환경영향평가 Vol. 35, No. 2(2026), pp.60-78 J. Environ. Impact Assess. 35(2): 60-78(2026) ISSN 1225-7184
https://doi.org/10.14249/eia.2026.35.2.60

Research Paper

Quantifying Changes in Particle Pathway Types Under Large-scale Coastal Development: A Lagrangian Trajectory Clustering Approach

HyeRyeon Gwon* · MinSun Kwon · Hoon Kang*** · JongGu Kim******

Kunsan National University Department of Environmental Engineering*
Ocean Physics Dept., Land & Ocean Environmental Eng.**
Marine Energy Division, MIT co., Ltd.***
Kunsan National University Department of Environmental Engineering Professor******

대규모 연안 개발에 따른 입자 이동 경로 유형변화의 정량화: 라그랑지안 궤적 군집화 접근법

권혜련* · 권민선 · 강 훈*** · 김종구******

군산대학교 환경공학과*, (주)국토해양환경기술단**, (주)해양정보기술***, 군산대학교 환경공학과 교수******

요약: 본 연구는 대규모 연안 개발이 물질 수송 패턴에 미치는 영향을 분석하기 위해 라그랑주 입자 역학(Lagrangian particle dynamics)에 머신러닝 기법을 적용하였다. 이러한 접근 방식을 통해 환경 영향 평가의 개선과 지속 가능한 해양 공간 계획 수립에 기여하고자 한다. 연구 대상지인 새만금은 간척과 항만 건설이 동시에 진행되는 반폐쇄형 하구로, 본 연구에서는 이를 개발 전, 개발 중, 개발 후 세 가지 시나리오로 수치 모형 실험을 수행하였다.

이때, 입자 운동의 구조적 복잡성을 정량화하기 위해 총 이동거리, 변위 비율, 지속 시간, 반경 평균, 곡률 등 5가지 궤적 기반 특징을 정의하였으며, 이를 기반으로 주성분 분석(PCA)과 K-means Clustering을 통해 입자를 이동 경로 유형별로 분류하였다. 분석 결과, 개발 전 시나리오에서는 다양한 이동 패턴이 혼재되어 나타났으나, 개발이 진행됨에 따라 재순환형(Recirculating-type) 입자의 비율이 점진적으로 증가하여 완공 후 시나리오에서는 68.1%까지 증가하였다. 반면, 확산형(Dispersive-type) 입자의 비중은 현저히 감소하였다. 이러한 유형별 빈도의 변화에도 불구하고, 각 유형의 평균 특징값(예: 이송형 입자의 이동 거리 98,000m 미만, 재순환형 입자의 체류 시간 약 28일, 확산형 입자의 반경 평균 6,500m 초과 등)은 시나리오 전반에 걸쳐 일관되게 유지되었다. 이는 본 연구에서 제안한 분류 체계의 견고성과 일반화 가능성을 입증한다.

따라서, 본 연구는 유속장이나 농도 분포와 같은 정적인 지표 대신 궤적 기반의 동적 패턴을 제시함으로써, 연안 개발에 따른 물질 수송 프로세스의 구조적 전이를 진단하는 새로운 프레임워크를 제공하였다. 또한, 머신러닝

First Author: HyeRyeon Gwon, Tel: +82-63-469-1871, E-mail: gwonhr@kunsan.ac.kr, ORCID: 0009-0005-3427-7272
Corresponding Author: JongGu Kim, Tel: +82-63-469-1874, E-mail: kjg466@kunsan.ac.kr, ORCID: 0000-0003-0451-5565
Co-Authors: MinSun Kwon, Tel: +82-31-695-3474, E-mail: mskwon@landocean.co.kr, ORCID: 0000-0003-0932-0070
Hoon Kang, Tel: +82-63-469-1871, E-mail: kh1023@daum.net, ORCID: 0000-0003-0867-5865

Received: 3 February, 2026. Revised: 13 April, 2026. Accepted: 17 April, 2026.

기반의 분류와 공간 분포 분석을 통합한 이 방법론은 향후 연안 환경 영향 평가 및 수질 관리 전략 수립에 있어 기초자료로 활용될 수 있다.

주요어 : 라그랑주 입자추적, 비지도 군집화, 연안 개발, 유동 경로 분류, 수동역학적 정체

Abstract : This study applied machine learning to Lagrangian particle dynamics to provide a detailed understanding of the effects of large-scale coastal development on material transport patterns. The approach contributes to improved environmental assessment and sustainable marine planning. The study area, Saemangeum, a semi-enclosed estuarine system undergoing simultaneous land reclamation and port construction, was modeled under three development scenarios: pre-, mid-, and post-development. Five trajectory-based features (total travel distance, displacement ratio, duration, radial average, and curvature) were defined to quantify the structural complexity of particle motions. These features were analyzed using principal component analysis and k-means clustering to classify the flow pathway types. The results showed that, although a mixture of movement patterns appeared in the pre-development scenario, the proportion of recirculating-type particles increased over time, reaching 68.1% in the post-development scenario. In contrast, the proportion of dispersive type particles markedly decreased. Despite these shifts in type frequency, the mean feature values for each type (e.g., travel distance < 98,000 m for advective particles, residence time ~28 days for recirculating particles, and radial average > 6,500 m for dispersive particles) remained consistent across scenarios, indicating the robustness and generalizability of the proposed classification scheme. By focusing on trajectory-based dynamic patterns rather than static metrics, such as velocity fields or concentration distributions, this study offers a novel framework for diagnosing structural changes in transport processes due to coastal development. Integrated machine learning-based classification and spatial distribution analysis provide practical insights for future coastal environmental impact assessments and water quality management strategies.

Keywords : Lagrangian Particle tracking, Unsupervised clustering, Coastal development, Flow pathway classification, Hydrodynamic stagnation

I. Introduction

The installation of artificial structures in coastal and estuarine areas induces physical alterations in natural hydrodynamic conditions such as tides, waves, and freshwater inflows. These changes affect both flow structure and material transport pathways. Alterations in flow regimes are directly linked to variations in water quality, ecosystem dynamics, sedimentation patterns, and pollutant stagnation. Therefore, a quantitative understanding of material transport mechanisms before and after the construction of such structures is essential for predicting and minimizing the environmental impacts of coastal development(Deng et al., 2025, Hong, 2020).

The Lagrangian particle-tracking approach offers a powerful tool for analyzing transport characteristics, such as stagnation, advection, and dispersion, by tracing individual particle trajectories within a continuum-based flow field. However, previous studies have typically reduced particle-tracking results to single summary metrics, such as mean residence time, total travel distance, or effective diffusivity, or have focused solely on whether particles reach predefined target locations. A recent study by Gwon et al. (2025) analyzed the effects on material transport characteristics by evaluating effective diffusivity based on concentration and particle experiments; however, it did not account for the multiple distinct and diverse particle trajectory patterns that manifest in complex flow fields.

These limitations become more pronounced in environments in which artificial structures or channel modifications lead to abrupt changes in the flow configuration, posing challenges in accurately documenting and comparing typological variability in particle transport pathways.

To overcome the inherent limitations of traditional statistical averaging, recent research trends in oceanography and coastal engineering have increasingly turned to unsupervised machine learning algorithms for analyzing high-dimensional Lagrangian data. Pioneering studies have successfully employed clustering techniques to autonomously classify dynamic movement patterns and identify dominant transport pathways within open oceans and natural estuarine systems (Jutras et al., 2024; Sonnewald et al., 2021). Unlike traditional approaches that rely on bulk statistical metrics, these data-driven clustering methods are particularly effective for isolating distinct movement types in highly nonlinear coastal flow environments. By grouping trajectories based on their structural and topological features, researchers can capture the underlying physical dynamics that are often smoothed out or lost through conventional macroscopic averaging.

Despite these significant methodological advancements, a critical academic gap remains in the current literature. While advanced machine learning techniques have proven highly effective in identifying particle trajectory types in natural flow regimes, their application to quantitatively diagnosing structural changes induced by massive anthropogenic coastal developments-such as large-scale land reclamation, seawall construction, or port expansion-remains largely unexplored. This limitation is particularly relevant and urgently needs to be addressed in the context of environmental impact assessments, where understanding the exact nature of flow obstruction and the resulting typological shifts in material transport is essential for sustainable marine spatial planning.

In this context, the aim of this study was to quantitatively evaluate the structural changes in particle transport patterns resulting from large-scale coastal development. To this end, principal component analysis and k-means clustering were applied to Lagrangian particle trajectories derived from a numerical hydrodynamic model. The resulting classification was used to compare the frequency and spatial distribution of each trajectory type across different stages of development. The study site was a semi-enclosed estuarine system that underwent simultaneous large-scale land reclamation and port construction. This study presents a novel diagnostic framework specifically designed to quantify structural transitions in transport patterns caused by large-scale coastal development by applying a machine learning-based clustering approach to Lagrangian particle trajectories. In contrast to previous studies, which focused on classifying particle motions under natural flow conditions, this study emphasizes the fundamental impact of anthropogenic modifications on the typology and spatial distribution of transport pathways.

II. Materials and Methods

1. Overview of Analytical Procedure

The methodology of this study consisted of the following steps:

(1) A hydrodynamic model, the Environmental Fluid Dynamics Code (EFDC), was constructed and validated using observed tidal levels and depth-averaged velocity data.

(2) A Lagrangian particle tracking simulation was conducted under three scenarios representing different stages of coastal development (Case 0: early stage development, Case 1: current development state, and Case 2: post-development).

(3) For each particle trajectory, five physical descriptors (total distance, displacement ratio, residence time, average radius, and average curvature) were calculated to characterize the transport behaviors.

(4) k-means clustering was applied to the standardized descriptors from Case 0 to define three representative transport types.

(5) Based on the derived cluster centroids, the particles in Cases 1 and 2 were classified into the nearest type. The proportion and spatial distribution of each transport type were compared across development scenarios to diagnose structural transitions in material transport pathways.

2. Numerical Model Setup

The area selected for this study was Saemangeum, a large artificial coastal lake located on the west coast of Korea. To simulate flow dynamics and particle transport between the inner lake and adjacent outer sea, a hydrodynamic model covering approximately 797 km² was constructed. A two-dimensional expanding grid system was applied in the horizontal domain, consisting of 141 grid lines in both the x- and y-directions, resulting in 9,244 active computational cells (Fig. 1).

The model incorporated the GATECONTROL subroutine to simulate the realistic operation of the tidal sluice gates (Park et al., 2023). Gate openings and closings were controlled twice daily during spring tide events to maintain an inner lake water level of approximately -1.5 m. The grid resolution near the sluice structures was refined to reflect the actual widths of the Sinsi (240 m) and Garyeok (300 m) sluices. Vertically, the water column was divided into five layers; bathymetric data from nautical charts were used as depth input. All bathymetric bottom elevations and water level boundary conditions were strictly referenced to the unified vertical datum of Elevation Level (E.L.). This standardization ensures physical consistency in calculating the total water column depth and accurately simulating subsequent hydrodynamic flows. Open boundary conditions were imposed using five dominant tidal constituents-M_2, S_2, N_2, K_1, and O_1-representing the semidiurnal and diurnal components. Harmonic constants for the amplitude (cm), phase (°), and frequency (cycles per hour) were applied at Gyeokpo Port (southern boundary) and Bieung Port (northern boundary), as summarized in Table 1 and shown in Fig. 1(b). Tidal forcing along the western open boundary was linearly interpolated between the two stations. The model was run for 30 days starting from January 27, 2023, with a time step of 2 s. To ensure the reproducibility and physical accuracy of the simulation, specific meteorological and freshwater boundary conditions were strictly implemented. For meteorological forcing, observed wind data from the Simpo weather station, perfectly synchronized with the simulation period, were applied. Additionally, constant dry-season freshwater inflows were prescribed for the upper boundaries, with 9.78 m³/s for the Mangyeong River and 4.38 m³/s for the Dongjin River, based on Park et al. (2023).

Three experimental scenarios were defined. Case 0 (Fig. 1a) represents 2013, prior to the construction of the East-West Road, when the inner lake remained

Table 1. Input tidal constituents at the open boundaries.

Name	Frequency	Gyeokpo-port*		Bieung-port*	
		Amp. (cm)	Phase (°)	Amp. (cm)	Phase (°)
M_2: Principal Lunar Semidiurnal	0.0805114	81.9	253.8	82.1	254.9
N_2: Elliptical Lunar**	0.0789992	13.4	237.6	13.5	238.7
S_2: Principal Solar	0.0833333	37.7	288.8	37.9	290.4
K_1: Declination Lunar-Solar	0.0417807	16.2	179.7	16.7	180.6
O_1: Principal Lunar Diurnal	0.0387307	10.9	146.4	11.0	149.0

* Four constituent harmonic constants from the Korea Hydrographic and Oceanographic Agency, 2025.
** Estimated from observations.

hydrologically connected. Case 1 (Fig. 1b) reflects the partial development status as of 2023, with the East-West Road and parts of the port constructed. Case 2 (Fig. 1c) represents future conditions in 2035, assuming full completion of reclamation and port development projects. The objective of this study was to assess whether hydrodynamic stagnation inside Saemangeum will be alleviated or exacerbated across these developmental stages. Except for changes in topography, all other boundary and initial conditions were maintained constant across all scenarios. The model was validated using observations at three locations: tidal currents at C-1, offshore water levels at L-1, and inner lake water levels at L-2 (see Fig. 1b). Additionally, a total of 1,000 particles (500 at each location) were released at two strategic locations, P-1 and P-2, for trajectory analysis. These locations were selected because they are situated in the primary central channel connecting the inner estuarine environment to the main sluice gates (Sinsi and Garyeok). As a hydrodynamic choke point, this area is highly sensitive to the structural alterations caused by the construction of the East-West Road, making it an optimal site to observe transitions

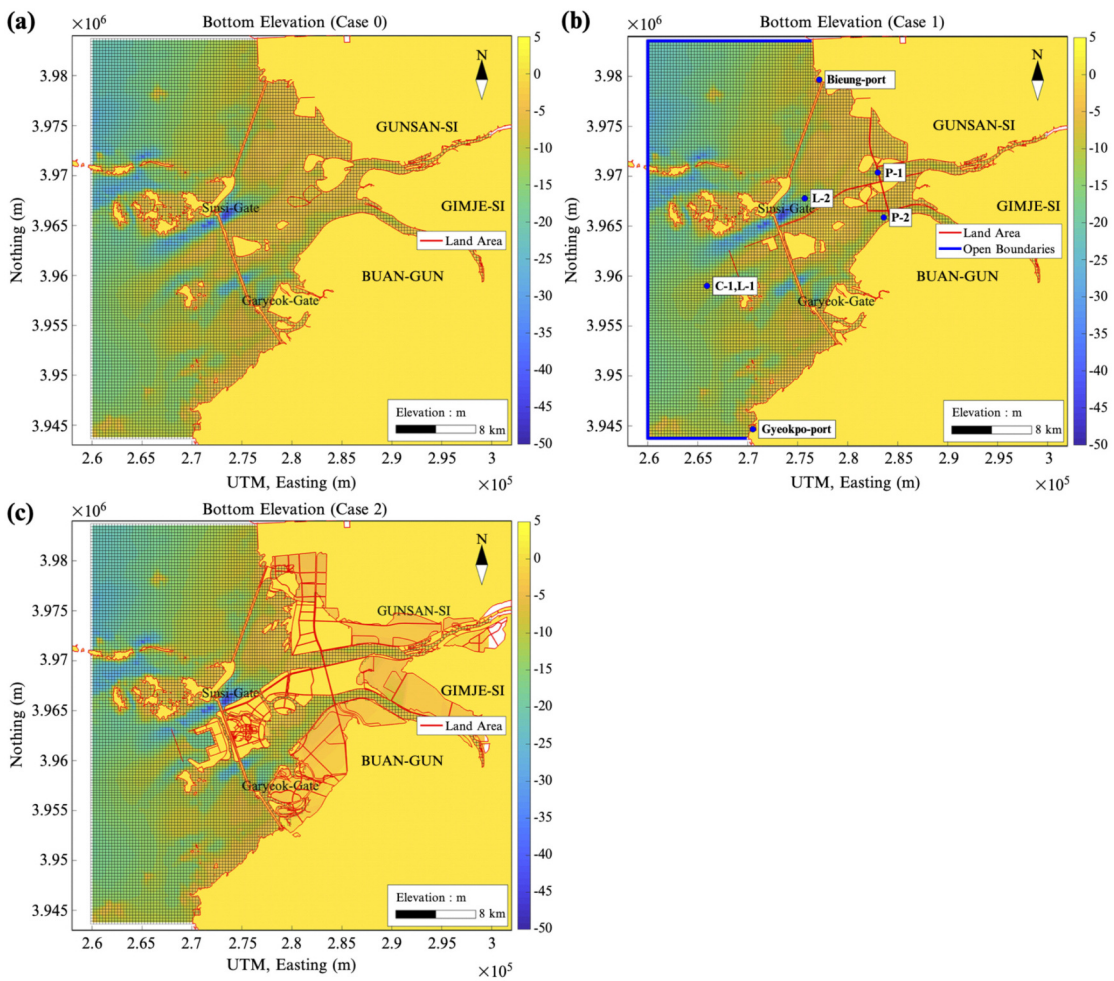


Figure 1. Model grids and bottom elevation contours for each development scenario: (a) Case 0, early-stage development (2013 conditions, prior to the internal road and port constructions); (b) Case 1, current development state (2023 conditions with partially completed roads and ports); (c) Case 2, post-development (all reclamation and port facilities completed by 2035).

between advective and recirculating transport behaviors. The total sample size of 1,000 particles (500 per location) was determined to be statistically sufficient for the unsupervised learning approach. Unlike conventional concentration-based dispersion models that require massive particle counts for mass conservation, this study focuses on the topological classification of trajectories. A sample of 1,000 trajectories provides a robust dataset for the K-means algorithm to stably converge and identify the representative cluster centroids without unnecessary computational overhead.

Although the EFDC hydrodynamic model was configured with five vertical sigma layers, the trajectory feature extraction and subsequent clustering in this study were primarily restricted to the horizontal 2D plane. This analytical scope was physically justified by the morphological and hydrodynamic characteristics of the inner Saemangeum basin. While localized deep channels (≈40m) exist immediately adjacent to the sluice gates, the vast majority of the inner lake remains exceedingly shallow and flat. In such environments, water movement is predominantly governed by intermittent sluice gate operations and limited freshwater inflows, making strong horizontal advection the primary driver of transport. Consequently, horizontal planar motions were sufficient to capture the macro-scale structural transitions of particle pathways (i.e., advection, localized looping, and multidirectional spreading) induced by the construction of massive artificial structures.

3. Classification of Particle Transport Types Using Machine Learning

To quantitatively characterize particle trajectories during the model simulation period, five feature variables were defined for each particle: total path length (d_p), displacement ratio (S_p), duration (τ_p), radial average distance (R_p), and curvature (C_p). The feature vectors were expressed as:

$$\overrightarrow{f_p} = \begin{bmatrix} d_p \\ s_p \\ \tau_p \\ R_p \\ C_p \end{bmatrix} \quad (1)$$

where d_p represents the total length of the path traveled by the particle throughout its trajectory p; in other words, the cumulative length of the actual movement trajectory, which reflects the complexity of the particle trace. Here, d_p was calculated as follows:

$$d_p = \sum_{k=1}^{T-1} | \overrightarrow{x_{k+1}} - \overrightarrow{x_k} | \quad (2)$$

where $\overrightarrow{x_k}$ is the particle position at time t_k and T indicates the number of trajectory points. Here, S_p represents the ratio of the actual path length to the straight-line distance between the start and end points of a trajectory (Gurarie et al., 2009):

$$s_p = \frac{| \overrightarrow{x_{end}} - \overrightarrow{x_{begain}} |}{\sum_{k=1}^{T-1} | \overrightarrow{x_{k+1}} - \overrightarrow{x_k} |} = \frac{d_{direct}}{d_p} \quad (3)$$

where τ_p represents the time difference between the start and end of a particle trajectory. This indicates the duration for which the trajectory tracking has been maintained, defined as follows:

$$\tau_p = t_T - t_1 \quad (4)$$

It should be noted that the Lagrangian particle tracking was conducted over a 30-day period, encompassing a full spring-neap-spring tidal cycle. For particles that failed to exit the domain through the sluice gates within this timeframe, their duration (τ_p) was inevitably right-censored (truncated) at the 30-day maximum. Consequently, the ultimate residence time for highly stagnant particles (e.g., Type B) is likely underestimated in absolute terms. However, because the primary objective of this study is not the quantitative calculation of absolute residence times, but rather the relative classification of trajectory

patterns using k-means clustering, this truncation does not compromise the analytical framework. Even with the 30-day cap, the bounded τ_p serves as a statistically robust feature in the high-dimensional space, effectively segregating rapidly advecting particles (Type A, which exit within a few days) from long-term recirculating particles (Type B) that hit the temporal limit.

where R_p indicates the density of the trajectory around the center point. The larger the dispersion shift, the smaller and more concentrated the concentration tendency, which is defined as follows(Coding et al., 2008):

$$R_p = \frac{1}{T}\sum_{k=1}^{T-1} | \overrightarrow{x_k} - \overrightarrow{x_{centroid}} | \quad (5)$$

where $\overrightarrow{x_{centroid}} = \frac{1}{T}\sum_{k=1}^{T-1} \overrightarrow{x_k}$, representing the central point of the particle trajectory. Furthermore, C_p indicates the average angle of rotation between the continuous vectors on a trajectory. It represents the degree of bending of the trajectory and is an indicator of vortex or rotational flow, defined as follows(Gautrais et al., 2009):

$$C_p = \frac{1}{T-2}\sum_{k=2}^{T-1}\theta_k, \quad (6)$$

where $\theta_k = \cos^{-1}\left(\frac{\overrightarrow{v_k}-1 \cdot \overrightarrow{v_k}}{|\overrightarrow{v_k}-1| \cdot |\overrightarrow{v_k}|}\right)$, representing the trajectory direction change angle, and is the vector of movement over time.

Prior to clustering, all feature variables were standardized to obtain a mean of zero and standard deviation of one. Classification was performed using the k-means algorithm implemented via the sklearn.cluster.KMeans library in Python.

To objectively determine the optimal number of clusters (k), quantitative evaluation metrics including the Elbow method (WCSS) and Silhouette score were applied to the standard feature space of Case 0. As shown in the Figure 2, the WCSS plot demonstrated a distinct reduction in variance up to k=3 and k=4. While the Silhouette score reached its absolute maximum at k=2 (≈0.415), it maintained a highly significant value at k=3 (≈0.390) before dropping sharply at k=4. The mathematically

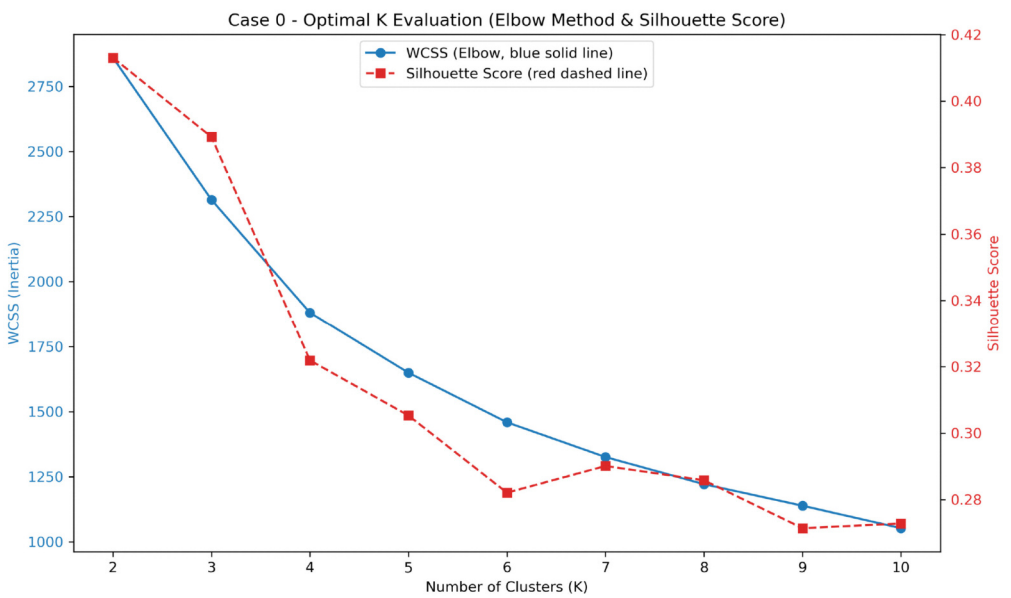


Figure 2. Quantitative evaluation for determining the optimal number of clusters (k) based on the trajectory feature space of Case 0. The blue solid line represents the Within-Cluster Sum of Squares (WCSS) for the Elbow method, demonstrating a reduction in variance up to k=3 and k=4. The red dashed line denotes the Silhouette score, which maintains a highly significant value at k=3 before dropping sharply at k=4.

optimized value of k=3 was strategically selected to balance statistical robustness with hydrodynamic validity. This value aligned perfectly with the three distinct physical transport mechanisms observed in the semi-enclosed estuarine system: advective (outward), recirculating (looping), and dispersive (spreading) behaviors, which would otherwise be artificially merged under k=2. Therefore, k=3 was robustly selected for the final k-means classification.

k-means assigns each data point to the nearest centroid based on the Euclidean distance. As the initial centroids were randomly selected, a fixed random seed was used to ensure the reproducibility of the results. The centroids were then updated iteratively by calculating the mean of all points assigned to each cluster. The process continued until the centroid positions converged.

The objective function of the k-means algorithm, which minimizes the within-cluster sum of squares, is defined as:

$$\underset{C_1, \cdots, C_k}{\min} \sum_{i=1}^{k} \sum_{\vec{x} \in C} \| \vec{x} - \vec{\mu_i} \|^2, \tag{7}$$

where $\vec{\mu_i}$ refers to the central point, i, of the cluster.

In this study, particle trajectory data from the pre-development scenario (Case 0) were used to derive clusters. The feature values of the resulting cluster centroids were used as fixed reference points. The particles from Cases 1 and 2 were then assigned to the nearest cluster based on the Euclidean distance from these centroids. In other words, rather than performing new clustering for each scenario, Cases 1 and 2 were classified using the fixed cluster structure derived from Case 0, enabling a consistent and quantitative comparison of trajectory-type frequencies across different developmental stages.

Furthermore, based on the average feature values of each cluster, the particle movement types (Types A-C) were defined using the following interpretation-based rules:

$$\text{Type A (Advective): } l = \underset{l}{\text{argmin}} \ \mu_l^{(duration)} \tag{8}$$

$$\text{Type B (Loop): } l = \underset{l \in A}{\text{argmax}} \ \mu_l^{(dist)} \text{ and} \tag{9}$$

$$\text{Type C (Spreading): } l = \underset{l \notin A \cup B}{\text{argmax}} \ \mu_l^{(radial\ avg)} \tag{10}$$

4. Model Validation Metrics

To evaluate the performance of the hydrodynamic model, we compared the simulated and observed water levels and current velocity components (u and v) using two statistical metrics: the root mean square error (RMSE) and the Pearson correlation coefficient (γ). These indicators quantified the level of agreement between the simulated (S_i) and observed (O_i) water levels at each time step, i, where N is the total number of observations:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (S_i - O_i)^2} \text{ and} \tag{11}$$

$$\gamma = \frac{\sum_{i=1}^{N} (S_i - \bar{S})\,(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N} (S_i - \bar{S})^2 \sum_{i=1}^{N} (O_i - O)^2}} \tag{12}$$

where $\bar{S}$ and $\bar{O}$ represent the mean values of the simulated and observed data, respectively. The RMSE provided an absolute measure of deviation in meters, whereas the correlation coefficient reflected the linear agreement between time-series, ranging from –1 to 1.

III. Results and Discussion

1. Model Validation of Hydrodynamic Simulation

Before classifying the particle trajectories, the results of the EFDC simulation were validated against field observations. EFDC is a widely used three-dimensional hydrodynamic and water quality modeling system developed by the U.S. EPA based on the finite-difference method; this system has the ability to simulate free-surface turbulent flows in rivers, estuaries, and coastal

regions(Hamrick, 1992). The model solves the Navier-Stokes equations under hydrostatic and Boussinesq assumptions, including modules for salinity, temperature, sediment, and constituent transport.

The selected observation data ensured robust temporal and spatial representativeness for the validation process. Temporally, the continuous 30-day period encompassed a full spring-neap-spring tidal cycle, capturing both extreme and weak hydrodynamic forcing conditions. Spatially, the validation stations were strategically selected to represent distinct hydrodynamic regimes. Stations L-1 and C-1, located at the same offshore position, validate the model's ability to capture the primary external tidal forcing (both elevation and currents) from the open sea. Conversely, station L-2, located within the inner basin, specifically verified the model's accuracy in reproducing the non-linear water level variations dictated by the intermittent operations of the sluice gates. Together, these locations comprehensively validated both the natural external boundary conditions and the artificial internal flow controls.

As shown in Fig. 3, the simulated and observed water levels exhibited strong agreement, with an RMSE of 0.29 m and a correlation coefficient of 0.98.

The depth-averaged current velocities of the simulations and observations were also compared (Fig. 4). The flow patterns in both the x-direction (u-component) and y-direction (v-component) were generally consistent. Quantitatively, the RMSE values were 0.06 m/s for the u-component and 0.12 m/s for the v-component, while the correlation coefficients were 0.88 and 0.64, respectively. Although the v component showed a relatively larger RMSE and lower correlation, this discrepancy was attributed to noise in the observed data, such as environmental disturbances or sensor limitations, which led to irregular fluctuations, as shown in Fig. 4.

Despite this, the simulation accurately captured major tidal characteristics, including periodicity, phase, and amplitude. Considering the objective of this study-to quantitatively compare particle trajectory types-the level of accuracy achieved by the hydrodynamic model was deemed sufficient for reliable trajectory-based analysis.

To verify whether the inner water level controlled by the GATECONTROL subroutine was appropriately simulated, the modeled results were compared with the observations. As a key component, the GATECONTROL subroutine is a user-defined control module implemented within the EFDC model to simulate the operation of

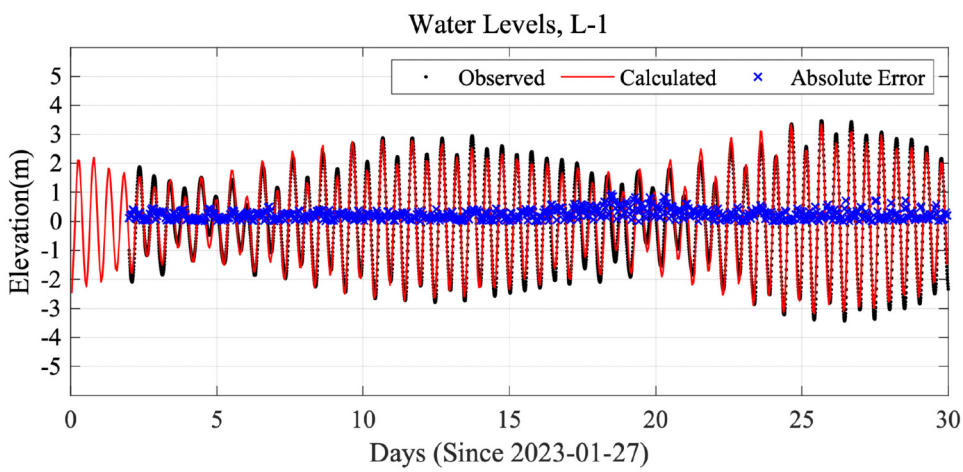
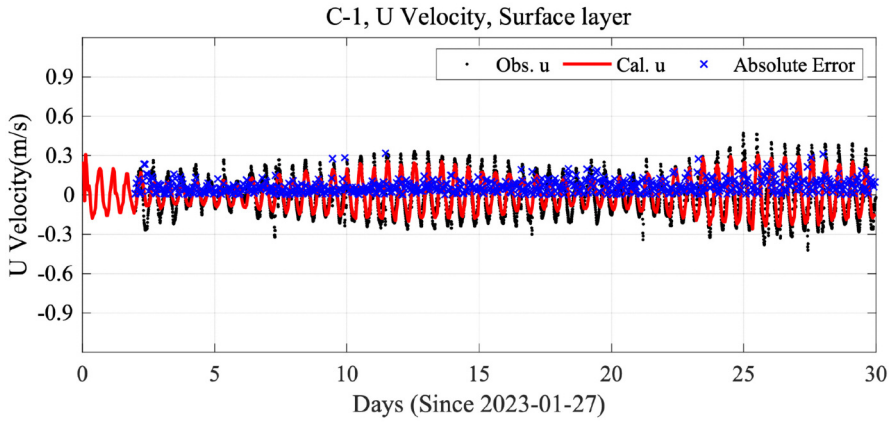
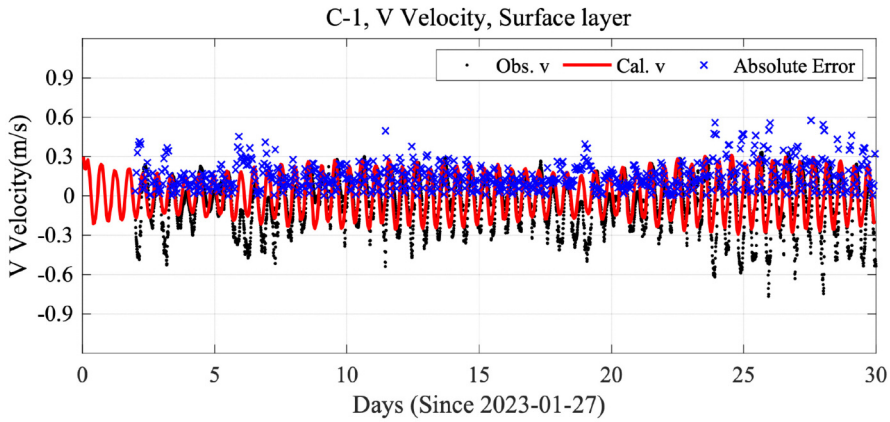


Figure 3. Comparison of observed and simulated tidal elevations at point L-1 over a 30-day period beginning on January 27, 2023. Black, red, and blue lines represent observed data, model results, and absolute error, respectively.



(a)



(b)

Figure 4. Comparison of observed and simulated surface u and v velocities at C-1. (a) u component, (b) v component. Black, red, and blue lines indicate observed, simulated, and absolute error values, respectively.

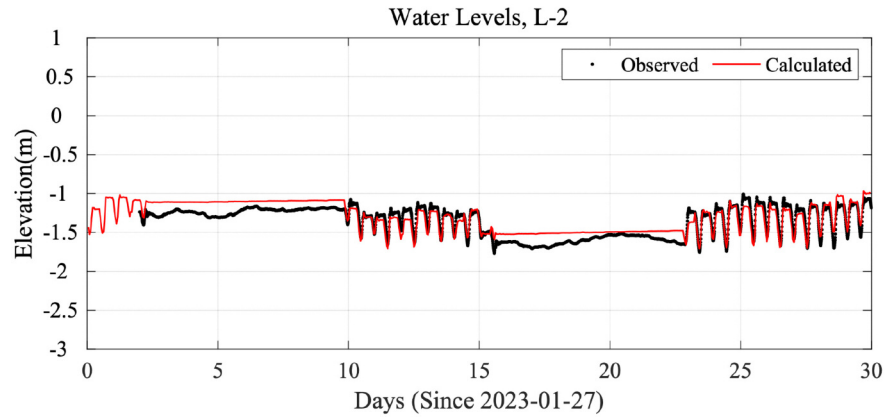


Figure 5. Comparison of observed and simulated water levels at point L-2 over a 30-day period beginning on January 27, 2023. The black line represents observed data and the red line represents model-simulated water levels.

water control structures, such as sluice gates or weirs. This routine dynamically modifies the boundary flow or water-level conditions based on predefined gate operation schedules or real-time water-level thresholds. In this study, a subroutine was used to replicate the actual operational behavior of the interior gates, which controlled the exchange between the inner basin and outer sea. By incorporating this routine, the model could accurately depict nonlinear flow behavior, including intermittent or pulsed exchange events, which could not be captured by static boundary conditions. The comparison of modeled and observed water levels (Fig. 5) demonstrated that the gate dynamics were accurately represented, with an RMSE of 0.11 m and a correlation coefficient of 0.87, indicating the effective functionality of the GATECONTROL routine within the simulation framework.

2. Simulated Hydrodynamic Flow Patterns

Figure 6 presents horizontal velocity vectors during flood and ebb tide conditions for spring and neap tides, respectively, reflecting the variation in flow structures arising from different tidal conditions and sluice gate operations. The study area connects to the open sea via two main sluices: the Sinsi sluice in the north and Garyeok sluice in the south. These gates are typically closed during neap tides and opened during spring tides, resulting in substantial water exchange between the inner and outer regions, primarily during spring tide

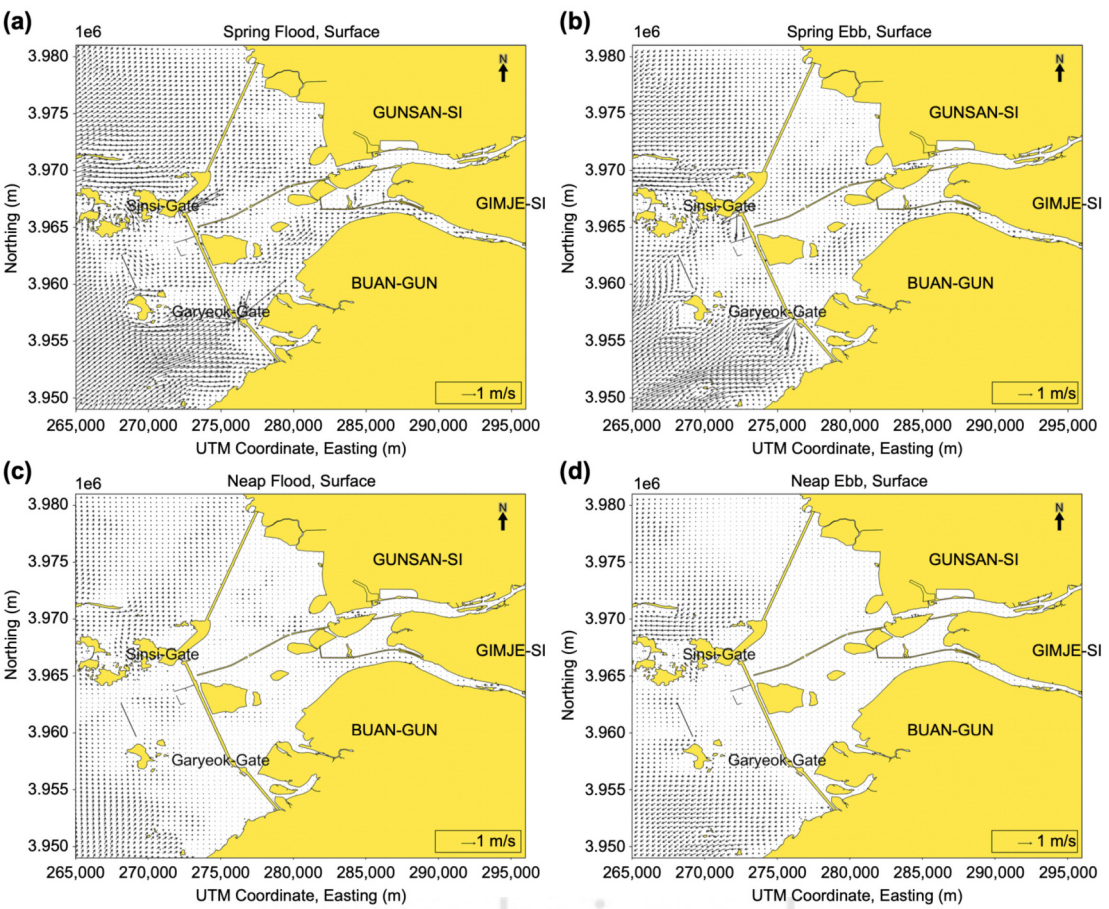


Figure 6. Simulated surface flow patterns under different tidal and temporal conditions. (a) Spring flood currents, (b) spring ebb currents, (c) neap flood currents, and (d) neap ebb currents.

conditions. In contrast, flow during neap tides is weak and irregular due to minimal water exchange.

During spring tides, well-defined flow patterns emerge in response to the timing of gate openings, peak floods, and ebb currents. These flows act as dominant drivers influencing particle transport routes and the formation of trajectory types. In contrast, under neap-tide conditions, the inner flow remains weak, resulting in limited particle displacement and a constrained distribution of trajectory types.

Differences in flow intensity and directionality between spring and neap tides significantly affect particle-level characteristics, such as straightness, curvature, and residence time. These tidally driven flow variations provide critical background conditions for the interpretation of transport structures based on the trajectory-type classification framework proposed in this study.

3. Classification of Particle Transport Types Using Machine Learning

Figure 7 illustrates the particle trajectories (grey dots), final positions (red dots), and initial release locations (blue stars) for three scenarios: Case 0 (pre-development condition in 2013), Case 1 (current partially developed condition in 2023), and Case 2 (fully developed condition projected for 2035). As development progressed from Cases 0 to 2, the overall extent and spatial pattern of particle movement changed in response to the configuration of the

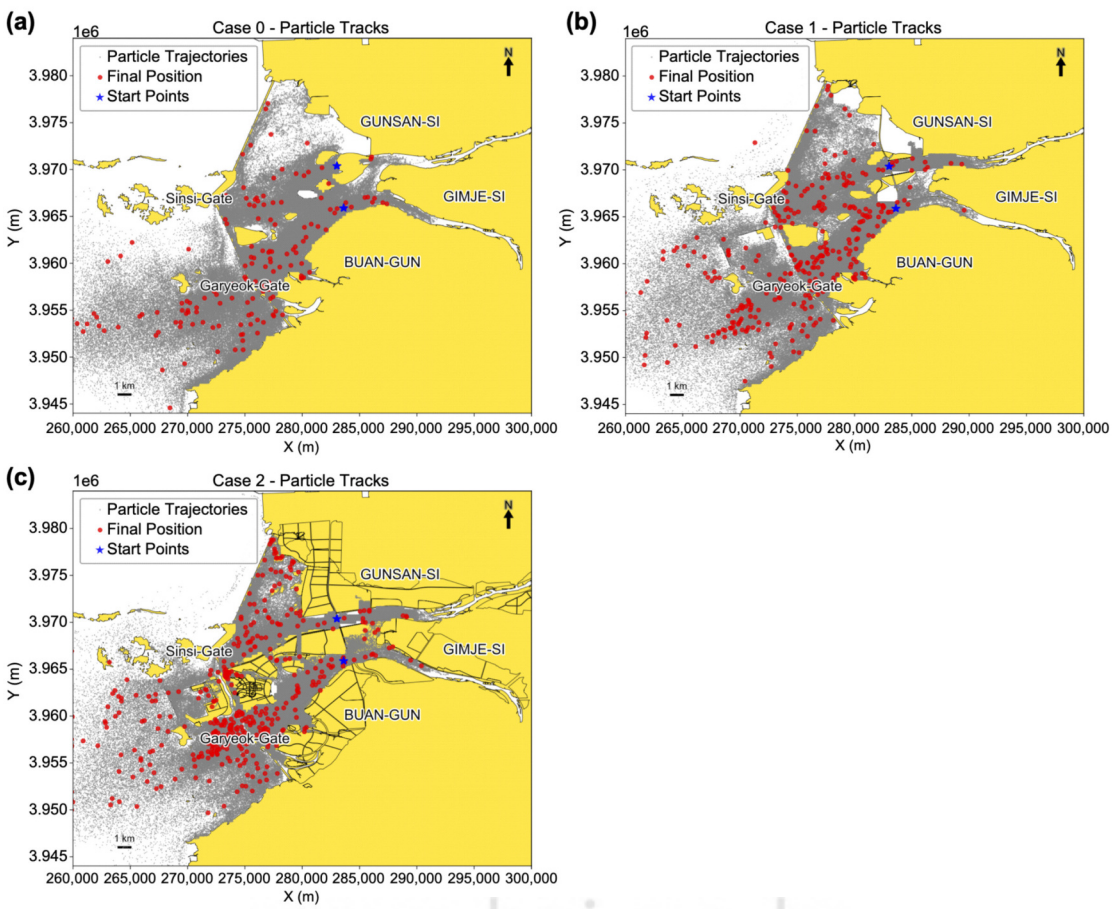


Figure 7. Simulated particle trajectories and final positions for each development scenario. (a) Case 0, early-stage development (2013); (b) Case 1, current development state (2023); (c) Case 2, post-development (2035).

artificial structures and alterations in the flow pathways. Particularly, a noticeable increase was observed in the number of particles retained within the inner region of Saemangeum (landward of the seawall).

However, these spatial distribution patterns alone are insufficient to fully explain the underlying transport behaviors and trajectory structures. Therefore, this study classified individual particle trajectories based on their characteristic features and quantitatively compared the spatial distribution and frequency of each transport type across the developmental stages. This approach enabled a more systematic interpretation of particle movement tendencies, such as advection-dominant, recirculating, or dispersive behaviors, beyond simple residence or escape outcomes. Ultimately, this allowed for a detailed evaluation of how material transport structures were reorganized in response to large-scale coastal development.

The particle trajectories were quantitatively characterized based on their physical transport properties and classified into distinct movement types using a machine learning approach. Five feature variables were used in the analysis: total travel distance (dist), displacement ratio, residence time (duration), radial average (rad_avg), and mean curvature. Using k-means clustering applied to the Case 0 data, three primary trajectory types-A, B, and C-were identified.

Figure 8 shows the clustering results for Case 0; the representative physical characteristics of each cluster

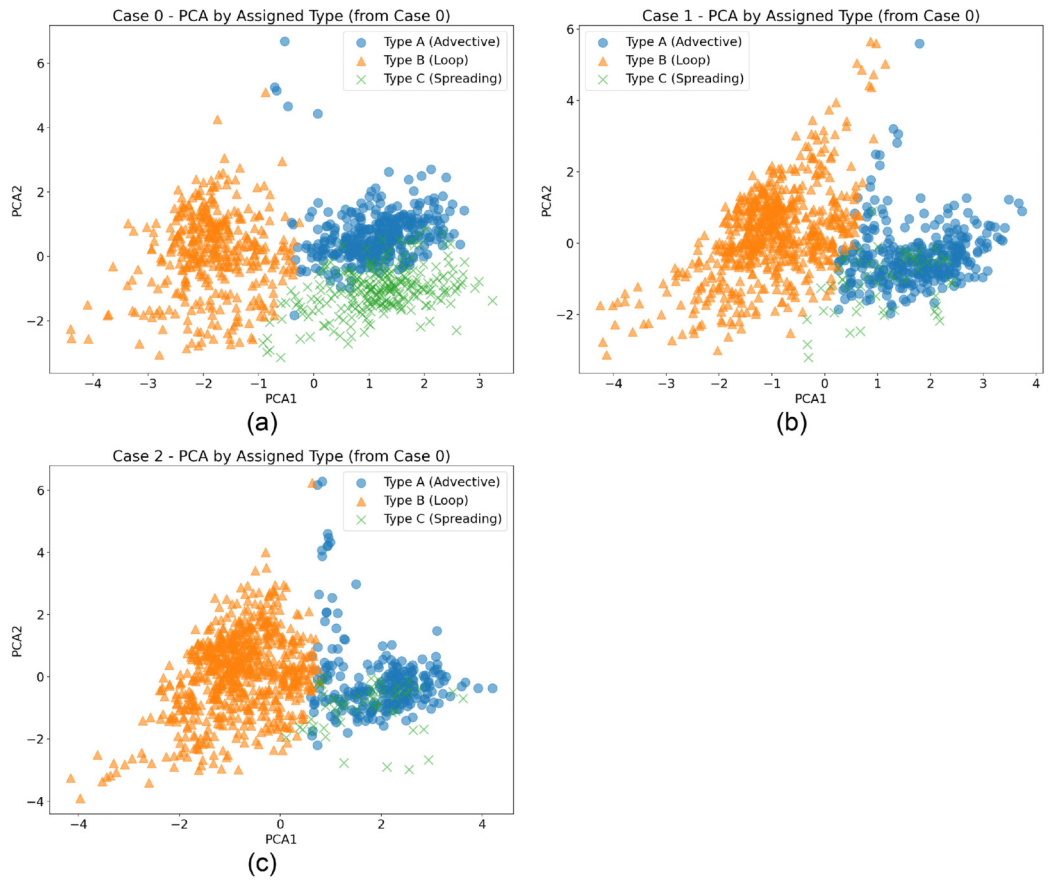


Figure 8. Principal component analysis scatter plots of particle trajectories classified by k-means clustering based on the Case 0 feature space. (a) Case 0, (b) Case 1, (c) Case 2. Each point represents a particle colored by its assigned transport type: Type A (advective), Type B (loop), and Type C (spreading).

centroid are summarized in Figure 9(a). Type A (advective) was characterized by rapid, unidirectional movement along dominant flow paths, typically toward the outer sea. This type exhibited the shortest distance, duration, and average radial values. Type B (loop-type) represented particles that recirculated within closed or semi-enclosed flow patterns and exhibited rotational behavior, prolonged residence,and local stagnation. It has the longest distance and duration, and the lowest displacement ratio. Type C (spreading type) described broadly dispersed trajectories that spread in multiple directions regardless of flow alignment. This type exhibited the largest radial average, reflecting widespread dispersion, and the lowest curvature, indicating minimal directional turning.

As described previously, the trajectory types (Types A-C) were initially defined by applying k-means clustering to the particle trajectory data from Case 0. The resulting classification scheme was then applied to Cases 1 and 2. As shown in Figure 7(b) and (c), the particle trajectories from both cases were successfully categorized into the same three types.

Moreover, as illustrated in Figure 9(a)-(c), the mean values of the key features for each type remained consistent across all scenarios-for example, the travel distance for Type A ranged from approximately 93,834 to 97,996 m, the average residence time for Type B was between 27.72 and 28.26 days, and the radial average for Type C ranged from 6,502 to 6,848 m. These results indicated that the relative differences among the three types were maintained across different developmental stages.

This consistency suggests that the proposed clustering framework was robust to variations in hydrodynamic conditions (e.g., changes in flow pathways across cases), providing a stable and generalizable basis for particle trajectory classification. This reinforced the reliability and applicability of the particle-based transport analysis performed in this study.

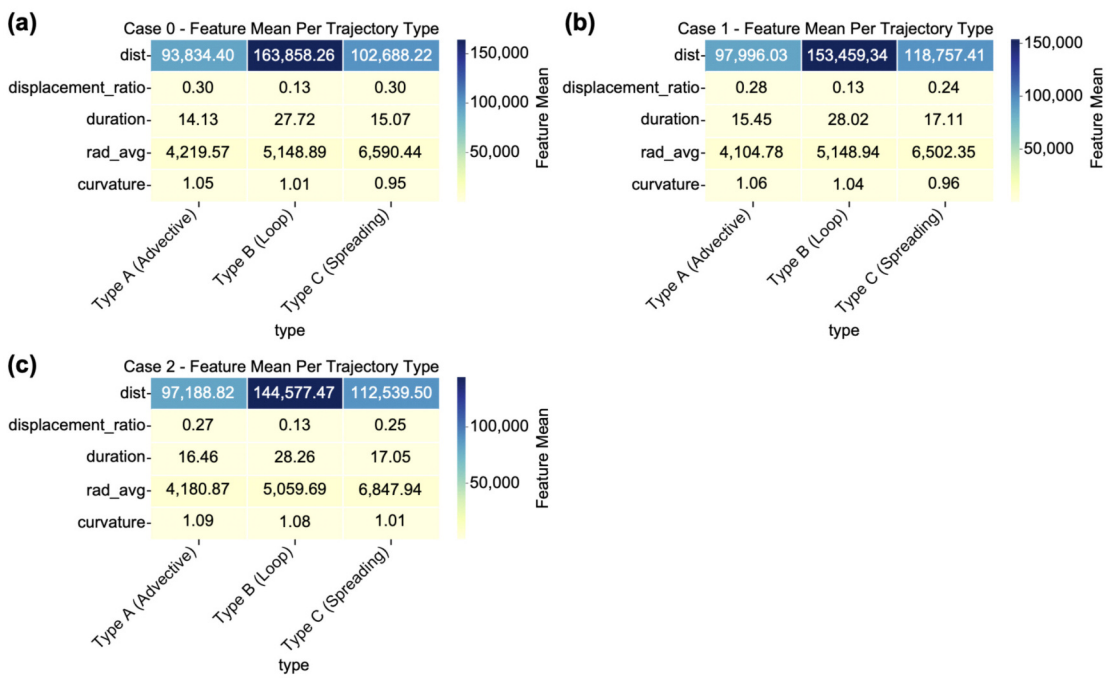


Figure 9. Mean values of trajectory features for each particle type across development scenarios. (a) Case 0, (b) Case 1, (c) Case 2. Each panel shows five representative features (distance, displacement ratio, duration, radial average, and curvature) averaged by transport type: Type A (advective), Type B (loop), and Type C (spreading).

4. Flow Pathway Analysis by Trajectory Type

The representative trajectories for each classified type, defined as the five particles closest to each cluster centroid, are shown in Figure 10. In Case 0, Type A (advective) particles exhibited straight outward-directed paths from both northern and southern regions, primarily following a strong seaward flow. Type B (loop) particles displayed recirculating behavior centered around the inner central channel, reflecting natural stagnation patterns in the absence of flow-blocking structures, such as cross-channel embankments. Type C (spreading) particles showed widely dispersed trajectories in both east-west and north-south directions, characterized by initial diffusion and complex spreading patterns.

In Case 1, Type A particles remained dominant in the southern channel, flowing outward to the sea; however, their presence decreased in the north due to reduced openness. Type B exhibited stronger recirculating patterns as structural installations reinforced inner-looped pathways, with a particularly well-defined recirculation center in the northern region. Type C exhibited intermediate characteristics between advection and looping, with a more spatially distributed pattern.

In Case 2, the Type A trajectories became even more pronounced with a clear, direct outflow from the inner area to the open sea, representing the most distinct

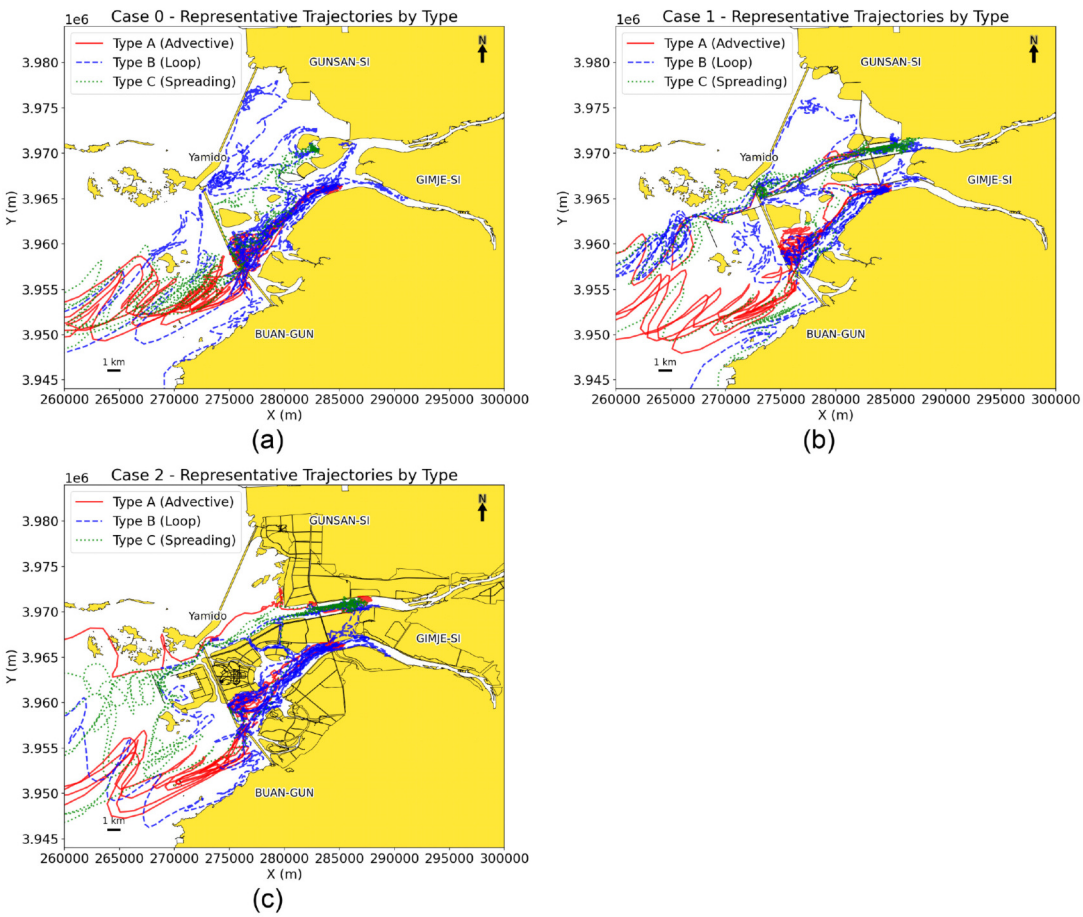


Figure 10. Representative particle trajectories for each transport type across development scenarios. (a) Case 0, (b) Case 1, (c) Case 2. Blue, red, and green lines represent Type A (advective), Type B (loop), and Type C (spreading) trajectories, respectively.

advective behavior. Type B revealed recirculation through narrow interior channels formed by reclamation and development activities. Type C showed the broadest dispersion radius, which was particularly influenced by the newly constructed outer port, forming complex and spatially expansive spreading patterns.

5. Changes in Flow Type Distribution Across Development Scenarios

Figure 11 summarizes the variation in the number of particles assigned to each trajectory type across development scenarios. In Case 0, the particle types were relatively evenly distributed: Type A (advective) accounted for approximately 37.8%, Type B (looping/recirculating) for 35.2%, and Type C (spreading) for 27.0%. This indicated a balanced mixture of transport behaviors, including outward advection, internal circulation, and multidirectional dispersion.

However, in Case 1, the proportion of Type B particles increased significantly to 59.0%, whereas that of Type C sharply declined to only 6.9%. As visually supported by the representative spatial trajectories in Fig. 10, this dramatic reduction in dispersive behavior was directly linked to the physical bifurcation of the inner basin. In Case 0, Type C particles (Figure 10a, green lines) exhibited unhindered, multidirectional spreading across the entire central basin, moving freely between the northern and southern areas. In contrast, the construction of the East-West Road in Case 1 acts as a massive transverse barrier that severed this north-south connectivity. Consequently,

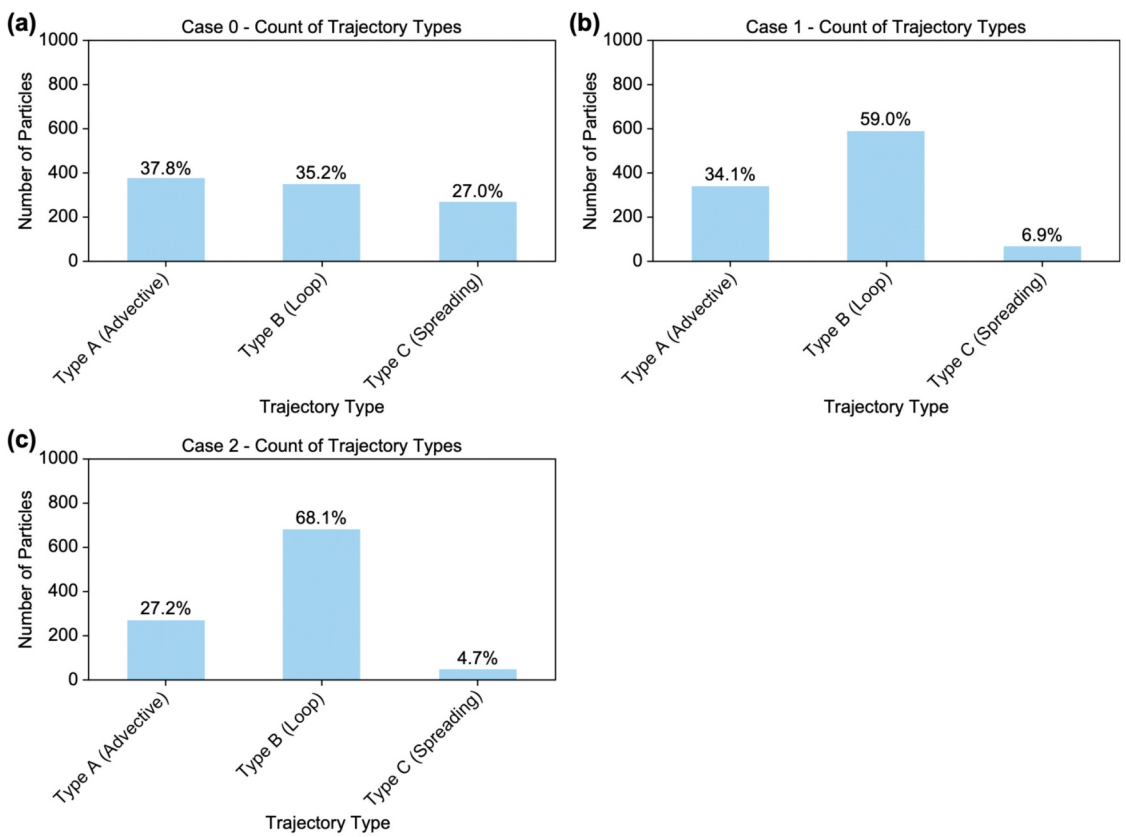


Figure 11. Distribution of particle trajectory types for each development scenario. (a) Case 0, (b) Case 1, (c) Case 2. The bars represent the number and percentage of particles assigned to each transport type: Type A (advective), Type B (loop), and Type C (spreading).

particles that would naturally disperse over a wide radius were obstructed by these internal embankments. Instead of maintaining complex and spatially expansive spreading patterns (Type C), these particles were dynamically forced into constrained, localized recirculating pathways (Type B). This physical constraint is clearly demonstrated in Fig. 10b, where the dense, localized blue loops are heavily concentrated adjacent to the road infrastructure, particularly in the northern region. Ultimately, these structural modifications induced severe hydrodynamic redirection, effectively converting dispersive particles into stagnating ones and structurally enhancing internal looping patterns.

In the fully developed Case 2, the dominance of Type B further intensified, comprising 68.1% of all particles; Type A decreased to 27.2% and Type C decreased to 4.7%. These results indicated that the large-scale development of Saemangeum has restricted seaward advection and promoted inner circulation and long-term particle retention within the basin.

Although these findings provide conclusive evidence of structural transitions in particle transport behavior, several limitations should be acknowledged to guide future research and interpretation. First, although the hydrodynamic model and particle tracking were conducted in three dimensions, the classification and visualization of the transport types were performed only in the horizontal plane. Consequently, vertical movement characteristics were not reflected in the analysis, potentially influencing the interpretation of transport behaviors in stratified or vertically dynamic environments. Second, a formal sensitivity analysis was not conducted. As the classification criteria were initially developed using Case 0 and then directly applied to Cases 1 and 2 without modification, the effectiveness of the classification scheme under substantially different flow regimes or variation in the trajectory feature values remains to be evaluated.

IV. Conclusion

This study offers a novel contribution by applying machine learning techniques to classify particle trajectories derived from numerical simulations and quantitatively comparing the frequency of each trajectory type across different development scenarios. Through this approach, this study provides a practical interpretation of the reorganization of material transport structures and stagnation characteristics under large-scale coastal development. Unlike previous studies that largely relied on static indicators, such as velocity fields or concentration distributions, this study focused on dynamic patterns derived from individual particle trajectories.

Specifically, the quantitative classification revealed a dramatic structural shift in transport pathways. In the pre-development scenario, particle typologies were relatively balanced; however, following the full-scale infrastructural development, the proportion of localized recirculating trajectories (Type B) severely surged to 68.1%, while widely dispersive pathways (Type C) drastically plummeted to 4.7%. These empirical numerical results conclusively demonstrated that the massive artificial barriers fundamentally restrict outward advection and structurally exacerbated hydrodynamic stagnation within the basin.

Particularly, the use of clustering-based classification and analysis of the spatial distribution across scenarios represents a rare methodological application, even at the global scale. By enabling the quantitative diagnosis of structural changes in transport behavior, particularly in complex and heavily modified coastal regions such as Saemangeum, this study moves beyond qualitative inference to offer actionable insights for coastal management. Practically, the proposed framework can serve as a robust decision-making support tool. For instance, by explicitly mapping the spatial hotspots of Type B (recirculating) particles, policymakers can optimize sluice gate operational schedules to strategically flush out localized water quality

stagnation zones. Furthermore, in the planning stages of future infrastructural projects, this methodology can guide the design of eco-friendly bridges or culverts that minimize the physical blockage of Type C (spreading) pathways. Ultimately, these specific applications demonstrate how this trajectory-based diagnostic approach can substantially elevate the precision and effectiveness of future coastal environmental impact assessments and sustainable marine spatial planning.

Although this study developed classification criteria based on the Saemangeum region, the overall methodology-extracting trajectory-based features, classifying transport types, and comparing their distributions across scenarios-can be adapted to other coastal systems. However, the implementation of this approach to different regions requires recalibration of the classification thresholds and validation against the specific hydrodynamic characteristics of the new environment.

Finally, a notable limitation of this study is the restriction of trajectory feature extraction and clustering to the horizontal (2D) plane. While this two-dimensional approach is largely justified by the shallow bathymetry and advection-dominated flows driven by sluice gate operations, the hydrodynamic simulation and subsequent Lagrangian particle tracking were fundamentally conducted in a fully three-dimensional space using the EFDC model. Because semi-enclosed estuarine systems like Saemangeum can also exhibit complex vertical dynamics, such as density-driven stratification and vertical mixing, neglecting vertical particle displacement may oversimplify the material transport processes. Therefore, extending this machine learning-based diagnostic framework to incorporate three-dimensional trajectory features remains an essential future research direction to provide a more comprehensive understanding of spatial transport mechanisms in dynamically stratified environments.

References

Codling E. A., Plank M. J., Benhamou S. (2008). Random walk models in biology. *J. R. Soc. Interface., 5*, 813-834. https://doi.org/10.1098/rsif.2008.0014

Deng X., Wang Z., Ma X. (2025). Impact of silted coastal port engineering construction on marine dynamic environment: A case study of Binhai Port. *J. Mar. Sci. Eng., 13*, 494. https://doi.org/10.3390/jmse13030494

Gautrais J., Jost C., Soria M., Campo A., Motsch S., Fournier R., Blanco S., Theraulaz G. (2009). Analyzing fish movement as a persistent turning walker. *J. Math. Biol., 58*, 429-445. https://doi.org/10.1007/s00285-008-0198-7

Gurarie E., Andrews R. D., Laidre K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecol. Lett., 12*, 395-408. https://doi.org/10.1111/j.1461-0248.2009.01293.x

권혜련, 권민선, 김종구, 강훈. (2025). 연안 개발에 따른 물질 확산 저해 메커니즘 분석: 농도·입자 기반 유효확산계수 비교, 환경영향평가학회지, 34(5), 314-328.

Gwon, H., Kim J., Kang, H., & Kwon, M. (2025). Assessing Mechanisms of Reduced Dispersion in Response to Coastal Development: Comparison of Concentration-and Particle-Derived Diffusion Indices. *Journal of Environmental Impact Assessment, 34*(5), 314-328

Hamrick, J. M. (1992). A three-dimensional environmental fluid dynamics computer code: *Theoretical and computational aspects. Special Report No. 317 in Applied Marine Science and Ocean Engineering*. Virginia Institute of Marine Science, The College of William and Mary, Gloucester Point, VA. https://doi.org/10.21220/V5TT6C

홍남식. (2020). 감천항 남방파제 및 도류제 건설이 유동장 변화 및 항내 해수순환에 미치는 영향 분석, 한국해양공학회지, 34(2), 136-146.

Hong, N. (2020). Analysis of effect on seawater flow change and circulation inside port due to the construction of south breakwater and weir at Gamcheon Port. *Journal of Ocean Engineering and Technology, 34*(2), 136-146. https://doi.org/10.26748/KSOE.2020.005

Jutras, M. P., Lanat, N., Dufour, C. O., & Talbot, L. C. (2024). Machine learning-based clustering of oceanic Lagrangian particles: Identification of the main pathways of the Labrador Current. *Journal of Advances in Modeling Earth Systems, 16*(7), https://doi.org/10.1029/2023MS003902

박성화, 김종구, 권민선. (2023). 새만금 배수갑문 운영에 따른 염분 변화와 저층수의 입자교환 모의. 해양환경안전학회지, 29(6), 562-575.

Park, S., Kim, J., & Kwon, M. (2023). Salinity Changes and Bottom Water Particle Exchange Simulations in Response to Sluice Gate Operations at Saemangeum Lake. *Journal of the Korean Society of Marine Environment & Safety, 29*(6), 562-575.

Sonnewald, M., Lguensat, R., Jones, D. C., Dueben, P. D., Brajard, J., & Balaji, V. (2021). Bridging observations, theory and numerical simulation of the ocean using machine learning. *Environmental Research Letters, 16*(7), https://doi.org/10.1088/1748-9326/ac0eb0