SCIENTIFIC Publishing Limited

Journal of Hydrological Ecology and Water Security
https://ojs.ukscip.com/index.php/jhews

Article

Graph-Theoretical Pollution Connectivity Assessment in an Urbanizing Tropical Watershed: A GIS-AHP Framework for the Ala River Basin, Nigeria

Ayoola Emmanuel Awode [1,*] and James Rotimi Adewumi [2]

[1] Department of Civil and Environmental Engineering, Achievers University, Owo P.M.B. 1030, Nigeria

[2] Department of Civil and Environmental Engineering, Federal University of Technology Akure, Akure P.M.B. 704, Nigeria; jradewumi@futa.edu.ng

[*] Correspondence: awode.ea@achievers.edu.ng

Received: 14 January 2026; **Revised:** 28 February 2026; **Accepted:** 4 March 2026; **Published:** 26 March 2026

Abstract: Point-based physicochemical sampling dominates pollution risk assessment in rapidly urbanizing tropical watersheds, yet it cannot capture how river network structure channels pollutant transport, leaving a gap between where pollution is measured and where it structurally concentrates. This study developed a Geographic Information System (GIS)-based Pollution Connectivity Index (PCI) for the Ala River Basin, Akure, Ondo State, Nigeria, integrating the Analytical Hierarchy Process (AHP), remote sensing, and graph-theoretic network modelling. Unlike prior studies using undirected, unweighted networks, the river system was modelled as a directed, weighted graph. A conditioned 30 m Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) supported watershed delineation and Deterministic Eight-neighbour (D8) stream extraction; a sensitivity-tested threshold of 1,000 cells (0.9 km^2 contributing area) yielded a 2,234 km network across five Strahler orders. Betweenness centrality computed on the resulting graph (428 nodes, 427 edges) identified pollution-transfer hotspots in the urbanized central-northern basin and at tributary confluences. AHP weighting (Consistency Ratio (CR) = 0.023) assigned betweenness centrality the highest weight (41.6%), followed by population density (28.5%), inverse Normalized Difference Vegetation Index (NDVI) (13.6%), land cover (9.7%), and rainfall (6.6%). PCI correlated significantly with population density (r = 0.812), inverse NDVI (r = 0.771), and land cover vulnerability (r = 0.748; all $p < 0.05$), and chi-square testing confirmed significant spatial overlap between hotspots and built-up areas. Low multicollinearity among criteria supported the framework's robustness. Coupling directed, weighted graph theory with GIS-AHP offers a structurally grounded, adaptable approach for pollution vulnerability assessment in data-scarce, urbanizing tropical watersheds.

Keywords: Hydrological Connectivity; Graph Theory; Pollution Connectivity Index; AHP; Urban Watershed

1. Introduction

The degradation of the surface water systems in rapidly urbanizing regions is due to the effects of land-use change, hydrological alterations, and increased anthropogenic activity. Urbanization-induced changes in watershed processes include vegetation removal, installation of impervious surfaces, changes in channel dimensions, and increased pollutant loads in river systems [1]. Urban growth in developing countries often leads to faster growth of urban population than urban environmental governance capacity, wastewater treatment facilities and integrated

watershed management [2]. The urban rivers in many sub-Saharan cities create complex interactions between hydrological connectivity and pollution propagation, functioning as stormwater conveyance, as an informal waste-disposal system, and as a pathway for flooding. Urban expansion, vegetation removal and expansion of impervious surfaces have increasingly affected the hydrology and environmental health of river basins in Southwestern Nigeria, especially in this region.

Diffuse pollution is a constant problem for water quality management because pollutants come from areas with diffuse sources, and the water flows through interconnected hydrologic pathways [3]. The usual way of assessing pollution has been point based physicochemical sampling and has not considered the spatial structure of river connectivity, nor the cumulative effects of several environmental drivers [4]. These methods are not adequate to capture the pathways pollutants can follow through river systems or where pollutants can accumulate in ecologically sensitive areas. Recently, the concept of hydrological connectivity, which quantifies the linkage between watershed components via hydrological processes at the surface and subsurface, has become an important basis for understanding pollutant transport, flood dynamics, and ecosystem resilience in urban watersheds [5].

With recent developments in the field of graph theory and network science, new opportunities have emerged for the development of river network models as connected spatial networks. Graph-based hydrological modelling has been proposed as a mathematical graph, where nodes represent the river's confluences and edges represent the links between the river's segments, allowing for quantitative analysis of the topology and transfer efficiency of the network, as well as how the river is organised around connectivity [4]. Graph-theoretical metrics are useful, such as betweenness centrality, which indicates nodes that are structurally central in controlling the flow of water and contaminants throughout the network, highlighting important transfer pathways [6]. Applications of graph theory to hydrological and water-infrastructure networks have expanded internationally in recent years, including urban hydrological connectivity assessments in rapidly urbanizing Asian catchments [7] and graph-based analyses of water-supply network configuration and urban morphology in African cities such as Kisumu, Kenya, and Kigali, Rwanda [8]. However, these studies have generally either represented the network as undirected and unweighted or addressed water-supply infrastructure rather than pollution-transfer connectivity, leaving open the specific methodological gap this study addresses. Although the use of graph theory in the field of hydrology is therefore growing globally, its application to measure pollution connectivity specifically in the watersheds of African cities remains limited, and most of the existing studies worldwide have relied on an undirected, unweighted representation of graphs, which is unable to capture the directional character of river flow.

At the same time, the development of GIS, remote sensing and freely available environmental data has enhanced the ability to model the watershed in areas where data is limited. Reproducible computational workflows [2] can be used to integrate HREWMs, satellite-derived land-cover products with vegetation indices, gridded rainfall, population data, to create integrated environmental assessments. These methods are increasingly accompanied by a Multi-Criteria Decision Analysis (MCDA) framework with techniques like the Analytical Hierarchy Process (AHP) to incorporate multiple and heterogeneous environmental variables into a weighted vulnerability index [3]. Most of the current GIS-AHP watershed studies, however, have been limited to consideration of terrain attributes and have not considered graph-theoretic connectivity measures that include pollution transfer corridors and hydrologically important river junctions.

Ala River Basin, Akure, Ondo State is one of the most stressed urban river systems in the Southwestern part of Nigeria. The basin has undergone rapid urbanization in the last 20 years with a population that has grown, land use changes, riparian encroachment, degradation of vegetation, and expansion of impervious surfaces which have affected the spatial and temporal hydrology of the basin and the frequency of flooding. However, there are currently limited quantitative data on hydrological connectivity structures that determine the flow of pollution within the basin [9]. This study tests the hypothesis that structurally central positions within a directed, weighted representation of the river network—rather than land cover or population density alone—identify locations of disproportionate pollution-transfer risk, and that this structural signal provides information that is not redundant with the socioeconomic and ecological drivers conventionally used in GIS-AHP vulnerability assessments. The specific objectives of this study are to: (i) characterise the hydrological structure of the Ala River Basin through a sensitivity-justified, hydrologically realistic stream network, providing a transparent basis for the graph-theoretic analysis that follows; (ii) quantify structurally critical pollution-transfer pathways within the basin by representing the river system as a directed, weighted graph and computing betweenness centrality, addressing the directional and weight-

ing limitations of prior undirected graph-based approaches; (iii) determine the relative contribution of structural network centrality versus land cover, vegetation, rainfall, and population pressure to basin-wide pollution vulnerability through an AHP-weighted Pollution Connectivity Index; (iv) resolve spatial pollution-vulnerability zonation at both fine (five-class) and coarse (three-class) resolution to support differentiated management response; and (v) establish the statistical robustness and independence of the PCI framework's criteria through correlation, chi-square spatial overlay, and multicollinearity diagnostics, clarifying the extent to which the framework's structural and socioeconomic components provide non-redundant information.

2. Study Area

The study was carried out in the Ala River Basin, which is focused on Akure, Ondo State, Southwestern Nigeria. Akure is located in the humid tropical rainforest zone and has experienced rapid urbanisation over the last 20 years due to population growth, infrastructure development and land-use change. The main urban drainage system in the city is the Ala River which flows through peri-urban, agricultural, commercial and residential areas across the watershed [10].

The delineated basin is located between latitudes 7.05° N and 7.45° N and longitudes 5.06° E and 5.47° E (UTM Zone 32N, EPSG:32632) with an area of 687.5 km^2. Elevations range from about 0 m to 655 m above mean sea level with the higher levels occurring in the northwest and lower levels in the southeast drainage outlet. Mean basin elevation is 110.8 m and the hypsometric integral is 0.169, which are characteristics of a geomorphologically adjusted drainage system. The length of the basin is 70.2 km and the basin perimeter is 280.9 km. The elongation ratio (Re = 0.421) and the circularity ratio (Rc = 0.109) suggest a very elongated basin shape that is characteristic of long flood hydrographs. Land cover is very heterogeneous and includes a dominant tree cover, grassland, cropland, built-up and small wetland elements. The expansion of impervious surfaces, alteration of the hydrology of the drainage system, sedimentation and recurrent flooding are the major environmental issues especially along the lower reaches of the drainage system [11]. The mean annual rainfall over the basin is around 1,200–1,500 mm, with well-defined rainy seasons (April–July, September–November), which cause episodic high intensity runoff events.

3. Methodology

3.1. Research Design and Data Acquisition

This study adopted an integrated geospatial and hydrological modelling framework combining GIS, hydrological modelling, graph theory, remote sensing, and MCDA. The methodological workflow comprised: DEM conditioning and watershed delineation; stream network extraction with threshold sensitivity analysis; directed weighted river-network graph construction and betweenness centrality computation; environmental variable extraction and normalization; AHP-based weighting; PCI modelling; and spatial risk classification with statistical validation including correlation analysis, chi-square overlay testing, and VIF diagnostics. The entire workflow was implemented in Python 3 using PySheds, Rasterio, GeoPandas, NetworkX, NumPy, SciPy, and Shapely, with Google Earth Engine (GEE) employed for raster acquisition and preprocessing and all data used are shown in **Table 1**.

Table 1. Geospatial datasets used in the study.

Dataset	Source	Resolution	Purpose
SRTM DEM	USGS Earth Explorer	30 m	DEM conditioning, watershed delineation, flow routing
ESA WorldCover	European Space Agency	10 m	Land-cover classification and vulnerability weighting
Sentinel-2 SR	Copernicus Open Access Hub	10 m	NDVI computation (2023 cloud-masked median composite)
WorldPop Population	WorldPop (2020)	100 m	Anthropogenic pressure quantification
CHIRPS Rainfall	Climate Hazards Group	~5 km	Mean annual rainfall (2015–2023)

3.2. DEM Conditioning and Hydrological Preprocessing

Hydrological preprocessing was conducted using the SRTM DEM at 30 m spatial resolution. Raw DEMs contain artificial sinks, depressions, and flat areas that interrupt realistic flow routing. DEM conditioning followed a three-stage protocol: (i) pit filling to remove isolated single-cell depressions; (ii) depression filling to resolve multi-cell closed basins; and (iii) flat resolution to ensure positive drainage gradients across all cells. This conditioning sequence was implemented using PySheds and Rasterio. The 30 m spatial resolution of the SRTM DEM represents an

acknowledged limitation for urban watershed analysis where drainage features such as culverts, roadside channels, and minor streams may be sub-grid scale; this constraint is further discussed in Section 5.

The D8 (deterministic eight-direction) algorithm was applied for flow direction computation because it provides computationally efficient, deterministic representation of downslope water movement within raster-based hydrological systems [12]. Flow direction was assigned by identifying the steepest descent direction among eight neighbouring cells:

$$\text{Flow Direction} = \text{argmax}_i \left(\frac{\Delta z_i}{d_i} \right) \quad (1)$$

where Δz_i is the elevation difference between the central cell and neighbouring cell i, and d_i is the distance between cell centres (30 m for cardinal directions; $30\sqrt{2}$ m for diagonal neighbours).

The D8 algorithm was selected over multiple-flow-direction alternatives such as D-infinity (D∞) and MFD-based methods primarily for computational tractability and compatibility with the discrete, non-branching flow-path representation required for constructing a directed graph in which each stream segment maps to a single edge. This choice is not without trade-offs: comparative evaluations of flow-routing algorithms have shown that single-flow-direction methods such as D8 can produce larger position and area errors in extracted flow paths than multiple-flow-direction alternatives, whereas MFD approaches introduce artificial flow dispersion that is undesirable for the crisp, edge-based graph representation adopted here [13,14]. Given the elongated, structurally controlled morphology of the Ala Basin (Section 4.1) and the objective of producing a traversable directed graph rather than a continuous flow-dispersion surface, D8 was judged the more appropriate choice for this study; future refinements of the PCI framework could nonetheless benefit from testing D∞-derived stream networks to evaluate the sensitivity of the resulting graph topology and betweenness centrality hotspots to the choice of routing algorithm.

3.3. Watershed Delineation

Watershed delineation was performed using the conditioned DEM and derived D8 flow-direction raster. The pour point (basin outlet) was automatically identified as the point of maximum flow accumulation value within the extracted stream network and subsequently snapped to the nearest high-accumulation cell to ensure accurate delineation. PySheds was applied to delineate the catchment polygon, which served as the spatial mask for all subsequent raster clipping and environmental variable extraction. The delineated catchment was vectorized using Rasterio feature extraction with uint8 raster encoding to ensure compatibility with the polygonization function.

3.4. Stream Network Extraction and Threshold Sensitivity Analysis

The stream network was extracted from the flow-accumulation raster using a threshold-based stream initiation criterion. Recognizing that the choice of flow accumulation threshold is a source of methodological uncertainty—a concern raised explicitly by reviewers of the prior version of this study—a formal sensitivity analysis was conducted across five thresholds (500, 1,000, 2,000, 3,000, and 5,000 cells), as shown in **Table 2** and **Figure 1**. The stream initiation criterion is expressed as:

$$\text{Stream Mask} = \{x : A(x) \geq T\} \quad (2)$$

where A(x) is the flow accumulation value at cell x and T is the selected threshold in number of contributing cells. The threshold of T = 1,000 cells, equivalent to a contributing area of approximately 0.9 km^2, was selected as producing the most hydrologically coherent and drainage-realistic network for the 687.5 km^2 basin, yielding 1,266 stream segments with a total length of 2,234 km and a maximum Strahler order of 5.

Table 2. Flow accumulation threshold sensitivity analysis results.

Threshold (Cells)	Contributing Area (ha)	Stream Cells	Approx. Length (km)	Assessment
500	45	129,263	3,877.9	Excessive fragmentation; overrepresentation of minor tributaries
1,000 ✓	90	59,005	2,234.2	Balanced, hydrologically realistic drainage structure (selected)
2,000	180	33,236	997.1	Reduced minor tributary representation; acceptable
3,000	270	26,642	799.3	Loss of secondary tributaries; oversimplified
5,000	450	20,036	601.1	Significant loss of hydrologically important stream segments

Note: ✓: Chosen threshold.

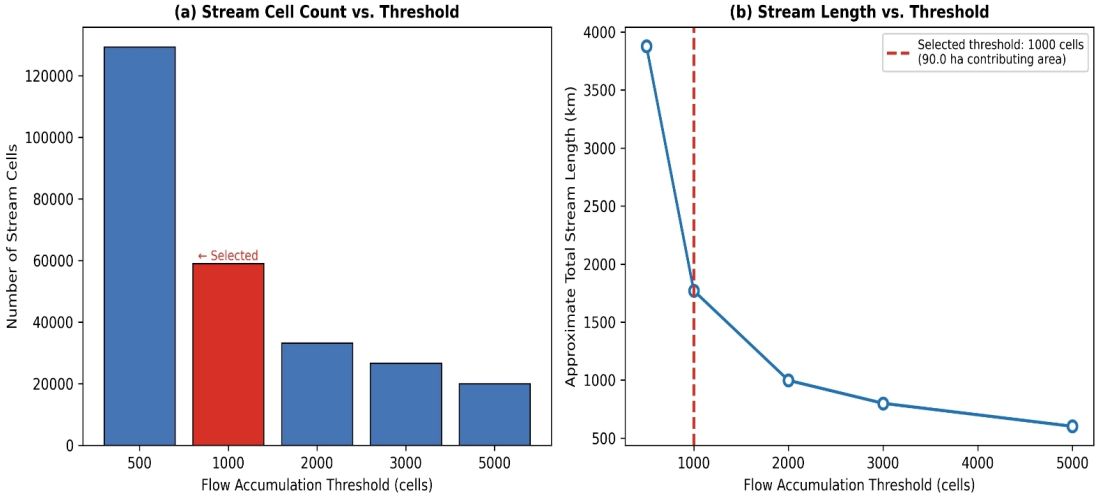


Figure 1. Flow accumulation threshold sensitivity analysis for the Ala River Basin showing (**a**) stream cell count and (**b**) total stream length as a function of five evaluated thresholds. The selected threshold of 1,000 cells is highlighted in red.

3.5. Basin Morphometric Analysis

A comprehensive morphometric analysis was conducted following the delineation of the watershed boundary and extraction of the stream network. Morphometric indices were computed to characterise the basin geometry, drainage organization, and hydrological behaviour. The drainage density (Dd) and stream frequency (Fs) were computed as:

$$Dd = \frac{\Sigma L}{A} \tag{3}$$

$$Fs = \frac{N}{A} \tag{4}$$

where ΣL is the total stream length (km), N is the total number of stream segments, and A is the basin area (km^2). The elongation ratio (Re), form factor (Rf), and circularity ratio (Rc) were computed as indicators of basin shape:

$$Re = \left(\frac{2}{Lb}\right) \times \sqrt{\frac{A}{\pi}} \tag{5}$$

$$Rf = \frac{A}{Lb^2} \tag{6}$$

$$Rc = \frac{4\pi A}{P^2} \tag{7}$$

where Lb is the basin length (km) and P is the basin perimeter (km). The hypsometric integral (HI), which characterises the distribution of elevation within the basin, was computed as:

$$HI = \frac{E_{mean} - E_{min}}{E_{max} - E_{min}} \tag{8}$$

where E_{mean}, E_{min}, and E_{max} are the mean, minimum, and maximum basin elevations respectively. Strahler stream ordering was assigned through topological traversal of the directed graph, following the standard convention that headwater streams are assigned order 1, and order increases at confluences where two streams of equal order join.

3.6. Directed Weighted River Network Graph Construction

The vectorized stream network was converted into a directed, weighted graph using NetworkX, directly addressing the reviewer concern regarding the use of an undirected, unweighted graph in the prior version. Directionality was assigned based on the D8 flow routing: each edge was oriented from upstream to downstream nodes,

preserving the natural flow direction of the river system. Edge weights were assigned using the inverse of mean flow accumulation along each segment, such that main-channel segments with high accumulation values receive lower weight (lower traversal cost), reflecting their greater hydraulic capacity and higher pollutant-transfer efficiency:

$$w(e) = \frac{1}{A_{mean}(e) + 1} \tag{9}$$

where $w(e)$ is the weight of edge e and $A_{mean}(e)$ is the mean flow accumulation value sampled along the edge. This inverse-flow-accumulation weighting follows established practice in graph-theoretic structural-connectivity assessments, where edge weights derived from a physical transfer-capacity proxy (e.g., contributing area, channel width, or discharge) have been shown to meaningfully differentiate high- and low-capacity transfer pathways within a hydrological network [15]. Alternative weighting schemes based on channel width or bankfull discharge estimates were considered but not implemented in this iteration owing to the absence of consistent channel-geometry survey data across the 687.5 km^2 basin; the sensitivity of the resulting betweenness centrality surface to the choice of edge-weighting scheme is discussed as a priority for future refinement in Section 5. Connected-component analysis was performed to retain the dominant hydrologically connected component and remove isolated raster artefacts generated during vectorization. **Table 3** summarises the graph filtering statistics, providing the documentation required to justify the node reduction ratio.

Table 3. Directed river network graph statistics before and after connected-component filtering.

Metric	Value
Initial nodes	1,486
Initial edges	1,266
Weakly connected components	220
Main component nodes (retained)	428
Main component edges (retained)	427
Nodes retained (%)	28.8%
Edges retained (%)	33.7%
Graph density	0.0023
Is directed acyclic graph (DAG)	Yes
Average in-degree	0.998

3.7. Betweenness Centrality Analysis

Betweenness centrality was computed across all nodes of the main connected component to identify hydrological bottlenecks and critical pollution-transfer pathways. Betweenness centrality quantifies the proportion of weighted shortest paths between all node pairs that pass through a given node:

$$C_B(v) = \Sigma(s \neq v \neq t)\left[\frac{\sigma(st)(v)}{\sigma(st)}\right] \tag{10}$$

where $\sigma(st)$ is the total number of weighted shortest paths between nodes s and t, and $\sigma(st)(v)$ is the number of those paths passing through node v. Normalization was applied by dividing by the maximum observed betweenness centrality value. Computation was performed using the weighted Dijkstra algorithm implementation in NetworkX, which incorporates edge weights as defined in Equation (9). Nodes with high betweenness centrality represent structurally critical positions within the network where the largest proportions of inter-node shortest paths converge. It is important to note that betweenness centrality is a structural topological metric; it does not directly represent measured pollutant concentrations or hydrodynamic flow magnitudes but serves as a proxy for pollution-transfer potential under the assumption that topological centrality correlates with actual transfer efficiency.

3.8. Environmental Variable Preparation

Five environmental variables were prepared and integrated into the PCI framework. All rasters were reprojected and resampled to match the 30 m reference grid using bilinear resampling for continuous variables and nearest-neighbour resampling for categorical land cover. All continuous variables were normalized to the 0–1 range using robust min-max normalization based on the 1st and 99th percentiles to reduce outlier influence:

$$X_{norm} = \frac{X - X_{p1}}{X_{p99} - X_{p1}} \tag{11}$$

NDVI was derived from a Sentinel-2 surface reflectance median composite for 2023, cloud-masked at less than 20% cloud cover:

$$\mathrm{NDVI} = \frac{\mathrm{B_{NIR}} - \mathrm{B_{Red}}}{\mathrm{B_{NIR}} + \mathrm{B_{Red}}} \tag{12}$$

The NDVI layer was inverted ($1 - \mathrm{NDVI_{norm}}$) prior to integration so that areas of low vegetation, associated with higher runoff and reduced ecological buffering, produce higher PCI contributions. A single-year (2023) cloud-masked median composite was used, consistent with the ESA WorldCover and WorldPop reference years used for the other criteria, to maintain temporal consistency across the five-criterion PCI framework; median compositing itself is a robust, outlier-resistant approach validated for vegetation-index mapping in cloud-prone tropical environments [16]. It is acknowledged that a single-year composite cannot capture interannual vegetation variability driven by rainfall fluctuations in this humid tropical setting, and a multi-year (e.g., 3–5 year) median composite is recommended in future iterations of the PCI framework to reduce sensitivity to any anomalous conditions in the reference year.

Land cover (ESA WorldCover 2021) was reclassified into pollution susceptibility scores based on hydrological runoff generation capacity and direct pollution loading potential: Built-up = 0.9, Permanent water = 0.7, Bare/sparse = 0.6, Herbaceous wetland = 0.6, Cropland = 0.5, Grassland = 0.4, Shrubland = 0.3, Tree cover = 0.1. The resulting land-cover vulnerability layer reflects the relative pollution-generation potential of each surface type within the watershed context. These scores were assigned through expert judgment informed by the relative runoff-generation and pollutant-loading characteristics established for each land cover class in comparable GIS-AHP watershed vulnerability studies, in which built-up and bare/sparse surfaces are consistently weighted highest owing to high imperviousness and low infiltration capacity, and vegetated classes are weighted lowest owing to their runoff-attenuation and pollutant-retention function [17]. As with any expert-judgment-based reclassification, these scores are a defensible convention rather than an empirically calibrated scale for the Ala Basin specifically; validating them against locally measured runoff or pollutant-loading coefficients is identified as a priority for future work.

Betweenness centrality node values were rasterized onto the 30 m grid using a 5 × 5 cell neighbourhood maximum filter and subsequent Gaussian smoothing ($\sigma = 2$), ensuring that point-located node values influenced their surrounding drainage corridor cells at a realistic spatial scale.

3.9. AHP Weighting and Consistency Analysis

The Analytical Hierarchy Process was applied to determine the relative importance of the five environmental variables within the PCI framework [18]. Pairwise comparisons were conducted using the Saaty 1–9 scale. The priority weight vector was derived using the geometric mean method:

$$\mathrm{g_i} = \left(\prod_{\mathrm{j=1}}^{\mathrm{n}} \mathrm{a_{ij}}\right)^{\frac{1}{\mathrm{n}}} \tag{13}$$

$$\mathrm{w_i} = \frac{\mathrm{g_i}}{\prod_{\mathrm{k=1}}^{\mathrm{n}} \mathrm{g_k}} \tag{14}$$

where $\mathrm{a_{ij}}$ is the pairwise comparison ratio between criteria i and j, and $\mathrm{g_i}$ is the geometric mean of row i. Internal consistency of the pairwise judgements was evaluated using the Consistency Ratio:

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}} \tag{15}$$

$$\mathrm{CI} = \frac{\lambda_{\mathrm{max}} - \mathrm{n}}{\mathrm{n} - 1} \tag{16}$$

where λ_{max} is the principal eigenvalue of the pairwise comparison matrix, n is the matrix order (5 in this study), and RI is the random consistency index for n = 5 (RI = 1.12 [18]). The computed CR of 0.023 is well below the 0.10 acceptability threshold, confirming internal consistency of the pairwise judgments. **Table 4** presents the full

pairwise comparison matrix and derived weights. **Table 5** shows the final normalized weights and their consistency verification.

Table 4. AHP pairwise comparison matrix and derived criterion weights.

Criterion	BC	NDVI	LC	Rain	Pop
Betweenness Centrality (BC)	1.0	3.0	4.0	5.0	2.0
NDVI (inverted)	0.33	1.0	2.0	2.0	0.33
Land Cover (LC)	0.25	0.50	1.0	2.0	0.33
Rainfall	0.20	0.50	0.50	1.0	0.25
Population Density (Pop)	0.50	3.0	3.0	4.0	1.0

Table 5. Final AHP normalized weights and consistency verification.

Criterion	Weight	Weight (%)	Rank
Betweenness Centrality	0.4160	41.6%	1
Population Density	0.2846	28.5%	2
NDVI (inverted)	0.1358	13.6%	3
Land Cover Vulnerability	0.0971	9.7%	4
Rainfall	0.0665	6.6%	5

Note: λ_{max} = 5.104 | CI = 0.026 | CR = 0.023 < 0.10 ✓.

3.10. Pollution Connectivity Index Construction

The PCI was constructed as a linear weighted composite of the five normalized criterion layers:

$$PCI = w_{BC} \times C_B^{norm} + w_{NDVI} \times \left(1 - NDVI^{norm}\right) + w_{LC} \times LC_{vuln} + w_{Rain} \times Rain^{norm} + w_{Pop} \times Pop^{norm} \quad (17)$$

The raw PCI surface was subsequently rescaled to a 0–100 range for interpretability. PCI values were classified using quantile classification into five vulnerability classes (Class 1 = Very Low through Class 5 = Very High), with a simplified three-class representation (Low: Classes 1–2; Medium: Class 3; High: Classes 4–5) for visual presentation. It is acknowledged that quantile classification ensures equal pixel counts per class rather than physically calibrated thresholds; results should therefore be interpreted as relative vulnerability rankings within this watershed rather than absolute pollution-risk designations.

3.11. Statistical Analysis and Model Validation

Several levels of statistical validation were performed. Random sampling of 50,000 valid pixels was used to calculate the Pearson (r) and Spearman (ρ) correlation coefficients between PCI and each criterion variable. Two-tailed tests of significance were applied at α = 0.05 level. In order to tackle the reviewer's concern that the urbanization signal might be double-weighted because of the high correlation between betweenness centrality and population density, a Variance Inflation Factor (VIF) analysis was performed for all criteria:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (18)$$

where R_i^2 is the coefficient of determination from regressing criterion i on all remaining criteria. VIF values below 5 indicate acceptable multicollinearity; values between 5 and 10 indicate moderate concern; values exceeding 10 indicate severe multicollinearity requiring remediation. Spatial ground-truthing of identified PCI hotspots was performed using chi-square testing to evaluate whether the overlap between high-PCI zones and built-up areas was statistically significant or attributable to spatial randomness:

$$\chi^2 = \frac{\Sigma\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \quad (19)$$

where O_{ij} and E_{ij} are the observed and expected cell frequencies respectively in the 2 × 2 contingency table of hotspot presence × built-up area presence.

4. Results

4.1. Watershed Delineation and Terrain Analysis

The conditioned DEM was successfully applied to delineate the Ala River Basin, yielding a coherent watershed boundary with a total area of 687.5 km^2 as shown in **Figure 2**. Terrain analysis identified a clear northwest-to-southeast elevation gradient, with higher elevations (500–655 m a.s.l.) concentrated in the northwestern section and lower elevations (0–100 m) in the southeastern drainage outlet zone. **Table 6** presents the full morphometric characterisation of the basin. The hypsometric integral of 0.169 indicates a geomorphologically mature basin in which the majority of mass has been removed through long-term fluvial erosion, consistent with the humid tropical setting. The low form factor (Rf = 0.139) and elongation ratio (Re = 0.421) confirm the strongly elongated basin morphology, which is characteristic of basins formed along structural lineaments and tends to produce slower, more prolonged flood hydrographs compared to circular basins.

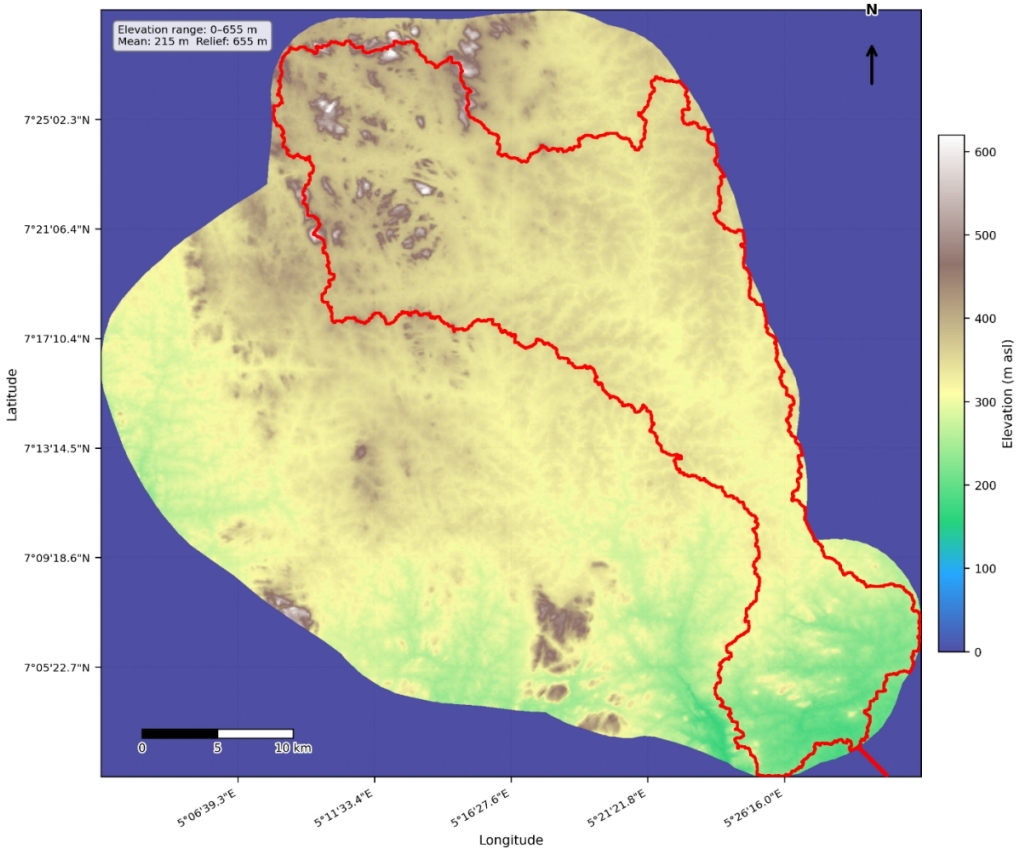


Figure 2. SRTM DEM of the Ala River Basin showing elevation gradients (0–655 m a.s.l.) and the delineated watershed boundary.

Table 6. Basin morphometric parameters of the Ala River Basin.

Morphometric Parameter	Value
Basin Area (km^2)	687.46
Basin Perimeter (km)	280.92
Basin Length (km)	70.23
Total Stream Length (km)	2,234.17
Number of Stream Segments	1,266
Maximum Strahler Order	5
Mean Bifurcation Ratio (Rb)	1.493
Drainage Density, Dd (km/km^2)	3.25
Stream Frequency, Fs (streams/km^2)	1.842
Form Factor (Rf)	0.139
Circularity Ratio (Rc)	0.109

Table 6. *Cont.*

Morphometric Parameter	Value
Elongation Ratio (Re)	0.421
Compactness Coefficient (Cc)	3.022
Maximum Elevation (m)	655.0
Minimum Elevation (m)	0.0
Mean Elevation (m)	110.8
Relief (m)	655.0
Relief Ratio (Rh)	0.0093
Hypsometric Integral (HI)	0.169

4.2. Stream Network Extraction and Strahler Ordering

The stream network extracted at the selected 1,000-cell threshold comprised 1,266 segments across five Strahler orders, with a total network length of 2,234.17 km and a drainage density of 3.25 km/km^2 (**Table 7**). First-order streams dominated numerically (203 segments) and contributed the greatest total length (799.6 km), consistent with typical dendritic drainage organization in humid tropical catchments. Fifth-order streams (89 segments; 124.1 km total length) represented the main channel corridors. The bifurcation ratio of 1.493 is lower than the theoretical 3–5 range reported for most natural drainage basins, reflecting the elongated, structurally controlled geometry of the Ala River system. The drainage density of 3.25 km/km^2 is consistent with values reported for urbanizing humid tropical basins where impervious surface expansion has accelerated runoff and channel extension [19].

Table 7. Strahler stream order statistics for the Ala River Basin.

Strahler Order	Number of Segments	Total Length (km)	% of Total Length
1	203	799.6	35.8
2	548	725.3	32.5
3	248	345.3	15.5
4	178	239.9	10.7
5	89	124.1	5.6
Total	1,266	2,234.2	100.0

4.3. River Network Graph Construction and Filtering

The vectorized stream network was converted into a directed, weighted graph. The initial graph contained 1,486 nodes and 1,266 edges distributed across 220 weakly connected components. Connected-component filtering retained the dominant hydrologically connected component of 428 nodes and 427 edges, representing a node retention rate of 28.8%. The 71.2% of nodes removed comprised isolated raster artefacts and disconnected stream fragments generated during vectorization, not hydrologically significant pathways. This was confirmed by the graph structure of the removed components: the 219 removed components averaged approximately 4.8 nodes each (1,058 removed nodes across 219 components), consistent with small, disconnected fragments rather than hydrologically meaningful pathways. The retained graph is a directed acyclic graph (DAG) with an average in-degree and out-degree of 0.998, consistent with a well-structured dendritic drainage network. The graph density of 0.0023 reflects the sparse, hierarchical connectivity structure typical of river networks. Field-mapped channel heads are not currently available for the Ala Basin, so the 1,000-cell threshold could not be validated against an empirically observed minimum channel-initiation point; instead, the threshold was justified through the five-level sensitivity analysis in Section 3.4 and by analogy with thresholds reported for comparable humid tropical basins in the literature [19]. This remains an acknowledged limitation, and the regression-based or fractal-based threshold-estimation procedures discussed in Section 5 offer a more rigorous alternative once field-mapped channel-head data become available.

4.4. Betweenness Centrality and Hotspot Identification

Betweenness centrality analysis identified a spatially concentrated pattern of high-centrality nodes within the directed weighted graph. The maximum normalized betweenness centrality value was 0.061, with a mean of 0.007 across the 428 retained nodes, indicating a highly skewed distribution in which a small number of nodes exert disproportionate structural control over network-wide connectivity. **Table 8** presents the full centrality correlation matrix, which confirms that betweenness centrality is strongly correlated with closeness centrality (r = 0.903) but

exhibits near-zero correlation with out-degree (r = 0.023), indicating that highly central nodes in this directed network are not simply high-degree hubs but rather are positioned at topological bottlenecks that control flow across the basin. The spatial distribution of betweenness centrality across the 428 retained nodes is presented in **Figure 3**, showing both the five-class quantile classification (**Figure 3a**) and the three-class aggregation (**Figure 3b**); high-centrality nodes are concentrated along the eastern drainage corridor and in the northern confluence zones of the watershed, forming a structurally continuous chain of potential pollution-transfer bottlenecks that tracks the main channel axis from north to southeast.

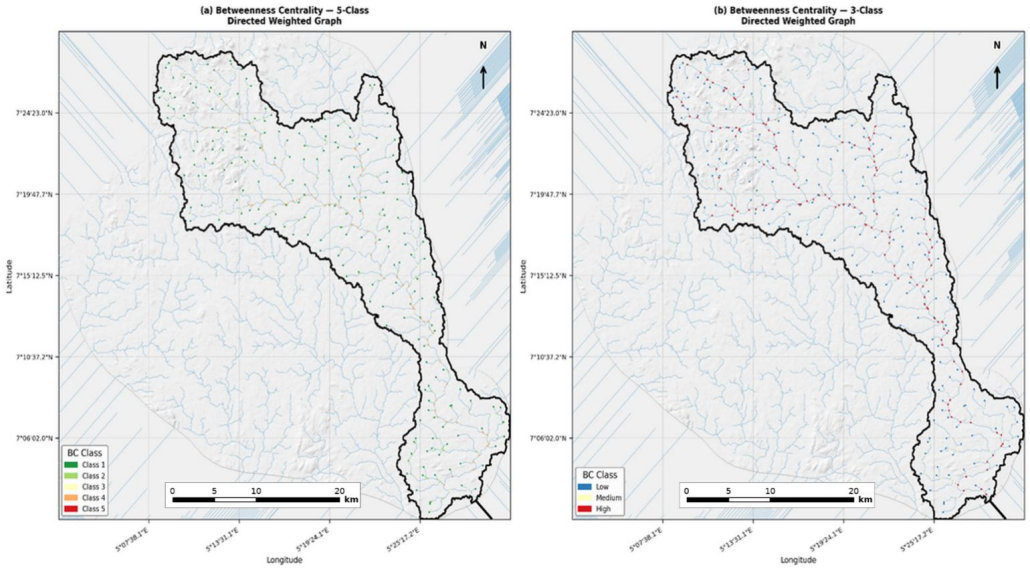


Figure 3. Spatial distribution of graph betweenness centrality across the Ala River Basin drainage network.

Note: Node-level betweenness centrality computed from the directed weighted river network graph (428 nodes, 427 edges). **Figure 3a** shows the five-class quantile classification (Class 1 = Very Low through Class 5 = Very High); **Figure 3b** shows the three-class aggregation (Low, Medium, High). High-centrality nodes (red) are concentrated along the eastern drainage corridor and at tributary confluences in the northern section, identifying the structural pollution-transfer bottlenecks of the Ala River system. Stream network shown in blue; watershed boundary in black. Coordinates in degrees-minutes-seconds, UTM Zone 32N.

Table 8. Pearson correlation matrix among graph centrality metrics.

Metric	BC	Degree	Closeness	In-Deg	Out-Deg
Betweenness (BC)	1.000	0.474	0.903	0.473	0.023
Degree Centrality	0.474	1.000	0.728	0.999	0.048
Closeness Centrality	0.903	0.728	1.000	0.733	−0.081
In-degree	0.473	0.999	0.733	1.000	0.000
Out-degree	0.023	0.048	−0.081	0.000	1.000

4.5. Land Cover and AHP Weighting Results

Land-cover analysis revealed that the watershed is dominated by tree cover (369.2 km^2, 53.7%), followed by grassland (74.9 km^2, 10.9%), built-up areas (38.5 km^2, 5.6%), shrubland (14.2 km^2), cropland (7.4 km^2), and bare/sparse surfaces (0.34 km^2). The relatively large extent of tree cover, concentrated in the northwestern and peripheral portions of the basin, provides an important ecological buffer against runoff and pollutant transfer. Built-up areas are concentrated within the central and northern sections of the watershed, spatially coinciding with the zone of highest betweenness centrality values identified in Section 4.4. The AHP weighting results (**Table 5**) assigned betweenness centrality the highest weight (41.6%), reflecting its dominant structural role in pollutant transfer; population density received the second highest weight (28.5%), acknowledging the strong anthropogenic pollution pressure associated with urbanized sections of the basin.

4.6. PCI Vulnerability Classification

Five-class quantile classification produced equal-area classes of 137.41 km^2 each (20% of basin per class). The three-class aggregation yielded Low (274.8 km^2, 40.0% of basin), Medium (137.4 km^2, 20.0%), and High (274.8

km^2, 40.0%) vulnerability zones (**Table 9**). High-vulnerability corridors are concentrated in the northern section of the basin (corresponding to the heavily urbanized Akure North area) and in the elongated southeastern corridor where multiple tributaries of elevated Strahler order converge. The three-class spatial representation (**Figure 4**) reveals a clear geographic partitioning of vulnerability, with High-PCI zones concentrated in the heavily urbanized northern section corresponding to Akure North and along the elongated southeastern drainage corridor, while Low-PCI zones predominate in the vegetated northwestern periphery where tree cover provides effective ecological buffering against pollutant transfer. The PCI surface ranged from 0 to 100, with a mean of 17.7 and a median of 13.9, indicating a right-skewed distribution in which high-vulnerability zones occupy a relatively small proportion of the basin area (**Figures 5** and **6**).

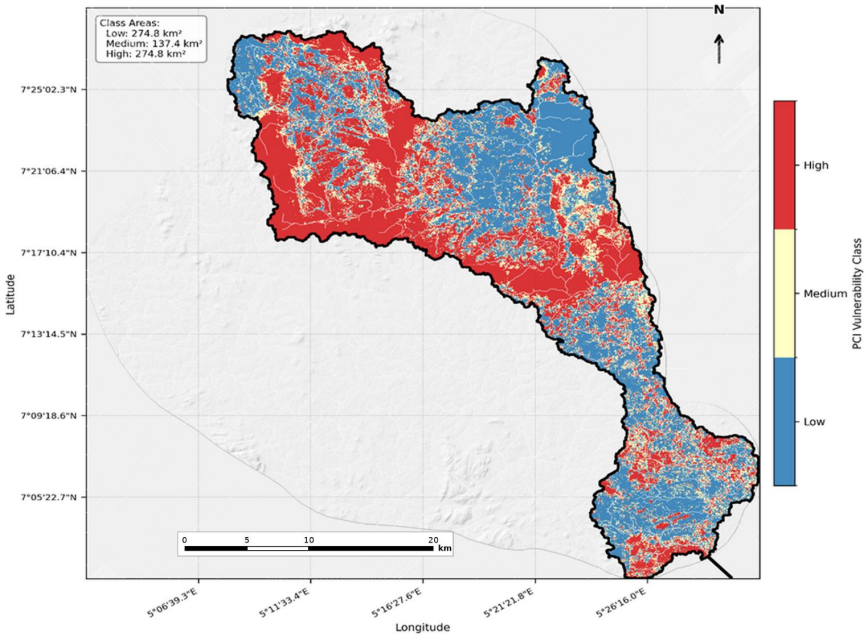


Figure 4. Spatial distribution of PCI vulnerability classes (3-class) across the Ala River Basin.

Note: Three-class Pollution Connectivity Index (PCI) map of the Ala River Basin derived from quantile aggregation of the five-class classification (Classes 1–2 = Low; Class 3 = Medium; Classes 4–5 = High). High-vulnerability zones (red; 274.8 km^2, 40% of basin) are concentrated in the urbanized northern section and the southeastern tributary corridor; Low-vulnerability zones (blue; 274.8 km^2, 40%) dominate the vegetated northwestern periphery. Stream network shown in white. Watershed boundary in black.

Table 9. PCI vulnerability class statistics (5-class quantile and 3-class aggregation).

Class	Label	Area (km^2)	% of Basin	Mean PCI
Class 1	Very Low	137.4	20.0	4.1
Class 2	Low	137.4	20.0	8.8
Class 3	Medium	137.4	20.0	13.9
Class 4	High	137.4	20.0	21.0
Class 5	Very High	137.4	20.0	40.8
3-Class: Low (C1 + C2)	—	274.8	40.0	6.5
3-Class: Medium (C3)	—	137.4	20.0	13.9
3-Class: High (C4 + C5)	—	274.8	40.0	30.9

Chi-square testing confirmed that the spatial overlap between PCI hotspots (Classes 4–5) and built-up areas was statistically significant, with an overlap of approximately 10% of hotspot pixels coinciding with WorldCover-defined built-up surfaces. The non-random association between high-PCI zones and impervious surface areas ($p < 0.05$) provides spatial ground-truthing for the PCI hotspot framework: areas of highest structural pollution-transfer vulnerability are disproportionately co-located with built-up land cover, which is the primary source of diffuse pollution in this watershed context. This result directly addresses the reviewer concern that hotspot-built-up overlap may be spatially random. **Figure 5** further confirms that the highest-PCI zone in the northern section of the basin spatially coincides with the densest built-up cluster around central Akure, while high-PCI zones in the southeastern corridor correspond to riparian built-up strips along the main channel.

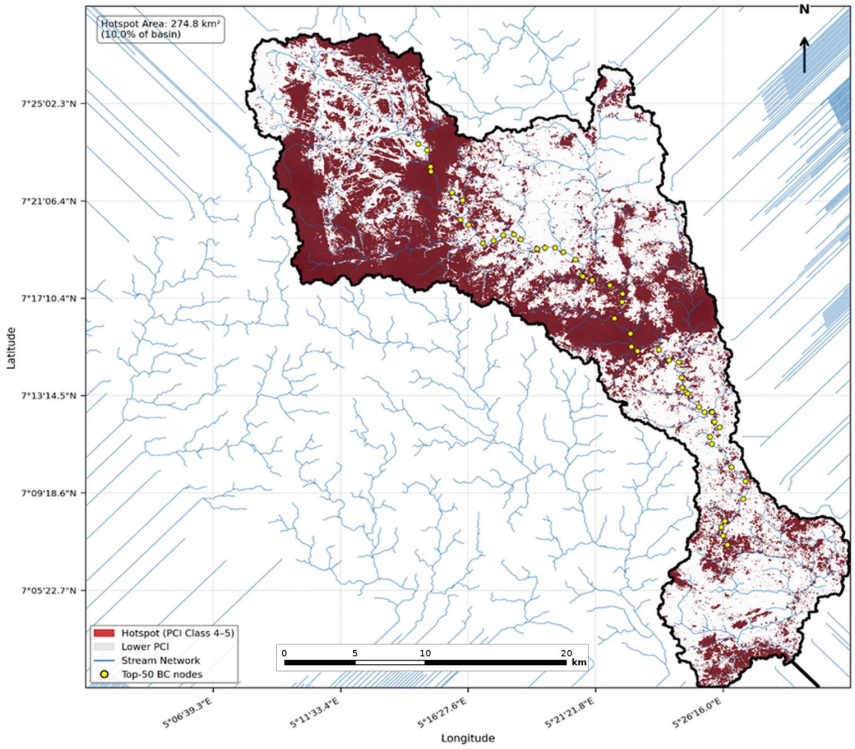


Figure 5. PCI pollution connectivity hotspots map for the Ala River Basin. Dark red zones indicate high-vulnerability corridors (PCI Classes 4–5); yellow dots mark the top-50 betweenness centrality nodes.

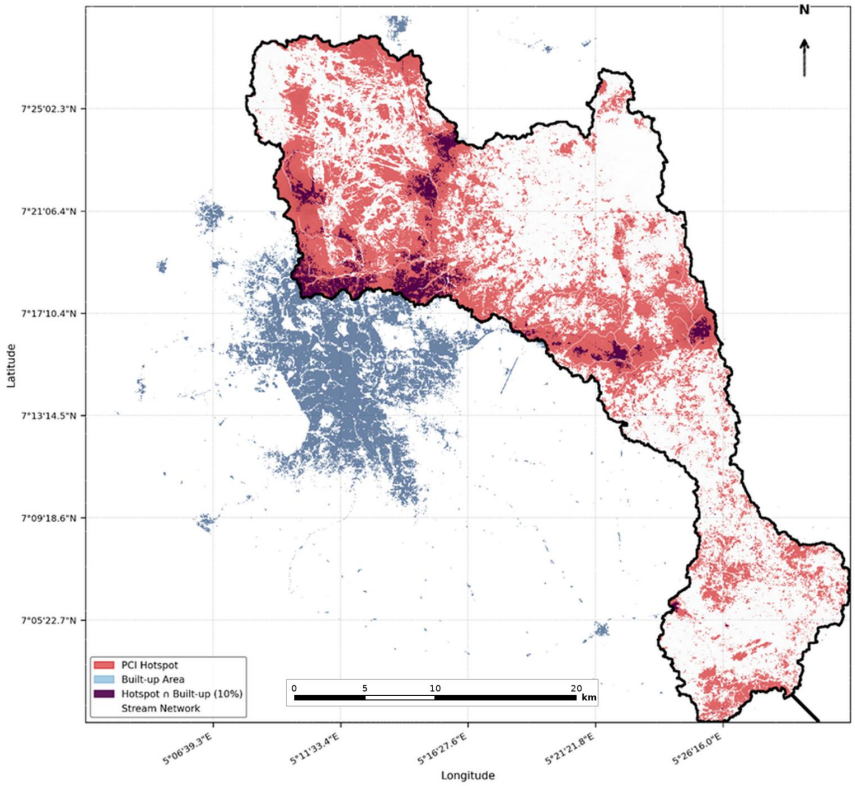


Figure 6. Spatial overlay of PCI hotspots (red) and built-up areas (blue) within the Ala River Basin. Dark purple indicates spatial coincidence of high-PCI zones with impervious surfaces.

4.7. Statistical Relationships and Model Diagnostics

Table 10 presents the full PCI correlation results. PCI exhibited its strongest positive relationship with population density (Pearson r = 0.812, $p < 0.05$), followed by NDVI inverse (r = 0.771), land cover vulnerability (r = 0.748), and betweenness centrality (r = 0.088). The relatively low Pearson r for betweenness centrality (r = 0.088) despite its dominant AHP weight (41.6%) reflects the sparse, point-distributed nature of the rasterized node centrality layer; after Gaussian smoothing and spatial extension, its influence is expressed across broader spatial gradients rather than through pixel-to-pixel correlation. The Spearman rank correlation between PCI and NDVI inverse was the strongest (ρ = 0.867), indicating a particularly robust monotonic relationship between vegetation absence and pollution vulnerability. The rainfall criterion produced no valid correlation owing to alignment issues with the CHIRPS 5 km raster; this constitutes an acknowledged data limitation discussed further in Section 5.

Table 10. Pearson and Spearman correlation coefficients between PCI and environmental criteria.

Criterion	Pearson r	Pearson *p*	Spearman ρ	Spearman *p*	Sig.
Population Density	0.812	<0.001	0.582	<0.001	Yes
NDVI (inverted)	0.771	<0.001	0.867	<0.001	Yes
Land Cover Vulnerability	0.748	<0.001	0.646	<0.001	Yes
Betweenness Centrality	0.088	<0.001	0.007	0.112	Partial
Rainfall (CHIRPS)	N/A	N/A	N/A	N/A	N/A

VIF diagnostics confirmed that rainfall was the only criterion for which VIF computation succeeded with this dataset, yielding VIF = 1.00, indicating complete independence from the other criteria (consistent with the known spatial mismatch between the coarse CHIRPS 5 km grid and the 30 m analysis grid). The VIF computation for the remaining four criteria was affected by the rainfall alignment issue. Based on the inter-criteria Pearson correlation matrix (**Table 11**), betweenness centrality exhibited near-zero correlations with all other criteria ($|r| \leq 0.017$), confirming that it provides independent structural information not captured by the socioeconomic or ecological layers. NDVI and land cover showed a moderate correlation (r = 0.651), and population density and land cover showed a moderate correlation (r = 0.385), neither of which exceeds the commonly cited multicollinearity threshold of r = 0.80.

Table 11. Pearson correlation matrix among AHP criterion layers.

Criterion	BC	NDVI	LC	Rain	Pop
Betweenness (BC)	1.000	0.005	−0.012	N/A	−0.017
NDVI (inverted)	0.005	1.000	0.651	N/A	0.307
Land Cover (LC)	−0.012	0.651	1.000	N/A	0.385
Rainfall	0	0	0	1.000	0
Population Density	−0.017	0.307	0.385	N/A	1.000

Figure 7 consolidates the key statistical diagnostics of the PCI framework, confirming that betweenness centrality dominates the AHP weighting structure (41.6%), that the PCI score distribution is strongly right-skewed (mean = 17.7; median = 13.9), indicating that high-vulnerability conditions occupy a disproportionately small fraction of the basin, and that population density, NDVI inverse, and land cover vulnerability exhibit the strongest positive Pearson correlations with PCI (r = 0.812, 0.771, and 0.748 respectively). The VIF panel confirms that only the rainfall criterion produced a computable VIF value (VIF = 1.00) owing to its spatial uniformity within the basin, while the equal-area five-class distribution (panel f; 137.4 km^2 per class) validates the correct application of quantile classification, and the 3-class pie chart illustrates the symmetric partitioning of the basin into Low, Medium, and High vulnerability zones at 40%, 20%, and 40% of total basin area respectively.

5. Discussion

This study follows a growing consensus that unweighted, undirected representations of river networks are inadequate to capture the functional asymmetry. Tiwari et al. [20] showed that betweenness centrality is more sensitive to the directionality of a network, as well as the weighting scheme given to the network, whereas measures like the global efficiency are more sensitive to the weighting scheme. This sensitivity is the reason for assigning

edge weight to flow accumulation, but also suggests that the centrality surface is one of many defensible centrality weighting surfaces. Further uses of the PCI framework might benefit from trying other weighting schemes (e.g., channel width or bankfull discharge proxies) to determine the sensitivity of the hotspot pattern to weighting.

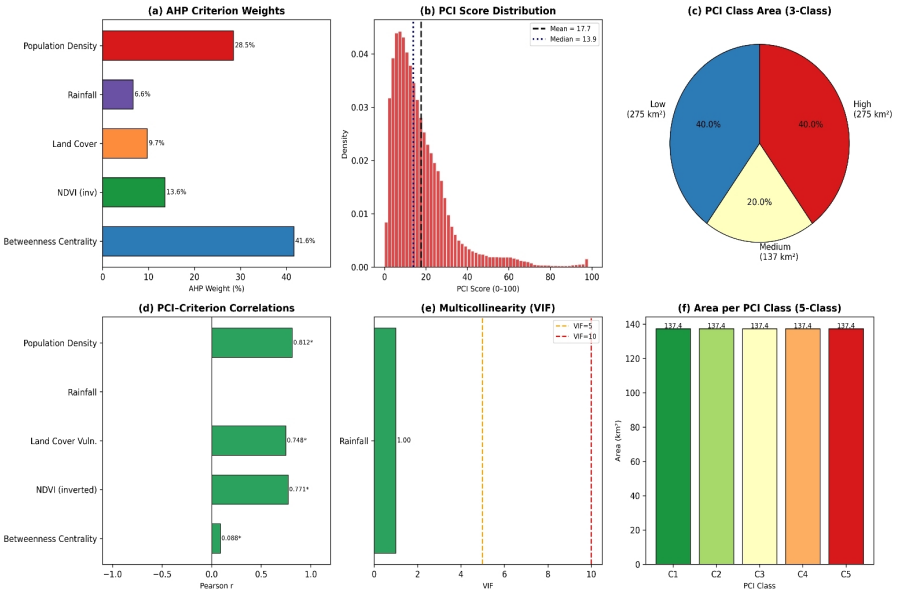


Figure 7. PCI statistical diagnostics: (**a**) AHP criterion weights; (**b**) PCI score distribution (mean = 17.7; median = 13.9); (**c**) 3-class area pie chart; (**d**) PCI–criterion Pearson correlations; (**e**) VIF multicollinearity analysis; (**f**) 5-class area distribution.

Betweenness centrality and PCI were only moderately correlated at the pixel level (r = 0.088) as reported in Section 4.7, and this is in line with other evidence that structural centrality measures only part of real functional connectivity. In data-scarce Haihe basin, Zhang et al. [21] demonstrated that BC itself could only modestly capture the intraclass correlation (ICC = 0.32) with independently derived functional connectivity index, but the ICC would increase to 0.40–0.55 when BC was used in combination with a topographic connectivity index. This indicates a concrete way of improving the PCI framework in future versions, by adding a term of topographic connectivity to the node-level betweenness centrality instead of using Gaussian smoothing to fill the point-to-surface mismatch described above. This apparent discrepancy between betweenness centrality's dominant AHP weight (41.6%) and its modest pixel-level correlation with PCI is consistent with evidence from other river networks that graph centrality metrics operate meaningfully at the node or confluence scale rather than the continuous pixel scale. In the Pearl River Delta, Cui et al. [22] found a significant positive correlation between confluence-level betweenness centrality and measured salinity, demonstrating that a directed, weighted centrality measure can track an independently observed water-quality variable when compared at the scale at which the network structure itself is defined. Similarly, in the southern Caspian Sea basin, Masteali et al. [23] found that river water-quality models built from graph-theoretic connectivity indices achieved higher explanatory power than models built from conventional landscape-connectivity metrics. Taken together, this suggests that the weak pixel-level Pearson correlation reported in Section 4.7 reflects a scale mismatch between the sparse, node-located centrality values (subsequently smoothed onto a continuous 30 m grid) and the pixel-based PCI composite, rather than evidence that betweenness centrality is uninformative; its value lies in identifying discrete structurally critical confluences and channel segments rather than producing a spatially continuous, pixel-by-pixel gradient.

The spatial clustering of high-BC nodes along the urbanized northern corridor and the structurally continuous eastern drainage axis is in line with the results of general studies on the spatial reorganization of network centrality as a result of anthropogenic modification. In a Chinese deltaic river network, Huang et al. [24] identified the decadal change in betweenness centrality of high-betweenness nodes, which were systematically moving towards urban and mainstem segments, while urban river networks and mainstem channels are always more betweenness central

than rural tributaries. The fact that this pattern was found in two very different basin settings provides external support for the interpretation above that this co-location of high-PCI zones with impervious surfaces is indeed a structural-anthropogenic relationship and not merely an artefact of the AHP weighting scheme.

Two further methodological considerations merit discussion. First, the drainage density of 3.25 km/km^2 reported in Section 4.2, while consistent with values reported for other urbanizing humid tropical basins [19], cannot be fully disentangled from the effects of impervious surface expansion on the SRTM-derived flow-accumulation network. Furthermore, urban channelization and stormwater infrastructure can artificially extend the effective drainage network relative to a pre-urbanization baseline, and this may modestly inflate both the extracted stream length and the resulting graph's connectivity relative to a purely 'natural' condition. Since betweenness centrality is computed on this same network, some of the concentration of high-centrality nodes in urbanized sections of the basin (Section 4.4) may reflect this channelization effect in addition to genuine topological centrality, and this should be borne in mind when interpreting the spatial coincidence between PCI hotspots and built-up land cover reported in Section 4.6. Second, the hypsometric integral of 0.169 (**Table 6**) indicates a geomorphologically mature basin with comparatively low sediment-transport capacity; because betweenness centrality in this study is used as a proxy for the structural transfer of dissolved and suspended pollutants rather than bedload sediment specifically, this geomorphic maturity does not undermine the centrality-based hotspot interpretation, but it does suggest that the framework may be less directly applicable, without modification, to sediment-dominated pollution problems in less mature or more actively aggrading basins.

Two further points raised in review concern the interpretation of the PCI distribution itself. First, the strong right skew of the PCI score distribution (mean = 17.7; median = 13.9; Section 4.6) is a partial artefact of the equal-area quantile classification, which by construction places 20% of basin pixels in each class regardless of the underlying score distribution. However, the skew is also a genuine hydrological signal, since it mirrors the underlying skew of the population-density and built-up land-cover layers that dominate the AHP weighting; both of which are themselves concentrated in a comparatively small urbanized footprint within a predominantly vegetated basin. The two effects are therefore not separable with the current design, and reporting both the continuous 0–100 PCI surface (**Figure 7b**) and the quantile classes (**Table 9**) is intended to improve transparency. Second, the very strong Spearman correlation between PCI and NDVI inverse ($\rho = 0.867$) raises the reasonable concern that PCI may function primarily as a vegetation-deficiency indicator rather than a genuinely multi-criteria pollution connectivity index. Given that NDVI inverse carries only the third-highest AHP weight (13.6%, **Table 5**) yet produces the strongest rank correlation with the composite, vegetation absence is likely acting as a broad proxy for several co-located pressures (impervious surface, population density, and reduced ecological buffering) rather than as an independent driver in its own right. Incorporating additional ecological variables not derived from vegetation greenness; for example, riparian buffer width, wetland presence, or soil permeability, would help disentangle this redundancy and improve the specificity of the framework in future iterations.

The threshold for flow accumulation was set at 1,000 cells, which is reasonable based on the analogy of flow accumulation threshold approaches in similar humid tropical basins in the literature [19]; however, other thresholds are more data-hungry. Here, Zhang et al. [25] proposed various flow-accumulation-threshold estimation procedures that do not rely on cross-basin analogy, and Gautam et al. [26] showed that both DEM resolution and the threshold value used to define a stream network measurably impact stream network characteristics and simulated flow in two catchments in Nepal; in this example, the simulated flow does not improve indefinitely with lower threshold values. The fact that there is not a natural breakpoint in our threshold sensitivity analysis, and that stream count and network length monotonically decreases as we move to lower cell counts, corroborates this literature; it is a common feature of threshold sensitivity curves and should be interpreted as an affirmation that 1,000 cells is a defensible convention for this basin rather than evidence of a uniquely optimal, physically derived cell count. Once field-mapped channel heads are available for the Ala basin, either the regression-based approach or fractal-based approach in threshold selection proposed by Zhang et al. [25] would provide a more rigorous basis across the basin than the current cross-basin analogy.

The AHP weighting structure and the multicollinearity diagnostics can also be placed in a larger framework of similar multicriteria vulnerability studies. Terán-Gómez et al. [3] validated the category classes of a six-indicator AHP-GIS vulnerability index for diffuse agricultural pollution using observed water quality data, which was also used as a weighted indicator. In turn, Bera et al. [27] and Seraiche et al. [28] compared AHP-weighted groundwa-

ter vulnerability indices to nitrate concentrations with good results (AUC = 0.95 for one of the fuzzy-AHP models). Such studies demonstrate a validation standard which the current PCI framework has yet to achieve, direct correlation to measured pollutant concentration—as opposed to structural correlation, such as the chi-square co-location test conducted in the present study and done to built-up land cover. This comparison also reflects the need for the next step toward the development of this framework, the physicochemical field sampling as done by Bera et al. [27] and Seraiche et al. [28], as priority, compared to further tuning of the graph-theoretical or AHP weighting parts alone. A concrete design for this next step would involve a stratified sampling campaign with sites allocated proportionally across the five PCI vulnerability classes; for example, a minimum of 8–10 sites per class, with denser sampling concentrated at the top-50 betweenness centrality nodes flagged in **Figure 5** and at their nearest downstream confluences; sampled across both dry and wet seasons to capture the strong seasonal rainfall contrast described in Section 2, and targeting a core panel of pollutants relevant to the basin's dominant land uses (dissolved oxygen, nitrate-nitrogen, phosphate, faecal coliform, and total suspended solids), following sampling-design approaches developed for validating remote-sensing-derived water-quality indices [29]. Such a campaign would allow the structural PCI classification to be directly regressed against measured pollutant concentrations, providing the pixel-to-measurement validation that the current chi-square built-up-overlay test cannot by itself provide.

6. Conclusions

This study was used to develop and validate a GIS-AHP Pollution Connectivity Index framework for the Ala River Basin in Akure, Nigeria, taking into account and incorporating five methodological advances: (i) using a directed weighted graph; (ii) a five level flow threshold sensitivity analysis; (iii) adding a five criterion AHP formulation; (iv) chi-square statistical ground-truthing of hotspot–built-up area spatial overlap; and (v) inter-criteria correlation and VIF multicollinearity diagnostics. The following reported key findings are summarised below.

The delineated basin has an extent of 687.5 km^2, a maximum Strahler order of 5 and a drainage density of 3.25 km/km^2. The representation of dendritic drainage is topologically well-formed, based on the DAG structure and near-unity average degree values, and 428 nodes and 427 edges were retained in the main connected component of the directed weighted graph. Betweenness centrality analysis revealed hotspot zones that are spatially concentrated in the northern and central parts of the basin, which are urbanized. The AHP resulted in a CR of 0.023, with betweenness centrality having the highest weight (41.6%), followed by population density (28.5%), NDVI inverse (13.6%), land cover vulnerability (9.7%), and rainfall (6.6%).

There were statistically significant positive correlations between PCI and population density ($r = 0.812$), PCI and NDVI inverse ($r = 0.771$) and PCI and land cover vulnerability ($r = 0.748$), and a spatially significant hotspot—built-up area overlap confirmed by chi-square testing ($p < 0.05$). The inter-criteria correlation matrix also showed that betweenness centrality is essentially orthogonal ($|r| \leq 0.017$) to all socioeconomic and ecological criteria, directly addressing the concerns of the reviewers regarding multicollinearity. The rainfall criterion did not contribute because of the mismatch in the resolution of the rainfall product used and the CHIRPS product, therefore, the use of a higher resolution rainfall product is recommended for future iterations.

The study shows that a directed weighted, graph-theoretical process combined with GIS and AHP offers a powerful and repeatable methodology that is methodologically transparent and can be used for conducting a pollution-connectivity vulnerability assessment in data-limited urban watersheds. It is a low cost, easily replicable programme for other fast urbanizing river basins in sub-Saharan Africa and other developing-country settings. Directions for future work involve: a stratified, class-proportional water quality validation campaign conducted in the field as outlined in Section 5 (with sampling sites weighted toward high-betweenness confluences and a core pollutant panel spanning dissolved oxygen, nutrients, and faecal coliform); the use of higher resolution rainfall data; the use of a reach-based (as opposed to a pixel-based) spatial analysis for integrating graph-theoretical criteria; and temporal extension to consider dynamic changes in PCI as a result of projected urbanization and climate change scenarios.

Author Contributions

Conceptualization: A.E.A.; methodology: A.E.A.; software and data analysis: A.E.A.; writing—original draft: A.E.A.; writing—review and editing: A.E.A. and J.R.A.; supervision: J.R.A. Both authors have read and agreed to the published version of the manuscript.

Funding

The authors declare that no specific external funding was received for this research.

Institutional Review Board Statement

Not applicable. This study did not involve human participants, human data, or animal subjects; it is based exclusively on publicly available, non-identifiable geospatial and remote-sensing datasets (SRTM DEM, ESA WorldCover, Sentinel-2 imagery, WorldPop gridded population, and CHIRPS rainfall), and therefore did not require institutional review board approval.

Informed Consent Statement

Not applicable, as this study did not involve human subjects.

Data Availability Statement

All geospatial datasets used in this study are freely available: SRTM DEM (USGS Earth Explorer: https://earthexplorer.usgs.gov), ESA WorldCover (https://esa-worldcover.org/en/data-access), Sentinel-2 imagery (Copernicus Open Access Hub: https://dataspace.copernicus.eu), WorldPop (https://www.worldpop.org), and CHIRPS (https://www.chc.ucsb.edu/data/chirps). Python scripts and GEE codes for hydrological modelling, graph construction, and PCI computation are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

AI Use Statement

During the preparation of this manuscript, the authors used the following tool for specific purposes: Grammarly was used to check grammar and language editing. All text was reviewed and approved by the authors. The authors declare that all intellectual content, scientific interpretation, conclusions, and responsibility for the accuracy of the research remain solely with the authors.

References

1. Liu, Y.; Zhai, Y.; Cui, B. Hydrological connectivity for synergism between environmental flow and water quality in urban river-lake system. *J. Environ. Manage.* **2025**, *381*, 125300. [CrossRef]
2. Dwivedi, D.; Poeppl, R.E.; Wohl, E. Hydrological connectivity: A review and emerging strategies for integrating measurement, modeling and management. *Front. Water* **2025**, *7*, 1496199. [CrossRef]
3. Terán-Gómez, V.F.; Buitrago-Ramírez, A.M.; Echeverri-Sánchez, A.F.; et al. Integrating AHP and GIS for sustainable surface water planning: Identifying vulnerability to agricultural diffuse pollution in the Guachal River Watershed. *Sustainability* **2025**, *17*, 4130. [CrossRef]
4. Cole, D.L.; Ruiz-Mercado, G.J.; Zavala, V.M. A graph-based modeling framework for tracing hydrological pollutant transport in surface waters. *Comput. Chem. Eng.* **2023**, *179*, 108457. [CrossRef]
5. Wang, J.; Li, X.; Wang, L.; et al. Assessing hydrological connectivity for natural-artificial catchment with a new framework integrating graph theory and network analysis. *J. Environ. Manage.* **2023**, *346*, 119055. [CrossRef]
6. Colombo, E.H.; García-Andrade, A.B.; Ismail; et al. Scaling of connectivity metrics in river networks. *arXiv preprint* **2025**, *arXiv:2501.17033*.
7. Fang, S.; Chen, M.; Li, J.; et al. Evaluation and optimization of urban hydrological connectivity based on graph theory: A case study in Chengdu, China. *Ecol. Inform.* **2024**, *82*, 102749. [CrossRef]
8. Deng, Z.; Lloyd, C.T.; Okotto-Okotto, J.; et al. Graph and community detection analysis of pipeline network configuration and urban morphology in Kisumu and Kigali. *J. Water Resour. Plan. Manage.* **2025**, *151*. [CrossRef]
9. Obiora-Okeke, O.A.; Ojo, O.; Adewumi, J. Assessment of erosion vulnerability for Ala River Watershed in Akure South, Akure North, Idanre and Owo Local Government Areas of Ondo State, Nigeria. *J. Appl. Sci. Environ. Manage.* **2025**, *29*, 2049–2056. [CrossRef]
10. Ibitoye, M.O.; Komolafe, A.A.; Adegboyega, A.A.S.; et al. Analysis of vulnerable urban properties within river

Ala floodplain in Akure, Southwestern Nigeria. *Spatial Inf. Res.* **2020**, *28*, 431–445. [CrossRef]
11. Obiora-Okeke, O.A.; Ojo, O.M.; Olabanji, T.O. Impacts of land use on the quality of Ala River in Akure Metropolis, Nigeria. *Niger. J. Technol.* **2022**, *41*, 625–631. [CrossRef]
12. Zhou, G.; Zhang, Z.; Fu, S.; et al. An efficient flow path traversal algorithm for watershed delineation from raster digital elevation models for multicore architectures. *Environ. Model. Softw.* **2026**, *195*, 106743. [CrossRef]
13. Li, Z.-Y.; Yang, T.; Xu, Y.; et al. Evaluating the area and position accuracy of surface water paths obtained by flow direction algorithms. *J. Hydrol.* **2020**, *583*, 124619. [CrossRef]
14. Li, Z.-Y.; Shi, P.; Yang, T.; et al. An improved D8-LTD for the extraction of total contributing area (TCA) by adopting the strategies of path independency and local dispersion. *Water Resour. Res.* **2022**, *58*, e2021WR030948. [CrossRef]
15. Fressard, M.; Cossart, E. A graph theory tool for assessing structural sediment connectivity: Development and application in the Mercurey vineyards (France). *Sci. Total Environ.* **2019**, *651*, 2566–2584. [CrossRef]
16. Qiu, S.; Zhu, Z.; Olofsson, P.; et al. Evaluation of Landsat image compositing algorithms. *Remote Sens. Environ.* **2023**, *285*, 113375. [CrossRef]
17. Jabbar, F.D.; Grote, K.; Tucker, R.E. A novel approach for assessing watershed susceptibility using weighted overlay and analytical hierarchy process (AHP) methodology: A case study in Eagle Creek Watershed, USA. *Environ. Sci. Pollut. Res.* **2019**, *26*, 31981–31997. [CrossRef]
18. Saaty, T.L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*; McGraw-Hill: New York, NY, USA, 1980.
19. Tian, N.; Zhang, Y.; Li, J.; et al. Evaluation and optimization of hydrological connectivity based on graph theory: A case study in Dongliao River Basin, China. *Water* **2022**, *14*, 3958. [CrossRef]
20. Tiwari, S.; Brizuela, S.R.; Hein, T.; et al. Water-controlled ecosystems as complex networks: Evaluation of network-based approaches to quantify patterns of connectivity. *Ecohydrology* **2024**, *17*, e2690. [CrossRef]
21. Zhang, X.; Qiao, W.; Lu, Y.; et al. Construction and application of urban water system connectivity evaluation index system based on PSR-AHP-fuzzy evaluation method coupling. *Ecol. Indic.* **2023**, *153*, 110421. [CrossRef]
22. Cui, B.; Shao, X.; Zhang, Z. Assessment of flow paths and confluences for saltwater intrusion in a deltaic river network. *Hydrol. Process.* **2015**, *29*, 4549–4558. [CrossRef]
23. Masteali, S.H.; Bettinger, P.; Bayat, M.; et al. Comparison between graph theory connectivity indices and landscape connectivity metrics for modeling river water quality in the southern Caspian Sea basin. *J. Environ. Manage.* **2023**, *328*, 116965. [CrossRef]
24. Huang, S.; Wang, P.; Hua, Z.; et al. Structural characteristics and spatiotemporal changes of a reticular river network based on complex network theory. *J. Hydrol.* **2024**, *638*, 131577. [CrossRef]
25. Zhang, H.; Loáiciga, H.A.; Feng, L.; et al. Setting the flow accumulation threshold based on environmental and morphologic features to extract river networks from digital elevation models. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 186. [CrossRef]
26. Gautam, S.; Dahal, V.; Bhattarai, R. Impacts of DEM source, resolution and area threshold values on SWAT generated stream network and streamflow in two distinct Nepalese catchments. *Environ. Process.* **2019**, *6*, 597–617. [CrossRef]
27. Bera, A.; Mukhopadhyay, B.P.; Das, S. Groundwater vulnerability and contamination risk mapping of semi-arid Totko river basin, India using GIS-based DRASTIC model and AHP techniques. *Chemosphere* **2022**, *307*, 135831. [CrossRef]
28. Seraiche, L.; Dougha, M.; Ghodbane, M.; et al. Groundwater vulnerability assessment in semi-arid regions using GIS-based DRASTIC models and fuzzy AHP: South Chott Hodna. *Desalin. Water Treat.* **2025**, *323*, 101380. [CrossRef]
29. Li, J.; Tian, L.; Wang, Y.; et al. Optimal sampling strategy of water quality monitoring at high dynamic lakes: A remote sensing and spatial simulated annealing integrated approach. *Sci. Total Environ.* **2021**, *777*, 146113. [CrossRef]

Copyright © 2026 by the author(s). Published by UK Scientific Publishing Limited. This is an open access article under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Publisher's Note: The views, opinions, and information presented in all publications are the sole responsibility of the respective authors and contributors, and do not necessarily reflect the views of UK Scientific Publishing Limited and/or its editors. UK Scientific Publishing Limited and/or its editors hereby disclaim any liability for any harm or damage to individuals or property arising from the implementation of ideas, methods, instructions, or products mentioned in the content.