Statistics Poland

TranStat System - An intelligent system for the production of maritime and road transport statistics using Big Data sources and technology

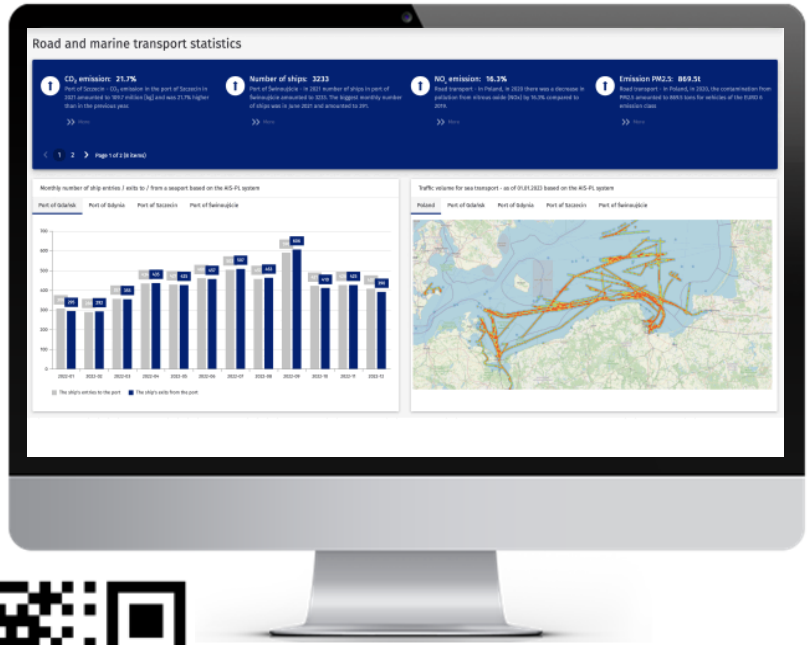
Beata Brzostowska, Bartosz Sidwa
Dominik Rozkrut, Michał Bis
Statistics Poland

Agenda

- TranStat – overview, assumptions, architecture, data flow and processing
- AIS - Automatic Identification System in a nutshell
- Statistics of traffic intensity, transportation volume and emissions in maritime transport – assumptions, results
- Conclusions

TranStat – overview

https://transtat.stat.gov.pl



TranStat

The project focused on:

- obtaining access to sensory data from the Automatic Identification System (AIS) and the e-TOLL electronic toll collection system;
- adaptation of modern Big Data methods and tools;
- development a methodology for estimating traffic intensity, transportation volume and the amount of emissions;
- implementation of experimental statistics in domain of road and maritime transport;
- lower costs by using non-statistical sources;
- speeding up the publication of statistics.

Project term: 2019 - 2021

Consortium: Statistics Poland, Maritime University of Szczecin, Cracow University of Technology

Statistics Poland

TranStat – assumptions & architecture

The general requirements for modern IT systems, including:

- implementation of open standards;
- technological neutrality (vendor lock-in);
- modular construction;
- easy expansion with new system functionalities in the future;
- ensuring an appropriate level of security

and **the requirements for scalable Big Data solutions:**

- model for the 3 V's of big data**:** volume, velocity and variety.

TranStat – architecture

The TranStat IT system has been developed and implemented in the production environment of Statistics Poland.

The following functional subsystems have been developed as part of the system:

- **Data collection and processing subsystem** responsible for the following subprocesses:
  - decoding AIS data,
  - processing stream data from sensors,
  - integration, validation, transformation and aggregation of data.

- **Data presentation and analysis subsystem - internal**, the purpose of which is to enable data exploration and visualization as well as statistical analyzes using the RStudio and Apache Zeppelin tools.

- **Data presentation and analysis subsystem - external,** intended for an external users, operating on the basis of calculated aggregates and indicators - https://transtat.stat.gov.pl

TranStat – the process of data flow and processing

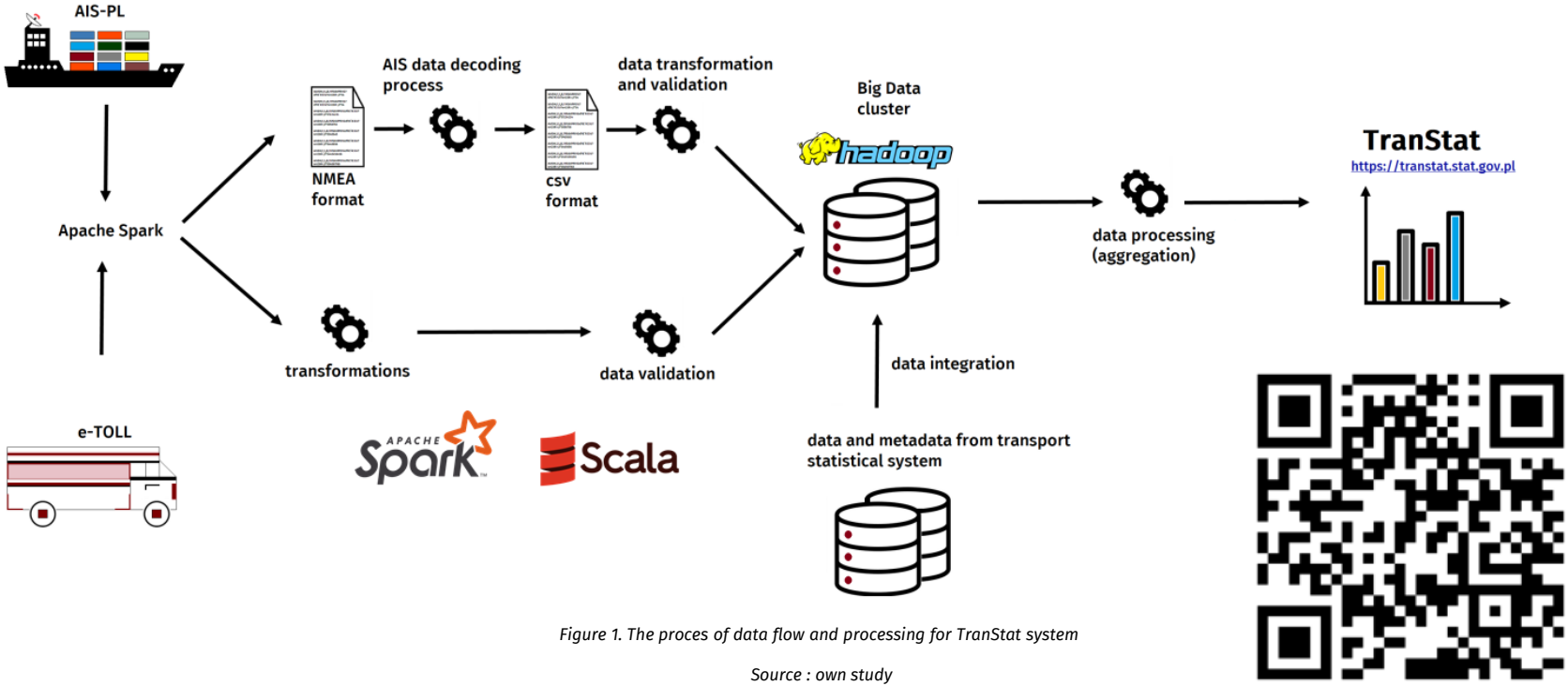


Figure 1. The proces of data flow and processing for TranStat system

Source : own study

Automatic Identification System in a nutshell

Automatic Identification System (AIS) - it is a system of automatic identification used on ships for the electronic exchange of information between nearby ships, AIS base stations and satellites.

According to the requirements defined in Chapter V of the SOLAS Convention developed by the IMO, the AIS system should be installed on:

- all ships of 300 gross tonnage and more used in international shipping,
- all ships of 500 gross tonnage and more not used in international shipping,
- all passenger ships, regardless of size.

The basic application of the AIS system:

- strengthening of navigation safety (anti-collision system),
- vessel traffic management support for coastal Vessel Traffic Service (VTS).

Data source availability - legal basis

Regulation of the Minister of Maritime Economy and Inland Navigation of September 26, 2018 on the National System for Monitoring Vessel Movement and Information Transmission.

Automatic Identification System in a nutshell

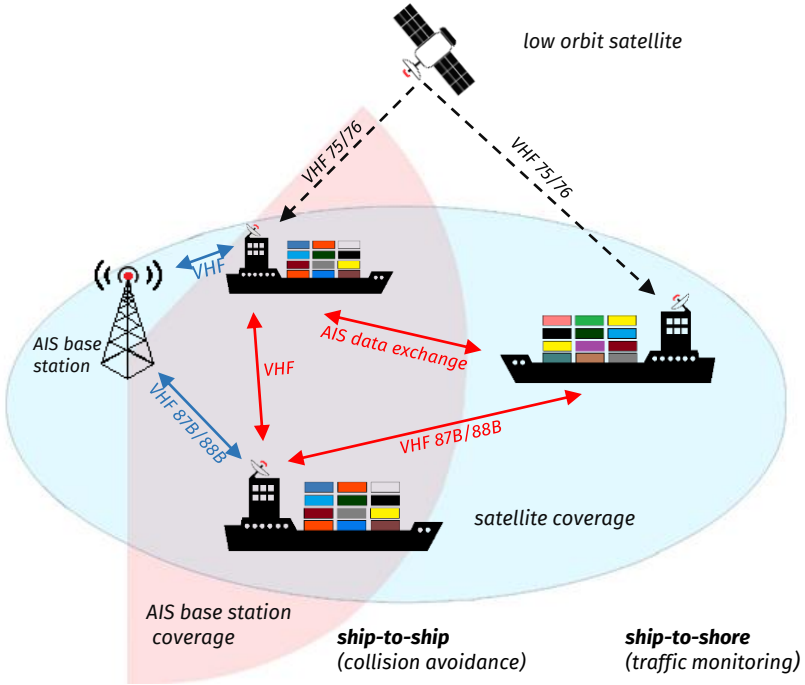


Figure 2. AIS – the principle of work

Source : own study

Dynamic data

- Information on ship movements
- Automatically transmitted
- Every 2 to 10 seconds depends on vessel's speed
- Every 3 to 6 minutes when anchored

- Maritime Mobile Service Identity number (MMSI)
- AIS navigational status
- Rate of turn
- Speed over ground
- Position coordinates (longitude/latitude)
- Course over ground
- Heading
- Bearing at own position
- UTC second

Static data

- Information on ship characteristic
- Manually transmitted
- Every 6 minutes

- International Maritime Organisation number (IMO)
- Call sign
- Name
- Type
- Dimensions
- Location of the positioning antenna on the vessel
- Type of positioning system
- Draught
- Destination
- ETA (estimated time of arrival)

Statistics of traffic intensity in maritime transport – assumptions

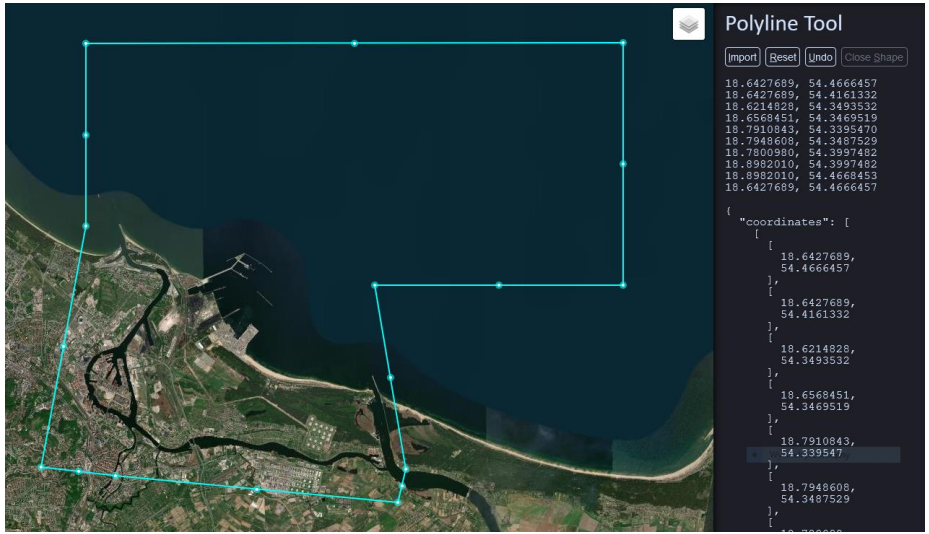
Traffic intensity is understood as the intensity of the stream, defined as the number of transport units passing through the line delimiting a given area in a certain period of time.

Implementation in TranStat application:

- **Location:** ports of Gdańsk, Gdynia, Szczecin, Świnoujście.
- **Data source:** Automatic Identification System (AIS)

Statistics of traffic intensity in maritime transport

Port of Gdańsk



Port of Gdynia

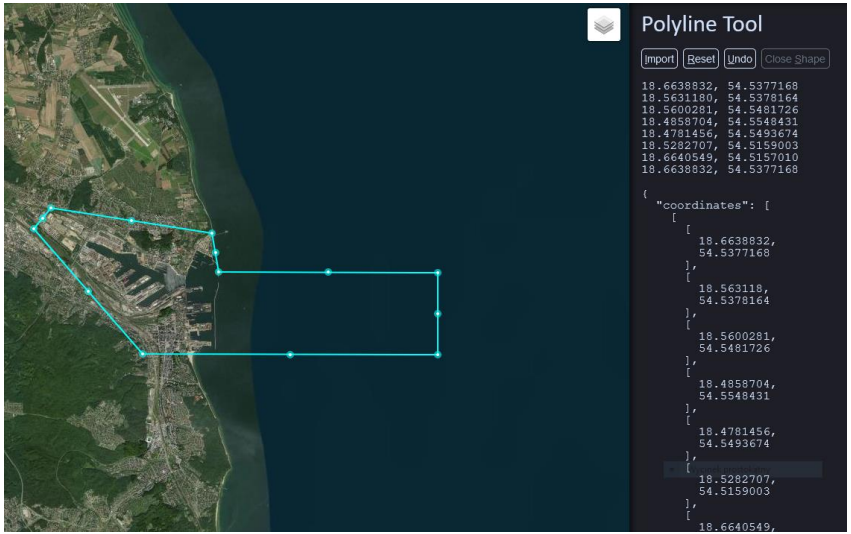
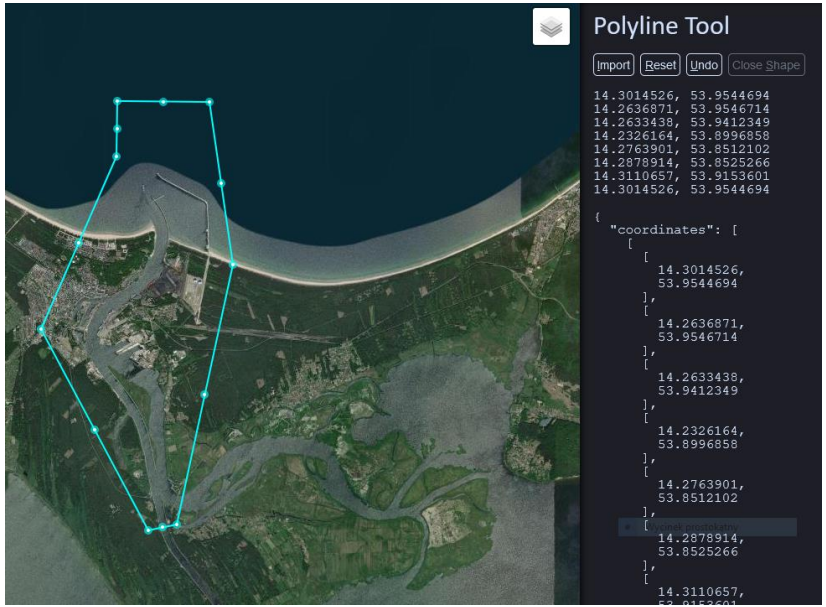


Figure 3. The ports of Gdańsk, Gdynia

Source: own study, generated on the basis of the tool:
https://www.keene.edu/campus/maps/tool/

Statistics of traffic intensity in maritime transport

Port of Świnoujście



Port of Szczecin

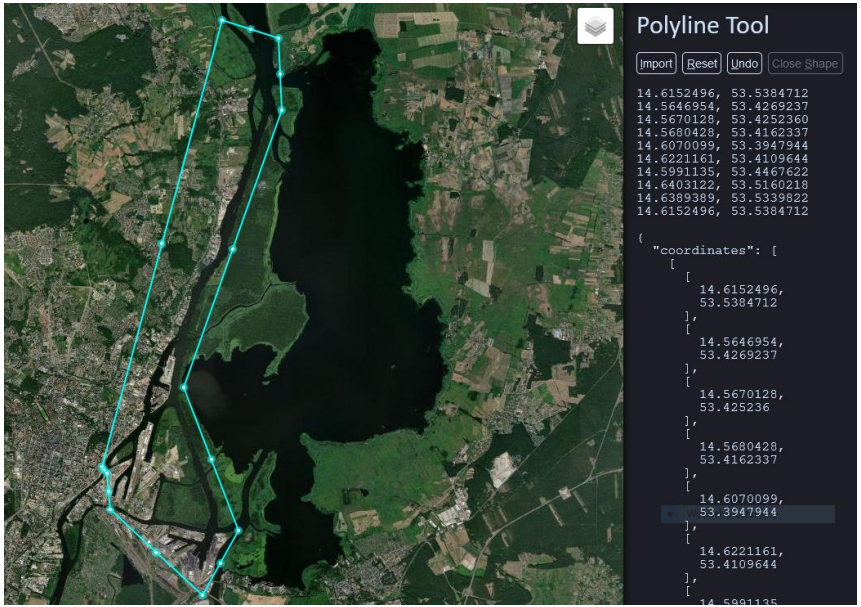


Figure 4. The ports of Świnoujścicie, Szczecin

Source: own study, generated on the basis of the tool:
https://www.keene.edu/campus/maps/tool/

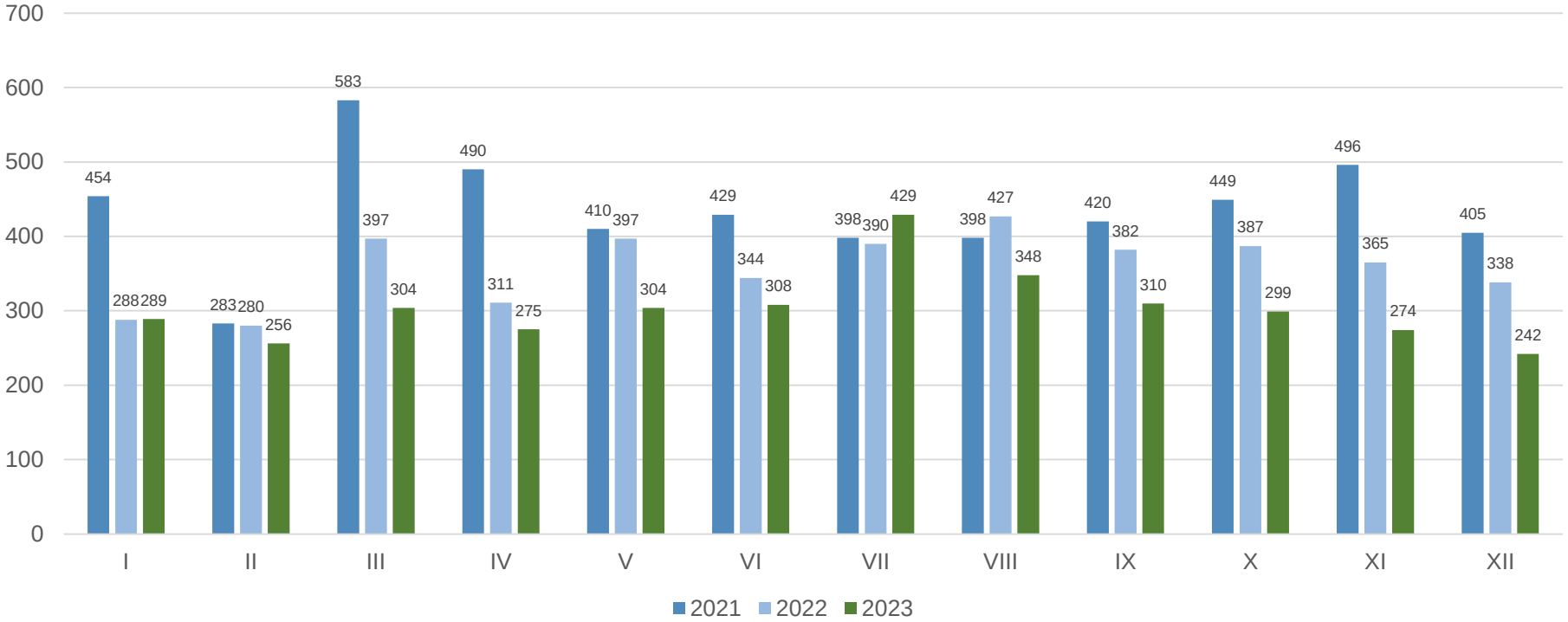
Statistics of traffic intensity in maritime transport – assumptions

As a result of the developed algorithms for the traffic intensity, the following variables and breakdowns are obtained, among others:

- **variables:**
  - number of ships at the port
  - number of calls by ships;
- **breakdowns:**
  - time: day, month, quarter, year
  - location: ports of Gdańsk, Gdynia, Szczecin, Świnoujście
  - means of maritime transportation: by type of ships, by country of flag

Statistics of traffic intensity in maritime transport – results

Number of calls of ships to the port of Szczecin by month in 2021-2023



Source: own study based on the results from the TranStat system

Statistics of traffic intensity in maritime transport

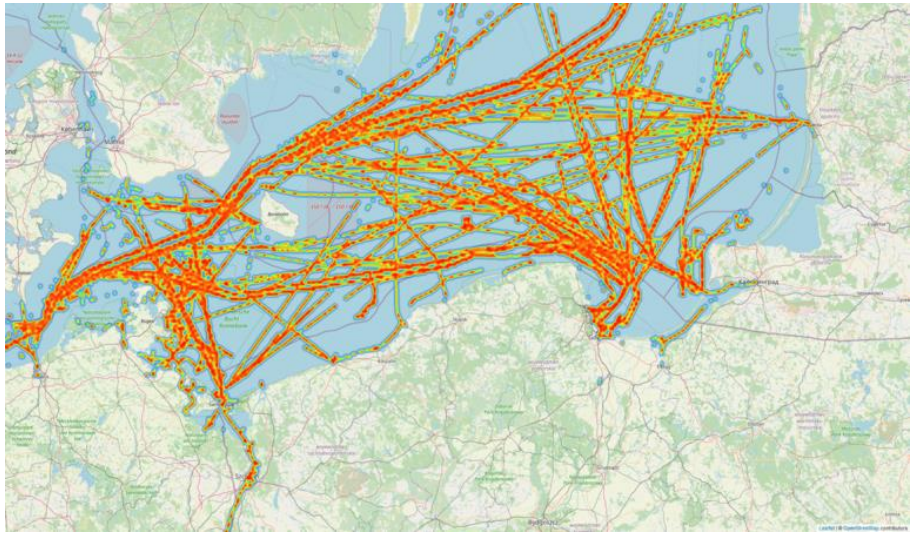


Figure 5. Traffic intensity of vessels for Poland - as of June 1, 2024.

Figure 6. Traffic intensity of vessels for port of Świnoujście - as of June 1, 2024.

Source: own study based on the results from the TranStat system

Statistics of emissions in marine transport

To estimate the amount of pollution emitted by means of sea transport, we analyzed ships with gross tonnage of GT 100 and more for ports of Gdańsk, Gdynia, Szczecin, Świnoujście.

Data source:

- Automatic Identification System,
- Information on ship parameters (a solution based on machine learning algorithms has been prepared, which requires access only to AIS data).

Statistics of emissions in marine transport

The adopted solution is based on the developed models, i.e .:

- reference model (detailed) - **REFERENCE**: requires the preparation of a matrix of characteristic technical parameters dedicated to the ship, enabling the determination of the value of individual emissions.

- specific model - **SPECIFIC**: using machine learning on a representative set of data from the REFERENCE model. The input parameters are the basic parameters of AIS messages and the output parameters are the emission values.

- generic model - **GENERIC**: used when the specific model obtains limit values or the input data are outside the accepted range, e.g. ship length over 300 m

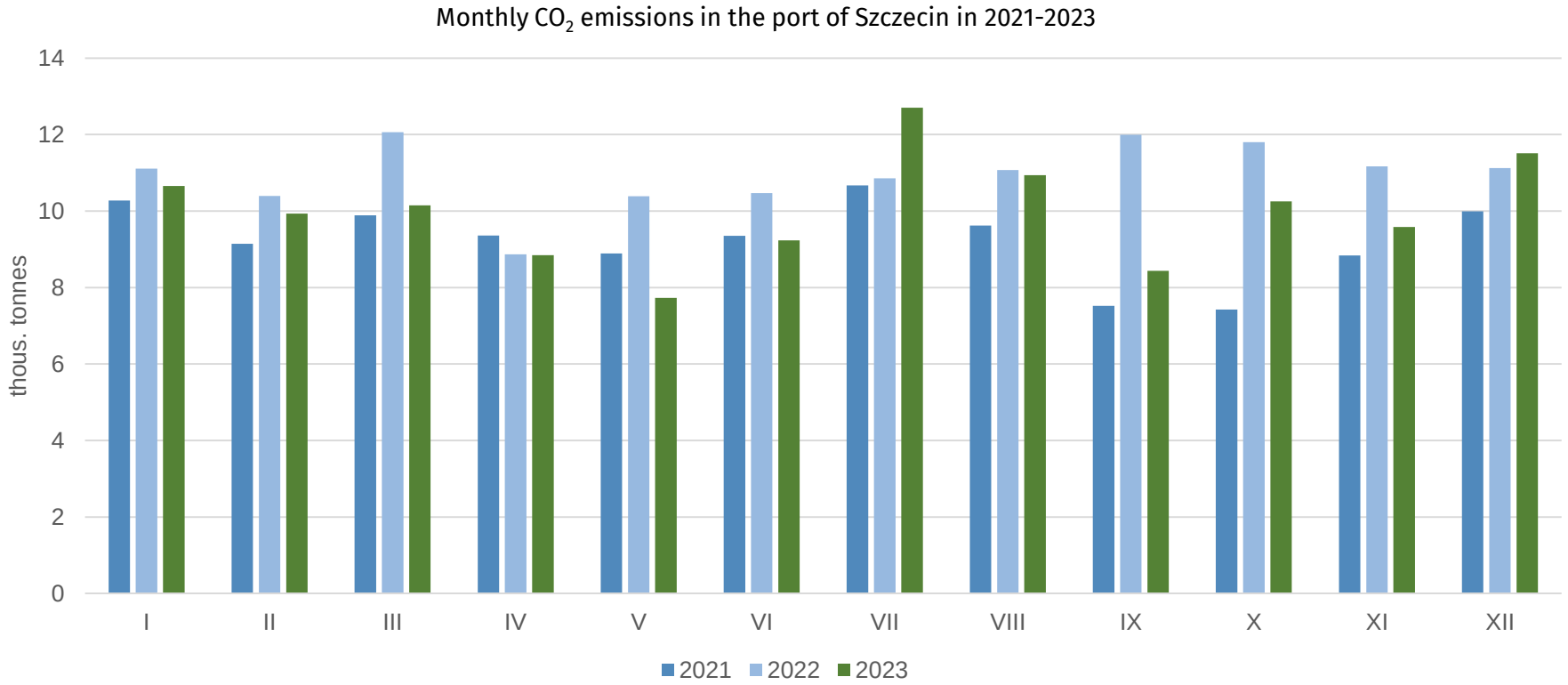
Statistics of emissions in marine transport

As a result of the developed algorithms for the emissions we obtained following variables and breakdowns (among others):

- **variables:**
  - NO_X emissions (nitrates, nitrites),
  - SO_X emissions (sulphates, sulphites),
  - CO_2 emission (carbon dioxide),
  - PM (particulate matter) emissions.

- **breakdowns:**
  - time: day, month, quarter, year
  - spatial - ports located on the Polish coast
  - ship types: means of sea transport by types

Statistics of emissions in marine transport – results



Source: own study based on the results from the TranStat system

Statistics of transportation volume in maritime transport

Tonne-kilometre (**tkm**) is the unit of measure representing the transport of one tonne of cargo in a ship over one kilometre.

Passenger-kilometre (**pkm**) is the unit of measure representing the transport of one passenger in a ship over one kilometre.

We need to know about : the amount of cargo (loaded/unloaded) and the ship's route.

Implementation in TranStat application:

- **Location:** ports of Gdańsk, Gdynia, Szczecin, Świnoujście.
- **Data source:** Automatic Identification System (AIS), Maritime transport data set based on Directive 2009/42/EC of the European Parliament and of the Council of 6 May 2009 on statistical returns in respect of carriage of goods and passenger by sea.

Statistics of transportation volume in maritime transport

The transportation volume estimation model implements the presentation of possible ship routes in the form of a directed (weighted) graph, where the vertices of the graph are navigation points or quays, and the edges are straight sections between them.

Each edge contains the coordinates of the start and end points, and the weight is the distance between individual nodes, calculated by the Haversine formula.

The graph consists of **9 859** vertices covering the entire globe.

There are **10 731** connections between the vertices.

There are **3 564 ports** included in the graph.

The sum of the weights of the edges of the graph is **1 088 864 km**.

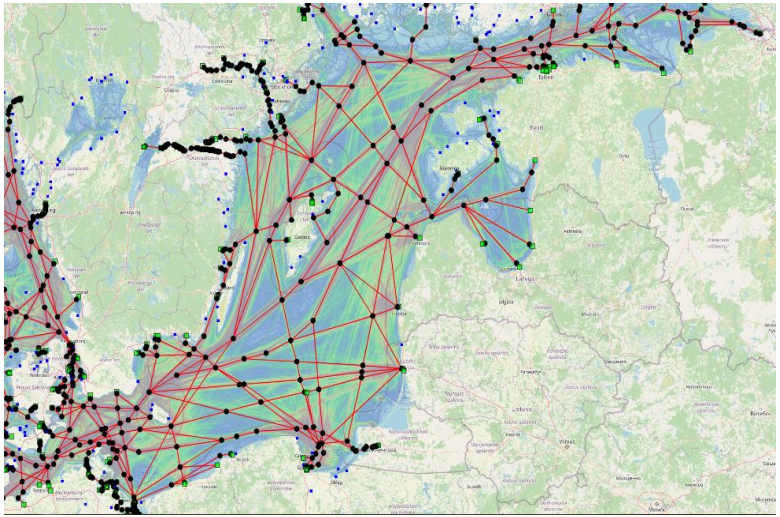


Figure 7. *Graph visualization for the Baltic Sea*

Source: Maritime University of Szczecin

Statistics of transportation volume in maritime transport

Implementation of port distance estimation based on directed graph:

1. determining the weights of the edges of a graph - **the Haversine formula**.

2. finding the shortest path in a graph- **the Dijkstra's algorithm**

A - start vertex , C - end vertex

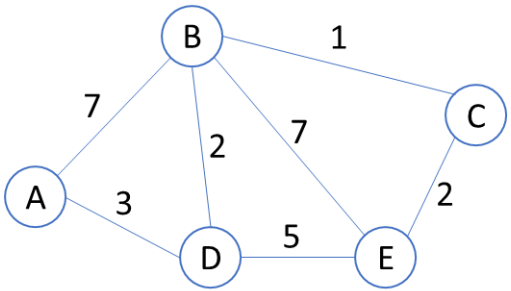


Figure 8. Graph with the representation of weights

Calculated shortest path from A to C is 6
and goes through vertices A, D, B and C

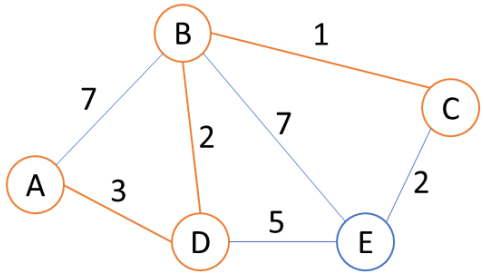


Figure 9. A graph with a representation of the shortest path

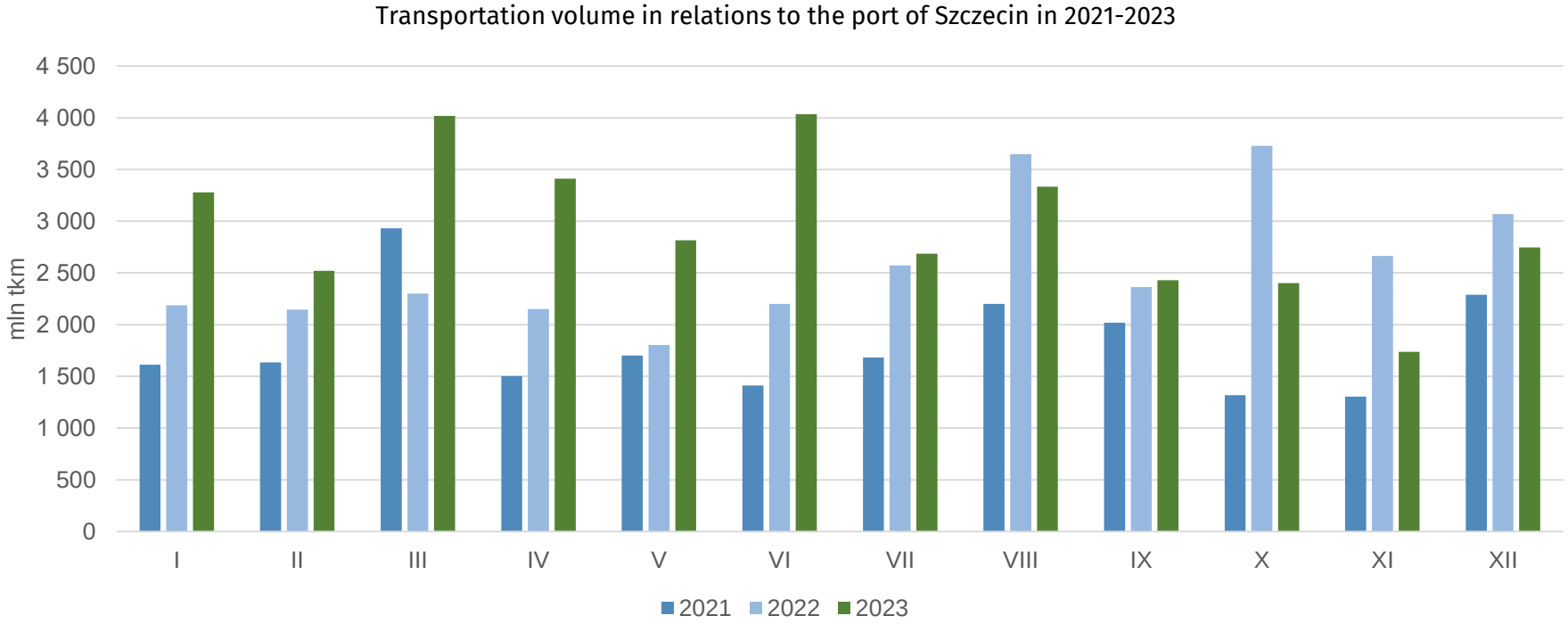
Statistics of transportation volume in maritime transport

As a result of the developed algorithms for the transportation volume we obtained following variables and breakdowns (among others):

- **variables:**
  - transportation volume for cargo and passengers,
  - average transport distance for 1 tonne of cargo in kilometers
  - average transport distance for 1 passenger in kilometers

- **breakdowns:**
  - time: day, month, quarter, year,
  - location: ports of Gdańsk, Gdynia, Szczecin, Świnoujście
  - means of maritime transportation - by type, by flag, by gross tonnage,
  - type of cargo – cargo group, commodity group.

Statistics of transportation volume in marine transport - results



Source: own study based on the results from the TransMor system

Conclusions

The implementation of the TranStat project in the field of maritime statistics has enriched the current statistical production carried out by Statistics Poland through:

- access to streaming Big Data source related to maritime transport (AIS);
- implementation of the necessary Big Data technology for sensory data enabling an automatic process of data flow, validation and processing;
- development of traffic intensity, transportation volume and emissions models in maritime transport with the use of sensory data;
- providing high-quality statistics in a short time (experimental statistics);
- reduction of research costs thanks to the use of modern technology in the collection and processing of non-statistical sources (AIS).

References

Rozkrut, D., Bilska, A., Bis, M., Pawłowska, J. (2023). TranStat: an intelligent system for producing road and maritime transport statistics using big data sources. *Wiadomości Statystyczne. The Polish Statistician, 68*(12), 1–24. https://doi.org/10.59139/ws.2023.12.1.

https://ws.stat.gov.pl/WS/2023/12/gus_ws_2023_12_dominik_rozkrut_et_al._transtat_an_intelligent_system_for_producing_road.pdf?v=180124

Statistics Poland

Thank you!