LASERI-trajectories

Jouni Tuomisto

2020-10-07

## R Markdown

This analysis produces a model for trajectories of human body burdens of persistent organic pollutants (POPs).

### Muistiinpanoja kokouksesta 2020-10-07

Rasvamassa on tärkeämpi kuin BMI, koska niitä on useammalta vuodelta lapsen kasvaessa (80, 83, 86). Lisäksi nimenomaan rasva vaikuttaa POPien kinetiikkaan (ja POPit rasva-aineenvaihduntaan?).

Rasvan määrää on arvioitu usealla eri kaavalla rasvapoimuista. Panu ja Noora tarkastavat parhaan laskentatavan. Myös vyötärönympäryksen ja leptiinipitoisuuden saaminen kaavaan tutkitaan.

Ritter on hyvä teoreettinen malli POP-kinetiikasta. Lisäksi ruotsalaisilla on jokin deskriptiivinen malli (mikä?).

Panu ja Noora kaivavat myös alkuperäiset pituus- ja painomittaukset 80, 83, 86. Näiden perusteella voisi suunnitella oman mallin, jolla kuvataan yksilöiden pituuden, painon ja rasvamäärän muutosta ajassa. Tähän voisi hyödyntää myös kasvukäyrien arvoja priorina. Pitääkö kehittää oma bayes-malli?

Hypoteesia pitäisi täsmentää. Tällä hetkellä ilmassa on useita, mutta mikä on se olennainen?

* Onko POP-pitoisuuksilla yhteyttä sairauksiin (kun vakioidaan rasvan suhteen)?
* Selittääkö POPien yleistyminen obesiteettia? (Tätä on kyllä todella vaikea testata, koska meillä ei ole dataa energiansaannista ja liikunnan määrästä)
* Mitkä ovat POP-pitoisuuden determinantit? (Tämä ei ole ensisijainen kysymys.)
* Miten huomioidaan ruumiin koko, joka on olennainen determinantti sairauksien synnyssä?
* Kemikaalikoktailien tutkiminen on olennainen osa alkuperäistä tutkimussuunnitelmaa. (Miten tarkalleen ottaen?)
* Onko vierasainemetabolia sekoittava tekijä? Joistakin POPeista metabolia tunnetaan aika hyvin. Lisäksi on SNIP-määrityksiä, joista voidaan tehdä genotyyppausta CYP-entsyymien suhteen. Pitäisi kehittää jokin hypoteesi jota tutkitaan. Panulla on tästä jo jokin raportti.

## Data preprocessing and preliminary analyses

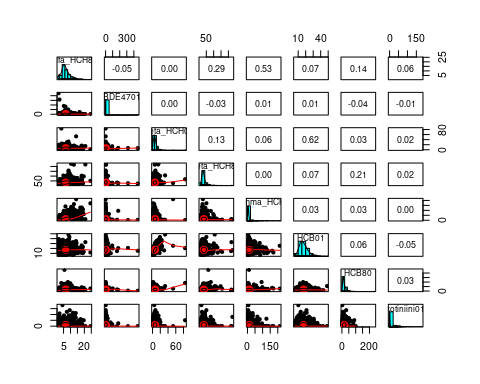
library(foreign)  
library(psych)  
  
df <- read.spss("~/Documents/priv/VS \_Henkilökohtaiset\_POP-trajektorit\_LASERI-aineistossa/updated\_korjattu\_LASERI\_data\_2020\_05\_WORKCOPY.sav", to.data.frame = TRUE)

## Warning in read.spss("~/Documents/priv/VS  
## \_Henkilökohtaiset\_POP-trajektorit\_LASERI-aineistossa/  
## updated\_korjattu\_LASERI\_data\_2020\_05\_WORKCOPY.sav", : ~/Documents/  
## priv/VS \_Henkilökohtaiset\_POP-trajektorit\_LASERI-aineistossa/  
## updated\_korjattu\_LASERI\_data\_2020\_05\_WORKCOPY.sav: Compression bias (0) is not  
## the usual value of 100

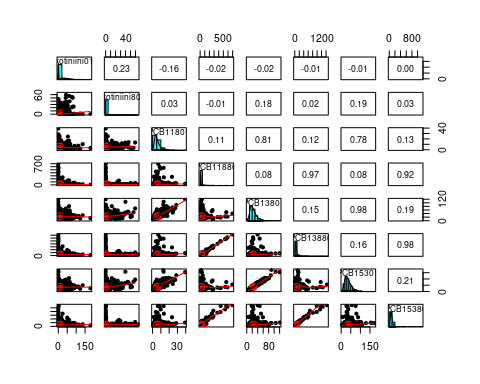
## Warning in read.spss("~/Documents/priv/VS  
## \_Henkilökohtaiset\_POP-trajektorit\_LASERI-aineistossa/  
## updated\_korjattu\_LASERI\_data\_2020\_05\_WORKCOPY.sav", : Undeclared level(s) 0  
## added in variable: bintup80

## Warning in read.spss("~/Documents/priv/VS  
## \_Henkilökohtaiset\_POP-trajektorit\_LASERI-aineistossa/  
## updated\_korjattu\_LASERI\_data\_2020\_05\_WORKCOPY.sav", : Undeclared level(s) 0  
## added in variable: bintup01

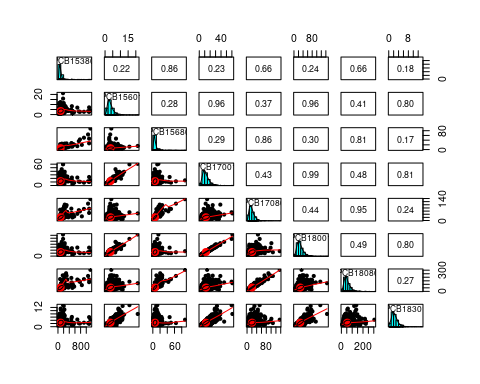
#colnames(df)[grep("PI",toupper(colnames(df)))]  
  
units <- read.csv("POP\_units.csv")  
  
# Get normalised (per fat?) concentrations  
tmp <- df[grepl("\_norm",colnames(df))][-2] # remove alfa\_HCH01, which contains no data  
colnames(tmp) <- gsub("\_norm","",colnames(tmp))  
# Compare 1980 (80) and 2001 (01) in pairwise correlations  
tmp <- tmp[sort(colnames(tmp))]  
pairs.panels(tmp[1:8])



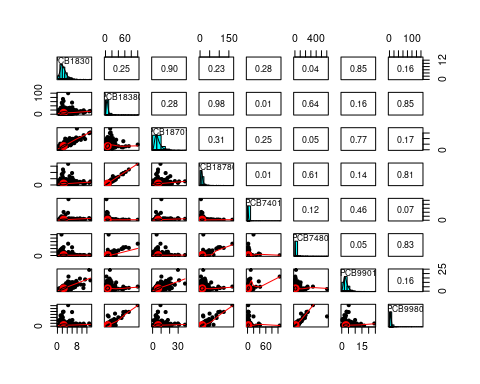
pairs.panels(tmp[8:15])



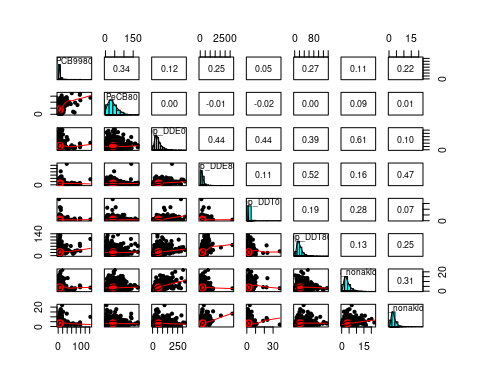
pairs.panels(tmp[15:22])



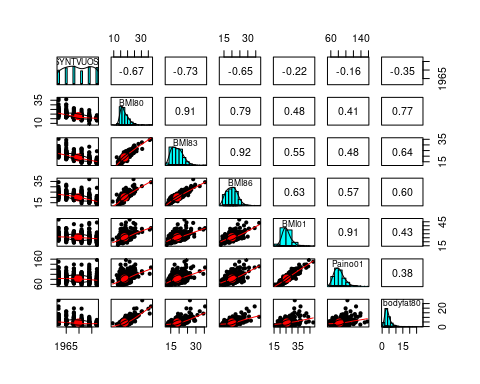
pairs.panels(tmp[22:29])



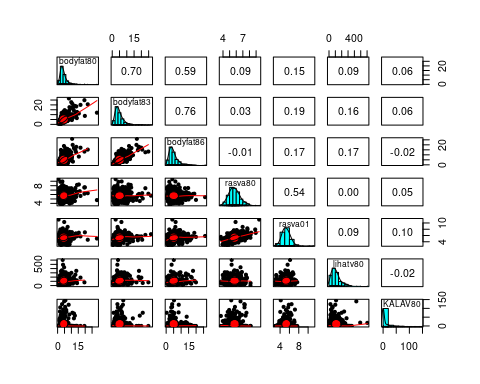
pairs.panels(tmp[29:36])



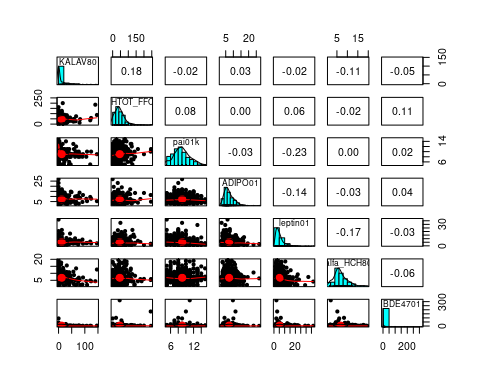
#colnames(tmp)  
# [1] "HCB01" "beta\_HCH01" "Trans\_nonakloori01"  
# [4] "pp\_DDT01" "pp\_DDE01" "PCB7401"   
# [7] "PCB9901" "PCB11801" "PCB15301"   
#[10] "PCB13801" "PCB15601" "PCB18701"   
#[13] "PCB18301" "PCB18001" "PCB17001"   
#[16] "BDE4701" "kotiniini01" "PeCB80"   
#[19] "HCB80" "alfa\_HCH80" "beta\_HCH80"   
#[22] "gamma\_HCH80" "Trans\_nonakloori80" "pp\_DDT80"   
#[25] "pp\_DDE80" "PCB7480" "PCB9980"   
#[28] "PCB11880" "PCB15380" "PCB13880"   
#[31] "PCB15680" "PCB18780" "PCB18380"   
#[34] "PCB18080" "PCB17080" "kotiniini80"   
  
POP <- c(  
 "HCB", # https://en.wikipedia.org/wiki/Hexachlorobenzene  
 "alfa\_HCH", # https://en.wikipedia.org/wiki/Alpha-Hexachlorocyclohexane  
 "beta\_HCH",  
 "gamma\_HCH", # also known as lindane  
 "Trans\_nonakloori", # https://en.wikipedia.org/wiki/Chlordane  
 "pp\_DDT", # para,para'-DDT https://en.wikipedia.org/wiki/DDT  
 "pp\_DDE",   
 "PCB74", # ug/kg fat, https://en.wikipedia.org/wiki/Polychlorinated\_biphenyl  
 "PCB99",  
 "PCB118",  
 "PCB153",  
 "PCB138",  
 "PCB156",  
 "PCB187",  
 "PCB183",  
 "PCB180",  
 "PCB170",  
 "BDE47", # https://en.wikipedia.org/wiki/Pentabromodiphenyl\_ether  
 "kotiniini" # https://en.wikipedia.org/wiki/Cotinine  
)  
  
# Take all-male panel of interesting variables  
df <- cbind(df[c(  
 "SYNTVUOSI", # birth year  
 "BMI80", # body mass index in 1980  
 "BMI83", # body mass index in 1983  
 "BMI86", # body mass index in 1986  
 "BMI01", # body mass index in 2001  
 "Paino01", # body weight kg in 2001  
 "bodyfat80", # body fat % in 1980  
 "bodyfat83", # body fat % in 1983  
 "bodyfat86", # body fat % in 1986  
 "rasva80", # serum fat content mg/ml in 1980  
 "rasva01", # serum fat content mg/ml in 2001  
 "lihatv80", # consumption of meat and meat products g/d in 1980  
 "KALAV80", # consumption of fish and fish products g/d in 1980  
 "FISHTOT\_FFQ07", # total fish consumption, g/d in 2007 food frequency questionnaire  
 "pai01k", # physical activity index in 2001  
 "ADIPO01", # adiponectin hormone in serum, ug/ml in 2001  
 "leptin01" # leptin hormone concentration ng/ml in 2001  
 )],  
 tmp  
)[df$SP=="Poika", ]  
# Very few observations in bodyfat92, Paino80, leptin80 --> omit  
# It would be good to obtain these variables:  
# ENERC07 energy consumption kJ/d from 2007 FFQ  
# FAT07 fat consumption g/d from 2007 FFQ  
  
pairs.panels(df[1:7])



pairs.panels(df[7:13])



pairs.panels(df[13:19])



### Select the population for analysis  
  
dfs <- na.omit(df[c(1:2,5:6,14,31:32)])  
  
## Indices used  
  
AGE <- 0:40  
H <- length(AGE)  
YEAR <- 1960:2020  
I <- nrow(dfs)  
SYNT <- dfs$SYNTVUOSI  
  
## Data used  
  
U <- 365 # unit conversion factor: 1/d --> 1/a  
A <- 0.9 # fraction, absorption fraction  
N <- params$N  
b <- 0.07 # 1/a, rate of decrease of the POP in the environment and intake  
c <- params$c # year of peak exposure  
P <- rep(1,H) # age-specific intake relative to lifetime average  
# Estimated from the Bayesian model:  
## k <- 0.35 # 1/a, elimination constant of the POP  
## a <- 1.1 # ug/kg/d, average intake at the peak exposure  
  
conc <- dfs[1:I,c("PCB15380","PCB15301")]  
fish <- dfs$FISHTOT\_FFQ07[1:I]  
  
# https://www.kasvuseula.fi/  
W <- t(matrix(rep(c(5,10,13,15,17,20,22,26,28,31,35,38,44,48,55,60,65,68,70,71,72,  
 rep(75,10),rep(80,10)),I),  
 nrow=H, dimnames=list(Age=AGE, Individual = 1:I)))  
  
# A simplistic model uses the population average as default until the weight reaches the  
# observed value of that individual. Then, actual value is used to replace 80 kg (the max default),  
# if larger. This leads to situation where people suddenly gain weight when at 31 or at 2001.  
W <- pmin(W, dfs$Paino01, na.rm = TRUE)  
W <- ifelse(W==80, pmax(W, dfs$Paino01, na.rm=TRUE), W)  
tmp <- dfs$SYNTVUOSI + matrix(rep(AGE,each=I),nrow=I)>=2001  
W <- ifelse(!is.na(dfs$Paino01) & tmp, dfs$Paino01, W)  
  
# A simplistic model that assumes that average BMI is equal to lifetime fat percentage.  
M <- W \* ifelse(is.na(dfs$BMI01), dfs$BMI80, (dfs$BMI80 + dfs$BMI01)/2) / 100  
  
#aggregate(df$Paino01, by=df[c("Age80")], FUN=function(x) mean(x, na.rm=TRUE))

The following conclusions can be made from the graphs above

* 1. PCBs correlate very well with each other.
  2. PCB correlation is much stronger between congeners at the same year (R>0.8) than the same congener 21 years apart (R = 0.2 - 0.4).
  3. DDT and DDE congeners correlate moderately with each other and across years (R ca. 0.4)
  4. Trans-nonachlor correlates well with DDE of the same year (R 0.45-0.6) and moderately across years (R ).
  5. HCH congeners correlate only with alpha-HCH in 1980 (R 0.3 - 0.5) but not across years nor beta vs gamma.
  6. HCB correlates well with alpha-HCH in 1980 but poorly across other congeners or years.
  7. Cotinine correlates poorly across years (R = 0.2) and not at all with POPs.

## Deterministic data analyses

library(ggplot2)

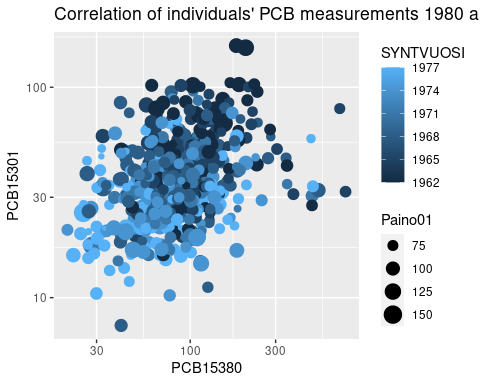
##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

mimax <- c(min(df[c("PCB15380","PCB15301")]),max(df[c("PCB15380","PCB15301")]))  
ggplot()+  
 geom\_point(data=df, aes(x=PCB15380, y=PCB15301, size=Paino01, colour=SYNTVUOSI))+  
 geom\_line(data=data.frame(x=mimax, y=mimax), aes(x=x, y=y),color="red")+  
 scale\_x\_log10()+scale\_y\_log10()+  
 labs(title="Correlation of individuals' PCB measurements 1980 and 2001")

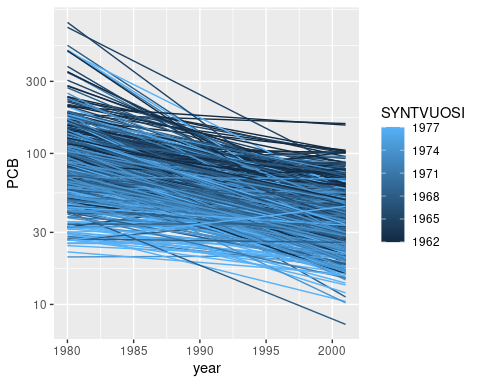
## Warning: Removed 27 rows containing missing values (geom\_point).

## Warning: Removed 2 row(s) containing missing values (geom\_path).

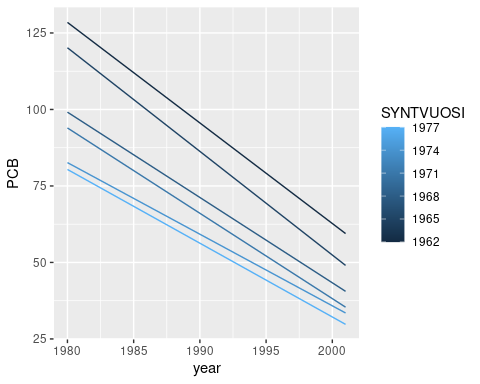


df$id <- 1:nrow(df)  
df$Paino80 <- NA  
df$bodyfat01 <- NA  
long <- reshape(df[c("SYNTVUOSI","id","PCB15380","PCB15301","BMI80","BMI01","Paino80","Paino01","bodyfat80","bodyfat01")],  
 idvar=c("SYNTVUOSI","id"),  
 varying=list(  
 PCB=c("PCB15380","PCB15301"),  
 BMI= c("BMI80","BMI01"),  
 weight=c("Paino80","Paino01"),  
 fat=c("bodyfat80","bodyfat01")  
 ),  
 v.names =c("PCB","BMI","weight","fat"),  
 timevar="year",  
 times=c(1980,2001),  
 direction="long")  
  
ggplot(long, aes(x=year, y=PCB, colour=SYNTVUOSI, group=id))+geom\_line()+  
 scale\_y\_log10()

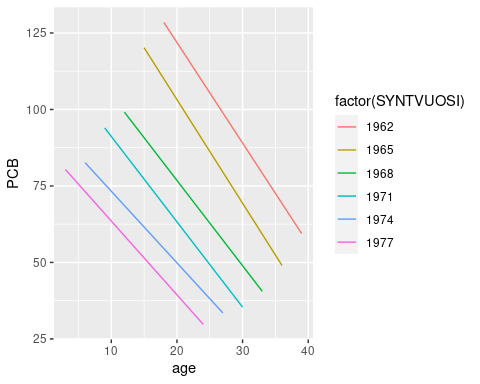
## Warning: Removed 27 row(s) containing missing values (geom\_path).



ggplot(aggregate(long["PCB"], by=long[c("SYNTVUOSI","year")], FUN=function(x) mean(x, na.rm=TRUE)),  
 aes(x=year, y=PCB, colour=SYNTVUOSI, group=SYNTVUOSI))+geom\_line()



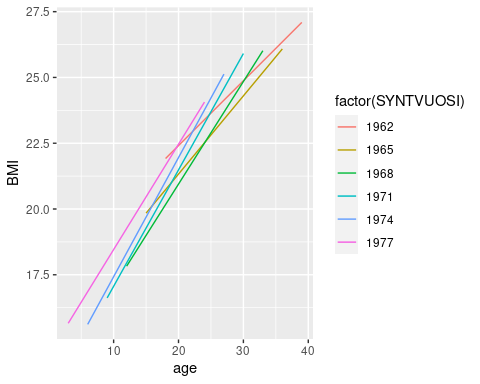
long$age <- long$year - long$SYNTVUOSI  
ggplot(aggregate(long["PCB"], by=long[c("SYNTVUOSI","age")], FUN=function(x) mean(x, na.rm=TRUE)),  
 aes(x=age, y=PCB, colour=factor(SYNTVUOSI), group=SYNTVUOSI))+geom\_line()



cat("Concentrations in 20-year-old people decrease", signif((1-(40/120)^(1/(1977-1962)))\*100,2), " % per year\n")

## Concentrations in 20-year-old people decrease 7.1 % per year

ggplot(aggregate(long["BMI"], by=long[c("SYNTVUOSI","age")], FUN=function(x) mean(x, na.rm=TRUE)),  
 aes(x=age, y=BMI, colour=factor(SYNTVUOSI), group=SYNTVUOSI))+geom\_line()

 ## Linear regressions to analyse impacts of variables

# Remove the least important variables from the model  
reg <- lm(PCB15380 ~ SYNTVUOSI + BMI80 + bodyfat80 + KALAV80 + lihatv80, data=df)  
reg <- lm(PCB15380 ~ SYNTVUOSI + BMI80 + bodyfat80 + KALAV80, data=df)  
reg <- lm(PCB15380 ~ SYNTVUOSI + BMI80 + bodyfat80, data=df)  
summary(reg)

##   
## Call:  
## lm(formula = PCB15380 ~ SYNTVUOSI + BMI80 + bodyfat80, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -79.72 -32.89 -14.55 12.24 593.65   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7552.2648 1534.7089 4.921 1.12e-06 \*\*\*  
## SYNTVUOSI -3.7926 0.7697 -4.927 1.09e-06 \*\*\*  
## BMI80 3.1155 1.7621 1.768 0.0776 .   
## bodyfat80 -8.4619 1.5169 -5.579 3.70e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 66.86 on 587 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.1261, Adjusted R-squared: 0.1217   
## F-statistic: 28.24 on 3 and 587 DF, p-value: < 2.2e-16

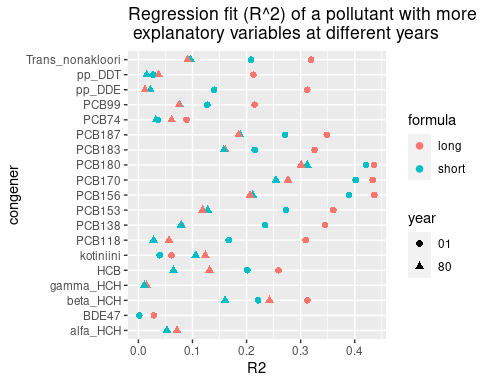
#reg <- lm(PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + Paino01 + PCB15380 + bodyfat80 + KALAV80 + lihatv80 + FISHTOT\_FFQ07, data=df)   
#reg <- lm(PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + Paino01 + PCB15380 + bodyfat80 + KALAV80 + FISHTOT\_FFQ07, data=df)   
#reg <- lm(PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + Paino01 + PCB15380 + bodyfat80 + FISHTOT\_FFQ07, data=df)   
#reg <- lm(PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + Paino01 + PCB15380 + FISHTOT\_FFQ07, data=df)   
reg <- lm(PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + PCB15380 + FISHTOT\_FFQ07, data=df)   
reg <- step(reg, direction = "both")

## Start: AIC=2142.71  
## PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + PCB15380 + FISHTOT\_FFQ07  
##   
## Df Sum of Sq RSS AIC  
## <none> 112865 2142.7  
## - FISHTOT\_FFQ07 1 1582.1 114447 2145.9  
## - BMI01 1 2232.2 115097 2148.0  
## - SYNTVUOSI 1 3231.0 116096 2151.2  
## - PCB15380 1 4062.6 116928 2153.9  
## - BMI80 1 11122.9 123988 2175.8

summary(reg)

##   
## Call:  
## lm(formula = PCB15301 ~ SYNTVUOSI + BMI80 + BMI01 + PCB15380 +   
## FISHTOT\_FFQ07, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -51.865 -10.461 -3.021 8.375 100.639   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1631.40537 505.96243 3.224 0.001376 \*\*   
## SYNTVUOSI -0.82494 0.25451 -3.241 0.001299 \*\*   
## BMI80 2.66412 0.44299 6.014 4.38e-09 \*\*\*  
## BMI01 -0.74422 0.27623 -2.694 0.007381 \*\*   
## PCB15380 0.04359 0.01199 3.635 0.000318 \*\*\*  
## FISHTOT\_FFQ07 0.06032 0.02660 2.268 0.023902 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 17.54 on 367 degrees of freedom  
## (219 observations deleted due to missingness)  
## Multiple R-squared: 0.3243, Adjusted R-squared: 0.3151   
## F-statistic: 35.22 on 5 and 367 DF, p-value: < 2.2e-16

det1 <- "SYNTVUOSI + BMI80 + BMI01 + Paino01 + bodyfat80"  
det2 <- " + KALAV80 + lihatv80 + FISHTOT\_FFQ07"  
  
df$alfa\_HCH01 <- NA  
df$BDE4780 <- NA  
df$gamma\_HCH01 <- NA  
  
  
out <- data.frame()  
for(j in POP) {  
 for(i in c("80","01")) {  
 for(k in 1:2) {  
 if(!all(is.na(df[[paste0(j,i)]]))) {  
 reg <- lm(as.formula(paste0(j, i," ~ ", c(det1, paste0(det1, det2))[k])), data = df)  
 out <- rbind(out,  
 cbind(  
 congener=j,  
 year=i,  
 formula = c("short","long")[k],  
 as.data.frame(summary(reg)[[4]]),  
 N = summary(reg)[[7]][2]+1,  
 R2 = summary(reg)[[8]][1]  
 )  
 )  
 }}}  
}  
  
tmp <- aggregate(out["R2"], by=out[c("congener","year","formula")],FUN=mean)  
tmp <- tmp[order(-tmp$R2),]  
ggplot(tmp, aes(x=congener, y=R2, colour=formula, shape=year))+geom\_point(size=2)+  
 coord\_flip()+  
 labs(title="Regression fit (R^2) of a pollutant with more or less\n explanatory variables at different years")



For PCB15380, we can see that

1. KALAV80 and lihatv80 bring little if any predictive power to the regression model
2. FISHTOT\_FFQ07 seems to have some predictive power

## Bayesian hierachical model for POP kinetics

# This was forked from Dioxdistboys\_individual\_rows.Rmd and  
# originally from code Op\_en3104/bayes on page [[EU-kalat]]  
  
## This code builds an iterative model with varying intake  
## https://github.com/jtuomist/dioxdisthuman/wiki#iterative-model-with-varying-intake-primary-choice  
  
#if(params$run\_bayes) {  
 library(OpasnetUtils)  
 library(reshape2)  
 library(rjags) # JAGS

## Loading required package: coda

## Linked to JAGS 4.3.0

## Loaded modules: basemod,bugs

library(MASS) # mvrnorm  
 library(car) # scatterplotMatrix

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

library(mcmcr) # as.mcarray

## Registered S3 method overwritten by 'mcmcr':  
## method from   
## as.mcmc.list.mcarray rjags

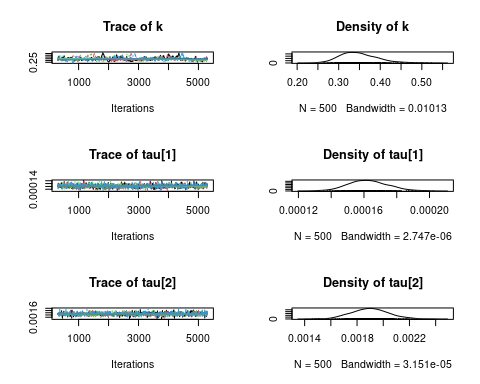
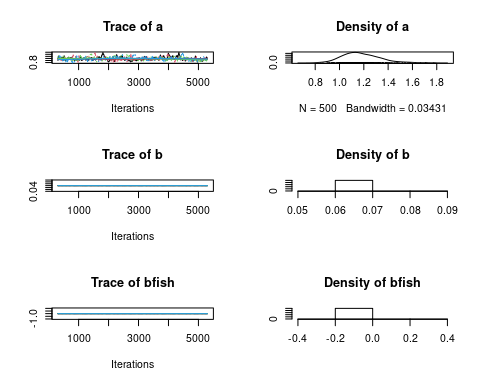
##   
## Attaching package: 'mcmcr'

## The following object is masked from 'package:OpasnetUtils':  
##   
## tidy

N <- params$N # defined as render parameter  
   
 # Hierarchical Bayes model.  
   
 mod <- textConnection(  
 "  
 var k, a, b, U, E[61], C[I,H], B[I,H], M[I,H], C1, C2, conc[I,2], SYNT[I], Efish[I],  
 B1[I], B2[I], B3[I], B4[I], B5[I], B6[I], B7[I], B8[I], B9[I], B10[I], B11[I],  
 B12[I], B13[I], B14[I], B15[I], B16[I], B17[I], B18[I], B19[I], B20[I], B21[I], B22[I],  
 B23[I], B24[I], B25[I], B26[I], B27[I], B28[I], B29[I], B30[I], B31[I], B32[I], B33[I],  
 B34[I], B35[I], B36[I], B37[I], B38[I], B39[I], B40[I], B41[I];  
 model{  
 # below.LOQ[i,j] ~ dinterval(-conc[i,j], -LOQ[j])  
 k ~ dnorm(0.3, 0.1)  
 a ~ dnorm(1, 0.1)  
# b ~ dnorm(0.1, 0.001)  
 tau[1] ~ dgamma(1.0,1.0)   
 tau[2] ~ dgamma(1.0,1.0)   
# bfish ~ dnorm(1.0,0.001)   
 bfish <- 0  
 Efish[] <- fish[] \* bfish  
 for(year in YEAR) {  
 E[year-1959] <- max(0,a\*b\*exp(1)\*(year-c+1/b)\*exp(-b\*(year-c+1/b)))  
 }  
 B2[] <- B1[] \* (1 - k) + U \* A \* W[,1] \* (E[SYNT[]-1960+1] + Efish[]) \* P[1]  
 B3[] <- B2[] \* (1 - k) + U \* A \* W[,2] \* (E[SYNT[]-1960+2] + Efish[]) \* P[2]  
 B4[] <- B3[] \* (1 - k) + U \* A \* W[,3] \* (E[SYNT[]-1960+3] + Efish[]) \* P[3]  
 B5[] <- B4[] \* (1 - k) + U \* A \* W[,4] \* (E[SYNT[]-1960+4] + Efish[]) \* P[4]  
 B6[] <- B5[] \* (1 - k) + U \* A \* W[,5] \* (E[SYNT[]-1960+5] + Efish[]) \* P[5]  
 B7[] <- B6[] \* (1 - k) + U \* A \* W[,6] \* (E[SYNT[]-1960+6] + Efish[]) \* P[6]  
 B8[] <- B7[] \* (1 - k) + U \* A \* W[,7] \* (E[SYNT[]-1960+7] + Efish[]) \* P[7]  
 B9[] <- B8[] \* (1 - k) + U \* A \* W[,8] \* (E[SYNT[]-1960+8] + Efish[]) \* P[8]  
 B10[] <- B9[] \* (1 - k) + U \* A \* W[,9] \* (E[SYNT[]-1960+9] + Efish[]) \* P[9]  
 B11[] <- B10[] \* (1 - k) + U \* A \* W[,10] \* (E[SYNT[]-1960+10] + Efish[]) \* P[10]  
 B12[] <- B11[] \* (1 - k) + U \* A \* W[,11] \* (E[SYNT[]-1960+11] + Efish[]) \* P[11]  
 B13[] <- B12[] \* (1 - k) + U \* A \* W[,12] \* (E[SYNT[]-1960+12] + Efish[]) \* P[12]  
 B14[] <- B13[] \* (1 - k) + U \* A \* W[,13] \* (E[SYNT[]-1960+13] + Efish[]) \* P[13]  
 B15[] <- B14[] \* (1 - k) + U \* A \* W[,14] \* (E[SYNT[]-1960+14] + Efish[]) \* P[14]  
 B16[] <- B15[] \* (1 - k) + U \* A \* W[,15] \* (E[SYNT[]-1960+15] + Efish[]) \* P[15]  
 B17[] <- B16[] \* (1 - k) + U \* A \* W[,16] \* (E[SYNT[]-1960+16] + Efish[]) \* P[16]  
 B18[] <- B17[] \* (1 - k) + U \* A \* W[,17] \* (E[SYNT[]-1960+17] + Efish[]) \* P[17]  
 B19[] <- B18[] \* (1 - k) + U \* A \* W[,18] \* (E[SYNT[]-1960+18] + Efish[]) \* P[18]  
 B20[] <- B19[] \* (1 - k) + U \* A \* W[,19] \* (E[SYNT[]-1960+19] + Efish[]) \* P[19]  
 B21[] <- B20[] \* (1 - k) + U \* A \* W[,20] \* (E[SYNT[]-1960+20] + Efish[]) \* P[20]  
 B22[] <- B21[] \* (1 - k) + U \* A \* W[,21] \* (E[SYNT[]-1960+21] + Efish[]) \* P[21]  
 B23[] <- B22[] \* (1 - k) + U \* A \* W[,22] \* (E[SYNT[]-1960+22] + Efish[]) \* P[22]  
 B24[] <- B23[] \* (1 - k) + U \* A \* W[,23] \* (E[SYNT[]-1960+23] + Efish[]) \* P[23]  
 B25[] <- B24[] \* (1 - k) + U \* A \* W[,24] \* (E[SYNT[]-1960+24] + Efish[]) \* P[24]  
 B26[] <- B25[] \* (1 - k) + U \* A \* W[,25] \* (E[SYNT[]-1960+25] + Efish[]) \* P[25]  
 B27[] <- B26[] \* (1 - k) + U \* A \* W[,26] \* (E[SYNT[]-1960+26] + Efish[]) \* P[26]  
 B28[] <- B27[] \* (1 - k) + U \* A \* W[,27] \* (E[SYNT[]-1960+27] + Efish[]) \* P[27]  
 B29[] <- B28[] \* (1 - k) + U \* A \* W[,28] \* (E[SYNT[]-1960+28] + Efish[]) \* P[28]  
 B30[] <- B29[] \* (1 - k) + U \* A \* W[,29] \* (E[SYNT[]-1960+29] + Efish[]) \* P[29]  
 B31[] <- B30[] \* (1 - k) + U \* A \* W[,30] \* (E[SYNT[]-1960+30] + Efish[]) \* P[30]  
 B32[] <- B31[] \* (1 - k) + U \* A \* W[,31] \* (E[SYNT[]-1960+31] + Efish[]) \* P[31]  
 B33[] <- B32[] \* (1 - k) + U \* A \* W[,32] \* (E[SYNT[]-1960+32] + Efish[]) \* P[32]  
 B34[] <- B33[] \* (1 - k) + U \* A \* W[,33] \* (E[SYNT[]-1960+33] + Efish[]) \* P[33]  
 B35[] <- B34[] \* (1 - k) + U \* A \* W[,34] \* (E[SYNT[]-1960+34] + Efish[]) \* P[34]  
 B36[] <- B35[] \* (1 - k) + U \* A \* W[,35] \* (E[SYNT[]-1960+35] + Efish[]) \* P[35]  
 B37[] <- B36[] \* (1 - k) + U \* A \* W[,36] \* (E[SYNT[]-1960+36] + Efish[]) \* P[36]  
 B38[] <- B37[] \* (1 - k) + U \* A \* W[,37] \* (E[SYNT[]-1960+37] + Efish[]) \* P[37]  
 B39[] <- B38[] \* (1 - k) + U \* A \* W[,38] \* (E[SYNT[]-1960+38] + Efish[]) \* P[38]  
 B40[] <- B39[] \* (1 - k) + U \* A \* W[,39] \* (E[SYNT[]-1960+39] + Efish[]) \* P[39]  
 B41[] <- B40[] \* (1 - k) + U \* A \* W[,40] \* (E[SYNT[]-1960+40] + Efish[]) \* P[40]  
   
 C[,1] <- B1 / M[,1]  
 C[,2] <- B2 / M[,2]  
 C[,3] <- B3 / M[,3]  
 C[,4] <- B4 / M[,4]  
 C[,5] <- B5 / M[,5]  
 C[,6] <- B6 / M[,6]  
 C[,7] <- B7 / M[,7]  
 C[,8] <- B8 / M[,8]  
 C[,9] <- B9 / M[,9]  
 C[,10] <- B10 / M[,10]  
 C[,11] <- B11 / M[,11]  
 C[,12] <- B12 / M[,12]  
 C[,13] <- B13 / M[,13]  
 C[,14] <- B14 / M[,14]  
 C[,15] <- B15 / M[,15]  
 C[,16] <- B16 / M[,16]  
 C[,17] <- B17 / M[,17]  
 C[,18] <- B18 / M[,18]  
 C[,19] <- B19 / M[,19]  
 C[,20] <- B20 / M[,20]  
 C[,21] <- B21 / M[,21]  
 C[,22] <- B22 / M[,22]  
 C[,23] <- B23 / M[,23]  
 C[,24] <- B24 / M[,24]  
 C[,25] <- B25 / M[,25]  
 C[,26] <- B26 / M[,26]  
 C[,27] <- B27 / M[,27]  
 C[,28] <- B28 / M[,28]  
 C[,29] <- B29 / M[,29]  
 C[,30] <- B30 / M[,30]  
 C[,31] <- B31 / M[,31]  
 C[,32] <- B32 / M[,32]  
 C[,33] <- B33 / M[,33]  
 C[,34] <- B34 / M[,34]  
 C[,35] <- B35 / M[,35]  
 C[,36] <- B36 / M[,36]  
 C[,37] <- B37 / M[,37]  
 C[,38] <- B38 / M[,38]  
 C[,39] <- B39 / M[,39]  
 C[,40] <- B40 / M[,40]  
 C[,41] <- B41 / M[,41]  
   
 for(i in 1:I) {   
 conc[i,1] ~ dnorm(C[i,1980 - SYNT[i]+1], tau[1]) # Age 0 is at position 1  
 conc[i,2] ~ dnorm(C[i,2001 - SYNT[i]+1], tau[2])  
 pred[i,1] ~ dnorm(C[i,1980 - SYNT[i]+1], tau[1])  
 pred[i,2] ~ dnorm(C[i,2001 - SYNT[i]+1], tau[2])  
 }  
 }  
 ")  
  
 jags <- jags.model(  
 mod,  
 data = list(  
 YEAR = YEAR,  
 W = W[1:I,1:H],  
 M = M[1:I,1:H],  
 H = H,  
 I = I,  
 U = U,  
 A = A,  
 b = b,  
 c = c,  
 P = P[1:H],  
 SYNT = SYNT,  
 conc = conc,  
 fish = fish,  
 B1 = rep(0,I)  
 ),  
 n.chains = 4,  
 n.adapt = 300  
 )

## Compiling model graph  
## Declaring variables  
## Resolving undeclared variables  
## Allocating nodes  
## Graph information:  
## Observed stochastic nodes: 746  
## Unobserved stochastic nodes: 750  
## Total graph size: 34815  
##   
## Initializing model

samps.c <- coda.samples(  
 jags,   
 variable.names=c("k","a","b","tau","bfish"),   
 n.iter=N\*10,  
 thin=10  
 )  
   
 samps.j <- jags.samples(  
 jags,   
 variable.names=c("k","a","b","tau","E","pred","bfish"),   
 n.iter=N\*10,  
 thin=10  
 )  
  
 #scatterplotMatrix(samps.j$k, main="Elimination constant k")  
   
 plot(samps.c)

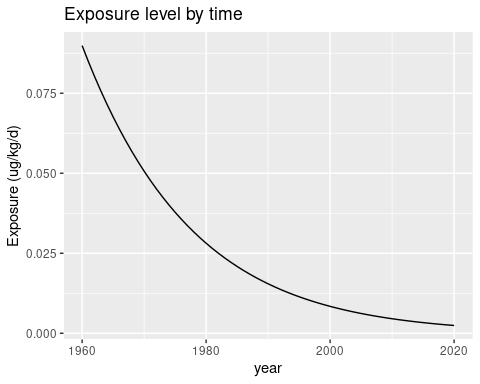


#}

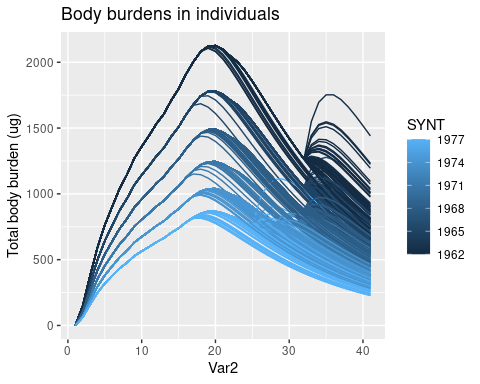
* k: elimination constant (1/a)
* a: maximum exposure at the year of peak production (ug/kg/d)
* b: nominal rate of environmental concentration decrease (1/a)
* tau: precision for individual concentration estimate at year x (1980 or 2001)
* E: exposure at a calendar year y (between 1960 and 2020)
* pred: predicted individual concentration (ug/kg fat)

## Results from the Bayesian kinetic model

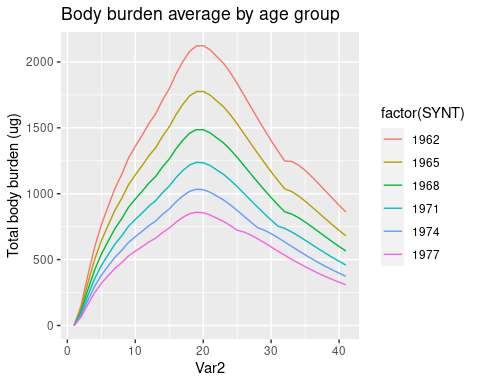
k <- mean(samps.j$k)  
a <- mean(samps.j$a)  
b <- mean(samps.j$b)  
  
E <- pmax(0,a\*b\*exp(1)\*(YEAR-c+1/b)\*exp(-b\*(YEAR-c+1/b)))  
B <- matrix(0, nrow=I, ncol=H)  
for(age in 2:H) {  
 B[,age] <- B[,age-1] \* (1 - k) + U \* A \* W[,age-1] \* E[SYNT-1963+age] \* P[age]  
}  
C <- B / M  
  
ggplot(data.frame(year=YEAR,exposure=E), aes(x=year, y=E))+geom\_line()+  
 labs(title="Exposure level by time",  
 y="Exposure (ug/kg/d)")



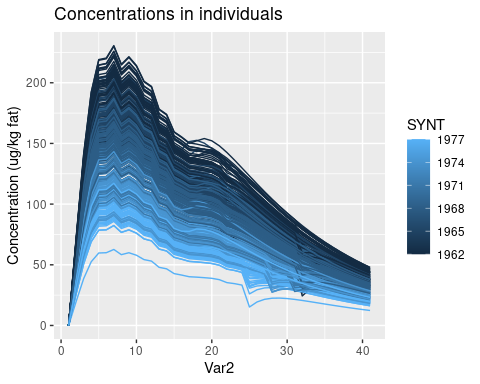
dat <- melt(B)#[seq(1,550, 10),])  
dat <- merge(dat, cbind(Var1= 1:I, SYNT=SYNT))  
  
ggplot(dat, aes(x=Var2, y=value, colour=SYNT, group=Var1))+geom\_line()+  
 labs(title="Body burdens in individuals",  
 y="Total body burden (ug)")



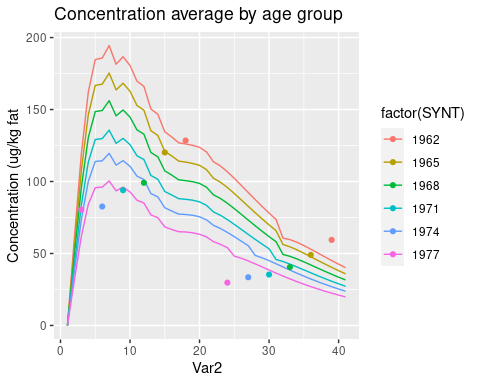
ggplot(aggregate(dat["value"], by=dat[c("Var2","SYNT")], FUN=function(x) mean(x, na.rm=TRUE)),   
 aes(x=Var2, y=value, colour=factor(SYNT), group=SYNT))+geom\_line()+  
 labs(title="Body burden average by age group",  
 y="Total body burden (ug)")



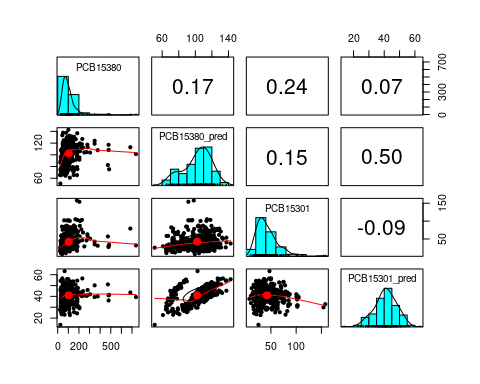
dat <- melt(C)#[seq(1,550, 10),])  
dat <- merge(dat, cbind(Var1= 1:I, SYNT=SYNT))  
  
ggplot(dat, aes(x=Var2, y=value, colour=SYNT, group=Var1))+geom\_line()+  
 labs(title="Concentrations in individuals",  
 y="Concentration (ug/kg fat)")



ggplot()+  
 geom\_line(data=aggregate(dat["value"], by=dat[c("Var2","SYNT")], FUN=function(x) mean(x, na.rm=TRUE)),   
 aes(x=Var2, y=value, colour=factor(SYNT), group=SYNT))+  
 geom\_point(data=aggregate(long["PCB"],by=long[c("SYNTVUOSI","age")],FUN=function(x) mean(x, na.rm=TRUE)),  
 aes(x=age, y=PCB,colour=factor(SYNTVUOSI)))+geom\_point()+  
 labs(title="Concentration average by age group",  
 y="Concentration (ug/kg fat")



tst <- rowMeans(samps.j$pred[,,,], dims=2)  
tst <- data.frame(  
 dfs[c("PCB15380")],  
 PCB15380\_pred = tst[,1],  
 dfs[c("PCB15301")],  
 PCB15301\_pred = tst[,2]  
)  
pairs.panels(tst)

 In conclusion, the model predicts OK the agegroup averages but does not predict individual variation. The correlation between measured and predicted PCB153 in 2001 are actually negative.

In addition: why do the predicted values form two subpopulations when 1980 and 2001 are scatterplotted?