

Vessel Path Identification in Short-Sea Shipping

iHelm Project

2023

Buro

MarineTraffic: Global Ship Tracking Intelligence | AIS Marine Traffic

Vessel Type: **Passenger Ship**

Size: (Length by Breadth) 19m * 6.41m;

MMSI: **265513810**

Area: **BALTIC - Kattegat**

Gross Tonnage: **68**

Average Speed recorded 8.2 knots(4.2 m/s)

Draught (Reported/Max): **2.5 m**

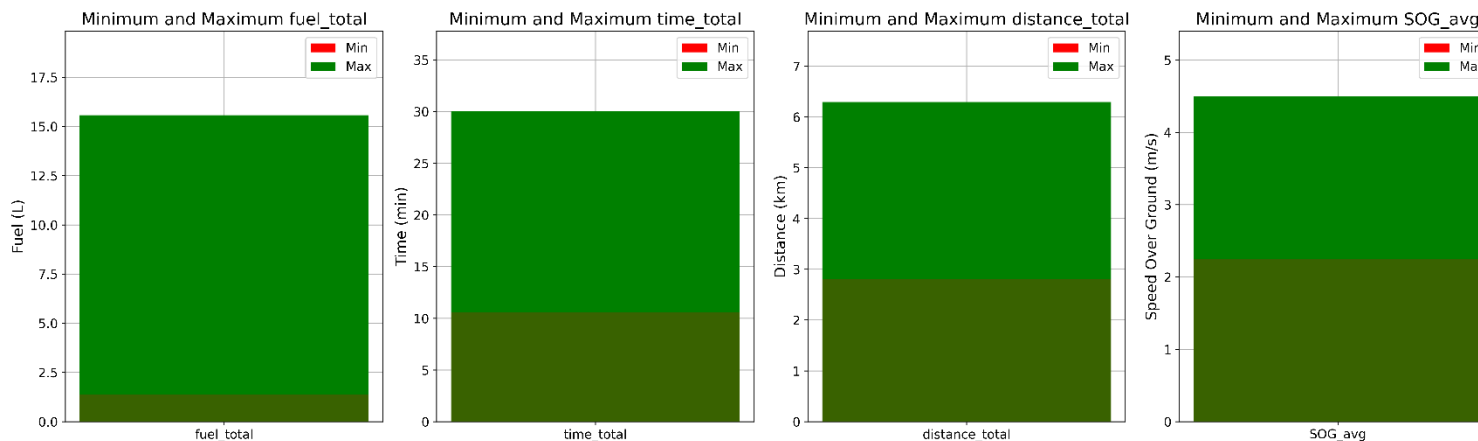
Flag: **Sweden (SE)**

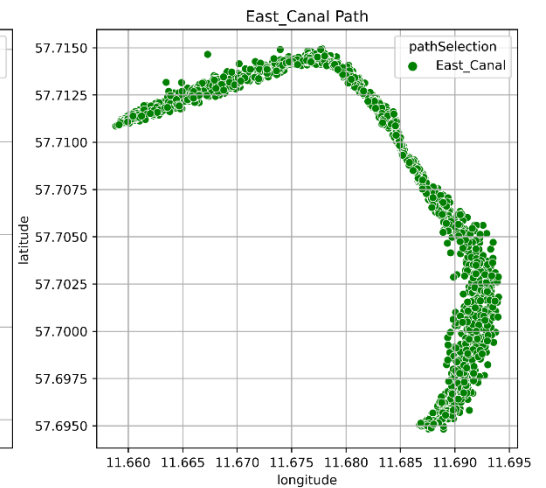
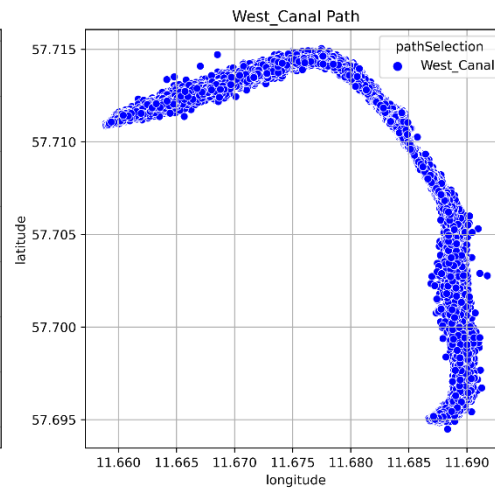
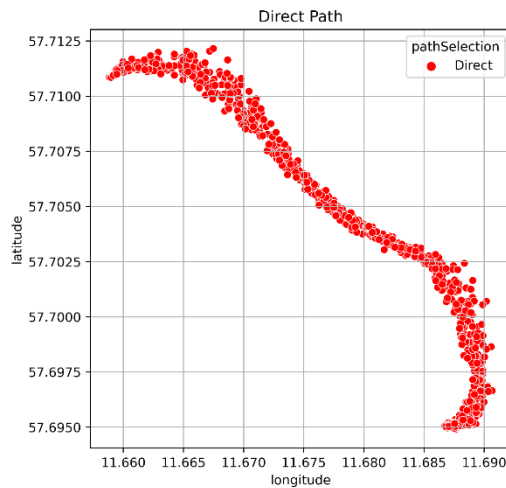
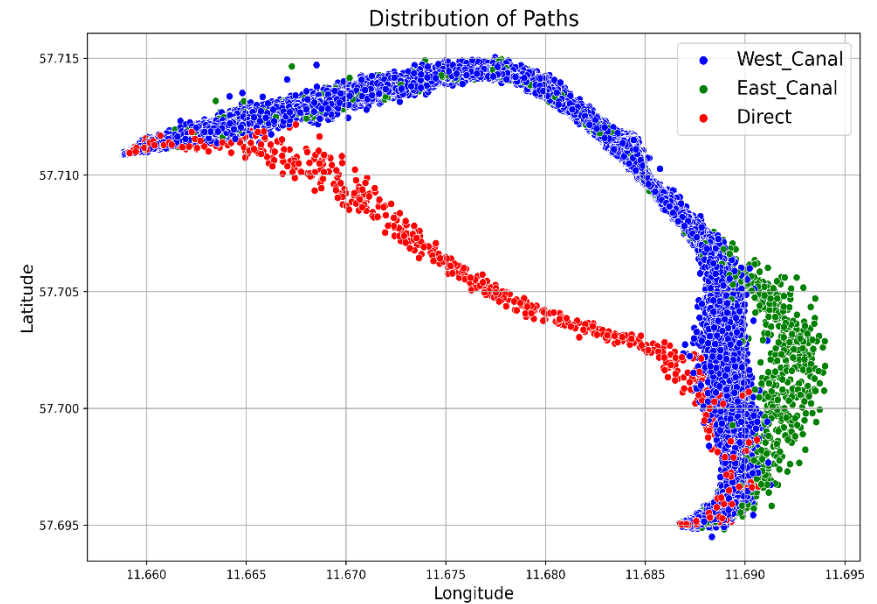
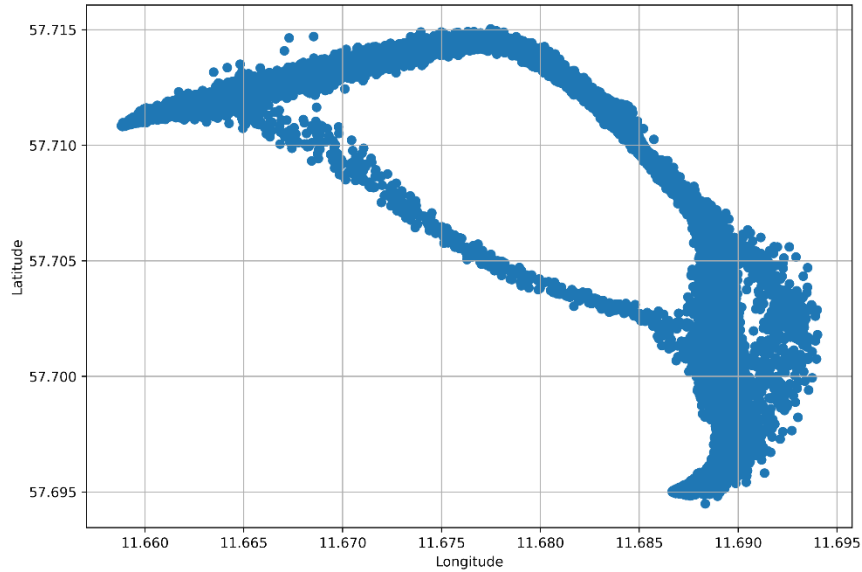


The onboard data have been received from our industry partner CetaSol AB in Gothenburg.

The data has been gathered over 15 months, 2020 and 2021

39193 Datapoints
1754 Trips





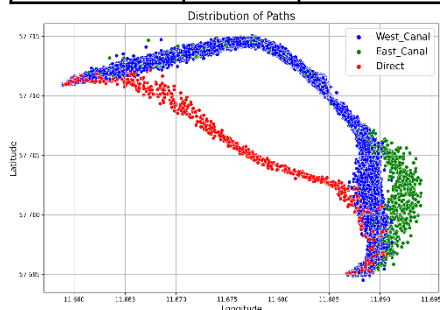
Using GMM

Confusion matrix:

	Direct	East_Canal	West_Canal
Direct	62	0	0
East_Canal	0	122	0
West_Canal	0	11	1560

Confusion matrix:

	Direct	East_Canal	West_Canal
Direct	62	0	0
East_Canal	0	122	0
West_Canal	0	0	1571



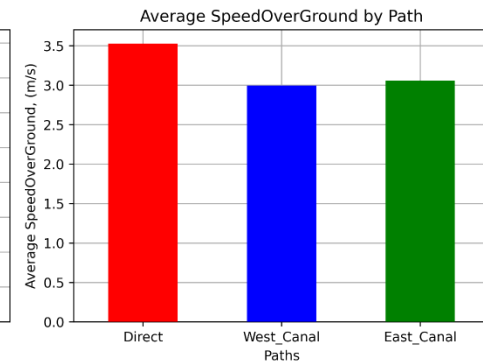
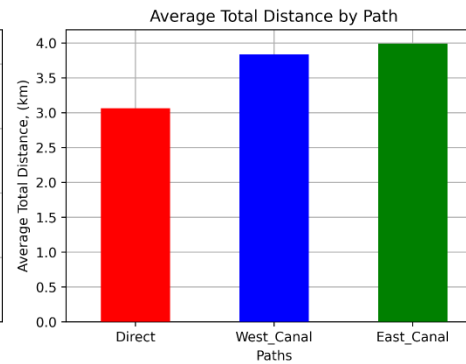
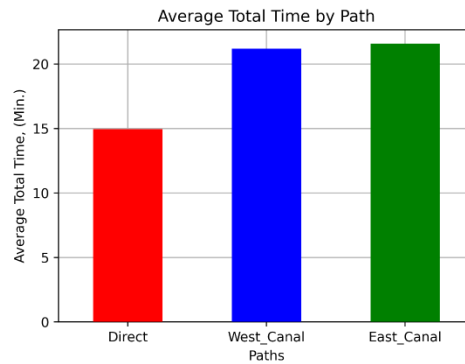
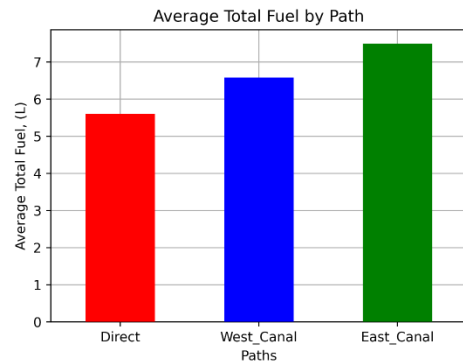
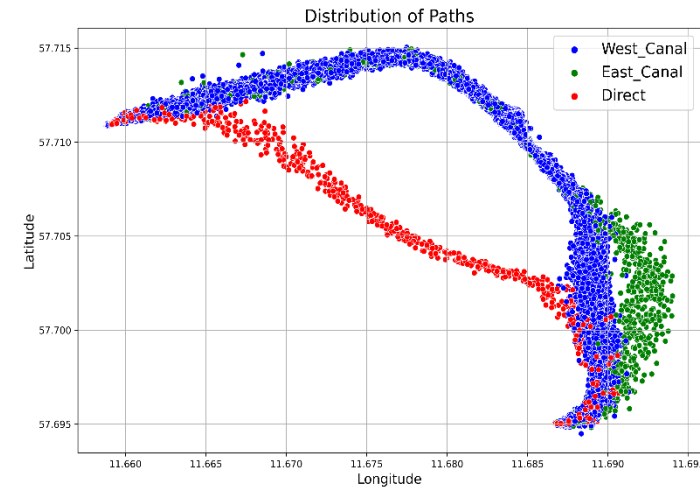
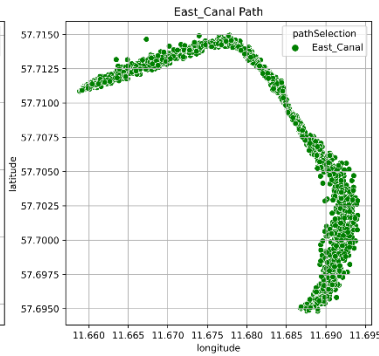
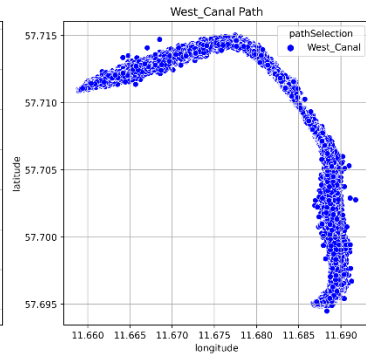
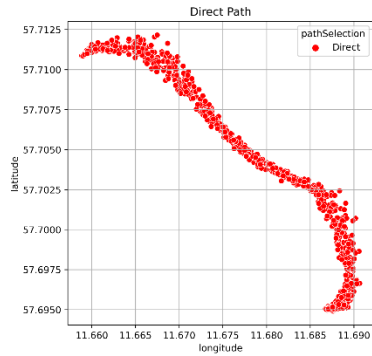
To Three Clusters by Distance Matrix

Statistics	Precision	Recall	F1-Score	Support
Direct	1.000	1.000	1.000	62.000
East_Canal	0.917	1.000	0.957	122.000
West_Canal	1.000	0.993	0.996	1571.000
accuracy	0.994	0.994	0.994	0.994
macro avg	0.972	0.998	0.984	1755.000
weighted avg	0.994	0.994	0.994	1755.000

To Three Clusters with Gaussian in Six Segments

Statistics	Precision	Recall	F1-Score	Support
Direct	1.0	1.0	1.0	62.0
East_Canal	1.0	1.0	1.0	122.0
West_Canal	1.0	1.0	1.0	1571.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	1755.0
weighted avg	1.0	1.0	1.0	1755.0

39193 Datapoints
1754 sequences (Routes)



Cinderella II

MarineTraffic: Global Ship Tracking Intelligence | AIS Marine Traffic

Vessel Type:

Passenger Ship

Length × Breadth:

41.76x7.68

MMSI: **265609540**

Gross Tonnage: **324**

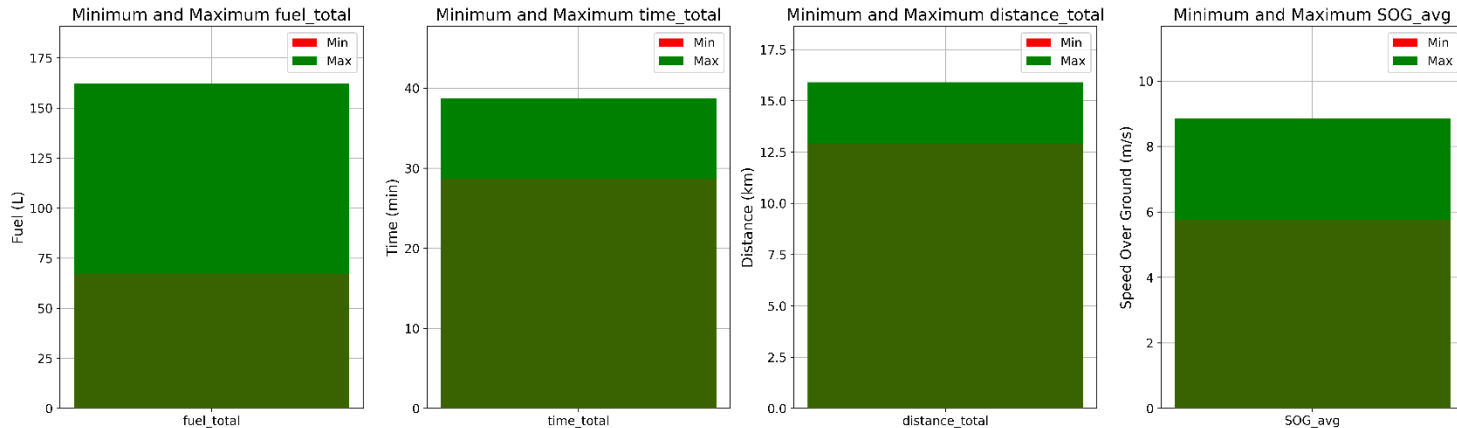
Speed recorded (Max / Average): **20.2 knots / 10.4 knots**

Draught (Reported/Max): **1.2 m**

Flag: **Sweden (SE)**

The onboard data have been received from our industry partner CetaSol AB in Gothenburg.

The data has been gathered over 5 months, 2022



243688 Datapoints
124 Trips

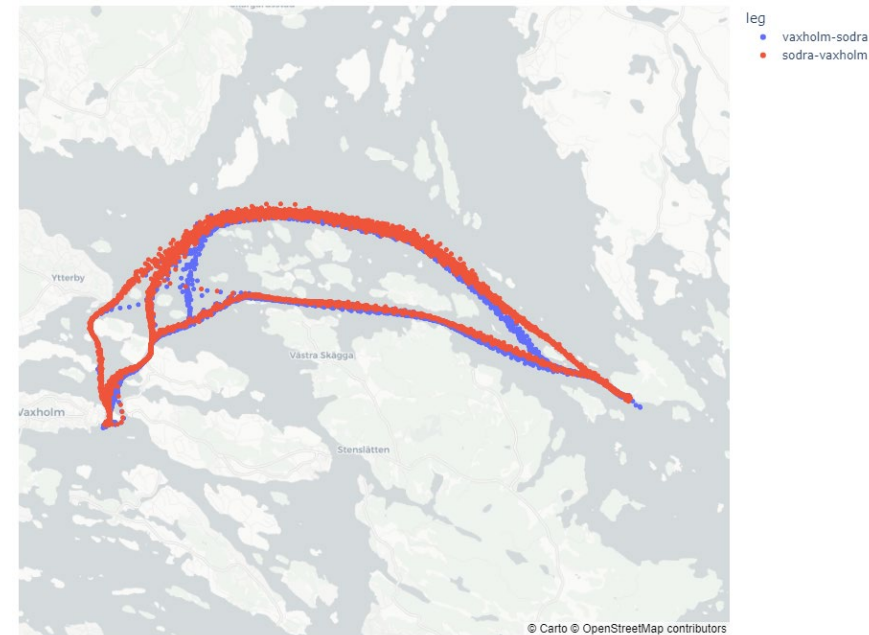
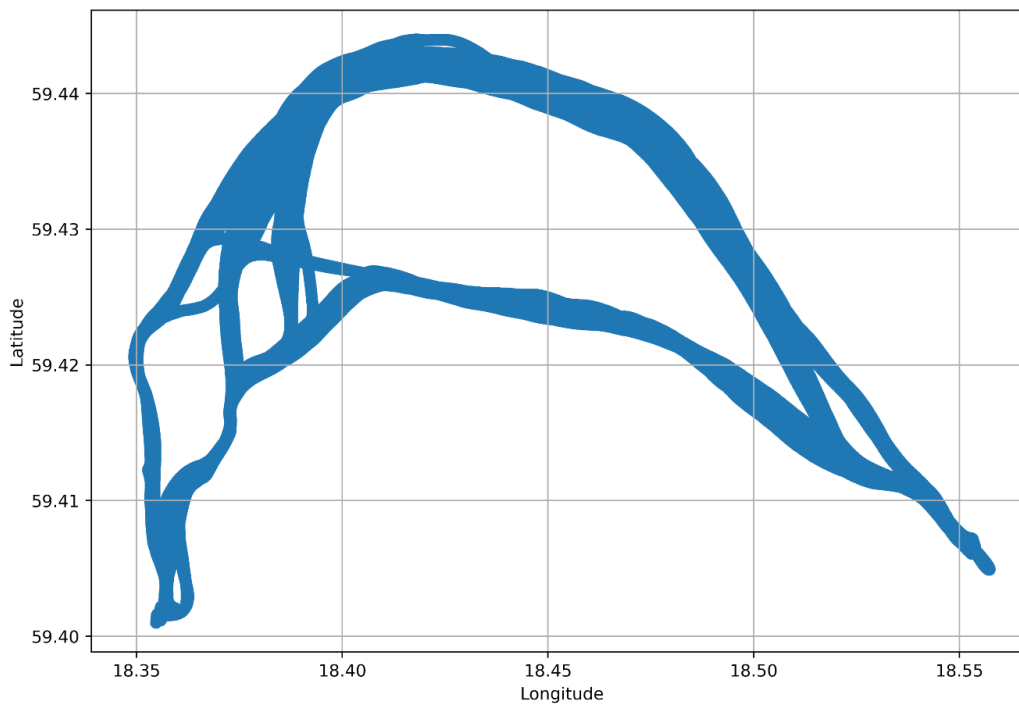
Vessel name: Cinderella II

Ports are Vaxholm and Sodra (East of Stockholm, Baltic Sea)

From July 1 to November 6th, 2022.

Temporal Resolution 1second

243688 Datapoints
124 Trips



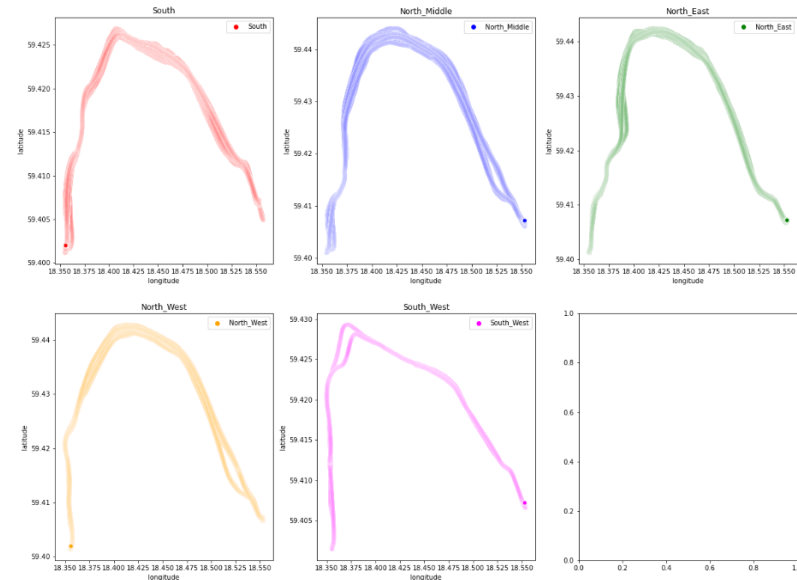
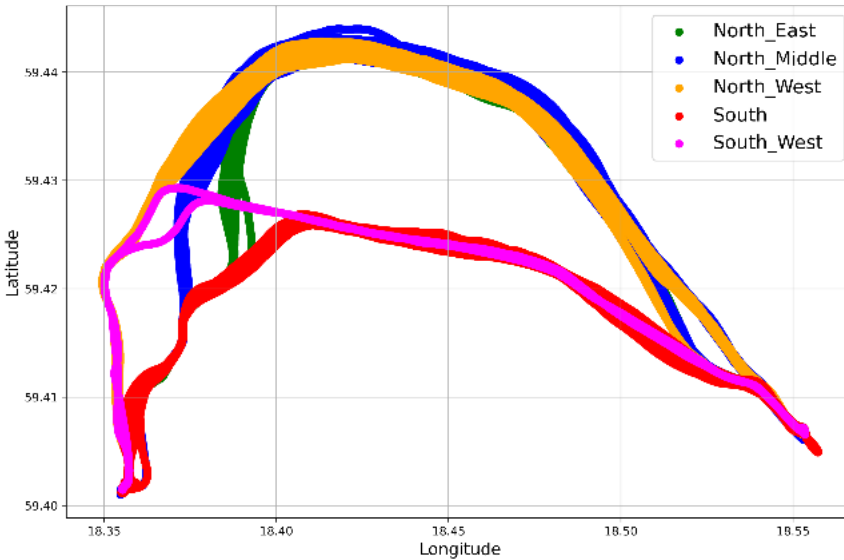
Cinderella_cluster_Weathers_Preprocessing_v1

Clustering of Ship Paths

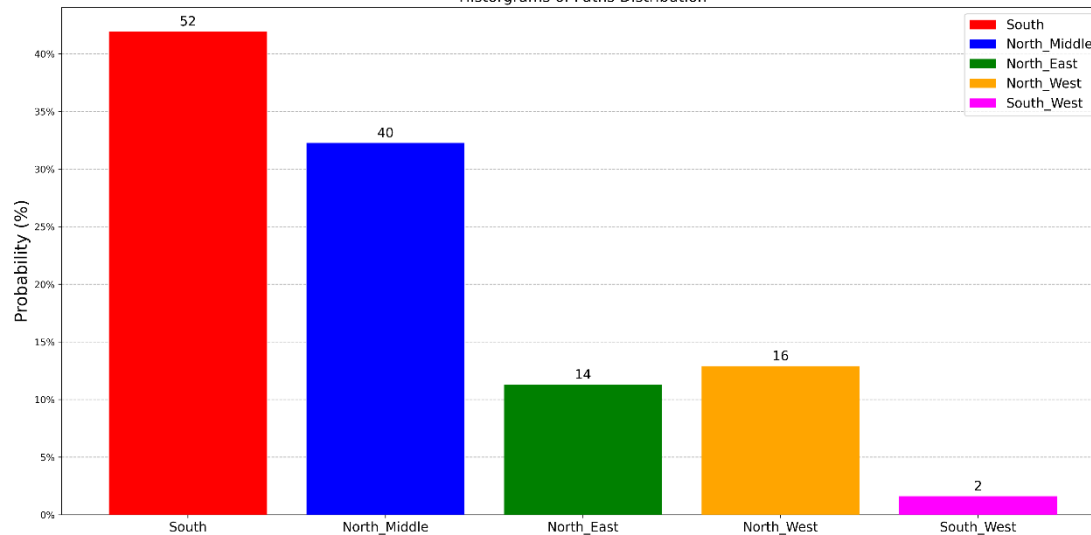
Cinderella

*Cluster ship paths
with a new data*

Paths Distribution

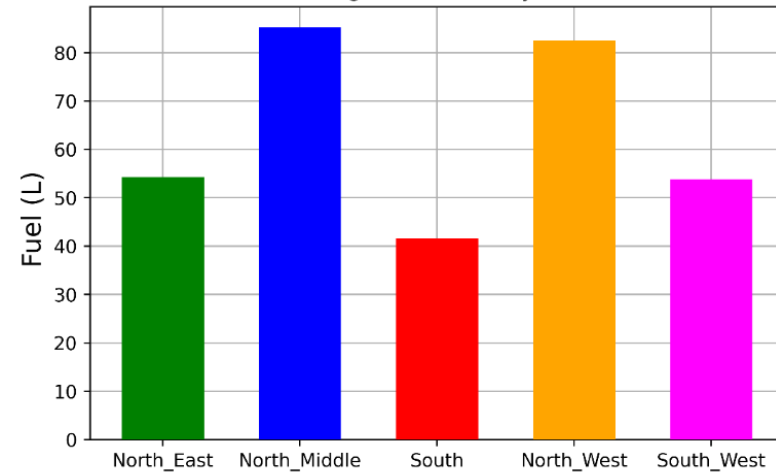


Histograms of Paths Distribution

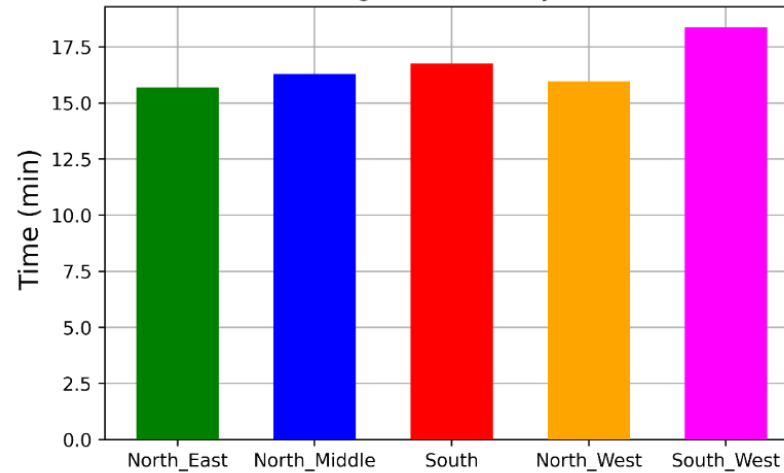


Statistics of Path Segments

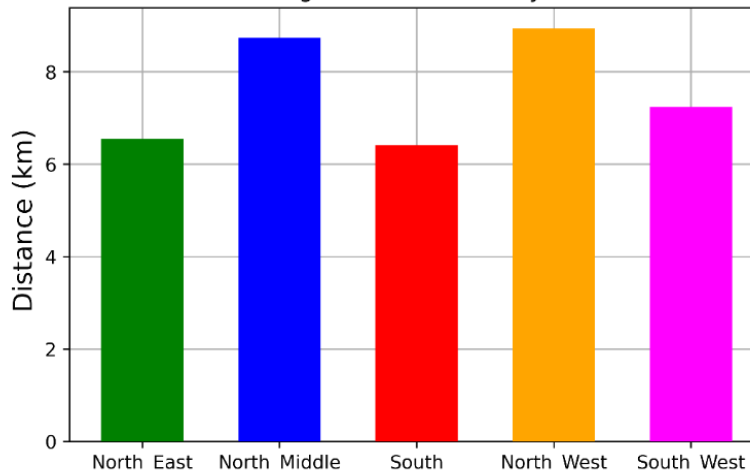
Average Total Fuel by Path



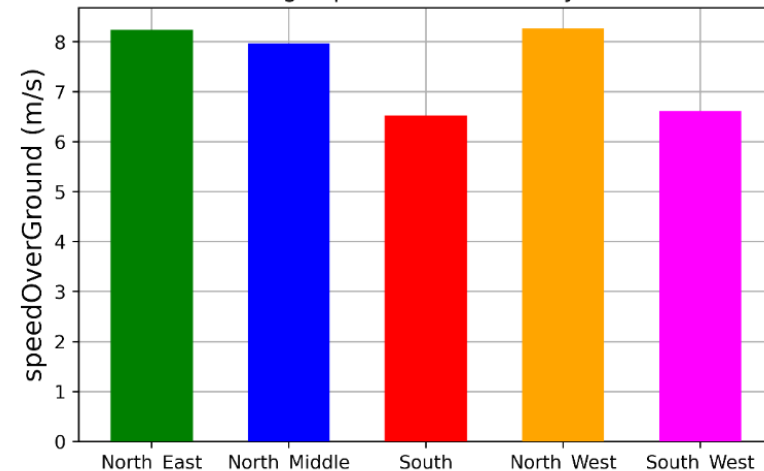
Average Total Time by Path



Average Total Distance by Path



Average speedOverGround by Path

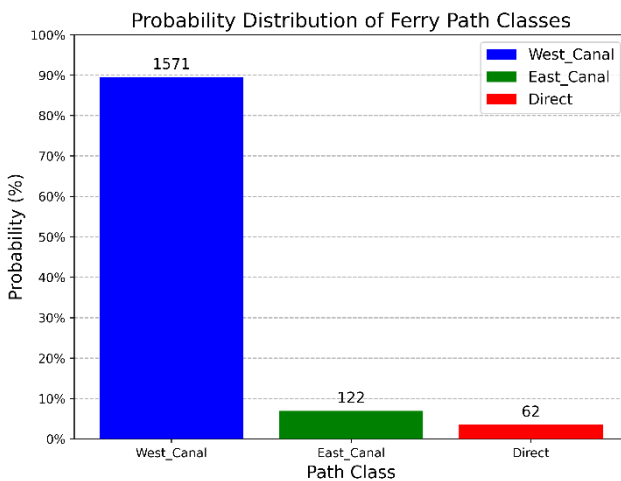
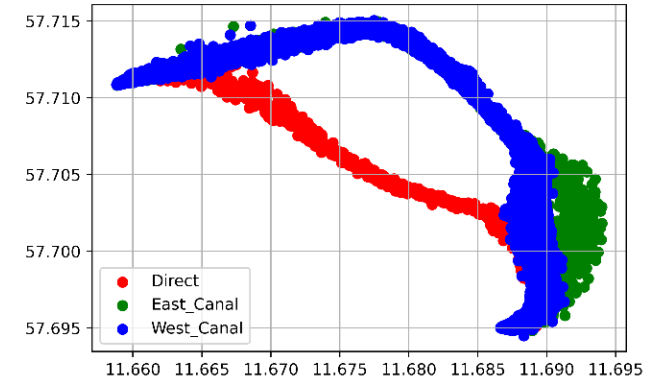
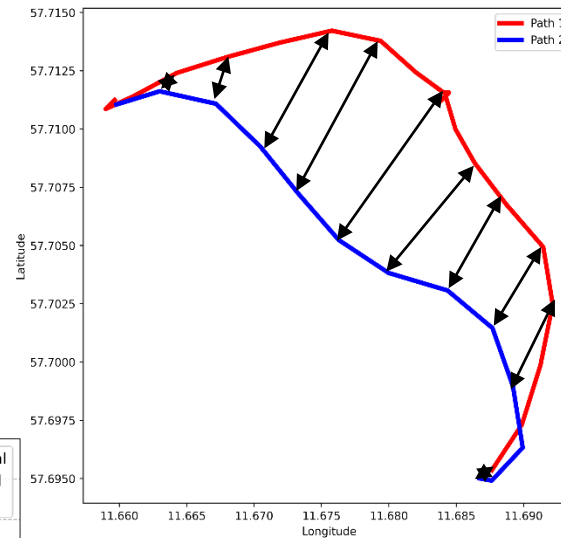
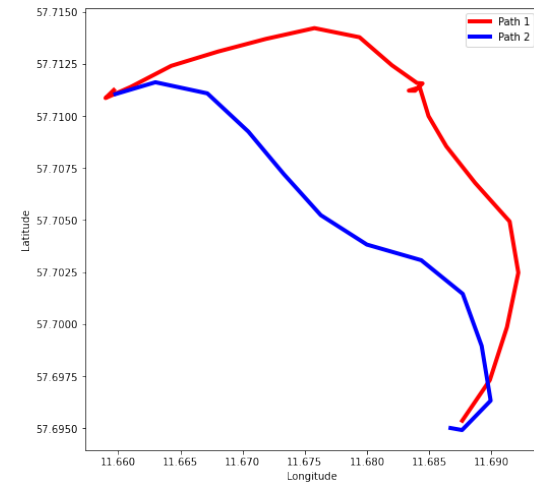


Clustering of Ship Paths

Buro's Data

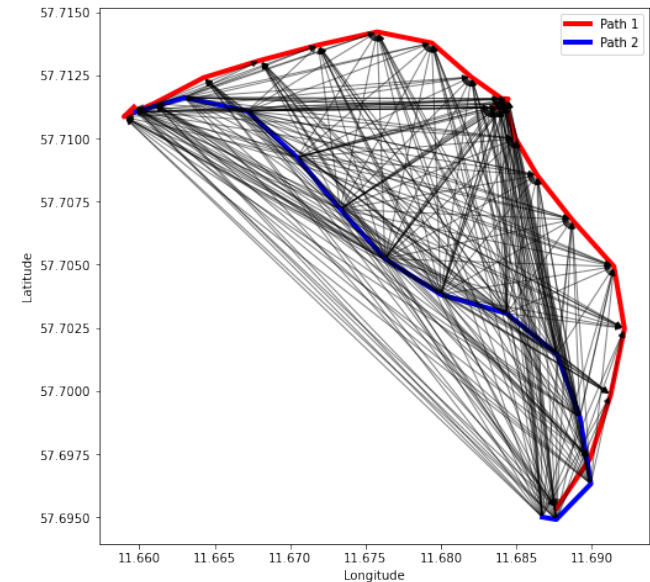
Clustering
by Distance Matrix

Buro's Data



$$\text{Distance} = \sum (P_i - \text{NN}(P_j))$$

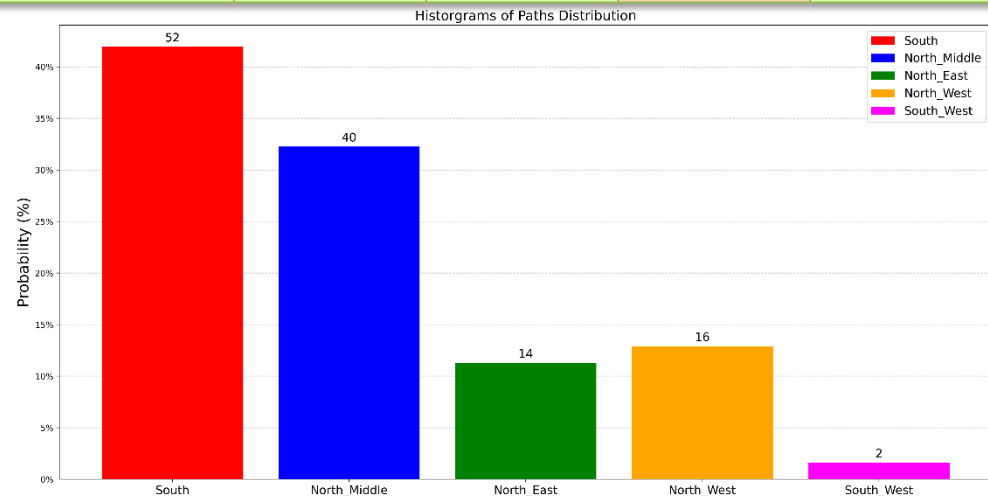
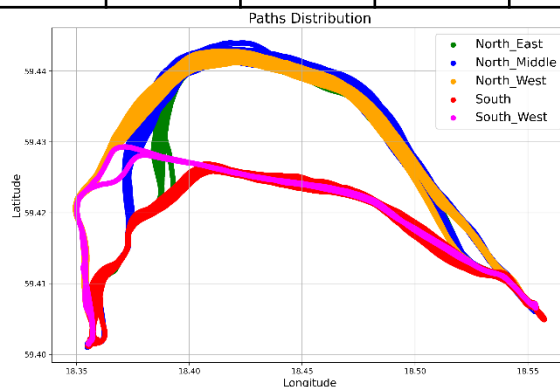
Path Class	West_Canal	East_Canal	Direct	West_Canal	East_Canal	Direct	West_Canal	West_Canal	West_Canal	East_Canal	West_Canal	Direct
West_Canal	0.0000	0.0034	0.0094	0.0004	0.0200	0.0083	0.0002	0.0134	0.0093	0.0207	0.0125	0.0063
East_Canal	0.0226	0.0000	0.0089	0.0171	0.0160	0.0479	0.0160	0.0173	0.0195	0.0180	0.0249	0.0064
Direct	0.0422	0.0458	0.0000	0.0423	0.0470	0.0097	0.0420	0.0389	0.0442	0.0422	0.0411	0.0117
West_Canal	0.0096	0.0158	0.1208	0.0000	0.0217	0.1199	0.0097	0.0096	0.0099	0.0214	0.0189	0.1180
East_Canal	0.0222	0.0141	0.1247	0.0228	0.0000	0.1233	0.0203	0.0239	0.0218	0.0136	0.0196	0.1226
Direct	0.0586	0.0421	0.0087	0.0570	0.0415	0.0000	0.0393	0.0376	0.0379	0.0382	0.0363	0.0100
West_Canal	0.0503	0.0163	0.1149	0.0096	0.0206	0.1146	0.0000	0.0091	0.0117	0.0210	0.0173	0.1122
West_Canal	0.0135	0.0164	0.1178	0.0087	0.0236	0.1195	0.0103	0.0000	0.0121	0.0202	0.0181	0.1154
West_Canal	0.0104	0.0154	0.1124	0.0092	0.0185	0.1101	0.0114	0.0112	0.0000	0.0177	0.0124	0.1079
East_Canal	0.0363	0.0189	0.1078	0.0345	0.0199	0.1047	0.0341	0.0138	0.0128	0.0000	0.0307	0.1030
West_Canal	0.0128	0.0218	0.1181	0.0177	0.0184	0.1181	0.0183	0.0180	0.0133	0.0172	0.0000	0.1183
Direct	0.0448	0.0488	0.0149	0.0443	0.0516	0.0143	0.0450	0.0417	0.0439	0.0446	0.0448	0.0000



K-Means

Confusion Matrix	Pred. North_East	Pred. North_Middle	Pred. North_West	Pred. South	Pred. South_West
North_East	14	0	0	0	0
North_Middle	6	34	0	0	0
North_West	0	0	16	0	0
South	0	0	0	52	0
South_West	0	0	0	0	2

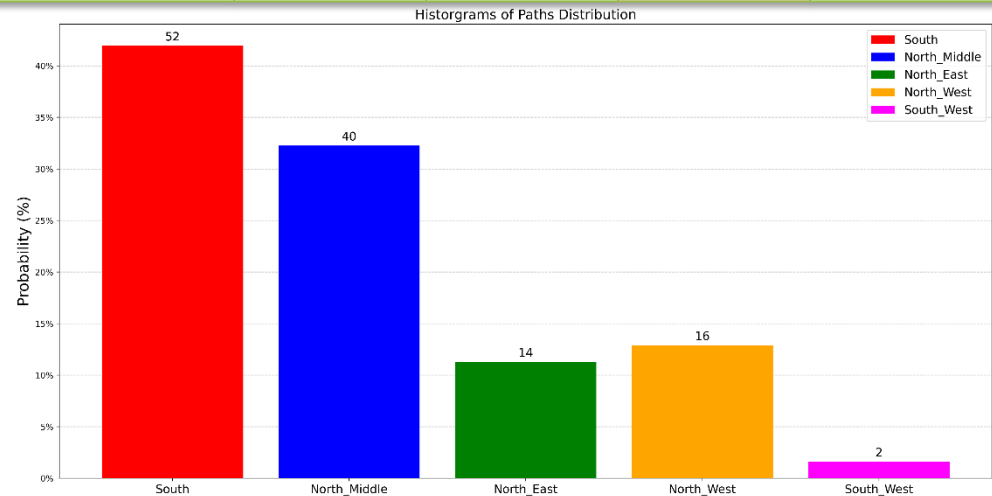
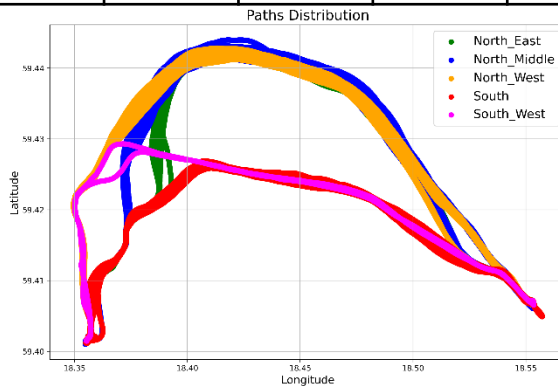
	Precision	Recall	F1-score	Support
North_East	0.700	1.000	0.824	14.000
North_Middle	1.000	0.850	0.919	40.000
North_West	1.000	1.000	1.000	16.000
South	1.000	1.000	1.000	52.000
South_West	1.000	1.000	1.000	2.000
accuracy	0.952	0.952	0.952	0.952
macro avg	0.940	0.970	0.948	124.000
weighted avg	0.966	0.952	0.954	124.000



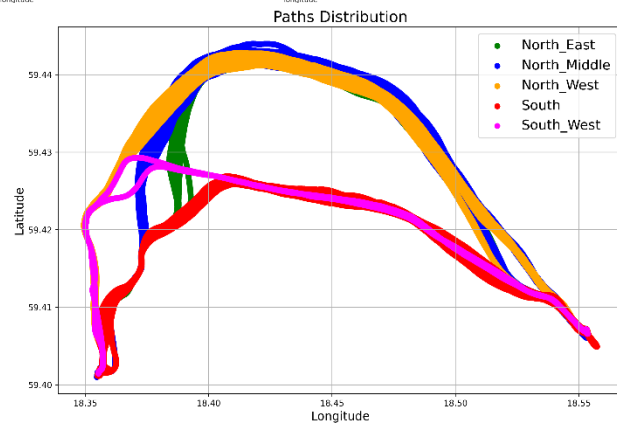
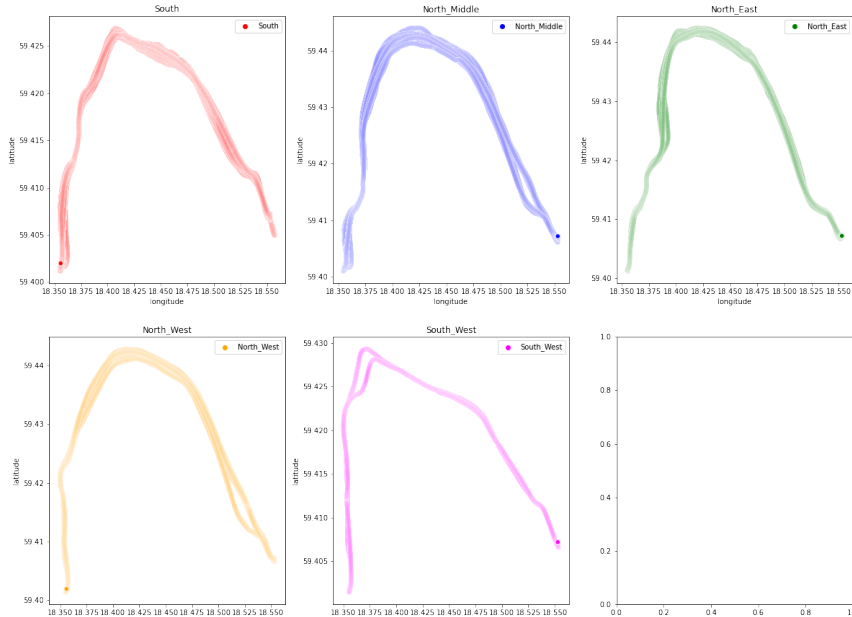
GMM

Confusion Matrix	Pred. North_East	Pred. North_Middle	Pred. North_West	Pred. South	Pred. South_West
North_East	14	0	0	0	0
North_Middle	6	34	0	0	0
North_West	0	0	16	0	0
South	0	0	0	52	0
South_West	0	0	0	0	2

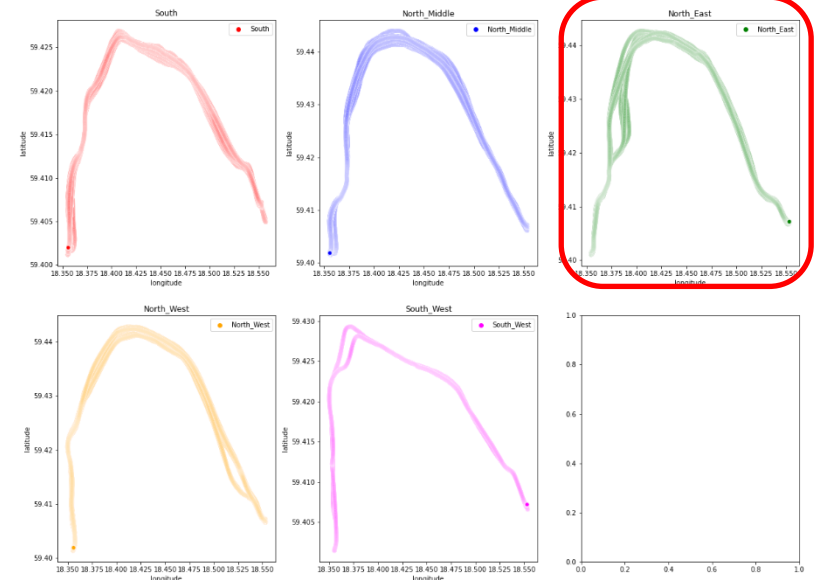
	Precision	Recall	F1-score	Support
North_East	0.700	1.000	0.824	14.000
North_Middle	1.000	0.850	0.919	40.000
North_West	1.000	1.000	1.000	16.000
South	1.000	1.000	1.000	52.000
South_West	1.000	1.000	1.000	2.000
accuracy	0.952	0.952	0.952	0.952
macro avg	0.940	0.970	0.948	124.000
weighted avg	0.966	0.952	0.954	124.000



Actual Paths

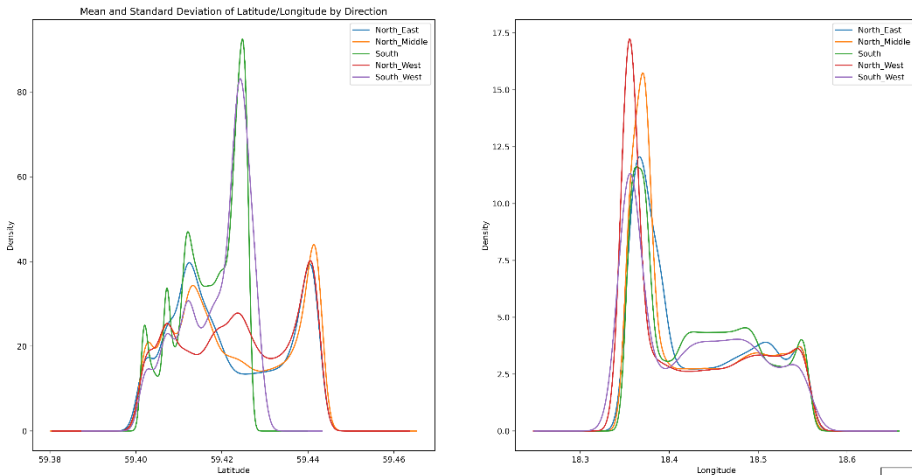


Predicted Paths from (K-means and GMM)

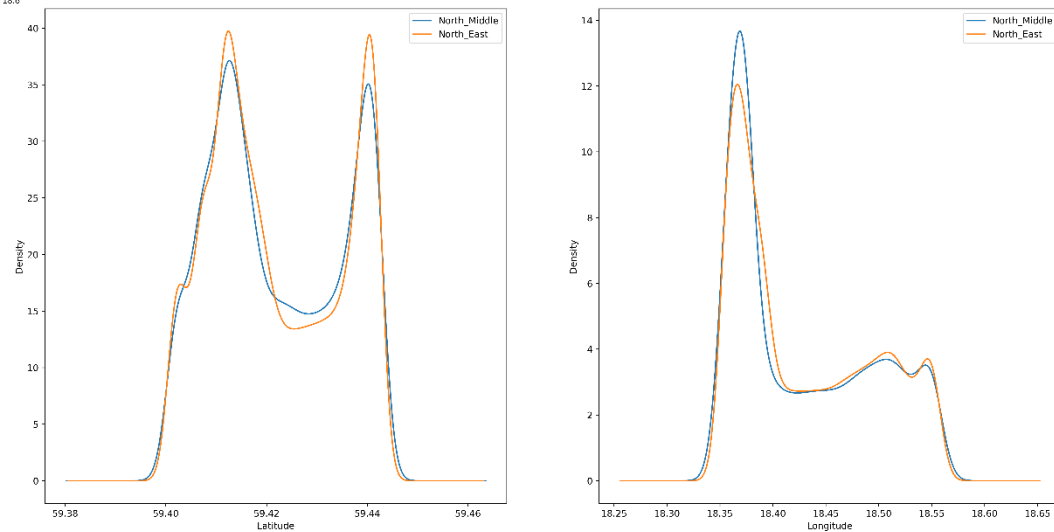


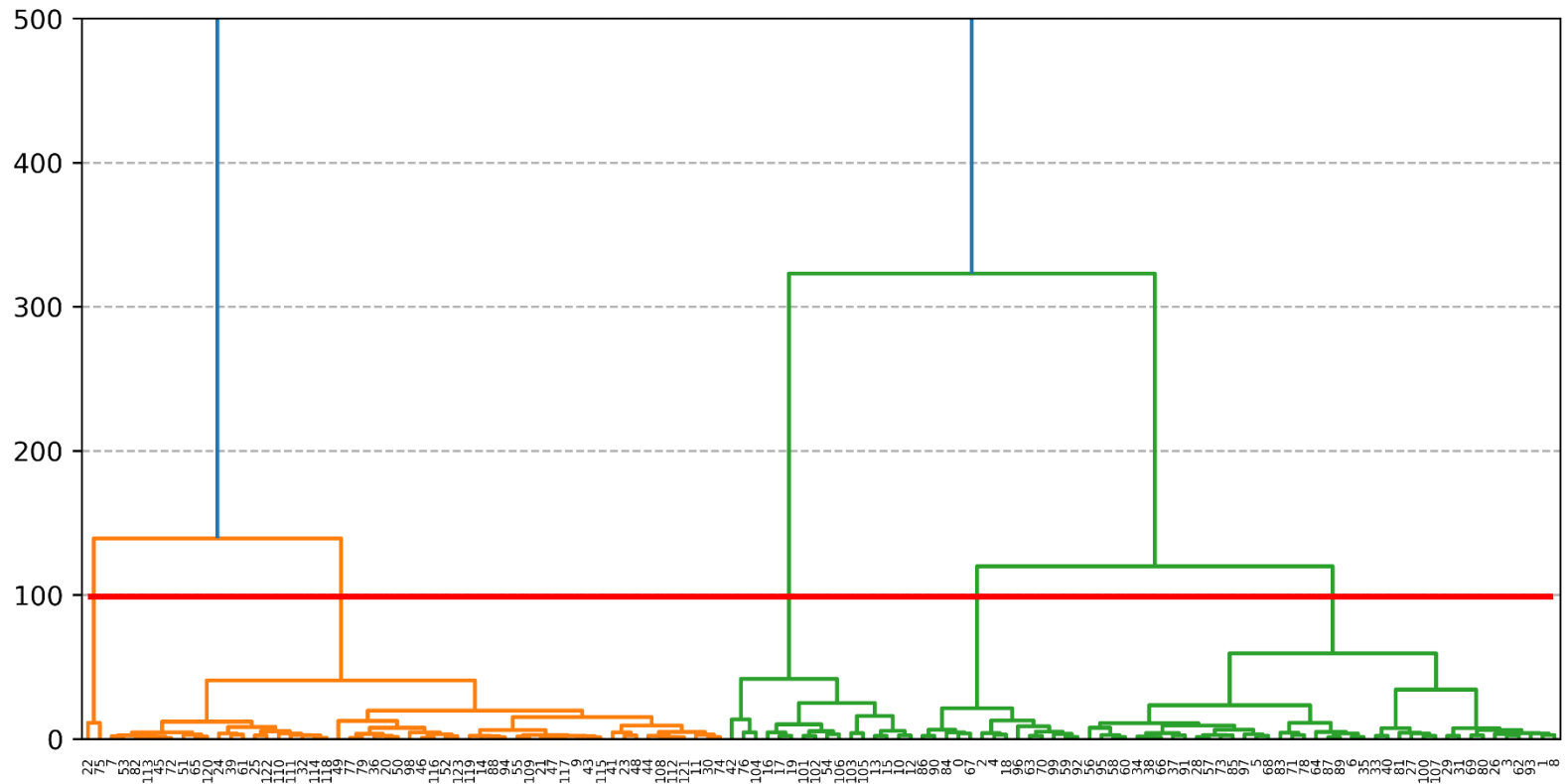
K-Means and GMM

PDF for correct-clustered paths



PDF for misclustered paths

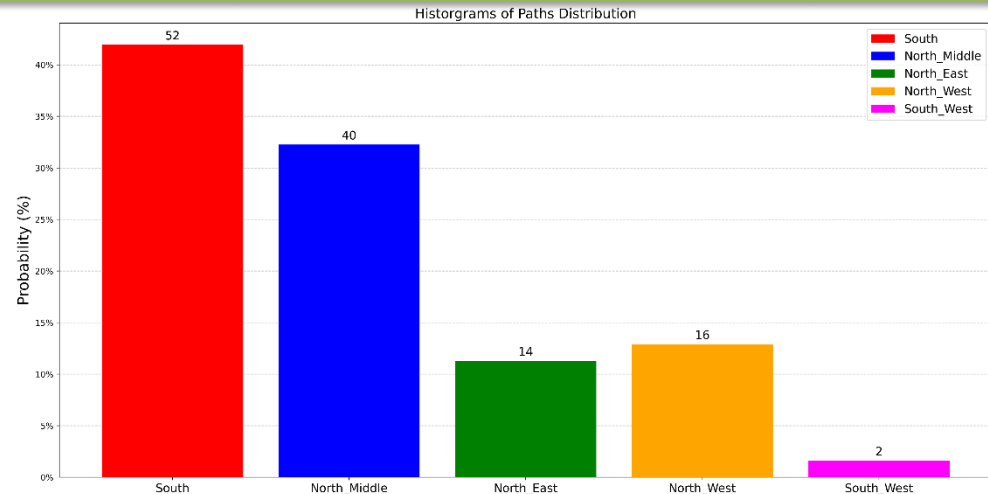
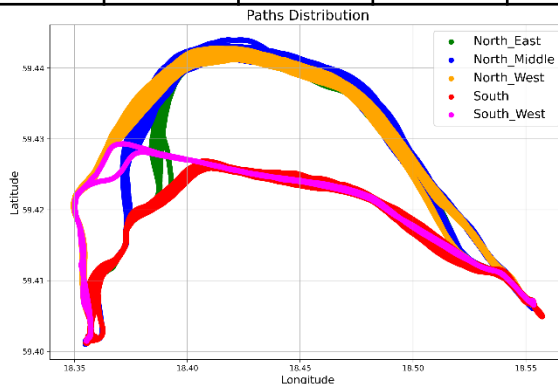


Hierarchical Clustering

Hierarchical Clustering

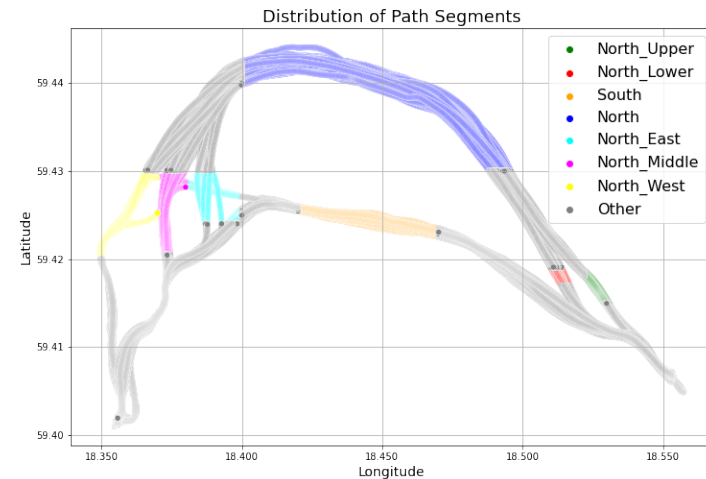
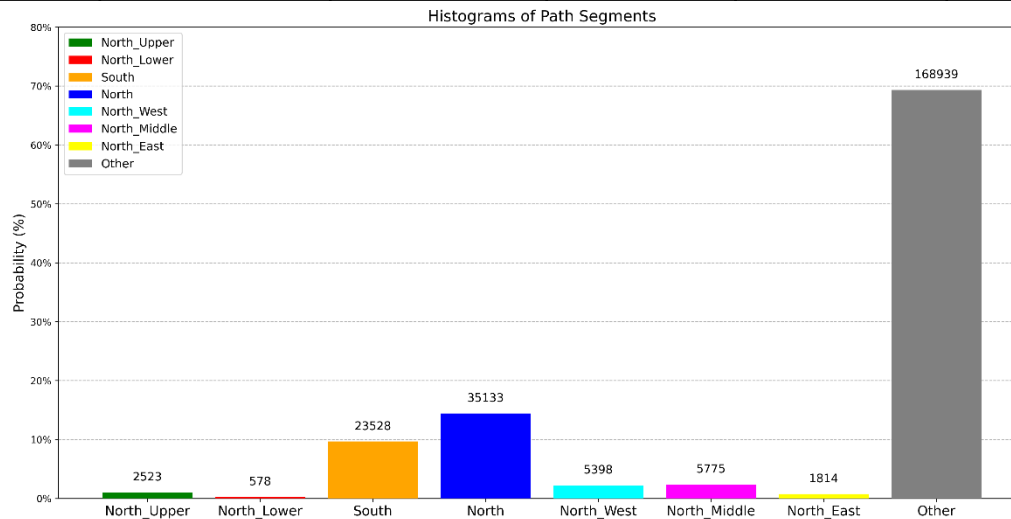
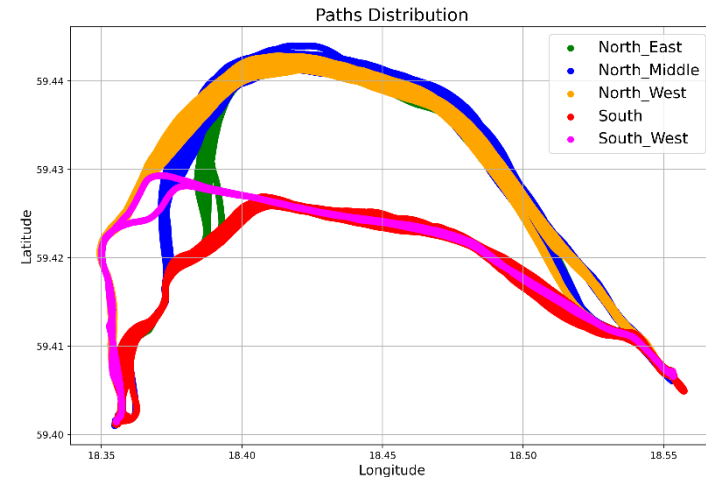
Confusion Matrix	Pred. North_East	Pred. North_Middle	Pred. North_West	Pred. South	Pred. South_West
North_East	14	0	0	0	0
North_Middle	0	40	0	0	0
North_West	0	0	16	0	0
South	0	0	0	52	0
South_West	0	0	0	0	2

	Precision	Recall	F1-score	Support
North_East	1.0	1.0	1.0	14.0
North_Middle	1.0	1.0	1.0	40.0
North_West	1.0	1.0	1.0	16.0
South	1.0	1.0	1.0	52.0
South_West	1.0	1.0	1.0	2.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	124.0
weighted avg	1.0	1.0	1.0	124.0

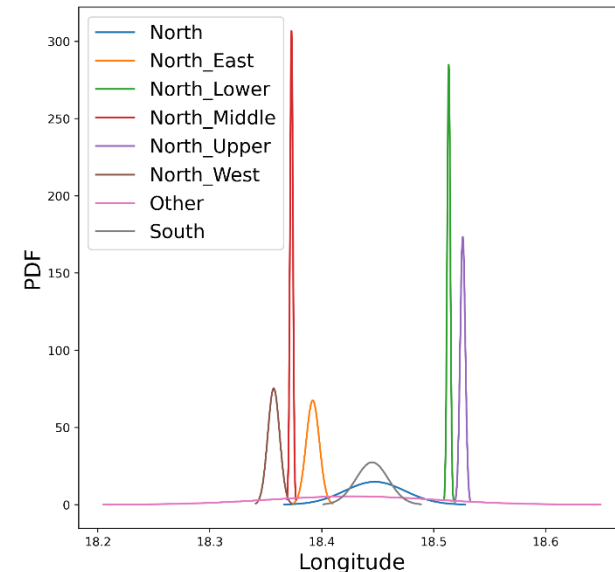
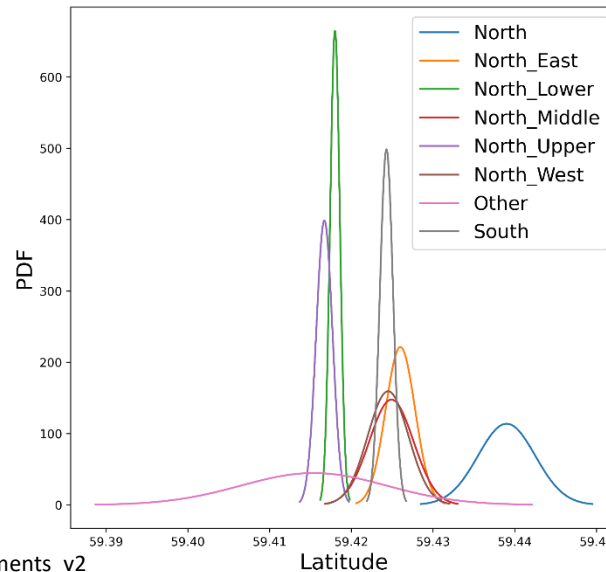
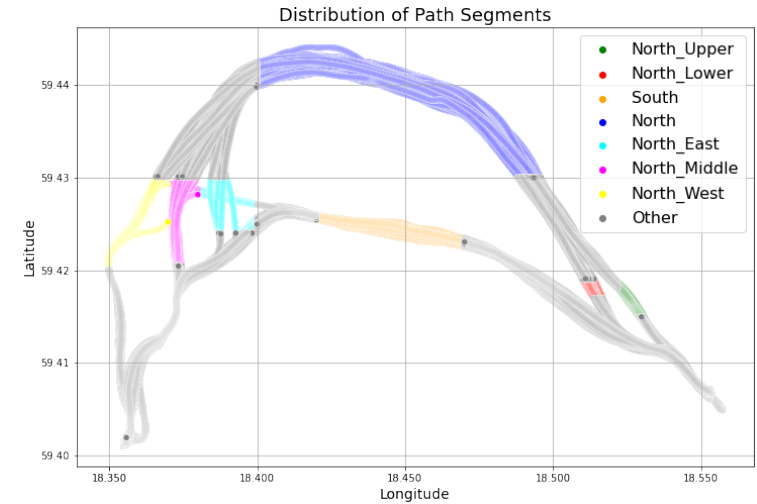
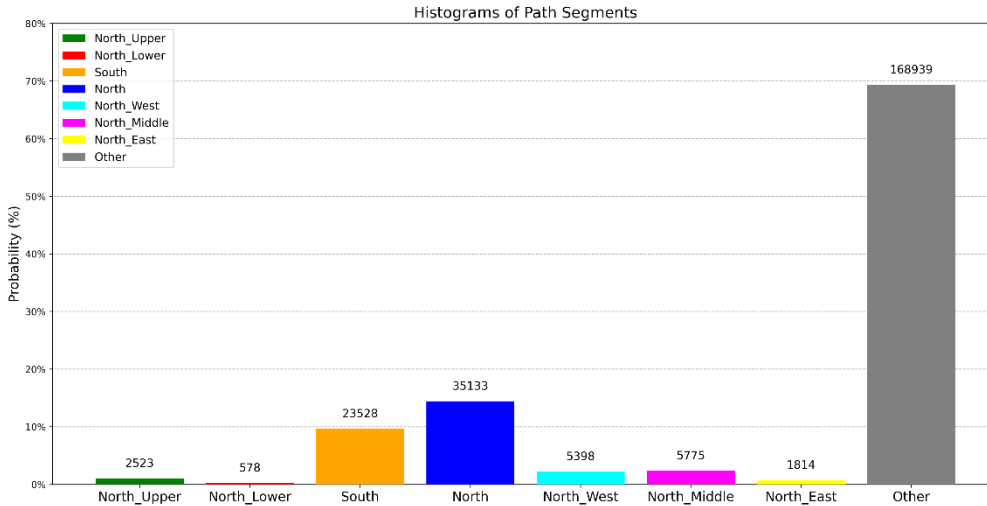


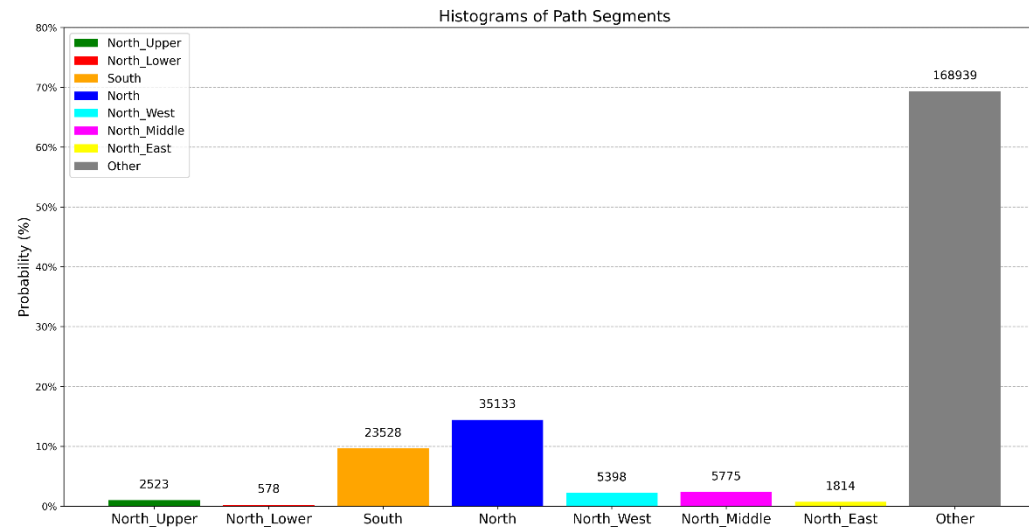
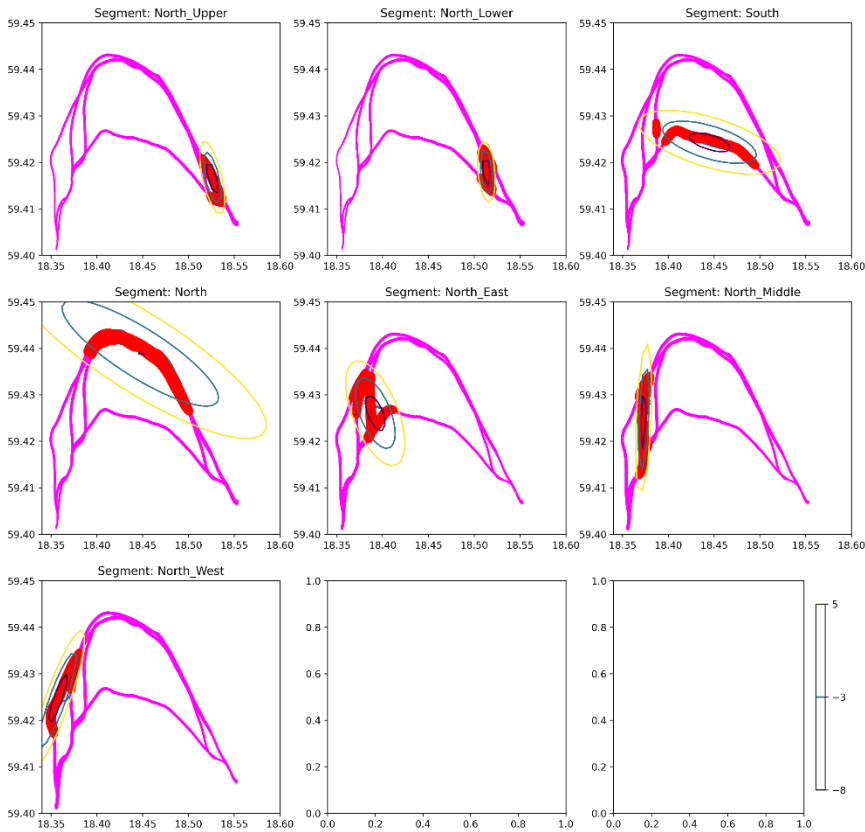
Statistics of Path Segments

index	segment	Latitude _mean	Latitude _std	Longitude _mean	Longitude _std
1	North	59.4390	0.0035	18.4474	0.0269
2	North_East	59.4260	0.0018	18.3920	0.0059
3	North_Lower	59.4180	0.0006	18.5135	0.0014
4	North_Middle	59.4249	0.0027	18.3730	0.0013
5	North_Upper	59.4167	0.0010	18.5260	0.0023
6	North_West	59.4245	0.0025	18.3571	0.0053
7	South	59.4243	0.0008	18.4450	0.0145
8	Other	59.4154	0.0089	18.4270	0.0740



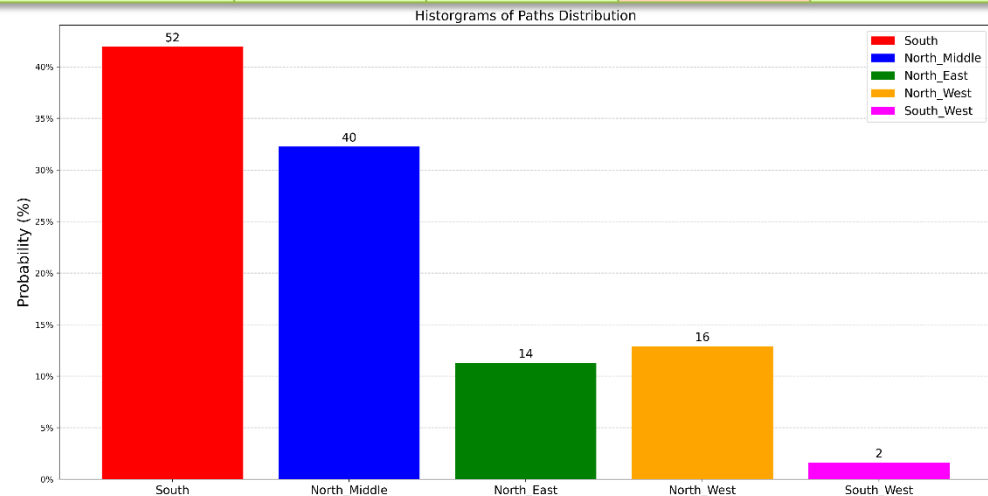
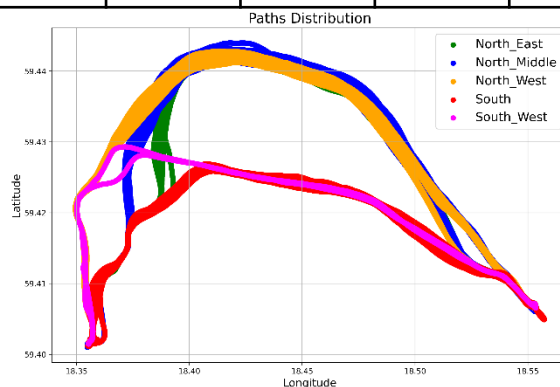
Statistics of the Clustering Results





Confusion Matrix	Pred. North_East	Pred. North_Middle	Pred. North_West	Pred. South	Pred. South_West
North_East	14	0	0	0	0
North_Middle	0	40	0	0	0
North_West	0	0	16	0	0
South	0	0	0	52	0
South_West	0	0	0	0	2

	Precision	Recall	F1-score	Support
North_East	1.0	1.0	1.0	14.0
North_Middle	1.0	1.0	1.0	40.0
North_West	1.0	1.0	1.0	16.0
South	1.0	1.0	1.0	52.0
South_West	1.0	1.0	1.0	2.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	124.0
weighted avg	1.0	1.0	1.0	124.0

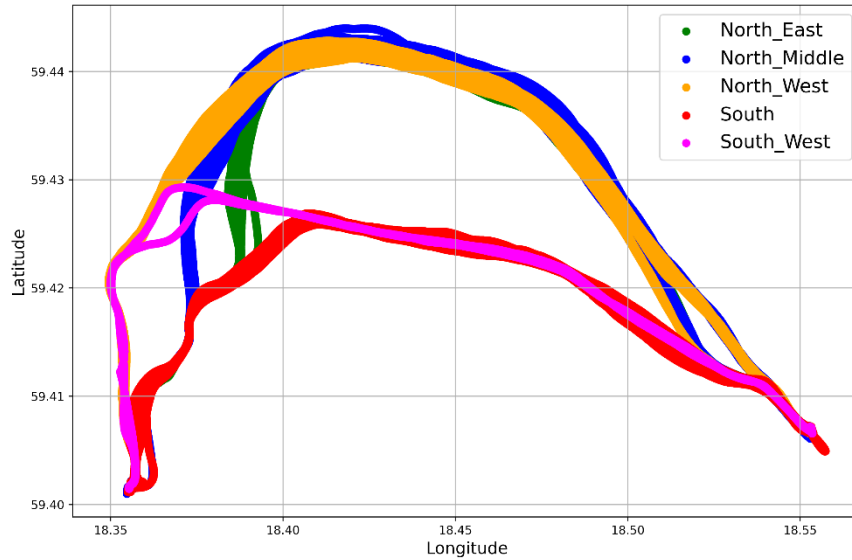


Classification of Ship Paths

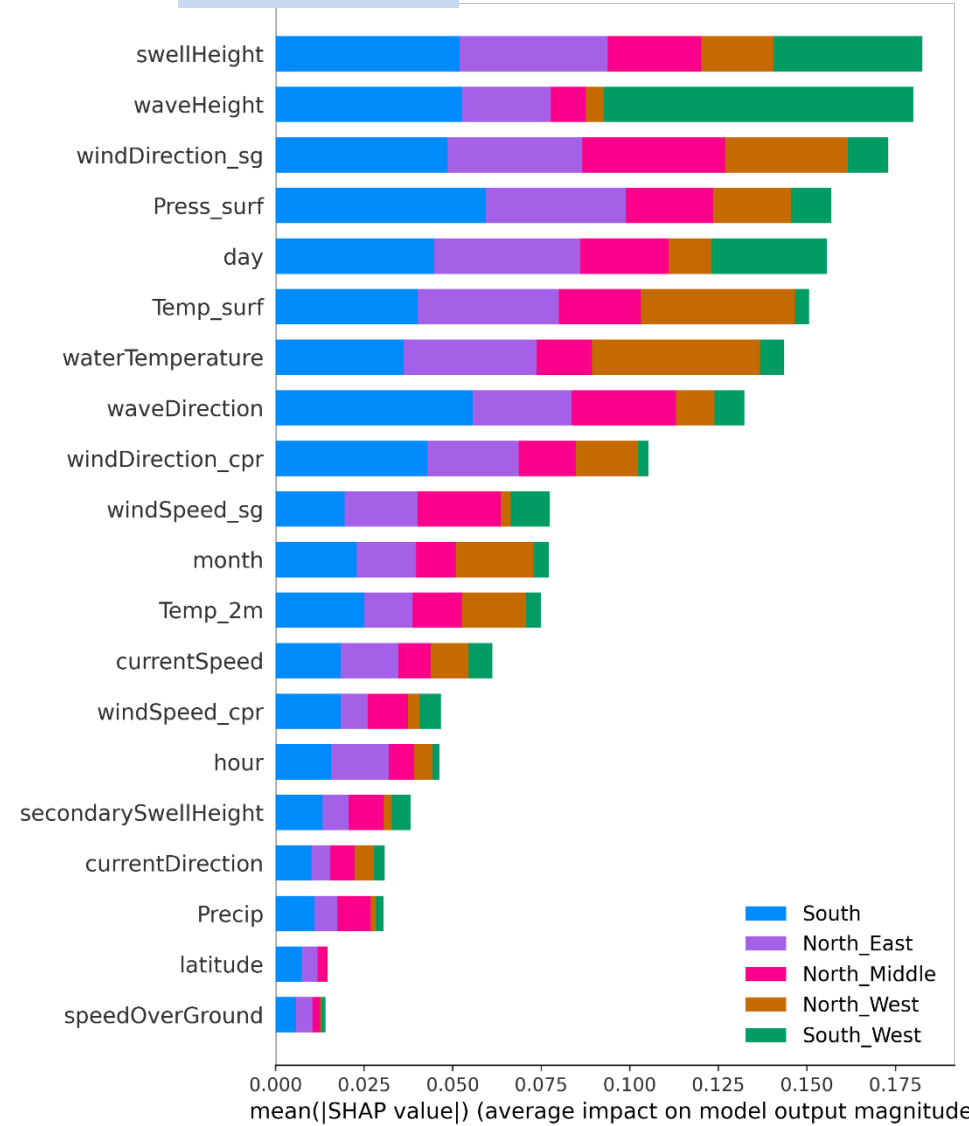
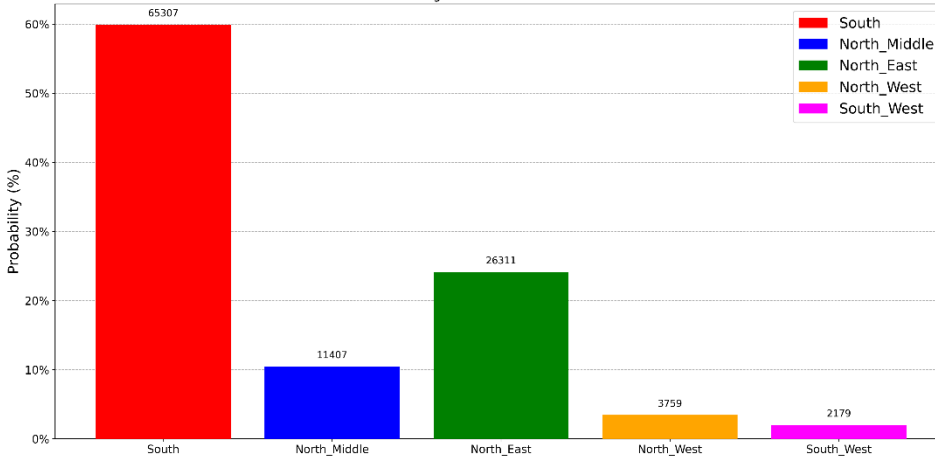
Cinderella

To 5 Classes

Paths Distribution



Histograms of Paths Distribution



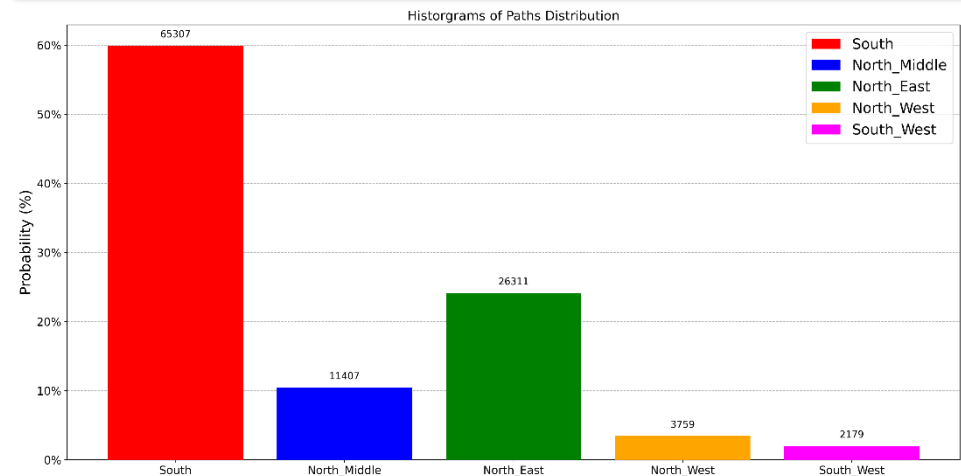
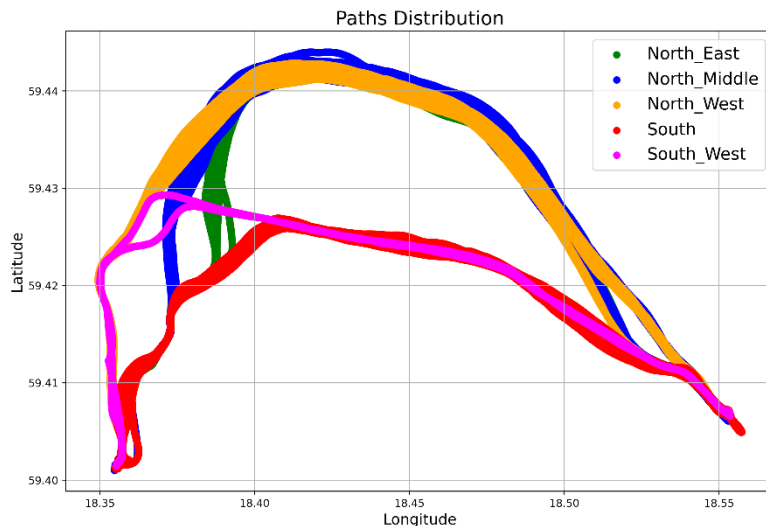
Shap Value

Size(y_test)=21793
out of 108963

Random Random Forest with Balanced Subsampling

Confusion Matrix	Pred. North_East	Pred. North_Middle	Pred. North_West	Pred. South	Pred. South_West
North_East	5263	0	0	0	0
North_Middle	0	2282	0	0	0
North_West	0	0	752	0	0
South	0	0	0	13062	0
South_West	0	0	0	0	436

	Precision	Recall	F1-score	Support
North_East	1.0	1.0	1.0	5263.0
North_Middle	1.0	1.0	1.0	2282.0
North_West	1.0	1.0	1.0	752.0
South	1.0	1.0	1.0	13062.0
South_West	1.0	1.0	1.0	436.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	21795.0
weighted avg	1.0	1.0	1.0	21795.0



Size(y_test)=21793
out of 108963

Random Random Forest with Balanced Subsampling

Confusion Matrix	Pred. North_East	Pred. North_Middle	Pred. North_West	Pred. South	Pred. South_West
North_East	5263	0	0	0	0
North_Middle	0	2282	0	0	0
North_West	0	0	752	0	0
South	0	0	0	13062	0
South_West	0	0	0	0	436

	Precision	Recall	F1-score	Support
North_East	1.0	1.0	1.0	5263.0
North_Middle	1.0	1.0	1.0	2282.0
North_West	1.0	1.0	1.0	752.0
South	1.0	1.0	1.0	13062.0
South_West	1.0	1.0	1.0	436.0
accuracy	1.0	1.0	1.0	1.0
macro avg	1.0	1.0	1.0	21795.0
weighted avg	1.0	1.0	1.0	21795.0

Uncertainty Analysis

Class North_East: Mean=1.0000, StdDev=0.0009, Median=1.0000, Min=0.9600, Max=1.0000
 Class North_Middle: Mean=0.9998, StdDev=0.0064, Median=1.0000, Min=0.7000, Max=1.0000
 Class North_West: Mean=0.9999, StdDev=0.0022, Median=1.0000, Min=0.9400, Max=1.0000
 Class South: Mean=1.0000, StdDev=0.0003, Median=1.0000, Min=0.9900, Max=1.0000
 Class South_West: Mean=1.0000, StdDev=0.0000, Median=1.0000, Min=1.0000, Max=1.0000

Thanks for your listening
Any Qs?