Engineering Faculty – School of Mechanical Engineering

Sub-Mesoscale Currents and Contaminant Analysis in the Mediterranean Sea Using Markov Chains

FINAL REPORT

FINAL PROJECT

By: Shai Pollak

Advisor: Professor Roy Barkan

July 2025

Contents

1. Abstract

This study presents a new use of Markov chain modeling to explore how sub-mesoscale currents affect surface contaminant transport in the Mediterranean Sea. We simulate Lagrangian particle paths using the Parcels Python framework. This allows us to analyze dispersion patterns at two spatial resolutions, 300 m and 3 km, and during two seasons, summer and winter. The transition matrices created from the simulations form the basis for a probabilistic transport model, which shows the chances of particle movement between spatial bins. Spectral analysis, including eigenvector decomposition and Sparse Eigenbasis Approximation (SEBA), identifies coherent transport areas and reveals dynamics that depend on resolution and season. The results show that finer-resolution simulations lead to better mixing due to sub-mesoscale variability, while coarser resolutions capture mesoscale features like boundary currents and eddies. The Markov framework allows for efficient prediction of contaminant paths from any starting point, serving as a valuable tool for environmental risk assessment and emergency response. Furthermore, interprovincial hitting-time metrics and spectral clustering provide a quantitative basis for defining natural transport barriers in the sea. This work demonstrates the critical role of sub-mesoscale currents in shaping pollutant dispersion and lays the groundwork for scalable, data-driven ocean management tools.

2. Introduction

2.1. The Contamination Challenge in the Mediterranean

Contaminants in the ocean originate from a variety of anthropogenic sources including oil spills, plastic debris, agricultural runoff, sewage discharge, and industrial pollutants. These substances not only pose a threat to marine ecosystems and biodiversity but can also enter the food chain, affecting human health. The Mediterranean Sea is particularly vulnerable due to its semi-enclosed nature, high coastal population density, oil and gas facilities, and limited water exchange with the Atlantic Ocean.

Effectively tackling marine contamination requires more than just clean-up efforts. It needs tools that can predict where pollutants will go and gather. Forecasting how contaminants move is important for quick responses, conservation planning, and reducing damage to ecosystems and the economy. By tracking the path of particles and understanding the flow patterns that influence their movement, scientists and policymakers can spot pollution hotspots and use resources wisely.

Developing quick, precise, and effective methods for modeling these trajectories, especially those affected by sub-mesoscale dynamics, is crucial. Using probabilistic tools like Markov chains allows us to measure the likelihood of transitions between ocean regions. This creates a statistical framework to predict how contaminants spread over time.

2.2. Circulation Patterns and Sub-Mesoscale Currents in the Mediterranean

The Mediterranean Sea has a complex and active circulation system made up of large-scale gyres, boundary currents, mesoscale eddies, and detailed sub-mesoscale structures. Key surface current systems include the Atlantic Water inflow through the Strait of Gibraltar. This water moves eastward along the North African coast as the Algerian Current and returns westward along the northern shores. The Levantine and Ionian sub-basins also show significant mesoscale activity, influenced by topography and seasonal wind patterns. In recent years, there has been more focus on sub-mesoscale currents, which are circulation features that cover spatial scales of 1 to 10 km and temporal scales of hours to days.

These sub-mesoscale currents, despite their relatively small spatial scales, play an important role in horizontal stirring, vertical fluxes, and the dispersion of passive

tracers such as pollutants. Sub-mesoscale features can create localized trapping zones, sharp transport barriers, and faster mixing that coarse-resolution models miss. In the semi-enclosed and highly stratified Mediterranean Sea, these effects are especially significant. Contaminants released at the surface can either stay confined to certain sub-regions or spread quickly, depending on the sub-mesoscale activity.

Figure 1 below shows the strong effect of sub-mesoscale currents on how surface-released particles move. The background displays relative vorticity normalized by the Coriolis parameter (ζ/f). This reveals a dynamic sub-mesoscale flow field filled with swirling filaments and eddies. In the left panel, a group of particles follows coherent paths, creating elongated patterns as the flow evolves. The right panel shows the same area before the movement, with particles still in their initial, compact release shape. This difference highlights how sub-mesoscale features quickly distort and spread contaminant plumes in just a few hours. It shows why capturing these dynamics in predictive models is essential.

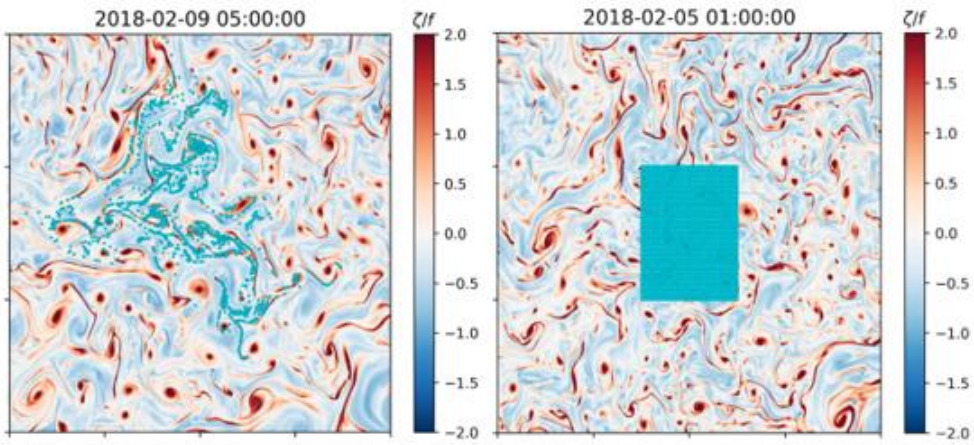


Figure 1 – Sub-Mesoscale currents and eddies carry surface released particles

3. State of the Art

Recent studies have highlighted the effectiveness of Markov chain frameworks in modeling ocean transport. Miron et al. (2019), for example, applied this approach to RAFOS float trajectories in the deep Gulf of Mexico. They constructed transition matrices representing the probability of transport between regions and used them to extract coherent Lagrangian provinces-zones where particles tend to remain for long durations. The Markov-based method automatically identified dynamically relevant areas, like recirculation cells and retention zones, by analyzing eigenvectors of the transition matrix. This enabled researchers to discover hidden structures in the Gulf's complex circulation influenced by bathymetry and measure retention and exchange between provinces.

The same approach was also used in the search for Malaysian Airlines flight MH370. Researchers built transition matrices from decades of surface drifter data in the Indian Ocean. They ran the model in reverse to identify the likely origin of debris found near Reunion Island. This method considered both average currents and unresolved small-scale dispersion, making it much more reliable than deterministic backtracking, which can quickly become inaccurate due to changes in the ocean. The Markov method not only significantly reduced the search area for the crash site but also naturally included uncertainty in the model output. This offered probabilistic source distributions that helped decision-makers during an investigation with limited data.

4. Gap of knowledge

Despite this work from a large-scale perspective on dispersion and ocean currents in the ocean, the role of sub-mesoscale features on the movement of surface contaminants has not been addressed very widely, particularly in semi-enclosed basins such as the Mediterranean. There is a need for high-resolution, seasonal assessments of transport dynamics that explicitly account for resolution-dependent dispersion and the identification of natural boundaries to inform environmental policy.

Markov chain models have shown strong potential in other oceanic regions, but their application to the Mediterranean Sea has been limited. This gap is due in part to the lack of long-term, high-resolution Lagrangian datasets in the region. Unlike the Gulf of Mexico or Indian Ocean, the Mediterranean has fewer float deployments and limited access to a broad surface drifter recordings.

Seasonal variability further complicates this issue. Circulation in the Mediterranean Sea drastically shifts between summer and winter because of shifts in wind forcing, temperature gradients, and stratification. Such shifts may change sub-mesoscale activity's strength and structure and consequently affect contaminant pathways and residence times. Nevertheless, several models are not able to resolve temporal variability. Without temporal resolution on the seasons, important dynamics-like summertime retention zones or deep mixing in winter- can be neglected and diminish the applicability and accuracy of forecast models. Therefore, to capture and control contaminant dispersion in the Mediterranean effectively, it is evident that temporally resolved data and seasonally varying modeling paradigms are crucially needed.

5. Experimental and Methods

The experiment involved collecting velocity data, running simulations, and applying mathematical tools to develop and analyze a statistical framework. All tasks were done using custom Python scripts.

5.1.Velocity Field and Grid Description

High-resolution, satellite-based ocean circulation models of the Mediterranean were used to obtain velocity data for two different spatial resolutions: 300 meters and 3 kilometers. Each of the datasets is made of horizontal velocity fields (u and v components) sampled hourly and spanning a 37-day duration in two representative seasons: winter (January 2018) and summer (July 2018). These gridded fields of velocity are saved in NetCDF files with variables representing horizontal components of velocity of the sea water (u and v), defined on a structured grid.

5.2. Lagrangian Simulation with Parcels

To simulate the transport of particles under oceanic surface flow, we used the *Parcels Python framework.* The simulation setup included releasing particles at every ocean psi-point within a 6° × 6° domain (31°N to 37°N latitude, 31°E to 37°E longitude). Land grid points were deleted using a land-sea mask. The number of released particles varied by resolution: 2,600,729 for the 300-meter grid and 159,733 for the 3-kilometer grid.

Particles were advected by surface currents for 37 days in both summer (July 2018) and winter (January 2018) simulations. Position data were recorded every 15 min. Parcels performs trilinear interpolation of the velocity fields in both time and space, after interpolating the staggered u and v components to the psi-points where particle velocities are computed. The framework allows the user to define simulation parameters such as duration, output frequency, particle release coordinates, oversampling rate (temporal interpolation), and number of concurrent MPI processes. The simulation outputs are saved in .zarr format, which efficiently stores particle positions over time and space and is well-suited for post-processing in large-scale studies. Later, custom Python scripts were used to analyze and visualize the resulting trajectories, as shown in *Figure 2*.

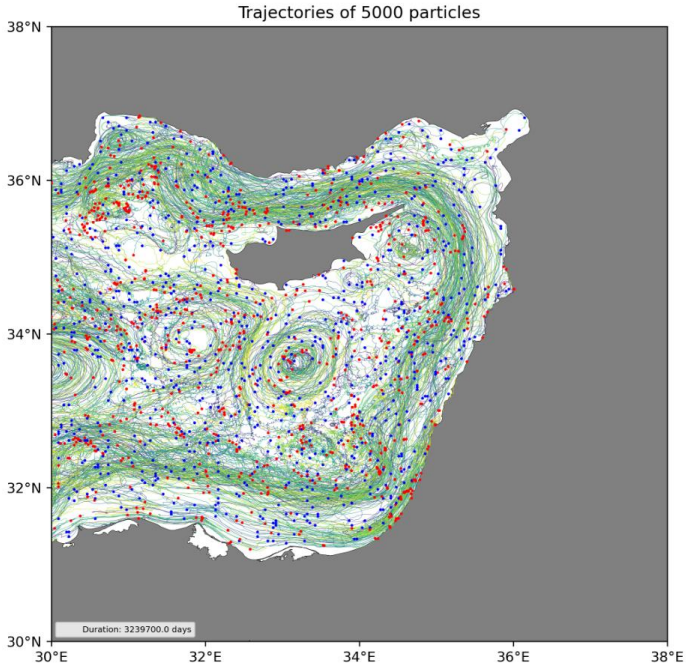


Figure 2 - Trajectories of 5000 random particles over 37 days. Blue – starting positions. Red – ending position.

5.3. Markov Chains and the Transition Matrix

Markov chains are mathematical models used to describe systems that undergo transitions from one state to another in a memoryless fashion. In a discrete-time Markov chain, the probability of transitioning to the next state depends only on the current state- not on past states. This is known as the Markov property. The system's behavior is captured by a transition matrix P, where each element P_{ij} represents the probability of transitioning from state i to state j in one time step:

$$(1)\quad P_{ij} = \Pr\left(X_{t+1} = s_j \mid X_t = s_i, X_{t-1} = s_{i_{t-1}}, \dots, X_0 = s_{i0}\right) = \Pr\left(X_{t+1} = s_j \mid X_t = s_i\right)$$

The matrix is row-stochastic, meaning the sum of each row is 1, representing the total probability distribution across possible next states:

$$(2)\quad \sum_{j=1}^{n} P_{ij} = 1$$

In this project, the Mediterranean Sea domain was divided into a grid of square bins (spatial cells) as shown in *figure 3*, each was considered as a distinct Markov state. The width of the bins is changable. While we used simulated Lagrangian particle trajectories, we enlarged statistical robustness by having multiple transitions from each trajectory: For each time step along a particle's path, we recorded its bin location and

its location after a fixed time interval *T*. This, effectively, treated each time step as a virtual particle, generating many transition observations per trajectory. These were then used to count transitions from bin *i* to bin *j*, n_{ij}, and compute the transition probability:

$$(3)\ P_{ij} = \frac{n_{ij}}{n_i}$$

where n_i is the total number of particles that started in bin *i*.

It is important to note that increasing the number of particles significantly improves the statistical robustness of the transition matrix. The high particle count ensures that even low-probability transitions are captured, reduces statistical noise, and enhances the accuracy of the resulting spectral analysis.

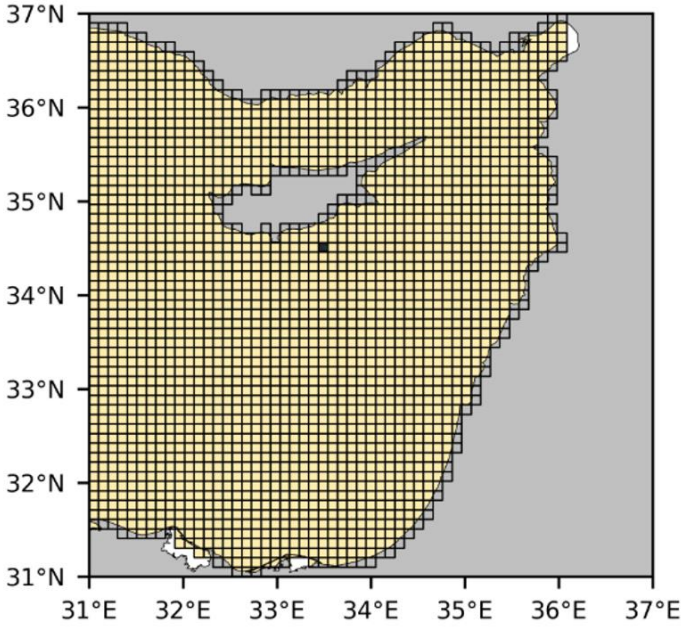


Figure 3 - 10km wide bins structure, with initial contamination within one bin (marked in purple)

This transition matrix enables prediction of future particle distributions. If p_t is a columb vector representing the probability distribution at time *t*, then the distribution after time interval *T* is:

$$(4)\ p_{t+T} = Pp_t$$

This sequence is called "*push forward*". As a result, from *t*=0, and after *n "push forward"* with time interval *T*, recursively:

$$(5)\ p_{0+nT} = P^n p_0$$

This recursive property is essential for modeling long-term contaminant spread or retention zones. In this experiment, we simulate the spread of contamination by initializing a probability vector p_0 with all zeros except for a one in the position corresponding to a single chosen bin, i.e., $p_0 = (0,0,0, \dots ,1, \dots ,0)$. This setup represents a contaminant initially localized at one specific location, as seen in *figure 3*. We then apply the transition matrix iteratively over *n* time steps, representing a total evolution over one month (31 days).

There is an important complication: it is not a closed basin, so some of the particles leave the domain when it's being simulated (namely, flow off over open boundaries), and some others can come in from elsewhere. It causes a violation of the row-stochastic nature of the transition matrix because of total probability not conserving within the confined region anymore. To correct for this, we introduce two additional "imaginary" bins: f_o, representing flux out of the region, and f_i, representing flux into the region. These bins are added as the last row and column of the transition matrix (*N+1)*. Outgoing particles that leave the domain will be directed into f_o, ensuring that the total transition probability from each row sums to 1. For incoming particles, (which is not mandatory, because it does not change the transition matrix) we split the f_i row such that it distributes probability uniformly across all valid spatial bins.

Additionally, one key assumption of this method is that the dynamics are Markovian-that is, particle transitions are independent of past movements. In order to validate this assumption, we evaluated the Lagrangian autocorrelation of particle velocities, which measures how correlated the velocity of a particle is with its past velocity over time lag *T*. The Lagrangian autocorrelation function (for 2 horizontal velocities *u, v*) is defined as:

$$(6)\quad R_{ij}(T) = \frac{\langle (u_i(t) - U_i)(u_i(t+T) - U_i) \rangle}{\langle (u_i(t) - U_i)^2 \rangle}$$

where $u_i(t)$ is the velocity vector in the the *i'th* direction of a particle at time *t (u,v)*. U_i is the lagrangian mean velocity in the *i'th* direction. A rapid decay of the function below $1/e$, and towards zero, tells us that there is loss of memory in the system, justifying the Markov assumption.

5.4. Spectral Analysis of the Transition Matrix

To understand the structure and dynamics of ocean transport, we perform a spectral analysis of the Markov transition matrix. Spectral analysis involves computing the eigenvalues and eigenvectors of the matrix, which provide insights into long-term behavior, mixing patterns, and the presence of coherent structures in the flow field.

Let *P* be the row-stochastic transition matrix constructed as described above. Some of the spectral properties *P* of are as follows:

- The largest eigenvalue λ_1 of *P* is always equal to 1. This corresponds to the *"stationary distribution"* - the eigenvector x_1 that satisfies:

  $$(7)\ Px_1 = \lambda_1 x_1 = x_1$$

  This stationary vector represents the long-term probability distribution of particles across the spatial domain, assuming the system is closed. It describes the steady-state in which the probabilities no longer change with time.
- The subdominant eigenvalues (i.e., those with $|\lambda_i| < 1$) and their corresponding eigenvectors capture almost-invariant sets - regions where particles tend to remain for long durations. The closer an eigenvalue is to 1, the more slowly the corresponding mode to decay over time.
- Each eigenvector represents a spatial pattern. When projected onto a probability vector p, these patterns determine how much influence each mode has on the evolution of the contaminant. Mathematically, the state at time step *n* can be shown as:

  $$(8)\ p_n = \sum_{k=1}^{N} a_k \lambda_k^n x_k$$

  where a_k are coefficients determined by the probability distribution and x_k are the eigenvectors of *P*. The first few eigenvectors - particularly those associated with the largest eigenvalues - reveal dominant flow structures and indicate where particles are most likely to accumulate or persist in the long term.

5.5. Clustering

In order to reveal dynamically consistent regions of the Mediterranean Sea, we applied clustering to dominant eigenvectors of transition matrices. The general idea is that it is probable for spatial bins to have similar values for dominant eigenvectors and will also exhibit similar long-term transport behavior. Such regions are prone to having

strong inner connectivity—particles pass through easily within them—and weak communication with adjacent regions and are therefore dynamically consistent. We implemented two main clustering approaches:

- *K-Means Clustering:* This standard algorithm partitions the spatial domain into K cluster. The eigenvectors are interpreted as feature vectors for each bin, and bins with similar spectral properties are grouped together. Mathematically, the algorithm seeks to minimize the function:

$$(10)\ J = \sum_{k=1}^{K} \sum_{x_i \epsilon C_k} \|x_i - \mu_k\|^2$$

  where x_i is the $i'th$ chosen eigenvector, C_k is the set of bins assigned to cluser number k, and μ_k is the controid of that cluster. K-means is simple and effective but can be sensitive to noise and the choice of K.
- Sparse Eigenbasis Approximation (SEBA): SEBA enhances the interpretability of the original eigenvectors by producing sparse and localized approximations of eigenvectors. It constructs a new set of orthogonal basis vectors, $\Phi = \{\phi_1, \phi_2, \phi_3, \dots, \phi_m\}$ that are linear combinations of the original eigenvectors $V = \{x_1, x_2, x_3, \dots, x_m\}$ and that minimize the overlap between spatial supports. This is written as a sparse rotation problem:

$$(11)\ \Phi = VR$$

  where R is a rotation matrix chosen such that the resulting basis Φ is maximally sparse. Sparsity is computed via the optimization problem:

$$(12)\ \min_{R} \sum_{j=1}^{m} \|\phi_j\|_1 \ \ subject\ to\ R^T R = I$$

5.6.Choosing the right eigenvectors

Choosing the right set of eigenvectors is crucial, as these vectors define used to extract coherent features. Including noisy or irrelevant eigenvectors can damage meaningful patterns or introduce artifacts in clustering. As a results, we follow these guidelines to choose the perfect set of eigenvectors:

- Visualization of Eigenvectors: The leading eigenvectors are visualized to identify coherent structures. The inspection focuses on ensuring that the vectors are not fragmented and that they reflect physically meaningful structures.

- Push forward Map Comparison: To validate the clustering output, we simulate contaminant spread from within selected regions. Pushfor ward (multipling with the Transition Matrix) maps are generated over time and compared with SEBA and K-Means clusters. The results confirm that contaminant probability follows the contours of these regions and rarely crosses the boundaries.
- Eigenvalue Certainty Graph: To assess the stability of the spectral features, we construct 50 transition matrices using random 80% subsets of the particles. For each matrix, the first 20 eigenvalues are computed. The variance of each eigenvalue across all trials is plotted. Low variance suggests a stable model and reliable identification of coherent structures. We choose the first eigenvalues with the lowest variance.
- Cumulative Sparsity Index (for SEBA clustering): This is a diagnostic score to assess how well the first *k* sparse vectors separate coherent features when we input *r* chosen eigenvectors into SEBA.

$$(13)\quad Max\left(\Phi^{(r)}, k\right) := -\sum_{j=1}^{k} \min_{i} \Phi_{ij}$$

Where $\Phi^{(r)}$ is sparse matrix output by SEBA when using r chosen eigenvectors as input. The max the better.

5.7. Hitting time and time connectivity provinces matrix

Markov chain theory enables the estimation of the mean hitting time-the average number of time steps required for a particle starting in one region to reach another. Let *A* be the starting region, and let *B* the target region (e.g., different provinces identified through clustering). To compute the mean time it takes a particle to reach *B* from each bin in *A*, we solve the system:

$$(14)\quad \left(I - P_{|\bar{B}}\right)\tau = 1$$

where $P_{|\bar{B}}$ is the transition matrix restricted to bins not in *B*, and τ gives the expected hitting time from each bin in $\bar{B}$ to region *B*. $\tau_j = 0$ for all $j \in B$, meaning the hitting time is zero for bins already in the same region. The solution yields a vector of hitting times from each bin in region *A* to region *B,* which can be then averaged to get a single value representing the average time it takes for particles to transition between

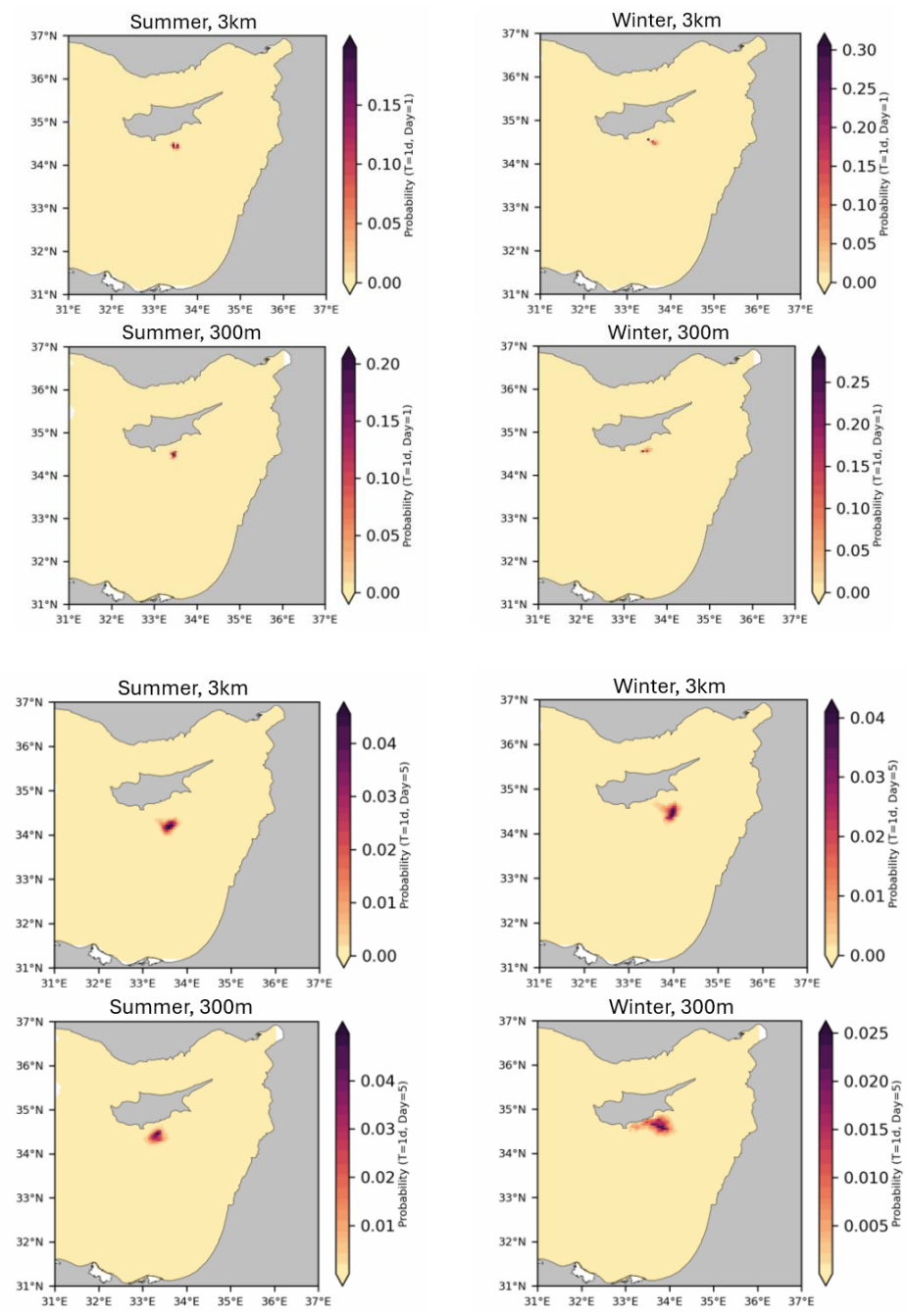
provinces. This method provides a measure of connectivity and transport efficiency between provinces.

6. Results

6.1. Lagrangian Time Lag (*T*)

In accordance with Equation (6), the Lagrangian autocorrelation was evaluated for each season and grid resolution. Transition time of ***T*=1 day** was chosen, for all configurations. A bin size is approximately *5km x 5km.*

6.2. Push Forward Maps at different resolutions and seasons:



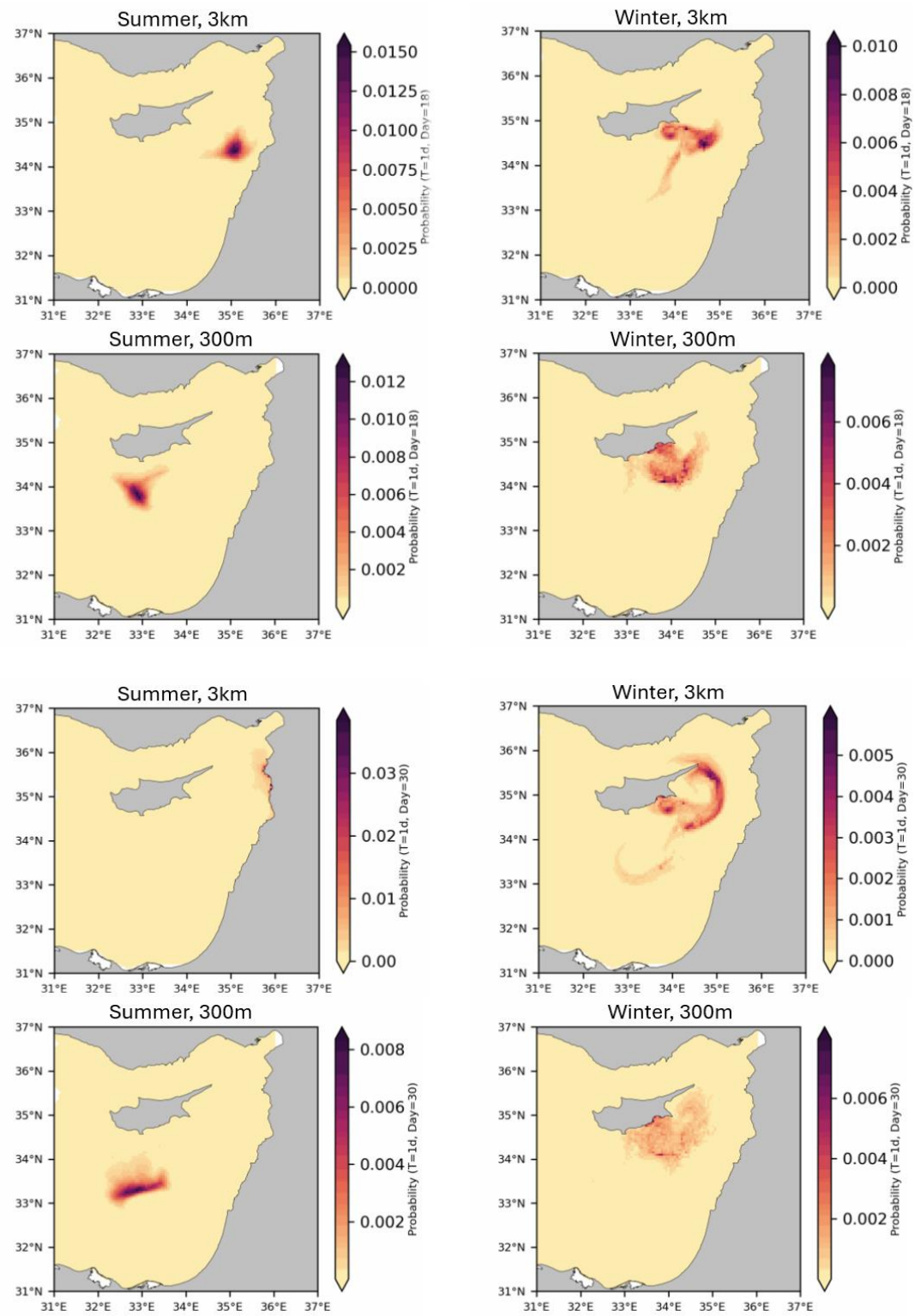


Figure 4 - Dispersion at different resolutions and seasons, after 1, 8, 18 and 30 days

6.3. Lagrangian Geography Maps

Based on the SEBA diagnostics and the certainty eigenvalue graph, we selected the following eigenvectors for each configuration:

	Summer	Winter
3km	*2 to 8*	*k = 2 to 9, r = 16*
300m	*2 to 5*	*2 to 6*

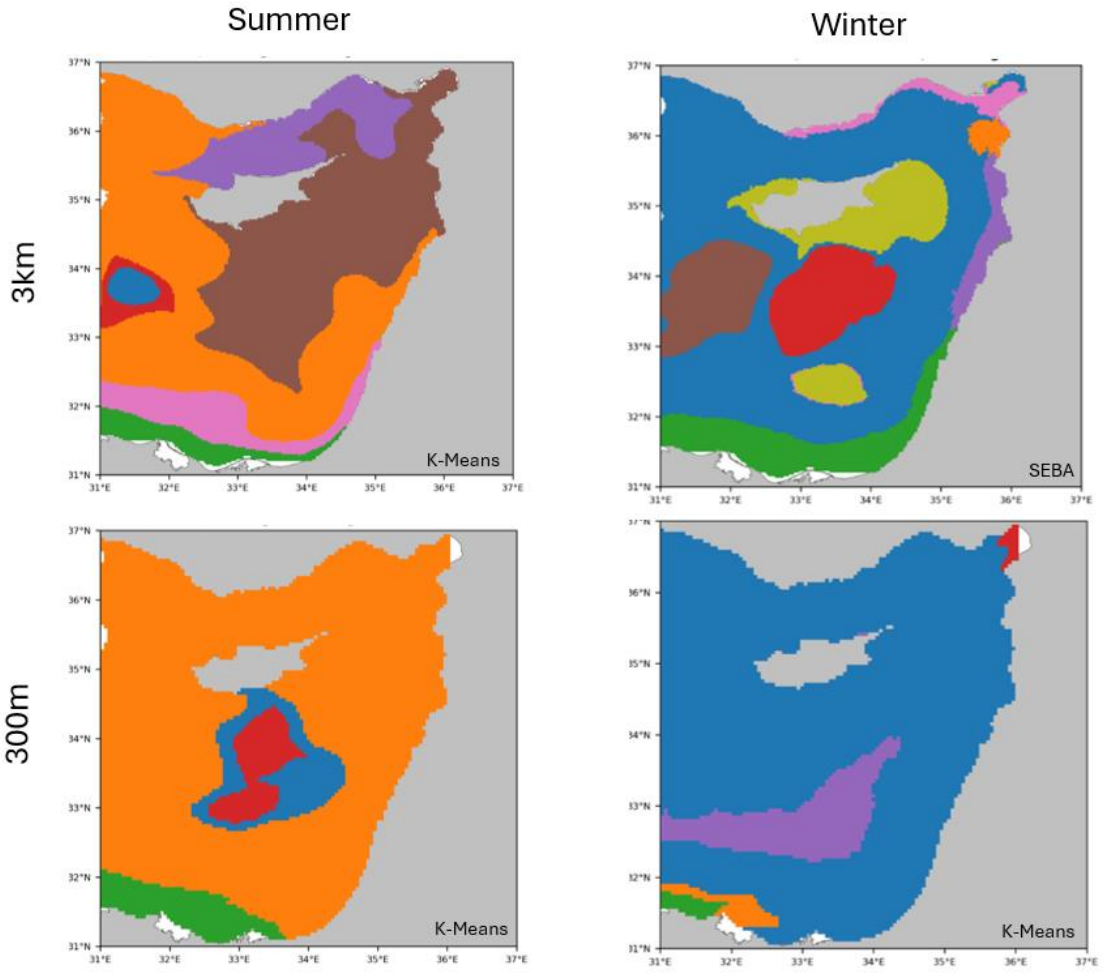


Figure 5 - Lagrangian Geography Maps, captured after using K-Means and SEBA algorithms

6.4. Hitting Time – has not fully implemented, yet

7. Discussion

7.1. Dispersion Patterns

Figure 4 demonstrate a clear difference in dispersion patterns between both resolution scales and seasonal conditions:

- **Summer (3 km resolution):** The contaminant moves eastward toward the Syrian coast and then curves northward. This trajectory likely reflects the influence of the alongshore Levantine Current, which flows cyclonically along the eastern boundary of the Mediterranean.
- **Summer (300 m resolution):** In comparison, the dispersion pattern exhibits a southwestward displacement with the plume initiating a left turn and traveling along the Israeli coast. Higher resolution better resolves smaller sub-mesoscale elements, probably eddies and filaments, by which particles are reoriented and distributed more laterally around the basin.
- **Winter (3 km resolution):** Dispersal is more localised with strong retention of the plume around the original release area and an evident recirculation feature to the north of Cyprus. Such a signature is characteristic of the well-known Cyprus Eddy, which is a geostrophic circulation becoming prominent throughout winter. Such 3 km resolution will resolve this mesoscale feature.
- **Winter (300m resolution):** The dispersion appears more uniform and widespread near the release site, with no indication of the Cyprus Eddy or any dominant mesoscale currents. The lack of a clear circulation pattern may be due to the increased presence of sub-mesoscale turbulence, which disperses the particles more eqally and takes them aways from larger-scale flows.

Here we see that while the 3 km grid for itself resolves mesoscale phenomena like boundary currents and eddies, sub-mesoscale dynamics are introduced by the 300m grid and enhance dispersion and alter trajectories. It thus highlights the two-fold significance of both the mesoscale and sub-mesoscale processes in outlining pollutant trajectories in the Mediterranean Sea with both scales having different aspects of the transport mechanism.

An important practical advantage we see in here is the efficiency in forecasting the spread of contamination once the transition matrix is computed. Although its

accuracy is still speculated, the matrix itself, extracted as a data structure, tells us the probability of particle movement. Given this matrix and a defined initial distribution of contamination (e.g., a unit vector indicating the release location), it becomes computationally easy to push forward the system through time by repeated matrix multiplication. This allows immediate estimation of future contaminant positions, in contrast to dynamic re-simulation.

7.2. Lagrangian Geography

The comparison of coherent Lagrangian regions derived from clustering in both the 300 m and 3 km simulations for summer and winter (*figure 5*) reveals several important differences in regional connectivity and transport dynamics.

SEBA and K-means clustering methods provided meaningful partitions across resolutions and seasons, but important differences appear when examining the eastern boundary current near Israel. In the 3 km simulations, both summer and winter maps clearly show a narrow, streched region near the Israeli coastline. This corresponds to the Levantine boundary current, a strong and persistent feature that transports water northward along the continental shelf. The existence of this current as a distinct transport pathway is well captured in the 3km resolution due to its more geostrophic (large-scale) behavior.

In contrast, the 300m high-resolution simulations do not show a similarly isolated boundary region. Instead, the clustering reveals more homogeneous connectivity along the coast. This change is likely due to the influence of submesoscale currents, which dominate at finer resolutions. Once again, these smaller-scale processes act to enhance mixing and break down transport barriers, and increasing the probability of exchange between regions. As a result, the boundary current is effectively gets bigger.

Overall, the connectivity in the 3 km resolution is more distinct, with clearly defined borders between provinces. Meanwhile, in the 300m simulations, many regions are more connected, which results in fewer, larger coherent areas due to higher rates of

inter-bin exchange. This also reinforces the assumption that submesoscale processes enhance lateral dispersion, particularly at boundaries and in eddy-influenced zones.

Beyond the most significant results, secondary patterns within the clustering maps enrich insight into Mediterranean Sea behaviour. At both summer and winter 3 km scales, an east–west separation appears most prominently in the northern basin near the Syrian and Turkish coasts. This separation reflects a dominance whereby powerful zonal current systems such as the Asia Minor Current establish barriers to transportation. Curiously, the barriers appear far less prominent at 300 m scale, when the cross-basin mixing characteristic of submesoscale behaviour reduces east–west coherence.

One pattern in particular jumps out in the winter 3 km SEBA map: two well-separated and tightly compacted eddy-like features in the Levantine Basin interior. These are typical of seasonal eddy strengthening in winter and create Lagrangian matter traps. In comparison with the finer-resolution 300 m model outputs, we have more fragmented and inter-connected regions, suggesting that motions at the submesoscale allow communication even among the mesoscale features.

Offshore, the 3 km K-means clustering shows a prominent separation of narrow shelf areas (e.g., Lebanese and southern Syrian shores) and neighboring offshore waters, indicating limited cross-shelf communication. At 300 m, coastal regions intermix more readily with offshore domains, in accord with increased lateral stirring that we expect through submesoscale movements.

7.3. Further Reseach

One important direction for further research is the extension of this framework to three-dimensional (3D) models that capture vertical transport processes. While the current study focuses on surface dynamics, vertical motion plays a critical role in ocean mixing, biological productivity, and pollutant dispersion - especially in submesoscale regimes where vertical velocities can be on the order of meters per day. Making a 3D Transition Matirx, informed by high-resolution data, would allow the detection of vertical retention zones, subsurface eddies, and upwelling/downwelling effects. Such

imporvements is essential for evaluating the fate of pollutants that undergo vertical cycling, including microplastics and oil droplets.

Another point for further tasks is generalizing the analysis across multiple years to establish anual consistency and assess interannual variability. The current results are based on single summer and winter simulations from 2018, which may not capture the full range of Mediterranean dynamical regimes. By repeating the methodology across a selected months and years, would allow to quantify how robust the identified coherent regions are under different forcing conditions (e.g., El Niño years, anomalous wind patterns, or changes in thermohaline structure). This would strengthen the applicability of the Markov chain-based framework for long-term environmental planning, marine spatial management, and emergency response forecasting.

8. References

- Froyland, G., Rock, C. P., & Sakellariou, K. (2019). Sparse eigenbasis approximation: Multiple feature extraction across spatiotemporal scales with application to coherent set identification. Communications in Nonlinear Science and Numerical Simulation, 77, 81–107.
- Taylor, J. R., & Thompson, A. F. (2023). Submesoscale dynamics in the upper ocean. Annual Review of Fluid Mechanics, 55, 103–127.
- Alhammoud, B., Béranger, K., Mortier, L., Crépon, M., & Dekeyser, I. (2005). Surface circulation of the Levantine Basin: Comparison of model results with observations. Progress in Oceanography, 66(4), 299–320.
- Miron, P., Beron-Vera, F. J., Olascoaga, M. J., & Froyland, G. (2017). A Lagrangian geography of the deep Gulf of Mexico. RAUGM Conference Presentation, October 24, 2017.
- Miron, P., Beron-Vera, F. J., Olascoaga, M. J., & Koltai, P. (2019). Markov-chain-inspired search for MH370. arXiv preprint.
- van Duinen, B., Kaandorp, M. L. A., & van Sebille, E. (2022). Identifying marine sources of beached plastics through a Bayesian framework: Application to Southwest Netherlands. Geophysical Research Letters, 49(17), e2022GL098447.
- Miron, P., Beron-Vera, F. J., Olascoaga, M. J., Froyland, G., Pérez-Brunius, P., & Sheinbaum, J. (2019). Lagrangian geography of the deep Gulf of Mexico. Journal of Physical Oceanography. https://doi.org/10.1175/JPO-D-18-0123.1
- LaCasce, J. H. (2008). Statistics from Lagrangian observations. Progress in Oceanography, 77(1), 1–29.
- Fadida, Y., Verma, V., Barkan, R., Biton, E., Solodoch, A., & Lehahn, Y. (n.d.). Submesoscale horizontal stirring enhances seasonal enrichment of low-chlorophyll surface waters. Nature Geoscience (preprint or in press).
- McWilliams, J. C. (2016). Submesoscale currents in the ocean. Proceedings of the Royal Society A, 472(2189), 20160117.
- Parcels Python Library: https://github.com/OceanParcels/Parcels

9. Appendices:

9.1. Lagrangian Lag-Autocorrelation example – 300m, winter

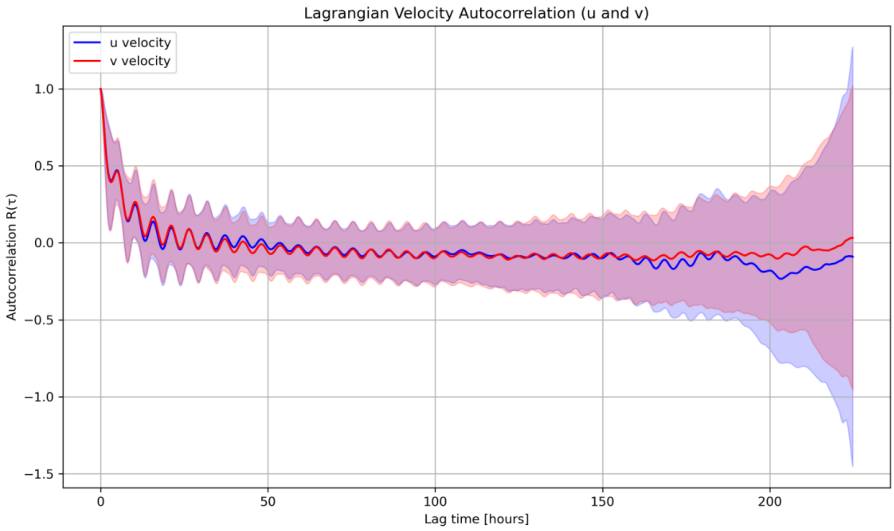


Figure 6 – Lagrangian Velocity Autocorrelation

9.2. Eigenvalues map (excluding $\lambda = 1$) - 300m, winter, 10 km wide bins

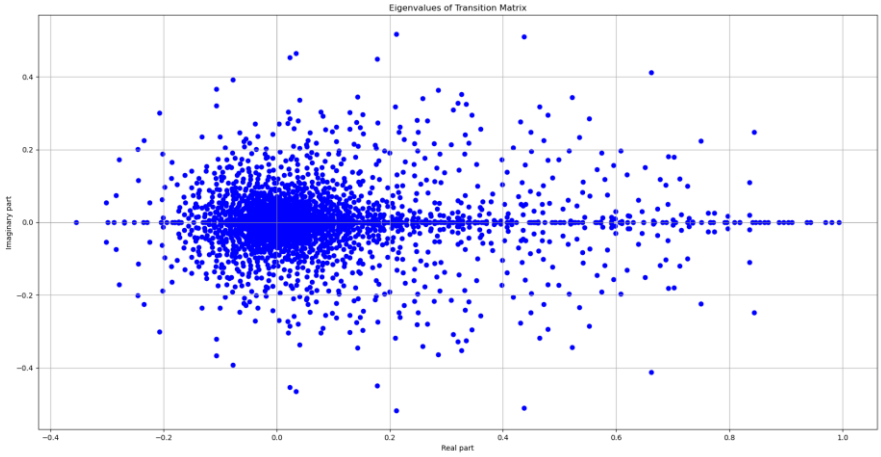


Figure 7 – Eigenvalues map example

9.3. Eigenvectors example – 300m, winter, 5km wide bin

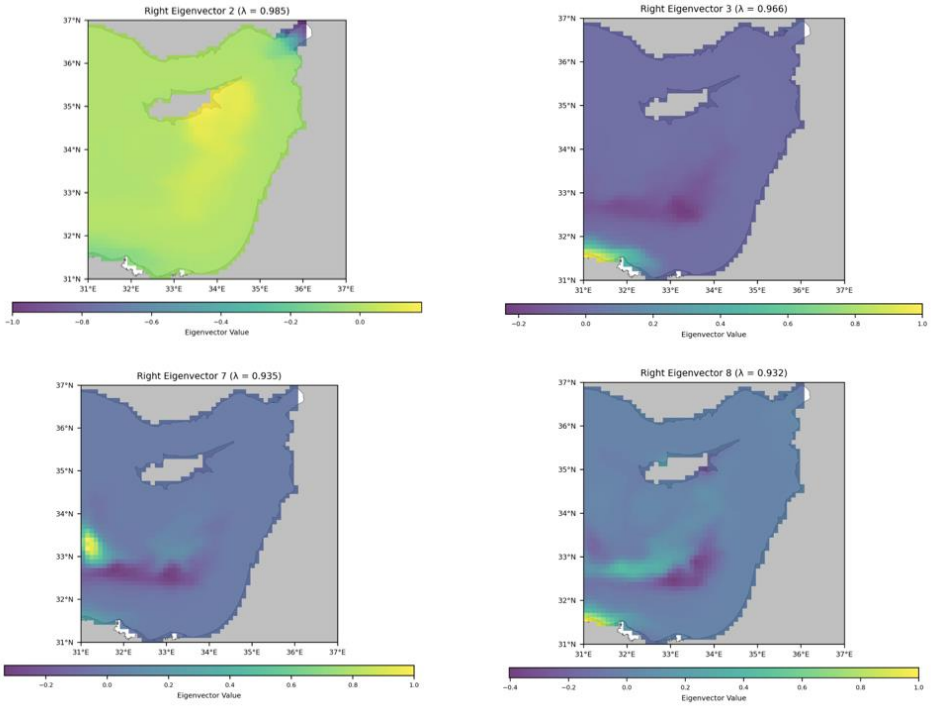


Figure 8 – 4 different eigenvectors corresponding to their eigenvalues. We can see different basins and structures forming for different eigenvalues, simbolizing different retention times of particles in these areas

9.4. Eigenvalue Cetainty example – 300m, winter, 5 km wide bins

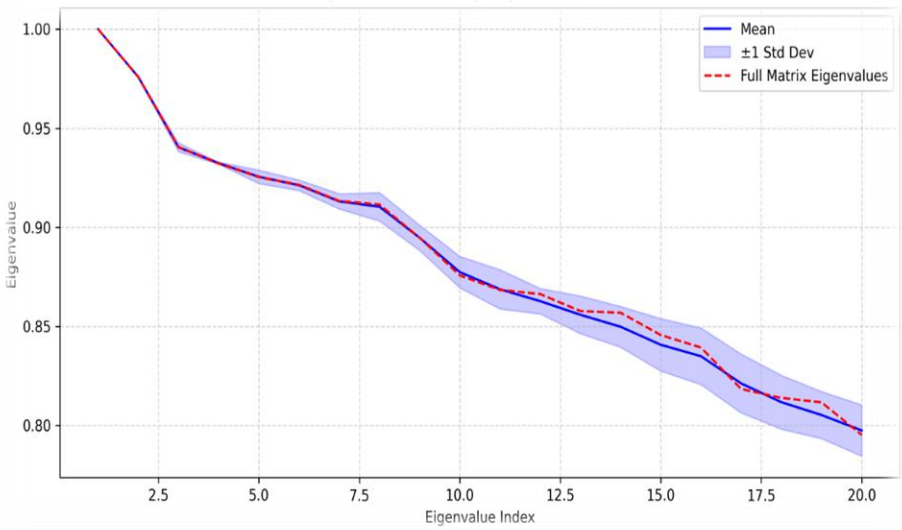


Figure 9 – Eigenvalues Certainty graph – for 50 sets of different 80% of the total particles. Dashed red line is the full matrix eigenvalues, and the blue line is the avg of the eigenvalues of different sets

9.5. Seba diagnostics example – 3km, summer, 5km wide bins:

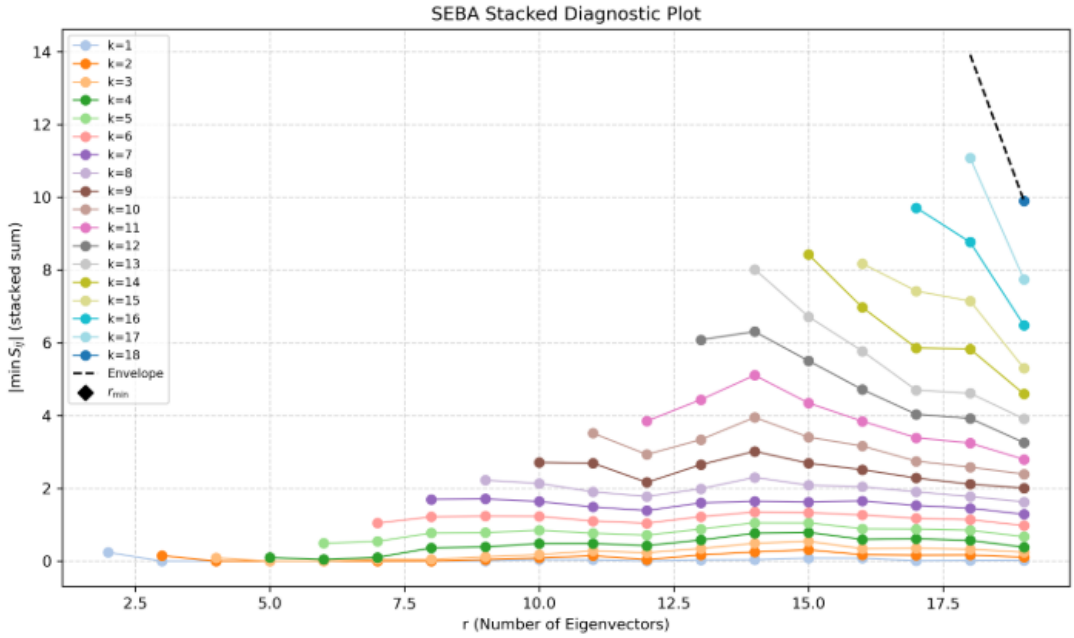


Figure 10 – Seba Dianostics. Each set is curve reference to different summasion methods. For k=9, for example, sum the first min sparse elements of the 9 first sparsed vectors. We are looking for rising trends. The curve with the most rising trend will be picked. That will determine k. After that, we are looking for the the max r in that trend – that will determine r.

9.6. Example of K-Means clusters – 3km, summer, 5km wide bins

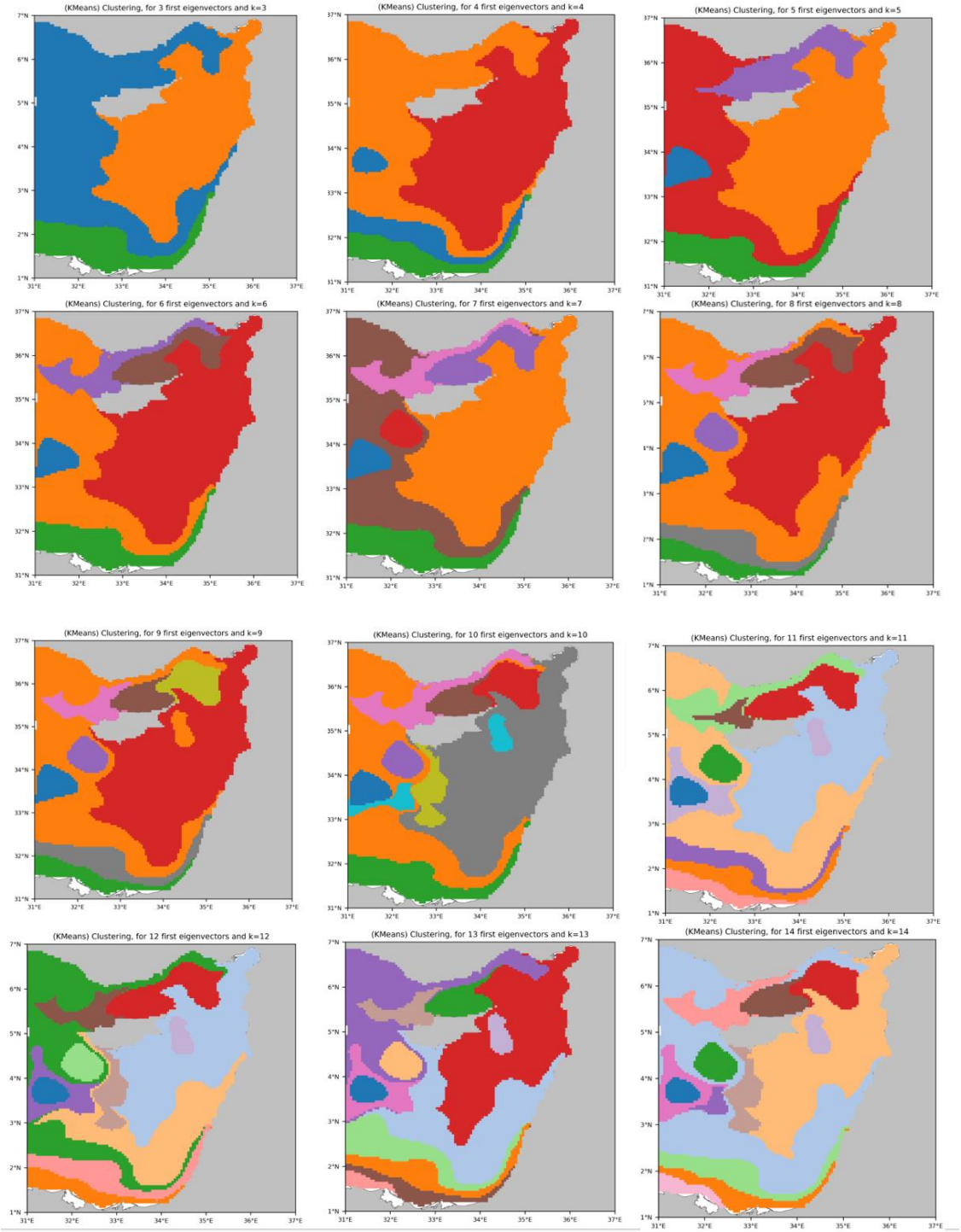


Figure 11 – different k-Means clusters, 3km, summer, 5km wide bins

9.7. Examples of SEBA clusters – 3km, summer, 5km wide bins

18 first eigenvectors were input to SEBA:

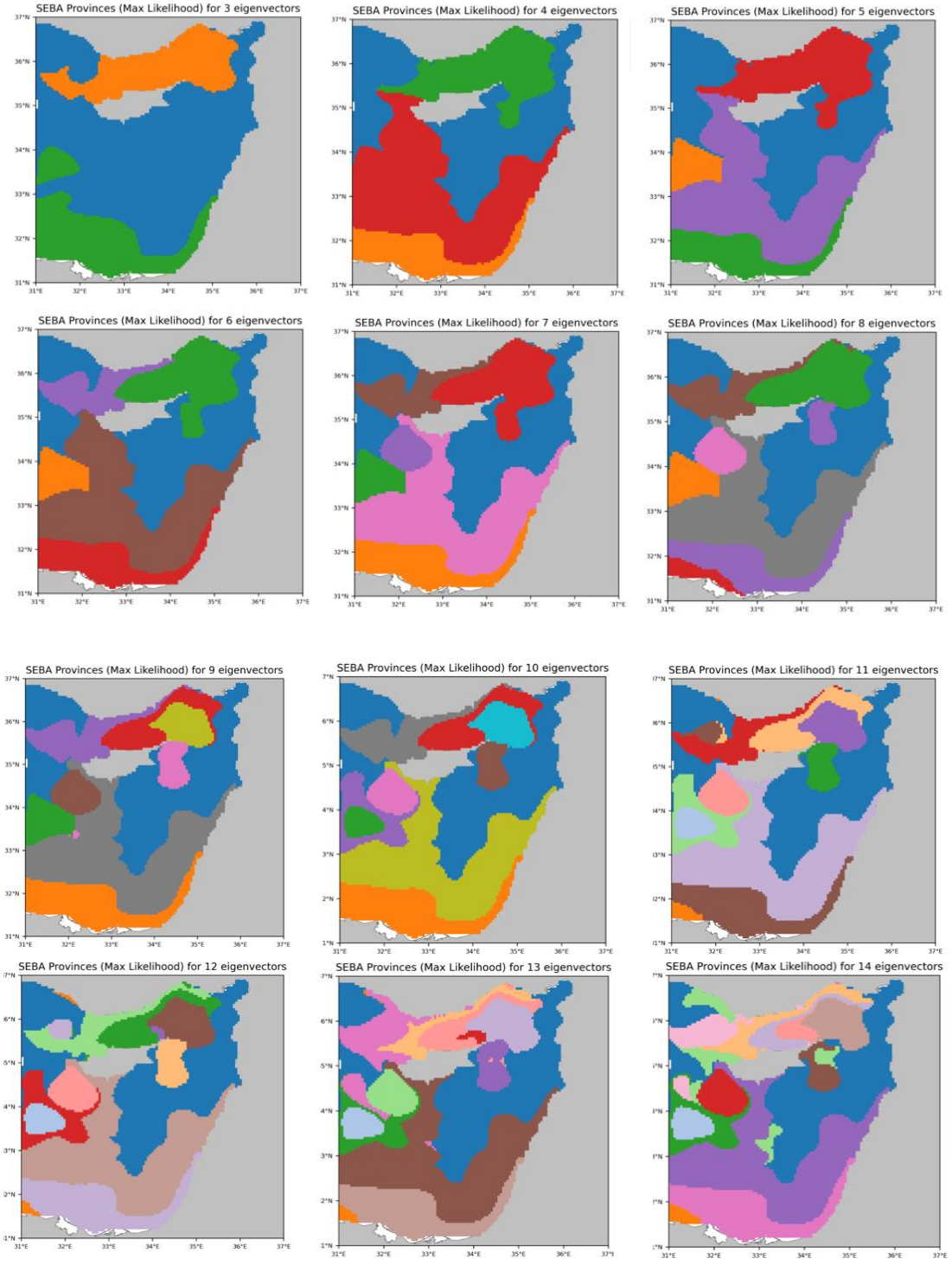


Figure 12 – Different SEBA clusters for different k, for r=18, 5 km wide bins

9.8. Python scripts documentation:

Link to GitHub: https://github.com/ShaiPollak/Final_Project#