

UNIVERSITEIT TWENTE

AFSTUDEERPROJECT

---

# Analyse van reizigersgedrag bij buitendienststellingen

---

*Auteur:*

Inge MASSELINK

*Faculteit:*

EWI

*Afdeling:*

Toegepaste Wiskunde

*Vakgroep:*

SOR

*Examencommissie:*

Dr. Judith TIMMER(UT)

Prof. Richard

BOUCHERIE(UT)

Dr.Ir Jasper

GOSELING(UT)

Dr. Georg STILL(UT)

Prof. Dr. Leo KROON(NS)

9 november 2012

# Inhoudsopgave

<b>1</b>	<b>Introductie</b>	<b>3</b>
<b>2</b>	<b>Probleemstelling</b>	<b>6</b>
2.1	Inleiding . . . . .	6
2.2	Buitendienststelling . . . . .	6
2.2.1	Vervangend vervoer . . . . .	7
2.2.2	Invloed op reistijd . . . . .	8
2.2.3	Verandering reizigersstromen . . . . .	10
2.3	OV-chipkaart gegevens . . . . .	10
2.4	NS-bussen . . . . .	10
2.5	Doel . . . . .	13
2.6	Afsluiting . . . . .	13
<b>3</b>	<b>Literatuuronderzoek</b>	<b>14</b>
3.1	Inleiding . . . . .	14
3.2	Smartcard . . . . .	14
3.3	Gedrag bij wegversperringen . . . . .	16
3.4	Modellen voor verkeersstromen . . . . .	16
3.5	TRANS . . . . .	18
3.6	Afsluiting . . . . .	18
<b>4</b>	<b>Methodologie</b>	<b>19</b>
4.1	Inleiding . . . . .	19
4.2	Dataset samenstellen . . . . .	20
4.3	Model aantal reizigers . . . . .	23
4.3.1	Verschilreeks . . . . .	25
4.3.2	Vinden passend model . . . . .	27
4.3.3	Aantal reizigers . . . . .	27
4.4	Model buitendienststellingen . . . . .	27

4.4.1	Logistische regressie model . . . . .	28
4.5	Gemiddelden per buitendienststelling . . . . .	29
4.6	Vuistregel NS . . . . .	30
4.6.1	Vuistregel en buitendienststellingsmodel . . . . .	30
4.6.2	Vuistregel en data . . . . .	30
4.7	Afsluiting . . . . .	31
<b>5</b>	<b>Resultaten</b>	<b>32</b>
5.1	Inleiding . . . . .	32
5.2	Model aantal reizigers . . . . .	32
5.2.1	Verschilreeks . . . . .	32
5.2.2	Aantallen reizigers . . . . .	38
5.2.3	Kwaliteitsverschil . . . . .	40
5.3	Model buitendienststellingen . . . . .	45
5.3.1	Kansen . . . . .	45
5.4	Voorspellen met gemiddelden per buitendienststelling . . . . .	46
5.5	Vuistregel . . . . .	47
5.5.1	Vuistregel en buitendienststellingsmodel . . . . .	47
5.5.2	Vuistregel en data . . . . .	48
5.6	Afsluiting . . . . .	48
<b>6</b>	<b>Conclusie en verder onderzoek</b>	<b>50</b>
6.1	Inleiding . . . . .	50
6.2	Conclusie . . . . .	50
6.2.1	Model aantal reizigers . . . . .	50
6.2.2	Model buitendienststellingen . . . . .	51
6.2.3	Model gemiddelden . . . . .	51
6.2.4	Vuistregel van de NS . . . . .	52
6.3	Verder onderzoek . . . . .	52

# Hoofdstuk 1

## Introductie

In deze scriptie is het afstudeerproject van Inge Masselink beschreven. Het onderzoek voor dit project is gedaan bij de Nederlandse Spoorwegen, ook wel de NS genoemd. De NS is de grootste gebruiker van het Nederlandse spoorwegnet, dat beheerd wordt door ProRail. Tegenwoordig rijdt de NS zo'n 5200 treinen per dag en brengt zo elke dag 1,1 miljoen mensen naar hun bestemming. 175 jaar geleden werd een van de rechtsvoorgangers van de NS opgericht, de Hollandsche IJzeren Spoorwegmaatschappij (HSM). Deze maatschappij werd opgericht door drie Amsterdammers. Pas in 1839 reed de eerste trein, dit was van Amsterdam naar Haarlem. Op de eerste spoorlijn in Nederland, die een lengte had van 16 kilometer.

Door de jaren heen zijn steeds meer kilometers spoor bijgebouwd, nu ligt er ongeveer 2800 kilometer spoor. Dit spoor vergt onderhoud en zijn elk weekend werkzaamheden nodig om de kwaliteit van het spoor hoog te houden. Deze werkzaamheden worden gepland door ProRail. Vroeger was het onderhouden van het spoor een taak van de NS, maar sinds 1 januari 1995 is besloten het bedrijf te splitsen in een gedeelte voor het materieel en de reizigers en een gedeelte voor de infrastructuur, wat dus later ProRail werd. Vanaf dat moment is de NS alleen verantwoordelijk voor het rijden van het materieel. Toch is er nog veel samenwerking nodig met ProRail, want het spoor is nodig voor het materieel.

De NS is onderverdeeld in vijf bedrijfsonderdelen: NS Reizigers, NS Hispeed, Abellio, NS Stations en NedTrain. Deze worden hieronder kort uitgelicht. NS Reizigers is verantwoordelijk voor alles wat met het rijden van treinen en serviceverlening naar de klant te maken heeft. Dus dienstregelingen maken, bij verstoringen oplossingen verzinnen, maar ook service naar de klant wordt geregeld door NS Reizigers. Bij het reizen met internationale en hogesnel-

heidsvervoer krijgen de reizigers te maken met NS Hispeed. NS Hispeed is de hogesnelheidsonderneming van de NS en biedt hogesnelheidsvervoer aan op internationale bestemmingen. Daarnaast zorgen ze voor service bij het boeken van internationale tickets, maar ook voor het verblijf op het station en in de trein, zoals de NS Hispeed Lounges op verschillende stations. De laatste tak die zich bezig houdt met reizigersvervoer is Abellio. Dit is een dochteronderneming van de NS en ze verzorgen bus- en railvervoer in Groot-Brittannië, Duitsland en Tsjechië.

Naast het reizigersvervoer zijn er nog twee takken bij de NS, namelijk NS stations en NedTrain. NS stations is exploitant van ruim 390 spoorwegstations in Nederland, ze bieden diensten en faciliteiten aan op het station en zorgen ze voor sociale veiligheid. Ook het beheer en schoonhouden van de stations is een kernactiviteit van NS Stations. Als laatste is er nog NedTrain, dit is een onderhouds- en revisiebedrijf voor spoorwegmaterieel.

De afstudeerstage is gedaan bij het eerstgenoemde bedrijfs onderdeel, namelijk NS Reizigers en dan bij de afdeling Proceskwaliteit en Innovatie (PI). Bij PI worden onderzoeken gedaan om verschillende processen bij de NS te optimaliseren/automatiseren met behulp van wiskundige modellen, zoals het maken van materieelplanning, personeelsplanning en de dienstregeling. Veel projecten worden uitgevoerd door promovendi en afstudeerders, waarvan er altijd wat op de afdeling aanwezig zijn. Sinds de invoering van de OV-chipkaart worden bij de NS ook onderzoeken gedaan om zoveel mogelijk kennis te krijgen met behulp van de OV-chipkaartgegevens.

De OV-chipkaart is een elektronische betaal- en toegangspas voor het openbaar vervoer per bus, tram en trein. Deze kaart is gebaseerd op de Hongkongse Octopus card en ontwikkeld door het bedrijf East-West, dat hiermee begon in 2003. Het bedrijf dat hoofdverantwoordelijk is voor de OV-chipkaart is Trans Link Systems.

De OV-chipkaart is pas een geldig vervoersbewijs als er ingecheckt is. Op het eind van de reis moet er weer worden uitgecheckt. Op deze in- en uitcheckmomenten worden de gegevens van de chipkaarthouder opgeslagen. Dit zijn gegevens als de locatie en het tijdstip van in- en uitchecken, abonnement van de reiziger en prijs van de reis. Met deze gegevens is het mogelijk meer inzicht te krijgen in het gedrag van reizigers.

Met dit onderzoek wordt gekeken of het mogelijk is om te voorspellen wanneer reizigers blijven reizen tijdens buitendienststellingen.

Deze scriptie is opgedeeld in verschillende de hoofdstukken die samen het onderzoek beschrijven. Hieronder een kort overzicht van de hoofdstukken. In hoofdstuk 2 wordt de probleembeschrijving gegeven. Hierin wordt meer

duidelijkheid gegeven over wat er gebeurt bij werkzaamheden en de data die gebruikt wordt. Ook wordt hier het doel van het onderzoek vermeld. Het hoofdstuk dat erop volgt, hoofdstuk 3, beschrijft onderzoeken die op een paar punten te vergelijken zijn met dit onderzoek, zo is er duidelijk wat er al gedaan is aan onderzoek in de afgelopen tijd. In hoofdstuk 4 wordt beschreven hoe er verschillende regressie modellen gemaakt zijn voor dit onderzoek. In hoofdstuk 5 staan de parameters die gevonden zijn bij de verklarende variabelen van de modellen uit hoofdstuk 4. In het laatste hoofdstuk, hoofdstuk 6 staan de conclusie en de adviezen voor verder onderzoek.

## Hoofdstuk 2

# Probleemstelling

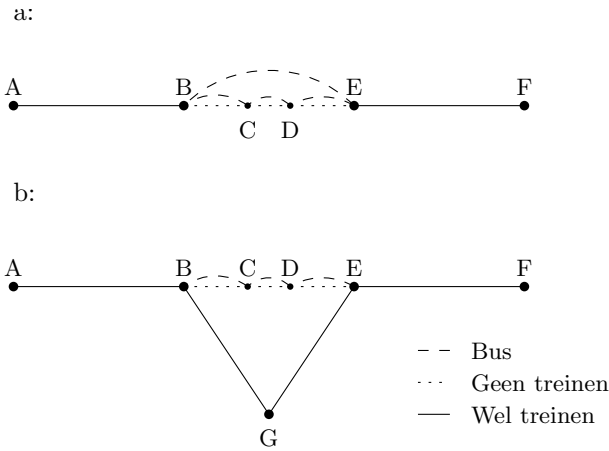
### 2.1 Inleiding

In dit hoofdstuk staat het probleem beschreven dat geleid heeft tot dit onderzoek. Het doel van dit onderzoek is namelijk meer inzicht te krijgen in reizigersstromen bij buitendienststellingen. De term buitendienststelling wordt beschreven in paragraaf 2.2. Ook wordt uitgelegd wat er verandert als er buitendienststellingen zijn. De reizigersstromen zijn verkregen uit de OV-chipkaartdata. Over de data wordt meer verteld in paragraaf 2.3. Bij buitendienststellingen wordt er vaak uitgecheckt voordat reizigers gebruik maken van NS-bussen, in paragraaf 2.4 wordt uitgelegd waarom dit een probleem geeft. De laatste paragraaf, paragraaf 2.5 beschrijft het doel van het onderzoek.

### 2.2 Buitendienststelling

In opdracht van ProRail wordt regelmatig gewerkt aan het spoor. De werkzaamheden zijn vaak gepland in nachten, weekenden of vakanties. Op die momenten zijn er minder reizigers die hinder ondervinden van de werkzaamheden. Wanneer gewerkt wordt aan een stuk spoor, dan rijden daar minder of geen treinen. Het laten vervallen van de treinen heet ook wel het buitendienststellen van die treinen. In dit onderzoek worden de termen buitendienststelling en werkzaamheden door elkaar gebruikt.

Om tijdens werkzaamheden de reizigers toch op plaats van bestemming te krijgen, wordt gebruik gemaakt van vervangend vervoer, hierover meer uitleg in paragraaf 2.2.1. Het reizen met vervangend vervoer kost extra tijd, in paragraaf 2.2.2 wordt hierover gepraat. Ook is daar beschreven dat de



Figuur 2.1: Voorbeelden vervangend vervoer

dienstregeling zo bewerkt wordt dat de reizigers buiten het werkzaamheden-traject er geen hinder van ondervinden. In paragraaf 2.2.3 wordt gesproken over reizigersstromen.

### 2.2.1 Vervangend vervoer

Tijdens werkzaamheden kunnen minder of geen treinen rijden. Als er minder treinen rijden is er vaak geen vervangend vervoer, er wordt dan geadviseerd op de reisplanner te letten en met treinen te gaan die wel rijden. Wanneer er geen treinen rijden worden bussen en/of omgeleide intercity's ingezet, zo kunnen treinreizigers toch op plek van bestemming komen.

Als voorbeeld staat in figuur 2.1 een stuk spoor getekend tussen de stations A en F, met intercitystations A, B, E en F en stopstations C en D. Stopstations zijn stations waar alleen stoptreinen stoppen. In dit voorbeeld zijn er werkzaamheden tussen B en E. Als alternatief vervoer is in figuur 2.1a alleen de mogelijkheid om bussen in te zetten. Dit betekent dat er stop- en snelbussen worden ingezet tussen B en E. Stopbussen stoppen op alle stations en snelbussen alleen op intercitystations. In het voorbeeld in figuur 2.1b is het mogelijk om een omreisroute te maken met de trein. Op een normale dag zou de intercity reizen van A via B en E naar F. Nu met de buitendienststelling wordt er gereden van A via B, G en E naar F. Voor de tussenstations C en D worden bussen ingezet.

Met behulp van deze maatregelen kunnen reizigers toch blijven reizen ondanks de werkzaamheden.



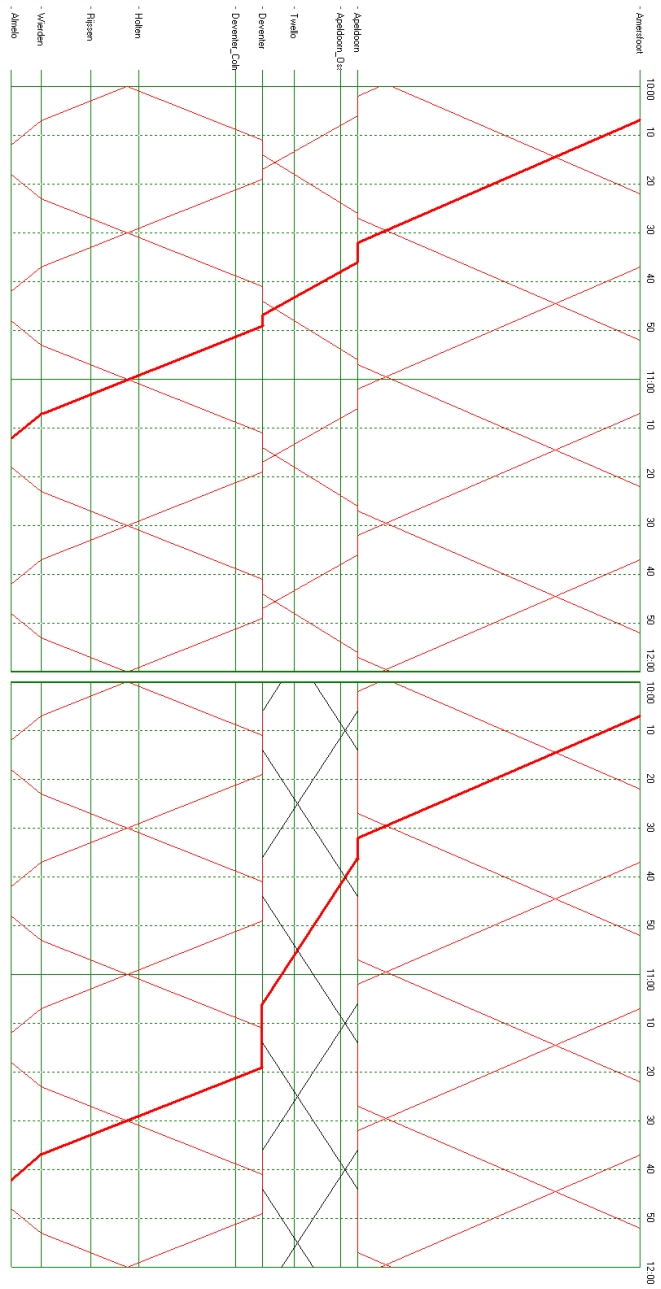
### 2.2.2 Invloed op reistijd

Het vervangend vervoer geeft de mogelijkheid dat er wel gereisd kan worden, maar kan wel zorgen voor extra reistijd voor reizigers die over het werkzaamhedentraject reizen. Reizigers die niet over het werkzaamhedentraject hoeven te reizen hebben nagenoeg geen last van de buitendienststellingen. Hoe deze hinder beperkt wordt terwijl treinen wel een ander reisschema hebben, staat in deze paragraaf uitgelegd.

In figuur 2.2 is een voorbeeld van een tijd-weg-diagram weergegeven van de dienstregeling tussen Amersfoort en Almelo, met (rechts) en zonder (links) buitendienststelling. Hier wordt de intercitydienst tussen Amersfoort en Almelo weergegeven. Dik gemarkeerd staat de intercity van 10.07 uit Amersfoort die om 11.12 arriveert in Almelo. Aan de horizontale gedeeltes is te zien dat gestopt wordt in Apeldoorn en Deventer.

Stel er wordt tussen Apeldoorn en Deventer aan het spoor gewerkt, de treinen kunnen daar dus niet rijden. In het rechter plaatje is de veranderde dienstregeling weergegeven. Er is te zien dat buiten het traject Apeldoorn-Deventer geen verandering is. Op het werkzaamheden traject zijn zwarte lijnen weergegeven die minder stijl lopen dan de lijnen in de andere figuur tussen Apeldoorn en Deventer. Deze zwarte lijnen stellen bussen voor. Bussen rijden langzamer dan treinen vandaar dat ze minder stijl zijn. Vanzelfsprekend rijden er meer bussen dan dat er zwarte lijnen weergegeven zijn. Voor mensen die over het traject gaan is er hinder. Een busreis duurt langer dan een treinreis, bijvoorbeeld van Apeldoorn naar Deventer is met de bus 30 minuten, en met de trein 10 minuten. Gewoonlijk wordt er om 10.36 vertrokken uit Apeldoorn en de aankomsttijd in Deventer is dan 10.47, zie linker plaatje. In het rechter plaatje vertrekt de bus ook om 10.36, maar de reis duurt een half uur, dus de intercity van 10.49 in Deventer wordt gemist. Er zal dus met de trein van 11.19 gegaan moeten worden. Wat een half uur extra reistijd oplevert. Bij sommige buitendienststellingen worden omgeleide intercity's ingezet, deze gaan via een ander station en doen dus langer over een traject. Dit levert extra reistijd op. De zwarte lijnen tussen Apeldoorn en Deventer in figuur 2.2 geven ook de dienstregeling weer van omgeleide intercity's.

Mensen die niet over het traject gaan ondervinden geen hinder en er wordt aangenomen dat die blijven reizen zoals ze altijd doen. In dit onderzoek wordt gekeken naar reizigers die over het traject reizen en dus daadwerkelijk hinder ondervinden.



Figuur 2.2: Voorbeeld verandering van dienstregeling

### 2.2.3 Verandering reizigersstromen

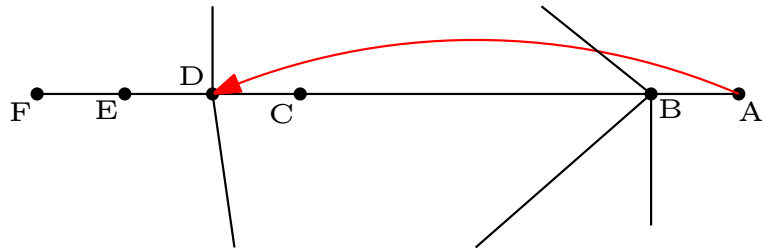
Als de treinen normaal rijden zijn veel vaste patronen te vinden in de aantallen reizigers die een bepaalde reis maken. Dit zijn de reizigersstromen die elke dag te vinden zijn. Tijdens een buitendienststelling zal er een verandering optreden in de reizigerstromen. Niet iedereen gaat reizen en dan zullen de stromen dus kleiner worden. Wat er precies gebeurt met het aantal reizigers als er een buitendienststelling is, is nog niet bekend. Toch wordt er in het materieel al enige tijd rekening gehouden met de werkzaamheden. Dit wordt gedaan door het aantal reizigers te schatten met een vuistregel. Deze vuistregel gaat er vanuit dat er 30% minder mensen reist tijdens buitendienststellingen. Nu er gebruik gemaakt kan worden van de OV-chipkaart gegevens is het interessant om te kijken of de vuistregel klopt.

## 2.3 OV-chipkaart gegevens

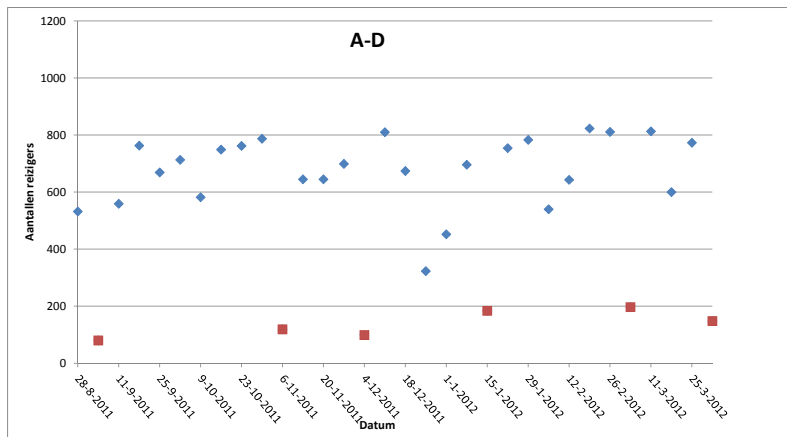
Sinds de invoering van de OV-chipkaart is bekend waar OV-chipkaart gebruikers in- en uitchecken en op welk tijdstip. Met behulp van deze gegevens kan nagegaan worden hoeveel mensen een bepaalde reis maken. De reis wordt ook wel OD (origin-destination) genoemd. Er kan bijvoorbeeld gevonden worden wat het aantal OV-chipkaartgebruikers is die op een dag een reis van Amersfoort naar Almelo maken. Naast het aantal reizigers is door middel van de OV chipkaartgegevens ook bekend wat de abonnementen zijn van de reizigers, zo kunnen bijvoorbeeld studenten en mensen zonder abonnement makkelijk gescheiden worden. Ook kan per reizigerstype onderzocht worden wat het effect van buitendienststellingen is.

## 2.4 NS-bussen

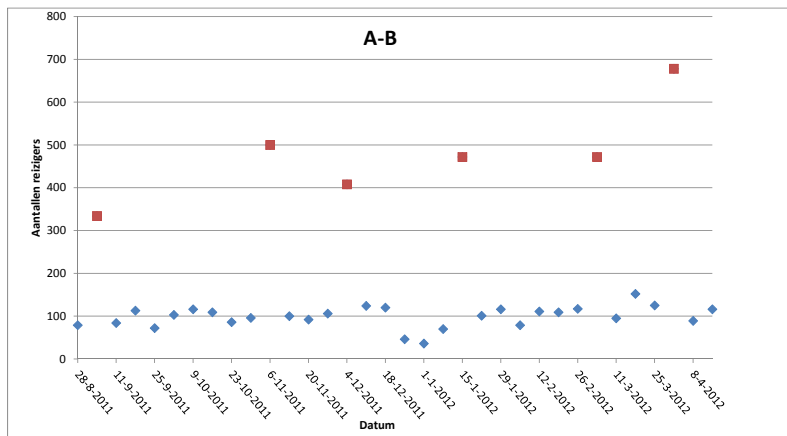
Tijdens buitendienststellingen worden vaak mensen met bussen vervoerd. Dikwijls gebeurt het dat mensen uitchecken als ze het station uitlopen naar de bus, terwijl ze dus nog een station verder moet. Als voorbeeld het traject in figuur 2.3. Hier op liggen de intercitystations  $A, B$  en  $D$ . In figuur 2.4 zijn de aantallen reizigers die de reis  $A$  naar  $D$  maken weergegeven. De rode vierkanten zijn ook aantallen reizigers maar dit zijn dagen waarop er een buitendienststelling is tussen  $B$  en  $D$  en daar gereisd moet worden met de bus. Er is dus duidelijk een verschil in aantallen reizigers te zien. Dan is er figuur 2.5 waarin het aantal reizigers die de reis van  $A$  naar  $B$  maken wordt weergegeven. Gewoonlijk liggen die zo rond de 100, maar wanneer er



Figuur 2.3: Traject



Figuur 2.4: Aantal reizigers tussen A en D



Figuur 2.5: Aantal reizigers tussen A en B

buitendienststelling is tussen  $B$  en  $D$  worden dit er tussen de 300 en 700. Dit wordt hoogstwaarschijnlijk veroorzaakt door onder andere mensen die van  $A$  naar  $D$  of  $C$  willen reizen, maar uitchecken op  $B$  en vandaar verder gaan met de NS-bus. In dit onderzoek is geen rekening gehouden met deze mensen die uitchecken.

## 2.5 Doel

Het doel van dit onderzoek is om de invloed van buitendienststellingen op reizigersstromen te analyseren. Dit wordt gedaan door te bepalen waar de kans om te blijven reizen vanaf hangt. Dus wat de verklarende variabelen zijn waardoor mensen ervoor kiezen om wel of niet te gaan reizen. Voorbeelden van variabelen als de dag waarop de buitendienststelling is, de lengte van de reis van de reizigers en het vervangende vervoer dat ter beschikking gesteld is. Dan is nog de vraag of de kans beter te voorspellen is per type reiziger of voor alle soorten reizigers samen. In dit onderzoek wordt dus gekeken of het percentage reizigers dat blijft reizen te voorspellen is per producttype.

Om de kans te berekenen dat mensen blijven reizen moet ook geschat worden wat het aantal reizigers per dag is, zonder rekening te houden met de buitendienststellingen. Bij het schatten van het aantal reizigers kunnen verschillende variabelen bekeken worden. Hierbij kan gedacht worden aan vergelijkingen met het aantal reizigers op andere dagen, maar ook het weer kan hier invloed op hebben. Bij het schatten van de aantallen kan ook gekeken worden of reizigers beter per type of allemaal samen geschat kunnen worden.

## 2.6 Afsluiting

In dit hoofdstuk is een beeld gegeven van buitendienststellingen en data die gebruikt kan worden. In het volgende hoofdstuk is beschreven welke literatuur gevonden is, dat past bij dit onderzoek.

## Hoofdstuk 3

# Literatuuronderzoek

### 3.1 Inleiding

Dit hoofdstuk beschrijft verschillende artikelen die gebruikt zijn in het literatuuronderzoek dat is gedaan voor deze scriptie. Tijdens het zoeken naar literatuur is maar één onderzoek gevonden dat grote overeenkomsten vertoont met dit scriptieonderzoek, namelijk dat van Van der Hurk. Dit is niet genoeg voor een heel literatuur onderzoek, vandaar dat gezocht is op verschillende onderwerpen die te maken hebben met het onderzoek. Het hoofdstuk is verdeeld in verschillende secties die hieronder kort beschreven worden. Paragraaf 3.2 beschrijft artikelen over het gebruik van smartcard(OV-chipkaart) gegevens. Artikelen over gedrag van weggebruikers bij werkzaamheden staan beschreven in paragraaf 3.3. Paragraaf 3.4 beschrijft verschillende modellen die gebruikt kunnen worden om de routekeuze van reizigers te modelleren. Tenslotte wordt in de laatste paragraaf 3.5, het reizigertoedelingsmodel TRANS beschreven.

### 3.2 Smartcard

In een artikel van Cheung [1] staat dat het sinds januari 2009 mogelijk is om in Nederland te reizen met een smartcard als geldig vervoersbewijs. Een voordeel van het gebruik van de smartcard is dat het reisgedrag gemakkelijker geanalyseerd kan worden. Bagchi et al. [2] beschrijft dit in zijn onderzoek. Door de extra kennis van het reisgedrag kan het openbaar vervoer beter aangepast worden aan de reizigers.

Niet alleen in Nederland wordt gebruik gemaakt van de smartcard in het openbaar vervoer. In een artikel van Paddington [3] staat een lijst met een

aantal steden die een smartcard gebruiken en mee willen werken aan zijn onderzoek naar smartcardgebruik. Deze steden liggen in de landen Groot-Britannië, Frankrijk, Duitsland, Italië, China, Japan, Brazilië, US, Canada, Australië, Baskenland, Singapore en Noorwegen. De smartcard wordt dus veel gebruikt in andere landen, wat resulteert in verschillende onderzoeken. Sinds 2005 worden er met regelmaat nieuwe onderzoeken gepubliceerd met analyses van smartcardgegevens. In een artikel van Pelletier [4] staat een lijst met voorbeelden van onderzoeken die gedaan zijn. Deze onderzoeken gaan over de voor- en nadelen van de smartcard, het gebruik in de strategische planning van openbaar vervoer en experimenten naar tarieven.

Daarnaast is onderzoek gedaan naar het gedrag van reizigers, zoals in het onderzoek Agard [5] waar de reizigers in vier groepen zijn verdeeld op basis van reispatronen. Van deze groepen is de samenstelling ( volwassenen, studenten, enz.) en het tijdstip van reizen bekeken. In een onderzoek van Bagchi en White [6] wordt gekeken naar het aantal oudere mensen in de bus, het aantal mensen dat moet overstappen, en de verandering van reizigers die een smartcard gebruiken. Bij het onderzoek van Morency et al [7] zijn de reizigers onderverdeeld in groepen gebaseerd op het soort abonnement dat ze hebben. Per groep wordt bekeken van hoeveel verschillende busstops ze gebruikmaken en wat de reistijden zijn overdag.

In een recent onderzoek van Bouman et al. [8] wordt gekeken naar het herkennen van patronen van reizigers in de OV-chipkaart data. Een voorbeeld van een patroon is de groep reizigers die van huis naar werk reist. Met deze patronen kan er een agent-based-microsimulation gemaakt worden.

Een onderzoek dat dichtbij dit scriptieonderzoek is van Van der Hurk et al. [9]. Hierin wordt bekeken hoe OV-chipkaart gegevens gebruikt kunnen worden voor Disruption Management. Disruption Management is het managen van materieel bij verstoringen. In dit onderzoek wordt beschreven hoe OD's (Origin-Destinations) geclusterd kunnen worden. Er wordt ook beschreven dat er onderzoek gedaan kan worden naar de invloed van verschillende producttypes.

Er is dus veel onderzoek gedaan naar het gedrag van smartcard gebruikers, maar geen van deze onderzoeken kijkt ook naar werkzaamheden. Namelijk in dit onderzoek wordt getracht de werkzaamheden in het openbaar vervoer, die van invloed zijn op het reisgedrag van de reiziger mee te nemen.



### 3.3 Gedrag bij wegversperringen

Onderzoeken die gevonden zijn over het gedrag van reizigers bij werkzaamheden gaan alleen over het gedrag van weggebruikers. In een onderzoek van Clegg [10] wordt een verkeerssituatie bekeken waarbij één rijrichting is afgesloten. Om het nieuwe reisgedrag van automobilisten te onderzoeken zijn er op verschillende plekken personen neergezet, die de nummerborden van voorbijkomende auto's noteren. Met deze gegevens is onderzocht hoe de reizigersstromen veranderen. Hierbij is als variabele de duur van de werkzaamheden meegenomen.

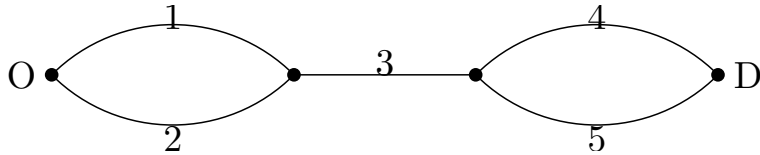
Naast werkzaamheden zijn verschillende onderzoeken gedaan naar het reisgedrag bij het instorten van de I-35W brug in Mississippi op 1 augustus 2007. Deze brug was een belangrijke verkeersader voor veel reizigers. Het is de route van 140.000 dagelijkse trips. Eén van die onderzoeken is gedaan door Zhu [11]. Om de data te verkrijgen heeft Zhu een enquête gehouden onder mensen die daar uit de buurt komen. Er is uitgezocht wat de nieuw gekozen routes zijn en of mensen van vervoersmiddel veranderd zijn.

In 2008 is er een groot onderhoud geweest aan Interstate-5 (I-5) in California. Met tussenpozen is een mijl van de I-5, negen weken dicht geweest. Deze werkzaamheden werden ook wel 'the Fix I-5 project' genoemd. Een paar weken van tevoren is besloten onderzoek te doen naar de verandering in reisgedrag van de reizigers. Via bedrijven zijn er mensen per mail uitgenodigd om mee te werken aan een onderzoek. Uit dit onderzoek kwam dat een groot percentage van de reizigers de spits ging vermijden of van route veranderde. Daarnaast is er een Binary Logit Model gemaakt, om de keuze voor thuiswerken en openbaar vervoer te modelleren.

### 3.4 Modellen voor verkeersstromen

Voor het modelleren van verkeersstromen zijn verschillende methoden te vinden in de literatuur. Een voorbeeld is het Flowmodel dat beschreven wordt in een artikel van He [12]. Het model is gemaakt bij het netwerk dat weergegeven is in figuur 3.1. In dit netwerk lopen vier verschillende paden die O en D verbinden door middel van links die een connectie met elkaar hebben. De verschillende paden zijn:

- **pad 1** via  $1 \rightarrow 3 \rightarrow 4$
- **pad 2** via  $1 \rightarrow 3 \rightarrow 5$
- **pad 3** via  $2 \rightarrow 3 \rightarrow 4$



Figuur 3.1: voorbeeld netwerk

- **pad 4** via  $2 \rightarrow 3 \rightarrow 5$

Alle links hebben dezelfde kosten (reistijd) dus het verkeer is gelijk verdeeld over het netwerk. In het netwerk kan de prijs van een link verhoogd worden. Het Flowmodel voorspelt dan opnieuw het aantal reizigers over de verschillende paden. Dit kan vergeleken worden met een verstoorde situatie. In een verstoorde situatie is de verstoorde link ook minder aantrekkelijk vanwege extra reistijd.

In een artikel van Chiang [13] worden verschillende manieren beschreven om verkeersstromen te modelleren. Een van die manieren is om verkeersstromen te modelleren als neurale netwerken. Wegen worden dan vergeleken met de zenuwbanen waar neuronen doorheen moeten. Een beknelde zenuwbaan kan dan vergeleken worden met een versperde weg. Daarnaast wordt in dit artikel gekeken naar methoden als regressie analyse en tijdreeks analyse.

Chorus [14] beschrijft in zijn onderzoek hoe het Regretmodel gebruikt kan worden als verkeersmodel. Dit is een model dat keuzes maakt op basis van de hoeveelheid spijt die er gekregen wordt na het maken van een keuze. Dit model wordt vaak gebruikt bij de analyse van de keuzes te modelleren om wel of niet mee te doen aan een loterij.

Een bekend statistisch model is het logistische regressiemodel beschreven in een artikel van Moore [15]. Dit wordt als model gebruikt waar de te verklaren variabele een kans voorstelt. Vaak wordt het gebruikt in het verkeer om route keuzes te modelleren. In een artikel van Cascetta [16] wordt een variant op de logistische regressie gebruikt, namelijk het C-logit model. Dit wordt gebruikt om de route keuze beter te modelleren dan met Multinomial Logit (MNL) model. Bij MNL worden kansen aan paden met meerdere links gehangen worden, die volgens de schrijvers van dit artikel ongeloofwaardig zijn.

### 3.5 TRANS

Bij de NS is het TRANS-model ontwikkeld. Dit model wijst reizigers toe aan verschillende routemogelijkheden tussen vertrek en aankomst. Dit model werd voornamelijk gebruikt voor het bestaan van de OV-chipkaart om te bekijken welke routes mensen gaan nemen.

Bijvoorbeeld bij de analyse van de reizigersstromen van de buitendienststelling van het station en delen van het spoor bij Nijverdal. Deze buitendienststelling duurt 3 jaar en de NS wilde weten hoe mensen gingen reizen naar stations in die omgeving.

De keuze van de route wordt gemaakt door te kijken naar het nut van een route. Als het nut groter is wordt de route door meer mensen gekozen. Het nut is afhankelijk van de variabelen reistijd, aantal overstappen, overstaptijden, vertrek- en aankomsttijden en toeslagen. Het TRANS-model is gebaseerd op een logit model dat de kans op een bepaalde keuze voor een route geeft afhankelijk van verschillende variabelen. In een intern rapport van Dam [17] wordt hier meer over geschreven.

### 3.6 Afsluiting

Met dit literatuuronderzoek is er bekend wat er in het verleden al onderzocht is over dit onderwerp. Uit dit literatuuronderzoek is gebleken dat er niet veel vergelijkbaars is gedaan en het onderzoek dus nieuwe inzichten geeft. In het volgende hoofdstuk worden de modellen die gemaakt zijn in dit onderzoek beschreven.

## Hoofdstuk 4

# Methodologie

### 4.1 Inleiding

Zoals vermeld in de probleemstelling, hoofdstuk 2, is van te voren niet precies bekend wat er verandert aan het aantal reizigers bij werkzaamheden. In dit hoofdstuk wordt beschreven welke modellen gemaakt zijn om de kans te voorspellen op het blijven reizen van reizigers bij buitendienststellingen, met behulp van de OV-chipkaartgegevens.

In paragraaf 4.2 wordt beschreven welke informatie uit de OV-chipkaart-dataset gebruikt wordt voor dit onderzoek en hoe de dataset bewerkt wordt tot verschillende datasets.

Deze verschillende datasets zijn nodig om zoveel mogelijk inzicht te kunnen geven in de keuze die de reiziger maakt bij buitendienststellingen. Om de keuze om te blijven reizen te voorspellen, is het nodig te weten hoeveel reizigers er gereisd hebben op een buitendienststellingsdag en hoeveel reizigers er zouden reizen op dezelfde dag als er geen buitendienststelling geweest zou zijn. De mensen die daadwerkelijk gereisd hebben, kunnen gehaald worden uit de OV-chipkaart gegevens. Het aantal reizigers dat gereisd zou hebben als er geen buitendienststelling zou zijn, is niet bekend. In paragraaf 4.3 wordt beschreven hoe die aantallen reizigers berekend worden met behulp van regressie modellen. Met deze verwachte aantallen reizigers wordt de kans berekend dat mensen blijven reizen. Dat wordt uitgelegd in paragraaf 4.4. Wanneer deze kansen zijn uitgerekend is het nodig om een model te maken dat in de praktijk gebruikt kan worden. Dit wordt gedaan door met de berekende kansen een regressiemodel te schrijven waarmee geschat wordt dat reizigers blijven reizen. Het model dat eruit komt is voor alle buitendienststellingen samen. In paragraaf 4.5 wordt beschreven dat er per

buitendienststelling uitgerekend kan worden wat de gemiddelde kans is dat mensen blijven reizen. Dit kan ook als model gebruikt worden. In paragraaf 4.6 wordt beschreven hoe de correctheid van de huidige vuistregel van de NS getest kan worden.

## 4.2 Dataset samenstellen

De OV-chipkaart dataset bevat informatie van het in- en uitchecken van reizigers met een OV-chipkaart sinds 2009. Bij het in- en uitchecken worden verschillende gegevens opgeslagen, o.a. het tijdstip en station van in- en uitchecken en het abonnement van de gebruiker. Uit de data is dan af te leiden welke reizen gemaakt zijn, op welke tijdstippen en met welke abonnementen. Het wil weleens voorkomen dat mensen alleen in- of uitchecken. In de data komen dan losse in- en uitchecks te staan. Deze zijn voor dit onderzoek verwijderd uit de dataset, omdat in dit geval niet bekend is waar de persoon gereisd heeft en die informatie is essentieel voor dit onderzoek. In dit onderzoek worden alleen de volledige reizen gebruikt vanaf vrijdag 6 januari 2012 tot en met 1 juli 2012. De datum 6 januari 2012 is gekozen vanwege het verplicht in- en uitchecken voor studenten vanaf 1 januari en 6 januari is de eerste vrijdag in de maand. Het verplichte in- en uitchecken veroorzaakte een grote stijging in het aantal gebruikers en gegevens. 1 juli is de einddatum vanwege de zomerluwte van de reizigers aantallen, en omdat 30 juni op een zaterdag valt is zondag 1 juli er nog bij genomen.

Uit deze verkleinde dataset, met volledige reizen in 2012, wordt per buitendienststelling een groep reizen gehaald. Deze deelverzameling per buitendienststelling wordt geselecteerd door alleen OD's te nemen die hinder kunnen ondervinden van de werkzaamheden. Alleen deze reisgegevens zijn interessant, want, bijvoorbeeld een buitendienststelling in Drenthe heeft geen invloed op reizigers in Zeeland, maar alleen op reizigers die over het desbetreffende werkzaamhedentraject reizen. De verschillende buitendienststellingen die bestudeerd worden, zijn

- Almere-Weesp
- Amersfoort-Utrecht
- Amsterdam Sloterdijk-Haarlem
- Den Haag Centraal-Gouda
- Hoofddorp-Leiden

- Leiden-Alpen aan de Rijn
- Sittard-Maastricht
- Zaandam-Amsterdam Sloterdijk

Deze buitendienststellingen zijn gekozen, omdat ze meerdere malen voorkomen in de onderzochte periode. Daarnaast vallen ze in het weekend en zijn ze overdag. Na deze groepering ontstaan er acht verschillende datasets, één passend bij elke buitendienststelling.

Deze datasets worden gesplitst in drie groepen afhankelijk van de lengte van de reis op dagen zonder buitendienststelling. Deze groepen zijn:

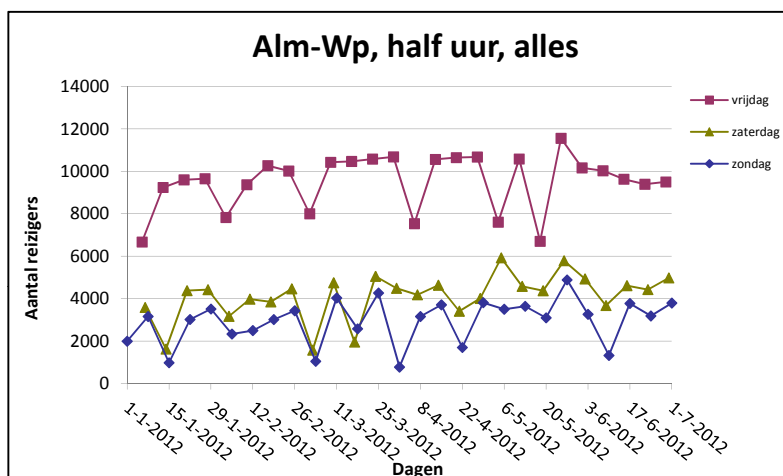
- reizen korter dan een halfuur
- reizen tussen een halfuur en een uur
- reizen langer dan een uur.

Zo ontstaan er  $8 \cdot 3 = 24$  datasets.

Naast de lengte van de reis kan het zijn dat reizigersaantallen beter te voorspellen zijn per soort reiziger, dus per producttype. Het reisgedrag, de keuze dat er gereisd blijft worden, van studenten kan namelijk anders zijn dan dat van reizigers met een dalabonnement. Om dit te onderzoeken is besloten aparte datasets te maken, om te kijken of hier verschillende modellen voor nodig zijn. De producttypes zijn opgedeeld in de volgende groepen:

- reguliere reizen, dit zijn mensen zonder abonnement of speciale kortingen.
- dalvrij of dalkorting, alle reizigers die een abonnement hebben dat in de daluren telt
- studenten
- alle reizigers

De 24 datasets die hierboven gemaakt zijn worden allemaal opgedeeld per productgroep, zo ontstaan  $8 \cdot 3 \cdot 4 = 96$  verschillende datasets die meer inzicht kunnen geven in het gedrag van reizigers bij buitendienststellingen. Elke dataset bevat het totaal aantal reizigers dat per dag gereisd heeft over de groep OD's die hinder ondervinden van de buitendienststellingen, verdeeld per lengte van de reis en de reizigers die voldoen aan de producttype-eis. In figuur 4.1 is de grafiek te zien die dan gemaakt kan worden. In deze

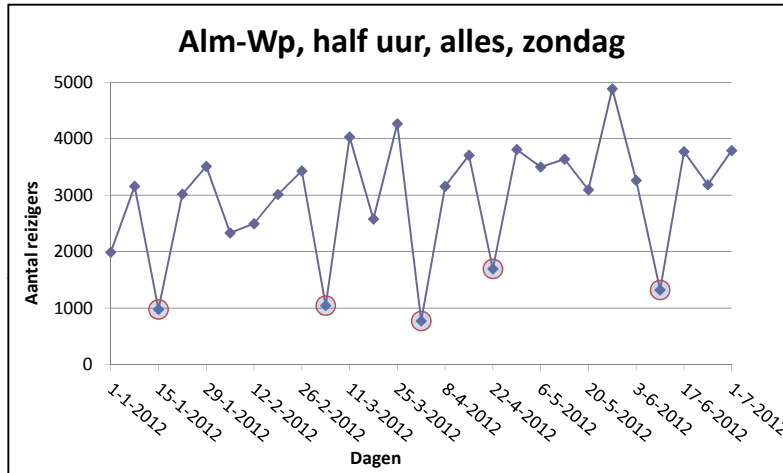


Figuur 4.1: Aantallen reizigers

grafiek is het aantal reizigers te zien voor de OD's die hinder hebben van de buitendienststelling Almere-Weesp, van reizen die korter zijn dan een half uur. Op de x-as de verschillende dagen en op de y-as het aantal reizigers. Met deze verdeling ontstaat er per buitendienststelling een dataset met het aantal reizigers per dag die over het traject gaan.

Op deze datasets wordt lineaire regressie uitgevoerd. Dus bij de aantallen reizigers wordt het best passende model gezocht.

De datasets bevatten het echte aantal reizigers dat gereisd heeft op een dag. Dus ook het echte aantal reizigers dat gereisd heeft op een dag waarop een buitendienststelling is. De modellen die gemaakt worden zijn om de aantallen te voorspellen als er geen buitendienststelling zou zijn geweest, dus moet de data die gebruikt wordt ook niet de buitendienststellingen bevatten. Dit zou namelijk invloed hebben op de voorspelkracht van de modellen voor de andere dagen. Daarom worden deze datasets nog bewerkt voordat de aantallen reizigers geschat kunnen worden. Het is besloten om het aantal reizigers op een dag met buitendienststelling aan te passen. De aangepaste



Figuur 4.2: Aantallen reizigers op zondagen

waarde is het gemiddelde van de aantallen op de dag een week ervoor en erna. Dus de aangepaste waarde van bijvoorbeeld zaterdag 12 mei wordt het gemiddelde van zaterdag 5 mei en zaterdag 19 mei. Er zijn verschillende manieren om de dataset aan te passen zodat het gebruikt kan worden om regressie op toe te passen. In dit onderzoek is gekozen voor gemiddelden omdat de aantallen reizigers vaak redelijk stabiel zijn. Voor het oplossen van dit probleem kunnen ook complexere methodes gebruiken, maar er zijn geen sterke argumenten om niet de gemiddelden te gebruiken.

### 4.3 Model aantal reizigers

De aantallen reizigers moeten worden geschat, omdat het nodig is te weten wat het aantal reizigers geweest zou zijn op een buitendienststellingsdag als er geen buitendienststelling geweest zou zijn. Als voorbeeld figuur 4.2 met de aantallen reizigers op zondag. Op de horizontale as staan de verschillende zondagen en op de verticale as het totaal aantal reizigers dat gereisd heeft op



die zondag over een bepaalde groep OD's. De vijf gemarkeerde punten zijn de zondagen met een buitendienststelling op het traject. Het aantal reizigers dat eigenlijk gereisd zou hebben, als er geen buitendienststelling zou zijn, wordt verwacht hoger te liggen dan wanneer er wel buitendienststellingen zijn. Het punt waar het aantal waarschijnlijk gelegen zou hebben is nodig om de kansen te berekenen dat mensen blijven reizen.

De enige manier om dat punt te benaderen is door te schatten wat het aantal reizigers zou zijn geweest. Om dit te schatten wordt een regressie model gemaakt dat het aantal reizigers op alle weekenddagen gaat schatten. Zo kan ter controle het voorspelde aantal reizigers vergeleken worden met het echte aantal reizigers, op niet buitendienststellingsdagen.

Het aantal reizigers is verzameld in de loop van een bepaalde periode en is afhankelijk van de dagen, daarom is besloten om tijdreeks modellen te gaan gebruiken. Daarnaast is er geen trend te vinden in de data, dus mag aangenomen worden dat het stationaire tijdreeksen zijn. In het boek van Heij et al [18] staat een iteratieve methode om stationaire tijdreeksen te modelleren. Deze methode is gebruikt om per dataset een model te vinden dat het aantal reizigers schat. Hieronder staan de stappen kort beschreven:

- **Stap 1: Grafieken van de data**

In de eerste stap wordt de data inzichtelijk gemaakt. Met grafieken komt er een duidelijk beeld van de gegevens. Zo kan bekeken worden hoe het reizigersverloop is en of er opvallende dagen zijn in de tijdsperiode die gekozen is.

- **Stap 2: Keuze van de Lag structuur**

Door de autocorrelatie tussen de dagen uit te rekenen kan er gekeken worden of de reizigersaantallen van verschillende dagen met elkaar gecorreleerd zijn.

- **Stap 3: Schatten van de parameters**

Tijdens de derde stap worden de parameters van het model geschat.

- **Stap 4: Diagnostische testen**

In deze stap wordt gekeken hoe goed het model past bij de data.

- **Stap 5: Verbeter het model**

Wanneer uit de vorige stap blijkt dat het model niet goed genoeg is dan moeten er verbeteringen komen. Dit kan gedaan worden door andere verklarende variabelen te kiezen.

- **Stap 6: Gebruik het model**

Wanneer het model goed genoeg voorspelt dan kan het gebruikt worden.

De termen tijdreeksmodellen en regressiemodellen worden door elkaar gebruikt in het verslag.

### 4.3.1 Verschilreeks

Er is besloten om de verschilreeks in plaats van de aantallen te gaan schatten. Met de verschilreeks wordt het verschil in het aantal reizigers tussen vorige week en deze week bedoeld. Er zit namelijk erg veel verschil in de aantallen reizigers op vrijdag en in het weekend. De verschilreeks ligt dicht bij elkaar. De basis van het model dat gebruikt wordt ziet er als volgt uit:

$$\Delta^7 \hat{y}_t = \alpha + \beta \cdot \Delta^7 y_{vr} + \gamma \cdot \Delta^7 y_{t-7} + \delta \cdot \Delta^7 y_{t+7} + \epsilon \cdot Tl_t + \zeta \cdot Nl_t + U_t.$$

Hieronder worden de variabelen nader toegelicht.

#### **Verschilwaarde zaterdag/zondag $\Delta^7 \hat{y}_t$**

Dit is de te verklaren variabele en is het verschil in het aantal reizigers tussen zater-/zondag vorige week en zater-/zondag deze week.

#### **Verschilwaarde vrijdag $\Delta^7 y_{vr}$**

De eerste variabele is de verschilwaarde van deze vrijdag en vorige week vrijdag. Op deze manier kunnen wekelijkse fluctuaties in de voorspelling worden meegenomen. Er is gekozen voor de vrijdag om de zaterdag en de zondag te voorspellen. Als de zaterdag gebruikt zou worden om de zondag te voorspellen dan werkt de buitendienststelling van zaterdag door op zondag. De data van de buitendienststeldingsdagen zijn natuurlijk wel bewerkt, maar de data van de vrijdagen zijn echte data en zijn dus netter om mee te werken.

#### **Verschilwaarde vorige week zaterdag/zondag $\Delta^7 y_{t-7}$**

Deze variabele is het verschil tussen de zaterdag/zondag twee weken geleden en vorige week.

### Verschilwaarde volgende week zaterdag/zondag $\Delta^7 y_{t+7}$

Hier wordt het verschil genomen tussen deze week en volgende week. Eigenlijk kan er niet in de toekomst gekeken worden, maar omdat het doel van dit onderzoek is om de reizigersstromen bij buitendienststellingen zo goed mogelijk te voorspellen, is toch gekozen voor deze verklarende variabele. Uiteindelijk wordt in de praktijk alleen het model gebruikt om de kansen te schatten.

### Weer op de dag $Tl_t \in \{0, 1\}$ , $Nl_t \in \{0, 1\}$ en $Nh_t \in \{0, 1\}$

In dit onderzoek wordt gekeken naar het weekend, een groot gedeelte van de reizigers zijn dagjesmensen. Mensen laten vaak een dagje weg afhangen van het weer. Via het KNMI weerarchief zijn de maximum temperaturen en de hoeveelheden neerslag per dag gevonden in de Bilt. Ook zijn daar de gegevens gehaald van wat de gemiddelde temperatuur is per maand en de verwachte hoeveelheid neerslag per maand, weergegeven in een interval van het aantal millimeters neerslag. Dan is gekeken of de temperatuur op die dag laag (onder gemiddelde) of hoog (boven gemiddelde) is. Als het laag is wordt variabele  $Tl_t = 1$ , anders is het nul. Bij de neerslag is er een interval dus als er minder dan het interval is gevallen op die dag dan is het weinig, in het interval is middel en boven het interval is veel. De variabele voor weinig neerslag is  $Nl_t \in 0, 1$ . Omdat er in de onderzochte periode geen enkele keer een gemiddelde hoeveelheid neerslag viel is er geen variabele nodig van gemiddelde of hoge neerslag. In het algemeen zou dit wel kunnen, maar in dit onderzoek is dit niet bekeken. Dan is het nog mogelijk dat het weer in de ene regio in Nederland anders is dan in de andere. Hier is geen rekening mee gehouden. Bij de korte reizen zou dit wel mogelijk zijn, de mensen reizen daar vaak binnen een regio en dan is bekend dat ze vaak in de regio moeten zijn van de buitendienststelling. De mensen met middellange en lange reizen kunnen overal vandaan komen en overal heengaan. Ze gaan alleen toevallig over de buitendienststelling. Er is hier dus niet bekend naar welke regio hun trip gaat dus wat het weer is daar.

### $U_t$

Voor elk datapunt is te vinden hoeveel de afwijking is met het model. Deze afwijking wordt ook wel de storingsterm genoemd. De verwachtingswaarde van de storingsterm is nul en normaal verdeeld.

### 4.3.2 Vinden passend model

Tijdens dit onderzoek wordt het best passende model gezocht voor de reizigers aantallen per dataset. Er wordt begonnen met alle verklarende variabelen. De variabele met de minste significantie wordt telkens verwijderd, tot de variabelen die overblijven een significantie hebben van 95% of meer. Wanneer dit niet gevonden wordt, is gekozen om alle variabelen te gebruiken voor het model. Dit is gekozen omdat in dit geval de  $R^2$  de hoogste waarde heeft. De  $R^2$  is de maat waarin een model de werkelijke data benadert, hoe dichter bij 1 hoe beter de benadering is. Bij elke dataset worden passende parameters gevonden, telkens met het basismodel. Vanwege de verschillende aantallen reizigers tussen zaterdag en zondag, worden verschillende modellen gemaakt per dagsoort. Er komen dus  $96 \cdot 2 = 198$  verschillende schattingen die allemaal gevormd zijn vanuit het basismodel dat besproken is in deze paragraaf.

### 4.3.3 Aantal reizigers

De geschatte verschillen in aantallen moeten nog omgerekend worden tot aantallen. Dit wordt weer per dataset gedaan door de formule

$$\hat{y}_t = \Delta^7 \hat{y}_t - y_{t-7} \quad (4.1)$$

te gebruiken. Waarin  $\hat{y}_t$  het geschatte aantal reizigers op dag  $t$  is,  $\Delta^7 \hat{y}_t$  is het geschatte aantal reizigers dat berekend is met het model uit de vorige paragraaf en  $y_{t-7}$  is het aantal reizigers op dag  $t - 7$ . In het resultaten hoofdstuk staat beschreven wat de parameters behorende bij de verklarende variabelen zijn.

## 4.4 Model buitendienststellingen

Met de tijdreeksmodellen uit de vorige paragraaf is uit te rekenen wat het aantal reizigers is dat gereisd zou hebben als er geen buitendienststelling zou zijn. Nu kan er gekeken worden naar de kans dat een persoon toch gaat reizen, ondanks buitendienststellingen. Per buitendienststellingsdag en per lengte van de reis wordt er voor elk producttype uitgerekend wat de kans is dat er gereisd gaat worden. Dit wordt uitgerekend met het daadwerkelijke aantal reizigers en het aantal reizigers dat berekend is met vergelijking 4.1. Dus het aantal reizigers dat werkelijk gereisd heeft wordt gedeeld door het

aantal dat geschat is. Dit resulteert in de kans dat er gereisd is, in formulevorm kan het als volgt worden weergegeven

$$p_t = \frac{y_t}{\hat{y}_t}.$$

Met deze berekende kansen  $p_t$  is er de benodigde data om regressie uit te voeren, zodat er per producttype een schatting gemaakt kan worden dat de kansen voorspelt dat er wel gereisd wordt  $\hat{p}_t$ . Er worden geen schattingen meer gemaakt per buitendienststelling en per lengte. Dit is gedaan omdat er per buitendienststelling niet genoeg gegevens zijn, vaak maar drie weekenden dat er buitendienststelling is. Nu worden alle buitendienststellingen met elkaar vergeleken. De waarde  $p_t$  wordt gebruikt om het model te maken dat de kansen schat op blijven reizen. Alleen de  $p_t$ 's op dagen van buitendienststellingen worden hier voor gebruikt.

#### 4.4.1 Logistische regressie model

Zoals eerder vermeld, wordt een model gezocht dat de kansen op blijven reizen per producttype schat. Dit betekent dat de waarden van de te verklaren factor tussen de 0 en 1 moeten zitten. Een goed passend model hiervoor is het logistische regressiemodel,

$$p = \frac{e^{\alpha + \beta_1 \cdot L_k + \beta_2 \cdot L_m + \gamma \cdot Z_a + \delta \cdot Tr + \epsilon \cdot Bus + Z_t}}{1 + e^{\alpha + \beta_1 \cdot L_k + \beta_2 \cdot L_m + \gamma \cdot Z_a + \delta \cdot Tr + \epsilon \cdot Bus + Z_t}},$$

waarbij  $\hat{p}_t$  de kans is dat mensen blijven reizen, dus de te verklaren variabele. De verklarende variabelen worden hieronder nader toegelicht:

**Lengte van de reis**  $L_k, L_m \in \{0, 1\}$

Deze variabele geeft aan of de groep reizigers een korte  $L_k$  of middellange  $L_m$  reis maken. De korte reis is korter dan een halfuur en een middellange reis is tussen een half uur en een uur.

**Dag in het weekend**  $Z_a \in \{0, 1\}$

In het weekend zijn er zaterdagen en zondagen, deze variabele is 1 als de buitendienststelling op zaterdag is. Bij een buitendienststelling op zondag is de variabele 0.

### Treinen $Tr \in \{0, 1\}$ of alleen bussen $Bus \in \{0, 1\}$

Bij buitendienststellingen zijn drie verschillende opties, of er blijven enkele treinen rijden, of er rijden alleen bussen of er zijn omgeleide intercity's en bussen. Als er alleen bussen rijden is de variabele  $Bus = 1$  en als er minder treinen rijden dan is  $Tr = 1$ . Het is niet mogelijk dat  $Tr = Bus = 1$ , als er treinen en bussen zijn als vervangend vervoer dan wordt dat weergegeven als  $Tr = Bus = 0$ , dus niet alleen treinen en niet alleen bussen.

Om te kunnen werken met het logistische model is gebruik gemaakt van het hoofdstuk *Logistische Regressie* uit het boek van Moore en McCabe [15], hierin wordt het model omgeschreven tot

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 \cdot L_k + \beta_2 \cdot L_m + \gamma \cdot Za + \delta \cdot Tr + \epsilon \cdot Bus + Z_t.$$

Waarbij  $Z_t$  de restterm is. Zo wordt het een lineair regressie model waarbij de kans dat iemand blijft reizen wordt omgeschreven tot de natuurlijke logaritme van de odds. Er worden vier verschillende schattingen uitgevoerd, namelijk één per producttype.

## 4.5 Gemiddelden per buitendienststelling

In het model wordt rekening gehouden met drie eigenschappen van de buitendienststelling, namelijk het vervangende vervoer, de lengte van de reis en de dag waarop de buitendienststelling valt. Om te kijken of er in verder onderzoek niet beter per buitendienststelling een model gemaakt moet worden, wordt er gekeken naar wat het gemiddelde percentage mensen is dat blijft reizen per buitendienststelling voor alle producttypes samen. De gemiddelden worden berekend vanuit de berekende percentages  $p_t$  dan mensen blijven reizen. Er is gekozen voor het gemiddelde, omdat er maar vaak maar erg weinig dagen zijn dat er buitendienststelling is. Als er maar drie dagen zijn dat er buitendienststelling is, geeft bijvoorbeeld een mediaan niet een betere weergave dan het gemiddelde. Deze gemiddeldes zijn berekend om een snel beeld te krijgen in de verschillende kansen op blijven reizen per buitendienststelling. Dit is gedaan omdat blijkt dat het andere model niet erg goed voorspelt. Met het berekenen van deze gemiddelden kan bekeken worden of het voor verder onderzoek van belang is om nog meer onderzoek te doen naar het verschil in buitendienststellingen en zo het model dat gemaakt is te kunnen verbeteren. Vandaar is gekozen voor de productgroep van alle reizigers, want deze nemen alle reizigers mee en dit geeft een snel

inzicht in wat er gebeurt op een dag met buitendienststellingen. In het resultaten hoofdstuk worden per werkzaamhedendag de gemiddelde percentages en standaarddeviaties weergegeven.

Bij de vuistregel wordt er voor elke buitendienststelling vanuit gegaan dat 70% van de reizigers blijft reizen. Nu per buitendienststelling bekend is wat er gemiddeld blijft reizen, is het interessant om te kijken wat de afwijking is als het gemiddelde wordt gebruikt om te zien hoeveel reizigers er blijven reizen. Bijvoorbeeld bij de buitendienststelling Amersfoort-Utrecht is de gemiddelde kans dat mensen blijven reizen op een buitendienststellingsdag 85%. Per buitendienststelling tussen Amersfoort en Utrecht wordt dus uitgerekend wat de afwijking is met de 85%. Dit wordt gedaan voor elke buitendienststelling en zo ontstaat er per buitendienststelling eigenlijk een vast percentage, met daarbij de afwijking. Het wordt niet aangeraden om deze methode in de praktijk te brengen, het geeft alleen inzicht in de verschillen per buitendienststelling. Daarnaast kan er vergeleken worden met de vuistregel of dit een beter of slechtere manier is om te kijken naar buitendienststellingen.

## 4.6 Vuistregel NS

Met het model en de data kan onderzocht worden hoe goed de vuistregel voorspelt. In sectie 4.6.1 wordt beschreven hoe de vuistregel vergeleken wordt met het buitendienststellingsmodel. Hoe er vergeleken wordt met de data staat beschreven in sectie 4.6.2.

### 4.6.1 Vuistregel en buitendienststellingsmodel

Met de modellen kan voor elk producttype uitgerekend worden wat de kans is dat mensen blijven reizen. De modellen hebben verschillende verklarende variabelen om de kansen te voorspellen, namelijk *Bus*, *Za*, *Tr*. Voor al deze variabelen kan alleen nul of één worden ingevuld. Voor het vergelijken met de vuistregel worden alle verschillende mogelijkheden die er te maken zijn met de variabelen uitgerekend. Deze kansen kunnen dan vergeleken worden met de vuistregel van de NS.

### 4.6.2 Vuistregel en data

Nu wordt de vuistregel gebruikt om in te schatten hoeveel mensen er gaan reizen tijdens buitendienststellingen. Dit komt niet altijd overeen met de werkelijkheid. De kansen dat mensen blijven reizen die berekend zijn in dit

onderzoek, komen het dichtst bij de werkelijkheid. Dus wordt de vuistregel vergeleken met deze kansen. De afwijking die er is per dag, per buitendienststelling wordt berekend. Zo kan er gekeken worden hoe goed de vuistregel de aantallen reizigers voorspelt.

## **4.7 Afsluiting**

In dit hoofdstuk zijn de modellen beschreven die de reizigers aantallen voorspellen en de kans dat reizigers blijven reizen. In dit hoofdstuk is uitgelegd waarom gekozen is voor deze twee modellen. Er volgt dus geen wiskundige analyse meer, omdat dat in dit hoofdstuk is verwerkt. In het volgende hoofdstuk worden de gevonden parameters bij de verklarende variabelen gegeven.



## Hoofdstuk 5

# Resultaten

### 5.1 Inleiding

In dit hoofdstuk worden de resultaten weergegeven van de verschillende regressiemodellen die beschreven zijn in het vorige hoofdstuk. Het eerste model is gemaakt om de aantallen reizigers te schatten op weekenddagen, in paragraaf 5.2 staan de resultaten hiervan. Deze resultaten zijn de parameters die horen bij de verschillende verklarende variabelen, de significantie en de betrouwbaarheid van de schatting. In paragraaf 5.3 staan de resultaten van de logistische regressie op een zelfde manier weergegeven. Om te kijken wat het verschil in het blijven reizen per buitendienststelling worden in paragraaf 5.4 de gemiddelde percentages per buitendienststelling gegeven. In paragraaf 5.5 staan de resultaten van het vergelijken van de vuistregel met de data en het model.

### 5.2 Model aantal reizigers

Deze paragraaf laat de resultaten zien van het model dat het aantal reizigers voorspelt, uit paragraaf 4.3. Deze resultaten zijn onderverdeeld in de weergave van de parameter waarden, paragraaf 5.2.1 en de betrouwbaarheid van de aantallen reizigers, paragraaf 5.2.2

#### 5.2.1 Verschilreeks

De tabellen in deze paragraaf geven de parameters weer die horen bij de verklarende variabelen van het lineaire regressiemodel dat de verschilreeks van de aantallen reizigers schat. Per dataset wordt telkens gezocht naar

de best passende parameters bij het basismodel. In paragraaf 4.3.3 is be-

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	5.95045	0.06642*	-0.34348*	-0.36622*			0.5513
	Reg	-0.22392	0.30690**		-0.41880*			0.4968
	Dal	-1.39899		0.2782**				0.332
	Student	3.3552		-0.5155**	-0.5191**			0.488
za:Middel	Alles	21.8643		-0.4612**	-0.4944**			0.6034
	Reg	2.4521		-0.4417**	-0.4979**			0.5324
	Dal	13.3499	0.4103*	-0.3980**	-0.4074**			0.7164
	Student	-6.6408		-0.5631***	-0.4352**	21.1915*		0.6705
za:Lang	Alles	6.08597	0.23870**	-0.35305*				0.5292
	Reg	1.9371		-0.4334*	0.4404*			0.3934
	Dal	4.4340		-0.4896**				0.4407
	Student	0.69437	0.09063**	-0.35731*				0.4222
zo:Kort	Alles	11.0428		-0.4722*	-0.4426*			0.4177
	Reg	1.2824		-0.5171**	-0.5225**			0.6279
	Dal	5.5950		-0.4672*	-0.4081*			0.4278
	Student	2.4833		-0.4624**	-0.4853**			0.4857
zo:Middel	Alles	14.4118		-0.5161**	-0.5189**			0.6599
	Reg	1.5918		-0.4608*	-0.4479**			0.497
	Dal	12.6818		-0.5246**	-0.5259***			0.701
	Student	-0.49121	0.04693**					0.3492
zo:Lang	Alles	6.02730	0.13512**	-0.62316***	-0.49105***			0.7436
	Reg	0.4842		-0.4713*	-0.4748**			0.3999
	Dal	3.8945		-0.5540**	-0.5879***			0.5784
	Student	0.7158		-0.4638*	-0.4649*			0.4366

Tabel 5.1: Almere - Weesp

schreven hoe dit in zijn werk gaat. In de tabel worden de bijbehorende parameters en het significantie niveau weergegeven, het significantieniveau wordt door het gebruikte programma gegeven. Met sterren is weergegeven wat het significantie niveau is voor de verschillende variabelen.

- \* dat de variabele in meer dan 95% van de gevallen significant is.
- \*\* dat de variabele in meer dan 99% van de gevallen significant is.
- \*\*\* dat de variabele in meer dan 95% van de gevallen significant is.

Als een cel van de tabel leeg is, betekent het dat die variabele in het bijbehorende model niet significant is en dus niet gebruikt wordt in het model. Voor de duidelijkheid wordt er een voorbeeld gegeven uit de eerste tabel, tabel 5.1. Deze tabel gaat over de buitendienststelling tussen Almere en Weesp. De bovenste regel geeft het model weer voor de OD's met korte reizen op een zaterdag waarbij gekeken is naar alle reizigers. In de derde kolom staat de waarde die hoort bij de  $\alpha$ , dit is de constante en is in veel gevallen niet significant, maar hoort bij het model. De vierde kolom is de parameter  $\beta$  behorende bij de variabele  $\Delta^7 y_{vr}$ . Deze variabele heeft een significantie niveau van één ster, wat overeenkomt met een significantie van meer dan 95%. Zo geven kolom vijf en zes de parameters voor resp. de variabelen  $\Delta^7 y_{t-7}$  en  $\Delta^7 y_{t+7}$ . De variabelen  $Tl_t$  en  $Nl_t$  zijn niet belangrijk voor dit model en daarom zijn de bijbehorende cellen leeg. De laatste kolom geeft de

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	-3.14485	0.11576**					0.39
	Reg	-5.9310	0.1415	-0.1833	-0.1637	16.6584	-0.7619	0.1508
	Dal	-5.84530	0.28199**					0.3918
	Student	-8.2067		-0.4537**	-0.5016**			0.5042
za:Middel	Alles	16.29981	0.10782**					0.4034
	Reg	1.4714	0.3595**					0.3403
	Dal	15.5370		-0.4726*	-0.5003*			0.3341
	Student	-7.0687		-0.5049***	-0.5726***	48.5887*		0.7668
za:Lang	Alles	123.0096		-0.5177***	-0.5174***			0.7543
	Reg	91.24480	0.06458	-0.34657	-0.34286	7.48938	-84.47088	0.2114
	Dal	-20.3966		-0.4055**	-0.5343***	177.3107*		0.7609
	Student	38.4118		-0.5421***	-0.5436***			0.8041
zo:Kort	Alles	27.1314		-0.3987*	-0.3956*			0.3248
	Reg	4.1284		-0.4097*	-0.4063*			0.371
	Dal	7.45290			0.19207**			0.329
	Student	11.0508		-0.4804**	-0.4922**			0.4894
zo:Middel	Alles	20.2802	0.20073**	0.5078**	-0.5169**			0.4925
	Reg	-0.22628						0.291
	Dal	10.1358		0.5258**	-0.5277**			0.4588
	Student	7.1617		0.5048**	0.5148**			0.5769
zo:Lang	Alles	108.5848		-0.6033***	-0.6061***			0.5586
	Reg	18.4054*		-0.5135***	-0.5390***			0.6688
	Dal	70.5955		-0.5797***	-0.5911***			0.6081
	Student	12.5751		-0.5018*	-0.5015*			0.3744

Tabel 5.2: Amersfoort - Utrecht

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	74.6424		-0.5487**	-0.5610**			0.3972
	Reg	123.0096		-0.5177***	-0.5174***			0.7543
	Dal	40.8259		-0.5078**	-0.5170**			0.41
	Student	20.0306		-0.6095**	-0.6141**			0.4906
za:Middel	Alles	50.4944		-0.4311*	-0.4471*			0.3615
	Reg	10.2578	0.4047**	-0.4579*	-0.4488*			0.4309
	Dal	13.4202						0.3502
	Student	16.0626		-0.5646**	-0.5823***			0.5669
za:Lang	Alles	25.98222		0.35522**				0.4336
	Reg	63.4204	0.7138*				-64.7442*	0.3391
	Dal	8.8559	0.9618***	-0.2764*				0.7144
	Student	8.9650		-0.4953*	-0.4841*			0.4069
zo:Kort	Alles	62.4006		-0.5349***	-0.5360***			0.6735
	Reg	108.5848		-0.6033***	-0.6061***			0.5586
	Dal	25.4853		-0.5280**	-0.5287**			0.6106
	Student	18.8603		-0.5405***	-0.5399***			0.7188
zo:Middel	Alles	35.4446		-0.5035**	-0.4941***			0.5866
	Reg	4.8540		-0.4804**	-0.4911**			0.5326
	Dal	15.5138		-0.5030**	-0.4687**			0.4989
	Student	13.3137		-0.4681**	-0.4996**			0.5953
zo:Lang	Alles	21.4682	0.4358***					0.4519
	Reg	-15.366	1.157**			36.095*		0.5144
	Dal	0.3896	1.3289***					0.6497
	Student	3.4977		-0.5028**	-0.5075**			0.5027

Tabel 5.3: Amsterdam Sloterdijk - Haarlem

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	-140.60832	0.01371	-0.38824	-0.32085	88.90145	121.66123	0.3064
	Reg	-67.97032	-0.09235	-0.29004	-0.23585	36.77358	60.46420	0.3503
	Dal	-73.49336	0.03223	-0.28539	-0.27164	40.96486	62.33229	0.2684
	Student	2.1840		-0.4784**	-0.4971**			0.4448
za:Middel	Alles	5.61083	0.15777**					0.3571
	Reg	2.2204	0.2717*	-0.4148*				0.4332
	Dal	1.3226	0.3558**					0.3692
	Student	3.61555	0.10222***					0.4204
za:Lang	Alles	18.16915	0.29675***					0.6747
	Reg	29.0866	0.2445	-0.3983	-0.3307	-3.1075	-22.2820	0.2497
	Dal	3.9986	0.6041***					0.6003
	Student	7.9550	0.1496***					0.5259
zo:Kort	Alles	8.3420			-0.4335*			0.1882
	Reg	36.28597	0.04642	-0.36853	-0.38933	-15.20177	-29.79837	0.2965
	Dal	-0.1733	-0.4745					0.2289
	Student	6.2637		-0.3939*	-0.3740*			0.3121
zo:Middel	Alles	16.5752		-0.4613**	-0.4865**			0.4719
	Reg	1.4567		-0.4719**	-0.4888**			0.5356
	Dal	3.9752		-0.4613**	-0.4769**			0.4956
	Student	9.8196		-0.4364*	-0.4561*			0.2673
zo:Lang	Alles	50.7577		-0.5693**	-0.5736***			0.566
	Reg	8.9521		-0.5717***	-0.5659***			0.7521
	Dal	24.2275		-0.5605***	-0.5934***			0.6025
	Student	12.6852		-0.4182*	-0.4065*			0.3094

Tabel 5.4: Den Haag Centraal - Gouda

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	4.2763	0.1074*					0.1791
	Reg	7.4240	0.1644	-0.3225	-0.2937	-12.9990	0.3368	0.3277
	Dal	-0.1339	0.5493*	-0.3669*				0.4748
	Student	3.8286		-0.4082*	-0.4409*			0.3678
za:Middel	Alles	3.4673		-0.4569*	-0.4910*			0.4201
	Reg	2.0126		-0.5115**	-0.5424**			0.4806
	Dal	-0.4927		-0.4568*	-0.4691*			0.3652
	Student	1.9315		-0.4146*	-0.4608*			0.4779
za:Lang	Alles	2.1682	0.2629*	-0.3940*	-0.4972*			0.357
	Reg	0.3917		-0.5135*	-0.4972*			0.3871
	Dal	-10.34659	0.02281	-0.33289	-0.26978	4.29987	10.10957	0.1641
	Student	16.97999	0.14881*	-0.5536***	-0.43551**		-16.56746*	0.793
zo:Kort	Alles	2.9231		-0.4709*	-0.4890*			0.3513
	Reg	19.667*					-22.417*	0.1889
	Dal	0.6237		-0.4818*	-0.4965*			0.2587
	Student	1.32721	0.03841***					0.3326
zo:Middel	Alles	2.5697		-0.4448*	-0.4292*			0.378
	Reg	7.3490	0.1559	-0.2109	-0.2429	-5.7805	-5.2049	0.1751
	Dal	1.1869		-0.4642*	-0.4612*			0.3234
	Student	0.6823		-0.5050***	-0.4965***			0.6963
zo:Lang	Alles	3.47391	0.04983	-0.30636	-0.33542	-0.64142	-3.31783	0.009928
	Reg	-0.4284		-0.5014*	-0.5230*			0.3538
	Dal	-4.39241	-0.01868	-0.30096	-0.27421	2.76543	3.56365	0.2041
	Student	-0.4284		-0.5014*	-0.5230*			0.3538

Tabel 5.5: Hoofddorp - Leiden

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	5.26640	0.03838**					0.3138
	Reg	1.5756		-0.4636*	-0.3602**			0.3801
	Dal	1.7117	-0.4172*	-0.4184*				0.3963
	Student	2.5830		-0.509**	-0.5256**			0.4897
za:Middel	Alles	3.8510		-0.4606**	-0.4670**			0.5315
	Reg	0.5071	0.3971**					0.3046
	Dal	3.1208	0.2776*	-0.3996**	-0.3122*			0.6817
	Student	-1.6482		-0.4185*	-0.4279*			0.415
za:Lang	Alles	11.6041		-0.5195**	-0.5304**			0.6035
	Reg	1.5756		-0.4636*	-0.3602*			0.3801
	Dal	4.6053		-0.5064**	-0.5177**			0.5302
	Student	4.5152		-0.5139**	-0.5153**			0.5161
zo:Kort	Alles	41.37510	-0.01367	0.03516	-0.08927	4.01577	-49.51039	0.1752
	Reg	0.3779		-0.6926***				0.473
	Dal	23.04417	-0.03504	0.02824	-0.07593	-5.55014	-25.92437	0.1284
	Student	9.659882	-0.003595	-0.251249	-0.330627	-2.264748	-10.963749	0.2112
zo:Middel	Alles	2.5705		-0.5216**	-0.5264**			0.5953
	Reg	-0.28702	0.28507**					0.3242
	Dal	0.3180		-0.5290**	-0.5285**			0.5561
	Student	2.8015		-0.4846***	-0.4905**			0.6004
zo:Lang	Alles	9.8760		-0.5964***	-0.5948***			0.6922
	Reg	0.3779		-0.6926***				0.473
	Dal	6.7354		-0.5343***	-0.5342***			0.7492
	Student	1.0601		-0.6888**	-0.6917***			0.6204

Tabel 5.6: Leiden - Alphen aan de Rijn

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	5.0428		-0.5180**	-0.4672**			0.6098
	Reg	1.9052		-0.5790***	-0.5305***			0.7172
	Dal	2.0066		-0.4769*	-0.4673*			0.4253
	Student	0.7482		-0.4310*	-0.4440*			0.4704
za:Middel	Alles	5.8379		-0.4930**	-0.5148**			0.4731
	Reg	1.7290		-0.4331*	-0.4607*			0.4952
	Dal	2.4761		-0.4184*	-0.4279*			0.3318
	Student	0.8953		-0.5384**	-0.5430**			0.4952
za:Lang	Alles	244.47884	0.05306	-0.01639	-0.21811	-24.79940	-229.84389	0.1501
	Reg	0.9072	0.4217*					0.2182
	Dal	89.02401	0.36543	-0.04229	-0.23018	-20.60454	-76.80923	0.2027
	Student	98.08610	0.04142	0.01203	-0.21096	-22.03465	-89.53307	0.14
zo:Kort	Alles	-63.41400	-0.04899	-0.27322	-0.11669	77.14381	0.4177	34.792550.2654
	Reg	-2.30964	-0.03606	-0.29491	-0.25677	8.17221	-1.69281	0.1175
	Dal	-48.8561	-0.3314	-0.1410	-0.1810	38.9727	37.6962	0.3318
	Student	-5.48591	-0.01544	-0.37470	-0.22847	24.65861	-5.14868	0.2493
zo:Middel	Alles	-14.5318		-0.4903**	-0.4867*			0.3282
	Reg	0.3729		-0.4460*	-0.4476*			0.2826
	Dal	12.62201	-0.08088	-0.49740	-0.46186	12.28007	-21.48442	0.274
	Student	-14.4953		-0.5441**	-0.5652*			0.3717
zo:Lang	Alles	40.6911		-0.5069*	-0.5059*			0.3137
	Reg	9.0110	0.2148	-0.3629	-0.2780	3.9148	-9.0448	0.1854
	Dal	27.1122	-0.2091	-0.3358	-0.2446	64.0002	-45.3137	0.1235
	Student	17.4325		-0.5122*	-0.5081*			0.3621

Tabel 5.7: Sittard - Maastricht

	Type	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$R^2$
za:Kort	Alles	14.52717	0.28157**					0.632
	Reg	2.7839	0.8309***					0.7016
	Dal	4.0385	0.7757***					0.7083
	Student	6.05439	0.08337***					0.6452
za:Middel	Alles	33.77712	0.07545*	-0.39628**	-0.42791**			0.6977
	Reg	1.5902	0.5750***					0.5796
	Dal	17.0743	0.4958***	-0.3416**	-0.2688*			0.8001
	Student	10.0561		-0.5445***	-0.6027***			0.7174
za:Lang	Alles	13.49211		0.27363***				0.6038
	Reg	10.9356	0.1140	-0.3730**	-0.3076	12.1650	-11.9526	0.3884
	Dal	9.8399	0.5809***	-0.5672***				0.7378
	Student	8.8570			-0.5559***			0.6497
zo:Kort	Alles	-2.86980	0.07076**					0.3753
	Reg	1.19951	0.25814***		-0.33709*			0.4833
	Dal	-9.51607	0.20509**					0.3315
	Student	17.20833	0.01369	-0.25909	-0.30868	-17.71275	-2.74781	0.2146
zo:Middel	Alles	14.1175		-0.5863***	-0.5728***			0.6592
	Reg	0.3159		-0.5296**	-0.5262**			0.4958
	Dal	4.18087	0.35750***					0.5013
	Student	7.2369		-0.4037*	-0.4018*			0.3998
zo:Lang	Alles	28.8431	0.13512**	-0.6348***	-0.6448***			0.5686
	Reg	2.1835		-0.5256**	-0.5492**			0.4821
	Dal	19.8088		-0.6167***	-0.6217***			0.606
	Student	4.3632		-0.4778*	-0.4780*			0.336

Tabel 5.8: Zaandam - Amsterdam Sloterdijk

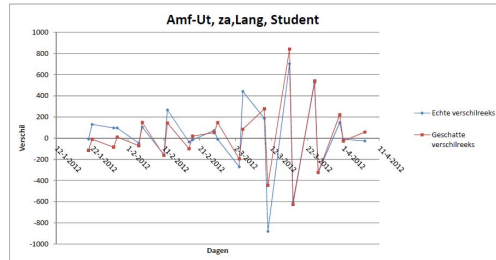
t van het model weer, dit is 0,5513. Het totale model wordt

$$\Delta^7 \hat{y}_t = 5.95045 + 0.06642 \cdot \Delta^7 y_{vr} - 0.34348 \cdot \Delta^7 y_{t-7} - 0.36622 \cdot \Delta^7 y_{t+7}.$$

Voor alle datasets is op deze manier het passende model beschreven. Wanneer geen enkele variabele een significantie niveau heeft staan, dan betekent het dat er geen model gevonden kon worden waarbij de variabelen genoeg significantie hadden. Op dat moment zijn alle variabelen gebruikt om het model te maken. In de tabellen 5.1-5.8 zijn de resultaten van alle datasets weergegeven. Wat opvalt is dat vaak de combinatie van de variabelen  $\Delta^7 y_{t-7}$  en  $\Delta^7 y_{t+7}$  of de variabele  $\Delta^7 y_{vr}$  significant is voor het model. De variabelen voor het weer zijn bijna nooit belangrijk voor het model. Over het algemeen is de  $R^2$  van het model niet erg hoog. Bij sommige buitendienststellingen is het zelfs erg laag. Per buitendienststelling verschilt het ook sterk welke productgroep het best te voorspellen is.

### Kwaliteitsverschil

Om een beter beeld te geven van het kwaliteitsverschil in de voorspellingen zijn grafieken met de beste en de slechtste voorspelling weergegeven. Dus degene met de hoogste en de laagste  $R^2$ . In grafiek 5.1 zijn de geschatte en echte verschilreeksen weergegeven voor het traject Amersfoort-Utrecht voor het reizigerstype studenten. Voor de zaterdagen heeft dit model de hoogste  $R^2$  waarde, namelijk 0,8041. In deze grafiek is duidelijk te zien dat



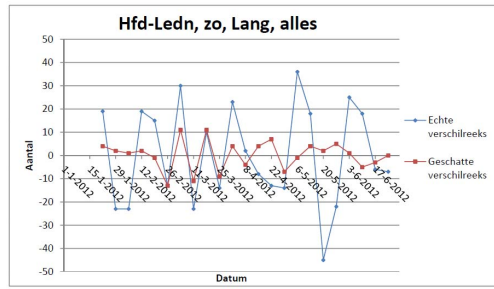
Figuur 5.1: Weergave geschatte en echte verschilreeks Amersfoort-Utrecht

de echte en geschatte punten erg dicht bij elkaar liggen. In grafiek 5.2 staan de geschatte en echte verschilreeksen voor alle reizigerstypes in de OD's over het traject Hoofddorp-Leiden. De zondagen hebben de laagste  $R^2$  namelijk 0.09928. het traject Hoofddorp-Leiden. Er is duidelijk te zien dat er best wel veel kwaliteitsverschil is tussen deze voorspellingen. Met deze grafieken is een duidelijk beeld gegeven van wat de  $R^2$  betekent.

### 5.2.2 Aantallen reizigers

De verschilreeks is nu voorspelt, maar het is interessant om te kijken naar de aantallen reizigers die hieruit volgen met vergelijking 4.1. Deze aantallen zijn namelijk nodig om de kansen te voorspellen dat mensen blijven reizen. In deze paragraaf worden de aantallen die berekend zijn vergeleken met de echte aantallen, maar dan zonder de dagen dat er buitendienststelling is. Het is zonder de buitendienststellingsdagen, omdat deze vergelijking kijkt hoe goed de modellen de aantallen reizigers voorspellen. In de tabellen hieronder is het gemiddelde en de standaard afwijking van de procentuele afwijking weergegeven.

Als voorbeeld tabel 5.9, deze hoort bij buitendienststelling Almere-Weesp. De eerste kolom geeft aan wat de lengte is van de OD's en de dag waarop de buitendienststelling is. De eerste regel heeft OD's die korter zijn dan een half uur op zaterdag en alle reizigers zijn meegerekend. Dan blijkt de gemiddelde procentuele afwijking tussen de echte aantallen en de geschatte aantallen  $-0.3\%$  te zijn en de standaardafwijking  $10.3\%$ . Voor elke buiten-



Figuur 5.2: Weergave geschatte en echte verschilreeks Hoofddorp-Leiden

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-0.3%	10.2%
	Reg	-1.1%	17.9%
	Dal	-0.1%	12.5%
	Student	-1.3%	13.2%
za:Middel	Alles	-2.6%	15.2%
	Reg	-3.6%	17.0%
	Dal	-1.4%	13.2%
	Student	-3.2%	16.7%
za:Lang	Alles	-3.4%	15.4%
	Reg	-12.6%	46.1%
	Dal	-5.1%	25.5%
	Student	-1.6%	16.7%
zo:Kort	Alles	-0.7%	10.1%
	Reg	-1.9%	21.3%
	Dal	-1.0%	9.2%
	Student	-1.4%	13.2%
zo:Middel	Alles	-2.4%	17.8%
	Reg	-5.0%	38.0%
	Dal	-3.1%	17.1%
	Student	-2.0%	23.8%
zo:Lang	Alles	-4.8%	14.9%
	Reg	-6.8%	33.6%
	Dal	-3.7%	14.4%
	Student	-3.1%	18.2%

Tabel 5.9: Afwijking Almere-Weesp



dienststelling zijn zo de percentages te vinden in de tabellen 5.9-5.16. De

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-0.9%	11.0%
	Reg	-1.3%	15.8%
	Dal	-1.3%	11.5%
	Student	-0.5%	10.4%
za:Middel	Alles	-0.8%	9.7%
	Reg	-2.3%	18.1%
	Dal	-0.4%	11.9%
	Student	0.0%	6.2%
za:Lang	Alles	-0.4%	6.2%
	Reg	-2.9%	13.9%
	Dal	-0.4%	5.5%
	Student	-0.3%	7.2%
zo:Kort	Alles	-0.4%	18.7%
	Reg	-3.9%	29.8%
	Dal	-2.7%	17.9%
	Student	-2.1%	13.1%
zo:Middel	Alles	-0.5%	17.0%
	Reg	-4.2%	29.5%
	Dal	-1.0%	14.6%
	Student	-1.1%	10.3%
zo:Lang	Alles	-2.0%	14.7%
	Reg	-7.1%	16.9%
	Dal	-2.5%	14.9%
	Student	-1.9%	18.6%

Tabel 5.10: Afwijking Amersfoort-Utrecht

meeste gemiddelden zijn negatief, dit betekent dat het model vaak de aantallen te hoog inschat. Dan is het bij de meeste buitendienststellingen zo dat de gemiddelden tussen de  $-4\%$  en  $0\%$  liggen, maar er zijn ook uitschieters naar boven de  $80\%$ . De standaard afwijking zit vaak tussen de  $15\%$  en  $20\%$ , maar ook hier komen erg grote uitschieters naar voren. Wanneer alle reizigers mee genomen worden zijn de voorspellingen het meest stabiel, er is wel altijd een afwijking maar die is vaak gelijk. Bij de productgroepen apart kan er erg veel afwijking ontstaan en zijn de modellen slecht. Daarnaast zijn er ook voorbeelden dat het juist beter te voorspellen is per productgroep. Zoals bijvoorbeeld de studenten op zaterdag bij Amersfoort-Utrecht. Wat het meest op valt is dat de voorspellingen bij een aantal buitendienststellingen zeer slecht zijn, zoals bijvoorbeeld de buitendienststelling Hoofddorp-Leiden.

### 5.2.3 Kwaliteitsverschil

Ook zijn voor het beeld van het verschil in kwaliteit de aantallen reizigers weergegeven in een grafiek. In de figuren 5.3 en 5.4 staan dezelfde groepen als in de vorige paragraaf maar dan met het aantal reizigers. Ook hier is weer te zien dat er erg veel verschil zit in de voorspelkwaliteit van de modellen.

In de grafieken staan de buitendienststeldingsdagen wel weergegeven. Er

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-1.1%	14.5%
	Reg	-1.9%	14.9%
	Dal	-0.8%	12.8%
	Student	-1.7%	27.1%
za:Middel	Alles	-1.6%	14.2%
	Reg	-2.0%	31.7%
	Dal	-0.8%	13.4%
	Student	-1.4%	13.5%
za:Lang	Alles	-4.8%	23.1%
	Reg	-5.2%	36.0%
	Dal	-1.1%	17.0%
	Student	2.6%	45.0%
zo:Kort	Alles	-1.8%	30.3%
	Reg	-5.9%	29.1%
	Dal	-3.2%	23.2%
	Student	-4.7%	24.3%
zo:Middel	Alles	-0.7%	21.3%
	Reg	-5.7%	24.2%
	Dal	-3.3%	16.5%
	Student	-1.8%	10.8%
zo:Lang	Alles	-10.0%	36.5%
	Reg	-21.3%	76.8%
	Dal	-6.9%	35.8%
	Student	-2.9%	17.4%

Tabel 5.11: Afwijking Amsterdam Sloterdijk-Haarlem

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-0.6%	9.8%
	Reg	-2.1%	61.2%
	Dal	-2.5%	25.4%
	Student	-3.3%	50.8%
za:Middel	Alles	-1.1%	8.5%
	Reg	-2.1%	75.6%
	Dal	-1.3%	13.0%
	Student	0.1%	15.6%
za:Lang	Alles	-1.0%	14.1%
	Reg	-4.8%	42.7%
	Dal	-2.6%	16.0%
	Student	-2.4%	18.8%
zo:Kort	Alles	-3.4%	31.1%
	Reg	-12.1%	54.7%
	Dal	-0.7%	24.0%
	Student	-6.3%	26.4%
zo:Middel	Alles	-1.4%	11.8%
	Reg	-2.1%	21.5%
	Dal	-0.8%	11.8%
	Student	-3.2%	13.8%
zo:Lang	Alles	-2.9%	12.4%
	Reg	-8.7%	22.0%
	Dal	-2.3%	12.1%
	Student	-3.8%	16.8%

Tabel 5.12: Afwijking Den Haag-Gouda

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-2.8%	29.4%
	Reg	17.2%	152.3%
	Dal	7.6%	70.2%
	Student	10.2%	120.5%
za:Middel	Alles	-0.8%	24.7%
	Reg	14.7%	334.9%
	Dal	4.2%	101.0%
	Student	3.7%	180.6%
za:Lang	Alles	-1.7%	26.7%
	Reg	-11.1%	59.2%
	Dal	3.3%	306.6%
	Student	105.7%	1067.1%
zo:Kort	Alles	-2.0%	11.9%
	Reg	-4.3%	21.9%
	Dal	-2.6%	17.2%
	Student	-1.5%	20.8%
zo:Middel	Alles	-2.9%	15.8%
	Reg	-9.4%	38.2%
	Dal	-3.0%	23.0%
	Student	-3.7%	15.8%
zo:Lang	Alles	-3.4%	19.5%
	Reg	-8.2%	39.4%
	Dal	-4.4%	21.2%
	Student	-2.3%	27.3%

Tabel 5.13: Afwijking Hoofddorp-Leiden

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-1.5%	8.8%
	Reg	87.1%	300.5%
	Dal	9.6%	32.2%
	Student	12.7%	60.6%
za:Middel	Alles	-0.2%	16.4%
	Reg	21.4%	71.7%
	Dal	1.3%	20.8%
	Student	4.3%	23.3%
za:Lang	Alles	1.4%	15.8%
	Reg	87.1%	300.5%
	Dal	9.8%	28.5%
	Student	16.0%	38.9%
zo:Kort	Alles	-1.7%	12.2%
	Reg	-3.5%	12.0%
	Dal	-0.5%	15.7%
	Student	-2.4%	15.3%
zo:Middel	Alles	0.1%	12.8%
	Reg	-1.9%	28.9%
	Dal	-0.1%	15.4%
	Student	-0.8%	13.1%
zo:Lang	Alles	-2.6%	16.7%
	Reg	-3.5%	12.0%
	Dal	-3.4%	16.7%
	Student	-3.3%	21.5%

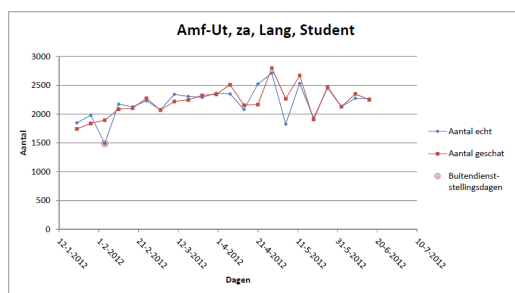
Tabel 5.14: Afwijking Leiden-Alphen aan de Rijn

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-1.1%	6.5%
	Reg	-9.1%	95.9%
	Dal	-5.1%	38.1%
	Student	-5.2%	46.5%
za:Middel	Alles	-1.0%	10.5%
	Reg	-76.1%	461.9%
	Dal	-9.9%	92.3%
	Student	-11.5%	75.2%
za:Lang	Alles	-1.3%	15.5%
	Reg	-16.3%	190.0%
	Dal	-2.7%	21.0%
	Student	-0.8%	16.0%
zo:Kort	Alles	-8.0%	32.4%
	Reg	-11.9%	54.2%
	Dal	-8.9%	33.1%
	Student	-8.2%	30.1%
zo:Middel	Alles	-1.7%	32.9%
	Reg	-9.8%	54.0%
	Dal	-6.0%	45.3%
	Student	-5.3%	51.8%
zo:Lang	Alles	-7.0%	26.0%
	Reg	-13.9%	44.2%
	Dal	-7.0%	22.6%
	Student	-10.8%	40.7%

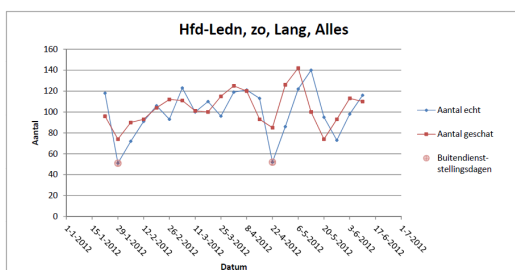
Tabel 5.15: Afwijking Sittard-Maastricht

	Type	Gemiddelde	St.dev.
za:Kort	Alles	-1.8%	17.5%
	Reg	-1.3%	16.6%
	Dal	-1.3%	14.2%
	Student	-1.9%	14.4%
za:Middel	Alles	-1.1%	13.4%
	Reg	-1.2%	15.7%
	Dal	-0.8%	9.9%
	Student	-0.6%	12.9%
za:Lang	Alles	-0.7%	10.5%
	Reg	-2.3%	15.0%
	Dal	-0.9%	9.6%
	Student	-0.1%	12.3%
zo:Kort	Alles	-0.1%	7.7%
	Reg	0.4%	12.7%
	Dal	0.1%	10.4%
	Student	-1.2%	13.5%
zo:Middel	Alles	-0.6%	7.8%
	Reg	-0.2%	17.5%
	Dal	-2.1%	15.6%
	Student	-2.6%	9.1%
zo:Lang	Alles	-1.1%	11.0%
	Reg	-2.9%	27.0%
	Dal	-0.9%	12.6%
	Student	-2.6%	14.5%

Tabel 5.16: Afwijking Zaandam-Amsterdam Sloterdijk



Figuur 5.3: Weergave geschatte en echte aantallen reizigers Amersfoort-Utrecht



Figuur 5.4: Weergave geschatte en echte aantallen reizigers Hoofddorp-Leiden

is te zien dat het model er geen rekening mee houdt, want het voorspelt vaak te hoog voor die dagen.

### 5.3 Model buitendienststellingen

In deze paragraaf zijn de resultaten weergegeven van het model uit paragraaf 4.4 voor het voorspellen van de kansen bij de buitendienststellingen. Zoals beschreven in paragraaf 4.4 is nu per producttype een schatting gemaakt. In tabel 5.17 worden de parameters weergegeven die gevonden zijn bij deze vier schattingen. Als voorbeeld uitleg bij de eerste regel van tabel 5.17. Deze regel geeft het model weer om voor alle reizigers te voorspellen wat de kans is dat ze blijven reizen. Voor deze groep zijn de variabelen  $Bus$ ,  $Tr$  en

Type	$\alpha$	$\beta_1$	$\beta_2$	$\gamma$	$\delta$	$\epsilon$	$R^2$
Alles	1.1117***	-1.4218***			1.7772***	-0.9205*	0.3569
Reg	0.4357				1.6613**	1.1127*	0.1667
Dal	1.0801***	1.4908***			1.8064***	-0.8350*	0.3348
Student	0.7068**	1.3669***			2.1818***		0.2665

Tabel 5.17: Parameters model buitendienststellingen

$L_k$  significant. De cellen voor de variabelen  $Za$  en  $L_m$  zijn leeg omdat deze niet genoeg significantie tonen voor het model. Zo zijn alle variabelen die besproken zijn in hoofdstuk 4 weergegeven in een kolom. Het model van de eerste regel ziet er als volgt uit

$$\ln\left(\frac{p}{1-p}\right) = 1.1117 - 1.4218 \cdot L_k + 1.7772 \cdot Tr - 0.9205 \cdot Bus.$$

Het valt op dat de variabelen  $Za$  en  $L_m$  niet significant zijn voor de verschillende modellen. Blijkbaar zit er geen groot verschil in kansen op blijven reizen op een zaterdag of zondag. De mensen met de middellange reizen en de mensen met de lange reizen vertonen ook hetzelfde gedrag bij buitendienststellingen. Parameters  $\beta_1$  en  $\epsilon$  zijn telkens negatief. Dit geeft aan dat buitendienststellingen waar alleen een bus als vervangend vervoer wordt ingezet voor reizigers met korte reizen een grote barière vormt. Als er gewoon treinen rijden, ook al zijn het er minder, is te zien aan parameter  $\gamma$  dat mensen gewoon blijven reizen. Wat nog meer opvalt is dat de kwaliteit van het voorspellen erg laag is, de  $R^2$  is lager dan 0.4.

#### 5.3.1 Kansen

De kansen die verkregen zijn uit de modellen van de vorige paragraaf worden vergeleken met de kansen uit de dataset. De procentuele afwijking van

de kansen uit de dataset vergeleken met de kansen die berekend zijn staan in tabel 5.18. Het is duidelijk dat er erg slecht voorspelt wordt met het

Type	Gemiddelde	St.dev.
Alles	-37.5%	81.5%
Reg	-113.1%	299.5%
Dal	-42.84%	90.6%
Student	-69.7%	176.0%

Tabel 5.18: Afwijking kansen op reizen

gemaakte model. Het model voorspelt te hoog en daarnaast zit er blijkbaar meer verschil tussen de buitendienststellingen dan alleen de dag en het vervangende vervoer, dit is te zien aan de standaarddeviatie.

## 5.4 Voorspellen met gemiddelden per buitendienststelling

Om te kijken of er veel verschil zit tussen de kansen om te blijven reizen per buitendienststellingen, is van elke buitendienststelling uitgerekend wat de gemiddelde kans is dat reizigers blijven reizen. In tabel 5.19 zijn deze gemiddelden en standaarddeviaties weergegeven. In de tabel is te zien dat het per buitendienststelling uitmaakt welk percentage van reizigers blijft reizen. De buitendienststelling waar de meeste mensen blijven reizen hebben vaak de inzet van bussen en omgeleide intercity's, zoals Amersfoort-Utrecht en Den Haag-Gouda. Toch zegt dit niet alles want bijvoorbeeld Hoofddorp-

Buitendienststelling	Gemiddelde	St.dev.
Almere-Weesp	24.3%	7.2%
Amersfoort-Utrecht	85.9%	23.7%
Asd Sloterdijk-Haarlem	58.5%	20.2%
Den Haag Centraal-Gouda	95.6%	10.8%
Hoofddorp-Leiden	21.8%	4.36%
Leiden-Alphen a/d Rijn	57.2%	6.1%
Sittard-Maastricht	35.0%	12.0%
Zaandam-Asd Sloterdijk	77.5%	23.2%

Tabel 5.19: Percentages reizigers dat blijft reizen

Leiden heeft ook bussen en omgeleide intercity's als vervangend vervoer, maar er blijven weinig reizigers over.

Wanneer deze gemiddelden als voorspelling gebruikt worden bij buitendienststellingen kan ook berekend worden wat de gemiddelde afwijking is. In de tabel 5.20 wordt per buitendienststelling weergegeven wat de betrouwbaarheid is als het gemiddelde percentage reizigers gebruikt wordt om de kansen op blijven reizen te schatten. Hier is te zien dat de gemiddelde afwijking een stuk kleiner is dan bij het model. Bij het model was de gemiddelde

Buitendienststelling	Gemiddelde	St.dev.
Almere-Weesp	-6.9%	31.2%
Amersfoort-Utrecht	-6.4%	34.9%
Asd Sloterdijk-Haarlem	-13.0%	48.7%
Den Haag Centraal-Gouda	-1.1%	12.4%
Hoofddorp-Leiden	-2.8%	21.7%
Leiden-Alphen a/d Rijn	-1.1%	12.6%
Sittard-Maastricht	-8.2%	32.6%
Zaandam-Asd Sloterdijk	-9.9%	45%

Tabel 5.20: Percentages reizigers dat blijft reizen

afwijking  $-37\%$  terwijl nu de maximale gemiddelde afwijking  $13\%$  is. De standaardafwijking is nog wel wat aan de hoge kant, maar beter dan bij de resultaten uit het model.

## 5.5 Vuistregel

In deze paragraaf wordt de vuistregel van de NS op twee manieren vergeleken met de resultaten die gevonden zijn in dit onderzoek. In de eerste sectie wordt de vuistregel vergeleken met de resultaten die uit het buitendienststellingsmodel komen. In de andere sectie worden de gemiddelden die berekend zijn in paragraaf 5.4 met de vuistregel vergeleken.

### 5.5.1 Vuistregel en buitendienststellingsmodel

Uit het buitendienststellingsmodel kunnen verschillende kansen berekend worden door verschillende variabelen in te vullen. De variabelen die significant bleken te zijn waren:  $L_k$ ,  $Tr$  en  $Bus$ .  $Tr$  en  $Bus$  kunnen niet tegelijk één zijn, maar wel beiden nul. De verschillende mogelijkheden die er overblijven staan boven tabel 5.21. Ze zijn weergegeven als  $L_k/Tr/Bus$ , dus  $1/0/0$  betekent  $L_k = 1$  en  $Tr = Bus = 0$ . In de tabel staan veel verschillende

Type	1/0/0	1/1/0	1/0/1	0/0/0	0/1/0	0/0/1
Alles	0.42	0.81	0.23	0.75	0.95	0.55
Reg	0.61	0.89	0.34	0.61	0.89	0.34
Dal	0.40	0.80	0.22	0.75	0.95	0.56
Student	0.34	0.82	0.34	0.67	0.95	0.67

Tabel 5.21: Kansen van buitendienststellingsmodel

kansen, gecreëerd door de verschillende mogelijkheden en producttypes. Er moet nog wel eerst gerekend worden voor er een vergelijking gemaakt kan worden met de vuistregel. Bijvoorbeeld de eerste kolom van de tabel geeft de kans dat mensen die korte reizen maken, bij een buitendienststelling waarbij bussen en omgeleide intercity's als vervangend vervoer worden ingezet, blijven reizen. Nu missen dus de mensen die lange en middellange reizen



maken, deze staan in de vierde kolom. Deze kansen kunnen dus zo niet vergeleken worden met de vuistregel, omdat die alle reislengtes meeneemt. De verdeling is ongeveer 40% korte reizen en 60% lange en middellange. De twee kolommen kunnen samengevoegd worden tot één. Bijvoorbeeld

$$0.42 \cdot 0.40 + 0.75 \cdot 0.60 = 0.618$$

dus 61.8% van de reizigers blijft reizen als er treinen en bussen worden ingezet. Dit kan gedaan worden met alle kolommen wat resulteert in tabel 5.22. In deze tabel staan de kansen voor reizigers van alle reislengtes samen

Type	$Tr = 0$ en $Bus = 0$	$Tr = 1$ en $Bus = 0$	$Tr = 0$ en $Bus = 1$
Alles	0.62	0.89	0.42
Reg	0.61	0.89	0.34
Dal	0.61	0.89	0.43
Student	0.54	0.90	0.54

Tabel 5.22: Parameters model buitendienststellingen

genomen. Nu is duidelijk te zien dat de optie wanneer er alleen een bus rijdt het minst populair is. Wanneer er gewoon treinen blijven rijden alleen wat minder blijft bijna iedereen rijden. De vuistregel is dat 70% van de reizigers blijven reizen, dat is wat meer dan wat dit model zegt bij twee types van de buitendienststelling en wat minder dan wanneer er treinen blijven rijden. De vuistregel is vergeleken met dit model niet een hele goeie maat. Er blijkt meer verschil te zijn tussen de verschillende buitendienststellingen.

## 5.5.2 Vuistregel en data

In tabel 5.19 is te zien dat het percentage mensen dat blijft reizen erg varieert. Bij sommige buitendienststellingen blijven er veel mensen reizen en bij sommige buitendienststellingen maar weinig. Als het gemiddelde van die percentages genomen wordt komen we op 57.0% van de reizigers blijven reizen bij een buitendienststelling. Dit is lager dan de vuistregel. Wanneer de vuistregel gebruikt zou worden als voorspeller voor alle buitendienststellingen, dan is er een gemiddelde procentuele afwijking van  $-60.0\%$ , wat dus aangeeft dat de vuistregel vaak te hoog inschat. De standaarddeviatie is 97.3%, er is dus vaak veel afwijking tussen het daadwerkelijke aantal reizigers en wat de vuistregel aangeeft.

## 5.6 Afsluiting

In dit hoofdstuk staan de resultaten weergegeven van de modellen die gemaakt zijn om de kansen te berekenen dat mensen blijven reizen. Ook wordt

er een vergelijking gemaakt met de vuistregel. In het volgende hoofdstuk worden aan de hand van deze resultaten de conclusies beschreven die hieruit getrokken zijn.

## Hoofdstuk 6

# Conclusie en verder onderzoek

### 6.1 Inleiding

In dit hoofdstuk wordt in paragraaf 6.2 de conclusies beschreven die uit het onderzoek gehaald kunnen worden. In paragraaf 6.3 wordt beschreven wat gedaan kan worden als verder onderzoek.

### 6.2 Conclusie

Deze paragraaf is opgedeeld in twee stukken, het eerste stuk, paragraaf 6.2.1, geeft conclusies over het schatten van de aantallen reizigers. In paragraaf 6.2.2 staan de conclusies beschreven die getrokken kunnen worden uit het schatten van de kansen op het blijven reizen. Paragraaf 6.2.4 geeft de conclusie die getrokken kan worden uit het onderzoek van de vuistregel.

#### 6.2.1 Model aantal reizigers

De meest significante combinatie van variabelen is vaak  $\Delta^7 y_{t-7}$  en  $\Delta^7 y_{t+7}$ . Wat wil zeggen dat er het best naar het verschil met dezelfde dagsoort gekeken kan worden. Een andere verklarende variabele die vaak significant is, is  $\Delta^7 y_{vr}$ . Dit wil zeggen dat wat door de week gebeurd ook wat zegt over het aantal reizigers in het weekend. De variabelen voor het weer zijn niet vaak significant, blijktbaar wordt in deze gevallen daar niet veel van aangetrokken.

Als er gekeken wordt naar de kwaliteit van de schattingen per buitendienst-

stelling is te zien dat dit erg verschilt per buitendienststelling. De eerste drie buitendienststellingen, Amersfoort-Utrecht, Almere-Weesp en Amsterdam Sloterdijk-Haarlem hebben in de meeste gevallen een afwijking die kleiner is dan 20%. De anderen hebben een paar hoge afwijkingen ertussen staan. Deze afwijkingen zijn vaak bij de reguliere kaarten. Deze groep reizigers is erg lastig te voorspellen. Dit komt waarschijnlijk omdat het mensen zijn die niet regelmatig reizen, als ze regelmatig zouden reizen was er namelijk wel een abonnement aangeschaft. De lengte van de reis maakt niet uit voor de mate van voorspelbaarheid.

Concluderend is te zeggen dat het aantal reizigers niet heel makkelijk te voorspellen is. Dit model zal ook niet gebruikt worden in de praktijk, het was alleen nodig voor het model van de kansen om te blijven reizen.

### **6.2.2 Model buitendienststellingen**

Bij buitendienststellingen wordt er liever niet gereisd wanneer er bussen ingezet worden, bij intercity's blijven veel mensen gewoon reizen. Deze schijnbare weerstand kan ook veroorzaakt worden omdat mensen die met de NS-bussen gaan vaak uitchecken en daarom in dit onderzoek niet meegeteld worden. Daarnaast telt ook nog mee wat de lengte van de reis is die gemaakt wordt. Als er namelijk een buitendienststelling is op een korte reis dan wordt er weinig gereisd. Ook dit kan samen hangen met het busprobleem want bij korte reizen wordt vaker met de bus gereisd.

De kwaliteit van de voorspellingen ligt erg laag. In het beste geval is er een gemiddelde afwijking van  $-37.5\%$ . Dit model kan nu nog niet gebruikt worden in de praktijk, er moet namelijk nog wel wat aan verbeterd worden, zodat de voorspellingen wat beter worden. Toch is het idee van zo'n model wel goed. Het aantal reizigers dat minder reist is namelijk afhankelijk van de soort buitendienststelling en dus niet overal gelijk zoals de vuistregel zegt.

### **6.2.3 Model gemiddelden**

Door de gemiddelden kansen om te blijven reizen te gebruiken per buitendienststelling worden beter voorspellingen gemaakt dan wanneer het buitendienststellingen model gebruikt wordt. Dit geeft aan dat elke buitendienststelling een eigen invloed heeft op de reizigers. Voor verder onderzoek na het oplossen van het busprobleem moet dus gekeken worden wat de eigenschappen zijn die veel invloed hebben op de reizigers. Dit gemiddelden idee is nog niet meteen toepasbaar in de praktijk, maar geeft een beter beeld van het reisgedrag bij buitendienststellingen.

#### **6.2.4 Vuistregel van de NS**

De vuistregel van de NS zit niet heel erg ver van de werkelijkheid af, maar door voor elke buitendienststelling hetzelfde percentage te gebruiken kan het voorkomen dat er verkeerd geschat wordt wat het aantal mensen is dat blijft reizen. Er kan beter per buitendienststelling gekeken worden wat er in het verleden gebeurde.

### **6.3 Verder onderzoek**

Voor verder onderzoek kan gezocht worden naar meer verklarende variabelen om de voorspellingen te verbeteren. Reizigers aantallen kunnen nog afhangen van evenementen, speciale NS-acties, vakanties, grootschalige verstoringen. Ook kunnen er nog andere combinaties van producttypes gemaakt worden. Dit kan ervoor zorgen dat voorspellingen van reizigers aantallen verbeterd worden. Voor het schatten van de kansen zou er nog verder gedoken moeten worden in de buitendienststellingen. Er kunnen op deze manier nog meer eigenschappen van buitendienststellingen gevonden worden.

Ook zou er vanuit een andere kant gekeken kunnen worden, namelijk vanuit een groep reizigers die elk weekend reist. Vanuit hier kan ook bekeken worden wanneer die niet meer gaan reizen. Dit kan hele andere inzichten bieden in de keuze voor het niet meer reizen bij buitendienststellingen.

Voordat er echt verdere acties ondernomen kunnen worden is het erg van belang dat er uitgezocht gaat worden hoeveel mensen er uitgecheckt in de NS-bussen reizen. Hierdoor kan het namelijk zo zijn dat het blijkt dat er veel meer mensen blijven reizen dan dat er nu gedacht wordt. Er wordt dus ook sterk aangeraden om hier onderzoek naar te doen.

# Bibliografie

- [1] F. Cheung. Implementation of nationwide public transport smart card in the netherlands: Cost-benefit analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 1971(-1):127–132, 2006.
- [2] M. Bagchi and P. White. What role for smart-card data from bus systems? *Municipal Engineer*, 157(1):39–46, 2004.
- [3] John Paddington. Study on public transport smartcards - final report. Technical report, European Commission, 2011.
- [4] M.P. Pelletier, M. Trépanier, and C. Morency. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568, 2011.
- [5] B. Agard, C. Morency, and M. Trépanier. Mining public transport user behaviour from smart card data. In *12th IFAC Symposium on Information Control Problems in Manufacturing–INCOM 2006*, pages 17–19, 2006.
- [6] M. Bagchi and PR White. The potential of public transport smart card data. *Transport Policy*, 12(5):464–474, 2005.
- [7] Catherine Morency, Martin Trépanier, and Bruno Agard. Measuring transit use variability with smart-card data. *Elsevier*, 14(3):193–203, 2007.
- [8] Paul Bouman, Milan Lovric, Ting li, Evelien van der Hurk, Leo Kroon, and Peter Vervest. Recognizing demand patterns from smart card data for agent-based micro-simulation of public transport. BRON:Complexity in Public Transport (ComPuTr).

- [9] Evelien van der Hurk, Leo G. Kroon, Gábor Maróti, and Peter Vervest. Dynamic forecast model of time dependent passenger flows for disruption management. BRON:Complexity in Public Transport (ComPuTr).
- [10] Richard G. Clegg. Empirical studies on road traffic response to capacity reduction. In *Transportation and Traffic Theory*, 2007.
- [11] Shanjing Zhu, David Matthew Levison, Henry Liu, and Katheleen Harder. The traffic and behavioral effects of the i-35w mississippi river bridge colapse. *Transport Policy*, 44(10):771–784, 2010.
- [12] Xiaozheng He. *Modeling the Traffic Flow Evolution Process after a Network Disruption*. PhD thesis, UNIVERSITY OF MINNESOTA, Minneapolis, MN, 2010.
- [13] W.C. Chiang, R.A. Russell, and T.L. Urban. Forecasting ridership for a metropolitan transit authority. *Transportation Research Part A: Policy and Practice*, 45(7):696–705, 2011.
- [14] C.G. Chorus, T.A. Arentze, and H.J.P. Timmermans. A random regret-minimization model of travel choice. *Transportation Research Part B: Methodological*, 42(1):1–18, 2008.
- [15] D.S. Moore, G.P. McCabe, and V. TransVorm. *Statistiek in de praktijk*. Academic Service wetenschap en techniek, 1994.
- [16] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta. A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 697–711. Lyon, France, Elsevier Science, 1996.
- [17] Suzanne Kieft Anneke Dam. Toedelingsalgoritme prolop-model. BRON: NSR Marketingonderzoek & Advies.
- [18] C. Heij, P. De Boer, P.H. Franses, T. Kloek, H.K. Van Dijk, et al. *Econometric methods with applications in business and economics*. OUP Oxford, 2004.