Dendrochronologia 71 (2022) 125904

Contents lists available at ScienceDirect

Dendrochronologia

journal homepage: www.elsevier.com/locate/dendro

Identifying southern yellow pine cross sections from the southeastern United States using quadratic discriminant analysis on pith and second annual ring diameters

Thomas L. Eberhardt [a,*], Patricia K. Lebow [a], Philip M. Sheridan [b], Arvind A.R. Bhuta [c]

[a] USDA Forest Service, USDA Forest Service, Forest Products Laboratory, 1 Gifford Pinchot Drive, Madison, WI 53726, USA
[b] Meadowview Biological Research Station, 8390 Fredericksburg Tnpk., Woodford, VA 22580, USA
[c] USDA Forest Service, 201 14th St SW, Washington D.C. 20250, USA

ARTICLE INFO

Keywords:
Loblolly pine
longleaf pine
Shortleaf pine
Statistical analysis
Tree ring
Wood anatomy

ABSTRACT

Southern yellow pine specimens collected from historical structures, stumps, and coarse woody debris in forests have been difficult to identify at the species level due to similar wood anatomy. This can be problematic for dendrochronologists when identifying the correct species used in the construction of historical structures, or reconstructing forest history on the landscape and using those specimens in the context of that history. We applied a quadratic discriminant analysis (QDA) to update a century-old method plotting pith diameters against second annual ring diameters to discern one species of southern yellow pine from others. Our analysis estimates error rates for false positive and false negative determinations when comparing longleaf pine (*Pinus palustris* Mill.) to shortleaf (*Pinus echinata* Mill.) and loblolly pine (*Pinus taeda* L.). The cross-validated false positive error rates for the smallest dataset ($n = 46$), was nearly twice (9.52%) that determined as a simple proportion by counting errant observations (4.76%). QDA of the largest dataset ($n = 206$) gave a flatter zero contour and false positive rate (3.13%) like the proportionally determined value (1.56%), despite one additional observation being falsely assigned to longleaf pine by QDA. An unknown, unearthed southern pine specimen from southeastern Virginia was radiocarbon dated up to 500 years prior and assigned as longleaf by our method (probability $\geq$ 0.9998). Thus, through a QDA, it is possible to greatly improve confidence in identifications of key unknown specimens that can provide evidence of discerning one species, longleaf pine, from other southern yellow pines.

1. Introduction

Dendrochronology allows scientists to use precise measurements and analyses of tree rings to untangle complex climatological, ecological, and/or archaeological phenomena associated with a tree's radial growth (Douglass, 1941; Speer, 2010). Overlapping present-day specimens from live trees with remnant specimens (e.g. preserved cross sections) builds a deeper knowledge of how these phenomena impacted radial growth further into the past. Using samples from the same genera in developing these chronologies is essential, if inferring how these events may have impacted a forest, or how and when a tree was cut for use in structures of archaeological interest. Furthermore, it may be necessary to identify the specimen down to the species level as different species might yield different climatic or ecological responses, or might help to discern why humans may have selected a certain species over another for use in historical and current structures. Such is the case when working with southern yellow pines from the southeastern United States.

Southern yellow pines have been used widely around the world for commercial forest products since European colonization of the Americas (Williams, 1989; Frost, 2006; Fox et al., 2007); species favored for past and present use include longleaf pine (*Pinus palustris* Mill.), loblolly pine (*Pinus taeda* L.), shortleaf pine (*Pinus echinata* Mill.) and slash pine (*Pinus elliottii* Engelm.). Longleaf pine was the predominant species used from European colonization up until the middle of the 20th century in this region, covering over 37 million hectares in pure or mixed forests prior to colonial settlement (Frost, 1993). This species' strong and dense wood from straight boles was suitable for ship masts as well as lumber and flooring used in commercial and residential structures (Kush et al., 2004). Longleaf pine was also a critical chemical resource for the naval stores industry. This species, along with slash pine, were intentionally

---

* Corresponding author.
E-mail address: thomas.l.eberhardt@usda.gov (T.L. Eberhardt).

https://doi.org/10.1016/j.dendro.2021.125904
Received 21 June 2021; Received in revised form 1 November 2021; Accepted 3 November 2021
Available online 9 November 2021
1125-7865/Published by Elsevier GmbH.

wounded for oleoresin collection and processed to provide a variety of resources (Gamble, 1921; Wahlenberg, 1960; Butler, 1998; Outland, 2004; Frost, 2006). Once the trees were harvested, their stumps were later pulled from the ground for steam/solvent extractions to separate the residual resin compounds from the stump wood (Gardner, 1989). The timber and naval stores industry resulted in the collapse of the Virginia longleaf pine ecosystem in 1843 (Rouse, 1988; Frost, 2006). Longleaf pine still grows on sites across its native range, from southeastern Virginia to eastern Texas; however, longleaf pine currently covers only a small fraction of the former land area (Frost, 1993; Outcalt and Sheffield, 1996; Sheridan et al., 1999; Oswalt and Guldin, 2021). Other southern yellow pines, such as loblolly pine, shortleaf pine, and slash pine have supplanted longleaf for commercial use throughout much of the 20th century and into the 21st century.

Differentiating southern yellow pines to the species level using chemotaxonomic or spectroscopic approaches has shown limited utility (Eberhardt et al., 2007, 2009a) while differentiating them by gross and microscopic anatomy has been ineffectual (Panshin and de Zeeuw, 1980). Determining southern yellow pines at the species level is especially important when dendrochronologists are trying to better understand the ecological, climatological, or archaeological history of the temperate forests of southeastern North America. Different southern yellow pine species may elicit varying responses to climate (Friend and Hafley, 1989; Zahner, 1989; Devall et al., 1991; Travis and Meentemeyer, 1991; Cook et al., 1998; Meldahl et al., 1999; Foster and Brooks, 2001; Bhuta et al., 2009; Henderson and Grissino-Mayer, 2009) and ecology (Bhuta et al., 2008; Pederson et al., 2008). Determining why one southern yellow pine species was chosen over another in construction might provide insight into the history of an archaeological site (van de Gevel et al., 2009). Longleaf pine was presumed to be present in many dendrochronological studies investigating climate, ecology, and archaeology (van de Gevel et al., 2009; Therrell et al., 2017; Harley et al., 2018; Patterson and Knapp, 2018); however, these investigators used qualitative assessments to identify their longleaf pine timbers or stumps. The only known method to identify longleaf pine timbers was developed by Koehler (1932), validated by Eberhardt et al. (2011), was not used in any of these studies.

The method for determining longleaf pine from other southern yellow pines (loblolly pine, shortleaf pine, and slash pine) involves measuring the pith and second annual ring diameters (if intact), plotting these diameters against each other, and fitting the measurements to a curve where longleaf pine lies above the curve and loblolly, shortleaf, and slash pines are below (Koehler, 1932). The method was tested and validated with longleaf, shortleaf, and loblolly pine samples (Eberhardt et al., 2009b, 2011). While this method uses repeatable and observable data, the curve or line lacked a statistical approach as it was simply drawn onto the plot to differentiate longleaf pine from other southern yellow pines. No data points or equation were provided with the original method (Koehler, 1915a, 1915b, 1932).

The first objective of our research was to determine if we could delineate a more statistically robust curve compared to the prior Koehler method (1932) in the determination of false positive (and false negative) error rates for differentiating a southern yellow pine, specifically longleaf pine, from shortleaf and loblolly pines. We used three different datasets to develop these curves: 1) cross sections at stump height (0.15 m, 0.5 m); 2) cross sections at and above stump height (5 m, 6.1 m); and 3) increment cores at stump height (0.2 m). The second objective of our research was to determine if these curves could be used to identify an unearthed radiocarbon-dated lightwood specimen from a mire in southeastern Virginia that had both intact pith and second annual ring. The quantitative measurements coupled with the statistical robustness in delineating these curves could potentially help scientists in the southeastern United States determine if unknown specimens they are working with are longleaf pine or one of the other southern yellow pines.

2. Materials and methods

2.1. Study tree sampling, preparation, and measurement

We obtained samples from cross sections and cores from throughout the southeastern United States. Cross sections collected for wood quality studies (e.g. Kelley et al., 2004), were sampled at a stump height of 0.5 m (n = 26), and in some cases at a height of 5 m (n = 11) above ground level, for loblolly (age range 13–39 years), shortleaf (22 years in age), and longleaf (55 years in age) pines. Seventy-year-old longleaf pine cross sections collected from a spacing, pruning, and thinning study site in Louisiana (Eberhardt et al., 2018, 2019) were measured at a stump height of 0.15 m (n = 20) and at a height of 6.1 m (n = 20) above ground level. Cross sections were air-dried and areas near the pith and second annual ring diameters were smoothed with a razor blade and/or fine sandpaper to improve measurement precision (0.01 mm resolution) using digital calipers (Eberhardt et al., 2011). When needed, a light application of water to the surface improved the contrast for the features of interest, particularly the first annual ring (Koehler, 1915a, 1915b). A hand lens (10 ×) was used when necessary to aid in our measurements. When annual rings were somewhat elliptical, the maximum and minimum diameters were averaged.

Increment cores (n = 206) collected at a stump height of 0.2 m above ground level from longleaf, loblolly, and shortleaf pines from unpublished, published, and ongoing research from across southeastern United States (Bhuta et al., 2008, 2009; Bhuta, 2011; Bhuta and Kennedy, 2021) were also used. The average age for these increment cores were 53.6 years for longleaf pine, 61.4 years for loblolly pine, and 50.8 years for shortleaf pine. Increment cores for these projects had already been prepared using progressively finer abrasive paper (200–15 μm) to reveal details of each core and were measured for ring widths. For our current study, these increment cores were examined to determine if their piths and second annual rings were intact. Cores with their pith and their first annual and second annual ring width diameters intact on both sides of the core were remeasured using a Velmex TA measuring system to determine both pith and second annual ring diameters (0.001 mm resolution).

A stem section of southern yellow pine was unearthed from a mire at Joseph Pines Preserve (Sussex County, VA) in 2018 (Fig. 1). This specimen was dried under ambient conditions and cut to expose the cross section for measuring values for the pith and second annual ring diameters. A small subsection (2 cm × 2 cm × 3 cm) was cleaved from rings 21–24 on the edge of the cross-section and sent to Beta-Analytic (Miami, FL) for Accelerator Mass Spectrometry (AMS) dating. Resin-soaked wood near the pith and irregular projections radiating from the pith were not included in our measurements as part of the pith for our cross sections, increment cores, and the unearthed pine sample (Koehler, 1932; Eberhardt et al., 2011). Transverse surfaces of longleaf (MADw 8859), shortleaf (MADw 15738), and loblolly (MADw 15729) pine wood specimens from the USDA Forest Products Laboratory xylarium (Madison, WI, USA) were polished by sanding with a series of sandpapers up to a grit of 1500, then imaged using the XyloTron (Ravindran et al., 2020), with each image covering a 6.35 mm × 6.35 mm area.

2.2. Quadratic discriminant analysis

Data points, nor an equation for delineating the initial curve, were ever provided by Koehler (1932). To recreate this curve, the graph from Koehler (1932) was enlarged on a photocopier to facilitate the selection of points along the curve that crossed the grid lines. We plotted these data points (18 total) and fit the best polynomial equation to these points in Microsoft Excel. This curve was plotted against our quadratic discriminant analyses to determine differences between our contours and Koehler's (1932) curve.

With observations that have known groups, supervised learning can

Fig. 1. Southern yellow pine tree section unearthed at The Joseph Pines Preserve in Sussex County Virginia: a) initial cut, b) cross section with rectangle indicating pith and second annual ring zone of interest, and c) closer view of pith with arrows indicating second annual ring diameter measurement.

be used to develop models to understand relationships between different group classifications and to help classify future unknown observations (James et al., 2013). Classical statistical methods, such as linear (or quadratic) discriminant analysis and logistic regression, provide simple functionals that can be applied to new observations to calculate probabilities of class membership. Other methods are available if warranted for more complex situations and data sets, such as when the distributional characteristics are unknown, or with relatively large (e.g., tall and/or wide) data sets that can present computational issues for traditional statistical methods (see "high-dimensional" data in James et al., 2013).

Based on initial assessments of previously collected data (Eberhardt et al., 2011) and the appearance of Koehler's curve on plots of those data, the statistical methods that are most comparable to Koehler's methodology were quadratic discriminant analysis (QDA) or logistic regression. A density plot for the first cross section dataset (n = 46) was generated to visualize the datapoint distributions and verify the applicability of QDA, and not linear discriminant analysis. QDA is applicable when groups are assumed to be normally distributed but have heterogenous covariance matrices and the resulting boundaries delineating the groups are quadratic functions. Alternatively, logistic regression provides a flexible approach, allowing the modeling of group probability membership as a function of polynomial equations, but is not discussed here since the results were similar to those following QDA.

Statistical analyses were conducted in SAS/STAT® V14.1 software (PROC DISCRIM, SAS Version 9.4 software, SAS Institute Inc., Cary, NC,

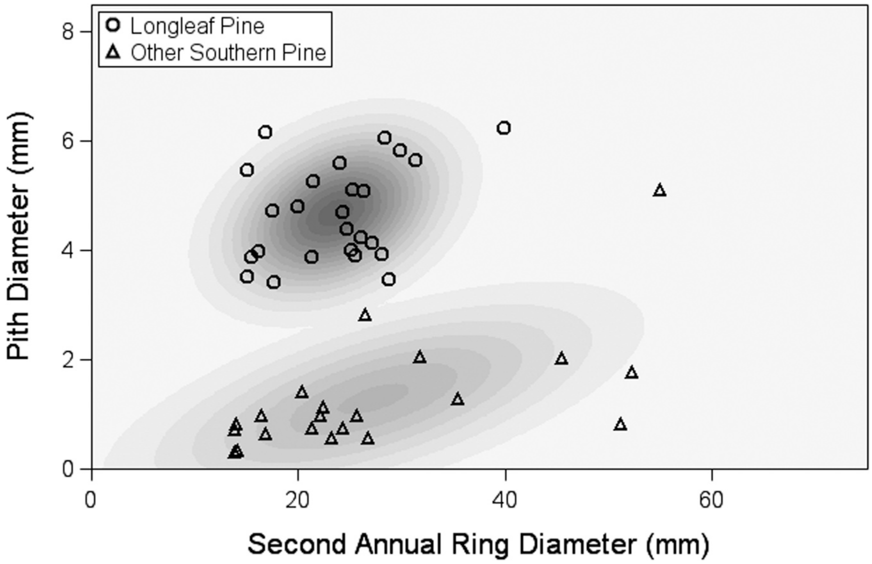

Fig. 2. Bivariate normal density plot for longleaf and other southern pines (loblolly pine and shortleaf pine) observations (pith and second annual ring diameters) at stump height (n = 46).

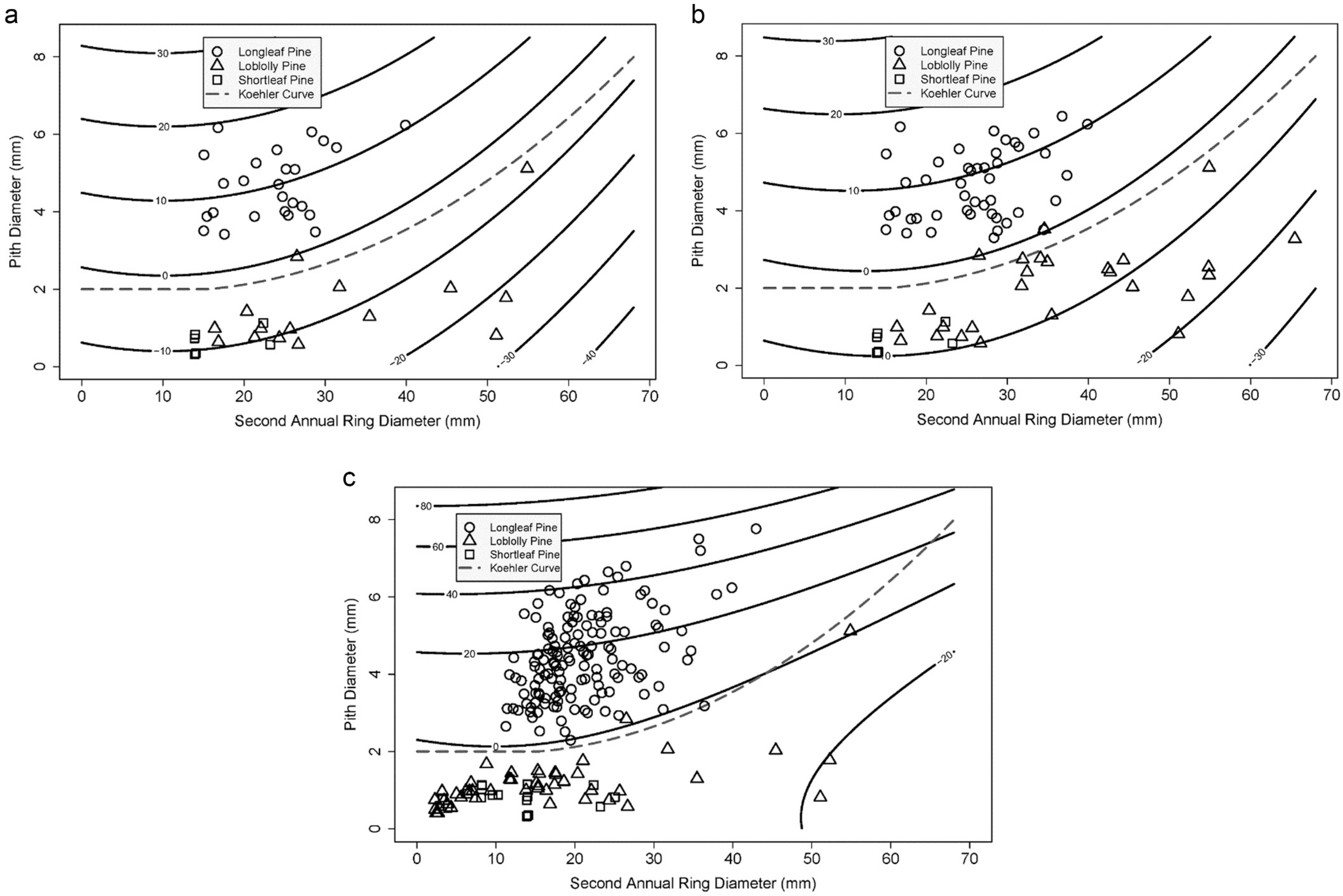

Fig. 3. Plot of pith against second annual ring diameters with quadratic discriminant analysis function contours and second order polynomial curve (Koehler, 1932) for longleaf and other southern yellow pines (loblolly pine and shortleaf pine) for: a) cross sections at stump height (n = 46) b) cross sections at and above stump height (n = 77) c) increment cores at stump height (n = 206). The discriminant function zero contour delineates longleaf pine from other southern yellow pines.

2016). Equal prior probabilities for class membership were assumed since the groups were not sampled based on proportional occurrence. A bivariate normal density plot generated for longleaf pine and the other southern yellow pine species (loblolly and shortleaf pines together) datapoints, all measured at stump height (0.15 m, 0.5 m; n = 46), exhibits distributions differing in shape and orientation (Fig. 2). Homogeneity of the within-group covariance matrices was tested with Bartlett's modification of the likelihood ratio test (POOL=TEST option in SAS's PROC DISCRIM, SAS Institute, Inc., 2016). Beyond SAS software documentation, Davis (1981) details QDA calculations, while Li (2006) also provides an overview of the methods in the context of a financial application. These discussions give the mathematics of the quadratic discriminant function in terms of measuring an observation's distance to a group using generalized squared distance and probability of group membership as a function of the squared distance. To evaluate the performance of the methods, "leave-one-out" cross-validation was used to estimate error rates as given in a confusion matrix (James et al., 2013). This type of cross-validation helps to reduce the bias in the estimation of error rates and is applied by leaving one observation out of the dataset, calculating the discriminant function without the observation, and then determining the classification for that observation using that newly calculated discriminant function. Contour figures of the generalized squared distances (based on the calculated quadratic discriminant functions) were generated in the statistical software package R, version 4.0.0 (R Core Team, 2020).

3. Results

3.1. Koehler and quadratic discriminant curves and analysis

We determined that a second order polynomial equation: $y = 0.00183x^2 - 0.0386x + 2.159$; where y = pith diameter (mm) and x = second annual ring diameter (mm) best fit Kohler's (1932) approach and assisted us when visually comparing this curve to our QDA. The minimum value for the second annual ring diameter on Koehler's graph is 6.35 mm. We assumed that any extension of the curve to lower second annual rings diameters would be with a constant pith diameter of 2 mm (Eberhardt et al., 2011).

The bivariate densities for longleaf pine and the other southern yellow pine species (loblolly and shortleaf pines together) datapoints, all measured at a stump height (0.15 m, 0.5 m; n = 46), show them to have different distributions based on shape and orientation (Fig. 2). The homogeneity of the within group covariance matrices was tested and rejected, with a highly significant probability value (0.0077). This indicated a difference in the covariances and variances within each of the groups, violating an assumption of linear discriminant analysis. QDA was henceforth applied to this dataset, and the larger datasets from cross sections (n = 77) and increment cores (n = 206), to give contours which bared resemblance to Koehler's curve (Fig. 3a–c). A confusion matrix based on our QDA provides corresponding estimates of the error rates (Table 1), which were determined by leave-one-out cross-validation and are based on an equal probability of being in either of the two groups (longleaf pine vs. the other southern pines).

Table 1
Confusion matrix with estimates of error rates as determined by cross validation for longleaf, loblolly and shortleaf pine data collected from cross sections at stump height, cross sections at and above stump height, and increment cores at stump height.

		Predicted Species								
		Longleaf	Loblolly or Shortleaf	Longleaf, Loblolly, or Shortleaf	Longleaf	Loblolly or Shortleaf	Longleaf, Loblolly, or Shortleaf	Longleaf	Loblolly or Shortleaf	Longleaf, Loblolly, or Shortleaf
		Cross sections at stump height (n = 46)			Cross sections at and above stump height (n = 77)			Increment cores at stump height (n = 206)		
Actual Species	**Longleaf**	25 (100%)	0 (**0%**)	25 (100%)	44 (97.8%)	1 (**2.2%**)	45 (100%)	140 (98.6%)	2 (**1.4%**)	142 (100%)
	Loblolly or Shortleaf	2 (**9.5%**)	19 (90.5%)	21 (100%)	2 (**6.3%**)	30 (93.8%)	32 (100%)	2 (**3.1%**)	62 (96.9%)	64 (100%)
	Longleaf, Shortleaf, or Loblolly,	27 (58.7%)	19 (41.3%)	46 (100%)	46 (59.7%)	31 (40.3%)	77 (100%)	142 (68.9%)	64 (31.1%)	206 (100%)
		Estimates for false positive (9.52%) and false negative (0.00%) error rates for classification as longleaf pine are shown in bold font. Estimated overall total error rate was determined to be 4.76%, based on equal probability of being in either of the two groups (longleaf vs. loblolly and shortleaf).			Estimates for false positive (6.25%) and false negative (2.22%) error rates for classification as longleaf are shown in bold font. Estimated overall total error rate was determined to be 4.24%, based on equal probability of being either of the two groups (longleaf vs. loblolly and shortleaf).			Estimates of false positive (3.13%) and false negative (1.41%) error rates for classification as longleaf are shown in bold font. Estimated overall total error rate was determined to be 2.27%, based on equal probability of being either of the two groups (longleaf vs. loblolly and shortleaf).		

3.2. *Identification and radiocarbon dating of unknown southern yellow pine mire sample*

We were able to determine the posterior probability of the unknown radiocarbon sample being identified as longleaf pine at 0.9998 (Table 2) based on the quadratic discriminant function calculated from our disk samples with observations taken at either a stump height or further up the tree (n = 77). This calculation is arrived at approximately by first calculating the generalized squared distance as shown in Eq. (1) (corresponding to the contours in Fig. 3b):

$$q(p,s) = -12.96 + 4.60p + 0.05p^2 + 0.26s - 0.01s^2 - 0.01ps \quad (1)$$

where p = pith measurement (mm) and s = second annual ring measurement (mm).

The posterior probability of being a longleaf pine is then calculated as shown in Eq. (2):

$$P(p,s) = \quad exp(q(p,s))/(exp(q(p,s)) + 1) \quad (2)$$

The posterior probabilities using the smaller (n = 46) and larger stump height (n = 206) datasets were 0.9999 and 1.0000, respectively (Table 2).

Radiocarbon dating provided five possible age ranges for the mire lightwood sample (Fig. 4, Table 3). One age (1940) could be excluded from consideration since no original longleaf pine existed at Joseph Pines in 2008, the calculated date of the tree for the last measurable rings assuming 1940 as the date for rings 21–24. In fact, no original longleaf pine existed in Sussex County, VA by that date. While none of the other four radiocarbon dates can be excluded, the most likely age of the mire sample is between 1624 and 1680 CE (53%). Regardless, the mire sample definitely dates to before 1800 CE. Anchoring this sample to the dendrochronological record could further validate the radiocarbon date, however there are no southern yellow pine chronologies in this area to which this sample could be anchored to.

4. Discussion

For clear wood specimens, there are no gross and microscopic anatomical features that allow even highly trained wood anatomists to definitively differentiate between the southern yellow pines on a species level (Panshin and de Zeeuw, 1980). This point is illustrated by the obvious similarity of cross sectional images (Fig. 5) from vouchered specimens of longleaf, shortleaf, and loblolly pines, all from the wood collection at the Forest Products Laboratory. Characteristic of the southern yellow pines as a group is the presence of wide latewood bands not found in the other hard pines (Panshin and de Zeeuw, 1980); Kukachka (1960) did note multiple latewood bands in individual annual rings are sometimes present for longleaf and slash pines. Tree cross sections, when available, do allow longleaf pine to be separated from the other southern pines by measurements of the pith and second annual ring diameters, the latter being narrow in longleaf pine and relatively wide for shortleaf and loblolly pines. Note that annual rings further out were not considered by Koehler (1932) because they would be less indicative of tree vigor around the time of pith formation.

The greatest accuracy for this empirical method was obtained at

Table 2
Posterior probabilities of unearthed southern yellow pine section being longleaf pine or the other southern pine (loblolly and shortleaf pines) group based on quadratic discriminant functions as determined by the different data sets.

Sampling height (m)	Total observations	Probability	
		Longleaf pine	Loblolly or shortleaf pine
0.15, 0.5	46	0.9998	0.0002
0.15, 0.5, 5, 6.1	77	0.9999	0.0001
0.2	206	1	0

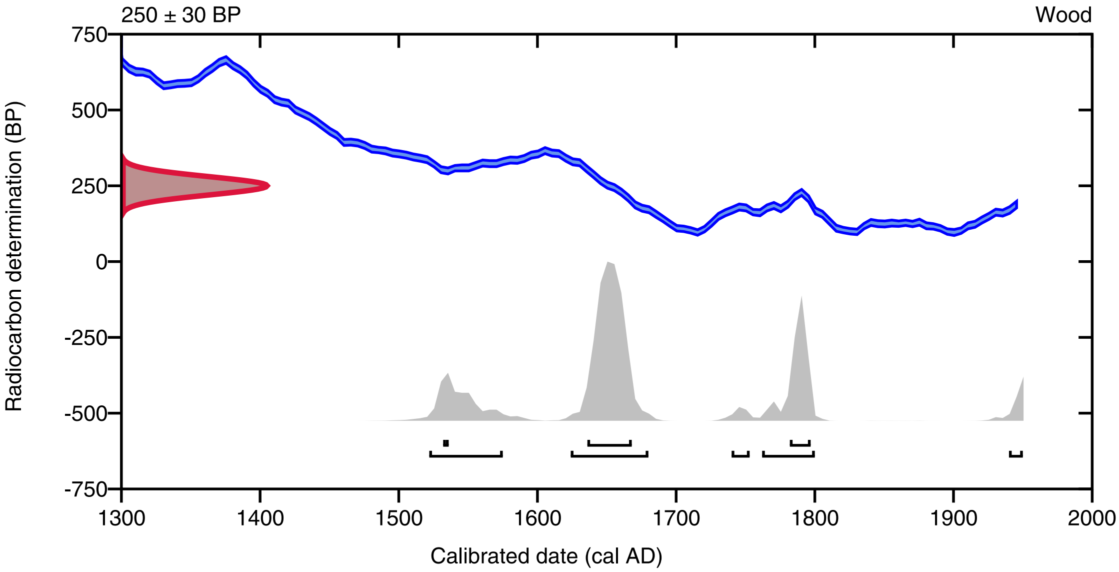

Fig. 4. Calibration of radiocarbon age to calendar years for unearthed southern yellow pine tree section.

Table 3
Calendar age ranges of unearthed southern yellow pine tree section as determined by radiocarbon dating.

Probability (%)[a]	Calibrated date (cal AD)	Radiocarbon determination (cal BP)
15.4	1522–1575	428–375
53	1624–1680	326–270
1.8	1740–1753	210–197
22	1762–1800	188–150
3.3	1940 to post 1950	10 to post 0

[a] 95.4% total probability determination values provided by analytical lab; all date ranges are possible unless eliminated by other chronologies.

stump height (Koehler, 1932) where the pith is the smallest for a given tree. Pith increases in size up the bole of the tree and decreases again at the crown (Koehler, 1915a, 1915b). The method is best applied to slow-growing timbers from natural forests, and not pine plantations, the latter having rapid growth for slash, loblolly and shortleaf pines that can give pith diameters approaching those typical of longleaf pine (Koehler, 1932). Initial research into this method (Koehler, 1915a, 1915b, 1917) describes a line with a positive slope, that is presented as a very shallow curve. An exception to this is a graph of a completely straight line with no data points shown (Koehler, 1921). Using a minimum differentiating pith diameter of 2.54 mm, the initial method was without exception (Koehler, 1915a, 1915b). Upon expanding the data set and applying an updated curve with a minimum differentiating pith diameter of about 2 mm, false positives for longleaf pine were reported to be 2.7%, 2%, and 3.7% for shortleaf, loblolly, and slash pine timbers, respectively (Koehler, 1932). Through a statistical approach, herein being discriminant analysis, the proximity of an unknown observation on the pith diameter versus second annual ring diameter plot, relative to clusters of known observations, can be factored into determining the probability of a false positive assignment to longleaf pine.

4.1. QDA of cross section data at stump height

Corresponding pith and second annual ring diameter values provided the observations for the discriminant analyses. A statistical approach was rationalized as having the potential to improve the separation of longleaf pine from the other southern pines, with loblolly and shortleaf pines grouped together. Another desired outcome was to determine classification error rates based on the positioning of individual observations from either group (longleaf pine or the other southern pines), as opposed to simply taking the percentage of the erroneous observations in the corresponding data group, the latter determined by counting the number of observations falling to the wrong side of Koehler's (1932)

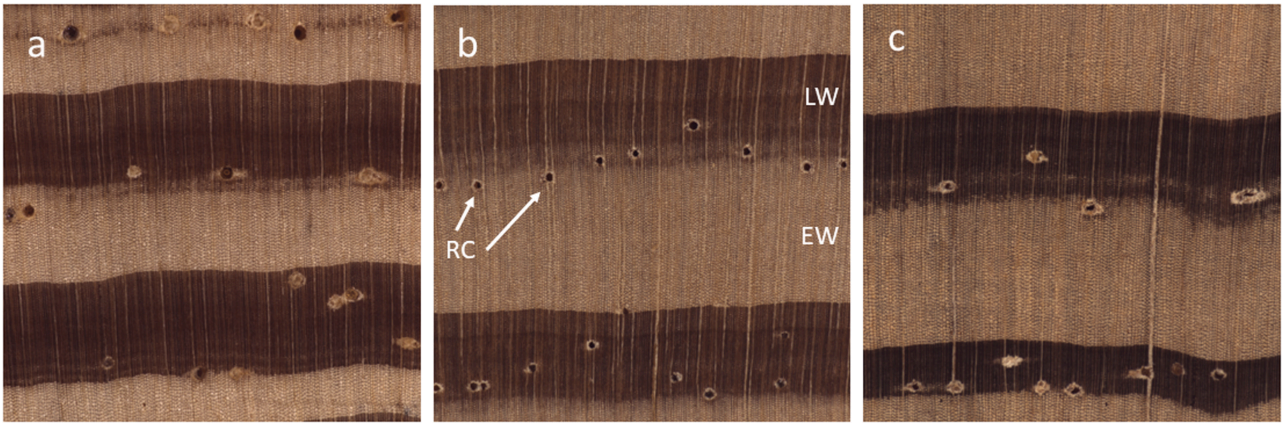

Fig. 5. Representative images of a) longleaf, b) shortleaf, and c) loblolly pine wood specimens showing earlywood growth (EW), latewood growth (LW), and resin canals (RC); each image is 6.35 mm × 6.35 mm.

delineating curve.

For the cross-section data collected at stump height (n = 46), the error rate for a false positive classification for longleaf pine by QDA was 9.52% and was attributed to the two observations nearest the zero-contour delineating the two groups (longleaf pine vs. other southern pines). The first data point attributed to a false positive (pith = 2.84 mm, second ring = 26.52 mm) is easy to rationalize, with it falling almost on the zero contour (Fig. 3a). The second false positive observation (pith = 5.12 mm, second ring = 54.88 mm) is particularly interesting because it had been accurately identified as not being longleaf pine by falling below Koehler's curve. Based on longleaf pine consistently having second annual ring diameters below 40 mm (Eberhardt et al., 2011), perhaps providing another threshold for species classification, the second ring measurement of 54.88 mm would also lead us to accurately identify this observation as one of the other southern pines (i.e., not longleaf pine). Within the context of this dataset of non-longleaf pine observations (n = 21), the observation in question (pith = 5.12 mm, second ring = 54.88 mm) appears to be influential and/or an outlier and calculating the discriminant function without it alters the discriminating contours substantially, resulting in it being classified incorrectly on cross-validation. Altogether, a validation measure of Koehler's method finds the odds of an overall false positive classification as a simple proportion (1/21) to be lower (4.76%) than the estimated error rate obtained by QDA with cross-validation (9.52%); this indicates that there is uncertainty in both Koehler's method and the estimate of the quadratic discriminant function, and further supports statistical approaches to characterize that uncertainty.

4.2. *QDA of cross section data including intermediate heights*

In our prior study (Eberhardt et al., 2011), data were also collected for longleaf pine at additional tree heights (0.75, 6.1, 12.2, and 18.3 m) while loblolly pine data was available for a limited number of trees at a height of 5 m. Adding the longleaf pine (height = 6.1 m) and loblolly pine (height 5 m) data to the statistical analysis gave similar positioning to the QDA contour lines, albeit with slightly wider generalized squared distances. The zero contour (Fig. 3b) for the larger dataset (n = 77) essentially traced the zero contour for the smaller dataset (n = 46), indicating that the addition of observations from multiple heights does not greatly alter the underlying bivariate densities.

There were still two false positive classifications for longleaf pine, but with the additional observations, the error rate decreased to 6.25% (Table 1). The peculiarity that arises is that by cross validation, we now see a false negative classification with an accompanying error rate of 2.22%. This value is 10-fold greater than what Koehler (1932) observed for the odds of a false negative classification, with only a single longleaf pine specimen out of a total of 505 (or 0.2%) being misidentified as being one of the other southern pines (i.e., not longleaf pine). Aside from Koehler attributing this occurrence to a "deformed" pith, it nonetheless is not a cross-validated error rate and likely biased downward (i.e., overly optimistic). Altogether, the application of QDA for both above datasets gave either one false positive or one false negative more than obtained with either dataset (n = 46 or n = 77) evaluated with Koehler's delineating curve.

Koehler's method is known to be most accurate at stump height. In our earlier study (Eberhardt et al., 2011), the data were also presented for longleaf pine at different tree heights (0.75, 6.1, 12.2, and 18.3 m) besides stump height (0.15 m). Plotting the pith and second annual ring diameter data against height for each tree supported the trend noted by Koehler (1915a, 1915b), particularly with second annual ring diameters, increasing rapidly and then gradually decreasing towards the crown (plots not shown). A linear model based on simple functionals of height (height and natural logarithm of height) was stronger with second annual ring diameter than pith diameter, the latter changing little with height. The linear model (which also considered a random tree effect and heterogeneity in pith and second annual ring diameters) indicated that higher pith diameters coincide with higher second annual ring diameters across the range of heights ($\widehat{\rho}= 0.767$). Although pith diameters did not appear to vary as much as the second annual ring diameters, wider pith diameters still coincided with wider second annual ring diameters. Thus, Koehler's curve corrects for rapid tree growth in southern pines other than longleaf pine by excluding the other southern pines with large pith diameters resulting from rapid growth as their second annual ring diameters would exceed the relationship exhibited by longleaf pine.

4.3. *QDA of increment core data at stump height*

Stump height data from increment cores collected during dendrochronological studies (Bhuta et al., 2008, 2009) was made available to give a new dataset with 5-fold higher number of observations (n = 206) from our initial assessment of cross sections at stump height (n = 46). The observations from this dataset gave a different contour plot of the generalized squared distances for group membership (Fig. 3c) such that the contour lines now deviate from the pattern of Koehler's curve. Specifically, we now observe that the zero contour is flatter, and crosses over the Koehler curve. While there was no effect on the absolute number of false positive classifications, the increase in the size of the dataset decreased the false positive rate error rate to 3.13% (Table 1). There was a slight increase in the false negative error rate from 0% to 1.41% from two observations from the dendrochronological dataset. The number of false negative observations is greater by QDA than obtained using the Koehler curve, with one of the observations being above the curve (pith, second ring; 2.290, 19.466 mm) and the other (pith, second ring; 3.179, 36.402 mm) exactly on the curve (pith, second ring; 3.179, 36.402 mm). The equation we generated for Koehler's curve was our best approximation for that presented in the 1932 publication. It is a rare occurrence for one of our observations to fall exactly on the curve down to 1 μm, and technically give neither a false positive nor false negative result.

Southern pine stumps and stem sections discovered in southeastern Virginia could be used to document the native range for longleaf pine if they could be identified, with a high degree of confidence, as longleaf pine. Up until this point, using the Koehler curve, we applied the simple probability of false positive results from our known sample datasets. Thus, in this case, because the southern yellow pine mire sample (pith = 4.47 mm, second annual ring = 22.25 mm) was not known to be at stump height, the error rate has the potential to be higher than the probability value of 0.0625 for a false positive when simply using the proportion of false positives classifications (2/32), even when using the dataset (n = 77) inclusive of data at stump height (0.15 m, 0.5 m) and higher up the tree (5 m, 6.1 m). However, such an oversimplified assessment does not consider the position of that observation relative to either group. Clearly, with an observation in the middle of a cluster of longleaf pine observations, well away from any of the observations for the other two southern pines, the probability value of 0.0625 does not appear to be excessively conservative. Indeed, even the false positive error rate of 3.13%, for the expanded dataset, would appear to be inappropriate.

4.4. *The southern yellow pine mire sample*

Within the context of identifying southern yellow pines, this methodology can aid dendrochronologists in identifying unknown samples as longleaf pine apart from other southern yellow pine species in archaeological or climatological research. This method could also help to solve other questions regarding the biogeography of longleaf pine. Our mire sample from southeastern Virginia demonstrates that lightwood and timber samples can be positively identified as longleaf pine, both by plotting the diameter values (pith = 4.47 mm, second annual ring = 22.25 mm) against Koehler's delineating curve (see Fig. 3a-c) and QDA. Since there is some debate on the range of longleaf pine in Virginia

(Little, 1971; Frost, 2006), this method can be used to identify longleaf pine relicts for accurate mapping.

The error of falsely identifying one of the other species as longleaf pine (a false positive) could lead to the unfortunate consequence of proposing range extensions for longleaf pine beyond its true historical range. Conversely, a false negative result, that is identifying a longleaf pine specimen as being one of the other southern pines, would be less impactful since there would be no challenge to the established historical range. By applying this statistical approach to an unknown southern pine tree section unearthed in a wetland mire at Joseph Pines Preserve in Sussex County, Virginia, we were able to provide probabilities of its false assignment to longleaf pine with greater confidence. To our knowledge, this is the first successful effort to radiocarbon date a suspected longleaf pine timber from antiquity and statistically identify it as longleaf pine with confidence. This is a remarkable finding and highlights the potential of identifying unknown samples as longleaf pine or not.

5. Conclusion

QDA provided statistically based estimated error rates for false positive and false negative determinations for longleaf pine when compared to other southern yellow pines. The cross-validated false positive error rates for the tree disk datasets were higher than that determined using the empirical plot. With a five-fold increase in the data set used for the QDA, similar false positive error rates (probabilities) were obtained. By calculating the posterior probabilities for an unknown southern yellow pine tree section, extremely high values (≥ 0.9998) for an assignment to longleaf pine were achieved. Altogether, while the statistical approaches used here did not appear to improve the partitioning of longleaf pine away from the other southern pines in the datasets, for individual unknown specimens, the assignment to longleaf pine can be done with an extremely high degree of confidence.

Funding

This research was funded in part by the USDA Forest Service and data contributed from author Bhuta was funded from a National Science Foundation (USA) Doctoral Dissertation Research Grant, Award Number 0927687.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank A.C. Wiedenhoeft and his research group at the USDA Forest Service, Forest Products Laboratory for providing images of southern yellow pine cross sections.

References

Bhuta, A.A.R., Kennedy, L.M., Copenheaver, C.A., Sheridan, P.M., Campbell, J.B., 2008. Boundary-line growth patterns to determine disturbance history of remnant longleaf pine (*Pinus palustris* P. Mill.) in mixed forests of southeastern Virginia. J. Torrey Bot. Soc. 135 (4), 516–529.

Bhuta, A.A.R., Kennedy, L.M., Pederson, N., 2009. Climate-radial growth relationships of northern latitudinal range margin longleaf pine (*Pinus palustris* P. Mill.) in the Atlantic coastal plain of southeastern Virginia. Tree-Ring Res. 65 (2), 105–115.

Bhuta, A.A.R., 2011. Variation in the structure, composition, and dynamics of a foundation tree species at multiple scales and gradients Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

Bhuta, A.R.R., Kennedy, L.M., 2021. Woody Plant Dynamics in a Foundation Conifer Woodland of the Appalachian Foothills, AL, USA Southeastern Naturalist, 20, 3, pp. 498–520.

Butler, C.B., 1998. Treasures of the Longleaf Pines: Naval Stores. Tarkel Publishing, Shalimar, Florida.

Cook, E.R., Nance, W.L., Krusic, P.J., Grissom, J., 1998. Modeling the differential sensitivity of loblolly pine to climatic change using tree rings. In: Fox, S., Mickler, R. A. (Eds.), The Productivity and Sustainability of Southern Forest Ecosystems in a Changing Environment. Springer, New York, pp. 717–739.

Davis, C.S., 1981. A SAS macro for calculating the quadratic discriminant function. In: Proceedings of the Sixth Annual SAS Users Group International Conference (SUGI 81). Orlando, FL. February 8–11, 1981. Paper 96, pp. 531–536.

Devall, M.S., Grender, J.M., Koretz, J., 1991. Dendroecological analysis of longleaf pine (*Pinus palustris*) forest in Mississippi. Vegetatio 93, 1–8.

Douglass, A.E., 1941. Crossdating in dendrochonology. J. For. 39 (10), 825–831.

Eberhardt, T.L., Sheridan, P.M., Mahfouz, J.M., So, C.-L., 2007. Old resinous turpentine stumps as an indicator of the range of longleaf pine in Southeastern, Virginia. In: Proceedings of the sixth Longleaf Longleaf Alliance Regional Conference. Longleaf Alliance Report No. 10. pp. 79–82.

Eberhardt, T.L., Sheridan, P.M., Mahfouz, J.M., 2009a. Monoterpene persistence in the sapwood and heartwood of longleaf pine stumps: assessment of differences in composition and stability under field conditions. Can. J. For. Res. 39, 1357–1365.

Eberhardt, T.L., Sheridan, P.M., Reed, K.G., 2009b. Surfing the Koehler curve: revisiting a method for the identification of pine stumps and logs. In: Bowersock, E.P., Hermann, S.M., Kush, J.S. (Eds.), Forestry in a Changing World: New Challenges and Opportunities. Longleaf Alliance Report No. 14, pp. 85–86.

Eberhardt, T.L., Sheridan, P.M., Bhuta, A.A.R., 2011. Revivification of a method for identifying longleaf pine timber and its application to relicts in southeastern Virginia. Can. J. For. Res. 41, 2440–2447.

Eberhardt, T.L., So, C.-L., Leduc, D.J., 2018. Wood variability in mature longleaf pine: differences related to cardinal direction for a softwood in a humid subtropical climate. Wood Fiber Sci. 50 (3), 323–336.

Eberhardt, T.L., So, C.-L., Leduc, D.J., 2019. Wood property maps showing wood variability in mature longleaf pine: Does getting old change juvenile tendencies? Wood Fiber Sci. 51 (2), 193–208.

Foster, T.E., Brooks, J.R., 2001. Long-term trends in growth of *Pinus palustris* and *Pinus elliottii* along a hydrological gradient in central Florida. Can. J. For. Res. 31 (10), 1661–1670.

Fox, T.R., Jokela, E.J., Allen, H.L., 2007. Silviculture in the southern United States. J. For. 105 (7), 337–347.

Friend, A.L., Hafley, W.L., 1989. Climatic limitations to growth in loblolly and shortleaf pine (*Pinus taeda* and *P. echinata*): a dendroclimatological approach. For. Ecol. Manag. 26 (2), 113–122.

Frost, C.C., 1993. Four centuries of changing landscape patterns in the longleaf pine ecosystem. In: Proceedings of the Tall Timbers Fire Ecology Conference. No. 18, pp. 17–43.

Frost, C.C., 2006. History and future of the longleaf pine ecosystem. In: Jose, S., Jokela, E.J., Miller, D.L. (Eds.), The Longleaf Pine Ecosystem, Ecology, Silviculture, and Restoration. Springer, New York, New York, pp. 9–42.

Gamble, T., 1921. Naval Stores: History, Production, Distribution, and Consumption. Review Publishing and Printing Company, Savannah, Georgia.

Gardner Jr., F.H., 1989. Wood naval stores. In: Zinkel, D.F., Russell, J. (Eds.), Naval Stores: Production, Chemistry, Utilization. Pulp Chemicals Association, New York, New York, pp. 143–157.

Harley, G.L., Maxwell, J.T., Oliver, J.S., Holt, D.H., Bowman, J., Sokolosky-Wixon, M., 2018. Precision dating and cultural history of the La Pointe-Krebs House (22JA526), Pascagoula, Mississippi, USA. J. Archaeol. Sci.: Rep. 20, 87–96.

Henderson, J.P., Grissino-Mayer, H.D., 2009. Climate tree growth relationship of longleaf pine (*Pinus palustris* Mill.) in the Southeastern Coastal Plain, USA. Dendrochronologia 27, 31–43.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R. Springer, New York.

Kelley, S.S., Rials, T.G., Groom, L.R., So, C.-L., 2004. Use of near infrared spectroscopy to predict the mechanical properties of six softwoods. Holzforschung 58 (3), 252–260.

Koehler, A., 1915a. A visual method of distinguishing longleaf pine. Am. Lumberm. 2104, 34–35.

Koehler, A., 1915b. Longleaf pine distinguished visually from loblolly or shortleaf. Eng. Record 71, 319–320.

Koehler, A., 1917. A method of distinguishing longleaf from shore and loblolly pine ties or timbers. Guidebook for the Identification of Wood Used for Ties and Timbers. Government Printing Office, Washington, pp. 73–74.

A. (anonymous) Koehler, 1921. A visual method of distinguishing longleaf from shortleaf [and] loblolly pine The New York Lumber Trade Journal, 70, 838, 29.

Koehler, A., 1932. The identification of longleaf pine timbers. South. Lumberman 145, 36–37.

Kukachka, B.F., 1960. Identification of coniferous woods. Tappi J. 43 (11), 887–896.

Kush, J.S., Meldahl, R.S., McMahon, C.K., Boyer, W.D., 2004. Longleaf pine: a sustainable approach for terrestrial carbon in the southern United States. Environ. Manag. 33 (Suppl. 1), s139–s147.

Li, J., 2006. Application of PROC DISCRIM and PROC LOGISTIC in credit risk modeling. In: Proceedings of the 31st Annual SAS Users Group International Conference (SUGI 31). San Francisco, CA. March 26–29, 2006. Paper 081–31. 9 pp.

Little, E.L., 1971. Atlas of United States Trees, Vol. 1: Conifers and Important Hardwoods. US Department of Agriculture Miscellaneous Publication 1146.

Meldahl, R.S., Pederson, N., Kush, J.S., Varner III, J.M., 1999. Dendrochronological investigations of climate and competitive effects on longleaf pine growth. In: Wimmer, R., Vetter, R.E. (Eds.), Tree Ring Analysis: Biological, Methodological and Environmental Aspects. CABI Publishing, Oxfordshire, UK, pp. 265–285.

Oswalt, C., Guldin, J.M., 2021. Status of Longleaf Pine in the South: An FIA Update. Unpublished Report, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina, 37 pp.

Outcalt, K.W., Sheffield, R.M., 1996. The longleaf pine forest: trends and current conditions. Resources Bulletin SRS-9, U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina, 23 pp.

Outland III, R.B., 2004. Tapping the Trees: The Naval stores Industry in the American South. Louisiana State University Press, Baton Rouge, Louisiana.

Panshin, A.J., de Zeeuw, C., 1980. Textbook of Wood Technology, Fourth ed. McGraw-Hill, New York, New York.

Patterson, T., Knapp, P., 2018. Longleaf pine stump in the Uwharrie Mountains of North Carolina. Southeastern Geogr. 58 (1), 1–3.

Pederson, N., Varner III, J.M., Palik, B.J., 2008. Canopy disturbance and tree recruitment over two centuries in a managed longleaf pine landscape. For. Ecol. Manag. 254 (1), 85–95.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL ⟨https://www.R-project.org/⟩.

Ravindran, P., Thompson, B.J., Soares, R.K., Wiedenhoeft, A.C., 2020. The XyloTron: flexible, open-source, image-based macroscopic field identification of wood products. Front. Plant Sci. 11, 1015.

Rouse Jr., P., 1988. The Timber Tycoons, the Camp Families of Virginia and Florida and their Empire 1887-1987. The William Byrd Press, Richmond, Virginia.

SAS Institute, Inc, 2016. SAS/STAT® 14.1 User's Guide. SAS Publishing. SAS Institute Inc., Cary, NC.

Sheridan, P., Scrivani, J., Penick, N., Simpson, A., 1999. A census of longleaf pine in Virginia. In: Kush, J.S. (Ed.), Longleaf Pine: A Forward Look, Longleaf Alliance Report No. 4. pp. 154–162.

Speer, J.H., 2010. Fundamentals of Tree-ring Research. University of Arizona Press,, Tucson, AZ.

Therrell, M.D., Mozayen, B.S., Gage, M.D., 2017. The Search for Fort Armstrong: Dendroarchaeology of the Williamson, 43. Dendrochronologia, "Snow Hill" Plantation, Cherokee County, Alabama USA, pp. 59–65.

Travis, D.J., Meentemeyer, V., 1991. Influence of glaze ice storms on growth rates of loblolly pine *Pinus taeda* and shortleaf pine *Pinus echinata* in the Southern Appalachian Piedmont. Clim. Res. 1, 199–205.

van de Gevel, S.L., Hart, J.L., Grissino-Mayer, H.D., Robinson, K.W., 2009. Tree-ring dating of old-growth longleaf pine (*Pinus palustris* Mill.) logs from an exposed timber crib dam, Hope Mills, North Carolina, U.S.A. Tree-Ring Res. 65 (1), 69–80.

Wahlenberg, W.G., 1960. Loblolly Pine: Its Use, Ecology, Regeneration, Protection, Growth and Management, School of Forestry. Duke University, Durham, North Carolina.

Williams, M., 1989. Americans & their Forests: A Historical Geography. Cambridge University Press, Cambridge.

Zahner, R., 1989. Tree-ring series related to stand and environmental factors in South Alabama *Pinus palustris* stands. In: Miller J.H. (Ed.), Proceedings of the Fifth Biennial Southern Silvicultural Research Conference, General Technical Report SO–74, Southern Forest Experiment Station, USDA, Forest Service, Asheville, North Carolina, pp. 193–197.