Obtaining Vessel Routes from VTOSS Data

Presentation by: Jason R. W. Merrick

GWU Personnel: Dr. J. Rene van Dorp

VCU Personnel: Dr. Jason R. W. Merrick

Puget Sound Harbor Safety Committee Presentation December 2012
Before Cleaning
Trip IDs
Identifying a Continuous Route

• Sort by Vessel Name and then Timestamp
• If the Trip ID is the same, then points are on the same route
• If the Trip ID changes, but the Vessel Name is the same and less than an hour between points, then points are on the same route

• This means you have to fix
  – Different spellings of the same Vessel Name
  – Multiple vessels with the same Vessel Name
One Route with Errors
What’s Wrong Here?
Finding 12:xx pm Points

- Parse the points sorted by Vessel Name and then Timestamp
- If the points are time stamped 12:xx pm
  - Search for the points on this route either side of 12:xx am
  - Determine whether the point is closer to those either side of 12:xx am or those either side of 12:xx pm
  - Move points if necessary
- Special case if the point would be the first on the route or if it is the last point where it is
- Repeat Code Sorted by Vessel Name then Trip ID then Timestamp
Still Errors
Fixing Errors

• Parse through points on a continuous route
• Take each consecutive set of three points and consider the triangle
• Calculate the perpendicular distance between the middle point and the line between the first and third
• If the perpendicular distance is greater than the maximum distance the vessel could travel in the time between the first and third point
  – Delete middle point

• You can also check that the perpendicular distance is greater than zero otherwise the middle point isn’t needed on the route
  – Don’t need three points to draw a straight line
Still Missing Data
Finding Some Missing Data

- Some routes where a new Trip ID is started by the same VTS, but then the original Trip ID resumes later
- Parse continuous route
- Look for these cases
- Change new Trip ID to the original

- Ensure these cases are not 12:xx pm points that should be 12:xx am points on a different route
Still Missing Data
Manual Fixing

- These will need to be manually fixed
Generating Heat

• Very computational
  – Database sizes require that data is separated into months (1-2 GB each)
  – Cleaning one month of data for just TUG TOW BARGE took 5-7 hours
• Tug data is cleaned
• Moving on to everything else but tankers and ferries
  – Vessel Name disambiguation and spelling correction complete
  – First month of data preliminary cleaning complete
  – Now tuning algorithms to maximize cleaning before cleaning remaining months
Moving Away from Representative Routes

- In previous analysis, we couldn’t clean every route
- Each bulk carrier going from Vancouver to Tacoma followed the same representative route
  - Had to choose the cleanest route we could get
- Now we are cleaning all routes
- Each vessel will follow its route
- Let’s look at some tug routes