Simon Wood EPSRC established career fellow statistics

- Smooth regression models. e.g. model y as sum of smooth functions of x_j's + 'noise':

$$y_i = f_1(x_{1i}) + f_2(x_{2i}) + \ldots + \epsilon_i$$

- Questions:
    1. Best representation of f_j's?
    2. How smooth should the f_j's be?
    3. What terms to include?
    4. How certain are the estimates?
- Extensions:
    1. y_i binary, discrete, multivariate, categorical, survival time, heavy tailed, autocorrelated etc.
    2. f_j's multivariate, on the surface of a sphere, adaptive, random fields, random effects.
    3. Methods for 10^7+ data and 10^4+ coefficients, exploiting sparsity, parallelism etc.

Software (mgcv) - part of R
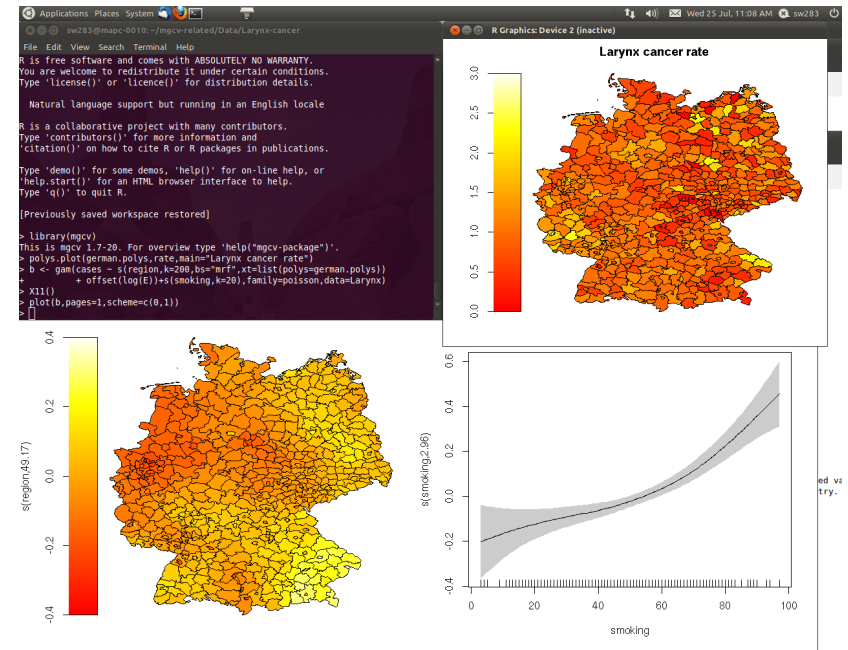
Applications with Matteo Fasiolo and students

- Electricité de France use the models and software to predict energy demand across France, we work with them on scalability and quantile regression.
- Quantitative Linguists in Tübingen and Groningen use the methods to analyse eye-tracking and EEG data from experiments on processing of written language. Again we collaborate on scalability and modelling of heavy tailed data.
- We collaborate on forest health modelling used to inform environmental and management policies in Germany and at the EU level, and with Fisheries scientists in CSIRO Australia, in particular on issues of modelling spatial correlation and spatial confounding.
- We developed scalable methods for modelling daily black smoke (pm10) from the UK monitoring network.