



Rome. 4-5 December 2024 THIRD WORKSHOP ON METHODOLOGIES FOR OFFICIAL STATISTICS

Statistical use of Automatic Identification System (AIS) data

Marco Di Zio, F. Arosio, A. Martino, G. Massacci, F. Ortame, A. Pappagallo, F. Pugliese, N. Salamone, F. Sisti, S. Talice, L. Valentino

Istat

- Context
- Data transformation
- Quality analysis
- Methods for dealing with gaps
- Final remarks



- Maritime transport survey aims to provide statistics on transport of goods and passengers carried out by vessels engaged in commercial activities
- Currently produced with a complete enumeration survey of shipping agencies/owners of ships, and administrative data of Port Authorities
- Eurostat request: improving timeliness and quality.



Data on the internet (using AIS)





Common feeling: it is easy!

When you do in practice: it is not easy

Problems. Data availability, data management, data transformation, quality

Data availability. Data are not free.

Fortunately, recently, UN made it available for free (for members of Committee of Experts on Big Data)

Data management is not straightforward: data in cloud, not possible to download, a huge amount of data,...

Now we made experience, and we are able to work with AIS data



Even with data at hand, it is not easy to produce (our) statistics

They are signals, need transformation to have "routes", for instance for the count of docked vessels for each port

First elements to compute:

- 1. Delimit an area defining a port
- 2. From signals, define when the vessel is docked.



Calculate each port area as a geo-referenced polygon (use H3 system).

Hexagon corresponding to port coordinates augmented with hexagons containing potential positions of the stationary ships (estimated through data)





From signals to route

Static data

CALLSIGN

ZAD4L

IMO

8401561

MMSI

20110115

Reconstruct all maritime voyages involving ships departing or arriving at an Italian port.

VESSEL

NAME

FROJDI II

VESSEL

TYPE

Cargo

TIME

04/06/2023 19:45

Dinamic data

COORDINATES

41.1323 16.8530

NAVIG.

STATUS

MOORED

SPEED

0

COURSE

258

A voyage is defined by two key events: departure and arrival.

In AIS these events are reflected in records that meet a specific criterion:

• Speed = 0 in an hexagon of a port

Routes built through the sequence of signals, ports and speed

NO	VESSEL TYPE	DEPARTURE PORT	ARRIVAL PORT	DEPARTURE DATE	ARRIVAL DATE
8401561	Cargo	ITBRI (Bari)	ITRAN (Ravenna)	04/09/2023	05/09/2023
9483712	Passenger	ITGOA (Genova)	ILOLB (Olbia)	05/09/2023	06/09/2023

Travel related data

DRAFT

null

DESTINATION

Ravenna



Data for (our) statistical use are now available.

- Are there any gaps?
- What is the quality?



1. Macroediting. Comparison of aggregated statistical figures

2. Record linkage for micro comparison

Problems:

- Short and frequent routes are under-counted
- Interruption of signals
- Anomalous routes





1st quarter of 2022. In 14% of the 98,127 voyages, the port of arrival was (potentially) missing.

The average velocity is less than a minimum threshold



Deterministic. If consistent, impute the port with the value specified in the '*destination*' field available in AIS. Otherwise, if acceptable, set most frequently destination of the ship, given the port of departure, over a specified time period.

Deep learning. Studied 2 models (1) a one-dimensional convolutional network (2) a long short-term memory network (LSTM). Estimated accuracy around 82%



Need to integrate space and time information....

we evaluate the average speed between pairs of data observations



Outlier. If average speed is greater than a threshold (max speed of a ship) .

This approach reduces the problem to a univariate analysis.



Implemented a solution involving a cyclical application of the algorithm. In the first cycle only the first outlier is removed, while in the subsequent cycle the second outlier is eliminated, and so on.





Methods for Anomalous routes - outliers





Improve some statistical figures, for instance the offshore platforms

The location of the offshore platforms has allowed the identification in AIS, for 2022, of almost 5,000 arrivals, the official procedure slightly more than 230.





Arrivals in TRAMAR survey (blue), in admin data PMIS (red) and in AIS (green) for each cruise ship (id ship replaced by numbers 1 to 20).





Now, we have AIS data available for the the statistical process.

For the moment, AIS can be used to supplement existing data sources:

o To deal with under-coverage of the other sources

o To check , validate and improve official procedure

However, further analysis are needed for introducing AIS data into the statistical production process



Thank you

MARCO DI ZIO | dizio@istat.it



