Computing and Artificial Intelligence 2024, 2(2), 1427.
https://doi.org/10.59400/cai.v2i2.1427

Article

The computational analysis of COVID-19-induced socio-economic, environmental, and climatic disruptions on the Indian food production system

Adya Aiswarya Dash*, Sonu Sharma

School of Engineering, University of Guelph, ON N1G 2W1, Canada
*** Corresponding author:** Adya Aiswarya Dash, adash@uoguelph.ca

CITATION

Dash AA, Sharma S. The computational analysis of COVID-19-induced socio-economic, environmental, and climatic disruptions on the Indian food production system. Computing and Artificial Intelligence. 2024; 2(2): 1427.
https://doi.org/10.59400/cai.v2i2.1427

ARTICLE INFO

Received: 15 April 2024
Accepted: 4 June 2024
Available online: 18 July 2024

Abstract: COVID-19 dominantly affected all the sectors of the Indian economy, surprisingly the impact is much lower with respect to the agricultural production (−2.7%) in India. The increase in yield of the crops can be attributed to the variables such as environmental, climatic, and socio-demographic factors. The study illustrates the relationship of the increase in crop yield in India during the first wave of COVID-19 along with the increase in the infection count and the land under cultivation attributed to supporting factors during the year 2020. The relation is explained by the method of ordinary least square (OLS) and geographically weighted regression (GWR). The distribution of the increase in crop yield across India is analyzed against COVID-19 infections along with other dominant factors. Useful intuitions against crop yield can be generated by studying the spatial relationships between them. The geographically weighted regression method depicted an increase in R^2 value as compared to the global ordinary least regression method. The Akaike information criterion in the geographically weighted regression method is also lower as compared to the ordinary least square therefore explaining GWR as a better model as compared to OLS. The combination of the various variables affecting agricultural yield in India is found to vary geographically as well as with the type of crops.

Keywords: agricultural yield; ordinary least square regression; geographically weighted regression; COVID-19

1. Introduction

The COVID-19 posed the most astounding impact on the Indian economy in comparison to other major climatic and socio-economic factors. The COVID-19 affected the global supply chain whereas climatic factors such as droughts are more localized ones [1]. As a shielding method against COVID-19 pandemic, the government announced a nationwide lockdown on 25 March 2020, which affected the agricultural economy. The subsequent effect of COVID-19 resulted in a growth of 3.4% in the agricultural sector during the month of April to June, which was though less as compared to the previous year of 2019–2020. The decline of 2.9% can be influenced by the COVID-19 pandemic. The explicit agricultural expansion for the year 2020–2021 can be attributed to the abundant crop harvest, increase in rainfall, more number of people being indulged in the agricultural sector, increase in the cultivated area under agriculture, and increase in the supply of agricultural input necessities such as fertilizers. The stupendous reserve of rice and wheat during the year 2019–2020 assisted the food system of India to face the pandemic [2,3].

During the initial phase of COVID-19 when the economy of all the major sectors remained stagnant, the agricultural sector was excluded with many restrictions but still, it faced large-scale interruptions. Though the pandemic posed major challenges it also opened ways for other possibilities, especially in the logistic agricultural sector. With an increase in migration of people due to lock down, a consistent growth in the number of people employed in agricultural sector has been noticed. The people with skills and knowledge got involved in farmer producer organization to provide necessary insights on demand and supply. During the onset of COVID-19, that is by April 2020, the harvesting of rabi crops turned out to be complete for which the agricultural yield has been less affected [4]. Whereas the output of other sectors such as poultry (−19.5%) and fisheries (−13.6%) got heavily reduced. India is the largest producer of pulses and the second-largest producer of rice and wheat globally. Due to disruption in movement and mobility, the other sector has been hit harder in comparison to agriculture. For the Kharif crops, the area under cultivation stood higher than the previous year. Therefore, the agricultural sector became a segment of reassurance for the Indian economy.

During the year 2020–2021, agriculture appeared to be a vivid field in the Indian economy. The record-breaking monsoon for the consecutive third year opened opportunities in agricultural production, therefore, leading to the growth of the whole sector. A huge amount of stocks for the staple food grains of India appeared to be higher for 6.5 times in rice and about twice in wheat [5]. The agricultural sector has been excluded from the lockdown guidelines and came to be aided by other components such as abundant rainfall, increasing employment in agriculture, boosting fertilizer production, and an increase in area under cultivation.

With the announcement of nationwide lockdown during the COVID-19 pandemic, about 45% of migrants shifted to their homes [6]. Both the agricultural and food sector got influenced by the COVID-19 epidemic [7]. A contemporary study considered the climatic impact on COVID-19 [8]. In a further study, the interrelation between COVID-19 infection with about 35 socio-economic and environmental variables has been explained through regression models such as Geographically weighted regression and (GWR) and multiscale GWR [9]. Many studies have explained the driving forces and their intuitions on agriculture. Few studies have elucidated the likely consequences on a local scale and accomplishing it to a socio-economic and environmental factor. In comparison to the other economies of the world, the Indian economy during the pandemic is sustained by growth in the agricultural sector during the first wave of COVID-19 through the many consequences of it which is not yet well explained on a local scale.

To explain the association between the variables, both global and local models can be used. Most of the factors explaining the relationships associated with COVID 19 emphasize the global models such as OLS (ordinary least regression) [10,11]. The global models consider the relation between the dependent and the explanatory variables to be homogenous that is constant over the study area. The results in the datasets of the global model infer the absence of spatial auto-correlation, whereas the relationship between the dependent and explanatory variables can be geographically inconsistent and heterogeneous [12]. Due to the prevalence of fragile associations between the dependent and the explanatory variable in the global models such as OLS, it results in low accuracy. On the contrary, local models such as GWR [13,14] can explain the geographically varying relationship on a local scale. The local method such as GWR can be used to explain the relationship between agricultural yield and other socio-environmental factors during the prevalence of COVID-19.

Our investigation into the ramifications of COVID-19 on the Indian food production system presents a methodologically innovative approach, distinguished by the integration of ordinary least square (OLS) and geographically weighted regression (GWR) techniques. This fusion allows for a detailed examination of the intricate relationships between environmental, climatic, socio-demographic factors, and variations in crop yield during the pandemic. Unlike conventional studies that solely rely on OLS regression, the incorporation of GWR facilitates the capture of spatially varying effects, thus offering a more comprehensive understanding of the localized impacts across different regions of India.

Moreover, our study showcases the superior performance of GWR over OLS, as evidenced by a notable increase in the R^2 value and lower Akaike information criterion (AIC). This enhancement in model performance underscores the efficacy of GWR in elucidating the complex interplay of factors influencing agricultural yield amidst the pandemic. By harnessing spatial analysis techniques, we unveil geographical patterns and disparities in the impact of COVID-19 infections and other dominant factors on crop yield distribution across India. This spatial perspective enables the identification of localized impacts that may be overlooked in traditional analyses, thereby enhancing the robustness and applicability of our findings.

In addition to our methodological innovation, our study is distinguished by the richness and specificity of the data utilized. Using a diverse range of data sources encompassing COVID-19 infection counts, agricultural statistics, environmental indicators, and socio-demographic variables, our dataset offers a comprehensive transparency of the multifaceted impacts of the pandemic on Indian agriculture. This comprehensive coverage ensures the relevance and applicability of our analysis to the unique challenges faced by the Indian food production system.

Furthermore, the temporal and spatial granularity of our dataset is a notable feature, capturing changes in crop yield and COVID-19 dynamics over the critical period of the pandemic's first wave in 2020. This temporal and spatial resolution enables a detailed examination of localized variations in agricultural performance and pandemic effects across different regions of India. Additionally, the customization of variables modified to the specific context of Indian agriculture, including land under cultivation, environmental conditions, socio-economic indicators, and crop-specific characteristics, further enhances the depth and accuracy of our analysis.

The research gap we aim to address revolves around the extensive understanding of the impact of COVID-19 on the Indian agricultural sector, particularly the intricate interplay between socio-economic, environmental, and climatic factors, and their influence on crop yield dynamics. While existing literature provides insights into the broader socio-economic results of the pandemic, there remains a notable gap in extensively examining the spatio-temporal dynamics of COVID-19-induced disruptions specifically on agricultural production in India. This gap includes the need to:

1) Explore the spatial variations in the relationship between COVID-19 infections and crop yield across different regions of India.

2) Investigate the specific environmental, climatic, and socio-demographic factors that contribute to variations in crop yield during the pandemic.
3) Assess the effectiveness of different regression models in capturing these complex relationships and spatial variations.

Our findings contribute actionable insights for public health and policy development by informing targeted interventions and resource allocation strategies aimed at safeguarding food security and livelihoods amidst the ongoing pandemic. Furthermore, our methodological advancements in spatial analysis pave the way for more robust and detailed assessments, thereby enhancing the evidence base for informed decision-making in public health and agricultural policy domains.

Thus, the main objectives of this study were framed as: (1) To retrieve the socio-economic, environmental, and climatic factors during COVID-19 pandemic driving the agricultural yield in India, (2) to illustrate the geographically varying relationship of agricultural yield with various factors by applying local model such as GWR, (3) to verify the outcome of both global (OLS) and local (GWR) models to establish which is more appropriate.

2. Study area and data

The spatial variation of agricultural yield is modeled based on state-level data across India. The agricultural yield data for the four major crops (rice, wheat, cereals, and pulses) have been cumulated from the Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India [15]. The yield data for the agricultural period of June to September 2020 has been assembled. Five dominant factors are exploited to understand the relationship of geographically varying yield of major crops during the first wave of COVID-19 with socio-economic, environmental, and climatic factors. The precipitation data has been acquired from IMD [16]. The total amount of rainfall for the agricultural period of June to September 2020 has been obtained. **Table 1** summarizes the sources for all the datasets with their link, description, and variable type.

The state-wise data for the number of COVID-19 infections is acquired from the website COVID19 INDIA [17] till the cumulative period of September 2020. The number of infections in each state is computed during the 1st wave of COVID-19. In other words, during the post-harvesting period of the Rabi crop. The infections at each state were calculated and standardized on a log 10 scale and used as a metric to define the variation of infection across the states of India. The geographical dispersal of COVID-19 infection is depicted in **Figure 1**. The third variable containing the amount of fertilizer supplied during the crop growing period of both rabi and Kharif crop has been acquired from the Department of Fertilizers, Government of India [18]. The fertilizer types such as urea, DAP, NPKS, and MOP are acquired in the study. The dataset on the number of people employed in the agricultural sector during the study period have been obtained from the Periodic Labor Force Survey [19] and the dataset for the area coverage under cultivation for the period of June to September 2020 is obtained from Department of Agriculture and Farmers Welfare [20].

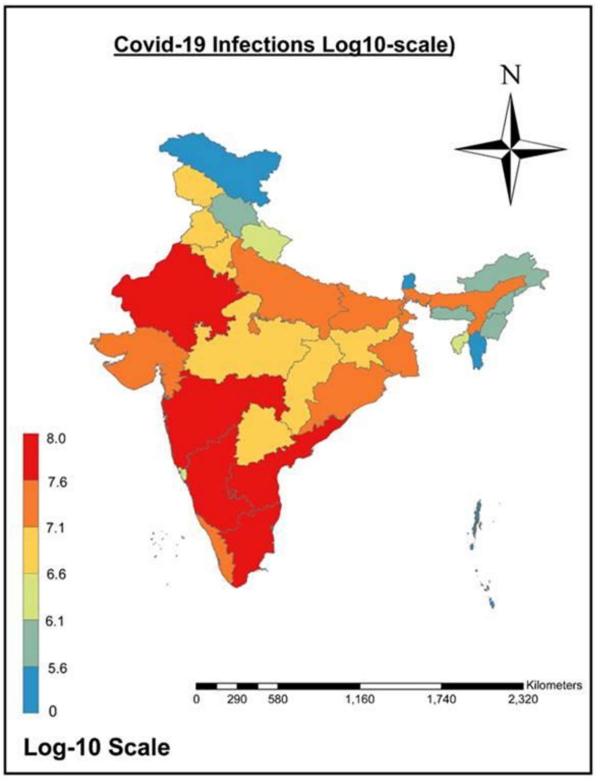


Figure 1. Spread of COVID-19 infections across India (represented on log10 scale).

Table 1. Datasets summary.

Dataset	Dataset explanation	Resource	Names of variables	Description
Agricultural yield data	Represents the yield of major crops up to September 2020.	[15,17]	Rice Wheat Cereals Pulses	State level yield for the four major crops were obtained.
Socio-economic data	Comprises the state wise socio-economic data	[15,19,20]	LOG10	State level COVID-19 infection rate till September 2020
			EAG	State level employment in agriculture for the year 2020
			L_UC	State wise area under cultivation for the four crops.
Environmental	The use of wide different kinds of fertilizers used and the nutrient content of the soil.	[18]	UREA DOP MOP NPKS	State wise amount of fertilizer consumed for the cultivation of the four major crops.
Climatic	Describes the abundant rainfall	[16]	Rainfall	State wise the amount of rainfall received in the year 2020.

Furthermore, the state-level data obtained from each of the states are linked with the geospatial vector data format.

3. Methodology

In this study, the spatio-temporal dynamics of COVID-19-induced disruptions on the Indian agricultural sector, a combination of regression techniques, including ordinary least square (OLS) and geographically weighted regression (GWR) has been used. These methodologies enable us to analyze the complex interactions between COVID-19 infections, environmental variables, socio-economic indicators, and crop yield variations across diverse regions of India. Moreover, insights have been drawn from recent advancements in spatio-temporal modeling of COVID-19 prevalence and mortality. Artificial neural network algorithms was used to model the spatial and temporal patterns of COVID-19 transmission and mortality rates [21]. By integrating findings from this study, the analysis with additional insights into disease prevalence and related variables, has enhanced the comprehensiveness of the study.

3.1. Data preparation

A sum of eight socio-economic, environmental, and climatic factors (**Table 1**) were chosen to describe the state-level geographical variation of the agricultural yield of each of the four major crops during the period of the COVID-19 pandemic. The socio-economic condition aimed to explain the agricultural yield included the employment of people in the agricultural sector for 2020–2021. The stupendous rainfall during 2020–2021 was taken as a climatic factor and the environmental factors included the fertilizer supplied to the crop i.e., urea, DAP, NPKS, and MOP. The area under cultivation was taken in terms of area in lakh hectares and the infection rate of COVID19 was taken as a summation from the period of June to September 2020–2021.

The exploratory regression model assisted in defining the variables chosen for the OLS model. In case we have a variable that doesn't match any of the criteria mentioned in the exploratory regression tool, it can help us to define the suitable variable that has a strong relationship with the dependent variable. We found that out of the four major crops chosen in our study, the yield of pulses and cereals showed an R^2 value less than 0.70, so we have not considered it as an input for study in the OLS model. Furthermore, the yield of wheat and rice showed a higher R^2 value of 0.70 with different explanatory variables thus, taken as an input for the OLS model.

3.2. Global model (ordinary least square)

The OLS described the relations between the explanatory and the dependent variables and can be expressed as in Equation (1) [22].

$$y_i = \beta_0 + x_i\beta + \varepsilon_i \tag{1}$$

where y_i is the yield of the crop at point location i in each of the states during 2020, the explanatory variable is expressed in terms of x_i with β as the regression coefficient with β_0 as intercept. ε_i is the error associated with it. The basic inference that global models such as OLS make is that the regression doesn't vary over space and there is an absence of correlation with the error term [23,24]. As OLS considers the results to be independent with respect to each other so it is mostly considered as a misinterpreted model as in our case agricultural yield is spatially correlated with other variables during COVID-19 [23].

Variance inflation factor (VIF) is used to extricate redundancy from the explanatory variables. VIF can be explained as in Equation (2):

$$\mathrm{VIF}^{i} = \frac{1}{1 - R_i^2} \tag{2}$$

VIF is the ratio of the inclusive variance of the model to a distinct explanatory variable. VIF is computed by regressing one explanatory variable against all others. The resulting R_i^2 value is then used for determining VIF. The coefficient of determination is expressed as R_i^2, where i-th is the explanatory variable regressing on others. The equation for coefficient of determination R_i^2 can be expressed as shown in Equation (3):

$$R_i^2 = 1 - \frac{SSE_i}{SST_i} \tag{3}$$

where SSE_i and SST_i stand for the sum of the square of errors and total variation respectively. The regression analysis of all the eight dependent variables is accomplished to estimate VIF. The variables with VIF greater than 7.5 are considered to be multicollinear. We found that the variables UREA, EAG, NPKS, and L_WHT have comparatively lower multicollinearity in the yield of wheat whereas low VIF is seen in the variables such as MOP and NPKS in the yield of rice. Accordingly, the variables are considered in the set of explanatory variables for the OLS model. The regression analysis is again conducted on the selected variables and their respective VIFs are listed in **Tables 2** and **3** for the yield of wheat and rice. We can verify that the variables are not exceeding the value of 7.5 during the second iteration. The variables aligned with larger VIF are rejected one after another until it has no more variable with larger VIF value.

Table 2. VIFs of the selected explanatory variable (wheat).

Variable	**VIF (*c*)**
Intercept	-
EAG	1.338825
UREA	8.127718
NPKS	2.536351
L_WHT	5.509770

Table 3. VIFs of the selected explanatory variable (rice).

Variable	**VIF (*c*)**
Intercept	-
MOP	4.394808
NPKS	4.394808

3.3. Local model (geographically weighted regression)

In contrary to the global models which state that the relationship between the explanatory and the dependent variable to be constant and doesn’t vary over space, local models such as GWR assume the variables to vary spatially [25]. Through GWR the regression parameter is considered at every point rather than as a whole like in the case of OLS [24]. Fotheringham and Oshan [26] explained GWR as denoted in Equation (4:)

$$y_i = \beta_{i0} + \sum_{j=1}^{m} \beta_{ij} X_{ij} + \varepsilon_i, i = 1, 2, \ldots, n \tag{4}$$

where at location i of the state, y_i is expressed as the yield of the crops for the period 2020, the intercept is expressed as β_{i0}, X_{ij} gives the value of the j-th explanatory variable. ε_i is the error term associated with it. The parameter estimates have been carried out as, Equation (5) [26]:

$$\widehat{\beta(i)} = (X'W(i)X)^{-1}X'W(i)y \tag{5}$$

where $\widehat{\beta}$ is the vector parameter comprising $(m \times 1)$ parameter estimates. The matrix X explains the ($n \times m$) variables. The matrix of spatial weights $(n \times n)$ is expressed as $W(i)$ and it is a diagonal matrix framed by assigning weights to the observation from the location i [23,24]. Proper specification of bandwidth and kernel function is required to determine $W(i)$. Mostly, Euclidean distance (number of nearest neighbors) is used to analyze the bandwidth. Section of different bandwidths is responsible for affecting the choice of neighborhoods where the local weights are assigned. For selecting the bandwidth the Akaike information criteria (AICc) is used with adjustment incorporated in the GWR method.

3.4. Model development

The given step has been carried out to determine the suitable GWR model with appropriate variables. Variables with high variance inflation factor (VIF) seemed to have high multicollinearity, therefore, discarded being unsuitable for the model. The selected variables through regression analysis are considered as an input variable for both OLS and GWR models. The following are the steps conducted for selecting the suitable model for GWR.

Step1: The eight explanatory variables in our study, each of them executed regression against an independent variable that is the yield of the crops such as pulses, cereals, rice, and wheat, respectively. The model with lower AICc and higher R^2 value has been selected and incorporated in the model.

Step 2: Similarly, the rest of the variables have been selected and incorporated into the model. The model with the lowest AICs and VIF less than 7.5 is considered.

The above method is conducted in ArcGIS 10.8 software. For specification of the bandwidth an adaptive bi-square kernel is employed and the AIC (Akaike information criteria) helps to select the suitable optimal bandwidth [24]. Thus, a total of four variables has been selected for the study of the yield of wheat and two variables to study the yield of rice. Moreover, the selected variables are used to evaluate and compare both local and global models.

4. Results

Exploratory regression has been conducted at first to find all suitable coalitions of the explanatory variables that best explain the dependent variable which therefore can be taken as an input for the OLS and GWR model. A flow chart explaining the work of exploratory regression tool, OLS, and GWR has been provided in **Figure 2**.

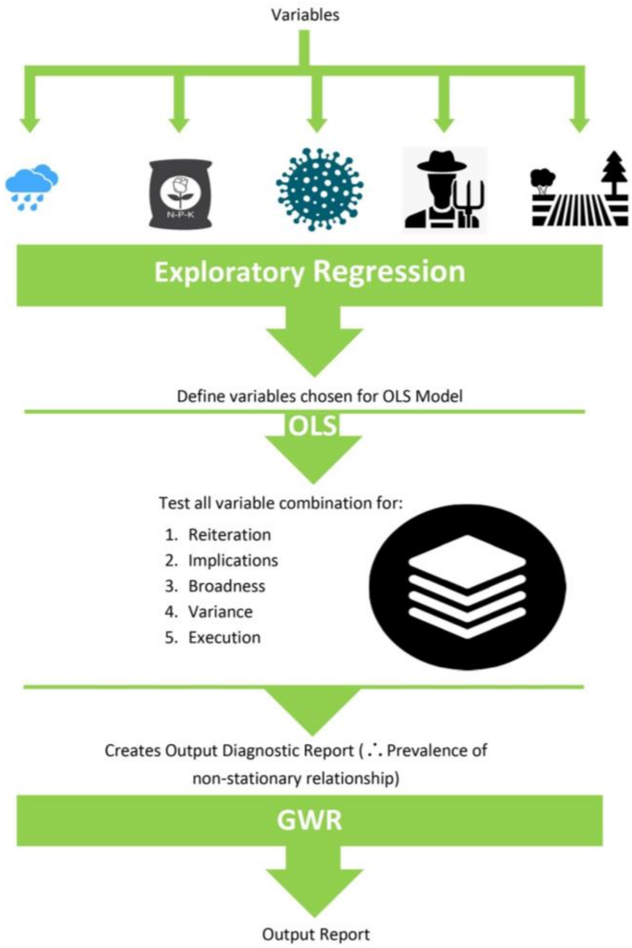


Figure 2. Flow chart depicting the work of exploratory regression, OLS and GWR model.

4.1. Results of exploratory regression tool

An elaborated summary of the suitable combinations of the explanatory variables has been provided in **Tables A1–A4** (see Appendix) for the crops such as pulses, rice, cereals, and wheat respectively. The main reason for using the explanatory regression tools is that it gives the information in the form of detailed analysis of the explanatory variable that is best suited for fitting in the OLS and the GWR model. The highest adjusted R^2 results of the yield of the four crops (pulse, rice, cereals, and wheat) are provided in **Tables A1–A4** (see Appendix). The adjusted R^2 value with the best possible combination can be seen there. Through the explanatory regression tools, it has been derived that variables explain the yield of rice and wheat better than the yield of pulses and cereals having a higher R^2 value along with desirable residual spatial autocorrelation. So, it can be incorporated into the OLS and GWR model. The exploratory regression tool describes the best possible variables that can explain the dependent variable.

4.2. Performance of OLS and GWR model

The final variables EAG, UREA, NPKS, and L_WHT for the yield of wheat and MOP, NPKS for the yield of rice were incorporated in the OLS model. Detailed output of the OLS model is given in **Tables 4** and **5** for wheat and rice respectively. All these variables considered have low VIFs. The OLS model for wheat has an R^2 value of 0.98 which describes that about 98% of the yield of wheat is explained by the given variables and for rice, the R^2 value is about 0.72 which explains that about 72% of the yield of wheat is explained by the dependent variables. Though, the R^2 value of rice is comparatively less than wheat but it is taken into consideration taking into the good residual spatial autocorrelation.

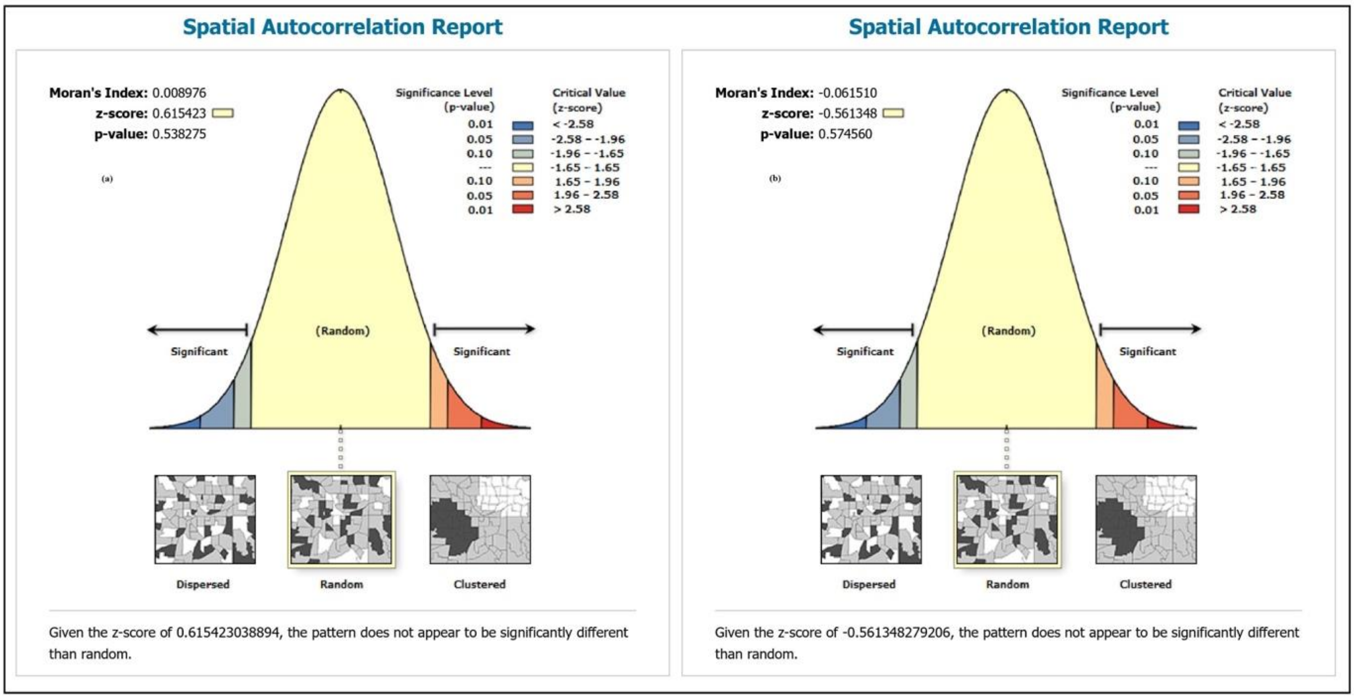
Table 4. Summary of OLS results—Model variables (wheat).

Variable	Coefficient (*a*)	Std error	*t*-statistic	Probability (*b*)	Robust SE	Robust_*t*	Robust_Pr (*b*)	VIF (*c*)
Intercept	14.941393	205.866115	0.072578	0.942595	76.182945	0.196125	0.845754	-
EAG	−120.054018	45.379331	−2.645566	0.012541*	52.384149	−2.291801	0.028645*	1.338825
UREA	106.342278	9.447711	11.255877	0.000000*	14.057580	7.564764	0.000000*	8.127718
NPKS	−69.900249	8.639342	−8.090923	0.000000*	14.051380	−4.974618	0.000021*	2.536351
L_WHT	277.033025	25.852112	10.716069	0.000000*	45.581898	6.077698	0.000001*	5.509770

Table 5. Summary of OLS results—Model variables (rice).

Variable	Coefficient (*a*)	Std error	*t*-Statistic	Probability (*b*)	Robust_SE	Robust_*t*	Robust_Pr (*b*)	VIF (*c*)
Intercept	308.693558	500.874242	0.616310	0.541794	290.356013	1.063155	0.295205	-
MOP	1019.001818	127.662920	7.981972	0.000000*	204.142366	4.991623	0.000017*	4.394808
NPKS	−137.775752	31.666428	−4.350846	0.000116*	50.685667	−2.718239	0.010257*	4.394808

To verify spatial autocorrelation, Moran's I has been analyzed which indicates the pattern to be random for both the yield of rice and wheat with Z score not being statistically significant in the yield of both. A spatial autocorrelation report of the OLS model for the yield of wheat and rice is given in **Figure 3a,b**.



(a) **(b)**

Figure 3. Spatial autocorrelation report—OLS model **(a)** wheat; **(b)** rice.

The Koenker Test is statistically significant in the case of the yield for both wheat and rice which explains the non-stationary relationship among the explanatory and dependent variables, therefore indicating that the model performance can be improved by moving towards the geographically weighted regression model. **Table 6** explains the output of the GWR model for the yield of wheat and rice. The performance matrices of the OLS and the GWR model are summarized in **Table 7**. To explain the yield of rice the OLS model defines only 72% of the relationship whereas the same is explained by 76% in a local model. Similarly, for the yield of wheat, the global model describes only 98% whereas the local model explains about 99% of the relationship between dependent and explanatory variables The AICc for the yield of wheat and rice in OLS is 613.89 and 686.76 which is reduced to 603.07 and 686.37 in case of GWR. Therefore, describing GWR as a better model than OLS due to the lower value of AICc and higher R^2 value.

Table 6. Output of GWR model.

Sr. No.	Variable name	Variable	Definition	Variable	Definition
1	Neighbours	17	-	24	-
2	Residual squares	5,210,795.28771	-	125,599,617.913876	-
3	Effective number	17.428555	-	10.284656	-
4	Sigma	515.98914	-	2168.272064	-
5	AICc	603.075612	-	686.372301	-
6	R^2	0.997218	-	0.827579	-
7	R^2Adjusted	0.994884	-	0.767656	-
8	Dependent field	0	Wheat	0	Rice
9	Explanatory field	1	EAG	1	Mop
10	Explanatory field	2	Urea	2	NPKS
11	Explanatory field	3	NPKS	-	-
12	Explanatory field	4	l_WHT	-	-

Table 7. Performance matrices of OLS and GWR model.

	R^2		AICc	
	OLS	**GWR**	**OLS**	**GWR**
Wheat	0.98	0.99	613.89	603.07
Rice	0.72	0.76	686.76	686.37

The observation of a random distribution pattern in the spatial autocorrelation analysis has several beneficial implications in this study. Firstly, it indicates that there are no spatially correlated clusters of high or low values of the variables, thereby suggesting a lack of spatial dependence or spatial structure in the data. This finding is valuable as it helps to identify regions where the variables exhibit spatial randomness, allowing for a more nuanced interpretation of the underlying spatial processes.

Moreover, the absence of significant spatial clustering implies that there are no distinct spatial clusters of COVID-19 infections or agricultural outcomes, which may have practical implications for policy and intervention planning. In regions with random distribution patterns, interventions may need to be more broadly targeted or tailored to specific local conditions rather than focused on clustered areas.

4.3. Relationship between the agricultural yield of wheat and rice to socio-economic, environment and climatic factors

The R^2 values for the yield of wheat and rice are expressed in **Figures 4** and **5**, respectively. The R^2 value for wheat falls in the range of 0.86–0.99. The northern, north-central, and northeastern parts of the country have a higher R^2 value as compared to the central, eastern, and southeastern parts of the country which have a moderate value. Similarly, for the yield of rice, the north-eastern and eastern parts of the country have higher R^2 values as compared to the northern and western parts of the country.

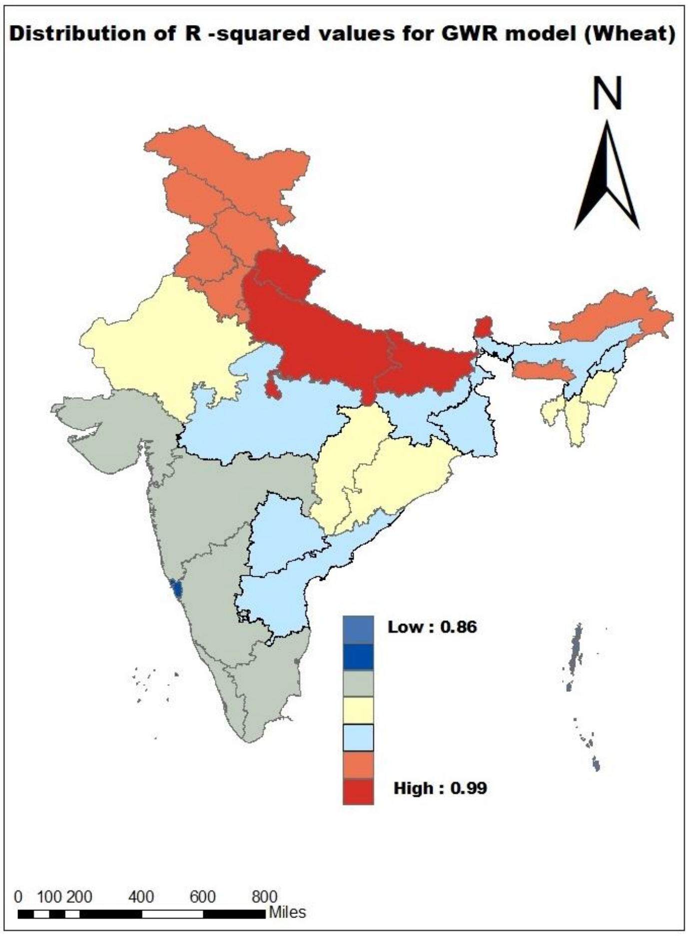


Figure 4. Distribution of R^2 values for GWR model—Wheat.

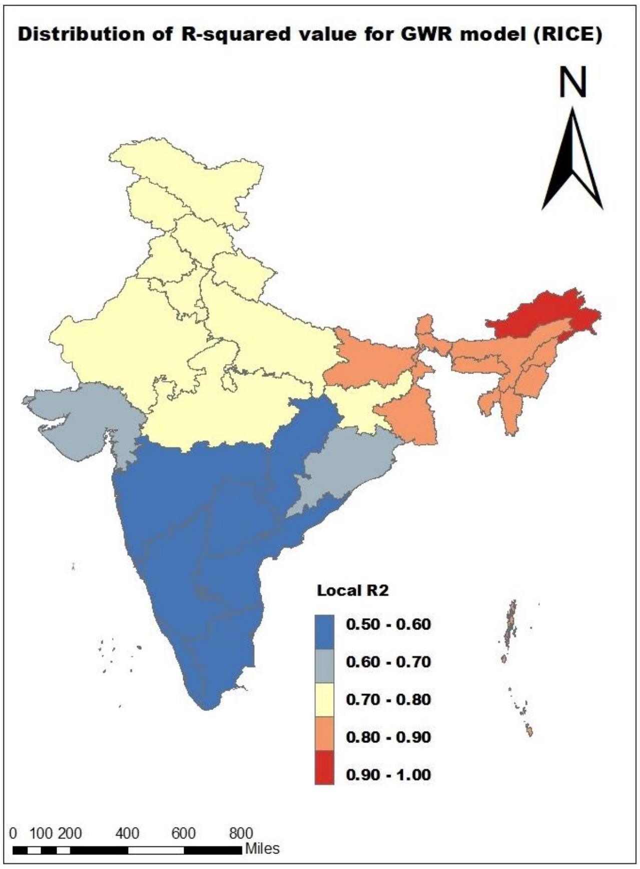


Figure 5. Distribution of R^2 values for GWR model—Rice.

The distribution of the coefficient of the explanatory variables with regards to the yield of wheat is shown in **Figures 6** and **7**. The dominant relationships have a higher coefficient value than the weaker ones. In **Figures 6** and **7**, the red color denotes a much stronger influence on the yield of wheat and rice than the rest in a decreasing manner. **Figure 6a** denotes the variable EAG that varies between the ranges of −3.42 to 0.94. A negative sign indicates a relationship where an increase in the value of the explanatory variable results in a decrease in the value of the dependent variable and the case of a positive sign an increase in the value of the explanatory variable results in an increase in the value of the dependent variable. **Figure 6a** shows the regions such as Punjab, Haryana, Rajasthan, parts of south India like Karnataka, Kerala, Tamilnadu, Andhra Pradesh, Telangana, and areas of NE states such as Nagaland, Manipur, Mizoram, and Tripura describing a strong positive relationship with the variable EAG.

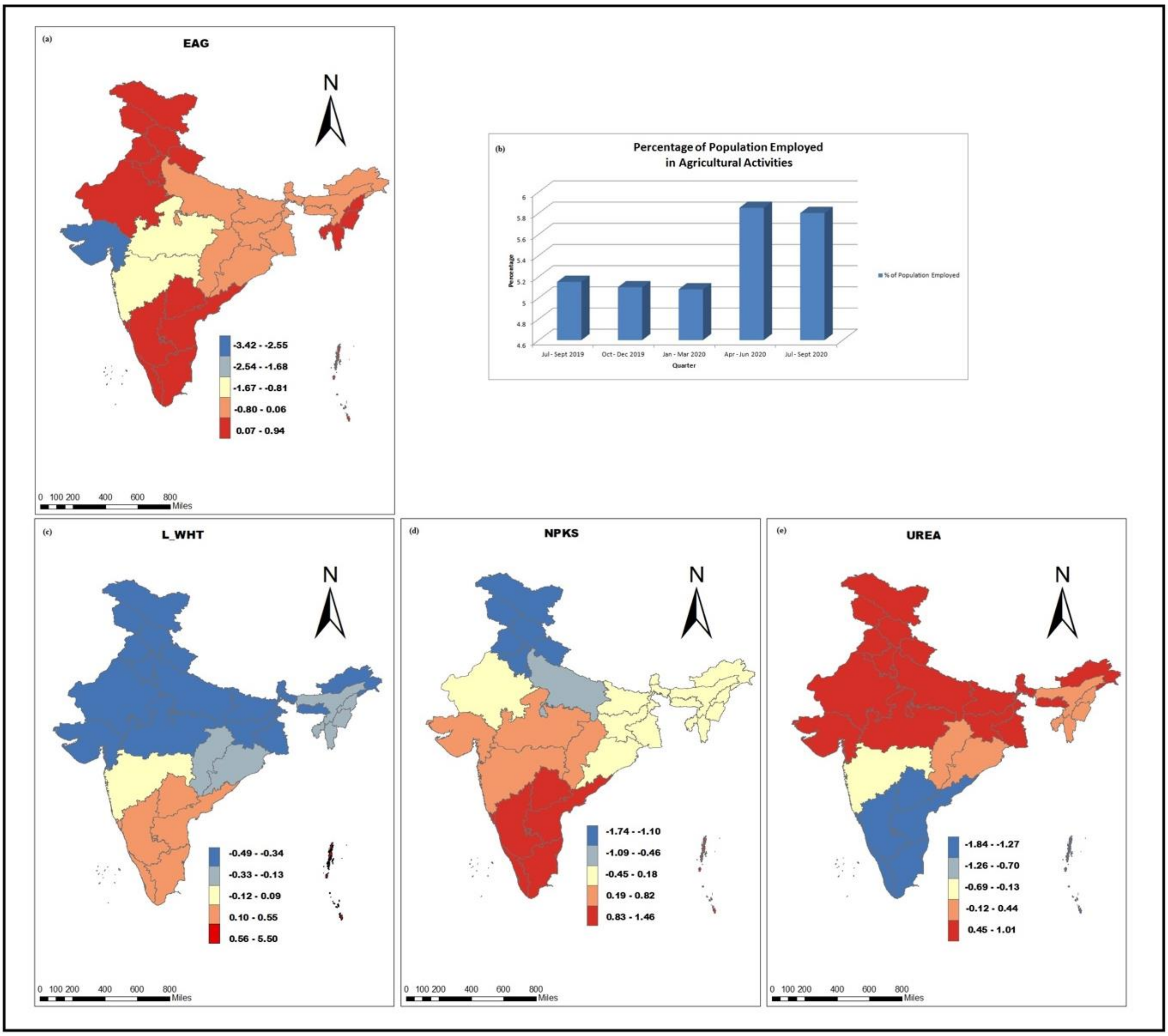


Figure 6. Assessment of local parameters through geographically weighted regression—Wheat; **(a)** EAG; **(b)** percentage of population employed in agricultural activities; **(c)** L_WHT; **(d)** NPKS; **(e)** UREA.

The prevalence of lockdown throughout the country resulted in the decline of major sectors of production such as fishery declined to about 19.6%, poultry by 19.3% but the agricultural showed a minimal decline of only 2.3%. According, to the Periodic Labor Force Survey (PLFS), report the agricultural sector has shown positive growth in labor employment than in the previous years as shown in **Figure 6b**. The percentage of people employed in the agricultural sector during the period July to September 2019 has been about 5.1% with a stupendous increase to 5.8% by the month July to September 2020. During, the period of crisis the agricultural sector has boosted the rural economy at large. Due to lockdown, the growth in migrant labor showed an increase with an increase of wage of about 8.36%. **Figure 6c** denotes the regions such as Karnataka, Kerala, Tamil Nadu, Andhra Pradesh, and Telangana where the variable L_WHT showed a strong relationship with wheat yield. Most of the area grown for the Kharif crops seems to be higher as compared to the last year. It may be because

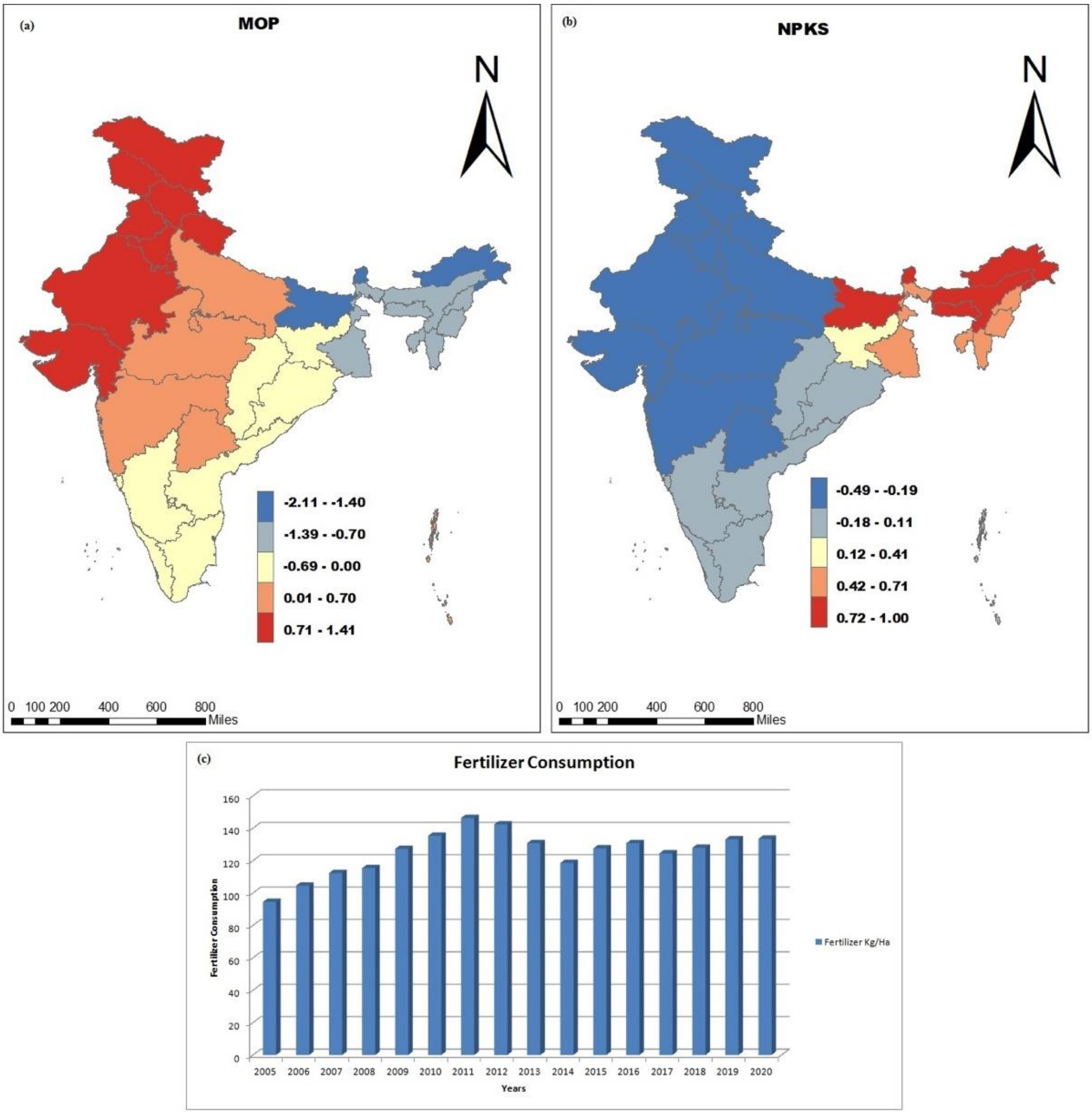


Figure 7. Assessment of local parameters through geographically weighted regression—Rice; **(a)** MOP; **(b)** NPKS; **(c)** Fertiliser consumption (2005–2020).

agriculture has been the only sector in boom during the pandemic. The overall consumption of fertilizers in the form of nutrients i.e., NPKS (Nitrogen, Phosphorous, Potash, and Sculpture) have shown a subsequent increase (see **Figure 6d**) mostly, in the southern parts of India. Similarly, the variable UREA (**Figure 6e**) has a stronger relationship with most of the northern parts of India. The consumption of fertilizer has played a vital role in both the yield of wheat and rice. For the yield of rice, the variable MOP (**Figure 7a**) has been more dominant in the northern and western part of the country whereas, in **Figure 7b** the variable NPKS has been significant in the north-western part of the country. According, to the yearly report by the Department of Fertilizers, GOI the overall fertilizer production has been increased to about 2.7% for the year 2020. With the onset of the southwest monsoon during the year 2020, there has been an increase in the sowing of crops, which increased the consumption, sale, and production of fertilizers. The fertilizer consumption (see **Figure 7c**) has increased from 127.9 kg/ha in the year 2018 to 133.4 kg/ha in the year 2020. Even during the first half of the year 2021, farmers has been piling up the stocks of fertilizers due to the prevalence of pandemic and the possibilities of a break on transportation along with an expectation of a good Kharif season.

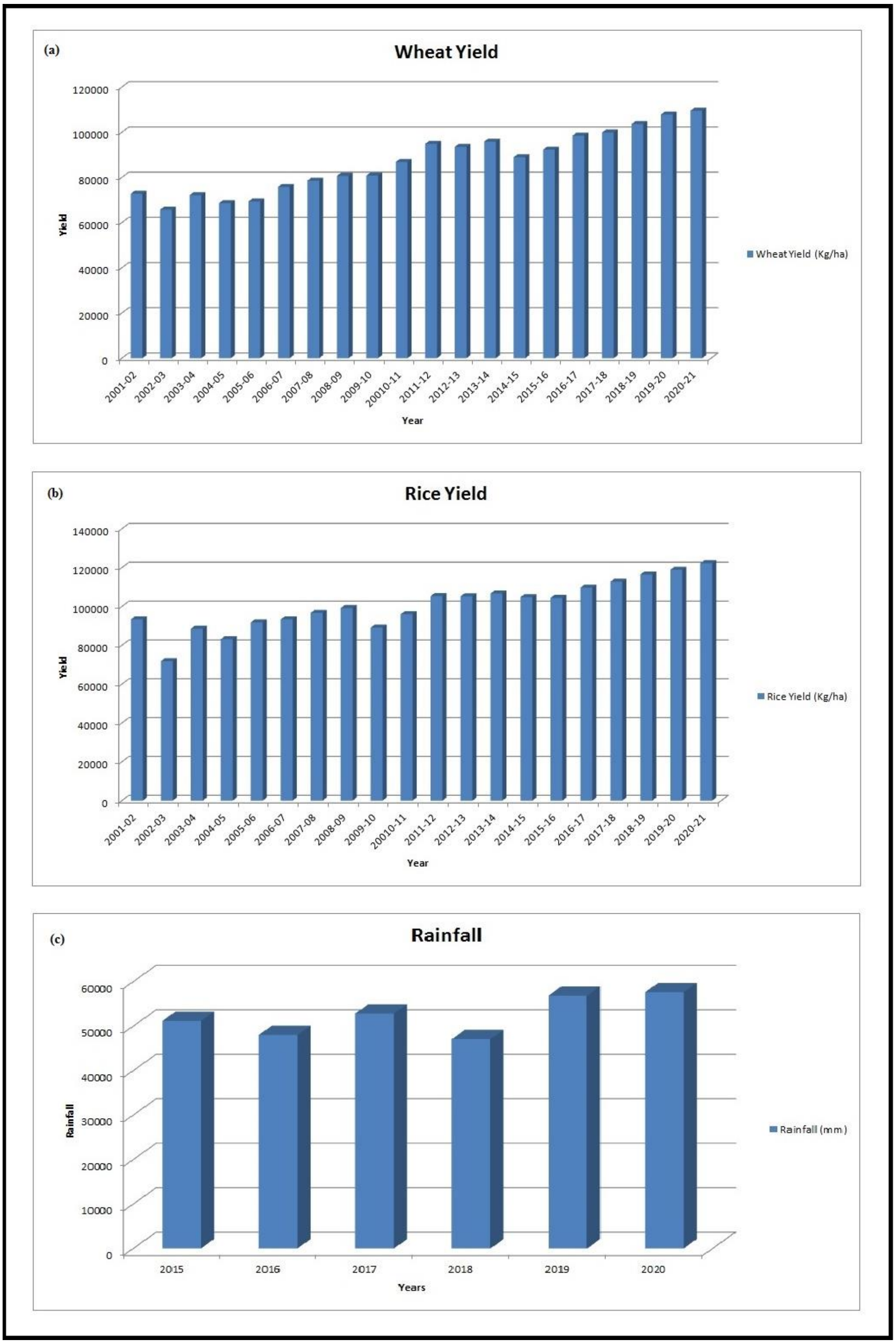


Figure 8. (a) historical yield—Wheat (2001–2020); **(b)** historical yield—Rice (2001–2020); **(c)** rainfall (2015–2020).

Agricultural is one of the pivotal sectors which have shown a positive increase in the gross value added in the production of food grains resulting in an increase of 3.0% in the year 2020–2021. The agricultural yield has seen consecutive positive growth in the last 10 years. The yield of wheat (see **Figure 8a**) has risen from 107,860.5 kg/ha in 2019–2020 to 109,517 kg/ha in 2020–2021. The yield of rice has also shown positive growth in the last 10 years. The yield of rice increased from 118,870.3 in 2019–2020 to 122,265.4 in 2020–2021 (see **Figure 8b**). One of the factors that aided such bumper harvest is rainfall. With the arrival of the southwest monsoon, it stood ahead at 9% from the past long average rainfall (**Figure 8c**). According to the area, much of the states received rainfall above past series except the north-western part of the country. Due to a good amount of rainfall, the area grown under Kharif crops increased by 4.8%. Even the area covered under rabi crops especially wheat has increased subsequently.

From **Figures 4** and **5** we can determine the variation in the R^2 values which shows the relative variation of the yield of wheat and rice. With the help of R^2 values, we can determine the regions where the yield is higher and the regions where the yield is lower. We can relate that the R^2 values (**Figure 4**) in the case of wheat are significantly not much affected in the COVID-19 infected areas (see **Figure 1**) and the range also stood higher from R^2 value 0.96 to R^2 value 0.99. Areas with highly infected COVID-19 cases such as Uttar Pradesh, Haryana, and Punjab do also have high R^2 values for the yield of wheat with 0.997, 0.996, and 0.996 respectively. In much of the high yield states for wheat, we can state that they have a stronger positive relationship with the variables EAG and UREA as compared to that of NPKS and L_WHT.

In the case of rice, the R^2 values for the yield usually varied to a wider range from 0.50 to 1.00. The states most dominantly affected by COVID-19 cases such as West Bengal also have a higher R^2 value of 0.859 (see **Figure 5**). The areas with moderate R^2 values such as Uttar Pradesh, Punjab, Bihar, and Assam also depict higher COVID-19 cases. On the contrary, the southern states such as Andhra Pradesh Telangana and Tamil Nadu have a much lower R^2 value of 0.538, 0.547, and 0.573 respectively. These southern states are also heavily infected by COVID-19 cases. Note, that for most of the high-yield rice areas the variable MOP has a negative relationship whereas the variable NPKS has a positive relationship. In areas with lower R^2 values for the yield of rice e.g., the southern part of India it shows a negative relationship with both variable MOP and NPKS.

5. Discussion

To evaluate the relationship between agricultural yields of two major crops in India such as wheat and rice with other factors during the COVID-19 pandemic, the distribution of yield and infection of COVID-19 has been studied. The western part of the country has more cases of COVID-19 as compared to the other parts of the country. Through, the local GWR model the relationship between the agricultural yield of wheat and rice to the socio-economic, environmental, and climatic factors during the COVID-19 pandemic has been successfully analyzed. Through this, we were able to find out the dominant factors influencing the yield of wheat such as (i) employment in agriculture (ii) land under cultivation of wheat (iii) use of urea as dominant fertilizer and (iv) the ratio of the nutrient percentage of the soil in the form of NPKS.

Similarly, for the yield of rice, we have found the most influencing dominant factors to be (i) use of MOP as fertilizer and (ii) the ratio of the nutrient percentage of the soil in the form of NPKS. We have also compared the global model and the local model. Previous literature has covered the relationship of COVID-19 with various other factors but to our optimum knowledge, this is the first study that encompasses the socio-economic and climatic factors affecting yield in India during COVID-19.

The local model that is GWR performs better as compared to that of the global OLS model with a higher R^2 and lower AICc value. We have also calculated Moran's I for both the yield of wheat and rice which explains the absence of spatial autocorrelation amidst the residuals of the local model. Therefore, it states that the relationships between the dependent and explanatory variables are not homogenous.

The outcome of the study presents the positive relationships among the yield of wheat and the variables EAG and UREA in the areas where there is a significantly higher COVID-19 infection. Earlier studies have shown favorable relation between the yield of wheat and EAG [27]. Similarly, earlier reports explain the positive relationship between the variable UREA and the yield of wheat [28]. Though a strong positive union is observed between these two variables there are other parts of India where a weaker relationship is observed. We can find a more dominant relationship between the yield of wheat and EAG in the areas more affected by COVID-19. Due to lockdown, more persons migrated to their native lands which have subsequently resulted in an increase in labor in those areas and employment in agriculture [29]. Previous literature has also been harmonious with the increase of yield during the pandemic mostly in the states of Punjab, Rajasthan, Telangana, and Gujarat. One of the surprising observations is that in most of the high R^2 values for the yield of wheat, a negative relationship is found with the variable L_WHT. The agricultural land may have declined due to floods in the areas of Assam and Bihar, because of irregular or deficient presence of nutrient NPKS in the soil, overuse of UREA that has led the soil barren, lack of advent technology in agriculture, and deficient of irrigation. In the case of the yield of rice, the variation in R^2 value is wider. The most dominant crop-growing states such as Andhra Pradesh and Tamil Nadu are also highly affected by COVID-19. Thus, shows a negative relationship with variables MOP and NPKS whereas areas such as West Bengal though have a higher amount of COVID cases also have a comparatively higher yield with a positive relationship with the variable NPKS. Areas such as Uttar Pradesh, Punjab, Haryana, and Telanagana which have moderate COVID-19 infections have moderate R^2 values for the yield of rice showing mostly a negative relationship with the variable NPKS. Therefore, it can be inferred that in the areas where rice cultivation is dominant, there is more use of urea as fertilizer. The green revolution has made dependence on high-yielding varieties of seeds which have resulted in the use of more urea ultimately leading to an imbalance of soil nutrients in the form of NPKS [30]. Areas such as Arunachal Pradesh have low COVID-19 infections, comparatively higher yield with a stronger relationship with the variable NPKS.

There are certain limitations in this research that can be resolute in future work. Due to the unavailability of the latest district-wise data for the yield of wheat and rice during the year 2020, the state-wise data has been incorporated so the distinctness of the data may have been reduced. Because of certain inaccessible data factors such as farmer's income, MSP prices, outcomes of various policies, Pradhan Mantri Fasal Bima Yojna, Kisan credit cards, the impact of groundwater level has not been included in the study. However, with all those limitations, this is one is the first study that has explained the geographically varying relationships of the major crops of India with various socio-economic, environmental, and climatic factors during the prevalence of COVID-19.

Limitations and future research

Researchers and policymakers should recognize the uncertainties associated with our data and consider alternative sources or supplementary analyses to validate the results. Moreover, future research should aim to conduct more detailed and temporally dynamic analyses to capture localized variations and temporal trends with greater precision. Addressing the methodological limitations of this study analysis requires sensitivity testing and exploration of alternative modeling approaches to improve the validity and generalizability of our findings.

Future research directions could focus on exploring causal mechanisms underlying COVID-19 impacts on agricultural productivity, investigating sector-specific vulnerabilities, and resilience strategies, and incorporating qualitative research methods to capture stakeholder perspectives. By addressing these gaps, researchers can advance the understanding of the complex dynamics of COVID-19 impacts on agriculture and inform evidence-based interventions to mitigate its effects.

6. Conclusion

In conclusion, our study highlights the substantial impact of COVID-19 on agricultural yield, revealing a geographically changing relationship with various influential factors across India. We observed that this relationship is not constant, attributed to the dynamic interplay of socio-economic, environmental, and climatic factors across different regions. Through the application of the geographically weighted regression (GWR) model, we elucidated the efficacy of this relationship, with GWR demonstrating superior performance compared to the global ordinary least squares (OLS) model. Specifically, GWR exhibited a higher R^2 value and lower AICc, underscoring its suitability for capturing localized variations and providing a more accurate representation of the spatial dynamics of agricultural yield. These findings underscore the importance of adopting spatially explicit modeling techniques like GWR to better understand and address the nuanced impacts of COVID-19 on agricultural productivity, thus informing targeted interventions and policy decisions to support agricultural resilience in India and beyond.

Author contributions: Conceptualization, AAD; methodology, AAD; software, AAD; validation, AAD, and SS; formal analysis, AAD; investigation, AAD; resources, AAD; data curation, AAD; writing—original draft preparation, AAD; writing—review and editing, AAD and SS; visualization, AAD; supervision, AAD and SS; project administration, AAD. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Mishra A, Bruno E, Zilberman D. Compound natural and human disasters: Managing drought and COVID-19 to sustain global agriculture and food sectors. Science of The Total Environment. 2021; 754: 142210. doi: 10.1016/j.scitotenv.2020.142210
2. Padhee A, Pingali P. Lessons from a pandemic to repurpose India's agricultural policy. Nature India. 2020; 1-9. doi: 10.1038/nindia.2020.83
3. FAO United Nations. Local Food Systems and COVID-19; A glimpse on India's responses First. Food and Agriculture Organization of the United Nations; 2020.
4. Department of Economic Analysis & Research. Impact Assessment of COVID-19 on Indian Agriculture and Rural Economy. Department of Economic Analysis & Research; 2020.
5. Reserve Bank of India. Reserve Bank of India Annual Report 2020-2021, RBI Annual Report 2020–2021. Reserve Bank of India; 2021.
6. Imbert C. Covid-19: Expected migrant movement as lockdown eases, Ideas for India. Available online: https://www.ideasforindia.in/topics/macroeconomics/covid-19-expected-migrant-movement-as-lockdown-eases.html (accessed on 7 January 2022).
7. Manzanedo RD, Manning P. COVID-19: Lessons for the climate change emergency. Science of The Total Environment. 2020; 742: 140563. doi: 10.1016/j.scitotenv.2020.140563
8. Jha S, Goyal MK, Gupta B, et al. A novel analysis of COVID 19 risk in India incorporating climatic and socioeconomic Factors. Technological Forecasting and Social Change. 2021; 167: 120679. doi: 10.1016/j.techfore.2021.120679
9. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Science of The Total Environment. 2020; 728: 138884. doi: 10.1016/j.scitotenv.2020.138884
10. Sun F, Matthews SA, Yang TC, et al. A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? Annals of Epidemiology. 2020; 52: 54–59. doi: 10.1016/j.annepidem.2020.07.014
11. Hutcheson G. Ordinary Least-Squares Regression. In: The Multivariate Social Scientist. SAGE; 2011. pp. 56–113. doi: 10.4135/9780857028075.d49
12. Cormack RM, Cressie N. Statistics for Spatial Data. Biometrics. 1992; 48(4): 1300. doi: 10.2307/2532724
13. Fischer MM, Getis A. Handbook of Applied Spatial Analysis. Springer Berlin Heidelberg; 2010. doi: 10.1007/978-3-642-03647-7
14. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships. Wiley; 2002.
15. Directorate of Economics and Statistics, Department of Agriculture and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, Government of India. Available online: https://eands.da.gov.in (accessed on 7 January 2022).
16. IMD. Available online: https://mausam.imd.gov.in/ (accessed on 7 January 2022).
17. Covid19 India. Coronavirus Outbreak in India. Available online: https://www.covid19india.org (accessed on 7 January 2022).
18. Department of Fertilizers, Government of India. Available online: https://www.fert.nic.in (accessed on 7 January 2022).
19. Schemes VS. Press Information Bureau. Available online: https://www.pib.gov.in/AllRelease.aspx?MenuId=30 (accessed on 7 January 2022).
20. Department of Agriculture & Farmers Welfare. Available online: https://www. agricoop.nic.in (accessed on 7 January 2022).
21. Kianfar N, Mesgari MS, Mollalo A, et al. Spatio-temporal modeling of COVID-19 prevalence and mortality using artificial neural network algorithms. Spatial and Spatio-temporal Epidemiology. 2022; 40: 100471. doi: 10.1016/j.sste.2021.100471
22. Ward M, Gleditsch K. Spatial Regression Models. SAGE Publications, Inc.; 2011. doi: 10.4135/9781412985888
23. Anselin L, Arribas-Bel D. Spatial fixed effects and spatial dependence in a single cross‐section. Papers in Regional Science. 2013; 92(1): 3‐18. doi: 10.1111/j.1435-5957.2012.00480.x
24. Oshan TM, Smith JP, Fotheringham AS. Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. International Journal of Health Geographics. 2020; 19(1). doi: 10.1186/s12942-020-00204-6
25. Brunsdon C, Fotheringham AS, Charlton ME. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geographical Analysis. 1996; 28(4): 281–298. doi: 10.1111/j.1538-4632.1996.tb00936.x

26. Fotheringham AS, Oshan TM. Geographically weighted regression and multicollinearity: dispelling the myth. Journal of Geographical Systems. 2016; 18(4): 303–329. doi: 10.1007/s10109-016-0239-5
27. Pinto A, Bhowmick A, Adlakha RK. How did India's rural economy fare through the COVID-19 lockdown and the re-opening? Available online: https://blogs.worldbank.org/endpovertyinsouthasia/how-did-indias-rural-economy-fare-through-covid-19-lockdown-and-re-opening (accessed on 7 January 2022).
28. PRS Legislative Research. Available online: https://prsindia.org/policy/analytical-reports/state-agriculture-india (accessed on 7 January 2022).
29. Department of Economic Analysis & Research. Available online: https://www.nabard.org/about-departments.aspx?id=5&cid=463 (accessed on 7 January 2022).
30. Chander S. Rationalization of Prices for Balanced Fertilization. Indian Journal of Fertilisers. 2009; 15(5).

Appendix

Table A1. Summary of exploratory regression—Wheat.

Adj R^2	AICc	JB	K (BP)	VIF	SA	Model
0.94	665.34	0.00	0.00	1.00	0.24	+L_WHT***
0.84	698.78	0.04	0.08	1.00	0.00	+DAP***
0.76	714.91	0.43	0.00	1.00	0.00	+UREA***

Table A2. Summary of exploratory regression—Pulses.

Adj R^2	AICc	JB	K (BP)	VIF	SA	Model
0.94	535.87	0.00	0.00	1.00	0.50	+L_PUL***
0.43	621.64	0.00	0.00	1.00	0.14	+DAP***
0.36	626.17	0.00	0.00	1.00	0.23	+UREA***

Table A3. Summary of exploratory regression—Rice.

Adj R^2	AICc	JB	K (BP)	VIF	SA	Model
0.58	700.23	0.01	0.00	1.00	0.14	+MOP***
0.52	704.68	0.00	0.25	1.00	0.05	+UREA***
0.39	714.09	0.00	0.49	1.00	0.16	+DAP***

Table A4. Summary of exploratory regression—Cereals.

Adj R^2	AICc	JB	K (BP)	VIF	SA	Model
0.46	657.28	0.00	0.67	1.00	0.18	+MOP***
0.45	657.90	0.00	0.80	1.00	0.13	+NPKS***
0.43	658.95	0.01	0.00	1.00	0.79	+LOG10***