Clustering GAM-Smoothed NFL Elo Ratings

STAT 448: Mixed Models

Paul Harmon

Fall 2017

What Are Elo Ratings?

Elo ratings are used by FiveThirtyEight to predict the outcome of head-to-head matchups between NFL teams. They were originally developed for prediction of chess matches but have been extended to sports analytics.

- A way to rank teams and predict outcomes
- Teams with higher Elo Ratings should beat teams with lower Elo Ratings
- Can be interpreted as point spreads for matchups (divide difference by 25 to get spread)
- Average Elo is 1500 for an NFL team, and they are calculated prior to every game

Why Do We Care About Classifying Elo Ratings?

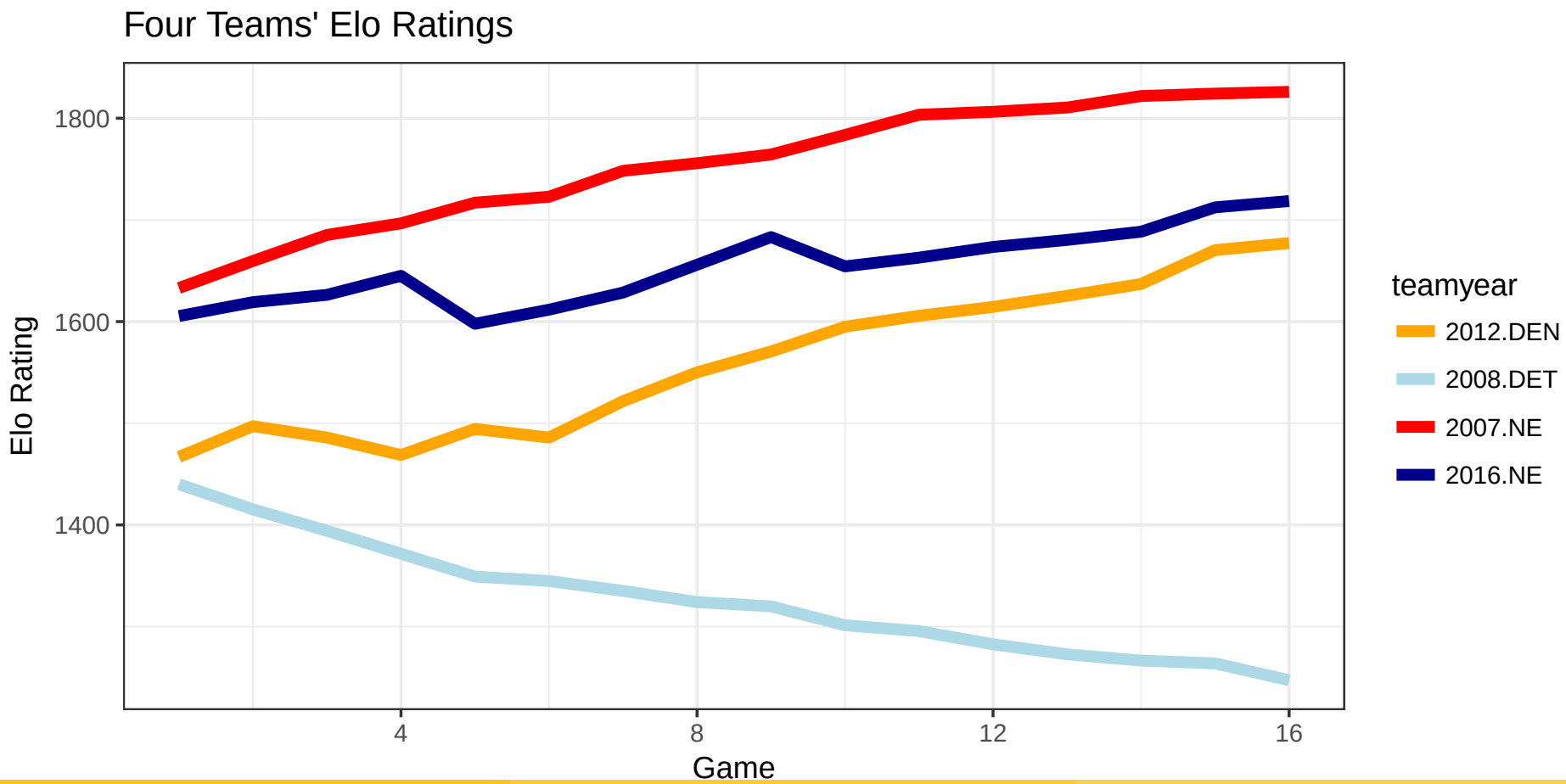
Team decision makers may want to develop a sense for the type of team that they have, given that not all teams realistically have a chance at winning a championship in every year. Sorry, Broncos and Browns fans...

- We want to make **comparisons** between different teams in different seasons.
- Allows for assessment of team **parity** over time.

Goal: To **smooth** season-long measurements of Elo ratings for each NFL team and **cluster** team-year Elos into 4 groups. Research by Hitchcock, Booth, and Casella (2007) indicates that smoothing improves cluster fits.

The Data

We have **32 teams** with 16 games per season over 15 seasons, leading to 480 curves that need to be smoothed with GAMs. Each team/year combination is considered independent.



GAM-Smoothed Estimates

- Elo Ratings can be noisy! It might be better to estimate a mean trend.
- I used GAMS to estimate smooth versions of these noisy trends in Elo.

A Very Complicated Model:

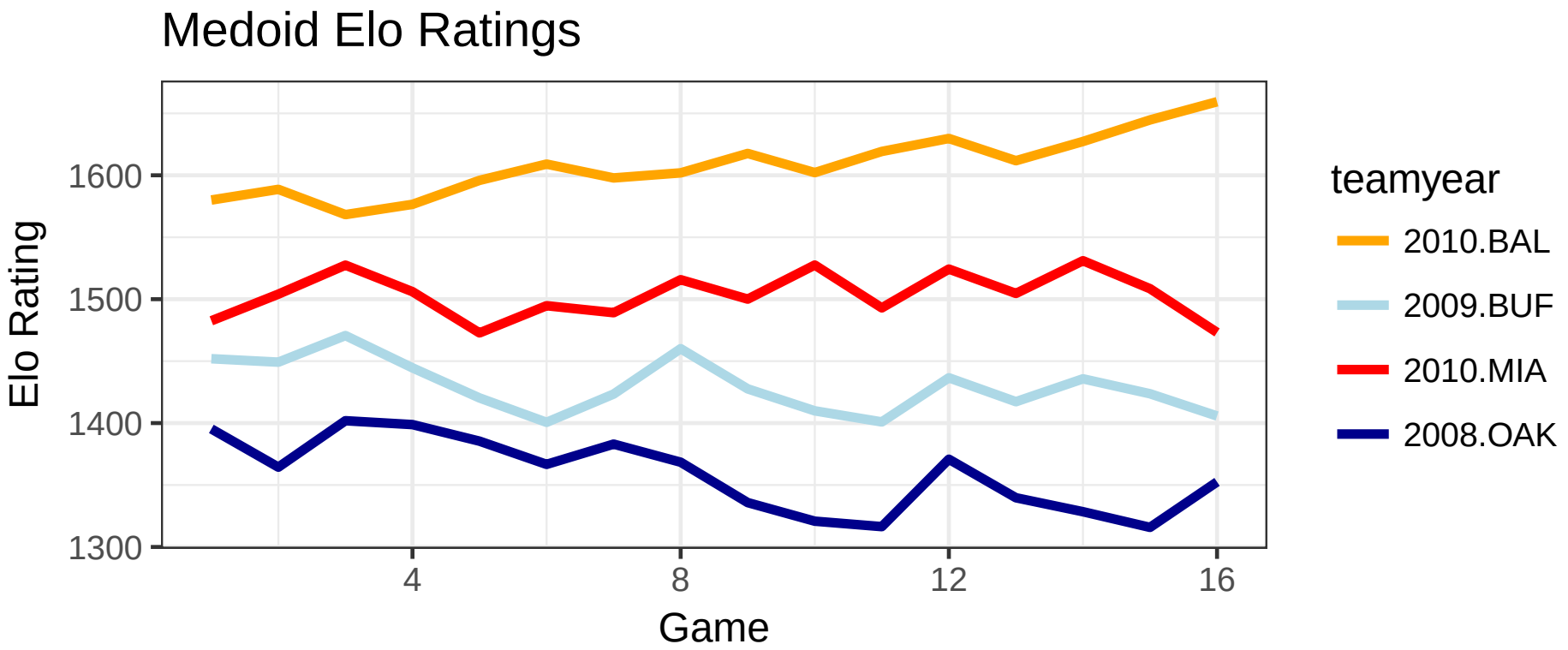
$$\hat{Elo}_{ij} = s_i(game_{ij})$$

Notationally, i refers to team/year combination (1..480) and j refers to game within season (1...16).

- Technically, there are 480 smoothed models fit here
- They may involve differing EDFs for each model depending on how wiggly things need to be
- Majority of the EDF were below 5.72, with 90 percent having small p-values

Functional Data clustering: 4-Cluster Solution

Contenders: 2010 Baltimore Ravens, **Status Quo**: 2010 Miami Dolphins, **Pretenders**: 2009 Buffalo Bills, **Loveable Losers**: 2008 Oakland Raiders



Selected References

Glickman, Mark E. and Jones, Albyn (1999).“Rating the Chess Rating System”. *Chance*. pp.21-28.

Silver,Nate (2014). “Introducing NFL Elo Ratings” _https://fivethirtyeight.com/features/introducing-nfl-elo-ratings_

Hitchcock, David B., Booth, James G., and Casella, George. (2007). “The Effect of Pre-Smoothing Functional Data on Cluster Analysis.” *The Journal of Statistical Computation and Simulation*. pp. 1043-1055.

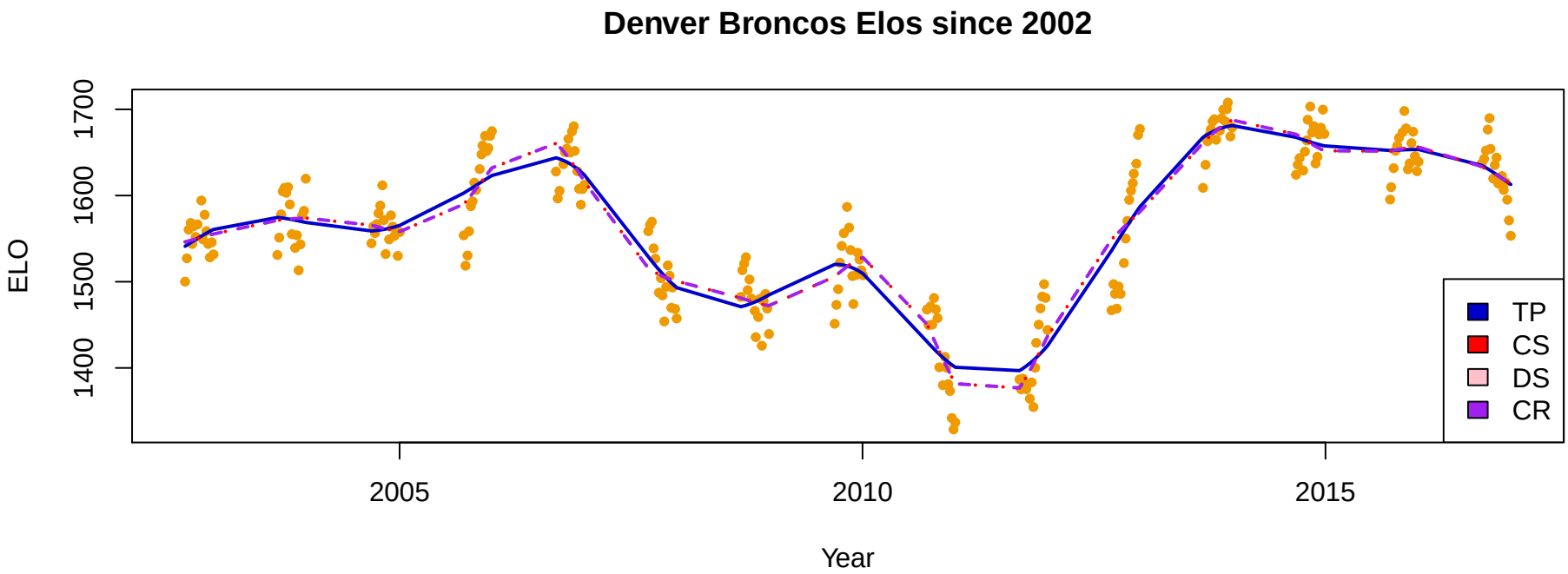
Wood, Simon.(2017). *Generalized Additive Models: An Introduction with R*. Ed 2.Chapman and Hall/CRC.

Questions

Figure 1

Supplementary Slides

Why not model long-term trends? We could ignore the fact that the Elos are normalized and try to model the long-term trends if we wanted to.



Some Interesting Team Results

	1	2	3	4
BUF	6	9	0	0
DEN	2	4	0	9
NE	0	0	0	15
NYG	2	10	0	3

Functional Data Clustering

We can then take these curves and calculate distances between each of the GAM smoothers. We can generate a 4-cluster solution based on a technique called **Partitioning Around Medoids** (PAM). The Idea:

- Determine 4 "medoid" GAM fits
- For each GAM, figure out which medoid is closest - that's the cluster that each observation goes in

Why Smooth?

Goal: To **smooth** season-long measurements of Elo ratings for each NFL team and **cluster** team-year Elos into 4 groups. Research by Hitchcock, Booth, and Casella (2007) indicates that smoothing improves cluster fits.