Passenger punctuality

An analysis of the method of calculation and describing models

G. J. Wolters

Technische Universiteit Delft

TU Delft Delft University of Technology

ProRail

Challenge the future

Passenger punctuality

An analysis of the method of calculation and describing models

by

G. J. Wolters

in partial fulfillment of the requirements for the degree of

Master of Science
in Civil Engineering

at the Delft University of Technology
to be defended publicly on Monday March 7, 2016 at 11:00 AM.

Supervisor:	Prof. dr. ir. S. P. Hoogendoorn	
Thesis committee:	Dr. R. M. P. Goverde,	TU Delft
	Ir. V. A. Weeda,	ProRail
	Dr. W. W. Veeneman,	TU Delft
	Ir. P. B. L. Wiggenraad,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Preface

Before you lies the thesis ‘Passenger punctuality – An analysis of the method of calculation and describing models’. It is the result of my research to fulfil the graduation requirements of the master Transport and Planning of the faculty of Civil Engineering and Geosciences of the Delft University of Technology. The research is performed between June 2015 and February 2016.

For eight months I did my research as an intern at ProRail in Utrecht. Here were most resources available for me to analyse and answer the research questions properly. Nonetheless, I could not fully examine passenger punctuality without the data, sources and knowledge available at the NS. Fortunately it was no problem to conduct a part of my research there.

I want to thank several people for their help during my graduation. First of all I want to thank my graduation committee: Serge Hoogendoorn, Rob Goverde and Paul Wiggenraad of the department of Transport and Planning; Wijnand Veeneman of the faculty of Technology, Policy and Management; and Vincent Weeda of ProRail. Their guidance, comments and suggestions were really helpful during my research and while writing my thesis.

Secondly I want to say thanks to the colleagues of the Prestatie Analyse Bureau at ProRail. Everyone was always interested in my work and they included me from the very first day. Special thanks go to Vincent Weeda. As my daily supervisor, Vincent was always prepared to answer questions, read and comment my work and give advice where necessary. He really improved my research. Peter van Waveren helped enormously in the formulating of my research question and guided me through my research. Jan-Martijn Egbers’ knowledge of the KPIs and the process indicators was essential to my research.

I could not have written this thesis without the help of some people of the NS, especially John Tuithof. He was constantly interested in my work and I could always ask anything to him. Besides, he provided me with data essential to this research. Marnix van den Broek was a big help as well. He gave me a better insight in the method of calculating passenger punctuality.

Finally, I want to thank my fellow students with who I worked closely during my time in Delft. I want say thanks to my family and friends as well, for supporting me during my study. My deep gratitude goes to my parents. Their interest, support and love throughout my entire time as a student were essential. Lastly, I want to say thanks to Him, whose unconditional love still amazes me every day.

Geert-Jan Wolters
Delft, February 2016

Summary

The maintenance and control of the Dutch rail network is the responsibility of ProRail. ProRail works closely with the largest train operator, the Nederlandse Spoorwegen (the Dutch Railways, NS). Both companies are accountable for their performance to the Dutch government. They measure their performance through Key Performance Indicators (KPI). In order to improve the performance of the rail network, ProRail and the NS adopted in 2015 the same KPI: passenger punctuality. The current definition is: the percentage of passengers arriving at their final destination within 5 minutes of the planned arrival time. The method of calculation is based on forecasts and the counting of passengers by train conductors.

With the introduction of the OV-chipkaart, the smart card for the Dutch public transport, more data about the passengers became available. Therefore the method of calculating passenger punctuality is going to change and in addition the definition as well. This will improve the KPI, but research in a better understanding of the KPI is lagging behind. A list of process indicators with a presumed correlation with the passenger punctuality is already drawn up, although the influence of these indicators on the KPI is unknown. This correlation might change in time or location as well. In Figure 1 a simplified scheme of the framework around passenger punctuality is visualised. Together this leads to the following main research question:

> **How is the passenger punctuality KPI constructed, and how is it described by the process indicators?**

Literature review

The performance of the NS and ProRail is measured by KPIs. To make the interests of passengers more important, in 2011 the NS introduced a new top KPI: passenger punctuality. In 2015 ProRail adopted this KPI as well. Herewith the Netherlands has quite a unique position in the world. Besides the NS, only the Danish State Railways (DSB) and the Swiss Federal Railways (SBB) measure passenger punctuality. Although the methods of calculation differ in details, the general idea is the same.

The current method of calculating passenger punctuality in the Netherlands is as follows. The number of arriving passengers with a delay less than 5 minutes plus the number of realised passenger transfers is divided by the total number of original scheduled passenger arrivals plus the total number of passengers transferring. Here the arrivals and transfers are measured at 35 stations and the number

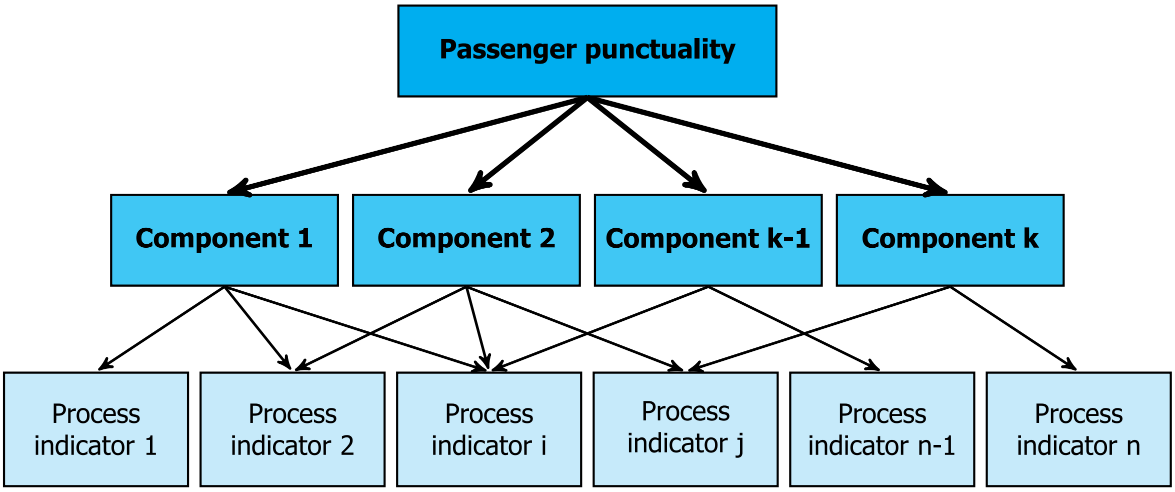


Figure 1: A simplified scheme of the framework around passenger punctuality

of passengers is based on forecasts. The planned arrivals and transfers are based on the timetable 48 hours before execution. Some remarks to this method can be given.

- Passengers transferring are explicitly included in the formula. These passengers are counted twice (and with more transfers even more). This implies that the weight of these passengers is higher than passengers without a transfer.
- Route choice is ignored. If due to a disruption passengers travel via another route, this passenger behaviour is not visible in the outcome of the calculation.
- Arrivals and transfers are measured at only 35 stations. Based on travel patterns, this will result in relatively more arrivals during morning rush hour in comparison with the evening rush hour.

Passenger punctuality 2.0

A new method of calculating the passenger punctuality is currently being developed at the NS. Thanks to the OV-chipkaart the departure and arrival stations and the check-in and checkout times are known. This suggests that for every passenger his or her journey is known. Based on the departure and arrival locations, the check-in time and the timetable two days prior to the execution, for almost every passenger a promised journey (in Dutch: Reisbelofte) can be determined. Combining the promised journey with the realisation data and the checkout time, the arrival time and delay on the final destination of each passenger is known. If the delay is less than 3 minutes, a passenger is considered on time.

The calculation of the passenger punctuality whereby the origin and destination (OD) and departure time of every passenger are known overcomes most disadvantages of the current method of calculating passenger punctuality.

- Every trip is counted once, so the length, number of transfers or direction does not matter, in contrast to the current method of calculation.
- Transfers are not explicitly included in to method of calculation any more. This leads to a more reliable outcome.
- Every possible transfer is taken into account while in the current method of calculation a transfer needs to satisfy a set of criteria based on transfer time and number of passengers transferring.

The new method of calculating the passenger punctuality KPI is called passenger punctuality 2.0. For approximately twenty thousand ODs every travel option is determined. After the execution, the travel options are compared with the realisation data. If a travel option could not have been completed, the journey should have been rescheduled. This data can be combined with the data of the OV-chipkaart. The result is the travel option with the earliest arrival time per trip on the main rail network: the promised journey. The final delay is the delay of the last train at the final destination of the promised journey. If a trip had to be rescheduled, the checkout time is used to determine the arrival time and thus the delay.

Every trip is categorised as well. The category explains why a trip is not used in the calculation or why the checkout time was used to determine the final delay. Eight categories can be distinguished.

1. **Other train operator:** Every passenger that travelled (partly) with another train operator is not taken into account.
2. **Bus transport:** There is no realisation data about bus transport available, so these trips are not used in the calculation.
3. **Train departs too early:** Passengers missing their train because the train departed earlier than planned should reschedule their journey.
4. **Train departs too late:** If a train had a delay of at least 15 minutes before departure, the assumption is made that the passengers rescheduled their journey.
5. **Train did not depart:** The checkout time is used if the promised train did not depart at the origin or transfer station.
6. **Train did not arrive:** If a train was cancelled midway its journey, passengers in this train had to reschedule their journey when the last reached station was not the destination or transfer station of the passengers.
7. **Missed transfer:** A journey needs to be rescheduled if the time between arrival and departure of two consecutive trains is less than the expected walking time between the platforms.
8. **Trip realised:** If the journey is completed as promised, the arrival time of the train is used to determine the arrival time of the passenger.

In the method of calculating passenger punctuality a lot of assumptions are made. In this research the assumptions are analysed and, if possible, the influence on the passenger punctuality is determined. Three assumptions stand out because the impact is large.

- **Departure and arrival:** If there is a disruption that causes large delays or the cancelling of trains, passengers might check in or check out at stations that are not their planned departure or arrival stations. The assumptions are made that these cases are negligible: every passenger checks in or out at his or her desired arrival or departure station. The overall impact of these assumptions is hard to determine because it depends on the location, the time of day and the duration of the disruption. Nonetheless, per case an estimation of the impact can be determined.
- **Walking time:** The assumption is made that the walking time between the card reader (used to check in or to check out) and the platform is 1 minute, independent of the station. If this walking time is specified per station, which implies more realistic walking times, the passenger punctuality is expected to change 0.42 percentage point.
- **Realisation data:** Based on more than 5000 measurements by hand, the average deviation between the exact arrival and departure times of trains at each station and the determined arrival and departure times used in the realisation data is 20 seconds. Moreover there are outliers of more than 1 minute. This difference between the realisation data and the exact realisation can lead to a maximum impact of 0.72 percentage point on the passenger punctuality KPI.

Altogether the estimation of the maximum impact of all assumptions on the passenger punctuality is about 2 percentage points. This implies that the real passenger punctuality can be 2 percentage point lower than the value of the KPI.

Breakdown of the passenger punctuality

From a passengers' perspective, the categories why a journey should be rescheduled are also reasons for a delay. These reasons, or components (see Figure 1), can give information about the performance of the rail network. Therefore, in this research, an algorithm is developed that can breakdown the passenger punctuality into the components. There are six components: the 5 rescheduling categories, where the final delay is at least a threshold value, and the component 'Train delayed'. This last component is within the trips categorized as 'Trip realised', where the final delay is at least a threshold value as well. An example of a resulting breakdown graph can be found in Figure 2. The size of the components is scaled such that the sum of the components at a delay threshold of 180 seconds is equal to 100%.

The breakdown can also be done for a part of the rail network, like an OD, an arriving or transfer station or a train series. Figure 3 shows the passenger punctuality and the shares of the components for passenger trips between stations Rotterdam Centraal and Breda. A sample of the data is taken to create this graph, so the 95% confidence interval is visualized by the two blue lines, just below and above the passenger punctuality line.

Models to describe passenger punctuality

Besides the new method of calculation, ProRail and the NS have created a model to describe passenger punctuality in order to increase the passenger punctuality. Here passenger punctuality is related to the number of Explainable Train Deviations (in Dutch: Te Verklaren Treinafwijkingen, TVTA). A TVTA is a deviation of a train with respect to the timetable. For passenger trains this is mostly due to the cancellation or the delay of the train. The model implies that the more trains are delayed or cancelled the lower the passenger punctuality is. In this research the correlation between the number of TVTAs and the passenger punctuality is calculated. For the business unit Transport control and Traffic control (TB / VL) the correlation is moderate. This implies that TB / VL do not have a large influence on the passenger punctuality.

In the same model, the number of TVTAs is related to process indicators. These process indicators measure the performance of specific processes. If the performance of a process increases, the number of TVTAs might go down and thus the passenger punctuality might improve. For the business unit TB / VL 20 process indicators can be distinguished, whereby five are not yet fully developed and therefore neglected in the remainder of the analysis. For four process indicators the correlation with the number of TVTAs is moderate. For six process indicators the exact definition is wrong: there is a correlation with the number of TVTAs but the value of the process indicator does not represent this. The remaining five process indicators do not have any correlation with the number of TVTAs.

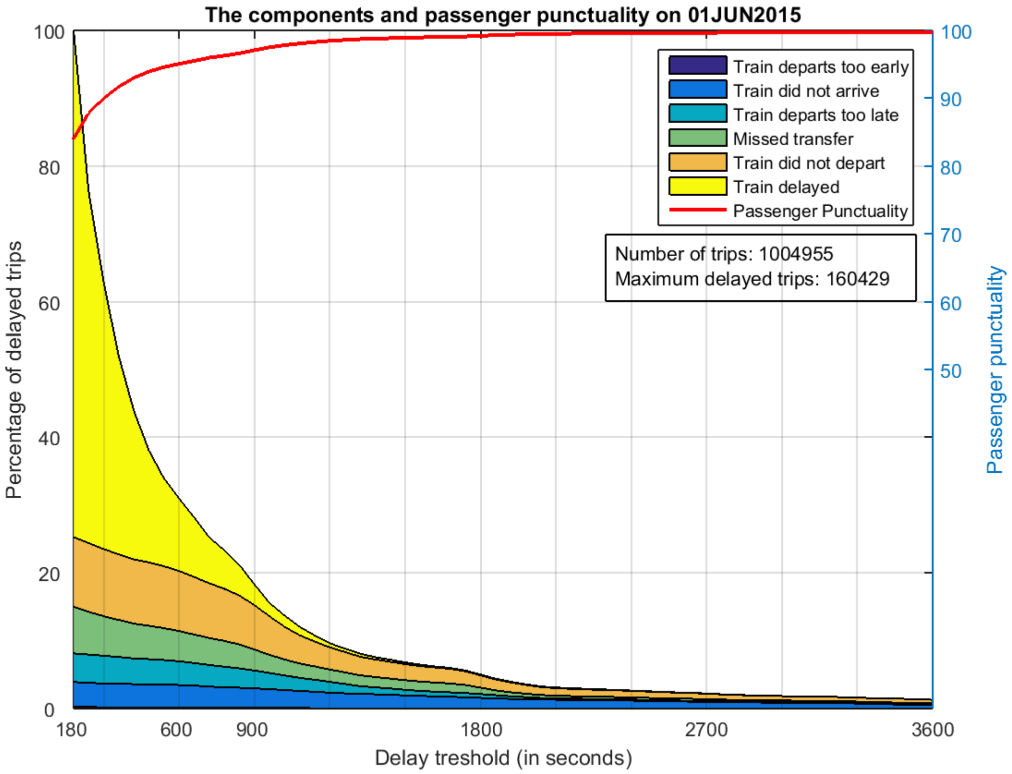


Figure 2: The passenger punctuality and the shares of the components on June 1, 2015

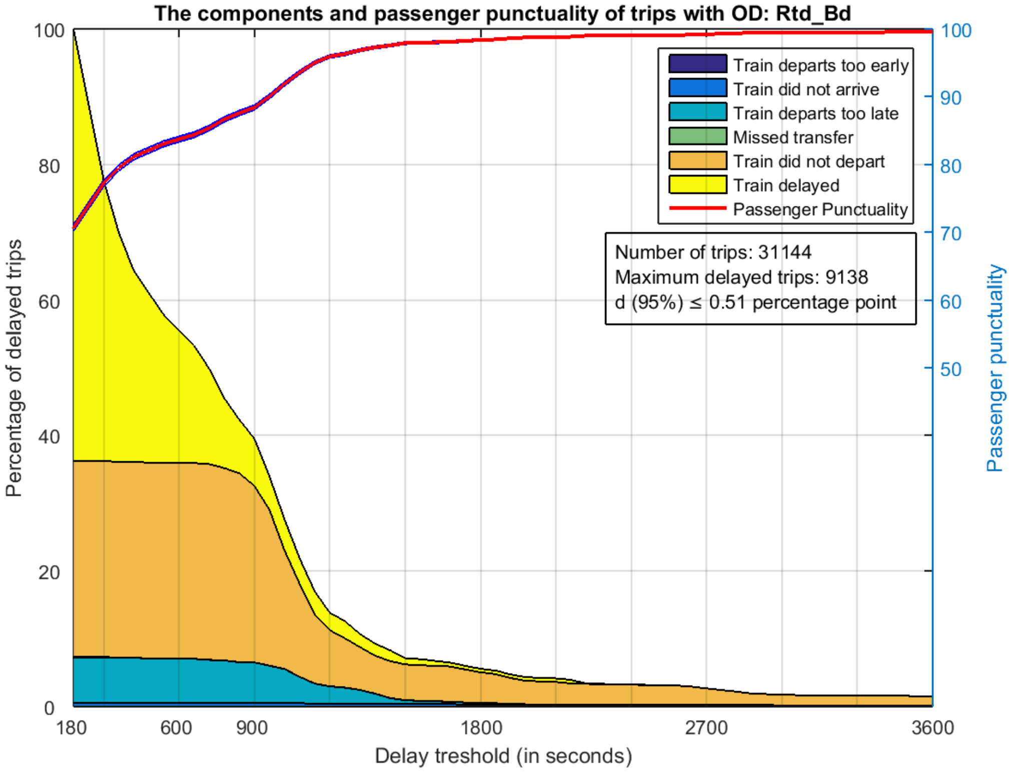


Figure 3: The passenger punctuality and the shares of the components for trips between stations Rotterdam Centraal and Breda

The largest disadvantage of this model is the use of TVTAs. TVTAs might not be very reliable and some types of TVTAs are not taken into account. The main drawback is the fact that TVTAs relate to trains and not to passengers. In other words, the TVTAs are not weighted with the number of passengers involved. Therefore this research proposes another model where the passenger punctuality is split into components. The process indicators are connected to the components and the passenger punctuality itself (see Figure 1). The process indicators are the same as in the model of ProRail and the NS.

The found correlations between the process indicators and the passenger punctuality in this new proposed model are comparable with the correlations between the process indicators and the number of TVTAs in the model op ProRail and the NS. The explanation is that this new model misses a translation between involved trains and involved passengers as well. The process indicators are based on delayed or cancelled trains instead of delayed passengers.

Process indicators for a subset of the rail network

The process indicators can only be calculated for the entire rail network. With the new method of calculating passenger punctuality, this can be calculated for subsets of the rail network, like a smaller region, a train series or a period of time, such as during rush hours as well. This is why an algorithm was developed in this research which is able to calculate the process indicators for a subset of the rail network as well.

In this algorithm the subset can be identified based on three options:

1. **Service control points** (in Dutch: Dienstregelpunt) A region, section of single station of the rail network can be identified via service control points, important geographical points in the rail network like a station or a place where a railway line branches off a main line.
2. **Train series** A single train series, in both or only one direction.
3. **Time of day** The whole day, during rush hours, off-peak hours or another specified time period.

Case study: Utrecht - Leeuwarden / Groningen

Both models are applied in a case study. A breakdown of the passenger punctuality on the study area between station Utrecht Centraal and stations Leeuwarden and Groningen shows that more passengers arrive on time than on average for the entire rail network. This is mostly due to the fact that most transfers are realised and the percentage of passengers that is delayed because of a cancelled train is relatively low.

Calculating the correlation for the study area between the process indicators and the number of TVTAs and the components shows large similarities with the found correlation for the entire network. Nonetheless there are some significant differences, such as the correlation with the departure punctuality of empty rolling stock and the correct uses of train handling documents. For both process indicators the correlation in the study area is non-existing while a moderate correlation applies for the entire network. It might therefore be more efficient to steer on the process indicators in a subset of the rail network than nationwide to increase passenger punctuality.

Conclusions and recommendations

In conclusion, the new method of calculating passenger punctuality is an improvement to the current KPI. Every trip is used and the weight of every trip is the same. The result of the assumptions in this new method is a calculated passenger punctuality that might significantly differ from the real passenger punctuality. A recommendation is to consider a couple of solutions to make the passenger punctuality more reliable, such as taking rescheduling into account, use more realistic walking times and use a delay threshold of 5 minutes instead of the 3 minutes currently used.

The effect of both models describing passenger punctuality could be improved. First of all, for some process indicators the definition should change to reflect the correlation with the number of TVTAs or components better. This will make steering more effective. In both models a translation between delayed and cancelled trains and delayed passengers is missing as well. Therefore a recommendation is to take the expected number of passengers involved into account in the definition of the process indicators. The last conclusion is that correlations might differ between subsets and the entire rail network.

Samenvatting

Het beheer en onderhoud van het Nederlandse spoornetwerk is de verantwoordelijkheid van ProRail. ProRail werkt nauw samen met de grootste reizigersvervoerder, de Nederlandse Spoorwegen (NS). Beide bedrijven moeten verantwoording voor hun prestaties afleggen bij de Nederlandse overheid. De presentaties worden gemeten via Kritieke Prestatie Indicatoren (KPI's). Om de prestaties van het spoornetwerk te verbeteren hebben ProRail en de NS in 2015 dezelfde KPI aangenomen: reizigerspunctualiteit. De huidige definitie is: Het percentage reizigers dat binnen 5 minuten van de geplande aankomsttijd arriveert op hun eindbestemming. De berekeningswijze is gebaseerd op prognoses en reizigerstellingen door conducteurs.

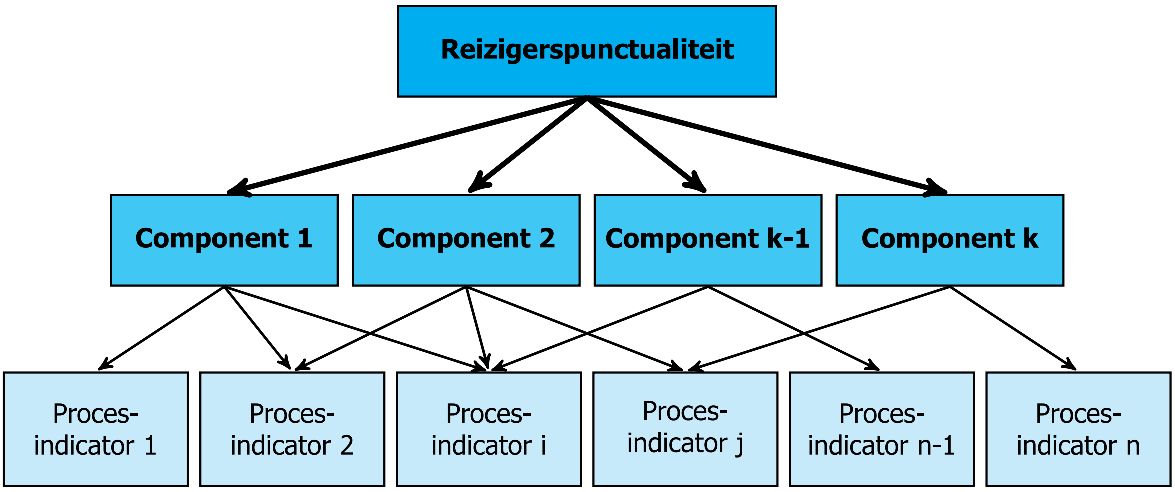
Met de introductie van de OV-chipkaart, de elektronische betaalpas voor het Nederlandse openbaar vervoer, kwam meer data over de reizigers beschikbaar. Hierdoor gaat de berekeningswijze van de reizigerspunctualiteit wijzigen en daardoor ook de definitie. Dit zal de KPI verbeteren maar onderzoek naar een beter begrip van de KPI blijft achter. Er is al een lijst gemaakt met procesindicatoren die een vermoedelijke correlatie met de reizigerspunctualiteit hebben, echter de invloed van deze indicatoren op de KPI is onbekend. Deze correlatie kan ook verschillen per locatie of in de tijd. In Figuur 4 een versimpeld schema van het raamwerk rondom reizigerspunctualiteit is weergegeven. Bij elkaar leidt dit tot de volgende onderzoeksvraag:

> **Hoe is de reizigerspunctualiteit KPI opgebouwd, en hoe kan het via de procesindicatoren beschreven worden?**

Literatuuronderzoek

De prestaties van de NS en ProRail worden gemeten via KPI's. Om de belangen van de reizigers belangrijker te maken, introduceerde de NS in 2011 een nieuwe top KPI: reizigerspunctualiteit. In 2015 begon ProRail deze KPI ook te gebruiken. Hiermee heeft Nederland een unieke positie in de wereld. Behalve de NS, meten alleen de Deense Nationale Spoorwegen (DSB) en de Zwitserse Federale Spoorwegen (SBB) reizigerspunctualiteit. Hoewel de berekeningswijzen in detail verschillen, is het algemene idee hetzelfde.

De huidige berekeningswijze van de reizigerspunctualiteit in Nederland is als volgt. Het aantal arriverende reizigers met een vertraging van minder dan 5 minuten plus het aantal overstappers met een gerealiseerde aansluiting wordt gedeeld door het totaal aantal arriverende reizigers plus het totaal



Figuur 4: Een versimpeld schema van het raamwerk rondom reizigerspunctualiteit

aantal overstappers. De aankomsten en overstaps worden op 35 stations gemeten en de aantallen reizigers zijn gebaseerd op prognoses. De geplande aankomsten en overstaps zijn gebaseerd op de dienstregeling, 48 uur voor de uitvoering. Enkele opmerkingen op deze methode zijn:

- Overstappers zijn expliciet opgenomen in de formule. Deze reizigers worden twee keer geteld (en bij meer overstaps nog vaker). Dit houdt in dat het gewicht van deze reizigers hoger is dan voor reizigers zonder overstap.
- Routekeuze wordt genegeerd. Als er door een versperring reizigers via een andere route reizen, wordt dit gedrag niet weergegeven in de resultaten van de berekening.
- Aankomsten en overstaps worden op 35 stations gemeten. Gebaseerd op reisgedrag leidt dit tot relatief meer aankomsten in de ochtendspit dan in de avondspits.

Reizigerspunctualiteit 2.0

Momenteel wordt er bij de NS een nieuwe berekeningswijze voor de reizigerspunctualiteit ontwikkeld. Dankzij de OV-chipkaart zijn de vertrek- en aankomststations en de check-in en check-uit tijden bekend. Dit suggereert dat voor elke reiziger zijn of haar reis bekend is. Via de vertrek- en aankomstlocaties, de check-in tijd en de dienstregeling van twee dagen voor de uitvoering, kan voor bijna elke reizigers een reisbelofte bepaald worden. Gecombineerd met de realisatiegegevens en de check-uit tijd wordt de aankomsttijd en vertraging op de uiteindelijke bestemming van elke reiziger bekend. Als de vertraging minder dan 3 minuten is, wordt de reiziger als op tijd beschouwd.

De berekening van de reizigerspunctualiteit waarbij de herkomst en bestemming (HB) en vertrektijd van elke reiziger bekend zijn, ondervangt de meeste nadelen van de huidige berekeningswijze van de reizigerspunctualiteit.

- Elke trip wordt één keer meegeteld, dus de lengte, aantal overstaps of richting maakt niet uit, in tegenstelling tot de huidige berekeningswijze.
- Overstaps zijn niet meer expliciet opgenomen in de berekeningswijze. Dit leidt tot een betrouwbaarder uitkomst.
- Elke mogelijke overstap wordt meegenomen, terwijl in de huidige berekeningswijze een overstap aan meerdere criteria rondom overstaptijd en aantal overstappers moet voldoen.

De nieuwe berekeningswijze van de reizigerspunctualiteit KPI wordt reizigerspunctualiteit 2.0 genoemd. For circa twintig duizend HB's wordt elke reismogelijkheid bepaald. Na de uitvoering worden de reismogelijkheden vergelijken met de realisatiegegevens. Als een reismogelijkheid niet voltooid kon worden, de reis zou herpland moeten worden. Deze gegevens kunnen gecombineerd worden met de OV-chipkaart data. Het resultaat is de reismogelijkheid met de vroegste aankomsttijd per trip op het hoofdrailnet: de reisbelofte. De uiteindelijke vertraging is de vertraging van de laatste trein op de eindbestemming van de reisbelofte. Als een trip herpland had moeten worden, wordt de check-uit tijd gebruikt om de aankomsttijd, en dus de vertraging, te bepalen.

Elke trip wordt ook gecategoriseerd. De categorie verklaard waarom een trip niet gebruikt wordt in de berekening of waarom de check-uit tijd was gebruikt om de vertraging te bepalen. Acht categorieën zijn te onderscheiden.

1. **Andere vervoerder:** Elke reiziger die (gedeeltelijk) reist met een andere vervoerder wordt niet meegerekend.
2. **Busvervoer:** Er zijn geen realisatiegegevens van busvervoer beschikbaar, en dus worden deze trips niet meegenomen in de berekening.
3. **Trein te vroeg vertrokken:** Reizigers die hun trein missen doordat deze eerder dan gepland vertrokken is, moeten hun reis herplannen.
4. **Trein te laat vertrokken:** Als een trein voor vertrek minimaal 15 minuten vertraging had wordt de aanname gemaakt dat de reizigers hun reis herplannen.
5. **Trein niet vertrokken:** De check-uit tijd is gebruikt als de beloofde trein niet vertrok van het herkomst- of overstapstation.
6. **Trein niet aangekomen:** Als een trein halverwege zijn reis opgeheven werd, moesten de reizigers in die trein hun reis herplannen, mits het laatste bereikte station niet hun eindbestemming of overstapstation was.
7. **Overstap gemist:** Een reis moet herpland worden als de tijd tussen aankomst en vertrek van twee opeenvolgende treinen minder is dan de verwachte looptijd tussen de perrons.

8. **Reis gerealiseerd:** Al de reis is voltooid zoals beloofd dan wordt de aankomsttijd van de trein gebruikt om de aankomsttijd van de reiziger te bepalen.

In de berekeningswijze van de reizigerspunctualiteit worden veel aannames gemaakt. In dit onderzoek zijn de aannames geanalyseerd en, indien mogelijk, is de invloed op de reizigerspunctualiteit bepaald. Drie aannames vallen op doordat de impact groot is.

- **Vertrek en aankomst:** Als er een verstoring is die leidt tot grote vertragingen of uitgevallen treinen, is het mogelijk dat reizigers in- of uitchecken op stations die niet de geplande vertrek- of aankomststations zijn. Aannames zijn gemaakt waardoor deze gevallen verwaarloosbaar zijn: Elke reiziger checkt in of uit op zijn of haar gewenste vertrek- of aankomststation. De algehele impact van de aannames is lastig te bepalen omdat het afhangt van de locatie, tijd en duur van de verstoring. Echter, per geval kan de impact geschat worden.
- **Looptijd:** De aanname is gemaakt dat de looptijd tussen kaartlezer (om in en uit te checken) en perron gelijk is aan 1 minuut, ongeacht het station. Als de looptijd gespecificeerd wordt per station, wat meer realistische looptijen impliceert, is de verachting dat de reizigerspunctualiteit 0.42 procentpunt veranderd.
- **Realisatiegegevens:** Gebaseerd op meer dan 5000 handmetingen is de gemiddelde afwijking tussen de exacte aankomst- en vertrektijden van treinen op elk station en de bepaalde aankomst- en vertrektijden in de realisatiegegevens 20 seconden. Bovendien zijn er uitschieters van meer dan 1 minuut. Het verschil tussen de realisatiegegevens en de exacte realisatie kan tot een maximum impact van 0.72 procentpunt op de reizigerspunctualiteit KPI leiden.

Alles bij elkaar is de schatting van de maximum impact van alle aannames op de reizigerspunctualiteit ongeveer 2 procentpunten. Dit betekend dat de echte reizigerspunctualiteit 2 procentpunten lager kan liggen dan de waarde van de KPI.

Uitsplitsing van de reizigerspunctualiteit

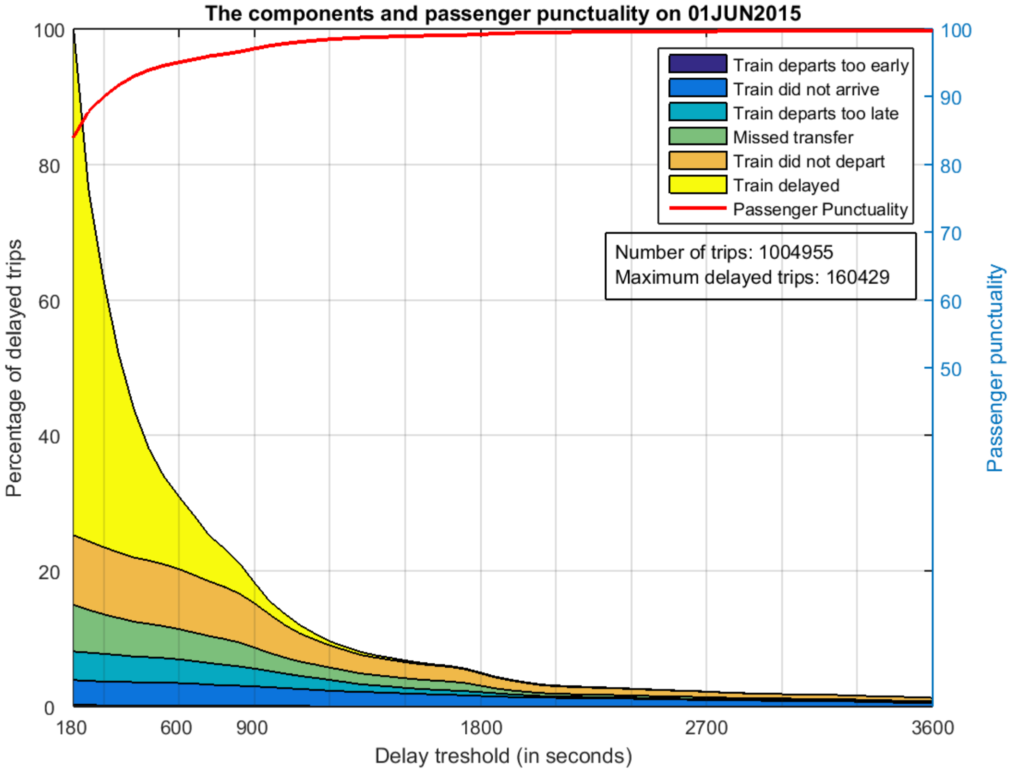
Vanuit een reizigersperspectief zijn de categorieën waarom een reis herpland moet worden ook redenen voor een vertraging. Deze redenen, of componenten (zie Figuur 4), kunnen informatie geven over de prestaties op het spoornetwerk. Daarom is in dit onderzoek een algoritme ontwikkeld dat de reizigerspunctualiteit kan uitsplitsen naar de componenten. Er zijn zes componenten: de 5 herplan-categorieën, waar de uiteindelijke vertraging minimaal gelijk is aan een grenswaarde, en de component 'Trein vertraagd'. Deze laatste component valt binnen de trips met de categorie 'Reis gerealiseerd', waar ook hier de uiteindelijke vertraging gelijk is aan een grenswaarde. Een voorbeeld van een uitsplitsingsgrafiek staat in Figuur 5. De grootte van de componenten is opgeschaald zodat de som van de componenten bij een grenswaarde van 180 seconden gelijk is aan 100%.

Een gedeelte van het spoornetwerk, zoals een HB, een aankomst- of overstapstation of een treinserie, kan ook uitgesplitst worden. Figuur 6 geeft de reizigerspunctualiteit en het aandeel van de componenten weer voor reizigers tussen station Rotterdam Centraal en Breda. Omdat een steekproef van de data is genomen om deze grafiek te maken, is het 95% betrouwbaarheidsinterval weergegeven via twee blauwe lijnen, net onder en boven de reizigerspunctualiteitlijn.

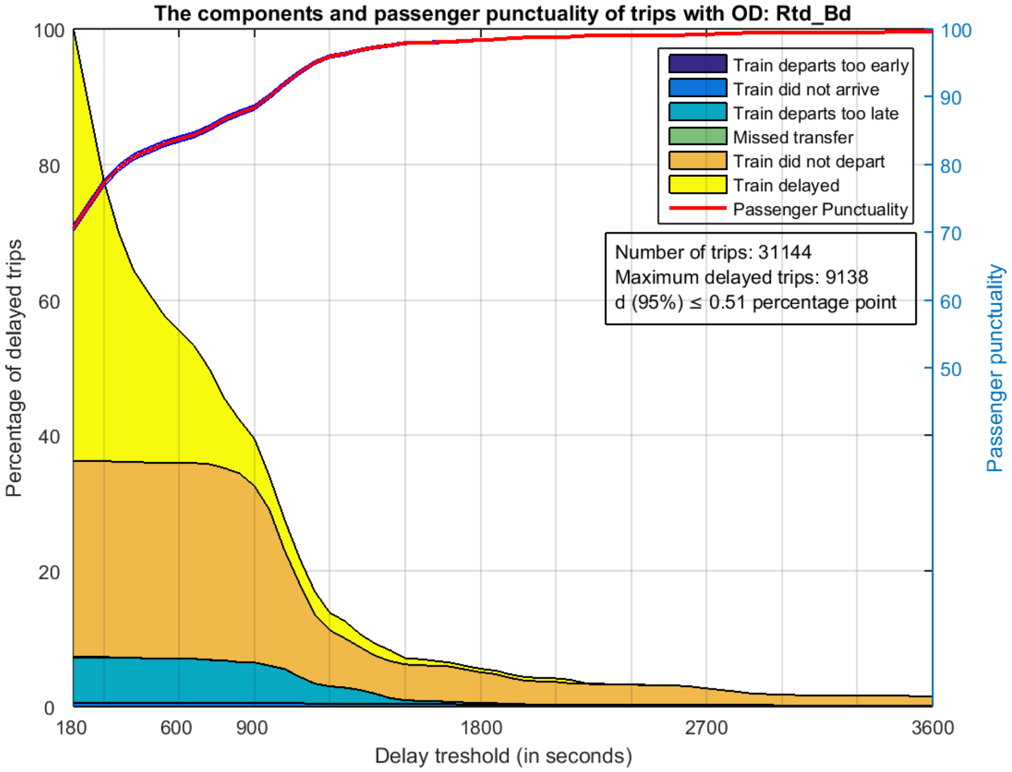
Modellen die de reizigerspunctualiteit beschrijven

Behalve de nieuwe berekeningswijze hebben ProRail en de NS ook een model gemaakt die de reizigerspunctualiteit beschrijft, om zo de reizigerspunctualiteit te verbeteren. Hierin is de reizigerspunctualiteit gerelateerd aan het aantal Te Verklaren Treinafwijkingen (TVTA). Een TVTA is een afwijking van een trein ten opzichte van de dienstregeling. Voor reizigerstreinen is dit meestal door opheffing of vertraging van de trein. Het model veronderstelt dat als er meer treinen vertraagd of opgeheven zijn, de reizigerspunctualiteit lager zal zijn. In dit onderzoek is de correlatie tussen het aantal TVTA's en de reizigerspunctualiteit berekend. Voor de bedrijfseenheid Transportbesturing en Verkeersleiding (TB / VL) is de correlatie is matig. Dit betekend dat TB / VL geen grote invloed op de reizigerspunctualiteit heeft.

In hetzelfde model wordt het aantal TVTA's gerelateerd aan procesindicatoren. Deze procesindicatoren meten de prestaties van specifieke processen. As de prestaties van een proces verbetert, zal het aantal TVTA's mogelijk minder worden en dus kan de reizigerspunctualiteit verbeterd worden. Voor de bedrijfseenheid TB / VL kunnen 20 procesindicatoren onderscheiden worden, waarvan 5 nog niet volledig ontwikkeld zijn en daardoor niet meegenomen zijn in het vervolg van de analyse. Voor



Figuur 5: De reizigerspunctualiteit en het aandeel van de componenten op 1 juni 2015



Figuur 6: De reizigerspunctualiteit en het aandeel van de componenten voor trips tussen station Rotterdam Centraal en Breda

vier procesindicatoren is de correlatie met het aantal TVTA's matig. Voor zes procesindicatoren is de precieze definitie verkeerd: er is een correlatie met het aantal TVTA's maar de waarde van de procesindicator geeft dit niet weer. De overige vijf procesindicatoren hebben geen correlatie met het aantal TVTA's.

Het grootste nadeel aan dit model is het gebruik van de TVTA's. TVTA's zijn niet heel betrouwbaar en sommige TVTA's zijn niet eens meegenomen. Het grootste bezwaar is het feit dat TVTA's gerelateerd zijn aan treinen en niet aan reizigers. Oftewel, de TVTA's worden niet gewogen met het aantal betrokken reizigers. Hierom wordt in dit onderzoek een ander model voorgesteld, waar de reizigerspunctualiteit is opgesplitst in componenten. De procesindicatoren zijn verbonden aan de componenten en aan de reizigerspunctualiteit zelf (zie Figuur 4). De procesindicatoren zijn hetzelfde als in het model van ProRail en de NS.

De gevonden correlaties tussen de procesindicatoren en de reizigerspunctualiteit in dit nieuwe model zijn vergelijkbaar met de correlaties tussen de procesindicatoren en het aantal TVTA's in het model van ProRail en de NS. De verklaring is dat dit nieuwe model ook een vertaalslag tussen betrokken treinen en betrokken reizigers mist. De procesindicatoren zijn gebaseerd op vertraagde of opgeheven treinen in plaats van vertraagde reizigers.

Procesindicatoren voor een deelgebied van het spoornetwerk

De procesindicatoren kunnen alleen berekend worden voor het hele spoornetwerk. Met de nieuwe berekeningswijze voor reizigerspunctualiteit, kan dit ook berekend worden voor een deelgebied van het spoornetwerk, bijvoorbeeld een klein gebied, een treinserie of een tijdsperiode zoals spitsuur. Dit is waarom er in dit onderzoek een algoritme was ontwikkeld dat in staat is om procesindicators uit te rekenen voor een deelgebied.

In dit algoritme het deelgebied kan geïdentificeerd worden via drie opties:

1. **Dienstregelpunt** Een gebied, spoorsectie of enkel een station kan geïdentificeerd worden met dienstregelpunten, belangrijke geografische punten in het spoornetwerk, zoals een station of een plek waar een zijspoorlijn afbuigt van een hoofdspoorlijn.
2. **Treinserie** Een treinserie, in beide of in een enkele richting.
3. **Tijdsmoment** De hele dag, gedurende spitsuren, daluren of een andere aangegeven tijdsperiode.

Casestudie: Utrecht - Leeuwarden / Groningen

Beide modellen zijn toegepast in een casestudie. Een uitsplitsing van de reizigers op het studiegebied tussen station Utrecht Centraal en stations Leeuwarden en Groningen laat zien dat meer reizigers op tijd aankomen dan gemiddeld voor het hele spoornetwerk. Dit komt voornamelijk doordat meer overstaps gehaald worden en het percentage reizigers dat vertraagd is door dat een trein is opgeheven relatief laag is.

De berekende correlaties voor het studiegebied tussen de procesindicatoren en het aantal TVTA's en de componenten laat grote overeenkomsten met de gevonden correlaties in het hele spoornetwerk zien. Er zijn echter een paar significante verschillen, zoals de correlatie met de vertrekpunctualiteit van leeg materieel en het correct uitvoeren van trein afhandelings documenten. Voor beide procesindicatoren is er geen correlatie in het studiegebied terwijl in het hele spoornetwerk een matige correlatie geldt. Het kan daardoor efficiënter zijn om op procesindicatoren in een deelgebied van het spoornetwerk te sturen dan landelijk om de reizigerspunctualiteit te verbeteren.

Conclusies en aanbevelingen

Concluderend, de nieuwe berekeningswijze van de reizigerspunctualiteit is een verbetering voor de huidige KPI. Elke trip wordt gebruikt en het gewicht van elke trip is hetzelfde. Het resultaat van de aannames in deze nieuwe methode is een berekende reizigerspunctualiteit die significant verschilt van de echte reizigerspunctualiteit. Een aanbeveling is om enkele oplossingen om de reizigerspunctualiteit betrouwbaarder te maken te overwegen, zoals het rekening houden met herplannen, meer realistische looptijden gebruiken en een grenswaarde van 5 minuten aanhouden in plaats van de 3 minuten waarmee nu gewerkt wordt.

Het effect van beide modellen die reizigerspunctualiteit beschrijven kan verbeterd worden. Allereerst, voor sommige procesindicatoren moet de definitie veranderen om de correlatie met het aantal TVTA's of de componenten beter weer te geven. Dit kan sturen effectiever maken. In beide modellen mist ook een vertaalslag tussen vertraagde en opgeheven treinen en vertraagde reizigers. Daarom een aanbeveling is om de verwachte aantal betrokken reizigers mee te nemen in de definitie van de procesindicatoren. De laatste conclusie is dat correlaties kunnen verschillen tussen deelgebieden en het hele spoornetwerk.

Contents

1

Introduction

Passenger punctuality is simply the percentage of passengers that arrive on time at their desired location. For the rail network, the Netherlands is one of the few countries calculating this performance indicator. With the introduction of the OV-chipkaart, the smart card for public transport, the passenger punctuality can be measured even more accurate. With the focus of the Nederlandse Spoorwegen (Dutch Railways, NS) to give interests of passengers the top priority, it becomes even more important to achieve the best passenger punctuality as possible. However the correlation between the underlying processes and the passenger punctuality is unknown. This makes steering and improving the passenger punctuality difficult.

This first chapter is an introduction to the research. First, in paragraph 1.1, the problem regarding the passenger punctuality, is defined. This problem definition ends with the drafting of the research question. In paragraph 1.2 this main research question is given, as well as the research approach and the five sub questions. In paragraph 1.3 the relevance of this research in described. Finally, in paragraph 1.4 the outline of this report can be found.

1.1. Problem definition

The responsibility of the Dutch rail network lies with ProRail. Operations include among others the design and building of new infrastructure, the maintenance and upgrading of the existing infrastructure, the building and improving of the timetables and the daily control of the trains on the network. ProRail works closely with the NS. This is the largest rail operator on the rail network. Responsibilities include domestic and international passenger rail transport, the purchase of new trains and the maintenance of the existing trains and also the development of the stations.

Both ProRail and the NS have to answer for their performance to the Dutch government. They are being reviewed based on a set of Key Performance Indicators (KPIs) including train punctuality. Train punctuality is the percentage of trains arriving within a certain threshold on a certain set of stations. The exact definition differs between ProRail and the NS. This KPI has some drawbacks. For example cancelled trains and the number of passengers per train are not included. Also not every station is being reviewed. Therefore the NS introduced in 2011 a new KPI: passenger punctuality, the percentage of passengers arriving with a maximum delay of 4:59 minutes at their final destination. In 2015 ProRail started to use this KPI as well.

The goal of both ProRail and the NS is to describe the passenger punctuality KPI in order to improve it; however research to give a better understanding in this KPI is lagging behind. The components that cause delays for individual passengers are for the greater part known (delayed or cancelled trains and missed transfers) though in which degree the influence of each component extent is not known. This influence can change in time or location as well. In the framework around passenger punctuality these components compose the first layer below the KPI, see Figure 1.1.

The second layer consists of the process indicators. The process indicators are composed by the Traffic control of ProRail and the Transport control of the NS. They are representing the performance of processes and can influence a component. The reasoning behind these indicators is as follows. Every

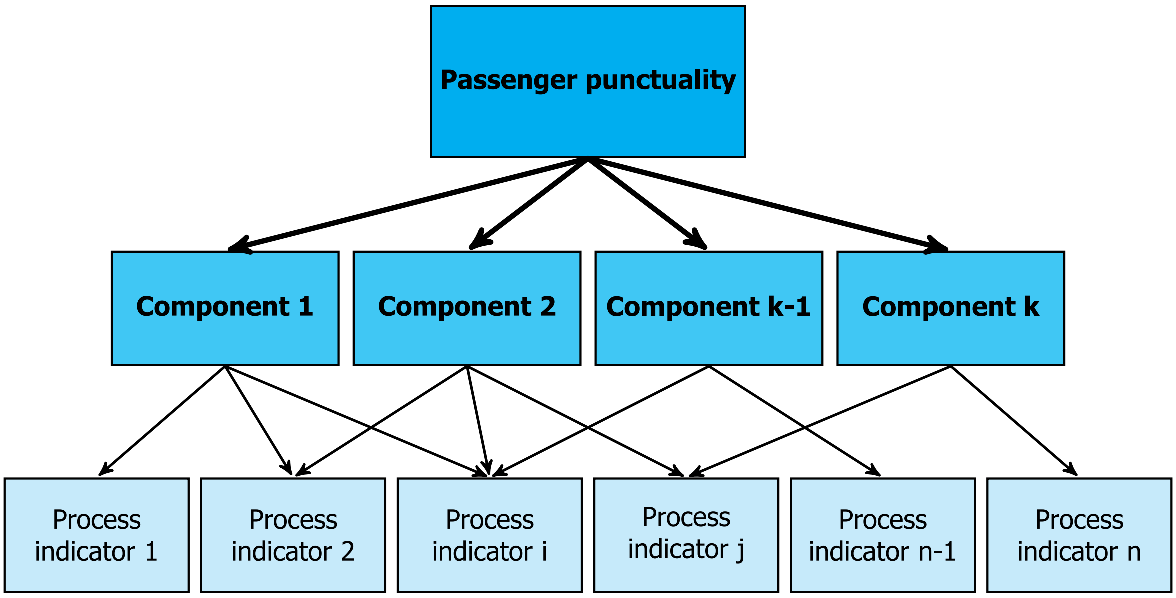


Figure 1.1: A simplified scheme describing passenger punctuality

indicator represents a part of the framework. If a process performs better, the associated indicator will go up. This will finally result in better passenger punctuality. Nonetheless the influence of each process indicator on the KPI is unknown. The same applies to the mutual relations.

1.2. Research question and approach

The need for a better insight in the passenger punctuality KPI and in particular in the framework around the KPI leads to the main research question of this research:

> **How is the passenger punctuality KPI constructed, and how is it described by the process indicators?**

Here, process indicators are the indicators associated with processes related to Traffic control and Transport control. Steering on these processes leads to a better performance and this might lead to higher passenger punctuality.

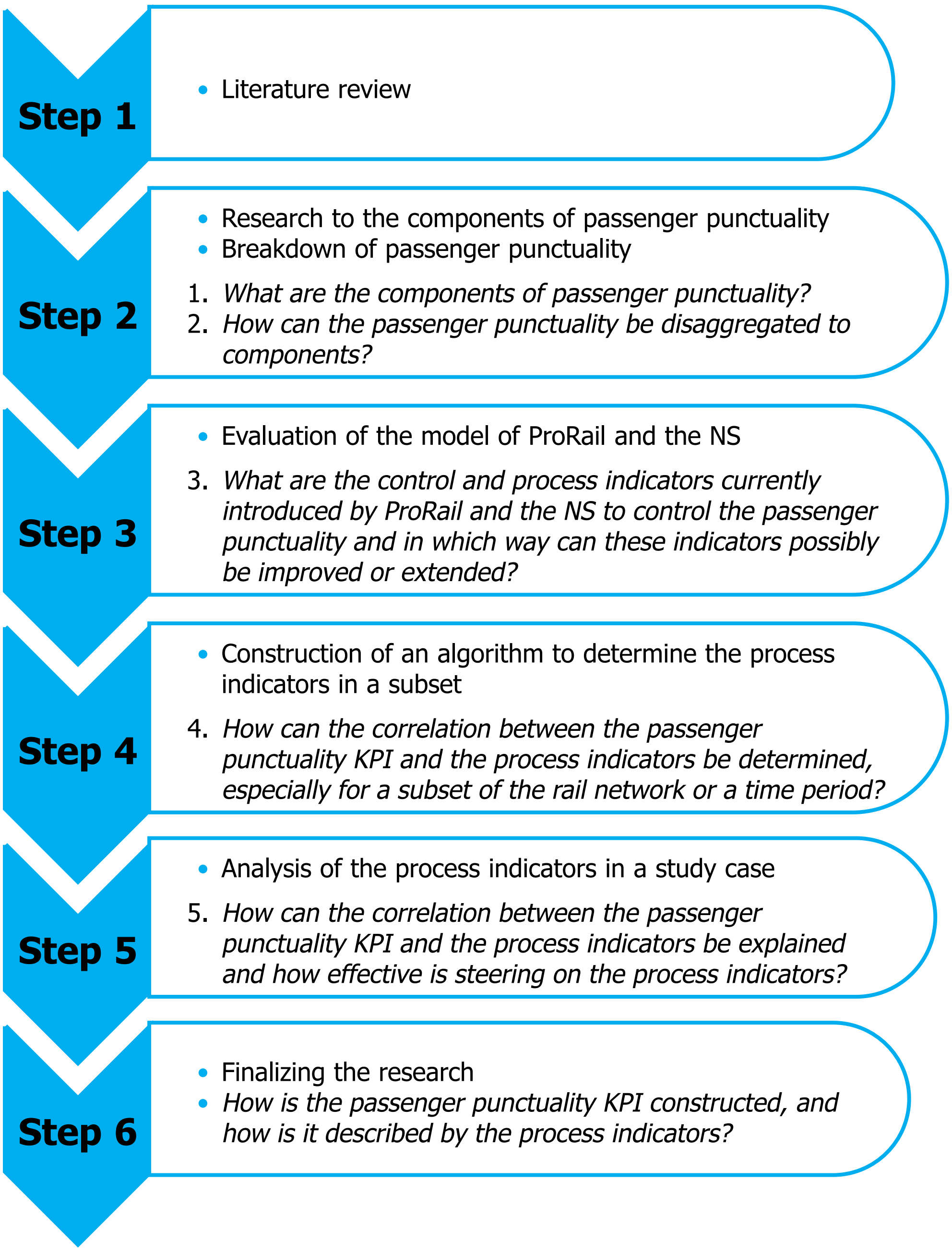
To achieve a solid answer there is need for some sub questions. The goal of the main research question is to explore the method of calculating passenger punctuality and to decompose the framework on which passenger punctuality is built from. This will eventually lead to ways to improve the passenger punctuality. The sub questions lay the foundation to the answer from the main research question.

A scheme of the whole research approach can be found in Figure 1.2. This research will start with a literature study. In this study the current state of research into passenger punctuality is analysed. The objective is to become familiar with the terminology and the background of the KPI. Special attention will be given to the methods of calculating passenger punctuality. The goal of this study is to gain insight in the passenger punctuality KPI: Why is it necessary, how is it defined, how is it structured and what are the benefits and the drawbacks? The literature study will answer these questions. With this a profound evaluation of the newest method of calculating passenger punctuality will be given. The goal is to find the margin of error and to determine how exact the passenger punctuality will be.

The first and second sub questions are formulated as follows:

1. What are the components of passenger punctuality?
2. How can the passenger punctuality be disaggregated to components?

The focus of these two sub questions lies on the different components of passenger punctuality. To start, every component needs to be detected. Secondly the passenger punctuality KPI will be disaggregated to every found component. For the passengers, there are two moments in time where the point



Step 1

- Literature review

Step 2

- Research to the components of passenger punctuality
- Breakdown of passenger punctuality

1. *What are the components of passenger punctuality?*
2. *How can the passenger punctuality be disaggregated to components?*

Step 3

- Evaluation of the model of ProRail and the NS

3. *What are the control and process indicators currently introduced by ProRail and the NS to control the passenger punctuality and in which way can these indicators possibly be improved or extended?*

Step 4

- Construction of an algorithm to determine the process indicators in a subset

4. *How can the correlation between the passenger punctuality KPI and the process indicators be determined, especially for a subset of the rail network or a time period?*

Step 5

- Analysis of the process indicators in a study case

5. *How can the correlation between the passenger punctuality KPI and the process indicators be explained and how effective is steering on the process indicators?*

Step 6

- Finalizing the research
- *How is the passenger punctuality KPI constructed, and how is it described by the process indicators?*

Figure 1.2: Scheme of the research approach

of view regarding their trip changes. The first moment is at a delay of 3 minutes. Before this moment the average passenger is satisfied with the journey time, after this moment their judgement decreases a bit. The second moment is at a delay of 15 minutes. After this moment, the judgement of the average passenger decreases strongly: their trip is disturbed. For both time moments, a breakdown of the passenger punctuality will be done. It is more desirable to improve the 15 minutes passenger punctuality because of the judgement drop after a 15 minutes delay. On the other hand, ProRail and the NS are not reviewed by the government based on the 15 minute passenger punctuality. Besides it is fair to assume that the 3 minute passenger punctually is lower and possibly offers more room for improvement.

In the next step, the focus will shift to the process indicators. These indicators are determined by the Traffic control of ProRail and the Transport control of the NS. First off, the indicators will be analysed to see how they are defined, how they are calculated and to which component they are related. Next they will be determined if they could be improved or otherwise be neglected. With this information the third sub question can be answered:

3. What are the control and process indicators currently introduced by ProRail and the NS to control the passenger punctuality and in which way can these indicators possibly be improved or extended?

The influence of every indicator might differ in time and place. Therefore it might be impossible to find the correct correlation between every single process indicator and the passenger punctuality KPI. Nonetheless it is possible to construct an algorithm to find the value of each indicator and the passenger punctuality for almost every subset. Herewith the correlations can be determined. This is the basis for the fourth sub question:

4. How can the correlation between the passenger punctuality KPI and the process indicators be determined, especially for a subset of the rail network or a time period?

At this point, the focus of the research shifts to a specific subset of the passenger punctuality. This subset will be used to validate the algorithm. The correlation with the indicators and the passenger punctuality in this subset are calculated as well. The results are analysed to determine the effectiveness of steering on the process indicators. With this the last sub question can be answered:

5. How can the correlation between the passenger punctuality KPI and the process indicators be explained and how effective is steering on the process indicators?

1.3. Relevance of the research

This research contributes in a better insight in the passenger punctuality KPI. First of the method of calculating passenger punctuality is examined. This will make clear where and when the calculated passenger punctuality differs from the actual passenger punctuality. It also points out if and where the method can be improved to achieve a more reliable KPI.

From a passengers' perspective, this research is also relevant. The causes of bad passenger punctuality are going to be decomposed. The result is not only a list of possible causes but also the degree of influence. This could be on a national level but for a local train line as well. A distinction in time is also possible. Difference between rush hours and off peak hours can be researched, but between work and weekend days as well. Altogether this breakdown of the passenger punctuality leads to a clear insight in the causes: the causes itself but also in the influence of every cause.

The relations between the control indicators and the passenger punctuality itself are researched as well. This offers a good insight in the control framework of influence to the passenger punctuality KPI. The influence of the indicators becomes known so a more optimal solution can be found to achieve the best passenger punctuality possible. This applies particularly to specific situations, in time or location. This research will investigate such a situation to find the influence of the indicators concerned. For the future, this will help by analysing other situations with bad passenger punctuality.

1.4. Outline of this report

This research will start with a literature review in chapter 2. Here the history of the passenger punctuality KPI is explained, other relevant research is analysed and a brief overview of the methods of calculating passenger punctuality is given. Chapter 3 will examine the latest method of calculating passenger punctuality more closely. Special attention will be paid to the underlying assumptions. Hereby the margin of error could be determined and thus the reliability of the passenger punctuality KPI.

In chapter 4 the passenger punctuality is decomposed into components for different subsets of the rail network. This can by a time period, but also an origin-destination combination, a train series or a station.

The analyses of the models to describe the passenger punctuality can be found in chapter 5. First the model of the NS and ProRail is analysed. Attention goes to the control indicators and the Explainable Train Deviations (TVTAs). Second a new model is proposed. Here the components of the passenger punctuality are used to connect the passenger punctuality with the process indicators. Next the process indicators are elaborated and the correlation with the TVTAs and the components is determined.

To perform a more detailed analysis of a subset of the rail network, an algorithm is created to calculate the process indicators for such a subset. This algorithm is discussed in chapter 6. Finally a small case study is performed to validate the results of this algorithm (chapter 7). Hereby also an analysis is performed to check the correlation between the process indicators and the TVTAs and the passenger punctuality. This report ends with some conclusions and recommendations, in chapter 8.

Five appendixes are included in this report. Appendix A is an extension to chapter 3. It provides some additional analyses and summary tables used in this chapter. To perform the breakdowns of the passenger punctuality a new type of graph is used. The algorithm to create these breakdown graphs is explained in appendix B. Appendix C provides a list of all the TVTAs in use by ProRail. The algorithm that is made to calculate the process indicators is explained in appendix D.

2

Literature review

This chapter describes the current developments regarding the passenger punctuality KPI (Key Performance Indicator). The focus will mainly be on the Netherlands, but when examining the current state of the research concerning the passenger punctuality the view will be widened. Altogether this chapter will provide some background information and will establish the starting point of the research.

First off, in paragraph 2.1, the origin of the rail network in the Netherlands, the NS and ProRail is explained. Paragraph 2.2 elaborates on the origin of the passenger punctuality in the Netherlands. The main question here will be: how is the passenger punctuality KPI originated? Not only the Netherlands, but also the Swiss and Denmark use passenger punctuality as a performance indicator. Research of these countries concerning passenger punctuality and more are reviewed in paragraph 2.3. Finally the methods to calculate the passenger punctuality in use in the Netherlands are explained and compared. This can be found in paragraph 2.4. This chapter ends with some conclusions (paragraph 2.5).

2.1. The rail network of the Netherlands

Since the first train line opened on the 20th of September 1839 the rail network in the Netherlands has grown and changed rapidly. In 2014 there were 3.032 kilometre of train tracks of which 2.307 km was electrified [1]. Daily around 5500 trains run [2] and the 410 stations handle more than a million passengers [3]. In 2009 Statistics Netherlands (Centraal Bureau voor de Statistiek, CBS) published a report concluding that the Netherlands had the busiest rail network of the whole European Union [4]. With yearly more than 20 thousand train kilometres per kilometre train track the network is almost twice as crowded as the European average, see Figure 2.1. By taking the number of passengers per train into account, this results in almost 2.5 million passenger kilometres per kilometre train track in the Netherlands. This is the highest position of the EU (see Figure 2.2). Looking at the train punctuality numbers, the Netherlands is at the third place worldwide [5]. Figure 2.3 shows that only Switzerland and Japan have higher train punctuality and a busier rail network.

In 1938, 100 years after the first train line opened, the Dutch Railways (Nederlandse Spoorwegen, NS) was founded as a fusion of the two largest railway companies in the Netherlands. All the shares were owned by the Dutch government. The NS exploited the main rail network and most of the regional lines. Until 1995 the management and the exploitation of the rail network were the responsibility of the NS. From this point on, due to European law, this could no longer be the case [6].

The NS continued to focus on the exploitation. Tasks include the transport of passengers, both domestic as international, purchase and maintenance of rolling stock and the development of stations. Former sections of the NS like infrastructure maintenance and rail control came directly in service of the Ministry of Infrastructure and the Environment. In 2005 these sections were combined into an independent company called ProRail. Responsibilities of ProRail include the construction of new rail connections or stations and the maintenance and the upgrading of the existing infrastructure but also safety issues are the concern of ProRail. Another field of operation is the distribution of the available train paths among the different train operators. Lastly the control of all the trains on the rail network is part of the responsibilities of ProRail.

Nowadays the NS is not the only rail operator in the Netherlands. The NS serves the main rail

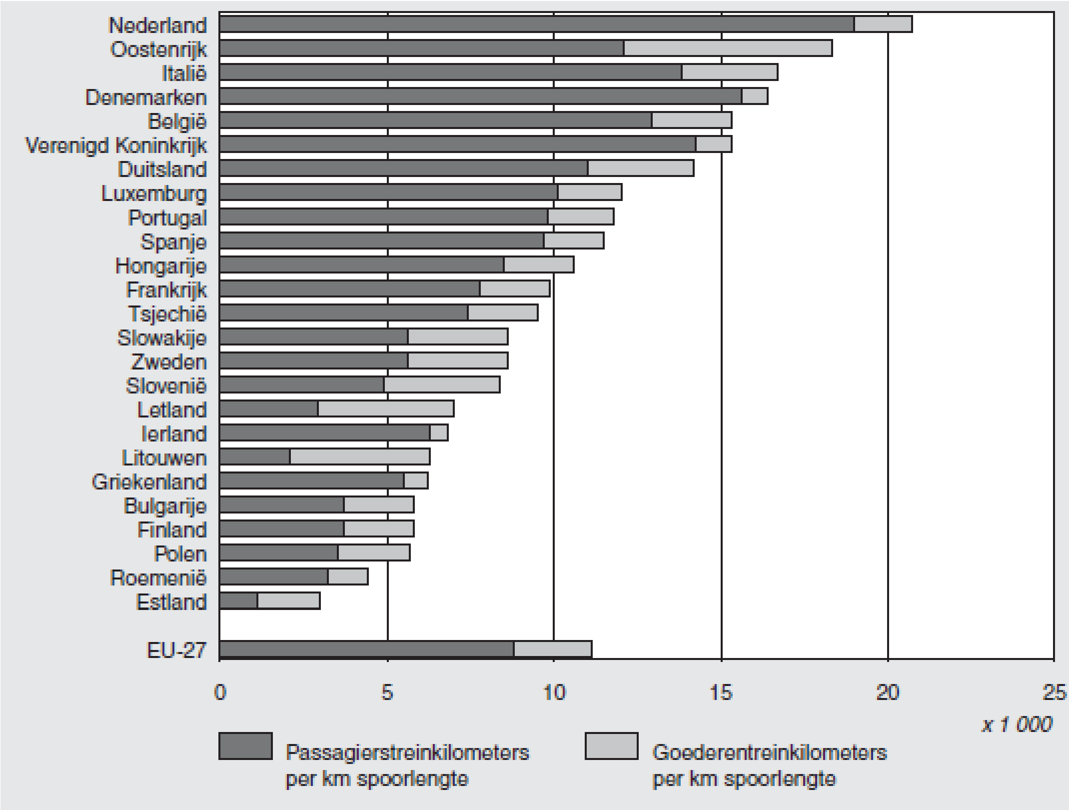


Figure 2.1: Train traffic per kilometre train track in the EU in 2006 [4]

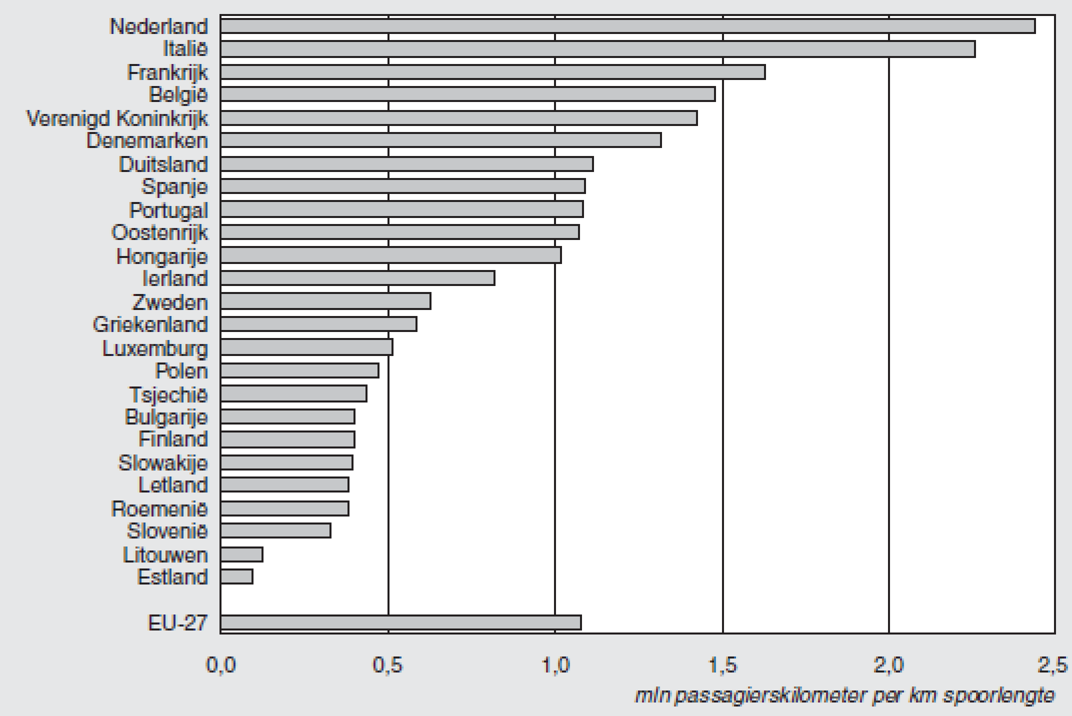


Figure 2.2: Passengers per kilometre train track in the EU in 2006 [4]

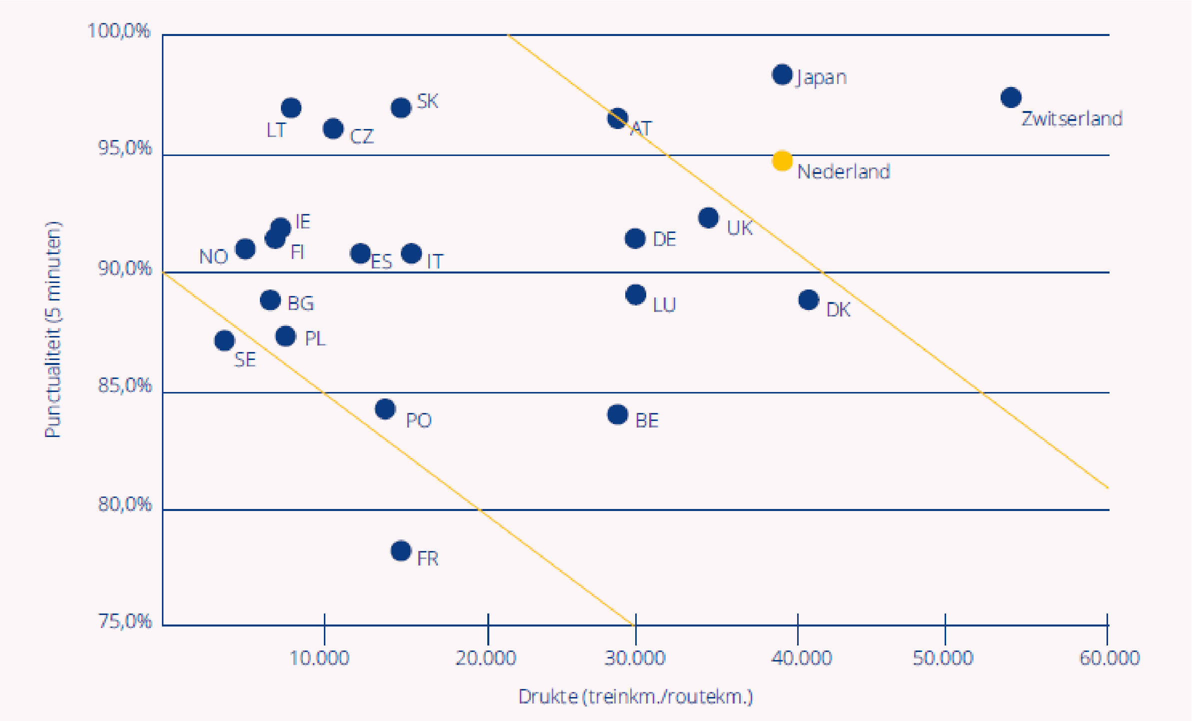


Figure 2.3: Punctuality versus crowdedness in different countries [5]

Table 2.1: The Key Performance Indicators of ProRail in 2014, with the threshold value and the realisation [8]

Key Performance Indicator	**Threshold**	**Realisation**
General opinion of the transports (both freight and passengers)	6.7	6.7
Overall passenger train punctuality (< 3 minutes)	87%	90.2%
Train punctuality main rail network (< 5 minutes)	93%	94.9%
Train punctuality regional lines (< 3 minutes)	93.1%	94.9%
Freight train punctuality (< 3 minutes)	82%	83%
Top 5 focus series	84%	87.3%
Delivered train paths	98%	97.9%

network and some regional train lines while the other regional lines are served by companies like Arriva, Connexxion, Veolia and Synthus. At this moment the NS is also only transporting passengers. The cargo transport is mainly operated by Railion and some smaller companies.

2.2. The origin of the passenger punctuality KPI

In the Netherlands the final responsibility of the railway system lies with the Dutch government or, more specific, the Ministry of Infrastructure and the Environment. This ministry determines the rail operator for passengers on the main rail network and the conditions it needs to meet. It also determines the company responsible for the construction, control and maintenance of the entire rail network (so not only the main rail network but the regional train lines as well) and the conditions this company needs to meet [7].

Up to 2025 the transport concession is assigned to the NS. This means that the NS is the only rail operator on the main rail network. The management concession is assigned to ProRail. These concessions are associated with conditions. If these conditions are not met every year, the NS or ProRail is ordered to pay a fine [8]. These conditions are formulated as a Key Performance Indicator (KPI) threshold. Such a KPI is a value representing a part of the performance. If a KPI threshold is not met in a certain year, the associated condition is not satisfied.

Table 2.1 and Table 2.2 shows the KPIs of 2014 for ProRail and the NS respectively, with the

Table 2.2: The Key Performance Indicators of the NS in 2014, with the threshold value and the realisation [8]

Key Performance Indicator	**Threshold**	**Realisation**
Passenger opinion on on-time-running	53.0%	49.9%
Train (arrival) punctuality (< 5 minute threshold)	93.0%	94.9%
Passenger punctuality	90.5%	92.3%
Passenger opinion on travel information by 0 to 15 minute delay	91.0%	94.4%
Passenger opinion on travel information by more than 15 minute delay	36.0%	39.0%
Information in train by a disruption	60.0%	70.8%
Information on the station by a disruption	80.0%	86.8%
Chance of meeting a conductor	65.0%	64.3%
Passenger opinion on cleanness train and station	55.0%	56.9%
Quality of cleanness train and station	90.0%	90.6%
Passenger opinion on social safety train and station	78.5%	80.2%
Passenger opinion on seating capacity in the train during rush hour	70.0%	67.8%
Transport capacity of passengers during rush hour	99.0%	98.9%

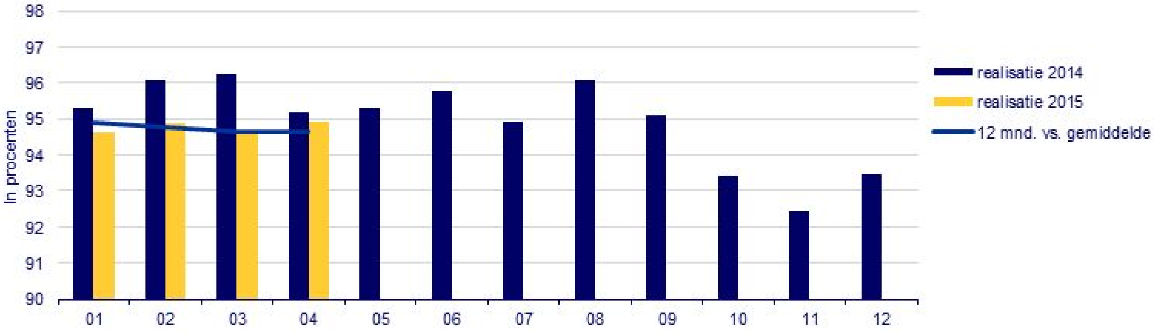


Figure 2.4: 5 minute train punctuality of the main rail network [10]

threshold and the realised values. In 2014 there were 13 KPIs to evaluate the performance of the NS. Some of those are related to hygiene at the stations and in the train, another is related to the number of available seats in a train. Another is related to the punctuality of the trains. Figure 2.4 shows the punctuality number of arrived trains for 2014 and the first four months of 2015. Here the train punctuality is calculated as follows: the percentage trains arriving at one of 35 measurement stations whereby the difference between the realised arrival time and the planned arrival time is 4:59 minutes or less [9]. The planned arrival time is the time as determined in the daily plan. The threshold value of 4:59 minutes is an international norm where trains with a delay less than 5 minutes are considered 'on time'. In 2014 the realised train punctuality was 94.9%. The threshold value of this KPI was 93.0% so this condition was satisfied [8].

Although the train punctuality is a good measuring tool to see whether and how much trains are delayed it has some drawbacks, certainly for the passengers [11][12]. First of all, trains that did not run are not taken into account. This seems logical: trains that do not run cannot be delayed so these trains should not be used in the calculation. However the passengers who wanted to take the cancelled train are delayed. They need to wait for the next connection.

A second disadvantage is that the number of passengers is neglected. A fully loaded delayed train during rush hour should have a larger impact on the performance KPI than an almost empty train late in the evening. However both trains are equally weighted in the train punctuality calculation.

Thirdly, missed transfers are not included. If a train has a delay of 4:59 minutes it is considered on time. Nonetheless, passengers in this train have 5 minutes less to make a transfer. Some tight transfers could not be made and thus some passengers experience a delay. In theory it could occur that every train runs with a delay less than 5 minutes. This implies a train punctuality of 100%, while

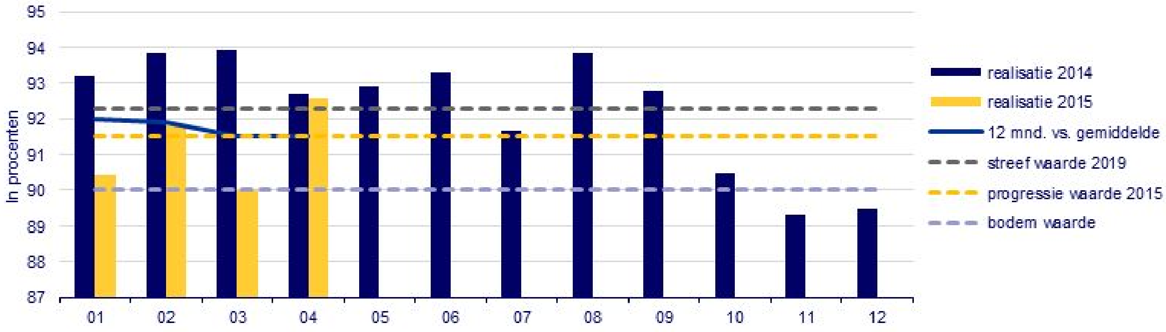


Figure 2.5: 5 minute passenger punctuality of the main rail network [10]

in reality all the passengers have a delay of more than 5 minutes. Note that this kind of reasoning can also be made for the exact opposite. If the transfer times are large, passengers arriving with a delay could still make the transfer. In theory the train punctuality could be low, but the number of passengers without a delay is high.

A last drawback of the train punctuality KPI is the number of stations where delays are measured. At this moment the NS measures at 35 stations, while it serves more than 350 stations. Trains could have a delay midway the journey but it is caught up when arriving at a measuring station. In the calculation of the train punctuality this train does not have a delay. Nonetheless every passenger with a final destination midway the journey does experience a delay. Here an opposite reasoning can be made as well. If a train runs on time but gets a delay just before its final destination every passenger departed before did not have a delay. The train itself though is registered as delayed.

In conclusion, from a passengers' perspective train punctuality is not the best KPI to measure the performance of the NS. Four drawbacks can be distinguished.

1. Trains that did not run are not taken into account
2. The number of passengers is neglected
3. Transfers are not included
4. Delays are measured at only 35 stations

Since 2011 the NS added a new KPI to evaluate their performance: passenger punctuality. In the Transport Plan of 2014 of the NS they defined the passenger punctuality KPI as follows [9]: 'The passenger punctuality indicator shows the percentage of passengers of which their train journey was successful.' In this case a successful journey is that the train was not cancelled, has a delay less than 5 minutes and the planned transfers are made. A more expanded definition is given in paragraph 2.4. Therefore the passenger punctuality deals with the most drawbacks of the train punctuality KPI. Figure 2.5 shows the passenger punctuality for 2014 and the first quarter of 2015. The mean passenger punctuality for 2014 was 92.3% while the threshold value was 90.5% [8]. So for 92.3% of the passenger their journey was 'successful'.

In February 2013 the Ministry of Infrastructure and the Environment published a report called 'Lange Termijn Spooragenda' (Long Term Rail agenda). Here it stated the vision and goals of the Dutch government for the rail network for 2025 and further [13]. The main goal was formulated as follows: 'Improve the quality of the rail network as a transport product in order for the passenger and the cargo shipper to see and use the train as an increasingly attractive transport option.' This main goal can be divided into three separate parts with each a set of sub goals: the attractiveness to the passenger, the quality of the rail network and the capacity of the rail network, see Table 2.3.

The NS and ProRail elaborated on this goal with their ambition to improve the performance of the rail network in Netherlands. In December 2013 they released a combined report: 'Beter en meer' (Better and more) [14]. Their ambitions include the improvement of reliability of the rail network, to increase the frequency of train and to offer, as much as possible, direct routes between origin and destination. To realise these ambitions ProRail and the NS concluded that the mutual collaboration should improve.

One solution to stimulate the cooperation between the NS and ProRail is to evaluate both companies

Table 2.3: The sub goals of the 'Lange Termijn Spooragenda' in view of passenger transport [13]

Attractiveness	**Quality**	**Capacity**
• Improve 'door-to-door' travel time • Improve travel comfort • Control of your own trip	• Improve safety • Improve reliability • Ensuring top position as durable means of transport	• Provide space for growth

based on the same KPIs with the same threshold values, if possible. Therefore ProRail adopted the passenger punctuality KPI. From 2015 onwards ProRail is also evaluated by this KPI [15]. The applied definition is the same as the definition of the NS. This means that the passenger punctuality is only calculated for the main rail network and not for the regional lines, despite the fact that ProRail is also responsible for these lines.

2.3. Current state of research concerning passenger punctuality

Since 2014, the motto of the NS is: the traveller on 1, 2 and 3 [5]. In other words, the interest of the passengers is the top priority. The passenger punctuality KPI is a good response to this view. Therefore in the annual report of 2014 the CFO of the NS, Engelhardt Robbe, stated that for the new concession for 2015-2024 the main focus lies on passenger punctuality instead of train punctuality. Remarkable however is the fact that the current knowledge of this KPI is lagging behind.

Passenger punctuality is not a new research subject. Simple passenger delay and punctuality models are in use since the 1990s in Denmark when the 0th generation model was introduced [16]. Commissioned by the Danish State Railways (DSB), Seest et al performed a research to determine the passenger punctuality of the subway system of Copenhagen [11]. The result was a model that can run during night time to calculate the passenger punctuality of the previous day. Route choice of passengers was based on a discrete choice model, where the route with the lowest cost was the preferred route. The model assumed the route choice was made at the origin of the trip when the possible delays of trains were known. However no data of passenger flows was used. That is why another Danish research concluded that at this moment passenger punctuality can only be derived theoretically by passenger delay models [17].

Based on the research of Seest et al, Nielson et al created the 3rd generation model in 2006 [18][12]. Here the origin, destination and desired arrival time are used to calculate the on time performance of a rail network. Their study of the Danish rail network shows that high train punctuality is not necessary equal to high passenger punctuality. However in 2007 the Danish Alex Landex concluded that a focus on passenger punctuality ensures that passengers are likely to be better off than with a focus on train punctuality [16]. Altogether there is some knowledge of passenger punctuality in Denmark but none of the reports investigated the causes of bad passenger punctuality or the indicators of influence to the passenger punctuality.

Aside from the metro of Copenhagen, the Swiss Federal Railways (SBB) also measures passenger punctuality. The passenger punctuality target is anchored as one of the top targets [19]. Here the passenger punctuality is defined as the percentage of passengers arriving within 3 minutes at their final destination [20]. Since 2008 this number is recorded [21]. In 2014 the passenger punctuality was 87.7% (see Figure 2.6). Just as in Denmark, no exact data of passenger flows between origins and destinations is used. The number of passengers is the average of traffic flow calculated using modelling data. The passenger punctuality of the SBB is calculated as follows. When a train arrives at a station the passengers are divided into three groups. The first group are the disembarking passengers. They are at their final destination in the network. The customer weighted arrival punctuality KPI (KAP) takes these passengers into account. The second group are the connecting passengers. They are taken into account for the customer weighted connection punctuality KPI (KCOR). The last group, the remaining

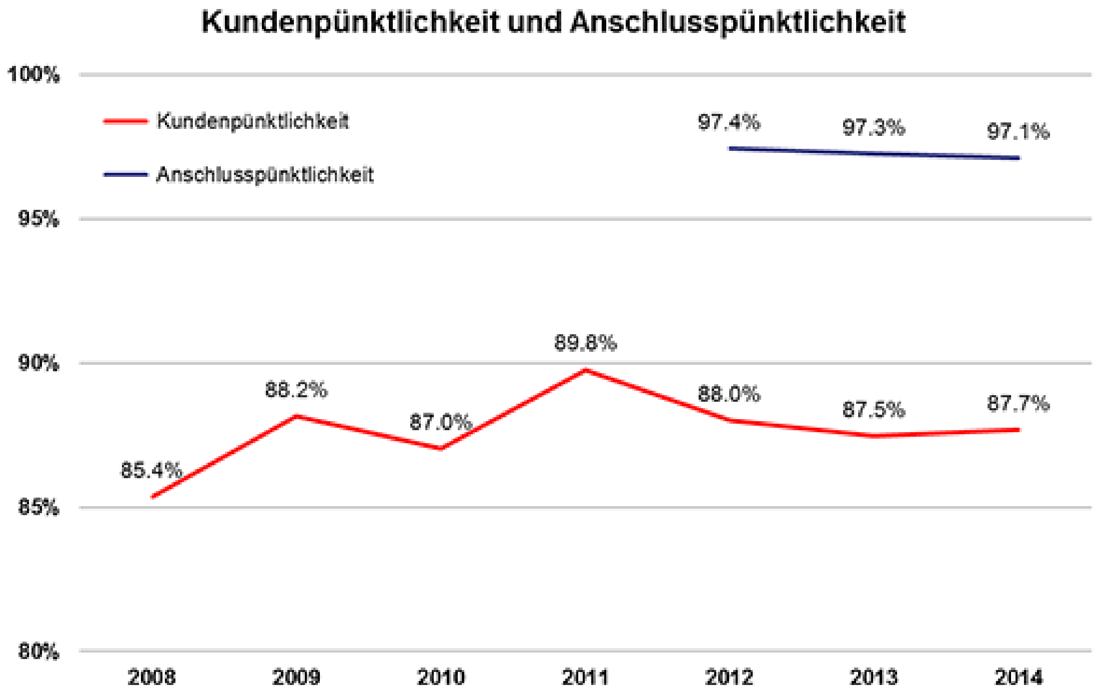


Figure 2.6: Passenger punctuality of the SBB [20]

passengers, do not depart and thus they stay in the train. The final passenger punctuality is a derived from the KAP and the KCOR.

The passenger punctuality in the Swiss is measured at 53 stations. Via this method approximately 90% of all trains are measured. Remarkable is the fact that the passenger punctuality is only measuring during working days. Saturdays and Sundays are neglected. Also at border points only the arrivals from within Switzerland are measured but the connections to and from trains outside Switzerland are counted as well [19].

The introduction of the passenger punctuality KPI in the Netherlands was motive for Dollevoet to do research towards delay management [22]. He introduced a set of models to minimize the overall passenger delays. These models were an expansion of the delay management model of Schöbel [23]. This model determines if a consecutive train needs to wait for a delayed predecessor so that the overall passenger delay is as small as possible. One of the models of Dollevoet extended the original model with the option of a different route to the final destination for the delayed passengers who needs to transfer between trains. A second model is also based on wait-depart decisions (so the decision if a train needs to wait for its delayed predecessor), but here the capacities of stations are taken into account. Solutions of this model can involve rerouting trains to a different platform than planned, if the overall passenger delays minimise. Note that for both models the number of passengers for every involved origin-destination combination should be known.

2.4. The Dutch methods of calculating passenger punctuality

At this moment the NS has two methods of calculating the passenger punctuality. In the first method, described in paragraph 2.4.1, the calculation is based on forecasts concerning the number of passengers and the number of passenger transferring. The second method, 'Passenger punctuality 2.0', is based on the OVCP data, derived from the OV-chipkaart, the smart card for the Dutch public transport [24]. This method is elaborated shortly in paragraph 2.4.2. A more extensive analysis of this second method can be found in chapter 3. Finally, in paragraph 2.4.3, the two methods of calculation are compared. Note that both methods are developed by the NS for the main rail network, thus regional lines are neglected. Although ProRail manages all the rail lines in the Netherlands, both the network used by the NS as by the other regional rail operators, the value of the passenger punctuality KPI of ProRail is the same as the value of the NS.

2.4.1. Method 1: Based on train punctuality and passenger forecast

The current definition of the passenger punctuality ($PP_{current}$) as used by the NS and ProRail is as follows [25]:

$$PP_{current} = \frac{A_{actual} + T_{actual}}{A_{scheduled} + T_{scheduled}} \tag{2.1}$$

Here, A_{actual} and $A_{scheduled}$ are the number of actual and scheduled passenger arrivals within a certain time interval. T_{actual} and $T_{scheduled}$ are the number of passengers transferring between two trains.

The total number of original scheduled passenger arrivals ($A_{scheduled}$) is the number of passengers arriving with the original scheduled trains, measured on 35 stations. The number of passengers is not known exactly, so these numbers are based on forecasts. So the number of passengers arriving is the number of planned train arrivals multiplied with the expected number of passengers per train. If there is not a passenger forecast for a train, a default value of 150 passengers is used.

For every planned train activity, a determination takes place to see if this train did run and if the realised arrival time was within 5 minutes of the planned arrival. If so, the train is considered to be arriving on time. All the non-delayed trains are multiplied with the passenger forecast. If a passenger forecast is not available, here also a default value of 150 passengers is used. Together this is the amount of passenger arrivals within 5 minutes of the scheduled plan (A_{actual}).

The total number of passengers transferring to performed connections ($T_{scheduled}$) is the number of passengers that will use a planned and realised connection.

- **Planned connection:** A connection is planned if the planned arrival time of the predecessor is 2 to 7 minutes earlier than the planned departure time of the connecting train. Also the daily expected number of passengers transferring should be 300 or more [26].
- **Realised connection:** A connection is realised if both the arriving train and the departing train did run. It could occur that a train activity is not realised with its planned train number but with a substitute number. In this case the substitute number is used because for the passengers it does not matter with which train number a train runs.

The exact number of passenger transferring is not known. Therefore forecasts are used. These forecasts are based on the time of day (working days between 7:00 and 9:00, working days between 16:00 and 18:00, the remaining hours of working days or weekend days) and the train series. The train number is thus not of influence. The combination of time of day and train series will give a percentage: the transfer percentage. This percentage is multiplied with the number of passengers arriving to give the estimation of the number of passengers transferring. To determine the time of day, the planned arrival time is used. If a transfer percentage is not known, a default value of 1 passenger transferring is used.

Per connection the NS has made a norm. This norm is the minimum time needed to transfers between two trains. If the time period between the realised arrival time and the realised departure time is at least the norm, the transfer is considered successful. The amount of passengers able to catch the performed connection (T_{actual}) is. Here also applies, if a substitute train number is used, this number is taken into account. If there is no forecast about the number of passengers transferring, here a default value of 1 passenger transferring is used as well.

Some remarks to this method of calculating the passenger punctuality can be made. First, the planned timetable is the timetable as known 48 hours before the day measured. This means that if the planned timetable is changed at the same day (because of unexpected disruptions) of the day before (because of extreme weather) the passenger punctuality KPI is affected and the performance will drop.

Second, the passenger punctuality is based on forecasts regarding the number of passengers and the number of passengers transferring. That implies that the reliability of this KPI is based on the reliability of the forecasts. This also implies that the outcome of this method is, in this regard, not more reliable, or precise, than the methods used in the metro of Copenhagen and in the Swiss. There the exact passenger flows are also unknown.

Thirdly, the KPI is not only the percentage of passengers arriving on time at their final destination. It differs because successful transfers are explicitly included. Passengers transferring are measure at least twice. Also the number of passengers transferring is a percentage of the total number of passengers in a train. This fact has some influence on the KPI and the result is a slightly incorrect passenger punctuality value.

A fourth remark to the method of calculation currently in use by the NS is the fact that route choice is ignored. For example, if travelling from station Delft to station Utrecht Centraal, there are two possibilities: via stations Den Haag or Rotterdam. The difference in travel time (and extra waiting time at Delft) lies within the threshold of 5 minutes. When there are no disruptions, a part of the travellers will travel via Den Haag and the other part will travel via Rotterdam. This passenger behaviour is included in the forecast for passengers transferring at stations Den Haag and Rotterdam. However when there is a disruption at Den Haag, almost all passengers will travel via Rotterdam. All these passengers will travel without a delay and are able to make their transfer. Though in the passenger punctuality calculation a part of the passengers will still travel via Den Haag and thus experiences delays and missed transfers.

Lastly, this method of calculating the passenger punctuality KPI deals with the most drawbacks of the train punctuality KPI (see paragraph 2.2). The impact for the passengers of trains that did not run are included in the KPI, because A_{actual} is the number of passengers that actually made a trip. The KPI will go down if this number becomes smaller because trains did not run. Although the number of passengers is a forecast, it is taken into account in the calculation. The same applies for the passengers transferring. If a transfer time is below the norm, transfers are not made and T_{actual} will drop. The only drawback still not handled is the number of measuring stations. The passenger punctuality KPI is measured only at the same 35 stations used for the train punctuality calculation. This means that the KPI is not as accurate as it could be. If a journey leads trough two or more measuring stations, the passenger is calculated at least twice in the passenger punctuality. On the other hand if a trip does not pass a measuring station, that trip is not taken into account in the calculation. It implies that a trip towards a measuring station outweigh a trip departing at a measuring station.

2.4.2. Method 2: Based on OVCP data

At this moment, the NS is developing a new method of calculating the passenger punctuality: 'passenger punctuality 2.0' [26]. The goal was to implement this method on time so in 2016 the KPI could be calculated via this method. In July 2015 the decision was made to postpone the introduction with at least one year. This new method is this in developing, so some details can change. The method currently is use has some drawbacks (see also paragraph 2.4.1) including:

- The number of passengers and the number of passengers transferring are forecasts.
- The passenger punctuality is measured at only 35 stations.
- Only planned transfers are taken into account
- The ratio between InterCitys (long distance trains) and Sprinters (local trains) is disproportional: InterCitys stop relatively more often at measuring stations than Sprinters.
- The ratio between the morning and evening rush hour is disproportional: relatively more passengers arrive at a measuring station in the morning rush hour than in the evening rush hour.

Passenger punctuality 2.0 is designed to deal with these disadvantages. The new KPI can be calculated as follows:

$$PP_{2.0} = \frac{T_{promised}}{T_{total}} \tag{2.2}$$

In this equation T_{total} is the total number of trips. This number is determined via the OCVP data [24]. This data contains the origin and destination of every passenger so every trip made can be determined. It is amassed via the OV-chipkaart, the smart card of the public transport system of the Netherlands. At the start of a trip, the passenger has to check-in by holding the smart card against a dedicated card reader. The location and time are registered. At the destination the passenger needs to do this again. Here also the location and time are registered. With this system the number of trips per origin-destination combination is known, as well the start and end time of each trip.

With more than 400 stations in the Netherlands, the number of origin-destination (OD) combinations is more than 160.000. Not all stations are always in use and most of the OD combinations are only in rare cases travelled on. Therefore a couple of conditions are drawn up to decrease the number of possible OD-combinations in the calculation of the passenger punctuality. First stations serviced only by other rail operators are neglected. OD combinations barely used are ignored [24] as well. Every quarter of a year a list is made with the most frequent used OD combinations. Here all combinations are included with at least 100 trips in the last 100 days, spread out over at least 20 days. This prevents also that event stations, like Rotterdam Stadion, are taken into account. These stations are only serviced during events and therefore not included in the general timetable. Altogether this will result in approximately twenty thousand OD combinations used in the calculation. This implies that around 87% of all OD combinations are not taken into account though still more than 98% of all trips in the network are used in the calculation, see also paragraph 3.2.1.

$T_{promised}$ is the number of trips where the total travel time is within the promised travel time plus a certain delay threshold. The promised travel time is the travel time as shown in the timetable before unplanned disruptions are known. When there is an adapted timetable due to extreme weather the original timetable is used for determining the planned arrival time as well. Although the delay threshold could be every value, at this moment 3 minutes is used in the official KPI. A threshold of 15 minutes could also be interesting [27], see also paragraph 4.2.

The calculation of $T_{promised}$ is as follows [24]. The OVCP data is used to determine the origin and destination and the check-in time of every trip. This check-in time is used to determine the promised journey from the origin to the destination. This promised journey (in Dutch: Reisbelofte) is the connection with the earliest possible arrival time with a departure after check-in time. Next the realisation data is examined. If a transfer was not made, or a train did not run or not arrive at the transfers station or destination, the trip needs to be rescheduled. Otherwise the realised arrival time is used to determine the final delay. If a trip needs be rescheduled, the checkout time is used to determine the final delay. A more detailed explanation of the calculation is given in chapter 3.

Note that the passenger punctuality 2.0 deals with every drawback of the train punctuality, as described in paragraph 2.2. Every trip is compared with the timetable, so trains that did not run are taken into account. In large extend the number of passengers are also known. The same applies for the number of passengers transferring. Finally, the delays are measured at every station for almost every origin-destination combination and not only at the 35 measuring stations. This implies that this passenger punctuality value gives a more realistic representation of the reality.

2.4.3. A comparison between the two methods

In this paragraph a comparison between the two Dutch methods of calculating passenger punctuality is made. Method 1 is the method currently in use, where the passenger punctuality is a combination of the train punctuality at 35 stations and the forecast of the number of passengers. Method 2, passenger punctuality 2.0, is based on the OVCP data. This method is still in development so details can change before the introduction. In an example the influence of the differences becomes clear. Here the passenger punctuality is calculated via both methods for different situations.

1. The weight distribution among the trips in the network is different. In method 1, the weight of a trip is dependent on the number of measuring stations passed and the number of transfers at a measuring station. In the worst case, this can imply that a trip is not taken into account at all, most common by the shorter journeys. On the other hand, the impact of longer journeys is severe. In method 2 every trip is weighted equally.

2. The transport plan of 2014 of the NS stated that passenger punctuality is the percentage of successful journeys, thus where all the trains run, all transfers are made and the final arrival time is within 5 minutes of the planned arrival [9]. Method 1 calculates the passenger punctuality slightly different: transfers are explicitly included, see situation 8 of the example case on page 19. In this example every passenger arrived with a delay. This implies that all the journeys were not successful. Nonetheless the passenger punctuality was not 0% because the transfers were made. In method 2 this is not possible. A journey is successful or not, but not partly because a transfer was made.

3. Method 1 uses forecasts of the number of passengers. These forecasts are based on passenger counting of conductors. Per train number and arrival a forecast is made. If such a forecast if not available, a default value of 150 passenger is used. Method 2 uses the exact number of passengers per train, retrieved from the OVCP data.

4. In method 1, the number of passenger transferring is a forecast. A percentage of all passengers arriving will transfer. Per train series and arrival stations there is a forecast of four percentages,

based on the arrival time [25]:

- Working days between 7:00 and 9:00 (morning rush hour)
- Working days between 16:00 and 18:00 (evening rush hour)
- Working days in the remaining time periods
- Weekend days

The forecast of the number of passengers is multiplied with the forecast of percentages of passengers transferring. The result is the number of transfers. Method 2 determines the travel option between the origin and destination with the earliest arrival time based on the departure time. The outcome is, among others, the probable location where a transfer took place.

5. Method 1 has strict rules about which transfers are included. Only trains with a planned departure 2 to 7 minutes later than the planned arrival are taken into account. Also the daily number of passengers transferring should exceed 300. Method 2 uses every official transfer irrespectively the transfer time and total number of passengers transferring. A transfer is official when it is shown in the timetable. Unofficial transfers are connections where the transfer time is less than the official walking time between the platforms. These transfers are not show in the timetable but some passengers are able to make the transfer nevertheless.

6. Method 1 uses forecasts of the number of passengers and transfers on only 35 stations to calculate the passenger punctuality. Method 2 uses the OVCP data of all passengers and calculates for every trip the expected journey. Method 1 can therefore be considered as a macroscopic model while method 2 is a microscopic model. This results in a longer calculation time and higher costs for method 2.

Example: The difference between train punctuality and the two methods of the passenger punctuality

Considering the following part of the main rail network of the Netherlands (see also Figure 2.7). The intercity between stations Rotterdam Blaak (RTB) and Den Haag Holland Spoor (GV) contains at departure 250 passengers. 100 passengers will make a transfer at station Rotterdam Centraal (RTD) to the Sprinter to station Woerden (WD) via station Gouda (GD). The other 150 passengers have GV as destination.

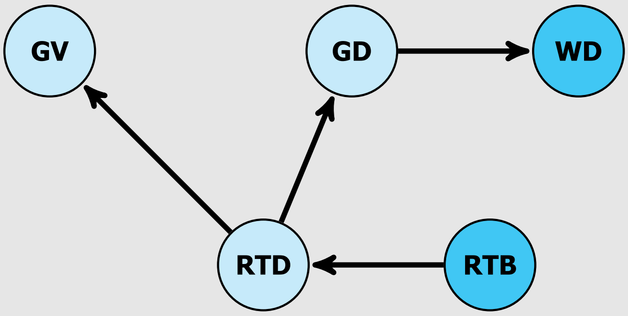


Figure 2.7: Part of the main rail network

For eight different situations the train punctuality and the passenger punctuality is calculated. The passenger punctuality is calculated via method 1 as well as method 2. RTD, GV and GD are measuring stations, so only these stations are used in the calculation of the train punctuality and the passenger punctuality via method 1. Note that, except for situation 1, the passenger punctuality of both methods is never the same. Situation 2 and 5 give some remarkable results: The train punctuality and the passenger punctuality via method 1 are not optimal, while according to method 2 the passenger punctuality is 100%. On the other hand, in situation 7 the opposite occurs. Here the passenger punctuality according to method 2 is not optimal, but it is 100% for the other two calculations.

Situation 1: No delays and no cancelation In the first situation there are no delays in the network and none of the trains is cancelled.

Train punctuality	There are 3 planned arrivals (RTD, GV and GD). Every train arrival is on time, so the train punctuality is $\frac{3}{3}$	**100%**
Method 1	There are 500 planned passenger arrivals (250 in RTD, 150 in GV and 100 in GD) and 100 planned transfers. Every train ran on time, so every arrival was on time and every transfer is made: $PP = \frac{500+100}{500+100}$	**100%**
Method 2	The total number of passengers is 250 and for every passenger the journey could be made as promised: $PP = \frac{250}{250}$	**100%**

Situation 2: The intercity has a delay of 5 minutes in RTD The intercity to GV has a delay of 5 minutes when arriving in RTD. The transfer to the sprinter can be made. In the onward journey to GV the intercity is able to decrease the delay to 2 minutes.

Train punctuality	The intercity does not arrive on time in RTD, so $TP = \frac{2}{3}$	**66.7%**
Method 1	There are 500 planned passenger arrivals. Of those arrivals, 250 arrivals were not on time (in RTD). Every transfer is made so: $PP = \frac{250+100}{500+100}$	**58.3%**
Method 2	The total number of passengers is 250 and for every passenger the journey could be made as promised: $PP = \frac{250}{250}$	**100%**

Situation 3: The intercity has a delay of 10 minutes in RTD The intercity to GV has a delay of 10 minutes when arriving in RTD. The transfer to the sprinter can therefore not be made. In the onward journey to GV the intercity is not able to decrease the delay.

Train punctuality	The intercity does not arrive on time in RTD and GV but the sprinter does arrive as planned in GD, so $TP = \frac{1}{3}$	**33.3%**
Method 1	There are 500 planned passenger arrivals. Of those arrivals, $250 + 150 = 400$ arrivals were not on time, because those passengers travelled with the intercity. This method works with forecasts and although the transfer was not made, the 100 arrivals in GD are thus considered on time: $PP = \frac{100+0}{500+100}$	**16.7%**
Method 2	None of the passengers arrived on time at their destination.	**0%**

Note that the outcome of situation is almost the same as when the intercity is cancelled. The only difference is the outcome of the train punctuality. Here, only trains that did run are taken into account, so $TP = \frac{1}{1} =$ **100%**.

Situation 4: The intercity is cancelled at RTD A disturbance on the route RTD – GV causes that the intercity is cancelled at RTD.

Train punctuality	Cancelled trains are not taken into account, so $TP = \frac{2}{2}$	**100%**
Method 1	Of the 500 planned arrivals, only the arrivals at GV were not realised. $PP = \frac{350+100}{500+100}$	**75%**
Method 2	The passengers with destination GV did not arrive as promised so $PP = \frac{100}{250}$	**40%**

Situation 5: The sprinter has a delay of 5 minutes in GD The sprinter to WD has a delay of 5 minutes when arriving in GD. In the onward journey to WD the sprinter is able to decrease the delay to 2 minutes.

Train punctuality	The sprinter does not arrive on time in GD, so $TP = \frac{2}{3}$	**66.7%**
Method 1	There are 500 planned passenger arrivals. Of those arrivals, 100 arrivals were not on time (in GD). Every transfer is made so: $PP = \frac{400+100}{500+100}$	**83.3%**
Method 2	The total number of passengers is 250 and for every passenger the journey could be made as promised: $PP = \frac{250}{250}$	**100%**

Situation 6: The sprinter has a delay of 10 minutes in GD The sprinter to WD has a delay of 10 minutes when arriving in GD. In the onward journey to WD the sprinter is not able to decrease the delay.

Train punctuality	The sprinter does not arrive on time in GD, so $TP = \frac{2}{3}$	**66.7%**
Method 1	There are 500 planned passenger arrivals. Of those arrivals, 100 where not on time, because those passengers travelled with the sprinter: $PP = \frac{400+100}{500+100}$	**83.3%**
Method 2	The passengers with destination WD did not arrive on time: $PP = \frac{150}{250}$	**60%**

Note that if the sprinter did not run at all, only the train punctuality would differ: $TP = \frac{1}{1} =$ **100%**.

Situation 7: The sprinter is cancelled at GD A disturbance on the route GD – WD causes that the sprinter is cancelled at GD.

Train punctuality	Cancelled trains are not taken into account, but WD is not a measuring station so $TP = \frac{3}{3}$	**100%**
Method 1	Of the 500 planned arrivals, every arrival is realised because the passenger with destination WD arrived on time in GD. $PP = \frac{500+100}{500+100}$	**100%**
Method 2	The passengers with destination GW did not arrive on time as promised, so $PP = \frac{150}{250}$	**60%**

Situation 8: Both the intercity and the sprinter are delayed When arriving in RTD the intercity has a delay of 10 minutes. A technical problem causes the sprinter to start its journey with a delay of also 10 minutes. Both trains are not able to decrease their delay further on their route.

Train punctuality	Both trains are delayed, so $TP = \frac{0}{3}$	**0%**
Method 1	Of the 500 planned arrivals, none of those is on time. However the transfer is made. $PP = \frac{0+100}{500+100}$	**16.7%**
Method 2	None of the passengers arrived on time as promised, so $PP = \frac{0}{250}$	**0%**

2.5. Conclusion

Both the NS, as the rail operator on the main rail network, as ProRail, as the rail network manager, are under concession of the Ministry of Infrastructure and the Environment. To measure their performance, for both companies a list of Key Performance Indicators (KPIs) is created. Each KPI gives the performance of a part of the operations of the NS or ProRail. If a value of a KPI is lower than a certain threshold, the performance is not as good as it should be and ProRail or the NS have to pay a fine. One of the KPIs is passenger punctuality: the percentage of passengers with a successful journey. Initially this KPI belongs only to the NS, but in 2015 ProRail adopted this KPI as well.

In Swiss and by the metro of Copenhagen the passenger punctuality is calculated as well. Although the exact calculation methods differ between the three countries, the basic idea is the same: the realisation data is combined with a forecast of the number of passengers. Together this leads to the passenger punctuality. Until a few years ago in none of the countries the actual number of passengers was known. Therefore more precise, microscopic, calculations were not possible.

With the introduction of the OV-chipkaart, data of almost every passenger in the Netherlands became available. With this data, a possible journey for every passenger can be determined. This implies the possibility of a more exact calculation of the passenger punctuality. Therefore the NS is developing a new method of calculating the passenger punctuality. Here the realised journeys of every passenger is compared with a promised journey, two days prior of the execution.

The NS made the passenger punctuality their most important KPI. It is therefore remarkable that at this moment it is unknown how precise this new method of calculation is. With only a short explanation of the method of calculation, it is already clear that some large assumptions are made. Also it is unknown how the KPI can be described, based on the operations of ProRail and the NS. Steering on the processes to improve the passenger punctuality is thus not possible.

3

Passenger punctuality 2.0

In paragraph 2.4.2 the new method of calculating passenger punctuality is briefly explained. Passenger punctuality 2.0 is the number of trips arriving with a delay less than a certain threshold, divided by all the trips in the network. Chapter 2 stated that some large assumptions are made in this calculation as well. Therefore this chapter will examine the method of calculation even further. Special attention goes to the assumptions made in this method.

This chapter will start with an explanation of the structure of the algorithm to calculate passenger punctuality, as developed at the NS. In paragraph 3.1 first the five modules that form the algorithm are explained. In the algorithm every trips gets categorised. These categories are also elaborated on. A total of 22 assumptions can be distinguished in the method of calculation. The assumptions are analysed in paragraph 3.2. If possible the impact of those assumptions is calculated as well. Some assumptions require an additional analysis. These analyses can be found in appendix A. Here an extensive overview of all assumptions and the calculated impact can be found as well (appendix A.6). The final concussions and some last remarks to the structure of the passenger punctuality calculation and the assumptions can be found in paragraph 3.3.

This chapter will provide a better insight in the method to calculate passenger punctuality. Moreover it will give an estimation of the impact of the assumptions in the method on the passenger punctuality KPI. This will place the outcome of the breakdown of the passenger punctuality (chapter 4) and the results of the models to describe passenger punctuality (chapter 5) in a more realistic perspective.

3.1. The structure of the passenger punctuality calculation

The NS has developed an algorithm to calculate the passenger punctuality 2.0 out of the raw data [24]. This algorithm can be divided into five modules. These modules can be found in Table 3.1. All modules have their place in the algorithm, though the third module it not developed by the NS. In this module trips are being rescheduled. There are multiple alternatives for rescheduling trips, see also paragraph 3.2.11. However the NS decided not to reschedule trips at all. Instead the checkout time is used to determine the realised arrival time and thus the final delay. Nonetheless the structure of the algorithm is built in such way, that if the third module is still developed the integration of this module is relatively easy. In paragraph 3.1.1 the five modules are elaborated.

Table 3.1: The five modules of the algorithm to calculate the passenger punctuality [24]

Module	**Function**
1	Retrieve and store travel options
2	Combining travel options with realisation data
3	*Reschedule: Determine new travel option*
4	Combining OVCP data with travel options
5	Generation data set

In the second module, every trip gets categorized. Based on the categorization, a trip is taken into

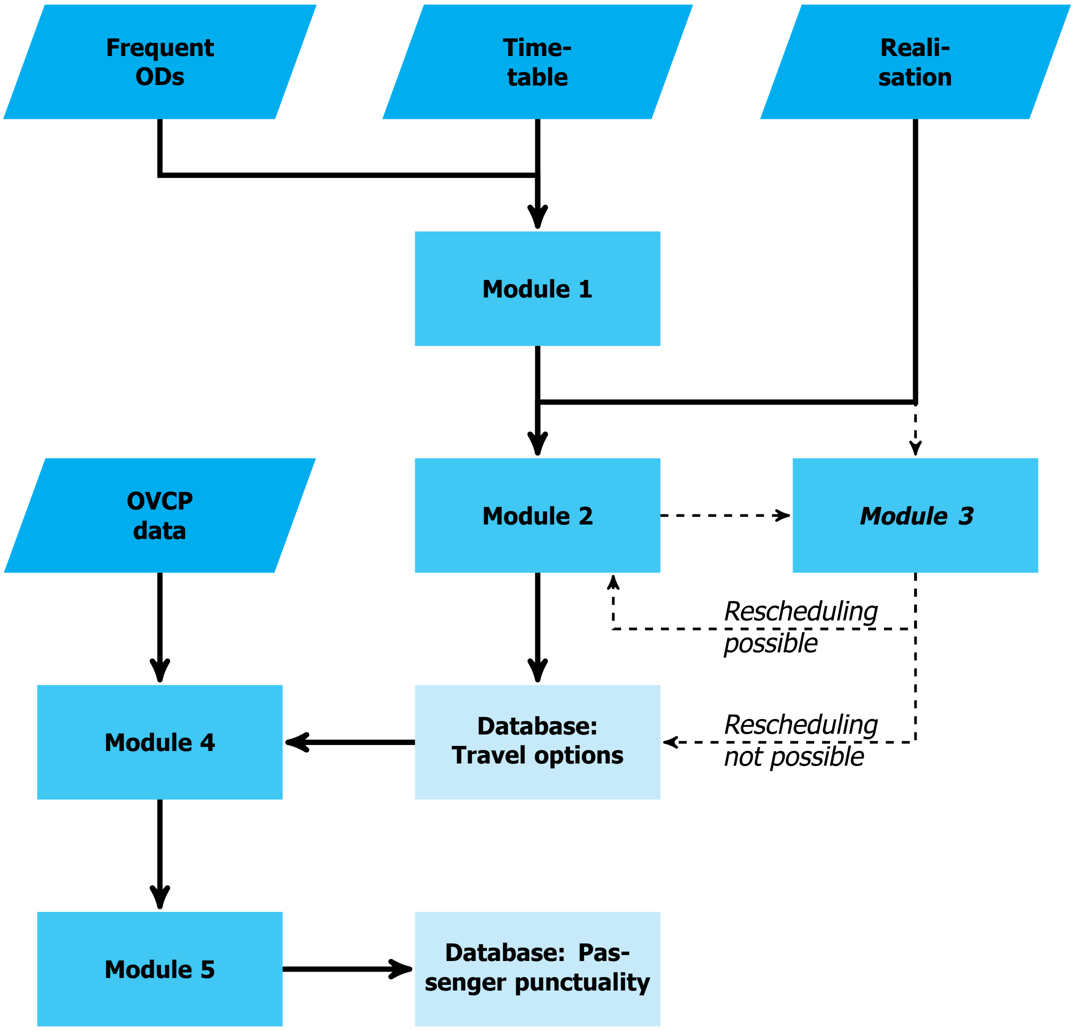


Figure 3.1: Simplified flow scheme of the algorithm to calculate the passenger punctuality 2.0 [24] (edited)

account in the calculation or not. It determines if a trips needs to be rescheduled as well. Eight different categories can be distinguished. In paragraph 3.1.2 the categories are analysed. The conditions for categorizing a trip, the size of each category and the effect of the category to the trip are given.

3.1.1. The five modules

A simplified flow scheme of the algorithm to calculate passenger punctuality is visualized in Figure 3.1. Here the relation between the input data, the modules and the output is displayed. As can be seen, module 3 is not essential to achieve the passenger punctuality. Nonetheless including this module in the calculation will make the outcome of the calculation more accurate (see paragraph 3.2.11). In this paragraph the four essential modules are explained, as well as the mutual relations and the relations between the modules and the input data.

Module 1: Retrieve and store travel options

The NS has built a journey planner. This planner determines the best travel options based on origin, destination and desired departure or arrival time. The input is the most recent timetable of the execution date because this timetable changes over time [28]. First there is the year plan. This is the least accurate timetable. Planned maintenance and last minute changes are not included. Three weeks before execution the timetable is updated with the week plan. Included in this plan is planned maintenance. From 16 hours for execution the timetable is constantly updated with last minute changes and delay predictions. All these last changes are due to unplanned maintenance and disturbances in the network. The passenger punctuality should include impacts of last minute deviations in the timetable but not the expected hindrance of planned maintenance. Therefore the travel options should be determined from the journey planner between three weeks and 16 hours before execution. In the documentation this is often simplified as the timetable 48 hours before execution.

Once a week all the travel options for at least two days later are derived from the journey planner [24]. The result is stored in a database: the, so called, historic journey planner. Note that this only applies for the travel options with an origin-destination (OD) combination frequently used. The list of frequent ODs is determined four times per year based on the OVCP data of the previous 100 days. An OD is labelled frequent if two conditions are met:

- An OD should appear on average once a day.
- An OD should appear on at least 20 different days

The result is a list of approximately twenty thousand different ODs. This list is updated every quarter of a year, so new stations can be included in the passenger punctuality and stations not in use anymore can be filtered out.

Module 2: Combining travel options with realisation data

After the execution date the realisation data is available (NVGB data, see paragraph 3.2.9. This data contains the realised arrival and departure times of the trains. For every travel option the module includes if a train was cancelled, the last reached station and if a train was replaced with bus transport as well. Finally this module categorizes every travel option. The exact categories are given in paragraph 3.1.2.

First a determination takes place to see if a travel option can be included in the passenger punctuality. A travel option cannot be included if it is (partly) serviced by another train operator or if a train was replaced with a bus. In both cases the NS does not have the realised departure and arrival times. This data is needed to determine if a travel option could be made as promised or if a travel option should be rescheduled. The second step is to determine for every remaining travel option if the travel option could be realised or not. A travel option is not realised if it is impossible to arrive at the final destination with the promised train. This could be due because a transfer could not be made or a promised train was cancelled. If so, the travel option should be rescheduled. The rescheduling is part of module 3. It is necessary to determine the realised arrival time. Nonetheless module 3 is not developed so no rescheduling is done. Instead the checkout time will be used to determine the final delay. This makes the whole process easier bus less accurate, see also paragraph 3.2.11.

The outcome is a database with every travel option combined with the realisation data. Every travel option is categorized as well. If the category implies that the module was not able to determine the realised arrival time of the travel option, this information is neglected.

Module 4: Combining OVCP data with travel options

The OVCP data contains, among others, the origin, destination and check-in and checkout time of every trip made. Based on the origin, destination and check-in time a travel option is found. The module searches per trip for all travel options with the same origin and destination. Each travel option must have a planned departure at least 1 minute after the check-in time and at most 1 hour later. The travel option arriving the earliest is the promised journey (in Dutch: Reisbelofte) for this trip. The threshold of 1 minute is the expected walking time between card reader and platform. Note that there are two reasons why a trip could not be combined with a travel option.

- The OD of the trip is not a frequent OD
- There is no travel option between 1 minute and 1 hour after check-in

At this moment the last category can be assigned. If the time between check in and realised departure is less than 1 minute, the category of the trip will be 'Train departs too early' (see paragraph 3.1.2). In this case it does not matter if the travel option had already a category. Note that this does not apply to trips not used in the passenger punctuality calculation, so trips without realisation data.

Based on the realised arrival time the final delay per trip can be calculated. This is the difference between the promised arrival time and the realised arrival time. No rescheduling is done, so for some trips there is no realised arrival time. Here the checkout time is used. The delay is the difference between the promised arrival time and the checkout time minus 1 minute. This 1 minute is the expected walking time between platform and card reader.

Module 5: Generation data set

This module takes the outcome of module 4 and stores it in a database. This database contains for every trip the promised journey and the realisation of this promise. This includes a categorization and the final delay. With this database the passenger punctuality can be calculated. That goes as follows. Take a time period. This can be for example a day or a year. Take also a delay threshold. If the goal is to calculate the 3 minute passenger punctuality this threshold is 180 seconds. Lastly count every trip within the chosen time period with a delay less than the chosen threshold and divide this with every trip within the chosen time period. The result is the passenger punctuality. Note that trips (partly) serviced by another train operator and trips with bus transport needs to be neglected. Note also that the analysed trips could also be filtered by, among others, origin-destination combination, train series or departure or arrival station. This will lead to a passenger punctuality value for a part of the network.

3.1.2. The trip categories as defined by the NS

As stated in paragraph 3.1.1 every travel option gets categorized. The category can determine why a trip is not used in the passenger punctuality or it determines why a trip needs to be rescheduled. For the latter the expected arrival time is based on the checkout time. The last category ('Trip realised') indicates that the trip can be realised as promised. The expected arrival time is the realised arrival time of the last train of the journey. This section examines every category and analyses which conditions must apply. Also a determination is made to see which percentage of the trips ends up with the analysed category.

Other train operator

It could be that a passenger start and end his or her trip at a station serviced by the NS. Therefore the passenger checks in and checks out by an OVCP card reader of the NS. However the found route associated with this trip involves another train operator. An example of such a trip could be between stations Amersfoort and Ede-Wageningen. This route is exploited by Valleilijn. However both stations are also serviced by the NS. Passengers are able to check in and out by the NS but travel with the Valleilijn. If this travel option is the fastest option as well, the trip gets the category: Other train operator.

The reason for this is that the NS does not have the realisation data of the other train operators. The realised departure and arrival times are therefore unknown. A correct determination whether a passenger arrives within a threshold value at the destination can thus not be made. Also the goal of the KPI is to give an indication of the performance of the NS. Using the trips made by another train operator is therefore not desirable. So trips with the component 'Other train operator' are not used in the passenger punctuality calculation.

For the first 25 days of June 2015 21.318.027 different check-in checkout combinations could be distinguished. Of those trips 99.61% (21.234.807) could be combined with a travel option. Of those trips with a promised journey, 80.555 (0.38%) were categorized as 'Other train operator.'

Bus transport

By work activities or long term disruptions bus transport is used instead of trains. This means that there is no realisation data available for these trips. Trips that encounter bus transport are categorized as 'Bus transport'. No realisation data is available, so rescheduling is not possible. Therefore this category is neglected in the passenger punctuality calculation. Passengers have an intent to check out before they get in the bus as well. Therefore the number of passengers with the category 'Bus transport' is not representative of the total number of passengers experience bus transport.

For the first 25 days of June 2015, 38.646 trips were categorized as 'Bus transport'. This is 0.18% of the total number of trips with a promised journey. 'Bus transport' and 'Other train operator' are the only two categories not used in the passenger punctuality. Together with the trips with no promised journey this means that 99.05% of all founded check-in checkout combinations are used in the calculation of the passenger punctuality.

Train departs too early

The earliest promised departure time is at least 1 minute (the walking time between card reader and platform) later than the check-in time. It could occur that a train leaves too early. The trip is categorized as 'Train departs too early' if the realised departure time is within 1 minute of the check-in time. The assumption is made that the passenger was not able to catch the train and thus the realised arrival time of the train is not the realised arrival time of the passenger. This implies a new journey needs to be planned. In this case the delay of the passenger is calculated via the checkout time. Here also a walking time of 1 minute is used. The delay of the passenger is the difference between the planned arrival time and the checkout time minus 1 minute.

Between June 1 and June 25 2015 21.338 trips were categorized as 'Train departs too early'. This is 0.10% of all trips used in the passenger punctuality calculation. Note that this category is only used when the first train of the promised journey left too early. If a transfer is missed because a consecutive train departs too early, the category 'Missed transfer' is used.

Train departs too late

It could occur that a promised train has a large delay. If this delay increases, the chance that passengers will take another train is also increasing. The timespan when a route will be rescheduled depends on the location in the network. On some routes the next train in the same direction departs only 30 minutes later while for other routes the maximum waiting time is 10 minutes or less. Therefore a default value of 15 minutes is used. If a train departs with a delay of 15 minutes or more the assumption is made that the passengers reschedule their trip. In this case the trip is categorized as 'Train departs too late'. The realised arrival time of the original planned train cannot be used anymore so the checkout time is used as a reference for the arrival time. The arrival time is the checkout time minus 1 minute.

For the first 25 days of June 2015, 155.655 trips were categorized as 'Train departs too late'. This is 0.73% off all the trips used in the calculation. Note that a train can depart too late at the origin of the trip but also at a transfer station.

Train did not depart

If a train did not depart at the origin or at a transfer station, the trip is categorized as 'Train did not depart'. Enclosed in this categorization is the fact that a train can run with a different train number as planned. Consider the following example. At the start of a route there is disturbance whereby the original train is not able to depart. On a station further on the route a substitute train is deployed. This train gets the train number of the original train plus 300000. For passengers waiting on the first station of the route, the train did not depart and their trips will be categorized as 'Train did not depart'. However this is not the case for passengers departing further on the route. The fact that the train departing is not the original planned train does not matter. They do not need to reschedule their trip. If the trip needs to be rescheduled, the checkout time is used to determine the final delay.

1.86% of all the trips of the first 25 days of June 2015 used in the calculation of the passenger punctuality has the category 'Train did not depart'. This is a total of 392.180 trips.

Train did not arrive

In some cases a train is not able to serve the second part of the route. This can be due to a disturbance further down the route or because the train itself has technical problems. Whatever the cause may be, passengers in this train are stranded on the last reached station. The planned trip needs to be rescheduled so this trip gets the category 'Train did not arrive'. The realised arrival time of every trip cannot be calculated based on the realisation data anymore so the checkout time is used. Note that the trips of the passengers waiting on the last reached station are getting the category 'Train did not depart'.

In the first 25 days of June 2015 129.646 trips were getting the category 'Train did not arrive'. This is 0.61% of all the used trips in the calculation of the passenger punctuality.

Missed transfer

A transfer is made when the difference between the realised arrival time of the first train and the realised departure time of the second train is at least equal to the norm time for transferring. This norm time is the walking time between the platforms. For the most stations and platforms, this is specified per unique combination of the station and the two platforms involved. When such a platform specific norm is not available, a station specific norm will be used. This is a norm time that differs only per station and not per platform combination. In the rare case such station norm is not available as well, a default value of 5 minutes is used. Note that the platforms used in the determination of the norm time are the planned platforms. The planned platforms are not necessarily the used platforms in the realisation. It is also possible that the arrival platform is equal to the departing platform. In such a case a check is made if the time between the realised arrival time and the departure time is at least 3 minutes. When the time difference is less, the conclusion is made that the planned platforms differs from the realisation. The minimum time between arrival and departure should be at least the specified station norm time for such a transfer to be made.

If the time difference between arrival and departure is less than the norm time, the transfer is not made. The trip is categorized as 'Missed transfer'. This is the case for 532.555 trips in the 25 first days of June 2015. This is 2.52% of all the used trips in the calculation. The transfer is missed, so the trip needs to be rescheduled. The realisation data is therefore not useful and the realised arrival time at the final destination is calculated based on the checkout time.

Trip realised

If the first train of a trip did not depart within 1 minute after check-in; and if all the trains of a trip did depart with a delay of less than 15 minutes; and if all the trains of a trip arrive at transfer stations or the destination; and if all the transfers were made; and if there was no bus transport during the trip and if there was no other train operator involved, than the trip was realised as promised. The arrival time can be derived from the realisation data. This does not mean that the departure and arrival times were as promised, only that the passenger was able to catch the trains as promised without rescheduling. In this case the trip gets the category 'Trip realised'.

By far the most trips get the category 'Trip realised'. In the first 25 days of June 2015 19.884.232 trips were 'realised'. This is 94.17% of all the trips. Note that this does not mean that the passenger punctuality is 94.17%. Trips with another category can still be on time and trips that are realised could still be delayed.

3.2. The assumptions and their impact on the KPI

In the method of calculating passenger punctuality 2.0 KPI 22 assumptions are made, see Table 3.2). These assumptions differ from passenger behaviour to the accuracy of the used data. This paragraph analyses the assumptions. If some assumptions can be grouped, this group is examined together. If possible the impact of the assumptions on the passenger punctuality KPI is calculated. A distinction is made between the maximum impact and the expected impact. The maximum impact is based on every trip influenced by the assumption. The expected impact is based on the trips influenced by the assumption and directly changes the passenger punctuality. For example, if due to an assumption some trips are considered delayed all those trips are the basis for the maximum impact value. On the other hand, only the trips with a final delay of more than 3 minutes are considered in the calculation of the expected impact value.

Some assumptions need additional analyses. If so, these analyses can be found in appendix A. An overall conclusion can be found in paragraph 3.3. An overview of all assumptions and the calculated impact values is given in appendix A.6.

3.2.1. The observed trips in the passenger punctuality calculation are representing the average passenger

The NS already examined the assumption that the observed trips in the passenger punctuality calculation are a correct representation of the average passenger on the main rail network [29]. A trip is a journey of a passenger where the origin is not the destination. The origin and destination are determined via the OVCP data. In the calculation of the passenger punctuality not all origin-destination (OD) combinations are taken into account. Every quarter of a year a selection is made to filter out the stations that are not always served: the event stations like Rotterdam Stadium and Heerenveen IJsstadion. Also very rare OD combinations are filtered out to speed up the calculation time. An OD is selected if it satisfies two conditions:

1. On average, in a period of 100 days at least once a day a passenger should make a trip with this OD-combination.
2. On at least 20 of the 100 days a passenger needs to make a trip with this OD-combination.

If both conditions are met, the OD is taken into the selection. The NS examined what the effect was on the passenger punctuality when using the selection of OD combinations instead of all OD combinations. This test was done in 2014 when the second condition was a bit stricter: On at least 50 of the 100

Table 3.2: The 22 assumptions in the calculation of the passenger punctuality [29]

	Assumptions	**Paragraph**
1	The observed trips in the passenger punctuality calculation are representing the average passenger	3.2.1
2	The closing of the OVCP terminals at stations do not have an influence on the representativeness of the used trips in the calculations	
3	Trips derived from the OVCP data are not with the Thalys, Eurostar and/or ICE	3.2.2
4	The checkout station is the desired destination of the passenger	3.2.3
5	The check-in station is the desired origin of the passenger	
6	The check-in time is the desired starting time of the trip of the passenger	
7	Trips (partly) conducted by bus transport are not included in the passenger punctuality	3.2.4
8	The passenger wants as quickly as possible from A to B, even if this is without a train of the NS	3.2.5
9	The passenger travels as advised by the timetable and hereby uses the default settings	
10	The passenger bases his expectation on the timetable two days prior to the journey	
11	Every trip obtained from the OVCP data can be coupled to a promised journey	3.2.6
12	The walking time between card reader and platform is 1 minute	3.2.7
13	The walking time between platform and card reader is 1 minute	
14	The clock in the card readers is exact	3.2.8
15	The realised arrival and departure times are measured exact on every station	3.2.9
16	The determination of 45 seconds as stopping time at a shortstop is accurate	
17	The promised transfers time corresponds with the realised transfer time	3.2.10
18	If a train has a delay of at least 15 minutes before departure, the realised arrival time is the checkout time minus the walking time between train and card reader	3.2.11
19	If a train does not depart, the realised arrival time is the checkout time minus the walking time between train and card reader	
20	If a train does not arrive at the destination or transfer station, the realised arrival time is the checkout time minus the walking time between train and card reader	
21	If a transfer is not made, the realised arrival time is the checkout time minus a walking time between train and card reader	
22	If the first train is missed because it departed too early, the realised arrival time is the checkout time minus a walking time between train and card reader	

days a passenger needs to make a trip with the OD-combination. For 100 days (between 2 June and 9 September 2014) all the OD-combinations occurring were distinguished. The result was a list of 66.423 different ODs. When applying the conditions a selection was made consisting of 19.755 ODs. Although the number of ODs in the selection is only 29.7% of the total number of ODs, 98.36% of all the trips are still taken into account in the calculation of the passenger punctuality.

For nine days the 3 minute passenger punctuality was calculated with and without the selection of ODs. For eight days, the difference was between 0.02 and 0.06 percentage point, with a mean of 0.05 percentage point. On one day (16 Augustus 2014) the difference was 0.15 percentage point. This outlier can be explained by the fact that on that day maintenance was scheduled. When this calculation was done, planned maintenance was not taken into account and could thus results in large outliers. Afterwards the method of calculation was adapted so planned maintenance is no reason for deviation of the passenger punctuality calculation anymore.

The second step is to check if every section of the main rail network is covered by the selection of ODs. If a section is not covered, disturbances on this section will not have an effect on the passenger punctuality. With VISUM, an assignment model, for every OD of the selection a determination is made which sections of the network are used. Then the same is done for every other OD. A check took place to see if there were sections covered by ODs not in the selection but not covered by ODs in the selection. This was not the case. In fact this was still not the case when using only the top 2000 ODs in the selection (instead of the original 19.755 ODs). When using the top 1000 ODs, only one section (between stations Veenendaal-Centrum and Rhenen) was not covered. The first OD covering this section was in the 1017th place.

Note that this check took place with the used ODs between 2 June and 9 September 2014. The selection of ODs can change over time. However the most frequent travelled ODs will stay in the selection. The fact that less than the top 1% of the selected ODs was enough was enough to cover the whole main rail network indicates that also in the future the selected ODs are representative for the whole network.

A third analysis to determine if the used trips are a correct representation of the average passenger is to check if the passengers using the OV-chipkaart are representative. Only the trips of the passengers who check in and check out can be used in the passenger punctuality calculation. According to the NS, in October 2014 80% of the passenger kilometres was made with trips with a check-in and a checkout.

The other 20% can be divided into two groups. The first group is the trips made with a train ticket that does not require a check-in or checkout: the unchipped train tickets, for example an international train ticket. Unfortunately the distribution of these tickets is not evenly in place and time. There is a slightly overrepresentation in the weekend and a small underrepresentation during rush hours. Note that the mean travel distance of this group is larger than the overall mean travel distance, thus the percentage of passenger kilometres travelled in this group is larger than the percentages of trips. On the other hand, the number of different unchipped train tickets is decreasing. The second group is the incomplete trips. A trip is incomplete if a check-in or checkout is missing. Research shows that between October 2012 and September 2013 2.0% of the train passengers did not check in or check out while this was required [30]. If these incomplete trips were evenly distributed among the network, this would be no effect to the overall passenger punctuality. Nonetheless this research shows a variation in the percentage of incomplete journeys among different regions as well (see Table 3.3). This means that the calculated passenger punctuality is slightly off. Note that the number of closed stations is increasing. This means that a passenger cannot enter or leave the station without checking in or checking out. Hereby the percentage of incomplete trips will decrease.

So for both groups the number of involved trips will decrease over time. This implies that the percentages of used trips in the passenger punctuality will increase. For now, the chance that a train is not included in the passenger punctuality is almost zero. With only 6 passengers, the chance that none of those passengers did check in and check out, is less than 0.01%.

In conclusion, by using a selection of ODs in the calculation of the passenger punctuality the maximum deviation will be approximately 0.05 percentage points. However every section of the main rail network is covered by those ODs. This implies that every disturbance is included in the calculation. At this moment, more than 80 per cent off all the trips in the network is used in the calculation and this

Table 3.3: The percentages of incomplete journeys [30]

Region	Total number of journeys (in millions)	Total incomplete journeys		Incomplete journeys, corrected for train
Not urban	2.5	0.2	6.4%	7.5%
Little urban	8.1	0.4	4.9%	5.8%
Moderate urban	23.0	0.7	3.2%	3.8%
Strong urban	98.5	1.8	1.9%	2.2%
Very strong urban	309.9	4.2	1.4%	1.6%
Total	**441.9**	**7.3**	**1.7%**	**2.0%**

percentage will increase over time. Hereby it is almost impossible that a single train is not included in the passenger punctuality. Unfortunately the distribution of the trips not included is not evenly spread out over place and time in the network. The differences are small though, but this still needs to be taken into account certainly when calculating the passenger punctuality on a detailed level. This assumption has a strong connection with another assumption: The closing of the OVCP terminals at stations does not have an influence on the representativeness of the used trips in the calculations. As concluded, the closing of the stations has some influence on the representativeness. However the result is a more accurate calculation of the passenger punctuality.

3.2.2. Trips derived from the OVCP data are not with the Thalys, Eurostar and/or ICE

The travel possibilities as shown in the timetable app and on the website of the NS as default do not include high speed trains like the Thalys, Eurostar or ICE. Also a reservation is required for these trains and the price of a train ticket is higher than on the same trip with a domestic train. Therefore the assumption is made that passengers will not take such a train.

Based on the OCVP data of the first 25 days of June, 52025 passengers checked in at Amsterdam Central station and checked out at Rotterdam Central station. For approximately 7578 of those passengers (14.6%) the fastest option would be to take the Thalys. Nonetheless probably only 1 passenger took this train. The influence of only one passenger on the passenger punctuality is negligibly small (for these 25 days it will be 4.71×10^{-6} percentage point). Of course in this example only the passengers between Amsterdam and Rotterdam are taken into account. The origin and destination can differ and the trip can also be in the other direction, but there is no reason to assume the number of passengers with other origins and destinations using the Thalys, Eurostar or ICE are much higher. Therefore the assumption that the trips derived from the OVCP data are not with the Thalys, Eurostar and/or ICE holds.

3.2.3. Departure and arrival

The following three assumptions are related to each other and therefore will be analysed together:

1. The checkout station is the desired destination of the passenger
2. The check-in station is the desired origin of the passenger
3. The check-in time is the desired starting time of the trip of the passenger

If there is no train traffic possible on a route due to a disturbance passengers need to reroute. If this is possible by the NS, the effect of the disturbance is visible in the passenger punctuality. However it possible that there is no alternative route by the NS and passengers will check out on the last reached station. With no other information it is not possible to determine if this last reached station is the desired checkout station, therefore the assumption is made that this is the case. However if the last reached station is not the desired checkout station the passengers experience a delay. This is not included in the calculation of the passenger punctuality.

Consider the following example. Due to a disturbance between the stations Delft and Den Haag Holland Spoor (HS) there is no train traffic possible. A passenger, with origin station Rotterdam, wants to go to station Leiden. Due of the disturbance, the passenger travels to Delft by train, checks out and takes the tram to Den Haag HS. In Den Haag HS the passenger checks in again to travel the final part to Leiden. Not only this passenger will experience a delay that is not visible in the passenger punctuality, in the calculation a second trip is added. This second trip is without a delay as well and the check-in station is not the origin of the passenger. If the card numbers of the OV-chipkaart are included into the calculation, it becomes visible when the same passenger makes two trips in quick succession. This suggests that this method will include rerouting via other train operators. For the sake of simplicity this is not implemented in the passenger punctuality calculation and the assumption is made that the check-in station is the desired origin of the passenger.

A second case of a disturbed trip is when a passenger checks if there is a disturbance just before the desired departure time. If so, the passenger will wait until the disturbance is resolved and then check in. In this case the passenger cannot leave because of a disturbance and the passenger experiences a delay. However based on the OVCP data the trip was successful. There is no method to correct the passenger punctuality calculation for this phenomenon and therefore the assumption is made that the check-in time is the desired starting time of the trip.

Both three cases have in common that by a disturbance the passenger punctuality is not measured correctly. The exact impact is hard to determine. It is dependent on time, place and duration. However for three disturbances a careful examination is made what the impact of the three assumptions is on the passenger punctuality. The calculation is given in appendix A.1, the outcome can be found in table 3.4. The impact lies between the 0.05 and 0.12 percentage point.

Table 3.4: The impact (in percentage point) of the assumptions on the passenger punctuality KPI

Route	**Impact**	**Duration**	**Remark**
Roermond – Sittard	0.12	312 minutes	Week day, during evening rush hour
Almelo – Deventer	0.05	204 minutes	Week day, during evening rush hour
Alkmaar – Zaandam	0.07	148 minutes	Week day, between rush hours

3.2.4. Trips (partly) conducted by bus transport are not included in the passenger punctuality

If a trip is partly conducted by bus instead of train, passengers tend to check out before getting in the bus or do not even check in. The number of passengers with this behaviour cannot be derived from the available data. Also the realised departure and arrival time is not measured. So a delay cannot be measured via the realisation data. Therefore these trips are excluded from the calculation at all.

As described in paragraph 3.1.2 all trips are categorized. If a trip is serviced (partly) by another train operator or by bus transport the trip is categorized as respectively 'Another train operator' and 'Bus transport'. These trips are neglected from the passenger punctuality calculation. If due to delays or disruptions the promised journey cannot be completed by the passenger, in other words the passenger is not able to arrive with the promised train at the final destination, the realised arrival time cannot be derived from the realisation data. The checkout time is used to measure the delay of each trip. The last category is 'Realised trip'. Here the passenger was able to catch the last train of the promised trip and this train arrives at the final destination of the passenger. In this case the delay is measured via the realised arrival time of the train.

To determine the impact of the assumption that bus transport is not included in the passenger punctuality the checkout time of those passengers is used to measure the delay. As stated this is not unusual. This method is already used if passengers experience some other disruption. Note that both scheduled and unscheduled bus transport are included in this determination. Unscheduled bus transport takes places by disruptions and last minute maintenance. Such causes are not included in the timetable two days before the trip so hindrance of passenger caused by this should be included in the passenger punctuality. Scheduled bus transport takes places by planned maintenance. This is included in the timetable so encountered hindrance should not be included in the passenger punctuality. However the travel time is also adjusted in the timetable so delays can still occur which can have some influence on the passenger punctuality.

The 3 minute passenger punctuality is calculated for the first 25 days of June 2015 as well as the passenger punctuality with bus transport included. If the checkout time is used to determine delays the difference between the planned arrival time and the checkout time is taken. Here 1 minute is abstracted. This 1 minute is the expected walking distance between platform and the card reader. The same methodology is applied to the trips categorized as 'Bus transport': The delay is the difference between the planned arrival time and the checkout time minus 1 minute. The results can be found Appendix A.2. The impact of the assumption that bus transport is not included in the passenger punctuality calculation is on average 0.07 percentage points. Note the large differences between June 6, 7, 13 and 14 and the other days. On these days some large maintenance was scheduled. This resulted in a relatively large amount of trips with scheduled bus transport. Excluding these days from the calculation leads to an average impact of only 0.005 percentage point. On the other hand, the impact of only those days is substantial: on average 0.42 percentage point. Trips categorized as 'Bus transport' are not included in the passenger punctuality calculation, so the maximum impact is the same as the expected impact.

The same calculation is done for the 15 minute passenger punctuality. Here the average impact was 0.04 percentage point. Excluding the days with scheduled maintenance decreases the impact to 0.002 percentage point. The average impact on June 6, 7, 13 and 14 is 0.25 percentage point.

In conclusion, the impact of bus transport on the passenger punctuality is on most days negligible. The reason is that the number of trips where bus transport is included is very small. If the share of trips with bus transport is more extensive the impact on the passenger punctuality increases fast.

3.2.5. Passenger behaviour with respect to the planning of their journey

There are multiple methods to plan a journey. All these methods lead to different choices concerning travel time, route and train operator. To deal with all these possibilities three assumptions are made to model the passenger behaviour regarding journey planning.

1. The passenger wants as quickly as possible from A to B, even if this is without a train of the NS
2. The passenger travels as advised by the timetable and hereby uses the default settings
3. The passenger bases his expectation on the timetable two days prior to the journey

The NS has a journey planner on their website and in their app. This journey planner will search for the best (train) connections based on an origin and destination. Based on the desired departure or arrival time the best solutions are shown. The basis for this solution is the timetable. If there is more time between the period of checking the journey planner and the actual journey, the solutions will be less accurate. The timetable gets a couple of updates towards the execution time [28]. For example small deviations due to planned maintenance are added around three weeks before the execution. And only minutes before execution the timetable can change with predicted departure times based on real time realisation data. Note the underlining assumption on all three analysed assumptions that the algorithm, searching for travel possibilities, will always get the best solutions.

The first assumption can be split into two separate assumptions: The passenger wants as quickly as possible from A to B and The passenger is willing to travel with another rail operator than the NS. To start with the latter, the impact of this assumption is hard to measure. Check-in and checkout data of other train operators is needed, as well as their realisation data to determine the impact. An impact value is therefore not possible to calculate. However there is no direct reason to assume this assumption does not hold. Most of the time travelling with the NS is more expensive than with another train operator. Also on every origin-destination combination served by the NS and another train operator, the travel time by the NS is equal or more than the travel time of the other train operator.

The impact of the assumption that the passenger wants as quickly as possible from A to B is less straight forward. Consider the route between stations Utrecht Centraal and Leiden. Sometimes the quickest route is via station Schiphol. Nonetheless, of all the passengers assumed to travel via Schiphol, 84.40% was actual travelling via a different connection (see appendix A.3.1). This connection had less transfers while the arrival time was only 5 minutes later. This has a maximum impact on the daily passenger punctuality KPI of 0.04 percentage points. The delay is measured via the realisation data when there is no disruption, so the actual impact is on average 0.01 percentage points. At this moment it is unclear which percentage of the passengers in the whole network takes not the quickest travel possibility. And thus the total impact of this assumption on the passenger punctuality KPI is unclear. An alternative might be to calculate the costs of every possible travel options, where the number of transfers, the waiting time, the cancellation rate and the train punctuality is included. The assumption than will be that the passenger travels with the connection with the lowest costs.

The impact of the first half of the second assumption is impossible to determine: The passenger travels as advised by the timetable. But assuming that the journey planner of the NS gives the most optimal options between origin and destination there is no reason for passenger to travel differently. Also if the passenger does not travel as advised by the timetable the question arises whether the NS is responsible when the passenger arrives with a delay. If there was another travel option, promoted by the NS via the journey planner, is it the fault of the NS when the passenger decides to travel otherwise and ending up with a delay?

As explained, the journey planner finds the best train connections. However there are some settings in the planner that can influence the outcome. The assumption is that every passenger uses the default settings to find their promised journey. The different settings are as follows:

- **Use high speed trains** In the default settings this option is off. As shown in paragraph 3.2.2 this is indeed the case in reality. Here all the passengers were observed who departed in Amsterdam and arrived in Rotterdam. Of those passengers, an determination took place who took the a high speed train. Only one passengers did probably took this train.
- **Use detours** Some train tickets allows the passenger to travel unlimited. Default detours are not shown in the options of the journey planner. However it could be beneficial to passengers with such a train ticket. In appendix A.3.2 this option is analysed. Although the percentage of passenger taking a possible detour is significant, the impact on the passenger punctuality is negligible. However more research on this travel option is necessary.
- **Transfer time** If the default settings are on, the journey planner will include travel options with the smallest possible transfer time. It is possible to change the minimum transfer time into 5, 10, 15 or 20 minutes. Given the available data, it is not possible to determine whether passengers change this setting.
- **Type of connection** The algorithm behind the journey planner searches for the most optimal route, which is a combination of travel time and number of transfers. This can be changed into the connection with the least transfers and into the fastest connections. In practice this leads to only a few minor changes. With the introduction of the new website of the NS in December 2015, this option was omitted though in the app of the NS this option is still present.

Four different settings can be changed. For two of those, the use of high speed trains and the use of detours, the impact on the passenger punctuality can be determined. In both cases this impact is negligible. For the other two setting, the transfer time and the type of connection, the impact cannot be determined. However the expectation is that here the impact is also too small to have significant impact. Therefore the assumption that the passengers use the default setting when using the journey planner holds.

As stated, the timetable used by the journey planner is updated several times before the execution. This means that with the same input, in time the journey planner can give different options. The passenger punctuality can thus change by using the timetable on different moment in time prior to the execution because the travel options change. The third assumptions states that the time between checking the journey planner and the trip itself is two days.

Consider the following. Every passenger checks the journey planner 10 minutes before departing. On that moment every disruption and delay is known so the promised journey is based on this information. The result is high passenger punctuality but this KPI does not give a correct indication on the performance of the NS because the performance is worse than the KPI. The opposite is also possible. Every passenger checks the journey planner two weeks before departing. Scheduled maintenance is not included yet so the promised journey takes this not into account. Although the NS clearly indicates in the two weeks prior to the execution that there is maintenance on the route, the passenger punctuality is very low. The KPI shows a lower performance than in reality.

The threshold of two days between checking the journey planner and the actual departure ensures that planned deviations to the timetable are uncounted for in the passenger punctuality but disruptions and last minutes changes due to extreme weather are included in the KPI. A higher threshold gives a worse KPI and a lower threshold gives a better KPI, while the performance of the NS stays the same. So in conclusion the assumption of a threshold of two days holds.

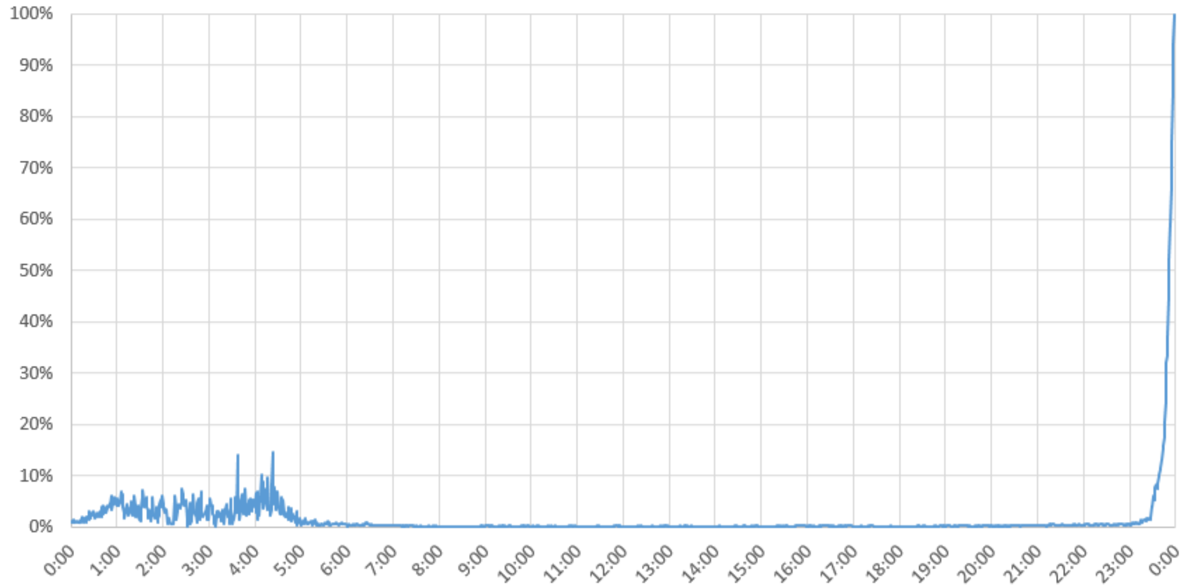


Figure 3.2: The percentage of trips without a promised journey for the first 25 days of June 2015

3.2.6. Every trip obtained from the OVCP data can be coupled to a promised journey

The number of records in the table with every trip obtained from the OVCP data (and where the OD is part of the selection of ODs, see paragraph 3.1.1) is always larger than the number of trips that can be coupled to a travel option, the promised journey. Three different reasons can be distinguished. To visualize these reasons, the percentage of trips without a promised journey is plotted against the time. The dataset used are all the trips of the first 25 days of June, in total more than 21 million records. The result is visible in Figure 3.2.

Three areas can be distinguished. The first area is between midnight and 6:00. Here a clear increasing in the percentage of trips without a promised journey is visible. On average this percentage is 2.68% but has peaks to more than 14%. This is due to the conditions included in the passenger punctuality. One of the conditions is that the planned departure time of the promised journey should be within the hour of the check in time. In the daytime this condition should not have any influence on the number of trips without a promised journey. In almost the whole network the minimal cycle time is 30 minutes. However during the night this is not the case. If a check-in time is later than one minute before the latest departure no travel option will be found. The maximum impact of this phenomenon is 0.02 percentage point.

The second area lies between 23:00 and midnight. Here the percentage of trips without a promised journey increases rapidly to 100% at 23:59. This is due to the fact that at this moment the algorithm behind the passenger punctuality KPI is not able to find travel options if the check in time is before midnight and the first possible travel option is after midnight. The average percentage of trips without a promised journey because of this reason is 16.79%. The maximum impact is 0.17 percentage point.

The last area is visible between 6:00 and 23:00 but it is present during the whole day. Here the percentage of trips without a promised journey is on average 0.22%. This means that there are always trips that cannot be coupled to a travel option regarding the circumstances. The maximum impact is 0.20 percentage point.

To check where this error originated, first a check took place to see the percentage of trips without a promised journey per departure station. For 81.23% of the stations the percentages was lower than 1%. For 96.39% of the stations the percentage of trips without a promised journey was lower than 2%. Ten stations had higher percentages. The maximum impact of the trips departing on these ten stations on the passenger punctuality is 0.02 percentage point. Nine of those stations had 8.4% or fewer trips without a promised journey. The last remaining station, Nijmegen Goffert, had a 100% score. This means that none of the trips departing at this station could be coupled to a travel option.

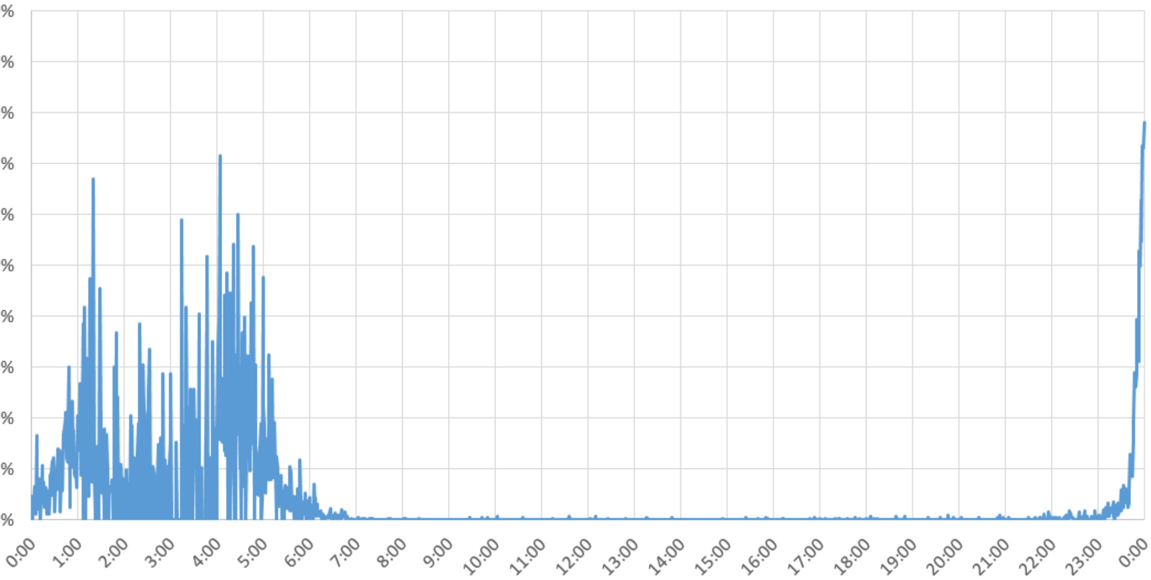


Figure 3.3: The percentage of trips without a promised journey for the first four days of August 2015

Analysing the arrival stations of the trips without a promised journey shows a same kind of result. Here also station Nijmegen Goffert had a 100% score. 14 stations (excluding Nijmegen Goffert) had a score of more than 2% with a maximum of 11.02% trips without a promised journey. The total impact of these stations is 0.03 percentage point.

A more remarkable result is when analysing the different origin-destination (OD) combinations. Of the 22613 possible ODs, all the trips on 1493 ODs (6.60%) could not be coupled to a promised journey. This has a maximum impact of 0.18 percentage point. For 89.97% off all ODs the percentage of trips without a promised journey is less than 1%.

Based on these findings the algorithm was adapted. There were some faults in the algorithm that could be fixed. Therefore the same kind of figure like Figure 3.2 was created, but now with the data of the first few days of August. Here some of the problems should be solved. The result can be found in Figure 3.3.

As can be seen the third problem is as good as solved. Every station is included and almost every OD combination. Nonetheless for 2 ODs 100% of all trips could not be combined with a travel option, but both those ODs had only one trip. Also visible is that the percentage of trips without a promised trip close to midnight is decrease to a maximum of 15.61%. So for the greater part this problem is also solved. The first problem, trips that could not made the last train of the day, is not yet solved. A solution could be to determine the travel options based on the checkout time when there is no travel option based on the check-in time. With this adaption to the algorithm the maximum impact of the assumption that every trip could be coupled to a travel option is 0.09 percentage point.

3.2.7. Walking time

In the passenger punctuality two assumptions are made concerning the walking time on stations, namely:

1. The walking time between card reader and platform is 1 minute
2. The walking time between platform and card reader is 1 minute

The NS already evaluate these assumptions [29]. A dataset of 4.6 million trips was used. For all these trips the realised arrival time could be determined. Taking the difference between the arrival time and the checkout time gives the walking time at the destination. Not every passenger would directly checkout after arriving or some passengers did not travel according to the promised journey. The average difference is thus not the best method to find the most likely walking time between platform and card reader, because taking the average would give every trip the same weight. Therefore the

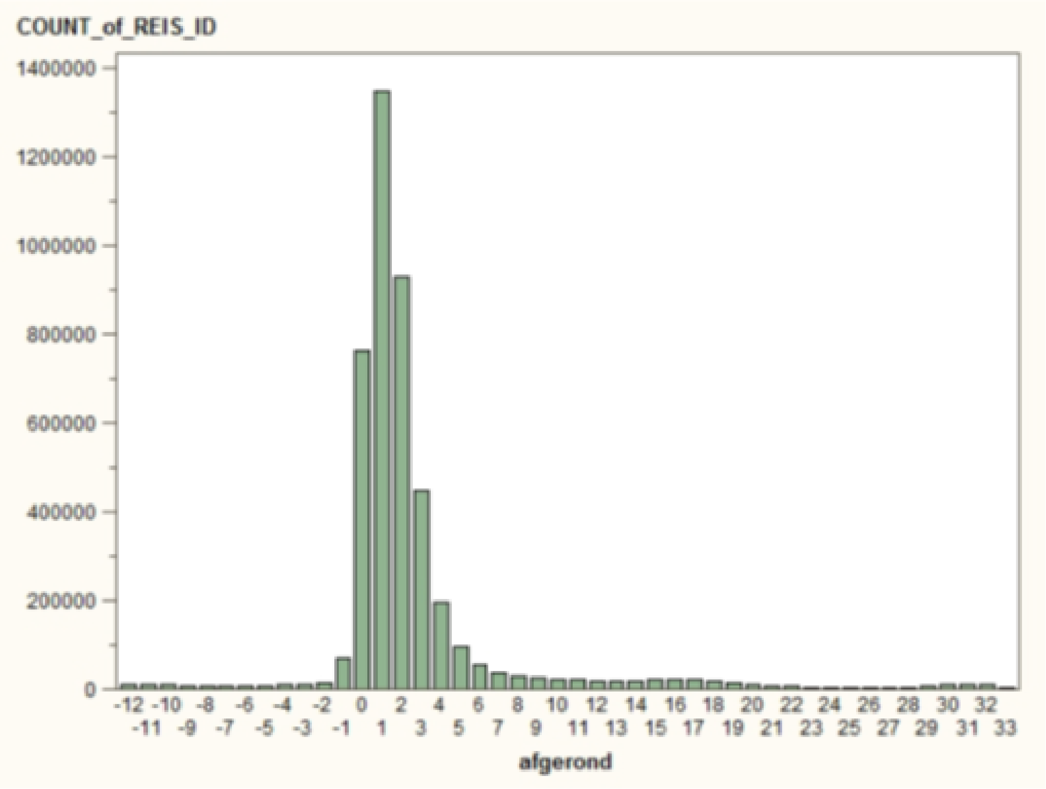


Figure 3.4: The number of trips per walking time interval [29]

mode was taken. Outliers would not have any effect on the found walking time. The results of the NS are given in Figure 3.4. The most trips had a walking time of 1 minute so the NS concluded that these two assumptions hold.

The method of the NS is based on, among others, the assumption that the walking distance of every station is the same. However on a station like Delft Zuid the card readers are placed on the platform while at station Utrecht Centraal the card readers are placed at the entrance of the station hall. The walking distance will differ per station. Although in the data set most trips had a walking distance of 1 minute, this group consist of only 29.26% of all the trips.

A better method would be to determine the walking distance per station. Therefore a dataset of almost 20 million trips was taken. For all these trips the realised arrival time of the last train could be determined. Here the difference between the realised arrival time and the checkout time was taken as well. Calculating the mode per station will lead to the expected walking time per station. The found values are given in appendix A.4. Note that with this method the walking time from platform to card reader is calculated. The checkout card reader is the same as a check-in card reader, so the walking time from check-in to the platform is assumed the same.

To determine the impact of the walking time of 1 minute on the origin side of a trip all the trips of the first 10 days of June 2015 were examined. The first possible departure was calculated based on the check-in time and the new determined walking times. When this first possible departure time was later than the promised departure time the assumption is made that the promised journey was not correct. This applies on average to 3.19% of the trips per day. Nonetheless for 80.26% of those trips the final delay was less than 3 minutes. This suggests that although the promised departure time was too early, this promised journey could still be completed. The result is that for 19.74% of the trips with an expected wrong promised journey the trips ends up with a delay. On average, this applies to 0.63% of all the trips per day. The maximum impact is thus 0.63 percentage points. If a trip needs to be rescheduled due a hindrance, the checkout time is used to determine the delay. For these trips the impact of the wrong promised journey on the passenger punctuality can be neglected. So the expected impact on the passenger punctuality are the trips categorized as 'Trip realised' where the promised departure time was too early and where the delay was at least 3 minutes. This applies to 0.42% of the trips per day.

The impact of the walking time of 1 minute on the destination side is determined by analysing the first 25 days of June 2015. Note that only trips that need to be rescheduled are influenced by different walking times. Per trip the final delay is calculated based on the new walking times. Two groups of trip can be distinguished.

1. This group contains trips that have a delay of at least 3 minutes with a default walking of 1 minute, but the delay is less than 3 minutes by a station specific walking time.
2. This group contains the trips without a delay of at least 3 minutes by a walking time of 1 minute but with a delay of at least 3 minutes by a station specific walking time.

More than 21 million trips were analysed but group 1 consist on only 13 trips. Group 2 consist of more than 24 thousand trips. The maximum trips impacted are the maximum of group 1 and group 2. The expected trips impacted are the difference between group 1 and group 2. Group 1 is very small, so the maximum and expected impact on the passenger punctuality is the same: on average 0.12 percentage points per day.

In this calculation, 28.91% of all the trips had a realised walking time of 1 minute. 38.99% had a realised walking time equal to the walking time associated by the station of destination. This percentage can further increased by determining the walking time between platform and (group of) card readers per station. Data to calculate these walking times is already available at the NS.

3.2.8. The clock in the card readers is exact

The passenger punctuality is based on both the realisation data and the OVCP data. The clock on each card reader should run parallel with the realisation data to prevent mistakes. It could occur that the clock of a card reader is not exact. The expectation is that the impact of this error on the passenger punctuality is negligible. Every card reader is calibrated at least once a day. Also there are always multiple card readers per station. If the clock on a card reader is not exact, this error is thus solved within a day and the number of passengers involved is only a fraction of the total number of passenger checking in or out at the station. Unfortunately the exact impact on the passenger punctuality is not calculable.

3.2.9. The realised arrival and departures times

Two assumptions are related to the realised arrival and departure times. Those assumptions are:

1. The realised arrival and departure times are measured exact on every station
2. The determination of 45 seconds as stopping time at a shortstop is accurate

There are two systems that determine the arrival and departure times. The first system calculates the arrival and departure time based on the time a train passes a signal [31]. This system is used in the calculation of the passenger punctuality. The arrival time is calculated as follows. The time a train passes the entry signal is registered. This passing time is added with a fixed correction factor based on the combination of arrival track and platform track. This new time is the expected arrival time. A visualisation of this method is given in Figure 3.5. The correction factor is the expected running time. Although this is a constant factor the actual running time can differ based on, among others, the speed of the train. This leads to an error in the calculated arrival time. Likewise if there is no correction factor for a certain combination of arrival track and platform track, a default value of 0 seconds is used. An example is station Bovenkarspel Grootebroek. The correction factor is here 0 seconds while the real time between signal and platform is approximately 2 minutes. The result of this is a calculated walking time between train and card reader of 3 minutes (see appendix A.4). This is much larger as expected for such a small station. Besides the departure time is calculated via the same method. Here the time the train passes the exit signal is taken and subtracted with a fixed correction factor. The calculated arrival and departure times are stored in the Nieuwe Vervoers Gegevens Bank (New Transport Data Bank, NVGB).

Another drawback of the NVGB is that on shortstop stations the arrival time is not calculated at all. The arrival time is set equal to the departure time. At some locations this departure time is based on extrapolation of the measured delay on the last past service control point (in Dutch: Dienstregelpunt, an important geographical point in the rail network).

Figure 3.5: Schematic visualisation of the arrival time calculation used in the NVGB [31] (edited)

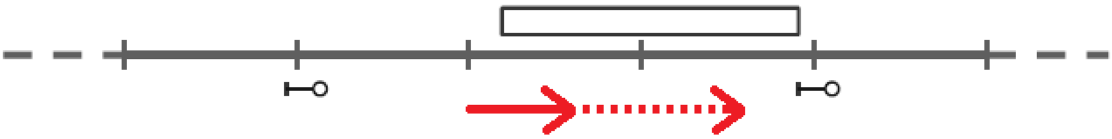


Figure 3.6: Schematic visualisation of the arrival time calculation used Trento [31] (edited)

The second system to determine arrival and departure times is called Trento. The arrival time is calculated as follows. The whole rail network is divided into sections. The sections are much smaller than the distance between two signals, or signal and station. The time that the last section before a platform is cleared is added with a variable correction factor to determine the arrival time. This correction time is based on the distance but also on the type of train and place of the train along the platform. A visualisation of this method is given in Figure 3.6.

This variable correction factor, plus the fact that the distance between the last measuring point and the platform is a lot smaller, makes Trento more reliable than the NVGB. Under the name 'Meten is Weten' (Measurement is Knowledge) in approximately 6 months more than 5000 measurements were done [31]. Here the exact arrival and departure times were measured by hand with atomic clocks. The outcome was compared with the calculated values of the NVGB and Trento. The average deviation of the NVGB was 20 seconds with outliers of more than a minute. The average deviation of Trento was 5 seconds with outliers of only 11 seconds. The main reason why the NVGB is used for the passenger punctuality is that Trento is not auditable. Another large disadvantage of Trento is the need for good management in comparison with the NVGB. Also not every station is included in Trento although the number of excluded stations is decreasing fast.

At this moment, ProRail is working on a new system called MCS. This system has some similarity with Trento. Here the arrival and departure times are also based on the time a section is cleared or occupied. However the correction factor is not based on speed, type of train and platform lay out. Over time MCS shall replace the NVGB. Trento might still be more accurate but the large benefit of MCS is that is it auditable.

As stated, in the calculation of the passenger punctuality data of the NVGB is used to determine the realised arrival and departure times. This data is not exact but has an error which implies that the passenger punctuality has an error. The delay for every passenger arriving without a disturbance is calculated based on the realised arrival time. So the realised arrival time has on average an error of 20 seconds. For the first 25 days of June the average 3 minute passenger punctuality lies hereby between the 82.24% and 87.20%. This means a difference of almost 5 percentage point. Hereby the influence of the error on missed transfers or trains departing too late or too early is neglected.

The largest disadvantage is the fact that the exact error on the realised arrival and departure times is unknown. However it is certain that the first assumption does not hold. An error of 20 seconds is likely which clearly influences the outcome of the passenger punctuality.

If Trento is used the average error is 5 seconds. The maximum difference in the 3 minute passenger punctuality will be 0.72 percentage point. For the 15 minute passenger punctuality this difference will be only 0.02 percentage point. The disadvantages of Trento can be overcome by using Trento if possible and NVGB otherwise.

As stated, the NVGB does not calculate the arrival times at shortstops. Therefore the determination of the arrival time at such a shortstop is the departure time minus 45 seconds. This value of 45 seconds is the median of more than 6.7 million stopping times at shortstops measured by Trento. This is thus a good estimation of the stopping time on an average station, but locally there can be some large deviations. An example is station Brummen where the mode of the difference between the arrival time and checkout time is negative (see appendix A.4). This means that the actual arrival time is earlier than 45 seconds before the departure time. Just like the walking times (see paragraph 3.2.7) a stopping time per station could be a more reliable value for the passenger punctuality. An even better solution would be to use the data of Trento. Trento does measure the arrival times on shortstops. The default value of 45 seconds (or better: the default value per station) can be used if Trento is not reliable.

3.2.10. The promised transfers time corresponds with the realised transfer time

The smallest possible transfer time between two platforms is determined as follows. First a check takes place to determine if a platform specific transfer time is available. This is a specified transfer time between two unique platforms. If such a transfer time is not available, the default transfer time per station is used. If this station specific transfer time is not available as well, a default value of 5 minutes is used. If the realised transfer time is smaller than the found value, the transfer is considered missed. The journey planner uses different values to determine if a transfer can be made. For example the minimum transfer time in the journey planner is 2 minutes. However by a cross platform transfer the used transfer time in the passenger punctuality calculation can be 1 minute. Although there are differences in transfer times between the journey planner and the passenger punctuality calculation, it can never occur that the journey planner suggest a transfer what cannot be made in the passenger punctuality calculation.

If a trip is categorised as 'Missed transfer' the difference between the realised arrival and departure time at the station is smaller than the corresponding transfer time. If a transfer is missed the conclusion is drawn that the passenger needs to reschedule. The result is at least an expected delay of the time between two consecutive trains on the second part of the promised journey. However in the first 25 days of June 2015 405.405 trips with only one transfer were categorized as 'Missed transfer'. 206.882 of those trips had a delay smaller than 3 minutes. This is 51.03%. This is partly because passengers are able to make unofficial transfers (transfers where the transfer time is less than the planned difference between arrival and departure) but in most cases the determined transfer time between platforms is too pessimistic.

A reason of this pessimistic behaviour is that if a platform specific transfer time is not available the station specific transfer time is used. This station specific transfer time is never smaller than a platform specific transfer time for the same station. An example is station Haarlem. It is possible that a transfer needs to be made between platforms 5 and 6. The transfer time between those two platforms is not specified. Therefore the default value of Haarlem is used. This value is 4 minutes. However the transfer time between platforms 5 and 6a and between platforms 5 and 6b are specified. These transfer times are 2 and 1 minute respectively. This means that in this case the used transfer time is 2 to 4 times higher than the real transfer time. Hereby trips can get the category 'Missed transfer' while in fact the transfer was not missed at all.

Another example is station Schiphol. If the planned arrival platform is equal to the planned departure platform the minimum time between realised arrival and departure should be at least 3 minutes. This is because due to safety measures, it is not possible to arrive and departure with a different train at the same platform within three minutes. If the planned arrival platform is equal to the planned departure platform and the realised transfer time is less than three minutes the conclusion is drawn that the planned and realised arrival or departure platform differs. The minimum transfer time is than the station specific default value. At Schiphol, this default value is 4 minutes. However at Schiphol the arrival platform tracks are known only a couple minutes before the actual arrival. This is due to the Dynamic Traffic Management (DVM) installed at Schiphol. Therefore in the timetable the arrival and departure platform are given as '1-2' or '5-6' instead of platform 1, 2, 5 or 6. It is possible that the planned arrival platform is '5-6' and the planned departure platform is '5-6'. If the realised transfer time is less than three minutes the default value of 4 minutes applies and the transfer is thus categorized as missed. In reality the arrival platform is 5 and departure platform is 6. The walking distance between those platforms is only 1 minute and thereby the transfer is made. The underlying problem is that the determination of transfer times is based on the planned arrival and departure platforms. The realisation data is derived from the NGVB (see paragraph 3.2.9). This system does not log the realised arrival and departure platforms. A solution is Trento. This system logs the realised arrival and departure platforms.

The impact of both causes on the large group of passengers without a missed transfer is hard to determine. As stated 51.03% of the trips categorized as 'Missed transfer' had a delay less than 3 minutes. For 42.44% of those trips the station specific transfer time is used instead of the platform specific transfer time and for 1.89% of the trips the planned arrival platform is the same as the planned departure platform. Moreover, the percentage of trips where the station specific transfer time is used instead of the platform specific transfer time off all the trips categorized as 'Missed transfer' is 39.87%. For the transfers where the arrival platform is the same as the departure platform this is 2.54%.

So for 44.33% of the trips where the transfer was made but was categorized as 'Missed transfer' the reason is one of above. Note that the only condition to assume if a transfer was made is if the final delay was less than 3 minutes. For the other 55.67% the reason is that the transfer times are possible too large and that the conditions used in the passenger punctuality calculations are too strict. Take for example a transfer time of 2 minutes. If the difference between arrival time and departure time is 1:59 minute the transfer is categorized as not made. Such a condition is harsh certainly when there is an error in the realised arrival and departure time of on average 20 seconds (see paragraph 3.2.9).

The difference between trips where the transfer was made and where the transfer was missed is the determination of the final delay. In the first case the realised arrival time is the basis for the delay calculation, in the second case the checkout time is used. In paragraph 3.2.11 the impact of this difference is calculated. In conclusion, the assumption that the promised transfers time corresponds with the realised transfer time does not hold. However the delay calculation by rescheduling does not take the reason of rescheduling into account. Hereby the mistakes made by wrongly categorizing trips as 'Missed transfer' are corrected.

3.2.11. Reasons to reschedule

There are five possible scenarios where a trip needs to be rescheduled, see paragraph 3.1.2. Instead of determining the exact route a passenger took, the final delay is determined via the checkout time. Note that in the official examination of the assumptions of the passenger punctuality only four reasons are mentioned: the scenario 'Train depart too early' is not mentioned [29]. The corresponding assumptions are as follows:

1. If a train has a delay of at least 15 minutes before departure, the realised arrival time is the checkout time minus the walking time between train and card reader
2. If a train does not depart, the realised arrival time is the checkout time minus the walking time between train and card reader
3. If a train does not arrive at the destination or transfer station, the realised arrival time is the checkout time minus the walking time between train and card reader
4. If a transfer is not made, the realised arrival time is the checkout time minus a walking time between train and card reader
5. If the first train is missed because it departed too early, the realised arrival time is the checkout time minus a walking time between train and card reader

For the latter four scenarios (train did not depart, train did not arrive, missed transfer and train depart too early) it is clear why a trips needs to be rescheduled. The passenger is not able to continue the trip as promised. This is not necessarily the case when the train departs with a delay of at least 15 minutes. The elaboration on the assumption that the threshold value of 15 minutes is correct is given in appendix A.5. In conclusion, the size of the initial delay on which the passengers starts to reschedule their trip differs per person, place and time. The chosen value of 15 minutes will results in an impact of approximately 0.09 percentage points on the 3 minute passenger punctuality and 0.18 percentage points on the 15 minute passenger punctuality. For the current timetable a default value of 15 minute as initial delay before rescheduling is acceptable because there is practically no route where the frequency of train series is more than 4 per hour. However if this changes a variable value, based on the frequency of the train series, might be a better solution.

If a trips needs to be rescheduled, five possible solutions can be distinguished [32]. These solutions are listed below. The order is the accuracy, so the most accurate solution is the most upper solution. The effort involved to implement each solution is directly related to the accuracy. The more accurate, the more effort it takes to implement the solution into the passenger punctuality calculation.

1. **Based on actual real time travel information** Rescheduling takes always place last minute. The passenger is thus able to use real time travel information to reschedule. Conditions to this solution are the need for the historic real time timetable of the journey planner and the time after which a passenger is going to reschedule.
2. **Based on the realisation data** The realisation data combined with a shortest path algorithm can lead to an alternative for the passenger. A condition is that the time after which a passenger is going to reschedule needs to be known. This solution uses hindrances further down the route to get the best alternative. However those hindrances are not necessarily known by the passenger while rescheduling.
3. **Based on promised journeys of other passengers** The available data consist of a list of all travel options. If a trip needs to be rescheduled the best alternative travel option off all the available travel options is taken. Here also the time after which a passenger is going to reschedule should be known. The largest disadvantage is that on the less frequent origin-destination combinations there is not always another travel option within a certain time limit, for example 1 hour. The same applies to the night hours.
4. **Promised journey based on the checkout time** A promised journey is the travel option with the earliest possible arrival time and where the first train departs after the check-in time. If a trip needs to be rescheduled, this new journey could be the travel option with the latest possible departure time where the last train arrives before the checkout time. The disadvantage is that the checkout time is not necessarily the actual arrival time.
5. **No rescheduling** This is the solutions currently in use. If a trips needs to be rescheduled, the checkout time is used to determine the expected arrival time. Nonetheless the checkout time is not necessarily the actual arrival time.

The rescheduling, in combination with the realisation data, is needed to determine the expected arrival time. With this new expected arrival time the delay can be calculated by taking the difference between this arrival time and the original promised arrival time. The more accurate the rescheduling takes place, the more accurate the delay can be calculated. The result is more accurate passenger punctuality.

In the method currently in use trips will not be rescheduled. Instead the checkout time is used to calculate the expected arrival time (this is the checkout time mines the walking time between platform and card reader). More than 5.6 million trips categorized as 'Trip realised' are examined. The advantages of those trips are that both the realised arrival and the checkout time are known. 86.97% of the trips had arrived within 3 minutes of the promised arrival time. If the arrival time based on the checkout time is used, only 77.05% of the trips will arrive within 3 minutes. This is a decrease of 11.41%. This could partly due to the fact that the used walking time of 1 minute might not be correct (see paragraph 3.2.7) or that there is a deviation in the realisation data (see paragraph 3.2.9) but this does not explain the whole difference. The same calculation can be done for a delay of 15 minutes. 99.36% of the trips arrived within 15 minutes of the promised arrival time. Based on the checkout time this is 96.45%: a decrease of 2.93%.

These calculated deviations apply to by trips without a hindrance. Under assumptions that these deviations are the same for the trips that needs rescheduling the impact on the passenger punctuality can be calculated. For the first 25 days of June the number of trips that needs rescheduling and with a delay of less than 3 minutes are determined. This delay is calculated via the checkout time. The expected number of trips with a delay less than 3 minutes is thus 11.41% higher. This expectation is added by the number of realised trips with a delay of less than 3 minute. The result is divided by the total number of trips used in the passenger punctuality to achieve the 3 minute passenger punctuality corrected for not rescheduling. The average result is a calculated passenger punctuality that is 0.23 percentage point higher than the original passenger punctuality. By the 15 minute passenger punctuality the maximum impact is on average 0.10 percentage point.

So by rescheduling the trips the expected passenger punctuality will be higher. The expected increase depends on the solution for rescheduling and the corresponding error. Further research to determine the impact of each solution is needed. The maximum impact will be 0.23 percentage point for the 3 minute passenger punctuality and 0.10 percentage point for the 15 minute passenger punctuality. Only trips with a delay of at least 3 or 15 minutes are influenced by these assumptions, so the expected impact is the same.

3.3. Conclusion

The goal of paragraph 3.1 is to give some background information about the algorithm that calculates the passenger punctuality. It provides the bases of some assumptions, examined in paragraph 3.2. It explains the trip categories as well. This is the bases for the components, as described in paragraph 4.1.

The difficulty of calculating passenger punctuality 2.0 is the fact that the needed data is not available at the same time. The data out of the timetable should be collected two days prior to the execution. This is due to the fact that there is no historic timetable. It is therefore impossible to determine the travel options, as given two days before execution, after the execution already took place. The OVCP data is available only after a couple of days after the execution. Altogether this explains why first all travel options are determined, than these travel options are compared with the execution data and finally every trip is combined with travel options to get the promised journey. This is the same order as the availability of the needed data.

The eight trip categories explain why a trip is being ignored in the calculation or why a trip needs to be rescheduled. Most categories are self-explaining. For example, the NS is not responsible for the performance of other rail operators. Trips that are made with another rail operator are therefore not taken into account. Also if a transfer could not be made, a passenger should search for an alternative: the trip is being rescheduled. The category 'Bus transport' is an exception. By planned maintenance it is possible that the NS scheduled busses instead of trains. These busses are included in the timetable at least two days prior to the execution and the travel times are adjusted. The busses are offered by the NS, so the performance of the busses should be included in the passenger punctuality. However due to the unavailability of realisation data of the busses, the realised arrival time cannot be determined and thus the delay cannot be calculated. Therefor trips with this category are neglected in the passenger punctuality. Nonetheless a better solution is to use the checkout time to calculate the final delay of each passenger. Via this method these trips are included in the passenger punctuality. Thereby this is not an unusual solution: the same is done for trips that need to be rescheduled (see paragraph 3.2.4).

22 assumptions are included in the method of calculating the passenger punctuality KPI. These assumptions are analysed in paragraph 3.2. An overview of the estimated impact values can be found in Table 3.5. In appendix A.6 a more extensive overview can be found. For some of the assumptions the impact could not be determined. Mostly this was due to the fact that the needed data are not available. For some assumptions the impact for a specific situation can be determined but not the overall impact on the passenger punctuality. An example is the impact when temporary no train traffic is possible on a route. Due to three assumptions (assumptions 4, 5 and 6) this hindrance is neglected in the calculation of the passenger punctuality. However the impact is time, place and duration dependent and the amount of such disruptions varies from day to day as well. Therefore it is not possible to give an overall impact value.

If possible a distinction is made between the maximum impact and the expected impact. The maximum impact is the impact caused by the trips influenced by the analysed assumption. The expected impact is caused by the trips influenced by the analysed assumption and where the delay changes by this assumption from less than 3 minutes to at least 3 minutes. These trips have thus a direct influence on the 3 minute passenger punctuality.

Assumption 7 states that the trips conducted by bus transport are not included in the passenger punctuality. The impact is in most cases negligible, though on days with planned maintenance the impact could be 0.42 percentage point. Assumption 18 (if a train has a delay of at least 15 minutes before departure, the checkout time is used to calculate the delay) is included twice in Table 3.5. The

Table 3.5: The 22 assumptions of the passenger punctuality 2.0 KPI and the corresponding impact

	Impact on 3 minute KPI	
	Maximum	*Expected*
1	0.05 percentage point	Undetermined
2	Undetermined	Undetermined
3	Negligible	Negligible
4 - 6	Time and place dependent	Time and place dependent
7	0.42 percentage point / negligible	
8	Undetermined	Undetermined
9	Negligible	Negligible
10	Negligible	Negligible
11	0.09 percentage point	Undetermined but probably negligible
12	0.63 percentage point	0.42 percentage point
13	0.12 percentage point	
14	Undetermined	Undetermined
15	0.72 percentage point	Undetermined
16	Undetermined	Undetermined
17	Undetermined	Undetermined
18	0.10 percentage point	0.09 percentage point
18 – 22	0.23 percentage point	

threshold of 15 minutes leads to a maximum impact of 0.10 percentage point while the fact that the checkout time is used leads to an impact value (together with assumptions 19 to 22) of 0.23 percentage point.

For 5 assumptions the maximum impact could not be determined and for 3 assumptions the maximum impact is negligible. For the remaining 14 assumptions the maximum impact could be determined although for some only per case or situation. The result is in any case a maximal impact of 1.94 percentage point. This means that the real 3 minute passenger punctually could be 1.94 percentage point lower than calculated.

For 8 assumptions the expected impact could not be determined and for 3 assumptions the expected impact is negligible. The 11 remainder assumptions result in a combined expected impact of at least 0.86 percentage points on the 3 minute passenger punctuality. Note that some assumptions result in an impact only per case or situation. These are not included in this combined value.

Some remarks to the found values can be given. These found impact values are based on the data of a limit set of days and number of trips. By taking a larger dataset the influence of outliers is minimalized but performing the same analysis on a different dataset might lead to a different result. Moreover the impacts are mostly calculated based on some new assumptions. These assumptions are mentioned in the analyses but the found values are therefore not the exact impact values. Nonetheless it will give a good impression of the order of magnitude of the impact values and therefore on the accuracy of the passenger punctuality KPI.

Second, every (group of) assumption is analysed independent of each other. However some assumptions have a relation with each other and thus the impact values depend on the other assumptions. An example is the walking times between card reader and platform. The impact is calculated by using a default walking time of 1 minute instead of a station specific walking time. However these station specific walking times are determent via the realised arrival time and the checkout time. The realised arrival time could be determined more accurate. The trips influenced by this inaccuracy could be the same trips influenced by the default walking time between platform and card reader.

Lastly, the impact on the daily KPI on national level is determined. On detail level the impact of some assumption could be larger. The best example is given in paragraph 3.2.5. 84.40% of the passengers between Utrecht and Leiden do not travel via the promised journey. On detail level this needs to be taken into account. Nonetheless the impact of this example on national level in only 0.04 percentage point.

4

Breakdown of passenger punctuality

In chapter 3 the structure of the new algorithm to calculate the passenger punctuality is described. This algorithm is divided into five modules. The fifth module generates multiple output tables [24]. With these tables the passenger punctuality can be calculated but these tables provided more information as well. This offers possibilities to additional analyses and performance checks of the rail network. This chapter will describe some of the output possibilities of the available data via performing different breakdowns of the passenger punctuality.

From a passengers' perspective, the most important information stored in the data is the categorisation of each trip (see paragraph 4.1). These categories form the basis of the components: reasons for the passengers why they are delayed. Therefore the passenger punctuality will be broken down into these components. Paragraph 4.1 describes these components. With the data the places in the rail network with the worst passenger punctuality can be found, as well as the places were the size of components is higher than on average in the rail network. A necessity is the need to know by which threshold a trip is delayed. Paragraph 4.2 will discuss two possible delay thresholds. Finally in paragraph 4.3 the breakdown of the passenger punctuality is executed. Hereby different options to perform a break down are taken into account. The results of the breakdowns are visualised via breakdown graphs. Appendix B describes how these graphs can be made. Finally in paragraph 4.4 some final conclusions are made.

4.1. The components

The passenger punctuality can be broken down into reasons why a trip is delayed. These reasons are based on the passengers' perspective and called components. This paragraph will introduce the components. In the first section the origin of the components and some background is described. In the second section the definitions of the components is gives, as used in this research.

4.1.1. The origin of the components

As stated in paragraph 3.1.2, all trips are categorized. Those categories are created to indicate why a trip should not be included or why a trip should be rescheduled. The categories for rescheduling are:

- If the time between check-in and departure is less than 1 minute the promised train could not made: the train departs too early.
- The train departed from the origin station of the trip or from a transfer station with at least a 15 minute delay: the train departs too late.
- The train did not depart at all.
- The last reached station of the train was not the destination station of the trip or a transfer station: the train did not arrive.
- The time between arrival and departure at a transfer station was less than the specified transfer time: missed transfer.

If a trip is included in the calculation and requires no rescheduling the trip is categorized as 'Trip realised'. Note that this does not mean that the trip is without any delays. However the final delay of the trip is the delay of the train at the final destination of the trip.

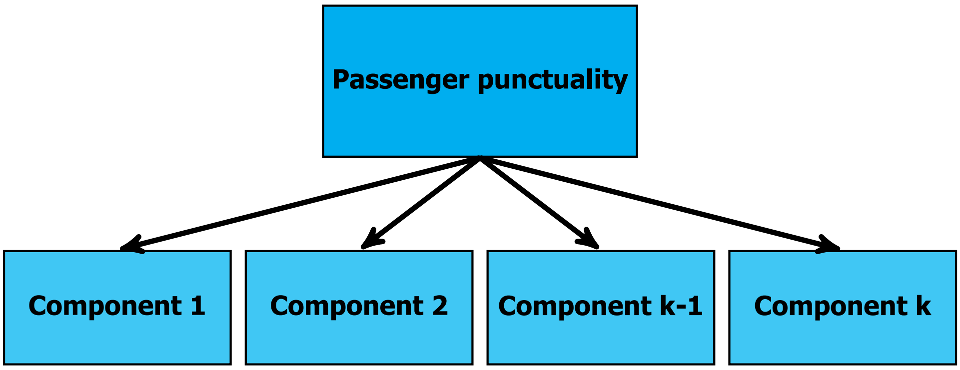


Figure 4.1: The upper part of the framework around passenger punctuality

The first layer of the framework of the passenger punctuality, as proposed in paragraph 1.1, contains the components (see Figure 4.1). Those components are the causes of a final delay, as experienced by the passengers. In basic these are the same as the reasons for rescheduling. That is why the categories are used to define the components. If a passenger experiences a delay because a train did not depart, the component will be 'Train did not depart'. The fact that there is an underlying reason why the train does not depart has no effect on the passenger and thus it has no effect on the component.

Two reasons can be distinguished why the definitions of the categories does not directly apply to the components. First of all, the categories are bound to some strict logic rules. For example if the specified transfer time is 4 minutes and the difference between arrival and departure is only 3:59 minutes, the trips get categorized as 'missed transfer'. However in reality the transfer is probably made. More than 50 per cent of the trips categorized as 'Missed transfer' are actually arrived within a 3 minute delay (see paragraph 3.2.11). In other words, every trip is categorized but only trips with a delay of at least the threshold are part of a component.

Secondly the category 'Trip realised' indicates that there was no need to reschedule the trip. However this does not imply that the trip was without any delay. So in principle this category is a component as well. Nonetheless the name of the category does not contain the meaning of this component. In addition, there is no category 'Train delayed' while there should be such a component. In the experience of the passenger, a delayed train is in most cases the cause for a delayed trip. So if a trip is realised and thus there is no other reason for a delayed trip, the component would be 'Train delayed' instead of 'Trip realised.

4.1.2. The definition of the components

This paragraph will describe the definition of the components used. As stated, the definitions are almost equal to the definitions of the corresponding categories. However there are some small differences. Also, if multiple components are applicable to a trip, the first component encountered that would explain the delay is the component used. The remaining components applicable might not have been a reason for delay if the first component did not occur at all.

- **Train departs too early** This component contains every trip where the time between check-in and realised departure is less than the specified walking time and where the final delay is at least equal to the delay threshold.
- **Train departs too late** If a trip is delayed, so if the final delay is at least the delay threshold, and a train in the promised journey (in Dutch: Reisbelofte) departed with at least a 15 minute delay, then this trip is included in this component
- **Train did not depart** A trip where a promised train did not depart from the origin station of the trip or from a transfer station is included in this component, given that the final delay of the trip is at least the delay threshold. The trains are cancelled.
- **Train did not arrive** If trains are cancelled halfway their journeys, all passengers in these trains encounter this component, on two conditions: the last reached station of the train is not the

promised destination or transfer station of the passenger and the final delay is at least the delay threshold.

- **Missed Transfer** This component includes every trip with a transfer where the time between arrival and departure at the transfer station is less than the specified transfer time, provided that the final delay is at least the delay threshold.
- **Train delayed** Every trip made as promised but with a final delay of at least the delay threshold is included in this component.

4.2. Delay threshold

In basic, the passenger punctuality is calculated as the number of trips without a delay divided by all the trips in the network. A trip is delayed if the final delay is at least the delay threshold. In theory every (positive) value could be the delay threshold. As can be seen in the figures in paragraph 4.3, the breakdowns are performed for every threshold between the 3 minutes and 1 hour. Nonetheless two thresholds get some extra attention.

First there is the 3 minute passenger punctuality. A threshold of 3 minutes is used in the official KPI. So every trip with a final delay of 2:59 minutes or less is considered on time. A trip with a final delay of 3 minutes or more is considered delayed. For these trips the component why the trip is delayed is important for the breakdown. To increase the passenger punctuality the breakdown of a threshold of 3 minutes is needed to determine the importance of each component by this threshold.

Another interesting threshold is by the 15 minute passenger punctuality. A research of the NS investigated the appreciation of the NS by the passengers during a disturbance [27]. The results were the following. Final delays of less than 15 minutes were perceived as a minor disturbance. However if the final delay was more than 15 minutes, the trips was perceived as disturbed. The rating of the NS dropped from sufficient to inadequate. A decrease in trips with a delay of at least 15 minutes is therefore desirable. Determining the size of each component by this threshold the 15 minute passenger punctuality becomes more insightful.

The border of 3 minutes and 1 hour of the analysed area of the breakdowns are chosen for the following reasons. The minimum border of 3 minutes is the first threshold with special attention. A higher minimum threshold was therefore not desirable. A lower threshold was possible but delays less than 3 minutes are not perceived as a disturbance. Also the maximum number of trips with a delay would be on average more than twice as high if the threshold would be only 1 minute as well. This would not benefit the clearness of the resulting graphs. Thereby due to the accuracy of the calculations the passenger punctuality with a delay threshold lower than 3 minutes would not been very reliable, see paragraph 3.2. The upper bound of 1 hour is chosen because after one hour most delayed trips are caused by wrong check outs of wrong promised journeys. Also between 45 minutes and 1 hour the size of each component decreases on average only 2.5 to 3 percentage points.

4.3. Results

With the available data, graphs like Figure 4.2 can be made. Appendix B describes the algorithms to create such breakdown graphs. Note that this figure is just an example: here the various elements of the breakdown graph are clearly visible. The red line is the passenger punctuality for every delay threshold between 3 minutes and 1 hour. The passenger punctuality is calculated as the number trips where the final delay is less than the threshold delay divided by all the trips. In this case on June 1 2015 1.004.955 trips are included in the calculation. The maximum number of delayed trips, at the 3 minute passenger punctuality, is 160.429. This means that on this day the 3 minute passenger punctuality is $\frac{1004955-160429}{1004955} = 84.04\%$. Those delayed trips can be divided into six components. A 100% score means that all the delayed trips at the 3 minute passenger punctuality are caused by the same component.

For June 1, the large share of trips where the train was delayed is clearly visible, however almost none of those trips had a delay longer than 15 minutes. Moreover, there were almost no trips where the train departed too early. If a train did not arrive, the final delay was at least 30 minutes, because this share stays almost the same for the first half hour.

A daily breakdown graph provides information about the performance of the NS and ProRail. It also

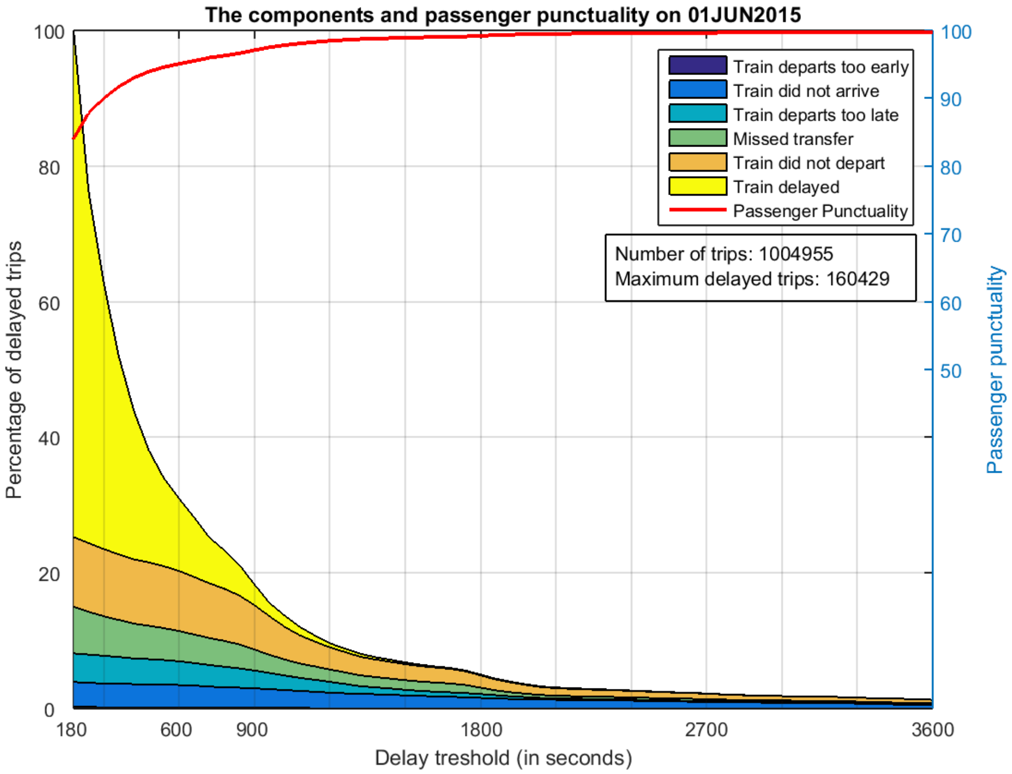


Figure 4.2: The breakdown graph for June 1, 2015

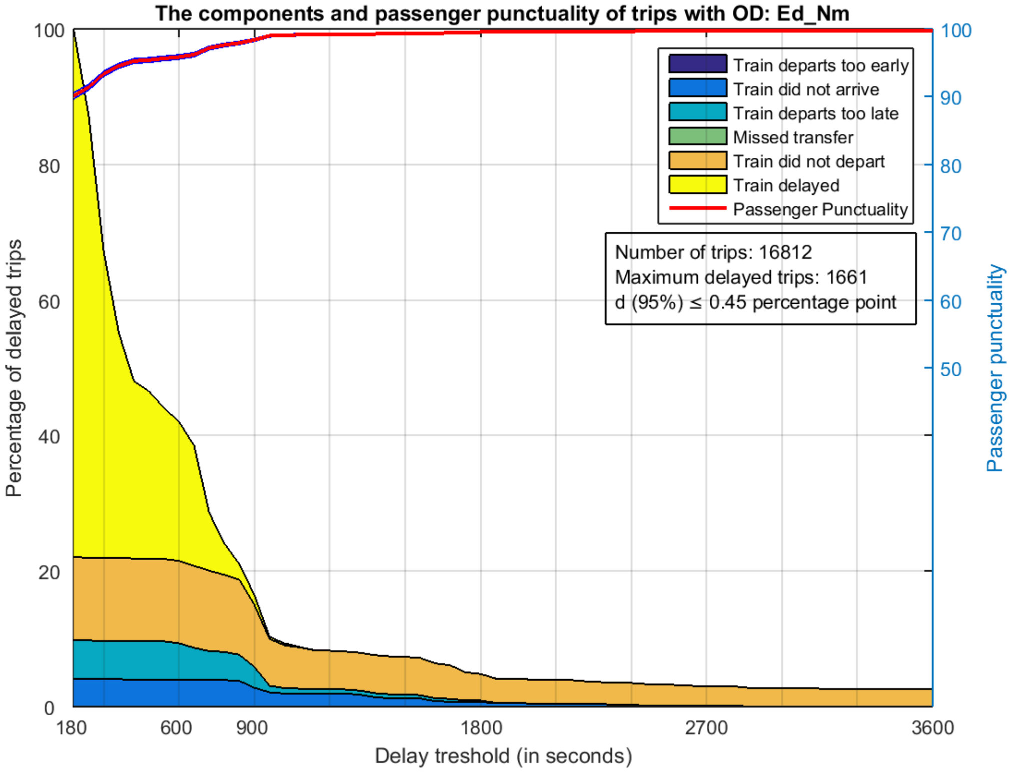


Figure 4.3: The breakdown graph for trips between Ede-Wageningen and Nijmegen

indicates the impact of the components and where the performance is lagging behind. Nonetheless the precise problems in the network are hard to determine: the graphs are too aggregated. Creating a breakdown graph dedicated for a station, origin-destination combination or train series, can give a better insight in the performance issues. To create such graphs, the algorithm must be slightly adjusted. The precise adjustments can be found in appendix B. The largest difference is the need to calculate the confidence interval when analysing the breakdown graphs for long term problems.

Creating a breakdown graph for, for example, an origin-destination combination, cannot be done without specifying a time period. The data used is the available data in that time period. The large drawback is that the breakdown graph provides only information about that time period. Even if the time period contains all the data available, the breakdown graph does not necessary predict the future, not to mention the computation time needed when using all data available. Therefore the data available in the time period is considered a sample of all data. This implies that calculated passenger punctuality is not necessarily the correct value. To determine how accurate the calculation is, the maximum deviation of the passenger punctuality should be calculated. The following formula is used [33].

$$N \geq \frac{Z^2}{d^2} * p(1-p) \tag{4.1}$$

This can be rewritten in:

$$d \leq \sqrt{\frac{Z^2}{n} * p(1-p)} \tag{4.2}$$

Here d is the maximum deviation, n the number of trips and p the expected fraction. Z depends on the desired reliability. Here the desired reliability is 95%. This means that Z is equal to 1.96. The expected fraction will be the calculated passenger punctuality based on the sample size. The hypothesis is that this calculated passenger punctuality is representative for the actual passenger punctuality. The found d is the maximum deviation where this is true with a 95% certainty. The larger the deviation, the larger the actual passenger punctuality can differ from the calculated passenger punctuality.

Note that the expected fraction p changes per delay threshold. That means that the maximum deviation changes. Plotting the passenger punctuality plus and minus the deviation will result in a 95% confidence interval. An example can be found in Figure 4.3. Clearly visible are the blue lines above and under the passenger punctuality plot. The lines represent the borders of the confidence interval. If the delay threshold increases, the interval will become smaller. This can be explained via equation 4.2. While Z and n are constant, p changes by a different delay threshold. If $p(1-p)$ increases, the deviation increases. $p(1-p)$ is 0 when p = 0 or 1 and the maximum is 0.25 when p = 0.5. So the deviation is the largest by a passenger punctuality that is the closest to 50%. The given deviation in the breakdown graph is the deviation by a delay threshold of 3 minutes.

This paragraph will show the parts of the network where the share of each component is the worst for both the 3 minute and the 15 minute passenger punctuality. For the components 'Train departs too early', 'Train departs too late' and 'Train did not depart' the trips are filtered by origin station. For the component 'Train did not arrive' and 'Train delayed' the trips are filtered by destination station. Filtering the trips by transfer station, will lead to the station where the share of 'Missed transfer' is the largest. The used data set is also filtered by origin-destination combination and train series. Other types of filtering are also possible, like time of departure, check-in time and train numbers as well as combinations of those. This paragraph will end with a comparison of the components. The part of the network is found where the size of the components is the largest, for a delay threshold of both 3 and 15 minutes.

4.3.1. Breakdown per origin station

In comparison with all stations, station Duivendrecht has the largest share of trips departing with the component 'Train departs too early' both for the 3 minute and the 15 minute passenger punctuality, see Figure 4.4. For the 3 minute passenger punctuality the share of trips with this component is 0.46%. Remarkable is the fact that for 68.36% of all trips departing here the realised departure time is exactly on the minute. For only trips with the component 'Train departs too early' this is 95.96%. Analysing the situation shows that for some train series the realised departure time is not measured at all. It is an extrapolation of the departure times or passing times of service control points close by. (A service

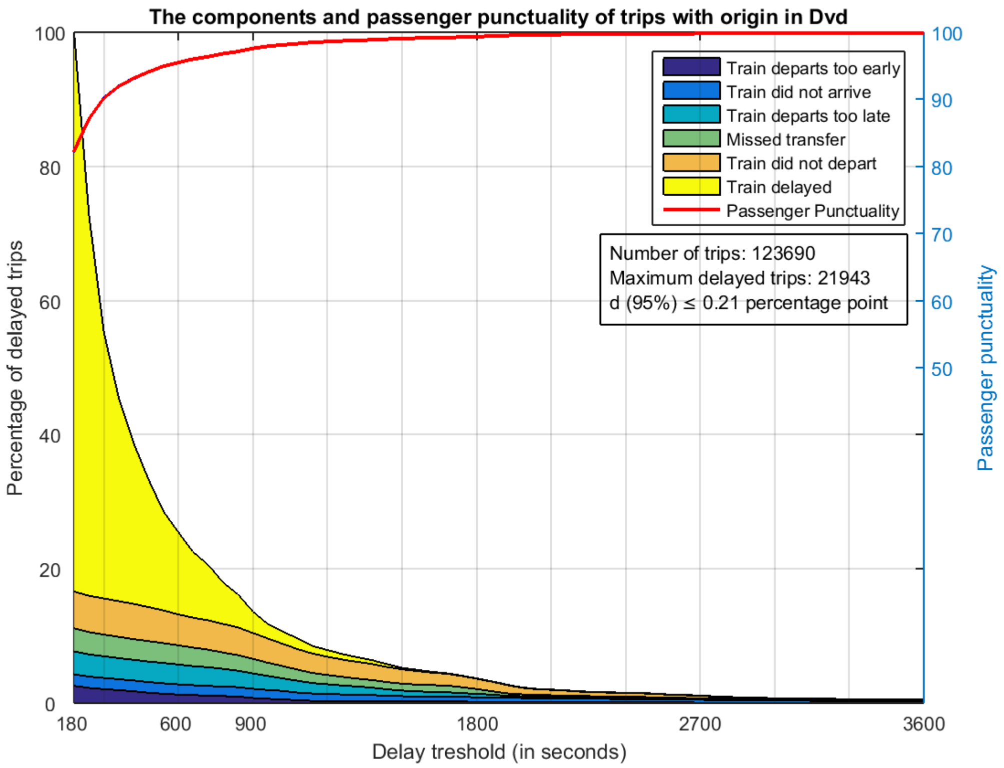


Figure 4.4: The breakdown graph for trips that originated from station Duivendrecht

control point is an important geographical point in the rail network, like a station.) Most likely this is the cause for the share trips with component 'Train departs too early'. Note that Trento does have a better estimation of the realised departure time (see also paragraph 3.2.9).

The station with the largest share of departing trips with the component 'Train departs too late' is station Hoek van Holland Strand. The percentage of trips with this component is by the 3 minute delay threshold almost equal to the percentage of trips by the 15 minute passenger punctuality, around 3.40% (see Figure 4.5). Hoek van Holland Strand is the start station of train series 4100 and no other train series runs here. The expectation is that the share of trips with the component 'Train departs too late' is almost equally high on the next station on the route: Hoek van Holland Haven. However the share by the 15 minute delay threshold is only 0.20%. This can be explained if analysing the situation in more depth. In the examined period only one train had a delay of at least 15 minutes. The number of passenger departing in this train at Hoek van Holland Strand was much larger than on average at this station, so the impact of this one train on the share of the components was extensive.

A remark is that in this analysis only the components of departing trips are taken into account. It could occur that a train departing from a transfer station has a delay of at least 15 minutes. This implies that the calculated share of the component per origin station is not per definition the share of the number of trips departing from that station with a delay of at least 15 minutes. On the other hand, 75.01% of the trips does not have a transfer. So the calculated share is representative for the actual share.

Ordering the origin stations by the share of trips with the component 'Train did not depart' does not give any striking outliers. In other words, there are no departing stations where the percentage of trips with this component is noteworthy. Nonetheless the station with the largest shares, both for the 3 minute (4.13%) and the 15 minute passenger punctuality (2.94%), is station Rotterdam Centraal, see Figure 4.6. For this station, only 2 train series were responsible for 48.23% of all cancelled trains. These were train series 900 and 2800. A trip can get the component 'Train did not depart', because a train was

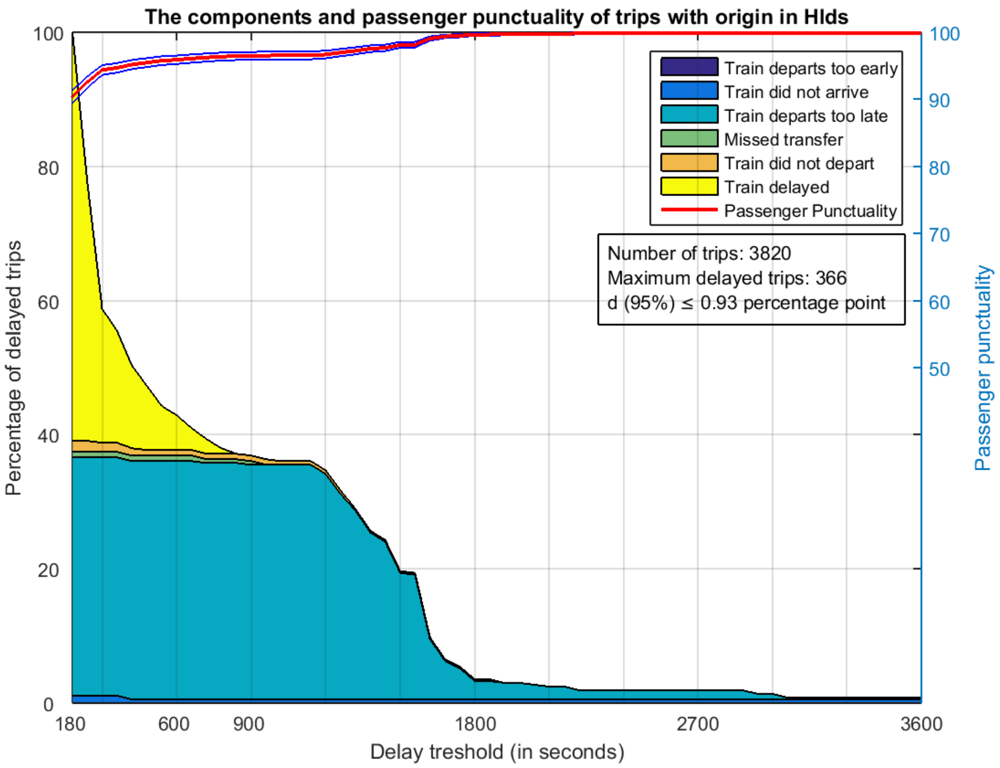


Figure 4.5: The breakdown graph for trips that originated from station Hoek van Holland Strand

cancelled at a transfer station and not at the origin station. This is just like by the component 'Train departs too late'. However for Rotterdam Centraal 94.95% of all trips with this component originated here.

Besides the large share of trips with the component 'Train did not depart' at Rotterdam, on an overall basis this origin station has the worst 15 minutes passenger punctuality: 94.80% with a maximum deviation of 0.05 percentage point by a 95% reliability. Station Amsterdam Lelylaan scores the worst for the 3 minute passenger punctuality. With a 95% certainty the passenger punctuality is within 0.24 percentage point of 74.45%. This means that, if departing in Amsterdam Lelylaan, the chance to arrive with a delay less than 3 minutes is 74.45%.

4.3.2. Breakdown per destination station

The destination station with the largest share of trips with the component 'Train did not arrive' for the 3 minute passenger punctuality is station Rotterdam Centraal as well. 1.56% off al trips with destination Rotterdam Centraal did not arrive on time because a train was cancelled before arriving in Rotterdam or before arriving at a transfer station. Nonetheless 72.53% of the trips arriving in Rotterdam Centraal and with the component 'Train did not arrive' encountered a cancelled train of train series 1700 and 11700. This is remarkable because those trains do not have a planned arrival in Rotterdam Centraal. The reason is as follows. These trains are part of a longer train from a different train series that get split up in stations Utrecht Centraal or Gouda. After those stations the part of the train with destination Rotterdam Centraal will get a new train number. This implies that passengers departing before Utrecht or Gouda will depart a train that changes the train number before arriving in Rotterdam Centraal. The algorithm that calculates the passenger punctuality cannot handle those changes in train number. The result is that those corresponding trips are categorized as 'Train did not arrive'.

This effect is visible if examining the destination stations by their share of trips with the component 'Train did not arrive' by the 15 minute delay threshold. Rotterdam Centraal is not in the top 10 of stations with the largest shares anymore. The destination station with the largest share is station

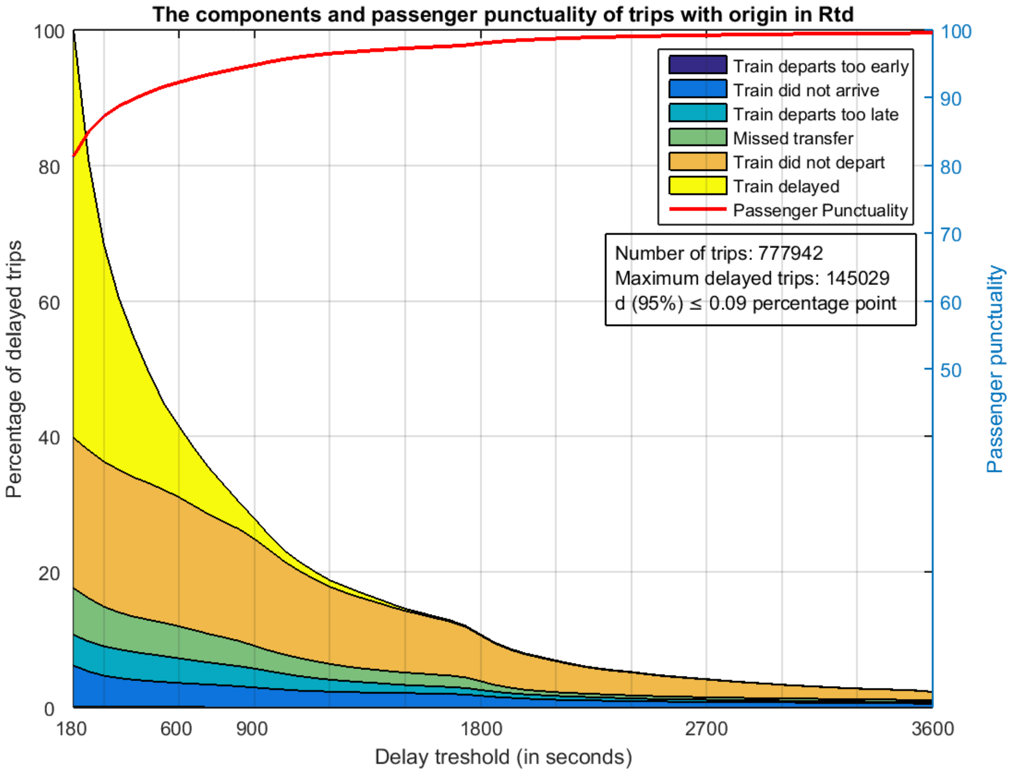


Figure 4.6: The breakdown graph for trips that originated from station Rotterdam Centraal

Arnhem Zuid, see Figure 4.7. 1.15% of all passengers with this station as destination does not arrive within 15 minute of their planned arrival time because a train is cancelled midway. Remarkable is the fact that the greater part of the trips where the train gets cancelled is of a train series that does not stop in Arnhem Zuid. This implies that of those cancelled trains the number of passengers with destination Arnhem Zuid was above average.

The first nine stations, where the share of trips with the component 'Train delayed' is the highest by the 3 minute passenger punctuality, are short stops. Of those nine stations, station Hoofddorp has the highest share (28.12%), see Figure 4.8. As explained in paragraph 3.2.9 the determination of the arrival times is not very accurate, certainly on short stops. By a small delay of only 1 or 2 minutes in reality, the registered delay can easily be larger than 3 minute. As suggested before, using Trento to determine the arrival times might be a solution for this phenomenon. A remark is that Hoofddorp is not always used as a shortstop. So not all delayed trips can be explained via the assumptions in the passenger punctuality calculations.

For the delay threshold of 15 minute station Den Helder Zuid has the largest share of arriving trips with the component 'Train delayed': 2.46%. This station is a short stop as well but the problems with the inaccurate measurements of the arrival times do not directly apply to larger delays. Also the second station with the highest share is station Den Helder, close to Den Helder Zuid and connected by the same train series. Den Helder is not a short stop. This implies that the trips with the component 'Train delayed' at station Den Helder Zuid is indeed caused by delayed trains and not due to other circumstances.

Hoofddorp is also the destination station with the worst 3 minute passenger punctuality of all stations. If a trip is made and the destination is Hoofddorp, the chance that the final delay is within 3 minutes is 68.02%, with a maximum deviation of 0.24 percentage point and with 95% certainty. For the 15 minute passenger punctuality the worst destination station is Rotterdam Centraal. The expectation is that 95.13% of all the trips with destination Rotterdam Centraal arrive within 15 minutes of the promised arrival time.

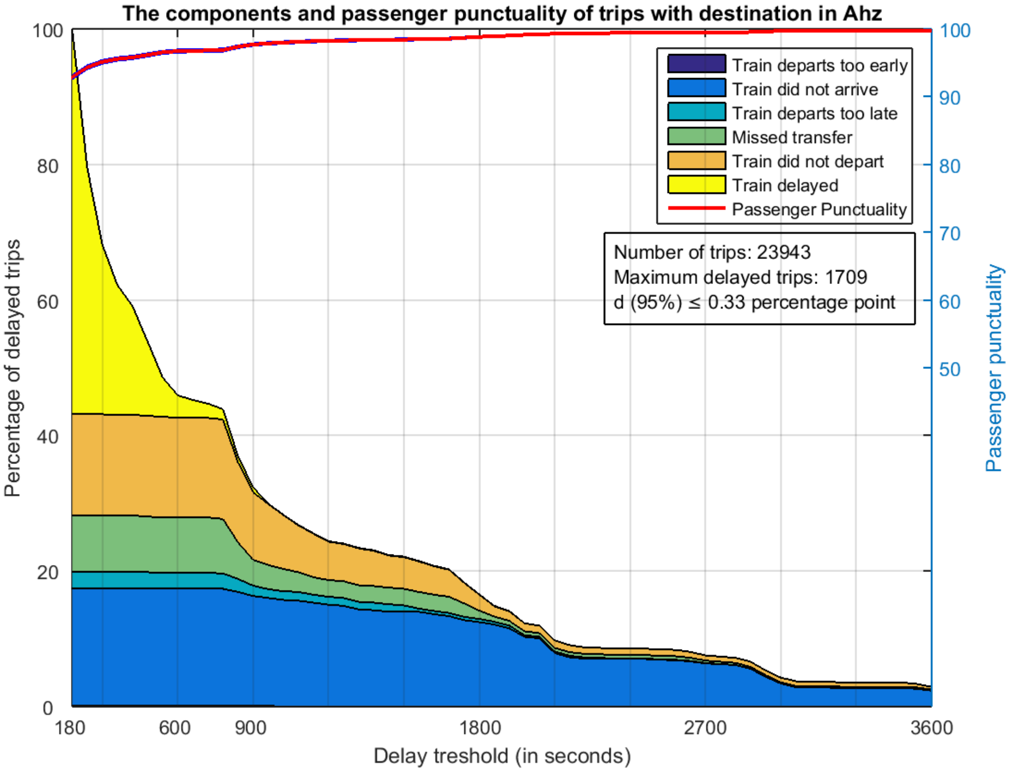


Figure 4.7: The breakdown graph for trips that have station Arnhem Zuid as destination

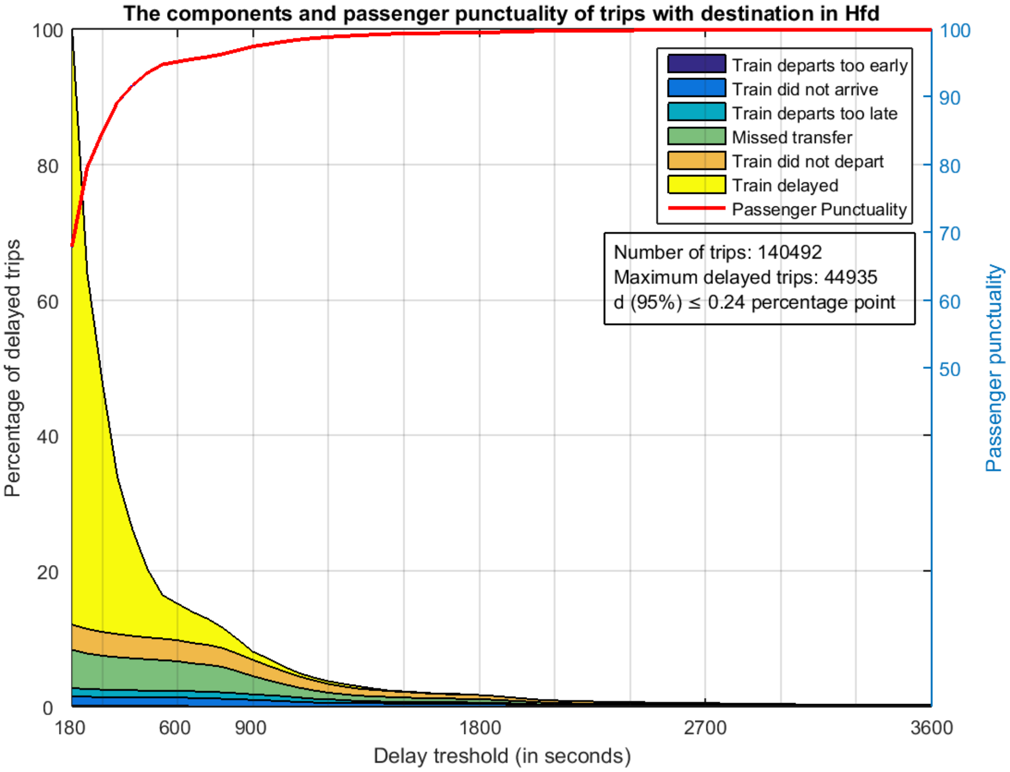


Figure 4.8: The breakdown graph for trips that have station Hoofddorp as destination

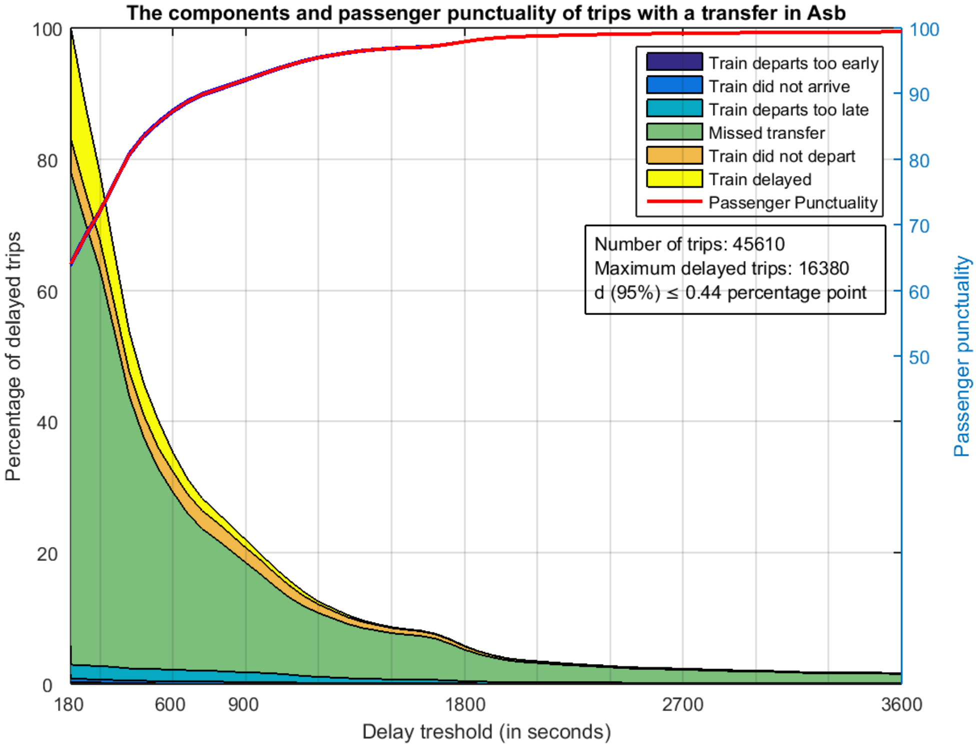


Figure 4.9: The breakdown graph for trips that have a transfer at station Amsterdam Bijlmer Arena

4.3.3. Breakdown per transfer station

Filtering the data by transfer station offers the opportunity to see at which stations the most transfers are missed. By the 3 minute passenger punctuality there are three stations where the share of trips where with the component 'Missed transfer' is remarkably high. One of those is station Bilthoven but here the number of passengers transferring was only 8. This is too low to draw any conclusions. The other two stations were stations Amsterdam Bijlmer Arena (26.96%, see Figure 4.9) and Lelystad Centrum (21.61%, see Figure 4.10). These high percentages can partly be explained via the used walking times on the stations. Both stations do not have platform specific walking times but only a station specific walking time. This station specific walking time is always the highest possible walking time. Shorter transfers are therefore categorized as 'missed' while there is a large chance that they are made. The delay threshold of 3 minutes is too low to filter these mistakes out of the delayed trips.

Nonetheless, the station specific walking time is not the whole explanation why the shares on these stations are higher than on the other transfer stations. By a delay threshold of 15 minutes these two stations have the highest share of trips with the component 'Missed transfer' as well. For station Amsterdam Bijlmer Arena this share is 6.01% and for station Lelystad Centrum this share is 7.05%.

4.3.4. Breakdown per origin-destination combination

For the analysis of the breakdown of the passenger punctuality per origin-destination (OD) combination the 200 ODs with the most passengers are used. These 200 ODs corresponds with 33.62% of all the trips in the network. Note that only trips with the exact origin and destination are used. Trips that are (partly) conducted on these ODs are not used.

The OD with the worst 3 minute passenger punctuality is Amsterdam Centraal – Hoofddorp (see Figure 4.11). 58.31% of the trips arrived within 3 minutes of the promised arrival time. This poor score is mostly due to the final delay of the last train because the largest component is 'Train delayed'. As already shown, station Hoofddorp is the destination station with the worst 3 minute passenger punctuality.

The OD with the second worst 3 minute passenger punctuality is Den Haag Centraal – Amsterdam.

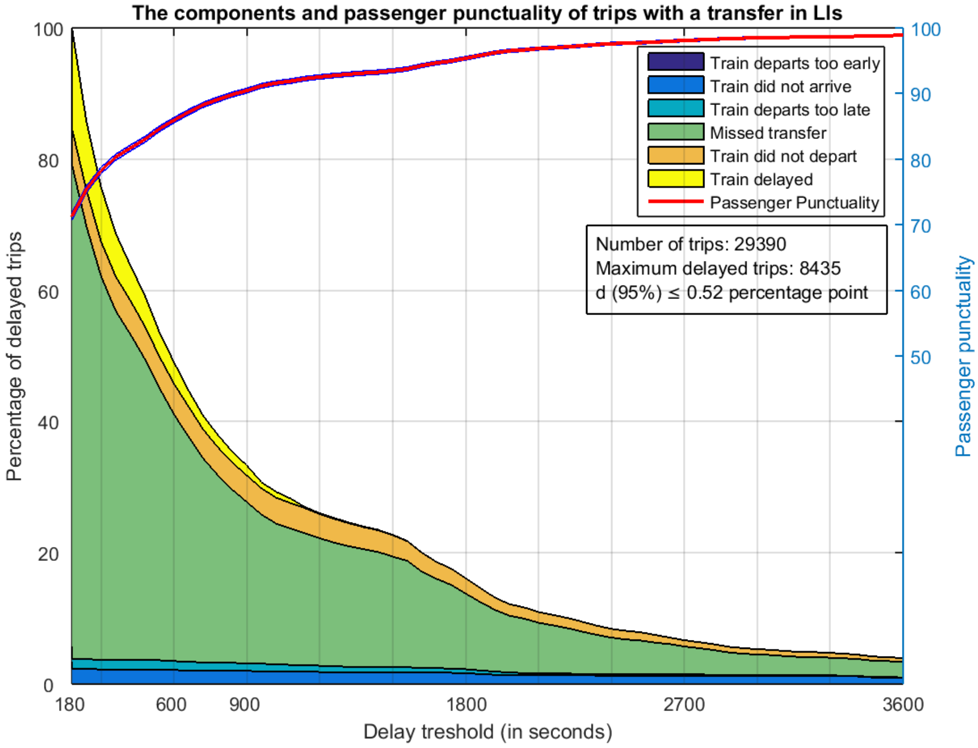


Figure 4.10: The breakdown graph for trips that have a transfer at station Lelystad Centrum

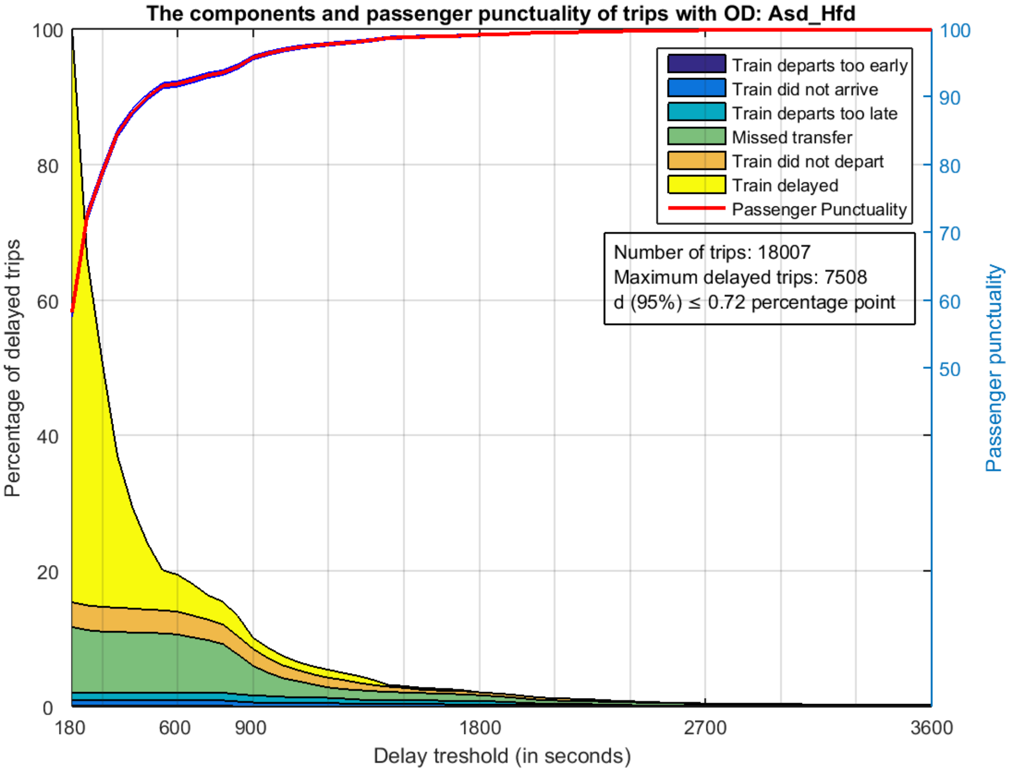


Figure 4.11: The breakdown graph for trips between Amsterdam Centraal and Hoofddorp

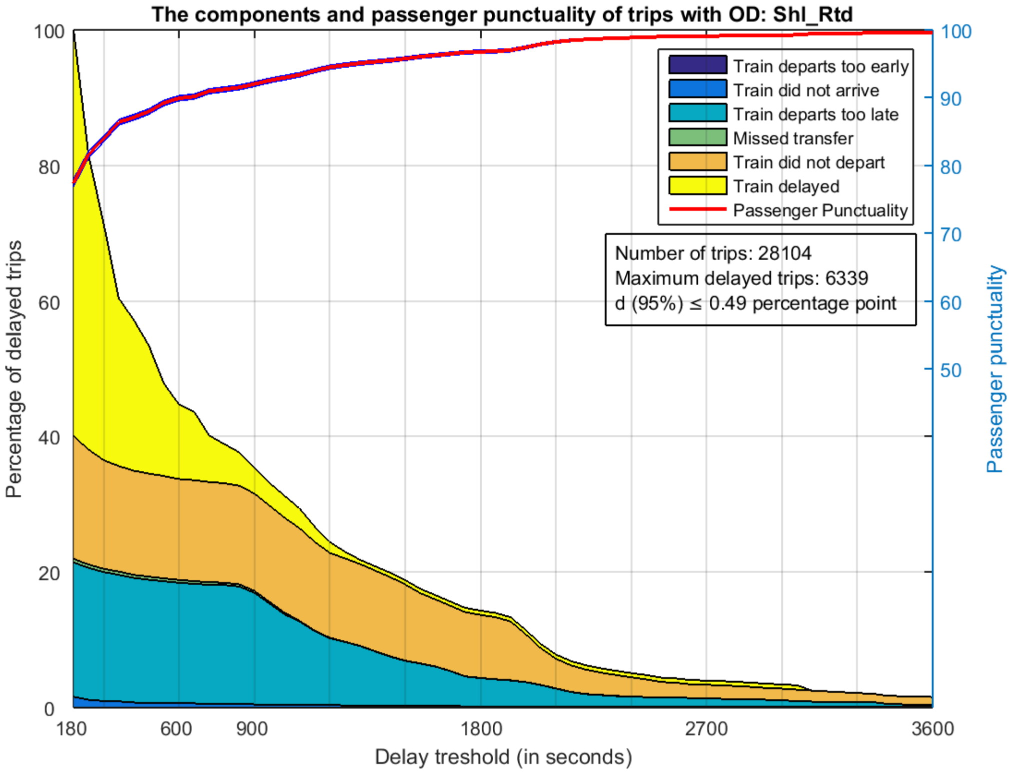


Figure 4.12: The breakdown graph for trips between Schiphol and Rotterdam Centraal

Here the 3 minute passenger punctuality is 63.87%. This OD has almost 2.5 times more passengers than there are between Amsterdam Centraal and Hoofddorp. The two main components are 'Train delayed' and 'Missed transfer'.

Based on the share of trips with the component 'Train departs too late' by the 3 minute passenger punctuality, the OD Schiphol – Rotterdam is an outlier (see Figure 4.12). Here 4.47% of the delayed trips is because the train had an initial delay of at least 15 minutes at Schiphol. Filtering these trips by train series gives a remarkable result: 91.99% of the trips was made with train series 900. The share of trips with the component 'Train did not depart' is also extensive (77.34%), although not a striking outlier. Moreover the OD with the largest share of trips with this component by the 3 minute passenger punctuality is Rotterdam Centraal – Breda (8.48%). Here 89.47% of the trips, that was delayed because the train did not depart, was caused by train series 900.

The OD Utrecht Centraal – Leiden is also a special case. Here the share of the trips with the component 'Missed transfer' is the largest of all ODs: 11.04%. This OD is already examined in paragraph 3.2.5. The conclusion was drawn that most passengers choose to travel via a different travel option than the promised journey. This alternative option has no transfers but arrives 5 minutes later in Leiden. The behaviour is visible in Figure 4.13. The share of trips with a missed transfer is extensive for a delay threshold of 3 minutes. However this share decreases rapidly after the delay threshold of 5 minutes: the time period after which the alternative train arrives in Leiden.

The most striking ODs if focussing on the 15 minutes passenger punctuality are again Schiphol – Rotterdam Centraal and Rotterdam Centraal – Breda (see Figure 4.14). This is not unlikely because the possible alternatives are limited, certainly when the promised train did not depart at all. In both cases the first alternative travel option arrives more than 15 minutes later than the promised train. This is visible is the breakdown graph of Rotterdam Centraal - Breda. The share of trips with the component 'Train did not depart' stays almost the same between the 3 minute and 15 minute delay threshold. Thereafter the share decreases rapidly.

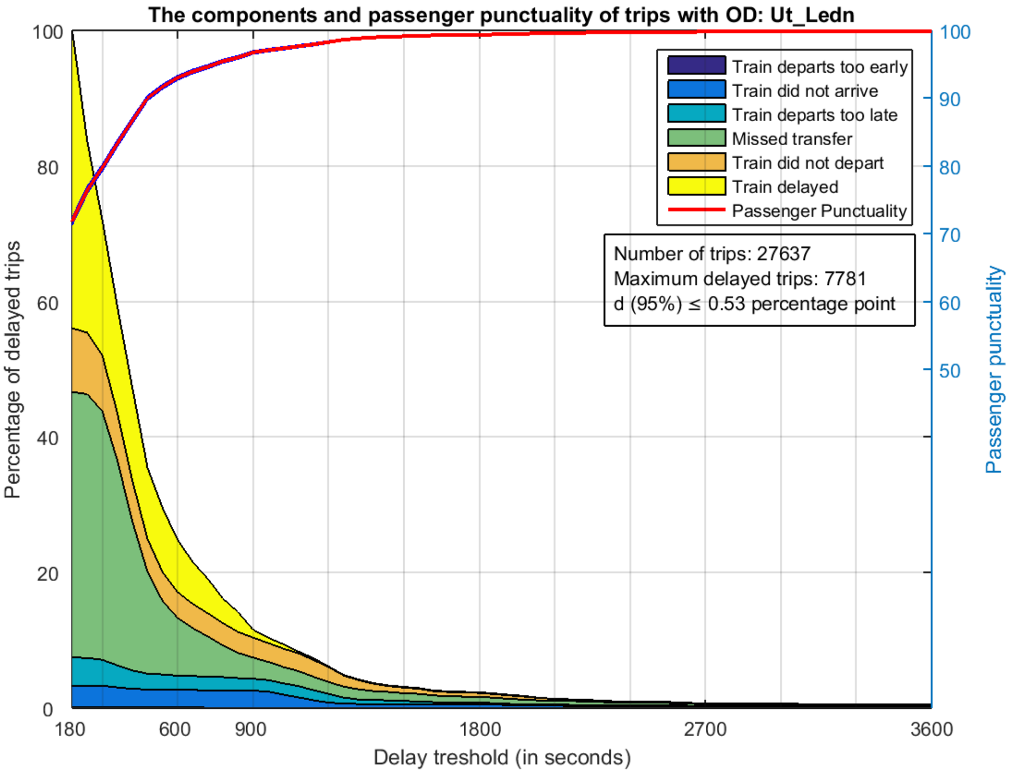


Figure 4.13: The breakdown graph for trips between Utrecht Centraal and Leiden

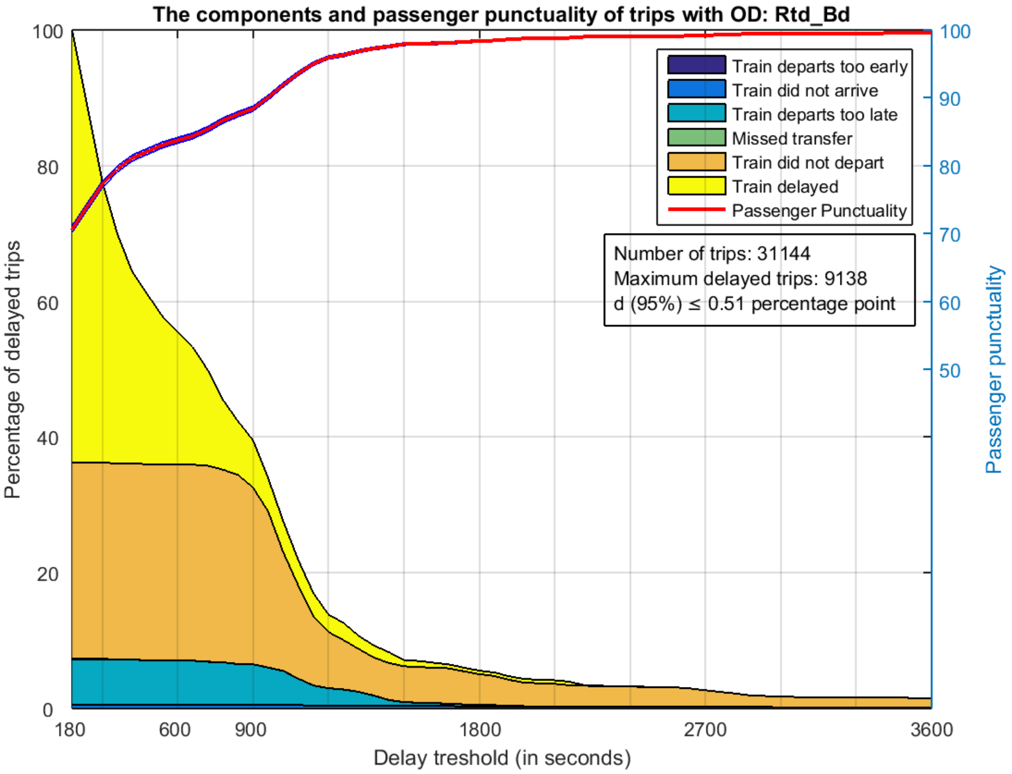


Figure 4.14: The breakdown graph for trips between Rotterdam Centraal and Breda

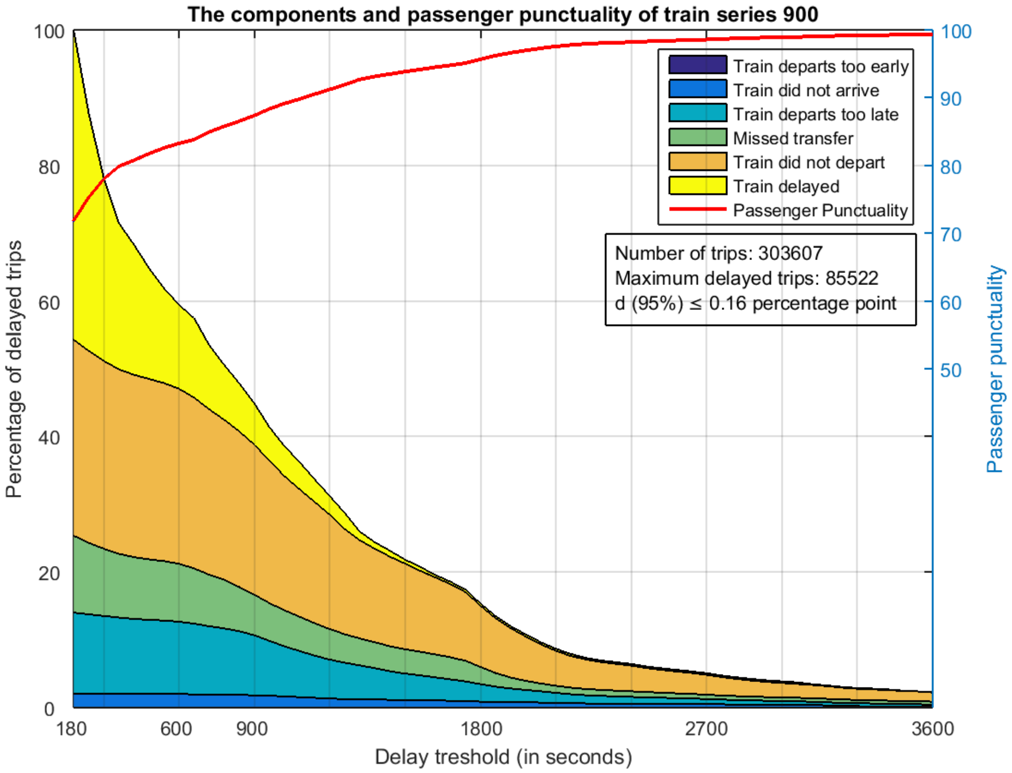


Figure 4.15: The breakdown graph for trips with train series 900

4.3.5. Breakdown per train series

Every promised journey, and thus every trip, can be combined with a train series. Filtering the trips by these train series leads to a breakdown per train series. If a promised journey includes a transfer two possible train series can be combined to the corresponding trip. If the trip can be made as promised, the used train series is the train series of the last train. If a trip needs to be rescheduled, the train that is the cause of the rescheduling is used to combine the trip to a train series. Note that if a transfer is missed, the first train series is used. Some train series are barely used, so only train series with a total share of at least 1% of the trips are used. This results in 47 different train series, which corresponds with 93.32% of all the trips.

Not surprisingly, train series 900 (between Amsterdam Centraal and Breda) has the worst 3 minute and 15 minute passenger punctuality: 71.83% and 87.36% respectively (See Figure 4.15). This is mostly due to trips with the component 'Train did not depart'. The second and third place on both delay thresholds are for the train series 2000 and 2800. Train series 2000 runs between stations Utrecht Centraal and Den Haag Centraal and train series 2800 runs between stations Utrecht Centraal and Rotterdam Centraal. The 3 minute passenger punctuality is 76.70% and 73.53% respectively. The largest group of delayed trips of these train series are trips with the component 'Train did not depart'.

Train series 3500 has the most trips with the component 'Missed transfer'. 34.58% of those trips has a planned transfer on Train series 2600 at Schiphol. Herewith this is the transfer that is most often missed. Here the planned transfer time is 4 minutes. With a walking time of 3 minutes this is a narrow transfer. The arriving train needs only a delay of only 1 minute and the transfer is considered missed.

4.3.6. Comparison of components

As stated, to check where the most passengers are delayed because their train departed too early or too late or where their train did not depart at all, a breakdown per origin station should be performed. The number of passengers per location with a delay because the train was delayed or the train did not arrive can be determined by a breakdown per arrival station. Finally, via a breakdown per transfer station, the number of passenger per location with a missed transfer can be determined. Combining these results can give an overview of the locations where the most delays occur. Also different causes, like missed transfers or cancelled trains, can easily be compared.

For the first five days of June 2015 such an analysis is made. In Table 4.1 the ten largest components by the 3 minute passenger punctuality are shown. Also the percentage of the passenger with this component at that destination, arrival or transfer station is shown. As can be seen, the main component of why a passenger has a delay of 3 minutes or more is because the train was delayed.

Table 4.1: The top 10 of largest components by a delay of 3 minutes or more

Station	Component	Number of passengers	Percentage
Amsterdam Centraal	Train delayed	61132	19.49%
Utrecht Centraal	Train delayed	49692	12.73%
Schiphol	Train delayed	32586	29.70%
Leiden Centraal	Train delayed	31474	20.40%
Den Haag Centraal	Train delayed	26633	14.92%
Rotterdam Centraal	Train delayed	25875	14.32%
Eindhoven	Train delayed	19412	15.32%
Utrecht Centraal	Missed transfer	16991	6.67%
Amsterdam Sloterdijk	Train delayed	16932	15.03%
's-Hertogenbosch	Train delayed	16072	17.22%

More interesting is the top 10 largest components by a delay of 15 minutes or more. This list can be found in Table 4.2. Note the fact that 6 of the 10 components are occurring at only 2 stations: Utrecht Centraal and Amsterdam Centraal, the two busiest station of the rail network.

Table 4.2: The top 10 of largest components by a delay of 15 minutes or more

Station	Component	Number of passengers	Percentage
Utrecht Centraal	Missed transfer	10378	4.07%
Utrecht Centraal	Train did not depart	7714	1.96%
Amsterdam Centraal	Train did not depart	5197	1.66%
Rotterdam Centraal	Train did not depart	4875	2.72%
Utrecht Centraal	Train departs too late	4038	1.03%
Amsterdam Centraal	Train departs too late	4035	1.30%
Den Haag Centraal	Train did not depart	3604	1.99%
Schiphol	Missed transfer	3409	4.99%
Leiden Centraal	Missed transfer	3290	3.55%
Amsterdam Centraal	Missed transfer	2851	3.72%

4.4. Conclusion

In this chapter the six different components were introduced. A component is a reason why a passenger is delayed, from the perspective of that passenger. If a passenger is delayed, the component is the same as the category of the trip. An exception is the component 'Train delayed'. This component consists of all delayed trips with the category 'Trip realised'.

If analysing the passenger punctuality and the components, two delay thresholds are important: 3 and 15 minutes. In the official KPI the threshold is 3 minutes, so analysing this passenger punctuality might lead to an improvement of the official KPI. The rating of the NS drops by a delay of 15 minutes. This is the main reason to improve the 15 minute passenger punctuality. On the other hand, visualising the passenger punctuality and the size of the components for an interval gives insight in some underlying dynamics.

Filtering the trips by origin, destination, transfer station, train series or OD can provide the basics for the breakdowns of the passenger punctuality for subsets of the rail network. This is done in paragraph 4.3. Analysing the outliers, thus parts of the network where the size of a components is much larger than on average, lead most often to drawbacks of the algorithm that calculates passenger punctuality than to actual performance problems in the rail network. An example is the stations where the size of the component 'Train delayed' by the 3 minute passenger punctuality is the largest. These stations are all short stops. The actual arrival and departure times on these stations differ from the arrival and departure times used in the passenger punctuality calculation.

There is one remarkable drawback of the passenger punctuality calculation, found during the breakdown analyses: the algorithm cannot handle trains that have multiple train numbers during a single run. This will lead to the wrong categorisation of trips. Not only has this an impact on the passenger punctuality, by small delay thresholds this will result in more trips with the component 'Train did not arrive' than actually the case. This drawback is remarkable because it is not mentioned in the documentation of the algorithm [24] or in the assumptions (paragraph 3.2).

A last remark to the breakdown of the passenger punctuality is that the found results should always require an explanation. The best example is the share of trips with the component 'Trains depart too late', departing at station Hoek van Holland Strand. This share is the highest of all stations in the rail network. Although the data of almost a month used, the size of the share was mostly due to only 1 train.

5

Models to describe passenger punctuality

Chapter 3 describes how the passenger punctuality is calculated. Based on the analyses of the assumptions, the calculation can be improved in order to achieve a more reliable KPI. Although a more reliable KPI would improve the passenger punctuality, it does not necessarily increase the passenger punctuality. Nonetheless this is a goal of ProRail and the NS (see chapter 2). Higher passenger punctuality implies that the performance of ProRail and the NS increases. In paragraph 5.1 two models are discussed. These models are created to provide insight in the processes related to the passenger punctuality.

The focus of the two models is the Transport control and Traffic control. Paragraph 5.2 analyses the underlying processes through process indicators. The indicators give information about the performance of the processes. Both models suggest that if the performance of the processes increases, eventually the passenger punctuality will increase. In paragraph 5.3 the correlation in both models is calculated. This will give information whether steering on the process indicators will increase the passenger punctuality or not. Finally in paragraph 5.4 the results of the two models are compared and some conclusion can be made.

5.1. Introduction of the two models

In order to increase passenger punctuality ProRail and the NS created a model that describes passenger punctuality. In this model the passenger punctuality is related to the delayed and cancelled trains (see paragraph 5.1.1). The processes of the Transport control and Traffic control should minimise the number of delayed and cancelled trains. This will than result in higher passenger punctuality. In paragraph 5.1.2 a second model is proposed. Here the delayed and cancelled trains are not used, mainly because it is unknown how many passengers every train contains. Instead the passenger punctuality is divided into the components (see chapter 4). The components are roughly the percentage of delayed trips because a train is delayed or cancelled.

5.1.1. Model 1: The model of ProRail and the NS

ProRail and the NS have developed a model where the passenger punctuality is related to underlying processes in separate units within ProRail and the NS (see Figure 5.1). This model can be used to describe passenger punctuality. In the top level, the management steers on passenger punctuality. A higher KPI means an improved performance of ProRail and the NS. To achieve this, on a lower level every involved business unit steers on 'Explainable Train Deviations' (in Dutch: Te Verklaren Trein Afwijking, TVTA). The model anticipate on the idea that these TVTAs are the main cause for the dispunctuality: by decreasing the number of TVTAs, the punctuality will increase.

Another level lower, the individual units corresponding with the business unit steer on the process indicators. If the performance increases, the number of TVTAs occurring will decrease and thus the passenger punctuality will become higher. An elaboration of the process indicators belonging to the business unit Transport control and Traffic control (TB / VL) can be found in paragraph 5.2. The

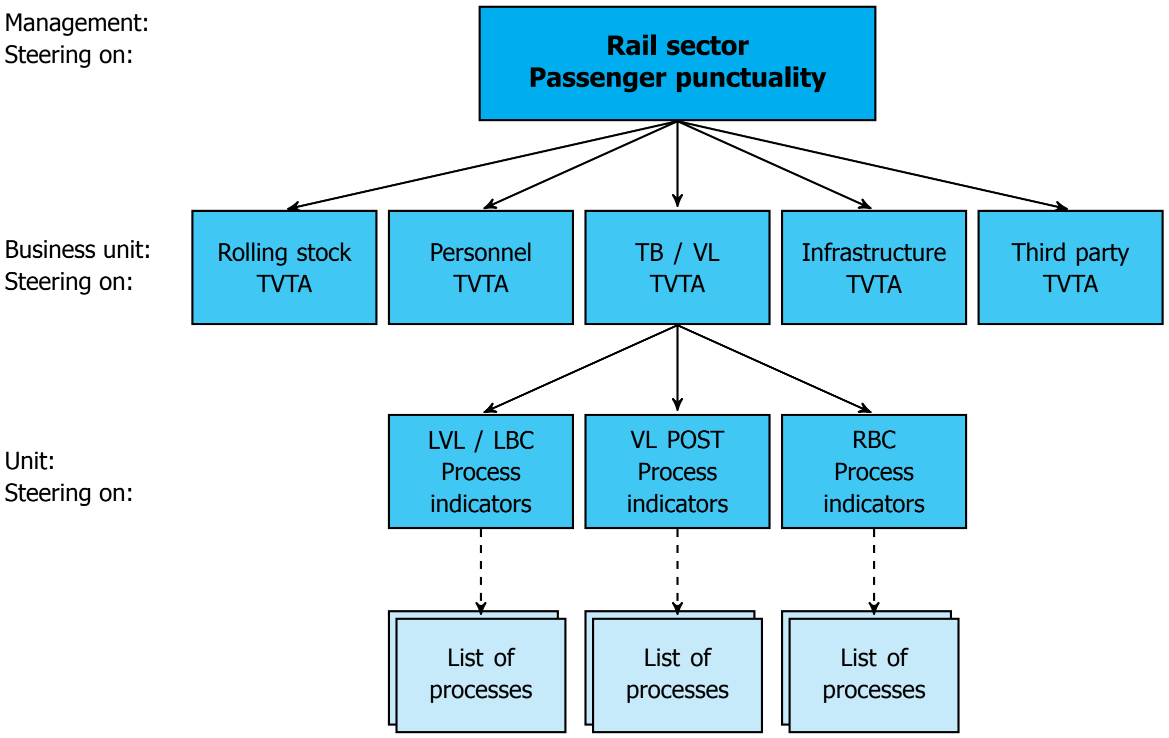


Figure 5.1: A scheme of the model of ProRail and the NS to describe passenger punctuality

correlation between the process indicators and the number of TVTAs can be found in paragraph 5.3. As stated, the first step in the model of ProRail and the NS is to translate passenger punctuality into 'Explainable Train Deviations' [34]. This paragraph will analyse these TVTAs. This is done firstly by an explanation on what a TVTA is, which TVTAs can be distinguished and how the TVTAs relate to the model of ProRail and the NS. Secondly some remarks to the use of TVTAs in the model of ProRail and the NS are given.

Explanation of the TVTAs

A train deviation is a deviation in the execution of a train according to the original traffic plan at a service control point (in Dutch: Dienstregelpunt) [35]. A traffic plan contains at least the planned trains between service control points and is used for the national network control. Note that a train deviation is not necessary a TVTA. For 4 categories of train deviations, the train deviation should be explained. In these cases, the train deviation becomes an 'Explainable Train Deviation', or a TVTA [36]. The four categories are listed below with the percentage of TVTAs belonging to this category for the first 6 months of 2015.

1. Cancelled train (16.83%)
2. Redirected train (0.99%)
3. Delayed train (73.42%)
4. Realignment of a train in time (8.76%)

The definition used for a cancelled train is a train that is cancelled for the whole or a part of the planned route [35]. A redirected train is a train that is directed via one or more different service control points than planned in the original traffic plan. A train is delayed if

- either the time difference between plan and reality at a service control point is at least 3 minutes (so the train is passing a service control point at least 3 minutes later than planned) and this time difference is at least 3 minutes larger than the time difference on the last passed service control point
- or the time difference between plan and reality by departure at a service control points is at least 3 minutes and this time difference is at least 3 minutes larger than the time difference by arrival at the same service control point.

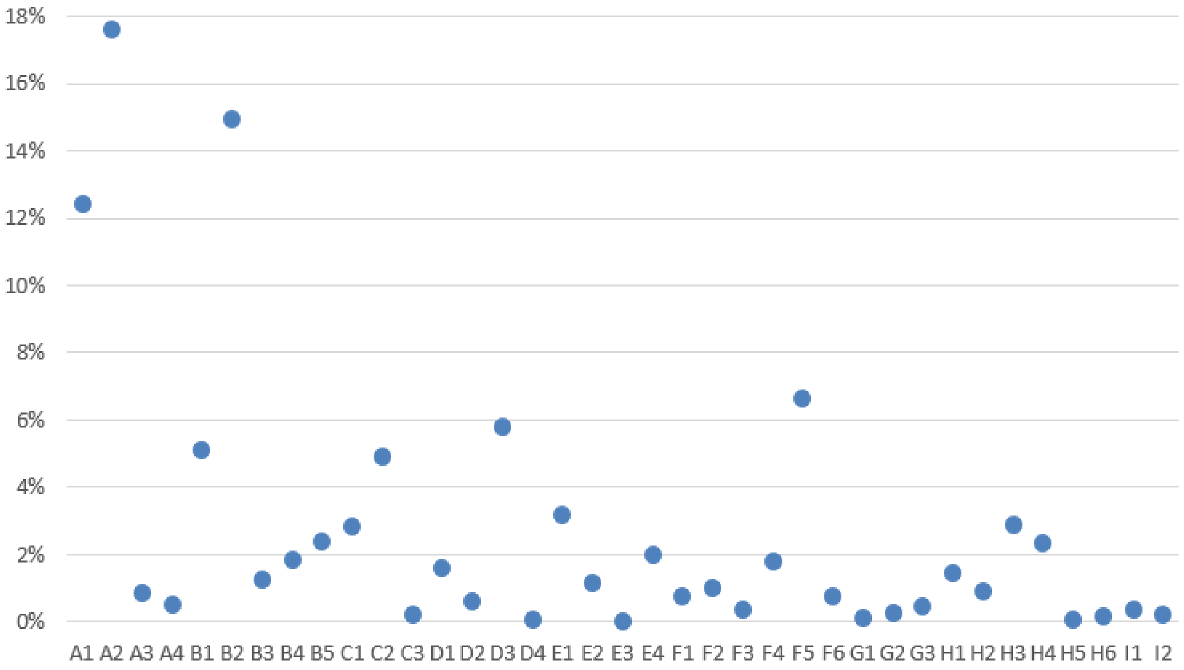


Figure 5.2: The probability of occurrence for each type of TVTA

It can occur that the planned arrival, departure and passage times are adjusted for the whole route but the route itself does not change. This is a realignment of a train in time. This rarely happens for passenger trains. However it is relatively common for freight trains. If there is a train deviation belonging to one of those four categories, a signaller of ProRail couples this to a type of TVTA. Basically this implies that the cause of the train deviation is registered. To make the possible causes more organized a list has been composed with 37 possible types of causes the signaller can choose from. This list can be found in appendix C. Here every type of TVTA is having a code (The first letter corresponds with the type category of the TVTA). These codes are used the remainder of this chapter for identifying a type of TVTA. In Figure 5.2 for each type of TVTA the probability of occurrence is visualized, based on the first 6 months of 2015.

The model of ProRail and the NS divides the TVTAs among the business units responsible for the TVTAs. Note that these responsible business units are in some cases different from the responsible parties for the category of TVTA (see appendix C). Five business units are defined (see Figure 5.1). First there is the 'Rolling Stock', responsible for all the rolling equipment. 'Personnel' includes all personnel in the trains and on the station. The Traffic control (VL) and Transport control (VB) are combined in 'TB / VL'. 'Infrastructure' is responsible for the infrastructure but also for the buildings and other civil objects. Lastly there is 'Third party'. The type of TVTAs corresponding with the business units are given in Table 5.1. In the third column the probability of occurrence of one of the TVTAs is given, based on the first 6 months of 2015. Note that not every TVTA has a responsible business unit. First there are three types of TVTAs that are caused by a disturbance somewhere else in the network. These types of TVTAs are not directly responsible for a train deviation. The second group of TVTAs have an external responsible party. Lastly there is one type of TVTA (D4) that was not assigned.

The difference between TVTAs corresponding to a third party and external TVTAs is that for the first group the effect can be influenced. Take for example TVTA H1. If there is a train deviation because of weather conditions, this TVTA is used. The weather itself cannot be influenced, but not only the weather itself determines the train deviations. By extreme weather measures are taken to minimise the effect of the weather on the performance on the rail network. So by optimising these measures the number of third party TVTAs can be minimised. This is in contrast to, for example, TVTA B5. ProRail or the NS cannot influence the situations on the rail network of Germany or Belgium. So if a train ends op with a delay abroad before arriving in the Netherlands, this is an external TVTA: even if the situation on the Dutch rail network is perfect, such a TVTA cannot be prevented.

Table 5.1: The business units and the corresponding type of TVTAs

Business unit	**Type of TVTA**	**Occurrence**
Rolling stock	D1, D2, D3	8.02%
Personnel	A4, B1, B3, B4, C2	13.65%
TB / VL	B2, C1, C3, E1, E2, E3, E4	24.36%
Infrastructure	F1, F2, F3, F4, F5, F6, G1, G2, G3	12.21%
Third party	H1, H2, H3, H4, H5, H6	7.83%
Indirectly	*A1, A2, A3*	*30.89%*
External	*B5, I1, I2*	*2.96%*
Not assigned	*D4*	*0.09%*

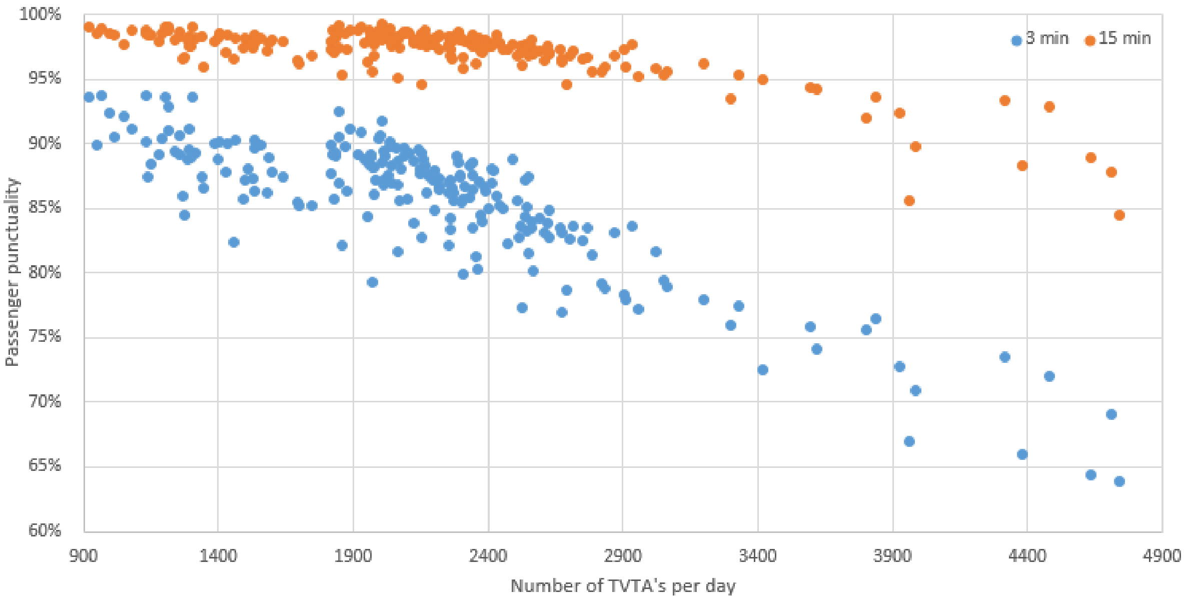


Figure 5.3: The number of TVTAs versus the passenger punctuality for 181 days

The model of ProRail and the NS assumes a clear correlation between the passenger punctuality and the number of TVTAs. The goal of each business unit is to decrease the number of TVTAs to increase the passenger punctuality. The relation between the passenger punctuality and the number of TVTAs can be represented via the Pearson correlation coefficient. This coefficient has a value between -1 and 1. If the coefficient has a value of -1, there is a perfect negative correlation. If the coefficient is 0, there is no correlation at all and if the coefficient is 1, there is total positive correlation.

With the data of the first seven months of 2015 the Pearson correlation coefficient between the 3 minute passenger punctuality and the number of TVTAs is determined (see Figure 5.3). This coefficient is -0.87. So there is indeed a strong relation between the number of TVTAs and the 3 minute passenger punctuality. The Pearson correlation coefficient between the 15 minute passenger punctuality and the number of TVTAs is -0.78. This means that there is still a correlation but less strong. Moreover, the fact that both calculated coefficients are negative can be explained by the fact that more TVTAs lead to more passenger delays which leads to a lower passenger punctuality.

Some remarks to the use of TVTAs

The model of ProRail and the NS relates the number of TVTAs to the passenger punctuality. This implies that to improve the passenger punctuality each business unit should steer on their TVTAs to decrease the number of TVTAs. However some remarks to the use of TVTAs and the relation with the passenger punctuality can be given.

- As stated in paragraph 2.2, an advantage of the passenger punctuality is that, in principle, every train is weighted with the number of passengers. Hereby the passenger punctuality gives a

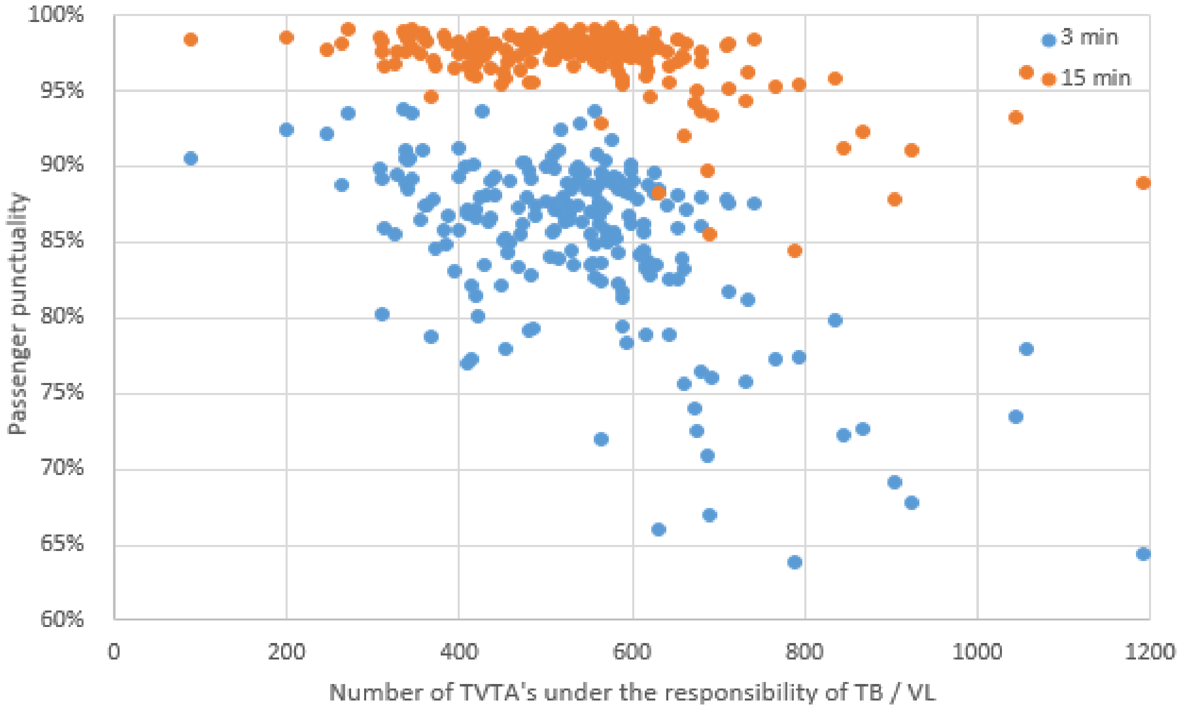


Figure 5.4: The number of TVTAs versus the passenger punctuality for 181 days for the business unit TB / VL

fairer view of the performance of ProRail and the NS than the train punctuality, according to the passengers. Nonetheless every TVTA is weighted the same. So the number of passengers affected by a TVTA is not taken into account. A TVTA that occurred late in the evening is treated the same as the same TVTA but occurring during morning rush hour whereby a fully loaded train is hindered. This implies that steering to decreasing the number of TVTAs is not as effective as the model suggests to increase passenger punctuality.

- The model of ProRail and the NS relates on the fact that the correct type of TVTA is coupled to every train deviation. This is the task of a signaller, but this is not his main priority. If the work load of the signaller is extensive (for example by a large disruption) the coupling of train deviations to the correct TVTA is put off. The question arises if after a while the signaller still knows the cause of every train deviation, certainly when there are many TVTAs. This makes the system less reliable. Besides it could occur that multiple types of TVTAs apply to the same train deviation though the signaller has to choose one. Take for example a train deviation because the doors of the train will not close. The signaller will label this train deviation as TVTA D3: Broken rolling stock. This incident causes the next train to deviate from its schedule as well. At this point the signaller can explain this train deviation the TVTA A2: Delay train - train, the second train is delayed because the first train is delayed. Nonetheless the signaller can choose to explain this train deviation with TVTA D3 as well: Broken rolling stock, the second train is delayed because the first train had a defect.

- The Pearson correlation coefficient between the 3 minute passenger punctuality and the number of TVTAs is -0.87. This suggests a clear and strong relation. Nonetheless in the model the TVTAs are divided among the business units. The Pearson correlation coefficient between the 3 minute passenger punctuality and the TVTAs for only the business unit TB / VL is -0.57. This implies a more moderate relation. The correlation coefficient for the 15 minute passenger punctuality is -0.52. In Figure 5.4 the number of TVTAs under the responsibility of TB / VL is related with the passenger punctuality.

- As can be seen in Table 5.1, 33.94% of all the TVTAs are not taken into account by the different business units. These TVTAs included two of the three types of TVTAs most occurring: A1 and A2. These TVTAs are indirect, which means that the actual cause of a train deviation is another

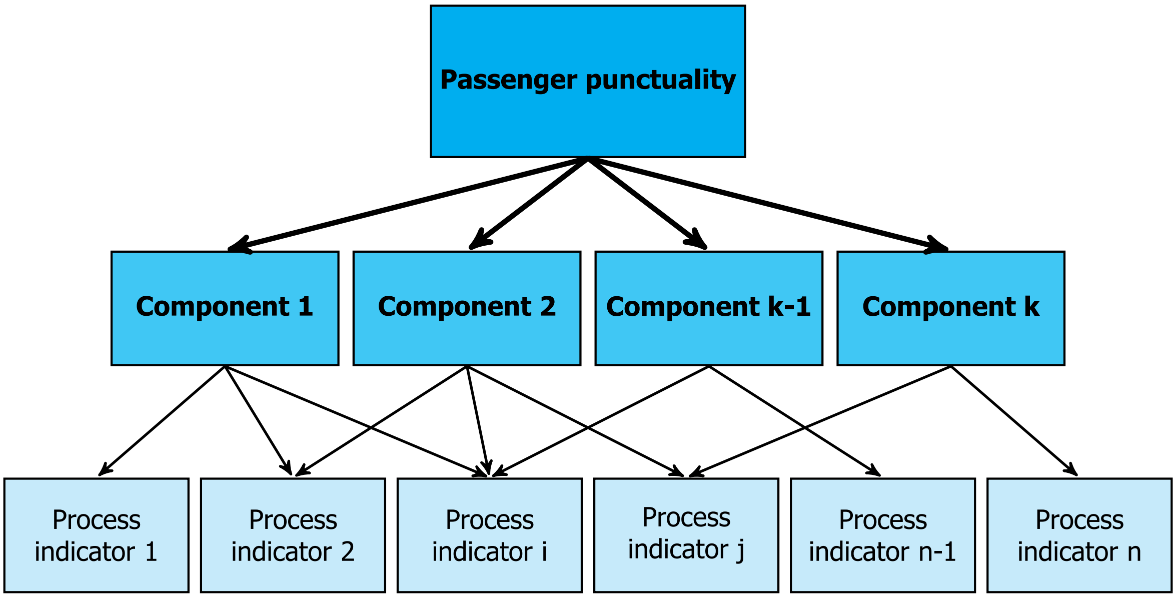


Figure 5.5: A simplified scheme of the framework around passenger punctuality

TVTA, possible occurred by another train. These indirect TVTAs indicate the knock-on effect. This effect describes the progression of train delays along the network after a disruption. Note that, if the disruption can be prevented, these TVTAs will also not occur. In the model of ProRail and the NS the focus lies on preventing disruption to occur. Hereby the knock-on effect is neglected. Nonetheless there are always disruptions that cannot be prevented so the knock-on effect should also be taken into account to minimise the passenger delays. Interesting here is the fact that the Pearson correlation coefficient between the number of A2 and A2 TVTAs and the passenger punctuality is -0.72, which implies a stronger relation than the relation between the TVTAs under the responsibility of the business unit TB / VL and the passenger punctuality.

5.1.2. Model 2: The model based on the components

The common factor in the disadvantages of the model of ProRail and the NS are the TVTAs (see paragraph 5.1.1). Therefore an alternative model will be proposed, where the process indicators of Transport control and Traffic control are linked with the passenger punctuality without a middle layer of TVTAs. This new model uses the framework around the passenger punctuality (see Figure 5.5).

The idea of this new model is the following. Bad passenger punctuality can be split into components, as examined in chapter 4. Process indicators, belonging to the Transport control and the Traffic control, are related to these components and therefore related to the passenger punctuality itself. Steering on these process indicators should therefore decrease the size of the corresponding components and thus improve the passenger punctuality.

The basis of the process indicators in this new model are the same process indicators in the model of ProRail and NS. As stated in paragraph 5.1.1, there are four categories of TVTAs. For passenger transport two categories are in particular important: cancelled and delayed trains. The other two categories are rare (redirected trains) or are not applicable to passenger trains (realignment in time). Note that almost all the components can be divided among two categories. The components 'train delayed', 'train departed too late' and 'missed transfer' occur because a train was delayed. If a train is cancelled the trips involved get the component 'train did not departed' or 'train did not arrive'. Together these five components form the majority of all delayed trips. For the 3 minute passenger punctuality the Pearson correlation coefficient between the number of TVTAs belonging to each category and the percentage of trips belonging to the two groups of components can be calculated. With a correlation coefficient of 0.77 between the TVTAs and percentage of trips because of delayed trains, and a correlation coefficient of 0.75 for the cancelled trains, the correlation is strong. This means that there is correlation between the components and the TVTAs. Therefore the process indicators used in the model of ProRail and the NS to reduce the number of TVTAs are used as basis for the process indicators in this new model

5.2. Introduction of the process indicators

Every business unit consists of a couple of individual units. For the Traffic control (VL) and Transport control (TB), four units can be distinguished:

1. National Traffic Control (LVL)
2. National Transport Control Centrum (LBC)
3. Regional Traffic Control (VL post)
4. Regional Transport Control (RBC)

Since 2010 the LVL and the LBC are combined in the Operational Control Centre Rail (OCCR). Here the LVL and the LBC work together to coordinate the handling of the disruptions on the rail network.

For the business unit TB / VL ProRail and the NS distinguished 20 different process indicators [37]. The process indicators reflect the performance of the corresponding processes and therefore the performance of the business unit itself. This is visualised in the model of ProRail and the NS (see Figure 5.1). These processes should have an effect on the number of TVTAs and thus on the passenger punctuality. However, as stated in paragraph 5.1.2, the same process indicators apply in the model based on the components.

The process indicators are given in Table 5.2. Per group the process indicators are explained in more detail. Process indicators 1 and 2 give the performance of the signallers on the VL post, and handles mostly small delays (paragraph 5.2.1). Punctuality of empty rolling stock is indicated via process indicators 3 and 4 in paragraph 5.2.2. The performance of the transport controllers is given with process indicators 5 and 6 (paragraph 5.2.3). Process indicators 7 to 12 are related to lead times between key moments by large disruptions. These indicators are explained in paragraph 5.2.4. Process indicators 13 to 15 (paragraph 5.2.5) are also related to large disruptions. The last five process indicators are not fully developed yet so these indicators are not used by calculating the correlations (paragraph 5.3). Nonetheless in paragraph 5.2.6 some background is given.

Table 5.2: The process indicators with the responsible units [37]

	Process indicator	**Unit**
1	Order processing VL-phase	VL post
2	Percentage of the correct use of the Train handling document	VL post
3	Departure punctuality of empty rolling stock at six rail yards	RBC, VL post
4	Departure punctuality of empty rolling stock at stations	VL post
5	Percentage of trains with the right composition at the morning start	RBC
6	Percentage of trains with the right number of personnel at the morning start	RBC
7	Lead time between 'Known VL' and 'Known BackOffice'	VL post
8	Lead time between 'Known BackOffice' and alarm	LVL-LBC
9	Lead time between alarm and Distribution decision	LVL-LBC
10	Lead time between Distribution decision and Blockage measure decision	LVL-LBC
11	Lead time between Blockage measure decision and running according to the Blockage measures	LVL-LBC, RBC, VL post
12	Lead time between end of Blockage and running according to the regular plan	LVL-LBC
13	Percentage of Blockage measure decision with a Distribution decision	LVL-LBC
14	Percentage of fully and correctly filled in report cards	LVL-LBC
15	Percentage of unadapted Blockage measures	LVL-LBC
16	*Red signal approaches due to not clearing the signals on time*	*VL post*
17	*Percentage of long lines whereby a new train is created*	*LVL-LBC*
18	*Blockage measures decision on time in VKL*	*LVL-LBC*
19	*The balance of the personnel on four moments per day*	*RBC*
20	*The balance of the rolling stock on four moments per day*	*RBC*

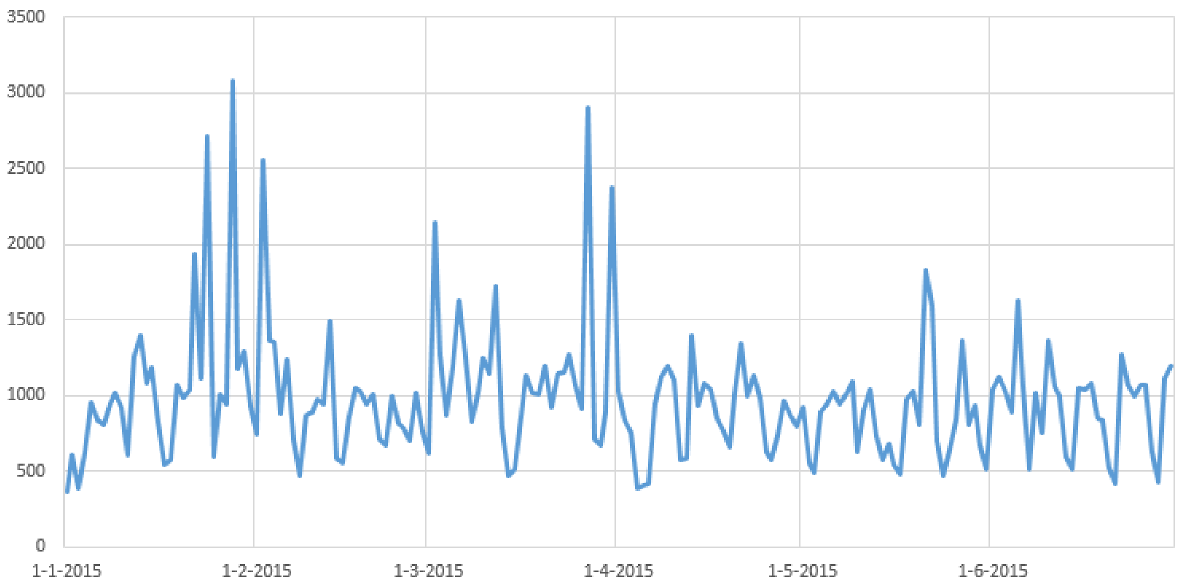


Figure 5.6: The number of mutations in VKL

5.2.1. Process indicators 1 and 2: Small delays

Dispatcher and signallers use a system called VKL. This system is designed to assist in the daily work activities. It provides the planned timetable and the corresponding deployment of rolling stock and personnel [38]. If something changes in for example the timetable, a mutation in VKL is made. The number of mutations in the last 36 hours before execution is the process indicator called 'Order processing VL-phase'. More mutations mean a higher workload for the signaller. And this will lead to a higher chance on delays. Note that only mutations in the regular plan are counted, like changes in the timetable or the cancelling of trains. Empty rolling stock runs are also included, but shunting movements are not. Changes in the balance of personnel or rolling stock are also not included. A side note is that this process indicator is meant as an information-item and not as in indicator to steer on [34].

Figure 5.6 shows the number of mutations for the first six months in 2015. Clearly visible are the large differences in the number of mutations. The highest number is more than 8 times as high as the lowest number (3078 versus 367). Moreover, a lower number of mutations do not guarantee lower passenger punctuality. It only implies a lower workload for the signaller.

If a train is delayed the train handling document (TAD) should be applied. This document consists of three components [39]:

1. **Waiting times passenger trains** For every node the maximum waiting time is determined for a delayed connecting train. Also the maximum waiting time for the last connection of the day is determined.
2. **Handling strategy per train or train series** The handling strategy describes per station what should happen by an increasing delay. The goal is to prevent the knock on effect of delays.
3. **If/then-scenarios** These scenarios are intended for delays in the train service originated at a node. If a train exceeds a certain delay threshold the signaller is authorized to use these scenarios.

A train handling document is made to minimise delays. It is therefore important for the signaller to execute this document in the correct manner. If not the delay is larger than possible and thus the passenger punctuality is not as high as it could be. The second process indicator shows the percentage of situations where a TAD applies and where this TAD is applied correctly.

The average of situations where the train handing document is applied correctly is the first six month of 2015 87.88% (see Figure 5.7). A remark is that in 1566510 of the 1638823 (95.59%) train handling documents situations the document did not apply. These situations are therefore not taken into account in the calculation of the process indicator.

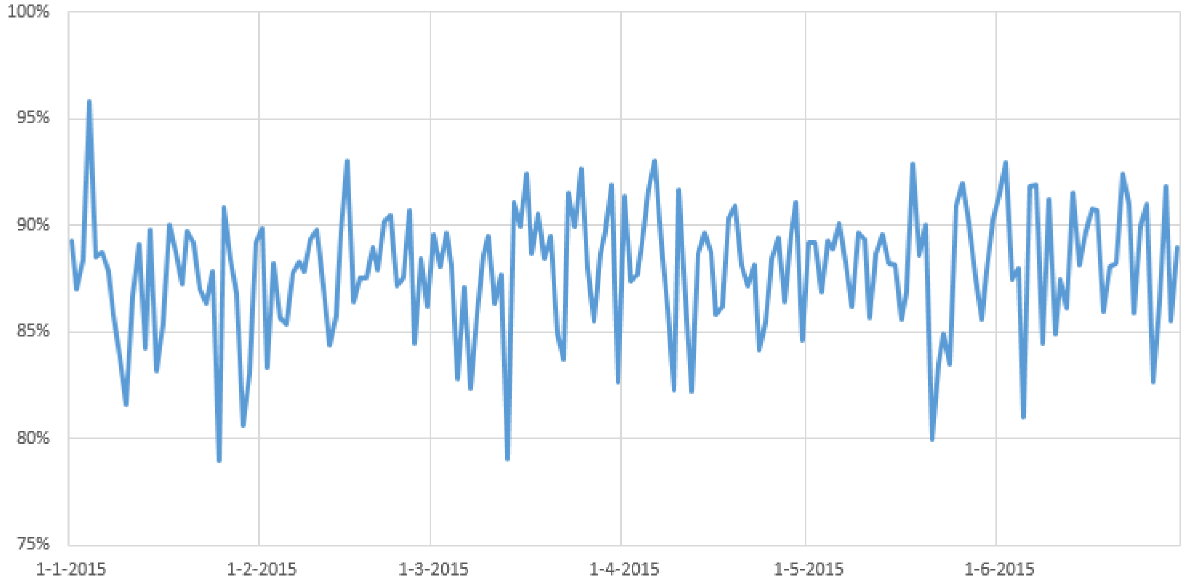


Figure 5.7: The percentage of situations where the train handling document is applied correctly

5.2.2. Process indicators 3 and 4: Empty rolling stock

Some extra attention in the process indicators goes to empty rolling stock runs. The third indicator is the departure punctuality of empty rolling stock at six rail yards. These rail yards are Amsterdam Lijnwerkplaats Zuid (Aswplz), Bokkeduinen (Bkd), Den Haag Binckhorst (Bkh), Hoofddorp Opstelterrein (Hfdo), Lelystad Opstelterrein (Llso) and Watergraafsmeer (Wgm). Most rail yards are located next to a station and have the same service control point. This implies that the empty rolling stock run from the rail yard to the station is not visible in the realisation data. For the six mentioned rail yards this is not the case. These rail yards do have their own service control point name and thus the runs are visible in the realisation data. To ensure that an empty rolling stock run belongs to this process indicator, the departure should be one of the rail yards and the running time needs to be less than 10 minutes. The process indicator is the 1 minute departure punctuality. If this number becomes higher, more runs are made on time and thus more rolling stock is on time at the station, ready for departure with passengers. Both the VL post and the RBC are responsible for this process indicator.

The fourth indicator is the departure punctuality of empty rolling stock runs between stations. To find those runs in the data, only empty rolling stock runs with a running time of at least 10 minutes are used. This prevents runs between rail yards and stations. However it also implies that empty rolling stock runs of less than 10 minutes are not taken into account. For the punctuality of all the trains it is important that the empty rolling stock runs on its reserved path. Therefore the process indicator is the two-sided 3 minute departure punctuality. This means that a train departing 3 minutes late is considered dispunctual, but a train departing 3 minutes early is considered dispunctual as well. The VL post is responsible for this indicator.

The calculated punctualities for both indicators for the first six months of 2015 can be found in Figure 5.8. As can be seen, both indicators are on average 70%. This suggest that the Pearson correlation coefficients between these process indicators and the TVTA under the responsibility of VL / TB are almost equal

5.2.3. Process indicators 5 and 6: Start conditions

The focus of process indicators 5 and 6 is to capture the performance of the start-up of each day. First there is the percentage of trains with the right composition of rolling stock at the morning start. Here the right composition is the composition according to plan. It concerns trains departing with passengers for the first time each day and where the rolling stocks come from the night transition and not from another train. Second there is the percentage of trains with the right number of personnel at

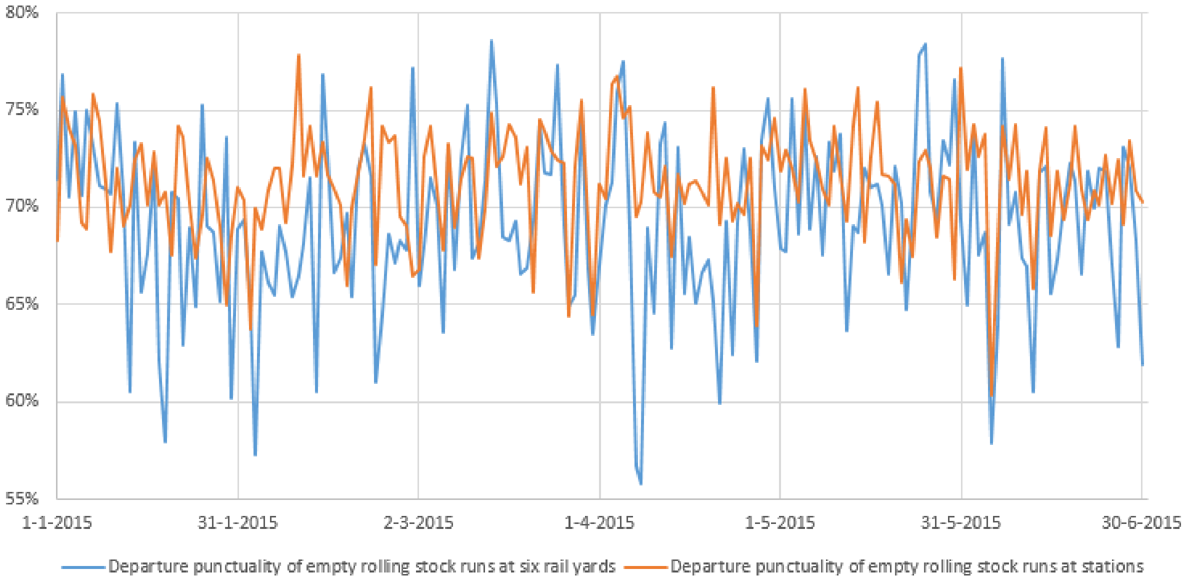


Figure 5.8: The values of process indicator 3 and 4

the morning start. Here it concerns the percentage of staffed first tasks of planned services between 4.00 a.m. and 7.59 a.m. For both indicators the RBC is responsible.

For the first six months in 2015, the values of these two indicators are given in Figure 5.9. Note that a bad performance of one of these indicators does not directly mean an increase in the number of the TVTAs or bad passenger punctuality. For example, a train with fewer coaches than planned can still transport all the passengers. This is thus no cause for a delay. However it is arguable that a low percentage leads to more cancelled trains. The dwell time might increase as well.

5.2.4. Process indicators 7 to 12: Lead times by disruptions

On a daily basis temporary measures are needed to ensure train service on the network. These measures are due to disruptions on the network. To inform all parties involved, a report card is made [40]. This is done by the BackOffice, an independent department at the OCCR. Two types of report cards can be distinguished.

1. **Communication:** If a train service cannot be executed by reasons other than limitations of the infrastructure maintained by ProRail such a report card is made. An example of an event where this type of report card is made is if personnel are on strike or if there is not enough rolling stock available.
2. **Calamity:** By a blockage, a disruption whereby the infrastructure cannot be used (like by a signal failure or collision), this type of report card is made.

In theory the function of those types of report cards is different, however in practice they are used in the same manner. Therefore the process indicators, based on (data of) the report cards, are calculated for both types combined. However some report cards are drawn up but not acted on, because it did not require any further action. These cards are neglected in the determination of the process indicators.

If there is an obstruction whereby a part of the infrastructure cannot be used a signaller at a VL post is notified. This signaller should contact the BackOffice so they can create a report card. The time that the BackOffice is notified is filled in on the report card. This is the 'Known BackOffice' time. The BackOffice notes also the 'Known VL' time: the time the signaller was notified.

To notify all parties involved, the BackOffice should alarm these parties. The time this is done is noted on the report card as well. After this a Distribution decision and then a Blockage measures decision should be made by the traffic controller and the transport controller at the OCCR. The Distribution decision is a decision about the distribution of the remaining capacity among the rail operators. This decision is made by the traffic controller. The transport controller should then make a Blockage measure decision. This is mostly a premeditate plan, but it can differ by exceptional circumstances.

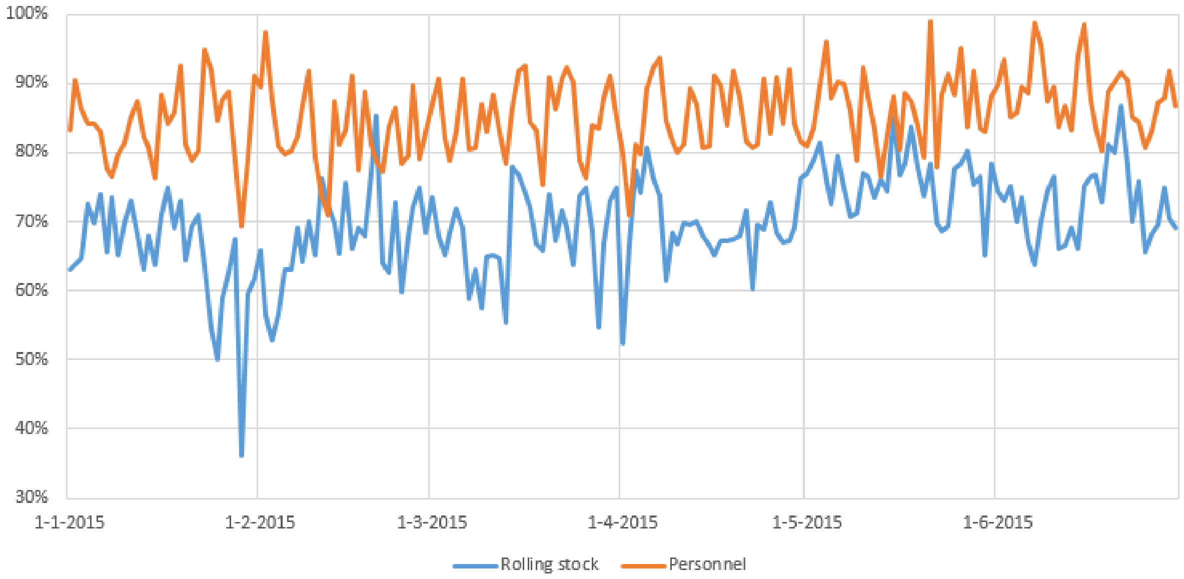


Figure 5.9: The values of process indicators 17 and 18

Blockage measures can include retiming or rerouting trains, cancelling trains or arranging alternative transport. Next the traffic controller can approve the plan or make some changes.

The time each decision is made is also noted on the report card. After the decision is made which Blockage measures needs to be taken, the signaller at the VL post and the transport controllers at the RBC can act accordingly. This implies that the trains can run according to the Blockage measures. The lead times between every time stamp on the report cards indicates how fast every process is. If a process becomes faster, measures are taken faster to minimise the total passenger delay. Therefore the lead times are a good indicator of how well each process performs. Here the exact definition of the process indicators is the 90th percentile of all the lead times of every process [34]. So 90% of every lead time is equal or less than the indicator. The 90th percentile is used to neglect outliers and to make large lead times more important: steering on the lead times leads in practise only to steering on the largest lead times.

The responsible unit to inform the BackOffice of a disruption is the VL post. The BackOffice is located at the OCCR, so the responsible unit is the LVL-LBC from the moment the BackOffice is notified until the decision about the Blockage measures is made.

When there is decided about the Blockage measures every unit should work together to take care that every train can run according to the Blockage measures. So the LVL-LBC, the RBC and VL post are all responsible for this process. Again, the indicator is the 90th percentile of all lead times between the time the decision is made and the time that every measure is applied at least once. The time the decision is made is noted on the report card. This is not the case for the time every measure is applied. This has to be determined later via the realisation data. Note that measures for infrequent and missing train series are neglected.

After the blockage is cleared and the entire infrastructure can be used again, the train services should start running according to the regular plan. The time that the infrastructure is cleared is noted on the report card, at the LVL-LBC. The time every train series runs according to plan is determined via the realisation data, at least one day later. A train series runs according to plan if two consecutive trains pass the location of the obstructed infrastructure. Infrequent and missing train series are neglected. The 90th percentile of all lead times between those key moments is used for the official process indicator.

5.2.5. Process indicators 13 to 15: Disruptions

Process indicators 13 to 15 do not indicate the lead times between steps in the process by a blockage of the infrastructure (see paragraph 5.2.4, though these indicators are connected to this process. First, every Blockage measure decision should only be made if a Distribution decision is made. Process indicator 13 gives the percentages of report cards where both decisions were made. A higher percentage will not necessary reduce the number of TVTAs or increase the passenger punctuality. However it shows that the developed process to deal with blockages is conducted as good as possible.

The same idea is behind the percentage of fully and correctly filled in report cards (process indicator 14). Steering on this indicator will have only an indirect effect on the number of TVTAs or the passenger punctuality. To goal is to minimise the number of wrongly filled in report cards. This includes also the report cards where negative lead times are registered. Steering on this indicator leads to less of such outliers and thus more reliable data. It is also possible that a lead time is too long to be correct. Process indicators 7 to 12 only use the 90th percentile of all lead times, so these outliers already are neglected [37].

Process indicator 15 gives the percentage of Blockage measures that are not adapted. As stated, the Blockage measures are mostly premeditating plans. These plans are made to minimise the delays. However due to the local circumstances, the Blockage measures might differ. This is for the transport controller to decide. Nonetheless if the Blockage measures differ from the predetermined plan, the traffic controller has to check the plan. This indicator gives the percentage of those measures that are adapted by the traffic controller. If the number of predetermined plans increases, this percentage will decrease. This will increase the lead time. This percentage can be measured with the report cards. If the measures are adapted, this will be noted on the report card.

5.2.6. Process indicators 16 to 20: Undeveloped

The last five process indicators are not yet fully developed. Therefore they are neglected in the remainder of this research, but this paragraph will give a quick overview.

If a train approaches a red signal, it has to slow down, possibly stop, and after the signal is cleared accelerate. This takes time and can cause a delay. Process indicator 16 gives the number of red signal approaches where there is a train present according to plan, there is no hindrance of other traffic, there is no 'Extra Control', the path is cleared by the signaller and the signal is not cleared. The less this occurs, the fewer delays and thus the better the passenger punctuality is. This process indicator is not yet in use because there is no clear definition of 'hindrance of other traffic'.

If a train has a large delay at the beginning of the journey, many passengers are delayed: not only the passengers departing at the start, but also the passengers departing later. This is especially a problem for long lines. In such cases the decision can be made to run an extra train on the second part of the journey on the original planned train path. The result is that the passengers departing in the second part are not delayed. The performance of this adjusting measure is indicated by the percentage of long lines whereby a new train is created (process indicator 17). The LVL-LBC is responsible for this indicator.

This indicator is not fully developed and thus not in use. For example, the series on which this measure can be applied are unknown at this moment. The same applies for the stations where a new train can be created and by which delays this should happen. Besides, if the indicator is in use the expected number of situation where this measure can be applied is low. This is due to the availability of rolling stock, personnel and shunting paths.

It is important that if a Blockage measure decision is made (see paragraph 5.2.4, the measures are included in VKL. Not only assist this system signallers in their work, this system provides the passenger information systems of the necessary input. This means for example that if Blockage measures are included in VKL, the impact of those measures is included in the journey planner of the NS. The responsibility lies by the LVL-LBC but also this indicator is not yet developed.

This indicator is not developed, so it is not possible to calculate the correlation between this indicator and the number of TVTAs or the passenger punctuality. However it is expected that there is no correlation. This process indicator has influence on passenger behaviour and not on the situation on the rail network. On the other hand, passenger punctuality is affected by passenger behaviour. Take the following example. A passenger sees that the train he was intended to take has a delay. If there was no delay, the passenger would check in on time and the promised train would match the train the passenger took. However because of the delay the passenger checks in later. This result in a promised train that differs from the train he was intended to take. If this second train is on time, the result in the passenger punctuality calculation would be that this passenger is considered on time, while in fact he had a delay equal to the delay of the first train. Or in other words, the passenger punctuality KPI can increase if more Blockage measure decisions are included in VKL on time, while the actual passenger punctuality not increases.

The last two indicators are the responsibility of the RBC. They give the balance of rolling stock and personnel on four moments on each day. Note that these two indicators have a large similarity with process indicators 5 and 6 (see paragraph 5.2.3). However, here the performance of the transport controllers is measured while with the right composition of the rolling stock and the right number of personnel at the morning start the start situation of the day is indicated.

5.3. Correlations in the two models

The correlation between two variables can be determined via the Pearson correlation coefficient. This coefficient is a value between -1 and 1. If the outcome if -1 or 1, the two variables are fully correlated. If the outcome is 0, there is no correlation at all. In this paragraph the correlation between the process indicators (paragraph 5.3 and the number of TVTAs (see paragraph 5.1.1) or the components and the passenger punctuality (see paragraph 5.1.2) is calculated. In general a Pearson correlation coefficient between -0.3 and 0.3 is interpreted as no correlation. By a higher of lower correlation coefficient the correlations becomes weak, moderate and finally strong.

For the model of ProRail and the NS and analyses about TVTAs, data of the first six months of 2015 is used. For the second model, based on the components, and analyses about passenger punctuality, data of August to October 2015 is used.

5.3.1. Process indicators 1 and 2: Small delays

The first two process indicators give the number of mutations in VKL and the average of situations where the train handing document is applied correctly (see paragraph 5.2.1. Table 5.3 gives the Pearson correlation coefficient between these process indicators and the number of TVTAs, according to model 1, and the components and the passenger punctuality, according to model 2.

Table 5.3: The Pearson correlation coefficient between the number of TVTAs, the 3 and 15 minute passenger punctuality and the underlying components and the first two process indicators

	1	**2**	**1**	**2**
TVTAs	0.61	-0.43		
	3 minute		*15 minute*	
Passenger punctuality	-0.56	0.40	-0.50	0.32
Train delayed	0.51	-0.39	0.47	-0.32
Train departed too early	-0.21	-0.01	-0.32	-0.03
Train departed too late	0.49	-0.35	0.47	-0.34
Train did not depart	0.64	-0.25	0.60	-0.24
Train did not arrive	0.04	-0.19	0.27	-0.18
Missed transfer	-0.09	-0.31	-0.21	-0.29

If there is a change in the original timetable, for example a train is cancelled, this results in a mutation in VKL. More mutations imply more TVTAs and more deviations in the timetable. It is expected that this leads to more delays for passengers because less passengers are able to travel according to their promised journey (in Dutch: Reisbelofte). Calculating the Pearson correlation coefficient confirms this expectation. The Pearson correlation coefficient between the number of mutations in VKL and the number of TVTAs under the responsibility of TB / VL is 0.61. This implies a moderate to strong relation. The coefficient between all TVTAs and the number of mutations is 0.85: a strong correlation.

The coefficient between the number of mutations and the 3 minute passenger punctuality is -0.56: a more than moderate correlation. For the 15 minute passenger punctuality the coefficient is -0.50, which implies a moderate correlation. Taking only passenger trains of the NS into account by calculating the number of mutations, the correlation coefficients become -0.67 and -0.77 for the 3 minute and 15 minute passenger punctuality respectively: an even stronger correlation.

The correlation coefficient can also be calculated for the percentage of trips with a specific component. For the component 'train did not depart' and the 3 minute passenger punctuality this coefficient is 0.64 and for the 15 minute passenger punctuality 0.60. This implies an almost strong relation between the number of mutations and the number of cancelled trains. The correlation coefficient with the component 'train did not arrive' is too low to have any meaning, certainly for the 3 minute passenger punctuality, so the conclusion can be made that most cancelled trains are cancelled before starting their trip and not during.

The correlation between the number of mutations and the percentage of trips with the component 'train delayed' is calculated as well. For the 3 minute passenger punctuality this coefficient is 0.51. This implies a moderate correlation. For the 15 minute passenger punctuality the correlation coefficient is 0.47. This implies that more mutations lead to more delays. Remarkable is that the correlation with the component 'missed transfer' is not-existing for the 3 minute passenger punctuality. This suggests that with more mutations, more passengers will get a delay, but the number of missed transfers (due to a delay) stays the same.

Process indicator 2 is intended to minimise delays. Therefore only the TVTAs under the responsibility of TB / VL and with the category 'Delayed train' (see paragraph 5.1.1) are used. The Pearson correlation coefficient between this set of TVTAs and the process indicator is -0.43. This implies a weak to moderate relation.

The intension to minimise delays is visible by calculating the correlation between the process indicator and the percentage of trips with the different components: A distinction is visible between components occurring due to delayed trains and due to cancelled trains. Table 5.3 shows these correlation coefficients. Although none of the correlations is more than moderate, the components 'train delayed', 'train departed too late' and 'missed transfer' have a higher correlation coefficient than the components 'train did dot depart' and 'train did not arrive'.

The Pearson correlation coefficient between the process indicator and the 3 minute passenger punctuality is 0.40: a weak to moderate correlation. For the 15 minute passenger punctuality this correlation is even weaker: 0.32. This is not surprising because a train handling document is mainly applicable is situations with small delays.

5.3.2. Process indicators 3 and 4: Empty rolling stock

Dispunctual behaviour of empty rolling stock runs can influence the punctuality of passenger trains (see paragraph 5.2.2. This can lead to more TVTAs and more passenger delays. It is therefore important that the departures of the empty rolling stock runs are on time. Within the empty rolling stock runs a distinction is made between runs between rail yard and station and between two stations. The first group is identified by a departure at one of the six rail yards that is not located directly next to a station and where the total running time is less than 10 minutes (process indicator 3). For this group the process indicator is the 1 minute departure punctuality. The second group is identified as the empty rolling stock runs with a running time of at least 10 minutes (process indicator 4). The process indicator is the percentage of departures within 3 minutes before or after the planned departure.

Table 5.4: The Pearson correlation coefficient between the number of TVTAs, the 3 and 15 minute passenger punctuality and its components and process indicator 3 and 4

	3	**4**	**3**	**4**
TVTAs	-0.42	-0.51		
	3 minute		*15 minute*	
Passenger punctuality	0.48	0.49	0.38	0.34
Train delayed	-0.47	-0.52	-0.39	-0.36
Train departed too early	0.16	0.15	0.18	0.11
Train departed too late	-0.39	-0.36	-0.40	-0.36
Train did not depart	-0.33	-0.22	-0.32	-0.24
Train did not arrive	-0.15	-0.20	-0.26	-0.27
Missed transfer	-0.28	-0.35	-0.18	-0.27

Table 5.4 shows the Pearson correlation coefficient between the number of TVTAs, the passenger punctuality and its components and the two process indicators. Dispunctual empty rolling stock runs only cause delayed passenger trains, so only TVTAs with this category (see paragraph 5.1.1 are used. The coefficient between the number of TVTAs under the responsibility of VL / TB is -0.42 for the departure punctuality at the rail yards and -0.51 for the departure punctuality at stations. This implies a weak to moderate correlation. The coefficient between the process indicators and the number of all TVTAs is -0.63 and -0.49 respectively. This implies a moderate to strong correlation.

For the 3 minute passenger punctuality applies a moderate correlation. The correlation with the component 'Train delayed' is almost the same and the correlation with 'Train departed too late' and 'Missed transfer' is stronger than for the other components as well. This suggests that lower departure punctuality leads to more delays, but not to more cancelled trains. This is in line with the reasoning behind the process indicators. The correlation coefficients by the 15 minute passenger punctuality are lower, but the same trend is visible.

Note that the (expected) number of passengers delayed by an empty rolling stock run is not taken into account in the process indicator. It is arguable that if the definition if the process indicators changes to the percentage of delayed trips as a result of an empty rolling stock run, the correlation will become stronger. However it is very hard, or maybe impossible, to determine which passengers are affected by a dispunctual empty rolling stock run and which passengers would be affected if an empty rolling stock run was dispunctual.

5.3.3. Process indicators 5 and 6: Start conditions

The percentage of trains with the right composition at the morning start (process indicator 5) and the percentage of trains with the right number of personnel at the morning start (process indicator 6) determines the start conditions of each day (see paragraph 5.2.3). Table 5.5 shows the Pearson correlation coefficient with the number of TVTAs and the components and passenger punctuality.

Table 5.5: The Pearson correlation coefficient between the number of TVTAs, the 3 and 15 minute passenger punctuality and its components and process indicator 5 and 6

	5	**6**	**5**	**6**
TVTAs	-0.19	0.084		
	3 minute		*15 minute*	
Passenger punctuality	-0.06	-0.04	-0.06	-0.05
Train delayed	0.06	0.02	-0.23	0.02
Train departed too early	0.05	0.06	0.10	0.09
Train departed too late	0.02	-0.03	0.02	-0.03
Train did not depart	0.003	0.07	0.01	0.06
Train did not arrive	0.16	0.06	0.16	0.03
Missed transfer	0.17	0.10	0.16	0.10

The Pearson correlation coefficient between the number of delay TVTAs under responsibility of TB / VL and the percentage of trains with the right composition is -0.19. This implies a weak to non-existing correlation. However it is arguable that a low percentage leads to more cancelled trains: if less rolling stock is available, the change that a train gets cancelled increases. Therefore the correlation coefficient between the number of TVTAs occurring because of cancelled trains and this process indicator is determined. This coefficient is also -0.19. The relation between the number of delay TVTAs under responsibility of VL / TB and the percentage of trains with the right number of personnel is -0.08: a non-existent correlation.

For both process indicators, and for both the 3 minute and 15 minute passenger punctuality the correlation with the components is calculated as well. Nonetheless almost every Pearson correlation coefficient is close to 0. This means that there is no correlation at all.

This outcome could be expected. If all trains run, but the length of the train is shorter or the type of train differs from the planned type, every passenger should be able to catch the train that was promised. Therefore the composition of the train does not have any effect on the passenger punctuality or on its components. And the same reasoning could be applied to the number of personnel at the morning start. On the other hand if a train is too short, the required dwell time can increase. Therefore an alternative process indicator can be the percentage of trains shorter than planned. Unfortunately this process indicator cannot be calculated with the current data available.

Note that this does not imply that the right composition of the trains and the right number of personnel is not important. Due to, among others, seat guarantee and safety, these factors are important. Nonetheless in combination with the passenger punctuality the process indicators could be neglected.

5.3.4. Process indicators 7 to 12: Lead times by disruptions

When there is a disruption multiple steps are taken by the regional traffic control post and at the national traffic control and national transport control centre to modify the timetable during disruptions and restore the planned timetable after the disruption (see paragraph 5.2.4). The goal is to minimise the lead times between each step. Two remarks can be given.

1. In the model of ProRail and the NS the goal is to minimise the number of TVTAs (paragraph 5.1.1). If a lead time is large, the number of TVTAs will also be large. These are, in larger extend, TVTAs because trains are cancelled. Therefor only TVTAs because of cancelled trains are used.
2. The first five lead times together are the lead times between the start and the disruption and running according to the Blockage measures. This group forms the reduction of the timetable because of a disruption, while the sixth lead time is the restoration of the timetable. The reason that the reduction of the timetable is split into five segments is the responsible party. Every lead time has a different responsible party. However for analyses those five lead times are grouped together as well.

As process indicator the 90th percentile of all lead times is used. Steering on this indicator will create the necessity to minimise the largest lead times (neglecting the outliers) in particular. However the 90th percentile does not indicate the total number of disruption and the duration of all disruption. This implies that there is no correlation between the process indicators and the number of TVTAs, the passenger punctuality or the percentage of trips with each component. Calculating the Pearson correlation coefficient confirms this, see table 5.6. For both models the correlation coefficient lies between -0.26 and 0.31. Two large remarks can be given. The first remark is regarding the number of TVTAs used, so specific for the model of ProRail and the NS. In the second remark the use of the 90th percentile in the definition of the process indicators is questioned.

Table 5.6: The 90th percentile lead times (in seconds) and the Pearson correlation coefficient between the number of cancelled trains TVTAs under the responsibility of TB / VL, the 3 minute passenger punctuality and the underlying components and process indicators 7 to 12: the lead times between the various steps in the process to restore the timetable after a disruption

Process indicator	**7**	**8**	**9**	**10**	**11**	**12**	**7-11**
Lead time	304.9	214.8	1741	1792.2	2816.5	268	4396.9
TVTAs	-0.017	0.009	0.011	0.019	0.023	0.11	-0.021
3 minute passenger punctuality	-0.005	0.08	-0.02	-0.05	-0.26	-0.19	-0.011
Train delayed	0.03	-0.10	0.08	0.06	0.20	0.15	0.034
Train departed too early	-0.11	-0.03	-0.19	-0.10	0.01	-0.14	-0.11
Train departed too late	-0.04	-0.04	-0.06	0.01	0.18	0.15	-0.030
Train did not depart	-0.04	-0.05	-0.10	0.05	0.31	0.30	0.033
Train did not arrive	-0.07	0.10	-0.04	-0.01	0.28	0.05	-0.046
Missed transfer	0.02	0.03	-0.05	-0.11	0.18	-0.08	0.032

The number of TVTAs

Table 5.6 shows a non-existent correlation between the process indicators and the number of TVTAs. Here only the cancelled trains TVTAs that are under responsibility of the business unit TB / VL are used, according to the model of ProRail and the NS. Nonetheless if there is a disturbance, train deviations occurring due to this disturbance are explained via TVTAs that explain the disturbance. This is often not a TVTA that is under responsibility of TB / VL. Hereby applies that a disturbance has, on first, only effect on the nearby region. Using all types of TVTAs to calculation the correlation might therefore give the wrong correlation coefficient as well.

The relation between each process indicator and the number of all cancelled trains TVTAs occurring during the lead times is examined. Per day the number of TVTAs occurring during the time a disturbance is known at the VL post and known at BackOffice is calculated. Next the correlation between this number and the process indicator is calculated. The same is done for the other five process indicators.
The expectation is that if the lead time becomes longer, more TVTAs occur. This implies initially a strong relation. A disturbance is local, so there might be some disturbance in the number of TVTAs due to train deviations that are not due to the disturbance. Therefore the expectation is that the actual correlation between the number of TVTAs occurring and the process indicators is present, although not very strong.
Table 5.7 shows this expected result occurs for the relation between the number of cancelled train TVTAs per lead time and the last three process indicators. However there is no correlation between the first three process indicators or the combined lead times and the number of TVTAs occurring. This implies that steering on these process indicators to decrease the lead times will not have an effect on the number of TVTAs.

Table 5.7: The Pearson correlation coefficient between the process indicators and the average number of all cancelled trains TVTAs occurring during the lead times for the first six month of 2015

Process indicator	**7**	**8**	**9**	**10**	**11**	**12**	**7-11**
Lead time	304.9	214.8	1741	1792.2	2816.5	268	4396.9
TVTAs	-0.013	-0.014	0.039	0.31	0.66	0.33	-0.028

As stated, the expectation is that if the lead time becomes longer, more TVTAs occur. This seems to apply for the last three process indicators. Although this does not mean that some TVTAs did not occur if the lead time was smaller. This implies that steering to decrease the length of the lead times might not necessary decrease the number of TVTAs. Therefore the number of TVTAs occurring during the lead times is calculated and divided by the lead time to achieve the number of TVTAs occurring during a lead time per minute. The same is done for all TVTAs occurring outside of a lead time. The mean number of TVTAs occurring outside of a lead time for the first six months of 2015 is 1.33 TVTAs / minute. The mean number of TVTAs occurring during a lead time is 1.93 TVTAs / minute. Under the null hypothesis that the found values comes from independent random samples with an equal mean a two-sample t-test is performed. The outcome indicates to reject this null hypothesis. This implies that, with a 95% certainty, the means are different. So the length of a lead time influences the number of TVTAs occurring. More TVTAs occur during a lead time, so steering on the process indicators to decrease the length will thus lead to less TVTAs.

The 90th percentile

The use of the 90th percentile in the definition of the process indicators might not be best method to calculate the value of the process indicators. Three arguments can be distinguished.

1. The number of blockage measures per day is limited. Presume a small dataset with only 10 values. If the 90th percentile is calculated, the largest value has a large influence on the calculated percentile. Such an example dataset is not uncommon when calculating the process indicators regarding the lead times. This implies that this calculation result in values which oppose the reasoning behind use of the 90th percentile: to find the highest lead time possible without taking outliers into account. The average of number of lead teams per day used to calculate the 90th percentile lies between 13.3 for the lead time between 'Known VL' and 'Known BackOffice' (process indicator 7) and only 3.1 for the lead time between the Blockage measures decision and running according to the Blockage measures (process indicator 11). Here days with no registered lead times where neglected.
2. Another disadvantages of the 90th percentile is the method of calculation. There is not a standard method. This leads to different values if using a different method. The differences may vary enormous, certainly if a small dataset is used when calculating the 90th percentile.
3. Table 5.6 shows that there is no correlation between the process indicators and the number of TVTAs, the components or the passenger punctuality. This goes against the logical reasoning

that the length of the lead time and the number of TVTAs is related. Also it is already proved that the number of TVTAs per minute occurring during a lead time is significantly higher than the number of TVTAs per minute occurring outside of a lead time. So, in any case, for the model of ProRail and the NS there should be some correlation, only the value of the process indicators does not reflect it.

For both models the correlation between the lead times and the number of TVTAs or the components is determined. Some suggestions for new definitions of the process indicators are given as well.

Model 1: Correlation with the number of TVTAs

As stated above, during a lead time more TVTAs per minute do occur than outside of a lead time. For process indicator 10 to 12 this correlation is already slightly visible in table 5.7. Therefore first a check takes place to see if this correlation is also present for process indicator 7 to 9. The number of TVTAs occurring per minute during the lead times (on average 1.89 TVTAs / minute) and number of TVTAs occurring per minute outside of the lead times (on average 1.45 TVTAs / minute) for only the first three process indicators are compared. The null hypothesis is that the means of the samples are equal. With a 95% certainty this hypothesis is rejected. So this suggests that there is a correlation between the number of TVTAs and the length of the first three lead times as well. A second check took place, now dividing a day in three parts: the part of the day without a lead time active (part 1), the first five minutes of each lead time (part 2) and the remainder of the lead times (part 3). Again the mean of the number of TVTAs occurring per minute for each part is calculated and compared. This shows, with a 95% certainty, that the means of part 1 and part 2 are different and that the means of part 1 and 3 are different. However, the means between part 2 and 3 are not different. This implies that the duration of the lead time matters, the number of TVTAs occurring per minute during the first 5 minutes is the same during the remainder of the lead time.

In conclusion, the duration of the lead times of the first three process indicators do have an effect on the number of TVTAs. And because this relation did not come up when comparing the 90th percentile of the lead times and the number of TVTAs, this implies that the 90th percentile is not a good indication of the lead times to steer on.

A solution can be to use the median instead of the 90th percentile in the definition of process indicator 7 to 12. With only a few data points a median can be calculated and there is only one method of calculating a median. Another advantage is that outliers more or less are neglected. Nonetheless, as can be seen in Table 5.8, there is no correlation between the average number of all cancelled trains TVTAs occurring during the lead times and the median of the lead times for the first 3 process indicators as well. This can be due to the fact that the median (just as the 90th percentile) does not give information about the number of lead times used to calculate the process indicator. Note that this does not mean that there is no correlation between the lead time and the number of TVTAs. As stated above, for the first three process indicators there is a correlation. Also, there is a weak correlation between the first five lead times combined and the number of TVTAs. Another disadvantage is that not only the outliers but the valid large lead times are neglected as well.

Table 5.8: The median of the lead times (in seconds) and the Pearson correlation coefficient between the process indicator and the average number of all cancelled trains TVTAs

Process indicator	**7**	**8**	**9**	**10**	**11**	**12**	**7-11**
Lead time	167	73	199.5	797	1568.8	1567.5	640
TVTAs	0.083	0.054	0.14	0.31	0.44	0.33	0.34

A second solution is to use the sum of the lead times. Table 5.9 shows the average sum of the lead times and the correlation with the number of all cancelled trains TVTAs occurring during the lead times. Almost all correlations are strong. Due to the manual input of the time stamps, negative or very large lead times can occur. In this calculation negative lead times and lead times longer than 24 hours were neglected.

This solution deals with the disadvantages of the different calculation methods of the 90th percentile as well as the problems with small datasets. All process indicators have a clear correlation with the number of TVTAs as well. A drawback to this method is the effect of outliers to the value of the process indicators. It is still possible that time stamps are wrongly filled in on the report card which leads to too large process indicators. On the other hand, steering on these process indicators increases the need to fill those report cards correctly because that will bring down the value of the process indicator. This can result in time in more realistic values. Another disadvantages is that this solution is heavenly related to the number of disruptions, a factor that cannot be influenced by the traffic control.

Table 5.9: The median of the lead times (in seconds) and the Pearson correlation coefficient between the process indicator and the average number of all cancelled trains TVTAs

Process indicator	**7**	**8**	**9**	**10**	**11**	**12**	**7-11**
Lead time	3725.2	1879.4	10016.7	7258.4	5688.5	6052.3	25911.5
TVTAs	0.35	0.53	0.70	0.56	0.84	0.73	0.58

Model 2: Correlation with the components

To determine if the length of a lead time influences the components or the passenger punctuality two analyses are performed. First the passenger punctuality is calculated for trips with a planned departure during and not during a lead time. The result is shown in Table 5.10. A check took place to test if the means of the found passenger punctualities were significantly different during a lead time and not during a lead time. For all combinations of 3 and 15 minute passenger punctuality and the lead time between the different steps a hypothesis was made that the passenger punctuality during a lead time was not significantly different from the passenger punctuality not during a lead time. With a 95% certainty, this hypothesis can be rejected for almost all combinations. An exception was the lead time between 'Known VL' and 'Known BackOffice' (process indicator 7) and between 'Known BackOffice' and alarm (process indicator 8) with the 3 minute passenger punctuality. An explanation could be the duration of these lead times. The percentage of the day that there was such a lead time was small, certainly in comparison with the other lead times. This analysis shows clearly higher passenger punctuality when there was no disruption, or at least when a lead time was not 'active'.

Table 5.10: The total duration of the disrupted situation as a percentage of the day and the passenger punctuality (3 and 15 minute) for trips with a planned departure during one or more lead times and with a planned departure in a undisrupted situation

Lead time	**Duration**	**3 minute**		**15 minute**	
		Disrupted	*Undisrupted*	*Disrupted*	*Undisrupted*
All	41.29%	82.30%	85.51%	96.40%	97.61%
Known VL - Known BackOffice	3.91%	82.86%	83.77%	96.24%	96.98%
Known BackOffice - Alarm	1.53%	82.80%	83.77%	96.28%	96.98%
Alarm - Distribution decision	10.14%	82.22%	83.99%	96.17%	97.06%
Distribution decision - Blockage measure decision	8.13%	81.68%	83.85%	95.77%	97.01%
Blockage measure decision - Running accordingly	6.64%	81.01%	84.10%	95.63%	97.09%
End of Blockage - Running according to regular plan	21.61%	82.09%	84.45%	96.45%	97.14%

Secondly the number of delayed trips per minute lead time is compared with the number of delayed trips per minute not during a lead time. A trip is allocated to a period during a lead time if the planned departed is during this lead time. If this planned departure is not during a lead time, the trip is used to calculate the delayed trips outside of a lead time. The number of delayed trips is influenced by the total number of trips, so the found value is corrected for this variable by correcting the total number of trips for each day. The found results can be found in table 5.11. Clearly visible is the large difference in the means of number of delayed trips during a lead time and not during a lead time. Here also an analysis took place to check if the means are significant different. For all lead times, with a 95% certainty, this is the case. This means that during a lead time more passengers will end up with a delay.

Table 5.11: The total duration of the disrupted situation as a percentage of the day and the number of delayed trips (3 and 15 minute) per minute during a lead time and in an undisrupted situation

Lead time	**Duration**	**3 minute**		**15 minute**	
		Disrupted	*Undisrupted*	*Disrupted*	*Undisrupted*
All	41.29%	154.77	80.46	31.56	13.06
Known VL - Known BackOffice	3.91%	155.30	112.73	33.70	20.97
Known BackOffice - Alarm	1.53%	200.39	111.78	42.07	20.82
Alarm - Distribution decision	10.14%	153.56	108.39	32.63	19.85
Distribution decision - Blockage measure decision	8.13%	162.09	109.91	36.87	20.21
Blockage measure decision - Running accordingly	6.64%	191.86	106.34	43.19	19.40
End of Blockage - Running according to regular plan	21.61%	173.93	93.70	34.00	16.96

In conclusion, the correlation between the process indicators 7 to 12 and the components of the passenger punctuality is not-existent at worst and low at best (see Table 5.6). However a large lead time leads to more delayed trips and to lower passenger punctuality. This implies that in this model the definition of the process indicator is not correct as well, if the goal is to steer to minimise the lead time. Therefore two new definitions of the process indicators are proposed.

1. For the model of ProRail and the NS the best solution was to define the process indicators as the sum of lead times. Using the sum will give the process indicator information about the total duration of all disruptions and still the largest lead times will affect the outcome the most. This suggests that steering on this new definition will also lead to decrease the length of the longest lead times.

   Table 5.12: The Pearson correlation coefficient between the 3 minute passenger punctuality and its components and the first alternative for process indicators 3 to 8: the sum of the lead times between the various steps in the process to restore the timetable after a disruption

Process indicator	**7**	**8**	**9**	**10**	**11**	**12**	**7-11**
3 minute passenger punctuality	-0.07	-0.02	-0.16	-0.17	-0.37	-0.23	-0.30
Train delayed	0.08	-0.03	0.14	0.17	0.28	0.25	0.27
Train departed too early	-0.07	0.03	-0.08	-0.07	0.00	-0.17	-0.096
Train departed too late	0.11	0.05	0.20	0.13	0.33	0.14	0.31
Train did not depart	-0.02	0.12	0.09	0.16	0.43	0.20	0.25
Train did not arrive	0.00	0.08	0.14	0.00	0.30	-0.02	0.15
Missed transfer	0.12	-0.01	0.16	-0.02	0.16	-0.04	0.17

   Table 5.12 shows the correlation between each component and the sum of the lead times for process indicators 7 to 12. Note that outliers in lead times are neglected. Here an outlier is defined as a lead time that is negative or longer than 1 day. For both types of outlier applies that almost always the date and time stamp on the report card is incorrect. Note also that the sum is not the sum if multiple lead times between the same steps overlap. A more precise description would be the total time one or more lead times were 'active'. As can be seen in the table, the correlation is not-existing or only low at best.

2. The disadvantage of the previous definition is that the sum of the lead times does not give any information about the number of passengers involved. A lead time during rush hour has the same impact as a lead time of the same length during the night. Nonetheless the number of passenger involved differs enormously. This might explain why there is not any correlation between the sum

of the lead times and the components: the components take the number of passenger involved into account and the sum of the lead times does not.

Therefore the second alternative is the sum of the maximum expected passengers involved during the lead times. This expected number of passengers can be the average number of departing passengers per section of the rail network. The period used to calculate the average number of departing passengers can be of influence as well. For example, the total number of passengers in the summer is lower than average but the number of passenger travelling towards the sea is higher. Nonetheless here the used dataset did not allow specifying the departures to a section of the rail network and the time period of three months is too short to take the influence of seasons into account. Therefore the average number of departing passenger per minute for the entire rail network is calculated, based on a period of three months. A distinction is made between working days and weekend days.

For every process indicator the expected number of passenger departing during the lead times is calculated. This number will be the value of this alternative for the process indicator. This implies that the value for a lead time during rush hour is higher than for a lead time of the same duration during the night. The Pearson correlation coefficient between each component and this variant of the process indicators can be found in table 5.13. Negative lead times and lead times longer than a day are neglected. Also if multiple lead times overlap, the overlapping part is not used multiple times but only once.

Table 5.13: The Pearson correlation coefficient between the 3 minute passenger punctuality and its components and the second alternative for process indicators 3 to 8: the sum of the maximum expected passenger involved during the lead times between the various steps in the process to restore the timetable after a disruption

Process indicator	**3**	**4**	**5**	**6**	**7**	**8**	**3-7**
3 minute passenger punctuality	-0.17	-0.16	-0.26	-0.17	-0.31	-0.21	-0.33
Train delayed	0.20	0.13	0.24	0.19	0.28	0.24	0.33
Train departed too early	-0.19	-0.10	-0.10	-0.09	-0.01	-0.21	-0.12
Train departed too late	0.19	0.18	0.34	0.14	0.25	0.15	0.34
Train did not depart	0.10	0.27	0.26	0.15	0.35	0.22	0.36
Train did not arrive	-0.15	-0.05	-0.06	-0.07	0.04	-0.15	-0.094
Missed transfer	-0.03	-0.15	-0.03	-0.09	-0.05	-0.20	-0.12

The overall results shows a correlation that is a little bit stronger than with the first alternative, however the correlation is still weak. On the other hand, this variant can be improved by a better determination of the expected number of passengers involved. Another advantage of this variant is prioritising. If there are multiple disruption that cannot be solved individually, the expected number of passengers involved might be of help to see which disruption gets a higher priority to achieve the highest passenger punctuality possible.

5.3.5. Process indicators 13 to 15: Disruptions

The last three process indicators are closely related to process indicators 7 to 12 (see paragraph 5.2.5). Process indicator 13 is the percentage of Blockage measures decision with a Distribution decision. This means that if there is a large disruption, a Blockage measures decision should be made, stating the various measures to ensure the best possible timetable during the disruption. To make the best possible Blockage measures decision a Distribution decision should be made. Here the distribution of the involved rail operators, or train series, among the available infrastructure is determined.

Process indicator 14 describes the percentage of fully and correctly filled in report cards. In practice this means that every report card is checked if all time stamps are filled in and that the lead times are all positive. A Blockage measures decision should be present as well. The last process indicator is the percentage of Blockage measures decision that is unadapted. Table 5.14 shows the Pearson correlation coefficient between the process indicators and the number of TVTAs and the passenger punctuality. The process indicators are dealing with disruptions and therefore with cancelled trains. Also, as stated in paragraph 5.3.4, the number of TVTAs not under the responsibility of TB / VL can also be influenced by these process indicators. Therefore all cancelled train TVTAs are used in the correlation calculation.

Table 5.14: The Pearson correlation coefficient between the number of TVTAs, the 3 and 15 minute passenger punctuality and its components and process indicator 13 to 15

Process indicator	**13**	**14**	**15**	**13**	**14**	**15**
TVTAs	-0.018	0.001	-0.027			
	3 minute			*15 minute*		
3 minute passenger punctuality	0.01	-0.07	-0.08	0.02	-0.1	0.05
Train delayed	0.00	0.05	0.16	-0.01	0.07	0.08
Train departed too early	0.03	0.17	-0.24	0.05	0.13	-0.22
Train departed too late	-0.02	0.02	0.02	0.00	0.01	0.00
Train did not depart	-0.07	0.10	-0.06	-0.07	0.11	-0.07
Train did not arrive	0.10	0.08	-0.11	0.08	0.10	-0.09
Missed transfer	0.04	0.03	-0.06	0.04	0.06	-0.10

For the first six months of 2015, the percentage of Blockage measures decisions with a Distribution decision is on average 75.57%. Nonetheless, in 69.27% of the days the value was 100% and in 22.35% of the days the value was 0%. This suggests a lack of variation and therefore no correlation with the number of TVTAs or the passenger punctuality. Table 5.14 shows indeed no correlation. As an additional check, for every blockage measures decisions the number of TVTAs per minute of lead time is calculated for both the cases with a Distribution decision and without. The null hypothesis is that the data in both groups comes from independent random samples with equal means. With a 95% certainty this null hypothesis is true. This means that there is no difference in the number of TVTAs if there is a Distribution decision or not. So there is definitively no correlation between this process indicator and the number of TVTAs.

There is no correlation between process indicator 14 and the number of TVTAs or the passenger punctuality as well. Nonetheless, the mean of the TVTAs per minute lead time for a blockage measure with a correctly filled in report card is 1.81 TVTAs per minute and for the other lead times the mean is 2.16 TVTAs per minute. These means are, with 95% certainty, different. This suggest that correctly and logically filled in report cards leads to more TVTAs so steering on this process indicator might not be effective to decrease the number of TVTAs. However in this last check the duration of the blockage measure is not taken into account. The mean of the numbers of TVTAs occurring during the whole blockage measure, is 156.8 TVTA per correctly and logically filled report card and 258.2 TVTA for the other report cards. So by a correctly and logically filled report card the number of TVTAs occurring per minute is higher than by the other report cards, however the total duration is smaller so the total number of TVTAs occurring is less. Note that it is possible that by report cards that where not correct or logical not all lead times where determined so the actual number of TVTAs occurring could be higher. Moreover, this conclusion does not state that steering on this process indicator will decrease the duration and thus the number of TVTAs occurring during a disruption. For the second model, based on the components applies that process indicator 10 does not have information about passengers, delays or lead times. It is therefore not surprising that there is no correlation with the components or the passenger punctuality. Nonetheless steering on this process indicator will make report cards more reliable which imply more reliable lead times by process indicators 7 to 12.

The effect of process indicator 15 on the performance of the rail network is unknown. As can be seen in Table 5.14, is has no influence on the passenger punctuality or on one or more components. An exception is the correlation with the component 'Train departed too early'. Here the correlation can be called 'weak'. However the number of trips with this component is low, on average 0.10% (see paragraph 3.1.2). This makes the outcome of this calculation not very reliable. The process indicator does not have any effect on the number of TVTAs as well. The number of TVTAs occurring per minute during lead times by unadapted Blockage measures is 2.12 TVTAs per minute. With a 95% certainty this is equal with the number of TVTAs occurring per minute during lead times by adapted Blockage measures. A last check is to compare the total number of TVTAs per report card with and without unadapted Blockage measures. Again, with a 95% certainty, the means of both groups is equal, so there if no effect on the duration of a disruption.

5.4. Comparison of the two models

The idea behind the model of ProRail and the NS to describe passenger punctuality is that steering on the processes that minimise the number of Explainable Train Deviations (paragraph 5.1.1) would increase the passenger punctuality. These TVTAs indicates discrepancies between the planned timetable and the execution. A discrepancy could be a cancelled train but also a rerouted train or a delayed train. Three disadvantages of this model can be distinguished.

1. As stated in paragraph 5.1.1, there is a very strong relation between this number and the 3 minute passenger punctuality. The Pearson correlation coefficient drops a bit by the 15 minute passenger punctuality, but still this correlation can be characterized as strong. However if focussing on only the business unit Transport control and Traffic control (TB / VL) this correlation between the passenger punctuality and the number of TVTAs is only moderate. Not all TVTAs are taken into account in de model as well, though the correlation between these neglected TVTAs and the passenger punctuality is strong. This suggests that neglecting these TVTAs might not be effective : by neglecting these TVTAs, a part of the rail network that could be improved in order to improve the passenger punctuality is neglected.
2. TVTAs are not weighted with the number of passengers affected by the TVTAs. This is a contradiction with the use of passenger punctuality above train punctuality: to make the interest of passengers top priority (see paragraph 2.3). As long as the TVTAs are not weighted with the number of passengers involved, steering to minimise the TVTAs is not the most effective method to increase passenger punctuality.
3. The focus of a signaller is safety and not the allocation of types of TVTAs to train deviations, so the TVTAs might not be very reliably. This is certainly the case by large disruptions, with many TVTAs and many delayed passengers as well.

The common factor in these three disadvantages is the use of TVTAs. Therefore a new model is proposed, where TVTAs are neglected (see paragraph 5.1.2. In this new model the passenger punctuality is split into components. The share of the component is based on the number of trips with that component, so the weight is included in this model. The process indicators relate to one or more components. The sum of the components equals the passenger punctuality, so the process indicators can also directly be related with the passenger punctuality as well. This is in contrast to the model of ProRail and the NS. In that model the sum of all the TVTAs does not equal the passenger punctuality. A disadvantage of the new model regarding the model of ProRail and NS is that the performance of the business units is not visible. Another drawback is that the correlation between the process indicators and the components or the passenger punctuality itself depends on the threshold value of a delay, so the correlation between the process indicators and the 3 minute passenger punctuality differs from the correlation with the 15 minute passenger punctuality.

ProRail and the NS composed a list of 20 process indicators within the business unit TB / VL (see Table 5.2). These process indicators are examined in paragraph 5.2. Steering on these indicators should affect the number of TVTAs or the passenger punctuality. Five of those process indicators are not fully developed yet and therefore not in use. A remark to process indicator 17 is that this indicator will not have an influence on the number of TVTAs or size of the components, but on the behaviour of the passengers. Steering on this indicator will therefore increase the passenger punctuality KPI but not the passenger punctuality itself.

For the remaining 15 process indicators, the correlation with the number of TVTAs and the passenger punctuality could be calculated. Note that the used TVTAs might differ between the process indicators, according to the goal of the process indicator. The results can be found in Table 5.15. Some specific notes can be given:

1. The Pearson correlation coefficient for the first process indicator (the number of mutations in VKL) is almost similar for both models. It is the process indicator with the strongest correlation with both the number of TVTAs and the passenger punctuality as well. Nonetheless this process indicator was intended to be an information-item and not an indicator that can be used to steer on [34].
2. Although for both models it is proven that a longer lead time leads to more TVTAs or more passengers with a delay, process indicators 7 to 12 do not reflect this correlation. This is because

Table 5.15: The Pearson correlation coefficient between the process indicators and the number of TVTAs and the 3 minute passenger punctuality (RPUN)

Process indicator	TVTA	RPUN	Note
1	0.61	-0.56	1
2	-0.43	0.4	
3	-0.42	0.48	
4	-0.51	0.49	
5	-0.19	-0.06	
6	-0.08	-0.04	
7	-0.017	-0.005	2
8	0.009	0.08	2
9	0.011	-0.02	2
10	0.019	-0.05	2
11	0.023	-0.26	2
12	0.11	-0.19	2
13	0.018	0.01	
14	0.001	-0.07	3
15	0.027	-0.08	

the process indicators are expressed as the 90th percentile of the lead times. Such a definition gives no information about the number of lead times and the overall duration. Together with the fact that there are multiple methods of calculating the percentile and the fact that the number of lead times during a day is too low to get reliable results, the need arises to change the definition of these process indicators.

For the model of ProRail and the NS the best solution is the total duration of the lead times, neglecting outliers. The correlation for all process indicators with this new definition will be strong. For the model based on the components, the best solution is the number of passengers departing during the lead time. Although the correlation is still weak to moderate, it is stronger than with the current definition.

3. Process indicator 14, the percentage of fully and correctly filled in report cards, has no correlation with both the number of TVTAs and the passenger punctuality. Nonetheless steering on this indicator will make process indicators 7 to 12 more reliable.

Based on the results, shown in Table 5.15, the two models seem very similar. Striking is the fact that none of the process indicators has a strong correlation with the number of TVTAs or the passenger punctuality. A large disadvantage of the model of ProRail and the NS is that the number of TVTAs gives no indication about the number of passenger involved. In the new model this disadvantages is resolved by not using TVTAs but components to divide the passenger punctuality. However all process indicators are still based on the number of trains involved. The result is that the mismatch between trains involved and passengers involved is shifted to a lower level. In other words, the results are similar because the TVTAs in the model of ProRail and the NS and the process indicators in the new model are not weighted with the number of passenger involved.

The solution is to redefine the definitions of the process indicators, where the outcome of the process indicator is weighted with the involved passengers. A preliminary start is already made in paragraph 5.3.4. Here a new definition for process indicator 7 to 12 for the second model was defined. The best solution was the sum of the departing passengers during a lead time. Although the correlation was still not strong, it was better than if the passenger where not taken into account in the definition. Besides, using the sum of departing passengers might be a bit too rough. It is possible to determine the number of passengers departing on every section of the rail network. Combining this determination with the section on which the Blockage measures apply leads to a more precise estimation of the involved passengers. If seasonality is taken into account, the estimation might get better as well. The expectation is that with a better estimation of the number of passengers involved the correlation with the passenger punctuality becomes stronger.

As stated, based on the found correlations the both models are very similar. Besides, the found correlations are moderate at best. This makes it difficult to claim that one model does describe the passenger punctuality better than the other. However the correlations in the model of ProRail and the NS are with the number of TVTAs. Decreasing the number of TVTAs will not necessarily increase the passenger punctuality with the same rate. This is in contrast to the model based on the components. If the size of one or more component will decrease, the passenger punctuality will increase with the same amount. This is an advantage the model of ProRail and the NS does not have. Nonetheless, as long as the process indicators do not include the number of passengers involved and the correlations do not change, the effect of both models can be improved.

For both models apply that the model is a simplification of the reality. The real performance of the rail network, and especially the business unit Transport control and Traffic control, is in reality different than the process indicators suggest. This implies that steering on the process indicator will never lead to a perfect rail network, or a passenger punctuality of 100%. Nonetheless this is no reason not to use these models. Correlations present in these models are also present in reality, although the Pearson correlation coefficient might differ. Besides these models help to evaluate and steer more effectively on the involved (business) units and are at this moment the best option to describe passenger punctuality.

6

Process indicators for a subset of the rail network

The official passenger punctuality KPI gives the performance of the whole network. Chapter 4 showed that the passenger punctuality can also be calculated for subsets of the rail network. A subset can be a station, a train series or an origin-destination combination. By calculating the passenger punctuality for a subset can reveal problems in the network that where not visible via the nationwide passenger punctuality. In chapter 5 two models are discussed that can describe the passenger punctuality in order to improve it. Both models use the process indicators, belonging to the Transport control and the Traffic control, to steer on the passenger punctuality. The correlation between these process indicators and the number of Explainable Train Deviations, the components and the passenger punctuality is determined. However this correlation is based on the entire network and might differ for a subset. Using the correlation of the entire network to steer on a subset of the network might therefore not be the most effective method to increase the passenger punctuality. At this moment it is not possible to calculate the process indicators for a subset of the rail network. Therefore an algorithm was created that is able to do this.

In order to calculate the process indicators, as described in paragraph 5.2, ProRail and the NS produces several data files each day. With this data, the values of the process indicators of the previous day can be calculated. Combining these daily files can give a process indicator for a longer period of time. The first part of the algorithm, created in this research, enriches each data file with some additional data. This is necessary because the data files do not contain all the information needed to calculate the process indicators for a subset of the rail network. This part of the algorithm is mostly written in MATLAB, but a small part is written in Structured Query Language. The second part of the algorithm, written entirely in MATLAB, can calculate the process indicators for a specified subset. In paragraph 6.1 the options to specify a subset are explained. Paragraph 6.2 describes in highlights the process to enrich the data files, which other data sources are needed and which part of the data can be enriched. In appendix D the algorithm can be found.

6.1. The options to create subsets

A subset of the rail network can be a station, a (part of a) route or a moment in time. A combination of these factors is possible as well. To ensure that for every subset every process indicator can be calculated, three options are distinguished. These options can be used to define the subset. Each option can add an additional constrain to the data used to calculate the process indicators. Altogether the constrains form the subset of the rail network.

1. **Service control points:** A service control point (in Dutch: Dienstregelpunt) is an important geographical point on the rail network. This can be a bridge, a rail yard, a place where a railway line branches off a main line or a station. Via this option, a specific point of the rail network can be examined. By the way of multiple points an area can be examined. In this last case, not every service control point in the area is needed, only the points at the border of the area and the

arrival and departure station of every train series running in that area. However to minimise the error, at least all stations in the area should also be included in the list of service control points.

An addition to this option is the possibility to define if one of these service control points is the start point of a train series. This has only effect on the process indicators concerning the right composition of the train and the right number of personnel at the morning start.

If multiple service control points are given instead of only one, at least two service control points should match the service control points present in the data files of the process indicator to use this data to calculate the process indicators. An exception is the data concerning train handling documents (TAD). If a train handling document applies to only one station, the number of service control points involved is also one. This makes it impossible to have at least two service control points in common with the examined area. As an addition, it is also possible to indicate that every service control point of the given area should be in common with data of the process indicators.

2. **Train series:** Via this option, the process indicators can be calculated for a single train series only. The direction of the train series (even or uneven) can also be distinguished, although this is not a prerequisite. It could occur that a disruption takes place during night time on a section that is on the route of the examined train series. Nonetheless the disruption is at night, so the train series do not have a planned passage of the section. It is therefore arguable that this disruption should not be used in calculating the associated process indicators. Therefore as an addition, it can be indicated if such situations should be taken into account or not. This applies not only to disruptions but also to the process indicators concerning empty rolling stock runs as well.

3. **Time of day:** A distinction can be made between the whole day, peak hours or off peak hours. It is also possible to define a more specific time period. Note that this option does not apply to the process indicators regarding the right composition of the train and the right number of personnel at the morning start.

6.2. The enriched data files

The process indicators can be calculated with data from six different data sources: VSM, PS, MS, TAD, VKL and VLM data. For all these data sources applies that the content of the data is enough to calculate the process indicators for the whole rail network, but not for a subset. In other words, each data source needs to be enriched with additional information in order to calculate the process indicators for a subset of the rail network. The NS is responsible for the PS and the MS data. The systems that generate these data files are B@P and VKL MAT respectively. The basis for the other four data sources, under the responsibility of ProRail, is three systems, called VKL/VOS, TROTS and ISVL. Another program, Sherlock, combines the output of these systems to produce the VSM, TAD, VKL and VLM data.

This paragraph will describe each data source, in particular the information stored and the extra information needed to calculate the process indicators for each subset. For the enriching of the original data sources, three additional sources were used, though not every additional source is used to enrich every original data source.

1. The ribbon data (in Dutch: Lintdata) of each train number. This source, an output of Sherlock, contains the planned and realised route for each day and each train number. The route is represented with a series of service control points, whereby every service control point the action of the train is shown: arrival, departure or passage. This data is available for ever train, so not only the passenger trains.
2. The ribbon data of each train series, as generated by Sherlock. If a train number is not known, this data source can be used. It contains per day the most frequent and the longest route of each train series. A route is represented with a series of service control points, whereby every service control point the action of the train is shown: arrival, departure or passage. The longest route is the route with the most service control points, not the route with the largest distance. This data is based on the realisation data only and not on the planned timetable. Using this information source instead of the ribbon data of each train number (see item 1) generates the assumption that the route of the individual train is the same as the most frequent route or the longest route of the corresponding train series.

3. The table 'EZ_D_RPU.FCT_RPUN_REISDEEL', one of the output tables of the passenger punctuality calculation. In this table the planned and realised departure and arrival times of every train number for every possible part of the route is stored. Note that this table only contains information of trains from the NS. This implies that if this data source is used, the process indicator does not contain data from other rail operators.

This paragraph ends with an overview of the percentage of data that can be enriched. The outcome depends on the data source. The exact algorithm to enrich the data sources can be found in appendix D.1.

VSM data

Nine process indicators can be calculated with the VSM data. This data contains information about the report cards, used by large disruptions. With this data the lead times between key moments in the process can be calculated. Also the percentage of Blockage measures decisions with a Distribution decision, the percentage of fully and correctly filled in report cards and the percentage of unadapted Blockage measures can be calculated.

Per report card the time period when the disruption occurred is known. Note that if the report card in not filled in correctly, or if the time when all trains runs according to the Blockage measures or to regular plan is missing, this time period might be shorter than in reality. Also the two service control points, giving the location of the disruption, is most of the time known. However it can occur that this information is missing. In this case the corresponding data cannot be enriched.

Based on the location where the disruption happened and the ribbon data per train number, the numbers of the trains passing the section of the disruption can be found. This information can be combined with the realisation data of all trains and the time the disruption occurred. The results are the trains that run by the disruption or that were cancelled during the disruption. In this last case, the conclusion is made that the train was cancelled because of the disruption.

The final result is the VSM data enriched with the involved train numbers (and thus train series) for almost every report card. Also the planned and realised route of each train number is known plus the fact if a train was cancelled.

PS data and MS data

The process indicator that calculates the percentage of trains with the right number of personnel at the morning start retrieves information from the PS data. The original data indicates per train series the planned and realised number of staff members. This indicator is time independent, so information about time is not needed. This implies that the data needs only be enriched with the route of the train series. This route can be found with the ribbon data per train series.

The layout and content of the MS data is very similar with the PS data. The MS data is used to calculate the percentage of trains with the right composition. Here also only the route of each train series needs to be known. For both data sets applies that the train series is used to find the route instead of every train number. If the actual route is not the most frequent or the longest route, this information cannot be retrieved and thus a small discrepancy between the enriched data and the reality occurs.

TAD data

The TAD data contains all train handling document (TAD) situations. The relevant information is the name of the TAD and the train that triggers the use of a TAD. In the name of the TAD information about the involved train series and the location of route where the TAD applies can be found. Note that information about time is not present. Another difficulty in enriching this data is the fact that it is possible that not every train series present in the name of the TAD is involved.

One of the few certainties is the involvement of the train that triggers the TAD. With this train number, the route of where the TAD applies and the ribbon data per train number the exact section of the TAD can be found. Next for every possible involved train series the arrival and departure times on station in this section is determined, based on the realisation data from the passenger punctuality calculation. Based on the time difference with the one certain train involved a determination can take place which train series or numbers are involved in the TAD. Here the planned arrival or departure times are used.

If all involved train numbers are known, the route for each train number can be found with the ribbon data. The minimum of all departure and arrival times of the involved train numbers on the section where the TAD applies is the start time of the TAD. The maximum is the end time. Here a distinction is made between the planned departure and arrival times and the realised departure and arrival times.

At the end, the enriched TAD data consist of the expected route where the TAD applies, the involved train numbers and train series and the time interval when the TAD applies. Note that because of the little available initial data, the data cannot be enriched without making some assumptions:

- Only the section of the route that all involved trains has in common is used to determine the arrival and departure times.
- If there are multiple trains from the same train series possible involved, the smallest difference in planned departure or arrival time with the only train certainly involved is used to make a choice between the options.
- The most frequent route of a train series is used. If the section where the TAD applies does not lie on this route, the route of that train series with the most service control points is used. If this route is also a mismatch with the section of the TAD, the train series is neglected.

The planned and realised departure and arrival trains are used as well. This data is retrieved from the 'EZ_D_RPU.FCT_RPUN_REISDEEL'-table. This table stores only data about trains from the NS.

VKL data

Every mutation in the VKL system is stored in the VKL data. This mutation consists of the mutation time and the involved train number. Only the location on the rail network is missing. So this additional information is needed to use this data source to calculate the process indicator for a subset of the rail network.

The route of the train that had a mutation can be found via the train number and the ribbon data per train number. In comparison with the other data sources, the high number of train numbers without a found route is striking. A remark to this is the fact that most trains without a found route are freight trains or empty rolling stock runs. Almost every passenger train can be linked with a ribbon of service control points.

VLM data

The data about empty rolling stock runs can be found in the VLM data. With this data the departure punctuality for empty rolling stock runs at rail yards and at stations can be found. The data consist of a train number, the departure service control point, the planned and realised departure time and the departure delay.

With the train number and the ribbon data per train number the route of each empty rolling stock run can be found. This route can be compared with every route of every passenger train in the ribbon data. If at least two service control points match, the conclusion is made that the passenger train did run (partly) along the same route of the empty rolling stock run.

The next step is to make a determination which of the found passenger trains did run together with the empty rolling stock. Hereby the planned and realised departure and arrival times are used. For each train the time period is calculated when the train did run on the same location as the empty rolling stock. This time period is compared with the departure time of the empty rolling stock run. If the departure time lies in the time period, the conclusion is drawn that the passenger train might be affected by the empty rolling stock trip.

Note that the method to determine if a passenger train is affected by an empty rolling stock method is not solid. If the matching section in the routes is at the end of the route of the empty rolling stock, the empty rolling stock run might pass this section long after its departure. It is unlikely that the passenger train that did run during the departure time on that section is really affected by the empty rolling stock run. Nonetheless the chance the corresponding train series is affected is higher. So this method distinguishes the train series affected by determining the train numbers possible affected.

Completeness of the enriched data

It is possible that not every row of data can be enriched, or not fully enriched. A reason can be that the original data or the additional data is not complete or correct. It is also possible that information of rail operators other than the NS is missing because this information is based on planned and realised departure and arrival times. Note that this data was stored in the 'EZ_D_RPU.FCT_RPUN_REISDEEL'-table. This table contains only information about trains of the NS. A last reason can be a bug or a missing exception in the algorithm to enrich the data. Some data is initially filled in by hand, so there might be some differences in the layout. Although the algorithm can handle a variety of types of input, it might occur that some data cannot be read, leading to not fully enriched data.

Table 6.1 shows for every data source which part can be fully enriched, which part can be partly enriched and the remainder. These values are based on the data of August, September and October of 2015. Note the large share of the PS data that cannot be enriched. This is mostly due to the fact that for the same share of personnel is it unknown which train series they are staffing. Another striking value is the share of partly enriched VLM data. This can be explained by the fact that for every empty rolling stock run only the departure time is known. This makes it complicated to find the involved train series.

Table 6.1: The size of the (partly) enriched data, per data source

Data file	**Fully enriched**	**Partly enriched**	**Not enriched**
VSM	99.41%	0.59%	0.00%
PS	79.34%	2.53%	18.13%
MT	88.80%	4.31%	6.89%
TAD	84.63%	15.36%	0.01%
VKL	51.93%	48.07%	0.00%
VLM	40.93%	59.07%	0.00%

7

Case study: Utrecht - Leeuwarden / Groningen

In chapter 5 two models describing passenger punctuality were analysed. In the first model, as developed by ProRail and the NS, the passenger punctuality is divided into "Explainable Train Deviations" (TVTAs), mostly cancelled and delayed trains. In the second model, the passenger punctuality is split up in the components: the reason why passengers are delayed, according to their perspective. For both models applies that the TVTAs or the components are related to the process indicators under the responsibility of Transport control and Traffic control (TB / VL). However this correlation is calculated for the entire network and might differ for a subset of the network.

Chapter 6 introduced an algorithm that is able to calculate the process indicators for a subset of the rail network. Such a subset can be a geographical area of the rail network, a train series or a moment in time, like rush hour. Herewith it becomes possible to determine the correlations in both models these subsets. Therefor this chapter will focus on a subset of the rail network: the part of the rail network between station Utrecht Centraal and stations Leeuwarden and Groningen. In paragraph 7.1 this area is elaborated. A breakdown of the passenger punctuality (see chapter 4 of this study area is performed as well. In paragraph 7.2 the correlations between the process indicators and the number of TVTAs, the components and the passenger punctuality are calculated. The outcome is compared with the correlations of the entire network. Finally some conclusions are drawn, described in paragraph 7.3.

7.1. Description and the passenger punctuality of the study area

The area of the case study is the main railway line to the north-east of the Netherlands. This line is schematically visualized in Figure 7.1. The line starts in the largest station in the Netherlands, Utrecht Centraal (Ut). After station Amersfoort and Zwolle the line splits into two branches near station Meppel (Mp). The western branch ends at station Leeuwarden (Lw) and the eastern branch ends at station Groningen (Gn). At all 5 intercity stations there are side branches of the rail network in multiple directions. There are also two sections splitting off between stations Utrecht Centraal and Amersfoort and one just before station Groningen. These side branches are left out of the study area. Note that not every station in the area is visualized, but they are all taken into account.

A subset of the rail network is examined, so there is a complication when calculating the passenger punctuality and the size of the components. Some passengers travel from a station in this study area to another station outside of the study area or vice versa. Three solutions can be distinguished.

1. All passengers that travel in this study area are taken into account. This also means that passengers travelling only partly in the area are taken into account as well, even if they end up with a delay in another part of the rail network.
2. For every passenger the train that causes his or her delay is determined. If that train runs in the study area, the passenger is taken into account in the punctuality calculation. If a passenger does not have a delay, their last train is used. This will result in lower passenger punctuality than

in reality because relative more passengers with a delay are taken into account.

3. Only passenger departing and arriving at one of the stations in the study area are taken into account. The advantage of this method is that no causes for delay outside the case area influence the passenger punctuality. A drawback is that the total number of passengers considered is lower than the reality, although the passenger punctuality is based on percentage of passengers and not on the total amount.

In this case study solution 3 is chosen, because it is presumed to give the most realistic results. These results are given in Table 7.1. As can be seen, the overall passenger punctuality is better than the average of the whole rail network. This is mostly due to less trips with the components 'Missed transfer', 'Train did not depart' and 'Train did not arrive'. These last two components suggest that fewer trains are cancelled in this study area than on average. However it is also possible that the number of passengers per train is less than average for the entire network.

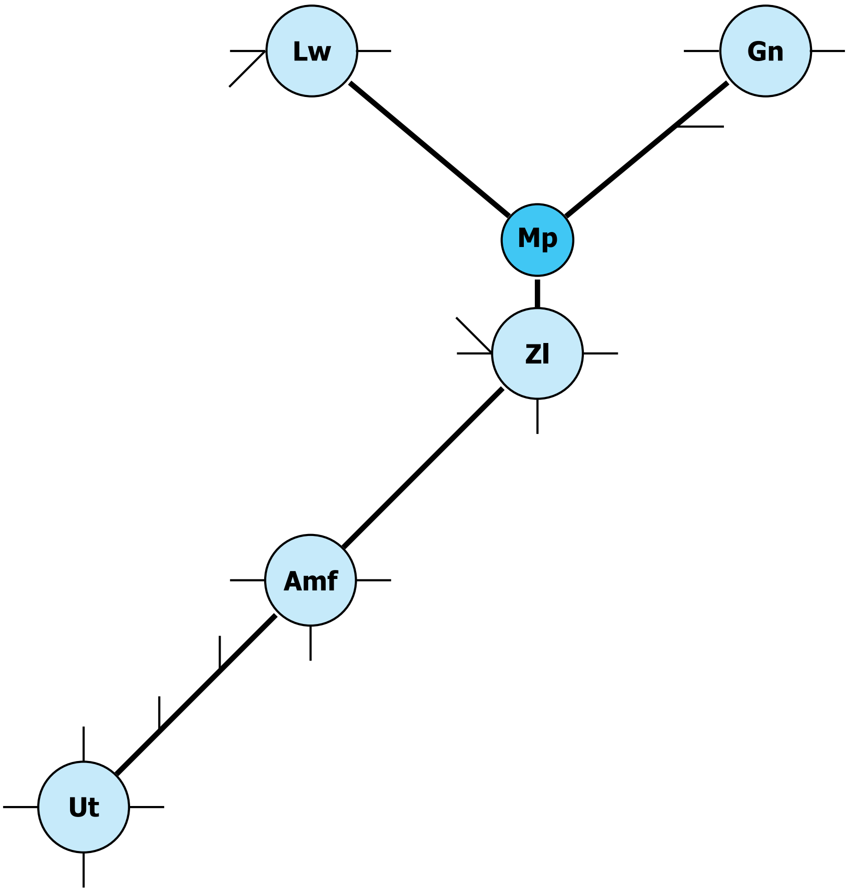


Figure 7.1: The part of the Dutch rail network between Utrecht (Ut) and Leeuwarden (Lw) and Groningen (Gn)

The breakdown graph can be found in Figure 7.2. Here can be seen that the largest share of delayed trips at the 3 minute passenger punctuality is due to delayed trains. By a delay threshold of 15 minutes, this share is decreased enormously while the shares of the other components are more or less stable. Thereafter each share starts decreasing gradually. An exception is the share of trips with the component 'Train did not arrive'. Although this share does decrease, the decline is less than by the other components. At a delay threshold of 30 minutes, the size of the share is almost the same as by delay threshold of 15 minutes.

7.2. The correlation with the process indicators

In chapter 5 the process indicators are analysed. Here the correlation with the performance of the whole rail network was calculated. This might differ from the correlation for a subset and in particular with this study area. Although some conclusions were made about the definition or the use of some

Table 7.1: The 3 and 15 minute passenger punctuality and the size of the components for the study area and the whole rail network between August and October 2015

	Case area		**The Netherlands**	
	3 minute	*15 minute*	*3 minute*	*15 minute*
Passenger punctuality	86.23%	98.43%	83.20%	96.94%
Train delayed	12.29%	0.32%	12.64%	0.45%
Train departed too early	0.01%	0.00%	0.03%	0.01%
Train departed too late	0.40%	0.33%	0.63%	0.43%
Train did not depart	0.70%	0.24%	1.55%	1.07%
Train did not arrive	0.27%	0.24%	1.55%	0.34%
Missed transfer	0.10%	0.08%	1.48%	0.75%

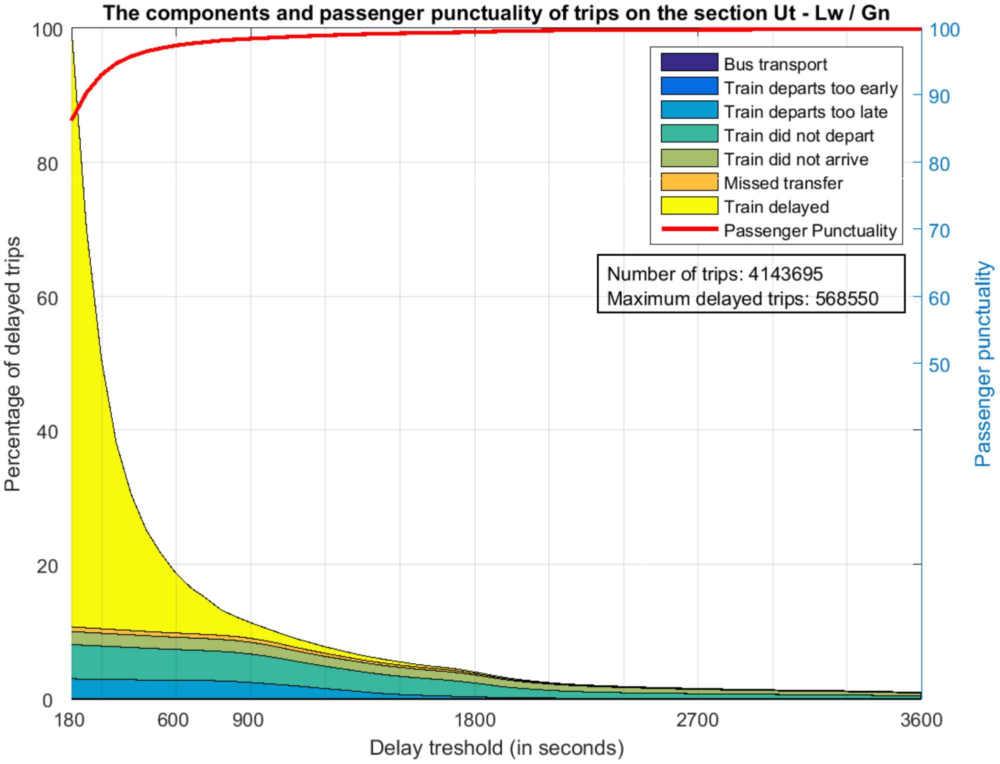


Figure 7.2: The passenger punctuality and the shares of the components for trips on the study area Ut - Lw / Gn for August to October 2015

process indicators, in first instance the correlation with all original process indicators is calculated. Two process indicators are not used.

3. **Percentage of Blockage measure decision with a Distribution decision** For all examined days, this percentage was 100%. This implies that the correlation cannot be calculated.
13. **Departure punctuality of empty rolling stock at six rail yards** None of the six rail yards is in, or near, the study area.

The remaining 13 process indicators used in this case study are given in Table 7.2.

Table 7.2: The available process indicators from the model of ProRail and the NS [37]

	Process indicator
1	Order processing VL-phase
2	Percentage of the correct use of the Train handling document
4	Departure punctuality of empty rolling stock at stations
5	Percentage of trains with the right composition at the morning start
6	Percentage of trains with the right number of personnel at the morning start
7	Lead time between 'Known VL' and 'Known BackOffice'
8	Lead time between 'Known BackOffice' and alarm
9	Lead time between alarm and Distribution decision
10	Lead time between Distribution decision and Blockage measure decision
11	Lead time between Blockage measure decision and running according to the Blockage measures
12	Lead time between end of Blockage and running according to the regular plan
14	Percentage of fully and correctly filled in report cards
15	Percentage of unadapted Blockage measures

The model of ProRail and the NS

In the model of ProRail and the NS, the process indicators are related to the number of Explainable Train Deviations (TVTAs), see paragraph 5.1. The group of TVTAs taken into account differs between the process indicators. In practice this means that in the correlation calculation with process indicators 1 and 7 to 15 a different set of TVTAs is used than by the remainder of the process indicators.

The analysed model was using only the TVTAs under the responsibility of the business unit Transport control and Traffic control (TB / VL). So in first instance, only these TVTAs are used. Process indicators 7 to 15, dealing with large disruptions, are an exception. Here train deviations are often given an explanation based on the cause of the disruptions, while the number of these TVTAs might differ based on the performance of TB / VL. TVTAs occurring due to a disruption are often not under the responsibility of TB / VL, so all TVTAs are used to determine the correlation.

Four categories of TVTAs can be distinguished, whereof mainly two apply to passenger trains: Cancelled trains and Delayed trains. Process indicators 7 to 15 are mainly based to minimise the number of cancelled trains. So only these TVTAs are used in the correlation calculation. Process indicator 1 does not have a direct relation with only delayed trains or only cancelled trains. So here all TVTAs are used. For the other process indicators applies that the main focus is to decrease the number of delayed trains.

Table 7.3 shows the Pearson correlation coefficient between the process indicators and the number of TVTAs for the study area and whole rail network. Here the group TVTAs used is mentioned as well. A Pearson correlation coefficient close to 0 means that there is no correlation while a coefficient of 1 or -1 means a perfect correlation. Some differences between the study area and the whole rail network can be distinguished. There is an almost moderate correlation between the percentage of the correct use of the train handling documents (process indicator 2) and the number of TVTAs in the whole rail network, but in the study area this correlation is non-existent. The same applies to process indicator 4: the departure punctuality of empty rolling stock at stations. For three process indicators the opposite applies: the correlation in the whole rail network is non-existent while in the study area the correlation is moderate. These process indicators are the lead time between a Blockage measure decision and running according to the Blockage measure (process indicator 11), the lead time between the end of the Blockage and running according to the regular plan (process indicator 12) and the percentage of unadapted Blockage measures (process indicator 15).

Table 7.3: The Pearson correlation coefficient between the process indicators and the number of TVTAs for the study area and the whole rail network

Process indicator	**Study area**	**The Netherlands**	**TVTA**
1	0.59	0.61	All TB / VL TVTAs
2	0.06	-0.43	Delay TB / VL TVTAs
4	0.09	-0.51	Delay TB / VL TVTAs
5	-0.04	-0.19	Delay TB / VL TVTAs
6	0.05	-0.08	Delay TB / VL TVTAs
7	0.11	-0.017	All Cancelled trains TVTAs
8	0.01	0.009	All Cancelled trains TVTAs
9	0.00	0.011	All Cancelled trains TVTAs
10	0.15	0.019	All Cancelled trains TVTAs
11	0.35	0.023	All Cancelled trains TVTAs
12	0.48	0.11	All Cancelled trains TVTAs
14	-0.04	0.001	All Cancelled trains TVTAs
15	-0.44	0.027	All Cancelled trains TVTAs

The model based on the components

The process indicators can also be linked directly with the passenger punctuality and its components (paragraph 5.1). In Table 7.4 the Pearson correlation coefficients between the 3 minute passenger punctuality, the components and the original process indicators are given. The correlation of the passenger punctuality of the study area with process indicators 11 and 12 (the lead time between a Blockage measure decision and running according to the Blockage measure and the lead time between the end of the Blockage and running according to the regular plan) are a bit stronger than for the whole rail network. The same applies to the percentage of fully and correctly filled in report cards (process indicator 14). The largest difference can be found by process indicator 4, the departure punctuality of empty rolling stock at stations. In the whole network the Pearson correlation coefficient with this process indicator is 0.49 while in the study area the coefficient was 0.10: a non-existing correlation.

By examining the correlation with the components, more differences can be found. Some are due to a lack of data. This applies to the differences by the component 'Train departed too early'. Too few trips had this component in the study area to conclude some changes in correlation. However a remarkable result is that the correlation coefficient between the component 'Train departed too late' and process indicator 8 (the lead time between 'Known VL' and 'Known BackOffice') in the study area is 0.47: a moderate correlation. This same correlation in the whole network is almost non-existing. The same applies to the correlation between the component 'Train did not arrive' and process indicator 12.

New process indicators

In paragraph 5.3.4 an alternative for process indicators 7 to 12 is given. Here the sum of all lead times per day is used to calculate the Pearson correlation coefficient with the number of TVTAs occurring during the lead times. Also process indicators 7 to 11 were combined because together they form the lead time of the reduction of the timetable after a disruption.

Table 7.5 gives the Pearson correlation coefficient for this alternative definition. For the process indicator 7 to 11 concerning the reduction of the timetable, the correlation is much stronger. This is also visible in the combined process indicator 7-11: this correlation is strong for the study area, while it is only moderate for the entire rail network. For the last process indicator, the restoration of the timetable after the disruption ended, the correlation is the same.

In the second model, based on the components, an alternative was proposed for process indicators 7 to 12 as well (paragraph 5.3.4). In this new definition the process indicator would be the expected number of passengers involved during all lead times. The expected number of passengers was all passengers departing during the lead time. For the study case, this estimation can be improved by

Table 7.4: The Pearson correlation coefficient between the 3 minute passenger punctuality and the process indicators

Study area							
Process indicator	**1**	**2**	**4**	**5**	**6**	**7**	**8**
Passenger punctuality	-0.58	0.33	0.10	0.014	-0.044	-0.072	-0.097
Train delayed	0.47	-0.40	-0.031	-0.013	0.093	0.11	0.044
Train departed too early	0.24	-0.055	0.017	-0.026	-0.23	0.26	-0.14
Train departed too late	0.32	-0.19	-0.16	-0.004	-0.11	0.034	0.43
Train did not depart	0.57	-0.022	0.19	-0.080	-0.14	-0.037	0.093
Train did not arrive	0.46	0.11	-0.21	0.066	-0.021	-0.006	0.088
Missed transfer	-0.062	0.047	0.034	0.19	0.30	-0.088	-0.013
Process indicator	**9**	**10**	**11**	**12**	**14**	**15**	
Passenger punctuality	0.027	0.16	-0.39	-0.36	-0.32	0.073	
Train delayed	0.019	-0.19	0.20	0.19	0.43	0.061	
Train departed too early	0.129	-0.18	0.53	0.27	0.077	-0.093	
Train departed too late	0.053	-0.14	0.29	0.19	0.031	-0.10	
Train did not depart	-0.093	-0.033	0.58	0.55	-0.053	-0.32	
Train did not arrive	-0.11	-0.022	0.53	0.44	0.013	-0.23	
Missed transfer	-0.088	-0.060	-0.058	0.005	0.15	0.11	
The Netherlands							
Process indicator	**1**	**2**	**4**	**5**	**6**	**7**	**8**
Passenger punctuality	-0.56	0.40	0.49	-0.063	-0.041	-0.005	0.080
Train delayed	0.51	-0.39	-0.52	0.055	0.024	0.033	-0.10
Train departed too early	-0.21	-0.009	0.15	0.046	0.058	-0.11	-0.034
Train departed too late	0.49	-0.35	-0.36	0.023	-0.027	-0.039	-0.041
Train did not depart	0.64	-0.25	-0.22	0.003	0.066	-0.045	-0.054
Train did not arrive	0.037	-0.19	-0.20	0.16	0.062	-0.066	0.096
Missed transfer	-0.090	-0.31	-0.35	0.17	0.096	0.021	0.026
Process indicator	**9**	**10**	**11**	**12**	**14**	**15**	
Passenger punctuality	-0.020	-0.046	-0.26	-0.19	-0.070	-0.085	
Train delayed	0.083	0.057	0.20	0.15	0.050	0.16	
Train departed too early	-0.19	-0.10	0.007	-0.14	0.17	-0.24	
Train departed too late	-0.058	0.008	0.18	0.15	0.015	0.018	
Train did not depart	-0.10	0.051	0.31	0.30	0.10	-0.059	
Train did not arrive	-0.040	-0.008	0.28	0.054	0.075	-0.11	
Missed transfer	-0.053	-0.11	0.18	-0.080	0.031	-0.056	

Table 7.5: The Pearson correlation coefficient between the sum of the lead times and the total number of cancelled trains TVTAs occurring during the lead times for the study area in August to October 2015

Process indicator	**Study area**	**The Netherlands**
7	0.64	0.35
8	0.85	0.53
9	0.78	0.70
10	0.66	0.56
11	0.97	0.84
12	0.71	0.73
7-11	0.86	0.58

taking only the passengers departing and arriving in the study area. This will be an underestimation of the total number of passengers in the study area, but the goal is not to find the exact number of passengers, only the relative distribution in time.

The outcome of this alternative for the study area can be found in Table 7.6. All correlations are weak or non-existing and mostly a little smaller than for the entire network. This difference might be due to the expected number of passengers involved. This number has been calculated differently for the case study.

Table 7.6: The Pearson correlation coefficient between the 3 minute passenger punctuality and its components and an alternative for process indicators 7 to 12: the sum of the expected passenger involved during the lead times

Process indicator	**7**	**8**	**9**	**10**	**11**	**12**	**7-11**
Passenger punctuality	-0.046	-0.003	-0.11	-0.011	-0.24	-0.25	-0.13
Train delayed	0.074	-0.018	0.18	0.014	0.187	0.16	0.14
Train departed too early	0.040	-0.166	0.091	-0.13	0.079	-0.037	0.071
Train departed too late	0.034	0.30	0.089	0.024	0.27	0.36	0.023
Train did not depart	0.001	-0.006	-0.017	0.031	0.24	0.30	0.069
Train did not arrive	-0.042	-0.012	-0.076	-0.026	0.21	0.22	0.098
Missed transfer	-0.103	-0.048	-0.081	-0.073	-0.10	-0.089	-0.051

7.3. Conclusion

For both models, analysed in chapter 5, the correlation with the process indicators for the part of the rail network between station Utrecht Centraal and stations Leeuwarden and Groningen is calculated. For most correlations, the difference between the study area and the whole network is relatively small. The conclusion here is that the influence of these process indicators on the number of TVTAs or on the passenger punctuality does not differ between the rail network and the study area.

On the other hand, for some correlation the difference is too large to be unnoticed. An example is the correlations between process indicators 7 to 12 and the number of TVTAs or the passenger punctuality. In both cases, the correlation in the study area was stronger than in the entire rail network. An explanation is that the number of lead times per day in the study area was on average only 1 to 3. This is too few to state some hard conclusions about the correlation.

Another remarkable result is that the influence of empty rolling stock runs (process indicator 4) is in the study area is far less than in the whole network. This indicates that results found in chapter 5 can change in detail for a subset of the rail network. For example, steering on the departure punctuality of empty rolling stock runs to improve the passenger punctuality might be effective overall, but in this study area it will not have any effect.

In chapter 5 the outcome was that both models are very similar. The correlations between the process indicators and the number of TVTAs and the passenger punctuality respectively were comparable. This differs in this case study. An example is process indicator 2: the percentage of correctly used Train Handling documents. While for the model of ProRail and the NS the correlation is non-existing (the first model), in the second model (based on the components) the correlation is almost moderate. This can be explained via the fact that less TVTAs does not directly imply higher passenger punctuality. This would suggest that the second model describes the passenger punctuality better than the first model.

8

Conclusions and recommendations

After the passenger punctuality KPI is analysed (chapter 3 and 4) and the models to describe the passenger punctuality are analysed (chapter 5 and 7) the answer to the main research question can be formulated. This is done in paragraph 8.1, where a small overview of the research is given as well. A deeper elaboration of the conclusion of this report can be found in paragraph 8.2. In the reflection (paragraph 8.3) the scope of this research is widened. From new perspectives some comments regarding the passenger punctuality and this research are given. This chapter ends with some recommendations regarding practical implication and further research of the passenger punctuality KPI and the models in paragraph 8.4.

8.1. Answer to the research questions

In order to achieve an answer to the main research question, five sub questions were formulated in chapter 1. The remainder of this report is an elaboration of these sub questions. This paragraph will formulate in short the answers to the sub questions to, in the end, find the answer to the main research question.

1. *What are the components of passenger punctuality?*

Passenger punctuality is presented as the percentage of the passenger arriving on time at their final destination (chapter 2). This slightly differs from the calculated passenger punctuality. This is the percentage of promised journeys that could be completed and where the last train arrived on time. A promised journey (in Dutch: Reisbelofte) is the travel option that a passenger was expected to take. The difference is that a passenger might travel differently from the promised journey or that the promised journey starts or ends at stations that are not the desired departure or arrival stations.

In both definitions the phrase 'on time' is used. The exact meaning can differ. In the Key Performance Indicator (KPI) currently in use, on time means within 5 minutes of the planned arrival time. In the new passenger punctuality KPI this threshold value is set at 3 minutes. So passengers arriving within 3 minutes of the planned arrival are considered on time. Another interesting threshold value is 15 minutes. By a delay of less than 15 minutes, the trip is considered delayed. If the delay exceeds this threshold value, the trip becomes disturbed. In other words, the rating of the performance drops.

A component is a reason why a trip is delayed, from a passengers' perspective (chapter 4). Delayed trips belong to a component. This implies that the sum of the passenger punctuality and the components is 100%. In the calculation of the passenger punctuality every trip is categorized. If a delay is at least the delay threshold value, the trip will belong to the component that with the same name as its category. Hereby is one exception. If a trip can be made as promised, the category is 'Trip realised' while the corresponding component is 'Train delayed'. Two categories ('Other rail operator' and 'Bus transport') are not taken into account because these trips are not used to calculate the passenger punctuality.

The six components of passenger punctuality are the following:

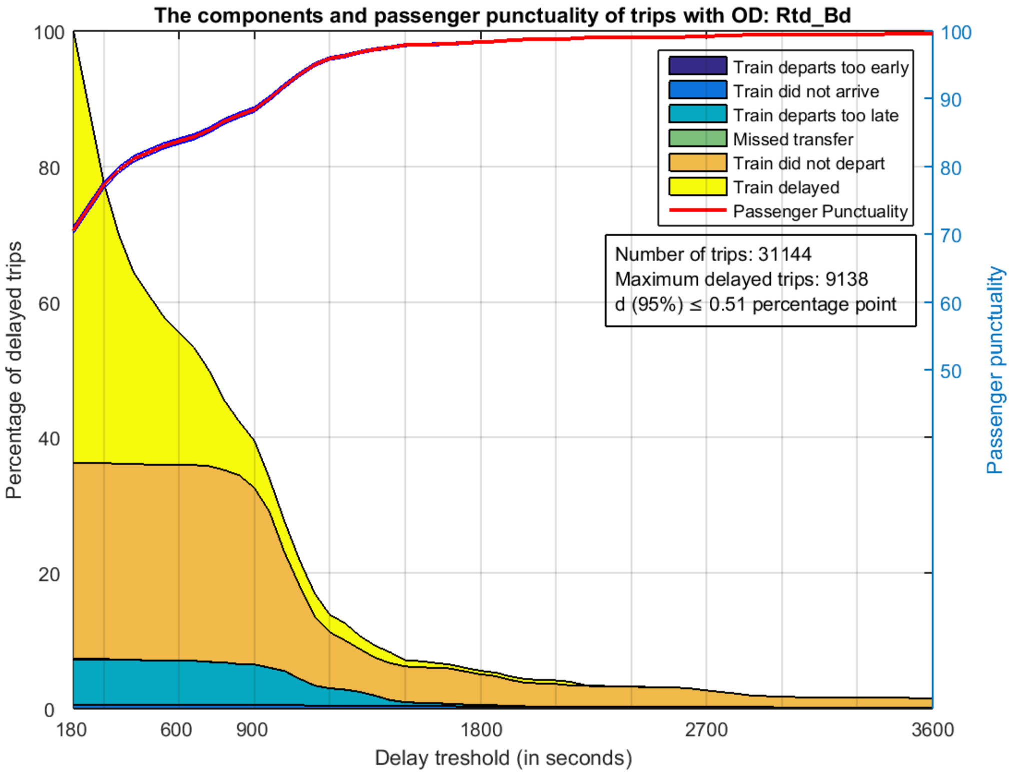


Figure 8.1: The passenger punctuality and the shares of the components for trips between Rotterdam Centraal and Breda

1. **Train departs too early** Every trip that is delayed because a promised train departed too early.
2. **Train departs too late** Delayed trips because a train departed at least 15 minutes later than scheduled.
3. **Train did not depart** Trips that are delayed because the promised train did not depart.
4. **Train did not arrive** If a train is cancelled during its journey, passengers in this train might be delayed.
5. **Missed transfer** Delayed trips because the walking time at a transfer station was less than the time difference between the realised arrival and departure.
6. **Train delayed** Every trip delayed because the train was delayed when arriving at the final destination of the passengers.

2. *How can the passenger punctuality be disaggregated to components?*

Of every trip the departure and arrival station are known, as well as the planned and realised departure and arrival times. This information can be used to disaggregate the passenger punctuality to the components. It is thus, for example, possible to calculate the passenger punctuality and the size of the components for the trips departing, arriving or even transferring at every station in the rail network.

In the available data of every trip, the promised journey is included. This promised journey consists not only of the departure, arrival and possible transfer station but the train numbers as well. It is therefore possible to perform a breakdown of the passenger punctuality for a train series or origin–destination combination.

Combining this information can lead to even more detailed passenger punctuality breakdowns. This can result in a better insight in the number of passengers on subsets of the rail network and the components of why they are delayed. As an example, Figure 8.1 shows the breakdown graph of the passengers travelling between stations Rotterdam Centraal and Breda.

3. *What are the control and process indicators currently introduced by ProRail and the NS to control the passenger punctuality and in which way can these indicators possibly be improved or extended?*

ProRail and the Dutch Railways (NS) created a model to describe the passenger punctuality (chapter 5). Here the passenger punctuality is related to the number of train deviations (TVTAs): deviations between the timetable and the realisation, mostly due to cancelled or delayed trains. These TVTAs can be distributed among the responsible business units. In this research the focus lies by the business unit Transport control and Traffic control (TB / VL). The numbers of TVTAs per business unit are the control indicators.

In the next layer of the model, the number of TVTAs under the responsibility of TB / VL is related to 20 process indicators. Five of those process indicators are currently not fully developed. The other 15 process indicators are as follows.

1. Order processing VL-phase
2. Percentage of the correct use of the Train handling document
3. Departure punctuality of empty rolling stock at six rail yards
4. Departure punctuality of empty rolling stock at stations
5. Percentage of trains with the right composition at the morning start
6. Percentage of trains with the right number of personnel at the morning start
7. Lead time between 'Known VL' and 'Known BackOffice'
8. Lead time between 'Known BackOffice' and alarm
9. Lead time between alarm and Distribution decision
10. Lead time between Distribution decision and Blockage measure decision
11. Lead time between Blockage measure decision and running according to the Blockage measures
12. Lead time between end of Blockage and running according to the regular plan
13. Percentage of Blockage measure decision with a Distribution decision
14. Percentage of fully and correctly filled in report cards
15. Percentage of unadapted Blockage measures

Not all TVTAs are used in the model of ProRail and the NS, the reliability of the TVTAs cannot be guaranteed and the number of TVTAs corresponds with the number of trains involved and not with the number of passengers involved. Therefore a second model is proposed. Here the process indicators are related to the components and therefore directly with the passenger punctuality. The TVTAs are neglected in this model.

4. *How can the correlation between the passenger punctuality KPI and the process indicators be determined, especially for a subset of the rail network or a time period?*

The passenger punctuality can be calculated for a subset of the rail network, offering potential to analyse specific parts of the rail network or calculate the performance during a specific time period, like during rush hour. With the two models that describe the passenger punctuality, the process indicators under the responsibility of TB / VL can be related to the passenger punctuality. Unfortunately it is not possible to calculate the process indicators for a subset if the rail network. Therefore a new algorithm is created that can enrich the six data sources used to calculate the process indicators (chapter 6). With this enriched data, it becomes possible to calculate the process indicators for a subset or time period.

5. *How can the correlation between the passenger punctuality KPI and the process indicators be explained and how effective is steering on the process indicators?*

Calculating the correlation between the passenger punctuality and the process indicators via the TVTAs (in the model of ProRail and the NS) or directly (via the second model, based on the components) shows similar results (chapter 5). For only four process indicators (1 to 4) the correlation with the number of TVTAs and the passenger punctuality is moderate. Several remarks can be given:

- For six process indicators (7 to 12) there is no correlation with the number of TVTAs or the passenger punctuality. However there is a correlation with the process itself, instead of the process indicator. This implies that the current definition of the process indicators is incorrect. For both models a new definition is proposed, which give better results.

- Process indicator 14 does not have a correlation with the passenger punctuality. However steering on this indicator to increase the performance leads to more reliable results of process indicators 7 to 12.
- Decreasing the number of TVTAs does not, per definition, leads to higher passenger punctuality. The correlation between the number of TVTAs under the responsibility of TB / VL and the passenger punctuality is only moderate.

The disadvantages of both models are the missing translation between the trains involved and the number of passengers involved. This is most clear in the number of TVTAs. These are trains that are delayed or cancelled, not passengers. This is one of the reasons why the second model was proposed. Nonetheless the same kind of reasoning applies to the process indicators. Steering on the processes will in first instance decrease the number of delayed or cancelled trains and not necessarily the number of delayed passengers.

A case study is performed to calculate the correlation for a subset of the rail network (chapter 7). As study area the main rail line between station Utrecht Centraal and stations Leeuwarden and Groningen is chosen. The found correlations are similar with the correlations of the entire rail network, although some differences can be found. This can partly be due to small data sets for the study area. Nonetheless this explains not all the differences. Some differences between the two models can be found as well. Effective steering to increase passenger punctuality might therefore differ between the entire rail network and the study area.

The main research question is formulated as follows:

> **How is the passenger punctuality KPI constructed, and how is it described by the process indicators?**

The passenger punctuality is the percentage of promised journeys that can be completed on time. Here on time is within 3 minutes of the planned arrival time. The promised journeys are based on the departure and arrival stations and the check-in time of passengers: the registration of each passenger at their departure station. Via the realisation data the arrival times can be determined. If a journey cannot be completed as promised, the checkout time is used to determine the delay of the passenger.

In the calculation of the passenger punctuality every trip is categorized based on the realisation of the promised journey. These categories are based to determine the reasons (the components), from a passengers' perspective, why a trip is delayed. In other words, the passenger punctuality can be broken down into components. Each component is a percentage of trips that is delayed with a specific reason, like a missed transfer or a train that did not depart.

Fifteen process indicators give the performance of the business unit Transport control and Traffic control. For most process indicators the correlation with the passenger punctuality is non-existent. Steering on these process indicators does not increase the passenger punctuality. For six process indicators applies that the current definition is not correct. The underlying process has a correlation with the passenger punctuality. Changing the definition can lead thus to more success in increasing the passenger punctuality.

For a subset of the network the correlation with the passenger punctuality is for most process indicators similar with the entire network. Nonetheless, for certain process indicator the correlation can change. The result is thus that specific steering by subsets of the rail network can lead to better passenger punctuality than steering on the same process indicator for the entire rail network.

8.2. Conclusions

This research can roughly be divided into two parts. In the first part the passenger punctuality itself is analysed: the origin, the calculation, the underlying assumptions and the breakdown to the components. In the second part two models to describe the passenger punctuality are examined. The focus lies by the correlation between the underlying process indicators and the train deviations or the passenger punctuality itself. This paragraph describes both parts of this research and the conclusion found. The main conclusions of this research are highlighted.

The analysis of the passenger punctuality itself started with an elaboration of the origin of the KPI in chapter 2. From a passengers' perspective, the new method of calculation is an improvement relative to the current method of calculation and certainly to the train punctuality. In the train punctuality every train is weighted the same, while in the passenger punctuality every train is weighted with the number of passengers. This number is an estimation at 35 stations in the current passenger punctuality method. In the new passenger punctuality method the number of passengers per train is based on data of all departure and arrival locations and times of every passenger. This method of calculation uses every trip only once while in the current method longer trips with more transfers have a larger impact. Besides it is also possible that short trips are not used at all in the current method.

In the new method of calculation data from the OV-chipkaart (the smartcard for public transport in the Netherlands) is used. This OVCP data contains the check-in and checkout times and locations. The timetable, as available two days before the execution, is combined with the realisation data of more than 20.000 origin-destination combinations. The result is that for almost every possible travel option the realised arrival time, and thus the delay, can be calculated. With the OVCP data every passengers can be combined with a travel option: the promised journey. If the realised arrival time is unknown the checkout time is used to determine the delay. The result is that the passenger punctuality is the percentage of promised journeys that can be completed with a delay less than 3 minutes.

In the new method of calculation 22 assumptions are made. These assumptions are analysed in chapter 3. If possible the impact on the passenger punctuality is calculated as well. For some assumptions, the impact is negligible. Nonetheless for other assumptions the impact can be more than 0.7 percentage point on the passenger punctuality. Combining all assumptions will result in a passenger punctuality KPI that can deviate almost 2 percentage points from the real passenger punctuality. In a subset of the rail network, the difference can even be larger. Hereby applies that by a delay threshold of 3 minutes the effect of the assumptions is more extensive than by a larger delay threshold.

The calculated passenger punctuality is based on the OVCP data and thus based on passenger behaviour. This behaviour is different during disruptions than during a normal service: passengers might check-in or checkout at stations that were not planned. Also the time a passenger checks in might differ from the desired check-in time. Disruptions are not consistent, so the overall impact of disruptions on the passengers cannot be determined. Nevertheless, per case the impact can be calculated.

During the calculation of the passenger punctuality every trip is categorized. These categories can be directly translated to components: reasons why passengers are delayed, like a train that did not depart or a missed transfer. Every category is based on a couple of conditions. If a trip complies with all those conditions, the trip is categorizes with the corresponding category. Here also some assumptions are made. This implies that some trips will end up with the wrong category and thus with the wrong component. Trips with the category 'Missed transfer' are most often wrongly categorized: up to 51% of the trips categorized as 'Missed transfer' did not have a missed transfer. A side-note is that by a larger delay threshold, the components will reflect the reality better.

Two models to describe the passenger punctuality are analysed in chapter 5. In the model of ProRail and the NS the passenger punctuality relates to the number of train deviations (TVTAs). Although there is a correlation between the passenger punctuality and the number of TVTAs, this correlation is not very strong. This is partly due because not every TVTA is included in the model. Another drawback is the subjectivity of the TVTAs. A signaller can have multiple options by choosing the type of TVTA (and thus if the TVTA is included in the model). Also the priority of the signaller is not the TVTAs. Certainly by large disruptions this can lead to mismatching between the reality and the type of TVTAs chosen. The last disadvantage of the TVTAs is the fact that TVTAs corresponds with trains, though the passenger punctuality is based on the number of passengers in the train. This number is not consistent in time and location, so decreasing the number of TVTAs does not implicit means increasing passenger punctuality.

The underlying process indicators give the performance of the business unit Transport control and Traffic control. Improving the corresponding processes should decrease the number of TVTAs. However, steering on the process indicators is not as effective as presumed. As stated in paragraph 8.1, of the 15 process indicators currently in use, five do not have any correlation with the number of TVTAs. For six process indicators the current definitions do not show a correlation while the corresponding processes do have a correlation with the number of TVTAs. The remaining four process indicators have only a moderate correlation with the number of TVTAs. Steering on only the process indicators that do have a correlation with the number of TVTAs might give better results, but this will only decrease the number of TVTAs. The correlation with the number of TVTAs and the passenger punctuality is not strong, so the overall effect on the KPI might therefore not have the desired effect. A remark is the accuracy of the passenger punctuality calculation. It is unknown to which extent this influences the results.

The second model neglects the TVTAs. Here the passenger punctuality is broken down into the components. The found correlation between the process indicators and the passenger punctuality is comparable with the correlations between the process indicators and the number of TVTAs. On the other hand, the correlation between the process indicators and the components was often worse. This can be due to the underlying assumptions of the components: the size of the components does not match the reality. Another explanation is the fact that most process indicators relates to more than one component. For example, process indicators linked to delayed trains have a correlation with both the component 'Train delayed' and the component 'Missed transfer'.

Just as in the model of ProRail, the strongest correlation in the new model is moderate at best. This can be explained with the fact that, just like the TVTAs, process indicators are related to trains and not to passengers. The translation between the trains and the passenger involved is missing.

The correlations in both models are calculated for the entire rail network of the Netherlands. This might differ from the correlations in a subset. Therefore in the case study of the route between station Utrecht and stations Leeuwarden and Groningen (chapter 7 the correlations in both models are calculated once again. For the majority of the process indicators the correlation with the number of TVTAs or the passenger punctuality in the study area is similar with the entire network. Nonetheless there were some significant differences. This suggest that steering on the process indicators with the strongest correlation in the study area has a larger effect on the passenger punctuality than steering on the process indicators with the largest correlation overall.

A remark to this conclusion is that the deviation between the calculated passenger punctuality and the real passenger punctuality for a smaller area can be larger than for the entire rail network. The margin of error by calculating process indicators for a subset of the rail network increases as well because the neglected data increases.

Based on the found results there is no model that seems better in describing the passenger punctuality than the other. Nonetheless, the second model has a preference above the model of ProRail and the NS. The reason is that in this model the TVTAs are neglected. TVTAs are related to trains while the components are related to passengers. Besides, decreasing the size of a component will directly increase the passenger punctuality. This is in contrast to the model of ProRail and the NS: decreasing the number of TVTAs will not necessarily increase the passenger punctuality.

8.3. Reflection

In this research the passenger punctuality KPI, as defined by the NS, is analysed. In this definition OVCP data is used to determine the promised journeys. There is no check to see if a passenger travels according to this journey. If this is not the case, the calculated passenger punctuality leads to a wrong performance value. With the help of additional data sources the allocation of promised journeys to passengers might lead to a higher reliability of the passenger punctuality.

A great additional data source might be the data originated from checking train tickets by the train conductor. In addition with the check-in and checkout data, this can lead to better promised journey. At this moment the travel option with the earliest arrival time is chosen as the promised journey. Waiting time, number of transfers and others aspects of influence are ignored. Calculating the costs of each travel option might lead to a better allocation of promised journeys to passengers. Not the travel option with the earliest arrival, but the travel option with the lowest costs will be the promised journey. Data of the train conductors can help to fine tune this model and to find the costs for all aspects involved.

There are other methods to monitor the trips made on the rail network. Examples are GSM data and Wi-Fi. Disadvantages of these methods are the privacy concerns, the contradiction with the basic principle that every passenger is weighted the same in the passenger punctuality, the impossibility to combine this data with the OVCP data and the fact that the GSM data does not belong to the NS or ProRail.

An aspect what made this research complicated is the need for a delay threshold when calculating the passenger punctuality. In every analysis a distinction is made between the 3 minute and 15 minute passenger punctuality. An option is to neglect the passenger punctuality and to adopt a new KPI that indicates the performance of the NS and ProRail while keeping the passengers' interests in mind.

This alternative KPI could be 'passenger delay minutes': the sum of every delay of every passenger. The big advantage is that every delay is taken into account into the calculation. There is no need to specify a delay threshold. On the other hand, in the delay calculation of every trip some large assumptions were made, for example regarding walking time and the promised journey. This results in a calculated delay that is a couple of minutes higher than the actual delay, leading to a large increase in the number of passengers with a small delay. The number of trips with a delay of 1 or 2 minutes is 1.3 times as high as the number of trips with a delay of 3 minutes or more. If there is no threshold value and the passenger delay minutes are calculated, trips with small delays will have an enormous influence on the outcome. This will make the final results unreliable.

Closely related is the fact that the delay threshold of 15 minutes might not be the best choice to calculate the passenger punctuality. The main reason is that on the larger part of the rail network the frequency of trains is 4 times per hour. If a train did not run, a passenger will take the next train and the final delay is 15 minutes. However this delay is calculated via the checkout time. This implies some deviation, because not every passenger checks out at the same moment and the used walking time might differ from the actual walking time. This deviation is visible is the breakdown graphs. There is often a larger decrease in the size of the components around the delay threshold of 15 minutes than by a delay threshold of 10 or 20 minutes. A good example can be found in Figure 8.2. Just after the delay threshold of 900 seconds, the size of all components decreases more rapidly than before this threshold and after a delay threshold of approximately 1000 seconds. Note that this same phenomenon can be spotted at a delay threshold of 1800 seconds (30 minutes).

The reason to use the 15 minute passenger punctuality is that after a delay of 15 minutes the passenger rating of the performance of the NS decreases rapidly. If the rating drops by a delay of 15 minutes, the threshold should be set at approximately 12 minutes to include all trips with a delay of least 15 minutes. If the rating drops after a delay of 15 minutes, the threshold should become approximately 18 minutes. This will result that only trips with at least a 15 minute delay are used.

8.4. Recommendations

Based on the conclusions of this research, eight recommendations can be made regarding further research and the implementation of the passenger punctuality or de models to improve the passenger punctuality. A distinction is made between recommendations regarding practical implication and further research.

Practical implication

1. The effect of the assumptions of the passenger punctuality calculation on the 3 minute passenger is much larger than by a higher delay threshold. Changing the KPI into the 5 minute passenger punctuality might create a more reliable KPI.

2. The passenger punctuality is the percentage of promised journeys that can be completed with a delay of less than 3 minutes. To create a reliable KPI, the promised journey should match the real journeys of the passengers. Most of the times, only the check-in time and location and the checkout location are not enough. Using more data sources can lead to a more reliable KPI.

3. The effect of disruptions on the passenger punctuality is larger than the KPI indicates. Disruptions are irregular, so it is hard to determine the exact impact on the passenger punctuality. But if the KPI should represent the performance of the rail network, disruptions should be included.

4. The walking time between check-in and platform and between platform and checkout is set at 1 minute, regardless the station or the location of the card readers. Also, the walking time between

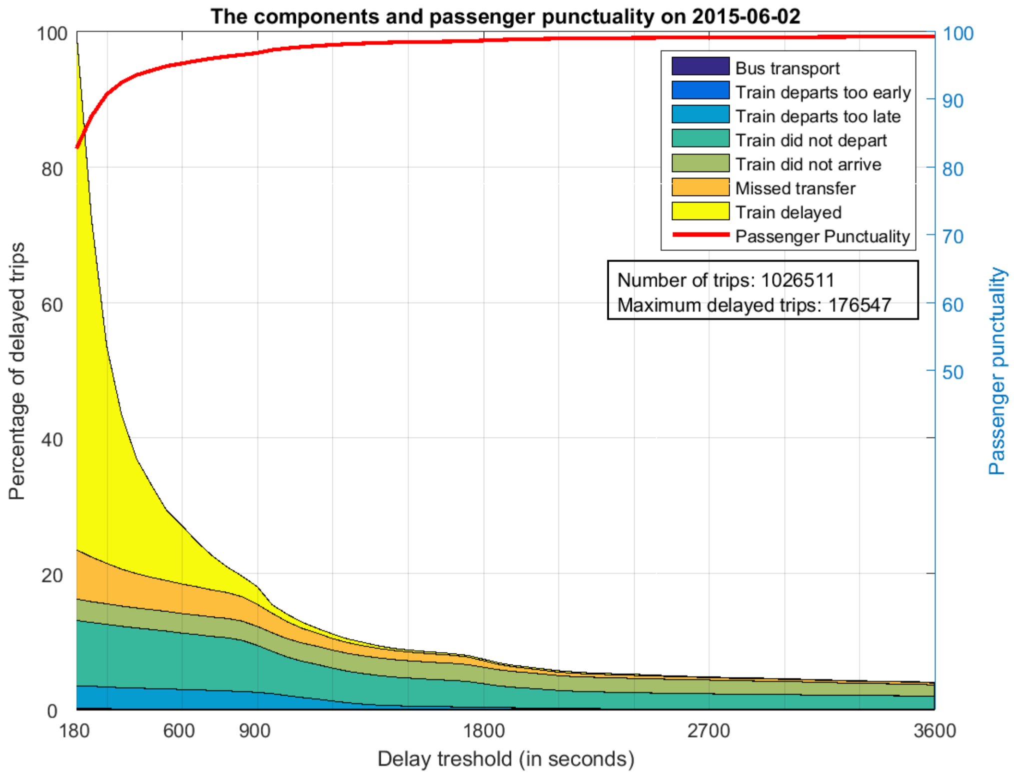


Figure 8.2: The passenger punctuality and the shares of the components on June 2, 2015

platforms is the maximum walking time possible. This leads to wrong promised journeys, wrong delays and more missed transfers than in reality. Implementing more realistic walking times will improve the passenger punctuality.

5. For six process indicators the value of the process indicator does not have a correlation with the number of TVTAs or the passenger punctuality, although the corresponding process influences the number of TVTAs and the passenger punctuality. Therefore the definition of these process indicators should be adjusted. This will make steering more efficient.

Further research

6. Although the threshold value of 3 minute should change into a delay threshold of 5 minutes, it is unknown how a different threshold changes the reliability of the KPI. Therefore further research in this subject is needed.

7. Rescheduling is not part of the passenger punctuality calculation. This means that if a trip could not be made as promised (at least the last train was missed or did not run), the checkout time is used instead of the realised arrival time to calculate the delay. This is a drawback because realisation data is more reliable than checkout times. Nonetheless there are methods to reschedule. Further research in these methods and if possible an implementation of such a method in the passenger punctuality calculation will increase the accuracy of the KPI.

8. For both models analysed apply that there is a missing translation between the passenger punctuality and the TVTAs or the process indicators. The latter group is focused on cancelled or delayed trains while the passenger punctuality is based on passenger delays. This can be adjusted by adapting the process indicators with the expected number of passenger involved. This expected number can be determined via the passenger punctuality data, although some analyses are needed to find the best estimation. It is possible to adjust the TVTAs as well. However this is not recommended because there are more reasons not to uses TVTAs in respect to the passenger punctuality.

9. In the created algorithm to enrich data to calculate process indicators for a subset, it is possible that some parts of the data are not enriched or only partly enriched. This has an effect on the outcome of the calculated process indicators, but the exact influence is unknown. This should be analysed. A better option is to improve the algorithms to increase the part of the data that is fully enriched. Related is the recommendation to analyse in more depth the correlations between the process indicators and the passenger punctuality in subsets of the rail network.

Bibliography

[1] CBS, *Lengte van spoorwegen; spoorwegkenmerken, provincie,* http://statline.cbs.nl/StatWeb/ (2014), [Online; accessed 11-06-2015].

[2] NS, *Veel gestelde vragen en antwoorden,* http://www.ns.nl/over-ns/campagnes/dossier-seizoenen/veel-gestelde-vragen.html (2015), [Online; accessed 11-06-2015].

[3] NS, *Facts & figures,* http://www.nsstations.nl/ns-stations/facts-figures.html (2014), [Online; accessed 11-06-2015].

[4] P. Ramaekers, T. De Wit and M. Pouwels, *Hoe druk is het nu werkelijk op het Nederlandse spoor,* Den Haag: CBS (2009).

[5] NS, *Jaarverslag Nederlandse Spoorwegen 2014,* Utrecht: NS Group NV (2015).

[6] ProRail, *Geschiedenis,* http://www.prorail.nl/reizigers/wie-zijn-we/geschiedenis (2015), [Online; accessed 11-06-2015].

[7] Rijksoverheid, *Rol Rijksoverheid, NS en ProRail,* http://www.rijksoverheid.nl/onderwerpen/openbaar-vervoer/rol-rijksoverheid-prorail-en-ns (2015), [Online; accessed 11-06-2015].

[8] W. J. Mansveld, *Prestaties NS en ProRail 2014,* Den Haag: Ministerie van Infrastructuur en Milieu (2015).

[9] NS, *Vervoerplan Nederlande Spoorwegen 2014,* Utrecht: NS Group NV (2013).

[10] NS, *Punctualiteit hoofdrailnet,* http://www.ns.nl/over-ns/wat-doen-wij/klantoordelen-en-prestaties/treinen-en-dienstregeling/punctualiteit-hoofdrailnet.html (2015), [Online; accessed 11-06-2015].

[11] E. Seest, O. A. Nielsen and R. D. Frederiksen, *Opgørelse af passagerregularitet i S-tog,* Aalborg: Trafikdage (2005).

[12] O. A. Nielsen, A. Landex and R. D. Frederiksen, *Passenger Delay Models for Rail Networks,* Schedule-Based Modeling of Transportation Networks: Theory and applications **46** (2008).

[13] IenM, *Lange Termijn Spooragenda,* Den Haag: Ministerie van Infrastructuur en Milieu (2013).

[14] NS and ProRail, *Beter en meer,* Utrecht: NS Group NV and ProRail (2013).

[15] ProRail, *Beheerplan ProRail 2015,* Utrecht: ProRail (2014).

[16] A. Landex, *Tog-og passagerregularitet på jernbaner,* in *Conference paper presented at TrafikDage* (Aalborg: Trafikdage, 2007).

[17] J. Parbo, O. A. Nielsen, A. Landex, and C. G. Prato, *Measuring Robustness, Reliability and Punctuality within passenger railway transportation,* KGS Lyngby, Denmark: Technical University of Denmark (2013).

[18] A. Landex, *Gis analyses of railroad capacity and delays,* in *Proceedings of ESRI User Conference. San Diego, California, USA* (San Diego: ESRI User Congerence, 2009).

[19] SBB, *Customer Punctuality. Arriving on time with secure connections,* Bern: SBB AG (2015).

[20] SBB, *Kundenpünktlichkeit,* http://www.sbb.ch/sbb-konzern/medien/dossier-medienschaffende/kundenpuenktlichkeit.html (2015), [Online; accessed 15-06-2015].

[21] SBB, *Der SBB Franken 2011,* Bern: SBB AG (2012).

[22] T. Dollevoet, *Delay management and dispatching in railways*, EPS-2013-272-LIS (Rotterdam: Erasmus Research Institute of Management (ERIM), 2013).

[23] A. Schöbel, *A model for the delay management problem based on mixed-integer-programming,* Electronic Notes in Theoretical Computer Science **50** (2001).

[24] T. Broersen, *Functioneel ontwerp RPUN,* Utrecht: NS Group NV (2014).

[25] B. v. d. Berg, *Reizigerspunctualiteit 2013 definitie,* Utrecht: NS Group NV (2013).

[26] NS, *Definitie klantgedreven KPI's op punctualiteit,* Utrecht: NS Group NV (2014).

[27] NS, *Klantonderzoek klantbegeleiding Q1 2015,* Utrecht: NS Group NV (2015).

[28] M. Pelger, *Memo Beloofd product en publicatiemomenten reisplanner,* Utrecht: NS Group NV (2014).

[29] NS, *Toetsdocument Aannames RPUN,* Utrecht: NS Group NV (2014).

[30] B. Schepers and G. H. K. Zwart, *Onderzoek incomplete transacties,* Zoetermeer: Panteia (2014).

[31] P. Bruning, *Trento – Nauwkeurige realisatiegegevens,* Utrecht: ProRail (2012).

[32] T. Broersen, *Memo afweging RFC voor herplannen,* Utrecht: NS Group NV (2014).

[33] R.J. Verhaeghe, M. Minderhoud and K. Lindveld, *Reader CIE4831 Empirical Analysis for Transport & Planning,* Delft: University of Technology (2014).

[34] R. Everhardt, *Gezamelijke PI's TB en VL,* Utrecht: ProRail (2015).

[35] M. Claessens, F. Brouwers and E. Plettenburg, *Spelregels Monitoring,* Utrecht: ProRail (2011).

[36] CVB, *Procesafspraken Monitoring,* Utrecht: ProRail (2012).

[37] J. M. Egbers, *Procesindicatoren in Qlikview,* Utrecht: ProRail (2015).

[38] P. Reitsma and H. Van der Spek, *VKL handleiding ProRail,* Utrecht: ProRail (2007).

[39] ProRail Verkeersleiding, *Werkwijze treindienstleider,* Utrecht: ProRail (2015).

[40] P. Reitsma, *ISVL Gebruikershandleiding,* Utrecht: ProRail (2007).

A

Impact of the assumptions on the passenger punctuality

This appendix is an elaboration on paragraph 3.2. This paragraph analyses the assumptions in the calculations of the passenger punctuality 2.0. In addition and if possible, the impact of this assumption is calculated. A distinction is made between the maximum impact and the expected impact. Paragraph A.1 determines the impact of three assumptions concerning arrivals and departures of passengers by disruptions. This is a completion of paragraph 3.2.3. Paragraph A.2 contains a table used in paragraph 3.2.4. In paragraph A.3 the subject is passenger behaviour by the planning of their journey. This is a completion of paragraph 3.2.5. Paragraph 3.2.7 states that station specific walking times between platform and card reader result in a more accurate passenger punctuality than the generic value of 1 minute. In paragraph A.4 the calculated station specific walking time is given. Paragraph A.5 is an elaboration of the assumption that passengers starts to reschedule when a train has an initial delay of 15 minutes or more. This is part of paragraph 3.2.11. Finally in paragraph A.6 an overview of all assumptions and the calculated impact values is given.

A.1. Departure and arrival

In the calculation of the passenger punctuality three assumptions are made to deal with passenger behaviour by disturbances:

1. **The checkout station is the desired destination of the passenger**
2. **The check-in station is the desired origin of the passenger**
3. **The check-in time is the desired starting time of the trip of the passenger**

The passenger punctuality cannot handle passengers not able to end their trip at their desired station, passengers rerouting by another rail operator and passengers leaving later than planned due to a disturbance on a part of the main rail network. These assumptions are made to deal with these situations. Unfortunately, hereby the passenger punctuality will reflect a better performance of the rail network than in reality.

The impact of these assumptions is time, place and duration dependent, so it is not possible to give a general impact number. Therefore three cases are examined. For every station on the route where the distribution took place, the number of checkouts is counted for the entire duration of the distribution. This number is compared with a mean of the number of checkouts of 15 other days, for the same station, same duration, same type of day (all working days) and without any other distribution. The absolute difference is the estimated number of passengers per station hindered by the disturbance.

If the difference is positive, if more passengers checked out than on an average day, the assumptions is made that those passengers did not reached their final destination. If the difference is negative less passengers than on average reached this station. These passengers could be the same as the passengers that did not reached their final station, so the absolute difference per station cannot be summed to find the total number of hindered passengers.

In the most positive case, all the passengers that did not reached their final destination are the same passengers that were missing on the other stations (where the difference was negative). Therefore to determine the total number of passengers hindered all the positive differences and all the negative differences are added together. The maximum between the absolute values between those two numbers is the total number of hindered passengers. This number divided by the total trips in the calculation of the passenger punctuality is the maximum impact on the KPI in percentage points.

Note that the outcome is a combination of the impacts of the first and second assumption and does not include the impact of the third assumptions. A distinction is not possible with the data available at this moment. Also, the impact is as estimation. It will give the order of magnitude and not an exact value because the used method to determine the impact is based on assumptions itself.

A.1.1. Case 1: Roermond - Sittard

On 11 June 2015 due to a disturbance there was no train traffic possible between Roermond and Sittard between 15:40 and 20:52. Table A.1 shows the average number of checkouts for the four stations on this route and the difference with June 11. Clearly visible are the increased number of checkouts on Roermond and Sittard. These are probably the passenger stranded with no possibility to reroute via the NS. In the intermediate stations the number of checkouts is lower than on average. This means that these stations are not reached by the passengers.

Table A.1: The number of checkouts between 15:40 and 20:52 for the four stations on the route Roermond – Sittard

	Roermond	**Echt**	**Susteren**	**Sittard**
1-jun-15	2287	554	210	1824
2-jun-15	2365	555	262	1897
3-jun-15	2236	508	223	1707
4-jun-15	2548	521	215	1978
5-jun-15	1968	408	173	1701
8-jun-15	2307	563	212	1791
9-jun-15	2582	573	245	1981
10-jun-15	2245	518	257	1699
12-jun-15	1947	398	185	1556
16-jun-15	2289	539	210	1727
17-jun-15	2169	504	202	1700
18-jun-15	2289	506	227	1780
19-jun-15	2168	449	199	1757
22-jun-15	2004	439	203	1508
23-jun-15	2122	521	205	1643
Mean	2235	504	215	1750
11-jun-15	3015	50	23	2231
Difference	780	-454	-192	481

The number of passengers that depart at another station than their final destination is $780 + 461 = 1261$. $454 + 192 = 646$ passengers did not arrive at their final destination. These groups might be partly the same so the maximum number of hindered passengers is the maximum of those groups. So 1261 passengers are hindered but not included in the passenger punctuality.

On 11 June the total number of trips included in the passenger punctuality is 1.046.785. The maximum impact of the disturbance on the KPI is $\frac{1261}{1046785} = 0.12$ percentage point for the route between Roermond and Sittard for a duration of 312 minutes during evening rush hour.

A.1.2. Case 2: Almelo - Deventer

On June 23, between 16:42 and 20:06, no trains run on the route between Almelo and Deventer. Table A.2 shows the number of passengers that checked out in this time span for 15 different days. The difference of the mean with the number of checkouts during the disturbance is the number of passengers hindered per station.

The maximum number of passenger hindered by the disturbance is the maximum between $39 + 173$ and $108 + 21 + 168 + 234$ and is thus 531. By a total of 966426 trips the maximum impact of the disturbance on the passenger punctuality is $\frac{531}{966426} = 0.05\%$ percentage point. This disturbance was

Table A.2: The number of checkouts between 16:42 and 20:06 for the six stations on the route Almelo – Deventer

	Almelo	**Wierden**	**Rijsen**	**Holten**	**Deventer Colmschate**	**Deventer**
1-jun-15	1161	293	355	209	281	2567
2-jun-15	1170	296	364	209	284	2681
3-jun-15	1087	270	334	195	293	2370
4-jun-15	1149	248	348	197	301	2540
5-jun-15	964	197	157	129	185	1873
8-jun-15	1152	301	324	213	306	2471
9-jun-15	1164	302	357	204	288	2634
10-jun-15	1058	250	303	181	283	2319
11-jun-15	1178	275	357	183	267	2544
12-jun-15	996	223	163	146	193	1782
15-jun-15	1092	247	370	180	273	2434
16-jun-15	1181	276	343	177	250	2634
17-jun-15	1014	237	295	164	232	2477
18-jun-15	1159	262	301	187	248	2474
19-jun-15	1082	226	262	157	211	1979
Mean	1107	260	309	182	260	2385
23-jun-15	999	239	348	14	26	2558
Difference	-108	-21	39	-168	-234	173

during the evening rush hour with a duration of 204 minutes.

A.1.3. Case 3: Alkmaar - Zaandam

On June 23, between 10:06 and 12:34, no trains run on the route between Alkmaar and Zaandam. Table A.3 shows the number of passengers that checked out in this time span for 15 different days. The difference of the mean with the number of checkouts during the disturbance is the number of passengers hindered per station.

The maximum number of passenger hindered by the disturbance is the maximum between 62 and $257+16+65+48+100+31+26+157$ and is thus 700. By a total of 989734 trips the maximum impact of the disturbance on the passenger punctuality is $\frac{700}{989734} = 0.07\%$ percentage point. This disturbance was during rush hours with a duration of 148 minutes.

A.2. Bus transport

Including bus transport in the passenger punctuality calculation can impact the KPI. For the first 25 days of June 2015 the 3 minute passenger punctuality with and without bus transport is calculated. The result can be found in table A.4.

A.3. Passenger behaviour with respect to the planning of their journey

There are three main assumptions regarding passenger behaviour with respect to the planning of their journey. The first assumption is: **The passenger wants as quickly as possible from A to B, even if this is without a train of the NS**. The first part of this assumption is elaborated in paragraph A.3.1. The second main assumptions (**The passenger travels as advised by the timetable and hereby uses the default settings**) is partly analysed in paragraph A.3.2.

A.3.1. The passenger wants as quickly as possible from A to B

Consider the following example route between the stations Utrecht Centraal and Leiden, see Figure A.1. In the first three weeks of June 2015 (excluding June 6 and 7, because of scheduled maintenance this example does not apply on these days) 21.513 trips were used in the passenger punctuality calculation.

Table A.3: The number of checkouts between 10:06 and 12:34 for the nine stations on the route Alkmaar – Zaandam

	Alkmaar	**Heiloo**	**Castricum**	**Uitgeest**	**Krommenie-Assendelft**	**Wormerveer**	**Koog-Zaandijk**	**Koog Bloemenwijk**	**Zaandam**
1-jun-15	1083	149	208	149	160	170	411	96	1075
2-jun-15	1132	170	247	153	163	106	137	98	1132
3-jun-15	1096	160	257	193	173	134	294	97	1115
4-jun-15	1188	148	297	131	146	128	198	93	1154
5-jun-15	1277	122	328	264	177	167	230	100	1130
9-jun-15	1205	167	260	166	134	119	167	93	1195
11-jun-15	1128	185	322	187	174	109	278	81	1112
12-jun-15	1366	197	432	231	180	127	254	97	1144
15-jun-15	950	141	260	188	161	134	217	77	996
16-jun-15	1073	139	218	179	174	114	209	97	1159
17-jun-15	1111	147	293	193	182	122	233	98	1161
18-jun-15	1000	158	291	184	197	127	237	86	1175
19-jun-15	1266	183	315	201	170	145	193	93	1137
22-jun-15	1000	143	221	149	124	123	137	90	1196
24-jun-15	1100	149	283	163	185	137	222	91	1230
Mean	1132	157	282	182	167	131	228	92	1141
8-jun-15	875	141	217	134	67	193	197	66	984
Difference	-257	-16	-65	-48	-100	62	-31	-26	-157

Table A.4: The 3 minute passenger punctuality inclusive and exclusive bus transport

	Passenger punctuality		
Datum	**Normal**	**With bus transport**	**Difference**
1-jun-15	84.04%	84.03%	0.005
2-jun-15	82.80%	82.80%	0.001
3-jun-15	84.42%	84.42%	0.001
4-jun-15	85.77%	85.77%	0.000
5-jun-15	67.72%	67.72%	0.001
6-jun-15	81.49%	81.02%	0.470
7-jun-15	88.29%	87.83%	0.456
8-jun-15	86.48%	86.48%	0.002
9-jun-15	89.38%	89.38%	0.000
10-jun-15	81.69%	81.69%	0.004
11-jun-15	83.12%	83.06%	0.052
12-jun-15	83.61%	83.61%	0.004
13-jun-15	87.37%	86.92%	0.443
14-jun-15	87.31%	87.01%	0.306
15-jun-15	86.57%	86.57%	0.000
16-jun-15	87.91%	87.91%	0.000
17-jun-15	87.40%	87.40%	0.000
18-jun-15	90.13%	90.13%	0.000
19-jun-15	91.78%	91.78%	0.001
20-jun-15	90.23%	90.22%	0.001
21-jun-15	91.29%	91.27%	0.025
22-jun-15	85.60%	85.59%	0.001
23-jun-15	86.31%	86.31%	0.000
24-jun-15	87.59%	87.59%	0.001
25-jun-15	87.92%	87.92%	0.000

7499 of those trips (34.86%) had a promised journey which included a transfer at Schiphol. The departure time is every 13 and 43 minutes past the whole hour and the total travel time is 49 minutes. That means that the expected arrival time is respectively 2 and 32 minutes past the next hour. There is another possibility to travel to Leiden. 12 minutes later another train departs in Utrecht travelling directly to Leiden. The travel time is 42 minutes so the expected arrival time is only 5 minutes later than the first option. Although the travel time and the number of transfers of the second option is less than the first option, the used promised journey in the passenger punctuality calculation is the first option if the check-in time is before 13 or 43 minutes past the whole hour.

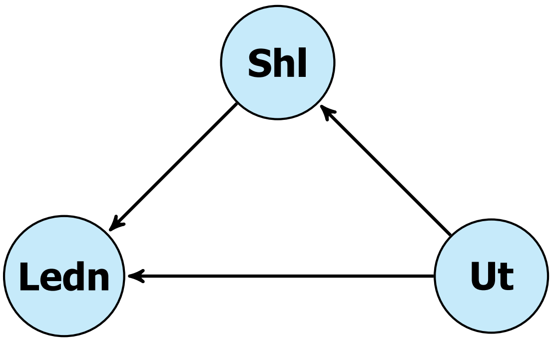


Figure A.1: The rail network between Utrecht (Ut) and Leiden (Ledn)

If the transfer in Schiphol is not made or if the second train did not run or had an initial delay of more than 15 minutes the passengers to Leiden are stranded in Schiphol and need to reschedule. The next train to Leiden departs 12 minutes later. The travel time is the same so the new arrival time is 12 minutes later than the original promised arrival time. In the first three weeks of June 2015 3084 trips to Leiden via Schiphol had a hindrance at Schiphol. Therefore the checkout time should be at least 12 minutes later than the original promised arrival time. Nonetheless 2603 trips had a smaller difference between promised arrival time and checkout time and the checkout time was at least 5 minutes later than the promised arrival time as well. There is thus a large chance that those passengers took the second option for travelling between Utrecht and Leiden. The percentage of passengers with an expected hindrance at Schiphol but based on their checkout time took the second travel option is $\frac{2603}{3084} = 84.40\%$. If the same percentage applies for all passengers between Utrecht and Leiden with a promised transfer at Schiphol than the total number of passengers who did not want as quickly as possible from Utrecht to Leiden for the first three weeks of June is 6329. Those passengers want a travel option with fewer transfers or with a higher reliability.

The maximum impact on the passenger punctuality is the number of passengers with a wrong promised journey divided by the total number of trips. For the first three weeks of June, this is on average 0.11 percentage points. The actual impact is based on only the passengers with a wrong promised journey and a disturbance on this promised journey. This is on average 0.01 percentage points for the first three weeks of June.

A.3.2. Settings in the journey planner: detours

In the journey planner of the NS the passengers can plan their journey. The assumption is that the passengers are using the default settings. This paragraph checks whether this is true for the detour setting. The default setting of the journey planner does not show detours. Nonetheless some train tickets allow passenger to travel unlimited. In those cases it could be beneficial to travel with a detour. In this section for two cases the number of passengers taking the detour is determined. In the end the conclusion is drawn of the impact of this setting in the journey planner on the passenger punctuality KPI.

Case 1: Zwolle - Groningen Europapark

Station Groningen Europapark is directly connection with Zwolle via a sprinter. However it is possible to take the intercity in Zwolle to Groningen and make a transfer to a sprinter in the reversed direction back to Groningen Europapark. This is a detour: the total travel time is 3 minutes longer and this route includes a transfer. However it can be the quickest option. The planned arrival time is 6 minutes earlier than the next option without a detour.

In the first 25 days of June 2015, 886 trips were made between Zwolle and Groningen Europapark. Based on the check-in time for 554 of those trips the quickest option would be to travel via Groningen and thus take the detour. However based on the checkout time, only 10 of those trips were made with the detour. This means that approximately 1.81% of the passengers on this route travels with a detour. The impact of this case on the daily passenger punctuality is 0.00005 percentage point.

Case 2: Utrecht - Amsterdam RAI

Per hour the journey planner shows six different options to travel between Utrecht and Amsterdam RAI, see Table A.5. Two of those options are detours so they are not shown if the default settings are used. Note the option with departure time :25 and :55. Here the travel time is 58 minutes. The next option is only 3 minutes later. Here the travel time is 39 minutes and thus this second option arrives 16 minutes earlier. The reason why the longer option is still shown is the fact that this option does not have any transfers while the faster option has two transfers.

Table A.5: Hourly timetable between Utrecht and Amsterdam RAI

Departure	Arrival	Duration	Detour
:13	:52	0:39	Yes
:25	:23	0:58	No
:28	:07	0:39	No
:43	:22	0:39	Yes
:55	:53	0:58	No
:58	:37	0:39	No

In the first 25 days of June 2015 353 trips were made between Utrecht and Amsterdam RAI. Based on the check-in time, for 49.3% of those trips (174 trips) the quickest option would be to take the detour. Based on the checkout time, 57 passengers took this option. This is 32.76% of the 174 trips and 16.15% of all the trips between Utrecht and Amsterdam RAI. Remarkable is the fact that of the other trips, only 18.39% travelled with the faster direct option (departure time :28 and :58) and 30.46% travelled with the longer option (departure time :25 and :55). For the other 18.39% the travel option could not be determined.

Although this case suggested that approximately 16.15% of the passengers travel with a detour if possible, the impact on the daily passenger punctuality of this case is negligible: 0.0003 percentage point.

Conclusion

In both cases the percentage of passengers taking a detour was significant. However the total impact on the passenger punctuality KPI was in both cases negligible. The total number of origin-destination combinations where a detour could be beneficial is unknown. More research into these travel possibilities is therefore necessary. The impact of the cases on the KPI was negligible, so for now the total impact on the KPI is considered to be negligible.

A.4. Walking time per station

Based on the difference between the realised arrival and the checkout time the walking time per trip at the destination station can be calculated. The mode of all trips per station gives the expected walking time per station. These are given in Table A.6 and A.7. The name of the station is abbreviated to the name of the associated service control point.

At station Brummen (Bmn) the mode of the walking time is -1 minute. This means that the most passengers with this destination checked out a minute before they arrived. This is of course not possible but this is due to the fact that Brummen is a shortstop station. By default at a shortstop station the arrival time is determined as 45 seconds before the departure time. If the actual arrival time is earlier, the difference between registered arrival time and checkout time could be negative. Moreover the median of the walking times at Brummen is 0 minutes. In seconds, the mode is -25 seconds. Negative walking times are not possible, so the walking time in Brummen is set to 0 minutes.

Table A.6: The mode of the walking times for station Abcoude (Ac) to Maarssen (Mas)

Station	Trips	Mode	Station	Trips	Mode	Station	Trips	Mode
Ac	11623	0	Brn	34493	0	Hde	10110	1
Ah	333170	1	Bsk	9830	1	Hdr	33147	1
Ahp	13001	0	Bsmz	33098	0	Hdrz	16276	0
Ahpr	22099	0	Btl	50220	1	Hfd	132471	1
Ahz	23087	1	Bv	46355	1	Hgl	110642	1
Akm	6172	0	Bzl	9066	0	Hgv	34248	0
Alm	214598	1	Cas	61883	1	Hil	18327	0
Almb	57982	1	Cl	71994	1	Hk	16115	1
Almm	59480	1	Cps	18211	0	Hks	18423	2
Almo	41411	1	Db	80836	0	Hld	12568	1
Almp	31114	1	Ddr	197021	2	Hlds	5032	2
Amf	350410	2	Ddzd	11078	0	Hlm	339390	2
Amfs	47463	0	Dld	14805	0	Hlms	30073	0
Aml	86714	0	Dmn	26316	0	Hlo	40106	0
Ampo	28528	1	Dmnz	30592	1	Hm	58227	0
Amr	164769	2	Dn	35860	1	Hmbh	14444	0
Amri	10068	1	Dr	29655	0	Hmbv	8416	0
Amrn	40965	0	Drh	7466	1	Hmh	9474	0
Ana	16745	1	Dron	26810	0	Hn	119113	1
Apd	123424	1	Dt	276397	1	Hnk	43414	1
Apdo	8470	1	Dtz	38033	1	Hno	5032	0
Apn	81433	1	Dv	169244	2	Hon	10124	1
Arn	3657	0	Dvc	11486	0	Hor	7005	0
Asa	254205	1	Dvd	131039	1	Hr	46409	0
Asb	213787	1	Dvnk	25551	1	Hrl	54583	1
Asd	1319739	2	Ec	17936	1	Hrn	9481	0
Asdl	118194	1	Ed	149325	0	Hrt	21620	0
Asdm	109386	0	Ehb	18001	0	Ht	371129	2
Asdz	389721	1	Ehv	519819	3	Htn	62594	0
Ashd	28603	1	Ekz	19779	1	Htnc	32689	0
Asn	76769	0	Eml	22908	1	Hto	11834	1
Ass	422460	1	Es	147725	1	Hvs	216156	1
Assp	32670	0	Esd	19226	1	Hvsm	30703	0
Avat	21022	0	Est	29197	1	Hvsp	55265	0
Bd	248119	1	Etn	23699	0	Hwd	62145	0
Bde	6856	0	Gd	179076	1	Hwzb	17134	1
Bdg	23625	2	Gdg	23838	0	Hze	12020	0
Bdpb	12479	0	Gdm	48084	1	Kbd	4405	1
Bet	44472	2	Gerp	10112	0	Kbw	23528	1
Bgn	57013	0	Gln	4486	0	Klp	31067	2
Bhv	40393	0	Gn	153193	1	Kma	42267	0
Bk	15865	0	Gp	11406	0	Kpn	33217	0
Bkf	5907	0	Gs	60446	1	Kpnz	11163	1
Bkg	18346	3	Gv	312281	1	Krg	6627	0
Bkl	41348	0	Gvc	651506	1	Kzd	25855	0
Bl	17084	0	Gvm	25059	1	Laa	141301	0
Bll	9306	2	Gvmw	20718	1	Ldl	33929	1
Bmn	8097	-1	Gw	7934	0	Ledn	603061	2
Bn	18218	1	Gz	21648	1	Lls	117728	1
Bnk	16051	0	Had	53110	0	Lut	11269	0
Br	9253	0	Hb	1305	0	Lw	80484	1
Brd	43943	1	Hd	47530	0	Mas	34977	0

Table A.7: The mode of the walking times for station Middelbrug (Mdb) to Zwijndrecht (Zwd)

Station	Trips	Mode	Station	Trips	Mode	Station	Trips	Mode
Mdb	38579	0	Rs	18486	1	Vb	18191	1
Mp	44184	0	Rsd	84805	1	Vdg	19570	0
Mrn	11498	0	Rsn	18784	1	Vdo	19116	1
Mss	15137	2	Rsw	59736	2	Vdw	13334	0
Msw	18718	0	Rta	133166	0	Vg	14323	0
Mt	122806	1	Rtb	206225	1	Vh	25078	1
Mtr	12941	1	Rtd	728800	2	Vl	26763	1
Mz	10493	1	Rtn	18107	1	Vndc	17091	2
Ndb	80887	0	Rtz	26596	2	Vndw	12393	0
Nh	3418	1	Rvs	11017	0	Vp	10770	1
Nkk	26921	0	Sbk	1740	0	Vs	22508	1
Nm	354307	2	Sd	5451	0	Vss	7879	1
Nmd	14225	0	Sdm	156338	1	Vst	25775	2
Nml	9052	0	Sgn	43817	0	Vtn	30857	0
Ns	22494	0	Shl	452540	1	Wad	11674	0
Nvd	22371	1	Sn	1685	0	Wadn	6275	1
Nvp	22704	0	Sptn	6910	0	Wc	32226	0
Nwk	26357	1	Sptz	6358	2	Wd	106303	1
Nwl	15928	1	Srn	7182	0	Wdn	13743	1
O	67024	1	Ssh	28206	0	Wf	4416	0
Obd	11190	2	St	1311	1	Wh	8714	1
Odb	9949	1	Std	102133	2	Wm	31722	0
Ost	9625	0	Stz	16416	1	Wp	83654	2
Ot	19659	0	Swk	25253	0	Wt	63314	0
Otb	3658	0	Tb	271898	1	Wv	12334	0
Ovn	19705	0	Tbr	19262	0	Wz	8442	1
Ow	14497	1	Tbu	45263	2	Ypb	17509	1
Pmo	16704	0	Tl	27021	1	Zbm	28920	1
Pmr	20025	0	Tpsw	10769	1	Zd	178362	0
Pmw	12859	0	Twl	10838	1	Zdk	11536	0
Pt	13718	0	Ut	1571908	3	Zl	330721	3
Rai	51705	1	Utg	42397	1	Zlw	7235	2
Rat	15742	0	Utl	26497	0	Zp	96331	1
Rb	3659	1	Utlr	6012	1	Ztm	43610	1
Rh	6099	1	Utm	381	3	Ztmo	27103	1
Rhn	10432	1	Uto	67363	1	Zvb	10172	1
Rlb	53215	1	Utt	27011	0	Zvt	48986	1
Rm	111732	1	Utzl	15674	1	Zwd	46892	1

A.5. The waiting time before rescheduling

There are several reasons to reschedule a trip: a train did not depart, a train did not arrive at the desired destination, a transfer is not made or the train departed too early. For those four reasons it is clear when the passenger starts rescheduling the trip. For the fifth and last reason this distinction is not so clear. The corresponding assumption is as follows: **If a train has a delay of at least 15 minutes before departure, the realised arrival time is the checkout time minus the walking time between train and card reader**. The focus of this paragraph is the first part of this assumption. It state that a trips needs to be rescheduled as the initial delay of a train is 15 minutes or more. This assumption is based on passenger behaviour and the characteristics of the main rail network. Most passengers will search for an alternative when the promised train has a certain initial delay. This threshold delay, the turning point when passengers reschedule, is not known. On the other hand, on most trajectories in the main rail network the frequency of trains is 4 times per hour, or 1 train per 15 minutes. For most trips the first alternative journey will thus be 15 minutes later than the original journey. Therefore the limit between no reschedule and reschedule is set to 15 minutes.

This is no guarantee that this threshold of 15 minutes is always correct. For example, on many routes multiple trajectories can be taken to get to the final destination. The first alternative is thus less than 15 minutes later than the original route. On the other hand, on some other routes this is not possible and the frequency is only 2 trains per hour. For the first 10 days of June the trips without a transfer and with the category 'Train departs too late' where examined. These are 90475 trips, which is 1.02% of all trips during those days. For 21.78% of those examined trips the expected arrival time based on the checkout time is later than the realised arrival time of the train. Most likely, the best option for those passengers was not to reschedule their journey. Also 22.40% of the examined trips had a delay, based on the checkout time, smaller than 15 minutes and the expected arrival time based on the checkout time was earlier than the arrival time of the train. These passengers probably took an alternative train departing within 15 minutes of the promised departure time.

The maximum impact of the threshold of 15 minutes is calculated as follows. For all trips categorized as 'Train departs too late' and with no transfers the delay is calculated via the realised arrival time and via the checkout time. The next step is to determine for how many trips the delay based on checkout time is less than 3 minutes but the delay based on the realised arrival time is at least 3 minutes (group 1). These are the trips where the value of the threshold has an influence on the 3 minute passenger punctuality. The opposite can also occur: trips where the delay based on the realised arrival time is less than 3 minutes and the delay based on the checkout time is 3 minutes or more (group 2). The maximum impact is calculated by taking the maximum of group 1 and group 2. The examined trips do not have transfers, so the outcome should be corrected for all the trips with the category 'Train departs too late'. Dividing the result by all the trips used in the calculation of the passenger punctuality leads to the maximum impact. This is on average 0.10 percentage points for the 3 minute passenger punctuality. To calculate the expected impact the difference of group 1 and group 2 is taken. The remainder of the calculation is the same. This leads to an expected impact on the passenger punctuality of 0.09 percentage point.

The threshold is 15 minutes, so the impact on the 15 minute passenger punctuality is expected to be larger. Calculating this impact shows that this expectation is true. The average maximum impact is 0.20 percentage point. The expected impact is 0.18 percentage point. Note that there will always be an impact on the passenger punctuality, independent of the threshold used. This is due to the fact that the time between planned departure and first alternative changes over time and place.

In Figure A.2 the impact by different thresholds before rescheduling is visualized. To calculate the impact by a threshold of 14 minutes or lower the trips categorized as 'Trip realised' and without transfers are used. For these trips the realised arrival time is known as well as the corresponding checkout time. As can be seen, by a threshold of approximately 10 minutes or lower the impact is considerably. After a threshold of 10 minutes the impact flattens out. However the impact is still decreasing. This is due to the fact that by a larger threshold more often trips are delayed based both on the checkout time and based on the realised arrival time of the delayed train. Based on this information alone a threshold closer to the 30 minutes would therefore decrease the impact on the passenger punctuality.

On the other hand, increasing the threshold by which the passengers are expected to reschedule will decrease the number of delayed trips with the correct category. This implies mostly to the trips categorized as 'Missed transfer'. More trips will get this category because the arriving train departed

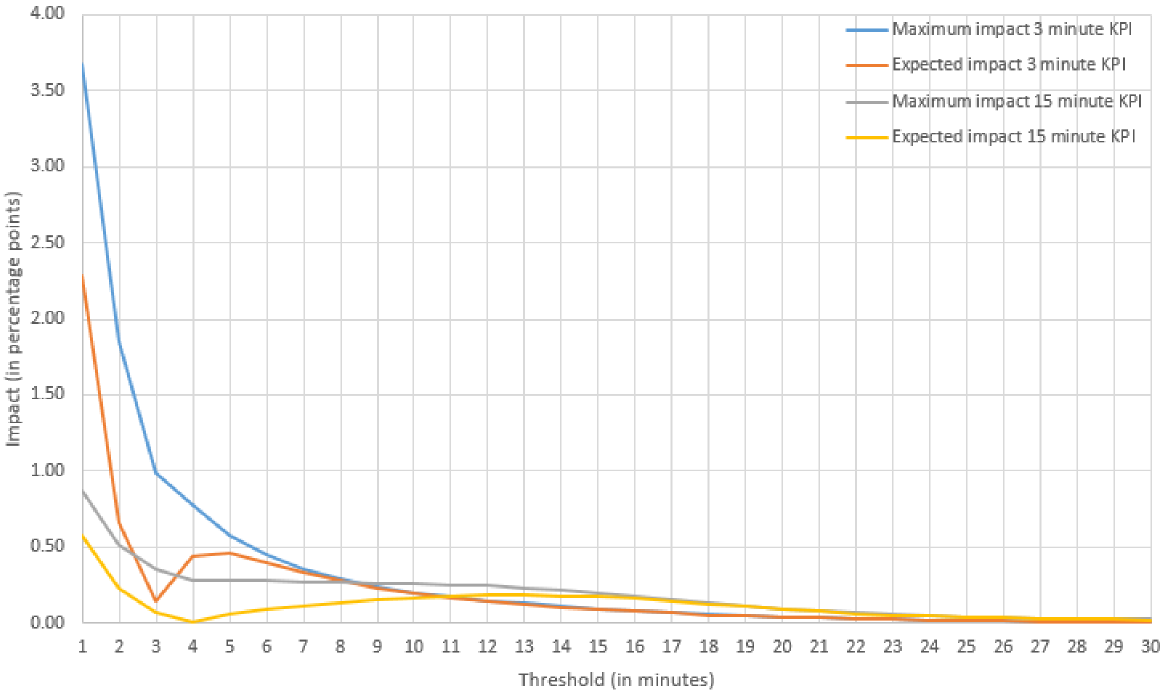


Figure A.2: The maximum and expected impact on the 3 and 15 minute passenger punctuality by different thresholds for rescheduling

with a large delay. A second reason is not to increase the threshold is already stated: the frequency on most routes is 1 train per 15 minutes. Passengers do not consider train numbers. So a consecutive train is, in the eyes of a passenger, the same of the original train. Neglecting the rare exceptions, at this moment the frequency of 1 train per 15 minutes is the highest possible. A lower threshold is therefore not useful.

Note that if the timetable changes on some routes will get a frequency of more trains per hour, this threshold value should change. A solution could be to change the definition of the trips with the category 'Train departs too late'. At this moment every trip whereby the train departs at least 15 minutes later than planned is categorized with this category (see also paragraph 3.1.2). If the default value of 15 minutes changes in a variable value based on the frequency of the train series more trips will be categorized correctly if the frequency changes to more trains per hour. An additional advantage is that the categorization of the trips with trains on routes with a lower frequency will also become better.

A.6. Summary

A summary of all assumption in the passenger punctuality 2.0 KPI with the corresponding impact is given in Table A.8 and Table A.9. The maximum impact is the number of trips influenced by the assumption divided by the total number of trips. The expected impact is the number of trips influenced by the assumption that will change the passenger punctuality divided by the total number of trips. For example if 10 of the 100 trips are influenced the maximum impact is $\frac{10}{100} = 10$ percentage point. Five trips already had a negative influence on the passenger punctuality, so the expected impact will be $\frac{10-5}{100} = 5$ percentage point.

Table A.8: The first 16 assumptions of the passenger punctuality 2.0 KPI and the corresponding impact

		Impact on 3 minute KPI		
	Assumption	**Maximum**	**Expected**	**Remark**
1	The observed trips in the passenger punctuality calculation are representing the average passenger	0.05 percentage points	Undetermined	Not all factors are included
2	The closing of the OVCP terminals at stations does not have an influence on the representativeness of the used trips in the calculations	Undetermined	Undetermined	The impact is undetermined but the effect is positive
3	Trips derived from the OVCP data are not with the Thalys, Eurostar and/or ICE	Negligible	Negligible	
4	The checkout station is the desired destination of the passenger	Time and place dependent	Time and place dependent	For three examples the maximum impact lies between the 0.05 and 0.12 percentage point
5	The check-in station is the desired origin of the passenger			
6	The check-in time is the desired starting time of the trip of the passenger			
7	Trips (partly) conducted by bus transport are not included in the passenger punctuality	On days with planned bus transport: 0.42 percentage point. Otherwise negligible		The maximum and expected impact are the same
8	The passenger wants as quickly as possible from A to B, even if this is without a train of the NS	Undetermined	Undetermined	Per case the impact is calculable.
9	The passenger travels as advised by the timetable and hereby uses the default settings	Negligible	Negligible	The impact of two factors could not be determined
10	The passenger bases his expectation on the timetable two days prior to the journey	Negligible	Negligible	
11	Every trip obtained from the OVCP data can be coupled to a promised journey	0.09 percentage points	Undetermined but probably negligible	The KPI will not change by a larger dataset
12	The walking time between card reader and platform is 1 minute	0.63 percentage points	0.42 percentage points	
13	The walking time between platform and card reader is 1 minute	0.12 percentage point		The maximum and expected impact are the same
14	The clock in the card readers is exact	Undetermined	Undetermined	Probably negligible
15	The realised arrival and departure times are measured exact on every station	0.72 percentage points	Undetermined	
16	The determination of 45 seconds as stopping time at a shortstop is accurate	Undetermined	Undetermined	

Table A.9: The last 6 assumptions of the passenger punctuality 2.0 KPI and the corresponding impact

		Impact on 3 minute KPI		
	Assumption	**Maximum**	**Expected**	**Remark**
17	The promised transfers time corresponds with the realised transfer time	Undetermined	Undetermined	Impact included in assumptions 18-22
18	If a train has a delay of at least 15 minutes before departure, the realised arrival time is the checkout time minus the walking time between train and card reader	0.10 percentage points	0.09 percentage points	The impact of rescheduling by an initial delay of 15 minutes
19	If a train does not depart, the realised arrival time is the checkout time minus the walking time between train and card reader			
20	If a train does not arrive at the destination or transfer station, the realised arrival time is the checkout time minus the walking time between train and card reader	0.23 percentage point		Only delayed trips are influenced by these assumptions. So the maximum impact is the same as the expected impact
21	If a transfer is not made, the realised arrival time is the checkout time minus a walking time between train and card reader			
22	If the first train is missed because it departed too early, the realised arrival time is the checkout time minus a walking time between train and card reader			

B
Algorithms to create breakdown graphs

Out of the output data of the calculation of the passenger punctuality the, so called, breakdown graphs can be created. These graphs show the passenger punctuality and the share of each component by a range of delay thresholds. These components are the reasons why a trip is delayed. The sum of all components is set equal to 100% by the delay threshold where the passenger punctuality is the lowest. These breakdown graphs show information about how much and why the investigated network is not punctual and by which delay threshold. The components, the delay threshold and the outcome of the graphs are discussed in chapter 4.

Two types of graphs can be distinguished. The first type is the 'Normal breakdown graph'. This breakdown graph can be used to determine the passenger punctuality and the share of the components for a certain time period. Here the time period is set on 1 day. The condition is that the data of all the trips of that time period are available. The outcome of this graph is than exact for that examined time period. The algorithm to create such a graph is explained in paragraph B.1.

It is also possible that the subject of an analysis is a station, train series or another part of the network. To create a graph were the displayed information is exact can only be done when all the data of all the trips is available. By using a sample size of the data however, an estimation can be made of the passenger punctuality and the shares of the components. The larger the sample size, the more accurate the breakdown graph becomes. Such a 'Special breakdown graph' is explained in paragraph B.2. This is a continuation of the previous paragraph.

B.1. Normal breakdown graph

This appendix describes the algorithm to create breakdown graphs for the daily passenger punctuality. In consist of two parts. The first part handles the data (paragraph B.1.1). In basis the data gets summarized into a small table. This is done via Structured Query Language (SQL). The resulting table is the input for the second part. Here the graphs are created (paragraph B.1.2). This is done via MATLAB. In theory MATLAB is able to deal with the actions of part 1, but because the size of the raw input is very large SQL is used. This programming language is better and faster to deal with large databases.

B.1.1. Part 1: Data handling

In paragraph 3.1 the structure of the passenger punctuality calculation is given. It consists of 5 modules. The fifth module generates the output. One table is called: 'EZ_D_RPU.FCT_RPUN_REIS' [24]. This table contains every trip with a promised journey. Other information is, among others, the date on which the trip was made, the final delay in seconds and the category of the trip (see paragraph 3.1.2). Every trip is included in this table, so the size is extensive. On average on a normal working day the number of entries is approximately one million (and less than half a million for a weekend day). Dedicated software is needed to handle such large tables. The NS uses SAS Enterprise Guide.

This software is based on the Structured Query Language (SQL), a programming language designed for managing databases. Therefore the presented code is written in SQL.

The first step is to define the database where the resulting table will be stored. In this case this is the same database where the input data ('EZ_D_RPU.FCT_RPUN_REIS') is stored, but this is not a necessity. Here the name of the database is 'Component_database'. Note that SQL is not case sensitive. A SELECT-statement is the same as a Select-statement or a select-statement. The same applies to names of tables or columns.

```
Use Component_database
Go
```

Secondly all the different days in the 'EZ_D_RPU.FCT_RPUN_REIS'-table needs to be found and be numbered. Each day should get a unique number so every day can be identified by that number. This is done in two steps. First all the different days in the table are identified by the Select Distinct-statement. The result is stored in a temporary table called 'Days_temp'. Lastly the days are ordered and given a unique number via the Row_Number-function. The result is a table called 'Days' with every data available with a unique number.

```
;
With Days_temp as (
Select Distinct RS_BELOFTE_H_DATUM
From EZ_D_RPU.FCT_RPUN_REIS )

Select *, Row_Number () Over (Order by RS_BELOFTE_H_DATUM) as Number
Into Days
From Days_temp
```

Next the resulting table is created. This table is called 'Components' and contains ten columns. The first column contains the date and the second the delay threshold. Together they form a unique couple. The next seven columns contain the number of trips per components, associated with the date and threshold. The last column contains the passenger punctuality.

```
Create Table Components (
Date nvarchar(40),
Threshold float,
Number_departstooearly float,
Number_delayed float,
Number_didnotarrive float,
Number_departstoolate float,
Number_missedtransfer float,
Number_didnotdepart float,
Number_realised float,
PPunc float )
```

Before this table can be filled, some variables needs to declared. Two variables provide the tools to loop the calculation part for every day ('Counter_date') and every threshold ('Counter_threshold'). Note that this thus means that there are two loops. Two other variables ('Date' and 'Total') are needed in the loops but a variable can only be declared once, so they are declared here.

```
Declare @Counter_date int = 1
Declare @Counter_threshold float
Declare @Date nvarchar(40)
Declare @Total float
```

Now the first loop can start. The content of the loop applies to every date in the table 'Days'. Here all the days are taken into account. Note that the counter is already set to 1. The last value is the highest value of the 'Number'-column of the 'Days' table. So the last examined date is the last date available. By changing the start value and the end value of the counter the days examined can be changed.

```
While @Counter_date <= (Select Max(Number) From Days)
Begin
```

The first step of the first loop is to select the entries from 'EZ_D_RPU.FCT_RPUN_REIS'-table where the date is equal to the date examined. So first the exact date must be identified. It is the date from the 'Days'-table where the unique number is equal to the counter. The selected entries are stored in a table called 'Selection'. Note that trips categorized as 'Bus transport' or 'Other rail operator' are not used in the passenger punctuality calculation. So those trips are neglected. Note also that the names of the categories are in Dutch. Finally the total number of trips in the selection are counted and stored in the corresponding variable.

```
Set @Date = (Select RS_BELOFTE_H_DATUM From Days Where Number = @Counter_date)

Select *
Into Selection
From EZ_D_RPU.FCT_RPUN_REIS
Where RS_BELOFTE_H_DATUM = @Date and RS_VTG_CATEGORIE not in ('Busvervoer' ,
'Andere vervoerder')

Set @Total = (Select Count(*) From Selection)
```

With the selection determined, the second loop can start. This loop will go through all the different threshold values. First off, the start value of this second loop needs to be determined. In this case, this is set to 1 minute. The delays are measured in seconds, so the start value is 60. The last threshold is here 1 hour, so 3600 seconds.

```
Set @Counter_threshold = 60

While @Counter_threshold <= 3600
Begin
```

In the second loop the 'Components'-table is being filled. First a unique entry needs to be created. As stated the combination of date and threshold is unique so those corresponding columns need to be filled first.

```
Insert Into Components (Date, Threshold)
Values (@Date, @Counter_threshold)
```

In the 'Selection'-table the involved trips are stored. However only the trips with a delay of at least the threshold needs to be counted. So this table needs to be filtered and counted by category, because the category is the basis for the component. The result is stored in a temporary table called 'Selection_temp'. Then the found values need to be stored in the correct cell in the 'Components'-table. Note that the names of the categories are in Dutch.

```
;
With Selection_temp As (
Select RS_VTG_CATEGORIE, Count(*) as Number
From Selection
Where RS_VTG_SEC >= @Counter_threshold
Group by RS_VTG_CATEGORIE )

Update Components
Set Number_departstooearly = (Select Number From Selection_temp Where RS_VTG_CATEGORIE =
'trein te vroeg vertrokken'),
Number_delayed = (Select Number From Selection_temp Where RS_VTG_CATEGORIE =
'reis gerealiseerd'),
Number_didnotarrive = (Select Number From Selection_temp Where RS_VTG_CATEGORIE =
'trein niet aangekomen'),
Number_departstoolate = (Select Number From Selection_temp Where RS_VTG_CATEGORIE =
```

```
'trein te laat vertrokken'),
Number_missedtransfer = (Select Number From Selection_temp Where RS_VTG_CATEGORIE =
'overstap gemist'),
Number_didnotdepart = (Select Number From Selection_temp Where RS_VTG_CATEGORIE =
'trein niet vertrokken')
Where Threshold = @Counter_threshold and Date = @Date
```

If by a threshold on a certain date no trips are with a certain component, the value of the corresponding cell is '*Null*'. This needs to be corrected. So for all components, if the value is '*Null*' it should be replaced with a 0.

```
Update Components
Set Number_departstooearly = 0
Where Number_departstooearly is Null

Update Components
Set Number_delayed = 0
Where Number_delayed is Null

Update Components
Set Number_didnotarrive = 0
Where Number_didnotarrive is Null

Update Components
Set Number_departstoolate = 0
Where Number_departstoolate is Null

Update Components
Set Number_missedtransfer = 0
Where Number_missedtransfer is Null

Update Components
Set Number_didnotdepart = 0
Where Number_didnotdepart is Null
```

The total number of trips realised and within the delay threshold is the total number of trips minus the trips per component. This is calculated in the next part of the script. The results are stored in the 'Components'-table.

```
Update Components
Set Number_realised = @Total - Number_departstooearly - Number_delayed - Number_didnotarrive -
Number_departstoolate - Number_missedtransfer - Number_didnotdepart
Where Threshold = @Counter_threshold and Date = @Date
```

At this moment the passenger punctuality corresponding with the threshold can be calculated. The formula is the number of realised trips within the delay threshold divided by the total number of trips, times 100%. The result is stored in the 'Components'-table.

```
Update Components
Set PPunc = (Number_realised / @Total) * 100
Where Threshold = @Counter_threshold and Date = @Date
```

Now all the content of the second loop is explained. However before the loop can be ended, the counter should be increased with the step size. Here the passenger punctuality is calculated for every minute so the step size is 60. After all the passenger punctualities for all threshold values are calculated, the 'Selection'-table can be deleted.

```
Set @Counter_threshold = @Counter_threshold + 60
End
```

```
Drop Table Selection
```

In the last part of the script also the first loop is ended. Here also the counter is increased by the step size. Every day in the original input table is here analysed, so the step size is 1. Finally the ’Days’-table can be deleted.

```
Set @Counter_date = @Counter_date + 1
End

Drop Table Days
```

The result of this script is a table with the size of the components and the passenger punctuality for all the days in the original table and for multiple delay thresholds. The number of different delay threshold is the last value minus the first value, divided by the step size, plus 1. So in this case this number is $\frac{3600-60}{60}+1=60$.

B.1.2. Part 2: Creating the breakdown graphs

The software used to create the graphs is MATLAB. This program can perform mathematical actions on datasets and visualise the results. However it cannot handle very large datasets. Therefore the first part is done via SQL and not in MATLAB though this is in theory possible. In contrast to SQL, statements, names and functions in MATLAB are case sensitive. This paragraph describes the MATLAB script to create the breakdown graphs.

The first part of the script is to import the data. First a connection with the SQL database needs to be made. Via a SQL-query the result of the first part can be imported. This will be stored in a table called ’data’. The first column of this table is used to find the unique days included in the data. The resulting list is stored in the table ’date_list’. The last action is to make a folder, named ’Day’, where the resulting graphs can be saved. If there is already a folder with this name, MATLAB will give a warning but continues running the script.

```
conn = database(’Component’, ”, ”, ’Vendor’, ’Microsoft SQL Server’, ’Server’, ’localhost’, ’AuthType’,
’Windows’, ’PortNumber’,1433);
query = ’Set Nocount on Select * From Components Order by Date, Threshold’;
data = fetch(conn,query);
date_list = unique(data(:,1));

mkdir(cd,’Day’);
```

A new graph needs to be created for every day in the data, so a loop is needed. In this case all the days are analysed so the start value of the counter is set to 1. The last value is the size of the ’date_list’-table. The step size is not determined so the default value of 1 is used.

```
for counter = 1:size(date_list)
```

The first part of the loop is to select the date that corresponds with the counter. With this date, the corresponding data can be determined. However to select the data first the rows that contains the data must be determined. The result is a table called ’selection’ containing only the data of the date.

```
date = date_list(counter);
row_selection = strcmp(data(:,1),date(1));
selection = cell2mat(data(row_selection == 1,2:10));
```

The resulting graphs show the share of each component for every delay threshold. The sum of each share on the 3 minute passenger punctuality is set to 100%. The ’selection’-table contains the number of trips with the same component. These values should therefore be divided by the sum of those values by the 3 minute delay threshold. So first this value is calculated. Also the total number of trips is calculated. At last, the shares of each component are calculated.

```
sum_share = sum([selection(3,2) selection(3,3) selection(3,4) selection(3,5) selection(3,6)...
selection(3,7)]);
total_trips = sum([selection(1,2) selection(1,3) selection(1,4) selection(1,5) selection(1,6)...
selection(1,7) selection(1,8)]);

selection(:,2) = selection(:,2)*100/sum_share;
selection(:,3) = selection(:,3)*100/sum_share;
selection(:,4) = selection(:,4)*100/sum_share;
selection(:,5) = selection(:,5)*100/sum_share;
selection(:,6) = selection(:,6)*100/sum_share;
selection(:,7) = selection(:,7)*100/sum_share;
```

At this moment the graph can be made. Note that the components are visualized as an area, the shares are stacked. The passenger punctuality is visualized as a line. The layout of the graph is set as well. This includes the title, the labels of the axis, the value of the axis, the grid lines and the legend. Note that the date is included in the title.

```
[ax,ar,lin] = plotyy(selection(:,1), [selection(:,2) selection(:,4) selection(:,5) selection(:,6) ...
selection(:,7) selection(:,3)], selection(:,1), selection(:,9),’area’,’plot’);
titlestring = strjoin([’The components and passenger punctuality on ’ date],”);
title(titlestring);
xlabel(’Delay threshold (in seconds)’);
ylabel(ax(1),’Percentage of delayed trips’);
ylabel(ax(2),’Passenger punctuality’);
lin.LineWidth = 2;
lin.Color = (’r’);
xlim(ax(1),[180 3600]);
xlim(ax(2),[180 3600]);
ylim(ax(1),[0 100]);
ylim(ax(2),[0 100]);
set(ax(1),’Ytick’,[0 20 40 60 80 100],’YtickLabel’,[0 20 40 60 80 100],’ygrid’,’on’);
set(ax(2),’Ytick’,[50 60 70 80 90 100],’YtickLabel’,[50 60 70 80 90 100]);
set(ax(2),’Xtick’,[180 300 600 900 1200 1500 1800 2100 2400 2700 3000 3300 3600],’XtickLabel’,...
[],’xgrid’,’on’);
set(ax(1),’Xtick’,[180 600 900 1800 2700 3600],’XTickLabel’,[180 600 900 1800 2700 3600]);
legend(’Train departs too early’,’Train did not arrive’,’Train departs too late’,’Missed transfer’,...
’Train did not depart’,’Train delayed’,’Passenger Punctuality’);
```

Below the legend a text box will be created with some additional information. Here the total number of trips are given, and the number of trip with a delay of three minutes or more. This text box will get a tag. MATLAB will not remove this text box automatically when making a new graph. The second graph will thus have two text boxes, projected above each other. Via the tag the text box can manual be removed later in the script.

```
content = [’Number of trips: ’ num2str(total_trips)], [’Maximum delayed trips: ’ num2str(sum_share)];
annotation(’textbox’, [0.61,0.6,0.28,0.08],’String’, content, ’Tag’ , ’Toberemoved’);
```

Now the graph is created, it can be saved. This is done in the folder, earlier created. The file name is the same as the date, so the graphs are saved individually. However, if there is already a file with the same name in the folder, this old file gets overwritten by the new file.

```
filename = strjoin([cd ’\Day\’ date ’.png’],”);
graph = figure(1);
saveas(graph,filename);
```

Before the loop is ended, the text box should be removed. This is done via the tag, as mentioned earlier. After all the graphs are made the window with the last graph is still open. The last command therefore is to close this window.

```
delete(findall(gcf,'Tag','Toberemoved'));

end

close all
```

This script results in a breakdown graph per analysed day. All the graphs are saved and stored in a folder called 'Day'. This folder can be found in the active folder in MATLAB. Normally this is the default folder. This folder can be found via the following command.

```
matlabroot
```

To change the active folder the following command can be used. Here is 'directory' the name of the new location.

```
cd('directory')
```

B.2. Special breakdown graphs

Instead of creating breakdown graphs for the whole network, it is also possible to create a graph for a part of the network. The used data comes from the 'EZ_D_RPU.FCT_RPUN_REIS'-table, a product of the fifth module of the passenger punctuality calculation, see paragraph 3.1. [24] As stated in paragraph B.1.1 this table consists data of every trip made. The columns 'RS_BELOFTE_H_DATUM', 'RS_VTG_SEC' and 'RS_VTG_CATEGORIE' are used. However there are more columns containing information about each trip. These columns can be used to filter the data to create breakdown graphs for origin, destination, transfer station, origin-destination combination and train series.

The algorithm to create such special breakdown graphs in in basic the same as the standard algorithm presented in paragraph B.1. The differences are explained in this part of the appendix. In paragraph B.2.1 the changes and additions in the first part of the algorithm are explained. A distinction is made between the different kinds of filtering of the data: arrival station, train series or origin-destination combination for example. The second part (paragraph B.2.2) describes the changes in the creation of the graph. This is the same for every type of filtering possible.

B.2.1. Changes in part 1: Data handling

The method to create the output table containing the number of trips per component per delay threshold and the corresponding passenger punctuality for a specific part of the network differs only on one point from the original algorithm, some changed column, table and variable names aside. In the original algorithm a table containing every date in the data is made. This part needs to changes. An associated change is the creations of the 'Selection' table. These changes are described in this paragraph, per type of special breakdown graph.

Breakdown per origin

The column named 'RS_DRP_HERKOMST' in the 'EZ_D_RPU.FCT_RPUN_REIS'-table contains the origin station of every trip. First every different origin station needs to be found. The result is stored in the temporary table 'Origin_temp'. After giving each station a unique number the final result is stored in 'Origin'.

```
;
With Origin_temp as (
Select Distinct RS_DRP_HERKOMST
From EZ_D_RPU.FCT_RPUN_REIS )

Select *, Row_Number () Over (Order by RS_DRP_HERKOMST) as Number
Into Origin
From Origin_temp
```

The part of the script where the 'Selection'-table is made differs also. Here the variable 'Origin' replaces the variable 'Date' of the original script. This variable is declared before the first loop but given a value just before the 'Selection'-table is made. This table consist of all trips originated in the origin associated with the counter of the loop.

```
Set @Origin = (Select RS_DRP_HERKOMST From Origin Where Number = @Counter_origin)

Select *
Into Selection
From EZ_D_RPU.FCT_RPUN_REIS
Where RS_DRP_HERKOMST = @Origin and RS_VTG_CATEGORIE not in ('Busvervoer' ,
'Andere vervoerder')
```

Breakdown per destination

The changes in the script to perform a breakdown per destination do not differ much from the breakdown per origin. In the 'EZ_D_RPU.FCT_RPUN_REIS'-table a column, named 'RS_DRP_BESTEMMING' contains the destination station of the trips. The list of every destination station with a unique number is created as follows.

```
;
With Destination_temp as (
Select Distinct RS_DRP_BESTEMMING
From EZ_D_RPU.FCT_RPUN_REIS )

Select *, Row_Number () Over (Order by RS_DRP_BESTEMMING) as Number
Into Destination
From Destination_temp
```

The changed creation of the 'Selection'-table is as follows.

```
Set @Destination = (Select RS_DRP_BESTEMMING From Destination Where Number =
@Counter_destination)

Select *
Into Selection
From EZ_D_RPU.FCT_RPUN_REIS
Where RS_DRP_BESTEMMING = @Destination and RS_VTG_CATEGORIE not in ('Busvervoer' ,
'Andere vervoerder')
```

Breakdown per origin-destination combination

A combination of origin and destination can also be used to create special breakdown graphs. The number of OD combinations is extensive: approximately 20.000 combinations can be found. Therefore here only the 200 combinations with the most passengers are analysed. Those top 200 combinations are stored first in a temporary table called 'OD_temp'. This table is used as input to create a new table with a unique name per OD combination and a unique number.

```
;
With OD_combination_temp as (
Select top 200 RS_DRP_HERKOMST, RS_DRP_BESTEMMING, Count(*) as Amount
From EZ_D_RPU.FCT_RPUN_REIS
Group by RS_DRP_HERKOMST, RS_DRP_BESTEMMING
Order by Amount Desc )

Select RS_DRP_HERKOMST + '_' + RS_DRP_BESTEMMING as OD, Row_Number () Over (Order by
RS_DRP_HERKOMST) as Number
Into OD_combination
From OD_combination_temp
```

To make the 'Selection'-table the 'Date'-variable changes in the 'OD'-variable. However before the loop two additional variables are declared: 'Origin' and 'Destination'. These variables are necessary because the OD combination cannot be directly obtained from the 'EZ_D_RPU.FCT_RPUN_REIS'-table.

```
Set @OD = (Select OD From OD_combination Where Number = @Counter_OD)
Set @Origin = Left(@OD, Charindex(’_’, @OD) -1)
Set @Destination = Right(@OD, Len(@OD) -Charindex(’_’, @OD))

Select *
Into Selection
From EZ_D_RPU.FCT_RPUN_REIS
Where RS_DRP_HERKOMST = @Origin and RS_DRP_BESTEMMING = @Destination and
RS_VTG_CATEGORIE not in (’Busvervoer’ , ’Andere vervoerder’)
```

Breakdown per transfer station

The column ’RS_STATION_TREINNR_COMB’ contains the promised journey. It is a string of stations and train numbers. An example of a value could be: ’Bsmz-5863-Hvs-4961-Ut-17460-Mas’. Here the promised journey started at station Bussum Zuid (Bsmz). The first train was train 5863 to Hilversum (Hvs). There a transfer was made to train 4961. At Utrecht Centraal (Ut) a second transfer was promised to the last train, train 17460, to the final destination: Maarssen (Mas). The first step is again to find all the transfer stations of all the trips. This can be done by splitting the value of the ’RS_STATION_TREINNR_COMB’-column. However there is no default function in SQL that can split a string of text. Therefore this function needs to be defined first.

```
Create Function [dbo].[Split]
( @String Nvarchar(4000), @Delimiter Nchar(1) )
Returns Table
as Return
( With Split(stpos, endpos) as (
Select 0 as stpos, Charindex(@Delimiter, @String) as endpos
Union all
Select endpos +1, Charindex(@Delimiter, @String, endpos +1)
From Split
Where endpos > 0 )
Select ’Id’ = Row_Number() Over (Order by (Select 1)),
’Data’ = Substring(@String, stpos, Coalesce(Nullif(endpos, 0), Len(@String) +1) -stpos)
From Split )
Go
```

To determine how many transfers are included in the promised journey the length of the value is analysed. This length minus the length of the value where all the hyphens are deleted gives the number of parts the promised journey consists of. Dividing this number by two gives the number of transfers and thus the number of transfer station of that trip. This idea is used to find all the station where a transfer is made. First all stations where the first transfer is made are determined. The result is stored in a temporary table called ’Transfer_temp’.

```
Select (Select Data
From dbo.Split(RS_STATION_TREINNR_COMB, ’-’)
Where Id = 3) as Transfer
Into Transfer_temp
From EZ_D_RPU.FCT_RPUN_REIS
Where (Len(RS_STATION_TREINNR_COMB) - Len(Replace(RS_STATION_TREINNR_COMB, ’-’, ”))) / 2
> 1
```

The table ’Transfer_temp’ should be extended with at the other transfer stations. This can be done via a loop. This implies the declaration of a new variable, called ’Counter_split’. For every transfer station what is not the first transfer station the name of the station is determined. Therefore the start value of the counter is 2. From an analysis the maximum number of transfer stations is determined to be 6. So this is the end value of the counter. After the loop is ended, the ’Split’-function can be deleted.

```
Declare @Counter_split int = 2

While @Counter_split <= 6
Begin

Insert into Transfer_temp
Select (Select Data
From dbo.Split(RS_STATION_TREINNR_COMB, ’-’)
Where Id = (@Counter_split * 2) + 1)
From EZ_D_RPU.FCT_RPUN_REIS
Where (Len(RS_STATION_TREINNR_COMB) - Len(Replace(RS_STATION_TREINNR_COMB, ’-’, ”))) / 2
> @Counter_split

Set @Counter_split = @Counter_split + 1
End

Drop Function Split
```

The table ’Transfer_temp’ contains now a list of all transfer stations, but it includes also many duplicate values. First all the distinct stations needs to be found and stored in a second temporary table, called ’Transfer_distinct_temp’. Then those stations can get a unique number. The final result is saved in the table ’Transfer’. The last step is to manual delete the first temporary table.

```
;
With Transfer_distinct_temp as (
Select Distinct Transfer
From Transfer_temp )

Select *, Row_Number () Over (Order by Transfer) as Number
Into Transfer
From Transfer_distinct_temp

Drop Table Transfer_temp
```

In contrast to the determination of the transfer stations, the creation of the ’Selection’-table is relatively easy. The algorithm determines every trip where the promised journey contains the name of the transfer station between two hyphens. The names of the origin and destination station are also in the promised journey but those names do not have a hyphen before and after the name.

```
Set @Transfer = (Select Transfer From Transfer Where Number = @Counter_transfer)

Select *
Into Selection
From EZ_D_RPU.FCT_RPUN_REIS
Where RS_STATION_TREINNR_COMB like ’%-’ + @Transfer + ’-%’ and RS_VTG_CATEGORIE not in
(’Busvervoer’ , ’Andere vervoerder’)
```

Breakdown per train series

The ’EZ_D_RPU.FCT_RPUN_REIS’-table contains a column with the train series. The name of this column is ’RS_VTG_TREINSERIE1’. This train series is the last train series used in the promised journey if the trip is realised. If not, and the trip needs to be rescheduled, the train series is the series which causes the disturbance. In the case of a missed transfer it is the arriving train series. This column is always filled with a value, except in some cases when the category of the trip is ’Bus transport’ or ’Other rail operator’. So the first step to perform a breakdown per train series is to determine all the train series in the network.

```
Select RS_VTG_TREINSERIE1
Into Train_temp
```

```
From EZ_D_RPU.FCT_RPUN_REIS
Where RS_VTG_CATEGORIE not in ('Busvervoer','Andere vervoerder')
```

If a train is cancelled, a substitute train can run on the same time and on the same route as the original train. In the case that this substitute train was not original planned, the train gets a train number which is 300000 higher than the train number of the original train. If the substitute train is planned, the train number is 700000 higher. The train series is then also 300000 or 700000 higher. However for a passenger this is not important. So the 'Train_temp'-table should be adjusted for this phenomenon.

```
Update Train_temp
Set RS_VTG_TREINSERIE1 = Case
When RS_VTG_TREINSERIE1 between 300000 and 400000
Then RS_VTG_TREINSERIE1 - 300000
When RS_VTG_TREINSERIE1 between 700000 and 800000
Then RS_VTG_TREINSERIE1 - 700000
Else RS_VTG_TREINSERIE1
End
```

At this point the 'Train'-table can be made. This table consist of every train series with a unique number. To create this table a temporary table called 'Train_series_temp' is used. At last, the 'Train_temp'-table can be deleted.

```
;
With Train_series_temp as (
Select Distinct *
From Train_temp )

Select *, Row_Number () Over (Order by RS_VTG_TREINSERIE1) as Number
Into Train
From Train_series_temp

Drop Table Train_temp
```

The creation of the 'Selection'-table does not differ very much from the original script. Here also the possibility for 300000- and 70000-train series needs to be taken into account. However it is possible that a train series is in use by another rail operator. In that case the number of trips in the 'Selection'-table is zero. Calculating the passenger punctuality is in this case useless. Therefore the second loop can be skipped. This is done via a Goto-statement.

```
Set @Trainseries = (Select RS_VTG_TREINSERIE1 From Train Where Number = @Counter_train)

Select *
Into Selection
From EZ_D_RPU.FCT_RPUN_REIS
Where (RS_VTG_TREINSERIE1 = @Trainseries or RS_VTG_TREINSERIE1 = @Trainseries + 300000 or
RS_VTG_TREINSERIE1 = @Trainseries + 700000) and RS_VTG_CATEGORIE not in ('Busvervoer' ,
'Andere vervoerder')

Set @Total = (Select Count(*) From Selection)
If @Total = 0 Goto Skip_loop
```

The code after the End-statement of the second loop is as follows.

```
Skip_loop:
```

B.2.2. Changes in Part 2: Creating of the graphs

This paragraph describes the changes in the original MATLAB script for creating special breakdown graphs. Aside for some variable, folder and file names and the title, the changes in all the types of special graphs are the same: the including of the 95% confidence interval in the figure. This is done by plotting the upper and lower bound of the interval and by giving the maximum deviation of the 3 minute passenger punctuality.

The first step is to calculate the upper and lower bounds of the confidence interval of the passenger punctuality. This is done via formula B.1.

$$d \leq \sqrt{\frac{Z^2}{n} * p(1-p)} \tag{B.1}$$

Here is p the passenger punctuality, n the number of trips and Z is 1.96 by a reliability of 95%. The upper bound of the interval is the passenger punctuality plus this deviation. However this cannot be more than 100%. The lower bound of the passenger punctuality is the passenger punctuality minus the deviation. Here the value cannot be smaller than 0%. The location of this part of the script is just before the graph is made.

```
p = selection(:,9)/100;
d = sqrt((1.96ˆ2).*p.*(1-p)./total_trips)*100;
upper = selection (:,9)+ d;
upper(upper > 100) = 100;
lower = selection (:,9)- d;
lower(lower < 0) = 0;
```

The next step is to change the content of the text box. The maximum deviation of the 3 minute passenger should be included. Note that thus the size and location of the text box also changes.

```
content = ['Number of trips: ' num2str(total_trips)], ['Maximum delayed trips: ' num2str(sum_share)],
['d (95%) ' char(8804) ' ' num2str(round(d(3)*100)/100) ' percentage point'];
annotation('textbox', [0.61,0.57,0.28,0.11],'String', content, 'Tag' , 'Toberemoved');
```

The last step is to plot the upper and lower bound of the confidence interval. This can be done just after the previous change. The already created figure should not be deleted, so the 'hold on'-statement is used.

```
hold on;
plot(selection(:,1),lower,'b');
plot(selection(:,1),upper,'b');
hold off;
```

C

Explainable Train Deviations

This appendix consists of a list of all possible types of 'Explainable Train Deviations' (TVTAs). All the TVTAs can be placed into nine type categories, divided over four responsible parties. This subdivision can be found in Table C.1. The code corresponds with the number of each type of TVTA. The list with all 37 types of TVTAs is split up into two parts: Table C.2 and Table C.3.

Table C.1: The subdivision of the TVTAs

Responsible party	**Type category**	**Code**
Transport companies	Train disruption	A
	Commercial reasons / process	B
	Personnel	C
	Rolling stock	D
Infrastructure manager	Capacity distributor / Traffic control	E
	Broken infrastructure	F
	Civil engineering's causes	G
Third parties	Third party causes	H
Extern	External causes	I

Table C.2: The first part of the list with types of TVTAs [36]

	Name	**Explanation**
A1	Delay own train	A delay caused because the same train is already delayed. Travel without conflict is therefore not possible any more
A2	Delay train - train	A delay caused because another train is delayed
A3	Adjustment to timetable because of large delay	Because of a large delay a train is (partly) cancelled
A4	Connection	A delay caused because the train has to wait for a delayed train to make a connection
B1	Exceeding stopping time, route set	A delay caused because the stopping time is exceeded while the route is already set
B2	Wish of rail operator to adjust timetable	The rail operator files a request to change the timetable
B3	Passengers/loading process	A delay caused by passengers or by problems with the loading process
B4	Preparing train / No wagon list	A delay caused because the train is not yet prepared by departure or the paperwork is not correct
B5	Import delay abroad	A delay caused abroad but offered in the Netherlands

Table C.3: The second part of the list with types of TVTAs [36]

	Name	**Explanation**
C1	Absence personnel	Because absence of the personnel a train is delayed or cancelled
C2	Work process travelling personnel	A delay caused by not or bad performing of the tasks by the travelling personnel
C3	Reschedule personnel	A delay caused because the rail operator did not or badly schedule the personnel
D1	Reschedule	A delay caused because rolling stock was not available because of a fault of the transport control
D2	Combine trains	A delay caused because the combining of trains was not ready on time
D3	Broken rolling stock	A delay caused because rolling stock is broken
D4	Hazardous materials incident	A train deviation due to an incident involving hazardous materials
E1	Conflict in timetable	A delay caused by a conflict in the timetable
E2	Work process signaller	A train deviation caused by the signaller
E3	Deviation of priority rules	A delay caused by deviations of the priority rules initiated by the traffic controller
E4	Unclear system behaviour	A delay caused due to unclear behaviour of the process control system or due to measurement errors
F1	Safety installations	A delay caused due to malfunction of the safety installations
F2	Malfunction level crossing	A train deviation because there is a malfunction at a level crossing
F3	IT / Telecommunication	A delay due to malfunction of the telecommunication system
F4	Power supply	A train deviation because of a power supply malfunction
F5	Tracks or switches	A train deviation due to malfunction of the railway tracks
F6	Buildings / Infrastructure object	A train deviation due to malfunction of a building or infrastructure object
G1	Unplanned maintenance	A delay caused because of unplanned maintenance
G2	Speed limit restrictions due to broken tracks	A delay caused due to lower speed limits because the railway track is broken.
G3	Irregularities in the execution of the possession planning	A delay due to irregularities in the execution of the possession planning
H1	Influence of weather on the infrastructure	A train deviation due to extreme weather
H2	Influence of a third party on the infrastructure	A train deviation due to influence of a third party
H3	Almost collision / increased risk along the tracks / SPAD	A train deviation due to an almost collision, a signal past at danger, or a person or animal along the tracks
H4	Collision	A train deviation due to a collision
H5	Delay due to assistance emergency services	A delay caused because there is a need for assistance of emergency services or the customs agency
H6	Strike	A train deviation due to a strike of the railway personnel
I1	Delay due to other infrastructure manager	A train deviation in an area controlled by ProRail due to problems at another infrastructure manager
I2	Delay due to a takeover of a train of another rail operator	A delay caused by the takeover of a train of another rail operator

D

Algorithms to calculate process indicators

To calculate the process indicators (PI) for the entire network multiple data sources are used. These data sources are tables where every row is unique. The information of each row is used to calculate the process indicator. To calculate the process indicators for a subset of the network, these data sources need to be filtered. However not all the information needed is present in the data. So before the process indicators for a subset can be calculated, the raw data needs to be enriched. The algorithm to do this is described in paragraph D.1.

The actual algorithm to calculate the process indicators is described in paragraph D.2. Here the enriched data is loaded and filtered to retrieve only the data needed to calculate the process indicators for a specific subset of the rail network.

D.1. Algorithm to enrich the PI data

This paragraph describes the algorithm to enrich the raw data to calculate the process indicators (PI). It consists of a main algorithm (paragraph D.1.1) and multiple functions. A function is a part of the algorithm in another file and can be called on in the main algorithm or in another function. The result of this algorithm is six data files per examined day.

There are two reasons to have a function. The first reason is because the code in the function is repetitive amongst the algorithm. To place this part of the code in a separate function, the overall algorithm becomes less extensive and less complicated. The second reason is because of clearness. The total algorithm is long. To place certain parts of the algorithm in a separate function, the total algorithm stays more clear, certainty when such a function has a goal on its own.

The main part of the algorithm is written in MATLAB. However at certain points data out of the passenger punctuality database is required. Here the output tables are quite extensive, so this data is stored in an SQL database. To retrieve the desired data, some parts of the algorithm are written in the Structured Query Language.

D.1.1. Main algorithm

In short the main algorithm is the user interface for the entire algorithm. Here the user can set some input values, like the settings of the SQL server and the dates of which the data needs to be enriched. After some checks if the input data is correct, multiple function are called on. Every function enriches the data of a single data source.

In the first part of the algorithm the user can alter five settings. The first two are the start and end date of the period of which the data needs to be enriched. The format of the input may vary. Correct values are '01JUN2015' or '2015-06-01'. The last three settings are settings of the SQL-server. Secondly, the inputs are checked by comparing the start date with the end date and by trying to make a connection with the SQL database.

```
Startdate = ’01JUN2015’;
Enddate = ’2015-06-01’;
server_name = ’localhost’;
database_name = ’RPUN’;
portnumber = 50637;

Startdate = datenum(Startdate);
Enddate = datenum(Enddate);
if Startdate <= Enddate

conn = database(database_name, ”, ”, ’Vendor’, ’Microsoft SQL Server’, ’Server’, server_name,
’AuthType’, ’Windows’, ’PortNumber’, portnumber);
if isopen(conn)
```

For almost every data source, there is a single file for every day. This is however not the case for the data about the rolling stock and personnel. Here only one file is available. So these files are loaded here. Note that the location of these files is the same as the location of this script. The output data, the raw data enriched with possibilities to filter, needs to be stored in separate folders. If those folders do not exist, these are created here.

```
fid = fopen(’MT1.csv’,’rt’);
MTdata = textscan(fid,repmat(’%s’,1,11),’Delimiter’,’;’,’CollectOutput’,1);
fclose(fid);
MTdata = MTdata{1,1}(2:end,:);

fid = fopen(’PS1.csv’,’rt’);
PSdata = textscan(fid,repmat(’%s’,1,10),’Delimiter’,’;’,’CollectOutput’,1);
fclose(fid);
PSdata = PSdata{1,1}(2:end,:);

if exist(’VSM_filter’,’dir’) ~= 7
mkdir(’VSM_filter’)
end
if exist(’VKL_filter’,’dir’) ~= 7
mkdir(’VKL_filter’)
end
if exist(’MT_filter’,’dir’) ~= 7
mkdir(’MT_filter’)
end
if exist(’PS_filter’,’dir’) ~= 7
mkdir(’PS_filter’)
end
if exist(’VLM_filter’,’dir’) ~= 7
mkdir(’VLM_filter’)
end
if exist(’TAD_filter’,’dir’) ~= 7
mkdir(’TAD_filter’)
end
```

At this point, the data can be enriched. For every data source, a unique function is called on. These functions are described in paragraph D.1.2 to D.1.7. The raw data is divided per day, so these functions are placed in a for-loop. Note that the format of the date in the personnel data is different from the other data sources. After the for-loop, the script can be ended.

```
for i = Startdate:Enddate
date = datestr(i,’dd-mm-yyyy’);
selection = PSdata(strcmp(date,PSdata(:,1)),:);
```

```
date = datestr(i,'yyyy-mm-dd');
PS_func(date,selection);
VSM_func(date,conn);
VKL_func(date);
selection = MTdata(strcmp(date,MTdata(:,1)),:);
MT_func(date,selection);
VLM_func(date, conn);
TAD_func(date, conn);
end

else
disp('There is no connection with the SQL database')
end
else
disp('The dates are not correct');
end
```

D.1.2. Function: PS_func

This function enriches the data concerning personnel. This data consist per day and per train series the number of planned staff members and the actual number of staff members. This data should be enlarged with the route of the train series. This route can change between the direction (odd or even), so a distinction between the directions should be made. The information about the route is stored in a datafile called 'list-linten_2015-06-01.dat', where of course the date can change. This file contains per train series and direction the route that is most often run that day and the longest route of that day. Here the route is presented as a list (a ribbon) of consecutive service control points and the action (arrival, departure or passage). The function starts with the loading of the data about the ribbons.

Next for every row in the data, the corresponding route should be found. This is done in multiple steps. For every row in the ribbon data a check is made if the train series is a passenger train. Then the number of the train series of the ribbon is compared with the number of the train series of the personnel data. Lastly the direction is determined.

Following, the route should be placed in the right format in the right variable. As stated, in the ribbon data per train series there are two routes: the most often travelled route and the longest route. At this point, the most often travelled route is used. Nonetheless if both routes are not the same, the longest route should also be put in the right format in the right variable.

The involved service control points are the service control stops on the route whit a departure or an arrival. These service control points can be found by adding all possible routes per train series and deleting those points with only a passage.

The found routes and the involved service control points should be added by the existing data. When this is done for every row in the personnel data, the enriched data should be exported to a new data file. Note that the result is a file per day, instead of only one overall file, like the raw data. The names of the columns of the exported file are in Dutch, like in the raw data. The format of the output file can be found in Table D.1.

Table D.1: The format of the enriched PS-file

Column	**Content**
1-10	Original PS data
11	Involved stations
12	Route, in the even direction
13	Route, in the odd direction

D.1.3. Function: VSM_func

There are nine process indicators concerning blockage measures. This function enriches the raw data concerning these measures. First of all, the lead times between the different steps between disruption via blockage measures to running according to plan again are calculated. Second the involved train series are determined. Per train series the route of the train series is determined.

The raw data consist of a single row per blockage measure. The enriched data has a row per combination of blockage measure and involved train series. If multiple train series are involved in a blockage measure, the enriched data has thus multiple rows for this blockage measure.

To start, the raw data needed is loaded. This is the data concerning the blockage measures and data about the routes (ribbons) of the trains. Of each train the planned route and the realised route is present in this data. This data is edited in the 'Ribbondata_func' function (see paragraph D.1.14). Later in this function some data is requested out of the passenger punctuality database. The query to do so is also loaded. This query is described in paragraph D.1.17.

For every blockage measure the time period and the lead times should be calculated. The time the blockage measure ended is determined in a new function: 'VSM_Maxtime', described in paragraph D.1.8. The lead times are calculated in a separate function as well, called 'VSM_leadtime_func' (paragraph D.1.9). With the location of each blockage measure the query can be updated and the involved trains with their route can be determined. This last bit is here also done in a separate function, described in paragraph D.1.10: 'VSM_trainseries_func'.

When the lead times and the involved train series are known, the new data file can be constructed and exported. In Table D.2 the format of the output file is described. In the original VSM-file every row contains of a single VSM. In this new file this is not the case. The number of rows for the same VSM is equal to the number of train series hindered by the VSM. In practise this means that column 1 to 37 and 46 contain the same information for multiple rows. Column 38 to 45 contains unique information per row.

Table D.2: The format of the enriched VSM-file

Column	**Content**
1-29	Original VSM-file
30	Lead time 'Known VL' and 'Known BackOffice'
31	lead time 'Known BackOffice' and alarm
32	lead time alarm and Distribution decision
33	lead time Distribution decision and Blockage measure decision
34	Lead time Blockage measure decision and running according to the Blockage measures
35	Lead time end of Blockage and running according to regular plan
36	Begin of the disruption
37	End of the disruption
38	Train series hindered by the disruption
39	Involved train numbers
40	Involved stations according to the timetable
41	Route according to the timetable
42	Involved stations according to the realisation
43	Route according to the realisation
44	Train did run during the VSM
45	Train was planned during the VSM
46	Unique number per VSM per day

D.1.4. Function: VKL_func

One process indicators counts the number of mutation is VKL, the system used by dispatchers and signallers. Such a mutation consists of a train number, a time and the start and end location of the train. This function will enrich this data with the route of the train, according to plan and realisation, if available. Therefore this function will start by loading not only the raw VKL data but also by loading the ribbon data, the planned and realised route of every train. The train numbers of this data will be stored in a separate variable to find the correct value later in the script.

For every mutation the train number and the start and end location should be determined. With this train number the planned and realised route can be found. Also, if the train number is not present in the ribbon data, the remainder of the script should be ignored.

The planned and realised route should be converted into the right format. Only the part of the route between the start and end location is needed as well. To convert and trim the route, two separate functions are used: 'route_func' (see paragraph D.1.15) and 'route_func_extra' (see paragraph D.1.16). Finally the result should be exported to a new data file. The format is shown in Table D.3.

Table D.3: The format of the enriched VKL-file

Column	Content
1-13	Original VKL-file
14	Involved stations according to the timetable
15	Route according to the timetable
16	Involved stations according to the realisation
17	Route according to the realisation

D.1.5. Function: MT_func

The function to enrich the data about the rolling stock is similar to the PS_func (see paragraph D.1.2). The differences are the location of the columns and the output file. The format of the output file can be found in Table D.4.

Table D.4: The format of the enriched MT-file

Column	Content
1-11	Original MT data
12	Involved stations
13	Route, in the even direction
14	Route, in the odd direction

D.1.6. Function: VLM_func

There are two process indicators related to empty rolling stock. In the VLM data all empty rolling stock trips are stored, with their planned and realised departure time. The first goal of this function is to find the route of each empty rolling stock trip. Information about the route (ribbon) of each train, even empty trips, can be found in an existing data source. So in the first part not only the raw VLM data is loaded, but also the ribbon data.

Next, the route of each empty rolling stock trip is determined. A distinction is made between the planned route and the realised route. The involved service control points are determined as well. These are the service control points where the train departed or arrived. Hereby the 'route_func' function is used (see paragraph D.1.15).

The second goal of this function is to find the passenger trains that might be hindered or affected by each trip. This is done in a separate function called 'VLM_trainnumber_func' (paragraph D.1.11). Besides the VLM data and the ribbon data this function needs data out of the passenger punctuality database as well. Therefore a query is loaded: 'VLM_sql'. This query is described in paragraph D.1.18. The result of the function can be exported to a new data file. In Table D.5 the format of this file can be found.

D.1.7. Function: TAD_func

A Train handling document (TAD) describes the priority rules of and waiting times of trains by small delays at a station or on a section of a route. The TAD data consist of all the names of the Train handling documents in combination with the train number of the train that triggered the use of the train handling document. A difficulty lays in the fact the location or route and the involved train series of each TAD are only present in the name of the TAD. Therefore the first step of this function is to load the TAD data and to filter the location or route and involved train series out of the name of each TAD.

Table D.5: The format of the enriched VLM-file

Column	Content
1-14	Original VLM-file
15	Involved stations according to the timetable
16	Route according to the timetable
17	Involved stations according to the realisation
18	Route according to the realisation
19	Trains that run during the empty rolling stock trip according to the timetable
20	Trains that did not run during the empty rolling stock trip according to the timetable
21	Trains that run during the empty rolling stock trip according to the realisation
22	Trains that did not run during the empty rolling stock trip according to the realisation
23	Train series that run during the empty rolling stock trip according to the timetable
24	Train series that did not run during the empty rolling stock trip according to the timetable
25	Train series that run during the empty rolling stock trip according to the realisation
26	Train series that did not run during the empty rolling stock trip according to the realisation

In the name of the TAD the involved trains are listed. The timetable repeats itself every 15, 30 or 60 minutes, so a TAD can be used multiple times in a day. In such cases not the involved train numbers, but the involved train series are listed. To determine the exact routes of the involved train series and train numbers, two data files are loaded. The first data file contains the routes (ribbons) of each train series. The second data file contains the route per train number. With these two data files for every involved train number of series the route can be found.

Lastly the departure and arrival times of every train on every station needs to be known. This information is available in the passenger punctuality database. With a query ('TAD_sql', see paragraph D.1.19) this data can be retrieved. To make calculations with this data easier the arrival and departure times (both according to plan and the realised time) are transformed to values. Also, the retrieved data out of the SQL-server is very large. To make calculations faster, the data is split by train number and stored in smaller tables. This implies that only smaller tables needs to be searched if departure or arrival times needs to be known.

The TAD data and ribbon data can be combined to find the exact route for each TAD. This is done in a separate function called 'TAD_route_func' (see paragraph D.1.12). Next the, already enriched, TAD data is combined with the arrival and departure times to find the time period of each TAD. This is done in a separate function as well: 'TAD_time_func' (paragraph D.1.13). Before the result can be exported to a new data file, the involved train series should be converted to the right format. The format of the new data file is shown in Table D.6.

Table D.6: The format of the enriched TAD-file

Column	Content
1-7	Original TAD-file
8	Borders of the TAD
9	Possible involved train series
10	Involved train series
11	Involved stations
12	Route
13	Begin of the TAD according to the timetable
14	End of the TAD according to the timetable
15	Begin of the TAD according to the realisation
16	End of the TAD according to the realisation

D.1.8. Function: VSM_Maxtime

This function is used in the function 'VSM_func' (see paragraph D.1.3). For each blockage measure and for multiple actions a time is logged. However is depends per blockage measure which times are logged. First this function finds all the logged time stamps. Then this function will compare all time stamps to find the latest logged time stamp. Herewith the time the blockage measure ends is found.

D.1.9. Function: VSM_leadtime_func

This function is an extension to 'VSM_func' (see paragraph D.1.3). This function calculates the lead times between the possible steps in the process between a disruption and the end of a blockage measure. Six lead times are possible so if the data is available, six times the lead time should be calculated.

D.1.10. Function: VSM_trainseries_func

The function 'VSM_func' (paragraph D.1.3) enriches the data about blockage measure with external data. This function is a part of that. It determines the involved trains, the section of their route what corresponds with the location of the blockage measure and whether the planned trains did run during the blockage measure.

First off, the train numbers involved are determined. This is done via the Ribbon data: data about the route of each train. If (a part of the) route corresponds to the location of the blockage measure, the data about this train is saved. This part of the function uses another function, called 'route_func_extra' (see paragraph D.1.16). Note that both the planned route as the realised route is examined.

At this point all possible train numbers and their stops on the section of their route where the blockage measure applies are known. The next step is therefore to determine which of these trains passed this section during the blockage measure. To do so, first all unique sections are determined. For every unique section a query is constructed to find which train numbers passed the blockage measure, according to the timetable. This query can be found in paragraph D.1.17. The data used is derived from the passenger punctuality database. The outcome of this query is combined with results of this function.

Next, the results can contain some rows where all the data is unique, except for the train number. This last part of the function finds those rows and combines it into one single row, where the train numbers are listed together.

D.1.11. Function: VLM_trainnumber_func

This function is part of 'VLM_func' (see paragraph D.1.6). Here the passenger trains that are hindered or affected by an empty rolling stock trip are determined. First data is retrieved from the passenger punctuality database. Paragraph D.1.18 describes the query used. this data consist of arrival or departure times of all the train on all the stations. To make the calculations easier, the planned and realised arrival or departure times are converted into values, instead of dates. The retrieved data is extensive, so it will be divided into a table per train series.

Next all possible routes of the passenger trains are determined. A distinction is made between planned routes and realised routes. The source is the Ribbon data, edited in another function: 'Ribbondata_func' (see paragraph D.1.14).

The following step is to find the trains corresponding with the list of unique routes as constructed previously. For every train the planned and realised arrival or departure time at every station is determined as well, based on the data out of the passenger punctuality database.

Just like the determination of the unique routes of the passenger trains, the same has to be done for the routes of the empty rolling stock trips. Here also a distinction is made between the planned route and the realised route.

For every combination of unique route of a passenger train and unique route of an empty rolling stock trip a determination should take place to check whether those routes overlap. Two routes overlap if at least two service control points are the same. The results are stored in a table. Here as well, a distinction is made between the planned routes and the realised routes.

For every empty rolling stock trip the trains that run (partly) on the same route can now be found. The next step is to determine the time the involved trains run on the same route as the empty rolling stock trip. If the departure time of empty rolling stock trip lies within the time period the involved train is on the same route, the conclusion is drawn that the train is affected by the empty rolling stock trip. Lastly the results should be stored in the correct place in the output table.

D.1.12. Function: TAD_route_func

As an extension to 'Tad_func' (see paragraph D.1.7) this function will find the exact route for each train handling document. This is done by first finding the routes of all involved trains or train series. Next the common service control points are determined. A condition hereby is that the service control points lies within the borders of the route of the train handling document. The first step to achieve this is to find the involved train numbers or train series and to determine the train series of the decision train, the train that triggered the use of the train handling document.

The second step is to determine the train series of all involved train numbers. If only the train series is known, it should be placed in the right format. Hereafter the route of each train or train series can be determined. This is done with the help of another function: 'route_func' (see paragraph D.1.15). A difficulty lies in the fact that there is a possibility that an involved train series in reality a list of train series. Of that list, only one of the train series is involved in the train handling document. An underlying assumption is that train to Germany have always an odd train number, while trains with a destination in Belgium have always an even train number.

The third step is to determine the borders of route where the train handing document applies. If the train handling documents applies to only one station, there will be no borders or the name of the borders will be the same. In this case, the results can be stored in the output table.

The remainder of the train handling documents apply to a section in the rail network. So the next step is to find which of the involved trains or train series has a route that goes to both borders of the section where the train handling document apply. via this method the exact route between the borders can be determined. Multiple trains or train series have such a route, so the largest option is chosen.

Finally the exact route can be determined. Here the route is the group of service control points that every involved train or train series have in common. These service control points are listed and the result is stored in the output table. Every station between the borders is determined as well. These are the service control points of the exact route where a train had an arrival or a departure.

D.1.13. Function: TAD_time_func

This function is an extension to 'TAD_func' (see paragraph D.1.7). Here the arrival and departure times of the involved trains on the involved stations are determined. These times are used to make decisions about which train number applies, when multiple trains from the same train series are possible. This will finally result in a time period per train handing document. Nonetheless first the involved trains and the borders of the route needs to be retrieved from the data.

Next for every involved train or train series the possible departure and arrival times needs to be determined. If the train handling documents applies to only one station, both the arrival and departure times are taken into account. If the train handling documents applies to a route, only a departure or arrival is taken into account when the station is a border station of the route.

It is possible, certainly when only an involved train series is known, that there are multiple arrival and departure times at the same station. The train that triggered the train handling document is known, so the arrival and departure times closest to the departure or arrival time of that train is chosen to be involved arrival and departure times.

If there are, per train series, still multiple train numbers possible, the train number is chosen where the departure or arrival is the closest to the mean of all involved planned times of the train that triggered the train handing document. Via this method only one train number per involved train series is determined.

At this point all involved departures and arrivals are known. The earliest of all these times is set to be the beginning of the train handing document. The latest involved arrival or departure is the end of the train handling document. A distinction is made between the planned times and the realised times. Finally the results are stored in the output table.

D.1.14. Function: Ribbondata_func

This function edits the ribbon data of each train. The ribbon data is the route of each train, both planned and realised. This function is used in 'VSM_func' (paragraph D.1.3) and 'VLM_trainnumber_func' (paragraph D.1.11). First this function will separate the passenger trains from the other trains, like freight trains or empty rolling stock trips. This implies that the output only consist of passenger trains. Next the stations are filtered out of the ribbon data. Here these stations are called involved service control points. This is partly done with another function called 'route_func' (see paragraph D.1.15). These service control points and the route itself are stored in the output table. A distinction is made between the planned routes and the realised routes. The last part of this function determines the direction of trains from and to Germany and Belgium.

D.1.15. Function: route_func

In six other functions this function is used: 'VSM_func' (paragraph D.1.3), 'VKL_func' (paragraph D.1.4), 'VLM_func' (paragraph D.1.6), 'VLM_trainnumber_func' (paragraph D.1.11), 'TAD_route_func' (paragraph D.1.12) and 'Ribbondata_func' (paragraph D.1.14). The goal of this function is to filter the service control points out of a route. The output contains this list of service control points, and a list of the service control points with only an arrival or departure.

D.1.16. Function: route_func_extra

This function is used in 'VSM_trainseries_func' (paragraph D.1.10). The input is a route and two service control points. A section is defined as the part of the route between the two service control points (the border service control points). This function will determine all the stations on the section. If one or both of those border service control points is not a station, the section will expand so the borders of the section will all be stations. So to start, this function will determine the location of the border service control points.

To find the involved stations between the borders (with or without expanding) a distinction is made between the fact if the border service control points are the same of not. Depending if the border service control points are the same of not, a different part of the function applies.

D.1.17. Query: VSM_sql

In 'VSM_func' (paragraph D.1.3) this query is needed to retrieve data out of the passenger punctuality database. It is written in Structured Query Language. This query retrieves the train numbers of the trains that during a blockage measure run, both according to the timetable as in the realisation.

D.1.18. Query: VLM_sql

In 'VLM_func' (paragraph D.1.6) this query is needed to retrieve data out of the passenger punctuality database. Here for a specific date for every station and every train the arrival time is determined. Is there no arrival time, for example because the trip started at a certain station, the departure time is used.

D.1.19. Query: TAD_sql

This query is used in 'TAD_func' (paragraph D.1.7). Here all departure and arrival times of all trains at all stations is retrieved from the passenger punctuality database.

D.2. Algorithm to calculate the process indicators

The structure of the algorithm to calculate the process indicators consist of a main algorithm and multiple functions. A function is a separate part of the algorithm and can be called on in the main algorithm or in another function. Here functions are used to make the algorithm more structured. The whole algorithm is written in MATLAB. The outcome is a file with the values of every process indicator with respect of the filtering for every date examined.

D.2.1. Main algorithm

This main algorithm functions as user interface for the entire algorithm. In the first part 10 different variables can be set. The first two variables determine the dataset used to calculate the process indicators. With the other eight the data can be filtered. If such a variable is left empty, it will be ignored when filtering the data.

- **Startdate:** The first day examined.

- **Number_days:** The number of days examined. Every day result in a unique value for each process indicators.
- **DRP:** Service control points.
- **StartDRP:** This variable only affects the process indicators concerning the performance of the start-up of the day. If this value is ’Yes’, the service control point is only taken into account if at this service control point one or more train series started.
- **Whole_Route:** If this value is ’Yes’, only data is used if all service control points are included in the route. Otherwise only 2 service control points are needed (only 1 if only a single service control point is given.
- **Trainseries:** Number of a train series. If an ’E’ (even) or ’O’ (odd) is included, the direction is taken into account.
- **Trains:** This variable only affects the process indicators concerning the Blockage measures and the empty rolling stock runs. If this value is ’Yes’, only data is used if during the Blockage measure or during the empty rolling stock run also a passenger train was planned.
- **TimeMoment:** The moment of the day. This value can be ’Day’ (the whole day), ’Peak’ (only peak hours) or ’Off peak’.
- **Starttime:** If the moment of the day is empty, with this variable the first border of a time period can be set.
- **Endtime:** If the moment of the day is empty, with this variable the second border of a time period can be set.

So in the example below, the data used to calculate the process indicators is from August, September and October 2015. Only trains of train series 800 between stations Utrecht Centraal and ’s-Hertogenbosch between midnight and 6 o’clock in the morning are taken into account.

```
Startdate = ’2015-08-01’;
Number_days = 92;
DRP = [’Ut’, ’Utl’, ’Htn’, ’Htnc’, ’Cl’, ’Gdm’, ’Zbm’, ’Ht’];
StartDRP = ”;
Whole_Route = ”;
Trainseries = ’800’;
Trains = ”;
TimeMoment = ”;
Starttime = ’00:00:00’;
Endtime = ’05:59:59’;
```

In the next part the process indicators are calculated. This is done via a separate function per enriched data file. So there are six separate functions. Finally, the outcome is exported to an output file.

```
Startdate = datenum(Startdate);
PI_output = cell(Number_days,16);
for i = Startdate:Startdate + Number_days-1 date = datestr(i,’yyyy-mm-dd’);
PI_output(i-Startdate + 1,1) = cellstr(date);
PI_output(i-Startdate + 1,2:10) = VSM_PI(date, DRP, Whole_Route, Trainseries, Trains,
TimeMoment, Starttime, Endtime);
PI_output(i-Startdate + 1,11) = VKL_PI(date, DRP, Whole_Route, Trainseries, TimeMoment, Starttime,
Endtime);
PI_output(i-Startdate + 1,12) = MT_PI(date, DRP, StartDRP, Whole_Route, Trainseries);
PI_output(i-Startdate + 1,13) = PS_PI(date, DRP, StartDRP, Whole_Route, Trainseries);
PI_output(i-Startdate + 1,14:15) = VLM_PI(date, DRP, Whole_Route, Trainseries, Trains,
TimeMoment, Starttime, Endtime);
PI_output(i-Startdate + 1,16) = TAD_PI(date, DRP, Whole_Route, Trainseries, TimeMoment, Starttime,
Endtime);
end

Output = cell2table(PI_output, ’VariableNames’, {’Date’, ’Leadtime_KnownVL_KnownBO’,
’Leadtime_KnownBO_Alarm’, ’Leadtime_Alarm_DistributionDecision’,
```

```
'Leadtime_DistributionDecision_BlockageMeasures',
'Leadtime_BlockageMeasures_RunningViaBlockage', 'Leadtime_EndBlockage_RunningPlan',
'BlockageMeasure_and_DistributionDecision', 'Complete_BlockageMeasure',
'Unadapted_BlockageMeasure', 'Number_Mutations', 'Correct_Composition', 'Correct_Staffing',
'DeparturePunctuality_Railyard', 'DeparturePunctuality_Station', 'TrainHandlingDocument'});
Outputname = 'PI_Output.dat';
writetable(Output,Outputname,'Delimiter',';');
```

D.2.2. Function: VSM_PI

This function calculates the process indicators based on the enriched VSM data. First the data is read. Every Blockage measures consist of multiple rows in the data, so the data should be split and per blockage measure a determination should took place if the blockage measure passes the filters.

So the next step is to filter the data. First the algorithm checks if the service control points are involved in the blockage measure. Here another function is used: 'DRPinRoute_func' (see paragraph D.2.8).

The final filters are the train series filter and the time moment filter. Filtering the data here is partly done via a second function, called 'VSM_timefilter_func' (paragraph D.2.11). If data gets through every filter, it is stored in a special table.

The last part of the function is to calculate the nine process indicators. The first six are the lead times. This is done in a separate function: 'VSM_LT_func' (see paragraph D.2.9).

D.2.3. Function: VKL_PI

This function, that calculates the process indicator out of the VKL data, is comparable with the function 'VSM_PI' (see paragraph D.2.2). First the data is read, than filtered and in the end the process indicator is calculated. By the filtering by service control points another function is used: 'DRPinRoute_func' (paragraph D.2.8). By the filtering by moment of day the function 'timefilter_func' (paragraph D.2.10) is used.

D.2.4. Function: MT_PI

This function determines the percentage of trains with the right composition at the morning start. After the data is read, it is filtered by service control points and train series. The value of this process indicator will not change by filtering on a time moment (like peak hours), so this is not possible. The filtering is partly done via another function, called 'DRPinRoute_func' (see paragraph D.2.8).

D.2.5. Function: PS_PI

The function to calculate the percentage of trains with the right number of personnel at the morning start is very similar with the 'MT_PI' function (see paragraph D.2.4). The format of the enriched data file is comparable as well. The only difference between the two functions is the location of some columns in the data file and the name of some variables.

D.2.6. Function: VLM_PI

Out of the VLM data two process indicators can be calculated: the departure punctuality for empty rolling stock runs at rail yards and at stations. First this function reads the data. Next it gets filtered. Here two other functions are used: 'DRPinRoute_func' (see paragraph D.2.8) and 'timefilter_func' (see paragraph D.2.10). In the end, the process indicators are calculated.

D.2.7. Function: TAD_PI

This function calculates the percentage of train handling documents that is applied correctly. This is partly done via two other functions: 'DRPinRoute_func' (see paragraph D.2.8) and 'timefilter_func' (see paragraph D.2.10).

D.2.8. Function: DRPinRoute_func

This function calculates the place of a service control point in a route. Here a route is a list of multiple service control points. In this function a distinction is made between a single service control point as input, and multiple service control points. In the second case, the output is multiple places, corresponding with every service control point.

D.2.9. Function: VSM_LT_func

This function is used in 'VSM_PI' (paragraph D.2.2). Here the 90th percentile is calculated for a set of lead times.

D.2.10. Function: timefilter_func

This function determines if a point in time lies within two time limits. This function is used in 'VKL_PI' (paragraph D.2.3), 'VLM_PI' (paragraph D.2.6) and 'TAD_PI' (paragraph D.2.7).

D.2.11. Function: VSM_timefilter_func

This function is based on the 'timefilter_func' function (paragraph D.2.10) and made especially for the function 'VSM_PI' D.2.2). In this function a determination takes place to check if a (part of a) time period lies within two time limits.