A Comparative Study of Detrending Methods on Crop Yield Time Series for Drought Studies

ABSTRACT

Detrending is a statistical technique that removes systematic variations or trends from time series data, allowing analysts to focus on the underlying patterns or fluctuations. While multiple detrending approaches have been applied but rarely discussed their consistency of outcomes and effectiveness in accurately capturing better yield trends. The validation of drought occurrences has proven to be a challenging task due to the non-stationary characteristics of time series data related to crop yield. This research utilizes time series of cotton yield data from the Marathwada region covering the period from 1998 to 2021. Three traditional trend models, including simple linear regression, second-order polynomial regression and central movingaverage were applied. Additionally, two machine learning models (random forest and support vector regression) were tested with a novel approach. Moreover, two decomposition models (additive and multiplicative) were used to remove non-linear trends in crop yield time series data. The performance of the chosen models was evaluated based on metrics such as root mean square error, mean absolute error, Nash-Sutcliffe efficiency, and index of agreement. The results suggest that the most effective detrending approach involves combining a random forest machine learning model with an additive decomposition model.

Keywords: Drought, Machine learning, Detrending, Decomposition model

1. INTRODUCTION

Drought is one of the most pervasive and impactful natural phenomena, exerting profound effects on agriculture, ecosystems, economies and societies worldwide [1]. Drought can bedefined as an extended period of abnormally low precipitation leading to water scarcity, droughts can manifest in various forms, from meteorological droughts characterized by reduced rainfall to hydrological droughts affecting water supply and agricultural droughts impacting crop growth and yield [2,3].

India's agricultural landscape is one of the world's most agrarian economies, with a substantial portion of its population reliant on agriculture for livelihood and sustenance. India's vulnerability to drought underscores the profound implications of water scarcity on agricultural productivity, rural livelihoods and national development. In the past, droughts have had a notable economic impact on the agricultural industry in India [4,5]. Prodhan et al. [6] have warned of an increased risk of yield reduction under extremely dry weather conditions. While past studies have highlighted the overall agricultural losses during specific incidents, challenges related to advancing technology and other non-climatic factors hinder the differentiation of these losses across various events impacting crop yields.

In the context of drought validation, detrending crop yield data serves multiple purposes. Firstly, it aids in establishing baseline conditions and historical trends in crop productivity, allowing for the identification of deviations from expected levels due to drought or other environmental stressors. Secondly, detrending facilitates the calibration and validation of drought indices and models by providing a standardized framework for

assessing model performance against observed data.Moreover, detrending crop yield data enhances the interpretability and reliability of drought monitoring and early warning systems, enabling stakeholders to differentiate between short-term fluctuations in yield attributable to weather variability and longer-term trends driven by factors such as technological advancements, land use changes, and market dynamics[7,8].

Previously, the detrending of crop yield was achieved by employing a predetermined function, such as a simple linear regression model (SLR) [9], polynomial regression model (PLR)[10] and locally weighted regression [11]. Finger [12] utilized a robust regression technique for detrending crop yield data. Tao et al. [16] conducted trials on simulated yield trends and observed that detrending techniques relying on smoothing methods displayed a general advantage over linear, loglinear, and autoregressive integrated moving average models (ARIMA). The most notable outcomes were obtained from the moving average and robust locally weighted regression models. Choudhary et al. [17] utilized piecewise regression to detrend crop yield for agricultural insurance and discovered it to be superior to simple linear regression. Lu et al. [18] employed six trend simulation models and two decomposition models for detrending crop yield, identifying the spatial and temporal impacts of drought on the United States. Irawan et al. [19] implemented locally weighted regression (LOWESS) to detrend crop yield for drought visualization in the West Java province of Indonesia.

Previously, detrending crop models using machine learning (ML) is not discussed. The strengths and limitations of these approaches are well understood, evident in their foundational algorithms and practical applications [20]. SLR, a widely used traditional technique, offers interpretability but lacks predictive power [21]. It struggles with the nonlinearity present in crop yield time series. While conventional regression methods struggle with these issues, ML models offer potential solutions. These models can handle complex functions and demonstrate better predictive accuracy [22].

The Marathwada region in India has been significantly impacted by drought events, leading to severe water scarcity and agricultural crises. Studies have highlighted the region's vulnerability to drought, with frequent and extreme drought events affecting the area [23,24]. The impact of drought in Marathwada has been profound, with reports of consecutive drought events from 2012 to 2016 [23]. The severity of the situation is evident from the fact that the region has witnessed a high number of farmer suicides, indicating the dire consequences of these drought events [25,26]. The Marathwada region of India faces significant challenges due to recurrent drought events, leading to water scarcity, agricultural crises, and socio-economic impacts. Research efforts have been directed toward understanding the dynamics of drought in the region, developing monitoring tools, and implementing mitigation measures to address the adverse effects of drought on the local population and environment.

This research compares conventional regression techniques with ML methods for detrending, evaluating their strengths and weaknesses in terms of applicability over space and time, reliability, and efficiency. Our study explores a method for self-adaptive detrending of data, able to automatically model long-term non-stationary and nonlinear yield trends influenced by technological progress, thus removing trends caused by extreme weather. The application of this method to detrend crop yield data and create standardized visual representations of drought is investigated.By synthesizing empirical evidence, case studies, and methodological frameworks, we aim to elucidate the importance of detrending in enhancing the accuracy and reliability of drought monitoring and prediction efforts, ultimately contributing to more effective drought management strategies and resilience-building initiatives in agricultural systems.

2. MATERIAL AND METHODS

2.1 Study Area and Data Collection

The state of Maharashtra is demarcated into six distinct divisions. The Marathwada region, also referred to as the Aurangabad division, encompasses a total of eight districts namely: Aurangabad, Beed, Hingoli,Jalna,Latur,Osmanabad and Parbhani. Positioned within the central plateau, Marathwada region spans from $17^0$37' North to $20^0$39' North latitude and $74^0$33' East to $78^0$22' East longitude, covering a geographical area of 64590sq. Km. This particular region is located in the rain shadow zone of the Sahyadri mountain range within the Western Ghats of Maharashtra. Characterized by a plateau topography with gentle undulations, Marathwada region is bordered by the Pune division to the south-west, Nashik division to the north-west, Amravati division to the Northeast, Telangana state to the south-east, and Karnataka state to the south. Additionally, the region includes extensions of the Ajantha and Balaghat hill ranges and is primarily drained by the prominent river Godavari. The Marathwada region experiences an average annual rainfall of 776 mm.

The time series data of crop area and yield for the years 2000 to 2021 were collected from Directorate of Economics and Statistics (DES), Ministry of Agriculture and Farmers Welfare, Government of India, (www.desagri.gov.in). This study selects the dominant crop grown in Marathwada region in the last two decades. We examined detrending techniques and showed how a drought year affects crop production during the selected year in Marathwada region.

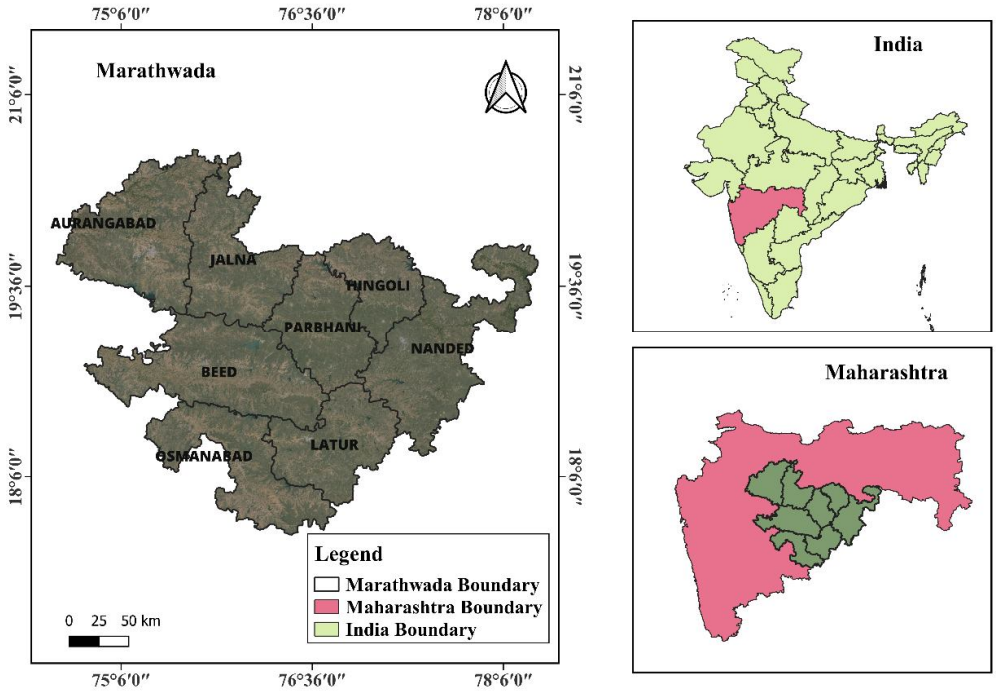


Fig.1 Location of Marathwada

2.2 Dominant crop selection

The major crop for the Marathwada region was selected based on the long-term average crop area during the Kharif season. we evaluated district-wisecrop-growing areas in the Kharif season. We found cotton is a major crop in Marathwada region (Fig. 2).

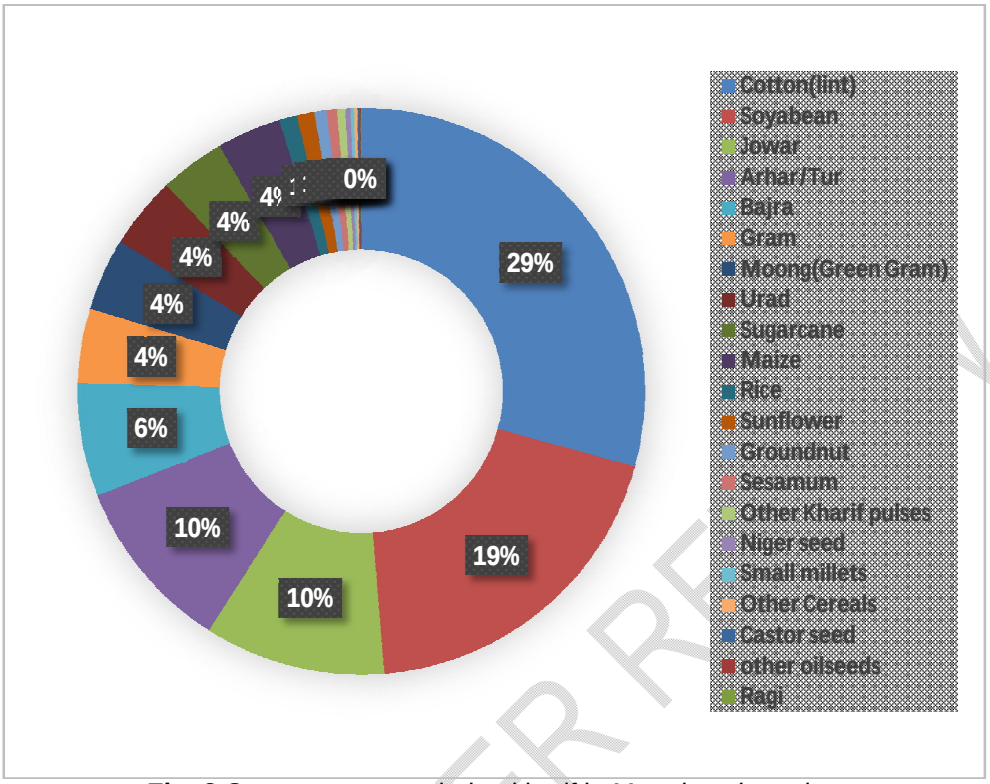


Fig. 2 Crop grown area during kharif in Marathwada region

2.3 Trend Simulation Methods

In this study, we apply three conventional regression techniques including simple linear regression, second-order polynomial model, Central Moving Average model (CMA) and two machine learning models namely Random Forest, and Support Vector Regressionto simulate the trend of cotton yield over time. After trend simulation, to detrend the data we applied two decomposition models (additive and multiplicative) and compared them. These techniques were tested individually for each district of Marathwada. All data analyses were conducted using R programming language with necessary packages.

2.3.1 Simple Linear Regression:

Simple linear regression is a statistical technique used to model the relationship between two variables, where one variable (the dependent variable) is predicted from the other variable (the independent variable) through a linear equation. The formula for simple linear regression is expressed as**:**

$$Y_t = \beta_0 + \beta_1 t \quad (1)$$

Where$Y_t$represents the dependent variable (cotton yield), t represents the independent variable (time), β_0 is the intercept term, $\beta_1$is the slope coefficient, representing the change in the dependent variable for a one-unit change in the independent variable.

2.3.2 Second-Order Polynomial Model:

A second-order polynomial model extends the simple linear regression model by allowing for a curvilinear relationship between the variables[27, 28]. It is expressed as:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad (2)$$

Where, Y_t, t, β_0 ,$\beta_1$andare defined as in simple linear regression. t^2 represents the squared term of the independent variable, allowing for curvature in the relationship.

2.3.3 Central Moving Average Model:

The Central Moving Average model is a simple smoothing technique used to identify trends in time series data. It calculates the average of a specified number of consecutive observations, centered around each data point. The application of the Central Moving Average technique to crop yield trend analysis holds significant implications for understanding the long-term dynamics of agricultural productivity, identifying cyclical patterns, and detecting anomalies indicative of environmental stressors or management interventions[18]. The formula for the Central Moving Average model is:

$$CMAY_t = \frac{1}{2m+1} \sum_{i=t-m}^{t+m} y_i \quad (3)$$

Where:$CMAY_t$represents the central moving average at time t, $y_i$represents the observed value at timei. m is the number of periods to include in the moving average calculation.

2.3.4 Random Forest:

Random Forest is a machine learning algorithm that utilizes an ensemble of decision trees to make predictions. It is particularly effective for regression tasks involving complex, non-linear relationships. The formula for Random Forest is not explicitly defined as it involves a collection of decision trees, each trained on a random subset of the data [29,30]. In this study, we aim to utilize RF as a tool for regression analysis. The term "decision tree" originates from the visual depiction of the algorithm's flow, mirroring tree structures. Initially, a set number of training sets are randomly chosen, leading to the creation of distinct trees. Each node employs a random selection of independent variables to divide the data into smaller units within each tree. These trees are expanded fully, and the predicted constant response value is calculated by averaging the fitted reactions from all trees. The success of the Random Forest algorithm in addressing overfitting trees contributes significantly to its efficacy as an ML technique [31,32].

The effectiveness of the RF algorithm largely depended on the number of decision trees used and the selection of potential attributes within a subset [33]. The relationship between the number of trees and the out-of-bag (OOB) error was analyzed to identify the optimal number of decision trees, resulting in the most favorable outcomes.

2.3.5 Support Vector Regression:

Support Vector Regression (SVR) is a machine learning algorithm known for its robustness and versatility. It works by mapping the input data into a higher-dimensional feature space and finding the hyperplane that best fits the data while maximizing the margin. The formula for SVR involves optimization techniques to find the optimal hyperplane, but it can be summarized as

$$Y = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b \quad (4)$$

Where Y represents the predicted value, x_i are the training data points,$\alpha_i$are the Lagrange multipliers, K is the kernel function and b is the bias term.This study investigates different kernel types such as linear, polynomial and radial basis functions and selects the optimal kernel based on minimal root mean square error. Grid search, a technique for hyperparameter optimization, is utilized to methodically explore a predefined set of hyperparameters and identify the combination that results in the best performance for the machine learning model.

2.4 Decomposition model

Upon employing suitable statistical models to simulate the trend, our next step involves utilizing decomposition models to eliminate this simulated trend and extract detrended data. Decomposition models play a pivotal role in data analysis by unveiling underlying patterns and trends, thereby furnishing valuable insights into intricate time series or observational data. Detrending stands out as a primary application of decomposition models, aiming to disentangle observed data from these underlying patterns or trends [34].

Two methods are available to accomplish this task: the additive and multiplicative models. The additive model derives detrended data by subtracting trend line values from the observed data. This method is suitable when variations in the data remain relatively consistent across different levels. Notably, the residuals maintain the same unit of measurement as the original data [18].

Conversely, the multiplicative decomposition approach involves deriving detrended data by calculating the ratio between the observed data and trend line values. These detrended data points lack dimensionality and illustrate the percentage discrepancies compared to the trend line values. This model is preferable when the magnitude of fluctuations in the data varies across different levels [18].

2.5Performance of trend simulation model

To assess the performance of the trend simulation models, we employ four commonly used metrics: Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (E), Mean Absolute Error (MAE), and Index of Agreement (IOA). These metrics provide valuable insights into the accuracy and reliability of the simulated trends compared to observed data.RMSE quantifies the average deviation between observed and simulated values. It is calculated as the square root of the mean squared differences between observed and simulatedvalues.Nash-Sutcliffe Efficiencymeasures the relative accuracy of the simulated values compared to the mean observed value[35]. MAE provides the average magnitude of errors between observed and simulated values. It is calculated as the mean of the absolute differences between observed and simulated values. IOA assesses the agreement between observed and simulated values, considering both the variability and bias of the data[36].These performance metrics provide comprehensive insights into the accuracy, precision, and overall agreement of the trend simulation models with observed data. By evaluating the RMSE, E, MAE and IOA, we can effectively assess the suitability of each model for capturing the underlying trends in cotton yield data and informing decision-making processes. The computational formula of evaluation metrics are as follows;

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

$$E = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (7)$$

$$IOA = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(|\hat{y}_i - \bar{y}_i| + |y_i - \bar{y}_i|)^2} \quad (8)$$

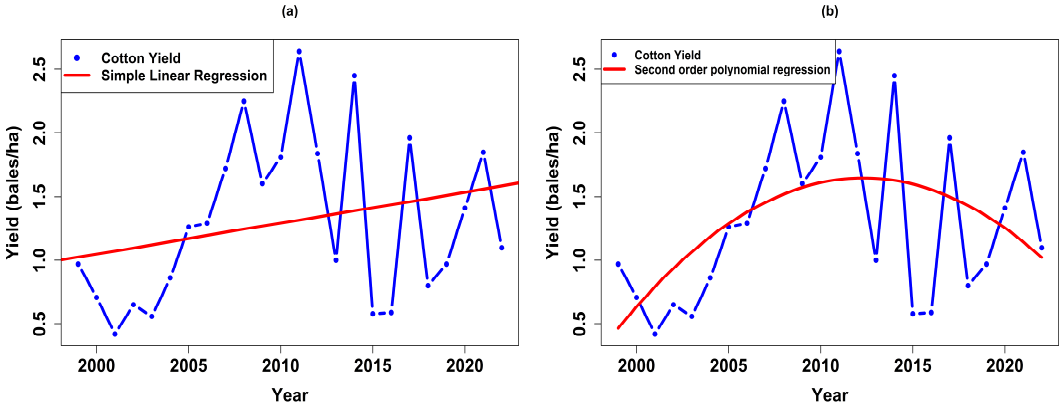
3. RESULTS AND DISCUSSION

The increase in cotton production observed can be linked to advancements in technology, improvements in management techniques and the collective influence of long-term climate changes. Removing the trend from the yield data is utilized as a strategy to address the temporal fluctuations in statistical analyses [37].

3.1 Trend simulation model comparison

Fig. 2 depicts the chronological progression of cotton yield in the Aurangabad district of Marathwada from 1998 to 2021 and presents the results of five models utilized for trend simulation.The cotton yield time series in each district showcases a non-linear upward trend that influences the long-term pattern of crop yield. This trend is mainly attributed to technological advancements and increased inputs.

The SLR model proves to be inadequate in explaining the fluctuations in cotton yield across the study areas (Fig. 3(a)) due to the non-linear nature of the technology trend, lacking coherence or rationale. While a quadratic trend improves the correlation, it still falls short of capturing the gradual rise evident in the trend (Fig. 3(b)). These predetermined models do not possess the necessary flexibility to remove non-linear and non-stationary trends consistently across all districts of Marathwada.



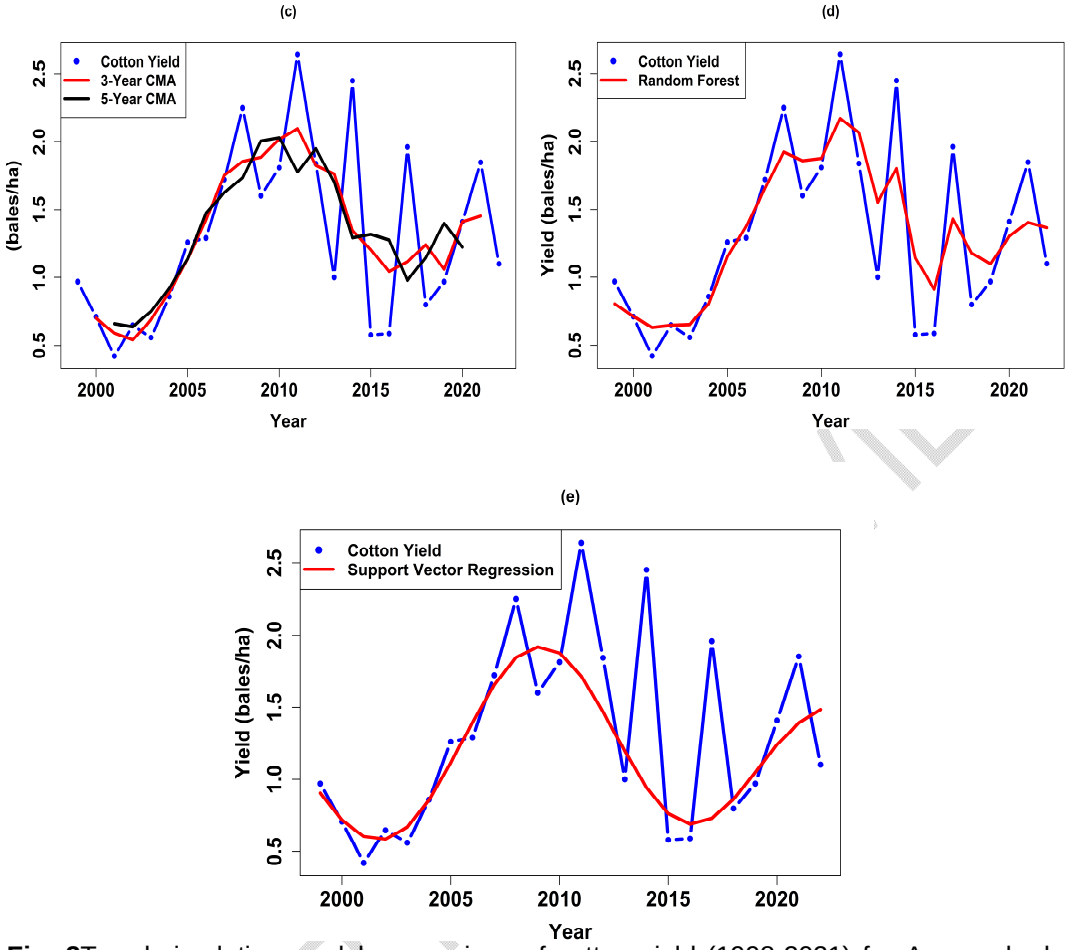


Fig. 3 Trend simulation model comparison of cotton yield (1998-2021) for Aurangabad: (a) simple linear regression model; (b) second-order polynomial model; (c) central moving average model of 3-year and 5-year; (d)) random forest model; (e) support vector regression model

Inspecting the visual representation of cotton yield reveals the need for implementing a CMA model to address the inconsistencies observed in the crop yield data (Fig. 3(c)). A specific time is required to perform the CMA model. Nevertheless, the process of determination and selection of time remains subjective. Moreover, the utilization of a CMA model presents a boundary issue. To implement a 3-year CMA model, it is essential to have 3-year data before and following the year of interest.

Consequently, when the data point shifts to the initial or latest years, the absence of adequate data for estimation leads to the identification of the first three and last three data points as missing data. Furthermore, the presence of a single missing data in the crop yield series can result in an additional 3 data points being designated as missing data for the CMA model. Even in the absence of missing data in the time series, the CMA model for three years remains loses two data points at the starting and ending positions in the time series. It's important to highlight that the effectiveness of the CMA model is compromised or biased near the border of time series data.

In the ML model, Random Forest performs best among all other models. We used out-of-bag error for an optimum number of decision trees. The trend curve predicted by the random forest model closely mirrors the time series of cotton yield and demonstrates strong performance for the study area (Fig. 3(d)). Support vector regression model performance is lesser than the random forest model. This study incorporates the most suitable kernel and user-defined hyperparameters C and ε (Table 1) for better accuracy of this model. The primary focus of this study is to explore regression models employing SVR and different kernel functions namely linear, radial basis functions and polynomial. Grid search optimization and k-fold cross-validation techniques are utilized to optimize these hyperparameters, aiming to minimize error estimates while reducing bias and variance in the dataset.

Table 1optimized hyperparameters used in support vector regression

Districts	Kernel	Cost function (C)	Epsilon (ε)
Aurangabad	Radial basis function	10	0.1
Beed	Radial basis function	1	0.1
Hingoli	Radial basis function	1	0.2
Jalna	Radial basis function	10	0.2
Latur	Radial basis function	1	0.5
Nanded	Radial basis function	1	0.5
Osmanabad	Radial basis function	1	0.5
Parbhani	Radial basis function	1	0.2

3.2 Evaluation metrics for model performance

Table 2 presents the metrics values utilized for evaluating the effectiveness of trend simulation models. The findings indicate that SLR models are the least precise in terms of fitting. Conversely, second-orderPLR models offer a superior fit compared to SLR models. The central moving average demonstrates enhanced performance when fitting cotton yield, outperforming SLR andPLR. In contrast to traditional regression methods, the ML model shows superior performance. Among all models, the Randomforest model achieves the highest accuracy of 78.95% forthe Latur district ofMarathwada.

Table 2 Trend simulation models evaluation results

District	Models	RMSE	MAE	IOA	E	District	Models	RMSE	MAE	IOA	E
Aurangabad	SLR	0.6063	0.5183	35.13	7.11	Latur	SLR	0.9387	0.7056	62.49	25.38
	PLR	0.5356	0.4363	65.56	27.50		PLR	0.8986	0.6944	68.75	31.61
	CMA3	0.4342	0.3141	83.45	55.55		CMA3	0.6029	0.3795	89.31	70.96
	CMA5	0.5260	0.4119	73.23	36.23		CMA5	1.0091	0.6568	65.97	21.07
	RF	0.2989	0.2406	92.18	77.42		RF	0.4878	0.3144	93.10	79.85
	SVR	0.4838	0.3008	78.80	40.85		SVR	0.9735	0.7931	44.91	19.74
Beed	SLR	0.4334	0.3385	34.73	6.15	Nanded	SLR	0.3314	0.2787	73.34	38.28
	PLR	0.4316	0.3287	35.43	6.96		PLR	0.3304	0.2802	73.71	38.67
	CMA3	0.2976	0.2177	82.26	57.97		CMA3	0.2651	0.2270	86.16	62.41
	CMA5	0.3683	0.2667	58.25	10.15		CMA5	0.3213	0.2738	75.17	41.29
	RF	0.2465	0.1702	88.32	69.65		RF	0.2004	0.1726	92.26	77.43
	SVR	0.3709	0.2489	62.43	31.26		SVR	0.3268	0.2728	73.17	40.01

Hingoli	SLR	0.6411	0.4386	47.62	13.69	Osmanabad	SLR	0.4947	0.4008	39.70	8.40
	PLR	0.6107	0.4047	57.64	21.69		PLR	0.4516	0.3894	60.81	23.67
	CMA3	0.4180	0.3308	86.59	64.55		CMA3	0.3303	0.2652	86.05	61.51
	CMA5	0.5760	0.4551	63.72	29.53		CMA5	0.3899	0.3220	78.49	46.78
	RF	0.3218	0.2349	92.04	78.25		RF	0.2498	0.2045	91.90	76.66
	SVR	0.5804	0.3928	57.81	29.26		SVR	0.4867	0.3913	54.02	11.35
Jalna	SLR	0.6536	0.5220	42.09	9.49	Parbhani	SLR	0.7643	0.4799	55.17	20.04
	PLR	0.6452	0.5205	46.08	11.79		PLR	0.7356	0.4030	62.15	25.94
	CMA3	0.5491	0.4179	73.20	39.90		CMA3	0.6473	0.3698	76.36	45.74
	CMA5	0.6943	0.5920	45.62	1.48		CMA5	0.8322	0.5226	52.35	13.70
	RF	0.4103	0.3255	85.08	64.33		RF	0.4598	0.2658	88.66	71.06
	SVR	0.5980	0.4331	71.25	24.23		SVR	0.7339	0.3997	59.54	26.27

3.3 Decomposition Models Comparison

Guttormsen [38] and Zhang and Qi [39] employed residuals derived from subtracting crops from the regression line to illustrate the deviation of crops from the norm, deeming these as detrended data. These studies operated on the assumption that fluctuations and trends were combined additively. Lu et al [18] proposed a multiplicative model for detrending long-term crop yield. By utilizing a multiplicative decomposition method to remove trends from the time series, the detrended cotton production in Marathwada displays an increasing variance over time (Fig. 4). As the variance of detrended data is standardized according to crop output scale, it progresses towards enhanced stationarity over time, favoring an additive decomposition model (Fig. 5). Hence, an additive decomposition model was applied to detrend cotton yield following the implementation of a suitable trend simulation technique.

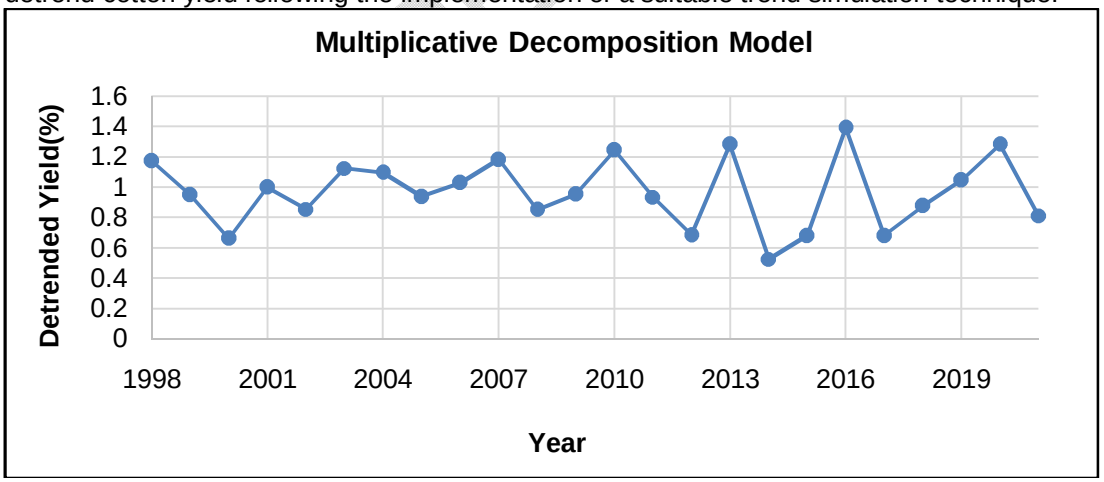


Fig. 4 Multiplicative decomposition model for Aurangabad (Random Forest Model)

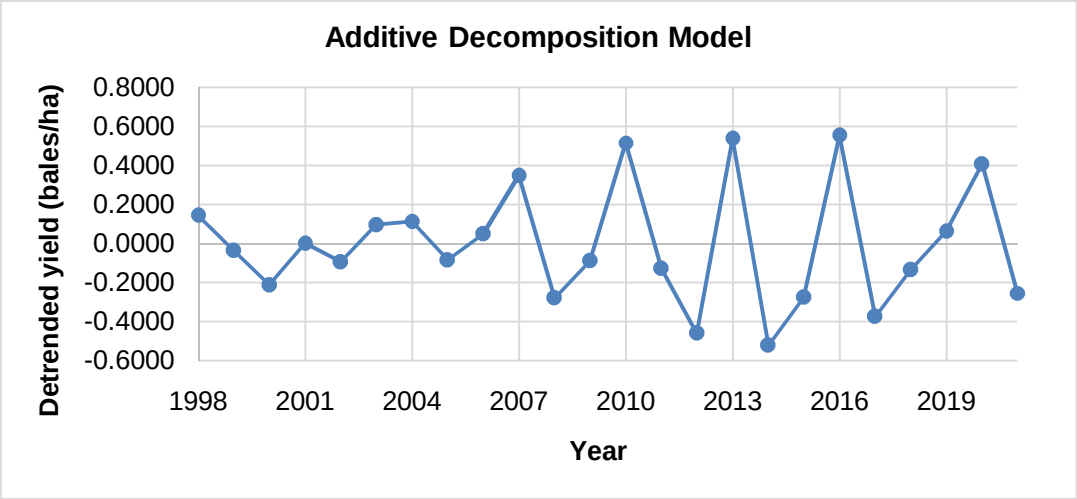


Fig. 5 Additive decomposition model for Aurangabad (Random Forest Model)

3.4 Selection of detrending model

In this research, the choice of detrending model is influenced by factors such as performance, reliabilityand efficiency. The analysis provided above emphasizes the unsatisfactory performance of SLR andPLR. Although the CMA model emerges as the most effective among traditional models, it is constrained by certain limitations and is notably impacted by missing data. On the other hand, two machine learning models including RF and SVRoutperform conventional detrending methods. These ML models excel in automatically identifying underlying patterns in nonstationary and non-linear time series of cotton yield, thereby providing a robust trend fitting for crop yields. These detrended crop yields are viewed as anomalies in crop yield, indicating values higher or lower than the average crop yield. Consequently, the integration of the random forest model with an additive decomposition technique is highlighted as the preferred detrending approach for cotton yield in the Marathwada region.

4. CONCLUSION

This study is focused on identifying the most appropriate detrending technique in cotton yield for the study area. The study evaluates various traditional and machine learning trend simulation models using four quantitative metrics. The findings suggest that SLR and PLRshow the weakest fit. In contrast, CMA surpasses the other conventional regression techniques but is limited by boundary issues. Among the machine learning models, the random forest model demonstrates superior accuracy. Furthermore, two decomposition models were assessed, revealing that the additive decomposition model is better suited for detrending crop yield. This choice is based on adjusting the variance of detrended crop yield according to the crop yield magnitude scale, leading to increased stationarity over time. Thus, the recommended approach for detrending cotton crop yield involves utilizing the random forest model alongside the additive decomposition model.

The methodology applied in this study entails standardizing cotton yield to enable a quantitative evaluation of its correlation with drought. This method also aids in visualizing the impact of drought on cotton yield. Further correlation analyses comparing various drought indicators and cotton yield anomalies can be conducted to validate drought occurrences. The study observed a relation between cotton yield anomalies and drought years, evident from significant drought years (2012, 2015 and 2018) illustrated in Fig. 5. By examining

cotton yield anomalies in Marathwada, variations in cotton yield below normal levels due to drought conditions were captured.

This detrending methodology is not exclusive to cotton and drought studies but can be applied to different crops and natural disasters. Our study offers a framework for assessing the impact of drought on crop production, aiding policymakers and stakeholders in developing robust strategies for risk management and mitigation of extreme weather effects on agriculture. This study can be enhanced by incorporating more machine learning and deep learning algorithms and used to make short-term predictions on the impact of technological advancements on crop yields. Additionally, the crop yield anomalies obtained through this method can be utilized in agricultural research focusing on climate change impacts.

CONSENT

Not applicable

ETHICAL APPROVAL

This article does not include any research with human or animal participants.

REFERENCES

1. Tate EL, Gustard A. Drought Definition: A Hydrological Perspective. In: Vogt, J V, Somma, F (Eds) Drought and Drought Mitigation in Europe. Advances in Nat Technol Hazards Res. 2000: 14.
2. Shi Y, Zhao L, Zhao X, Lan H, Teng H. The Integrated Impact of Drought on Crop Yield and Farmers' Livelihood in Semi-Arid Rural Areas in China. Land. 2022; 11:2260. https://doi.org/10.3390/land11122260.
3. Orimoloye IR. Agricultural Drought and Its Potential Impacts: Enabling Decision-Support for Food Security in Vulnerable Regions. Front Sustain Food Syst. 2022; 6. https://doi.org/10.3389/fsufs.2022.838824.
4. Gautam RC, Bana RS. Drought in India: Its impact and mitigation strategies – A review. Indian Agron. 2014;59(2):179 -190.
5. Chuphal DS, Kushwaha AP, Aadhar S, Mishra V. Drought Atlas of India, 1901–2020. Sci Data. 2024;11:7. https://doi.org/10.1038/s41597-023-02856-y.
6. Prodhan FA, Zhang JH, Sharma TPP, Nanzad L, Zhang D, Seka AM, Ahmed N, Hasan SS, Hoque MZ, Mohana HP. Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach. Sci Total Environ. 2022; 807: 151029.
7. Waseem M, Jaffry AH, Azam M, Ahmad I, Abbas A, Lee JE. Spatiotemporal Analysis of Drought and Agriculture Standardized Residual Yield Series Nexuses across Punjab, Pakistan. Water. 2022; 14(3): 496. https://doi.org/10.3390/w14030496.
8. Panwar V, Sen S. Economic Impact of Natural Disasters: An Empirical Re-examination. Margin: J Appl Econ Res. 2019; 13(1): 109-139.https://doi.org/10.1177/0973801018800087
9. Johari R, Li H, Liskovich I, Weintraub GY. Experimental design in two-sided platforms: An analysis of bias. Management Science. 2022.
10. Liu Y, Wang Y, Du Y, Zhao M, Peng J. The application of polynomial analyses to detect global vegetation dynamics during 1982–2012. Int J Remote Sens. 2016; 37(7): 1568-1584.
11. Hendrawan VSA, Kim W, Touge Y, Ke S, Komori D. A global-scale relationship between crop yield anomaly and multiscalar drought index based on multiple precipitation data. Environ Res Lett. 2022; 17:014037. 10.1088/1748-9326/ac45b4.
12. Finger R. Revisiting the Evaluation of Robust Regression Techniques for Crop Yield Data Detrending. Am J Agric Econ. 2010; 92: 205-211. https://doi.org/10.1093/ajae/aap021.

16. Tao Y, Jianliang N, Peijun S, Zhu W. Performance of detrending models of crop yield risk assessment: evaluation on real and hypothetical yield data. Stoch Environ Res Risk Assess. 2015; 29. https://doi.org/10.1007/s00477-014-0871-x.
17. Choudhury AH, Jones JR, Choudhury R, Spaulding AD. Association of rainfall and detrended crop yield based on piecewise regression for agricultural insurance. J Econ Econ Educ Res. 2015; 16(2): 31-43.
18. Lu J, Carbone G, Gao P. Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. Agric For Meteorol. 2017;237-238:196-208. https://doi.org/10.1016/j.agrformet.2017.02.001
19. Irawan ANR, Komori D, Hendrawan VSA. Correlation analysis of agricultural drought risk on wet farming crop and meteorological drought index in the tropical-humid region. Theor Appl Climatol. 2023;153: 227–240.https://doi.org/10.1007/s00704-023-04461-w
20. Sidhu BS, Mehrabi Z, Ramankutty N, Kandlikar M. How can machine learning help in understanding the impact of climate change on crop yields? Environ Res Lett. 2023; 18(2): 024008.
21. Leng G, Adan RA, Belot M, Brunstrom JM, de Graaf K, Dickson SL, Hare T, Maier S, Menzies J, Preissl H. The determinants of food choice. Proc Nutr Soc .2017; 76: 316–327.
22. Hu T, Zhang X, Bohrer G, Liu Y, Zhou Y, Martin J, Li Y, Zhao K. Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. Agric For Meteorol. 2023; 336:109458. https://doi.org/10.1016/j.agrformet.2023.109458
23. Khetwani S, Singh R. Drought vulnerability of Marathwada region, India: A spatial analysis. GeoScape. 2020; 14:108-121.https://doi.org/10.2478/geosc-2020-0010.
24. Dhawale R, Paul SK. A comparative analysis of drought indices on vegetation through remote sensing for Latur region of India, Int Arch Photogramm Remote Sens Spatial Inf Sci, XLII-5. 2018: 403–407.https://doi.org/10.5194/isprs-archives-XLII-5-403-2018
25. Aher M, Yadav S. Assessment of rainfall trend and variability of semi-arid regions of Upper and Middle Godavari basin, India. J Water Clim Change. 2021; 12. https://doi.org/10.2166/wcc.2021.044.
26. Kulkarni A, Gadgil S, Patwardhan S. Assessment of rainfall trend and variability of semi-arid regions of Upper and Middle Godavari basin, India. Current Science. 2016; 111 (7): 1182–1193.
27. Bahrami M, Mahmoudi MR. Long-term temporal trend analysis of climatic parameters using polynomial regression analysis over the Fasa Plain, southern Iran. Meteorol Atmos Phys. 2022; 134(2): 42.
28. Huang H, Wang Z, Xia F, Shang X, Liu Y, Zhang M, Mei K. Water quality trend and change-point analyses using integration of locally weighted polynomial regression and segmented regression. Environ Sci Pollut Res. 2017; 24:15827-15837.
29. Dang C, Liu Y, Yue H, Qian J, Zhu R. Autumn crop yield prediction using data-driven approaches: support vector machines, random forest, and deep neural network methods. Can J Remote Sens. 2021; 47(2): 162-181.
30. Jhajharia K, Mathur P, Jain S, Nijhawan S. Crop yield prediction using machine learning and deep learning techniques. Procedia Comput Sci. 2023; 218: 406-417.
31. Bhatnagar R, Gohain GB. Crop yield estimation using decision trees and random forest machine learning algorithms on data from Terra (EOS AM-1) & Aqua (EOS PM-1) satellite data. Stud ComputIntell. 2020; 835. https://doi.org/10.1007/978-3-030-20212-5_6
32. Everingham Y, Sexton J, Skocaj D, Bamber G I. Accurate prediction of sugarcane yield using a random forest algorithm. Agron Sustain Dev. 2016; 36:(2). https://doi.org/10.1007/s13593-016-0364-z.
33. Ramosaj B, Pauly M. Consistent estimation of residual variance with random forest Out-Of-Bag errors. Stat Probab Lett. 2019;151: 49-57.
34. Hikmah N, Kartikasari MD. Decomposition Method with Application of Grey Model GM (1, 1) for Forecasting Seasonal Time Series. Pak J Stat Oper Res. 2022; 18(2).

35. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol. 1970;10(3):282–290.
36. Willmott CJ. On the validation of models. Phys Geogr.1981; 2(2):184–194.
37. Xu R, Li Y, Guan K, Zhao L, Peng B, Miao C, Fu B. Divergent responses of maize yield to precipitation in the United States. Environ Res Lett. 2022;17(1): 014016.
38. Guttormsen AG. Forecasting weekly salmon prices: Risk management in fish farming. Aquac Econ Manag.1999;3(2):159-166. DOI: 10.1080/13657309909380242
39. Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. Eur J Oper Res.2005;160(2): 501-514. https://doi.org/10.1016/j.ejor.2003.08.037.