

Comparative Performance of Response Surface Methodology and Linear Regression Models in Predicting Sugarcane Yield in Zimbabwe

Edwin Rup^{1,2*}, Philimon Nyamugure³, Peter Chimwanda², Precious Mdlongwa³, Thomas Musora²

¹Sciences Department, Masvingo Teachers College, Masvingo, Zimbabwe

²Department of Mathematics and Statistics, Chinhoyi University of Technology, Chinhoyi, Zimbabwe

³Department of Statistics and Operations Research, National University of Science and Technology, Bulawayo, Zimbabwe

Email: *edwinrupi@gmail.com

How to cite this paper: Rup¹, E., Nyamugure, P., Chimwanda, P., Mdlongwa, P. and Musora, T. (2026) Comparative Performance of Response Surface Methodology and Linear Regression Models in Predicting Sugarcane Yield in Zimbabwe. *Open Journal of Statistics*, 16, 382-400.

<https://doi.org/10.4236/ojs.2026.165019>

Received: July 27, 2026

Accepted: September 27, 2026

Published: September 30, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Accurate prediction of sugarcane yield is essential for optimizing agronomic inputs and enhancing productivity in commercial and smallholder production systems. This study compared the predictive performance of Response Surface Methodology (RSM) and conventional Multiple Linear Regression models for estimating sugarcane yield from agronomic and environmental variables. Field-level data on soil pH, organic matter content, fertilizer application rate, irrigation level, and planting density were analysed. A second-order RSM model was fitted using a Central Composite Design with 32 factorial points, 10 axial points and 2 centre points, while a multiple linear regression model was estimated under the conventional framework. Model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and Akaike Information Criterion (AIC). Results indicated that RSM outperformed conventional regression models in capturing interaction and quadratic effects, yielding higher predictive accuracy and lower error metrics. The quadratic Response Surface Model ($R^2 = 0.975$, adjusted $R^2 = 0.954$) performed well compared to the Multiple Linear Regression ($R^2 = 0.922$, adjusted $R^2 = 0.912$), achieving lower prediction errors (RMSE = 1.26, MAE = 0.98) and a substantially lower Akaike Information Criterion (AIC = 29.7 vs. 62.0), demonstrating greater efficiency for sugarcane yield estimation. The findings demonstrate the robustness of RSM for modelling complex agronomic systems and highlight its suitability for yield optimization strategies in sugarcane production. The study concluded that Response Surface Methodology is a superior modelling approach for sugarcane yield prediction and optimization, and recommends its integration into decision-support systems to guide fertilizer,

irrigation, soil pH, organic matter content, and planting density management for improved productivity and climate-resilient sugarcane production.

Keywords

Response Surface Methodology, Multiple Regression, Predictive Modelling, Agronomic Optimization, Ordinary Least Squares

1. Introduction

Sugarcane is one of the most economically significant cash crops in tropical and subtropical regions, contributing substantially to rural livelihoods, agro-industrial development, and national revenue generation. In countries such as Zimbabwe, sugarcane production supports major estates and out-grower schemes, particularly in irrigated low-veld regions. Optimizing yield is therefore a strategic priority for both commercial producers and policymakers. Complex interactions among soil properties, climatic variables, and agronomic inputs influence crop yield. Traditional modelling approaches, particularly multiple linear regression, have been widely used to estimate yields. However, linear models often fail to adequately capture the nonlinearities and interaction effects inherent in biological systems.

Response Surface Methodology (RSM), originally developed for industrial process optimization, has gained increasing application in agricultural research due to its ability to model curvature and interactions amongst factors. RSM employs designed experiments and second-order polynomial models to identify optimal input combinations and predict responses with improved accuracy. Despite the widespread use of both approaches, limited empirical studies have systematically compared the predictive performance of RSM and conventional regression models in sugarcane yield modelling. This study, therefore, aims to evaluate and compare these two modelling frameworks in predicting sugarcane yield using agronomic and environmental variables. Specifically, the study seeks to:

Develop yield prediction models using RSM and conventional regression approaches.

Compare their predictive accuracy using standard statistical performance metrics.

Assess their suitability for agronomic decision-making and yield optimization.

2. Literature Review

Accurately predicting sugarcane yield is critical in making informed decisions about agronomic inputs, especially in intensive production areas such as Chiredzi District, where irrigation-based farming systems dominate [1]. Traditionally, researchers have relied on conventional regression models, particularly multiple linear regression (MLR), to examine how yield responds to factors such as soil pH,

fertilizer application rates, irrigation levels, and organic matter content. Multiple linear regression is an extension of the simple linear regression to include more than one explanatory variable. The equation for multiple linear regression has the same form as that for simple linear regression, but has more terms:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i \quad (1)$$

These models are popular because they are straightforward to apply, easy to interpret, and require relatively modest computational effort. However, MLR assumes that relationships amongst variables are linear and that predictors are independent, which may be unrealistic in biological systems where factors interact nonlinearly [2]. As a result, conventional regression models may lose predictive accuracy under complex field conditions.

Response Surface Methodology (RSM), as outlined by [3], builds on traditional linear regression by incorporating quadratic and interaction terms within structured experimental designs such as the Central Composite Design (CCD) and Box-Behnken Design [4]. Equation (2) gives the form of the RSM model:

$$Y_i = \beta_0 + \sum \beta_i X_i + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j + \varepsilon \quad (2)$$

Compared to simple linear regression approaches, RSM is designed to model curvature and interactions amongst variables, making it particularly useful for optimization studies where the goal is to determine the best combination of input values [4]. In agricultural research, RSM has demonstrated improved performance in identifying interactive effects among soil nutrients and other factors, often producing higher coefficients of determination and better model fit than linear regression [4].

Historically, regression-based approaches have dominated crop yield modeling because they quantify how yield changes with explanatory variables in a clear mathematical form [5]. However, the approaches' reliance on linear assumptions limits their ability to capture interaction and nonlinear effects that are common in field conditions [2]. By contrast, RSM systematically estimates linear, quadratic, and interaction effects, enabling graphical interpretation (using contour plots and response surfaces) and optimization of conditions [4]. Comparative studies in other crops suggest that RSM generally outperforms conventional regression when nonlinear relationships dominate. Nonetheless, the comparative performance of RSM and traditional regression in sugarcane yield prediction under field-level variability is still relatively underexplored. Consequently, more flexible modelling techniques, including machine learning and advanced response-based modelling approaches, are increasingly applied to improve prediction accuracy and capture interactions among factors influencing crop productivity [5]. [6] employed diverse methods, in addition to multiple linear regression (MLR), three penalized regression techniques like the Least absolute shrinkage selection operator (LASSO), Ridge regression, Elastic Net (ELNET), and advanced machine learning approaches for sugarcane yield prediction.

Overall, existing literature indicates that RSM tends to provide more precise predictions and stronger optimization capacity when agronomic factors interact, while conventional regression remains useful for exploratory analysis and situations with limited or observational data. The choice between these methods should be based on the research goals, the organization of the data, and whether the main aim is straightforward prediction or optimization.

3. Materials and Methods

3.1. Study Area and Data Collection

Data used were from 44 researcher-established irrigated sugarcane experimental plots in Chiredzi, Zimbabwe, over one sugarcane production season. The plots were established and managed under strictly controlled experimental conditions, with the levels of the selected agronomic and environmental factors systematically manipulated according to the experimental design. The 44 experimental runs comprised 32 factorial points, 10 axial points, and 2 centre points, forming the Central Composite Design (CCD) used in the study. This controlled experimental arrangement enabled the systematic assessment of the individual and combined effects of the selected factors on sugarcane yield while providing sufficient experimental runs for fitting and evaluating the response surface model.

The following variables were included:

- Soil pH (X_1)
- Organic matter content (%) (X_2)
- Nitrogen fertilizer rate (kg/ha) (X_3)
- Irrigation level (mm) (X_4)
- Planting density (stalks/ha) (X_5)
- Sugarcane yield (tons/ha) (Response variable- Y)

3.2. Regression Models

A multiple linear regression model was estimated:

$$Y = \beta_0 + \sum_{i=1}^5 \beta_i X_i + \varepsilon \quad (3)$$

The model parameters were estimated using ordinary least squares (OLS).

A second-order RSM model was fitted using a Central Composite Design (CCD) to structure the experimental data and was specified as:

$$y = \beta_0 + \sum_{i=1}^5 \beta_i X_i + \sum_{i=1}^5 \sum_{j=i+1}^5 \beta_{ij} X_i X_j + \sum_{i=1}^5 \beta_{ii} X_i^2 + \varepsilon \quad (4)$$

where y = sugarcane yield, β_i = the model coefficients, β_{ii} = quadratic coefficients, β_{ij} = interaction coefficients, and X_i = the predictor variables, $X_i X_j$ are the interaction terms and X_i^2 are the quadratic terms in the model. To evaluate the predictive performance of the two modelling approaches, the two models were compared using the coefficient of determination (R^2), the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE) and the Akaike Information Criterion (AIC).

4. Results and Discussion

4.1. Descriptive Statistics of the Study Variables

Data comprising the five explanatory variables, namely soil pH, organic matter content, nitrogen fertilizer rate, irrigation level, and planting density, and sugarcane yield (tons/ha) considered as the response variable produced the descriptive statistics in **Table 1**.

Table 1. Descriptive statistics of variables used in sugarcane yield modelling.

	Descriptive Statistics				
	<i>N</i>	Minimum	Maximum	Mean	Std. Deviation
X_1	44	5.4	6.4	5.900	0.4446
X_2	44	1.2	2.5	1.850	0.5780
X_3	44	120	210	165.00	40.015
X_4	44	650	800	725.00	66.691
X_5	44	11000	12500	11750.00	666.909
Y	44	70	96	83.75	5.942
Valid <i>N</i> (list-wise)	44				

The results indicate moderate variability among the explanatory variables. Soil pH ranged from 5.4 to 6.4, suggesting that the soils in the sampled plots were slightly acidic to near neutral, which is generally suitable for sugarcane production. Organic matter content varied between 1.2% and 2.5%, reflecting differences in soil fertility across the farms. Nitrogen fertilizer application rates ranged from 120 to 210 kg/ha, while irrigation levels varied between 650 mm and 800 mm. Planting density ranged from 11,000 to 12,500 stalks per hectare. The average sugarcane yield across the experimental plots was 83.75 t/ha.

4.2. Conventional Multiple Linear Regression Model

To establish a baseline predictive model, an Analysis of Variance (ANOVA) was carried out to check on the significance of the individual variables to yield prediction and optimization. The results are shown in **Table 2**.

The Analysis of Variance (ANOVA), **Table 2**, evaluated the significance of the multiple linear regression model used to predict Yield (t/ha) from the five agro-nomic factors: Soil pH (X_1), Organic Matter (X_2), Nitrogen (X_3), Irrigation (X_4), and Plant Density (X_5). Since the p-value for each variable is less than 0.05, the regression model was statistically significant. This means that the combined effects of Soil pH, Organic Matter, Nitrogen, Irrigation, and Plant Density significantly influence sugarcane yield. The multiple linear regression model fitted to the data using the five explanatory variables was:

$$\text{Yield (t/ha)} = -16.70 + 5.176X_1 + 4.842X_2 + 0.09281X_3 + 0.02863X_4 + 0.002118X_5 \quad (5)$$

Table 2. Analysis of variance for yield versus soil pH, organic matter, nitrogen, irrigation, plant density.

Source	DF	Adj SS	Adj MS	F-Value	p-Value
Regression	5	1400.06	280.012	90.03	0.000
Soil pH (X_1)	1	227.76	227.765	73.23	0.000
Organic Matter (X_2)	1	336.74	336.735	108.26	0.000
Nitrogen (Kg/ha) (X_3)	1	593.06	593.059	190.68	0.000
Irrigation (mm) (X_4)	1	156.74	156.735	50.39	0.000
Plant Density (stalks/ha) (X_5)	1	85.76	85.765	27.57	0.000
Error	38	118.19	3.110		
Lack-of-Fit	37	117.69	3.181	6.36	0.306
Pure Error	1	0.50	0.500		
Total	43	1518.25			

The model accounts for a significant amount of the variability in yield, suggesting that it effectively predicts the experimental data. Each factor was tested to determine whether it significantly affected yield. Because all p-values are less than 0.05, each variable significantly contributed to explaining yield variation. The Mean Squared Error (MSE) = 3.110 represents unexplained variability in the model. A relatively small error variance indicates that the model predictions are close to the observed yield values. Root Mean Square Error (RMSE) measures the average magnitude of prediction errors, giving more weight to larger errors. This was calculated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} = \sqrt{\frac{797.0147}{44}} = 4.25605104 = 4.26 \quad (6)$$

The RMSE of 4.26 t/ha means that the model's predicted sugarcane yield differs from the observed yield by about ± 4.26 tonnes per hectare on average. This indicates high prediction accuracy, especially if the yield values range widely (for example, 70 - 96 t/ha in typical sugarcane studies). A small RMSE relative to the yield magnitude shows the model fits the data well.

The Mean Absolute Error (MAE) is one of the most widely used measures for evaluating the predictive accuracy of statistical and machine learning models because it quantifies the average magnitude of prediction errors without considering their direction. MAE is calculated as the average of the absolute differences between observed and predicted values, providing an intuitive measure of model performance expressed in the same units as the response variable [7]. Unlike error met-

rics based on squared residuals, such as the Root Mean Square Error (RMSE), MAE is less sensitive to extreme prediction errors, making it particularly useful when assessing the typical prediction accuracy of models in agricultural studies where biological and environmental variability may generate occasional outliers [8]. When comparing Multiple Linear Regression (MLR) and Response Surface Methodology (RSM), a lower MAE indicates that the model produces predictions that are, on average, closer to the observed crop yields, thereby demonstrating superior predictive performance. In this study, the lower MAE obtained for the RSM model confirms its enhanced ability to capture the complex nonlinear relationships and interaction effects among agronomic and climatic variables influencing sugarcane yield. Consequently, MAE complements other model evaluation statistics, such as the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Akaike Information Criterion (AIC), by providing a robust and easily interpretable measure of average prediction error, thereby facilitating the selection of the most reliable model for crop yield prediction and optimization [9] [10].

The Mean Absolute Error (MAE) was used to measure the average absolute difference between predicted and observed values and was approximated assuming normally distributed errors by:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |e_i| = \frac{1}{44} (167.41176) = 3.80481 = 3.80 \text{ tonnes/ha} \quad (7)$$

The value $\text{MAE} = 3.80 \text{ t/ha}$ is slightly smaller than RMSE, indicating that on average, the predicted sugarcane yield differs from the actual yield by about 3.80 tonnes per hectare. This confirms that large prediction errors are rare, meaning the regression model performs consistently. **Table 3** shows the model performance summary.

Table 3. Summary of model performance.

Model Summary									
Model	R	R Square	Adjusted R Squared	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	0.960 ^a	0.922	0.912	1.764	0.922	90.027	5	38	0.000

The lack-of-fit test (F-value = 6.36, p-value = 0.306) was used to check whether the regression model adequately represented the experimental data. Since $p > 0.05$, the lack of fit is not significant. This means that the regression model fits the data well and there is no evidence that the model is missing important terms. As shown in **Table 3**, about 92.2% of the variation in sugarcane yield is explained by the model, indicating a strong relationship between yield and the agronomic factors.

The Akaike Information Criterion (AIC) is a widely used model selection criterion for comparing competing statistical models, including Multiple Linear Regression (MLR) and Response Surface Methodology (RSM). Developed from information theory, AIC evaluates the relative quality of alternative models by sim-

ultaneously considering model goodness-of-fit and model complexity, thereby minimizing the risk of overfitting while maintaining predictive accuracy [11]. Unlike the coefficient of determination (R^2), which generally increases as additional parameters are included, AIC imposes a penalty for model complexity, favouring models that achieve an optimal balance between explanatory power and parsimony [7]. Consequently, models with lower AIC values are considered more efficient because they explain the observed data with fewer unnecessary parameters. In agricultural modelling, where crop yield is often influenced by complex non-linear relationships and interactions among agronomic and environmental variables, AIC provides an objective basis for selecting the most appropriate predictive model [10]. When comparing MLR and RSM, a lower AIC for the RSM model indicates that incorporating quadratic and interaction terms substantially improves model performance without introducing excessive complexity. Therefore, AIC serves as a valuable criterion for identifying models that are not only statistically adequate but also more likely to provide reliable predictions and robust optimization of crop yield under varying production conditions [7] [9].

The AIC for the Multiple Linear Regression model was calculated using the formula:

$$AIC = n \ln \left(\frac{RSS}{n} \right) + 2k \quad (8)$$

where n = number of experimental runs;

RSS = residual sum of squares;

k = number of model parameters (including the intercept).

In this case $MSE = \frac{RSS}{n}$, $RSS = n \times MSE = 3.11 \times 44 = 136.84$

$$AIC = 44 \ln \left(\frac{136.84}{44} \right) + 2(6) = 44 \ln(3.11) + 12 = 61.92 \approx 62.$$

The calculated Akaike Information Criterion (AIC) value of approximately 62.0 gives the information loss associated with the regression model. In model evaluation, lower AIC values indicate better models, as they achieve a better balance between goodness-of-fit and model complexity. Given that the model already demonstrates a high coefficient of determination $R^2 = 0.922$ and a relatively low standard error of estimate (1.764), the computed AIC value suggests that the model provides a strong fit to the data without excessive complexity. **Table 4** gives a summary of model coefficients.

4.3. Model Specification

Table 4, of coefficients, showed that all five predictors had a statistically significant positive effect on yield. Soil pH significantly increased sugarcane yield ($B = 5.176$, $t = 8.557$, $p < 0.001$), indicating that a unit increase in soil pH (X_1) increases yield by approximately 5.176 t/ha when other factors are held constant. Organic matter (X_2) also showed a strong positive effect ($B = 4.842$, $t = 10.405$, $p < 0.001$). Nitrogen (X_3) had the largest standardized effect ($\beta = 0.625$, $t = 13.809$, $p < 0.001$),

suggesting it is the most influential predictor of sugarcane yield. Irrigation (X_4) significantly contributed to yield improvement ($B = 0.029$, $t = 7.099$, $p < 0.001$), while plant density (X_5) had a smaller but still significant effect ($B = 0.002$, $t = 5.251$, $p < 0.001$).

The standardized coefficients indicate that nitrogen had the strongest influence on yield, followed by organic matter, soil pH, irrigation, and plant density. Furthermore, the 95% confidence intervals for all predictors did not include zero, confirming their statistical significance. The collinearity diagnostics showed tolerance values of 1.000 and variance inflation factors (VIF) of 1.000 for all predictors, indicating the absence multicollinearity among the independent variables and confirming the stability and reliability of the regression estimates.

Table 4. Table of coefficients.

Coefficients									
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	-16.699	6.766		-2.468	0.018	-30.396	-3.002		
X_1	5.176	0.605	0.387	8.557	0.000	3.952	6.401	1.000	1.000
X_2	4.842	0.465	0.471	10.405	0.000	3.900	5.784	1.000	1.000
X_3	0.093	0.007	0.625	13.809	0.000	0.079	0.106	1.000	1.000
X_4	0.029	0.004	0.321	7.099	0.000	0.020	0.037	1.000	1.000
X_5	0.002	0.000	0.238	5.251	0.000	0.001	0.003	1.000	1.000

4.4. Response Surface Methodology (RSM) Model

To capture the complex, nonlinear relationships between sugarcane yield and agronomic management practices, a second-order polynomial model based on Response Surface Methodology (RSM) was fitted. Response Surface Methodology is widely applied in agricultural optimization because it simultaneously estimates the individual, quadratic, and interaction effects of multiple factors, thereby providing a more realistic representation of biological production systems than linear models [12] [13]. In sugarcane production, crop responses to inputs such as soil fertility, nitrogen fertilizer, irrigation, and plant density are inherently nonlinear, with yield increasing up to an optimum level before declining due to diminishing returns or resource limitations. The inclusion of quadratic terms enables the model to capture this curvature, while interaction terms quantify the synergistic or antagonistic effects among agronomic factors, facilitating the identification of optimal management combinations under varying environmental conditions [14] [15]. Consequently, RSM provides a robust framework for modelling and optimizing sugarcane yield, supporting data-driven decision-making for sustainable

and climate-resilient crop production. **Table 5** shows the Analysis of Variance (ANOVA) for the response surface model where OM, N, and Density represent Organic matter, Nitrogen Fertilizer, and Plant Density, respectively. **Table 5** gives the Analysis of Variance for the Response Surface Methodology model.

Table 5. ANOVA table for quadratic RSM model.

Source	DF	Adj SS	Adj MS	F-Value	p-Value
Model	20	1436.92	71.846	45.37	<0.001
Linear	5	1148.36	229.672	145.01	<0.001
Soil pH	1	104.72	104.72	66.12	<0.001
Organic Matter	1	162.54	162.54	102.61	<0.001
Nitrogen	1	452.81	452.81	285.88	<0.001
Irrigation	1	264.43	264.43	166.94	<0.001
Plant Density	1	163.86	163.86	103.44	<0.001
Square	5	169.81	33.962	21.44	<0.001
pH × pH	1	18.47	18.47	11.66	0.002
OM × OM	1	34.11	34.11	21.53	<0.001
N × N	1	58.32	58.32	36.82	<0.001
Irrigation × Irrigation	1	37.66	37.66	23.77	<0.001
Density × Density	1	21.25	21.25	13.41	0.001
2-Way Interaction	10	118.75	11.875	7.50	<0.001
pH × OM	1	8.23	8.23	5.20	0.032
pH × N	1	6.41	6.41	4.05	0.056
pH × Irrigation	1	4.18	4.18	2.64	0.118
pH × Density	1	3.24	3.24	2.05	0.166
OM × N	1	17.56	17.56	11.09	0.003
OM × Irrigation	1	11.22	11.22	7.08	0.014
OM × Density	1	7.43	7.43	4.69	0.041
N × Irrigation	1	26.14	26.14	16.51	<0.001
N × Density	1	20.33	20.33	12.84	0.002
Irrigation × Density	1	14.01	14.01	8.85	0.007
Error	23	36.30	1.578		
Lack-of-Fit	22	35.30	1.605	1.61	0.48
Pure Error	1	1.00	1.000		
Total	43	1473.22			

The model summary statistics is shown in **Table 6**.

Table 6. Model summary statistics.

Statistic	R ²	Adjusted R ²	Predicted R ²	RMSE	MAE	AIC
Value	0.975	0.954	0.938	1.26 t/ha	0.98 t/ha	29.7

The quadratic response surface model is highly significant ($F = 45.37$, $p < 0.001$), indicating that the combination of soil pH, organic matter, nitrogen fertilizer, irrigation, and plant density explains a substantial proportion of the variation in sugarcane yield. All five linear effects are statistically significant, with nitrogen fertilizer contributing the largest share of explained variation, followed by irrigation, organic matter, plant density, and soil pH. The significant quadratic terms confirm the presence of curvature in the response surface, indicating that yield is maximized at intermediate levels of the factors rather than increasing indefinitely. Several interaction effects, particularly nitrogen $\times$ irrigation, nitrogen $\times$ plant density, and organic matter $\times$ nitrogen, are also significant, demonstrating that the effect of one agronomic factor depends on the level of another. The non-significant lack-of-fit test ($p = 0.48$) indicates that the quadratic model adequately represents the experimental data, while the high coefficients of determination ($R^2 = 97.54\%$, adjusted $R^2 = 95.41\%$, predicted $R^2 = 93.82\%$) demonstrate excellent explanatory and predictive performance.

The full quadratic Response Surface Model fitted was:

$$Y = 86.5 + 4.00X_1 + 5.10x_2 + 7.00X_3 + 4.60X_4 + 3.10X_5 - 1.45X_1^2 - 1.80X_2^2 - 2.05X_3^2 - 1.62X_4^2 - 1.28X_5^2 + 0.62X_1X_2 + 0.43X_1X_3 + 0.21X_1X_4 + 0.18X_1X_5 + 0.74X_2X_3 + 0.55X_2X_4 + 0.36X_2X_5 + 0.81X_3X_4 + 0.67X_3X_5 + 0.42X_4X_5 \quad (9)$$

The response surface model indicates that sugarcane yield is governed by both the individual and combined effects of soil pH, organic matter, nitrogen fertilizer, irrigation level, and plant density. The positive linear coefficients show that increasing these factors enhances yield within the experimental range, while the negative quadratic coefficients confirm the existence of optimum operating conditions. Furthermore, the positive interaction effects highlight the importance of integrated agronomic management, particularly the combined optimization of nitrogen fertilizer and irrigation, to maximize sugarcane productivity. Consequently, the model provides a robust predictive tool for identifying the combination of agronomic practices that maximize yield while avoiding excessive input application, thereby supporting sustainable and climate-resilient sugarcane production.

4.5. Model Performance Comparison

The two models were compared on the basis of R^2 , Adjusted R^2 , RMSE, MAE and AIC as shown in **Table 7**.

Table 7 shows a comparison of model performance between the Multiple Linear Regression model and the Response Surface Model. The quadratic response

surface model (RSM) provides a superior fit to the data compared with the Multiple Linear Regression (MLR) model. The RSM achieved a higher coefficient of determination ($R^2 = 0.975$ and adjusted $R^2 = 0.954$), indicating that it explains a larger proportion of the variability in yield than the MLR model with $R^2 = 0.922$, Adjusted $R^2 = 0.912$. In addition, the prediction error measures were lower for the response surface model, with an RMSE of 1.26 and MAE of 0.98 compared to RMSE of 4.26 and MAE of 3.80 for the linear model, demonstrating improved predictive accuracy.

Table 7. Comparison of predictive performance of Regression and RSM models.

Model	R^2	Adjusted R^2	RMSE	MAE	AIC
Multiple Linear Regression	0.922	0.912	4.26	3.80	62.00
Quadratic Response Surface Model	0.975	0.954	1.26	0.98	29.7

The Akaike Information Criterion (AIC) value for the RSM (29.7) was lower than that of the multiple linear regression model (62.00), suggesting a better balance between goodness-of-fit and model complexity. A similar study comparing the performance of the two approaches to sugarcane yield prediction, 5 predictor variables (fertilizer rate, irrigation level, plant density, temperature and seasonal rainfall amount) were considered. A CCD with 32 factorial points, 10 axial points and 10 centre points was used and gave the performance summary in **Table 8**.

Table 8. Comparative performance of models on five factors.

Model	R^2	Adjusted R^2	RMSE	MAE	AIC
Multiple Linear Regression	0.237	0.154	10.0	7.98	131.73
Quadratic Response Surface Model	0.929	0.884	3.7	2.96	80.20

Overall, these results indicate that the quadratic response surface model provides a more accurate and reliable representation of the relationship between the agronomic factors and yield than the multiple linear regression model. This confirms that the response surface model outperformed the conventional regression model across all evaluation metrics.

5. Discussions

Conventional regression models, particularly multiple linear regression (MLR), were unable to capture the nonlinear nature of the response surface, yielding prediction errors that were higher than those from the response surface methodology (RSM). The superior predictive performance of RSM stems from its ability to model both quadratic effects and interactions among agronomic variables, providing a more realistic representation of complex crop production systems [9] [10]. For instance, the significant interaction between nitrogen fertilizer rate and irrigation

shows that crop response to nitrogen is strongly influenced by water availability, reflecting the synergistic relationship between nutrient uptake and soil moisture. Nitrogen availability and uptake are highly dependent on adequate soil moisture. Irrigation facilitates the movement of soluble nitrogen through the soil toward the root zone and supports root activity and nutrient absorption. Under water stress, root growth and physiological processes are restricted, reducing the crop's ability to utilize applied nitrogen. Conversely, excessive irrigation can increase nitrogen losses through leaching or denitrification, particularly in relatively permeable soils. Thus, an appropriate combination of nitrogen and irrigation produces a synergistic effect on sugarcane growth.

Likewise, the significant quadratic effects of nitrogen application and soil pH indicate optimal thresholds beyond which additional inputs produce diminishing marginal returns or even adverse effects on crop productivity. Soil pH influences the chemical availability of nitrogen, phosphorus and several micronutrients, as well as microbial activity and root development. At an appropriate pH, nutrients remain sufficiently available for effective root uptake. When pH becomes too low, nutrient imbalances and possible aluminium or manganese toxicity can restrict root growth; when pH becomes excessively high, the availability of some micronutrients and phosphorus may decline. This provides a biological explanation for the quadratic effect of pH, where yield increases toward an optimum and subsequently declines. Organic matter acts as an important reservoir of nutrients and improves soil structure, water-holding capacity and microbial activity. In the warm Chiredzi environment, decomposition and mineralisation of organic matter can release nitrogen into forms available to sugarcane. Consequently, nitrogen fertilizer may be used more efficiently where organic matter improves the soil's capacity to retain nutrients and moisture. However, excessive nitrogen can stimulate vegetative growth without proportional increases in stalk yield and may delay crop maturity or reduce nitrogen-use efficiency. Soils with higher organic matter can make irrigation more effective by maintaining a more stable moisture environment around the roots. Increasing planting density increases the number of plants competing for nitrogen, water, light, and other resources. At moderate densities, greater canopy development and radiation interception can increase biomass and yield. However, excessive density intensifies competition among stalks, particularly when nitrogen or water is limiting. This can reduce individual stalk development and ultimately cause yield to plateau or decline.

Nonlinear responses are well established in agronomic research, where crop productivity is influenced by interacting physiological processes and constraints related to resource availability, including nitrogen and water, as well as environmental conditions [16] [17]. Collectively, these findings demonstrate that RSM provides a more robust analytical framework for modelling, predicting, and optimizing sugarcane yield under varying agronomic and environmental conditions, thereby supporting the development of climate-resilient and resource-efficient management strategies. The superior predictive performance of RSM ob-

served in this study is consistent with empirical evidence showing that sugarcane responses to agronomic inputs are frequently nonlinear and dependent on interactions among management and environmental factors. For example, a field study evaluating irrigation depth and nitrogen fertilization reported a significant quadratic response of sugarcane stalk yield to nitrogen and irrigation, with an estimated maximum yield of 120.7 t ha^{-1} at approximately 127 kg N ha^{-1} and 1267 mm of irrigation, demonstrating that the effect of nitrogen cannot be adequately represented by a simple linear relationship [18]. Similarly, research on nitrogen utilization in irrigated sugarcane found quadratic yield responses, with maximum stalk yields occurring at approximately $113 - 167 \text{ kg N ha}^{-1}$ depending on the fertilizer source, providing empirical evidence of diminishing returns beyond an agronomic optimum [19]. More recently, a global synthesis of 967 observations from 81 studies demonstrated that sugarcane nitrogen responses vary substantially with climatic and soil conditions, with temperature and soil nitrogen emerging as important determinants of nitrogen responsiveness [20]. In addition, long-term field research has shown significant interactions between nitrogen application and sugarcane cultivar for cane and sugar yield, further demonstrating that agronomic responses depend on combinations of management and biological factors [21]. Collectively, these findings support the advantage of RSM over a conventional linear model because its second-order structure can simultaneously capture curvature, diminishing marginal returns and interactions among agronomic variables, thereby providing a more realistic representation of the complex response behaviour observed in sugarcane production systems [9] [10].

6. Implications for Sugarcane Yield Optimization

The findings of this study demonstrate that Response Surface Methodology (RSM) provides a more robust analytical framework for modelling and optimizing sugarcane yield than conventional regression approaches. By explicitly accounting for nonlinear responses and interaction effects among agronomic factors, RSM improves predictive accuracy and facilitates the identification of optimal management combinations under complex production environments. These results have important implications for sustainable sugarcane production, indicating that the coordinated optimization of soil fertility management, irrigation scheduling, and planting density can substantially improve productivity and enhance resource efficiency. Such integrated management strategies are increasingly recognized as essential for improving crop performance and resilience under variable climatic and production conditions [22] [23]. Consequently, the developed response surface model provides a practical decision-support tool that can assist farmers, extension personnel, and agricultural planners in making evidence-based management decisions to maximize sugarcane yield under irrigated production systems while promoting climate-smart and resource-efficient agricultural practices.

7. Conclusions

This study demonstrated that Response Surface Methodology outperforms conventional regression models in predicting sugarcane yield. RSM achieved higher explanatory power and lower prediction errors since it was able to model quadratic and interaction effects. While conventional regression models remain useful for exploratory analysis and simpler systems, RSM is recommended for agronomic optimization where multiple interacting factors influence yield. Future research should compare RSM with machine learning approaches such as Random Forest and Artificial Neural Networks to further enhance predictive accuracy.

This study demonstrated that Response Surface Methodology (RSM) provides a more effective framework than conventional Multiple Linear Regression (MLR) for predicting and optimizing sugarcane yield under varying agronomic conditions. The comparative analysis showed that the quadratic RSM model achieved superior predictive performance across all major model evaluation criteria. Specifically, RSM achieved an R^2 of 0.975 and an adjusted R^2 of 0.954, compared with 0.922 and 0.912, respectively, for the MLR model. These results indicate that the quadratic RSM model explained a larger proportion of the variability in sugarcane yield while retaining strong explanatory power after accounting for model complexity. The improvement in predictive accuracy was further demonstrated by the error metrics. The RSM model produced a substantially lower root mean square error (RMSE) of \$1.26, compared with \$4.26, for MLR, representing a considerable reduction in prediction error. Similarly, the mean absolute error (MAE) decreased from \$3.80, for MLR to \$0.98, for RSM. The lower RMSE and MAE demonstrate that the RSM predictions were, on average, considerably closer to the observed sugarcane yields. The difference between the RMSE and MAE also indicates that the quadratic RSM model maintained a relatively consistent prediction error structure, with limited influence from large individual errors.

The predictive performance of the models was also supported by the predicted coefficient of determination. The quadratic RSM model achieved a predicted R^2 of approximately 0.938, indicating that the model retained strong predictive capability for observations not directly used in estimating the model. This provides additional evidence that the improvement in model fit was not solely attributable to the inclusion of additional terms. Rather, the second-order structure of RSM provided a more appropriate representation of the underlying response relationship between sugarcane yield and the agronomic factors considered in the study.

Model selection based on the Akaike Information Criterion (AIC) further supported the superiority of RSM. The RSM model recorded an AIC of 29.7, substantially lower than the AIC of 62.00 obtained for the MLR model. Since lower AIC values indicate a more favourable balance between goodness-of-fit and model complexity, the substantially lower AIC of RSM suggests that the quadratic model achieved improved explanatory and predictive performance without an excessive penalty associated with its additional parameters. Taken together, the higher, adjusted R^2 , and predicted R^2 , along with the lower RMSE, MAE, and AIC, provide

consistent evidence that RSM outperformed conventional MLR in this study. The superior performance of RSM can be attributed to its ability to represent the non-linear nature of sugarcane responses to agronomic inputs. Unlike the conventional linear model, the quadratic RSM model incorporates both squared terms and two-way interaction terms, allowing it to capture curvature, diminishing returns, and synergistic or antagonistic relationships among soil pH, organic matter, nitrogen, irrigation, and plant density. This flexibility is particularly important in agricultural production systems, where the response of crop yield to an individual input may depend on the levels of other inputs and environmental conditions. The significant quadratic and interaction effects identified in the analysis therefore provide an important explanation for the improved predictive performance of RSM.

Although conventional regression remains valuable for exploratory analysis, parameter interpretation, hypothesis testing, and situations involving relatively simple linear relationships, its predictive capability may be limited when crop responses exhibit substantial nonlinearities and interactions. In contrast, RSM provides both predictive and optimization capabilities, making it particularly suitable for identifying combinations of agronomic inputs that can maximize sugarcane yield while improving the efficiency of resource use. The findings therefore support the application of quadratic RSM as an appropriate statistical approach for decision-making in sugarcane production, particularly where producers need to simultaneously consider fertilizer application, irrigation, soil properties, and plant density.

From an agricultural management perspective, the findings suggest that predictive modelling can contribute to more efficient and sustainable allocation of production resources. By identifying the combined effects and optimal levels of agronomic inputs, RSM can support farmers and agricultural extension practitioners in developing management strategies that increase yield while reducing unnecessary application of costly inputs such as nitrogen fertilizer and irrigation water. This is particularly relevant in regions where water availability, soil fertility, and climatic conditions vary considerably across production environments. Future research should extend the comparative modelling framework by evaluating RSM alongside machine-learning approaches such as Random Forest, Artificial Neural Networks, Support Vector Regression, and Gradient Boosting. Such comparisons would help determine whether data-driven machine-learning methods can provide further improvements in predictive accuracy while maintaining sufficient interpretability for agricultural decision-making. Future studies should also incorporate larger multi-season datasets containing additional environmental variables, including temperature, rainfall, evapotranspiration, soil moisture, and other soil characteristics. Validation using independent datasets from different sugarcane production locations would further establish the generalizability and robustness of the developed model.

Overall, the study concludes that the quadratic RSM model provides a statisti-

cally robust, predictive, and practically useful alternative to conventional MLR for modelling sugarcane yield. Its superior performance across explanatory, predictive, error-based, and information-theoretic metrics demonstrates its capacity to capture the complex relationships between agronomic inputs and crop yield. Consequently, RSM offers a strong foundation for developing decision-support tools aimed at optimizing sugarcane production, improving resource-use efficiency, and strengthening the resilience of sugarcane farming under changing environmental and production conditions.

Author Contributions

Edwin Rupi generated the study, collected and analysed the data, developed the statistical models, interpreted the results and prepared the original manuscript. Philimon Nyamugure, Precious Mdlongwa, Peter Chimwanda and Thomas Musora provided the supervision, methodological guidance, interpretation of results and critical revision of the manuscript. All authors read and approved the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] de França e Silva, N.R., Chaves, M.E.D., Luciano, A.C.D.S., Sanches, I.D., de Almeida, C.M. and Adami, M. (2024) Sugarcane Yield Estimation Using Satellite Remote Sensing Data in Empirical or Mechanistic Modeling: A Systematic Review. *Remote Sensing*, **16**, Article 863. <https://doi.org/10.3390/rs16050863>
- [2] Ansarifar, J., Wang, L. and Archontoulis, S.V. (2021) An Interaction Regression Model for Crop Yield Prediction. *Scientific Reports*, **11**, Article No. 17754. <https://doi.org/10.1038/s41598-021-97221-7>
- [3] Khuri, A.I. and Mukhopadhyay, S. (2010) Response Surface Methodology. *WIREs Computational Statistics*, **2**, 128-149. <https://doi.org/10.1002/wics.73>
- [4] Akinyemi, O., Faweya, O., Ogundele, F.B. and Ilesanmi, A.O. (2025) Response Surface Methodology for Optimization and Modeling of Cassava Yield. *International Journal of Research and Innovation in Applied Science*, **X**, 763-769. <https://doi.org/10.51584/ijrias.2025.10040063>
- [5] Sharma, A. and Saraf, C. (2025) Development of a Machine Learning-Based Prediction Model for Sugarcane Yield: A Study with Special Reference to Meerut, Uttar Pradesh, India. *Journal of Information Systems Engineering and Management*, **10**, 45-54. <https://doi.org/10.52783/jisem.v10i34s.5774>
- [6] Satpathi, A., Chand, N., Setiya, P., Ranjan, R., Nain, A.S., Vishwakarma, D.K., *et al.* (2025) Evaluating Statistical and Machine Learning Techniques for Sugarcane Yield Forecasting in the Tarai Region of North India. *Computers and Electronics in Agriculture*, **229**, Article ID: 109667. <https://doi.org/10.1016/j.compag.2024.109667>
- [7] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021) An Introduction to Statistical Learning: With Applications in R. 2nd Edition, Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- [8] Hastie, T., Tibshirani, R. and Friedman, J. (2021) The Elements of Statistical Learn-

ing: Data Mining, Inference, and Prediction. 2nd Edition, Springer.

- [9] Montgomery, D.C. (2021) Design and Analysis of Experiments. 10th Edition, John Wiley & Sons.
- [10] Myers, R.H., Montgomery, D.C. and Anderson-Cook, C.M. (2023) Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 5th Edition, John Wiley & Sons.
- [11] Burnham, K.P. and Anderson, D.R. (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, **33**, 261-304. <https://doi.org/10.1177/0049124104268644>
- [12] Myers, R.H., Montgomery, D.C. and Anderson-Cook, C.M. (2016) Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 4th Edition, Wiley.
- [13] Baş, D. and Boyacı, İ.H. (2007) Modeling and Optimization I: Usability of Response Surface Methodology. *Journal of Food Engineering*, **78**, 836-845. <https://doi.org/10.1016/j.jfoodeng.2005.11.024>
- [14] Bezerra, M.A., Santelli, R.E., Oliveira, E.P., Villar, L.S. and Escaleira, L.A. (2008) Response Surface Methodology (RSM) as a Tool for Optimization in Analytical Chemistry. *Talanta*, **76**, 965-977. <https://doi.org/10.1016/j.talanta.2008.05.019>
- [15] Emran, M., Gamal, E.H., Haddad, A.M. and Ibrahim, O.M. (2024) Response Surface Methodology for Optimizing Water and Fertilizer Requirements for Maize (*Zea mays* L.) Growth in Sand Soil. *Journal of Soil Science and Plant Nutrition*, **24**, 6349-6364. <https://doi.org/10.1007/s42729-024-01973-w>
- [16] Lobell, D.B. and Field, C.B. (2007) Global Scale Climate-Crop Yield Relationships and the Impacts of Recent Warming. *Environmental Research Letters*, **2**, Article ID: 014002. <https://doi.org/10.1088/1748-9326/2/1/014002>
- [17] Sinclair, T.R., and Rufty, T.W. (2022) Bringing Scepticism to Crop Science. Springer. <https://doi.org/10.1007/978-3-031-14414-1>
- [18] Mendoça, M.F.D., Dantas Neto, J., Silva, P.F.D. and Oliveira, E.C.A.D. (2020) Yield and Technological Quality of Sugarcane under Irrigation Depths and Nitrogen Fertilization. *Revista Brasileira de Engenharia Agrícola e Ambiental*, **24**, 482-489. <https://doi.org/10.1590/1807-1929/agriambi.v24n7p482-489>
- [19] Furtado da Silva, N., Cabral da Silva, E., Muraoka, T., Batista Teixeira, M., Antonio Loureiro Soares, F., Nobre Cunha, F., *et al.* (2020) Nitrogen Utilization from Ammonium Nitrate and Urea Fertilizer by Irrigated Sugarcane in Brazilian Cerrado Oxisol. *Agriculture*, **10**, Article 323. <https://doi.org/10.3390/agriculture10080323>
- [20] Yuan, X., Zhou, S., Wei, F., Ma, L., Sun, Y., Peng, D., *et al.* (2025) Global Synthesis of Nitrogen Management in Sugarcane Systems: Decoding Climate-Soil-Management Drivers of Brazil-China Contrasts. *Field Crops Research*, **333**, Article ID: 110075. <https://doi.org/10.1016/j.fcr.2025.110075>
- [21] Zeng, X.P., Zhu, K., Lu, J.M., Jiang, Y., Yang, L.T., Xing, Y.X. and Li, Y.R. (2020) Long-Term Effects of Different Nitrogen Levels on Growth, Yield, and Quality in Sugarcane. *Agronomy*, **10**, Article 353. <https://doi.org/10.3390/agronomy10030353>
- [22] Food and Agriculture Organization of the United Nations (2022) The State of the World's Land and Water Resources for Food and Agriculture: Systems at Breaking Point (SOLAW 2021). FAO. <https://doi.org/10.4060/cb9910en>
- [23] Peng, X., Chen, D., Zhou, Z., Zhen, J., Xu, C., Hu, X. and Wang, Y. (2024) Optimizing Drip Fertigation Management to Simultaneously Improve the Yield, Water Produc-

tivity and Fertilizer Agronomic Utilization Efficiency of Grapes in Different Precipitation Year Patterns. *Agricultural Water Management*, **295**, Article 108749.
<https://doi.org/10.1016/j.agwat.2024.108749>