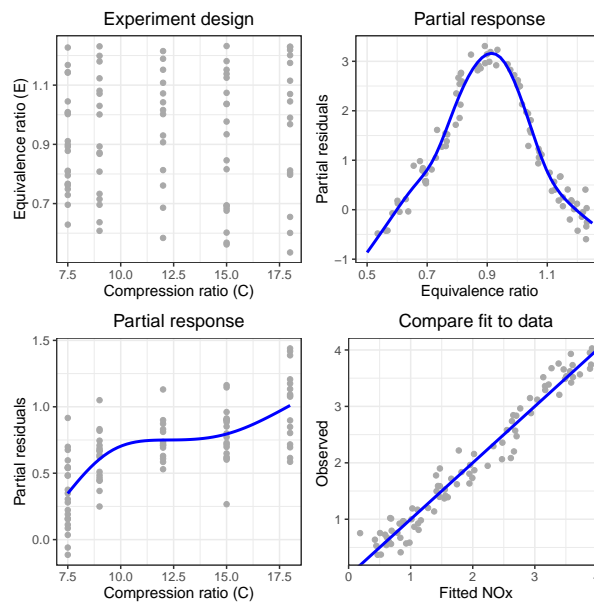


GAM smoothing (Ethanol data)



Fitting a generalized additive model (GAM) to the Ethanol data. Design points for compression ratio (C) and equivalence ratio (E) (top, left); the additive smooth components (top, right) and (bottom, left); observed vs. fitted values (bottom, right). R code in `f-ethanol-gam.R`

```
# GAM smoothing (Ethanol data)
# A graph in the book 'Practical Smoothing. The Joys of P-splines'
# Paul Eilers and Brian Marx, 2019

library(ggplot2)
library(SemiPar)
library(gridExtra)
library(JOPS)

# Get the data
data(ethanol)

# Basis matrix for compression ratio (C)
clo = 7.5
chi = 18
ngrid = 50
B1 = bbase(ethanol$C, clo, chi)

# Basis matrix for equivalence ratio (E)
elo = 0.5
ehi = 1.25
B2 = bbase(ethanol$E, elo, ehi)

# Penalty matrix
n = ncol(B1)
D = diff(diag(n), diff = 2)
r1 = 1:n
r2 = r1 + n
lambdas = c(1, 0.1)
P = kronecker(diag(lambdas), t(D) %*% D)
P = P + diag(2 * n) * 1e-06

# Fit the model
B = cbind(B1, B2)
```

```

y = ethanol$NOx
a = solve(t(B) %*% B + P, t(B) %*% y)
mu = B %*% a
cat('SD of residuals:', sd(y - mu), '\n')

# Add fitted components to data frame (for ggplot)
Fmod = ethanol
Fmod$f1 = B1 %*% a[r1]
Fmod$f2 = B2 %*% a[r2]

# Model fit on fine grid
cgrid = seq(clo, chi, length = ngrid)
Bg1 = bbase(cgrid, clo, chi)
egrid = seq(elo, ehi, length = ngrid)
Bg2 = bbase(egrid, elo, ehi)
zc = Bg1 %*% a[r1]
ze = Bg2 %*% a[r2]

# Fill data frames for ggplot
Fc = data.frame(cgrid, zc)
Fe = data.frame(egrid, ze)

# Create plots
plt1 = ggplot(aes(x = C, y = E), data = ethanol) +
  geom_point(color = "darkgrey") +
  xlab("Compression ratio (C)") +
  ylab("Equivalence ratio (E)") +
  ggtitle("Experiment design") +
  JOPS_theme()

plt2 = ggplot(aes(x = E, y = NOx - f1), data = Fmod) +
  geom_point(color = "darkgrey") +
  geom_line(aes(x = egrid, y = ze), data = Fe, size = 1, color = "blue") +
  xlab("Equivalence ratio") +
  ylab("Partial residuals") +
  ggtitle("Partial response") +
  JOPS_theme()

plt3 = ggplot(aes(x = C, y = NOx - f2), data = Fmod) +
  geom_point(color = "darkgrey") +
  geom_line(aes(x = cgrid, y = zc), data = Fc, color = "blue", size = 1) +
  xlab("Compression ratio (C)") +
  ylab("Partial residuals") +
  ggtitle("Partial response") +
  JOPS_theme()

plt4 = ggplot(aes(x = f1 + f2, y = NOx), data = Fmod) +
  geom_point(color = "darkgrey") +
  geom_abline(slope = 1, intercept = 0, color = "blue", size = 1) +
  xlab("Fitted NOx") +
  ylab("Observed ") +
  ggtitle("Compare fit to data") +
  JOPS_theme()

# Make and save pdf
grid.arrange(plt1, plt2, plt3, plt4, ncol = 2, nrow = 2)

```
