Science Forum (Journal of Pure and Applied Sciences)

Journal Homepage: https://atbuscienceforum.com.ng/index.php/jpas

Spatial Anomaly Identification utilizing an Adjacency-Weighted Spatial Outlier Factor (AWSOF)

Ali Mohammed Baba[1*], Isah Haruna[1], Abubakar Bishir[2]

[1]*Department of Biostatistics, Federal University of Health Sciences, Azare, Bauchi State, Nigeria*

[2]*Department of Statistics, Abubakar Tafawa Balewa University, Bauchi, Bauchi State, Nigeria*

Abstract

Spatial outlier detection remains a challenging problem because anomalous observations are inherently localized and strongly influenced by spatial coordinates and neighborhood structure. Conventional detection techniques are particularly susceptible to masking and swamping effects, whereby extreme neighboring observations either conceal true spatial anomalies or incorrectly classify normal observations as outliers. Furthermore, many existing approaches rely on directional neighborhood comparisons, such as left-right or unidirectional evaluations, limiting their ability to capture the multidirectional spatial relationships that characterize complex geographic patterns. To address these limitations, this study proposes a novel Adjacency-Weighted Spatial Outlier Factor (AWSOF) framework based on an Exponential Decay Weight (EDW) model that incorporates local topological adjacency into the construction of spatial proximity weights. The proposed weighting scheme quantifies the influence of spatial proximity on variations in behavioral attributes by assigning exponentially decreasing influence with increasing neighborhood distance. Absolute attribute differences are projected across coordinate-based neighborhood vectors to derive the AWSOF, thereby providing a robust multidirectional measure of spatial deviation while minimizing the influence of localized extremes. The performance of the proposed framework was rigorously evaluated through extensive simulation experiments and validated using multiple real-world spatial datasets. Comparative analyses against conventional spatial outlier detection methods demonstrate that the proposed exponential weighting framework consistently delivers superior detection accuracy, robustness, and reliability under varying spatial configurations. Empirical results reveal that the EDW-based AWSOF effectively mitigates masking and swamping effects while preserving neighborhood integrity, leading to more accurate identification of genuine spatial anomalies. In contrast, the standardized approach, which behaves similarly to an arithmetic mean, remains highly sensitive to extreme localized observations, resulting in distorted neighborhood influence and reduced diagnostic performance.

Keywords

Spatial Anomaly, Spatial Contiguity, Adjacency Weight, Masking Effect, Swamping Effect.

ISSN 1119-4618 JPAS Faculty of Science, ATBU Bauchi

Corresponding Author: Ali Mohammed Baba, *Department of Biostatistics, Federal University of Health Sciences, Azare, Bauchi State, Nigeria*

Email: ali.mohammed@fuhsa.edu.ng

1. Introduction

In classical statistical literature, an outlier is famously characterized by Hawkins (1980) as "an observation which deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism." As noted by Harris et al. (2014), these anomalies can stem from measurement errors, reporting discrepancies, or genuine physical phenomena that represent highly atypical behaviours. Failing to identify and isolate these observations can distort classical estimators, leading to biased parameter estimates and invalid statistical inferences.

Unlike their classical counterparts, spatial outliers are fundamentally defined by their geographic or coordinate positioning. They represent highly localized anomalies rather than global ones, disrupting the expected spatial autocorrelation and continuous distribution of attributes across neighbouring locations (Kou and Lu (2008)). Aggarwal (2013) characterizes a spatial outlier as a localized entity whose behavioural attributes differ significantly from those of its immediate geographical neighbours. Haining and Haining (2003) further observe that an attribute value might only appear extreme when analysed relative to its local coordinate neighbourhood on a map. Consequently, an observation can function as a spatial outlier even if its value lies well within the typical range of the global data distribution.

Efficient identification of these spatial anomalies has profound real-world implications. As emphasized by Kou and Lu (2008), detecting spatial outliers is critical for predicting extreme weather phenomena (such as hurricanes and tornadoes), identifying anomalous genetic markers or oncological cell formations, mapping traffic bottlenecks on highways, locating military targets via satellite imagery, discovering untapped oil reservoirs, and monitoring water contamination incidents.

Typically, spatial datasets contain two distinct categories of attributes:

1. Behavioural Attribute: The primary variable of interest measured at a given location (e.g., crime rates, wind speed, or surface temperature).
2. Contextual Attribute: The spatial or coordinates metrics (such as longitude, latitude, or boundary topology) that define the object's physical or network position.

While classical outlier detection methods are numerous, they cannot be directly applied to spatial contexts due to the presence of spatial dependence. Spatial anomaly detection approaches span both visual/graphical and rigorous quantitative tests. Common graphical tools include traditional scatterplots (Anselin (1994)) and Moran scatterplots (Anselin (1994)). Anselin (1994) established a quantitative alternative by extending the local anomaly tests to global measures of spatial autocorrelation, which gives rise to the Local Indicator of Spatial Autocorrelation (LISA). Under the null hypothesis of zero spatial autocorrelation, the sum of LISA statistics remains proportional to global autocorrelation metrics, allowing the derivation of a threshold to flag anomalies.

However, this method is highly susceptible to masking and swamping when neighbouring points possess highly polarized attribute values (Kou and Lu (2008)). To mitigate this, Shekhar et al. (2002) introduced the x spatial statistic, which computes the difference between an attribute value at location x and the average of its neighbours. Lu et al. (2003) identified that the primary vulnerability of Shekhar's approach lies in its susceptibility to masking and swamping too; clean observations are often misclassified as outliers, while true outliers are obscured by the aggregate neighbourhood mean.

In response, they proposed the Iterative Z, Iterative R, and Median Z algorithms. While Iterative Z and Iterative R identify outliers through recursive feedback loops, the Median Z method is a single-pass algorithm that replaces the vulnerable mean with the median. Nonetheless, these approaches were evaluated only on small-scale datasets, leaving their reliability on larger datasets unverified. Using a different strategy, Liu et al. (2010) focused on the masking/swamping challenge by designing two random-walk models: Random Walk on Bipartite Graph (RW-BP) and Random Walk on Exhaustive Combination (RW-EC).

These models use dual weighted graphs to compute and rank outlier scores, flagging the top k objects. However, these formulations lack a standardized, statistically derived threshold for declaring outliers. A recent work by Ren et al. (2025) introduced the second-dimension outliers (SDO) models that incorporate local outlier information at unsampled locations to enhance prediction accuracy. However, they focused on unsampled locations which are outside area of study to detect outliers. Singh and Lalitha (2018) introduced a Local Quotient (LQ) method to compare an attribute's value with its local

spatial neighbourhood and subsequently applying the iterative algorithms from (Lu et al. (2003)) to isolate outliers. Hadi and Imon (2018) introduced a spatial distance metric calculated across all n observations simultaneously.

However, this metric calculates differences using only the immediate left and right neighbours of a sequential observation. While useful for identifying linear clusters of contiguous anomalies, it fails to capture complex, multi-directional spatial relationships, exposing the model to masking and swamping. Driven by these theoretical and computational challenges, we propose a new spatial anomaly detection framework: the Adjacency Weighted Spatial Outlier Factor (AWSOF).

This method constructs spatial outlier indicators that capture spatial contiguity by integrating relative coordinate distances with the behavioural variations of neighbouring attributes. The AWSOF framework is designed specifically to address the prevalent issues of masking and swamping in large-scale spatial data settings.

2. Material Methods

2.1 Robust Spatial Z-Test

Hadi and Imon (2018) demonstrated that the traditional spatial Z-test formulated by Shekhar et al. (2002) struggles to identify multiple spatial outliers. They proposed an alternative spatial distance measure, d_i, defined for attribute values $Z_1, Z_2, \dots, Z_n$ as:

$$d_i = \begin{cases} |Z_{i+1} - Z_i|, & i = 1 \\ |Z_i - Z_{i-1}|, \ \ i = 2,3,\dots, & n-1 \\ |Z_i - Z_{i-1}|, & i = n \end{cases}$$

The standard spatial Z-test is then computed as:

$$z = \frac{d_i - \bar{d}_i}{std(d_i)} \quad (2)$$

where $\bar{d}_i$ and $std(d_i)$ denote the sample mean and standard deviation of d_i, respectively.

Recognizing that the sample mean and standard deviation are non-robust estimators easily distorted by outliers, Hadi and Imon (2018) formulated a robust Z_i-statistic by replacing these components with the median and the Normalized Median Absolute Deviation ($NMAD(d_i)$):

$$z = \frac{d_i - median(d_i)}{NMAD(d_i)} \quad (3)$$

where: $NMAD = \frac{MAD(d_i)}{0.6745}$ and $MAD(d_i) = Median\ (|d_i - Median(d_i)|)$. The scale factor 0.6745 ensures that the estimator is asymptotically unbiased for normally distributed data. In both formulations, an observation is flagged as an outlier if$|z > 3|$.

.

2.2 Proposed AWSOF Framework

We argue that the standard robust spatial Z-test is fundamentally limited because it does not incorporate complete spatial neighborhood contiguity. To address this, the proposed Adjacency Weighted Spatial Outlier Factor (AWSOF) utilizes pre-specified spatial weights to model the relative positions of behavioural attributes, so as mitigate the problem of masking and swamping.

2.2.1 Topological Contiguity Formulation

Define an *n* by *n* adjacency matrix, A_{ij}, such that:

$$A_{ij} = \begin{cases} 1 & if\ \ i\ \ and\ \ j\ \ are\ \ neighbours \\ 0 & otherwise \end{cases} \quad (4)$$

Let $Z_1, Z_2, \ldots, Z_n$denote the behavioral attributes measured across n coordinates. We construct an absolute pairwise variation matrix D where each element represents the absolute variation between locations i and j:

$$D = \begin{bmatrix} |Z_1 - Z_1| & \cdots & |Z_1 - Z_n| \\ \vdots & \ddots & \vdots \\ |Z_n - Z_1| & \cdots & |Z_n - Z_n| \end{bmatrix} \quad (5)$$

The diagonal entries represent self-comparison, therefore, they are identically zero ($d_{ii} = 0$).

2.2.2 Exponential Weight (EWR)

The exponential weight models spatial decay using a negative exponential function of distance. The Exponential Weight is given by:

$$exp(-(md)_{ij})$$

In matrix form, the weight matrix is given by the product:

$$W = A_{ij} * e_{ij}. \qquad (6)$$

2.3 Adjacency-Effect Matrix and Outlier Factor Calculation

Let W represent the weight matrix derived from either Equation (6). We define the Adjacency-Effect Matrix E as:

$$E = W * D. \qquad (7)$$

The Adjacency Weighted Spatial Outlier Factor, f_i, is then calculated as the sum of the elements in the i-th row of E:

$$f_i = \sum_i e_i \qquad (8)$$

Where e_{IJ} represents the elements of the adjacency-effect matrix$E$.

This formulation successfully integrates local spatial topology with attribute differences, in agreement with Tobler's First Law of Geography (Tobler (1970)). To execute the anomaly test, we substitute the traditional difference metric d_i in Equation (2) or (3) with our newly defined spatial factor f_i:

$$z_{awsof} = \frac{f_i - median(f_i)}{NMAD(f_i)} \qquad (9)$$

Any location exhibiting $|z_{awsof}| > 3$ is classified as a spatial outlier.

AWSOF Algorithm:

1. Construct the adjacency matrix A_{ij} using Equation (4).
2. Compute the matrix of absolute attribute differences D using Equation (5).
3. Select the weight matrix W in Equation (6).
4. Compute the Adjacency-Effect Matrix $E = W \bullet D$.
5. Obtain the spatial outlier factors f_i using Equation (8).
6. Calculate the robust z_i statistics using Equation (3) (replacing d_i with f_i).
7. Flag all locations with $|z_i| > 3$ as spatial outliers.

2.4 Simulation Framework

We generate synthetic spatial random fields on regular square grids of size $n \in \{36, 81, 100, 225\}$ following the formulation proposed by Marchetti et al. (2018):

$$Z(S) = Y(S) + \epsilon(S). \quad (10)$$

Where $Y \sim N(0, \sigma^2 \Sigma)$ and $\epsilon \sim N(0, \sigma^2 I)$. The covariance matrix, $\Sigma = [\sigma^2]_{N \times N}$ is constructed from an exponential model

$$\sigma_{ij}^2 = \begin{cases} \sigma_Y^2 \exp(-\left\|S_i - S_j\right\| & if \quad S_i \neq S_j \\ \sigma_Y^2 + \sigma_\epsilon^2 & if \quad S_i \neq S_j \end{cases}$$

with $\sigma_Y^2 = 1$ and $\sigma_\epsilon^{2^{\epsilon}} = 0.01$.

The spatial fields are parameterized to generate spatial autocorrelations of 0.51, 0.60, 0.78, and 0.86 for sample sizes of 36, 81, 100, and 225, respectively.

2.4.1: Contamination Procedure and Modelling

We introduce artificial anomalies within each dataset. The contamination values are placed such that low values are surrounded by high values, and vice-versa.

To evaluate classification performance, we train a Random Forest classifier (Breiman (2001)) using the contaminated datasets. Uncontaminated points are labelled as 0 (clean) and contaminated points as 1 (outliers). The classifier's predictions are evaluated using the Receiver Operating Characteristics (ROC) curve and Area Under the Curve (AUC) metrics (Fawcett (2006)).

3. Results and Discussion

3.1 Simulation Results

The robust spatial Z-test (Hadi & Imon (2018)) and our proposed Robust Exponential Weight (EWR)—were evaluated across all simulated spatial fields. The calculated AUC values are summarized in Table 2.

Table 2: Comparative Area Under the Curve (AUC) performance metrics across different grid configurations.

Sample Size (n)	Outliers Introduced	Hadi & Imon (2018)	EWR
36	4	0.91	0.94
81	8	0.47	0.87
100	10	0.90	0.94
225	15	0.88	0.94

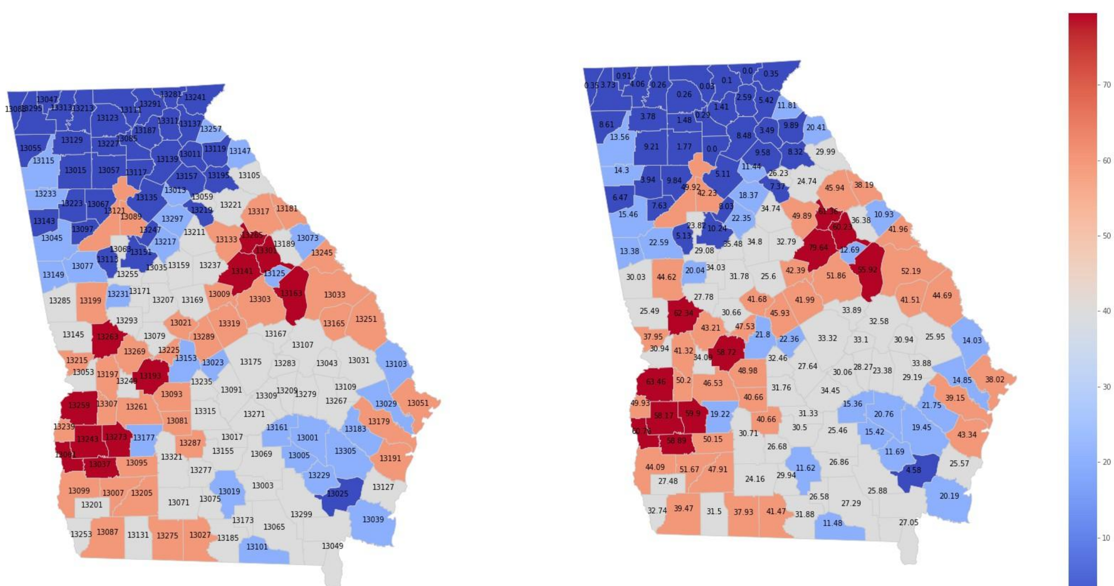
Case-by-Case Analysis:

- For $n = 36$: The standard robust Z-test detected 3 out of 4 anomalies resulting in an AUC of 0.91. The EWR frameworks achieved perfect detection of all 4 anomalies with no false positives, yielding an AUC of 0.94.
- For $n = 81$: The standard robust Z-test suffered severely from masking, identifying only 3 genuine anomalies while incorrectly flagging 1, resulting in an AUC of 0.47. Conversely, EWR detected all 8 anomalies, with AUC values of 0.87.
- For $n = 100$: The standard robust Z-test achieved an AUC of 0.90, detecting all 10 anomalies but incorrectly flagged 2 locations. The EWR model successfully identified all 10 anomalies with zero false positives, resulting in AUC values of 0.94.
- For $n = 225$: The standard robust Z-test correctly identified the 15 anomalies but suffered from severe swamping, incorrectly flagging 28 typical coordinates and resulting in an AUC of 0.88. The EWR frameworks exhibited superior control over swamping, flagging only 9 false positives and achieving AUC values of 0.94.

In summary, the proposed AWSOF formulations consistently outperformed the standard robust spatial Z-test across all grid sizes, demonstrating excellent resilience to masking and swamping.

3.2 Real-World Application: Georgia County Dataset

We validated our proposed methodologies using the proportion of African-American residents (PctBlack) across 159 counties in Georgia, originally compiled by Fotheringham et al. (2003) and processed using the PySAL library (R

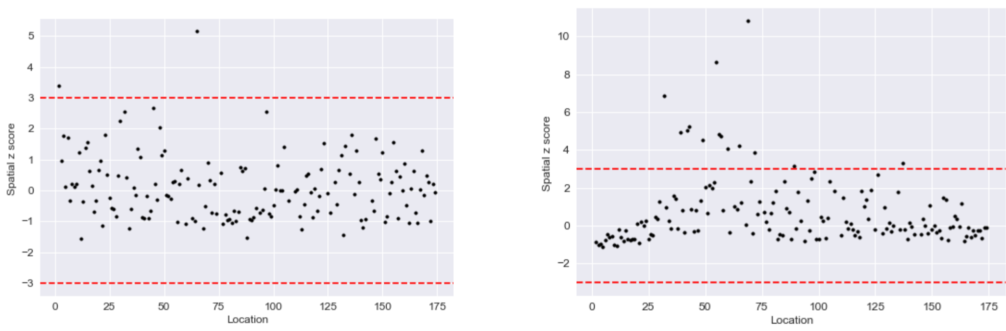


(a) With Location AreaKey **(b)** With Proportion of African Americans

Figure 1. Map of Georgia showing of African−Americans in Counties of Georgia

The map has some locations that look completely different from their neighbours; they lie on different extremes of the colourmap. Spatial z test, robust spatial z test and our proposed method were used to detect spatial anomalies in the PctBlack of the Georgia data.

The robust spatial z test method could only identify two outliers (see Figure 2a). The AreaKey for these locations are 13141 and 13241.



(a) Robust spatial test **(b)** Robust Exponential Weight (EWR)

Figure 2. Scatter plots with cut off points for the real example for Robust spatial z test and Robust

Key Findings from the Empirical Analysis:

1. False Positive Isolation (County 13241): The standard spatial Z-test and its robust version flagged County 13241 as an outlier (z = 3.15, attribute value = 0.35). However, a neighborhood checks reveals adjacent values of 0.00 and 5.42, demonstrating that 0.35 is entirely typical of its local environment. This highlights the swamping issue in Hadi & Imon's model, which was successfully avoided by all our proposed AWSOF variants.
2. True Outlier Identification (County 13141): All tested methods flagged County 13141 (attribute value = 79.64) as an outlier. This county is surrounded by much lower values (12.69, 32.79, 42.39, 49.89, 51.86, 60.23, and 61.36), confirming its status as a genuine spatial anomaly.
3. Neighbourhood Contrast Delineation (Counties 13121 & 13125): Our proposed methods successfully identified County 13121 (value = 49.92) and County 13125 (value = 12.69) as outliers. County 13121 is surrounded by low values (ranging from 0.00 to 23.82), while County 13125 is surrounded by high values (ranging from 51.86 to 79.64). These genuine anomalies were completely missed by the standard Z-tests due to masking.
4. Superior Performance of Robust Exponential Weight (EWR): The EWR method identified the most comprehensive list of localized anomalies (14 counties), capturing subtle variations in dense urban-rural boundaries. This confirms that incorporating exponential distance decay significantly improves the model's sensitivity to localized attribute variations.

4. Conclusion and Recommendation

This paper introduced the Adjacency Weighted Spatial Outlier Factor (AWSOF), a robust spatial anomaly detection framework designed to overcome the limitations of traditional spatial Z-tests. By projecting absolute attribute differences along coordinate-based adjacency vectors, our proposed framework successfully accounts for localized spatial contiguity. Simulation results and empirical evaluations using the Georgia dataset show that our proposed model provides superior protection against masking and swamping. The Exponential Weight model achieved the best diagnostic accuracy in real-world scenarios. These findings confirm that integrating multidirectional geographic distance decay is essential for robust spatial anomaly detection.

References

1. Aggarwal, C. C. (2013). Spatial outlier detection. In *Outlier Analysis*, Springer, 345–368.
2. Anselin, L. (1994). Exploratory spatial data analysis and geographic information systems. *New Tools for Spatial Analysis*, 17, 45–54.
3. Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115.
4. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
5. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
6. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
7. Hadi, A. S., & Imon, A. R. (2018). Identification of multiple outliers in spatial data. *International Journal of Statistical Sciences*, 16, 87–96.
8. Haining, R. P., & Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.
9. Hawkins, D. M. (1980). *Identification of Outliers*. Chapman and Hall.
10. Kou, Y., & Lu, C.-T. (2008). Outlier detection, spatial. *Encyclopedia of GIS*, 1539–1546.
11. Lu, C.-T., Chen, D., & Kou, Y. (2003). Detecting spatial outliers with multiple attributes. *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*, 122–128.
12. Marchetti, Y., Nguyen, H., Braverman, A., & Cressie, N. (2018). Spatial data compression via adaptive dispersion clustering. *Computational Statistics & Data Analysis*, 117, 138–153.
13. Rey, S. J., & Anselin, L. (2007). PySAL: A Python library of spatial analytical methods. *Handbook of Applied Spatial Analysis*, Springer, 175–193.
14. Shekhar, S., Lu, C.-T., & Zhang, P. (2002). Detecting graph-based spatial outliers. *Intelligent Data Analysis*, 6(5), 451–468.
15. Singh, A. K., & Lalitha, S. (2018). *A novel spatial outlier detection technique*. *Communications in Statistics - Theory and Methods*, 47(1), 247–257.
16. Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1), 234–240.
17. Ren, K., Song, Y., & Yu, Q. (2025). Second-dimension outliers for spatial prediction. *International Journal of Geographical Information Science*, 1–28.