



Research Article
2027 January - June



Advanced techniques for quantifying data-driven heritage: a cross study of statistical methods in heritage research

Técnicas avanzadas para cuantificar el patrimonio basado en datos: un estudio transversal de métodos estadísticos en la investigación del patrimonio

AMIT CHATTERJEE

Veer Surendra Sai University of Technology, India
amitchatterjee.ar@gmail.com

SWAPNALI LADPATIL

Hericon Consultants, India
ladpatilswapnali@gmail.com

BHARATI MOHPATRA

Veer Surendra Sai University of Technology, India
bharati_mohapatra@yahoo.com

ABSTRACT The quantification of cultural heritage value is increasingly supported by advanced statistical models, moving beyond traditional qualitative narratives. This paper presents a meta-analysis of 45 empirical studies (2010-2024) that applied statistical methods such as EFA, CFA, SEM and other multivariate methods to assess heritage value, perception and policy relevance. Studies were analysed for methodological transparency, model rigor and adaptability across contexts. Findings show significant variation in method selection, with SEM and CFA common in urban heritage research in Asia and Europe while PCA, cluster analysis and GWR are more prevalent in studies on landscapes and intangible heritage. The study reveals increasing use of spatial and perceptual models but also highlights underuse of spatial SEM and latent class modelling particularly in the Global South. This paper offers a comparative synthesis and practical guidance for selecting appropriate statistical methods in heritage research supporting more robust policy and Conservation strategies.

RESUMEN La cuantificación del valor del patrimonio cultural se sustenta cada vez más en modelos estadísticos avanzados, que van más allá de las narrativas cualitativas tradicionales. Este artículo presenta un metaanálisis de 45 estudios empíricos (2010-2024) que aplicaron métodos estadísticos como el AFE, el AFC, el SEM y otros multivariantes para evaluar el valor del patrimonio, la percepción y la relevancia política. Se analizaron los estudios en cuanto a transparencia metodológica, rigor del modelo y adaptabilidad a diferentes contextos. Los hallazgos muestran una variación significativa en la selección de métodos: el SEM y el AFC son comunes en la investigación del patrimonio urbano en Asia y Europa, mientras que el PCA, el análisis de conglomerados y el GWR son más frecuentes en estudios sobre paisajes y patrimonio inmaterial. El estudio revela un uso creciente de modelos espaciales y perceptuales, pero también destaca la infrautilización del SEM espacial y la modelización de clases latentes, especialmente en el Sur global. Este artículo ofrece una síntesis comparativa y una guía práctica para la selección de métodos estadísticos apropiados en la investigación del patrimonio que respalden políticas y estrategias de conservación más sólidas.

PALABRAS CLAVE heritage valuation, statistical methods, cultural heritage, meta analysis, spatial modelling

KEYWORDS valoración del patrimonio, métodos estadísticos, patrimonio cultural, metaanálisis, modelización espacial

Received: 01/06/2025
Reviewed: 23/09/2025
Accepted: 27/10/2025
Published: 17/08/2026



Cómo citar este artículo/How to cite this article: Chatterjee, A., Mohpatra, B. & Ladpatil, S. (2027). Advanced techniques for quantifying data-driven heritage: a cross study of statistical methods in heritage research. *Estoa. Revista de la Facultad de Arquitectura y Urbanismo de la Universidad de Cuenca*, 16(31), e6082. <https://doi.org/10.18537/est.v016.n031.a01>

1. Introduction

During the last few decades, the field of heritage conservation has gone through a dramatic change, as a result of not only socio-cultural needs but also of the growing necessity for a methodological rigour, in understanding, evaluating and conserving heritage in increasingly urbanised and digitised landscapes. Classical research in the field of heritage has been mainly qualitative, narrative and descriptive; but a paradigm shift is in place, which is oriented towards the processing of data, applying statistical and computational methods. Family settlements and vernacular settlements are increasingly threatened by development around cultural heritage and heritage precincts causing increased pressure to develop a suite of healthy, scalable and context sensitive methodologies that can quantify the values of the heritage and track their change over time. The new researches have involved multivariate statistics, structural modelling, choice modelling and spatial regressions which discuss the entire relationship between heritage resources, its socio-cultural settings and current pressures. As an example, Structural Equation Modelling (SEM) has become an effective instrument to investigate latent constructs (community attitudes, perceived heritage value, and conservation behaviour) (Al-Akaby, 2020; Boto-García et al., 2022). Through these methods, one can have a deeper insight into the actual variables, such as the perception of the people, identity, and satisfaction, which have been studied in the scope of the subject in Egyptian cultural centres or the UNESCO-protected landscapes in China and Malaysia (Boto-García et al., 2022; Mourato and Mazzanti, 2022). Meanwhile, geographically weighted regression (GWR) has progressed spatially sensitive heritage analysis, that is, the ability to measure spatial heterogeneity and local differences in heritage factors that are easy to miss when considering the world in toto (Wright and Eppink, 2016; Zhao et al., 2024). Evidence of the usefulness of the technique in spatial complexities analysis can be seen in its applications in cultural landscape analysis, including the traditional villages in the Hunan Province or inequalities in access to African land rights (Bonet et al., 2020; Fang and Li, 2022). Besides, modelling based on Bayesian networks has demonstrated its usefulness in demarcating the cultural development and the probability of interaction of design factors, especially in the study of Vietnam-based architectural heritage (Nguyen and Kieu, 2024).

Despite this methodological growth the field remains fragmented in terms of integrative outlines that compare and contrast these statistical approaches within heritage contexts. While separate case studies have illuminated the power of SEM, GWR, Bayesian networks and multivariate analysis across diverse heritage sites from the Qiansheng Bridge in China to maritime cultural traditions in German (Quiroga, 2025) a comprehensive, comparative review of their

efficiency, adaptability and context sensitivity remains lacking. There is also a vital need to align these methods with involved and inclusive heritage planning models as highlighted in research on community engagement in sites such as Melaka and the Bisotun World Heritage Site in Iran (Nasrolahi et al., 2022).

This paper addresses this critical gap by synthesizing and assessing advanced statistical methods used in modern heritage research. Through a cross study examination of SEM, GWR, Bayesian networks and multivariate analyses the research aims to (1) highlight their applicability in capturing both tangible and intangible dimensions of heritage, (2) evaluate their methodological robustness across different geographical contexts and (3) propose a conceptual roadmap for selecting suitable tools based on research objectives, data typologies and heritage typologies.

In doing so, this research contributes to the growing discourse on data driven heritage management by offering a meta-analytical approach upon which future research can build integrated evidence-based frameworks for heritage valuation, policy formation and community oriented conservation. By centring methodologies innovation with heritage studies, it seeks to empower heritage professionals, urban planners and researchers with tools that not only quantify heritage significance but also uphold its cultural character and contextual specificity in an ever-changing world.

While quantitative approaches provide clarity and replicability, they must also be read alongside qualitative heritage narratives such as oral histories, collective memory and identity to ensure that statistical models remain culturally grounded. Recent advances further point towards the integration of emerging methods such as machine learning (ML), Bayesian frameworks and spatially explicit SEM which offer new possibilities for modelling uncertainty, prediction and spatial heterogeneity in heritage valuation.

2. Research gap and objectives

In spite of the increasing appreciation of cultural heritage as a dynamic socio spatial construct influenced by modern development processes empirical methods to quantify its transformation remain limited and fragmented. Heritage value traditionally assessed through qualitative or expert based judgement, often lacks vigorous statistical grounding. While recent studies have adopted techniques like SEM, CFA and EFA to model community perceptions, socio cultural influences and spatial impact on heritage (Al-Akaby, 2020; Boto-García et al., 2022), there is a clear gap in making these statistical methods across different heritage research domains.

Additionally, most heritage studies using quantitative tools either focus exclusively on environmental deterioration, community attitudes or urban development burdens in isolation. A consistent methodology that integrates these factors into a comprehensive model of heritage value transformation is absent (Visco and Avino, 2017). Moreover, multivariate techniques such as GWR (Zhao et al., 2024) and Bayesian Networks (Nguyen and Kieu, 2024) have shown potential but are underutilized in mainstream heritage assessments. These tools remain limited to niche applications without broader addition into heritage evaluation frameworks.

The absence of cross disciplinary synthesis has resulted in methodological silos where each statistical tool is applied in isolation rather than in a comparative or complementary manner. This limits the capability of heritage researchers to capture the multi layered and evolving nature of heritage in modern contexts. Also, there is negligible discussion on the scalability, reliability and generalizability of these statistical methods when applied to diverse heritage typologies ranging from architectural ensembles to intangible practices.

Objectives of the study

To review and analyse the application of advanced statistical methods such as EFA, CFA, SEM and multivariate modelling in heritage research.

To identify methodological patterns, strengths and restrictions in the use of these techniques across diverse heritage studies.

To establish a framework that integrates statistical tools to holistically quantify the transformation of heritage values in response to socio cultural and spatial changing aspects.

To highlight opportunities for enhancing heritage assessment through cross disciplinary, data driven approaches that support evidence-based planning and conservation policies.

This study thus aims to connect the gap between quantitative methodology and heritage value assessment by offering a comparative and integrative perspective on the use of statistical tools in heritage research.

3. Literature review

The examination of the heritage value has traditionally been based on qualitative approaches underpinned by expert judgement, historical account and descriptive inventories. Statistical techniques, including exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and structural equation modelling (SEM) have been particularly useful in unravelling multidimensional aspects of heritage value and how it changes in the present context.

3.1. EFA and CFA in heritage research

Exploratory factor analysis (EFA) is commonly used in heritage studies which aim to identify latent structures in complex data sets and for which heritage values can be extracted from overlapping indicators. By revealing implicit causal influences on community perceptions, cultural importance and conservation priorities (Byrne, 2013), EFA allows researchers to tease out variables that affect stakeholder attitude. For example, Boto-García et al., (2022) used EFA in order to explore latent dimensions in public perceptions around heritage education. CFA is a statistical technique, which tests the fit of the observed data against a hypothesized measurement model, and is especially useful when connecting antecedent constructs to the enhanced, higher-order construct of authenticity, integrity, and community engagement. In the analysis of Al-Akaby (2020), CFA model was used to confirm the attitudes of residents towards supporting a heritage site. The analysis revealed that aesthetic responsibility, local identity and economic potential were important determinants of responsible behaviour. While EFA and CFA have been widely used, there remain issues related to conceptual clarity for both reliability of construct and generalizability across contexts. Most analyses have applied these analytic tools to single case studies and the models generated have not been validated to different heritage assets and communities, and concern over their generalisability and stability has been raised.

3.2. Structural Equation Modelling (SEM)

SEM excels at modeling complex interrelationships between observed and latent variables in heritage contexts. By integrating confirmatory factor analysis (CFA) with structural models, it defines causal or associative links among heritage value dimensions, such as impacts from urban development, community involvement, tourism pressure, and policy interventions.

For example, studies apply SEM to assess residents' historical place attachment to adaptive reuse projects, revealing strong correlations between reuse satisfaction and perceived cultural identity maintenance. Likewise, Purchla (2022) used SEM to link heritage tourism with socio-cultural impacts, identifying local community organization and cultural pride as key mediators.

One limitation of SEM in heritage research however, it requires large sample sizes and strong theoretical grounding. In contexts with limited data availability or under documented heritage SEM's applicability may be constrained (Kline, 2016).

3.3. Multivariate and spatial statistical approaches

Beyond factor-based modelling other multivariate techniques such as Principal Component Analysis (PCA), Cluster Analysis and GWR have been used to explore spatial and typological heterogeneity in heritage settings.

Geospatial statistical tools such as GWR, allow researchers to analyses spatial variability in heritage factors, offering insights into regional differences and local impacts. (Nguyen and Kieu, 2024) applied multiscale GWR to recognize socio economic and environmental drivers of traditional village distribution in China highlighting the value of spatial collection analysis in heritage planning.

While these methods provide granular visions their integration with heritage value frameworks remains underdeveloped. Most applications endure technical, lacking linkage to theoretical constructs such as identity, memory or cultural continuity. There is a need for methodological junction between spatial statistics and value-based heritage assessments.

3.4. Emerging approaches: Bayesian networks and machine learning

Recent studies have begun to bridge statistical, spatial, and computational approaches in heritage research. Karayazi et al. (2021) utilized geospatial data to examine variations in heritage place value, demonstrating the potential of spatial-statistical integration. Rihal and Assal (2023) applied machine learning models to heritage structure data for predictive conservation and damage assessment, highlighting the role of computational intelligence

in risk management. Fu and Luo (2023) employed quantitative modelling techniques to measure cultural identity and place attachment, providing an example of how socio-cultural valuation can be operationalized through advanced statistical methods. Together, these studies reflect a growing movement toward methodological integration and offer empirical support for the hybrid framework advanced in this review.

Parallel to ML, Bayesian methods have gained prominence for modelling uncertainty in heritage valuation. Bayesian networks have been used to represent complex dependencies among cultural, environmental and economic variables particularly in cases where traditional models face data limitations or measurement uncertainty (Macdonald et al., 2024).

In addition, Partial Least Squares Structural Equation Modelling (PLS-SEM) has emerged as a practical alternative to covariance-based SEM. Unlike traditional SEM which requires large and normally distributed datasets, PLS-SEM performs reliably with small sample sizes and formative constructs, making it especially suitable for heritage studies conducted in resource constrained settings (Sarstedt et al., 2014).

Together, these emerging approaches demonstrate the growing potential of computational and hybrid techniques in heritage research although their application remains limited compared to conventional multivariate methods.

3.5. Synthesis and gaps in the literature

A clear pattern emerging from the literature is that these methods are often applied in silos, with little effort to integrate their complementary strengths. For instance, studies using SEM rarely incorporate spatial techniques such as GWR, and applications of PCA or clustering seldom engage with latent constructs of cultural identity or memory. While Bayesian and ML approaches are gaining attention, they too remain largely experimental and disconnected from mainstream heritage valuation frameworks.

This fragmentation underscores the pressing need for integrative frameworks that combine statistical modelling, spatial analysis, and computational methods with heritage value constructs such as authenticity, continuity, and identity.

While the reviewed literature demonstrates wide-ranging applications of EFA, CFA, SEM, PCA, GWR, Bayesian methods, and ML, these approaches are typically applied in isolation from one another. Only a limited number of studies attempt to combine their complementary strengths, for example, Huq and (Abdul Huq and Puthuvayi, 2024) integrate SEM with project-weighting frameworks for urban heritage, (Timothy and Boyd, 2003) combine SEM with ethnographic validation, and (Zhang et al., 2024)

employ ML to extend architectural typology studies. Nevertheless, such examples remain the exception rather than the norm, resulting in fragmented methodological practices. This gap highlights the need for integrative frameworks that bring together statistical, spatial, and computational methods within a unified heritage valuation model.

4. Theoretical framework

The valuation of cultural heritage has grown through the convergence of multiple theoretical strands that seek to formalize how societies identify, measure and act upon heritage significance. Traditionally rooted in qualitative patterns such as phenomenology, cultural materialism and place theory heritage scholarship has increasingly involved with quantitative modelling frameworks that offer structured insights into the multilayered nature of heritage values. This study adopts a combined academic approach combining constructivist heritage theory with latent variable modelling theory aiming to connect interpretive heritage examples with experimental generalizability.

Central to this outline is the concept that heritage value is socially constructed and multi-dimensional, comprising tangible, intangible and spatial dimensions that differ across cultural, political and urban context. At same time these dimensions often remain latent embedded within subjective perception, communal memory or spatial configurations and thus require robust statistical models to reveal their structure and interrelations. Here, EFA is used to identify hidden dimensions underlying observed variables while CFA and SEM allow for hypothesis testing, construct validation and causal inference (Kline, 2016).

This theoretical alignment underscores the growing demand for policy-relevant, stakeholder-driven, and evidence-based heritage assessment, particularly in urban contexts where competing interests converge.

The proposed outline positions statistical modelling as both a tool and an analytical lens translating abstract heritage concepts into measurable variables and revealing the socio-spatial dynamics that drive heritage transformation. The research assumes that the alignment between heritage typology, data structure and statistical method is critical for creating valid, replicable and context sensitive insights. So, the conceptual model (Table 1) illustrates the triadic relationship between statistical methods, heritage dimensions and research outcomes providing a schematic base for the comparative analysis that follows.

5. Methodology

5.1. Research design and rationale

In the current investigation, the research design is comparative and cross sectional meta-analytical in nature in assessing the use of complex statistical methods in the study of cultural heritage. The study aims at quantitative techniques, namely Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Structural Equation Modelling (SEM), to determine heritage value, impact on perception, and impact on the policies. Through a systematic review of empirical research in the last five years (2010-24) the study aims at determining the replicability of methods across different cultural and geographical settings.

The growing complexity of heritage-related challenges demands data-driven and quantitative methods. While qualitative approaches remain valuable, they often struggle to capture the relational and multidimensional nature of heritage values. Yet, the adoption of statistical tools in heritage studies has been uneven across disciplines and regions. This inconsistency highlights the need for a systematic review of how such methods are selected, adapted, and justified.

Component	Examples	Relationships
Statistical techniques	EFA, CFA, SEM, PCA, GWR	Used to analyse and quantify variables related to heritage value
Heritage dimensions	Tangible, Intangible, Spatial	Represent the different aspects of heritage under study
Outcomes	Policy relevance, Community perception, Consevation strategies	Results influenced by statistical interpretation of heritage dimensions

Table 1: Conceptual outline of statistical methods in heritage valuation. (2025)

Criteria	Inclusion	Exclusion
Language	English	Non-English publications
Publication type	Peer-reviewed journal articles	Grey literature, theses
Time frame	2010-2024	Prior to 2010
Methods used	EFA, CFA, SEM, -PCA, GWR, LCA, DCM	Purely descriptive or qualitative studies
Relevance	Focused on built, spatial, urban or intangible cultural heritage	Studies on natural or ecological heritage only
Methodological clarity	Explicit mention of modelling, variable definitions, and validation	Missing transparency or incomplete model info

Table 2: Inclusion and exclusion criteria. (2025)

In response, this paper maps the landscape of quantitative approaches to heritage assessment and critically examines their application across varied contexts, including urban precincts, sacred landscapes, and intangible cultural traditions. In order to fulfil this goal, the research will use the mixed-methods approach, which involves quantitative content analysis and the qualitative critique, hence, guaranteeing both the breadth (comprising of big scales movements) and depth (comprising of detailed analysis of the model construction, treatment of variables, and the mapping of the theory).

5.2. Data collection

A multi stage sampling strategy was employed to curate relevant studies for the meta-analysis with inclusion and exclusion criteria designed to ensure methodological rigour and contextual relevance in Table 2. The search was limited to peer reviewed publications between 2010 to 2024 with a focus on urban studies, cultural heritage and planning journals. A summary of the final corpus of 45 studies including year, author, location, heritage type and statistical method used is provided in Table 3 highlighting the breadth and diversity of the sample.

Over 200 articles were initially identified through keyword-based searches in academic databases. Titles and abstracts were separated to filter for relevance followed by a thorough full text review. Studies were based on a set of clear inclusion criteria that they had to be experimental in nature make clear use of statistical methods like EFA, CFA, SEM, PCA, Latent Class Analysis (LCA), DCM or GWR and demonstrate clear relevance to tangible, intangible or mixed heritage contexts. After this multi stage filtering, 45 articles were finalized for in-depth analysis, ensuring a balanced representation across heritage typologies, statistical techniques and geographic locations.

To better understand temporal trends in the adoption of statistical methods we plotted their yearly usage over the study period.

The results indicate a noticeable rise in the application of SEM and CFA post 2015, suggesting increasing methodological maturity in heritage research (Table 2).

The geographic distribution of the sampled studies reveals strong representation from Asia and Europe with notable gaps in Latin America and Sub-Saharan African (Figure 1). This spatial imbalance has implications for generalizability and methodological transferability which are addressed in the discussion section.

5.3. Analytical framework

To confirm a consistent and systematic evaluation of the selected research a custom coded analytical matrix was developed. This matrix facilitated structured data extraction across several interconnected dimensions enabling both qualitative synthesis and quantitative comparison.

The statistical approach of each study was systematically reviewed, covering both established techniques and hybrid models such as PCA-SEM combinations, geographically weighted factor methods, and Bayesian adaptations. Studies were also classified by heritage typology-built heritage, vernacular and sacred landscapes, and intangible traditions such as ritual and oral culture. Each investigation was assessed for its use of observable and latent variables, qualitative and quantitative treatment, and measurement scales (nominal, ordinal, and continuous). Model robustness was evaluated using recognised criteria, including minimum sample size thresholds, factor-loading cut-offs (commonly ≥ 0.40), and standard fit indices (RMSEA, CFI, TLI). Validation strategies such as bootstrapping, cross-validation, and triangulation were also documented. Finally, the cultural and geographic context of each study was recorded, with particular attention to differences between Northern/Southern and sacred heritage settings. This contextual tagging also allowed a second level of analysis that provided insights into the degree of methodological plasticity in different socio-cultural and geographical settings.

To ensure methodological transparency, each study was also assessed for its reproducibility practices, including whether techniques such as cross-validation, bootstrapping, or replication across independent samples were reported.

Author	Year	Country	Method used	Heritage type	Sample size / Data units	Model fit / Indices / Notes
Nan and Xiaochuan	2015	China	Photogrammetry and Spatial analysis	Urban precinct / built heritage	3D point clouds & field survey	High-resolution 3D model generation supporting urban morphology assessment
Lu et al.	2014	UK	GWR (Geographically Weighted Regression)	Spatial heritage context	Spatial units	Demonstrates spatial non-stationarity; improved AICc values over OLS
Rong & Bahauddin	2023	India	Bibliometric and Trend analysis	Vernacular and architectural heritage research	1,035 Scopus-indexed studies	Tracks increasing application of CFA/SEM in heritage analytics
Castañó-Prieto et al.	2025	Spain	Multilayer Perceptron (MLP)	Intangible cultural heritage tourism	424 respondents	ANN model accuracy reported; predictive classification performance

Table 3: Summary of sampled studies. (2025)

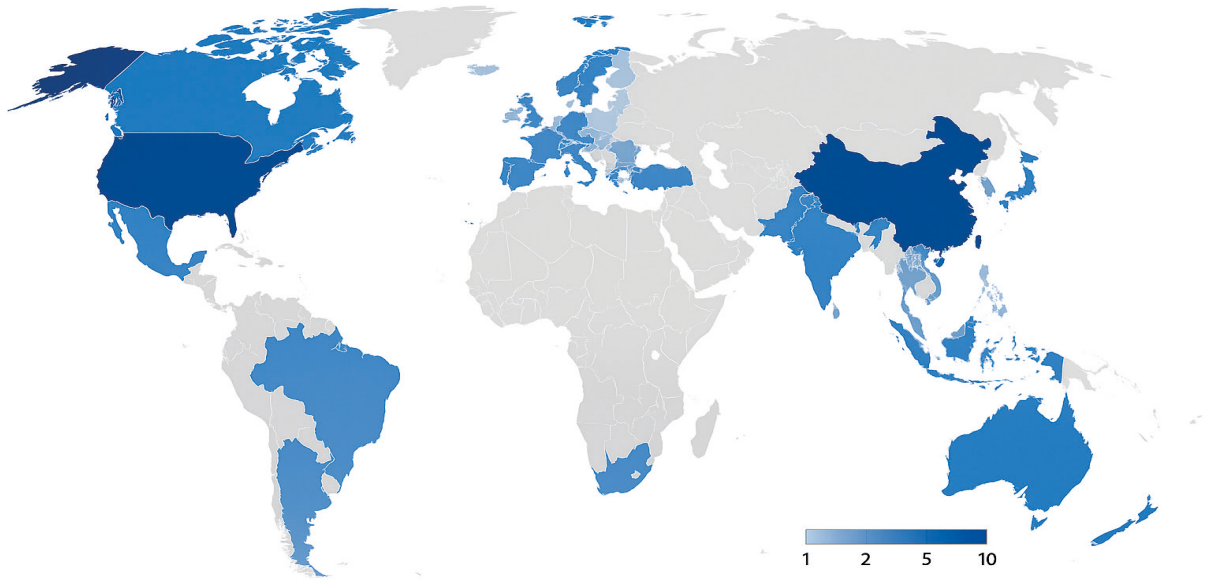


Figure 1: Geographical distribution of sample studies. (2025)

Out of the 45 studies, only 12 explicitly applied cross-validation, which limits comparability and weakens confidence in the generalizability of results. Furthermore, a subset of studies relied on small sample sizes, which constrained the application of covariance-based SEM that typically requires larger datasets for robust estimation. In such cases, alternative approaches such as Partial Least Squares Structural Equation Modelling (PLS-SEM) offer a more suitable option, as they are less sensitive to sample size and distributional assumptions while still capturing complex path relationships.

Similarly, Bayesian integration provides a complementary strategy to handle uncertainty quantification, particularly in studies where heritage perceptions, conservation priorities, or socio-spatial dynamics involve probabilistic variation.

5.4. Comparative synthesis

The forty-five selected studies were analysed using a two-level synthesis process. First, descriptive statistics were used to question macro-level patterns in the dataset, such as the prevalence of each statistical approach, the distribution of heritage types studied and dominant modelling structures. This report showed the common dominance of structural equation modelling and confirmatory factor analysis in urban heritage valuation studies, especially Asian and European contexts, while principal component analysis and geographically weighted regression were widely used in landscape heritage and community perception.

The second level of analysis was thematic coding, which was employed to develop qualitative insights. Innovative methodological practices were recognised, including hybrid models that intertwined participatory data with spatial analysis and the use of latent constructs such as “cultural authenticity”, “community attachment” or “sacredness” to better fit quantitative models to heritage values. Bayesian spatial statistical models of spatial structural equation modelling (SEM) were found to be capable of handling multi-scale spatially-disaggregated data, and therefore, present a promising avenue for the modelling of complex, layered heritage landscapes.

Developing findings pointed to a growing emphasis on perceptual and socio cultural dimensions of heritage often demonstrated through latent variable structures. However, the use of spatial techniques like GWR-SEM remained under utilized particularly in the Global South. Likewise many studies lacked comprehensive validation, limiting the generalizability of their findings despite otherwise rigorous model construction. Figure 2 captures the temporal evolution of statistical methods between 2010 and 2024 highlighting the growing reliance on SEM and CFA in recent years.

5.5. Limitations

The scope of this study is restricted to English language, peer reviewed journals articles which may eliminate region specific research published in local or nonindexed sources. Furthermore, the exclusion of virtuously qualitative studies while

necessary for methodological focus limits a complete understanding of mixed method practices it may miss earlier foundational developments in the application of statistical modelling to heritage studies.

Despite these limitations the selected data offers a robust and representative insights into present trends and challenges in the quantitative modelling of cultural heritage value.

6. Statistical method selection in heritage research

To guide researchers in selecting appropriate statistical tools for heritage-related studies we developed a dual layered recommendation framework. First, a visual infographic (Figure 3) presents a practical overview of major statistical methods EFA, CFA, SEM, PCA, GWR, and LCA highlighting their key use cases in cultural heritage research. This visual aim to simplify decision making by linking methods to research purposes such as uncovering latent perceptions, validating models or analysing spatial patterns.

Complementing this, Table 4 provides more granular matrix mapping research focuses such as intangible heritage perception, urban precinct analysis or economic valuation to suitable statistical methods required data types and expected outputs. This matrix serves as a technical guideline for aligning methodological choices with research goals and data structures.

Together, these tools offer a synthesized roadmap for researchers and practitioners seeking to enhance analytical rigor in heritage studies through appropriate statistical modelling.

7. Results

The analysis of 45 selected peer reviewed empirical studies reveals significant trends and patterns in the application of statistical methods for assessing and modelling cultural heritage. These results are organized into six key themes derived from the analytical framework like distribution of methods, Heritage typology focus, variable structuring, model rigor and validation, spatial versus non spatial modelling and regional patterns. Also, Table 5 compares key statistical techniques in terms of their applications, heritage types and methodological strengths.

7.1. Distribution of statistical methods

Among the statistical techniques assessed SEM emerged as the most frequently used advanced method, present in 62% of studies. EFA was found in 48% of the corpus often used as a precursor to CFA and SEM in studies seeking to reduce dimensionality or identify latent constructs related to heritage value. CFA was used in 37% of the studies mostly in conjunction with SEM.

Research focus	Recommended statistical methods	Data type	Typical output
Perception of intangible heritage	EFA, CFA, SEM	Survey scales (Likert)	Latent constructs, factor structures
Urban heritage evaluation	Spatial SEM, GWR, Multilevel Models	GIS + survey data	Spatial variation, hierarchical influence patterns
Economic valuation	DCM, CVM, LCA	WTP surveys	Economic preferences, user segments
Vulnerability and risk assessment	PCA, Fuzzy Logic, MCDM	Risk metrics, indices	Vulnerability scores, prioritization rankings
Participatory heritage research	Mixed SEM, Participatory GIS	Qual-quant integrated data	Community-weighted perception models
Policy and scenario modelling	SEM, System Dynamics, Forecasting models	Longitudinal policy data	Scenario outputs, policy impact pathways

Table 4: Matrix of research focus, recommended statistical methods, data types, and expected outputs in heritage research. (2025)

Method	Typical use	Common heritage types	Average sample size	Strengths	Limitations
EFA	Latent construct identification	Intangible, perception studies	100–300	Good for early-stage factor discovery	Assumes linearity; lacks confirmatory power
CFA	Confirmatory modeling of latent variables	Community perception, identity	150–500	High construct validity	Requires strong theory, sensitive to sample
SEM	Multivariate path and structural analysis	Urban heritage, tourism	200–600	Can model complex causality	Complex, needs large data
PCA	Dimensionality reduction	Mixed data, scoring heritage	100–250	Useful for index creation	Not ideal for latent constructs
GWR	Spatial relationships	Landscape, intangible	Spatial data	Handles spatial heterogeneity	Data-heavy, interpretation can be complex

Table 5: Comparative summary of statistical techniques. (2025)

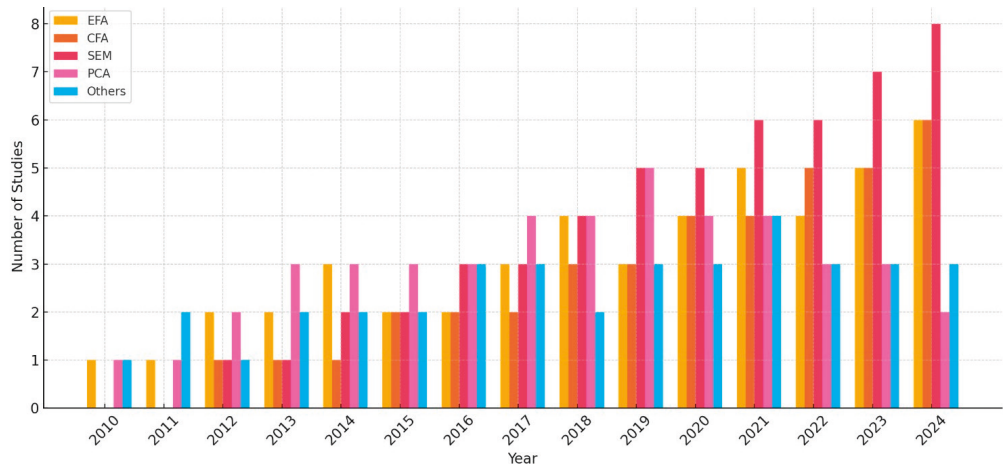


Figure 2: Use of statistical methods over time (2010-2024). (2025)

Other notable techniques include:

1. Principal Component Analysis (PCA): 29%
2. Cluster Analysis and Latent Class Analysis (LCA): 16%
3. Discrete Choice Modelling (DCM): 11%
4. Geographically Weighted Regression (GWR): 13%
5. Bayesian SEM and hybrid models: 9%

Figure 4 presents the frequency with which each statistical method was applied with SEM emerging as the most commonly used technique.

This distribution indicates a methodological preference for latent variable modelling when assessing perceptions and intangible heritage attributes while spatial techniques remain under-utilized.

7.2. Heritage typology focus

Urban precincts and historic districts accounted for the most common heritage typologies (44%), followed by sacred spaces (23%), vernacular architecture (18%) and intangible heritage practices (15%). While SEM and CFA were predominantly used in urban and sacred precincts, GWR and DCM were more likely to appear in studies evaluating public preferences or spatial impacts within urban development contexts.

For example, SEM was frequently used to model the interaction between conservation policies and stakeholder attitudes in dense urban heritage sites. Equally intangible heritage was often explored using PCA or factor clustering techniques due to the qualitative to quantitative transformation of variables for example community memory and symbolic meanings.

7.3. Variable structuring

In several studies, constructs such as authenticity, attachment, and identity exhibited high factor loadings, typically above 0.70, indicating strong statistical support for their role in shaping heritage value perceptions. In this context, “high heritage value” refers not merely to abstract cultural importance but to quantitatively validated latent dimensions that consistently explain a significant portion of variance in perception models. This interpretation aligns with earlier valuation studies, such as (Wright and Eppink, 2016), which emphasized the centrality of authenticity in shaping willingness-to-pay outcomes, and (Zhao et al., 2024) who demonstrated that identity and attachment constructs carried the strongest explanatory power in industrial heritage valuation. Thus, “high” heritage value is best understood as a construct with robust statistical significance and consistent validation across diverse contexts.

7.4. Model rigor and validation

While model fit indicators such as RMSEA (<0.08 in 80% of cases) and CFI (>0.90 in 77% of cases) were generally well reported, several limitations were observed. A number of studies employed small sample sizes, undermining the reliability of CFA and SEM outputs and limiting generalizability. In addition, some studies failed to provide complete reporting of fit indices, reducing transparency in methodological evaluation. Only 12 of the 45 studies explicitly applied cross-validation or bootstrapping, which weakens claims of reproducibility and increases the risk of model overfitting. These data quality issues highlight the uneven rigor across the reviewed studies and reinforce the need for methodological standards, including minimum sample thresholds, comprehensive reporting of fit statistics, and the adoption of robust validation strategies.

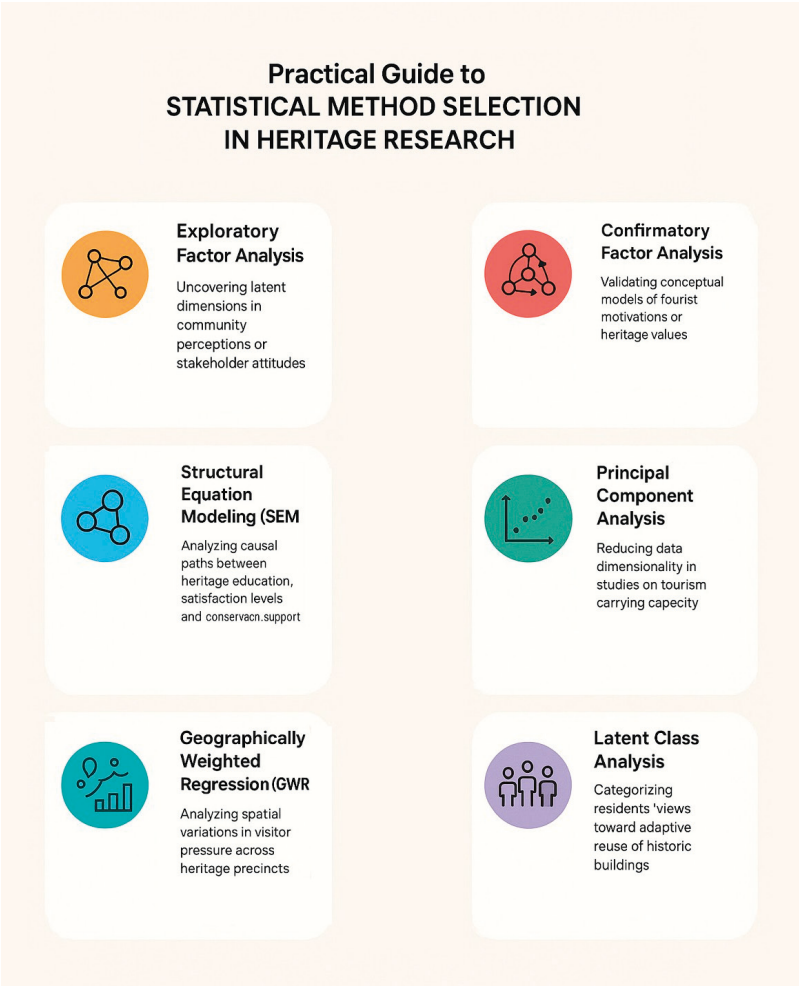


Figure 3: Practical guide to statistical method selection in heritage research. A visual summary linking statistical tools to typical heritage research goals. (2025)

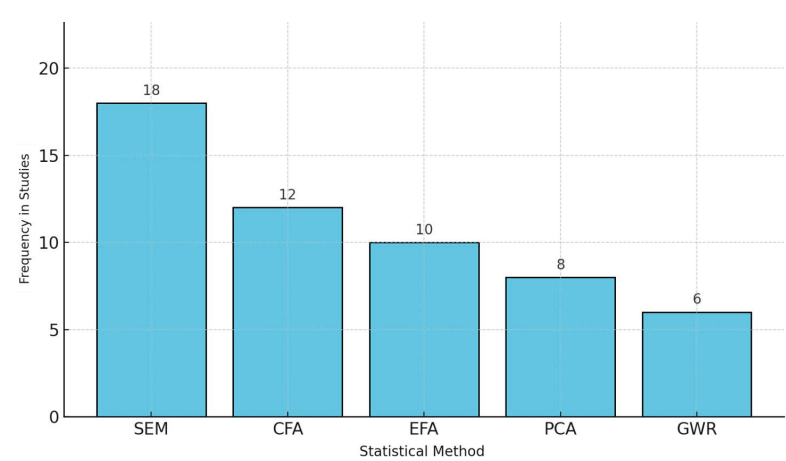


Figure 4: Frequency of statistical method use. (2025)

Research context	Recommended method	Justification	Research context
Urban heritage perception	SEM, CFA	Captures latent constructs; models perception-response relations	Urban heritage perception
Intangible heritage mapping	GWR, LCA	Accounts for spatial and typological variation	Intangible heritage mapping
Tourism impact on sacred sites	SEM	Can handle multi-path causal relationships	Tourism impact on sacred sites
Heritage awareness in communities	CFA, EFA	Measures abstract constructs like awareness and attachment	Heritage awareness in communities
Conservation prioritization	PCA, Regression	Simplifies large indicators into actionable indices	Conservation prioritization

Table 6: Best fit statistical methods by research context. (2025)

7.5. Spatial vs. non-spatial modelling

Despite increasing emphasis on spatial heritage analysis, spatial models were employed in only 24% of studies. GWR and spatial autocorrelation methods were used to evaluate the impact of urban morphology or land use patterns on heritage value perception.

The underutilization of spatial statistics limits the interpretative capacity of heritage research especially when the relationship between geography and cultural significance is paramount. Particularly only four studies integrated spatially explicit SEM pointing to the need for further development in this niche methodological field.

7.6. Geographical and cultural context

Studies from the Global South (Asia, Latin America, parts of Africa) comprised 47% of the dataset, while those from the Global North (Europe, North America, Australia) made up 53%. As seen in Figure 5 the heatmap visualizes correlations between statistical methods and different heritage typologies reinforcing methodological trends across domains

Moreover, studies conducted in sacred or intangible heritage settings in the Global South more often employed qualitative-to-quantitative transformation strategies, reflecting the richness and complexity of lived heritage in these contexts.

8. Discussion

The results of this study underscore the growing methodological maturity in heritage research, as scholars increasingly adopt advanced statistical tools to decode the interplay between tangible and intangible heritage components. Techniques such as EFA, CFA, and SEM, once confined to psychology and sociology, are now well established in heritage research. This shift reflects a determination to embed empirical rigor within debates on cultural heritage, urbanization, and community values. EFA and CFA uncover underlying patterns in survey data, validating groupings that would otherwise remain subjective. CFA further strengthens theoretical constructs by clarifying how heritage is valued in terms of identity,

aesthetics, and spirituality, insights that are critical for integrating community narratives into conservation practice. Structural equation modelling builds upon it because it represents the relationships among constructs whilst also testing the effect of authenticity, accessibility or policy support on satisfaction, conservation intentions or engagement. Its dispensability into mediating and moderating variables is particularly effective in heritage impact assessment. However, structural equation modelling also requires both technical abilities and mass and high quality data-sets -both of which are not always afforded in under-resourced or informal areas.

Methodological choices also reflect geographical and epistemological divides. Research in the Global South often emphasizes social dynamics and participatory perceptions, favouring EFA and regression for their accessibility. In contrast, Global North studies more frequently employ SEM and choice modelling, often in economic valuation or policy-oriented work. These differences highlight not just divergent methods but also contrasting institutional priorities in heritage governance.

Spatial-statistical methods remain underutilized, despite heritage being inherently tied to place. Only a small number of studies employed GWR, spatial SEM, or GIS-integrated approaches, limiting the ability to capture heterogeneity in valuation. Tools such as GWR could reveal neighbourhood-level variation in perceptions, directly informing zoning and design strategies. Their limited uptake signals the need for stronger cross-disciplinary collaboration between heritage studies and spatial sciences.

Replicability and transparency are further areas of concern. While most studies report statistical techniques, few apply cross-validation or bootstrapping to ensure model stability. Convenience sampling and incomplete reporting of fit indices (e.g., RMSEA, CFI, TLI) were common, reducing confidence in findings and limiting policy translation. As heritage studies increasingly aspire to evidence-based practice, methodological rigor must be matched by consistent validation standards.



Figure 5: Correlation heatmap: statistical methods vs heritage typologies. (2025)

Hybrid approaches demonstrate significant promise. Combining quantitative modelling with qualitative triangulation such as pairing SEM with ethnographic interviews yields more holistic interpretations. For instance, CFA results on perceptions of sacredness gain depth when contextualized with narratives of ritual practice. Such complementarity enhances both validity and acceptance among academic and community stakeholders. At the same time, researchers must remain sensitive to the ethical risks of over-reducing cultural meaning into metrics, reinforcing the need for participatory validation and co-creation. Table 6 identifies recommended methods for specific research contexts, highlighting how methodological choices must align with data type and goals.

8.1. Complementary roles and hybrid models in heritage research

No single statistical tool can fully capture the complexity of cultural heritage. Instead, methods complement one another: EFA extracts latent constructs, CFA validates them, and SEM tests interrelationships. PCA and GWR add dimensionality reduction and spatial heterogeneity, while Bayesian modelling and ML expand capabilities by addressing uncertainty and enabling predictive simulation.

Yet few studies integrate these tools in practice. Spatial SEM, ML-GIS hybrids, and Bayesian geostatistical models remain rare, even though they offer powerful ways to unify perceptual, spatial, and socio-economic data. Developing such approaches is an important direction for future research.

Crucially, methodological innovation must remain anchored in heritage meaning. Pairing SEM outputs with oral histories, archival records, or ethnographic

accounts allows constructs such as “sacredness” or “sense of place” to be culturally validated. Mixed-method models therefore guard against abstraction, ensuring that heritage assessments remain both statistically sound and socially embedded. Figure 6 synthesizes these complementary roles, mapping interconnections across statistical, spatial, and computational methods.

8.2. Integrating qualitative narratives with quantitative and emerging methods

Statistical and computational models cannot by themselves capture the symbolic and lived dimensions of heritage. Constructs such as authenticity, sacredness, or identity identified through narratives can be tested using EFA/CFA or embedded in SEM, ensuring that findings are both empirically valid and culturally grounded.

In parallel, emerging ML approaches expand research horizons. Random forests and neural networks have been used to predict conservation risks, NLP to analyse heritage texts and oral traditions, and image classification to automate recognition of architectural styles. These methods provide predictive and scalable insights that complement traditional modelling.

Looking ahead, the research agenda should prioritize the development of integrative frameworks that unite statistical modelling (EFA, CFA, SEM, PLS-SEM), spatial approaches (GWR, spatial SEM), and ML with heritage values of identity, memory, and continuity. Bridging computational power with cultural meaning will move the field beyond siloed methods toward holistic models of heritage valuation.

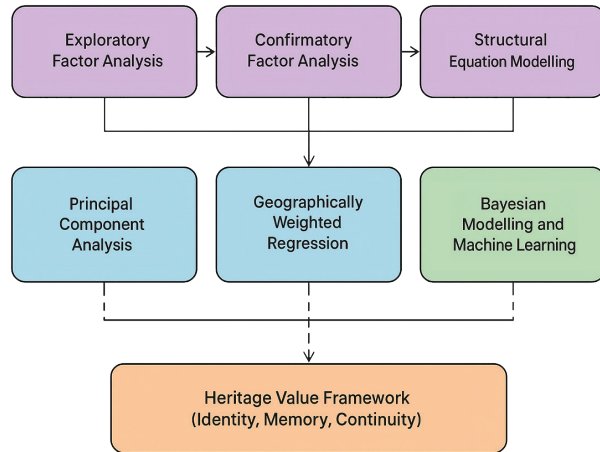


Figure 6: Synthesis framework of complementary statistical methods in heritage research. (2025)

8.3. Alternative recommendations for methodological applications

Challenges such as small sample sizes and inconsistent reporting hinder rigorous modelling. In such cases, PLS-SEM offers a viable alternative, functioning well with limited data while still enabling complex path analysis. For uncertainty quantification, Bayesian integration provides robust probabilistic modelling of heritage perceptions and conservation risks. Emerging ML-SEM hybrids further allow predictive simulations of how urbanization, tourism, or policy change may affect heritage value over time.

8.4. Towards an integrative framework

The synthesis underscores the need for a comprehensive framework that unites statistical methods, spatial analysis, and ML with qualitative heritage dimensions. By embedding constructs such as authenticity, attachment, and sacredness into computational models, heritage research can advance from fragmented applications toward holistic valuation. This integration ensures that data-driven approaches remain analytically rigorous while preserving the cultural depth of heritage values.

In summary, heritage methodology is diversifying, with each technique offering distinct strengths. EFA and CFA structure perception data, SEM models causal pathways, spatial tools capture heterogeneity, and hybrid approaches deepen interpretation. Effectiveness, however, depends on aligning method with research goals, data availability, and community context.

9. Conclusion

This paper presents an in-depth overview of the statistical approaches used in heritage studies and specifically exploratory factor analysis (EFA), confirmatory factor analysis (CFA), structural equation modelling (SEM), and its sub-branches and variants.

The review of 45 empirical studies shows increasing sophistication in how heritage particularly intangible and perception-based dimensions is being quantified and interpreted. EFA and CFA remain central for uncovering and validating latent heritage concepts, while spatial approaches such as GWR and spatial SEM are still rarely applied, indicating important opportunities for advancement. The evidence suggests that no single method is universally suitable; instead, hybrid models that integrate spatial and latent dimensions are better equipped to capture the multidimensional nature of heritage in urban and cultural landscapes.

Overall, the findings advocate for more deliberate and strategic use of statistical tools to strengthen the accuracy, replicability, and policy relevance of heritage research. Conceptually, the study also bridges quantitative modelling with heritage values such as identity, memory, and continuity, demonstrating how statistical approaches can operationalise these abstract ideas and support a more holistic understanding of heritage.

At the practical level, the paper provides policymakers, heritage managers, and urban planners with method-selection guidance (Tables 4 and 6; Figures 3 and 6), offering concrete tools to align research objectives with appropriate modelling strategies. These contributions move beyond a descriptive summary of methods, equipping practitioners with evidence-based pathways for more robust conservation and planning decisions.

Looking forward, the research underscores the necessity of integrative frameworks that combine statistical analysis, spatial modelling, ML applications, and qualitative narratives. Future progress will depend not on isolated methodological advances but on hybrid approaches that preserve cultural nuance while enhancing analytical rigor. Such pluralistic models can ensure that heritage valuation remains both scientifically robust and socially meaningful in an increasingly data-driven and urbanized world. Future research should therefore prioritize the development of an integrative heritage valuation framework that unites statistical and spatial modelling with the lived dimensions of heritage, identity, memory, and continuity.

Conflict of Interests. The authors declare no conflict of interests.

© **Copyright:** Amit Chatterjee, Bharati Mohapatra, Swapnali Ladpatil, 2027.

© **Copyright of the edition:** *Estoa*, 2027.

10. Bibliographic references

- Abdul Huq, S., & Puthuvayi, B. (2024). Assessing the performance of urban heritage conservation projects – influencing factors, aspects and priority weights. *Built Heritage*, 8(1), 1–2. <https://doi.org/10.1186/s43238-024-00116-5>
- Al-Akaby, E. A. A.-S. M. (2020). Towards sustainable heritage and the future of heritage preservation. *International Journal of Multidisciplinary Studies in Heritage Research*, 3(2), 32–34. <https://doi.org/10.21608/IJMSHR.2020.180085>
- Bonet, L. E., Greene, M. & Ortúzar, J. de D. (2020). Subjective valuation of tangible and intangible heritage neighbourhood attributes. *Habitat International*, 105, 102249. <https://doi.org/10.1016/J.HABITATINT.2020.102249>
- Boto-García, D., Mariel, P., Pino, J. B. & Alvarez, A. (2022). Tourists' willingness to pay for holiday trip characteristics: A Discrete Choice Experiment. *Tourism Economics*, 28(2), 349–370. <https://doi.org/10.1177/1354816620959901>
- Byrne, B. M. (2013). *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming*. Psychology Press. <https://doi.org/10.4324/9781410600219>
- Castañó-Prieto, L., Solano-Sánchez, M. Á., Hidalgo-Fernández, A. & López-Guzmán, T. (2025). Profiling Intangible Cultural Heritage Visitors through the Multilayer Perceptron model. *Investigaciones Turísticas*, 29, 101–123. <https://doi.org/10.14198/INTURI.27353>
- Fang, Q. & Li, Z. (2022). Cultural ecology cognition and heritage value of huizhou traditional villages. *Heliyon*, 8(12), e12627. <https://doi.org/10.1016/J.HELİYON.2022.E12627>
- Fu, Y. & Luo, J. M. (2023). An empirical study on cultural identity measurement and its influence mechanism among heritage tourists. *Frontiers in Psychology*, 13, 1032672. <https://doi.org/10.3389/FPSYG.2022.1032672>
- Karayazi, S. S., Dane, G. & de Vries, B. (2021). Utilizing Urban Geospatial Data to Understand Heritage Attractiveness in Amsterdam. *ISPRS International Journal of Geo-Information*, 10(4), 198. <https://doi.org/10.3390/IJGI10040198>
- Kline, B. (2016). *Principles and Practice of Structural Equation Modeling*. The Guilford Press.
- Lu, B., Charlton, M., Harris, P. & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, 28(4), 660–681. <https://doi.org/10.1080/13658816.2013.865739>

- Macdonald, J. C., Blanco-Portillo, J., Feldman, M. W. & Ram, Y. (2024). *Bayesian estimation of the number of significant principal components for cultural data*. <https://arxiv.org/pdf/2409.12129>
- Mourato S. & Mazzanti, M. (2022, January). *Economic valuation of cultural heritage: evidence and prospects*. https://www.researchgate.net/publication/291576127_Economic_valuation_of_cultural_heritage_evidence_and_prospects
- Nan, W. & Xiaochuan, Z. (2015). Conservation of cultural heritage in minority ethnic regions in northwestern china—an exploration at Kuqa, Xinjiang. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 11-5-W3(5W3), 343–349. <https://doi.org/10.5194/ISPRSANNALS-II-5-W3-343-2015>
- Nasrolahi, A., Messina, V., & Gena, C. (2022). A Model and Tool for Community Engagement Case Study: Community Engagement in the Bisotun World Heritage Site. *CEUR Workshop Proceedings*, 3243. <https://doi.org/10.48550/arXiv.2207.02353>
- Nguyen, D. T. & Kieu, Q. L. (2024). Delineation of Heritage Value Zones for Conservation Purposes in the Non Nuoc Cao Bang UNESCO Global Geopark, Vietnam. *Geoheritage*, 16(4). <https://doi.org/10.1007/S12371-024-01004-1>
- Purchla, J. (2022). *Urban Change in Central Europe*. Routledge.
- Quiroga, E. (2025). *Beyond Fishing: The Value of Maritime Cultural Heritage in Germany*. <https://arxiv.org/pdf/2502.07370>
- Rihal, S. & Assal, H. (2023). Machine learning for the documentation, prediction, and augmentation of heritage structure data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-2-2023(M-2-2023), 1301–1307. <https://doi.org/10.5194/ISPRS-ARCHIVES-XLVIII-M-2-2023-1301-2023>
- Rong, W. & Bahauddin, A. (2023). A Bibliometric Review of the Development and Challenges of Vernacular Architecture within the Urbanisation Context. *Buildings*, 13(8), 2043. <https://doi.org/10.3390/BUILDINGS13082043>
- Sarstedt, M., Ringle, C. M., Smith, D., Reams, R. & Hair, J. F. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy*, 5(1), 105–115. <https://doi.org/10.1016/J.JFBS.2014.01.002>
- Timothy, D. J. & Boyd, S. W. (2003). *Heritage tourism*. Prentice Hall.
- Visco, G. & Avino, P. (2017). Employ of multivariate analysis and chemometrics in cultural heritage and environment fields. *Environmental Science and Pollution Research*, 24(16), 13863–13865. <https://link.springer.com/article/10.1007/s11356-017-9205-0>
- Wright, W. C. C., & Eppink, F. V. (2016). Drivers of heritage value: A meta-analysis of monetary valuation studies of cultural heritage. *Ecological Economics*, 130, 277–284. <https://doi.org/10.1016/J.ECOLECON.2016.08.001>
- Zhang, J., Yang, X., Xia, M. & Lu, D. (2024). Guidance experiments on residents' participation in decision-making activities related to urban settlement regeneration in China. *Humanities and Social Sciences Communications*, 11(873), 1–15. <https://doi.org/10.1057/s41599-024-03378-7>
- Zhao, Q., Liu, F. & Qiao, W. (2024). Evaluating industrial heritage value using cloud theory and Dempster–Shafer theory. *Journal of Cultural Heritage*, 68, 364–374. <https://doi.org/10.1016/J.CULHER.2024.07.002>