Identification of Flood-Vulnerable Village Clusters in the Southwestern Fringe of Guwahati City Using GIS-Based Spatial Autocorrelation Technique and Real-Time SAR Imagery

Rakesh Kumar Sarmah[1], Santanu Sarma[2]

Abstract

This study employs a multifaceted approach, combining remote sensing, GIS-based spatial autocorrelation techniques, and Sentinel-1 Synthetic Aperture Radar (SAR) imagery to comprehensively evaluate flood-affected areas in the southwestern fringe of Guwahati city. The unplanned urban expansion within the adjacent main city is causing issues like flooding and waterlogging, impacting the city's residents. Critical assessment is needed in the fringe areas to avoid similar consequences. The historical Annual Flood Inundation Layers available in the ISRO-BHUVAN portal are analyzed in a GIS environment village-wise considering 10 maps spanning from 2001 to 2010 to identify the vulnerable villages. The National Remote Sensing Centre (NRSC) generates these layers by integrating different flood waves in a calendar year considering the maximum flood extent. The village-wise inundation statistics are analyzed using the Global Moran's I to understand whether the inundation is clustered, dispersed, or random during this timeframe. The results show very high z-score and very low p-values indicating the presence of statistically significant spatial clusters. The historical data are further analyzed using Local Moran's I and Getis-Ord Gi statistics and vulnerable village clusters are delineated towards the study area's north, west, and central portion during these 10 years. The clusters show temporal dynamism and interconnection in major flood years. To assess these areas in real-time after a decade, the Sentinel-1 Synthetic Aperture Radar (SAR) imagery was used during a major flood event in 2020. The village-wise inundation is determined by extracting the flood signatures through the intensity*

[1] Research Scholar, Department of Geology, Cotton University, Panbazar, Guwahati – 781001, Assam, India
[2] Professor, Department of Geology, Cotton University, Panbazar, Guwahati – 781001, Assam, India

thresholding method. The method segregates land pixels against water pixels based on the backscattered value range of the Radar signal. The SAR data analysis is compared with the 2010 results due to their similar areal extent. The Area Under the Receiver Operating Characteristic (ROC) Curve is 70.38%, indicating that the analysis demonstrates good prediction accuracy. The study provides a roadmap for urban planning, disaster management, water resource management, and related disciplines.

Keywords: *Urban Expansion, Fringe Area Development, Flood-prone Villages, Spatial Autocorrelation, Synthetic Aperture Radar*

1. Introduction

It is crucial to address the issue of flooding in the peri-urban or fringe areas of developing cities, especially in the face of climate change. These areas are at risk due to the prevalence of informal settlements in low-lying agricultural lands, floodplains, protected areas such as Ramsar sites, and areas with high biodiversity (Winter & Karvonen, 2022). The dynamism in these areas with a gradual transition of green spaces into impervious surfaces with the blocking of natural waterways makes them more vulnerable to flood disasters in comparison to rural areas (Wang et al., 2023). Due to lower property and service tax compared to the city, peri-urban areas attract industries, commercial establishments, etc. amplifying the urbanization process (Nallathiga, 2015). Urbanization indicates economic growth but at the same time, it has negative impacts in terms of the physical environment such as loss of agricultural land, surface, and groundwater depletion, changes in geomorphic features, flooding, and landslides (Mahapatra et al., 2014). Proper planning is required to make these areas resilient and sustainable to ensure supporting quality standards of life as they are carriers of urban expansion and act as a distribution centre of goods and energy between the city and external regions (Dong et al 2022; Mahanta and Borgohain; 2022).

Guwahati, the gateway to northeast India, is continuously expanding in terms of population and urbanization. The city's growth is due to the development of infrastructure to support the ever-increasing population. However, this growth has caused encroachment in hills, wetlands, and government lands. Flash floods and waterlogging occur annually in Guwahati city during monsoon season due to overflowing feeder drains (ASDMA, 2014). Urban flooding causes socio-economic impacts such as education, health, transportation, drinking water, sanitation, and livelihood insecurity (Kashyap & Mahanta, 2020). Anthropogenic activities magnify the inundation (Rashiq & Prakash, 2023), and poor urban planning makes the city vulnerable to natural disasters. Spatial planning with proper land use allocation is crucial in minimizing exposure to natural hazards (Abdrabo et al., 2021). Remote sensing and GIS technology are extensively used in site suitability analysis for urban expansion (Saleh & Rawashdeh, 2007, Huang et al., 2010, Sun et al., 2020, Sahin et al., 2024). Flood vulnerability assessment regarding people and places is essential for emergency management (Chen et al 2021). The city's Southwest direction is a potential site for a sustainable satellite township. The study aims to identify spatially significant clusters of flood-vulnerable villages, by adopting spatial autocorrelation techniques by analyzing past data and real-time SAR imagery in a GIS environment. The concept of spatial autocorrelation is based on Waldo Tobler's first law of geography, which emphasizes that "near things are more related than distant things" (Manning et al., 2023). This spatial autocorrelation technique is commonly used in various fields to understand spatial clustering of variables, for example in health sciences (Shariati et al., 2020, Vilinová & Petrikovičová., 2023, Yin et al., 2024), traffic analysis (Le et al., 2022), rainfall pattern analysis (Rousta et al., 2017), pollution concentration (Liu et al., 2013), waste management (Tsui et al., 2022), and natural hazards (Lin et al., 2017). To confirm these vulnerable villages in real-time, Sentinel 1 Synthetic Aperture Radar (SAR) imagery of a particular flood date, after a decade is used. Having up-to-date information on flooding is important for emergency response and reducing the risk of flood damage. SAR is particularly useful as it can penetrate through clouds and provide day and night images making it an effective remote sensing method for mapping real-time flood events (Garg et al. 2024). The village-wise real-time inundation statistics obtained from SAR are compared with the vulnerable villages demarcated for multiple major flood years to ensure the clustered pattern.

1.1 Study Area:

The study area encompasses the administrative blocks of Rani and Chayani Barduar in Kamrup – Metro and Rural districts of Assam, spanning a total of 423.03 km^2 with a total population of 184411 in 129 villages. The southwestern region of the city is primarily utilized for agriculture, while the southern area is a narrow valley encircled by hills and the picturesque Chandubi Lake. The Guwahati Metropolitan Development Authority (GMDA) Master Plan aims to build a new town featuring an Information Technology hub and Special Economic Zone. The study area is bordered by the Brahmaputra River in the north, the Kulshi River in the west, Meghalaya plateau hills in the south and southeast, and Guwahati city in the northeast (Figure 1). The major geological formations within the study area are alluviums of the Brahmaputra River and its tributaries and Precambrian rocks of the Meghalaya plateau. The distinguishable geomorphic features within the study area are floodplains, alluvial plains, piedmont zones, and denudational hills. The Kulsi river system is also present in the area, with all streams flowing towards the north to merge with the Brahmaputra River. The region is prone to annual flooding due to heavy rainfall. As per the Indian Meteorological Department report (IMD), the average annual rainfall in the study area betwee 2018-2022 is 1583.07 mm. The region experiences a tropical monsoon climate with rainfall from June to September. The average temperature varies from 9°C in winter to 35°C in summer.

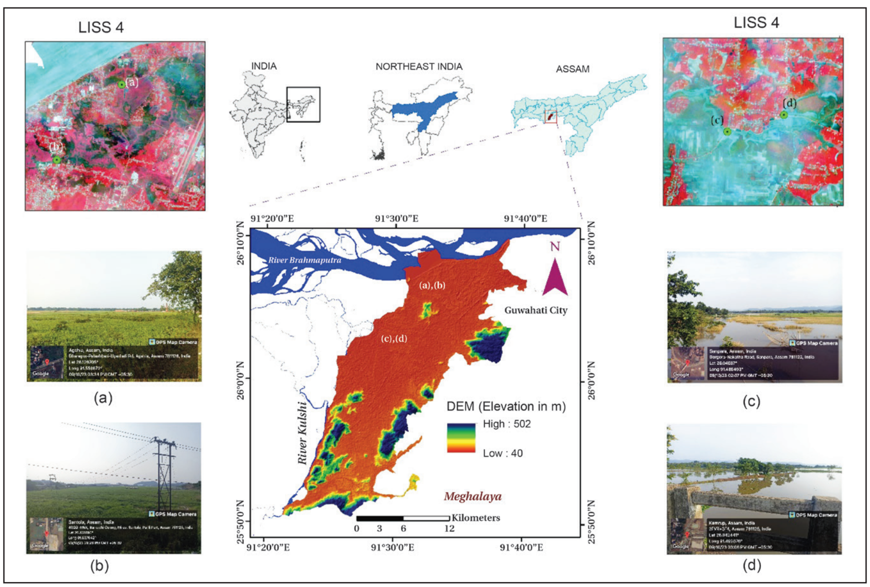

Figure 1: Study area map with recent field photographs of flood-vulnerable sites

2. Materials and Methods:

2.1 Flood History

The Decision Support Centre (DSC) under the National Remote Sensing Centre (NRSC) in Hyderabad, India is responsible for incorporating space technology into the Disaster Management Support (DMS) program by the Indian Space Research Organization (ISRO). The DSC has released an annual flood inundation atlas (1998-2010) for Assam and Bihar using optical and microwave remote sensing techniques to analyze different flood waves throughout the year with an accuracy of up to a scale of 1:250000. This archive is freely available in ISRO-BHUVAN portal. Hailin et al. (2009) have used the same scale to assess flood risk in Hubei Province. Out of 13 available maps, the inundation history of the study area has been analyzed for 10 years (2001-2010). The available maps are georeferenced using Arc GIS 10.8 in WGS 84 UTM 46 N projections, the raster maps are converted to vector polygons, and year-wise flood inundation is calculated. The village layer of the study area is collected from Assam Remote Sensing Application Centre, Govt of Assam, India, and the population data are collected from the District Census Handbook (DCHB) published by Census of India (2011). The village layer is superimposed in the GIS environment, and the village-wise inundation area is calculated for the specific period.

2.2 Flood Clusters Using a Spatial Autocorrelation Technique

Spatial autocorrelation is a type of correlation that measures the degree of alignment between two attributes based on their relative magnitudes (Griffith & Chun, 2018) or their similar patterns within a geographic area (Lo et al., 2022). Positive spatial autocorrelation occurs when closer geographic locations have similar attributes (Griffith & Lea, 2005), whereas negative spatial autocorrelation is observed when distant values are more similar than nearby values, contradicting Tobler's law (Fischer & Wang, 2011).

Spatial correlation can be global or local, with global Moran's I being a common measure for evaluating whether the pattern expressed is clustered, dispersed, or random for a given set of features and associated attributes in a study area (Fasona et al., 2011, Prasannakumar et al., 2011). The village-wise inundation statistics in terms of percentage of inundation are considered as the numerical field used in assessing spatial autocorrelation. A small village with a large inundation area is considered more vulnerable than a large village with a relatively small inundation.

The global Moran's I can be calculated for n observations on a variable x at locations (i,j)

$$I = \frac{n}{S_0} \frac{\sum_i^n \sum_j^n w_{ij} \quad (x_i - \bar{x}) \quad (x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2}$$

Where,

$\bar{x}$ = the mean of the variablex, x_i = the value of the variable x at location i, x_j = the value of the variable x at location j, w_{ij} = the elements of the weight matrix, n = number of observations, S_0 = Sum of the elements of the weight matrix = $\sum_i^n \sum_j^n w_{ij}$

The Global Moran's I value ranges from -1 to +1. A value of '-1' indicates perfect clustering of dissimilar values or perfect dispersion. A value of 0 indicates no autocorrelation or perfect randomness. A value of +1 indicates perfect clustering of similar values. The Global Moran's I use inferential statistics, meaning the results of the analysis are interpreted within the context of its null hypothesis, considering a random distribution of the attributes under scrutiny. A significant p-value, coupled with a positive Moran's I, suggests that the high and/or low values in the dataset are more spatially clustered than would be expected under such a random distribution, leading to the rejection of the null hypothesis. Conversely, a significant negative Moran's I indicate that the spatial distribution of high and low values is more dispersed than expected (Raza et al., 2020). For spatial relationships, the 'contiguity edges only' option in ArcGIS is used where only neighboring polygon features that share a boundary, or overlap will influence the computations of the target polygon.

Global Moran's I measures the clustering strength of a map as a whole, whereas local Moran's I identifies the location of clusters on the map. The formula for local Moran's I is similar to its global counterpart, with the only difference being that it only iterates through the neighbors of one polygon instead of all pairs of polygons (Tsui et al. 2022).

$$I_i = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1}^{n} w_{ij} \, (x_j - \bar{x})$$

Where, $S_i^2 = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^2}{n-1} - \bar{x}^2$

$$E(I) = \frac{-1}{n-1}$$

Moran's I value greater than E(I) implies positive autocorrelation and smaller implies negative spatial autocorrelation. The clusters can be either high-high (meaning high values in a high-value neighborhood) or low-low (meaning low values in a low-value neighborhood), a high-low (meaning a high value in a low-value neighborhood), or a low-high (meaning a low value in a high-value neighborhood) outlier (Fu et al., 2014).

Another reliable geostatistical technique for mapping clusters and determining statistically significant hot or cold spots is the Gi* statistic, also known as the Getis-Ord Gi. It assesses the Gi* for each feature in a dataset and identifies where features with high or low z-scores and p-values are spatially clustered. A positive z-score indicates a higher clustering of high values, while a negative z-score indicates a higher clustering of low values (Hazaymeh et al., 2022).

$$Gi^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{s\sqrt{\frac{n \sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}}$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{X}^2}$$

Where x_j is the attribute value for the feature j, $w_{i,j}$ is the weight between features *i* and *j* and n is equal to the total number of features. TheG*i** statistics is the z score.

The vulnerable village clusters are finally delineated year-wise integrating Asselin's Local Moran's I and the Getis-Ord Gi statistics.

2.3 Real-time validation using SAR after 10 years on a significant flood date

To identify the real-time extent of flood inundation within the study area a decade later, a significant flood date, July 16, 2020, was selected based on heavy rainfall in Kamrup Rural and Metro districts of Assam and the adjacent RiBhoi district of Meghalaya, as per INSAT-3D images provided by the India Meteorological Department (IMD). The state's leading newspaper, The Assam Tribune, published a report on the same date detailing the severity of the flood conditions in Kamrup Rural and Kamrup Metro, where the study area is an integral part (Figure 6).

2.3.1 SAR Data

The Sentinel 1 mission involves two operational satellites called Sentinel 1A and Sentinel 1B that were launched by the European Space Agency in April 2014 and April 2016, respectively. These satellites carry a C Band Synthetic Aperture Radar (C-SAR) payload and follow a sun-synchronous orbit with a repeat cycle of 12 days (Raspini et al., 2018; Potin et al., 2016). The data they collect is available for free on the Copernicus Open Access Hub platform of the European Space Agency (ESA) and is used extensively in critical decision-making during emergencies such as natural disasters or humanitarian crises (Skoda & Adam, 2020). These RADAR datasets can be analyzed using the Sentinel Application Platform (SNAP), an open-source architecture developed by the European Space Agency. To download the necessary product, users only need to create a login ID. To determine the exact extent of flooding on a specific date, it is crucial to have data on permanent bodies of water and their extent before the flood season. Therefore, two images depicting pre-flood and during-flood conditions are required. The pre-flood image is a Sentinel 1A Ground Range Detected (GRD) product from February 15th, 2020, while the during-flood image is a Sentinel 1B GRD product captured on July 16th, 2020.

Table 1: Details of Sentinel 1 products considered for the study

Satellite	Product Name	Date and Time	Product Type	Acqui-sition Mode	Spatial Resolution
Sentinel 1A	S1A_IW_GRDH_1SD-V_20200215T115719_031263_031263_03988A_A076	15 Feb 2020, 14:31:22	Level-1 GRD	IW	10×10 m
Sentinel 1B	S1B_IW_GRDH_1SD-V_20200716T234628_20200716T234657_022503_02AB64_13F1	16 Jul 2020, 17:40:35	Level-1 GRD	IW	10×10 m

2.3.2 Processing of the data

To process the downloaded data, SNAP open-source software is used. The image needs to be pre-processed to remove any distortion and enhance its quality. The pre-processing involves several steps that are performed sequentially. The raw data, which comes in a zip format from Copernicus Open Access Hub, is directly processed using the SNAP software step by step. To ensure the accuracy of the orbital data, improve geocoding, and the subsequent processing steps, the orbit file is applied (Hong et al., 2017). To convert the pixel intensity or radar reflectivity (DN number) to radar cross-section (RCS), or scattering cross-section, the radiometric calibration technique is used (Schmidt et al., 2020; WeiB, 2018). The RCS is the measure of the reflective strength of the radar target, and it is indicated by the symbol 'sigma'. The value of 'sigma' depends upon the shape, dielectric properties, orientation, roughness, etc. of the target, which also varies for different observation angles, frequencies, and polarizations. Since the radar return signifies the proportion of energy scattered by the distributed area corresponding to the size of the image pixel, a normalized measure is required to relate the inferred target area, derived from the scattering cross-section, to the actual geometrical area on the ground surface. This unit-less normalized measure is defined as 'backscatter coefficient', 'differential radar cross-section,' or 'normalized radar cross-section,' denoted by 'sigma nought' (Woodhouse 2006; Sentinel user guide). Due to the low backscattering response, flooded areas appear dark, and strong backscattering from rough soil surfaces and vegetation causes land surfaces to appear bright (Manjusree et al., 2012). 'Speckles' are random, high-frequency noises present in SAR images, causing a granular, salt, and peppery pattern with degradation of image quality, as well as loss of information (Choi & Jeong 2019; White et al., 2015). To enhance the image quality, the 'speckle filtering' technique is applied. Different speckle filters are available in SNAP, such as Gamma Map, Lee, Refined Lee, and Lee Sigma, with different window sizes. Comparing different filters, the Lee (5*5) window size is found to be suitable for the study. The side-looking geometry of RADAR and topographic relief causes several geometric distortions, such as foreshortening, layover, and shadow (Chen et al., 2018; Esmaeilzade et al., 2015). The geometric distortion caused by topography is corrected using the Range Doppler terrain correction, which uses a digital elevation model to rectify the location of each pixel. The process utilizes available orbital information in the metadata, radar timing annotations, and slant-to-ground range conversion parameters with the reference DEM (Filipponi, 2019). For this study, the SRTM DEM with 3 arcsec and bilinear interpolation resampling technique has been used for geometrical correction. The final product is converted to decibel (dB) units using logarithmic transformation. Conversion of linear amplitude to decibels (dB) improves the image display by compressing a wide range of values and reducing the impact of multiplicative noise (Filipponi, 2019).

3. Results

3.1 Historical Flood Vulnerable Villages (2001-10)

Based on the area of inundation, the villages are classified into five categories: Very High (above 75%), High (75-50%), Moderate (50-25%), Low (25-10%), and Very Low (below 10%). The year-wise frequency distribution for different categories is determined for 10 years. Figure 2 shows the significant distribution of flood events in all five categories. During peak flooding years, such as 2002 and 2004, six to eight villages fall under the 'high' category, while twelve to sixteen villages fall under the 'moderate' category, indicating the severity of flood hazards in the study area.

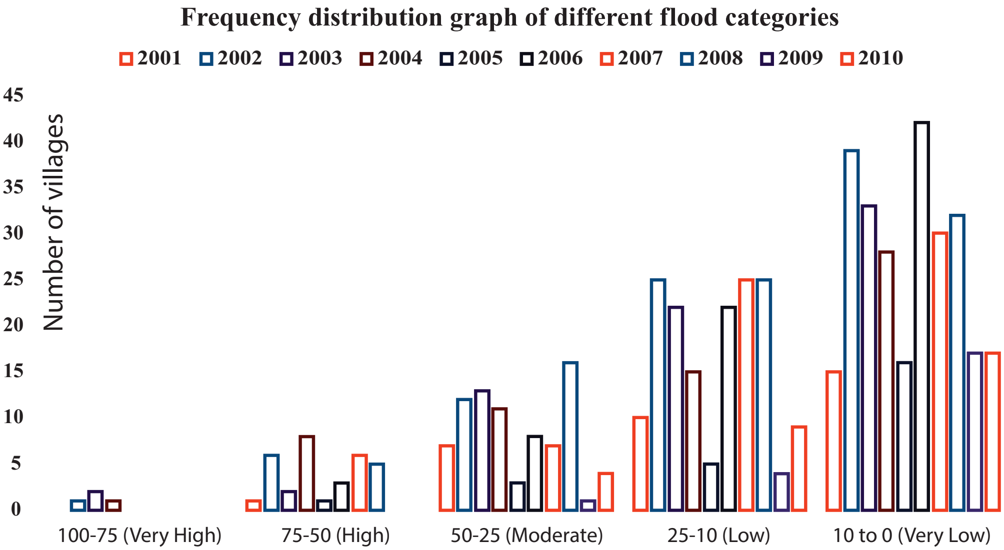

Figure 2: Frequency distribution graph of different flood categories

Out of 10 years of analysis, the highest flood inundation is observed for the year 2003 followed by 2007, 2002, and 2004. The area undergoing inundation is 67.23, 64.33, 61.35, and 58.49 km2 respectively. The lowest inundation observed within this timeframe was in 2009 with only 4.75 km2 area undergoing inundation. Table 2 depicts the number of villages and populations affected during each flood event. The figures show that more than 1 lakh people are vulnerable during major flood years.

Table 2: The vulnerability of villages and population during the study period, and z-score, p-values of the spatial autocorrelation analysis.

Year	Total Area under flood (km^2)	Total Village	Total population	Global Moran's I			Getis–Ord Gi*			Type of distribution
				Moran's I	z-score	p-value	General G	z-score	p-value	
2001	15.94	26	69688	0.31	8.23	0	0.019599	7.83	0	Clustered
2002	61.35	97	166139	0.34	8.87	0	0.010899	8.88	0	Clustered
2003	67.23	87	158527	0.32	8.34	0	0.010530	8.81	0	Clustered
2005	10.16	29	78721	0.11	3.49	0.00049	0.015892	3.36	0.000793	Clustered

Year	Total Area under flood (km^2)	Total Village	Total popula-tion	Global Moran's I			Getis–Ord Gi*			Type of distribu-tion
				Moran's I	z-score	p-value	General G	z-score	p-value	
2008	33.07	76	135240	0.24	6.24	0	0.011231	7.24	0	Clustered
2009	4.75	25	70693	0.11	3.26	0.00109	0.014244	3.04	0	Clustered
2010	10.96	33	64261	0.16	4.55	0.000005	0.013685	3.81	0	Clustered
2020 SAR	13.39	50	83359	0.16	4.71	0.000002	0.010994	4.62	0.000004	Clustered

3.2 Spatial Autocorrelation Results

Based on the positive z-score and significant p-value of Global Moran's I, it seems likely that the null hypothesis can be rejected. This means that the distribution of high and/or low values in the dataset is more spatially clustered than expected by chance, with less than a 1% likelihood that the pattern is random. To identify flood vulnerable clusters, we need to examine the local level of clustering and identify villages where the percentage of inundation is particularly high. Figure 3 displays the results of Anselin's Local Moran'I statistics during major flood years. High-High clusters are concentrated in the north, west, and central parts of the study area, while low-low clusters are mainly found in the east and south. Spatial outliers, such as low-high clusters, are prominent near-high clusters, as these are locations with low values and high-value neighbors.

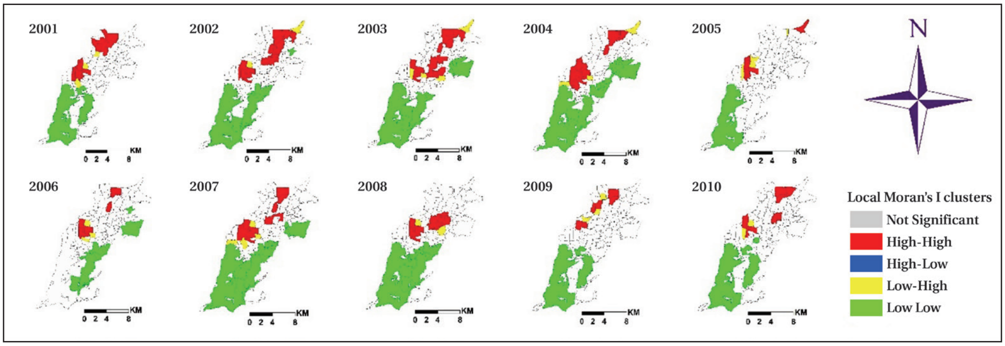

Figure 3: Local Moran's I cluster analysis of the study area between 2001 and 2010

According to the analysis using the Getis-Ord Gi* statistics, there are both flood hot spots and cold spots. These spots have confidence levels of 90%, 95%, and 99%. The hot spots indicate an intense clustering of high positive z values, while the cold spots show an intense clustering of low negative z values (Figure 4). These spots are located spatially in accordance with the clusters identified by the local Moran's I.

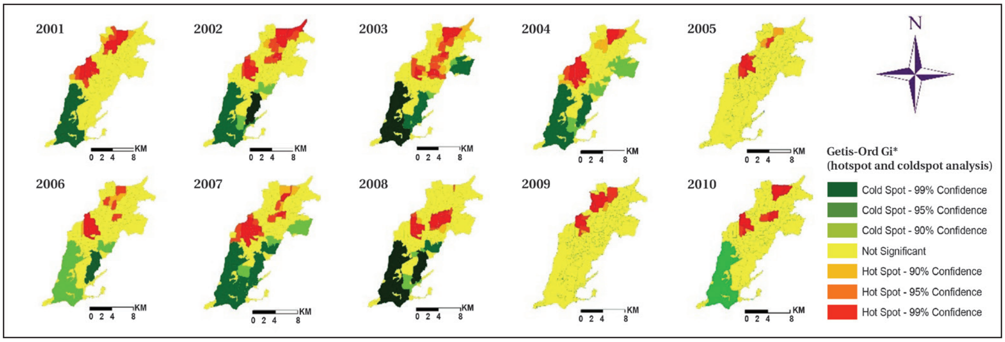

Figure 4: Getis-Ord Gi* hot spot and cold spot analysis of the study area between 2001 and 2010

The flood-vulnerable clusters are finally delineated by integrating the Local Moran's I High-High clusters and hotspots from Getis-Ord Gi* analysis (Figure 5)

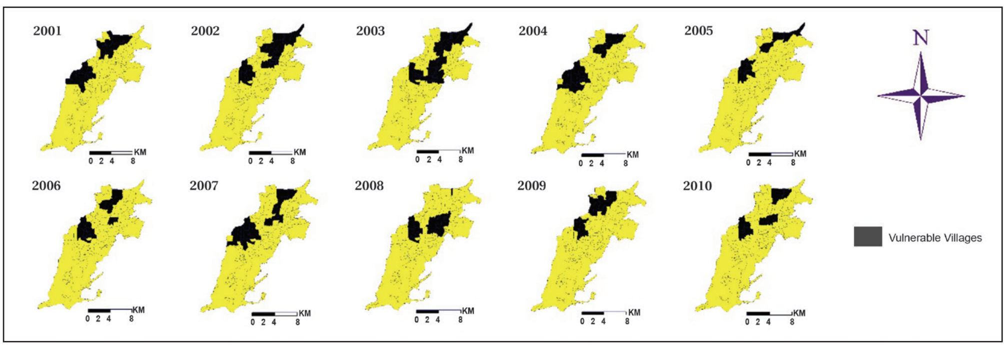

Figure 5: Vulnerable village clusters during 2001 and 2010

3.3 Results of SAR data analysis

The pre-flood image analysis provides information about permanent water bodies so that there won't be any overestimation of flooded areas during the flood. Permanent water bodies contain smooth open surfaces acting as specular reflectors, scattering the radar energy away from the sensor. Thus relatively dark pixels appear in the image in comparison to non-water areas (Martinis and Rieke, 2015). The pre-flood imagery of 15th February 2020 is processed following the steps as mentioned earlier and the final terrain corrected product in dB for both VV and VH polarization is obtained. VH polarization generates correctly defined surfaces and is found more suitable for demarcating flooded areas in comparison to VV polarization (Conde & Munoz, 2019). The most widely accepted technique in differentiating land pixels vs. water is the intensity thresholding method (López-Caloca et al., 2018). The backscattered value range can provide significant discrimination between water and non-water regions. Careful observation of the histogram provides a threshold value to separate water bodies from other features (Conde and Munoz 2019; Zhang et al. 2020). By using a suitable band maths expression (Sigma0_VH_db < -21.22), the water pixels can be extracted. The permanent water bodies within the study area are mainly stretches of river Brahmaputra, river Kulshi, Deepor Beel Lake, Chandubi Lake, and some ponds. The total area of permanent water bodies is found to be 14.03 km^2. Flood extent delineation is done using Sentinel 1B GRD product captured on 16th July 2020. The image is pre-processed to the final terrain-corrected dB product of both VV and VH polarization. The histogram has been analyzed and a threshold backscatter value (dB = -24.26) has been identified to separate water and non-water pixels from the image. The total area undergoing inundation is calculated in the GIS environment and found 27.42 km^2. This also includes the area of permanent water bodies, misclassified pixels, and topographic shadows. Misclassified pixels mainly occur in and around airport premises, flat surfaces such as airport runways, parking, and new construction sites. These areas are showing low backscattering which is not related to flooding. In many works of literature, misclassified pixels in airport runways and other flat-constructed surfaces are mentioned and validated using suitable techniques. Topographic shadow also shows dark tone not necessarily water pixels and are mainly found in forest areas at higher elevation. Validation is done using LISS-4 imagery and Google Earth. Nearly 0.107 km^2 areas are found under misclassified pixels due to airport runways and other flat construction sites and 0.023 km^2 due to topographic shadow. Excluding these areas, the total inundation on 16th July 2020 was found to be 13.39 km^2.

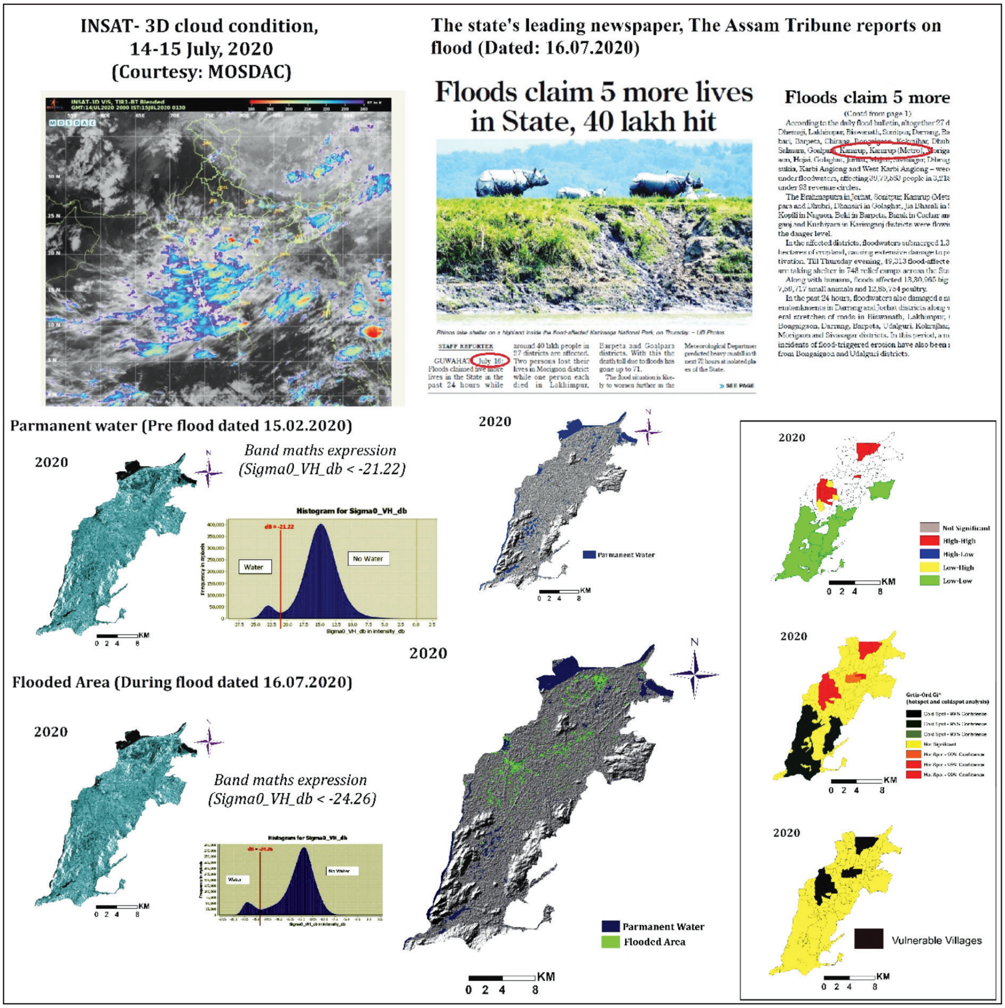

Figure 6: Sentinel-1 SAR-derived flood area estimation during a major flood date (evident from the cloud condition by INSAT-3D and news report) using histogram-based intensity thresholding method, cluster analysis using Local Moran's I and Getis Ord G* statistics

3.4 Validation

The Receiver Operating Characteristic (ROC) curve is a proven technique to ensure the classifier's efficiency and sensitivity of the method (Avand et al. 2021). Due to their similar area coverage, the village-wise inundation statistics obtained from SAR data analysis are compared with the 2001 results. In Figure 7, the Area under the ROC curve (AUC) shows 70.38% indicating good predictive ability (Tao et al. 2021).

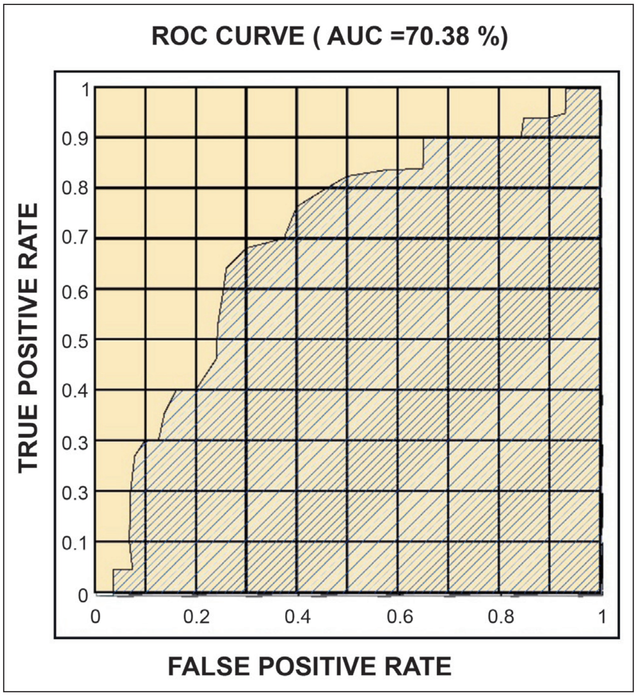

Figure 7: ROC analysis between 2001 historical data and 2020 SAR data

4. Discussion

The Local Moran's I and Getis-Ord Gi* analysis results reveal prominent flood clusters in the study area's north, west, and central parts. A similar trend of clustering is also evident from the SAR analysis. Figure 5 shows the dynamics of these vulnerable village clusters during the study period. The north and west clusters are seen throughout the analysis period and there is a chance of flood accumulation initiation. The central cluster is prominent in 2003, 2008 and 2010. During major flood events, the clusters tend to get interconnected. Though the vulnerable villages show interconnectivity, for the administrative and socio-economic point of view, the villages are classified into three major clusters: the clusters in the west consisting of 19 villages with an area coverage of 49.06 km^2 with total population of 26578; the clusters in the north consisting of 18 villages covering 61.03 km^2 with a total population of 74442; and the clusters in the central portion consisting of 17 villages with 37.29 km^2 coverage with a total population of 16915.

Figure 8 shows the location of the clusters considering land use land cover, existing stream network, and a digital elevation model in the background. The analysis of the cluster in the west shows a complex network of interconnecting streams and wetlands making this area flood-prone. The Kulshi and Batha Rivers from the south, and 'Rani Nadi' from the east interconnect here. In Figure 1 (c) and (d), a recent field photograph within this cluster, taken in October 2023, shows the receding floodwaters, providing evidence of the devastating floods during the monsoon season. The prominent hill ranges visible in the photograph are part of the elevated Meghalaya plateau upstream of this catchment, known for receiving the world's heaviest rainfall.

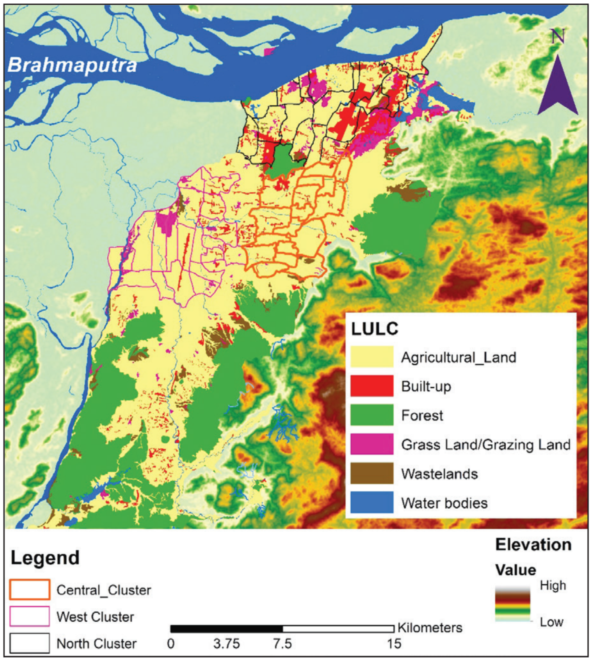

Figure 8: Cluster location considering land use land cover, stream network, and DEM

Similar flood concentrations are observed in the northern part of the study area. This region is gradually urbanizing with the establishment of industrial facilities and apartment complexes near the Guwahati Lokpriya Gopinath Bordoloi International Airport. Field photographs in Figure 1 (a) and (b) show the presence of water hyacinth and wetland-like features, which get inundated during the rainy seasons. These natural features are crucial for floodwater mitigation. It is essential to regulate concrete construction in these eco-sensitive areas to ensure future sustainability. The cluster in the central portion of the study area is mainly agricultural lands. The location of the study area is a narrow valley adjacent to the Shillong plateau and River Brahmaputra with a complex fluvial-morphological scenario, based on the intensity of rainfall in the upper catchments of River Brahmaputra as well as in the Meghalaya, the flood situations vary, and the dynamism may be witnessed in vulnerable clusters.

5. Conclusion

The primary objective of this study is to identify areas in the southwest vicinity of Guwahati city that are susceptible to flooding and natural waterlogging, with a focus on future urbanization. Identifying flood-vulnerable clusters through historical data and real-time imagery is crucial for urban planning. The fringe or peri-urban landscape is vulnerable in terms of rapid urbanization due to the rural-urban policy dichotomy. Towards west and north, the clusters are seen in interconnected networks of natural drains, wetlands, and swampy-marshy features adjoining the Brahmaputra floodplain, these features are important for flood mitigation, acting as storage for major flood events, filtering sediments, and recharging aquifers. A proper hydro-ecological assessment is needed before any urban intervention. The landscape of Guwahati city was like its fringe areas before urbanization with an abundance of wetlands, swamps, and marshes. The unplanned urbanization has led to the encroachment in wetlands, blockage of feeder drains intensifying waterlogging, and flash floods. The clusters identified in the central portion of the area are mainly vast agricultural lands. Due to the proximity of the city and relatively cheap land prices, these parcels are easy targets for investors for future ventures. The government should keep a critical eye on the infrastructure projects and strong policy decisions to ensure future sustainability to avoid water-related hazards. The study will provide a holistic overview of integrating remote sensing and GIS addressing flood vulnerability mapping in the context of urban developments.

6. Acknowledgement

The authors would like to thank different government and non-government organizations for providing valuable data.

References

1. Abdrabo, K. I., Kantosh, S. A., Saber, M., Sumi, T., Elleithy, D., Habiba, O.M., & Alboshy, B. (2022). The role of urban planning and landscape tools concerning flash flood risk reduction within arid and semiarid regions. In Sumi, T., Kantoush, S.A., Saber, M. (Eds.), *Natural Disaster Science and Mitigation Engineering:* DPRI Reports. Wadi Flash Floods Challenges and Advanced Approaches for Disaster Risk Reduction (pp. 283-316). *Springer: Singapore*.
2. Annual Flood Inundation- Assam and Bihar. NRSC/ISRO. (2012). Retrieved from https://bhuvan-app1.nrsc.gov.in/2dresources/documents/al.pdf
3. Assam State Disaster Management Authority. (2014). Review of Studies on urban floods in Guwahati from flood knowledge to urban action. Retrieved from https://asdma.assam.gov.in/sites/default/files/Guwahati%20floods.pdf
4. Avand M., Moradi H.R., Lasboyee M.R.: Spatial Prediction of Future Flood Risk: An Approach to the Effects of Climate Change. *Geosciences*, vol. 11, 2021, 25. https://doi.org/10.3390/geosciences11010025
5. Bharali, K.K., & Sarma, S. (2012). Bank erosion of the Brahmaputra River in the downstream of Palasbari and Soalkuchi, Kamrup district, Assam. *Journal of Indian Geomorphology*, 1(2012):103-106.
6. Carreno Conde, F., De Mata Munoz, M. (2019). Flood monitoring based on the study of Sentinel-1 SAR images: the Ebro River case study. *Water*, 11(12) :2454.
7. Chen, X., Sun, Q., & Hu, J. (2018). Generation of Complete SAR Geometric Distortion Maps Based on DEM and Neighbor Gradient Algorithm. *Applied Sciences*, 8(11):2206.
8. Chen, Y., Ye, Z., Liu, H., Chen, R., Liu, Z., & Liu, H. (2021). A GIS-based approach for flood risk zoning by combining social vulnerability and flood susceptibility: a case study of Nanjing, China. *International Journal of Environmental Research and Public Health*, 18(21): 11597.
9. Choi, H., & Jeong, J. (2019). Speckle noise reduction technique for SAR images using statistical characteristics of speckle noise and discrete wavelet transform. *Remote Sensing*, 11(10):1184.
10. District Census Handbook published by Census of India. Retrieved from https://censusindia.gov.in/census.website/data/handbooks
11. Dong, Q., Qu, S., Qin, J., Yi, D., Liu, Y., & Zhang J. (2022). A method to identify urban fringe area based on the industry density of POI. *International Journal of Geo-Information*, 11(2):128.
12. Esmaeilzade, M., Amini, J., & Zakeri, S. (2015). Georeferencing on synthetic aperture RADAR imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W5:179-184.
13. Fasona, M., Omojola, A., Soneye, A. (2011). A study of land degradation pattern in the Mahin mud-beach coast of southwest Nigeria with spatial-statistical modelling geostatistics. *Journal of Geography and Geology*, 3(1): 141-159.
14. Fischer, M.M., & Wang, J. (2011). Spatial data analysis: models, methods and techniques. *Springer*, Heidelberg.
15. Filipponi, F. (2019). Sentinel-1 GRD preprocessing workflow. *Proceedings*, 18(1):11.
16. Fu, W.J., Jiang, P.K., Zhou, G.M., & Zhao, K.L. (2014). Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences*, 11(8), 2401–2409.
17. Garg, S., Dasgupta, A., Motagh, M., Martinis, S., & Selvakumaran, S. (2024). Unlocking the full potential of Sentinel-1 for flood detection in arid regions. *Remote Sensing of Environment*, 315: 114417.
18. Griffith, D.A., & Lea A.C. (2005). Locational Decision Making. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 559-575). University of California: Elsevier Science.
19. Griffith, R.B., & Chun, Y. (2018). GIS and Spatial Statistics/Econometrics: An Overview. In B. Huang (Ed.), *Comprehensive Geographic Information Systems* (pp. 1-26). *Elsevier*, Netherlands.
20. Hailin, Z., Yi, J., Xuesong, Z., Gaoliao, J., Yi, Y., & Baoyin, H. 2009. GIS-Based Risk Assessment for Regional Flood Disaster. *International Conference on Environmental Science and Information Application Technology*, 2009: 564-567.

21. Hazaymeh, K., Almagbile, A., & Alomari, A.H. (2022). Spatiotemporal analysis of traffic accidents hotspots based on geospatial techniques. *International Journal on Geo-information*, 11(4):260.
22. Hong, S., Choi, Y., Park, I., & Sohn H.G. (2017). Comparison of orbit-based and time-offset-based geometric correction models for SAR satellite imagery based on error simulation. *Sensors*, 17(1): 170.
23. Huang, B., Zhang, L., & Wu, B. (2010). Spatiotemporal analysis of rural–urban land conversion. *International Journal of Geographical Information Science*, 23(3): 379-398.
24. Kashyap, S., & Mahanta, R. (2021). Socioeconomic vulnerability to urban floods in Guwahati, northeast India: an indicator-based approach. In T. Chaiechi (Ed.), *Economic Effects of Natural Disasters* (pp. 457-475). United Kingdom: Academic Press.
25. Le, K. G., Liu, P., & Lin, L. (2020). Traffic accident hotspot identification by integrating kernel density estimation and spatial autocorrelation analysis: a case study. *International Journal of Crashworthiness*, 27(2): 543-553.
26. Lin, S.C., Ke, M.C., & Lo, C.M. (2017). Evolution of landslide hotspots in Taiwan. *Landslides*, 14:1491-1501.
27. Liu, G., Bi, Rutian., Wang, S., Li F., & Guo, G. (2013). The use of spatial autocorrelation analysis to identify PAHs pollution hotspots at an industrially contaminated site. *Environmental Monitoring and Assessment*, 185(11): 9549–9558.
28. Lo, D., Chau, K.W., Wong, S.K., McCord, M., & Haran, M. (2022). Factors affecting spatial autocorrelation in residential property prices. *Land*, 11(6):931.
29. López-Caloca, A.A., Tapia-Silva, F. O., Rivera, G. 2018. Sentinel-1 satellite data as a tool for monitoring inundation areas near urban areas in the Mexican tropical wet. In M. Glavan (Ed.), *Water Challenges of an Urbanizing World* (pp. 127-144). Croatia: Intech.
30. Mahanta, A., & Borgohain, P. (2022). Urban livability and contextual uncertainties: an assessment of livability through the lens of urban dwellers in Guwahati, India. *Journal of Infrastructure, Policy and Development*, 6(1): 1395.
31. Manjusree, P., Kumar, L.P., Bhatt, C.M., Rao, G.S., & Bhanumurthy, V. (2012). Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. *International Journal of Disaster Risk Science*, 3: 113-122.
32. Manning, N., Li Y., & Liu, J. 2023. Broader applicability of the metacoupling framework than Tobler's first law of geography for global sustainability: A systematic review. *Geography and Sustainability*, 4(1): 6-18.
33. Martinis, S., & Rieke, C. (2015). Backscatter analysis using multi-temporal and multi-frequency SAR data in the context of flood mapping at River Saale, Germany. *Remote Sensing*, 7(6):7732-7752.
34. Mahapatra, S.N., Pani, P., & Sharma, M. (2014). Rapid Urban Expansion and Its Implications on Geomorphology: A Remote Sensing and GIS Based Study, *Geography Journal*, 2014: 361459.
35. Nallathiga, R., Taneja, S., Gupta, A., & Gangal B. (2018). Sustainability of Urban Fringe Development and Management in NCT-Delhi: A Case Study. In: J. Mukherjee (Ed.), Sustainable Urbanization in India: Challenges and Opportunities Exploring Urban Change in South Asia (pp. 109-133). *Springer* : Singapore.
36. Potin, P., Rosich, B., Miranda, N., & Grimont, P. (2016). Sentinel-1 Mission Status. *Procedia Computer Science*, 100: 1297-1304.
37. Prasannakumar, V., Vijith, R., Charutha, N., Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia- Social and Behavioral Sciences*. 21: 317-325.
38. Rashiq, A., & Prakash, O. (2023). Assessment of spatio-temporal variability of climate in the lower Gangetic alluvial plain. *Environmental Monitoring and Assessment*, 195: 945.
39. Raspini, F., Bianchini, S., Ciampalini, A., Soldato, M.D., Solari, L., Novali, F., Conte, S.D., Rucci, A., Ferretti, A., & Casagli, N. (2018). Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites. *Scientific Reports*, 8: 7253.
40. Raza, O., Mansournia, M.A., Foroushani, A.R., & Holakouie-Naieni, K. (2020). Exploring spatial dependencies in the prevalence of childhood diarrhea in Mozambique using global and local measures of spatial autocorrelation. *Medical Journal of the Islamic Republic of Iran*, 34: 59.
41. Rousta, I., Doostkamian, M., Haghighi, E., Malamiri, H. R. G., & Yarahmadi, P. (2017). Analysis of spatial autocorrelation patterns of heavy and super-heavy rainfall in Iran. *Advances in Atmospheric Sciences*, 34: 1069-1081.
42. Sahin, E., Iban, M. C., & Bilgilioglu, S. S. (2024). Remote Sensing-Enabled Urban Growth Simulation Overlaid with AHP-GIS-Based Urban Land Suitability for Potential Development in Mersin Metropolitan Area, Türkiye. *Applied Sciences*, 14(8): 3484.
43. Saleh, B., & Rawashdeh, S. A. (2007). Study of urban expansion in Jordanian cities using GIS and remote sensing. *International Journal of Applied Science and Engineering*, 5(1):41-52.
44. Schmidt, K., Schwerdt, M., Miranda, N., & Reimann, J. (2020). Radiometric Comparison within the Sentinel-1 SAR constellation over a wide backscatter range. *Remote Sensing*, 12(5):854.
45. Sentinel User Guide. Retrieved from https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/definitions
46. Shariati, M., Mesgari, T., Kasraee, M., & Jahangiri-rad, M. (2020). Spatiotemporal analysis and hotspots detection of

COVID-19 using geographic information system (March and April, 2020). *Journal of Environmental Health Science and Engineering*, 18(2):1499-1507.
47. Skoda, P., & Adam, F. (2020). Knowledge Discovery in Big Data from Astronomy and Earth Observation Astrogeoinformatics. USA: *Elsevier Science*.
48. Sun, W., Shan, J., Wang, Z., Wang, L., Lu, D., Jin, Z., & Yu K. (2020). Geospatial analysis of urban expansion using remote sensing methods and data: a case study of Yangtze River delta, China. *Complexity*, 2020: 3239471.
49. Tao S., Zhang T., Wang K., Xie F., Ni L., Mei Z., Song S. (2021): Identification of the risk factors in perioperative respiratory adverse events in children under general anesthesia and the development of a predictive model. *Translational Pediatrics*, vol. 10(7), 2021, pp. 1877-1882. doi: 10.21037/tp-21-257
50. Tsui, T., Derumigny, A., Peck, D., Timmeren, A. V., & Wandl, A. (2022). Spatial clustering of waste reuse in a circular economy: A spatial autocorrelation analysis on locations of waste reuse in the Netherlands using global and local Moran's I. *Frontiers in Built Environment*, 8: 954642.
51. Vilinová, K., & Petrikovicová, L. (2023). Spatial autocorrelation of COVID-19 in Slovakia. *Tropical Medicine and Infectious Disease*, 8(6): 298.
52. Wang, M., Fu, X., Zhang, D., Chen, F., Su, J., Zhou, S., Li, J., Zhong, Y., & Tan S.K. (2023). Urban flooding risk assessment in the rural-urban fringe based on a bayesian classifier. *Sustainability*, 15:5740.
53. Weiβ, T. (2018). SAR pre-processing documentation. Retrieved from https://buildmedia.readthedocs.org/media/pdf/multiply-sar-pre-processing/get_to_version_0.4/multiply-sar-pre-processing.pdf
54. White, L., Brisco, B., Dabboor, M., Schmitt, A., & Pratt, A. (2015). A collection of SAR methodologies for monitoring wetlands. *Remote Sensing*, 7(6):7615-7645.
55. Winter, A.K., & Karvonen, A. (2022). Climate governance at the fringes: Peri-urban flooding drivers and responses. Land Use Policy, 117: 106124.
56. Woodhouse, I.H. (2006). Introduction to Microwave Remote Sensing. *CRC Press*:Florida.
57. Yin, Z., Dong, Y., Wang, Q., Ma, Y., Gao, Z., Ling, Z., Aihaiti, X., Abudusaimaiti, X., Qiu, R., Chen, Z., & Wushouer, F. (2024). Spatial-temporal evolution patterns of influenza incidence in Xinjiang Prefecture from 2014 to 2023 based on GIS. *Scientific Reports*, 14(1):21496.
58. Zhang, W., & Hu B. Brown G.S. (2020). Automatic surface water mapping using polarimetric SAR data for long-term change detection. *Water*, 12(3):872.