Geosciences

AIMS Geosciences, 12(2): 302–334.
DOI: 10.3934/geosci.2026012
Received: 26 September 2025
Revised: 21 January 2026
Accepted: 26 February 2026
Published: 11 March 2026

https://www.aimspress.com/journal/geosciences

Research article

Urban heat island assessment using Landsat-derived indices and geographically weighted regression in Sonipat district, India

Diksha[1], **Maya Kumari**[2,*], **Rayudu Venkata Komali Supraja**[3], **Varun Narayan Mishra**[3,*]

[1] Division of Environment Science, ICAR-Indian Agricultural Research Institute, IARI, Pusa campus, New Delhi-110012. India
[2] Amity School of Natural Resources & Sustainable Development, Amity University Uttar Pradesh, Noida, Sector 125. India
[3] Amity Institute of Geoinformatics and Remote Sensing, Amity University Uttar Pradesh, Noida, Sector 125. India

* **Correspondence:** Email: vnmishra@amity.edu; mkumar10@amity.edu; Tel: +91120 4395601.

Abstract: Urban heat islands (UHI), a confined rise in temperature relative to surrounding areas, are one environmental problem caused by urbanization, despite being essential for societal growth. By analyzing land surface temperature (LST) data from Landsat-7 (L7) and Landsat-8 (L8) Earth observation data and processing them using ArcGIS, SAGA GIS, and Google Earth Engine (GEE) software and cloud platform for 2011 and 2021, this study sought to quantify the UHI in the Sonipat district of Haryana. Results indicated that areas with high Normalized Difference Built-Up Index (NDBI) values correspond to elevated LST, while reduced Normalized Difference Vegetation Index (NDVI) values across all land use types reveal a strong negative spatial correlation with LST, further supported by increasing Moran's index values. Following the mapping and identification of locations exhibiting varying UHI intensities, the Urban Thermal Field Variance Index (UTFVI) indicated that regions with higher UHI intensity were predominantly situated in residential areas, whereas zones with diminished intensity were located in agricultural lands. An ordinary least squares (OLS) analysis demonstrated that Surface albedo (SA), solar radiation (SR), sky view factor (SVF), and NDVI accounted for 37.6% of LST variability in 2011, which considerably increased to 92.8% in 2021, demonstrating a more robust and consistent influence of these components over time. Due to the nonstationarity of LST, a geographically weighted regression (GWR) model was also considered and demonstrated superior performance compared to the OLS model. This study will aid policymakers, stakeholders, and local governments in effectively developing new urban and constructed environments to mitigate significant UHI impacts.

Keywords: autocorrelation; geographically weighted regression; land surface temperature; normalized difference vegetation Index; ordinary least square; sky view factor; urban heat island; urban thermal field variance index

1. Introduction

The global population has grown significantly in recent years, hastening urbanization, particularly in developing countries where people are migrating from rural to urban areas in search of better living conditions, infrastructure, and job opportunities [1]. Urban areas currently contain nearly 50% of the global population, and estimates suggest that it may rise to about two-thirds by the year 2050 [2]. According to [3], the rapid relocation of people from urban and rural spaces to urban spaces has not simply led to growth in new megacities but has also increased the population density of currently inhabited ones. The large-scale migration causes extensive pressure on the ecosystem and environmental outcomes at both global and localized scales [4]. The replacement of natural landscapes with constructed and man-made entities causes urban regions to hold warmer temperatures relative to other neighboring rural zones, an effect called the urban heat island (UHI) effect [5,6]. Elevated temperatures in urban spaces have been attributed to a decrease in living comfort [7], as well as wider ecological concerns, including air quality decreases, increased global warming, and health hazards [8]. In addition, UHI imprints upon weather conditions, including an effect on the level of humidity, cloud covering, fog development, and even rainfall [9,10]. Residents of such locations are exposed to increased vulnerability to heat stress, dehydration, and heat sickness, while pre-existing conditions may be exacerbated with elevated temperatures [11–13]. Moreover, the greater need for cooling devices like air conditioning aggravates energy consumption and leads to the release of harmful compounds, including chlorofluorocarbons (CFCs) [14,15].

The reduction of evapotranspiration as a result of the availability of impenetrable surfaces, combined with heat accumulation in buildings and infrastructure, aggravates the problems [16]. In the past couple of decades, the strength of the heat island urban effect has significantly escalated in large metropolitan regions [17]. Even if urban growth is often seen as a development [18], it has simultaneously altered natural landscapes, increased soil sealing, changed surface albedo, and disrupted the local energy balance [19]. Other processes, including slower rates of cooling, radiation trapping, and variations in land coverage, are equally responsible for generating the urban heat island (UHI) effect [20]. Local dynamics linked to landscape transformation generate these variations in surface temperature [21,22]. Main drivers include the radiating of heat due to anthropogenic activity, the usage of heat-retentive building materials, inadequate air circulation between tightly situated buildings, and the dominance of low-albedo surfaces that absorb instead of reflect sunshine [23]. To evaluate these interactions, researchers utilize a combination of satellite images, ground measurements [24,25], and atmospheric instruments such as transportable thermal cameras for mapping and quantifying UHI patterns.

The combination of varied datasets has given a new understanding of surface urban heat islands (SUHI) and broader processes of urban heat islands (UHI) [26]. While direct field measurement is not always feasible, particularly over a broad area of study, satellite-based observation constitutes an important mode of analysis of UHI in developing and advanced contexts [27]. Some of the most utilized thermal infrared sensors are TIRS-2, TIRS, ETM+, TM, and MODIS, which have been utilized extensively in the observation of land surface temperature (LST) and associated UHI patterns [28,29]. These sensors allow a series of applications, including the estimation of surface temperature, evapotranspiration evaluation, vegetation health evaluation, hydrological cycles analysis, and climatic processes observation [30]. Depending on available spectral bands, scientists would either use the single infrared channel approach or the split-window approach of analysis [31]. Impacts and drivers of UHI, and increasing levels of severity of its consequences, have attracted a series of studies [32].

Numerous others have investigated the processes behind urban heat differences, and the findings have been used by policymakers in developing mitigation plans to lower heat illness risk levels [33]. Significant research has also been conducted on the underlying mechanisms that cause variations in the air temperature in urban settings. The findings of this research give local governments crucial knowledge on the health effects of UHI, which they can utilize for urban planning and mitigation.

Despite the rapid growth of cities in areas near cities, like Sonipat, most UHI research is still short-term or focused on one city, missing out on how temperatures change over decades in secondary urban areas. Moreover, the simultaneous use of OLS and GWR over a decade to analyze both global trends and localized nonstationary drivers of UHI has not been sufficiently studied in the Indian context. The aim of this decadal study is to analyze long-term variations in land surface temperature and UHI intensity in Sonipat district through multi-temporal remote sensing and GIS methodologies. This research seeks to elucidate the effects of urban expansion and land use/land cover modifications on surface thermal dynamics, assess the correlation between land surface temperature and significant urban-environmental variables via ordinary least squares regression, and identify spatial discrepancies in these correlations using geographically weighted regression to pinpoint decade-specific UHI hotspots and their underlying causes.

The effect of explanatory variables, namely the Normalized Difference Vegetation Index (NDVI), solar radiation (SR), surface albedo (SA), and sky view factor (SVF), on land surface temperature was assessed. Urban Thermal Field Variance Index (UTFVI) was used to estimate the strength of UHI, and ordinary least squares (OLS) and geographically weighted regression (GWR) models were used in the analysis of spatial variation [34]. Additionally, the findings highlight the role of vegetation cover in worsening thermal conditions. By investigating explanatory variables spatially, this study hopes to clarify patterns of UHI and provide suggestions for mitigation measures in rapidly urbanizing places.

2. Study area

The study site, Sonipat district, lies within the jurisdiction of India's National Capital Region, as evident in Figure 1. Sonipat is a district of 2213 km^2 and is situated between Delhi and Uttar Pradesh [35]. It occupies the period of the Quaternary and the epoch of the Pleist, which created the Indo-Gangetic Plains. With an average elevation of roughly 224 m, Sonipat's elevation varies. The difficulties experienced by many developing regions around the world are best illustrated by Sonipat's distinctive combination of fast urbanization, reliance on agriculture, and varied landscape typologies within the NCR [36]. By examining these patterns, results can be extrapolated to similarly agrarian and urbanizing regions, guiding sustainable land management strategies in a variety of settings. The primary water supply system in Haryana comprises the Yamuna River and the irrigation canals that run from it. There are no perennial rivers in the district, and the subsurface water supplies vary by region. At ten feet, the water table is lowest in the Khader region along the Yamuna. Depths of 30–40 feet can be found in the district's southeast and western regions. Groundwater in some places is saline and brackish. Although there are some uneven areas, the region is generally a continuous portion of the Haryana–Punjab plain. The district's soil is richly colored and fine loam. While some regions have Kallar soil, others have sandy soil [37]. To the east and south, the plain descends slightly. Khadar, the upland plain, and the sandy region are the three separate regions that make up the district. The district experiences cold winters and dry summers, with the monsoon season beginning in July and lasting until September [38]. December is when winter officially begins. The months with the highest temperatures are May and June, while January has the lowest temperatures. Light rains fall in December, January, and February. Sonipat receives the majority of its annual precipitation from the southwest monsoon, which accounts for nearly 70–80% of yearly rainfall. Mean annual rainfall ranges roughly between 550 and 700 mm, though strong interannual variability is common due to monsoon dynamics.

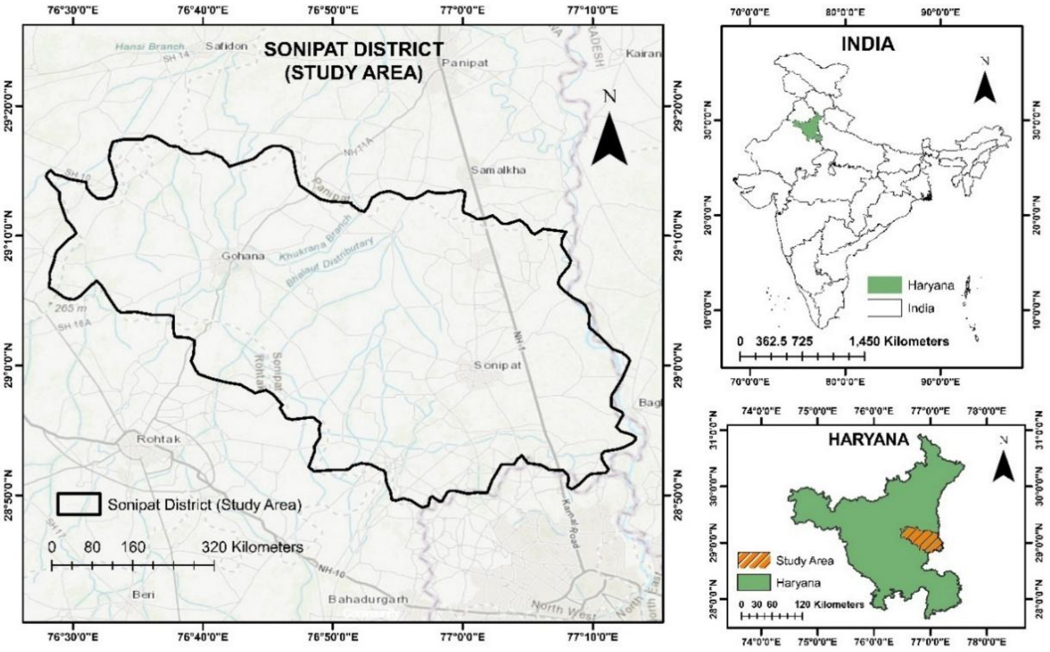


Figure 1. Geographical location of the study area.

3. Data and software used

As indicated in Table 1, the L7 TM and L8 (TIRS and OLI) datasets in GeoTIFF format were downloaded from the United States Geological Survey (USGS) website. A variety of spectral bands, including thermal and optical bands, were included in these downloaded files. These bands were used to generate NDVI and SA maps, from which the values of these parameters were extracted to identify the UHI effects in the study area. Additionally, ASTER DEM data were used to calculate the SVF and SR. All calculations and data processing were done using ArcMap 10.8, QGIS software, and Google Earth Engine (GEE) cloud platform.

Table 1. Data and software used.

Data	Resolution	Cloud cover	Acquisition time	Process/software/cloud platform
L7	30 m (spatial resolution)	Less than 10%	Mean pixel images were considered for the month of February.	ArcMap 10.8, GEE, and QGIS https://earthexplorer.usgs.gov/
L8	30 m (spatial resolution)	Less than 10%	Mean pixel images were considered for the month of February.	
ASTER (DEM)	30 m (spatial resolution)	Not applicable	Not applicable	

4. Methodology

The approach used to assess and estimate the UHI effect in the Sonipat district is shown in Figure 2.

4.1. Data preprocessing

During the stage of data preparation, a satellite image was acquired and refined. Landsat images had a 30 m spatial resolution and were 28-day mean pixel images. With a kernel window of 30, the study concentrated on February, the month with the least cloud cover, and scan line fixes on GEE. In Google Earth Engine (GEE), the Landsat-7 SLC-off gaps were made smaller by using a focused (kernel) filter to fill in missing stripes. Then, several images were mosaicked or temporally composited to fill in any remaining gaps. This kernel-based smoothing helps bring back the thermal band's spatial continuity while keeping the LST gradients realistic. Preprocessing helped with finding UHIs in the research region. High-temperature pixel clusters were used to characterize locations as UHIs. Using properties like SA and SVF, the connections between these clusters and UHI intensity were explored. LST changes were explained by NDVI, SA, SVF, and SR.

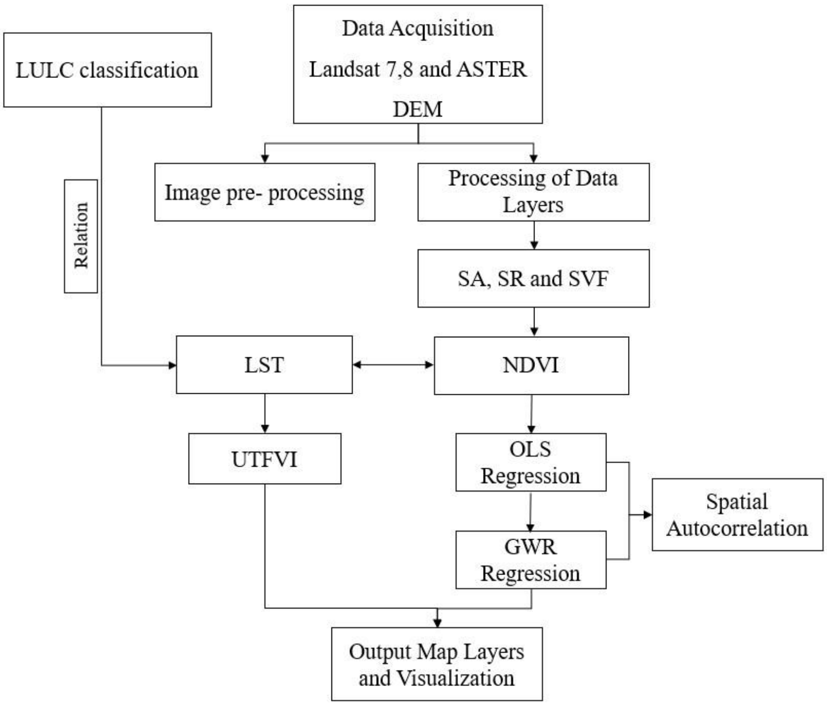


Figure 2. Methodology flowchart.

4.2. Computation of LST

LST was computed using the formula below, particularly from L8:

$$T_S = \frac{BT}{\left\{1+\left[\left(\frac{\lambda \mathrm{BT}}{p}\right) In\varepsilon_\lambda\right]\right\}} \tag{1}$$

Where T_S is LST in °C [39]. BT is the temperature of the brightness. λ is the average wavelength of Landsat 8 image, band 10. p is [h × (c/σ)] = 1.438 × 10^{-2} mK. h = 6.62607004 × 10–34 m^2 kg/s (Planck's constant). c = speed of light ~ 3.00×108 m/s. σ = 1.38064852 × 10–23 m^2 kg s^{-2} K^{-1} (Boltzmann Constant) [40].

To estimate LST using images from the Landsat 7 satellite, the equation for the thermal band of the image's digital numbers (DN) was used, after converting to spectral radiation (Lλ) [41]:

$$L\lambda\ = Gain\ X\ DN\ +\ of\ f\ set, \tag{2}$$

The conversion formula is

$$T_B = \frac{K2}{In\left(\frac{K1}{L_\lambda}+1\right)} \tag{3}$$

The following equation was used to calculate emissivity-corrected LST.

$$LST = \frac{T_B}{1+\left(\frac{\lambda T_B}{P}\right) In\epsilon} \tag{4}$$

The aforementioned formulas outline the typical process for obtaining land surface temperature (LST) from Landsat thermal data, which involves converting digital values to brightness temperature and spectral radiance first and then correcting for emissivity. To provide consistency across sensors, a similar radiance–temperature–emissivity structure is used for Landsat-7 ETM+ thermal data, whereas wavelength-specific emissivity is used to correct Band-10 brightness temperature for Landsat-8.

4.3. Computation of surface albedo

SA stands for the percentage of incoming solar energy, which is then reflected into space. It is a significant variable or parameter for determining the energy of the surface equilibrium, especially for studies of ground and climate conditions. SA ranges between 0 and 1, where 0 is a totally absorptive surface, usually a deep-water body or dense jungle, and 1 is an ideal reflective surface, for example, snow [42]. The surface albedo from Landsat 7 and 8 data can be calculated by the following equation:

$$SA = c_o + c_1B_1 + c_2B_2 + \ldots\ldots + c_7B_7 \tag{5}$$

Where SA is the surface albedo, and B_1 are band values.

4.4. Computation of sky view factor

The ratio of the visible sky area from a ground-based observation site to the entire hemispherical sky dome above is referred to as the SVF. This is one of the concepts used in the analysis of energy balance and climatology of cities. It regulates the exchange of longwave radiation between the atmosphere and the ground surface. SVF ranges between 1 and 0, where 0 signifies deep valleys or limited urban canyons where the sky is almost completely blocked by nearby structures or natural formations, and 1 implies broad, flat plains under the full, open sky [43]. SVF is a significant factor influencing UHI since it represents the geometry of urban surfaces. The form of an urban surface greatly influences the circulation of surface air temperature. SAGA GIS, an open-source program, was used to calculate the SVF, utilizing DEM raster as input data. The general formula for SVF at a point P is as follows:

$$SVF\ (P) = \frac{1}{\pi}\int_0^{2\pi} cos^2\theta(\phi)d\phi \tag{6}$$

Where $\theta(\phi)$ is the elevation angle of the horizon in the direction ϕ (azimuth).

The integral computes the average sky visibility across all directions.

4.5. Solar radiation values

The intensity of the UHI is reliant on the solar radiation emitted by the sun, the season, and the landscape of the region [44]. A raster of the digital surface model was used as data for the Spatial Analyst of ArcGIS in order to compute SR and deduce its impact and interaction via UHI.

4.6. NDVI calculation for vegetation analysis

Among the different spectral indices used in the field of vegetation studies is the Normalized Difference Vegetation Index (NDVI) [45]. NDVI is computed as a normalized ratio of the near-infrared (NIR) and red bands, which serve as critical pointers of plants' vitality and health. In this study, NDVI values were generated in ArcMap through the use of the Raster Calculator, using the following expression:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \tag{7}$$

Where $NDVI$ is the Normalized Difference Vegetation Index, NIR is the near-infrared band, and RED is the visible red band. For L7 data, NDVI = (Band 4 – Band 3) / (Band 4 + Band 3). For L8 data, NDVI = (Band 5 – Band 4) / (Band 5 + Band 4).

4.7. UTFVI calculations

To assess how UHI conditions influence the quality of life in cities, the Urban Thermal Field Variance Index (UTFVI) was applied as an ecological indicator [46]. This index helps quantify the extent to which heat radiation from UHI affects specific locations. In this study, UTFVI was computed using the following formula:

$$UTFVI = \frac{Ts - Tmean}{Ts} \tag{8}$$

Where $UTFVI$ is the urban thermal field's variance; Ts is the LST of a particular pixel in Kelvin. $Tmean$ is the average LST for the study area.

Table 2 provides an overview of the UTFVI measurements by summarizing the intensity of the estimated values. This is accomplished by measuring the strength of the UHI in the research region.

Table 2. UTFVI classes used to compute UHI in the study area.

S. No.	UHI phenomenon	UTFVI
1.	None	<0
2.	Weak	0–0.005
3.	Middle	0.005–0.010
4.	Strong	0.010–0.015
5.	Stronger	0.015–0.020
6.	Strongest	>0.020

4.8. OLS regression analysis for the evaluation of LST variation comparison

Using the OLS tool in ArcMap, an OLS regression was carried out to look at variations in LST. The input data contained feature classes with an ID field, the dependent variable (LST), and explanatory variables (NDVI, SR, SVF, and SA) [47]. The overall performance of the model was evaluated using the R2 and adjusted R2 scores. Furthermore, model nonstationarity and residual normality were evaluated using the Koenker and Jarque–Bera statistics, respectively [48]. OLS is a global linear regression model that calculates the relationship, under the assumption that it remains constant across space, between a dependent variable and one or more independent variables. The OLS formula is as follows:

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon \tag{9}$$

Where:

- Y: $n \times 1$ vector of dependent variable
- X: $n \times k$ matrix of independent variables (including intercept)
- β: $k \times 1$ vector of coefficients
- ε: $n \times 1$ vector of residuals

OLS estimation (closed-form solution):

$$\hat{\beta} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathrm{Y} \tag{10}$$

4.9. Computation with the GWR model

By enabling the development of numerous localized models that can reflect the variation in LST, the geographically weighted regression is typically used when the Koenker (BP) test shows a nonstationary process. It manages nonstationary dynamics appropriately and allows coefficients of influencing elements to change over space [49]. By determining various bandwidth distances for the parameters, three models were created. The crucial elements are:

1. LST (dependent variable).
2. SR, SA, SVF, NDVI (explanatory variable).
3. 30, 100, and 200 m (bandwidth distances).

GWR is a local regression model where the parameters are allowed to vary across geographic space. It captures spatial heterogeneity in relationships.

The GWR model at location i is as follows:

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)X_{ik} + \varepsilon_i \tag{11}$$

Where:

- (u_i, v_i): coordinates of location i;
- $\beta_k(u_i, v_i)$: location-specific coefficients;
- Each point i has a different set of coefficients based on its location.

4.10. NDBI calculation for identifying built-up areas

The NDBI can be computed using the following formula:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \tag{12}$$

Where $SWIR$ is the short-wave infrared [50], and NIR is near infrared.

4.11. Quantifying spatial clustering of LST with Moran's I

A Pearson correlation analysis between NDVI and LST for the years 2011 and 2021 was conducted in order to evaluate the link between green cover and LST across various LULC types. The four main LULC categories [cropland, water bodies, urban and built-up, and others (fallow/barren lands)] were examined independently. To ascertain the strength and direction of the association, a random sample of spatial points (n = 12–30) was taken for each class. The correlation coefficient (R), coefficient of determination (R^2), and significance level ($p \leq 0.01$) were then computed [51].

$$I = \frac{N}{W} \cdot \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{13}$$

Where I is the Moran's Index (measure of spatial autocorrelation). N is the total number of spatial units (e.g., pixels, zones, and districts). x_i, x_j are observed values at locations i and j. $\bar{x}$ is the Mean of the observed variable (e.g., mean LST). w_{ij} is the spatial weight between location i and j (often 1 if locations are neighbors, and 0 otherwise). W is the sum of all spatial weights: $\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$.

Interpretation of Moran's I:

$I > 0$: Positive spatial autocorrelation (clustering); $I \approx 0$: Random spatial pattern; $I < 0$: Negative spatial autocorrelation (dispersion).

Using Moran's I statistic, spatial autocorrelation analysis was performed to look at the spatial pattern and clustering of LST. Whether high or low LST values were diffused, geographically grouped, or randomly distributed over the research area was determined by this investigation. For both years, ArcGIS or comparable spatial statistical tools were used. Spatial clustering in the LST distribution was validated by statistically significant z-scores ($p < 0.01$) and high positive Moran's Index values.

5. Results and discussion

5.1. NDBI and LST analysis

Built-up intensity and surface temperature have a substantial spatial relationship, according to an examination of the LST and NDBI maps for Sonipat in 2021 (Figures 3 and 4). Densely populated areas like urban centers and industrial zones are indicated by red areas with high NDBI values. These regions, especially those with temperatures between 24.54 and 31.46 °C, closely match places with elevated LST. This alignment suggests that the UHI effect, a phenomenon whereby urbanized areas tend to retain more heat, contributes to greater surface temperatures. The cooling effect of natural land cover is highlighted by the lower LST values found in places with low NDBI values, which are usually vegetated or agricultural areas. In quickly developing regions like Sonipat, this link highlights the necessity for climate-resilient urban design and the effects of urban expansion on local microclimates.

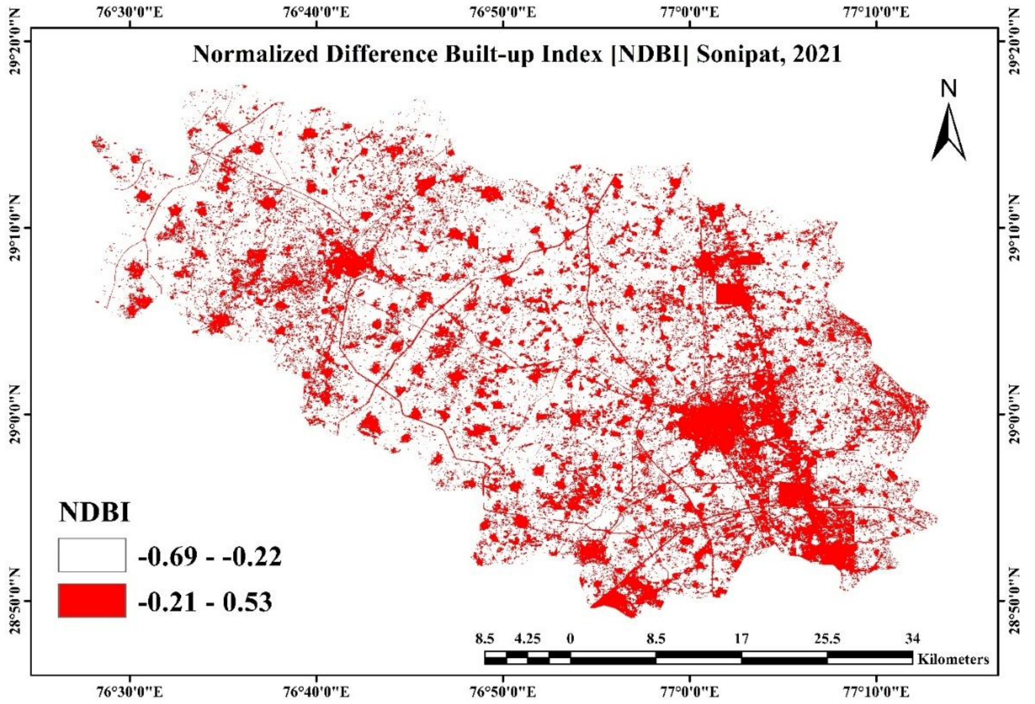


Figure 3. NDBI for 2021 in the Sonipat district.

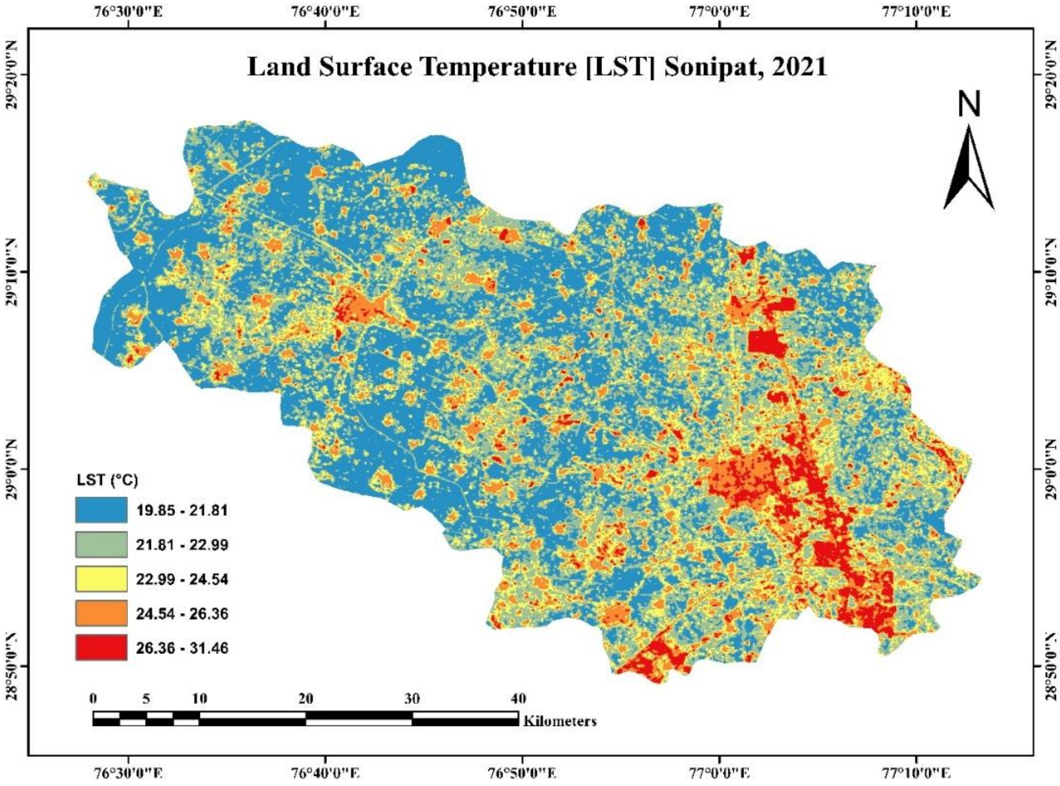


Figure 4. LST in February 2021.

5.2. UHI and UTFVI analysis

Using 2021 UHI and UTFVI maps (Figures 5 and 6), the Sonipat district's spatial analysis shows a notable overlap between ecologically stressed areas and high surface temperature zones. Urban heat islands are defined by the UHI map as areas with LST values higher than 24.54 °C. These areas are primarily found in the district's central, eastern, and southeast regions and signify significant anthropogenic heat retention. By classifying regions into six classifications, ranging from "Excellent" to "Worst", according to their thermal variance, the UTFVI map further measures thermal stress. The ecological evaluation indicates that roughly 27.14% of the land is in the "Excellent" category, meaning that there is no thermal stress (UTFVI < 0), while approximately 3.27% of the area is in the "Worst" ecological state (UTFVI > 0.020) (Table 3). The UTFVI map shows a distinct and spatially coherent pattern (Figure 6), with low UTFVI classes predominating in agricultural and outlying areas and high UTFVI values regularly clustered in dense urban and industrial pockets.

The UHI regions and high UTFVI zones closely align, indicating that regions experiencing extreme urban heat stress also have deteriorated ecological conditions. When taken as a whole, the UHI and UTFVI studies show the environmental cost of urbanization and the pressing need for thermal mitigation techniques, including reflective structures and increased green cover, as well as sustainable urban planning.

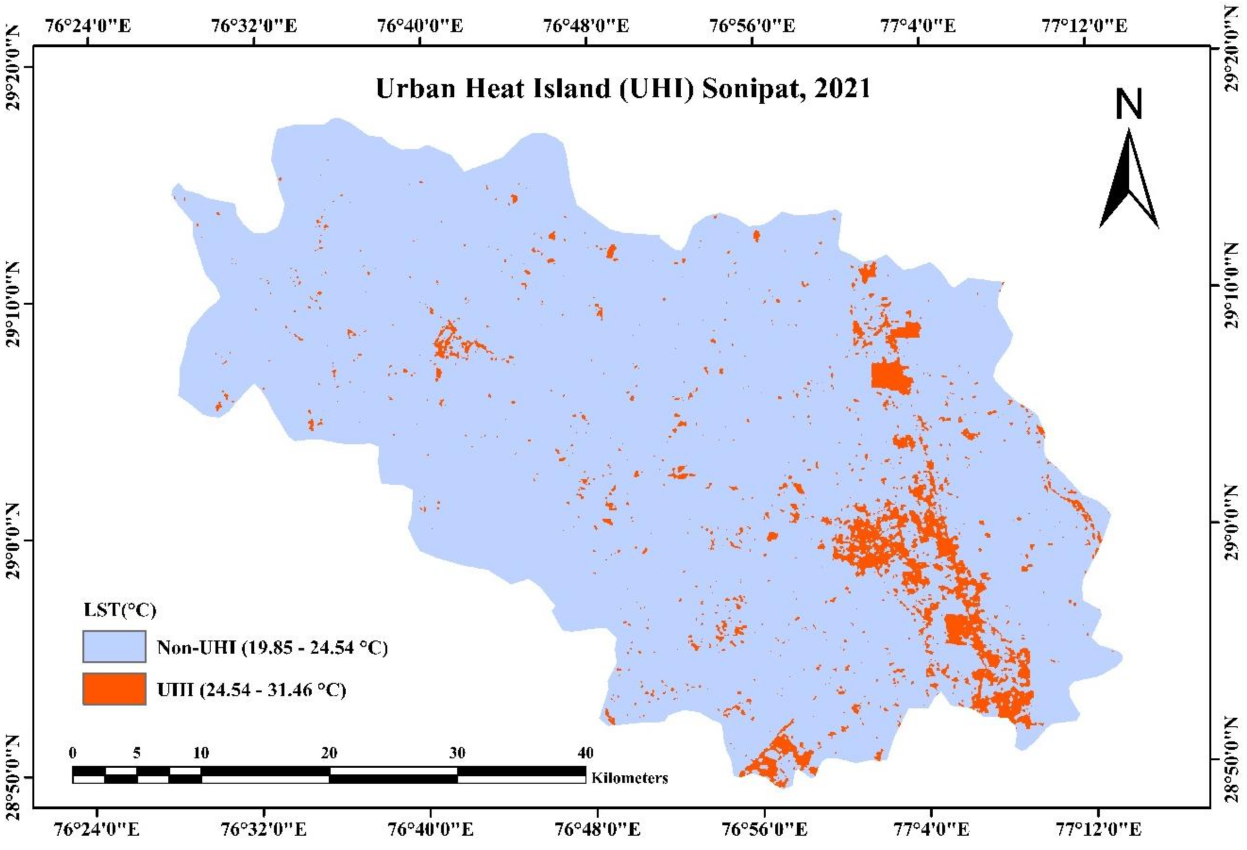


Figure 5. UHI for February 2021.

Table 3. UTFVI phenomenon in the study area.

UTFVI	UHI phenomenon	Ecological evaluation index	Area (km^2)	% of area
			Sonipat	
<0	None	Excellent	600.7302	27.14008%
0–0.005	Weak	Good	701.3331	31.68517%
0.005–0.010	Middle	Normal	398.9466	18.0238%
0.010–0.015	Strong	Bad	241.0713	10.89124%
0.015–0.020	Stronger	Worse	199.0404	8.992344%
>0.020	Strongest	Worst	72.3204	3.267326%

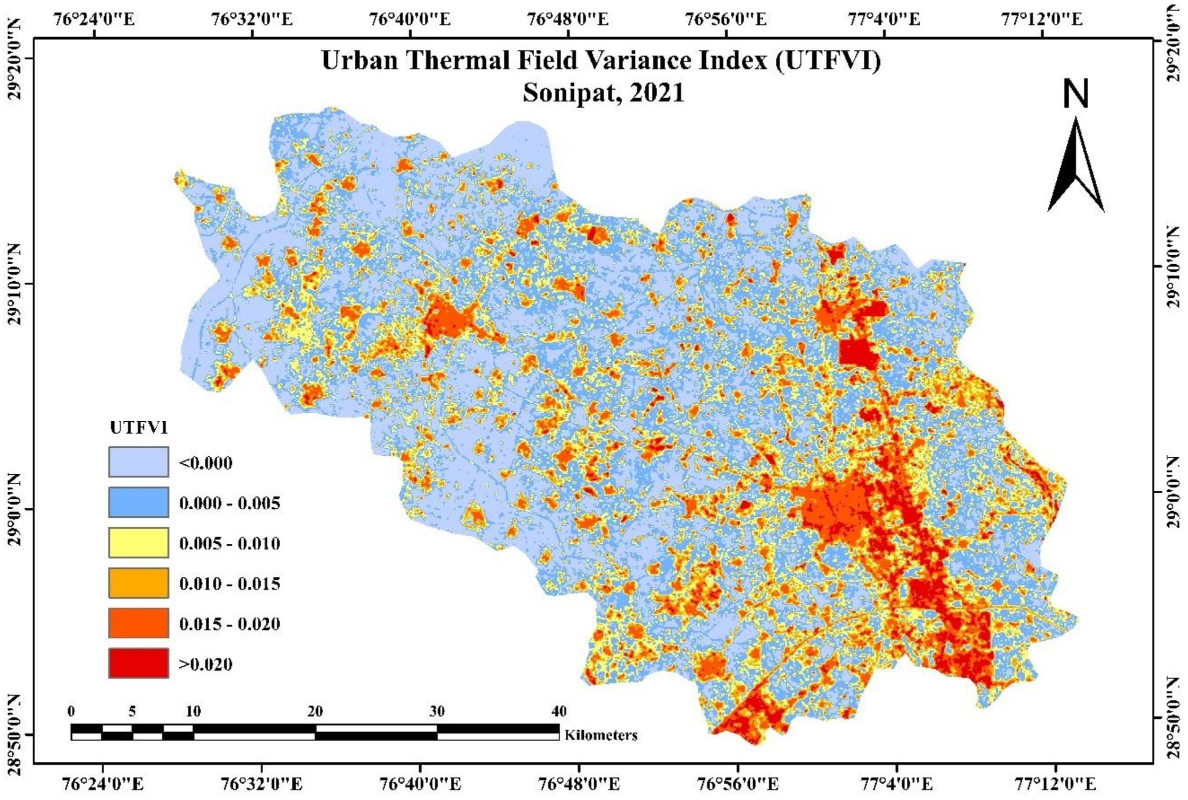


Figure 6. UTFVI distribution in the study area.

5.3. *Variations in NDVI and LST over time by land use type (2011–2021)*

The comparison of NDVI maps between 2011 and 2021, alongside the NDVI and LST values for various land use and land cover (LULC) categories, underscores a significant alteration in vegetation health and surface thermal dynamics in Sonipat throughout the decade. A reduction in NDVI values and an increase in LST were seen across all LULC types, signifying deteriorating vegetative conditions and exacerbating urban heat effects. Urban and built-up regions had a minor decline in NDVI from 0.128 (2011) to 0.115 (2021) (Figures 7 and 8), accompanied by a notable rise in LST from 23.12 to 27.37 °C, indicating heightened heat absorption attributable to urban expansion. Likewise, croplands underwent a significant decline in NDVI from 0.594 to 0.447, along with an increase in LST from 16.85 to 20.95 °C, indicating either vegetation stress or a reduction in agricultural activities. Notably, fallow or barren fields experienced a pronounced reduction in NDVI from 0.573 to 0.202 and a substantial increase in LST from 16.56 to 23.91 °C, signifying considerable land degradation (Table 4). Water bodies exhibited significant variance in NDVI, attributable to variations in water levels, and a rise in LST from 17.4 to 21.06 °C. These trends cumulatively underscore the detrimental effects of urbanization and land use alterations on the thermal and ecological integrity of the region. During the current study period, slightly increased LST conditions may inhibit vegetation productivity, suggesting that the cooling effects of vegetation are limited to specific seasons. Recent studies indicate that increased surface temperatures can diminish GPP, underscoring that urban heat island mitigation via vegetation is intricately associated with preserving optimal thermal conditions for sustained ecosystem productivity.

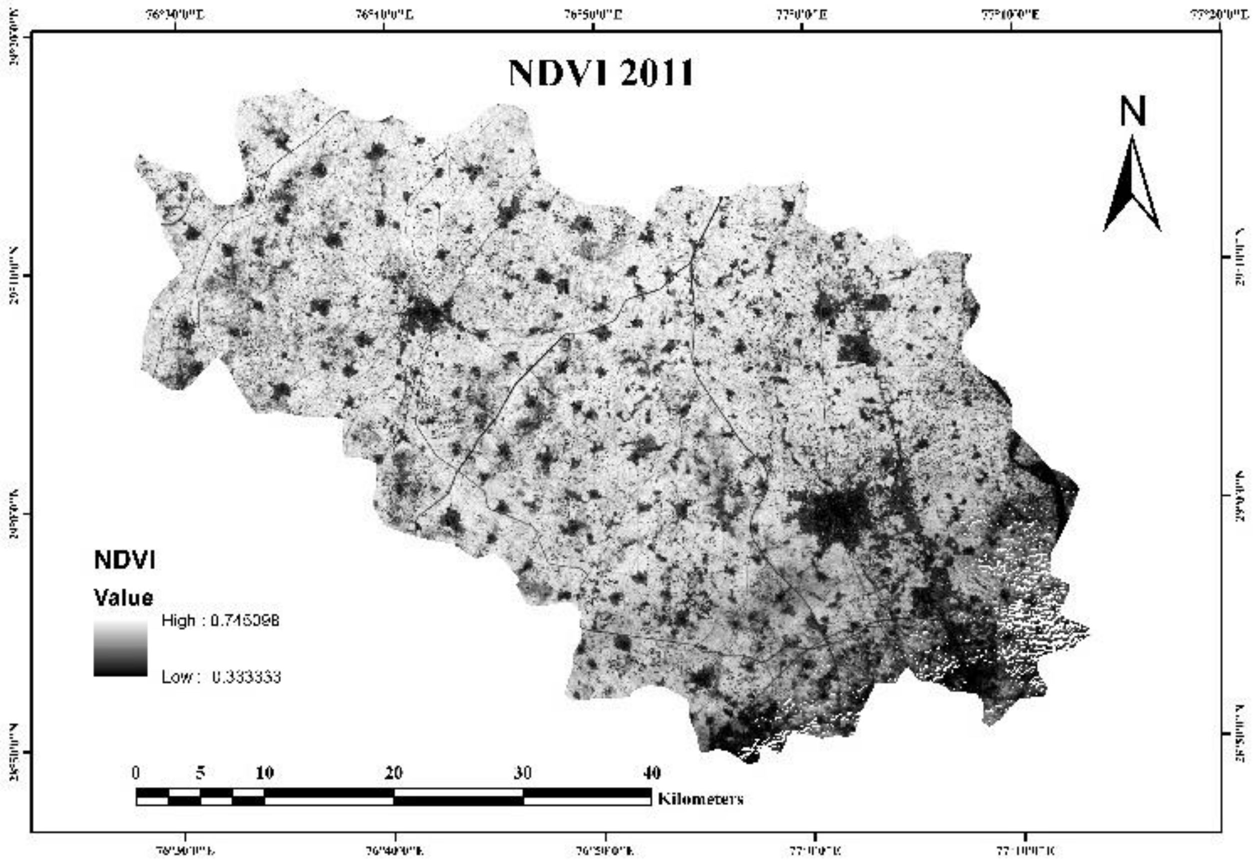


Figure 7. NDVI for the year 2011.

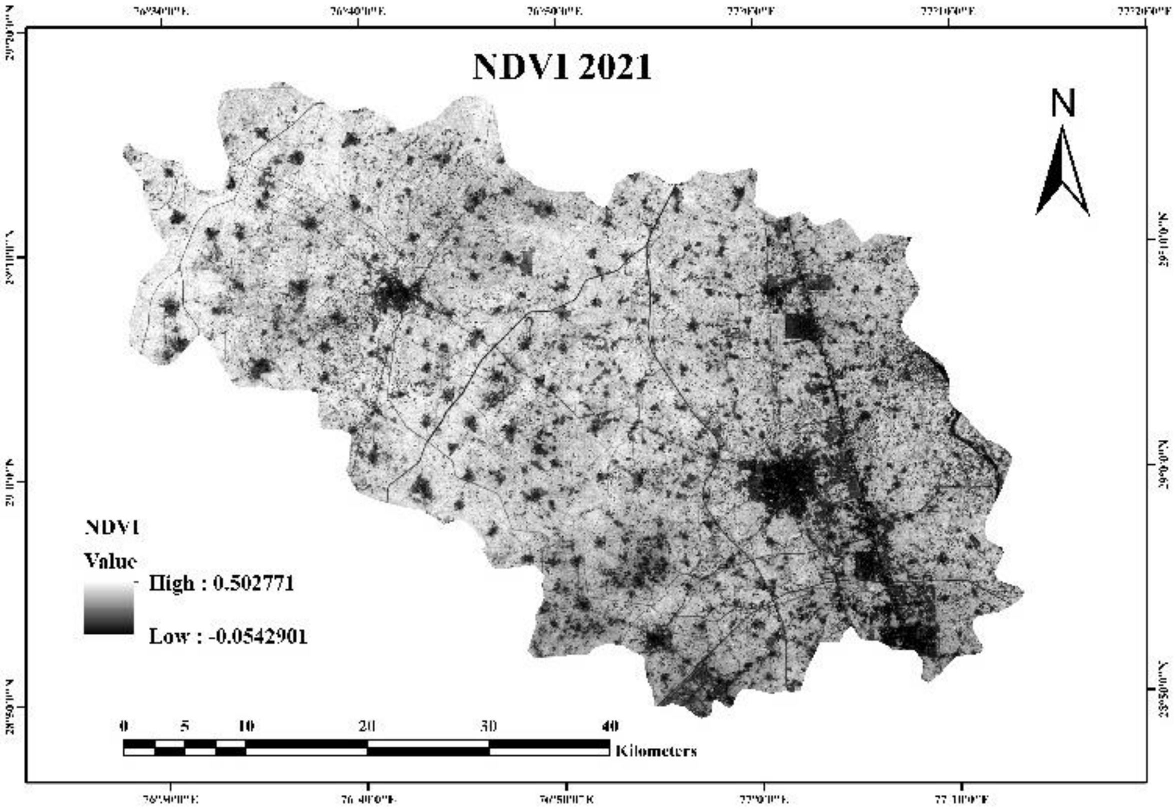


Figure 8. NDVI for the year 2021.

Table 4. Comparative NDVI and LST of 2011 and 2021 by land use category.

Land use	2011		2021	
	NDVI	LST (°C)	NDVI	LST (°C)
Urban and built-up	0.128	23.119	0.115	27.371
Water bodies	−0.203	17.4	0.03	21.062
Cropland	0.594	16.848	0.447	20.948
Others fallow/barren	0.573	16.563	0.202	23.908

5.4. Spatial autocorrelation of LST and vegetation-temperature relationship

The statistical analysis of Moran’s Index offers quantifiable evidence of geographic clustering in land surface temperature trends. In 2011, Moran’s Index was 0.4131, accompanied by a high z-score of 28.20 and a p-value of 0.00 (Table 5), signifying a robust and considerable spatial autocorrelation in the distribution of LST. By 2021, the Moran’s Index increased to 0.4969, accompanied by a z-score of 33.66, indicating a heightened spatial dependence in temperature values, possibly attributable to the proliferation of heat-retaining constructed environments. The autocorrelation findings validate the visual analysis of the NDVI and LST maps, indicating that urban clusters demonstrate elevated temperatures and diminished vegetation density. The inverse relationship between NDVI and LST across various LULC types further substantiates the ecological premise that dense vegetation aids in regulating and cooling surface temperatures (Table 6). This analysis emphasizes the imperative to maintain green cover in urban and rural areas to alleviate the impacts of increasing land surface temperatures and to uphold ecological equilibrium.

Table 5. LST autocorrelation statistics.

Year	Moran's Index	z-score	p-value
2011	0.413117	28.19545	0.00
2021	0.496888	33.66412	0.00

Table 6. Correlation between LST and NDVI for selected LULC types.

Year	Land use	n	R	R^2	Significance level
2011	Urban and built-up	20	−0.341	$R^2 = 0.1166$	0.01
	Water bodies	15	0.159	$R^2 = 0.0255$	0.01
	Cropland	30	−0.759	$R^2 = 0.5761$	0.01
	Others fallow/barren	12	−0.846	$R^2 = 0.7158$	0.01
2021	Urban and built-up	20	−0.490	$R^2 = 0.2404$	0.01
	Water bodies	15	0.537	$R^2 = 0.2884$	0.01
	Cropland	30	−0.598	$R^2 = 0.3585$	0.01
	Others fallow/barren	15	−0752	$R^2 = 0.5655$	0.01

5.5. Ordinary least squares regression (OLS)

OLS is a commonly used statistical method in LST studies to analyze the relationship between LST and various independent variables. OLS regression is a common method of linear modeling aiming at estimating the coefficients of an equation by minimizing squared differences between observed and estimated values. Applied in the field of LST studies, OLS is used for determining the effect of several explanatory variables on LST and quantifying these relationships. LST serves as a dependent variable in these models, while independent variables may include land cover, vegetation indices, climatic variables, topographic variables, and anthropogenic variables. The use of OLS regression provides a number of benefits. It first allows researchers to estimate the effect of each independent factor through regression coefficients, which indicate both the strength and direction of impact. Positive coefficients indicate that a variable increases LST, whereas negative coefficients indicate that the variable reduces it. Second, OLS includes statistical tests of significance to determine the variables that substantially describe LST variation. Third, after the model is developed and evaluated, it can be used, at least in principle, to predict LST values from available data, which is extremely valuable for mapping and surveillance in places where no direct measurement data is available. Finally, the use of OLS within a spatial framework sheds light on geographic patterns, hotspots, and cool places, and departures from typical LST trends at varying landscapes.

Figures 9 and 10 show histograms and scatter plots used to explore the distribution of variables and their relationships with LST. Histograms provide a general idea of the distribution of each explanatory variable: NDVI, solar radiation (SR), sky view factor (SVF), and surface albedo (SA). By examining the shape and the spread of these graphs, we can infer the distribution and spectrum of values belonging to each variable. The scatter plots here provide a counterpart by indicating how LST behaves against single explanatory variables. For SR, 2011 plots exhibit a strong positive correlation, so higher SR levels tend to correspond with increased LST. For 2021, the correlation remains positive, though again the trend looks more gradual. There is a good linear correlation between SA and LST for both years, indicative of higher albedo levels associated with rising LST in the area of study. The strength of these associations may be estimated from the slope of the regression line and the compactness of the data points around it. Sharper slopes and closer groupings indicate stronger associations. Overall, the results highlight the role of SR and SA as important sources of variation in the study area. These outcomes provide a good foundation for the development of policies in land use planning, city development, and climate adaptation.

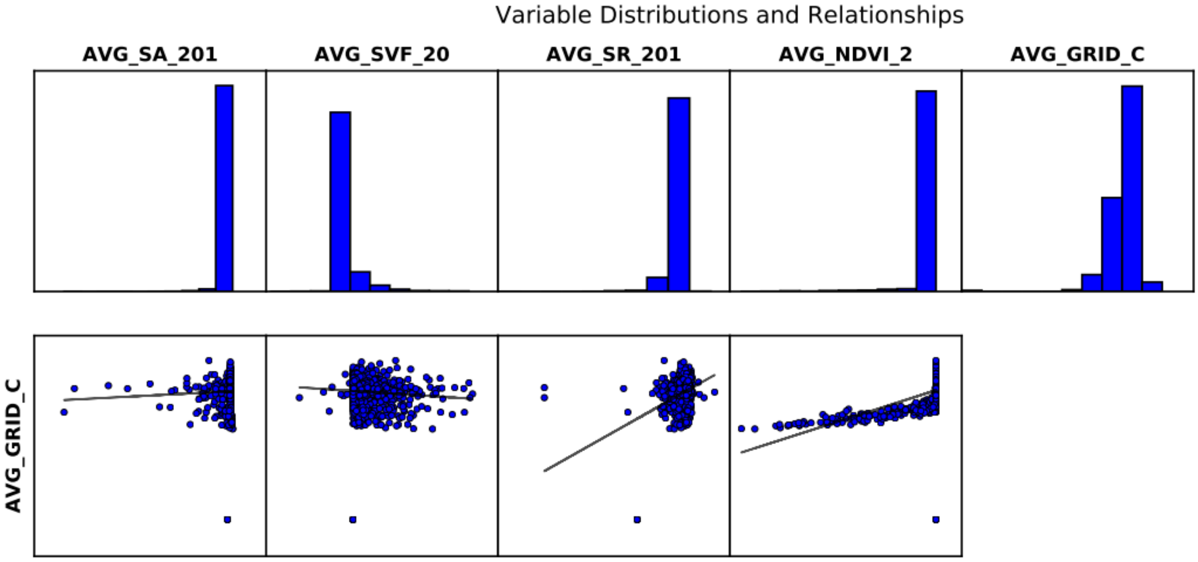


Figure 9. Scatter plot diagrams for LST vs. SR, SVF, SA, and NDVI. The lines show the linear relationship between the variables listed in 2011.

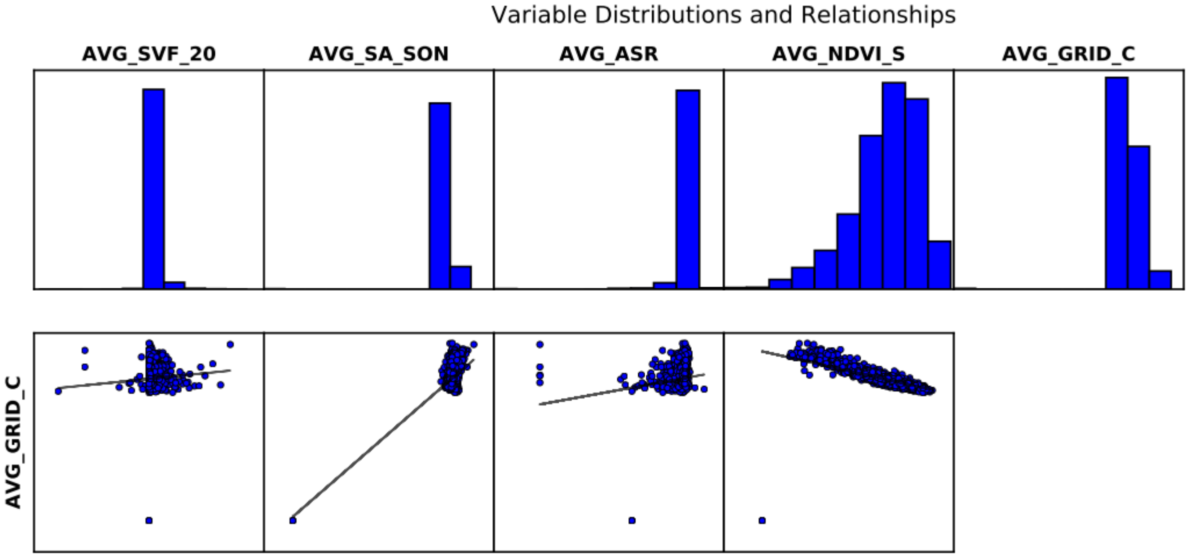


Figure 10. Scatter plot diagrams for LST vs. SR, SVF, SA, and NDVI. The lines show the linear relationship between the variables listed in 2021.

Table 7. Summary of ordinary least squares results: model variables for LST regression model in the Sonipat district for 2011 and 2021.

Variable 2011	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	11.971786	0.241194	49.635599	0.000000*	2.018196	5.931925	0.000000*	-
SVF	−0.000642	0.00015	−4.26506	0.000025*	0.000485	−1.321946	0.186323	1.281757
SR	17.211468	1.383573	12.439871	0.000000*	4.906801	3.507676	0.000476*	1.641406
NDVI	0.001034	0.000044	23.315522	0.000000*	0.000376	2.749021	0.006023*	1.679933
SA	0.001129	0.000038	29.50693	0.000000*	0.000050	22.664657	0.000000*	1.029003
Variable 2021	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	1.831558	0.183361	9.988784	0.000000*	0.713751	2.566101	0.010337*	-
SVF	−0.378056	0.301198	−1.255173	0.209544	0.503286	−0.751174	0.452612	1.121539
SA	0.002318	0.000017	135.116060	0.000000*	0.000063	36.613383	0.000000*	1.085811
SR	0.000114	0.000010	11.139859	0.000000*	0.000025	4.555660	0.000008*	1.170291
NDVI	−17.180213	0.135295	−126.983566	0.000000*	0.192258	−89.360254	0.000000*	1.043764

The adjusted R-squared is a statistical measure that represents the proportion of the variability in the dependent variable (LST) that is explained by the independent variables in the regression model. In this case, the adjusted R-squared was 0.37 in 2011, indicating that the model explains 37% of the variability in LST in that year. However, the adjusted R-squared value increased by 0.55 in 2021, indicating that the model explains 92% of the variability in LST in 2021. This suggests that the model improved significantly in capturing and explaining the variations in LST over time. Table 7 provides information about the coefficients of the explanatory variables (NDVI, SVF) and their relationships with LST. In 2011, the coefficient for the SVF variable was negative, indicating a negative correlation between SVF and LST. Similarly, in 2021, both the coefficients for NDVI and SVF were negative, suggesting negative correlations between these variables and LST. The negative correlations imply that higher values of NDVI and SVF are associated with lower LST values in the study area. The presence of asterisks in the table indicates that certain coefficients are statistically significant, except for the SVF robust probability values in 2021. Statistical significance suggests that the relationship between the respective variables and LST is unlikely to have occurred by chance and is likely to be meaningful. Although no clear linear relationships are observed in the data, the results highlight the importance of other variables in the modeling of LST. It also presents additional statistical information. The variance inflation factor (VIF) values, all below 7.5, indicate that there is no significant multicollinearity issue among the independent variables. The Koenker (BP) and Jarque–Bera statistics are significant in both years (Table 8). The significance of the Koenker (BP) statistic suggests the presence of nonstationarity or heteroskedasticity in the modeled relationships. The Jarque–Bera significance indicates that the model predictions may be biased, and residuals may not have a normal distribution. Overall, these findings highlight the performance and limitations of the OLS regression model in explaining LST variations (Figures 11 and 12). It suggests that while the model captures a significant portion of the variability in LST, there may be other models at play that influence LST, and further investigations are warranted. An adaptive kernel was used to choose the GWR bandwidth, which was then optimized using AICc-based cross-validation. Several bandwidths were assessed, and the one that minimized AICc was selected. This method takes spatial nonstationarity into consideration while guaranteeing an ideal balance between model fit and complexity.

Table 8. Ordinary least squares diagnostics for the LST regression model in the Sonipat district for 2011 and 2021.

Input features 2011	Aver_point	Dependent variable	AVG_GRID_C
Number of observations	2380	Akaike's information criterion (AICc) [d]	7726.101106
Multiple R-squared [d]	0.376	Adjusted R-squared [d]	0.374949
Joint F-statistic [e]	357.772942	Prob(>F), (4,2375) degrees of freedom	0.000000*
Joint Wald statistic [e]	608.971604	Prob(>chi-squared), (4) degrees of freedom	0.000000*
Koenker (BP) statistic [f]	1481.409815	Prob(>chi-squared), (4) degrees of freedom	0.000000*
Jarque–Bera statistic [g]	150650.0081	Prob(>chi-squared), (2) degrees of freedom	0.000000*
Input features 2021	Aver_point	Dependent variable	AVG_GRID_C
Number of observations	2387	Akaike's information criterion (AICc) [d]	2872.966517
Multiple R-squared [d]	0.928418	Adjusted R-squared [d]	0.928298
Joint F-statistic [e]	7723.651441	Prob(>F), (4,2382) degrees of freedom	0.000000*
Joint Wald statistic [e]	9950.730700	Prob(>chi-squared), (4) degrees of freedom	0.000000*
Koenker (BP) statistic [f]	439.906425	Prob(>chi-squared), (4) degrees of freedom	0.000000*
Jarque–Bera statistic [g]	439.906425	Prob(>chi-squared), (2) degrees of freedom	0.000000*

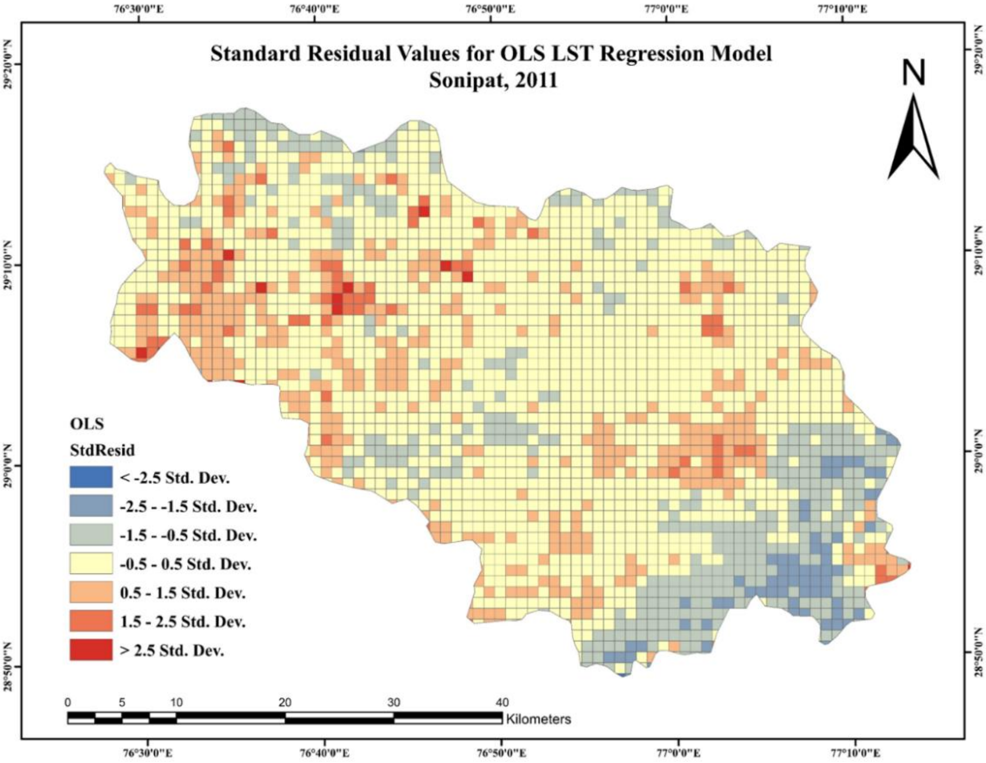


Figure 11. Ordinary least squares for the LST regression model in the Sonipat district for 2011.

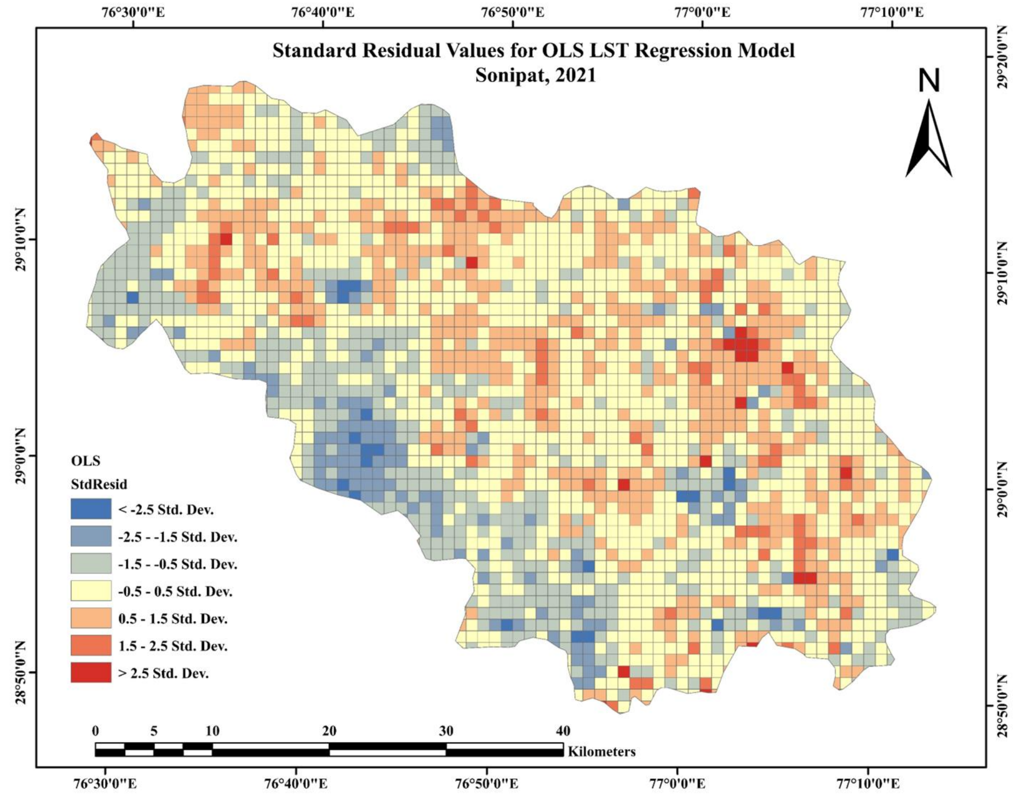
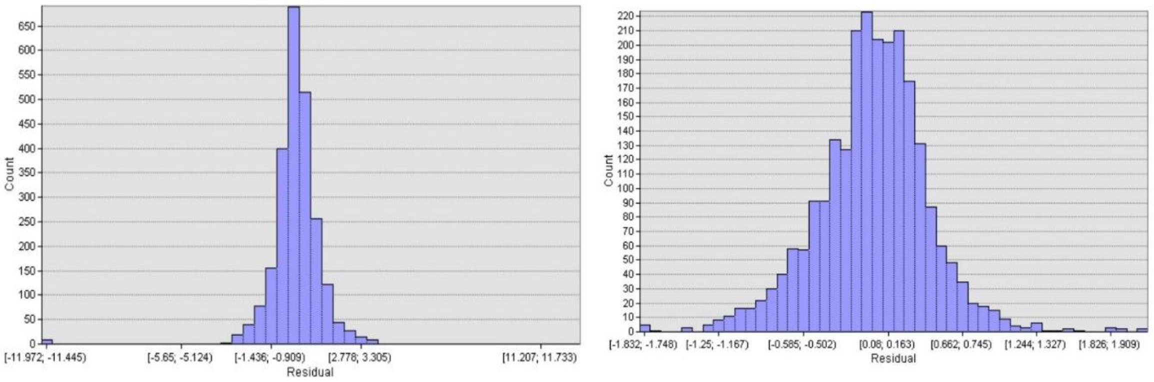


Figure 12. Ordinary least squares for the LST regression model in the Sonipat district for 2021.

The implementation of OLS models has been effective in capturing the disparities observed in Figures 11 and 12 for the years 2011 and 2021. These models have been able to predict the locations where higher or lower LST values are likely to occur in the study area. However, it is important to note that there are large residuals in the OLS regressions, as depicted in the OLS figures, which deviate from a normal distribution. This indicates that the OLS model fails to fully meet the assumption of normality (Figure 13). Thus, the OLS approach may not provide a detailed description or explanation of the whole data set. To mitigate this problem, it is advisable to consider a different modeling strategy that yields residual plots resembling a bell-shaped curve, meaning better correspondence with the standard distribution. One approach is a geographic regression approach, such as geographically weighted regression. GWR takes into consideration spatial variation and nonstationarity within the data, facilitates localized interrelationships, and describes the spatial variation inherent in the interrelationships among variables. By using GWR, it will be possible to better capture the spatial attributes of data and get more accurate estimations of LST over the area of study. GWR is capable of revealing the conditions under which relationships between LST and explanatory variables vary spatially, thus giving a better understanding of the variables determining LST distribution at varying places.



(a) (b)

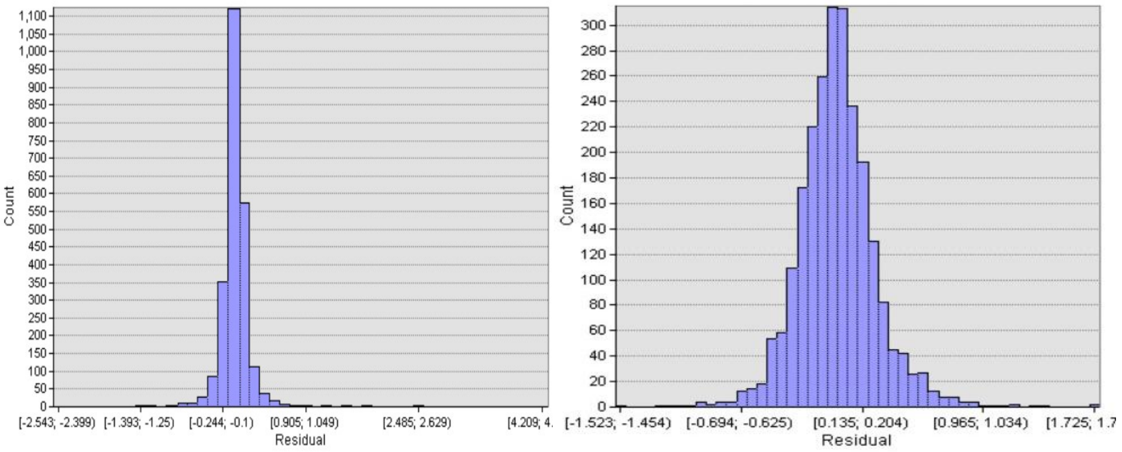
Figure 13. Histograms of OLS regression residuals for (a) 2011 and (b) 2021.

5.6. Geographically weighted regression

Geographically weighted regression (GWR) is a spatial regression method created to capture data heterogeneity and nonstationarity. Whereas ordinary least squares (OLS) assumes a constant relationship between variables throughout the study area, GWR considers the possibility of differing relationships based on geographic space. It operates by assigning weights to observations according to the distance from the focal area in a manner where nearer observations have more influence on the regression analysis than observations farther away. Such a system of weights enables the model to capture localized differences and produce regression coefficients that vary spatially. In practical applications, GWR generates a series of coefficients that become place-specific, so the connection between LST and explanatory variables, including NDVI, SR, SVF, and SA, does not always remain the same but rather differs depending on differing geographical conditions. Through the use of spatial data, GWR gains a better understanding of variables influencing LST at varying places in the study area. An important parameter of GWR is the bandwidth, which determines the spatial resolution of the analysis. Bandwidth selection decides the number of neighboring observations included in each of the local regressions. A smaller bandwidth will capture detailed fine-scale variation, whereas a large bandwidth smooths more of the patterns but may miss localized details. Choosing a good value thus becomes significant in order to trade off accuracy and overfitting. Outputs of a GWR model most commonly include local regression coefficients, diagnostic statistics, and residual patterns, through which model performance can be evaluated. Such outputs reveal points of variable associations getting stronger or weaker and provide informative indications of processes spatially varying and operating on LST.

In the GWR model, the nonstationarity of the variables is dealt with through the simulation of the OLS model. After examining the statistical results of the OLS regression equation and grasping its weakness as a continuation of capturing spatial nonstationarity, the application of the GWR model follows.

The GWR is an appropriate approach for dealing with the nonstationarity of the variables. It accounts for observed spatial variation in the relationships between the dependent variable (LST) and the independent variables. It is then, through the running of the GWR model, possible to generate maps of specific parameters, which attempt to provide the best LST model. Within the GWR model, three maps are created in ArcMap through the application of the GWR tool. The maps contain the spatial connections among LST and explanatory variables. Accounting for spatial variations, the GWR model is able to capture the nuances of the connections between variables that are specific to varying places in the study region. A significant advantage of the GWR model over the OLS regression is that the distribution of the GWR residuals is standard, as shown in Figure 14. This demonstrates that the GWR model provides superior data fit and rectifies the problem of non-normal errors of the OLS regression. With the application of the GWR model, it is possible to get a precise idea of the spatial relationships and interactions between LST and explanatory variables. GWR takes into account the spatial environment and enables localized modeling, which provides valuable data on the controlling elements of LST deviations over the study region.



(a) (b)

Figure 14. Histograms of GWR regression residuals for (a) 2011 and (b) 2021.

In GWR models, we have relatively small residuals, since these models provide a better explanation for the transformation of LST vis-à-vis the OLS regression model. Included in the report generated by ArcMap are GWR model outputs, as shown in Table 9. Among the key aspects of GWR is the calculation of the bandwidth parameter to determine the number of neighboring entities included in the local regression estimation. GWR allows selecting different bandwidth values for investigating the spatial influence at different scales. Three of the GWR models used bandwidth parameters of 30, 100, and 200. All GWR models accounted for a precise number of neighboring sites delineated by the adopted bandwidth. This allows for a more regionally specific analysis, taking into account the geographic environment and regional differences in the relationships between LST and explanatory variables. With variable bandwidth parameters, the GWR models explore spatial LST variation and draw out knowledge of the localized variables affecting LST patterns. The selection of the bandwidth value depends on the particular research question and the spatial scale of focus. The findings in Table 9 summarize the models of GWR and the respective coefficients and statistical indicators. These results aid in understanding spatially distributed LST and interrelationships with explanatory variables, containing valuable information supporting subsequent investigation and decision-making processes.

Table 9. GWR diagnostics for LST models in 2011 and 2021.

2011	GWR 30	GWR 100	GWR 200
Bandwidth parameter	30 neighbors	100 neighbors	200 neighbors
Residual squares	78.956446	352.665805	549.845759
Sigma	1.079962	1.254414	1.23713
AICc	401.016059	918.101399	1304.533742
R2	0.957685	0.851819	0.779955
2021	GWR 30	GWR 100	GWR 200
Bandwidth parameter	30 neighbors	100 neighbors	200 neighbors
Residual squares	3.743411	7.661439	5.435638
Sigma	0.604822	0.51877	0.51307
AICc	59.643691	66.823433	45.997353
R2	0.996767	0.993954	0.995465

After taking local relationships into account, the unexplained variance is represented by residual squares in the context of the GWR results (Table 9); lower values indicate a better local model fit. Smaller sigma values imply more accurate geographic estimates, and sigma represents the standard deviation of residuals and the level of local forecast uncertainty. A lower AICc indicates that GWR offers a better balance between goodness-of-fit and model parsimony. AICc is used to compare model performance while penalizing complexity. In GWR, the model's improvement over OLS underscores the significance of geographical nonstationarity in elucidating UHI drivers. R2 is a measure of the proportion of variation explained by the model.

In regression, the total errors' squares are termed the residual sum of squares (RSS). Specifically, it calculates the distance between the estimated and the actual values. A smaller RSS value relates to an increase in the congruence between the model and the data; Figures 15, 16, 17, 18, 19, and 20 document these increases. Among the three GWR models, the minimum RSS was obtained for GWR 30 (78.95) in 2011. In 2021, RSS was 3.74, indicating that this particular model outperformed the other two models. The standard deviation of the residuals, called Sigma, provides a measurement of the typical discrepancy between the predicted and actual values. For GWR 30, the Sigma value was 1.079962 in 2011 and 0.60 in 2021, showing an average deviation of 1.079 and 0.604 °C between the estimated and the measured LST in the model. If the residuals converge toward the observed values, it indicates a better prediction accuracy. To assess the model-related performance, the adjusted Akaike information criterion (AICc) acts as an evaluative tool. The AICc evaluates the model's quality, regarding both goodness of fit and the number of estimating parameters. Lower AICc values indicate a better fit of the model to the data. In this case, GWR 30 has the lowest AICc, suggesting a better fit compared to models with larger bandwidths. The coefficient of determination (R-squared) measures how well the model represents the real values, with values ranging from 0.0 to 1.0. A higher R-squared indicates a better representation of the real values. In both years, the R-squared is highest for GWR 30, implying a better overall fit compared to the other models. Overall, the GWR 30 model demonstrates superior performance in terms of lower residual squares, a closer agreement between predicted and observed values (Sigma), lower AICc values, higher R-squared, and spatial patterns of residuals indicating positive influences in certain areas.

The GWR residual maps for 2011 show that the size of a neighborhood has a clear effect on how well a model works and how stable it is over time (Figures 15, 16, and 17). The 30-neighbor model shows that most of Sonipat has near-zero residuals, which means that the model fits well in that area. However, there are some areas with high positive and negative residuals, especially in the east and south, which shows that the model does not fully capture localized urban heterogeneity at very small bandwidths. When the neighborhood size is increased to 100 neighbors, the residual pattern becomes more balanced and spatially coherent, with a clear drop in extreme residual values. This suggests that the model is more stable while still keeping relevant local fluctuations. The 200-neighbor model has the most even distribution of residuals, with very few extremes. This means that it is less likely to overfit, but it does lose some fine-scale spatial sensitivity in densely populated urban and industrial areas.

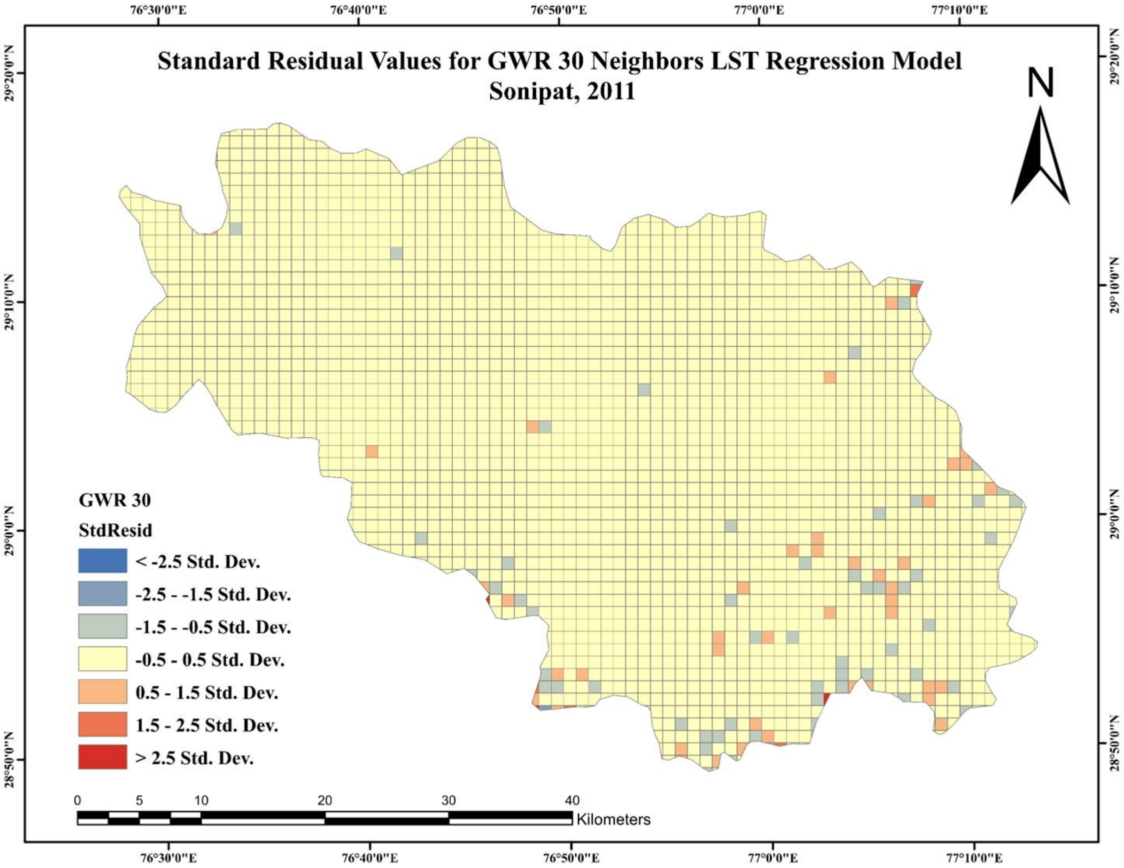


Figure 15. GWR 30 neighbors LST regression model in the Sonipat district for 2011.

The residual patterns from 2021 show that LST dynamics became more spatially complex because cities are growing faster (Figures 18, 19, and 20). The 30-neighbor model shows more variability in the residuals than it did in 2011. There are clear positive residual clusters in the eastern and southeastern areas of Sonipat. This shows how hard it is to represent quickly changing urban thermal conditions on very small scales. When the neighborhood size is extended to 100 neighbors, the residuals become more equally spread out, and extreme deviations are limited to certain areas. This shows that the model is better at finding local nonstationarity while also reducing random noise. The 200-neighbor model smooths out residual surfaces much more and significantly cuts down on extreme results. This shows that the model is very powerful generally, but localized UHI effects may be partially averaged out.

The comparison between different years and neighborhood sizes shows that increasing bandwidth consistently lowers residual clustering and uncertainty, making the model more reliable. Smaller neighborhoods capture fine-scale temperature variability, but they are more vulnerable to local noise. Larger neighborhoods, on the other hand, improve generalization but lose some local information. The findings indicate that a modest bandwidth, especially the 100-neighbor GWR model, offers an optimal balance between spatial sensitivity and model stability for the decadal LST and UHI study in the Sonipat district.

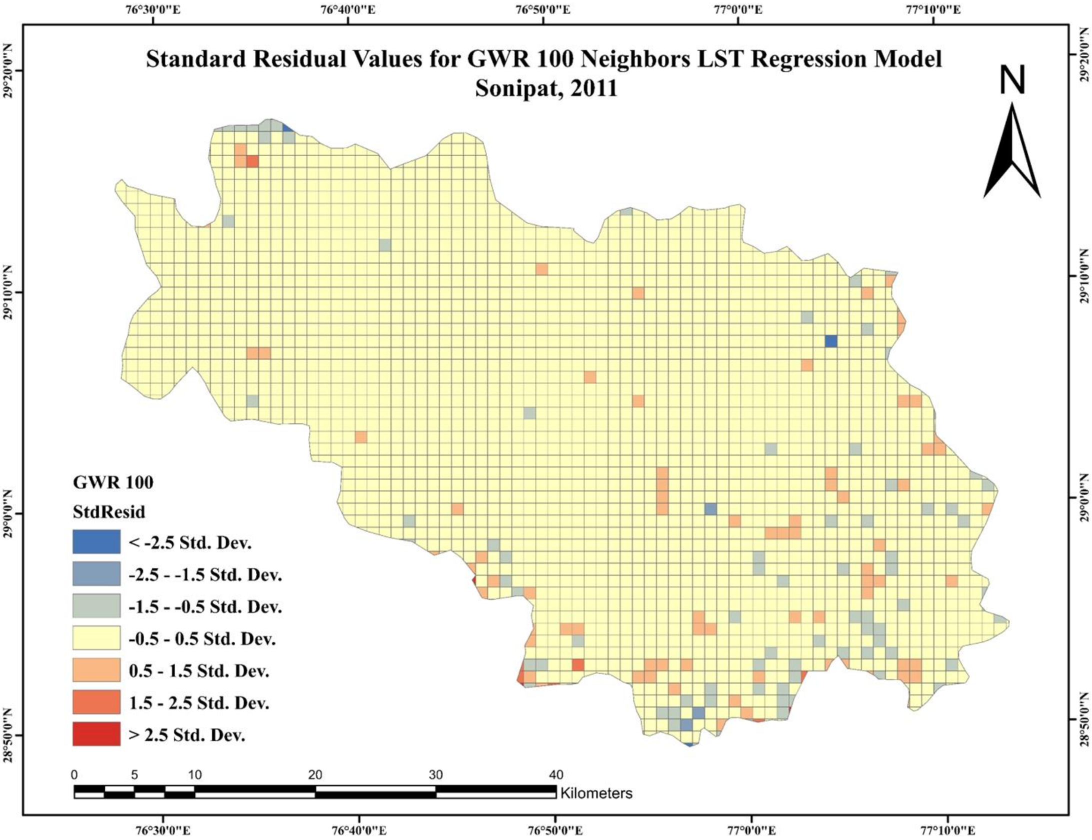


Figure 16. GWR 100 neighbors LST regression model in the Sonipat district for 2011.

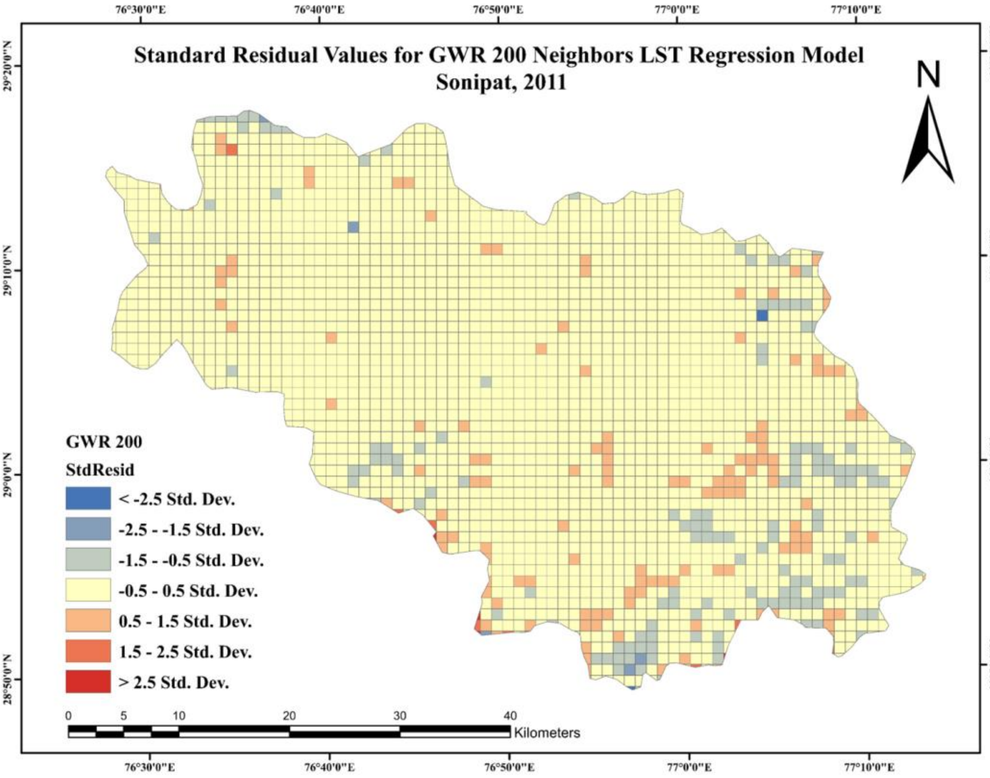


Figure 17. GWR 200 neighbors LST regression model in the Sonipat district for 2011.

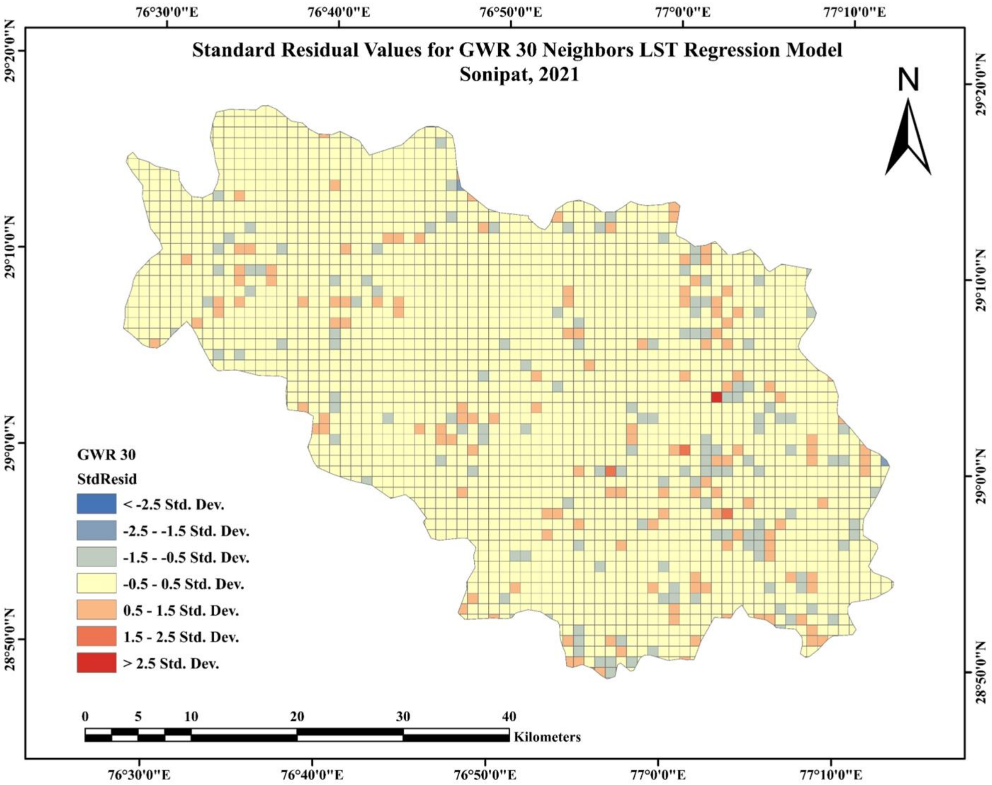


Figure 18. GWR 30 neighbors LST regression model in the Sonipat district for 2021.

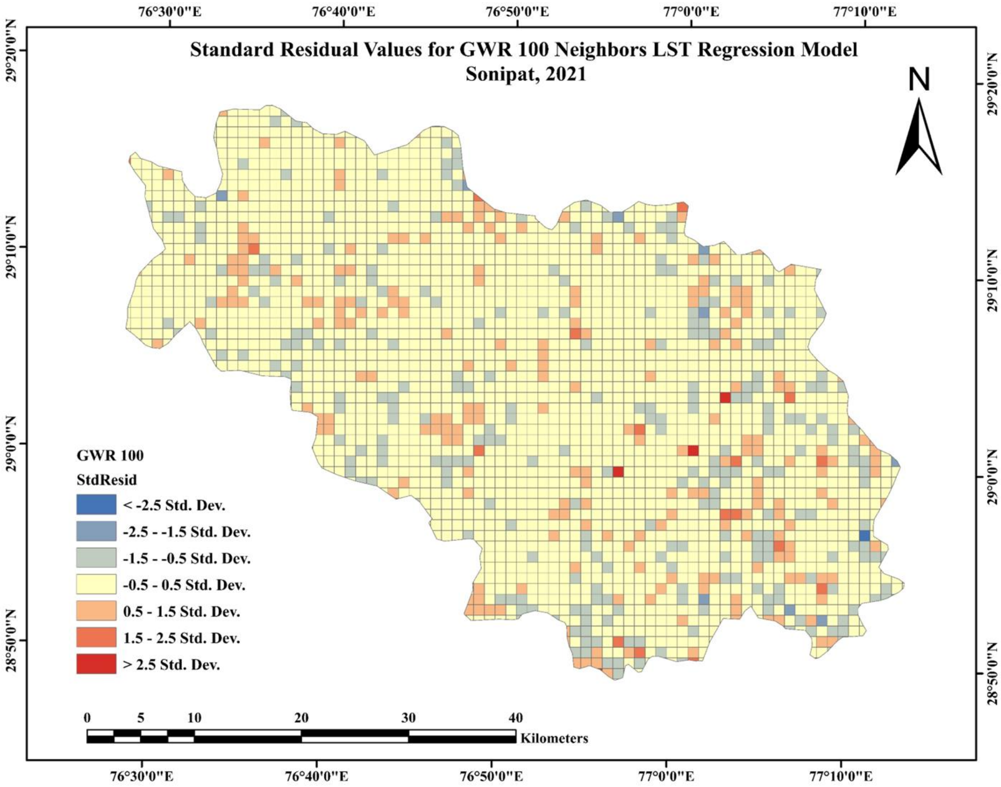


Figure 19. GWR 100 neighbors LST regression model in the Sonipat district for 2021.

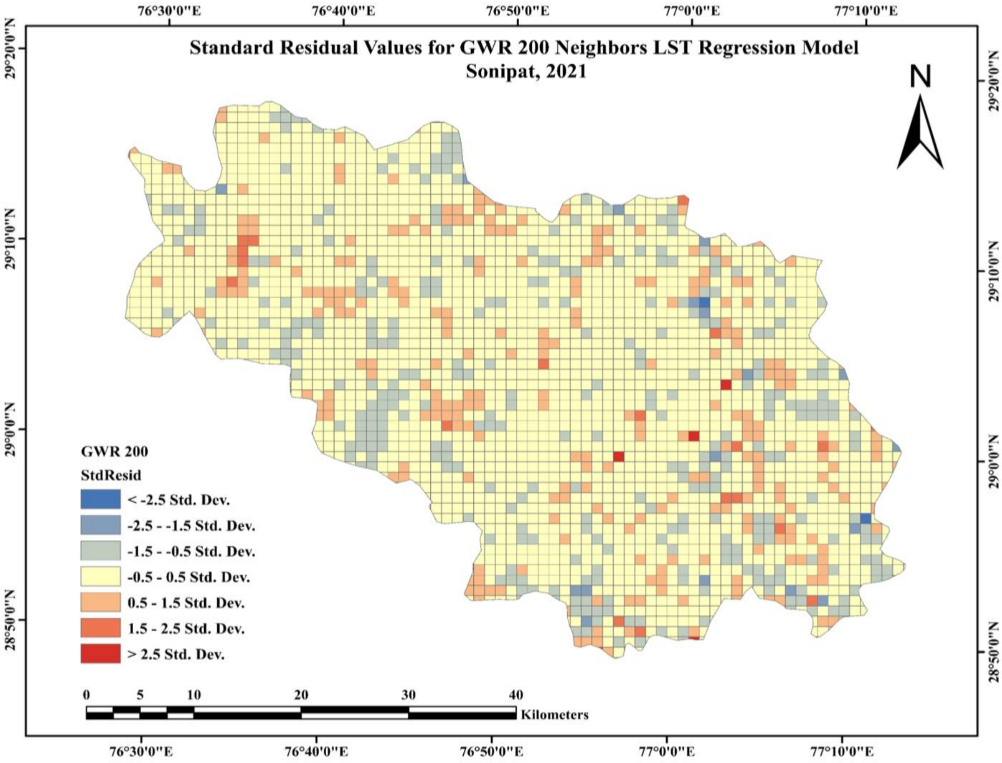


Figure 20. GWR 200 neighbors LST regression model in the Sonipat district for 2021.

GWR performed better than OLS because it takes geographical nonstationarity into account. It also implicitly incorporates multiscale variability by choosing a bandwidth that balances local and global effects. Sensitivity analysis employing various neighborhood sizes illustrates that UHI drivers function across many spatial scales, aligning with multiscale regression methodologies documented in the literature [52]. The high UHI and positive residual zones show where urban design changes could be made to reduce heat stress. These changes could include adding more green spaces, planting trees along the road, using cool roofing materials, and using reflective pavements. These results can help with planning strategies that are specific to a certain area by helping with zoning rules, making the best use of property, and building infrastructure that can withstand climate change in regions of Sonipat that are growing quickly.

6. Conclusions

In the study region, UHI was evaluated, and the LST effect was modeled using a methodical methodology. L7 and L8 imagery provided the LST, NDVI, and SA layers, while ASTER DEM provided the SR data. The investigation included key parameters—SR, SVF, SA, and NDVI. LST values were higher in the northeastern and southeastern parts of the district. Places with crops, green cover, and water bodies consistently showed lower temperatures. The highest LST values were recorded in areas with no vegetation or green cover or very little vegetation, such as concrete areas, urban, built-up, and barren terrain. In line with the UHI effect, the spatial analysis employing NDBI and LST maps verified that surface temperatures were greater in built-up and urban areas. While areas with greater NDVI values, such as croplands and water bodies, were linked to lower surface temperatures, highlighting the cooling effect of vegetation, areas with high NDBI values matched zones of elevated LST. Particularly in agriculture and barren land classes, where R^2 values were as high as 0.71 in 2011 and 0.56 in 2021, Pearson's correlation demonstrated a strong negative link between NDVI and LST.

These results show that surface heat is greatly modulated by plant cover, and its reduction over time directly helps to raise LST. Furthermore, suggesting increasing geographic clustering of high-temperature zones, presumably driven by increased urbanization, Moran's I statistic indicated a jump from 0.414 in 2011 to 0.497 in 2021.

Moreover, the OLS regression models provide a closer understanding of the factors affecting LST. Using the same criteria, NDVI, SVF, SR, and SA explained 37.6% of the variability in LST in 2011; this considerably increased to 92.8% in 2021. This reflects not just a more predictable thermal pattern in the urban setting but also the increasing predominance of land surface features on LST, resulting from fast land cover change. Apart from OLS, GWR was used to improve spatial heterogeneity in the link between LST and its affecting elements. Especially in 2021, the GWR model exceeded the global OLS model. With an R^2 of 0.96 and the lowest AICc (401.02), the model with a 30-neighbor bandwidth performed the best for 2011, therefore suggesting a strong localized explanation of LST fluctuation. GWR results in 2021 were even more striking; R^2 values across all bandwidths exceeded 0.99. At 200 neighbors, the best model fit was obtained; the AICc was lowest (45.99), and R^2 reached 0.995, stressing the improved spatial predictability of LST in recent times. Furthermore, the decrease in residual squares and sigma across years reflects increased model quality and lower spatial error. These findings underline the need for spatially adaptive modeling techniques for urban heat analysis since they indicate that the link between LST and variables such as NDVI, SA, SR, and SVF is not spatially constant and fluctuates across the terrain.

Author contributions

Conceptualization, D. and M.K.; methodology, D., M.K. and V.N.M.; software, D., R.V.K.S.; validation, M.K., and V.N.M.; formal analysis, D., R.V.K.S., and M.K.; investigation, D., and M.K.; data curation, D., and R.V.K.S., visualization, V.N.M.; writing—original draft preparation, D., and M.K.; writing—review and editing, V.N.M., and M.K. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors gratefully acknowledge the support and resources provided by the Amity University Uttar Pradesh during the course of this research. The study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors also thank colleagues and reviewers for their valuable suggestions that helped improve the quality and clarity of the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Rai CK, Goyal SK, Raj K, et al. (2025) Impact of Urbanization on Peri-Urban Agriculture. *Indian Agriculture: Challenges, Priorities and Solutions*, Singapore: Springer Nature Singapore, 105–119. https://doi.org/10.1007/978-981-96-5273-0_6
2. Anestis G, Stathakis D (2024) Urbanization trends from global to the local scale, *Geographical Information Science*, Elsevier, 357–375. https://doi.org/10.1016/B978-0-443-13605-4.00010-2
3. Yangtianzheng Z, Ying G (2024) Spatial patterns and trends of inter-city population mobility in China—Based on Baidu migration big data. *Cities* 151: 105124. https://doi.org/10.1016/j.cities.2024.105124

4. Han Q, Kumar R, Kumar A (2024) Climate change and human migration: Perspectives for environmentally sustainable societies. *J Geochem Explor* 256: 107352. https://doi.org/10.1016/j.gexplo.2023.107352
5. Abbas W (2024) Urban climate of the middle east and north Africa region: a comprehensive anatomization, *Hydroclimatic Extremes in the Middle East and North Africa*, Elsevier, 177–228. https://doi.org/10.1016/B978-0-12-824130-1.00004-7
6. Mangal S, Kumar D, Dhupper R, et al. (2024) Identifying influential climatic factors for urban risk studies in rapidly urbanizing Region. *Comput Urban Sci* 4: 9. https://doi.org/10.1007/s43762-024-00121-5
7. Panda J, Mukherjee A, Choudhury A, et al. (2023) Urban Heat: UHI and Heat Stress Threat to Megacities, In: Chatterjee U, Shaw R, Kumar S, et al. Eds., *Climate Crisis: Adaptive Approaches and Sustainability,* Sustainable Development Goals Series, Springer, Cham, 425–445. https://doi.org/10.1007/978-3-031-44397-8_22
8. Bao Y, Li Y, Gu J, et al. (2025) Urban heat island impacts on mental health in middle-aged and older adults. *Environ Int* 199: 109470. https://doi.org/10.1016/j.envint.2025.109470
9. Kutty SG, Jain M (2025) Urban Fog and its Impact on Atmospheric Visibility Degradation. In: Saxena P, Gupta AK, Sorí R, Eds., *Fog: Formation, Chemistry, Dynamics, Impact and Dissipation*, Singapore: Springer Nature Singapore, 83–107. https://doi.org/10.1007/978-981-96-4422-3_5
10. Diksha, Kumari M, Mishra VN, et al. (2024) Unveiling pollutants in Sonipat district, Haryana: Exploring seasonal, spatial and meteorological patterns. *Phy Chem Earth* 135. https://doi.org/10.1016/j.pce.2024.103678
11. Heidari H, Mohammadbeigi A, Khazaei S, et al. (2020) The effects of climatic and environmental factors on heat-related illnesses: A systematic review from 2000 to 2020. *Urban Clim* 34: 100720. https://doi.org/10.1016/j.uclim.2020.100720
12. Simpson CH, Brousse O, Taylor T, et al. (2025) The mortality and associated economic burden of London's summer urban heat island effect: a modelling study. *Lancet Planet Health* 9: e219–e226. https://doi.org/10.1016/S2542-5196(25)00025-7
13. Shekhar S, Chauhan MS, Omar PJ, et al. (2021) River Discharge Study in River Ganga, Varanasi Using Conventional and Modern Techniques, In: Chauhan MS, Ojha CSP, Eds., *The Ganga River Basin: A Hydrometeorological Approach,* Society of Earth Scientists Series, Springer, Cham. 101–113. https://doi.org/10.1007/978-3-030-60869-9_7
14. Han L, Bao Y, Gu J, et al. (2025) Urban heat island and the risk of schizophrenia spectrum disorders in middle-aged and older adults. *Urban Clim* 61: 102435. https://doi.org/10.1016/j.uclim.2025.102435
15. Joshi K, Kumari M, Mishra VN, et al. (2025) Geoinformatics based evaluation of heat mitigation strategies through urban green spaces in a rapidly growing city of India: implications for urban resilience. *Theor Appl Climatol* 156: 188. https://doi.org/10.1007/s00704-025-05411-4
16. Huang B, He BJ (2025) Lawn and irrigation cooling from ground longwave radiation reduction: Understanding the climate-driven variability in cooling performance. *Urban Clim* 60: 102360. https://doi.org/10.1016/j.uclim.2025.102360

17. Wang H, Yi T, Lu Y, et al. (2025) Patterns of nighttime surface urban heat island patch in mega urban agglomerations: a case study in the Pearl River Delta, China. *Sustain Cities Soc* 128: 106465. https://doi.org/10.1016/j.scs.2025.106465
18. Omar PJ, Gupta N, Tripathi RP, et al. (2017) A Study of Change in Agricultural and Forest Land in Gwalior City Using Satellite Imagery. *SAMRIDDHI* 9: 109–112. https://doi.org/10.18090/samriddhi.v9i02.10870
19. Perini L, Colantoni A, Renzi G, et al. (2017) Urban sprawl, soil sealing and impacts on local climate, *Urban Expansion, Land Cover and Soil Ecosystem Services*, London, Boston: Routledge, Routledge, 193–203. https://doi.org/10.4324/9781315715674-11
20. Chakrabortty R, Pramanik M, Hasan MM, et al. (2025) Mitigating Urban Heat Islands in the Global South: Data-driven Approach for Effective Cooling Strategies. *Earth Syst Environ* 9: 447–474. https://doi.org/10.1007/s41748-024-00449-2
21. Fotoohi O, Jahani N, Lee S (2024) A method for regulating land surface temperature by creating balance in urban landscape heterogeneities: the case of Tehran Metropolitan. *Environ Dev Sustain*, 1–23. https://doi.org/10.1007/s10668-024-05334-6
22. Kumari M, Somvanshi S, Sharma R, et al. (2022) Analysis, of multi-temporal remotely sensed spectral indices influence on ecology of Singrauli sub-district, Madhya Pradesh using an ecological impact index. *Egypt J Remote Sens Space Sci* 25: 863–871. https://doi.org/10.1016/j.ejrs.2022.08.005
23. Islam S, Karipot A, Bhawar R, et al. (2024) Urban heat island effect in India: a review of current status, impact and mitigation strategies. *Discov Cities* 1: 34. https://doi.org/10.1007/s44327-024-00033-3
24. Omar PJ, Gupta P, Wang Q (2023) Exploring the rise of AI-based smart water management systems. *AQUA Water Infrastruct Ecosyst Soc* 72: iii–iv. https://doi.org/10.2166/aqua.2023.005
25. Shalwee, Dhupper R, Kumari M, et al. (2025) Drought vulnerability assessment with meteorological drought indices for ensuring environmental justice. *Discov Sustain* 6: 1181. https://doi.org/10.1007/S43621-025-01178-W
26. Assenova IA, Vitanova LL, Petrova-Antonova D (2024) Urban heat islands from multiple perspectives: Trends across disciplines and interrelationships. *Urban Clim* 56: 102075. https://doi.org/10.1016/j.uclim.2024.102075
27. Tengku Ibrahim TNB, Feisal NAS, Cheah WY, et al. (2025) Monitoring, Assessing, and Forecasting Urban Climate Issues and Challenges, *Urban Climate and Urban Design*, Urban Sustainability, Springer, Singapore, 19–47. https://doi.org/10.1007/978-981-96-1521-6_2
28. Singh S, Mall RK, Chaturvedi A, et al. (2024) Advances in remote sensing in measuring urban heat island effect and its management. *Earth Observation in Urban Monitoring*, Elsevier, 113–132. https://doi.org/10.1016/B978-0-323-99164-3.00011-2
29. Somvanshi SS, Kumari M, Sharma R (2024) Spatio-temporal analysis of rural–urban transitions and transformations in Gautam Buddha Nagar, India. *Int J Environ Sci Technol* 21: 5079–5088. https://doi.org/10.1007/s13762-023-05336-3

30. Derardja B, Khadra R, Abdelmoneim AAA, et al. (1927) Advancements in Remote Sensing for Evapotranspiration Estimation: A Comprehensive Review of Temperature-Based Models. *Remote Sens* 16: 1927. https://doi.org/10.3390/rs16111927
31. Asghari Saraskanroud S, Faramarzi Ouri B, Zeinali B, et al. (2024) Estimation of land surface temperature (LST) using single-channel and multi-band methods in Sablan mountainous region. *Adv Space Res* 74: 2915–2929. https://doi.org/10.1016/j.asr.2024.06.027
32. Srivastava VT, Sharma A, Jadon SS (2024) A review of the formation, mitigation strategies from 50 years of global urban heat island studies. *Environ Dev Sustain* 28: 97–114. https://doi.org/10.1007/s10668-024-04966-y
33. Petrou I, Kassomenos P (2024) Estimating the importance of environmental factors influencing the urban heat island for urban areas in Greece. A machine learning approach. *J Environ Manage* 368: 122255. https://doi.org/10.1016/j.jenvman.2024.122255
34. Mishra VN, Kumar V, Prasad R, et al. (2021) Geographically Weighted Method Integrated with Logistic Regression for Analyzing Spatially Varying Accuracy Measures of Remote Sensing Image Classification. *J Indian Soc Remote Sens* 49: 1189–1199. https://doi.org/10.1007/s12524-020-01286-2
35. Neeraj (2024) Expansion of Irrigation Facilities and Its Impact on Cropping Intensity, *Disaster Management and Environmental Sustainability*, Wiley, 121–130. https://doi.org/10.1002/9781394167463.ch9
36. Satpal, Nirmala, Kumar B (2024) Dynamics of Industrialization in Sonipat and Rewari Districts of Haryana, *Latest Trends in Engineering and Technology*, London: CRC Press, 387–394. https://doi.org/10.1201/9781032665443-55
37. Verma P, Jangra R, Kaushik SP (2024) Geospatial measurement of urban sprawl and land transformation using multi-temporal datasets: A case study of Sonipat-Kundli urban agglomeration. *Sustainable Environ* 10. https://doi.org/10.1080/27658511.2024.2366556
38. Singh J, Kataria D, Rani B, et al. (2025) Association Between Different Type of Addictions And Oral Cancer: A Case Control Study in Sonipat District of Haryana. *Indian J Public Health Res Dev* 16: 212–218. https://doi.org/10.37506/jhv8y868
39. Pande CB, Egbueri JC, Costache R, et al. (2024) Predictive modeling of land surface temperature (LST) based on Landsat-8 satellite data and machine learning models for sustainable development. *J Cleaner Prod* 444: 141035. https://doi.org/10.1016/j.jclepro.2024.141035
40. Gök DT, Scherler D, Wulf H (2024) Land surface temperature trends derived from Landsat imagery in the Swiss Alps. *Cryosphere* 18: 5259–5276. https://doi.org/10.5194/tc-18-5259-2024
41. Walawender JP, Szymanowski M, Hajto MJ, et al. (2014) Land Surface Temperature Patterns in the Urban Agglomeration of Krakow (Poland) Derived from Landsat-7/ETM+ Data. *Pure Appl Geophys* 171: 913–940. https://doi.org/10.1007/s00024-013-0685-7
42. da Silva BB, Braga AC, Braga CC, et al. (2016) Procedures for calculation of the albedo with OLI-Landsat 8 images: Application to the Brazilian semi-arid. *Rev Bras Eng Agríc Ambient* 20: 3–8. https://doi.org/10.1590/1807-1929/agriambi.v20n1p3-8

43. Bernard J, Bocher E, Petit G, et al. (2018) Sky View Factor Calculation in Urban Context: Computational Performance and Accuracy Analysis of Two Open and Free GIS Tools. *Climate* 6: 60. https://doi.org/10.3390/cli6030060
44. Ismaila ARB, Muhammed I, Adamu B (2022) Modelling land surface temperature in urban areas using spatial regression models. *Urban Clim* 44: 101213. https://doi.org/10.1016/j.uclim.2022.101213
45. Guha S, Govil H, Dey A, et al. (2018) Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city, Italy. *Eur J Remote Sens* 51: 667–678. https://doi.org/10.1080/22797254.2018.1474494
46. Nasar-u-Minallah M, Haase D, Qureshi S, et al. (2023) Ecological monitoring of urban thermal field variance index and determining the surface urban heat island effects in Lahore, Pakistan. *Environ Monit Assess* 195: 1212. https://doi.org/10.1007/s10661-023-11799-1
47. Kashki A, Karami M, Zandi R, et al. (2021) Evaluation of the effect of geographical parameters on the formation of the land surface temperature by applying OLS and GWR, A case study Shiraz City, Iran. *Urban Clim* 37: 100832. https://doi.org/10.1016/j.uclim.2021.100832
48. Njoku A, Tenenbaum DE (2022) Quantitative assessment of the relationship between land use/land cover (LULC), topographic elevation and land surface temperature (LST) in Ilorin, Nigeria. *Remote Sens Appl Soc Environ* 27: 100780. https://doi.org/10.1016/j.rsase.2022.100780
49. Hama Rash AJ, Khodakarami L, Muhedin DA, et al. (2024) Spatial modeling of geotechnical soil parameters: Integrating ground-based data, RS technique, spatial statistics and GWR model. *J Eng Res* 12: 75–85. https://doi.org/10.1016/j.jer.2023.10.026
50. Badapalli PK, Gugulothu S, Nakkala AB (2025) Geospatial and NDBI approaches for the Musi River basin morphometric studies in the metropolitan urban Cities of India. *Adv Space Res* 75: 3375–3396. https://doi.org/10.1016/j.asr.2024.12.010
51. Diksha, Kumari M, Kumari R (2023) Spatiotemporal Characterization Of Land Surface Temperature In Relation Landuse/Cover: A Spatial Autocorrelation Approach. *J Landsc Ecol* 16: 1–18. https://doi.org/10.2478/jlecol-2023-0001
52. Wang A, Al Kafy A, Feng Y, et al. (2025) Controlling Spatial Heterogeneity of Land Transfer Dynamics Using Multiscale Regression Approach for Sustainable Resource Management. *Land Degrad Dev* 36: 1018–1033. https://doi.org/10.1002/ldr.5409

AIMS Press