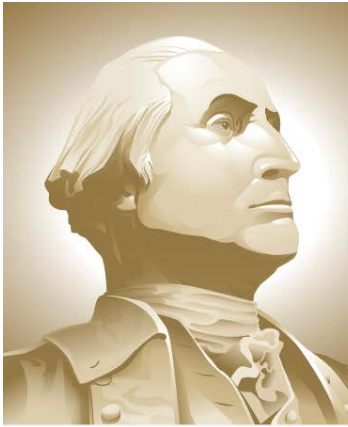


Obtaining Vessel Routes from VTOSS Data

Presentation by: Jason R. W. Merrick



THE GEORGE
WASHINGTON
UNIVERSITY

WASHINGTON, DC

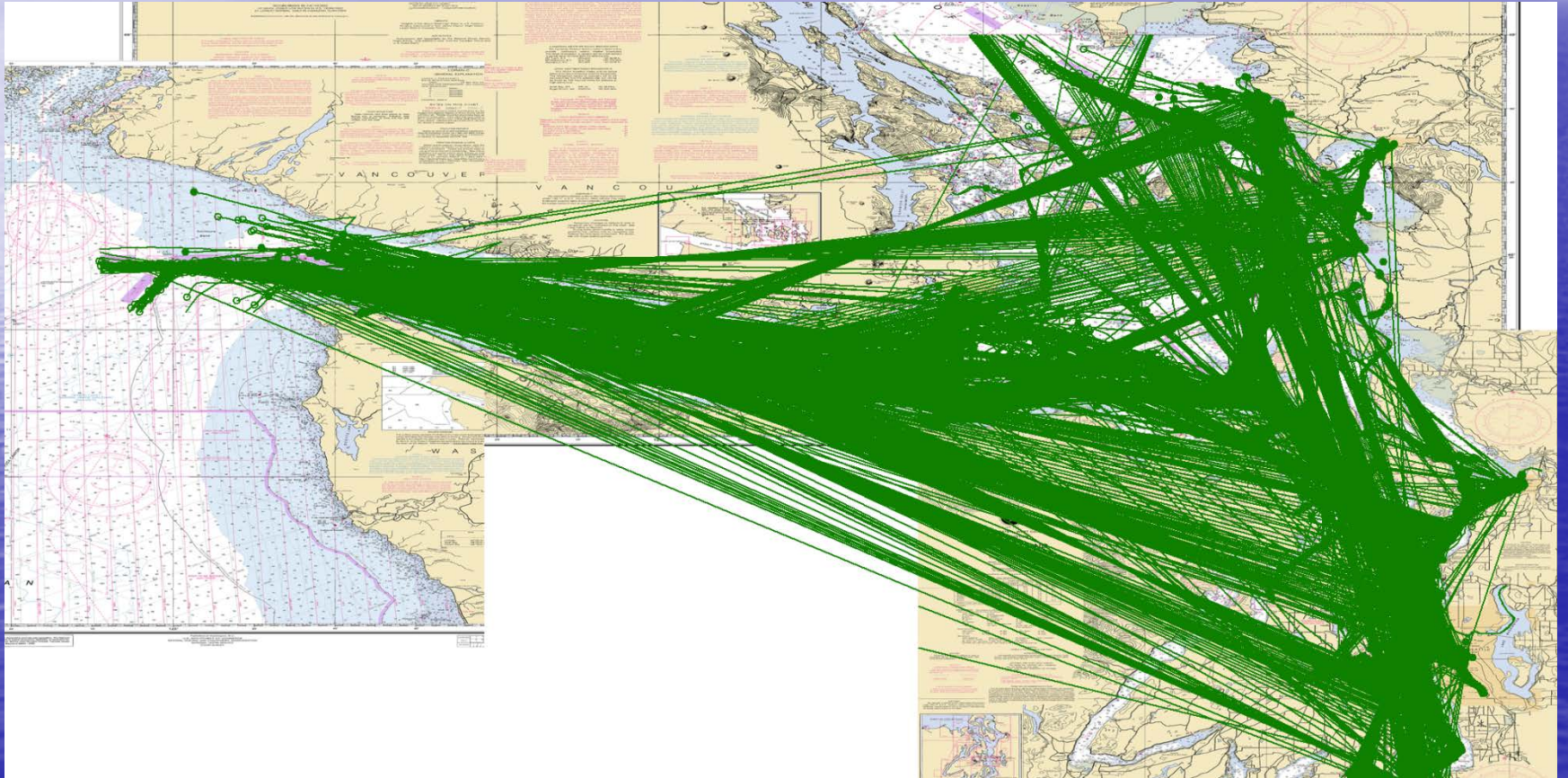
VCU

GWU Personnel: Dr. J. Rene van Dorp

VCU Personnel: Dr. Jason R. W. Merrick

Puget Sound Harbor Safety Committee Presentation December 2012

Before Cleaning

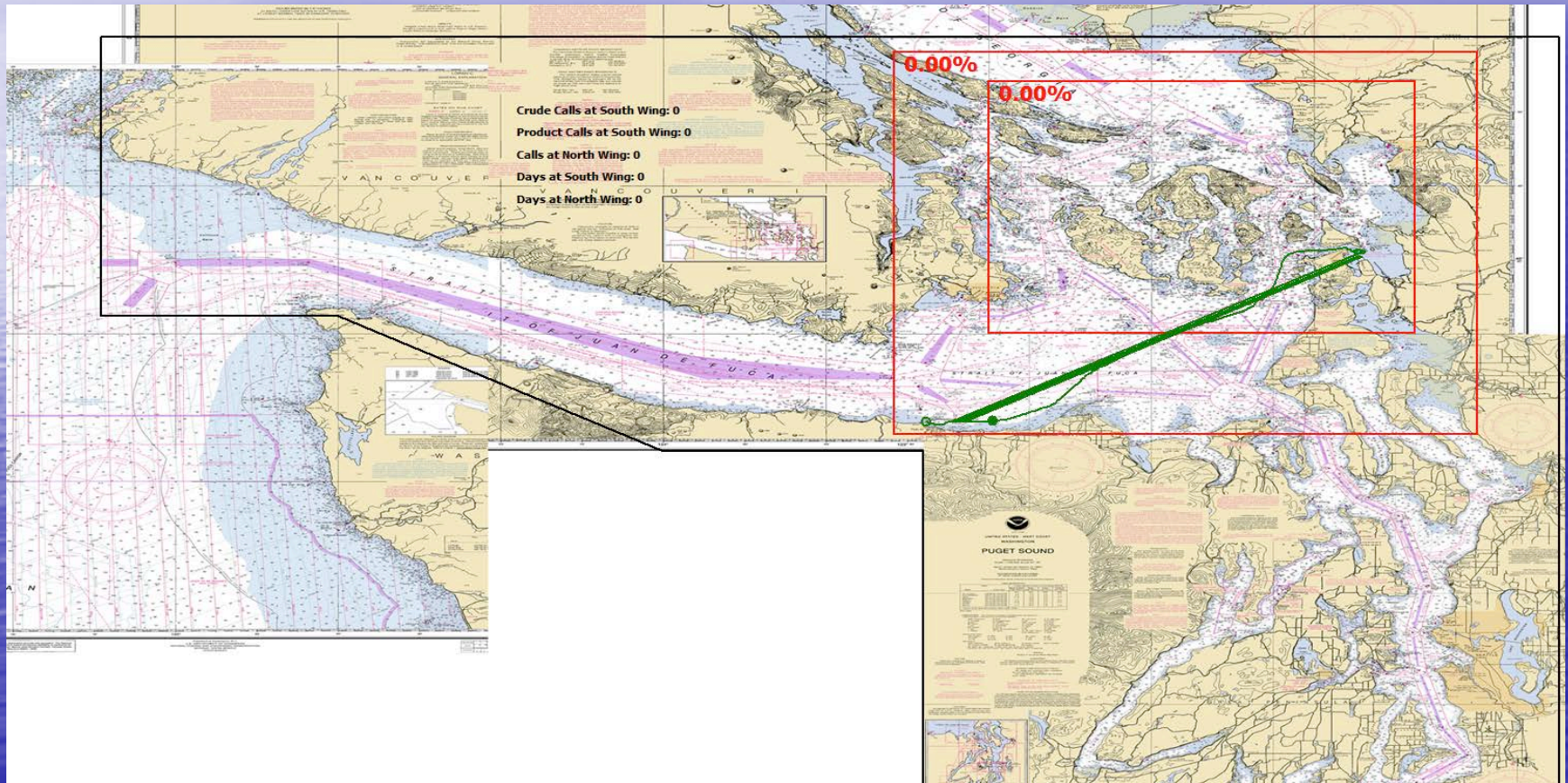


Trip IDs

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?

The screenshot shows Microsoft Access with a table named 'SortedGoodData' open. The table contains the following data:

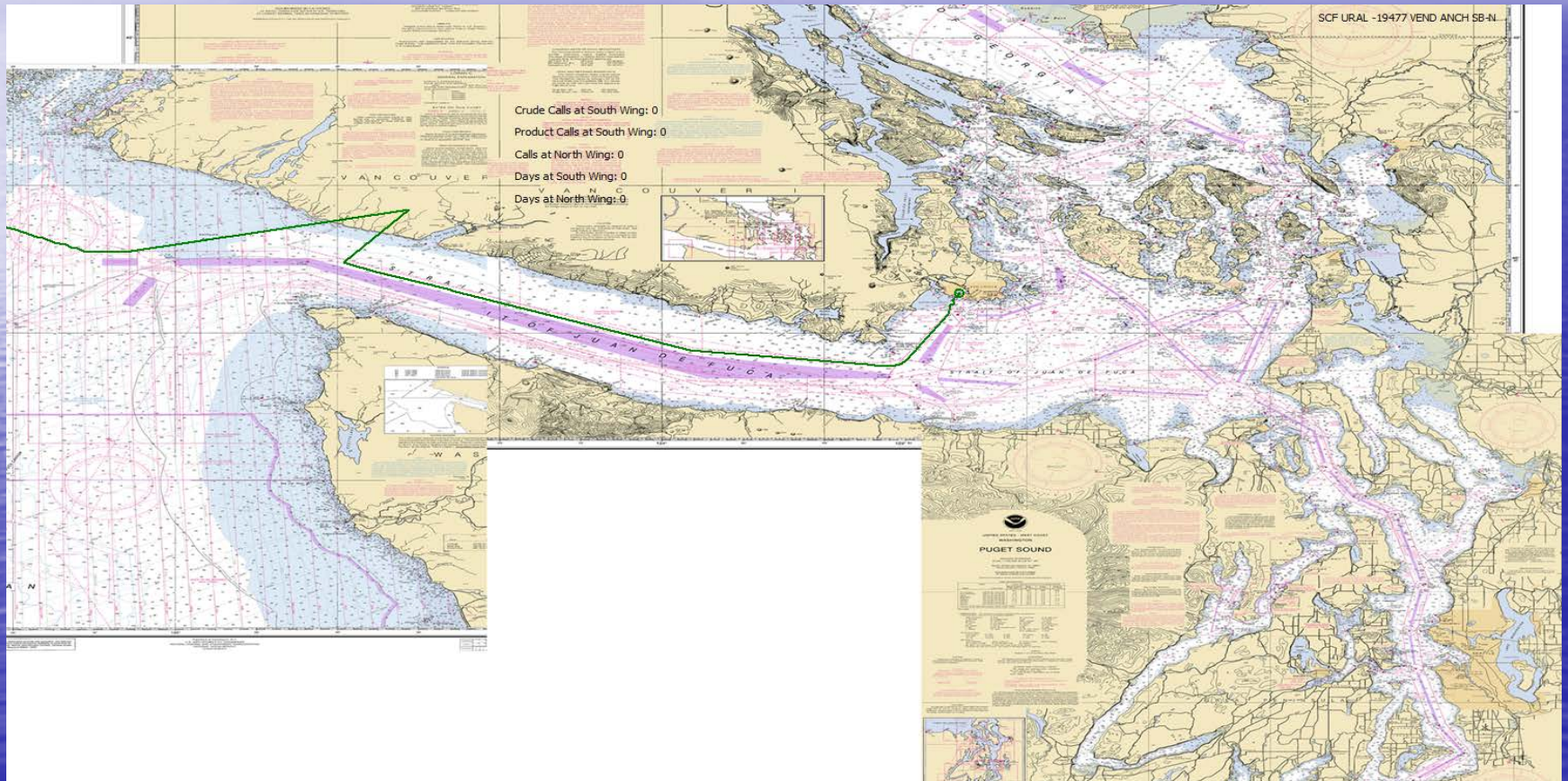
LAST_UDDTC	NAME	TIMESTAMP	POS_LAT	POS_LONG	TRIP
201002121258	ADMIRAL PETE	2/12/2010 12:58:00 PM	47.587	122.413	2010088812
201002071201	ADRIATIC BRID	2/7/2010 12:01:00 PM	47.274	122.409	2010087273
201002071201	ADRIATIC BRID	2/7/2010 12:01:00 PM	47.274	122.409	2010087273
201002071202	ADRIATIC BRID	2/7/2010 12:02:00 PM	48.139	123.38	2010087210
201002071204	ADRIATIC BRID	2/7/2010 12:04:00 PM	48.137	123.386	2010087210
201002071204	ADRIATIC BRID	2/7/2010 12:04:00 PM	48.137	123.386	2010087210
201002071205	ADRIATIC BRID	2/7/2010 12:05:00 PM	47.274	122.409	2010087273
201002071205	ADRIATIC BRID	2/7/2010 12:05:00 PM	47.274	122.409	2010087273
201002071207	ADRIATIC BRID	2/7/2010 12:07:00 PM	48.135	123.395	2010087210
201002071207	ADRIATIC BRID	2/7/2010 12:07:00 PM	48.135	123.395	2010087210
201002071207	ADRIATIC BRID	2/7/2010 12:07:00 PM	47.274	122.409	2010087273
201002071207	ADRIATIC BRID	2/7/2010 12:07:00 PM	47.274	122.409	2010087273
201002071210	ADRIATIC BRID	2/7/2010 12:10:00 PM	47.274	122.409	2010087273
201002071210	ADRIATIC BRID	2/7/2010 12:10:00 PM	47.274	122.409	2010087273

The second row with timestamp '2/7/2010 12:01:00 PM' is highlighted in red, indicating a duplicate record. The interface includes a ribbon with 'FILE', 'HOME', 'CREATE', 'EXTERNAL DATA', and 'DATABASE TOOLS' tabs. The 'HOME' tab is active, showing options for View, Clipboard, Sort & Filter, Records, and Find. The left pane shows a list of tables (VTOSSData03 to VTOSSData12) and a 'Queries' section.

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

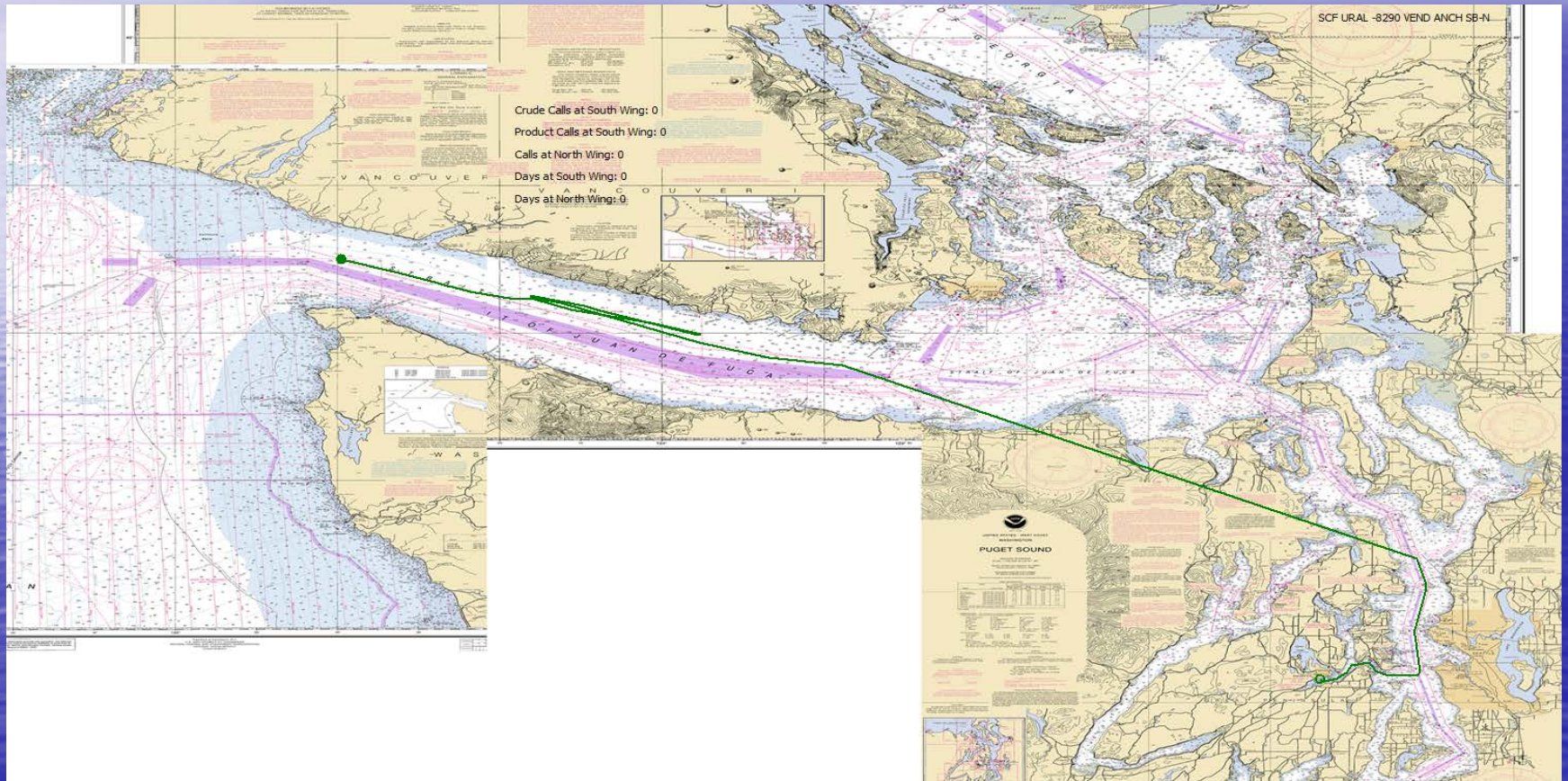
Still Errors



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

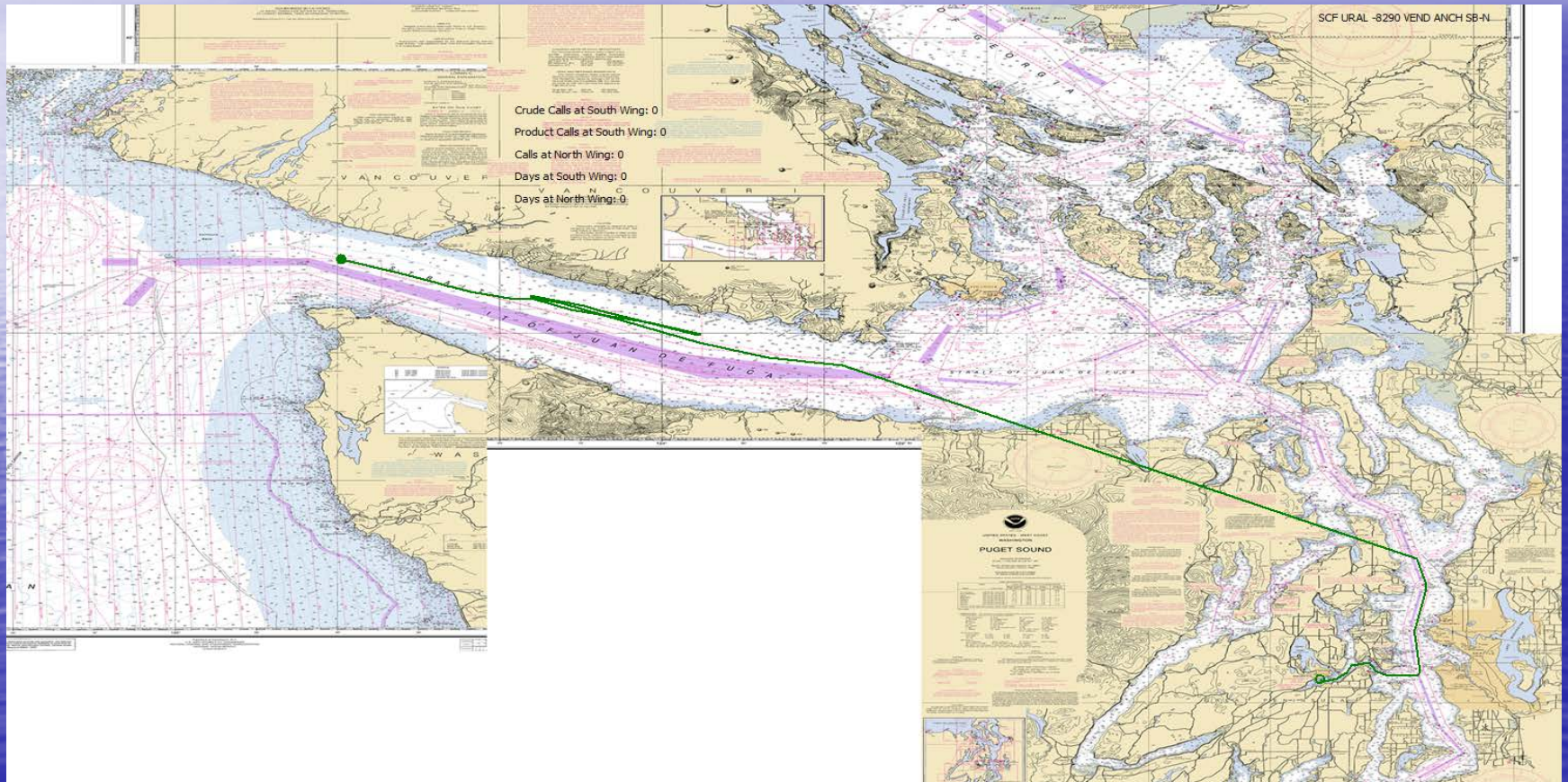
Still Missing Data



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