Computer Exercise 3

Christian Glissov, s146996

November 28, 2018

DTU Compute - Institute for Mathematics and Computer science
Advanced Time Series Analysis 02427 November 28, 2018
Henrik Madsen & Erik Lindstrøm

Technical University of Denmark

Made in collaboration with Frederik Boe Hüttel, s151689, and Jonas Gottlieb Svendsen, s152992.

Contents

Question 1a: Simulation and discretization of diffusion processes

Using the Euler-Marayama approximation given by the two-dimensional stochastic difference equation in (3a) and (3b) from the assignment:

$$
\begin{aligned}
Y_{n+1}^1 &= Y_n^1 + \theta_3(Y_n^2 + Y_n^2 - \frac{1}{3}(Y_n^1)^3 + \theta_4)\Delta + \sigma \Delta W_{n+1}^1 \\
Y_{n+1}^2 &= Y_n^2 - \frac{1}{\theta_3}(Y_n^1 + \theta_2 Y_n^2 - \theta_1)\Delta
\end{aligned}
\tag{1}
$$

Here the Wiener process is generated by the following:

$$
\begin{aligned}
W_{t+1} &= W_t + \epsilon_i, \quad \epsilon_i \sim N(0, \Delta) \\
W_0 &= 0
\end{aligned}
\tag{2}
$$

Using the specified parameters:

Δ	θ_1	θ_2	θ_3	θ_4	Y_0^1	Y_0^2
2^{-9}	0.7	0.8	3.0	-0.34	-1.9	1.2

The realisations of the approximation in Equation 1 with the above parameter values can be seen in Figure 1:

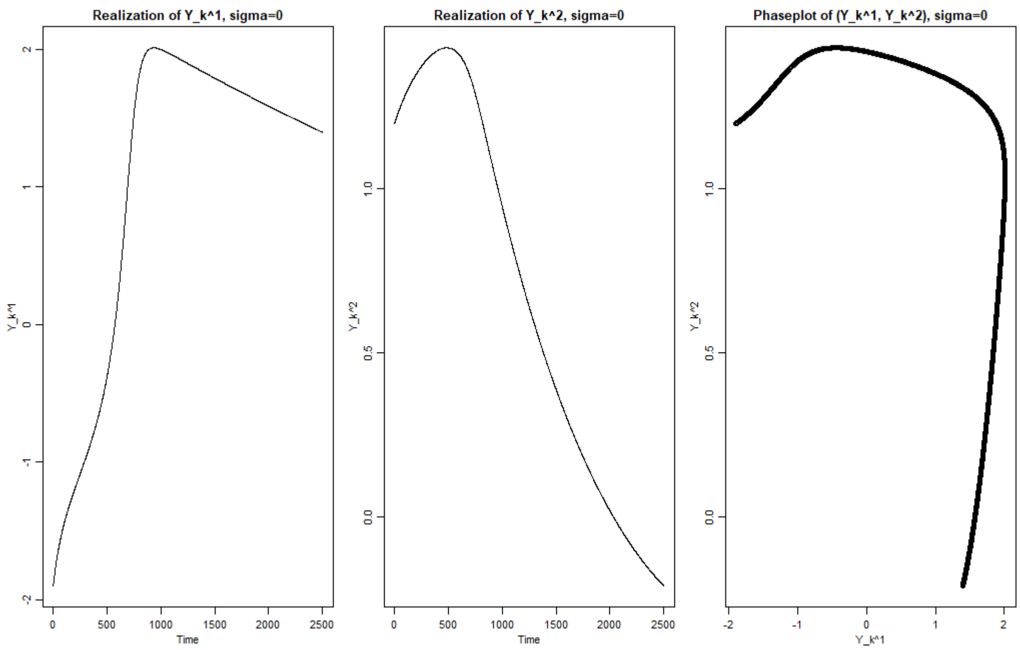

Figure 1: Realisation of Equation 1, $\sigma = 0$, and the corresponding phaseplot

One can see the nature of the negative potential and permeability of respectively (Y_k^1,Y_k^2). Due to imperfections it is necessary to add additive noise. This is done by using a Wiener process, $\Delta W_{n+1}^1 \in \mathbb{N}(0, \Delta)$. Equation 1 will now be simulated with $\sigma = 0.1, 0.2, 0.3$ and 0.4.

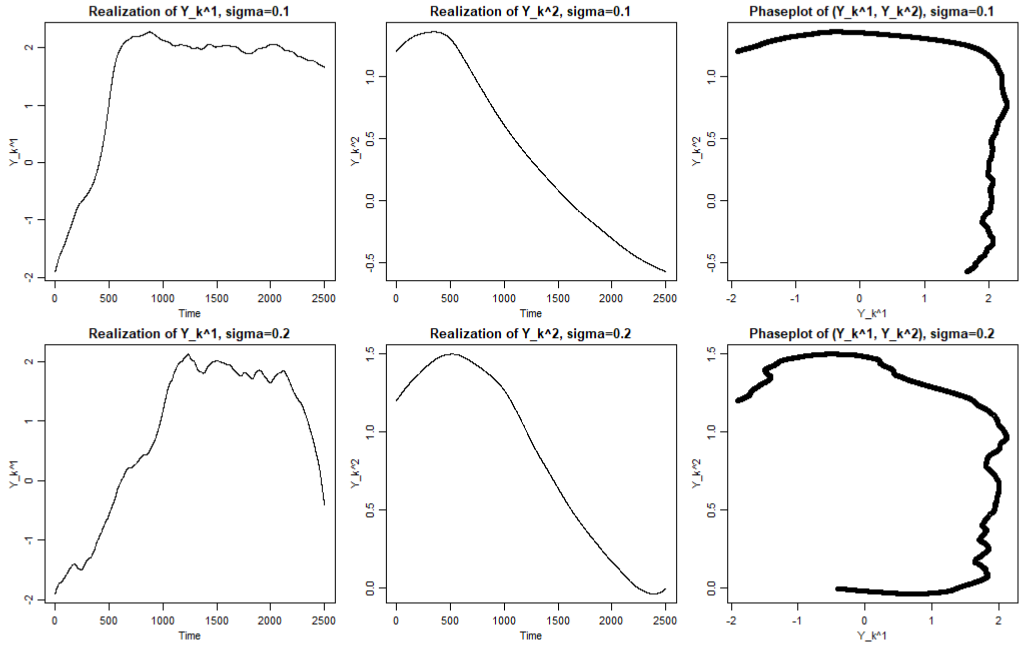

Figure 2: Realisation of Equation 1, $\sigma = (0.1, 0.2)$, and the corresponding phase-plot

Now for $\sigma = 0.3$ and 0.4

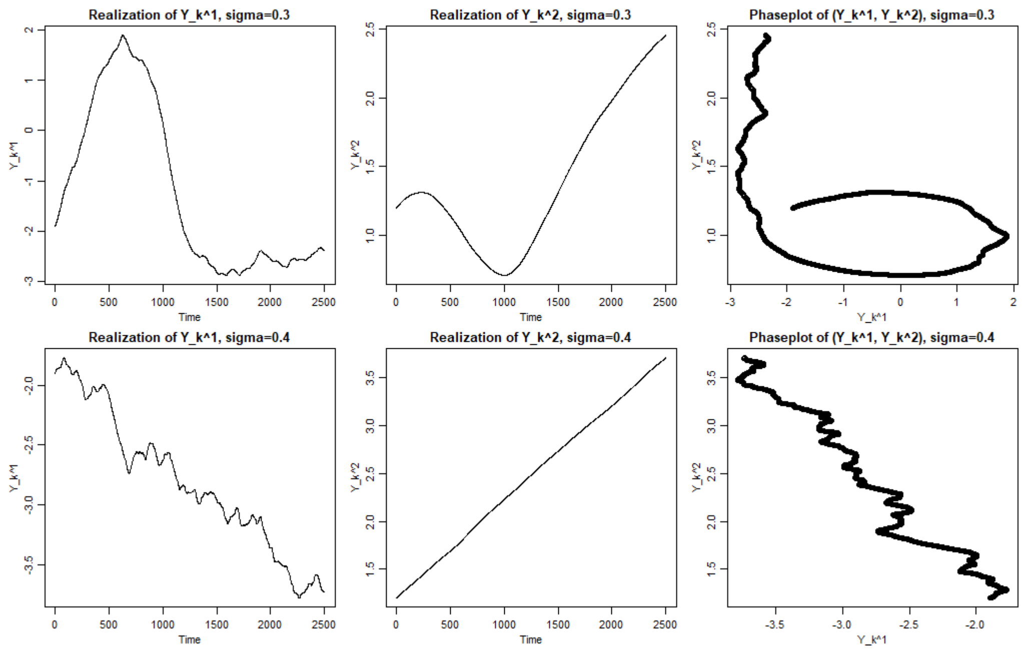

Figure 3: Realisation of Equation 1, $\sigma = (0.3, 0.4)$, and the corresponding phase-plot

One might notice that the greater σ is, the more volatile and random do the process look. This is because σ acts like a weighting of the Wiener process. This means a larger σ will make the impact of the randomness from the Wiener process more emphasised in the simulations. For the permeability (Y_k^2) it still looks fairly smooth, this is due to the very low Δ, decreasing the impact of Y_k^1, however it is still seen that when the negative potential, $Y_k^1 < 0$, the permeability increases and for $Y_k^1 > 0$ it decreases, acting like a predator-prey function.

Question 1b

In the 2D case the simulations of the process in Equation 1 was shown, however the distribution of the simulation was difficult to observe. To get more information a 3D plot is made, which can be seen in Figure 4 for $\sigma = 0.1$.

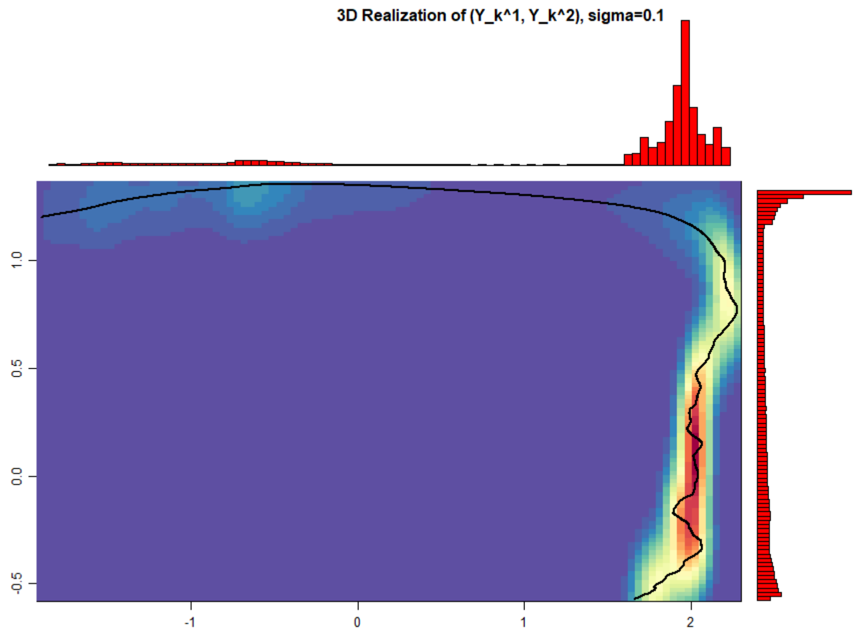

Figure 4: By adding another dimension, the density of the simulations can be observed, the warmer the colour, the higher the density, $\sigma = 0.1$.

One can see that the histograms indicates where the highest distribution of data can be found. Looking at $\sigma = (0.2, 0.3, 0.4)$:

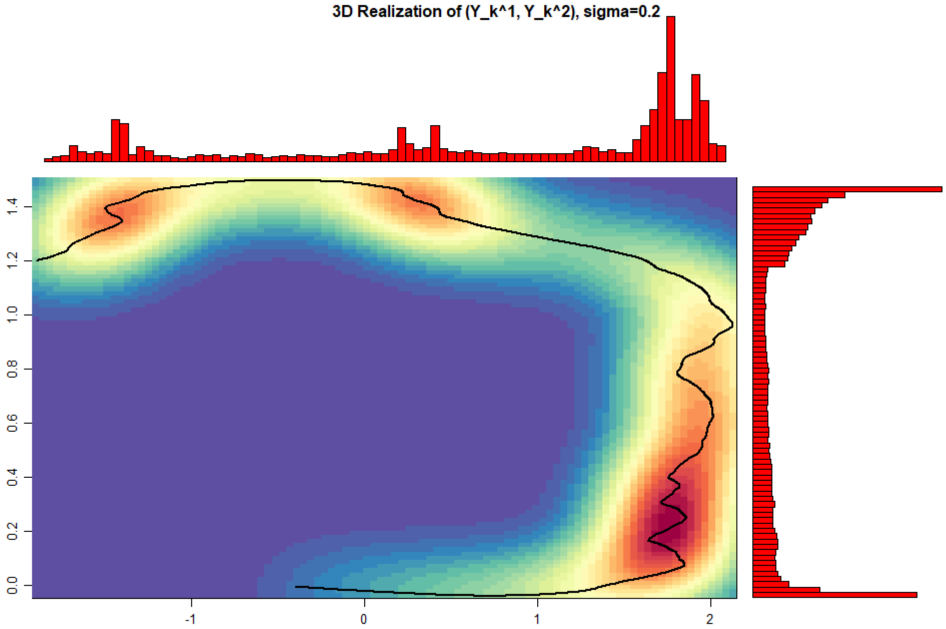

Figure 5: $\sigma = 0.2$.

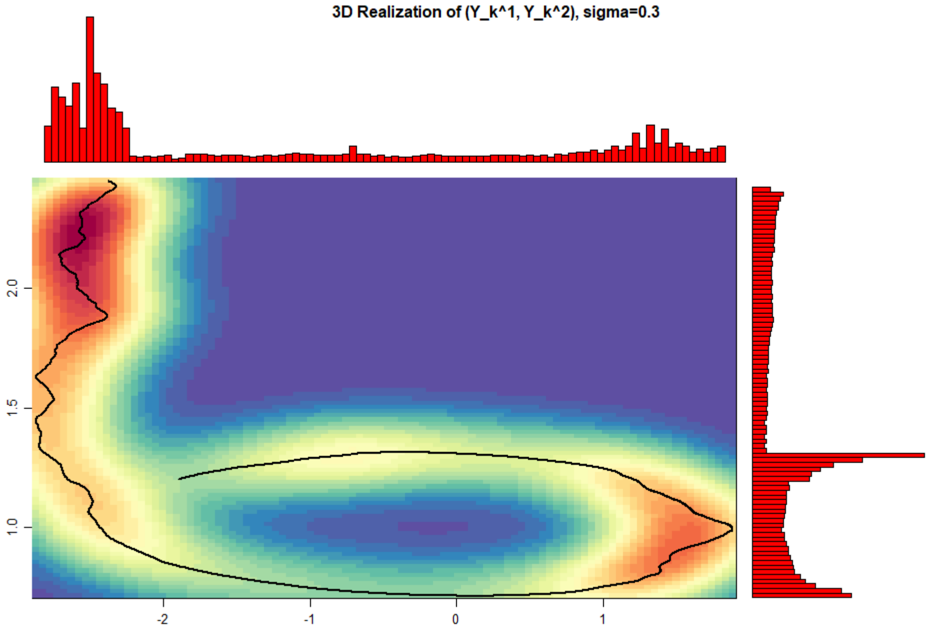

Figure 6: $\sigma = 0.3$.

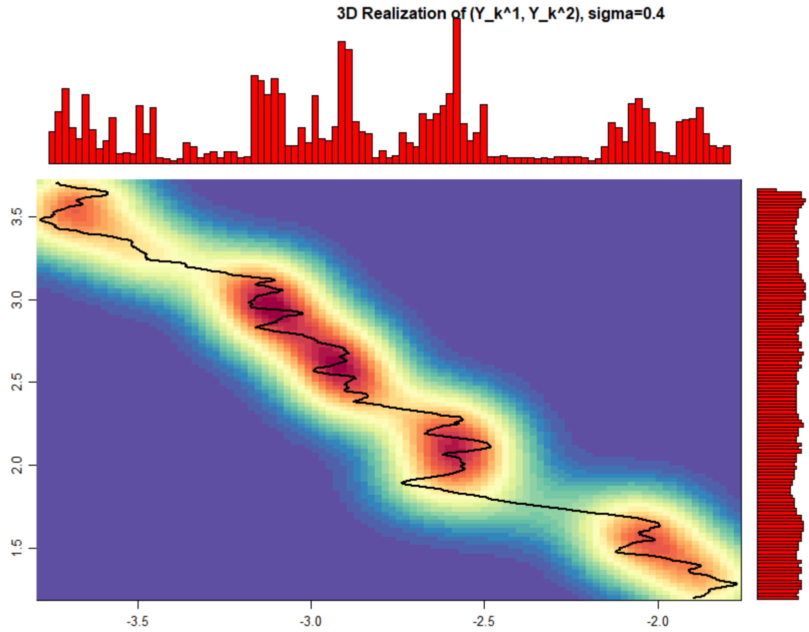

Figure 7: $\sigma = 0.4$.

From the 4 above figures it can be seen that they are quite different due to the Wiener process. Compared to the phaseplots, the 3D plot contains information about the distribution. In the phaseplot the behaviour of the simulation is observed, but it is not possible to see the distribution of the simulations or where most of the activation of the function appear. As an example in Figure 4, the phaseplot indicates the behaviour of Y_k^1 against Y_k^2, but it doesn't show the distribution of the simulations. This might also tell something about the support and form of the density function of the solution as mentioned in the assignment.

Task 2a: 2-state model of single room

In part 2 of the assignment a building will be modelled. The R-package `CTSM-R`, [3], will be used to find an adequate model. The data contains the following:

Table 1: Description of data contained in `Exercise3.RData`.

Name	Description	Type	Unit
`date`	Date and Time	string	Date, hour, min and sec
`t`	Time since start of data set	Integer	Hour
`yTi1`	Temperature 1	Continuous	C°
`yTi2`	Temperature 2	Continuous	C°
`yTi3`	Temperature 3	Continuous	C°
`yTi4`	Temperature 4	Continuous	C°
`Ta`	Ambient temperature	Continuous	C°
`Gv`	Global Horizontal Solar Radiation	Continuous	W/m^2
`Ph1`	Heating power in northern circuit	Continuous	W
`Ph2`	Heating power in southern circuit	Continuous	W

In this assignment a model of room 1 will primarily be investigated. The model is already given in the assignment as:

$$\begin{aligned} dT_i &= \frac{1}{C_i}(\frac{1}{R_{ia}}(T_a - T_i) + \frac{1}{R_{im}}(T_m - T_i) + \Phi + A_w G_v)dt + \sigma_i dw_i \\ dT_m &= \frac{1}{C_m}(\frac{1}{R_{im}}(T_i - T_m))dt + \sigma_m dw_m \\ yT_i &= T_i + e_1 \end{aligned} \tag{3}$$

Where i indicates the interior, i.e. indoor air, a indicates the ambient, i.e. outdoors and the medium is indicated by m, i.e. interior walls, furniture etc. T_i is therefore temperature of the interior and T_m is the thermal medium. C_x indicates the heat capacities of x. The R_{xk}'s indicate the resistance between x and k.

The approach

As an attempt to find an optimal model the following approach from [4] is used:

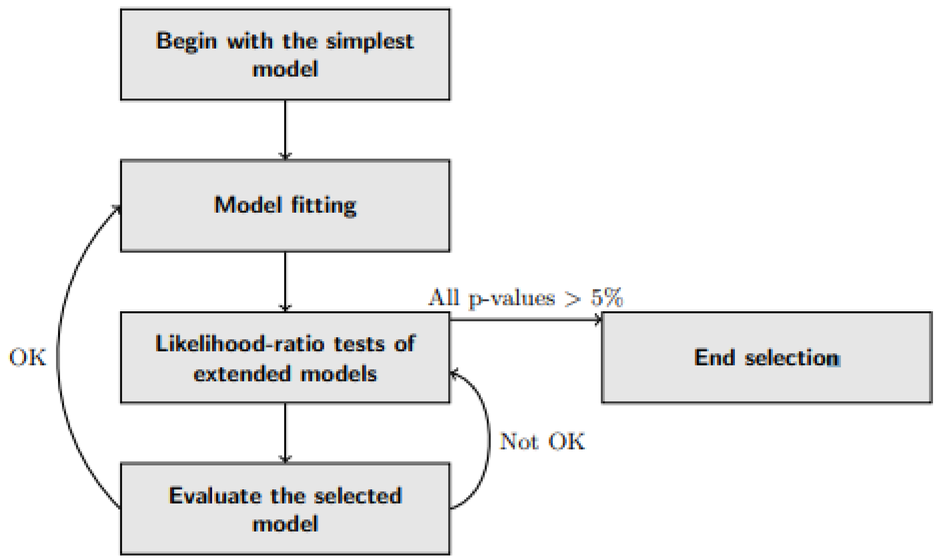

Figure 8: Model selection procedure.

From Figure 8 it is seen that first the simplest model is chosen. This is the model given in Equation 3. The model is fitted and likelihood-ratio tests will be performed on model extensions, if relevant. Where the end selection will be satisfied by removing non-significant parameters based on the p-values, using a forward or backward selection. In each step the model needs to be evaluated, this procedure will be done by following the criteria given in the CTSM-R user-guide, [5]:

- The assumption of white noise residuals should be inferred upon using the auto-correlation function (ACF) and the cumulated periodogram (CP).

- Plots of the inputs, outputs, and residuals. To show effects being improperly described by the model.
- Evaluation of the estimated physical parameters. The results should be consistent among different models and should make sense, by being consistent with reality. An initial ambient temperature $T_{a0} = 100C^{\circ}$ wouldn't make sense.

If satisfied, the final model will be used for model fitting.

Modifying the model

Using the CTSM package to estimate the initial model in Equation 3, it is clear that it's a poor model. The noise can't be considered white, based on the periodograms and ACF Figure 9

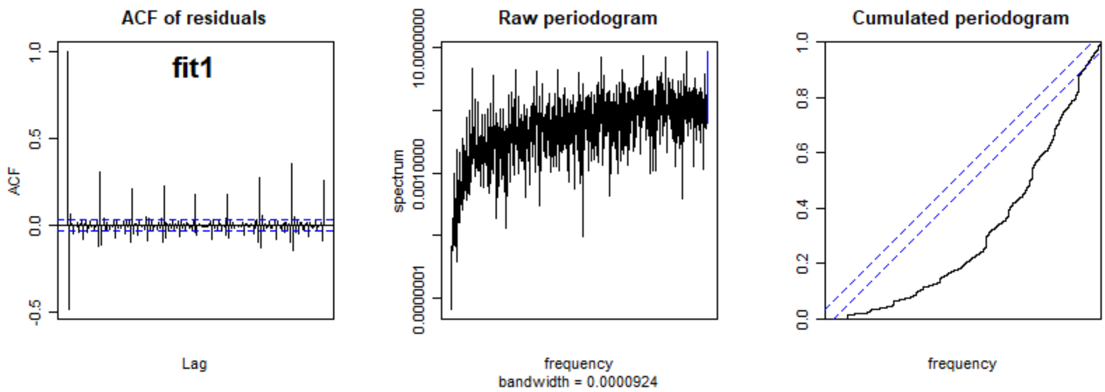

Figure 9: ACF and periodograms of TiTm model.

The log-likelihood of the first model is found to be $ll_{TiTm} = -1100$. As an attempt to improve the model, the effective window area, which is assumed constant in the initial model, will be made dynamic by using splines. This is done since it is clearly observed in Figure 10 that the solar radiation is not constant.

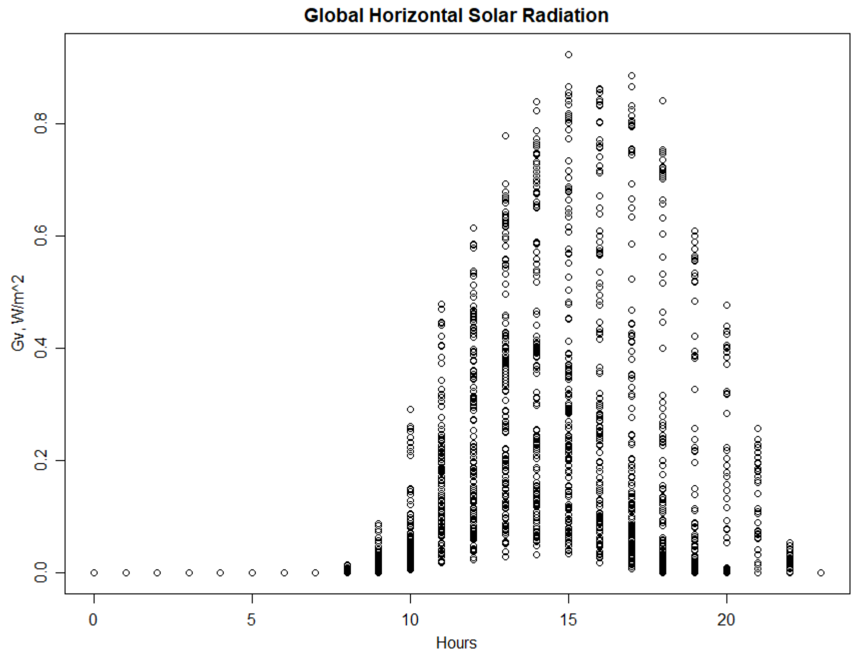

Figure 10: The global horizontal solar radiation (Gv) varying over time.

Now including the splines, A_w can be written as follows:

$$A_w = \sum_{k=1}^{N} a_k bs_k(t)$$

With bs_k being the spline and a_k a parameter. The splines used are qubic polynomial splines and 5 splines will be used. Doing this gives a log-likehood of $ll_{STiTm} = 70$, an improvement compared to the initial model. By looking at the ACF and periodograms, Figure 11, an improvement can clearly be seen.

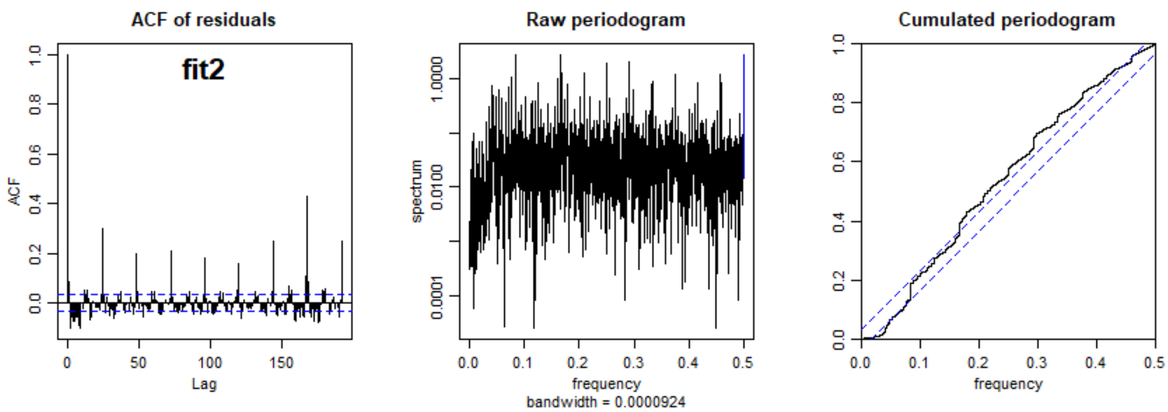

Figure 11: ACF and periodograms of STiTm model.

However there are still improvements to be made.

Task 2b: Improving the single-room model

As was seen from Figure 11 improvements can still be made. The first thing to improve will be the splines.

Splines

Adding more splines the likelihood seems to increase up until 7 splines. As is seen in Table 1.

$N_{Splines}$	5 Splines	6 Splines	7 Splines	8 Splines
log-likelihood	70	108	127	122

Table 1: Log-likelihood of STiTm with different splines.

Doing a likelihood ratio test to compare the models:

Test	Models	df	$\lambda(y)$	p-value
1	5 splines vs 6 splines	1	76	$< 10^{-16}$
2	6 splines vs 7 splines	1	38	$5.6 \cdot 10^{-}9$

Table 2: Likelihood ratio test of models with different splines.

Adding a higher degree to the splines didn't seem to benefit the model neither did adding more than 7 splines or less than 5. 7 splines therefore seems to be an optimal amount of splines to include in the model.

Including states

The model STiTm is a very simple model. It would be interesting to include other relevant states. Such as to describe the temperature of the heaters, which should have a large effect on the indoor temperature. The same applies to the temperature of the envelope of the building. First the state of the temperature of the heater is defined:

$$dT_h = \frac{1}{C_h}(\frac{1}{R_{ih}}(T_i - T_h) + \Phi)dt + \sigma_h dw_h \tag{4}$$

The temperature of the envelope can then be defined as:

$$dT_e = \frac{1}{C_e}(\frac{1}{R_{ie}}(T_i - T_e) + \frac{1}{R_{ea}}(T_a - T_e))dt + \sigma_e dw_e \tag{5}$$

The above states are taken from [6]. The likelihood for the new models including the states is:

Model	STiTmTh	STiTmTe	STiTeTh
Log-Likelihood	271	288	286
N_{param}	20	21	21

Table 3: Likelihood for each added state.

Both states improve the log-likelihood, therefore an idea might be to combine them. Doing this it is seen that the likelihood doesn't change a lot and actually got a bit worse, $ll_{STiTmTeTh} = 287.77$, the residuals didn't seem to improve either. A likelihood ratio test of the STiTmTe and STiTeTh model gives a p-value of $1.9 \cdot 10^{-9}$. Therefore we choose the $STiTmTe$ model. A model containing a state for the temperature of the heater therefore doesn't seem necessary in this case. The ACF/Periodogram plot of $STiTmTe$ can be seen in Figure 12 and the relevant information about residuals, input and outputs can be seen in Figure 13

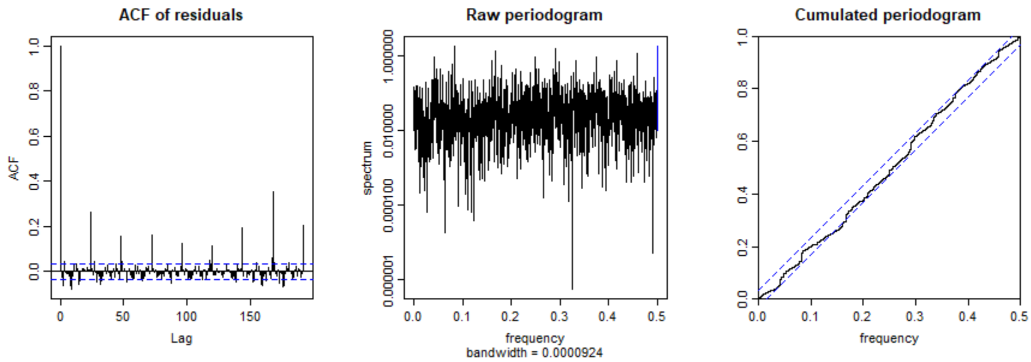

Figure 12: ACF and periodograms of STiTmTe model.

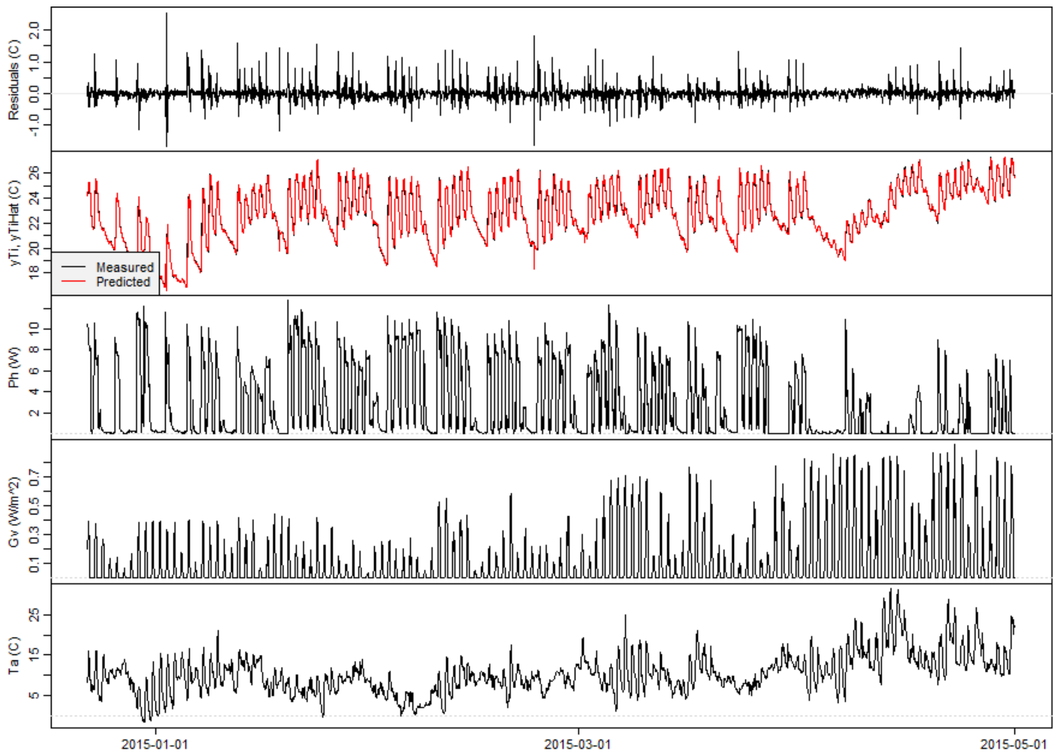

Figure 13: Inputs, outputs and residuals of STiTmTe model.

It's clear that the residuals seem to be more like white noise, based on the periodograms. However there are still significant lags in the ACF, especially for for the daily variation creates seasonality in the ACF plot, which can also be seen in the residuals.

```
Coefficients:
              Estimate           Std. Error              t value            Pr(>|t|)             dF/dPar dPen/dPar
Ti0    23.58403411682561     0.13127353861904   179.65565920537401    0.00000000000000     0.00019874253197    0.0004
Tm0    32.25950513436126    10.13856472036260     3.18186114347826    0.00147796703551    -0.00004694961591    0.0022
Te0    20.37896112344749     0.51135661128849    39.85273813532530    0.00000000000000    -0.00042963111016    0.0002
a1    116.02876605217493     7.65365352367009    15.15991881437255    0.00000000000000     0.00029444773595    0.0028
a2      5.45708149132117     2.10782720717915     2.58896055271259    0.00967183820346    -0.00002914730644    0.0001
a3      8.05895050521839     1.28918057752702     6.25121929829065    0.00000000046527    -0.00007970464448    0.0001
a4      1.62446054094161     0.99607644616247     1.63085930522709    0.10302191238724     0.00002554791080    0.0000
a5      4.99403967475103     1.25989600444745     3.96385071237784    0.00007545460539     0.00002240181785    0.0000
a6      0.01299719306050     0.06084847484006     0.21359932347787    0.83087368200591    -0.00001012907802   -0.0014
a7     14.16298981008809     5.16241626850008     2.74348077982543    0.00611437298740    -0.00004396531946    0.0008
Ce     78.58497918933935     9.17622698537726     8.56397507543875    0.00000000000000    -0.00015943334182    0.0000
Ci      8.91418397263486     0.24431039889201    36.48712462941521    0.00000000000000     0.00463533706447    0.0000
Cm      0.00029280312343     0.00064778768816     0.45200476758386    0.65129731411444    -0.00317605791359    0.0000
e11    -5.16207952347791     0.07518624924352   -68.65722888714097    0.00000000000000     0.00296482359459    0.0000
p11   -13.88030334623518     4.95822826054657    -2.79944823369328    0.00515091686495     0.00011708986548    0.0001
p22     9.45353503642246     2.20131915102830     4.29448634561075    0.00001806226362    -0.05864817784338    0.0316
p44    -2.59078336876886     0.16460398386154   -15.73949371084574    0.00000000000000    -0.00003808697272    0.0000
Rea     5.18513979990759     0.96328873852995     5.38274724131056    0.00000007902529     0.00034135901970    0.0000
Ria     7.95897592762440     1.65221747387412     4.81714789576834    0.00000152844499     0.00017925307434    0.0000
Rie     0.53330481979993     0.01788707973771    29.81508595143854    0.00000000000000    -0.00103570504156    0.0000
Rim  4809.77367781896828 10645.75604488735917     0.45180198170414    0.65144339227794     0.00306696965736    0.0000

Correlation of coefficients:
    Ti0   Tm0   Te0   a1    a2    a3    a4    a5    a6    a7    Ce    Ci    Cm    e11   p11   p22   p44   Rea   Ria   Rie  
Tm0  0.06                                                                                                                 
Te0  0.17  0.07                                                                                                           
a1   0.02  0.13  0.07                                                                                                     
a2   0.04  0.11 -0.02 -0.19                                                                                               
a3   0.01 -0.04 -0.02  0.25 -0.53                                                                                         
a4   0.01  0.09 -0.05 -0.11  0.44 -0.58                                                                                   
a5   0.00  0.05 -0.06  0.04 -0.17  0.51 -0.49                                                                             
a6  -0.02 -0.80  0.07 -0.06 -0.02  0.01 -0.06 -0.07                                                                       
a7   0.02  0.33  0.04 -0.13  0.06 -0.14  0.14  0.01 -0.26                                                                 
Ce   0.03 -0.03  0.10  0.05 -0.07  0.01 -0.01  0.01 -0.03 -0.03                                                           
Ci   0.00  0.04  0.03  0.47  0.06  0.12  0.01 -0.02  0.00 -0.11  0.21                                                     
Cm  -0.08 -0.70 -0.26  0.17 -0.26  0.47 -0.25  0.25  0.69 -0.62 -0.14  0.09                                               
e11  0.02  0.05 -0.01  0.00  0.13 -0.07  0.10 -0.01 -0.05  0.03 -0.05  0.04 -0.05                                         
p11  0.08  0.70  0.26 -0.16  0.26 -0.46  0.25 -0.25 -0.69  0.61  0.18 -0.08 -1.00  0.05                                   
p22  0.08  0.70  0.26 -0.16  0.26 -0.47  0.25 -0.25 -0.69  0.62  0.13 -0.08 -1.00  0.06  1.00                             
p44  0.01  0.08  0.07  0.02  0.02 -0.04  0.01 -0.03 -0.09 -0.03  0.36  0.02 -0.08 -0.09  0.16  0.07                       
Rea -0.04 -0.03  0.26  0.06 -0.08 -0.22 -0.20 -0.29  0.05 -0.04 -0.16 -0.02 -0.14 -0.03  0.14  0.14 -0.10                 
Ria  0.04  0.03 -0.31 -0.11  0.05  0.21  0.18  0.28 -0.07  0.05  0.09 -0.03  0.13  0.04 -0.14 -0.13  0.03 -0.94           
Rie  0.00 -0.10 -0.05 -0.03 -0.17 -0.17 -0.23 -0.20  0.07 -0.05  0.03 -0.04 -0.05 -0.13  0.03  0.05 -0.05  0.29 -0.26     
Rim  0.08  0.70  0.26 -0.17  0.26 -0.47  0.26 -0.25 -0.69  0.62  0.13 -0.09 -1.00  0.07  1.00  1.00  0.07  0.14 -0.13  0.04
```

Figure 14: *R* summary of STiTmTe model.

Looking at the parameters, then it can be seen that a few parameters are close to the boundary based on the `dF/dPar` to `dPen/dPar` ratio, however it doesn't look too bad. The correlation matrix seems to be fairly bad at places, with quite a few of the resistance parameters correlated. This might be due to unnecessary parameters or bad estimations. Most of the parameters look fairly realistic, the temperatures seem reasonable. The resistance of *Rim* seems to be rather high, but it doesn't seem unrealistic that the resistance between the interior and the interior thermal medium can be high. It is also noticed that *Rim* is the parameter being rather highly correlated with other parameters and being non-significant according to the t-test.

Reducing insignificant parameters

Before evaluating the model, insignificant parameters will be removed. I was able to remove one of the parameters, a_6, removing more made the model not able to converge properly. Due to time limits i was unable to complete this stage of the modelling. A possible solution to this problem, might have been to tweak the model parameters or identify a superior model. It would have been interesting to see if removing certain parameters would remove the highly correlated parameters in the correlation matrix.

Seasonality

Looking at the residuals it's clear that there are still seasonality dependent lags. An idea to further improve the model, might be to add a term which explains the seasonality. The seasonality is due to a daily variation in the data. It might be because of the daily automation of the heaters. When doing this one should also take into account the physical interpretation. It might not be necessary to model the seasonality, since one is not interested in explaining this automation, such as when doing control, but rather the heating dynamics of the building. We did not succeed in removing the seasonality.

Single-room model: Conclusion

The full model, STiTmTe looks like the following:

$$\begin{aligned}
dT_i &= \frac{1}{C_i}\left(\frac{1}{R_{ia}}(T_a - T_i) + \frac{1}{R_{ie}} * (T_e - T_i) + \frac{1}{R_{im}}(T_m - T_i) + \Phi + \sum_{i=1}^{7}(a_i bs_i)G_v\right)dt + \sigma_i dw_i \\
dT_m &= \frac{1}{C_m}\left(\frac{1}{R_{im}}(T_i - T_m)\right)dt + \sigma_m dw_m \\
dT_e &= \frac{1}{C_e}\left(\frac{1}{R_{ie}}(T_i - T_e) + \frac{1}{R_{ea}}(T_a - T_e)\right)dt + \sigma_e dw_e \\
yT_i &= T_i + e_1
\end{aligned} \tag{6}$$

The model is an improvement compared to the simple TiTm model, but far from optimal. It does adhere to some of the desired properties, such as white noise based on the cumulative periodogram and also the fit seem to be decent, except at peaks, due to daily seasonality or other factors not being modelled. A concern is also the highly correlated resistances. There are definitely room for improvement.

Task 3b: Making a multi-room model

In this exercise, the single room model will be extended to a multi-room model. Modelling all of the rooms. One of the benefits of `CTSM-R` is that it's fairly simple to model a fairly complex mathematical problem like this. However one of the drawbacks is that due to a vast amount of parameters, the optimization will require heavy computation. Therefore some assumptions need to be made.

- Splines: Due to a lack of window space and the size of room 2 and 3, it is assumed that Gv is fairly constant and splines will therefore not be used for these two rooms. Furthermore the splines for room 1 and 4 are the same.
- Resistance: The resistance in each room is assumed to be identical.
- Capacity: The same goes for the capacity, it's assumed that material, air etc. is identical across the rooms.
- Ambient temperature: It's assumed Ta is the same for all the rooms.

- Noise of Wiener process: The noise terms σ_{xk} is assumed to be the same across all states, except for the indoor temperature state Ti.
- Only 500 observations will be used. Larger data sets either took too long to run or simply didn't converge.

The following model used is: Room 1:

$$\begin{aligned} dT_{i1} &= \frac{1}{C_i}(\frac{1}{R_{ia}}(T_a - T_{i1}) + \frac{1}{R_{im}}(T_{m1} - T_{i1}) + \Phi_1 + p \cdot \sum_{k=1}^{7}(a_k bs_k)G_v)dt + \sigma_{i1}dw_{i1} \\ dT_{m1} &= \frac{1}{C_m}(\frac{1}{R_{im}}(T_{i1} - T_{m1}))dt + \sigma_m dw_{m1} \\ yT_{i1} &= T_{i1} + e_1 \end{aligned} \tag{7}$$

Room 2:

$$\begin{aligned} dT_{i2} &= \frac{1}{C_i}(\frac{1}{R_{ia}}(T_a - T_{i2}) + \frac{1}{R_{im}}(T_{m2} - T_{i2}) + \Phi_1 + A_s G_v)dt + \sigma_{i2}dw_{i2} \\ dT_{m2} &= \frac{1}{C_m}(\frac{1}{R_{im}}(T_{i2} - T_{m2}))dt + \sigma_m dw_{m2} \\ yT_{i2} &= T_{i2} + e_2 \end{aligned} \tag{8}$$

Room 3:

$$\begin{aligned} dT_{i3} &= \frac{1}{C_i}(\frac{1}{R_{ia}}(T_a - T_{i3}) + \frac{1}{R_{im}}(T_{m3} - T_{i3}) + \Phi_2 + A_s G_v)dt + \sigma_{i3}dw_{i3} \\ dT_{m3} &= \frac{1}{C_m}(\frac{1}{R_{im}}(T_{i3} - T_{m3}))dt + \sigma_m dw_{m3} \\ yT_{i3} &= T_{i3} + e_3 \end{aligned} \tag{9}$$

Room 4:

$$\begin{aligned} dT_{i4} &= \frac{1}{C_i}(\frac{1}{R_{ia}}(T_a - T_{i4}) + \frac{1}{R_{ie}}(T_e - T_{i4}) + \Phi_2 + \sum_{i=1}^{7}(a_k bs_k)G_v)dt + \sigma_{i4}dw_{i4} \\ dT_{m4} &= \frac{1}{C_m}(\frac{1}{R_{im}}(T_{i4} - T_{m4}))dt + \sigma_m dw_{m4} \\ yT_{i4} &= T_{i4} + e_4 \end{aligned} \tag{10}$$

The model achieves a log-likelihood of 663 with 28 parameters. Looking at the ACF in Figure 15 it is clear that there are still some information left to be modelled. The model is far from perfect, especially the lags at lag 1 in all of the figures seem to be better explained. Adding more states, while still keeping the model simple enough to converge might do the trick.

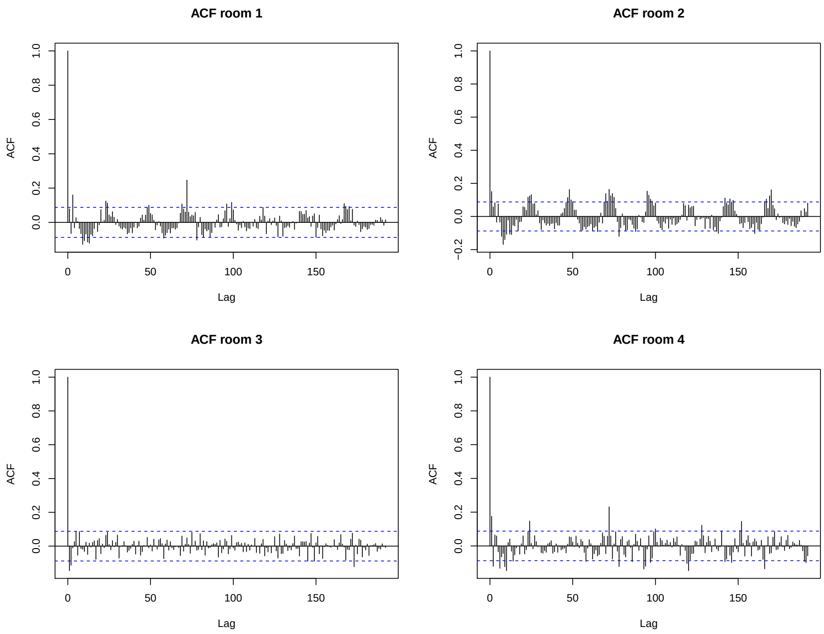

Figure 15: ACF of multi-model.

The cumulative periodograms in Figure 16 also tell that not quite all of the residuals seem to be white noise, especially for room 4.

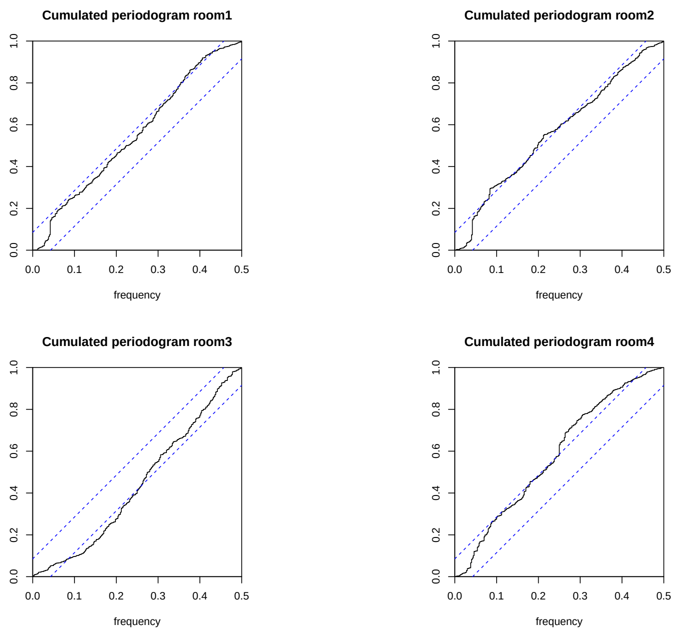

Figure 16: Cumulative Periodogram of multi-model

In the spectrum Figure 17 it seems like there is a slight trend to be observed.

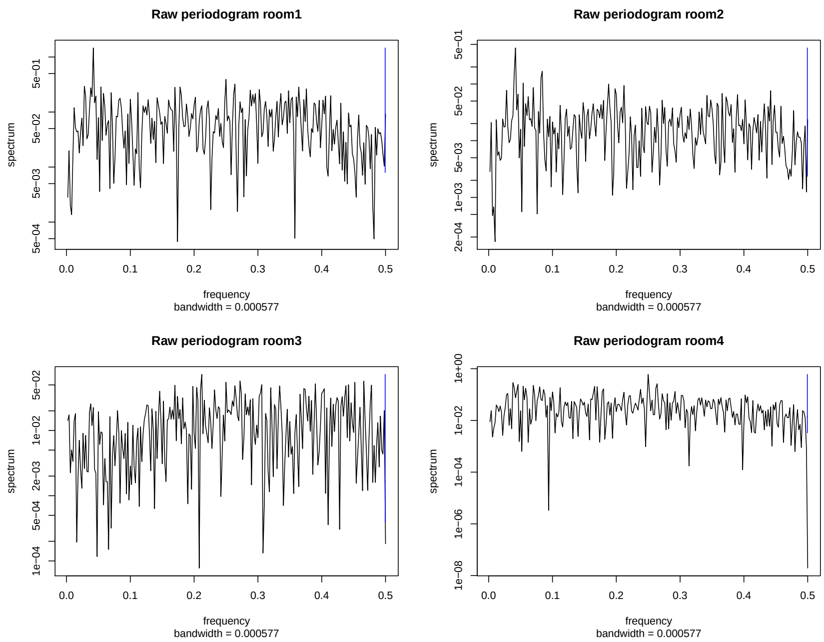

Figure 17: Spectogram of multi-model

Finally the fit Figure 18 seems to be good, however the model still overestimates the peaks and valleys.

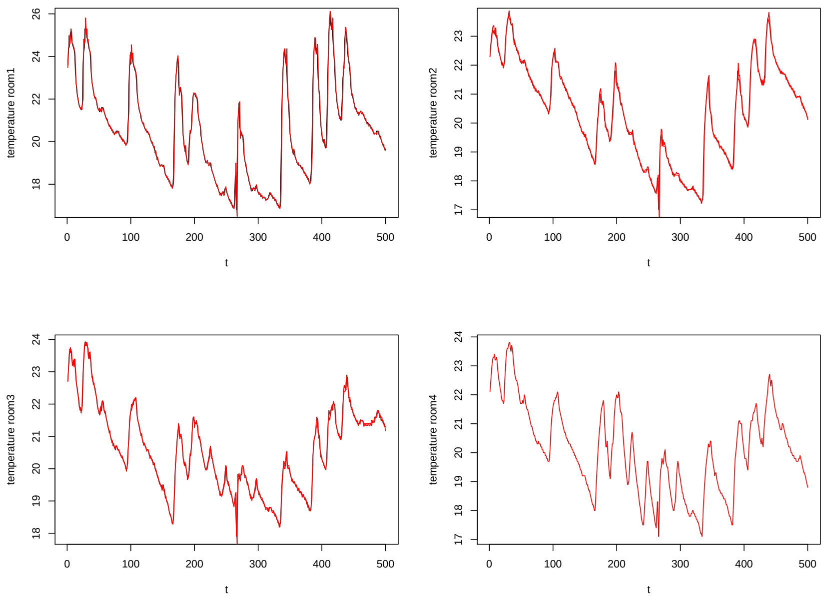

Figure 18: Prediction of multi-model

Clearly this is not a very adequate model. More states and/or modifications to the current states are needed to properly explain the heating dynamics of each room in the building. This can be tricky due to too heavy computation. Given time constraints, a lot of other modelling attempts have been left out of this report. A better understanding of CTSM-R have been achieved and given more time, a more adequate model could surely have been found.

References

[1] Henrik Madsen, *Time Series Analysis*, Chapman & Hall/CRC, 2008.

[2] Henrik Madsen & Jan Holst, *Modelling Non-Linear and Non-Stationary Time Series*, IMM, 2006, December.

[3] ctsmr: CTSM for R, Rune Juhl, 2018, R package version 0.6.17

[4] Identifying suitable models for the heat dynamics of buildings, Energy and Buildings, July 2011, Peder Bacher & Henrik Madsen. Informatics and Mathematical Modelling, Richard Pedersens Plads, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark

[5] CTSM-R User Guide, Rune Juhl, Niels Rode Kristensen, Peder Bacher, Jan Kloppenborg, Henrik Madsen, April 3, 2013

[6] Grey-box modeling of the heat dynamics of a building with CTSM-R, Rune Juhl, Niels Rode Kristensen, Peder Bacher, Jan Kloppenborg, Henrik Madsen, September 18, 2017, ctsm.info.