Spatial Analysis of Maritime Traffic for Maritime Security

Ronald Pelot, Ph.D., P.Eng.

Research assistants
Dong Lin, Ph.D.       Yan Wu, Ph.D.
Casey Hilliard       David Wootton
1. REPORT DATE  
OCT 2009

2. REPORT TYPE

3. DATES COVERED
00-00-2009 to 00-00-2009

4. TITLE AND SUBTITLE
Spatial Analysis of Maritime Traffic for Maritime Security

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. Author(s)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dalhousie University, Networks of Centres of Excellence (NSERC), Halifax, Nova Scotia, Canada B3H 4R2,

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Maritime Domain Awareness and Counter Piracy, 26-29 October 2009, Ottawa, Canada. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:
   a. REPORT unclassified
   b. ABSTRACT unclassified
   c. THIS PAGE unclassified

17. LIMITATION OF ABSTRACT
   Same as Report (SAR)

18. NUMBER OF PAGES 58

19a. NAME OF RESPONSIBLE PERSON

Standard Form 298 (Rev. 8-98)  
Prepared by ANSI Std Z39-18
Traffic Modelling
### Activity Types

<table>
<thead>
<tr>
<th>Activity Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial shipping</td>
</tr>
<tr>
<td>Commercial fishing</td>
</tr>
<tr>
<td>Commercial recreational</td>
</tr>
<tr>
<td>Aquaculture</td>
</tr>
<tr>
<td>Ferries</td>
</tr>
<tr>
<td>Cruise ships</td>
</tr>
<tr>
<td>Private recreational</td>
</tr>
<tr>
<td>Class</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>
Maximize use of information:
- source: location, date, time
- intermediate points: way points, fishing efforts
- final destination: location, date, time

Extrapolate using more general information
- feasible range: distance, time
- feasible location; eg. fishing grounds; tourist sites
- typical locations: cluster analysis
Trip Waypoints
Direct Connection Results
Node Network
Land Avoidance
Dispersion Algorithm
Trips to feasible region
Grid the area
Define feasible destinations

Nova Scotia

New Brunswick
Generate tracks

Nova Scotia

New Brunswick

Nova Scotia
Abnormal ship trajectories
Objectives

(1) To mathematically define commonly used shipping routes from historical shipping data.

(2) To detect unusual vessel trajectories based on norms of typical shipping routes using Statistical Process Control (SPC).
Line pattern analysis in this research involves the description of the relationship among different trajectories and the classification of similar trajectories into routes and paths.

**Trajectory**: a line generated from a series of waypoints by an individual moving object.

**Route**: a frequently used track followed by moving objects from the time when they enter the study area until the time they exit.

**Path**: a feature grouping common sections of routes.
Traffic Monitoring System

Vessel movement report (Radar/AIS)

Traffic Monitoring System

Alarm?

Anomaly Detection

Route Formation

Major Routes

Prepared Data

Current Trajectories

Data preparation

Keep only the "normal" trajectories

Anomalous Data
Three main steps:

- Route-trajectory matching;
- Route-trajectory updating;
- Route-route merging.

Each route node here is characterized by a vector format with coordinates $\vec{x}_i = [x_i, y_i]$.

Attributes of each route: central axis, left and right boundaries (also called “envelopes”), normal vector and weight.

Resample distance – equally spaced nodes are obtained on each trajectory and route.
Label each trajectory with a unique identifier and resample it. The first trajectory is initialized as the first existing route in the route database.
A second trajectory is taken from the database, resampled, and Hausdorff distance is calculated between the trajectory (T) and all the existing routes (R).
Compare the distances $d_i(TR)$ from the trajectory to all the routes in the route database and get the minimum distance. The index of this route is written as $I$.

$$d_I(TR) := \min_{i=1}^{m} d_i(TR), \quad I := \arg \min_i d_i(TR)$$
If $d_i(\text{TR})$ is greater than a given threshold (‘trajectory - assimilate-threshold’), the trajectory will be created as a new route.

Otherwise, a match between the trajectory and route $I$ is detected. Go to 2\textsuperscript{nd} step, route updating.
Central axis node: \( \overrightarrow{X_{lk}} := \frac{w_I}{w_I + 1} \cdot \overrightarrow{x_{lk}} + \frac{1}{w_I + 1} \cdot \overrightarrow{x_{lkt}} \)
Route extension: The longer part of the trajectory is used to extend the route I. If the extension length is more than half of the length of the entire route, the trajectory is created as a new route.
**Boundaries**: The farthest intersections on both sides are the new left and right boundary nodes.

**Weight of route**: incremented by 1.
Sort all the routes by descending order of weights.

Step 1: Start from the route with the highest weight (\(R_h\)). Calculate the Hausdorff distance between \(R_h\) and the route with the smallest weight in the database (\(R_s\)).

Step 2: If the distance is less than a predefined threshold (\(\text{route-merge-threshold}\)), then the secondary route is qualified to be merged into the main route \(R_h\). Otherwise, they are not merged.

Step 3: Go to the next lower weighted route and repeat the above two steps until all the routes in the database are compared with \(R_h\).

Step 4: Go to the next higher weighted route and start a new loop (the above 3 steps), until all the routes (except the least weighted route) are used as \(R_h\) to be compared.
User file control

Route Formation

Trajectory Data
- Skip Trajectory Data Loading (Route Merge and Path Formation Only)
- Input Trajectory Data:
  - Text: C:\research\RouteFormation\Sample_Data\Div_TK030106_tracks_prj_BInnt.txt
  - Trajectory ID:
  - Output Trajectory Data:
    - C:\research\RouteFormation\RouteFormation\2008-05-04_1547\TrajOut(2008-06-03_2156).shp
- Calculate Matches for Trajectories
  - Output Statistics Data:
    - C:\research\RouteFormation\RouteFormation\2008-05-04_1547\MatchStatistic(2008-06-03_2156).shp
  - Test Trajectories (Text):
    - C:\research\RouteFormation\AlarmTest_data\AlarmTest(2008-06-03_2100).txt
  - Flagged Test Trajectories:
    - C:\research\RouteFormation\AlarmTest(2008-06-09_2200)\Alarm(2008-06-09_2100).shp

Route Data
- Use Input Route Data:
  - C:\research\RouteFormation\Output(2008-06-06_1122)\Sorting\RoutesOut(2008-06-06_1122).shp
  - Output Route Data:
    - C:\research\RouteFormation\RouteFormation\2008-05-04_1547\RoutesOut(2008-06-09_2156).shp

Path Data
- Output Path Data:
  - C:\research\RouteFormation\RouteFormation\2008-05-04_1547\PathsOut(2008-06-09_2156).shp
- Output Junction Data:
  - C:\research\RouteFormation\RouteFormation\2008-05-04_1547\JunctionsOut(2008-06-09_2156).shp
## User parameter control (metres)

### Constants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENVELOPE_OFFSET</td>
<td>0.1</td>
</tr>
<tr>
<td>RESAMPLE_INTERVAL</td>
<td>2000</td>
</tr>
<tr>
<td>MIN_TRAJECTORY_LENGTH_IN_RESAMPLES</td>
<td>3</td>
</tr>
<tr>
<td>P_Q_INTERSECTION_DIST_THRESHOLD</td>
<td>3000</td>
</tr>
<tr>
<td>MIN_COMMON_PROPORTION_FOR_MERGING</td>
<td>0.50</td>
</tr>
<tr>
<td>MAX_UNCOMMON_PROPORTION_FOR_MERGING</td>
<td>0.50</td>
</tr>
<tr>
<td>TRAJECTORY_ASSIMILATE_THRESHOLD</td>
<td>2500</td>
</tr>
<tr>
<td>ROUTE_TRAJ_UPDATE_THRESHOLD</td>
<td>2500</td>
</tr>
<tr>
<td>ROUTE_TRAJ_BOUND_THRESHOLD</td>
<td>2500</td>
</tr>
<tr>
<td>ROUTE_MERGE_THRESHOLD</td>
<td>2500</td>
</tr>
<tr>
<td>ROUTE_ROUTE_UPDATE_THRESHOLD</td>
<td>2500</td>
</tr>
<tr>
<td>ROUTE_ROUTE_BOUND_THRESHOLD</td>
<td>5000</td>
</tr>
<tr>
<td>MAX_ROUTE_EXTENSION_WEIGHT</td>
<td>15</td>
</tr>
<tr>
<td>TRIM_NODE_PROPORTION</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Implementation of the algorithm

- Data source: VTOSS database (Vessel Traffic Operations and Support Systems)
- Study area: the ships that approach within 50km of the British Columbia (BC) shoreline.
- Study period: Fourteen daily datasets from Jan. 1\textsuperscript{st} to 14\textsuperscript{th}, 2003; three weekly datasets – Jan. 1\textsuperscript{st} to 7\textsuperscript{th}, 8\textsuperscript{th} to 14\textsuperscript{th}, and Aug. 6\textsuperscript{th} to 12\textsuperscript{th}, 2003.
- The following figures are displayed for a one-week traffic data (from Jan. 1\textsuperscript{st} to 7\textsuperscript{th}, 2003).
Historical trajectories
(resample_interval = 2500, trajectory_assimilate_threshold = 3000, route_merge_threshold = 3000)
Kernel Density of one week of traffic
Kernel Density vs Generated Routes
Cumulative sum (CUSUM) control chart is applied in maritime traffic to monitor the process (i.e. movement) of any new individual trajectory. Let $x_i$ be the $i^{\text{th}}$ observation on the process. $\mu_0$ is the target value for the quality characteristic $x$.

The tabular CUSUM works by accumulating deviations from $\mu_0$ that are above target with one statistic $C^+$ and accumulating deviations from $\mu_0$ that are below target with $C^-$. 

$$C^+_i = \max[0, x_i - (\mu_0 + K) + C^+_{i-1}]$$
$$C^-_i = \max[0, (\mu_0 - K) - x_i + C^-_{i-1}]$$
$x_i =$ the minimum distance from node $i$ on the trajectory to the route envelope.

The overall mean and standard deviation (unbiased) are calculated for the route across all incorporated trajectory nodes to route envelope distances.

Only $C^+$ statistic is used, because distance is a positive variable.

Two levels of alarm are set.
Alarm test of two individual trajectories

Traj: 5-380

Traj: 2-516
Discriminating recreational boat types
Characterizing of Recreational Boats based on GPS Trajectory Points
This Study examines Global Positioning System (GPS) trajectory data to characterize spatial patterns, providing insight into recreational boat movements and amount of exposure to risk.
Spatial Pattern I

Patterns of Different Recreational Boat Types

(a) c-0707-08-03  (b) c-0721-11-04  (c) s-0722-14-10
Patterns of Different Recreational Boat Types at Different Locations

(a) s-0722-14-10  (b) s-0824-16-03
### Scope of Study

<table>
<thead>
<tr>
<th>Sample size</th>
<th>NB</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canoe</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Kayak</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>Motorboat</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Sailboat</td>
<td>5</td>
<td>47</td>
</tr>
</tbody>
</table>
Methodology

- Data Pre-cleaning
  - Points on land
  - Time gap
  - Resting/Stopping events
- Attributes Extracting/Calculation
  - Speed
  - Total distance
  - Bounding Box
  - Dedensified Trajectory
  - Distance from shore
- Discrimination & Classification
Attributes to Variates

- Speed
- Total distance
- Bounding Box
- Dedensified Trajectory
- Distance from shore

- Max Speed
- Max 1/20 Speed
- Mean Speed
- Total Distance
- Complexity
- Aspect Ratio
- Mean Turning Angle
- Distance to Shore
Discrimination

Kayak Trajectories

Canoe Trajectories

K-Avg. Speed

C-Avg. Speed

Null Hypothesis:
Ho: K-Avg. S = C-Avg. S
Alternative Hypothesis:
Ha: K-Avg. S ≠ C-Avg. S

Conclusion

Classification

Trajectories

Pattern Classification Procedure

Predicted:
Kayak: T1, T5...
Canoe: T3, T4...
Sailboat: T6, ...
Motorboat: T2...

Actual:
Kayak: T1, T5...
Canoe: T3, T4...
Sailboat: T6, ...
Motorboat: T2...

Correction?

Conclusion
## Discriminant Analysis

*Do different boat types have significant different attributes?*

<table>
<thead>
<tr>
<th>Attribute</th>
<th>NB</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td></td>
<td>C, K, S, M</td>
</tr>
<tr>
<td>Mean.S</td>
<td>C, K, S, M</td>
<td>C, K, M, S</td>
</tr>
<tr>
<td>Max.S</td>
<td>C, K, S, M</td>
<td>C, K, M, S</td>
</tr>
<tr>
<td>TAQ</td>
<td>C, K, S, M</td>
<td>C, K, M, S</td>
</tr>
<tr>
<td>TAH</td>
<td>C, K, S, M</td>
<td>C, K, M, S</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>S</td>
</tr>
<tr>
<td>DTS</td>
<td>C, K, M, S</td>
<td>C, K, S</td>
</tr>
<tr>
<td>AR</td>
<td>K</td>
<td>S</td>
</tr>
</tbody>
</table>
**Discriminant Analysis**

*Do geography characteristics influence boating trajectories?*

<table>
<thead>
<tr>
<th></th>
<th>Canoe</th>
<th>Kayak</th>
<th>Motorboat</th>
<th>Sailboat</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAH</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TAQ</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DTS</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Max.S</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Max 1/20.S</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Mean.S</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TD</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AR</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

P=0.1

X: Not reject $H_0$: There is no significant difference between the two areas.
Problem: Whether can we determine the type of recreational boat solely through the trajectory data without reference to the weather and operator factors, even geography?

Objectives:

- Determining whether statistically significant differences exist between attributes of different boating trajectories of different boat types
- Determining which of the independent variables account the most for the differences in the average score profiles of the four groups.
- Establishing procedures for classifying boating trajectories on the basis of their score on a set of attributes.
Estimation of the Simultaneous Discriminant Analysis

**Standardized Canonical Discriminant Functions:**

\[ Z_1 = 0.047x_1 + 1.009x_2 + 0.585x_3 + 0.093x_4 \]
\[ Z_2 = 0.454x_1 - 0.584x_2 + 0.469x_3 + 0.671x_4 \]
\[ Z_3 = -0.744x_1 + 0.202x_2 + 0.732x_3 + 0.174x_4 \]

**Fisher’s Linear Classification Functions:**

- **Canoe**
  \[ = -6.941 - 5.223E - 05X_1 + 3.301X_2 + 0.125X_3 + 5.894E - 04X_4 \]

- **Kayak**
  \[ = -5.629 - 1.516E - 05X_1 + 3.251X_2 + 0.103X_3 + 3.209E - 04X_4 \]

- **Motorboat**
  \[ = -27.16 - 6.223E - 05X_1 + 9.891X_2 + 0.191X_3 + 7.617E - 05X_4 \]

- **Sailboat**
  \[ = -13.74 + 3.057E - 05X_1 + 5.156X_2 + 0.179X_3 + 1.498E - 03X_4 \]

\(x_1=\text{Total distance}, \ x_2=\text{Mean speed}, \ x_3=\text{mean turning angle}, \ x_4=\text{farthest distance to shore.}\)
## Classification Results

<table>
<thead>
<tr>
<th>Original Count</th>
<th>TYPE</th>
<th>canoe</th>
<th>kayak</th>
<th>motorboat</th>
<th>sailboat</th>
</tr>
</thead>
<tbody>
<tr>
<td>canoe</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>kayak</td>
<td>3</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>motorboat</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>sailboat</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>47</td>
<td>47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>%</th>
<th>canoe</th>
<th>kayak</th>
<th>motorboat</th>
<th>sailboat</th>
</tr>
</thead>
<tbody>
<tr>
<td>canoe</td>
<td>88.2</td>
<td>11.8</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>kayak</td>
<td>14.3</td>
<td>85.7</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>motorboat</td>
<td>10.0</td>
<td>.0</td>
<td>80.0</td>
<td>10.0</td>
</tr>
<tr>
<td>sailboat</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

92.6% of original grouped cases correctly classified.
Key results

- Stepwise discriminant analyses for the coastal and river study areas retain the same independent variables
  - Mean Speed, Mean Turning Angle, Farthest distance to shore

- Geography does not influence boating trajectories much

- Mean Speed, Mean Turning Angle, Farthest distance to shore are significantly different among different groups
P(Kayak) = 0.733
P(Canoe) = 0.240
P(Motorboat) = 0.899
P(Sailboat)  = 0.101
THANK YOU!

QUESTIONS?