



# Comments on: Inference and computation with Generalized Additive Models and their extensions

Sonja Greven<sup>1</sup> · Fabian Scheipl<sup>2</sup>

Published online: 23 April 2020  
© The Author(s) 2020

**Mathematics Subject Classification** 62-08 · 62J05 · 62J99 · 62G08 · 62G99 · 62P99

## 1 Introduction

Simon Wood describes a very general framework for additive regression modeling. We wholeheartedly would like to congratulate him not only on this well-written overview but also on the work that it summarizes, much of it his own. In particular, this includes the methodological and theoretical developments, but also the availability of an implementation of much of what is described in the R package `mgcv` (Wood 2019). This allows these versatile modeling tools to be the basis for a whole ecosystem of follow-up work by other researchers. It also ensures that the methods are not only used by statisticians, but are truly useful for researchers with all kinds of applications ranging from ecology (Pedersen et al. 2018) to linguistics (Winter and Wieling 2016; Baayen et al. 2018).

The model class that Wood describes in Section 3.3, as it is based on the general concept of penalized regression, is even larger than might be apparent from the many examples given. Together with the comprehensive and extendable implementation, this means that many further models can be fitted. In the following sections, we describe two such extensions from our own work, which rely on the inferential techniques presented here: regression with functional data in Sect. 2 and time-to-event models in Sect. 3. We close with some comments on statistical infer-

---

This comment refers to the invited paper available at: <https://doi.org/10.1007/s11749-020-00711-5>.

---

✉ Sonja Greven  
sonja.greven@hu-berlin.de  
Fabian Scheipl  
fabian.scheipl@stat.uni-muenchen.de

<sup>1</sup> Chair of Statistics, Humboldt-Universität zu Berlin, Spandauer Str. 1, 10178 Berlin, Germany

<sup>2</sup> Department of Statistics, LMU Munich, Ludwigstr. 33, 80799 Munich, Germany

ence and thoughts on potential extensions from our own perspective in Sects. 4 and 5.

## 2 Functional regression

As an example of the extensions possible with the discussed model class, we briefly discuss a general framework of functional regression models that we proposed in Scheipl et al. (2015, 2016), Brockhaus et al. (2015, 2016) and summarized in Greven and Scheipl (2017), with the accompanying R packages `refund` (Goldsmith et al. 2018) and `FDboost` (Brockhaus et al. 2020) internally using the `mgcv` and `mboost` (Hothorn et al. 2010) packages for the model fitting, respectively.

For functional responses, the general idea is to acknowledge that these are only discretely observed and to model these discrete observations, with the functional data structure accounted for in the predictor. For the mean, a general additive model then is

$$E(Y_i(t_{id})|X_i = \mathbf{x}_i) = \sum_{j=1}^J h_j(\mathbf{x}_i, t_{id}) \quad (1)$$

for curve  $i$ ,  $i = 1, \dots, n$ , observed at  $t_{id}$ ,  $d = 1, \dots, D_i$ , in some interval  $T$ . Each partial effect  $h_j$  can depend on one or several scalar and/or functional covariates  $\mathbf{x}_i$  and vary with  $t_{id}$ . This can also be generalized to some other feature than the mean, like, e.g., the mean composed with a link function (Scheipl et al. 2016), a quantile (Brockhaus et al. 2015), or even several features like the mean and variance (Brockhaus et al. 2018; Stöcker et al. 2018). As covariate effects are unlikely to explain all structure in the data and there will thus be remaining auto-correlation along  $t$ , it is usually necessary to include a functional residual  $E_i(t_{id})$  (or functional random intercept per curve) as one of the  $h_j(\mathbf{x}_i, t_{id})$  in (1). Conditional on this, we can then reasonably assume uncorrelatedness of the remaining white noise ‘measurement error.’

All model terms are expanded in suitable penalized basis expansions (see Greven and Scheipl 2017), directly building on the penalized basis expansions discussed by Simon Wood and extending these for the additional dimension over  $t$ . Functional covariates can be included with linear effects as in the signal regression example in Wood’s Section 3.3, illustrating the usefulness of the discussed ‘linear functionals of smooths as model components.’ These functional covariate effects can also be extended to the case of functional responses, to nonlinear effects (Scheipl et al. 2015) or varying smoothness of the coefficient function along the functional covariate. Setting up the basis expansions and penalties, we have to be careful to impose suitable and interpretable identifiability constraints on all model terms (Brockhaus et al. 2015), which is very much related to the discussion of main and interaction effects in smooth ANOVA models in Wood’s Section 3.2.1. Once this is achieved, model fitting essentially reduces to an application of the discussed penalized regression framework and can be achieved using `mgcv` or `mboost`.

Additive models of the kind described here are also important for functional data for estimating mean functions and covariance operators as inputs for functional principal component analysis (FPCA). To estimate covariances  $C_X(t, t') = \text{Cov}(X(t), X(t'))$

from noisy, potentially sparse realizations of functional data  $\tilde{x}_i(t_{id}) = x_i(t_{id}) + \epsilon_{id}$ , centered cross-products are typically used, as their expectation corresponds to the covariance. This reduces the problem of estimating  $C_X(t, t')$  to smoothing—by means of bivariate penalized splines and leaving out the diagonal—the mean surface for the products  $(\tilde{x}_i(t_{id}) - \hat{\mu}(t_{id}))(\tilde{x}_i(t_{id'}) - \hat{\mu}(t_{id'}))$ , where  $\hat{\mu}(t)$  is the estimated smooth mean function. This is a fairly challenging problem even for simple settings due to the quadratically increasing number of cross-products and the constraints of symmetry and positive definiteness on the estimated surface. The problem is exacerbated for nested, crossed or hierarchical functional data in which the overall covariance is a superposition of the covariances on the different grouping levels as well as the auto-covariances along  $t$ . Additive model-based surface estimates can be used very effectively in both the simpler (Cederbaum et al. 2018) and more complex settings (Cederbaum et al. 2016).

Subsequently, FPCA can be an end in itself for exploratory or descriptive analysis. The estimated functional principal components can also be used as empirical,  $L_2$ -optimal compact basis for (penalized) basis representations of the functions in subsequent analysis steps. For the FPCA scores, the used ridge-type penalty is inversely proportional to the eigenvalues of the covariance operator, as these eigenvalues correspond to the score variances due to the Karhunen–Loève theorem (Scheipl et al. 2015). Used in functional additive mixed models, this approach greatly increases computational efficiency compared to penalized spline-based basis expansions (Cederbaum et al. 2016). This also shows that the discussed penalized regression framework is not limited to bases such as splines, but that empirical bases learned from the data can be useful additions as well.

### 3 Modeling time-to-event data

Due to the additional challenges posed by (partial) likelihood inference for censored and/or truncated data, the available implementations for time-to-event models have historically lagged behind their counterparts for conventional regression problems in both flexibility and performance. We expect that subsuming Cox-type proportional hazard regression models into a framework and implementation as flexible, fast and well validated as the one described by Wood to be a boon to researchers and practitioners in this field alike, and allow them to free themselves from unexamined linearity and independence assumptions. In our own work on piecewise-exponential additive mixed time-to-event models (PAMM) (Bender et al. 2018a; Bender and Scheipl 2018), we have followed the Poisson likelihood-based data augmentation approach (Argyropoulos and Unruh 2015, e.g.,) alluded to in Section 3.5 of the discussed article, with convincing results on large real-world data sets. While naive implementations of this approach would incur a penalty in computational effort roughly proportional to the size of the data set, due to the highly repetitive structure of the augmented pseudo-data, this is compensated to some extent by the tremendous efficiency gains that `mgcv` and boosting implementations like `mboost` achieve via the ‘unique value compression’ of the model matrix (Lang et al. 2014; Wood et al. 2017; Li and Wood 2020, cf. Section 5.4 of the discussed article). Combining this representation of time-to-event

models with the inferential strategies laid out in the discussed article opens up the possibility of routine use of penalized estimators for very complex nonlinear and/or smoothly *time-varying* effects as well as time-varying covariates for time-to-event and even competing-risk or recurrent-event data sets, also under complex censoring schemes. Note that neither time-varying effects nor time-varying covariates are possible in conventional proportional hazards models, at least not without incurring an increase in computational effort equivalent to that of the Poisson pseudo-data approach. In our work, we have employed the PAMM approach to re-analyze a large heterogeneous database ( $\sim 20,000$  patients,  $\sim 600$  study locations, Heyland et al. 2010) of ICU cases to investigate the time-varying, delayed and time-limited as well as cumulative-over-time association of nutrition with mortality risk, correcting for heterogeneity between study locations and complex nonlinear and time-varying confounder effects (Bender et al. 2018b). To the best of our knowledge, fitting such complex time-to-event models on data sets of this size is not currently feasible with any other approach. Once more, this demonstrates the versatility and methodological fecundity of the general additive model framework.

## 4 Comments on inference

### 4.1 Boosting and inference

Model-based boosting (Hothorn et al. 2010) as described in Section 5.3 is a useful approach to model estimation in particular for large and complex models, as it includes the selection of model terms and scales well to high-dimensional data. We have ourselves used this estimation approach for the functional regression models discussed in Sect. 2 (Brockhaus et al. 2015, 2016, 2020). While Simon Wood laments the ‘inability’ of boosting ‘to drop a term, once included’, we have found that stability selection (Shah and Samworth 2013) is useful to overcome this problem, and to only choose the model terms that are stably selected into the model under subsampling.

A fact maybe underappreciated in the context of boosting is that standard bootstrap-based uncertainty quantification will not lead to confidence intervals or bands with proper coverage (discussed, e.g., in Rügamer et al. 2018). This is due to the shrinkage bias of boosting-based estimators, which will lead each fit in each bootstrap sample to be shrunk (biased) towards zero, and also has to do with the model selection aspect of boosting. Thus, confidence bands, say, based on percentiles of nonparametric bootstrap fits, will not be correctly centered.

We have proposed inference for  $L_2$ -boosting in the special case of linear, grouped and penalized additive models in Rügamer and Greven (2020). The used framework is selective inference (Fithian et al. 2014; Tibshirani et al. 2016; Yang et al. 2016), a recent development in post-selection inference, which conditions inference on the observed model selection event. This follows the principle that ‘the answer must be valid given that the question was asked’ (Fithian et al. 2014), i.e., given that the parameters chosen for testing or confidence bands were previously selected into the model based on the same data that is then used for inference. The conditioning leads to truncated distributions for, e.g., test statistics, which for  $L_2$ -boosting we characterize

using a Monte Carlo approximation based on ideas of Yang et al. (2016). Further work is needed to also obtain valid inference in mixed or generalized models estimated using boosting.

## 4.2 Smoothing parameter uncertainty

As discussed in Sections 2.3 and 5.1 of Simon Wood's article, to obtain proper inference for the  $\beta$  coefficients, it is important to incorporate smoothing parameter uncertainty. A simple correction from Wood et al. (2016) is presented in 5.1, using a normal approximation to the posterior of  $\beta$  with a correction term for the covariance matrix.

As smoothing parameters can tend to infinity if function estimates tend to smooth functions in the null space of the penalty, there is a phase transition that we discussed in Greven and Scheipl (2016). It turns out that the correction in 5.1 works well for estimates far from the null space (e.g., for clearly nonlinear functions if deviations from linearity are penalized). As the smoothing parameter approaches infinity, however, the corrected covariance matrix actually approaches the uncorrected one and smoothing parameter uncertainty is not accounted for. At the same time, the posterior for  $\beta$  deviates from a normal distribution and can show, e.g., heavier tails or bimodality. This means that for function estimates close to the penalty null space, confidence bands are too narrow and can incorrectly ignore the possibility of functions outside the penalty null space.

We also wonder how this affects the performance of the AIC. If smoothing parameter uncertainty is not accounted for, the (conditional) AIC tends to select too many model terms into the model (Greven and Kneib 2010). As the correction term in 5.1 nearly disappears for the most interesting cases close to the null space, it would be interesting to further investigate how this influences model selection behavior. Finally, note that if confidence bands and tests are constructed in models selected using the AIC, a post selection inference problem occurs similarly to the one discussed for boosting in Sect. 4.1 above, which it would be interesting to study further.

## 5 Outlook and challenges

To achieve even more widespread adoption, the powerful and extremely flexible models based on the presented framework will need to be accompanied by similarly powerful and versatile model visualization and model diagnostic tools. As the fitted models become more complex and feasible alternatives for model specifications multiply, practitioners will require such tools more and more. One important step in this direction, implementing the ideas in Figure 9 of the discussed article and many more, is already implemented in the add-on package `mgcViz` (Fasiolo et al. 2018). This tool allows to build very problem specific visual model diagnostics for identifying model misspecification and iteratively improving fitted models. It would be great if such tools were developed further and became even more accessible.

Another important challenge for the discussed model class is the as yet unresolved question on how to incorporate general (marginal) dependence structure of

the responses that cannot be captured by conditioning on effects of covariates and/or the data's grouping structure. Our experience indicates that this question is more general than including 'short range auto-correlations.' In functional regression, it will require coming to grips with heterogeneity and auto-covariance of residuals along the functional responses' domain in a principled way. Much more generally, the question is how to subsume the estimation of (semi-)structured residual covariance in this framework in a computationally efficient and theoretically sound way. The goal should be that analysts have a choice between modeling such dependence structures either by conditioning on grouping-level specific effects (e.g., a functional random intercept  $E_i(t)$  for each curve as in the functional models of Sect. 2), which becomes computationally challenging if the number of levels is large and potentially limits the scope of implied marginal correlation structures, or by marginalizing them out. While simple random intercepts imply a compound symmetry correlation matrix marginally, marginalization becomes even more challenging for more complex random effects structures. A related point appears in the context of GAMLSS-type (generalized additive models for location, scale and shape) functional response models (Stöcker et al. 2018), where conditioning on functional residuals means that only the measurement error variance is allowed to vary according to an additive predictor. Ideally, we would like to model both the residual (marginal) variance as well as the remaining (marginal) functional covariance structure depending on covariates. This, however, would require even further extensions to the framework.

Finally, much has been done to extend the flexibility of the discussed additive models to outcomes beyond scalar responses, including multivariate response vectors discussed in Wood's Section 3.4 and functional observations discussed in our Sect. 2. Extending the approach even further to what is sometimes called 'object data', where the outcome is an object such as, e.g., a 2D or 3D curve, a density, a composition or a shape is one of the interesting challenges lying ahead in the further development of this field.

**Acknowledgements** Open Access funding provided by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Argyropoulos C, Unruh ML (2015) Analysis of time to event outcomes in randomized controlled trials by generalized additive models. *PLoS ONE* 10(4):e0123784
- Baayen RH, van Rij J, de Cat C, Wood S (2018) Autocorrelated errors in experimental data in the language sciences: some solutions offered by generalized additive mixed models. In: Speelman D, Heylen K, Geeraert D (eds) *Mixed effects regression models in linguistics*. Springer, Berlin, pp 49–69

- Bender A, Scheipl F (2018) Pammtools: piece-wise exponential additive mixed modeling tools. Arxiv preprint [arXiv:1806.01042](https://arxiv.org/abs/1806.01042)
- Bender A, Groll A, Scheipl F (2018a) A generalized additive model approach to time-to-event analysis. *Stat Model* 18(3–4):299–321
- Bender A, Scheipl F, Hartl W, Day AG, Küchenhoff H (2018b) Penalized estimation of complex, non-linear exposure-lag-response associations. *Biostatistics* 20(2):315–331
- Brockhaus S, Rügamer D, Greven S (2020) Boosting functional regression models with FDboost. *J Stat Softw.* <https://www.jstatsoft.org/index>
- Brockhaus S, Scheipl F, Hothorn T, Greven S (2015) The functional linear array model. *Stat Model* 15(3):279–300
- Brockhaus S, Melcher M, Leisch F, Greven S (2016) Boosting flexible functional regression models with a high number of functional historical effects. *Stat Comput* 27:913–926
- Brockhaus S, Fuest A, Mayr A, Greven S (2018) Signal regression models for location, scale and shape with an application to stock returns. *J R Stat Soc Ser C* 67(3):665–686
- Cederbaum J, Pouplier M, Hoole P, Greven S (2016) Functional linear mixed models for irregularly or sparsely sampled data. *Stat Model* 16(1):67–88
- Cederbaum J, Scheipl F, Greven S (2018) Fast symmetric additive covariance smoothing. *Comput Stat Data Anal* 120:25–41
- Fasiolo M, Nedellec R, Goude Y, Wood SN (2018) Scalable visualisation methods for modern generalized additive models. *J Comput Graph Statist.* Arxiv preprint [arXiv:1707.03307](https://arxiv.org/abs/1707.03307) (to appear)
- Fithian W, Sun D, Taylor J (2014) Optimal inference after model selection. Technical report, [arXiv:1410.2597](https://arxiv.org/abs/1410.2597)
- Goldsmith J, Scheipl F, Huang L, Wrobel J, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C, Reiss PT (2018) Refund: regression with functional data. R package version 0.1-17
- Greven S, Kneib T (2010) On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika* 97(4):773–789
- Greven S, Scheipl F (2016) Comment on “Smoothing parameter and model selection for general smooth models” by Wood, Pya and Säfken. *J Am Stat Assoc* 111(516):1568–1573
- Greven S, Scheipl F (2017) A general framework for functional regression modelling. *Stat Model* 17(1–2):1–35
- Heyland DK, Cahill NE, Dhaliwal R, Sun X, Day AG, McClave SA (2010) Impact of enteral feeding protocols on enteral nutrition delivery: results of a multicenter observational study. *J Parenter Enter Nutr* 34(6):675–684
- Hothorn T, Buehlmann P, Kneib T, Schmid M, Hofner B (2010) Model-based boosting 2.0. *J Mach Learn Res* 11:2109–2113
- Lang S, Umlauf N, Wechselberger P, Harttgen K, Kneib T (2014) Multilevel structured additive regression. *Stat Comput* 24(2):223–238
- Li Z, Wood SN (2020) Faster model matrix crossproducts for large generalized linear models with discretized covariates. *Stat Comput* 30:19–25
- Pedersen EJ, Miller DL, Simpson GL, Ross N (2018) Hierarchical generalized additive models: an introduction with mgcv. *Peer J Prepr* 6:e27320v1
- Rügamer D, Greven S (2020) Inference for  $L_2$ -boosting. *Stat Comput* 30(2):279–289
- Rügamer D, Brockhaus S, Gentsch K, Scherer K, Greven S (2018) Boosting factor-specific functional historical models for the detection of synchronization in bioelectrical signals. *J R Stat Soc Ser C Appl Stat* 67(3):621–642
- Scheipl F, Staicu AM, Greven S (2015) Functional additive mixed models. *J Comput Graph Stat* 24(2):477–501
- Scheipl F, Gertheiss J, Greven S (2016) Generalized functional additive mixed models. *Electron J Stat* 10(1):1455–1492
- Shah R, Samworth R (2013) Variable selection with error control: another look at stability selection. *J R Stat Soc Ser B* 75:55–80
- Stöcker A, Brockhaus S, Schaffer S, von Bronk B, Opitz M, Greven S (2018) Boosting functional response models for location, scale and shape with an application to bacterial competition. *Statist Model.* [arXiv:1809.09881](https://arxiv.org/abs/1809.09881)
- Tibshirani R, Taylor J, Lockhart R, Tibshirani R (2016) Exact post-selection inference for sequential regression procedures. *J Am Stat Assoc* 111(514):600–620

- Winter B, Wieling M (2016) How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *J Lang Evol* 1(1):7–18
- Wood SN, Pya N, Säfken B (2016) Smoothing parameter and model selection for general smooth models. *J Am Stat Assoc* 111(516):1548–1563
- Wood S (2019) mgcv: mixed GAM computation vehicle with automatic smoothness estimation. R-package version 1.8–31. <https://CRAN.R-project.org/package=mgcv>
- Wood SN, Li Z, Shaddick G, Augustin NH (2017) Generalized additive models for gigadata: modeling the UK black smoke network daily data. *J Am Stat Assoc* 112(519):1199–1210
- Yang F, Barber RF, Jain P, Lafferty J (2016) Selective inference for group-sparse linear models. In: *Advances in neural information processing systems*, pp 2469–2477

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.