Environmental Data and Gaussian Processes
Written Analysis Report

Introduction

The Philippine Archipelago is a highly complex ocean region composed of many islands, narrow straits, shallow shelves, and deep basins.
Because of this complicated geometry, ocean currents in this area show strong variations across space and time.
Understanding these flows is important for applications such as debris tracking, environmental monitoring, and search and rescue operations.

In this project, we analyze a time-varying two-dimensional ocean flow dataset collected in January 2009.
The data consists of horizontal and vertical velocity components measured at fixed spatial grid points over time.
Using this dataset, we first explore basic flow properties and correlations, then simulate the movement of particles under the influence of ocean currents.
Finally, we use Gaussian Process (GP) models to estimate unobserved flow values and extend simulations to longer time scales.
The main goal is to demonstrate how statistical modeling and simulation can be combined to make informed predictions in environmental systems.

Problem 2: Identifying Long-Range Correlations

The goal of this problem is to identify spatially distant regions within the Philippine Archipelago that show similar flow behavior over time.
Because the dataset contains over 500 by 500 spatial points, computing correlations between all possible pairs of points would be computationally infeasible.
Instead, we adopted a random sampling strategy.

I randomly selected a subset of spatial locations across the map and computed Pearson correlation coefficients between pairs of points.
For each pair, correlations were computed separately for the horizontal (u) and vertical (v) velocity components using their 100-time-point vectors.
To combine directional information, we used the average of the u and v correlations.
We defined a high correlation as an absolute correlation value greater than 0.7.

Using this approach, we identified two regions that are far apart spatially but still exhibit strongly correlated flow behavior.

Prepared by: Esmatullah Hassanzada

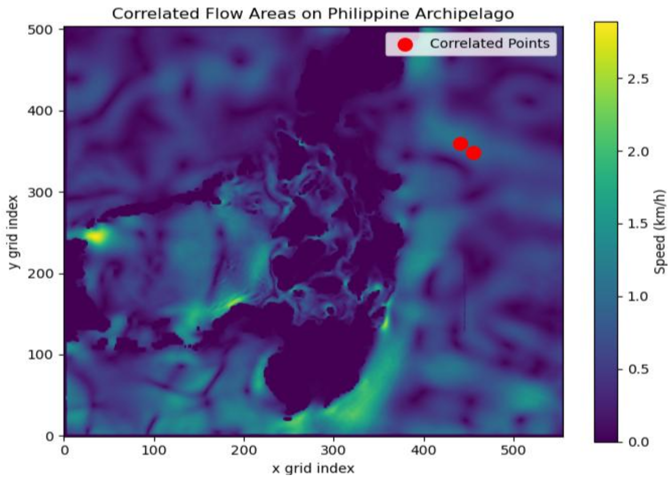

Figure 1. Map of the Philippine Archipelago showing two spatially distant regions with highly correlated flow behavior.

These regions were marked on a map of the archipelago.
Such correlations may arise due to large-scale circulation patterns or connected ocean basins that influence flow behavior across distant areas.

Problem 3.a: Simulating Particle Movement

In this problem, we developed a simulator to track the movement of particles under time-varying ocean flows.
The simulator assumes that a particle located at a given position moves with the same velocity as the ocean flow at the nearest grid point.
Particle positions are updated using a simple forward Euler method, where displacement is computed as the product of velocity and time step.

Particles were initialized uniformly across the entire spatial domain, including land areas, as instructed.
The time step was set to three hours, matching the temporal resolution of the dataset.
At each step, particle positions were updated using the flow velocity corresponding to the current time index.

I simulated particle trajectories for a total duration of 300 hours.
Plots were generated to show the initial particle positions, two intermediate states, and the final state.

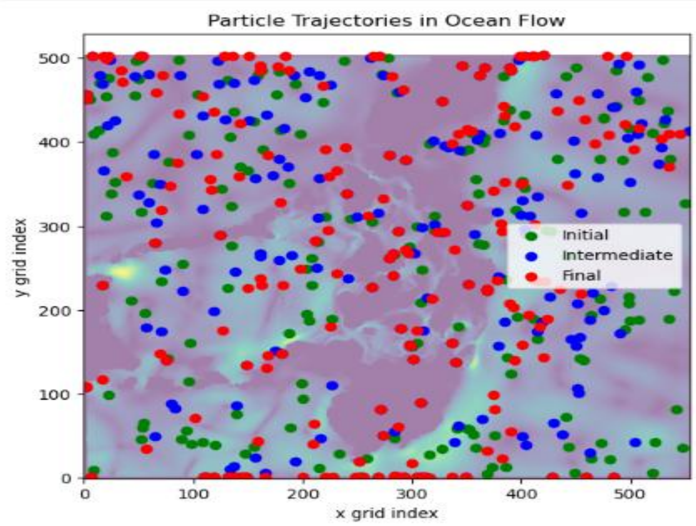

Figure 2. Particle trajectories in the ocean flow. Blue points indicate initial particle positions, green points show an intermediate state, and red points represent final positions after 300 hours of simulation.

The results demonstrate that particles tend to follow dominant flow pathways and exhibit coherent motion driven by the underlying ocean circulation.

Problem 3.b: Plane Crash Debris Simulation

In this scenario, we modeled a hypothetical plane crash north of Palawan, where the exact crash location is uncertain.
The initial debris locations were modeled using a two-dimensional Gaussian distribution with a specified mean and variance.
Particles sampled from this distribution were then advected using the particle simulator developed earlier.

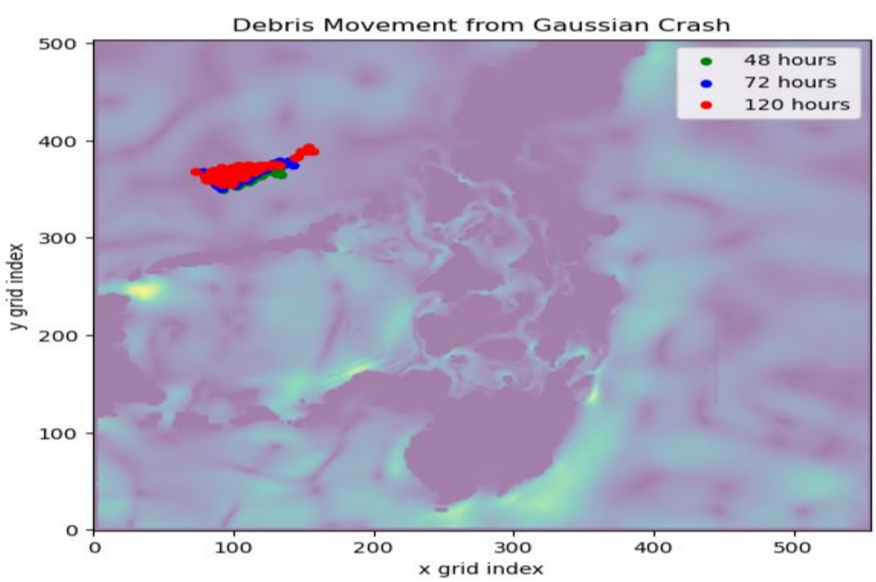

Figure 3. Figure 6: Simulated debris locations at 48, 72, and 120 hours for a fixed initial variance.

I simulated debris movement and examined expected debris locations at 48, 72, and 120 hours after the crash.
To study uncertainty, we varied the variance of the initial Gaussian distribution.
We observed that smaller variances resulted in more concentrated debris clusters, while larger variances produced wider spatial dispersion.
These results highlight how uncertainty in initial conditions can significantly affect search areas.
Larger uncertainty requires broader search coverage, while smaller uncertainty allows for more focused search efforts.

Problem 4: Gaussian Process Modeling of Flow

To model temporal variations in ocean flow at a fixed location, we used Gaussian Process regression.
For each selected spatial location, we modeled the horizontal (u) and vertical (v) velocity components independently as one-dimensional time series.

I selected the squared exponential kernel due to its smoothness and suitability for modeling physical processes that change gradually over time.
Kernel parameters, including the signal variance and length scale, were estimated using cross-validation and log-likelihood maximization.
The noise variance was treated as a small fixed parameter to ensure numerical stability.

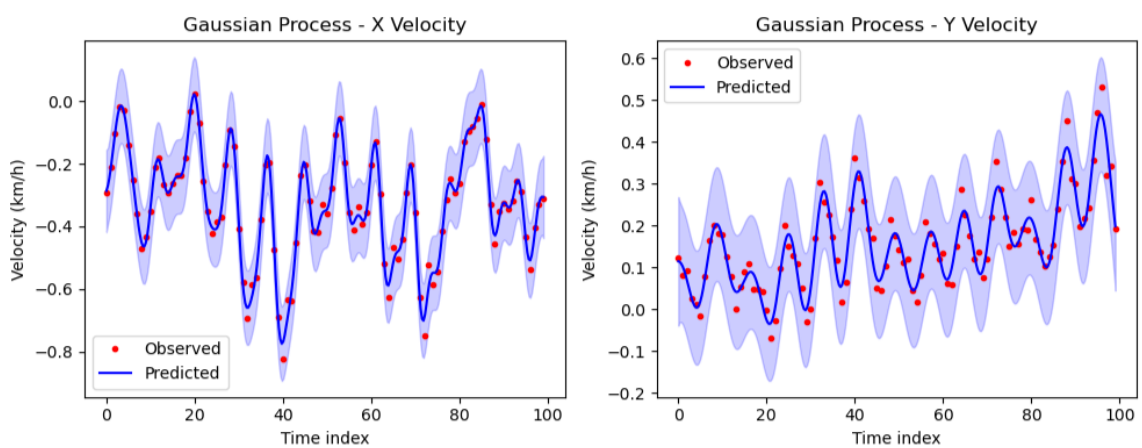

Figure 4. Log-likelihood values across kernel parameter choices used for Gaussian Process model selection..

Prepared by: Esmatullah Hassanzada

I tried to repeat this parameter estimation process for multiple spatial locations.
Across different locations, we observed that length-scale parameters were relatively consistent, suggesting similar smoothness.
However, variance parameters varied more significantly, reflecting spatial differences in flow intensity.

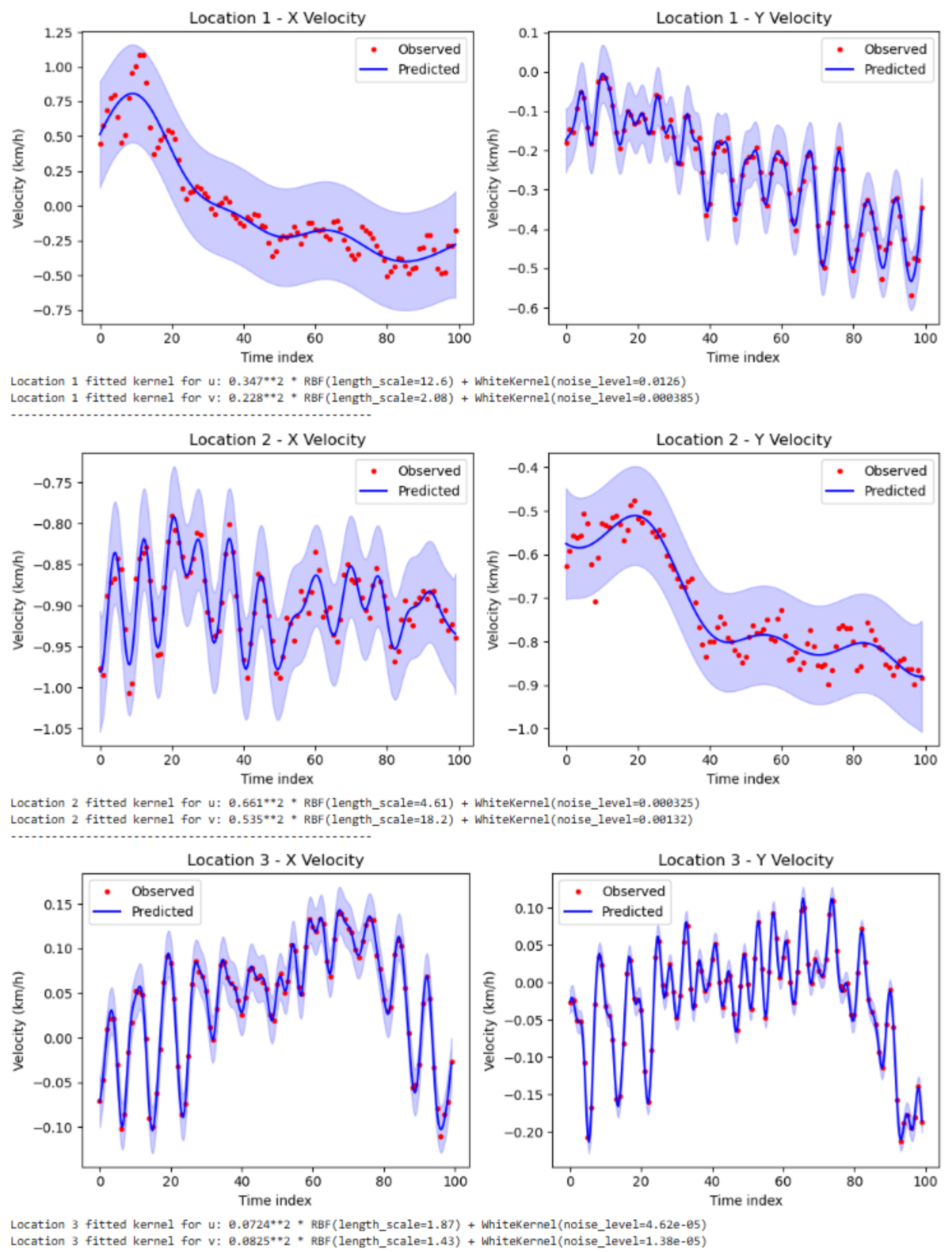

Figure 5. Comparison of estimated Gaussian Process kernel parameters across multiple spatial locations

Prepared by: Esmatullah Hassanzada

Problem 5: Estimating Unobserved Flow Data

The Gaussian Process models obtained in the previous problem were used to estimate flow values at unobserved time points.
Because the original data is assumed to be sampled every three days, we interpolated flow values at daily intervals.

Using the conditional distribution of the Gaussian Process, we computed predicted mean velocities and associated uncertainties.
Results were visualized by plotting the predicted mean along with three standard deviation confidence bands and observed data points.

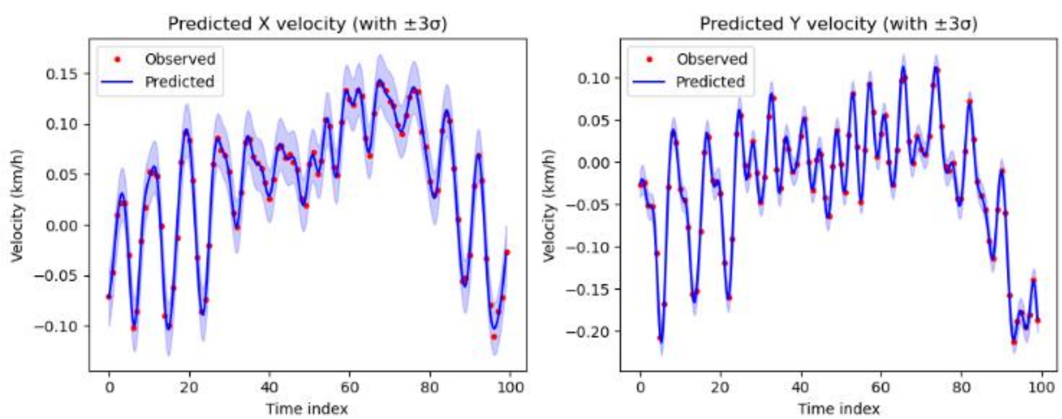

Figure 6. Gaussian Process interpolation of flow velocity showing predicted mean, uncertainty bands, and observed data.[i]

The interpolated flows were smooth and physically reasonable, with uncertainty increasing farther away from observed data points.
This interpolation step is essential for enabling simulations with finer temporal resolution.

Problem 6.a: Long-Term Debris Simulation

Using Gaussian Process-interpolated flow data, we extended the particle simulator to model debris movement over 300 days.
Representative kernel parameters were selected based on results from multiple spatial locations.

Particles were initialized using a Gaussian distribution representing the crash location and evolved forward in time.
Some particles eventually reached land, while others remained in the ocean.
By tracking landfall events, we identified coastal regions where debris is more likely to accumulate.

Plots were produced showing initial, intermediate, and final particle positions.
Based on these results, we identified one coastal location and one offshore location as priority search areas.

Problem 6.b: Placement of Monitoring Stations

In the final problem, we proposed locations for three coastal monitoring stations designed to track ocean debris.
I tried to initialize a large number of particles uniformly across the ocean domain, excluding land points, and simulated their movement using the long-term flow model.

Many particles eventually reached the coast.
By analyzing regions with high concentrations of particle landings, we identified three coastal locations that would be effective monitoring sites.

These locations were selected to maximize coverage of debris accumulation zones and to capture dominant flow pathways.
The results demonstrate how simulation-based analysis can support practical decision-making in environmental monitoring.

Conclusion

This project combined data analysis, particle simulation, and Gaussian Process modeling to study ocean flows in the Philippine Archipelago.
Through correlation analysis, we identified long-range flow relationships.
Particle simulations revealed dominant transport pathways and debris accumulation regions.

Gaussian Process models allowed us to estimate unobserved flow values and extend simulations to longer time scales.
Overall, this work demonstrates the value of statistical modeling and simulation for understanding and predicting complex environmental systems.

Prepared by: Esmatullah Hassanzada