FLEXIBLE REGRESSION AND SMOOTHING: Using GAMLSS in R by M.D. Stasinopoulos, R.A. Rigby, G.Z. Heller, V. Voudouris, F. De Bastiani. CRC Press The R Series (2017) Boca Raton FL, 549 pages.

A.H. Welsh
Australian National University

This well-written book is an introduction to Generalized Additive Models for Location, Scale and Shape (GAMLSS) and the use of the R package `gamlss` developed by the authors for fitting and using these models. The focus is mainly on the R package, making applied statisticians the primary target audience, although the Preface states that two further volumes on GAMLSS models are forthcoming. This is a difficult type of book to write because of the need to steer a path between writing a manual for the package and a book on applied statistics. The authors have obviously taken considerable care and produced a very readable book; it contains a lot of information about the package but it does not feel like just a manual.

The basic GAMLSS class of models includes 4 parameter (conditional) response distributions with parameters representing location, scale and shape (one for skewness and one for kurtosis), and allows some or all the parameters in the distribution to be additive, smooth functions of explanatory variables and/or of random effects. The `gamlss` package allows smooth functions of explanatory variables to be estimated in various ways, allows the use of penalized likelihood methods, including ridge regression and the lasso, and also implements fitting finite mixtures of distributions (including zero inflated and zero adjusted models as special cases) and centile or quantile estimation. In all of this, the smoothing parameters can be chosen automatically. It is very flexible, potentially very useful and highly extendable. This is a good reason for looking into the book and considering using the package.

The presentation of statistical material would be much easier if we could administer an initial massive vaccination of concepts, ideas, techniques and experience and then build on this by adding deeper layers of knowledge, subtlety and experience. Unfortunately, this is not (at least yet) possible and books in particular require a linear presentation of material. This raises difficult questions (to which there are no perfect answers) of what to assume of readers, where to start the presentation and how to deal with the inevitable gaps? The Preface states that readers should be familiar with basic regression and have a working knowledge of R. I think a reader with just that level of background would find it a very challenging book and the more you already know the more you are likely to get out of it. For example, it helps if you already know a great deal about modern statistical modelling, including regression and generalized linear modelling, and it is useful to have some experience of smoothing and generalized additive models. The book uses some of the language of statistical learning, some of the techniques (e.g. we are presented with neural networks as a method of non-parametric smoothing in Chapter 2) and mentions big data in passing but is not primarily about these topics and is more solidly positioned in applied statistics.

The book starts in Part I with the analysis of a real data set to motivate the use of GAMLSS models by illustrating their fitting with `gamlss` and then introduces the `gamlss` package. This is a very appealing way to start but requires a lot of the reader. The authors deal with the unavoidable gaps related to these topics by referring ahead in the book for explanations and giving various warnings. We are referred to Part V for a definition and explanation of the choice of residuals which are probability integral transformed residuals (p 418 ff) and worm plots (p 426 ff) which are detrended Gaussian QQ plots enhanced with approximate pointwise 95% confidence intervals. The importance and value of the early warnings inevitably depend on the background knowledge of the reader who has to interpret them. Already in Chapter 1, we read that: the default links are not canonical; residuals are normalized (randomized) quantile residuals; when smoothers are fitted all standard errors shown should be treated with caution. This can be annoying but, if you keep going, you do get past it.

The authors state that the book is not intended to be read right through (applied readers are recommended to start with the first and last Parts I and VI, then III-V and finally Part II) but, if you want to use the package seriously, you probably do need to read it through before coming back and dipping into specific parts as you need to. Following the advice to the applied reader and jumping to Part VI, means jumping from an introduction to the package to a chapter on centile or quantile regression and then some further general examples. It feels different from the rest of the book (in much the same way as finite mixture distributions feel different from the rest of the book) so unless this topic is very important to you, you'd be better leaving it and going to the last chapter which presents examples that fit more with the rest of the book. The examples are good; the authors are careful to always keep their statistical purpose in front of the reader (stating it in a grey box at the start of each example) but I felt that the substantive purpose of each analysis was less clear. This is probably inevitable when the focus is illustrating the use of the package, but it distracts from the idea of using the book in an applied statistics course. As exercises are included, the authors are obviously keen for the book to be used in this way too. Also, the authors sometimes just fit a whole set of models without any explanation of why they have chosen the set (see for example the fish species data pp498-507). In practice, I think it is preferable to proceed logically step by step and try to do no more and no less than needs to be done.

Parts III-V are the statistical core of the book. Part III is concerned with the distributions that can be fitted and the link functions that can be used for their parameters. It explains how to modify and add distributions and link functions to extend the package. The package includes finite mixtures of distributions and allows modelling the mixture probabilities as functions of covariates. In finite mixture models, there is a distinction in the package between parameters not in common and parameters in common in the mixture components. For parameters in common, all distributions and links must be the same for all components and mixing probabilities are not allowed to depend on covariates (p207). The two types of mixtures are fitted by `gamlssMX()` and `gamlssNP()`, respectively, the latter being the function used to fit GAMLSS models with random effects. The fitting of finite mixtures is connected to fitting random effects through the fact that conceptually both can be viewed as creating mixtures of distributions and computationally both can be fitted using the EM algorithm. This results in a potentially confusing use of the random formula on p209: "This formula is also used for fixed effect mixture models to define interactions of the factor MASS with continuous explanatory variables x in the predictor for μ. If, for example, different coefficients in x in the predictor of mu are needed for the $\mathcal{K}$ different components of the finite mixture, use random = x." It is then unclear whether you can include additional random effects in a mixture model? (It seems possible with parameters in common.)

Part IV explains what the package can do with the explanatory variables and random effects in the model terms. This includes fitting polynomials, piecewise polynomials, B-splines and free-knot models as well as different kinds of smoothers. A key unifying idea is the use of penalized smoothers, both univariate and multivariate, but the package allows much more, including using monotonic and cyclic smooth functions, ridge and lasso regression, Gaussian Markov random fields, multivariate smoothers as well as other smoothers (including neural networks, loess etc). The last chapter in this Part discusses random effects and the functions for fitting them. One very interesting feature is that you can interface with `lme()` to access further possible analyses.

Part V deals with model selection, diagnostics and functions for doing both of these. The effective degrees of freedom is defined as the trace of the smoother matrix and this is used in AIC and BIC (p379). Model selection of or with random terms seems not to be discussed. A useful function when analysing large data sets is the `gamlssVGD()` function which allows the fitting of models on a test data set and their evaluation on a training set.

Part II discusses the algorithms used to fit the models, the `gamlss()` function to do the actual fitting and then various functions used for inference and prediction. Several different bootstraps can be used within the package to obtain standard errors and confidence intervals. Prediction is always complicated in R but is quite nicely explained (p135ff). It is not completely clear to me but it seems that functions for prediction require the parameters to not include random effects, making prediction actually just parameter estimation.

There is a huge amount of material in this book and it is generally well-presented. With the high level of flexibility, it is not always clear from the book what combinations of specifications are possible, how these should be interpreted and whether a numerical result that is returned means anything e.g. random effects in zero-inflated models fitted with the lasso or included in centiles and whether you can fit these using the lasso? As someone who finds that Gaussian linear mixed models with a single random effect still have their mysteries and, while widely used, are not as well understood as they should be, I found the treatment of random effects a bit cavalier. All parameters are allowed to include random terms (p346) but there is no discussion of how to interpret them. It seems that the random effects for a particular parameter are assumed to be independent and the random effects in different parameters are assumed to be independent. There is no warning that this is not always desirable (see for example Cantoni et al, 2017). Some limitations on what can and cannot be done with random effects in the package are given on p325. While the function `random()` is limited, as noted above, the `lme()` function can be accessed through `re()` and PQL used for fitting more general models. The absence of warnings about the use of PQL seems an oversight. The `gamlss` package allows normal and discrete (nonparametric) distributions for the random effects and the book does not discuss the implications of these choices. Generally, there seem to be fewer warnings to the reader later in the book. An obvious one given in the context of an example on p396 but that I would have liked to see emphasised more is that there is often very little information for estimating shape parameters, making it unwise to put very complicated structures onto these parameters unless you have very large data sets. Sometimes advice is given but with no explanation: writing about the `find.hyper()` function on p459, the authors state that it is preferable to use cubic splines (`cs()`) rather than P-splines (`pb()`). Every now and again there is a surprising piece of information about a limitation such as on p455 that `lms()` only fits a single explanatory variable (though there are other functions that can fit multiple explanatory variables) and then again on p480 quantile sheets only allow one explanatory variable. Human nature being what it is, the more we are given, the more we want.

Each chapter ends with bibliographic notes giving some historical context and references. The references are a mixture of those to the original work and those just to a convenient recent text. For my taste, there is not quite enough distinction between these. More seriously, there is no reference to closely related work such as vector generalized additive models (VGAM) which has an eponymous R package (Yee and Wild 1996, Yee, 2015, 2019) and multiparameter regression (Pan and Mackenzie, 2003, Burke and MacKenzie, 2017). More recently distributional regression (Umlauf et al 2017, Umlauf et al 2018) which is very similar to GAMLSS (even based on it) but takes a Bayesian approach and has an R package `bamlss` has appeared. There may well be other related work. It would be useful to have a summary of the differences and similarities with VGAM and the other relatives, precisely because the relationship is complicated.

References

Burke, K. and MacKenzie, G. (2017). Multi?parameter regression survival modeling: An alternative to proportional hazards. *Biometrics* **73** 678–686. https://doi.org/10.1111/biom.12625

Cantoni, E. Mills-Flemming, J. and Welsh A.H. (2017). A random-effects hurdle model for predicting bycatch of endangered marine species. *Annals of Applied Statistics* **11** 2178–2199.

Pan, J.X. and MacKenzie, G. (2003). On modelling mean?covariance structures in longitudinal studies. *Biometrika* **90** 239–44.

Umlauf, N, Klein, N and Zeileis, A. (2018). BAMLSS: Bayesian Additive Models for Location, Scale and Shape (and Beyond). *Journal of Computational and Graphical Statistics* **27** 612–627 doi:10.1080/10618600.2017.1407325

Umlauf, N., Klein, N., Zeileis, A. and Simon, T. (2018). Bamlss: Bayesian Additive Models for Location Scale and Shape (and Beyond). http://CRAN.R-project.org/package=bamlss.

Yee, T.W. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R*. New York, USA: Springer.

Yee, T.W. (2019). VGAM: Vector Generalized Linear and Additive Models. R package version 1.1-1. URL https://CRAN.R-project.org/package=VGAM

Yee, TW and Wild, CJ (1996). Vector Generalized Additive Models. *Journal of Royal Statistical Society, Series B* **58** 481–493. doi:10.1007/978-1-4939-2818-7