Remote Sensing Applications: Society and Environment 37 (2025) 101494

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

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

# Towards a framework for monitoring crop productivity in agroforestry parklands of the Sudano-Sahel using Sentinel-1 and 2 time series

Julianne Oliveira [a,*], Martin Karlson [b], Abraham Sotongo Ouédraogo [c], Hugues Roméo Bazié [c], Madelene Ostwald [b]

[a] Department of Crop Production Ecology, Swedish University of Agricultural Sciences, 90183, Umeå, Sweden
[b] Department of Thematic Studies/Environmental Change and Centre for Climate Science and Policy Research, Linköping University, 58183, Linköping, Sweden
[c] Laboratoire Biosciences, Unité de Formation et Recherche en Sciences de la Vie et de la Terre, Université Joseph KI-ZERBO, 03 BP 7021, Ouagadougou, Burkina Faso

ARTICLE INFO

*Keywords:*
Crop productivity
Agroforestry
Satellite remote sensing
Sentinel 1–2
Parkland

ABSTRACT

The agroforestry parklands in the Sudano-Sahelian zone are of critical importance for food security, but face several challenges in terms of changes in climate and land use. The ability to systematically monitor crop productivity in these systems is therefore of importance for both informing land management policies and studying long-term trends. This study, conducted in two different agroecological areas in southern and central Burkina Faso covering two climate-wise very contrasting years (2020–2021), is an initial step to designing a system based on satellite remote sensing that enables national-scale monitoring of crop productivity. In these two sites, we collected large field datasets of crop productivity (150 plots) for use in model training and validation. The main assessments focused on how to best process and combine remote sensing data sources, including time series from the Sentinel-1 and Sentinel-2 satellite systems, as well as soil properties, elevation and tree cover. Other key focuses were evaluating different regression modelling algorithms (multilinear and machine learning) and clarifying the potential benefits of performing the modelling in specific geographic regions and years or if the modelling can be generalized. Overall, the results show that accurate estimates of crop productivity are achievable using the proposed modelling framework, with encouragingly high $R^2$ (0.49–0.82) and low root mean square errors (11.80–19.35%). Sentinel-2 was the most important data source, but our results also demonstrate the potential of Sentinel-1, which has the benefit of not being affected by clouds. Another encouraging aspect is that the results were stable both between the years, which differed significantly in terms of rainfall and crop productivity, and between the sites that are characterized by contrasting crop compositions. This study shows that the development of a national-level crop monitoring system in Burkina Faso or countries with similar environmental conditions is within reach.

* Corresponding author.
*E-mail address:* julianne.oliveira@slu.se (J. Oliveira).

https://doi.org/10.1016/j.rsase.2025.101494
Received 25 May 2024; Received in revised form 8 February 2025; Accepted 20 February 2025
Available online 23 February 2025

## 1. Introduction

The Sudano-Sahelian zone stretches from the Atlantic coast to the Red Sea, with the Sahel running as a dryland band north of the semi-arid Sudanian zone. This region faces severe challenges, including widespread poverty, population growth, high reliance on rainfed agriculture with relatively low output and armed conflicts. Erratic rainfall and high pressure of human activities on land use systems, such as cropland and grazing areas, make the population highly susceptible to short-term food scarcity and long-term environmental degradation. Recent climate projections for the region anticipate increasing temperatures and decreasing rain (Trisos et al., 2023), further aggravating the situation. Specifically, the livelihood of the people in the region is highly dependent on locally produced resources such as food, fodder, and fuel, which all are components of the net primary production (NPP). Estimates from Sahel show a 41% increase in the demand for NPP from 2000 to 2010 (Abdi et al., 2014) or an annual increase of 2.2%, with higher demand for food (20.4%) followed by fodder (16.7%) and fuel (5.5%). This reliance on the local supply of NPP is also pressured by a demographic trend of high population growth, putting additional demand on natural resources. From 2010 to 2022, the population increased from 312 to 429 million (37% increase) in West Africa as a whole, while Burkina Faso increased from 16 to 22 million (40%) over the same period (UNISTAT, 2023).

Smallholder farming is the dominant livelihood strategy in the Sudano-Sahel and produces up to as much as 90% of the food consumed (Carletto et al., 2015; Morton, 2007; Samberg et al., 2016). Furthermore, agroforestry in parklands is the dominating agricultural practice (Bayala et al., 2014) and, consequently, the primary NPP source. Parklands are traditional land-use systems where trees, grass and crops coexist (Boffa, 1999). Specifically, farmers retain preferred trees in the fields because they positively impact soil quality and erosion along with non-timber forest products, including food and fodder. In addition, parklands supply other ecosystem services in terms of soil porosity, groundwater recharge and shade (Bayala et al., 2014; Ilstedt et al., 2016; Sinare and Gordon, 2015). However, the tree canopy cover in the fields is usually kept below 20% to limit the competition with crops for light, water and nutrients (Karlson et al., 2023).

Many countries dealing with food insecurity, including those in the Sudano-Sahelian zone, lack capacity for systematic monitoring their agricultural resources (Burke and Lobell, 2017; Carletto et al., 2015). The lack of crop statistics is especially prevalent for smallholder farms in Africa. Reliable and up-to-date statistics and forecasts on crop productivity represent key information to support effective agricultural management and decision-making controlling food availability (Bégué et al., 2020). While traditionally being the main method for acquiring such information, field surveys are time-consuming and labour intensive, as well as associated with high costs and considerable limitations in spatiotemporal coverage and reliability. However, geographical datasets with wide spatiotemporal coverage at the pixel or field scale can help understand causes of spatial variability in crop productivity and yield gaps related to management practices, such as fertilizer application, selection of crop types and varieties and modifications of the tree cover in the case of agroforestry systems, particularly in parkland systems (Bayala et al., 2015; Burke and Lobell, 2017; Karlson et al., 2023; Leroux et al., 2020). On regional and national scales, the availability of these datasets can uncover long-term trends in agricultural productivity related to various causes, including changes in climate, soil characteristics, management practices and armed conflicts. Such information can also help identify areas to prioritize for land and soil rehabilitation (Lobell et al., 2009), and for the targeted distribution of food aid.

Satellite remote sensing represents an alternative for acquiring spatially distributed crop statistics with the benefits of covering large areas at high repeatability and low cost (Lobell, 2013). In fact, a large body of research stretching several decades back in time has demonstrated that satellite remote sensing can provide accurate estimates of crop productivity (Wu et al., 2023), especially in regions characterized by large monocrop fields, for example, the Great Plains in the United States (Weiss et al., 2020). However, such favourable conditions do not apply to smallholder landscapes common throughout the tropics, where the fields are usually small and irregular in shape, often planted with mixed crops characterized by different structural traits, phenology and contain trees (Burke and Lobell, 2017). An additional challenge is the extensive cloud cover during the growing season and the relatively low spatial resolution of the free-available data, which previously constrained the use of optical satellite systems, such as MODIS and Landsat. Developments in remote sensing systems have helped overcome many of these limitations, including the Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 optical systems becoming operational and providing high resolution data in both the spatial and temporal domains (Duncan et al., 2015). The Sentinel-2 system represented a game changer for crop productivity mapping, especially in African smallholder landscapes. This has been demonstrated in Tanzania and Kenya (Azzari et al., 2017; Burke and Lobell, 2017; Jin et al., 2019), Burkina Faso (Karlson et al., 2020; Leroux et al., 2019), Senegal (Leroux et al., 2020) and Mali (Lambert et al., 2018; Lobell et al., 2020b). Less focus has been directed to the potential of Sentinel-1 to complement optical systems in the mapping of crop productivity, as well as the suitability of using image in-season composition methods, which are contributions of this work. Moreover, this work focused on a substantial field data collection, covering areas with different climates, growing conditions, and cropping systems.

For crop productivity estimations, remote sensing data can be used by applying different models, such as light use efficiency (LUE) models and empirical models. LUE models are based on the theory that plant productivity is a function of solar radiation absorbed by green leaves and the photosynthetic rate of the plant, which is controlled by environmental factors, in particular, availability of water and nutrients (Potter et al., 1993). On the other hand, empirical models rely on regression techniques for relating remotely sensed predictor variables to the target variable to be estimated. While the LUE modelling approach is attractive from a theoretical perspective, there are important challenges related to the acquisition of the required spatial datasets as well as model parameterization. Empirical modelling is less complex compared to LUE models in terms of input data requirements and parameterization. They can be developed using both traditional statistical regression analyses that assume a linear relationship between the predictors and the target variable, and more recent techniques, including machine learning and artificial intelligence (Morais et al., 2021). Traditional

statistical regressions are easily repeatable and capable of future prognoses. In contrast, a potential benefit of using machine learning for regression is the ability to characterize complex non-linear relationships in the data that may improve the predictive performance of the model. However, a higher demand for input data for model training may make it less suitable, especially in situations where the availability of field data is a limitation. Well-established machine learning algorithms, such as Random Forest and Support Vector Machine, and several other up-to-date methods have crossed over the barriers for accurately predicting crop productivity in different types of landscapes (Muruganantham et al., 2022). Moreover, empirical modelling approaches using temporally dense time-series data combining data from different sensors, e.g., optical and SAR, together with other spatial datasets on landscape characteristics, such as soil properties and elevation, may represent a powerful alternative to assess spatial-temporal dynamics of complex crop systems like the agroforestry parklands in Sudano-Sahel.

In this study, we evaluated the performance of statistical and machine learning modelling methods to estimate crop productivity in two agroforestry parkland regions in Burkina Faso using a multi-sensor approach. Comprehensive field data was collected in smallholder agroforestry farm systems over two consecutive years (2020 and 2021) and two regions (Cassou and Saponé communes) in Burkina Faso. The main aim was to provide key insights into how a framework for national-scale crop productivity monitoring can be designed using remote sensing. To test our framework, we calibrated the models using predictor variables based on Sentinel-1 and Sentinel-2 time series over two growing seasons (2020 and 2021), as well as datasets representing different soil properties, tree canopy cover and elevations.

The key questions we wanted to address were.

i) How do statistical and machine learning modelling methods perform in terms of estimating the crop productivity in two contrasting parkland areas?
ii) How can site-specific and annually trained models be applied to multiple years and sites without calibration?
iii) What is the most suitable period for acquiring satellite data and which is the preferable method for temporal composition during the growing season to estimate crop productivity?
iv) What is the contribution of the different remotely sensed predictor variables in the estimation of crop productivity?

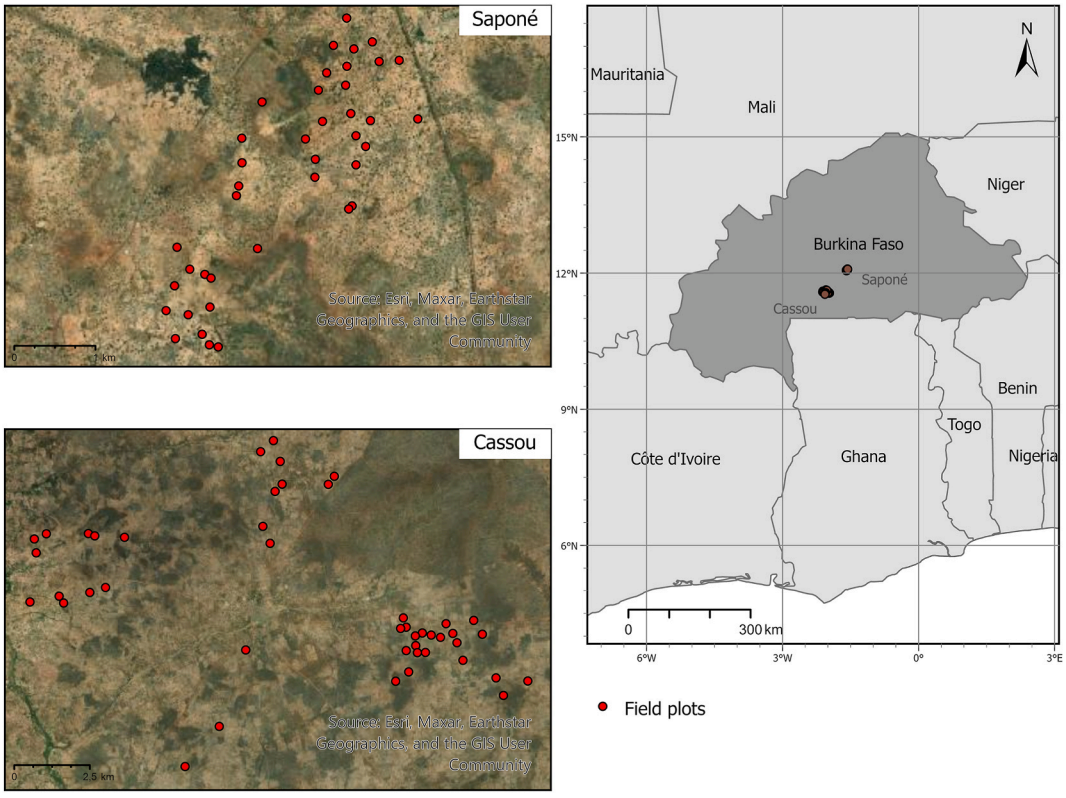


**Fig. 1.** Map of the study areas in the Saponé and Cassou communes and locations of the field plots. Source of the background satellite images: World Imagery Metadata, Maxar (Vivid) image, 0.6 m of spatial resolution acquired in 2018–05-01 and 2022-03-11 for Saponé and Cassou, respectively. More information about the background satellite images can be found in: https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9.

## 2. Material and methods

### 2.1. Study area

The study was conducted in two agroforestry parkland landscapes in the rural communes of Saponé (12°04′48″ N, 1°34′00″ W) and Cassou (11°34′50″ N, 2°02′57″ W), located in the central and southern parts of Burkina Faso (Fig. 1), respectively. Field data were collected at these sites in 2020 and 2021. Saponé and Cassou are smallholder landscapes dominated by rainfed fields interspersed by fallow areas used for grazing. The fields' size varied, ranging from 1 to 10 ha (Paré et al., 2010). The main crops in Saponé are sorghum (*Sorghum bicolor* (L.) Moench), pearl millet (*Pennisetum glaucum* (L.) R. Br.) and legumes, primarily cowpea (*Vigna unguiculata)* (Karlson et al., 2020). In Cassou, sorghum and maize (*Zea mays* L.*)* are the main crops (Foli and Abdoulaye, 2016). Both Cassou and Saponé farmers rely on manual labour to work the land and on local organic nutrient sources, including compost and manure, to replenish soil fertility. Specifically, organic household waste is composted with dung from poultry, goats and sheep, and applied to the fields during land preparation. In addition, cattle manure is deposited in the fields when livestock graze on crop residues during the dry season.

Cassou belongs to the south-Sudanian ecological zone where rainfall is relatively high, ranging between 900 and 1200 mm annually. The rainy season is unimodal and usually lasts from May through September (Etongo et al., 2015). Saponé has a drier climate with a mean annual rainfall of 730 mm, but the inter-annual variability is high (Bayala et al., 2008). The rainy season in Saponé generally starts in June and ends in October. Rainfall conditions for both sites in 2020 and 2021 are described in Fig. 2. The between-site difference in climate is also reflected in the composition of the tree cover. In Saponé, mean tree canopy cover is around 15% and the most common species include *Vitellaria paradoxa*, *Parkia biglobosa* and *Lannea microcarpa* (Karlson et al., 2015). The tree cover in Cassou is generally denser and more diverse in terms of species composition, with key tree species including *Combretum* sp., *Deuterium microcarpum* and *Vitellaria paradoxa.* The main soil types in both Cassou and Saponé are silt-clay cambisols, sandy lixisols, and loamy ferric luvisols (Bayala et al., 2002; Bazié et al., 2012) with very low nutrient contents. Furthermore, soil quality and fertility are degrading in both areas, representing a major challenge for the local farmers. An important reason for this is the reduced fallow periods resulting from high pressure on productive lands (Söderström et al., 2003).

### 2.2. Reference data

#### 2.2.1. Crops

The reference data used in the modelling include crop-cuts collected in 20 × 20 m plots located within agricultural fields. These fields were selected considering a spectrum of soil characteristics, topographical features, prior crop rotations, accessibility to the area, as well as alignment with farmers regarding methodology and data utilization. The plots included fields with different crop combinations, productivity levels and tree canopy cover, where agreements with the local landowners were organized. The centre and four corner point locations of the plots were geo-referenced using a Garmin Oregon 700 GPS device. In 2020, 79 plots were installed and monitored over the growing seasons in the Cassou and Saponé sites (Table 1). From these 79 plots, 61 (24 in Cassou and 37 in Saponé) were revisited in 2021 and 17 new plots (16 in Cassou and 1 in Saponé) were installed, totaling 71 plots in 2021. The new plot locations in 2021 were the result of farmers' decision to change the crop system in the field or because of damages that had occurred in the plot between the years, as well as governmental restrictions related to security issues.

The sowing of the plots was carried out between mid-June and July at both sites. In Saponé, three different crops were cultivated in the sampled fields: sorghum, pearl millet and cowpea. The plots included fields with single crops (millet and sorghum) and mixed crops (millet + cowpea and sorghum + cowpea). In contrast, in Cassou, only single crop fields were sampled, including fields cultivated with maize and sorghum (Table 1).

We quantified aboveground biomass (AGB) of the crops in all plots after harvest (late October to early November). In each main plot, we randomly allocated 5 to 9 square subsamples (2 × 2 m; see the subsample scheme in Fig. A1 of Appendix A) and cut the aboveground material for biomass estimation. The material was then sent to the laboratory, separated by plant components, dried for

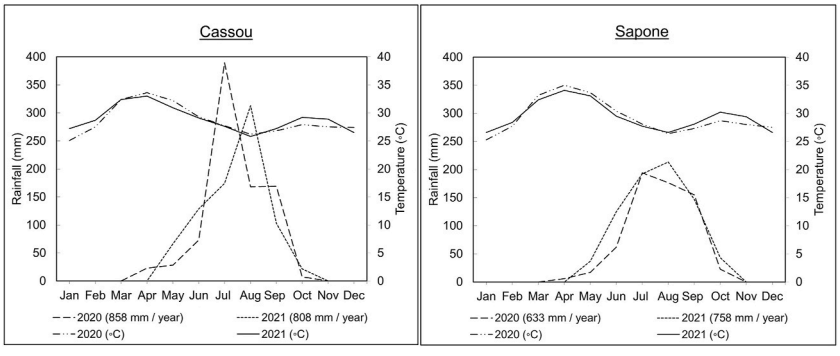


**Fig. 2.** Annual rainfall distribution and temperature for 2020 and 2021 in Cassou and Saponé. Rainfall data from Agence Nationale de la météorologie du Burkina Faso and temperature data from https://en.tutiempo.net/.

**Table 1**
Number of field plots by commune and crop type for the years 2020 and 2021.

| Commune | Crop | Num. Plots 2020 | Num. Plots 2021 |
|---|---|---|---|
| Cassou | Maize | 18 | 16 |
| | Sorghum | 23 | 21 |
| Saponé | Millet | 9 | 7 |
| | Millet-cowpea | 10 | 10 |
| | Sorghum | 10 | 7 |
| | Sorghum-cowpea | 9 | 10 |
| Total | | 79 | 71 |

24 h in an oven at 60 °C, and weighed. The drying under controlled conditions was done to ensure more consistency and precision in the data, avoiding the fluctuations between harvests due to water content. The total AGB represents all dried plant components, including the grain. For maize, millet and sorghum, the plant components included ears, grain and straw, and for cowpea, the plant components are pods, grain and residues.

While most previous efforts of remote sensing-based crop production estimation have focused on the crop yield, we used AGB as the target variable for several reasons. Firstly, in this area and in this land use system, all crop components contribute to the livelihood, including as fodder and for soil enrichment. Secondly, remotely sensed vegetation indices are sensitive to crop properties such as density and chlorophyll content, both of which are more closely related to AGB than to the yield. Lastly, there is usually a strong relationship between AGB and grain yield, generally operationalized by the harvest index (Hay, 1995). This strong relationship has been demonstrated in Burkina Faso, including for millet and sorghum ($R^2 = 0.8$; Karlson et al., 2020), which makes it possible to accurately predict yield from AGB using local knowledge about cropping practices and environmental conditions.

#### 2.2.2. Tree cover

The plots were in fields with different tree cover densities, which may influence the ability to model crop productivity from the remote sensing data, especially in cases of mixed pixels (representing crops and trees) in optical data and surface roughness in SAR data. For example, Leroux et al. (2020) showed that their remote sensing-based model produced more accurate crop productivity estimates when including tree density as a predictor variable. Thus, we quantified the tree canopy cover within each plot through manual digitization using the base-map imagery from 2018 available in ArcGIS Pro 2.8 as reference in Saponé. In Cassou, the base-map imagery in ArcGIS Pro was deemed too old (years 2011–2013) and the tree cover digitization was instead done using high-resolution imagery available in Google Earth (year 2020) as reference. The tree canopy cover in the sampled fields ranged between circa 0 and 15% in Cassou (mean = 5%, std = 7%) and between 0 and 28% (mean = 9%, std = 7%) in Saponé. The influence of tree cover on the crop production estimation was evaluated by including it as a predictor variable in the crop productivity models.

### 2.3. Satellite data and processing

The satellite remote sensing datasets comprised the Sentinel-1 and Sentinel-2 missions, two twin polar-orbiting satellite systems from the Copernicus Earth Observation program of the European Space Agency (ESA). The Sentinel-1 system has a dual-polarization C-band SAR instrument and consists of a two-satellite constellation (Sentinel-1A, launched in 2014; and Sentinel-1B; launched in 2016). Each Sentinel-1 scene provides a single (HH or VV) and dual polarization (HH and HV or VV and VH) at one of three spatial resolutions (10, 25, or 40 m), depending on the acquisition mode (IW: Interferometric Wide Swath; EW: Extra Wide Swath; or SM: Strip Map). The mission acquires images with a revisit time of 12 (one satellite) or 6 (two satellites) days at the equator.

The Sentinel-2 mission is also a twin-satellite constellation (Sentinel-2, launched in 2015; and Sentinel-2B, launched in 2017). Both satellites carry the Multi Spectral Instrument (MSI) sensor that collects 13 spectral bands covering visible to shortwave infrared (SWIR) wavelengths. The revisit time of Sentinel-2 is 10 (one satellite) or 5 (two satellites) days at the equator. The spectral bands have a spatial resolution of 10 m (blue, green, red and near infrared wavelengths), 20 m (four red edge bands and two SWIR bands), or 60 m (aerosol, water vapour and cirrus bands).

Time series of Sentinel-1 C-band SAR data and Sentinel-2 multispectral images were acquired in the years 2020 and 2021 over the two sites, between the months of June and November. This period covers the growing season of the crops at the two sites. The satellite data were accessed and processed in the Google Earth Engine (GEE) cloud computing platform (Gorelick et al., 2017), using python API and the "geemap" package (Wu, 2020). In GEE, the available Sentinel-1 scenes have already been pre-processed for thermal noise removal, radiometric calibration and terrain correction using Shuttle Radar Topography Mission (SRTM) 30 digital elevation model (DEM) and tools from the Sentinel-1 Toolbox provided by ESA (S1TBX - ESA Sentinel-1 Toolbox, http://step.esa.int). However, despite the pre-processed corrections, border artefacts, speckle noise and topographic effects may remain and can compromise their usage. Here we used the VV and VH polarization bands from the IW acquisition mode in ascending orbit direction, with 10 m spatial resolution. The border noise was removed based on the incidence angle and backscattering values of neighbouring pixels in the border region (Hird et al., 2017; Stasolla and Neyt, 2018). For the speckle noise removal, we applied the Refined Lee filter to each individual scene independently (Lee et al., 1999). We further mitigated topographic effects by radiometric terrain normalization using a simplified angular-based correction approach (Hoekman and Reiche, 2015) according to the volume and surface scattering models implemented in GEE by Vollrath et al. (2020), using SRTM DEM to represent surface elevation. The border noise correction, speckle filtering and radiometric terrain normalization removal was performed using the comprehensive GEE framework proposed by Mullissa

et al. (2021) that is available at https://github.com/adugnag/gee_s1_ard (Accessed on September 14, 2022).

For the Sentinel-2 data, we used the Level-2A product available in the GEE data catalog, which is atmospherically corrected with the Sen2Cor processor (Main-Knorn et al., 2017). One limitation of the Sentinel-2 images during growing periods is the high occurrence of clouds. To deal with that without unnecessarily losing the high temporal resolution of these data by defining a cloud percentage threshold for the whole scenes, we masked out the clouds and shadows pixels by using a function (s2cloudness) based on the cloud probability band from Sentinel-2 available in GEE. In this function, clouds are identified using the Sentinel-2 cloud probability band and shadows are defined by the cloud projection intersection with low-reflectance near-infrared (NIR) pixels. We masked out the clouds and shadows of each image in the time-series during the growing period (June–November) in both years using a cloud probability threshold of 40% (values greater than the threshold are considered clouds). We also used 1 km of maximum distance from cloud edges to search for cloud shadows, a NIR reflectance threshold of 0.15 (lower reflectance values than this threshold are considered as potential cloud shadow), and a buffer distance of 100 m to expand the edge of the identified cloud object. We chose these parameters after several iterative tests over the two sites during the study period. A detailed description of this function is available on the GEE developer's community webpage (https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless). Following cloud and cloud shadow masking, we selected the bands in the visible (B2, B3 and B4), red edge (B5, B6, B7), NIR (B8) and SWIR (B11 and B12) spectral regions, which are particularly sensitive to vegetation properties. The bands with 20 m spatial resolution (B5, B6, B7, B8A, B11, B12) were resampled to 10 m using the nearest neighbor interpolation method.

We used satellite-based products of soil properties and elevation as additional information sources to characterize the sites and, consequently, potentially improve the crop AGB estimation. Soil properties were extracted from the iSDASoil digital soil map (https://isda-africa.com/isdasoil), which provides soil information for over 20 different physical and chemical properties for the African continent with 30 m spatial resolution based on comprehensive compilations of soil samples and modelling with Earth observation data (Hengl et al., 2021; Miller et al., 2021). Specifically, we used the sand, clay and silt mean content (% unit) and the mean organic

**Table 2**
Vegetation indices used for crop biomass estimation in Cassou and Saponé.

| Vegetation Index (VI) | Equation[a] | Reference |
|---|---|---|
| **Sentinel-1** | | |
| Polarization Ratio (PR) | $\frac{\sigma^0_{VH}}{\sigma^0_{VV}}$ | Blaes et al. (2006) |
| Normalized Difference Polarization Index (NDPI) | $\frac{\sigma^0_{VV}-\sigma^0_{VH}}{\sigma^0_{VV}+\sigma^0_{VH}}$ | (Cao et al., 2008; Nasirzadehdizaji et al., 2019) |
| Radar Vegetation Index (RVI) | $\frac{4*\sigma^0_{VH}}{\sigma^0_{VV}+\sigma^0_{VH}}$ | (Kim and Zyl, 2009; Nasirzadehdizaji et al., 2019) |
| Normalized Product Polarization Index (NPI) | $\frac{\sigma^0_{VV}*\sigma^0_{VH}}{\sigma^0_{VV}+\sigma^0_{VH}}$ | Adapted from Li et al. (2012) |
| Dual Polarization SAR Vegetation Index (DPSVI) | $\sigma^0_{VH}*\frac{\left[\begin{array}{l}\left(\sigma^0_{VV_{max}}*\sigma^0_{VH}-\sigma^0_{VV}*\sigma^0_{VH}+{\sigma^0_{VH}}^2\right)\\+\left(\sigma^0_{VV_{max}}*\sigma^0_{VV}-\sigma^0_{VH}*\sigma^0_{VV}+{\sigma^0_{VV}}^2\right)\end{array}\right]}{\sqrt{2}*\sigma^0_{VV}}$ | Periasamy (2018) |
| Dual Polarization SAR Vegetation Index m (DPSVIm) | $\frac{{\sigma^0_{VV}}^2+\sigma^0_{VV}*\sigma^0_{VH}}{\sqrt{2}}$ | dos Santos et al. (2021) |
| **Sentinel-2** | | |
| Normalized Difference Vegetation Index (NDVI) | $\frac{B8-B4}{B8+B4}$ | Rouse et al. (1974) |
| Enhanced Vegetation Index (EVI) | $\frac{g*(B8-B4)}{B8+C_1*B4-C_2*B2+L}$ | Huete et al. (1997) |
| Soil Adjusted Vegetation Index (SAVI) | $\frac{B8-B4}{B8+B4+L}(1+L)$ | Huete (1988) |
| Soil Adjusted Vegetation Index 2 (SAVI2) | $\frac{B8}{B4+\frac{b}{a}}$ | Major et al. (1990) |
| Modified Soil-Adjusted Vegetation Index (MSAVI) | $\frac{2*B8+1-\sqrt{(2*B8+1)^2-8*(B8-B4)}}{2}$ | Qi et al. (1994) |
| Triangular Vegetation Index (TVI) | $\frac{120*(B8-B3)-200*(B4-B3)}{2}$ | Broge and Leblanc (2001) |
| Red Edge NDVI 1 (RENDVI1) | $\frac{B8-B5}{B8+B5}$ | Forkuor et al. (2018) |
| Red Edge NDVI 2 (RENDVI2) | $\frac{B6-B4}{B6+B4}$ | Forkuor et al. (2018) |
| Normalized Multi-band Drought Index (NMDI) | $\frac{B8-(B11-B12)}{B8+(B11-B12)}$ | Wang and Qu (2007) |

[a] Terms of the equation are explained, as follows: $\sigma^0_{VV}$ is the VV backscattering coefficient (in linear power units). $\sigma^0_{VH}$ is the VH backscattering coefficient (in linear power units). $\sigma^0_{VV_{max}}$ is the maximum value of the VV backscatter coefficient over the study region (in linear power units). $B8$: NIR band; $B4$: Red band; $B2$: Blue band; $g$: adjusted gain factor ($g = 2.5$); $C_1$ and $C_2$: factors for atmosphere effect correction ($C_1 = 6$ and $C_2 = 7.5$); $L$: factor for soil effect correction ($L = 1$); $a$: soil line slope; $b$: soil line intercept; $B3$: Green band; $B5$: Red-edge 1 band; $B6$: Red-edge 2 band; $B11$: SWIR 1 band; and $B12$: SWIR 2 band.

carbon maps (g/kg), all representing 0–20 cm of soil depth. We selected these soil properties for being less prone to change over short time intervals and spatial scales. Detailed information about the methodology and accuracy assessment of the iSDASoil data is provided in Hengl et al. (2021). Elevation data were obtained from the Copernicus Digital Elevation Model (GLO-30), available with 30 m spatial resolution (https://doi.org/10.5270/ESA-c5d3d65). The iSDAsoil and GLO-30 datasets were both acquired through GEE data collection, followed by resampling to 10 m using the nearest neighbor interpolation method to allow integration with the Sentinel time-series. The elevation and soil maps showing the variations of these properties over the two communes are presented in Fig. A2 of Appendix A.

### 2.4. Satellite-based predictor variables

Overall, 33 predictor variables from five different sources (Sentinel-1, Sentinel-2, soil properties, elevation and tree canopy cover) were selected to be included in the crop AGB modelling. These include 15 vegetation indices (VIs) calculated from Sentinel-1 and Sentinel-2 time series as predictor variables in the crop AGB modelling (Table 2). These VIs were selected according to their previous usage in vegetation assessment and potential relevance to quantify crop AGB at different spatiotemporal scales, particularly in African regions (Ibrahim et al., 2021; Jin et al., 2019; Karlson et al., 2020; Lambert et al., 2018). For Sentinel-1 indices, the polarization ratio-based index (PR), the Normalize Difference Polarization Ratio (NDPR) and the Normalization Polarization Index (NPI) were used based on their potential to maximize the sensitivity to vegetation targets (Nasirzadehdizaji et al., 2019; Veloso et al., 2017; Vreugdenhil et al., 2018), as well as the Radar Vegetation Index (RVI), intended to assess vegetation growth by characterizing volume scattering (Kim et al., 2012). We also used the Dual Polarization SAR Vegetation Index (DPSVI), proposed to be sensitive to the physical scattering of ground targets by integrating co- and cross-polarized parameters (Periasamy, 2018). Lastly, we used the DPSVIm, which is a simple adaptation of the DPSVI to deal with seasonality of more dense vegetation (dos Santos et al., 2021). All Sentinel-1 VIs were calculated within GEE.

For Sentinel-2 VIs we included the Normalized Difference Vegetation Index (NDVI), which is one of the most widely used VIs for vegetation analysis based on the normalized difference between the NIR and red bands (Rouse et al., 1974). Further, we included the Enhanced Vegetation Index (EVI), which is a modified version of NDVI in which factors to correct for soil and atmospheric effects have been added (Huete et al., 1997). We also included the Soil Adjusted Vegetation Index (SAVI) and two of its adapted versions (SAVI2 and Modified Soil-Adjusted Vegetation Index; MSAVI) since they all were designed to reduce effects from soil conditions (Huete, 1988; Qi et al., 1994). We also included the Triangular Vegetation Index (TVI), which adds the influence of the leaf pigments using the green band (Broge and Leblanc, 2001). Furthermore, we included two VIs based on red edge (RE) bands (RENDVI1 and RENDVI2; Forkuor et al., 2018), as well as the Normalized Multi-band Drought Index (NMDI), which uses the reflectance behaviour of the SWIR bands to enhance the sensitivity to soil and vegetation moisture (Wang and Qu, 2007). Sentinel-2 VIs were computed using the "eeMont" python package available in GEE (Montero, 2021), with the exception of RENDVI1 and RENDVI2.

We also included the individual bands from Sentinel-2 (B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12) and the incidence angle (the incident radar beam according to the surface), the vertical transmit-vertical receive polarization (VV) and the vertical transmit-horizontal receive polarization (VH) information from Sentinel-1. The VV and VH were considered as backscattering coefficients in linear power units, respectively, $\sigma^0_{VV}$ and $\sigma^0_{VH}$. Soil properties at 20 cm depth (sand, clay, silt, and organic carbon), elevation and tree canopy cover (%) were also included as predictor variables for their potential to account for spatial variability in crop growth conditions.

The predictor variables based on Sentinel-1 and Sentinel-2 were computed for each available date in the period of analysis (June to November 2020 and 2021). However, due to the extensive cloud cover, there were several occurrences of missing information from masked pixels. To overcome this lack of data, we calculated temporal statistical compositions for each month (June, July, August, September, October and November) and the growing season (all available images between June and November). For this approach, we performed a comparison of the predictive potential between the mean, median and maximum statistical compositions. Pixel values of the predictor variables were extracted according to the delineation of the field plots, i.e., $2 \times 2$ pixels. Since satellite-based predictor variables presented different levels of magnitude, we standardized them by max–min procedure (ranging between 0 and 1) for further calibrating the models (Shahhosseini et al., 2021).

### 2.5. Modelling approaches and scenarios

To identify the most effective modelling framework for estimating crop AGB, we tested two main modelling techniques: multiple linear regression (MLR) and three machine learning (ML) algorithms. MLR was based on the parametric Ordinary Least Squares (OLS) technique and for the ML approach, we tested the performance of the nonlinear methods Multivariate Adaptive Regression Splines (MARS) and Support Vector Machine (SVM), as well as the tree-based algorithm Random Forest (RF). These ML methods were chosen due to their wide usage, well-documented availability, and good suitability to estimate AGB of several vegetation types in different edaphic-climatic regions (Morais et al., 2021). It is important to mention that these modelling methods (MLR and ML) are not meant to be exhaustive, but a starting point covering simple to complex methods.

#### 2.5.1. Multilinear regression models

MLR is a widely used statistical method for attributing linear relationships of predictor variables ($X_1$, $X_2$, …, $X_k$) to perform predictive analysis of the response variable ($Y$). Due to its simplicity and common use, it has been the standard procedure used for

quantitative agricultural applications of remote sensing for the last decades (e.g., Monteiro et al., 2022; Olson and Olson, 1986). Considering all the satellite-based predictor variables to define the best MLR model, we used the backward stepwise method, which is a feature selection method based on a p-value threshold (Kuhn and Johnson, 2019). MLR models take the general form of Eq. (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \quad (1)$$

where, $Y$ is the response variable, $X_1, X_2, \ldots X_k$ are predictors variables, $\beta_0, \beta_1, \beta_{02} \ldots \beta_k$ are parameters, and $\varepsilon$ is the error component. Based on this general equation, the backward stepwise regression method uses a sequence of steps to allow potential predictor variables (features) to leave the regression model one-at-a-time. In the first step, it considers a model that contains all the predictor variables, and then gradually eliminates them to maintain the best model with the highest coefficient of determination ($R^2$) and minimise the mean squared deviation from the regression model while maintaining the significant variables. Finally, this procedure converges to a subset of predictor variables used in the model.

Because of the potential of high correlation between some predictor variables, we considered MLR models as a combination of variables grouped by, as follow (Eq. (2)):

$$Biomass_k = f(S2_{Band}, S2_{Index}, S1, Soil, Elev, Tree\%) \quad (2)$$

where, $Biomass_k$ is the AGB of crops in the field plot (sample), $S2_{Band}$ is a band from Sentinel-2 (10 possible variables), $S2_{Index}$ is an index from Sentinel-2 (9 possible variables), $S1$ is a band or index from Sentinel-1 (9 possible variables), *Soil* is soil-related data (4 possible variables), *Elev* is the elevation data (1 variable) and *Tree*% is the tree cover data. Then we applied the backward stepwise regression method for each model resulting from the possible combination of these groups of predictor variables, totaling 3240 models ($10 \times 9 \times 9 \times 4 \times 1$). For this approach, we used the method "leapBackward" through the "caret" package (Kuhn, 2008) available in the R software package (R Development Core Team, 2020).

### 2.5.2. Machine learning

The ML algorithms chosen here, MARS, RF and SVM, each follow different procedures. A common constraint for some ML methods is the negative effect of including irrelevant predictor variables in the model (Kuhn and Johnson, 2019). To overcome this effect, before calibrating each ML model, we performed a feature selection procedure to use just the most important case-specific predictor variables nested for each group of analysis. To do so, we applied the Recursive Feature Elimination algorithm (Guyon et al., 2002) from the "caret" package in R software using the RF function with a 10-fold cross validation process repeated five times.

MARS is a flexible modelling algorithm for high-dimensional data (Friedman, 1991). It verifies nonlinear relationships in the data by assessing piecewise functions and selecting the most significant basis functions from the model. It can identify additive contributions of different multivariable interactions. MARS was implemented using the "caret" package (Kuhn, 2008), and applying the "gcvEarth" method from the "earth" package (Milborrow et al., 2024).

RF is one of the most popular machine learning algorithms. It is a decision tree model based on the ensemble learning of multiple random trees that can be used for both regression and classification problems (Breiman, 2001). For regression problems, the algorithm outputs the average of all trees built. For RF implementation, we used the "ranger" method available in the "ranger" package (Wright and Ziegler, 2017) also from "caret" package.

SVM is designed to find an optimal hyperplane that maximizes the distance between observations (samples or classes) for forming groups with similar characteristics (Cortes and Vapnik, 1995). It is a widely used and robust nonlinear ML algorithm. SVM was also implemented using the "caret" package, where a radial basis kernel was adopted by using the "svmRadial" method from the "kernlab" package (Karatzoglou et al., 2004).

All the models were developed through a grid search approach to optimize tuning parameters with different combinations of values. For MARS, the tuning parameter "degree" (product degree) was evaluated with values ranging from 1 to 2. For RF, the model was built with 500 trees and the tuning parameters included "mtry" (randomly selection of predictor; ranging from 1 to the total number of predictor variables), "splitrule" (splitting rule; either "variance" or "extratrees"), and "min.node.size" (minimal node size; set to 5). For SVM, the parameters tuned were "C" (cost; with a sequence of 20 values ranging from 1 to 2) and "sigma" (spanning from 0.0 to 0.05 in increments of by 0.01).

**Table 3**
Modelling scenarios used to evaluate the framework for crop productivity monitoring.

| Scenario | Type | Description |
|---|---|---|
| 1 | General model | Saponé and Cassou sites for year 2020 |
| 2 | General model | Saponé and Cassou sites for year 2021 |
| 3 | General model | Saponé and Cassou sites for both years 2020 and 2021 |
| 4 | Site-specific model | Saponé site for year 2020 |
| 5 | Site-specific model | Saponé site for year 2021 |
| 6 | Site-specific model | Cassou site for year 2020 |
| 7 | Site-specific model | Cassou site for year 2021 |
| 8 | Site-specific model | Saponé site for both years 2020 and 2021 |
| 9 | Site-specific model | Cassou site for both years 2020 and 2021 |

### 2.5.3. Modelling scenarios

We used several scenarios when performing the AGB modelling to compare prediction performance and address advantages, limitations and generalization potential of the framework. Specifically, modelling was first done using only field data from the individual sites, hereafter referred to as "site-specific" models. We also created models where the field data from two sites were combined, and these are referred to as "general" models. Site-specific and general models were created both for the individual years (2020 and 2021) and when field data from both years were aggregated. Overall, this amounted to nine modelling scenarios, including three general and six site-specific models (Table 3). Another important analysis in this modelling framework was to identify the most suitable period for extracting satellite predictor variables over the growing period, i.e., the month or season when the correlation between satellite and crop AGB is the strongest. Thus, we compared the performance of the AGB estimation using the predictor variables grouped by statistics composition of the period (single months and season) for each modelling scenario.

## 2.6. Evaluation and estimation of the crop production

Fig. 3 illustrates the summary flowchart of the modelling framework used for estimating the crop AGB of the agroforestry parklands (Sections 2.2 to 2.6). Models for crop AGB estimation were calibrated using the satellite-based predictor variables and modelling

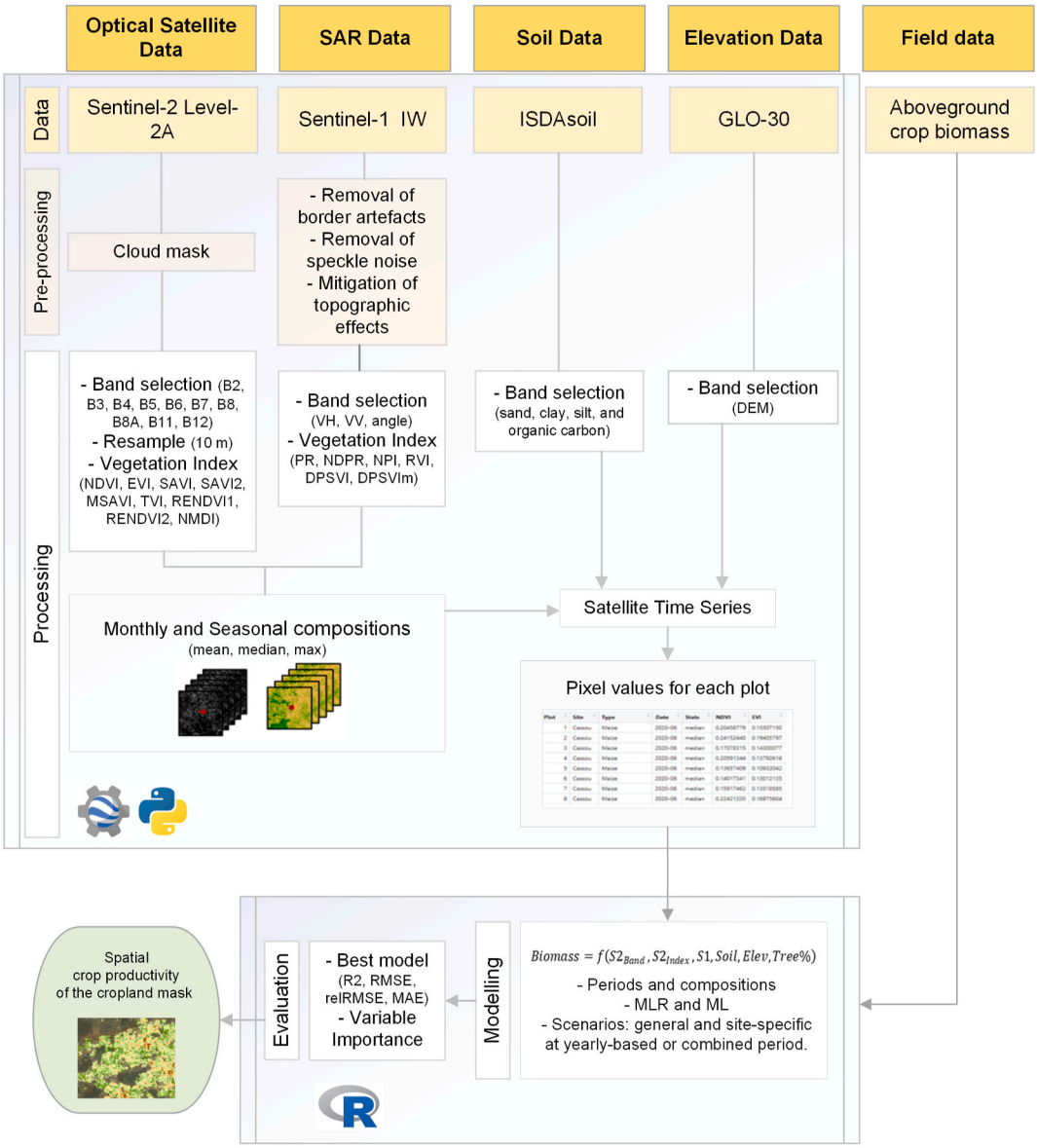


**Fig. 3.** Flowchart of the main steps used to build and evaluate the performance of the modelling to estimate biomass.

approaches described above (Sections 2.4 and 2.5). Given the limited reference sample size (field data; Table 1), we used a 10-fold cross validation (10-fold CV) procedure repeated five times for model calibration and calculation of accuracy metrics. This approach was chosen for its relatively low computational cost across varying the number of observations in the modelling scenarios, while the methodology remains feasible for future updates with larger datasets.

The performance of the models was evaluated according to a number of accuracy metrics, including the coefficients of determination ($R^2$), the root mean square error (RMSE, ton $ha^{-1}$; Eq. (3)), relative root mean squared error (relRMSE, %; Eq. (4)) and mean absolute error (MAE, ton/ha; Eq. (5)).

$$RMSE\ (ton\ ha^{1}) = \sqrt{\sum_{i=1}^{N} \frac{(\widehat{Y}_i - Y_i)}{N}} \tag{3}$$

Where, $\widehat{Y}_i$ is the estimated crop AGB at sample $i$, $Y_i$ is the observed crop AGB at sample $i$, and $N$ is the total number of samples.

$$relRMSE(\%) = 100 \times \frac{RMSE}{(\max Y_{obs} - \min Y_{obs})} \tag{4}$$

Where, max $Y_i$ and min $Y_i$ are the maximum and minimum observed crop AGB, respectively, and $N$ is the total number of samples.

$$MAE\ (ton\ ha^{1}) = \frac{1}{N} \sum_{i=1}^{N} |\widehat{Y}_i - Y_i| \tag{5}$$

Where, $\widehat{Y}_i$ is the estimated crop AGB at sample $i$, $Y_i$ is the observed crop AGB at sample $i$, and $N$ is the total number of samples.

The relative importance of predictor variables to the respective AGB models was also quantified to gain further insights into the modelling process. We used the "varImp" function of the "caret" package in R, which calculates the variable importance of the training dataset according to the specific methods of each modelling algorithm.

The best modelling approach in each scenario was then used to estimate the spatial crop productivity for Saponé and Cassou in 2020 and 2021. A cropland mask based on Potapov et al. (2022) was used to create the maps at the commune level.

## 3. Results

### 3.1. Field data summary

Crop cuts from the field plots in Cassou and Saponé provided input to the AGB modelling and Fig. 4 shows the range of the total AGB at the sites and in the two years. The plots in Cassou generally had higher mean AGB compared to Saponé for all crop types. It is also apparent that the growing conditions differed between the two years, with higher crop AGB in 2020 compared to 2021. The mean AGB was around 50% lower in 2021 for both sites and most crop types. The only exception was the sorghum plots in Cassou, with a reduction in AGB of around 80% in 2021. A reason for this inter-annual variability in AGB can be explained by the one-month delay in the onset of the rainy season in 2021 (Fig. 2). Tables A1 and A2 (Appendix A) show detailed statistics of the crop AGB for both sites in 2020 and 2021, respectively.

After revising the field campaign protocols and data exploration, and talking with field experts in Burkina Faso, we decided to remove 12 plots from our crop-cut dataset, which were considered outliers according to their AGB values; more precisely, the plots with AGB placed on the first and third quartile of the frequency distribution. In the end, 72 and 66 plots were used in the modelling and

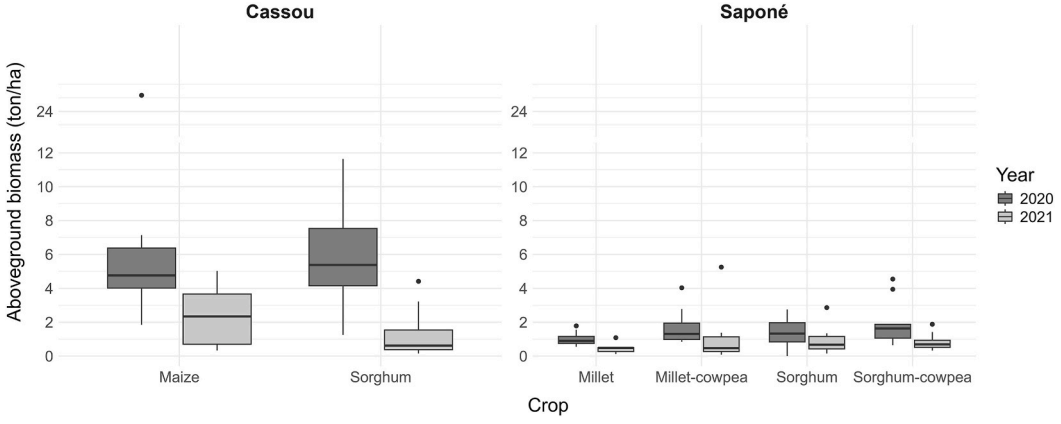


**Fig. 4.** Boxplots of the distribution of the observed aboveground biomass for the years 2020 and 2021 in Saponé and Cassou communes. The dots represent outlier values that were removed prior to the modelling.

analysis in 2020 and 2021, respectively. Even though we recognise that not always observations out of statistical quartiles should be removed from the analysis, our case demonstrated that these values could be related to errors in data tabulation.

### 3.2. Sentinel-1 and 2 observations

Sentinel-1 and Sentinel-2 predictor variables extracted for the dates available over the two years show the variability over the sites, including growth stages and crop systems. Fig. 5 shows an example of the NDVI (Sentinel-2) and PR (Sentinel-1) values and their variation over this period. For Sentinel-2, the cloud cover was more extensive in August, so this month was excluded from the modelling analysis. Also, according to the NDVI, it is possible to identify the growth peak of the vegetation around September–October. PR values showed higher variations between the plots all over the growth period and more available images in Cassou than in Saponé.

Working with monthly and seasonal mean, maximum and median compositions was an alternative to overcome the lack of satellite data, as well as evaluate the influence of the different growth stages on the total AGB. The NDVI and PR compositions illustrate an example of spectral and backscattering changes between sites, crops, months and years (Fig. 6). The NDVI shows the general trend over the years, with higher values in September and October, similar as in Fig. 5. Also, PR compositions show higher variability between

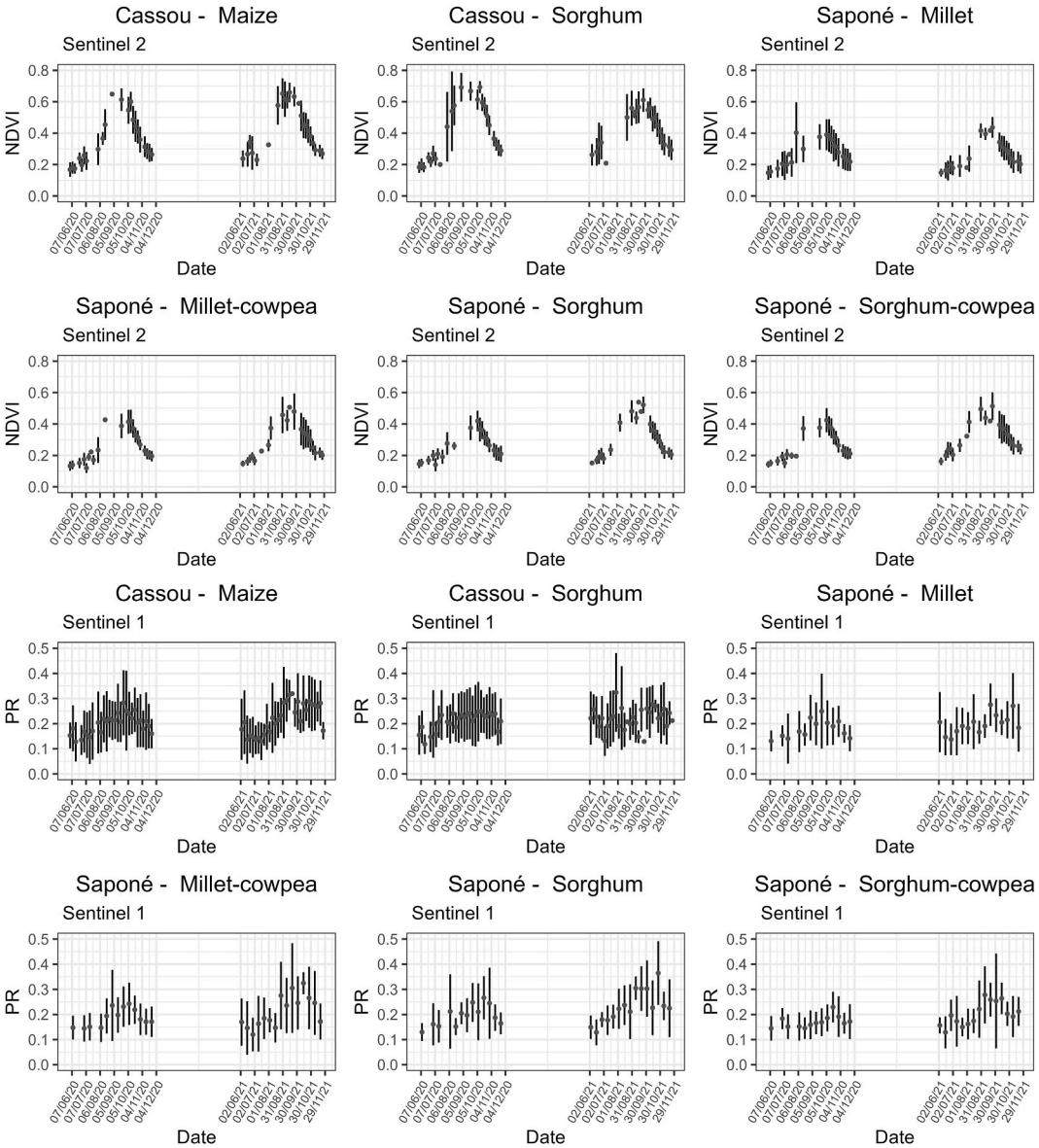


**Fig. 5.** Normalized Difference Vegetation Index (NDVI) and Polarization Ratio (PR) average (grey circles) and standard deviation values (grey vertical bars) acquired from, respectively, Sentinel-2 and Sentinel-1 satellites for the plots in Cassou and Saponé communes over the two growing seasons in 2020 and 2021.

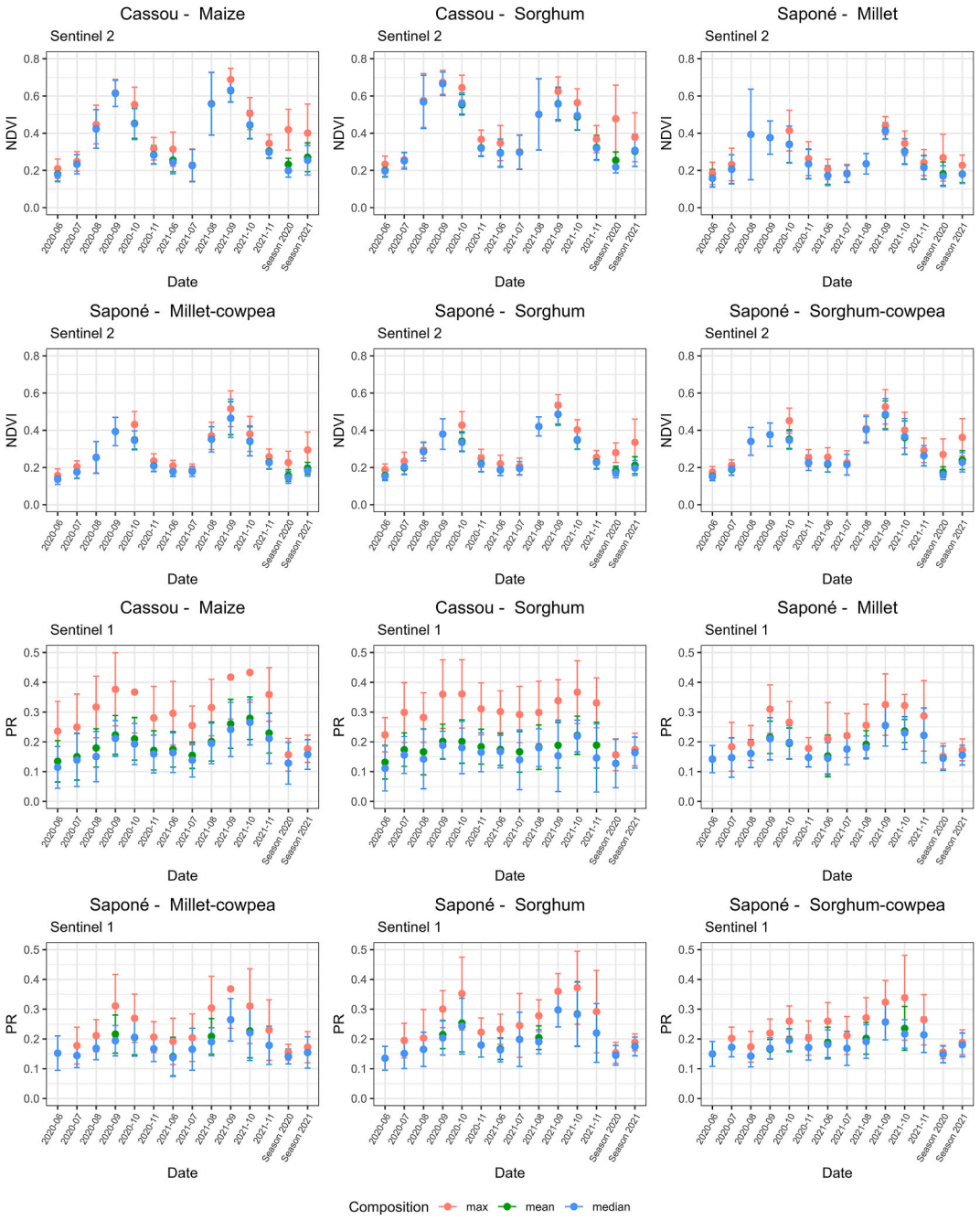


**Fig. 6.** Normalized Difference Vegetation Index (NDVI) and Polarization Ratio (PR) monthly and seasonal compositions (maximum, mean and median) acquired from, respectively, Sentinel-2 and Sentinel-1 satellites for the plots in Cassou and Saponé communes over the two growing seasons in 2020 and 2021.

plots, but with a slight trend of peak values around September–October. Comparatively to Fig. 5, using time-series compositions (Fig. 6) showed potential to fill missing information when estimating crop AGB.

### 3.3. *Performance of the different modelling scenarios*

The modelling was done using nine different scenarios (Table 3) based on how the field data were used, including general models where the field data from both sites were combined and site-specific models. Models were also developed using field data from each year separately, and when both years were combined. The results from these nine modelling scenarios are presented in Table 4 and

**Table 4**
Most accurate combinations of modelling algorithms, composition period and method for the nine modelling scenario. MARS - Multivariate Adaptive Regression Splines, MLR – Multi Linear Regression, RF – Random Forest, SVM – Support Vector Machine.

| Scenario | Model type | Site | Year | Algorithm | Best month | Composition | $R^2$ | RMSE (tons/ha) | relRMSE (%) | MAE (tons/ha) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | General | Both | 2020 | MARS | September | Median | 0.64 | 1.83 | 16.54 | 1.39 |
| | | | | MLR | September | Median | 0.76 | 1.42 | 12.79 | 1.12 |
| | | | | RF | September | Max | 0.75 | 1.44 | 13.04 | 1.11 |
| | | | | SVM | September | Median | 0.72 | 1.51 | 13.64 | 1.24 |
| 2 | General | Both | 2021 | MARS | September | Max | 0.76 | 0.59 | 12.07 | 0.43 |
| | | | | MLR | September | Mean | 0.63 | 0.69 | 13.93 | 0.54 |
| | | | | RF | September | Max | 0.68 | 0.69 | 14.00 | 0.52 |
| | | | | SVM | September | Median | 0.70 | 0.58 | 11.80 | 0.44 |
| 3 | General | Both | Both | MARS | September | Mean | 0.69 | 1.28 | 15.12 | 0.97 |
| | | | | MLR | September | Mean | 0.64 | 1.36 | 16.00 | 1.08 |
| | | | | RF | September | Mean | 0.74 | 1.13 | 13.37 | 0.87 |
| | | | | SVM | September | Mean | 0.76 | 1.10 | 12.93 | 0.79 |
| 4 | Site-specific | Saponé | 2020 | MARS | September | Mean | 0.67 | 0.55 | 24.88 | 0.47 |
| | | | | MLR | September | Mean | 0.72 | 0.43 | 19.52 | 0.37 |
| | | | | RF | October | Mean | 0.72 | 0.40 | 18.17 | 0.35 |
| | | | | SVM | September | Mean | 0.68 | 0.43 | 19.41 | 0.35 |
| 5 | Site-specific | | 2021 | MARS | September | Max | 0.66 | 0.45 | 34.24 | 0.38 |
| | | | | MLR | September | Median | 0.75 | 0.27 | 20.46 | 0.23 |
| | | | | RF | September | Max | 0.82 | 0.31 | 19.35 | 0.26 |
| | | | | SVM | September | Max | 0.77 | 0.31 | 22.91 | 0.25 |
| 6 | Site-specific | Cassou | 2020 | MARS | September | Median | 0.53 | 2.13 | 20.40 | 1.77 |
| | | | | MLR | October | Median | 0.57 | 1.75 | 16.87 | 1.43 |
| | | | | RF | September | Median | 0.55 | 1.78 | 17.81 | 1.60 |
| | | | | SVM | September | Median | 0.55 | 1.85 | 17.83 | 1.60 |
| 7 | Site-specific | | 2021 | MARS | September | Mean | 0.71 | 0.77 | 15.76 | 0.61 |
| | | | | MLR | September | Max | 0.80 | 0.71 | 14.62 | 0.59 |
| | | | | RF | September | Mean | 0.76 | 0.72 | 14.85 | 0.54 |
| | | | | SVM | September | Mean | 0.77 | 0.70 | 14.31 | 0.58 |
| 8 | Site-specific | Saponé | Both | MARS | October | Median | 0.32 | 0.73 | 26.13 | 0.55 |
| | | | | MLR | October | Median | 0.47 | 0.51 | 18.34 | 0.40 |
| | | | | RF | September | Max | 0.47 | 0.53 | 19.03 | 0.44 |
| | | | | SVM | September | Max | 0.49 | 0.51 | 18.61 | 0.42 |
| 9 | Site-specific | Cassou | Both | MARS | September | Median | 0.63 | 1.57 | 18.72 | 1.29 |
| | | | | MLR | September | Mean | 0.70 | 1.47 | 17.41 | 1.16 |
| | | | | RF | September | Mean | 0.69 | 1.40 | 16.69 | 1.38 |
| | | | | SVM | September | Mean | 0.72 | 1.30 | 15.51 | 1.04 |

they show that the accuracy of the models is high overall, but with important differences. Specifically, $R^2$ and relRMSE ranges from 0.49 to 0.82 and 11.80–19.35%, respectively, when considering the best model alternative in each scenario. Another clear observation in Table 4 is that September is the preferable composition period in most of the scenarios and models. For the site-specific models, using October as the composition period produced the most accurate AGB estimates in 2020 in both Cassou and Saponé. However, the difference to using the September composition in these two cases is marginal in both sites. Lastly, performance of the composition methods differed widely between the year and modelling scenarios, with the mean and median compositions being the most common outcomes in the models. Fig. A3 in Appendix A shows examples of predicted vs. observed scatterplots for the training and validation datasets.

The differences in performance between the four algorithms used for building the models were marginal, especially when comparing MLR, RF and SVM (Fig. 7). On the other hand, MARS produced the least accurate AGB estimates in 7 out the 9 modelling scenarios. The differences between the algorithms can also be demonstrated when calculating the average $R^2$ and relRMSE of the most accurate models for MARS ($R^2 = 0.65$; relRMSE = 18.43%), MLR ($R^2 = 0.70$; relRMSE = 16.33), RF ($R^2 = 0.71$; relRMSE = 15.76) and SVM ($R^2 = 0.70$; relRMSE = 16.0). Fig. 7 further details the differences in model performance when using different composition periods, as well as variability in $R^2$ introduced by the three composition methods and modelling algorithms (indicated by the error bars). A key observation is that using September as the composition period consistently produces the most accurate AGB estimates, but that the differences often are small in several modelling scenarios, especially in relation to using the October composition. The performance improvement using the September composition was slightly larger in 2021 compared to 2020. In addition, the error bars in Fig. 7 show that September produced more stable results in terms of $R^2$, with lower variability introduced by the different modelling algorithms and composition methods.

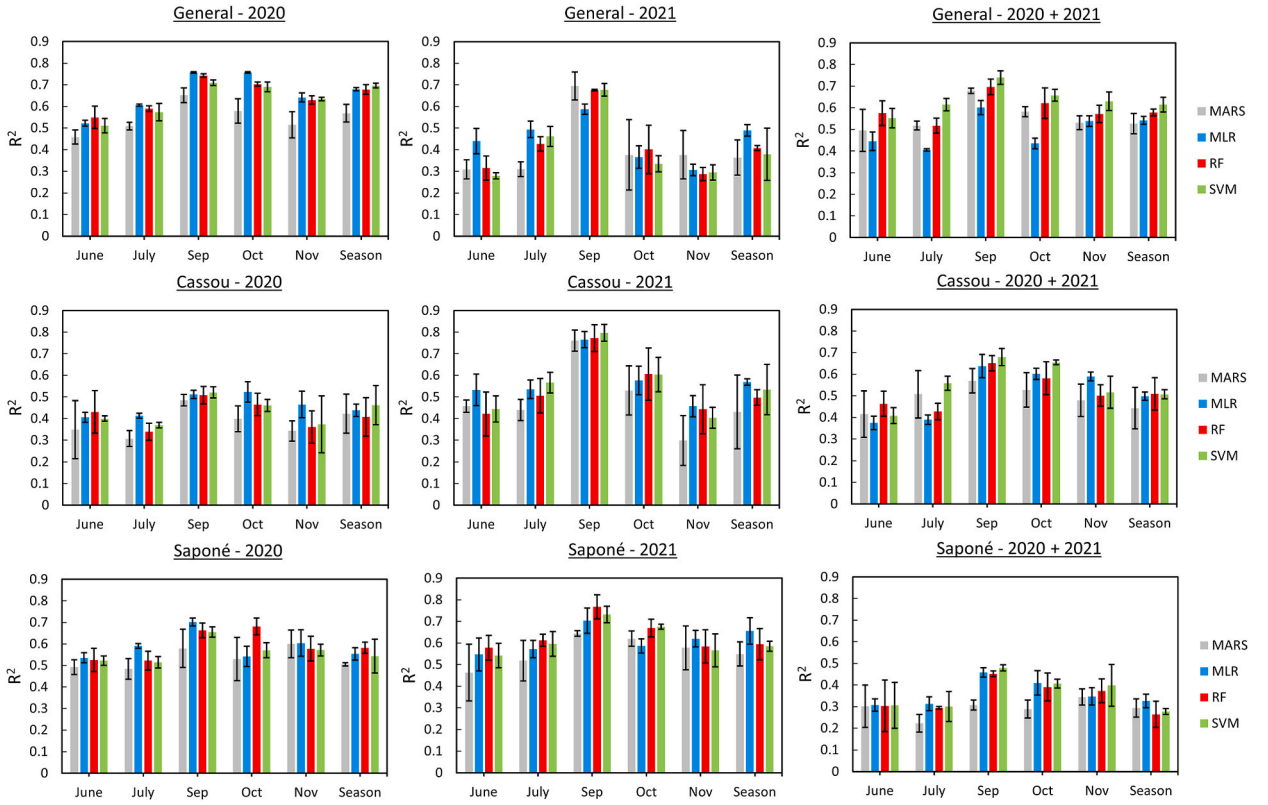


**Fig. 7.** Graph showing variations in model fit ($R^2$) in the nine scenarios when using different modelling algorithms based on different composition periods. The bars present the averaged value of the three composition methods (max, mean, median), whereas the error bars show the standard deviation.

### *3.4. Variable importance*

After the model selection, we quantified the importance of predictor variables in the respective AGB models for both the MLR and ML approaches (Table 5 and Fig. 8, respectively). Significant differences in predictor variable importance were observed between the approaches and scenarios. For example, in the most generalized scenario (both years and sites; scenario 3 in Table 4) and the MLR approach, the Sentinel-2 individual red-edge band (B5) and the soil brightness compensated VI (SAVI) were selected as important variables, together with soil organic carbon and tree cover (Table 5); demonstrating the influence of the landscape components (crop vegetation, soil and trees) to the crop AGB. When considering the most accurate ML approach, more predictor variables were ranked as important, from both Sentinel 1 and 2, as well as soil-related predictors, elevation and tree cover. However, Sentinel-2 variables were more critical, top-ranked by the individual red-edge band (B5) (Fig. 8).

In general, Sentinel-2 predictor variables were important in most scenarios and approaches. However, the addition of Sentinel-1 variables was also relevant in several scenarios, as well as the tree cover, soil and elevation data sources. These results indicate the potential of using satellite and ancillary raster datasets when considering crop monitoring at the national level.

**Table 5**
Predictor variables selected by the multilinear regression (MLR) approach for each scenario (site and year).

| Scenario | Site and period (and number of selected predictor variables) | Predictor ranking |
|---|---|---|
| 1 | Both sites 2020 (4) | B5, EVI, VH, silt |
| 2 | Both sites 2021 (3) | B5, RENDVI1, TC |
| 3 | Both sites combined years (4) | B5, SAVI, SOC, TC |
| 4 | Saponé 2020 (3) | PR, clay, elevation |
| 5 | Saponé 2021 (3) | B3, MSAVI, TC |
| 6 | Cassou 2020 (2) | B6, TVI |
| 7 | Cassou 2021 (4) | B8, RENDVI1, SOC, TC |
| 8 | Saponé combined years (4) | EVI, NL, clay, elevation |
| 9 | Cassou combined years (3) | B5, NMDI, TC |

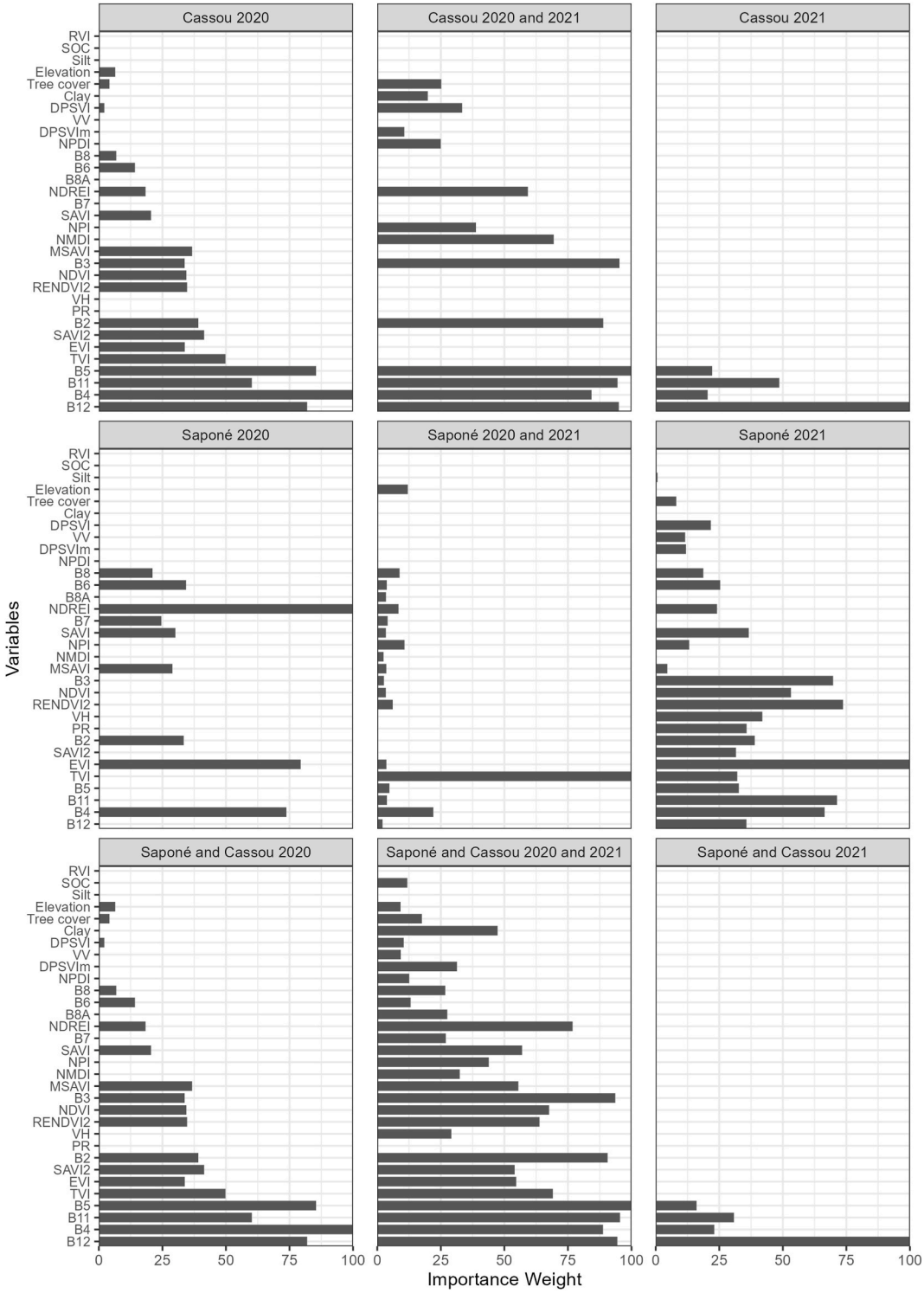


**Fig. 8.** Variable Importance of the predictor variables for the best machine learning (ML) approaches in the modelling scenarios (RF, SVM, SVM, RF, SVM, RF, RF, SVM and SVM, respectively, for the left to right subplots sequentially from upper to bottom part).

### 3.5. Spatial crop productivity

The spatial crop productivity can then be estimated based on these most accurate AGB models and scenarios. For example, the spatial differences in annual crop AGB for 2020 and 2021 can be seen in Fig. 9, where commune-level maps were derived using the MLR described in Table 4 for 2020 and 2021 (Scenarios 1 and 2). MLR was selected for mapping considering its accuracy and simplicity compared to the other methods. The regression equations for the most accurate MLR models in each modelling scenario are available in Table A3 of Appendix A. The maps in Fig. 9 highlight the lower crop production in 2021 compared to 2020, more so in Cassou compared to Saponé. The proportion of cropland was 43% (50820 ha) in Cassou and around 48% (32326 ha) in Saponé. In terms of crop AGB, in Cassou, this amounts to a total of 269350 tons in 2020 (mean = 5.3 tons/ha) and 116888 tons in 2021 (mean = 2.6 tons/ha). For Saponé, the crop AGB totaled 62712 tons in 2020 (mean = 1.9 tons/ha) and 46226 tons in 2021 (mean 1.4 tons/ha).

## 4. Discussion

### 4.1. Interannual and between-site variability in crop productivity

The field data used for calibrating remote sensing models in smallholder production systems is primarily collected either through household surveys or crop-cuts of plots. While crop-cuts provide more reliable data, an obvious tradeoff must be made between quality and quantity. Here, we used a nested plot approach with crop-cuts done in subsections of the main plots to capture spatial variability in crop productivity while limiting workload. Previous research has shown that this approach to field data collection is as effective as whole-plot crop-cuts to calibrate remote sensing-based models (Lobell et al., 2020a). The number of plots used (79 in 2020 and 71 in 2021) was a compromise between capturing variability of rainfalls, sites, soil properties, crop type and productivity, as well as labor costs and is within the range of earlier studies conducted in agroforestry systems in West African countries (Lambert et al., 2018; Leroux et al., 2019, 2020).

A strength of our study is that field data were collected at two contrasting sites (soil types, vegetation, cropping system and rainfall) covering two years. In an ideal situation, field data should have been collected in the drier parts of northern Burkina Faso to capture national coverage. This was the original plan, but unfortunately, we had to refrain from it due to severe security problems. The two sites we could work on differed in both the crop types and crop combinations, as well as in the growth conditions and crop productivity.

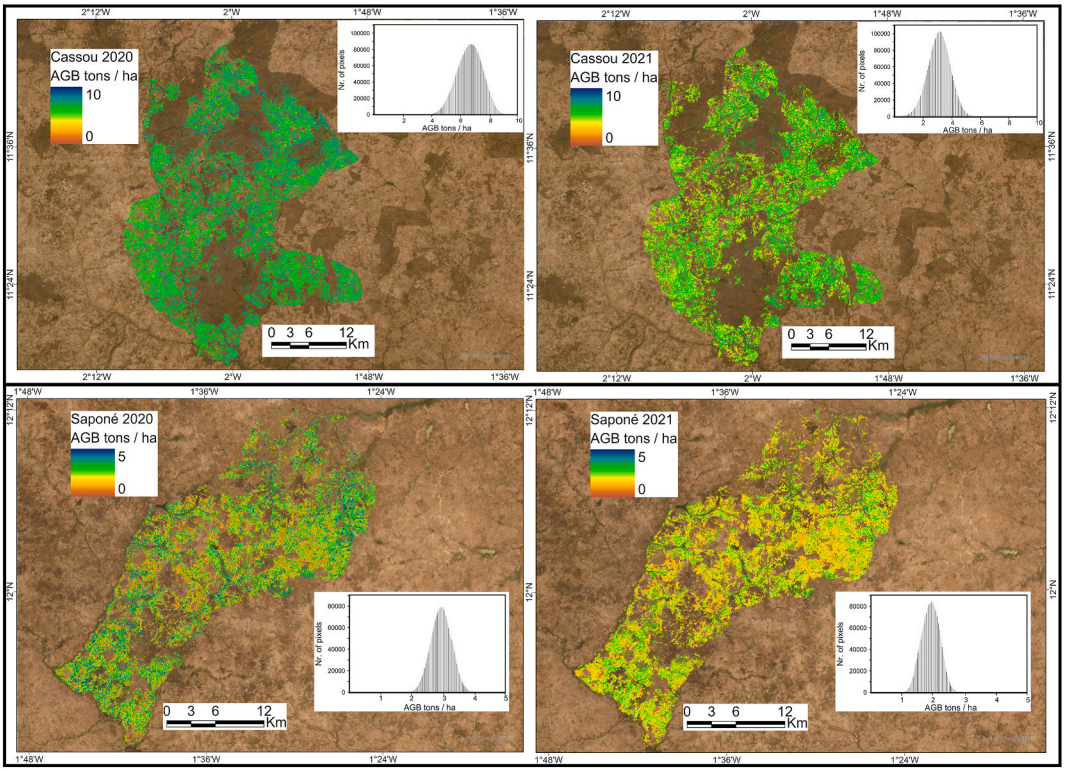


**Fig. 9.** Examples of commune-level maps of annual crop AGB from both sites in 2020 and 2021 based on the most accurate Multi Linear Regression (MLR) models (Table 4 and Table A3 in Appendix A) and a cropland mask (Potapov et al., 2022). These models were trained using the general models based on combined field data from both sites (Scenarios 1 and 2). The histograms show the frequency distribution of pixel-level AGB (tons/ha) estimated by the MLR models.

The crop productivity was higher in Cassou in both years, which was expected due to higher rainfall, higher soil fertility and the dominance of maize. This was especially seen in 2020 when the mean AGB was around 5 tons/ha compared to 1.5 tons/ha in Saponé. The between-site differences were less pronounced in 2021, with mean crop AGB around 2 tons/ha in Cassou and 1 ton/ha in Saponé. The main reason for the lower crop productivity in 2021 was the delay in the rain onset (ca one month) and drought periods that negatively affected the crop growth. The start date and rainfall distribution over the rainy season are key-factors for failure in the Sudano-Sahel region (Oguntunde et al., 2014; Sivakumar, 1988).

Moreover, the focus here was especially on total AGB, since in the smallholder production system in Burkina Faso all parts of the plant are used. Sorghum, cowpea and maize grains are used for human consumption, while the tops and crop residues (stalks and leaves) are used for animal feed. In some cases, the residues are used as cooking fuel (Yaméogo et al., 2014). The relationship between AGB and grain yield was strong overall in both sites and in both years, with $R^2$ ranging between 0.7 and 0.8.

### 4.2. Combined use of Sentinel-1 and Sentinel-2 for estimating single and mixed crop AGB

The combined use of SAR (Sentinel-1) and optical (Sentinel-2) time series for crop productivity modelling is a main contribution of this work. It has not been researched in this type of agricultural system earlier where woody perennial species are grown with annual crops. This aspect, together with potential differences in growing conditions (soil and climate), crop types and combinations, cropping practices, differences in input data for model training, as well as methods for modelling and accuracy assessment, makes it not a straightforward task to make cross-study comparisons. In addition, previous studies have mainly estimated crop yield, whereas we used crop AGB as the target variable. Furthermore, most previous studies have focused on single crop fields, which is an easier case to model with remote sensing data than mixed crop fields with differences in canopy architecture and phenology. With that being said, the results achieved here are promising with relatively high $R^2$ (0.49–0.82) and low relRMSE (11.80–19.35%), and stability between the two years. For example, Lambert et al. (2018) achieved an $R^2$ of 0.48–0.8 and RMSE of 0.5–1.0 ton/ha (based on the full field dataset) for fields with maize, millet and sorghum in Mali, which is quite similar to our results. But Lambert et al. (2018) required a prior classification of the crop types in the field, which can add a layer of uncertainty to the analysis. Also in Mali, but focusing specifically on sorghum, Lobell et al. (2020b) reported a correlation coefficient ($R^2$) of 0.48 between crop-cuts and satellite estimates and an RMSE of 0.3 ton/ha (full field dataset). Two other studies worth mentioning in this context are those by Leroux et al. (2020, 2019) that, similar to us, used cross validation in the accuracy assessments. In their 2019 study, the focus was on maize fields in Burkina Faso with results of $R^2$ of 0.59 and relRMSE of 25.9% when including drought indices and surface soil moisture as predictor variables in the model, in addition to those derived from 250 m MODIS data. Comparable results were reported for estimation of millet productivity in Senegal using Sentinel-2 data and tree density as an additional predictor variable in the model, with $R^2$ of 0.7 and relRMSE of 28% (Leroux et al., 2020).

Studies that have included mixed crop fields in the analysis are fewer in number, with two exceptions. Burke and Lobell (2017) evaluated the influence of mixed crop fields on model performance compared to single crop fields but did not find a significant effect on the explained variance ($R^2 = 0.39$; full field dataset). Similar conclusions were also drawn in our previous study in Saponé, Burkina Faso (2017–2018; Karlson et al., 2020). Model performance in Karlson et al. (2020), however, was considerably poorer than what was achieved in the present study, with $R^2$ between 0.4 and 0.8 and relRMSE between 31 and 66%. The improved performance provides an indication of the potential benefits that can be achieved when including SAR-based predictors, as well as other data sources (e.g., soil, elevation and tree cover) in the crop productivity modelling.

Besides the potential of using Sentinel-1 and 2 products for crop ABG estimations, some drawbacks and pre-processing requirements still need attention. Sentinel-1 and 2 indices used in our approach may not be directly linked to dry AGB, especially compared to their correlation with wet vegetation characteristics, such as pigment and moisture content. Therefore, caution is required when using them without supplementary ancillary data. Our modelling framework includes additional predictable variables, such as soil properties, elevation and tree cover, along with multi-temporal data, which may help capture environmental variations, moisture dynamics, and phenological factors that influence satellite signals. This study focuses on smallholder production systems in Burkina Faso, where all post-harvest parts of the plants are used, as discussed in the previous section, which reinforces the importance of assessing the dry AGB. Furthermore, the correlation between wet and dry AGB supports the transferability of these satellite data sources for estimating dry AGB. Previous studies have also investigated dry AGB and grain yield using satellite data (e.g., Hosseini et al., 2019; Karlson et al., 2020; Leroux et al., 2019), which supports our findings.

In terms of pre-processing requirements, for Sentinel-2, clouds were especially crucial in our study region because the rainy season overlaps with the growing season. We used a cloud mask function (s2cloudness) to remove the cloudy and shadowed pixels. This approach has shown good performance in agricultural areas (Gao et al., 2024) but relies on defining thresholds, such as for cloud probability. In our case, these values (e.g., 40% for cloud probability) were defined after a visual inspection, which may not be replicable to other areas and, therefore, requires a re-analysis when changing locations. Sentinel-1, on the other hand, requires more comprehensive pre-processing, which can limit its use by non-experts. The vegetation structure influences SAR data, and trees in the plots could have affected the signal. However, it was not the case in the plots we used in this study since the tree coverage was low (on average 4.3%) and relatively similar between crop types and sites, even considering that we did not use the same plots both years. However, this issue needs further evaluations when applied on broader scales.

### 4.3. Insights from the modelling framework and mapping crop productivity

We evaluated several scenarios in which models were trained using different subsets of the field data to gain insights to guide the

design of a potential national system for monitoring crop productivity. The comparison between site-specific and general models (both sites combined) indicates the prospective benefits of using an agroecological stratification that reflects regional differences in growing conditions and crop types in the modelling framework, but the results are not conclusive. Such an approach has, for example, shown promise for cropland classification in Mali (Vintrou et al., 2012) because it counteracts regional variability in climate conditions and crop phenology. The small differences in prediction accuracy between the site-specific and general models in terms of the proportion of variance explained for both scenarios and years suggest that the agricultural systems at our two sites are relatively similar at the spatial scale observed by the satellite systems we used despite some differences in crop types and combination in the fields. Even though intercropping of cereals (millet or sorghum) and cowpea was practiced in a substantial portion of the fields in Saponé, this did not negatively affect the accuracy of the AGB predictions, as has been seen in previous research (Li et al., 2022; Lobell et al., 2020a). This can be explained by the relatively low contribution of cowpea (10–20%) to overall plot AGB and that cowpea is a low growing crop obscured by the cereal canopy from the perspective of the satellite sensor.

The monthly and seasonal compositions started in June since this was generally when the growing season starts and is, therefore, relevant for crop establishment and resilience. Moreover, satellite data shortage or misfunctioning could happen and evaluating the relation of early season data to the total ABG could be beneficial. However, there is still a long period for plant growth, and the weather and management conditions would impact the final ABG biomass, suggesting that June is probably not the best month to be individually considered when evaluating monthly satellite information to estimate crop AGB, which was shown in our results.

The fact that September was the composition period that generated the most accurate results for both site-specific and general models shows that the timing of crop maturity occurs during similar time windows at both sites. While most comparable studies conducted in the Sudano-Sahel have used the whole growing season as composition period (Karlson et al., 2020; Lambert et al., 2018; Leroux et al., 2020; Lobell et al., 2020b), our results suggest that improvements in prediction accuracy may be achieved if a narrower time window (e.g., a month) is used. However, the possibility of employing such a narrow time window when using optical satellite data (e.g., Sentinel-2) will depend on cloud cover conditions, which can vary from year to year and, by extension, the size of the area to be mapped. We speculate that the reason for September being the best month for doing the composition is that it is the period when the chlorophyll content of the crops is highest, just before the time when the crops start to ripen. Focusing on this smaller time window provides a more precise representation of this phenological stage compared to the full growing season composition that can be influenced by less suitable satellite observations.

Another key aspect of the modelling framework design is the selection of a modelling algorithm (Lu, 2006; Morais et al., 2021). While the performance of an algorithm in terms of prediction accuracy and stability are obviously key criteria, other aspects also need to be considered, including input data requirements and computation efficiency. We evaluated the performance of four algorithms that are among the most common in the literature focusing on crop productivity estimation from remote sensing data (Jeong et al., 2016; van Klompenburg et al., 2020). More advanced methods, such as deep learning, are available and could be an important topic for future research on crop productivity estimation in agroforestry landscapes using remote sensing (Joshi et al., 2023; Murugananthan et al., 2022). Overall, our results demonstrate that the estimation performance in terms of accuracy of the algorithms does not differ substantially, except for MARS that was ranked at the bottom in several modelling scenarios. On the other hand, MLR, RF and SVM produced consistently high estimation accuracies for most modelling scenarios. A key benefit of MLR is that it is simple and produces model equations that can be observed and easily implemented, given that required input data are available. These equations can also be used to better understand the relationships between predictors and crop productivity, as well as give indications on how the model can be improved.

Assessments of predictor variable importance provide insight into the modelling process and can guide choices about which input data sources to include, as well as those that are redundant (Wei et al., 2015). This can make the modelling more efficient by limiting unnecessary data handling and processing, as well as facilitating the construction of parsimonious models. For the MLR, SAR-based predictors contributed to models developed for 2020 and in the combined years-scenarios. A reasonable explanation is that SAR predictors better captured crop structure in 2020 when AGB was considerably higher compared to 2021 (McNairn and Shang, 2016). Thus, including SAR predictors in the modelling can compensate for the potential saturation of vegetation indices based on optical data that can occur when crop AGB is high (Aklilu Tesfaye and Gessesse Awoke, 2021). While the use of SAR data for crop AGB modelling has been limited in Africa and non-existent in West African parkland landscapes (Parag et al., 2024), our results show that this is an interesting area for further investigation.

Optical predictors, including red-edge wavelengths from Sentinel-2, provide important contributions in most of the modelling scenarios for both MLR and ML. Red-edge wavelengths are known to be sensitive to the chlorophyll content of leaves and canopy (Clevers and Kooistra, 2012; Dash and Curran, 2004) and have been successfully applied in crop productivity modelling focused on agroforestry parklands with millet and sorghum previously (Karlson et al., 2020; Lobell et al., 2020b).

Moreover, the contribution of predictor variables characterizing soil properties and tree cover was not consistent in our results. Soil properties, such as soil carbon, should have a positive influence on crop productivity (Ma et al., 2023), but the quality of the ISDASoil data may be insufficient in this geographical area and at this spatial scale. In addition, the inclusion of tree cover as a predictor in crop productivity modelling in agroforestry parklands has been shown to improve the performance previously (Leroux et al., 2020), and our results confirm that the inclusion can be beneficial, with the prerequisite concerning the availability of a high-resolution tree cover product. Many legacy tree cover products have performed relatively poorly in areas with open tree cover, such as parkland landscapes (Heiskanen, 2008; Herrmann et al., 2013). Recent potentially improved products include the updated version of Hansen et al. (2013) and Brandt et al. (2022). Lastly, elevation was an important predictor in models, possibly due to some topographic variability and the fact that the high-yielding fields are located where runoff concentrates. The effect of topography in crop productivity modelling should be investigated in more detail, preferably by computing terrain indices from high-quality DEM for inclusion as predictor variables.

Overall, our results regarding predictor variable importance suggest that all data sources used in this study should be considered in the modelling process and be subject to the variable selection process to create optimal conditions for achieving high accuracy crop AGB estimates.

### 4.4. Potential ways forward

This study has provided key insights about several important aspects related to the development of a framework for estimating and monitoring crop productivity in agroforestry parklands, including how different datasets should be managed and processed and how to develop models that provide accurate estimates at the pixel scale. There are, however, things that need further attention, both in terms of practical considerations and scientific progress.

Firstly, one prerequisite for such a framework is the availability of field data over the years, preferably collected in a way that characterizes geographical variations in crop productivity and crop types (Lobell et al., 2020a). The pooling of resources for creating open-access databases of both old and new field data collected during different initiatives includes scientific projects and other public and private agricultural initiatives (Bégué et al., 2020). Another way to achieve this could be with crowdsourcing, where local farmers are trained in the collection and reporting of crop-cut data (Wu et al., 2023). The use of crowdsourcing has been proven successful for related tasks, including farmers collecting vast amounts of labeled data for training and validating crop classification models (Wang et al., 2020; Yu et al., 2023). The required tools, including custom-build applications and GPS can relatively easily be disseminated through smartphone technology, which is growing rapidly in Africa (Ezeoha et al., 2019). Farmers could support wet biomass estimations, and the dry biomass could be extrapolated by using water content ratios already available for the different crops. However, the challenges associated with making farmers adopt the technology and gaining sufficient knowledge should not be understated. Such an initiative will probably require that farmers are incentivized economically. However, the overall costs would likely be manageable considering the value of the resulting product for both monitoring and scientific applications. The costs obviously will be related to the number of field plots that are included in the effort. This need does not consider a high number of plots per se but the acquisition of field data that represents the variability of the region over contrasting geographical zones and variations in the weather conditions and length of the growing season over the years. Recent research has shown that the number of plots required to enable accurate crop productivity modelling over large areas is smaller than what is usually assumed. For example, Lobell et al. (2020b) showed that a fraction as small as 5% of field data plots (30 out of their original 595 plots) was sufficient to achieve accurate crop productivity estimates in Mali. Lobell et al. (2020a) provide similar conclusions for a study conducted in Uganda.

Secondly, there is a practical consideration with the high computational demands for processing remote sensing data when applying this framework in large areas. We used the Google Earth Engine (Gorelick et al., 2017), which is an excellent platform for cloud computing. However, to overcome processing restrictions inherent in the platform, we suggest including a tiling procedure in the framework where data processing is done in subsections of the focus area, where the size of the tiles needs to be determined practically.

A final practical consideration is the availability of a cropland mask that is updated annually. While we used the global land cover product by Potapov et al. (2022), optional datasets for deriving annual cropland masks include the Dynamic World land cover product (Brown et al., 2022) that provides near real-time land cover maps, as well as the cropland product of Digital Earth Africa. It is also worth noting that the development of Earth observation satellites is rapid and that new data options are soon to be operational. One interesting example is the NASA-ISRO SAR mission (NISAR), which will deliver L-band and S-band SAR data at a spatial resolution between 3 and 10 m depending on acquisition mode (Torbick et al., 2022). NIRSAR could represent an important complement to the Sentinel systems with potential benefits in terms of modelling performance improvements.

## 5. Conclusions

This study demonstrates the feasibility of accurate AGB estimation in the complex agroforestry parklands in Burkina Faso, utilizing primarily Sentinel-1 and Sentinel-2 time series imagery. The evaluation of different modelling scenarios, including site-specific and general models, offers important considerations for the design of a consistent crop productivity monitoring system. The findings contribute to developing effective and scalable frameworks for accurate and timely crop productivity assessment, which is crucial for informed decision-making in agriculture and food security, as well as for scientific enquiry.

The results suggest that an agroecological stratification of the area in focus, i.e., in different strata reflecting regional differences in growing conditions, could enhance modelling accuracy and generalization. The modelling algorithms emphasize the comparable and better performance of multilinear regression (MLR), as well as the RF and SVM machine learning algorithms, which outperformed the MARS algorithm. MLR could represent a more straightforward alternative for further practical applications and a transparent evaluation of the importance of predictor variables and their contribution to the models. The combination of SAR and optical data contributes to estimating crop productivity and its spatial variability in agroforestry landscapes with both single and mixed crop systems, providing promising results with stable performance both across geographical regions and between years.

The exploration of the predictor variables importance also offers valuable insights. In addition to the importance of those derived from Sentinel-1 and Sentinel-2 time series, including other data sources in the modelling process, needs further exploration, with variable results for the different modelling scenarios, sites and years. Given this, our recommendation is to include the data sources tested here (soil, elevation and tree cover) in the modelling process and let the variable selection procedure make the decision about inclusion.

The potential ways forward include the use of crowdsourcing for field data collection, addressing computational challenges through tiling procedures or data cube initiatives, and considering the availability of an updated cropland mask. The ongoing

development of Earth observation satellites presents possibilities for enhancing performance and expanding the scope of crop productivity estimation.

## CRediT authorship contribution statement

**Julianne Oliveira:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Martin Karlson:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis. **Abraham Sotongo Ouédraogo:** Methodology, Data curation. **Hugues Roméo Bazié:** Writing – review & editing, Resources, Methodology, Funding acquisition, Data curation. **Madelene Ostwald:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition.

## Ethical statement

The authors declare that all ethical practices have been followed in relation to the development, writing, and publication of the article.

## Declaration of competing interest

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

## Acknowledgements

We would like to thank the farmers of Cassou and Saponé for their willingness to allow us to use their fields and for their fruitful collaboration. Funding: this work was supported by the Swedish Research Council 2018–03722 and Formas 2018–00570.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rsase.2025.101494.

## Data availability

Data will be made available on request.

## References

Abdi, A.M., Seaquist, J., Tenenbaum, D.E., Eklundh, L., Ardö, J., 2014. The supply and demand of net primary production in the Sahel. Environ. Res. Lett. 9, 94003. https://doi.org/10.1088/1748-9326/9/9/094003.

Aklilu Tesfaye, A., Gessesse Awoke, B., 2021. Evaluation of the saturation property of vegetation indices derived from sentinel-2 in mixed crop-forest ecosystem. Spat. Inf. Res. 29, 109–121. https://doi.org/10.1007/s41324-020-00339-5.

Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: testing multiple methods and satellites in three countries. Remote Sens. Environ. 202, 129–141. https://doi.org/10.1016/j.rse.2017.04.014.

Bayala, J., Sanou, J., Teklehaimanot, Z., Kalinganire, A., Ouédraogo, S.J., 2014. Parklands for buffering climate risk and sustaining agricultural production in the Sahel of West Africa. Curr. Opin. Environ. Sustain. 6, 28–34. https://doi.org/10.1016/j.cosust.2013.10.004.

Bayala, J., Sanou, J., Teklehaimanot, Z., Ouedraogo, S.J., Kalinganire, A., Coe, R., Noordwijk, M. va, 2015. Advances in knowledge of processes in soil–tree–crop interactions in parkland systems in the West African Sahel: a review. Agric. Ecosyst. Environ. 205, 25–35. https://doi.org/10.1016/j.agee.2015.02.018.

Bayala, J., Teklehaimanot, Z., Ouedraogo, S.J., 2002. Millet production under pruned tree crowns in a parkland system in Burkina Faso. Agrofor. Syst. 54, 203–214. https://doi.org/10.1023/A:1016058906682.

Bayala, J., van Noordwijk, M., Lusiana, B., Ni'matul, K., Teklehaimanot, Z., Ouedraogo, S.J., 2008. In: Jose, S., Gordon, A.M. (Eds.), Separating the Tree–Soil–Crop Interactions in Agroforestry Parkland Systems in Saponé (Burkina Faso) Using WaNuLCAS BT - toward Agroforestry Design: an Ecological Approach. Springer, Netherlands, Dordrecht, pp. 285–297. https://doi.org/10.1007/978-1-4020-6572-9_17.

Bazié, H.R., Bayala, J., Zombré, G., Sanou, J., Ilstedt, U., 2012. Separating competition-related factors limiting crop performance in an agroforestry parkland system in Burkina Faso. Agrofor. Syst. 84, 377–388. https://doi.org/10.1007/s10457-012-9483-y.

Bégué, A., Leroux, L., Soumaré, M., Faure, J.-F., Diouf, A.A., Augusseau, X., Touré, L., Tonneau, J.-P., 2020. Remote sensing products and services in support of agricultural public policies in Africa: overview and challenges. Front. Sustain. Food Syst. 4. https://doi.org/10.3389/fsufs.2020.00058.

Blaes, X., Defourny, P., Wegmuller, U., Della Vecchia, A., Guerriero, L., Ferrazzoli, P., 2006. C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model. IEEE Trans. Geosci. Rem. Sens. 44, 791–800.

Boffa, J.M., 1999. Agroforestry Parkland in Sub-saharan Africa: FAO Conservation Guide 34. Rome, Italy.

Brandt, J., Ertel, J., Spore, J., Stolle, F., 2022. The extent of trees in the tropics. https://doi.org/10.21203/rs.3.rs-2109093/v1.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76, 156–172. https://doi.org/10.1016/S0034-4257(00)00197-8.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic World, near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4.

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. 114, 2189LP–2194. https://doi.org/10.1073/pnas.1616919114.
Cao, Y., Yan, L., Zheng, Z., 2008. Extraction of information on geologu hazard from multi-polarization SAR images. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. XXXVII, 1529–1532. Part B4. Beijing.
Carletto, C., Jolliffe, D., Banerjee, R., 2015. From tragedy to renaissance: improving agricultural data for better policies. J. Dev. Stud. 51, 133–148. https://doi.org/10.1080/00220388.2014.968140.
Clevers, J.G.P.W., Kooistra, L., 2012. Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 5, 574–583. https://doi.org/10.1109/JSTARS.2011.2176468.
Cortes, C., Vapnik, V., 1995. Support-Vector networks. Mach. Learn. 20, 273–297.
Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Rem. Sens. 25, 5403–5413. https://doi.org/10.1080/0143116042000274015.
dos Santos, E.P., Da Silva, D.D., do Amaral, C.H., 2021. Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. Int. J. Rem. Sens. 42, 7581–7609. https://doi.org/10.1080/01431161.2021.1959955.
Duncan, J.M.A., Dash, J., Atkinson, P.M., 2015. The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. Front. Environ. Sci. 3.
Etongo, D., Djenontin, I.N.S., Kanninen, M., Fobissie, K., 2015. Smallholders' tree planting activity in the Ziro province, southern Burkina Faso: impacts on livelihood and policy implications. Forests. https://doi.org/10.3390/f6082655.
Ezeoha, A.E., Obi, A., Igwe, A., Ezeruigbo, C., 2019. The mobile phone revolution, the Internet and rural electricity: what are the implications for food security in Africa? Inf. Dev. 36, 603–622. https://doi.org/10.1177/0266666919884991.
Foli, S., Abdoulaye, R., 2016. Drivers and outcomes of changing land use in parkland agroforestry systems of central Burkina Faso. In: Deakin, E.L., Kshatriya, M., Sunderland, T.C.H. (Eds.), Agrarian Change in Tropical Landscapes. Center for International Forestry Research (CIFOR), Bogor, Indonesia, pp. 269–301.
Forkuor, G., Dimobe, K., Serme, I., Tondoh, J.E., 2018. Landsat-8 vs. Sentinel-2: examining the added value of Sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. GIScience Remote Sens. 55, 331–354. https://doi.org/10.1080/15481603.2017.1370169.
Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141.
Gao, X., Chi, H., Huang, J., Han, Y., Li, Y., Ling, F., 2024. Comparison of cloud-mask algorithms and machine-learning methods using Sentinel-2 imagery for mapping paddy rice in Jianghan plain. Remote Sens. https://doi.org/10.3390/rs16071305.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.
Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422. https://doi.org/10.1023/A:1012487302797.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853. https://doi.org/10.1126/science.1244693.
Hay, R.K.M., 1995. Harvest index: a review of its use in plant breeding and crop physiology. Ann. Appl. Biol. 126, 197–216. https://doi.org/10.1111/j.1744-7348.1995.tb05015.x.
Heiskanen, J., 2008. Evaluation of global land cover data sets over the tundra–taiga transition zone in northernmost Finland. Int. J. Rem. Sens. 29, 3727–3751. https://doi.org/10.1080/01431160701871104.
Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G.E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.-M., Chamberlin, J., Silatsa, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci. Rep. 11, 6130. https://doi.org/10.1038/s41598-021-85639-y.
Herrmann, S.M., Wickhorst, A.J., Marsh, S.E., 2013. Estimation of tree cover in an agricultural parkland of Senegal using rule-based regression tree modeling. Remote Sens. https://doi.org/10.3390/rs5104900.
Hird, J.N., DeLancey, E.R., McDermid, G.J., Kariyeva, J., 2017. Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. Remote Sens. https://doi.org/10.3390/rs9121315.
Hoekman, D.H., Reiche, J., 2015. Multi-model radiometric slope correction of SAR images of complex terrain using a two-stage semi-empirical approach. Remote Sens. Environ. 156, 1–10. https://doi.org/10.1016/j.rse.2014.08.037.
Hosseini, M., McNairn, H., Mitchell, S., Dingle Robertson, L., Davidson, A., Homayouni, S., 2019. Synthetic aperture radar and optical satellite data for estimating the biomass of corn. Int. J. Appl. Earth Obs. Geoinf. 83, 101933. https://doi.org/10.1016/j.jag.2019.101933.
Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X.
Huete, A.R., Liu, H.Q., Batchily, K., van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sens. Environ. 59, 440–451. https://doi.org/10.1016/S0034-4257(96)00112-5.
Ibrahim, E.S., Rufin, P., Nill, L., Kamali, B., Nendel, C., Hostert, P., 2021. Mapping crop types and cropping systems in Nigeria with Sentinel-2 imagery. Remote Sens. https://doi.org/10.3390/rs13173523.
Ilstedt, U., Bargués Tobella, A., Bazié, H.R., Bayala, J., Verbeeten, E., Nyberg, G., Sanou, J., Benegas, L., Murdiyarso, D., Laudon, H., Sheil, D., Malmer, A., 2016. Intermediate tree cover can maximize groundwater recharge in the seasonally dry tropics. Sci. Rep. 6, 21930. https://doi.org/10.1038/srep21930.
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.
Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. Remote Sens. Environ. 228, 115–128. https://doi.org/10.1016/j.rse.2019.04.016.
Joshi, A., Pradhan, B., Gite, S., Chakraborty, S., 2023. Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: a systematic review. Remote Sens. https://doi.org/10.3390/rs15082014.
Karatzoglou, A., Hornik, K., Smola, A., Zeileis, A., 2004. Kernlab – an S4 package for kernel methods in R. J. fo Stat. Softw. 11.
Karlson, M., Bolin, D., Bazié, H.R., Ouedraogo, A.S., Soro, B., Sanou, J., Bayala, J., Ostwald, M., 2023. Exploring the landscape scale influences of tree cover on crop yield in an agroforestry parkland using satellite data and spatial statistics. J. Arid Environ. 218, 105051. https://doi.org/10.1016/j.jaridenv.2023.105051.
Karlson, M., Ostwald, M., Bayala, J., Bazié, H.R., Ouedraogo, A.S., Soro, B., Sanou, J., Reese, H., 2020. The potential of Sentinel-2 for crop production estimation in a smallholder agroforestry landscape, Burkina Faso. Front. Environ. Sci. 8. https://doi.org/10.3389/fenvs.2020.00085.
Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., Mattsson, E., 2015. Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and Random Forest. Remote Sens. 7, 10017–10041. https://doi.org/10.3390/rs70810017.
Kim, Y., Jackson, T., Bindlish, R., Lee, H., Hong, S., 2012. Radar vegetation index for estimating the vegetation water content of rice and soybean. IEEE Geosci. Remote Sens. Lett. 9, 564–568. https://doi.org/10.1109/LGRS.2011.2174772.
Kim, Y., Zyl, J.J. van, 2009. A time-series approach to estimate soil moisture using polarimetric radar data. IEEE Trans. Geosci. Rem. Sens. 47, 2519–2527. https://doi.org/10.1109/TGRS.2009.2014944.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. fo Stat. Softw. 28.
Kuhn, M., Johnson, K., 2019. Feature Engineering and Selection: A Practical Approach for Predictive Models. Chapman & Hall/CRC Press, Boca Raton, FL.
Lambert, M.-J., Traoré, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. Remote Sens. Environ. 216, 647–657. https://doi.org/10.1016/j.rse.2018.06.036.
Lee, J.-S., Grunes, M.R., Grandi, G. de, 1999. Polarimetric SAR speckle filtering and its implication for classification. IEEE Trans. Geosci. Rem. Sens. 37, 2363–2373. https://doi.org/10.1109/36.789635.

Leroux, L., Castets, M., Baron, C., Escorihuela, M.-J., Bégué, A., Lo Seen, D., 2019. Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. Eur. J. Agron. 108, 11–26. https://doi.org/10.1016/j.eja.2019.04.007.

Leroux, L., Falconnier, G.N., Diouf, A.A., Ndao, B., Gbodjo, J.E., Tall, L., Balde, A.A., Clermont-Dauphin, C., Bégué, A., Affholder, F., Roupsard, O., 2020. Using remote sensing to assess the effect of trees on millet yield in complex parklands of Central Senegal. Agric. Syst. 184, 102918. https://doi.org/10.1016/j.agsy.2020.102918.

Li, C., Chimimba, E.G., Kambombe, O., Brown, L.A., Chibarabada, T.P., Lu, Y., Anghileri, D., Ngongondo, C., Sheffield, J., Dash, J., 2022. Maize yield estimation in intercropped smallholder fields using satellite data in southern Malawi. Remote Sens. https://doi.org/10.3390/rs14102458.

Li, G., Lu, D., Moran, E., Dutra, L., Batistella, M., 2012. A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. ISPRS J. Photogrammetry Remote Sens. 70, 26–38. https://doi.org/10.1016/j.isprsjprs.2012.03.010.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field Crops Res. 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008.

Lobell, D.B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., Murray, S., 2020a. Eyes in the sky, boots on the ground: assessing satellite- and ground-based approaches to crop yield measurement and analysis. Am. J. Agric. Econ. 102, 202–219. https://doi.org/10.1093/ajae/aaz051.

Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: their importance, magnitudes, and causes. Annu. Rev. Environ. Resour. 34, 179–204. https://doi.org/10.1146/annurev.environ.041008.093740.

Lobell, D.B., Di Tommaso, S., You, C., Yacoubou Djima, I., Burke, M., Kilic, T., 2020b. Sight for sorghums: comparisons of satellite- and ground-based sorghum yield estimates in Mali. Remote Sens. https://doi.org/10.3390/rs12010100.

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Rem. Sens. 27, 1297–1328. https://doi.org/10.1080/01431160500486732.

Ma, Y., Woolf, D., Fan, M., Qiao, L., Li, R., Lehmann, J., 2023. Global crop production increase by soil organic carbon. Nat. Geosci. 16, 1159–1165. https://doi.org/10.1038/s41561-023-01302-3.

Main-Knorn, M., Plug, B., Louis, J., Debaecker, V., Muller-Wilm, U., Gascon, F., 2017. Sen2Cor for sentinel-2. In: Proceedings Volume 10427, Image and Signal Processing for Remote Sensing XXIII, 1042704. Wasaw, Poland.

Major, D.J., Baret, F., Guyot, G., 1990. A ratio vegetation index adjusted for soil brightness. Int. J. Rem. Sens. 11, 727–740. https://doi.org/10.1080/01431169008955053.

McNairn, H., Shang, J., 2016. A review of multitemporal synthetic aperture radar (SAR) for crop monitoring, 317–340. https://doi.org/10.1007/978-3-319-47037-5_15.

Milborrow, S., Hastie, T., Tibshirani, R., Miller, A., Lumley, T., 2024. earth: Multivariate Adaptive Regression Splines. R package version 5.3.4, pp. 1–53.

Miller, M.A.E., Shepherd, K.D., Kisitu, B., Collinson, J., 2021. iSDAsoil: the first continent-scale soil property map at 30 m resolution provides a soil information revolution for Africa. PLoS Biol. 19, e3001441.

Monteiro, L.A., Ramos, R.M., Battisti, R., Soares, J.R., Oliveira, J.C., Figueiredo, G.K.D.A., Lamparelli, R.A.C., Nendel, C., Lana, M.A., 2022. Potential use of data-driven models to estimate and predict soybean yields at national scale in Brazil. Int. J. Plant Prod. 16, 691–703. https://doi.org/10.1007/s42106-022-00209-0.

Montero, D., 2021. eemont: a Python package that extends Google Earth Engine. J. Open Source Softw. 6, 3168.

Morais, T.G., Teixeira, R.F.M., Figueiredo, M., Domingos, T., 2021. The use of machine learning methods to estimate aboveground biomass of grasslands: a review. Ecol. Indic. 130, 108081. https://doi.org/10.1016/j.ecolind.2021.108081.

Morton, J.F., 2007. The impact of climate change on smallholder and subsistence agriculture. Proc. Natl. Acad. Sci. 104, 19680LP–19685. https://doi.org/10.1073/pnas.0701855104.

Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N., Reiche, J., 2021. Sentinel-1 SAR backscatter analysis ready data preparation in Google Earth Engine. Remote Sens. https://doi.org/10.3390/rs13101954.

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N.H., Islam, N., 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sens. https://doi.org/10.3390/rs14091990.

Nasirzadehdizaji, R., Balik Sanli, F., Abdikan, S., Cakir, Z., Sekertekin, A., Ustuner, M., 2019. Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage. Appl. Sci. https://doi.org/10.3390/app9040655.

Oguntunde, P.G., Lischeid, G., Abiodun, B.J., Dietrich, O., 2014. Analysis of spatial and temporal patterns in onset, cessation and length of growing season in Nigeria. Agric. For. Meteorol. 194, 77–87. https://doi.org/10.1016/j.agrformet.2014.03.017.

Olson, K.R., Olson, G.W., 1986. Use of multiple regression analysis to estimate average corn yields using selected soils and climatic data. Agric. Syst. 20, 105–120. https://doi.org/10.1016/0308-521X(86)90062-4.

Parag, M., Lottering, R., Peerbhay, K., Agjee, N., Poona, N., 2024. The use of synthetic aperture radar technology for crop biomass monitoring: a systematic review. Remote Sens. Appl. Soc. Environ. 33, 101107. https://doi.org/10.1016/j.rsase.2023.101107.

Paré, S., Savadogo, P., Tigabu, M., Ouadba, J.M., Odén, P.C., 2010. Consumptive values and local perception of dry forest decline in Burkina Faso, West Africa. Environ. Dev. Sustain. 12, 277–295. https://doi.org/10.1007/s10668-009-9194-3.

Periasamy, S., 2018. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: an attempt on Sentinel-1. Remote Sens. Environ. 217, 537–549. https://doi.org/10.1016/j.rse.2018.09.003.

Potapov, P., Hansen, M.C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., Stolle, F., Harris, N., Song, X.-P., Baggett, A., Kommareddy, I., Kommareddy, A., 2022. The global 2000-2020 land cover and land use change dataset derived from the landsat archive: first results. Front. Remote Sens. 3. https://doi.org/10.3389/frsen.2022.856903.

Potter, C.S., Randerson, J.T., Field, C.B., Matson, P.A., Vitousek, P.M., Mooney, H.A., Klooster, S.A., 1993. Terrestrial ecosystem production: a process model based on global satellite and surface data. Glob. Biogeochem. Cycles 7, 811–841. https://doi.org/10.1029/93GB02725.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1.

R Development Core Team, 2020. R: A Language and Environment for Statistical Computing. R foundation for statistical computing, Vienna.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. NASA Special Publication, Washington DC, pp. 309–317.

Samberg, L.H., Gerber, J.S., Ramankutty, N., Herrero, M., West, P.C., 2016. Subnational distribution of average farm size and smallholder contributions to global food production. Environ. Res. Lett. 11, 124010. https://doi.org/10.1088/1748-9326/11/12/124010.

Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S.V., 2021. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Sci. Rep. 11, 1606. https://doi.org/10.1038/s41598-020-80820-1.

Sinare, H., Gordon, L.J., 2015. Ecosystem services from woody vegetation on agricultural lands in Sudano-Sahelian West Africa. Agric. Ecosyst. Environ. 200, 186–199. https://doi.org/10.1016/j.agee.2014.11.009.

Sivakumar, M.V.K., 1988. Predicting rainy season potential from the onset of rains in Southern Sahelian and Sudanian climatic zones of West Africa. Agric. For. Meteorol. 42, 295–305. https://doi.org/10.1016/0168-1923(88)90039-1.

Söderström, B., Kiema, S., Reid, R.S., 2003. Intensified agricultural land-use and bird conservation in Burkina Faso. Agric. Ecosyst. Environ. 99, 113–124. https://doi.org/10.1016/S0167-8809(03)00144-0.

Stasolla, M., Neyt, X., 2018. An operational tool for the automatic detection and removal of border noise in Sentinel-1 GRD products. Sensors. https://doi.org/10.3390/s18103454.

Torbick, N., Huang, X., Chapman, B., Kellndorfer, J., Siqueira, S.S.P., 2022. NISAR: open access and operational L-band data for agricultural science. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 6049–6052. https://doi.org/10.1109/IGARSS46834.2022.9883318.

Trisos, C.H., Adelekan, I.O., Totin, E., Ayanlade, A., Efitre, J., Gemeda, A., Kalaba, A., Lennard, C., Masao, C., Mgaya, Y., Ngaruiya, G., Olago, D., Simpson, N.P., Zakieldeen, S., 2023. Africa, in: (IPCC), I.P. On C.C. Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth

Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 1285–1456. https://doi.org/10.1017/9781009325844.011.

UNISTAT, 2023. Population, surface area and density [WWW Document]. https://data.un.org. (Accessed 14 February 2024). https://data.un.org/_Docs/SYB/PDFs/SYB65_1_202209_Population,SurfaceAreaandDensity.pdf.

van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: a systematic literature review. Comput. Electron. Agric. 177, 105709. https://doi.org/10.1016/j.compag.2020.105709.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. Remote Sens. Environ. 199, 415–426. https://doi.org/10.1016/j.rse.2017.07.015.

Vintrou, E., Desbrosse, A., Bégué, A., Traoré, S., Baron, C., Lo Seen, D., 2012. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. Int. J. Appl. Earth Obs. Geoinf. 14, 83–93. https://doi.org/10.1016/j.jag.2011.06.010.

Vollrath, A., Mullissa, A., Reiche, J., 2020. Angular-based radiometric slope correction for sentinel-1 on Google earth engine. Remote Sens. https://doi.org/10.3390/rs12111867.

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., Strauss, P., 2018. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: an Austrian case study. Remote Sens. https://doi.org/10.3390/rs10091396.

Wang, L., Qu, J.J., 2007. NMDI: a normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. Geophys. Res. Lett. 34. https://doi.org/10.1029/2007GL031021.

Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., Lobell, D.B., 2020. Mapping crop types in southeast India with smartphone crowdsourcing and deep learning. Remote Sens. https://doi.org/10.3390/rs12182957.

Wei, P., Lu, Z., Song, J., 2015. Variable importance analysis: a comprehensive review. Reliab. Eng. Syst. Saf. 142, 399–432. https://doi.org/10.1016/j.ress.2015.05.018.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402. https://doi.org/10.1016/j.rse.2019.111402.

Wright, M.N., Ziegler, A., 2017. ranger: a fast implementation of Random Forests for high dimensional data in C++ and R. J. Stat. Software 77, 1–17. https://doi.org/10.18637/jss.v077.i01.

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A.B., Qin, X., Yan, N., Chang, S., Zhao, Y., Dong, Q., Boken, V., Plotnikov, D., Guo, H., Wu, F., Zhao, H., Deronde, B., Tits, L., Loupian, E., 2023. Challenges and opportunities in remote sensing-based crop monitoring: a review. Natl. Sci. Rev. 10, nwac290. https://doi.org/10.1093/nsr/nwac290.

Wu, Q., 2020. geemap: a Python package for interactive mapping with Google Earth Engine. J. Open Source Softw. 5, 2305.

Yaméogo, G., B, Y., Yélémou, L., Kaboré, O., Traoré, D., 2014. Combustibles des ménages et modes d'utilisation à l'échelle du terroir de Vipalogo, en zone nord soudanienne du Burkina Faso. Int. J. Biol. Chem. Sci. 8, 1742–1756.

Yu, Q., Duan, Y., Wu, Q., Liu, Y., Wen, C., Qian, J., Song, Q., Li, W., Sun, J., Wu, W., 2023. An interactive and iterative method for crop mapping through crowdsourcing optimized field samples. Int. J. Appl. Earth Obs. Geoinf. 122, 103409. https://doi.org/10.1016/j.jag.2023.103409.