

Improved Maritime Statistics with Big Data

Session 15

June 2, 2016

Abboud Ado, Transport Analysis, Abboud.Ado@Trafal.se

Annica Isaksson de Groote, Statistics Sweden, Annica.Isaksson@scb.se

Ingegerd Jansson, Statistics Sweden, Ingegerd.Jansson@scb.se

Marcus Justesen, Statistics Sweden, Marcus.Justesen@scb.se

Jerker Moström, Statistics Sweden, Jerker.Mostrom@scb.se

Fredrik Söderbaum, Transport Analysis, Fredrik.Soderbaum@Trafal.se

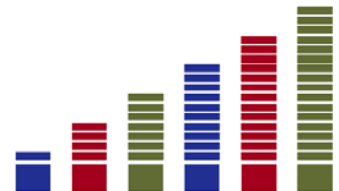
Big Data and AIS

Big Data:

- Huge volumes of data
- Generated at short intervals and often unstructured
- Require innovative methods and technological solutions to extract useful information

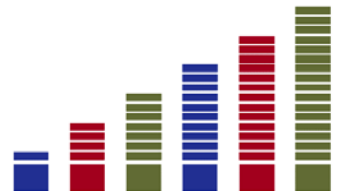
Automatic Identification System, AIS:

- Makes it possible for ships to identify and track other ships' movements
- Provides detailed information about the ships
- Relies on digital information transmitted by the ships at short intervals



Pilot study

- Conducted jointly by Statistics Sweden and the Swedish government agency Transport analysis in 2015
- Partly funded by Sweden's Innovation Agency Vinnova
- **Aim:** Investigate the potential of AIS data to improve maritime statistics



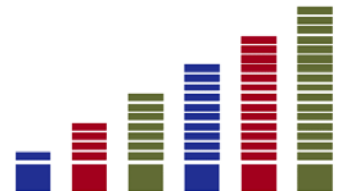
Pilot study

- **Problem:** Official statistics on maritime transports are based on survey data from Swedish ports. One variable of interest is

$$\text{Transport Performance (TP)} = \text{Amount of transported goods} \times \text{Travel distance}$$

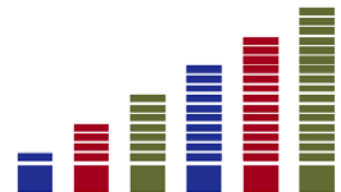
The TP requires information about the distances between ports. The flat-rates used today are known to be flawed.

- Can AIS data simplify the data collection and improve the quality of the distance measures?



Data

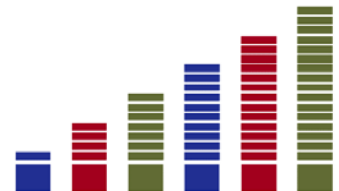
- Provided by the Swedish Maritime Administration
- Two weeks in 2014: 7-20 September
- The entire Baltic Sea and Skagerrak and Kattegat
- About one million data points per day
- Passenger traffic excluded from analysis



Creating port areas

A data set of port areas was created, containing:

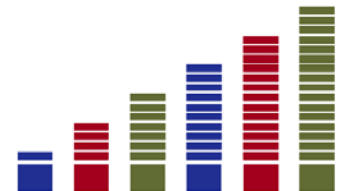
- Swedish ports areas,
- a dummy area representing foreign ports, and
- water (outside the ports)



Creating a population of routes

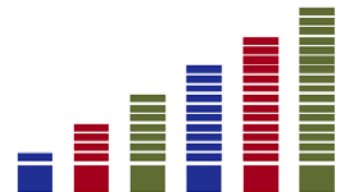
Two approaches:

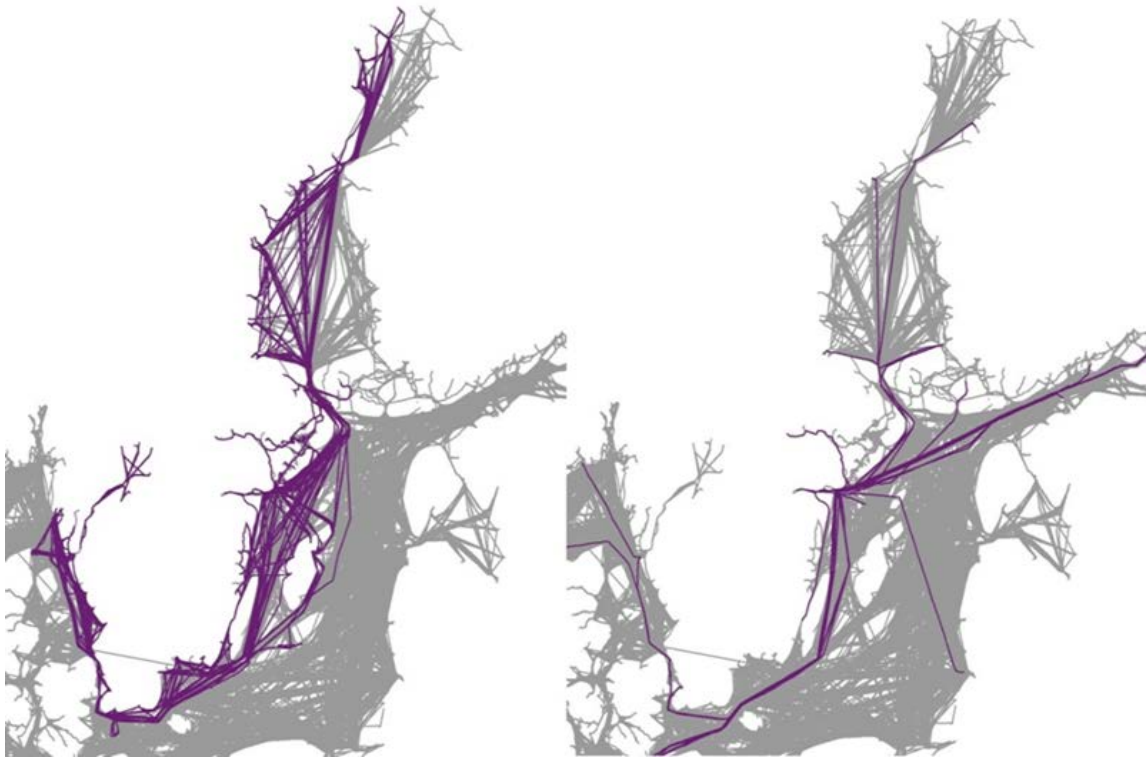
- Vector-based (direct)
- Raster-based (model-based)



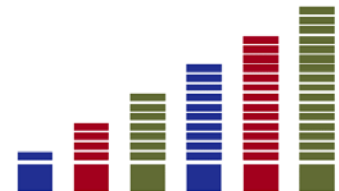
A vector-based approach

- All AIS data points are categorized to fall into one of the three port areas
- For each transport between ports, a line connecting the AIS data points is created based on port, time stamp and Maritime Mobile Service Identity (MMSI)
- Each line is updated with information about origin and destination of the vessel





Left: all traffic between Swedish ports (in purple). Right: all traffic to Norrköping (in purple). Grey lines represent all traffic in the area.

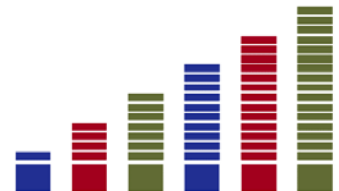


A raster-based approach

- Consider using *raster*, defined as:

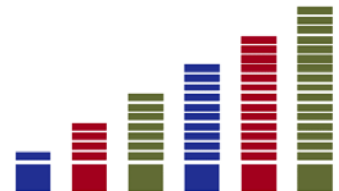
A surface composed of cells with a specific definition, e.g. 1 x 1 m, which contains a value for each cell

- The value in each cell is chosen to represent any information necessary for modelling the network



A raster-based approach

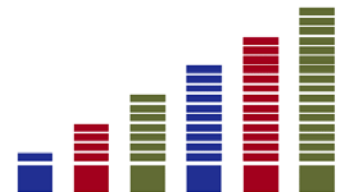
1. All lines are converted to a **raster**.
2. For each raster cell, line density is calculated. Line density determines the **waterway**.
3. The waterway combined with destination defines the **cost** to travel through each cell.
4. A distance raster is created from the cell to the port of interest. Each cell in the raster is assigned a **value** equal to the least accumulated cost of being transported from the cell to the port.
5. A **route** is calculated as the path with the least accumulated cost from specified port or ports to the port of interest based on the distance raster.



Vectors vs. raster

A vector-based approach:

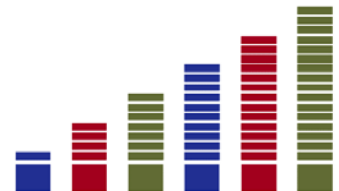
- Produces very accurate results
- Is very time consuming as it involves repeated processing of large quantities of data
- The population of routes needs to contain enough data to cover all possible routes, ideally with a large number of observations



Vectors vs. raster

A raster-based approach:

- Produces accurate result, however based on ideal distances between ports rather than actual measured routes
- Is based on AIS data, but instead of recalculating the actual measured routes every time, the calculations can be made once and then fed into the model
- Makes sense since maritime transport operations take place on a *surface*, they are not limited to the roads the way road traffic is
- Can handle lots of data fast
- Flexible and suitable for experiments



Summary

- We have developed preliminary methods for identifying routes between Swedish ports, and between Swedish and foreign ports
- Our models for calculating distances, and especially the raster-based approach, have shown promising results
- Further development work is needed

