

Beyond mean modelling: multi-parameter GAMs

Matteo Fasiolo

matteo.fasiolo@bristol.ac.uk

Material available at:

https://github.com/mfasiolo/GAM_Workshop_Dortmund_25

These slides cover:

- 1 Multi-parameter GAMs and GAMLSSs
- 2 Multivariate Gaussian GAMs
- 3 Stacking and aggregation of experts

Recall GAM model structure:

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

where

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

Multi-parameter GAM structure (Wood et al., 2016):

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

where

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

The role of $\theta_1, \dots, \theta_p$ determines the type of model.

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

Example: **Gaussian location-scale model**

Model is

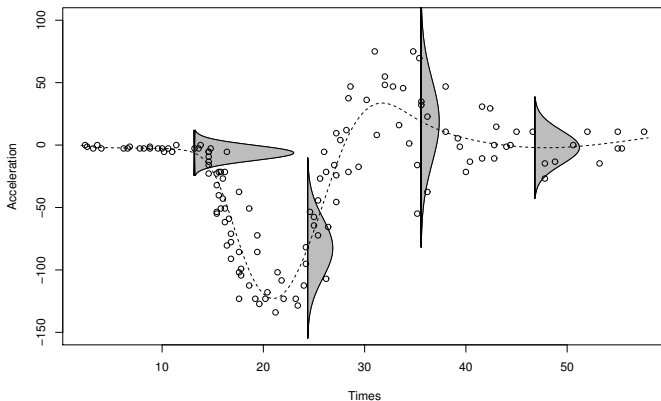
$$y|\mathbf{x} \sim N\{y|\mu(\mathbf{x}), \sigma^2(\mathbf{x})\}$$

where

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

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

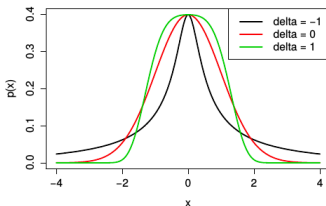
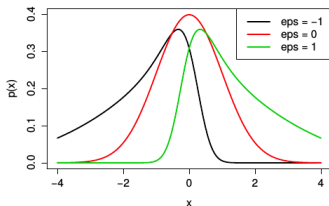
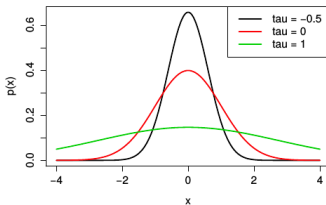
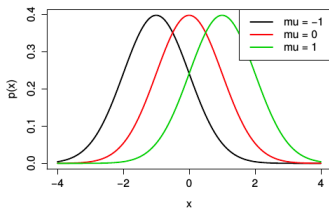
See appendix for complete list of distributions in `mgcv`.

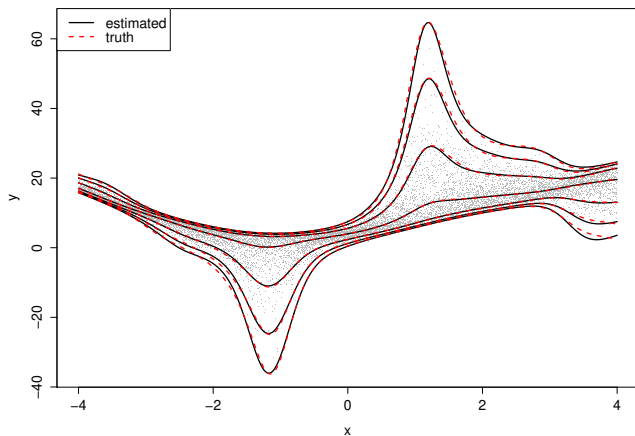


In mgcv: `gam(list(y ~ s(x), ~ s(x)), family=gaulss).`

Example: Sinh-arcsinh (shash) distribution

Four parameter distribution where location, scale, skewness (asymmetry) and kurtosis (tail behaviour) can depend on x (Jones and Pewsey, 2009).





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

Let's look at [R_demos/3_multi_gams.html](#)

These slides cover:

- 1 Multi-parameter GAMs and GAMLSSs
- 2 Multivariate Gaussian GAMs
- 3 Stacking and aggregation of experts

Non-GAMLSS example: **multivariate normal GAMs**

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_d \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1(\mathbf{x}) \\ \mu_2(\mathbf{x}) \\ \mu_3(\mathbf{x}) \\ \vdots \\ \mu_d(\mathbf{x}) \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \cdots & \Sigma_{1d} \\ \Sigma_{12} & \Sigma_{22} & \cdots & \cdots & \Sigma_{2d} \\ \Sigma_{13} & \Sigma_{23} & \cdots & \cdots & \Sigma_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1d} & \Sigma_{2d} & \cdots & \cdots & \Sigma_{dd} \end{pmatrix} \right).$$

In `mgcv` you can do:

```
fit <- gam(list(y1 ~ s(x1),
               y2 ~ s(x3),
               y3 ~ s(x3)),
           family = mvn(3))
```

With the SCM package we can model Σ as well

$$\mathbf{y} \sim N(\boldsymbol{\mu}(\mathbf{x}), \Sigma(\mathbf{x})).$$

$\Sigma(\mathbf{x})$ must be positive definite so we can **not** write $\Sigma_{jk} = \sum_j f_j(\mathbf{x})$.

One option is the modified Cholesky decomposition (Pourahmadi, 1999)

$$\Sigma^{-1} = \mathbf{T}^\top \mathbf{D}^{-2} \mathbf{T},$$

where \mathbf{D}^2 is a diagonal matrix and \mathbf{T} is upper triangular.

Why the MCD?

If $\mathbf{r} = \mathbf{y} - \boldsymbol{\mu}$ then $\text{cov}(\mathbf{r}) = \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1} = \mathbf{T}^\top \mathbf{D}^{-2} \mathbf{T}$.

Consider simulated residuals s.t. $\text{var}(\tilde{r}_1) = D_{1,1}^2$ and

$$\tilde{r}_l = - \sum_{k=1}^{l-1} T_{l,k} \tilde{r}_k + \epsilon_l, \quad \text{var}(\epsilon_l) = D_{l,l}^2,$$

where

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ T_{2,1} & 1 & 0 & \cdots & 0 \\ T_{3,1} & T_{3,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_{14,1} & T_{14,2} & \cdots & T_{14,13} & 1 \end{pmatrix},$$

then $\text{cov}(\tilde{\mathbf{r}}) = \boldsymbol{\Sigma}$.

Available via the SCM package (Gioia et al., 2025):

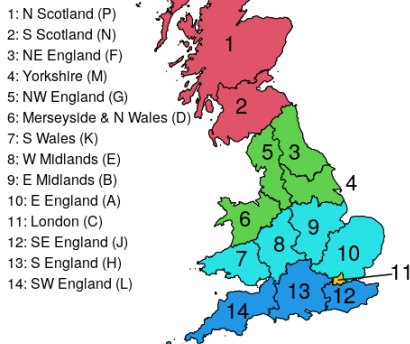
<https://github.com/VinGioia90/SCM>:

To install type:

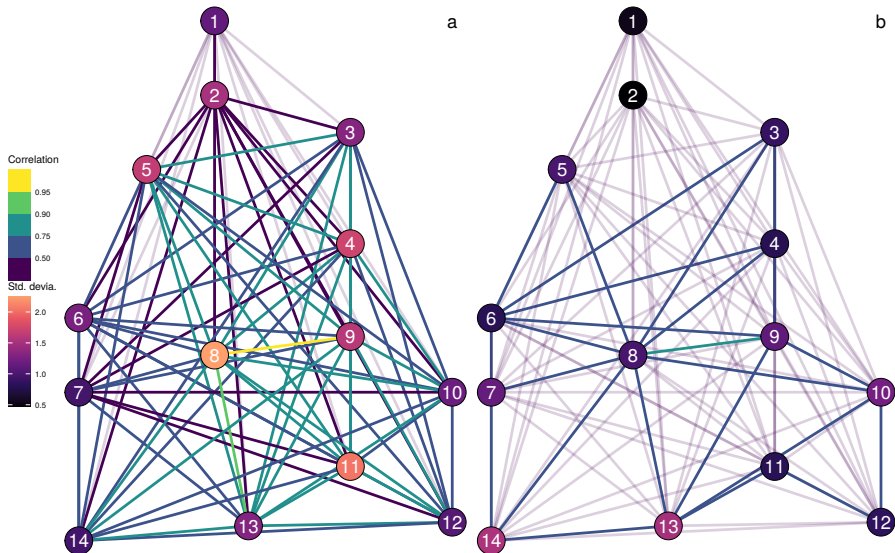
```
library(devtools)
install_github("VinGioia90/SCM")
```

See Gioia et al. (2024) for details.

The UK electricity grid is divided into 14 grid supply groups (GSP).



Gioia et al. (2024) produce **joint** probabilistic forecasts of net-demand.

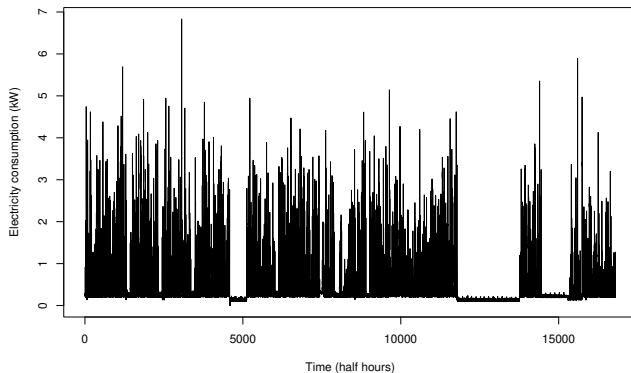


Left: 7am 31/12/2018. Right: midnight 20/08/2018.

Now continue example on `3_multi_gams.html`.

These slides cover:

- 1 Multi-parameter GAMs and GAMLSSs
- 2 Multivariate Gaussian GAMs
- 3 Stacking and aggregation of experts

Non-GAMLSS example: **additive stacking** or **aggregation of experts**

Two models for log-demand at time t or day d :

- ① $\log \text{dem}_{td} \sim N(\mu = f(\text{time_of_day}_t), \sigma^2)$
- ② $\log \text{dem}_{td} \sim N(\mu = \text{mean}(\log \mathbf{dem}_{d-1}), \sigma^2 = \text{var}(\log \mathbf{dem}_{d-1}))$

where $\mathbf{dem}_d = \{\log \text{dem}_{1d}, \dots, \log \text{dem}_{48d}\}$.

We want to predict $y|\mathbf{x}$ and we have models $p_1(y|\mathbf{x}), \dots, p_K(y|\mathbf{x})$.

Build mixture with covariate-dependent weights

$$p_{\text{mix}}(y|\mathbf{x}) = \sum_{k=1}^K w_k(\mathbf{x}) p_k(y|\mathbf{x});$$

where $0 \leq w_k \leq 1$ and $\sum_k w_k = 1$.

Capezza et al. (2021) use the multinomial parametrisation

$$w_k = \frac{e^{\eta_k}}{\sum_j e^{\eta_j}},$$

with $\eta_1 = 0$ and $\eta_k = \sum_l f_{lk}(\mathbf{x})$ for $k = 2, \dots, K$.

Available via the gamFactory:

```
install_github("mfasiolo/gamFactory")
```

For more advanced methods see gamstackr package:

<https://github.com/eenticott/gamstackr>

by Euan Enticott.

Now continue example on 3_multi_gams.html.

For related approaches see Yao et al. (2022) and Rügamer et al. (2022).

References I

- Capezza, C., B. Palumbo, Y. Goude, S. N. Wood, and M. Fasiolo (2021). Additive stacking for disaggregate electricity demand forecasting. *The Annals of Applied Statistics* **15**(2), 727–746.
- Gioia, V., M. Fasiolo, R. Bellio, and S. N. Wood (2025). Scalable fitting methods for multivariate gaussian additive models with covariate-dependent covariance matrices. *arXiv preprint arXiv:2504.03368*.
- 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.
- Jones, M. and A. Pewsey (2009). Sinh-arcsinh distributions. *Biometrika* **96**(4), 761–780.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86**(3), 677–690.
- 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.

References II

- Rügamer, D., F. Pfisterer, B. Bischl, and B. Grün (2022). Mixture of experts distributional regression: Implementation using robust estimation with adaptive first-order methods. *arXiv preprint arXiv:2211.09875*.
- 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.
- Yao, Y., G. Pirš, A. Vehtari, and A. Gelman (2022). Bayesian hierarchical stacking: Some models are (somewhere) useful. *Bayesian Analysis* 17(4), 1043–1071.

List of distributions in mgcv

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

Wood et al. (2016) allows $\text{Distr} \not\in$ 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.

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

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 `gumb1s` → 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).