Introducing mgcViz

Matteo Fasiolo

Joint work with:

Simon N. Wood (University of Bristol, UK)
Yannig Goude (EDF R&D)
Raphaël Nedellec (Talend, formerly EDF R&D)

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/workshop_EDF19

Workshop plan

First session: basic mgcViz framework and tools

Talk + 30min hands-on session.

Second session: qgam and mgcViz

Talk + 30min hands-on session.

mgcViz: talk structure

The `mgcv` ecosystem:

- `gamm4` generalized additive mixed model using `lme4` for estimation
- `refund` functional GAMs using `mgcv` estimation
- `qgam` quantile GAMs (last session)
- `mgcViz` GAM visualization using `ggplot2` (next)
- and many others `scam`, `vagam`, `GJRM`, `itsadug`, ...

mgcViz: talk structure

mgcViz: why do we need it and how does it work?

- limitations of `mgcv` plotting methods
- `mgcViz` layer-based solution
- smooth effects plots
- diagnostic plots

For detailed account see

Fasiolo, M., R. Nedellec, Y. Goude, and S. N. Wood (2018). Scalable visualisation methods for modern generalized additive models. arXiv:1809.10632

or online training material

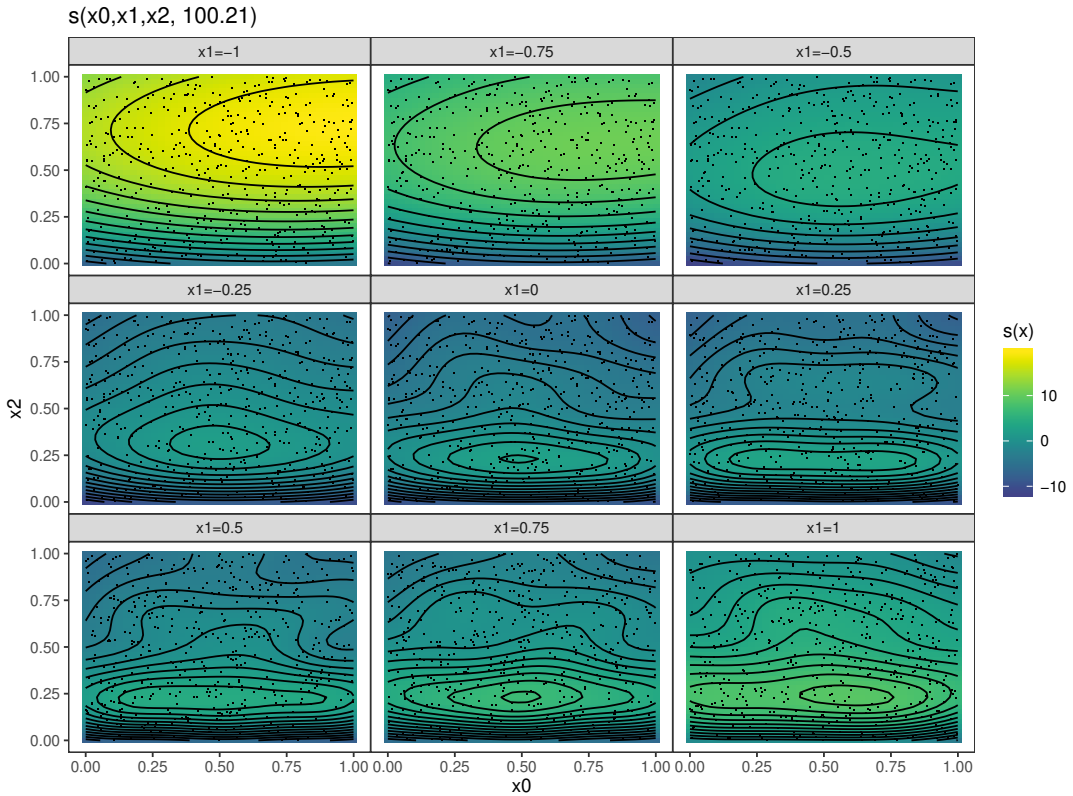
`https://mfasiolo.github.io/mgcViz/index.html`

mgcViz: why and how

Now we go to "mgcViz.html" for demonstration

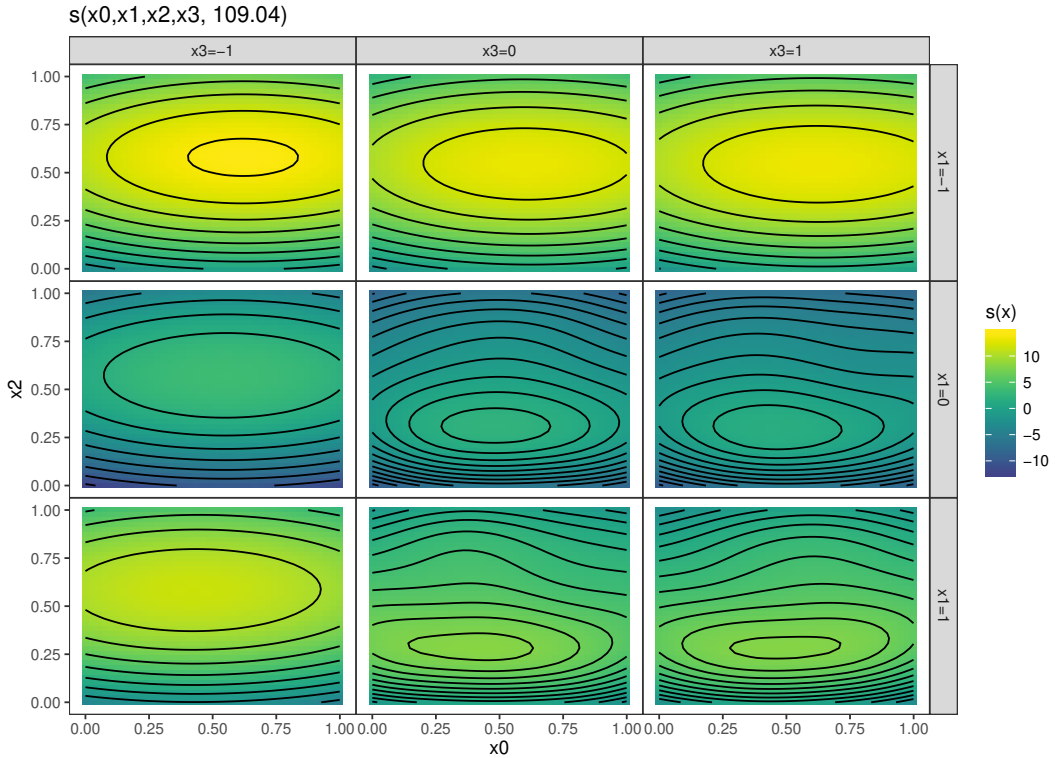
mgcViz: why and how

```
dat <- gamSim(1, n=5000, dist="normal", scale=2)
b <- bamV(y ~ s(x0, x1, x2), data = dat)
pl <- plotSlice(x = sm(b, 1),
                fix = list("x1" = seq(-1, 1, len = 9)))
pl + l_fitRaster() + l_fitContour() + l_points()
```



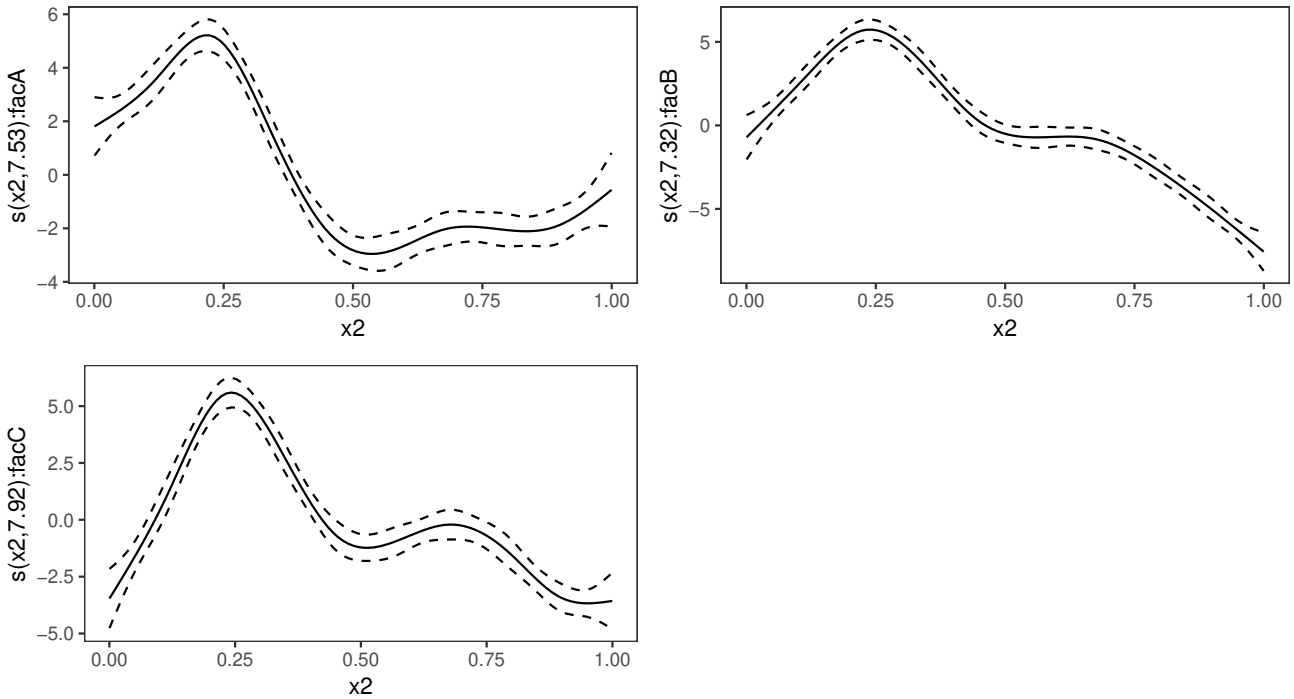
mgcViz: why and how

```
b <- bamV(y ~ s(x0, x1, x2, x3), data = dat)
pl <- plotSlice(x = sm(b, 1),
                fix = list("x1" = seq(-1, 1, len = 3),
                           "x3" = seq(-1, 1, len = 3)))
pl
```



mgcViz: why and how

```
dat <- gamSim(1,n=2000,dist="normal",scale=2)
dat$fac <- factor(sample(c("A","B","C"),nrow(dat),r=TRUE))
dat$y <- with(dat, dat$y + 30*(fac=="A")*(x2-0.5)^2 -
                           30*(fac=="B")*(x2-0.5)^3)
b <- gamV(y ~ s(x2, bs="cr", by = fac), data=dat)
print(plot(b), pages = 1)
```

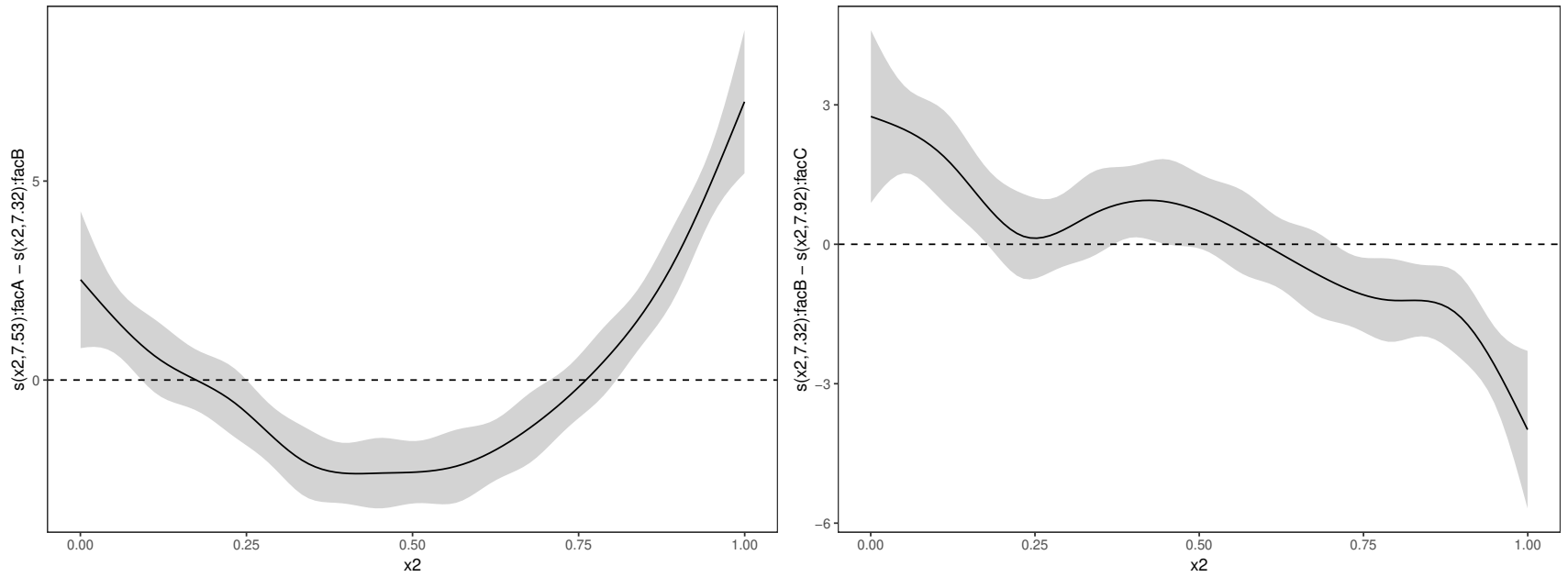


mgcViz: why and how

```
d1 <- plotDiff(s1=sm(b, 1), s2=sm(b, 2)) + l_ciPoly() +
      l_fitLine() + geom_hline(yintercept=0, linetype=2)

d2 <- plotDiff(s1=sm(b, 2), s2=sm(b, 3)) + l_ciPoly() +
      l_fitLine() + geom_hline(yintercept=0, linetype=2)

gridPrint(d1, d2, ncol = 2)
```



mgcViz: why and how

What if I want to plot the effects of the same model, estimated on different data?

We can use `plot.mgamViz`.

See "plot_mgamViz.html".

References I

Fasiolo, M., R. Nedellec, Y. Goude, and S. N. Wood (2018). Scalable visualisation methods for modern generalized additive models. *arXiv preprint arXiv:1809.10632*.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*, Volume 43. CRC Press.

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 54*(3), 507–554.

Wickham, H., W. Chang, et al. (2008). ggplot2: An implementation of the grammar of graphics. *R package version 0.7, URL: http://CRAN. R-project. org/package= ggplot2*.

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.