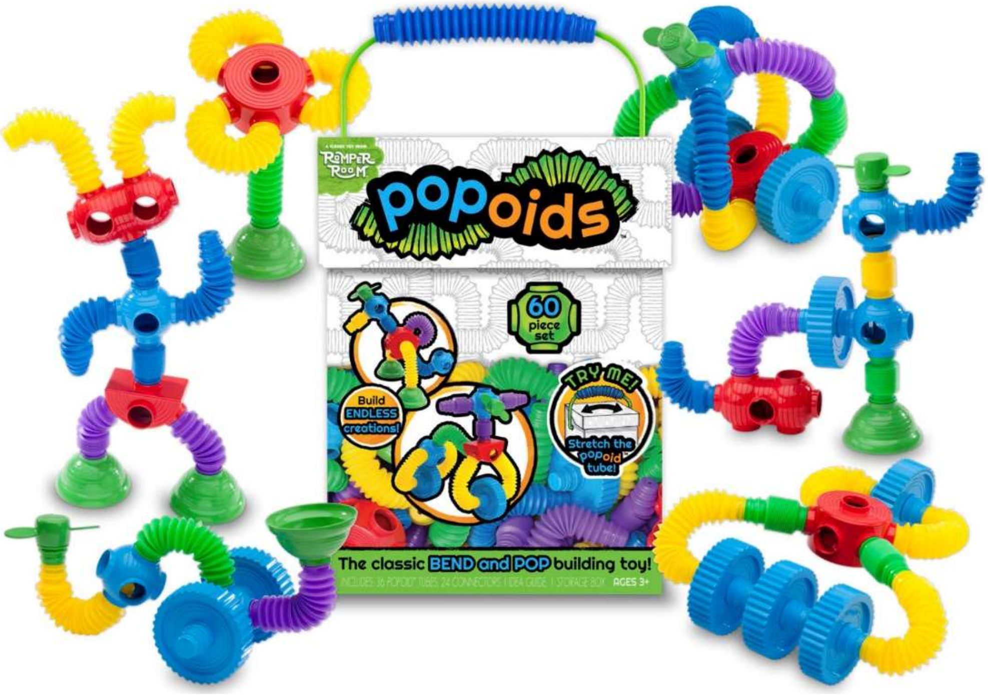
May 21, 2020 – IEP Data Science PWT

Introduction to Generalized Additive Models (GAMs) with R

Vanessa Tobias, PhD

✉ vanessa_tobias@fws.gov

@marshprincess


ROMPER ROOM
popoids
60 piece set
Build ENDLESS creations!
TRY ME!
Stretch the popoid tube!
The classic BEND and POP building toy!
INCLUDES 36 POPOID TUBES 24 CONNECTORS 1 IDEA GUIDE 1 STORAGE BOX
AGES 3+

The use of popoids or any other brand in this presentation isn't an endorsement. The generic concept of bendable tubes with connectors is what we're going for here.

What is a GAM?

SPECIAL SUPPLIES
FUN TUBES
HOURS OF SENSORY AND FINE MOTOR FUN
TACTILE AND AUDITORY SENSORY PLAY
STRETCH, BEND, BUILD, AND CONNECT
HOME OR CLASSROOM FUN
8 BRIGHT COLORS
16 PACK

Generalized → many response distributions

Additive → adding the terms together

Models

Intro slides are plagiarized liberally from this excellent workshop webpage: https://eric-pedersen.github.io/mgcv-esa-workshop/slides.html

GAMs are not that different from linear regression (GLM)

They have most of the same components

GLM $$\underline{y_i} = \underline{\beta_0} + \underline{x_{1i}}\beta_1 + \underline{\epsilon_i}$$

GAM $$\underline{y_i} = \underline{\beta_0} + \sum_j s_j(\underline{x_{ji}}) + \underline{\epsilon_i}$$

Dependent variable

Intercept

Explanatory variable(s)

Error term

GAMs are not that different from linear regression (GLM)

This part is different, though.

GLM $y_i = \beta_0 + x_{1i}\beta_1 + \epsilon_i$

GAM $y_i = \beta_0 + \boxed{\sum_j s_j}(x_{ji}) + \epsilon_i$

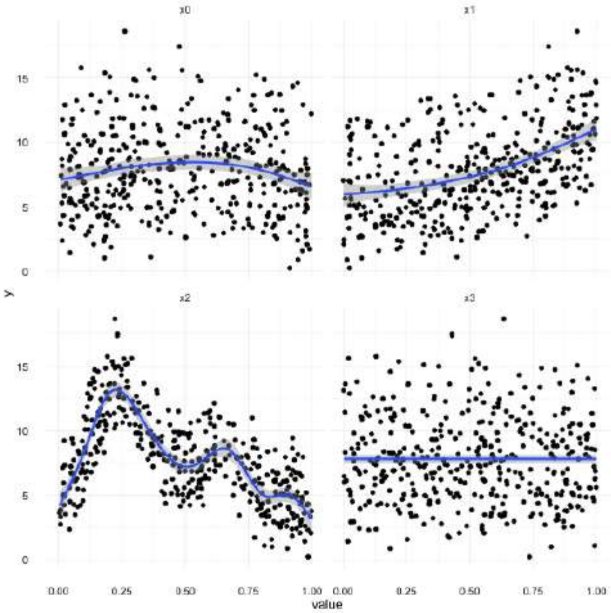
Dependent variable

Intercept

Explanatory variable(s)

Error term

Okay, but what about these ”s” things?

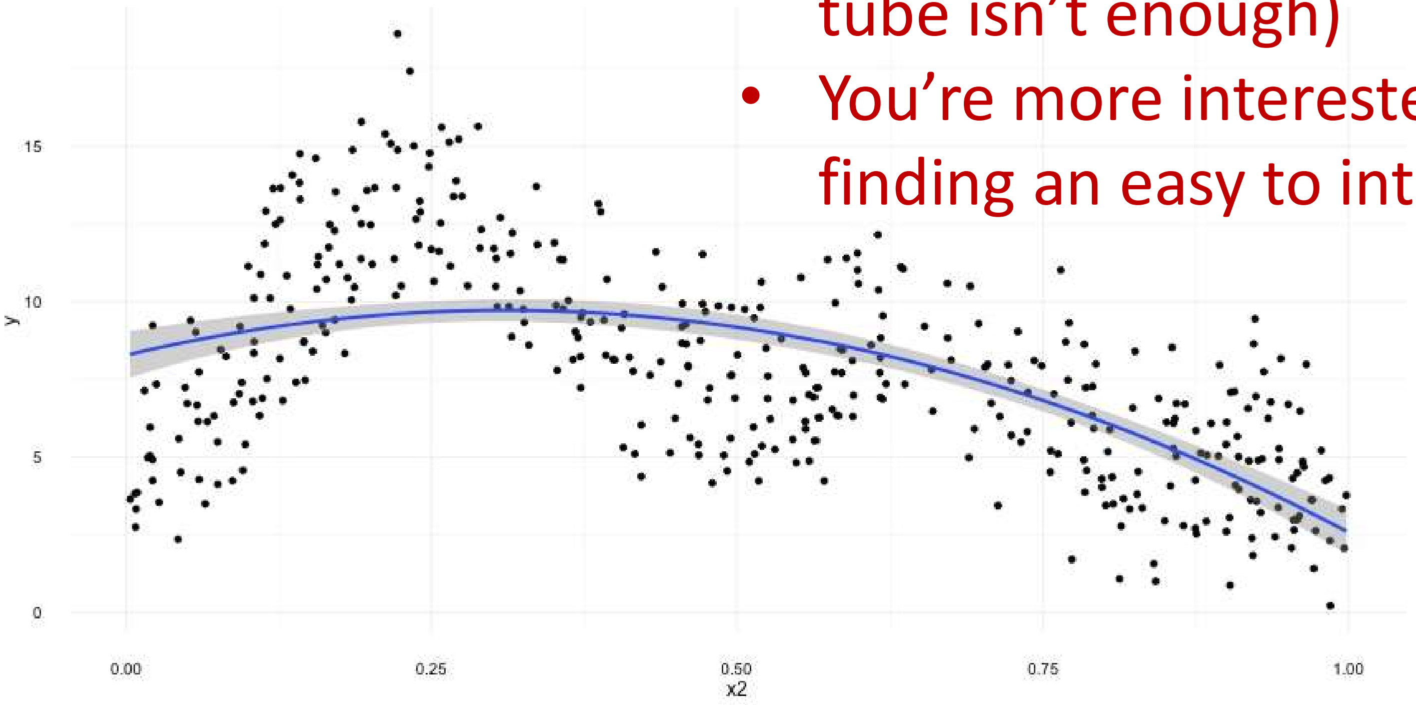


- Think s=**smooth**
- Want to model the covariates flexibly
- Covariates and response not necessarily linearly related!
- Want some “wiggles”

Intro slides are plagiarized liberally from this excellent workshop webpage: https://eric-pedersen.github.io/mgcv-esa-workshop/slides.html

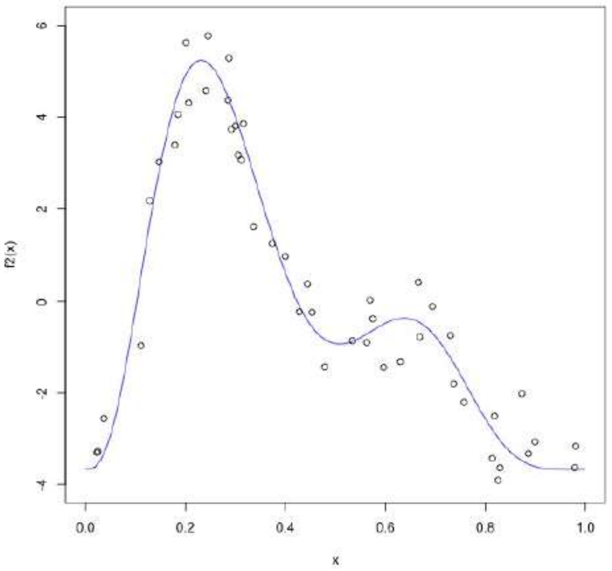
Why use a GAM instead of a GLM?

- Your data are too wiggly! (One bendy tube isn’t enough)
- You’re more interested in prediction than finding an easy to interpret explanation



Intro slides are plagiarized liberally from this excellent workshop webpage: https://eric-pedersen.github.io/mgcv-esa-workshop/slides.html

The Art of Smoothing

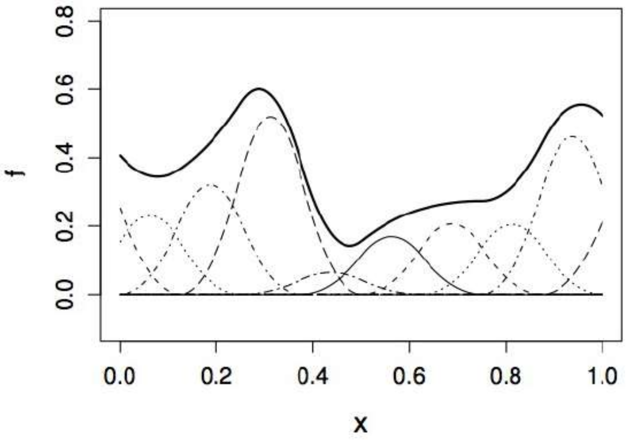


- Want a line that is “close” to all the data
- Don't want interpolation – we know there is “error”
- Balance between interpolation and “fit”

Intro slides are plagiarized liberally from this excellent workshop webpage: https://eric-pedersen.github.io/mgcv-esa-workshop/slides.html

Basis Functions

- Smooths are functions made of simpler functions
- You choose the maximum number (k) of basis functions to use & what type
- Types include:
  - Splines (so many kinds of splines!)
  - Cubic polynomials
  - Fourier series



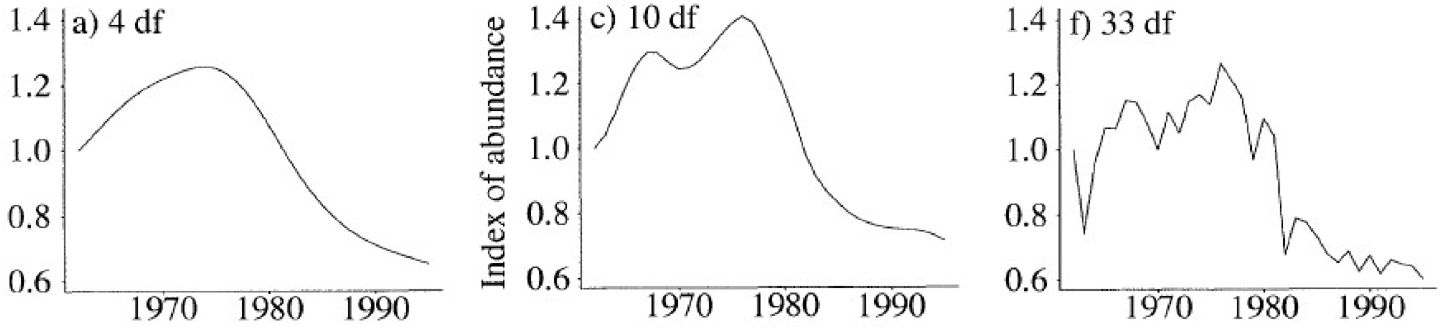
- Minimum: k = 2
  – s is a straight line
- Maximum: k = number of data points
  – s is an Interpolation

https://stat.ethz.ch/R-manual/R-patched/library/mgcv/html/smooth.terms.html

How to choose k

- Higher k → more wiggly line
- It is possible to have too many wiggles ("over fit") or too few ("over smooth")
- Choose your k to match your goals

Higher k implies higher degrees of freedom (df)



Graphs: Fewster et al. 2000

Think of a smooth function as being made up of bendy tubes. Then "k" is the number of sections of bendy tubes you need to describe your data.

What can you do with the output from a GAM?

STIMULATING CREATIVITY
The tactile feel of each tube is great for physical stimulation while the connectable, buildable design can be used to strengthen fine motor skills
CREATE UNIQUE SOUNDS*
These stretch tubes also become sound tubes! Push or pull them to different lengths, lightly swing them in circles, and kids can create music

* GAMs don't make sounds, but beep() from the beepr package does!

- Interpret patterns
- Make predictions with confidence intervals
- Select predictor variables
- Compare nested models
- Average a suite of models
- Functional Data Analysis (FDA)

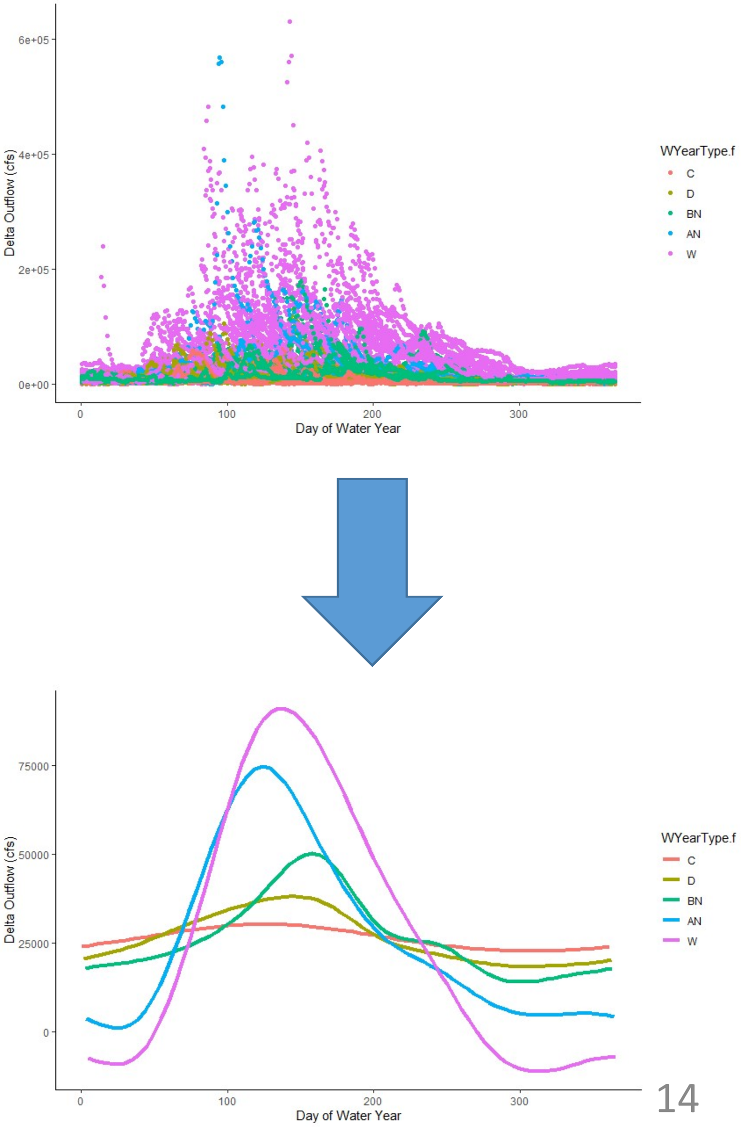
Image: https://www.mysensorytools.com/products/pop-tubes

Things to read!

- Fewster, RM, ST Buckland, GM Siriwardena, SR Baillie, JD Wilson. 2000. Analysis of population trends for farmland birds using generalized additive models. Ecology 81(7): 1970-1984. DOI: [10.2307/177286](https://doi.org/10.2307/177286)
- Pedersen EJ, Miller DL, Simpson GL, Ross N. 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ 7:e6876 https://doi.org/10.7717/peerj.6876
- Kain MP, Bolker BM, McCoy MW. 2015. A practical guide and power analysis for GLMMs: detecting among treatment variation in random effects. PeerJ 3:e1226 https://doi.org/10.7717/peerj.1226
- The documentation for mgcv::gam is very good for learning details about GAMs.

On to some R code!

- Data: dayflow
  - https://data.ca.gov/dataset/dayflow
  - Daily estimate of historical mean daily water flows relating to the Delta
- Packages
  - mgcv
    - https://cran.r-project.org/web/packages/mgcv/index.html
    - We'll mainly be using the function "gam" but there are others that you might want to look into.
  - lubridate – for using dates
  - tidyverse – for some plotting


6e+05
4e+05
2e+05
0e+00
Delta Outflow (cfs)
Day of Water Year
WYearType.f
C
D
BN
AN
W
Delta Outflow (cfs)
Day of Water Year
WYearType.f
C
D
BN
AN
W