공학석사 학위논문

Littoral Cell Clustering of the East Sea Based on Coastal Sediment Transport Parameters Using Unsupervised Machine Learning

비지도 기계학습을 활용한 연안 유사이송 매개변수 기반 대한민국 표사계의 경계 설정

2025년 2월

서울대학교 대학원

건설환경공학부

김도현

Littoral Cell Clustering of the East Sea
Based on Coastal Sediment Transport Parameters
Using Unsupervised Machine Learning

비지도 기계학습을 활용한 연안 유사이송 매개변수
기반 대한민국 표사계의 경계 설정

지도교수 박 용 성

이 논문을 공학석사 학위논문으로 제출함

2025년 02월

서울대학교 대학원
건설환경공학부
김 도 현

김도현의 석사 학위논문을 인준함

2025년 02월

위 원 장 황 진 환 (인)

부위원장 박 용 성 (인)

위 원 양 수 현 (인)

ABSTRACT

Littoral Cell Clustering of the East Sea Based on Coastal Sediment Transport Parameters Using Unsupervised Machine Learning

By

Do Hyun Kim

Master in Civil and Environmental Engineering
Seoul National University

Advisor: Yong Sung Park

The coastline is increasingly vulnerable to environmental changes, such as sea level rise, global warming, and human activities, which exacerbate coastal erosion and pose significant social challenges. Due to the high utilization of coastal spaces and the growing risk of natural disasters, effective coastal management is crucial in

South Korea. The littoral cell approach offers a promising framework for managing sediment transport and mitigating erosion. Existing systems in Korea need more regional oceanographic characteristics and sediment transport mechanisms in demarcation, which can be an objective reason for divided results.

This study aims to cluster the boundaries of the East Sea littoral cells in South Korea using unsupervised machine learning techniques, focusing on coastal sediment transport parameters. A comprehensive dataset was constructed, standardized, and subjected to dimensionality reduction. Clustering methods were applied to identify spatially distinct littoral cells, and results were saved and visualized using GIS (Geographic Information System) data. The newly defined littoral cells are utilized to examine the feasibility of defining mesoscale littoral cells. They also serve as a basis to suggest a management scope for coastal erosion grade D, which represents the most severely eroded areas as reported in The Comprehensive Report on Coastal Erosion Status. This research will contribute to sustainable coastal management and adaptive strategies for addressing coastal erosion.

Key words: Coastal Erosion, Littoral Cell, Sediment Transport, Unsupervised Machine Learning, Coastal Management

Student Number: 2023-28970

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Latin Uppercase

$A, B, C, C_r, C_s, U, V, W, S, T$	Clustering structure
C_i	i^{th} subsets
K	Number of clusters
M	Eigenvector matrix
$M_{d\phi}$	Mean diameter
S_{all}	The set of all instances
$\{T_i\}$	Dendrogram
N_T	Total number of points within the new cluster
$\|U\|, \|V\|, \|S\|, \|T\|$	Number of points in clusters A and B

Latin Lowercase

a, b, i, j, x, y	Point in the cluster
$a_{sum(i)}, b_{sum(j)}$	Sum of points in true cluster I and predicted cluster j
$cov(X)$	Covariance matrix of the data X
c_k	Centroid of cluster k
d	Grain diameter
$d(C_i)$	Maximum distance within a cluster i
$d(C_i, C_j)$	Distance between the two closest points from different clusters i and j
$d(i, j)$	Euclidean distance between the i^{th} and j^{th} points
d_{50}	Grain diameter at the 50% of the cumulative distribution
n	The number of dimensions

$t_{den}(i,j)$	Dendrogramatic distance between the model points T_i and T_j
x_i, y_i	i[th] point in the cluster

Greek Lowercase

λ	Eigenvalue corresponding to M
ϕ_{16}	Grain diameter at the 16% of the cumulative distribution
ϕ_{50}	Grain diameter at the 50% of the cumulative distribution
ϕ_{84}	Grain diameter at the 84% of the cumulative distribution

LIST OF ABBREVIATIONS

Abbreviation	Description
DBSCAN	Density Based Spatial Clustering of Applications with Noise
GIS	Geographic Information System
HAC	Hierarchical Agglomerative Clustering
PCA	Principal Components Analysis
t-SNE	t-Distributed Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation Projection

1. Introduction

1.1 Background and necessities of study

The coastline is the land that meets the sea and is the first area to be impacted by oceanic processes. Korea has an extensive coastline with a high density of coastal space utilization. Increased human activities led to diverse and concentrated demands on coastal areas, including those for industry, ports, and residential complexes. Also, the coastal area plays a crucial role not only in helping humans maintain life with dwellings and economic resources but also in protecting the natural environment. However, environmental changes such as sea level rise, global warming, and human activities have made coastal regions susceptible to erosion.

In South Korea, coastal erosion sites are classified into grades based on the severity of erosion. As defined by the Ministry of Oceans and Fisheries, Grade D erosion sites represent the most critically eroded coastal areas requiring urgent intervention due to their high vulnerability to natural disasters and the potential for various unforeseen issues. This classification system helps prioritize areas for coastal management and restoration. Therefore, the preservation of the coastline is important, while the increasing risk of natural disasters along the coast can become a serious social issue. To prevent this from happening, understanding the coastal system is crucial. Although structures like groins and breakwaters are constructed to prevent shoreline erosion, these efforts sometimes fail and even accelerate shoreline changes because of the designs that do not fit well with the characteristics of the coastal

sediment transport system (Lim et al., 2021; Rahmawati et al., 2021).

Factors causing coastal erosion can be broadly categorized into three types: long-term background erosion, midterm redistribution erosion, and short-term episodic erosion (Lim et al., 2021). First, background erosion primarily occurs due to long-term changes in the sand budget of the sediment system (Foley et al., 2017). Second, redistribution erosion occurs due to mid-term changes in the wave environment at the breaking point, caused by alterations in wave direction or the construction of structures, leading to consistent coastal sediment transport and subsequent shoreline changes (Komar & Inman, 1970; Hsu & Evans, 1989; Kamphuis, 2003). Lastly, short-term episodic erosion is caused by cross-shore sediment transport due to high waves striking perpendicular to the shoreline, leading to coastal retreat. However, because this is a short-term change, the shoreline usually recovers to its original state (Wright et al., 1985; Miller & Dean, 2004; Yates et al., 2009; Lim et al., 2021).

The shoreline changes due to these complex factors can be managed more efficiently with littoral cell units. South Korea's current littoral cell system is divided into three scales: macroscale, mesoscale, and unit littoral cells, which are categorized based on their size and management purpose. Among these, mesoscale littoral cells are classified according to watersheds, while unit littoral cells are established along the coast that are detailed enough to allow for sediment budget analysis based on natural headlands or protruded rocky coastlines. However, these currently used systems did not consider the objective data from distinct marine characteristics, regional conditions, and sediment transport mechanisms of the East Sea

simultaneously. This limitation hinders the ability to explain hydrodynamic and sedimentary phenomena specific to each region and to identify the causes of erosion. Therefore, by considering both the coastal environment and sediment transport parameters together, the coastal sediment transport system will be better understood, and the cause of erosion will be more easily identified. Moreover, effective coastal management will be available.

Various studies have been conducted focusing on the mechanisms of coastal sediment dynamics and littoral cells (Table 1); however, few studies explored the relationship between coastal sediment transport parameters and littoral cells using unsupervised machine learning techniques. In north-central Oregon, the dynamics of littoral cell rotations were analyzed using a statistical downscaling framework and shoreline change model to assess the longshore sediment transport (Anderson et al., 2018). California's coastal headlands were classified using k-means clustering to improve coastal management by understanding headland dynamics and littoral cell boundaries (George et al., 2015). Reef (2023) employed hierarchical clustering to classify coastal cells based on sediment assemblages, revealing the sediment sharing among sub-embayment beaches. Southwestern Australia's coast was examined using cluster analysis and auto-correlation to define littoral cell boundaries and sediment compartmentalization (Sanderson & Eliot, 1999), while the impact of climate variability on the longshore sediment transport in southeast Queensland was modeled using the spectral wave analysis and regression models correlating climate indices with the sediment transport (Splinter et al., 2012).

Table 1. Literature review of coastal sediment dynamics and littoral cell

References	Research Object	Region
Sanderson & Eliot, 1999	Investigate the applicability of auto-correlation and cluster analysis for defining littoral cell boundaries	Australia
Anfuso et al., 2008	Analyze the morphodynamic behavior and short-term coastal trends by monitoring beach profiles to calculate volumetric budgets and group beaches into different erosive/accreting sectors	Spain
Romans et al., 2009	Evaluate the relative contributions and history of controls on sediment flux through a source-to-sink system	U.S.
Splinter et al., 2012	Understand the influence of large-scale climate variability on longshore sediment transport and resulting coastal evolution	Australia
George et al., 2015	Classify headlands along the coast to enhance the understanding of headland dynamics and littoral cell boundaries	U.S.
Sickmann et al., 2016	Understand the controls that sedimentary processes exert on detrital-zircon provenance	U.S.
Utizi et al., 2016	Assess the impacts of a mixed intervention (beach nourishment, feeder berm construction, and artificial reef placement) on coastal erosion and beach stability	Italy
Garzanti et al., 2018	Investigate the causes, modalities, and obstacles of sediment transfer along the longest documented cell of littoral sand drift	Nambia to Angola
Díez et al., 2018	Quantify and understand the intra-annual and interannual variations in the dry-beach profile shape	U.S.
Anderson et al., 2018	Investigate the dynamics resulting in persistent (multi-decadal) rotations of littoral cells Develop a framework to characterize historical wave conditions for understanding longshore sediment transport	U.S.
Antolínez et al., 2019	Develop and apply a hybrid shoreline change and foredune erosion model (COCOONED) to inform coastal planning and adaptation	U.S.
Avnaim-Katav, 2021	Explore the impact of the Nile River damming on the recent benthic ecology by analyzing the spatial distribution and diversity patterns	Israeli
Reef, 2023	Elucidate sediment dynamics and sharing among sub-embayment beaches using a coastal sediment cell framework combined with high-resolution satellite-derived rates of coastal change	Australia

1.2 Research objectives

Establishing the boundary of each littoral cell and understanding sediment transport systems are essential for effective coastal erosion management. Managing coasts by littoral cells offers several benefits, including ease of local management, varied and targeted research opportunities, increased efficiency in resource allocation, and enhanced stakeholders' cooperation. While international studies often incorporate sediment budget analyses that reflect the characteristics of each littoral cell, domestic research is lacking, particularly in incorporating physical and environmental properties into littoral cell classifications.

This study aims to cluster the unit littoral cells in the East Sea based on coastal sediment transport parameters through the application of unsupervised machine learning. The research focuses on South Korea's eastern coastal regions, including the Gangwon (Gangwon-do) and BUG (Busan-Ulsan-Gyeongsang-do) coastal areas. As shown in Figure 1, a database was built, standardized, and subjected to dimensionality reduction to analyze each various parameter across space. Dendrograms were drawn based on each maximum value of the cophenetic correlation coefficient to visualize hierarchical linkages between data. The optimal cluster number was determined by considering the silhouette index and used to divide the whole data point. Results were stored in GIS DB and illustrated on a map, indicating an assigned cluster of each littoral cell that reflects distinct oceanographic characteristics and sediment transport mechanisms. The new boundaries of the littoral cell were defined when an allocated cluster of each littoral cell was altered.

Additionally, the mapped results were compared with previously divided unit littoral cells. The newly defined littoral cells examine the feasibility of defining mesoscale littoral cells and suggest management scope for coastal erosion ‘Grade D’. This research will contribute to sustainable coastal management and adaptive strategies for addressing erosion.

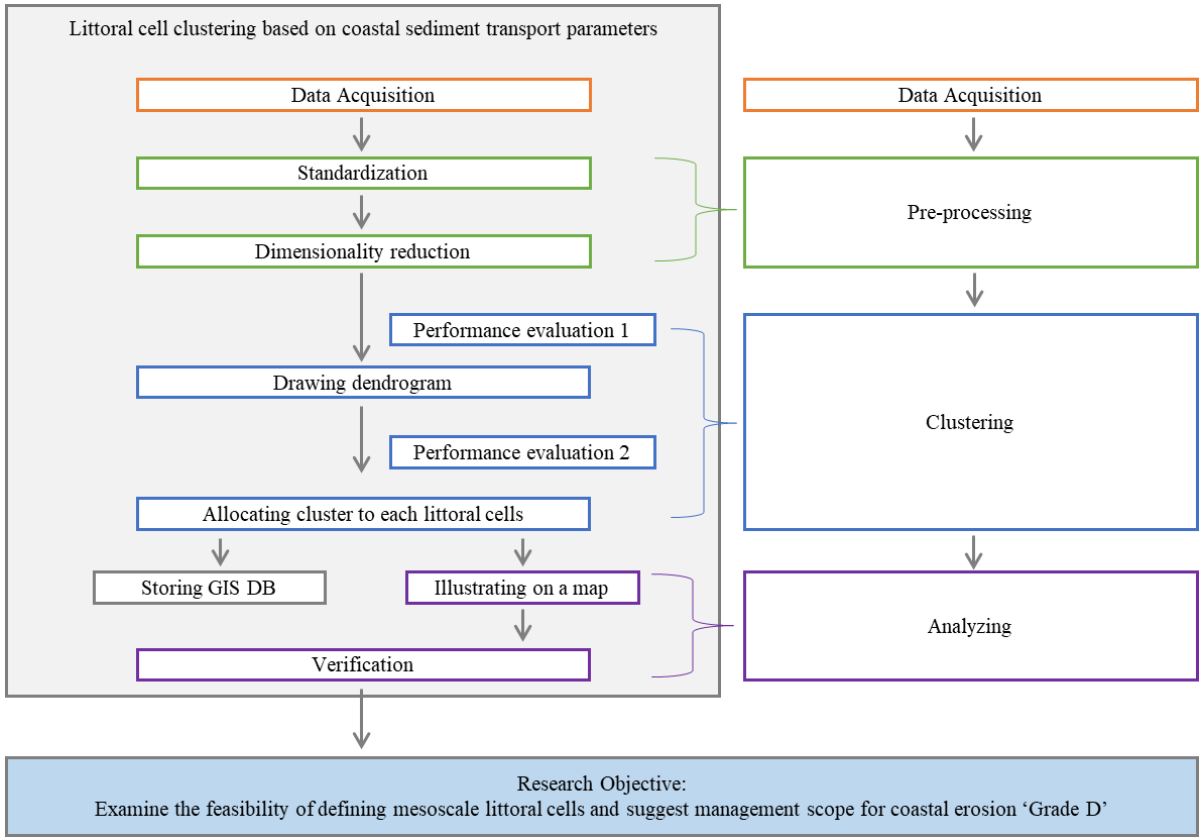


Figure 1. Flowchart of the research

2. Theoretical background

2.1 Sediment transport

2.1.1 Transport processes

Coastal features, such as beach profiles and planform shapes, are primarily shaped by the interaction of waves and nearshore currents. Waves suspend sediment and generate currents that transport these sediments alongshore or cross-shore. Longshore currents, induced by waves breaking obliquely to the shoreline, typically flow in the same direction as the waves and can occasionally turn offshore, forming rip currents that transport sediment away from the coast.

The movement of sediment along the coast, referred to as littoral drift, is measured as littoral transport or longshore sediment transport, commonly expressed in cubic meters per year. Although the direction of sediment transport can shift due to seasonal wave changes, most coastlines have a dominant direction of transport, known as downdrift, with the opposite referred to as updrift. Additionally, cross-shore transport, driven by wave- or wind-induced flows, plays a critical role in forming sandbars and altering beach profiles. These changes may occur gradually over the years or rapidly during storm events within hours.

2.1.2 Sediment transport

Sediment transport in coastal environments is a crucial process that shapes shorelines, with most coastlines featuring a predominant direction of sediment movement. This transport is influenced by various mechanisms and actions, primarily classified into three types: longshore sediment transport, cross-shore sediment transport, and aeolian sediment transport. Longshore sediment transport, known as littoral drift, is the movement of sediments such as sand and gravel along the shoreline due to various mechanisms, including bedload, suspended load, and swash load. Littoral drift is influenced by wave actions and longshore currents and plays a significant role in shaping coastal landscapes, influencing the formation and evolution of various coastal features like beaches, spits, and barrier islands.

In Longshore sediment transport, sediment along the coastline can vary in two directions, depending on the direction of the waves. At any given site, longshore transport comprises both positive and negative drift; the net drift represents the algebraic sum of these components, while the gross drift is the aggregate of the absolute values of the drift magnitudes. Understanding these dynamics is crucial for coastal management. While longshore sediment transport primarily occurs due to currents generated by waves, cross-shore transport is influenced by wave actions and undertow and is the movement perpendicular to the shoreline. Undertow is the oceanward movement of wave-driven mass transport, which carries suspended sediment out to sea and generates seaward-directed shear stresses that facilitate the transport of bedload offshore. Seasonal variations in shoreline are typically linked to

increased storm activity during the winter season, leading to enhanced seaward movement and deposition of sand in nearshore bars and other beach profiles. Elevated wave heights during storms cause these bars to form at greater offshore depths, increasing their overall size and necessitating a larger amount of sand, which is partly sourced from the erosion of the beach's subaerial profile.

Aeolian transport, the wind-driven movement of sand, also plays a critical role in coastal sediment dynamics. This process not only removes sand from the beach but also deposits it into dunes, which can either stay fixed or migrate significantly inland, sometimes threatening infrastructure and settlements. Typically, aeolian transport replenishes dunes with sand eroded during storms. Storms cause high water levels and vigorous waves that erode dunes and move sand to offshore bars. After a storm event, as normal wave conditions return, this sand is shifted back towards the land, forming a broad, dry berm ideal for wind transport. The prevalent landward sea breezes facilitate this movement. Strong onshore winds during storms help transport sand landward over the berm, where it is often captured by dune vegetation, promoting dune recovery. Aeolian processes typically leave behind a lag layer of coarser particles, having selectively removed finer sands from the berm.

Given the complexity of sediment transport dynamics, understanding the spatial and temporal variations in sediment movement requires physical observations and advanced analytical approaches. While effective in describing general transport processes, traditional methods often fall short in capturing patterns influenced by multiple interacting factors such as wave energy, current variability, and seasonal changes. Data-driven techniques such as clustering algorithms enable the

identification of hidden patterns within large datasets (Xu & Tian, 2015), providing new interpretations about sediment transport that might not be apparent through conventional analysis. Applying clustering algorithms makes it possible to classify and interpret sediment transport trends more effectively, supporting coastal management and erosion mitigation policies.

2.2 Littoral cell

2.2.1 Littoral cell

“A littoral cell is a coastal compartment that contains a complete cycle of sedimentation, including sources, transport paths, and sinks" (Inman, 2003). The cell boundaries sketch the geographical boundary within the balanced sediment budget that helps the quantitative analysis of coastal erosion and accretion. Thus, a littoral cell represents the smallest component of a cyclic mass system, affected by the input of sand from rivers and the erosion caused by wave activity, and is identified as an independent area with distinct information on the sediment's entry and exit.

Inman and Chamberlain introduced the concept at the International Geological Conference in 1960s(Inman & Frautschy, 1965; Chamberlain, 1968). In 1971, Inman and Nordstrom classified coasts based on tectonic settings and linked the littoral cell concept with geomorphological classifications, enhancing the understanding of coastal complexities. Inman and Jenkins (1984) highlighted how human interventions disrupt the natural balance of littoral cells, while Inman and Masters (1991) conducted a comprehensive analysis of sediment sources, transport mechanisms, and deposition areas, providing insights into effective coastal management strategies. Cooper and Pontee (2006) evaluate the impact of the sediment cell concept on coastal management, emphasizing its practicality through case studies of its application in England and Wales.

Bowen and Inman (1966) categorized the southern California coast into

various subareas, defining them as littoral cells based on the presence of headlands and the patterns of coastal sediment movement. Inman (2003) defined a littoral cell as a region experiencing a distinct cycle of sediment transportation. Motyka and Brampton (1993) suggested that the limits of a littoral cell are marked by headlands with contrasting directions of sediment transport and by submarine canyons serving as sediment traps. Meanwhile, Sanderson and Eliot (1999) developed a statistical approach to identify littoral cells using cluster analysis that considers the regional similarities in geomorphological characteristics, the physical attributes of seabed materials, and prevailing natural conditions like wave patterns.

2.2.2 Importance of littoral cell in coastal management

Sediments enter a littoral cell from various sources, such as riverine inputs, sedimentation, and erosion. The quantity and quality of sediments introduced into the cell significantly influence the coastal morphology. The balance between sediment deposition and erosion within a littoral cell determines the shape and stability of the coastline. Areas experiencing net deposition may see the growth of beach width and coastal landforms, while areas with net erosion may require intervention to prevent loss of land and infrastructure (Motyka & Brampton, 1993; Inman, 2003).

The notion of littoral cells assists in precisely understanding behaviors and quantitatively assessing current conditions, proving advantageous for engineering

purposes. Nonetheless, since sediment transport is just one aspect of the coastal system, focusing solely on littoral cells is inadequate for forecasting the long-term dynamics of larger coastal systems. Therefore, it is crucial to consider the interactions among all elements that form the coastal framework, and understanding interactions between sedimentation and erosion requires a comprehensive understanding of sediment sources, transport mechanisms, and deposition patterns, which is crucial for effective coastal management.

2.2.3 Littoral cell system in Korea

In Korea, littoral cells are classified into macro-scale littoral cells, meso-scale littoral cells, and unit littoral cells, depending on their size and management objectives. Nine macro-scale littoral cells are divided based on the jurisdictional boundaries of local governments: three in the East Sea, three in the West Sea, and three in the South Sea and Jeju. Specifically, these are Gangwon-do, Gyeongsangbuk-do, and the East Sea region of Gyeongsangnam-do, which are along the East Sea; Incheon and Gyeonggi-do including Anmyeon island, Chungcheongnam-do below Anmyeon island centered around the Geum River estuary, and the West Sea region of Jeollanam-do, which are along the West Sea; and the South Sea region of Jeollanam-do, the South Sea region of Gyeongsangnam-do, and Jeju Island which are along the South Sea. Unit littoral cells are divided in detail along the coast based on natural headlands or protruding rocky coastlines to enable

detailed sediment budget analysis.

2.2.4 Littoral cell system in other nations

Region-specific sediment management based on littoral cell systems has been adopted in coastal nations, such as the U.S., U.K., and Japan, where coastal erosion is a significant concern. In the U.S., According to the Virginia Institute of Marine Science and Bureau of Ocean Energy Management, Southern California, the Chesapeake Bay in Virginia, and San Francisco have addressed coastal erosion with littoral cell management since the 1970s. In the U.K., The Shoreline Management Plan (SMP) was established from 1996 to 1999, dividing the coasts of England and Wales into 11 major littoral cells. Each major littoral cell contains 3 to 7 sub-cells (DEFRA, 2006). Japan has implemented a comprehensive and large-scale approach through Sato-umi, a comprehensive approach to integrated coastal management that emphasizes the harmonious coexistence of humans and nature since 2000 (Ministry of the Environment, 2010).

2.3 Factors related to sediment transport

2.3.1 Sediment size

Sediment size of surficial sediments differs based on the sampling location. These spatial variations and trends in grain size are influenced by sediment transport processes, including abrasion, selective transport, and the mixing of sediments from various sources (Russell, 1939). This suggests that specific grain size trends are likely to be related to sediment transport pathways (Gao et al., 1994). Geologists have created size classifications to define what constitutes sand, gravel, and other sediment types. One widely used classification system is the Wentworth scale, which categorizes sediment sizes in millimeters based on powers of two. Krumbein (1936) introduced the phi (ϕ) scale as an alternative size measure. The phi size is related to the grain size by the formula $\phi = -\log_2 d$, which is $2^{-\phi} = d$, where d is the grain diameter in millimeters. The phi scale is extensively used, especially in coastal geology, because it provides a convenient way to represent grain size distributions.

A common measure of grain samples is d_{50} (or φ_{50}), which represents the median grain size. This size can be directly obtained from the cumulative distribution curve, as it indicates the size at which half of the sample's weight is coarser and the other half is finer. Consequently, the phi sizes φ_{84} and φ_{16} correspond to $\varphi_{(50\pm\frac{68}{2})}$. Otto (1939) and Inman (1952) suggested that the mean diameter be defined as

$$M_{d\phi} = \frac{(\phi_{84} + \phi_{16})}{2}$$

Other methods for determining the mean diameter have been proposed. Folk and Ward (1957), who analyzed sand samples with both large and small sizes, suggested the following measurement for bimodal distributions:

$$M_{d\phi} = \frac{(\phi_{84} + \phi_{50} + \phi_{16})}{3}$$

These two definitions differ little for distributions that approximate a log-normal distribution. The mean and median sizes are identical in the case of sand with a symmetrical size distribution.

Offshore sands are generally finer than those found in the nearshore region, which is more dynamic due to the impact of shoaling and breaking waves. Variations along the beach can result from various nonuniform processes. Differences in wave energy affecting the beach can lead to variations in sand size and beach face slope. Bascom (1951) demonstrated that sand diameter decreases with reduced exposure to waves and that beach face slope decreases as sand size diminishes.

2.3.2 Constructive and destructive forces acting on beach

The beach profile is the variation of water depth with distance offshore from the shoreline. The equilibrium profile is a conceptual result of the balance of destructive versus constructive forces (Dean, 1991). The equilibrium profile is considered a dynamic concept in nature, as the incident wave field and water level change continuously in nature; therefore, the profile responds continuously. A mean

equilibrium can be defined by averaging these profiles over a long period (Lanzoni & Seminara, 2002). An equilibrium beach profile represents a balance of destructive and constructive forces acting on the beach (Dean, 1991; Dean & Dalrymple, 2004). Many different destructive and constructive forces affect beach profiles.

Beaches with gentle slopes experience lower turbulence because wave breaking is distributed across a wider area of the surf zone. In contrast, steep-sloped beaches release concentrated wave energy over a smaller area, resulting in more substantial turbulent fluctuations that penetrate deeper into the water column. Different sediment types' ability to withstand varying levels of turbulent energy impacts beach stability, which results in fine sediments typically found on milder slopes and coarse sediments on steeper slopes. Gravity is a persistent destructive force, continuously working to even out irregular profiles. Additionally, high turbulence in the surf zone, where breaking waves transform structured wave energy into turbulent fluctuations, is a significant destructive element. These fluctuations dislodge sediment particles, helping gravity move them seaward (Dean, 1991; Dean, 1995; Dean & Dalrymple, 2004).

Three distinct constructive forces influence the formation of beach profiles. One is the average streaming velocities at the seabed, which result from energy dissipation within the bottom boundary layer and cause local momentum transfer. Another force comes from the net onshore shear stresses at the seabed caused by shallow-water waves. Lastly, the intermittent suspension and selective transport of particles driven by shoreward velocities contribute to constructive forces. If the fall time of the particles is less than half a wave period, there is a net onshore sediment

transport. However, if the fall time is greater than half but less than an entire wave period, it results in net offshore sediment transport (Dean & Galvin, 1976; Dean, 1991; Dean, 1995).

2.3.3 Coastal landforms: Spit, Tombolo, inlet

Coastlines exhibit remarkable diversity in their forms and characteristics, shaped by the local environments and geological processes that influence the formation and evolution of coastal landforms. Classifying shorelines is beneficial because it reveals consistent patterns in how different shorelines react to natural forces (Stive et al., 2002). Consequently, analyzing shorelines based on their classification helps comprehension of how they respond to various environmental pressures. This approach forms a fundamental part of understanding the dynamics of shoreline changes.

A spit is a subaerial depositional feature formed by waves and currents transporting sediment from sources like eroding headlands into an elongated shape parallel to the shoreline. Its growth depends on sediment supply and water depth at the terminal end. A faster sediment supply promotes spit growth, while a slower supply may lead to bayside or bayhead beach formation. Spits can develop into various shapes, including recurved spits, which form when wave directions change, or complex spits, where protected sections curve further back. Elevation differences on spits, influenced by wave run-up, can create features like beach ridges or broader

formations like beach planes (Dean & Dalrymple, 2004).

Tidal currents at inlets often limit spit growth. Strong currents may balance sediment deposition, forming equilibrium inlets or preventing inlet closure. If currents are weak, the inlet may close, forming bay mouth barriers. Other depositional features include midbay barriers along embayment shorelines and unique formations like winged headlands or lagoons around isolated hills submerged by rising sea levels. These processes highlight the dynamic interaction of sediment transport, wave energy, and tidal currents in shaping coastal landforms.

A tombolo is a depositional feature that forms a subaerial connection between an offshore island and the shoreline. This occurs due to wave sheltering by the island, which modifies wave crests to curve inward toward the sheltered area behind the island. Depending on sediment availability, a tombolo can grow either from the shoreline toward the island or vice versa. In some cases, elongated islands close to the shore can lead to the formation of double tombolos, and even triple tombolos are known to exist. If the offshore island is too far from the coastline, a salient (a shoreline protrusion) may form instead of a tombolo (Dean & Dalrymple, 2004).

Tidal inlet morphology varies due to interactions between sediment supply, wave characteristics, and tidal flow. Inlets typically migrate downdrift with longshore sediment transport as sediment accumulates on the updrift side, narrowing the channel and increasing current velocity, which erodes the downdrift side. Occasionally, deposition on the updrift side overlaps the downdrift shoreline, causing inefficiencies and potential breakthroughs during storms, and it forms barrier

islands or modifies spit features. Offsets between updrift and downdrift shorelines are common, with ebb-tidal shoals playing a key role. These shoals trap sediment, reduce wave energy, and promote downdrift shoreline growth.

2.4 Dimensionality reduction

Dimensionality reduction techniques convert a dataset X with a dimensionality of D into a new dataset Y with a dimensionality of W, while trying to preserve the original geometry of the data as much as possible (Van der Maaten et al., 2009). As can be seen in Figure 2, the dimensionality reduction method is divided into two methods: linear methods and nonlinear methods. Linear methods reduce dimensionality by embedding the data into a lower-dimensional linear subspace. Principal Components Analysis (PCA) is one of the most commonly used linear dimensionality reduction techniques. It utilizes an orthogonal transformation to project the data onto a low-dimensional space while preserving as much variance as possible (Van der Maaten et al., 2009; Reddy et al., 2020).

Nonlinear methods are often more effective than linear ones, as the relationship between the latent and observed variables can be more complex than just a simple matrix multiplication. However, these models usually involve numerous parameters, which require a large dataset for accurate identification (Lee & Verleysen, 2007). Nonlinear dimensionality reduction techniques can be classified into three main categories: (1) methods that preserve local structures of the original data in the low-dimensional representation, (2) methods that focus on capturing global structures, and (3) methods that achieve global alignment through a combination of linear models (Van Der Maaten et al, 2009).

Coastal sediment datasets often have multiple correlated variables,

reflecting the complex interactions between environmental factors such as wave energy, current patterns, and sediment properties. This high dimensionality can disturb data analysis by introducing noise, redundancy, and increased computational demands, which may degrade the quality of outcomes and reduce the efficiency of clustering algorithms. Therefore, dimensionality reduction techniques are crucial in enhancing the performance of subsequent clustering analyses by reducing noise. In this section, Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation Projection (UMAP) will be discussed, which are widely considered representatives of dimensionality reduction.

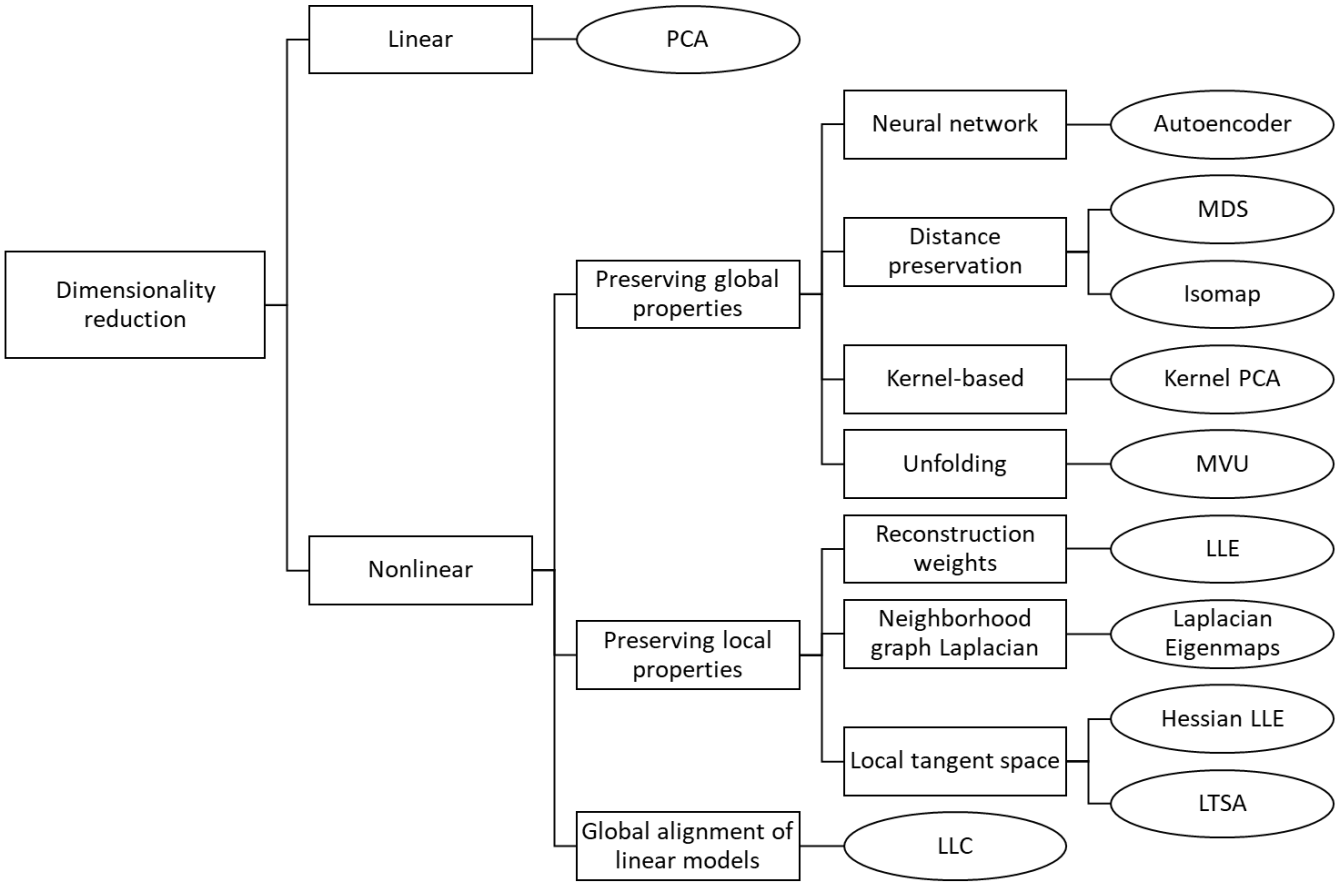


Figure 2. Taxonomy of dimensionality reduction

2.4.1 Principal Component Analysis (PCA)

The most widely recognized linear dimensionality reduction technique is PCA (Principal Component Analysis), which creates a low-dimensional representation of data by identifying a linear basis of reduced dimensionality that maximizes variance. PCA addresses this by solving the eigenvalue problem that is

$$cov(X)M = \lambda M$$

to determine a linear mapping M that maximizes $M^T cov(X)M$, where $cov(X)$ represents the covariance matrix of the data X, M is eigenvector matrix, and λ is the eigenvalue corresponding to M. PCA has been widely applied in various fields. However, as the dimensionality of the data points increases, the covariance matrix grows larger, making computation difficult for very high-dimensional datasets. Iterative methods such as Simple PCA or probabilistic PCA can be employed as alternatives (Van Der Maaten et al., 2009; Hasan & Abdulazeez, 2021)

2.4.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)

Stochastic Neighborhood Embedding (SNE), introduced by Hinton and Roweis (2002), is a well-known non-linear technique. Despite producing reasonably good visualizations, SNE is limited by a cost function that is difficult to optimize and a crowding issue. t-SNE addresses these challenges by employing a symmetrized version of the SNE cost function with simpler gradients (Cook et al., 2007) and by

using a fatter-tailed student-t distribution with one degree of freedom instead of a Gaussian for similarity calculation between two points in the low-dimensional space (Pal & Sharma, 2020; Wang et al., 2021; Ghojogh et al., 2020) and overcome the crowding problem (van der Maaten & Hinton, 2008; Sainburg et al., 2021).

t-SNE, a variant of Stochastic Neighbor Embedding, is much easier to optimize as it reduces the tendency to cluster points in the center of the map, resulting in significantly improved visualizations. Typically, t-SNE employs random walks on neighborhood graphs to visualize the structure of large datasets. This approach ensures that the inherent structure of the entire dataset influences how the subset of data is displayed (Pal & Sharma, 2020).

2.4.3 Uniform Manifold Approximation Projection (UMAP)

UMAP is another non-linear dimensionality reduction technique used for visualization. It is based on a theoretical framework grounded in Riemannian geometry and algebraic topology. A Riemannian manifold is a mathematical concept in Riemannian geometry, representing a smooth and curved space where curvature, distance, and angles can be defined at each point. It extends the notion of Euclidean space, which is a flat space, enabling geometric computations on curved surfaces and higher-dimensional curved spaces. There are three assumptions about the data: (1) The data is uniformly distributed on Riemannian manifold, (2) The Riemannian metric is locally constant, and (3) The manifold is locally connected (McInnes et al.,

2018; Trozzi et al., 2021; UMAP-learn, n.d.). UMAP offers several useful features, including the capability to project new data into an existing embedding space and the ability to utilize labels for dimensionality reduction (Trozzi et al., 2021). Biology, Medicine, and Virology use UMAP to visualize cellular development processes, analyze mutation data, and assist clustering (Packer et al., 2019; Hurley et al., 2019; Cao et al., 2020; Hozumi et al., 2020)

The UMAP algorithm considers a few hyper-parameters: n-epochs, n, min-dist, and d. n-epochs refers to the number of training epochs, a training cycles used for optimization in the low-dimensional subspace. n is the number of neighbors considered while approximating the local metric. Lower values of n focus more on local structure, while higher values capture more global structure. min-dist represents the required separation in the embedding space between close points. Lower min-dist values create more distinct clusters, while higher values result in less distinct clusters. d is the target embedding dimension, which is generally set to 2 or 3 for visualization (Pal & Sharma, 2020).

2.4.4 Comparison between PCA, t-SNE, and UMAP

t-SNE and UMAP are types of manifold learning, while PCA is not. Manifold learning assumes that there is a subspace that encompasses high-dimensional data well when it is represented in data space. With adequate manifold, a space that can effectively represent high-dimensional data in lower dimensions

(Han et al., 2022). This means t-SNE and UMAP are better suited than PCA. UMAP and t-SNE operate on similar concepts and are both commonly used dimensionality reduction techniques for visualization. However, several points make UMAP superior to t-SNE: (1) It better preserves the global structure in high-dimensional topological space by capturing both local structure and global structure between distinct clusters, while t-SNE preserves only local structure, (2) It can also be utilized for model training, including supervised learning. In particular, Parametric UMAP leverages neural networks to learn data embeddings, and (3) It offers better runtime performance and no computational limitations regarding embedding dimensions. (McInnes et al., 2018; Pal & Sharma, 2020; Sainburg et al., 2021; Wang et al., 2021).

2.5 Clustering algorithm

2.5.1 Clustering algorithms

Clustering plays a vital role in understanding data structure and discovering patterns. Clustering organizes data instances into subsets such that similar instances are grouped together, while dissimilar instances are placed in separate groups. This process arranges the instances into an efficient representation that reflects the sampled population. The clustering structure is defined as a set of subsets $C = C_1, \dots, C_k$ of S_{all}, such that $S_{all} = U_{i=1}^{k} C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. As a result, each instance in S belongs to precisely one subset (Reddy et al., 2020).

Figure 3 is a taxonomy of clustering algorithms, each employing different approaches to categorize data based on its properties and goals. These have been classified into two categories: the Hierarchical Clustering Algorithm and the Partitional Clustering Algorithm. In hierarchical clustering algorithms, data are divided into levels in a hierarchical format and clusters are designed to generate a top-down or bottom-up approach (Saxena et al., 2017). In a partitional clustering algorithm, data is arranged in a series of nested groups that do not follow any hierarchical structure. (Jain and Dubes, 1988; Ezugwu et al., 2022)

Table 2 describes hierarchical and partitional clustering methods. Hierarchical clustering is divided into agglomerative and divisive methods. In agglomerative clustering, clusters are merged into larger clusters, while in divisive clustering, every cluster is divided into smaller levels. Partitional clustering includes hard/crisp, mixture resolving, and the Fuzzy methods. In hard/crisp clustering, data

belongs to only one cluster, while Fuzzy method assigns data to two or more clusters based on fuzzy sets (Ezugwu et al., 2022). Mixture resolving clusters objects based on probabilities from predefined probabilistic clusters.

The most commonly used clustering methods are hierarchical clustering, DBSCAN, mean shift clustering, and k-means clustering. Table 3 explains each method's description, advantages, and disadvantages. Hierarchical clustering organizes data points into a hierarchical structure specialized in understanding the hierarchical relationship between data points. It can visualize cluster structures using dendrograms and is widely used in bioinformatics and document clustering. DBSCAN categorizes data points based on density and is frequently employed in fields related to density, such as geographic information systems and anomaly detection. Mean shift clustering forms clusters by shifting data points toward higher-density areas. It is known for its simplicity and stability in computation. It is mainly used in real-time processing tasks like video tracking. K-means clustering group data points into clusters by minimizing the distance between data points with centroid. It is relatively simple to implement compared to other clustering methods, making it widely used in various fields such as image segmentation, computer vision, geostatistics, astronomy, and agriculture. The clustering method should be chosen appropriately based on the characteristics of the dataset and research objective to achieve the best results.

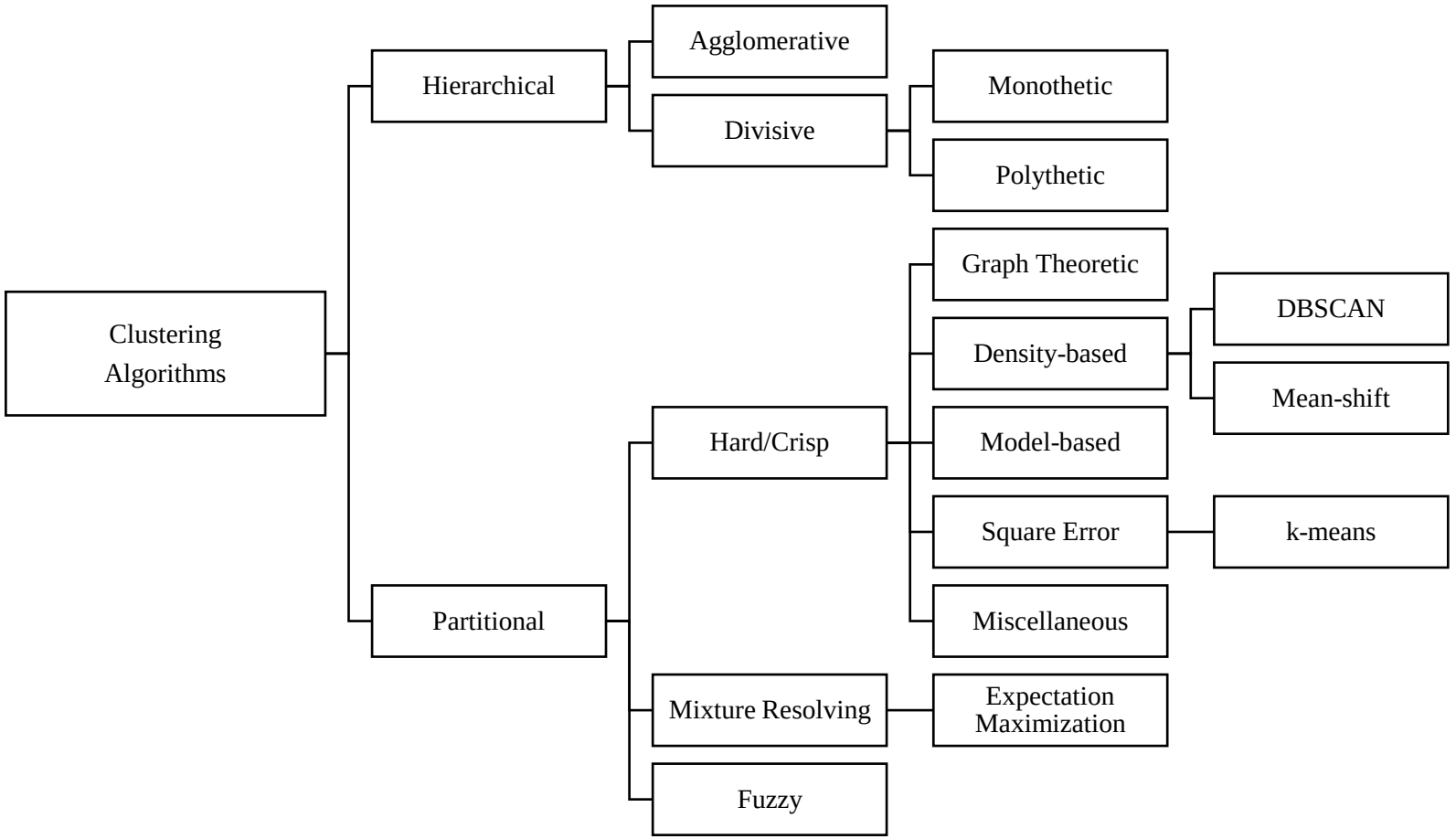


Figure 3. A taxonomy of clustering algorithms (Ezugwu et al., 2022)

Table 2. Hierarchical (Agglomerative, Divisive) and Partitional (Hard/Crisp, Fuzzy, and Mixture Resolving) clustering methods and descriptions

Clustering Method		Description
Hierarchical		
	Agglomerative	Merge into larger clusters
	Divisive	Divide into smaller level
Partitional		
	Hard/Crisp	Assigned to only one cluster
	Fuzzy	Assigned to two or more clusters, Defined in fuzzy sets
	Mixture Resolving	Clustered based on probabilities from predefined probabilistic clusters

Table 3. Description, advantages, and disadvantages of Clustering Algorithms (Hierarchical, DBSCAN, Mean-shift, and k-means)

Clustering Algorithm	Description	Advantages	Disadvantages
Hierarchical	Organizes data points into a hierarchical structure	No need to pre-define the number of clusters Visualize the relationship between clusters Versatile and applicable to various data types Robust to outliers	Computationally expensive Challenging at large datasets
DBSCAN	Categorize data points based on density	No need to pre-define the number of clusters No limitation on the size or shape of the data	Difficult to find parameters with high-density data
Mean-shift	Form cluster by shifting data points toward areas of higher density	No need to pre-define the number of clusters	Slower than DBSCAN and k-means Unsmooth behavior of the kernel density
k-means	Group into clusters by minimizing the distance between data points with centroid	Fast and simple Suitable for large datasets	Require a prior number of clusters(K) Results can vary depending on the initial centroid selection

2.5.2 Hierarchical clustering algorithm

Hierarchical clustering is a cluster analysis method that shows a hierarchy between clusters. This method is divided into two categories: (1) Agglomerative, a bottom-up approach where each observation starts in its own cluster and merge pairs of clusters as one moves up the hierarchy (Seid & Mulatu Mengesha, 2022) and (2) Divisive, a top-down approach where all observations start in one cluster and then splits are formed recursively as one moves down the hierarchy. Generally, the merges and splits are determined in a greedy manner, which means the algorithm decides which cluster to merge and split at each stage of the process (Nielsen, 2016; Murtagh & Contreras, 2012). The results of hierarchical clustering are represented in a dendrogram (Nielsen, 2016; Murtagh & Contreras, 2012; Jain et al., 1999), which illustrates the nested grouping of objects and the similarity levels at which these groupings change. Clustering of the data is achieved by cutting the dendrogram at the chosen similarity level. The merging or splitting of clusters is performed with similarity measurement.

Maimon and Rokach (2005) stated that hierarchical methods generally possess two strengths and two weaknesses. The first strength is versatility. For example, single-link methods perform well on datasets containing non-isotropic clusters, such as well-separated, chain-like, and concentric clusters. The second strength is multiple partitions. Hierarchical methods produce multiple nested partitions rather than a single partition, enabling different users to select similarity levels according to their desire. However, hierarchical methods also have

weaknesses. The first one is an inability to scale well. The time complexity of hierarchical algorithms is at least $O(m^2)$, where m is the total number of objects. Clustering a large number of objects using a hierarchical algorithm incurs substantial costs. In addition, it is lack of reversibility. Hierarchical methods cannot undo previous steps, meaning there is no backtracking capability.

Figure 4 is an example result of hierarchical agglomerative clustering, where a sequence of irreversible algorithmic steps constructs a structured dendrogram from n objects. Each step involves merging pairs of clusters into progressively larger groups. This process results in two equivalent representations. One is a series of n-1 partitions evolving from individual data to collective groups in a dendrogram that offers a binary tree. The other is an ultrametric topology, which emphasizes the hierarchical relationships within the data. These frameworks are crucial for analysis, as they enable a detailed examination of the data's intrinsic hierarchical organization (Murtagh & Contreras, 2012).

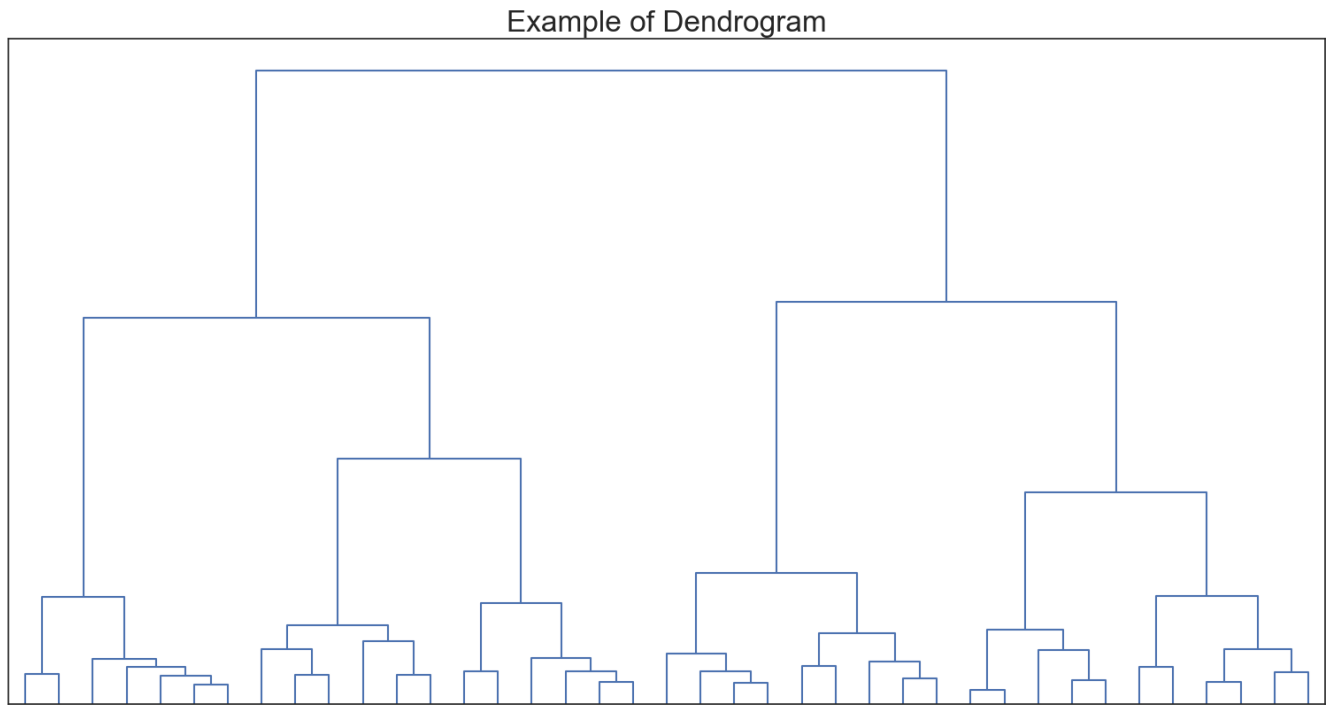


Figure 4. Example result of hierarchical clustering, Dendrogram

2.5.3 Similarity measurement

In hierarchical clustering, similarity measurement between clusters can be performed by setting linkage and metric. Linkage refers to setting a datum point that will be measured, and metric measures distance. Linkage includes Single, Complete, Average, Weighted, and Ward (Murtagh & Contreras, 2012). Table 4 has formulas for each linkage, which calculate the distance between the newly formed cluster U, V, and W while i and j are points in the cluster. In the case of Ward, U denotes the newly joined cluster consisting of clusters S and T, V is an unused cluster, and $Total = |v| + |s| + |t|$ defines the total number of points within the new cluster. Metric includes Euclidean, Manhattan, Chebyshev, and Cosine. Table 5 has formulas for each distance measurement metric for two points, $a(x_1, y_1)$ and $b(x_2, y_2)$.

Single, known as the minimum method or nearest neighbor method, merges clusters based on minimum distance (Sneath & Sokal, 1957), while complete, also referred to as furthest neighbor method, merges based on the maximum distance between two clusters and average, which is known as the minimum variance method is based on average distance literally (Sokal & Rohlf, 1962). Weighted gives equal weight to each cluster (Sokal & Rohlf, 1962), which helps balance the influence of differently sized clusters. Ward is another clustering method that is based on a classical sum-of-squares criterion, merging clusters that minimize the increase in total within-cluster variance (Ward, 1963). It has been widely used and generalized since its first description (Murtagh & Legendre, 2014).

Numerous studies have compared various Hierarchical Agglomerative Clustering (HAC) methods. Most findings suggest that despite being the only feasible option for large document sets, the single-link method generally produces suboptimal results (El-Hamdouchi & Willett, 1989; Willett, 1988; Steinbach et al., 2000). Among the HAC methods, the group average method is often considered superior to both the complete link and Ward's methods (El-Hamdouchi & Willett, 1989; Steinbach et al., 2000; Zhao & Karypis, 2002). This advantage is likely due to the single link method relying on minimal information and the complete link method treating clusters as highly dissimilar. In contrast, the group average method mitigates these issues by calculating the average distance between clusters (Steinbach et al., 2000).

Euclidean is the most common metric that measures the straight-line distance between two points. Manhattan measures distance by summing the absolute differences between the coordinates, which works well when movement is restricted to grid-like paths, such as city navigation or text processing. Chebyshev measures distance using the maximum absolute difference between coordinates, while cosine takes the angle between two vectors that cannot measure the actual physical distance.

Table 4. Linkage (Single, Complete, Average, Weighted, and Ward) and its formula

Linkage	Formula
Single	$d(U,V) = \min\left(d(U[i], V[j])\right)$
Complete	$d(U,V) = \max\left(d(U[i], V[j])\right)$
Average	$d(U,V) = \sum_{ij} \frac{d(U[i], V[j])}{(\lvert U\rvert * \lvert V\rvert)}$
Weighted	$d(U,V) = \frac{(d(U,W) + d(V,W))}{2}$
Ward	$d(U,V) = \sqrt{\frac{\lvert V\rvert + \lvert S\rvert}{N_T} d(V,S)^2 + \frac{\lvert V\rvert + \lvert T\rvert}{N_T} d(V,T)^2 - \frac{\lvert V\rvert}{N_T} d(S,T)^2}$

Table 5. Metric (Euclidean, Manhattan, Chebyshev, and cosine) and its formula

Metric	Formula
Euclidean	$d(a,b) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
Manhattan	$d(a,b) = \sum_{i=1}^{n}\lvert x_i - y_i\rvert$
Chebyshev	$d(a,b) = \max\left(\lvert x_2 - x_1\rvert, \lvert y_2 - y_1\rvert\right)$
Cosine	$d(a,b) = 1 - cos\theta = 1 - \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}}$

2.6 Performance evaluation criteria

Validating the quality of clustering is essential to ensure the accuracy and reliability of the outcomes. This section includes two performance evaluation indices: the cophenetic correlation coefficient to evaluate how well each similarity measurement combination explains the whole data and Silhouette scores to determine the optimal number of clusters.

2.6.1 Cophenetic correlation coefficient

Cophenetic correlation coefficients evaluate clustering solutions by determining how well a dendrogram retains the pairwise distances from the original distance matrix (Saracli et al., 2013). These coefficients are computed by comparing the original distance matrix to the cophenetic distance matrix, in which the original object distances are substituted with the distances between clusters when they are merged. Each clustering solution has its cophenetic correlation coefficient calculated separately, with higher values indicating superior clustering. This allows for quick and easy comparison of different clustering combinations. However, this method is sensitive to outliers (Sokal & Rohlf, 1962). The cophenetic correlation coefficient CPH is given by

$$CPH = \frac{\sum_{i<j}\left[d(i,j) - \bar{d}\right]\left[t_{den}(i,j) - \bar{t}\right]}{\sqrt{\sum_{i<j} d(i,j) - \overline{\bar{d}^2 \sum_{\iota<j} t_{den}(\iota,j)} - \bar{t}^2}}$$

Original data $\{D\}$ was clustered to produce a dendrogram $\{T_i\}$. Here, $d(i,j)$ represents the Euclidean distance between the i[th] and j[th] points. At the same time, $t_{den}(i,j)$ is the dendrogramatic distance between the model points T_i and T_j. The averages of $d(i,j)$ and $t_{den}(i,j)$ are denoted by $\bar{d}$ and $\bar{t}$, respectively. This value ranges from 0 to 1, meaning a higher value stands for a better combination in similarity measurement.

2.6.2 Cluster validity indices and Silhouette index

A few cluster validity measures evaluate clustering results, such as the Dunn index, Adjusted Rand index, Calinski-Harabasz index, and Silhouette index (Table 6). In Dunn index, $d(C_i, C_j)$ is the distance between the two closest points from different clusters i and j. $d(C_i)$ is the maximum distance within a cluster i. In Adjusted Rand index, n_{ij} denotes the number of points in predicted and true cluster i, and a_i and b_j are the sum of points in true cluster i and predicted cluster j, respectively. In the Calinski-Harabasz index, K is the number of clusters, n_k is data in cluster k, c_k is the centroid of cluster k, x_i^k is the i[th] data point in cluster k, and $\|a\|_2^2$ is the squared Euclidean distance. In the

Silhouette index, $a(i) = \frac{\sum_{j \in \{C_r \setminus i\}} d_{ij}}{n_r - 1}$ is the average dissimilarity of the i[th] object to all other objects within cluster C_r, which means cohesion. $b(i) = max_{s \neq r}\{d_{iCs}\}$, and $d_{iCs} = \frac{\sum_{j \in C_s} d_{ij}}{n_s}$ are the average dissimilarity of the i[th] object to all other objects within cluster C_s, representing separation (Rousseeuw, 1987; Gere, 2023).

The Dunn index evaluates the compactness between two clusters by the ratio of the minimum inter-cluster distance to the maximum intra-cluster (Dunn, 1973). The value increases when the distance between clusters is large and the variance within clusters is small. A higher value indicates better clustering quality. The adjusted Rand index measures the similarity between two clustering results by the proportion of pairs consistently classified in both clustering (Hubert & Arabie, 1985). The Rand index's accuracy decreases as the number of clusters increases (Rand, 1971), so the Adjusted Rand index adjusts for the expected similarity, compensating for the Rand index's limitations. The Calinski-Harabasz index considers both the cohesion within clusters and the separation between clusters by putting each in numerator and denominator (Chou et al., 2004). The silhouette index is used to evaluate the quality of clustering for individual data points by measuring how similar an object is to its own cluster compared to others. It assesses the quality of individual data points within clusters. (Hubert & Arabie, 1985; Chou et al., 2004; Pakhira et al., 2004; Rui Xu et al., 2012; Arbelaitz et al., 2013).

Cluster validity indices can be employed to determine the number

of clusters to mathematically validate the clustering. The Silhouette index measures how well an object is matched to its own cluster compared to other clusters (Rousseeuw, 1987). The Silhouette index varies between -1 and 1, with a score of 1 showing that the object is correctly grouped within its cluster.

Table 6. Clustering validity indices (Dunn, Adjusted Rand index, CS index, Silhouette index) and its formula

Cluster Validity Indices	Formula
Dunn Index (Dunn, 1973)	$DI = \frac{\min\limits_{1 \le i < j \le k} d(C_i, C_j)}{\max\limits_{1 \le i \le k} d(C_i)}$
Adjusted Rand Index (Hubert & Arabie, 1985)	$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}$
Calinski-Harabasz Index (Chou et al., 2004)	$CH = \left[\frac{\sum_{k=1}^{K} n_k \lVert c_k - c \rVert_2^2}{K-1}\right] / \left[\frac{\sum_{k=1}^{K} \sum_{i=1}^{nk} \lVert x_i^k - c_k \rVert_2^2}{n-K}\right]$
Silhouette Index (Rousseeuw, 1987; Charrad et al, 2014)	$S(i) = \frac{1}{N} \sum_{i=1}^{N} \frac{b_i - a_i}{\max(a_i, b_i)}$

3. Data

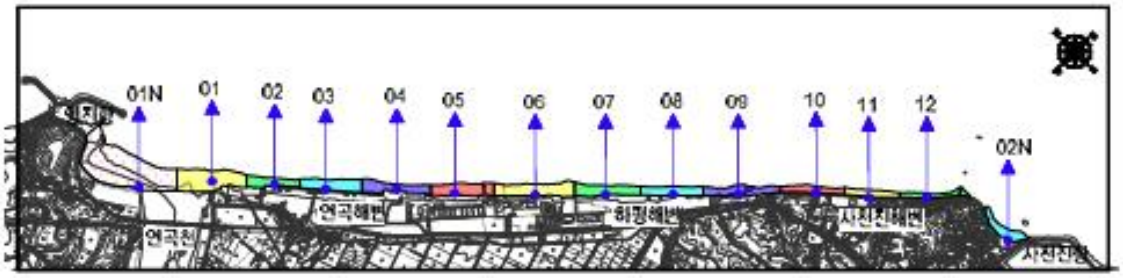
The data for this study were obtained from the annual first-year report titled *Cyclic Adaptive Coastal Erosion Management Technology Development* (Ministry of Oceans and Fisheries, 2023) and the annual report titled *Coastal Erosion Status Report* (Ministry of Oceans and Fisheries, 2023). Table 7 summarizes the list of data for each previously divided unit littoral cell, and Figures 5 and 6 represent the example tables from which the data were extracted in the report. In this study, we define the right with criterion facing the coast and the left side as the opposite direction within the unit littoral cell. For the East Coast, data were extracted, including left coordinates, right coordinates, length of coastline, beach width, representative sediment characteristics, grain size (d_{50}) of the left, center, and right side of the site, coastal morphology, and beach slope of the left, center, and right side for each littoral cell.

As distinct data existed for Gangwon and BUG, the analysis was conducted by dividing the East Coast into two regions: Gangwon and BUG, which encompasses Busan, Ulsan, and Gyeongbuk. The number of analyzed littoral cells was 42 for Gangwon and 62 for BUG (three for Busan, five for Ulsan, and 54 for Gyeongbuk). Figure 7 shows the distribution of unit littoral cells for Gangwon and BUG.

Table 7. List of available data for Gangwon and BUG region

Region	Available Data	
	Distinct	Common
Gangwon	Coastal morphology	Coordinates of coast, Length of coastline, Beach width, Representative sediment characteristic, d_{50}
BUG	Beach slope	

1.11.60 강릉시 사천면 사천진해변



해안선 길이	해빈폭	유입하천	대표저질특성	관측기선	침식등급	기선번호
870m	11~66m	-	모래	4개	C	10~12.02N

해빈폭	관측 전기간 대비 평균 8.5% 감소 전년 대비 평균 1.9 감소
단면적	관측 전기간 대비 평균 15.9 감소 전년 대비 평균 5.9% 증가
기울기	전년 대비 평균 기울기가 완만해짐 (10.5° → 8.9°)
표층퇴적물	전년 대비 평균 중앙입경 세립화 (1.13mm → 0.97mm)
연안정비사업	-
침식현황	고파랑 내습으로 인한 해안도로 월파(2012, 2020), 해변 침식(2022)

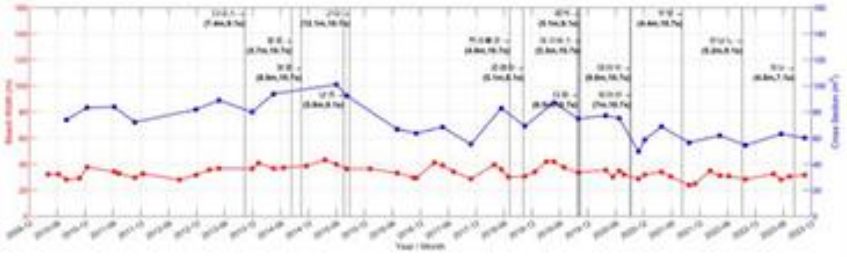
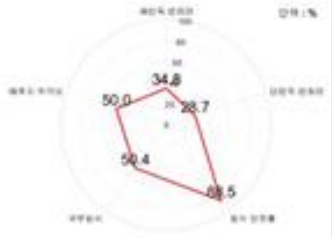
2010년~2023년 평균 해빈폭 및 단면적 변화		
고찰		○ 전년 대비 해빈폭은 감소하였으나, 단면적은 증가하였음 ○ 계절적인 연안표사에 의한 해빈폭과 단면적의 변화가 크게 발생하고 있음 ○ 고파랑 내습시 월파 및 해안도로 모래 퇴적 피해가 빈번함 ○ N02번 기선구간 침식으로 인해 해안절벽이 형성되었으며, 경사가 가파르게 형성됨

Figure 5. Example table of the report titled *Coastal Erosion Status Report* (Ministry of Oceans and Fisheries, 2023)

Figure 6. Example table of the report titled *Cyclic Adaptive Coastal Erosion Management Technology Development* (Ministry of Oceans and Fisheries, 2023)

Figure 7. Distribution of unit littoral cells for Gangwon and BUG

Figure 10 visualizes the length of the coastline for each littoral cell and Figure 11 shows the beach width at each site. Figures 12, 13, and 14 show the d_{50} for each unit littoral cell's left, center, and right sides, connected by line graphs. Figures 15, 16, and 17 show the Beach Slope for each unit littoral cell's left, center, and right side, connected by line graphs.

Pearson correlation coefficients were calculated on the raw data used in the study. The sample sizes were 42 for Gangwon and 62 for BUG. The results are presented in Figures 8 and 9. In Gangwon and BUG, Leftside Latitude, Leftside Longitude, Rightside Latitude, and Rightside Longitude exhibited strong positive and negative correlations. As such, high correlations can lead to redundancy problems, so latitude and longitude data were averaged, reducing redundancy and improving data integrity for the analysis.

Correlation analysis of D50(L), D50(M), and D50(R) yielded values of 0.67, 0.80, and 0.78 in Gangwon, and 0.74, 0.75, and 0.78 in BUG, respectively. However, these values were not entirely reliable due to the presence of numerous missing values in the raw data. To maintain consistency in the sample size, missing values were imputed with the average d_{50} values of D50(L), D50(M), and D50(R).

It is important to note that Pearson correlation measures linear relationships, which limits its effectiveness in capturing non-linear relationships. Therefore, hasty decisions about data selection based solely on these correlation values could lead to misinterpretations in data analysis. Moreover, since UMAP does not work with missing values (cases with missing data are excluded from the analysis), it was essential to fill in. Therefore, d_{50} was newly introduced, representing the average of D50(L), D50(M), and D50(R) to avoid redundancy problems and to enhance accuracy.

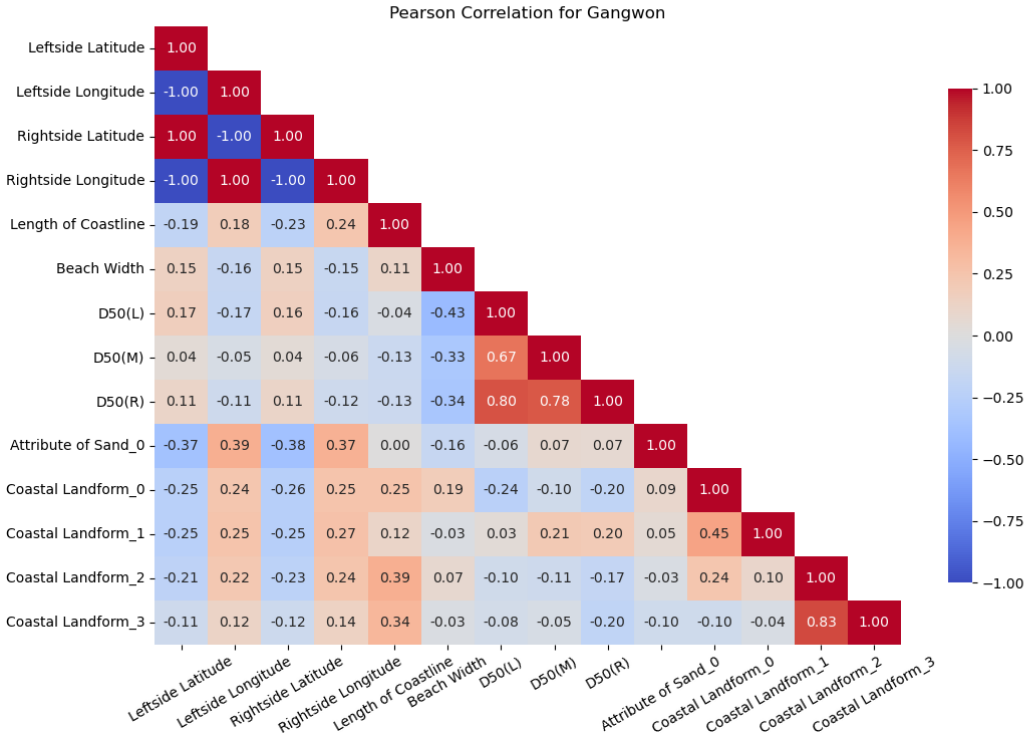


Figure 8. Pearson Correlation Matrix for Gangwon

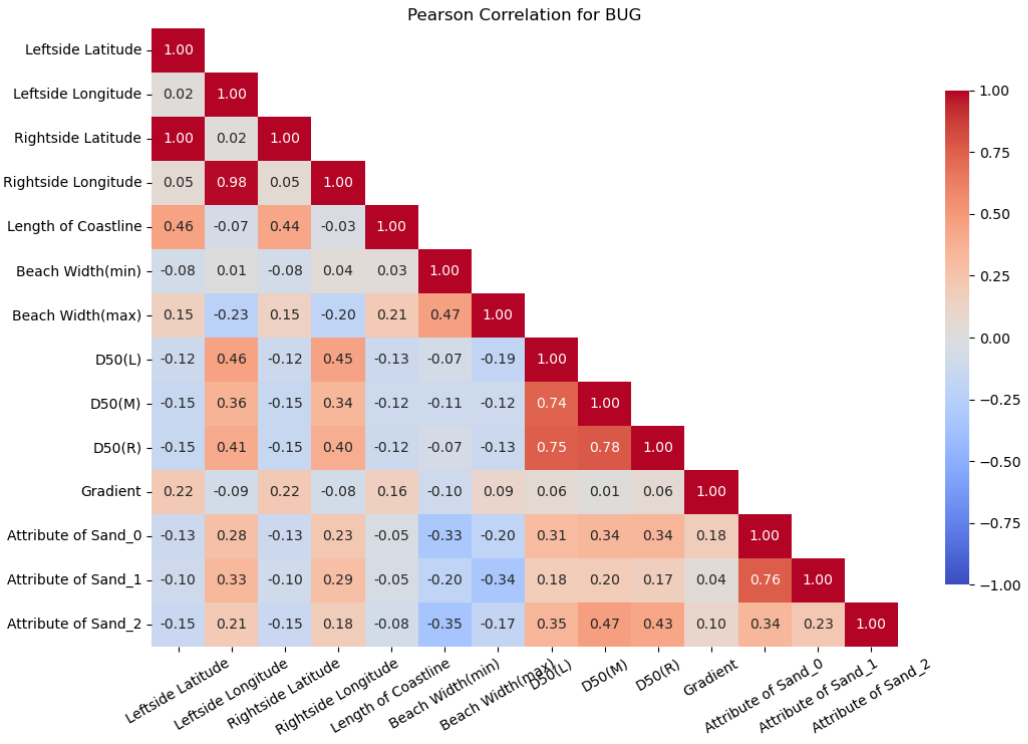


Figure 9. Pearson Correlation Matrix for BUG

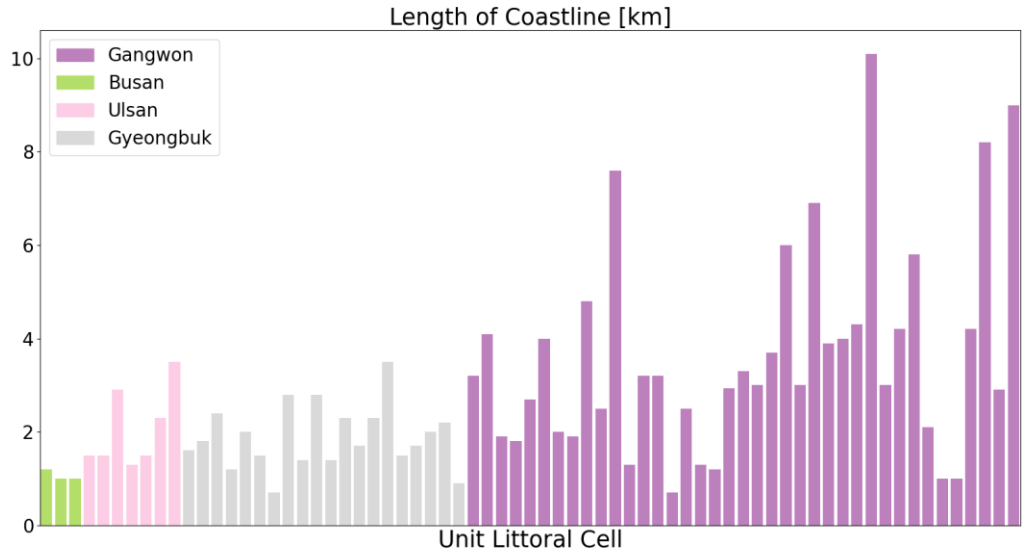


Figure 10. Length of coastline for Gangwon and BUG

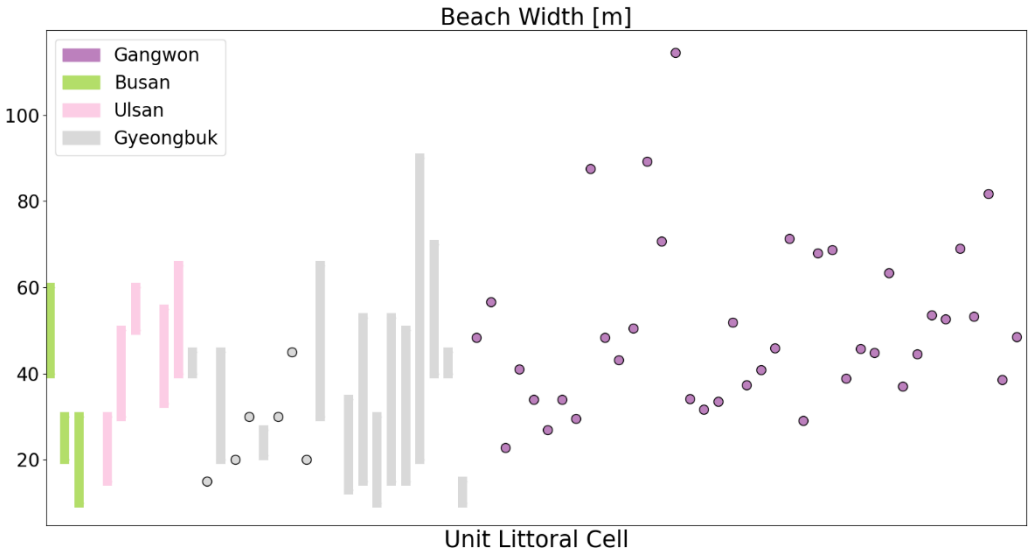


Figure 11. Beach width for Gangwon and BUG

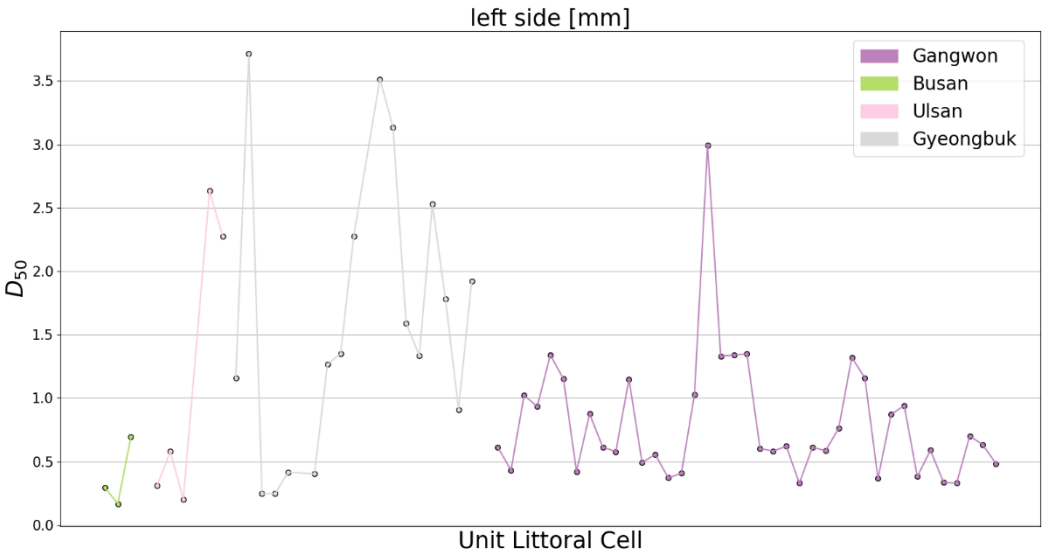


Figure 12. D_{50} of left side of unit littoral cell for Gangwon and BUG

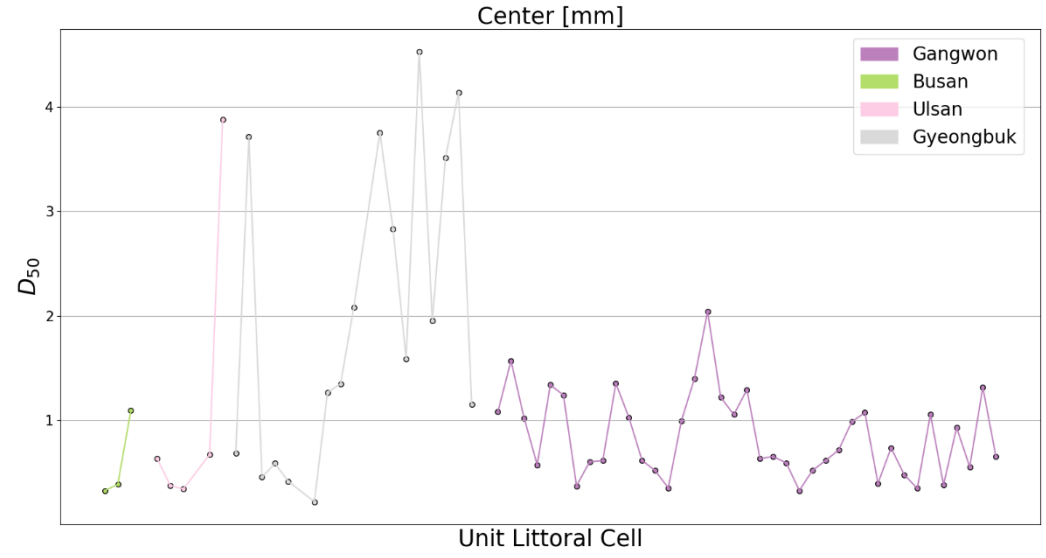


Figure 13. D_{50} of center of unit littoral cell for Gangwon and BUG

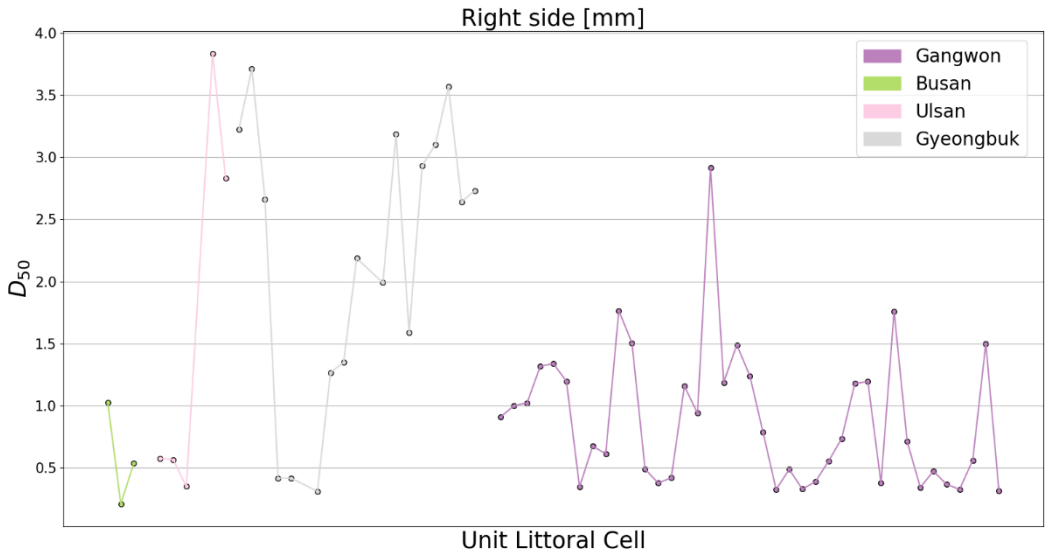


Figure 14. D_{50} of right side of unit littoral cell for Gangwon and BUG

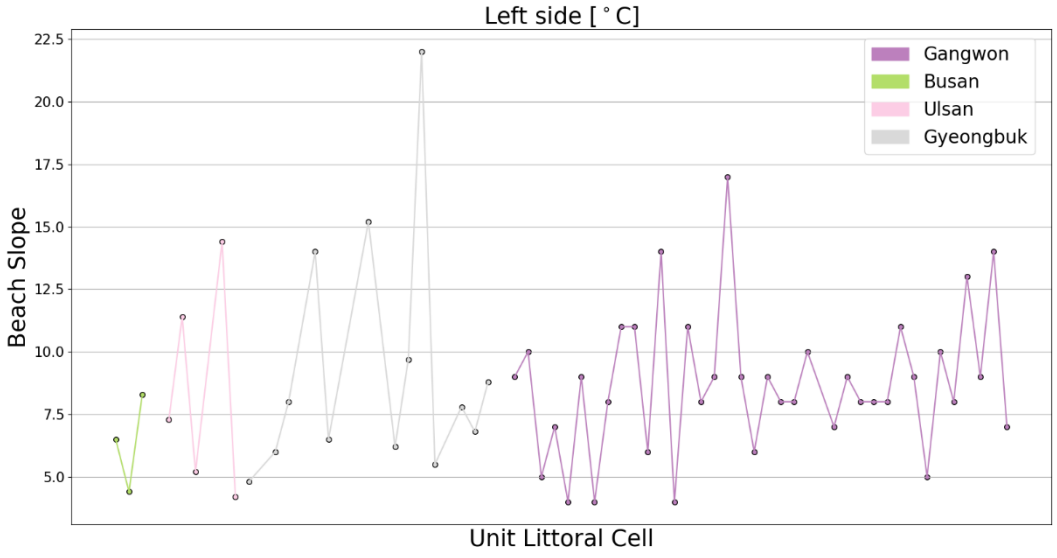


Figure 15. Beach slope of left side of unit littoral cell for Gangwon and BUG

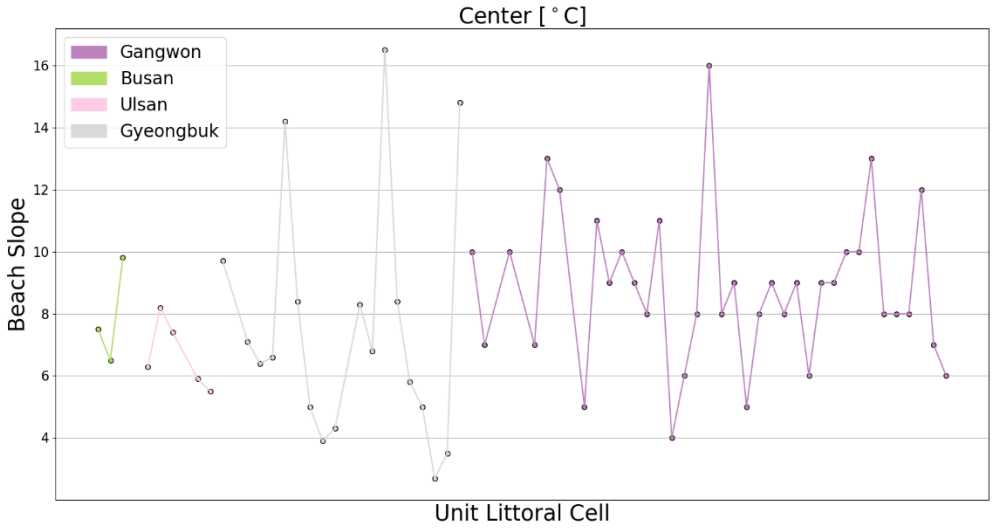


Figure 16. Beach slope of center of unit littoral cell for Gangwon and BUG

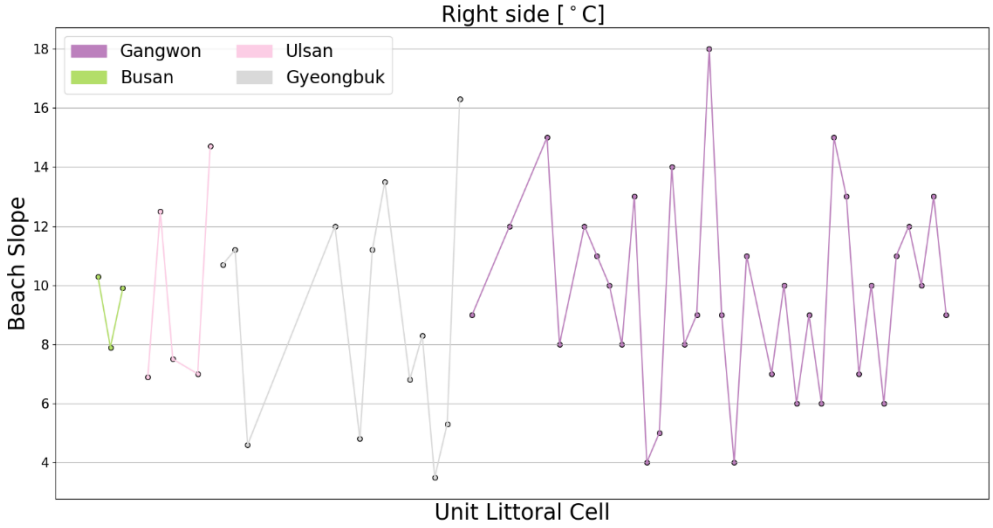


Figure 17. Beach slope of right side of unit littoral cell for Gangwon and BUG

4. Results

This section presents the results of the analysis conducted on clustering unit littoral cells. The findings are organized into three key steps: dimensionality reduction, clustering and its validation, and GIS-based revising. First, the results of dimensionality reduction using UMAP are presented to reduce high-dimensional sediment transport parameter data in a lower-dimensional space. Next, clustering and its validation are conducted using the cophenetic correlation coefficient and silhouette scores to enhance the quality of outcomes. Finally, GIS-based spatial revising is employed to double-check the spatial distribution of designated clusters of unit littoral cells with observed coastal features.

4.1 UMAP projection

To facilitate algorithm execution and improve performance, UMAP reduced high-dimensional datasets with different features into new-dimensional data objects. The results were visualized on a two-dimensional plane, as shown in Figures 18 and 19. Overall, clusters are formed by data points from the same or adjacent regions grouping together.

To evaluate how well UMAP preserves the distance structure of the original data, we compared the Mean Squared Error (MSE) between the UMAP embedding results and random embedding. The comparison showed that UMAP outperformed random embedding regarding distance preservation in both Gangwon and BUG. Specifically, the MSE for UMAP was 7.916 in Gangwon, compared to 9.465 for random embedding. Similarly, in BUG, the MSE for UMAP was 5.510, while random embedding resulted in an MSE of 6.598. These results indicate that UMAP maintains the original data structure better after dimensionality reduction.

4.2 Cophenetic correlation coefficient

Tables 8 and 9 show the calculated cophenetic correlation coefficient for different linkage methods: Single, Complete, Average, Weighted, and Ward, with the four metrics: Euclidean, Manhattan, Chebyshev, and Cosine. Table 8 shows the Cophenetic correlation coefficient of Gangwon. As a higher value means better performance combination, Single-Manhattan shows the worst combination value of 0.35. The Complete-Chebyshev, Average-Chebyshev, Single-Cosine, and Average-Cosine metrics showed the highest value of 0.67. Compared with other metrics, including Euclidean, the Complete-Euclidean metric exhibited the highest score of 0.66. Table 9 is the result of BUG. Euclidean metrics, except for single-Euclidean, showed the highest value of 0.85, while cosine metrics showed the worst.

Random seeds were controlled using tf.random.set_seed(42) for TensorFlow, np.random.seed(42) for NumPy, and random.seed(42) for Python's random module. This approach guarantees reproducibility, as repeated executions consistently yield the same results. Therefore, the observed cophenetic correlation coefficients are not the outcome of random chance but reflect stable and meaningful clustering patterns in the data.

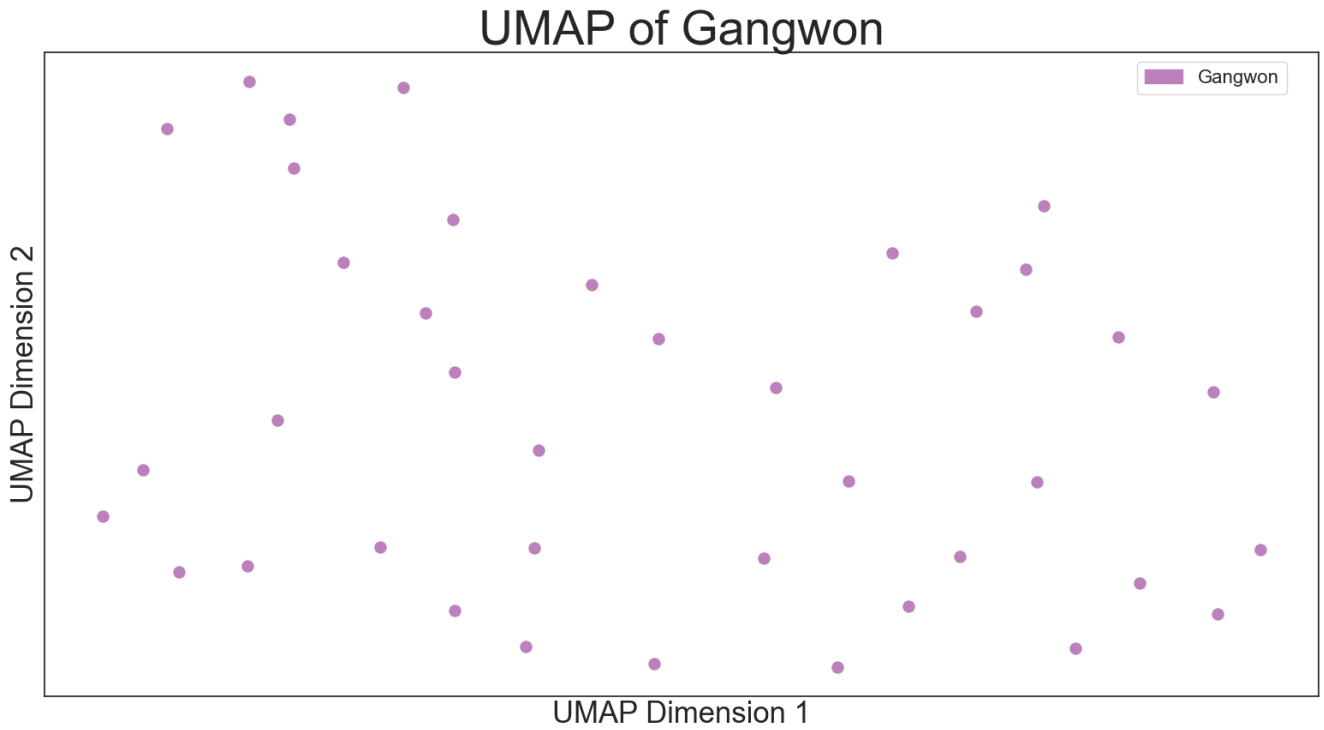


Figure 18. UMAP of Gangwon

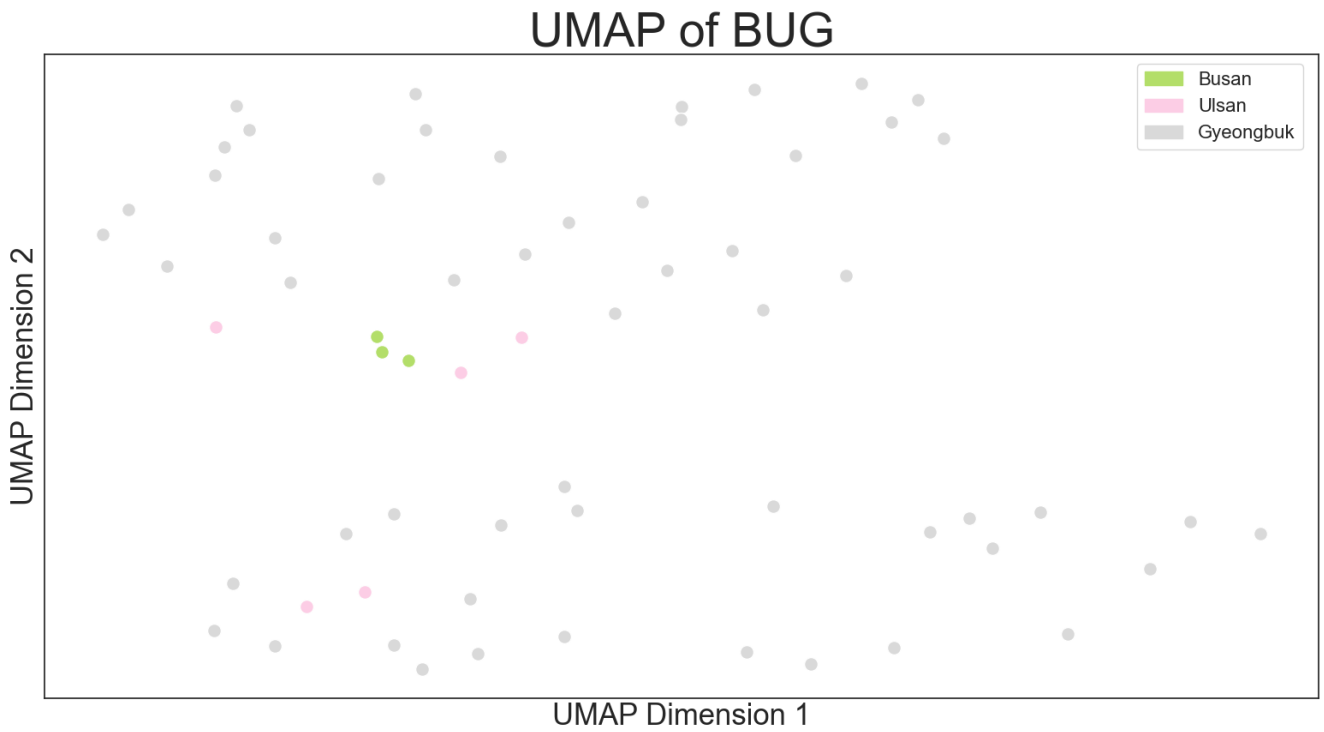


Figure 19. UMAP of BUG

Table 8. Cophenetic correlation coefficient for Gangwon

	Euclidean	Manhattan	Chebyshev	Cosine
Single	0.84	0.83	0.83	0.81
Complete	0.85	0.84	0.84	0.81
Average	0.85	0.84	0.84	0.81
Weighted	0.85	0.84	0.84	0.81
Ward	0.85	Euclidean only	Euclidean only	Euclidean only

Table 9. Cophenetic correlation coefficient for BUG

	Euclidean	Manhattan	Chebyshev	Cosine
Single	0.41	0.35	0.52	0.67
Complete	0.66	0.62	0.67	0.54
Average	0.65	0.64	0.67	0.67
Weighted	0.64	0.6	0.66	0.54
Ward	0.65	Euclidean only	Euclidean only	Euclidean only

4.3 Dendrogram

By clustering sites using hierarchical agglomerative clustering, a dendrogram that visualizes hierarchical linkage and clusters directly can be obtained. Since only the Euclidean metric can be applied to Ward, the cophenetic correlation coefficient values were compared by fixing the metric to Euclidean in both cases. For Gangwon, the Complete linkage showed the highest value, while for Bug, all linkage methods except for Single exhibited equally high values. Therefore, linkage as Complete and metric as Euclidean was used for both cases. The site number with colored index represents the region allocated to each unit littoral cell (site). The colored linkage in the dendrogram is set to have its own region color when data points from the same region meet. Figure 22 is leaf cut dendrogram of BUG for closure observation.

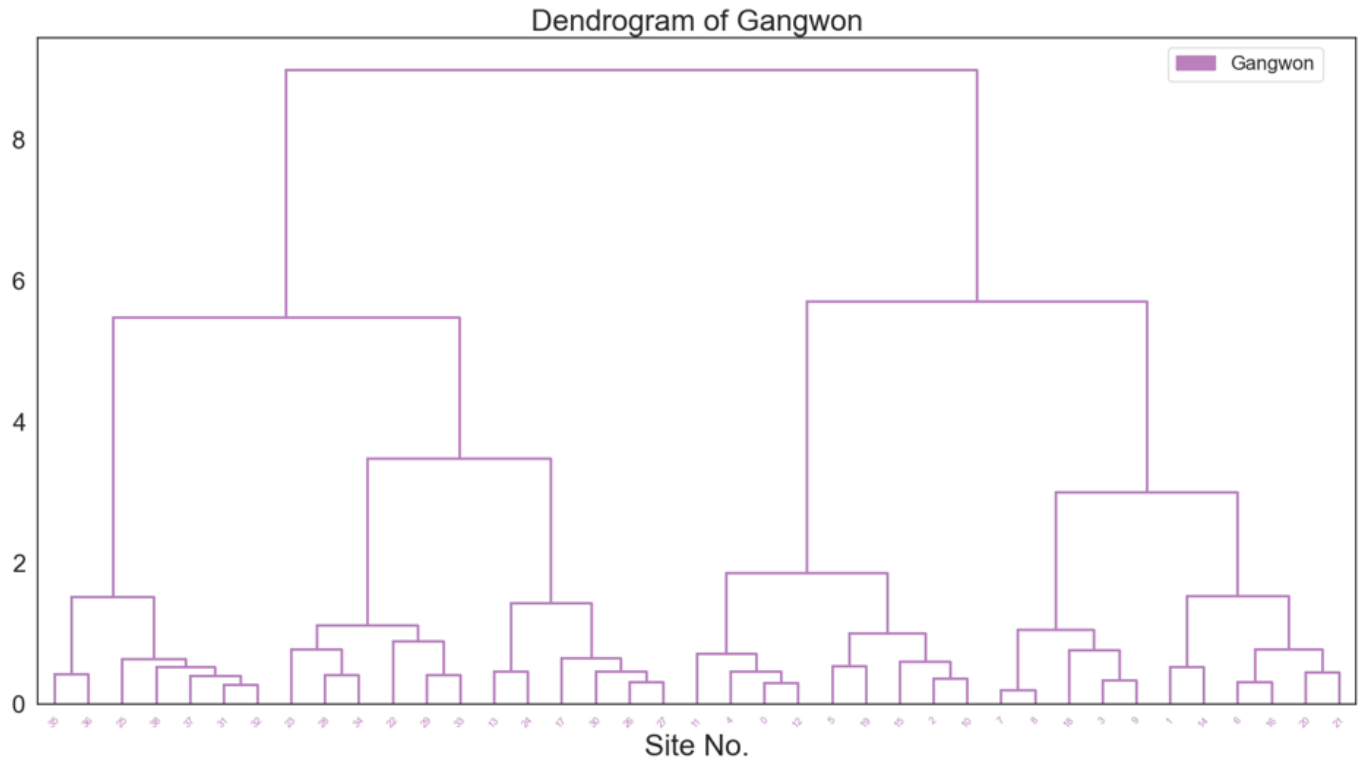


Figure 20. Dendrogram of Gangwon using linkage as Ward and metric as Euclidean

Figure 21. Dendrogram of BUG using linkage as Ward and metric as Euclidean

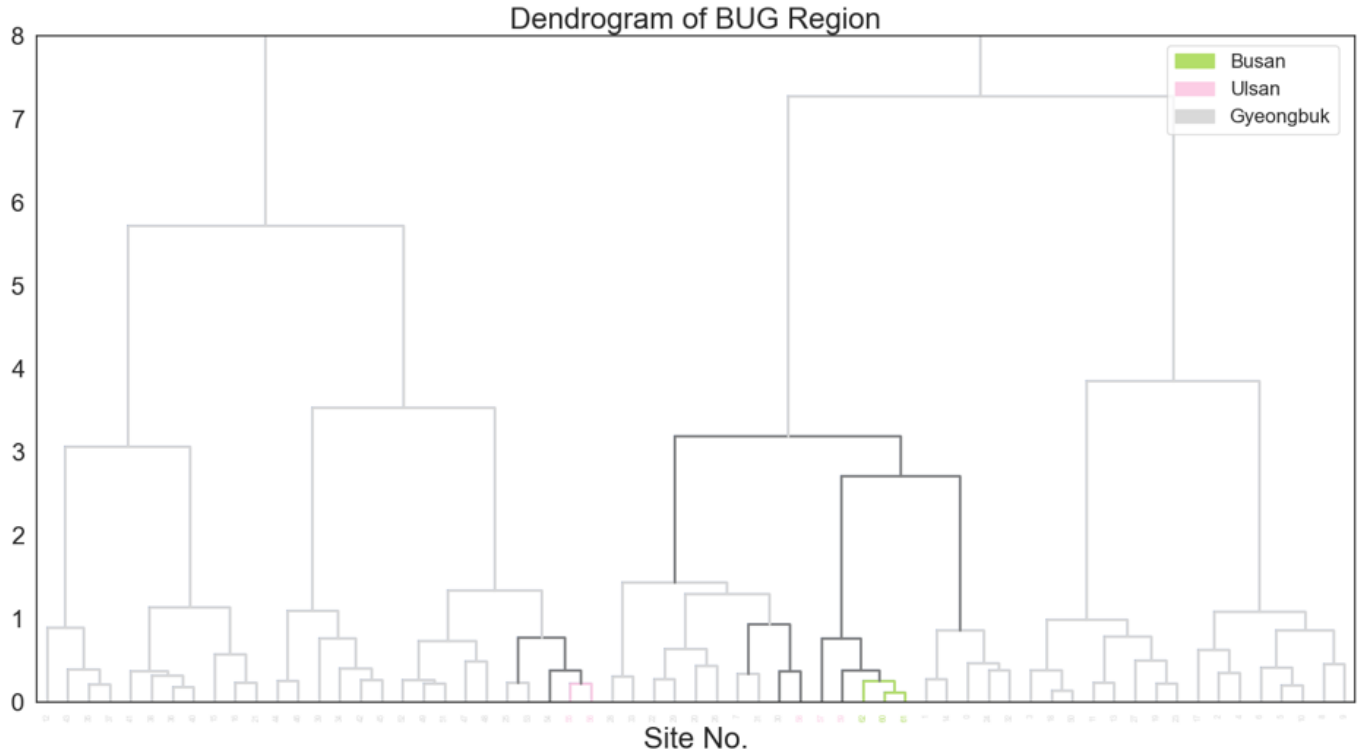


Figure 22. Enlarged dendrogram of BUG using linkage as Ward and metric as Euclidean

4.4 Silhouette score and optimal cluster number

The silhouette index was calculated to determine the optimal number of clusters that best explain the individual data points and clusters. The number of tested clusters range from two to a whole number of data points, with each littoral cell have its own cluster. Figures 23 to 24 are the silhouette scores of each number of clusters. Top five results with the highest reliability were found for each case. For Gangwon, silhouette scores were 0.42, 0.42, 0.41, 0.41, and 0.40 when the whole data was divided into 6, 5, 7, 4, and 8 clusters, respectively (Table 10). As clusters with high silhouette scores align with well-divided clustering results, Gangwon was divided into 6 clusters. For BUG, silhouette scores were 0.63, 0.50, 0.49, 0.47, and 0.46 when the whole data was divided into 2, 3, 9, 8, and 10 clusters, respectively (Table 11). Although BUG ranks third, it was divided into 9 clusters to see the detailed clustered results because they are better for examining the feasibility of defining mesoscale littoral cells.

Table 10. Top 5 optimal number of clusters and silhouette score of Gangwon

Top 5 Clusters by Silhouette Score
Rank 1: 6 clusters, Silhouette Score: 0.4159
Rank 2: 5 clusters, Silhouette Score: 0.4131
Rank 3: 7 clusters, Silhouette Score: 0.4085
Rank 4: 4 clusters, Silhouette Score: 0.4052
Rank 5: 8 clusters, Silhouette Score: 0.3999

Table 11. Top 5 optimal number of clusters and silhouette score of BUG

Top 5 Clusters by Silhouette Score
Rank 1: 2 clusters, Silhouette Score: 0.6284
Rank 2: 3 clusters, Silhouette Score: 0.4987
Rank 3: 9 clusters, Silhouette Score: 0.4923
Rank 4: 8 clusters, Silhouette Score: 0.4684
Rank 5: 10 clusters, Silhouette Score: 0.4598

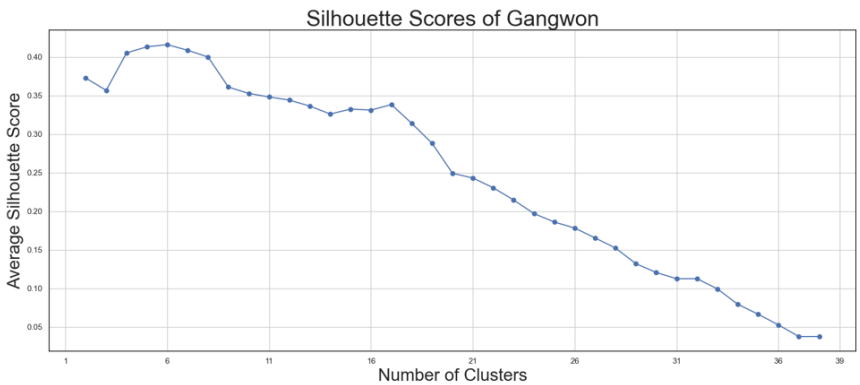


Figure 23. The silhouette score of each number of clusters for Gangwon

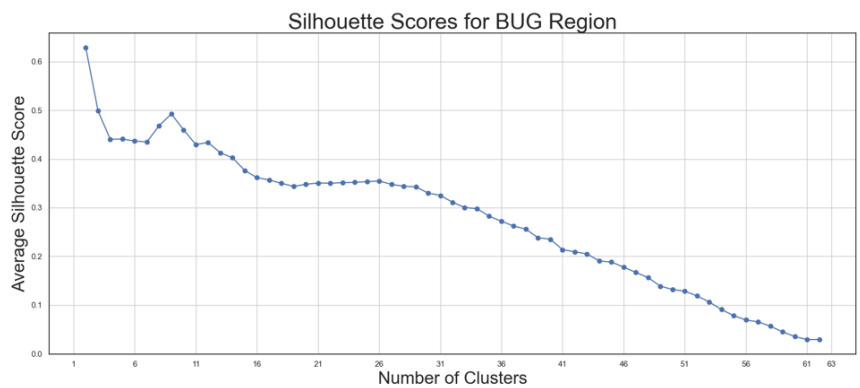


Figure 24. The silhouette score of each number of clusters for BUG

4.5 Mapping and comparing with previously divided unit littoral cell

Figures 25 and 26 illustrate the mapped results of the allocated clusters based on latitude and longitude data. To compare these results with previously divided unit littoral cell, interpretation rule and steps were established: (1) If the allocated cluster changes between adjacent littoral cells, the boundary is designated as a new boundary for the newly demarcated littoral cells, (2) The clustering results were refined using satellite imagery. Natural capes and man-made structures, such as harbors and sediment control facilities, were considered as the factor that limit the sediment transport. (3) The clustering outcomes are meticulously examined through close observation to ensure accuracy and alignment with these geomorphological characteristics. Figure 27 is an example of a refinement process.

A chart was drawn to visualize the differences between the two systems by aligning the data points and assigning varying heights to each based on their allocated clusters (Figures 28 and 29). Each circle indicates the original unit littoral cells demarcated in the report. When a single unit littoral cell needed to be subdivided into smaller units, a line across the circle was drawn. In cases where multiple unit littoral cells were assigned to the same cluster (indicating they were considered as part of the same system) but were determined to need separation into different clusters (indicating they should be distinguished as distinct sedimentary systems), a line between the circles was drawn. Additionally, when multiple unit littoral cells were determined to form a single littoral cell, pink labels were assigned to distinguish them. Below, the clustering results using machine learning are

presented with pink labels, each representing a single littoral cell.

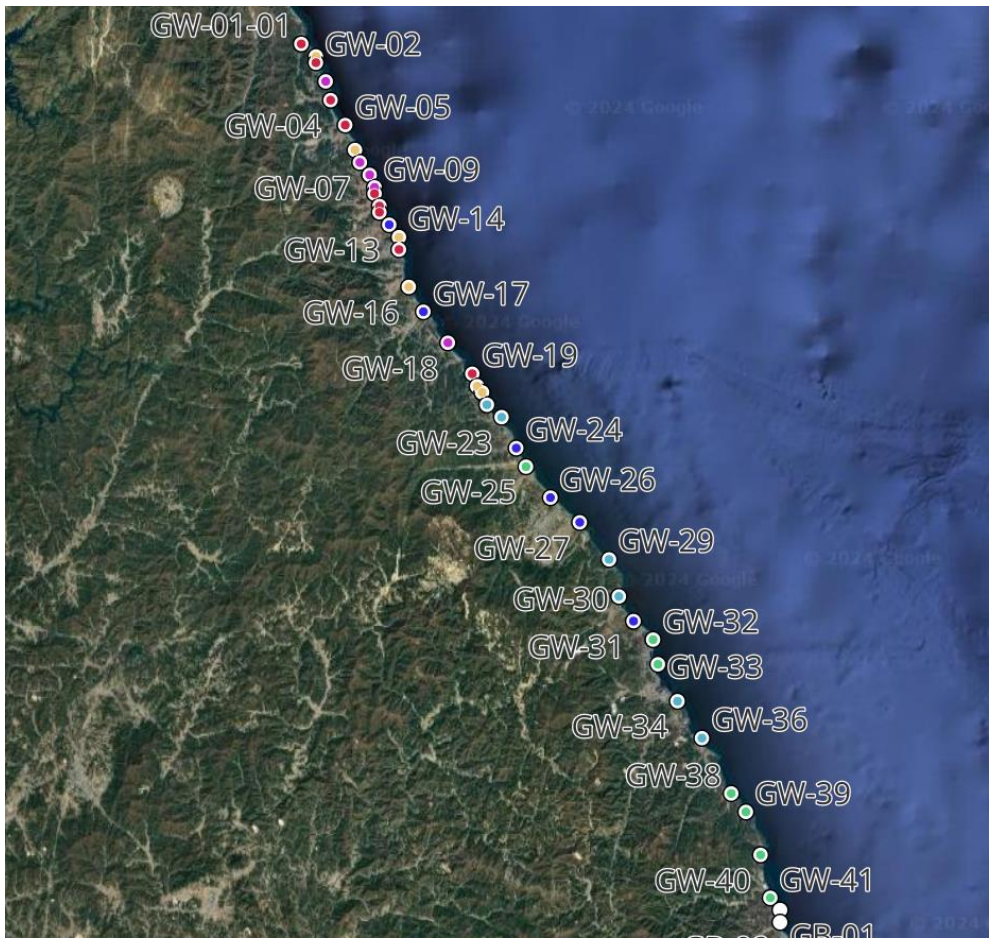


Figure 25. Mapping with 6 clusters of Gangwon, Silhouette Score: 0.4159

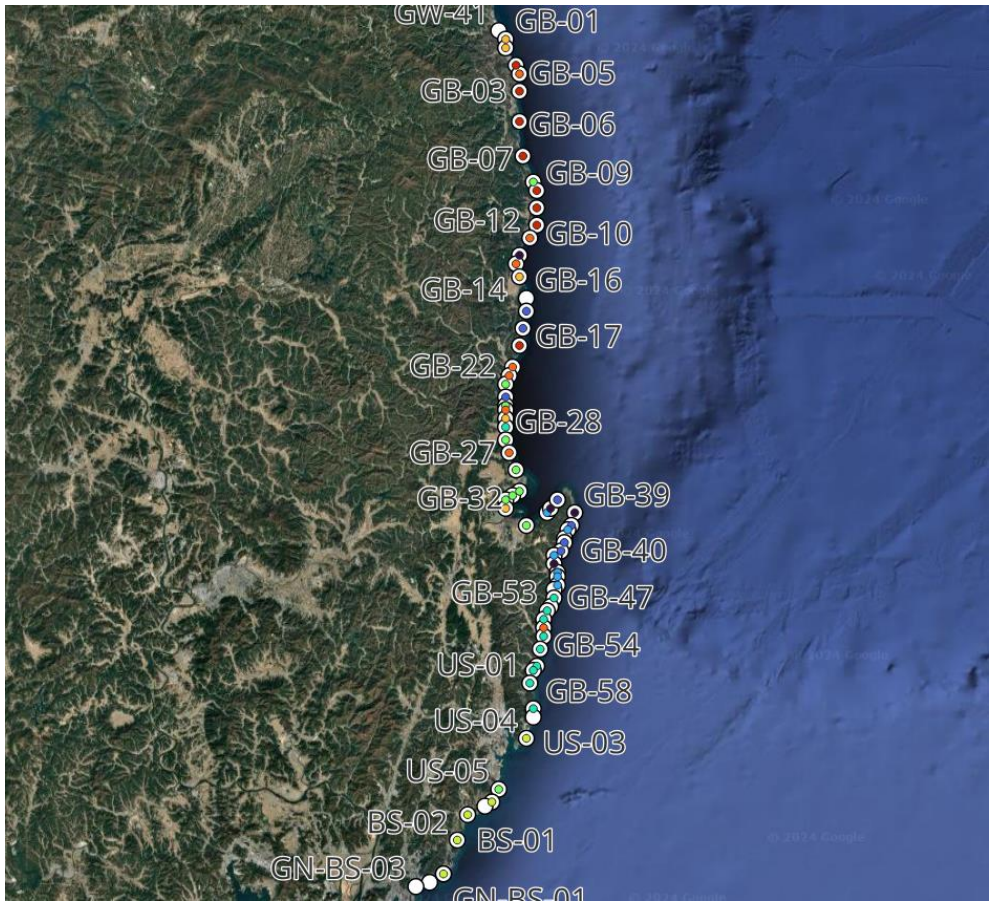


Figure 26. Mapping with 9 clusters of BUG, Silhouette Score: 0.4923

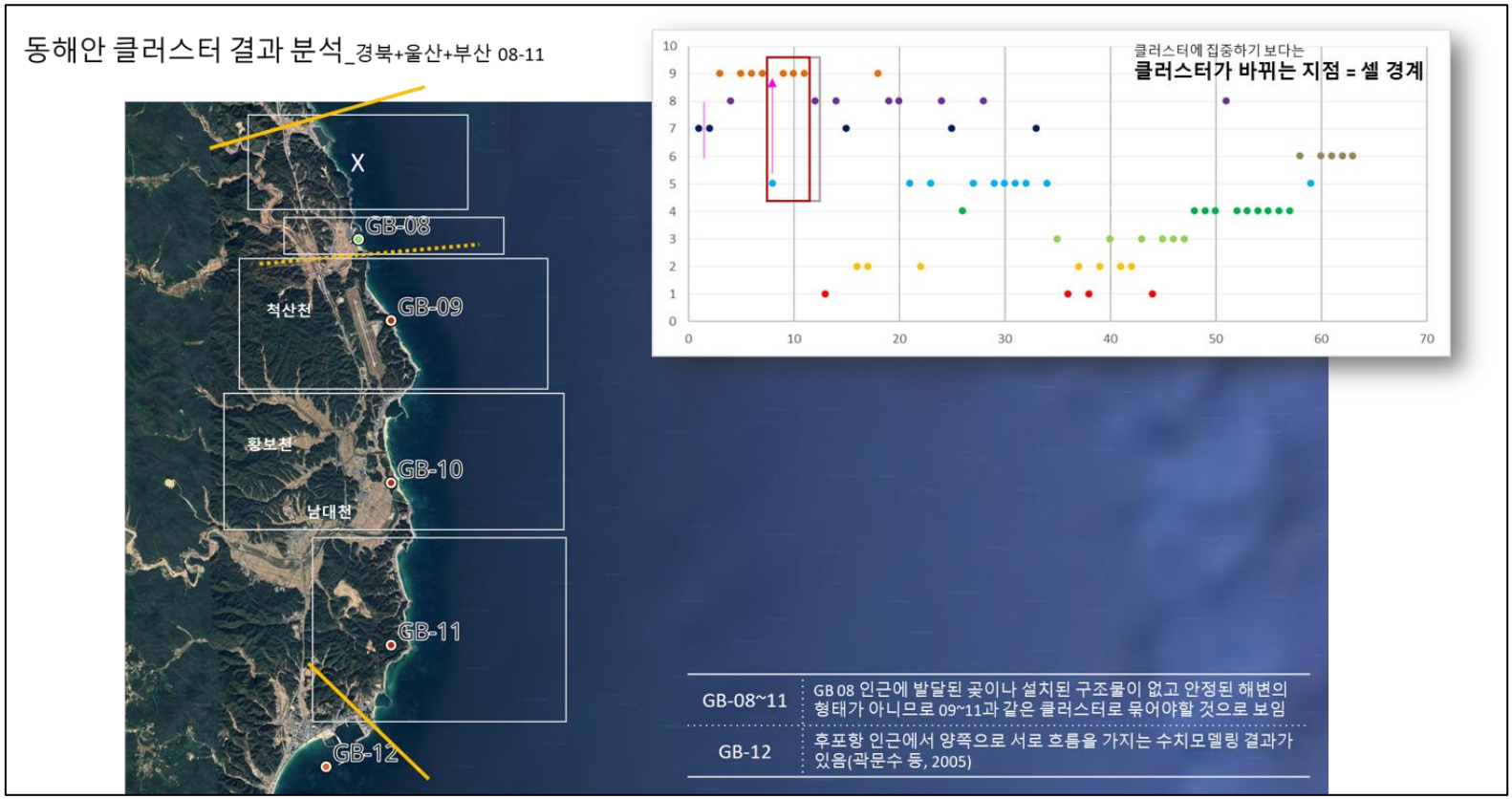


Figure 27. Example of a refinement using geomorphological images and information

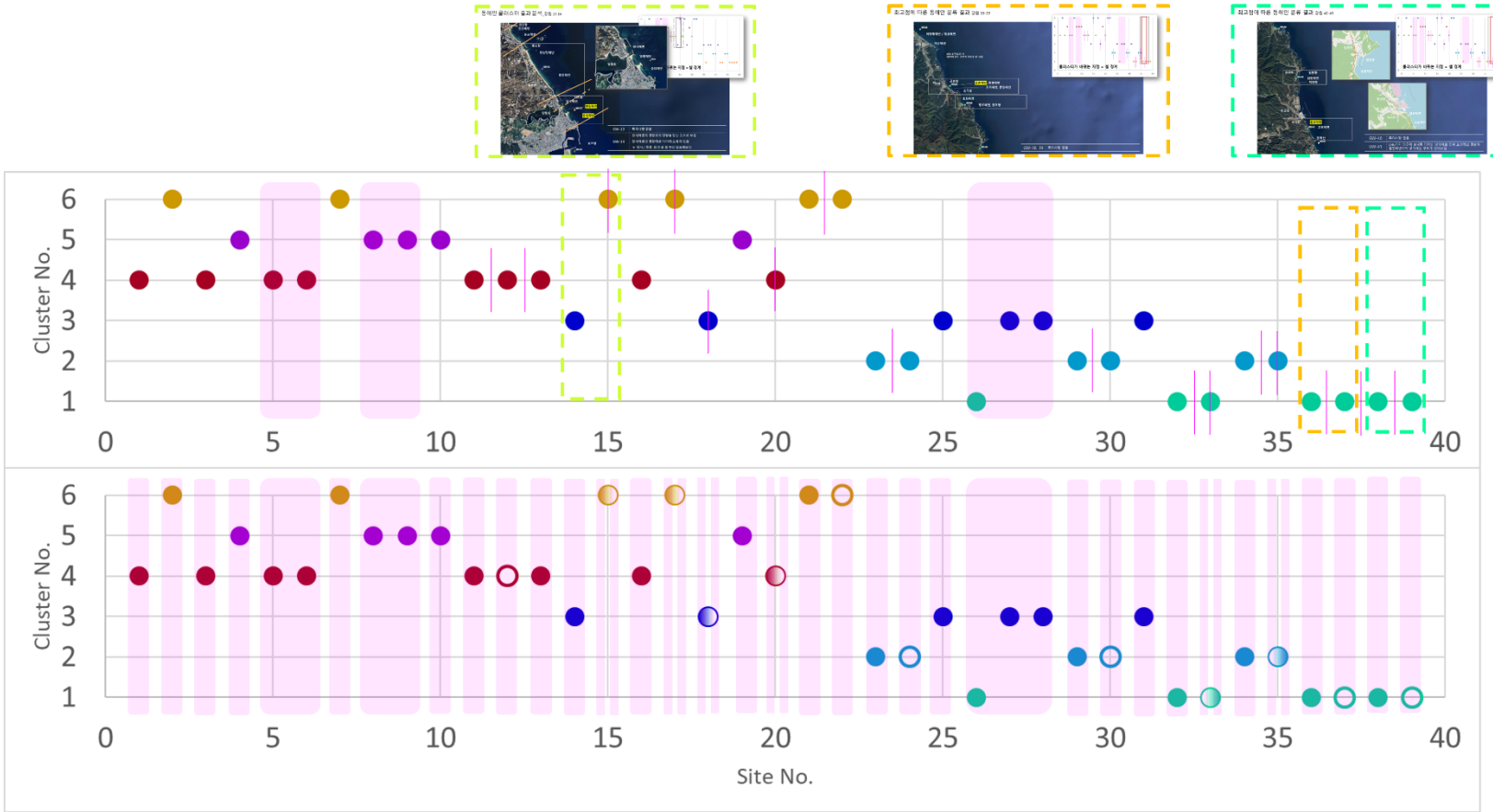


Figure 28. Refined result of unit littoral cells and comparison with current littoral cell system for Gangwon

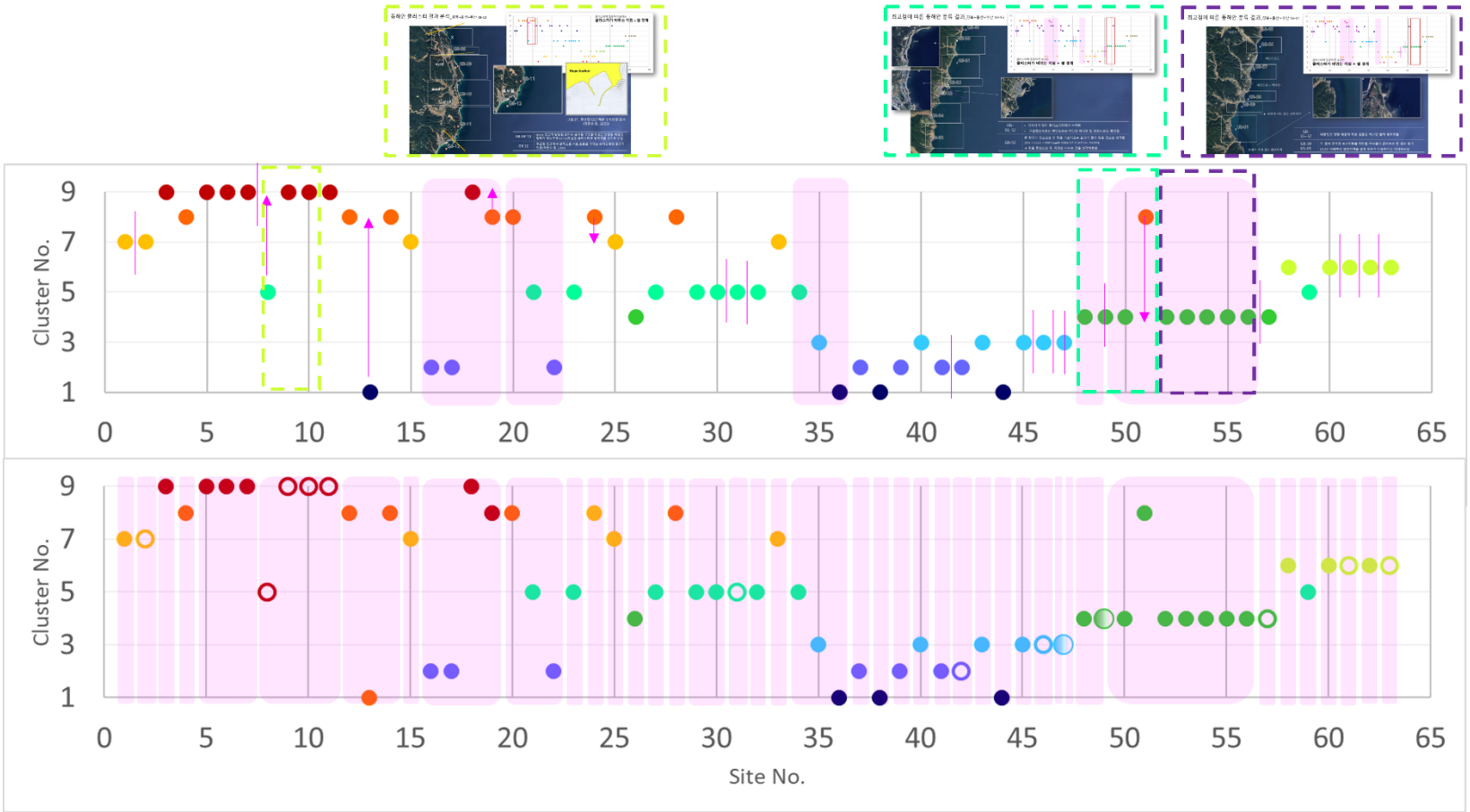


Figure 29. Refined result of unit littoral cells and comparison with current littoral cell system for BUG

5. Discussion

5.1 Coastal erosion grade D

The Comprehensive Report on Coastal Erosion Status established evaluation criteria to assess the sedimentation and erosion conditions of each monitoring site, which are 360 primary coastal erosion monitoring areas. Based on these criteria, the report assigned coastal erosion grades from A to D to each site. Sites classified as Grade D erosion are shown in Table 12.

Table 12 List of Beaches Classified as Grade D in Coastal Erosion Evaluation

Administrative District	Region Name	Type of Erosion
Gangwon	Goseong-gun Cheonhakjeong	Beach Erosion
	Sokcho-si Yeongnang-dong	Beach Erosion
	Sokcho-si Deungdae	Beach Erosion
	Gangneung-si Hasi-dong	Beach Erosion
	Samcheok-si Gungchon	Beach Erosion
	Samcheok-si Wolcheon	Beach Erosion
Gyeongbuk	Uljin-gun Jiksan	Beach Erosion
	Yeongdeok-gun Geumjin ~ Hajeo	Beach Erosion
	Yeongdeok-gun Woncheok ~ Buheung	Sediment Loss
	Gyeongju-si Naa	Beach Erosion
Total	10	

5.2 Management scope suggestion for Gangwon

In this section, the clustering results for beaches with erosion grade D were revised using satellite imagery, cross-checking the initial clustering outcomes with specific coastal characteristics and existing research for accuracy and relevance. Natural capes were considered as they significantly influence sediment transport patterns compared to open coastlines. Additionally, the impact of coastal structures was considered, especially areas reported as having sheltered zones due to harbors or other artificial structures. Such structures can create localized changes in current flow and sediment deposition. Furthermore, existing studies on current flow dynamics near D-grade erosion beaches or ports were incorporated to validate the clustering. Through this methodology, management boundaries were suggested for each erosion grade D site and six key regions in Gangwon were examined: (1) Goseong-gun Cheonhakjeong, (2) Sokcho-si Yeongnang-dong, (3) Sokcho-si Deungdae, (4) Gangneung-si Hasi-dong, (5) Samcheok-si Gungchon, and (6) Samcheok-si Wolcheon.

Figure 30. Satellite image of Chunhakjeong Beach

Figure 30 is a satellite image of the Chunhakjeong Beach. According to Kim and Song (2019a), Munamcheon, which used to be a sediment source to Chunhakjeong Beach, altered its flow direction to Baekdo Beach, blocking sediment supply. The construction of the Munam2ri Port disrupted the seasonal equilibrium state, accelerating coastal erosion in Gyoam Beach. Based on the long-period wave data in Gyoam Beach, the dominant wave direction is northeast (Kim and Song, 2019b). Therefore, sediment is unlikely migrate beyond Ayajin Beach toward Ayajin Port. The boundary of the littoral cell that includes Chunhakjeong Beach can be inferred from the yellow dotted line in Figure 30.

Figure 31. Satellite image of Yeongnang Beach and Deungdae Beach

Figure 32. Zoomed-in satellite image of Yeongnang Beach and Deungdae Beach

Figures 31 and 32 are satellite images of Yeongrang Beach and Deungdae

Beach. The yellow dotted line indicates the boundaries of the littoral cell used for the analysis. However, Yeongnang Beach and Deungdae Beach exhibit stable pocket-shaped beach morphology and the sediment transport from Yeongnang Lake is restricted by the groin, observed from the deposits on the northern side of Yeongnang Beach. Therefore, when managing Yeongnang Beach and Deungdae Beach, which are classified as Grade D erosion sites, it is assumed that a more efficient management boundary would range from the magenta dotted line to the yellow dotted line on the southern side of Deungdae Beach, rather than adhering to the original littoral cell boundaries.

Figure 33. Satellite image of Hasidong Beach

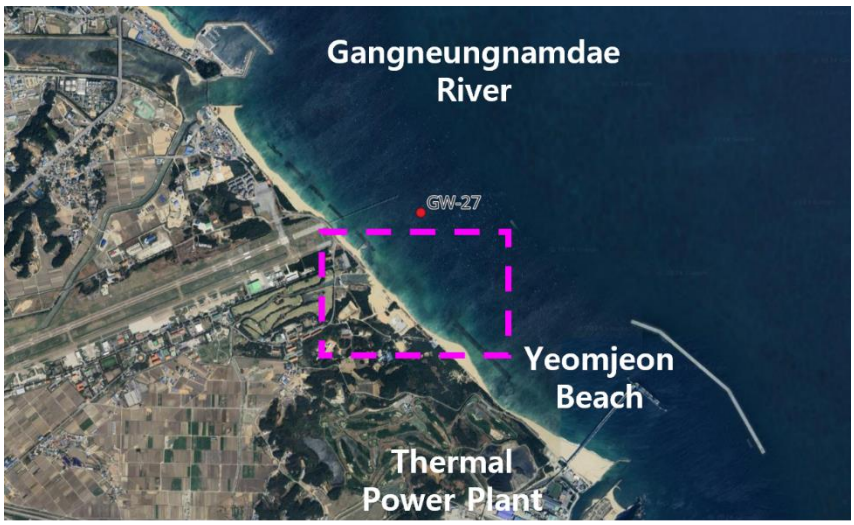


Figure 34. Zoomed-in satellite image of Hasidong Beach

Figures 33 and 34 are satellite images of Hasidong Beach. In Figure 34, the area marked by the magenta dotted line indicates Hasidong Beach, where the cause of erosion is identified as beach erosion. According to the Comprehensive Report on Coastal Erosion Status, continuous erosion damage has been caused by the thermal power plant. Hasidong Beach receives sediment from the Gangneungnamdae River and the structures installed at the thermal power plant, located south of the Hasidong beach, are likely to restrict sediment movement. Therefore, to efficiently manage Hasidong Beach, it can be inferred that the management boundary should extend from Gangneung Port to the thermal power plant as indicated by the yellow dotted line in the Figure 33.

Figure 35. Satellite image of Gungchon Beach

Figure 36. Zoomed-in satellite image of Gungchon Beach

Figures 35 and 36 are satellite images of Gungchon Beach. The cause of erosion at Gungchon Beach is identified as beach erosion, and it receives sediment supply from the Chu River. Gungchon Port restricts the movement of northward

littoral drift, while a natural cape is located at Chogok Port. Therefore, to efficiently manage erosion at Gungchon Port, the management boundary should be set from Gungchon Port to Chogok Port, as indicated by the yellow dotted line in Figure 35.

Figure 37. Satellite image of Wolchon Beach

Figure 38. Zoomed-in satellite image of Wolcheon Beach

Figures 37 and 38 are satellite images of Wolcheon Beach. The cause of erosion at Wolcheon Beach is beach erosion, which has been continuously classified as Grade D following the construction of a nearby LNG production facility. Although the erosion grade temporarily improved after restoration work in 2022, sediment supply to Wolcheon Beach remains restricted due to the surrounding structures. To efficiently manage Wolcheon Beach, the management boundary should include the area within the dotted lines in Figure 37, which encompasses GagokStream.

5.3 Management scope suggestion for BUG

In this section, the clustering results for beaches with erosion grade D in BUG were revised using the same criteria applied in the Gangwon. Four key regions for BUG requiring targeted management strategies were identified: (1) Uljin-gun Jiksan, (2) Yeongdeok-gun Geumjin to Hajeo, (3) Yeongdeok-gun Woncheok to Buheung, and (4) Gyeongju-si Naa. However, unlike the Gangwon, the littoral cells in the Gyeongbuk were initially set on a larger scale, which means surveys have not been conducted on all beaches. As a result, even if multiple beaches are included within a single littoral cell, only one or several beaches were often surveyed. Therefore, it is assumed that the surveyed beach represents the entire littoral cell in this study and all analyses were conducted based on this assumption.

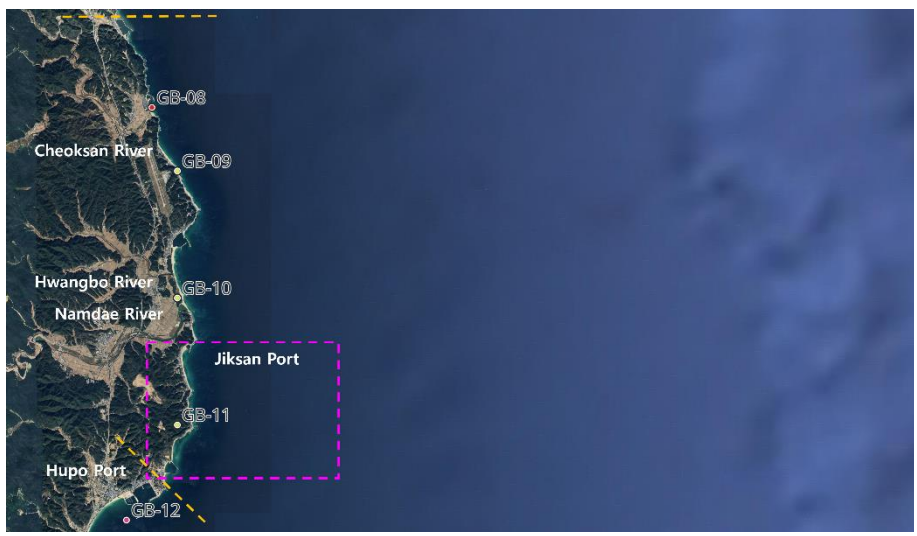


Figure 39. Satellite image of Littoral cell that contains Jiksan Port

Figure 40. Zoomed-in satellite image of Jiksan

Figures 39 and 40 are satellite images of Jiksan, Uljin-gun. As shown in Figure 39, there are no natural capes or structures along the coastline that restrict sediment movement, indicating that combining multiple littoral cells for new littoral cell is appropriate for effective management. Based on clustering result that was refined using satellite imagery and objective information, the littoral cell that includes Jiksan extends to the yellow dotted line encompassing GB-08. In addition, according to Kwak et al. (2005), the numerical modeling results predicting current direction near Hupo Port are divided into opposite direction, suggesting that between GB-11 and GB-12 should serve as boundary of the new littoral cell. Jiksan is represented by the area marked with a magenta dotted line in Figure 39 and Figure 40 is enlarged image of magenta dotted box.

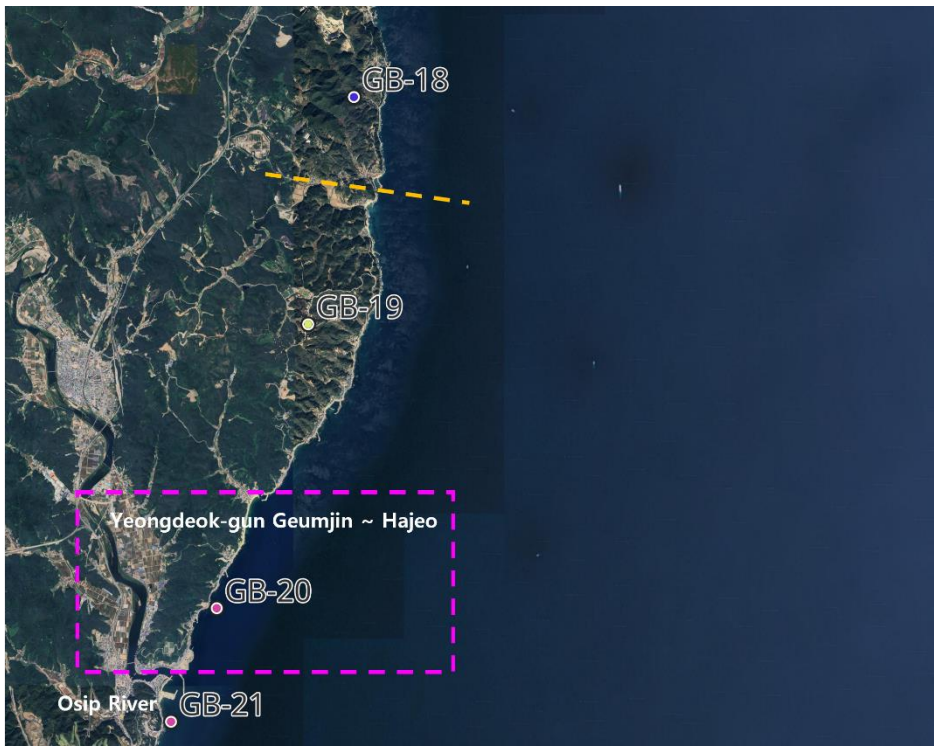


Figure 41. Satellite image of Littoral cell of Yeongdeok-gun Geumjin ~ Hajeo

Figure 42. Zoomed-in satellite image of near Osip River

Figure 41 is a satellite image of the Geumjin to Hajeo in Yeongdeok-gun. GB-19 and GB-20 were assigned to different clusters based on the clustering results. However, since the coastline directions are similar each other and there are no structures or natural capes restricting sediment flow, they were adjusted into the same cluster. Additionally, as shown in Figure 42, both GB-20 and GB-21 receive sediment supply from the Osip River. Therefore, when managing erosion in the Geumjin to Hajeo, GB-19, 20, and 21 should be considered together.

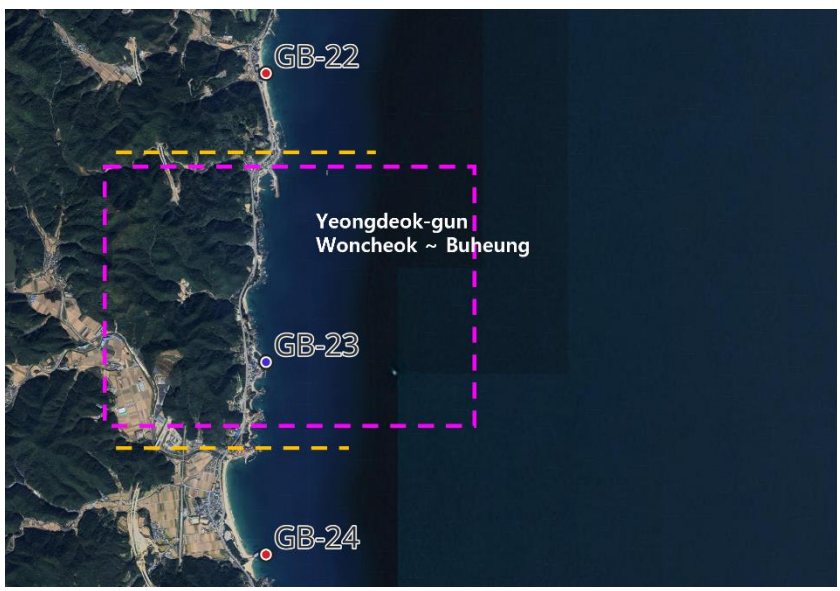


Figure 43. Satellite image of Littoral cell of Yeongdeok-gun Woncheok ~ Buheung

Figure 44. Zoomed-in satellite image of Yeongdeok-gun Woncheok ~ Buheung

Figures 43 and 44 are satellite images of the Woncheok to Buheung in Yeongdeok-gun. In Figure 43, the magenta dotted box indicates the Woncheok to Buheung area, which includes multiple beaches. However, only the area shown in Figure 44 was surveyed. While natural capes are well-developed, restricting sediment supply, erosion is likely to occur during high wave events that may affected its classification as a severely eroded area. Since natural capes are well-developed, the littoral cells can be delineated based on each cape, resulting in a more detailed segmentation compared to the original littoral cells.

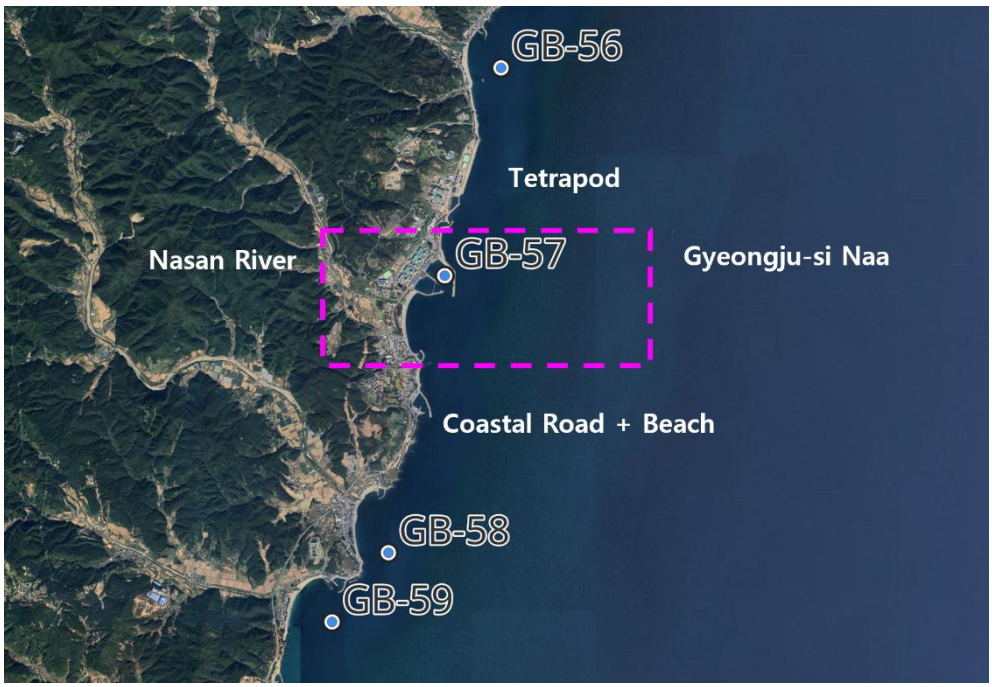


Figure 45. Satellite image of Littoral cell of Naa, Gyeongju-si

Figure 45 is a satellite image of Naa, Gyeongju-si, with the area outlined in magenta dotted lines representing Naa. According to the clustering results, it was assigned to the same cluster as GB-56, GB-58, and GB-59. Therefore, to manage Naa, it is necessary to manage the surrounding areas together that is different way from the past.

6. Conclusion

A littoral cell refers to an independent zone where there is no sediment exchange with adjacent littoral cells or where information about the inflow and outflow of sediment at the boundaries is clear. Therefore, a littoral cell is the minimum unit of a cyclic mass system influenced by the supply of sand through rivers and wave action. Korea's current littoral cell system is classified into macroscale, mesoscale, and unit littoral cells based on size and management purposes. Mesoscale littoral cells are divided based on watersheds, while unit littoral cells are divided based on natural formations such as capes or bays along the coastline, allowing for sediment budget analysis.

With the acceleration of coastal erosion due to climate change and human activities, the importance of sustainable and effective erosion management and response and the increased necessity for littoral cells have been emphasized. Including physical mechanisms that represent coastal characteristics in establishing littoral cell systems is expected to help manage the coast efficiently and stably and better explain various sediment transport phenomena occurring along the coast.

This study constructed a dataset comprising coastal environmental characteristics and sediment transport parameters for each unit littoral cell, followed by standardization and dimensionality reduction. Unit littoral cells with high similarity were clustered with HAC, and their distribution was evaluated by the Cophenetic correlation coefficient and the Silhouette score, with six clusters for

Gangwon and nine for BUG. Results were mapped with GIS to visualize each allocated unit littoral cell cluster and compared with previously divided littoral cells. This research examines the feasibility of defining mesoscale littoral cells, providing a theoretical basis for clustering unit littoral cells, and suggests management scope for coastal erosion 'Grade D'. The methodology and outcomes will contribute to setting effective, adaptive, and sustainable strategies for addressing coastal management.

Reference

곽문수, 이홍규, 이동수, & 편종근. (2005). 후포항 주변의 흐름장 해석. 대한토목학회 학술대회, 760-763.

Anderson, D., Ruggiero, P., Antolínez, J. A., Méndez, F. J., & Allan, J. (2018). A climate index optimized for longshore sediment transport reveals interannual and multidecadal littoral cell rotations. Journal of Geophysical Research: Earth Surface, 123(8), 1958-1981.

Anfuso, G., Benavente, J., Del Río, L., & Gracia, F. J. (2008). An approximation to short-term evolution and sediment transport pathways along the littoral of Cadiz Bay (SW Spain). Environmental geology, 56, 69-79.

Antolínez, J. A., Méndez, F. J., Anderson, D., Ruggiero, P., & Kaminsky, G. M. (2019). Predicting climate-driven coastlines with a simple and efficient multiscale model. Journal of Geophysical Research: Earth Surface, 124(6), 1596-1624.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. Pattern recognition, 46(1), 243-256.

Avnaim-Katav, S., Almogi-Labin, A., Herut, B., Kanari, M., & Guy-Haim, T. (2021). Benthic foraminifera from the Southeastern Mediterranean shelf: Dead assemblages and living-dead comparisons recording consequences of Nile River damming. Marine Micropaleontology, 164, 101977.

Bascom, W. N. (1951). The relationship between sand size and beach-face slope. Eos, Transactions American Geophysical Union, 32(6), 866-874.

Bowen, A. J., & Inman, D. L. (1966). Budget of littoral sands in the vicinity of Point Arguello, California (No. 19). US Army Coastal Engineering Research Center.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., ... & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature, 566(7745), 496-502.

Chamberlain, T. (1968). The littoral sand budget, Hawaiian Islands.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. Journal of statistical software, 61, 1-36.

Chou, C. H., Su, M. C., & Lai, E. (2004). A new cluster validity measure and its application to image compression. Pattern Analysis and Applications, 7, 205-220.

Cook, J., Sutskever, I., Mnih, A., & Hinton, G. (2007, March). Visualizing similarity data with a mixture of maps. In Artificial intelligence and statistics (pp. 67-74). PMLR.

Cooper, N. J., & Pontee, N. I. (2006). Appraisal and evolution of the littoral 'sediment cell'concept in applied coastal management: experiences from England and Wales. Ocean & coastal management, 49(7-8), 498-510.

Dean, R. G. (1991). Equilibrium beach profiles: characteristics and applications. Journal of coastal research, 53-84.

DEFRA, H. (2006). Shoreline management plan guidance. Volume 1: Aims and requirements.

DEAN, R. G. (1995). Cross-shore sediment transport processes. In Advances In Coastal And Ocean Engineering: (Volume 1) (pp. 159-220).

Dean, R. G., & Dalrymple, R. A. (2004). Coastal processes with engineering applications. Cambridge University Press.

Dean, R. G., & Galvin Jr, C. J. (1976). Beach erosion: causes, processes, and remedial measures. Critical Reviews in Environmental Science and Technology, 6(3), 259-296.

Díez, J., Cohn, N., Kaminsky, G. M., Medina, R., & Ruggiero, P. (2018). Spatial and temporal variability of dissipative dry beach profiles in the Pacific Northwest, USA. Journal of Coastal Research, 34(3), 510-523.

Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S., & Trapnell, C. (2020). Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nature communications, 11(1), 1537.

Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2(1), 20-30.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

El-Hamdouchi, A., & Willett, P. (1989). Comparison of hierarchic agglomerative clustering methods for document retrieval. The Computer Journal, 32(3), 220-227.

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743.

Foley, M. M., Warrick, J. A., Ritchie, A., Stevens, A. W., Shafroth, P. B., Duda, J. J., ... & Cubley, E. S. (2017). Coastal habitat and biological community response to

dam removal on the Elwha River. Ecological Monographs, 87(4), 552-577.

Folk, R. L., & Ward, W. C. (1957). Brazos River bar [Texas]; a study in the significance of grain size parameters. Journal of sedimentary research, 27(1), 3-26.

Gao, S., & Collins, M. B. (1994). Analysis of grain size trends, for defining sediment transport pathways in marine environments. Journal of Coastal Research, 70-78.

Gao, S., Collins, M. B., Lanckneus, J., De Moor, G., & Van Lancker, V. (1994). Grain size trends associated with net sediment transport patterns: An example from the Belgian continental shelf. Marine Geology, 121(3-4), 171-185.

Garzanti, E., Dinis, P., Vermeesch, P., Andò, S., Hahn, A., Huvi, J., ... & Vezzoli, G. (2018). Sedimentary processes controlling ultralong cells of littoral transport: Placer formation and termination of the Orange sand highway in southern Angola. Sedimentology, 65(2), 431-460.

George, D. A., Largier, J. L., Storlazzi, C. D., & Barnard, P. L. (2015). Classification of rocky headlands in California with relevance to littoral cell boundary delineation. Marine Geology, 369, 137-152.

Gere, A. (2023). Recommendations for validating hierarchical clustering in consumer sensory projects. Current Research in Food Science, 6, 100522.

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2020). Stochastic neighbor embedding with Gaussian and student-t distributions: Tutorial and survey. arXiv preprint arXiv:2009.10301.

Han, H., Li, W., Wang, J., Qin, G., & Qin, X. (2022). Enhance explainability of manifold learning. Neurocomputing, 500, 877-895.

Hinton, G. E., & Roweis, S. (2002). Stochastic neighbor embedding. Advances in

neural information processing systems, 15.

Hsu, J. R., & Evans, C. (1989). Parabolic bay shapes and applications. Proceedings of the Institution of Civil Engineers, 87(4), 557-570.

Hubert, L., & Arabie, P. (1985). Comparing partitions. Journal of classification, 2, 193-218.

Hurley, N. C., Haimovich, A. D., Taylor, R. A., & Mortazavi, B. J. (2022). Visualization of emergency department clinical data for interpretable patient phenotyping. Smart Health, 25, 100285.

Inman, D. L. (1952). Measures for describing the size distribution of sediments. Journal of Sedimentary Research, 22(3), 125-145.

Inman, D. L. (2003). Littoral cells.

Inman, D. L., & Frautschy, J. D. (1965, October). Littoral processes and the development of shorelines. ASCE.

Inman, D. L., & Jenkins, S. A. (1984). The Nile littoral cell and man's impact on the coastal zone of the southeastern Mediterranean. In Coastal Engineering 1984 (pp. 1600-1617).

Inman, D. L., & Masters, P. M. (1991). Coastal sediment transport concepts and mechanisms.

Inman, D. L., & Nordstrom, C. E. (1971). On the tectonic and morphologic classification of coasts. The Journal of geology, 79(1), 1-21.

Jain, A. K., & Dubes, R. C. (1988). Algorithms for clustering data. Prentice-Hall, Inc..

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

Kamphuis, J. W. (2003). Alongshore transport rate of sand. In Coastal Engineering 2002: Solving Coastal Conundrums (pp. 2478-2490).

Komar, P. D., & Inman, D. L. (1970). Longshore sand transport on beaches. Journal of geophysical research, 75(30), 5914-5927.

Krumbein, W. C. (1936). Application of logarithmic moments to size-frequency distributions of sediments. Journal of Sedimentary Research, 6(1), 35-47.

Lanzoni, S., & Seminara, G. (2002). Long-term evolution and morphodynamic equilibrium of tidal channels. Journal of Geophysical Research: Oceans, 107(C1), 1-1.

Lee, J. A., & Verleysen, M. (2007). Nonlinear dimensionality reduction (Vol. 1). New York: Springer.

Lim, C., Kim, T. K., Lee, S., Yeon, Y. J., & Lee, J. L. (2021). Assessment of potential beach erosion risk and impact of coastal zone development: a case study on Bongpo–Cheonjin Beach. Natural Hazards and Earth System Sciences, 21(12), 3827-3842.

Maimon, O., & Rokach, L. (Eds.). (2005). Data mining and knowledge discovery handbook (Vol. 2, No. 2005). New York: Springer.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Miller, J. K., & Dean, R. G. (2004). A simple new shoreline change model. Coastal Engineering, 51(7), 531-556.

Ministry of Oceans and Fisheries. (2023) Coastal Erosion Status Report

Ministry of Oceans and Fisheries. (2023). Cyclic adaptive coastal erosion management technology development

Ministry of the Environment, Japan. (2010). Sato-umi: A coastal sea area in which there is a harmonious coexistence of nature and human beings. http://www.env.go.jp/water/heisa/satoumi/common/satoumi_panf_e.pdf

Motyka, J. M., & Brampton, A. (1993). Coastal management: mapping of littoral cells.

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(1), 86-97.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. Journal of classification, 31, 274-295.

Nielsen, F. (2016). Introduction to HPC with MPI for Data Science. Springer.

Otto, G. H. (1939). A modified logarithmic probability graph for the interpretation of mechanical analyses of sediments. Journal of Sedimentary Research, 9(2), 62-76.

Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. Pattern recognition, 37(3), 487-501.

Pal, K., & Sharma, M. (2020, October). Performance evaluation of non-linear techniques UMAP and t-SNE for data in higher dimensional topological space. In 2020 fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC) (pp. 1106-1110). IEEE.

Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., ... & Murray, J. I. (2019). A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. Science, 365(6459), eaax1971.

Rahmawati, R. R., Putro, A. H. S., & Lee, J. L. (2021). Analysis of long-term shoreline observations in the vicinity of coastal structures: a case study of south Bali beaches. Water, 13(24), 3527.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association, 66(336), 846-850.

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of dimensionality reduction techniques on big data. Ieee Access, 8, 54776-54788.

Reef, R. E., Perry, P. V., & Wong, V. N. (2023). A new approach to mapping littoral cells within a tide-dominated embayment. Marine Geology, 461, 107068.

Romans, B. W., Normark, W. R., McGann, M. M., Covault, J. A., & Graham, S. A. (2009). Coarse-grained sediment delivery and distribution in the Holocene Santa Monica Basin, California: implications for evaluating source-to-sink flux at millennial time scales. Geological Society of America Bulletin, 121(9-10), 1394-1408.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.

Xu, R., Xu, J., & Wunsch, D. C. (2012). A comparison study of validity indices on swarm-intelligence-based clustering. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(4), 1243-1256.

Russell, R. D. (1955). Effects of transportation on sedimentary particles.

Sainburg, T., McInnes, L., & Gentner, T. Q. (2021). Parametric UMAP embeddings for representation and semisupervised learning. Neural Computation, 33(11), 2881-2907.

Sanderson, P. G., & Eliot, I. (1999). Compartmentalisation of beachface sediments along the southwestern coast of Australia. Marine Geology, 162(1), 145-164.

Saraçli, S., Doğan, N., & Doğan, İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of inequalities and Applications, 2013, 1-8.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. Neurocomputing, 267, 664-681.

Seid, M. A., & Mulatu Mengesha, Y. (2022). Intraspecific morphological variations among the populations of Milicia excelsa, Pouteria adolfi-friedericii, and Prunus africana in different natural forests of southwest Ethiopia. International Journal of Forestry Research, 2022(1), 9335428.

Sickmann, Z. T., Paull, C. K., & Graham, S. A. (2016). Detrital-zircon mixing and partitioning in fluvial to deep marine systems, central California, USA. Journal of Sedimentary Research, 86(11), 1298-1307.

Sneath, P. H., & Sokal, R. R. (1973). Numeral taxonomy. WH: Freemam, San Francisco, California.

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. Taxon, 33-40.

Splinter, K. D., Davidson, M. A., Golshani, A., & Tomlinson, R. (2012). Climate controls on longshore sediment transport. Continental Shelf Research, 48, 146-156.

Steinbach, M. (2000). A Comparison of Document Clustering Techniques. Technical Report# 00_034/University of Minnesota.

Stive, M. J., Aarninkhof, S. G., Hamm, L., Hanson, H., Larson, M., Wijnberg, K. M., ... & Capobianco, M. (2002). Variability of shore and shoreline evolution. Coastal engineering, 47(2), 211-235.

Trozzi, F., Wang, X., & Tao, P. (2021). UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: A comparison study. The Journal of Physical Chemistry B, 125(19), 5022-5034.

Utizi, K., Corbau, C., Rodella, I., Nannini, S., & Simeoni, U. (2016). A mixed solution for a highly protected coast (Punta Marina, Northern Adriatic Sea, Italy). Marine Geology, 381, 114-127.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Van Der Maaten, L., Postma, E. O., & Van Den Herik, H. J. (2009). Dimensionality reduction: A comparative review. Journal of machine learning research, 10(66-71), 13.

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(201), 1-73.

Willett, P. (1988). Recent trends in hierarchic document clustering: a critical review. Information processing & management, 24(5), 577-597.

Wright, L. D., Short, A. D., & Green, M. O. (1985). Short-term changes in the morphodynamic states of beaches and surf zones: an empirical predictive model. Marine geology, 62(3-4), 339-364.

Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. Annals of data science, 2, 165-193.

Yates, M. L., Guza, R. T., & O'reilly, W. C. (2009). Equilibrium shoreline response: Observations and modeling. Journal of Geophysical Research: Oceans, 114(C9).

Zhao, Y., & Karypis, G. (2002, November). Evaluation of hierarchical clustering algorithms for document datasets. In Proceedings of the eleventh international conference on Information and knowledge management (pp. 515-524).

Appendix. A

Python Code for the analysis

```
import numpy as np
from sklearn.preprocessing import StandardScaler, LabelEncoder
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import tensorflow as tf
import random
from tensorflow.keras.layers import Input, Embedding, Flatten
from tensorflow.keras.models import Model
import umap
from scipy.cluster.hierarchy import linkage, cophenet, dendrogram, fcluster
from scipy.spatial.distance import pdist
from sklearn.metrics import silhouette_score

# Set random seeds for reproducibility
tf.random.set_seed(42)
np.random.seed(42)
random.seed(42)

# Load Data (User should replace 'data.xlsx' with their own dataset file)
file_path = 'path_to_your_data.xlsx'  # Replace with your data file path
sheet_name = 'Sheet1'  # Replace with your sheet name if necessary
factor = pd.read_excel(file_path, sheet_name=sheet_name, header=1)

# Drop missing values
factor = factor.dropna()

#%% Encode categorical data (example: attribute of sand)
label_encoder = LabelEncoder()
factor['sand_label'] = label_encoder.fit_transform(factor['attribute of sand'])
```

```
# Create embedding model for encoded categorical feature
input_dim = factor['sand_label'].nunique()
output_dim = min(50, input_dim // 2)

input_layer = Input(shape=(1,))
embedding_layer = Embedding(input_dim=input_dim, output_dim=output_dim,
input_length=1)(input_layer)
flatten_layer = Flatten()(embedding_layer)

embedding_model = Model(inputs=input_layer, outputs=flatten_layer)
embedding_model.compile(optimizer='adam', loss='mse')
embedding_model.fit(factor['sand_label'], np.zeros((factor.shape[0], output_dim)),
epochs=50, verbose=0)

embedding_matrix = embedding_model.predict(factor['sand_label'])

# Reset factor and drop temporary label column
factor_reset = factor.reset_index(drop=True).drop(columns=['sand_label'])

# Select numeric columns
numeric_cols = factor_reset.select_dtypes(include=[np.number]).columns.tolist()

# Concatenate embeddings with other numerical data
full_data = pd.concat(
    [factor_reset[numeric_cols],
     pd.DataFrame(embedding_matrix, columns=[f'attribute_{i}' for i in
range(output_dim)])], axis=1
)

#%% Apply UMAP for dimensionality reduction
reducer = umap.UMAP(random_state=0, n_components=2)
factor_data = full_data.values
scaled_data = StandardScaler().fit_transform(factor_data)
embedding = reducer.fit_transform(scaled_data)
```

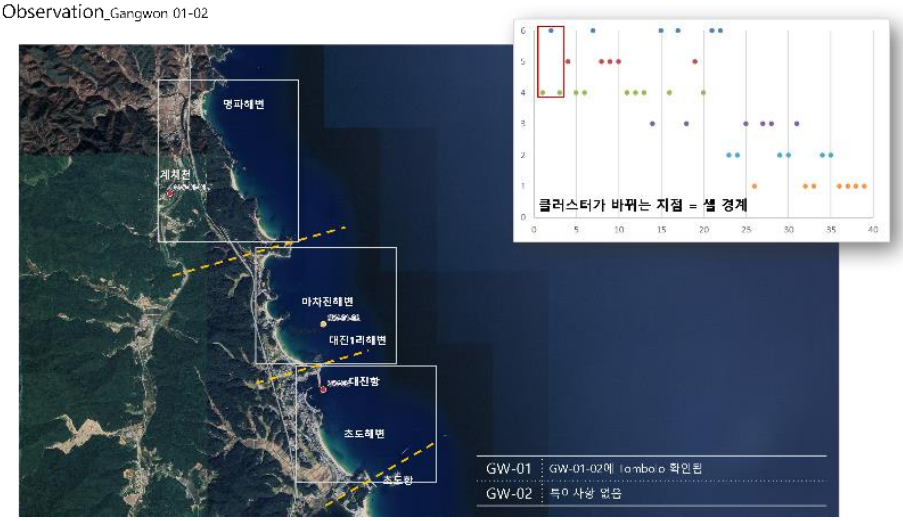
```
# Plot UMAP result (generalized for categorical groups)
plt.figure(figsize=(20, 14))
sns.scatterplot(x=embedding[:, 0], y=embedding[:, 1], s=130, edgecolor='w')
plt.title('Dimensionality Reduction Result by UMAP')
plt.xticks([])
plt.yticks([])
plt.show()

#%% Calculate and display cophenetic correlation for linkage and distance metrics
linkage_methods = ['single', 'complete', 'average', 'weighted', 'ward']
distance_metrics = ['euclidean', 'cityblock', 'chebyshev', 'cos']
cph_results = []

for method in linkage_methods:
    for metric in distance_metrics:
        try:
            Z = linkage(embedding, method=method, metric=metric)
            c, _ = cophenet(Z, pdist(embedding, metric=metric))
            cph_results.append([method, metric, round(c, 2)])
        except Exception as e:
            cph_results.append([method, metric, np.nan])

cph_df = pd.DataFrame(cph_results, columns=['linkage_method', 'distance_metric',
'Cophenetic Correlation'])

#%% Dendrogram Plot (generalized, example linkage: complete and euclidean)
plt.figure(figsize=(10, 9))
dendro = dendrogram(linkage(embedding, method='method', metric='metric'))
plt.title("Hierarchical Clustering Dendrogram")
plt.xlabel('Data Points')
plt.show()

#%% Silhouette analysis for optimal clusters(example: )
silhouette_avgs = []
for n_clusters in range(2, 10):
```

```
    cluster_labels = fcluster(linkage(embedding, method='complete',
metric='euclidean'), n_clusters, criterion='maxclust')
    silhouette_avg = silhouette_score(embedding, cluster_labels)
    silhouette_avgs.append((n_clusters, silhouette_avg))
    print(f"Clusters: {n_clusters}, Silhouette Score: {silhouette_avg:.4f}")

# Plot silhouette scores
plt.figure(figsize=(20, 8))
plt.plot([x[0] for x in silhouette_avgs], [x[1] for x in silhouette_avgs], marker='o')
plt.title('Silhouette Scores by Number of Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('Average Silhouette Score')
plt.grid(True)
plt.show()

#%% Assign clusters and save to Excel
cluster_labels = fcluster(linkage(embedding, method='complete',
metric='euclidean'), 2, criterion='maxclust')
full_data['Cluster'] = cluster_labels
output_file_path = 'generalized_clustered_data.xlsx'
full_data.to_excel(output_file_path, index=False)

print(f"Clustered data saved to {output_file_path}")
```

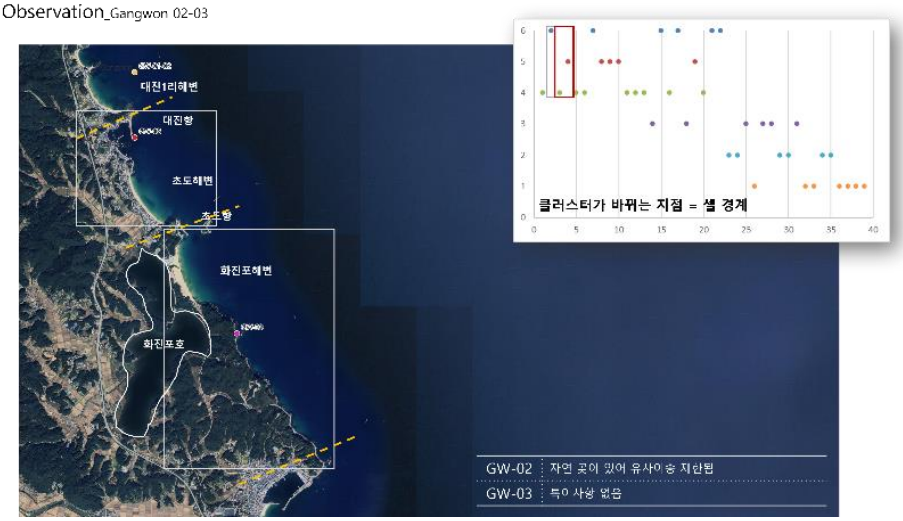
Appendix. B

Gangwon

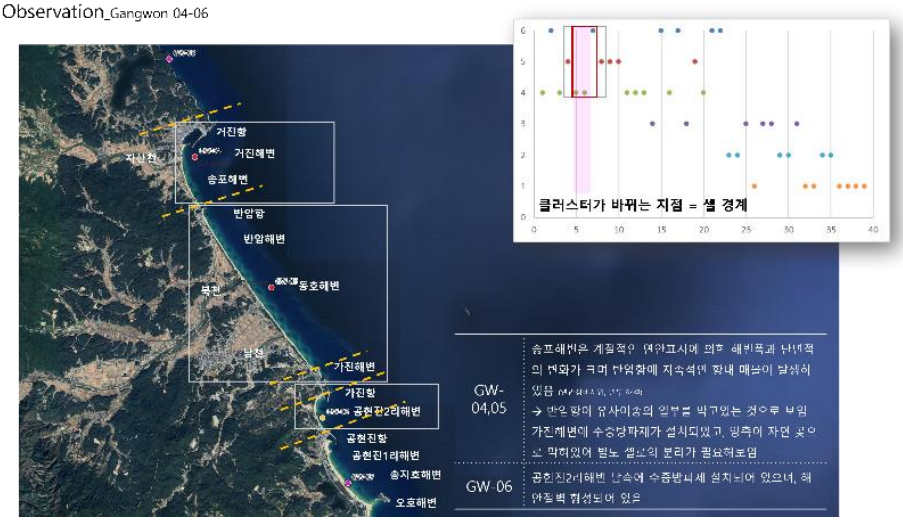
Close Observation_Gangwon 01-02



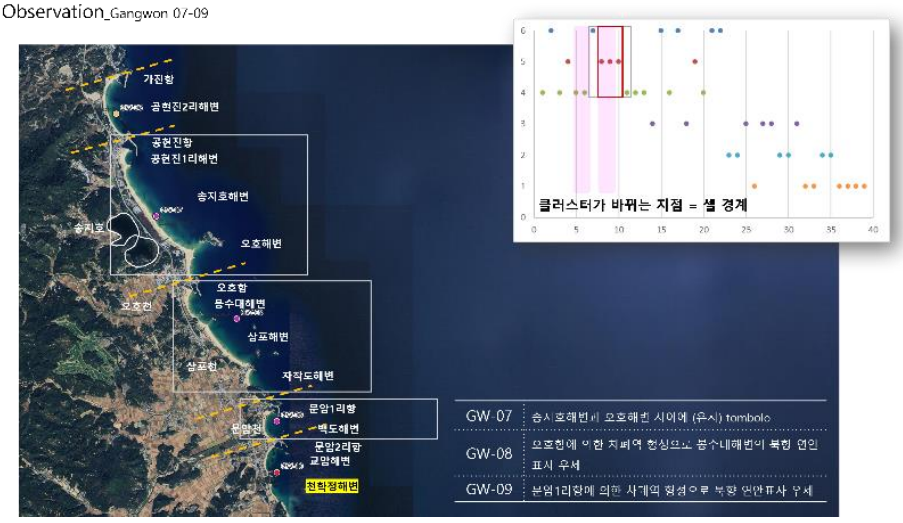
Close Observation_Gangwon 02-03



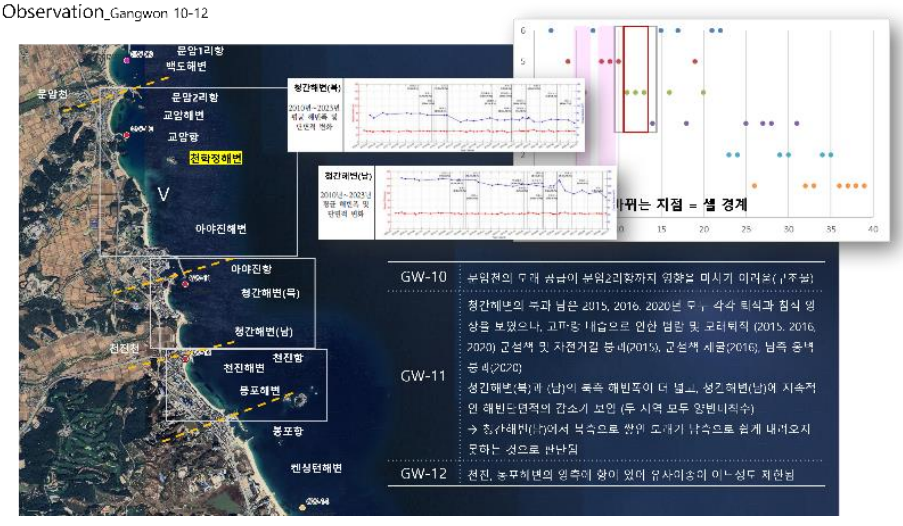
Close Observation_Gangwon 04-06



Close Observation_Gangwon 07-09

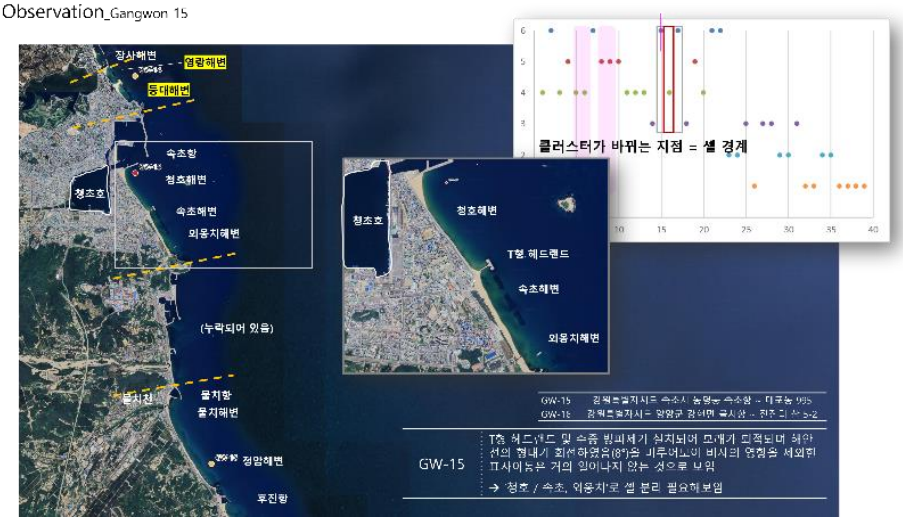


Close Observation_Gangwon 10-12

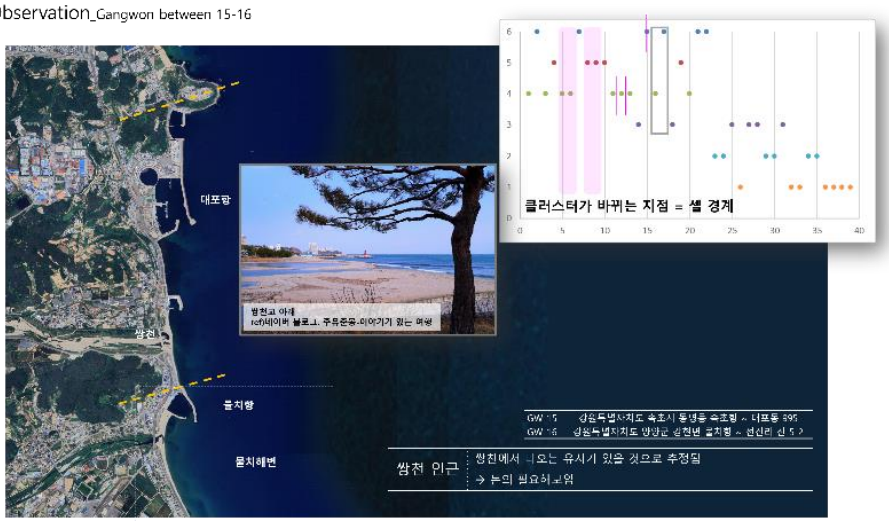


Close Observation_Gangwon 13-14

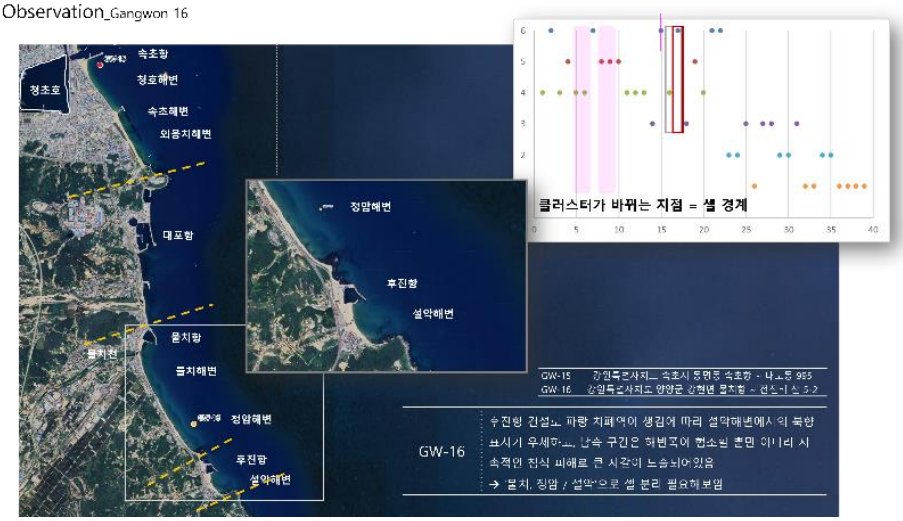
Close Observation_Gangwon 15



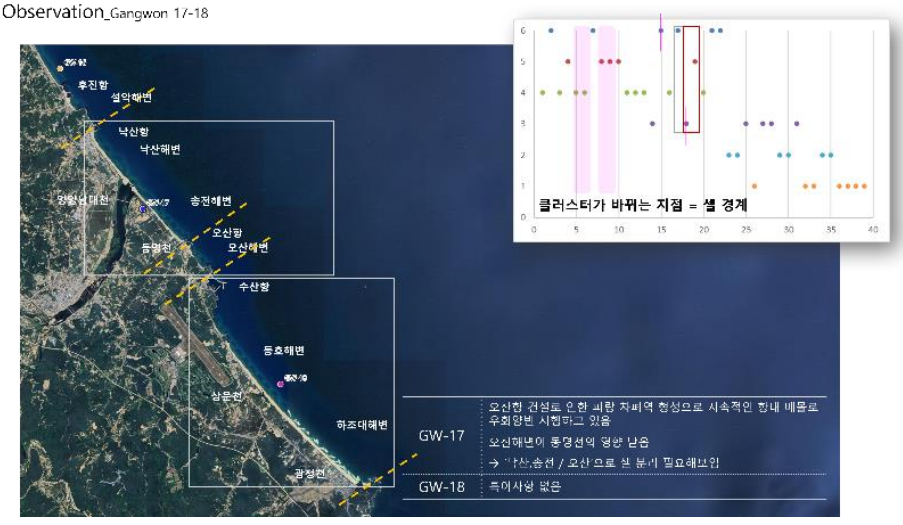
Close Observation_Gangwon between 15-16



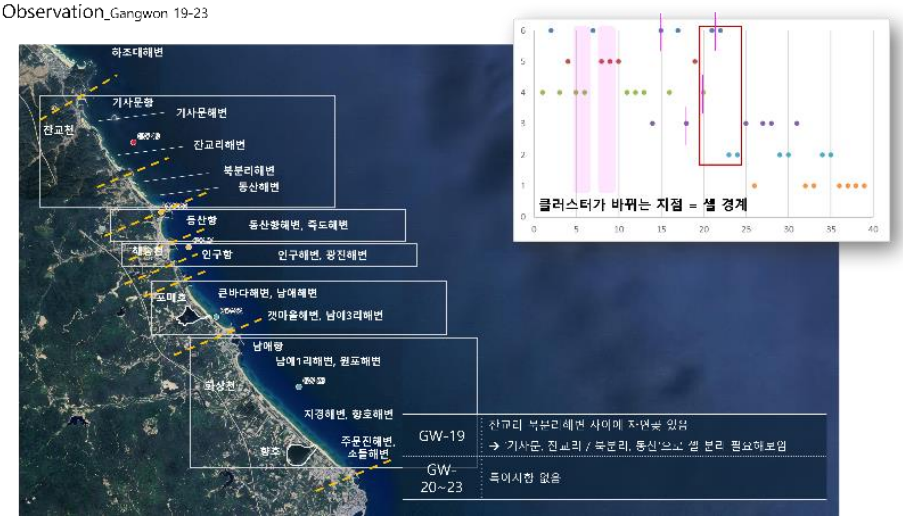
Close Observation_Gangwon 16



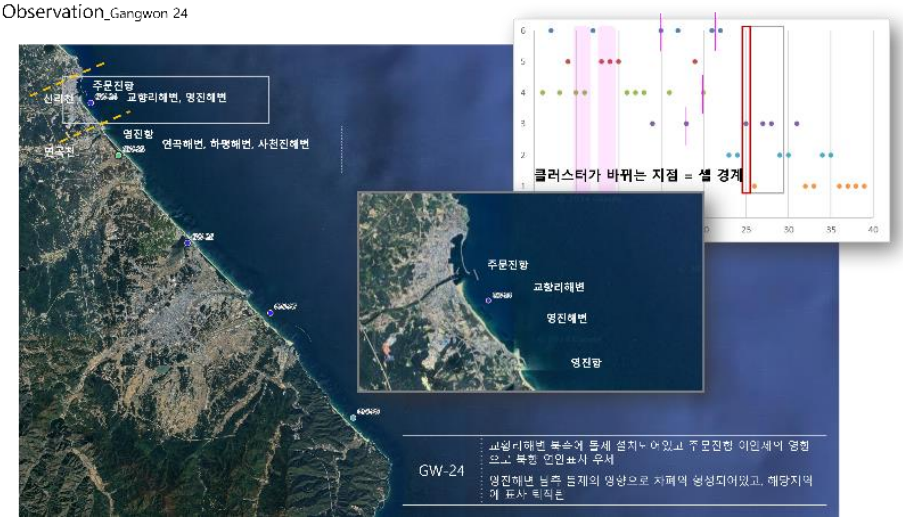
Close Observation_Gangwon 17-18



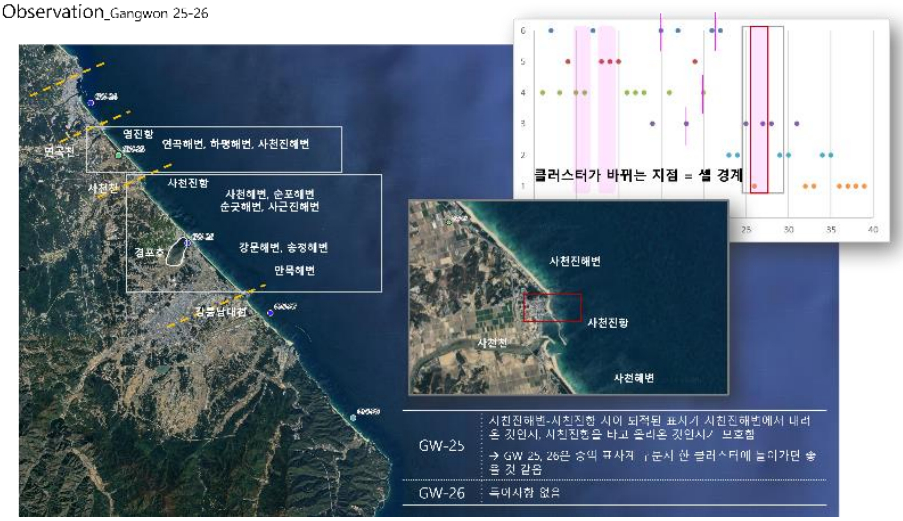
Close Observation_Gangwon 19-23



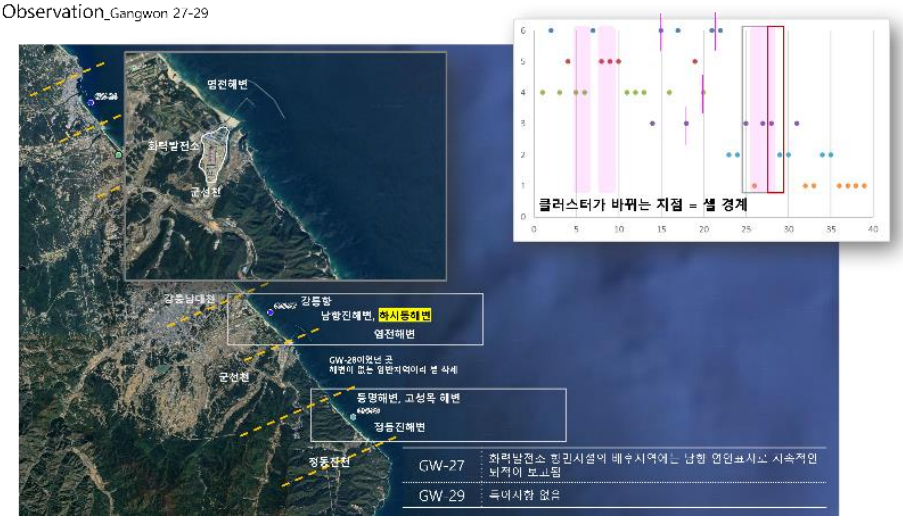
Close Observation_Gangwon 24



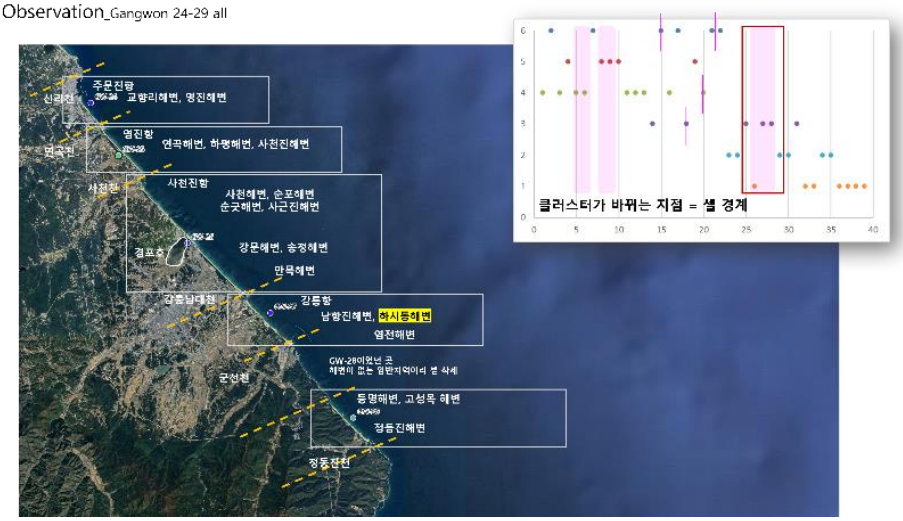
Close Observation_Gangwon 25-26



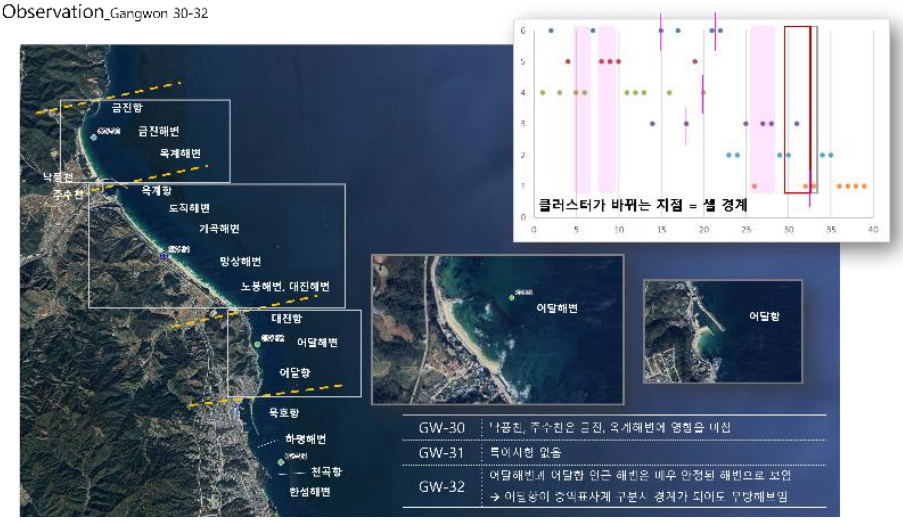
Close Observation_Gangwon 27-29



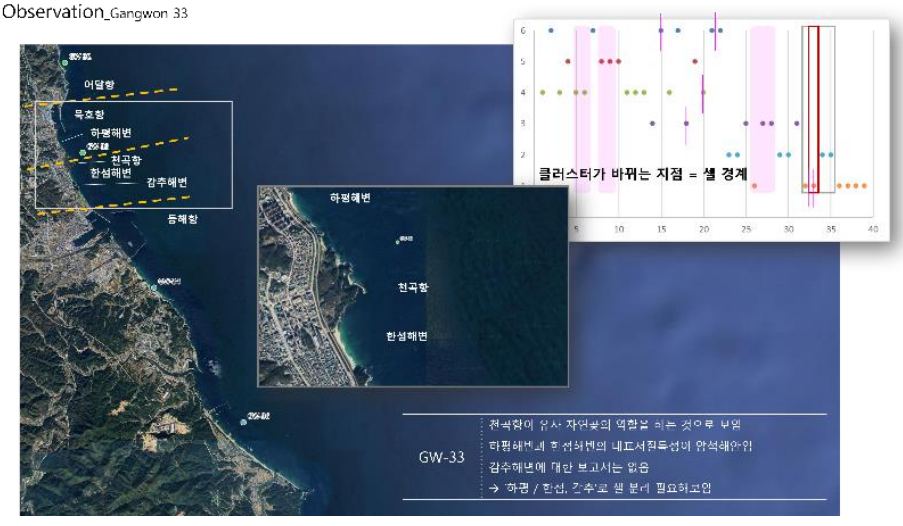
Close Observation_Gangwon 24-29 all



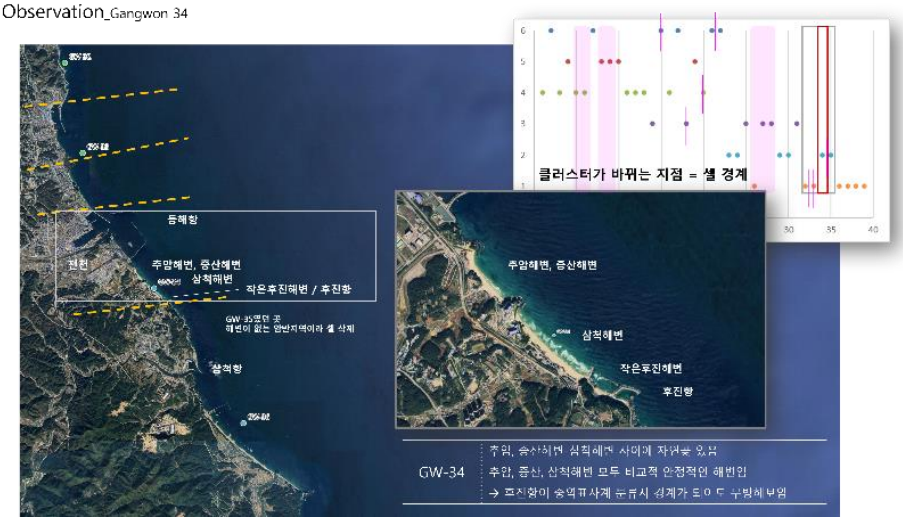
Close Observation_Gangwon 30-32



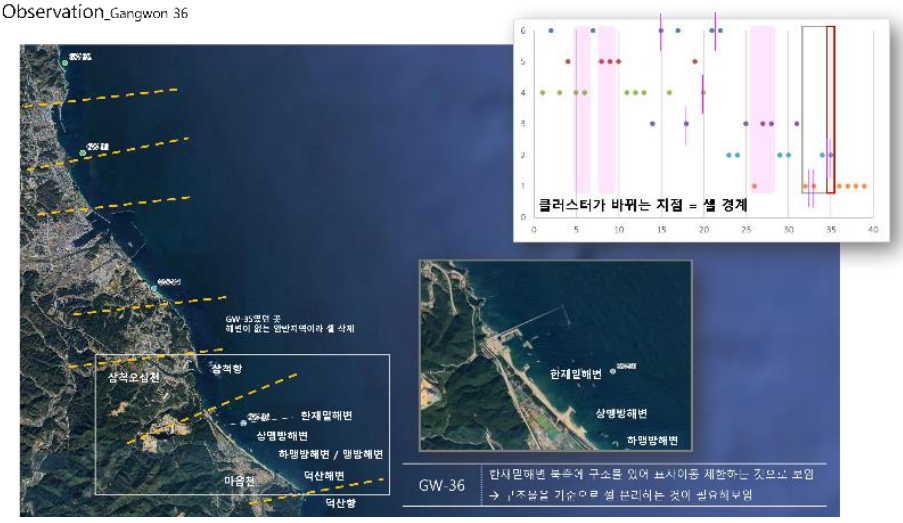
Close Observation_Gangwon 33



Close Observation_Gangwon 34

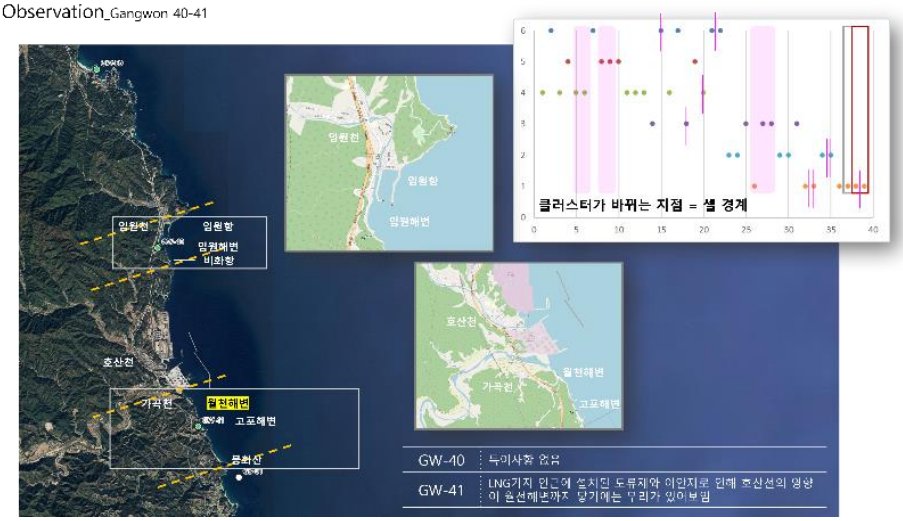


Close Observation_Gangwon 36



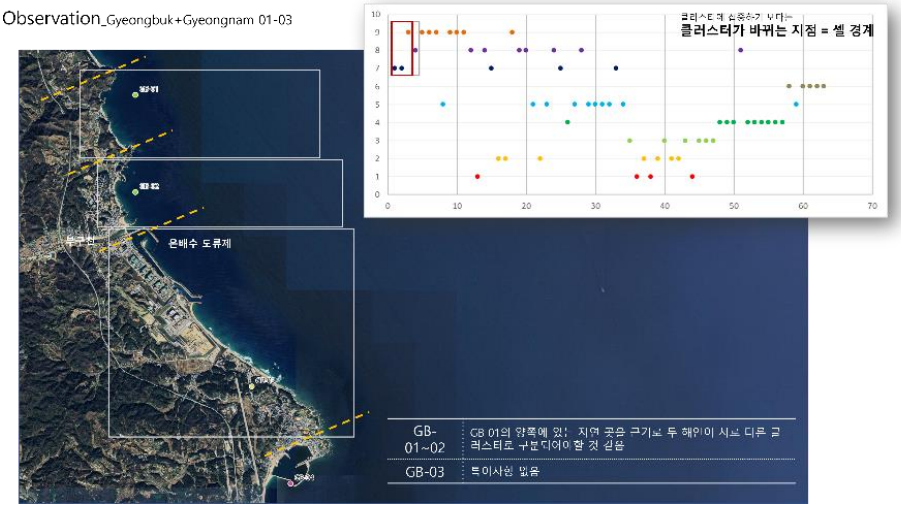
Close Observation_Gangwon 38-39

Close Observation_Gangwon 40-41

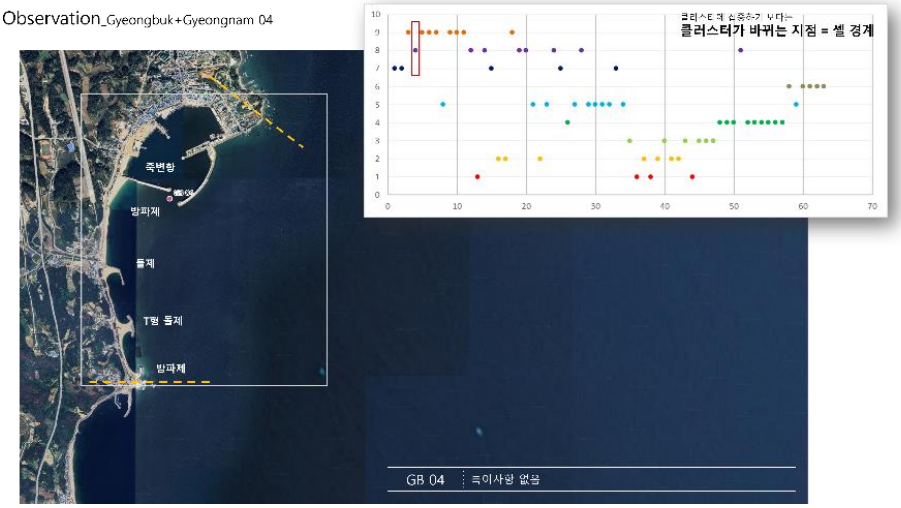


BUG

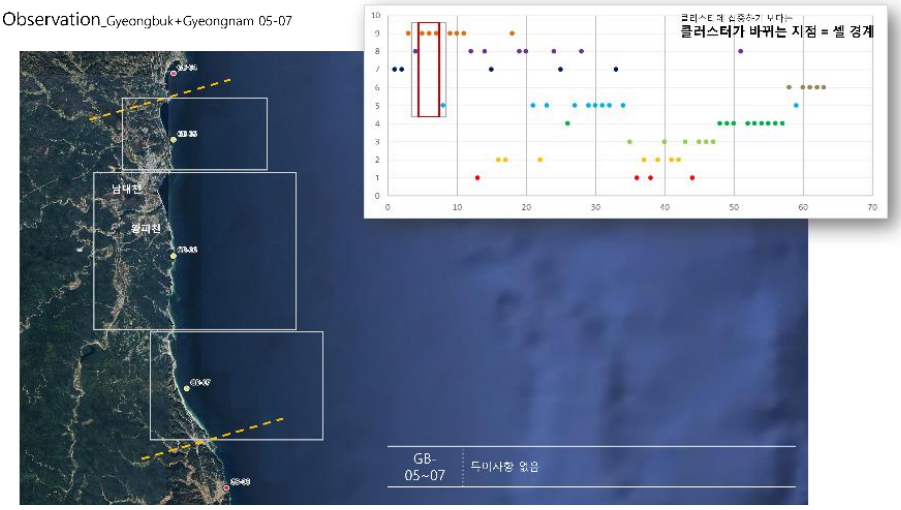
Close Observation_Gyeongbuk+Gyeongnam 01-03



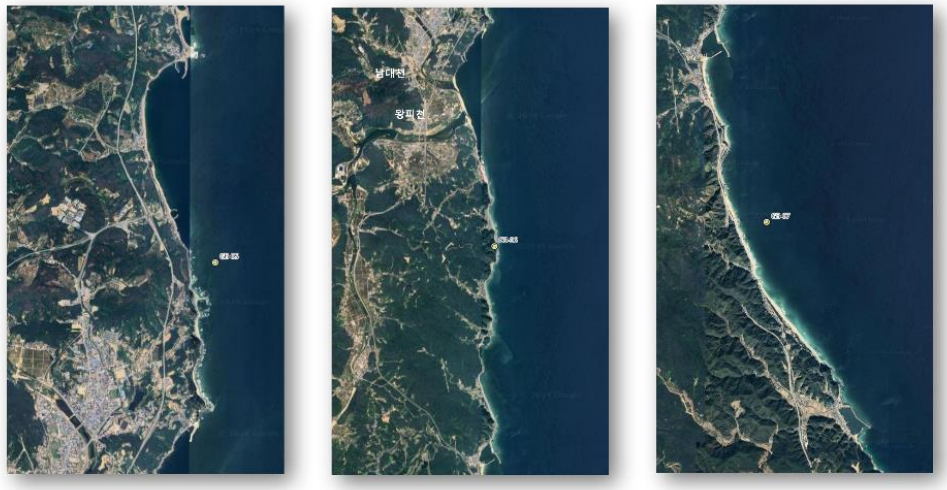
Close Observation_Gyeongbuk+Gyeongnam 04



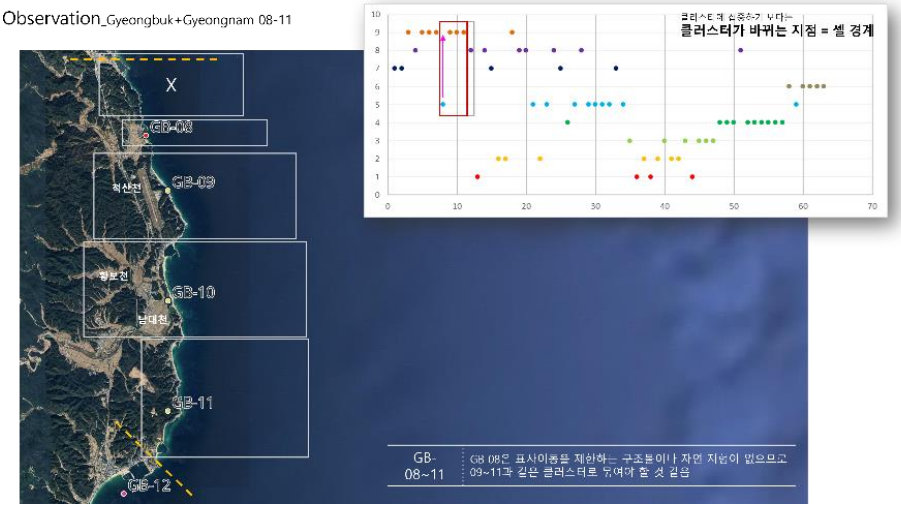
Close Observation_Gyeongbuk+Gyeongnam 05-07



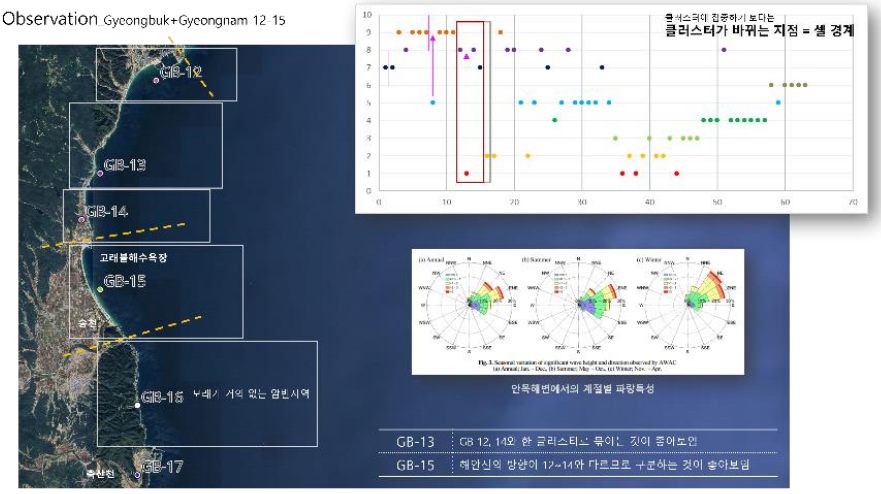
Close Observation_Gyeongbuk+Gyeongnam 05-07 (from left)



Close Observation_Gyeongbuk+Gyeongnam 08-11

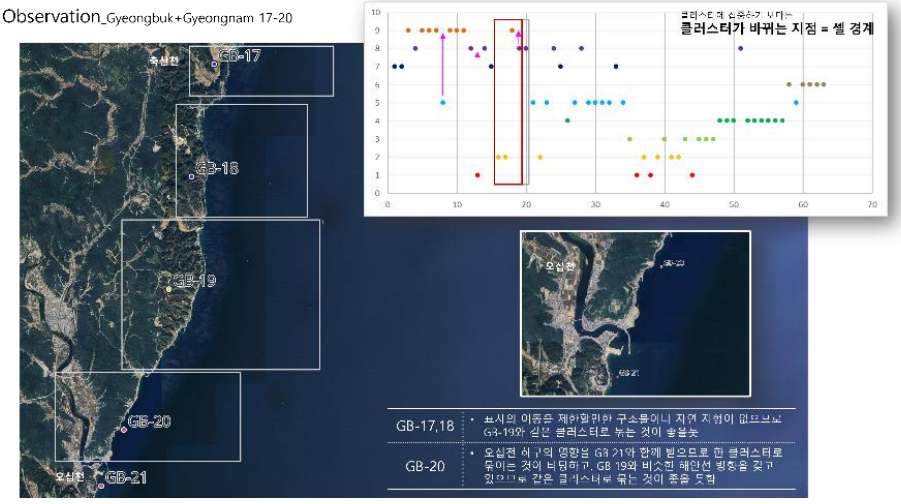


Close Observation_Gyeongbuk+Gyeongnam 12-15

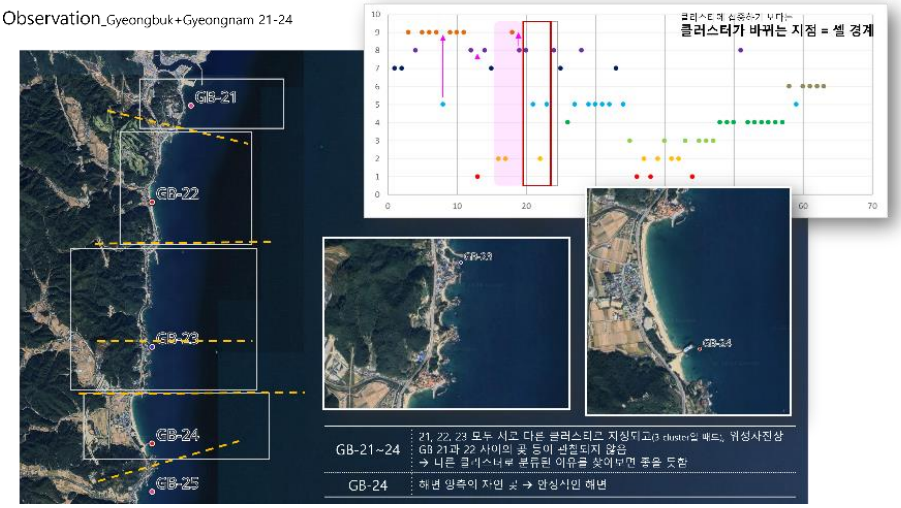


Kim, M.J., Son, D.H., Yoo, J.S., 2018. Analysis of Seasonal Morphodynamic Patterns using Delft3D in Anmok Coast. Journal of Coastal Disaster Prevention, 5(4), 183-192. http://doi.org/10.20481/kscdp.2018.5.4.183

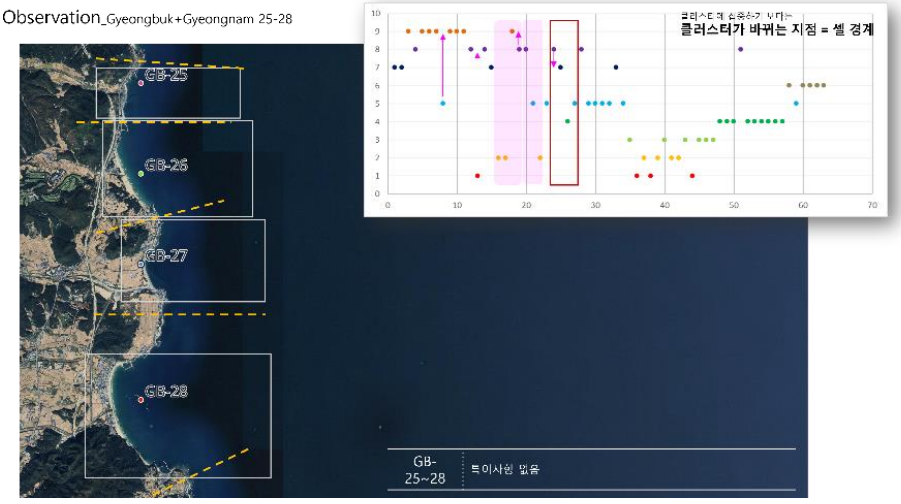
Close Observation_Gyeongbuk+Gyeongnam 17-20



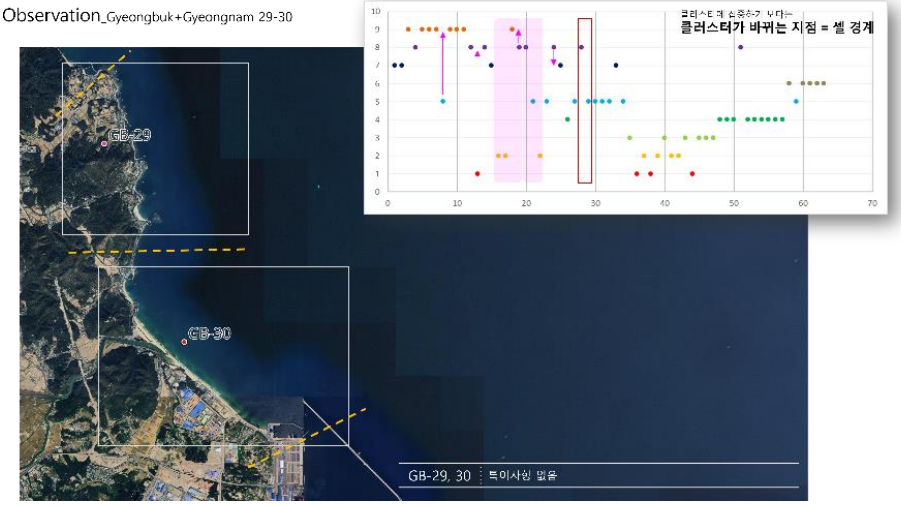
Close Observation_Gyeongbuk+Gyeongnam 21-24



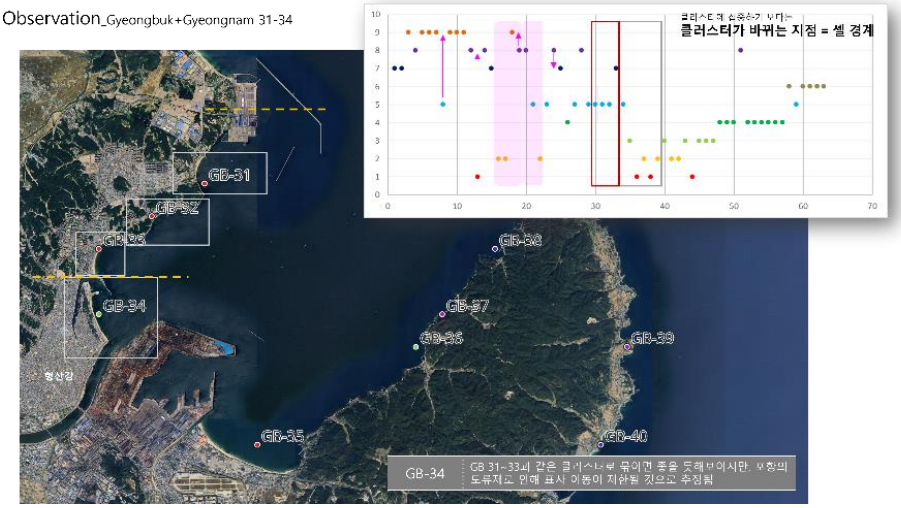
Close Observation_Gyeongbuk+Gyeongnam 25-28



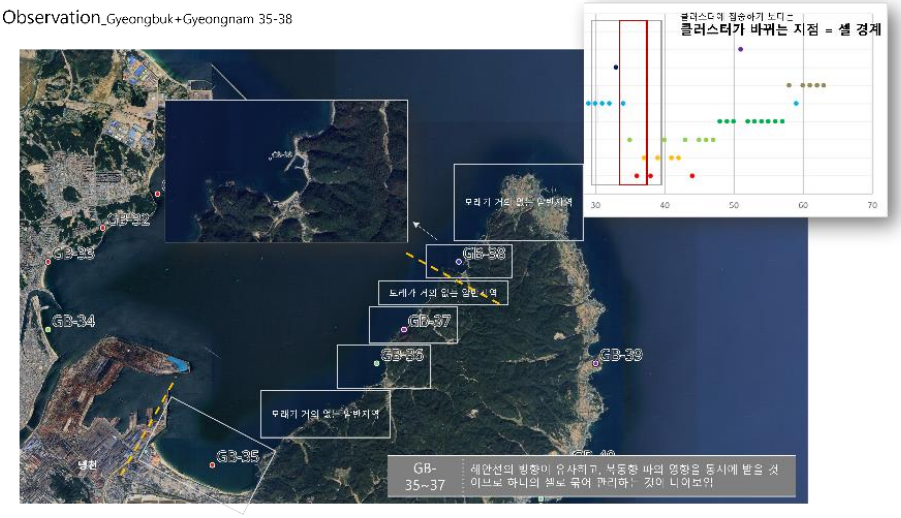
Close Observation_Gyeongbuk+Gyeongnam 29-30



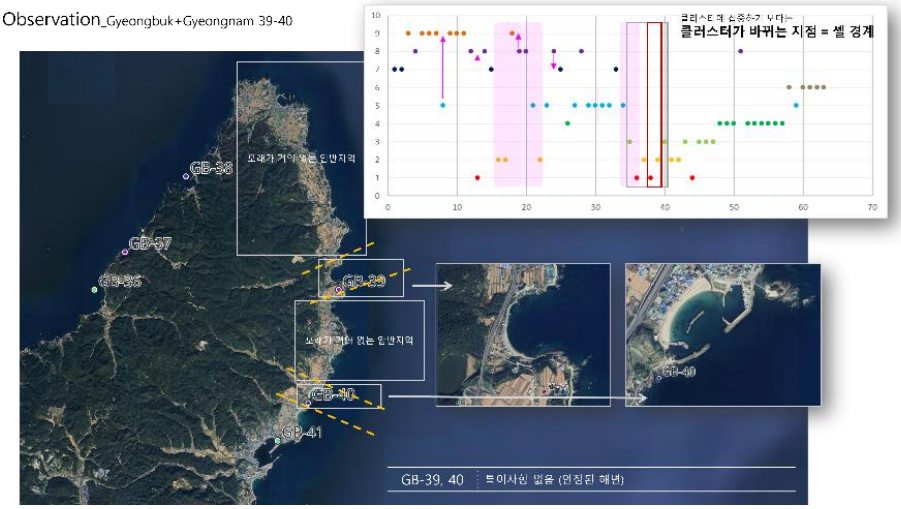
Close Observation_Gyeongbuk+Gyeongnam 31-34



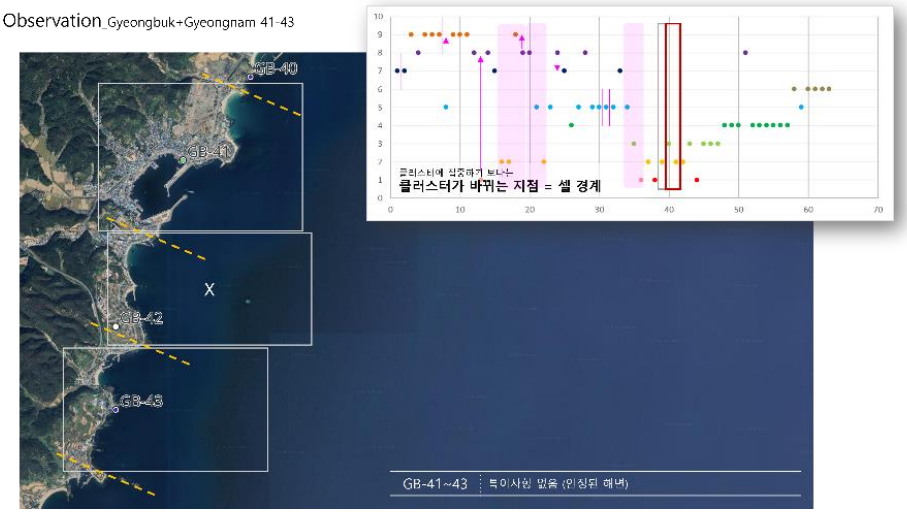
Close Observation_Gyeongbuk+Gyeongnam 35-38



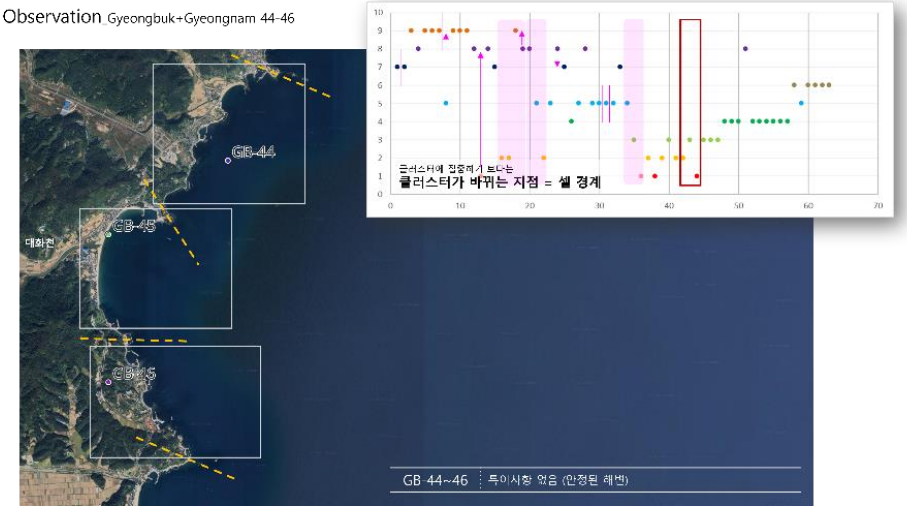
Close Observation_Gyeongbuk+Gyeongnam 39-40



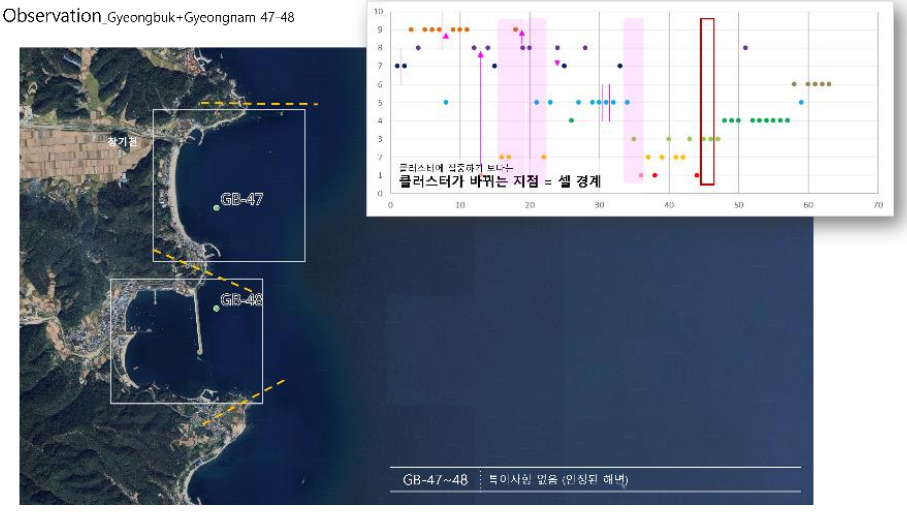
Close Observation_Gyeongbuk+Gyeongnam 41-43



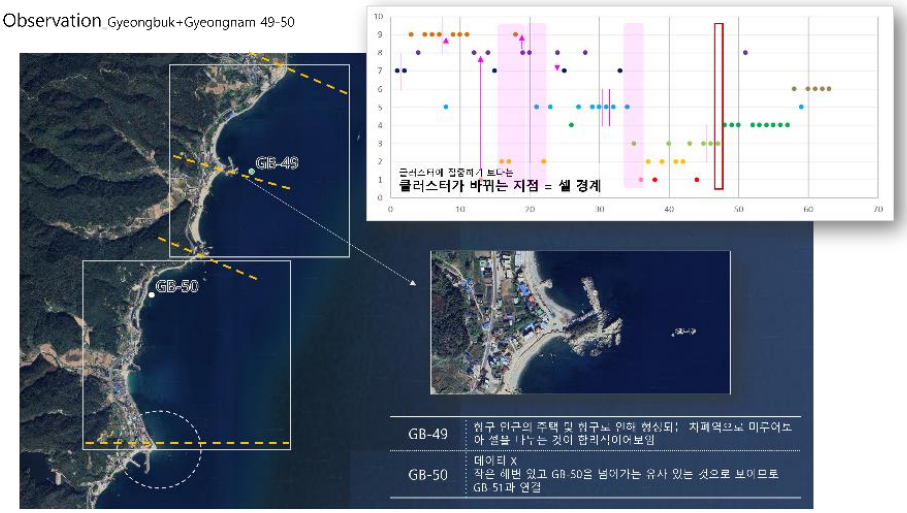
Close Observation_Gyeongbuk+Gyeongnam 44-46



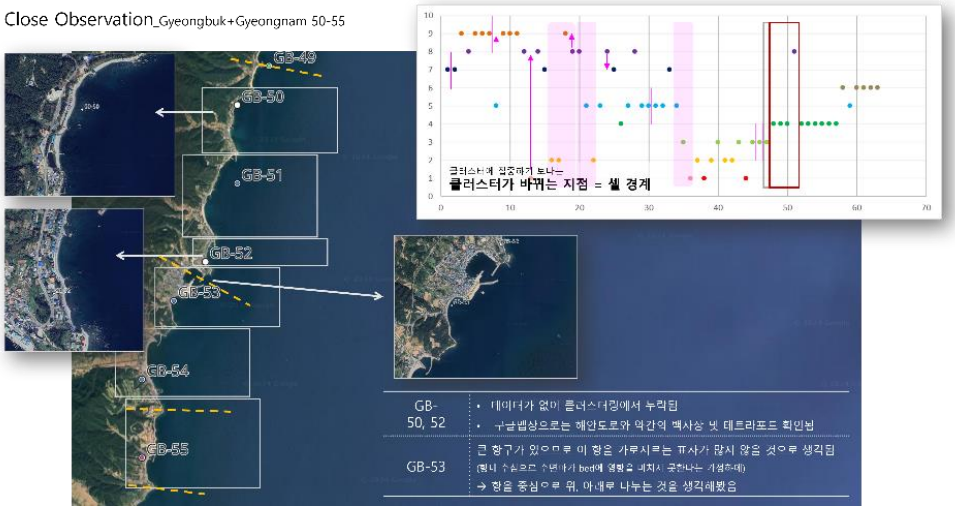
Close Observation_Gyeongbuk+Gyeongnam 47-48



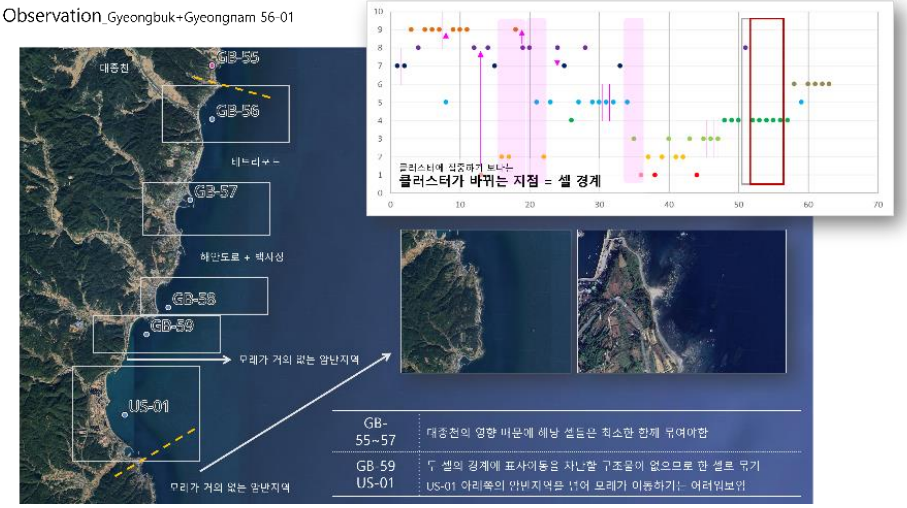
Close Observation_Gyeongbuk+Gyeongnam 49-50



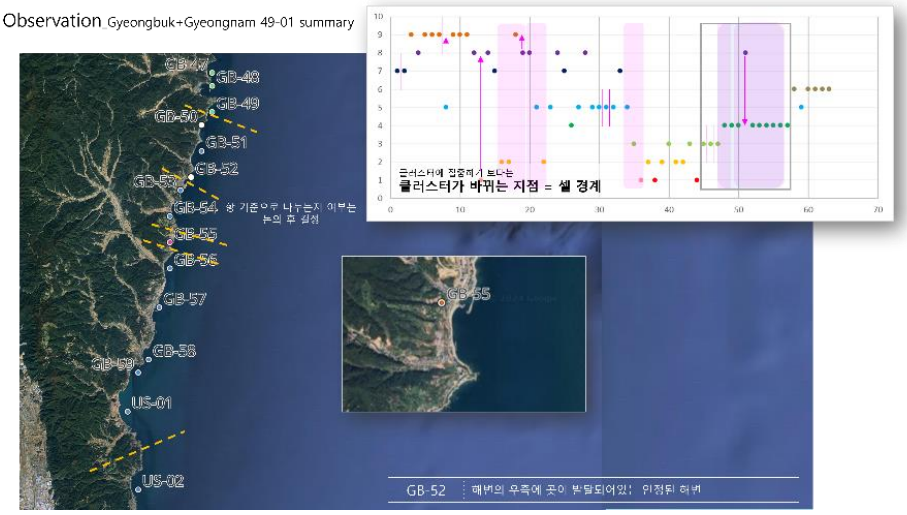
Close Observation_Gyeongbuk+Gyeongnam 50-55



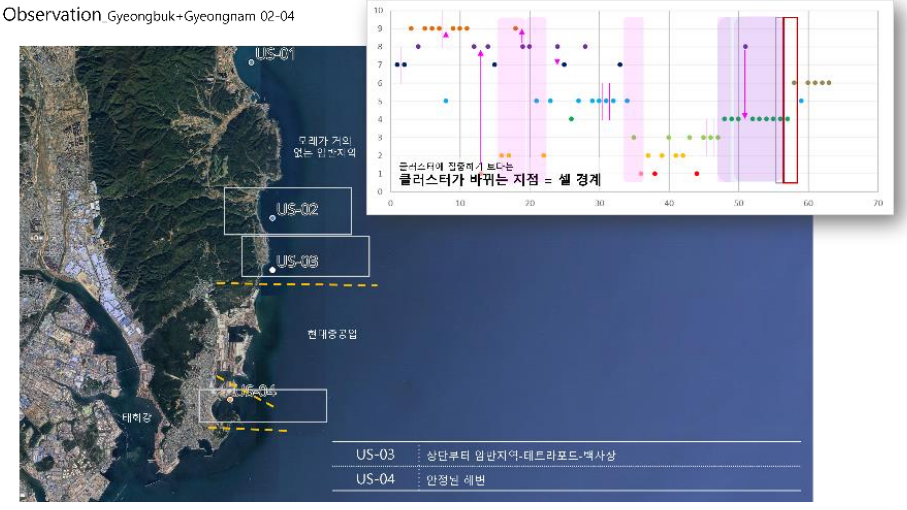
Close Observation_Gyeongbuk+Gyeongnam 56-01



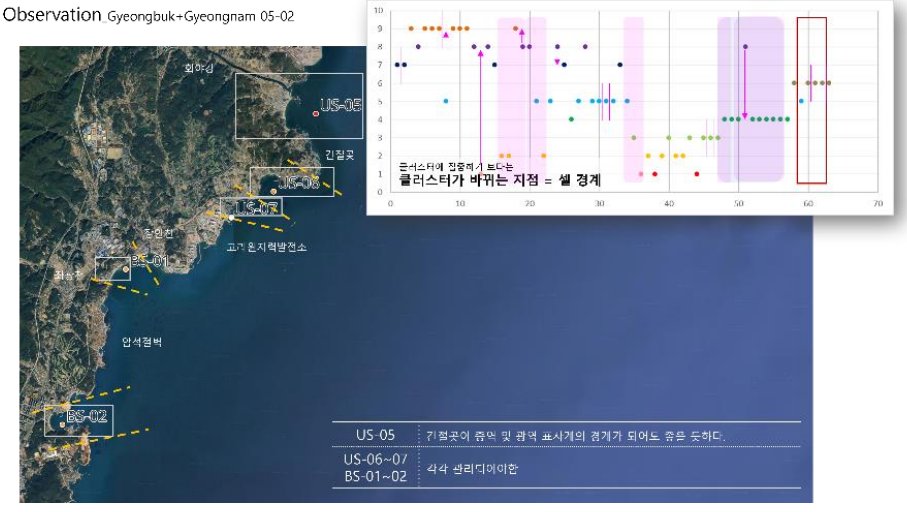
Close Observation_Gyeongbuk+Gyeongnam 49-01 summary



Close Observation_Gyeongbuk+Gyeongnam 02-04



Close Observation_Gyeongbuk+Gyeongnam 05-02



국문초록

비지도 기계학습을 활용한 연안 유사이송 매개변수 기반 대한민국 표사계의 경계 설정

서울대학교 대학원

건설환경공학부

김 도 현

기후 변화와 인간활동으로 인해 연안침식이 가속화됨에 따라 지속 가능하고 효율적인 침식 관리 및 대응의 필요성이 증대되었다. 표사계 단위로 연안을 구분하여 관리하는 것은 효율적인 연안 관리 방법 중 하나이며, 국부적 관리와 이해관계자 간 협력을 용이하게 하고 대상지에 특화된 국지적 연구를 활성화하는 등의 이점이 있다. 현재 국내의 표사계는 크기와 관리 목적에 따라 광역 표사계, 중역 표사계, 단위 표사계의 세 단계로 나뉘며, 특히 강원도권 표사계의 경우 중역 표사계는 유역을 중심으로, 단위 표사계는 연안에 위치한 자연 곶이나 돌출된 암석해안을 기준으로 표사 수지분석이 가능하도록 상세하게 나누어져 있다. 본 연구는 표사계 설정에 있어서 새로운 방법론을 정립하기 위한 기초연구로서 연안 유사 이송 변수와 비지도 기계학습을 활용하여 동해안의 단위표사계를 군집화한 결과를 해석하고, 해당 결과로 중역표사계의 경계를 설정하는 것이 가능한지 탐구하였다.

유사이송 매개변수로 위도, 경도, 중앙입경, 해빈폭, 해안선 길이, 대표저질특성, 전빈경사, 해안형태를 취득하였다. 차원 축소를 위해 비선형 차원

축소 기법 중 하나인 UMAP 을 채택하였으며, 데이터 간 상대적 거리 구조가 잘 보존되도록 차원 축소된 결과를 얻었다. 이후, 평균제곱오차 값에 대하여 원자료를 무작위로 임베딩하였을 때와 비교하여 UMAP 의 차원 축소 결과가 신뢰할만함을 증명하였다. 연안의 크기와 관리 목적에 따라 표사계의 크기를 쉽게 설정할 수 있도록 군집간 계층적 관계가 덴드로그램으로 나타나는 계층적 병합형 군집화 기법을 선택하였다. 군집화 결과는 성능 평가를 거쳐 GIS DB 로 저장된 후 그 결과를 알기 쉽게 시각화하였다. 결과 해석을 위해 유사도가 높은 개체끼리 동일한 군집으로 분류된다는 군집화의 특징을 활용하였는데, 결과 분석시 해안선을 따라 단위 표사계를 차례로 나열하였을 때 할당된 군집이 서로 다른 두 인접한 표사계의 사이를 표사계간 경계로 두며 표사 이동에 경계가 생기는 지점이라 해석하였다. 또, 위성 영상으로 확인되는 자연 곶이나 인공 구조물 및 하천의 영향 등을 종합적으로 고려하여 모든 군집화 결과를 재검토하였다. 검토 결과는 연안침식 실태조사 종합보고서에 보고된 연안침식 등급 ‘D’ 지역에 대한 연안 관리 범위를 제안하는 데에 활용되었다. 본 연구의 방법론은 국내 표사계 설정시 실용 표사계와 함께 이론적 기반으로 활용되어, 효율적인 해안 지형 변화 예측 및 대응 방안 수립에 기여할 것으로 기대된다.

주요어: 연안 침식, 표사계, 퇴적물 이동, 비지도 학습, 해안 관리

학번: 2023-28970