Analysing ship NOx emissions with Hadoop A Case Study for the Port of Rotterdam

Dr. Ir. Frank Cremer
Geomatik
The Hague, The Netherlands
frank@geomatik.nl

Muktha Muralee, PDEng
MFasize
The Hague, The Netherlands
m.muralee@mfasize.com

***Abstract*— Atmospheric emissions such as NO_x from ship engines have a drastic impact on the environment. Controlling them is crucial for maintaining a sustainable growth for any logistics company. The Port of Rotterdam (The Netherlands) is using big data analytics to gain actionable insights into these emissions. Our case study deals with the implementation of the emission calculations and reporting implemented in Hadoop. In the analytical setup we introduce the method for estimating emissions based on recorded ship position data and information about its engines. We present a flexible approach that stores intermediate results allowing different levels of aggregation. These levels of aggregations are per geographical area, per grid or for a whole journey attributed to each visited berth. The results are visualized in a Geographical Information System (GIS). The estimated atmospheric emissions also serves as input for the deposition model. We present some selected results of emissions per grid as well as for pre-defined areas. These results are used by the port to make strategic decision. For future work we recommend to also implement the deposition model in Hadoop as this model is also calculative intensive and therefore it currently only accepts aggregated emissions as input, whereby its accuracy is most likely reduced.**

Keywords—atmospheric emissions, emission model, GIS, Hadoop, Mapreduce

I. Introduction

Big Data solutions like Hadoop can provide insights that were previously not possible. This has been demonstrated in many case studies related to the Internet before. However, in this case study for the Port of Rotterdam we describe how Hadoop has been used to extract actionable analytical information from the dataset of aerial emissions from ship traffic. First, we explain why this study is relevant to the Port of Rotterdam. Next, we introduce the analytical setup used to model the emission per ship, based on each individual ship characteristics, its actual route. We then describe how it is aggregated into actionable results and visualized. Then, we show the implementation in software and discuss some selected results. At the end of the paper we present the conclusions and options for future work.

II. Impact of ship emissions

The Port of Rotterdam is located near the metropolitan city of Rotterdam and is close to protected and sensitive nature areas. Since it is one of the ten largest ports in the world, it has a very intensive ship traffic. This ship traffic is the source of large amounts of atmospheric emissions of NO_x, CO_2, SO_x and particulate matter (PM) [1], even though less than other sources (such as road traffic and agriculture).

For the Port of Rotterdam authority, it is a necessity to manage the ship emissions in order to be able to grow whilst reducing the footprint on the environment. The emission per terminal is restricted by its environmental license. Hence, it is imperative that the Port Authority have information about the emission attributed to each terminal. Only then operational decisions can be made to reduce the emission and/or incentives be given to less polluting ships.

Up until now the emission was calculated on a large grid, by RIVM, a government body. The results were not at a fine-grained level to be useful for determining emissions per terminal. Hence, the Port decided to implement their own model based on their dataset of ship positions, which was stored in Hadoop. The following sections of this paper show the implementation of the emission model in Hadoop and its results.

III. Analytical SETUP

In this section, we describe the analytical model setup to determine the ship emissions and to provide aggregated results that can be visualized and used as input for other models, for example, the deposition model. The first part deals with the emission model, the second part describes the stored ship position data. How data is aggregated is described in the third part and the visualization is described in the last part.

An overview of the processing from ship position data to storing intermediate results for further aggregation and visualization is shown in Fig. 1. Examples of the input and output data are shown for illustration. There are three steps involved: (1) creating tracks from individual position updates, (2) matching these tracks with predefined areas and (3) calculating the emission and storing these for these areas and separately for segments (two consecutive points).

IJRS

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

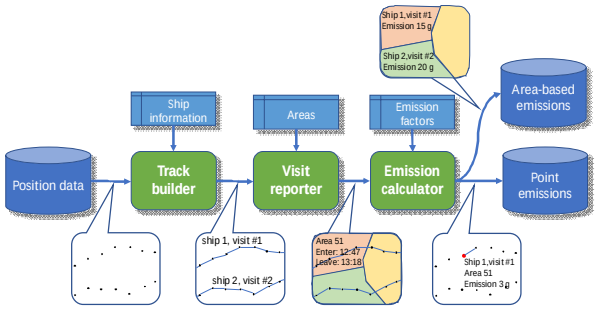


Fig. 1: Overview of the processing steps.

A. *Emission model*

The emission model is a model that estimates the various emissions into the air from sea ships. It is modelled, as opposed to measured, since as of date ships do not have sensors for measuring it in real time.[1]

The model uses as input the speed, the duration and the navigational status of the ship. The navigational status of a ship can be three states:

- Cruising at sea
- Maneuvering in the port
- Anchored or at a berth

For each state, different configurations of the main engine and auxiliary engines are used. The speed of the ship determines how many engines are used (if there is more than one) and to what percentage each engine is used.

The emission from the main engine during cruising at sea and maneuvering is given by the following formula:

$$E_{ME,subst} = EF_{subst}(n_{eng}) * t * CRS(v) * CEF_{subst}(v), \quad (1)$$

with

- $E_{ME,subst}(n_{eng})$ – the emission (in g), which depends on n_{eng}, the number of engines in use (-)
- t – the time (in s)
- v – the speed (in m/s)
- CRS(v) – the speed correction factor (-):

$$CRS(v) = ((v/v_{cruise})^3 + 0.2) / (1 + 0.2) \quad (2)$$

- $CEF_{subst}(v)$ – the substance correction factor (-)

When the ship is moored at a berth or at anchor, the main engine is considered to be off, so

$$E_{ME,subst} = 0 \quad (3)$$

For the auxiliary engine, the formulas are more straightforward:

$$E_{AE,subst} = EF_{AE,subst,<state>} * t, \quad (4)$$

with

- $E_{AE,subst}$ – the emission for the auxiliary engine (in g)
- $EF_{AE,subst,<state>}$ – the emission factor for the auxiliary engine, given for the states “anchor”, “berth”, “sea” and “maneuver” (in g/s)

The emission factors for each individual ship has been provided by TNO (National research institute). TNO has determined these factors based on the ship data (such as engine, cruise speed, gross tonnage) obtained from the Lloyds register of ships.

B. *Ship position data*

The Port maintains a continuously updated dataset of all the ship traffic in its area. This data set is used operationally for traffic management and stored in Hadoop for traffic analyses. The data set consists of the ship position along with another 60 parameters (such as ship name, identifier, length, speed, course, etc.) and is stored every 10 seconds. Given an average number of ships in the range of a thousand ships, this gives a total of about 10 million records per day. It is obvious that processing these many records (especially for a whole year) is way beyond the reach of traditional solutions.

Based on the emission model (as described in part A) and the ship positions, the emission for the ship between any subsequent 10 seconds is calculated. The distances and speeds are calculated based on the coordinates of the subsequent time intervals.

These emission results are stored back into Hadoop. With a 10 second interval for measurements, we have created a very detailed emission model. However, this model is too fine-grained to provide any insights, hence we need to aggregate the data as described in the next paragraph.

C. *Aggregating results (port/area/berth visits)*

There are four different selections within the emission data set that are of importance to the Port:

- Total emission from activities within the Port.
- The emission attributed to the individual terminals, e.g. container terminals, refineries, etc.
- The emission per administrative area, e.g. a port basin.
- The emission per visit per ship. A visit is defined as a ship entering the port’s operational area at 60 km from the entrance of the port, visiting one or

IJRS

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

more terminals and then leaving outside of the 60 km from the entrance.

An overview of the administrative areas is shown in Fig. 2. These areas range from 60 km out in sea (with designated anchor areas) to the port area itself (at a length of more than 40 km). In Fig. 3 a selected area of the port is shown. In this figure the port basins are visible (side arms of the river) as well the berths at the terminals. For illustration, a track of ship – actually its individual positions – is shown. This ship enters the port from the sea, sails up to the river and into the port basin called Maasvlakte to moor at the berth of terminal. After unloading and loading (not visible here) the ship leaves again the port to set sail for (presumably) the next port.

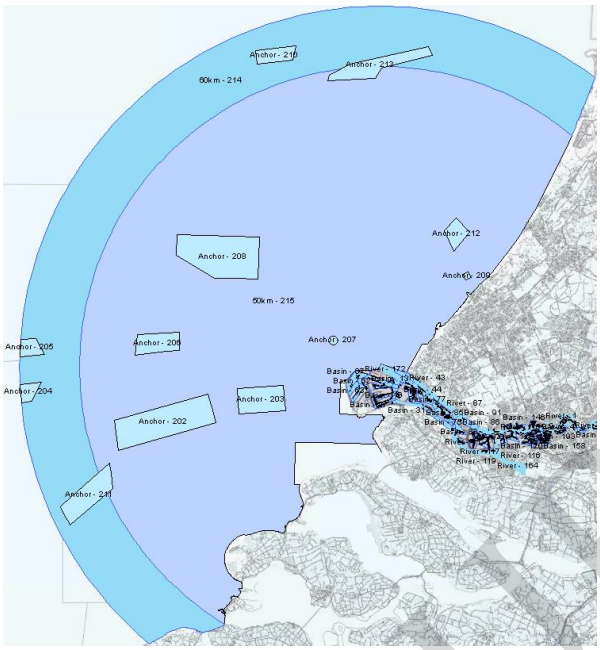


Fig. 2 Overview of the administrative areas

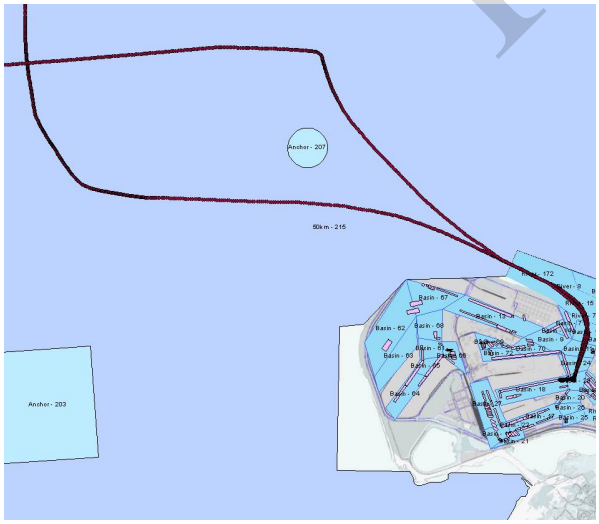


Fig. 3: Selection of the river, port basin and terminal area, with the track of a ship.

For each of these selections, the emissions can be aggregated in different ways:

- A uniform grid.
- The grid used by RIVM for yearly country-wide emission reporting.
- A grid which is finer (i.e. has smaller cells) closer to sensitive areas.
- Emissions per administrative area.
- Emissions for complete port visits attributed to specific berths belonging to the terminals.

Due to these many combinations of selections and aggregations, a flexible computational storage structure is necessary. The following intermediate data is stored:

- Visit information, containing the following:
  - Unique identifier for the ship
  - Unique visit identifier
  - Per area:
    - Area type (e.g. port area, anchor area, port basin and berth)
    - Area identifier
    - Timestamp of entry
    - Timestamp of exit
- Segment emission data (10 second interval), containing the following:
  - Timestamp
  - Unique ship identifier
  - Visit identifier
  - Navigational status (anchor/berth, cruising at sea or maneuvering)
  - Emission for the main and auxiliary engines for the eight different emission substances.
  - Funnel height.
  - Warmth output

Using these two data sets, it's possible to query the data to obtain the required combinations of selections and aggregations. For instance, the emissions of all the ships visiting a specific terminal is calculated in two steps. First, all the visit identifiers are determined by selecting the visits that have as type "berth" and as area identifier the required berth identifier. Second, a selection is made of all segment emissions which is aggregated onto the required grid.

IJRS

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

D. Visualisation

Insights into emissions is not only obtained by the numerical values of the emission estimates, but also through visualization. Obviously, any statistics can be put into graphs, e.g. comparing terminals or the same terminal over several years. However, in this paper, we limit the results to how we visualize the data set geographically.

Although the individual segment emission can be easily displayed geographically, there are too many points to provide any insights. Therefore, the visualization uses the emission per grid cell. The output of the aggregation is a set of 8 different emissions per grid cell. These grid cells are converted into a geographical database and visualized into a geographical information system (GIS).

E. Input for deposition model

The output of our model, i.e. the emission per grid cell, can be used for the deposition model AERIUS [2]. This model calculates the deposition for the substance NO_x, i.e. where it falls on the ground. The deposition depends on the height of the funnel (chimney) of the ship and the warmth output of the engine. The higher the funnel and the greater the warmth output, the further the NO_x will be deposited. Nature areas are especially sensitive to NO_x as it acts as a fertilizer affecting the natural vegetation.

Ideally, the aggregation should be done separately for each combination of funnel height and warmth output, to serve as input to the deposition model. However, this leads to more input points than AERIUS can handle. To reduce the number of points for each grid cell, 9 different combinations of funnel height and warmth output are used (3 for each parameter).

IV. Software Implementation

The analytical setup as described in the previous section has been implemented using various software components. In this section, we describe how it is implemented.

The software implementation is divided into three parts: (1) the analytics for modelling the emissions and determining ship visits, (2) the selection and aggregation of these results and (3) the visualization of the results. For our work, we have used the following components:

- Hadoop filesystem for storage of (1) and (2)
- Custom developed MapReduce [3] Java code for (1) and in part for (2)
- Hive queries for (2)
- FME for converting the results from (2) in a geographical format.
- ArcGIS for geographical visualization of the results (3).

The custom developed Java code has been developed using common IT standards. Especially for the emissions, model test cases have been created and the results have been validated externally by TNO.

V. Selected Results

In this section, some selected results are shown to demonstrate the insights into the emission calculations. Fig. 4 gives an overview of the total NO_x emissions per grid cell for all ships visiting the port (as the actual numbers are not relevant for this paper, the scale has been omitted). Although the emission is not normalized by the grid size, travel patterns can clearly be identified.

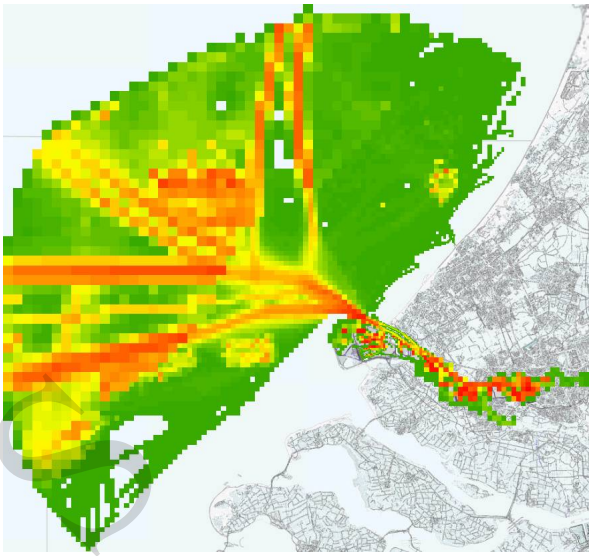


Fig. 4: Overview of the NO_x emission from ship traffic that visits the port. The colour scale ranges from green at low emission, via yellow to red at high emissions.

Fig. 5 gives similar results for ships that do not visit the port from seaside (there is some residual traffic between the port and the neighboring ports in Dordrecht and Moerdijk). In this graph, there is clearly ship traffic passing along Rotterdam at seaside (please note that the scales are not the same as in Fig. 4).

IJRS

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

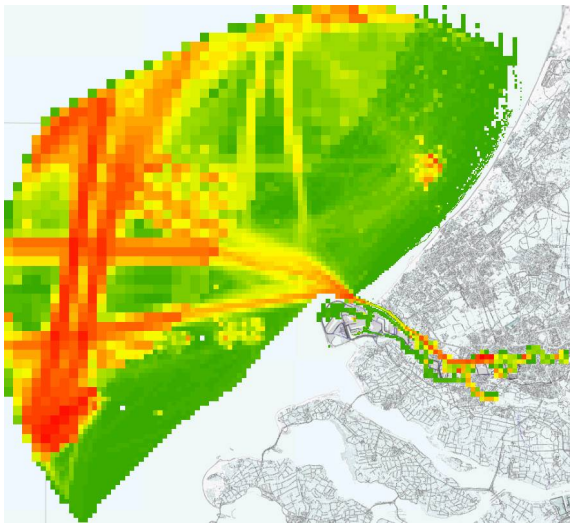


Fig. 5: Overview of the NO_x emission from ship traffic that does not visit the port. The colour scale ranges from green at low emission, via yellow to red at high emissions.

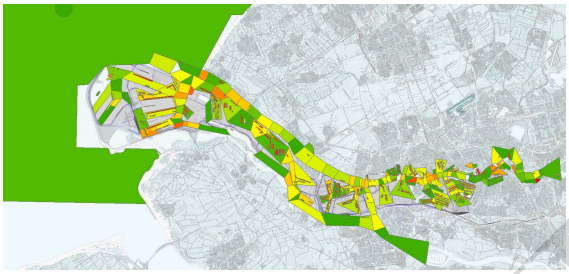


Fig. 6: The emission per area normalised to area size.

Fig. 6 Shows the emission per administrative area. The emission is normalized by its area to provide emissions in kg per m^2. Obviously, these normalized emissions appear to be high for berths at terminals as ship need to spend some time at the same location for loading and unloading. The emission model allows to estimate the savings in emissions when onshore power supply is used, so that ships do not need to run their auxiliary engines while being moored at the berth.

VI. Conclusions

With the approach followed in this paper, we have demonstrated that actionable insights can be gained by deploying Big Data methodologies for emissions calculations for the Port of Rotterdam. Specifically, the Port now has insights in the emissions per terminal, the emission close to sensitive areas, and the emission per ship per visit to the operational area of the port. A fine-grained emission model has been deemed essential for providing these results.

Based on these results actions can be defined to reduce the ship emissions while increasing the throughput of cargo, such as providing electricity for moored ships and/or providing incentives (e.g. by reducing the mooring fees) for cleaner ships and/or enforcing change or routes to diminish the impact on sensitive areas.

VII. Future Work and Recommendations

The results obtained using this methodology are more accurate than has previously accomplished by less fine-grained models, due to (1) better position estimates (2) shorter time intervals and (3) using emission factors for individual ships. Therefore, these results could be used for the Dutch National report on emissions. Providing these inputs for Rotterdam area should give a better estimate country wide. Moreover this methodology could be used for other Dutch areas as well, e.g. the Port of Amsterdam and the North Sea even further improving the estimates. Obviously, the same approach can be used for other countries.

A recommendation for improvement is for AERIUS to be implemented to run on Hadoop, thereby lifting the restriction for the number of input points. As the deposition model is additive, i.e. the total deposition is the sum of the deposition of each point emission, it should be able to parallelize the computations.

Acknowledgment

We would like to acknowledge the Port of Rotterdam for the work carried out in this project. We would also like to state on record our appreciation to TNO for providing the emission model and the ship-specific emission factors.

References

[1] H. Denier van der Gon, J. Hulskotte, “Methodologies for estimating shipping emissions in the Netherlands”, The Netherlands Research Program on Particulate Matter, Report 500099012, ISSN: 1875-2322 (print), ISSN: 1875-2314 (on line), 2010

[2] J. Duyzer, P. Zandveld, W. Lohman, "Doelmatigheidsonderzoek AERIUS Calculator (betaversie 8) en Monitor (versie 2014)", TNO Rapport 2015 R10211, 2015

[3] J. Dean and S. Ghemawat,"MapReduce: Simplified Data Processing on Large Clusters", OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)