remote sensing

MDPI

Review

Sugarcane Yield Estimation Using Satellite Remote Sensing Data in Empirical or Mechanistic Modeling: A Systematic Review

Nildson Rodrigues de França e Silva [1,*], Michel Eustáquio Dantas Chaves [2], Ana Cláudia dos Santos Luciano [3], Ieda Del'Arco Sanches [1,4], Cláudia Maria de Almeida [1,4] and Marcos Adami [1,4]

[1] Remote Sensing Postgraduate Program (PGSER), Coordination of Teaching, Research and Extension (COEPE), National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil; ieda.sanches@inpe.br (I.D.S.); claudia.almeida@inpe.br (C.M.d.A.); marcos.adami@inpe.br (M.A.)

[2] São Paulo State University (Unesp), School of Sciences and Engineering, Tupã 17602-496, Brazil; michel.dantas@unesp.br

[3] Department of Biosystems Engineering, Graduate School of Agriculture Luiz de Queiroz (ESALQ), University of São Paulo (USP), Piracicaba 13418-900, Brazil; analuciano@usp.br

[4] Earth Observation and Geoinformatics Division (DIOTG), General Coordination of Earth Science (CG-CT), National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil

* Correspondence: nildson.silva@inpe.br

Citation: de França e Silva, N.R.; Chaves, M.E.D.; Luciano, A.C.d.S.; Sanches, I.D.; de Almeida, C.M.; Adami, M. Sugarcane Yield Estimation Using Satellite Remote Sensing Data in Empirical or Mechanistic Modeling: A Systematic Review. *Remote Sens.* **2024**, *16*, 863. https://doi.org/10.3390/rs16050863

Academic Editor: Nicolas R. Dalezios

Received: 17 January 2024
Revised: 19 February 2024
Accepted: 26 February 2024
Published: 29 February 2024

Abstract: The sugarcane crop has great socioeconomic relevance because of its use in the production of sugar, bioelectricity, and ethanol. Mainly cultivated in tropical and subtropical countries, such as Brazil, India, and China, this crop presented a global harvested area of 17.4 million hectares (Mha) in 2021. Thus, decision making in this activity needs reliable information. Obtaining accurate sugarcane yield estimates is challenging, and in this sense, it is important to reduce uncertainties. Currently, it can be estimated by empirical or mechanistic approaches. However, the model's peculiarities vary according to the availability of data and the spatial scale. Here, we present a systematic review to discuss state-of-the-art sugarcane yield estimation approaches using remote sensing and crop simulation models. We consulted 1398 papers, and we focused on 72 of them, published between January 2017 and June 2023 in the main scientific databases (e.g., AGORA-FAO, Google Scholar, Nature, MDPI, among others), using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology. We observed how the models vary in space and time, presenting the potential, challenges, limitations, and outlooks for enhancing decision making in the sugarcane crop supply chain. We concluded that remote sensing data assimilation both in mechanistic and empirical models is promising and will be enhanced in the coming years, due to the increasing availability of free Earth observation data.

Keywords: crop modeling; text mining; crop monitoring; systematic literature review; crop yield

1. Introduction

Sugarcane (*Saccharum officinarum*) is a semi-perennial crop grown in tropical and subtropical countries that have economic, social, and environmental importance due to its use in the production of sugar, bioethanol, and bioelectricity [1,2]. Currently, annual world production is 2 billion tons in 27.5 million hectares (Mha), most of which is derived from developing countries [3]. Although China, Pakistan, and Thailand have relevant production, 57% of this amount is concentrated in Brazil (36%) and India (21%). These five countries have the largest harvested areas: Brazil (10 Mha), India (5.2 Mha), China (2.3 Mha), Thailand (1.5 Mha), and Pakistan (1.3 Mha) [3].

Due to the area extent of sugarcane production and its economic importance [1–3], technologies to anticipate yield information are essential for different phases of the supply chain, including crop management and decision making. Empirical or mechanistic models are useful tools for collecting this information in advance. The empirical models are built on statistical relationships between variables of interest (dependent) and predictors (independent variable, in this study represented by sugarcane yield), and the mechanistic ones simulate sugarcane cultivation development using a set of equations that represents its physiological responses under different environmental variables [4,5]. Mechanistic models are prominent approaches to estimating it, but they require a lot of input information [6], most of them from the field. However, as complex systems, local peculiarities and spatial–temporal scales may create divergences and noise in the models' results.

Mechanistic models simulate processes such as, for example, photosynthesis, soil moisture, phenology, temperature dynamic, biomass growth or grain yield formation, and gas exchanges between the canopy and the atmosphere. Due to this, these models need a large amount of input data (usually in a daily frequency) that can be difficult to parameterize for large scales and where there is broad agricultural variability, e.g., types of soil, crop practices, or varieties [7,8]. However, because mechanistic models allow different processes that influence crop development to be simulated, it is possible to assess the crop in a variety of ways, e.g., in relation to soil moisture, leaf area index, biomass, and yield [9].

Empirical models do not need a calibration process as mechanistic models do, and they need a large available dataset as input. They are also useful in studies where the aim is to obtain the crop yield at regional or global scales, either for the close future or the past. Nevertheless, they are limited in future scenario extrapolation due to the fact that they do not have a reliable physical mechanism to estimate yields in the future since they are dependent on historical input data [10].

Field data are expensive and time-consuming, especially for large areas. In this regard, remote sensing-based estimation methods can increase the efficiency of yield models [11,12]. Remote sensing (RS) data as input allow for their application on larger spatial–temporal scales and could replace field data that are hard to obtain in a sound and timely manner. RS is also a nondestructive method that enables monitoring vegetation temporally and spatially [13,14]. Two kinds of RS data stand out in estimating crop yield. Firstly, mechanistic models take the nature of the soil, climatic variables, and so on, as input parameters that can estimate yield by simulating crop physical processes or combine this approach using the assimilation of RS data into crop growth models. And according to [8], such data allow information from Earth observations to be incorporated into a model. Secondly, low-resolution satellite images, which refer to optical sensors with a resolution above 250 m, are essential to predict yields at the regional level. In the last decade, medium-resolution satellite images also started to be used for this purpose [15–17].

Differently from previous review articles in this line, which tend to focus on multiple crops indistinctly, this work aimed to perform a systematic literature review specifically on sugarcane yield estimation using RS data in empirical or mechanistic models. Seeking to organize the state-of-the-art related to this theme and identify opportunities for new studies, we considered papers published between January 2017 and June 2023 in the main scientific databases.

2. Empirical and Mechanistic Crop Yield Models

Crop yield models are basically divided into empirical and mechanistic (also referred as deterministic) models, in which the first group relies on conventional statistical, machine learning (ML), and deep learning (DL) approaches, and the second group uses formal equations relating parameters associated with meteorological and soil conditions, crop physiological status, and management practices to yield. There is also a third group which concerns hybrid approaches, which merge empirical and mechanistic methods, lying in a fuzzy zone between pure empirical and pure deterministic approaches. However, in the case of this study, which focuses on the sugarcane crop in a specific timeframe, this strategy was extremely rare, and only one paper dealing with such a hybrid approach was found, namely the study of [18].

2.1. Empirical Models

Empirical models, also known as "regressions", are developed based on a linear or nonlinear relationship, calibrating a numerical association between a specific variable or several multi-predictor biophysical variables and RS data or a transformation of these data [14]. An example is the estimation of crop yields (dependent attribute) using meteorological, soil, and management data (independent attributes). As an advantage, this type of modeling allows the user to test different variables that are not common in crop simulation models, for example, different satellite vegetation indices and image sensor time series, weather indices, or other variables, depending on the experience of the researcher. In addition, most models are trained with field observations, allowing for the use of reliable data on crop management to make predictions. The empirical model's robustness increases when analyzing variables of interest over large spatiotemporal scales [6,19,20].

Computational advances that lead to the use of machine learning and deep learning algorithms have expanded the development of agricultural crop yield models using empirical approaches and RS data [13,21]. Different strategies have been used to obtain sugarcane yield using empirical models, such as Linear Regression, Multiple Linear Regression, and Stepwise Multiple Regression [11,22–25], Support Vector Machine (SVM) [11,18,26,27], Artificial Neural Networks (ANN) [11,28,29], and Random Forest (RF) [12,18,22,26,27,30–32]. As input, they use RS, field, agrometeorological, and terrain data, among others. The main variables are listed in the Supplementary Materials.

As for disadvantages, empirical models can hardly extrapolate beyond their training region [6,20]. They are subject to collinearity problems between the predictor variables (temperature and altitude, for example). Another possible problem is the stationarity of the used data when past relationships may not happen in the future [33]. A point of emphasis is the quality of the reference data. The final model may perform at the same level as the reference data's quality.

Empirical models are highly dependent on both reference data availability and access, since the existing data are not always rendered to modelers [21]. These models are also sensitive to data quality and demand efforts for a sound management of their database. Despite that, they are less data-intensive as compared to mechanistic models, and they do not require a complex parameterization either. They are also suitable for large-scale studies [20], given the fact that they are commonly driven by remote sensing data.

Such models do not require a high level of data handling as mechanistic models do [6], although both empirical and mechanistic models have different degrees of computational cost, which tend to vary on a case-by-case basis. These two categories of models are not easily transferrable to a business model, although empirical models present the advantage of being flexible to the inclusion of manifold variables, while mechanistic models follow a pre-defined list of input data. In terms of model performance, both categories can achieve high accuracy, provided they are skillfully executed. Empirical models can increase their accuracy in cases where remote sensing data are associated with field data [12].

2.2. Mechanistic Models

Mechanistic models are formed by equations collections that aim to correlate crop physiological responses to environmental conditions and estimate how this affects their development [5]. Usually, they are more complex than empirical models and may require a large amount of input data. Examples are the Decision Support System for Agrotechnology Transfer (DSSAT) [34], Agricultural Production Systems Simulator (APSIM) [35], World Food Studies (WOFOST) [36], FAO Agroecological Zone Model (FAO-AZM) [37], AquaCrop [38], Agricultural Land Management Alternatives with Numerical Assessment Criteria (ALMANAC) [39], CROPWAT [40] and Agronomic Modular Simulator for Sugarcane (SAMUCA) [41]. We selected DSSAT, APSIM, WOFOST, FAO-AZM, and AquaCrop to discuss because they are the most widespread and cited models.

2.2.1. Decision Support System for Agrotechnology Transfer (DSSAT)

DSSAT is composed of more than 40 crop simulation models [34], including the CANEGRO model, specifically developed for sugarcane by [42], refined by [43,44]. In summary, CANEGRO is a module implemented from DSSAT version 3.5 that simulates sugarcane's growth and development using data from the sugarcane variety, meteorological conditions, soil properties, and management information [44].

The DSSAT/CANEGRO model has been globally applied. In Pakistan, ref. [45] calibrated, validated, and analyzed their results in industrial and non-industrial sugarcane areas. The authors also evaluated the impacts of climate change on sugarcane. This application was unprecedented for a semi-arid region. In the United States, ref. [46] assessed the feasibility of simulating sugarcane growth and estimating biomass yield for type II energy sugarcane genotypes, which are characterized by having a low sugar level (sucrose less than 6%) and very high fiber content. In the study, the authors concluded that calibrated DSSAT/CANEGRO could provide good estimations of energy sugarcane biomass (Mean Absolute Error, MAE = 2.9 ton ha^{-1}; % Root Mean Square Error, %RMSE = 16.5 ton ha^{-1}; Coefficient of Determination, R^2 = 0.94), and emphasized that the modeling could be improved using specific genotype data for energy sugarcane in the simulation process.

In Brazil, ref. [47] determined the best planting date for sugarcane for a producing region in the state of Alagoas, northeast Brazil. The authors performed simulations for different dates, observing that the model can simulate crop growth variables and indicate the best planting window in different Brazilian regions in regular years and years affected by El Niño and La Niña events. In regular years, the best date for planting sugarcane in the region was 30 October; however, in El Niño and La Niña years, these dates were, respectively, shifted to 15 January and 30 September.

2.2.2. Agricultural Production Systems Simulator (APSIM)

The APSIM model, developed by [35] at the Commonwealth Scientific and Industrial Research Organization (CSIRO) and Agricultural Production Systems Research Unit (APSRU), is one of the most used simulation models for agricultural systems [48,49]. The main component of the sugarcane module in APSIM is its ability to estimate crop dry matter accumulation and sugar production. Also, the model can estimate the crop water use efficiency, nitrogen accumulation, and the dry and fresh biomass weight of the plant or ratoon cane, considering climate, soil type, genotype, and management [35,49,50]. Aiming to maximize sucrose production in Brazil, ref. [51] used the model to determine the best periods for irrigation interruption in irrigated areas. The authors concluded that the drying-off periods can vary according to their locations, soil type, and harvest month, but generally occur at the beginning and end of the harvesting season, when higher rainfall interannual variability is noticed. In Australia, ref. [52] used APSIM to propose a bioeconomic model that related water productivity and profit. The authors obtained sugarcane yields under different climate conditions, scheduling scenarios for irrigation and estimating the expected profit with less uncertainty.

2.2.3. World Food Studies (WOFOST)

WOFOST is a simulation model integrating the Monitoring Agriculture with Remote Sensing (MARS) system as a central component of the crop monitoring and yield estimation system in Europe [36]. With a strong biophysical basis, WOFOST has been used in different studies related to inter-annual variability and risk of crop yields, crop yield variation to soil type and agro-hydrological conditions, evaluation of differences between cultivars, factors impacting crop development, detection of adverse conditions in crop development, and prediction of crop yields on a regional scale [53]. The model output variables are leaf area, water use, and the simulated total crop biomass and yield. As input, it demands meteorological, soil, crop, location, and management data [54]. Regarding sugarcane cultivation, the model has already been, for instance, applied in Ethiopia [55] and China [56,57]. However, it has still been limitedly tested and validated [53]. Its generic implementation allows for its application to different crops using the same principles and algorithms, changing only parameter values. WOFOST has recently been incorporated into the Python Crop Simulation Environment (PCSE) [36,53].

2.2.4. FAO Agroecological Zone Model (FAO-AZM)

The FAO-AZM [37] presents a much simpler formulation [58,59]. It allows for estimating the potential crop yield if water and nutritional needs are satisfied, disregarding losses due to pests or diseases. The actual yield is calculated by penalizing the potential yield by water deficit [60]. Over the years, the model has undergone several improvements [59,61,62]. Presenting satisfactory results for assessing continental-scale regions, the FAO-AZM is widely used in countries with large production areas, such as Brazil. The best yield estimates for sugarcane cultivation using the FAO-AZM were obtained in studies where adjustments and calibration of the model to local climate conditions were executed [60]. Ref. [58] applied FAO-AZM to assess the impacts of irrigation systems on sugarcane yields, considering land and water use efficiency. In the study, the authors said that the irrigation system can reduce the yield variability in the different producing regions in Brazil and decrease the land demand because higher yields were obtained in the irrigated areas, and conventional agriculture needs to undergo a transformation to sustainable intensive agriculture.

On the other hand, ref. [63] combined agrometeorological (potential yield) and economic (sugarcane prices and rural credit concession) approaches to estimate the actual sugarcane yield in 18 producing regions of São Paulo state, Brazil, between 1995 and 2012, assessing the impact of economic variables by means of a statistical model. In the studied region, the sugarcane actual yield was indeed influenced by the above-mentioned agrometeorological and economic variables.

2.2.5. AquaCrop

AquaCrop [38] evolved from the FAO-AZM model [37] to estimate crop biomass and yield but is very useful for assessing water management and irrigation [64]. The simulation model divides evapotranspiration (ET) into crop transpiration (Tr) and soil evaporation (Epsoil), obtaining the crop yield as a function of biomass and harvest index (HI). AquaCrop demands climate, soil, crop, and management information as input variables [38,65]. AquaCrop is composed of a database of 17 crops (cotton, maize, potato, quinoa, rice, soybean, sugarbeet, sunflower, tomato, wheat, barley, sugarcane, sorghum, teff, dry beans, cassava, and alfalfa), each one with its respective parameters derived from calibration/validation processes with field data. The model was tested on 46 crops, especially maize, wheat, and rice [66,67].

Table 1 summarizes the empirical and mechanistic models previously presented in terms of data requirements, spatial implementation, complexity, and application for different crops.

Table 1. Main aspects of the mechanistic and empirical models revised in this work.

Model Type	Name/Method	Data Requirements	Spatial Implementation	Complexity	Application
Mechanistic	DSSAT APSIM WOFOST FAO-AZM AquaCrop	Weather, soil, crop information, and management practices obtained in the field or from RS data.	Forcing, recalibration, updating	Highly complex in the data processing and model operation.	Specific crops

Table 1. *Cont.*

Model Type	Name/Method	Data Requirements	Spatial Implementation	Complexity	Application
Empirical	Linear Regression, SVM, ANN, and RF	Features extracted from field and RS data. For further information on features, please see the Supplementary Materials.	Implemented on a pixel basis.	(*) Linear Regression: less complex than ML in data processing and model operation. ML: Moderately complex in data processing and less complex in model operation.	Any crops

* These models handle massive volumes of data well, except for linear regression models, which tend to present underfitting in such cases. On the other hand, ML (SVM, ANN, RF) methods do not work well with a limited amount of input data.

3. Material and Methods

The methodological procedure adopted to perform the literature review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method [68]. To avoid unbiased results, PRISMA divides the selection into four steps: (1) identification of all papers that will be filtered from scientific databases; (2) screening of identified papers; (3) eligibility of papers; and (4) selection of papers that will compose the systematic review (Figure 1). We filtered all papers published between January 2017 and June 2023 in the scientific databases AGORA (FAO), Directory of Open Access Journals (DOAJ), Google Scholar, Multidisciplinary Digital Publishing Institute (MDPI), Nature, Science Direct (Elsevier), Taylor & Francis, Wiley Online Library, Scopus (Elsevier), and Web of Science (Clarivate Analytics).

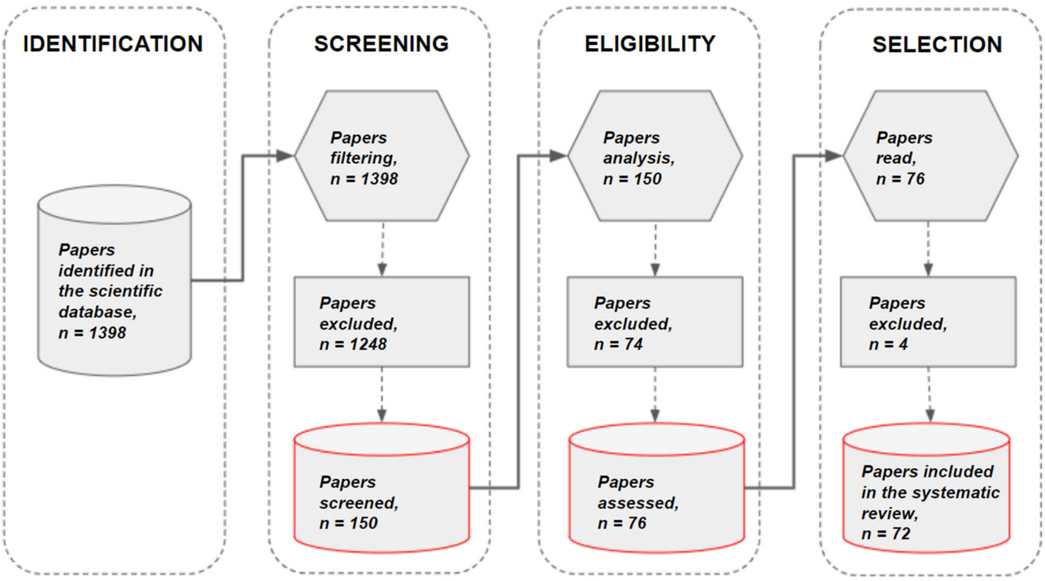

Figure 1. Flowchart with the steps taken to select the papers used in the survey.

In the identification step, we applied search criteria (Figure 2) to select papers published between 2017 and 2023 stored in the scientific databases. As a precondition, we analyzed only papers published in the English language. In the literature, yield estimation and yield prediction are sometimes used interchangeably and they may refer to past, present, and future timeframes, while yield forecasting is exclusively employed for yield assessment in future time horizons. In this manuscript, we strived to gather the greatest number of papers lying within the theme of our systematic review, regardless of their time settings. As the paper search is semiautomated, the input keywords need to be diverse to cope with the heterogeneity of terms found in such review topic.

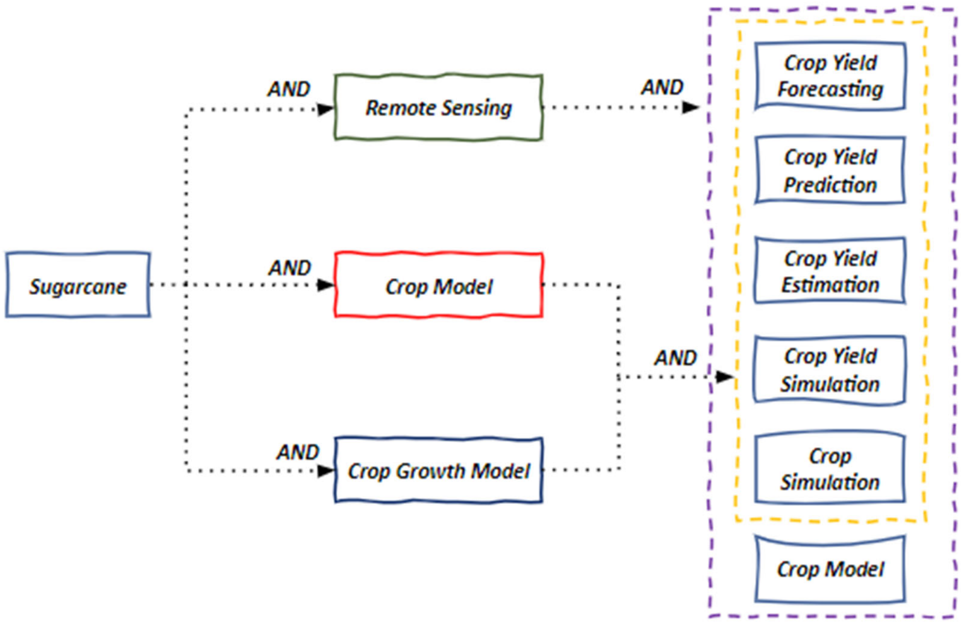

Figure 2. Flowchart with the logic of terms used in the papers survey.

We identified 1398 papers and stored them on the Mendeley platform [69]. In the screening step, we excluded 1248 of them, the core topic of which was not about sugarcane crop and yield estimate. In the eligibility step, we assessed whether empirical or mechanistic models were used to estimate sugarcane yield. Among the papers in which some empirical method was used, we selected those that used RS data and information from agrometeorological or agrometeorological–spectral models. Among the papers in which mechanistic models were used, we selected only those in which the model cited appeared in more than one paper in the reference set. During this process, we discarded papers in which the sugarcane yield estimation was based on drone data, because we focused only on satellite RS. In the eligibility step, we removed another 74 papers with scope and objectives contrasting with our search. Finally, in the selection step, an in-depth and critical reading guided the selection to compose the systematic review. This step removed 4 papers and proceeded with the remaining 72.

We created a database in Mendeley for each step. Then, the papers were analyzed using the VOS viewer program [70]. In addition, we calculated the influence of the selected papers based on [71], which considers the following relationship: Influence = number of citations/(base year—publication year). Here, 2023 was taken as the base year and the number of citations in Scopus were considered. According to [71], the influence metric aims to normalize the number of citations in the evaluated time window.

4. Results and Discussion

4.1. Overview

Sugarcane yield estimation using RS data by empirical or mechanistic models has sparked the interest of research groups across all continents (Figure 3). Considering statistics from the Food and Agriculture Organization [3], the predominance of papers identified by our methodology is in countries with large sugarcane production. Yet, studies have been conducted by research groups of countries with little or no production. The top five countries with papers identified by our methodology were China, the United States, Brazil, India, and Australia, with 260, 182, 163, 124, and 78 papers, respectively. Brazil has the most significant sugarcane production and ranks third in the number of publications. On the other hand, China leads the number of publications despite being only the third largest producer.

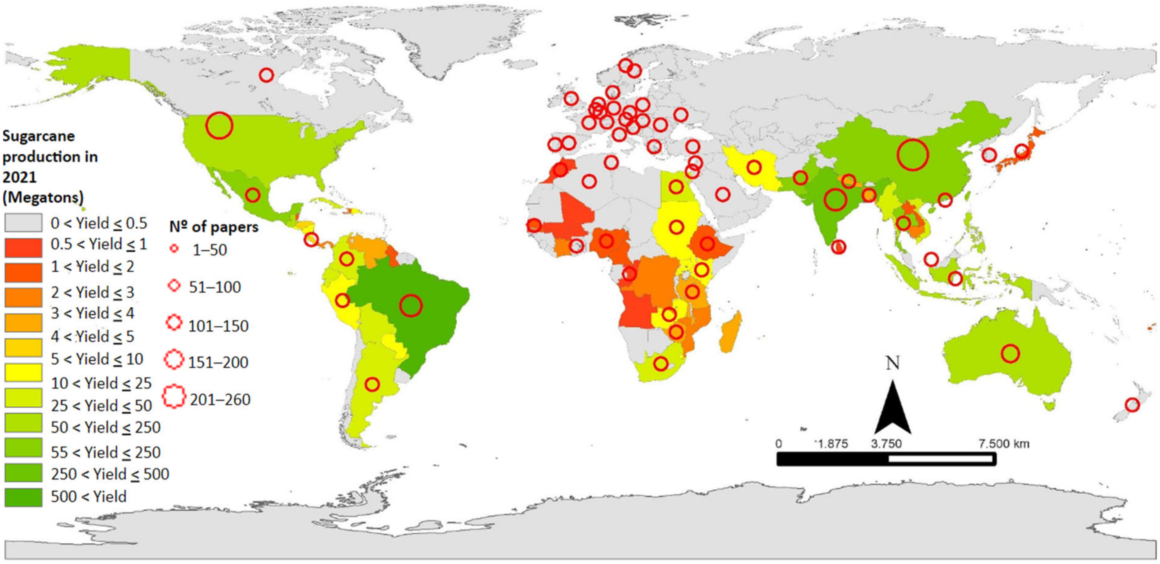

Figure 3. Spatial distribution of research groups responsible for the identified papers and production (megatons) of sugarcane for 2021 [3].

The spatial distribution of research groups accounting for the 72 selected papers shows a predominance of tropical regions (Figure 4), especially Brazil, India, and Australia, with 28, 12, and 7 papers, respectively. The commonly used mechanistic yield models for sugarcane estimation were DSSAT, APSIM, and FAO-AZM. DSSAT and APSIM were globally used, while FAO-AZM was used exclusively in Brazil. An explanation cited by the authors is that FAO-AZM does not demand so much input data and has a simpler methodology, and is suited for supporting decision making in countries with continental size, limited data coverage (e.g., agrometeorological), and low-scale mapping (e.g., soils), such as Brazil [58,61,72–76].

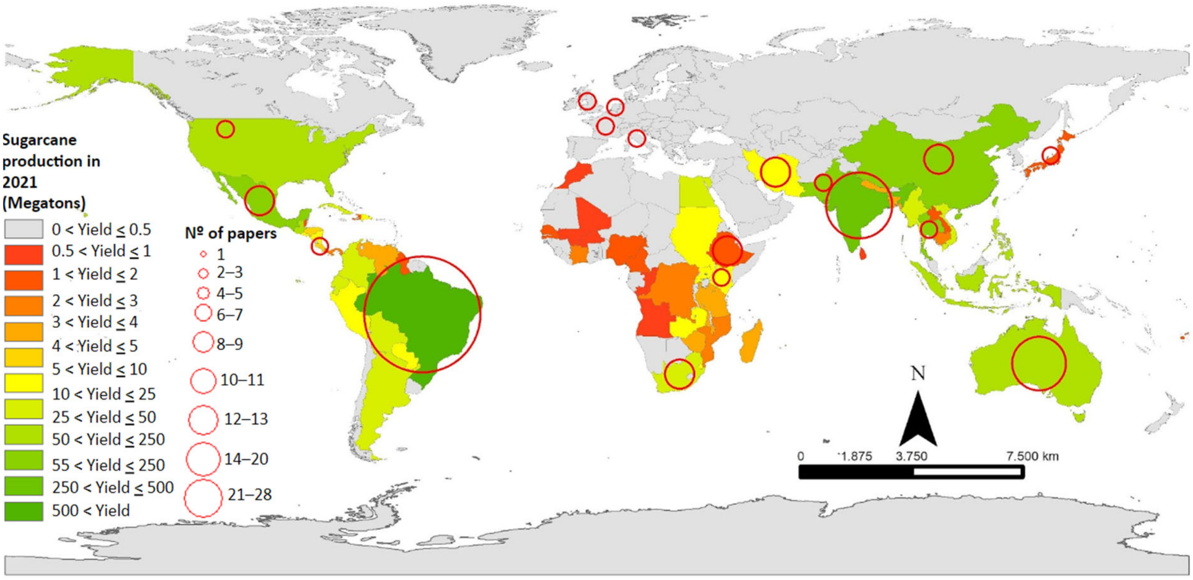

Figure 4. Spatial distribution of research groups responsible for the selected papers and production (megatons) of sugarcane for 2021 [3].

The total number of publications per year more than doubled from 2017 to 2021 (Figure 5). However, by the end of the survey (June 2023), 64 papers had been published. Generally, we selected approximately 6% of the publications for each year to compose the review. Regarding the selected papers, the publication proportion was highest in 2018 (21%) and lowest in 2019 (7%), and 2023 accounts for 8% of the total selected papers. Figure 6 presents the proportion of the selected papers with respect to the model type (mechanistic, empirical, and hybrid models) per analyzed year.

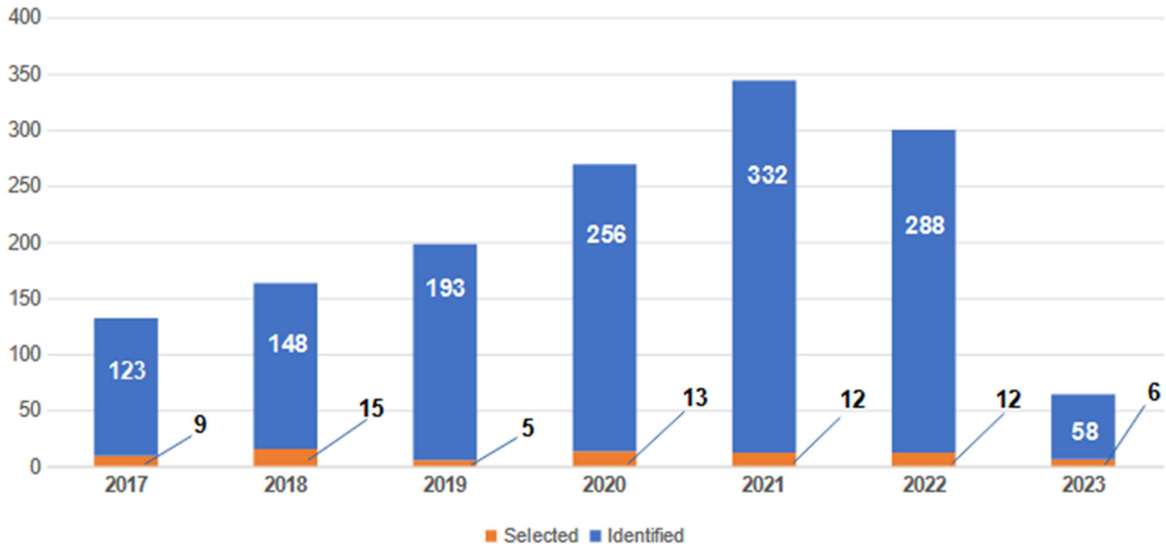

Figure 5. Total number of papers identified and selected for the studied period.

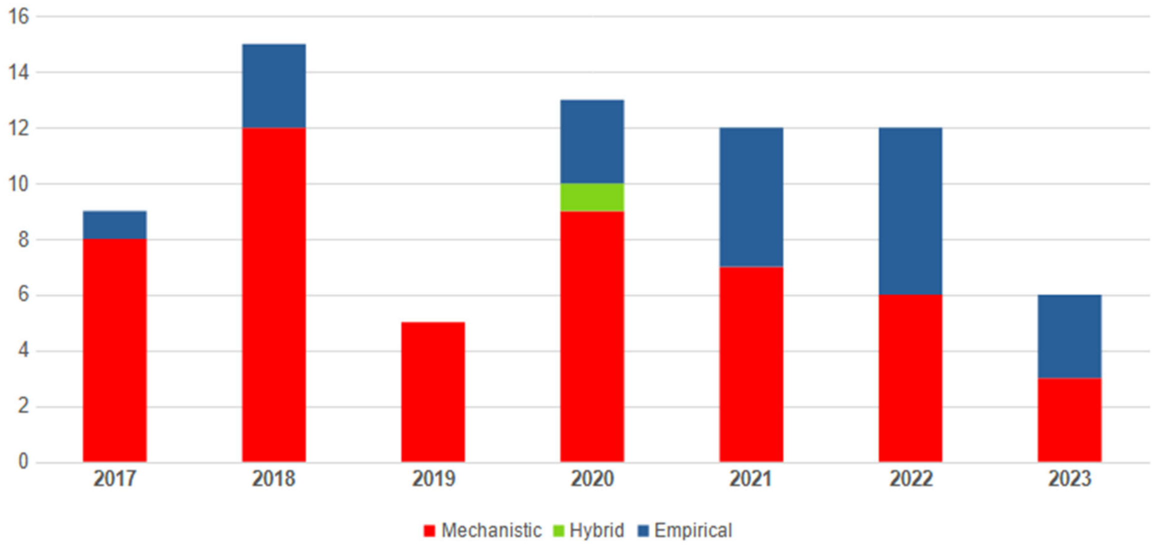

Figure 6. Total number of papers selected for the studied period per model type. The red color indicates the mechanistic models; the green color is associated with the hybrid model; and the blue color represents the empirical models.

The journals containing the greatest number of selected papers were the European Journal of Agronomy (8 papers), Field Crops Research (6), and the Journal of the Indian Society of Remote Sensing (6) (Figure 7). Eleven journals from different areas, ranging from irrigation and drainage to agricultural sciences or RS, comprised all the relevant literature for our review.

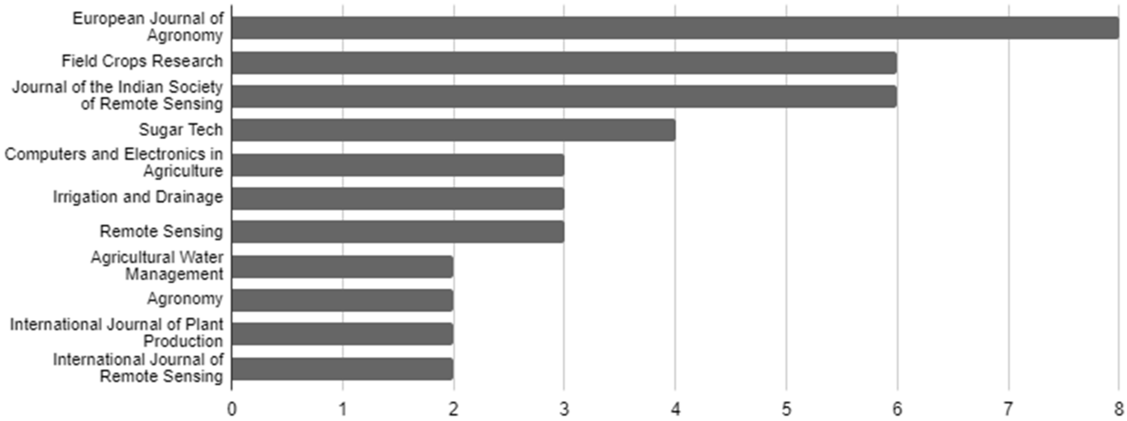

Figure 7. Journals where the selected papers were published.

Table 2 shows the 6 most influential papers out of the 72 that were selected to compose the review based on [71].

Table 2. Most influential selected papers.

Author (Model Type)	Title	Year	Journal	Influence
Monteiro et al. (2018) [76] (Mechanistic)	"Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation"	2018	*International Journal of Climatology*	14.20
Shendryk et al. (2021) [32] (Empirical)	"Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning"	2021	*Field Crops Research*	12.00
Dias and Sentelhas (2018) [73] (Mechanistic)	"Sugarcane yield gap analysis in Brazil—A multi-model approach for determining magnitudes and causes"	2018	*Science of the Total Environment*	11.80
Rahman and Robson (2020) [77] (Empirical)	"Integrating Landsat 8 and Sentinel-2 time series data for yield prediction of sugarcane crops at the block level"	2020	*Remote Sensing*	11.33
Fernandes et al. (2017) [28] (Empirical)	"Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble"	2017	*International Journal of Remote Sensing*	11.00
Canata et al. (2021) [22] (Empirical)	"Sugarcane yield mapping using high-resolution imagery data and machine learning technique"	2021	Remote Sensing	10.50

The study by [76] was considered the most influential; the authors compared the National Aeronautics and Space Administration/Prediction of World Wide Energy Resources (NASA/POWER) product with the weather stations of the National Institute of Meteorology (INMET) of Brazil, assessing the potential of NASA/POWER as meteorological input data to estimate potential and actual yields for sugarcane crops using the FAO-AZM model. They recommended NASA/POWER on the national and regional scales, but also emphasized that regional data would be better in areas with higher latitudes and elevations.

After, refs. [32,76] used satellite data from Sentinel-1 and Sentinel-2/MultiSpectral Instrument (S2/MSI) and ancillary data about climate, soil, and elevation to develop a predictive model for sugarcane yield estimation in Australia using machine learning. The data combination improved the detection of sugarcane yield (ton ha^{-1}), sugar yield (ton ha^{-1}), and commercial sugar yield (%) at field and mill area levels about four months before the harvest. Using a multi-model approach (FAO-AZM, DSSAT/CANEGRO, and APSIM-Sugarcane), ref. [73] concluded that water deficit and a poor crop management caused sugarcane yield gaps in Brazil, which can be mitigated by irrigation, deep soil profile, and drought-tolerant cultivars. Also in Brazil, ref. [22] used S2/MSI time series and RF to estimate sugarcane yield, achieving an RMSE of 4.63 ton ha^{-1} in a commercial site. Ref. [28], in turn, obtained sugarcane yield at a municipality level three months before the harvest using Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) time series. Ref. [77], in a sugarcane producer region in Australia, developed a linear regression model integrating Landsat 8 Operational Land Imager (L8/OLI) and S2/MSI time series to obtain accurate sugarcane yield predictions (RMSE = 11.33 ton ha^{-1}) at the block level.

4.2. Accuracy of the Methodologies Discussed in the Selected Papers

The modeling yielded results with higher discrepancy with respect to field observations derived from the FAO-AZM model. The mean RMSE was 25 ton ha^{-1}, and the values ranged from 13.8 ton ha^{-1} [75] to 46.1 ton ha^{-1} [60]. The most accurate model was AquaCrop, with an average RMSE of 0.96 ton ha^{-1} and RMSE values varying between 0.44 ton ha^{-1} [64] and 1.7 ton ha^{-1} [65]. In mechanistic models, there may be a need for calibrating one or more input variables.

In studies that used the AquaCrop model, the authors performed sensitivity analyses of input variables and model calibration. Ref. [65] carried out a model calibration to decrease the RMSE from 39.69 ton ha^{-1} to 1.6 ton ha^{-1}. In addition, they analyzed whether the different models for estimating sugarcane yields have statistically different means. The AquaCrop models present an average RMSE equal to the ones obtained by WOFOST and data mining (DM). Likewise, the FAO-AZM, DM/FAO-AZM, and APSIM models have statistically equal RMSE means. According to the results of Tukey's test, the RMSE of the DSSAT model is statistically different from FAO-AZM and AquaCrop but has a mean error equal to the other studied models (Figure 8). Figure 8 also shows the relation in the percentage of the average RMSE for each model evaluated versus the average yield for Brazil in 2021 according to [78]. The highest RMSEs were found in papers that used the FAO-AZM model with an average of 25 ton ha^{-1} corresponding to 35% of the average Brazilian yield. The average RMSE observed in the DM-derived models is 17% of the average sugarcane yield. However, such models need high-quality and representative reference data and imply higher computational costs.

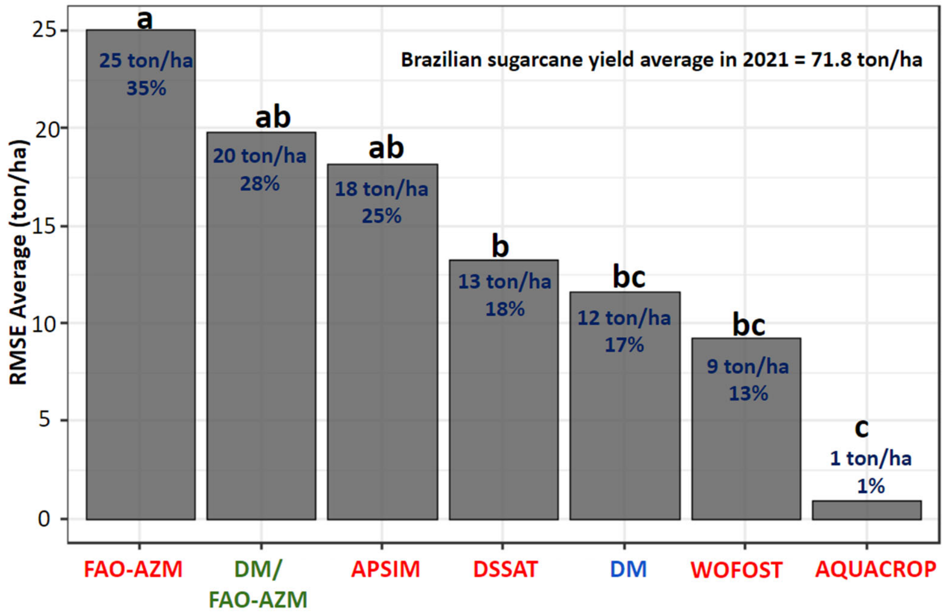

Figure 8. Mean RMSE (ton ha^{-1}) of the models in the selected papers. Equal letters indicate statistically equal means by the Tukey test at a 5% significance level. Below the letters, the approximate RMSE averages and percentages concerning the average sugarcane yield production for Brazil in 2021 according to IBGE [78] are presented. The red color indicates the mechanistic models; the green color is associated with the hybrid model; and the blue color represents the empirical models. Please see Table S1 in the Supplementary Materials for further information.

4.3. Attributes Used in the Selected Papers That Made Use of Statistical Modeling

Analyzing attributes used to generate predictive models for estimating sugarcane yield via RS, it was possible to identify trends, such as field data, based on spectral bands and vegetation indices, meteorological, synthetic-aperture radar (SAR), and terrain data, and other attribute types. Supplementary Tables S2–S7 show these attributes and their respective references. Many attributes stand out (Figure 9). Among the vegetation indices, the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) are the most used, followed by precipitation data, the number of cuts, and the variety used. About the cited vegetation indices, i.e., NDVI, EVI, Green Normalized Difference Vegetation Index (GNDVI), Soil Adjusted Vegetation Index (SAVI), Leaf Area Index (LAI), and Normalized Difference Water Index (NDWI) [79], they demand one or more spectral bands of RED, near infrared (NIR), shortwave infrared (SWIR) 1, and GREEN, mainly RED and NIR. Refs. [11,12,27] showed the importance of these spectral bands in the prediction of sugarcane yield, and the study of [27], in particular, cited the BLUE band. Moreover, these studies do not focus only on vegetation indices, but also on adopting spectral bands as attributes in the modeling process.

In Figure 9, variables derived from field information, spectral bands, vegetation indices, meteorological data, and terrain information are presented. Ref. [12] accomplished three comparative experiments, in which the best results were obtained when the sugarcane yield prediction was estimated by a model driven by satellite data, field information, and the harvest date (RMSE = 9.4 ton ha^{-1}). The second experiment used satellite and field data (RMSE = 9.9 ton ha^{-1}), and the third experiment, representing the worst-case estimates, regarded a model solely relying on field data (RMSE = 13.6 ton ha^{-1}). Ref. [31] used climate data (rainfall and temperature) and satellite products from a Moderate-Resolution Imaging Spectroradiometer (MODIS) (NDVI and EVI), and they concluded that satellite variables were helpful features in sugarcane yield models. According to them, just EVI could explain 43% of yield variations during the crop season.

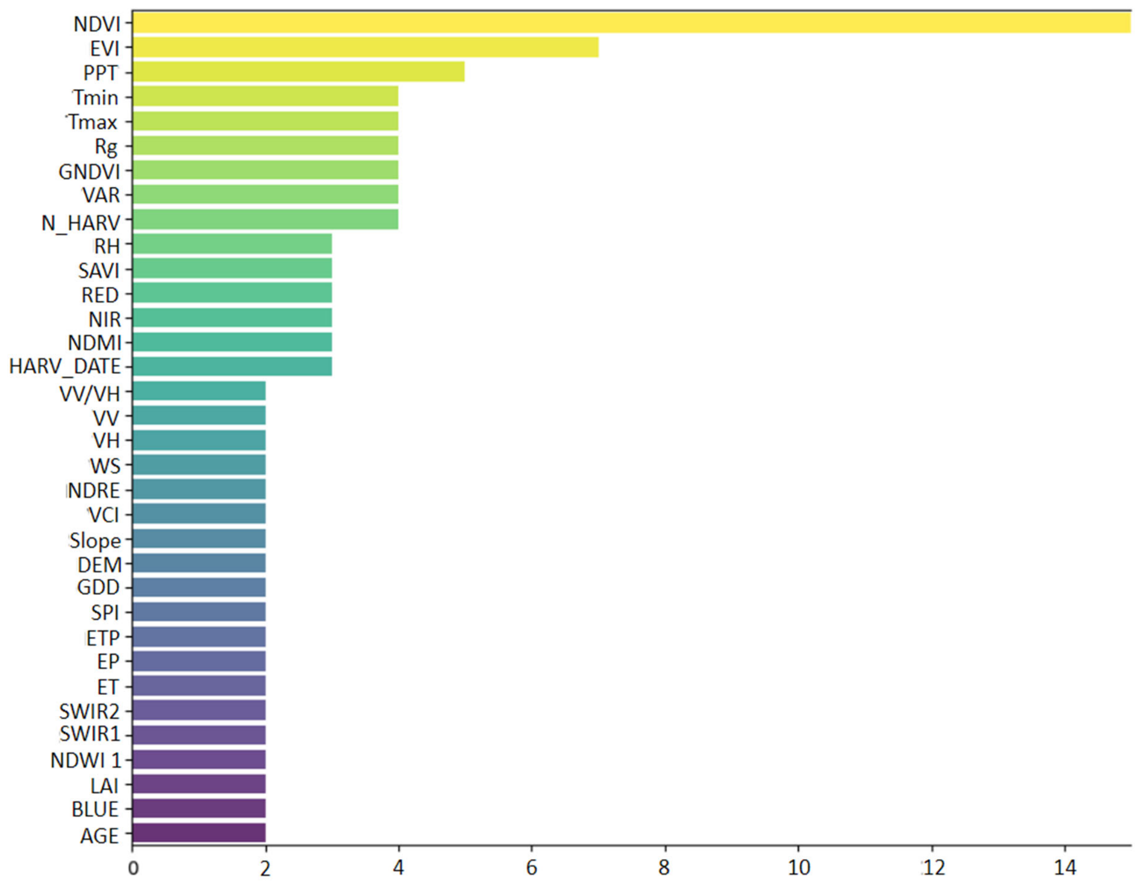

Figure 9. Frequency, considering a minimum of 2 selected papers, of the variables used to create sugarcane yield estimation models using data mining techniques. Note: we considered the acronym of each attribute and the number of selected papers in which it was used. The acronyms' meanings and respective references can be found in the Supplementary Materials.

In the selected list of papers, SAR data were also used in the work of [32], who integrated Sentinel-1 and Sentinel-2 image time series, obtaining predictions of sugarcane and sugar yield with an RMSE at mill level, respectively, of 4.6 ton ha^{-1} and 1 ton ha^{-1}. Ref. [80] integrated RS data from Sentinel-1 and Sentinel-2 and different machine learning models to estimate sugarcane yield at field level in India, presenting Normalized Root Mean Square Errors (NRMSE, %) of 18% and 32%. Ref. [23] also considered adopting satellite metrics from SAR in their future work.

Figure 10 shows the number of publications by satellite in the selected papers. Landsat, S2/MSI, and Terra-Aqua are the most used ones. This was expected, given their importance to RS research and ease of accessing data. The studies of [22,28,77] were already discussed in Section 4.1 and classified in Table 2 as some of the most influential papers in the selected list.

It is worth discussing the studies of [12,30], because both of them used Landsat images and the second one also evaluated Sentinel-2 image time series. In addition to satellite images, these studies tried to evaluate the contribution of agronomical, meteorological, terrain attributes (i.e., slope and elevation), and radiometric information to the model's performance. Ref. [12] understood that integrating agronomic, meteorological data, and Landsat image time series vegetation indices improved the yield model (attaining an RMSE less than 17 ton ha^{-1}), and the most important variables were the number of harvests and the Normalized Difference Moisture Index (NDMI) [81], which was computed using the NIR and SWIR bands.

In [30], the most important variables were related to the soil and terrain (Radiation Absorbed Dose, Gamma Radiometric Potassium, and a Digital Elevation Model), and the most important image index was the Normalized Difference Built-up Index (NDBI) [82], complying with the findings of [12].

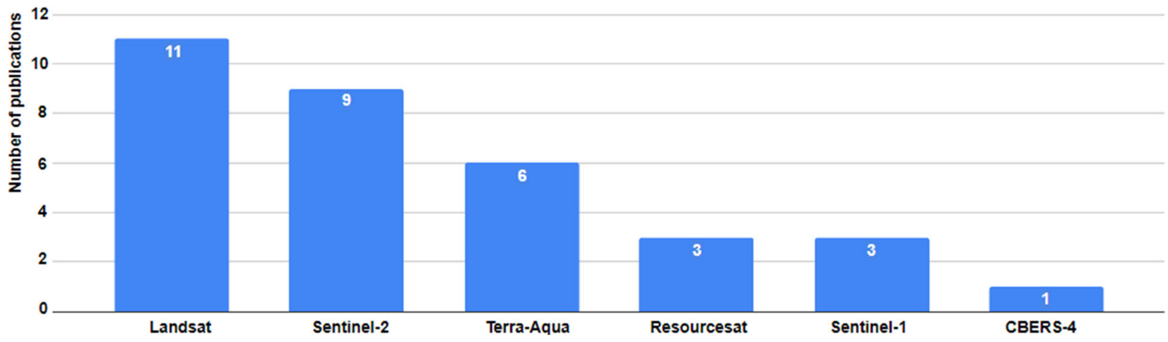

Figure 10. Number of publications separated by type of satellite used in the selected papers that generated yield models using data mining.

Another point to emphasize is the use of SAR data from Sentinel-1, even though it is found in only three of the selected articles. The use of such data is expected to increase in the coming years. One of these articles is the work of [32], which was particularly ranked as the second most influential article, as discussed in Section 4.1.

In addition to the Landsat, Terra-Aqua, Sentinel-1, or Sentinel-2 satellites, it is interesting to note the appearance of studies in the selected articles that made use of the China–Brazil Earth Resources Satellite CBERS-4 [24] and Resourcesat [23,25,83]. In [24], the authors understood that the use of NDVI images from CBERS-4 were better compared to NDVI from a field hyperspectral sensor (FieldSpec Spectroradiometer, Malvern Panalytical, Almelo, The Netherlands) to predict sugarcane yield, and NDVI from CBERS-4 in combination with information about leaf tissue nitrogen and phosphorus concentrations were useful to generate a yield model for sugarcane in Brazil.

Ref. [83] used Resourcesat images to obtain a crop map that was used in the extraction of the evaluated vegetation indices for the different studied districts in India. In their study, the authors predicted the sugarcane yield at district level in a statistically significant way. Both [23,25] also used Resourcesat images as input in their models. Ref. [23] derived a sugarcane yield model using Resourcesat images at mill level. They could estimate the yield of the crop two months before harvest with a deviation of less than 10% from the reported sugarcane production values. Furthermore, the authors emphasized the use of agrometeorological products and satellite metrics derived from SAR in future studies. Different from [23], ref. [25] obtained a sugarcane yield model based on the relationship between the farm scale values of yield and LAI. The LAI in the paper was derived from the relation of Resourcesat NDVI images with LAI values obtained using ground measurements using LP-80 AccuPAR Ceptometer, achieving an R^2 = 0.714.

4.4. Research Trends

Figure 11 presents a Venn diagram listing the keywords of the identified and selected papers. We used only papers with at least one citation and keywords with at least another two occurrences. Search terms used to identify papers were removed. Additionally, synonymous terms were standardized (e.g., NDVI and Normalized Difference Vegetation Index became just NDVI).

The overlapping area between the keywords in the identified and selected papers is small. In addition, the terms water stress, water deficit, and water productivity were highlighted in the selected papers and are most related to better understanding the yield gap or irrigation management in sugarcane crops. Along the same line, the terms climate change and variability were derived from studies related to explain how climate change affects sugarcane crop yields.

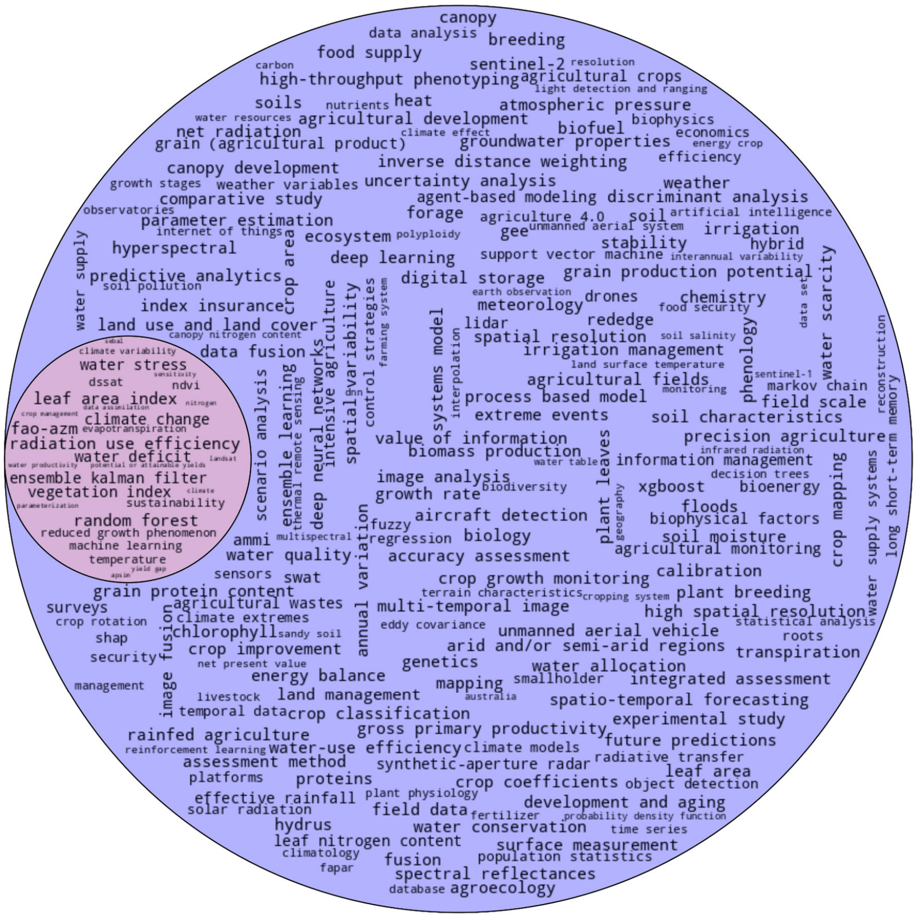

Figure 11. Venn diagram listing the keywords (minimum occurrence in 2 papers) found in the identified (at least 1 citation in Scopus) (in blue) and selected (in red) papers.

Another highlight is the term vegetation index, derived from RS data. An example is the study of [4] that used, for example, Sentinel-2-derived vegetation index time series and phenological metrics from NDVI as input in several regression models to estimate sugarcane yield in an irrigated region in Ethiopia. In the study, the sugarcane yield model, based on RF regression, presented an $R^2 = 0.84$ and up to 0.82 to estimate the sugar quantity. Also in the experiment, the authors mentioned that phenological metrics derived from NDVI were useful features to estimate sugarcane yield, and in the future, they want to integrate multisensor data from different satellites (e.g., Sentinel-1) or aerial image platforms.

The cited term ensemble Kalman filter (EnKF) is a data assimilation (DA) method that refers to the integration between satellite and terrestrial sensor data into crop simulation models. When the assimilation was performed using satellite data, it used information from LAI in the assessed crop. For example, ref. [55] performed the DA of LAI from Landsat-8/OLI and Sentinel-1A in the WOFOST model. In the same way, both [56,84] assimilated information from field sensors. Ref. [56] evaluated three different assimilation methods (forcing, calibration, and EnKF) assimilating soil water content (SWC) and LAI observations into the SWAP/WOFOST model with the aim to understand the contribution of these assimilated variables to improve sugarcane simulation. As a result, the authors pointed out that the assimilation of SWC and LAI contributed to the sugarcane simulation on SWAP/WOFOST, and the EnKF method was the most effective to estimate SWC, LAI development, and sugarcane yield. Ref. [84] evaluated the performance of sugarcane yield estimations on DSSAT/SAMUCA, coupling to the model LAI observations from three different DA methods (EnKF, ensemble smoother–ES, and weighted mean–WM). According to the authors, the sugarcane yield estimations based on assimilation methods presented better performances than using the model without DA, with the best results being obtained by the ES (RMSE = 20.27 ton ha^{-1}), followed by the EnKF (RMSE = 20.28 ton ha^{-1}) and WS (RMSE = 21.59 ton ha^{-1}) methods, respectively. Similarly, [84] point out that when the sugarcane cultivar in the field was different from the genotype-specific calibration used, they had a higher improvement in the model performance adopting EnKF and ES, while WS had the opposite results.

Regarding the models' names, FAO-AZM, DSSAT, and APSIM were very commonly used keywords, and this can be explained by the fact that a great number of the selected models that used mechanistic approaches made use of them. For example, in relation to the use of these models, we can cite the selected studies of [74,85], which, respectively, used FAO-AZM and DSSAT.

Ref. [74] compared the potential and attainable productivity estimated for sugarcane in a Brazilian municipality using three different meteorological datasets ((i) Xavier; (ii) NASA/POWER, and (iii) a meteorological station) as input for the FAO-AZM model. In conclusion, the authors recommend the use of the Xavier database in the prediction of productivity penalty for water deficit and the management of the studied crop, with an adjustment range varying between 63% and 88% with the meteorological station. Also, with up to 87% of adjustment to the reference data, NASA POWER can be used in the modeling process. Using DSSAT in India, ref. [85] simulated variety-wise sugarcane yield models and obtained a good agreement with the reference data. They also observed the genetic potential of the different sugarcane varieties in the studied region, named CoS-767 (lowest yield), CoSe-95422, CoS-8436, CoSe-92423, and CoSe-98231 (highest yield) in the three planting evaluated dates. Moreover, in the study, the authors highlighted that the sugarcane yield model is sensitive to the maximum temperature, minimum temperature, solar radiation, and CO_2 concentration level.

Figure 12 shows the keywords cluster network in the selected papers and their relationship. The WOFOST model is strongly related to RS, not just using meteorological data from RS products, but integrating into the model LAI derived, for instance, from Landsat or Sentinel-1 [55]. One reason for this can be the development of PCSE/WOFOST, implemented in the Python language, and its combination with Jupyter notebooks [36,53].

In the selected studies involving APSIM and sugarcane yield estimations [52,61,86–91] in Figure 12, there is a link between the model and the cluster with the keywords vegetation index and data cube, even though in the mentioned papers, none of them use vegetation indices products in the sugarcane simulation. We can explain this link as an opportunity for study, because LAI is a vegetation index estimated in APSIM and can be assimilated via RS or field data. In addition, the studies involving APSIM are also related with sugarcane yield, for example, the estimations in relation to the climate variability, temperature, and bio-economic modeling. The green cluster is a link between the yellow and red cluster. In this group we can see opportunities for study using DSSAT, APSIM, or WOFOST, for example, in evapotranspiration, water stress, water productivity, sugar, ethanol production, and using RS data as input.

Figure 13 shows the temporal evolution of the keywords during the study period. Studies that make use of machine learning, RS data, and their integration with crop simulation models are more recent. For example, the terms Sentinel-1A, data cube, EnKF, and Landsat integration with crop models (e.g., WOFOST) were more cited in recent studies in the selected papers, while those focused on ethanol production, temperature, carbon dioxide or storage, and water stress using DSSAT or APSIM models are older.

We can separate the clusters per year in two periods, one between 2017 and 2020, and another one from 2020 to 2023. In the first period, 2017–2020, 45% of the selected papers use mechanistic models, different from the second period (2020–2023), with 32% of the selected papers. The decrease in the number of articles that use mechanistic models to estimate sugarcane yield is due to the greater number of publications that seek to estimate the crop yield using empirical models based on machine learning and the greater availability of free RS data. In addition, we can observe more published papers about the use of RS data or field sensors as input for the mechanistic models [55,84,88,92].

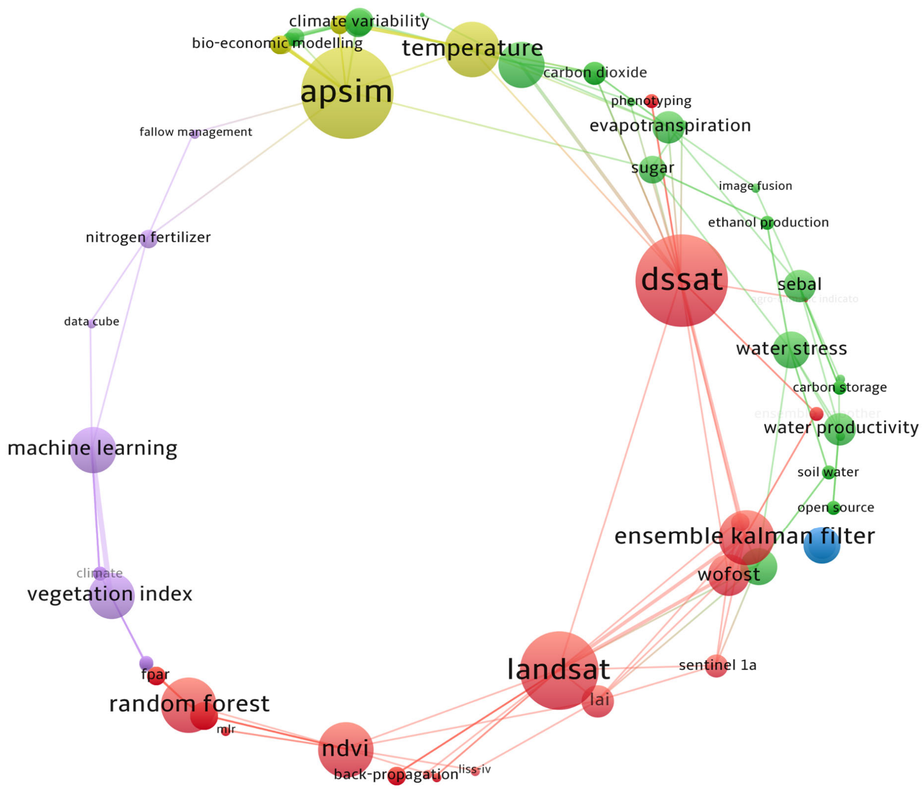

Figure 12. Cluster network of the selected papers keywords.

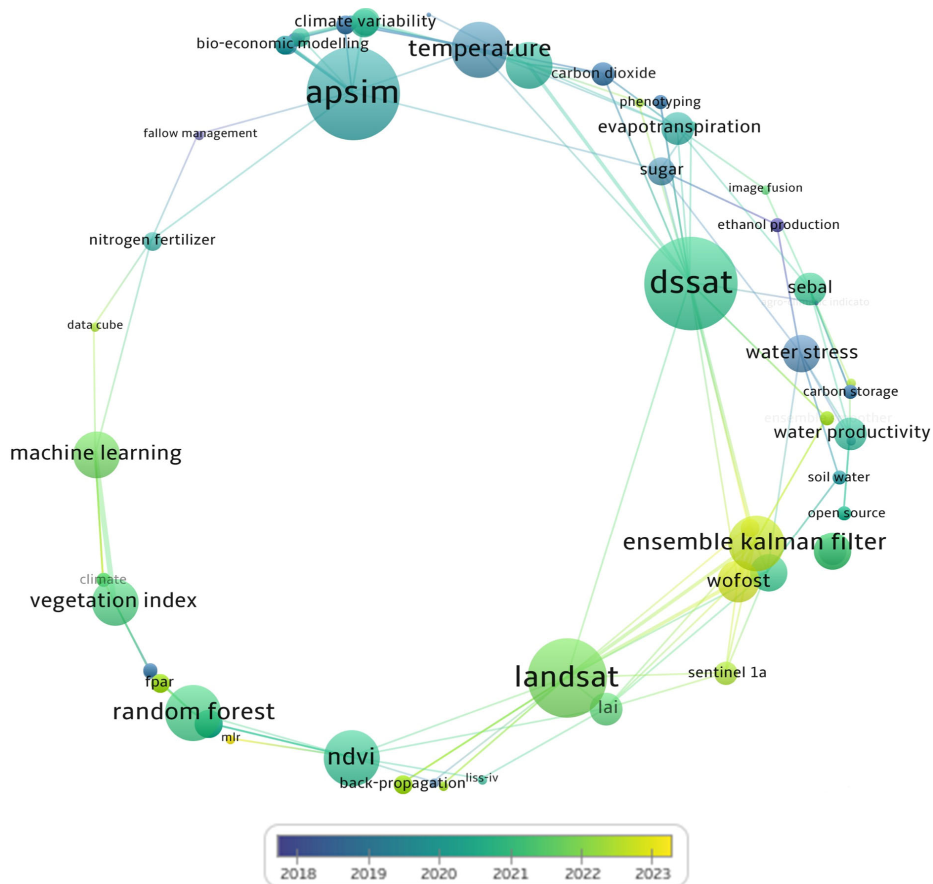

Figure 13. Time trend of the keywords in the selected papers.

Specifically concerning the use of top-edge artificial intelligence (AI) methods, we found no papers dealing with deep learning (DL) and related approaches. A unique paper employing DL for sugarcane yield prediction in particular [93] was published immediately after the upper threshold of our time frame, and hence, it was not included in our review. In this work, the authors proposed a novel hybrid CNN-Bi-LSTM_CYP (Convolutional Neural Network—Bidirectional Long Short-Term Memory—Crop Yield Prediction) deep learning-based approach that includes convolutional layers to extract the relevant spatial information in a sequence to Bi-LSTM layers, which recognize the phenological long-term and short-term bidirectional dependencies in the dataset to predict the sugarcane crop yield. It was concluded that the proposed approach was superior to other empirical models (either statistical or machine learning—ML) and even outperformed conventional DL approaches.

Other papers relating DL to yield estimation models in general have been published in recent years, and they were designed mainly for corn, soybean, and wheat [94–102], among other crops, and hence, they were excluded from our review. These studies are still limited in number, for this is still an incipient area in the field of yield estimation and forecasting, a point also noted in the systematic reviews of [13,103,104]. The minority of such studies regards the combination of ML and DL approaches [94,95,98,99], while most of them exclusively deals with DL methods, especially the most recent studies, considering that there is an ongoing trend to migrate to pure DL approaches, since this is the state-of-the-art in the field of empirical yield estimation models.

Finally, we ought to mention that DL presents several advantages for yield estimation and forecasts, like the ability to handle large and complex data; its independence on hand-engineered features; its capacity to deal with sequential data (time series); the ability to handle missing data and also non-linear relationships; its scalability, which allows for data to be deployed on cloud platforms and edge devices; its generalization ability, since DL methods are able to learn abstract and hierarchical representations of data; and its improved performance, as it is able to deliver highly accurate results.

Nevertheless, DL approaches are data-intensive and do not work well with limited data; they are dependent on human expertise for defining the optimal network architecture and the ideal settings for the parameterization and hyperparameterization processes, and demand high computational costs, since they require significant hardware resources, including powerful GPUs and large amounts of memory, which can be costly and time-consuming. In brief, DL remains a "black-box" model, as it is difficult to understand how the model makes predictions and identifies the factors that influence the predictions [13,105]. However, it is expected that empirical yield estimation models relying on DL, including sugarcane yield estimation models, will grow and gain increasing importance and visibility.

4.5. Limitations

Remote sensing-based sugarcane yield estimation models offer an opportunity to access information at the field scale and for extensive areas, with a low cost and in a well-timed manner, which is relevant for sugarcane producers to improve crop yield production, reduce costs, and help crop management and logistics [12]. However, these models differ in the degree of parameterization needed and the ability to simulate different cultivars and different stress conditions, hindering their application for sugarcane, given the lack of understanding of their capabilities, limitations, and difficulties, and the general lack of model credibility [106]. Ref. [32] concluded that most remote sensing-based models to predict sugarcane yield have been limited in their scope, often only detecting reasonably strong correlations between satellite imagery and sugarcane yield when considering yields and imagery averaged over large regions. They pointed out limitations related to their use over large areas and to provide early information. Also, high-quality field data are required for model development, and more effort is needed to parameterize and validate models to improve the reliability of crop simulations. Different physiological and growth parameters used in models vary among sugarcane cultivars, and therefore need to be estimated from data to appropriately predict growth and yields. Region-specific calibrations of models are also essential [106].

In line with this, different limitations emerged in the literature. Among the main ones, Refs. [72,73,107] discussed problems related to water deficit. As sugarcane yield increment varied among planting dates, because of the time and intensity of water deficit and the phenological phase, water deficit during the crop phases when leaves were expanding and stalks were growing caused higher impacts on the final yield than during the other phases. Ref. [92] assessed sugarcane for water-limited environments by considering the FAO-AZM model. They highlighted that climate-related limitations affect yield in response to annual variability, soil type, intensity, and duration of water deficit, consequently affecting the model. Ref. [11] concluded that one of the main limitations of regression methods for estimating crop yields is that they are only implemented for specific crop growth stages or certain geographic regions. Their results confirmed that the continuous addition of Earth observation data into the modeling (jointly with multivariate data, such as soil moisture, canopy nitrogen content, and evapotranspiration) can help to overcome these limitations. The authors cited that multi-source satellite data can improve information and overcome the limitations of data from individual sensors. In this line, ref. [22] demonstrated that the sugarcane yield mapping based on yield monitors was limited when compared to grains because of the high-resolution data (due to the slow traveling speed of the harvester and narrow row spacing), high biomass variability, and noise from the yield monitor system. The authors justify that these factors have guided the interest in using remote sensing-based methods to monitor sugarcane yield.

Even anticipating the crop forecast three months before the harvest, ref. [28] exposed the limitations of predicting sugarcane yield from moderate-resolution satellite images, averaged municipal yield data, and only NDVI spectral variables. The authors explained that more spectral and temporal variables with better resolutions improved estimations. This issue was also cited by [24]. Ref. [108] presented complementary limitations regarding the use of RS: spatial resolution; land cover noise of non-sugarcane land use, such as farm roads and irrigation and drainage infrastructures within a pixel; the number of cloud-free images on which the analysis and numerical interpolation are based; the time of day when images are taken; and the angle of image capture and its correction function. Ref. [47] indicate that the DSSAT/CANEGRO model has some limitations for crop simulations under rainfed conditions; therefore, further studies on the effect of drought on the development of LAI and canopy should be conducted. Ref. [109] presented limitations related to the population size, sugarcane area under cultivation, climatic factors, and exposure to drought, flood, heat, and cold waves. Furthermore, the CANEGRO-Sugarcane model did not consider the impact of pest infections, weeds, and diseases on a crop. Ref. [30] attested that current sugarcane yield forecasts predict a single, averaged yield value for an entire district or region and are poorer when late-season satellite images of the crop are excluded from the model. Ref. [1] discussed the natural challenges of reliably estimating trait parameter values from limited experimental data.

Limitations regarding climatic and intrinsic sugarcane variables also emerged. Ref. [106] pointed out limitations related to uncertainties associated with downscaled outputs of global climate models and changes in the standard patterns of temperatures and incidence of pests and diseases, factors that affect production. They concluded that it is urgent to integrate predicted new scenarios with the incidence of primary and secondary pests affecting sugarcane before estimating sugarcane growth and yields. Ref. [49] applied a sophisticated statistical downscaling method to generate daily climate data as inputs for crop models. The downscaling presented limitations, such as changes in the frequency of extreme events (e.g., droughts or heat waves). They discussed that models to predict the effects of future climate change on yield have presented simplified processes of plant growth and soil processes. Although necessary, simplifications limit the accuracy of the simulated crop response to environmental conditions, disregarding the effect of extreme weather events on soil conditions. Ref. [52] discussed the limitations of the biophysical APSIM-Sugar model, such as the simulated responses of leaf area expansion and radiation use efficiency to transpiration efficiency, the modeling of diurnal interactions between transpiration, photosynthesis, vapor pressure deficit, and water stress, and the non-inclusion of weeds, pest, and diseases on yield performances. Considering that soil temperature and moisture affect sugarcane tillering and physical, chemical, and biological processes, ref. [110] assessed the integrated use of evapotranspiration, soil moisture, and soil temperature data with requirements for dimension irrigation. They concluded that the inclusion of the effects of nutrient-limited environments in sugarcane growth is an emergent opportunity for future improvements of the SAMUCA model.

Ref. [73] also discussed how the lack of information regarding water deficit and biological, biochemical, and biophysical aspects linked to crop yield can affect yield gap estimation. Ref. [84] discussed limitations related to the update of only a few state variables, citing that this situation may affect the model integrity and cause undesired model states in some circumstances. Also, the authors discussed the need for improvements in model calibration, no interference of reducing factors, soil and climate characterization and climate data, and the use of state variables, such as aboveground biomass, plant height, soil moisture, canopy nitrogen accumulation, and canopy cover to enhance the model accuracy. Yet, they cited that further studies could explore the allometric relations between LAI with the number of stalks, stalk height, and other related crop variables to simultaneously update these variables without direct measurements. Using the APSIM sugarcane model, ref. [90] exposed limitations related to the rigor of model validation for large areas, not recommending validation for only a limited number of sites. They do not recommend disregarding the impact of pests and other disasters caused by meteorological factors in the real production process. In addition, they discussed spatial resolution issues and recommend assessing sugarcane cultivation impact on the ecological environment before estimating any model.

4.6. Future Research Directions

The above-mentioned limitations need to be addressed by the scientific community. Regarding potentialities that can fuel future research directions, water-related assessments, such as irrigation, water management, coupling crops models and hydrologic models, footprint, consumption, and use efficiency, focused on their relations with sugarcane yield, emerge as themes that can be further explored, as well as the relationship between crop yields and land use and land cover change, different planting windows, and plant phenology. In addition, since we can collect large spatiotemporal datasets for the same area, precision agriculture as well as the use of time series of crop yields in different years and their relations with several environmental variables and agricultural aptitude of the land (by considering zoning plans, crop calendars, and soil suitability) could also be explored. Another future line of research concerns the use of DL approaches to estimate sugarcane yield since these are the state-of-the-art in agriculture studies and particularly within the domain of empirical sugarcane yield modeling.

Due to the fact that mechanistic models are not developed to offer spatial information results, the RS data are very valuable to combine with these models. The assimilation of RS data in mechanistic models based on combining one or more satellite sensors is very promising. Thus, it would be interesting to verify the feasibility of generating sugarcane productivity results in these environments, identifying limitations, opportunities for improvement, and potentialities.

5. Conclusions

We observed the current state-of-the-art approaches related to sugarcane yield estimation using RS data in empirical or mechanistic models. These approaches have benefits and drawbacks that can be pondered, such as input data, region of interest, computational cost, and the available time to obtain information. As for outlooks, there are some related to sugarcane yields, such as studies related to water resources, climate change, land use and land cover change, irrigation management, carbon, and nitrogen use, integration of RS data with crop simulation models, cloud processing, and the impact of the spatial resolution and clouds on the estimations in mechanistic models.

Specifically in RS, the contribution of SAR and optical image time series and the use of spectral indices other than NDVI or LAI as input data ought to be investigated. Due to the difficulty in obtaining meteorological, soil, and other physical terrain data, it is recommended to assess the possibility of testing different remotely sensed datasets. In summary, there are many research opportunities related to sugarcane yield estimation, and each day, new demands that need to be addressed come on the scene.

The assimilation of RS data in mechanistic models and as features in empirical models is promising and will increase in the upcoming years following the development and increasing availability of free Earth observation data. As we discussed in this study, there are many themes of research and challenges to take advantage of the wide potential of this technology, including emerging lines of research in precision agriculture, model coupling, and AI. The use of mechanistic models without assimilation will continue in some regions and applications, but in others, e.g., when there is a necessity to obtain sugarcane yield on a global or regional scale, or even in places where it is difficult to obtain meteorological observations, RS integration will tend to be more commonly employed.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs16050863/s1, Table S1: Basic statistics on RMSE (ton ha^{-1}) of the selected papers' models, where DM means Data Mining; Table S2: Attributes based on field information; Table S3: Attributes based on spectral bands and vegetation indices; Table S4: Attributes based on meteorological data; Table S5: Attributes based on SAR data; Table S6: Attributes based on terrain information; Table S7: Other attribute types [111–116].

Author Contributions: N.R.d.F.e.S.: conceptualization, formal analysis, investigation, methodology, project administration, validation, visualization, writing—original draft, writing—review and editing; M.E.D.C.: conceptualization, formal analysis, investigation, methodology, supervision, validation, visualization, writing—original draft, writing—review and editing; A.C.d.S.L.: writing—review and editing; I.D.S.: writing—review and editing; C.M.d.A.: supervision, writing—review and editing; M.A.: conceptualization, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Coordination for the Improvement of Higher Education Personnel—CAPES), Brazil, Finance Code 001. The authors also are grateful to the São Paulo Research Foundation (FAPESP) through research grant N° 2021/07382-2 (Chaves, M.E.D.), and the Brazilian National Council for Scientific and Technological Development (CNPq) through research grants N° 310042/2021-6 (Sanches, I.D.), N° 311324/2021-5 (Almeida, C.M.), and N° 306334/2020-8 (Adami, M.).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hoffman, N.; Singels, A.; Patton, A.; Ramburan, S. Predicting Genotypic Differences in Irrigated Sugarcane Yield Using the Canegro Model and Independent Trait Parameter Estimates. *Eur. J. Agron.* **2018**, *96*, 13–21. [CrossRef]
2. Pagani, V.; Stella, T.; Guarneri, T.; Finotto, G.; Van Den Berg, M.; Marin, F.R.; Acutis, M.; Confalonieri, R. Forecasting Sugarcane Yields Using Agro-Climatic Indicators and Canegro Model: A Case Study in the Main Production Region in Brazil. *Agric. Syst.* **2017**, *154*, 45–52. [CrossRef]
3. FAOSTAT. FAO Global Statistical Yearbook, FAO Regional Statistical Yearbooks—2021. Available online: https://www.fao.org/faostat/en/#data/QCL (accessed on 22 August 2022).
4. Dimov, D.; Uhl, J.H.; Löw, F.; Seboka, G.N. Sugarcane Yield Estimation through Remote Sensing Time Series and Phenology Metrics. *Smart Agric. Technol.* **2022**, *2*, 100046. [CrossRef]
5. Estes, L.D.; Bradley, B.A.; Beukes, H.; Hole, D.G.; Lau, M.; Oppenheimer, M.G.; Schulze, R.; Tadross, M.A.; Turner, W.R. Comparing Mechanistic and Empirical Model Projections of Crop Suitability and Productivity: Implications for Ecological Forecasting. *Glob. Ecol. Biogeogr.* **2013**, *22*, 1007–1018. [CrossRef]

6. Kern, A.; Barcza, Z.; Marjanović, H.; Árendás, T.; Fodor, N.; Bónis, P.; Bognár, P.; Lichtenberger, J. Statistical Modelling of Crop Yield in Central Europe Using Climate Data and Remote Sensing Vegetation Indices. *Agric. For. Meteorol.* **2018**, *260–261*, 300–320. [CrossRef]
7. Hansen, J.W.; Jones, J.W. Scaling-up Crop Models for Climate Variability Applications. *Agric. Syst.* **2000**, *65*, 43–72. [CrossRef]
8. Huang, J.; Gómez-Dans, J.L.; Huang, H.; Ma, H.; Wu, Q.; Lewis, P.E.; Liang, S.; Chen, Z.; Xue, J.-H.; Wu, Y.; et al. Assimilation of Remote Sensing into Crop Growth Models: Current Status and Perspectives. *Agric. For. Meteorol.* **2019**, *276–277*, 107609. [CrossRef]
9. Knowling, M.J.; White, J.T.; Grigg, D.; Collins, C.; Westra, S.; Walker, R.R.; Pellegrino, A.; Ostendorf, B.; Bennett, B.; Alzraiee, A. Operationalizing Crop Model Data Assimilation for Improved On-Farm Situational Awareness. *Agric. For. Meteorol.* **2023**, *338*, 109502. [CrossRef]
10. Feng, X.; Tian, H.; Cong, J.; Zhao, C. A Method Review of the Climate Change Impact on Crop Yield. *Front. For. Glob. Chang.* **2023**, *6*, 1198186. [CrossRef]
11. Abebe, G.; Tadesse, T.; Gessesse, B. Combined Use of Landsat 8 and Sentinel 2A Imagery for Improved Sugarcane Yield Estimation in Wonji-Shoa, Ethiopia. *J. Indian. Soc. Remote Sens.* **2022**, *50*, 143–157. [CrossRef]
12. Luciano, A.C.D.S.; Picoli, M.C.A.; Duft, D.G.; Rocha, J.V.; Leal, M.R.L.V.; Le Maire, G. Empirical Model for Forecasting Sugarcane Yield on a Local Scale in Brazil Using Landsat Imagery and Random Forest Algorithm. *Comput. Electron. Agric.* **2021**, *184*, 106063. [CrossRef]
13. Muruganantham, P.; Wibowo, S.; Grandhi, S.; Samrat, N.H.; Islam, N. A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. *Remote Sens.* **2022**, *14*, 1990. [CrossRef]
14. Weiss, M.; Jacob, F.; Duveiller, G. Remote Sensing for Agricultural Applications: A Meta-Review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
15. Atzberger, C. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sens.* **2013**, *5*, 949–981. [CrossRef]
16. Chao, Z.; Liu, N.; Zhang, P.; Ying, T.; Song, K. Estimation Methods Developing with Remote Sensing Information for Energy Crop Biomass: A Comparative Review. *Biomass Bioenergy* **2019**, *122*, 414–425. [CrossRef]
17. Rembold, F.; Atzberger, C.; Savin, I.; Rojas, O. Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. *Remote Sens.* **2013**, *5*, 1704–1733. [CrossRef]
18. Hammer, R.G.; Sentelhas, P.C.; Mariano, J.C.Q. Sugarcane Yield Prediction Through Data Mining and Crop Simulation Models. *Sugar Tech.* **2020**, *22*, 216–225. [CrossRef]
19. Roberts, M.J.; Braun, N.O.; Sinclair, T.R.; Lobell, D.B.; Schlenker, W. Comparing and Combining Process-Based Crop Models and Statistical Models with Some Implications for Climate Change. *Environ. Res. Lett.* **2017**, *12*, 095010. [CrossRef]
20. Shi, W.; Tao, F.; Zhang, Z. A Review on Statistical Models for Identifying Climate Contributions to Crop Yields. *J. Geogr. Sci.* **2013**, *23*, 567–576. [CrossRef]
21. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop Yield Prediction Using Machine Learning: A Systematic Literature Review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]
22. Canata, T.F.; Wei, M.C.F.; Maldaner, L.F.; Molin, J.P. Sugarcane Yield Mapping Using High-Resolution Imagery Data and Machine Learning Technique. *Remote Sens.* **2021**, *13*, 232. [CrossRef]
23. Kumar, M.; Das, A.; Chaudhari, K.N.; Dutta, S.; Dakhore, K.K.; Bhattacharya, B.K. Field-Scale Assessment of Sugarcane for Mill-Level Production Forecasting Using Indian Satellite Data. *J. Indian. Soc. Remote Sens.* **2022**, *50*, 313–329. [CrossRef]
24. Pinheiro Lisboa, I.; Melo Damian, J.; Roberto Cherubin, M.; Silva Barros, P.; Ricardo Fiorio, P.; Cerri, C.; Eduardo Pellegrino Cerri, C. Prediction of Sugarcane Yield Based on NDVI and Concentration of Leaf-Tissue Nutrients in Fields Managed with Straw Removal. *Agronomy* **2018**, *8*, 196. [CrossRef]
25. Verma, A.K.; Garg, P.K.; Hari Prasad, K.S.; Dadhwal, V.K. Modelling of Sugarcane Yield Using LISS-IV Data Based on Ground LAI and Yield Observations. *Geocarto Int.* **2020**, *35*, 887–904. [CrossRef]
26. Nihar, A.; Patel, N.R.; Danodia, A. Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. *J. Indian. Soc. Remote Sens.* **2022**, *50*, 1519–1530. [CrossRef]
27. Singla, S.K.; Garg, R.D.; Dubey, O.P. Ensemble Machine Learning Methods to Estimate the Sugarcane Yield Based on Remote Sensing Information. *RIA* **2020**, *34*, 731–743. [CrossRef]
28. Fernandes, J.L.; Ebecken, N.F.F.; Esquerdo, J.C.D.M. Sugarcane Yield Prediction in Brazil Using NDVI Time Series and Neural Networks Ensemble. *Int. J. Remote Sens.* **2017**, *38*, 4631–4644. [CrossRef]
29. Krupavathi, K.; Raghubabu, M.; Mani, A.; Parasad, P.R.K.; Edukondalu, L. Field-Scale Estimation and Comparison of the Sugarcane Yield from Remote Sensing Data: A Machine Learning Approach. *J. Indian. Soc. Remote Sens.* **2022**, *50*, 299–312. [CrossRef]
30. Han, S.Y.; Bishop, T.F.A.; Filippi, P. Data-Driven, Early-Season Forecasts of Block Sugarcane Yield for Precision Agriculture. *Field Crops Res.* **2022**, *276*, 108360. [CrossRef]
31. Pignède, E.; Roudier, P.; Diedhiou, A.; N'Guessan Bi, V.H.; Kobea, A.T.; Konaté, D.; Péné, C.B. Sugarcane Yield Forecast in Ivory Coast (West Africa) Based on Weather and Vegetation Index Data. *Atmosphere* **2021**, *12*, 1459. [CrossRef]
32. Shendryk, Y.; Davy, R.; Thorburn, P. Integrating Satellite Imagery and Environmental Data to Predict Field-Level Cane and Sugar Yields in Australia Using Machine Learning. *Field Crops Res.* **2021**, *260*, 107984. [CrossRef]

33. Lobell, D.B.; Burke, M.B. On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change. *Agric. For. Meteorol.* **2010**, *150*, 1443–1452. [CrossRef]
34. Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Shelia, V.; Wilkens, P.W.; Singh, U.; White, J.W.; Asseng, S.; Lizaso, J.I.; Moreno, L.P.; et al. The DSSAT crop modeling ecosystem. In *Advances in Crop Modeling for a Sustainable Agriculture*; Boote, K.J., Ed.; Burleigh Dodds Science Publishing: Cambridge, UK, 2019; pp. 173–216.
35. Keating, B.A.; Robertson, M.J.; Muchow, R.C.; Huth, N.I. Modelling Sugarcane Production Systems I. Development and Performance of the Sugarcane Module. *Field Crops Res.* **1999**, *61*, 253–271. [CrossRef]
36. De Wit, A.; Boogaard, H.; Fumagalli, D.; Janssen, S.; Knapen, R.; Van Kraalingen, D.; Supit, I.; Van Der Wijngaart, R.; Van Diepen, K. 25 Years of the WOFOST Cropping Systems Model. *Agric. Syst.* **2019**, *168*, 154–167. [CrossRef]
37. Doorenbos, J.; Kassam, A.H.; Bentvelsen, C.I.M. *Yield Response to Water, FAO Irrigation and Drainage Paper*; Food and Agriculture Organization of the United Nations: Rome, Italy, 1979.
38. Steduto, P.; Hsiao, T.C.; Raes, D.; Fereres, E. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts Underlying Principles. *Agron. J.* **2009**, *101*, 426–437. [CrossRef]
39. Kiniry, J.R.; Williams, J.R.; Gassman, P.W.; Debaeke, P. A General, Process-Oriented Model for Two Competing Plant Species. *Trans. ASAE* **1992**, *35*, 801–810. [CrossRef]
40. FAO. Land & Water—CropWat. Food and Agriculture Organization of the United Nations. Available online: https://www.fao.org/land-water/databases-and-software/cropwat/en/ (accessed on 14 August 2022).
41. Marin, F.R.; Jones, J.W. Process-Based Simple Model for Simulating Sugarcane Growth and Production. *Sci. Agric.* **2014**, *71*, 1–16. [CrossRef]
42. Inman-Bamber, N.G. A Growth Model for Sugar-Cane Based on a Simple Carbon Balance and the CERES-Maize Water Balance. *S. Afr. J. Plant Soil.* **1991**, *8*, 93–99. [CrossRef]
43. Singels, A.; Bezuidenhout, C.N. A New Method of Simulating Dry Matter Partitioning in the Canegro Sugarcane Model. *Field Crops Res.* **2002**, *78*, 151–164. [CrossRef]
44. Singels, A.; Jones, M.; Van Der Berg, M. *DSSAT v.4.5 DSSAT/CANEGRO: Sugarcane Plant Module: Scientific Documentation*; South African Sugarcane Research Institute, International Consortium for Sugarcane Modeling: Mount Edgecombe, South Africa, 2008.
45. Nadeem, M.; Nazer Khan, M.; Abbas, G.; Fatima, Z.; Iqbal, P.; Ahmed, M.; Ali Raza, M.; Rehman, A.; Ul Haq, E.; Hayat, A.; et al. Application of CSM-CANEGRO Model for Climate Change Impact Assessment and Adaptation for Sugarcane in Semi-Arid Environment of Southern Punjab, Pakistan. *Int. J. Plant Prod.* **2022**, *16*, 443–466. [CrossRef]
46. Pokhrel, P.; Rajan, N.; Jifon, J.; Rooney, W.; Jessup, R.; Da Silva, J.; Enciso, J.; Attia, A. Evaluation of the DSSAT-CANEGRO Model for Simulating the Growth of Energy Cane (*Saccharum* spp.), a Biofuel Feedstock Crop. *Crop Sci.* **2022**, *62*, 466–478. [CrossRef]
47. Leonaldo De Souza, A.L.D.C.; Leonaldo de Souza, S.; Santos Almeida, A.C.; Lyra, G.B.; Iedo Teodoro, G.B.L.; Ivomberg, R.A.F., Jr.; Rodrigues Santos, D.M. Sugarcane Productivity Simulation under Different Planting Times by DSSAT/CANEGRO Model in Alagoas, Brazil. *Emir. J. Food Agric.* **2018**, *30*, 190–198. [CrossRef]
48. Marin, F.R.; Thorburn, P.J.; Nassif, D.S.P.; Costa, L.G. Sugarcane Model Intercomparison: Structural Differences and Uncertainties under Current and Potential Future Climates. *Environ. Model. Softw.* **2015**, *72*, 372–386. [CrossRef]
49. Ruan, H.; Feng, P.; Wang, B.; Xing, H.; O'Leary, G.J.; Huang, Z.; Guo, H.; Liu, D.L. Future Climate Change Projects Positive Impacts on Sugarcane Productivity in Southern China. *Eur. J. Agron.* **2018**, *96*, 108–119. [CrossRef]
50. Lisson, S.N.; Robertson, M.J.; Keating, B.A.; Muchow, R.C. Modelling Sugarcane Production Systems. *Field Crops Res.* **2000**, *68*, 31–48. [CrossRef]
51. Dias, H.B.; Sentelhas, P.C. Drying-Off Periods for Irrigated Sugarcane to Maximize Sucrose Yields Under Brazilian Conditions. *Irrig. Drain.* **2018**, *67*, 527–537. [CrossRef]
52. An-Vo, D.-A.; Mushtaq, S.; Reardon-Smith, K.; Kouadio, L.; Attard, S.; Cobon, D.; Stone, R. Value of Seasonal Forecasting for Sugarcane Farm Irrigation Planning. *Eur. J. Agron.* **2019**, *104*, 37–48. [CrossRef]
53. De Wit, A.; Boogaard, H. A Gentle Introduction to WOFOST. WUR. 2021. Available online: https://www.wur.nl/en/research-results/research-institutes/environmental-research/facilities-tools/software-models-and-databases/wofost/documentation-wofost.htm (accessed on 22 August 2023).
54. Boogaard, H.L.; De Wit, A.J.W.; Te Roller, J.A.; Van Diepen, C.A. WOFOST Control Centre 2.1 and WOFOST 7.1.7. In *User's Guide for the WOFOST Control Centre, 2*; Alterra, Wageningen University & Research Centre: Wageningen, The Netherlands, 2014.
55. Abebe, G.; Tadesse, T.; Gessesse, B. Assimilation of Leaf Area Index from Multisource Earth Observation Data into the WOFOST Model for Sugarcane Yield Estimation. *Int. J. Remote Sens.* **2022**, *43*, 698–720. [CrossRef]
56. Hu, S.; Shi, L.; Huang, K.; Zha, Y.; Hu, X.; Ye, H.; Yang, Q. Improvement of Sugarcane Crop Simulation by SWAP-WOFOST Model via Data Assimilation. *Field Crops Res.* **2019**, *232*, 49–61. [CrossRef]
57. Shi, L.; Hu, S.; Zha, Y. Estimation of Sugarcane Yield by Assimilating UAV and Ground Measurements Via Ensemble Kalman Filter. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 8816–8819.
58. Cardozo, N.P.; De Oliveira Bordonal, R.; La Scala, N. Sustainable Intensification of Sugarcane Production under Irrigation Systems, Considering Climate Interactions and Agricultural Efficiency. *J. Clean. Prod.* **2018**, *204*, 861–871. [CrossRef]
59. Monteiro, L.A.; Sentelhas, P.C. Potential and Actual Sugarcane Yields in Southern Brazil as a Function of Climate Conditions and Crop Management. *Sugar Tech.* **2014**, *16*, 264–276. [CrossRef]

60. Caetano, J.M.; Casaroli, D. Sugarcane Yield Estimation for Climatic Conditions in the State of Goiás. *Rev. Ceres* **2017**, *64*, 298–306. [CrossRef]
61. Dias, H.B.; Sentelhas, P.C. Evaluation of Three Sugarcane Simulation Models and Their Ensemble for Yield Estimation in Commercially Managed Fields. *Field Crops Res.* **2017**, *213*, 174–185. [CrossRef]
62. Marin, F.R.; Carvalho, G.L.D. Spatio-Temporal Variability of Sugarcane Yield Efficiency in the State of São Paulo, Brazil. *Pesq. Agropec. Bras.* **2012**, *47*, 149–156. [CrossRef]
63. Figueira, S.R.F.; Rolim, G.D.S. Economic and Agrometeorological Modeling of Sugarcane Productivity in São Paulo State, Brazil. *Agron. J.* **2020**, *112*, 4836–4848. [CrossRef]
64. Farooq, N.; Gheewala, S.H. Assessing the Impact of Climate Change on Sugarcane and Adaptation Actions in Pakistan. *Acta Geophys.* **2020**, *68*, 1489–1503. [CrossRef]
65. Bahmani, O.; Eghbalian, S. Simulating the Response of Sugarcane Production to Water Deficit Irrigation Using the AquaCrop Model. *Agric. Res.* **2018**, *7*, 158–166. [CrossRef]
66. FAO. AquaCrop Version 7.0, Reference Manual, Annexes. Available online: https://www.fao.org/3/br244e/br244e.pdf/ (accessed on 26 July 2023).
67. FAO. *The AquaCrop Model—Enhancing Crop Water Productivity*; FAO: Rome, Italy, 2021; ISBN 9789251352229.
68. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ* **2009**, *339*, b2535. [CrossRef] [PubMed]
69. Mendeley. Mendeley Reference Manager—2023. Available online: https://www.mendeley.com/reference-management/reference-manager (accessed on 20 June 2023).
70. Van Eck, N.J.; Waltman, L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef] [PubMed]
71. Herrmann, P.B.; Nascimento, V.F.; Freitas, M.W.D.D. Sensoriamento Remoto Aplicado à Análise de Fogo Em Formações Campestres: Uma Re-Visão Sistemática. *Rev. Bras. Cartogr.* **2022**, *74*, 437–458. [CrossRef]
72. Dias, H.B.; Sentelhas, P.C. Dimensioning the Impact of Irrigation on Sugarcane Yield in Brazil. *Sugar Tech.* **2019**, *21*, 29–37. [CrossRef]
73. Dias, H.B.; Sentelhas, P.C. Sugarcane Yield Gap Analysis in Brazil—A Multi-Model Approach for Determining Magnitudes and Causes. *Sci. Total Environ.* **2018**, *637–638*, 1127–1136. [CrossRef] [PubMed]
74. Dos Anjos, J.C.R.; Casaroli, D.; Alves Júnior, J.; Paixão, J.S.; Silva, G.C.D.; Moraes, J.M.F.; Anjos Neto, J.G.D.; Medrado, L.D.C.; Almeida, F.D.P.; Santos, D.P. Productivity and Penalty in Sugarcane from Three Meteorological Databases in Jataí-GO. *Sci. Elec. Arch.* **2023**, *16*. [CrossRef]
75. Monteiro, L.A.; Sentelhas, P.C. Sugarcane Yield Gap: Can It Be Determined at National Level with a Simple Agrometeorological Model? *Crop Pasture Sci.* **2017**, *68*, 272. [CrossRef]
76. Monteiro, L.A.; Sentelhas, P.C.; Pedra, G.U. Assessment of NASA/POWER Satellite-based Weather System for Brazilian Conditions and Its Impact on Sugarcane Yield Simulation. *Intl J. Climatol.* **2018**, *38*, 1571–1581. [CrossRef]
77. Rahman, M.M.; Robson, A. Integrating Landsat-8 and Sentinel-2 Time Series Data for Yield Prediction of Sugarcane Crops at the Block Level. *Remote Sens.* **2020**, *12*, 1313. [CrossRef]
78. Brazilian Institute of Geography and Statistics—IBGE. Municipal Agricultural Production (PAM)—2021. Available online: https://sidra.ibge.gov.br/pesquisa/pam/tabelas (accessed on 22 August 2023).
79. McFEETERS, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
80. Das, A.; Kumar, M.; Kushwaha, A.; Dave, R.; Dakhore, K.K.; Chaudhari, K.; Bhattacharya, B.K. Machine Learning Model Ensemble for Predicting Sugarcane Yield through Synergy of Optical and SAR Remote Sensing. *Remote Sens. Appl. Soc. Environ.* **2023**, *30*, 100962. [CrossRef]
81. Wilson, E.H.; Sader, S.A. Detection of Forest Harvest Type Using Multiple Dates of Landsat TM Imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [CrossRef]
82. Zha, Y.; Gao, J.; Ni, S. Use of Normalized Difference Built-up Index in Automatically Mapping Urban Areas from TM Imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]
83. Dubey, S.K.; Gavli, A.S.; Yadav, S.K.; Sehgal, S.; Ray, S.S. Remote Sensing-Based Yield Forecasting for Sugarcane (*Saccharum officinarum* L.) Crop in India. *J. Indian. Soc. Remote Sens.* **2018**, *46*, 1823–1833. [CrossRef]
84. Fattori Junior, I.M.; Dos Santos Vianna, M.; Marin, F.R. Assimilating Leaf Area Index Data into a Sugarcane Process-Based Crop Model for Improving Yield Estimation. *Eur. J. Agron.* **2022**, *136*, 126501. [CrossRef]
85. Verma, A.K.; Garg, P.K.; Prasad, K.S.H.; Dadhwal, V.K. Variety-Specific Sugarcane Yield Simulations and Climate Change Impacts on Sugarcane Yield Using DSSAT-CSM-CANEGRO Model. *Agric. Water Manag.* **2023**, *275*, 108034. [CrossRef]
86. Dias, H.B.; Inman-Bamber, G.; Bermejo, R.; Sentelhas, P.C.; Christodoulou, D. New APSIM-Sugar Features and Parameters Required to Account for High Sugarcane Yields in Tropical Environments. *Field Crops Res.* **2019**, *235*, 38–53. [CrossRef]
87. Dias, H.B.; Inman-Bamber, G.; Sentelhas, P.C.; Everingham, Y.; Bermejo, R.; Christodoulou, D. High-Yielding Sugarcane in Tropical Brazil—Integrating Field Experimentation and Modelling Approach for Assessing Variety Performances. *Field Crops Res.* **2021**, *274*, 108323. [CrossRef]

88. Dias, H.B.; Sentelhas, P.C. Assessing the Performance of Two Gridded Weather Data for Sugarcane Crop Simulations with a Process-Based Model in Center-South Brazil. *Int. J. Biometeorol.* **2021**, *65*, 1881–1893. [CrossRef] [PubMed]
89. Dias, H.B.; Sentelhas, P.C.; Inman-Bamber, G.; Everingham, Y. Sugarcane Yield Future Scenarios in Brazil as Projected by the APSIM-Sugar Model. *Ind. Crops Prod.* **2021**, *171*, 113918. [CrossRef]
90. Peng, T.; Fu, J.; Jiang, D.; Du, J. Simulation of the Growth Potential of Sugarcane as an Energy Crop Based on the APSIM Model. *Energies* **2020**, *13*, 2173. [CrossRef]
91. Sexton, J.; Everingham, Y.L.; Inman-Bamber, G. A Global Sensitivity Analysis of Cultivar Trait Parameters in a Sugarcane Growth Model for Contrasting Production Environments in Queensland, Australia. *Eur. J. Agron.* **2017**, *88*, 96–105. [CrossRef]
92. Paixão, J.S.; Casaroli, D.; Dos Anjos, J.C.R.; Alves Júnior, J.; Evangelista, A.W.P.; Dias, H.B.; Battisti, R. Optimizing Sugarcane Planting Windows Using a Crop Simulation Model at the State Level. *Int. J. Plant Prod.* **2021**, *15*, 303–315. [CrossRef]
93. Saini, P.; Nagpal, B.; Garg, P.; Kumar, S. CNN-BI-LSTM-CYP: A Deep Learning Approach for Sugarcane Yield Prediction. *Sustain. Energy Technol. Assess.* **2023**, *57*, 103263. [CrossRef]
94. Agarwal, S.; Tarar, S. A Hybrid Approach for Crop Yield Prediction Using Machine Learning and Deep Learning Algorithms. *J. Phys. Conf. Ser.* **2021**, *1714*, 012012. [CrossRef]
95. Bi, L. Deep Learning Approaches for Yield Prediction and Crop Disease Recognition. Ph.D. Thesis, Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, USA, 2022.
96. Cunha, R.L.D.F.; Silva, B. Estimating Crop Yields with Remote Sensing and Deep Learning. In Proceedings of the 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 22–26 March 2020; pp. 273–278.
97. Kaneko, A.; Kennedy, T.; Mei, L.; Sintek, C.; Burke, M.; Ermon, S.; Lobell, D. Deep Learning for Crop Yield Prediction in Africa. In Proceedings of the the International Conference on Machine Learning AI for Social Good, Long Beach, CA, USA, 9–15 June 2019; pp. 33–37.
98. Shetty, S.A.; Padmashree, T.; Sagar, B.M.; Cauvery, N.K. Performance Analysis on Machine Learning Algorithms with Deep Learning Model for Crop Yield Prediction. In *Data Intelligence and Cognitive Informatics*; Jeena Jacob, I., Kolandapalayam Shanmugam, S., Piramuthu, S., Falkowski-Gilski, P., Eds.; Springer: Singapore, 2021; pp. 739–750, ISBN 9789811585296.
99. Srikamdee, S.; Rimcharoen, S.; Leelathakul, N. Sugarcane Yield and Quality Forecasting Models: Adaptive ES vs. Deep Learning. In In Proceedings of the 2nd International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Phuket, Thailand, 24 March 2018; pp. 6–11.
100. Vignesh, K.; Askarunisa, A.; Abirami, A.M. Optimized Deep Learning Methods for Crop Yield Prediction. *Comput. Syst. Sci. Eng.* **2023**, *44*, 1051–1067. [CrossRef]
101. Wang, A.X.; Tran, C.; Desai, N.; Lobell, D.; Ermon, S. Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. In Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, Menlo Park/San Jose, CA, USA, 20 June 2018; pp. 1–5.
102. Zhu, Y.; Wu, S.; Qin, M.; Fu, Z.; Gao, Y.; Wang, Y.; Du, Z. A Deep Learning Crop Model for Adaptive Yield Estimation in Large Areas. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *110*, 102828. [CrossRef]
103. Joshi, A.; Pradhan, B.; Gite, S.; Chakraborty, S. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sens.* **2023**, *15*, 2014. [CrossRef]
104. Oikonomidis, A.; Catal, C.; Kassahun, A. Deep Learning for Crop Yield Prediction: A Systematic Literature Review. *N. Z. J. Crop Hortic. Sci.* **2023**, *51*, 1–26. [CrossRef]
105. Ahmed, S.F.; Alam, M.S.B.; Hassan, M.; Rozbu, M.R.; Ishtiak, T.; Rafa, N.; Mofijur, M.; Shawkat Ali, A.B.M.; Gandomi, A.H. Deep Learning Modelling Techniques: Current Progress, Applications, Advantages, and Challenges. *Artif. Intell. Rev.* **2023**, *56*, 13521–13617. [CrossRef]
106. Baez-Gonzalez, A.; Kiniry, J.; Meki, M.; Williams, J.; Alvarez-Cilva, M.; Ramos-Gonzalez, J.; Magallanes-Estala, A.; Zapata-Buenfil, G. Crop Parameters for Modeling Sugarcane under Rainfed Conditions in Mexico. *Sustainability* **2017**, *9*, 1337. [CrossRef]
107. Thorburn, P.J.; Biggs, J.S.; Palmer, J.; Meier, E.A.; Verburg, K.; Skocaj, D.M. Prioritizing Crop Management to Increase Nitrogen Use Efficiency in Australian Sugarcane Crops. *Front. Plant Sci.* **2017**, *8*, 1504. [CrossRef] [PubMed]
108. Chukalla, A.D.; Mul, M.L.; Van Der Zaag, P.; Van Halsema, G.; Mubaya, E.; Muchanga, E.; Den Besten, N.; Karimi, P. A Framework for Irrigation Performance Assessment Using WaPOR Data: The Case of a Sugarcane Estate in Mozambique. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 2759–2778. [CrossRef]
109. Sonkar, G.; Singh, N.; Mall, R.K.; Singh, K.K.; Gupta, A. Simulating the Impacts of Climate Change on Sugarcane in Diverse Agro-Climatic Zones of Northern India Using CANEGRO-Sugarcane Model. *Sugar Tech.* **2020**, *22*, 460–472. [CrossRef]
110. Vianna, M.D.S.; Nassif, D.S.P.; Dos Santos Carvalho, K.; Marin, F.R. Modelling the Trash Blanket Effect on Sugarcane Growth and Water Use. *Comput. Electron. Agric.* **2020**, *172*, 105361. [CrossRef]
111. Prado, H.D. Ambientes de produção de cana-de-açúcar na região Centro-Sul do Brasil. *Informações Agronômicas* **2005**, *110*, 12–17.
112. Zhu, L.; Liu, X.; Wang, Z.; Tian, L. High-Precision Sugarcane Yield Prediction by Integrating 10-m Sentinel-1 VOD and Sentinel-2 GRVI Indexes. *Eur. J. Agron.* **2023**, *149*, 126889. [CrossRef]
113. Barnes, E.M.; Clarke, T.R.; Richards, S.E.; Colaizzi, P.D.; Haberland, J.; Kostrzewski, M.; Waller, P.; Choi, C.; Riley, E.; Thompson, T.; et al. Coincident Detection of Crop Water Stress, Nitrogen Status, and Canopy Density Using Ground-Based Multispectral Data. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, 16–19 July 2000; pp. 1–15.

114. Gitelson, A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [CrossRef]
115. Saini, P.; Nagpal, B.; Garg, P.; Kumar, S. Evaluation of Remote Sensing and Meteorological Parameters for Yield Prediction of Sugarcane (*Saccharum officinarum* L.) Crop. *Braz. Arch. Biol. Technol.* **2023**, *66*, e23220781. [CrossRef]
116. Rogers, A.S.; Kearney, M.S. Reducing signature variability in unmixing coastal marsh Thematic Mapper scenes using spectral indices. *Int. J. Remote Sens.* **2004**, *25*, 2317–2335. [CrossRef]