Sugar Tech (Mar–Apr 2026) 28(2):667–674
https://doi.org/10.1007/s12355-026-01737-z

RESEARCH ARTICLE

Estimating Sugar Yield in Sugarcane Using Green Normalized Difference Vegetation Index Derived from Imagery Obtained by Remotely Piloted Aircrafts

Julio Cezar Souza Vasconcelos[1] · Caio Simplicio Arantes[1] · Eduardo Antonio Speranza[2] · João Francisco Gonçalves Antunes[2] · Luiz Antonio Falaguasta Barbosa[2] · Geraldo Magela de Almeida Cançado[2]

Received: 29 September 2025 / Accepted: 5 February 2026 / Published online: 3 March 2026
© The Author(s) 2026

Abstract

Sugarcane (*Saccharum officinarum* L.) is one of the largest crops in Brazil, and its productivity varies according to the environment and management practices adopted. In this study, tons of sugar per hectare (TSH) are estimated using a heteroscedastic gamma (GA) regression model, which considers several explanatory variables, one of which is the normalized difference green vegetation index (GNDVI), obtained from multispectral images in two locations over two consecutive growing seasons. The modeling considers regression structures in the parameters representing the mean and coefficient of variation, respectively. The results show that there is an influence of location, cultivar, cycle, accumulated precipitation, and GNDVI. To verify if the model is well-fitted to the data, the analysis of quantile residuals shows that the model is adequate. Therefore, the results indicate that heteroscedastic GA regression is an alternative model for predicting TSH and can assist in decision-making in sugarcane cultivation.

Keywords GNDVI · Heteroscedastic GA regression model · *Saccharum officinarum* L · TSH

Introduction

Sugarcane (*Saccharum officinarum* L.) is a vital crop for Brazilian agriculture, serving as a primary source for sugar, biofuel, bio-electricity, and renewable raw materials. As the world's largest producer, Brazil's success is underpinned by the continuous adoption of new technologies. Monitoring the performance of this extensively cultivated crop in real time using traditional methods is impractical, thereby driving the need for modern, efficient approaches.

Remote sensing, using satellites or remotely piloted aircraft systems (RPAS), provides a viable solution for large-scale crop monitoring. A standard methodology involves calculating vegetation indices (VIs) to track canopy variations and correlate them with agronomic variables. While the normalized difference vegetation index (NDVI) is widely used, the green normalized difference vegetation index (GNDVI) (Gitelson et al. 1996) leverages the green spectral band. It is often more sensitive to chlorophyll concentration and nitrogen status, making it a powerful tool for assessing crop health and photosynthetic activity.

These indices are strongly linked to key biophysical properties. Beyond predicting biomass yield, there is a growing

These authors contributed equally to this work.

✉ Geraldo Magela de Almeida Cançado
geraldo.cancado@embrapa.br

Julio Cezar Souza Vasconcelos
juliocezarvasconcelos@hotmail.com

Caio Simplicio Arantes
caio.arantes@colaborador.embrapa.br

Eduardo Antonio Speranza
eduardo.speranza@embrapa.br

João Francisco Gonçalves Antunes
joao.antunes@embrapa.br

Luiz Antonio Falaguasta Barbosa
luiz.barbosa@embrapa.br

[1] Fundação de Apoio à Pesquisa e ao Desenvolvimento – FAPED, Sete Lagoas, Minas Gerais, Brazil

[2] Embrapa Digital Agriculture, Campinas, São Paulo, Brazil

interest in using them to forecast sugar content–a critical factor for milling efficiency and profitability. GNDVI, in particular, has shown promise for estimating sucrose accumulation because it correlates with photosynthetic efficiency and plant vigor. Although the normalized difference vegetation index (NDVI) remains the most commonly employed VI in these studies, VIs such as GNDVI are more effective in representing the properties of chlorophyll in green plants (Taddeo et al. 2019; Dixon et al. 2021). Considering the final stages of the sugarcane production cycle, such as growth and maturation, GNDVI may present high correlations with measures related to the presence of sucrose with and without impurities (brix and pol) (Leandro et al. 2024).

Typically, linear regression is commonly used to model the relationship between explanatory and response variables. Nevertheless, standard models may be insufficient when residual variability is not constant across observations, a phenomenon known as heteroscedasticity. To address this, heteroscedastic regression models have been widely applied in agricultural research, extending classical regression by allowing the dispersion of the response variable to depend on covariates, in addition to modeling its mean. For instance, Prataviera et al. (2022) modeled location and scale parameters of the skew-t distribution for soil chemistry data, accommodating asymmetric residuals and non-constant variance, while Vasconcelos et al. (2021) proposed an extension of the exponential Gaussian distribution with two systematic components to model bimodal seed data. Other studies have also used heteroscedastic regression models, for example, to analyze soil clay dispersion (Batista et al. 2022), crop productivity under microbial inoculants (Santos et al. 2024), and the effects of chemical treatments on soybeans (Vasconcelos et al. 2025a).

Research demonstrates that unmanned aerial vehicle (UAV)-based hyperspectral imaging, coupled with machine learning, provides a non-invasive and efficient method for predicting tons of sugar per hectare (TSH) in sugarcane breeding programs. In a study by Poudyal et al. (2022), which integrated aerial imagery with ground data in the final selection stage, a gradient boosting regression tree model was selected for its low prediction error to estimate key yield variables, including TSH, tons of cane per hectare (TCH), and sucrose concentration. The results indicated that sucrose percentage could be predicted with high accuracy (94% in plant cane and 93% in first ratoon), outperforming the predictive accuracy for TSH and TCH themselves (Poudyal et al. 2022)

Given this, this work uses a heteroscedastic gamma (GA) regression model to predict tons of sugar per hectare (TSH), with one of the explanatory variables the GNDVI index derived from multispectral images. By applying the model to these data, we provide a new approach for predicting tons of sugar per hectare of sugarcane.

Materials and Methods

Field Experiments

This study was conducted during the agricultural seasons of 2020/2021 and 2021/2022 using two commercial sugarcane fields in Brazil, located in the municipalities of Piracicaba (−22.773005, −47.580135) and Tambaú (−21.708543, −47.246643), both in the state of São Paulo as indicated in the Fig. 1. In the Piracicaba field, the cane plant cycle was planted in March 2020 and harvested in April 2021, with a growing period of thirteen months. While the first ratoon cycle was harvested in February 2022, with a growing period of 10 months. In the field in Tambaú, the cane plant cycle was planted in June 2020 and harvested in February 2021, with a growing period of 9 months, while the first ratoon was harvested in March 2022, with a growing period of 13 months.

Both fields were planted with four commercially relevant Brazilian sugarcane varieties, selected to represent different maturation cycles (Table 1). The experimental site in Piracicaba used two-month-old pre-sprouted seedlings (PSS) for each variety, which were initially cultivated in a greenhouse. A drip irrigation system was implemented during the first 3 months after transplanting the plantlets into the field to promote robust root development. In contrast, the experimental site in the rural area of Tambaú employed sugarcane stalks with an average length of 1.5 m as plantlets for direct field planting. All management practices and agronomic operations adhered to regional commercial standards and were maintained uniformly across treatments in both experimental sites during the plant cane cycle and the first ratoon cycle.

The soil of the rural area of Piracicaba, São Paulo, is predominantly characterized by deep, highly weathered Latosols (Oxisols in the Brazilian Soil Classification System, SiBCS), which are typical of tropical regions on ancient erosion surfaces (de Sousa Mendes and Dematte 2022). While naturally acidic and of low to moderate fertility due to the prevalence of kaolinite and iron oxides in the clay fraction, these Latosols form the foundation for extensive sugarcane cultivation (Solos et al. 2013). In contrast, the soils in the rural area of Tambaú, also within São Paulo State, exhibit greater variability due to local differences in parent material and relief. While initially cultivated in a greenhouse, Oxisols may also be present; the landscape likely includes Argisols, identified by an increase in clay content in the subsurface horizon, and other orders, such as Cambisols. Texture can correspondingly vary from sandy to clayey, influencing water and nutrient retention capacities. As with the broader region, many of these soils require careful fertility management

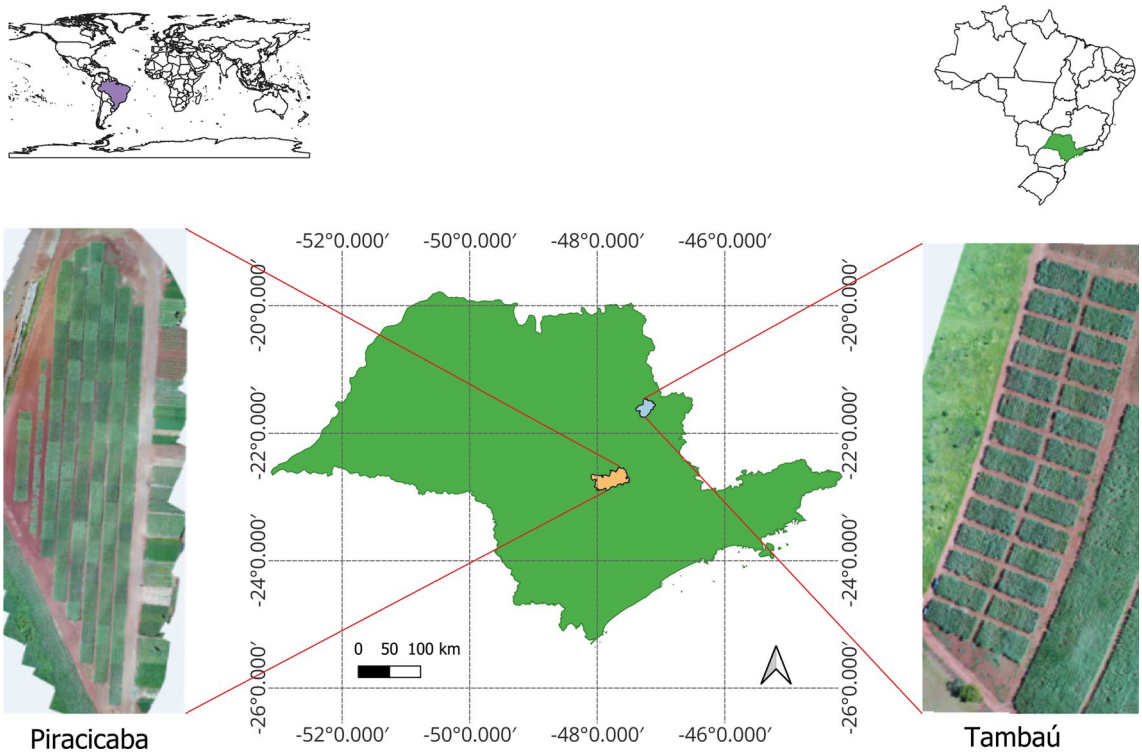

Fig. 1 Sugarcane experimental fields location (Tambaú and Piracicaba) during two consecutive growing seasons: 2020/2021 and 2021/2022

Table 1 Sugarcane varieties

Sugarcane genotype	Code	Piracicaba	Tambaú	Variety traits
1	V1	CTC 1007	CTC 1007	Normal cycle
2	V2	RB 966,928	RB 966,928	Early cycle
3	V3	CV 0618	CV 0618	Early to normal cycle
4	V4	CV 7870	CV 7870	Normal cycle

to overcome natural acidity and low base saturation for effective agricultural use (Da Silva et al. 2025). Table 2 presents the average soil fertility and texture properties measured before the installation of experimental sugarcane fields located in Piracicaba and Tambaú in 2020.

Both experiments adopted a randomized complete block design with 28 replicates per variety. The plots in Piracicaba consisted of four 9-m rows (54 m^2), while in Tambaú, each plot had four 10-m rows (60 m^2). Each experiment consisted of 112 plots, and the total number of experimental plots throughout two growing seasons and two locations was 448. At the end of each cycle, after harvest, two productivity-related metrics were evaluated: the tons of cane per hectare (TCH), as described in Vasconcelos et al. (2025b), and the total recoverable sugars (TRS), expressed as kilograms of sugar per sample and obtained by randomly collecting 10 stalks from each experimental plot. Subsequently, the total sugar per hectare (TSH) was calculated.

According to Gilbert et al. (2006), the TSH is calculated as the product of TCH, representing crop biomass yield, and the kilograms of sugar per ton of cane (KST), representing the technological quality or sucrose concentration of the harvested cane. The value of TSH levels is calculated using the Eq. 1 as follows below:

$$\text{TSH}\ [t/ha] = \frac{\text{TCH}\ [t/ha] \times \text{KST}\ [kg/t]}{1000} \quad (1)$$

Table 2 Average fertility and physical texture of soils from experimental fields

Location	pH	Ca	Mg	K	P	S	Mn	Fe	Cu	Zn	B	OM	Clay	Silt	Sand
	$CaCl_2$	mmolc/dm^3			mg/dm^3							g/dm^3	g/kg		
Piracicaba	5.7	49	27	3	35	7	90	22	1.9	3.6	0.93	28	535	170	295
Tambaú	5.2	35	19	6.9	61	17	45	130	2.6	2.9	0.52	27	358	176	466

Weather Data

The climate data for the experimental fields were compiled from the NASA POWER database, as indicated by Monteiro et al. (2018), for the entire period during which the sugarcane crop was growing. The dataset was uploaded directly from (Power 2022). Accumulated precipitation (in millimeters) was quantified during the phenological growth phase of the cane plant and first ratoon cycles. For the experiment in Piracicaba, this phase was from October 2020 to December 2020 for the first harvest and from November 2021 to January 2022 for the second harvest (Fig. 2a). For the experiment in Tambaú, this phase was from October 2020 to January 2021 for the first harvest and from October 2021 to February 2022 for the second harvest (Fig. 2b). Figure 2a and b show the GNDVI profile over the two harvests at the two locations and guided the selection of the growth phases. The accumulated precipitation indicator (Fig. 2a and b) enabled the estimation of water availability in the environment during the phenological growth phase and the evaluation of the role of this occurrence in the TSH variable.

Vegetation Indexes

The GNDVI (Eq. 2) was computed for all experimental plots during the late developmental phenological stage. A high-resolution orthomosaic was constructed from a set of aerial images acquired monthly using a DJI Phantom 4 Pro unmanned aerial vehicle (UAV). This platform was equipped with a dual-sensor payload, comprising an RGB camera for the visible spectrum and a multispectral camera Sentera Single Sensor NDVI (https://senterasensors.com/) capturing the following wavebands: blue (B): 450 nm ± 16 nm; green (G): 560 nm ± 16 nm; red (R): 650 nm ± 16 nm; red edge (RE): 730 nm ± 16 nm; and near-infrared (NIR): 840 nm ± 26 nm. Flights were conducted following a pre-defined flight plan between 11:00 AM and 1:00 PM at a standardized altitude, with no clouds.

$$\frac{(NIR - G)}{(NIR + G)} \tag{2}$$

From the monthly acquisitions, only the imagery that best represented the target phenological stage (approximately 240 days after planting) was selected for subsequent index extraction. Orthomosaics for the studied indices, including the GNDVI, were generated at a spatial resolution of approximately 5 cm/pixel. For the following analysis, the mean GNDVI of all pixels within the boundaries of each experimental plot was used.

Flights were conducted following a pre-defined flight plan, scheduled between 11:00 AM and 1:00 PM, at a standardized altitude, and with no cloud cover; therefore, the radiometric correction of images was not necessary. The images were evaluated using R software with the FieldImageR package (Matias et al. 2020).

Statistical Model

All explanatory variables, including the response variable (TSH), were initially evaluated using a normal linear model. However, the Shapiro–Wilk test for normality (Shapiro and Wilk 1965) and the Breusch–Pagan test for heteroscedasticity (Breusch and Pagan 1979) indicated violations of model assumptions. As documented in the literature, such violations can impair the adequacy of the normal linear model (Antunes et al. 2016; Ng and Cribbie 2019; Feng et al. 2020; Amado et al. 2025), thereby justifying the use of alternative approaches. For this reason, a heteroscedastic GA regression model was adopted, with regression structures specified for both the mean (μ) and the coefficient of variation (CV) (σ).

It is worth noting that this formulation is a fully parametric case of the more general semiparametric heteroscedastic GA regression model proposed by Santos et al. (2024), in which nonlinear covariate effects can also be incorporated via smooth functions. Furthermore, the same modeling framework was previously applied by Vasconcelos et al. (2025b) to predict sugarcane yield, where, in both studies, the response variable considered was TCH. In the present work, only linear effects were taken into account, leading to the following regression structures:

$$\begin{aligned}
g(\mu_i) = \log(\mu_i) &= \beta_0 + \beta_1 \text{Block2}_i + \beta_2 \text{Block3}_i + \beta_3 \text{Block4}_i \\
&+ \beta_4 \text{Location_Tambau}_i \\
&+ \beta_5 \text{Cultivar_V2}_i + \beta_6 \text{Cultivar_V3}_i + \beta_7 \text{Cultivar_V4}_i \\
&+ \beta_8 \text{Cycle_Ratoon}_i + \beta_9 \text{Precipitation}_i + \beta_{10} \text{GNDVI}_i, \\
g(\sigma_i) = \log(\sigma_i) &= \gamma_0 + \gamma_1 \text{Block2}_i + \gamma_2 \text{Block3}_i + \gamma_3 \text{Block4}_i \\
&+ \gamma_4 \text{Location_Tambau}_i \\
&+ \gamma_5 \text{Cultivar_V2}_i + \gamma_6 \text{Cultivar_V3}_i + \gamma_7 \text{Cultivar_V4}_i \\
&+ \gamma_8 \text{Cycle_Ratoon}_i + \gamma_9 \text{Precipitation}_i + \gamma_{10} \text{GNDVI}_i.
\end{aligned}$$

where $g(\cdot)$ is the logarithmic link function, i.e., $g(\mu_i) = \log(\mu_i)$ and $g(\sigma_i) = \log(\sigma_i)$, and the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ represent the coefficients to be estimated associated with the covariates in the mean and the coefficient of variation, respectively.

Predictors include categorical factors - block (1 to 4, Block 1 as reference), location (Piracicaba or Tambaú, Piracicaba as reference), cultivar (V1 to V4, V1 as reference), and cycle (cane plant or first ratoon, cane plant as reference); and continuous variables - cumulative precipitation (mm) during the growth phenological phase and mean GNDVI calculated throughout this phase. The

response variable (TSH) has a GA(μ,σ) distribution and corresponds to the value measured at the end of each cycle.

For an independent sample $y_1, \ldots, y_n$, the log-likelihood under this GA(μ,σ) model can be written as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \left(\frac{1}{\sigma_i^2} - 1 \right) \log(y_i) - \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2 \mu_i} - \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \log(\sigma_i^2 \mu_i) - \sum_{i=1}^{n} \log \Gamma\left(\frac{1}{\sigma_i^2} \right)$$

where μ_i and σ_i follow the regression structures defined above.

Covariate selection was performed using the *stepGAICAll.A()* function of the `gamlss` package in the **R** software (Stasinopoulos and Rigby 2008). Each variable is examined separately with respect to distribution parameters using the generalized Akaike information criterion (GAIC), which makes it possible to identify different sets of covariates for both the mean and variability. This procedure uses the `stepAIC()` function from the 'MASS' package (Stasinopoulos et al. 2017; Ripley et al. 2013).

Table 3 Mean and coefficient of variation (CV) of TSH by Location, Cultivar, and Planting Cycle

Category	Level	Mean	CV (%)
Location	Piracicaba	12.26	33.89
	Tambau	13.96	28.15
Cultivar	V1	12.28	28.63
	V2	15.04	28.14
	V3	13.18	30.73
	V4	11.91	33.53
Cycle	Cane Plant	9.96	24.02
	First Ratoon	16.25	17.97

Residuals Analysis

Analyzing the residuals is important for verifying whether the model adequately explains the data. In the case of GA regression with heteroscedasticity, the quantile residuals can be obtained by the transformation:

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i),$$

where Φ^{-1} represents the quantile function of the standard normal distribution and $\hat{u}_i$ corresponds to the estimated quantile residuals (Stasinopoulos and Rigby 2008).

Results and Discussion

Descriptive Analysis for Model Implementation

Table 3 summarizes the estimates of mean and coefficient of variation (CV) of sugar productivity per hectare (TSH, t/ha), organized by location, cultivar, and planting cycle. Considering the evaluated locations, Tambau presented the highest average TSH value (13.96 t/ha), whereas Piracicaba registered a lower yield (12.25 t/ha). The dispersion was more pronounced in Piracicaba (CV = 33.89%), while Tambau exhibited less variation (CV = 28.15%), suggesting greater productive uniformity in this environment. Among the evaluated cultivars, material V2 stood out with the highest average TSH (15.04 t/ha), whereas V4 presented the lowest (11.91 t/ha). The coefficients of variation indicate greater variability for V4 (33.53%) and less for V2 and V1 (28.14% and 28.63%,

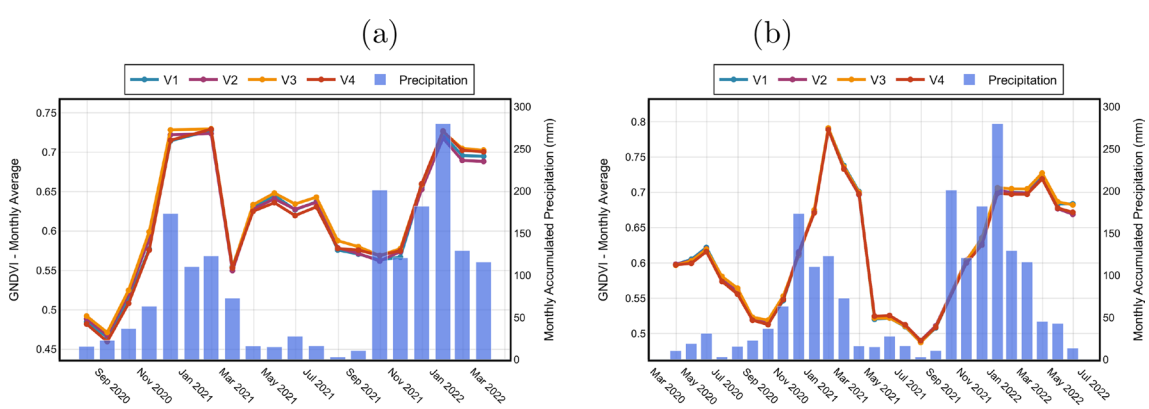

Fig. 2 Monthly GNDVI and precipitation (mm) for the locations of (**a**) Piracicaba and (**b**) Tambaú, during two consecutive growing seasons

respectively), evidencing greater productive stability of these two materials.

Finally, examining the Planting Cycles, the first ratoon cycle had the highest average (16.25 t/ha), whereas the cane plant cycle had the lowest (9.96 t/ha). The variability was greater in the cane plant cycle (CV = 24.02%) and lower in the first ratoon cycle (CV = 17.97%), indicating more consistent production during the latter.

Figure 2 shows the average monthly GNDVI pattern for each sugarcane cultivar over the two harvests and the rainfall distribution for the locations of Piracicaba (Fig. 2a) and Tambaú (Fig. 2b). The early stages of sugarcane growth, as illustrated by peaks in GNDVI readings, precede periods of higher precipitation, which, to a certain extent, favor crop growth due to the increased availability of soil moisture.

Results Obtained from Fitting the Heteroscedastic GA Regression Model

The explanatory variables included in the systematic components for the mean μ_i and for the coefficient of variation σ_i were selected using the *stepGAICAll.A()* method (Stasinopoulos and Rigby 2008), as previously described in the Materials and Methods section. During the selection process, some covariates were removed from the variability component.

The final fitted model can be expressed as:

$$\begin{aligned} g(\mu_i) = {} & \beta_0 + \beta_1 \text{Block2}_i + \beta_2 \text{Block3}_i + \beta_3 \text{Block4}_i + \beta_4 \text{Location_Tambau}_i \\ & + \beta_5 \text{Cultivar_V2}_i + \beta_6 \text{Cultivar_V3}_i + \beta_7 \text{Cultivar_V4}_i \\ & + \beta_8 \text{Cycle_Ratoon}_i + \beta_9 \text{Precipitation}_i + \beta_{10} \text{GNDVI}_i, \\ g(\sigma_i) = {} & \gamma_0 + \gamma_1 \text{Location_Tambau}_i + \gamma_2 \text{Precipitation}_i. \end{aligned}$$

observed that factors such as location, cultivar (with V3 and V4 differing statistically from V1, used as the reference), ratoon cycle compared to the plant cane cycle (used as reference), accumulated precipitation, and the GNDVI index have a significant effect on the average sugar yield per hectare.

Sugarcane is very sensitive to temperature, rainfall, solar radiation, and other environmental factors, which significantly affect both its production and sugar yield (Srivastava and Rai 2012). Rainfall plays a fundamental role in crop development, especially during the vegetative phase, when water availability favors stem elongation and internode formation. On the other hand, lower rainfall during the maturation stage improves juice quality and reduces tissue moisture. Given this contrasting behavior throughout the cycle, accumulated rainfall during the growth period was included as a response variable in the TSH modeling.

Estimates of the parameters associated with variability (σ) indicate that the TSH coefficient of variation is influenced by location, suggesting that local conditions affect the consistency of TSH yield. This result aligns with practical expectations, as different locations vary in environmental characteristics, topography, and soil composition, which can influence sugarcane growth and sugar accumulation.

Figure 3 shows the quantile residuals (QRs) along with the corresponding envelope for the fitted heteroscedastic GA regression model. No outlier observations were detected, suggesting that the model provides an adequate fit to the TSH data.

Table 4 presents the estimates of the parameters of the heteroscedastic GA regression model fitted for TSH. It can be

Table 4 Estimates of the parameters of the heteroscedastic GA regression model for TSH

Component	Effect	Parameter	Estimate	Std. Error	p-value
μ	Intercept	β_0	0.5021	0.3034	0.0990
	Block2	β_1	−0.0524	0.0220	0.0175
	Block3	β_2	−0.0537	0.0220	0.0152
	Block4	β_3	−0.0469	0.0217	0.0312
	Location_Tambau	β_4	0.2899	0.0337	<0.0001
	Cultivar_V2	β_5	0.2027	0.0214	<0.0001
	Cultivar_V3	β_6	0.0824	0.0233	0.0004
	Cultivar_V4	β_7	−0.0307	0.0220	0.1634
	Cycle_Ratoon	β_8	0.5452	0.0193	<0.0001
	Precipitation	β_9	−0.0004	0.0001	<0.0001
	GNDVI	β_{10}	3.0360	0.5242	<0.0001
σ	Intercept	γ_0	−1.4675	0.1395	<0.0001
	Location_Tambau	γ_4	−0.3270	0.0791	<0.0001
	Precipitation	γ_9	−0.0004	0.0003	0.1040

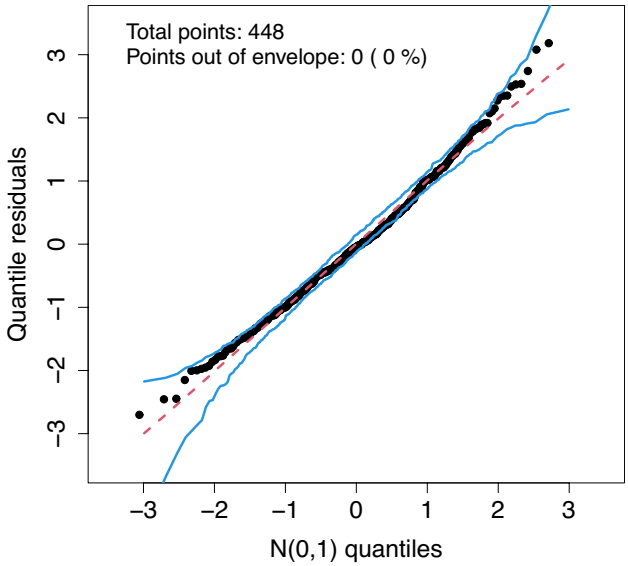

Fig. 3 Normal probability plot of quantile residuals with envelope for the heteroscedastic GA model on TSH data

Table 5 Multiple comparison results from the heteroscedastic GA regression model applied to TSH data

Hypotheses H_0	Estimate	SE	p-value
Cultivar_nV2 - Cultivar_V3	−0.1203	0.0224	<0.0001
Cultivar_V2 - Cultivar_V4	−0.2333	0.0217	<0.0001
Cultivar_V3 - Cultivar_V4	−0.1131	0.0227	<0.0001

Table 5 presents the results of multiple comparisons obtained from the heteroscedastic GA regression model adjusted for TSH. In the main table, V1 was used as the reference for comparison purposes. Afterward, the references were shuffled, first considering V2 relative to V3 and V4, then V3 relative to V4. The results show statistically significant differences at the 5% level between V2 and V3, V2 and V4, and V3 and V4, highlighting that cultivar choice has a substantial impact on productivity in TSH.

Conclusions

The use of the heteroscedastic GA regression model allowed us to identify how different agronomic and environmental variables can influence TSH (total sugar yield). Variables such as location, cultivar, cycle, accumulated precipitation, and the GNDVI index have significant influences, indicating that management practices adapted to each environment and genetic material are essential. Residual analysis confirms that the model is adequate to explain these data. Thus, the adopted method constitutes a consistent strategy for estimating TSH and making management decisions, in addition to contributing to the advancement of precision agriculture applications. Other research is underway and addresses the practical limitations associated with the use of aerial images obtained by RPAS (Remotely piloted aircraft systems), favoring the adoption of orbital products with greater coverage and efficiency for monitoring extensive sugarcane areas.

Funding The Article Processing Charge (APC) for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614). This research was funded through collaborative grants from the Brazilian Agricultural Research Corporation (Embrapa) and the São Paulo State Sugarcane Growers' Cooperative (Coplacana) [Grant SEG 30.19.90.005.00.00], as well as Embrapa, the Simbiose Company, and the Brazilian Funding Authority for Studies and Projects (Finep) [Grant SEG 20.24.00.110.00.00]. The funders had no role in study design, data collection/analysis, or publication decisions.

Declarations

Conflict of interest The authors declare that there is no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Amado, C., A. M. Bianco, G. Boente, and I. M. Rodrigues. 2025. Robust estimation of heteroscedastic regression models: a brief overview and new proposals. *Statistical Papers* 66 (3): 1–30.

Antunes, F., A. O'Sullivan, F. Rodrigues, and F. Pereira. 2016. A review of heteroscedasticity treatment with gaussian processes and quantile regression meta-models. Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics: 141–160.

Batista, A. M., P. L. Libardi, M. E. Alves, F. Prataviera, and N. F. B. Giarola. 2022. Electrochemical effects on clay dispersion in rhizo-and non-rhizospheric soils. *Journal of Soil Science and Plant Nutrition* 22 (3): 3518–3526.

Breusch, T.S. and A.R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the econometric society: 1287–1294.

da Silva, J.P., J.Z. Junior, L.A.S. Romani, and E.D. Assad. 2025. Soil fertility characterization in brazilian pastures through soil chemical and physical analysis. Geoderma Regional: e01010.

de Sousa Mendes, W., and J. A. Dematte. 2022. Digital soil mapping outputs on soil classification and sugarcane production in brazil. *Journal of South American Earth Sciences* 116 : 103881.

Dixon, D. J., J. N. Callow, J. M. Duncan, S. A. Setterfield, and N. Pauli. 2021. Satellite prediction of forest flowering phenology. *Remote Sensing of Environment* 255 : 112197.

Feng, C., L. Li, and A. Sadeghpour. 2020. A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology* 20 (1): 175.

Gilbert, R. A., J. Shine Jr., J. D. Miller, R. W. Rice, and C. Rainbolt. 2006. The effect of genotype, environment and time of harvest on sugarcane yields in florida, usa. *Field Crops Research* 95 (2–3): 156–170.

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak. 1996. Use of a green channel in remote sensing of global vegetation from eos-modis. *Remote sensing of Environment* 58 (3): 289–298.

Leandro, E. R., M. K. Heenkenda, and K. F. Romero. 2024. Estimating sugarcane maturity using high spatial resolution remote sensing images. *Crops* 4 (3): 333–347.

Matias, F. I., M. V. Caraza-Harter, and J. B. Endelman. 2020. Field-imager: an r package to analyze orthomosaic images from agricultural field trials. *The Plant Phenome Journal* 3 (1) : e20005.

Monteiro, L. A., P. C. Sentelhas, and G. U. Pedra. 2018. Assessment of nasa/power satellite-based weather system for brazilian conditions and its impact on sugarcane yield simulation. *International Journal of Climatology* 38 (3): 1571–1581.

Ng, V. K., and R. A. Cribbie. 2019. The gamma generalized linear model, log transformation, and the robust yuen-welch test for analyzing group means with skewed and heteroscedastic data. *Communications in Statistics-Simulation and Computation* 48 (8): 2269–2286.

Poudyal, C., L. F. Costa, H. Sandhu, Y. Ampatzidis, D. C. Odero, O. C. Arbelo, and R. H. Cherry. 2022. Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based hyperspectral imaging and machine learning. *Agronomy Journal* 114 (4): 2320–2333.

Power, N. 2022. Data access viewer available online: https://power larc. nasa. gov/data-access-viewer. Last accessed 11(10).

Prataviera, F., A. M. Batista, P. L. Libardi, G. M. Cordeiro, and E. M. M. Ortega. 2022. Joint regression modeling of location and scale parameters of the skew t distribution with application in soil chemistry data. *Journal of Applied Statistics* 49 (1): 195–213.

Ripley, B., B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth, and M. B. Ripley. 2013. Package 'mass'. *Cran r* 538 (113–120): 822.

Santos, D.P.d., A. Soares, G. de Medeiros, D. Christofoletti, C.S. Arantes, J.C.S. Vasconcelos, E.A. Speranza, L.A.F. Barbosa, J.F.G. Antunes, and G.M.d.A. Cancado. 2024. Evaluation of sugarcane yield response to a phosphate-solubilizing microbial inoculant: Using an aerial imagery-based model. *Sugar Tech* 26 (1): 143–159.

Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52 (3–4): 591–611.

Solos, E. et al. 2013. Sistema brasileiro de classificação de solos. Centro Nacional de Pesquisa de Solos: Rio de Janeiro 3.

Srivastava, A.K. and M.K. Rai. 2012. Sugarcane production: Impact of climate change and its mitigation. Biodiversitas Journal of Biological Diversity 13(4).

Stasinopoulos, D. M., and R. A. Rigby. 2008. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software* 23:1–46.

Stasinopoulos, M.D., R.A. Rigby, G.Z. Heller, V. Voudouris, and F. De Bastiani. 2017. Flexible regression and smoothing: using GAMLSS in R. CRC Press, Taylor & Francis Group.

Taddeo, S., I. Dronova, and N. Depsky. 2019. Spectral vegetation indices of wetland greenness: Responses to vegetation structure, composition, and spatial distribution. *Remote sensing of Environment* 234 : 111467.

Vasconcelos, J.C., A.K. Suzuki, É.M.d. Rezende, E.M. Ortega, G.M. Cordeiro, J.A. Oliveira, and R. Vila. 2025. Bimodal heteroskedastic bivariate regression with application to soy data. Communications in Statistics-Simulation and Computation: 1–26.

Vasconcelos, J. C. S., C. S. Arantes, E. A. Speranza, J. F. G. Antunes, L. A. F. Barbosa, and G.M.d.A. Cançado. 2025. Predicting sugarcane yield through temporal analysis of satellite imagery during the growth phase. *Agronomy* 15 (4): 793.

Vasconcelos, J. C. S., G. M. Cordeiro, E. M. M. Ortega, and É.M.d. Rezende. 2021. A new regression model for bimodal data and applications in agriculture. *Journal of Applied Statistics* 48 (2): 349–372.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.