

**Innovatieve toepassingen van OV chipkaartdata:  
buiten de lijntjes van data-gedreven OV onderzoek**

dr. ir. Niels van Oort  
Goudappel Coffeng/ TU Delft  
[NvOort@Goudappel.nl](mailto:NvOort@Goudappel.nl)

drs. Martijn Ebben  
Goudappel Coffeng  
[MEbben@Goudappel.nl](mailto:MEbben@Goudappel.nl)

Peter Kant Msc  
DAT.Mobility  
[pkant@dat.nl](mailto:pkant@dat.nl)

**Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk  
19 en 20 november 2015, Antwerpen**

## **Samenvatting**

### *Innovatieve toepassingen van OV chipkaartdata: buiten de lijntjes van data-gedreven OV onderzoek*

Afgelopen CVS congressen is er veel gesproken over nieuwe databronnen die helpen bij de uitdagingen in de OV wereld. De OV chipkaart is één van de bronnen, waarmee we het OV beter en efficiënter kunnen maken. Maar tot nog toe gebruikten we deze data vooral om binnen de lijntjes te kleuren: de data als vervanging van eerdere handmatig verkregen data. In dit paper gaan we een stap verder. Aan de hand van drie innovatieve cases laten we zien dat er ook buiten de lijntjes te kleuren valt.

Met OV chipkaart data stelden wij een OV-model op voor Den Haag voor korte termijn prognoses, zoals vorig jaar gepresenteerd op het CVS. Dit is de basis geweest voor de drie cases:

De vraag voor eerste case was: zijn elasticiteitsparameters af te leiden uit revealed preference data voor verschillende praktijksituaties? Wij merken dat dit goed mogelijk is. En dat het gedrag van reizigers verschilt per context: reizigers reageren heftiger op 'tijdelijk ongemak' dan in een vergelijkbare structurele situatie. De elasticiteitsparameter kan tot 25% hoger liggen.

Ook kijken wij naar een belangrijk, maar vaak in modellen genegeerd aspect van reisbeleving: comfort. Voor de regio Den Haag nemen wij expliciet comfort op in de (model) kostenfunctie door rekening te houden met de capaciteit van voertuigen. De bestaande vraag leiden wij direct af uit OV chipkaartgegevens. Onze studieresultaten tonen aan dat het niet beschouwen van capaciteit en comfort kan leiden tot een onderschatting van de vervoerwaarde-effecten tot 30%. We laten ook zien dat deze aanpak kan worden toegepast in de praktijk: de rekentijd is kort en het leidt tot een betere vraagraming van openbaar vervoer.

Tot slot kijken we naar de bruikbaarheid en inzet van andere databronnen. Als pilot hebben we een vergelijkende analyse tussen OV chipkaart- en GSM data uitgevoerd voor de regio Emmen. We tonen aan dat de GSM data aanvullend is: deze is namelijk ook bruikbaar voor analyse van de niet-ov-reizigers. Tot slot laten we zien dat het combineren van de twee databronnen inzicht verschaft in de potentie voor OV op specifieke HB relaties. Zo benoemen wij een aantal relaties in de regio Emmen waar op basis van de data het OV gebruik (vooralsnog) achter blijft en dus potentie heeft.

Alle drie de cases laten innovatie zien op onderzoek en toepassing van OV chipkaartdata. Wij gaan door met buiten deze lijntjes te kleuren voor een beter en efficiënter OV!

## 1. Inleiding

Het openbaar vervoer ziet zich gesteld voor vele uitdagingen. Het is een redelijk conservatieve bedrijfstak in een tijd van snelle verandering. "De reiziger" bestaat niet meer. Het gaat om een persoonlijke beleving en die kan voor iedereen anders zijn. Reizigers stellen daarbij steeds hogere eisen aan de kwaliteit van het vervoer. De reiziger "reist" daarbij ook digitaal met smartcards en apps en zo komen er in rap tempo nieuwe databronnen beschikbaar. Tegelijkertijd is er steeds meer aandacht voor kosteneffectiviteit door afnemende overheidsbudgetten voor openbaar vervoer. Op het gebied van kwaliteitsverbeteringen winnen robuustheid en betrouwbaarheid aan aandacht (zie bijv. Cats et al. 2015, Yap et al. 2015, Van Oort en Lee et al. 2014).

Het openbaar vervoer is in toenemende mate uitgerust met systemen die automatisch en continue data verzamelen over de uitvoering en het gebruik van het vervoer. Deze data kan de bedrijfstak helpen om goed om te gaan met de genoemde uitdagingen:

- GPS-data uit voertuigen is al lange tijd beschikbaar (e.g. Van Oort et al. 2013, Van Leusden et al. 2012, Hickman 2004) en biedt, in combinatie met de dienstregeling, inzicht in stiptheid, betrouwbaarheid en reistijden.
- De beschikbaarheid van databronnen zoals GSM-data (e.g. Van der Mede 2015, Calabrese et al. 2011) kan meer inzicht geven in reispatronen, reisbehoeften en daarmee de potentie van nieuwe verbindingen.
- Recent is ook veel meer data over het gebruik van het openbaar vervoer beschikbaar gekomen (via OV-chipkaart en automatische telsystemen; e.g. Neema et al. 2015, Van Oort 2014, Pelletier et. al 2011,).

Deze databronnen ondersteunen het ontwerp van het vervoersysteem en de besluitvorming over maatregelen, met als einddoel efficiënt vervoer van hoge kwaliteit. De beschikbare data kan worden gebruikt om de behoeften van reizigers beter te doorgronden, de prestaties van het systeem nauwkeurig in beeld te brengen en de kansen voor verbeteringen vast te stellen. Bovendien stelt de data planners in staat om de kosten van verstoringen in de dienstuitvoering te bepalen en om de potentiële besparing van verbeteringen te becijferen. Deze kosten en besparingen kunnen worden opgenomen in kosten-batenanalyses ( Van Oort en Van Leusden 2015) om daarmee het besluitvormingsproces te ondersteunen.

In dit paper gaan we een stap verder dan de traditionele datatoepassingen. We laten zien dat er ook buiten de lijntjes te kleuren valt. De OV chipkaart biedt ons nieuwe mogelijkheden voor onderzoek; onderzoek dat we voorheen niet deden, maar wel een grote bijdrage in OV planning en uitvoering kan leveren. Met dit paper hopen we u te inspireren om niet alleen binnen de lijntjes, maar vooral ook daarbuiten te gaan kleuren: daar liggen de echte innovaties!

Dit paper laat drie voorbeelden zien van nieuw, innovatief onderzoek, voortbouwend op onze CVS bijdrage in 2014 (Een verkeersmodel voor de korte termijn met OV-chipkaartdata als basis ( Van Oort et al. 2014)). Het gaat om de volgende cases:

- 1) Schatten van gedrag bij omleidingen voor betere vervoerwaardeprognoses;
- 2) Meenemen van comfort en capaciteit bij vervoerwaardeprognoses;
- 3) Datafusie van chipkaart- en GSM-data om kansen en bedreigingen voor het OV gedetailleerd in kaart te brengen.

Deze voorbeelden worden uitgebreid behandeld in hoofdstuk 3. Eerst staan wij in hoofdstuk 2 stil bij ontwikkelingen op het gebied van Big Data in de OV sector.

## 2. Big data in openbaar vervoer

De inzet van "big data" in de vervoerssector neemt toe. Verschillende ontwikkelingen zorgen voor een toename aan data. Enkele voorbeelden zijn:

- Een bus uitgerust met GPS zendt zijn positie om de 15 seconden uit wat resulteert in circa 3.000 positierecords per dag.
- In groot stedelijke gebieden zijn dagelijks miljoenen smartcardhandelingen.
- Video analyse technieken maken het mogelijk om automatisch passanten en passagiers te tellen. Dit maakt bewakingscamera's op stations en in voertuigen een interessante databron.
- Social media activiteiten van reizigers kunnen inzichten opleveren in het gebruik van het openbaar vervoer (zie bijv. Bregman 2012).
- De digitalisering van de maatschappij resulteert in meer mensen met een smartphone en bredere aanwezigheid van netwerktechnieken als Bluetooth en Wi-Fi in de openbare ruimte, waaronder ook in voertuigen. Dit maakt onder andere de analyse van voetgangersstromen mogelijk (zie bijv. Van den Heuvel et al. 2015).
- Sensoren die verbonden zijn aan infrastructuuronderdelen, zoals seinen en wissels, maken het mogelijk om onderhoudsschema's te optimaliseren.

### *Toepasbaarheid van big data in relatie tot privacy discussie*

Dat big data bronnen bestaan, betekent nog niet direct dat dit tot relevant inzicht leidt. Op wettelijk en maatschappelijk vlak lopen discussies rond privacy die van invloed zijn op de inzetbaarheid van databronnen. Dit creëert een spanningsveld tussen het belang van analyse (meer data op een fijner niveau is doorgaans beter) en de reiziger (anoniem blijven). Juist dit spanningsveld dwingt ons als onderzoekers te kijken naar de inzetbaarheid van data in geaggregeerde vorm: hoe kunnen wij vanuit samengevoegde datasets toch inzichten afleiden en uitspraken doen over prestaties en kansen in het OV systeem? Dit doen wij onder meer door het opstellen van algoritmen en toepassen van analysetechnieken. In dit paper kijken we naar de mogelijkheden van (geaggregeerde) OV chipkaartdata, al dan niet in combinatie met andere databronnen.

### *Eerdere toepassingen met OV chipkaart gegevens*

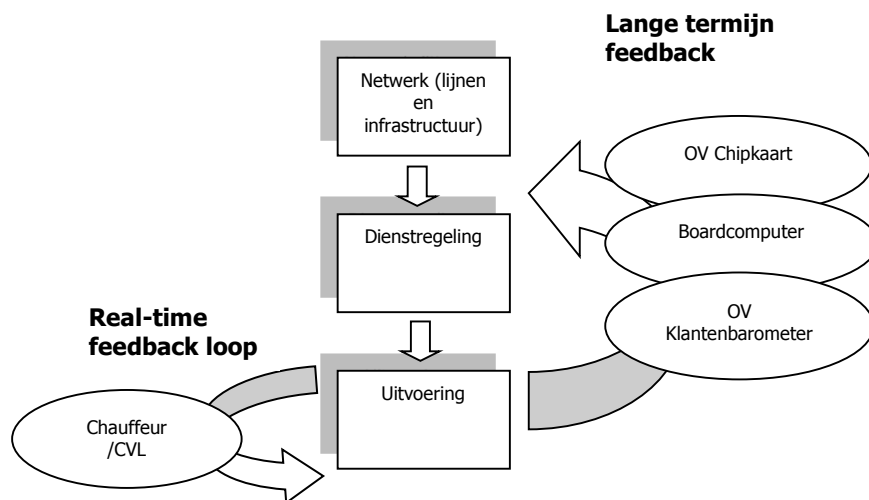
Eerder onderzoek laat de belangrijkste voordelen zien van smartcards voor wat betreft data-analyse (Bagchi en White 2005). Zij stellen vast dat met de grote beschikbare hoeveelheden data de volgende inzichten in reisgedrag kunnen worden afgeleid:

- Trends in reisgedrag over de tijd;
- Soort OV-gebruik van klanten, zoals onderverdeling naar reisfrequentie;
- Belangrijkste herkomsten en bestemmingen.

Het aantal gebieden waar de smartcard wordt toegepast en wordt geanalyseerd neemt snel toe. Voorbeelden hiervan zijn Londen (Oyster card), Hong Kong (Octopus card), Seoul, Beijing, Santiago, Shenzhen en Brisbane (Kurauchi et al. 2014, Ma et al. 2013, Munizaga en Palma 2012, Park et al. 2008, Wang et al. 2011). Ook Nederland staat internationaal op de kaart voor wat betreft smartcard data analyse (Van Oort 2014).

### *Inzet van big data voor het managen van het OV*

Het planningsproces rondom openbaar vervoer bestaat uit strategisch, tactisch, operationeel en controle beslissingen (zie figuur 1). Big data kan zowel de off-line processen tijdens deze planningsfasen als real-time tijdens de exploitatie bevorderen.



Figuur 1. Planning en exploitatie van OV

Pelletier et al. (2011) en Van Oort (2014) laten verschillende internationale toepassingen zien, van chipkaartdata in bovenstaande processen, waarbij de focus ligt op het schatten van de HB-matrix (bijv. Munizaga en Palma 2012), het afleiden van routekeuzegedrag (bijv. Schmöcker et. al 2013) en overstapanalyses (bijv. Seaborn et. al 2009). Het inzetten van historische data voor het beter voorspellen van toekomstige ontwikkelingen, wordt nog beperkt onderzocht.

### 3. Nieuwe toepassingen op basis van big data

#### 3.1 Een verkeersmodel voor de korte termijn met OV-chipkaartdata als basis

Voor het voorspellen van het aantal openbaar-voerreizigers in nieuwe of toekomstige situaties wordt veelal gebruik gemaakt van verkeersmodellen. Dergelijke modellen berekenen vaak de vraag in generatie- en distributiemodellen op basis van zonale kenmerken en weerstanden. De berekende vraag (een HB-matrix) wordt vervolgens toegevoerd aan het netwerk.

Door de introductie van de OV-Chipkaart is het huidige gebruik van openbaar vervoer bekend op een hoog detailniveau. Hierdoor is het mogelijk om direct de vraagmatrix af te leiden uit de OV-Chipkaartdata. Dit blijkt een ideale basis voor korte termijn prognoses, zoals we eerder uitgebreid beschreven (Van Oort et al. (2014)). Door het toevoegen van een zone bij elke halte(groep) is een halte-halte matrix direct een herkomst-bestemmingsmatrix. Daar komt bij dat het OV-netwerk inclusief de volledige dienstregeling vrij beschikbaar is als open data. Het resultaat van de toedeling bij deze aanpak is een basissituatie van een verkeersmodel die perfect aansluit bij de waarnemingen op straat.

Op basis van de HB-matrix zoals waargenomen in de OV-chipkaart is het mogelijk korte termijn voorspellingen te doen. Het is bijvoorbeeld mogelijk om de netwerk-effecten in kaart te brengen als gevolg van een verandering in frequentie op een lijn, de verandering van een route van een lijn, het introduceren van nieuwe routes of het verhogen van de snelheid van een lijn (bijvoorbeeld door doorstromingsmaatregelen). Dit kunnen zowel tijdelijke als permanente maatregelen zijn. Het model (Van Oort et al. (2014)) rekent de kwaliteit van een reis om naar gegeneraliseerde kosten, zowel voor het huidige vervoeraanbod als voor de situatie met het gewijzigde vervoeraanbod. Door middel van elasticiteiten wordt het effect van het gewijzigde vervoeraanbod berekend. We hebben inmiddels voortgebouwd op deze basis, zoals we in de volgende paragrafen laten zien.

### 3.2 Schatten van gedrag bij omleidingen voor betere vervoerwaardeprognoses

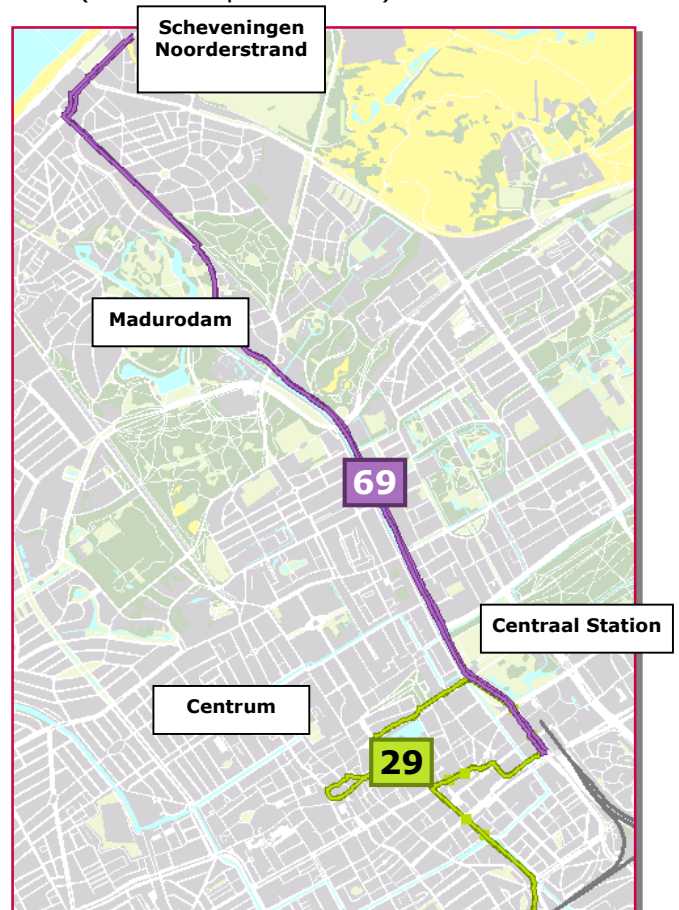
Om het keuzegedrag van reizigers met prognosemodellen goed te voorspellen, maken modellen, zoals hiervoor geïntroduceerd, gebruik van specifieke parameters. Deze parameters zijn tot nu toe veelal gebaseerd op enquête-onderzoek ("stated preference") of ervaring uit de praktijk. Bij enquête-onderzoek wordt reizigers gevraagd welke keuze zij zouden maken in een bepaalde situatie, gegeven een vaste set opties. Bij ervaring uit de praktijk wordt er gebruik gemaakt van het inzicht van experts die op basis van eerder voorgedane situaties hun voorspellingen bepalen. De validatie van de resultaten van dit type onderzoek is echter moeilijk; mensen gedragen zich soms anders dan men in een enquête aangeeft. De komst van de OV-chipkaart biedt hier meerwaarde: we weten nu exact welke reizen er daadwerkelijk gemaakt worden ("revealed preference").

Deze basis hebben we gebruikt voor de bepaling van gedragsparameters in het elasticiteitenmodel van de HTM in Den Haag. Het onderzoek spitste zich toe op het reisgedrag bij tijdelijke, langdurige onderbrekingen in het netwerk, met als doel om de effecten van tijdelijke wijzigingen in het netwerk in het vervolg beter te kunnen voorspellen. Daar is momenteel weinig over bekend.

Voor deze praktijkstudie is er voor een grootschalige, langdurige onderbreking van één lijn gekozen: de onderbreking van tramlijn 9 in Den Haag in verband met werkzaamheden aan de Koninginnegracht in het najaar van 2014 (zie ook figuur 2).

Deze onderbreking had de volgende wijzigingen in het netwerk tot gevolg:

- Tramlijn 9 Vrederust – Scheveningen werd tijdelijk opgeknipt in een tramlijn Vrederust – Centraal Station – centrum (lijn 29) en een buslijn Centraal Station – Scheveningen (lijn 69).
- Bij het Centraal Station sloten de lijnen op elkaar aan.



Figuur 2. Selectie lijnen Den Haag tijdens werkzaamheden

Voor de reguliere- en de verstoringssituatie is chipkaartdata geanalyseerd met statistische en visuele technieken. Uit dit onderzoek blijkt dat de elasticiteitsparameter voor de twee situaties afwijkt: de waarde van de elasticiteit voor omleidingen ligt circa 25% hoger, wat impliceert dat de reizigers heftiger reageren op tijdelijke aanpassingen dan op structurele veranderingen. We raden dan ook aan om in modellen bij langdurige tijdelijke omleidingen een hogere absolute elasticiteitswaarde te gebruiken.

### 3.3 Meenemen van comfort en capaciteit bij vervoerwaardeprognoses

In stedelijke OV-systemen nemen de problemen rondom comfort en capaciteit toe. Vollere voertuigen hebben een nadelig effect op de comfort beleving van reizigers. In huidige vervoermodellen wordt het aspect comfort niet of nauwelijks meegenomen, terwijl dit de keuze van de reiziger wél beïnvloedt. Zowel voor de vraagrading als bij het doorrekenen van toekomstsituaties, bijvoorbeeld ten gevolge van stremmingen en

omleidingen, wordt aangeraden wel rekening te houden met comfort en capaciteit. De OV chipkaartdata biedt de unieke kans om hier grote stappen in te zetten.

### Bestaande literatuur

In de bestaande literatuur zijn er verschillende manieren beschreven om reizigers toe te delen aan een OV-netwerk rekening houdend met de vervoerscapaciteit. Cepeda et al. (2006) zet een harde capaciteitsbeperking op het OV-systeem, met als gevolg dat een reiziger een omweg maakt zodra de maximale capaciteit wordt bereikt. Florian (2002) maakt gebruik van een crowdingfunctie waarbij link kosten afhankelijk zijn van de vervoersstroom. Net als in een statische auto-toedeling betekent dit in de praktijk dat de capaciteit van de links onbeperkt blijft, maar dat de kosten sterk toenemen zodra de vervoersstroom in de buurt van of zelfs over de capaciteit gaat. Pel et al. (2014) gaan een stap verder door een crowdingfunctie te introduceren zowel op het instappen als tijdens de rit in het voertuig. Schmöcker et al. (2011) gebruiken een soortgelijke aanpak door de mogelijkheid te introduceren dat een reiziger geen zitplaats kan vinden. Pel et al. (2014), Schmöcker et al. (2011) en Cepeda et al. (2006) passen hun werkwijze toe in casestudies van bestaande situaties, maar zonder een groot netwerk te beschouwen.

### Onze methodiek

Onze methodiek gebruikt een crowdingfunctie, waarin de zitplaatscapaciteit en de maximale capaciteit van een voertuig worden meegenomen. Het doel van deze aanpak is snelle, realistische voorspellingen te maken van toekomstige OV-gebruik, rekening houdend met de beschikbare capaciteit.

In onze aanpak wordt gerekend met een gegeneraliseerde kostenfunctie waarin verschillende componenten worden meegewogen, bijvoorbeeld wachttijd en reisafstand. De crowdingfunctie heeft invloed op de waardering van de reistijd in het voertuig en houdt daarbij specifiek rekening met de zitplaats- en maximale capaciteit voor iedere lijn, periode van de dag en in te zetten voertuigtype. De waarde van de crowdingfunctie (Crowdfactor) is als volgt van invloed op de reistijdwaardering:

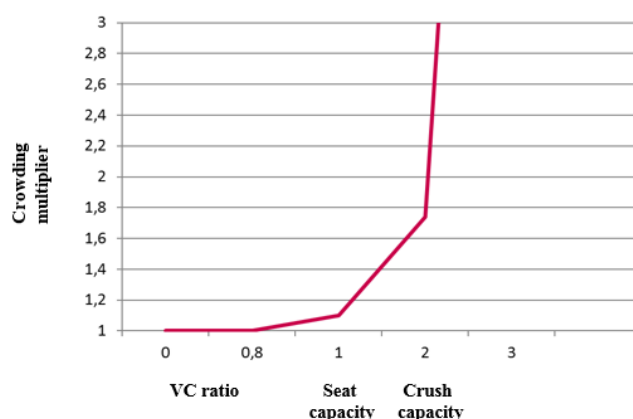
$$GC = \dots + \dots + VOT_{voertuig} \times Crowdfactor \times Reistijd_{voertuig} + \dots + \dots$$

De Crowdfactor hangt samen met een functie die wordt afgeleid van de bezetting:

$$VC = \left\{ \frac{\frac{L}{C_{seats}}}{1 + \frac{L - C_{seats}}{C_{crush} - C_{seats}}} \right.$$

Hierbij staat L voor de voertuigbezetting,  $C_{seats}$  voor de zitplaatscapaciteit en  $C_{crush}$  voor de maximale voertuigcapaciteit. De factor neemt progressief toe naarmate de bezetting van het voertuig verder toeneemt, zie figuur 3.

De steile curve in figuur 3 zorgt ervoor dat in het model de



Figuur 3. Crowding function

aantrekkelijkheid van een reis met een overvol voertuig sterk afneemt en de reiziger een minder drukke route zal kiezen. Via één of meer iteratieslagen bereikt het model een optimale verdeling van reizigers die rekening houdt met de beschikbare capaciteit (en dus reiscomfort).

#### *Toepassing: frequentieverhoging tramlijn 15 in Den Haag*

Deze methode is toegepast in verschillende casestudies met betrekking tot het bus- en tramnetwerk van Den Haag. Een belangrijk verschil met eerdere onderzoeken is toepassing van de capaciteitsafhankelijke toedelingstechniek in het prognosemodel OmniTRANS. We presenteren hier de case waarbij de frequentie van tramlijn 15 is verhoogd van 6 naar 8 trams per uur gedurende de spitsuren. In de casestudy zijn de effecten van de frequentieverhoging berekend zonder en met capaciteitsafhankelijke toedeling.

Tabel 1 laat de verwachte toename van het gebruik van tramlijn 15 zien op werkdagniveau, tabel 2 laat de verwachte toename van OV-gebruik op netwerkniveau zien, waarbij onderscheid is gemaakt tussen ochtend- en avondspits.

	Model zonder comfort	Model met comfort
Werkdag	+8%	+10%

*Tabel 1: verwachte relatieve toename in gebruik tramlijn 15 na frequentieverhoging*

	Model zonder comfort	Model met comfort
Ochtendspits	+165	+ 240
Avondspits	+165	+ 200

*Tabel 2: verwachte reizigerstoename op netwerkniveau na frequentieverhoging lijn 15.*

In tabel 2 zien we dat wanneer alleen wordt gerekend met een kortere wachttijd door de frequentieverhoging een toename van circa 165 reizigers per spitsperiode optreedt. Nemen we ook in ogenschouw dat de frequentieverhoging leidt tot een toename van comfort, dan blijkt de reizigerstoename per spitsperiode 200 tot 240 te bedragen. Aangezien in de ochtendspits de vervoervraag meer geclusterd is in een korte periode, profiteren reizigers in die periode meer van capaciteitsuitbreiding door frequentieverhoging dan in de avondspits. We kunnen concluderen dat traditionele verkeersmodellen, die geen rekening houden met comforteffecten, het effect van een frequentieverhoging op de vervoervraag onderschatten met 20% (avondspits) tot 30% (ochtendspits).

### **3.4 Datafusie van chipkaart- en GSM-data om kansen en bedreigingen voor het OV gedetailleerd in kaart te brengen**

De vorige voorbeelden laten zien dat de OV chipkaart een rijke databron is voor analyse en optimalisatie van het huidige openbaar vervoer. Wat die voorbeelden nog niet laten zien is hoe de OV chipkaart bijdraagt aan inzicht in de potentie van OV. Niet-ov-reizigers zijn immers niet opgenomen in die specifieke databron.

#### *Pilot Emmen*

GSM-data bevat veel informatie over reizigersstromen, ongeacht de vervoerwijze. De zeer wijde verspreiding van mobiele telefoons maakt dat de verplaatsing van telefoons een goede maat is voor de verplaatsingen in Nederland. We combineerden beide databronnen in een pilotproject in de regio Emmen.

De gemeente Emmen telt ruim 100.000 inwoners en heeft met het Dierenpark Emmen een belangrijke publiekstrekker. De stad heeft veel ambitie. Met de aanstaande verhuizing en uitbreiding van het dierenpark in het vooruitzicht wil de stad meer bezoekers trekken en een aantrekkelijke vestigingslocatie zijn. Kijken we naar de modal split, dan is – zoals in zoveel gebieden met een landelijk karakter – het autogebruik relatief hoog.



Onze onderzoeksvraag luidde: *wat kunnen we uit 'big data' leren over aan te boren potentieel aan reizigers met bestemming Emmen voor bestaand of wellicht anders vormgegeven openbaar vervoer?*

Belangrijk hierbij was onze ambitie zoveel mogelijk 'te meten' om daarmee schattingen zo realistisch mogelijk te laten zijn en vermoedens van een feitelijke onderbouwing te voorzien dan wel op grond van feiten te kunnen weerleggen. Dat is goed voor de kwaliteit van de vervolgbeslissingen. En het biedt procesmatige voordelen: minder improductieve discussie en meer draagvlak.

Om de onderzoeksvraag te beantwoorden hebben we het OV-gebruik naar Emmen afgezet tegen het verplaatsingspatroon van mensen naar Emmen. Met andere woorden: we hebben uitgezocht hoe groot het aandeel openbaar-vervoergebruik is op het totaal aantal verplaatsingen van/naar Emmen. Een klein aandeel openbaar-vervoergebruik op zware relaties wijst de weg naar een wellicht nog aan te boren doelgroep.

Voor het feitelijke verplaatsingspatroon met het openbaar vervoer maakten we gebruik van OV chipkaart data. We volstonden voor deze eerste verkenning met de gegevens van één vervoerder (QBuzz). QBuzz rijdt namelijk vrijwel alle busvervoer in en rondom Emmen. Alleen de trein naar Zwolle (Vechtdallijnen, zie Courtz en Van Waveren, 2014) en buslijn 305 naar Groningen worden gereden door Arriva. Dit geeft dus een ondermeting op die reisrelaties die bijvoorbeeld ook door QBuzz lijnen 59, 300 en 600 worden bediend. Qbuzz leverde ons – volledig anoniem - van één maand alle gecombineerde ritten (dus inclusief overstappers) per uurblok met herkomst of bestemming Emmen: in totaal 85.619 reizen. Ruim 90% betrof reizen op werkdagen.

#### *Gebruikte data OV-chipkaart*

Van de OV Chipkaart data maakten we een aantal standaard visualisaties (drukte per dag, druke per uurblok, instappers naar halte etc.). Deze grafieken en kaartbeelden kwamen grotendeels overeen met het verwachte beeld:

- Weekdagen fors drukker dan zaterdag en zondagen;
- Een piek op de maandagen;
- Twee heel drukke ochtendspitsuren en een meer gespreide, maar vroege avondspits;
- Zaterdag en zondag zijn veel gelijkmatiger en de druke komt veel later op gang;
- Veel instappers op Emmen Station, in het centrum, de 'scholierenhaltes' en de overstappunten.

Naast bevestiging van onze verwachtingen zaten er ook verrassingen in de grafieken. Zo is het zondag om 19.00 uur opvallend druk in het OV. Een groot deel van deze reizigers reist naar Groningen: dit zijn waarschijnlijk studenten die na het weekend weer naar hun studiestad terugreizen.

#### *Gebruikte data GSM*

Als tweede databron keken we naar de Herkomst- en Bestemmingspatronen (HB-patronen) zoals deze naar voren komen uit de GSM-data van View.DAT, Deze inzichten zijn gebaseerd op een zeer groot aantal metingen van daadwerkelijk gemaakte individuele verplaatsingen.

View.DAT is gebaseerd op GSM data van het Vodafone netwerk. Landelijk gaat het daarbij om ruim 6 miljoen mobiele telefoons. Deze telefoons nemen we als representatief voor de eigenaar (zie Van der Mede, 2015). Het landelijke netwerk werkt met antennes is daarbij opgedeeld in ruim 35.000 cellen. Apparaten als GSM's, smartphones of iPad's melden zich 20 – 200 keer per dag, ook als ze niet actief gebruikt worden, aan bij een specifieke cel in het GSM-netwerk. Samen met partner Mezero combineert en filtert DAT.Mobility deze metingen tot een kennisbron met feitelijke druke- en verplaatsingspatronen. De telefoonverplaatsingen worden daarbij met

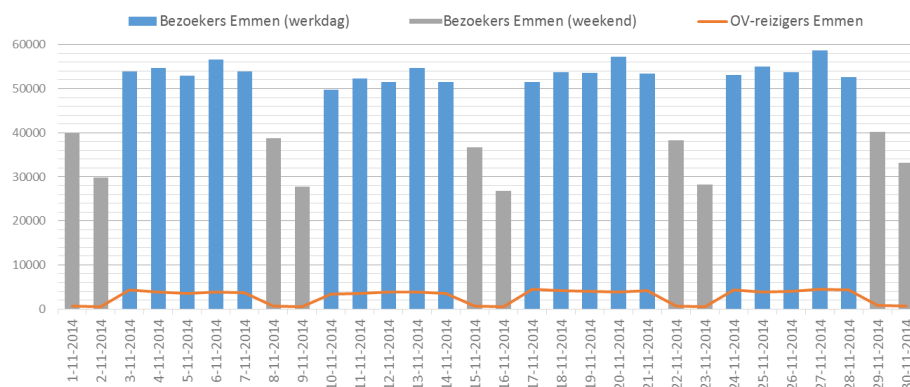
ophoogheurstieken vertaald naar aantallen personen die zich verplaatsen (o.a. gebaseerd op marktaandeel van Vodafone en data uit het OViN). Het detailniveau is 'per uur' en 'per kern'. Binnen de data is het mogelijk om onderscheid te maken tussen frequente, reguliere en incidentele bezoekers. Het onderscheid inwoner/bezoeker wordt eveneens 'gemeten': waar een telefoon 's nachts in de regel aanwezig is wordt aangemerkt als zijn woonplaats.

Met View.DAT bekeken we het mobiliteitsgedrag van bezoekers aan Emmen. In de onderzoeksmaand ging het om 1.117.870 bezoeken aan Emmen. Ook hier komt het gros door de week. Net als bij de chipkaartdata draaiden we ook op deze databron een aantal standaardvisualisaties (zoals de drukte per dag, bezoekfrequentie en een herkomstenkaart). De data liet het volgende beeld zien:

- een relatief stabiel weekpatroon qua aantallen en verdelingen over soorten bezoekers;
- bijna 2/3 van de bezoekers komt frequent in Emmen;
- op zaterdag komen bijna net zoveel bezoekers naar Emmen als op weekdays;
- op weekenddagen is het aantal incidentele en reguliere bezoekers hoger dan op weekdays;
- Emmen Centrum is verantwoordelijk voor bijna de helft van de bezoekers.

#### *Verbanden aanbrengen leidt tot meer inzicht*

Ook zonder inzage in individuele verplaatsingen zijn View.DAT en OV-Chipkaartdata met elkaar in verband te brengen op een manier die past bij onze onderzoeksvraag. We deelden per herkomstgebied het aantal OV-verplaatsingen op het totaal aantal verplaatsingen: de OV-verhouding op de betreffende HB-relatie. Zowel de OV-verhouding als het absolute aantal verplaatsingen visualiseerden we geografisch naar herkomst. Die laatste indicator is belangrijk omdat bij lage absolute aantallen de betrouwbaarheid een punt van aandacht is. Reguliere spreiding kan dominantier zijn dan een door de visualisatie gesuggereerd effect. Bovendien zoeken we naar OV-potentie, daarvoor is ook de absolute omvang van de potentie relevant.



**Figuur 4: dagelijkse mobiliteit naar Emmen stad (OV-reizigers QBuzz)**

Het eerste wat opvalt, is de afwijkende OV-verhouding op zaterdagen. Gemiddeld ontvangt Emmen op werkdagen circa 50.000 bezoekers, van incidentele dagtoeristen tot frequente werknemers en scholieren die niet in Emmen woonachtig zijn. Daarvan komen er op werkdagen tussen de 3.500 en 4.500 met het openbaar vervoer. Het OV-aandeel in de modal split is dus circa 8%. Kijken we naar de zaterdagen, dan daalt het aantal bezoekers naar ongeveer 40.000 terwijl het aantal OV-reizigers daalt naar ongeveer 750. Het OV-aandeel bedraagt op zaterdag nog slechts 2% (zie figuur 4).

In figuur 5 is de volgende informatie samengebracht:

- Per herkomst is het OV-aandeel aangegeven (OV-gebruik uit OV-chipkaart afgezet tegen totaal aantal verplaatsingen uit GSM-data);
- Met balken is aangegeven het totaal aantal bezoekers aan Emmen;
- Per herkomst is met taartdiagrammen de bezoekfrequentie van bezoekers aan Emmen aangegeven.

We kunnen uit de figuur vele zaken afleiden:

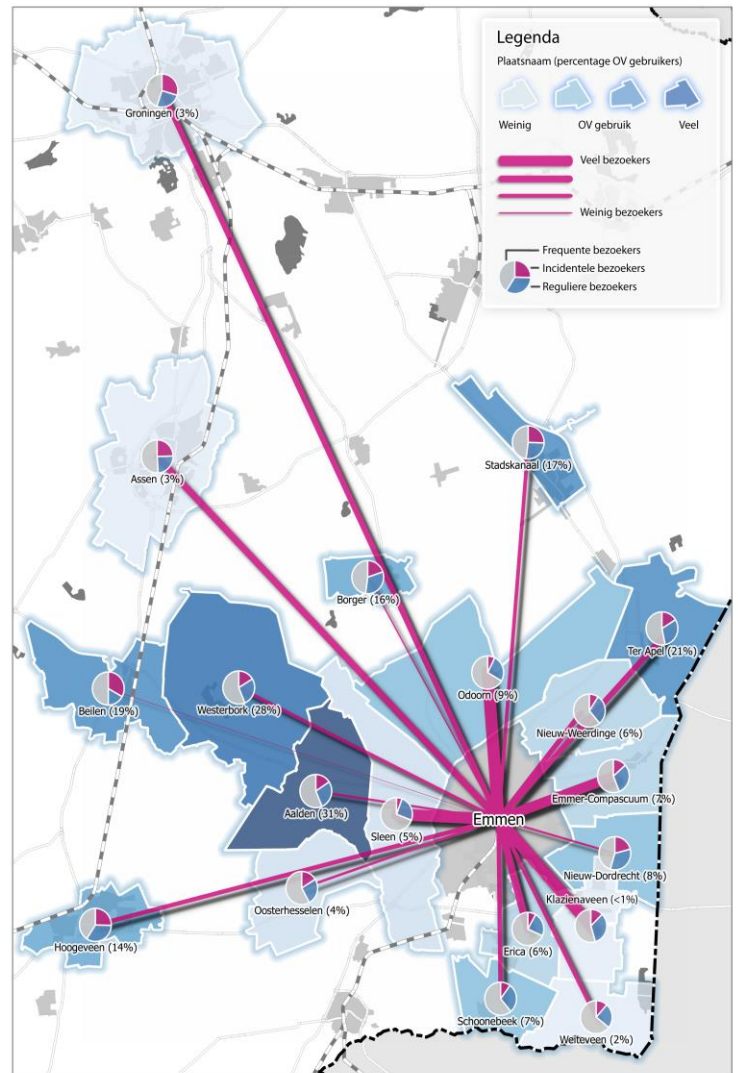
- Kernen die net iets verder liggen dan Emmen laten een relatief sterk 'OV-profiel' zien, zoals Ter Apel, Borger en Westerbork. De kernen liggen dicht genoeg bij Emmen om een grote stroom bezoekers te genereren, maar te ver weg om met de fiets te komen.
- Klazienaveen is een uitzondering hierop: een erg laag aandeel OV in combinatie met een groot aantal verplaatsingen. Een verklaring ligt mogelijk in de goede fietsverbindingen met Emmen.
- Schoonoord, Sleen en Erm vallen ook op: hun aandeel OV ligt lager dan veel gebieden die op dezelfde afstand liggen. Gezien het aantal verplaatsingen zijn deze relaties interessant om nader te onderzoeken, hier ligt immers OV-potentieel!
- Geen eenduidig beeld, maar wel interessante onderzoeksinformatie levert de confrontatie van het OV-aandeel met het bezoekersprofiel vanuit een gebied. Frequente bezoekers komen meer dan 10 maal per maand in Emmen, incidentele bezoekers minder dan 3 maal.

Een interessant vervolgonderzoek lijkt ons de frequentie van de OV-gebruikers hiermee te confronteren. Hebben deze een totaal ander patroon? Het lijkt er wel op gegeven het achterblijvende OV-gebruik op zaterdagen, maar nadere analyse van de OV-chipkaartdata kan hierin inzicht geven.

## 5. Conclusies

Dit paper laat zien dat er met OV chipkaartdata nog veel meer te ontdekken valt dan het huidige kleuren binnen de lijntjes. Met drie verschillende voorbeelden hebben we laten zien dat er nog grote stappen te zetten zijn in het verbeteren van het OV en de planning daarvan.

- OV chipkaart data is rechtstreeks bruikbaar in analyses. Inzet van een vraagmodel is niet nodig. Zo wordt meer gewerkt met waargenomen gedrag ("revealed preference").
- De elasticiteitsparameter verschilt tussen reguliere en tijdelijke veranderingen. Analyse van de OV chipkaart toont aan dat de waarde voor elasticiteit bij tijdelijke



Figuur 5: Analyse bezoekersstromen Emmen

veranderingen circa 25% hoger ligt. Reizigers reageren dus heftiger op tijdelijke verstoringen. Het wordt aangeraden om in modellen bij langdurige tijdelijke omleidingen een hogere absolute elasticiteitswaarde te gebruiken.

- Voor reizigers is comfort een belangrijk aspect. Het meenemen van comfort en capaciteit in prognosemodellen is daarom relevant. Onze case studie met capaciteitsafhankelijke modellen bij vervoerbedrijf HTM in Den Haag toont aan dat deze aanpak bruikbaar is, voldoende snel doorrekent en dat resulterende waarden interessant zijn voor het bepalen van vervoerseffecten.
- Big data bronnen laten zich binnen de verkeerskunde goed combineren. Analyse in de regio Emmen toont aan inzet van OV chipkaart én GSM goed mogelijk is. Verschil in onderzoekspopulatie en differentiatieniveaus is hierbij niet relevant en individuele reisbewegingen zijn niet nodig. Met slim gekozen aggregaten zijn al heel waardevolle analyses te doen. In de regio Emmen lieten wij zien dat een aantal HB relaties in absolute zin een grote stroom reizigers kent, maar dat hier (nog) geen significant aandeel OV is. Zo wordt objectief zichtbaar waar de potentie zit. Dat is een grote stap voorwaarts voor vervoerbedrijven en OV- autoriteiten. Zij kunnen aan de slag met bijvoorbeeld een slimmer tarievenaanbod of andere vormen van mobiliteitsbeleid. De ingezette verdere verfijning van de gebiedsindeling van View.DAT maakt deze analyses steeds beter.

Bovenstaande cases tonen aan hoe belangrijk het is de werkvelden 'big data' en de (openbaar) vervoersector met elkaar te verbinden. Deze bevindingen staan niet op zichzelf en markeren in onze ogen de start van een 'revolutie' met meer inzicht in reizigersgedrag. Meer data zal beschikbaar komen, vragen om slimme oplossingen resulteren in kortere beleids- en sturingscycli. De TU Delft, Goudappel Coffeng en DAT.Mobility blijven zoeken naar manieren om bij te dragen aan betere planvorming en bijsturing van het openbaar vervoer!

### **Acknowledgements**

Pionieren doen je nooit alleen. Dit onderzoek is uitgevoerd in samenwerking met OV bureau Groningen-Drenthe, Provincie Drenthe, gemeente Emmen, QBuzz en HTM. Maarten Oud deed het onderzoek naar gedrag bij omleidingen i.s.m. Goudappel Coffeng, HTM en TU Delft.

### **Referenties**

Bagchi, M, P. White (2005). The Potential of Public Transport Smart Card Data. Transport Policy, Vol. 12, No. 5, pp 464-474.

Bregman, S. (2012). Uses of Social Media in Public Transportation. Transit Cooperative Research Program (TCRP) Synthesis 99. Transportation Research Board, Washington,

Calabrese, Francesco (2011). Estimating origin-destination flows using mobile phone location data. In: Pervasive Computing, IEEE.

Cats, O., M.D. Yap, N. van Oort (2015) Exposing the role of exposure: identifying and evaluating critical links in public transport networks, INSTR conference, Nara, Japan.

Cepeda, M., Cominetti, R., Florian, M. (2006) A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. Transportation Research Part B: Methodological, 40, pp 437-459.

Courtz, M., H.D. van Waveren (2014). Databoogjes geven inzicht, OV-magazine.

Florian, M. (2002). Frequency based transit route choice models. Chapter 6 in: *Advanced Modeling for Transit Operations and Service Planning*, ed. William H.K. Lam, Michael G. H. Bell.

Heuvel, J. van den, J., A. Voskamp, W. Daamen, S.P. Hoogendoorn (2015). Using bluetooth to estimate the impact of congestion on pedestrian route choice at train stations, *Traffic and Granular flow '13*, M.Charaibi et. al (eds), Switzerland.

Hickman, M. (2004) *Evaluating the Benefits of Bus Automatic Vehicle Location (AVL) Systems*, in: D. Levinson and D. Gillen (eds.), *Assessing the Benefits and Costs of Intelligent Transportation Systems*, Chapter 5, Kluwer, Boston.

Kurauchi, F., J.D. Schmöcker, H. Shimamoto, S.M. Hassan (2014). Variability of commuters' bus line choice: an analysis of oyster card data. *Public Transport*, Vol. 6, No. 1-2, pp. 21-34.

Lee, A., N. van Oort, R. van Nes (2014). Service Reliability in a network context: impacts of synchronizing schedules in long headway services, *Transportation Research Record*, No. 2417, 2014, pp. 18-26.

Leusden, R van, Oort, N. van, Ebben, M. (2012). Robuuster, goedkoper en beter openbaar vervoer door gebruik va GOVI-data. *Bijdrage aan het colloquium vervoersplanologisch speurwerk* (pp. 1-12).

Ma, X., Y.J. Wu, Y. Wang, F. Chen, J. Liu (2013). Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 36, 2013, pp. 1-12.

Munizaga, M., C. Palma (2012). Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smart card data from Santiago, Chile. *Transportation Research C*, Vol. 24, pp. 9-18.

Neema, N., Hickman M. and Z-L. Ma (2015). Activity detection and transfer identification for public transit fare card data, *Transportation*, Vol. 42, No. 4, 2015, pp. 683-705.

Park, J., D.J. Kim, D.J., Y. Lim (2008). Use of Smart Card Data to Define Public Transit Use in Seoul, South Korea. *Transportation Research Record*, No. 2063, pp.3-9.

Pel, A. J., Bel, N. H., Pieters, M. (2014) Including passengers' response to crowding in the Dutch national train passenger assignment model. *Transportation Research Part A: Policy and Practice*, 66, 111-126.

Pelletier, M., M. Trepanier en C. Morency (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 4, 2011, pp. 557-568.

Schmöcker, J. D., Fonzone, A., Shimamoto, H., Kurau chi, F., Bell, M. G. (2011) Frequency-based transit assignment considering seat capacities. *Transportation Research Part B: Methodological*, 45(2), 392-408.

Schmöcker, J.D., H. Shimamoto, F. Kurauchi (2013). Generation and calibration of transit hyperpaths. *Transportation Research Part C: Emerging Technologies*, Vol. 36, pp. 406-418.

Seaborn, C., J. Attanucci, N.H.M. Wilson (2009). Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record*, No.2121.-1, pp. 55-62.

Van der Mede, P. (2015), GSM-data niet meer weg te denken voor mobiliteitstoepassingen, GIS-Magazine.

Van Oort, N & R. van Leusden (2015). Reizigerseffecten van onbetrouwbaar ov in maatschappelijke kostenbatenanalyses. Tijdschrift Vervoerswetenschap, 51(1), pp.67-81.

Van Oort, N. (2014). OV-data wereldwijd omarmd. OV Magazine, augustus.

Van Oort, N., M.P. Drost, T. Brands (2014). Betere OV prognoses met anonieme OV-Chipkaartdata. CVS congres Eindhoven.

van Oort, N., D. Sparing, T. Brands, R. Goverde (2013). Optimizing Public Transport Planning and Operations Using Automatic Vehicle Location Data: The Dutch Example. Presented at the 3rd International Conference on Models and Technology for ITS, Dresden, Germany.

Wang, W., Attanucci, J.P., N.H.M. Wilson (2011). Bus Passenger Origin-Destination Estimation and Related Analyses. Journal of Public Transportation, Vol. 14, No. 4, pp. 131-150.

Yap, M, N. van Oort, R. van Nes & B. van Arem (2015). Robuustheid van multi-level openbaar vervoer netwerken. Tijdschrift Vervoerswetenschap, 51(1), pp. 82-99.