Forecasting II

Generalized Additive Models

Valérie Chavez-Demoulin

April 14, 2024

GAM

In classical regression analysis, a popular and flexible set of models is the class of generalized additive models which link the mean behaviour of a random variable Y with a set of covariates $\mathbf{X} \in \mathbb{R}^q$ through

$$\mathrm{E}\left(\mathrm{Y} \mid \mathbf{X} = \mathbf{x}\right) = \mathrm{g}\left\{\mathbf{u}^{\mathrm{T}}\beta + \sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{h}_{\mathrm{k}}(\mathrm{t}_{\mathrm{k}})\right\}, \tag{1}$$

where

- g is a link function,
- $(u_1, \ldots, u_s)$ and $(t_1, \ldots, t_K)$ are subsets of $\{x_1, \ldots, x_q\}$,
- $\beta \in \mathbb{R}^s$ is a vector of parameters, and
- $h_k : \mathbb{T}_k \to \mathbb{R}$ are smooth functions supported on closed $\mathbb{T}_k \subset \mathbb{R}$, for all k.

Smooth functions and R package

We assume that each smooth function $h_k \in \mathcal{C}^2(\mathbb{T}_k)$ admits a finite m_k-dimensional basis parametrized by $\mathbf{h}_k = (h_{k,1}, \ldots, h_{k,m_k})^T \in \mathbb{R}^{m_k}$ and a quadratic penalty representation $\int_{\mathbb{T}_k} h_k''(t)^2\, dt = \mathbf{h}_k^T S_k \mathbf{h}_k$, where S_k is a uniquely determined symmetric matrix.

The class of C^2 smoothers with finite quadratic penalty representation is broad and encompasses, among many flexible smoothers, the natural cubic splines, the tensor product splines, and the cyclic cubic splines which are all included in the `R` package `mgcv`

Penalized log-likelihood

- Considering a sample of n observations $\{y_i, \mathbf{x}_i\}_{i=1}^n$, such models are estimated by maximizing a penalized log-likelihood

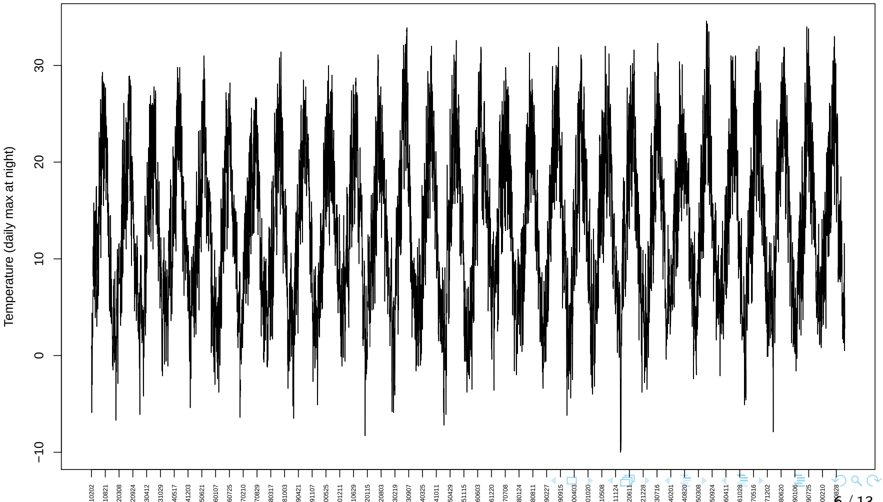
$$\ell_p(\boldsymbol{\theta}, \gamma) = \ell(\boldsymbol{\theta}) - \frac{1}{2}\sum_{k=1}^{K} \gamma_k \int_{\mathcal{A}_k} h_k''(t_k)^2 dt_k = \ell(\boldsymbol{\theta}) - \frac{1}{2}\sum_{k=1}^{K} \gamma_k \mathbf{h}_k^T S_k \mathbf{h}_k,$$

where $\boldsymbol{\theta}$ is the set of parameters of the model to be estimated and the first term $\ell(\boldsymbol{\theta})$ is the log-likelihood of the distributed random variables $\{y_i, \mathbf{x}_i\}_{i=1}^n$.

Smoothing parameters

- The penalty term controls the roughness of the smoothers through a vector of smoothing parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$ with higher values yielding smoother curves.
- The related effective degrees of freedom of each smooth function h_k are then defined as $\mathrm{tr}\,(\mathrm{I} + \gamma_\mathrm{k}\mathrm{S}_\mathrm{k})$.
- The smoothing parameters are chosen based on the Akaike Information Criterion (AIC).
- Given $\boldsymbol{\gamma}$, the penalized log-likelihood is maximized using an iterative weighted least squares procedure based on a Newton–Raphson algorithm.

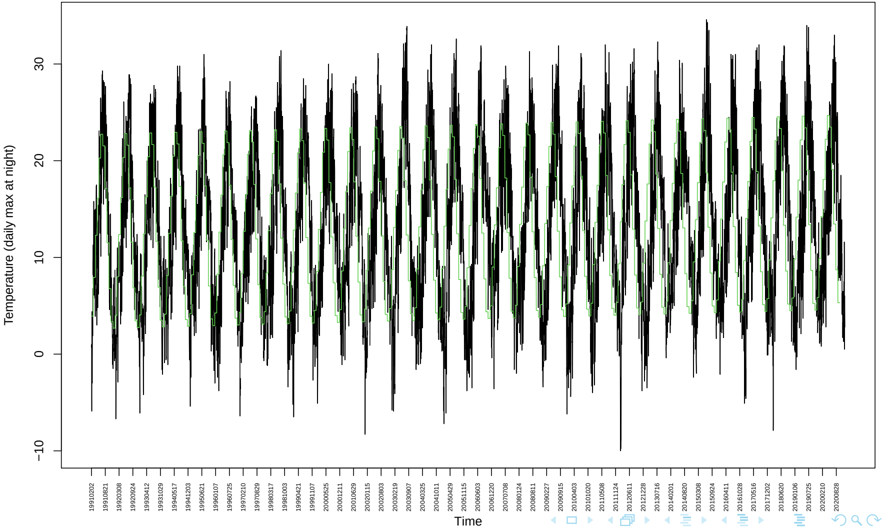
GAM in practice
Lausanne temperature data



First model: Smoothing months + linear model for year

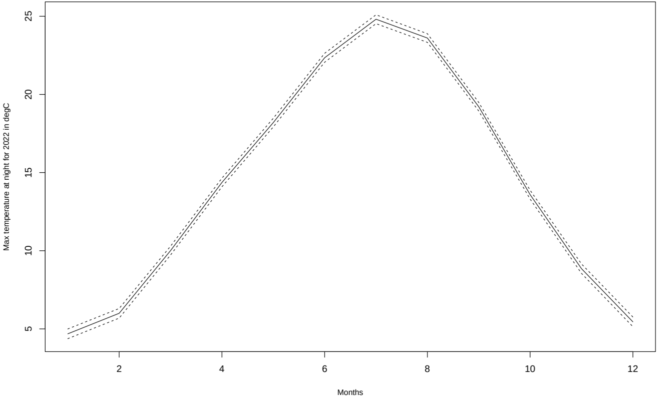
```
Family: gaussian
Link function: identity

Formula:
y ~ s(months) + years

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.209e+02  9.231e+00  -13.10   <2e-16 ***
years        6.687e-02  4.602e-03   14.53   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
            edf Ref.df    F p-value
s(months) 8.694  8.973 3489  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.744   Deviance explained = 74.4%
GCV =  17.05  Scale est. = 17.033    n = 10854
```

Fitted values



Prediction for 2023

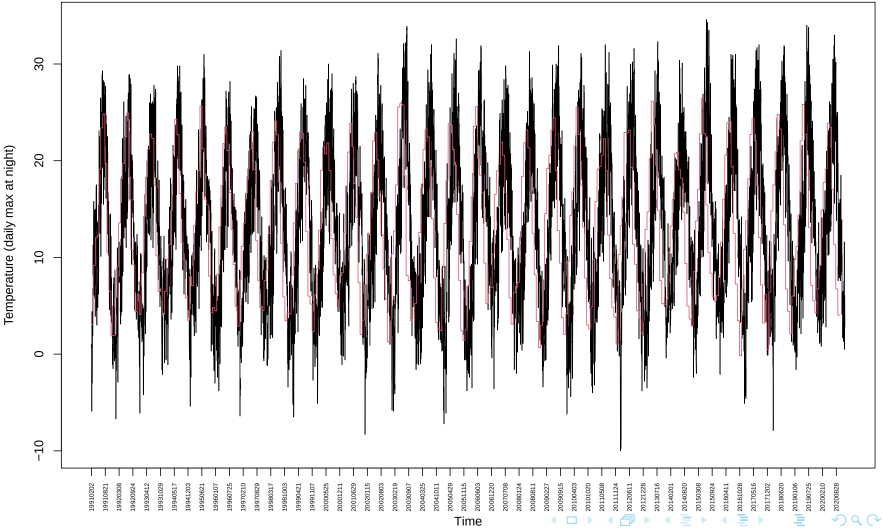
```
newdata <- data.frame("years"=rep(2022,12),"months"=1:12)
pred1 <- predict.gam(fit.gam1,newdata,se=T)
```



Second model: Smoothing months by year

```
Family: gaussian
Link function: identity

Formula:
y ~ s(months, by = as.factor(years))

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.25489    0.03791   349.6   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                    edf Ref.df      F p-value
s(months):as.factor(years)1991 8.889  8.996 133.41  <2e-16 ***
s(months):as.factor(years)1992 8.563  8.945 132.51  <2e-16 ***
s(months):as.factor(years)1993 7.933  8.705 123.59  <2e-16 ***
s(months):as.factor(years)1994 8.048  8.762 115.59  <2e-16 ***
...
s(months):as.factor(years)2015 8.290  8.863 154.78  <2e-16 ***
s(months):as.factor(years)2016 6.635  7.766 158.06  <2e-16 ***
s(months):as.factor(years)2017 6.872  7.972 178.44  <2e-16 ***
s(months):as.factor(years)2018 8.587  8.952 161.48  <2e-16 ***
s(months):as.factor(years)2019 8.257  8.851 147.42  <2e-16 ***
s(months):as.factor(years)2020 8.403  8.902 126.72  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.772   Deviance explained = 77.7%
GCV = 15.529  Scale est. = 15.181    n = 10854
```

Fitted values



Functional dependence of months per years