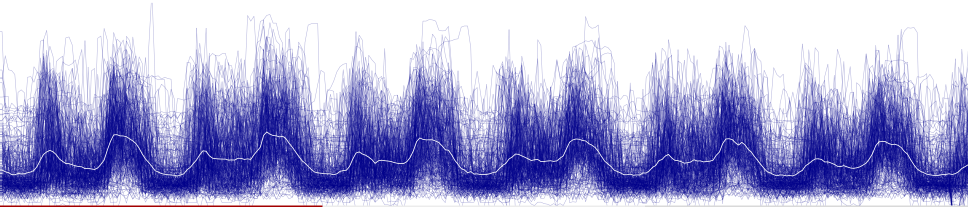
Electricity demand forecasting with Generalised Additive Models

Matteo Fasiolo

matteo.fasiolo@bristol.ac.uk

Material available at:

`https://github.com/mfasiolo/2025_GAM_Workshop_DS_Soc`

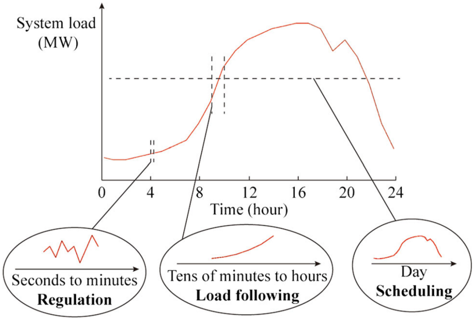


Electricity demand forecasting

Electric power system present some unique challenges:

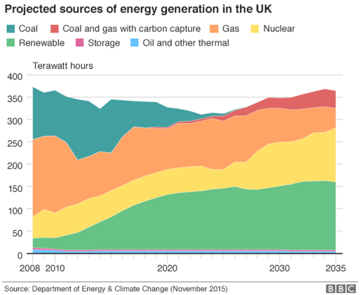
- very limited storage
- supply must meet demand in near-real time

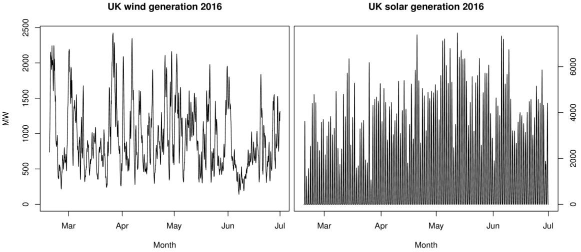
Short-term operations rely on accurate forecasts at several time scales:



Kroposki (2017)

Importance of forecasts increased by developments on **supply side**:





- increased renewable penetration → **more uncertainty**
- fossil fuel replaced by nuclear plants → **less flexibility**

Supply-side uncertainty & less tolerance to forecasting errors.

Demand-side and IT trends:

- emergence of prosumers

  $\rightarrow$ embedded solar & wind production

- smart meters in every home

  $\rightarrow$ handle larger and more complex data sets

- electric vehicles

  $\rightarrow$ higher load & uncertainty

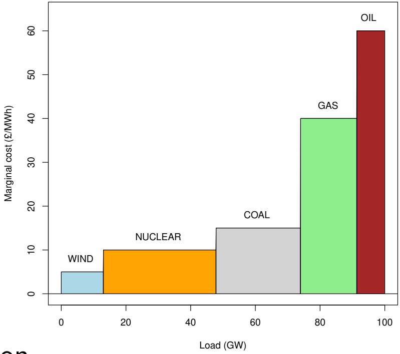
Automated grid management systems to avoid distribution grid overload.

Forecasts at multiple levels of aggregation will be key input.

Focus on day-ahead forecast of aggregate UK demand.

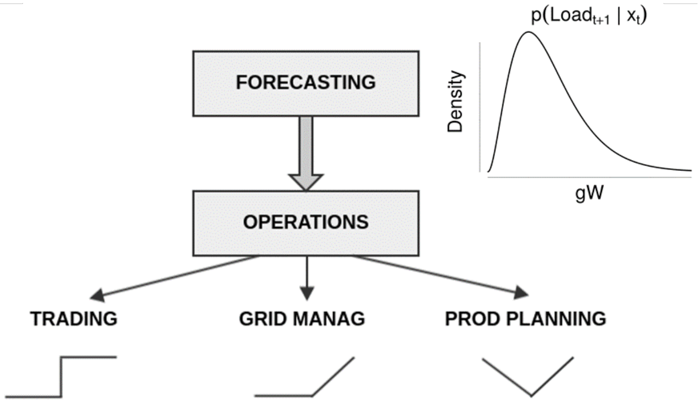
Forecast is main input for decision in operations:

- trading
- grid management
- production planning



Each has specific loss functions:

- regulation
  $\rightarrow$ sanctions for over/under production
- technological factors
  $\rightarrow$ availability of storage and generation cost



Optimization process:

$$D^* = \operatorname*{argmin}_D \mathbb{E}\{L(\text{Load}_{t+1}, D)\}.$$

What is an additive model

Regression setting:

- y is our response or dependent variable (demand here)
- $\mathbf{x}$ is a vector of covariates or independent variables

In **distributional regression** we want a good model for $\text{Dist}(y|\mathbf{x})$.

Model is $\text{Dist}_m\{y|\theta_1(\mathbf{x}), \theta_2, \dots, \theta_q\}$, where $\theta_1, \dots, \theta_q$ are param.

We assume that $\theta_2, \dots, \theta_q$ do not depend on $\mathbf{x}$.

Gaussian additive model:

$$y|\mathbf{x} \sim N\{y|\mu(\mathbf{x}), \sigma^2\},$$

where

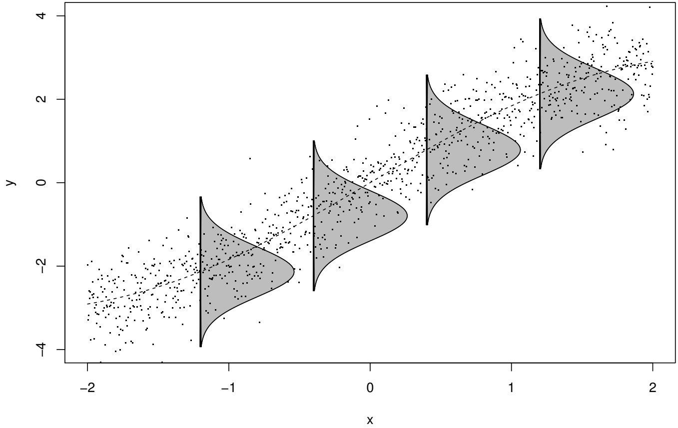
$$\mu(\mathbf{x}) = \sum_{j=1}^{m} f_j(\mathbf{x}),$$

and

$$\sigma^2 = \text{Var}(y|\mathbf{x}).$$

f_j's can be linear or non-linear (smooth) effects.

NB: we call $\sum_{j=1}^{m} f_j(\mathbf{x})$ **linear predictor** because it is linear in $\boldsymbol{\beta}$.



Gaussian model with variable mean.
In mgcv: `gam(y~s(x), family=gaussian)`.

Generalized additive model (GAM) (Hastie and Tibshirani, 1990):

$$y|\mathbf{x} \sim \text{Distr}\{y|\theta_1 = \mu(\mathbf{x}), \theta_2, \dots, \theta_p\},$$

where

$$g\{\mu(\mathbf{x})\} = \sum_{j=1}^{m} f_j(\mathbf{x}),$$

and g is a one-to-one function.

Poisson GAM:

- $y|\mathbf{x} \sim \text{Pois}\{y|\mu(\mathbf{x})\}$
- $\mu(\mathbf{x}) = \exp\left\{ \sum_{j=1}^{m} f_j(\mathbf{x}) \right\}$
- $g = \log$ assures $\mu(\mathbf{x}) > 0$

Recall GAM model structure:

$$y|\boldsymbol{x} \sim \text{Distr}\{y|\theta_1 = \mu(\boldsymbol{x}), \theta_2, \dots, \theta_p\}.$$

Multi-parameter GAM structure:

$$y|\boldsymbol{x} \sim \text{Distr}\{y|\theta_1 = \mu_1(\boldsymbol{x}), \theta_2 = \mu_2(\boldsymbol{x}), \dots, \theta_p = \mu_p(\boldsymbol{x})\},$$

where

$$g_1\{\mu_1(\boldsymbol{x})\} = \sum_{j=1}^{m} f_j^1(\boldsymbol{x}), \quad \dots \quad g_p\{\mu_p(\boldsymbol{x})\} = \sum_{j=1}^{m} f_j^p(\boldsymbol{x}).$$

Special case are **Generalized Additive Models for Location Scale and Shape** (GAMLSS) (Rigby and Stasinopoulos, 2005).

See appendix for complete list of distributions in `mgcv`.

Example: **Gaussian location-scale model**

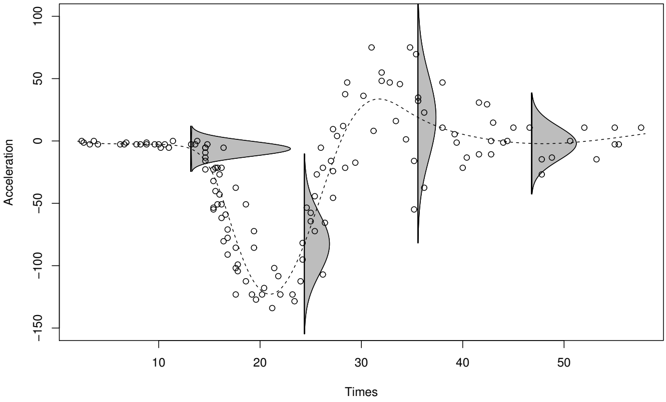
Model is

$$y|\boldsymbol{x} \sim N\{y|\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x})\}$$

where

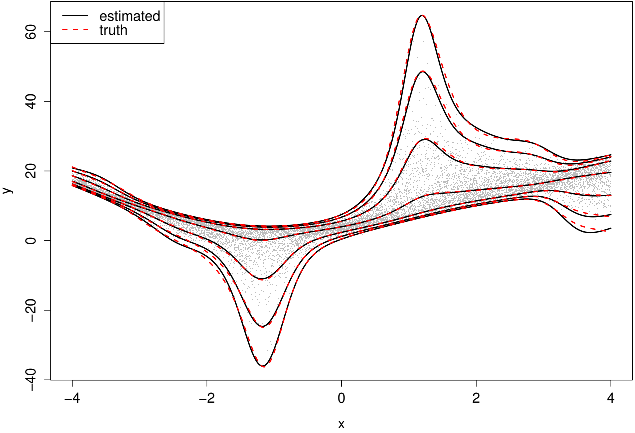
$$\mu(\boldsymbol{x}) = \sum_{j=1}^{m} f_j^1(\boldsymbol{x})$$

$$\log \sigma(\boldsymbol{x}) = \sum_{j=1}^{m} f_j^2(\boldsymbol{x})$$



In mgcv: `gam(list(y ~ s(x), ~ s(x)), family=gaulss)`.

Example: **Sinh-arcsinh (shash) distribution** (Jones and Pewsey, 2009)



`gam(list(y~s(x), ~s(x), ~s(x), ~s(x)), family=shash).`

Introducing smooth effects

Consider additive model

$$g\{\mu(\mathbf{x})\} = f_1(\mathbf{x}) + f_2(\mathbf{x}) + f_3(\mathbf{x}),$$

where

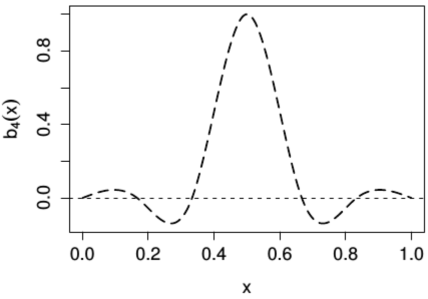
- $f_1(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$
- $f_2(\mathbf{x}) = \begin{cases} 0 & \text{if } x_2 = \text{FALSE} \\ \beta_4 & \text{if } x_2 = \text{TRUE} \end{cases}$
- $f_3(\mathbf{x}) = f_3(x_3)$ is a non-linear smooth function.

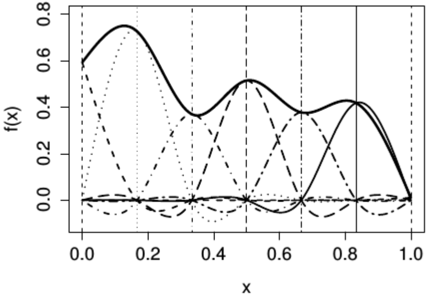
Smooth effects built using spline bases

$$f_3(x_3) = \sum_{k=1}^{r} \beta_k b_k(x_3)$$

where β_k are unknown coeff and $b_k(x_3)$ are known spline basis functions.

```
s(x, bs = "cr", k = 20)
```

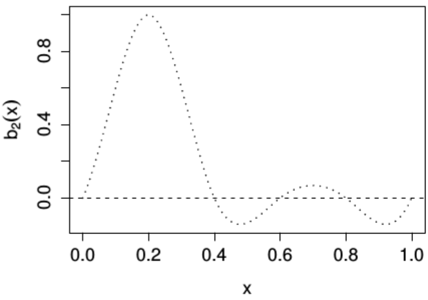


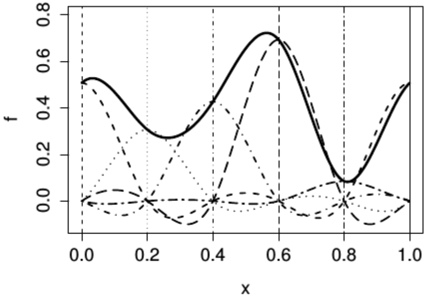


Cubic regression splines are related to the optimal solution to

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \gamma \int f''(x)^2\, dx.$$

`mgcv` offers many smooths (see `?smooth.terms`). `s(x, bs = "cc")`:

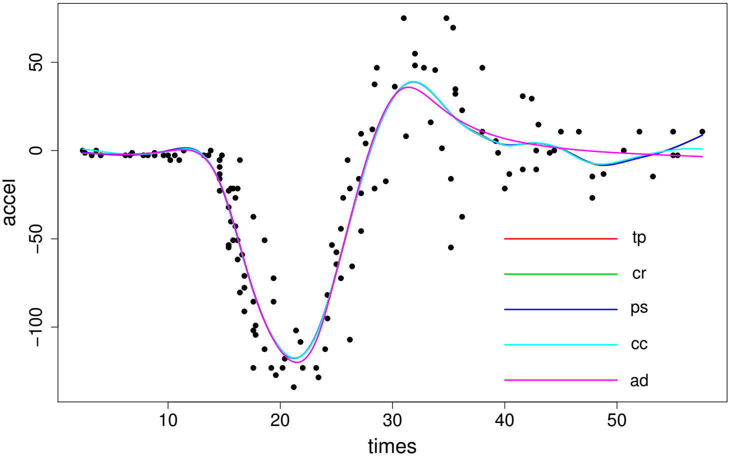




Cyclic cubic regression splines make so that

- $f(x_{min}) = f(x_{max})$
- $f'(x_{min}) = f'(x_{max})$
- $f''(x_{min}) = f''(x_{max})$

`s(x, bs = "ad")`



The wiggliness or smoothness of $f(x)$ depends on x.

Recall structure of smooth effects:

$$f(\mathbf{x}) = \sum_{j=1}^{k} \beta_j b_j(\mathbf{x}).$$

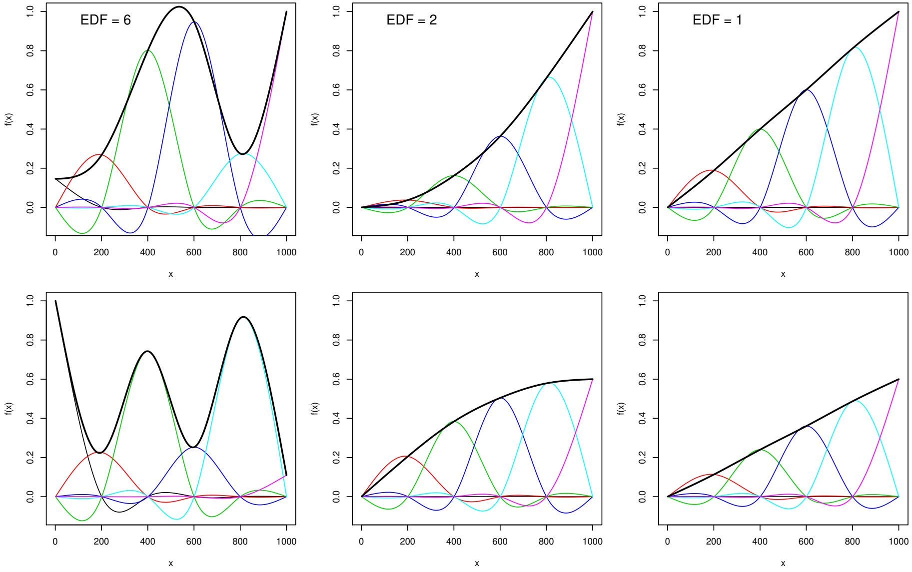
where $\boldsymbol{\beta}$ shrunk toward smoothness by penalty

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmax}} \Big\{ \overbrace{\log p(\mathbf{y}|\boldsymbol{\beta})}^{\text{goodness of fit}} - \underbrace{\text{Pen}(\boldsymbol{\beta}|\gamma)}_{\text{penalize complexity}} \Big\}$$

Exact k is unimportant, we choose it large and let penalty work.

Effective number of parameters we are using is $\leq k$.

Approximation is **Effective Degrees of Freedom** (EDF) $\leq k$.


EDF = 6
EDF = 2
EDF = 1
f(x)
x
0
200
400
600
800
1000
0.0
0.2
0.4
0.6
0.8
1.0

Learning more

Here we cover basic GAM and GAMLSS modelling with `mgcv` and `mgcViz` (Fasiolo et al., 2018).

`mgcv` is a recommended R package, included in R by default.

There are alternatives to `mgcv` for GAM modelling, such as:

- `mboost` (Hothorn et al., 2010)
- `gamlss` (Rigby and Stasinopoulos, 2005)
- `brms` (Bürkner et al., 2017)
- `BayesX` (Brezger et al., 2003)
- `INLA` (Rue et al., 2009)

The `mgcv` ecosystem is still expanding.

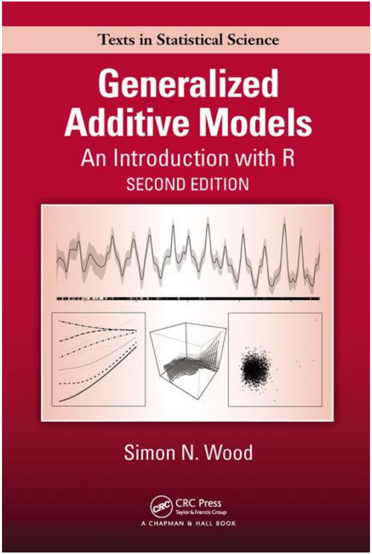
New developments are multivariate forecasting models with SCM package of Gioia et al. (2024):

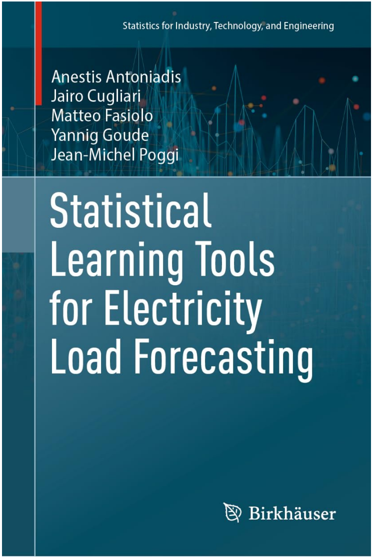
`https://github.com/VinGioia90/SCM`

Quantile GAMs of Fasiolo et al. (2018) can be used to forecast demand without assuming its distribution.

See the `qgam` package on CRAN.

See also the Big Data GAM methods of Wood et al. (2017) and `mgcv::bam`.


Texts in Statistical Science
Generalized Additive Models
An Introduction with R
SECOND EDITION
Simon N. Wood
CRC Press
A CHAPMAN & HALL BOOK


Statistics for Industry, Technology, and Engineering
Anestis Antoniadis
Jairo Cugliari
Matteo Fasiolo
Yannig Goude
Jean-Michel Poggi
Statistical Learning Tools for Electricity Load Forecasting
Birkhäuser

References I

Brezger, A., T. Kneib, and S. Lang (2003). Bayesx: Analysing bayesian structured additive regression models. Technical report, Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians .

Bürkner, P. C. et al. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software 80*(1), 1–28.

Fasiolo, M., Y. Goude, R. Nedellec, and S. N. Wood (2018). Fast calibrated additive quantile regression. *arXiv preprint arXiv:1707.03307*.

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

Gioia, V., M. Fasiolo, J. Browell, and R. Bellio (2024). Additive covariance matrix models: Modeling regional electricity net-demand in great britain. *Journal of the American Statistical Association*, 1–13.

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

Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner (2010). Model-based boosting 2.0. *The Journal of Machine Learning Research 11*, 2109–2113.

References II

Jones, M. and A. Pewsey (2009). Sinh-arcsinh distributions. *Biometrika 96*(4), 761–780.

Kroposki, B. (2017). Integrating high levels of variable renewable energy into electric power systems. *Journal of Modern Power Systems and Clean Energy 5*(6), 831–837.

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.

Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology) 71*(2), 319–392.

Wood, S. N., Z. Li, G. Shaddick, and N. H. Augustin (2017). Generalized additive models for gigadata: modeling the uk black smoke network daily data. *Journal of the American Statistical Association 112*(519), 1199–1210.

Wood, S. N., N. Pya, and B. Säfken (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association 111*(516), 1548–1575.

List of distributions in mgcv

Type `?mgvc::family` on the R console to see a list of distributions from the exponential family.

Wood et al. (2016) allows Distr $\notin$ exponential family (**extended GAMs**):

1. `scat` $\rightarrow$ scaled Student-t;
2. `betar` $\rightarrow$ beta for $y \in (0, 1)$;
3. `ziP` $\rightarrow$ zero-inflated Poisson;
4. `tw` $\rightarrow$ Tweedie;
5. `ocat` $\rightarrow$ order categorical;
6. `nb` $\rightarrow$ negative binomial.

Note that these EGAMs have one linear predictor, i.e.

$$y_i|\boldsymbol{x}_i \sim \text{Distr}\{y_i|\theta_1(\boldsymbol{x}), \theta_2, \dots, \theta_p\}.$$

hence are not GAMLSSs.

Available GAMLSS families:

1. `gammals` → 2-par gamma;
2. `gaulss` → 2-par Gaussian;
3. `shash` → 4-par sinh-arsinh;
4. `ziplss` → 2-par zero-inflated Poisson;
5. `gevlss` → 3-par generalised extreme value distribution (GEV);
6. `gumbls` → 2-par Gumbel (special case of GEV);
7. `twlss` → 3-par Tweedie.

Further models are:

1. `multinom` → multinomial categorical;
2. `cox.ph` → Cox Proportional Hazards model;
3. `mvn` → multivariate Gaussian model (fixed covariance).