Approximating Dendrochronology Smoothing Splines Using Conventional Techniques

Eva Nolan[1], Emmeline Riendeau[2], and Dr. Nicholas Bussberg[3]

[1]Hamilton College, [2]Haverford College, [3]Elon University

Abstract

Dendrochronologists study tree rings to reconstruct past climatic conditions. The first step in processing tree-ring data is to use smoothing splines to separate climatic versus non-climatic trends. Smoothing splines are piecewise cubic polynomials which offer the flexibility needed to fit tree-ring data. A smoothing parameter is used to balance bias and variance in smoothing spline models—changing its value can make the models more linear or more interpolating. While smoothing splines are used in many other fields, the standard practice in dendrochronology is to choose the smoothing parameter using a method developed by Cook and Peters in 1981, which is a selection method unique to the field (Cook & Peters, 1981). This method will be referred to as the Cook and Peters method (CP method). For this method, the Fourier transform is employed to filter out the different frequencies present in the growth trend. However, as the Cook and Peters method is so specific to dendrochronology, a traditional spline method for choosing the smoothing parameter may be more widely understood, making tree-ring data processing more accessible to a wider audience. This project aims to find equivalence between traditional smoothing spline selection methods and the Cook and Peters method. Comparing the performance of multiple smoothing parameter selection methods with the Cook and Peters method determined that for specific values the traditional selection method degrees of freedom produces approximately equal splines to the Cook and Peters method. Additionally, this research identifies a direct relationship between the smoothing parameters chosen by the Cook and Peters method and this degrees of freedom method.

Introduction

- Tree core samples used to gather ring width data
- Width influenced by climate, age, stand dynamics, and disturbances
- Data is standardized to isolate climatic variance
- Smoothing splines are flexible method to fit complex growth trends
  - Piecewise function of cubic polynomials
  - Smoothing parameter (λ) balances closeness and smoothness of fit
- Smoothing parameter selection is mathematically dense
  - Can be approximated with Frequency Response Curves
- **Goal: Find relationship between the dendrochronology smoothing parameter selection method and conventional selection methods**

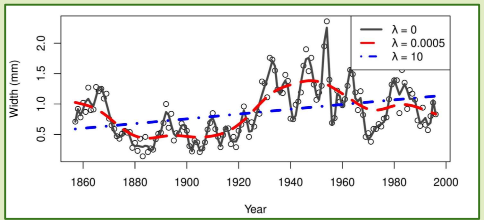

Figure 1: Effect of λ on spline fit

Materials & Methods

- dplR package
  - *detrend* function
  - Original Cook and Peters method (CP method)
- Base R Stats package
  - *smooth.spline* function
  - Traditional Smoothing splines
- Root mean squared difference (RMSD) for comparing function performance

Results

Estimating CP method splines using degrees of freedom

- Most common frequency response (*fr*) used in CP method is 0.5
  - Degrees of freedom (*df*) of 4 closely matches this default spline
  - See Figure 2 for example
- Different *fr* can be mimicked by *df* spline matches
  - *df* increments by 1
  - Each *df* choice is assigned a specific *fr* that creates the most similar spline (Figures and 4 show best and worst case fits for these matches)
  - Each *df* choice is assigned an *fr* range that it best approximates

Best Frequency Response (fr) given degrees of freedom (df)

df	fr range	closest fr	Mean RMSD
2	0.01 - 0.03	0.01	4.68×10^{-5}
3	0.04 - 0.25	0.11	4.92×10^{-7}
4	0.26 - 0.54	0.40	2.25×10^{-7}
5	0.55 - 0.76	0.68	1.01×10^{-6}
6	0.77 - 0.87	0.84	2.07×10^{-6}
7	0.88 - 0.93	0.91	2.41×10^{-6}
8	0.94 - 0.96	0.95	9.09×10^{-7}
9	0.97 - 0.99	0.97	1.11×10^{-6}

Table 1

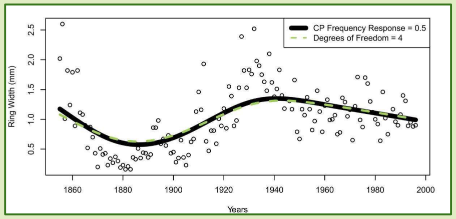

Figure 2:
Spline match producing the worst RMSD among matches for the default *fr* choice of 0.5

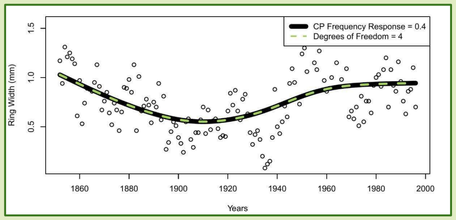

Figure 3:
Spline match producing the lowest RMSD for the swed320 dataset

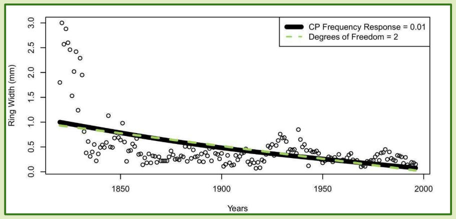

Figure 4:
Spline match producing the worst RMSD amongst the "closest *fr*" range of frequency responses

Results

Calculating *detrend* λ using degrees of freedom = 4

- More precise method
  - Can calculate *detrend* λ (λ_d) using *smooth.spline* λ when *df* = 4 (λ_{s4})
  - Equation 1 works when *fr* = 0.5
  - Table 2 contains similar equations (and their accuracy) for different *df*s

$$\lambda_d = (781\lambda_{s4} - 0.439)^4$$

Equation 1:
Converting λ_{s4} to λ_d

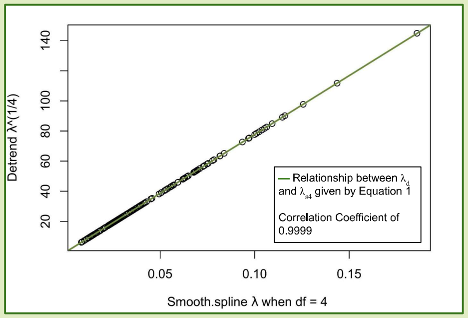

Figure 5:
Relationship between λ_d and λ_{s4}

df	Equation	Correlation
2	$\lambda_d = (-0.04024\lambda_{s2} + 29.2684)^4$	-0.857
3	$\lambda_d = (153.86\lambda_{s3} - 0.4433)^4$	0.999
4	$\lambda_d = (782.25\lambda_{s4} - 0.4370)^4$	0.999
5	$\lambda_d = (2471.18\lambda_{s5} - 0.4416)^4$	0.999
6	$\lambda_d = (2471.18\lambda_{s6} - 0.4416)^4$	0.999

Table 2: Converting λ_s to λ_d for various values of *df* when *fr* = 0.5

- Equations can be altered to accommodate different *fr* choices
- Multiplying any equation from Table 2 by an identical first term will adapt it to any desired frequency responses.
  - Example shown in Equation 2

$$\lambda_d(fr) = \left(\frac{1}{1-fr} - 1\right)(781\lambda_{s4} - 0.439)^4$$

Equation 2: Conversion when df=4 for desired frequency response

Inability to mimic *detrend* with Generalized Cross-Validation (GCV)

- Average GCV smoothing spline fit *fr* = 0.95
- Most common *fr* are 0.99 and 1.00; Table 3 provides an example of this
- Can be large variations in *fr* created, shown in Table 4
- Ultimately, *fr* is too high and varies too much for GCV to be an adequate substitute for the CP method

Core ID	RMSD	Equivalent *fr*
HBTH0039	0.03021	0.99
HBTH0049	0.01273	1.00
HBTH0059	0.01381	1.00
HBTH0069	0.01871	0.99
HBTH0079	0.02007	1.00
HBTH0089	0.00428	0.99
HBTH0099	0.02734	1.00

Table 3:
Sampled GCV *fr* for dataset swed320

Core ID	RMSD	Equivalent *fr*
BAO09A	0.00220	0.99
BAO09B	7.53279×10^{-9}	0.10
BAO10A	1.34091×10^{-5}	0.01
BAO10B	1.47302×10^{-7}	0.18
BAO10C	0.00904	0.99
BAO11A	2.17830×10^{-5}	0.01
BAO11C	2.32128×10^{-6}	0.01

Table 4:
Sampled GCV *fr* for dataset zimb001

Conclusion

- This research details two versatile alternative smoothing parameter selection methods using degrees of freedom to the CP method
  - Smoothing Spline Fit Approximation using *df* in *smooth.spline*
    - Close approximation
    - Provides some control over *fr*
  - Conversion equation to calculate CP method smoothing parameter
    - Gives precise smoothing parameter value
    - Provides significant control over *fr*
- Benefits of the Degrees of Freedom Method
  - Can be used in the *smooth.spline* function
    - *Smooth.spline* is available in the base-R stats package making it very accessible
    - *Smooth.spline* returns much more detailed spline fit information than *detrend*
  - Uses traditional smoothing parameter selection method
    - More widely used giving researchers better understanding of their research

Discussion and Next Steps

- The correlation of 0.9999 in Figure 3 suggests that a direct proof could be written to connect the Cook and Peters smoothing parameter selection method and the degrees of freedom method. Doing so would strengthen this research's findings.
- This research only focused on mimicking CP method spline fits with the 67%*n* criterion as that is the most common option used. However, researchers can select between 30%*n* and 75%*n*. Further testing could be performed to find ways to mimic CP method splines with different %*n* values. This would add to the versatility of the degrees of freedom method.
- The method of testings of *fr* values resulted in gaps between *fr* ranges in Table 1. Further testing with smaller iterations could be performed to shrink or eliminate these *fr* range gaps.
- This research focused on the original CP method; however Bussberg et. al. (2020) identified a complete derivation of the CP method. Further work could be done to establish a relationship between the complete CP method and this degrees of freedom method.

References

Bunn, A., Korpela, M., Biondi, F., *dplR: Dendrochronology Program Library in R.* R package version 1.7.2 2021

Bussberg, Nicholas W. Environmental Applications of Temporal, Spatial, and Spatio-Temporal Statistics. July 2020. Indiana University, PhD dissertation.

Cook, E.R., Peters, K. 1981. The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. Tree-Ring Bulletin 41:45-53.

DeYoung, Joan. "4.5: Increment Boring." Biology LibreTexts, Libretexts, 14 July 2020, bio.libretexts.org/Bookshelves/Botany/Forest_Measurements_-_An_Applied_Approach_(DeYoung)/04%3A_Tree_Age/04.5%3A_Increment_Boring.

H. Linderholm, *Noaa/wds paleoclimatology - linderholm - stockholm - pisy - itrdb swed320*

R Core Team, *R: A Language and Environment for Statistical Computing,* R package version 4.1.0, Foundation for Statistical Computing, Vienna, Austria, 2021.

Stahle, D.W.; Gabor, D.; Haynes, G.; Klimowicz, J.; Ngwenya, P. (2005-08-25): NOAA/WDS Paleoclimatology - Stahle - Baobab Holes - CABU - ITRDB ZIMB001. NOAA National Centers for Environmental Information. https://doi.org/10.25921/c3wk-1988. Accessed 06-2021.

Acknowledgements

This research was conducted as a part of the North Carolina A&T State University and Elon University Joint Summer REU Program in Mathematical Biology and was funded by the National Science Foundation DMS# 1851957/1851981.

Thank you to Dr. Karen Yokley and Dr. Nicholas Luke for their guidance and leadership in the REU.