Online appendix for "The Functional Linear Array Model"

Sarah Brockhaus [1], Fabian Scheipl [1], Torsten Hothorn [2], and Sonja Greven [1]

[1] Institut für Statistik, Ludwig-Maximilians-Universität München, Germany
[2] Institut für Epidemiologie, Biostatistik und Prävention, Abteilung Biostatistik Universität Zürich, Switzerland

---

Address for correspondence: Sarah Brockhaus, Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, D – 80539 München, Germany.
E-mail: sarah.brockhaus@stat.uni-muenchen.de.
Phone: (+49) 89 2180 2248.
Fax: (+49) 89 2180 5308.

Abstract: In the online appendix we show how the identifiability constraints are enforced by using suitable transformations of the design matrix. Then we give an example of data simulated for the simulation study in the paper and we show results on run-time. As an application for function-on-function regression the Canadian climate data is analyzed, regressing precipitation curves on temperature curves and climatic regions, incorporating spatially dependent residual curves. In order to compare the framework of FLAMs to existing models we compare results for the Canadian weather data with results from Scheipl et al. (2014). In addition, a table summarizing characteristics of the FLAM, the penalized likelihood approach based on mixed models (Scheipl et al., 2014) and the Bayesian wavelet approach (Meyer et al., 2013) is given.

Key words: boosting; functional data analysis; smoothing; structured additive regression; varying coefficient models

A   Identifiability constraint

Consider a model $\boldsymbol{\xi}(Y_i(t)) = \beta_0(t) + h_j(x)(t)$, with smooth intercept $\beta_0(t)$ and an effect $h_j(x)(t)$ that contains an intercept $\beta_0(t)$ as special case. For the effects in Table 1, this is the case for the smooth effect $\gamma(z,t)$ and smooth interaction $f(z_1, z_2, t)$, the group-specific intercept $b_g(t)$ and smooth residual $e_i(t)$. The problem is that such a model is not identifiable as

$$\boldsymbol{\xi}(Y_i(t)) = \{\beta_0(t) + \bar{h}_j(x)(t)\} + \{h_j(x)(t) - \bar{h}_j(x)(t)\} = \tilde{\beta}_0(t) + \tilde{h}_j(x)(t)$$

yields the same fit with a different parametrization for $\bar{h}_j(x)(t) = E_X(h_j(X)(t))$, or replacing the expectation by the mean for concrete data, for $\bar{h}_j(x)(t) = N^{-1}\sum_i h_j(x_i)(t)$. Scheipl et al. (2014) pointed out that standard sum-to-zero constraints $\sum_{i,t} h_j(x_i)(t) = 0$ are not suitable for regression models with functional response. A suitable constraint is that the mean effect of each covariate should be zero in each point t:

$$N^{-1}\sum_i h_j(x_i)(t) = 0 \ \forall t.$$

We now show how to embed this constraint within the array framework of the FLAM. We define $\boldsymbol{B}$ as the $NG \times K_Y K_j$ design matrix with rows $\left(\boldsymbol{b}_j(x_i)^\top \otimes \boldsymbol{b}_Y(t_g)^\top\right)$ defined as in equation (2.2). $\boldsymbol{B}$ is the tensor product of the two marginal design matrices $\boldsymbol{B} = \boldsymbol{B}_j \otimes \boldsymbol{B}_Y$, with $\boldsymbol{B}_j$ having $\boldsymbol{b}_j(x_i)^\top$ as rows and $\boldsymbol{B}_Y$ having $\boldsymbol{b}_Y(t_g)^\top$ as rows. In this notation the response would be concatenated to a $1 \times NG$ vector $(Y_1(t_1), \ldots, Y_N(t_G))^\top$. Then the sum-to-zero-constraint over t can be represented as a linear constraint on the coefficient vector by enforcing $\boldsymbol{C\theta} = \boldsymbol{0}$, with $\boldsymbol{C} = (\boldsymbol{1}_N^\top \otimes \boldsymbol{I}_G)\boldsymbol{B}$, where $\boldsymbol{1}_N$ is the vector of length N containing ones and $\boldsymbol{I}_G$ is the G-dimensional identity matrix. Wood (2006, sec. 1.8.1) implements linear constraints by rewriting the model in terms of $K_Y(K_j-1)$ unconstrained parameters through a change of basis for the design matrix $\boldsymbol{B}$. For this the full QR decomposition of $\boldsymbol{C}^\top$ is needed:

$$\boldsymbol{C}^\top = [\boldsymbol{Q} : \boldsymbol{Z}]\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{QR},$$

where $[\boldsymbol{Q} : \boldsymbol{Z}]$ forms a $K_Y K_j \times K_Y K_j$ orthonormal matrix and $\boldsymbol{R}$ is a $G \times G$ upper triangular matrix. The transformed design matrix is obtained by the multiplication of the original design matrix with the transformation matrix $\boldsymbol{Z}$ yielding $\boldsymbol{BZ}$.

To apply this method for implementing a linear constraint in a FLAM, it is necessary to do the transformation on the marginal design matrices $\boldsymbol{B}_Y$ and $\boldsymbol{B}_j$. Therefore we rewrite $\boldsymbol{C}^\top$ depending on the marginal bases as $\boldsymbol{C}^\top = ((\boldsymbol{1}_N^\top \otimes \boldsymbol{I}_G)(\boldsymbol{B}_j \otimes \boldsymbol{B}_Y))^\top = (\boldsymbol{B}_j^\top \boldsymbol{1}_N) \otimes \boldsymbol{B}_Y^\top$ and use the QR decompositions of $\boldsymbol{B}_j^\top \boldsymbol{1}_N$ and $\boldsymbol{B}_Y^\top$ whose components are indexed by j and Y respectively:

$$\boldsymbol{C}^\top = \left([\boldsymbol{Q}_j : \boldsymbol{Z}_j]\begin{bmatrix} \boldsymbol{R}_j \\ \boldsymbol{0} \end{bmatrix}\right) \otimes \left([\boldsymbol{Q}_Y : \boldsymbol{Z}_Y]\begin{bmatrix} \boldsymbol{R}_Y \\ \boldsymbol{0} \end{bmatrix}\right)$$
$$= ([\boldsymbol{Q}_j : \boldsymbol{Z}_j] \otimes [\boldsymbol{Q}_Y : \boldsymbol{Z}_Y])\left(\begin{bmatrix} \boldsymbol{R}_j \\ \boldsymbol{0} \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{R}_Y \\ \boldsymbol{0} \end{bmatrix}\right) = [\boldsymbol{Q} : \boldsymbol{Z}]\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}.$$

Thus we can calculate the transformation matrix $\boldsymbol{Z}$ as $\boldsymbol{Z} = (\boldsymbol{Z}_j \otimes [\boldsymbol{Q}_Y : \boldsymbol{Z}_Y])$ and we obtain the transformed design matrix as

$$\boldsymbol{BZ} = (\boldsymbol{B}_j \otimes \boldsymbol{B}_Y)(\boldsymbol{Z}_j \otimes [\boldsymbol{Q}_Y : \boldsymbol{Z}_Y]) = (\boldsymbol{B}_j \boldsymbol{Z}_j) \otimes (\boldsymbol{B}_Y[\boldsymbol{Q}_Y : \boldsymbol{Z}_Y]).$$

As multiplication by the orthonormal matrix $[\boldsymbol{Q}_Y : \boldsymbol{Z}_Y]$ only rotates the basis, this rotation can be omitted. Thus only the transformation by $\boldsymbol{Z}_j$ is necessary and it suffices to compute the QR decomposition of $\boldsymbol{B}_j^\top \boldsymbol{1}_N \in \mathbb{R}^{K_j \times 1}$. It is not necessary to

compute the QR decomposition of the complete design matrix $\boldsymbol{B}$ or even to construct $\boldsymbol{B}$ explicitly. A basis transformation of the design matrix, $\boldsymbol{B}_j\boldsymbol{Z}_j$, involves that the penalty matrix $\boldsymbol{P}_j$ has to be transformed accordingly to $\boldsymbol{Z}_j^\top \boldsymbol{P}_j \boldsymbol{Z}_j$.

Please note that while unrelated to identifiability, centering the covariates by subtracting their mean (function) can in some cases additionally lend itself to nice interpretations of the intercept as the overall mean.

B Simulation details: data example and run-time overview

Figure 1 shows the coefficient functions and simulated responses for a setting with $N = 100$ observations, $G = 30$ grid points per trajectory and a signal-to-noise-ration of 2.

In Figure 2 the computation times of the model fits in the simulation are given. The optimal stopping iteration for FLAM is determined by 10-fold bootstrap over curves, which was parallelized on 10 cores of a 64-bit linux platform. For each model fit the optimal m_{stop} was searched on a grid up to a maximum of 2000. One sees clearly that FLAM scales better for a growing number of observations than the PFFR algorithm.

C Function-on-function regression: Canadian weather data

The Canadian weather data is a well known functional data example (Ramsay and Silverman, 2006). The data contains monthly temperature and precipitation at 35 different locations in Canada averaged over 1960 to 1994, see Figure 3. The weather stations are assigned to four climatic zones (Atlantic, Pacific, Continental, Arctic) and for each weather station the latitude and the longitude are given. The goal is to look at the association between precipitation and temperature curves, taking into account the climatic zones and the locations of the weather stations. As the precipitation and temperature curves are averaged over several years they are no time series but typical profiles and models relating precipitation to temperature thus do not have the problem of the future influencing the past. For the same reason values in the end and the beginning of a year should be similar.

We use this example to compare the results obtained by boosting with those of the PFFR method (Scheipl et al., 2014). In the online appendix, Scheipl et al. (2014) use the logarithm of precipitation as response variable and fit amongst others a model with smooth effects of the four climatic zones, functional effect of temperature and
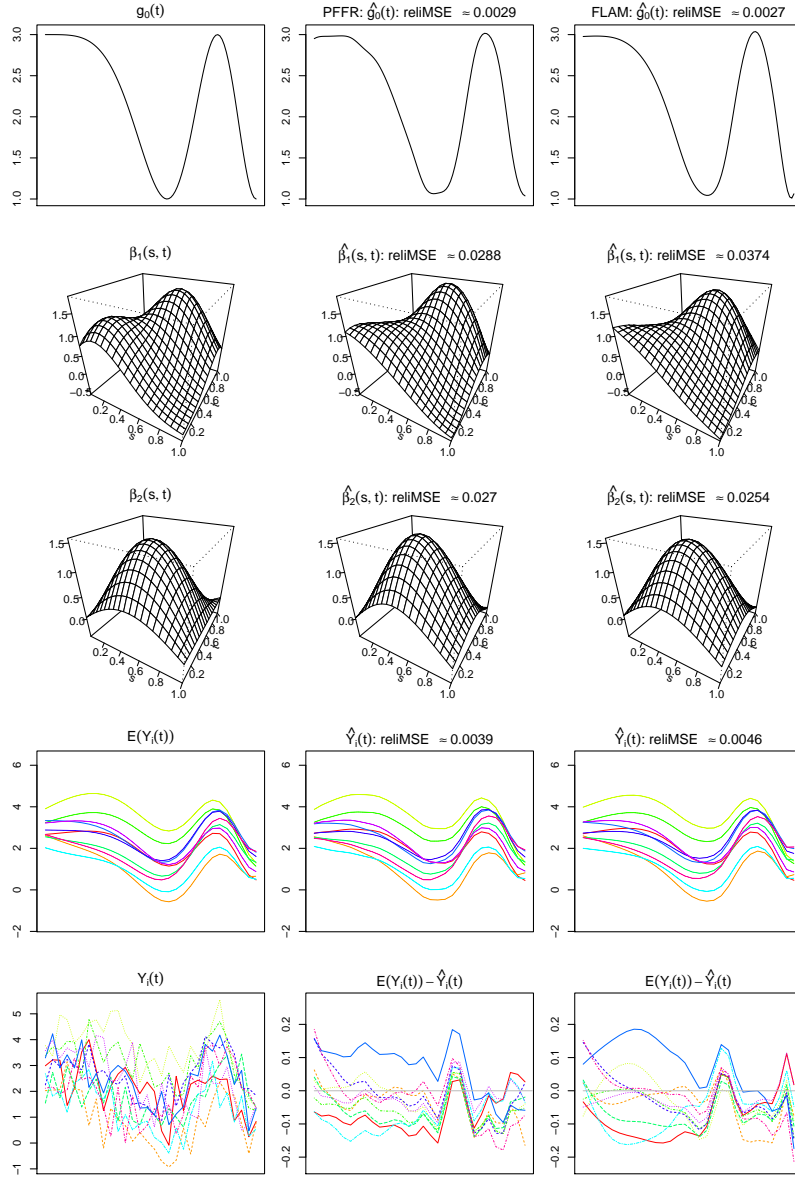
Figure 1: Data example of simulation. Simulated data and estimates with number of observations $N = 100$, number of grid-points $G = 30$ and signal-to-noise-ratio $\mathrm{SNR}_\epsilon = 2$. True coefficients and response are depicted in the left column. Estimated coefficients, predictions and residuals obtained by PFFR and by boosting are given in the middle and right column. The upper three rows show the true coefficient functions and their estimates. The forth row shows true and predicted response for ten observations, the lowest panel the response with errors and the residuals for the same observations.
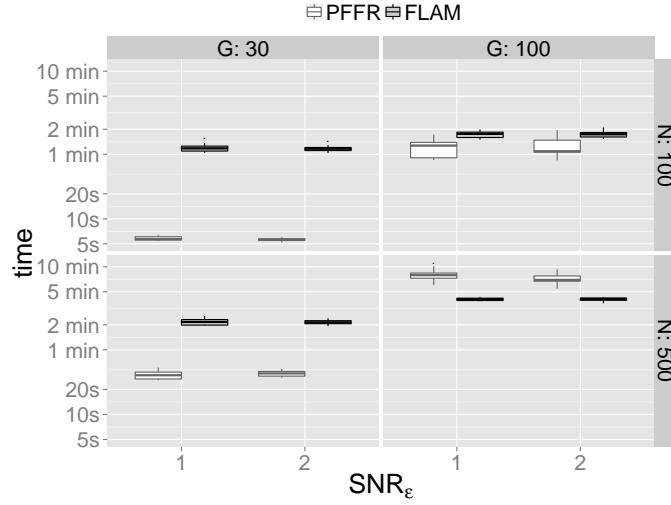
Figure 2: Computation time. The boxplots show the computation times for all combinations of sample size N, number of grid points G and signal to noise-ratio SNR_ε.

smooth spatially correlated residuals:

$$\mathbb{E}(Y_i(t)|\text{rg}_i,\text{temp}_i,i) = I(\text{rg}_i = k)\beta_k(t) + \int \text{temp}_i(s)\beta(s,t)ds + e_i(t),$$

where $Y_i(t)$ is the log-precipitation over month $t = 1,\ldots,12$, I is the indicator function, rg_i is the region of the i-th station, $\beta_k(t)$ are the smooth effects per region, $\text{temp}_i(s)$ is the centered temperature over the month $s = 1,\ldots,12$, so that $N^{-1}\sum_i \text{temp}_i(s) = 0 \ \forall s$, $\beta(s,t)$ is the coefficient surface and $e_i(t)$ are smooth spatially correlated residual curves. This model for a functional response thus depends on a scalar and a functional covariate and includes smooth spatially correlated residuals in addition to allowing for error terms $\varepsilon_{it} = Y_i(t) - E(Y_i(t)|x_i)$ uncorrelated along $\mathcal{T}$.

To set up effects in the model with functional response we use formula (2.2). The basis over the domain of the response $\boldsymbol{b}_Y(t)$ is a cyclic P-spline basis in all effects to achieve similarity between the coefficient estimates for January and December. For the effect in the functional covariate temperature, we set up the basis $\boldsymbol{b}_j(x(s))$ analogously to equation (6.2). We use a cyclic B-spline basis over time combined with the functional observations and their integration weights plus a squared difference matrix as penalty. The region-specific smooth effects and the smooth residuals are linear effects in the covariates –dummies for the regions and for each curve, respectively– of the form $x\beta(t)$. As the intercept is included in the region-specific effect, the base-leaner is $\boldsymbol{b}_j(x)^\top = (x)$, with x being a dummy-vector of length 4. For the smooth residuals $e_i(t)$ we enforce the default sum-to-zero-constraint at each t, c.f. Section 2 and the online appendix A. For the region specific effect we use a Ridge-penalty by setting the penalty to the four-dimensional identity matrix $\boldsymbol{P}_j = \boldsymbol{I}_4$. The spatial correlation of the
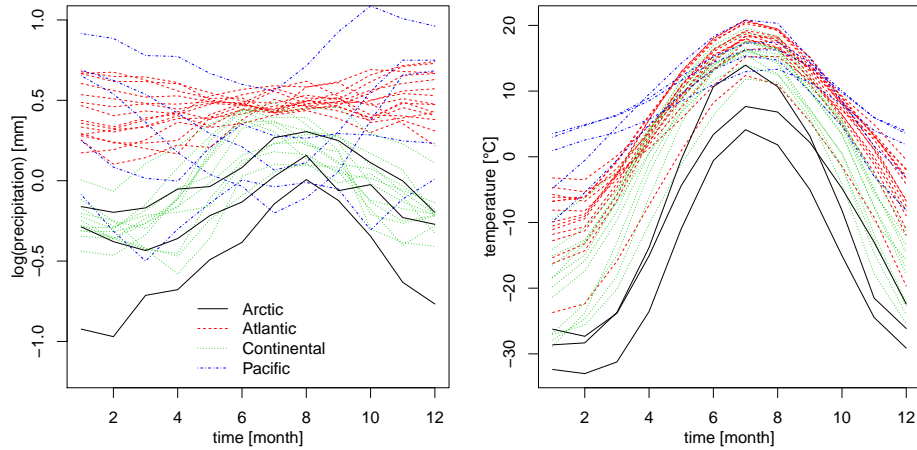
Figure 3: Canadian weather data. Monthly average temperature and log-precipitation at 35 locations in Canada. Regions are coded by colors and different line types.

residual curves is accommodated by using the inverse of a spatial correlation matrix as penalty matrix $\boldsymbol{P}_j$. Following Scheipl et al. (2014), we use a Matérn correlation matrix with smoothness parameter 0.5 and range 310 kilometers, which implies a rather strong correlation.

The optimal stopping iteration for 25-fold bootstrapping over curves is so small that the base-learner for smooth residual curves is not selected (m_{stop}=49). If the optimal stopping iteration is determined by leaving-one-curve-out cross-validation, it can be seen that for three weather stations the out-of-bag prediction is quite bad and getting worse for higher m_{stop}, causing the optimum of the mean to be very small. Those three stations are Pr. Ruppert (14), Kamploos (15) and Resolute (34) (numbers in brackets correspond to numbers in Figure 5). When looking at the median over the squared errors the optimal stopping iteration is much higher (m_{stop}=750) so that the base-learner for the smooth residuals is selected into the model. As the effects for region and temperature are very similar irrespective of the number of boosting iterations, we limit the representation of results to the model with m_{stop}=750, as it includes all effects. Figure 4 shows the estimated coefficients for the regions and the effect of temperature on log-precipitation. In general the precipitation is lowest in spring and highest in summer. An exception is the Pacific region where the highest precipitation values are measured in winter. Stations in the Atlantic region have the highest precipitations during the whole year. In the Continental region differences between the seasons are most pronounced. The effect of temperature on log-precipitation changes over the year. Higher temperatures in spring and summer are associated with lower precipitation over the whole year whereas higher temperatures in autumn and winter are associated with higher precipitation values. The association of temperature and precipitation is stronger in the winter than in the summer. Figure 5 shows the smooth residual curves. They vary in the range of -0.4 to 0.4. The most extreme smooth residuals are estimated for the Pacific region, where precipitation is relatively
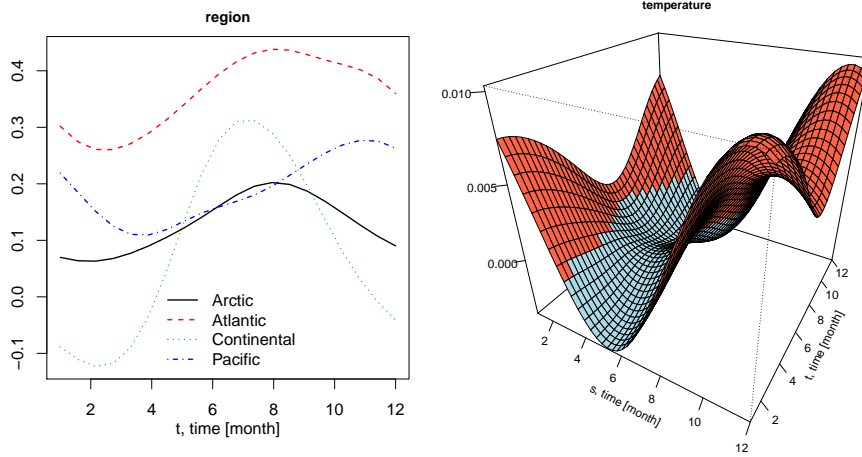
Figure 4: Estimated coefficient functions. The estimated coefficients for the four climatic regions are plotted with color coded regions (left panel). The coefficient function for the functional effect of temperature is colored in red for positive values and in blue for negative values (right panel).

variable on a small spatial scale indicating additional unmeasured covariates besides the regional and spatial effects. The uncorrelated error terms ε_{it} are quiet small compared to the smooth residual curves $e_i(t)$. Comparing the results obtained by PFFR (Scheipl et al., 2014) and boosting, the estimates for the effects of region and temperature are very similar. The estimated spatially correlated smooth residuals are similar in part but different for some stations.

D Characteristics of different general frameworks for regression with functional data

Table 1 gives an overview of the three very general different frameworks for regression with functional data, the newly presented FLAM model estimated by boosting, the PFFR model by Scheipl et al. (2014) and the Bayesian function-on-function regression model by Meyer et al. (2013).
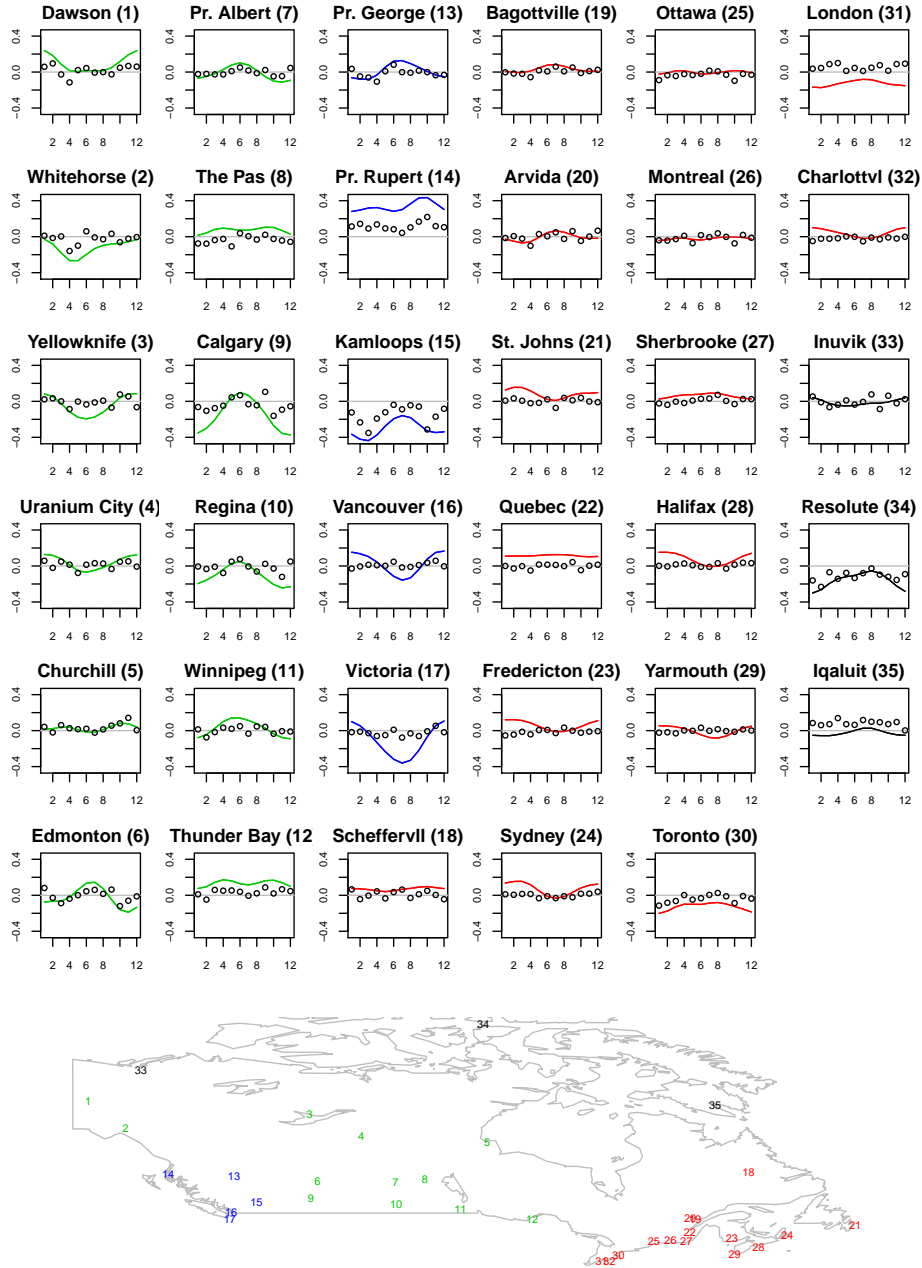
Figure 5: Estimated residuals. The estimated smooth spatially correlated residual curves $e_i(t)$ (lines) and the uncorrelated error terms ε_{it} (circles) are plotted with regions color-coded. The locations of the weather stations are given in the map at the bottom.

Table 1: Overview table for general frameworks for regression with functional data summarizing some characteristics of the FLAM models estimated by boosting, the penalized function-on-function regression (PFFR) by Scheipl et al. (2014) and Bayesian function-on-function regression by Meyer et al. (2013).

Characteristic	boosting FLAM	PFFR	Bayesian FFR
GLM for functional response	yes	yes	yes
general loss functions, e.g. quantile regression	yes	no	no
scalar response	yes	(yes)[1]	(distributed lag models)[2]
types of covariate effects	many	many	many
variable selection, $n > p$	yes	no	no
missing values	yes	yes	no[3]
inference based on	bootstrap	Likelihood / mixed models	Bayesian methods
computational scalability			
- for large n	good	fair	fair
- for large G	good	fair	good

[1] for scalar-on-function regression see penalized functional regression models by Goldsmith et al. (2011)

[2] for distributed lag models see Malloy et al. (2010)

[3] but see Morris et al. (2006) for an imputation scheme in particular cases of missingness

References

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics*, **20**(4), 830–851.

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010). Wavelet-based functional linear mixed models: an application to measurement error–corrected distributed lag models. *Biostatistics*, **11**(3), 432–452.

Meyer, M. J., Coull, B. A., Versace, F., Cinciripini, P., and Morris, J. S. (2013). Bayesian function-on-function regression for multi-level functional data. Technical report, The SelectedWorks of Jeffrey S. Morris. Available at `http://works.bepress.com/jeffrey_s_morris/52`.

Morris, J. S., Arroyo, C., Coull, B. A., Ryan, L. M., Herrick, R., and Gortmaker, S. L. (2006). Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *Journal of the American Statistical Association*, **101**(476), 1352–1364.

Ramsay, J. O. and Silverman, B. W. (2006). *Functional Data Analysis*. Springer-Verlag, New York.

Scheipl, F., Staicu, A.-M., and Greven, S. (2014). Functional additive mixed models. *Journal of Computational and Graphical Statistics.* In press, DOI 10.1080/10618600.2014.901914.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman & Hal/CRC, Boca Raton, Florida.