GCB Bioenergy

WILEY

Global Change Biology Bioenergy
BIOPRODUCTS FOR A SUSTAINABLE BIOECONOMY

RESEARCH ARTICLE OPEN ACCESS

A Global Short Rotation Coppice (SRC) Willow Dataset for the Bioeconomy: Implications for the Yield in the United Kingdom

Antonio Castellano Albors[1] | Anita Shepherd[1] | Ian Shield[2] | William Macalpine[2] | Kevin Lindegaard[3] | Ian Tubby[4] | Astley Hastings[1]

[1]School of Biological Sciences, University of Aberdeen, Aberdeen, UK | [2]Rothamsted Research, Harpenden, UK | [3]Crops for Energy, Bristol, UK | [4]Forestry Commission, Bristol, UK

Correspondence: Antonio Castellano Albors (a.castellanoalbors.22@abdn.ac.uk)

Received: 12 March 2025 | **Revised:** 16 June 2025 | **Accepted:** 1 July 2025

Funding: This work was supported by the UK Research and Innovation (BB/V0115553/1 and BB/V011588/1), UK Energy Research Centre (EP/S000771/1).

Keywords: best practice management | genetic diversity | genotype by environment (GxE) | Random forest | short rotation coppice (SRC) | variable importance | willow | yield

ABSTRACT

Short rotation coppice (SRC) willow is a second-generation lignocellulosic energy crop with a background of research and breeding programmes carried out globally for more than three decades. While commercial standards include planting in mixtures of 6–8 willow genotypes of genetic diversity, much research to date has focused on monoculture trials. Research has found significant differences in willow performance through different management methods, soil properties and environmental interactions (GxE), when applied locally. However, global analysis of these interactions remains a challenge. We present a global SRC willow dataset to facilitate researchers and growers with a resource not available to date to help in closing the gap between research and industry. Data has been collected through literature review and personal communications with key researchers on willow in the United Kingdom. Global annual average yield is 9 Mg Dry Matter (DM) ha^{-1} $year^{-1}$ with 17 genotypes, including two types of mixtures, above the economic threshold of 10 Mg DM ha^{-1} $year^{-1}$. Canada and the United States are the best and worst performers with 10.6 and 6.7 Mg DM hr^{-1} $year^{-1}$, respectively. We expect this dataset to provide an efficient way of estimating yields at a smaller scale by multiple combinations of GxE interactions. Biomass production from 1-year-old stems in the first harvest cycle is significantly lower than for the second and third year of the first harvest cycle (ANOVA, $p < 0.001$). Harvest cycles of 2 and 3 years did show significant but small differences in final yield ($t = 3.87$, $p < 0.001$). A random forest statistical procedure was applied to test for the association of the predictor variables with biomass production. The model explained up to 63.65% of the variance observed in yield for all genotypes and sites, with genetic diversity among the most important variables.

1 | Introduction

Globally, bioenergy accounts for 55% of renewable energy (e.g., transport fuels, biogas, heat and power generation, etc.) and 6% of the total energy supply (IEA 2023).

The potential benefit of biomass production as a renewable source of energy to alleviate the burden of fossil fuels in the power system has been long praised by multiple organisations such as the International Energy Agency (IEA), Intergovernmental Panel on Climate Change (IPCC) or the Climate Change Committee

(CCC) (Smith and Porter 2018; CCC 2020; IEA Bioenergy 2020) in the United Kingdom. However, newest reports call for caution on previous estimates, and more strict regulations have been put in place in order to make the production of energy from biomass sustainable (Lecocq et al. 2022; IEA 2023; DESNZ 2023; Clifton-Brown et al. 2023).

Current concerns about energy crops comprise environmental issues regarding soil health, biodiversity, water quality and invasive species (Powlson et al. 2005; Haughton et al. 2009; Schmidt-Walter 2020; Zitzmann and Rode 2021), competition for the same agricultural land with food versus energy production (Karp 2011; Karp and Richter 2011; Don et al. 2011; Valentine et al. 2011). Previous land use is another important factor when considering the benefits of bioenergy crops, along with the potential impacts from the use of fertilisers and herbicides (McCalmont et al. 2015; Milner et al. 2015). Perennial bioenergy crops are better suited for overcoming this challenge than annual crops due to their ability to grow in marginal lands and provide wider environmental benefits (Karp and Shield 2008; Clifton-Brown et al. 2018).

Willows (*Salix* spp.) are one of the second-generation lignocellulosic energy crops aimed at contributing towards achieving net zero emissions by 2050 in the United Kingdom (Karp 2011; Hastings et al. 2014; Clifton-Brown et al. 2018; Albanito et al. 2019). Along with the genus *Populus*, genus *Salix* comprises the family *Salicaceae*. Lack of agreement regarding the taxonomy of willows still prevails (Shield et al. 2014; Dickmann and Kuzovkina 2014). Their origin is considered to be Eastern Asia (Karp 2011), but they are mainly distributed in the northern hemisphere in temperate climates (Dickmann and Kuzovkina 2014). Shrub willows (subgenus *Vetrix*) and tree willows (subgenus *Salix*) are the best suited and used for bioenergy production. Willows are catkin-bearing perennial crops grown as short rotation coppice (SRC), characterised by a fast growth rate after coppicing (Shield et al. 2014).

Phenology stages of willow grown as SRC in a full growing season would include budburst (late winter/early spring), leaf elongation, stem growth, leaf senescence leading to stem growth stop and leaf fall in the autumn, and dormancy during the winter until the next season.

Research on willow as a bioenergy crop dates back to early programmes in the 1970s in Sweden (Gullberg 1993; Volk et al. 2006) and the 1980s in United Kingdom (ETSU 1998) and United States (Volk et al. 2006), anticipated by uncertainty in oil prices and energy distribution (Karp 2014; Karp and Richter 2011). Despite this effort, willow is lacking a commercial and industrial scale that allows it to be economically viable for growers in the United States, United Kingdom and Poland (Volk et al. 2016; Macalpine 2019; Ziety et al. 2022). Research has continued in the form of funded programmes (Evans et al. 2007; Valentine et al. 2009; Volk et al. 2016; Clifton-Brown et al. 2018, 2023) that determine the potential yields of willow along with the associated environmental, social and economic impacts through the supply chain (Volk et al. 2006; Hastings et al. 2014; Tallis et al. 2013; Aylott et al. 2008).

This extensive research in the United Kingdom enabled knowledge in the agronomy of willow and the most influential factors on its productivity (Tubby and Armstrong 2002; Evans et al. 2007; Valentine et al. 2009; Caslin et al. 2023).

Best practice guidance in the United Kingdom (Caslin et al. 2023) comprises the suitability of the terrain alongside climate variables more suitable for crop growth (e.g., rainfall, slope, altitude, temperature and land shape). Ground preparation is an important factor when considering willow as a bioenergy crop, as commercial plantations have a life span of 20+ years (Shield et al. 2014; Powlson et al. 2005; Nordh 2005). This usually takes place the year before planting is due. Issues such as soil compaction, herbicide application, soil testing and pest control must be considered to attain optimum conditions for crop establishment. Planting material usually comes in the form of approximately 20 cm long dormant cuttings from 1-year-old plants and is stored between −4°C and −2 (Welc et al. 2017, 2018; Volk et al. 2004). Planting takes place during early spring to allow the plant to take advantage of longer days and higher radiation exposure in order to compete with weeds until May/June. Herbicide is usually applied post-planting to prevent weed growth. Fencing might be needed if herbivores are a potential issue. Cutback (coppicing) after the first growing season remains best practice to induce vigorous regrowth and allow for more herbicide application if needed. According to best practice, the plant is then left to grow for three (Caslin et al. 2023) more seasons when it is harvested again in the winter months once leaf drop has occurred. Fertilisation is advised only if needed (marginal land) or when soil analysis confirms so. Willow should be planted in mixtures of 6–8 genotypes (or clones—called variety if it was marketed) to increase their resilience against Leaf Rust (*Melampsora* spp.) among other pathogens (McCracken and Dawson 2003; Begley et al. 2009). Genotype selection for the mixture and planting density can also influence final yields across different harvest cycles along with water availability (Bullard et al. 2002; Stolarski et al. 2019). Harvesting, drying and storage methods are also compiled in best practice guidance along with site restoration to grassland or arable.

Extensive research confirms the importance that the establishment period has on later development and performance of willow (Welc et al. 2017; Verwijst et al. 2012; Larsen et al. 2014) and represents 20%–25% of total costs (Volk et al. 2016). Field trials with different planting methods (vertical vs. horizontal), materials (cuttings, rods and billets) and lengths have been carried out to gain knowledge about establishment methods and potential yields (Bergante et al. 2016; Larsen et al. 2016; Edelfeldt et al. 2015; Cao et al. 2011; Vigl and Rewald 2014; McCracken et al. 2010; Welc et al. 2018). Bush et al. (2015) studied the establishment constraints in a commercial context in the United States by comparing the performance of various planting machines.

Weed competition is a major problem if not treated appropriately. Survival rates and poor performance, along with upfront costs, could make the whole commercial operation economically unviable (Valentine et al. 2009; Jureková et al. 2011; Albertsson et al. 2014). Weeds compete with willow in the establishment years for water and nutrients and can halt their capacity to survive and grow. Albertsson et al. (2014) and Schlepphorst et al. (2017) found survival rates and biomass production critically impacted by this competition. Cutback after the first year (establishment year) is a common practice, widespread globally (Cunniff et al. 2015; McKenzie 2011; Fabio 2017; Caslin et al. 2023). However, some studies (and now guidance) suggest that coppicing at the end of the first year should be considered with precaution and only if the plant shows vigorous growth in early establishment (Caslin et al. 2023; Albertsson et al. 2014). McKenzie (2011) compared the yield of two identical experiments, but with one of them not following the standard practice of cutback. Results show that no cutback increased the yield in the first harvest cycle (3 years) due in part to cutback plants being unable to compete in the same conditions because of the harder climate in Orkney, Scotland.

Breeding programmes since the 1980–1990s have been aimed at improving yields and disease resilience (Lindegaard et al. 2001, 2011; Macalpine et al. 2011; Macalpine et al. 2014). Genetic research in willow has brought to light promising findings to improve the genotype by environmental interactions, including phenological and physiological traits by new available molecular techniques (Clifton-Brown et al. 2018; Hallingbäck et al. 2016; Hanley and Karp 2014; Berlin et al. 2017).

In this paper, we have collected and produced a comprehensive dataset of willow phenology and physiology from literature and trial data with the objective of using this dataset to understand *Salix* phenology and physiology to develop a statistical willow yield model.

2 | Materials and Methods

2.1 | Data Collection

Yield, crop management, environmental, soil and willow's phenology data was collected from a literature search internationally and key players in willow research in the United Kingdom. Keywords 'SRC', 'willow', '*Salix*', 'biomass', 'yield' or 'harvest' were used in the Web of Science and Scopus databases. In some cases, data from published studies were complemented with data from personal communications. In addition, some papers found in the search were part of a long-term experiment, with various published papers for the same site. In this case, a specific search was conducted to gather as many harvest cycles as possible for the same site. Best crop management practice as defined by best practice guidelines (Caslin et al. 2023) was used as a criterion for inclusion in the dataset. Establishment methods such as horizontal planting and studies on soil remediation and wastewater treatment or similar were discarded. Besides location, two data variables were considered indispensable in the screening process for inclusion in the dataset: yield and planting/harvest dates to calculate the average yield per year of harvest cycle. The values of yield in the dataset are presented as dry matter (DM). Yield data that was only reported graphically was extracted using PlotDigitizer (https://plotdigitizer.com/).

This willow dataset (Castellano Albors et al. 2025) comprises soil characteristics, management methods for several genotypes, both commercial and those never released to the market. A summary of the variables included and their characteristics can be found in Table S1. It was often found that one site hosted different research trials. In these cases, a new column (id) was added where different numbers for the same location indicate a different research trial in time. A total of 3429 ID entries comprising 99 independent sites (Figure 1) and 113 genotypes from different breeding stages can be found in this dataset. Individual yield for each of the years included in a harvest cycle was preferred; however, it was not always reported. Willow's growth and yield can vary dramatically from year to year within a cycle due to impacts by drought, late frost, etc., and these events were not always reported.

Soil properties were divided into three soil layers with all associated characteristics to be as representative as possible and to avoid any soil profile averages that could cause misleading results when analysed. Terrain elevation that was not reported in the papers was extracted from Google Earth Pro (Google 2024).

FIGURE 1 | Global map showing the site trials included in SRC willow dataset.

2.2 | Calculations and Assumptions

The purpose of this dataset is to understand the explanatory variables that relate to aboveground biomass in *Salix* spp. Data from all variables were extracted directly from papers when available. Some minor calculations were performed when data were reported in different units across papers to keep consistency within the dataset (i.e., soil organic carbon (SOC) in reported units → SOC in %). Soil stocks for carbon and nitrogen were calculated on a per hectare basis to match with the yields (also in Mg ha^{-1}). The following equations were used:

$$\mathrm{SOC}\left(\mathrm{Mg\ ha^{-1}}\right) = \mathrm{SOC}\,(\%) \times \text{bulk density}\left(\mathrm{g\ cm^{-3}}\right) \times \text{soil depth (cm)}$$

$$\mathrm{SN}\left(\mathrm{Mg\ ha^{-1}}\right) = \mathrm{SN}\,(\%) \times \text{bulk density}\left(\mathrm{g\ cm^{-3}}\right) \times \text{soil depth (cm)}$$

Soil nitrogen (SN) above refers to both soil available nitrogen (SAN) and soil total nitrogen (ST). SAN is the nitrogen in the soil readily available for the crop to uptake, and STN is the nitrogen in the soil that has the potential to be available for the crop after chemical modification. When soil organic matter (SOM) was reported, SOC was calculated as a 50% of the total SOM as per Pribyl (2010). When soil texture in percentage was reported, this data was transposed to the soil textural triangle (USDA system) to know the soil type associated with it. However, this was not reciprocal due to the large variation in percentage for sand, silt and clay for some soil types (e.g., clay soils could have between 60% and 100% of clay). Bulk density, when not reported, was calculated using the Saxton calculator (Saxton et al. 1986) by adding the sand and clay percentage. Plant available water (PAW), field capacity (FC) and permanent wilting point (PWP) were calculated with the model proposed by Campbell (1985) modified by Hastings et al. (2014). Despite some of the papers reporting PAW values, the calculator was used to keep consistency throughout the dataset.

If not reported, ArcMap v. 10.8.1 (ESRI 2020) was used to calculate raster data of the aspect derived from two data elevation models. EuroDEM height data (EuroGeographics 2023) was used for sites located in Europe. The North America Elevation 1 km resolution data from the Commission for Environmental Cooperation (CEC) was used for the United States and Canada (CEC 2007).

Climatic variables such as temperature and precipitation were often reported annually in papers or for the growing season from May to September. Due to the variation in latitude and phenology between different sites, a growing season from April to October was included in the dataset. Growing season length has been considered according to phenology data (I. Shield, personal communications, July 2023) and supported by previous research (Tallis et al. 2013; Holman and Hess 2014). Climate data was extracted from the nearest weather station when available and otherwise from average monthly data provided by the CRU TS v. 4.08 dataset (Harris et al. 2020).

Following Armstrong (1997) and Tallis et al. (2013), a similar approach was considered by creating a 'climate' variable according to cumulative degree days over 5.6°C (DDcum) and soil moisture deficit (SMD) levels for the growing season (April—October). The MiscanFor model (Hastings et al. 2009) was used for the calculations, subsequently transferred to a map in ArcMap v. 10.8.1 (Esri 2020). Six climatic zones were differentiated according to Pyatt et al. (2001), with slightly modified thresholds to be adapted to the output data (Table 1). This classification is intended to provide an objective comparison between sites and takes into consideration the difference in DDcum and SMD for geographical areas in proximity due to altitude.

A new column 'group' was included to differentiate between genotypes grown in monoculture and those grown in mixtures. Within 'variety', an average of genotypes grown in monoculture has been calculated and stored as 'mix'. According to Begley et al. (2009) and McCracken et al. (2001), the yield produced by different genotypes grown in mixtures is equal or larger than the yield produced when the same genotypes are grown in monoculture. Therefore, the 'mix' values could be considered as an estimate of the potential yield of the genotypes grown in monoculture for the same trial.

European (Caslin, Finnan, and McCracken 2012) and US (Cameron et al. 2007) variety guides were used to ascertain the parentage of the genotypes included in the dataset when no information was reported in the papers or data was received from personal communications. The breeding stage was defined by following the crossings as per the diagram in Figure 2 from Caslin et al. (2023). The first stage consists of genotypes found naturally in the environment and were the first collected, or clones derived, by researchers to start willow trials (e.g., S. *dasyclados*, S. *schwerinii*, L78183, Germany, etc.). Any crosses between genotypes pertaining to stage 1 will be considered as stage 2. Crosses of genotypes included in stage 2 with genotypes included in stage 1 or 2 will be included in stage 3, and successively. This approach is good for capturing the breeding effort but will miss other factors such as ploidy (number of complete sets of chromosomes in a cell) or diversity (W. Macalpine, personal communication, July 2024). For instance, intraspecific (between same species) and interspecific (between different species) crosses will rank the same even though performance in the field might differ.

The genotypes were grouped according to their genetic diversity to facilitate the statistical analysis. Following the approach included in Rothamsted Research (2022, 2023), a total of 15 groups were created.

TABLE 1 | Climatic zones and their thresholds for the growing season based on monthly means from DDcum (5.6°) and soil moisture deficits. These have been adapted to the current data for better representation.

Climatic zone	Accumulated degree days above 5.6°C	Soil moisture deficit (mm)
Cool wet	0–1375	0–20
Cool moist	0–1375	20–50
Cool dry	0–1375	> 50
Warm wet	> 1375	0–20
Warm moist	> 1375	20–50
Warm dry	> 1375	> 50

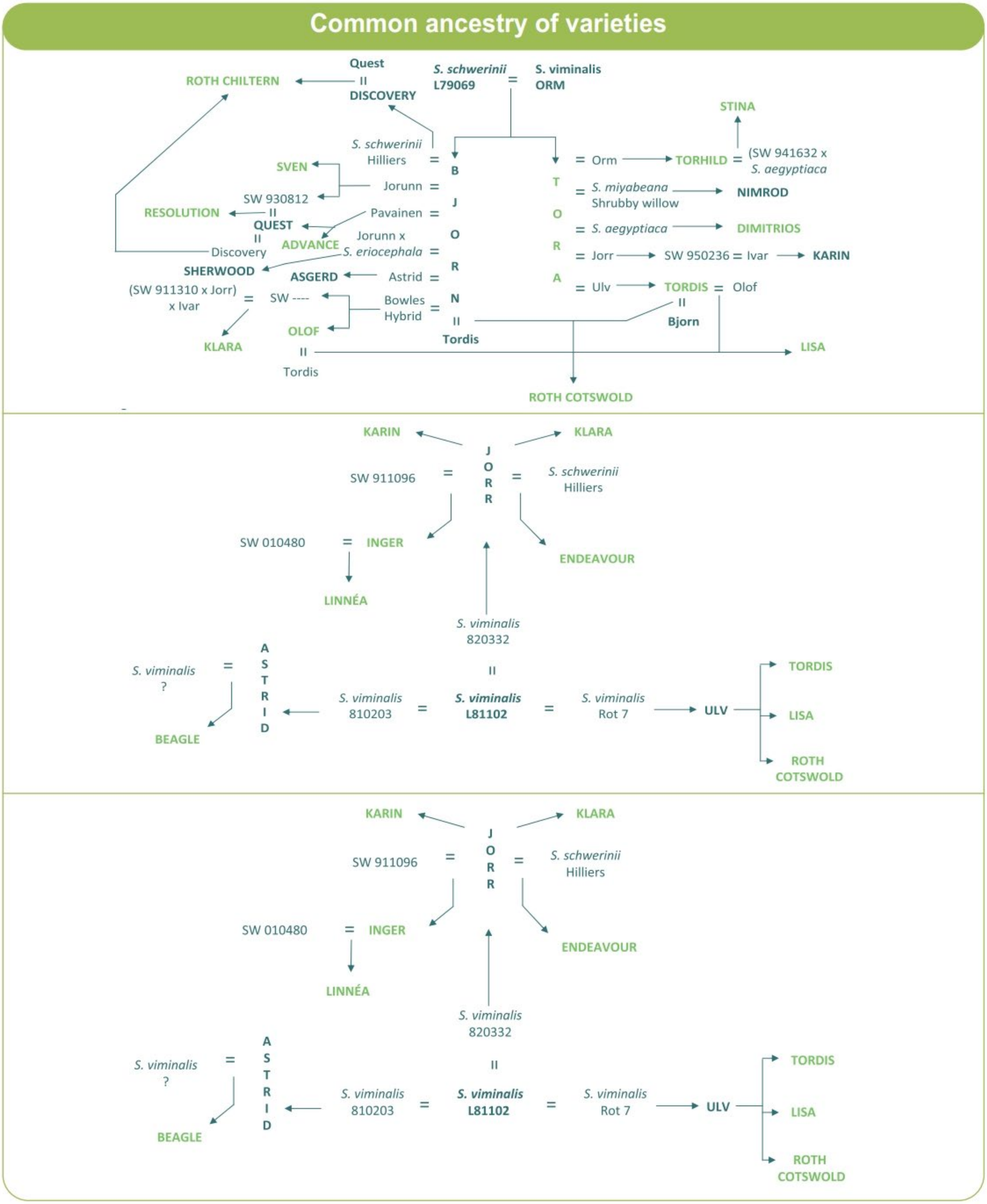

FIGURE 2 | Diagram used as a base for elaborating the parentage (breeding) of the genotypes collected in the willow metadata set. As presented in Caslin, Finnan, and McCracken (2012).

2.3 | Dataset Analysis

2.3.1 | Data Exploration

R (R Core Team 2022) and R Studio (R Studio Team 2020) were used to explore the data (Table S1) and carry out statistical analysis. For this purpose, some variables were coded differently to be able to run statistics on them. For instance, previous land use was coded as 'arable/grassland', fertilisation was set as 'yes/no', etc. Harvest usually takes place when coppiced willow is dormant (i.e., between autumn/winter). As most of the trials were harvested in the year after last growing season, harvesting carried out in the same year as the last growing season was changed to the following year to keep consistency.

There is a noteworthy difference in the data between final yield averaged over the length of the cycle and data independently for each of the cycle years. To fix this inconsistency in terms of statistical analysis, cumulative data was averaged per cycle length and establishment year data was removed. This led to a reduced number of 2178 observations. An issue detected when exploring the data is the large number of missing values (NA in dataset) leading to unbalanced data. Graphical exploration and descriptive statistics were performed on all the variables to remove those with little representation over all the data. Due to the data structure, soil and environmental factors associated with a particular site are increased by the number of genotypes included in the study. This can be overcome by subsetting only for the 'mix' (average) and 'mixtures' in the group column since exploratory analysis tends to provide similar results (Roback and Legler 2021).

2.3.2 | Statistical Analysis

The main objective for the statistical analysis in this paper was to find which of the 66 variables of interest were associated (Shmueli 2010) with the variation in yield. After exploring the dataset, the Random forests procedure (Breiman 2001) was considered the best fit as no linear relationship was observed when plotting the independent variables against the dependent variable, yield.

Random forests deal proficiently with non-linearity and provide more flexibility in regard to correlation and multicollinearity between explanatory variables (Estévez et al. 2023; Hanberry 2024). Random forests of multiple trees are created by resampling the data (bootstrapping), usually with repetition. Since not all variables are chosen for all trees, the algorithm for regression or classification is able to identify the most influential variables associated with the dependent variable and their relative importance. Random forests are also very effective when dealing with both continuous and categorical predictor variables presenting unbalanced data between levels (Díaz-Uriarte and de Alvarez Andrés 2006), as found in this analysis.

R package *randomForest* (Liaw and Wiener 2001) was used as a regression tool in two datasets derived from the main one. Firstly, a subset (cumulative dataset) of cumulative only data was used to show the relationship between explanatory variables and annual above-ground biomass (AGB) production over a full harvest cycle. The second subset (average dataset) is the modification of the original dataset by calculating the average per harvest cycle of the cumulative data to match it with the rest of the data. Both datasets had most variables in common except for planting density and fertiliser, only included in the average data and 'rootage' only included in the 'cumulative' model.

After screening and selecting the variables with a lower percentage of missing values, multiple combinations of variables and levels were considered, with small differences in variance explained and prediction correlation (PC) (data not shown). After consideration, it was decided to include all the variables in an initial model despite correlations found between some variables in the dataset (Figure S1), in case some information could be lost if left out following Hanberry (2024). Columns containing missing values up to 30% were also included since entire rows would be removed if any of the columns had a missing value, and potential useful information for the model could be lost (Hapfelmeier and Ulm 2014).

The variables chosen for the initial models in the cumulative and average subsets are shown in Table 2. The same combination of 57 variables was included in the initial model for the cumulative subset, with one including missing values (NA) and the other without missing values.

Root year ('*rootage*') was not considered for the analysis of the average subset. Shoot year ('*shootage*') was included since it would show if the cycle length of 2 versus 3 years had an impact on yield. Soil profiles number two and three were removed from the initial model, containing for the average subset to compare the initial model with and without missing values as we did in the cumulative subset. A total of 27 variables were included for analysis and four models were analysed (Table 2).

Continuous predictors were chosen over categorical when they provided the same information (e.g., sand, silt and clay percentage instead of soil type) for variables to be assessed as independently as possible. Despite random forests handling overfitting successfully (Breiman 2001), the datasets were split into train (80%) and test (20%) sets to be used for extra protection in avoiding overfitting.

A simple process of variable selection was followed by eliminating the 20% of variables that showed the lowest scores of variable importance (%incMSE) in the model. This was followed until the variance explained (VE) dropped and the root mean square error (RMSE) increased, indicating that important variables were missing for association with AGB production. The number of trees to be grown (ntree) and the number of predictors (mtry) to be chosen at each split of the data are two key parameters in random forest models that can be tuned or modified to achieve a higher accuracy or prediction. After some preliminary tests on how the initial model was affected by their modification, it was decided to leave it as default (ntree = 500; mtry = $p/3$, where p is the total number of predictors available at each step), since not much variation was observed (data not shown).

An ANOVA test and a t-test were carried out in both cumulative and average subsets, respectively, to show the possibilities derived from the main dataset. The shoot (stem) age must be interpreted differently for both sets. In the cumulative subset, each shoot year indicates the AGB production for that particular year. Shoot year in the average subset indicates the length of the harvest cycle.

3 | Results

3.1 | Summary and Statistical Analysis

The average temperature for all sites and harvest cycles was 12.91°C with a standard deviation (±1.3), with the maximum mean temperature of 18.9°C reached in Savoy (United States) and the minimum of 9.1°C in Aberystwyth (Wales). The average

TABLE 2 | Variables used in the initial model for both cumulative subset and average subset.

Variables cumulative subset				Variables average subset		
Latitude	Silt3	SoilAN3	Pwp3	Latitude	Bulkdensity1	Fctotal
Altitude	Clay1	SoilANstock1	Pawtotal	Altitude	Soilcarbon1	Pwptotal
Climate	Clay2	SoilANstock2	Fctotal	Climate	SoilCstock1	Density
Aspect_f	Clay3	SoilANstock3	Pwptotal	Aspect_f	SoilCtotal	Shootage
Avgtemp	Bulkdensity1	pH 1	SoilCtotal	Avgtemp	Fertiliser	Cyclenumber
Avgprecip	Bulkdensity2	pH 2	SoilANtotal	Avgprecip	pH 1	Diversity
Soildepth	Bulkdensity3	pH 3	Density	Thickness1	Paw1	Breeding
Thickness1	Soilcarbon1	Paw1	Shootage	Sand1	Fc1	
Thickness2	Soilcarbon2	Paw2	Rootage	Silt1	Pwp1	
Thickness3	Soilcarbon3	Paw3	Cyclenumber	Clay1	Pawtotal	
Sand1	SoilCstock1	Fc1	Diversity			
Sand2	SoilCstock2	Fc2	Breeding			
Sand3	SoilCstock3	Fc3				
Silt1	SoilAN1	Pwp1				
Silt2	SoilAN2	Pwp2				

precipitation for all sites was 545.28 (±137.29) mm. Aberystwyth was the site with the maximum precipitation over a harvest cycle with 1009.3 mm, with Geneva in United States reaching a minimum amount of precipitation with only 167 mm. As expected, average precipitation increases from dry to wet climatic zones and average temperatures are higher in the warm than in cool climatic zones (Table 3).

The annual average yield across all data is 9.01 (±4.05) Mg DM $ha^{-1} year^{-1}$ for all harvest cycles and genotypes (Figure 3). Yield varies along the six climatic zones, achieving a higher production in warm and wet areas. Data show that willow seems to perform better in wetter areas with warm conditions but not when temperatures are cooler under the same soil moisture conditions.

Annual production of biomass per country ranged from 6.87 (±3.13) to 10.6 (±6.35) Mg DM $ha^{-1} year^{-1}$ in the United States (US) and Canada, respectively. Maximum annual yield of 32.7 Mg DM $ha^{-1} year^{-1}$ was achieved in Canada (La Pocatière), located in a cool and wet area, for the genotype 'SX67' (*Salix miyabeana*) (Figure 4). Annual yield for genotypes grown in mixtures is slightly lower than for monocultures (Figure 5). However, mixtures were only grown in cool moist and wet climatic zones. After comparing only between these two areas, genotypes grown in mixtures perform slightly better than in monoculture.

Previous land use can be an important factor affecting yield (W. Macalpine, personal communication, July 2024). When subsetting for this, crop production where the previous land was grassland was slightly higher than when arable land, for a total of 23 sites. Application of fertilisers on land previously classified as arable had a higher impact on annual yield than when applied on land previously used as grassland (Figure 6). Best practice guidance in the United Kingdom (Caslin et al. 2023) advises planting the cuttings in spring to take advantage of the sun radiation early in the growing season. Data on planting dates comprising 15 sites and nine countries show higher yields when planted in March and April rather than in May and June (Figure 7).

Yield increase was among the main objectives of breeding programmes globally. Figure 8 shows data for genotypes bred in the United Kingdom, Sweden and United States. Mean values are similar across all stages, but median values increase, suggesting a yield increase of up to 1 Mg DM $ha^{-1} year^{-1}$ for genotypes included in the 4th stage. Planting densities were moved into a total of 11 groups due to some of them having differences of less than 100 plants ha^{-1} (Figure 9).

Annual yield in Mg (dry matter) DM $ha^{-1} year^{-1}$ (se) for all genotypes and harvest cycles included in the average dataset, is highest for 'Paramore' (13.16 ± 1.63), 'Ashton Parfitt' (12.35 ± 0.85), 'SX67' (12.31 ± 1.11), 'SW930984' (12 ± 2.4) and 'Endurance' (11.71 ± 0.59). The five lowest yielding genotypes were 'India' (3.9), 'Gustav' (3.35 ± 0.31), 'V7511' (2.73 ± 0.27), 'V794' (2.38 ± 0.24) and 'Baldwin' (0.8), with only one sample for 'India' and 'Baldwin' in the dataset (Figure S2).

A total of 20 genotypes (Figure 10), including two types of mixtures, are above the economical and commercially viable threshold of 10–12 Mg DM $ha^{-1} year^{-1}$ (Lindegaard et al. 2001). They range through four breeding stages (1 to 4, as described in methods) and six diversity groups (*S. dasyclados* hybrid, *S. miyabeana*, *S. viminalis x S. miyabeana*, *S. viminalis*, *S. viminalis x S. miyabeana* (b), *S. dasyclados*). Genotypes of *S. miyabeana*, grown only in the United States and Canada, achieved an annual yield of 10.42 (±0.5) Mg DM $ha^{-1} year^{-1}$ for all harvest cycles and sites (n = 134), followed by *S. spaethii* (n = 14) and *S. sitchensis* (n = 14) with 9.86 (±0.74) and

TABLE 3 | Average precipitation and temperature for the growing season (April–October) for each climatic zone. It shows the mean and standard deviation, the median with minimum and maximum values, and the number of missing values and its percentage.

	Climatic zone					
	Cool dry	**Cool moist**	**Cool wet**	**Warm dry**	**Warm moist**	**Warm wet**
Variables	**(*N* = 60)**	**(*N* = 91)**	**(*N* = 64)**	**(*N* = 1)**	**(*N* = 54)**	**(*N* = 47)**
Avgprecip (mm)						
Mean (SD)	439 (72.8)	487 (46.4)	597 (149)	584 (NA)	577 (128)	653 (111)
Median [min, max]	452 [207, 584]	492 [368, 586]	531 [384, 1010]	584 [584, 584]	575 [167, 848]	649 [410, 898]
Missing	0 (0%)	0 (0%)	2 (3.1%)	0 (0%)	2 (3.7%)	0 (0%)
Avgtemp (°C)						
Mean (SD)	13.4 (0.908)	12.6 (0.764)	11.6 (0.798)	16.3 (NA)	14.2 (1.21)	13.5 (1.26)
Median [min, max]	13.5 [9.80, 16.0]	12.8 [10.3, 13.9]	11.7 [9.10, 12.9]	16.3 [16.3, 16.3]	13.7 [12.9, 18.9]	13.1 [11.2, 15.3]
Missing	0 (0%)	0 (0%)	2 (3.1%)	0 (0%)	5 (9.3%)	0 (0%)

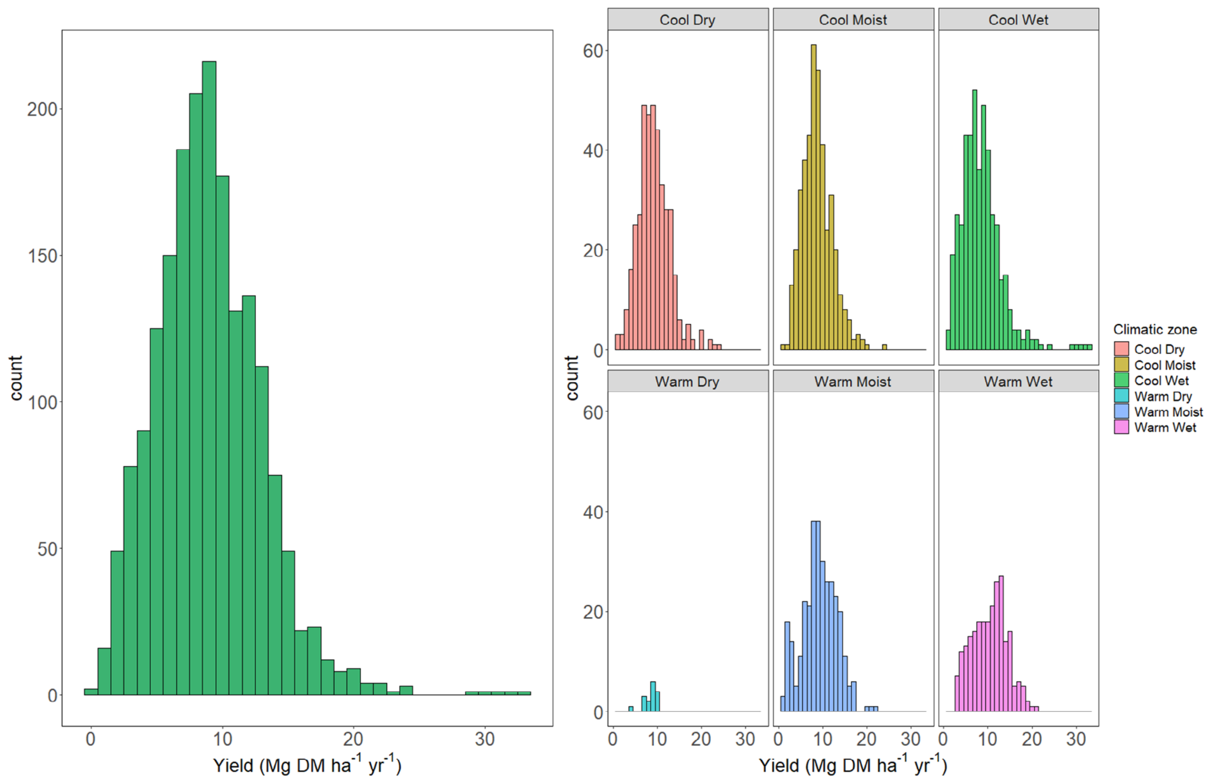

FIGURE 3 | Annual yield for all genotypes over all sites and cycles in the average subset (left). This includes genotypes grown in monoculture and mixtures. On the right, histograms of the annual yields (Mg DM ha^{-1} year^{-1}) achieved for each climatic zone.

9.81 (±0.88) Mg DM ha^{-1} year^{-1}, respectively. Genotypes from the breeding stage number two ranked first with an annual yield of 9.08 (±0.13) Mg DM ha^{-1} year^{-1}, with genotypes at the breeding stage number five ranked the last, achieving 8.63 (±0.95) Mg DM ha^{-1} year^{-1} (Figure 11).

Analysis of shoot years in the cumulative dataset showed the AGB production independently for each year of the harvest cycle (Figure 12). In the first cycle of 3 years, 1-year-old shoots produce lower AGB than 2- and 3-year-olds ($p < 0.0001$). Shoots of 2- and 3-year-olds are not significantly different from each other ($p > 0.05$). When looking at cycle number two, AGB from 2-year-old shoots was significantly lower than 1- and 3-year-old shoots ($p < 0.001$), but no significant difference was found between the latter ($p > 0.05$).

When comparing shoots of the same age between the two cycles, a significant difference was found between all of them ($p < 0.05$).

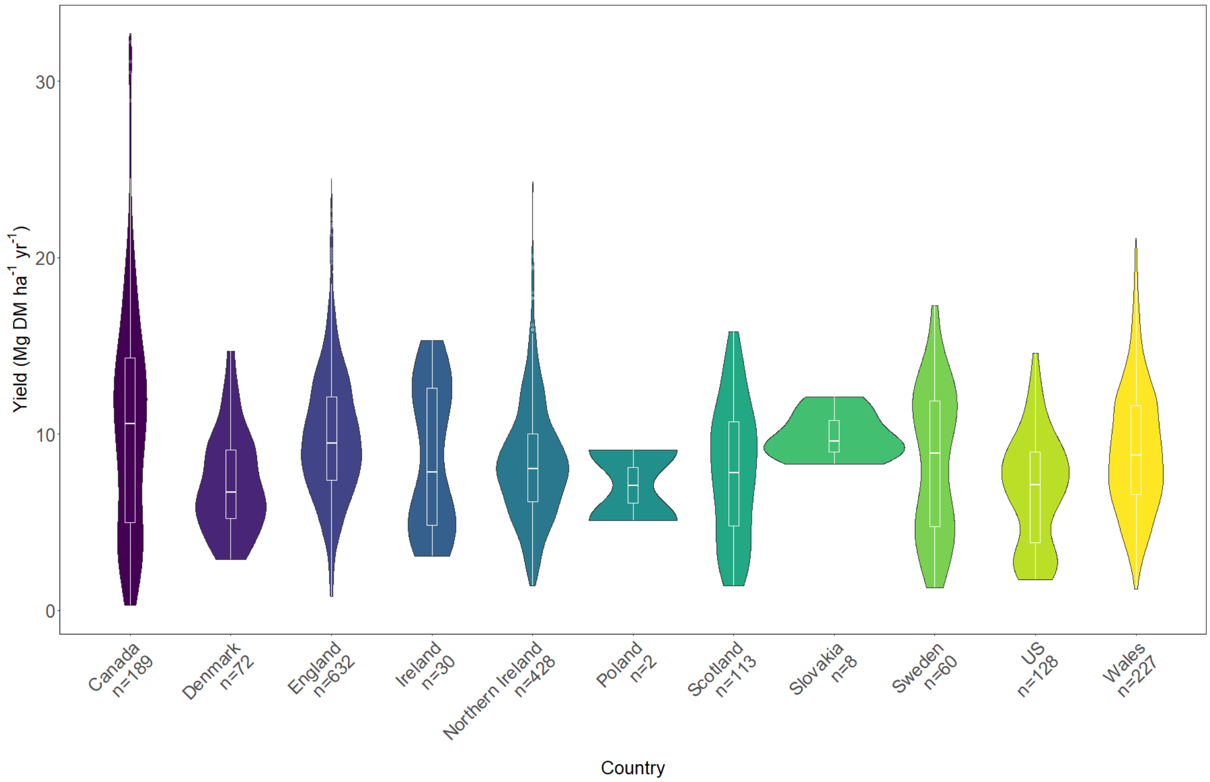

FIGURE 4 | Violin plots of annual yield (Mg DM ha^{-1} year^{-1}) achieved per country in the average subset. *N* indicates the number of samples (from Castellano Albors et al. 2025).

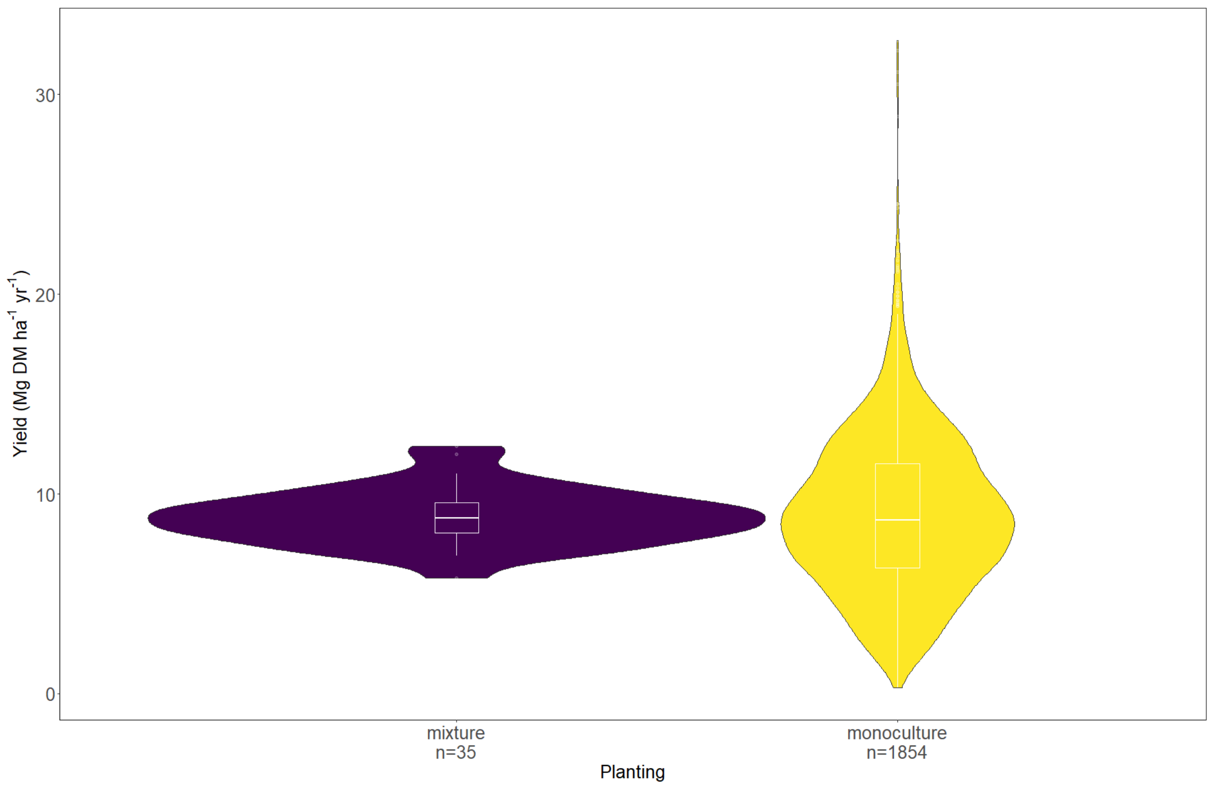

FIGURE 5 | Violin plot showing the annual yield (Mg DM ha^{-1} year^{-1}) for genotypes planted in monoculture and mixtures. Differences in shape accounts for the high number of samples for monoculture in comparison with the willow grown in mixtures. N indicates the number of samples.

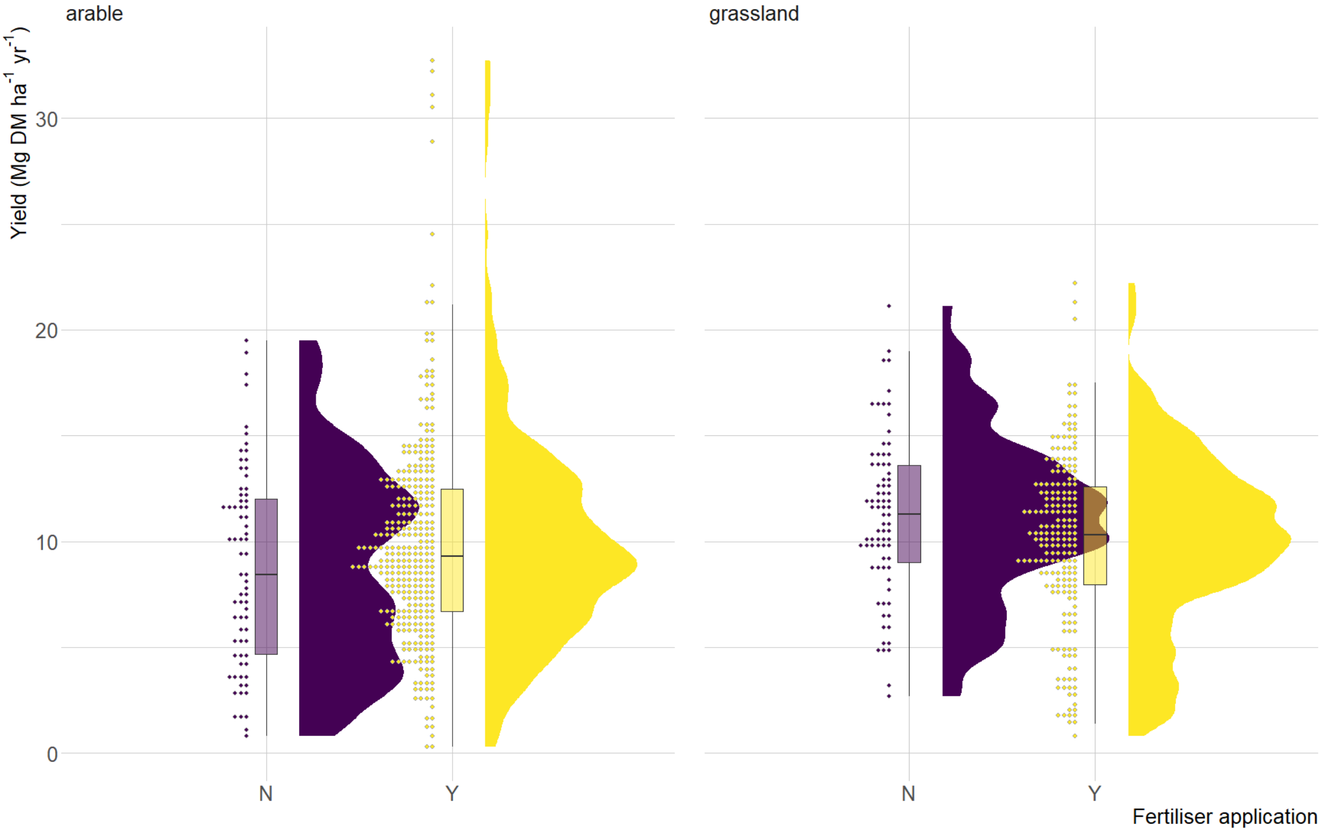

FIGURE 6 | Raincloud plot showing boxplot, sample distribution and density for the annual yield (Mg DM ha^{-1} year^{-1}) achieved by application of fertiliser Y (YES) or N (NO) for land previously used as arable or grassland.

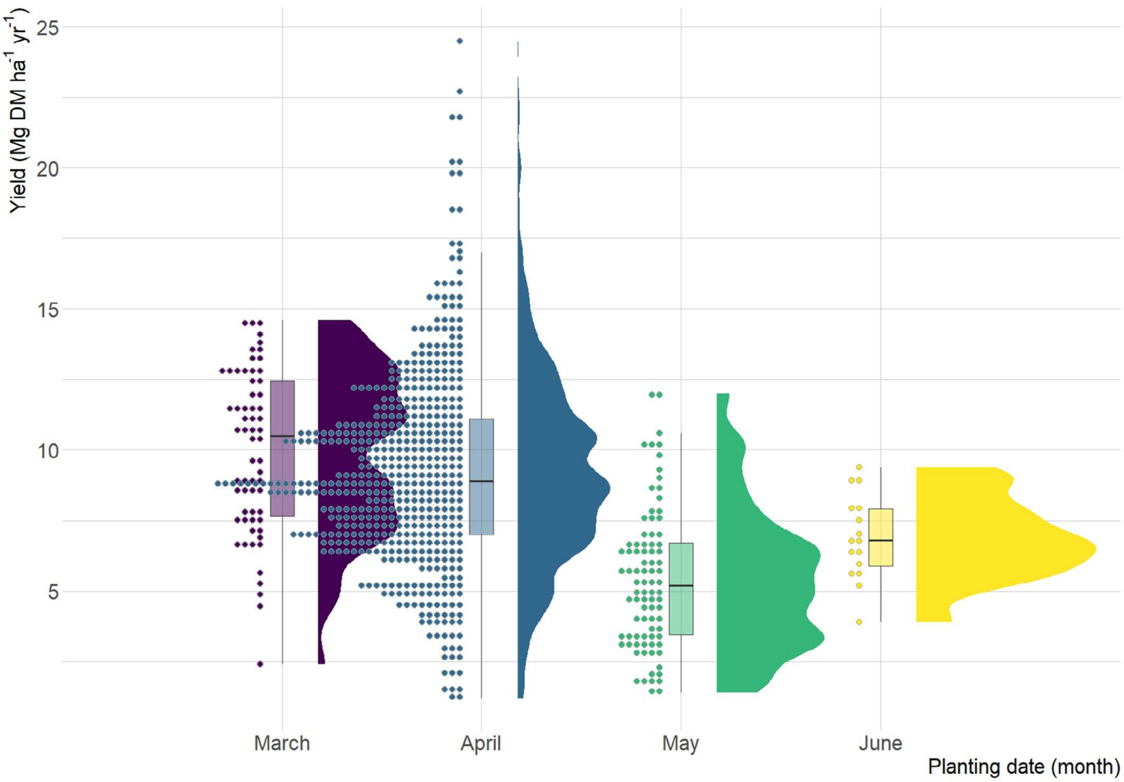

FIGURE 7 | Annual yield (Mg DM ha^{-1} year^{-1}) for all cycles when planting dates were provided. These will include cutback and non-cutback crops. Cutback happened when crops were coppiced during dormancy after the first growing season after planting. Non-cutback crops were left without coppicing.

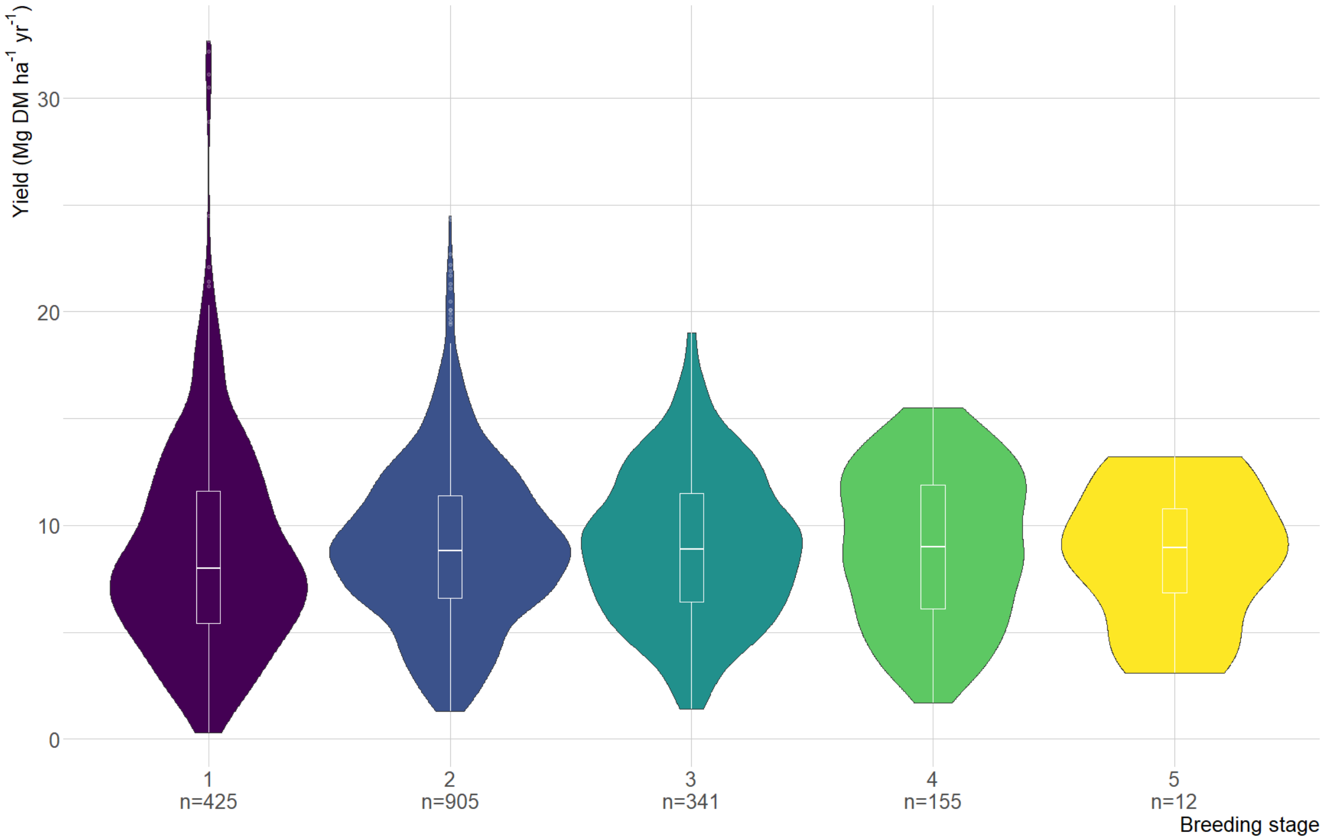

FIGURE 8 | Annual yield (Mg DM ha^{-1} year^{-1}) achieved by all genotypes in relation to the breeding stage. Oldest genotypes are included in group one and new crosses are found in group five (non-marketed). Stage 1 included genotypes as found in nature; Stage 2 includes any crosses between genotypes of Stage 1; Stage 3 included any crosses with genotypes from Stage 2; Stage 4 included any crosses with genotypes from Stage 3; and Stage 4 included any crosses with genotypes from Stage 5.

Shoots of 1- and 3-year-olds in the second cycle performed better in general than shoots of the same age in the first cycle. The opposite happened for 2-year-old shoots.

For the average subset, the mean yield for all sites when coppiced willow was harvested after two growing seasons was higher than those harvested after 3 years with 9.6 and 8.76 Mg DM ha^{-1} year^{-1}, respectively ($t = 3.87$, $p < 0.001$) (Figure 13).

3.2 | Random Forest on Cumulative Subset

3.2.1 | With Missing Values

The total number of models and associated variables, prediction correlation and RMSE are shown in Table S2. For the set including missing values, the maximum variance explained (58.9%) with the highest PC and lowest RMSE is model number six (optimum model), comprising 19 independent variables. It can predict up to 73% of the test set and the variance not explained by the model is 2.81. The VE by the different models is very similar until the optimum model. The next two models see a slight decrease in the VE and PA and a slight increase in the RMSE. Since we are interested in the GxE interaction, we would not choose the model number nine or subsequent models, due to the removal of the variable 'diversity'. However, VE and PA decreases substantially from here, and the RMSE increases. Thus, the variable selection process continues until there is only one explanatory variable in the model.

Model number eight is the last model with the variable 'diversity' included. The difference in VE between this and the optimum model (number six) is less than 2%, and more than 12% with model number 10. Models with less variables are preferred under similar conditions, thus, model number eight would be chosen over number six in this case.

3.2.2 | Without Missing Values

Model number three achieved the highest score in VE (56.7%) and the second highest score in PA (0.78) and RMSE (3.16). It includes 36 independent variables, 17 more than the optimum model, including missing values. The variance explained drops more than 8% when diversity is removed from the model (model number eight).

3.3 | Random Forest on Average Subset

3.3.1 | With Missing Values

Number five is the optimum model with higher VE (63.65%), approximately 5% higher than the initial model, PC (0.72) and lower RMSE (2.89) including 10 independent variables. The first model

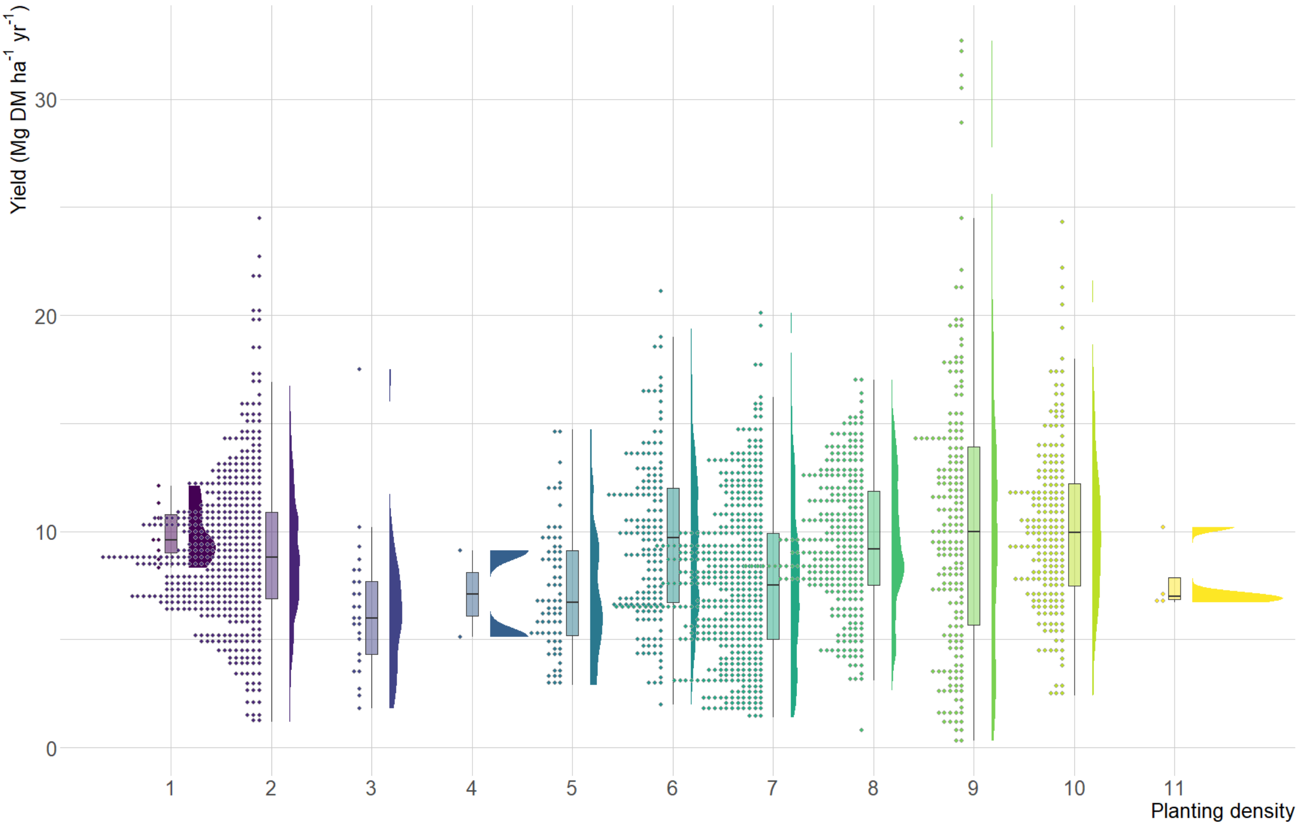

FIGURE 9 | Raincloud plots showing boxplot, sample distribution and density for each of the plant density groups. Density groups relates to the following planting densities (plants ha^{-1}): 1 (8.889); 2 (9875); 3 (10,000); 4 (11,000); 5 (12,000); 6 (13,000; 13,333); 7 (14,813; 15,000; 15,408); 8 (16,000; 16,600; 16,666; 16,667); 9 (18,000; 18,500); 10 (20,000); and 11 (25,000).

after removing 'diversity' is the number eight with only five variables and explaining up to 52.11% of the variance observed in yield.

3.3.2 | Without Missing Values

The initial model when soil profiles number two and three were not included explained almost 54% of the variance, with a prediction correlation of 0.81 and RMSE of 2.43 with 27 independent variables. Model number four was the optimum model with a VE of 62.99%, PC of 0.77 and RMSE of 2.41, including 13 independent variables. The 10 most important variables (highest association with biomass production) for the initial and optimum model for the four different options are shown in Figure 14. Figure 15 shows the correlation between predicted and measured values applied to the test data for the optimum models.

4 | Discussion

4.1 | Agronomy

With approximately 5000 ha of SRC willow in the United Kingdom, the dataset and meta-analysis presented here aim to better understand what the associated factors are with above-ground biomass production of SRC willow at a global scale, in order to help in the mass-scale deployment until 2050. Research trials spanning 30 years from 1992 until 2022 were collected. Site classification into different climatic zones by differences in accumulated degree days over 5.6°C and soil moisture deficit allowed us to consider the regional variation when factors such as elevation are included, among others.

Here, we presented biomass production on a year-by-year basis as measured in the field (Mg DM ha^{-1} year^{-1}) and also averaged over all the years of the cycle (Mg DM ha^{-1} year^{-1}). The former included 49 sites from Forest Research trials established in the mid-90s in the United Kingdom (Aylott et al. 2008).

SRC willow presents many challenges that have stimulated interest and driven research since the 1990s. A key limitation of the extrapolation of the results provided here is the unbalanced research to date in regard to willow grown in mixtures versus monoculture. As seen in Figure 4 and consistent with previous studies, the overall performance of individual genotypes within a mixture will be at least equal to or higher than the yield achieved by the same genotypes grown in monoculture. Only three studies with reported yields from mixtures were found, all located in Northern Ireland. In the United Kingdom, current projects such as UKRI-funded Perennial Biomass Crops Greenhouse Gas Removal (PBC4GGR) and DESNZ-funded Biomass Connect include up to six genotypes of willow grown in mixtures, with promising knowledge gain for future research.

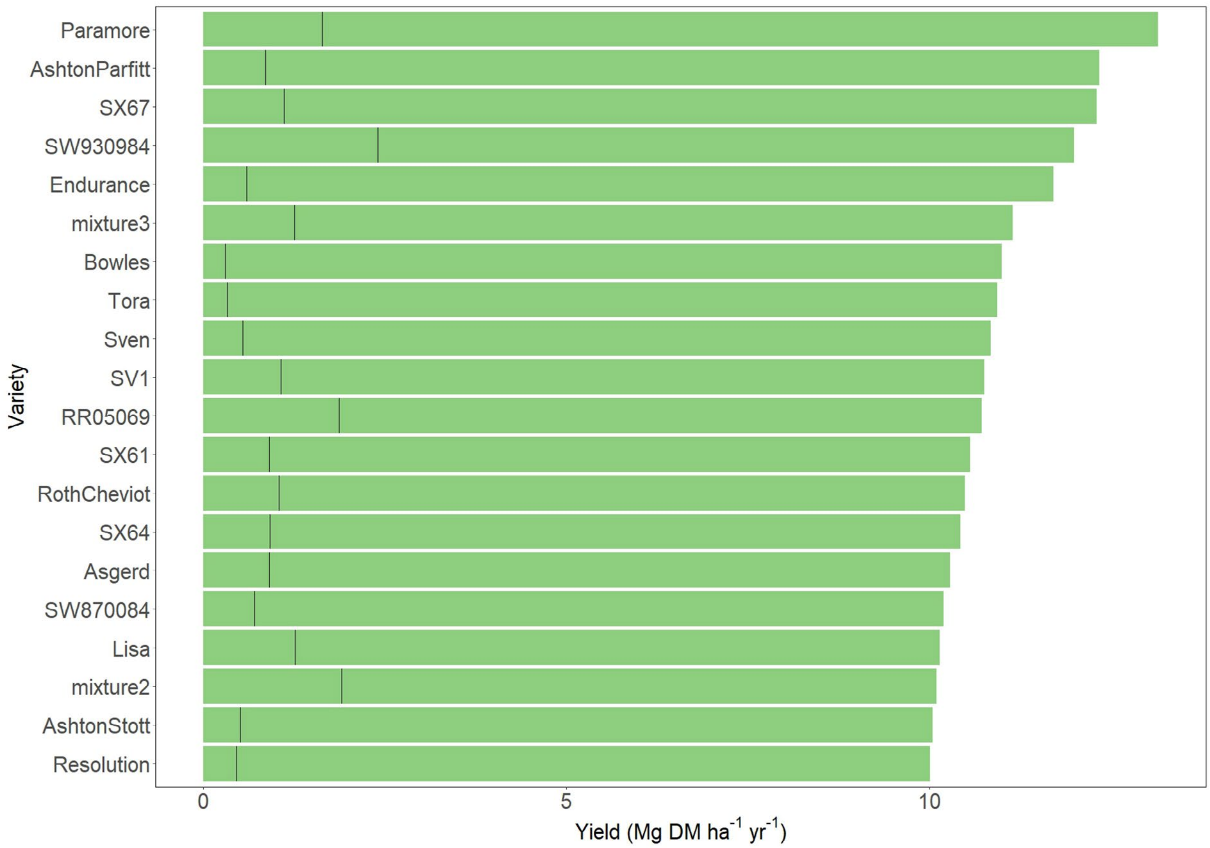

FIGURE 10 | Average yield (Mg DM $ha^{-1} year^{-1}$) for the 20 most productive genotypes (including two types of mixtures) in the dataset. Standard error bars are shown in black inside the columns.

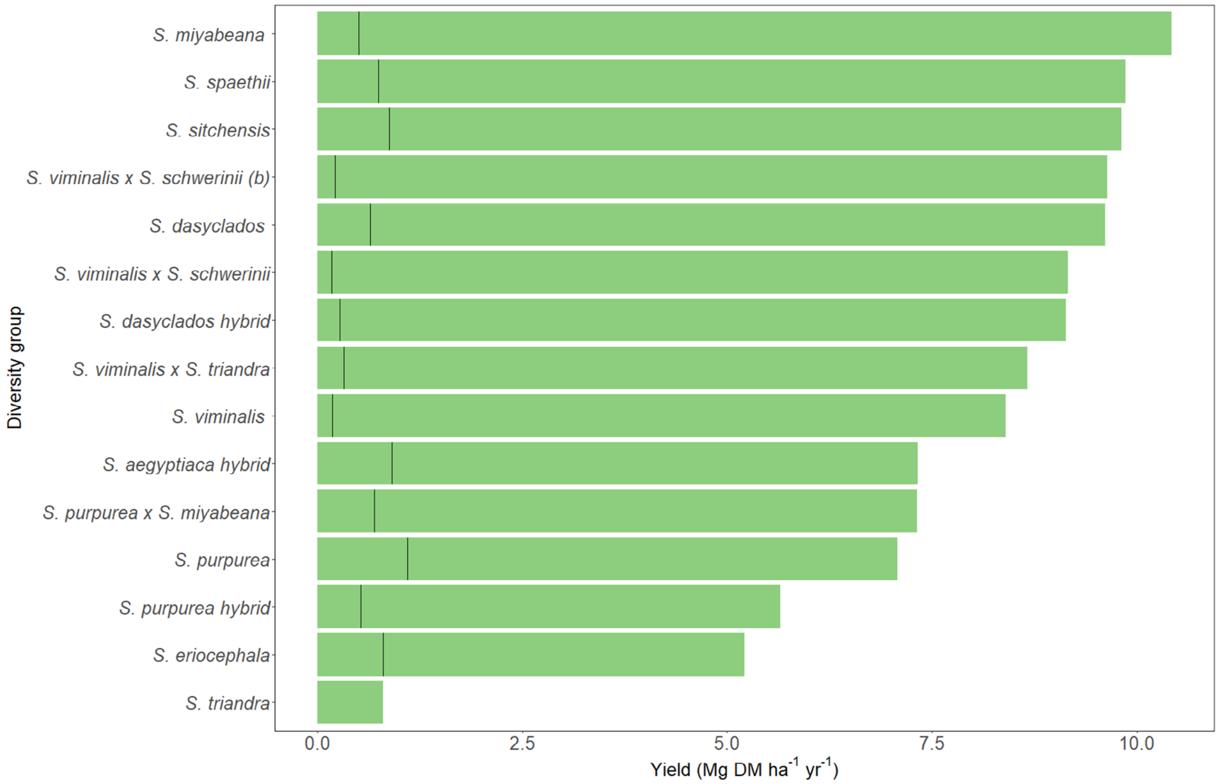

FIGURE 11 | Annual yield (Mg DM $ha^{-1} year^{-1}$) achieved by diversity group. Standard error bars are shown in black inside the columns except for *S. triandra* with only one value in the dataset.

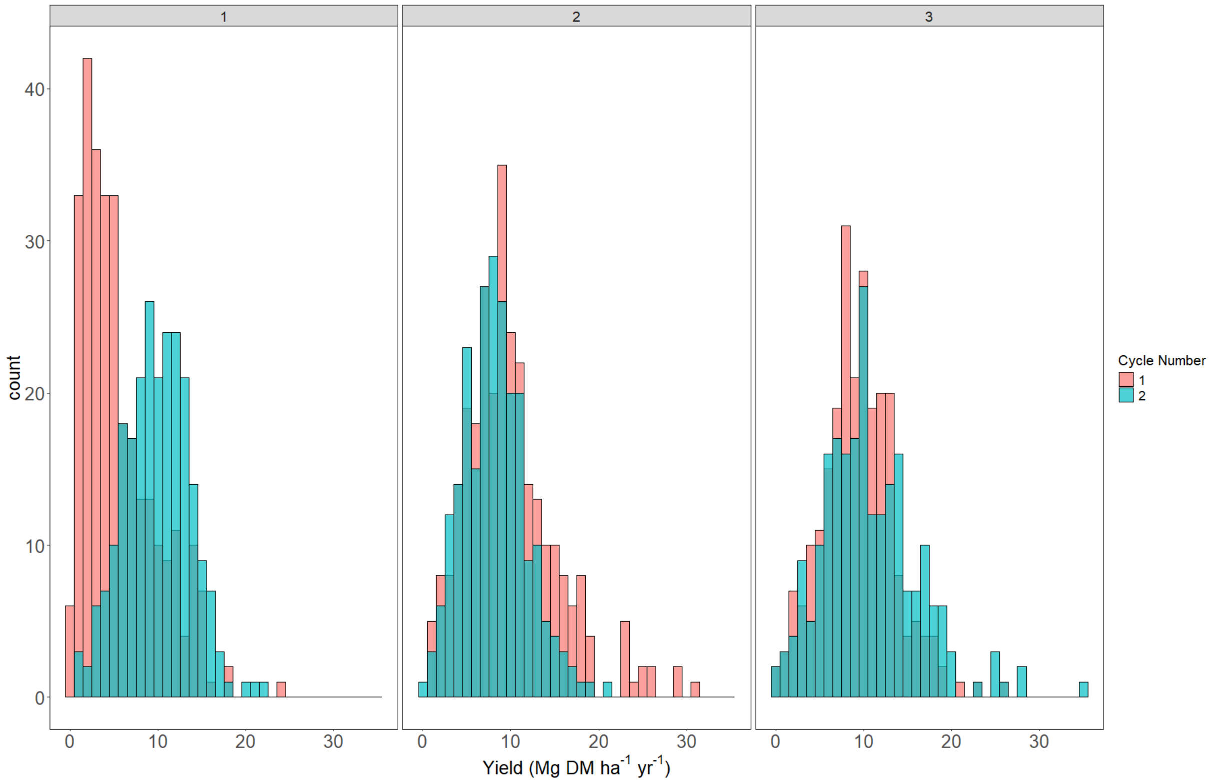

FIGURE 12 | Histograms of annual yield (Mg DM ha^{-1} year^{-1}) by shoot year for the first two cycles in the cumulative subset.

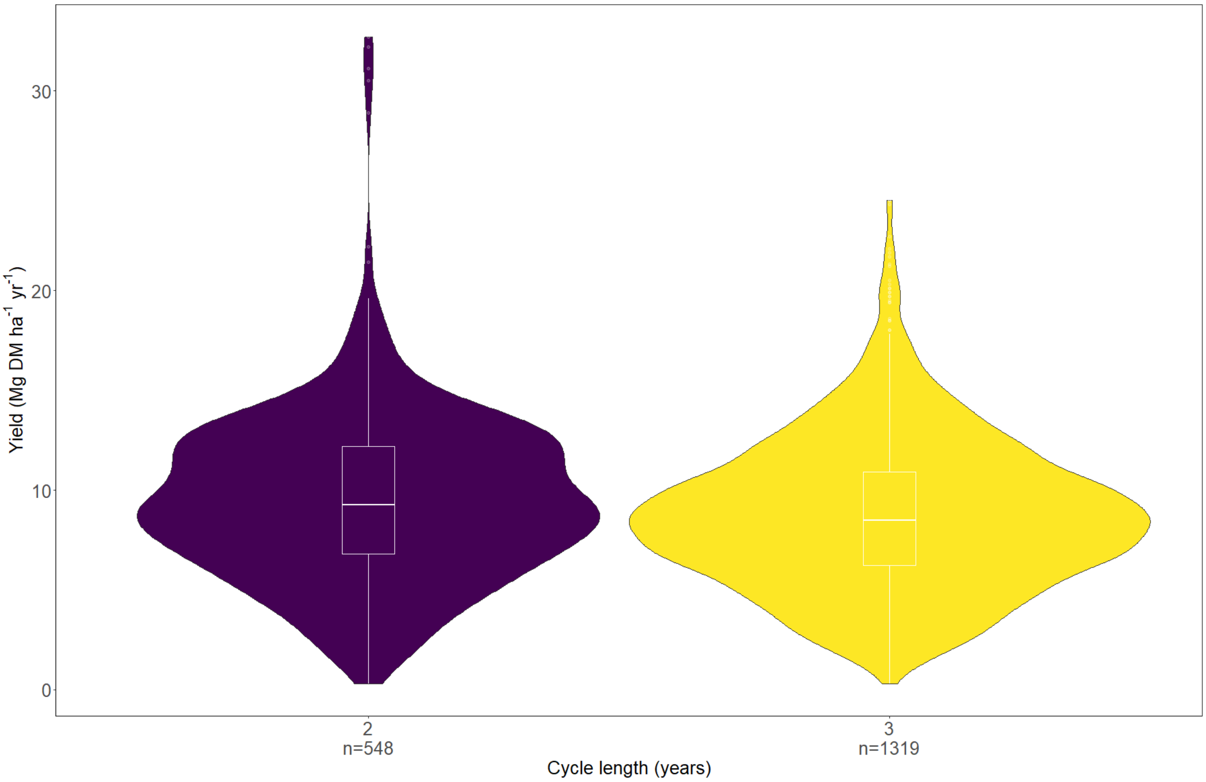

FIGURE 13 | Violin plots showing the difference in yield (Mg DM ha^{-1} year^{-1}) between coppiced willow harvested after 2- and 3-year cycles. N indicates the number of samples.

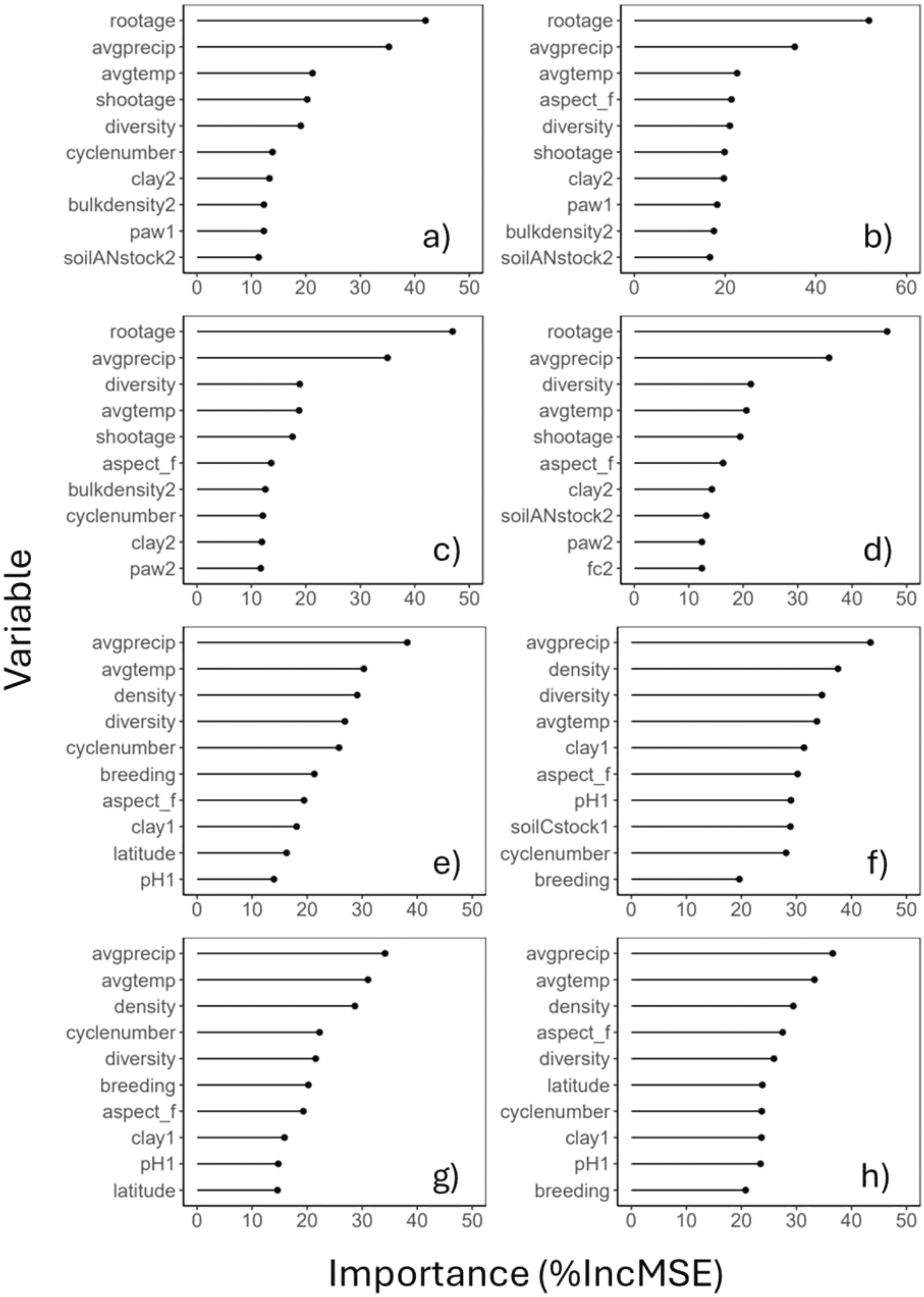

FIGURE 14 | Plots showing the 10 variables that scored highest according to randomForest's importance measure (%IncMSE). The initial models are shown on the left and the optimum models are shown on the right for the following subsets: (a, b) cumulative with missing values; (c, d) cumulative without missing values; (e, f) average (only upper soil profile) with missing values; (g, h) average (only upper soil profile) without missing values.

According to Lindegaard et al. (2011), yields in commercial plantations are expected to be 20% lower than those achieved in experimental trials, suggesting a bigger challenge to achieve the economical threshold of 10–12 Mg DM $ha^{-1} year^{-1}$. Yields presented here may also differ from commercial activities, due to crop trial yield estimates sometimes being upscaled from plot level and with associated survival rates. Low yields due to poor establishment, rust infection or other setbacks were reported

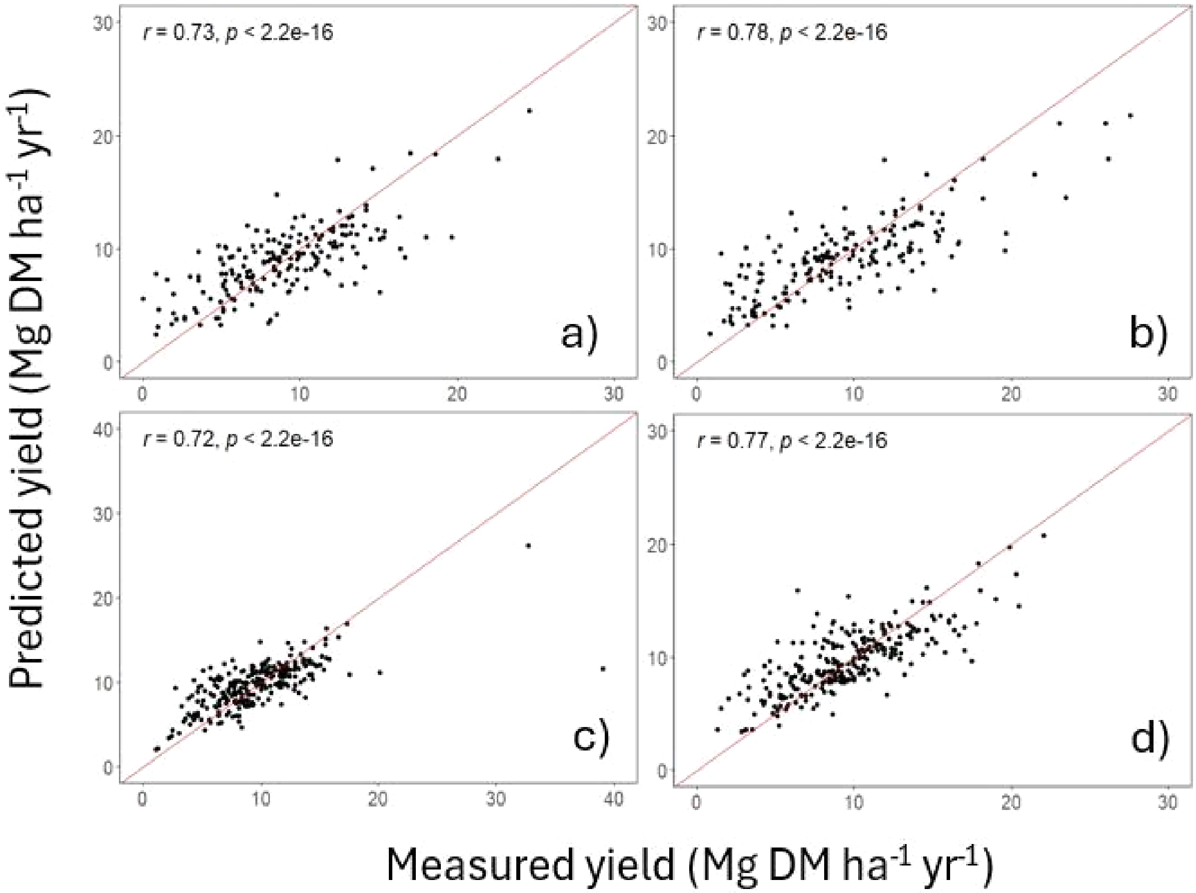

FIGURE 15 | Correlation plots between yield predicted by random forest model in the test set. Only optimum models shown as follows: (a) cumulative with missing values; (b) cumulative without missing values; (c) average (only upper soil profile) with missing values; and (d) average (only upper soil profile) without missing values.

and kept in the dataset. Despite challenging the understanding of the relationship between explanatory variables and yield, the impact of these issues was considered part of the commercial reality growers face.

Globally, breeding programmes increased yield over 1 Mg DM $ha^{-1} year^{-1}$ over four crosses. However, genotypes in the United States and Canada are only at their second breeding stage. When considering European genotypes only, across all climatic zones, an increment in yield was observed of 1.8 Mg DM $ha^{-1} year^{-1}$ from the first breeding stage to the second breeding stage but no further improvement since then. When climatic zones are included, an improvement in yield was observed for all climatic zones except for 'Cool Wet' (Table 4).

American genotypes show a decline in yield from the first breeding stage to the second breeding stage of 3.8 Mg DM $ha^{-1} year^{-1}$. These observations must be considered with caution due to the high yields of genotypes from first breeding stage observed in trials planted in Canada from Labrecque et al. (2023). Factoring in climatic zones did not show any yield improvements.

Different management methods can not only produce a significant difference in yield, but also a difference in the economics and sustainability of the plantation. The use of fertilisers, pesticides and herbicides can increase the running costs but might offer a better economic return via higher yields (Nordborg et al. 2018). While significant differences in yield due to fertilisation were found in the individual studies, Fabio and Smart (2018) did not find the same correlation for all studies combined in a meta-analysis. They suggested that different environments and genotypes (European vs. American) could lead to differences in nitrogen effectiveness. Similarly, no difference in yield was found between plots with nitrogen added from those that were not fertilised, 9.31 and 9.33 Mg DM $ha^{-1} year^{-1}$, respectively. As seen in Figure 6, not only were yields higher for former grassland but also the yield was higher for non-fertilised crops. Fertilised crops in former arable land benefited from fertilisation, with yields increased by almost 1 Mg DM $ha^{-1} year^{-1}$.

The importance of weed control through mechanical or chemical methods is captured in the dataset where only one site in Canada, Boisbriand, did not apply any weeding method. Mean yield achieved on this site was 10.77 (±1.88) for 16 American genotypes, which is within the economically viable threshold. Albertsson et al. (2014) reported reductions in stem dry weight of up to 94.5% on average across three sites in Sweden. When mechanical and herbicide application methods were compared, mean yields for the former were 3.4 Mg DM $ha^{-1} year^{-1}$ higher than the latter (7.5 Mg DM $ha^{-1} year^{-1}$). If mechanical weed control were applied, energy intensity and carbon footprint could be lower than when applying

TABLE 4 | Average yield (Mg DM ha^{-1} $year^{-1}$) for each breeding stage and climatic zones.

	Climatic zone					
Breeding	**Cool dry**	**Cool moist**	**Cool wet**	**Warm moist**	**Warm wet**	**Total**
1	8.59	6.96	7.31	8.25	7.76	7.49
2	8.31	9.35	8.53	10.97	9.58	9.16
3	9.86	8.25	6.64	11.24	10.54	9.21
4	10.22	8.40	5.64	12.67	9.05	8.90
Total	9.67	8.31	7.21	10.78	8.91	8.71

herbicide (Nordborg et al. 2018). Trial plots are much smaller than commercial plantations and mechanical weeding may only be feasible at small-scale plantations.

During the first harvest cycle, 1-year-old stems were significantly different in dry weight from 2- and 3-year-old stems. Although yields from the establishment year (biomass production in the growing season following planting) were included, there were only a few samples of data, and the difference between stem years was still significant when removed. During the first growing season after planting, roots started developing from cuttings holding back stem growth. The 1-year-old stems that were cutback (coppiced) after the first growing season following planting (2-year-old roots) produced higher yields than non-cutback crops, due to roots being developed during the establishment year. Yields in the second harvest cycle were higher than in the first cycle for all but the second shoot year. While higher yields for the second harvest cycle have been reported in previous studies, it is not possible to know from our current data why 2-year-old stems did not follow this trend.

Harvest cycles' length may vary between countries. In Sweden, 3–5-year cycles are common (Mola-Yudego and Aronsson 2008), whereas best practice guidance in the United Kingdom suggests 3-year cycles for an optimum balance between yield and costs associated with harvest operations (Caslin et al. 2023). Only six sites presented results on 4-year cycles and were left out from the analysis, as shown in Figure 12. Johnston et al. (2022) and Labrecque et al. (2023) studied the difference in yield between 2-year and 3-year cycles in Northern Ireland and Canada, respectively. While yields from 3-year cycles in Northern Ireland were about 10% higher than 2-year cycles, trials in Canada did not find generally any differences between them. The *t*-test showed a significant but not large difference (less than 1 Mg DM ha^{-1} $year^{-1}$) between the two cycles' lengths ($t = 3.87$, $p < 0.001$). It included 22 sites for 2-year cycles and 88 sites for 3-year cycles.

Planting densities showed variability, but no correlation between yield and density was observed. Bullard et al. (2002) found significant differences and higher yields from higher densities; however, some of these were beyond the current best practice (e.g., 111,000 plants ha^{-1}). Planting dates (shown by month in Figure 6) show the importance of getting the cutting in the soil at the right time of the year to increase the yield by taking advantage of the solar radiation as early as possible in the growing season. The data include annual yields averaged across the cycle years, with most of the trials being cutback after the first growing season and only four sites being left to grow without coppicing. Cutback after the establishment year has been the standard practice globally for SRC willow, as it allows the willow to grow more stems in the following year and to apply herbicide to last the full harvest cycle. Tubby and Armstrong (2002) advise to cutback only if the plants have grown healthy and are able to shade the surrounding weeds. Conversely, McKenzie (2011) and Albertsson et al. (2014) found that coppicing after establishment negatively impacted biomass production and survival in plots in Scotland and Sweden, respectively. It would be interesting to gather some more data on non-cutback trials to understand under what conditions cutback could be beneficial or negative and transfer this advice to the growers, since the current commercial standard in the United Kingdom is to not cutback after the establishment year (Lindegaard, pers. comm., January 2024).

4.2 | Random Forest

Performance of SRC willow vary according to genotype by environment interaction, agronomy and management methods. These factors were analysed through random forest's algorithm for two different subsets (Table 3), using the maximum number of variables as possible as described in methods. For a meta-analysis conducted on data collected from multiple and diverse sources, the data usually entails lack of standardisation and homogeneity between values that in most cases can be overcome (Ledo et al. 2019). For a sample point in the dataset, values collected for other variables associated with a missing value will be discarded, with the risk of losing some important information in the process. Similar results were obtained for all models with this 'default' approach where missing values were handled with the default mode in the R package.

Due to both subsets not having the exact set of variables included, a direct comparison of the importance of all variables was not possible. However, some common variables were found to be highly associated with AGB production as shown below. The results observed for all five models suggest that overfitting was overcome and correlated variables did not inflate the variance explanation. While Hapfelmeier and Ulm (2014) agree with the inclusion of missing values as a much more reliable approach than leaving the analysis to complete cases only, they suggested that variable importance scores would benefit from an extra method to help random forest handle missing values. Similarly, it has been suggested that variable importance selection may be biased by how the random forest algorithm favours the selection of categorical variables with a large number of levels against fewer levels (Strobl et al. 2007), and the lack of a significant threshold for separating important from non-important variables (Janitza et al. 2018).

R package *randomForest* was unable to provide an importance measure or ranking of the levels included in categorical predictors. Thus, this interpretation was left to subjectivity and previous knowledge. The Caret package in R was tried but discarded since it did not handle missing values. It would be advised to try other random forest packages or use more classic statistical tests to check for these differences within predictors.

4.2.1 | Annual Yield

Initial random forest models applied to the cumulative subset showed the same five most important variables with the growing season mean precipitation ('*avgprecip*') and the root year ('*rootage*') as the two most important in the same order. The age of the stems ('*shootage*'), the diversity group ('*diversity*') and the growing season mean temperature ('*avgtemp*') were all among the five most important variables for the initial and optimum models with and without missing values.

Root year ('*rootage*') indicates the year since cuttings were planted is the most important variable according to the model and indicates the year since the cuttings were planted. Rootage is the most important variable associated with AGB production according to the model and consistent with previous studies from Pacaldo et al. (2013) and Cunniff et al. (2015) where yield increase along the years for every cycle and within cycles. Similarly, our data shows similar results with an increase in yield over the first three harvest cycles (data not shown). However, shoot year (shootage) did not show any increasing pattern for any of the harvest cycles. Conversely, when harvest cycle was not included, yield showed an increase from beginning to the end of cycle.

Precipitation and temperature are important factors associated with the performance of SRC willow (Rothamsted Research 2023). Richard et al. (2019) simulated the performance under different climate scenarios of different genotypes, with the latter showing clear differences in yield due to water availability and temperature. Under the same conditions, different genotypes performed differently due to differences in water use efficiency (biomass gained by unit of water used) and radiation use efficiency (biomass gained by unit of solar radiation), among others. Lindroth and Båth (1999) estimated AGB production based entirely on precipitation, achieving good results. Although plant available water in soil did not reach a high score, Henner et al. (not published) found an important association between AGB production and proximity to water bodies.

Despite including in the model several variables highly correlated, the variance explained, prediction correlation between train and test data and RMSE associated with it do not vary in great proportion, as observed in Table S2. These results are consistent with Hanberry (2024), where models with correlated variables not only did not overfit but also provided the better association with the dependent variable. Despite only being interested in models including the variable associated with the genotype (GxE), the backward selection was continued as described in methods. A drop in variance explained (VE) was found in both with and without missing values models. While this drop in VE in the model with missing values was associated with a drop in PC and an increase in RMSE, we did not observe the same variation for the model without missing values.

4.2.2 | Yield Averaged Over Full Cycle

When only the upper soil profile was included, due to a high percentage of missing values in the other two soil profiles, the VE was higher for the models including missing values than for those without them. Conversely, models without missing values performed better in terms of PC and RMSE. Similarly to models applied in the cumulative subset, the two initial models with and without missing values share the same most important variables with slight variation in the order observed. Their two optimum models share eight variables within the 10 most important (Figure 13).

In summary, we have showed the versatility of a global SRC willow dataset by running an ANOVA test on a subset for annual biomass production and a *t*-test for final yield averaged over the full harvest cycle. Random forest was applied to both subsets with the maximum number of samples possible. The large number of locations with different types of management and environmental characteristics will enable the analysis of the variability in the performance of willow due to GxE interactions. This will allow growers to know the potential yield for different genotypes according to the site characteristics and agronomy to be applied. Despite some of the genotypes being outclassed or not marketed, a diversity group variable was created to help compare them to newer genotypes with similar backgrounds. Different combinations of genotypes per site could be used to assess the yield from a mixture of genotypes and fill the gap between research and industry.

With most of the genotypes showing a mean yield below the economical threshold of 10–12 Mg DM ha^{-1} $year^{-1}$ and a potential 20% yield reduction when applied in commercial plantations, willow's purpose is not all about the yield; it is important to consider the environmental benefits reported from willow plantations, such as soil remediation and Carbon Dioxide Removal (CDR) for carbon markets, nutrient retention buffers for agriculture and wastewater treatment, and flood control as alternative incentives for planting, along with ongoing breeding efforts.

The lack of standardisation and homogeneity in the reporting of soil and environmental factors led to some assumptions that must be considered when analysing this data. When not reported, precipitation and temperature were collected from the closest weather stations, some located miles away or from half-degree estimations from CRU TS. Although random forest regression showed high association between the latter and final yield, in addition to being consistent with previous studies, caution must be taken. Easy access to complete environmental data ranging from soil properties to climatic factors from research trials could improve our understanding of the complex association between biomass production and genotype by environment

(GxE) interaction. This dataset could also be improved by the inclusion of newer genotypes and mixtures to be more representative of the commercial practice.

Author Contributions

Antonio Castellano Albors: conceptualization, data curation, formal analysis, investigation, methodology, writing – original draft. **Anita Shepherd:** conceptualization, supervision, writing – review and editing. **Ian Shield:** resources, writing – review and editing. **William Macalpine:** resources, writing – review and editing. **Kevin Lindegaard:** resources, writing – review and editing. **Ian Tubby:** resources, writing – review and editing. **Astley Hastings:** conceptualization, funding acquisition, resources, supervision, writing – review and editing.

Acknowledgements

This work is part of the Demonstrator Programmes PBC4GGR (Perennial Bioenergy Crops for Greenhouse Gas Removal), and Net Zero Plus funded by UK Research and Innovation (BB/V0115553/1 and BB/V011588/1, respectively). This research was also made possible by UKERC (UK Energy Research Centre) Phase 4 research programme (EP/S000771/1).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in figshare at [https://doi.org/10.6084/m9.figshare.28570385.v1] and [https://doi.org/10.6084/m9.figshare.28570433].

References

Albanito, F., A. Hastings, N. Fitton, et al. 2019. "Mitigation Potential and Environmental Impact of Centralized Versus Distributed BECCS With Domestic Biomass Production in Great Britain." *GCB Bioenergy* 11, no. 10: 1234–1252. https://doi.org/10.1111/gcbb.12630.

Albertsson, J., D. Hansson, N. Bertholdsson, I. Ahman, and C. Kempenaar. 2014. "Site-Related Set-Back by Weeds on the Establishment of 12 Biomass Willow Clones." *Weed Research* 54, no. 4: 398–407. https://doi.org/10.1111/wre.12086.

Armstrong, A. 1997. "The United Kingdom Network of Experiments on Site/Yield Relationships for Short Rotation Coppice." *Forest Research* The UK Network of Experiments on Site/Yield Relationships for Short Rotation coppice (forestresearch.gov.uk): 12.

Aylott, M. J., E. Casella, I. Tubby, N. R. Street, P. Smith, and G. Taylor. 2008. "Yield and Spatial Supply of Bioenergy Poplar and Willow Short-Rotation Coppice in the UK." *New Phytologist* 178, no. 2: 358–370. https://doi.org/10.1111/j.1469-8137.2008.02396.x.

Begley, D., A. R. McCracken, W. M. Dawson, and S. Watson. 2009. "Interaction in Short Rotation Coppice Willow, *Salix viminalis* Genotype Mixtures." *Biomass & Bioenergy* 33, no. 2: 163–173. https://doi.org/10.1016/j.biombioe.2008.06.001.

Bergante, S., M. Manzone, and G. Facciotto. 2016. "Alternative Planting Method for Short Rotation Coppice With Poplar and Willow." *Biomass & Bioenergy* 87: 39–45. https://doi.org/10.1016/j.biombioe.2016.02.016.

Berlin, S., H. R. Hallingbäck, F. Beyer, N. Nordh, M. Weih, and A. Rönnberg-Wästljung. 2017. *Genetics of Phenotypic Plasticity and Biomass Traits in Hybrid Willows Across Contrasting Environments and Years*. Oxford University Press (OUP). https://doi.org/10.1093/aob/mcx029.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. https://doi.org/10.1023/A:1010933404324.

Bullard, M. J., S. J. Mustill, S. D. McMillan, P. M. I. Nixon, P. Carver, and C. Britt. 2002. "Yield Improvements Through Modification of Planting Density and Harvest Frequency in Short Rotation Coppice Salix spp.—1. Yield Response in Two Morphologically Diverse Varieties." *Biomass and Bioenergy* 22, no. 1: 15–25. https://doi.org/10.1016/S0961-9534(01)00054-X.

Bush, C., T. A. Volk, and M. H. Eisenbies. 2015. "Planting Rates and Delays During the Establishment of Willow Biomass Crops." *Biomass & Bioenergy* 83: 290–296. https://doi.org/10.1016/j.biombioe.2015.10.008.

Cameron, K., L. Smart, B. Ballard, T. A. Volk, and L. P. Abrahamson. 2007. "Willow cultivars (Outreach and Extension, No. 2)." https://core.ac.uk/download/233603555.pdf.

Campbell, G. S. 1985. *Soil Physics With BASIC: Transport Models for Soil-Plant Systems*. Vol. 14. Elsevier.

Cao, Y., T. Lehto, T. Repo, R. Silvennoinen, and P. Pelkonen. 2011. "Effects of Planting Orientation and Density of Willows on Biomass Production and Nutrient Leaching." *New Forests* 41, no. 3: 361–377. https://doi.org/10.1007/s11056-010-9228-6.

Caslin, B., J. Finnan, C. Johnston, et al. 2023. "Envirocrops: Short Rotation Coppice Willow Best Practice Guidelines." [online] Envirocrops. https://envirocrops.com/.

Caslin, B., J. Finnan, and A. McCracken. 2012. "Willow Identification Guide." https://teagasc.ie/wp-content/uploads/media/website/publications/2012/Willow_Identification_Guide_2012.pdf.

Castellano Albors, A., A. Hastings, A. Shepherd, K. Lindegaard, I. Shield, and W. Macalpine. 2025. "SRC Willow Global Dataset for the Bioeconomy." Figshare. Dataset. https://doi.org/10.6084/m9.figshare.28570385.

CCC. 2020. "Land Use: Policies for a Net Zero UK." Land-use-Policies-for-a-Net-Zero-UK.pdf(theccc.org.uk).

Clifton-Brown, J., A. Harfouche, M. D. Casler, et al. 2018. *Breeding Progress and Preparedness for Mass-Scale Deployment of Perennial Lignocellulosic Biomass Crops Switchgrass, Miscanthus, Willow and Poplar*. Wiley. https://doi.org/10.1111/gcbb.12566.

Clifton-Brown, J., A. Hastings, M. Von Cossel, et al. 2023. "Perennial Biomass Cropping and Use: Shaping the Policy Ecosystem in European Countries." *GCB Bioenergy* 15: 538–558. https://doi.org/10.1111/gcbb.13038.

Commission for Environmental Cooperation. 2007. "North American Environmental Atlas." Retrieved from: Explore the North American Environment with the Atlas.

Cunniff, J., S. J. Purdy, T. J. P. Barraclough, et al. 2015. "High Yielding Biomass Genotypes of Willow (*Salix* spp.) Show Differences in Below Ground Biomass Allocation." *Biomass and Bioenergy* 80: 114–127. https://doi.org/10.1016/j.biombioe.2015.04.020.

DESNZ. 2023. "Biomass Strategy 2023." UK Government. https://assets.publishing.service.gov.uk/media/64dc8d3960d123000d32c602/biomass-strategy-2023.pdf.

Díaz-Uriarte, R., and S. de Alvarez Andrés. 2006. "Gene Selection and Classification of Microarray Data Using Random Forest." *BMC Bioinformatics* 7, no. 3: 13. https://doi.org/10.1186/1471-2105-7-3.

Dickmann, D. I., and J. Kuzovkina. 2014. *Poplars and Willows: Trees for Society and the Environment*. CABI. http://www.cabi.org/cabebooks/ebook/20143048413.

Don, A., B. Osborne, A. Hastings, et al. 2011. "Land-Use Change to Bioenergy Production in Europe: Implications for the Greenhouse Gas Balance and Soil Carbon." *GCB Bioenergy* 4, no. 4: 372–391. https://doi.org/10.1111/j.1757-1707.2011.01116.x.

Edelfeldt, S., A. Lundkvist, J. Forkman, and T. Verwijst. 2015. "Effects of Cutting Length, Orientation and Planting Depth on Early Willow Shoot Establishment." *Bioenergy Research* 8, no. 2: 796–806. https://doi.org/10.1007/s12155-014-9560-3.

Esri. 2020. *ArcGIS (Version 10.8.1) [Software]*. Environmental Systems Research Institute, Inc.

Estévez, V., S. Mattbäck, A. Boman, A. Beucher, K.-M. Björk, and P. Österholm. 2023. "Improving Prediction Accuracy for Acid Sulfate Soil Mapping by Means of Variable Selection." *Frontiers in Environmental Science* 11: 1213069. https://doi.org/10.3389/fenvs.2023.1213069.

ETSU. 1998. "Energy From Biomass." ENERGY FROM BIOMASS (forestresearch.gov.uk).

EuroGeographics. 2023. "Euro DEM Height Data." Retrieved from: Euro Dem | Eurogeographics.

Evans, S., M. Baldwin, P. Henshall, et al. 2007. "Final Report: Yield Models for Energy: Coppice of Poplar and Willow. Volume A—Empirical Models." Report to DTI (B/W2/00624/00/00 URN). Ed: I Tubby and J Poole. 91 pp. Microsoft Word—Volume A final IT.doc (forestresearch.gov.uk).

Fabio, E. S. 2017. *Environmental and Management Factors Influencing Shrub Willow Biomass Yield and Quality (Ph.D.)* ProQuest Dissertations & Theses A&I: Science & Technology (1960627738). Cornell University. https://www.proquest.com/dissertations-theses/environmental-management-factors-influencing/docview/1960627738/se-2?accountid=8155.

Fabio, E. S., and L. B. Smart. 2018. "Effects of Nitrogen Fertilization in Shrub Willow Short Rotation Coppice Production—A Quantitative Review." *GCB Bioenergy* 10: 548–564. https://doi.org/10.1111/gcbb.12507.

Google. 2024. "Google Earth Pro."

Gullberg, U. 1993. "Towards Making Willows Pilot Species for Coppicing Production." *Forestry Chronicle* 69, no. 6: 721–726. https://doi.org/10.5558/tfc69721-6.

Hallingbäck, H. R., J. Fogelqvist, S. J. Powers, et al. 2016. "Association Mapping in *Salix viminalis* L. (*Salicaceae*)—Identification of Candidate Genes Associated With Growth and Phenology." *GCB Bioenergy* 8, no. 3: 670–685. https://doi.org/10.1111/gcbb.12280.

Hanberry, B. B. 2024. "Practical Guide for Retaining Correlated Climate Variables and Unthinned Samples in Species Distribution Modeling, Using Random Forests." *Ecological Informatics* 79: 102406. https://doi.org/10.1016/j.ecoinf.2023.102406.

Hanley, S. J., and A. Karp. 2014. "Genetic Strategies for Dissecting Complex Traits in Biomass Willows (Salix spp.)." *Tree Physiology* 34, no. 11: 1167–1180. https://doi.org/10.1093/treephys/tpt089.

Hapfelmeier, A., and K. Ulm. 2014. "Variable Selection by Random Forests Using Data With Missing Values." *Computational Statistics & Data Analysis* 80: 129–139. https://doi.org/10.1016/j.csda.2014.06.017.

Harris, I., T. J. Osborn, P. Jones, and L. David. 2020. "Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset." *Scientific Data* 7: 109. https://doi.org/10.1038/s41597-020-0453-3.

Hastings, A., J. Clifton-Brown, M. Wattenbach, C. P. Mitchell, and P. Smith. 2009. "The Development of MISCANFOR, a New Miscanthus Crop Growth Model: Towards More Robust Yield Predictions Under Different Climatic and Soil Conditions." *GCB Bioenergy* 1, no. 2: 154–170. https://doi.org/10.1111/j.1757-1707.2009.01007.x.

Hastings, A., M. J. Tallis, E. Casella, et al. 2014. "The Technical Potential of Great Britain to Produce Ligno-Cellulosic Biomass for Bioenergy in Current and Future Climates." *GCB Bioenergy* 6, no. 2: 108–122. https://doi.org/10.1111/gcbb.12103.

Haughton, A. J., A. J. Bond, A. A. Lovett, et al. 2009. "A Novel, Integrated Approach to Assessing Social, Economic and Environmental Implications of Changing Rural Land-Use: A Case Study of Perennial Biomass Crops." *Journal of Applied Ecology* 46, no. 2: 315–322. https://doi.org/10.1111/j.1365-2664.2009.01623.x.

Holman, I. P., and T. M. Hess. 2014. *Land Use, Climate Change and Water Availability (Phase 2a)—Task B: Development of a Range of Plausible Future Land Use, Land Management and Growing Season Changes*. Unpublished. Cranfield Water Science Institute. http://dspace.lib.cranfield.ac.uk/handle/1826/8532.

IEA. 2023. "Tracking Clean Energy Progress 2023." https://www.iea.org/energy-system/renewables/bioenergy#tracking.

IEA Bioenergy. 2020. "Roles of Bioenergy in Energy System Pathways Towards a WB2 World: Workshop Report." https://www.ieabioenergy.com/wp-content/uploads/2020/07/Roles-of-bioenergy-in-energy-system-pathways-towards-a-WB2-world-Workshop-Report.pdf.

Janitza, S., E. Celik, and A. L. Boulesteix. 2018. "A Computationally Fast Variable Importance Test for Random Forests for High-Dimensional Data." *Advances in Data Analysis and Classification* 12: 885–915. https://doi.org/10.1007/s11634-016-0276-4.

Johnston, C. R., L. R. Walsh, and A. R. McCracken. 2022. "Effect of Two vs. Three Year Harvest Intervals on Yields of Short Rotation Coppice (SRC) Willow." *Biomass and Bioenergy* 156: 106303. https://doi.org/10.1016/j.biombioe.2021.106303.

Jureková, Z., G. Drazic, J. Milovanovic, et al. 2011. *External and Internal Factors Influencing the Growth and Biomass Production of Short Rotation Woods Genus Salix and Perennial Grass Miscanthus*. Faculty of Applied Ecology FUTURA Singidunum University Belgrade. https://www.researchgate.net/publication/313741524_External_and_internal_factors_influencing_the_growth_and_biomass_production_of_short_rotation_woods_genus_Salix_and_perennial_grass_Miscanthus.

Karp, A. 2011. "Hypes, Hurdles and Hopes: Energy and Biofuels From Biomass." *Biochemist (London, England)* 33, no. 2: 30–33. https://doi.org/10.1042/BIO03302030.

Karp, A. 2014. "Willows as a Source of Renewable Fuels and Diverse Products." In *Challenges and Opportunities for the World's Forests in the 21st Century. Forestry Sciences*, edited by T. Fenning, vol. 81. Springer. https://doi.org/10.1007/978-94-007-7076-8_27.

Karp, A., and G. M. Richter. 2011. "Meeting the Challenge of Food and Energy Security." *Journal of Experimental Botany* 62, no. 10: 3263–3271. https://doi.org/10.1093/jxb/err099.

Karp, A., and I. Shield. 2008. "Bioenergy From Plants and the Sustainable Yield Challenge." *New Phytologist* 179, no. 1: 15–32. https://doi.org/10.1111/j.1469-8137.2008.02432.x.

Labrecque, M., S. Daigle, and S. Olishevska. 2023. "Comparing Biomass Yields of Various Willow Cultivars in Short Rotation Coppice Over Six Growing Seasons Across a Broad Climatic Gradient in Eastern Canada." *Canadian Journal of Forest Research* 53: 533–543. https://doi.org/10.1139/cjfr-2022-0205.

Larsen, S., D. Jaiswal, N. S. Bentsen, D. Wang, and S. P. Long. 2016. "Comparing Predicted Yield and Yield Stability of Willow and Miscanthus Across Denmark." *GCB Bioenergy* 8: 1061–1070. https://doi.org/10.1111/gcbb.12318.

Larsen, S. U., U. Jørgensen, J. B. Kjeldsen, and P. E. Laerke. 2014. "Long-Term Yield Effects of Establishment Method and Weed Control in Willow for Short Rotation Coppice (SRC)." *Biomass & Bioenergy* 71: 266–274. https://doi.org/10.1016/j.biombioe.2014.10.001.

Lecocq, F., H. Winkler, J. P. Daka, et al. 2022. "Mitigation and Development Pathways in the Near- to Mid-Term." In *IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by P. R. Shukla, J. Skea, R. Slade, et al. Cambridge University Press. https://doi.org/10.1017/9781009157926.006.

Ledo, A., J. Hillier, P. Smith, et al. 2019. “A Global, Empirical, Harmonised Dataset of Soil Organic Carbon Changes Under Perennial Crops.” *Scientific Data* 6, no. 1: 57. https://doi.org/10.1038/s41597-019-0062-1.

Liaw, A., and M. Wiener. 2001. “Classification and Regression by RandomForest.” *Forest* 23: 18–22.

Lindegaard, K., M. M. Carter, A. McCracken, et al. 2011. “Comparative Trials of Elite Swedish and UK Biomass Willow Varieties.” *Aspects of Applied Biology* 112: 57–65. (Biomass and Energy Crops IV). (PDF) Comparative Trials of Elite Swedish and UK Biomass Willow Varieties. researchgate.net.

Lindegaard, K., R. I. Parfitt, G. Donaldson, and T. Hunter. 2001. “Comparative Trials of Elite Swedish and UK Biomass Willow Varieties.” *Aspects of Applied Biology* 62: (Biomass and Energy Crops II). (PDF) Comparative Trials of Elite Swedish and UK Biomass Willow Varieties. researchgate.net.

Lindroth, A., and A. Båth. 1999. “Assessment of Regional Willow Coppice Yield in Sweden on Basis of Water Availability.” *Forest Ecology and Management* 121, no. 1: 57–65. https://doi.org/10.1016/S0378-1127(98)00556-8.

Macalpine, W. J. 2019. “Identifying Drought Tolerant Short Rotation Coppice Willows.” Doctoral dissertation, Harper Adams University. https://hau.repository.guildhe.ac.uk/id/eprint/17661.

Macalpine, W. J., I. F. Shield, L. R. E. Walsh, et al. 2011. “Rothamsted Research Willow (*Salix* spp.) Breeding Programme: Multi-Site Yield Trial Results 2011.” *Biomass and Energy Crops IV. Aspects of Applied Biology* 112: 263–270.

Macalpine, W. J., I. F. Shield, L. R. E. Walsh, et al. 2014. *Rothamsted Research Willow (Salix spp.) Breeding Programme; Multi-Site Yield Trial Results 2014*. ETA-Florence Renewable Energies. https://doi.org/10.5071/22ndeubce2014-1do.1.2.

McCalmont, J. P., A. Hastings, N. P. McNamara, et al. 2015. “Environmental Costs and Benefits of Growing Miscanthus for Bioenergy in the UK.” *GCB Bioenergy* 9, no. 3: 489–507. https://doi.org/10.1111/gcbb.12294.

McCracken, A. R., and W. M. Dawson. 2003. “Rust Disease (*Melampsora epitea*) of Willow (*Salix* spp.) Grown as Short Rotation Coppice (SRC) in Inter- and Intra-Species Mixtures.” *Annals of Applied Biology* 143, no. 3: 381–393. https://doi.org/10.1111/j.1744-7348.2003.tb00308.x.

McCracken, A. R., W. M. Dawson, and G. Bowden. 2001. “Yield Responses of Willow (*Salix*) Grown in Mixtures in Short Rotation Coppice (SRC).” *Biomass & Bioenergy* 21, no. 5: 311–319. https://doi.org/10.1016/S0961-9534(01)00046-0.

McCracken, A. R., J. P. Moore, L. R. E. Walsh, and M. Lynch. 2010. “Effect of Planting Vertical/Horizontal Willow (*Salix* spp.) Cuttings on Establishment and Yield.” *Biomass & Bioenergy* 34, no. 12: 1764–1769. https://doi.org/10.1016/j.biombioe.2010.07.008.

McKenzie, F. 2011. *The Potential of Short Rotation Coppice (SRC) Willow (Salix L.) as a Biomass Crop in Orkney* Available from ProQuest Dissertations & Theses A&I. ProQuest. https://search.proquest.com/docview/2296375369.

Milner, S., R. A. Holland, A. Lovett, et al. 2015. “Potential Impacts on Ecosystem Services of Land Use Transitions to Second-Generation Bioenergy Crops in GB.” *GCB Bioenergy* 8, no. 2: 317–333. https://doi.org/10.1111/gcbb.12263.

Mola-Yudego, B., and P. Aronsson. 2008. “Yield Models for Commercial Willow Biomass Plantations in Sweden.” *Biomass & Bioenergy* 32, no. 9: 829–837. https://doi.org/10.1016/j.biombioe.2008.01.002.

Nordborg, M., G. Berndes, I. Dimitriou, A. Henriksson, B. Mola-Yudego, and H. Rosenqvist. 2018. “Energy Analysis of Willow Production for Bioenergy in Sweden.” *Renewable and Sustainable Energy Reviews* 93: 473–482. https://doi.org/10.1016/j.rser.2018.05.045.

Nordh, N.-E. 2005. “Long Term Changes in Stand Structure and Biomass Production in Short Rotation Willow Coppice.” Doctoral dissertation, Swedish University of Agricultural Sciences. https://core.ac.uk/download/pdf/11694963.pdf.

Pacaldo, R. S., T. A. Volk, and R. D. Briggs. 2013. “Greenhouse Gas Potentials of Shrub Willow Biomass Crops Based on Below- and Aboveground Biomass Inventory Along a 19-Year Chronosequence.” *Bioenergy Research* 6: 252–262. https://doi.org/10.1007/s12155-012-9250-y.

Powlson, D. S., A. B. Riche, and I. Shield. 2005. “Biofuels and Other Approaches for Decreasing Fossil Fuel Emissions From Agriculture.” *Annals of Applied Biology* 146, no. 2: 193–201. https://doi.org/10.1111/j.1744-7348.2005.040056.x.

Pribyl, D. W. 2010. “A Critical Review of the Conventional SOC to SOM Conversion Factor.” *Geoderma* 156, no. 3: 75–83. https://doi.org/10.1016/j.geoderma.2010.02.003.

Pyatt, G., R. Duncan, and J. Fletcher. 2001. “An Ecological Site Classification for Forestry in Great Britain.” https://cdn.forestresearch.gov.uk/2001/03/fcbu124.pdf.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

R Studio Team. 2020. *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/.

Richard, B., G. M. Richter, M. Cerasuolo, and I. Shield. 2019. “Optimizing the Bioenergy Water Footprint by Selecting SRC Willow Canopy Phenotypes: Regional Scenario Simulations.” *Annals of Botany* 124, no. 4: 531–542. https://doi.org/10.1093/aob/mcz006.

Roback, P., and J. Legler. 2021. *Applied Generalized Linear Models and Multilevel Models in R*. Retrieved from: Beyond Multiple Linear Regression. Chapman and Hall/CRC.

Rothamsted Research. 2022. *Accelerating Willow Breeding and Deployment (AWBD) Final Report From a Lot 1 Phase 1 Project for the BEIS Biomass Feedstocks Innovation Programme*. BEIS Retrieved From: Accelerating Willow Breeding and Deployment—Phase 1 Report.

Rothamsted Research. 2023. “Growers Guide to Short Rotation Coppice Willow (SRCw) Varieties for Biomass.” https://www.rothamsted.ac.uk/sites/default/files/Project_files/SRC%20Willow%20Variety%20Selection.pdf.

Saxton, K. E., W. J. Rawls, J. S. Romberger, and R. I. Papendick. 1986. “Estimating Generalized Soil-water Characteristics From Texture†.” *Soil Science Society of America Journal* 50: 1031–1036. https://doi.org/10.2136/sssaj1986.03615995005000040039x.

Schlepphorst, R., H. Hartmann, and D. Murach. 2017. “Yield of Fast-Growing Tree Species in Northeast Germany: Results of Research With Experimental Plots (2006 to 2015)/Versuchsflachen Mit Schnellwachsenden Baumarten in Brandenburg—Ertragskundliche Ergebnisse fur 2006 Bis 2015.” *Landbauforschung* 67, no. 2: 93+. https://link.gale.com/apps/doc/A535420840/AONE?u=anon~20ace797&sid=googleScholar&xid=a7902160.

Schmidt-Walter, P. 2020. “Evaluation of Environmental Impacts of Short Rotation Coppice With Regard to the Amount and Quality of Groundwater Recharge.” Doctoral dissertation, Georg-August-University Göttingen. https://doi.org/10.53846/goediss-8186.

Shield, I., W. Macalpine, S. Hanley, and A. Karp. 2014. “Breeding Willow for Short Rotation Coppice Energy Cropping.” In *Industrial Crops*, 67–80. Springer New York. https://doi.org/10.1007/978-1-4939-1447-0_4.

Shmueli, G. 2010. “To Explain or to Predict?” *Statistical Science* 25, no. 3: 289–310. https://doi.org/10.1214/10-STS330.

Smith, P., and J. R. Porter. 2018. “Bioenergy in the IPCC Assessments.” *GCB Bioenergy* 10, no. 7: 428–431. https://doi.org/10.1111/gcbb.12514.

Stolarski, M. J., S. Szczukowski, J. Tworkowski, J. Tworkowski, M. Krzyżaniak, and D. Załuski. 2019. "Willow Production During 12 Consecutive Years—The Effects of Harvest Rotation, Planting Density and Cultivar on Biomass Yield." *Global Change Biology* 11, no. 4: 635–656. https://doi.org/10.1111/gcbb.12583.

Strobl, C., A. L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution." *BMC Bioinformatics* 8: 25. https://doi.org/10.1186/1471-2105-8-25.

Tallis, M. J., E. Casella, P. A. Henshall, et al. 2013. "Development and Evaluation of ForestGrowth-SRC a Process-Based Model for Short Rotation Coppice Yield and Spatial Supply Reveals Poplar Uses Water More Efficiently Than Willow." *GCB Bioenergy* 5, no. 1: 53–66. https://doi.org/10.1111/j.1757-1707.2012.01191.x.

Tubby, I., and A. Armstrong. 2002. "Establishment and Management of Short Rotation Coppice. Forestry Commission Practice Note FCPN7." https://cdn.forestresearch.gov.uk/2022/02/fr_bec_establishment_and_management_of_src_fcpn7_2002.pdf.

Valentine, J., J. Clifton-Brown, A. Hastings, P. Robson, G. Allison, and P. Smith. 2011. "Food vs. Fuel: The Use of Land for Lignocellulosic 'Next Generation' Energy Crops That Minimize Competition With Primary Food Production." *GCB Bioenergy* 4, no. 1: 1–19. https://doi.org/10.1111/j.1757-1707.2011.01111.x.

Valentine, J., C. J. Duller, M. Hinton-Jones, et al. 2009. "The Development of Sustainable Heat and Power Fuelled by Biomass From Short Rotation Coppice in Wales." Aberystwyth University Report of the Helyg i Gymru/Willow for Wales 2004-2008 Project. 92 pp. cdn.forestresearch.gov.uk/2022/02/willow_for_wales_2009.pdf.

Verwijst, T., A. Lundkvist, S. Edelfeldt, J. Forkman, and N. Nordh. 2012. "Effects of Clone and Cutting Traits on Shoot Emergence and Early Growth of WIllow." *Biomass and Bioenergy* 37: 257–264. https://doi.org/10.1016/j.biombioe.2011.12.004.

Vigl, F., and B. Rewald. 2014. "Size Matters?—The Diverging Influence of Cutting Length on Growth and Allometry of Two *Salicaceae* Clones." *Biomass & Bioenergy* 60: 130–136. https://doi.org/10.1016/j.biombioe.2013.11.020.

Volk, T. A., L. P. Abrahamson, C. A. Nowak, L. B. Smart, P. J. Tharakan, and E. H. White. 2006. "The Development of Short-Rotation Willow in the Northeastern United States for Bioenergy and Bioproducts, Agroforestry and Phytoremediation." *Biomass and Bioenergy* 30, no. 8: 715–727. https://doi.org/10.1016/j.biombioe.2006.03.001.

Volk, T. A., B. Ballard, D. J. Robison, and L. P. Abrahamson. 2004. "Effect of Cutting Storage Conditions During Planting Operations on the Survival and Biomass Production of Four Willow (*Salix* L.) Clones." *New Forests* 28, no. 1: 63–78. https://doi.org/10.1023/B:NEFO.0000031334.86593.4e.

Volk, T. A., J. P. Heavey, and M. H. Eisenbies. 2016. "Advances in Shrub-Willow Crops for Bioenergy, Renewable Products, and Environmental Benefits." *Food and Energy Security* 5, no. 2: 97–106. https://doi.org/10.1002/fes3.82.

Welc, M., A. Lundkvist, and T. Verwijst. 2017. "Effects of Cutting Phenology (Non-Dormant Versus Dormant) on Early Growth Performance of Three Willow Clones Grown Under Different Weed Treatments and Planting Dates." *Bioenergy Research* 10, no. 4: 1094–1104. https://doi.org/10.1007/s12155-017-9871-2.

Welc, M., A. Lundkvist, and T. Verwijst. 2018. "Effects of Propagule Phenology (Non-Dormant Versus Dormant) and Planting System (Vertical Versus Horizontal) on Growth Performance of Willow Clones Grown Under Different Weeding Regimes." *Bioenergy Research* 11, no. 3: 703–714. https://doi.org/10.1007/s12155-018-9929-9.

Ziety, J. J., E. Olba-Ziety, M. J. Stolarski, et al. 2022. "Legal Framework for the Sustainable Production of Short Rotation Coppice Biomass for Bioeconomy and Bioenergy." *Energies* 15, no. 4. https://doi.org/10.3390/en15041370.

Zitzmann, F., and M. Rode. 2021. "Short-Rotation Coppice Managed According to Ecological Guidelines—What Are the Benefits for Phytodiversity?" *Forests* 12, no. 5: 646. https://doi.org/10.3390/f12050646.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Correlation diagram between numerical variables used in random forest analysis. Positive numbers indicate a positive relationship. Negative numbers indicate a negative correlation. The bigger the circle the higher the correlation. **Figure S2:** Annual yield (Mg DM ha^{-1} $year^{-1}$) of all the genotypes and mixtures included in the willow SRC global dataset. Plot has been divided in two for better visualisation. Standard error bars are shown as vertical black bars inside the green bars. **Table S1:** Metadata collected on *Salix* spp. from literature review and personal communications. It includes columns with the variables and associated units, number of parameters or levels for categorical variables and a brief description. **Table S2:** gcbb70069-sup-0003-TableS1-S2.docx. Table including all the models and associated values for the four situations considered. Cumulative subset including three soil profiles with and without missing values. Average subset including only the upper soil profile with and without missing values. **Data S1:** Global SRC willow dataset R code.