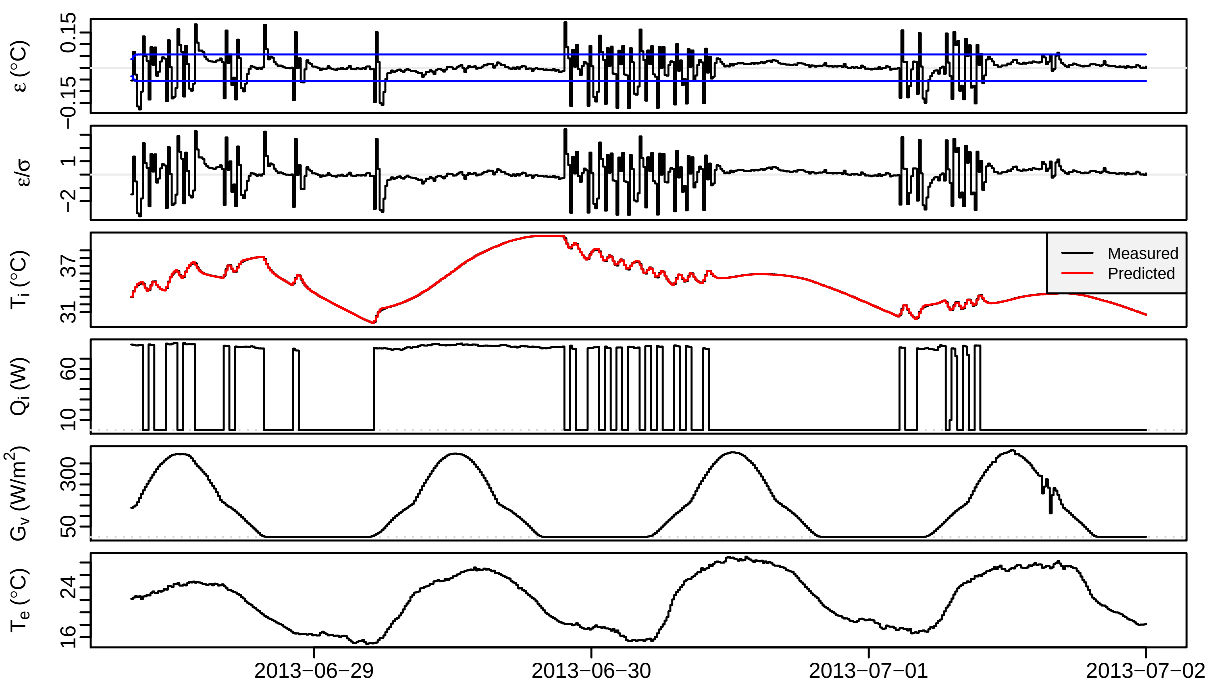
- Now compare the two models implemented in **functions/sdeTiTw.R** and **functions/sdeTiTw__sigmalevels.R**. What is the difference?

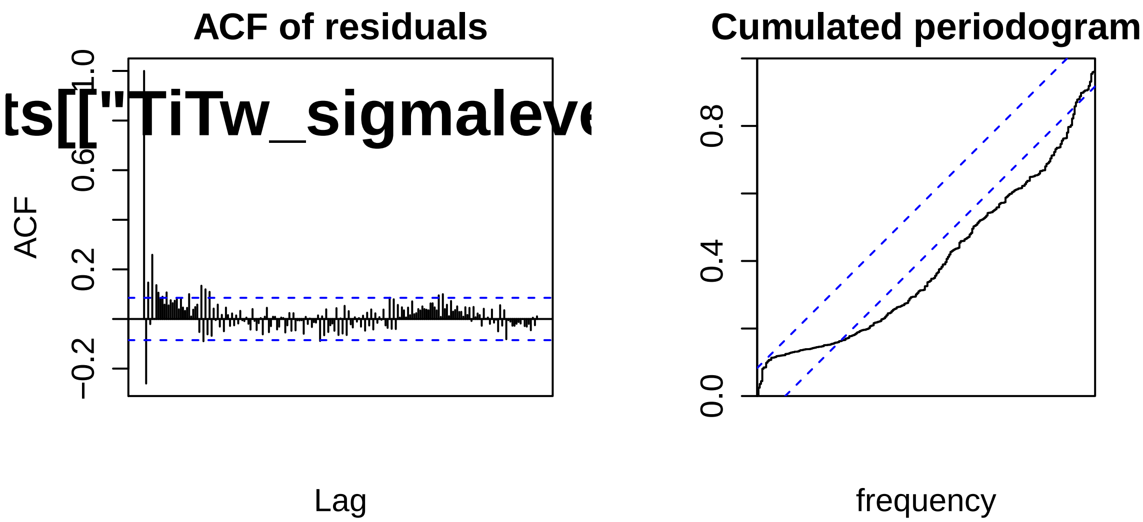
The second model includes with sigma levels include: $(1 + (stepQi * sigmalevel))$ *in the error term allowing to account for the noise of turning ON and OFF the heater.*

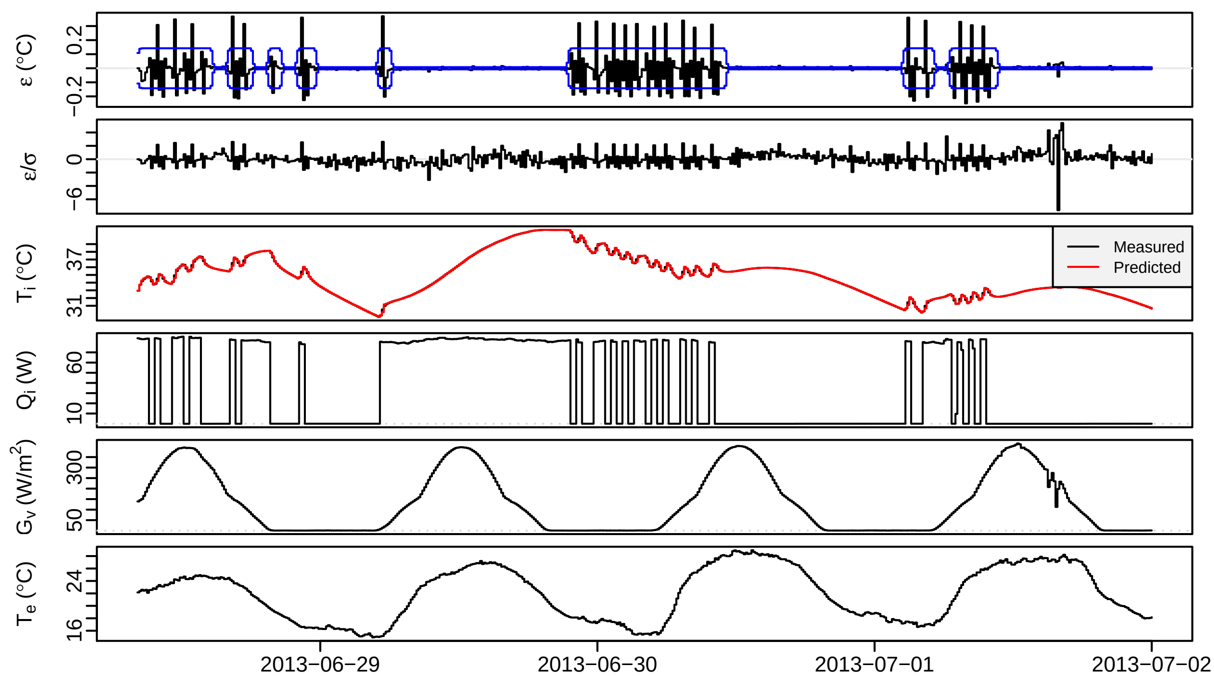
Now go to the script and fit the two models. Compare the results:


ACF of residuals
Fits[["TiTw"]]
ACF
Lag
0
50
100
150
−0.2
0.2
0.4
0.6
0.8
1.0


Cumulated periodogram
frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8
1.0


ε (°C)
ε/σ
Measured
Predicted
2013−06−29
2013−06−30
2013−07−01
2013−07−02





- What is plotted in the upper two plots? (You maybe have to look into the analyzeFit() function).

The upper plot is the residuals of the predicted variable **yTi**. *The lower plot is the standarized residuals (residuals / sd(yTi)), the y-axis changes as the errors are now divided by the standard deviation (σ)*

- What is indicated by the blue lines in the upper plot? Step back in the plots and compare the results, and look at the summary output.

The blue lines in the upper plot are the standard deviation of the predicted variable **yTi** *by which the residuals are divided to be standarized in the second graph below.*

- Step back in the plots and compare the results, and look at the summary output. Which of the two models will you prefer and why?

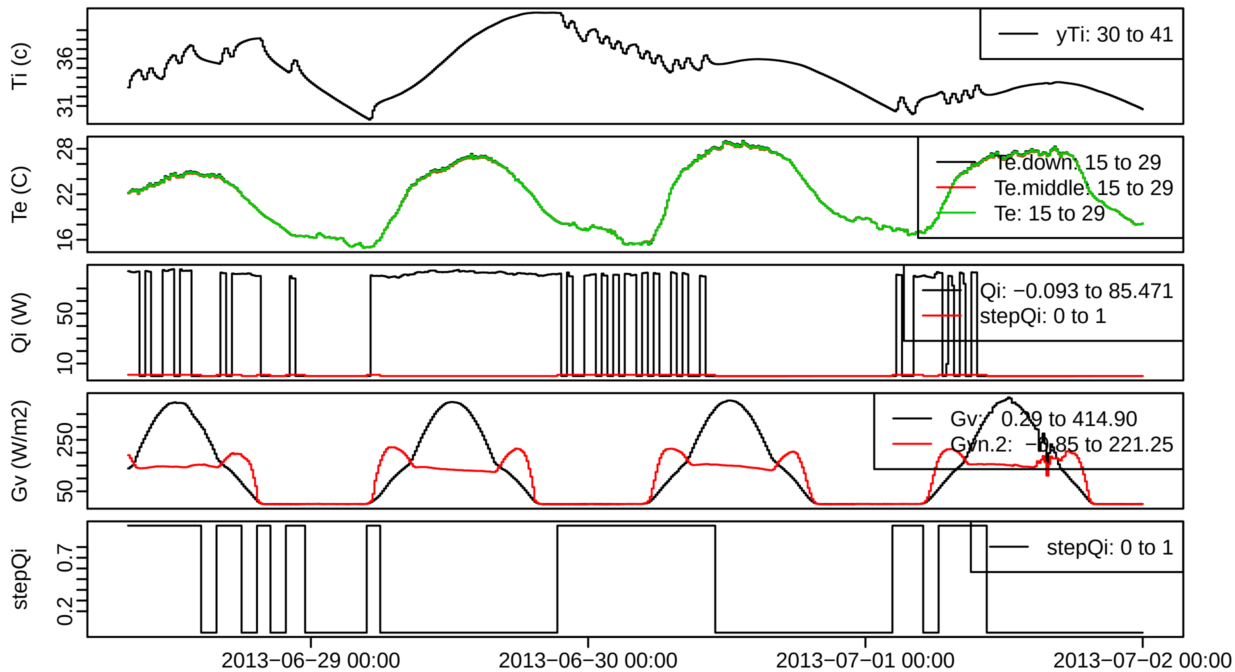
The one with sigma level because it allows the variance to change with time and captures better the variability of the process and that is shown in its residuals. When comparing their Log-Likelihood, also the model with sigma-levels has a larger likelihood.

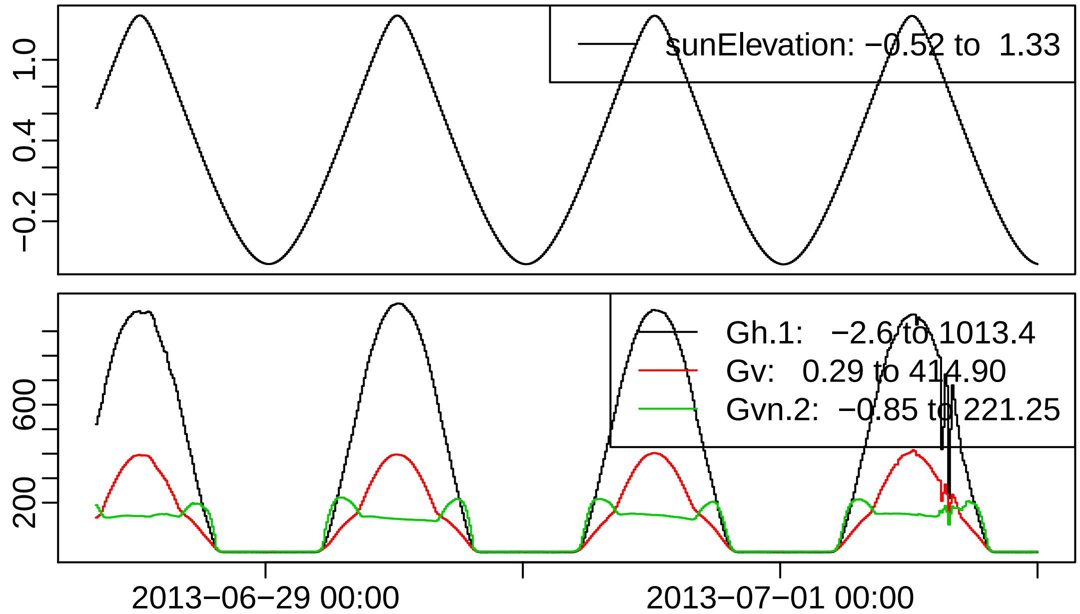
So it becomes clear that we have some (possible non-linear) dynamics when the heating turns on and off, which our models doesn't predict so well. But **instead of adding a more detailed description to the deterministic part of the model, we simply vary the system noise, or in other words, change the uncertainty level of our states under under different conditions**. This is a very useful thing, since there will be many phenomena in buildings, especially occupied buildings, which will lead different to levels of noise, e.g., solar radiation and occupants doing funny things.

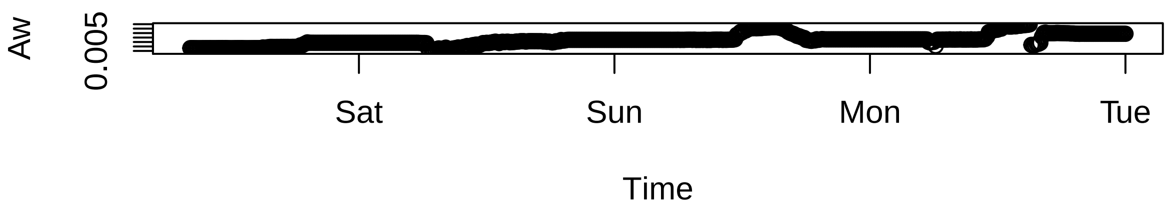
Question 2

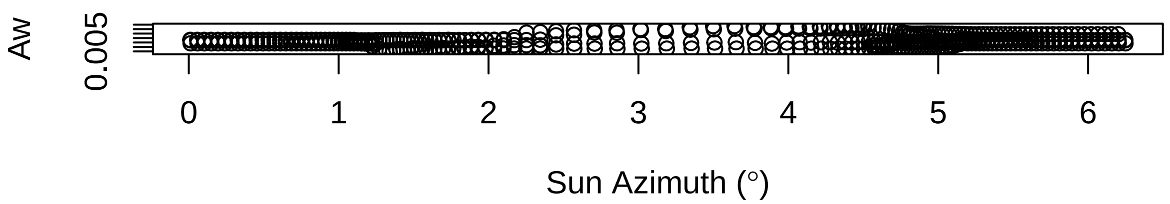
So far we have assumed that the solar gain is proportional to the radiation outside the test box. In reality, the heat gain from the sun depends highly on building geometry, surroundings, window properties, etc. In this part of the exercise, we will apply splines to estimate the solar heat gain as a function of solar position.

First we make a hidden state for Aw (also called the gA-value) to investigate if it changes over time and as the function of the sun position.





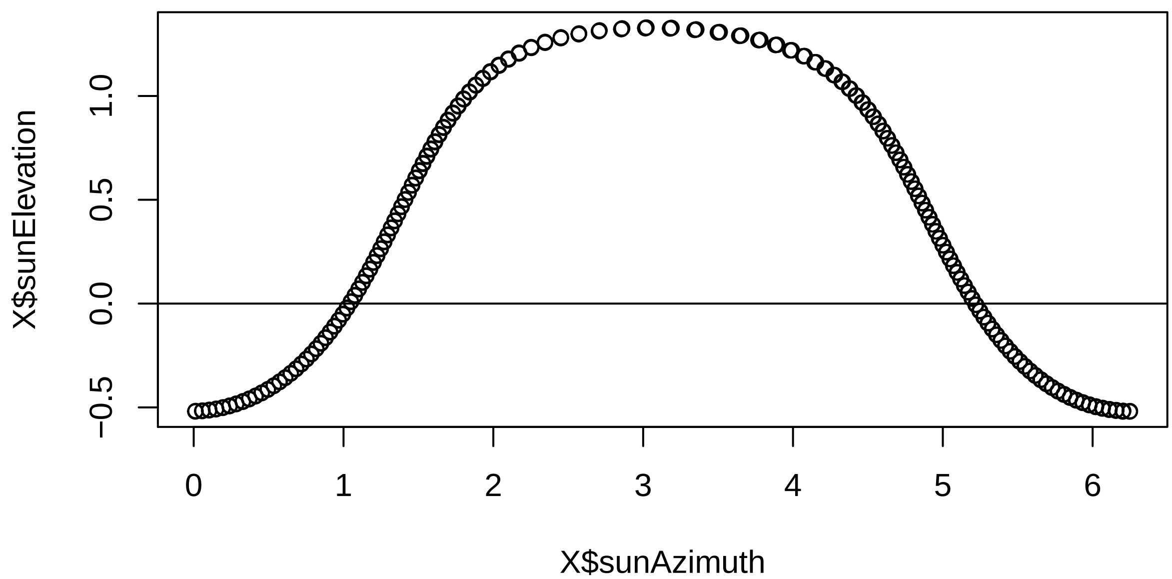




It is clear that the state of Aw is not constant, but changes. Furthermore, it seems like there could be a relation to the sun azimuth.

Now, answer the questions below as you progress in modelling the solar radiation with use of splines. The

spline function we want to estimate is the **gA value** (e.g. the percentage of solar heat that enters through the window, multiplied with the window area) as a function of the sun azimuth.
First, plot the sun elevation as a function of the sun azimuth, as well as a horizontal line through 0 (notice that the angles is in radians).



- Find the azimuth angles (in radians) that corresponds to the sunrise and sunset, and assign them to azumith_bound <- c(… , …) below. These two angles will in a moment be our boundary azimuth angles. Outside the boundaries the gA value is 0, as the sun is below the horizon and the radiation is zero. Thus, we are only interested in the gA values from sunrise to sunset.

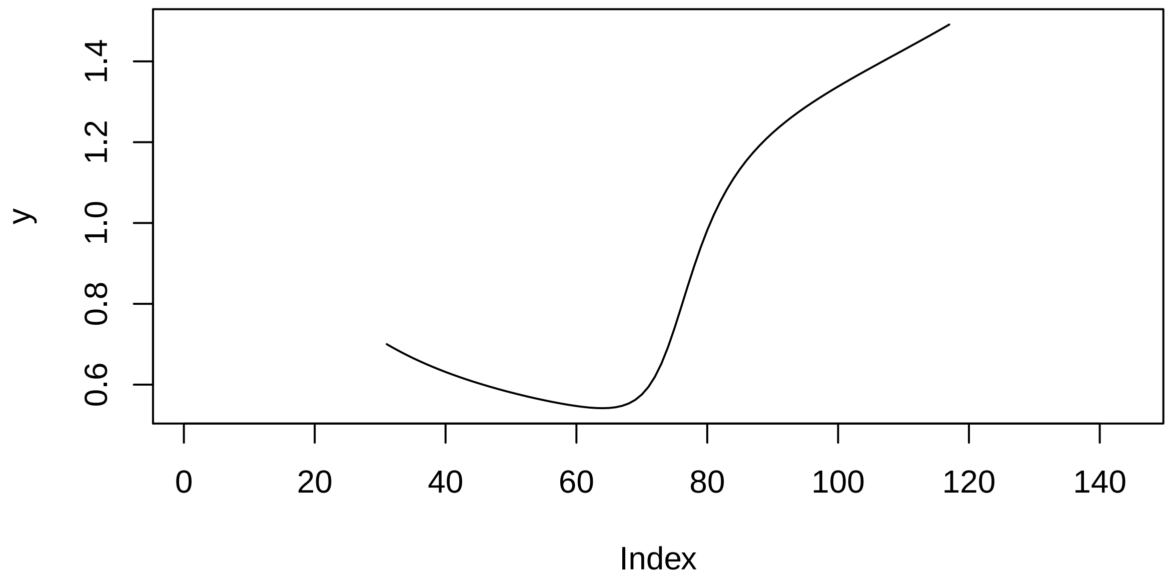
```
## Inset the boundary angles (in radians) below.
a <- X %>% select(sunAzimuth, sunElevation) %>%  #select relevant columns
      mutate(sunAz = round(sunAzimuth, 2), #round the azimuth to 2 decimals
             #obtain the sign of the elevation and add to previous/lagged sign
             #when there is a change it will be = 0
              sign_change = sign(sunElevation) + lag(sign(sunElevation))) %>%
      filter(sign_change == 0) %>%  #find the rows where it changes
      summarise(min_bound = min(sunAz), max_bound = max(sunAz)) #summarise to the min, max

#convert the above data frame to a vector of 2 values
azumith.bound<- as.vector(t(a)) #t() transposes the matrix to do it by columns

#is the same as:#azumith.bound <- c(1.05, 5.22)
```

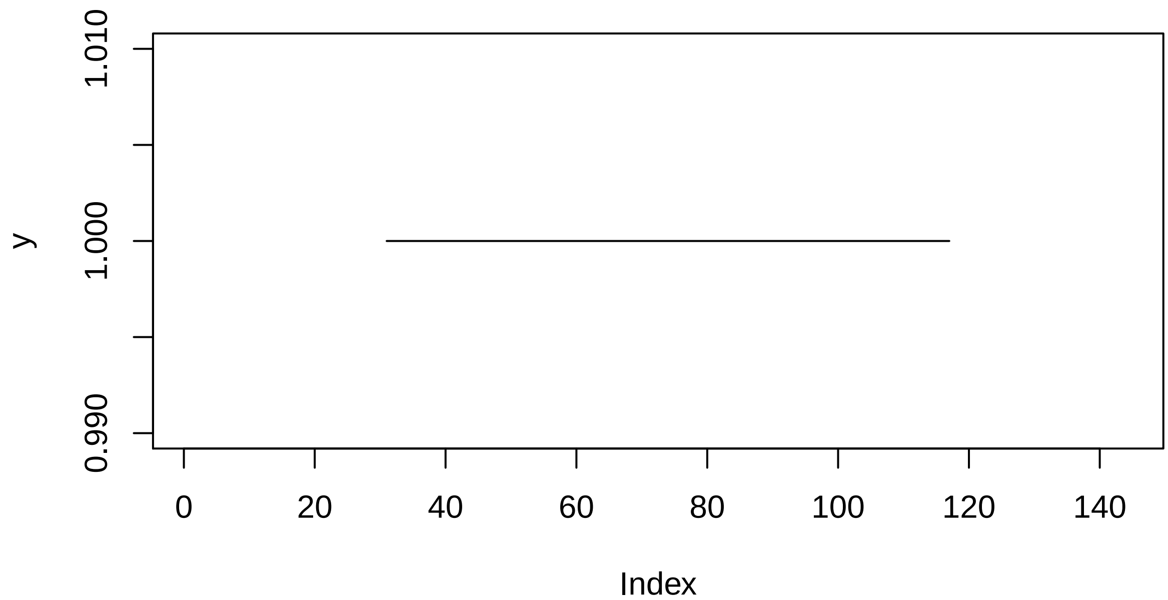
Define the base splines in the following lines of the script and stop after you have assigned the base splines to the data frame with the command X <- cbind(X,Xbs). Now play around with the four parameters in in the vector Aw, and plot the resulting spline function to get an understanding of how the base splines and the resulting spline function work.

(Pro tip: the package lubridate is very useful when working with dates and time. Which often is the case for when dealing with time series!)



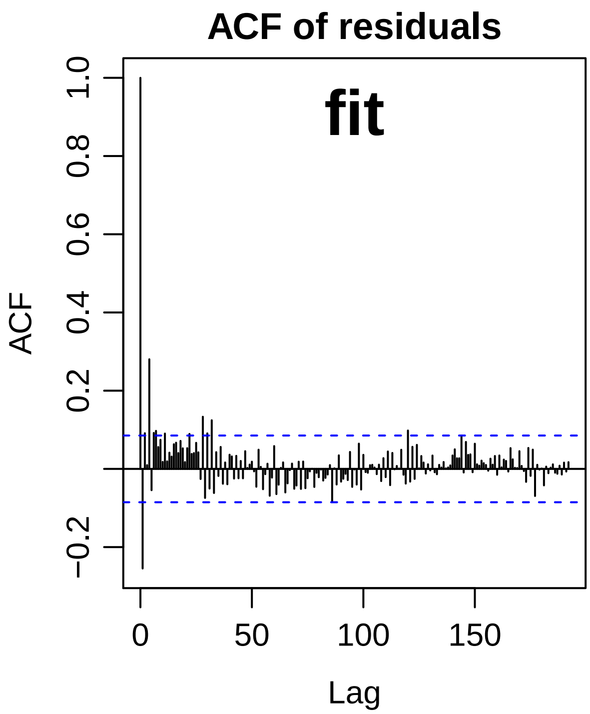
- What happens if Aw only consists of 1's?

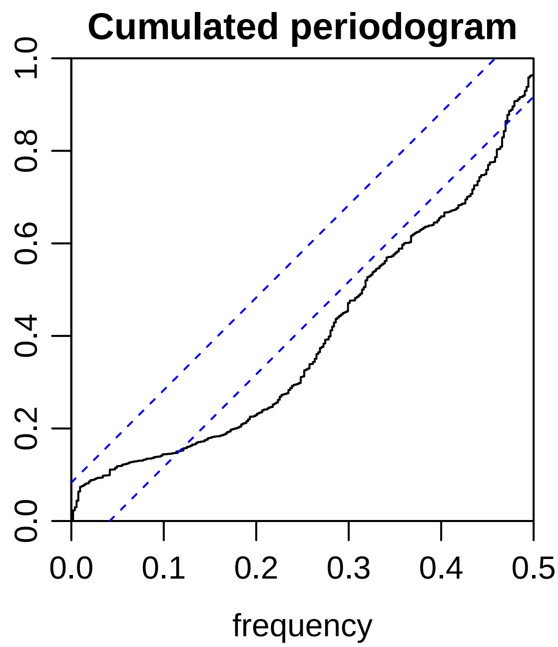
The result is just a straight line. These parameters determine how the spline function will look like. By multiplying the base splines to a solar radiation series and using this as input to an ARX model, the solar absorption coefficient (**gA-value** *) can vary as a smooth funiton of the sun azimuth angle.*

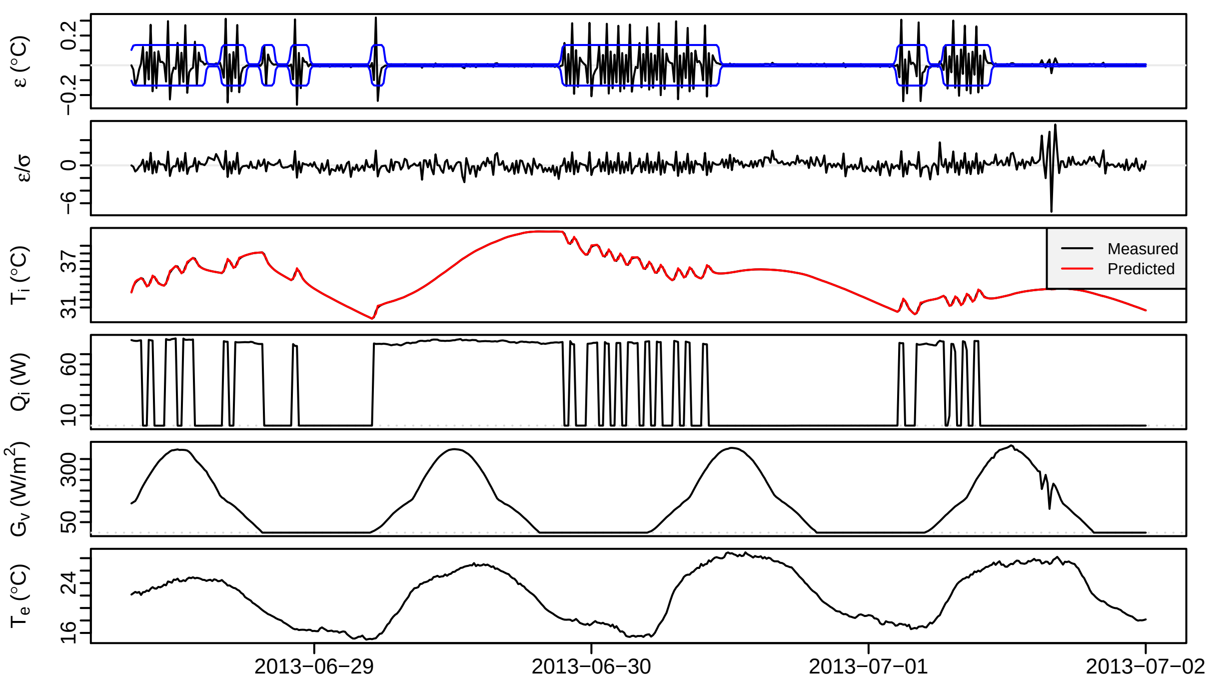


- Fit the model and investigate the estimated parameters. Is the parameters Aw1, Aw2, Aw3 and Aw4 significant, and is the magnitude reasonable when the actual glazed area is 52 x 52 cm?

The parameters Aw2 nad Aw4 are highly significant. Aw3 is significant but its value is very low and Aw1 is not significant. For the significant values of Aw2 and Aw4, the coefficient of 0.1 to 0.13 is close to the constant gA estimated previously and since this is the time where most solar radiation is entering the box through the window, the estiamted levels are found to be very reasonable.

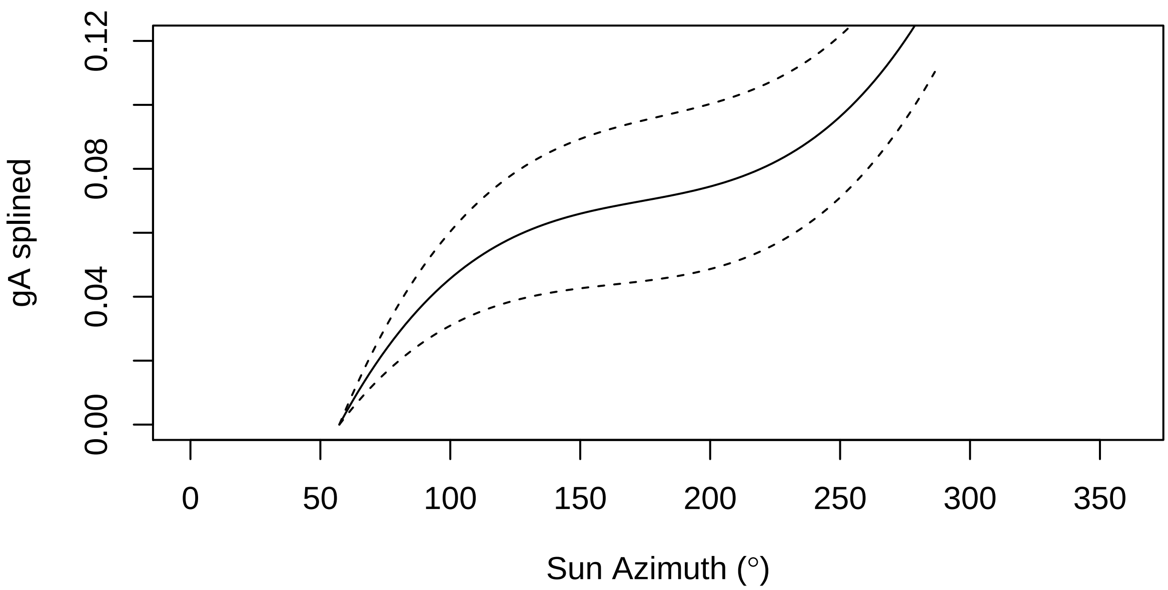




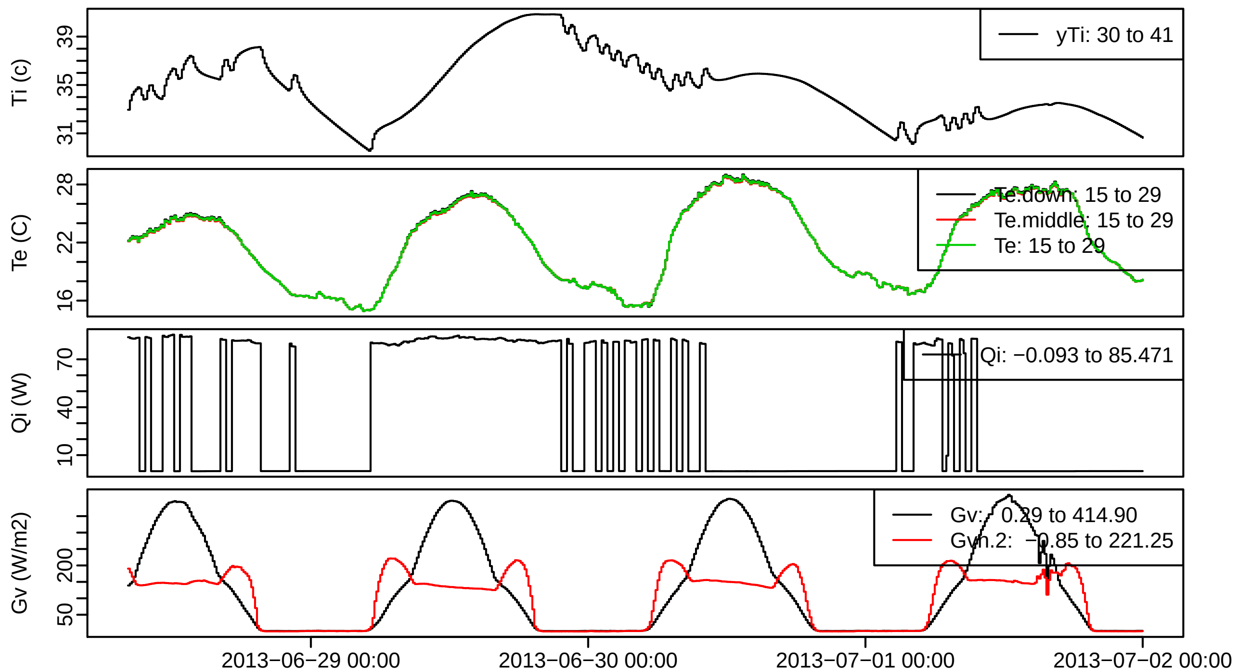


- Plot the gA curve and the 95 % confidence interval. The window in the test box is facing south towards an open area, and should therefore be rather unobstructed. Why do the gA curve then have a shape which is asymmetrical around the south (180 degrees)?

This could be due to the position of the vertical irradiance sensor which is on the east side of the box. This means that it would recieve radiation in the morning but in the afternoon the box itself would block some of the diffuse light since the Sun is not going down exactly at the West.

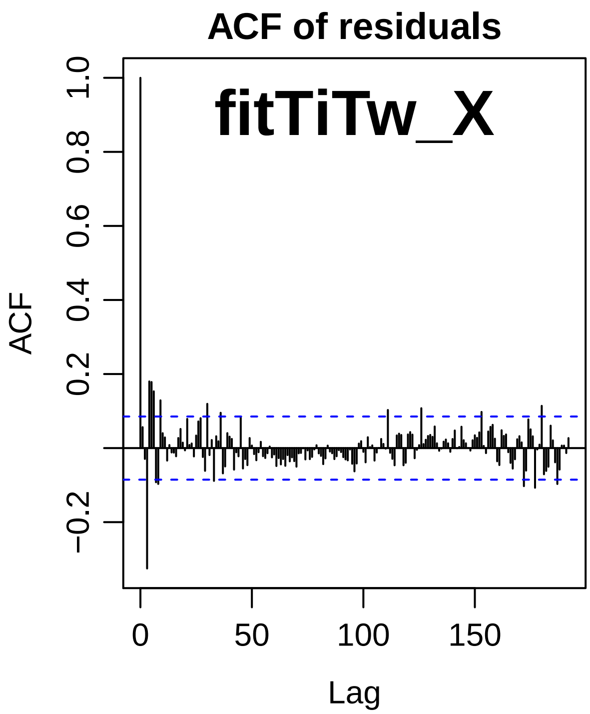


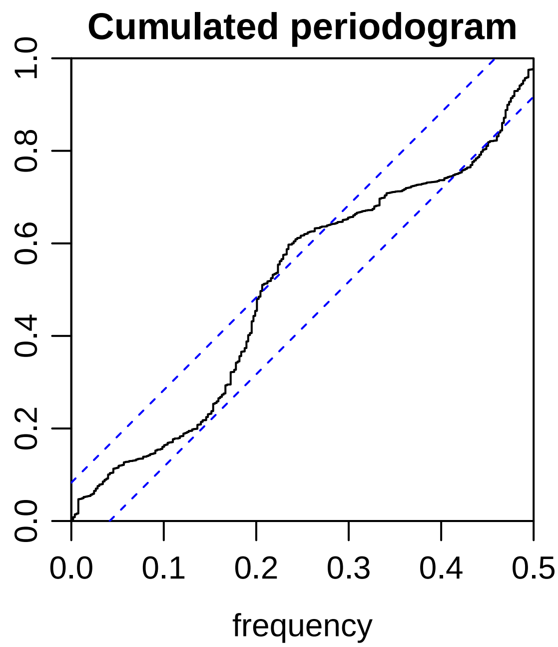
Question 3

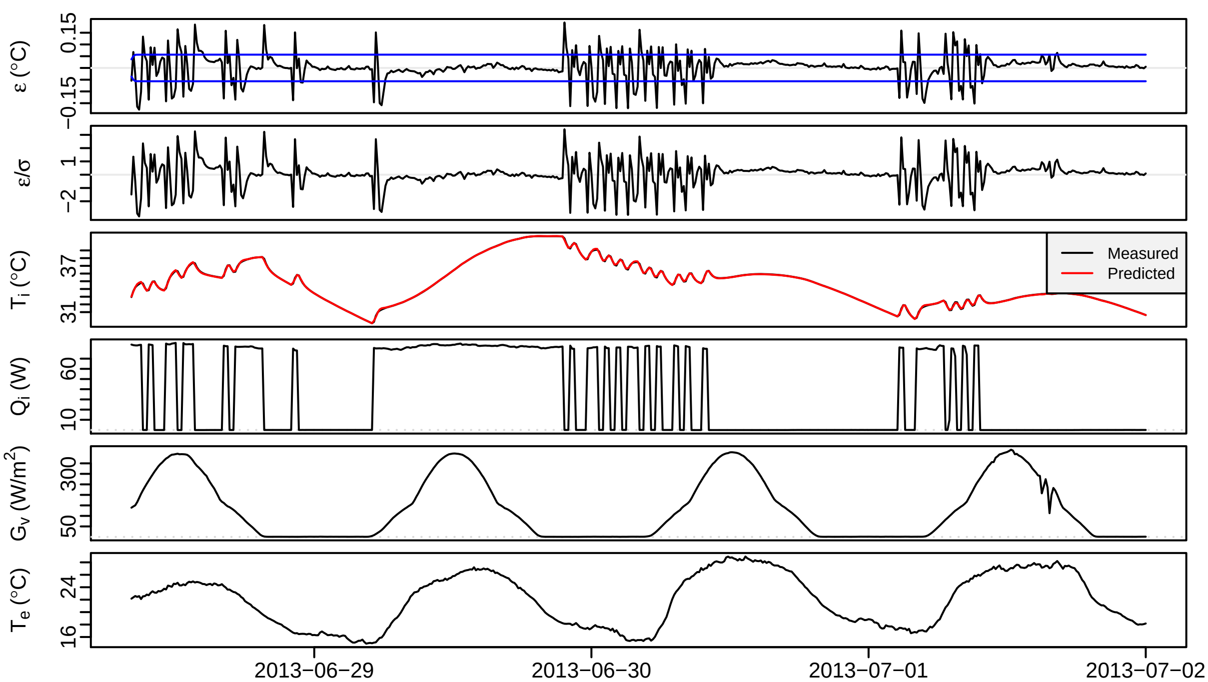


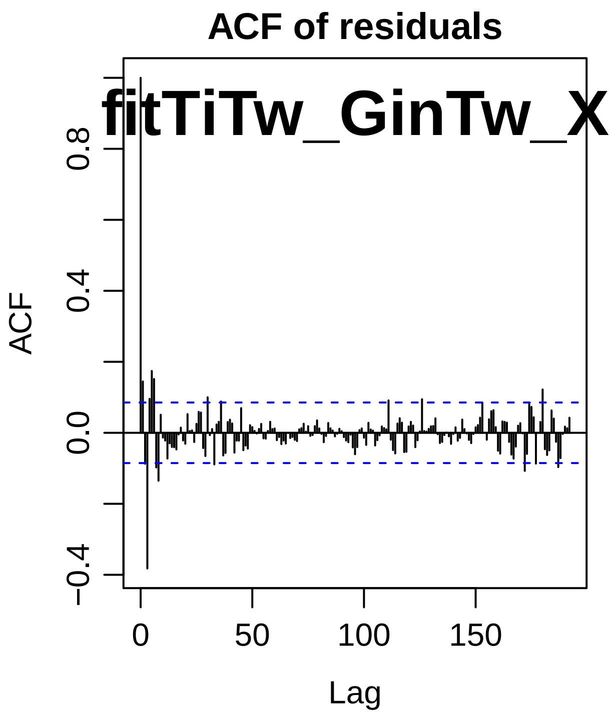
Estimate the parameters in the model **fitTiTw_X** and the following model, **fitTiTw_GinTw_X**. Which of the models has the highest (log) likelihood?

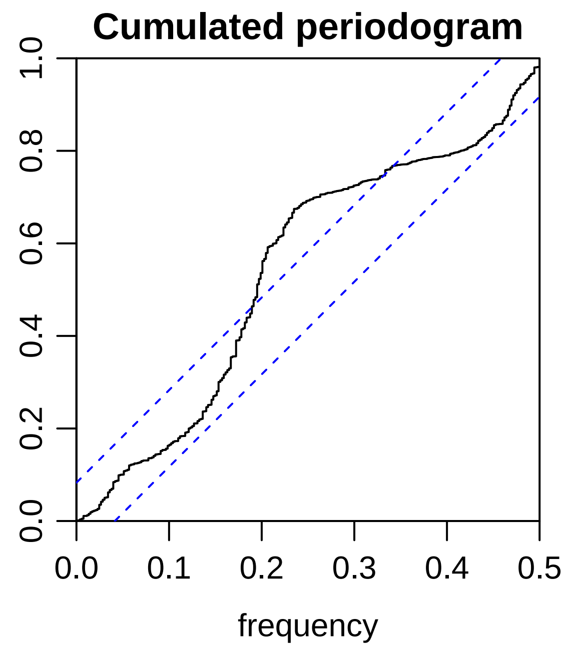
The model fitTiTw_X has a Loglikelihood of ***765.59*** *while fitTiTw_GinTw_X has a higher a Loglikelihood of* ***778.83****.*

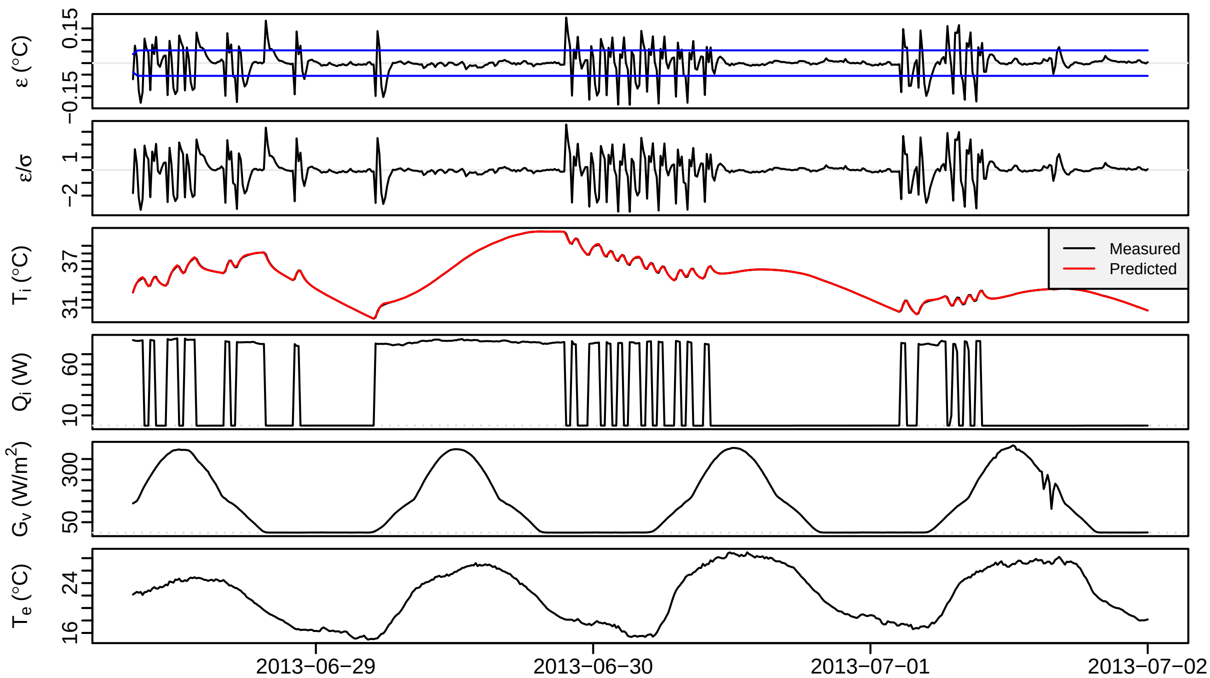

ACF of residuals
fitTiTw_X
ACF
Lag
0
50
100
150
−0.2
0.2
0.4
0.6
0.8
1.0


Cumulated periodogram
frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.2
0.4
0.6
0.8
1.0


ε (°C)
ε/σ
Measured
Predicted
2013−06−29
2013−06−30
2013−07−01
2013−07−02


ACF of residuals
fitTiTw_GinTw_X
ACF
Lag


Cumulated periodogram
frequency


Measured
Predicted
2013−06−29
2013−06−30
2013−07−01
2013−07−02

The second model has included a parameter, **p**, which is used to balance how much of the solar radiation that should be assigned to each of the two states, namely Ti and Tw.
Open the script for the function **sdeTiTw** and **sdeTiTw_GinTw** and compare them.

- How is the balancing actually done?

The fitTiTw_GinTw_X model includes p in the first state: $(Aw \cdot p/Ci \cdot Gv)$ *instead of* $(Aw/Ci \cdot Gv)$ *and includes a new term in the second state with (1-p) for doing the balancing:* $(Aw \cdot (1-p)/Cw \cdot Gv)$

- For which value of **p** does the model sdeTiTw_GinTw correspond to the first model sdeTiTw?

When $p = 1$ *both models are equal.*

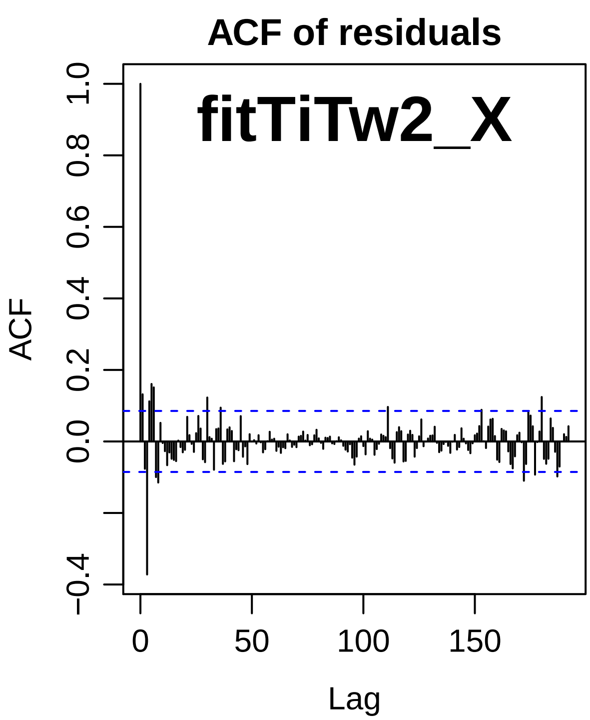
- What is the estimate of the parameter, **p**, and is it significantly different from 0?

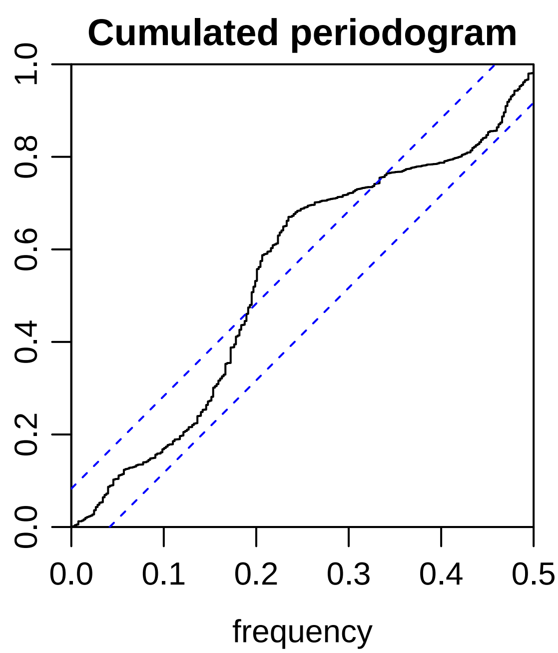
The estimate of the parameter is $p = 0.27$ *and is significantly different from 0*

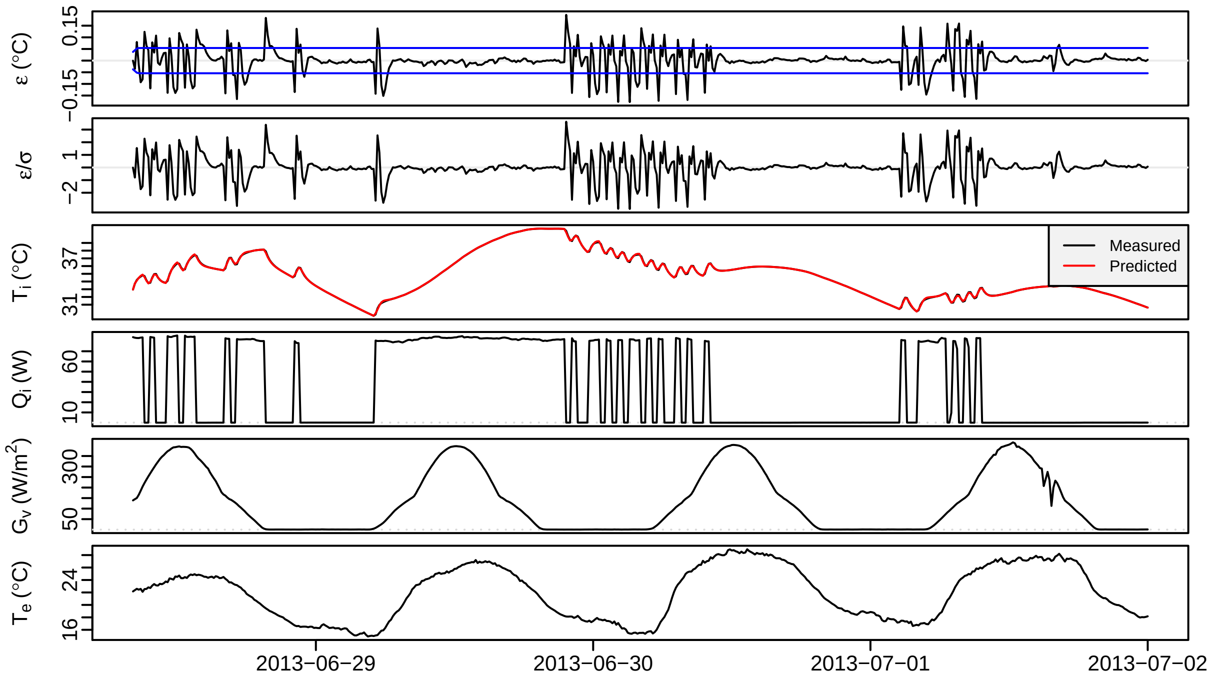
- Explain in words what the meaning of a small and a large value of p means. Based on the estimated parameter, **p**, is it reasonable to assume that the solar radiation entering the building solely should be assigned directly to the air temperature?

*The value of **p** determines how much of the entering solar radiation is heating the air in the room and how much is heating the walls. It should two be balanced between both, air and walls, because it is not reasonable to assume that only the air is affected.*

With the model **sdeTiTw_sigmalevels** as starting point, setup a model that includes two layers in the wall and a balancing parameter, p, to balance the solar radiation between Ti and Tw:



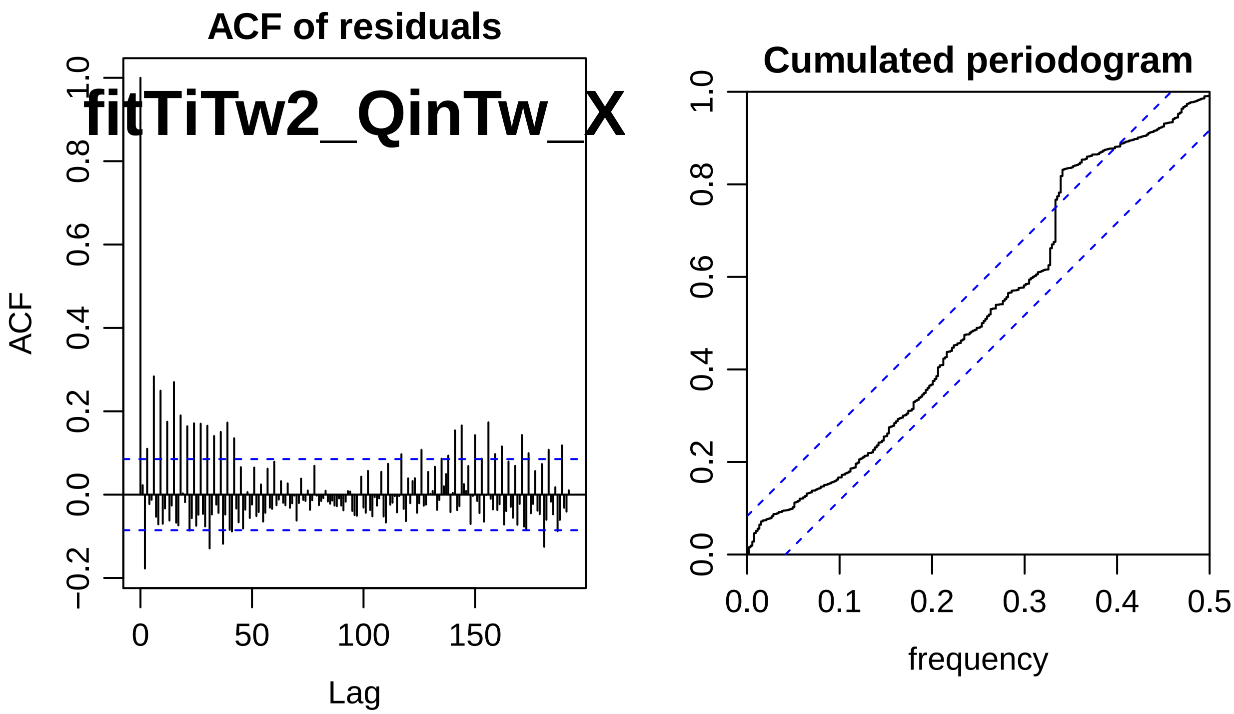


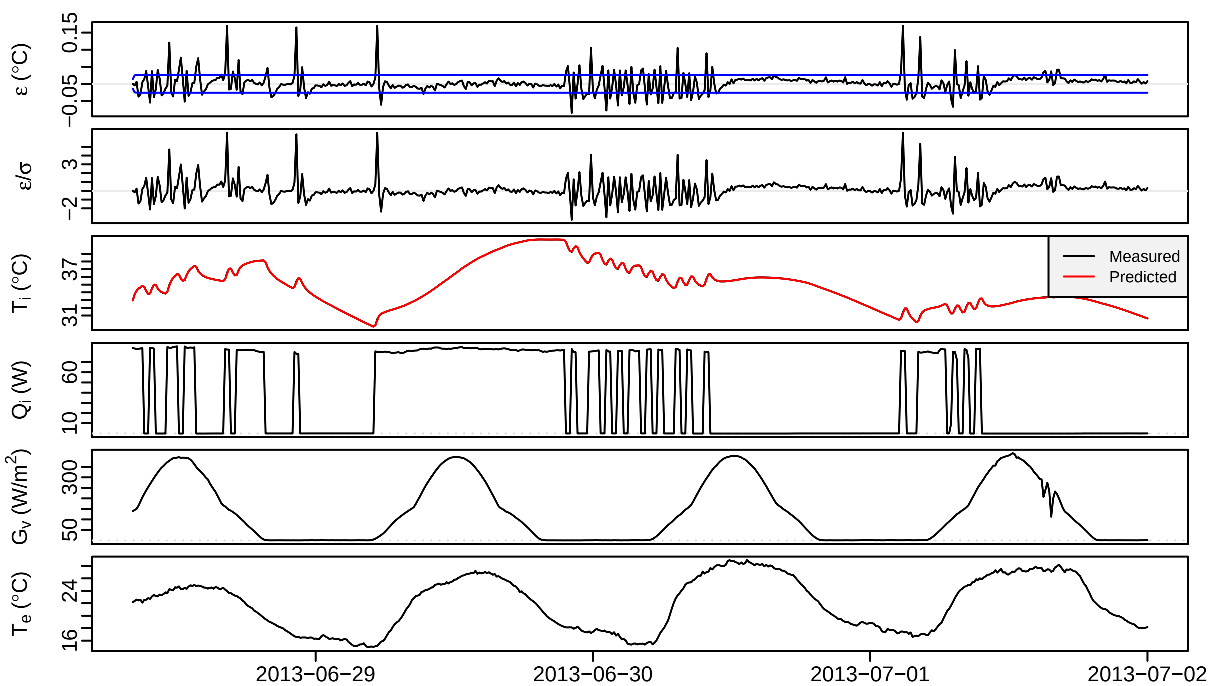


As it is the case for the solar radiation, the heat input from the heating system can also be assigned to different thermal capacities. Until now, we have assigned it directly to the indoor air temperature, *Ti*. To which state the heat should be assigned depends on the reaction time of the heating system. E.g. an electrical heat blower has a much faster responds time than a built-in floor heating system, and should most likely not be assigned to a state with very slow heat dynamics.

- Open the script of the function sdeTiTw2_QinTw and see how the model is made. Compared to the previous model, we have introduced an additional layer in the wall and assigned the heat input from the heating system to the inner wall.

Compared to the previous model sdeTiTw2, the sdeTiTw2_QinTw model removes $1/Ci \cdot Qi$ *from the indoor temperature state and instead adds Qi in form of* $1/Cw1 \cdot Qi$ *to the first wall state Tw1.*

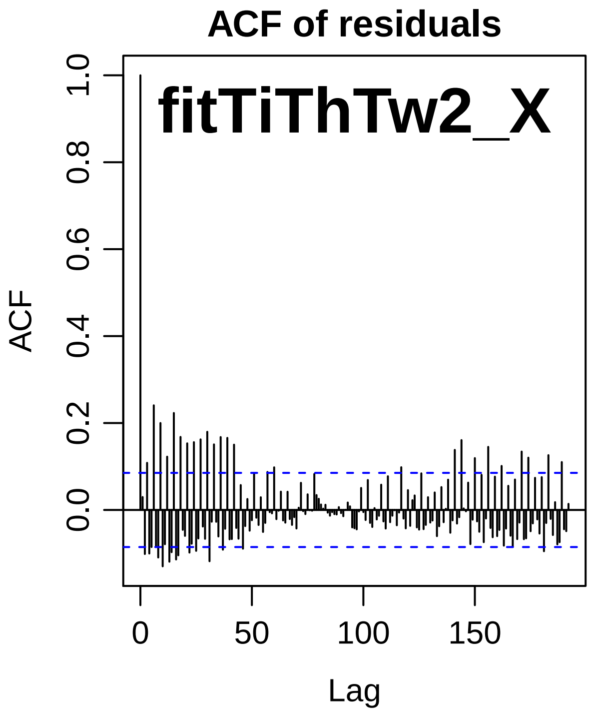


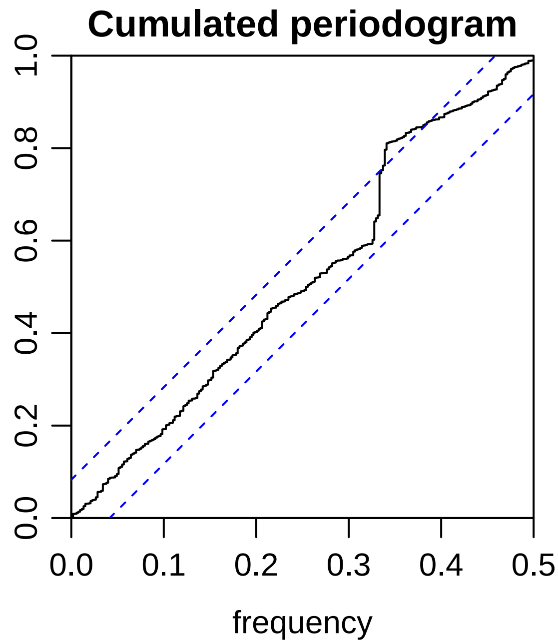


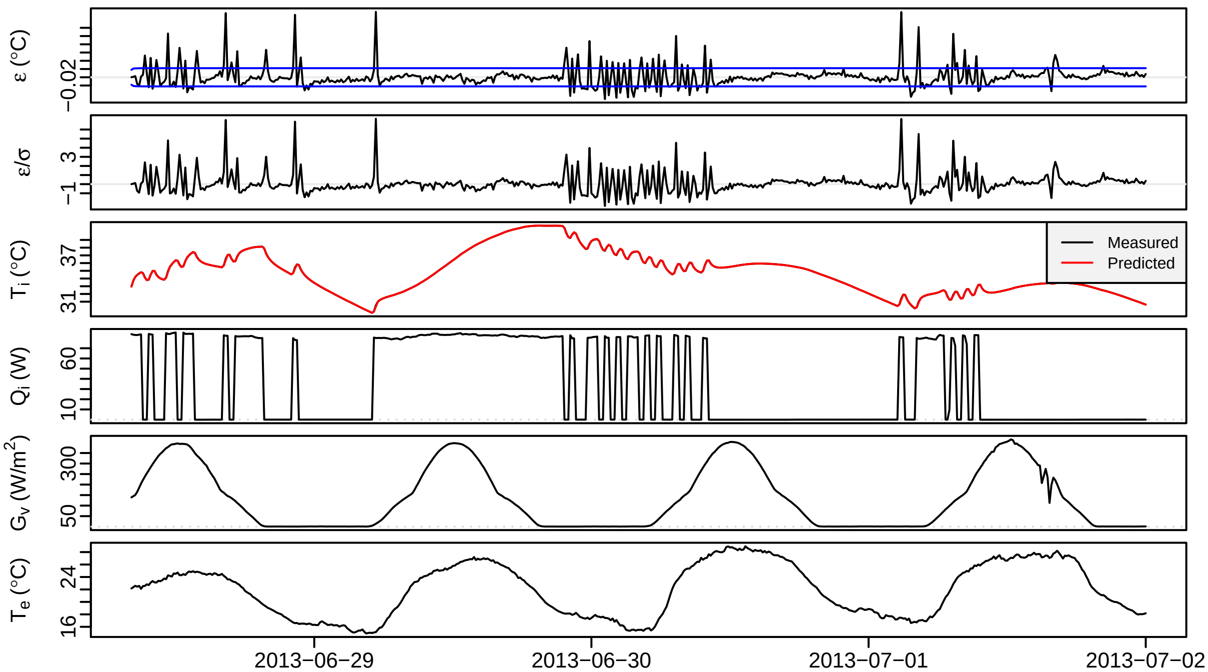
- Fit the model fitTiTw2_QinTw_X and assess—only from the loglikelihood—if it has improved compared to the previous model.
  *There was an improvement in terms of the Loglikelihood from **787.37** to **1150.74** with two states in the wall and heating goes term in wall.*

- Eyeball the residual plots (ACF, cummulated periodogram, and the residuals as function of time). Why does it not seem reasonable to conclude that the model has improved?

Because there is still autocorrelation in the ACF graph, the frequency is not evenly distributed n the peridogram and the residuals are not White Noise, still change their varince corresponding to the times when the heating system is turned on.







- What does it mean in physical terms when we include an additional state for the heating system, *Th*, as done in the model sdeTiThTw2?

For the indoor temperature stare, the sdeTiThTw2 model includes a term $(1/(Ci \cdot Rih) \cdot (Th - Ti))$ *compared to just* $(1/Ci \cdot Qi)$ *included in the sdeTiTw2 model. Then the state for the heating system* $(1/(Ch \cdot Rih) \cdot (Ti - Th) + 1/Ch \cdot Qi)$ *contains the oposite term to balance it out. In physical terms this new state for the heating system accounts for the energy of the heating system that is going to the indoor temperature and tries to account for that fluctuations.*

- Look closely at the plots for the fit fitTiThTw2_X. What seems to drive the large fluctuations in the residuals, and what can the reason be?

The large fluctuations in the residuals seem to be driven by the ON and OFF fast cylcing of the heating system and that the system is not able to predict that fast enough.

References

JJ Allaire, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, and Winston Chang. *rmarkdown: Dynamic Documents for R*, 2018. URL https://CRAN.R-project.org/package=rmarkdown. R package version 1.10.

Peder Bacher and Philip Delff. *IEA Common Exercise 4: ARX, ARMAX and Grey-box Models for Thermal Performance Characterization of the Best Box.* DTU Compute, 2014.

Peder Bacher and Henrik Madsen. Experiments and data for building energy performance analysis: Financed by the danish electricity saving trust. 2010.

Peder Bacher and Henrik Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43(7):1511–1522, 2011.

Dirk Eddelbuettel. *digest: Create Compact Hash Digests of R Objects*, 2018. URL https://CRAN.R-project.org/package=digest. R package version 0.6.15.

Keith R Godfrey. Correlation methods. In *System Identification*, pages 527–534. Elsevier, 1981.

Rune Juhl. *ctsmr: CTSM for R*, 2018. R package version 0.6.17.

Henrik Madsen. *Time series analysis.* Chapman and Hall/CRC, 2007.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

Vitalie Spinu, Garrett Grolemund, and Hadley Wickham. *lubridate: Make Dealing with Dates a Little Easier*, 2018. URL https://CRAN.R-project.org/package=lubridate. R package version 1.7.4.

Hadley Wickham. *tidyverse: Easily Install and Load the 'Tidyverse'*, 2017. URL https://CRAN.R-project.org/package=tidyverse. R package version 1.2.1.

Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2018. URL https://CRAN.R-project.org/package=knitr. R package version 1.20.