RESEARCH ARTICLE OPEN ACCESS

Beyond Photo Counts: A Bias-Aware Multi-Indicator Model for Measuring Digital Tourism Visibility from Geo-Tagged Flickr Metadata in Data-Scarce Contexts—Evidence from Sudan

Mohamed Gafar Elhag Elamin[1]* and Dieter Fritsch[2]

[1] Sudan University of Science and Technology, Khartoum, Sudan; m.gaafer.sd@gmail.com

[2] Institute for Photogrammetry and Geoinformatics, University of Stuttgart, Stuttgart, Germany; dieter.fritsch@ifp.uni-stuttgart.de

* Correspondence: m.gaafer.sd@gmail.com

ABSTRACT

Geo-tagged social-media photographs offer fine-grained spatial evidence in destinations where official tourism data are incomplete, but raw photo counts can be strongly distorted by intensive contributors. This study develops a Bias-Aware Digital Tourism Visibility Index (BA-DTVI) and evaluates how conclusions change when photo-level evidence is separated from contributor-normalised spatial evidence. Flickr metadata for Sudan from 23 November 2002 to 17 February 2019 are analysed as a pre-war historical baseline. The photo-level branch contains 8,162 in-boundary records from 298 contributors and 1,291 contributor-days. DBSCAN (ε = 2 km; MinPts = 15; Haversine distance) identifies 46 candidate hotspots. A complementary contributor-cell branch retains one record per contributor and approximately 110 m coordinate cell, producing 2,686 records and 20 conservative core hotspots. BA-DTVI combines photographic volume, contributor breadth, and contributor-day participation through a geometric mean and adjusts the result for contributor concentration using the Herfindahl-Hirschman Index. The reproduced results show that 2018 leads photo volume, whereas 2013 ranks first under DTVI and BA-DTVI. March has the largest pooled monthly photo count, January the highest pooled BA-DTVI, and November leads after within-year normalisation. Spatially, Khartoum ranks first, followed by the Nile Valley heritage corridor and Red Sea nodes. All 20 conservative cores correspond to photo-level candidates, while 25 candidates disappear and one is almost entirely removed after normalisation. Contributor-removal, alternative-weight, concentration-adjustment, and duplicate-sensitivity tests confirm the robustness of the principal rankings. Fifteen of the 20 cores lie within 5 km of a curated tourism-relevant site. Confidence-aware cross-platform validation shows significant convergence with Wikimedia Commons image availability in the primary high- and medium-confidence matches (Spearman $\rho = 0.665$, $p = 0.0095$, $n = 14$), confirmed in the all-match sensitivity analysis ($\rho = 0.711$, $p = 0.0004$, $n = 20$), while the other Wikipedia/Wikidata indicators are not statistically significant. The model therefore provides a transparent and reproducible approach for distinguishing content volume from distributed digital attention without interpreting Flickr traces as tourist arrivals.

Keywords: Digital tourism visibility; geo-tagged Flickr metadata; contributor bias; composite indicator; DBSCAN; data-scarce contexts

1. INTRODUCTION

Tourism is spatially and temporally structured: destinations attract attention at different times, tourism-related activity concentrates around particular places, and these interactions may leave geo-referenced digital traces. Conventional monitoring systems use border records, accommodation statistics, ticket sales, transport data, surveys, and field counts. These sources remain indispensable, but they are often delayed, spatially coarse, expensive, or unavailable at the local resolution required for destination analysis.

Geo-tagged photographs provide a complementary source because they connect coordinates, timestamps, contributor identities, and often descriptive metadata. Previous studies have used such traces to map urban tourism spaces, reconstruct spatial patterns and movement, monitor protected areas, and compare destination prominence (Kádár & Gede, 2013; Kádár, 2014; Önder et al., 2016; Heikinheimo et al., 2017; Domènech et al., 2020). Their value is particularly evident in data-scarce contexts, although platform adoption, demographic selectivity, uneven connectivity, and changing posting practices prevent them from functioning as neutral censuses (Tenkanen et al., 2017; Wilkins et al., 2021; Ghermandi, 2022).

Photo count is intuitive and computationally inexpensive, but it assumes that every photograph contributes independent evidence. In practice, a single account may upload hundreds of images from one expedition or destination, while many other contributors provide only one or two records. A high-volume site may therefore reflect intensive documentation by a small number of contributors rather than broadly distributed digital attention. Contributor-day aggregation reduces same-day bursts, while contributor-cell normalisation constrains repeated photography by the same contributor within a small spatial cell. Neither operation, however, automatically removes inequality among highly active or spatially mobile contributors.

This study uses the term digital tourism visibility to describe the observable intensity, contributor breadth, temporal participation, and distribution of tourism-related digital traces within the analysed platform. It is not equivalent to visitor volume, destination demand, tourist arrivals, or official tourism statistics. Urban hotspots such as Khartoum may additionally reflect broader digital place visibility, including resident, professional, transit, and general urban photography. Tourism-related interpretations therefore require both geographic plausibility checks and appropriately qualified cross-platform validation.

Sudan provides a demanding testbed because major urban, archaeological, riverine, desert, coastal, and cultural resources coexist with limited fine-scale tourism-monitoring data. The study period, from 23 November 2002 to 17

February 2019, is treated as a historical pre-war baseline and not as a representation of current tourism conditions.

The present study reuses the cleaned geospatial corpus and the complementary photo-level and contributor-cell branches established within the broader research programme, but addresses a distinct measurement problem: how to combine content volume, contributor breadth, temporal participation, and contributor concentration into an interpretable bias-aware measure of digital tourism visibility. The photo-level branch provides the complete temporal baseline and candidate-hotspot structure, whereas the contributor-cell branch tests which spatial structures remain when repeated local photography is constrained. The resulting design therefore separates evidence volume from the robustness of spatial visibility under contributor normalisation.

1.1 Research aim and questions

The aim of this study is to develop and evaluate a bias-aware multi-indicator model for measuring digital tourism visibility and to determine how contributor normalisation, contributor concentration, and alternative analytical assumptions affect the temporal and spatial patterns obtained from geo-tagged Flickr metadata.

- RQ1: What annual, monthly, seasonal, state-level, and hotspot patterns emerge from the 8,162-record photo-level baseline?
- RQ2: How does contributor-cell normalisation alter the number, structure, and correspondence of the 46 photo-level candidate hotspots?
- RQ3: To what extent do photo volume, unique contributors, contributor-days, DTVI, and BA-DTVI produce different rankings of digital tourism visibility?
- RQ4: How robust are BA-DTVI rankings relative to photo-volume rankings under contributor removal, alternative weighting, concentration adjustment, and duplicate-removal sensitivity tests?
- RQ5: To what extent do the conservative core hotspots correspond with tourism-relevant geographic features and show convergence with independent digital-visibility indicators?

1.2 Scientific and computational contributions

1. A dual-unit analytical architecture that preserves the 8,162-record photo-level branch for temporal analysis and candidate-hotspot discovery while using a 2,686-record contributor-cell branch for conservative spatial analysis and bias control.
2. An interpretable bias-aware composite indicator, BA-DTVI, which integrates photographic volume, contributor breadth, and contributor-day participation through a geometric mean and explicitly penalises within-unit contributor concentration using the Herfindahl-Hirschman Index.
3. An auditable and reproducible data-processing procedure that resolves legacy contributor-identification errors, distinguishes 8,162 analytical rows from 8,132 distinct photo IDs, applies spatial administrative assignment through ADM1 point-in-polygon operations, and defines contributor-cell normalisation using the first retained occurrence in authoritative source order without temporal reordering.
4. A direct candidate-to-core correspondence procedure that traces retained contributor-cell observations back to the 46 photo-level candidate hotspots and quantifies shared records, candidate coverage, core purity, and centroid displacement, thereby distinguishing persistent cores, removed candidates, and border-point absorption.
5. A structured computational robustness protocol combining contributor-removal ablation, alternative weighting schemes, finite-sample HHI adjustment, duplicate sensitivity, Spearman and Kendall rank agreement, maximum rank shifts, top-five overlap, and leader persistence.
6. A confidence-aware validation strategy combining geographic correspondence with curated tourism-relevant sites and cross-platform comparison with Wikipedia, Wikidata, and Wikimedia indicators, while distinguishing primary higher-confidence matches from broader sensitivity analysis and avoiding interpretation of digital-platform measures as official visitor statistics.

2. RELATED WORK AND RESEARCH GAP

2.1 User-generated geographic information and tourism analytics

Tourism research increasingly combines surveys and administrative data with passively generated digital records, reflecting the broader transition toward data-driven geographic analysis (Miller & Goodchild, 2015). Li et al. (2018) distinguish user-generated content, device-generated traces, and transaction data, each of which requires different assumptions about sampling, behaviour, and representativeness. Geo-tagged photographs are particularly useful for spatial analysis because geographic position, time, and contributor identity can be represented within a single observational record.

Digital footprints have been used to identify tourist spaces and compare the spatial patterns captured by different platforms. Salas-Olmedo et al. (2018) found that digital platforms represent overlapping but non-identical dimensions of urban tourism, while Önder et al. (2016) demonstrated that geo-tagged photographs can reveal spatial and temporal structures associated with tourism activity. These studies establish the analytical value of digital traces but also demonstrate that the platform itself forms part of the measurement process rather than functioning as a neutral census.

This limitation is consistent with volunteered geographic information research. Senaratne et al. (2017) distinguish completeness, positional accuracy, contributor heterogeneity, provenance, and fitness for use as separate dimensions of data quality. For geo-tagged tourism photography, this distinction is critical: a geographically valid photograph may still provide disproportionate evidence if it originates from an unusually intensive contributor. Consequently, spatial validity alone does not guarantee behavioural representativeness.

2.2 Flickr, contributor-days, and contributor bias

Flickr has been widely used for identifying urban areas of interest, comparing spatial clustering approaches, and reconstructing spatial patterns and trajectories (Hu et al., 2015; Höpken et al., 2020; Domènech et al., 2020). Country-scale research also shows that preprocessing, contributor identification, and visitor-local classification decisions can materially affect both the size of the usable dataset and its interpretation (Derdouri & Osaragi, 2021; Martins & Santos, 2024).

User-day or photo-user-day measures provide one means of reducing the influence of intensive same-day photography. Wood et al. (2013) demonstrated that Flickr photo-user-days can explain variation in observed recreational activity, while later studies compared social-media indicators with visitor surveys and other platform-derived measures in protected areas and public lands (Heikinheimo et al., 2017; Tenkanen et al., 2017; Hausmann et al., 2018; Barros et al., 2020; Wood et al., 2020).

However, contributor-day aggregation addresses only one form of repetition. A contributor who documents many places across multiple dates can still exert substantial influence on a spatial or temporal ranking. Similarly, retaining only one observation for a contributor within a small spatial cell can reduce repeated local photography without eliminating inequality among contributors at broader spatial scales. This distinction motivates separating local repetition control from within-unit contributor-concentration adjustment.

Geolocated social-media counts are often positively associated with visitation-related measures, but their predictive value varies with platform penetration, site type, spatial resolution, and analytical specification (Song et al., 2020; Ghermandi, 2022). Such indicators should therefore be treated as complementary observational evidence rather than replacements for representative visitor-monitoring systems (Wilkins et al., 2021).

2.3 Spatial clustering and model-dependent hotspots

DBSCAN is well suited to geo-tagged observations because it detects irregularly shaped clusters, explicitly represents noise, and does not require the number of clusters to be specified in advance (Ester et al., 1996; Schubert et al., 2017). These properties make it useful for identifying geographically concentrated structures in heterogeneous social-media datasets.

Nevertheless, DBSCAN results depend on the neighbourhood radius, minimum-sample threshold, sampling density, preprocessing decisions, and analytical unit. A fixed parameter configuration may fragment a large urban structure, merge neighbouring attractions, or fail to identify sparsely represented heritage locations. Hotspots derived from density-based clustering should therefore be interpreted as model-dependent analytical units, not as directly observed tourism destinations.

The present study consequently does not claim a new clustering algorithm. DBSCAN provides spatial comparison units for the subsequent measurement framework. The methodological contribution lies instead in testing how those spatial structures change when intensive local photography is constrained through contributor-cell normalisation and when contributor concentration is incorporated into hotspot ranking.

2.4 Composite indicators, contributor concentration, and robustness

Composite indicators provide a means of summarising multidimensional phenomena, but they introduce methodological choices concerning feature selection, normalisation, weighting, aggregation, and uncertainty. Saisana et al. (2005) and Greco et al. (2019) emphasise that composite-index design should therefore be transparent and accompanied by sensitivity analysis.

For digital tourism visibility, photographic volume represents only one component of the available evidence. Contributor breadth indicates how widely that evidence is distributed across accounts, while contributor-days capture repeated temporal participation. A geometric mean provides a transparent aggregation mechanism in which a weak component constrains the resulting score rather than being completely compensated by a very high value in another dimension.

However, even a multidimensional index may remain biased if observations are highly concentrated among a small number of contributors. Contributor concentration therefore represents a distinct property from volume, contributor breadth, or contributor-day participation. The Herfindahl-Hirschman Index provides an interpretable means of quantifying this concentration and can be incorporated as an explicit reliability adjustment rather than assuming that contributor-cell normalisation alone eliminates user inequality.

Robustness is equally important because a composite ranking can appear plausible while remaining sensitive to individual contributors, weighting assumptions, finite-sample concentration effects, or duplicate observations. Rank correlations, contributor-removal experiments, alternative weighting schemes, concentration-adjustment sensitivity, duplicate-removal tests, and persistence of leading units therefore provide complementary evidence about whether the main conclusions depend on a narrow analytical configuration.

2.5 Research gap

The research gap addressed by this study is more specific than the general use of Flickr or geo-tagged social media in tourism analysis.

First, many studies rely primarily on photo counts or user-day measures, leaving the distinction between content volume, contributor breadth, temporal participation, and contributor concentration only partially addressed.

Second, controlling repeated photography at the record level does not necessarily control inequality at the analytical-unit level. Contributor-cell normalisation can limit repeated local observations, but substantial concentration may remain among contributors. Few approaches explicitly combine these two levels of bias control.

Third, composite digital-visibility measures require systematic robustness assessment. The effect of dominant-contributor removal, alternative component weights, concentration correction, duplicate observations, and changes in hotspot structure is not always reported together within a single reproducible analytical protocol.

Fourth, spatial hotspots identified from the full photo-level record set are commonly treated as final outputs. Less attention has been given to tracing how individual candidate hotspots persist, disappear, or are absorbed after contributor-aware normalisation.

Fifth, auxiliary digital platforms can support convergent validation, but different platforms measure different dimensions of digital visibility. Agreement should therefore be tested rather than assumed, and matching uncertainty should be explicitly incorporated into the validation design.

Finally, empirical evidence from African and other data-scarce contexts remains comparatively limited, even though these settings have the greatest need for complementary fine-scale sources of spatial evidence.

BA-DTVI addresses this intersection by combining dual analytical units, multidimensional visibility features, explicit contributor-concentration adjustment, candidate-to-core correspondence, systematic robustness testing, and uncertainty-aware geographic and cross-platform validation within a reproducible computational GIS workflow.

3. MATERIALS AND METHODS

3.1 Research design and dual analytical architecture

The study follows a dual analytical architecture designed to separate evidence volume from contributor-aware spatial robustness. The two analytical branches are complementary and are not treated as interchangeable datasets.

The photo-level branch retains all 8,162 records located inside the adopted post-2011 Sudan boundary. It provides the complete annual, monthly, seasonal, and year-normalised temporal baseline, the photo-level state baseline, and the input to candidate-hotspot detection.

The contributor-cell branch constrains repeated local photography by retaining one observation for each contributor and approximately 110 m rounded coordinate cell across the observation period. This produces 2,686 records and is used for conservative state-level comparisons, contributor-concentration analysis, core-hotspot detection, and spatial robustness assessment.

This dual structure allows the analysis to preserve the full temporal information of the Flickr archive while separately evaluating which spatial structures persist after intensive local posting is constrained

.

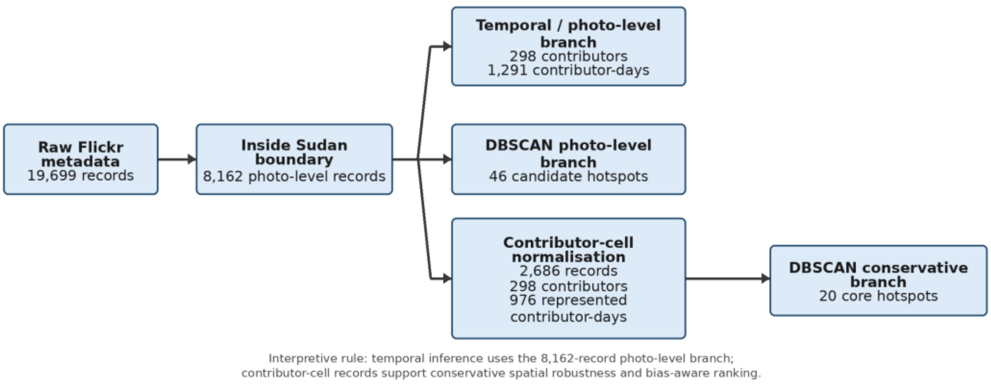

Figure 1. Data lineage and complementary analytical roles of the photo-level and contributor-cell branches.

Figure 1 clarifies the analytical separation: the 8,162-record branch preserves temporal and candidate-hotspot evidence, whereas the 2,686-record contributor-cell branch tests conservative spatial persistence, yielding 46 candidate hotspots and 20 conservative cores respectively.

3.2 Data source, temporal coverage, spatial filtering, and audit

The original relational database contained 19,699 Flickr metadata records collected for the Sudan study area. Spatial point-in-polygon filtering against the adopted post-2011 national boundary retained 8,162 records inside Sudan. Capture dates range from 23 November 2002 to 17 February 2019, and the resulting dataset is interpreted as a historical pre-war baseline rather than a description of current tourism conditions. The principal fields used in the analysis include photo identifier, latitude, longitude, capture time, URL, contextual place information, and administrative-location information.

A pre-analysis audit identified two important data-quality issues. First, the legacy owner field did not contain reliable Flickr account identifiers; in the analytical source file it contained place or city labels. Contributor identifiers were therefore reconstructed deterministically from the Flickr URL pattern /photos/<account>/<photo>/. This procedure identified 298 contributors and 1,291 distinct contributor-date pairs in the complete photo-level branch.

Second, the 8,162 analytical rows contained 8,132 distinct photo IDs, corresponding to 30 duplicate-excess rows. The 8,162-row branch was retained as the principal photo-level dataset for consistency with the preceding analyses, while a duplicate-free sensitivity branch retained the first occurrence of each photo ID in authoritative source order. The duplicate-sensitivity analysis was conducted separately to determine whether these records affected the temporal rankings or hotspot partition.

Administrative state assignment was not based on the raw Flickr region attribute because those metadata contained inconsistent and occasionally neighbouring-country region names. Instead, state units were assigned spatially by point-in-polygon intersection with the adopted Sudan ADM1 administrative boundary layer used in the reproducibility workflow. The final conservative branch contains 19 administrative units.

Contributor identifiers were used only for aggregation, concentration measurement, and robustness testing. Individual contributor identities, profiles, and trajectories are not reported.

3.3 Temporal photo-level branch

All archive-wide temporal analyses use the complete 8,162-record photo-level branch. For each temporal unit g, P_g is the number of photo-level records, U_g is the number of distinct contributors, and UD_g is the number of distinct contributor-date pairs.

Annual analysis retains all available years, although the partial years 2002 and 2019 are interpreted cautiously. Pooled monthly analysis aggregates all records belonging to the same calendar month across the observation period. Calendar seasons are reported as descriptive temporal groupings rather than causal climatic categories.

To reduce domination by unusually high-volume Flickr years, a complementary within-year normalisation was calculated for the full years 2003–2018. For each year, monthly values were expressed as shares of the corresponding annual total before the monthly shares were averaged across years. This representation measures recurrence within years rather than total pooled platform volume.

The contributor-cell branch is not used for general annual, monthly, or seasonal inference because its contributor-cell key does not preserve repeated appearances of the same contributor in the same spatial cell across different dates.

3.4 Contributor-cell conservative spatial branch

Contributor-cell normalisation was used to constrain repeated local photography while preserving contributor identity. Latitude and longitude were rounded to three decimal places, corresponding to an approximate spatial cell size of 110 m at the study latitudes. For each contributor u and rounded coordinate cell c, only one observation was retained across the complete observation period.

The representative record was selected as the first occurrence in authoritative source order. No temporal sorting was applied before contributor-cell normalisation.

This procedure reduced the 8,162 photo-level records to 2,686 contributor-cell records from the same 298 contributors. The retained contributor-cell sample contains 976 contributor-date pairs represented by the selected source records. This value is a property of the retained sample and is not interpreted as 976 visits, tourists, or independent tourism events.

Contributor-cell normalisation controls repeated local photography but does not eliminate inequality among contributors who photograph multiple distinct locations. Contributor concentration is therefore addressed separately in BA-DTVI.

$$A(u,c) = 1 \text{ if contributor } u \text{ has at least one record in cell } c; \text{ otherwise } A(u,c) = 0 \quad (1)$$

3.5 Digital visibility features and contributor concentration

For each analytical unit g, three primary visibility features were calculated. P_g represents the number of observations applicable to the analytical branch; U_g represents contributor breadth; and UD_g represents distinct contributor-date participation within that branch and unit.

$$U\_g = |\{u\_i : i \in g\}| \quad (2)$$

$$UD\_g = |\{(u\_i, d\_i) : i \in g\}| \quad (3)$$

For temporal and photo-level candidate-hotspot analyses, P_g is based on photo-level records. For conservative state and core-hotspot analyses, P_g is based on contributor-cell records.

Contributor inequality was evaluated using measures including dominant-contributor share, the Gini coefficient, normalised entropy, effective contributor number, and the Herfindahl-Hirschman Index (HHI). HHI provides the concentration term used directly in BA-DTVI. If s_ug is the share of records contributed by contributor u within analytical unit g:

$$HHI\_g = \Sigma\_u\, s\_ug^2 \quad (4)$$

$$N\_effective,g = 1 / HHI\_g \quad (5)$$

Higher HHI values indicate that the digital evidence within an analytical unit is concentrated among fewer contributors.

3.6 DTVI and BA-DTVI construction

Within each analytical scope, P, U, and UD were independently normalised by their maximum values:

$$P^*\_g = P\_g / \max(P); \quad U^*\_g = U\_g / \max(U); \quad UD^*\_g = UD\_g / \max(UD) \quad (6)$$

The base Digital Tourism Visibility Index (DTVI) is defined as the equal-weight geometric mean:

$$DTVI\_g = (P^*\_g \times U^*\_g \times UD^*\_g)^\wedge(1/3) \quad (7)$$

Equal weighting is used as a transparent baseline rather than as an assumption that the three dimensions necessarily have equal substantive importance in all applications. Contributor concentration is incorporated through R_g = 1 − HHI_g and:

$$BA\text{-}DTVI\_g = DTVI\_g \times (1 - HHI\_g) \quad (8)$$

The resulting BA-DTVI rewards units supported simultaneously by photographic volume, contributor breadth, and temporal participation while reducing the score when evidence is concentrated among a small number of contributors. Scores are normalised within each analytical scope and consequently should not be compared numerically across different scopes.

Because the theoretical minimum of HHI depends on contributor count, an adjusted concentration measure was also evaluated:

$$HHI\_adj,g = (HHI\_g - 1/U\_g) / (1 - 1/U\_g), \text{ for } U\_g > 1 \quad (9)$$

$$BA\text{-}DTVI\_adj,g = DTVI\_g \times (1 - HHI\_adj,g) \quad (10)$$

This sensitivity analysis tests whether units containing relatively few contributors are mechanically penalised through both U and unadjusted HHI.

3.7 DBSCAN hotspot detection and candidate-to-core correspondence

DBSCAN was applied separately to the photo-level and contributor-cell branches using the same reference parameters: Haversine distance, $\varepsilon = 2$ km, and MinPts = 15. Coordinates were transformed to radians before Haversine-distance clustering. The photo-level branch defines the candidate-hotspot configuration, while the contributor-cell branch provides the conservative core-hotspot configuration. DBSCAN is used here as a density-based spatial partitioning method rather than as a newly proposed clustering algorithm.

Candidate-to-core correspondence was established by tracing retained contributor-cell observations back to the candidate-cluster membership of their source photo-level records. For candidate i and core j, shared records S_ij count the retained observations assigned to both units. Candidate coverage is the proportion of retained contributor-cell records originating from candidate i that are shared with core j, and core purity is the proportion of core j supplied by candidate i. Centroid displacement was calculated as the Haversine distance between corresponding candidate and core centroids.

$$Coverage\_ij = S\_ij / retained\_i \times 100 \qquad (11)$$

$$Purity\_ij = S\_ij / core\_j \times 100 \qquad (12)$$

Candidates were subsequently classified as persistent one-to-one structures, removed after contributor-cell normalisation, or mostly removed with only a border observation absorbed into an adjacent core.

3.8 Indicator agreement, ablation, and robustness analysis

Robustness was evaluated at multiple levels rather than relying on a single sensitivity test. Spearman rank correlation and Kendall's tau were used to quantify agreement among P, U, UD, DTVI, BA-DTVI, and adjusted BA-DTVI.

Ablation was operationalised as a staged comparison from raw photographic volume (P), through contributor breadth (U) and contributor-day participation (UD), to the composite DTVI and finally BA-DTVI. This sequence was used to determine how multidimensional aggregation and the contributor-concentration penalty altered agreement with the original photo-volume ranking.

3.8.1 Contributor-removal experiments

Contributor dominance was evaluated by successively removing the single largest contributor and the top 1%, 5%, and 10% of contributors. Contributors were ranked according to their number of records within each analytical scope. For percentage-based scenarios, the number removed was rounded upward to the next integer. The analytical units themselves were held fixed during these tests; DBSCAN was not rerun after contributor removal. The purpose was to test ranking stability within previously established analytical units rather than to estimate a new hotspot configuration.

Stability was evaluated using Spearman rank correlation, Kendall's tau, maximum absolute rank shift, top-five overlap, and persistence of the highest-ranked unit. For rank-shift calculations, ties were assigned using dense ranking.

3.8.2 Alternative weighting

Three alternative weighting specifications were evaluated against the equal-weight model: user-priority (0.20, 0.40, 0.40), volume-priority (0.50, 0.25, 0.25), and persistence-priority (0.20, 0.30, 0.50) for P, U, and UD respectively. For a general weight combination:

$$DTVI\_w,g = (P^*\_g)\text{^}wP \times (U^*\_g)\text{^}wU \times (UD^*\_g)\text{^}wUD, \text{ with } wP + wU + wUD = 1 \qquad (13)$$

$$BA\text{-}DTVI\_w,g = DTVI\_w,g \times (1 - HHI\_g) \qquad (14)$$

3.8.3 Duplicate sensitivity

A duplicate-free sensitivity branch was constructed by retaining the first source-order occurrence of each photo ID, reducing the photo-level dataset from 8,162 rows to 8,132 distinct photo IDs. Annual BA-DTVI ordering and DBSCAN partition stability were then compared with the principal 8,162-row branch.

3.9 Geographic and cross-platform validation

Validation was conducted at two complementary levels. First, the centroids of the 20 conservative core hotspots were compared with a curated inventory of 69 tourism-relevant, archaeological, heritage, urban, and natural sites assembled from open geographic and knowledge sources. A centroid-to-site distance of 5 km or less was treated as close geographic correspondence. This test evaluates geographic plausibility rather than estimating visitor numbers.

Second, cross-platform convergent validation compared hotspot-level BA-DTVI with auxiliary Wikipedia, Wikidata, and Wikimedia indicators: Wikimedia Commons image availability, Wikipedia article-image counts, Wikipedia page views, article length, number of language editions, and Wikidata sitelinks. These variables are interpreted as independent dimensions of digital visibility, not as official tourism or visitation statistics.

3.9.1 Confidence-aware entity matching

To reduce ambiguity in linking core hotspots to external knowledge entities, the validation was recalculated using a confidence-aware matching procedure. Candidate entities were evaluated using a combination of place-name correspondence, geographic distance from the hotspot centroid, entity type, consistency with geographic context, and duplicate-aware allocation where competing hotspots could otherwise be assigned to the same external entity.

Name-based candidates located more than 25 km from the corresponding hotspot were rejected rather than forced into the analysis. No destination-specific hard-coded assignment was imposed. Each final match was assigned a confidence class of High, Medium, or Low.

The primary validation analysis uses High- and Medium-confidence matches only. A second sensitivity analysis uses all available matches, including Low-confidence cases, to determine whether the direction and strength of associations depend on matching uncertainty. Missing page-view observations, including unavailable or unresolved pages, remain missing rather than being converted to zero.

Spearman rank correlation was used to assess association between BA-DTVI and each external digital-visibility indicator. This validation design therefore tests convergent rather than interchangeable measurement: substantial correspondence with one external platform does not require all digital-platform indicators to produce identical rankings.

4. RESULTS

4.1 Data lineage and analytical coverage

Table 1 summarises the audited data lineage and the complementary roles of the two analytical branches. The original retrieval contained 19,699 Flickr metadata records, of which 8,162 were retained after spatial filtering to the adopted Sudan boundary. These 8,162 analytical rows contain 8,132 distinct photo IDs, corresponding to 30 duplicate-excess rows. Contributor identifiers reconstructed from Flickr URL paths yielded 298 verified contributors and 1,291 contributor-days in the complete photo-level branch.

Contributor-cell normalisation reduced the dataset to 2,686 records while retaining the same 298 contributors. The selected source records represent 976 contributor-date pairs within this retained sample. This value is specific to the contributor-cell branch and is not used for general temporal inference. DBSCAN applied to the photo-level branch produced 46 candidate hotspots, containing 6,907 clustered records and 1,255 noise records (15.38%). Applying the same reference parameters to the contributor-cell branch produced 20 conservative cores, containing 1,709 clustered records and 977 noise records (36.37%).

Table 1. Data lineage, audit findings, and analytical roles.

Stage or measure	Value	Analytical role
Raw retrieval	19,699	Starting metadata collection
Inside Sudan boundary	8,162 rows	Temporal baseline, state baseline, candidate DBSCAN
Distinct photo IDs	8,132	30 duplicate-excess rows audited through sensitivity analysis
Correct Flickr contributors	298	Parsed from Flickr URL paths
Photo-level contributor-days	1,291	Annual, monthly, and seasonal temporal participation
Contributor-cell sample	2,686	Conservative state and core-hotspot analysis
Contributor-days represented in retained contributor-cell sample	976	Retained-sample value only
Photo-level DBSCAN	46 hotspots	6,907 clustered; 1,255 noise (15.38%)
Contributor-cell DBSCAN	20 hotspots	1,709 clustered; 977 noise (36.37%)

4.2 Contributor inequality before and after normalisation

Contributor-cell normalisation reduced the contributor-activity Gini coefficient from 0.836 to 0.759, demonstrating that repeated local photography accounted for part, but not all, of the observed inequality. Even after normalisation, the top 10% of contributors generated 68.7% of the 2,686 retained records. Contributor-cell normalisation should therefore be interpreted as a control for repeated local representation rather than as a complete correction for contributor inequality. This remaining imbalance provides the empirical justification for the explicit HHI adjustment incorporated into BA-DTVI.

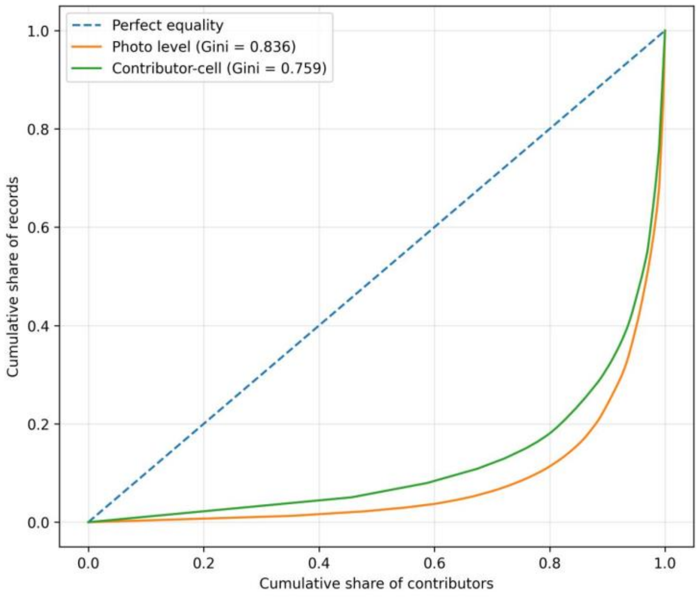

Figure 2. Contributor Lorenz curves before and after contributor-cell normalisation.

Figure 2 shows that contributor-cell normalisation moves the Lorenz curve closer to equality, but the remaining curvature confirms substantial contributor imbalance and supports the subsequent HHI-based concentration adjustment.

4.3 Corrected annual digital visibility

The annual results demonstrate that photographic volume and distributed digital visibility are related but not interchangeable. The largest photo-level volume occurs in 2018 (1,873 records), but this evidence is generated by only 24 contributors across 79 contributor-days and is highly concentrated (HHI = 0.482). Consequently, 2018 falls to eighth position under BA-DTVI. In contrast, 2013 combines 888 records, 50 contributors, 178 contributor-days, and a substantially lower concentration level (HHI = 0.225), producing the highest DTVI and BA-DTVI (0.593). Photo count and BA-DTVI remain strongly correlated (Spearman $\rho = 0.862$), but the rank changes confirm that photographic volume alone does not preserve the same interpretation.

Table 2. Leading years according to corrected photo-level BA-DTVI.

Year	Photos	Contributors	Contributor-days	HHI	BA-DTVI	Photo rank	BA rank
2013	888	50	178	0.225	0.593	3	1
2012	551	48	134	0.118	0.516	8	2
2011	701	52	128	0.200	0.514	6	3
2014	763	36	181	0.226	0.507	4	4
2016	1000	34	112	0.179	0.493	2	5
2015	660	36	128	0.194	0.449	7	6
2017	702	21	114	0.138	0.394	5	7
2018	1873	24	79	0.482	0.304	1	8

Table 2 highlights the central annual contrast: 2018 has the highest photo volume, whereas 2013 leads BA-DTVI because its evidence is broader across contributors and contributor-days and is less concentrated.

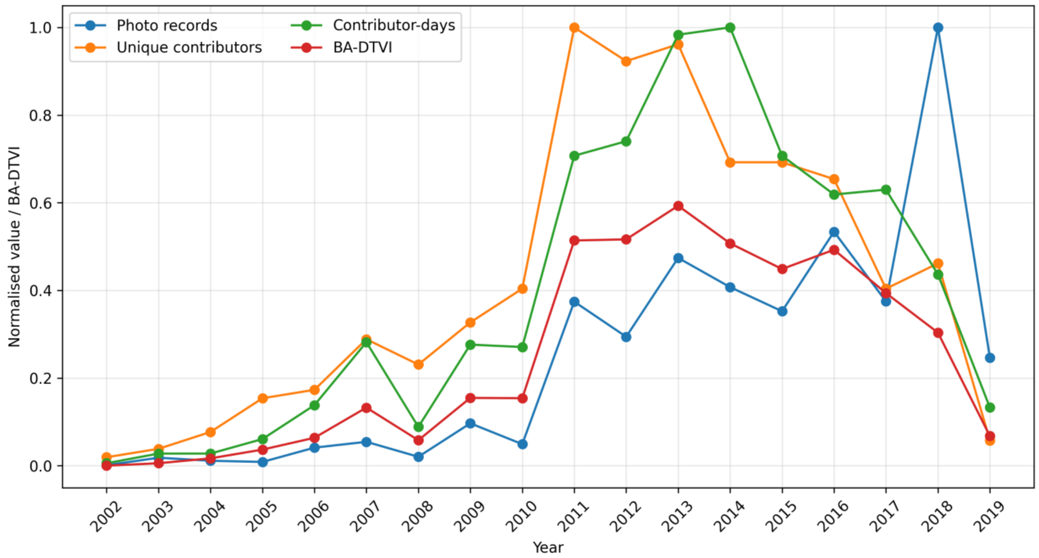

Figure 3. Corrected annual photo volume, unique contributors, contributor-days, and BA-DTVI using 8,162 photo-level records.

Figure 3 visualises the annual divergence between volume and distributed visibility: the 2018 photo-volume peak does not become a BA-DTVI peak, whereas 2013 combines broader contributor participation and contributor-day support with lower concentration.

4.4 Pooled monthly, year-normalised monthly, and seasonal patterns

March contains the largest pooled monthly photo volume (1,601 records), whereas January ranks first under pooled BA-DTVI (0.830). March falls to second (0.627), followed by December (0.607). These differences reflect the additional influence of contributor breadth, contributor-day participation, and concentration.

Table 3. Leading pooled calendar months according to BA-DTVI.

Month	Photos	Contributors	Contributor-days	HHI	BA-DTVI	Photo rank	BA rank
January	1448	59	175	0.114	0.830	2	1
March	1601	65	157	0.350	0.627	1	2
December	1296	52	131	0.228	0.607	3	3
November	588	46	121	0.097	0.509	6	4
February	437	51	135	0.076	0.507	7	5
October	599	47	102	0.196	0.435	5	6

Table 3 shows that the month with the most photographs is not the month with the highest bias-aware visibility: March leads photographic volume, whereas January leads BA-DTVI.

A different pattern emerges after within-year normalisation. November ranks first, followed by December, January, and March. This indicates that pooled monthly prominence is partly influenced by exceptionally high-volume years. The two representations answer different questions: pooled results describe the accumulated Flickr evidence across the archive, whereas year-normalised results describe the recurrence of monthly prominence across complete years.

Table 4. Calendar-season photo-level indicators.

Calendar season	Photos	Contributors	Contributor-days	HHI	BA-DTVI	Photo rank	BA rank
Winter	3181	118	441	0.110	0.890	1	1
Autumn	1544	100	306	0.072	0.611	3	2
Spring	2519	112	311	0.260	0.599	2	3
Summer	918	88	233	0.085	0.443	4	4

Table 4 shows that winter leads both photographic volume and BA-DTVI, while autumn overtakes spring after the concentration adjustment despite its lower photo volume.

At the calendar-season level, winter ranks first in photographic volume, contributor breadth, contributor-days, and BA-DTVI. Autumn rises above spring after concentration adjustment despite lower photographic volume, illustrating again that distributed visibility is not determined by volume alone. These calendar-season labels are descriptive and are not interpreted as evidence of climatic causation.

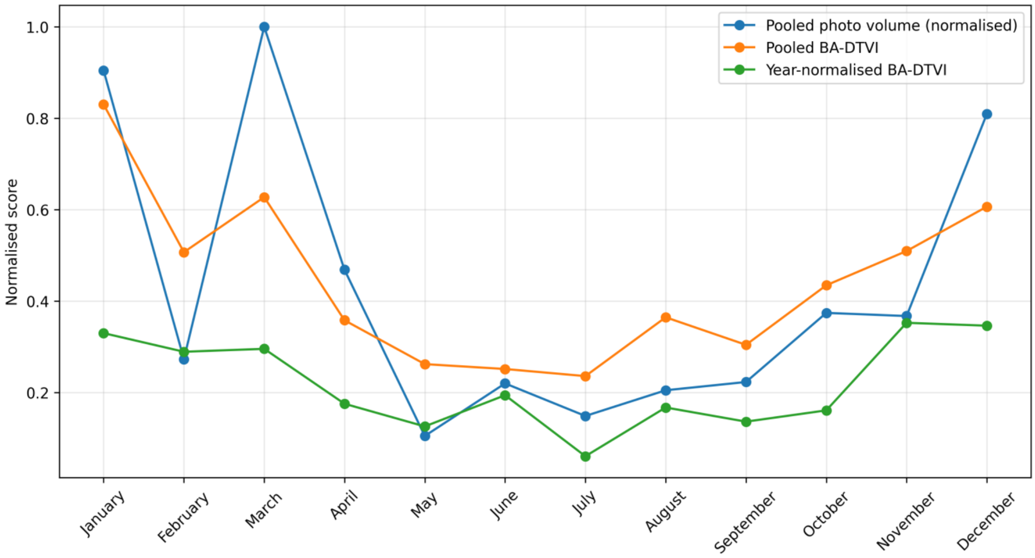

Figure 4. Pooled monthly photo volume and BA-DTVI compared with the year-normalised monthly BA-DTVI pattern.

Figure 4 demonstrates why the pooled and within-year normalised monthly views are complementary: pooled evidence emphasises January and March, whereas year-normalisation shifts recurrent BA-DTVI prominence toward November and December.

4.5 State-level photo baseline and conservative ranking

The photo-level and conservative contributor-cell state analyses identify the same leading geographic hierarchy. Khartoum ranks first under BA-DTVI in both branches, followed by Northern State, River Nile, and Red Sea. The principal contrast is between Khartoum and Northern State. Their conservative record volumes are similar (946 versus 914), but Khartoum is supported by 198 contributors and 520 contributor-days, compared with 66 contributors and 188 contributor-days in Northern State. This broader and less concentrated participation produces BA-DTVI values of 0.981 and 0.458, respectively. The result illustrates why similar platform-content volumes should not automatically be interpreted as equivalent distributed visibility. State membership is based on spatial ADM1 assignment rather than the raw Flickr region metadata.

Table 5. Leading states under the photo-level baseline and conservative contributor-cell analysis.

State	Photo P	Photo U	Photo UD	Photo BA	Photo rank	Core P	Core U	Core UD	Core BA	Core rank
Gedaref	69	14	27	0.032	6	43	14	22	0.041	6
Gezira	74	12	29	0.033	5	—	—	—	—	—
Khartoum	2,404	198	722	0.918	1	946	198	520	0.981	1
North Darfur	—	—	—	—	—	27	18	24	0.045	5
Northern	2,899	66	218	0.413	2	914	66	188	0.458	2
Red Sea	301	23	50	0.077	4	116	23	40	0.083	4
River Nile	1,755	51	140	0.287	3	429	51	116	0.280	3

Table 5 indicates a stable leading geographic hierarchy across the two analytical branches: Khartoum ranks first under BA-DTVI in both, even though Northern State leads photo-level volume, reflecting Khartoum's broader contributor participation.

Note. The table presents the union of the six highest-ranked states in each analytical branch. An em dash indicates that the state falls outside the top six for that branch, not that the underlying state-level data are missing. State membership is assigned by spatial ADM1 point-in-polygon intersection.

4.6 Correspondence between 46 candidate hotspots and 20 conservative cores

The 20-core solution is not an independent alternative clustering pattern but a conservative subset of the 46-candidate structure. Of the 46 photo-level candidates, 20 persist as dominant one-to-one correspondences after contributor-cell normalisation, 25 disappear entirely, and one candidate—Merowe—is almost entirely removed, with only one retained border record absorbed into the neighbouring Tangasi core. All 20 conservative cores have an identifiable dominant photo-level candidate source, and the 20 dominant sources are unique.

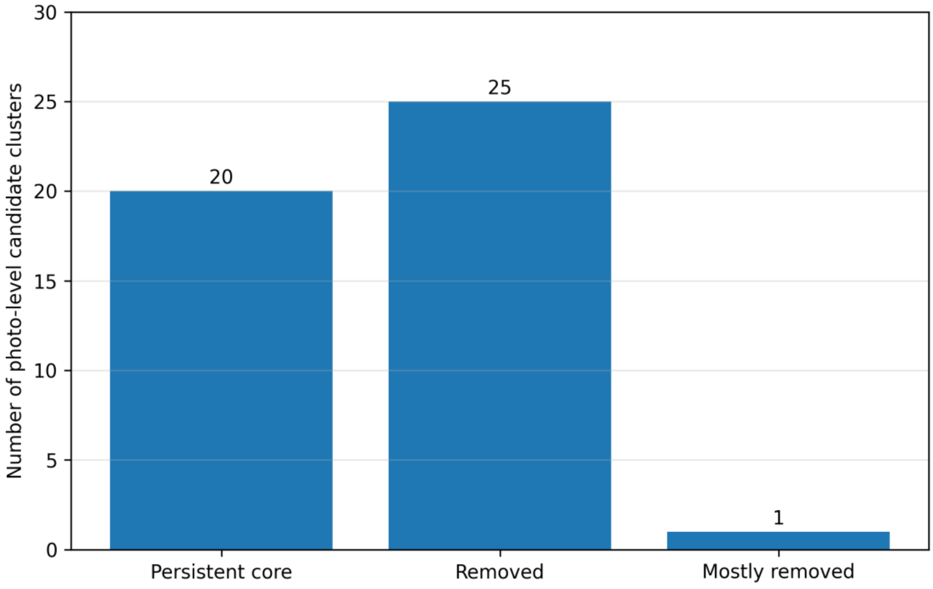

Figure 5. Outcomes of the 46 photo-level candidate hotspots after contributor-cell normalisation.

Figure 5 summarises structural persistence: 20 candidates remain as conservative cores, 25 disappear after normalisation, and one is almost entirely removed, demonstrating the effect of repeated local photography on the original hotspot inventory.

The strongest examples include Marawi (C46-00 → C20-00; 172/172 retained observations, 100% coverage) and Khartoum (C46-09 → C20-05; 796/810, 98.3%). By contrast, the Merowe candidate retains 23 contributor-cell observations but contributes only one observation to the Tangasi core (4.3% coverage). These results demonstrate that contributor-cell normalisation primarily removes spatial structures supported by intensive local photography while preserving the principal urban, Nile-corridor heritage, and Red Sea structures.

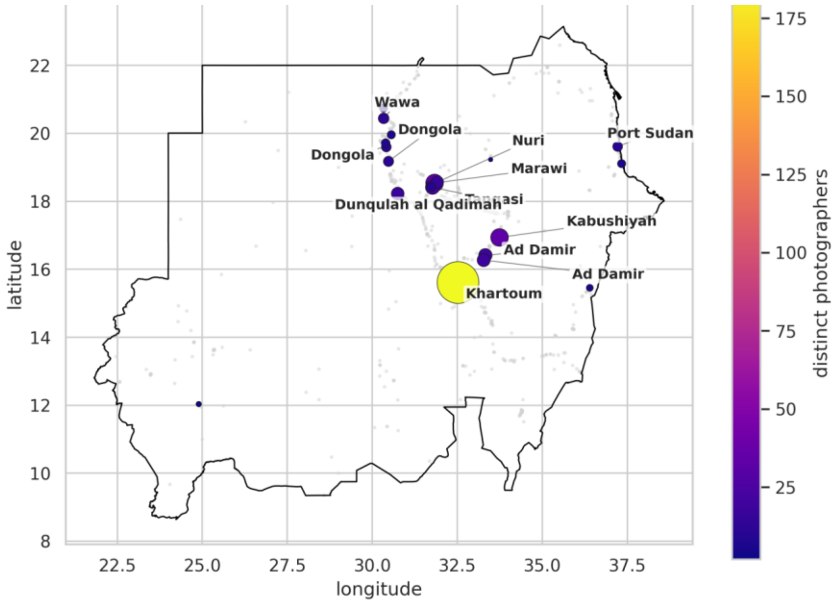

Figure 6. Geographic distribution of the 20 contributor-normalised core hotspots; marker size and colour represent contributor breadth.

Figure 6 shows that the conservative cores remain concentrated around Khartoum, the Nile Valley corridor, and the Red Sea, while sparse peripheral representation remains visible as a limitation of the Flickr-derived evidence.

4.7 Conservative core-hotspot visibility

Khartoum is the dominant conservative core under every principal visibility indicator, with 796 retained observations, 185 contributors, 444 contributor-days, HHI = 0.019, and BA-DTVI = 0.981. The next two cores—Kabushiyah/Meroë (0.160) and

Marawi/Karima (0.156)—form a substantially smaller but clearly identifiable heritage tier. These values represent relative digital visibility within the contributor-cell analytical scope and should not be interpreted as rankings of intrinsic tourism importance, archaeological value, or actual visitor numbers.

Table 6. Leading conservative core hotspots according to BA-DTVI.

ID	Place / validation anchor	State	P	U	UD	HHI	BA-DTVI	Rank
C20-05	Khartoum / Greek Orthodox Church of the Annunciation	Khartoum	796	185	444	0.019	0.981	1
C20-02	Kabushiyah / Pyramids of Meroe	River Nile	189	33	54	0.074	0.160	2
C20-00	Marawi / Karima Museum	Northern	172	33	52	0.058	0.156	3
C20-06	Dunqulah al Qadimah / Old Dongola	Northern	86	17	20	0.109	0.068	4
C20-04	Ad Damir / Naqa	River Nile	54	18	22	0.080	0.063	5
C20-03	Ad Damir / Musawarat	River Nile	50	19	22	0.073	0.063	6
C20-07	Tangasi / Q1430459	Northern	46	21	21	0.078	0.062	7
C20-01	Nuri	Northern	53	16	16	0.097	0.053	8
C20-16	Wawa / Temple of Soleb	Northern	34	12	14	0.145	0.038	9
C20-10	Port Sudan / Badi	Red Sea	28	10	13	0.250	0.029	10
C20-17	Dongola / Tabo	Northern	16	11	12	0.109	0.028	11
C20-15	Jazirat Say / Saï	Northern	31	7	13	0.197	0.028	12

Table 6 confirms the strong separation between Khartoum and the remaining conservative cores; the next highest BA-DTVI scores are associated with Nile Valley heritage locations rather than approaching the metropolitan score.

The two low-support cores illustrate why concentration adjustment is necessary. The unnamed River Nile core has only two contributors, and Nyala has three, with dominant-contributor shares of 0.55 and 0.91 respectively. Their BA-DTVI values are therefore markedly lower than their record counts alone would imply.

4.8 Ablation and contributor-removal robustness

Ablation confirms that the index is not a disguised photo count. At the annual level, P correlates with DTVI at Spearman ρ = 0.926 but with BA-DTVI at ρ = 0.862; DTVI and BA-DTVI correlate at ρ = 0.967. The concentration adjustment therefore changes rankings most where volume and participation are concentrated. At the pooled-monthly level, P and BA-DTVI show a lower agreement of ρ = 0.839.

Contributor-removal experiments provide the strongest direct evidence that BA-DTVI is less dependent on dominant contributors than raw record volume in several analytical scopes. After removing the top 10% of contributors, annual photo-volume stability falls to ρ = 0.807, whereas BA-DTVI retains ρ = 0.941. For the conservative state analysis, the corresponding values are 0.764 and 0.985. The largest contrast occurs among the 46 candidate hotspots: photo-volume stability falls to ρ = 0.373 with a maximum rank shift of 24 positions, while BA-DTVI retains ρ = 0.798 with a maximum shift of 16. The 20-core configuration is more sensitive because several cores contain small contributor populations, but the leading Khartoum core remains unchanged. These experiments are ranking-robustness tests within fixed analytical units; DBSCAN was not rerun after contributor removal.

Table 7. Rank stability after removing the top 10% of contributors.

Scope	P ρ	BA ρ	P max shift	BA max shift	P leader	BA leader
Annual	0.807	0.941	7	5	2012	2013
States—core	0.764	0.985	7	2	Khartoum	Khartoum
46 candidates	0.373	0.798	24	16	9	9
20 cores	0.616	0.694	11	10	5	5

Table 7 shows that BA-DTVI generally retains stronger rank stability than raw photo volume after removal of the most active contributors, with the largest relative advantage occurring in the 46-candidate-hotspot scope.

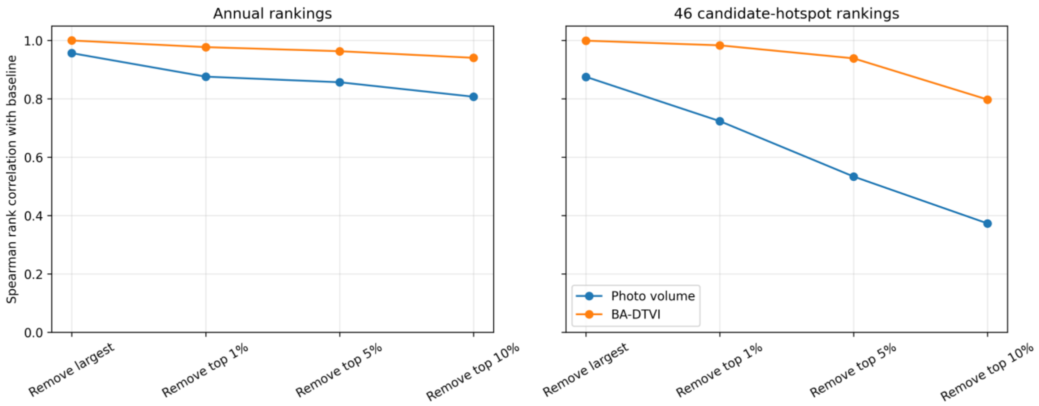

Figure 7. Spearman rank stability of photo volume and BA-DTVI under progressive contributor removal.

Figure 7 makes the robustness advantage visible across progressive removal scenarios: BA-DTVI correlations decline more slowly than photo-volume correlations, particularly for the 46-candidate-hotspot ranking.

4.9 Weight, adjusted-HHI, and duplicate sensitivity

BA-DTVI remains highly stable under all three alternative weighting specifications. Annual Spearman correlations relative to the equal-weight baseline range from 0.979 to 0.992, all alternative specifications preserve the same annual top five, and 2013 remains first. For the 20 conservative cores, correlations range from 0.982 to 0.992 and Khartoum remains first.

Finite-sample HHI adjustment produces no rank changes for annual, pooled-monthly, calendar-season, or conservative state analyses. Reordering is confined to lower-support hotspot units, with maximum shifts of six positions among the 46 candidates and two positions among the 20 conservative cores.

Duplicate sensitivity provides an additional data-audit robustness check. Removing the 30 duplicate-excess rows reduces the photo-level branch from 8,162 analytical rows to 8,132 distinct photo IDs but leaves the annual BA-DTVI ordering unchanged (Spearman $\rho = 1.000$). Rerunning DBSCAN on the duplicate-free branch again produces 46 candidate clusters, and the duplicate-free partition is identical to the retained-record partition of the principal solution (adjusted Rand index = 1.000). Together, these tests indicate that the principal rankings are not artifacts of a particular reasonable weighting specification, the minimum HHI implied by contributor count, or the 30 audited duplicate-excess rows.

4.10 Geographic plausibility and confidence-aware cross-platform validation

Geographic plausibility remains strong: 15 of the 20 conservative core centroids fall within 5 km of a site in the curated 69-site inventory, with a median nearest-site distance of 1.02 km. The inventory provides particularly strong coverage of archaeological and Nile-corridor heritage locations, whereas several urban, coastal, and broad regional cores remain less well represented.

Close geographic matches include Meroë/Kabushiyah, Marawi/Jebel Barkal, Nuri, Old Dongola, Musawarat, Naqa, El-Kurru, Soleb, Kerma, Tombos, Saï, Sabu, and Kurgus. More distant or broad-context matches for Port Sudan, Suakin, Kassala, and Nyala reinforce that the curated inventory is more comprehensive for archaeological and Nile-corridor heritage sites than for urban, coastal, marine, or broad regional destinations.

Table 8. Confidence-aware cross-platform associations between core-hotspot BA-DTVI and external digital-visibility indicators.

Indicator	Primary n	Primary ρ	Primary p	Sensitivity n	Sensitivity ρ	Sensitivity p
Wikimedia Commons images	14	0.665	0.0095	20	0.711	0.0004
Wikipedia article images	14	−0.035	0.9044	20	0.213	0.3680
Wikipedia page views	12	0.168	0.6021	17	0.287	0.2644
Wikipedia article length	14	0.218	0.4549	20	0.371	0.1069
Language editions	14	0.106	0.7194	20	0.191	0.4204
Wikidata sitelinks	14	0.095	0.7477	20	0.134	0.5730

As shown in Table 8, the primary analysis uses High- and Medium-confidence entity matches, while the sensitivity analysis uses all available matches. Confidence-aware cross-platform validation provides selective rather than universal convergence. Wikimedia Commons image availability shows the only statistically significant association with BA-DTVI in the primary analysis (Spearman $\rho = 0.665$, $p = 0.0095$, $n = 14$). The association remains strong in the broader all-match sensitivity analysis ($\rho = 0.711$, $p = 0.0004$, $n = 20$). In contrast, Wikipedia article-image counts, page views, article length, language editions, and Wikidata sitelinks are not statistically significant in either analysis. These indicators capture related but non-interchangeable forms of online prominence, and none is interpreted as a measure of actual tourist arrivals or visitor demand.

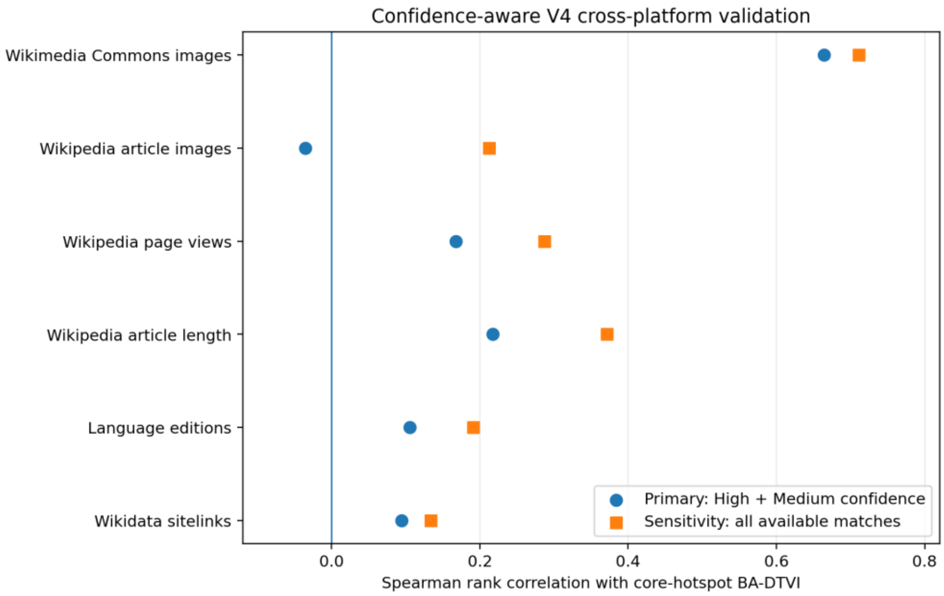

Figure 8. Confidence-aware V4 cross-platform validation of conservative core-hotspot BA-DTVI. Circles show the primary High- and Medium-confidence analysis; squares show the all-match sensitivity analysis.

Figure 8 visually confirms that the positive cross-platform association is concentrated in Wikimedia Commons, while the other Wikipedia and Wikidata indicators remain comparatively weak in both the primary and all-match sensitivity analyses.

5. DISCUSSION

5.1 Photo counts are informative but incomplete

The results do not suggest that photographic volume should be discarded. Photo count remains an informative measure of the intensity of platform representation and provides the full observational basis from which the candidate spatial structure is identified. Its limitation is interpretive: similar numbers of photographs can be generated by substantially different participation processes.

BA-DTVI makes these differences explicit by requiring simultaneous support from photographic volume, contributor breadth, and contributor-day participation, while reducing the resulting score when observations are highly concentrated among a small number of contributors. The contrast between 2018 and 2013 demonstrates this distinction particularly clearly. In 2018, the Flickr archive reaches its highest photographic volume, but the evidence is generated by relatively few contributors and contributor-days and is strongly concentrated. In 2013, a smaller photo total is supported by broader participation, greater temporal representation, and lower contributor concentration, producing the highest DTVI and BA-DTVI.

The implication is therefore not that raw counts are incorrect, but that they answer a narrower question. Photo counts measure how much content is present, whereas BA-DTVI assesses how broadly and persistently that content is distributed across contributors. High correlation between the two measures does not guarantee equivalent ranking or interpretation.

5.2 Contributor-cell normalisation reduces repetition but does not eliminate contributor inequality

The reduction from 8,162 photo-level records to 2,686 contributor-cell records demonstrates that repeated local photography forms a substantial component of the Flickr archive. However, contributor-cell normalisation should not be interpreted as complete bias removal.

The contributor Gini coefficient decreases from 0.836 to 0.759 after normalisation, confirming that repeated local observations contribute to inequality. Nevertheless, the top 10% of contributors still account for 68.7% of the retained contributor-cell records. A highly active contributor can therefore remain influential by documenting many distinct spatial cells even when repeated observations within any one cell are constrained.

This finding supports a layered bias-control strategy. Contributor-cell normalisation operates at the observation level by limiting repeated local representation, whereas HHI operates at the analytical-unit level by measuring the remaining concentration of evidence among contributors. The two operations address different sources of bias and should not be treated as substitutes.

The distinction also clarifies terminology. The 2,686-record transformation is not conventional database deduplication: valid photographs are not declared erroneous duplicates. Instead, it is an analytical normalisation designed to test how spatial structures change when repeated local contribution is constrained.

5.3 Temporal interpretation depends on the aggregation perspective

The temporal results demonstrate that digital visibility is sensitive not only to indicator choice but also to temporal aggregation. March produces the largest pooled monthly photo count, whereas January ranks first under pooled BA-DTVI. After within-year normalisation, November becomes the leading month, followed by December, January, and March. These results are not contradictory. They describe different aspects of the archive.

Pooled monthly values measure the cumulative platform evidence accumulated across the entire observation period and can therefore be strongly influenced by unusually high-volume years. Within-year normalisation instead asks whether a month repeatedly occupies a prominent position across complete years and reduces domination by years containing exceptionally large Flickr volumes.

At the broader calendar-season level, winter remains the leading period across photographic volume, contributor breadth, contributor-days, and BA-DTVI. Autumn rises above spring after concentration adjustment despite having lower photographic volume, further demonstrating that distributed visibility and content volume are not identical constructs.

These patterns should nevertheless remain descriptive rather than causal. The Flickr metadata do not directly observe climate, motivation, trip purpose, nationality, residence, or travel constraints. Consequently, the results establish a temporal structure in platform-derived digital visibility but do not demonstrate that cooler weather caused tourism activity. The distinction is particularly important because the observation period is treated as a historical pre-war baseline, not as a representation of current conditions in Sudan.

5.4 Spatial persistence and wider place visibility

The change from 46 photo-level candidate hotspots to 20 contributor-cell conservative cores should not be interpreted as evidence that 26 photo-level hotspots were simply false. Rather, the two solutions measure spatial concentration under different observational units. The photo-level configuration preserves the full photographic evidence and therefore identifies structures supported both by broad participation and by intensive local documentation. Contributor-cell normalisation asks a more conservative question: which spatial structures retain sufficient density after repeated local contributions by the same contributor are constrained?

The candidate-to-core tracing provides stronger evidence than a visual comparison of two maps. Twenty candidate structures persist as dominant one-to-one correspondences, 25 disappear after normalisation, and one candidate is almost entirely removed, with only a border observation incorporated into a neighbouring core. All 20 conservative cores retain identifiable dominant candidate sources.

This distinction supports the interpretation of the Khartoum metropolitan concentration, the Nile Valley heritage corridor, and the Red Sea nodes as persistent structures in the Flickr-derived evidence. At the same time, the disappearance of smaller candidates demonstrates that hotspot inventories remain sensitive to analytical unit and contribution behaviour.

Khartoum requires particularly careful interpretation. Its leading position is supported by a very broad contributor base, many contributor-days, and low contributor concentration, so its prominence cannot be explained by one or two intensive accounts. However, Khartoum is simultaneously an administrative, residential, commercial, transport, and cultural centre. Its BA-DTVI should therefore be interpreted as strong and broadly distributed digital place visibility, not as a direct estimate of tourism intensity.

The heritage cores along the Nile corridor—including the Meroë/Kabushiyah area, Marawi/Karima, Old Dongola, Naqa, and Musawarat—have considerably smaller digital footprints. Lower BA-DTVI values do not imply lower cultural, archaeological, or tourism significance. They indicate that these sites generated less broadly distributed Flickr evidence during the observed period. Similarly, sparse Flickr observations in western and southern Sudan should be interpreted as an evidence gap rather than proof that the corresponding locations have low tourism value.

5.5 Robustness strengthens the interpretation of BA-DTVI

The strongest support for BA-DTVI comes from the convergence of several independent robustness tests rather than from any single correlation. Contributor-removal experiments demonstrate that raw photographic-volume rankings are more sensitive to intensive contributors than BA-DTVI rankings. This contrast is particularly strong among the 46 candidate hotspots, while annual and conservative state rankings also show substantially greater stability under BA-DTVI.

Alternative weighting produces annual rank correlations of approximately 0.979–0.992 relative to the equal-weight model and approximately 0.982–0.992 for the 20 conservative cores. The leading annual unit remains 2013 under every tested weighting scheme, while Khartoum remains the leading conservative core. Thus, the main conclusions do not depend on assigning exactly one-third weight to each component.

The adjusted-HHI sensitivity analysis provides a complementary test. Correcting HHI for its contributor-count-dependent minimum produces no rank changes at the annual, monthly, seasonal, or conservative state levels. The effects are confined primarily to smaller hotspot units, with a maximum shift of six positions among the 46 candidates and two among the 20 cores.

The duplicate sensitivity analysis further strengthens the audit. Removing the 30 duplicate-excess rows produces 8,132 distinct photo IDs but leaves the complete annual BA-DTVI ordering unchanged, with Spearman $\rho = 1.000$. Rerunning DBSCAN also produces the same 46-cluster partition, with an adjusted Rand index of 1.000.

Taken together, these tests show that the main findings are not artifacts of one reasonable weighting scheme, the conventional HHI formulation, the audited duplicate rows, or a small number of highly active contributors. However, robustness should not be equated with absence of bias. The tests assess whether the conclusions are stable under explicitly defined analytical perturbations. They do not remove demographic, platform-adoption, connectivity, visitor-local, or geographic-coverage biases.

5.6 External validation supports selective, not universal, cross-platform convergence

The geographic validation provides an important first layer of contextual support. Fifteen of the 20 conservative core centroids lie within 5 km of a site in the curated tourism-relevant inventory. This correspondence is particularly strong for archaeological and Nile-corridor heritage locations, although the inventory is less comprehensive for some urban, coastal, and broad regional structures.

The updated confidence-aware cross-platform validation provides a more selective result. Wikimedia Commons image availability is significantly associated with BA-DTVI in the primary High- and Medium-confidence matches ($\rho = 0.665$, $p = 0.0095$, $n = 14$), and the relationship remains strong when all 20 available matches are included in sensitivity analysis ($\rho = 0.711$, $p = 0.0004$).

In contrast, Wikipedia article-image counts, page views, article length, language editions, and Wikidata sitelinks are not statistically significant under the improved matching procedure. The appropriate interpretation is therefore selective rather than universal cross-platform convergence.

The strong Wikimedia Commons relationship is conceptually plausible because both Flickr and Commons are visually oriented systems in which geographic places are represented through contributed imagery. Wikipedia readership, encyclopaedic article development, language coverage, and Wikidata connectivity reflect different processes. Their lack of significant association should not be treated as a failure of BA-DTVI; instead, it demonstrates that digital visibility is multidimensional and that different platforms are not interchangeable measurement instruments.

The external comparison also does not validate actual tourist numbers. Flickr and Wikimedia Commons remain user-generated digital systems subject to overlapping geographic, participation, and connectivity biases. The evidence therefore supports convergence in a dimension of visual digital place visibility, not equivalence with visitor arrivals or tourism demand.

5.7 Computational, methodological, and practical implications

The principal computer-science contribution is not a new clustering algorithm. Instead, the study demonstrates how a biased and sparse user-generated geospatial archive can be transformed into an auditable analytical system through explicit data lineage, dual analytical units, feature engineering, density-based spatial partitioning, interpretable composite-index construction, contributor-concentration adjustment, and structured robustness evaluation.

This transparency is important because every stage can be inspected independently. The 8,162-record branch preserves the complete photo-level evidence; the 2,686-record contributor-cell branch provides a conservative spatial representation; DBSCAN defines model-dependent comparison units; BA-DTVI decomposes into interpretable P, U, UD, and HHI components; and the final rankings are subjected to multiple sensitivity tests.

The approach is therefore more readily auditable than a black-box predictive model. It also requires only metadata that are common to many volunteered or location-based datasets: coordinates, timestamps, and persistent contributor identifiers. Transfer to another platform or country nevertheless requires recalibration of spatial cell size, clustering parameters, temporal interpretation, index normalisation, and validation sources rather than mechanical replication of the Sudan configuration.

For tourism and heritage applications, the model is most useful as a screening and prioritisation instrument. It can distinguish locations that are broadly represented across contributors from locations whose apparent prominence depends on intensive documentation by a small number of accounts. It can also identify locations where weak platform visibility may justify further investigation. It should not, however, be used to infer tourist arrivals, destination demand, economic value, or cultural importance in the absence of representative external data.

6. LIMITATIONS AND FUTURE WORK

6.1 Platform and population representativeness

Flickr contributors constitute a self-selected population and cannot be treated as a statistically representative sample of tourists, residents, or the general population. Platform participation depends on internet access, device ownership, photographic interests, language, nationality, age, and willingness to share geotagged content publicly. Platform popularity and posting behaviour also changed substantially during the observation period.

Consequently, the absence of Flickr records does not imply an absence of tourism activity, and high digital visibility may reflect tourists, residents, researchers, professional photographers, journalists, transit passengers, or other contributors. BA-DTVI reduces the influence of intensive contributors within the observed platform but does not correct demographic, geographic, connectivity, or platform-access bias.

Future studies should compare platform-derived visibility with representative sources such as official visitor statistics, accommodation records, ticketing systems, surveys, transport data, protected-area counts, or other institutional monitoring sources where these are available.

6.2 Tourist, resident, and contributor classification

The available metadata do not provide a definitive classification of contributors as international tourists, domestic tourists, residents, workers, professional photographers, or transit passengers. This limitation is particularly relevant to large urban centres such as Khartoum, where digital place visibility may combine tourism-related photography with broader urban documentation.

Future work could examine residence inference, repeated presence, trip duration, temporal gaps, and movement histories to distinguish visitor-like from resident-like patterns. Such extensions would require explicit validation and strict privacy safeguards and should remain at aggregate analytical levels rather than attempting to profile identifiable individuals.

6.3 Analytical units and the meaning of the contributor-cell record

The 2,686-record contributor-cell branch represents contributor-location observations, not visits, trips, tourists, or independent tourism events. A single journey may contribute records to multiple spatial cells, whereas repeated appearances by the same contributor in the same approximately 110 m cell are collapsed across the full study period.

The 976 contributor-days represented in this branch therefore describe the dates attached to the retained source observations only. They should not be interpreted as a complete temporal participation count for the contributor-cell branch or as independent visits.

The separation of the 8,162-record photo-level temporal branch from the 2,686-record conservative spatial branch resolves this limitation for the present study. Future research could introduce contributor-cell-day or other date-preserving analytical units to evaluate local temporal persistence while retaining control over intensive same-location photography.

6.4 Duplicate records and source-data uncertainty

The principal photo-level branch retains 8,162 analytical rows for consistency with the established data lineage, although these rows contain 8,132 distinct photo IDs and 30 duplicate-excess observations.

This issue is now empirically bounded rather than unresolved. Removing the 30 duplicate-excess rows leaves the annual BA-DTVI ordering unchanged (Spearman ρ = 1.000) and reproduces the same 46-cluster DBSCAN partition (adjusted Rand index = 1.000). The duplicate rows therefore do not alter the principal conclusions of this study.

Nevertheless, future applications should enforce photo-ID uniqueness as early as possible in the data-ingestion pipeline and preserve explicit audit logs so that database duplication, behavioural repetition, and analytical normalisation remain conceptually distinct.

6.5 Spatial scale and density-based clustering

The three-decimal contributor-cell resolution, corresponding to approximately 110 m, and the DBSCAN reference parameters of ε = 2 km and MinPts = 15 are documented modelling choices rather than universally optimal spatial scales.

Different cell sizes or clustering parameters may split neighbouring attractions, merge adjacent structures, or change the representation of sparsely documented locations. The present analysis demonstrates that the broad Khartoum, Nile Valley, and Red Sea structures remain robust, but the exact hotspot inventory remains dependent on the analytical unit and density threshold.

Future work should therefore evaluate multiple spatial resolutions, adaptive cell definitions, and variable-density clustering approaches such as HDBSCAN (Campello et al., 2013) or related methods. These alternatives should be treated as comparative extensions rather than assumed improvements, particularly because a single national dataset contains both dense metropolitan observations and sparse archaeological, desert, and coastal locations.

6.6 Temporal coverage and historical interpretation

The observation period begins on 23 November 2002 and ends on 17 February 2019. The first and final years are therefore partial, and Flickr adoption and geotagging practices changed during the study period.

The results should be interpreted as a historical pre-war digital-visibility baseline. They should not be extrapolated directly to contemporary tourism conditions in Sudan, especially across subsequent periods of political transition, pandemic disruption, armed conflict, population displacement, site inaccessibility, and changing platform use.

Future post-2019 analysis should be constructed as a separate, conflict-sensitive temporal dataset with explicit consideration of changes in platform use, population displacement, site accessibility, safety, and the interpretation of location-based digital traces. Direct combination with the historical baseline without such controls would risk conflating tourism change with platform and societal disruption.

6.7 Composite-index design and statistical uncertainty

BA-DTVI uses max normalisation, an equal-weight geometric mean, and an HHI-based concentration adjustment. These choices are transparent and interpretable but are not uniquely optimal for every application.

The robustness analysis shows that the main annual and leading spatial conclusions remain stable under alternative weights and finite-sample HHI adjustment. Nevertheless, small analytical units may still obtain apparently favourable concentration measures from only a few contributors, and within-scope max normalisation limits direct numerical comparison across different analytical scopes or independently collected datasets.

Future work should investigate robust scaling, logarithmic transformation, fixed-reference normalisation, minimum-support thresholds, bootstrap confidence intervals, uncertainty propagation, and out-of-sample calibration. Expert-derived, entropy-based, or empirically calibrated weighting schemes could also be compared where stronger external reference data are available.

6.8 Geographic and external validation

The curated 69-site inventory provides useful geographic plausibility evidence but is not a complete tourism inventory. Its coverage is stronger for archaeological and Nile Valley heritage sites than for urban, coastal, marine, natural, and less extensively mapped destinations. Consequently, long centroid-to-site distances in underrepresented regions may reflect limitations of the reference inventory rather than invalid hotspot detection.

Wikipedia, Wikidata, Wikimedia Commons, and Flickr are also digitally mediated systems with their own contributor, language, editorial, geographic, and iconicity biases. The confidence-aware V4 matching procedure reduces entity-matching ambiguity, but cross-platform correlations remain evidence of convergent digital visibility rather than validation of physical visitor numbers.

The updated results demonstrate selective rather than universal convergence: Wikimedia Commons provides a robust external association with BA-DTVI, whereas the other tested Wikipedia and Wikidata indicators do not show statistically significant associations under the improved matching procedure.

Future validation should therefore prioritise independent physical and institutional evidence, including official visitor counts, heritage-management records, accommodation data, field observation, expert assessment, locally curated destination inventories, and other appropriately matched sources. Cross-platform matching should continue to report confidence levels and sensitivity analyses rather than forcing uncertain entity correspondences.

6.9 Transferability and future computational development

The framework is designed to be transferable to other countries and location-based platforms when coordinates, timestamps, and persistent contributor identifiers are available. However, transferability should be evaluated rather than assumed. Different platforms may exhibit substantially different contributor inequalities, spatial behaviours, API restrictions, temporal coverage, and population biases.

Future computational development should examine multi-country and cross-platform replication; adaptive spatial and temporal analytical units; variable-density clustering benchmarks; minimum-support and uncertainty-aware BA-DTVI reporting; scalable spatial indexing and parallel or cloud-based processing for substantially larger datasets; reproducible provenance tracking from API acquisition through final analytical outputs; and interactive GIS visualisation for exploring uncertainty, component scores, and hotspot persistence.

These extensions should preserve the central principle of the present study: computational complexity should increase only when it provides measurable analytical benefit and should not obscure the provenance, assumptions, uncertainty, or interpretation of the resulting digital-visibility measures.

7. Conclusion

This study developed and evaluated a bias-aware multi-indicator model for measuring digital tourism visibility from geo-tagged Flickr metadata in a data-scarce context. The central methodological premise is that photographic volume alone cannot distinguish broadly distributed digital attention from intensive documentation by a small number of contributors. The proposed BA-DTVI therefore integrates content volume, contributor breadth, contributor-day participation, and within-unit contributor concentration within an interpretable computational framework.

The final analytical design separates two complementary units of evidence. The complete photo-level branch retains 8,162 in-boundary records for temporal analysis and produces 46 candidate hotspots, whereas contributor-cell normalisation reduces repeated local representation to 2,686 records and identifies 20 conservative core hotspots. Direct record-level correspondence confirms that 20 photo-level candidates persist as dominant one-to-one cores, 25 disappear after normalisation, and one candidate is almost entirely removed, with only a border observation absorbed into an adjacent core. This dual-unit design demonstrates that hotspot structure depends not only on geographic density but also on how contributor behaviour is represented.

The temporal results further show that content volume and distributed visibility provide related but non-equivalent interpretations. The year 2018 contains the largest photographic volume, whereas 2013 ranks first under both DTVI and BA-DTVI. March has the largest pooled monthly photo count, January the highest pooled BA-DTVI, and November leads after within-year normalisation. Spatially, Khartoum remains the strongest conservative core, while persistent Nile Valley heritage and Red Sea structures demonstrate that the principal national geography is retained under contributor-aware normalisation. These outcomes describe patterns in platform-derived digital visibility and should not be interpreted as estimates of tourist arrivals or destination demand.

The robustness analyses provide additional support for the measurement framework. BA-DTVI rankings are generally more stable than raw photo-volume rankings when dominant contributors are removed. The leading conclusions also remain stable under alternative component weights and finite-sample HHI adjustment. Removing the 30 duplicate-excess rows leaves the annual BA-DTVI ordering unchanged and reproduces the same 46-cluster spatial partition, providing further evidence that the principal findings are not driven by the audited duplicates or by one specific reasonable model parameterisation.

Geographic validation shows that 15 of the 20 conservative core hotspots lie within 5 km of a site in the curated tourism-relevant inventory. The updated confidence-aware cross-platform analysis provides selective rather than universal convergence. Wikimedia Commons image availability is significantly associated with BA-DTVI in the primary High- and Medium-confidence matches (Spearman $\rho = 0.665$, $p = 0.0095$, $n = 14$), and the relationship remains significant when all available matches are included in sensitivity analysis ($\rho = 0.711$, $p = 0.0004$, $n = 20$). The other tested Wikipedia and Wikidata indicators are not statistically significant. This result supports convergence in visual digital place visibility while also demonstrating that different digital platforms capture related but non-interchangeable dimensions of place prominence.

The principal contribution is therefore not a claim that Flickr provides a census of tourism activity, nor the introduction of a new clustering algorithm. Rather, the study contributes a reproducible computational GIS protocol for transforming sparse and unequal user-generated geospatial data into more interpretable evidence. Explicit data lineage, dual analytical units, density-based spatial partitioning, multidimensional feature construction, contributor-concentration adjustment, candidate-to-core traceability, and structured robustness testing together make the analytical process auditable rather than dependent on a single opaque score or raw content count.

Sudan serves as the empirical testbed rather than the boundary of the contribution. The framework can be transferred to other data-scarce territories and location-based platforms where persistent contributor identifiers, geographic coordinates, and timestamps are available, provided that spatial scales, platform behaviour, clustering parameters, normalisation choices, and validation sources are recalibrated for the new context.

The main conclusion is therefore not that photo counts should be discarded. They should instead be interpreted alongside who generated the content, how broadly participation was distributed, how persistently contributors appeared across dates, and how concentrated the available evidence remained. BA-DTVI provides one transparent means of making those distinctions while preserving the uncertainty required when digital traces are used to support spatial knowledge in data-scarce contexts.

REFERENCES

[1]. Barros, C., Moya-Gómez, B., & Gutiérrez, J. (2020). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. Current Issues in Tourism, 23(10), 1291–1310. https://doi.org/10.1080/13683500.2019.1619674

[2]. Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), Advances in Knowledge Discovery and Data Mining (pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

[3]. Derdouri, A., & Osaragi, T. (2021). A machine learning-based approach for classifying tourists and locals using geotagged photos: The case of Tokyo. Information Technology & Tourism, 23(4), 575–609. https://doi.org/10.1007/s40558-021-00208-3

[4]. Domènech, A., Mohino, I., & Moya-Gómez, B. (2020). Using Flickr geotagged photos to estimate visitor trajectories in World Heritage Cities. ISPRS International Journal of Geo-Information, 9(11), 646. https://doi.org/10.3390/ijgi9110646

[5]. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, & U. Fayyad (Eds.), Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (pp. 226–231). AAAI Press.

[6]. Ghermandi, A. (2022). Geolocated social media data counts as a proxy for recreational visits in natural areas: A meta-analysis. Journal of Environmental Management, 317, 115325. https://doi.org/10.1016/j.jenvman.2022.115325

[7]. Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. Social Indicators Research, 141(1), 61–94. https://doi.org/10.1007/s11205-017-1832-9

[8]. Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. Conservation Letters, 11(1), e12343. https://doi.org/10.1111/conl.12343

[9]. Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. ISPRS International Journal of Geo-Information, 6(3), 85. https://doi.org/10.3390/ijgi6030085

[10]. Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. Computers, Environment and Urban Systems, 54, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

[11]. Höpken, W., Müller, M., Fuchs, M., & Lexhagen, M. (2020). Flickr data for analysing tourists' spatial behaviour and movement patterns: A comparison of clustering techniques. Journal of Hospitality and Tourism Technology, 11(1), 69–82. https://doi.org/10.1108/JHTT-08-2017-0059

[12]. Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. Tourism Geographies, 16(1), 88–104. https://doi.org/10.1080/14616688.2013.868029

[13]. Kádár, B., & Gede, M. (2013). Where do tourists go? Visualizing and analysing the spatial distribution of geotagged photography. Cartographica: The International Journal for Geographic Information and Geovisualization, 48(2), 78–88. https://doi.org/10.3138/carto.48.2.1839

[14]. Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. Tourism Management, 68, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

[15]. Martins, M., & Santos, A. (2024). Exploring the potential of Flickr user-generated content for tourism research: Insights from Portugal. European Journal of Tourism, Hospitality and Recreation, 14(2), 258–272. https://doi.org/10.2478/ejthr-2024-0019

[16]. Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. GeoJournal, 80(4), 449–461. https://doi.org/10.1007/s10708-014-9602-6

[17]. Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2016). Tracing tourists by their digital footprints: The case of Austria. Journal of Travel Research, 55(5), 566–573. https://doi.org/10.1177/0047287514563985

[18]. Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society: Series A (Statistics in Society), 168(2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

[19]. Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. Tourism Management, 66, 13–25. https://doi.org/10.1016/j.tourman.2017.11.001

[20]. Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems, 42(3), Article 19, 1–21. https://doi.org/10.1145/3068335

[21]. Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. International Journal of Geographical Information Science, 31(1), 139–167. https://doi.org/10.1080/13658816.2016.1189556

[22]. Song, X. P., Richards, D. R., He, P., & Tan, P. Y. (2020). Does geo-located social media reflect the visit frequency of urban parks? A city-wide analysis using the count and content of photographs. Landscape and Urban Planning, 203, 103908. https://doi.org/10.1016/j.landurbplan.2020.103908

[23]. Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific Reports, 7, 17615. https://doi.org/10.1038/s41598-017-18007-4

[24]. Wilkins, E. J., Wood, S. A., & Smith, J. W. (2021). Uses and limitations of social media to inform visitor use management in parks and protected areas: A systematic review. Environmental Management, 67(1), 120–132. https://doi.org/10.1007/s00267-020-01373-7

[25]. Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. Scientific Reports, 3, 2976. https://doi.org/10.1038/srep02976

[26]. Wood, S. A., Winder, S. G., Lia, E. H., White, E. M., Crowley, C. S. L., & Milnor, A. A. (2020). Next-generation visitation models using social media to estimate recreation on public lands. Scientific Reports, 10, 15419. https://doi.org/10.1038/s41598-020-70829-x