

**Wat Big Data ons niet vertelt:
De waarde van enquêtes**

Wouter Kuhlman
TU Delft/Panteia BV
wouterkuhlman@gmail.com

Jan Kiel
Panteia BV
j.kiel@panteia.nl

**Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk
20 en 21 november 2014, Eindhoven**

Samenvatting

Wat Big Data ons niet vertelt: De waarde van enquêtes

De digitalisering van de maatschappij zorgt voor grote nieuwe datastromen (Big Data) die veel waardevolle informatie bevatten. Een voorbeeld hiervan zijn de OV-chipkaart data. Deze data bieden in potentie veel gebruiksmogelijkheden, bijvoorbeeld in transportmodellen. In de praktijk is dat echter lastiger dan vaak wordt gedacht. Dat komt onder andere door de beschikbaarheid van de data en de kwaliteit van de aanwezige variabelen. Zo ontbreekt bij de OV-chipkaart bijvoorbeeld informatie over het reismotief en de vervoerwijze van voor- en natransport. Omdat er ook andere bronnen van informatie zijn, roept dit de vraag op of de OV-chipkaart data wellicht verrijkt kunnen worden.

Deze paper presenteert een methode om OV-chipkaart data te verrijken met informatie uit de enquête onderzoeken WROOV en OViN. Het resultaat kan gebruikt worden om bijvoorbeeld HB matrices voor transportmodellen te maken. Het resultaat van de verrijking zijn matrices die zijn uitgesplitst naar vervoerwijze, dagdeel en motief. De hier gepresenteerde methode richt zich op het verbeteren van basismatrices voor BTM als hoofdtransportmiddel. Maar het biedt ook ideeën die op andere vlakken kunnen worden toegepast.

De OV-chipkaart data leveren halte-halte matrices die een goede beschrijving van het huidige reisgedrag vormen, maar nog niet toepasbaar zijn als basismatrices. Voor de conversie naar basismatrices worden modellen opgesteld om het motief en het voor- en natransport te schatten. Deze modellen volgen uit analyses op de WROOV data. Tevens wordt een vergelijking gemaakt met de OViN data om de modellen te kunnen actualiseren met correcties voor eventuele trends. De evaluatie van de methode vindt plaats aan de hand van een vergelijking met de resultaten van bestaande methoden.

Enkele eerste analyses hebben al interessante informatie opgeleverd. De verschillende motieven zijn bijvoorbeeld goed te onderscheiden naar de verplaatsingsafstand, het dagdeel en de stedelijkheidsgraad. Daarnaast is een methode getest om verplaatsingen toe te kennen aan zones op basis van invloedgebieden van haltes. Hiermee kan de conversie van halte-halte matrices naar HB matrices gemaakt worden.

Omdat Big Data over het algemeen niet optimaal zijn afgestemd op het inwinnen van informatie voor uiteenlopende doeleinden, is het verrijken van deze data met informatie uit andere bronnen een interessante oplossing. In de hier gepresenteerde case kunnen enquêtes als WROOV en OViN de benodigde achtergrondinformatie leveren voor het verrijken van OV-Chipkaart data, zodat deze Big Data beter benut kunnen worden.

1. Inleiding

1.1 Achtergrond

De introductie van de OV-chipkaart heeft gezorgd voor een enorme datastroom die het reisgedrag in het Nederlandse openbaar vervoer (OV) beschrijft. Elke check-in en check-out met de OV-chipkaart wordt opgeslagen in een database. In 2013 werden ruim 1.9 miljard transacties geregistreerd (Trans Link Systems, 2014). Deze Big Data¹ zijn onder andere interessant voor strategische transportmodellen waarin de basis bestaat uit herkomst-bestemming (HB) matrices die het huidige reisgedrag beschrijven (Pelletier, Trépanier, & Morency, 2011).

De beschikbaarheid van data over het reisgedrag was voor de vervoerders één van de redenen voor het invoeren van de OV-chipkaart (Bergmans, Bottenberg, & Hilferink, 2012). De OV-chipkaart data hebben de WROOV enquêtes vervangen voor zowel de verdeling van opbrengsten als informatiebron van het reisgedrag. Toch worden de data nog niet of nauwelijks gebruikt in strategische planning. Hiervoor zijn meerdere oorzaken aan te wijzen: de beschikbaarheid van de data is ontoereikend en bovendien bevatten ze niet alle benodigde informatie (Bagchi & White, 2005).

1.2 Doel

Deze paper tracht de mogelijkheden en beperkingen van Big Data te beschrijven aan de hand van een case: het gebruik van OV-chipkaart data in transportmodellen. Voor deze case wordt een methode gepresenteerd, waarmee de voordelen van zowel Big Data als enquêtes gecombineerd worden. Tevens presenteert de paper enkele eerste resultaten en een opzet voor het vervolg van deze studie.

1.3 Opzet van de paper

De opzet van deze paper is als volgt. Hoofdstuk 2 geeft een indicatie van de (on)mogelijkheden van het gebruik van OV-chipkaart data, waaronder de discrepantie tussen aanwezige en benodigde informatie. Hoofdstuk 3 beschrijft twee alternatieve bronnen voor reizigersinformatie over het OV: de WROOV enquêtes en de MON/OViN onderzoeken. Hoofdstuk 4 behandelt een methode om een toepasbaar resultaat te ontwikkelen uit een combinatie van bronnen. Enkele eerste resultaten van deze methode worden gepresenteerd in hoofdstuk 5. Hoofdstuk 6 bevat conclusies ten aanzien van het gebruik van Big Data en de geschetste methode. Tot slot wordt afgesloten met een discussie in hoofdstuk 7.

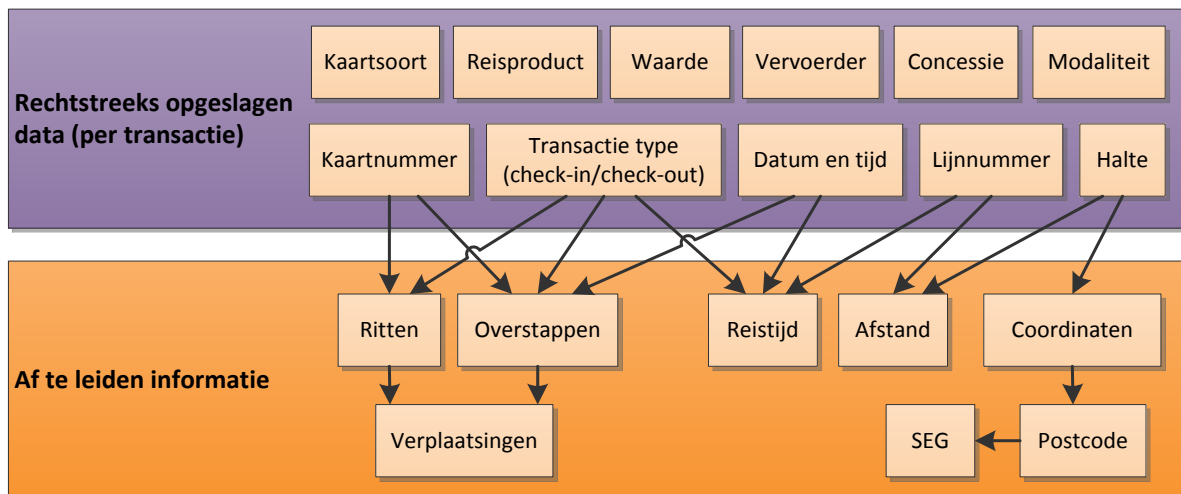
¹ Over de exacte definitie van Big Data bestaat nog geen consensus. Veelgenoemde eigenschappen zijn grote hoeveelheden, hoge snelheid waarmee de data binnen komt en grote verscheidenheid binnen de data. Deze eigenschappen maken de data lastig en te analyseren. OV-chipkaart data voldoen aan al deze eigenschappen.

2. OV-chipkaart data

2.1 Aanwezige informatie

Zonder diep in te gaan op de technische details van de OV-chipkaart data, is de informatie die verkrijgbaar is op te delen in rechtstreeks opgeslagen informatie en daaruit af te leiden informatie (zie Figuur 1).

In principe worden alle ritten gemaakt in het OV rechtstreeks opgeslagen. Er zijn enkele uitzonderingen: in sommige regio's zijn nog papieren kaartjes verkrijgbaar bij de chauffeur, daarnaast zorgen vergeten check-ins en check-outs voor incomplete ritten in de data. Begin 2013 was het aantal incomplete transacties bij alle vervoerders onder de 2% (TK, 2013/2014). De OV-chipkaart data bevatten dus een schat aan informatie, die door de hoge dekkingsgraad ook nauwkeurig is.



Figuur 1: Informatie in de OV-chipkaart data

2.2 Publieke beschikbaarheid

De beschikbaarheid van OV-chipkaart data voor onderzoek en beleid laat tot op heden te wensen over. Dit is toe te schrijven aan twee zaken: de privacy-gevoeligheid van de data en de structuur van het databeheer. De privacy-gevoeligheid van de data ontstaat doordat reispatronen van individuen zijn af te leiden wanneer de data over langere tijd geanalyseerd worden. Het is dan ook verboden om OV-chipkaart data langer te bewaren dan 18 maanden. Het College Bescherming Persoonsgegevens (CBP) ziet hier scherp op toe waardoor eigenaren erg behoedzaam zijn.

Het databeheer is op te delen in verschillende niveaus. Dit komt erop neer dat vervoerders alleen over hun eigen OV-chipkaart data beschikken. Overheden krijgen alleen sterk geaggregeerde data terug van de vervoerders, tenzij er specifiek om bepaalde data gevraagd is in de concessievoorwaarden. Translink Systems (TLS) slaat de data van alle vervoerders op in een centrale backoffice. TLS is voornamelijk echter zeer terughoudend in het verstrekken van data voor onderzoeksdoeleinden.

2.3 Benodigde informatie voor transportmodellen

De input voor transportmodellen omvat netwerkgegevens, sociaaleconomische gegevens (SEG's), basismatrices en tellingen. De basismatrices bevatten aantallen verplaatsingen per HB combinatie per modaliteit voor het recente verleden. De SEG's bevatten informatie over de bevolking, economie en geografische gegevens. Om de invloeden van SEG's beter te modelleren worden de basismatrices uitgesplitst naar dagdeel en motief (Rijkswaterstaat, 2012). Voor deze case ligt de focus op de basismatrices voor bus, tram en metro (BTM).

2.4 Discrepantie tussen aanwezige en benodigde informatie

De OV-chipkaart data zijn een ander type informatiebron dan de bronnen die de input leveren in de huidige modelopzet (zoals WROOV en OViN). Omdat de OV-chipkaart vrijwel alle ritten registreert, is het niet zomaar een telling van reizigers waarmee een basismatrix gekalibreerd kan worden: het geeft een goede beschrijving van de huidige situatie. Het inpassen van OV-chipkaart data in de modelopzet is echter gecompliceerd. Het geeft een nauwkeurig beeld van het huidige reisgedrag, maar er ontbreken ook enkele essentiële onderdelen.

Zoals hierboven uiteengezet, is het reismotief een belangrijke schakel bij het vormen van basismatrices. Deze informatie is niet aanwezig in OV-chipkaart data. Daarnaast ontbreekt er een onderdeel in het vertalen van de OV-chipkaart data naar HB matrices: HB matrices gaan uit van een herkomst en een bestemming, meestal het adres van een woning, kantoor of school, terwijl de OV-chipkaart data alleen verplaatsingen tussen haltes beschrijven. Hierin ontbreken het voor- en natransport van BTM verplaatsingen. De WROOV onderzoeken en MON/OViN onderzoeken zijn mogelijke bronnen voor het toevoegen van de ontbrekende informatie aan de OV-chipkaart data, om zo verrijkte matrices te krijgen.

Het invoeren van deze Big Data als bron voor transportmodellen is dus niet eenduidig en zal nauwkeurig moeten worden afgestemd op de modelopzet. De vraag is wat hiervoor de beste methode is, hoofdstuk 4 gaat hier verder op in.

3. Nederlandse mobiliteit enquêtes

3.1 WROOV

De WROOV onderzoeken (Werkgroep Reizigers Omvang en Omvang Verkopen) zijn gestart in 1984 in opdracht van het toenmalige ministerie van Verkeer en Waterstaat (nu Infrastructuur en Milieu, I&M), om de opbrengsten van de Nationale Vervoer Bewijzen (NVB) te verdelen over de vervoerbedrijven. Door middel van de WROOV onderzoeken werd een verdeelsleutel opgesteld om de opbrengsten te verdelen over de vervoerders, voornamelijk van BTM en enkele trajecten met de trein.

De meest recente WROOV onderzoeken bestonden uit een jaarlijkse enquête. Van 2003 t/m 2009 heeft dit geresulteerd in een jaarlijks aantal ritten en reizen met NVB, op basis

van 100.000 tot 150.000 ingevulde enquêtes per jaar. In 2010 is het onderzoek stopgezet vanwege de invoering van de OV-chipkaart.

De enquêtes leverden informatie over de herkomst, de bestemming, gebruikte haltes, het vertrektijdstip en het motief van de reis, alsook het geslacht en de leeftijd van de respondent. De resultaten van de enquêtes werden vervolgens gekoppeld aan hulpgegevens, zoals informatie over lijnen, exploitanten en geografische informatie.

De WROOV onderzoeken hebben een zeer grote dataset opgeleverd die veel informatie bevat voor het gebruik in transportmodellen. Een kanttekening is dat de WROOV onderzoeken alleen het reisgedrag voor de NVB dekken. Een deel van de verplaatsingen, gemaakt met kaartsoorten buiten de NVB om, valt dus buiten het onderzoek. Deze categorie bestaat uit Regionale Vervoerbewijzen (RVB), wagenverkoop en de Studenten OV-kaart (SOV). Ter indicatie: de NVB dekken ruim 70% van alle reizigerskilometers (Ballhaus, 2012). De WROOV data bevatten dus zeer waardevolle en betrouwbare informatie over het reisgedrag met BTM.

3.2 MON/OViN

OViN (Onderzoek Verplaatsingen in Nederland) is een jaarlijks onderzoek naar de mobiliteit in Nederland. Sinds 2010 heet het onderzoek OViN, daarvoor heeft het bestaan als MON (Mobiliteits Onderzoek Nederland, 2004 t/m 2009) en OVG (Onderzoek Verplaatsings Gedrag, 1978 t/m 2003). Het doel van deze onderzoeken is het geven van inzicht in de mobiliteit in Nederland voor het ministerie van I&M en andere beleids- en onderzoeksinstanties (Rijkswaterstaat, 2010) (CBS, 2013).

De MON/OViN onderzoeken dekken alle verplaatsingen van inwoners binnen Nederland, maar niet van buitenlanders in Nederland. Het onderzoek dekt verplaatsingen met alle vervoermodaliteiten. Het vervoer met BTM is hierdoor minder nadrukkelijk aanwezig in deze onderzoeken in vergelijking met de WROOV onderzoeken. Zowel MON als OViN is opgezet als een continu, dagelijks onderzoek naar de verplaatsingen in Nederland.

De enquêtes leveren gegevens op over de herkomst, de bestemming, het vertrek- en aankomsttijdstip en het motief van de door respondenten gemaakte verplaatsingen. Daarnaast wordt binnen zowel MON als OViN ook gevraagd naar persoonsgegevens zoals geslacht, leeftijd en beroep. De MON/OViN onderzoeken bevatten daarmee alle benodigde informatie voor het opzetten van basismatrices. De beperkende factor is het aantal waarnemingen voor BTM. Dit ligt een factor 50 lager dan in WROOV. Dit zorgt ervoor dat MON/OViN een goed beeld geeft van het reisgedrag op grote schaal, maar op kleinere schaal is de betrouwbaarheid van BTM onvoldoende. Daarnaast heeft de overgang van MON op OViN in 2010 geresulteerd in een methodebreuk en daarmee een trendbreuk (Wouters & Brakel, 2010).

3.3 Bruikbaarheid van deze enquête onderzoeken

Beide onderzoeken bevatten waardevolle informatie over de reizigers (geslacht, leeftijd) en de verplaatsingen (motief, voor- en natransport) die niet aanwezig is in de OV-chipkaart data. Ten aanzien van de bruikbaarheid hebben deze bronnen specifieke voor-

en nadelen. Het grote aantal waarnemingen is een groot voordeel van de WROOV data. De data is echter verouderd. Het OViN onderzoek geeft een mogelijkheid tot correcties voor een trend in het reisgedrag, maar het aantal waarnemingen is relatief laag.

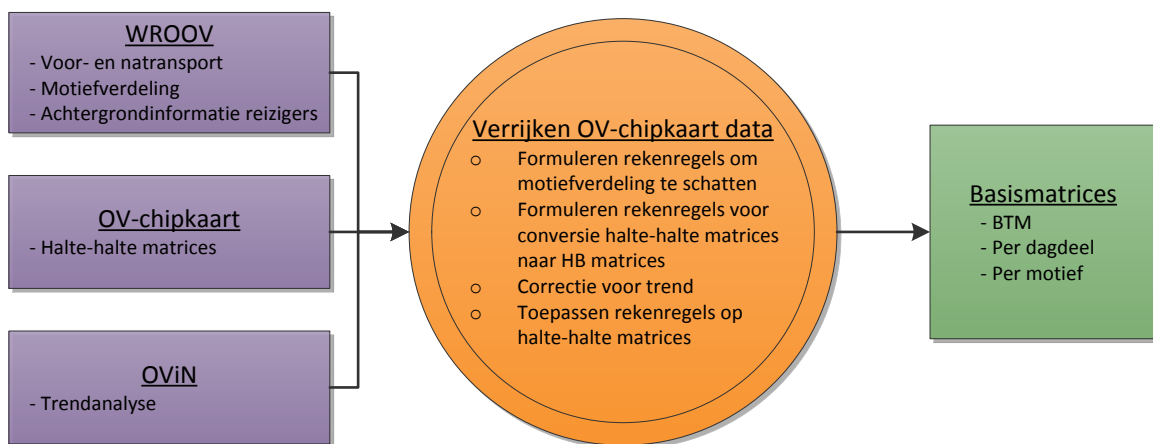
4. Methode voor implementatie van OV-chipkaart data in transportmodellen

De vraag is nu, of en hoe de bronnen kunnen worden gecombineerd. In dit hoofdstuk worden de stappen voor het combineren van bronnen uiteengezet. Van enkele onderdelen is de precieze invulling nog helemaal duidelijk, hiervoor zullen de verschillende opties benoemd worden als input voor discussie.

4.1 Welke informatie uit welke bron?

Het doel van deze methode is het combineren van OV-chipkaart data met WROOV en OViN data, voor het formuleren van basismatrices BTM. Zoals aangegeven in hoofdstuk 2 geven de OV-chipkaart data een goed beeld van het reisgedrag met BTM, maar ontbreekt een deel van de noodzakelijke informatie. De OV-chipkaart data kunnen voor dit doel verrijkt worden door deze te koppelen aan informatie uit enquêtes (zie Figuur 2). De WROOV onderzoeken leveren hier de beste data voor, terwijl OViN kan zorgen voor een continue evaluatie van die informatie.

De OV-chipkaart leveren in deze aanpak het aantal verplaatsingen BTM tussen haltes per dagdeel. De verplaatsingen worden vervolgens verdeeld over de in het model aanwezige motieven aan de hand van rekenregels. Deze rekenregels worden bepaald door middel van analyses op de WROOV data. Vervolgens worden de halte-halte matrices omgezet tot HB matrices door het voor- en natransport te schatten. Deze conversie geschiedt tevens op basis van de WROOV data.



Figuur 2: Concept onderzoek

De OV-chipkaart data kan per jaar geaggregeerd worden tot halte-halte matrices per dagdeel voor de gemiddelde werkdag. In deze geaggregeerde vorm zijn geen personen te herleiden en kunnen de matrices van elk jaar bewaard blijven. De WROOV onderzoeken raken steeds meer verouderd en de geldigheid van deze relaties zal dan ook getest moeten worden. Dit kan door een vergelijking met de reeks van MON/OViN data. Hiermee kan jaarlijks een actualisatie gemaakt worden door middel van correctiefactoren.

4.2 Sleutelvariabelen

De informatie uit WROOV dient omgezet te worden in rekenregels, die vervolgens zijn toe te passen op de OV-chipkaart data. De rekenregels worden opgesteld aan de hand van een aantal verklarende variabelen die ook aanwezig zijn in de OV-chipkaart data. De volgende sleutelvariabelen zijn in beide bronnen aanwezig:

- Haltes (plus afgeleide geografische informatie)
- Verplaatsingsafstand
- Tijdstip van instappen
- Kaartsoort

Daarnaast wordt een *permutatietest* toegepast om de bronnen te vergelijken ten aanzien van deze sleutelvariabelen. Het reisgedrag kan immers veranderd zijn in de tijd tussen het stopzetten van de WROOV onderzoeken en de huidige OV-chipkaart data. Het zou bijvoorbeeld zo kunnen zijn dat de gemiddelde reistijd veranderd is na de laatste WROOV onderzoeken. Afhankelijk van deze test kan de informatie rechtstreeks worden gekoppeld of dienen correctiefactoren toegepast te worden.

4.3 Uniformeren van data

Bij het koppelen van de verschillende bronnen dienen de data op elkaar afgestemd te worden. Een punt van aandacht hierbij is de analogie van definities. Het gaat dan onder meer om definities zoals ritten, reizen en verplaatsingen. Daarnaast moeten ook definities van verklarende variabelen zoals spitsuren op elkaar afgestemd worden.

De daadwerkelijke data-analyses worden toegepast op verplaatsingsniveau, waarbij een verplaatsing is gedefinieerd als een aaneenschakeling van ritten met BTM, zonder onderbreking van een treinreis of een activiteit. De data worden opgesplitst in BTM als hoofdtransport en BTM als voor- of natransport van een treinreis. Veel transportmodellen bevatten een aparte module voor het treinverkeer, inclusief het voor- en natransport. De focus voor het formuleren van basismatrices BTM ligt in deze case op BTM als hoofdtransport.

4.4 Schatten van relaties

Motiefverdeling

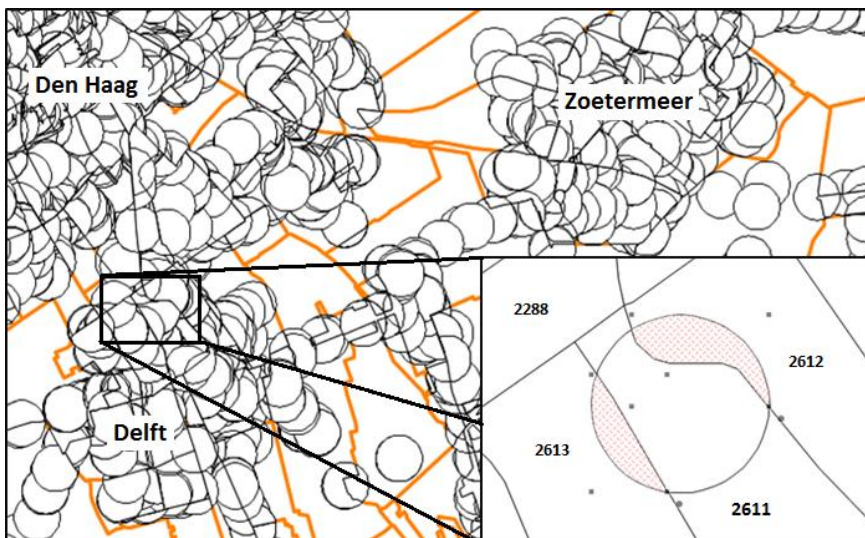
Voor het schatten van de motiefverdeling wordt de invloed van de sleutelvariabelen op de reismotieven onderzocht. Hiervoor zijn verschillende regressietechnieken mogelijk. De meest eenvoudige optie is een *multinomial logistic regression*. Deze methode identificeert significante verklarende variabelen en bepaalt weegfactoren voor deze variabelen. Voor elk motief wordt de kans bepaald dat een verplaatsing met dat motief is gemaakt. De kansverdeling over de motieven bepaalt hoe vaak elk motief wordt toegekend aan de verplaatsingen.

Wanneer de verklarende variabelen een grote onderlinge correlatie hebben kan dit de kwaliteit van de regressie verlagen. Een alternatief is dan om een *principal component analysis* toe te passen om variabelen te clusteren en daar vervolgens gewichten aan te

hangen met een *generalized linear model*. Beide methoden leveren een model op dat kan worden toegepast op de OV-chipkaart data om de verplaatsingen te verdelen over de motieven.

Voor- en natransport

Voor het schatten van de afstand van voor- en natransport wordt dezelfde methode toegepast als voor de motiefverdeling, met als verschil dat het hier om een continue variabele gaat in plaats van een categorische variabele. Voor de conversie naar HB matrices wordt een invloedgebied bepaald per halte, afhankelijk van de afgelegde afstand bij voor- en natransport. Deze afstanden zijn niet per definitie gelijk, bijvoorbeeld door de beschikbaarheid van een fiets bij voortransport van de BTM verplaatsing.



Figuur 3: Opdeling van invloedgebied halte in drie postcode-4 gebieden

De invloedgebieden worden vervolgens in een GIS programma over een indelingskaart, bijvoorbeeld van postcode gebieden, gelegd om de overlap met deze gebieden te bepalen. Reizigers kunnen vervolgens verdeeld worden over de onderliggende indeling op basis van de aandelen in oppervlak (zie Figuur 3).

Tevens is het mogelijk om SEG's te koppelen aan de indelingskaart om de toedeling aan gebieden niet op basis van oppervlak te verrichten maar bijvoorbeeld op basis van aanwezige arbeidsplaatsen. Er is veel onderzoek gedaan naar de invloed van SEG's op motieven. Op basis van literatuur kan een toedelingsmodel opgezet worden voor deze conversie.

4.5 Implementatie basismatrices in transportmodellen

De OV-chipkaart data leveren halte-halte matrices. Na het verrijken met voor- en natransport vormen ze HB-matrices, die worden opgesplitst naar motief en dagdeel. Deze set basismatrices dient als input voor een transportmodel. De nieuw opgestelde basismatrices kunnen vergeleken worden met basismatrices die zijn gebaseerd op bestaande methoden. Bijvoorbeeld, de methoden die in modellen zoals VENOM worden gebruikt. VENOM heeft wel basismatrices voor BTM als input, die vooral (maar niet

uitsluitend) gebaseerd zijn op enquêtes en tellingen. (Kieft, Herder, & Pieters, 2013). Een indicatie voor een verbeterde modelopzet is een vermindering van het aantal correcties en ophoogfactoren dat in de opzet gebruikt wordt.

4.6 Continuïteit

De kwaliteit van de schattingen op basis van WROOV data verminderen mogelijk doordat het onderzoek is gestopt. Om de continuïteit van deze methode te waarborgen kan daarom gecorrigeerd worden op basis van trends in de MON/OViN data. Globaal zijn deze data te vergelijken met de WROOV voor de periode 2003-2009. Wanneer een trend waarneembaar is voor een van de verklarende variabelen, dan kan deze gecorrigeerd worden door middel van correctiefactoren. Het is mogelijk om dit op jaarlijkse basis te doen aan de hand van OViN resultaten. In dat geval worden zowel de halte-halte matrices uit de OV-chipkaart data als het model om deze matrices te verrijken met de ontbrekende informatie jaarlijks geactualiseerd.

5. Eerste resultaten WROOV data analyses

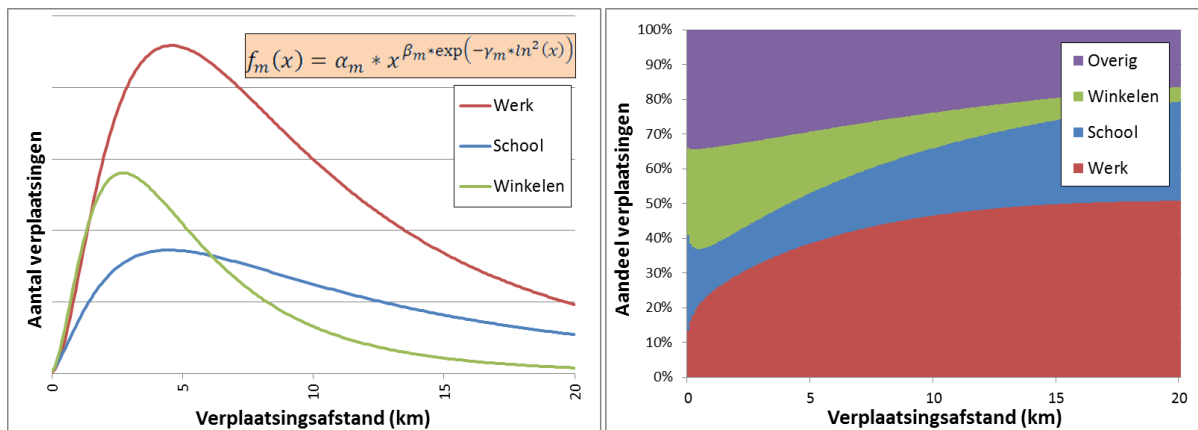
De eerste stappen van dit onderzoek zijn reeds gezet, waaronder enkele analyses op de WROOV data uit de periode 2003 t/m 2009. De invloed van enkele sleutelvariabelen op de motiefverdeling is onderzocht. Dit geeft hoopvolle resultaten voor het opstellen van rekenregels. Daarnaast is de methode om halte-halte matrices om te zetten naar HB matrices getest in een GIS programma en is een trendanalyse gedaan op de WROOV data. Dit hoofdstuk presenteert enkele eerste resultaten.

5.1 Motiefverdeling

Verplaatsingsafstand

De hypothese is dat reizigers voor sommige motieven bereid zijn verder te reizen als voor andere motieven en dat motieven daarom te onderscheiden zijn op basis van de afgelegde afstand. Om dit te onderzoeken is gekozen het aantal reizigers per afstandsklasse te schatten met een top-lognormaal verdeling. Deze methode wordt ook toegepast op de vervoerwijze keuze. Hierbij is de aanname gemaakt dat vervoermiddelen een bepaalde afstand kennen, waarop ze relatief vaker de voorkeur krijgen boven andere vervoermiddelen. Zo is de fiets populair voor korte afstanden en de trein voor lange afstanden. Voor elke verplaatsingsafstand wordt zo een kans berekend per vervoermiddel, om daarmee de reizigers over de vervoermiddelen te verdelen.

De schatting van de top-lognormaalverdeling is gedaan met de non-lineaire regressie module in SPSS. Hierbij zijn drie parameters geschat die elk een specifieke invloed hebben op de vorm van de verdeling. De eerste parameter (α) bepaalt de hoogte van de piek, de tweede (β) de locatie van de piek op de x-as en de derde (γ) de steilheid van de piek.



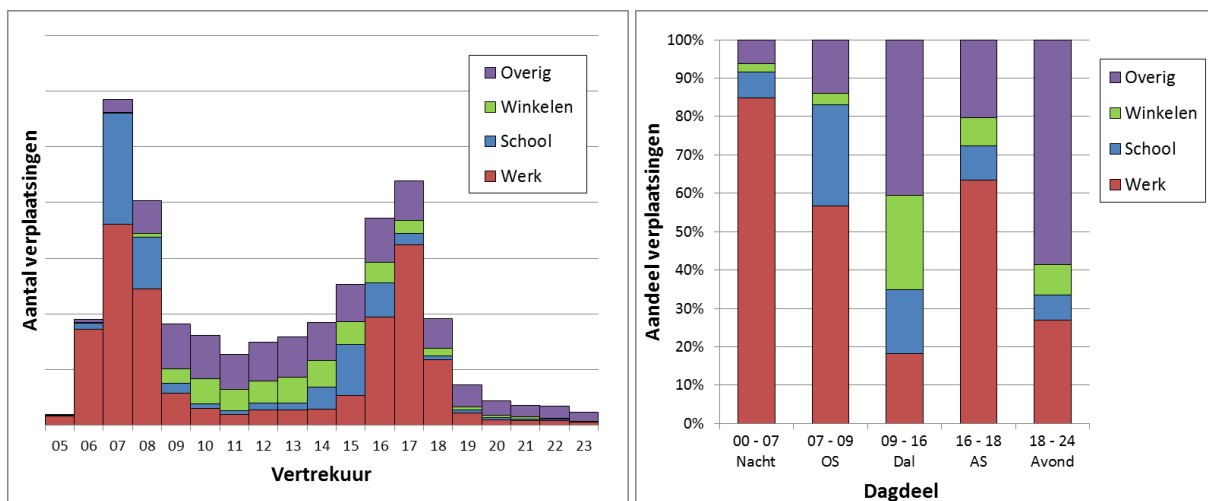
Figuur 4: Schatting aantal verplaatsingen (links) en aandeel verplaatsingen (rechts) over afstand

De geschatte benaderingsformules laten een duidelijk onderscheid zien tussen de drie meest voorkomende motieven: werk, winkelen en school. Het totale aantal verplaatsingen bestaat voor ruim 40% uit werkverplaatsingen, 20% met motief school en een kleine 15% met motief winkelen. De overige vijf motieven beslaan samen 25% van alle verplaatsingen en zijn daarom samengevoegd tot de categorie overig. Voor de motieven werk en school ligt de piek van het aantal verplaatsingen bij een langere verplaatsingsafstand dan voor de overige motieven (zie Figuur 4). Dit is te verklaren doordat reizigers eerder bereid zijn tot langere verplaatsingen bij deze reismotieven. Het motief winkelen, waaronder ook de dagelijkse boodschappen vallen, heeft een minder dwingend karakter. Een steile piek, op een korte verplaatsingsafstand, kenmerkt de formule voor het aantal verplaatsingen met dit motief. De formules resulteren voor korte verplaatsingen in een groot aandeel voor het motief winkelen en voor lange verplaatsingen grote aandelen voor werk en school.

Vertrektijdstip

In Figuur 5 komen de piekuren duidelijk naar voren. De motieven werk en school veroorzaken de piekuren, overige motieven zijn vooral aanwezig in de middag. Voor het motief werk zijn de ochtendspits en de avondspits redelijk symmetrisch. Voor het motief school is de ochtendspits duidelijk sterker en scherper. Veel scholieren reizen eerder terug dan de gebruikelijke avondspits. Daarnaast gaan ze ook meer gespreid terug, wat te verklaren is door de verschillende roosters van scholieren.

De meeste transportmodellen onderscheiden drie dagdelen: een ochtendspits (OS), een avondspits (AS) en de restdag. Binnen de restdag zijn verschillen zichtbaar tussen de nacht, dal en avond (zie Figuur 5). Het aantal verplaatsingen in de nacht is laag, daarom is het wellicht een betere optie om de nacht en de avond samen te voegen. De spitsperiodes bevatten voornamelijk verplaatsingen met het motief werk. Het aandeel van het motief school is in de ochtendspits groter dan in de avondspits. Verplaatsingen met het motief winkelen vinden grotendeels plaats in de dalperiode.

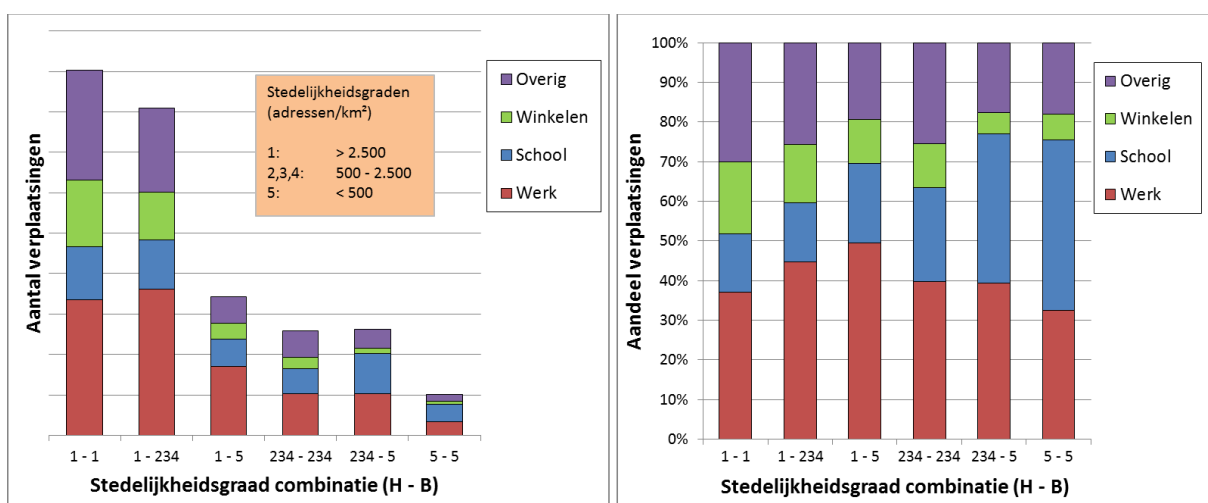


Figuur 5: Aantal verplaatsingen per uur (links) en motiefverdeling per dagdeel (rechts)

Stedelijkheidsgraad

Zowel de herkomst als de bestemming van een verplaatsing kan worden ingedeeld naar stedelijkheidsgraad. De stedelijkheidsgraad is een indeling naar 5 klassen van het aantal adressen per km², waarbij klasse 1 zeer sterk stedelijk is en klasse 5 niet stedelijk. De stedelijkheidsgraden 2, 3 en 4 zijn samengevoegd om het aantal mogelijke combinaties te verkleinen van 25 naar negen. Deze gebieden onderscheiden zich het minst van elkaar qua stedelijkheid. Daarnaast zijn verplaatsingen in tegengestelde richting (bijvoorbeeld van 1 naar 5 en van 5 naar 1) samengenomen. Dit resulteert in zes combinaties. Hierdoor is het aantal waarnemingen ook enigszins gelijk getrokken. De verhouding tussen het grootste aantal en het kleinste aantal verplaatsingen per categorie is afgenomen van een factor 60 naar een factor 9.

De motiefverdeling laat een verband zien over de verplaatsingen met een herkomst of bestemming met stedelijkheidsgraad 1. Uitgaande van een herkomst met stedelijkheidsgraad 1, neemt het aandeel verplaatsingen met motief werk toe naarmate de bestemming minder stedelijk wordt. Dit gaat voornamelijk ten koste van de aandelen winkelen en overig (zie Figuur 6).



Figuur 6: Aantal verplaatsingen (links) en motiefverdeling (rechts) per stedelijkheidsgraad

Voor de verplaatsingen met herkomst en bestemming in landelijke gebieden is een omgekeerde trend waarneembaar. Hiervoor neemt het aandeel verplaatsingen met motief werk af naarmate de verplaatsing zich bevindt in meer landelijk gebied. Het aandeel verplaatsingen met het motief school neemt daarentegen toe. Dit is waarschijnlijk gecorreleerd met de verplaatsingsafstand, die gemiddeld groter is in meer landelijke gebieden.

5.2 Voor- en natransport haltes

De gemiddelde hemelsbrede afstand ligt bij voortransport op 400 meter, bij natransport is dit 377 meter. Het aandeel verplaatsingen dat bij het voortransport van BTM een postcode-4 (PC4) grens overschrijdt ligt op 27%, voor natransport ligt dit hoger: 30%. Hier speelt mee dat heenreizen in vaker naar stedelijk gebied gaan, waar de PC4 gebieden kleiner zijn dan in landelijk gebied.

5.3 Trendanalyse

De onderzochte variabelen zijn op basis van WROOV per jaar geanalyseerd om mogelijke trends in de data te ontdekken. Eerst is gekeken naar de motiefverdeling in zijn geheel. Alleen voor het motief werk wijken de jaren 2003 en 2009 af. In 2003 is het percentage werk 8% lager dan het gemiddelde, in 2009 5% hoger. Voor 2009 is dit te verklaren met de invoering van reizen op saldo met de OV-chipkaart: de OV-chipkaart werd voornamelijk gebruikt voor meer incidentele verplaatsingen als vervanger van de strippenkaart. Het voornaamste reisproduct voor verplaatsingen met motief werk, het abonnement, werd nog wel toegerekend aan de WROOV onderzoeken. De afwijking in 2003 is op dit moment nog niet te verklaren.

Op grote schaal lijken de WROOV data per motief vrij constant te zijn, op enkele verschuivingen in 2009 na. Een vergelijking met de MON/OViN data kan meer duidelijkheid geven over trends in het reisgedrag, ook voor de periode na WROOV.

6. Conclusies

Beperkingen Big Data

De decentralisatie van OV beleid heeft geleid tot een versplintering van informatie. De OV-chipkaart data is minder goed beschikbaar omdat data verspreid zijn over overheden en vervoerders en de informatiedeling beperkt is. Daarnaast bemoeilijken de verschillende opslag- en verwerkingsmethodes bij verschillende partijen de verwerkbaarheid van de data. De interpretatie van grote aantallen variabelen vergt veel tijd, helemaal wanneer verschillende coderingen gebruikt worden. De bruikbaarheid van Big Data zoals de OV-chipkaart is om deze redenen voor verbetering vatbaar. Om de bruikbaarheid te verbeteren dient al tijdens het opzetten van een coderingsjabloon rekening gehouden te worden met de informatieverwerking.

Mogelijkheden enquêtes

De WROOV onderzoeken hebben een grote database opgeleverd die alle benodigde informatie bevat voor het formuleren van basismatrices in transportmodellen. Voorlopige analyses geven een goed beeld van de mogelijkheid om de informatie die ontbreekt in

OV-chipkaart data te schatten op basis van WROOV data. De WROOV onderzoeken zijn echter wel verouderd. Om de geldigheid van deze informatie te testen, en waar nodig te corrigeren, kunnen de WROOV data vergeleken worden met MON/OViN data. Deze onderzoeken omvatten een kleinere hoeveelheid data, maar vormen een doorlopende reeks. De combinatie van WROOV en MON/OViN enquêtes met OV-chipkaart data bevat alle benodigdheden voor het formuleren van basismatrices die geschikt zijn voor transportmodellen.

Toepassing in transportmodellen

Basismatrices geformuleerd op basis van OV-chipkaart data kunnen de huidige situatie beter beschrijven dan synthetische basismatrices gekalibreerd met tellingen. In hoeverre deze methode de huidige modellering van BTM in transportmodellen kan verbeteren is op dit moment nog niet duidelijk. Het onderzoek is in volle gang. De volgende stap omvat het samenstellen van modellen op basis van WROOV data. Daarna volgt het toepassen van deze modellen op de OV-chipkaart data om basismatrices te vormen. Vervolgens dienen de basismatrices ingevoerd te worden in het model om de resultaten te evalueren.

7. Discussie

Publieke beschikbaarheid data

De publieke beschikbaarheid van OV-chipkaart data laat op dit moment te wensen over. Vervoerders leveren alleen sterk geaggregeerde data aan opdracht gevende overheden, aan de hand van het MIPOV (KpVV, 2008) (KpVV, 2011) format, terwijl geargumenteed kan worden dat de overheden als opdrachtgever eigenaar zijn van de data. Alleen wanneer specifiek genoemd in de concessievoorwaarden moeten vervoerders meer data leveren. Dit is nu zelden het geval. Overheden zouden benodigde informatie voor strategische planning van OV beschikbaar moeten hebben.

Uniformiteit van data

De open structuur van het OV-chipkaart systeem zorgt voor verschillende verwerkingsmethoden van de data. Dit komt de uniformiteit van de data niet ten goede en bemoeilijkt de interpretatie. Eén uniforme methode van coderen en verwerken van data zou de bruikbaarheid van de OV-chipkaart data voor onderzoek bevorderen.

Continuïteit van onderzoeken

Na het stopzetten van de WROOV onderzoeken is geen vergelijkbare opvolger als informatiebron opgezet. De overgang van MON op OViN viel gelijk met het einde van de WROOV onderzoeken, waardoor nu alleen het OViN onderzoek doorloopt, dat een methodebreuk met MON bevat. Het OViN zou nu uitgebreid kunnen worden om extra informatie over het OV te verkrijgen. Een alternatief is een apart onderzoek om informatie te koppelen aan OV-chipkaart data, bijvoorbeeld door reizigers online een motief op te laten geven bij gemaakte reizen in hun chipkaart overzicht. Dit is echter een gevoelige kwestie vanwege de privacy-gevoeligheid en online data beveiliging.

Vervolgonderzoek

De kwaliteit van de modellen hangt af van het aantal significante variabelen. Wanneer dit er weinig zijn gaat dit ten koste van de voorspellende waarde, wanneer er veel variabelen significant zijn kan er ook een schijnnaauwkeurigheid gecreëerd worden door *overfitting* op de WROOV dataset. De balans hiertussen is lastig te controleren, aangezien er geen rechtstreekse controlemogelijkheden zijn voor de resultaten van deze methode.

Literatuur

- Bagchi, M., & White, P. (2005). The potential of public transport smart card data. *Transport Policy*(12), pp. 464-474.
- Ballhaus, F. (2012). *Ontwikkeling reizigerskilometers*. Zoetermeer: NEA.
- Bergmans, j., Bottenberg, W., & Hilferink, P. (2012). *WROOV evaluatie*. Ministerie van Infrastructuur en Milieu.
- CBS. (2013). *Onderzoek Verplaatsingen in Nederland 2012: Onderzoeksbeschrijving*. Den Haag: Rijkswaterstaat.
- Kieft, S., Herder, J., & Pieters, M. (2013). Openbaar Vervoer Matrices in VENOM. *Colloquium Vervoersplanologisch Speurwerk (CVS)*. Rotterdam.
- KpVV. (2008). *Model Informatieprofiel Openbaar Vervoer(MIPOV)* . Rotterdam: Kennisplatform Verkeer en Vervoer.
- KpVV. (2011). *Model Informatieprofiel Openbaar Vervoer (MIPOV): Technische uitwerking MIPOV 2008 m.b.t. OV-chipkaart*. Rotterdam: Kennisplatform Verkeer en Vervoer.
- Pelletier, M., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C*(19), pp. 557-568.
- Rijkswaterstaat. (2010). *Mobiliteitsonderzoek Nederland 2009 Het onderzoek*. Ministerie van Verkeer en Waterstaat.
- Rijkswaterstaat. (2012). *Documentatie Groeimodel 2011 Deel 1*. Rijkswaterstaat.
- Tavasszy, L., Snelder, M., Duijnsveld, M., Haaijer, R., Meurs, H., Nes, R. V., . . . Jansen, B. (2012). *Audit LMS en NRM Syntheserapport*. Delft: TNO.
- TK. (2013/2014). 8e Voortgangsbrief Aanvalsplan OV Chipkaart. *bijlage 251861 Percentage incomplete transacties*. Tweede Kamer, vergaderjaar 2013/2014: Kamerstuk 23645, nr. 546.
- Trans Link Systems. (2014). *Feiten & cijfers - Trans Link Systems*. Retrieved 08 2014, from [www.translink.nl: https://www.translink.nl/nl-NL/Persinformatie/Feiten-cijfers](https://www.translink.nl/nl-NL/Persinformatie/Feiten-cijfers)
- Wouters, B., & Brakel, J. v. (2010). Simulatiestudie naar Methodebreuken in het Onderzoek Verplaatsingen in Nederland. *Colloquium Vervoersplanologisch Speurwerk (CVS)*. Roermond.