

mgcv tips sheet

What is mgcv ?

`mgcv` is an R package for fitting generalized additive models (GAMs). That means we can fit models where the predictors are smooth functions of the covariates. Often these smooth functions are splines, but that's not all they can be.

The main functions in mgcv

`gam`
For fitting GAMs

`gamm`
For fitting generalized additive mixed models. Can include correlation structures and performance can be better for random effects. You can specify random effects using `lme` syntax.

`bam`
For fitting big additive models. Includes some special tricks for fitting to large datasets.

Formula

formula=

We can write a model formula in `mgcv` just as we can when we use `lm` or `glm`, with some additions.

`s()` is the general setup for a smooth.

`te()` interaction via tensor product.

Distributions

family=

Binomial	<code>binomial</code>
Normal	<code>gaussian</code>
Gamma	<code>Gamma</code>
Inverse normal	<code>inverse.gaussian</code>
Poisson	<code>poisson</code>
Quasi	<code>quasi</code>
Quasi-binomial	<code>quasibinomial</code>
Quasi-Poisson	<code>quasipoisson</code>
Tweedie	<code>tw/Tweedie</code>
Negative binomial	<code>nb/negbin</code>
Beta	<code>betar</code>
Censored normal	<code>cnorm</code>
Ordered categorical	<code>ocat</code>
Scaled t	<code>scat</code>
Zero inflated Poisson	<code>zip</code>
Zero inflated Poisson location-scale	<code>zipLSS</code>
Cox proportional hazards	<code>cox.ph</code>
Generalized extreme value location-scale	<code>gevlss</code>
Normal	<code>gaulss</code>
location-scale model	<code>mvn</code>
Multivariate normal	<code>gammals</code>
Gamma	<code>gumbLs</code>
location-scale	<code>gumbLs</code>
Gumbel	<code>gumbLs</code>
location-scale	<code>gumbLs</code>
Multinomial	<code>multinom</code>
Tweedie	<code>twlss</code>
location-scale	<code>twlss</code>
Sinh-arcsinh	<code>shash</code>
location-scale-shape	<code>shash</code>
General family	<code>gfam</code>

Smoothers

Using the `bs=` argument in `s()`, `te()`, etc. Further details can be found in `?smooth.construct.*.smooth.spec`

Univariate only smoothers

Cubic regression splines `cr`

Cubic regression splines with shrinkage `cs`

Cyclic cubic splines `cc`

B-splines `bs`

P-splines `ps`

Special smoothers

Adaptive smoothers `ad`

Factor-smooth interactions `sz`

Random factor-smooth interactions `fs`

Smoothers in ≥ 1 dimension

Thin plate regression splines `tp`

Thin plate regression splines within shrinkage `ts`

Duchon splines `ds`

Random effects `re`

Markov random fields `mrf`

Gaussian process smooths `gp`

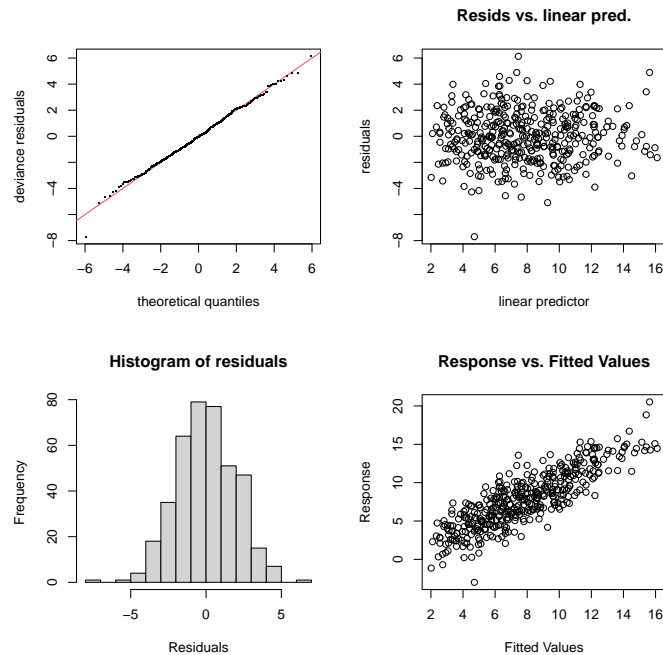
Smoothers in 2 dimensions

Splines on the sphere `sos`

Soap film smoothing `so` (`sw` and `sf`)

Model checking

gam.check



Top left: Quantile-quantile plot: points should be close to the line, meaning residuals are normally distributed.

Bottom left: Histogram of residuals: again, looking for normal(ish) distribution.

Top right: Residuals vs. linear predictor: looking for no increase or decrease in spread with increasing linear predictor value, otherwise we have heteroskedasticity.

Bottom right: Response vs. linear predictor, expecting tight line relationship indicating good agreement between the predictions and data.

Knots and basis complexity

General strategy: check k and double if too small.
When do we know k is too small?

Fitting criterion `method=`

"GCV.Cp"	Generalized cross validation, default
"REML"	REstricted Maximum Likelihood, preferred
"ML"	Maximum Likelihood
"NCV"	Neighbourhood Cross-Validation

Extras

<code>gam.mh</code>	Metropolis-Hastings sampling of the posterior
<code>concurvity</code>	Assess concavity between terms
<code>gam.vcomp</code>	Random effects style output
<code>gamSim</code>	Simulate GAM-type data
<code>inSide/in.out</code>	point-in-polygon test
<code>jagam</code>	Generate JAGS/Nimble code
<code>new.name</code>	Generate a variable name
<code>place.knots</code>	Place knots evenly
<code>rmvn</code>	Generate multivariate normal deviates

Extra help

<code>?gam.models</code>	Fitting fancy models
<code>?linear.functionals</code>	How to use <code>by=</code>
<code>?random.effects</code>	Random effects syntax
<code>?mgcv.FAQ</code>	frequently asked questions
<code>?mgcv.parallel</code>	Info on parallelisation
<code>?missing.data</code>	What to do about missing data
<code>?choose.k</code>	How to select basis size
<code>?one.se.rule</code>	Making smoother smooth models

Other packages

<code>scam</code>	Shape constrained smoothing
<code>gratia</code>	Plotting with <code>ggplot2</code>
<code>mgcViz</code>	Fancy plotting
<code>qgam</code>	Quantile GAMs
<code>gamm4</code>	Random effects based on <code>lme4</code>

Useful references

Wood. Generalized Additive Models. An Introduction with R. 2nd ed. CRC Press, 2017

Pedersen, Miller, Simpson and Ross. Hierarchical Generalized Additive Models in Ecology: An Introduction with `mgcv`. PeerJ (2019). <https://doi.org/10.7717/peerj.6876>