Part 2: Modelling a building using CTSM-R

2a: 2-state model of a single room

We will estimate the following model where the impact of the measured radiation G_v is scaled relative the sun's angle through the window. Since we do not have access to this information, a non-parametric fit will be done using B-splines.

$$
\begin{aligned}
dT_i &= \frac{1}{C_i}\left(\frac{1}{R_{ia}}\left(T_a - T_i\right) + \frac{1}{R_{im}}\left(T_m - T_i\right) + \Phi + \left(\sum_{k=1}^{N} a_k bs_k(t)\right) G_v\right) dt + \sigma_1 dw_1 \\
dT_m &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_i - T_m\right)\right) dt + \sigma_2 dw_2. \\
yT_i &= T_i + e_1,
\end{aligned}
$$

```r
#install.packages("ctsmr", repo = "http://ctsm.info/repo/dev")
#install.packages("pkgbuild")

# For git pushing
## git push https://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@github.com/davidripsen/3-Assignment.git

library(ctsmr)
library(splines)
source("CompEx3_E18/sdeTiTm.R")
# Load data
if (Sys.info()[7] == "davidipsen")
  {path <- "~/Documents/DTU/3. Semester (MSc)/Advanced Time Series/Assignments/3-Assignment/CompEx3_E18/
} else {path <- "CompEx3_E18/"}
load(paste0(path,"Exercise3.RData"))
#AllDat

########## Initial model ############
  fit1 <- sdeTiTm(AllDat,AllDat$yTi1,AllDat$Ph1) # Original model

  summary(fit1,extended=TRUE)
  fit1$loglik

  Hour <- as.numeric(strftime(AllDat$date, format="%H"))

  Pred <- predict(fit1)
  #plot(Pred[[1]]$state$pred$Ti - AllDat$yTi1 ~ Hour)

  # Fit only splines for radiation hours
  #plot(AllDat$Gv ~ Hour) #


  idx <- (Hour>8 & Hour < 23) # It is impossible to fit a window area for the hours without any sun, so
  bs = bs(Hour[idx],df=5,intercept=TRUE) # Dvs. 4 knots / 5 basis splines

  # What does the splines look like?
  #plot(bs[14:27,1],type='l')
  #lines(bs[ 14:27,2])
  #lines(bs[ 14:27,3])
  #lines(bs[ 14:27,4])
```

```
  #lines(bs[ 14:27,5])

  bs1 <- bs2 <- bs3 <- bs4 <- bs5 <- bs6 <- numeric(dim(AllDat)[1])

  bs1[idx] = bs[,1]
  bs2[idx] = bs[,2]
  bs3[idx] = bs[,3]
  bs4[idx] = bs[,4]
  bs5[idx] = bs[,5]

  AllDat$bs1 = bs1
  AllDat$bs2 = bs2
  AllDat$bs3 = bs3
  AllDat$bs4 = bs4
  AllDat$bs5 = bs5


### IMPLEMENT THE MENTIONED MODEL ###
source(paste0(path,"sdeTiTmAv.R"))
fit2 <- sdeTiTmAv(AllDat,AllDat$yTi1,AllDat$Ph1)
```

Let's compare the two models

```
sprintf('Model 1: logL = %f', fit1$loglik)
sprintf('Model 2: logL = %f', fit2$loglik)
```

I.e. we see a very large improvement in likelihood (for only 4 extra parameters).

```
summary(fit2, extended=T)
```

We see that all parameters are significant, except for a2. For completeness, a2 is kept in the model. Let's visualize the spline-fit

```
plot(9:22, bs[14:27,1]*fit2$xm[3]+bs[14:27,2]*fit2$xm[4]+bs[14:27,3]*fit2$xm[5]+bs[14:27,4]*fit2$xm[6]+
```

Remember, that the above fit is not the actual input radiation to the room, but it is a *weighting* of the actual radiation. Apparently, the model would like to have more emphasis on the radiation when the sun is low.

2b: Improving the single-room model

In this part, I will try to expand the model. First, note how the effect of the temperature in room 3 and 4 affects only room 1 through the medium of room 2. In other words: T_1 is *conditionally independent* of T_3 and T_4 when T_2 is known.

Initially, I added the temperature of the neighboring room (T_2) as an additional state and then had it impact dT_1. However, I realized that the temperature of room 2 only affects room 1 through *the wall* as a medium, i.e. the thermal mass of room 1. And since we have direct measurements of $T2$ and only focus on modelling the single-room ($T1$), I add it as input to the model with the term:

$$\frac{1}{R_{2,1}}(T_2 - T_1)$$

thereby letting the change in thermal mass of room 1 (dT_m) being proportional to the difference in temperature between the two rooms.

The model is then

$$
\begin{aligned}
dT_1 &= \frac{1}{C_1}\left(\frac{1}{R_{ia}}\left(T_a - T_1\right) + \frac{1}{R_{im}}\left(T_m - T_1\right) + \Phi + \left(\sum_{k=1}^{N} a_k b s_k(t)\right) G_v\right) dt + \sigma_1 dw_1 \\
dT_m &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_1 - T_m\right) + \frac{1}{R_{2,1}}(T_2 - T_1)\right) dt + \sigma_2 dw_2. \\
yT_1 &= T_1 + e_1,
\end{aligned}
$$

Let's estimate it

```r
set.seed(101)
source(paste0(path,"2b-sdeT1T2TmAv.R"))
fit3 <- sdeT1T2TmAv_2b(AllDat,AllDat$yTi1,AllDat$Ph1)
```

Let's have a look at the model.

```r
sprintf('Model 3: logL = %f', fit3$loglik)
sprintf('Likelihood ratio test: p-value = %f', 1-pchisq(abs(2*(fit2$loglik - fit3$loglik)),1))
summary(fit3, extended=T)
```

Again, a large improvement in likelihood is seen (at the cost of 1 extra parameter). Since the two models are nested, at likehood ratio test is performed and shows a very strong signficant difference. Again, all parameters are significant, except for the observation error which might in fact have 0 as the true value, however we do see some strong correlations in the coefficients (!).

Now expanding further on the model, I see that the radiator is placed close to the window in room 1, i.e. some of the heating might go straight out of the window. Additionally, I note that the North Heating Circuit and South Heating Circuit are piped together before entering the building, so the actual heating in room 1 is a mix of the two (however, the two are higly correlated with $\hat{\rho} = 0.91$). I.e. denote $= \Phi = (\Phi_1 + \Phi_2)/2$ and change then the Φ term in dT_1 to the term $(1-c)\Phi - c\Phi\frac{T_1 - T_a}{T_1}$, where $c \in [0, 1]$. This makes a proportion c of the heating load escape directly through the window relative to the temperature difference[1] between inside and outside[2]. Note that using this formulation, the hypothesis H0: $c = 0$ is of interest (meaning no loss of heating) and when the outdoor and indoor temperature is the same, the second term cancels (meaning no loss of heating).

The model then becomes

$$
\begin{aligned}
dT_1 &= \frac{1}{C_1}\left(\frac{1}{R_{ia}}\left(T_a - T_1\right) + \frac{1}{R_{im}}\left(T_m - T_1\right) + \left((1-c)\Phi - c\Phi(T_1 - T_a)\right) + \left(\sum_{k=1}^{N} a_k b s_k(t)\right) G_v\right) dt + \sigma_1 dw_1 \\
dT_m &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_1 - T_m\right) + \frac{1}{R_{2,1}}(T_2 - T_1)\right) dt + \sigma_2 dw_2. \\
yT_1 &= T_1 + e_1,
\end{aligned}
$$

Where $\Phi = (\Phi_1 + \Phi_2)/2$.

```r
#plot(AllDat$date, AllDat$Ta, type='l')
#plot(AllDat$date, (AllDat$Ph1 + AllDat$Ph2)/2, type='l')
#plot(AllDat$Ph1, AllDat$Ph2); cor(AllDat$Ph1, AllDat$Ph2)
source(paste0(path,"2b-2-sdeT1T2TmAv.R"))
#fit4 <- sdeT1T2TmAvfit4(AllDat,AllDat$yTi1)
```

However, I can't get his model to converge properly, even with reduced convergence criteria and increased function evaluation allowance.

[1] I would have liked it to be a ratio instead of difference, but that makes it much harder for CTSM to solve.

[2] The meaning of T_a is not explicit from the assignment. From plots I've concluded to assume it to be the outside temperature.

```r
## Evaluate model
#fit4$loglik
#summary(fit4, extended=T)
#sprintf('Likelihood ratio test from previous model: p-value = %f', 1-pchisq(-2*(fit3$loglik - fit4$log
```

For this reason, I'm satisfied with model 3. Of course there are yet many possible extensions to try out, but for this single-room model, model 3 will suffice.

I will end this section with model validation of model 3.

```r
# To assses the model fit, calculate the one-step predictions and residuals. Do residual analysis for m

D = AllDat; D$Ph <- AllDat$Ph1; D$T2 <- AllDat$yTi2; D$yT1 <- AllDat$yTi1;
preds <- predict(fit3, newdata=D)
yTi1Hat <- preds$output$pred$yT1
residuals <- AllDat$yTi1 - yTi1Hat

par(mfrow=c(1,2))
acf(residuals)
pacf(residuals)

hist(residuals, breaks=100)
qqnorm(residuals)
qqline(residuals)

plot(residuals)
cpgram(residuals)
```

There is clearly further work to be done - the residuals are far from white noise. Inspecting the autocorrelation and partial autocorrelation, it would clearly be beneficial to adopt this autoregression further, particularly in lag 24. There is a strong over-dispersion in the residuals.

3 Making a multi-room model

In this exercise, I will fit a continuous-discrete state space model to there 4 rooms. I will use extend model 3 to the multi-room case with the following considerations and simplifications (for computational feasibility/identifiability):

1. The thermal diffusion from other rooms is only present for the exact neighbouring rooms - and enters **through** the thermal mass.
2. Only room 1 and 4 are neighbours to the outside (T_a)
3. Use $\Phi = (\Phi_1 + \Phi_2)/2$
4. Use R_{iw} for the resistance of interior walls as a generalization of $R_2, 1$.
5. Assume shared parameters where relevant, e.g. assume similar walls $R_i w$ (simplification)
6. Initially, assume equal variances across rooms $\{\sigma_1^2, \sigma_2^2\}$ (simplification)
7. Use A_w instead of splines (simplification)

The model is as follows

$$
\begin{aligned}
dT_1 &= \frac{1}{C_i}\left(\frac{1}{R_{ia}}\left(T_a - T_1\right) + \frac{1}{R_{im}}\left(T_{1m} - T_1\right) + \Phi + \left(\sum_{k=1}^{N} a_k b s_k(t)\right) G_v\right) dt + \sigma_1 dw_{1,1} \\
dT_{1m} &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_1 - T_{1m}\right) + \frac{1}{R_{iw}}(T_2 - T_1)\right) dt + \sigma_2 dw_{1,2}. \\
dT_2 &= \frac{1}{C_i}\left(\frac{1}{R_{im}}\left(T_{2m} - T_2\right) + \Phi\right) dt + \sigma_1 dw_{2,1} \\
dT_{2m} &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_2 - T_{2m}\right) + \frac{1}{R_{iw}}(T_1 - T_2) + \frac{1}{R_{iw}}(T_3 - T_2)\right) dt + \sigma_2 dw_{2,2}. \\
dT_3 &= \frac{1}{C_i}\left(\frac{1}{R_{im}}\left(T_{3m} - T_3\right) + \Phi\right) dt + \sigma_1 dw_{3,1} \\
dT_{3m} &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_3 - T_{3m}\right) + \frac{1}{R_{iw}}(T_2 - T_3) + \frac{1}{R_{iw}}(T_4 - T_3)\right) dt + \sigma_2 dw_{3,2}. \\
dT_4 &= \frac{1}{C_i}\left(\frac{1}{R_{ia}}\left(T_a - T_4\right) + \frac{1}{R_{im}}\left(T_{4m} - T_4\right) + \Phi + \left(\sum_{k=1}^{N} a_k b s_k(t)\right) G_v\right) dt + \sigma_1 dw_{4,1} \\
dT_{4m} &= \frac{1}{C_m}\left(\frac{1}{R_{im}}\left(T_4 - T_{4m}\right) + \frac{1}{R_{iw}}(T_3 - T_4)\right) dt + \sigma_2 dw_{4,2}. \\
\forall_{i \in \{1,2,3,4\}} y T_i &= T_i + e_i
\end{aligned}
$$