

Mekelweg 2
2628 CD Delft
the Netherlands
Phone +31 (0)15-2782889
Fax +31 (0)15-2781397
www.mtt.tudelft.nl

Specialization: Transport Engineering and Logistics

Report number: 2015.TEL.7965

Title: **Modelling route choice of inland
shipping vessels in the port of
Rotterdam**

Author: E.J. Gerritse

Title (in Dutch) Modelleren van routekeuze van binnenvaartschepen in de haven van Rotterdam

Assignment: Masters thesis

Confidential: no
Initiator (university): Prof.dr.ir. G. Lodewijks

Initiator (company): Ir. M. Ruijgers (Witteveen+Bos, Deventer)

Supervisor: MSc. H. Zheng
2nd supervisor: Dr. ir. R.R. Negenborn

Date: September 17th, 2015

This report consists of 65 pages and 4 appendices. It may only be reproduced literally and as a whole. For commercial purposes only with written authorization of Delft University of Technology. Requests for consult are only taken into consideration under the condition that the applicant denies all legal rights on liabilities concerning the contents of the advice.

Mekelweg 2
2628 CD Delft
the Netherlands
Phone +31 (0)15-2782889
Fax +31 (0)15-2781397
www.mtt.tudelft.nl

Student: Erik Gerritse
Supervisor (TUD): Huarong Zheng
Supervisor (TUD): Rudy Negenborn
Specialization: TEL
Creditpoints (EC): 35

Assignment type: MSc graduation project
Report number: 2015.TL.7965

Confidential: no

Subject: Modelling route choice of inland shipping vessels in the Port of Rotterdam

In this project you will work at the engineering and consulting firm Witteveen + Bos and the Port of Rotterdam Authority (Havenbedrijf Rotterdam) to investigate route choice in the Port of Rotterdam. Witteveen + Bos is working on the development of a traffic model of the port of Rotterdam. Based on economic forecasts, the model translates goods flows to the required number of ships, how many ships are sailing from A to B, and finally the route of the ships, so that traffic volumes can be determined.

In road traffic, route choice is often modelled based on travel time. However, the inland shipping sector is different in many respects from road traffic (for example, with regard to traffic volume, homogeneity, congestion, interaction, manoeuvrability, etc.). This probably also applies to the choice of route, where travel time may be less dominant than in road traffic. To model the traffic in the port of Rotterdam an understanding of the drivers of route choice is necessary.

The study consists of two parts:

1. Determine what the main drivers for route choice are for inland shipping vessels;
2. Create a model for route choice in the port of Rotterdam based on the findings from part one.

Possible drivers are: time, distance, flow, bridge heights, interactions with other vessels, and so on. Based on AIS and radar data the current route choices are known, but the reasons for those choices are not. That means we know how the traffic is currently divided over the routes, but not why. To model the effect of future developments on route choice this insight is necessary.

The network is clear: there are eleven origins (and destinations) and depending on the origin-destination relation, one to four routes between them. The results of 1 part will determine the complexity of the necessary modelling. The starting point is that the current model is extended with an improved component for route selection.

You will need to report your results and recommend future research opportunities and applications in a scientific way. Your report must be written in English and must comply with the guidelines of the section. Details can be found on the website. For more information, contact Huarong Zheng or Rudy Negenborn.

The supervisors and professor,

Msc. H. Zheng, Dr. ir. R.R. Negenborn, Prof. dr. ir. G. Lodewijks



Abstract

The Port of Rotterdam is one of the busiest ports in the world and Europe's largest sea port. Each year, tens of thousands of ships visit the port. This number is expected to grow in the foreseeable future. In order to make sure the port remains accessible, meticulous planning is necessary. The creation of realistic scenarios for future traffic flows is a requirement for planning the future of the port area. Engineering company Witteveen + Bos has worked in close cooperation with the department of Capacity Management at the Rotterdam Port Authority to create scenarios for predicting traffic flows in the port area. Based on general predictions about future cargo volumes and modal split, a model of traffic flows in the port area is being created. Part of this model details the inland shipping sector, an important sector for hinterland transportation.

Currently the inland shipping traffic model does not incorporate a model for route choice. Given the fact that a vessel will travel from a certain location A to another location B, there is no model that predicts what route that vessel will take. Instead, historical data is used to predict future traffic flows. This report contains the results of research conducted with the goal of creating a model for route choice of inland shipping vessels. The following research question has been answered:

Main research question

How can route choice of inland shipping vessels, given the origin and destination of the vessel and network topography, be modelled?

Sub questions:

1. How can the Port of Rotterdam area be modelled?
2. How can the route choice process be modelled?
3. What factors have an observable influence on route choice of inland shipping vessels?
4. How can traffic data be used to calibrate a route choice model?
5. How does the modelled behaviour compare to traffic data from the Port of Rotterdam?
6. How can the route choice model be applied to the prediction of inland shipping traffic in the Port of Rotterdam?

Research method

Route choice for inland shipping vessels can be modelled with Multinomial Logit method. This method, developed in econometrics to study consumer decision making, is now widely used to study route choice for cars and pedestrians. The port of Rotterdam was modelled as a graph. This was used to create route choice sets for each origin-destination pair. A multinomial logit model was created to calculate the probability of choosing each route in a choice set.

In order to find out what the main reasons for choosing routes are, a survey was conducted. 76 inland shipping captains replied to a questionnaire about route choice in the port of Rotterdam area and in general. This led to a list of factors that influence route choice. These were used as factors in the logit model. The following factors were included:

- Distance
- Bridges
- Locks
- Wind
- Waves
- Narrow or shallow waterways

Traffic data from the port of Rotterdam was used to calibrate and test the model. All traffic from inland shipping vessels in the 1st quarter of 2014 was used to calibrate the model. Calibration was done according to the maximum likelihood principle. This calibration process quantified the influence on route choice of the various factors mentioned above.

The influence of vessel size and cargo type was investigated by calibrating the model not only for all route choices, but also for trips only made by vessels with a certain size or cargo type.

Results and conclusions

Using a multinomial logit model calibrated with data from the 2st quarter from 2014, the 2nd quarter of that year was predicted. This resulted in 95% of choices being predicted correctly.

The cargo flow predictions using the model were tested further by looking at data from a week in 2014 where an important waterway was closed. For this period the model also predicts route choice with 95% accuracy. Using historical data such predictions are not possible.

The models calibrated for specific vessel sizes and cargo types were also used to predict part of the traffic from the 2nd quarter of 2014. Predictions made using these models had a higher likelihood then predictions made using a general model.

In short, the multinomial logit model is proven to be capable of predicting route choices for inland shipping vessels in the port of Rotterdam accurately, even when the port area changes. Distance, presence of locks and bridges, wind and waves and narrow or shallow waterways are the most important reasons for preferring one route over another.

List of symbols and key formulas

This section lists all mathematical symbols used in the report, as well as the most important formulas.

Route description and choice set

\mathcal{U} = Universal choice set containing all paths between two nodes

\mathcal{M} = Set containing all paths between two nodes

\mathcal{C} = Choice set of feasible routes between two nodes

Γ_i = Set of links for route i

Λ_i = Set of nodes for route i

i = route number

a = link number

n = vessel number

$Z_{ij} := \begin{cases} 1 & \text{if } \Lambda_i \supset \Lambda_j \\ 0 & \text{if } \Lambda_j \not\supset \Lambda_i \end{cases}$ Dummy variable that determines whether or not a path from set \mathcal{M} is in the feasible choice set \mathcal{C} .

Utility function

$$U_i = V_i + PSC_{i,C} + \epsilon$$

U_i = total utility of route i

$V_i = \sum_{\alpha \in i} V_\alpha$ = Utility of route i

$PSC_{i,C}$ = Path-Size Correction factor

ϵ = GEV distributed random variable (scale 1, location 0)

Path-size Correction function

$$PSC_{i,C} = - \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \ln \sum_{j \in \mathcal{C}} \delta_{aj}$$

l_a = length of link a

L_i = length of path i

$$\delta_{aj} = \begin{cases} 1 & \text{if } a \in \Gamma_j \\ 0 & \text{if } a \notin \Gamma_j \end{cases}$$

$\sum_{j \in \mathcal{C}_n} \delta_{aj}$ = number of paths in choice set \mathcal{C} sharing link a .

Calculating likelihood

Likelihood (β) :

$$L(\beta) = \prod_{n=1}^n \prod_{i \in C(n)} P(i|C)_n^{g_{in}}$$

$$g_{in} = \begin{cases} 1 & \text{if route } i \text{ was chosen by ship } n \\ 0 & \text{otherwise} \end{cases}$$

$$l(\beta) = \log(L(\beta)) = \sum_{n=1}^n \sum_{i \in C_n} g_{in} \log(P(i|C)_n)$$

Contents

Abstract	1
Main research question.....	1
Sub questions:	1
Research method	1
Results and conclusions.....	2
List of symbols and key formulas	3
Route description and choice set	3
Utility function.....	3
Path-size Correction function.....	3
Calculating likelihood	4
1 Introduction.....	8
1.1 Inland shipping traffic forecasts	8
1.2 Problem definition.....	8
1.3 Research questions.....	9
2 Modelling the Port of Rotterdam area	10
2.1 Dividing the port into areas.....	10
2.2 Creating a graph to model the port areas and waterways	14
3 Techniques for modelling route choice.....	15
3.1 Selection of general modelling method	15
3.1.1 Discrete Choice Modelling.....	15
3.1.2 Utility Maximization and Regret Minimization paradigms.....	15
3.2 Logit models	15
3.2.1 Limitations of the MNL model.....	16
3.2.2 Adapted Logit models.....	17
3.2.3 Selecting a Logit modelling method	18
3.3 Modelling assumptions	20
3.4 Creating a Path-Size Logit model.....	20
3.4.1 Path generation and choice set formation.....	20
3.4.2 Path-size factor.....	22
3.4.3 Summary.....	23
4 Choosing parameters for the route choice model	25
4.1 Poll results	26

4.1.1	Influence of vessel characteristics on route choice	27
4.2	Investigating of possible drivers for route choice	27
4.2.1	Distance	28
4.2.2	Travel time.....	28
4.2.3	Narrow / shallow waterways.....	28
4.2.4	Tides (water levels).....	29
4.2.5	Bridges	29
4.2.6	Currents	31
4.2.7	Wind and waves	31
4.2.8	Locks	32
4.2.9	Berths	32
4.2.10	Speed limits	33
4.2.11	Congestion.....	33
4.3	Conclusion	34
4.3.1	Defining the utility function	34
5	Using traffic data to calibrate the route choice model	35
5.1	Traffic data collection by the port authority	35
5.2	Processing the data	36
5.2.1	Example	37
5.3	Python data analysis.....	37
5.4	Errors in traffic data.....	38
5.4.1	Example	39
5.4.2	Example 2	41
5.5	Error reduction	41
5.6	Summary.....	42
6	Calibration methods and performance indicators	44
6.1	Calibration method.....	44
6.1.1	Maximum likelihood.....	44
6.2	Assessing model performance	44
7	Calibration	46
7.1	Assigning a unique penalty to each obstacle	47
7.1.1	Distance	47
7.2	Values of $[\beta]$	48

8	Validation	51
8.1.1	Comparison between model and historical data	52
8.1.2	O-D pairs predicted with less than 90% accuracy	54
8.2	Testing predictive capabilities of the route choice model	56
9	Influence of cargo type and ship size on route choice	58
9.1.1	Difference in route choice preference per vessel or cargo type explained	60
9.1.2	Conclusion	60
10	Results and conclusions.....	61
10.1	Results	61
10.2	Conclusions.....	61
10.3	Answers to research questions	61
11	Suggestions for further research.....	63
11.1	Investigate applicability to other situations.....	63
11.2	Divide port map into smaller areas	63
11.3	Look into the whole visit to the port area, not just trips from 1 origin to 1 destination	63
12	References	64

1 Introduction

This report describes the creation, validation and use of a model for route choice of inland shipping vessels in the Port of Rotterdam. This route choice model was developed for the engineering company Witteveen+Bos in close cooperation with the Port of Rotterdam Authority.

1.1 Inland shipping traffic forecasts

The Port of Rotterdam Authority (Havenbedrijf Rotterdam, or HbR) employs a number of people at a department called Capacity Management. Their job is to make sure the infrastructure in the port area (waterways, roads, railways, pipes and power lines) has enough capacity to accommodate the traffic and cargo flows going through the port. In order to manage traffic on the waterways, HbR wants to forecast inland shipping traffic flows in each waterway.

In order to predict traffic flows of inland shipping vessels, a model called 'Binnenslag' is being developed by Witteveen+Bos. This model uses general forecasts for cargo flows in the port of Rotterdam as input and produces a forecast for inland shipping traffic in the waterways in the port area. In order to get from general cargo flow forecasts to detailed traffic forecasts, it uses the steps described in the diagram in Figure 1. Each step represents a more detailed look at the traffic in the port area than the last.



Figure 1 Steps in the Binnenslag model used to make a forecast about traffic in the port area

1.2 Problem definition

In order to know the amount of inland shipping traffic in each waterway (the final step in Figure 1), one would need to know that routes the inland shipping vessels will take to get from a certain point in the port to another location, given a potential future layout of the port area. Traffic data shows that inland shipping vessels sometimes take different routes to get from the same origin to the same destination, but sometimes they do not. An example of different possible routes is shown in Figure 2.

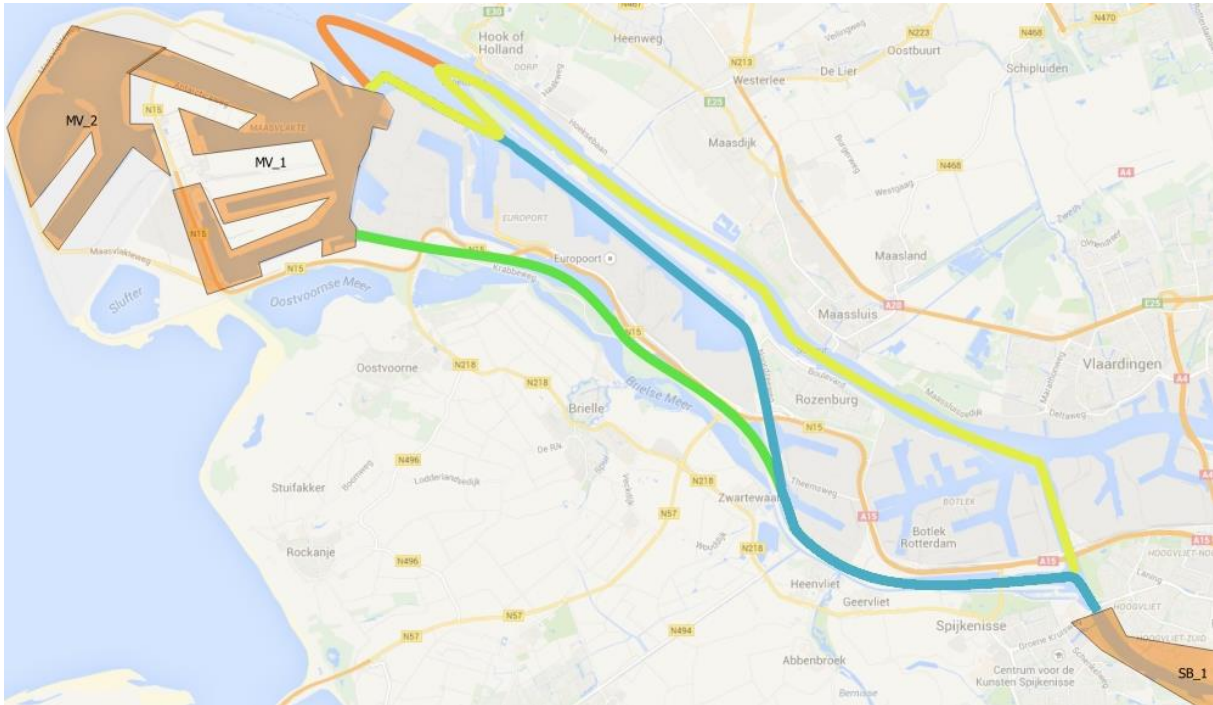


Figure 2 Example of different routes vessels can (and do) take from one place to another. Map adapted from Google Maps

Currently the Binnenslag model has no way to predict route choice for inland shipping vessels and assumes that traffic will be divided across the waterways in the same way it currently is. Changes in the layout of the port area cannot be taken into account, as there is no historical information upon which to base such a forecast.

In order to make forecasts about inland shipping traffic and assess if infrastructure capacity will be sufficient to handle future traffic flows, a model for route choice of inland shipping vessels is needed. This investigation will create and validate such a model and show how it can be applied to make forecasts about inland shipping traffic.

1.3 Research questions

Main research question:

How can route choice of inland shipping vessels, given the origin and destination of the vessel and network topography, be modelled?

Sub questions:

1. How can the Port of Rotterdam area be modelled?
2. How can the route choice process be modelled?
3. What factors have an observable influence on route choice of inland shipping vessels?
4. How can traffic data be used to calibrate a route choice model?
5. How does the modelled behaviour compare to traffic data from the Port of Rotterdam?
6. How can the route choice model be applied to the prediction of inland shipping traffic in the Port of Rotterdam?

2 Modelling the Port of Rotterdam area

In order to model route choice in the port of Rotterdam, we first need a model of the port itself. This chapter describes how the port area is analysed and modelled.

The map of the port (see Figure 3) is divided into a number of areas. Table 1 contains a list of these areas. The division into areas was made by the Port of Rotterdam Authority and adapted slightly for this investigation. As can be seen in Figure 3, these defined areas do not cover the entire map but only places deemed to be of interest by HbR.

Table 1 Areas in the Port of Rotterdam

	Area Code	Full name	Area in graph
1	MV_1	Maasvlakte 1	MV
2	MV_2	Maasvlakte 2	MV
3	NW_1	Nieuwe Waterweg	NWA
4	NW_2	Zee	ZEE
5	EU_1*	Europoort – Calandkanaal	EP_C / EP_BEN / EP_VII
6	EU_2	Hartelhaven	EP_H
7	EU_3	Beerhaven	EP_B
8	HA_1	Hartelkanaal	HART
9	CA_1	Rozenburgsesluis	ROOS
10	BT_1	Brittaniëhaven	BV_BRIT
11	BT_2	Seinehaven	BV_SEIN
12	BT_3	Botlek	BV
13	ZM_1	Zevenmanshaven	NO
14	KW_1	Koningin Wilhelminahaven	NO
15	WT_1	Wiltonhaven	NO
16	WM_1	Wilhelminahaven	NO
17	KL_1	Keilehaven / Lekhaven	NO
18	ST_1	Merwehaven	NO
19	ST_2	IJsselhaven	NO
20	VL_1	Vulcaanhaven	NO
21	MH_1	Maashaven	CITY
22	RH_1	Rijnhaven	CITY
23	PS_1	Parksluis	PARK
24	EH_1	Eemhaven	WEH
25	EH_2	Eemhaven	WEH
26	EH_3	Eemhaven	WEH
27	WH_1	Waalhaven	WEH
28	WH_2	Waalhaven	WEH
29	BB_1	Van Brienenoordbrug	BRIEN
30	SB_1	Spijkenissebrug	SPIJK
	**	Ligplaatsen Hartelkanaal – Oude Maas	HART_OM
	**	Ligplaatsen Noordereiland / Katendrecht	CITY

*This area is registered as one area, but was divided into three areas for this investigation.

**This area does not correspond to a location in Figure 3 but was defined especially for this investigation. It will probably be included in future HbR models.



Figure 3 Map of the port area, with the HaMIS areas in orange. Map adapted from Google Maps

Modelling route choice of inland shipping vessels in the port of Rotterdam
2015.TEL.7965

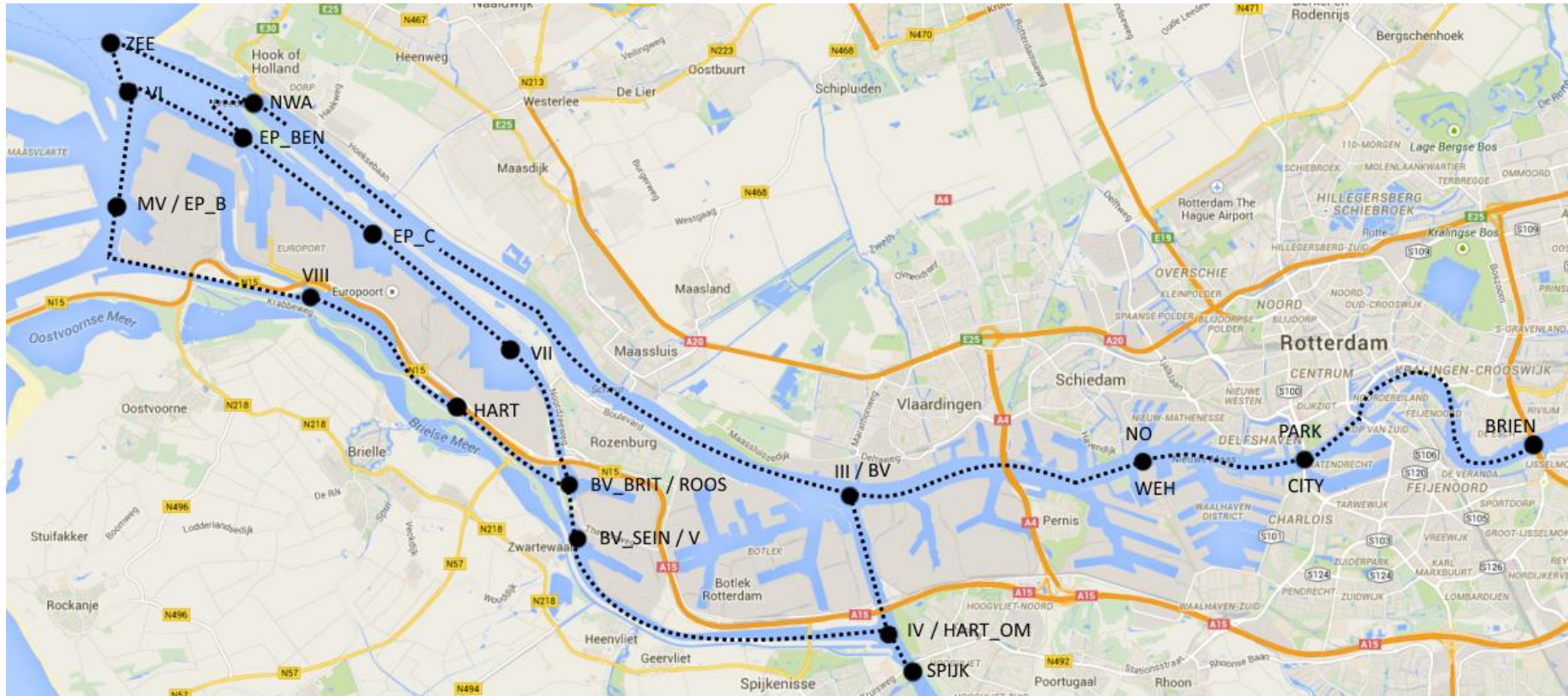


Figure 4 Map of the port with the assumed locations of areas and waterways in black. Map adapted from Google Maps

Modelling route choice of inland shipping vessels in the port of Rotterdam
2015.TEL.7965

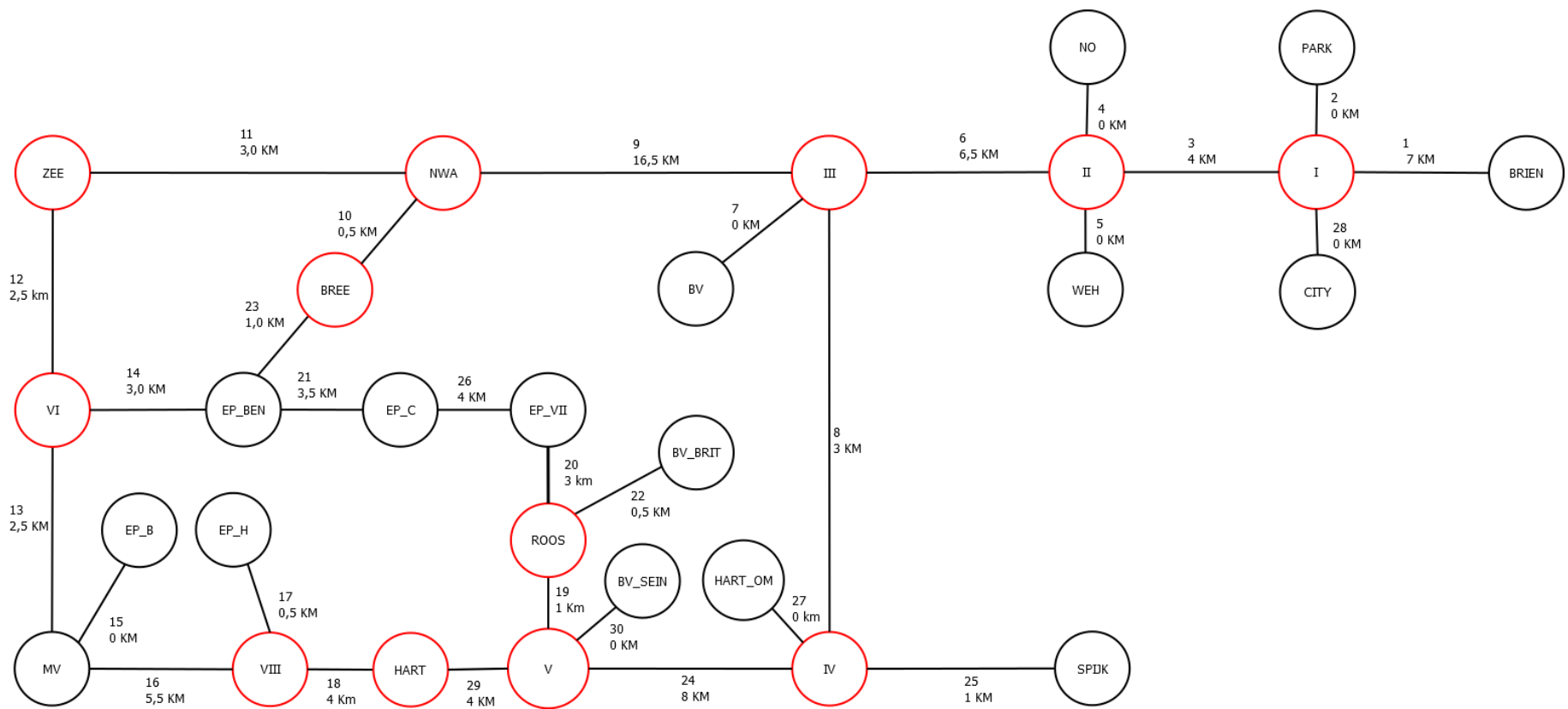


Figure 5 Graph of the port area corresponding to the map in Figure 4

2.1 Creating a graph to model the port areas and waterways

In order to model route choice, a representation of the area in which the origin and destination are located is necessary. Therefore, the port area needs to be modelled in order to study route choice. This is done using graph theory. Please see Figure 3 for a map of the port of Rotterdam, Figure 4 for a simplified version of that map and Figure 5 for the resulting graph.

A graph consists of links and nodes, where each link connects two nodes. The links have been drawn in such a way that they closely resemble the actual waterways. The black nodes represent port basins, which can be an origin or destination for inland shipping vessels. The red nodes represent important locations in the waterways.

The distances visible in Figure 5 were measured using Google Maps and rounded towards the nearest 500 m. Note that because of the size of some port areas, significant simplifications were necessary to compute the distances. For example, the whole northern bank of the Nieuwe Maas (see Figure 3) is represented by a single point in the graph (Figure 4). Therefore the true length of a trip between two areas can only be approximated with this model. This will undoubtedly impact the accuracy of the route choice model.

The areas that are used in the graph are almost the same as the ones used in the Binnenslag model, which makes the route choice model easily compatible with other work done by Witteveen+Bos and HbR. The only changes made were dividing the EP_C area into three parts and adding the areas CITY and HART_OM. Adding the two new areas was done to incorporate two locations where many berths are located, but that are not included in the current models used by the port authority.

The original area 'EP_C' was cut into separate pieces to investigate if this would improve the accuracy of the route choice model. This turned out to be the case. Therefore it is recommended to divide other large areas into smaller pieces as well. The EP_C (Caland channel) area was used to test this because it accommodates a lot of traffic and was thought to be the very inaccurately represented by just one point for its entire 16 KM length.

3 Techniques for modelling route choice

This section explains the criteria used to select a modelling technique, presents the different techniques for modelling route choice found in literature and presents the chosen modelling technique. It answers sub question 2: ‘How can the route choice process be modelled?’

3.1 Selection of general modelling method

A literature review reveals there are multiple ways of modelling route choice. All of them are members of the family of Discrete Choice models. Discrete choice models use assumptions about actor behaviour to calculate the chances different options are chosen. These calculations can be used to forecast route choice in future situations.

3.1.1 Discrete Choice Modelling

There are two types of modelling discrete choice processes: Logit and Probit. Solving probit models cannot be done directly but requires many iterative steps, making them computationally much more complex than Logit models [1]. For discrete choice problems, Logit models have been shown to produce more accurate results than Probit models [2]. For these two reasons, Logit models are used much more often. Therefore it is decided to use a Logit type of discrete choice model for the investigation into route choice.

3.1.2 Utility Maximization and Regret Minimization paradigms

All variants of the logit model can be based either on the paradigm of Utility Maximization (standard) or Regret Minimization (non-standard but potentially more accurate). Identical models have been shown to yield different results when UM and RM equations are used. Because at this moment no significant advantages of applying RM are known [3] it is proposed to use a utility maximization model.

In short, a discrete choice model is the best way to model the route choice process in the Port of Rotterdam. A Logit model based on the Utility Maximization paradigm is the most suitable type of discrete choice model.

3.2 Logit models

The standard Discrete Choice Logit model is the so-called Multinomial Logit (MNL) model. This model assumes the following:

- there are a finite number of options that can be picked. Together they are called the choice set C .
- It assumes the Independent of Irrelevant Alternatives (IIA) property, which means there is no unobserved correlation between the available options. All alternatives are assumed to be independent, and adding one or taking one away from the choice set does not influence the others.
- Each alternative has a so-called utility value (U), which is expressed as follows:

$$U_{i,n} = V_i + \epsilon \quad (1)$$

Where

$U_{i,n}$ = Utility of alternative i to individual n

V_i = Utility of the route i

ϵ = GEV random value for individual α , modelling true randomness and unobserved parameters

The probability of choosing a route i from the choice set C is then computed as follows:

$$P(i|C) = \frac{\exp(U_i)}{\sum_{j \in C} \exp(U_j)} \quad (2)$$

3.2.1 Limitations of the MNL model

MNL is not applicable to the route choice problem due to its inability to deal with routes that are very similar or have partial overlap [4] [5]. Because MNL assumes all alternatives are uncorrelated, overlapping or similar routes will be treated as completely different. The route choice problem thus violates the IIA assumption. An example of this follows below.

Consider the following network consisting of two nodes (A and B) and two links connecting them. The link utilities (U_1 and U_2) are the same.

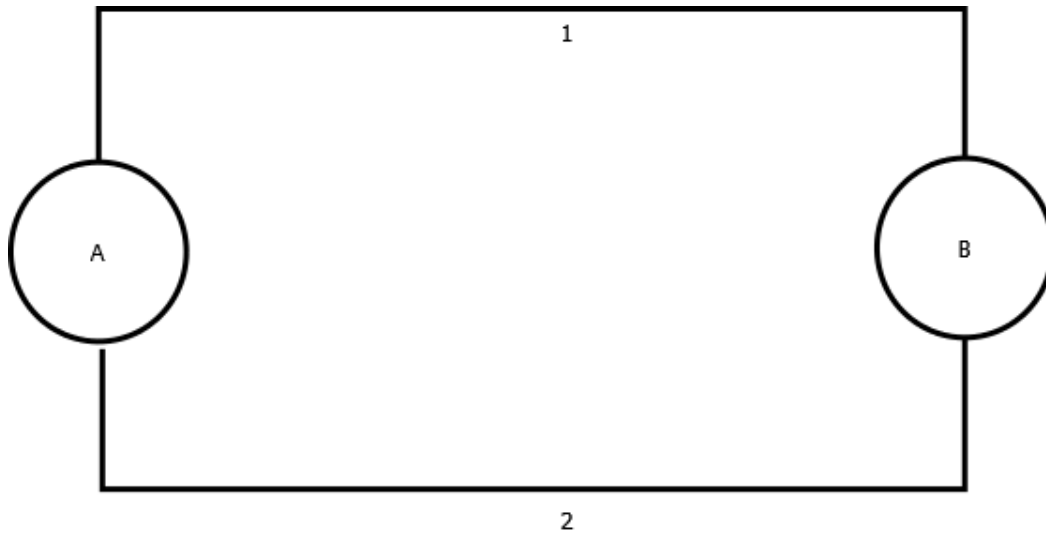


Figure 6 a network with two equal routes from A to B

Using a MNL model, the probabilities of choosing route 1 and route 2 from the choice set C are also the same:

$$P(1|C) = P(2|C) = 0.50 \quad (3)$$

Now imagine adding a small section to route 2 (see Figure 7). Route 2a and route 2b overlap for a large amount of their total distance and are equal in length to each other and to route 1.

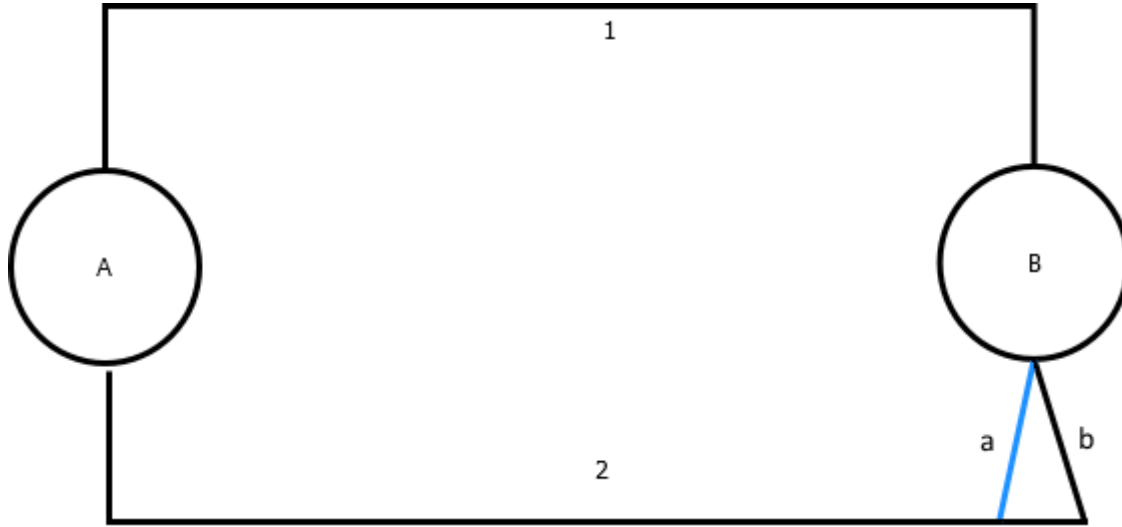


Figure 7 The new network from A to B with the addition of a nearly similar route (3)

Intuitively, we would expect the people that preferred route 1 previously still do so. The people that chose route 2 before will now choose either route 2a or route 2b, so the probabilities in reality will be as follows:

$$\begin{aligned} P(1|C^*) &= 0.50 \\ P(2|C^*) &= 0.50 \\ P(2_a|C^*) &= P(2_b|C^*) = 0.25 \end{aligned} \quad (4)$$

However, MNL models view each alternative on its own. This would lead to assigning equal probabilities to all alternatives, even if the length of segments a and b $\rightarrow 0$:

$$\begin{aligned} P(1|C^*) &= 0.\overline{33} \\ P(2|C^*) &= 0.\overline{66} \\ P(2_a|C^*) &= P(2_b|C^*) = 0.\overline{33} \end{aligned} \quad (5)$$

This answer is obviously not realistic. Route 2 and route 3 are almost equivalent, and someone who chose route 1 before is unlikely to make another decision now that the new route 2b is available. The only people who might take route 2b are people that would have taken route 2 before, and since routes 2a and 2b are equal in length the division of travellers across them will probably be 50/50.

In the port of Rotterdam area many routes have partial overlap (as route 2 and 3 do here). The inability to take this into account makes the standard MNL model unsuitable for analysing route choice in the port area. Therefore a variant of the Logit model has to be used that can deal with overlapping routes (or, more generally, correlated alternatives).

3.2.2 Adapted Logit models

Many adaptations of MN Logit models exist that are made specifically to deal with overlapping routes. A list of them compiled from academic literature follows below.

Mixed Logit is a field standard [6] that does not suffer from the limitations of regular MNL models. It is very flexible, does not suffer from the constraints of the MNL model. The integrals used in Mixed Logit models do not have a closed form solution so the probabilities for each route will have to be

approximated by solving algorithms, just like for probit models. This makes them computationally much more complex. Also, this makes the calculation results difficult to check by hand. This is important for debugging and testing of solution algorithms.

Nested Logit [6] [7] allows for routes that are similar to be grouped together ('nested'). This allows for modelling of correlation between alternatives. Not all types of nested logit are suitable for application to route choice problems, though some, like **C-Logit** and **Link-nested Logit**, are.

Path-size Logit is another variant of the MNL model that takes into account the effects of partially overlapping routes. It has been shown to be applicable to the route-choice problem by Ben-Akiva et al [8]. Tests prove that Path-Size Logit performs better than C-Logit and Link-nested Logit for route choice problems [1]. A Path-Size component will have to be computed for each O-D pair and route, but the equation used is quite straightforward and this process can be automated. Therefore it will not take a lot of time.

A number of improvements to the path-size logit model have been suggested. Many of those deal with model improvements that are not useful for studying traffic in the Port of Rotterdam. The so-called **Error Component model** proposed by Frejinger and Bierlaire in 2007 [9] significantly improves the predictions made with the model compared to the Path-Size Logit method, while adding complexity. The EC model introduces additional terms in the utility function to account for paths that are not overlapping, but still correlated. Computing the Error Component for each O-D pair will mean a significant amount of extra work. The situation in the Port of Rotterdam is not deemed sufficiently complex to justify increasing model complexity, since it does not seem to contain paths that are closely correlated but do not overlap. Therefore, the additional model complexity will probably not yield better predictions.

The **Recursive Logit** model with correction for link size as proposed by Fosgenau et al in 2013, though not yet tested in practice, is theoretically superior to path-size logit models [10]. However, RSL models are computationally much more complex and require sophisticated solving algorithms to estimate. Whether or not they produce more accurate forecasts is still uncertain, because both PSL and RSL models use a proxy for correlation between paths. Which proxy is better remains the subject of investigation and debate.

3.2.3 Selecting a Logit modelling method

One method will need to be selected from the different options. This will be done based on the following criteria.

- Setting up the model for analysis of different O-D pairs should be relatively easy. This is important because many origin-destination pairs (around 50) need to be analysed and parts of the setup need to be done by hand, as they cannot be automated.
- Models where results can be computed directly are preferred to models that have to be solved by iteration and/or approximation. Using complex computational methods will mean that the software used to evaluate the model will have to be much more complex as well. It will also mean that implementing the route choice model into the larger project for Witteveen+Bos and the Port Authority will be more difficult.

- The model should be able to deal with partially overlapping routes. This seems obvious but is important to mention, because dealing with overlapping routes (or correlating alternatives to use a more general term) is impossible or difficult for many discrete choice models.

The modelling methods found in literature are evaluated according to the criteria presented in the beginning of this section. The results are in Table 2.

Table 2 Evaluation of modelling techniques

	Easy to set up	Computationally simple	Can model route overlap
Multinomial Logit	Yes	Yes	No
Mixed Logit	Yes	No	Yes
Error component	No	Yes	Yes
Recursive logit	No	No	Yes
Nested Logit	Yes	Yes	Yes
Path-Size Logit	Yes	Yes	Yes

In short, Nested Logit and Path-Size Logit satisfy all requirements. Path-Size Logit models have been shown to yield more accurate results for route choice modelling than Nested Logit models. Therefore a Path-Size Logit model will be used in this investigation.

3.3 Modelling assumptions

The Path Size Correction Logit model we use for this investigation contains the following assumptions:

- **Utilities are link-additive.** This means the total utility of a route is equal to the sum of the utilities of all links. Mathematically this is expressed as:

$$V_i = \sum_{a \in \Gamma_i} V_a \quad (6)$$

- **Links are bi-directional and have the same utility in both directions.** This means it is assumed that for route choice, it does not matter whether a vessel goes from A to B or from B to A. This assumption was deemed acceptable by interviewed inland shipping captains.

- **Vessels know all alternatives in each choice set and the corresponding utilities. They make rational choices based upon this knowledge. Any irrational behaviour is modelled adequately by the random factor ϵ .** This assumption is deemed acceptable since inland shipping captains are professionals that often visit the port area many times during the year, and also there is a lot of information available on maps and via the internet. Water levels, wind, current speeds and direction, availability of berths and quays, location of other vessels are just a few examples of information that is available real-time, for free, online.

3.4 Creating a Path-Size Logit model

The creation of a logit model for route choice is generally done in three steps [11]. They are listed below

Path generation means creating a number of paths (\mathcal{M}) that will be examined from the set of all possible paths (\mathcal{U}).

Choice set formation is the selection of the subset of paths that the user can chose (\mathcal{C}), based on awareness of the possible routes, traffic information and a host of other factors. It is proposed to assume that inland shipping vessels are aware of all alternatives, as they are professional captains, many visit the port often and the amount of possibilities is relatively limited.

Formulating the **Route Choice Model** means setting up the equations that assign traffic volumes to each route based on that route's utility function. The function $P(i|C)$ describes the chance that a vehicle will choose path i from choice set C , where C contains routes between O-D combination n .

3.4.1 Path generation and choice set formation

The formation of the choice set is important, not only because it needs to represent reality but also because the choice set influences the outcomes of the model. Adding nonsensical routes to the choice set can significantly decrease the quality of the route choice predictions [12]. Obviously, leaving realistic routes out of the choice set will also decrease the predictive qualities of the model.

Path set \mathcal{U} is called the full choice set. It contains all routes between origin s_o and destination s_d . For many route choice models, \mathcal{U} cannot be explicitly generated because there are too many possible routes or not all aspects of the networks are known. For those types of problem a set $\mathcal{M} \subseteq \mathcal{U}$ is

defined, where \mathcal{M} contains all known or observable routes. In this report, \mathcal{M} contains all routes between s_o and s_d that do not contain the same link more than once.

Because the number of waterways is very limited and completely known, we do not need advanced techniques to determine the composition of \mathcal{M} , although they are often employed for other route choice models [4].

Choice set $C \subseteq \mathcal{M}$ contains what are deemed feasible paths. Path $i \in C$ if the following condition holds:

$$\sum_{j \in \mathcal{M}} Z_{ij} = 0 \quad (7)$$

With

$$Z_{ij} := \begin{cases} 1 & \text{if } \Lambda_i \supset \Lambda_j \\ 0 & \text{if } \Lambda_j \not\subset \Lambda_i \end{cases}$$

Where

Λ_i = set of nodes that form route i .

This means that a path i is excluded from C if it contains detours. A route is said to contain a detour if it contains all nodes of another route in the universal choice set \mathcal{M} . Choice set C represents the set of realistic alternative paths.

Whether we should use C or \mathcal{M} as the choice set is unclear from literature, but is it theoretically preferable to use \mathcal{M} because it contains all possible choices. If C is used instead, a sampling correction can be used to account for the fact that not all possible routes are taken into account [9]. Researcher Michiel Bliemer has shown that the Path-Size Logit method is very sensitive to the inclusion of irrelevant routes [13] [12] and that adding those routes will significantly worsen the predictive qualities of a PS-L model. Therefore it is proposed to use C and not \mathcal{M} as the choice set for this investigation.

Theoretically, it is possible to exclude a valid choice by using C as the choice set. In some situations where a direct route contains an obstacle with a very large penalty, it might be rational to use another route that contains a detour but avoids the penalty. No example of this was found when studying the Port of Rotterdam graph but the possibility cannot be ruled out. This might impact the accuracy of the model. However, since the Logit model is very sensitive to the inclusion of irrelevant routes, the loss of accuracy from using \mathcal{M} will likely be far greater than the loss of accuracy resulting from using C .

3.4.1.1 Example

If a ship wants to go from the Maasvlakte (MV) to the van Brienenoordbrug (see Figure 3), it can take a route (1) that has the following Λ : ['MV', 'VI', 'ZEE', 'NWA', 'III', 'II', 'I', 'BRIEN']. Theoretically,

another route (2) is ['MV', 'HART', 'V', 'ROOS', 'VII', 'EP_C', 'VI', 'ZEE', 'NWA', 'III', 'II', 'I', 'BRIEN']. This second route is excluded from the choice set because $\Lambda_2 \supset \Lambda_1$.¹

In Figure 8, routes 1 to 6 represent the feasible choice set \mathcal{C} and route 1 to 12 constitute the full choice set \mathcal{M} .

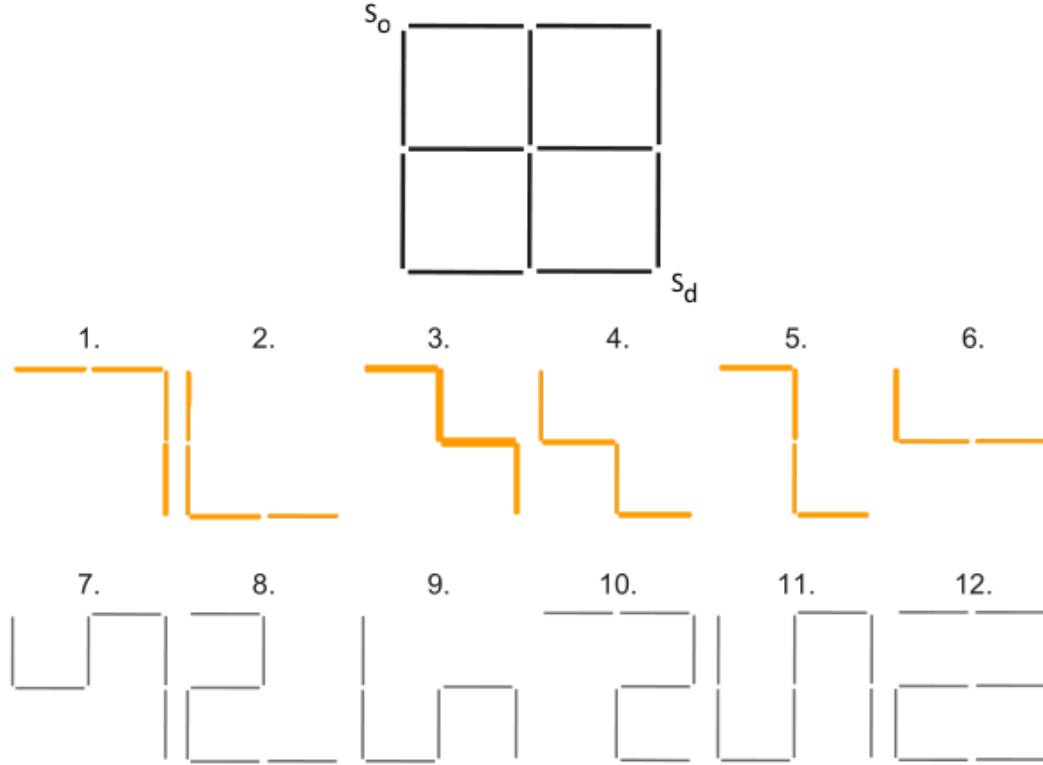


Figure 8 Example of an O-D Pair and the possible (1 – 12) and feasible (1 – 6) routes between them

3.4.2 Path-size factor

In order to model overlap between paths in the Logit model, a path-size factor is added to the utility function. The path-size factor can be formulated in a number of different ways [1]. There are two generally accepted formulations of the Path-Size correction factor. They are described and compared below and the most suitable alternative is selected.

3.4.2.1 Path-size factor and sampling correction (Frejinger and Bierlaire, 2007)

According to research by E. Frejinger [1], the general Path Size formulation is the most appropriate one to use, both because it seems to perform best and because it has a theoretical basis. It is defined as follows:

$$PS_i = \sum_{a \in \Gamma_i} l_a / L_i * 1/\Delta \quad (8)$$

l_a = length of link a

L_i = length of route i

¹ Note that in such a case, the statement $\Gamma_1 \subset \Gamma_2$ is not necessarily true.

Γ_i = set of links in route i

$$\delta_{ai} := \begin{cases} 1 & \text{if } a \in \Gamma_i \\ 0 & \text{if } a \notin \Gamma_i \end{cases}$$

$\Delta = \sum_{i \in C_n} \delta_{ai}$ = number of paths in choice set C_n sharing link a .

3.4.2.2 Path-Size Correction (Bovy and Bekhor, 2008)

The Path-Size Correction (PSC) method was proposed by Piet Bovy in his 2008 paper ‘The Factor of Revisited Path Size: Alternate Derivation’ [13].

The PSC formulation is as follows:

$$\begin{aligned} PSC_i &= - \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \ln \Delta \\ \Delta &= 1 / \sum_{i \in C} \delta_{ai} \end{aligned} \quad (9)$$

Utility is defined as

$$U_i = V_i + PSC_{i,C} + \epsilon \quad (10)$$

Both Bovy [13] and Frejinger [1] conclude that the performance of both methods in terms of model fit and prediction accuracy is roughly equal. However, the theoretical foundation for the PSC method is more robust. Therefore it is proposed to use the Path-Size Correction method.

3.4.3 Summary

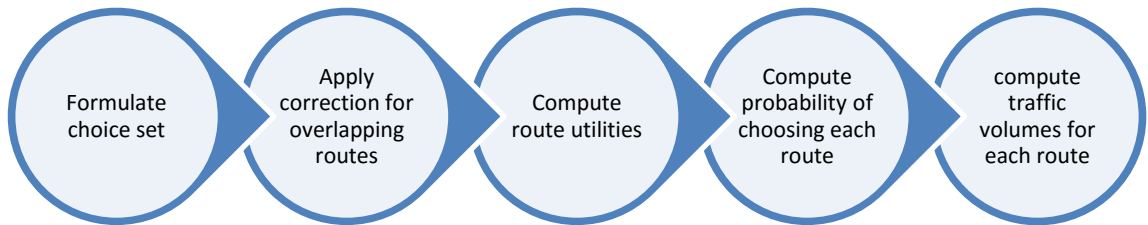


Figure 9 Schematic representation of route choice modelling

The proposed modelling approach is as follows (see Figure 9 for a schematic representation):

- Use C as the choice set

- Compute the utility V_i of each route by adding the utilities of all links in the route, so that the utility of the route is calculated in the following way:

$$V_i = \sum_{a \in \Gamma_i} V_a \quad (11)$$

- Apply the PSC path-size correction to each utility function so that $U_i = V_i + PSC_{i,C}$

-Compute probability that vessels will choose a route with the following formula:

$$P(i|C) = \frac{\exp(U_i)}{\sum_{j \in C} \exp(U_j)} \quad (12)$$

Where

$$U_i = V_i + PSC_{i,C} + \epsilon \quad (13)$$

And

$$\begin{aligned} PSC_i &= -\sum_{a \in \Gamma_i} \frac{l_a}{L_i} \ln \Delta \\ \Delta &= \sum_{i \in C} \delta_{ai} \end{aligned} \quad (14)$$

- Compute traffic flows for each route by multiplying the total amount of traffic by $P(i|C)$. Appendix 1 is a fully worked-out example of how this is done.

4 Route choice modelling

Each route in the choice set has its own utility value, V_i . This is the sum of the utilities of all links in the route:

$$V_i = \sum_{a \in \Gamma_i} V_a \quad (15)$$

The full utility function is then

$$U_i = V_i + PSC_{i,C} + \epsilon \quad (16)$$

The full utility function consists of three parts:

Path size correction factor (PSC) depending on path I and choice set C

Random factors (ϵ) – a generalized extreme value distributed number with scale 1 and location 0 [11].

Utility (V)

V_i consists of a number of parameters, each with its own weight factor:

$$V = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (17)$$

Or, alternatively:

$$V = \beta \cdot X \quad (18)$$

Where

$$\begin{aligned} \beta &= [\beta_1, \beta_2, \dots, \beta_n]^T \\ X &= [X_1, X_2, \dots, X_n]^T \end{aligned} \quad (19)$$

The number of parameters is often kept fairly small to reduce the risk of overfitting the model. Models with between two [11] and four [9] parameters are often found in literature. More parameters can be added at will but will often lead to mistaking ‘noise’ for ‘signal’ (see Figure 10), not to better quality of predictions. As the 20th century mathematician John von Neumann tells us “With four parameters I can fit an elephant, and with five I can make him wiggle his trunk”².

An overfit model will yield predictions that are worse than those of a well fit model, despite giving the impression of being more detailed. Note that a good statistical fit of the model to the data used for calibration does not mean it has not been overfit [14]. Indeed, correlation and model fit alone do

² As it turns out, this is actually true – see the paper “Drawing an elephant with four complex parameters” by Mayer, Khairy and Howard [21].

not mean anything, as can be seen clearly from the hundreds of meaningless correlations documented by Tyler Vigen on his ‘Spurious Correlations’ website [15].

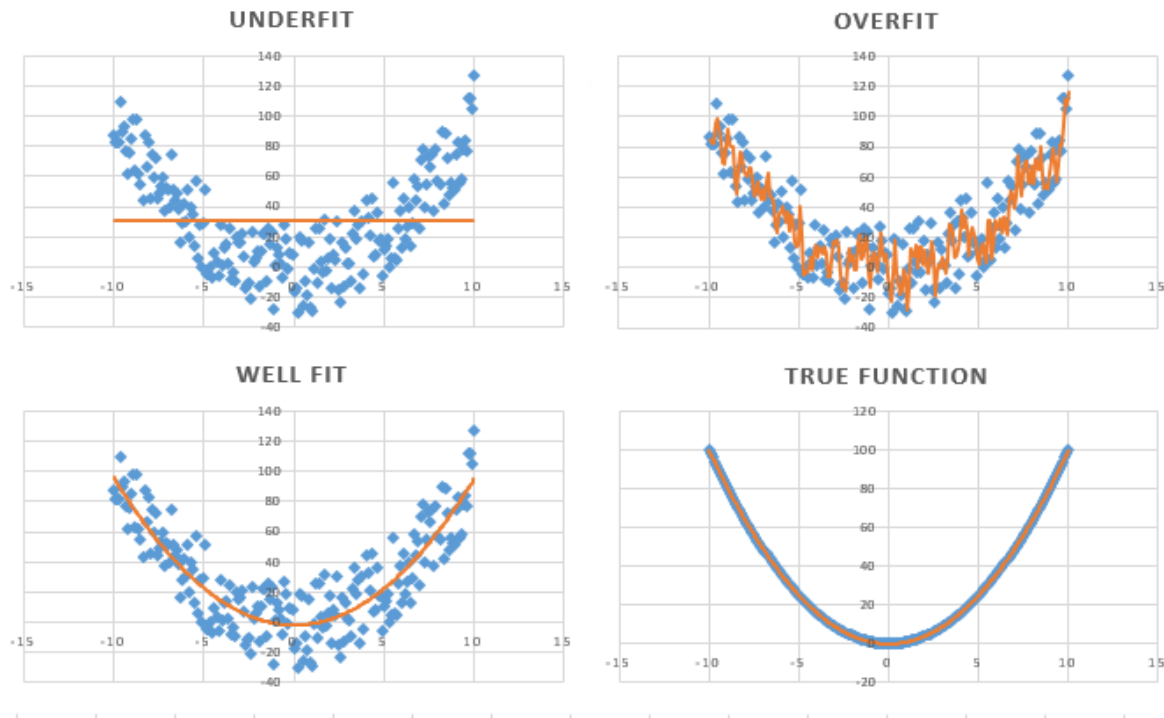


Figure 10 - Illustrating underfit (top left), well fit (bottom left) and overfit (top right) models. Data are in blue, the model prediction in orange. The true function from which the samples were generated is shown in the bottom right figure.

4.1 Poll results

In order to discover what inland shipping captains think are important factors for route choice, a poll was conducted. 77 captains responded. Three short follow-up interviews were conducted at the shipping fair in Gorinchem on the 3rd of June 2015. The full poll and results can be found in appendix 3. Amongst other things, the captains were first asked what factors they thought were most important for route choice. This was an open question with room for written responses. They were also asked to choose from a list of factors those they thought were important. This was a closed question with a limited number of possible answers. The responses are summarised below.

Question: what are the most important considerations when choosing a route? (more than one answer possible)

Factor	# times chosen	% of times chosen
tides and currents	42	55%
distance and travel time	27	35%
weather / wind / waves	14	18%
bridges	11	14%
Obstructed waterways	7	9%
personal reasons	6	8%
berths and utilities	5	6%
depth of waterways	3	4%
congestion	1	1%

Which of the following factors do you think are important when choosing a route? (multiple answers possible)

Factor	# times chosen	% of times chosen
Tides	62	81%
Currents	59	77%
Distance	44	57%
Travel time	43	56%
Wind	37	48%
Bridges	36	47%
Locks	26	34%
Berths	26	34%
Waves	21	27%
Speed limits	17	22%
Other infrastructure	11	14%
Narrow / shallow waterways	10	13%
Route of friends / colleagues	6	8%
Congestion	4	5%
None of the above	0	0%

Based on the responses, the following factors will be investigated.

- Distance
- Tides
- Currents
- Wind
- Waves
- Bridges
- Locks
- Berths
- Speed limits
- Narrow / shallow waterways
- Congestion

4.1.1 Influence of vessel characteristics on route choice

Different ships could react to conditions on the waterways in different ways. Therefore, the influence of cargo type and ship size on route choice will also be investigated. This is discussed chapter 9.

4.2 Investigating of possible drivers for route choice

In order to decide if a factor can be part of the route choice model, two questions must be answered.

1. Will the factor have impact on the route choice in the Port of Rotterdam area?
2. Can this influence be detected?

If the answer to both questions is positive, the factor will be incorporated in the route choice model.

4.2.1 Distance

Distance will probably be an important factor in the choice of route as it impacts the travel cost and travel time. The distance between different points in the port can easily be measured in km because many high-quality maps are available. Because origins and destinations are known only approximately, some inaccuracy can be expected in the distance measurement.

4.2.2 Travel time

Travel time depends on distance and speed. The distance part has been discussed above. Travel speed is dependent on the winds, currents and perhaps speed limits, which will be discussed below. Therefore the factor 'travel time' is not used as a separate part in the model, because it is effectively incorporated by using these other parameters.

4.2.3 Narrow / shallow waterways

The waterways in the port of Rotterdam have been constructed to allow passage of inland and deep-sea vessels. Therefore all main waterways are deep enough to accommodate any known inland shipping vessel. The same goes for the width of the waterways. The one exception here is the area called 'Breddiep' which ironically is neither wide (breed) nor deep (diep).

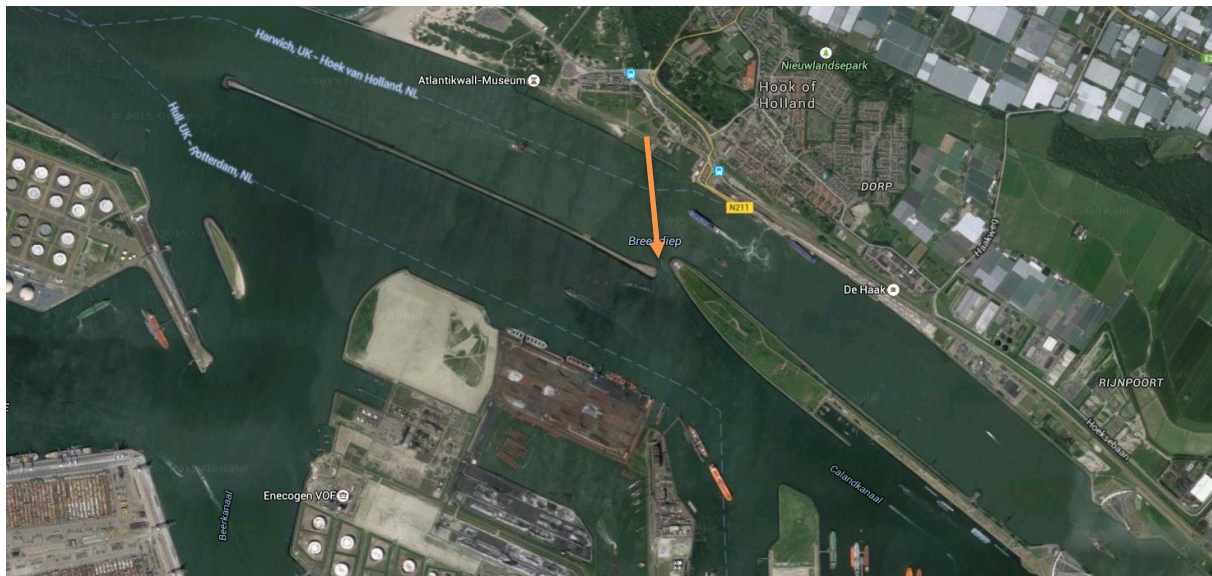


Figure 11 the Breddiep (source: Google maps)



Figure 12 - the Breeddiep. Picture by Leo van der Stok

This small passage between the Nieuwe Waterweg and Caland channel was never meant to be used for water traffic, but a court case brought by inland shipping captains forced the Port Authority and Rijkswaterstaat to allow them passage. The passage is 75 m wide and 6 m deep. While this is still wide and deep enough to comfortably accommodate any inland shipping vessel, the combination of fast currents, heavy winds, bad visibility and relative lack of manoeuvring space make this area difficult to navigate. Therefore the Breeddiep will be included in the model as an obstacle.

4.2.4 Tides (water levels)

The influence of the tides on the water can be felt in two ways: changing water levels and changing currents. The influence of changing currents will be discussed below in the 'currents' section. This section will focus on the effect of changing water levels.

The Port Authority measures the water levels at a number of points in the port (see appendix 2). These measurements show that fluctuations in water levels are roughly equal (ranging from -0.5 to +1.5 m NAP) in all major waterways in the port. They also rise and fall at the same time in the different waterways. Therefore the water levels cannot be a reason to choose one waterway over another. It will not be included in the model.

Rising or falling water levels are not, by themselves, problematic. They can influence inland shipping traffic by making it harder to pass bridges. This will be discussed in the 'bridges' section.

4.2.5 Bridges

Bridges can be an obstacle for inland shipping vessels if they are too low to allow for passage without opening of the bridge. Most bridges are opened on request but the Botlek bridge has a so-called

‘rush hour regimen’ which means that it will only open once every 20 minutes during rush hour so as not to slow down road traffic too much.

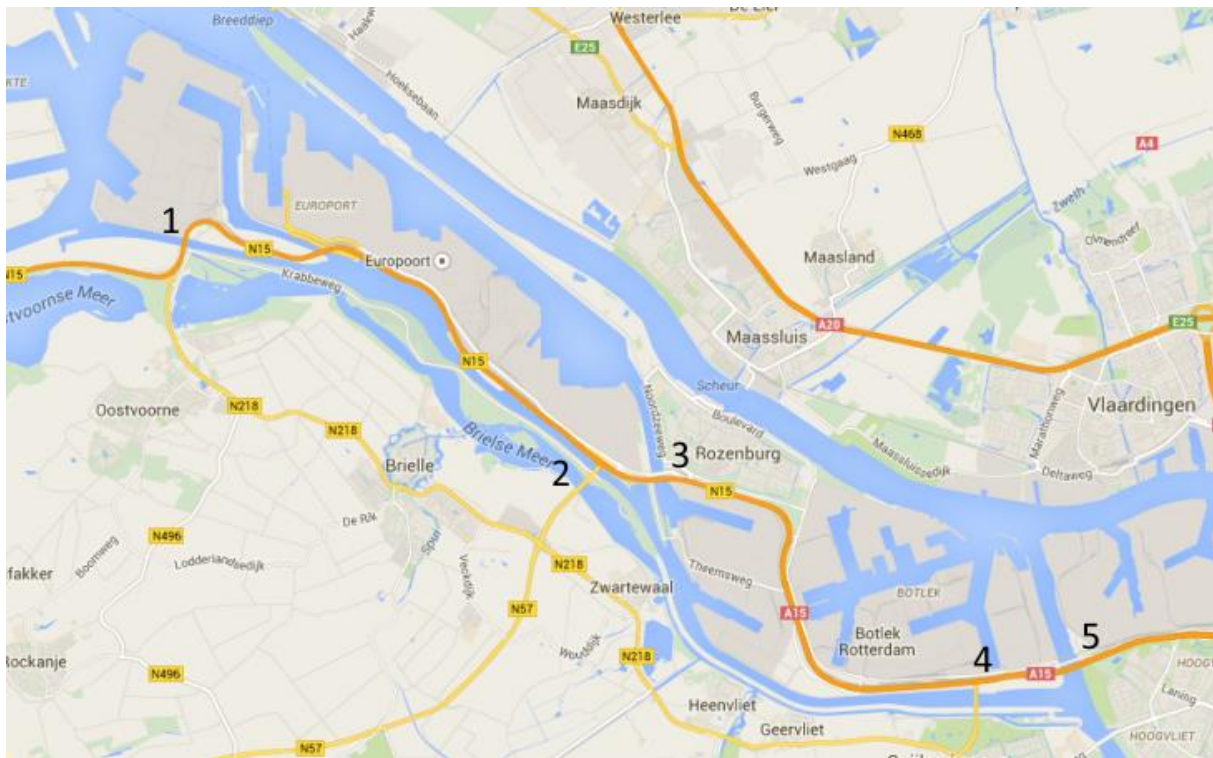


Figure 13 Bridges in the port area that could influence route choice (map adapted from Google Maps)

Table 3 bridges and heights. Distance from bridge to water level depends on the tides

Number	name	Height (above NAP)	Height to water level (-0.5 ... 1.5 m + NAP)
1	Suurhoffbrug	11.5 m	10 – 12 m
2	Harmsenbrug	11.5 m	10 – 12m
3	Calandbrug	11.7 m	10.2 – 12.2 m
4	Hartelbrug	10.8 m	9.3 – 11.3 m
5	Botlekbrug	8.0 m	6.5 – 8.5 m

The report “rapportage containerhoogtemetingen” written for Rijkswaterstaat [16] states that bridges with a height of 11.3 m or more are sufficient to allow all inland shipping vessel to pass that are the same height or lower than a class VI container ship with four empty high-cube containers – currently the highest type of inland shipping vessel. Vessels with three layers of containers can pass with a height of 8.4 meters, and half-loaded vessels with four layers of containers can pass with 10.4 meters³. This means most vessels will be able to pass the first three bridges on the list with little to no trouble at any time.

³ These numbers are higher than the ones usually used by Rijkswaterstaat from the ‘Richtlijnen Vaarwegen 2011’. They reflect the growing numbers of high-cube containers in use.

The Hartelbrug and Botlekbrug are significantly lower than the other two bridges. They will have to be opened more often to allow inland shipping vessels to pass, and thus might be a significant obstacle for inland shipping traffic.

A number of respondents specifically name the Botlek Bridge as being a major obstacle and a reason to alter their route. Interviews reveal a main cause of this is not only the bridge height but also frequent delays in opening because of technical problems or the rush hour regimen. Therefore the Botlek Bridge will be added to the model as an obstacle.

The Hartelbrug was named by none of the respondents or interviewees as being an obstacle. A number of tests calibrating the model while including the Hartelbrug always produce optimal fit when its penalty is set to 0. Therefore it was decided not to include it in the model as an obstacle.

4.2.6 Currents

Currents in the port area can be the results of the flow of the river towards the sea, and the results of the influence of the tides. The tides have the biggest influence, as can be seen from several graphs created with the current model (Operationeel Stromings Model, OSR) from the Port of Rotterdam Authority (see appendix 2). A small offset relative to the 0 axis represents the continuous outflow of the river towards the sea.

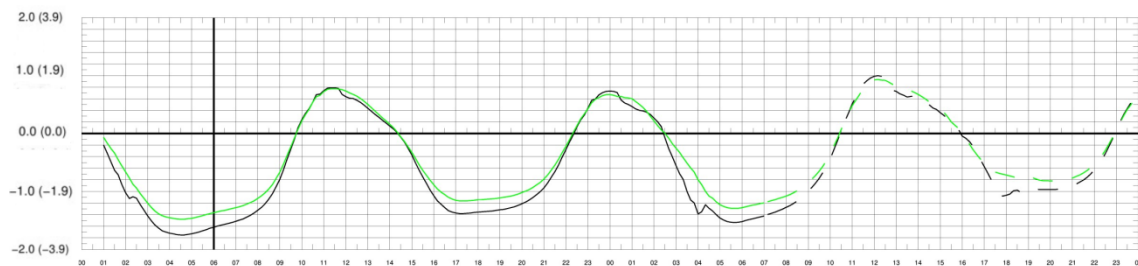


Figure 14 Graph created by OSR showing currents near the Spijkenissebrug

Changing current because of tides might influence the departure time of vessels but not the route, since current speeds are roughly the same on all waterways. The exception is the Caland channel, which has almost no current because of the lock at its end. The tides and currents also occur at the same time in the different waterways in the port.

Both the Hartelkanaal and the Nieuwe Waterweg feature some locations with strong currents, but interviews with inland shippers indicate that for them, this would not be a reason to choose another route. The captains I spoke to told me currents caused by the tide mainly influence departure time and choice of destinations. Captains often need to visit more than one location in the port area and choose their point of entry and sequence of visits in such a way they can travel in the direction of the currents as much as possible. However when destination and origin are taken as a given (as is the case in this model) currents will not influence route choice. Therefore, they are not included in the model.

4.2.7 Wind and waves

Interviews with captains reveal that there are two areas where wind and waves are a big issue – the Breeddiep and the Lage Licht (sea), both located at the western end of the Nieuwe Waterweg. Both

locations are relatively close to the open sea and have little cover from the winds, which cause slower travel speeds and waves. The Hartelkanaal and Calandkanaal are considered calmer with respect to the wind and waves.

Because waves and wind are often cited as reasons to choose another route we will include the wind and waves at the Breddiep and the sea in the model.

4.2.8 Locks

The port of Rotterdam contains only one lock, the Rozenburgsesluis (lock near Rozenburg). It is situated at the end of the Caland channel, where it joins the Oude Maas. This lock takes at least 20 minutes to fill or empty and has fixed width and length, meaning only a certain number of vessels can fit in it at the same time. It therefore presents a significant obstacle and will be included in the model.

4.2.9 Berths

Some captains cite location and availability of berths as a factor that influences route choice. In interviews, the captains said this was more of a factor when deciding on a destination rather than one that influenced the route they took towards that destination. A visit to the port area is made up of a number of trips, and locations of berths might influence the destination of those trips.



Figure 15 Location of public berths in the port area. Source: Port of Rotterdam Authority

Apart from these public berths there are also private berths, buoys and dolphins that can be used to moor vessels. The map shows that vessels can never be more than half an hour's worth of travel away from a public berth. Therefore it is unlikely that a ship would choose a different route in the port area based on the location of berths along the route. Choosing a different destination (and thus planning a different route for the whole trip) is a possibility but the choice of destination is not studied in this research.

4.2.10 Speed limits

The port of Rotterdam has imposed speed limits on parts of the Hartelkanaal and the Nieuwe Maas. This limits the speed of inland shipping vessels to 13 km/h relative to the water flow speed. The speed limit is only applicable when wind speeds are below 7 knots.

The speed limit in the Nieuwe Maas is only applicable to a small zone (4 km in length) in the centre of Rotterdam. There is no alternative route to take in this area so this speed limit will not be included in the route choice model.

The speed limit in the Hartelkanaal applies to a zone of roughly 10 km in length, from the Oude Maas to the Harmsenbrug (see Figure 16). Inland shipping vessels normally have speeds that are somewhere between 10 and 20 km/h, meaning the speed limit could increase travel time by somewhere between 0 and 30 minutes. The speed limit could be avoided by choosing the Nieuwe Waterweg instead of the Hartelkanaal.

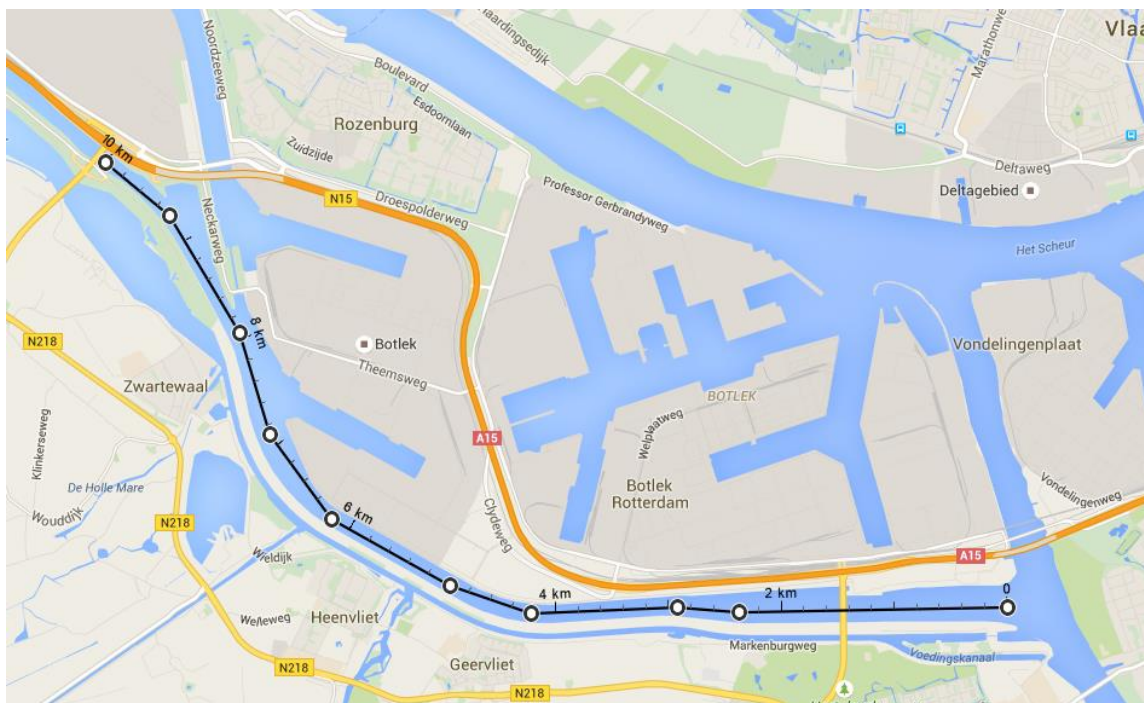


Figure 16 Hartelkanaal – the zone where the speed limit is in place is marked with the black line. Map adapted from Google Maps

Tests taking the speed limit into account as an obstacle on link 8 (see Figure 5) show the model fit is best when the penalty is set to 0. This suggests the speed limit does not influence route choice enough to warrant inclusion in the model.

4.2.11 Congestion

Interviews with Port Authority officials and inland shipping captains suggest that congestion is not currently a problem for inland shipping vessels in the Port of Rotterdam. The only places where vessels might have some hindrance because of congestion are the lock near Rozenburg and the Breddiep, which are both included in the model as obstacles. Therefore congestion in general is not included in the model.

4.3 Conclusion

The following factors are likely influential for the route choice:

- Distance
- Bridges
- Locks
- Wind
- Waves
- Narrow – shallow waterways

Influences can be incorporated in Logit models as penalties assigned to certain links. These factors will be included in the model as ‘obstacles’ at the following locations:

- Breiddiep (link 10)
- Lage Licht (sea) (link 11)
- Rozenburgsesluis (link 19)
- Botlekbrug (link 8)

These links will receive a penalty greater than their distance.

4.3.1 Defining the utility function

Having decided which additional penalties to include in the route choice model, it is now possible to write down the full utility function. The function for route i is defined as follows:

$$U_i \triangleq V_i + PSC_{i,C} + \epsilon \quad (20)$$

Where

$$V_i = \beta \cdot X_i \quad (21)$$

PSC and ϵ are the path-size correction and random factors.

Vector X is defined as:

$$X_i = [X_D, X_{ROOS}, X_{BZEE}, X_{BREE}, X_{BOT}]^T$$

Where X_D is the length of the route in km and the other X 'es are dummy variables that are either 1 or 0, depending on whether the obstacle that is referred to is in the route or not.

And β is defined as

$$\beta = [\beta_D, \beta_{ROOS}, \beta_{ZEE}, \beta_{BREE}, \beta_{BOT}]^T$$

5 Using traffic data to calibrate the route choice model

Traffic data supplied by the Port of Rotterdam Authority will be used to calibrate the parameters of the route choice model and to verify it. However, there is no dataset that contains route choice for inland shipping vessels. The only available datasets contain vessel positions over time. Therefore, some data analysis is required in order to reach the information on route choice.

5.1 Traffic data collection by the port authority

The port of Rotterdam uses Automatic Identification System (AIS) and radar data to track the position, orientation, identity and speed of all vessels in the port area. This information is stored every 9 seconds for each vessel. In order to make the dataset more accessible, a processing step is applied by the Port Authority. In this process, the full list of positions is reduced to a list of areas the vessel has visited. For a list of possible areas, see Table 1. The time when the vessel enters and exits the area is also recorded, as well as the time when vessels moor at and leave berths. The other data is discarded. Each visit to the port (all area visits between entering and exiting the port area) is given a unique identification number. The vessels are identified by their MMSI, short for Maritime Mobile Service Identity.

The processed dataset, produced by a system called HaMIS (Haven Management en Informatie Systeem) is used in this investigation. An example of this data can be seen in Figure 17.

The dataset used for calibration contains all registered visits of inland shipping vessels to the port of Rotterdam for the first 3 months of 2014.

SB_1	22/01/14 12:30	22/01/14 12:45		PASSAGE_AREA
PORT	22/01/14 12:45	24/01/14 0:13		PORT
CA_1	22/01/14 13:20	22/01/14 13:36		PORT_BASIN
BT_1	22/01/14 13:36	23/01/14 22:57		PORT_BASIN
Z100/10/93	22/01/14 13:36	23/01/14 19:54		TERMINAL
Z100/10/95	22/01/14 19:54	23/01/14 20:23		TERMINAL
Z100/10/94	23/01/14 20:04	23/01/14 21:06		TERMINAL
Z100/10/95	23/01/14 21:06	23/01/14 21:36		TERMINAL
CA_1	23/01/14 22:57	23/01/14 23:24	23/01/14 23:21	PORT_BASIN
SB_1	24/01/14 0:13	24/01/14 0:36		PASSAGE_AREA

Figure 17 Some of the data recorded for a visit of a vessel to the port of Rotterdam. Not all data is shown – more information is included in each row of the actual database. Only the most relevant information is shown here.

The screenshot in Figure 17 shows only part of the information that is recorded. The real database has many additional columns. Each full record in the HaMIS spreadsheet contains the information described in Table 4.

Table 4 fields in the HaMIS database with explanation

mmsiNumber	The MMSI number of the vessel- blurred in the screenshot for privacy reasons
bargeCategory	Category inland shipping vessel
Length	Length of the vessel
Beam	Width of the vessel
maximumDraught	The maximum depth of a loaded vessel in the water, taken from the level of the waterline to the lowest point of the hull.

Tonnage	Maximum tons the ship can carry
Flagcode	Country where the ship is registered
Vesselvisitid	Unique code created by HaMIS for each visit – consists of mmsinumber-visitidentificationnumber
Areacode	Code identifying the area where the ship was sighted
Datetimestart	Moment the ship was first seen in the area
Dateteimeend	Moment the ship was last seen outside the area
lastseen	Moment the ship was last seen in the area – can be quite different from dateimeend if the AIS- and/or radar signals were lost at some point
Numberofbunkerings	The number of bunkerings – only applicable to bunker ships
areatype	The type of area (terminal, passage_area etc.)
mainsegment	Main commercial activity in the area – not necessarily commercial purpose of the ship
Marketsegment	More specific description of commercial activity in the area
productgroup	Even more specific description of commercial activity in the area
Shipnumber	ENI-number, European ship identification number.
startendtime	

5.2 Processing the data

In order to gather information on route choice, the database from the port of Rotterdam will have to be analysed and manipulated. A software package was written in the Python language for this purpose.

The traffic database has already been divided into visits- each with its own unique number called “vesselvisitid”. All records with the same vesselvisitid together form a description of a visit to the port area. An example of such a series of records can be seen in Figure 17 and is illustrated in Figure 18.

In order to study route choice however we are interested in trips from one location to the other. This means that each visit will have to be subdivided into one or more trips. Trips will consist of an origin, a destination, and a route (list of places visited between origin and destination). The first trip starts when the vessel enters the port and ends at the first berth. The endpoint of one trip is automatically the starting point of the next, which ends when there is another berth, and so on until the vessel leaves the port area.

5.2.1 Example

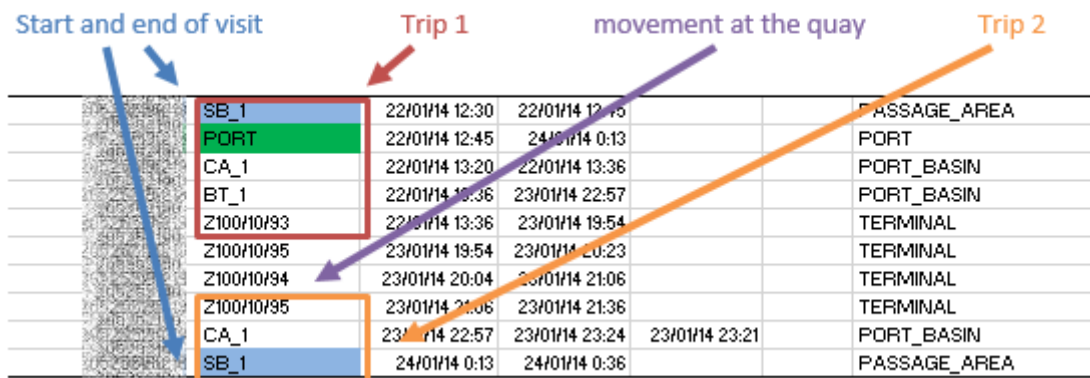


Figure 18 Illustrated example of traffic data – mmsi numbers have been blurred for privacy reasons

This example from the traffic data describes a visit to the port of Rotterdam in January 2014. The MMSI-number has been blurred out for privacy reasons. The vessel takes two trips during its visit – one from the Spijkenissebrug (SB_1) through the Calandkanaal and the lock near Rozenburg (CA_1) to the Brittaniëhaven (BT_1), and a return trip to the Spijkenissebrug using the same route. The vessel has moored at a number of berths in the Brittaniëhaven. Berths are recognizable by the fact that all berth codes start with 'Z100/'. Movements between berths in the same port basin are not counted as trips.

5.3 Python data analysis

The traffic data for the first quarter of 2014 contains roughly 20,000 visits and almost 100,000 trips. In order to determine from traffic data how many trips were made and what routes were taken, a significant amount of data processing needs to be done. This is very difficult with Excel. Therefore, the programming language Python was used to build custom software to process and analyse this dataset.

The Python program takes the traffic data from the excel sheet (in .csv format) as input and stores the visits and trips using the methods described in the previous sections. For each O-D pair, it outputs an Excel spreadsheet that detail all routes recorded between them and the amount of time they were chosen. Python also computes the feasible routes (C) for each O-D pair, the PSC-factor and the total utility for each route. Using the location data discussed here as input, the Python scripts deliver an excel file containing the following information:

- amount of trips made in the port area
- origin and destination of each trip
- vessel size and cargo type
- route taken and alternative routes
- PSC factor and Utility for each route
- predicted chance of choosing each route
- amount of vessels that took each route

5.4 Errors in traffic data

This section describes some errors in the traffic data. To illustrate this section a simple graph is presented in Figure 19.

When analysing the traffic data, frequently routes are found in the data that are not in choice set C for the corresponding O-D relation. Often, the registered route is not even part of the possible choice set \mathcal{M} . This can mean two things:

1. The choice set is incorrect
2. The recorded route is incorrect

After a study of the discrepancies between the data and the theory and discussions with Witteveen+Bos and the Port Authority, it was concluded the second option was more likely.

Two things can be wrong with storing the location of a vessel. A vessel can be registered in a place it has not visited, or it can be not registered in a place it has visited. Sometimes these errors are invisible, when they make a choice of one route appear to be a choice of another (but similar) route. For example, when a ship going through one area (C) is mistakenly registered in another (E). These errors are undetectable and therefore cannot be removed from the data.

Other errors are visible. One type of error is omitting a place from a route. For example, route $A \rightarrow B \rightarrow C \rightarrow$ will sometimes show up in traffic data as " $A \rightarrow C$ " – that would still allow us to determine which route the vessel chose between O and D. If the route, on the other hand, is " $A \rightarrow B$ ", that does not tell us anything – did the vessel reach its destination via C or E?

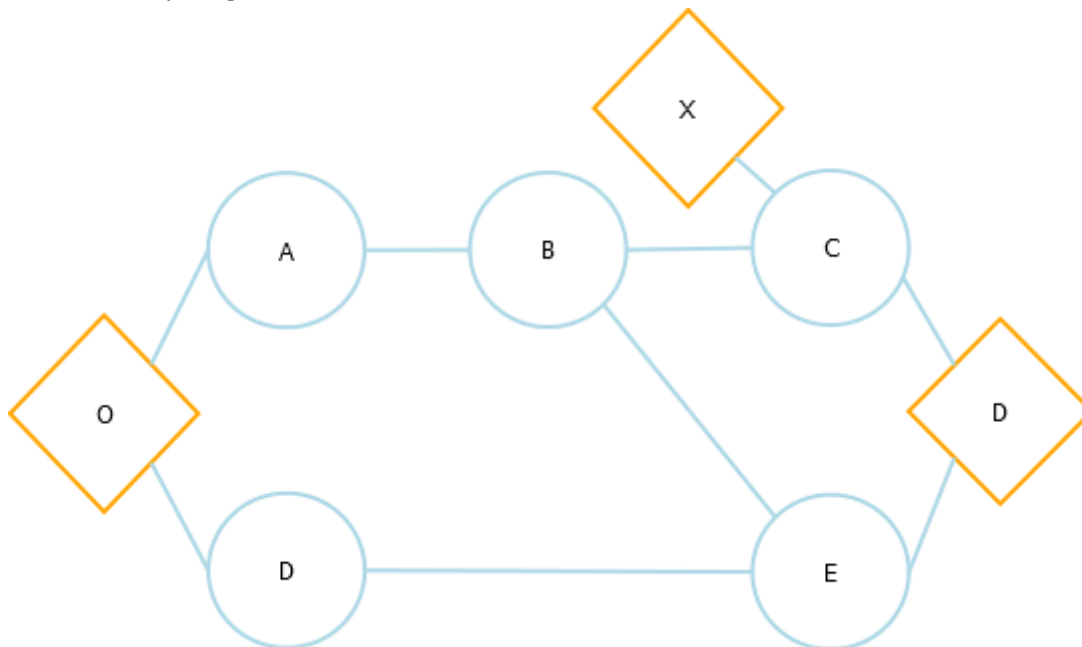


Figure 19 A simple graph

Sometimes, multiple places are missing from the route, so that a trip from O to D will be registered as having route "A" or even "...".

Another type of error that occurs is data that is visibly incorrect. For example, a route might be registered as “A → B → D → E” even though that is impossible. Another common error is areas which are close getting mixed up, so that a vessel might be seen as visiting X when in fact it was merely passing by.

Finally, not all berths in the port area are currently registered in HaMIS. This means that some visits to a terminal are not registered, giving the impression of a vessel merely passing through an area when in fact it has moored there (making the area the destination and not part of the route).

5.4.1 Example

mmsi	lumb	barge	length	berth	area	vesselVisitId	areaCode	dateTimeStart	dateTimeEnd
		8430	16	4	NL	-1395662911000	SB_1	24-03-14 12:50	24-03-14 13:08
		8430	16	4	NL	-1395662911000	PORT	24-03-14 13:08	26-03-14 12:16
		8430	16	4	NL	-1395662911000	NW_1	24-03-14 13:38	24-03-14 14:01
		8430	16	4	NL	-1395662911000	NW_1	24-03-14 15:30	24-03-14 15:48
		8430	16	4	NL	-1395662911000	CA_1	26-03-14 8:09	26-03-14 9:23
		8430	16	4	NL	-1395662911000	Z100/10/90	26-03-14 8:24	26-03-14 8:59

Figure 20 Traffic data with an obvious error – note the time difference between the 4th and 5th line.

An example of errors in the traffic data is shown in Figure 20. According to the Python algorithm, this data represents a trip from SB_1 to BV_BRIT (where the berth z100/10/90 is located) via NW_1 and CA_1. This route is not in the choice set C , or even in the possible choice set \mathcal{M} .

A manual inspection reveals something interesting – the time difference between line 4 and line 5 is 1.5 days! Presumably the vessel has done something during this time. It has at least moved to the CA_1 area, but via which route we cannot tell, nor do we know anything about any other trips it may have taken.

The Port Authority has software that lets users plot a vehicle's trajectory for a given time period. A plot for this trip is shown below.



Figure 21 Plot from the HbR Vessel Tracking System

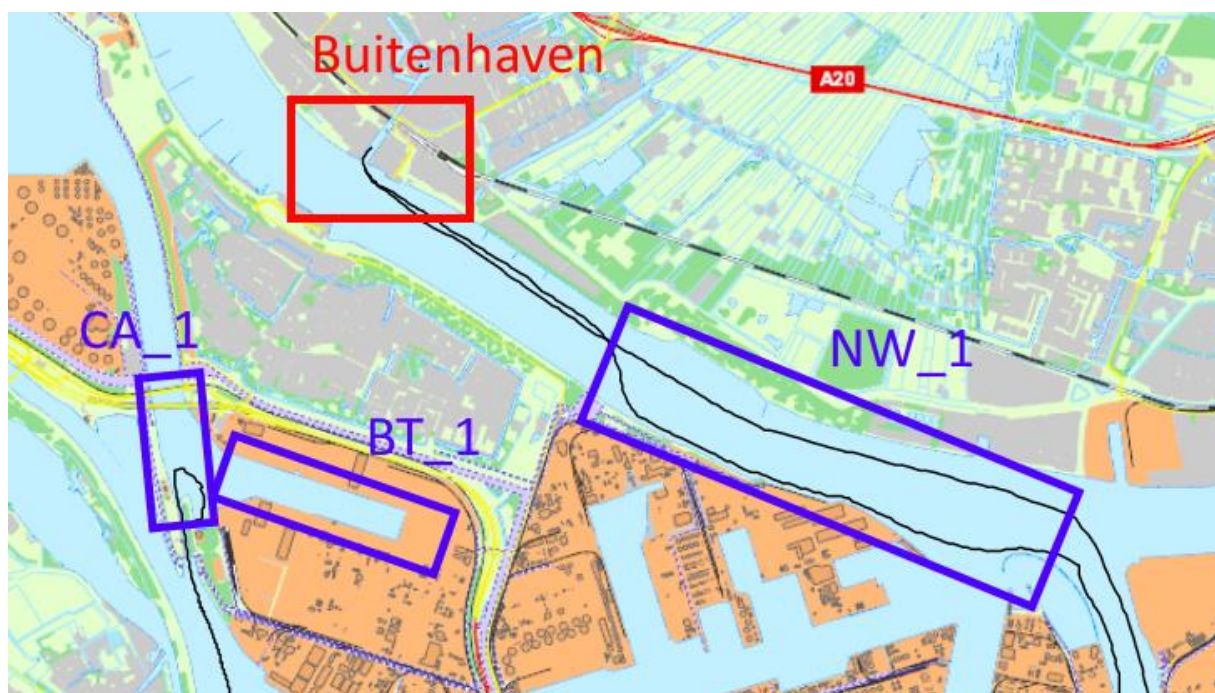


Figure 22 Detail from the previous figure - areas marked in blue, unknown area marked in red

As can be seen from Figure 21 and Figure 22, the vessel did indeed visit the areas registered in the traffic data. However, it stayed in the Buitenhaven for a time, where no berths are registered. In reality, the vessel made 2 trips: one from the Spijkenissebrug to the Buitenhaven, and one from the Buitenhaven to the Brittaniëhaven. Traffic data was only recorded for a part of the actual journey, providing a distorted picture of reality.

5.4.2 Example 2

mmsiNumb	bargeC	length	beam	maximi	tonnag	flagCoc	vesselVisitId	areaCode	dateTimeStart	dateTimeEnd
		77	8				000054445-1393938022000	Z100/10/47	05/03/2014	06/03/2014
		77	8				000054445-1393938022000	Z100/10/112	06/03/2014	06/03/2014
		77	8				000054445-1393938022000	SB_1	06/03/2014	06/03/2014

Figure 23 Screenshot from the HaMIS traffic data showing a trip from BT_1 to SB_1 - note that the areas in between are missing

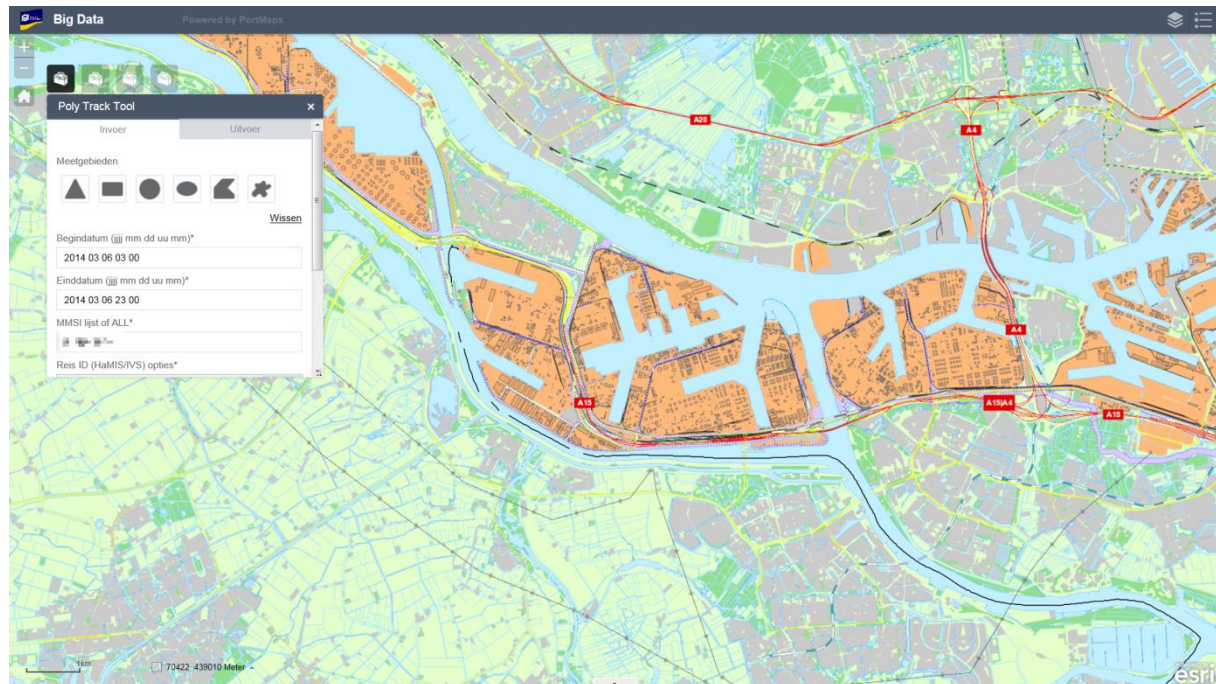


Figure 24 Plot from the HbR Vessel Tracking System – note the interruptions in the tracking line near BT_1

Another example is shown above, also for a trip between the Spijkenissebrug and Brittaniëhaven. For this trip, the route is empty – from looking at the HaMIS data it is not possible to say which route was taken. Plotting the course for this vessel using the Vessel Tracking System, we see the AIS and/or radar signal was lost a few times – most importantly during the time the vessel was in the lock in the CA_1 area. This explains why the CA_1 entry is missing from the HaMIS data.

5.5 Error reduction

A few algorithms were developed in order to remove errors from the traffic data. They are described below.

- Remove all port bassins from the route. Port bassins are not part of the normal waterways and can therefore not be part of a route, but only an origin or a destination. Sometimes a vessel is registered as being in a port basin because it merely passes it on one of the waterways.

For example (see Figure 19), the route $[A, B, X, C]$ would be reduced to $[A, B, C]$ because we know that X cannot be part of any route between O and D.

- If a recorded route (R_i) is not part of the route choice set ($R_i \notin C_n$) but can be identified as a part of route $R_j \in C_n$ and is only part of route R_j and not of all other routes in C_n , then we assume $R_i \equiv R_j$.

$$R_i \equiv R_j \text{ if } \begin{cases} R_i \subset R_j \\ R_i \not\subset R_n \forall R_n \in (C_n - R_j) \end{cases} \quad (22)$$

For example: if $R_i = [A, C]$ then we know that in reality, $R_i = R_c = [A, B, C]$. However, if $R_i = [B, C]$ we do not know what route was chosen: R_c could be either $[A, B, C]$ or $[D, E, B, C]$. Trips that have such incomplete routes are registered as errors.

-The area 'CA_1' is often not registered. This is because a part of the lock near Rozenburg is not in the area, so many ships that pass through the lock are only technically in the CA_1 area for a very short time. See Figure 25. This causes many ships that pass through the lock to be registered incorrectly in the traffic data.

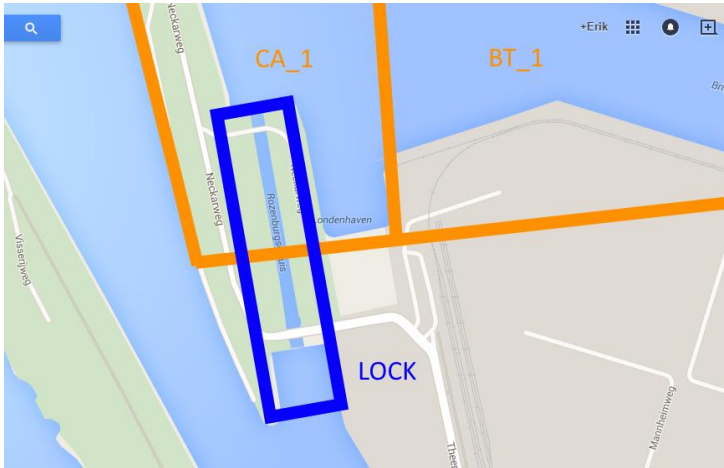


Figure 25 The Rozenburgsluis is only partially in area CA_1 (map from Google maps)

5.6 Summary

The port of Rotterdam Authority collects very detailed traffic data from all vessels in the port area. A processed version of this database containing only the inland shipping vessels is used for this investigation. Additional data processing was done using Python and Excel in order to distil information on route choice from the traffic data. Sometimes the data contains errors. It is possible to correct some, but not all of those.

Using the dataset, a custom data analysis tool written in Python and MS Excel, the route choices made by inland shipping vessels can be extracted from the traffic data for the first quarter of 2014. This data can be used to calibrate the parameters of a PS-LOGIT model as described in chapter 3.

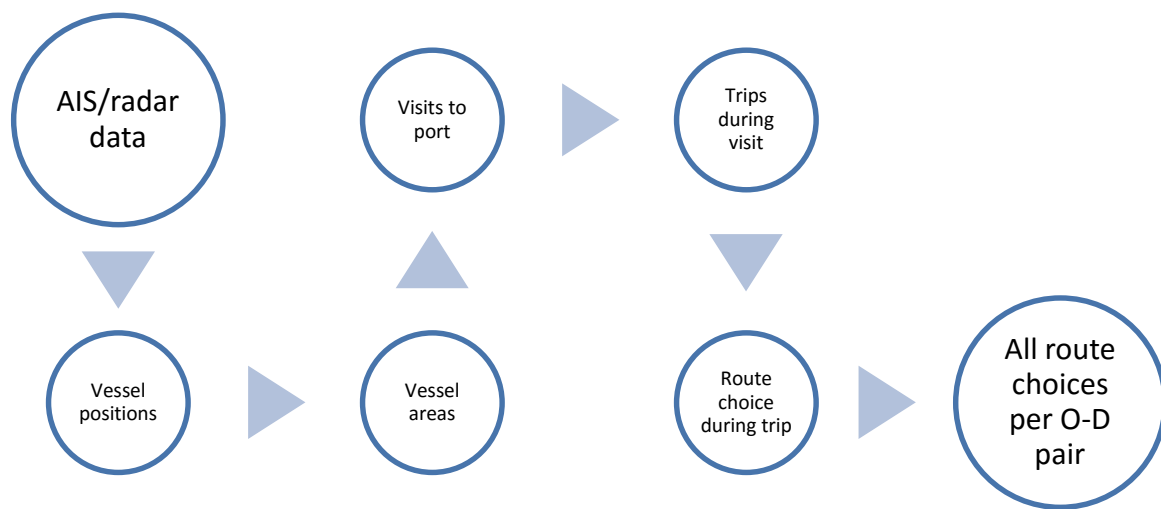


Figure 26 The data analysis process

6 Calibration methods and performance indicators

This sections describes how the model is compared to reality as observed from the traffic data.

6.1 Calibration method

When calibrating the model traffic data from the 1st quarter of 2014 was used. All trips and route choices made during that time by inland shipping vessels were analyzed. The calibration process aims to find the value of the vector β that gives the best match between model predictions and reality. From literature on other route choice models and discrete choice models in general, it is know that the calibration method best suited for this task is the so-called Maximum Likelihood method [6] [1] .

When using this method, the choice model is calibrated in such a way that the likelihood of the observation resulting from the model is maximised. This calibration method assigns more penalty to wrong predictions if they are very unlikely.

6.1.1 Maximum likelihood

Let likelihood $L(\beta)$ be defined as:

$$L(\beta) = \prod_{n=1}^n \prod_{i \in C(n)} P(i|C)_n^{g_{in}} \quad (23)$$

Here, g is a dummy variable defined as follows:

$$g_{in} = \begin{cases} 1 & \text{if route } i \text{ was chosen by ship } n \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

Remember $P(i|C)_n$ is the probability of vessel n choosing alternative i from choice set C .

The function we seek to maximise is $l(\beta) = \log(L(\beta))$.

$$l(\beta) = \log(L(\beta)) = \sum_{n=1}^n \sum_{i \in C_n} g_{in} \log(P(i|C)_n) \quad (25)$$

Maximizing this function is much simpler and will yield the same optimum β as maximizing $L(\beta)$ [6]. The log likelihood is used often in this report instead of likelihood because for large numbers of trips, the number $L(\beta)$ will quickly become very small.

6.2 Assessing model performance

This section details what metrics were used to assess how well the model corresponds to reality.

Accuracy is often used to assess model quality [17]. Overlap between the prediction for the 2nd quarter of 2014 and actual traffic (i.e. how many choices were predicted correctly?) can be used to assess the performance of the route choice model. This will result in a percentage of route choices that have been predicted correctly. The percentage of correct predictions can be calculated for the whole dataset or for each Origin-Destination pair individually. Looking at each Origin-destination pair individually will give a more detailed view of how accurate the predictions using the model are. This measure was chosen as a main performance indicator after discussion with Witteveen + Bos.

Likelihood or log likelihood can also be used as an indicator of how well the model corresponds to reality. Using formula 24 to calculate the log likelihood provides a dimensionless number that is

useful for comparing different models and for calibration. It is not a very intuitive measure of how well the model performs when predicting traffic or how well it compares to the real world.

Keep in mind that the highest percentage of correct predictions does not necessarily occur at the value of $[\beta]$ associated with the maximum likelihood. Theoretically, a certain $[\beta]$ could produce many correct predictions, but some erroneous ones that are very unlikely. This could make the likelihood of a model using this $[\beta]$ less than that of another model producing more, but more likely errors.

7 Calibration

Remember from the previous section that the Logit model has a number of factors that influence the route choice. They are collected in vector X . The weight vector β determines how each factor impacts the final decision. Calibration is the process of determining the optimal value for β . Here ‘optimal’ is taken to mean a value of β that results in the least squares solution, or lowest sum of the squares of all errors made. An ‘error’ is defined as a difference between predicted and observed traffic.

Remember that the probability of choosing a route from a set of choices is calculated with the following formula:

$$P(i|C) = \frac{\exp U_i}{\sum_j \exp(U_j)}$$

Where

$$U \triangleq V + PSC + \epsilon$$

And

$$V = \beta \cdot X$$

The model uses the vector X defined as follows:

$$X = [X_D, X_{ROOS}, X_{BZEE}, X_{BREE}, X_{BOT}]^T \quad (26)$$

Where X_D is the length of the route in km and the other X ’es are dummy variables that are either 1 or 0, depending on whether the obstacle that is referred to is in the route or not.

And β is defined as

$$\beta = [\beta_D, \beta_{ROOS}, \beta_{ZEE}, \beta_{BREE}, \beta_{BOT}]^T \quad (27)$$

The β values represent the penalties associated with respectively the distance, the lock near Rozenburg, the waves and wind at sea, waves and wind near the Breiddiep and the passage of the Botlek Bridge.

For this research there are two problems that one encounters when attempting to calibrate the model parameters.

1. The parameters in $[\beta]$ are not independent. The value of one influences the optimal value of the others. Therefore, calibrating each parameters separately is not possible. See for example the figure below, where it is impossible to calculate optimal values for obstacles 1, 2 and 3 from this data alone. The only value that can be calibrated from this figure is the absolute difference between (1+2) and 3. Therefore, it is necessary to look at multiple Origin-Destination pairs in order to get an accurate approximation of the right value for $[\beta]$.

- The route choice for the port of Rotterdam consists of many Origin-Destination pairs, each with its own optimal set of β values (or optimal value of $[\beta] \triangleq [\beta_1, \beta_2, \dots, \beta_n]^T$). However because of the large number of O-D pairs present in the dataset, only one value of $[\beta]$ is optimal for the whole model.

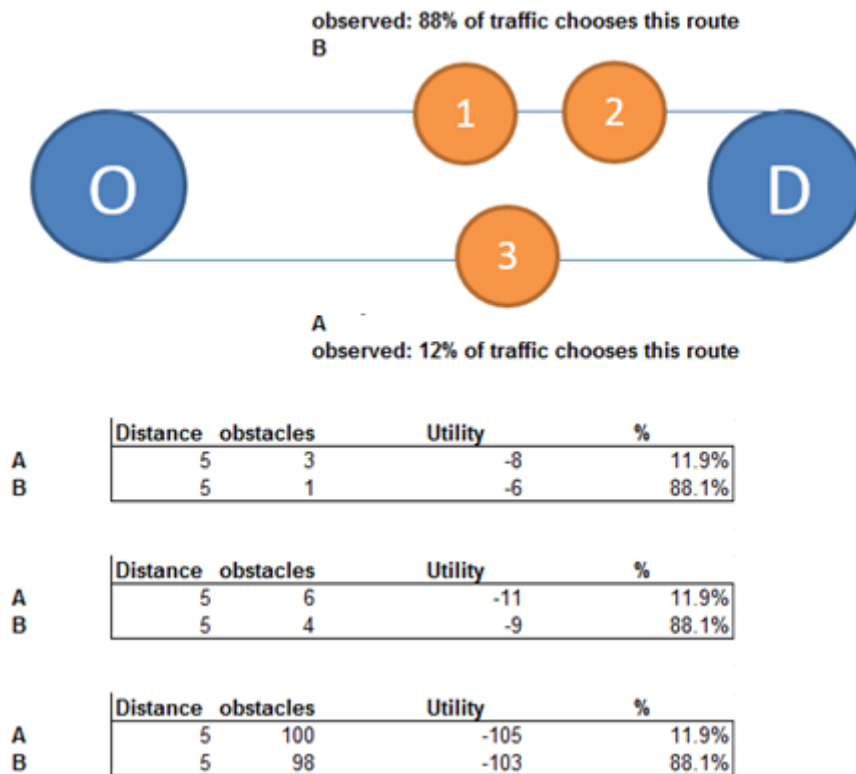


Figure 27 Example of O-D pair with multiple bridges (orange dots) in its routes – and calculation showing that different penalty values can lead to the same model predictions. Which values are correct cannot be determined from this O-D pair alone!

7.1 Assigning a unique penalty to each obstacle

In an effort to calibrate the penalties for the various obstacles discussed in chapter 4, we look for Origin-destination pairs that contain only one of those parameters so we can look at it in an isolated fashion. Ideally, the route choice set should contain only one unknown obstacle so the penalty value can be estimated from the revealed preference of the shipping captains.

Also, the weight factor for the distance will need to be calibrated. This will be done first because there are some routes that do not contain obstacles. This allows for observing the influence of distance in isolation.

7.1.1 Distance

Distance is a part of all routes. As we have seen, the absolute difference in the utilities is used to compute the likelihood of choosing each route. Therefore the distance has to be multiplied by a factor β_D that represents how much influence a distance difference has on the route choice process.

How can we find out what the proper weight factor (β_D) for distance is?

7.1.1.1 Finding a lower boundary value for β_D

Look at EP_BEN – EP_VII. These areas are very close together but there are multiple routes possible between them. Data shows that all vessels take the direct route, which is a straight line from one to the other. The same goes for the O-D pairs EP_BEN – EP_B, EP_BEN – EP_H and EP_H – SPIJK.

To establish a lower bound for β_D we will try the following:

- make β_O equal to β_D .

- set penalties for waves, bridges and locks to 13, or the equivalent of 1 hour of travel at top speed. This is considered an upper bound for the penalties for all obstacles.

- look for the lowest value where model results will approximate observed vessel behavior with at most 5% error.

This happens at the value of 0.2. Therefore, 0.2 is assumed to be the lower bound for β_D .

In this imaginary scenario, all penalties for the alternative routes are very high. If the model still predicts some vessels taking those routes, its weight factor for distance (β_D) is too low.

7.1.1.2 Finding an upper boundary value for β_D

In order to establish an upper bound, we look at the following (theoretical) case.

Consider an origin (O) and a destination (D). There are two routes between them with a length of 20 km (L_A) and 20.5 km (L_B). This difference is chosen because 0.5 km is also the accuracy with which the locations of areas on the map are chosen, or what one might call the ‘resolution’ of the graph used for this research.

A difference of 0.5 km (less than 3 minutes travel time) will not have a noticeable influence on route choice. This should be reflected in the weight assigned to the distance difference. The higher the weight factor assigned to distance, the more influence a small difference in distance will have.

The probability of choosing these routes should be 50% - 50% or values not further removed than 5% from this distribution. This results in an upper bound for β_D of 0.4.

This means:

$$0.2 \leq \beta_D \leq 0.4 \quad (28)$$

7.2 Values of $[\beta]$

Remember from the beginning of this section that the different parts of $[\beta]$ cannot be estimated independently – if the value of one changes, the optimal values of the others will also be different.

There are no Origin-Destination pairs that allow for estimation of boundary values for the remaining β values, the way this was done for β_D . There were also no methods to mathematically determine the value of $[\beta]$ that produces the model with the maximal likelihood. Therefore the ‘brute force’

method was used - many combinations of different parameters were tried and the one with the best model fit was selected. A step size of 0.01 was used for practical purposes.

Because it is known from literature that this kind of maximum likelihood optimization has only one optimum [6]. Therefore one can adjust the values of parameters with a relatively large step size to get an idea of the location of this optimum, then use a smaller step size to approximate the optimal value to a chosen amount of decimals. This approach is illustrated in Figure 28, where the dots represent values that were tested (the curve is interpolated for clarity). Near the optimum a number of values were tried, while further away the step size is larger.

For testing, the value of β_D was fixed to values between 0.2 and 0.4⁴. The penalty values for β_{ZEE} , β_{BREE} , β_{ROOS} and β_{BOT} were adjusted until likelihood started to decrease – a sign that the optimum value had been passed. The maximum likelihood method produces only one optimal $[\beta]$ so there is no need to increase values to search for other optimal values [6]. See the plot in Figure 28 for an example.

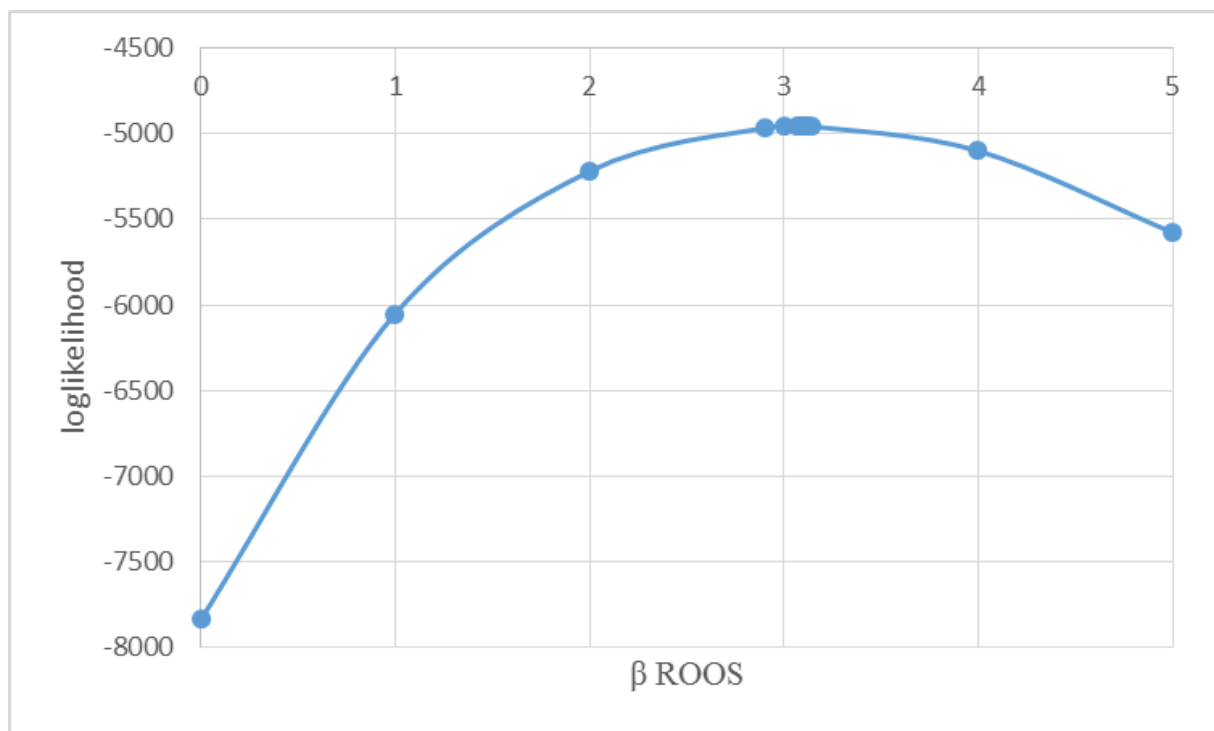


Figure 28 Graph showing the relation between β_{ROOS} and log likelihood. Dots represent data points – the curve was interpolated for better readability.

The optimal values for the different penalties are the following:

⁴ Experiments were also done calibrating the model with $\beta_D = 0.2$ and $\beta_D = 0.5$ but as expected, this resulted in worse model fit than using 0.3 or 0.4 or values in between.

$$\begin{aligned}
\beta_D &= 0.34 \\
\beta_{BREE} &= 0.78 \\
\beta_{ZEE} &= 1.80 \\
\beta_{ROOS} &= 3.11 \\
\beta_{BOT} &= 1.21
\end{aligned}$$

These parameters result in 94.4 % of trips being predicted correctly.

The utility function becomes:

$$U_i = V_i + PSC_{i,C} + \epsilon = \beta \cdot X_i + PSC_{i,C} + \epsilon$$

$$\beta = [0.34, 0.78, 1.80, 3.11, 1.21]^T$$

$$X = [D_i, X_{BREE}, X_{ZEE}, X_{ROOS}, X_{BOT}]^T$$

Here D_i is the distance in km of route i and the various X 'es are dummy variables, which are 1 if the obstacle is included in route i and 0 otherwise.

Table 5 obstacles and the links that contain them

Link number	obstacle
19	Lock (Rozenburgsesluis)
8	Botlek bridge
10	Breeddiep
11	Sea / Lage Licht

8 Validation

In order to see if the route choice model actually works it is important to see if it can be used to make predictions. The traffic between different Origin-destination pairs in the first quarter of 2014 was used to calibrate the model. As an experiment, the route choices of vessels in the 2nd quarter will be forecasted using the calibrated model. This is done by looking at all trips in the 2nd quarter and the origins and destinations, and forecasting the route choices using the Logit model.

This experiment results in 94.58% or 26,748 of 28,259 route choices being predicted correctly. This is an aggregate number that reflects accuracy for the total dataset (containing all O-D pairs).

The graph in Figure 29 shows the number of trips per O-D pair on the x axis and the percentage of trips predicted accurately on the y axis. As can be seen, bad predictions (<80% accurate) occur only for O-D pairs with small amounts of traffic. The dotted trend line shows that the average prediction quality goes up as there are more trips between the O-D pairs.

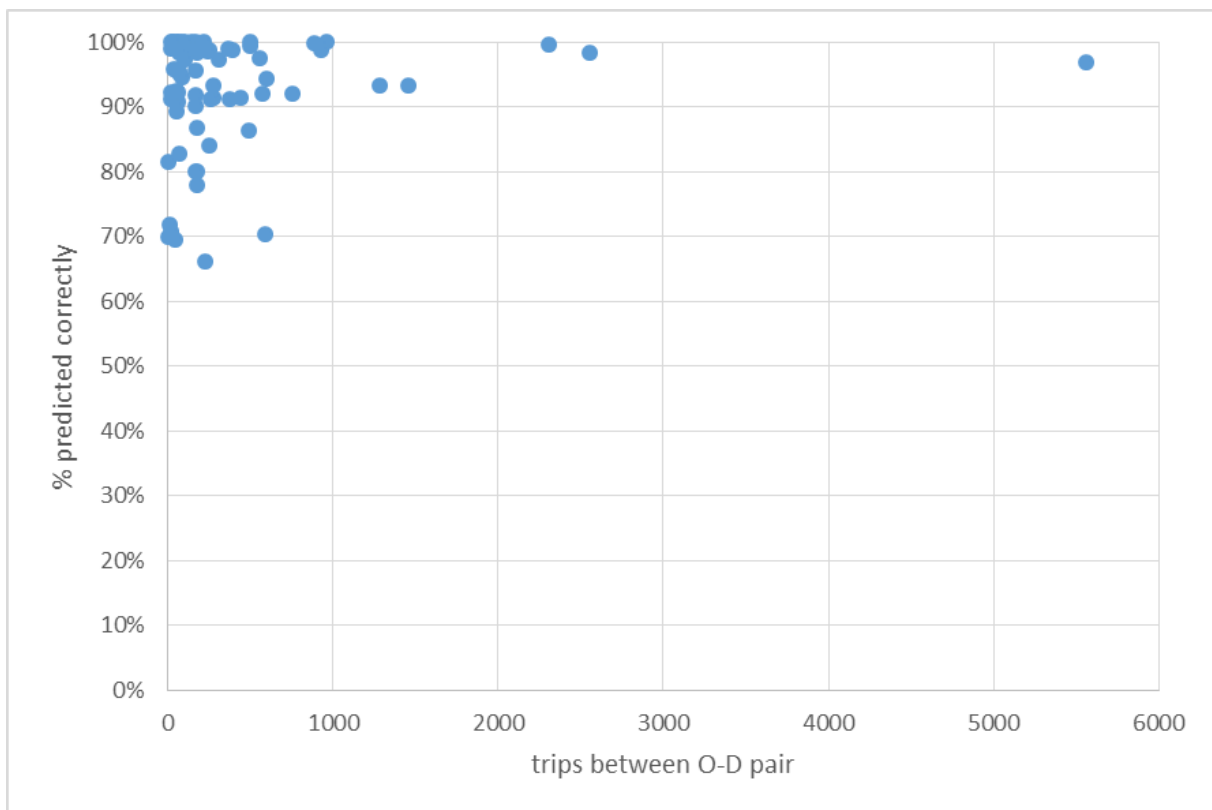


Figure 29 accuracy of prediction vs number of trips per O-D pair – graph containing all O-D pairs

This graph shows that for most Origin – Destination pairs, the model can predict route choice accurately. There are a few O-D pairs, however, for which this is not the case. They will be discussed in section 8.1.1. Errors in prediction are roughly normally distributed with $\mu = -1\%$ and $\sigma = 2.8\%$ (see Figure 30). The reason μ is not 0 is because the probabilities of choosing each route is calculated using an exponential function which can approximate, but never be equal to 0.

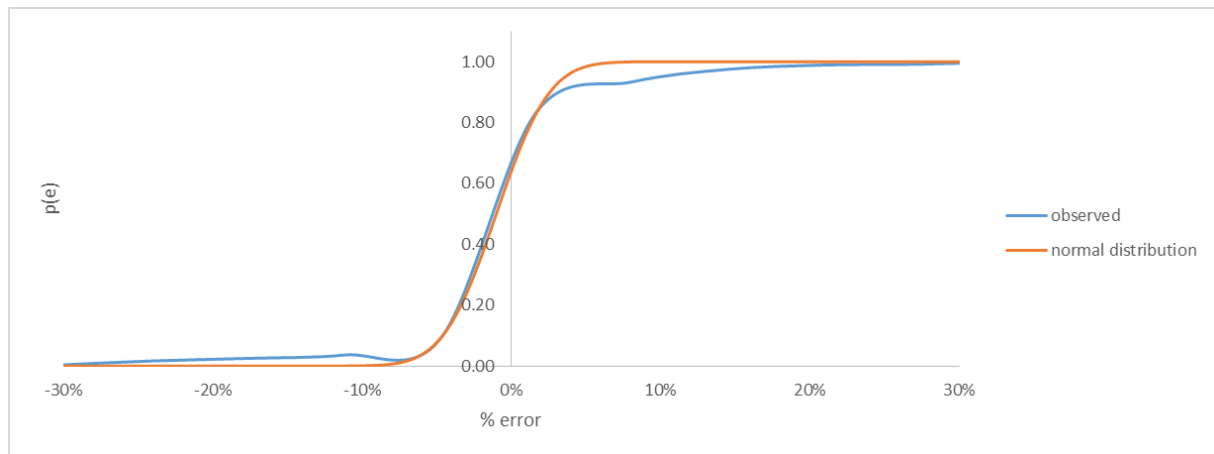


Figure 30 Comparison between normal distribution ($\mu = -0.01$, $\sigma = 0.03$) and observation

8.1.1 Comparison between model and historical data

If route choice data from the 1st quarter was used to predict traffic in the 2nd quarter, 97.48% of traffic would have been predicted correctly. Computing log likelihood for historical data is mathematically impossible. Figure 31 shows the accuracy vs amount of trips per O-D pair.

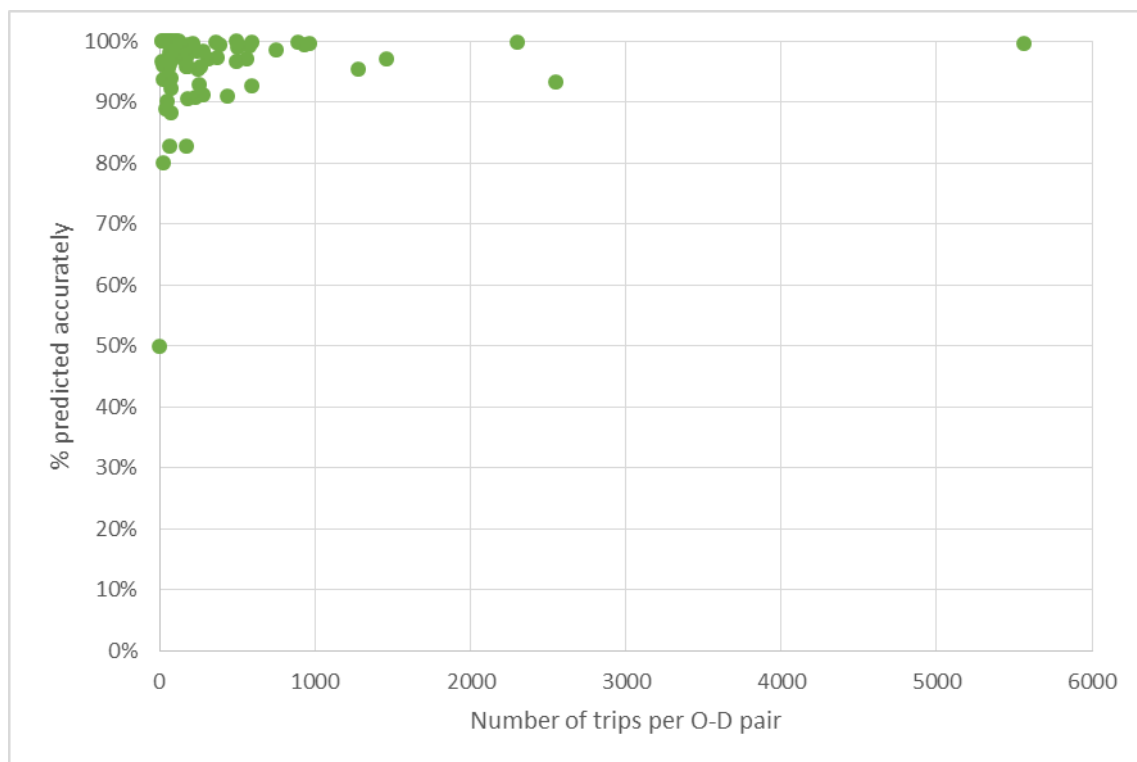


Figure 31 Accuracy of predictions vs number of trips - predictions made using historical data

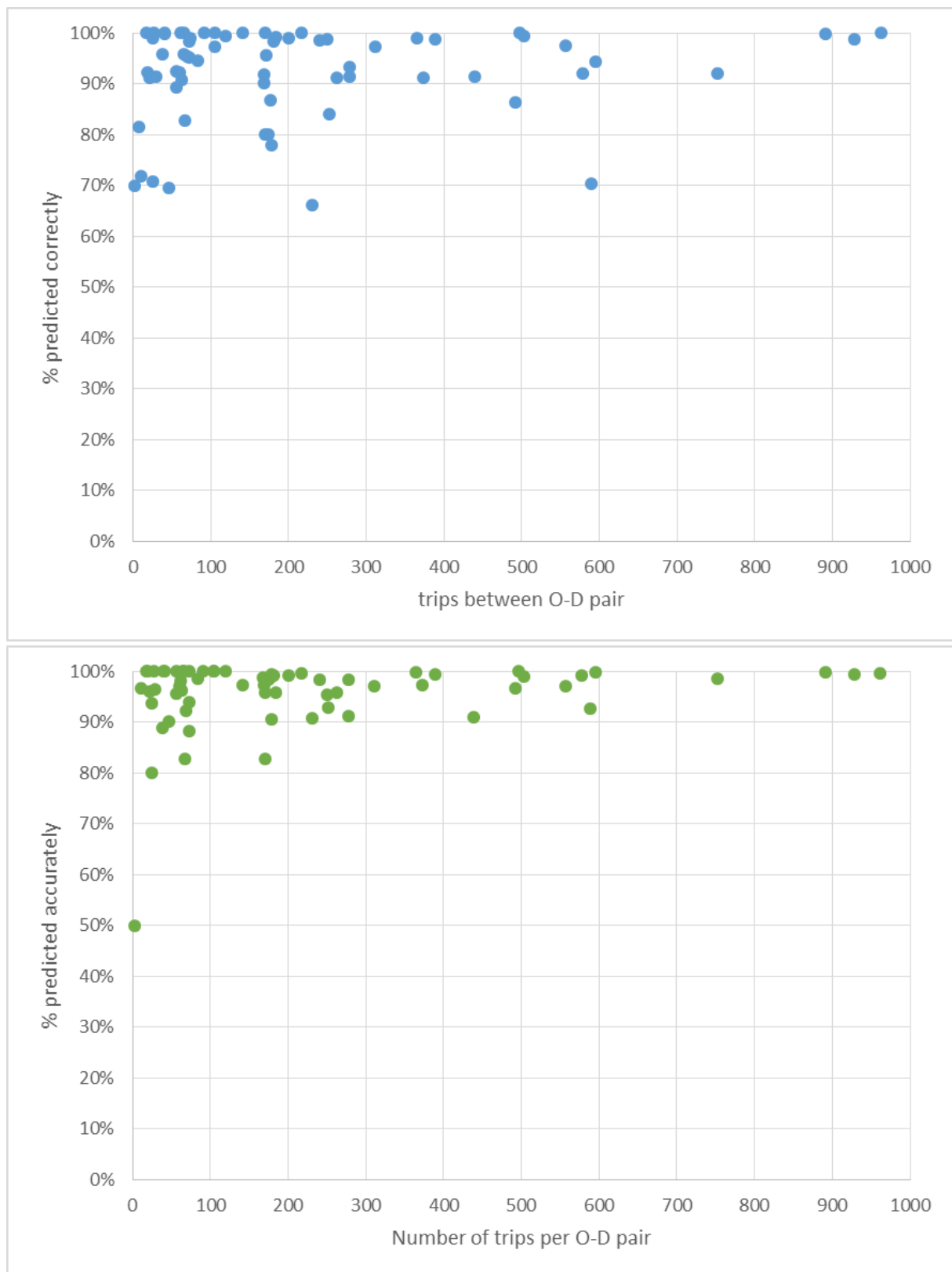


Figure 32 Comparison between model (left) and historical data (right) (outliers with # trips>1000 removed for readability)

The comparison shows a model is not better than simply using historical data to predict future traffic. It is however much more useful when trying to predict situations for which no historical data is available.

8.1.2 O-D pairs predicted with less than 90% accuracy

In order to investigate the predictive qualities and shortcomings of the model, the origin – destination pairs for which less than 90 % of choices are predicted correctly are investigated. They are shown in Table 6. Table 7 shows those pairs which also have more than 100 trips between them, making it likely the bad prediction is not just due to lack of data.

We will look at some factors and try to determine if there is a common cause for bad predictions.

8.1.2.1 Randomness

First, let's look at the data for the first and second quarters. Are the errors just caused by randomness? Table 7 shows accuracy in q1 and q2 are strongly correlated. The O-D pairs which are predicted with low accuracy also have a relatively poor fit in the calibration dataset. This suggests randomness is not the main cause of the poor predictions.

Looking at the amount of different MMSI-numbers per O-D pair tells us that the bad predictions are not caused by behaviour of a few vessels. Many different vessels travel between these O-D pairs. Therefore, random variations in the choices of individuals also cannot explain poor prediction accuracy.

Table 6 O-D pairs predicted with less than 90% accuracy

O-D pair	% predicted correctly	Number of trips	Historical prediction accuracy
PARK - MV	68%	11	97%
EP_B - WEH	70%	46	90%
EP_C - EP_H	71%	25	80%
EP_VII - WEH	72%	231	91%
BV - EP_VII	75%	589	93%
PARK - BV_BRIT	75%	2	50%
PARK - EP_C	78%	7	#N/A
BV - EP_B	78%	178	91%
EP_BEN - SPIJK	80%	492	97%
BV_BRIT - WEH	81%	177	99%
EP_B - BRIEN	82%	170	83%
NO - EP_B	83%	67	83%
HART_OM - EP_BEN	84%	63	96%
EP_VII - NO	86%	174	98%
EP_BEN - BV_SEIN	86%	29	96%
HART_OM - EP_C	86%	56	100%
BV - EP_C	86%	252	93%
BV_BRIT - NO	86%	60	97%
NO - MV	88%	439	91%
EP_C - SPIJK	88%	578	99%

PARK - EP_BEN	89%	19	100%
BV_BRIT - BRIEN	89%	69	92%

Table 7 O-D pairs predicted with less than 90% accuracy and more than 100 trips between them

O-D pair	% predicted correctly (q2)	% overlap model and data (q1)
EP_VII - WEH	72%	73%
BV - EP_VII	75%	76%
BV - EP_B	78%	82%
EP_BEN - SPIJK	80%	86%
BV_BRIT - WEH	81%	87%
EP_B - BRIEN	82%	65%
EP_VII - NO	86%	79%
BV - EP_C	86%	79%
NO - MV	88%	90%
EP_C - SPIJK	88%	92%

8.1.2.2 Historical data

Historical data from the 1st quarter of 2014 has a few origin-destination pairs which predict behaviour in the 2nd quarter of 2014 poorly (accuracy < 90%). Some of those are also poorly predicted by the model, which was calibrated based on data from 2014 1st quarter. Others are not. Some O-D pairs which are predicted poorly by historical data are predicted fine with the model, and vice versa. Therefore it is concluded that the historical sample used for calibration is not the cause of bad model predictions.

8.1.2.3 Cargo type or vessel size

Perhaps the general model would predict certain O-D pairs poorly is the traffic was only caused by a certain cargo or vessel type.

For example, traffic between EP_VII and BV is almost exclusively tank or bunker vessels. If models for tank and bunker cargo are used to predict this traffic, the accuracy is 81 % as opposed to 75% using a general model. For the EP_VII and WEH, most traffic is caused by bunker vessels. The bunker traffic is predicted with 88% accuracy, as opposed to 72% for all traffic using the general model.

The O-D pair MV – NO is visited by vessels carrying all sorts of cargo, but predominantly small and medium sized ones. Using the small and medium size models to predict cargo results in respectively 88% and 89%, the same as the one using the general model. No other O-D pairs in the list have specific vessel sizes that visit them.

This shows that some inaccuracy can be explained by the influence of cargo type on route choice.

8.1.2.4 Graph

The port area is modelled as a graph. This is a simplification of reality which might lead to prediction errors. For example, all O-D pairs involving the area 'EP_VII' (7e Petroleumhaven in the Caland channel, see Figure 4) suffer from bad predictions because they underestimate the amount of traffic

which will go through the Breddiep and Nieuwe Waterweg. This could indicate a problem with the composition of the graph which leads in turn to incorrect calibration and therefore incorrect predictions. There is no way of checking if this is indeed the case short of changing the graph and comparing results, which is not feasible due to time constraints.

8.1.2.5 Summary

Errors in prediction could be due to small amounts of traffic, influence of cargo type or errors in the graph. They could also result from other, undiscovered flaws in the model. The study of these poor predictions suggests some interesting areas for further research, which will be discussed in chapter 11. From studying the errors it is observed that while the MNL model can yield good predictions for route choice, it is not perfect. Also keep in mind that the percentage of correct predictions is a performance indicator not directly proportional to maximum likelihood, which was used to calibrate the model.

8.2 Testing predictive capabilities of the route choice model

In order to see if the route choice model can be used to predict traffic for situations other than the exact same one it was based on, an experiment was conducted. The calibrated version of the route choice model was used to predict traffic in the Port of Rotterdam for a period where there was a large maintenance project going on at the Rozenburgsesluis, the lock and bridge that separate the Caland channel from the Hartel Channel. This meant that the lock was closed and no ships could pass.



The screenshot shows the Port of Rotterdam website interface. The top navigation bar includes links for Home, Nieuws, Business, **Scheepvaart** (highlighted), Haven, and Havenbedrijf. A language selector shows 'nl | en | de'. Below the navigation bar, a breadcrumb trail reads: Home > Scheepvaart > Actueel > Bekendmakingen > Calandkanaal / Hartelkanaal- Rozenburgsesluis.

The main content area is titled 'Calandkanaal / Hartelkanaal- Rozenburgsesluis'. It includes a 'Geen bediening' (No service) notice with the following details:

- Nr.** 068/2014
- Van** 30-6-2014
- Tot** 6-7-2014

The notice text states: 'Geen bediening' and '1. Vanaf maandag 30 juni 07:00 uur tot zondag 6 juli 18:00 uur zullen in verband met dringende werkzaamheden de Rozenburgsesluis en bruggen niet worden bediend.' It also mentions that the shipping schedule will be affected and provides contact information for the HAVENMEESTER van ROTTERDAM (Tevens Rijkshavenmeester regio Rotterdam-Rijnmond) as of 20 mei 2014.

A sidebar on the left under 'Actueel' lists various categories: Bekendmakingen, Aankomst & Vertrek, Hydrometeo, Webcams, Zeevaart, Binnenvaart, Recreatievaart, Nautisch beheer Zeehaven Dordrecht, Wet- en regelgeving, Communicatie en VTS, Havengelden, DHMR visie, and Contact & Service.

Figure 33 - Note in the Port of Rotterdam website announcing the closing of the lock for maintenance [18]

The lock near Rozenburg was supposed to be closed from the 30th of June until the 6th of July 2014. However, AIS data shows that the first ships already started crossing the lock on the afternoon of the 5th of July. Probably the works on the lock and bridges took less time than was scheduled. Therefore this experiment uses data from the 30th of June, 07:00 AM to the 4th of July, 23:59 AM.

The general route choice model was used to predict the flow of traffic while the lock was closed. This was done by setting the penalty for passage of the lock to 1000, which is practically equal to infinity. The predictions made with the model predicted 95.24% of the trips accurately. This is roughly the same level of accuracy as when the model is used for predicting the normal traffic flows.

This example suggests the route choice model is able to accurately forecast the changes in traffic that result from changes in the port infrastructure.

9 Influence of cargo type and ship size on route choice

In order to investigate the influence of cargo type and ship size on route choice, the choices made by vessels of a certain size and type were isolated. The values of $[\beta]$ were calibrated again using only those samples. Results are shown in Table 8. The tables contain the calibration results for $[\beta]$ and also the loglikelihood value, which is computed with formula 24.

The log likelihood values show that the models calibrated for specific cargo types produce predictions with a higher likelihood compared to predictions made using the general model. The exception is the breakbulk traffic. This is probably because there is so little breakbulk traffic recorded that calibration produces unreliable results.

Table 8 beta values and log likelihood using the general model and specific calibrations per cargo type

cargo type	Distance	bree	zee	roos	bot
general model	0.34	0.78	186	3.11	1.21
tank	0.32	0.61	1.16	2.51	1.33
container	0.35	0.73	2.5	3.66	1.08
bulk	0.46	1.78	3.49	4.4	0.92
bunker	0.27	0.00	0.35	1.74	1.22
breakbulk	0.44	1.08	7.25	2.96	1.10

cargo type	number of trips in q2	Loglikelihood standard model	Loglikelihood specific model
tank	5871	-1111.22	-1090.16
container	9173	-1855.99	-1829.67
bulk	3810	-411.78	-384.343
bunker	4797	-1095.07	-996.209
breakbulk	557	-67.3291	-74.8742

Table 9 beta values and prediction accuracy for different ship sizes

Vessel size	Distance	bree	zee	roos	bot
small	0.33	0.38	2.60	2.95	1.21
medium	0.36	1.21	2.08	3.30	1.51
large	0.33	0.57	0.79	2.88	1.02
push barge	0.33	2.15	3.89	3.45	1.15

Vessel size	number of trips in q2	Log likelihood standard model	Log likelihood specific model
Small	9947	-1518.47	-1465.90
Medium	11130	-2006.15	-2003.68
Large	4803	-1100.86	-1036.43
Push barge	2375	-287.53	-276.44

An example of the difference between a general model and a more specific one is shown below. The charts in Figure 34 show the accuracy of predictions per O-D pair for predictions made using a specific model for push barges (left) and the general model applicable to all traffic (right). Note that the amount of trips predicted right and the least squared difference between prediction and data are not necessarily achieved at the same value of $[\beta]$. The percentage of trips predicted correctly is not a scientific measure of the quality of the model but rather an indication of the real-world applicability. Nevertheless, it is interesting to see what amount of route choices is predicted correctly.

The graph shows that using the general model results in more O-D pairs with relatively low prediction accuracy. The cargo type-specific model can be said to perform better than the general model. It also has a higher likelihood value (see Table 9), though not very much higher (note the rounded numbers in the 1st and 2nd columns from the right of the table are the same). The same general result is obtained for other sizes and cargo types. The exception here is breakbulk, which is predicted better with the general model than with the specific model.

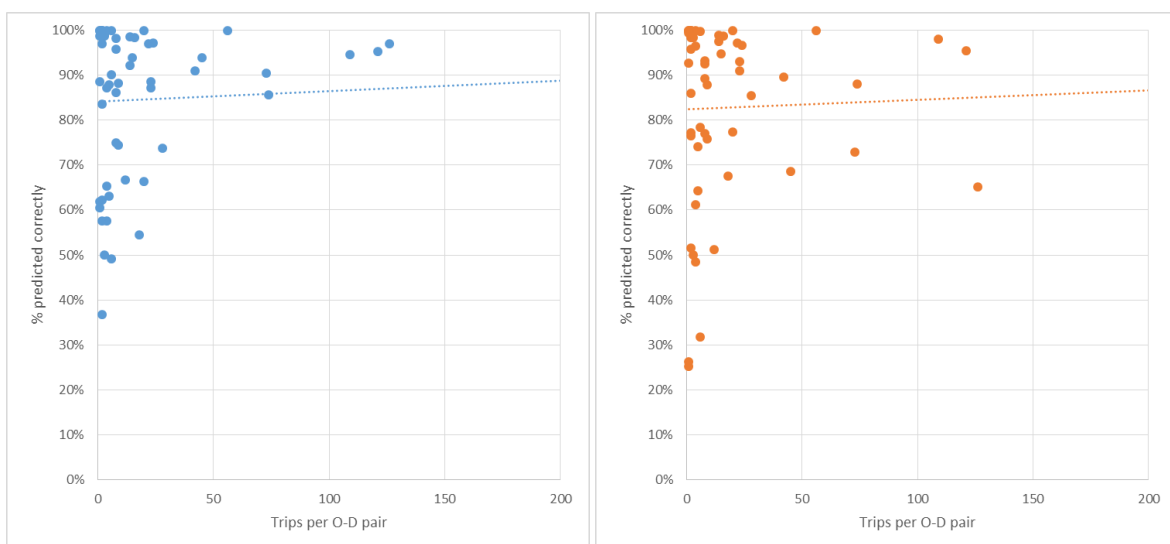


Figure 34 Accuracy vs number of trips for the push barge subset using calibration for push barge (left) and general model (right).

The main difference between the models for the separate cargo types are the penalties assigned to sea and Breddiep due to waves and wind. If one wants to predict traffic for these areas specifically it is recommended to use the cargo type-specific or size-specific models.

9.1.1 Difference in route choice preference per vessel or cargo type explained

The main interesting point here is that penalties for passing the zee and Breddiep areas vary significantly between vessel and cargo types. Small vessels and vessels with solid cargo (bulk, breakbulk, or containers) are much more likely to avoid the wind and waves. Small vessels will experience more severe tilt from high waves, so this is understandable. Vessels which have their cargo exposed to sea water (container and dry bulk carriers) will likely want to avoid exposure to seawater as this corrodes the vessel and cargo. Also, vessel movement due to waves can cause cargo to start moving or even fall overboard. Therefore dry cargo vessels, too, will be more likely to avoid the waves near the sea end Breddiep.

Splitting the dataset into multiple subsets means the amount of data available for calibration is much smaller. Therefore results will be less reliable than those obtained using the whole data set. This is especially true for categories with small amounts of traffic, such as 'breakbulk'.

9.1.2 Conclusion

Cargo type and ship size do influence route choice. Models calibrated for a specific cargo type or specific vessel size have a higher log likelihood than a general model applicable to all vessels.

Whether or not it is advantageous to use the models per cargo type or ship length mainly depends on what one wants to predict. If route choice for the port area is predicted the general model functions very well. If one is interested in the passage of specific obstacles like the zee and Breddiep area, using the models calibrated per cargo type or ship length might be advantageous. This can be seen in Figure 34, where the graph of results using the general model shows more O-D pairs with very low levels of prediction accuracy even though the overall accuracy percentage is not that different.

It should be noted that for many trips, cargo type and/or vessel size are not registered. It is also known that not all cargo types and vessel sizes are registered correctly. This could potentially distort the calibration process and means that not all traffic can be predicted with cargo- or size-specific models. Using general model will therefore always be necessary for modelling at least parts of traffic.

10 Results and conclusions

10.1 Results

The goal of this investigation was to develop a model for predicting route choice of inland shipping vessels in the Port of Rotterdam. This model has been developed and tested. It was able to predict the validation dataset with 95% accuracy.

10.2 Conclusions

The conclusions from this investigation are the following:

- logit models are able to predict route choice of inland shipping vessels accurately
- Main drivers for route choice are length of the route and presence of bridges and locks, waves and wind and shallow / narrow waterways. Models of different sizes and cargo types will react in different ways to these factors.
- Using separate models to forecast route choice based on cargo type or vessel size produces slightly more accurate predictions. This mainly concern vessels that can choose routes over sea or the Breiddiep.

10.3 Answers to research questions

Sub questions:

1. **How can the Port of Rotterdam area be modelled?**
Using graph theory, a simplified model of the port of Rotterdam is obtained (Figure 5)
2. **How can the route choice process be modelled?**
Using a multinomial Logit model with Path Size Correction, the route choice process can be modelled. This discrete choice model can be used to predict the probability of choosing a route for each ship and route.
3. **What factors have an observable influence on route choice of inland shipping vessels?**
Distance, wind and waves, locks, bridges, narrow/shallow waterways.
4. **How can traffic data be used to calibrate a route choice model?**
Using Python, Excel and traffic data from the first quarter of 2014, the route choice model is calibrated so its prediction has a maximal likelihood of producing the outcome observed in the traffic data.
5. **How does the modelled behaviour compare to traffic data from the Port of Rotterdam?**
The route choice model predicts route choice in the 2nd quarter of 2014 with roughly 95% accuracy. Route choice for unusual situations, such as the closing of the Rozenburgsesluis, is also predicted with roughly 95% accuracy.
6. **How can the route choice model be applied to the prediction of inland shipping traffic in the Port of Rotterdam?**
Using the route choice model produces forecasts which are nearly as accurate as using predictions based in historical data (95.0% vs 96.7 %). However the choice model also yields accurate predictions for situations where the infrastructure of the port is altered, something

which historical data cannot do. This is illustrated by predicting traffic for the period where the lock near Rozenburg was closed for maintenance.

11 Suggestions for further research

Researching route choice for inland shipping vessels sheds light on other interesting subjects for research. There are also ways in which this model could be improved and expanded upon.

11.1 Investigate applicability to other situations

It would be very interesting to build a version for another port area. Possibilities include nearby ports such as Amsterdam, Antwerp or Hamburg. Also it would be interesting to see if the model can be used to predict longer trips from the hinterland to the port. This would probably shed some more light on the influence of tides and currents on route choice.

11.2 Divide port map into smaller areas

The current way of dividing the map of the port of Rotterdam into areas (see Figure 3) is very rough. This undoubtedly produces inaccuracies when modelling route choice. The current areas Botlek – Vondeligenplaat, Noordoever, Maasvlakte, and Europoort are very large and the model would probably benefit from splitting them into smaller areas. Europoort was divided into three sub-areas to improve the model for this investigation; the other areas were not. This division resulted in much increased accuracy, which suggests that similar increase in model detail for other areas would also be beneficial.

11.3 Look into the whole visit to the port area, not just trips from 1 origin to 1 destination

This report concerns only trips by vessels in the port of Rotterdam between a single origin and destination. It is likely that looking at the entire visit to the port area and the order and direction of those trips would lead to more insight into route choices of inland shipping traffic. The choice and sequence of destinations is assumed to be a given for this research; in reality, this is not the case. Travel plans can be, and are, adapted mid-journey in reaction to changing circumstances.

References

- [1] E. Freijinger, "Route Choice Analysis: Data, Models, Algorithms and Applications," Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2008.
- [2] A. Ghareib, "Evaluation of Logit and Probit Models in Mode-Choice Situation," *Journal of Transportation Engineering*, vol. 122, no. 4, pp. 282-290, 1996.
- [3] C. Chorus, T. Dekker and S. van Cranenburgh, "Towards more behaviourally robust travel demand forecasts - catering to utility maximisers and regret minimizers," in *Proceedings of the 1st International Conference on Uncertainties in Transport Infrastructure Evaluation*, Copenhagen, Denmark, 2013.
- [4] C. G. Prato, "Route Choice Modelling: Past, Present and Future Research," *Journal of choice modelling*, vol. 2, no. 1, pp. 65-100, 2009.
- [5] F. Koppelman and V. Sethi, "Closed-form Discrete Choice Models," in *Handbook of Transport Modelling*, Elsevier Science , 2000, pp. 235-252.
- [6] J. d. D. Ortúzar and L. Willumsen, *Modelling Transport*, 4th Edition ed., Wiley, 2011.
- [7] E. Cascetta, A. Nuzzolo, F. Russo and A. Vitetta, "A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and Some Calibration Results for Interurban Networks," in *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, 1996.
- [8] S. Bekhor, M. Ben-Akiva and M. Scott Ramming, "Adaptation Of Logit Kernel To Route Choice Situation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1805, pp. 78-85, 2002.
- [9] E. Frejinger and M. Bierlaire, "Capturing Correlation with subnetworks in route choice models," *Transportation Research part B*, no. 41, pp. 363-378, 2007.
- [10] M. Fosgerau, E. Frejinger and A. Karlstrom, "A link based network route choice model with unrestricted choice set," *Transportation Research Part B*, vol. 56, pp. 70-80, 2013.
- [11] M. Ben-Akiva, M. Bierlaire and E. Frejinger, "Expanded path size attribute for route choice models including sampling correction," in *Proceedings of the International Choice Modelling Conference*, Harrogate, UK, 2009.
- [12] M. Bliemer, "Impact of Route Choice Set on Route Choice Probabilities," the University of Sydney, 1st April 2008. [Online]. Available: http://sydney.edu.au/business/___data/assets/pdf_file/0015/27150/mikeb08.pdf. [Accessed 02 18 2015].

- [13] P. Bovy and S. Bekhor, "The Factor of Revisited Path Size: Alternate Derivation," *Transportation Research Record*, vol. 2076, pp. 132-140, 2008.
- [14] D. Leinweber, "Stupid Data Miner Tricks - Overfitting the S&P 500," *The Journal of Investing*, vol. 16, no. 1, pp. 15-22, 2007.
- [15] T. Vigen, "Spurious Correlations," [Online]. Available: <http://www.tylervigen.com/>. [Accessed 24 02 2015].
- [16] J. Brolsma, "Rapportage Containerhoogtemetingen," Rijkswaterstaat, Driebergen, 2013.
- [17] J. Mark and M. Goldberg, "Multiple Regression Analysis and Mass Assessment: A review of the issues," *Appraisal Journal*, vol. 56, no. 1, p. 89, 1988.
- [18] Port of Rotterdam, [Online]. Available: <http://www.portofrotterdam.com/nl/Scheepvaart/actueel/Bekendmakingen/Pages/calandkanaal-hartelkanaal-rozenburgsesluis.aspx>. [Accessed 06 27 2015].
- [19] J. Brolsma, "Enquête naar de routekeuze van de binnenvaart in de haven van Rotterdam," Rijkswaterstaat, Rotterdam, 2005.
- [20] T. Thomas and B. Tutert, "Een nieuwe kijk op routekeuze? Bevindingen uit het kentekenonderzoek Enschede," Enschede, 2009.
- [21] J. Mayer, K. Khairy and J. Howard, "Drawing an elephant with four complex parameters," *American Journal of physics*, vol. 78, no. 6, p. 648, 2010.

Appendices

Appendix 1 – Research paper

Appendix 2 – Example – analysing route choice

Appendix 3 – Tides and currents in the Port of Rotterdam area

Appendix 4 – Questionnaire

Appendix 1

Modelling route choice of inland shipping vessels in the Port of Rotterdam

E. Gerritse, BSc., H. Zheng, MSc., Dr.Ir. R. Negenborn, Prof.Dr.Ir. G. Lodewijks

Department of Maritime and Transport Technology, Delft University of Technology, Delft, the Netherlands.

1. Introduction

The Port of Rotterdam is one of the busiest ports in the world and Europe's largest sea port. Each year, tens of thousands of ships visit the port. This number is expected to grow in the foreseeable future. In order to make sure the port remains accessible, meticulous planning is necessary. The creation of realistic scenarios for future traffic flows is a requirement for planning the future of the port area. Engineering company Witteveen + Bos is working in close cooperation with the department of Capacity Management at the Rotterdam Port Authority to create scenarios for predicting traffic flows in the port area. Based on general predictions about future cargo volumes and modal split, a model of traffic flows in the port area is being created. Part of this model details the inland shipping sector, an important sector for hinterland transportation.

Currently the inland shipping traffic model does not incorporate a model for route choice. Given the fact that a vessel will travel from a certain location A to another location B, there is no model that predicts what route that vessel will take. Instead, historical data is used to predict future traffic flows. This report contains the results of research conducted with the goal of creating a model for route choice of inland shipping vessels. The following research questions have been answered: **Main research question:** How can route choice of inland shipping vessels, given the origin and destination of the vessel and network topography, be modelled?

Sub questions:

How can the route choice process be modelled?

How can the Port of Rotterdam area be modelled?

What factors have an observable influence on route choice of inland shipping vessels?

How can traffic data be used to calibrate a route choice model?

How does the modelled behaviour compare to traffic data from the Port of Rotterdam?

How can the route choice model be applied to the prediction of inland shipping traffic in the Port of Rotterdam?

The structure of the paper is as follows: section 2 will outline the research method, section 3 shows the results of the research, section 4 contains conclusions based on those results, and section 5 contains suggestions for further research.

2. Research method

In order to answer the research questions posed in section 0, the following research steps were taken.

2.1 Determining what factors influence route choice

In order to determine what influences route choice, a questionnaire was sent out to inland shipping captains. 76 of them replied and listed factors that influenced their route choice. They also picked one or more factors they thought would be influential from a pre-prepared list. This resulted in the following factors to be studied:

- Distance
- Tides
- Currents
- Wind
- Waves
- Bridges
- Locks
- Presence of berths
- Speed limits
- Narrow / shallow waterways
- Congestion
- Cargo type
- Vessel size

No evidence for the influence of congestion, tides and currents, speed limits and berths was found for the specific area of the port of Rotterdam. Though these are all real obstacles, they either do not apply to the port area (congestion, lack of berths) or are present in all major waterways in roughly equal amounts (speed limits, tides and currents). From all bridges in the port area, only the Botlek Bridge is specifically identified as being an obstacle by inland shipping captains. This leaves the following factors to be included in the model:

- Distance
- Wind and waves at sea and the Breiddiep
- The Botlek bridge
- The lock near Rozenburg
- Narrow / shallow waterways at the Breiddiep
- Cargo type
- Vessel size

2.2 Modelling the route choice process

Many different ways of modelling choice exist in literature [1] [2]. This research uses the Multinomial Logit (MNL) method because it is simple, commonly used in practice and has a closed-form mathematical formulation, making it relatively easy to compute and estimate. To biggest problem with MNL models is that they cannot model partially overlapping routes. To address this problem the Path Size Logit method is used [3].

The multinomial logit model works as follows. First, a choice set (C) is created that contains all routes between origin (O) and destination (D). (More on this in section 2.3). For route i a utility (U_i) is computed with the formula:

$$U_i \triangleq V_i + PSC_{i,C} + \epsilon \quad (1)$$

Where

- V = Route utility, see section 2.3.3.
- $PSC_{i,C}$ = Path Size Correction factor, see section 2.3.2
- ϵ = Random factor (GEV distribution with location 0 and scale 1) [4]

The odds of choosing route i from the available routes in choice set C are computed with the following formula:

$$P(i|C) = \frac{\exp(U_i)}{\sum_{j \in C} \exp(U_j)} \quad (2)$$

2.3 Modelling the Port of Rotterdam

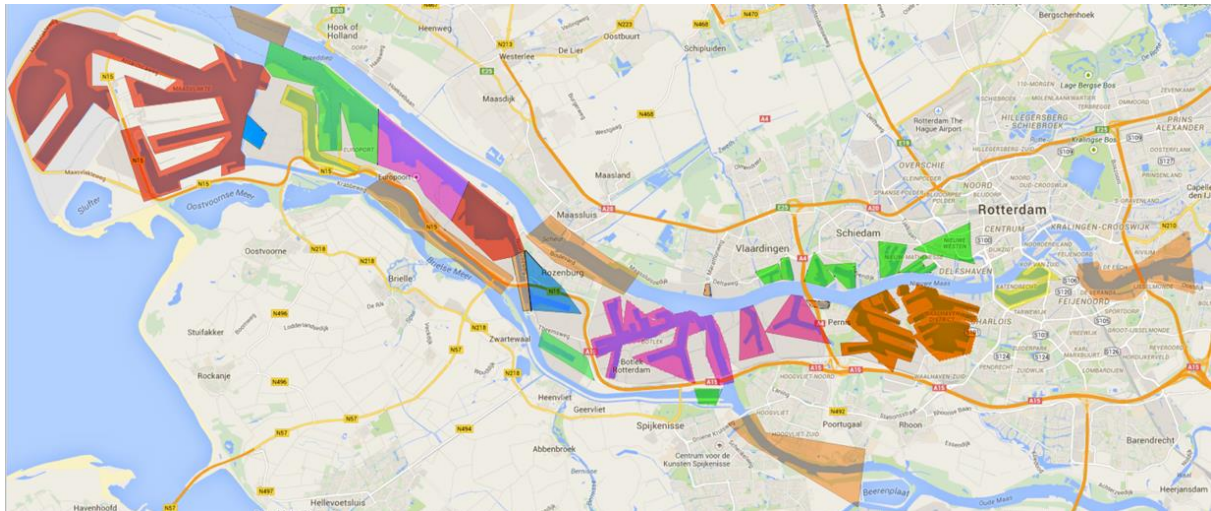


Figure 1 Port of Rotterdam divided into areas

Witteveen + Bos and the Port Authority already use a model of the port of Rotterdam area. Here, the port is divided into a number of areas (see the coloured areas in Figure 1). The areas represent port basins where terminals and berths are located and which can be origins and destinations for traffic. The port of Rotterdam has to be modelled in such a way that route choice sets can be determined and each route can be assigned a value for utility U . Also, the representation of the port area needs to be at least roughly compatible with work done earlier by the Port of Rotterdam Authority and Witteveen + Bos. Graph theory is used to accomplish this, yielding the graph in Figure 2. Graphs consist of nodes (circles, representing locations in the port) and links (lines connecting nodes, representing the waterways).

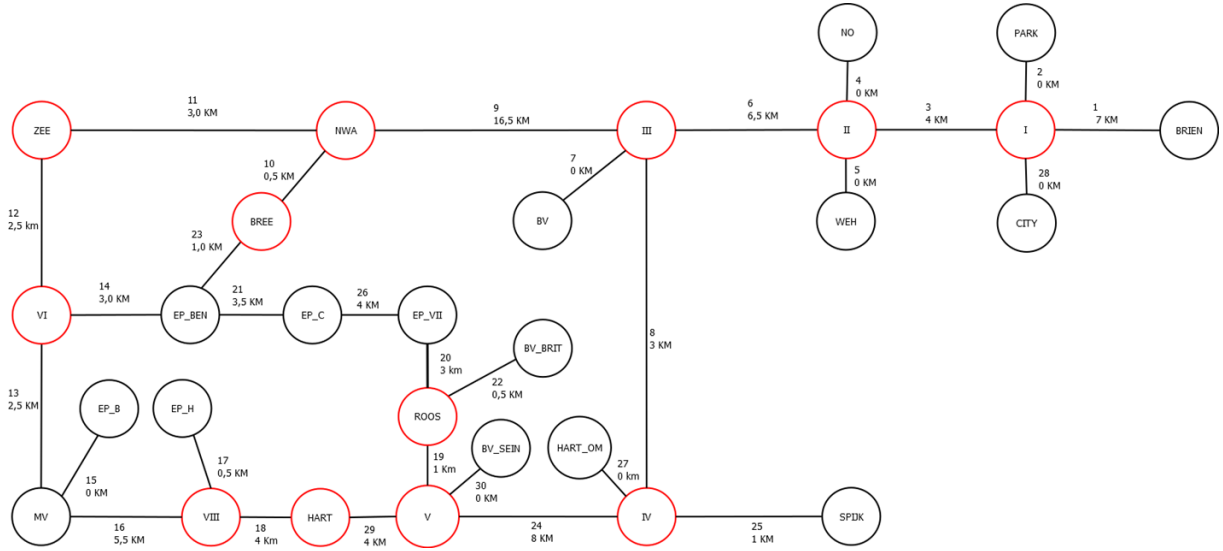


Figure 2 Graph representing the Port of Rotterdam area

2.3.1 Creating the choice set

Because the graph is relatively simple, it is feasible to analyse all possible routes between a given origin and destination on the condition that one only includes routes that do not feature a link more than once¹. We call this set the Universal Choice Set (\mathcal{M}). From these routes, only the ones are selected for feasible choice set \mathcal{C} that do not contain detours. A route is considered to contain a detour if it contains all nodes present in another route in \mathcal{M} . Only the choice set containing feasible paths is used because research has shown outcomes of Logit models are very easily distorted by including non-feasible alternatives in the choice set [5].

Choice set $\mathcal{C} \subseteq \mathcal{M}$ contains what are deemed feasible paths. Path $i \in \mathcal{C}$ if the following condition holds:

$$\sum_{j \in \mathcal{M}} Z_{ij} = 0 \quad (3)$$

With

$$Z_{ij} := \begin{cases} 1 & \text{if } \Lambda_i \supset \Lambda_j \\ 0 & \text{if } \Lambda_j \not\supset \Lambda_i \end{cases}$$

Where

Λ_i = set of nodes that form route i .

2.3.2 Path size correction

The Path size correction (PSC) factor is attempts to model partial overlap of paths in choice set \mathcal{C} . The PSC factor is computed based on the portion of route i that overlaps with other routes. It is computed as follows.

$$PSC_i = -\sum_{a \in \Gamma_i} \frac{l_a}{L_i} \ln \Delta \quad (4)$$

¹ Without this condition, there is an infinite number of possible routes between every origin and destination.

Where

Γ_i = set of links in route i

l_a = length of link a

L_i = length of route i

$$\delta_{ai} := \begin{cases} 1 & \text{if } a \in \Gamma_i \\ 0 & \text{if } a \notin \Gamma_i \end{cases}$$

$$\Delta = 1 / \sum_{i \in C} \delta_{ai}$$

2.3.3 Assigning Utility to a route

The utility U is a penalty assigned to a route that represents its attractiveness relative to the other options in the choice set. A penalty is assigned to each route based on the length of all links in the route. Additional penalties are added if the route includes the links in Table 1. Link numbers refer to the graph in Figure 2.

Table 1 Links receiving additional penalty

Link number	Location	Reason for additional penalty
10	Breeddiep	Narrow waterway / wind and waves
11	Sea / Lage Licht	Wind and waves
19	Lock near Rozenburg	Lock
8	Botlek bridge	Bridge

The penalty based on distance and each of the obstacles are multiplied by a weight factor β which represents how much each property of a route influences the route choice. The utility is computed as follows:

$$U_i = V_i + PSC_{i,C} + \epsilon \quad (5)$$

With the following definitions for the parts of equation 5.

$$V = [\beta] \cdot [X] \quad (6)$$

$$X = [X_{Distance}, X_{BREE}, X_{ZEE}, X_{ROOS}, X_{BOT}]^T \quad (7)$$

$$\beta = [\beta_{Distance}, \beta_{BREE}, \beta_{ZEE}, \beta_{ROOS}, \beta_{BOT}]^T \quad (8)$$

2.4 Calibrating the Logit model

Now that the model equations have been established, calibration is necessary to ensure the model resembles reality as closely as possible. The weight factors in β have been calibrated to match a traffic data set using the maximum (log) likelihood principle [2]. The dataset used for calibration contains all trips between the origins and destinations in Figure 1 during the first quarter of 2014. A

second dataset containing traffic for the second quarter of 2014 will be used to see if the model can make accurate predictions for route choice.

3. Results

Calibrating the model based on all trips made in the 1st quarter of 2014 yields the following values for $[\beta]$:

$$\begin{aligned}\beta_D &= 0.34 \\ \beta_{BREE} &= 0.78 \\ \beta_{ZEE} &= 1.80 \\ \beta_{ROOS} &= 3.11 \\ \beta_{BOT} &= 1.21\end{aligned}\tag{9}$$

This model was used to ‘predict’ route choices for all trips in the 2nd quarter of 2014. The log likelihood of this prediction was -4919.34. The model predicted 95% of all choices correctly. This is slightly lower than the accuracy achieved when using route choices in Q1 to predict those in Q2. This approach resulted in 97% of route choices being predicted correctly.

Below, Figure 3 shows the accuracy of the prediction for each Origin-Destination pair individually. As can be seen in Figure 4 the main difference is using historical data results in less O-D pairs being predicted with low accuracy.

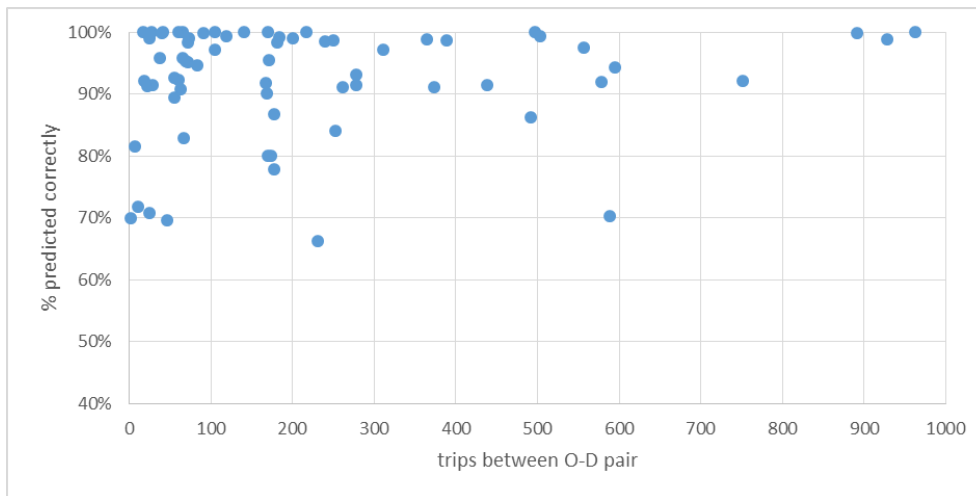


Figure 3 Accuracy of predictions for all O-D pairs with less than 1000 trips between them

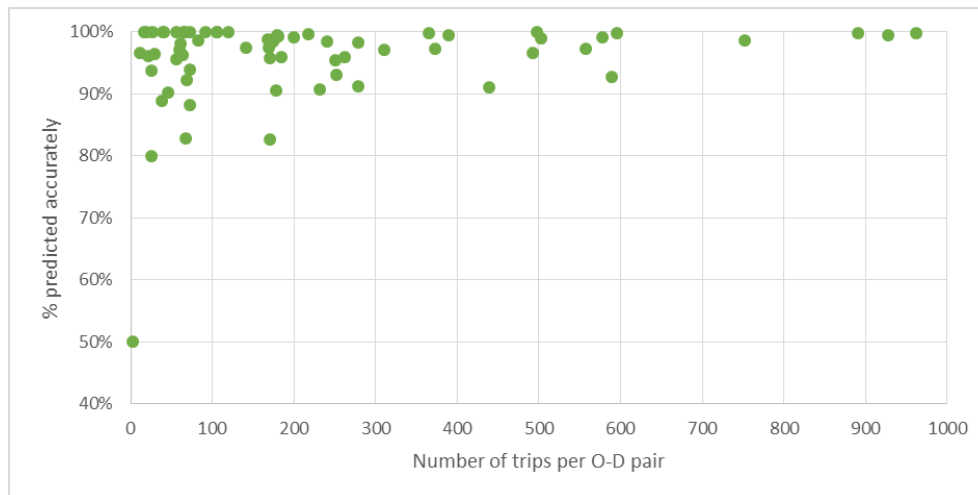


Figure 4 Accuracy of predictions based on historical data for all O-D pairs with less than 1000 trips between them

3.1.1 Influence of vessel size and cargo type

The same logit model was also calibrated using only data from trips with a certain cargo type or vessel size. This was done to investigate the influence of these properties on route choice. These models were used to predict route choices in 2014 Q2 for the appropriate trips in that period. The calibrated values and log likelihood results are in Table 2 below. Note that cargo type and vessel size were not known for all trips in the database.

Table 2 Calibration for different vessel and cargo types

cargo type	$[\beta]$	number of trips in q2	Log likelihood standard model	Log likelihood specific model
Tank	[0.32, 0.61, 1.16, 2.51, 1.33]	5871	-1111.22	-1090.16
Container	[0.35, 0.73, 2.50, 3.66, 1.08]	9173	-1855.99	-1829.67
Bulk	[0.46, 1.78, 3.49, 4.40, 0.92]	3810	-411.78	-384.34
Bunker	[0.27, 0.00, 0.35, 1.74, 1.22]	4797	-1095.07	-996.21
Breakbulk	[0.44, 1.08, 7.25, 2.96, 1.10]	557	-67.32	-74.87

Vessel size		number of trips in q2	Log likelihood standard model	Log likelihood specific model
Small	[0.33, 0.38, 2.60, 2.95, 1.21]	9947	-1518.47	-1465.90
Medium	[0.36, 1.21, 2.08, 3.30, 1.51]	11130	-2006.15	-2003.68
Large	[0.33, 0.57, 0.79, 2.88, 1.02]	4803	-1100.86	-1036.43
Push barge	[0.33, 2.15, 3.89, 3.45, 1.15]	2375	-287.53	-276.44

These results show that, with the exception of breakbulk vessels, specific models result in predictions with a higher log likelihood than using the general model for subsets of traffic. This does not necessarily result in a higher prediction accuracy, since high predictive accuracy and high log likelihood appear to be correlated but are not directly linked to one another.

3.2 Experiment – predicting traffic for closed lock near Rozenburg

During 6 days in July of 2014, the lock near Rozenburg was closed for maintenance. The Logit model was used to predict route choice during those 6 days by assigning a weight of 10,000 to β_{ROOS} , reducing the probability of selecting a route containing the lock to effectively 0%. The other parts of $[\beta]$ were left unchanged. This model predicted route choices made during those 6 days with 95% accuracy, the same level of accuracy as was observed when predicting regular route choices. This suggests the model is also applicable to situations other than the exact one used to calibrate the model and shows it is applicable to practical problems, such as forecasting traffic flows during future maintenance operations in the port area.

4. Conclusions

This research proves that a Multinomial Logit model with Path Size Correction can be used to forecast route choice in the port of Rotterdam accurately for 95% of trips. This was tested on a dataset spanning three months and more than 25,000 trips. The model was also used to forecast traffic for a changed situation, where it also predicted 95% of route choices accurately.

When using historical data from 2014 Q1 to predict route choices in 2014 Q2, 97% accuracy was achieved. This increase is mainly due to the fact that using historical data produces less O-D pairs predicted with relatively low accuracy. The difference with historical data suggests that the model does not succeed in predicting all aspects of route choice correctly, and there is room for improvement. Using a more detailed graph is the most promising option for improving prediction quality.

Using models specifically calibrated for a certain cargo type or vessel size improves the likelihood of the predictions slightly. However, the increase in likelihood is not so large it is recommended in practice. Only for forecasting traffic near certain obstacles (sea and Breiddiep) real practical gains are expected. For other applications using the general model is suggested.

5. Suggestions for further research

Researching route choice for inland shipping vessels sheds light on other interesting subjects for research. There are also ways in which this model could be improved and expanded upon.

Investigate applicability to other ports

It would be very interesting to build a version for another port area. Possibilities include nearby ports such as Amsterdam, Antwerp or Hamburg. Also it would be interesting to see if the model can be used to predict longer trips from the hinterland to the port. This would probably shed some more light on the influence of tides and currents on route choice.

Divide port map into smaller areas

The current way of dividing the map of the port of Rotterdam into areas (see Figure 1) is very rough. This undoubtedly produces inaccuracies when modelling route choice. Using a more detailed model would probably yield more accurate predictions.

Look into the whole visit to the port area, not just trips from 1 origin to 1 destination

This report concerns only trips by vessels in the port of Rotterdam between a single origin and destination. It is likely that looking at the entire visit to the port area and the order and direction of those trips would lead to more insight into route choices of inland shipping traffic. The choice and sequence of destinations is assumed to be a given for this research; in reality, this is not the case. Travel plans can be, and are, adapted mid-journey in reaction to changing circumstances.

Quantify the influence of changing weather conditions on $[\beta]$

In the current model, $[\beta]$ is assumed to be static. This assumption is probably not valid for the penalties associated with wind and waves, since they change continually based on weather conditions. It would be interesting to investigate if and how the penalties change for varying weather conditions.

References

- [1] E. Freijinger, "Route Choice Analysis: Data, Models, Algorithms and Applications," Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2008.
- [2] J. d. D. Ortúzar and L. Willumsen, *Modelling Transport*, 4th Edition ed., Wiley, 2011.
- [3] E. Freijinger and M. Bierlaire, "Capturing Correlation with subnetworks in route choice models," *Transportation Research part B*, no. 41, pp. 363-378, 2007.
- [4] M. Ben-Akiva, M. Bierlaire and E. Freijinger, "Expanded path size attribute for route choice models including sampling correction," in *Proceedings of the International Choice Modelling Conference*, Harrogate, UK, 2009.
- [5] M. Bliemer, "Impact of Route Choice Set on Route Choice Probabilities," the University of Sydney, 1st April 2008. [Online]. Available: http://sydney.edu.au/business/__data/assets/pdf_file/0015/27150/mikeb08.pdf. [Accessed 02 18 2015].

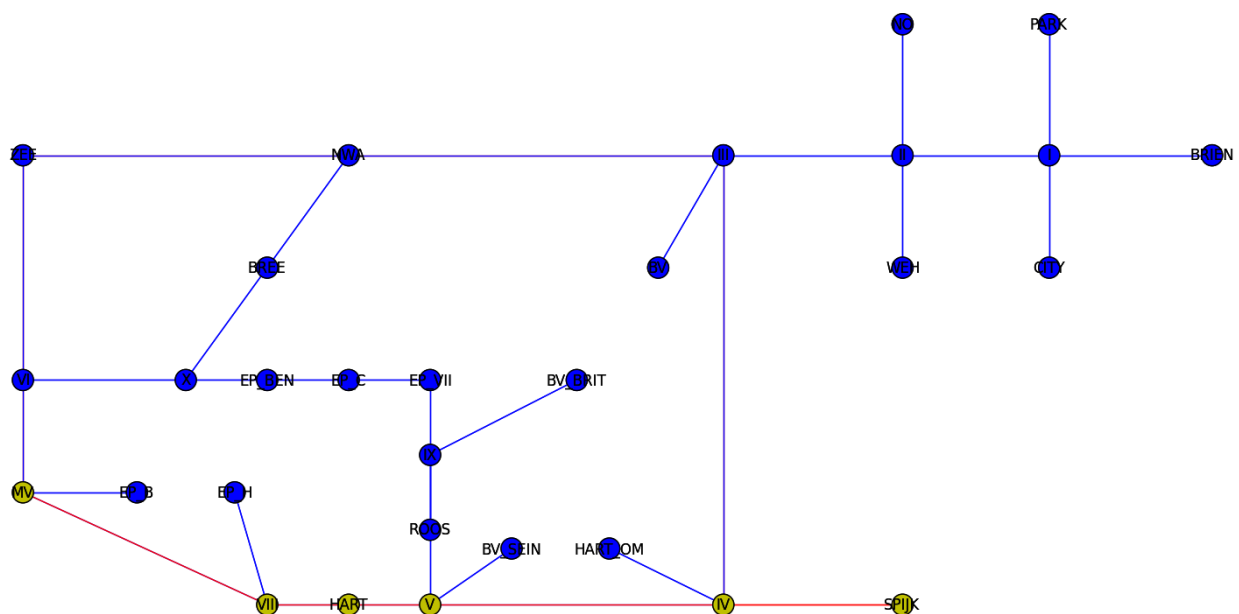
Appendix 2 Example – analysing route choice for a single O-D relationship

The complete process of calculating route choice probabilities for an Origin – Destination pair is described in full in this appendix. Everything in this section could be reproduced using the formulas and algorithms in the regular report but for anyone wanting to use this route choice model, a worked-out example might be useful.

For this example the routes between the Maasvlakte (MV) and Spijkenissebrug (SPIJK) are studied.

Choice set generation

The choice set consists of 4 routes. Other routes are possible but they do not satisfy the criterion that no subroutes can exist which also connect the origin and destination. The feasible routes are illustrated below.



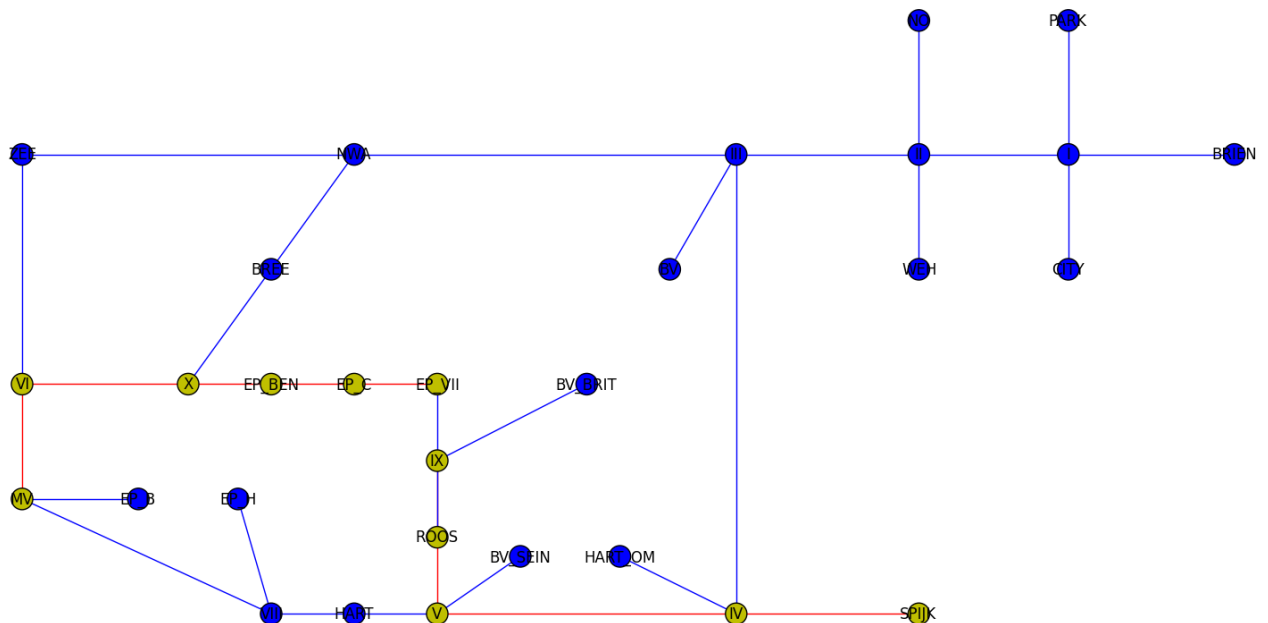
nodes: ['MV', 'VIII', 'HART', 'V', 'IV', 'SPIJK']

Links: [16, 18, 29, 24, 25]

Length: 22.5 km.

PSC: 0.308

Route number: 633



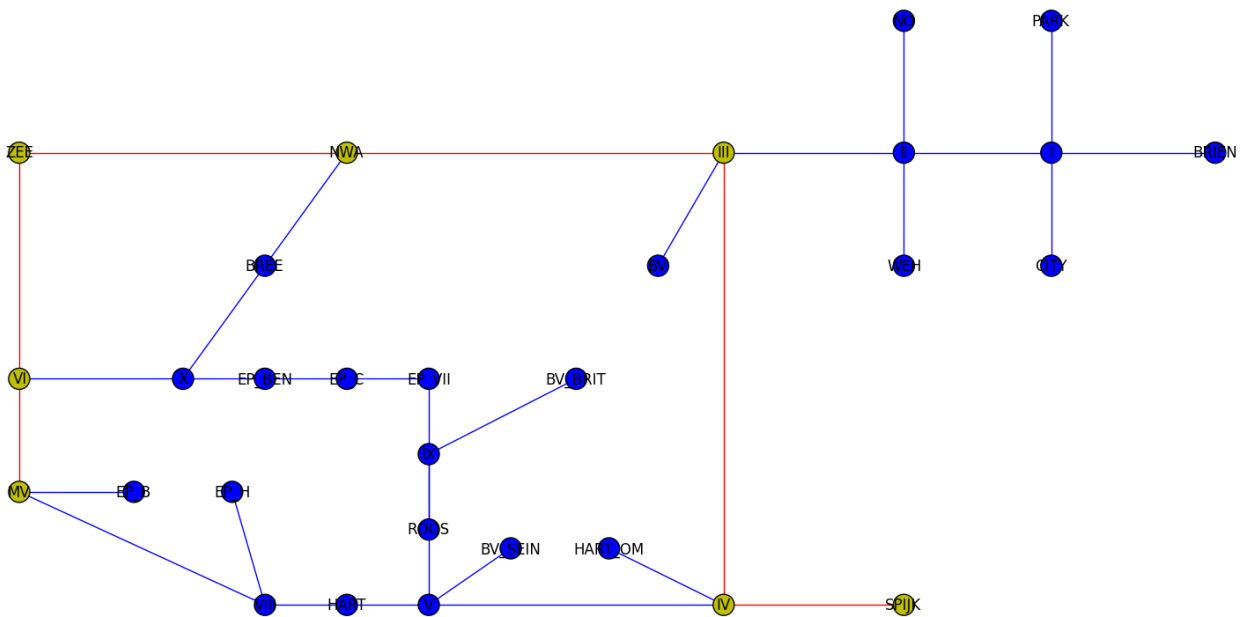
nodes: ['MV', 'VI', 'X', 'EP_BEN', 'EP_C', 'EP_VII', 'IX', 'ROOS', 'V', 'IV', 'SPIJK']

Links: [13, 14, 31, 21, 26, 32, 19, 24, 25]

Length: 21.5 km.

PSC: 0.515

Route number: 634



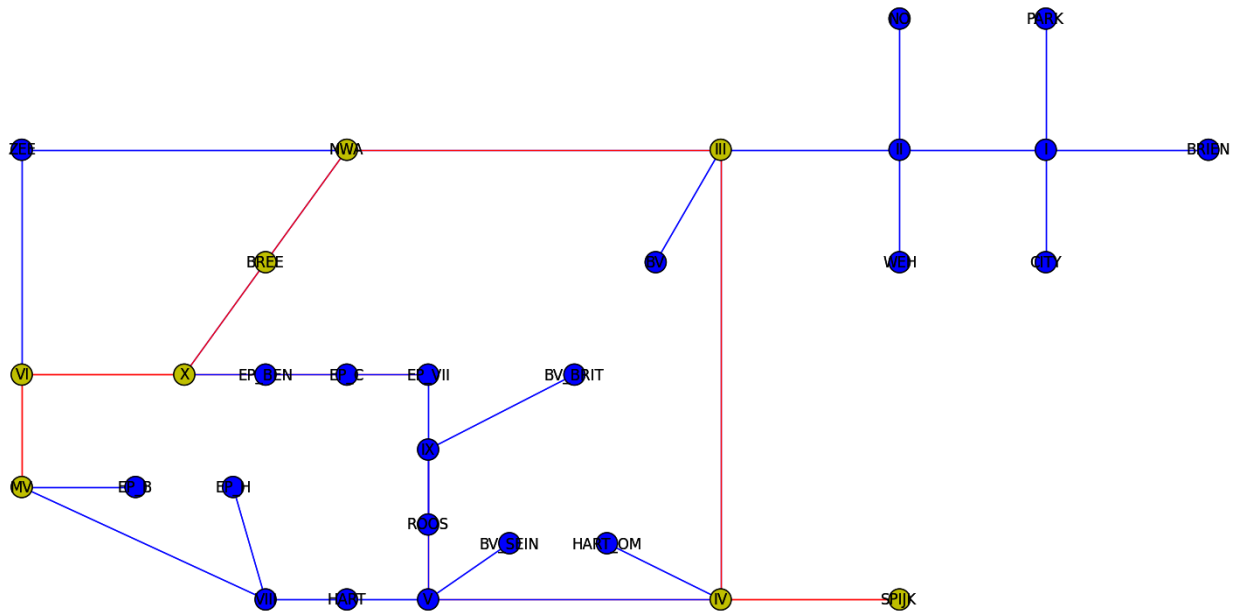
nodes: ['MV', 'VI', 'ZEE', 'NWA', 'III', 'IV', 'SPIJK']

Links: [13, 12, 11, 9, 8, 25]

Length: 28.0 km.

PSC: 0.618

Route number: 635



nodes: ['MV', 'VI', 'X', 'BREE', 'NWA', 'III', 'IV', 'SPIJK']

Links: [13, 14, 23, 10, 9, 8, 25]

Length: 26.5 km.

PSC: 0.71

Route number: 636

Computing probabilities

All routes can be assigned a utility value U_i based on their distance and special penalties assigned to certain links (like for example the Breddiep). The different properties and utilities for each route are shown in Table 3 below.

Table 3 Properties of routes between SPIJK and MV

Route number	Significant links (i.e. visible in traffic data)	Distance	PSC	Obstacles	Utility
633	['HART']	22.50	0.31	0	-7.96
634	['ROOS', 'EP_VII', 'EP_C']	21.50	0.51	3.11	-10.93
635	['NWA', 'ZEE']	28.00	0.62	3.01	-13.15
636	['NWA']	26.50	0.71	1.99	-11.71

Now using formula 11, reprinted below for convenience:

$$P(i|C) = \frac{\exp(U_i)}{\sum_{j \in C} \exp(U_j)}$$

The probabilities of choosing each route can be computed. They are printed below in Table 4.

Table 4 Probabilities of choosing each route in choice set C

Route number	Significant links (i.e. visible in traffic data)	P(i C)
633	['HART']	93%
634	['ROOS', 'EP_VII', 'EP_C']	5%
635	['NWA', 'ZEE']	1%
636	['NWA']	2%

Comparing the model to the real world

From traffic data collected by the Port of Rotterdam, it is possible to look at all trips between the MV and SPIJK areas and look at what areas the vessels visited in between. This list of areas then needs to be matched to one of the four possible routes. The table below () shows the recorded areas between SPIJK and MV, the amount of times each record was found, and the route number assigned to each record. Route number 0 means this route is unknown.

Table 5 Routes recorded for trips between MV and SPIJK

Chosen route	Times counted	Route number
['HART']	5116	633
['ROOS', 'EP_C']	58	634
[]	42	0
['NWA']	38	636
['ZEE', 'NWA']	33	635
['EP_C', 'HART']	24	633
['EP_C', 'NWA']	24	636
['EP_C']	4	0
['EP_C', 'ZEE', 'NWA']	4	635
['ZEE', 'NWA', 'HART']	2	0
['EP_C', 'NWA', 'ZEE', 'HART']	1	0
['ZEE', 'HART']	1	0
['NWA', 'HART']	1	0

Total	Total - errors	error percentage
5348	5297	0.95%

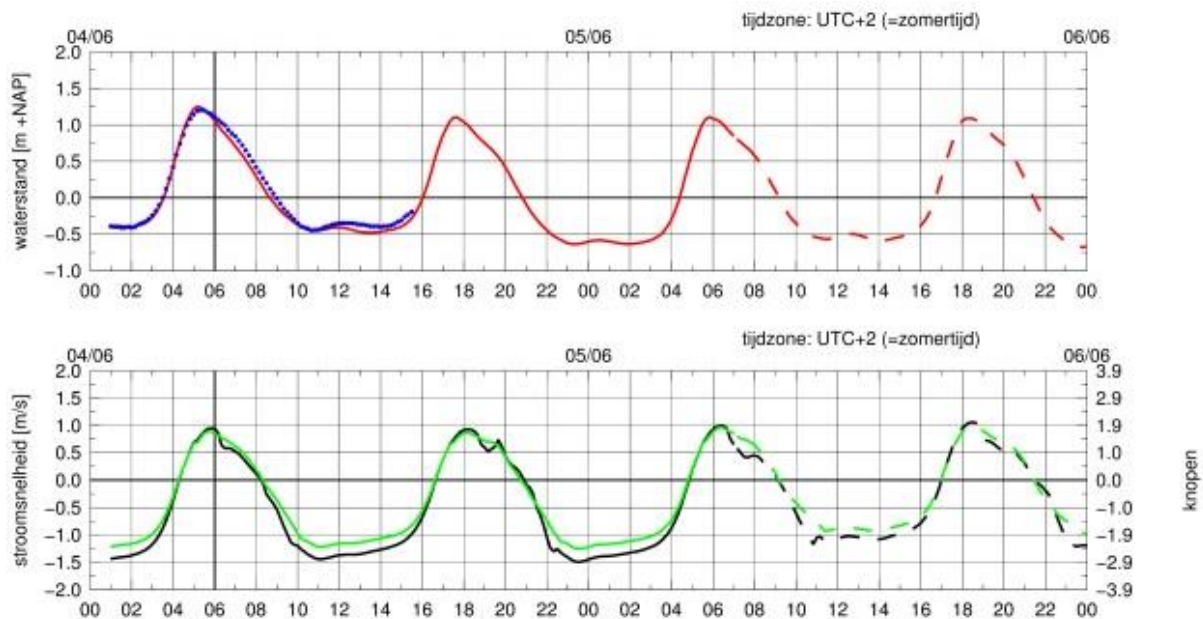
		P(i C)	Observed	% right
633	['HART']	93%	97%	93%
634	['ROOS', 'EP_VII', 'EP_C']	5%	1%	1%
635	['NWA', 'ZEE']	1%	1%	1%
636	['NWA']	2%	1%	1%

As can be seen from Table 5, 96% of trips are predicted correctly using this model.

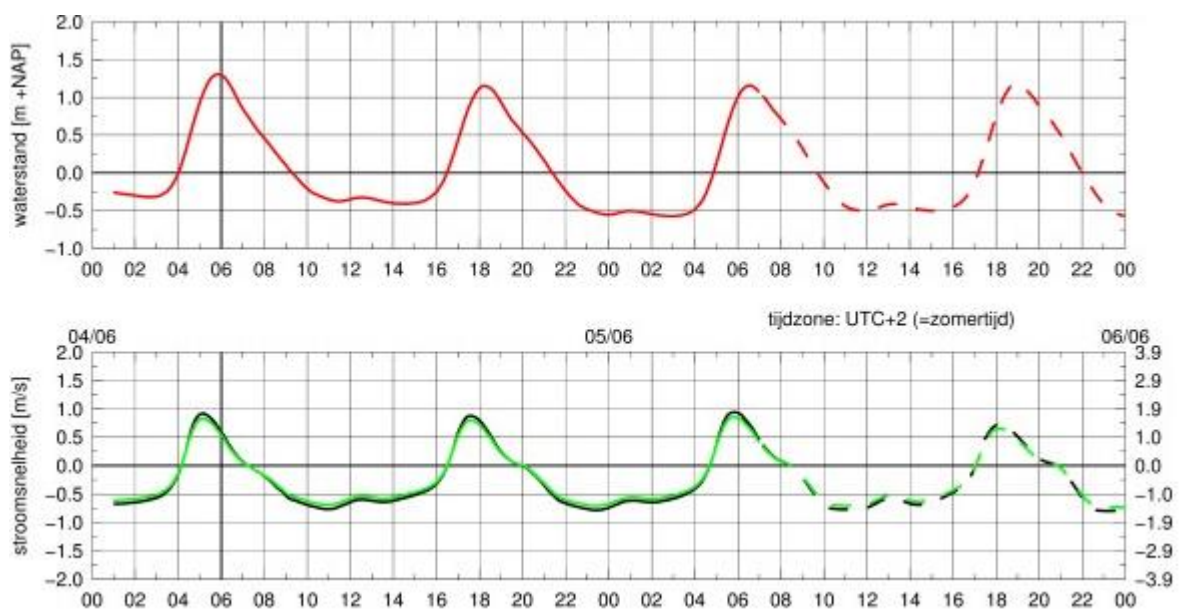
Appendix 3 - Tides and currents in the port of Rotterdam area

This section contains information on the tides in the port of Rotterdam waterways. All graphs are taken from the website 'Operationeel Stormingsmodel Rotterdam' from the Port of Rotterdam. These are shown here as examples of how water levels and currents are roughly the same in the different waterways.

Comparison tides and currents in different waterways



Oude Maas at the Spijkenissebrug (SPIJK)



Nieuwe Maas at Brienenoordbrug (BRIEN)

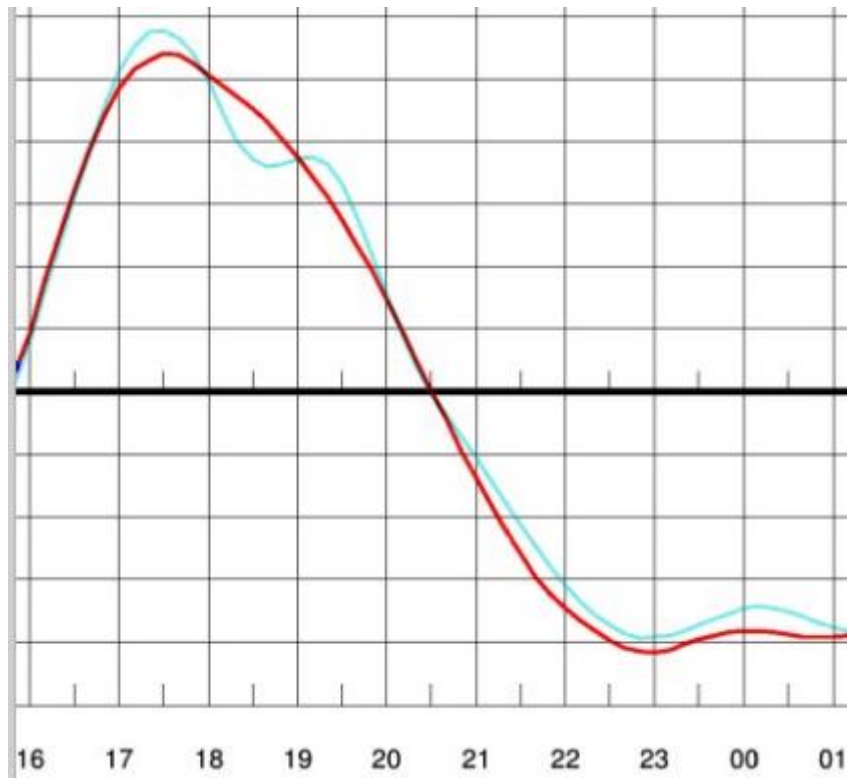


Figure 5 Water level near Harmsenbrug (blue) and Maassluis (Red)

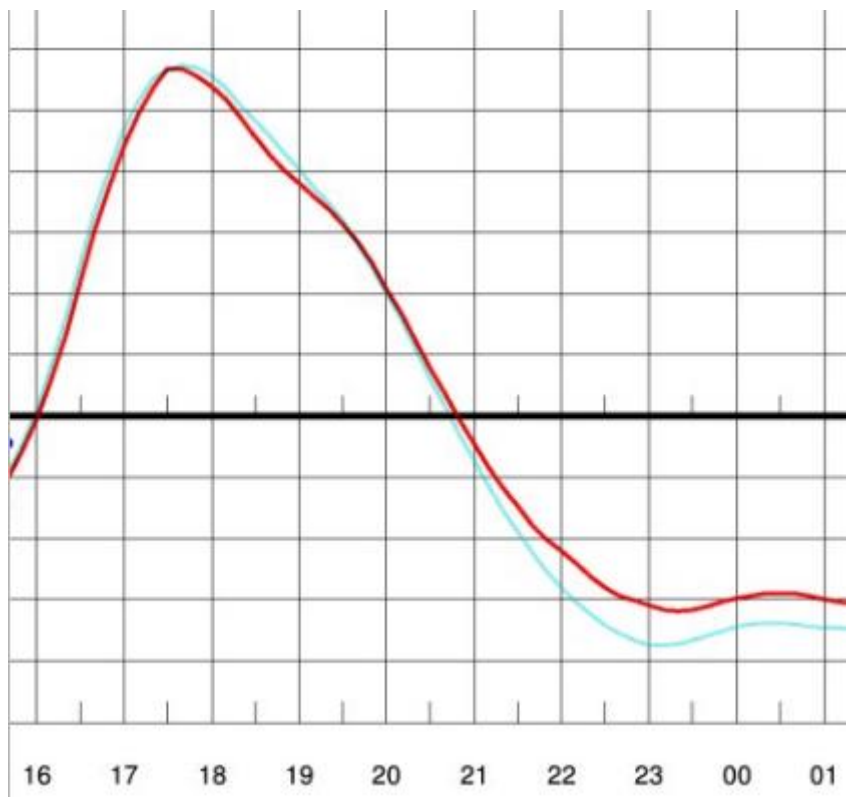


Figure 6 water level near Botlekbrug (blue) and Hartelbrug (red)

Figure 5 and Figure 6 show clearly that water levels are almost the same in different waterways.

Interviews with inland shippers at the inland shipping fair

During the inland shipping fair 'Maritime Industry 2015' a number of inland shipping captains were asked for their opinion on route choice and specifically the influence of currents and tides on route choice in the port of Rotterdam. They stated that there is no difference in tides between waterways. One shipper mentioned it, but all others say the difference is non-existent or negligible.

The OSR graphs show that no clear advantage with respect to tides can be gained from changing routes. This is confirmed by the interviewed shippers. Tides are important in the sense that shippers try to arrive in time to be able to sail with the tide. This saves fuel and time and therefore saves money. However by the time the port area is reached this decision has already been made.

Influence of currents

Comparing the water flow in two pairs of locations.

	Hoek van Holland (Nieuwe Waterweg)	Suurhofbrug (Hartelkanaal)	Scheurkade (Nieuwe Waterweg)	Hartelbrug (Hartelkanaal)
Vloedstroom max(gemiddeld)	0.96	1.35	1.10	1.34
Ebstroom max (gemiddeld)	-0.57	-0.77	-0.64	-1.34

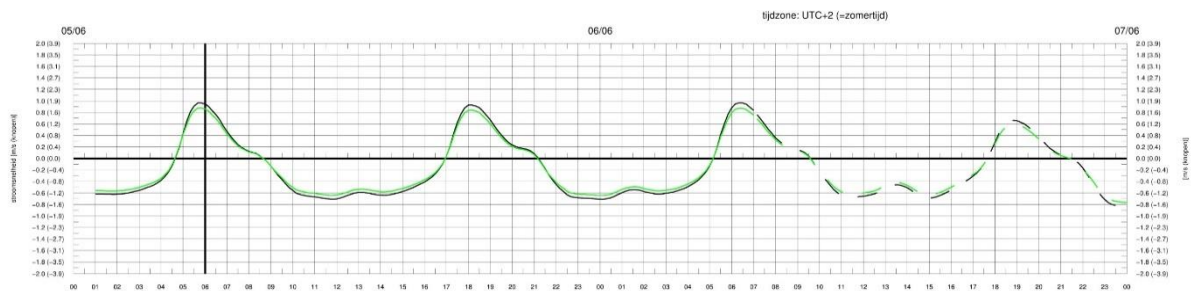


Figure 7 Locations of measurement points. Source: HydroMeteo report

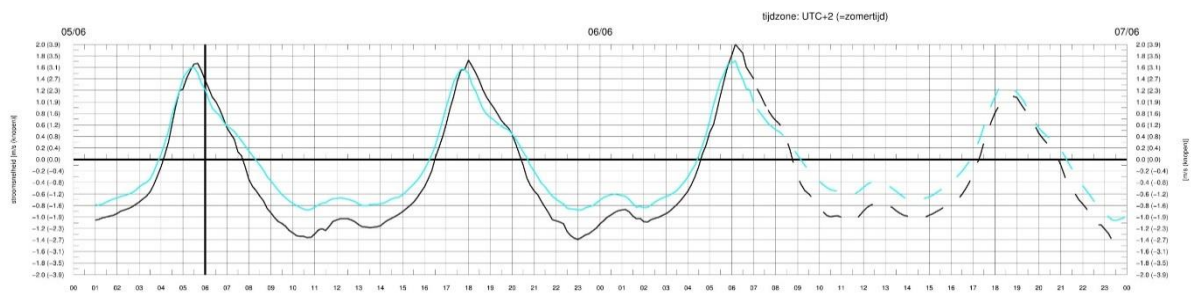
Positive numbers mean water flows upstream, negative numbers mean water flows downstream (towards the sea).

Please note that the water flow speeds are the average flow speeds over the depth of the waterway – different depths levels can and do exhibit different flow speeds and directions.

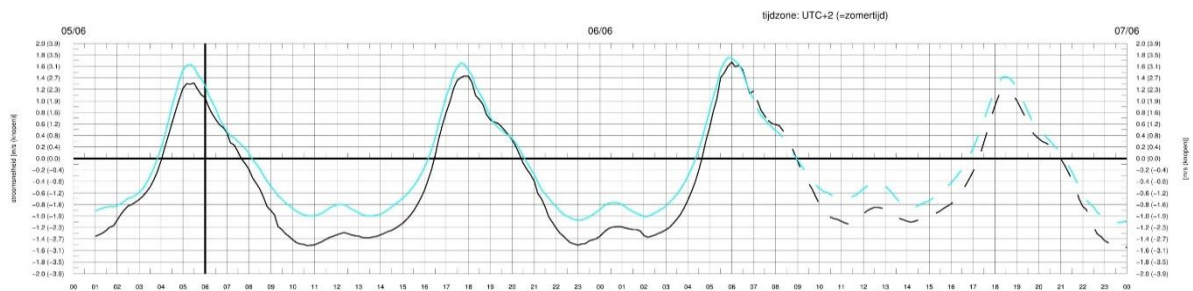
Currents from OSR – Nieuwe Waterweg



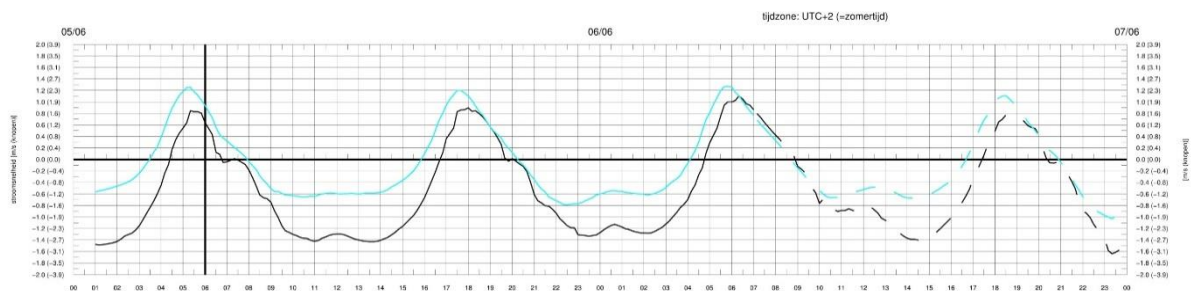
Brienoordbrug



Scheurkade



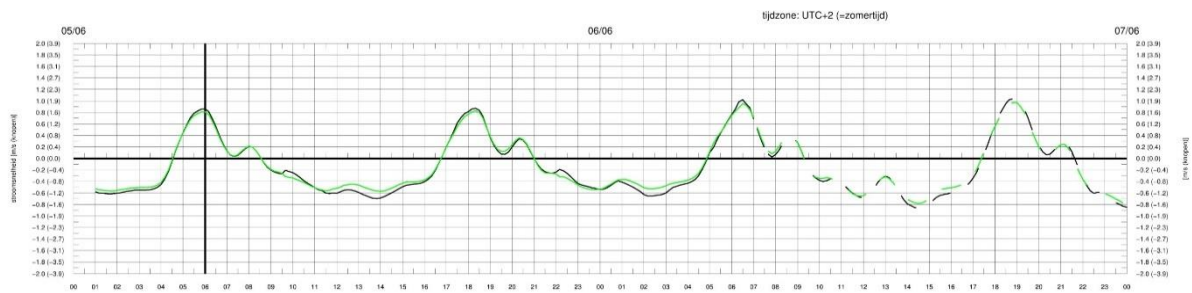
Maassluis



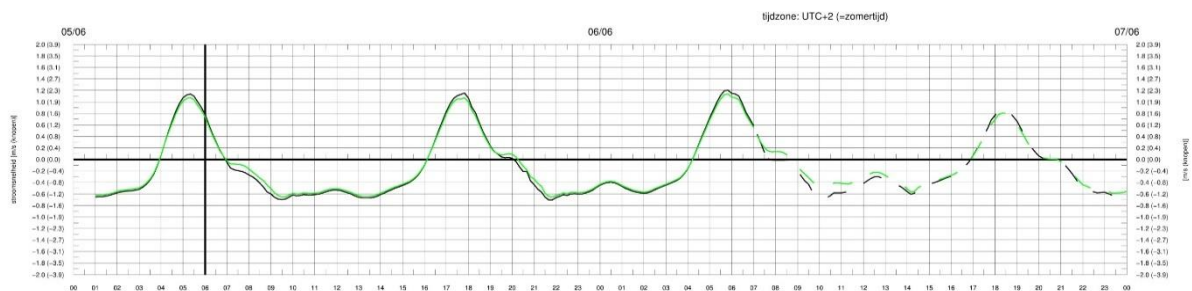
Hoek van Holland

Black line is currents 0-5m, blue line currents 0-15 m, green line is depth averaged.

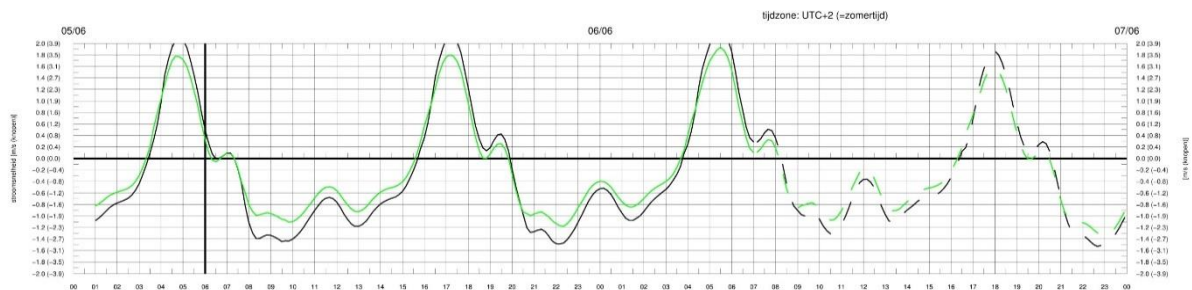
Currents from OSR – Hartelkanaal



Hartelbrug



Harmsenbrug



Suurhoffbrug

Analysis

Currents do not vary significantly across the different channels. Current is somewhat faster in one place in the the Hartelkanaal but is not the case for large parts of the channel. There is no time difference between the current peaks in the Nieuwe Waterweg and Hartelkanaal.

Hoek van Holland has really different currents for different depths, probably because of the different salt levels of the sea and inland waterways.

Flood currents are roughly equal on both waterways. On some parts of the Nieuwe Waterweg, the ebb currents (towards the sea) are notably quicker than those on the Hartelkanaal. On the Nieuwe Waterweg these currents can reach a speed of 1,5 m/s (5,4 km /h) whereas on the Hartelkanaal this current is only 0.7 m/s (2.5 km/h). This can mean a speed difference of 2.5 km/h. Maximum speeds

on these channels are 13 km/h faster than the water, so this can mean a speed increase of about 15% when going downstream with the tide. Normally, one would sail the 17 km long track of the Nieuwe Waterweg in roughly 1 hour and 10 minutes (assuming a speed of 15 km/h). Adding 2.5 km/h to this speed would mean the journey takes about 60 min. This would mean saving roughly 10 minutes of travel time, which is probably not a reason to change routes. Sailing with the currents would also mean saving on fuel, which may be a reason to change routes.

If currents had an influence on the route choice in the port of Rotterdam, we would expect the Nieuwe Waterweg to be relatively more attractive (have a higher utility) when going towards the sea. This is not observed in traffic data. This further suggests that currents are not an important influence on route choice within the port of Rotterdam.

Appendix 4 – Questionnaire

A questionnaire was sent to inland shipping captains on Facebook. The questions and results are collected in this appendix. The original questionnaire was in Dutch. Translations of the questions and answers are used for this report.

Questions in Dutch

- Wat voor lading vervoert u meestal?
- Hoe lang is uw schip?
- Waar komt u meestal het havengebied van Rotterdam binnen?
- Hoe kiest u welke route u vaart?
- Bepaalt u de route van tevoren of tijdens de reis?
- Kiest u, bij gelijke herkomst en bestemming, altijd dezelfde route?
- Wat zijn voor u de belangrijkste zaken bij het kiezen van een route?
- Met welke van de onderstaande factoren houdt u rekening bij uw routekeuze?
- Zouden deze mogelijke toekomstige ontwikkelingen invloed hebben op uw routekeuze?
- Heeft u nog opmerkingen?
- Zou ik voor aanvullende vragen contact met u mogen opnemen?

Questions in English

- What kind of cargo do you usually carry?
- What size is your ship?
- Where do you usually enter the port of Rotterdam?
- How do you determine what route you take?
- Do you choose a route beforehand or during your journey?
- Do you, if origin and destination are the same, always use the same route?
- What do you think is most important when choosing a route?
- Which of the following factors do you take into account when choosing a route?
- Would these future developments influence your route choices?
- Do you have any comments?
- Could we contact you for further questions?

Answers

What kind of cargo do you usually carry?

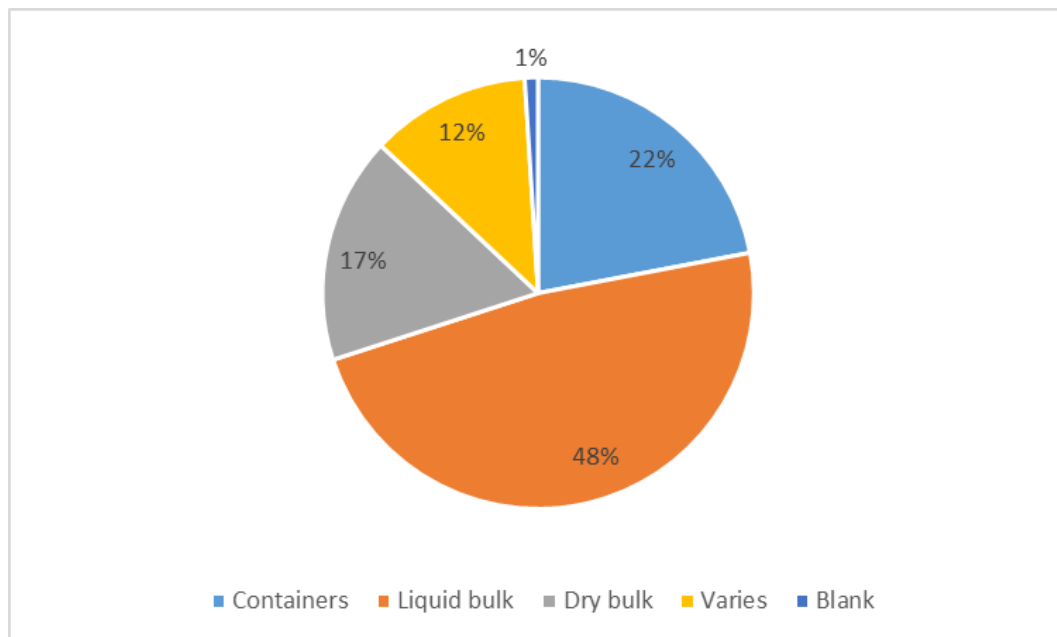


Figure 8 Responses to question 1

What size is your ship?

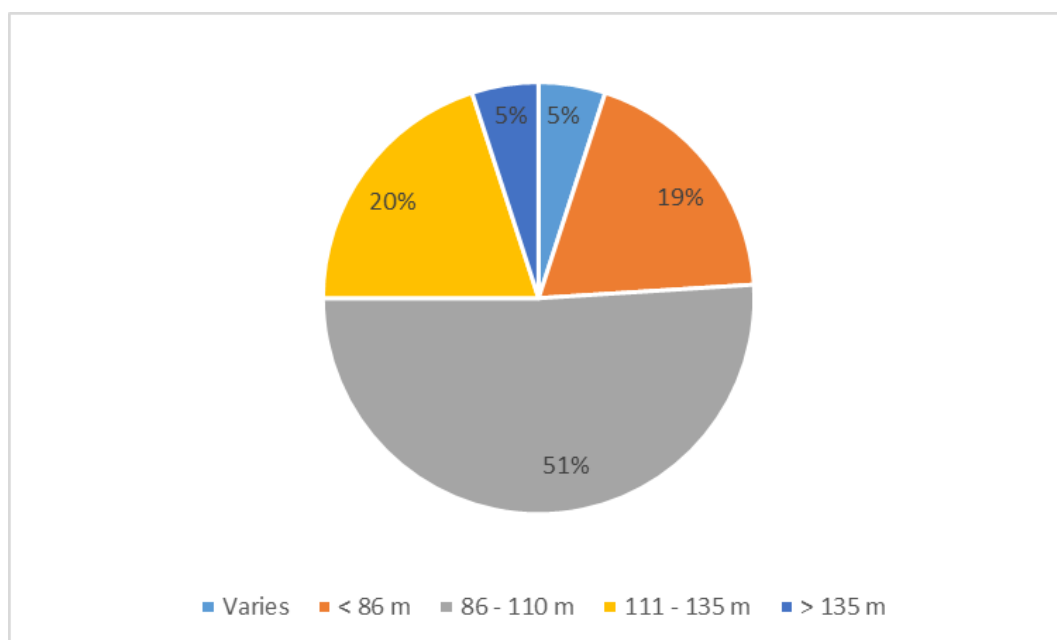


Figure 9 Responses to question 2

Where do you usually enter the port of Rotterdam?

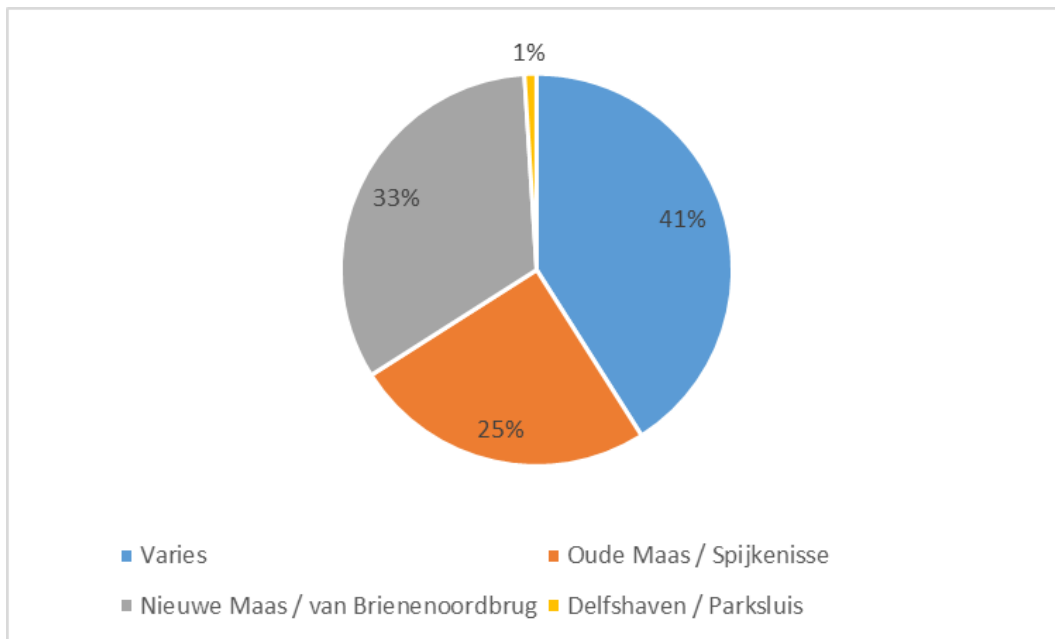


Figure 10 Responses to question 3

How do you determine what route you take?

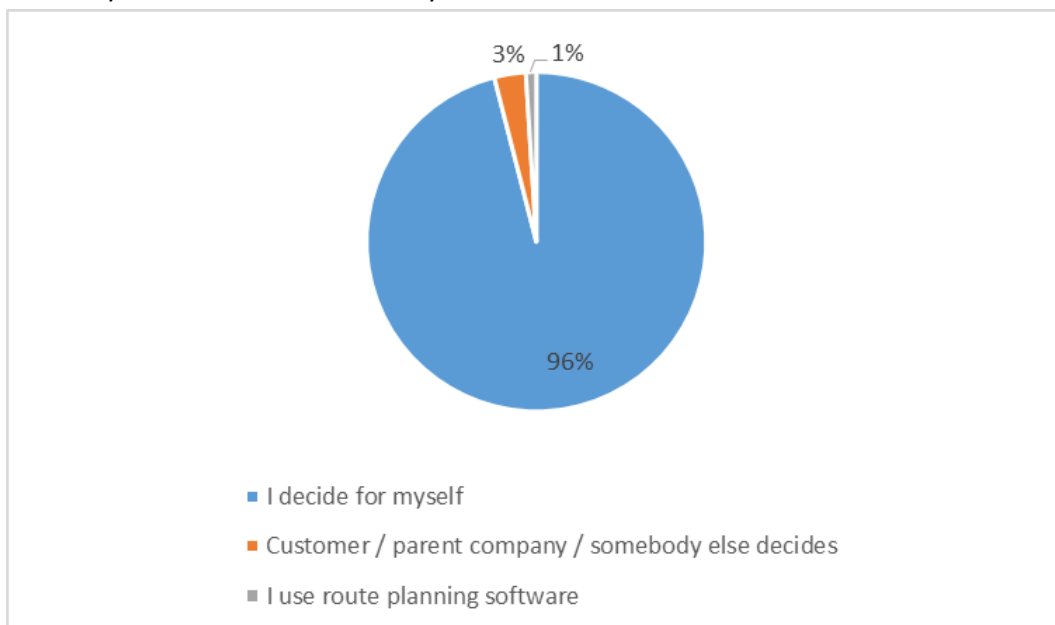


Figure 11 Response to question 4

Do you choose a route beforehand or during your journey?

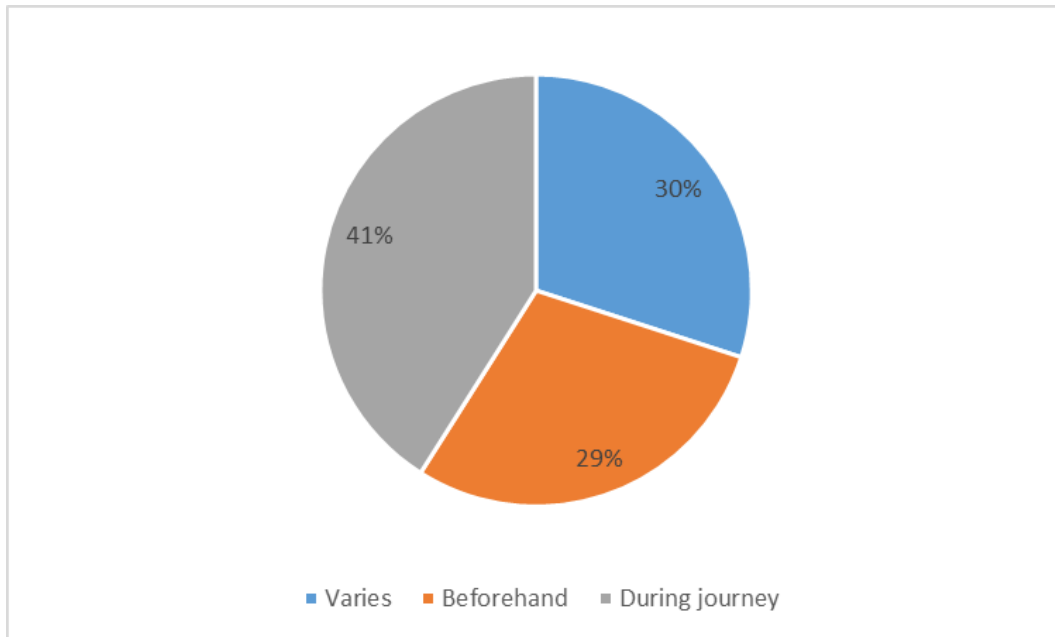


Figure 12 Response to question 5

Do you, if origin and destination are the same, always use the same route?

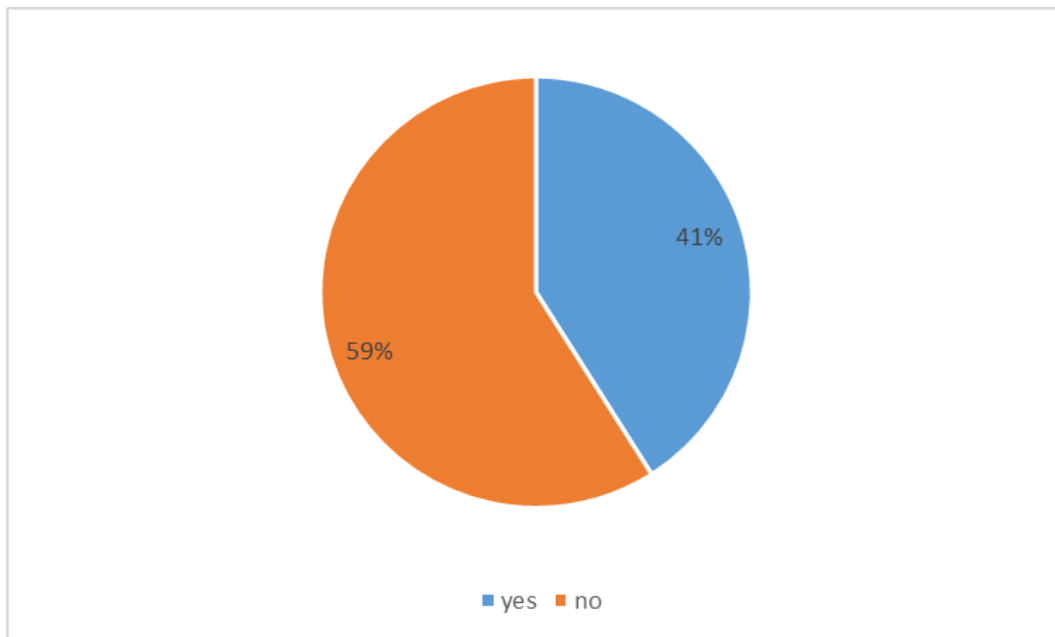


Figure 13 Response to question 6

What do you think is most important when choosing a route? (open question)

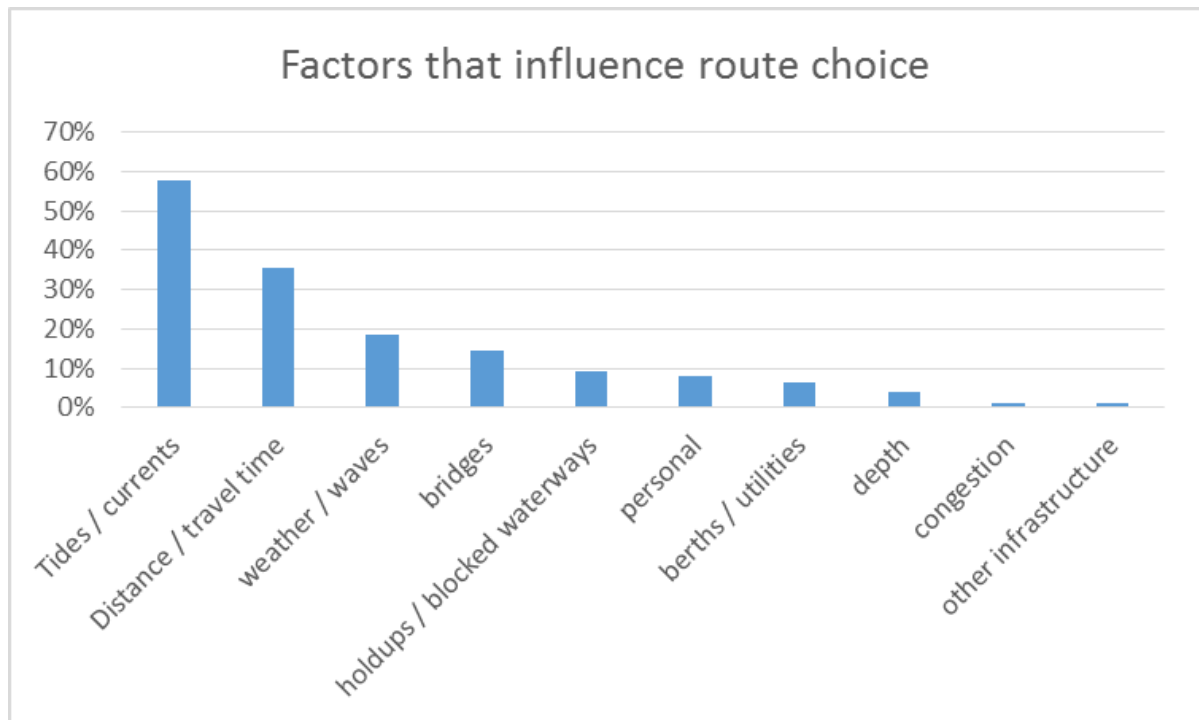


Figure 14 Response to question 7

Do these factors influence your route choice? (list of choices)

Tides	62	82%
Currents	59	78%
Distance	44	58%
Travel time	43	57%
Wind	37	49%
Bridges	36	47%
Locks	26	34%
Berths	26	34%
Waves	21	28%
Speed limits	17	22%
Other infra	11	14%
Depth / width waterways	10	13%
Friends / colleagues	6	8%
Congestion	4	5%
none of the above	0	0%

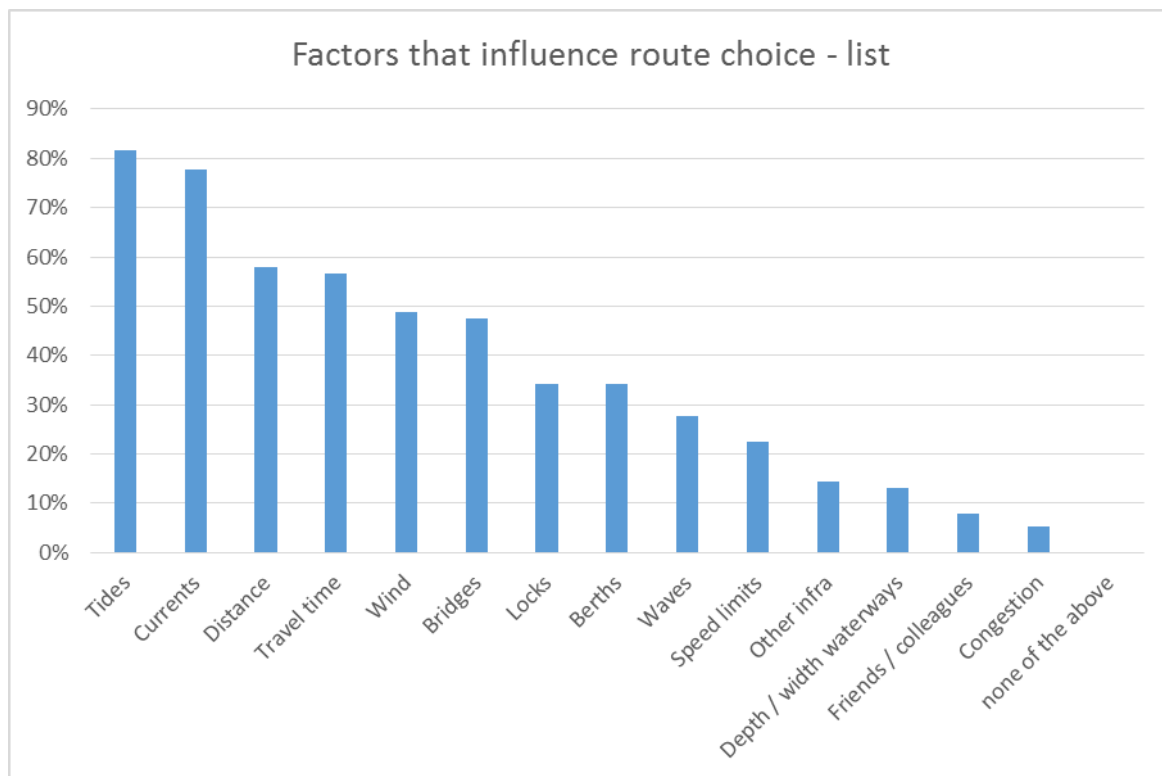


Figure 15 Response to question 8

Would these future developments influence your route choices?

Answer	#	%
Removing Rozenburgsesluis	32	42%
Remove speed limits	32	42%
Widening Breddiep	28	37%
None of the above	21	28%
large increase in traffic volume	8	11%

The results from this questionnaire were used to create a list of potential influences on route choice. These potential influences were further investigated in order to determine which ones to include in the route choice model. Also the list was used to search for an experiment to test the route choice model, which was found in the closing of the lock near Rozenburg.