Approximating Dendrochronology Smoothing Splines Using Conventional Techniques

Emmeline Riendeau

Haverford College

Eva Nolan

Hamilton College

Dr. Nicholas Bussberg

Elon University Department of Mathematics and Statistics

July 2021

Approximating Dendrochronology Smoothing Splines Using Conventional Techniques

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. [9]. 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.

1 Introduction

1.1 Dendrochronology

This research is centered in the field of dendrochronology, which is essentially the study of tree rings. Analysis of tree rings can be used to date archaeological artifacts and major historical weather events, and to reconstruct past climatic conditions [8]. This research focuses on the use of tree-ring data for climate reconstruction. The nature of tree-ring data is

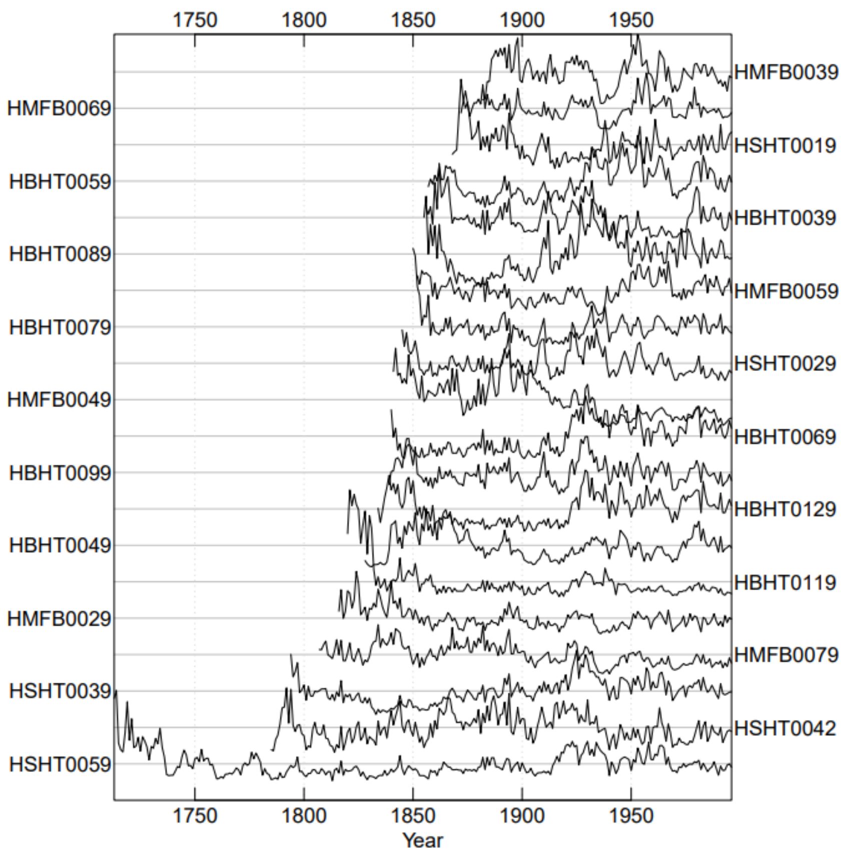

Figure 1:
An example tree-ring chronology depicting a sample of cores (stacked on the y-axis) aligned by date (x-axis). Cross dating—finding similar recurrent patterns in tree rings—allows dendrochronologists to build tree-ring chronologies that span far into the past. This plot type can be used to visually identify similarities in the cores' growth trends and also possible disturbances. The amplitude of a line represents the ring width, which is typically measured in millimeters. Data collected from Stockholm, Sweden as part of the Swed320 dataset [14].

advantageous to reconstructing past climate; tree-ring data can span farther back in history than human climate records (see Figure 1), and tree growth is dependent on climate, allowing tree rings to serve as a proxy climate record.

Each year, a tree adds a new layer of growth around its perimeter. Each year's growth

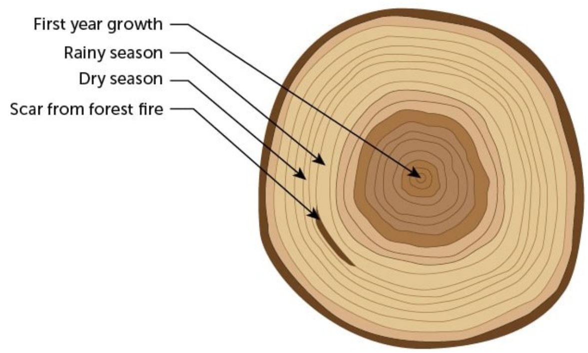

Figure 2:
A tree-ring diagram illustrating the growth pattern of tree rings and the effects of different influences on growth [6].

is separated from the next by a thin dark line—these demarcations create the tree rings. The oldest tree ring is at the center of a tree, and the youngest tree ring is the outermost layer of a tree. To study tree rings, dendrochronologists use a tree coring tool to remove a pencil-width sample that runs at least halfway through the tree. This allows dendrochronologists to measure the width of each tree ring, without having to cut the tree down (illustrated tree cross-section is shown in Figure 2).

Tree rings vary in width from year to year. For years when the tree grows more, tree rings are thicker, and for years when the tree grows less, tree rings are thinner. Climate can influence the amount a tree grows. Warmer and wetter conditions may support tree growth. For example a tree may grow more under these conditions, resulting in thicker rings. Conversely, in cooler, drier years, rings may be thinner. Analyzing the change in tree ring widths over time can be used to reconstruct past climate. However, as climate is not the only factor that influences ring growth, other non-climatic factors need to be removed from the data first to isolate climate data.

One non-climatic factor that needs to be considered is the natural declining growth trend of trees. Simple geometry helps explain this phenomenon. Suppose a similar amount of new volume is grown each year by a given tree, but as the tree ages this volume is spread over an increasingly large circumference, resulting in decreased ring-width over time. This trend can be modeled with a negative exponential equation. In semi-arid environments where trees are not in close proximity to each other—known as open canopy forests—this model adequately fits tree ring growth over time [8].

This natural declining growth trend can be present in closed-canopy forests, but there are additional factors that make growth trends in these forests more complex. Closed canopy forests are forests where the trees are in close proximity to each other—examples include forests in Northeast America and Europe. The closeness of the trees causes stand dynamics—competition for resources among a group of trees. An example of stand dynamics is the competition for sunlight [8]. A taller tree may block sunlight from smaller trees in close proximity, suppressing their growth. When this taller tree eventually dies and falls down, the trees around it have a sudden increase in access to sunlight and experience a spike in growth. These disturbances cause a complex growth trend with peaks and valleys that a negative exponential does not adequately model.

It is crucial to be able to model tree growth trends in closed-canopy forests, as they represent a large portion of forests. Without an adequate way to model tree growth trends in these forests, dendrochronologists are blocked from accurately studying climate in these areas. In 1981, Cook and Peters revolutionized tree ring growth modeling in closed-canopy forests with the implementation of smoothing splines. Smoothing splines create a more flexible model that can adequately fit more complex growth trends.

1.2 Traditional Smoothing Splines

The term spline comes from the "thin flexible strips used in drafting for interpolating new values" [9] (Figure 3). Spline functions in math are used in the same way—for interpolation. In 1967, Reinsch developed a variation on traditional splines, called the smoothing spline function. Reinsch's variation allows for more smoothing, which is especially important for identifying trends in experimental data[17].

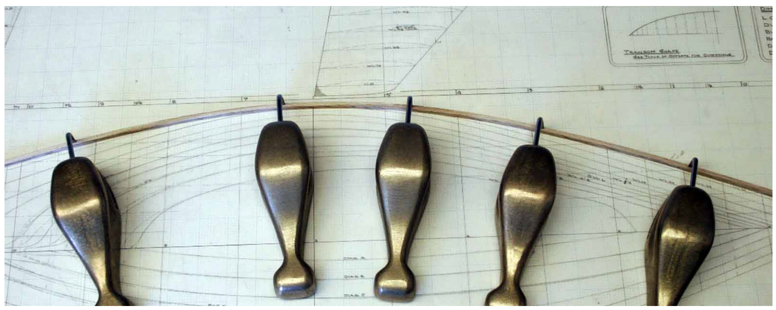

Figure 3:
A physical drafting spline bent to model a smooth curve [15].

In traditional smoothing splines the following expression is minimized:

$$\sum_{i=0}^{n-1}[g(x_i) - y_i]^2 + \lambda \int_{x_o}^{x_{n-1}} [g^{''}(x)]^2 dx,$$

where $g(\cdot)$ is fitted function, x_i is the raw independent data, y_i is the raw dependent data, and n is the total number of data points. When used in tree-ring analysis x_i is the year, y_i is the ring width for a given year, and n is the number of rings in the core, which is equivalent to the number of years in a tree's life span.

The first term in the expression calculates the residual between the experimental data and the model. This term is responsible for the closeness of the spline fit. If this term were minimized on its own, $g(x)$ would fit the data perfectly as an interpolating function. This term is identical to the minimization expression used to determine linear regression fits, in which $g(x)$ is forced to be linear.

The second term of the expression can be thought of as an aggregate measure of the curviness of the model. The second derivative of the spline function $g(x)$ is taken at each point for two reasons. The second derivative gives a measure of the curviness of a function. Thus taking the summation of $(g''(x))^2$—squaring to avoid negatives and positives canceling—gives the total measure of curivness of $g(x)$. Additionally, taking the second derivative ensures a continuous function $g(x)$. Ultimately, as this term is being minimized, it is responsible for the smoothness in the expression.

The last element of the minimization expression is the smoothing parameter λ. By

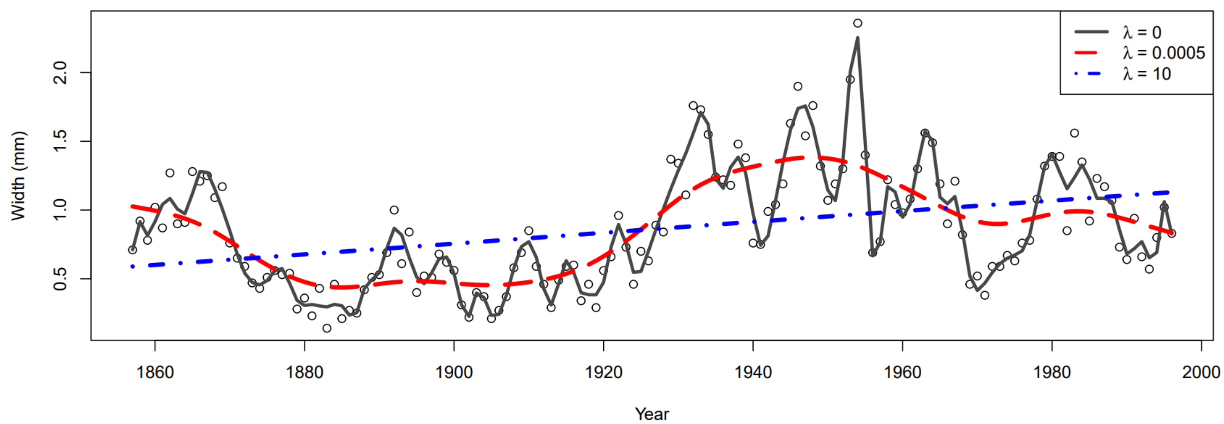

Figure 4: Three splines with varying λ values fit to tree-core HBTH0059 from the Swed320 dataset from Stockholm, Sweden[14].

weighting the relationship between the two terms of the expression, λ balances how smooth the spline is with how closely it fits the data. Therefore, λ also controls the balance between variance and bias in the model. Due to the nature of minimization, λ acts as a penalty to curviness. The larger λ is, the straighter and more regression-like the smoothing spline function will be. Conversely, the smaller λ is, the curvier and closer fitting the spline function will be to the data. Figure 4 illustrates the effect of different λ values.

There are several common ways to select the smoothing parameter λ in traditional smoothing splines. Our research focuses on two of these procedures, degrees of freedom and generalized cross-validation, which are detailed in Sections 2.3 and 2.4 respectively.

1.3 Cook and Peters Method

In their 1981 paper, Cook and Peters detailed the application of smoothing splines in the field of dendrochronology[9]. A more detailed report in given in Peters and Cook (1981).Before smoothing splines were used in the field, the complex growth trends in closed canopy forests were poorly modeled, if modeled at all. The flexibility and control offered by smoothing splines now presented a way to adequately fit these more complex growth trends. The minimization expression used by Cook and Peters is the following:

$$2p\sum_{i=0}^{n-1}[g(x_i) - y_i]^2 + \int_{x_o}^{x_{n-1}} [g^{''}(x)]^2 dx$$

As we will discuss, the Cook and Peters approach (abbreviated CP method for convenience in this paper) to smoothing splines is slightly different than traditional smoothing splines. The expression that the CP Method uses is a slight variation on standard smoothing splines. In the CP Method the smoothing parameter p is placed on the summation term in contrast to standard smoothing splines where the smoothing parameter λ is placed on the integral term. A minimization expression can be multiplied by a constant without altering the result; therefore the CP method variation does not affect the resulting spline fit. Hence the constant 2, which is added in some variations of the CP method to simplify mathematical derivations, does not affect the resulting spline fit either.

The smoothing parameter p is different from the traditional smoothing parameter λ. In the CP method, p weights the first term, meaning that increasing p will put more emphasis on closeness of fit, opposite to the relationship seen with λ in traditional splines. Therefore the larger p is the curvier the spline fit is, and the smaller p is the straighter the spline fit is. Later in the paper, λ_d will be discussed, which is a conversion of p into an equivalent form of the λ seen in traditional smoothing splines, referred to in this paper as λ_s. To achieve this conversion, the entire minimization expression is divided by $2p$, so $\lambda_d = \frac{1}{2p}$

The main difference between the CP method and traditional spline procedures is how the smoothing parameter is chosen. The smoothing parameter selection technique employed by the CP method was developed specifically for dendrochronology and is not used in other fields. The process utilizes the Fourier transform, which converts data from the time domain to the frequency domain, allowing the different frequencies present in the data to be decomposed (Figure 5 gives a visual representation of the Fourier transformation). After pulling apart the different frequencies in the data, the CP method typically chooses to reduce the frequency response by 0.5 of a frequency of interest. A frequency response of 0.5 means an amplitude of 1, for example, would be reduced to an amplitude of $\frac{1}{2}$. Note that frequency response is typically denoted as a decimal, for example a 50% reduction in frequency response is often referred to as a frequency response (fr) of 0.5. Since climate data is low frequency and disturbances have high frequencies, reducing the frequency response aims to remove high frequency disturbances while preserving low frequency climate data.

The closed form solution provided by the Cook and Peters method is as follows,

$$p = \frac{(6(cos(2\pi f) - 1))^2}{(cos(2\pi f) + 2)},$$

where p is the smoothing parameter and f is the frequency of the core data. In this closed form solution the frequency response is assumed to be 0.5. A frequency response of 0.5 is most commonly used.

The equation is a mathematically convenient calculation of p; however, it is not the complete complete solution (for more details, see Bussberg et al. 2020 [5]). Since there is a known relationship between the complete method and this closed-form solution, and the closed-form solution is easier to work with, this research works with the closed-form solution.

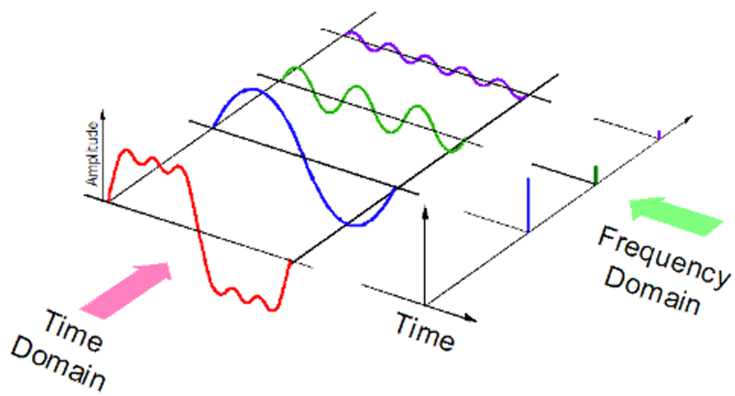

Figure 5:
A visual representation of the Fourier transform. The red plot at the front is the amalgamation of frequencies, i.e., the full time series being analyzed. The Fourier transform separates out all frequencies that contribute to the red plot, and identifies their amplitudes [7].

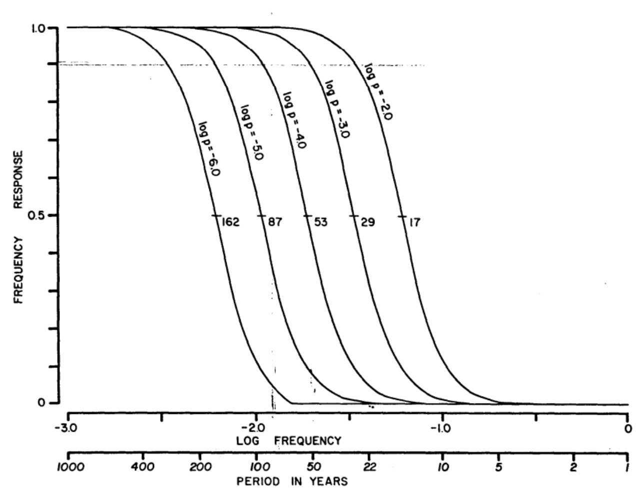

Figure 6:
Frequency response curves (graph from Cook and Peters, 1981) that are used to estimate the smoothing parameter p that would be calculated by the CP method. Five different smoothing parameters and their associated curves are represented. To use, the period in years and the desired frequency response are chosen, and the curve they intersect on indicates the smoothing parameter that should be used. [9]

If frequency response other than 0.5 is desired the following equation can be used,

$$fr = 1 - \frac{1}{1 + \frac{p(cos(2\pi f)+2)}{p(cos(2\pi f)-1)^2}}$$

where p is the smoothing parameter, fr is the desired frequency response and f is the frequency of the core data. Note in Cook and Peter's (1981) original notation, fr is referred to as $u(f)$ in their Equation 2.

The frequency of the tree core data is $\frac{1}{n}$, where n is the series length, also known as the period. When calculating the frequency, researchers use a $\%n$ criterion that is typically chosen to be between 30% of the series length and 75% of the series length. Dendrochronologists can, however, choose any $\%n$ they wish. In this research $67\%n$ is used as this is the most common $\%n$ used for the CP method.

As an example of how to select the smoothing parameter using the Cook and Peters method, say that a tree-core sample is 150 years old and a researcher is looking to fit a smoothing spline with a frequency response of 0.5 with $67\%n$. First 67% would be taken of the series length, giving a period of 100 years. This period would be converted to frequency by taking the inverse, giving a frequency of $\frac{1}{100}$. Since the research is using a frequency response of 0.5 they can simply use the closed form solution to solve for p. Then, they simply plug in the frequency found above of $\frac{1}{100}$. If the researcher desired a different frequency response they would have to use the *fr* equation and solve for p. This researcher could also use the frequency response curves in Figure 6 to estimate the smoothing parameter. The researcher would find where on the graph 0.5 on the y-axis and 100 on the x-axis meet, and the curve that crosses this intersection would give them their smoothing parameter.

2 Methods

2.1 Tree Core Data

Tree core data for this research was sourced from the International Tree-Ring Data Bank. This data bank is managed by the National Center for Environmental Information's Climatology Team and the World Data System for Climatology under the National Oceanic and Atmospheric Administration [1]. Data sets provide the width of tree rings with the corresponding year for all tree-cores sampled in a stand.

Graphs and models in this paper primarily use dataset Swed320. Swed320 is a closed canopy tree core data set collected from Stockholm, Sweden by H.W. Linderholm [14].

Analysis was tested on a total of 9 different tree-stands for robustness. These data sets include closed-canopy and open-canopy forests. Though smoothing splines are not commonly used in open-canopy settings, open-canopy data can still be used to check the robustness of the results.

2.2 R software

This research uses R version 4.1.0 [16]. The R packages used are the stats package from base-R and the dplR package [4].

The stats package is used primarily for the function smooth.spline. This function creates smoothing splines using the traditional smoothing spline minimization expression. Users can choose from multiple different smoothing parameter selection methods including degrees of freedom and generalized cross-validation.

The dendrochronology specific package dplR is used primarily for the functions read.rwl and detrend. In dendrochronology, it is standard to store tree-ring datasets as rwls, which have a unique formatting. Read.rwl helps to read these unique datasets into R. Detrend creates spline fits using the Cook and Peters smoothing spline minimization expression and smoothing parameter selection method.

The two functions, smooth.spline and detrend, can be manipulated to create equivalent spline fits. However, smooth.spline may be preferable because of its accessibility, its multitude of options for selecting the smoothing parameter, and the detailed information it returns about spline fits. The root mean squared difference test (RMSD) was used to compare these two functions and determine how the two functions can be used to obtain equivalent splines.

2.3 Degrees of Freedom

Degrees of freedom is another method of selecting the smoothing parameter in traditional smoothing splines. A larger value of degrees of freedom indicate more flexibility for the spline thus creating a more interpolating spline. Conversely, a smaller value of degrees of freedom indicate less flexibility for the spline thus creating a more linear spline. The equation for degrees of freedom is defined as,

$$DF = tr(S_\lambda),$$

where S_λ is an $n \times n$ smoother matrix, also known as the hat matrix (the form for S_{lambda} can be found in [2]).

Exploratory graphs were created using the degrees of freedom (*df*) option in the R function smooth.spline. During this exploratory phase, it quickly became apparent that the smoothing splines produced by a *df* of 4 are very similar to those created using the CP method with a frequency response (*fr*) of 50% with 67%n. Recall that 67%n means the frequency of interest is calculated using a period that 67% of the series length. Additionally, recall that a *fr* of 0.5 means the frequency of interest is reduced to 50% of its original amplitude. For notational simplicity the *fr* and %n criterion used will be referred to as, *fr* of x with $y\%n$, where x and y are given values.

Further testing was performed to quantify how close $df = 4$ splines matched those created using the CP method with a fr of 0.5 with $67\%n$. Smoothing splines were fit to multiple tree stands using both $df = 4$ for the smooth.spline function and $fr = 0.5$ with $67\%n$ for the detrend function. The root mean squared difference (RMSD) was used to compare the two different sets of spline fits.

The research was then extended to determine the fr of splines created using other values of df. One set of smoothing spline fits was created using the CP method with fr values between 0 and 1 in iterations of 0.01. Another set of smoothing spline fits was created using df values between 0 and 9 in iterations of 1. These two sets of spline fits were subsequently compared using the root mean squared difference test or RMSD. The RMSD was calculated between each fr smoothing spline and each df smoothing spline. The df value that produced the smallest RMSD for a given fr value was then considered the best choice.

2.4 Generalized Cross-Validation

Generalized cross-validation is a traditional smoothing spline method for selecting the smoothing parameter (see example Wahba 1990 [19]). Generalized cross-validation (GCV) essentially splits the raw data into two groups, a group for learning and a group for testing. The learning group is used to create the model and the testing group is used to determine the quality of the model's fit. The equation for GCV is as follows,

$$GCV(\lambda) = n^{-1}\frac{\sum_{i=1}^{n}(y_i - g(x_i))^2}{(1 - n^{-1}trS_\lambda)^2}$$

where n is the series length, y_i is the raw independent data, $g(\cdot)$ is the spline function, and S_λ is an $n \times n$ smoother matrix, also known as the hat matrix (the form for S_{lambda} can be found in [2]).

In the present study, GCV was tested first as a substitute for the Cook and Peters method, because GCV is one of the most popular techniques for smoothing parameter selection.. Using smooth.spline with the option cv set to false, preliminary spline fits were created using GCV to calculate the smoothing parameter as seen in Figures 7 and 8. In these preliminary plots, the smoothing spline fits were more interpolating, suggesting a high frequency response.

This research then aimed to find the exact frequency response (*fr*) of smoothing splines

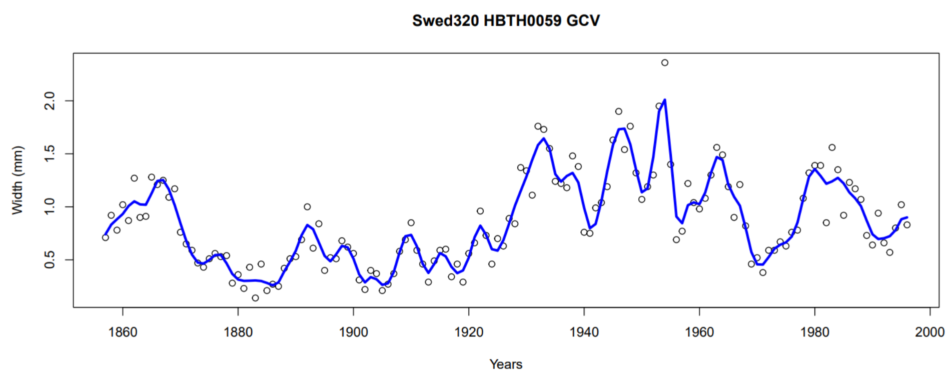

Figure 7: Smoothing spline for dataset Swed320 tree core HBTH0059 using GCV to select smoothing parameter [14]

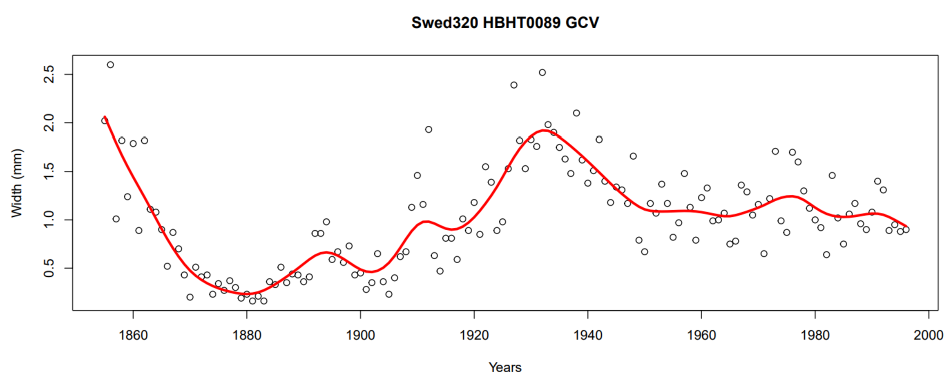

Figure 8: Smoothing spline for Swed320 tree core HBTH0089 using GCV to select smoothing parameter [14]

created using GCV to select the smoothing parameter. To find the exact *fr*, GCV spline fits were compared to CP spline fits with varying *fr*. For each tree core being analyzed, a smoothing spline was fit to the data using GCV to select the smoothing parameter. Additionally, for each tree core being analyzed, smoothing splines using the CP method to calculate the smoothing parameter, with *fr* values from 1 to 100 in increments of 1, were fit. In this set of smoothing spline fits 67%n was used in the smoothing parameter calculation.

After the two sets of smoothing spline fits were created, they were compared using the root mean squared difference test (RMSD). RMSD essentially tests the difference between two sets of data. The closer RMSD is to zero, the lower the difference between the two sets. The CP method spline with the lowest RMSD to the GCV spline was considered equivalent and the corresponding *fr* was noted. Ultimately this process established the *fr* of each GCV created smoothing spline tested. This process was tested on multiple different closed canopy tree-ring data sets for robustness.

3 Results

Degrees of freedom and generalized cross-validation, both traditional smoothing spline smoothing parameter selection methods, were tested as substitutes for the Cook and Peters method. This process revealed that degrees of freedom is an adequate substitute as it can consistently create equivalent smoothing splines to the Cook and Peters method, for multiple different frequency responses. Conversely, this process found that generalized cross-validation is not a good substitute for the Cook and Peters method because it primarily yields a high frequency response and lacks consistency.

3.1 Degrees of Freedom

There are two ways to use degrees of freedom to create smoothing splines equivalent to those produced by the Cook and Peters method. The first way is to use the R function smooth.spline and simply choose a spline with a particular number of degrees of freedom. The second, more precise way is to use a conversion equation on the smoothing parameter calculated by smooth.spline given a particular df. The equation created through this research connects the Cook and Peters method smoothing parameter, λ_d, with the smoothing parameter produced by degrees of freedom in traditional smoothing splines, λ_s. The first way of using degrees of freedom is faster and simpler, but the second method is more precise.

3.1.1 Approximating Cook and Peters' Smoothing Splines using Degrees of Freedom

Setting the degrees of freedom in the R function smooth.spline can create approximately equivalent smoothing splines to the Cook and Peters method. A very commonly desired frequency response used in the Cook and Peters method is 0.5 in conjunction with $67\%n$. This research finds that using smooth.spline with degrees of freedom set to 4 creates a very close approximation of the CP method smoothing spline when using a frequency response of 0.5 using $67\%n$.

Testing of this degrees of freedom method on the Swed320 data set serves as an example of the method's ability to closely approximate CP method smoothing splines. The root mean squared difference between the smoothing splines with four degrees of freedom and the CP method using a frequency response of 0.5 using $67\%n$ was on average 2.53×10^{-4}, supporting that the two method's smoothing splines are nearly identical (Figure 9). All tree-ring series tested behaved in a similar manner; the highest RMSD observed in the Swed320 dataset with this combination of df and fr was 1.11×10^{-3}, which on the scale of the ring data is still a close fit. This can be seen in Figure 9, which is detrend spline, created with default parameters, that had the largest RMSD when compared with the $df = 4$ spline.

Further results are detailed in Table 1, which shows the frequency response each degrees of freedom choice in smooth.spline corresponds to. To approximate a detrend spline using

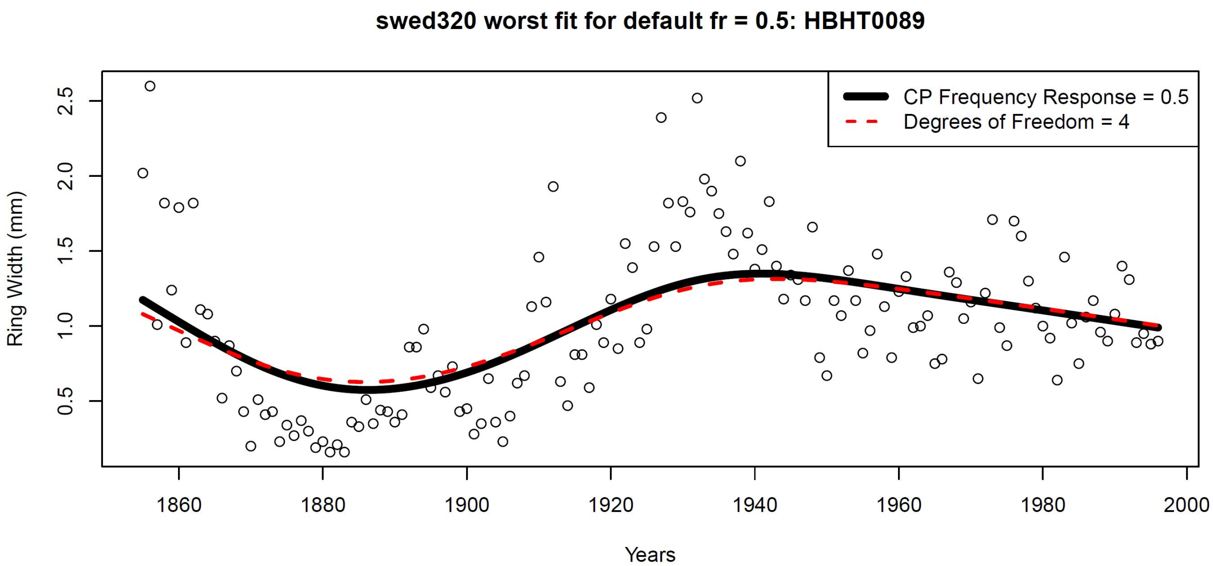

Figure 9: Detrend spline with the default $fr = 0.5$ and period $67\%n$ graphed beneath the spline created by smooth.spline with $df = 4$.

this table, first note what frequency response is desired (what would be normally be entered in detrend's frequency response parameter). Locate the range that captures this fr from the column fr range. To the left of the range, the column df contains its corresponding degrees of freedom. So, for example, to approximate the detrend spline created when $fr = 0.85$, select the fr range of $[0.77, 0.87]$. The degree of freedom in that row is 6. Setting degrees of freedom to 6 in smooth.spline will create a good approximation of the detrend spline.

Table 1: 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}

The fr range column indicates the range of frequency responses which are best approximated by each possibly df. In addition, the closest fr column in Table 1 indicates the specific frequency response that is primarily created by each degrees of freedom choice. The RMSD values illustrate the near equivalency of the two methods' smoothing spline fits. The mean RMSD column is the average performance of the degrees of freedom method for all tree cores tested.

Values in this table were calculated using the Swed320 dataset; however robustness testing was performed using multiple other tree-core datasets [14]. In the robustness tests, there were slight variations in the average RMSD, with the worst case being an RMSD of 1.04×10^{-3} when $df = 2$ in the Pola017 dataset. In all stands tested, the only RMSD that varried this much was when $df = 2$. For all other df values, the RMSD was nearly always a value multiplied by 10^{-7} or 10^{-6}, which are similarly small RMSDs to those in Table 1. In the robustnest test, the frequency response best matched by each degrees of freedom value remained consistent. This suggests that the degrees of freedom method consistently creates nearly equivalent spline fits to the Cook and Peters methods for the given frequency responses.

The highest and lowest RMSDs produced by this method for the Swed320 dataset can be seen in Figure 10 [14]. The spline fits with the lowest and highest RMSD to their corresponding Cook and Peters method spline fit are graphed. The spline fit with the lowest RMSD is a very precise match to its corresponding Cook and Peters method spline fit. The spline fit with the highest RMSD, although it deviates slightly, is still a close match to its corresponding Cook and Peters spline fit. While there is some variability in the degrees of freedom method, even the worst performing spline fits are adequate substitutes for their corresponding Cook and Peters spline fits.

It should be noted that since degrees of freedom can only be set to integers, this parameter does not have the same sliding scale as frequency response, which is a continuous variable. The table indicates the range of frequency responses that would be best approxi-

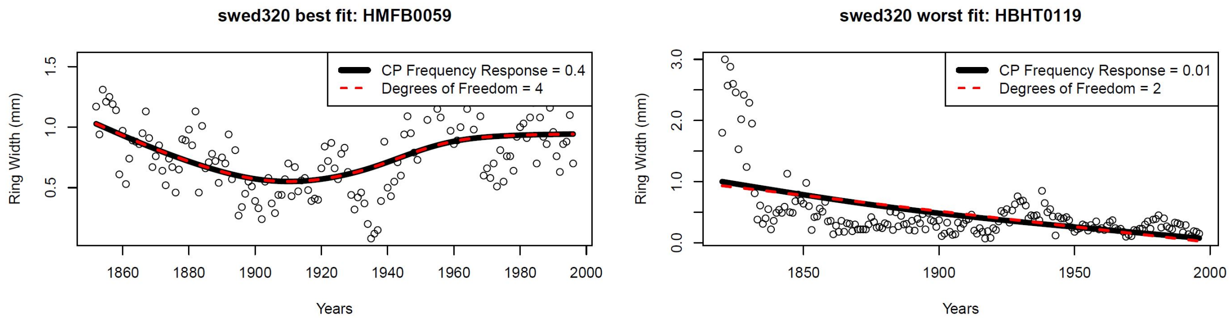

Figure 10: The closest frequency response (fr) match for $df = 4$ (left) and $df = 2$ (right). $df = 4$ has the lowest (best) average RMSD among the closest fr column in 1. Among the Swed320 dataset, tree HMFB0059 produces the best RMSD given given Table 1's suggested fr. $df = 2$ has the highest (worst) RMSD with its closest fitting fr. Among the Swed320 dataset, tree HMFB0119 produces the worst RMSD given Table 1's suggested fr.

mated by a choice of df. Even at the edges of these ranges, the smooth.spline graph is a still good approximation of the corresponding Cook and Peters graph (e.g., Figure 11).

Note that, for the ranges specified in the Table 1, the frequency response values tested were between 0 and 1, with iterations of 0.01. This manner of testing created small gaps between the frequency response ranges. Thus, for example, frequency response values in the range $f = [0.04, 0.25]$ correspond to $df = 3$ and in the next range $fr = [0.26, 0.54]$ correspond to $df = 4$. Frequency response values that fall in between these ranges will be referred to as intermediate fr values. Further research could provide a finer scale if needed, but the precision in Table 1 provides a good fit. Frequency response values that fall in these gaps can be graphed using the df value of either range it falls between. Doing so should still create a decent spline fit match to the corresponding Cook and Peters spline fit.

The two examples provided in Figure 12 indicate that between fr ranges, either choice of df results in a small RMSD between the two the df method spline and the Cook and Peters method spline. These graphs demonstrate that both the upper and lower df values create relatively close fits to the Cook and Peters method for cases where frequency response falls between two of the ranges from Table 1. Figure 12 shows two cases from the Swed320 dataset where the chosen fr is an intermediate value. Further information about the two

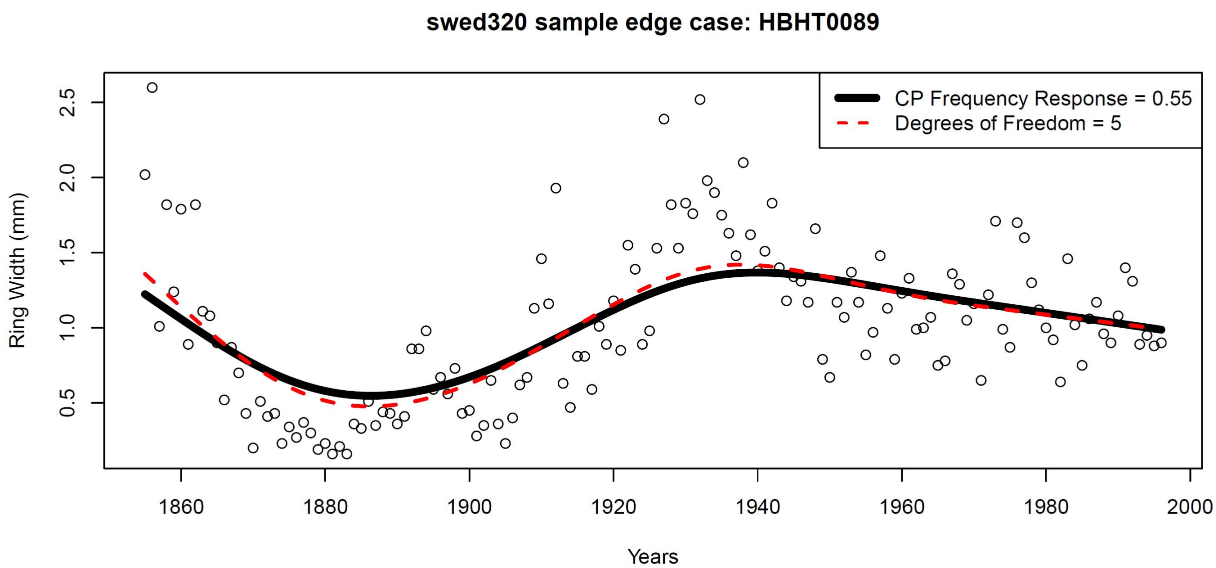

Figure 11: Cook and Peters method spline with frequency response of 0.55 (at the bottom edge of a fr range) graphed beneath the spline created by its suggested smooth.spline *df* choice. Tree-core selected from the Swed320 dataset [14].

gap cases graphed can be found in Table 2. It should be noted that the choice of *df* has some minimal effect as the larger *df* will create a slightly curvier spline, and the smaller *df* will create a slightly smoother spline.

Table 2: Worst case RMSD for intermediate fr values

fr	*df*	Worst RMSD	Core ID
0.225	3	3.28×10^{-3}	$HBHT0119$
0.225	4	2.63×10^{-3}	$HBHT0119$
0.875	6	3.33×10^{-4}	$HBHT0119$
0.875	6	5.13×10^{-4}	$HBHT0119$

For many practical purposes, the approximations provided in Table 1 will be sufficient. However, Section 3.2 contains a more precise technique which calculates the Cook and Peters method smoothing parameter using the degrees of freedom smoothing parameter selection method.

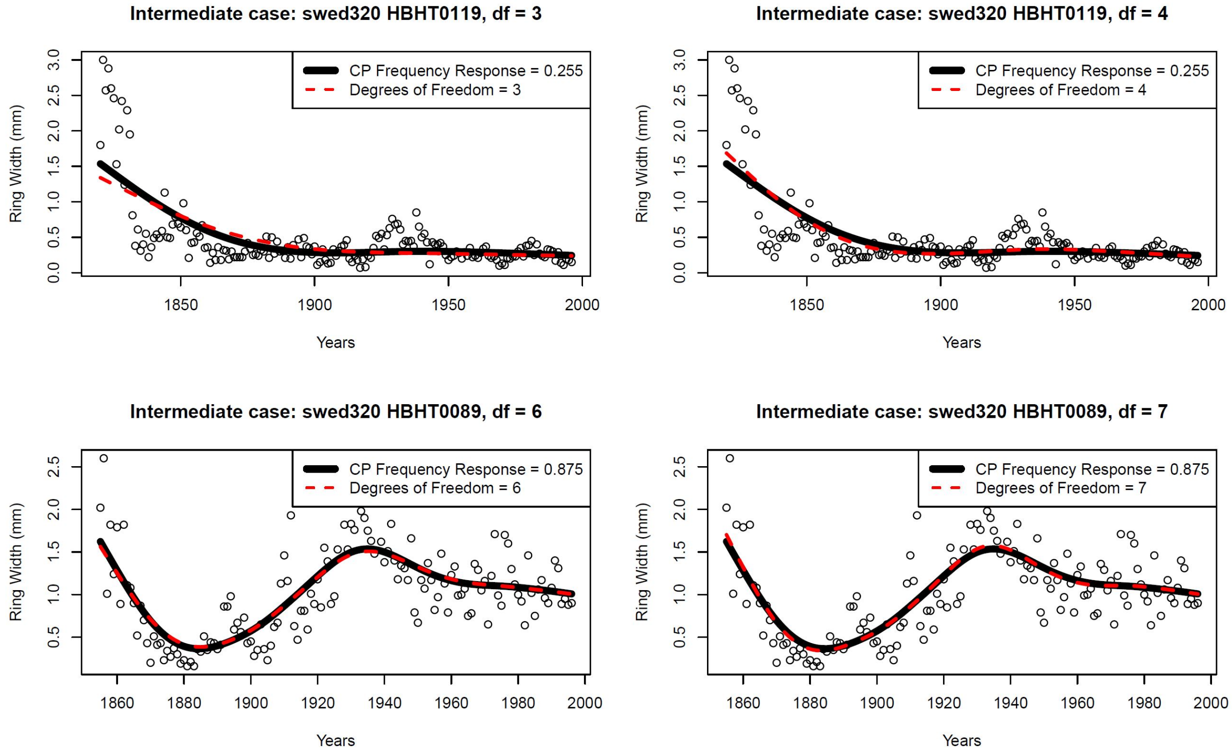

Figure 12: Intermediate frequency responses' resultant CP method splines plotted against their the spline created by using either possible degree of freedom choice with smooth.spline

3.1.2 Precise conversion of smooth.spline's λ to CP's λ with degrees of freedom

There is a direct relationship between the smoothing parameters chosen by the two methods, which can be used to almost exactly calculate the smoothing parameter of the detrend function using the smoothing parameter chosen by smooth.spline.

Equation 1 shows the relationship between λ_d, detrend's chosen lambda value, and λ_{s4}, smooth.spline's chosen lambda value when $df = 4$. It should be noted that the detrend function uses the Cook and Peters method however this research refers to the detrend λ which is $\frac{1}{p}$, where p is the Cook and Peters smoothing parameter, based on the expression from Section 1.3.

$$\lambda_d = (781\lambda_{s4} - 0.439)^4 \tag{1}$$

This equation is based on the linear relationship between (λ_d) and λ_{s4}^4 (Figure 13). This

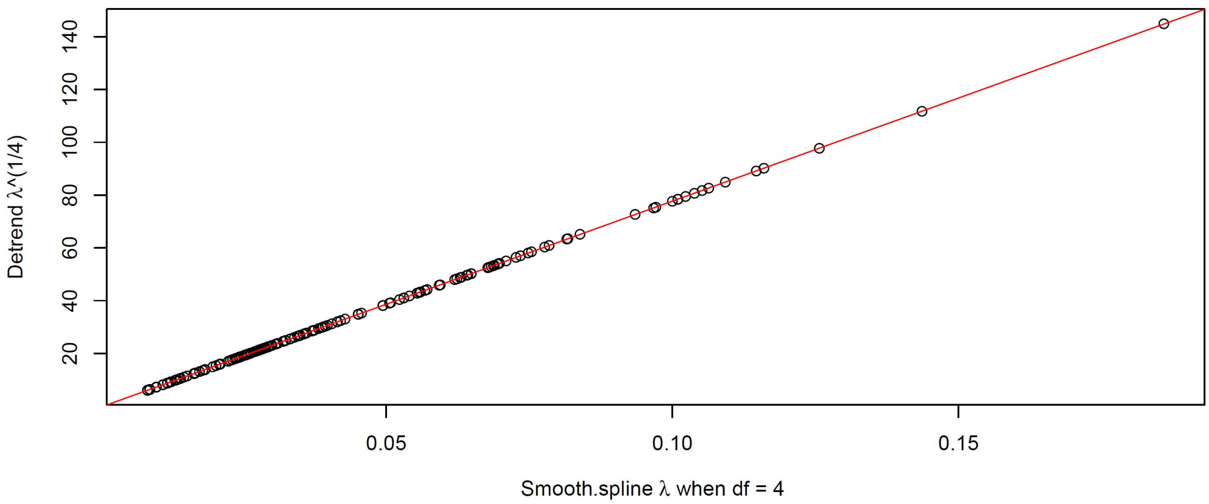

Figure 13: The relationship between λ_d and λ_{s4}. The redline superimposed has the equation $\lambda_d = (781\lambda_{s4} - 0.439)^4$.

relationship was obtained using the Swed320 dataset and seven other stands for robustness [3, 10, 11, 12, 13, 14, 18, 20]. The correlation coefficient was found to be 0.999. Unlike other relationships previously explored, there were no outliers in this relationship. Further, results did not vary with the quantity or selection of data being used. Running this relationship on a single stand versus many stands at once produces the same results, indicating that this relationship should be widely applicable to any dendrochronology data set.

There are similar equations that can be used on some, but not all, *df* value choices. Table 3 shows the equations representing the relationship between the lambda chosen by each degree of freedom and the lambda chosen by detrend, and confirms which relationships have a strong correlation.

Table 3: Conversion Equations from λ_{s4} to λ_d

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

Note that these equations are specific to the case where $fr = 0.5$. However, if a different choice of fr is desired, this can be calculated with a slight alteration of the equation. Let $\lambda_d(fr)$ denote the function to calculate detrend's λ given some frequency response, fr. Equation 2 can be used to obtain any fr choice's lambda using the lambda generated by $df = 4$. This additional term in the equation can be successfully applied to all of the equations in Table 3.

$$\lambda_d(fr) = \left(\frac{1}{1 - fr} - 1\right)(781\lambda_{s4} - 0.439)^4 \tag{2}$$

3.2 Generalized Cross-Validation

Another traditional technique evaluated in this research was Generalized Cross Validation (GCV). Averaging the frequency response value of all tree cores analyzed finds that using GCV to select the smoothing parameter creates a smoothing spline fit with a frequency response of about 0.95. The mean RMSD was 0.032 with a maximum of 0.24. Since, on average using GCV yields a high frequency response and therefore a low frequency reduction, it would not be an adequate substitute for the CP method. When researchers use smoothing splines in dendrochronology, some frequency reduction is desired to reduce high-frequency signals, with 0.5 frequency reduction being the most common choice.

Additionally, the variability in the frequency response of smoothing splines created using GCV, makes the method an inadequate substitution for the CP method. Although many of the smoothing spline fits created using this method had a frequency response of 0.99 or 1.00 (Table 4), there would occasionally be fits with much lower frequency responses (Table 5).

Table 4: A sample of the frequency response of GCV smoothing splines for dataset Swed320 from Stockholm, Sweden [14].

Core ID	RMSD	Equivalent fr
HBHT0039	0.03021	0.99
HBHT0049	0.01273	1.00
HBHT0059	0.01381	1.00
HBHT0069	0.01871	0.99
HBHT0079	0.02007	1.00
HBHT0089	0.00428	0.99
HBHT0099	0.02734	1.00
HBHT0119	0.01788	0.99
HBHT0129	0.02074	0.99
HMFB0029	0.00714	1.00

Table 5: A sample of the frequency response of GCV smoothing splines for dataset ZIMB001 from Zimbabwe [18].

Core ID	RMSD	Equivalent fr
BAO07B	0.00390	0.99
BAO08A	0.00594	0.99
BAO08B	0.02478	0.99
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

4 Discussion

4.1 Significance of Results

When Cook and Peters developed their smoothing parameter selection method in 1981, smoothing splines were a fairly nascent technique. Shortly after Cook and Peters developed their smoothing parameter selection method specifically designed for tree-ring analysis, many other techniques were developed for selecting the smoothing parameter. These techniques are what this research refers to as traditional smoothing spline smoothing parameter selection methods. These selection methods have been thoroughly studied over the past 4 decades, with a wide range of applications. Due to the versatility and widespread use of these traditional smoothing parameter selection methods, they are more widely understood than the Cook and Peters smoothing parameter selection method.

In addition to being more widely understood, the use of traditional smoothing splines is more accessible. The smooth.spline function, which creates traditional smoothing splines, is available in base R under the stats package. In contrast, to access detrend the function that creates Cook and Peter method splines, the dplR package must be installed. Moreover, smooth.spline is a more flexible function and its output offers far more information about the nature of the spline fit created than detrend, as smooth.spline returns data such as the smoothing parameters and a summary of the spline function. Using traditional smoothing splines in conjunction with the function smooth.spline may give dendrochronologists more control over spline fits and will offer them more quantitative information on the nature of their data. For these reasons, this research aimed to find a traditional smoothing parameter selection method available in smooth.spline that can create equivalent splines to the Cook and Peters method, ultimately allowing for the utilization of smooth.spline in dendrochronology.

This research offers two alternative methods, based on degrees of freedom, to the Cook and Peters method to fit smoothing splines to tree-ring data. Degrees of freedom allows researchers to create the same splines as the detrend function, with increased versatility and accessibility. Being able to convert splines between the two methods allows a wider audience to interpret tree-ring data. This research bridges the gap between dendrochronology and other field that use smoothing splines, making these traditional spline methods even more universal.

4.2 Potential Next Steps

Due to the strength of the correlation between λ_{s4} chosen by $df = 4$ and λ chosen by detrend, it is likely that this relationship can be derived mathematically, in addition the the results provided in this research.. The correlation was extremely close to, yet not exactly, 1. This discrepancy is potentially due to rounding between steps in R which would have added error to the calculated smoothing parameters. Thus it would be important to show this relationship mathematically. Proving this relationship would further establish the strength of this research's findings. If this relationship cannot be proven, further would be needed to understand the true connection between the Cook and Peters method and the degrees of freedom method and where the two methods possibly differ.

Additional research could be performed to enhance the precision of the degrees of freedom method. This research tested frequency response values from 0 to 1 in iterations of .01. It would have been too computationally expensive for this research to test frequency response values in smaller iterations. Doing so however would help to shrink or eliminate the gap frequency response values outlined in the results section.

This research focused on using degrees of freedom to mimic the Cook and Peters method for different frequency values, but no analysis was performed for different $\%n$ criteria. This research used $67\%n$ for all Cook and Peters smoothing parameter calculations as it is the most common choice in dendrochronology. Although $67\%n$ is the most common choice, dendrochronologists can select between $30\%n$ and $75\%n$. Examining the relationship between the Cook and Peters method and the degrees of freedom method for varying $\%n$ could make the degrees of freedom method more robustly applicable as there would be more variability in parameter options.

Lastly, further work could be done to connect the degrees of freedom method with the complete version of the Cook and Peters method as derived in Bussberg et. al. (2020) [5]. Bussberg et. al details how the derivation of the closed form solution in the Cook and Peters method is not complete. The complete solution provided by Bussberg et. al. yields slightly different smoothing parameter values than the original Cook and Peters method. This research chose to use the original Cook and Peters method because its closed form solution was much easier to implement in testing. There is a relationship between the smoothing parameters selected by the original and complete Cook and Peters methods which could be used to connect the degrees of freedom method with the complete Cook and Peters method.

5 Acknowledgments

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 the REU program heads Dr. Karen Yokley and Dr. Nicholas Luke for your guidance and leadership.

References

[1] *International Tree-Ring Data Bank.* URL: https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring.

[2] *Smoothing Splines and BIC*, (2018).

[3] H. P. Borgaonkar, *NOAA/WDS Paleoclimatology—Borgaonkar—Narangathara—TEGR—ITRDB INDI025.* URL: https://doi.org/10.25921/4j49-4y22, May 2013. Accessed 2021-07.

[4] A. Bunn, M. Korpela, F. Biondi, F. Campelo, P. Mérian, F. Qeadan, and C. Zang, *dplR: Dendrochronology Program Library in R*, 2021. R package version 1.7.2.

[5] N. Bussberg, *Environmental applications of temporal, spatial, and spatio-temporal statistics*, PhD Thesis, Indiana University (2020).

[6] A. Camerino, *Tree-ring diagram.* URL: snowbrains.com/dendrochronology-history-from-a-trees-perspective/, Jun 2020.

[7] N. Chen, *Denoising data with fast fourier transform.* URL: https://kinder-chen.medium.com/denoising-data-with-fast-fourier-transform-a81d9f38cc4c, Mar 2021.

[8] E. R. Cook and L. Kairiukstis, eds., *Methods of dendrochronology: applications in the environmental sciences*, Kluwer Academic Publishers, Dordrecht, the Netherlands Norwell, MA, 1989.

[9] E. R. Cook and K. Peters, *The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies*, Tree-Ring Bulletin, (1981), p. 41–45.

[10] D. Granato-Souza, G. Pereira, D. Stahle, and A. Barbosa, *NOAA/WDS Paleoclimatology—Granato-Souza—Rio Paru Site A—CESP–ITRDB BRA001.* URL: https://doi.org/10.25921/86cd-n444, May 2018. Accessed 2021-07.

[11] D. A. Graybill, *NOAA/WDS Paleoclimatology—Graybill—Onion Valley—PIBA—ITRDB CA531*. url: https://doi.org/10.25921/rf8w-2t60, Apr 2002. Accessed 2021-07.

[12] R. L. Holmes and A. J. E, *NOAA/WDS Paleoclimatology—Holmes—El Maiten—AUCH—ITRDB ARGE010*. url: https://doi.org/10.25921/qxc2-g248, May 1996. Accessed 2021-07.

[13] V. C. Lamarche, *NOAA/WDS Paleoclimatology—Lamarche—Coolangatta Road—PHAS—ITRDB AUSL012*. url: Put URL here, Apr 2002. Accessed 2021-07.

[14] H. W. Linderholm, *NOAA/WDS Paleoclimatology—Linderholm—Stockholm—PISY—ITRDB SWED320*. url: https://doi.org/10.25921/n1n3-fh42. Accessed 2021-07.

[15] R. Noe, *When splines were physical objects*. url: https://www.core77.com/posts/55368/When-Splines-Were-Physical-Objects, Aug 2016.

[16] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021.

[17] C. H. Reinsch, *Smoothing by spline functions*, Numerische mathematik, 10 (1967), pp. 177–183.

[18] D. W. Stahle, D. Gabor, G. Haynes, J. Klimowicz, and P. Ngwenya, *NOAA/WDS Paleoclimatology—Stahle—Baobab Holes—CABU—ITRDB ZIMB001*. url: https://doi.org/10.25921/e3wk-1988, Aug 2005. Accessed 2021-07.

[19] G. Wahba, *Spline models for observational data*, Society for Industrial and Applied Mathematics, 1990.

[20] T. Wazny, *NOAA/WDS Paleoclimatology—Borgaonkar—Narangathara—TEGR—ITRDB INDI025*. url: https://doi.org/10.25921/e5t9-xa37, Apr 2002. Accessed 2021-07.