

Supplementary material

```
library(MASS);library(dplyr);library(nlme);library(mgcv);library(gdata);library(reshape);library(psych)
library(ggplot2);library(reshape2);library(plyr)
```

```
dfaF <- subset(PSVO2max_excl_missing_data_HR_RER_BMI_rokers_en_sporters_21018)
dfaF065 <- subset(dfaF,subset = Age<65.1)
femalesF065 <- subset(dfaF065,subset = Gender=="0")
malesF065 <- subset(dfaF065,subset = Gender=="1")
dfaF065$Gender <- with(dfaF065,factor(ifelse(Gender == 1, "male", "female")))
describe(dfaF065)
describe(malesF065[1-14])
describe(femalesF065[1-14])
hist(dfaF065$Age, main= "Age distribution", xlab= "Age")
```

```
linmodF065 <- gam(VO2max~Gender+ Age+Length+Weight,data=dfaF065)
summary(linmodF065)
polmodF065 <-
lm(VO2max~Gender+Age+l(Age^2)+l(Age^3)+Length+l(Length^2)+l(Length^3)+Weight+l(Weight^2)+
l(Weight^3),data=dfaF065)
summary(polmodF065)
polmodF065minAIC <- stepAIC(polmodF065, trace=0, k = log(nrow(dfaF065)))
summary(polmodF065minAIC)
gammodF065 <- gam(VO2max~Gender + s(Age,by=Gender) + Length+ Weight ,data=dfaF065)
summary(gammodF065)
cbind(AIC(linmodF065,polmodF065minAIC,gammodF065),BIC(linmodF065,polmodF065minAIC,gammodF065))
```

```
dfaF065.all <- subset(dfaF065,Age<100)
predictive_explanatory_variables_allF065 <- dfaF065.all[,1:14]
dfaF065.adult <- subset(dfaF065,Age>20)
predictive_explanatory_variables_adultF065 <- dfaF065.adult[,1:14]
dfaF065.child <- subset(dfaF065,Age<=20)
predictive_explanatory_variables_childF065 <- dfaF065.child[,1:14]
```

#R2 & SEE Training

#linear

```
yhat <- predict(linmodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_allF065)
cor(yhat$fit,dfaF065.all$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.all$VO2max)^2)/nrow(dfaF065.all))
yhat <- predict(linmodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_adultF065)
cor(yhat$fit,dfaF065.adult$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.adult$VO2max)^2)/nrow(dfaF065.adult))
yhat <- predict(linmodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_childF065)
cor(yhat$fit,dfaF065.child$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.child$VO2max)^2)/nrow(dfaF065.child))
```

#Poly

```

yhat <- predict(polmodF065minAIC,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_allF065)
cor(yhat$fit,dfaF065.all$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.all$VO2max)^2)/nrow(dfaF065.all))
yhat <- predict(polmodF065minAIC,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_adultF065)
cor(yhat$fit,dfaF065.adult$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.adult$VO2max)^2)/nrow(dfaF065.adult))
yhat <- predict(polmodF065minAIC,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_childF065)
cor(yhat$fit,dfaF065.child$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.child$VO2max)^2)/nrow(dfaF065.child))

```

#GAM

```

yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_allF065)
cor(yhat$fit,dfaF065.all$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.all$VO2max)^2)/nrow(dfaF065.all))
yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_adultF065)
cor(yhat$fit,dfaF065.adult$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.adult$VO2max)^2)/nrow(dfaF065.adult))
yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_childF065)
cor(yhat$fit,dfaF065.child$VO2max)^2
sqrt(sum((yhat$fit-dfaF065.child$VO2max)^2)/nrow(dfaF065.child))

```

#CROSS VALIDATION

```

dfacvF065 <- CV_compleet2_met_filters_RER_HF85_smokers_bmi
dfacvF065 <-
with(dfacvF065,data.frame(cbind(VO2max,Gender,Age,Length,Weight,BMI,HRmax,RER)))
dfacvF065 <- with(dfacvF065,data.frame(cbind(VO2max,Gender,Age,Length,Weight,BMI)))

dfacvF065$BMI <- (dfacvF065$Weight) / ((dfacvF065$Length/100)*(dfacvF065$Length/100))
dfacvF065$VO2max <- (dfacvF065$VO2max*1000)
femalescvF065<-subset(dfacvF065,subset = Gender=="0")
malescvF065<-subset(dfacvF065,subset = Gender=="1")
dfacvF065$Gender <- with(dfacvF065,factor(ifelse(Gender == 1, "male", "female")))
describe(dfacvF065[1-14])
describe(malescvF065[1-14])
describe(femalescvF065[1-14])

```

```

dfacvF065.all <- subset(dfacvF065,Age<=100)
predictive_explanatory_variables_allF065 <- dfacvF065.all[,2:6]
dfacvF065.child <- subset(dfacvF065,Age<=20)
predictive_explanatory_variables_childF065 <- dfacvF065.child[,2:6]
dfacvF065.adult <- subset(dfacvF065,Age>20)
predictive_explanatory_variables_adultF065 <- dfacvF065.adult[,2:6]

```

#cross validation Linear

```

yhat <- predict(linmodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_allF065)

```

```

cor(yhat$fit,dfacvF065.all$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.all$VO2max)^2)/nrow(dfacvF065.all))
yhat <- predict(linmodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_adultF065)
cor(yhat$fit,dfacvF065.adult$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.adult$VO2max)^2)/nrow(dfacvF065.adult))
yhat <- predict(linmodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_childF065)
cor(yhat$fit,dfacvF065.child$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.child$VO2max)^2)/nrow(dfacvF065.child))

```

#cross validation Poly

```

yhat <- predict(polmodF065minAIC,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_allF065)
cor(yhat$fit,dfacvF065.all$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.all$VO2max)^2)/nrow(dfacvF065.all))
yhat <- predict(polmodF065minAIC,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_adultF065)
cor(yhat$fit,dfacvF065.adult$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.adult$VO2max)^2)/nrow(dfacvF065.adult))
yhat <- predict(polmodF065minAIC,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_childF065)
cor(yhat$fit,dfacvF065.child$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.child$VO2max)^2)/nrow(dfacvF065.child))

```

#cross validation Gam

```

yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_allF065)
cor(yhat$fit,dfacvF065.all$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.all$VO2max)^2)/nrow(dfacvF065.all))
yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_adultF065)
cor(yhat$fit,dfacvF065.adult$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.adult$VO2max)^2)/nrow(dfacvF065.adult))
yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_childF065)
cor(yhat$fit,dfacvF065.child$VO2max)^2
sqrt(sum((yhat$fit-dfacvF065.child$VO2max)^2)/nrow(dfacvF065.child))

```

#Fitting on whole sample

```

dfa <- subset(PSVO2max_excl_missing_data_HR_RER_21018)
dfa065 <- subset(dfa,subset = Age<65.1)
females065 <- subset(dfa065,subset = Gender=="0")
males065 <- subset(dfa065,subset = Gender=="1")
dfa065$Gender <- with(dfa065,factor(ifelse(Gender == 1, "male", "female")))

```

```

yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_all065)
cor(yhat$fit,dfa065.all$VO2max)^2
sqrt(sum((yhat$fit-dfa065.all$VO2max)^2)/nrow(dfa065.all))
yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_volw065)

```

```

cor(yhat$fit,dfa065.volw$VO2max)^2
sqrt(sum((yhat$fit-dfa065.volw$VO2max)^2)/nrow(dfa065.volw))
yhat <- predict(gammodF065,type="response",
se.fit=TRUE,newdata=predictive_explanatory_variables_kinder065)
cor(yhat$fit,dfa065.kinder$VO2max)^2
sqrt(sum((yhat$fit-dfa065.kinder$VO2max)^2)/nrow(dfa065.kinder))

#comparison jones
YhatjoMale <- with(subset(dfaF065, Age>20),(-3.76 + 0.034*Length + 0.022*Weight -
0.028*Age)*1000)
YMale <- with(subset(dfaF065, Age>20), VO2max)
cor(YhatjoMale,YMale)^2
YhatjoMale <- with(subset(dfacvF065, Age>20),(-3.76 + 0.034*Length + 0.022*Weight -
0.028*Age)*1000)
YMale <- with(subset(dfacvF065, Age>20), VO2max)
cor(YhatjoMale,YMale)^2

#comparison harkel
YhatTHMale <- with(subset(dfaF065, Age<20.1),((0.66*Age) + 38.6)*1000)
YMale <- with(subset(dfaF065, Age<20.1), VO2max)
cor(YhatTHMale,YMale)^2
YhatTHMale <- with(subset(dfacvF065, Age<20.1),((0.66*Age) + 38.6)*1000)
YMale <- with(subset(dfacvF065, Age<20.1), VO2max)
cor(YhatTHMale,YMale)^2

# visualisatie GAM
plot_smooth(gammodF065, view="Age", plot_all = "Gender", rug = FALSE)
plot_smooth(gammodF065, view="Weight", plot_all = "Gender", rug = FALSE)
plot_smooth(gammodF065, view="Length", plot_all = "Gender", rug = FALSE)

format(round(482.4332, 2), nsmall = 2)

"PREDICTIONS"
x1 <- c(0,10,143,33.5)
x2 <- c(0,15,164,52 )
x3 <- c(0,20,168.8,63.2)
x4 <- c(0,30,169.3,68.5)
x5 <- c(0,40,169.3,70.3)
x6 <- c(0,50,167.7,70.5)
x7 <- c(0,60,166.6,71.6)
x8 <- c(1,10,143,34)
x9 <- c(1,15,168,53 )
x10 <- c(1,20,183.5,78.1)
x11 <- c(1,30,183.7,83.8)
x12 <- c(1,40,182.4,85.1)
x13 <- c(1,50,181.3,86.4)
x14 <- c(1,60,179.2,84.4)
predval <- data.frame(rbind(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14))
colnames(predval) <- c("Gender", "Age", "Length", "Weight")
predval
predval$Gender <- with(predval, as.factor(ifelse(Gender==0, "female", "male")))

```

```
p <- c(0.025,0.05,0.10,0.25,0.50,0.75,0.90,0.95,0.975)
quantile <- qnorm(p)
install.packages("mgcv")
x <- matrix(NA,14,9)
x
yhat <- predict(gammodF065,type="response", se.fit=TRUE,newdata=predval)
??qr.lm(object)
for (i in 1:14) {x[i,] <- yhat$fit[i] + yhat$se.fit[i]*quantile}
x
stats::qr.lm
```