Growth and yield estimation of banana through mathematical modelling: a systematic review

S. L. Jayasinghe, C. J. K. Ranawana, I. C. Liyanage and P. E. Kaliyadasa

Crops and Soils Review


Received: 29 August 2021
Revised: 13 March 2022
Accepted: 14 April 2022
First published online: 23 May 2022

Key words:
Co-occurrence analysis; fibre; machine learning; mechanistic models; predictions; regression models

Author for correspondence:
P. E. Kaliyadasa,
E-mail: ewon@uwu.ac.lk

Abstract

Banana is one of the main fruit crops in the world as it is a rich source of nutrients and has recently become popular for its fibre, particularly as a raw material in many industries. Mathematical models are crucial for strategic and forecasting applications; however, models related to the banana crop are less common, and reviews on previous modelling efforts are scarce, emphasizing the need for evidence-based studies on this topic. Therefore, we reviewed 75 full-text articles published between 1985 and 2021 for information on mathematical models related to banana growth and fruit and fibre yield. We analysed results in order to provide a descriptive synthesis of selected studies. According to the co-occurrence analysis, most studies were conducted on the mathematical modelling of banana fruit production. Modellers often used multiple linear regression models to estimate banana plant growth and fruit yield. Existing models incorporate a range of predictor variables, growth conditions, varieties, modelling approaches and evaluation methods, which limits comparative evaluation and selection of the best model. However, the banana process-based simulation model ‘SIMBA’ and artificial neural network have proven their robust applicability to estimate banana plant growth. This review shows that there is insufficient information on mathematical models related to banana fibre yield. This review could aid stakeholders in identifying the strengths and limitations of existing models, as well as providing insight on how to build novel and reliable banana crop-related mathematical models.

Introduction

Banana (Musa spp.) is one of the world’s oldest cultivated fruit plants (Perrier et al., 2011; De Langhe et al., 2015), originating in South East Asia and Indochina (Simmonds, 1962; Nyombi, 2010). Banana is also considered the fifth most valuable commercialized agricultural food crop, with a cultivated area of 3.8 million hectares over 122 countries worldwide (Hossain et al., 2016; FAO, 2019, 2021). More than 1000 banana varieties are grown across the globe (FAO, 2021); among Asian countries, India is prominent for producing bananas, contributing approximately 25.7% of the total return (Rathod and Mishra, 2018). The Asia-Pacific region has a 61% share of global consumption in the banana market. Banana is an important nutritional supplement since they contain 67 calories per 100 g of fruit (Sharrock and Lusty, 2000). Banana is rich in calcium, phosphorus and nitrogen and, supplies 23% of our daily potassium requirement (Mohapatra et al., 2010; Hossain et al., 2016).

Bananas and plantains (i.e. cooking banana) are rhizomatous herbs whose inflorescence is produced by the terminal bud. The sequence of cycles is repeated for one to fifty generations or longer, indicating that is perennial (Turner, 1994; Tixier et al., 2004). The key developmental phases of banana plants contain sucker emergence, vegetative growth, flowering and fruiting (Tixier et al., 2004).

The economically vital part of the banana plant is a ripe fruit; however, unripen fruit, inflorescence, leaves, stem and rhizome parts are used in many nations as a cooked vegetable and animal feed. Furthermore, banana can be known as a herbal plant as it is enriched with medicinal benefits. All parts of the banana have nutritional and therapeutic value (Kumar et al., 2012), and prior studies have shown that the banana is abundant in bioactive compounds (carotenoids, flavonoids, phenolics, amines, phytosterols, vitamins) that have antioxidant qualities and can be exploited to provide pharmaceutical and health benefits (Pereira and Maraschin, 2015; Singh et al., 2016; Sidhu and Zafar, 2018).

Moreover, banana fibre has become a more popular type of natural fibre due to its many beneficial properties such as biodegradability, recyclability, chemical-free, low cost, low weight, high strength, non-toxic and odour-free (Vinoth et al., 2018). Besides, it is a better alternative for synthetic and other types of natural fibres (Subagyo and Chafidz, 2018). Fibre output is greatly influenced by variety; Musa textiles, for example, is known for its superior fibre characteristics. The extractable pseudostem and fibre yield percentage in the desert group (Musa acuminate) were found to be 46.4 and 0.53%, while in the cooking group (Musa paradisiaca),...
extractable pseudostem and fibre yield were 55.2 and 0.78%, respectively (Preethi and Balakrishna, 2013). The high percentage of fibre consists of the pseudostem and the banana peduncle thrown away after harvesting fruits (Vinoth et al., 2018). These fibres are used in many fields, such as textiles, paper industry as well as for bio and synthetic composites to be used in wide applications (Rosentrater et al., 2009; Vinoth et al., 2018; Priyaradshana et al., 2020). Also, banana fibre has been used to manufacture marine ropes as it is highly resistant to seawater due to the buoyancy properties (Subagyo and Chafidz, 2018). Banana fibres have also been used as a raw material in building construction and handicraft manufacturing in the automotive industry because of their more robust physical characteristics. Banana fibre is already a more commercialized product in Japan, Germany and Australia as a natural fibre source. Currently, there is a higher tendency in promoting fibre as a by-product of the desert type banana cultivations in other Southeast Asian countries such as Sri Lanka and India (Vinoth et al., 2018; Priyaradshana et al., 2020).

In the agricultural sector, modelling techniques are crucial in predicting crop growth and yield, allowing for more efficient and precise decision-making at present and in the future too (Jayasinghe et al., 2018; Basso and Antle, 2020; Hammer et al., 2020). In comparison, a large number of crop models have been developed and used for the main cereal crops (Rosenzweig et al., 2014; Beza et al., 2017), with little attention paid to tropical perennials (Rozendaal et al., 2020) such as banana. Crop modelling approaches are based on three key concepts: system, model and simulation (de Wit et al., 2019; Silva and Giller, 2021). A ‘model’ is a simplified depiction of a system distinguished as physical, conceptual, pictorial and mathematical (Putri et al., 2020); a ‘system’ is a specific portion of an actuality that comprises interrelated elements, and the term ‘simulation’ refers to the use of computer models to mimic a condition or process (de Wit et al., 2019).

Among the different models, a mathematical model provides a description of the behaviour of real-world systems in mathematical concepts, terms, and languages such as equations, inequalities, functions, variables and constraints (Chaturvedi, 2017). Moreover, these models allow for making crop predictions under specific environmental conditions (Medina-Ruiz et al., 2011). Mathematical models are divided into various categories or types based on the different features and purposes for which they have been constructed (Fig. 1). To investigate crop reactions in various cropping systems, researchers plan to use hybrid models that integrate various models (Wang et al., 2001; Jayasinghe et al., 2021; Shahhosseini et al., 2021).

Crop simulation models (CSMs) that are process-based mathematical representations of the mechanisms that contribute to crop growth, development, and yield in response to genotype, environmental variables, and management (Stöckle et al., 2003; Antle et al., 2017). Crop modelling and machine learning are combined to provide crop yield forecasts and externalities quickly (Drouttas et al., 2019; Folberth et al., 2019; Silva and Giller, 2021). Hence, mathematical models have been used to estimate crop yield gaps (Van Ittersum et al., 2016; Schils et al., 2018), the gap between food demand and supply (Keating et al., 2014), and the projected land area required to feed the world’s population (Yin et al., 2003; Chenu et al., 2009).

According to the literature, mathematical modelling approaches have rarely been employed in banana agro-systems (Jannoyer, 1995). Nonetheless, the few studies that have used mathematical models to forecast banana bunch yield revealed a lack of specificity in modelling outcomes, inadequate information on model development, small sample size, unclear assumptions and bias in model evaluation. Soltani et al. (2010), for example, used the water displacement approach to construct a novel mathematical methodology for forecasting the volume of banana fruit, claiming that the method is sufficiently accurate because the regression coefficient was 0.974. They did not, however, fully describe the method, regression equation, and evaluation, as well as the fact that the amount of water absorbed by the banana impacts its attributes and that the results cannot be generalized to other situations. Ganry and Chillet (2008) used the thermal time to establish a model for forecasting the harvest time of banana bunches. In their experiment, improper position of temperature sensors resulted in imprecise harvesting date estimation, as well as no detailed estimations or computations were included. Yamaguchi and Araki, (2004) estimated the biomass of banana plants using linear regression models based on pseudostem volume for EAHB cultivars under rainfed cultivation. Despite having high fits ($R^2 = 0.93$), a small sample size ($n = 14$), and the exclusion of moisture stress from their model, the applicability of the model for future work in biomass prediction remains dubious. Woomer et al. (1999) used bunch volume to estimate bunch weights in rainfed cultivation of cooking banana (EAHB cv. Mbwazirume) with great precision ($R^2 = 0.85–0.94$), but yields were not predicted as the parameters changed until harvest (Jullien et al., 2001). Their model is inaccurate for use in bunch yield prediction, and no details on possible stress were provided in their model development.

Furthermore, little effort has been made to model the fibre yield potential of banana plants. As mathematical modelling becomes a more ubiquitous and useful tool to understand plant responses, attention must be paid to reviewing previous studies related to modelling approaches. Such efforts would bridge the knowledge gap and provide a foundation for further investigations in mathematical modelling regarding the growth and yield of banana crops. Therefore, the main objective of the present review was to explore studies related to mathematical models that were used to estimate the growth and yield of bananas for both fruit and fibre and to synthesize the available information.

**Method**

*Search strategy, extraction and recording data*

The systematic review was based on the Preferred Reporting Items guidelines for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2010; Jayasinghe and Kumar, 2021). We primarily employed the Scopus database to search related studies. The relevant articles and the search terms were iterated and refined to locate the most specific studies. We developed three sets of search terms (TITLE-ABS-KEY (banana OR musa OR plantain OR banana OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plantain* OR banana* OR false AND banana) AND TITLE-ABS-KEY (model* OR simulate* OR modelling OR prediction*) OR OR model* OR or OR OR plan
to our topic to improve the sensitivity of the search for locating relevant literature from the Scopus database (Sargeant and O’Connor, 2020). Conference papers, conference abstracts and short surveys were excluded, retaining only peer-reviewed journals, books, book chapters and reviews. Retaining only peer-reviewed publications can help ensure that the articles are reviewed by experts in the field (Meline, 2006) and include sufficient facts about the methodologies they used in their studies. We also excluded articles written in languages other than English. The advanced search option available in Google Scholar was also used to find the full-reviewed articles and grey literature related to our scope. The search information yielded a total of 1082 articles, including 854 and 228 articles from the Scopus database and Google scholar, respectively. All the search information was initially recorded in a ‘comma-delimited (CSV)’ document with authors’ details, titles, published years, abstract, keywords and journal details. Duplicates were deleted from all of the references using the Endnote software (version X9). After removing duplicate records, the sample was reduced to 567 records.

Prospective studies for systematic review were then appraised for eligibility based on relevance and acceptability, following the exclusion protocol suggested by previous work (Meline, 2006; Linnenluecke et al., 2020). Accordingly, some studies (n = 445) were excluded from the further review because they (a) were clearly not in the scope of studies related to banana, Musa spp., or other fibre species (b) included incomplete or ambiguous methods and (c) failed to report sufficient statistics or data for estimating growth, development, fruit yield or fibre yield. As a result, a total of 122 full-text articles were eligible for further screening. Finally, the review team carefully selected 75 full-text articles for inclusion in the present review based on their relevance to the topic of the present study (Fig. 2). After screening related full-text articles, details were extracted from the manuscripts. The information on authors, year of publication, country of the study, statistical approach and the key findings were tabulated. Then, the articles were divided into three tables (Tables are given in the supplementary material as Table S1, S2, and S3), including articles related to mathematical modelling to estimate (1) fruit yield, (2) fibre yield and (3) growth and development of a banana plant.

Mapping of keywords to visualize research topic

The thematic content of selected studies’ keywords was visualized on a co-occurrence map using the VOSviewer software 1.6.5 (http://www.vosviewer.com, accessed on 9 February 2022) (Moral-Muñoz et al., 2019; Marchiori and Franco, 2020). The positioning of keywords on the map is based on the co-occurrence of the corresponding keywords related to mathematical models used to estimate bananas’ growth and yield. The co-occurrence map was visualized based on 75 papers on mathematical models to estimate banana growth and yield.

Results

Co-occurrence analysis based on keywords

As per Fig. 3, the positioning of 652 keywords on the map was based on a clustering algorithm allocated by VOSviewer software 1.6.18. The similarity of any two keywords is inversely proportional to the distance between, with a smaller distance indicating a stronger similarity or relationship between the keywords (McAllister et al., 2021). The frequency co-occurrence clearly shows the main keywords as well as how keywords link to each other within our field of study. We found that, the highest number of articles have been focused on the mathematical modelling of banana fruit...
yield. The smallest occupancy is visualized by 'natural-fibre' within the context of our review title, signifying that a lower number of studies have been executed on this subject (Fig. 3).

The map identified six keyword clusters [(1) prediction/crop yield/growth rate, etc, (2) fruits, (3) natural fibre/tensile strength, (4) musa/banana (5) fruit production and (6) numerical model/regression] based on probabilistic latent semantic analysis (McAllister et al., 2021). Except for some noticeable overlap on the left side of the map where clusters of 5 and 6 terms are related and had tied together with clusters 1 and 4, most of the clusters...
(e.g., 1, 2, 3, 4) are distinguishable from one another. These overlapping map clusters indicate that the clusters are related but not similar enough to be grouped together (Fig. 3). Furthermore, it indicates a strong association between clusters and the connection to the same field of study, ensuring that banana modelling is more closely associated with fields such as fruit production, growth and yield, the later in the context of our systematic review. Natural fibre, tensile strength and related modelling aspects are further away and even have small clusters (Fig. 3), indicating that natural fibres and related modelling aspects relevant to the banana crop are scarce in publications.

Basic statistics of published papers related to banana crop modelling

The number of publications related to mathematical modelling on banana fruit production (a, b, c), growth and development (d, e, f), and fibre-related characteristics (g, h, i) by year, country, and mathematical approach is shown in Fig. 4. In 2020, the highest number of papers related to banana yield modelling attempts were reported (Fig. 4(a)), and India and Brazil, are the major countries where many scientists are interested in the topic (Fig. 4(b)). Multiple linear regression (MLR) approaches have been widely employed in the development of banana yield models, followed by machine learning techniques such as artificial neural networks (ANN) (Fig. 4(c)). The years 2009 and 2020, followed by 2008, had the largest number of publications among the efforts made in developing models to predict growth and development (Fig. 4(d)), and France being the top country that significantly contributed for developing growth-related models (Fig. 4(e)). As shown in Fig. 4(f), the majority of published models were built using MLR and CSM. Most articles on banana fibre were published in 2019, followed by 2017 (Fig. 4(g)), and India was the leading country for these publications (Fig. 4(h)). There were no particular fibre-related models generated using mathematical approaches (Fig. 4(i)), but the majority of articles in the literature indicated the use of mechanical measures to assess tensile strength, and several other characteristics were linked to banana fibre processing.

Modelling of banana fruit production

Thirty four of the full-text articles were related to banana fruit production. The current review highlights how the fruit yield of banana has been estimated through mathematical modelling approaches in previous studies. In this context, several previous studies have shown evidence of how they used simulation, mechanistic, and/or statistical models to estimate the yield of bananas (Fig. 4 and Table S1).
MLR has been more prevalently used (38%) in modelling banana fruit/bunch yield according to the content analysis of previous studies (Fig. 4). Salvacion (2020) used time series analysis and a regression model to investigate the impact of climate (i.e., annual rainfall, frequency of wet days, precipitation seasonality, annual mean temperature and temperature seasonality) on banana yield at the provincial level in the Philippines. MLR analysis revealed that 10% of banana-producing areas in the country are affected by local weather conditions (Salvacion, 2020). Sharath (2016) used stepwise multiple regression and it showed that plant height at the fourth month, girth at the fifth month, the number of fruits per bunch, and fruit girth at the final stage were the most significant factors in predicting total banana yield, while sucker parameters and number of leaves had no significant impact on banana yield. Zaculoto et al. (2013) found that morphological characteristics such as the third leaf’s width, number of leaves per tree, the bunch’s diameter, and the number of bananas per bunch estimated the bunch weight of banana cv. Prata Ana using MLR models with moderate accuracy (Coefficient of Determination (R²) = 0.58).

Olivares et al. (2022) converted soil morphological factors in banana plantations to numerical scale and created a multiple linear regression model between those parameters and the banana crop Productivity Index (PI), with R² = 0.645 as the model accuracy. Robinson and Human (1988) used regression allometries to forecast banana harvest based on seasonal variations in bunch development rate and bunch mass. However, 15% coefficient of variation (CV) reflected a 100-day harvest time, limiting the prediction capacity of the derived model.

Wairegi et al. (2009) developed a mathematical model with high accuracy (R² = 0.73) to estimate the bunch weights of bananas using parameters such as log-transformed girth of pseudostem at the base and 1 m height, a number of hands, and a number of fingers in the lower row of the second-lowest hand using an MLR method. The same authors employed a quadratic or linear regression model (R² = 0.58) to quantify the substantial geographic variations in banana yield including biotic constraints (e.g., percentages of nematodes and weevil) in highland banana production in Uganda (Wairegi et al., 2010).

The statistical models developed by Venugopalan and Shamasundaran (2005) showed that at 70 days after planting (DAP), the number of leaves and plant girth with optimum values as eight leaves and 15.1 cm were the best yield indicators for bananas. Further, MLR showed that at 250 DAP, plant height and plant girth with optimum values as 159.2 and 67.8 cm and at 315 DAP, leaf breadth and leaf length with optimum values as 67.2 and 164.1 cm were the significant yield predictors. Finally, during the harvest stage, i.e. at 375 DAP, fingers per bunch and number of hands per bunch with optimum values as 26 fingers hand-1 and 13 hands bunch-1 were the best crop yield indicators. All models were robust as the coefficient of determination ranged from 0.81 to 0.99. Kuneepong et al. (2020) used a general linear model for predicting the growth of a new banana variety (Kasetsart 2) in Thailand with aid of data related to local weather, irrigation systems and soil types. However, model predictions of banana yields for some locations were under-estimated in their model.

Scientists have also used MLR in conjunction with multivariate analysis to develop mathematical models for banana crops. For example Villegas-Santa and Castañeda-Sánchez (2020) identified the relationship between soil variables and crop performances of bananas using the multivariate statistical tool. Three clusters of sites were evaluated based on dry mean weight, pH, and Ca + Mg/K ratio of soils, and all these soil properties highly correlated with banana yield, however the principal component analysis (PCA) was unable to identify the production quartiles due to the lack of significant causal relationships.

Jaiswal et al. (2012) used a non-destructive method using calibration models in the wavelength range of 299–1100 nm and partial least square (PLS) as well as MLR to predict dry matter (DM), total soluble solids, and acid-Brix ratio, pH for a banana at their maturity and/or ripening stage. The result indicated that DM of banana could be accurately estimated in the best range of wavelength of 106.2–1089.4 nm with a correlation coefficient (R) value of 0.83. Salazar-Díaz and Tixier (2021) investigated the impact of the plant community around each cacao tree and banana plant on their growth and yield parameters. First, they used linear mixed models to identify the radius of a specific species’ zone of influence that best explained the proportion of attainable yield (PAY), using the number of plants from each category (banana, cocoa, wood trees, fruit trees) as a predictor. The models were relatively accurate in predicting the average effect of all plant communities, accounting for almost 60% of the variance in PAY, which is a good result. The Bayesian technique was used by Yeasin et al. (2021) to forecast banana yield. They compared traditional regression models with modified Bayesian regression model, and observed that Bayesian regression [root mean square error (RMSE) = 216.9] predicts banana yields more accurately than traditional regression (RMSE = 223.1).

Olivares et al. (2020) developed a model to predict the banana Productivity Index (PI) in two main cultivation areas in Venezuela using soil properties of Magnesium (Mg), penetration resistance (PR), total microbial respiration (TMRc), bulk density (BD), and free-living omnivorous nematodes (NVLomc). The soil variables were selected through the random forest and the final model was derived from MLR (stepwise) models. The following model performed well, with a coefficient of determination (R²) of 0.55, the root mean squared error (RMSE) of 1.0 and the mean absolute error (MAE) of 0.8 indicating the capability of estimate banana production using soil properties:

\[
PI = 9.03 + 0.19 \text{(Mg)} - 8.78 \text{(PR)} - 0.40 \text{(TMRc)} - 4.70 \text{(BD)} + 0.02 \text{(NVLomc)}
\]  

(1)

In certain models, simple linear regression has been used to estimate banana yield or biomass using only one independent variable. For instance, Alcudia-Aguilar et al. (2019) demonstrated that above-ground banana biomass (AGB) was strongly correlated with the diameter of pseudostem at the height of 30 cm (DBH) and to a certain extent with height data. Considering cross-evaluation, the best-derived model was:

\[
AGB = -0.0927 + 0.0203 \text{DBH}^2 \quad \text{with} \quad R^2 = 0.88 \quad \text{and} \quad \text{MSE} = 1.9
\]  

(2)

In the context of modelling of growth and yield of banana, non-linear mathematical models were also used by many authors that take complicated non-linear relationships of predictor variables and banana yield into account. Laskar et al. (2020) used mathematical models to estimate biomass accumulation of wild Musa spp. The predictors of diameter, height and combination of diameter and height of pseudostem were used to estimate total biomass using a non-linear, seemingly unrelated regression.
analysis. Further, mathematical forecasting approaches such as regression and time series analysis were used for linear data, but most real-world data are non-linear. Rathod and Mishra (2018) have tackled this problem in their study by proposing a new hybrid model to forecast mango and banana crop yield in Karnataka, India which is consisted of both linear and non-linear components. They have used banana and mango yield as dependent variables and weather variables, socioeconomics, and some other agricultural variables as independent variables.

Bugaud et al. (2011) developed a model to estimate the effects of source/sink parameters by simulating the increase in pulp dry weight as the source/sink ratio changes. To simulate the cell filling rate in the bunch according to the source/sink ratio during cell filling, they used a Michaelis-Menten relationship, which is a popular non-linear equation. The exponential regression model proposed by Ortiz-Ulloa et al. (2020) that employs circumference at breast height (CBH) has the best biomass prediction capacity for Ecuadorian banana, with a $R^2$ of 0.85 (Table S1). Negash et al. (2013) used allometric biomass models with linear and nonlinear regressions to estimate above and below-ground biomass of Ensete’s false banana plant, and found that the simple power model with trunk basal diameter at 10 cm (d10) and total plant height (H) performed well, explaining 90% of the total biomass:

$$\text{Biomass} (Y) = 0.0007d10^{0.571}H^{0.101}$$  (3)

Weighted least square regression models were built by Stevens et al. (2020) to estimate aboveground vegetative biomass over time and to forecast bunch potentials. Pseudostem volume as a predictor for bunch weight forecasting led to good model performance in both tested cultivars in their study [relative root mean squared error (RRMSE) 0.13–0.14].

Gaussian functions are commonly employed in statistics as a stochastic process to explain normal distributions using a set of random variables indexed by time or space. For example, de Deus et al. (2020) developed models to estimate dry matter partitioning in banana mat (collection of fruit-bearing stems). Dry matter estimation models were generated for the different plant organs in mother and daughter plants of banana as a function (i.e., Gaussian) of the dry matter weight of the banana mat (Fig. 4 and Table S1).

Process-based crop simulation models were also found in our review. For instance, Tixier et al. (2004) made a tremendous effort to describe banana crop growth, development, and yield as a function of plant physiological parameters, weather conditions, soil properties and crop management-related variables. This model gives a reliable estimation of the harvesting time and the number of harvested bunches. The fruit development curve was used to estimate sink demand, which was calculated as the fruit’s requirement multiplied by the number of fruits per bunch.

The ANN approach could handle non-linear relations among the biometrical characteristics in crop modelling research for a realistic representation of nonlinearity. An Autoregressive Integrated Moving Average (ARIMA) model was developed by Hossain et al. (2016) that analysed univariate time series data and estimated a value in a response time series to forecast the development of bananas in Bangladesh. Moreover, Hossain et al. (2016) mentioned that ARIMA (0, 2, 1) could be considered the best model to forecast banana production in Bangladesh. Soares et al. (2013) estimated the bunch weight of banana using a Tropical (YB42-21) hybrid cultivar. The weight of the rachis, stalk length and diameter, weight of the second hand, the total number of hands per bunch, the weight of the fruit, number of fruits per bunch, length of the fruit, diameter of fruit, and number of fresh leaves at harvest were considered as input layers. They employed the ANNs method to predict banana yield with 10:1:0 architecture, and the model performed well with high accuracy ($\text{MPE} = 1.40, \text{MSD} = 2.29$ and $R^2 = 91\%$).

de Souza et al. (2019) developed a model using ANNs to find the relationship between climatic variables (average temperature, minimum temperature, maximum temperature, relative humidity, precipitation and photoperiod) and banana bunch gestation period to predict the harvesting time. Network training indicated reliable results (RMSE = 0.3% and $R^2 = 0.89$). For a AAAB tetraploid banana hybrid, Soares et al. (2014) used ANN and MLR to determine bunch weights. Although the models were reliable ($R^2 > 0.71$), the model consisted with the predictors that could only be measured destructively at harvest.

Guimarães et al. (2021) used ANNs to assess the yield of ‘Prata-An’ and ‘BRS Platina’ banana plants. The model was built using data on growth-reproductive traits. For ‘Prata-An’ ($R^2 = 0.99$) and ‘BRS Platina’ ($R^2 = 0.97–1$), the optimal adjustments were obtained with two and three intermediate layer neurons, respectively. Eyduran et al. (2020) investigated various ANN methods, including ARIMA (0,1,1), ARIMA (1,1,0) and ARIMA (1,1,1), as well as Exponential Smoothing (Holt, Brown and Damped) models, and discovered that the Brown exponential smoothing model was the best for forecasting banana harvest area and production ($R^2 = 0.99$ and mean absolute percentage error (MAPE) = 19). Khan et al. (2020) employed three potential deep ANN networks [Levenberg-Marquardt optimisation (LM), scale conjugate gradient back propagation (SCG) and Bayesian regularisation back propagation (BR)] to estimate future fruit production in Pakistan, including banana, from 1980 to 2025. Although the accuracy of these algorithms varies, the best accuracy is seen with BR, which has a 75% accuracy. To estimate banana harvest yields, a deep multilayered system that included a number of recurrent neural network-long term memory (RNN-LSTM) layers was also used. When compared to models that used multiple LSTM layers and models that used a single LSTM layer, the enhanced model performed better, with an RMSE of 34.8 and error rates of 45 and 43.5%, respectively (Rebortera and Fajardo, 2019a, 2019b).

**Modelling of banana growth**

A total of 20 studies presented mathematical modelling techniques to estimate plant growth and development. In particular, many attempts were executed to assess the leaf area of banana crops using regression analyses ($n = 9$) and crop simulation models (Fig. 4 and Table S2).

MLR approaches ($n = 7$) are commonly used regression techniques in banana growth models, as shown in Fig. 4. A more realistic leaf area estimation model was developed by Donato et al. (2020) for ‘Prata-An’ and ‘BRS Plantina’ banana plants using the variables width ($W$), length ($L$) and width/length ratio ($WLR$). The following models gave precise results ($R^2$ around 0.96, and $R = 0.98$, respectively):

$$LA (\text{Prata-An } \sim a) = -0.0133624 + 0.000489859L + 0.00183182W$$  (4)
LA (Platina) = 0.00237026 + 0.00478116 W − 0.0968020WLR
(5)

Vinson et al. (2018) developed a regression model to determine flower emergence and to assess vegetative and reproductive growth of Cavendish and non-Cavendish banana using the circumference of pseudostem (CM), the height-to-circumference ratio (HCR), and the number of days from planting to inflorescence emergence (DPE). Separate models were developed to predict leaf area of medium and tall banana varieties using regression models.

In addition, Mekwatanakarn and Turner (1989) constructed a robust model ($R^2 = 0.96$) to evaluate the leaf emergence rate in bananas in the subtropics by assessing the relation of leaf emergence rate to temperature and ontogeny. Allen et al. (1988) created a model for estimating the leaf initiation rate (LIR) of 17 banana cultivars based on average monthly air temperature, day length, age of planting, plant density and cultivar stature. The model was evaluated with an independent data set from South Africa and it gave a reliable prediction of LIR with $R^2 = 0.78$. Arantes et al. (2016) used MLR equations to find a correlation between chlorophyll index and nutrient contents of leaves and estimate the nutritional status of ‘Prata’ banana. The selected models reliably predict the leaf nutrient content of Prata’ banana ($R^2 = 0.84$ to 0.97).

Nyombi et al. (2009) studied different phenological stages of highland bananas grown in East Africa and assessed growth and yield using simple regression as well as MLR algorithms. Individual area was estimated as:

$$
LA (m^2) = \text{length} (m) \times \text{maximum lamina width} (m) \\
\times 0.68 (R^2 = 0.99)
$$

The product of the measured middle leaf area (MLA) and the number of active leaves was used to calculate the total plant leaf area (TLA) and estimated as:

$$
MLA (m^2) = -0.404 + 0.381 \text{ height} (m) + 0.411 \text{ girth} (m) \\
(R^2 = 0.67)
$$

The mathematical relationship between above-ground biomass (AGB in kg DM) and girth (cm) during the vegetative stage followed a power function, AGB = 0.0001 (girth) $^{2.35}$ ($R^2 = 0.99$), but followed exponential functions at flowering, AGB = 0.325e$^{0.036}$ (girth) ($R^2 = 0.79$) and at harvest, AGB = 0.069e$^{0.068}$ (girth) ($R^2 = 0.96$). Demirsoy (2009) constructed models for the leaf area of fruit trees, including two banana varieties using simple linear regression. Several subsets of the independent variables were used in their leaf area prediction model, such as leaf length ($L$), leaf width ($W$), $L^2$, $W^2$ and $L^2/W^2$.

Potdar and Pawar (1991) applied multiple linear regression to predict leaf area (LA) in the banana cultivars ‘Ardhapuri’ and ‘Basrai’ using various combinations of leaf length ($L$) and width ($W$), and found that the models had a predictive power with $R^2$ of around 0.96. Tauliya (2013) used multiple linear regression to estimate the fresh mass (FW) of the plants from the growth parameters (plant height ($H$); number of functional leaves ($L$); and the mean pseudostem volume with $R^2 = 0.59$ accuracy:

$$
PV(FW) = 0.03873H + 0.281L + 3.169 \times 10^{-6} P_v - 2.905
$$

Latif et al. (2020) investigated the use of a simple logistic growth model to predict banana vegetative growth in response to foliar fertilizer. To simulate pseudostem height, pseudostem girth, and leaf area at harvest ($Y$), variables of time (week) ($t$), carrying capacity (maximum growth (cm) ($K$), constant ($A$) and growth rate ($r$) were used:

$$
Y = Y = K/1 + Ae - rt
$$

Mendes et al. (1999) applied adjusted Poisson regression models to study multiplication rates ($Y$) of in vitro banana plants during successive subcultures using number of shoots ($N$) as:

$$
Y = \exp (3.75 + 2843N - 0.2312 \times N^2) (R^2 = 0.98)
$$

Stochastic models are also widely exploited as mathematical models where random variables are considered. For example, Lamour et al. (2020) used a novel stochastic model to estimate the average time gap between two flowering occurrences on the banana plant (CD) using population dynamics and banana phenological development stages. The results indicated differences in CD with a median of 209 days ± 24 days. There was a positive effect of elevation, cultivar and irrigation on model fitting and the CD estimation. Moreover, some other crop simulation models deal with plantain (Musa × paradisiaca) crop systems to simulate growth or plantain yield. In the given context, Chaves et al. (2009) introduced models to estimate the potential yield of plantains using the models already developed by Spitters and Schapendonk (1990) and Kooman and Haverkort (1995), especially considering the involvement of dry matter partitioning, light interception and daily conversion of light in dry matter production. Mechanistic modelling and non-linear regression analysis were used to develop the above models. Chaves et al. (2009) observed that leaf, stem and corn dry matter increased in equal proportions during the vegetative stage. However, only the stem was observed to extend its dry-matter content during the reproductive period, while the leaves and corn were found to reduce dry weight. Tixier et al. (2008) developed a model called SIMBA with different sub-modules, design for a sustainable banana-based cropping system that mimics several cropping cycles at the field level.

A modified version of the SIMBA-GROW module was developed by Tixier et al. (2008) to simulate growth of banana plant considering four growth stages (planting, sucker initiation, flower initiation and harvesting) while taking the bunch as the sink. The SIMBA-GROW model is constructed using radiation interception, biomass conversion and dry matter partitioning to leaves, suckers and bunches. The physiological development of banana is determined by thermal degree days taking base temperature as 14°C. SIMBA-GROW calculates the plant growth for every cohort defined in the SIMBA POP module (Tixier et al., 2008). SIMBA-POP is the first crop model for long-term simulation in non-synchronized banana cropping systems (Tixier et al., 2004). The SIMBA-POP sub-module was developed based on a cohort population concept (i.e., a group of individuals characterized by the same phenological stage) that simulates the banana plant.
population and its growth and development patterns. Further, this model considered the details about sucker selection which affects plant management (Tixier et al., 2004). They used log-normal distribution for the sucker appearance and flowering patterns in the field as explained by Cottin et al. (1987).

Using a compaction score, the SIMBA-SOIL module simulates soil structure. This accounts for data on essential management practices in the banana field, including the use of fertilizers, harvest trailers and ploughing. The SIMBA-WAT, a water balance module that simulates the content of the soil, water, flow-off and nutrients leaching (Tixier et al., 2008). Tixier et al. (2011) developed the SIMBA-CC model to select cover crops for banana cover-cropping systems using 11 cover crop species, LAI, biomass, N content in the biomass, light interception traits, and values of optimal photosynthetically active radiation (PARopti). The SIMBA-COV sub-module simulates soil cover, including weed growth and the impact of herbicides, mulching, and crop residues on weed growth in banana cultivation. Weed growth is represented as a percentage of soil cover that calibrates using a logistic function and it has an inverse proportion to the leaf area index of banana. Damour et al. (2012) used simplified indicators of soil water and nitrogen availability, as well as integrated plant characteristics, to estimate the growth of banana (Musa spp.) cultivated on cover-crop. Dorel et al. (2008) designed the SIMBA-Nitrogen (N) model to simulate N dynamics in the banana cropping system with reliable evaluated results that can use in N fertilizer management in banana cultivation.

The STICS (Standardized mulTIdisciplinary Crop Simulator) model was used by Brisson et al. (1997) to estimate the impact of soil and water management on banana growth between planting and flowering. Solar radiation, minimum and maximum temperatures, rainfall, and evaporation were taken into the model development. Soil parameters of carbon and nitrogen were also considered as input variables. The model simulates the carbon, nitrogen and water balance of the banana cropping system and calculates growth using environmental data. A comprehensive description of model calibration and evaluation was not given in their paper.

ANN models are widely applied in crop modelling studies and indicate their applicability to estimate banana plants’ growth rate. Revathi et al. (2019) developed a non-destructive model for tissue-cultured banana plantlets during the acclimatization period using bootstrapped artificial neural network (BANN). Both non-destructive and destructive growth parameters, greenhouse temperature, radiation and carbon dioxide concentration were recorded on a regular basis as independent data. The projected growth statistics indicate the reliability of the model to predict the growth with satisfactory Nash and Sutcliffe efficiency (NSE) coefficient, RMSE and MAE.

**Modelling of banana fibre**

No previous studies have specifically (Jayasinghe and Kumar, 2021) addressed the modelling of fibre yield of the banana plant according to our review results. However, many efforts have been made to predict the tensile strength (n = 14) and other mechanical properties (n = 4) of banana fibre (Fig. 4 and Table S3). Two studies that use mathematical approaches to assess the fibre yield of false bananas (Ensete ventricosum) have been identified, and we have included them in our review. Researchers employed a range of statistics to investigate fibre properties and a diverse applications in the search (Fig. 4), demonstrating how predictive models are used to determine the orientation and, mechanical and chemical properties of banana fibre using different regressions, numerical measures (i.e., mean, median, mode, percentiles, range, variance and standard deviation), and ANN statistics.

For instance, Chokshi et al. (2020) used mathematical models (e.g. exponential, linear, logarithmic, polynomial and power models) to predict the tensile strength at different strain rates of natural fibre. The polynomial model yielded higher accuracy (R² = 0.96) to predict tensile strength at low and high strain rates. Mwesigwa and Mwasiagi (2019) used a MLR model with strong model performance (R² = 0.9) to study the parameters influencing the tensile and compressive characteristics of banana bio-composites. Monzón et al. (2019) used a statistical approach to characterize the mechanical properties of newly developed technical textile with a composite of banana fibre and compare the results with numerical simulation. Madhusudan et al. (2018) used mathematical equations to predict hygro strain coefficients for natural hybrid composites of banana and pineapple.

The basic mathematical model was used by Mizera et al. (2017) to describe the dying behaviour of false bananas and fibre strength. Venkateshwaran et al. (2012) predicted tensile properties of hybrid-natural fibre composites. Besides, Venkateshwaran and ElayaPerumal (2011) assessed the tensile properties of banana fibre. de Oliveira et al. (2020) developed nine mathematical models to study the dying behaviour, moisture diffusivity, activation energy and thermodynamic properties of banana pseudostem fibre using thermogravimetric, morphological and spectroscopic methods. Devireddy and Biswas (2018) developed a mathematical model to calculate thermal conductivity and the predicted thermal conductivity was very close to actual values. Chokshi and Gohil (2018) developed mathematical models to predict the tensile strength of the banana fibre. According to their study, linear relationships of exponential, linear, logarithmic, polynomial and power models were used to study the behaviour of tensile strength and they found that tensile strength increases with an increase in strain rate.

Patwari et al. (2019) proposed a quadratic model to forecast the compressive load of moulded green composite materials (containing banana fibre):

\[ Y = aX^2 + bX + c \]  

Bhaskar et al. (2020) developed a novel mathematical model to predict the transverse thermal conductivity of hybrid composites based on microstructural characteristics. Mizera et al. (2016) used general exponential models to describe the effect of gauge length of false banana fibre for tensile strength.

Natural fibres have many advantages over synthetic fibres, but the reduction of quality is a drawback of natural fibres due to climate and growing conditions (Rao et al., 2017). When predicting the tensile strength of the fibre, the issue has arisen due to the variation of the cross-sectional area of the fibre with the length (Sia et al., 2018). Sia et al. (2018) produced a Weibull distribution with modification to predict the tensile strength of the fibre by providing the solution to the issue of cross-sectional variation. As per their findings, the fibre tensile strength decreases when the gauge length of the fibre increases. We found some other statistical approaches related to fibre processing that was adopted by researchers. For example, Macedo et al. (2020) and Veeramanipriya et al. (2019) developed simple mathematical models to identify the drying kinetics of banana fibre. Devireddy and Biswas (2018) and developed a mathematical...
model to calculate thermal conductivity with minimum errors (>1%).

Pujiari et al. (2017a) compared the performances of ANN and regression analysis for predicting the water absorption behaviour of jute and banana fibre reinforced epoxy composites and indicated that ANN gives better results for physical properties of natural fibre composites of banana than the regression analysis. Moreover, Pujiari et al. (2017b) predicted the volume changes of jute and banana fibre composites by using ANN and regression analysis and they indicated that ANN performs better than regression analysis.

For the enset or false banana crop (*Ensete ventricosum*), a close relative of the banana crop, Haile (2014) attempted to estimate enset fibre content from aboveground plant traits using a linear regression model, but even when fibredata was log transformed, he was unable to produce significant regression equations ($R^2 = 0.01$). For the first time, Yemataw et al. (2021) developed an allometric model that can correctly estimate enset fibre yield. The PCA was selected using a biplot and the Kaiser–Meyer–Okin (KMO) method of eigenvalues over one, and a backwards regression analysis was utilised to automatically establish the most significant model to explain fibre yield. Accordingly, fibre yields of enset can be estimated using leaf length, petiole length and plant height (adj. $R^2 = 0.35–0.57$).

**Discussion**

As a traded fruit and natural fibre source, the continuous production of banana should be maintained to cater to the demand, and banana-growing systems should be managed with effective strategies and techniques. Modern agricultural systems have inspired the mathematical modelling approach for growth and yield with the technological revolution (Jones et al., 2017). Based on this premise, mathematical modelling (Fig. 1) plays a crucial role in providing a link between the concept and the actual biological unit of banana systems (Connor et al., 2011). Some scientists in different countries have modelled banana growth and fruit yield using mathematical models (Fig. 4). To understand how mathematical models have been used in estimating banana yield and growth, we have presented our results based on three separate aspects of the mathematical modelling of banana: (1) fruit/banana yield, (2) growth and (3) fibre yield.

In the present review, 34 and 20 previous studies were found to estimate fruit yield and growth, respectively. However, no studies have been performed to assess banana fibre yield to date based on our search. The review also confirmed the less frequent occurrence of ‘fibre yield’ and its distant link to aspects of banana crop modelling in the keyword co-occurrence map (Fig. 3). This may be because the application of modelling techniques has not been fully exploited for banana fibre yield as it is an emerging venture being explored recently.

To understand and exploit the mechanism of mathematical models, we extracted the modelling approaches of selected studies. A range of techniques has been implemented for the mathematical modelling of bananas (Fig. 4). Regression analysis is a simple, commonly used mathematical technique used by many authors to estimate banana growth, yield and fibre properties. Our review explicated that most of the developed models are linear. Many authors have developed simple linear and/or MLR equations to estimate banana productivity (Potdar and Pawar, 1991; Wairegi et al., 2009; Zucoloto et al., 2013; Alcudia-Aguilar et al., 2019; Olivares et al., 2020), and growth (Potdar and Pawar, 1991; Demirsoy, 2009; Nyombi, 2010; Vinson et al., 2018; Donato et al., 2020). In all these models, the growth and yield of bananas have been modelled as a function of either plant characteristics or weather, soil, and/or management practices by integrating agronomic knowledge. Some scientists applied true eigenvector-based multivariate analysis (eg., PCA) combined with MLR (e.g., Sharath, 2016; Villegas-Santa and Castañeda-Sánchez, 2020) to explore major patterns of the physiological predictors and yield of bananas. The problem of component collinearity can be avoided with PCA, but the challenge of determining the ideal number of eigenvectors still exists, hence some modellers (e.g., Jaiswal et al., 2012) used PLS regression in their modelling efforts.

Rathod and Mishra (2018) proposed a new hybrid model consisted of both linear and non-linear components. The hybrid model with the ARIMA and Support Vector Regression (SVR) model’s conflation performed well in both model development and evaluation. Some authors, such as Yeasin et al. (2021) and Khan et al. (2020) used Bayesian regression techniques to predict banana production since Bayesian statistics are often more powerful because they provide a whole distribution of regression parameters rather than just point estimates. Stevens et al. (2020) estimated banana crop growth and yield using weighted least squares regression. When modelling the behaviour of random errors in a model, WLS can be utilised with functions that have either linear or nonlinear parameters. When modellers want to model the average number of occurrences per unit of time, they utilise modified Poisson regression, which combines a log Poisson regression model with robust variance estimation. Mendes et al. (1999) used this allometrics to investigate the multiplication rates of *in vitro* propagated banana plants over time. In linear mixed models, simple linear models are extended to include both fixed and random effects, making it a flexible method for analysing complicated datasets with repeated or clustered observations. Salazar-Díaz and Tixier (2021) used LMM to evaluate banana and cacao interactions in heterogeneous multi strata agroecosystems, with the aim of predicting banana productivity. Furthermore, Gompertz, Gaussian, and non-linear logistic models have provided a reasonable representation of banana growth and yield. For instance, de Deus et al. (2020) studied dry matter partitioning of the banana plant using the Gaussian process model.

Our review explicated that most developed models are linear (Tables S1 and S2), but a significant issue with developing a regression model is the many independent variables that may consume a degree of freedom (Harrell Jr. et al., 1996). We observed that linear regression models have been used to predict the effect of some plant-related variables (age, height, weight and number of shoots), which may not give a real picture of the system. In the development of models, the modeller should select appropriate predictors that describe the dependent variables (growth, yield and fibre properties) of banana using numerical values. Moreover, the data demand of regression models is lower that of crop simulation models (Irwin et al., 2008). However, multiple regression models require well-defined descriptors to predict growth or yield.

Crop simulation models have become an essential tool during the last decade to model growth and yield of banana plants. Tixier et al. (2008) developed the process-based SIMBA model with several sub modules design for a sustainable banana-based cropping system that can be simulated at the field level over various cropping cycles (Tixier et al., 2013). This model serves as a powerful tool to choose the best genotypes for cultivating bananas. It includes almost all essential environmental variables, cohort...
dynamics, soil parameters, cropping patterns and eco-
physiological factors. It should be noted that this model requires
complex information to understand the model principles. Some
findings (e.g. Chaves et al., 2009) have highlighted the importance
of incorporating the source-sink relationship with plant growth.

There are some other process-based models (e.g., APSIM, CROPSYST, DAISY, FASSET, HERMES, DNDC, WOFOSt, DSSAT, AquaCrop and STICS) (Sannagoudar et al., 2019; Hoogenboom et al., 2021) that can be utilized as banana crop simulation models. Effective banana crop simulation model outputs (i.e., growth, fruit and fibre yield) are only possible if the model’s parameters are accurately represented in the simulation and the input cultivar related data are accurate and reliable (Grassini et al., 2015; Geethanjali and Muralidhara, 2021). Future research should focus on developing crop models for banana and expanding crop model capabilities to include nutrient limitation methods as well as yield reductions due to pests and diseases.

Machine learning techniques have enabled substantial progress in recent crop model development (Droutsas et al., 2019; Folberth et al., 2019). Christian (2020) showed that crop meta-models based on machine learning algorithms may accurately mimic biophysical crop models for yield potential. The machine learning approaches include linear regression models, feedforward, recurrent and convolutional neural networks, decision tree and random forest, support vector regression, agricultural deep learning, autoregressive integrated moving average, and Bayesian regression (Chakraborty and Joseph, 2017). Recently, ANN has been widely accepted in crop modelling as a reliable computer-based nonlinear empirical model (Poznyak et al., 2019). Considering given benefits of ANN, many scientists have utilized this technique to model the growth, yield and fibre properties of banana.

Several limitations need to be overcome in some mathematical models, leading to more extensive use and acceptance by general users. Even though some models (SIMBA, ANN models) provide a robust simulation of the growth and yield of banana, these models could be more straightforward and evaluated with independent data sets obtained from other banana-growing environments that involve tedious and time-consuming processes. Moreover, the main constraint is that many mathematical models disregarded plant and environmental interactions and only focused on a single plant to estimate specific processes. Crouch and Haines (2004) stated that a successful mathematical model needs to use suitable mathematical techniques, principles and predictors. Some previous studies have identified the importance of plant, soil, climate and cropping conditions as predictor variables to build mathematical models. In the given context, assembling and measuring some predictor variables is a challenging task in the model development, calibration and evaluation steps. For instance, the comprehensive crop models developed by Tixier et al. (2008) to stimulate the growth and yield of banana based on cropping cycles are complex and, costly that require long-term monitoring to obtain measurements. Hence, model should be simplified to some extent to reduce the amount of data required to build and assess the models while avoiding clumsy and excessive modelling techniques (Yin and Struik, 2017; Basso and Antle, 2020).

Sufficient meaningful data should be selected from an empirical data set for model calibration using the identification analysis as proposed by De Swaef et al. (2019), with the aid of sensitivity functions and collinearity analysis (Brun et al., 2001). Every stage of modelling is riddled with uncertainty (i.e. errors in observations, prediction uncertainty, model uncertainty) (Chapagain et al., 2020). Model evaluation is a vital part of mathematical modelling that compares measured values and predictions from mathematical models (Giordano et al., 2013; Strano et al., 2013). Different parameters have been used to validate models as shown in the findings of the current review (Table S1, S2 and S3), such as RMSE, MAPE, MAE, NSE, relative root mean square error (RRMSE), correlation coefficient (R), coefficient of variation (CV), SD and determination coefficient ($R^2$). From a statistical perspective, RMSE is the same as standard error that measures the extent of change in mean absolute difference. RMSE alone is insufficient to use for model evaluation as it is not enough to conclude a model’s predictive abilities, and RMSE cannot estimate the mean absolute difference between observed and measured values (Cao et al., 2012; Gramatica and Sangion, 2016). Therefore, it is better to use other parameters such as $R^2$, NSE and MAE to show its accuracy. Most of the reviewed studies included evaluation results. However, the model evaluation procedures were not clearly stated in some of the studies used in our review.

In future research, mathematical models of banana can help in the development of novel concepts, parameters, techniques and the design of new trials. There are no crop models in the existing literature that can predict banana production for future climate scenarios. Climate trends, the Extended Virtual Riesling model (Schmidt et al., 2019), high-quality weather data (Hayman et al., 2012; Bracho-Mujica et al., 2021), and the bias-corrected regional climate model (RCM) (Webber et al., 2018; Falconnier et al., 2020) should be taken into consideration for designing process-based banana yield simulation models. Furthermore, banana crop models can be combined with geospatial systems, artificial intelligence and remote sensing approaches (Ruane et al., 2015; Hersbach et al., 2020), where a large variety of global gridded data sets can be used in modelling efforts in the future. Crop models have recently been combined with phenomics, breeding and testing in other crops (Cooper et al., 2007; Hammer et al., 2020), which we can apply to banana modelling in the future. Existing genotypes should be screened via models (Marshall-Colon et al., 2017) for high fruit and fibre yields, and banana crop ideotypes should then be designed accordingly. Despite the fact that there are various allometric models for calculating false banana (Enset) fibre yields (Yemataw et al., 2021), no work on banana fibre yield modelling has been done yet, indicating the need for modellers to fill this gap in the future.

The mathematical models presented in this review are beneficial to analyse the growth and yield of bananas, and the process-based simulation models demonstrate the methodologies with precise detail, offering outstanding notions and inspiration to other modellers to engage with modelling techniques. Some models have been embedded with subtle details that provide an opportunity to leverage the development of reliable, versatile and generalized banana-plant-inspired algorithms.

Conclusions

Mathematical models have been highly exploited to estimate the growth and yield of bananas, acting as a tool for predicting, planning and managing resources and important in bridging the conceptualization and realization. However, the cropping pattern of the banana plant is very intricate, and quantitative integration of appropriate variables is essential to develop a robust model with high accuracy. Some recent studies also utilized machine learning approaches (i.e. ANN), which is advantageous as it
optimize complex growing systems and simulate the growth and yield of banana. No studies have previously been conducted to estimate banana fibre yield using mathematical approaches, thus highlighting the necessity of developing models to investigate the mechanism of banana fibre production. Overall, the present systematic review assesses evidence-based modelling efforts made by previous researchers to estimate banana growth and yield. However, different predictor variables, growing conditions, varietal differences, modelling approaches and evaluation procedures used in the existing models restrict the comparative assessment and selection of the best model. Hence, accurate, novel and informative models for estimating banana growth and yield of both fruit and fibre are encouraged, utilizing appropriate predictors and algorithms.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/S0021859622000259

**Author contributions.** P. E. K. conceived the idea for the topic and P. E. K. and S. L. J. contributed to the conception and systematic bibliography search. S. L. J. retrieved the data. P. E. K., S. L. J., C. J. R., and I. C. L. performed statistical analyses. P. E. K. and S. L. J. and I. C. L. drafted the manuscript; P. E. K. and C. J. R. revised the manuscript and approved the final version. All authors read and approved the final manuscript.

**Financial support.** This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education funded by the World Bank.

**Conflict of interest.** The authors have no conflicts of interest to declare.

**Ethical standards.** Not applicable.

**References**


Demirsoy H (2009) Leaf area estimation in some species of fruit tree by using models as a non-destructive method. *Fruits* 64, 45–51.


De Swaef T, Bellocci G, Aper J, Lootens P and Roldán-Ruiz I (2019) Use of identifiability analysis in designing phenotyping experiments for model-


and evaluation of cropping potential in selected banana cultivars (Musa sp.) cultivated in subtropical conditions of coastal Alabama. *HortScience* 53, 1634–1639.


