Forest Ecology and Management 556 (2024) 121732

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Forest soil classification for intensive pine plantation management: "Site Productivity Optimization for Trees" system

Rachel Cook [a,*], Thomas R. Fox [b], Howard Lee Allen [c], Chris W. Cohrs [d], Vicent Ribas-Costa [a,e], Andrew Trlica [a], Matthew Ricker [f], David R. Carter [g], Rafael Rubilar [h], Otávio Campoe [i], Timothy J. Albaugh [g], Pete Kleto [j], Ed O'Brien [j], Kirk McEachern [b]

[a] Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695-8008, USA
[b] Rayonier, Inc. Forest Research Center, Yulee, FL 32097, USA
[c] ProFor Consulting, Cary, NC, USA
[d] Manulife Investment Management, Chinquapin, NC 28521, USA
[e] Centro para la Conservación de la Biodiversidad y el Desarrollo Sostenible (CBDS), ETSI Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Calle José Antonio Novais 10, 28040 Madrid, Spain
[f] Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695-8008, USA
[g] Virginia Tech, Department of Forest Resources and Environmental Conservation, 228 Cheatham Hall, Blacksburg, VA 24061, USA
[h] Cooperativa de Productividad Forestal, Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepción, Victoria 631, Casilla 160-C, Concepción, Chile
[i] Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil
[j] Forestry and Land Resource Consultants, Inc., Grantville, GA 30220, USA

ARTICLE INFO

Keywords:
Loblolly pine
Soil classification
Geology
Site productivity
Site index
Forest fertilization

ABSTRACT

Forest productivity and response to silvicultural treatments are dependent on inherent site resource availability and limitations. Trees have deeper rooting profiles than agronomic crops, so evaluating the impacts of soils, geology, and physiographic province on forest productivity can help guide silvicultural management decisions in southern pine plantations. Here, we describe the Forest Productivity Cooperative's "Site Productivity Optimization for Trees" (SPOT) system which includes: texture, depth to increase in clay content, drainage class, soil modifiers (*i.e.*, surface attributes, mineralogy, and additional limitations such as root restrictions), geologic formations, and physiographic province. We quantified the total area for each SPOT code in the native range of loblolly pine (*Pinus taeda* L.), the region's most commercially important species, and used a remotely-sensed layer to quantify SPOT code areas in managed southern pine (approximately 14 million ha). The most common SPOT code in the native range is also the most planted, a B2WekoGgPD (fine loamy, shallow depth to increase in clay, well-drained, eroded, kaolinitic, granitic, Piedmont soil), spanning 1.1 million ha total, but only 12% in managed southern pine. However, the SPOT code with the greatest percentage of managed southern pine (61%; a D4PoioAmAF, spodic, deep to increase in clay, siliceous, middle Atlantic Coastal Plain, Flatwoods soil) was the 20th most common in the native range with 474,662 ha. We used machine learning and data from decades of "Regionwide" trials to assess the variable importance of SPOT constituents, climate, planting year, and N rate on site index (base age 25 years) and found that planting year was the most important variable, showing an increase of 17 cm site index per year since 1970, followed by maximum vapor pressure deficit, and precipitation. Geology was the top-ranking SPOT variable to explain site index followed by physiographic province. The Regionwide trials represent 72 unique SPOT codes (out of over 10,000 possible in the pine plantations) and approximately one million ha (or about 7% of all soils identified as supporting managed pine). To extrapolate site index values outside of the unique soil and geologic conditions empirically represented, we created a predictive model with an R^2 of 0.79 and an RMSE of 1.38 m from SPOT codes alone. With this extrapolation, the Regionwide data predicts 10.5 million ha, or 74%, of all soils under loblolly pine management in its native range. Overall, this system will allow managers to assess their current site productivity, and

Abbreviations: FPC, Forest Productivity Cooperative; RW, Regionwide; SPOT, Site Productivity Optimization for Trees.

* Corresponding author.
E-mail address: rlcook@ncsu.edu (R. Cook).

https://doi.org/10.1016/j.foreco.2024.121732
Received 20 November 2023; Received in revised form 22 January 2024; Accepted 23 January 2024
Available online 8 February 2024
0378-1127/

recommend silvicultural treatments, thus, providing a framework to optimize forest productivity in pine plantations in the southeastern US.

1. Introduction

Loblolly pine (*Pinus taeda* L.) is the most economically important and widely planted tree species in the southeastern US, making up 71% of planted timberland in the US and producing more timber than any other country in the world (Johnston et al., 2022; Oswalt et al., 2019; Prestemon and Abt, 2002). Vast improvements in productivity over the decades have been achieved through silvicultural management such as fertilization, vegetation control, and site preparation (Carter et al., 2015; Fox et al., 2007a) in combination with improvements in genetics (McKeand et al., 2021). However, to continue to increase productivity to meet increasing global demand, appropriate management practices must be applied on a site-specific basis.

Forest managers need to know observed and potential productivity, and likely site resource limitations, in order to evaluate if it is feasible to increase growth with silvicultural tools. Site index is a useful measure of productivity because it is generally independent of stand density and sensitive to inherent site quality and silvicultural management (Roth, 1916; Tesch, 1980). Site index is defined as the dominant height of a given species at a particular "base" age, usually age 25 years for planted, managed loblolly pine. Currently, base age 50 site index values are available from the US Department of Agriculture's Natural Resource Conservation Service (USDA NRCS) at a soil series level, but are typically based on decades-old information from unimproved, unmanaged stands and do not accurately represent current managed loblolly pine forests.

Furthermore, the USDA's soil taxonomic system, while an exceptional soil classification system for a broad range of uses, was primarily developed for agricultural soils. Soil series in NRCS are problematic as they were not classified to differentiate forest potential productivity. Long-lived and deeply rooted forest systems are more sensitive to subsoil properties than agricultural systems and series can be mapped across a variety of geologic parent material or formations. Parent material greatly influences soil nutrient availability (Moore et al., 2022), and while in some ways this is part of the USDA Soil Taxonomy, is not always directly incorporated. Unfortunately, non-priority landscapes such as forests, wetlands, and rangelands were typically mapped at a coarser scale (Order 3 soil survey) than intensively managed agricultural areas (Order 2 soil survey, Soil Science Division Staff, 2017), resulting in larger map units that include more dissimilar soil components (complexes, associations, and undifferentiated soil units) which creates greater uncertainty within forested landscapes of the southeastern US. Fortunately, variables can be extracted from map units within the NRCS' Soil Survey Geographic Database (SSURGO; Soil Survey Staff, 2021) that relate to forest productivity and geology can be extracted from other sources.

Much of the early work to specifically classify soils for managed southern pines was performed by Theodore Coile who predicted site productivity with the "factorial method" using continuous variables and relied heavily on subsoil texture and drainage class (Coile, 1952). The limitation of the factorial method lies in the fact that different resources may be limiting at different sites. This system was modified by Fisher and Garbett (1980) into the Cooperative Research in Forest Fertilization (CRIFF) system, specifically targeting soil management in the Atlantic Flatwoods physiographic region. The Strategically Aligned Integrated Silvicultural System (SAISS), developed by Jim Gent and others for Champion and later International Paper, was built on the CRIFF soil system and became one of the most widely utilized and recognized soil systems in plantation forestry. SAISS included soil mineralogy, soil modifiers, and a decision support system relating soils and management to potential growth response across the 2.8 million ha landbase of International Paper, but was never published. Many forestry companies developed soil systems internally to aid in forest management (Morris and Campbell, 1991), but most were regional efforts specific to a relatively small number of soils.

To address the need for southeast-wide site productivity information, the Forest Productivity Cooperative (FPC), a university-public-private partnership that focuses on management of site resources for intensively managed plantation forestry, developed a system that includes factors known to influence forest productivity and response to management from decades of empirical research. Codes, similar to the Coile-CRIFF-SAISS tradition, were developed to capture potential resource availability and limitations such as texture, depth to an increase in clay content, drainage class, mineralogy, root restrictions, geology, and physiographic province across the range of loblolly pine. Now, with decades of field trials spanning the southeastern US and improved GIS capabilities, we have the capacity to test predictions of site productivity and create recommendations for silvicultural management on a site-specific basis.

The goal of this effort was to create a system specifically designed for production forestry that can be used to predict site productivity and inform management decisions. The objectives for this study were specifically to 1) describe the classification system, 2) evaluate the variable importance of soils, geology, physiographic province, climate, nutrient addition, and planting year on site index, 3) evaluate the area of unique codes in loblolly pine management and coverage of empirical trials with site index information, and 4) assess the central tendency and variation of site index values and develop a model to predict site index to extrapolate outside of the unique empirical observations.

2. System development

The FPC Site Productivity Optimization for Trees (SPOT) system classifies site codes from soils, geologic information, and physiographic province (Table 1; Fig. 1). Data sources for SPOT codes included NRCS

Table 1
Example SPOT code categories (*e.g.*, A2WekoGgPD) with descriptions, data sources, and relative scale of each source. NRCS SSURGO: Natural Resource Conservation Service Soil Survey Geographic Database: NRCS SSURGO; SGMC: US Geologic Survey State Geologic Map Compendium; 3DEP DEM: 3 Dimensional Elevation Program Digital Elevation Model; USDA MLRA: US Department of Agriculture Major Land Resource Areas.

SPOT Code Categories	Example	Example description	Data source	Scale
Major soil group (dominant texture)	A	Clay dominant	NRCS SSURGO	1:12,000 to 1:63,360
Depth to increase in clay content (argillic/kandic)	2	12.5–25 cm (5–10 in)		
Drainage class	W	Well drained		
Nature of surface	e	Eroded		
Nature of subsurface (mineralogy)	k	Kaolinitic		
Additional limitations or resources	o	Other		
Geocode (geology, geologic formation, or coastal plain terrace)	Gg	Granite and gneiss	Geology: SGMC Coastal plain terraces: 3DEP DEM	1:50,000 to 1:1000,000 30 × 30 m
Physiographic province	PD	Piedmont	USDA MLRA	1:2000,000

SSURGO (Soil Survey Staff, 2021), the US Geologic Survey State Geologic Map Compendium (SGMC; Horton et al., 2017), elevation-derived coastal plain terraces (U.S. Geological Survey, 2020), and Major Land Resource Areas (MLRA; United States Department of Agriculture, 2022). There is not a direct translation from NRCS soil series or phase to SPOT codes for two reasons, 1) some series have enough variation to code into different groups, and 2) the map units within NRCS SSURGO data are made up of multiple polygons that often span multiple geologic and/or physiographic province codes.

This classification system covers the current range of managed loblolly pine plantations across the southeastern US. Soils were classified in the states of AL, AR, FL, GA, LA, MS, NC, SC, TN, TX, VA, and OK. While portions of these states are not under loblolly pine production due to climate or suitability for agricultural crops, we classified the entire area for regional continuity and future potential increase in planted range.

Each variable within the SPOT system code is detailed below, with tables providing notes for each based on decades of experience from researchers, forest managers, and forest soil mappers. It is important to note that while these general considerations are given for each variable, combinations of soil modifiers, land-use history, influence of geologic formations, or other local conditions may override the general notes. Additionally, soils should be field validated to ensure the given codes are a valid reflection of what is actually present within a particular forest stand. Soils are inherently spatially variable and mapping resolution varies greatly (Table 1).

2.1. Major soil groups

The first letter of the SPOT code primarily corresponds to whole profile texture to a restrictive layer (Table 2). Soil profile texture is calculated from a weighted average of all horizons down to a 200 cm (80 in) depth and corresponds to the USDA soil texture codes. Groups A, B, and C are in order from fine-textured clays to coarse loams. Spodic soils (group D) tends to be particularly responsive to sufficient fertilization. The SPOT system also includes silty texture soils (group E), and deep sands either with an increase in clay content below 100 cm (40 in; groups F) or no clay subsoil (group G). Organic soils, or Histosols, are in group H.

2.2. Depth code

The depth code (Table 3) is based on where in the profile there is an increase in clay content according to the USDA Taxonomic definition of an argillic or kandic diagnostic subsurface (a Bt horizon, Soil Survey Staff, 2022):

Case 1). a 3% absolute increase if overlying horizons are < 15% clay;.

Case 2). a 20% relative increase if overlying horizons are 15–40% clay;.

Case 3). an 8% absolute increase if overlying horizons are > 40% clay;.

For Case 1, a minimum clay percentage threshold of 10% exists to circumvent scenarios where total clay percent never exceed 10% throughout the entire profile. In some cases, if there is missing information as to where this increase occurs within the top 50 cm (0–20 in.), then a depth code of “0” is given for “unknown, but within 20 in..”

2.3. Drainage class codes

Drainage class is important for operational management decisions, such as species selection, bedding, and harvest operability. A shallow depth to water table can restrict rooting volume due to lack of soil oxygen. Definitions of drainage class (Table 4) can differ somewhat, but are generally defined by color and volume of redox depletions (>2%), which can indicate seasonal high water table. However, the water table will be lower in a mature pine stand than one that has recently been harvested, create operational challenges for site preparation operations. Poor drainage is usually indicated by gray soil redox depletions, Munsell Color book as a value of 4 or more and a chroma of 2 or less (Soil Survey

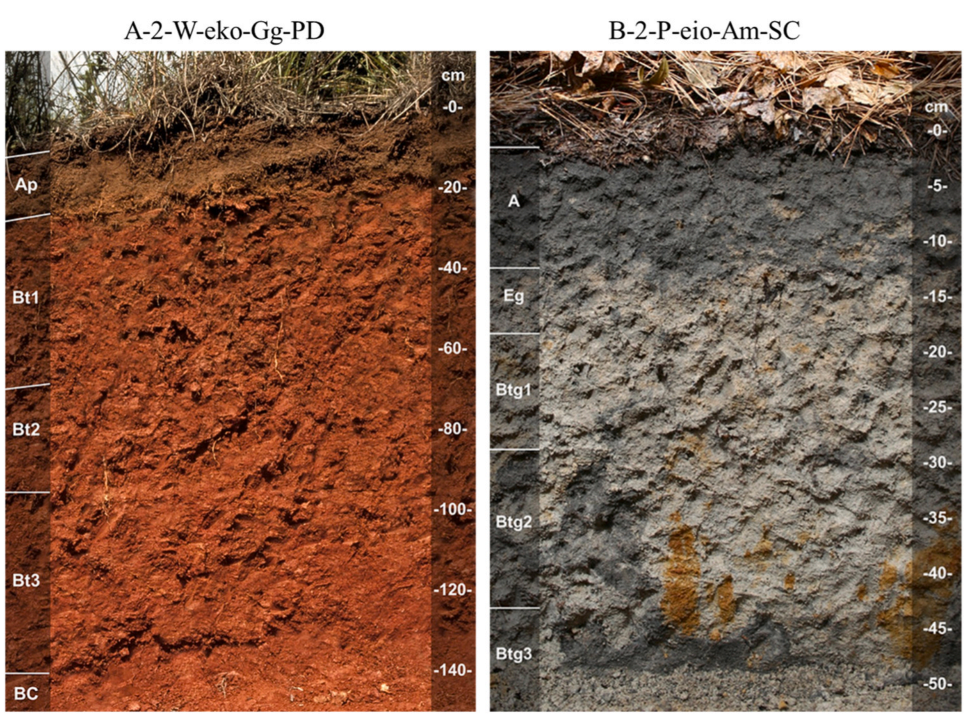

Fig. 1. An NRCS Cecil series (left, Fine, kaolinitic, thermic Typic Kanhapludult, granitic and gneiss geology, Piedmont physiographic province) and a Rains series (right, Fine-loamy, siliceous, semiactive, thermic Typic Paleaquult, middle Atlantic Coastal Plain geology, Southern Coastal Plain physiographic province) can be coded into multiple SPOT codes depending on dominant profile texture, depth to increase in clay content, and geologic parent material. (Photo credit John Kelley, USDA NRCS).

Table 2
Major soil group factor with USDA textures included and general comments included. Notes regarding observed productivity issues or opportunites for each group are included for background.

Soil Group	Dominant Profile	Textures Included and Special Characteristics	General comments regarding relation to loblolly pine
A	Clayey	Clay, sandy clay, silty clay	May have issues with water infiltration/permeability unless good structure is present. Soil strength may be high when dry.
B	Fine loamy	Sandy clay loam, clay loam	Few root growth restrictions, and potentially higher fertility.
C	Coarse loamy	Loam, sandy loam	Sandier texture means more likely to need fertilization, but provides better internal drainage on wetter sites.
D	Spodic	Spodic or Spodosol, usually sandy	Highly responsive to fertilization.
E	Silty	Silt, silt loam	Higher potential for erosion and compaction during harvest.
F	Deep subsoil	Sandy clay loam, clay loam, subsoil grossarenic (>100 cm or 40 in)	Deep sands, low potential productivity, fertilization critical to success.
G	Sandy	Sand, loamy sand, no clay subsoil	Extremely deep sands, low potential productivity. Responsive to fertilization, but best suited for longleaf pine.
H	Organic	Organic (>60 cm or 24 in. of organic),histic or Histosol	High potential productivity when drained, and bedded.
Miscellaneous Soil Groups (no soil attribute information)			
P	Borrow Pit/Mine/Disturbed Area/Industrial Waste Pit		
Q	Sandy Alluvial land		
R	Rock Outcrop		
S	Swamp/Marsh		
U	Gullied Land		
V	Dumps/Dams/Quarries/Urban land/Udorthents		
W	Water		

Table 3
Depth groups relate the appearance of an increase in clay content (argillic or kandic horizon). Definition of an "increase" is based on USDA Soil Taxonomic definition.

Depth Group	Depth to Clay Increase	General comments regarding relation to loblolly pine
1	0-12.5 cm (0–5 in.)	Thin topsoil (A horizon); subsoil close to surface; may be eroded; may need fertilization; if A horizon is dark and well-developed, may not have nutrient limitations.
2	12.5-25 cm (5–10 in.)	Generally, an ideal depth from surface, providing ideal growing environment.
3	25-50 cm (10–20 in.)	If surface is sandy, may need additional nutrient additions. If surface is finer textured and dark, should have relatively high productivity.
4	50-100 cm (20–40 in., arenic)	If surface is sandy, may have nutrient or water limitations.
5	100-200 cm (40–80 in., grossarenic)	Likely to be very nutrient limited.
6	None within 200 cm (80 in.)	Low productivity potential; may be a candidate site for longleaf or sand pine.
0	Unknown within 0–50 cm (0–20 in.)	Information missing from database, field validation required.

Staff, 2022). Very poorly drained mineral soils may also have a dark surface or mollic, umbric, or histic epipedons (20–40 cm of organic soil materials). These sites can be extremely productive where historically ditched and drained and are bedded to control for excess water. As soil becomes more poorly drained, bedding height and quality become increasingly important.

Table 4
Drainage class soil groups are classified by depth of redox depletions (*i.e.*, mottles or matrix with gray color of Munsell value of 4 or more and a chroma of 2 or less) with notes regarding restrictions and opportunities for loblolly pine management or potential alternate conifer species.

Class	Drainage	Definition	General comments regarding relation to loblolly pine
E	Excessively drained	Tend to be sandy with no redox depletions; water is removed very rapidly.	Potential for water limitations; may be a candidate site for longleaf or sand pine.
D	Somewhat excessively drained	Tend to be loamy capped over sandy with no redox depletions; water is removed rapidly.	Potential for water limitations; may be a candidate site for longleaf pine.
W	Well drained	Redox depletions at 100-150 cm (40–60 in); water is removed readily.	No excess water limitations; wetness does not inhibit root growth; generally provides sufficient soil moisture
M	Moderately well drained	Redox depletions at 50-100 cm (20-40 in); water removed somewhat slowly during brief periods.	No excess water limitations; wetness does not inhibit root growth; provides good soil moisture; wet for short period during growing season.
S	Somewhat poorly drained	Redox depletions at 25–50 cm (10–20 in); surface 0-25 cm (0–10 in) does not have gray colors and < 50% gray matrix anywhere between 25-50 cm (10-20 in); not hydric; water removed slowly.*	Single bedding needed; wet for significant periods during growing season, may inhibit growth in young stands; wet upland sites with sufficient slope would likely not benefit from bedding.
P	Poorly drained	Redox depletions at 0–25 cm (0–10 in), Ochric epipedons over > 50% depleted/gleyed matrix from 25 cm (10 in) down; hydric; water removed very slowly.	* *Double bedding, or single beds with equivalent heights needed; wet for long periods during growing season; free water at surface.
V	Very poorly drained	Free water constantly at 0–25 cm (0–10 in); organic soils or mineral soils with dark surfaces or mollic, umbric, or histic epipedons; > 50% depleted/gleyed matrix from 25 cm (10 in) down; hydric.	Double bedding and usually ditching and draining required for survival and growth response; soils almost always wet, swampy.

*Sandy Spodosols in the coastal plain may not have iron in the parent material and follow different conventions to determine drainage class. **Double bedding refers to two passes of a bedding plow to achieve greater bed height. Some operations can also achieve sufficient bed height with a single pass.

2.4. Soil modifiers

Soil modifiers are meant to capture additional information that may influence operational decisions, benefits, or risks, and/or potential productivity. Soil modifiers are organized into three categories: Modifier 1) nature of surface soil (Table 5), Modifier 2) nature of subsoil (Table 6), and Modifier 3) additional limitations or resources (Table 7). Each table contains notes regarding why the modifier is informative and what operations may be affected. The degree of influence in many cases will depend on the severity of the situation (*e.g.*, severely eroded sites might have lower productivity than less eroded sites).

Table 5
Modifier 1: Nature of surface soil (in order of expected importance) describes characteristics in the soil surface that may impact pine productivity positively or negatively.

Modifier	Nature of Surface	General comments regarding relation to loblolly pine
d	Dark surface (Mollic, Umbric, Organic)	Increases productivity potential; high level of organic matter increases water and nutrient holding capacity and nutrient availability.
y	Silty surface (top 15 cm/ 6 in.)	Greater potential for erosion and difficulty maintaining beds. Logging on loess silt, when wet, can create a slurry that dries with high bulk density and seals soil surface. Loess caps in West Gulf region tend to be more productive.
e	Eroded (moderate-severe)	Slight to moderate erosion common in Piedmont with low likelihood of negative effects in productivity. Severe erosion can decrease productivity.
g	Gullied	Heavily eroded; lower productivity can be found where severe erosion occurred; potential problems with machinery operations.
r	Rocky (skeletal to gravelly, coarse fragments)	Negative potential effects in productivity by reduction in rooting volume; can increase harvest operability in wet conditions.
o	Other or NA	No documented features to consider.

Table 6
Modifier 2: Nature of subsoil (in order of expected importance) describes characteristics in the clay mineralogy that may impact pine productivity positively or negatively.

Modifier	Nature of Subsoil	General comments regarding relation to loblolly pine
a	Alfic	High base saturation (CEC is > 35% Ca, Mg, K) reduces potential nutrient limitations.
m	Micaceous	Mica supplies potassium, but clayey subsoils near surface may cause operability issues (*i.e.,* slippery on slopes).
x	Mixed (2:1 clays)	Higher potential productivity, these clays have a relatively high cation exchange capacity.
p	Plastic (smectitic/ shrink-swell)	Sticky clays can create operability issues. Churning during harvest can create impermeable layers. Plastic subsoil near surface can be slippery.
k	Kaolinitic (1:1 clays)	Relatively less cation exchange capacity, highly weathered, and ubiquitous throughout the Southeast.
i	Siliceous	Formed from parent material with silica (*e.g.,* quartz sands, granite, quartz sandstone). Often relatively low in cation exchange capacity and buffering potential.
o	Other or NA	No information available or undefined.

2.5. Geocodes

The SPOT system "geocode" combines lithology (*i.e.,* rock type), geologic formations, and coastal plain terraces into a two-letter code (Table 8). Lithology is applicable in the Piedmont, Mountains, and regions where parent materials influences nutrient availability or soil formation (residuum). Coastal plain terraces supersede lithology where marine and fluvial sediment deposition has buried rock and become the parent material. Formations and terraces are grouped according to similar site limitations or resource availability and mapped across state lines based on the US Geologic Survey SGMC (Horton et al., 2017).

Coastal plain terraces were created from 30 × 30 m digital elevation map. Some Atlantic Coastal Plain terrace names change across state lines, so we assigned consistent names at the following breaks in elevation: Silver Bluff/Princess Anne (up to 3 m), Pamlico (3.1–10.7 m), Talbot (10.8–13.7 m), Penholoway (13.8–21.3 m), Wicomico (21.4–30.48 m), Sunderland (30.5–51.8 m), Coharie (51.9–60.9 m), and Hazelhurst (61–97.5 m) (Cooke, 1931).

Table 7
Modifier 3: Additional limitations or resources (in order of expected importance) describes characteristics in the soil surface that may impact pine productivity positively or negatively.

Modifier	Additional Limitations or Resources	General comments regarding relation to loblolly pine
c	Alkaline, calcareous	Must check pH, expect reductions in productivity if pH greater than 6.
f	Floods (fluvic)	Unsuitable or high risk for intensive management.
l	Lamella	Narrow (6–22 mm) horizontal layers of clay in sandy soils, commonly found between 72–155 cm depth (28–61 in; Bockheim and Hartemink, 2013), provides additional productivity.
n	Salt affected (natric, saline, or sodic)	May have issues with permeability, bedding in poorly drained soils may help; may have some issues with operability when wet.
s	Root limited (densic, lithic, paralithic) < 25 cm (10 in)	Reduced rooting volume may limit productivity.
t	Root limited (densic, lithic, paralithic) 20–50 cm (10-20 in)	Reduced rooting volume may limit productivity.
u	Root limited (densic, lithic, paralithic) 50–100 cm (20-40 in)	Reduced rooting volume may limit productivity.
v	Root limited (densic, lithic, paralithic) 100–200 cm (40-80 in)	Reduced rooting volume may limit productivity.
q	Restrictions within 100 cm (40 in) (fragic, cemented, plinthic)	Restrictions are less severe than s, t, and u, but may still limit productivity. Roots may be able to bypass restrictions in old root channels.
w	Ponded Water	Unsuitable for intensive management.
o	Other or NA	No additional soil features to include.

Over 200 rock types that occur in the Southeast were assessed based on their weatherability and inherent nutrient availability (Colpitts et al., 1995) and assigned a fertility rating on a scale of 1–9 similar to Hennigar et al. (2017) where 1 is low fertility and 9 is high fertility. For example, quartzite received a "1", granite received a "4", gabbro received a "6", and marl received a "9." We used the "low_lith" column in the SGMC which can have up to three lithological units per polygon that get assigned a fertility ranking code and then averaged. For each geocode, fertility ranks were averaged across all lithologies or terraces per polygon. For coastal plain terraces, we performed a spatial join to the underlying geologic formation and assigned fertility rank based on the geologic formations, but appended the coastal plain notes with additional nutrient deficiency information based on experience from field trials, such as K deficiency in the Pliocene-Pleistocene terraces (Carlson et al., 2014).

2.6. Physiographic Province

Physiographic provinces can be useful for grouping soils for general management guidelines as they have some predominant soil and climate factors at a regional scale (Morris and Campbell, 1991). Major Land Resource Areas have historically been designated by the USDA for agricultural planning. These geographically associated areas are thousands of hectares in extent and are characterized by a particular pattern of soils, climate, water resources, land uses, and agricultural practices. The MLRA codes are grouped here according to similarities for management considerations for loblolly pine (Table 9). As these areas are geographically large (Fig. 2), guidelines for management should consider more specific soils information for a given site.

Table 8
Geocodes include geology, geologic formations, parent material, and coastal plain terraces. Fertility ratings range from 1 (low fertility) to 9 (high fertility). For example, the "grand" average fertility rating per geocode combines all rock types to a group. However, fertility ratings are assigned at a polygon level in GIS (not the grand average). Fertility ratings and groupings may be updated in future versions with more training data.

Geocode	Formation/Rock type	Average Fertility Rating	General comments regarding relation to loblolly pine
Av	Alluvium / Deposits - floodplain, levee, terrace	2.3	Variable, moderately P deficient; potentially high productivity; productivity may vary depending on source of alluvium.
Pa	Pamlico Terrace (Pamlico, Princess Anne, Silver Bluff)	2.7	Severely P-deficient; typically wet and needs bedding.
Al	Atlantic Lower Coastal Plain Terraces (Wicomico/ Windsor/Waccamaw, Penholoway, Talbot)	3.3	P, K, B deficient; typically wet and needs bedding.
Am	Atlantic Middle Coastal Plain Terraces (Argyle, Claxton, Pearson, Sunderland, Coharie, Okefenokee, Waycross)	2.9	P, K deficient.
Au	Atlantic Upper Coastal Plain Terrace (Hazelhurst, Yorktown, Brandywine, Neogene)	4.1	Moderate fertility.
Cb	Claibourne Group (Queen City, Carrizo Sands, Weches, Cockfield, Cook Mt, Sparta, Cane River, Gosport, Lisbon, Tallahatta, McBean)	3	Can have glauconitic rock or marl (rich in K); moderate to good fertility.
Cs	Coarse-textured Sediments (Sandy to Loamy)	2.3	Beach sands, dune sands, sand; low fertility.
Ms	Medium-textured Sediments (Silty)	2.9	Loess or silt; low to moderate fertility.
Fs	Fine-textured Sediments (Clayey)	3.4	Alluvium, clay or mud, delta, peat.
Lo	Loess derived parent material	2.5	Potentially higher productivity, but silty soils may not hold beds and operability is constrained when wet; if sandier, will be more operable.
Ct	Citronelle Formation	1.9	Severely P deficient, often massive to weak Bt structure; well-drained to xeric.
Le	Lumbee Group (Middendorf, Black Creek, Pee Dee)	2.8	K, B deficient, potentially low productivity but some areas have phosphatic parent material.
Ba	Blackland Group (acid, Porters Creek, Naheola, Clayton, Tuscaloosa)	3.2	More acidic soils of Blackland.
Bb	Blackland Group (basic, Navarro, Taylor, Austin, Eagle Ford, Prairie Bluff, Providence, Ripley, Demopolis, Cussetta, Mooreville, Eutaw, Bluffton, Kemp, Corsicana, Nacatoch)	4.2	More basic soils of Blackland; micronutrient deficient; Bluffton often good for wet weather logging (ancient sand dunes)
Dw	Deweyville Formation	2.1	Moderately P-deficient, Texas only; separated from Beaumont because alluvial in nature (3.5–5 m better SI than Beaumont and not as P deficient).
Ch	Catahoula Group (Paynes Hammock, Oligocene undifferentiated)	2.4	Severe to Moderately P-deficient, extremely variable; in TX has numerous problems (shallow, rocky, *etc.*); summits may be productive with a Willis cap.
Fl	Fleming Group (Pascagoula, Hattiesburg, Logarto, Oakville)	2.6	Moderately P-deficient.
Jk	Jackson Group (Caddell, Manning,Yazoo, Moody Branch, Wellborn)	3.8	Variable fertility.
Lb	Lissie-Bentley Formation (Montgomery, Bentley, Prairie, Beaumont)	3	Severely P-deficient; Beaumont and Prairie commonly poorly drained
Vk	Vicksburg Group (Chickassawhay, Nash Creek, Byram, Mint Spring, Forest Hill, Red Bluff, Whitsett)	5.8	Moderate to high fertility.
Ws	Willis Formation	2.4	Moderately-to-Severely P deficient; Willis caps on Catahoula formation in LA can be very productive.
Wx	Wilcox Group (Carrizon, Sabinetown, Pendleton, Hatchetigbee, Tuscahoma, Nanafalia)	2.7	Can have glauconitic rock (rich in K), may have sticky clays, poor structure, ironstone.
Yg	Yegua	2.5	Higher productivity than Jackson group, best soils in East Texas
Md	Midway Group (Naheola, Clayton, Wills Point, Kincaid)	4.3	Moderate to good nutritional properties.
Gg	Granite and Gneiss (Felsic–light colored & low mafic, coarse-grained igneous and metamorphic rocks dominated by feldspar and quartz)	4.8	High K availability, potential Ca availability.
Sc	Schist and Phyllite (fine-grained metamorphic rocks dominated by mica)	5	Ca, Fe, K, Mg availability.
Sd	Sandstone	2.9	Nutrient availability depends on cementing agent.
Sh	Shale/Siltstone (quartz/ feldspar/mica)	4.2	K availability with presence of mica.
Mr	Metamorphic Rock with extremely variable parent rock	4.4	K availability with presence of mica.
Sr	Sedimentary Rock with extremely variable clasts and cement mineralogy	4.2	K availability with presence of mica.
St	Slate (mica)	4.8	K availability with presence of mica.
Bg	Basalt, Gabbro, Greenstone (Moderate to high mafic–dark colored, fine-and-coarse-grained igneous rocks dominated by olivine, pyroxene, biotite, hornblende)	6.3	Ca, Fe, K, Mg availability.
Lm	Limestone (calcite) and Marl	8.1	Ca, Mg availability.
Um	Ultramafic (Igneous and meta-igneous rocks with a very low silica and	6.4	High or excess Mg availability; may affect Ca and K availability.

(*continued on next page*)

Table 8 (*continued*)

Geocode	Formation/Rock type	Average Fertility Rating	General comments regarding relation to loblolly pine
	potassium content, ex: peridotite, komatite, serpentinite)		
Ui	Unimportant or non-applicable (water, ice, landslide, indeterminate, *etc.*)	3.3	Other

3. Methods

3.1. Field site selection and site productivity

We selected 64 study sites contributing a total of 927 field plots from several series of “Regionwide” (RW) studies implemented by the Forest Productivity Cooperative which tested response to nutrient additions across the geographic range in which loblolly pine is managed (Fig. 3; Carter et al., 2021). All included study sites had control (non-fertilized) and fertilized treatment plots with the addition of elemental nitrogen (N) rates ranging from 112 kg ha^{-1} to 538 kg ha^{-1}. All N applications rates had at least 10% accompanying elemental P (Table 10). Control and treatment plots received other silvicultural treatments such as bedding or vegetation control. In the RW18, frequency had no effect on growth response as long as the cumulative dose applied was the same, so treatments were grouped by cumulative rate. Study sites were selected that had at least eight years of consecutive field measurements though fertilization response can last longer (Albaugh et al., 2021).

To calculate site index, each observation in the dataset consisted of one field plot where dominant height was defined as the average height of the biggest 100 trees per ha by stem volume. Site index was calculated using the (Diéguez-Aranda et al., 2006) model:

$$SI = \frac{26.14 + X_0}{1 + (1455/X_0) \times (ba)^{-1.107}} \quad (1)$$

Where X_0 is given by:

$$X_0 = \frac{1}{2}\left(dh - 26.14 + \sqrt{(dh - 26.14)^2 + 4 \times 1455 \times dh \times age^{-1.107}}\right) \quad (2)$$

Where *dh* is the dominant height in meters, *age* is the forest stand age in years, and *ba* is the base age (25 years). This site index model is commonly used in the southeastern US (Allen and Burkhart, 2015) but was parameterized with unthinned, extensively managed stands, so there may be some underestimation of site index in intensively managed stands (Trim et al., 2020). However, as some of the studies represented here were planted in the 1970 s, the earlier Regionwide trials would be more representative of stands used to parameterize the model.

3.2. Variable importance selection with random forest

We used random forest using the randomForest package in R (Liaw and Wiener, 2002; R Core Team 2023), to rank the relative importance of variables for climate, SPOT codes (soils, geocode, physiographic province), fertility rating from geology (see Table 1 for spatial resolution), nitrogen rate, and planting year. Climate variables were calculated as 30-year averages from 1991–2020 on an 800 m grid (PRISM, 2014) and assigned to each Regionwide study. Precipitation was calculated as average annual rainfall in mm. Mean, minimum, and maximum temperature were calculated from daily averages in °C. Maximum and minimum vapor pressure deficit (kPa) and mean dew point temperature (°C) were also daily annual averages. SPOT soils variables included: Major code, depth to clay, drainage, soil modifiers (nature of surface, nature of subsurface, and additional limitations &

Table 9
Physiographic provinces derived from Major Land Resources Areas (MLRA). Geocodes shared refer to areas designated as under loblolly pine management.

Code	Physiographic Province	MLRA Code	Geocodes unique to Phys. Prov.	Geocodes Shared across Phys. Prov.	Loblolly Pine Management Notes
AA	Alluvium	150, 151, 152, 133 A, 133B, 134, 135 A, 135B, 136, 116, 117, 118, 119, 122, 123, 124, 125, 128, 129, 130, 131ABCD	Av		Depending on source, material deposited by rivers and streams often have a higher productivity than soils formed in place. Drainage class influences suitability for pine.
AF	Atlantic Coastal Plain Flatwoods	138, 153, 154, 155, 156	Pa, Al, Am	Cs, Ms, Fs	Many low-lying areas that will require bedding and are often P-deficient. Waxy-leaf competing vegetation common.
BP	Blackland Prairie	135 A, 135B	Ba, Bb		Divided into acid and basic Blackland Prairie in geocodes. Basic can have high alkalinity resulting in poor growth
GF	Gulf Coastal Plain Flatwoods	150, 151, 152	Dw, Lb	Cs, Ms, Fs	Many low-lying areas that will require bedding and are often P-deficient. Waxy-leaf competing vegetation common.
LP	Mississippi Valley Loess Plain	134	Lo		Loess caps can add additional productivity but can be inoperable when wet.
MT	Mountains (Blue Ridge, Ridge and Valley, Appalachian Plateau, Highland Rim, Arkansas Ridge and Valley, Ouachita Mountains, Sand Mountain)	116, 117, 118, 119, 122, 123, 124, 125, 128, 129, 130		Bg, Gg, Lm, Mr, Sc, Sh, Sr, St, Um	Slope and soils need to be considered for harvest operability. Generally avoid slopes > 30%, otherwise manage based on soil codes. Coarse fragments may reduce

(*continued on next page*)

Table 9 (*continued*)

Code	Physiographic Province	MLRA Code	Geocodes unique to Phys. Prov.	Geocodes Shared across Phys. Prov.	Loblolly Pine Management Notes
					rooting volume.
PD	Piedmont	136			Typically well-drained, except for lower and toe slope positions. Rolling landscape. Generally, needs N + P additions.
SH	Sandhills	137	Le		Coarse sandy soils will require substantial nutrient additions to increase productivity.
SC	Southern Coastal Plain	133 A	Au	Cb, Cs, Ct	Refer to Coastal Plain
WG	Western Gulf Coastal Plain	133B	Ch, Fl, Jk, Vk, Ws, Wx, Yg, Md		Terraces in geocode for nutrient status

resources).

Three different combinations of covariates were tested: (i) each group within the SPOT code (soil groups, geocodes, and physiographic province), (ii) SPOT plus climatic variables, and (iii) SPOT, climatic variables, and nitrogen rate. For some variables, we performed a simple linear regression to explore the relationship with site index.

Categorical data, in general, have inherent limitations in that any new level not previously seen by a model cannot be estimated. Similarly, random forest cannot extrapolate beyond the underlying data (Hennigar et al., 2017; Jeong et al., 2016), however it still has benefits: to implicitly deal with auto-correlation and highly dimensional data, to handle interactions between variables, to identify informative inputs using a permutation-based variable importance index, and to capture complicated phenomena and reveal non-linear relationships (Antoniadis et al., 2021; Cheng et al., 2020).

Random forest generally does not overfit data due to its algorithm design (Loecher, 2022) but we still prepared the dataset to optimize the random forest procedure and protect from overfitting. Categorical variables needed to be represented in at least 10 field plots to be included in the analysis (which excluded 135 observations). In addition, 19 SPOT variables, which exist in loblolly pine soils had zero observations in the Regionwide dataset (Table 11). Consequently, for modeling purposes, we had 792 observations and 57 unique SPOT codes. Miscellaneous major group codes (P, Q, R, S, U, V, and W) were not included in the Regionwide data as they do not include sufficient SSURGO information and represent a very small proportion of area.

We followed a sequential model-based optimization directly implemented *via* the *tuneRanger* R package (Probst et al., 2019). However, as we also wanted to consistently control overfitting, we manually established certain thresholds based on our research interests and knowledge of the dataset. We used 2000 trees, two drawn candidate variables in each split, 0.6565809 sample fraction, no replacement when resampling, minimum splitting node size and minimum terminal node size of five, tree depth of 14, splitting rule set to variance, and variable importance criteria set to permutation. The final random forest model was built and implemented *via* the *ranger* R package (Wright and Ziegler, 2017). The final model was evaluated following a 10-fold cross-validation procedure.

To assess the relative stability of variables in the model, the random forest modelling procedure was performed ten separate times. From those ten runs, we calculated the 95% confidence interval for both variable importance and rank order of each variable and tested for differences among variable importance using Tukey's honest significant difference with an alpha value of 0.05.

3.3. Managed southern pine distribution

To assess the relative extent of SPOT codes mapped under managed southern pine, we first determined the total area of each unique SPOT code within the traditional loblolly pine native range area (Little, 1971). We then used a raster layer that identifies intensively managed evergreen stands for the states of AL, AR, FL, GA, LA, MS, NC, SC, TN, TX, VA,

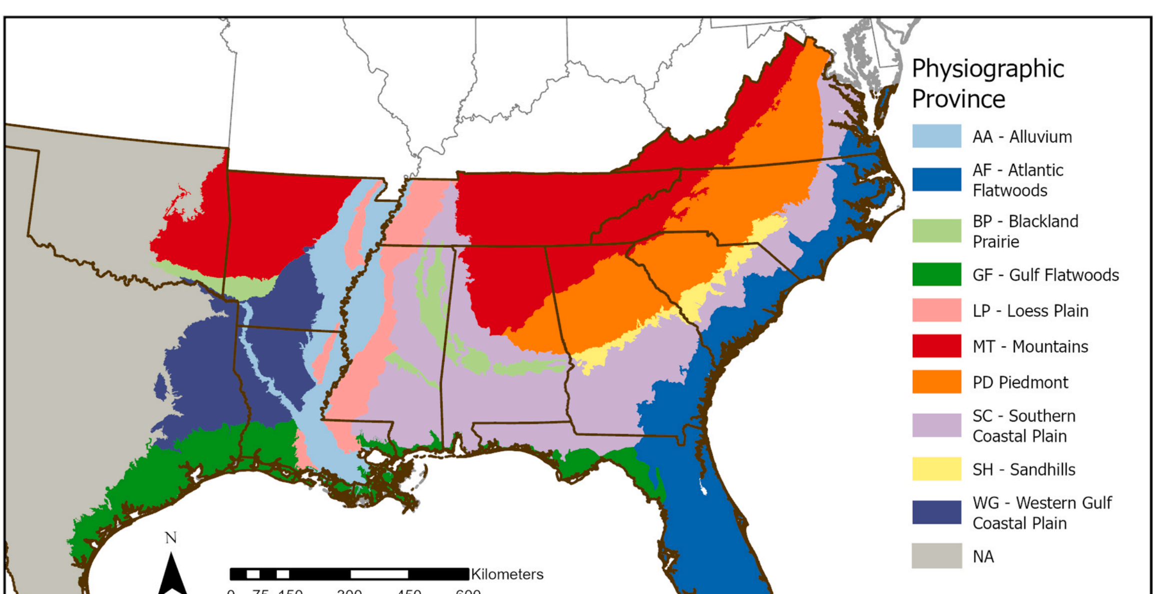

Fig. 2. Physiographic provinces derived from Major Land Resource Areas for loblolly pine plantation management.

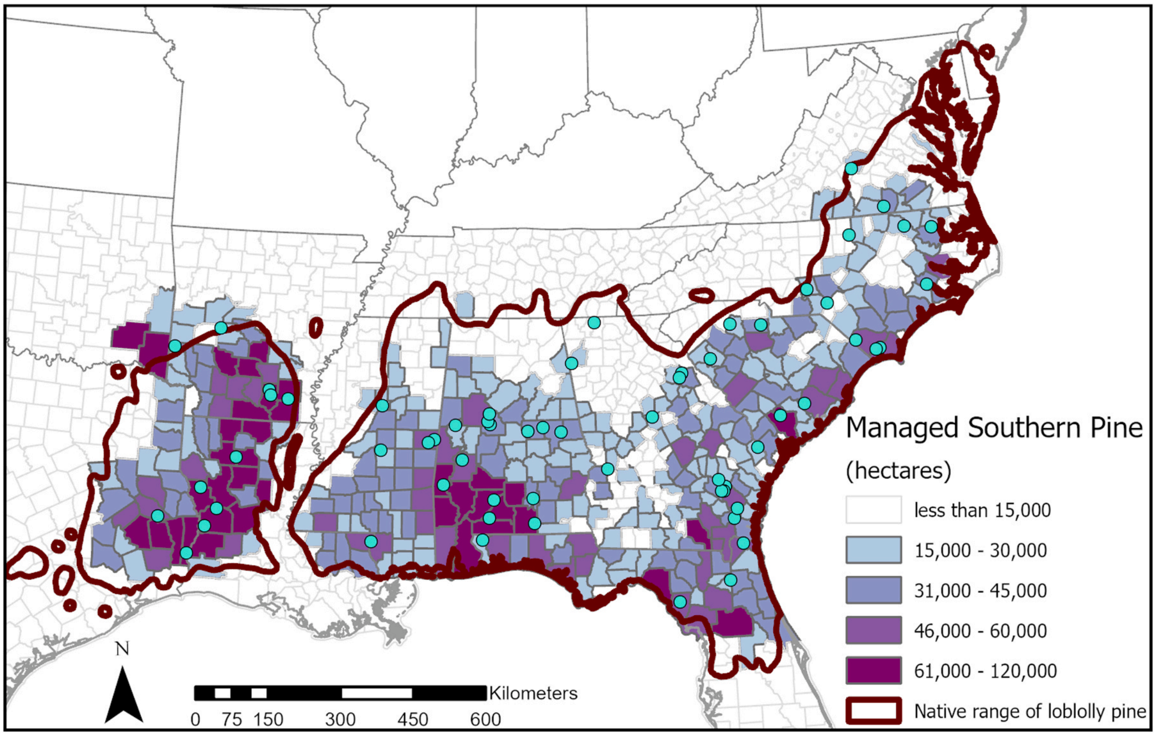

Fig. 3. All 64 study sites are located within or near the natural range of loblolly pine (*Pinus taeda* L.; Little, 1971) in the southeastern United States. Data for model development were collected from field plots measured 8 years after treatment. Shading for each county represents the area considered managed southern pine generated through remotely-sensed data (Thomas et al., 2021).

Table 10
Details of five Regionwide (RW) studies spanning the southeastern US included in the analysis of modeling *Pinus taeda* (L.) site index (base age 25). Tree ages eight years since treatment were the ages used to calculate site index. Nutrient application rates are cumulative rates for the RW18 which were applied at different frequencies and rates.

Study name	Years of planting	Tree ages at 8 years since treatment (years)	Elemental N application rates (kg ha^{-1})	Elemental P application rates (kg ha^{-1})	No. Study Sites	No. Plots	Reference
RW13: Midrotation fertilization with rates of N and P	1970–1977	18–24	112, 224, 336	28, 56	16	296	Fox et al. (2007b)
RW15: Additions of N + P, K, and micronutrients	1970–1985	17–33	224	56	15	219	Carlson et al. (2014)
RW17: Fertilization x vegetation control at Midrotation	1975–1988	17–30	224	56	8	116	Albaugh et al. (2012)
RW18: Rate x frequencies of N + P at juvenile application	1993–2000	10–14	269, 538	27, 54	20	210	Albaugh et al. (2015)
RW19: Thinning x fertilization response	1992–2000	20–24	224	28	5	71	Albaugh et al. (2017)

Table 11
Categorical variables that were not included in the random forest model, therefore could not be included in assessing variable importance or in predictive modeling. Obs. = Observations in the Regionwide trial dataset; Var.= Variable within a SPOT code group.

SPOT Groups	Var.	Name	Obs.	SPOT Code Level	Var.	Name	Obs.
Major	H	Organic soil	0	Geocode	Ch	Catahoula Group	0
Depth	6	No argillic/kandic within 200 cm	0		Dw	Deweyville Formation	0
Drainage	E	Excessively drained	8		Fs	Fine-textured Sediments (Clayey)	0
Nature of Surface	g	Gullied	0		Lm	Limestone (calcite) and Marl	0
Nature of Subsoil	m	Micaceous	6		Md	Midway Group	0
Additional Limitations or Resources	c	Alkaline, calcareous	0		Ms	Medium-textured Sediments (Silty)	0
	f	Floods (fluvic)	0		Pa	Pamlico Terrace	0
	l	Lamella	0		Sd	Sandstone	0
	n	Salt affected (natric)	0		Sc	Schist and Phyllite	1
	s	Root limitation < 25 cm	0		St	Slate (mica)	0
	w	Ponded Water	0		Ui	Unimportant or non-applicable	0
Physiographic Province	LP	Mississippi Valley Loess Plain	0		Vk	Vicksburg Group	0
					Yg	Yegua	0

and OK (Thomas et al., 2021). We refer throughout the manuscript to these remotely-sensed areas as “managed southern pine” as it could include a small amount of other southern pines on short rotations that received a thinning. We know that the native range of loblolly pine is a conservative estimate of planted loblolly pine area, whereas observations outside of the native range, but within the southeastern states, may contain false positives (*e.g.*, misidentification of thinned, evergreen stands as loblolly pine). We excluded 112,205 ha with 12,840 unique codes under managed pine with less than 40 ha (100 ac). We also excluded a total of 1113 ha and 105 unique SPOT codes identified with miscellaneous major codes P, R, U, and V. Each of these miscellaneous codes had less than 40 ha per unique code.

3.4. Predictive modeling with ordinary least squares

We first examined the site index values for all observed RW-trial SPOT codes with at least 10 observations to assess the central tendency, 95% confidence intervals, and range of values. We then tested how accurately we could predict site index in meters (SI_m) by using only the “SPOT model” with an ordinary least squares linear regression (Equation 3).

Equation 3:

$$SI_m \sim Major\ Soil\ Group + Drainage + Depth + Nature\ of\ Surface + Nature\ of\ Subsoil + Additional\ Resources\ or\ Limitations + Geocode + Physiographic\ Province$$

This method can extrapolate SI in SPOT codes not empirically observed in the Regionwide data. We then tested several iterations of ordinary least squares models to assess the relative additional improvement by adding planting year, climate, and N rate to the “SPOT model,” as well as climate and N rate alone, since the random forest model indicated their importance.

We assessed the sample size and the corresponding risk of overfitting based on the following criteria (Riley et al., 2019): 1) Small optimism in predictor effect estimates as defined by a global shrinkage factor of ≥ 0.9, 2) small absolute difference of ≤ 0.05 in the apparent and adjusted R^2, 3) precise estimation (a margin of error $\leq$ 10% of the true value) of the model’s residual standard deviation, and 4) precise estimation of the mean predicted outcome value (model intercept).

To validate all ordinary least squares linear regressions, we performed a 10-fold cross validation approach using the RMS R package (Harrell, 2023), following a similar approach to other similar studies (*i.e.*, Hennigar et al., 2017; Pahlavan-Rad et al., 2020). We also performed a penalization analysis of the ordinary least squares linear regression model *via* the RMS R Package. We used the penalized maximum likelihood estimation, checked the optimal value yielded by the *pentrace* function and applied it to our final ordinary least squares linear regression model. We assessed the results using a 5-fold cross validation procedure as well, which uses more data for each fold but found similar performance to the 10-fold cross validation approach.

For the final prediction model, we assessed standard error in the ordinary least squares model indicating multi-collinearity for some coefficients with small sample size. Therefore, we removed the following levels from the model to prevent multi-collinearity: major group: F and G; depth: 5; drainage: D; geocode: Sh; and physiographic province: MT and SH. These removals reduced the total number of observations to 759 for the model. We used an alpha value of 0.05 for the significance of model and calculated the relative root mean square error (RRMSE).

It should be noted that the Random Forest approach and the Ordinary Least Squares approach may not select the same variables. Additionally, the variables selected may not behave in the same way in each approach. We chose to proceed with using the Ordinary Least Squares approach for predictive modeling because it is a simpler and more interpretable method given our somewhat constrained dataset.

4. Results

4.1. Variable importance for site index

Planting year was the most important variable for predicting site index (Fig. 4), showing a linear increase of 17 cm of site index per year since 1970 ($p < 0.01$). After planting year, maximum vapor pressure deficit was second most important. Both of these variables, in 10 model runs, always remained in first and second rank of importance. Geocode ranked in position 3 or 4 in all model runs and was consistently the highest ranking variable out of any of the SPOT variables. Almost all of the climate variables, except minimum vapor pressure deficit (VPD), fell between geocode and physiographic province. Physiographic province was the second most important SPOT variable and had an average rank of 7.9. Nitrogen rate tended to fall in the central ranks (average rank 11.6) and was of higher importance than all other soil variables. Major code always ranked higher than the other soil codes. Nature of subsoil and drainage ranks overlapped, but depth to increase in clay, nature of surface, and additional limitations and resources were always in the last three ranks, respectively.

4.2. Loblolly pine and regionwide study distribution

Remotely-sensed, managed southern pine made up approximately 17% of total area within the native range of loblolly pine (Table 12). Only about 1% of the total area in the southeastern US that is outside the native range of loblolly was classified as managed southern pine. A large part of the area outside the native range was made up by west Texas, which does not support loblolly pine growth due to climate. SPOT coverage totals are slightly less than total area per state as areas such as water, roads, or urban land, were excluded from the total area.

We identified 10,461 unique SPOT codes with at least 40 ha (100 ac) of managed pine within the native range. Regionwide studies are found on 73 unique SPOT codes but there were only 57 unique SPOT codes in the Regionwide database with at least 10 observations. SPOT codes represented by Regionwide trials covered 1,097,556 ha, or 7% of all area in planted pine in the native range.

The SPOT code most commonly planted in pine is also the most common SPOT codes within the native range. The top five most prevalent soils in the native range of loblolly (light yellow bars, Fig. 5) were fine-textured, well-drained, eroded, kaolinitic soils in granitic geology (B2WekoGgPD, B1WekoGgPD, A2WekoGgPD; 346,859 ha) or in the upper Atlantic Coastal Plain (B3WekoAuSC, B2WekoAuPD; 169,449 ha). However, the largest percentage of loblolly area (dark green bars) for a given SPOT code was made up by poorly drained, sandy Spodosols in the middle Atlantic Coastal Plain (D4PoioAmAF) at 61%. This SPOT code ranks 20th in total soil area for the native range.

4.3. Predictive modeling with ordinary least squares

Site index values for soils that had at least 10 observations ranged between 11.6 and 31.2 m with an average of 21.6 m in total height and a 95% CI of [21.4, 21.9] (Fig. 6). The lowest site index was observed in a D4SoioAmAF, a nutrient deficient Spodosol, planted in 1975.

Since the random forest variable importance analysis showed that geocode was more important than fertility rating, we continued modeling geocode instead of fertility rating for the ordinary least squares predictive modeling. The predictive ordinary least squares model with the highest R^2 and lowest RMSE was SPOT + Planting Year + N rate + Climate (Table 13). The addition of Planting Year + N rate + Climate to the SPOT-only model (soils + geocode + physiographic province) increased the R^2 by 0.08 and decreased RMSE by 0.28 m.

The predicted *versus* the observed analysis for the SPOT-only model of Regionwide data showed a reasonably good fit around the one-to-one line with some over predictions at lower values and under predictions at higher values (Fig. 7). For the SPOT-only model, the R^2 and RMSE for

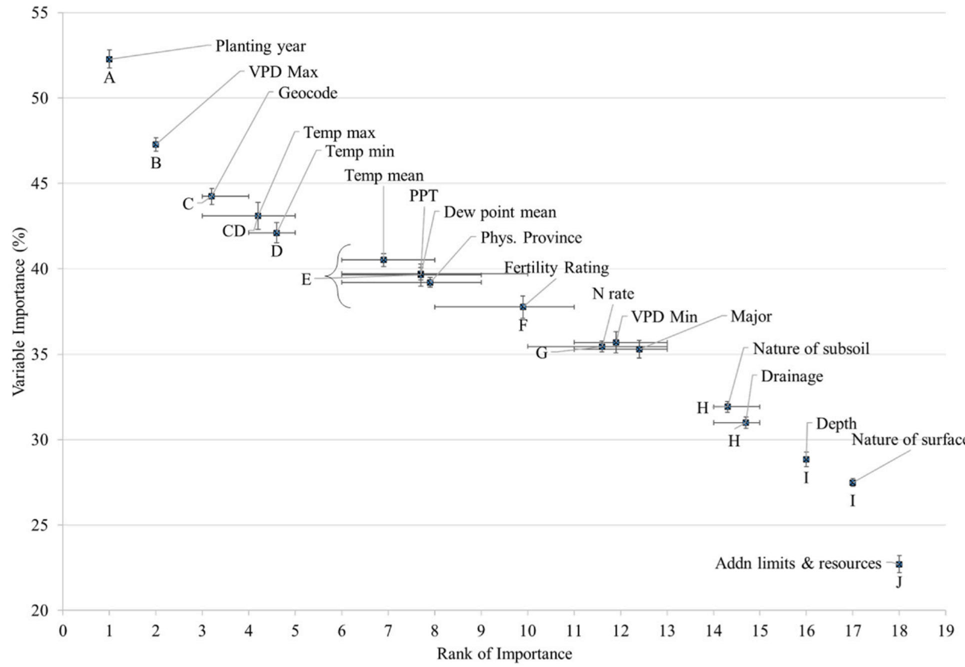

Fig. 4. Random forest variable importance ranks show the mean percent importance (y axis) and mean rank (x axis). Vertical error bars represent the 95% confidence interval for variable importance and horizontal error represent range of rank values across the 10 model runs. Letters separating variable importance are significantly different at a 0.05 level.

Table 12
Summary of loblolly pine area inside and outside of native range in the southeastern US (SEUS).

	Managed Southern Pine	Total SPOT Coverage
	Hectares	
Inside Native Range	14,270,533	82,177,815
Outside Native Range	1,377,299	121,084,015
SEUS Total	15,647,832	203,261,830

original were 0.79 and 1.91 m, for the training data 0.79 and 1.90 m, and for the test data 0.76 and 2.16 m. The 10-fold cross-validation results were very stable across all 10 runs with an average and standard deviation in R^2 of 0.79 ($\pm$ 0.0056) and an average and standard deviation for RMSE of 1.38 m ($\pm$ 0.0159). When computing the maximum likelihood estimation penalization, we obtained an optimal penalty value of 0.00013, showing a low need for penalization. Additionally, all the recommendations about sample size (Riley et al., 2019) were satisfied: global shrinkage factor of 0.9803; absolute difference between R^2 and R^2-adjusted of 0.014, the model residual standard deviation of 6.24%, and mean predicted outcome (model intercept) of 1.05%. Additionally, the RRMSE of the model was 6.81%. We used B2WekoGgPD as the "base" level and found the following spot variables significantly different: Major codes A, C, D, E; Drainage S and V; Depth to increase in clay 1, 4, and 0; Nature of surface r; Nature of subsoil a, I, p, and x; Additional limitations and resources q, u, and v; Geocodes Al, Am, Au, Av, Bb, Cs, Fl, Lb, Le, Lo, Um, Wx; and Physiographic province AF, BP, GF, SC, and WG.

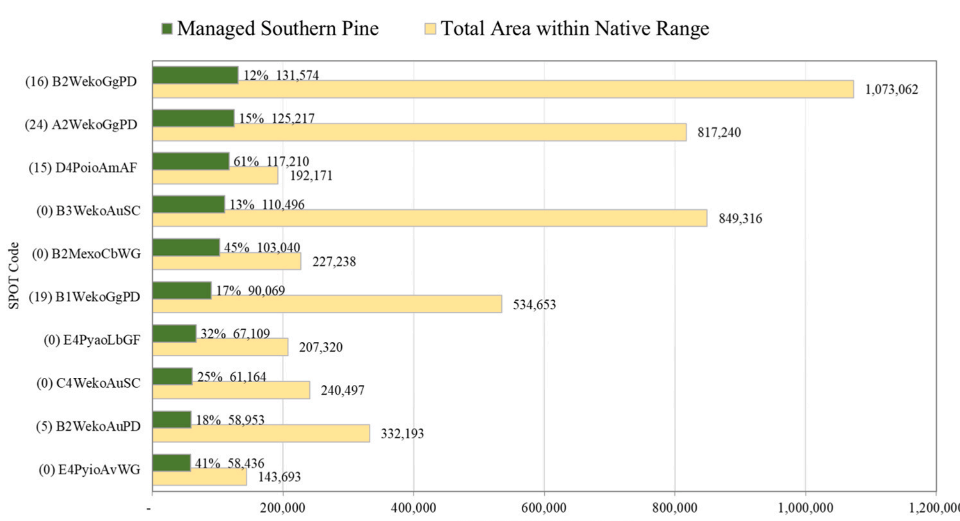

Fig. 5. Summary of ten most representative soils in managed southern pine (green bars) and total area (yellow bars). Numbers within green bars indicate the area in managed southern pine derived from Thomas et al. (2021), numbers in yellow bars indicate total SPOT code area in the native range. Numbers in parentheses before SPOT codes indicate number of Regionwide trial observations per code.

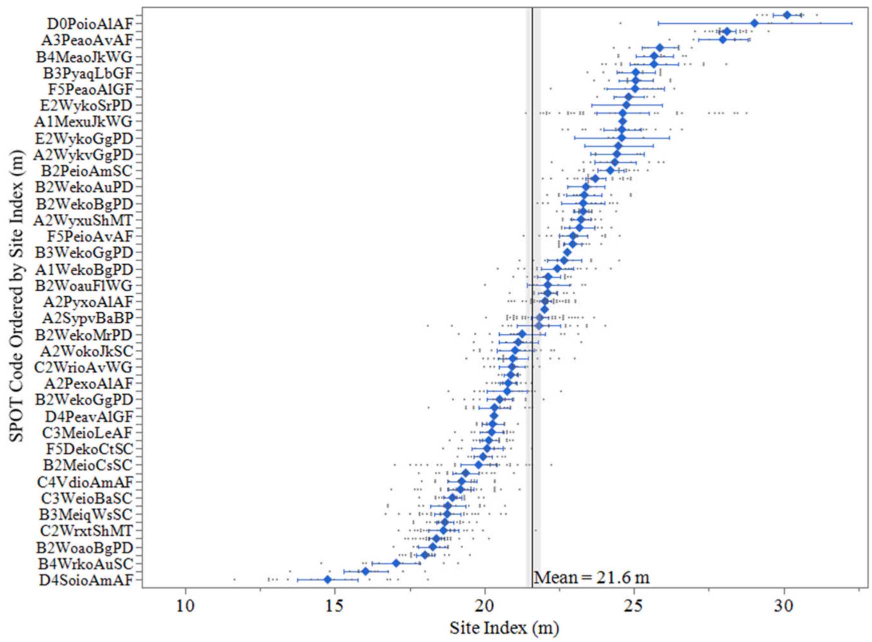

Fig. 6. Site index values in meters with at least 10 observations for unique SPOT codes in the Regionwide data. Blue diamonds are mean values and error bars are 95% confidence intervals. Small dots are individual observations (plot-level). The vertical black line is the overall mean site index (21.6 m) with shaded gray 95% confidence interval [21.4, 21.9].

Table 13
Summary of candidate models from an ordinary least squares regression 10-fold cross-validation outcomes predicting site index with penalization when needed. SPOT factors include: major group, depth, and drainage, nature of surface, nature of subsurface, and additional resource and limitations, geocode, and physiographic province. Climate includes: precipitation (mm), minimum, maximum, and mean temperature (°C), minimum and maximum vapor pressure deficit (kPa), and mean dew point temperature (°C). N rate indicates nitrogen rate (kg ha^{-1}).

Ordinary Least Squares Model	R^2	RMSE (m)
N Rate	0.11	2.84
Climate	0.12	2.84
SPOT	0.79	1.38
SPOT + Planting year	0.81	1.30
SPOT + Climate	0.79	1.40
SPOT + Planting year + N rate	0.82	1.25
SPOT + Planting year + N rate + Climate	0.87	1.10

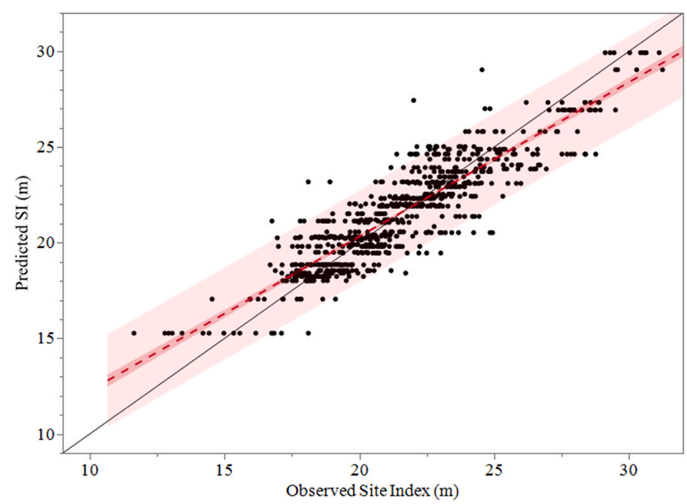

Fig. 7. Predicted *vs.* observed site index (m) base age 25 yr values for the SPOT code predictive model. The black line is the one-to-one line and the red dotted line is the model. Dark red shading represents 95% confidence interval. Light red shading represents the prediction interval.

5. Discussion

In this study, we were able to identify and summarize the relative importance of soils (by SPOT code) under pine plantation management in the southeastern region. It is the first time that such a comprehensive analysis has been conducted within the native range of loblolly pine but also throughout all southeastern states. We found that most of the area identified as managed pine plantation (91%) was located within the native range of loblolly pine.

The Forest Productivity Cooperative SPOT system grew out of decades of research and experience by soil mappers, academic researchers, and forest managers and expanded previously developed soil classification systems. It will continue to evolve as new information becomes available. It is important to emphasize that because the SPOT system uses underlying soil boundaries and information from NRCS SSURGO at a map unit level, and intersects polygons with USGS geology, coastal plain terraces, and physiographic provinces, many soil series are divided and some soil series are functionally grouped. These differences become useful for understanding variation in site productivity that soil series do not capture.

The most important variable in the random forest model was planting year. Site index has increased since the 1970 s due to a variety of management factors, such as genetic improvement (McKeand et al., 2021), improved site preparation (both mechanical and chemical), fertilization (Fox et al., 2007b), and environmental factors, such as temperature, deposition of nitrogen, and improvements in soil physical properties after reforestation of eroded agricultural lands (Albaugh et al., 2022; Davis et al., 2022). Atmospheric CO_2 has increased by 75 ppm from 1970 to 2015 and may be responsible for a 27% increase in wood volume in planted loblolly/shortleaf pine at age 25 in the Southeast (Davis et al., 2022). An important question for the future will be: how long can this continued improvement of site productivity be expected to last? Results from a regional modeling effort suggested increasing temperature and CO_2 may improve pine productivity in the future, unless rainfall decreases or nutrient limitations are not addressed (Thomas et al., 2017). Fortunately, silvicultural practices, such as fertilization, can override the negative effects of moderate drought conditions (Bracho et al., 2018). Vapor pressure deficit, our second most important model variable, has increased over the decades with increasing air temperature and decreasing relative humidity and is expected to continue increasing, which may negatively affect potential future productivity by reducing stomatal conductance (Ficklin and Novick, 2017). Overall, it is currently difficult to predict to what extent site index will continue to increase in the future due to environment and/or management, but there is still substantial room for improvement as the expected physiological maximum (32 m) is still far from the current average productivity. For this reason it is critical to understand the soil-site characteristics that limit growth.

Parent material (geocode) is the most important soil-site factor to influence forest productivity as it drives many soil physical and chemical characteristics (*e.g.*, Littke et al., 2011; Moore et al., 2022). Hennigar et al. (2017) found in the Northeast that geology and temperature explained 65% of total biomass growth variation. Lower coastal plain terraces, which have younger marine deposits are typically more productive than older more highly weathered terraces (Everett and Thorp, 2008). Deficiencies in K manifest in the Pleistocene-Pliocene terraces (Carlson et al., 2014). However, there may be some differences in the relative fertility of rock types across regions, particularly due to age and weathering. For example, in the Northeast, granite is associated with the lowest site productivity, because it weathers slower, is low in base cations, occurs at higher elevations, and tends to be sandy and rocky (Hennigar et al., 2017). Whereas, in the southeast region, granitic soils provide K, are more highly weathered, and have large soil volumes that can be important for site productivity.

Physiographic province ranked highly, before soil characteristics, likely because each province is grouped by similar soils, elevation, topography, climate, and land use. Therefore, much of the commonalities within regions are captured by this variable. This grouping likely explains the overlap of physiographic province with several climate variables.

Interestingly, N rate was of greater importance than soil variables derived from NRCS (major group, depth, drainage, and modifiers). In preliminary analyses, we found that when only lower N rates were included, the N rate variable became less important than soil variables. When higher N rates were included it moved up in the variable importance ranks. These results suggest that inherent site productivity due to soil characteristics can be superseded with sufficient nutrient inputs. Nitrogen rate has been found in many studies to be an important variable in site productivity (*e.g.*, Albaugh et al., 2021) and there are interactions such that some soils respond more than others (Albaugh et al., 2015).

Major group, which is defined primarily by profile texture, was the most important of all the soil factors. Soil texture is important because it relates to a variety of attributes critical to productivity such as nutrient and water holding capacity. Many studies have found texture to be a critical soil variable in forest productivity (*e.g.*, Fisher and Garbett, 1980; Hacker and Bilan, 1992; Subedi and Fox, 2016). While the order of variable importance from the random forest model may well describe our particular Regionwide dataset, we expect that the relative importance of each variable will likely change with expanded observations. For example, our trial location selection, very purposely, avoided soils with shallow restrictive layers, as the objectives were generally to evaluate fertilizer response. Hence we have no soils with root limitations at $<$ 25 cm (Additional limitations group "s"). Therefore, while our trial network is extensive, a wider dataset would capture more operational variability across the landscape.

The prediction models we tested, though limited by linear relationships and lack of interactions, provides the advantage to estimate site productivity of soils where we do not have empirical data. The SPOT-based model provided a reasonably parsimonious and satisfactory explanatory model that can be extended across the southeastern US, either with or without climate data. With the mean and range in variation for mapped soils, forest managers can make better-informed decisions regarding silvicultural management. The SPOT codes can help indicate the limitations or opportunities for a given site.

Whether for mitigating climate change through improved carbon sequestration and storage, or for economic incentives, understanding the drivers of site productivity is critically important to optimize forest management. Increasing site productivity across the region means targeting sites that have the capacity to improve and, of course, knowing how to improve them. This system allows us to provide observed and attainable site productivity values, and to understand what soil resources may be limiting to help guide management decisions. Granted, past land use practices can either improve site productivity above and beyond expectations (*e.g.*, from legacy fertilizer, Everett and Palm-Leis, 2009) or decrease potential such as from soil compaction (Aust and Blinn, 2004) or poor competition control (Albaugh et al., 2012). Silvicultural inputs are generally limited by costs and financial returns. Therefore, understanding which inputs can be optimized on a site-specific basis is critical to achieve landowner objectives. To our knowledge, this study is the most comprehensive effort to map managed loblolly pine site index across the entire native range.

5.1. Limitations and future directions

Forest soils were often mapped by NRCS to a much lower resolution than agricultural soils, often representing forest soils in complexes, associations, or undifferentiated map units which can add greater uncertainty when predicting site productivity or growth responses to particular management practices. Future work could take advantage of LiDAR digital elevation models to extract minor changes in landforms within map units to identify inclusions, such as depressions, that should managed differently. Additionally, repeat LiDAR flights in the future could provide for observed site productivity data at much greater resolution. The relative differences in the scale of maps certainly has led to some level of inaccuracy and caution should be taken when applying information at a small scale. Future iterations should also investigate the addition of landforms, such as water shedding *vs* water collecting features. Additionally, geologic maps often list three rock types but without percent contribution or information regarding depth, leaving some uncertainty as to the relative importance of each lithological unit.

Managed pine is grown across an incredible diversity of soils and geologies in the southeastern US. While our Regionwide trial network is likely one of the most extensive of its kind in the nation, it covered only 72 out of thousands of unique SPOT codes. It would be impossible to have field trials to assess fertilization and/or other management responses on every unique soil code, therefore we must extrapolate from our observations. Future steps will be to acquire site productivity datasets across a broader gradient of soils under intensively managed loblolly pine.

While this study focused primarily on managed loblolly pine productivity, the system can be used also for identification of slash, longleaf, or sand pine appropriate sites. While outside of the scope of this study, there may be future opportunities to use SPOT codes for other forest systems. Additionally, while SPOT was developed on the southeastern US, there may be opportunities to expand the system to other regions that also grow pine plantations, such as Brazil, or to other regions of the US. In these cases, the system would need to be adapted to recognize soil characteristics important for productivity and management that do not occur in the southeastern US.

Finally, previous land-use practices, particularly P fertilizer applications, can affect available P, even across rotations (Everett and Palm-Leis, 2009). Soil chemical properties, particularly soil extractable P, should be assessed at planting to properly guide establishment P decisions.

6. Conclusion

The SPOT system provides a comprehensive framework for improving pine management. While previous soil systems focused on specific regions in the southeastern US, this is the first system to span the entire range of loblolly pine and include soils, geology, and physiographic province. The explanatory and predictive capabilities for site productivity will provide critical information for sustainable management of site resources, whether for timber, carbon, or other objectives. Additionally, these data provide a baseline to observe sustained productivity and sustainable site resource management. Modifications or evolutions may occur as new information and future studies continue to inform the SPOT system.

CRediT authorship contribution statement

Albaugh Timothy J.: Writing – review & editing, Investigation, Formal analysis, Data curation. **Kleto Pete:** Writing – review & editing, Conceptualization. **Fox Thomas R.:** Writing – review & editing, Conceptualization. **O'Brien Ed:** Writing – review & editing, Conceptualization. **Cook Rachel:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Carter David R.:** Writing – review & editing, Funding acquisition. **Ricker Matthew:** Writing – review & editing. **Campoe Otavio:** Writing – review & editing, Funding acquisition. **Rubilar Rafael:** Writing – review & editing, Funding acquisition. **Cohrs Chris:** Writing – original draft, Visualization, Software, Methodology, Conceptualization. **McEachern Kirk:** Conceptualization. **Allen Howard Lee:** Writing – review & editing, Conceptualization. **Trlica Andrew:** Writing – review & editing, Validation, Methodology, Formal analysis, Conceptualization. **Ribas-Costa Vicent:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rachel Cook reports financial support was provided by National Science Foundation. Chris Cohrs reports financial support was provided by US Department of Agriculture. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgments

We would like to acknowledge the tremendous amount of effort of forest soil experts that contributed to the creation of the Forest Productivity Cooperative Site Productivity Optimization for Trees System. In addition, many contributed a significant amount of time to commenting on and revisions including: Connor Fristoe, Lewis Williams, John Torbert, Jim Gent, Jim Rakestraw, John Thorp, and Jim DeWitt. Students and researchers contributed to the coding and creation of the database including Rob Austin, Emily Stoll, Jacob Dearing, Ellis Ackerman, Morgan Schulte, and Zach Grover. The project would not have been possible without the dedication of Chris Cohrs from his undergraduate degree to PhD.. We appreciate the support for over 50 years from the Forest Productivity Cooperative and gratefully acknowledge support provided by the Department of Forestry and Environmental Resources at NC State University, the Department of Forest Resources and Environmental Conservation at Virginia Tech, the Departamento de Silvicultura, Facultad de Ciencias Forestales, Universidad de Concepcion, and the Universidade Federal de Lavras. Funding for this work was provided in part by the McIntire-Stennis Program of the National Institute of Food and Agriculture, U.S. Department of Agriculture Accession No. 1025518, and the National Science Foundation Center for Advanced Forestry Systems Award 1916552. Vicent Ribas-Costa was funded by the Doctoral INPhINIT Retaining fellowship n. º LCF/BQ/DR22/11950023 from 'la Caixa' Foundation (ID 100010434).

References

Albaugh, T., Fox, T., Allen, H., Rubilar, R., 2015. Juvenile southern pine response to fertilization is influenced by soil drainage and texture. Forests 6, 2799–2819. https://doi.org/10.3390/f6082799.

Albaugh, T.J., Stape, J.L., Fox, T.R., Rubilar, R.A., Allen, H.L., 2012. Midrotation vegetation control and fertilization response in *Pinus taeda* and *Pinus elliottii* across the southeastern United States. South. J. Appl. For. 36, 44–53. https://doi.org/10.5849/sjaf.10-042.

Albaugh, T.J., Fox, T.R., Rubilar, R.A., Cook, R.L., Amateis, R.L., Burkhart, H.E., 2017. Post-thinning density and fertilization affect *Pinus taeda* stand and individual tree growth. For. Ecol. Manag. 396, 207–216. https://doi.org/10.1016/j.foreco.2017.04.030.

Albaugh, T.J., Albaugh, J.M., Carter, D.R., Cook, R.L., Cohrs, C.W., Rubilar, R.A., Campoe, O.C., 2021. Duration of response to nitrogen and phosphorus applications in mid-rotation *Pinus taeda*. For. Ecol. Manag. 498, 119578 https://doi.org/10.1016/j.foreco.2021.119578.

Albaugh, T.J., Carter, D.R., Cook, R.L., Campoe, O.C., Rubilar, R.A., Creighton, J.L., 2022. Improving *Pinus taeda* site index from rotation to rotation with silvicultural treatments. For. Ecol. Manag. 526, 120581 https://doi.org/10.1016/j.foreco.2022.120581.

Antoniadis, A., Lambert-Lacroix, S., Poggi, J.-M., 2021. Random forests for global sensitivity analysis: a selective review. Reliab. Eng. Syst. Saf. 206, 107312 https://doi.org/10.1016/j.ress.2020.107312.

Aust, W.M., Blinn, C.R., 2004. Forestry best management practices for timber harvesting and site preparation in the eastern United States: an overview of water quality and productivity research during the past 20 years (1982–2002). Water Air Soil Pollut. Focus 4, 5–36. https://doi.org/10.1023/B:WAFO.0000012828.33069.f6.

Bockheim, J.G., Hartemink, A.E., 2013. Classification and distribution of soils with lamellae in the USA. Geoderma 206, 92–100. https://doi.org/10.1016/j.geoderma.2013.04.014.

Bracho, R., Vogel, J.G., Will, R.E., Noormets, A., Samuelson, L.J., Jokela, E.J., Gonzalez-Benecke, C.A., Gezan, S.A., Markewitz, D., Seiler, J.R., Strahm, B.D., Teskey, R.O., Fox, T.R., Kane, M.B., Laviner, M.A., McElligot, K.M., Yang, J., Lin, W., Meek, C.R., Cucinella, J., Akers, M.K., Martin, T.A., 2018. Carbon accumulation in loblolly pine plantations is increased by fertilization across a soil moisture availability gradient. For. Ecol. Manag. 424, 39–52. https://doi.org/10.1016/j.foreco.2018.04.029.

Carlson, C.A., Fox, T.R., Allen, H.L., Albaugh, T.J., Rubilar, R.A., Stape, J.L., 2014. Growth responses of loblolly pine in the southeast United States to midrotation applications of nitrogen, phosphorus, potassium, and micronutrients. For. Sci. 60, 157–169. https://doi.org/10.5849/forsci.12-158.

Carter, M.C., Kellison, R.C., Wallinger, R.S., 2015. Forestry in the US South: A history. LSU Press.

Cheng, L., De Vos, J., Zhao, P., Yang, M., Witlox, F., 2020. Examining non-linear built environment effects on elderly's walking: A random forest approach. Transp. Res. Part Transp. Environ. 88, 102552 https://doi.org/10.1016/j.trd.2020.102552.

Coile, T.S., 1952. Soil and the Growth of Forests. in: Advances in Agronomy. Elsevier, pp. 329–398. https://doi.org/10.1016/S0065-2113(08)60313-2.

Colpitts, M.C., Fahmy, S.H., MacDougall, J.E., Ng, T.T.M., Mclnnis, B.G., Zelazny, V.F., 1995. Forest Soils of New Brunswick.

Cooke, C.W., 1931. Seven coastal terraces in the southeastern states. J. Wash. Acad. Sci. 21, 503–513.

Davis, E.C., Sohngen, B., Lewis, D.J., 2022. The effect of carbon fertilization on naturally regenerated and planted US forests. Nat. Commun. 13, 5490 https://doi.org/10.1038/s41467-022-33196-x.

Diéguez-Aranda, U., Burkhart, H.E., Amateis, R.L., 2006. Dynamic site model for loblolly pine (*Pinus taeda* L.) plantations in the United States. For. Sci. 52, 262–272.

Everett, C.J., Palm-Leis, H., 2009. Availability of residual phosphorus fertilizer for loblolly pine. For. Ecol. Manag. 258, 2207–2213. https://doi.org/10.1016/j.foreco.2008.11.029.

Everett, C.J., Thorp, J.H., 2008. Site quality evaluation of loblolly pine on the South Carolina Lower Coastal Plain, USA. J. For. Res. 19, 187–192. https://doi.org/10.1007/s11676-008-0033-4.

Ficklin, D.L., Novick, K.A., 2017. Historic and projected changes in vapor pressure deficit suggest a continental-scale drying of the United States atmosphere. J. Geophys. Res. Atmosph. 122, 2061–2079. https://doi.org/10.1002/2016JD025855.

Fisher, R.F., Garbett, W.S., 1980. Response of semimature slash and loblolly pine plantations to fertilization with nitrogen and phosphorus. Soil Sci. Soc. Am. J. 44, 850–854. https://doi.org/10.2136/sssaj1980.03615995004400040039x.

Fox, T.R., Jokela, E.J., Allen, H.L., 2007a. The development of pine plantation silviculture in the southern United States. J. For. 105, 337–347. https://doi.org/10.1093/jof/105.7.337.

Fox, T.R., Lee Allen, H., Albaugh, T.J., Rubilar, R., Carlson, C.A., 2007b. Tree nutrition and forest fertilization of pine plantations in the southern United States. South. J. Appl. For. 31, 5–11. https://doi.org/10.1093/sjaf/31.1.5.

Hacker, W.D., Bilan, M.V., 1992. Site factors affecting growth of loblolly pine in the post oak belt. South. J. Appl. For. 16, 197–200. https://doi.org/10.1093/sjaf/16.4.197.

Harrell, F., 2023. rms: Regression Modeling Strategies.

Hennigar, C., Weiskittel, A., Allen, H.L., MacLean, D.A., 2017. Development and evaluation of a biomass increment based index for site productivity. Can. J. For. Res. 47, 400–410. https://doi.org/10.1139/cjfr-2016-0330.

Horton, J.D., San Juan, C.A., Stoeser, D.B., 2017. The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (USGS Numbered Series No. 1052), The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States, Data Series. U.S. Geological Survey, Reston, VA. https://doi.org/10.3133/ds1052.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.-M., Gerber, J.S., Reddy, V.R., Kim, S.-H., 2016. Random forests for global and regional crop yield predictions. PLOS ONE 11, e0156571. https://doi.org/10.1371/journal.pone.0156571.

Johnston, C.M.T., Guo, J., Prestemon, J.P., 2022. RPA forest products market data for U. S. RPA Regions and the world, historical (1990–2015), and projected (2020–2070) using the Forest Resource Outlook Model (FOROM). https://doi.org/10.2737/RDS-2022–0073.

Littke, K.M., Harrison, R.B., Briggs, D.G., Grider, A.R., 2011. Understanding soil nutrients and characteristics in the Pacific Northwest through parent material origin and soil nutrient regimes. Can. J. For. Res. 41, 2001–2008. https://doi.org/10.1139/x11-115.

Little, E.L., 1971. Atlas of United States trees. Volume 1. Conifers and important hardwoods. U.S. Dept. of Agriculture, Forest Service, Washington D.C.

Loecher, M., 2022. Unbiased variable importance for random forests. Commun. Stat. - Theory Methods 51, 1413–1425. https://doi.org/10.1080/03610926.2020.1764042.

McKeand, S.E., Payn, K.G., Heine, A.J., Abt, R.C., 2021. Economic significance of continued improvement of loblolly pine genetics and its efficient deployment to landowners in the southern United States. J. For. 119, 62–72. https://doi.org/10.1093/jofore/fvaa044.

Moore, J.A., Kimsey, M.J., Garrison-Johnston, M., Shaw, T.M., Mika, P., Poolakkal, J., 2022. Geologic soil parent material influence on forest surface soil chemical characteristics in the Inland Northwest, USA. Forests 13, 1363. https://doi.org/10.3390/f13091363.

Morris, L.A., Campbell, R.G., 1991. Soil and site potential. In: Duryea, M.L., Dougherty, P.M. (Eds.), Forestry Sciences, Forest Regeneration Manual. Springer Netherlands, Dordrecht, pp. 183–206. https://doi.org/10.1007/978-94-011-3800-0_10.

Oswalt, S.N., Smith, W.B., Miles, P.D., Pugh, S.A., 2019. Forest resources of the United States, 2017: a technical document supporting the Forest Service 2020 RPA Assessment. (No. WO-GTR-97). U. S. Dep. Agric., For. Serv., Wash., DC. https://doi.org/10.2737/WO-GTR-97.

Pahlavan-Rad, M.R., Dahmardeh, K., Hadizadeh, M., Keykha, G., Mohammadnia, N., Gangali, M., Keikha, M., Davatgar, N., Brungard, C., 2020. Prediction of soil water infiltration using multiple linear regression and random forest in a dry flood plain, eastern Iran. CATENA 194, 104715. https://doi.org/10.1016/j.catena.2020.104715.

Prestemon, J.P., Abt, R.C., 2002. Timber products supply and demand. South. For. Resour. Assess. 53, 299–325.

Probst, P., Wright, M.N., Boulesteix, A.-L., 2019. Hyperparameters and tuning strategies for random forest. WIREs Data Min. Knowl. Discov. 9, e1301 https://doi.org/10.1002/widm.1301.

Riley, R.D., Snell, K.I.E., Ensor, J., Burke, D.L., Harrell Jr, F.E., Moons, K.G.M., Collins, G. S., 2019. Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. Stat. Med. 38, 1262–1275. https://doi.org/10.1002/sim.7993.

Roth, F., 1916. Concerning site. J. For. 14, 3–12. https://doi.org/10.1093/jof/14.1.3.

Soil Science Division Staff, 2017. Soil survey manual. C.Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Soil Survey Staff, N.R.C.S., 2021. United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for southeastern US. Available online.

Soil Survey Staff, 2022. Keys to Soil Taxonomy, 13th edition. USDA Natural Resources Conservation Service.

Subedi, S., Fox, T.R., 2016. Predicting loblolly pine site index from soil properties using partial least-squares regression. For. Sci. 62, 449–456. https://doi.org/10.5849/forsci.15-127.

Tesch, S.D., 1980. The evolution of forest yield determination and site classification. For. Ecol. Manag. 3, 169–182. https://doi.org/10.1016/0378-1127(80)90014-6.

Thomas, R.Q., Brooks, E.B., Jersild, A.L., Ward, E.J., Wynne, R.H., Albaugh, T.J., Dinon-Aldridge, H., Burkhart, H.E., Domec, J.-C., Fox, T.R., Gonzalez-Benecke, C.A., Martin, T.A., Noormets, A., Sampson, D.A., Teskey, R.O., 2017. Leveraging 35 years of *Pinus taeda* research in the southeastern US to constrain forest carbon cycle predictions: regional data assimilation using ecosystem experiments. Biogeosciences 14, 3525–3547. https://doi.org/10.5194/bg-14-3525-2017.

Thomas, V.A., Wynne, R.H., Kauffman, J., McCurdy, W., Brooks, E.B., Thomas, R.Q., Rakestraw, J., 2021. Mapping thins to identify active forest management in southern pine plantations using Landsat time series stacks. Remote Sens. Environ. 252, 112127 https://doi.org/10.1016/j.rse.2020.112127.

Trim, K.R., Coble, D.W., Weng, Y., Stovall, J.P., Hung, I.-K., 2020. A new site index model for intensively managed loblolly pine *Pinus taeda* plantations in the West Gulf coastal plain. For. Sci. 66, 2–13. https://doi.org/10.1093/forsci/fxz050.

2020 U.S. Geological Survey, 2020. 3D Elevation Program 1-Meter Resolution Digital Elevation Model.

United States Department of Agriculture, N.R.C.Service., 2022. Land resource regions and major land resource areas of the United States, the Caribbean, and the Pacific Basin. U.S. Department of Agriculture, Agriculture Handbook 296.

Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77. https://doi.org/10.18637/jss.v077.i01.