UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A GENERIC FRAMEWORK FOR DELINEATING THE BASIC UNITS OF SOCIAL-ENVIRONMENTAL SYSTEMS: ENSURING USER CONTROL AND REPRODUCIBILITY

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

MOHAMMAD SHAHRIYAR PARVEZ

Norman, Oklahoma

2024

A GENERIC FRAMEWORK FOR DELINEATING THE BASIC UNITS OF SOCIAL-ENVIRONMENTAL SYSTEMS: ENSURING USER CONTROL AND REPRODUCIBILITY

A THESIS APPROVED FOR THE

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY THE COMMITTEE CONSISTING OF

Dr. Xin (Selena) Feng, Chair

Dr. Jennifer Koch

Dr. Thomas M. Neeson

DEDICATION

To the scars in heart,

all the unexpected experiences,

and the paths that took new directions.

To the friends and family, I discovered on my journey.

"O you who have believed, persevere and endure and remain stationed and fear Allah that you may be successful."

— *Surah Al-Imran 3: Ayah 200*

Acknowledgement

First and foremost, I express my deepest gratitude to Allah, the Almighty, for the blessings throughout my research journey enabling me to successfully complete this work.

I am profoundly indebted to my research supervisor, Dr. Xin (Selena) Feng, Assistant Professor in the Department of Geography and Environmental Sustainability at the University of Oklahoma, for her unwavering support and invaluable guidance. Her expertise, vision, and dedication have been instrumental in shaping this research. The opportunity to work under her mentorship has been both a privilege and an honor, for which I am extremely grateful.

I extend my heartfelt appreciation to my parents for their unwavering love, prayers, and sacrifices, which have been the foundation of my academic and personal growth. I also wish to acknowledge the support and encouragement of my friends and colleagues: Ahmed Shatil Alam, Mohammad Shadman Rahman Anjon, S.M. Al Imran Hossain, Himel Talukder, Sree Sujon Sutradhar, Md. Kamal Uddin, Abdullah Al Tasim, Sagor Ghos, Md. Nurul Islam,Wahida Alam, and Mansura Khatun Roly. Your support has been invaluable.

A special acknowledgment is due to an unnamed individual whose impact has significantly contributed to my resilience and strength, encapsulated by the adage, "What does not kill you makes you stronger."

Finally, I extend my sincere thanks to all those who have supported me, both directly and indirectly, in the completion of this research.

Abstract

This study investigates the dynamic interplay between environmental services (ES) and human systems across multiple spatial scales, examining the supply of ES by natural ecosystems, the impacts of human actions mediated by socio-political institutions, and their effects on environmental and social subsystems. Emphasizing the pivotal role of these interactions within landscape planning, the research highlights the absence of explicit boundary mapping for fundamental Social-Environmental System (SES) units. To address this gap, a unified, structured framework is introduced, integrating Geographic Information Systems (GIS), dimension reduction, and regionalization techniques to effectively delineate and characterize socio-environmental units. This framework uniquely combines raster and vector data across various scales and dimensions, utilizing spatial optimization techniques to control the spatial properties of the resulting SES units. Advanced dimension reduction algorithms are incorporated to accommodate the non-linear characteristics of SES, enhancing the precision of the delineation process.

Utilizing the socio-environmental geodatabase of the Rio Grande/Bravo basin, the research demonstrates the practical application of the framework. This basin, encompassing diverse cultures, ecosystems, and economies, serves as an ideal case study for testing the methodology. The delineation process considers various factors, including administrative boundaries, estimated total quantities, compactness, spatial contiguity, and similarity in socio-environmental characteristics. A key objective is to enhance the accessibility, reproducibility, and scalability of the methodology by employing open-source Python packages. Addressing computational demands, the study employs the Uniform Manifold Approximation and Projection (UMAP) algorithm for dimension reduction, facilitating efficient processing.

This methodological framework advances the understanding of interactions between environmental and socio-economic subsystems, promoting sustainable resource governance. The proposed framework supports sustainable landscape planning and resource management through robust regionalization and interdisciplinary synthesis, making it transferable to other research contexts using diverse data formats and spatial scales.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This chapter serves as the introductory segment of the thesis, providing a foundational overview of the Social-Environmental System (SES) and its critical role within the fields of environmental and social sciences. It delves into the complex interactions between human societies and their environmental contexts, which have positioned SES as a pivotal area of global scholarly discussions. This exploration includes examining the dynamics of these interactions and the challenges and methodologies involved in delineating the basic units of SES. Key concepts such as adaptability, resilience, and transformability are discussed to highlight the importance of understanding and managing these systems effectively. The chapter also addresses the significance of regionalization in capturing local interactions and concerns, which aids policymakers in implementing effective strategies and managing natural resources efficiently. Through this comprehensive introduction, the thesis sets the stage for a deeper investigation into the nuances of SES and the innovative approaches employed to sustain and enhance these integral systems.

1.1 Social-Environmental System (SES) and Dynamics of its Interactions

The concept of the SES has emerged as a significant area of inquiry within the environmental and social sciences. This framework, which articulates the intricate interactions between human societies and their environmental contexts, has become a central focus of global academic discourse (Berkes & Folke, 1998; Liu et al., 2007; Martín-López et al., 2017; Ostrom, 2009; Yang et al., 2023). Introduced by Berkes and Folke in 1998, the SES framework has undergone substantial evolution, reflecting theoretical and methodological advancements (Colding & Barthel, 2019; Herrero-Jáuregui et al., 2018). It is now viewed as a dynamic and adaptable construct that integrates diverse environmental elements—ranging from biophysical and ecological to geomorphological and climatic—with the structures of human societies (Martín-López et al., 2017). This holistic approach extends the role of the environment beyond a mere service provider, highlighting the reciprocal and bidirectional interactions between human activities and both environmental and social spheres (Berkes & Folke, 1998; Díaz et al., 2018).

The sustainability of SES is deeply influenced by human activities, manifesting in its spatio-temporal complexities (Price et al., 2011; Wilkinson, 2005). This complexity prompts a focused effort by the research community to decode the intricate interplay and feedback mechanisms within SES, highlighting its critical significance (Griggs et al., 2013; Liu et al., 2007; Verburg et al., 2016). Recognizing that neither environmental nor socio-economic factors alone can fully explain SES dynamics, scholars advocate for a comprehensive approach to grasp its subtleties fully (Ostrom, 2009).

In response to these complexities, researchers have developed and applied a variety of innovative methodologies aimed at elucidating the intricate dynamics of SES. These methodologies include mapping SES through anthropogenic biomes (Ellis & Ramankutty, 2008), identifying land system archetypes (Václavík et al., 2013), delineating ecoregions (Castellarini et al., 2014). Additionally, studies have explored bundles of ecosystem service use (Hamann et al., 2015), and integrated both social and environmental components to characterize the evolving dynamics of social-ecological systems. These diverse methodologies not only deepen our understanding of SES but also enhance our capacity to manage and sustain these systems effectively.

1.2 SES in the Lens of Regionalization

The challenges of sustainable development present intricate social-environmental dynamics that defy solitary disciplinary or subsystem solutions (Liu et al., 2007; Wu, 1991; Yalcin, 2017). Understanding the nexus between environmental and social subsystems, crucial for addressing issues like global climate change and ecological footprints, underscores the need for integrated approaches (Cetin, 2020; Cetin et al., 2021; Osman & Sevinc, 2019). Likewise, domains such as landscape, tourism, and urban planning demand comprehensive decision-making frameworks that merge environmental and social perspectives to tackle multifaceted challenges effectively (Cetin, 2016; Kilicoglu et al., 2021). To navigate these complex social-environmental issues, a unified framework is imperative (Cetin, 2015; Wang et al., 2011; Xue-Sen & Jing-Yuan, 1990). The SES research framework provides such a holistic approach, facilitating deeper insights into the dynamic interactions between human and natural systems (Cetin, 2015; Cumming et al., 2006; Wang et al., 2011), thereby enhancing our capacity to manage these interconnections (Aminpour et al., 2020; Kilicoglu et al., 2020; Li et al., 2018).

The concept of a SES unit presents a distinct classification, setting it apart from other commonly used categories such as 'land-use' (Gregorio, 2005), 'service providing unit' (Andersson et al., 2015; Luck et al., 2009), 'social-ecological hotspots' (Alessa et al., 2008), and 'social-environmental patches' (Sinare et al., 2016). Studying SES dynamics often involves looking at nested scales, particularly focusing on regional levels capable of capturing local interactions and concerns. This regional focus is crucial for enabling policymakers to enact impactful measures (Brunckhorst et al., 2006; Fischer et al., 2015; Wu, 2013). Organizing large spatial data sets into smaller, more manageable, and geographically cohesive groups is a key strategy in this process (AssunÇão et al., 2006). Regions can be delineated by grouping areas with similar socio-economic, cultural characteristics, and types of land use. This process, known as regionalization, aims to define socio-economic units that exhibit consistency in their socio-economic and cultural traits as well as their land use patterns (Castillo-Eguskitza et al., 2018). Regionalization methods support this by consolidating numerous smaller areas into fewer larger regions based on criteria such as internal homogeneity, attribute equality, and geographical continuity (Shortt, 2009; Wei et al., 2021). These methods are particularly effective in managing large data sets by minimizing fine-scale variations and maintaining important patterns, thereby improving the utility of the data for discovering spatial patterns (Alvanides & Openshaw, 1999).

Regionalization offers flexible, multipurpose spatial framework suitable for a variety of applications including inventory management, assessment, monitoring, and management, fostering the integration of ideas across consistent geographic units. The primary goal of most regionalization methods, from a mathematical point of view, is to solve a constrained optimization problem. These methods strive to enhance the uniformity within regions and increase the diversity between them, while adhering to spatial constraints like contiguity and compactness (Duque et al., 2011; Feng & Koch, 2024; Li et al., 2014; Wei et al., 2021). This approach enhances analytical and management consistency and has proven essential in optimizing resource allocation across sectors such as healthcare, eco-efficiency, and sustainable economic development (Li et al., 2023; Ramos et al., 2020; Song et al., 2023). Moreover, regionalization can be scaled to accommodate environmental phenomena. Delineating ecoregions, for instance, provide a relevant geographic framework for ecosystem management by representing ecosystem patterns across different scales and offering a mechanism for extrapolating site-specific knowledge about ecosystem behavior to broader areas (Omernik & Bailey, 1997).

Chapter 2: Literature Review

This chapter provides an in-depth literature review on SES, focusing on the delineation and mapping of SES boundaries. It examines the integration of environmental and social data within traditional and advanced methodologies, such as machine learning techniques. The chapter critiques these traditional approaches and proposes a new integrative framework that utilizes spatial optimization and dimension reduction algorithms.

2.1 Background and Research Gaps

The SES consists of two primary subsystems: the environmental system (ES) and the social/human system (SS). Efforts to model the ES are prioritized to support the conservation of ES services, leveraging global-scale satellite data to map the spatio-temporal dynamics of these services (Di Minin et al., 2017; Duarte et al., 2016; Egoh et al., 2011; Izquierdo & Clark, 2012; Liu et al., 2013; Naidoo et al., 2008; Qu & Lu, 2018; Reid et al., 2005). Significant resources have been invested in employing remotely sensed data to effectively capture these dynamics (Choudhary et al., 2018; Junge et al., 2010; Martínez-Harms et al., 2016; Nizeyimana, 2012), underscoring that the benefits of ecosystem services extend beyond local administrative boundaries and often have transregional impacts across both time and space (Pascual et al., 2017).

Conversely, vector data are utilized to develop detailed models that interpret human behavior by aggregating and categorizing social and economic functions at various scales (Niu & Silva Elisabete, 2020). However, data concerning the social system are often outlined at different administrative levels, such as county, state, or national. Information related to human subsystems is not consistently available in vector formats; for instance, population density is commonly represented as raster data, whereas environmental data might be captured in vector formats, such as droughts or soil conditions (Elsawah et al., 2020).

The delineation of basic unit boundaries is crucial in defining SES. Given the integrative nature of SES, analysis and modeling typically involve a diverse array of data, influenced by various factors including spatiotemporal environmental data (e.g., land use/cover, elevation, and climate) and components of the social system (e.g., population, household) (Feng & Koch, 2024). Initial efforts in this area often involved comparing human-perceived values with measurable environmental factors (Alessa et al., 2008). Traditional SES delineation frameworks have relied heavily on subjective human judgment or complex, lasagna-style spatial superposition analysis. These methods, however, often fall short in effectively managing the intricate interrelations among social and environmental services, frequently resulting in suboptimal resource allocation. Ideally, frameworks for delineating unit boundaries should strive to minimize trade-offs while maximizing the overall functionality of the SES as much as possible (Song et al., 2023).

Numerous scholars worldwide have proposed various methodologies for identifying the basic units of SES. These methodologies include delineating areas where human-perceived landscape values coincide with physically measured ecological values (Alessa et al., 2008), defining social-ecological patches that correspond to landscape units referred to by local terminology (Sinare et al., 2016), and distinguishing social-ecological units that categorize different village types by their unique species diversity patterns (Hanspach et al., 2016). SES can also be analyzed through case studies that illustrate the impact of regional attributes on local dynamics, emphasizing the need for customized management strategies. These studies identify enduring characteristics such as aridity, topography, and a unique political economy of land that consistently define the SES region (Jones et al., 2019). An alternative method for analyzing SES involves examining the social perception and valuation of a broad array of ecosystem services and disservices in peri-urban communal forests, differentiating between the viewpoints of landowners and various types of visitors. It was proposed that a socio-cultural approach incorporating semi-quantitative surveys be integrated into a basic Public Participation Geographic Information System. Communal forests were conceptualized as social-ecological units (SEU) and recognized as complex adaptive systems where social and biophysical components interact at a local scale (Rodríguez-Morales et al., 2020). Advancements in the environmental and social sciences have also propelled the delineation of the SES framework, emphasizing attributes like adaptability, resilience, and transformability (Leslie et al., 2015; Ostrom, 2009; Walker et al., 2004).

The frameworks proposed for delineating SES generally adopt a systematic approach to data pre-processing and post-processing. These methods initially treat the two subsystems (ES and SS) independently, categorizing them as distinct regions. Subsequently, SES is defined by overlapping the regions of ES and SS. For the environmental subsystem, predefined environmental districts or hydro-systems are utilized to identify and delineate ES boundaries (Castro et al., 2014; Klijn & de Haes, 1994; Martín-López et al., 2011). In contrast, the delineation of social system regions embraces a variety of scholarly approaches. Multivariate analysis (Zhang et al., 2011), combined qualitative and quantitative analysis (Kušová et al., 2008), and the use of semi-structured social surveys (Alessa et al., 2008; Brunckhorst et al., 2006) are prevalent. Additionally, statistical tests such as ANOVA and Kruskal-Wallis, followed by post hoc analyses like Bonferroni and Dunn’s pairwise comparisons, are employed to identify socio-economic units (Martín-López et al., 2017).

In addressing the multifaceted nature of SES, a variety of machine learning techniques—including Principal Component Analysis (PCA) (Martín-López et al., 2017; Rattan & Hsieh, 2005; Zhang et al., 2011), Random Forest (RF) (Ellis et al., 2012; Grossmann et al., 2010), and K-means (Deng & Cao, 2023; Ropero et al., 2021; Yang et al., 2021)—are employed. These methods facilitate the quantification of similarities among different SES and the categorization of geographic units that exhibit high levels of similarity into appropriate clusters (Yang et al., 2021). Such techniques have been broadly recognized as effective in managing multiple SES concurrently. Nevertheless, numerous studies have demonstrated that the spatial distribution and temporal dynamics of SES are influenced by a multitude of factors, rendering SES inherently nonlinear with data distributions that are typically non-spherical (Li et al., 2022; Mandal & Pal, 2022). Techniques like K-means, PCA, and RF may not be optimally suited for the efficient classification of SES. In contrast, Self-Organizing Maps (SOM) leverage the adaptive properties of neural networks to identify nonlinear, non-spherical clustering structures, offering superior capabilities in spatial pattern recognition (Kim et al., 2023; Song et al., 2023). However, delineation of the basic unit boundaries based on SOM often encounters issues with heterogeneity and fragmentation of clustering units (Gao et al., 2014; Song et al., 2023), which complicates the establishment of precise boundary definitions. Traditional approaches relying on watersheds or administrative districts often fall short in effectively guiding the development and enhancement of SES within regions due to significant heterogeneity in geographic and anthropogenic features (Edwards et al., 2010; Prato, 2009; Song et al., 2023), and the challenges in maintaining consistency in SES interrelationships across dividing areas (Peng et al., 2019; Song et al., 2023). Moreover, the integration of expert knowledge in delineation of the basic unit boundaries can introduce a high degree of subjectivity. In contrast, Support Vector Machine (SVM) employs linear classifiers to delineate hyperplanes that meet classification criteria effectively, emerging over time as a robust method due to its generalization capabilities, optimal solution finding, and discriminative abilities (Cervantes et al., 2020).

Current research on SES structures identifies six primary types: "balanced," "collaborative," "hierarchical," "open space," "comprehensive interactive," and "point-axis network" and each type places emphasis on different characteristics. Notably, these structures are recognized for their cyclical mutual feedback mechanisms, yet the analysis of these cycles and the interactions among system elements remains incomplete (An et al., 2014; Barnes et al., 2019). This reveals a significant scientific challenge: to determine which structural system effectively integrates these cyclical feedbacks between and within elements to maintain ongoing cyclical processes, a question that warrants deeper investigation (Arnaiz-Schmitz et al., 2018; Morzillo et al., 2014). Again, despite the development of these methods, issues remain, especially in extending local insights to address broader socio-economic and environmental changes, thereby enriching the global knowledge base (Thorn et al., 2021). Precisely identifying and mapping SES boundaries remains critical for effective natural resource management and policy implementation, posing ongoing challenges for researchers.

While many researchers have proposed general frameworks for delineating SES boundaries that integrate environmental conditions, socioeconomic indicators, and land-use patterns (Kumar et al., 2021), as well as biodiversity considerations (Lazzari et al., 2019) or separate treatments of ES and SS (Martín-López et al., 2017), there is limited research integrating both vector and raster data on a single platform for preprocessing and boundary delineation, with only one known exception (Feng & Koch, 2024). Furthermore, there is a noticeable absence of evidence supporting the reproducibility of these frameworks across different regions. Additionally, there is a lack of user control over the selection of regions or the ability to adjust the dimensions of data to influence the weighting of factors considered in delineating regions. Moreover, the operational and computing costs associated with delineating SES units have not been thoroughly explored in the literature (Alessa et al., 2008; Hanspach et al., 2016; Sinare et al., 2016). This oversight highlights a gap in addressing the practical challenges of implementing these frameworks efficiently and effectively.

2.2 Objective

Addressing the research gaps identified in the literature review, this research proposes a generic framework designed to delineate the basic units of SES boundaries effectively. This framework aims to integrate vector and raster data across various scales and dimensions, utilizing spatial optimization techniques to control the spatial properties of the resulting SES units. It will also incorporate advanced dimension reduction algorithms to accommodate the non-linear characteristics of SES. Furthermore, the study seeks to enhance the accessibility and reproducibility of the methodology across different regions by employing open-source Python packages. Additionally, it aims to optimize computational resource requirements to ensure the framework's feasibility within the constraints of available resources.

This chapter discusses the methodology developed to delineate the basic units of SES. The methodology follows a structured four-phase approach. The initial phases focus on data preparation, involving the extraction and conversion of raster data into vector format, followed by the integration and cleaning of the vector data. The subsequent phase applies dimension reduction to streamline the dataset for further analysis. Finally, spatial optimization techniques are used to accurately delineate the basic units of SES, ensuring the methodology is efficient and effective for environmental analysis.

3.1 Method

To develop a generic approach for delineating the basic units of SES, this research integrates Geographic Information Systems (GIS), advanced dimension reduction algorithms, and spatial optimization techniques. GIS provides a comprehensive suite of tools for acquiring, storing, managing, and visualizing spatio-temporal data, making it an indispensable resource for spatial modeling and the analysis of geospatial data across various dimensions (Feng & Koch, 2024).

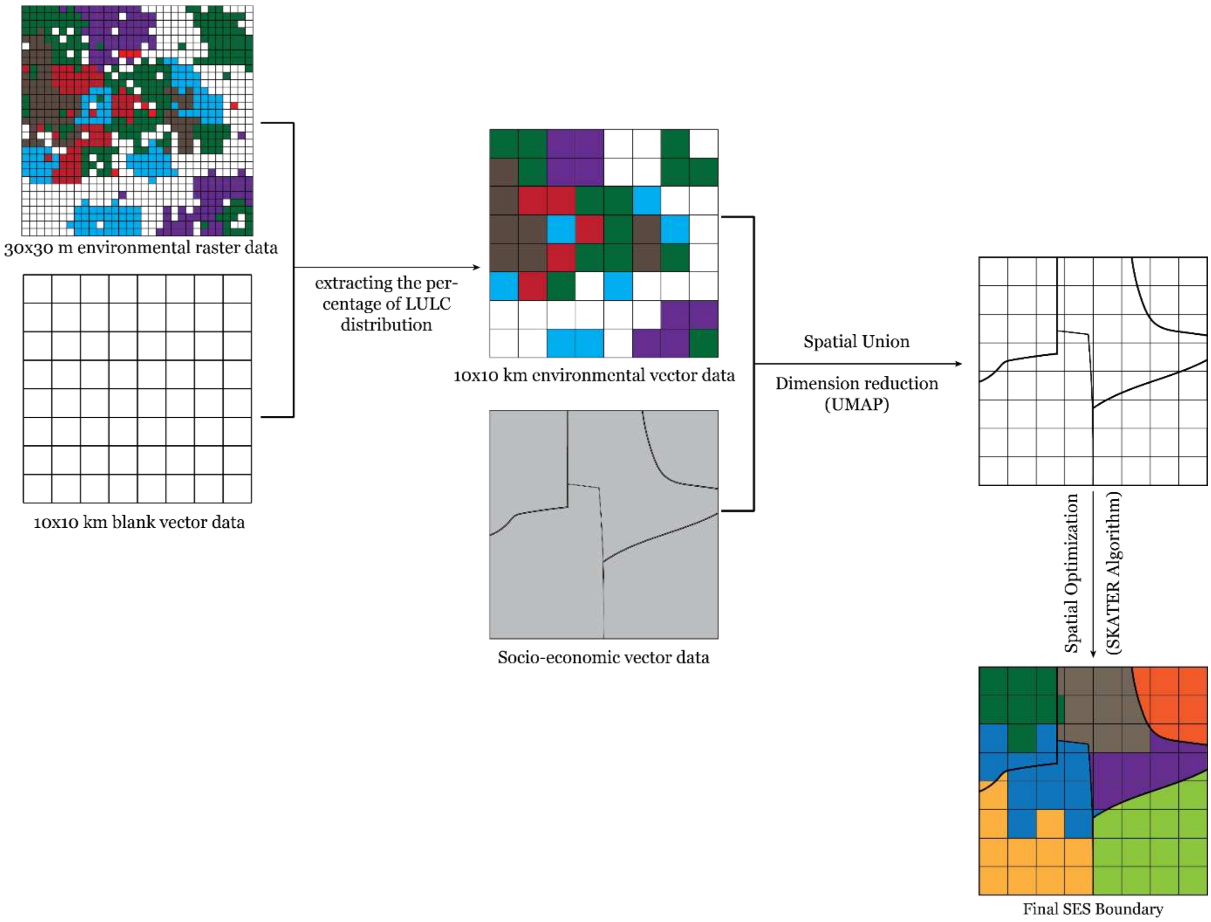

Figure 3.1: Methodology of the Study

Figure 3.1 outlines the methodological framework proposed by this research. This methodology is structured into a four-phase approach aimed at effectively delineating the basic units of SES. The initial two phases focus on data preparation necessary for the delineation process, beginning with the extraction of information from raster data and its conversion into vector format. Following this, the second phase involves overlaying all vector data and subjecting it to data cleaning processes. In the third phase, the cleaned data is processed through a dimension reduction algorithm to prepare it for regionalization algorithms. The final phase applies spatial optimization techniques to precisely delineate the basic units of SES, ensuring that the methodology is both efficient and effective in managing and utilizing spatial data for environmental analysis.

3.1.1 Phase 1: Extracting Information from Raster Data

Environmental data typically incorporates raster structures characterized by regularly tessellated grids, each grid-cell endowed with distinct attributes and clear relational dynamics with adjacent cells (DeMers, 2002; Pike et al., 2009). This structure often incorporates varied tessellations that showcase unique geometric properties (DeMers, 2002; Pike et al., 2009). Moreover, homogenous grids employing vector data structures facilitate precise sampling and offer greater flexibility in modeling the altitude field between data points, allowing for a detailed and accurate representation of geographic features (Tachikawa et al., 1994; Wilson, 2012). The process of modeling the land surface, regardless of the conceptual model or form of representation, involves the transformation and interpolation of sampled topographic data to match a specific reference frame and resolution that aligns with research objectives (Cavazzi et al., 2013; Hengl & Evans, 2009; Li et al., 2005).

Phase 1 of the methodology adopts a method akin to the fishnet technique (Xu et al., 2017) and involves the tessellation of raster data (Bishop et al., 2018). This phase meticulously overlays a grid to analyze the spatial characteristics of raster data, followed by comprehensive statistical computations within each grid cell. Initially, a structured grid is created, tailored to the dimensions and coordinate system of the raster dataset. Each cell within this grid serves as a distinct unit of analysis, enabling the localized examination of the raster's pixel values, as detailed in the pseudocode provided in Figure 3.2.

Upon establishing the grid, the raster data undergoes a systematic analysis where the composition of pixels within each grid cell is examined. Various statistics are computed, such as the count of different pixel values, the total area these values occupy within each cell, and their proportional coverage relative to the cell's total area. This quantitative analysis is pivotal for deciphering spatial patterns and variability within the raster data, facilitating the identification of regions characterized by high or low concentrations of specific values, which denote geographical or environmental attributes.

Algorithm: Analyzing Raster Data with Grid Overlay

Input: raster_path, raster_bounds = $(xmin, ymin, xmax, ymax)$, cell_size = c $(meters)$, raster_crs

Output: grid_gdf with pixel statistics

Algorithm:

1. Function Generation_Grid (raster_bounds, c, raster_crs):

   Calculate $n_x = \left|\frac{xmax-xmin}{c}\right|$, $n_y = \left|\frac{ymax-ymin}{c}\right|$

   For i in $[0, n_{x-1}]$, j in $[0, n_{y-1}]$:

   boxes = $\{[xmin + i \cdot c, ymin + j \cdot c, xmin + (i+1) \cdot c, ymin + (j+1) \cdot c] \mid 0 \leq i < nx$ $, 0 \leq j < ny\}$

   Return: \text{GeoDataFrame}(\text{boxes}, \text{raster_crs})

2. Function Calculate_Pixel_Statistics (raster_path, grid_gdf):

   Open raster, read U $(unique\ pixel\ values)$, \text{pixel_area} = \text{resolution}_x \times \text{resolution}_y

   $mask, data = mask(raster, box)$

   $counts = count\_pixels(data, U)$

   $A_u = \backslash text\{counts\}_u \times \backslash text\{pixel_{area}\} for\ u \in U$

   $A_{total} = \sum_{u \in U} A_u$

   $P_u = \frac{A_u}{A_{total}} \times 100\%\ for\ u \in U$

3. Execute:

   Load raster_ data to get \text{raster_bounds}, \text{raster_crs}.

   grid_gdf = Generate Grid $(raster\_bounds, c, raster\_crs)$

   Calculate Pixel Statistics $(raster\_path, grid\_gdf)$

   Output: grid_gdf

End of Algorithm

Figure 3.2: Raster Data Extraction with Grid Overlay

3.1.2 Phase 2: Integrating Vector and Raster Data

GIS offers a variety of data integration tools that facilitate the combination of raster and vector data. Among these, GIS overlay techniques are extensively utilized across various modeling approaches due to their effectiveness in handling multi-criteria applications that involve specific environmental thresholds (Faisal & Shaker, 2017; Weng, 2002). In this study, the GIS overlay integration method is employed to merge environmental data collected in Phase 1 with socio-economic data, creating a unified dataset for delineating the boundaries of Social-Environmental Systems (SES).

The methodology in Phase 2 follows a structured processing flow, beginning with the resolution of duplicated column names within GeoDataFrames. This process involves identifying duplicates and appending a sequence number to the first seven characters of each name to ensure their uniqueness. To achieve spatial consistency, the algorithm standardizes the coordinate reference systems (CRS) across all GeoDataFrames, aligning them with the CRS of the initial frame in the series. The spatial analysis is then enhanced through the intersection of these GeoDataFrames, which helps determine common geographic areas across the datasets.

Algorithm: GeoDataFrame Processing and Intersection

Inputs: Set of GeoDataFrames: $GDFs$
Outputs: Processed GeoDataFrame $result$

Algorithm:

1. Function: Rename_Duplicated_Columns(gdf)
   For each duplicated column dup in gdf:
   indices = find all indices where $gdf.columns\ =\ dup$
   $gdf.columns[indices] \leftarrow [dup\,[:7] +\ ``\ " + i\ if\ i\ \neq 0,$
   $else\ dup\ for\ i\ in\ range\ (length\ of\ indices)]$
2. Load GeoDataFrames:
   $GDFs == load\ each\ GDF\ from\ shapefiles$
3. Standardize CRS:
   $common\_crs \leftarrow GDFs[0].crs$
   For each gdf in $GDFs$:
   $gdf.to\_crs(common\_crs)$
4. Intersect GeoDataFrames:
   $result \leftarrow GDFs[0] result \leftarrow GDFs[0]$
   For $gdf\ in\ GDFs[1:]$:
   $result \leftarrow gpd.overlay(result, gdf, how = 'intersection')$
5. Shorten and Rename Columns:
   $result.columns \leftarrow [col[:10] for\ col\ in\ result.columns]$
   \text{Rename_Duplicated_Columns(result)}
6. Process Geometry:
   Explode MultiPolygons:
   $result \leftarrow result.explode(index\_parts = False).reset\_index(drop = True)$
   Convert MultiPolygon to Polygon by maximum area:
   $result['geometry'] \leftarrow result.apply(\lambda row: max(row.geometry, key = \lambda a: a.area)\ if\ isinstance(row.geometry, MultiPolygon) else\ row.geometry, axis = 1)$

End of Algorithm

Figure 3.3: Integrating Vector and Raster Data

Following the intersection, column names are truncated to the first ten characters for improved clarity, and any remaining duplications are systematically addressed. The final stages of this phase involve the geometric processing of the data, where MultiPolygon geometries are simplified into individual polygonal components. In instances where MultiPolygons remain, the polygon with the largest area is selected for detailed analysis. This methodical approach not only streamlines the data preparation process but also ensures that the datasets are optimally configured for comprehensive geographic analysis, as detailed in the pseudocode illustrated in Figure 3.3.

3.1.3 Phase 3: Dimension Reduction

Dimension reduction techniques are extensively employed in GIS to streamline complex data and enhance processing efficiency. These techniques not only simplify data but also improve the accuracy and speed of data extraction and processing (Song et al., 2019; Wang et al., 2022). Among the various dimension reduction methods used in GIS,, PCA (Martín-López et al., 2017), Stochastic Neighbor Embedding (SNE), t-distributed Stochastic Neighbor Embedding (t-SNE) (Liu et al., 2021; Shi et al., 2023), Multidimensional Scaling (MDS) (Mueller, 2004), Isomap (Kanishka & Eldho, 2017) are the most prevalent.

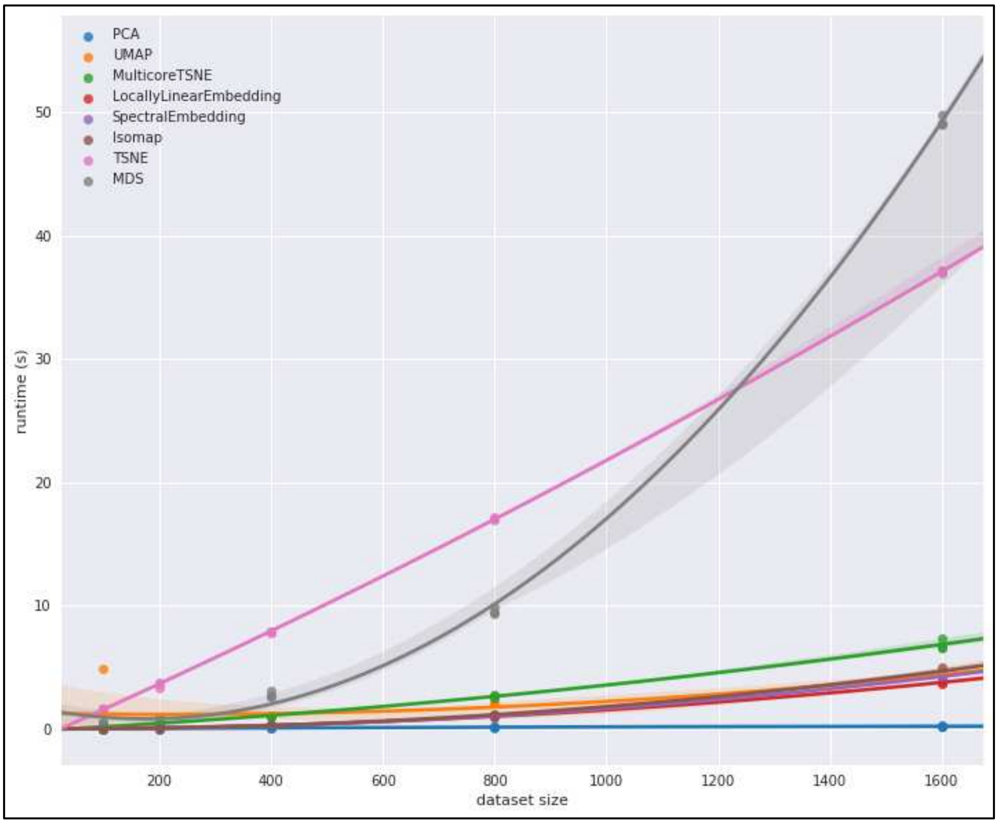

Figure 3.4: Performance Testing by Dataset Size. From "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction" by McInnes, L, Healy, J, 2018, ArXiv e-prints. Copyright 2018 by ArXiv e-prints.

Figure 3.4 compares the processing times of these algorithms for large datasets, highlighting that most dimension reduction techniques are relatively slow except for PCA, which is faster but limited by its reliance on linear data transformation (McInnes & Healy, 2018). Given the non-linear relationships inherent in the fractal characteristics of SES, there is a need for a dimension reduction algorithm that can handle such complexities more efficiently.

Uniform Manifold Approximation and Projection (UMAP) is proposed for this purpose, as it aims to preserve both local and global data structures, making it well-suited to the non-linear nature of SES. This method offers a balance between computational efficiency and the ability to capture intricate data relationships, thereby enhancing the overall analysis and interpretation of SES datasets.

UMAP is constructed from a theoretical framework based in Riemannian geometry and algebraic topology. The result is a practical scalable algorithm that is applicable to real world data. The UMAP algorithm preserves more of the global structure with superior run time performance. Furthermore, UMAP has no computational restrictions on embedding dimensions, making it viable as a general-purpose dimension reduction technique for machine learning (McInnes & Healy, 2018).

The UMAP algorithm operates in two main phases: graph construction in the high-dimensional space and optimization of the graph layout in the low-dimensional space. The first phase involves identifying the k nearest neighbors for each observation based on a distance metric, typically Euclidean (McInnes & Healy, 2018; Wang; et al., 2021). A minimal positive distance (equation 1), ρ_i, and a scaling parameter (equation 2), δ_i, are then computed for each observation to normalize distances and preserve relative proximities. In the second phase, the algorithm initializes the low-dimensional representation using spectral embedding (equation 5) and iteratively applies attractive and repulsive forces (equation 6) to adjust the positions of observations. The attractive force pulls observations closer together based on their high-dimensional relationships, while the repulsive force pushes non-neighboring observations apart (McInnes & Healy, 2018; Wang; et al., 2021). The UMAP dimension reduction considers the following notations:

$$input\ dataset, \chi = \{x_1, x_2, x_{3,} \dots \dots \dots \dots, x_N\}$$

$$dissimilarity\ measure, d: \chi \times \chi \longrightarrow \mathbb{R}_{\geq 0}$$

$$\kappa = number\ of\ nearest\ neighbors$$

$$\epsilon = small\ constant$$

$$\rho_i = \min\left\{d\left(x_i, x_{i_j}\right) 1 \leq j \leq \kappa, d\left(x_i, x_{i_j}\right) > 0\right\} \tag{1}$$

$$\sum_{j=1}^{\kappa} \exp\left(\frac{-\max\left(0, d\left(x_i, x_{i_j}\right) - \rho_i\right)}{\delta_i}\right) = \log_2(\kappa) \tag{2}$$

$$set\ of\ directed\ edges, E = \left\{\left(x_i, x_{i_j}\right) | \leq j \leq \kappa, 1 \leq i \leq N\right\}$$

$$weight\ function, \omega\left(\left(x_i, x_{i_j}\right)\right) = \exp\left(\frac{-\max\left(0, \left(x_i, x_{i_j}\right) - \rho_i\right)}{\delta_i}\right) \tag{3}$$

$$symmetric\ weight\ of\ an\ edge\ (i,j),\ \varpi_{i,j} = \omega(x_i, x_j) + \omega(x_j, x_i) - \omega(x_i, x_j).\omega(x_j, x_i) \quad (4)$$

$$embedded\ approximation\ of\ the\ gradient, y_i = y_i + \alpha.\frac{-2ab\|y_i - y_j\|_2^{2(b-1)}}{1+a\left(\|y_i - y_j\|_2^2\right)^b}\varpi_{i,j}(y_i - y_j) \quad (5)$$

$$repulsive\ force, y_i = y_i + \alpha.\frac{b}{(\epsilon + \|y_i - y_k\|_2^2)\left(1+a(\|y_i - y_k\|_2^2)^b\right)}(1 - \varpi_{i,k})(y_i - y_k) \quad (6)$$

Algorithm: GeoDataFrame Processing for UMAP Dimensionality Reduction
Inputs: gdf: GeoDataFrame loaded with data. **Outputs:** new_gdf: GeoDataFrame with UMAP components.
Algorithm: 1. Function: Is_Categorical(column, threshold) return $column.dtype = 'object'\ or\ \mid unique(column) \mid \leq threshold$ 2. Separate Columns: Categorical Columns: \text{categorical_cols} = [col \; \text{for} \; col \; \text{in gdf.columns} \; \text{if Is_Categorical}(gdf[col])] Quantitative Columns: $quantitative\_cols = [col\ for\ col\ in\ gdf.columns\ if\ col \notin categorical\_cols\ and\ gdf[col].dtype \in 'biufc']$ Exclude 'geometry': $quantitative\_cols = [col\ for\ col\ in\ quantitative\_cols\ if\ col \neq 'geometry']$ 3. Handle NaNs in Quantitative Columns: For col in quantitative_cols: if $NaN\ or\ Inf$ in $gdf[col]$: $median = median(gdf[col])median = median(gdf[col])$ $Replace\ Inf\ with\ NaN, Fill\ NaN\ with\ median$ 4. Scale Quantitative Variables: \text{scaled_quant_vars} = \text{StandardScaler().fit_transform}(gdf[\text{quantitative_cols}]) Dummy Variables for Categorical Columns: \text{dummy_vars} = \text{get_dummies}(gdf[\text{categorical_cols}], \text{dummy_na=True}) 5. Combine Data: combined_data=concat(dummy_vars,scaled_quant_vars)combined_data=concat(dummy_vars,scaled_quant_vars) Raise error if NaN in combined_datacombined_data 6. Impute Missing Values: \text{imputed_data} = \text{KNNImputer(n_neighbors=5).fit_transform}(\text{combined_data}) 7. UMAP Dimensionality Reduction: a. Initialize UMAP: i. umap_reducer=UMAP($n\_neighbors$=15,min_dist=0.1,$n\_components$=10,$random\_state$=10)umap_reducer=UMAP($n\_neighbors$=15,min_dist=0.1,$n\_components$=10,$random\_state$=10) b. Apply UMAP: i. \text{umap_embedding} = \text{umap_reducer.fit_transform}(\text{imputed_data}) c. umap_df=DataFrame(umap_embedding,columns=['UMAP_'+i+1])umap_df=DataFrame(umap_embedding,columns=['UMAP_'+i+1])
End of Algorithm

Figure 3.5: Dimension Reduction Pseudocode

Phase 3 of the methodology focuses on preparing the dataset for dimensionality reduction, ensuring it is clean and ready for advanced analysis. The process began with the categorization of the data into categorical and quantitative columns using a function to determine whether a column's data type is object or if it has a limited number of unique values. Categorical columns have been identified, and quantitative columns have been filtered out, with the 'geometry' column excluded to focus on relevant numerical data.

Handling missing values in the quantitative columns is also crucial. If any column contained NaN or Inf values, these have been replaced with the column's median value to maintain data integrity. The quantitative variables have then been scaled using standard scaling techniques to ensure uniformity across the dataset. For categorical data, dummy variables have been created to represent categorical values numerically. These dummy variables have been combined with the scaled quantitative variables to form a comprehensive dataset. Any remaining missing values have been imputed using a K-Nearest Neighbors (KNN) imputer to ensure no gaps in the data.

The final step involved application of an open-source python package called UMAP for dimensionality reduction. UMAP has been initialized with specific parameters and applied to the imputed data, reducing its dimensions while preserving its essential structure. The resulting UMAP embeddings have been organized into a new DataFrame, ready for further analysis and interpretation. This methodical approach has ensured that the data is accurately processed and optimized for subsequent analytical phases.

3.1.4 Delineating SES Boundaries through Regionalization

Regionalization can be approached through three primary methodologies. The first method employs a two-step algorithm, starting with clustering non-spatial attributes and subsequently dividing non-contiguous objects into separate regions. Although this approach facilitates a quick evaluation of spatial dependence, it does not directly account for spatial adjacency. Consequently, it is limited in its capacity to capture spatial patterns effectively. This often results in more regions than desirable, particularly in datasets with low spatial autocorrelation (Haining et al., 2000). The second approach integrates both geographical positions and non-spatial features, using centroid coordinates as additional attributes. Similarity is measured as a weighted mean of feature space and geographical proximity. This method, employed by the SAGE system, uses iterative processes to achieve connected clusters, balancing homogeneity, compactness, and equality to produce comparable regions, typically based on population data (Haining et al., 2000; Martin, 1998). However, some critics argue that homogeneity should be the sole objective function, with compactness and equality treated as constraints (Openshaw et al., 1998). The third approach employs adjacency constraints in clustering. The AZP (Automatic Zoning Procedure) algorithm begins with a random partition and reallocates objects to minimize an objective function while maintaining adjacency. Although improvements like the ZDES system address the computational expense of AZP, more efficient methods utilizing adjacency relations are discussed in subsequent sections (Alvanides et al., 2002). In the final phase of the framework, spatial optimization techniques are implemented to delineate the basic units of SES. One such technique is the Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) algorithm. SKATER is an efficient method for regionalizing socio-environmental units represented as spatial objects, combining the use of a minimum spanning tree with combinational optimization techniques (AssunÇão et al., 2006). SKATER is highly flexible, enabling users to specify their criteria and identify an optimal regionalization scheme. Consequently, this algorithm has significant potential for widespread application in defining basic spatial units across various use cases.

The SKATER algorithm transforms the regionalization problem into a graph partitioning problem. It is a constrained spatial regionalization algorithm based on spanning tree pruning, where a pre-specified number of edges are cut in a continuous tree to group spatial units into contiguous regions. The first step involves creating a connectivity graph that captures the neighborhood relationships between spatial objects, with the cost of each edge inversely proportional to the similarity between the regions it connects. This neighborhood structure is organized by a minimum spanning tree (MST), a connected tree without circuits. The subsequent step is partitioning the MST by successively removing edges that link dissimilar regions, resulting in connected regions with maximum internal homogeneity (AssunÇão et al., 2006; Feng et al., 2022). Following are the notations of SKATER algorithm:

$i = The\ region\ lebel\ for\ each\ observation, z_i\ \epsilon\ \{1, 2, 3, \dots, k\}$

$z = Cluster\ Lables$

$N = Number\ of\ observations$

$P = Number\ of\ features$

$X\ \varepsilon\ \mathbb{R}^{N \times P}$

$W = weights\ matrix\ (binary\ and\ symmetric\ matrix\ expressing\ neighbor\ relationships)$

$L_k = Laplacian\ Matrix$

$\lambda_i = Eigenvalues$

$$Dissimilarity\ function, d(i, j) = \left\| x_i - x_j \right\|^2 \quad (7)$$

Given a set of dissimilarities $\{d_i\}$ within a cluster, the reduction function R can be generally formulated as:

$$R\ (\{d_i\}) = Aggregate\ (\{d_i\}) \tag{8}$$

Given a set of attribute vectors $\{x_i\}$ within a cluster R_k, the center function C can be generally formulated as:

$$C(R_k) = Center(\{x_i \mid i \in R_k\}) \tag{9}$$

The SKATER model is formulated as follows:

$$min_z \sum_{k=1}^{K} R\left(\sum_{i \in R_k} d\left(x_i, C(R_k)\right)\right) \tag{10}$$

Subject to:

$$\sum_{k=1}^{K} 1\,\{z_i = k\} = 1 \quad \forall i \qquad where \begin{cases} 1\,\{z_i = k\} = 1\ if\ z_i = k \\ 0, otherwise \end{cases} \tag{11}$$

$$|R_k| \geq quorum \quad \forall k \tag{12}$$

$$\lambda_1(L_k) = 0\ and\ \lambda_i(L_k) > 0\ for\ i = 2, \dots, |R_k| \tag{13}$$

The primary objective (10) of the SKATER algorithm is to enhance the optimization of spatial clustering, specifically the delineation of the basic units of Social-Ecological Systems (SES), by minimizing the total intra-cluster dissimilarity. The objective function is meticulously crafted to accumulate the dissimilarities between individual observations and their respective cluster centers. Constraint (11) guarantees that each observation is exclusively allocated to one cluster. To avoid the formation of unduly small clusters and to maintain statistical significance, each cluster is required to encompass a minimum number of observations, as dictated by a predetermined quorum. Thus, constraint (12) imposes a lower limit on cluster size to ensure adequacy in terms of statistical representation. A pivotal feature of the SKATER algorithm is the necessity for each cluster to constitute a connected subgraph within the spatial weights graph, as determined by the adjacency matrix. Constraint (13) ensures that each cluster forms a single, contiguous entity within the spatial graph, reinforcing the spatial coherence of the clustering process.

An open-source Python package, ‘spopt,’ developed by Feng et al. (2022), has been utilized to delineate the basic units of Social-Ecological Systems (SES) through regionalization. The algorithm detailed in the pseudocode (Figure 3.6) employs the SKATER method, which is specifically designed to discern and visualize patterns in geographic data. The process commences by setting up the computational environment, suppressing warnings to streamline output, and recording the start time to evaluate the efficiency of the analysis. A critical element of this procedure involves constructing a spatial weights matrix using Queen contiguity, which defines the adjacency relationships between geographic units and fundamentally influences the clustering dynamics. These weights are essential for identifying spatial connections that are crucial for subsequent clustering phases. The clustering operation is governed by parameters tailored to measure dissimilarities among geographic units using the SKATER algorithm, ensuring the analysis accurately reflects the unique spatial attributes of the dataset.

The algorithm iteratively clusters the geographic units into varying numbers of groups, ranging from two to ten, allowing for a comprehensive exploration of potential spatial organizations. Post-clustering, the results for each configuration of cluster sizes are visualized through a series of plots, providing a clear visual representation of the spatial distribution of clusters. This visualization is integral to the analysis, facilitating an intuitive understanding of how geographic units group together under different scenarios. Finally, the algorithm concludes by calculating the elapsed time from the start to the end of the process. This measurement provides valuable insights into the computational efficiency of the clustering operation, highlighting the performance of the algorithm under the given data and parameter settings.

Algorithm: Spatial Clustering with SKATER

Inputs: GeoDataFrame gdf, Spatial weights W, Attributes $attrs$, Cluster sizes $N = [2,4,6,8,10]$

Outputs: gdf with cluster labels, Visual plots for cluster sizes.

Algorithm:

1. Initialization
   $W \leftarrow libpysal.weights.Queen.from\_dataframe(gdf)$
2. Set SKATER Parameters:
   $params \leftarrow \{dissimilarity: skm.manhattan\_distances, reduction: np.sum, center: np.mean, verbose: False\}$
3. Define Clustering Attributes:
   $attrs \leftarrow [UMAP_i\ for\ i\ in\ range(10)]$
4. Cluster Formation Loop:
   For n in N:
   $model \leftarrow spopt.region.Skater(gdf, W, attrs, n, floor = 50, trace = True, islands = 'ignore', params)$
5. Visualization
   $(f, axarr) \leftarrow plt.subplots\ (1, len\ (N), figsize = (15, 10))$
   For i, n in enumerate (N):
   $gdf.plot\ (column = \text{“}clusters\text{”}_n, ax = axarr\ [i], cmap = \text{“}tab20\text{”})$
   $axarr[i].set\_title(\text{"}Clusters: \text{"} + n)$
   $axarr[i].set\_axis\_off()$

End of Algorithm

Figure 3.6: Delineating SES Boundaries through Regionalization

4.1 Study Area

To implement the proposed framework in a practical scenario, the Rio Grande/Bravo Basin (RGB) area (figure 4.1) has been selected as the case study. The socio-environmental geodatabase of RGB, curated by Plassin et al. in 2020, provides an ideal dataset to test this methodology. The RGB is a transboundary region spanning the United States and Mexico, offering diverse environmental and socio-economic characteristics, making it suitable for large-scale, spatially explicit mapping of Social-Ecological Systems (SES).

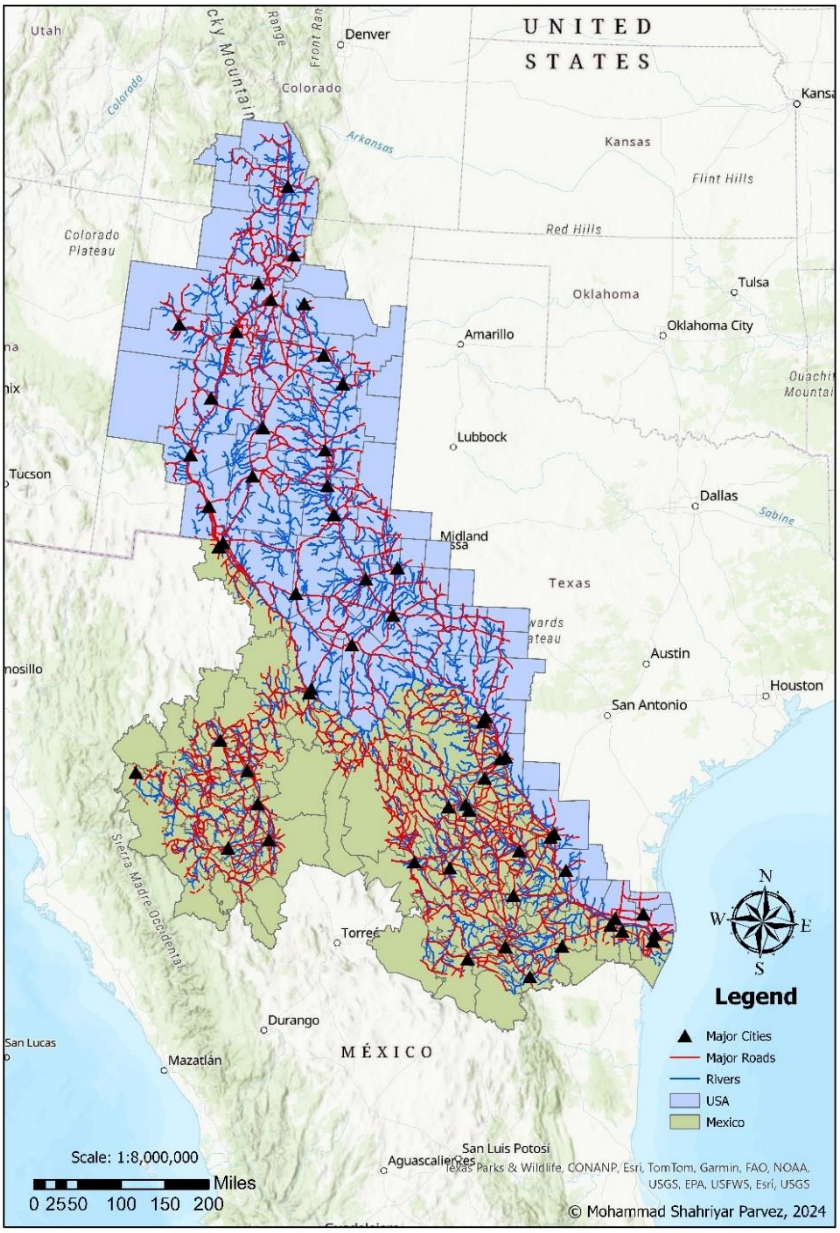

Figure 4.1: The Rio Grande Basin

The Rio Grande River, partially forming the border between the U.S. and Mexico, is the fifth-longest river in North America. The basin covers an area of 552,382 km² and supports a population of approximately 10.4 million people. Key economic sectors include agriculture and oil and gas production, both of which heavily depend on hydrological resources. Due to the limited availability of surface water, these sectors increasingly rely on groundwater. The RGB encompasses varied landscapes, agricultural systems, and political and administrative units, presenting a complex SES with the common issue of limited water availability (Koch et al., 2021).

The basin is also a critical source of irrigation water for the southwestern U.S. Historical records indicate a 25% decline in source waters between 1958 and 2015 (Chavarria & Gutzler, 2018). Future projections suggest an 18% or greater reduction in runoff by the end of the 21st century (Elias et al., 2015), with climate change expected to have a significant impact, particularly in the upper RGB (Dettinger et al., 2015). The RGB is over-appropriated, meaning that water availability frequently fails to meet the demands of all water rights holders. The downstream segment of the river, from El Paso to Ojinaga, often runs dry due to drought and misuse, earning the moniker "The Forgotten River" to highlight its dire condition (CNN, 2001).

4.2 Input Data

Geospatial data for the Rio Grande/Bravo Basin (RGB) application has been sourced from a socio-environmental geodatabase developed by Plassin et al. (2020). Accessible via the Open Science Framework, this comprehensive geodatabase includes 145 datasets (125 vector and 20 raster) covering the transboundary basin. Categorized into themes such as Water and Land Governance, Hydrology, Water Use and Hydraulic Infrastructures, Socioeconomics, and the Biophysical Environment, the total size of the geodatabase is 1.40 gigabytes (GB).

To demonstrate the efficacy and potential of the newly developed approach, seven representative datasets have been selected. These include two raster datasets: land use/land cover data (figure 4.2a) and elevation data (figure 4.2b). The land use categories encompass temperate or sub-polar needleleaf forest, tropical or sub-tropical broadleaf evergreen forest, tropical or sub-tropical broadleaf deciduous forest, temperate or sub-polar broadleaf deciduous forest, mixed forest, tropical or sub-tropical shrubland, temperate or sub-polar shrubland, tropical or sub-tropical grassland, temperate or sub-polar grassland, wetland, cropland, barren lands, urban and built-up, and water. Additionally, four vector datasets have been chosen: composite land management zones (figure 4.2c), eco-regions (figure 4.2d), administrative boundaries consisting of 68 U.S. counties and 135 Mexican municípios (figure 4.2e), hydrological sub-basin boundaries (figure 4.2f), These selections provide diverse data models (raster and vector), data types (categorical and numerical), and key determinants for hydrological resource management and governance. The study area contains between 10^6 to 10^9 pixels per raster layer and 10^2 to 10^5 polygons per vector layer, depending on the resolution. By employing these datasets, the practical application and robustness of this approach in analyzing and managing socio-environmental systems within the RGB are effectively demonstrated.

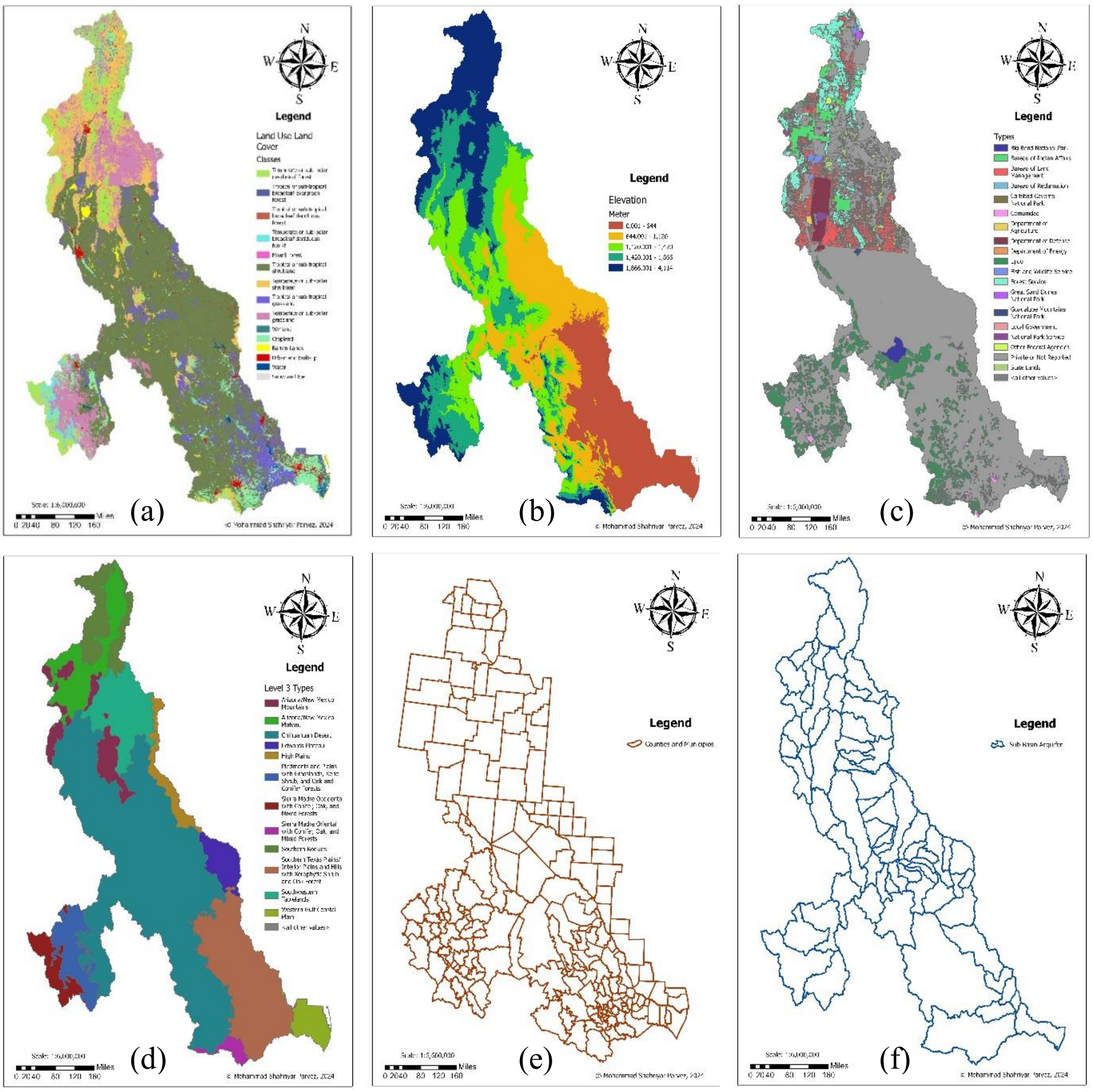

Figure 4.2: Selected Geospatial Dataset from Plassin et al. (2020)

Table 4.1 and 4.2 gives a brief description of the vector dataset selected for the analysis.

Table 4.1: Brief Description of the Selected Vector Datasets

Name of the dataset	Data Type	No of Variables
Administrative Boundaries	Nominal and Ratio	224
Eco-regions	Nominal	18
Land Management	Nominal	9
Sub-basin	Nominal and Ratio	11

Table 4.2: Brief Description of the Selected Raster Datasets

Name of the dataset	Spatial Resolution	No of classes
Land Use Land Cover	30 meters	17
Elevation	916 meters	NA

4.3 Data Processing and SES Delineation

The objective of this application is to delineate SES regions based on user preferences, while adhering to strict constraints that prevent these regions from crossing county and sub-basin boundaries. The generation of SES units is governed by criteria such as estimated total quantity, compactness, spatial contiguity, and similarity in socio-environmental characteristics. These constraints enable policymakers to implement targeted policies and distribute resources efficiently based on regional needs.

The proposed framework begins by extracting raster data information using the fishnet technique. LULC and elevation data are integrated into the fishnet vector data, followed by overlaying additional vector datasets to create a comprehensive base dataset for SES delineation. To maintain the critical constraint of preserving county and sub-basin boundaries, the next step involves partitioning the entire RGB area into distinct, non-overlapping polygons. Each polygon uniquely combines information from sub-basins, counties/municípios, LULC, and other vector datasets with various boundaries beyond administrative lines.

To illustrate and visualize the framework effectively, two distinct polygons have been selected (Figure 4.3). One polygon is located in Mexico, following the Bravo/Sosa sub-basin boundary and covering approximately 62,656 km², while the other is in New Mexico, USA, following the Eddy County boundary and covering approximately 10,873 km². These polygons have been chosen for their diverse LULC types, with Eddy County containing 13 of the original 17 LULC categories, and Bravo/Sosa sub-basin containing 11. This selection process ensures the preservation of most environmental data, despite segmenting the study area into smaller units.

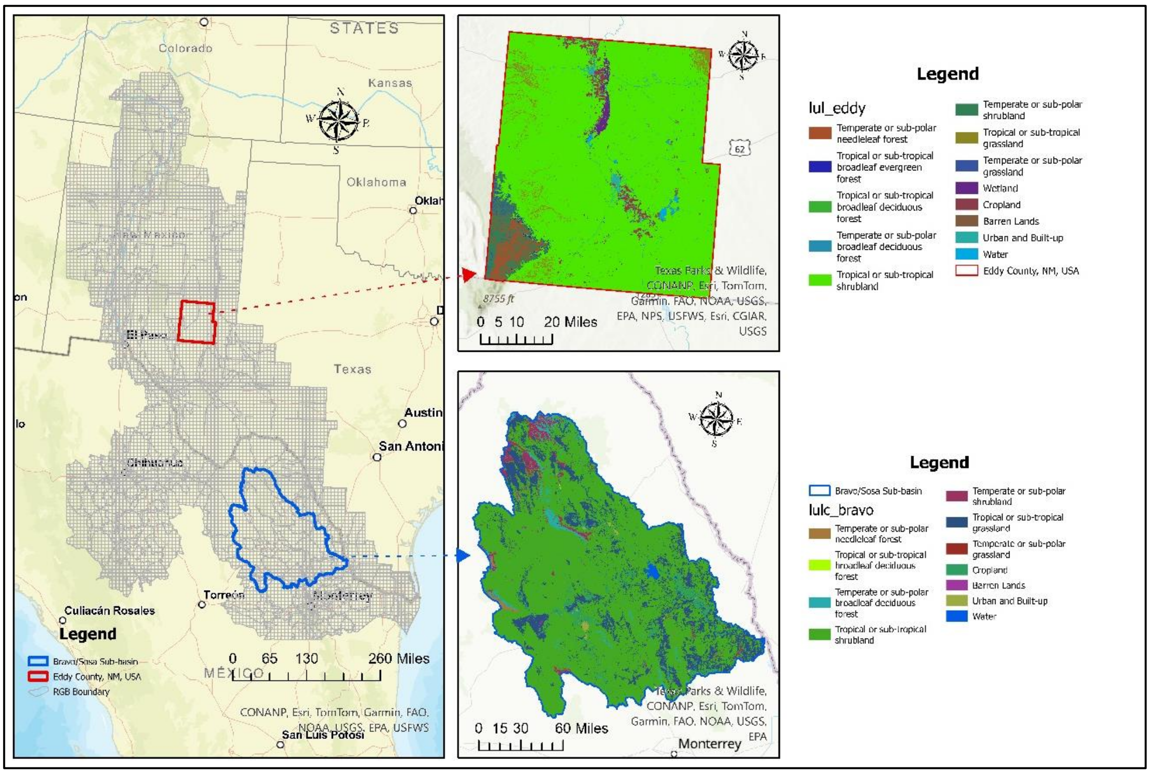

Figure 4.3: Land Use Land Cover Classes of Eddy County and Bravo/Sosa Sub-basin

Figures 4.4, 4.5, and 4.6 respectively show the elevation, land management types, and ecoregions of Eddy County and Bravo/Sosa Sub-basin. The elevation of both polygons has been categorized into five classes using the Jenks Natural Breaks Method (North, 2009) to enhance data visualization and understanding. The ranges indicate that Bravo/Sosa sub-basin has greater elevation variance, ranging from 83 meters to 2860 meters. Conversely, Eddy County has an elevation range from 866 to 2225 meters, indicating a flatter terrain.

In terms of land management, both polygons are managed by either public or private entities, reflecting a symmetry in the data. Regarding ecoregions, Bravo/Sosa sub-basin is nearly divided between two ecoregions, whereas Eddy County spans four different ecoregions, predominantly the Chihuahuan Desert Ecoregion. Despite these differences, both polygons share a similar profile in terms of input data, demonstrating the robustness of the proposed framework in analyzing and managing socio-environmental systems within the RGB.

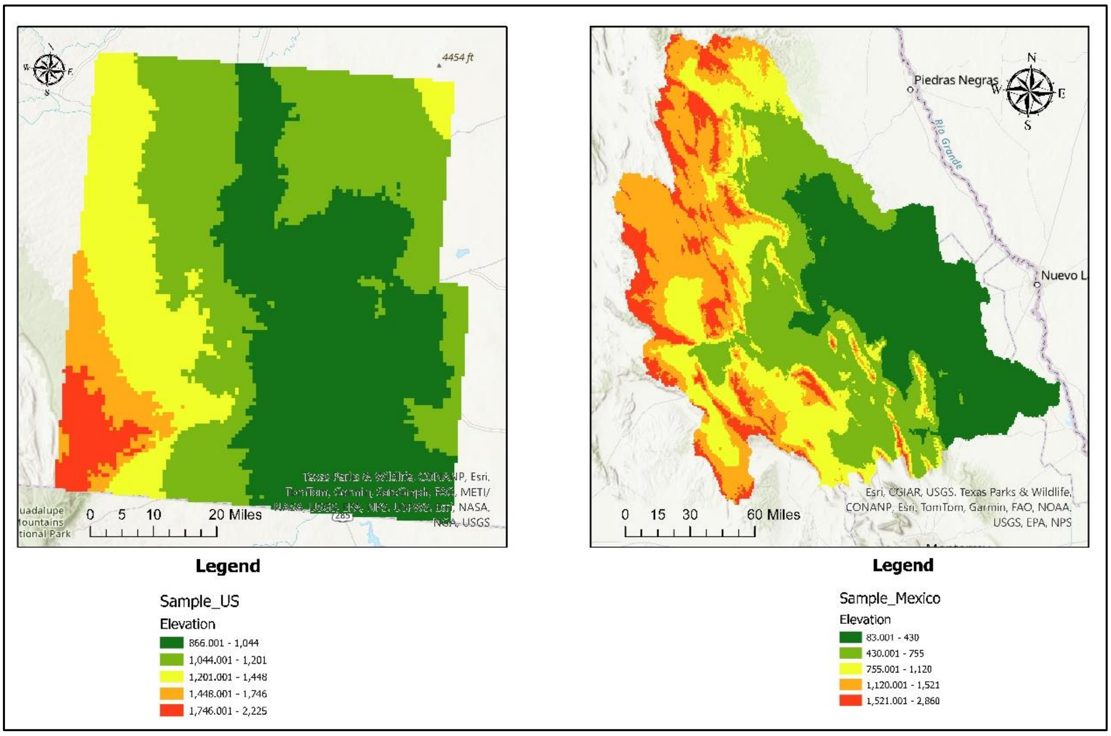

Figure 4.4: Elevation Data of Eddy County and Bravo/Sosa Sub-basin

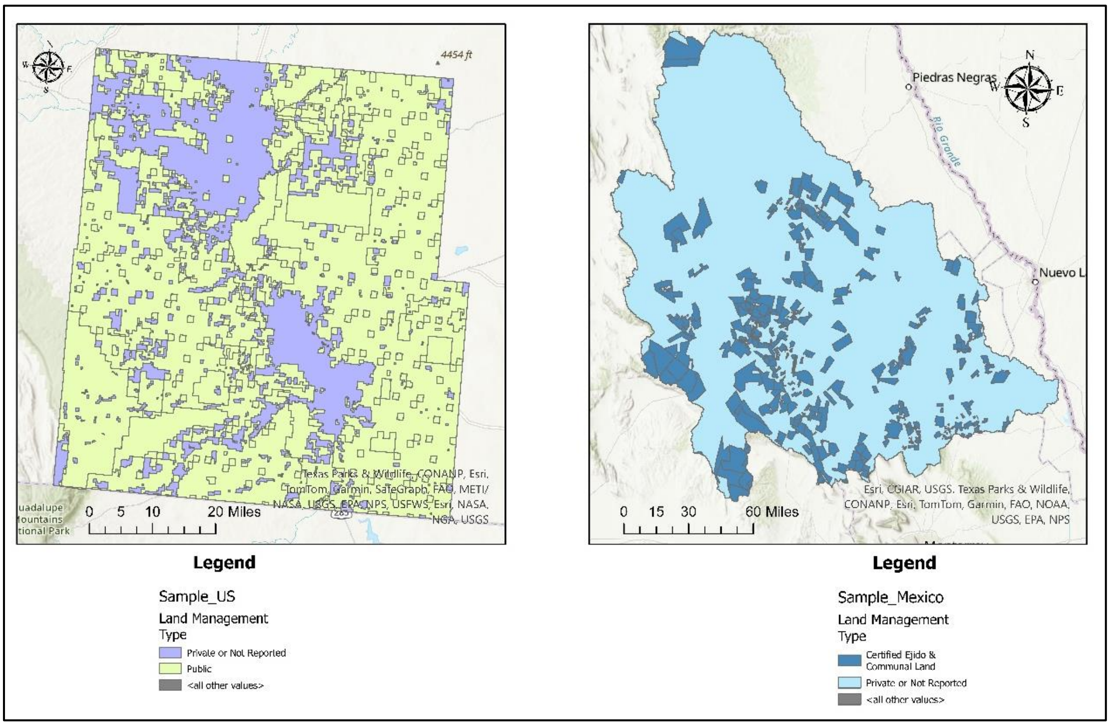

Figure 4.5: Land Management Data of Eddy County and Bravo/Sosa Sub-basin

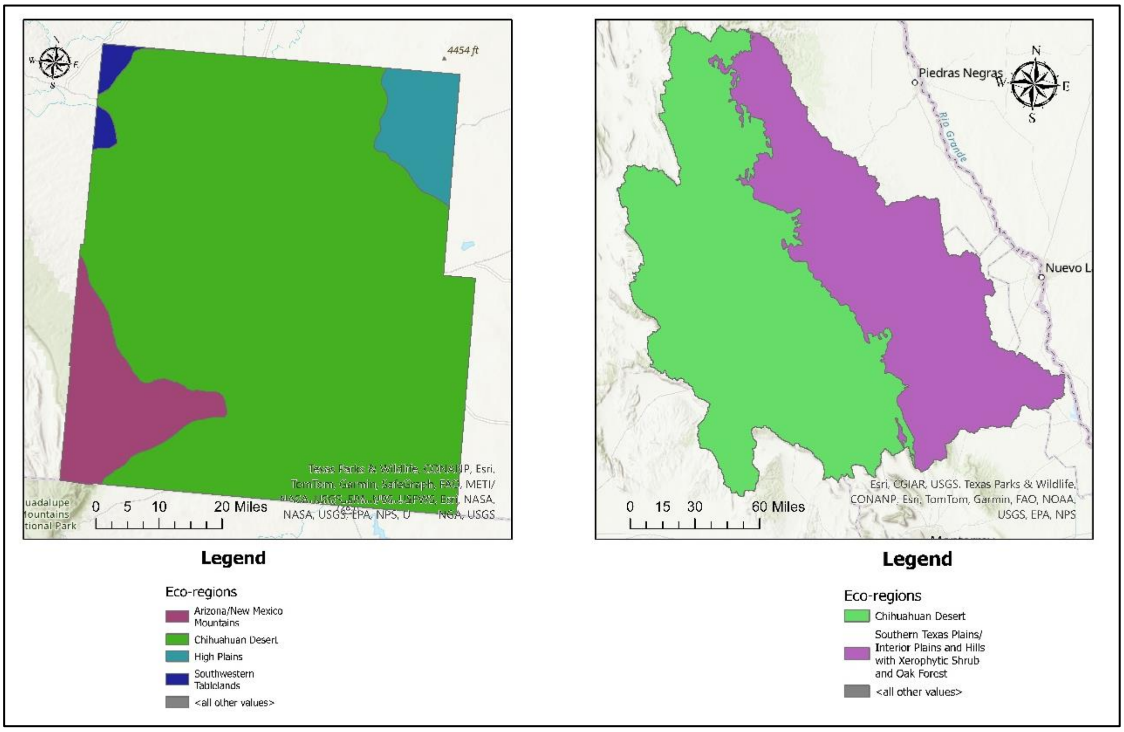

Figure 4.6: Eco-regions of Eddy County and Bravo/Sosa Sub-basin

All datasets for each sample polygon have been integrated according to the second phase of the methodology. After integration, the number of variables for Eddy County expanded to 136 with 251 features, while for the Bravo/Sosa sub-basin, the number of variables increased to 102 with 1115 features. Given the high-dimensional nature of these datasets, the SKATER algorithm—renowned for its constrained spatial regionalization capabilities—requires significant computational power and time to process and provide accurate solutions.

To mitigate these computational demands, the dimension reduction algorithm has been employed. UMAP is particularly well-suited for this task due to its ability to handle complex and non-linear relationships within the data, effectively preserving essential patterns and structures. This algorithm not only reduces the computational load but also maintains the integrity of the dataset's inherent characteristics. By implementing UMAP, the complexity of the high-dimensional data is distilled into a more manageable form, ensuring efficient processing without sacrificing the quality of the analysis.

UMAP offers flexibility by allowing users to specify the number of output dimensions. For this demonstration, the variable size for both Eddy County (originally 136 variables) and the Bravo/Sosa Sub-basin (originally 102 variables) has been reduced to 10 new variables using the UMAP process, minimizing data distortion. Greater emphasis has been placed on sub-basin-related variables for Eddy County and municípios-related variables for the Bravo/Sosa sub-basin to ensure that the final SES units prioritize existing sub-basin boundaries for Eddy County and municípios administrative boundaries for the Bravo/Sosa sub-basin. This significant reduction streamlines the data, making it more manageable for subsequent analysis while retaining the critical information necessary for accurate SES delineation.

In the next step, the SKATER algorithm has been implemented using the 10 variables derived from the UMAP dimension reduction. The SKATER algorithm, known for its ability to manage spatial regionalization with constraints, enables the regulation of regional divisions and ensures a specified minimum number of spatial entities within each designated region. The algorithm environment has been configured to allocate a minimum of 50 and 10 features to each region within Eddy County and Bravo/Sosa Sub-basin polygons respectively. This configuration ensures that each SES region has enough spatial entities to maintain statistical and practical significance.

To explore different regional configurations, the SKATER algorithm has been run to generate 2, 4, 6, 8, and 10 regions for each of the sample polygons. This iterative approach allows for a comprehensive examination of how varying the number of SES units affects the overall delineation and spatial distribution. Both the minimum number of features per region and the number of regions (i.e., SES units) can be adjusted through user input, providing a high degree of control over the final spatially explicit units. This flexibility is crucial for tailoring the analysis to specific user needs and policy objectives.

Additionally, the time required to delineate these SES boundaries has been meticulously recorded to provide insights into the computational efficiency of the process (Tables 4.3 and 4.4). By documenting the computational time, the performance of the algorithm under various configurations can be rigorously assessed, enabling the optimization and refinement of the methodology. This comprehensive approach ensures that the framework not only generates accurate and meaningful SES regions but also operates efficiently within practical time constraints. It is important to note, however, that the recorded times for delineating SES units do not include data processing time. The recorded time specifically represents the duration required by the SKATER algorithm after all data preprocessing has been completed and dimensions have been reduced using the UMAP algorithm.

Table 4.3: Delineation of SES units of Eddy County, NM, USA

Number of SES Units	Number of Features in each SES	Time Required to Solve
2	(160, 91)	0.29 seconds
4	(27, 195, 18, 11)	0.58 seconds
6	(27, 149, 26, 20, 18, 11)	1.01 seconds
8	(27, 122, 26, 11, 20, 18, 11, 16)	1.20 seconds
10	(27, 84, 28, 26, 11, 20, 18, 10, 11, 16)	1.51 seconds

Table 4.4: Delineation of SES units of Bravo/Sosa Sub-basin

Number of SES Units	Number of Features in each SES	Time Required to Solve
2	(505, 610)	1.12 seconds
4	(198, 307, 345, 265)	2.20 seconds
6	(198, 195, 51, 345, 61, 265)	3.48 seconds
8	(85, 113, 195, 51, 243, 102, 61, 265)	4.62 seconds
10	(85, 113, 195, 51, 188, 102, 55, 61, 53, 212)	5.75 seconds

Identifying SES units necessitates the aggregation of smaller polygons into larger, contiguous regions to minimize internal dissimilarity across five key attributes within each region. Each region is uniformly color-coded to ensure geographic continuity. Although some regions may initially appear fragmented, closer inspection—upon magnification—reveals that they are interconnected through vertices or narrow lines, thereby maintaining their integrity as single entities. (figure 4.7 and 4.8).

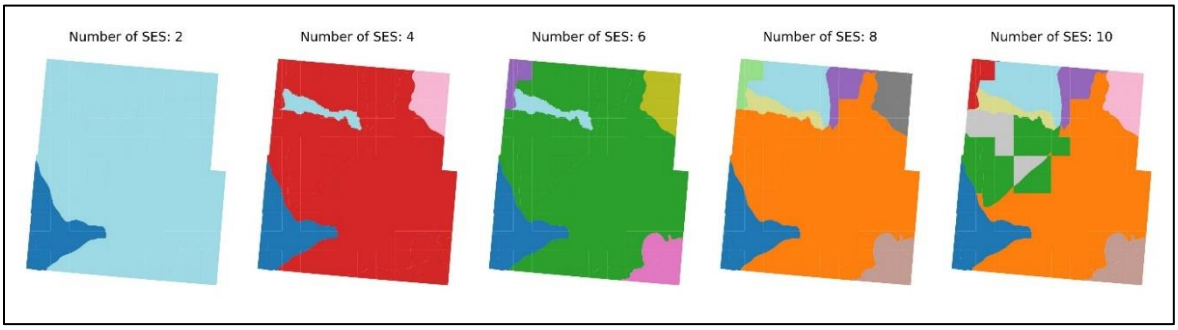

Figure 4.7: Different SES Units of Eddy County

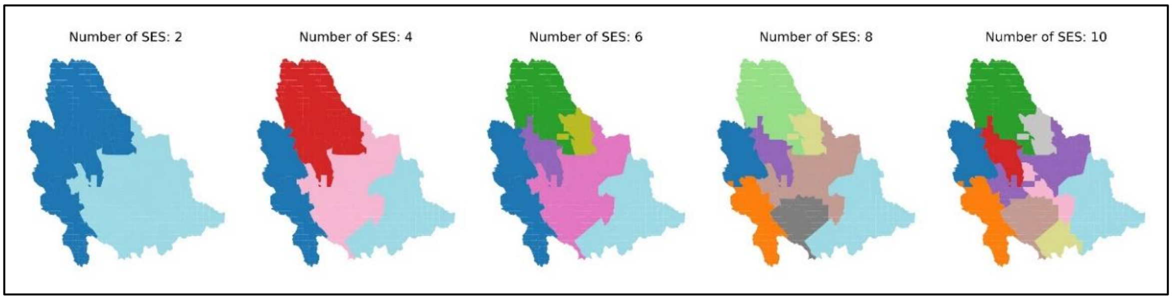

Figure 4.8: Different SES Units of Bravo/Sosa Sub-basin

A detailed examination of the SES unit delineation within the sample polygon of Mexico reveals that municipios attributes predominantly influence the delineation when a greater number of SES units are derived. Conversely, with fewer SES units, topography and LULC data become the primary factors guiding the delineation process. This indicates a shift in the driving factors of delineation depending on the granularity of the segmentation.

For the Eddy County, the delineation process exhibits a different pattern. When fewer SES units are delineated, the dominance of ecoregions is more apparent. However, as the number of SES units increases, LULC types and ecoregions become more significant in determining the boundaries. This demonstrates how the delineation criteria can vary significantly based on the number of regions being considered. Eddy County, which lacks a major city or significant population density, predominantly sees its SES delineation driven by ecological attributes. This observation underscores the variability in dominant factors influencing SES delineation, depending on the number of units and the specific regional characteristics.

4.4 Discussion

The delineation of SES unit boundaries is influenced by several critical factors, including the raster data extraction process, the choice of regionalization methods, and the fine-tuning of parameters within these selected methods. Each of these factors plays a pivotal role in ensuring that the delineation of SES units is both accurate and meaningful.

The raster data extraction process lays the groundwork for segmentation and subsequent analysis. This involves selecting appropriate raster datasets, such as elevation and LULC, and configuring parameters like resolution and grid size. These initial choices significantly impact the granularity and precision of the delineated units, as they determine the level of detail captured in the spatial data.

Following the extraction, the choice of dimension reduction and regionalization methods further shapes the delineation outcome. Both dimension reduction and regionalization methods necessitate meticulous parameter tuning to control the process effectively. Dimension reduction methods, such as UMAP, employ parameters like the number of nearest neighbors, minimum distance, and the number of output dimensions to manage complexity while preserving important data structures. These parameters aid in reducing high-dimensional data into a lower-dimensional space, retaining essential patterns necessary for accurate regionalization.

Regionalization methods involve various parameters to further refine the delineation process. The number of minimum features ensures each SES unit contains a sufficient number of data points to be statistically significant. Distance measures for quantifying pixel similarity are crucial for forming coherent regions by assessing the similarity between pixels. Edge smoothness helps create continuous boundaries, preventing jagged edges that could complicate analysis. Convergence criteria determine when the algorithm should stop iterating, ensuring computational efficiency and preventing overfitting. Various algorithms and techniques, such as the SKATER algorithm or the use of superpixels in image segmentation, offer different approaches to grouping spatial entities. Each method possesses inherent strengths and weaknesses, and the selection must align with the study's specific objectives and constraints. For example, some methods may prioritize geographic contiguity and compactness, which are crucial for creating cohesive and manageable units, while others focus on minimizing attribute dissimilarity within regions to ensure that SES units are as homogeneous as possible in terms of socio-environmental characteristics.

Careful adjustment of these parameters optimizes the delineation process, producing SES units that are both accurate and aligned with the study's specific goals and contextual requirements. The interplay between these factors underscores the complexity and necessity of a tailored approach in the effective delineation of socio-environmental units. This detailed and context-specific optimization ensures that the resulting SES units are robust, reliable, and useful for subsequent environmental and policy analyses.

Chapter 5: Conclusion

The exploration and analysis of complex SES necessitates the integration of diverse datasets and multiple modeling approaches. Ensuring the congruence of spatial scales across different data sources is crucial for accurate geospatial analyses. This thesis introduces a structured and reproducible methodology that integrates GIS, dimensionality reduction, and spatial optimization techniques to effectively delineate geospatial units, demonstrating its application through an expansive study area utilizing high-resolution datasets and diverse data models.

Spatial optimization emerges as a robust tool for delineating regions by accommodating diverse application requirements while managing large volumes of detailed input data. Preprocessing steps, including data extraction from raster datasets and vector-based overlays, ensure that the polygons aggregated in the regionalization model remain manageable. This process preserves the intrinsic properties and natural distributions of the data, facilitating the identification and characterization of socio-environmental units within a feasible timeframe.

The methodology relies exclusively on open-source Python modules and a publicly accessible geodatabase for the Rio Grande/Bravo Basin, promoting transparency and encouraging widespread adoption in SES analysis and modeling. This openness invites further refinement and expansion by the research community. Each application must customize the processing steps to align with specific research objectives.

Future research should refine parameter settings for segmentation and regionalization processes to enhance the accuracy and efficiency of SES delineation. Additionally, expanding the framework to incorporate a wider array of datasets and applying it to different regions will help validate its robustness and adaptability.

Alessa, L., Kliskey, A., & Brown, G. (2008). Social–Ecological Hotspots Mapping: A Spatial Approach for Identifying Coupled Social–Ecological Space. *Landscape and Urban Planning, 85*(1), 27-39. https://doi.org/https://doi.org/10.1016/j.landurbplan.2007.09.007

Alvanides, S., & Openshaw, S. (1999). *Zone Design for Planning and Policy Analysis*. Springer

Alvanides , S., Openshaw , S., & Rees , P. (2002). *Designing Your Own Geographies*. Wiley.

Aminpour, P., Gray, S. A., Jetter, A. J., Introne, J. E., Singer, A., & Arlinghaus, R. (2020). Wisdom of Stakeholder Crowds in Complex Social–Ecological Systems. *Nature Sustainability, 3*(3), 191-199. https://doi.org/10.1038/s41893-019-0467-z

An, L., Zvoleff, A., Liu, J., & Axinn, W. (2014). Agent-Based Modeling in Coupled Human and Natural Systems (Chans): Lessons from a Comparative Analysis. *Annals of the Association of American Geographers, 104*(4), 723-745. https://doi.org/10.1080/00045608.2014.910085

Andersson, E., McPhearson, T., Kremer, P., Gomez-Baggethun, E., Haase, D., Tuvendal, M., & Wurster, D. (2015). Scale and Context Dependence of Ecosystem Service Providing Units. *Ecosystem Services, 12*, 157-164. https://doi.org/https://doi.org/10.1016/j.ecoser.2014.08.001

Arnaiz-Schmitz, C., Díaz, P., Ruiz-Labourdette, D., Herrero-Jáuregui, C., Molina, M., Montes, C., Pineda, F. D., & Schmitz, M. F. (2018). Modelling of Socio-Ecological Connectivity. The Rural-Urban Network in the Surroundings of Madrid (Central Spain). *Urban Ecosystems, 21*(6), 1199-1212. https://doi.org/10.1007/s11252-018-0797-z

AssunÇão, R. M., Neves, M. C., Câmara, G., & Da Costa Freitas, C. (2006). Efficient Regionalization Techniques for Socio‐Economic Geographical Units Using Minimum Spanning Trees. *International Journal of Geographical Information Science, 20*(7), 797-811. https://doi.org/10.1080/13658810600665111

Barnes, M. L., Bodin, Ö., McClanahan, T. R., Kittinger, J. N., Hoey, A. S., Gaoue, O. G., & Graham, N. A. J. (2019). Social-Ecological Alignment and Ecological Conditions in Coral Reefs. *Nature Communications, 10*(1), 2039. https://doi.org/10.1038/s41467-019-09994-1

Berkes, F., & Folke, C. (1998). *Linking Social and Ecological Systems: Management Practices and Social Mechanisms for Building Resilience*. Cambridge University Press.

Bishop, M. P., Young, B. W., & Huo, D. (2018). Geomorphometry: Quantitative Land-Surface Analysis and Modeling☆. In *Reference Module in Earth Systems and Environmental Sciences*. Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.11469-1

Brunckhorst, D., Coop, P., & Reeve, I. (2006). 'Eco-Civic' Optimisation: A Nested Framework for Planning and Managing Landscapes. *Landscape and Urban Planning, 75*(3), 265-281. https://doi.org/https://doi.org/10.1016/j.landurbplan.2005.04.001

Castellarini, F., Siebe, C. T., Elena ; , de la Tejera, B., Cotler, H. P., Carlos Boege, E., Moreno, A. R., Saldívar, A., Larrazabal, A. G., Carla ; , Married, J. M., & Balvanera, P. (2014). A Social-Ecological Spatial Framework for Policy Design Towards Sustainability: Mexico as a Study Case. *Environmental research. Science and public policy, 6*(2), 45-59.

Castillo-Eguskitza, N., Martín-López, B., & Onaindia, M. (2018). A Comprehensive Assessment of Ecosystem Services: Integrating Supply, Demand and Interest in the Urdaibai Biosphere Reserve. *Ecological Indicators, 93*, 1176-1189. https://doi.org/https://doi.org/10.1016/j.ecolind.2018.06.004

Castro, A. J., Verburg, P. H., Martín-López, B., Garcia-Llorente, M., Cabello, J., Vaughn, C. C., & López, E. (2014). Ecosystem Service Trade-Offs from Supply to Social Demand: A Landscape-Scale Spatial Analysis. *Landscape and Urban Planning, 132*, 102-110. https://doi.org/https://doi.org/10.1016/j.landurbplan.2014.08.009

Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., & Fealy, R. (2013). Are Fine Resolution Digital Elevation Models Always the Best Choice in Digital Soil Mapping? *Geoderma, 195-196*, 111-121. https://doi.org/https://doi.org/10.1016/j.geoderma.2012.11.020

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends. *Neurocomputing, 408*, 189-215. https://doi.org/https://doi.org/10.1016/j.neucom.2019.10.118

Cetin, M. (2015). Using Gis Analysis to Assess Urban Green Space in Terms of Accessibility: Case Study in Kutahya. *International Journal of Sustainable Development & World Ecology, 22*(5), 420-424. https://doi.org/10.1080/13504509.2015.1061066

Cetin, M. (2016). Determination of Bioclimatic Comfort Areas in Landscape Planning: A Case Study of Cide Coastline. *Turkish Journal of Agriculture - Food Science and Technology, 4*(9), 800-804. https://doi.org/10.24925/turjaf.v4i9.800-804.872

Cetin, M. (2020). The Changing of Important Factors in the Landscape Planning Occur Due to Global Climate Change in Temperature, Rain and Climate Types: A Case Study of Mersin City. *Turkish Journal of Agriculture - Food Science and Technology, 8*, 2695-2701. https://doi.org/https://doi.org/10.24925/turjaf.v8i12.2695-2701.3891

Cetin, M., Agacsapan, B., Cabuk, S. N., Senyel Kurkcuoglu, M. A., Isik Pekkan, O., Baran Argun, E., Dabanlı, A., Kucukpehlivan, T., Yilmazel, B., & Cabuk, A. (2021). Assessment of the Ecological Footprint of Eskisehir Technical University–Iki Eylul Campus. *Journal of the Indian Society of Remote Sensing, 49*(10), 2311-2327. https://doi.org/10.1007/s12524-021-01395-6

Chavarria, S. B., & Gutzler, D. S. (2018). Observed Changes in Climate and Streamflow in the Upper Rio Grande Basin. *JAWRA Journal of the American Water Resources Association, 54*(3), 644-659. https://doi.org/https://doi.org/10.1111/1752-1688.12640

Choudhary, K., Boori, M. S., & Kupriyanov, A. (2018). Spatial Modelling for Natural and Environmental Vulnerability through Remote Sensing and Gis in Astrakhan, Russia. *The Egyptian Journal of Remote Sensing and Space Science, 21*(2), 139-147. https://doi.org/https://doi.org/10.1016/j.ejrs.2017.05.003

CNN. (2001). *Rio Grande Sucked Dry for Irrigation, Industry*. CNN. Retrieved May 15, 2024 from http://www.cnn.com/TRANSCRIPTS/0106/09/smn.08.html

Colding, J., & Barthel, S. (2019). Exploring the Social-Ecological Systems Discourse 20 Years Later. *Ecology and Society, 24*(1), Article 2. https://doi.org/10.5751/ES-10598-240102

Cumming, G. S., Cumming, D. H. M., & Redman, C. L. (2006). Scale Mismatches in Social-Ecological Systems: Causes, Consequences, and Solutions. *Ecology and Society, 11*(14). https://www.ecologyandsociety.org/vol11/iss1/art14/

DeMers, M. N. (2002). *Gis Modeling in Raster*. Wiley. https://cir.nii.ac.jp/crid/1130000797982204416

Deng, Z. Y., & Cao, J. S. (2023). Incorporating Ecosystem Services into Functional Zoning and Adaptive Management of Natural Protected Areas as Case Study of the Shennongjia Region in China. *Sci Rep, 13*(1), 18870. https://doi.org/10.1038/s41598-023-46182-0

Dettinger, M., Udall, B., & Georgakakos, A. (2015). Western Water and Climate Change. *Ecological Applications, 25*(8), 2069-2093. https://doi.org/https://doi.org/10.1890/15-0938.1

Di Minin, E., Soutullo, A., Bartesaghi, L., Rios, M., Szephegyi, M. N., & Moilanen, A. (2017). Integrating Biodiversity, Ecosystem Services and Socio-Economic Data to Identify Priority Areas and Landowners for Conservation Actions at the National Scale. *Biological Conservation, 206*, 56-64. https://doi.org/https://doi.org/10.1016/j.biocon.2016.11.037

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R. T., Molnár, Z., Hill, R., Chan, K. M. A., Baste, I. A., Brauman, K. A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P. W., van Oudenhoven, A. P. E., van der Plaat, F., Schröter, M., Lavorel, S., Aumeeruddy-Thomas, Y., Bukvareva, E., Davies, K., Demissew, S., Erpul, G., Failler, P., Guerra, C. A., Hewitt, C. L., Keune, H., Lindley, S., & Shirayama, Y. (2018). Assessing Nature's Contributions to People. *Science, 359*(6373), 270-272. https://doi.org/10.1126/science.aap8826

Duarte, G. T., Ribeiro, M. C., & Paglia, A. P. (2016). Ecosystem Services Modeling as a Tool for Defining Priority Areas for Conservation. *PLOS ONE, 11*(5), e0154573. https://doi.org/10.1371/journal.pone.0154573

Duque, J. C., Church, R. L., & Middleton, R. S. (2011). The P-Regions Problem. P-区域问题. *Geographical Analysis, 43*(1), 104-126. https://doi.org/https://doi.org/10.1111/j.1538-4632.2010.00810.x

Edwards, H. J., Elliott, I. A., Pressey, R. L., & Mumby, P. J. (2010). Incorporating Ontogenetic Dispersal, Ecological Processes and Conservation Zoning into Reserve Design. *Biological Conservation, 143*(2), 457-470. https://doi.org/https://doi.org/10.1016/j.biocon.2009.11.013

Egoh, B. N., Reyers, B., Rouget, M., & Richardson, D. M. (2011). Identifying Priority Areas for Ecosystem Service Management in South African Grasslands. *Journal of Environmental Management, 92*(6), 1642-1650. https://doi.org/https://doi.org/10.1016/j.jenvman.2011.01.019

Elias, E. H., Rango, A., Steele, C. M., Mejia, J. F., & Smith, R. (2015). Assessing Climate Change Impacts on Water Availability of Snowmelt-Dominated Basins of the Upper Rio Grande Basin. *Journal of Hydrology: Regional Studies, 3*, 525-546. https://doi.org/https://doi.org/10.1016/j.ejrh.2015.04.004

Ellis, E. C., & Ramankutty, N. (2008). Putting People in the Map: Anthropogenic Biomes of the World. *Frontiers in Ecology and the Environment, 6*(8), 439-447. https://doi.org/https://doi.org/10.1890/070062

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient Forests: Calculating Importance Gradients on Physical Predictors. *Ecology, 93*(1), 156-168. https://doi.org/https://doi.org/10.1890/11-0252.1

Elsawah, S., Filatova, T., Jakeman, A. J., Kettner, A. J., Zellner, M. L., Athanasiadis, I. N., Hamilton, S. H., Axtell, R. L., Brown, D. G., Gilligan, J. M., Janssen, M. A., Robinson, D. T., Rozenberg, J., Ullah, I. I. T., & Lade, S. J. (2020). Eight Grand Challenges in Socio-Environmental Systems Modeling. *Socio-Environmental Systems Modelling, 2*, 16226. https://doi.org/10.18174/sesmo.2020a16226

Faisal, K., & Shaker, A. (2017). An Investigation of Gis Overlay and Pca Techniques for Urban Environmental Quality Assessment: A Case Study in Toronto, Ontario, Canada. *Sustainability, 9*(3).

Feng, X., Barcelos, G., Gaboardi, J. D., Knaap, E., Wei, R., Wolf, L. J., Zhao, Q., & Rey, S. J. (2022). Spopt: A Python Package for Solving Spatial Optimization Problems in Pysal. *7*. https://doi.org/10.21105/joss.03330

Feng, X., & Koch, J. (2024). Combining Vector and Raster Data in Regionalization: A Unified Framework for Delineating Spatial Unit Boundaries for Socio-Environmental Systems Analyses. *International Journal of Applied Earth Observation and Geoinformation, 128*, 103745. https://doi.org/https://doi.org/10.1016/j.jag.2024.103745

Fischer, J., Gardner, T. A., Bennett, E. M., Balvanera, P., Biggs, R., Carpenter, S., Daw, T., Folke, C., Hill, R., Hughes, T. P., Luthe, T., Maass, M., Meacham, M., Norström, A. V., Peterson, G., Queiroz, C., Seppelt, R., Spierenburg, M., & Tenhunen, J. (2015). Advancing Sustainability through Mainstreaming a Social–Ecological Systems

Perspective. *Current Opinion in Environmental Sustainability, 14*, 144-149. https://doi.org/https://doi.org/10.1016/j.cosust.2015.06.002

Gao, Y., Feng, Z., Wang, Y., Liu, J.-L., Li, S.-C., & Zhu, Y.-K. (2014). Clustering Urban Multifunctional Landscapes Using the Self-Organizing Feature Map Neural Network Model. *Journal of Urban Planning and Development, 140*(2), 05014001. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000170

Gregorio, A. D. (2005). *Land Cover Classification System: Classification Concepts and User Manual*. Food and Agriculture Organization of the United Nations.

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). Sustainable Development Goals for People and Planet. *Nature, 495*(7441), 305-307. https://doi.org/10.1038/495305a

Grossmann, E., Ohmann, J., Kagan, J., May, H., & Gregory, M. (2010). Mapping Ecological Systems with Random Forest Model: Tradeoffs between Error and Bias. *GAP Analysis, 17*. https://www.fs.usda.gov/pnw/pubs/journals/pnw_2010_grossman001.pdf

Haining, R., Wise, S., & Ma, J. (2000). Designing and Implementing Software for Spatial Statistical Analysis in a Gis Environment. *Journal of Geographical Systems, 2*(3), 257-286. https://doi.org/10.1007/PL00011458

Hamann, M., Biggs, R., & Reyers, B. (2015). Mapping Social–Ecological Systems: Identifying 'Green-Loop' and 'Red-Loop' Dynamics Based on Characteristic Bundles of Ecosystem Service Use. *Global Environmental Change, 34*, 218-226. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2015.07.008

Hanspach, J., Loos, J., Dorresteijn, I., Abson, D. J., & Fischer, J. (2016). Characterizing Social–Ecological Units to Inform Biodiversity Conservation in Cultural Landscapes. *Diversity and Distributions, 22*(8), 853-864. https://doi.org/https://doi.org/10.1111/ddi.12449

Hengl, T., & Evans, I. S. (2009). Chapter 2 Mathematical and Digital Models of the Land Surface. In T. Hengl & H. I. Reuter (Eds.), *Developments in Soil Science* (Vol. 33, pp. 31-63). Elsevier. https://doi.org/https://doi.org/10.1016/S0166-2481(08)00002-0

Herrero-Jáuregui, C., Arnaiz-Schmitz, C., Reyes, M. F., Telesnicki, M., Agramonte, I., Easdale, M. H., Schmitz, M. F., Aguiar, M., Gómez-Sal, A., & Montes, C. (2018). What Do We Talk About When We Talk About Social-Ecological Systems? A Literature Review. *Sustainability, 10*(8).

Izquierdo, A. E., & Clark, M. L. (2012). Spatial Analysis of Conservation Priorities Based on Ecosystem Services in the Atlantic Forest Region of Misiones, Argentina. *Forests, 3*(3), 764-786.

Jones, K., Abrams, J., Belote, R. T., Beltrán, B. J., Brandt, J., Carter, N., Castro, A. J., Chaffin, B. C., Metcalf, A. L., Roesch-McNally, G., Wallen, K. E., & Williamson, M. A. (2019). The American West as a Social-Ecological Region: Drivers, Dynamics and Implications for Nested Social-Ecological Systems. *Environmental Research Letters, 14*(11), 115008. https://doi.org/10.1088/1748-9326/ab4562

Junge, B., Alabi, T., Sonder, K., Marcus, S., Abaidoo, R., Chikoye, D., & Stahr, K. (2010). Use of Remote Sensing and Gis for Improved Natural Resources Management: Case Study from Different Agroecological Zones of West Africa. *International Journal of Remote Sensing, 31*(23), 6115-6141. https://doi.org/10.1080/01431160903376415

Kanishka, G., & Eldho, T. I. (2017). Watershed Classification Using Isomap Technique and Hydrometeorological Attributes. *Journal of Hydrologic Engineering, 22*(10), 04017040. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001562

Kilicoglu, C., Cetin, M., Aricak, B., & Sevik, H. (2020). Site Selection by Using the Multi-Criteria Technique—a Case Study of Bafra, Turkey. *Environmental Monitoring and Assessment, 192*(9), 608. https://doi.org/10.1007/s10661-020-08562-1

Kilicoglu, C., Cetin, M., Aricak, B., & Sevik, H. (2021). Integrating Multicriteria Decision-Making Analysis for a Gis-Based Settlement Area in the District of Atakum, Samsun, Turkey. *Theoretical and Applied Climatology, 143*(1), 379-388. https://doi.org/10.1007/s00704-020-03439-2

Kim, S.-H., Cho, K.-J., Kim, T.-S., Lee, C.-S., Dhakal, T., & Jang, G.-S. (2023). Classifying Habitat Characteristics of Wetlands Using a Self-Organizing Map. *Ecological Informatics, 75*, 102048. https://doi.org/https://doi.org/10.1016/j.ecoinf.2023.102048

Klijn, F., & de Haes, H. A. U. (1994). A Hierarchical Approach to Ecosystems and Its Implications for Ecological Land Classification. *Landscape Ecology, 9*(2), 89-104. https://doi.org/10.1007/BF00124376

Koch, J., Friedman, J. R., Paladino, S., Plassin, S., & Spencer, K. (2021). Conceptual Modeling for Improved Understanding of the Rio Grande/Río Bravo Socio-Environmental System. *Socio-Environmental Systems Modelling, 1*, 16127. https://doi.org/10.18174/sesmo.2019a16127

Kumar, P., Fürst, C., & Joshi, P. K. (2021). Socio-Ecological Systems (Sess)—Identification and Spatial Mapping in the Central Himalaya. *Sustainability, 13*(14).

Kušová, D., Těšitel, J., Matějka, K., & Bartoš, M. (2008). Biosphere Reserves—an Attempt to Form Sustainable Landscapes: A Case Study of Three Biosphere Reserves in the Czech Republic. *Landscape and Urban Planning, 84*(1), 38-51. https://doi.org/https://doi.org/10.1016/j.landurbplan.2007.06.006

Lazzari, N., Becerro, M. A., Sanabria-Fernandez, J. A., & Martín-López, B. (2019). Spatial Characterization of Coastal Marine Social-Ecological Systems: Insights for Integrated Management. *Environmental Science & Policy, 92*, 56-65. https://doi.org/https://doi.org/10.1016/j.envsci.2018.11.003

Leslie, H. M., Basurto, X., Nenadovic, M., Sievanen, L., Cavanaugh, K. C., Cota-Nieto, J. J., Erisman, B. E., Finkbeiner, E., Hinojosa-Arango, G., Moreno-Báez, M., Nagavarapu, S., Reddy, S. M. W., Sánchez-Rodríguez, A., Siegel, K., Ulibarria-Valenzuela, J. J., Weaver, A. H., & Aburto-Oropeza, O. (2015). Operationalizing the Social-Ecological

Systems Framework to Assess Sustainability. *Proceedings of the National Academy of Sciences, 112*(19), 5979-5984. https://doi.org/10.1073/pnas.1414640112

Li, K., Hou, Y., Andersen, P. S., Xin, R., Rong, Y., & Skov-Petersen, H. (2022). An Ecological Perspective for Understanding Regional Integration Based on Ecosystem Service Budgets, Bundles, and Flows: A Case Study of the Jinan Metropolitan Area in China. *Journal of Environmental Management, 305*, 114371. https://doi.org/https://doi.org/10.1016/j.jenvman.2021.114371

Li, P., Li, X., & Yuan, G. (2023). Cross-Regional Allocation of Human Capital and Sustainable Development of China's Regional Economy—Based on the Perspective of Population Mobility. *Sustainability, 15*(12).

Li, W., Church, R. L., & Goodchild, M. F. (2014). The P-Compact-Regions Problem. *Geographical Analysis, 46*(3), 250-273. https://doi.org/https://doi.org/10.1111/gean.12038

Li, X., Cheng, G., Lin, H., Cai, X., Fang, M., Ge, Y., Hu, X., Chen, M., & Li, W. (2018). Watershed System Model: The Essentials to Model Complex Human-Nature System at the River Basin Scale. *Journal of Geophysical Research: Atmospheres, 123*(6), 3019-3034. https://doi.org/https://doi.org/10.1002/2017JD028154

Li, Z., Zhu, C., & Gold, C. (2005). *Digital Terrain Modeling: Principles and Methodology*. CRC Press. https://doi.org/https://doi.org/10.1201/9780203357132

Liu, H., Yang, J., Ye, M., James, S. C., Tang, Z., Dong, J., & Xing, T. (2021). Using T-Distributed Stochastic Neighbor Embedding (T-Sne) for Cluster Analysis and Spatial Zone Delineation of Groundwater Geochemistry Data. *Journal of Hydrology, 597*, 126146. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126146

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H., & Taylor, W. W. (2007). Complexity of Coupled Human and Natural Systems. *Science, 317*(5844), 1513-1516. https://doi.org/10.1126/science.1144004

Liu, Y., Zhang, H., Yang, X., Wang, Y., Wang, X., & Cai, Y. (2013). Identifying Priority Areas for the Conservation of Ecosystem Services Using Gis-Based Multicriteria Evaluation. *Polish Journal of Ecology, 61*.

Luck, G. W., Harrington, R., Harrison, P. A., Kremen, C., Berry, P. M., Bugter, R., Dawson, T. P., de Bello, F., Díaz, S., Feld, C. K., Haslett, J. R., Hering, D., Kontogianni, A., Lavorel, S., Rounsevell, M., Samways, M. J., Sandin, L., Settele, J., Sykes, M. T., van den Hove, S., Vandewalle, M., & Zobel, M. (2009). Quantifying the Contribution of Organisms to the Provision of Ecosystem Services. *BioScience, 59*(3), 223-235. https://doi.org/10.1525/bio.2009.59.3.7

Mandal, I., & Pal, S. (2022). Assessing the Impact of Ecological Insecurity on Ecosystem Service Value in Stone Quarrying and Crushing Dominated Areas. *Environment, Development and Sustainability, 24*(10), 11760-11784. https://doi.org/10.1007/s10668-021-01918-8

Martín-López, B., García-Llorente, M., Palomo, I., & Montes, C. (2011). The Conservation against Development Paradigm in Protected Areas: Valuation of Ecosystem Services in the Doñana Social–Ecological System (Southwestern Spain). *Ecological Economics, 70*(8), 1481-1491. https://doi.org/https://doi.org/10.1016/j.ecolecon.2011.03.009

Martín-López, B., Palomo, I., García-Llorente, M., Iniesta-Arandia, I., Castro, A. J., García Del Amo, D., Gómez-Baggethun, E., & Montes, C. (2017). Delineating Boundaries of Social-Ecological Systems for Landscape Planning: A Comprehensive Spatial Approach. *Land Use Policy, 66*, 90-104. https://doi.org/https://doi.org/10.1016/j.landusepol.2017.04.040

Martin, D. (1998). Optimizing Census Geography: The Separation of Collection and Output Geographies. *International Journal of Geographical Information Science, 12*(7), 673-685. https://doi.org/10.1080/136588198241590

Martínez-Harms, M. J., Quijas, S., Merenlender, A. M., & Balvanera, P. (2016). Enhancing Ecosystem Services Maps Combining Field and Environmental Data. *Ecosystem Services, 22*, 32-40. https://doi.org/https://doi.org/10.1016/j.ecoser.2016.09.007

McInnes, L., & Healy, J. (2018). Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*. https://doi.org/1802.03426

Morzillo, A. T., de Beurs, K. M., & Martin-Mikle, C. J. (2014). A Conceptual Framework to Evaluate Human-Wildlife Interactions within Coupled Human and Natural Systems. *Ecology and Society, 19*(3), Article 44. https://doi.org/10.5751/ES-06883-190344

Mueller, K. (2004). What's Your Position? Using Multi-Dimensional Scaling (Mds) and Geographical Information Systems (Gis) for Locating Ancient Settlements in the Meris of Polemon/Graeco-Roman Fayum. *50*(2), 199-214. https://doi.org/doi:10.1515/apf.2004.50.2.199 (Archiv für Papyrusforschung und verwandte Gebiete)

Naidoo, R., Balmford, A., Costanza, R., Fisher, B., Green, R. E., Lehner, B., Malcolm, T. R., & Ricketts, T. H. (2008). Global Mapping of Ecosystem Services and Conservation Priorities. *Proceedings of the National Academy of Sciences, 105*(28), 9495-9500. https://doi.org/10.1073/pnas.0707823105

Niu, H., & Silva Elisabete, A. (2020). Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods. *Journal of Urban Planning and Development, 146*(2), 04020007. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000566

Nizeyimana, E. (2012). *Remote Sensing and Gis Integration*. https://doi.org/https://doi.org/10.1201/9781351235860

North, M. A. (2009). *A Method for Implementing a Statistically Significant Number of Data Classes in the Jenks Algorithm* Proceedings of the 6th international conference on Fuzzy systems and knowledge discovery - Volume 1, Tianjin, China.

Omernik, J. M., & Bailey, R. G. (1997). Distinguishing between Watersheds and Ecoregions1. *JAWRA Journal of the American Water Resources Association, 33*(5), 935-949. https://doi.org/https://doi.org/10.1111/j.1752-1688.1997.tb04115.x

Openshaw, S., Alvanides, S., & Whalley, S. (1998). *Some Further Experiments with Designing Output Areas for the 2001 Uk Census.*

Osman, M. M., & Sevinc, H. (2019). Adaptation of Climate-Responsive Building Design Strategies and Resilience to Climate Change in the Hot/Arid Region of Khartoum, Sudan. *Sustainable Cities and Society, 47*, 101429. https://doi.org/https://doi.org/10.1016/j.scs.2019.101429

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science, 325*(5939), 419-422. https://doi.org/10.1126/science.1172133

Pascual, U., Palomo, I., Adams, W. M., Chan, K. M. A., Daw, T. M., Garmendia, E., Gómez-Baggethun, E., de Groot, R. S., Mace, G. M., Martín-López, B., & Phelps, J. (2017). Off-Stage Ecosystem Service Burdens: A Blind Spot for Global Sustainability. *Environmental Research Letters, 12*(7), 075001. https://doi.org/10.1088/1748-9326/aa7392

Peng, J., Hu, X., Qiu, S., Hu, Y. n., Meersmans, J., & Liu, Y. (2019). Multifunctional Landscapes Identification and Associated Development Zoning in Mountainous Area. *Science of The Total Environment, 660*, 765-775. https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.01.023

Pike, R. J., Evans, I. S., & Hengl, T. (2009). Chapter 1 Geomorphometry: A Brief Guide. In T. Hengl & H. I. Reuter (Eds.), *Developments in Soil Science* (Vol. 33, pp. 3-30). Elsevier. https://doi.org/https://doi.org/10.1016/S0166-2481(08)00001-9

Plassin, S., Koch, J., Paladino, S., Friedman, J. R., Spencer, K., & Vaché, K. B. (2020). A Socio-Environmental Geodatabase for Integrative Research in the Transboundary Rio Grande/Río Bravo Basin. *Scientific Data, 7*(1), 80. https://doi.org/10.1038/s41597-020-0410-1

Prato, T. (2009). Fuzzy Adaptive Management of Social and Ecological Carrying Capacities for Protected Areas. *Journal of Environmental Management, 90*(8), 2551-2557. https://doi.org/https://doi.org/10.1016/j.jenvman.2009.01.015

Price, S. J., Ford, J. R., Cooper, A. H., & Neal, C. (2011). Humans as Major Geological and Geomorphological Agents in the Anthropocene: The Significance of Artificial Ground in Great Britain. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369*(1938), 1056-1084. https://doi.org/10.1098/rsta.2010.0296

Qu, Y., & Lu, M. (2018). Identifying Conservation Priorities and Management Strategies Based on Ecosystem Services to Improve Urban Sustainability in Harbin, China. (2167-8359 (Print)).

Ramos, M. C., Barreto, J. O. M., Shimizu, H. E., Moraes, A. P. G. d., & Silva, E. N. d. (2020). Regionalization for Health Improvement: A Systematic Review. *PLOS ONE, 15*(12), e0244078. https://doi.org/10.1371/journal.pone.0244078

Rattan, S. S. P., & Hsieh, W. W. (2005). Complex-Valued Neural Networks for Nonlinear Complex Principal Component Analysis. *Neural Networks, 18*(1), 61-69. https://doi.org/https://doi.org/10.1016/j.neunet.2004.08.002

Reid, W. V., Mooney, H. A., Cropper, A., Capistrano, D., Carpenter, S. R., Chopra, K., Dasgupta, P., Dietz, T., Duraiappah, A. K., Hassan, R., Kasperson, R., Rik Leemans, May, R. M., McMichael, T. A. J., Pingali, P., Samper, C., Scholes, R., Robert T. Watson, Zakri, A. H., Shidong, Z., Ash, N. J., Bennett, E., Kumar, P., Lee, M. J., Raudsepp-Hearne, C., Simons, H., Thonell, J., & Zurek, M. B. (2005). *Ecosystems and Human Well-Being: Synthesis*.

Rodríguez-Morales, B., Roces-Díaz, J. V., Kelemen, E., Pataki, G., & Díaz-Varela, E. (2020). Perception of Ecosystem Services and Disservices on a Peri-Urban Communal Forest: Are Landowners' and Visitors' Perspectives Dissimilar? *Ecosystem Services, 43*, 101089. https://doi.org/https://doi.org/10.1016/j.ecoser.2020.101089

Ropero, R. F., Maldonado, A. D., Uusitalo, L., Salmerón, A., Rumí, R., & Aguilera, P. A. (2021). A Soft Clustering Approach to Detect Socio-Ecological Landscape Boundaries Using Bayesian Networks. *Agronomy, 11*(4).

Shi, Z., Zuo, R., Xiong, Y., Sun, S., & Zhou, B. (2023). Revealing Geochemical Patterns Associated with Mineralization Using T-Distributed Stochastic Neighbor Embedding and Random Forest. *Mathematical Geosciences, 55*(3), 321-344. https://doi.org/10.1007/s11004-022-10024-y

Shortt, N. K. (2009). Regionalization/Zoning Systems. In R. Kitchin & N. Thrift (Eds.), *International Encyclopedia of Human Geography* (pp. 298-301). Elsevier. https://doi.org/https://doi.org/10.1016/B978-008044910-4.00506-X

Sinare, H., Gordon, L. J., & Enfors Kautsky, E. (2016). Assessment of Ecosystem Services and Benefits in Village Landscapes – a Case Study from Burkina Faso. *Ecosystem Services, 21*, 141-152. https://doi.org/https://doi.org/10.1016/j.ecoser.2016.08.004

Song, S., Yu, H., Zhang, Q., Zhang, L., Zhong, Q., & Zhang, G. (2023). Functional Regionalization of Land Resources Considering Eco-Efficiency in Nanjing Metropolitan Area, China. *Ecological Indicators, 155*, 110964. https://doi.org/https://doi.org/10.1016/j.ecolind.2023.110964

Song, W., Wang, L., Liu, P., & Choo, K.-K. R. (2019). Improved T-Sne Based Manifold Dimensional Reduction for Remote Sensing Data Processing. *Multimedia Tools and Applications, 78*(4), 4311-4326. https://doi.org/10.1007/s11042-018-5715-0

Tachikawa, Y., Shiiba, M., & Takasao, T. (1994). Development of a Basin Geomorphic Information System Using a Tin-Dem Data Structure1. *JAWRA Journal of the American Water Resources Association, 30*(1), 9-17. https://doi.org/https://doi.org/10.1111/j.1752-1688.1994.tb03268.x

Thorn, J. P. R., Klein, J. A., Steger, C., Hopping, K. A., Capitani, C., Tucker, C. M., Reid, R. S., & Marchant, R. A. (2021). Scenario Archetypes Reveal Risks and Opportunities for Global Mountain Futures. *Global Environmental Change, 69*, 102291. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2021.102291

Václavík, T., Lautenbach, S., Kuemmerle, T., & Seppelt, R. (2013). Mapping Global Land System Archetypes. *Global Environmental Change, 23*(6), 1637-1647. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2013.09.004

Verburg, P. H., Dearing, J. A., Dyke, J. G., Leeuw, S. v. d., Seitzinger, S., Steffen, W., & Syvitski, J. (2016). Methods and Approaches to Modelling the Anthropocene. *Global Environmental Change, 39*, 328-340. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2015.08.007

Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, Adaptability and Transformability in Social–Ecological Systems. *Ecology and Society, 9*(2). http://www.jstor.org/stable/26267673

Wang, R., Zhou, T., Hu, D., Li, F., & Liu, J. (2011). Cultivating Eco-Sustainability: Social–Economic–Natural Complex Ecosystem Case Studies in China. *Ecological Complexity, 8*(4), 273-283. https://doi.org/https://doi.org/10.1016/j.ecocom.2011.03.003

Wang, Y., Li, Y., Zhang, B., & Xing, P. (2022). Dimensionality Reduction and Extraction of Engineering Remote Sensing Data Based on Building Information Modeling and Geographical Information System. *Scientific Programming, 2022*, 1319087. https://doi.org/10.1155/2022/1319087

Wang;, Y., Huang;, H., Rudin;, C., & Shaposhnik;, Y. (2021). Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering T-Sne, Umap, Trimap, and Pacmap for Data Visualization. *Journal of Machine Learning Research, 22*, 1-73.

Wei, R., Rey, S., & Knaap, E. (2021). Efficient Regionalization for Spatially Explicit Neighborhood Delineation. *International Journal of Geographical Information Science, 35*(1), 135-151. https://doi.org/10.1080/13658816.2020.1759806

Weng, Q. (2002). Land Use Change Analysis in the Zhujiang Delta of China Using Satellite Remote Sensing, Gis and Stochastic Modelling. *Journal of Environmental Management, 64*(3), 273-284. https://doi.org/https://doi.org/10.1006/jema.2001.0509

Wilkinson, B. H. (2005). Humans as Geologic Agents: A Deep-Time Perspective. *Geology, 33*(3), 161-164. https://doi.org/10.1130/G21108.1

Wilson, J. P. (2012). Digital Terrain Modeling. *Geomorphology, 137*(1), 107-121. https://doi.org/https://doi.org/10.1016/j.geomorph.2011.03.012

Wu, C. (1991). Research Core of Geography: The Human-Earth Areal Systems. *Economic Geography, 11*(3), 1-6.

Wu, J. (2013). Landscape Sustainability Science: Ecosystem Services and Human Well-Being in Changing Landscapes. *Landscape Ecology, 28*(6), 999-1023. https://doi.org/10.1007/s10980-013-9894-9

Xu, Y., Wang, L., Fu, C., & Kosmyna, T. (2017). A Fishnet-Constrained Land Use Mix Index Derived from Remotely Sensed Data. *Annals of GIS, 23*(4), 303-313. https://doi.org/10.1080/19475683.2017.1382570

Xue-Sen, Q., & Jing-Yuan, Y. (1990). Dai Ru-Wei. *A new discipline of science—the study of open complex giant system and its methodology. Chinese Journal of Nature, 13*(1), 3-10.

Yalcin, G. (2017). *Using of Remote Sensing/Gis Technology Onspecial Environmental Protection Areas in Turkeyand Sustainable Land Management*. OMICS International Limited.

Yang, X., Liu, S., Jia, C., Liu, Y., & Yu, C. (2021). Vulnerability Assessment and Management Planning for the Ecological Environment in Urban Wetlands. *Journal of Environmental Management, 298*, 113540. https://doi.org/https://doi.org/10.1016/j.jenvman.2021.113540

Yang, Y., Bao, W., & de Sherbinin, A. (2023). Mapping Fine-Resolution Nested Social-Ecological System Archetypes to Reveal Archetypical Human-Environmental Interactions. *Landscape and Urban Planning, 239*, 104863. https://doi.org/https://doi.org/10.1016/j.landurbplan.2023.104863

Zhang, Q.-f., Wu, F.-q., Wang, L., Yuan, L., & Zhao, L.-s. (2011). Application of Pca Integrated with Ca and Gis in Eco-Economic Regionalization of Chinese Loess Plateau. *Ecological Economics, 70*(6), 1051-1056. https://doi.org/https://doi.org/10.1016/j.ecolecon.2011.01.016