Center for
Big Data Statistics
Working paper

Combining data sources to gain new insights in mobility

A case study

Working paper no.: 03-20

Y.A.P.M. Gootzen
M.R. Roos
B.O. Mussmann

July 2020

Abstract

This paper describes how multiple data sources that relate to mobility in the city of Rotterdam were combined to create a model for scenario analysis. Sources originate from RET, Gemeente Rotterdam and Statistics Netherlands.

A collection of sources measuring the real-life situation and a collection of administrative sources are converted to the granularity of metro segment. Based on the ratio between these groups, administrative sources are calibrated towards an expectation of real-life implications.

The methods used in this paper are applicable for analysis of any type of scenario that can be described by means of administrative data. This is demonstrated by two population growth scenarios, in which the number of residents is artificially increased for specific neighbourhoods. Each scenario is then applied to the understandings learned from the various data sources to illustrate the effects of population growth on the metro network. This paper illustrates the added value of combining sources that individually would not have allowed to reach the same insights.

Contents

1 Introduction

The City of Rotterdam has a multimodal transportation network that gets more crowded every year, while the population of the city is expected to grow in the coming years (Gemeente Rotterdam, 2017). Insights into the effects of population growth on the city's mobility are valuable for policy makers. Though sources on traffic counts and scenarios of population growth are available, their relation is not straightforward and requires extensive modelling (Rijkswaterstaat, 2017). This paper describes the joint efforts of a public transportation carrier, a local government and a national statistical office to integrate administrative data, traffic counts and future scenarios on a small scale.

In 2019, Rotterdam Elektrische Tram (RET), City of Rotterdam and Statistics Netherlands collaborated within the context of the Leadership Challenges in Big Data and Analytics program at Erasmus University. Combining a multitude of sources led to a model that can estimate the influence of a fictive (future) scenario on the metro network in Rotterdam. The methods used in this paper are applicable for analysis of any type of scenario that can be described by means of administrative data. This is demonstrated by two population growth scenarios, in which the number of residents is artificially increased for specific neighbourhoods. The scenario is then applied to the understandings learned from the various data sources to illustrate the effects of such population growth on the metro network.

The concept of mobility can be divided into categories based on different aspects such as transportation mode (e.g. car, train, bus, metro, bicycle, by foot) or motive (e.g. home-work commute, home-education, free time, tourism). The work presented in this paper focused on the transportation mode metro and the travel motive home-work commute. We train a model that can convert expected commuters from administrative data sources to observed travellers in the metro network. The model was trained on data from June 2018.

This paper combines sources with no overlapping variables by a stepwise conversion so variables can eventually be compared on the same level of granularity. By granularity, we mean the identifying concepts and their level of detail. For example, both a segment and a route are valid levels of granularity for the concept metro infrastructure. To apply a scenario to the final model, the key is to express it on a level of granularity (e.g. geographical regions) that not only allows for an accurate description of the scenario (e.g. population growth) but also has a relation to the granularity of at least one of the other sources. Let us define a scenario definition as the the formalisation of a scenario on a level of granularity, whereas a scenario is often initially described by merely words. The multitude of sources (e.g. administrative, survey, observational counts) allows for a wider range of potential meaningful levels of granularity for the scenario definition and thus a wider range of acceptable scenario definitions compared to when the scenario definition should match the granularity of a single source (e.g. observational counts on an infrastructural level). The population growth scenarios considered in this paper are intuitively defined at neighbourhood level, which resembles the granularity level of one of our administrative data sources. So if we relate administrative data sources to our metro network source and we also relate increasing amounts of residents to administrative data sources then we can translate increasing amounts of residents into observed metro network counts more reliably. The goal of this paper is to illustrate the added value of combining sources that individually would hot have allowed to obtain outcomes at more detailed levels and richer in information content.

2 Research Background

Statistics Netherlands has been studying new data sources and models to provide more insight into travel motives and modalities (CBS, 2018b). Register based data, survey data and new data sources are combined in a stepwise approach. Examples of these new data sources are traffic loop data and anonymised data from mobile phone providers. Register based data sources include the register on wages (in Dutch: 'Polis'), the education register, the register of addresses of utility services (malls, day-care centres) and the population register (in Dutch: 'bevolkingsregister').

The focus of this research is whether information in register data can be used as a proxy for different travel motives: the register on wages for commuting to work, the education register for travel to the place of study and the register of addresses of utility services for a mixture of other motives.

Combining addresses from the population register with destinations from the other registers results in origin-destination (OD) matrices, in which each value represents an estimate of the number of potential commuters between the origin neighbourhood and destination neighbourhood. A subset of commuters was included in the sample of the transportation behaviour survey (CBS, 2018a). For these people, we know whether they travelled and if so, we know their motive and transportation mode. This sample provides valuable insights in transportation mode choices.

A Bayesian model was developed that estimates the probability distribution of the various transportation modes based on background information of a person, such as age, level of education, social economic class, neighbourhood urbanity and possession of a driver's license. This model was trained on seven years of data from the transportation behaviour survey. The experimental model allows us to generalise the individual observations from the sample to estimate the probabilities of transportation modes for those who were not included in the sample. Summarising the individual probabilities per origin neighbourhood and destination neighbourhood results in estimated Origin-Destination (OD) matrices for both travel motive and transportation mode.

With travel motives and transportation modes derived from these register and survey sources, these data are combined with actual traffic counts from new data sources. The combination of sources should provide a more detailed insight into reasons why people are travelling from A to B, the choices they make for departure times, and the transportation mode. The use case studied in this paper is the next step in testing the validity of the underlying assumptions and models.

3 Data

The data described in this paper was combined only for the educational purpose of the collaboration. This section describes each of the used sources in detail.

3.1 RET data

RET shared two types of data: observed counts of travellers and network structure.

3.1.1 Anonymised observation data

Based on unique check-in and -out events of travellers walking through gates at metro-stations, RET is able to report counts of the total number of travellers between any two stations. These counts are provided per hour for different types of days (workday, Saturday and Sunday). No individual observation data was shared, since these were aggregated into anonymised counts prior to sharing and analysing. Let us define a trip as the consecutive segments travelled when going from the check-in station to the next check-out station such that no check-in or -out occur in between. The journey of a traveller may consist of multiple trips if they are required to pass through gates in the middle of their journey. The variables included in the RET count data and their notation, as used in Section 4, are presented in Table 3.1.

3.1 Relevant variables in the RET observation data.

Variable	Notation	Description
Departure station	v	Check-in station, where the trip starts.
Arrival station	v'	Check-out station, where the trip ends.
Hour	h	Value of the hour of day, starting at 3 and ending at 26 because the first two hours of the night are considered part of the previous day for administrative purposes.
Type of day	d	Workday, Saturday or Sunday.
Count	$n(v, v', d, h)$	Number of people whose trip started during the specified hour of the specified type of day, originating from the departure station and ending at the arrival station.

3.1.2 Network structure

The network structure is viewed as a graph, consisting of nodes and edges that represent stations and segments, respectively. Let a segment be defined as a piece of rail between two stations with no station in between. Two data sets are used to define the graph, one containing information on stations and another containing information on segments between the stations.

3.2 Relevant variables in the RET station data, used to define nodes in the graph database.

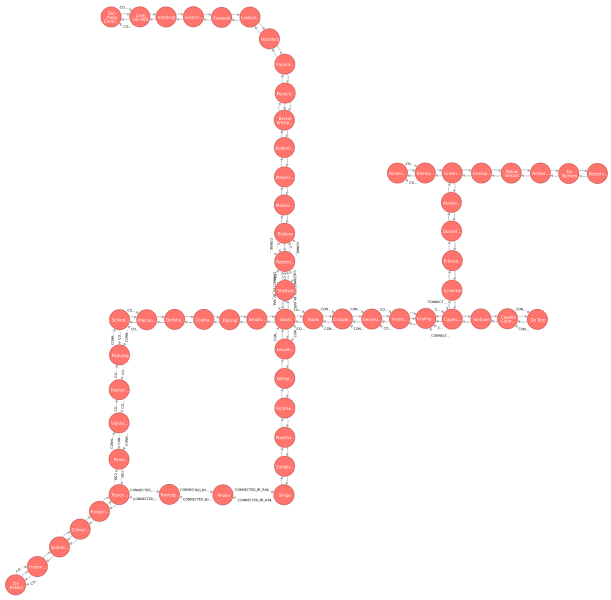
Variable	Notation	Description
Station ID	v	A unique code used to distinguish one station from another.
Station name		The name used when announcing a stop at the station.
Latitude		Latitude of station.
Longitude		Longitude of station.

3.3 Relevant variables in the RET network data, used to define edges in the graph database.

Variable	Notation	Description
Station ID from	v	A unique code used to distinguish the departure station.
Station ID to	v'	A unique code used to distinguish the arrival station.
Distance		The physical length of the segment between the departure and arrival stations measured by rail.

Table 3.2 and Table 3.3 are used to define the graph that is build in a neo4j environment. Neo4j is a graph database management system (Neo4j, 2019). A schematic visualisation of the entire graph for the metro system is depicted in Figure 3.1.

3.1 Schematic overview of the RET metro network.



3.2 Municipality of Rotterdam data

Data from the municipality of Rotterdam consists of an expected relative growth rate for different regions within the municipality. We call the description of the expected relative population growth a *scenario*: it holds information on the phenomenon mobility, but its exact relation to traveller counts is unclear. Table 3.4 shows the contents of a data set describing a scenario.

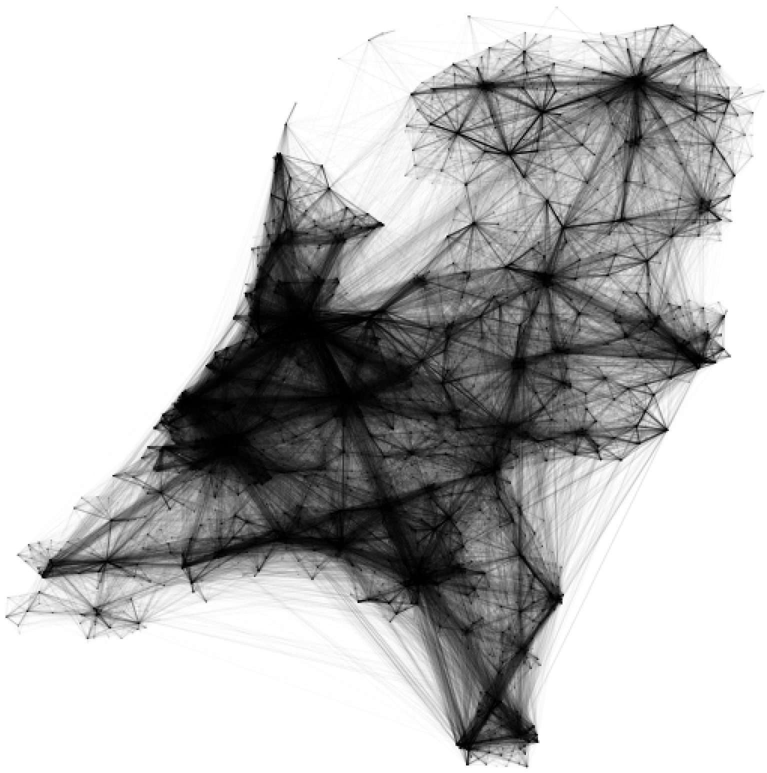
3.4 Relevant variables in the scenario from municipality of Rotterdam.

Variable	Notation	Description
Region	R	Geographical subdivision of the municipality, similar to neighbourhood.
Prognosis 2020		Relative population growth rate per region in 2020.
Prognosis 2025		Relative population growth rate per region in 2025.
Prognosis 2030		Relative population growth rate per region in 2030.

3.3 Statistics Netherlands data

Statistics Netherlands has access to data that is often not collected with mobility as its primary focus. The municipal personal records database contains information on neighbourhoods in which a person lives. The Polis contains information on employees and the companies they work for, with information about their respective locations being available in other registers. Assumptions based on this information are made to assign each commuter with a working location.

3.2 A schematic overview of potential commuter trips in the Netherlands based on the Polis. The darker a line, the more expected travellers. Lines represent the relation between two locations, not the actual travel trajectory of a citizen.



With some uncertainty, these data can be combined into an OD matrix for the entire country containing the number of employees for combinations of living and working neighbourhoods. Figure 3.2 shows a schematic overview of the commuter OD matrix of the Netherlands. This matrix is based on several assumptions such as: the working location is stable, commuting is the exclusive purpose of the trip and no commuters are unregistered. By a stable working location, we mean that a person does not alternate different establishments, for example on different days of the week. These assumptions are most likely untrue and in order to learn to which extend they are responsible for errors and how to correct accordingly, the OD matrix must be compared to real-life observations.

For all commuters in the OD matrix, it is unsure whether they will travel and in case they do, when they will and which mode of transportation they use. With the help of data on mobility behaviour from the transportation behaviour survey, a model was developed that estimates mode of transportation based on person characteristics (such as age, urbanisation of neighbourhood, ownership of a driver's licence). When applied to all potential Dutch commuters, this results in an OD-matrix for each mode of transportation. This model incorporates some errors due to the previously mentioned assumptions.

3.5 Variables in the OD-matrix for public transportation commuters.

Variable	Notation	Description
Neighbourhood departure	R	Unique code for the home location neighbourhood.
Neighbourhood arrival	R'	Unique code for the destination (working location) neighbourhood.
Intensity	$k(R, R')$	Expected number of commuters between the departure and arrival neighbourhoods for public transportation.

Table 3.5 shows the contents of the OD-matrix of commuting via public transportation. Travel motives other than commuting to work and transportation modes other than public transportation are not included in this OD-matrix.

3.4 Additional data

The OD-matrix based on administrative data sources and a survey is defined at neighbourhood level. To translate the administrative data onto the network, we needed to learn the relation between neighbourhood-based trips and the network infrastructure. We call this process *projecting* the OD-matrix onto the network. We used the Google Maps Directions API for the additional data needed for this purpose (Google, 2019).

The data was obtained by requesting directions between the centroids of each combination of neighbourhoods that exists in the OD-matrix with at least 15 commuters. The API provides this data in JSON format and often contains multiple route options between two neighbourhoods via public transportation. Each route option consists of a number of legs, describing a part of the journey in a particular modality such as a walking, train, bus, metro or waiting period. We project each metro leg of the route on to the metro network. The route description often did not include a straightforward list of segments but textual instructions and a polyline. A polyline is a series of coordinates that gives a schematic overview of the leg and is not guaranteed to contain all exact coordinates of the stations that are passed during the leg of the route. We combined a matching algorithm, the efficiency of a route and the distance between the polyline and stations in the network to determine the most plausible route as intended by each polyline. The matching algorithm calculates text similarity between google maps instructions and station names. It is an adapted version of Ratcliff-Obershelp pattern recognition algorithm implemented in the Python *difflib* package (Python Software Foundation, 2019; Ratcliff & Metzener, 1988). The resulting data set contains the variables shown in Table 3.6.

3.6 Variables in the OD-matrix for public transportation.

Variable	Notation	Description
Neighbourhood departure	R	Unique code for the home location neighbourhood.
Neighbourhood arrival	R'	Unique code for the working location neighbourhood.
Station ID from	v	Unique code for the starting station of the segment.
Station ID to	v'	Unique code for the ending station of the segment.
Duration	$t(R, R', i)$	Travel time of the entire route from the departure neighbourhood to the arrival neighbourhood.
Route index	i	Unique indicator for each route.
Station index	j	Indicator for the order of stations within a route.

4 Methods

This section describes how all of the previously mentioned sources are combined for validation and scenario analysis. First, the RET counts and expected number of commuters are translated to segment level. For every segment, we calibrate a correction factor based on the ratio between expected number of commuters and counts. The scenario of the municipality of Rotterdam is applied to the OD matrix to create a fictive version, which is then projected onto the segments. Applying the previously calculated correction factors to the segments results in an expected amount of travellers for the scenario.

4.1 RET counts to segments

The graph model in Figure 3.1 is used in the translation from counts per trip to counts per segment. We introduce the property *intensity* and set it to zero for all segments. For every trip, the shortest path algorithm is applied to break it down into used segments. Dijkstra's shortest path algorithm is used to determine which segments are part of the shortest trip between two non-neighbouring stations based on the distance of segment (Dijkstra, 1959). We assume that each trip is executed along the shortest available path. The intensity variable for a segment is set to the total of the observed number of travellers over all trips that include the segments in their path. The resulting intensities resemble the number of travellers that used the segment in their trip.

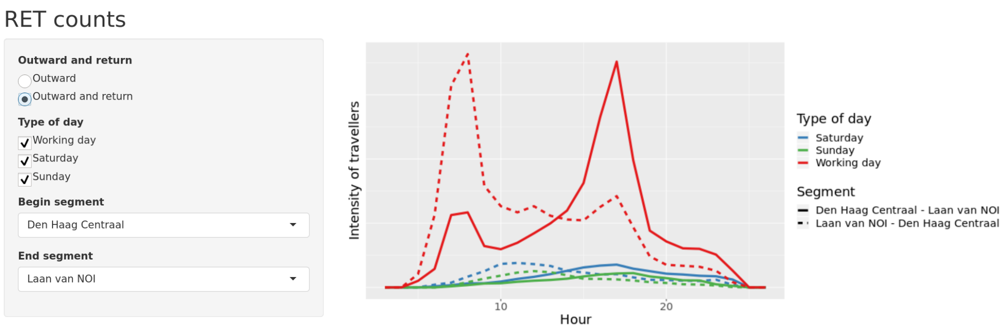
The above process is executed for all days and hours, resulting in a data set described by Table 4.1. At first sight, this data set might seem to resemble Table 3.1, which is based on full trips in the network. The difference between these two data sets is that Table 3.1 is based on trips, whereas Table 4.1 is based on railway segments.

4.1 Variables of the projection of RET counts on the network.

Variable	Notation	Description
Station ID from	v	Unique code for the starting station of the segment.
Station ID to	v'	Unique code for the ending station of the segment.
Hour	h	Value of the hour of day, starting at 3 and ending at 26 because the first two hours of the night are considered part of the previous day for administrative purposes.
Type of day	d	Workday, Saturday or Sunday.
Intensity	$X(v, v', d, h)$	Number of people whose trip included this segment within the specified hour of the specified type of day.

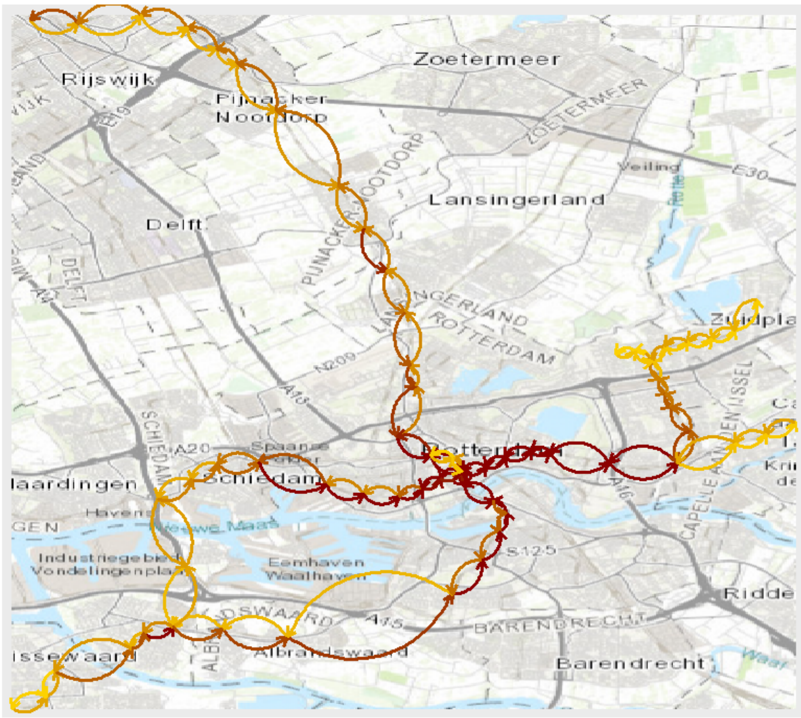
The spatiotemporal data in Table 4.1 is visualised in two ways. Focusing on the temporal component, Figure 4.1 shows the number of travellers for one given segment over time for different days in either direction. One can observe that the morning and afternoon rush hour peaks result in different loads for either direction. Most commuters seem to travel toward Den Haag Centraal in the morning, coming back in the afternoon.

4.1 Visualisation of the information in Table 4.1, the observed number of travellers over the course of a day, for different types of types. Observations are shown for both directions of the segment from Den Haag Centraal to Laan van NOI.



Second, the information is visualised focusing on the spatial component. Figure 4.2 shows the observed number of travellers for all directed segments in the network at a given hour for a given type of day. A video or gif of this visualisation where the hour value ranges over all hours, gives insight into the temporal component (but this is left for the readers fantasy, as it impractically conveyed in pdf format).

4.2 **Visualisation of the information in Table 4.1, the observed number of travellers for both directions of all segments at a specific hour and a specific type of day. Yellow and red arrows indicate few or many travellers, respectively.**



4.2 Statistics Netherlands expectations to segments

The transformation of the neighbourhood-based OD matrix to the expected commuters per segment consists of multiple steps. First, the network infrastructure data was used to construct a graph of the Rotterdam metro network.

The segments used for a journey between neighbourhoods are expressed in relative intensity. When different route options are available, they are weighed based on the inverse of their total travel time of the entire journey:

$$w(R, R', i) := \frac{\frac{1}{t(R,R',i)}}{\sum_{j=1}^{N_R} \frac{1}{t_j(R,R',j)}} \quad i \in \{1, 2, ..., N_R\}, \tag{1}$$

where N_R is the total amount of available routes from neighbourhood R to neighbourhood R', $t(R, R', i)$ is the travel time of route i, and $w(R, R', i)$ is the weight of route i from neighbourhood R to neighbourhood R'. The motivation behind this weighting formula is that a person is likely to choose the fastest route. However, we recognise that the fastest route might not always be the most optimal, due to timing of subsequent activities such as the start of a shift or opening hours. For example, consider someone who has to arrive at their destination at 9:00 and has the options to depart at 8:40 and travel for 20 minutes arriving exactly on time or depart at 8:30 and travel for only 15 minutes but arrive 15 minutes early. They might prefer travelling for 20 minutes over 15 because the time between departure and 9:00 is minimised. To accommodate for this, routes with a longer duration are given a non-zero weight that is smaller than the weight fastest route. The weight is inversely proportional to the route duration. Let us define:

$$w(v, v', R, R') := \sum_{j=1}^{N_R} w(v, v', R, R', j) \mathbb{1}\{\text{segment } (v, v') \text{ is part of route } j \text{ from } R \text{ to } R'\}. \tag{2}$$

The value $w(v, v', R, R')$ is called the relative intensity and models the expected fraction of persons travelling from neighbourhood R to neighbourhood R' that include the segment from station v to station v' in their route.

Combining the expected counts and relative intensities on segments between neighbourhoods, leads to an expected total count per segment:

$$m(v, v') := \sum_{R \in \mathcal{R}} \sum_{R' \in \mathcal{R} \setminus \{R\}} w(v, v', R, R') \cdot k(R, R'), \tag{3}$$

where $k(R, R')$ is the expected amount of public transportation commuters travelling from neighbourhood R to neighbourhood R' and $m(v, v')$ is the expected amount of public transportation commuters travelling from station v to station v'. Note that this expectation is for an unspecified period of time, since no day or time is given. This period can only vaguely be described as the interval of time in which most people travel to their work, and will henceforth be referred to as morning commuting period.

4.3 Validation

Since the expected counts from Statistics Netherlands relate to the morning commuting period, an assumption is necessary on the aggregation of RET hourly data to resemble this period. The morning commuting period is pragmatically assumed to be all morning hours, before noon. The morning starts at hour 3, because of the artificial date line in the data. Whether this is realistic is up for discussion, but the number of travellers is so small in the early morning hours that the result of this discussion is considered irrelevant at this stage. This selection is expected to include the majority of commuter trips towards working locations while excluding the commuter trips towards home. Let us define the morning commuting period as observed in RET data as:

$$n(v, v', d) := \sum_{h=3}^{11} n(v, v', d = \text{Workday}, h) \tag{4}$$

Now that the information of (expected) counts from RET and Statistics Netherlands have been transformed by projecting them onto segments, they can be compared on the same level of granularity. We investigate the ratio c between RET count and Statistics Netherlands expectation per segment, defined as:

$$c(v, v', d) := \frac{n(v, v', d)}{m(v, v', d)}, \tag{5}$$

for all segments form station v to station v'.

The ratio $c(v, v')$ comprises of at least the following shortcomings of our modelling approach:

- The RET counts cover all travellers while the expected counts only cover commuters. Other travel motives such as education and tourism are not included in the expectations. What population bias is introduced by only including commuters?
- Both the route planner and shortest path methods are not necessarily correct in the estimated path through the network.
- The transportation modality model comes with a margin of error (which needs further studying) originating from at least two causes; the fact that it is based on only a sample of the population and its probabilistic nature. It also assumes independence between background variables such as age and driver's license ownership.

Further validation is needed for determining the share of each of the above causes to the error. We expect that the population bias is the biggest contributor and the most inconsistent when comparing ratios between segments. Segments close to a university or touristic location might have a significantly different ratio than a segment in a business park. For example, on a segment near Kralingse Zoom station (next to the university) we observed a relatively high correction factor. The farther away from university, the smaller the correction factor became. We argue that this is because the commuters data does not capture the majority of people that are travelling from and to university. This is an indication that educational data would be beneficial for reducing the location bias in the correction factor.

All additional new real-life data is easily added to the model and helps towards reducing the shortcomings of our model by improving correction factors.

If the correction factors were constant over time and between locations, this could be interpreted as that the ratio shows no location or temporal bias. If each correction factor were equal to 1, this would resemble equality between the expected and observed amount of travellers.

4.4 Calibration

Let us consider the goal of preparing for the analysis of a scenario. We assume that a scenario can be described by a data set that resembles a copy of administrative data with all of its general properties, but with values reflecting the fictive scenario rather than the current real-world situation. We call such a data set a fictional instance of the administrative data. An estimation based on administrative data is valuable because no real-world counts are available for a scenario that has not (yet) occurred in the real world.

The intrinsic information from validating the expectations against the counts is preserved in the ratios defined in Equation (5). Since multiplying the Statistics Netherlands expectation for each segment by the segment ratio equals the RET count, we can say that the expectation can be calibrated by the ratio to create more realistic expectations than those based on the (fictive) administrative data. One could consider calling the ratio a calibration constant.

The calibration constants may be reapplied for new (fictive) administrative data if one assumes that they remain constant (or at least within the same order of magnitude) for each combination of segment, hour and type of day. This assumption might not be valid in all scenarios. For example, when the metro capacity is not scaled according to the growth, travellers might consider alternative modes of transportation. Calibrating is a way of simultaneously compensating for all of the possible errors listed in Section 4.3 that are constant over time.

5 Results

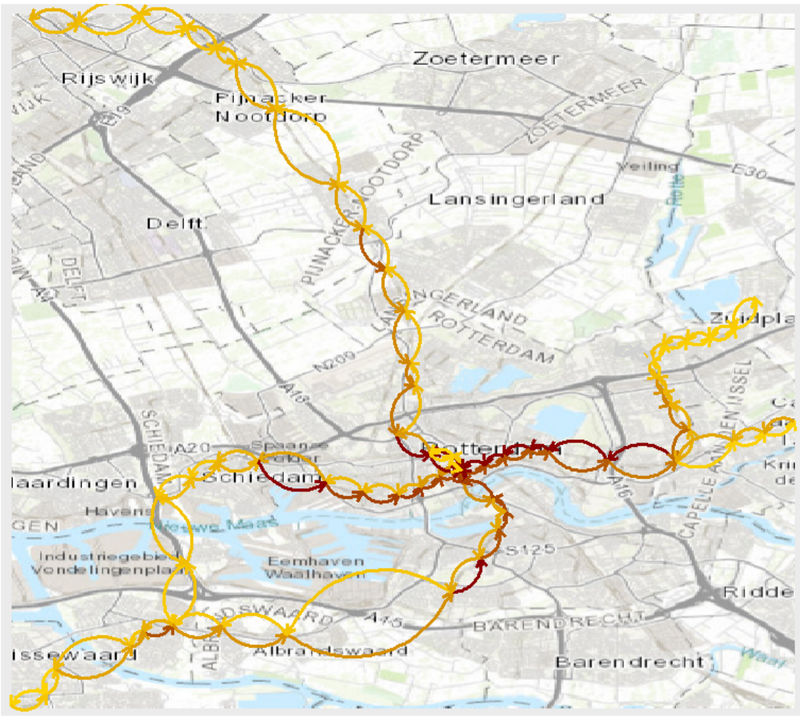
Let us now consider two specific scenarios, that each describe a 10% increase of commuters. In the *pull scenario*, all of these extra commuters are assumed to work in the centre of the city of Rotterdam. In the *push scenario*, the extra commuters are assumed to live in the city centre and work in other neighbourhoods. These rather extreme and unrealistic scenarios are designed to illustrate their effect on the network.

For each scenario, a fictional instance of the original OD matrix of commuters expected to travel with public transport was created by adjusting the original such that the overall increase in commuters is 10%. The fictional OD matrix was then subjected to the same steps as the original OD matrix, projecting the commuters onto the network such that a number of expected commuters were available for each segment. These values per segment were multiplied by the calibration factors to correct for the collection of errors, which is assumed to be constant per segment.

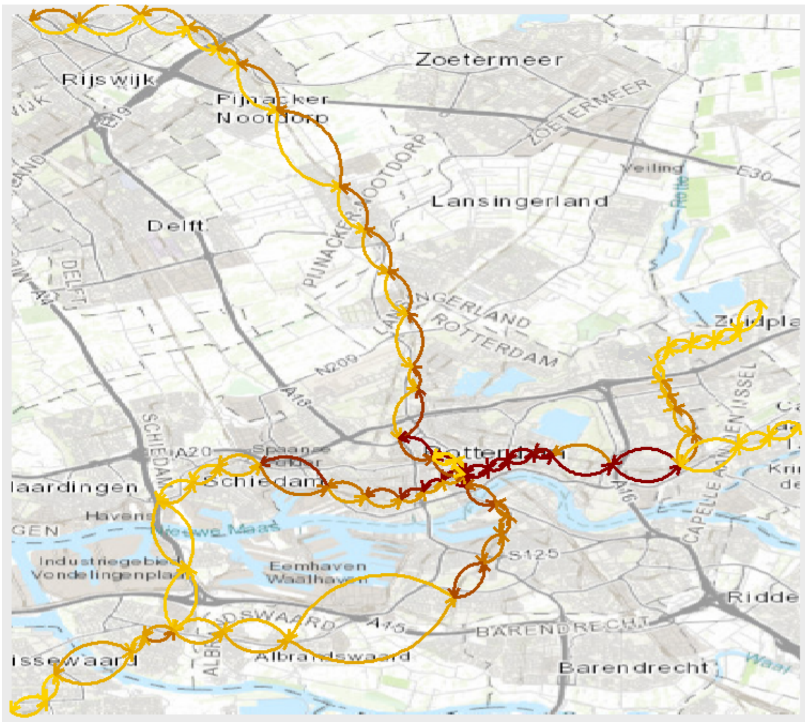
Population growth is but one dimension of expected changes that can be used as an input scenario. Not only can many more be defined on administrative data, scenarios can also be defined on other parts of the input data. For example, infrastructural and policy changes could be modelled to influence transportation mode and route choices.

Figure 5.1 and 5.2 illustrate the results of the calibrated fictional pull and push scenarios. When comparing the expected outcome of the two scenarios, the effects seem to be consistent with what could be expected based on the initial scenario definitions.

5.1 Expected traveller flow for the morning rush hour period in the pull scenario. Yellow and red arrows indicate few or many travellers, respectively.



5.2 Expected traveller flow for the morning rush hour period in the push scenario. Yellow and red arrows indicate few or many travellers, respectively.



6 Conclusion

The visualised results of the scenario analysis correspond to their original scenario descriptions. Though they might help in visualising the definitions of these simple test scenarios, the images themselves only confirm our expectations. However, the pipeline of combining multiple sources and models developed during this study can be reused for the analysis of complicated scenarios whose effect on the city wide mobility are less than obvious.

This study combines multiple sources, that each measure a different side of mobility on different aggregation levels, towards a model for analysing fictional and/or future scenarios. The scenario analysis would not be possible on any single one of the separate sources and we argue that the creation of new insights by combining, calibrating and projecting different sources is the most important contribution of this study.

Future work is envisioned in two directions. First, we plan to include other sources in the presented data ecosystem, expanding on transportation modes, travel motives and region. Potential sources include traffic loop data, mobile phone data and more detailed traveller counts with (aggregated) background characteristics.

Second, we envision a more abstract theory on combining data sources by navigating different granularity levels of the phenomena covered by each source. This framework is a formal method of capturing the synergy between sources and their ability to create new insights that none of the sources could provide individually.

References

CBS (Mar. 2018a). "Onderzoek Verplaatsingen in Nederland 2017". In: *Centraal Bureau voor de Statistiek*.

CBS (2018b). *Towards motives behind mobility*. URL: %5Curl%7Bhttps://www.cbs.nl/en-gb/our-services/innovation/project/towards-motives-behind-mobility%7D (visited on 09/18/2019).

Dijkstra, Edsger W (1959). "A note on two problems in connexion with graphs". In: *Numerische mathematik* 1.1, pages 269–271.

Gemeente Rotterdam (2017). *Openbaar vervoer als drager van de stad, OV-visie Rotterdam 2018-2040*.

Google (2019). *Google Maps Directions API*. URL: %5Curl%7Bhttps://developers.google.com/maps/documentation/directions/start%7D (visited on 10/09/2019).

Neo4j (2019). *The Neo4j Operations Manual v3.5*. URL: %5Curl%7Bhttps://neo4j.com/docs/operations-manual/current/%7D (visited on 11/18/2019).

Python Software Foundation (2019). *difflib - Helpers for computing deltas, difflib.get_close_matches*. URL: %5Curl%7Bhttps://docs.python.org/3.7/library/difflib.html%7D (visited on 11/18/2019).

Ratcliff, John W and David E Metzener (1988). "Pattern-matching-the gestalt approach". In: *Dr Dobbs Journal* 13.7, page 46.

Rijkswaterstaat (2017). *Het Landelijk Model Systeem*.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands

Design
Edenspiekermann

Enquiries
Telephone: +31 88 570 70 70
Via contact form: www.cbs.nl/infoservice