

## **Busbehoefte op basis van OV-chipkaartdata**

M.W. Smit MSc  
m.w.smit@alumnus.utwente.nl

Ing. K.M. van Zuilekom  
Universiteit Twente  
k.m.vanzuilekom@utwente.nl

Ing. C. Doeser  
Grontmij  
Cees.doeser@grontmij.nl

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

## Samenvatting

### *Busbehoefte op basis van OV-chipkaartdata*

De invoering van de OV-chipkaart heeft geleid tot een nauwkeurig inzicht in het reisgedrag van reizigers. Hoewel niet alle gegevens beschikbaar zijn om privacy en concurrentie overwegingen, biedt de beschikbare informatie veel onderzoeksmogelijkheden. Wanneer het mogelijk is om het verplaatsingsgedrag van busreizigers nauwkeurig te voorspellen met modellen, kunnen vervoerders hier gebruik van maken voor het ontwerpen van hun lijnennet en de dienstregeling.

De vraag is of het mogelijk is om met de beschikbare OV-chipkaartdata en andere databronnen het reisgedrag van busreizigers nauwkeurig te beschrijven. Daarnaast is het de vraag hoe generiek dit reisgedrag is. Deze paper gaat in op deze vragen.

Met behulp van OV-chipkaartdata en sociaaleconomische en ruimtelijke data is een unimodaal vierstapsmodel ontwikkeld. Er is gebruik gemaakt van OV-chipkaartdata van Breda en Tilburg, beschikbaar gesteld door Veolia Transport. Daarnaast is data van onder andere het CBS en DUO gebruikt. De modellen zijn ontwikkeld voor Breda en vervolgens toegepast op Tilburg om de generaliseerbaarheid van de modellen te testen.

De resultaten van het onderzoek zijn veelbelovend. De ritproductie en –attractie in Breda bleek heel goed overeen te komen met de werkelijkheid met een  $R^2$  van ruim boven de 0,9 voor zowel productie als attractie. Ook de gekalibreerde distributiefunctie geeft een zeer goed resultaat in Breda. De uiteindelijke toedeling van de reizigersdistributie wijkt ongeveer 5% af van de waarnemingen.

Ook het toepassen van het Breda model in Tilburg geeft vrij goede resultaten. De ritproductie en –attractie hebben beiden een  $R^2$  van ruim boven de 0,9. De resultaten zijn iets minder goed dan bij toepassing in Breda, maar nog altijd behoorlijk goed. Ook toepassing van het distributiemodel geeft een zeer goed resultaat. De toedeling van de reizigersdistributie wijkt ongeveer 12% af van de waarnemingen.

Op basis van dit onderzoek kan geconcludeerd worden dat er met de OV-chipkaart data mogelijk is om een vrij nauwkeurig unimodaal model te ontwikkelen. Aangezien in dit onderzoek slechts twee steden zijn beschouwd is verder onderzoek noodzakelijk. Uiteindelijk zou dit moeten leiden tot een *tool* waarmee vervoerders inzicht krijgen in het reisgedrag van hun passagiers.

## 1. Inleiding

Nu de OV-chipcard in geheel Nederland is geïntroduceerd wordt het feitelijk gedrag van de reiziger nauwkeurig vastgelegd. Vanwege privacy en concurrentie overwegingen is er weliswaar geen compleet beeld bij de vervoerder van de verplaatsingsketens op individueel niveau van A naar B, maar de nu beschikbare informatie staat niet in verhouding tot de situatie ten tijde van de strippenkaart. De OV-chipkaart informatie is ook niet geheel sluitend aangezien in het busvervoer nog papieren producten bestaan. Dat neemt niet weg dat met de OV-chipkaart informatie het mogelijk zou moeten zijn om het verplaatsingsgedrag goed te beschrijven. Dit opent nieuwe perspectieven. Immers als het verplaatsingsgedrag goed te beschrijven is, kan hier in het ontwerpstadium van een lijnennet en dienstregeling al gebruik van worden gemaakt.

Idealiter is de beschrijving van het verplaatsingsgedrag nauwkeurig en generiek. Met nauwkeurig wordt verstaan dat de verschillende stadia in het verplaatsingsgedrag (bepaling van bestemming, vertrektijdstip, vervoerswijze, route, overstappen) het keuzeprocess en de uitkomsten van die keuzes goed worden beschreven. Als het verplaatsingsgedrag generiek is dan is het mogelijk om de resultaten van de ene situatie te gebruiken in een ander situatie.

Al met al is het de vraag of met de beschikbaarheid van de OV-chipkaart informatie er een goede beschrijving van het verplaatsingsgedrag mogelijk is. Het is ook de vraag of de beschrijving van het verplaatsingsgedrag generiek is waardoor het mogelijk is de gevolgen van grotere aanpassingen in het OV te beschrijven.

Deze vragen stonden centraal in het onderzoek dat in opdracht van de Grontmij en in samenwerking met Veolia in de steden Breda en Tilburg is uitgevoerd. Het onderzoek is uitgevoerd door Marijn Smit in het kader van zijn afstuderen aan de Universiteit Twente bij de vakgroep Verkeer, Vervoer- en Ruimte. Het onderzoek is vanuit de Grontmij begeleid door M.E. Kraan en C. Doeser en vanuit de UT begeleid door K.T. Geurs en K.M. van Zuilekom.

Met het onderzoek is een eerste stap gezet naar een instrument, *tool*, waarmee het voor de vervoerder mogelijk wordt om inzicht te krijgen op belangrijke dimensioneringsvraagstukken: (1) wat zijn de herkomsten en bestemmingen van onze klanten; (2) hoe verdelen de klanten zich vanuit de herkomsten over de bestemmingen; (3) waar moeten de haltes komen; (4) welk lijnennet past het beste bij de vervoersvraag en (5) wat moet de capaciteit van de vervoermiddelen worden?

## 2. Methodologie

Gekozen is om het unimodale zwaartekrachtmodel als uitgangspunt te nemen. De stappen in het unimodale zwaartekrachtmodel sluiten aan bij de dimensioneringsvraagstukken van de vervoerder, namelijk waar komen reizigers vandaan en waar gaan ze naar toe, welke haltes worden gebruikt, welke route wordt gevolgd en wat betekent dat voor de bezettingsgraad. Met de keuze voor één vervoerswijze, *unimodaal*, is de concurrentie van en uitwisseling met andere vervoerswijzen vooralsnog niet meegenomen. In het onderzoek zijn de OV-chip data centraal gesteld en daarmee informatie van overige vervoerswijzen (lopen, (brom/snor)fiets en auto) nog buiten beschouwing gebleven.

Gebruikt zijn alle TLS data zoals de vervoerder Veolia beschikbaar had over de periode 01-01-2012 t/m 31-12-2012. Aanvullend is sociaaleconomische en ruimtelijke informatie verzameld voor de steden Breda en Tilburg. In hoofdzaak is deze informatie via het CBS, DUO en het Kadaster verkregen.

Alle data is omgezet naar een 100 meter grid in een Grafisch Informatie Systeem (GIS).

Voor Breda zijn de ontwikkelingsstappen van een unimodaal zwaartekrachtmodel doorlopen:

- Bepaling van de coëfficiënten voor riteindberekening voor verkeersproductie en – attractie
- Bepaling van de distributiefunctie, de coëfficiënten van de distributiefunctie en daarmee tevens de herkomsten- bestemmingen matrix (HB-matrix)
- Toedelen van de HB-matrix, via de haltes, aan de buslijnen

Vervolgens zijn de coëfficiënten van het model Breda toegepast op de situatie van Tilburg:

- Toepassing van de coëfficiënten van de riteindberekening van Breda op de sociaaleconomische gegevens van Tilburg
- Distributie van de herkomsten en bestemmingen van Tilburg (vorige stap) gebruikmakend van de distributiefunctie van Breda
- Toedelen van de HB-matrix (vorige stap), via de haltes, aan de buslijnen

### 3. Resultaten voor Breda

#### 3.1 Ritproductie en -attractie

Er is een ritproductie- en een ritattractiemodel voor doordeweekse dagen ontwikkeld voor Breda. Met deze modellen kan het aantal in- en uitstappers van elke bushalte worden geprognostiseerd. Voor de ontwikkeling was de beschikking over de geregistreerde in/uitstappen en 55 variabelen met sociaal-economische en ruimtelijke gegevens uit de directe omgeving van elke bushalte beschikbaar. De uiteindelijke formules voor de ritproductie en ritattractie bevatten 10 variabelen. Er is een stapsgewijze aanpak gevolgd, waarbij de invloedrijkste variabele als basis is gebruikt en iteratief variabelen zijn toegevoegd waarbij het resultaat zowel voor Breda als voor Tilburg verbeterde. Variabelen waarbij het resultaat verslechterde zijn niet toegevoegd.

De ritgeneratie en -attractie hebben de volgende formules opgeleverd:

$$\text{Ritproductie} = -128.654 + 3.721x_1 - 0.612x_2 + 0.135x_3 + 0.155x_4 - 0.044x_5 + 0.002x_6 + 1.074x_7 + 0.019x_8 + 9281.374x_9 - 152.157x_{10}$$

$$\text{Rit attractie} = -127.775 + 3.979x_1 - 0.643x_2 + 0.096x_3 + 0.126x_4 - 0.029x_5 + 0.002x_6 + 0.886x_7 + 0.015x_8 + 8127.618x_9 - 141.667x_{10}$$

Tabel 1 licht toe welke variabelen gebruikt zijn. De ritproductie en ritattractie bevatten beide dezelfde variabelen.

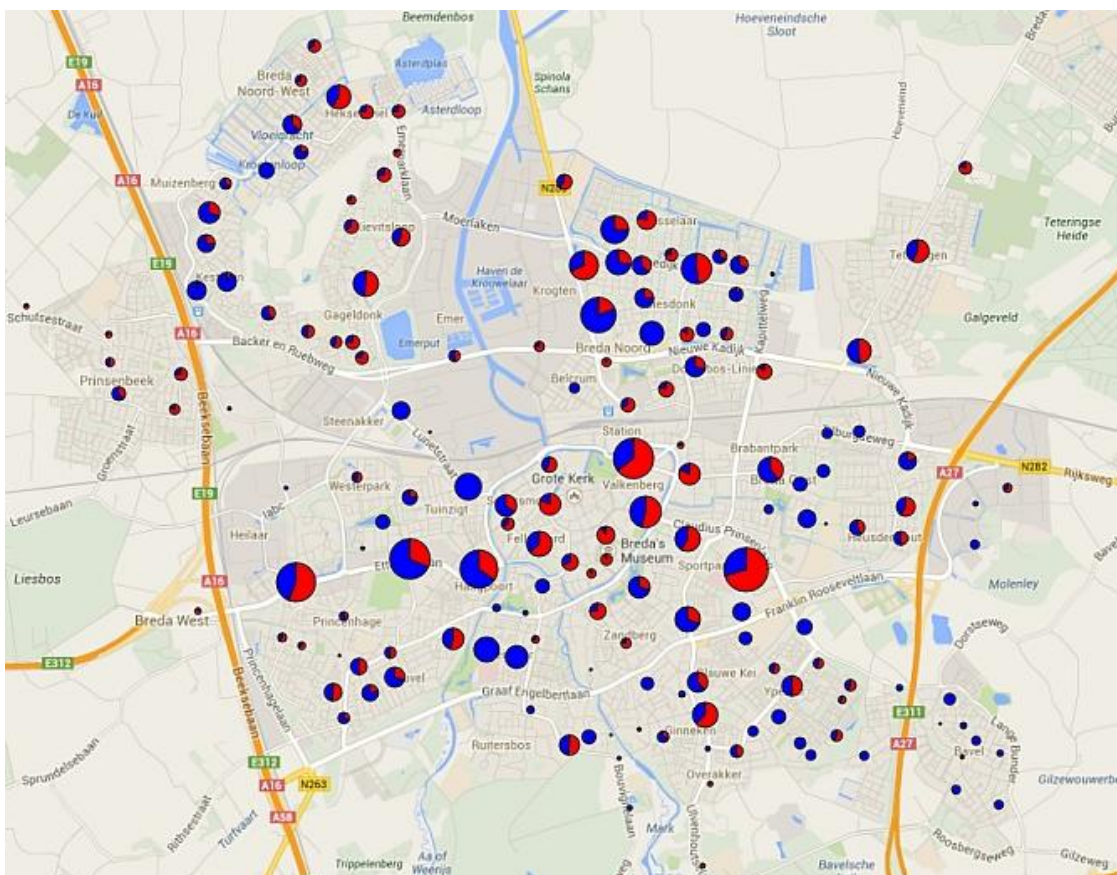
Tabel 1: Variabelen in de formules van de ritproductie en -attractie

Variabele	Beschrijving variabele	Toelichting
$x_1$	Bereikbaarheidsindicator	Aantal haltes dat bereikbaar is vanaf een halte zonder overstappen
$x_2$	Percentage laag-inkomenshuishoudens	
$x_3$	Aantal auto's	
$x_4$	Aantal studenten	MBO, HBO en WO
$x_5$	Omgevingsadressendichtheid	
$x_6$	Oppervlakte kantoorruimte	
$x_7$	Percentage woningen	
$x_8$	Oppervlakte winkels en horeca	
$x_9$	Dummy voor aanwezigheid van Centraal Station	Waarde 0 bij geen aanwezigheid, waarde 1 bij wel aanwezig
$x_{10}$	Dummy voor aanwezigheid Hogeschool of Universiteit	Waarde 0 bij geen aanwezigheid, waarde 1 bij wel aanwezig

Opvallend aan beide formules is dat de constante term negatief is. Het is dus mogelijk dat bij bepaalde haltes een negatief aantal in- of uitstappers wordt voorspeld. Dit wordt veroorzaakt door het grote verschil in aantallen in- en uitstappers bij de haltes; er zijn haltes waar dagelijks duizenden mensen gebruik van maken, terwijl er ook haltes zijn waar slechts enkele mensen in of uit de bus stappen.

Nog iets opvallends is dat de coëfficiënt voor de variabele voor de aanwezigheid van een Hogeschool of Universiteit negatief is. De aanwezigheid van een Hogeschool of Universiteit zou dus een negatieve invloed hebben op het aantal in- en uitstappers. Zonder deze variabele wordt het aantal in- en uitstappers voor haltes bij een Hogeschool of Universiteit blijkbaar overschat. Mogelijk heeft dit te maken met de variabele met het aantal studenten.

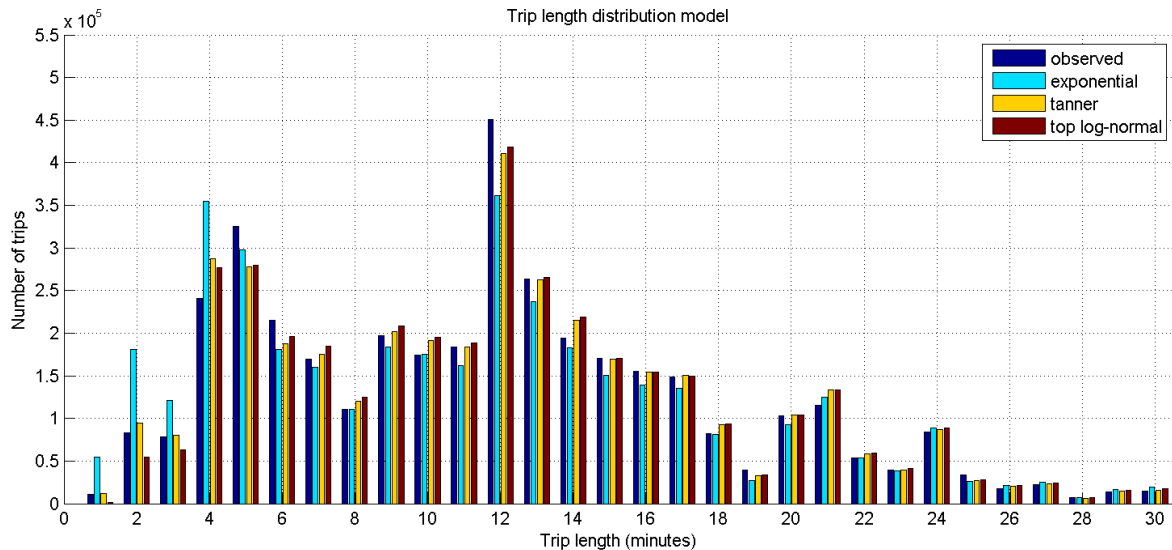
De ritteindberekening vertont een zeer goede fit met de waarneming conform de TLS-data. Het ritproductieberekening heeft een fit met een  $R^2$  van 0,981 en het attractiemodel een  $R^2$  van 0,997. De resultaten worden getoond in figuur 1. De cirkels vertegenwoordigen de bushaltes, waarbij het blauwe deel het aantal instappers volgens de TLS data is en het rode deel het gemodelleerde aantal. Bij haltes waarbij het rode en blauwe deel even groot zijn, komt het model zeer goed overeen met de werkelijkheid.



*Figuur 1: Bushaltes in Breda met aantal instappers volgens TLS-data (blauw) en gemodelleerd (rood). Niet alle haltes worden weergegeven.*

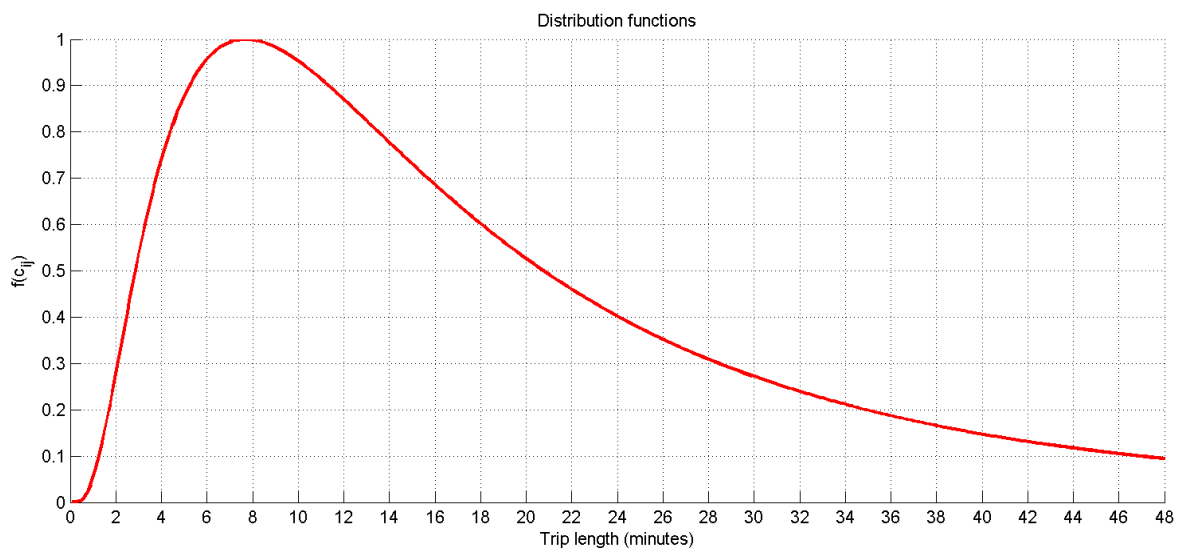
### 3.2 Ritdistributie

Er zijn drie verschillende distributiefuncties onderzocht: exponentieel, tanner en top log-normaal. De exponentiele distributiefunctie gaf geen goed resultaat. De tanner en top log-normale distributiefunctie gaven een zeer goed en zo goed als identiek resultaat (zie figuur 2). Gekozen is om met de top log-normale distributiefunctie verder te gaan aangezien de top log-normale distributiefunctie beter te interpreteren is (één parameter legt de plaats van de top vast, de andere parameter de afstandgevoeligheid).



*Figuur 2: Ritlengtefrequentieverdelingen volgens de TLS-data en volgens de geschatte distributiefuncties*

Figuur 3 laat de top log-normale distributiefunctie zien. De grafiek laat de bereidheid zien om binnen Breda met de bus te reizen. De top van de grafiek ligt op een ritlengte van 7,7 minuten.

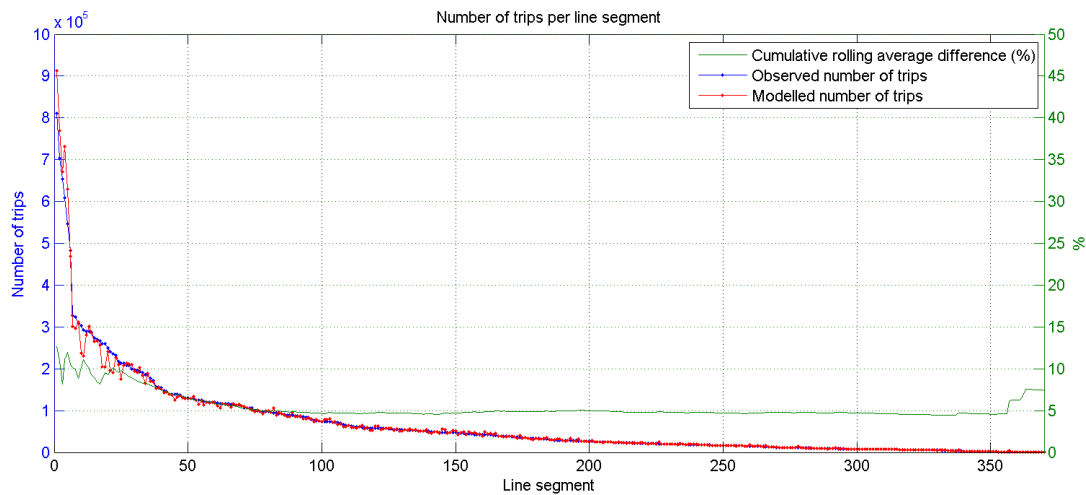


*Figuur 3: Plot van de gekalibreerde top log-normale distributiefunctie*

### 3.3 Toedelen

Tijdens het schatten van de distributiefunctie wordt een HB-matrix geconstrueerd. Deze HB-matrix die het resultaat is van de distributie is toegedeeld aan het busnetwerk volgens het principe van de kortste route gemeten naar minste aantal haltes. De resultaten zijn gevisualiseerd in de grafiek in figuur 4. De figuur laat elk lijnsegment (deel van het busnetwerk tussen twee opeenvolgende haltes) in het busnetwerk zien met de gemeten en gemodelleerde aantallen reizigers. De lijnsegmenten zijn gesorteerd naar het aantal waargenomen verplaatsingen op de segmenten. Gemiddeld verschilt de

toedeling ongeveer 5% van de waargenomen verplaatsingen (weergegeven in de groene lijn).



Figuur 4: Gemeten en gemodelleerde aantallen reizigers per lijnsegment in Breda

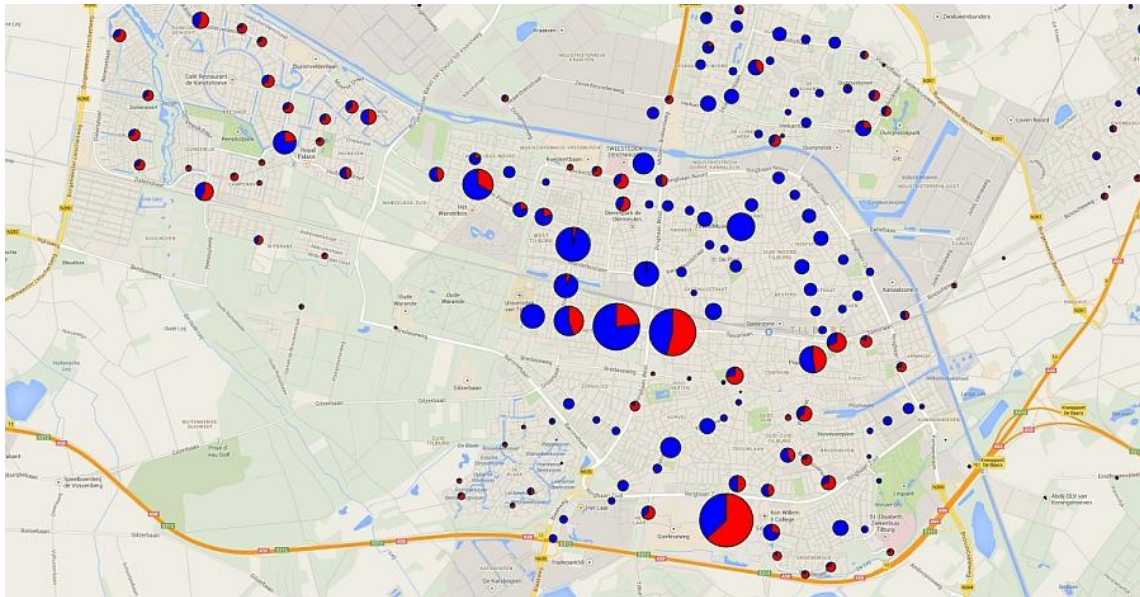
#### 4. Toepassing van model coëfficiënten van Breda op de situatie van Tilburg

Om de algemene toepasbaarheid van de modellen te toetsen, zijn de voor Breda ontwikkelde model coëfficiënten toegepast op Tilburg. Verwacht kon worden dat de modelresultaten meer afstand van het waargenomen gedrag zou hebben. De steden zijn niet identiek, alhoewel er ook wel overeenkomsten zijn. De steden hebben beide hoger onderwijsinstellingen binnen de grenzen en kennen een vrij vergelijkbare leeftijdsopbouw van de bevolking. Ook wat aantal inwoners en arbeidsplaatsen betreft ontlopen beide steden elkaar niet veel. Relevant voor dit onderzoek is dat het aantal treinreizigers op het Centraal Station van beide steden van vergelijkbare grootte is.

##### 4.1 Ritproductie en -attractie

Figuur 5 toont de resultaten van het ritgeneratiemodel wanneer deze wordt toegepast op Tilburg. De verschillen tussen het model en de werkelijkheid zijn duidelijk groter dan in Breda, hoewel er ook haltes zijn waar beide vergelijkbaar zijn.



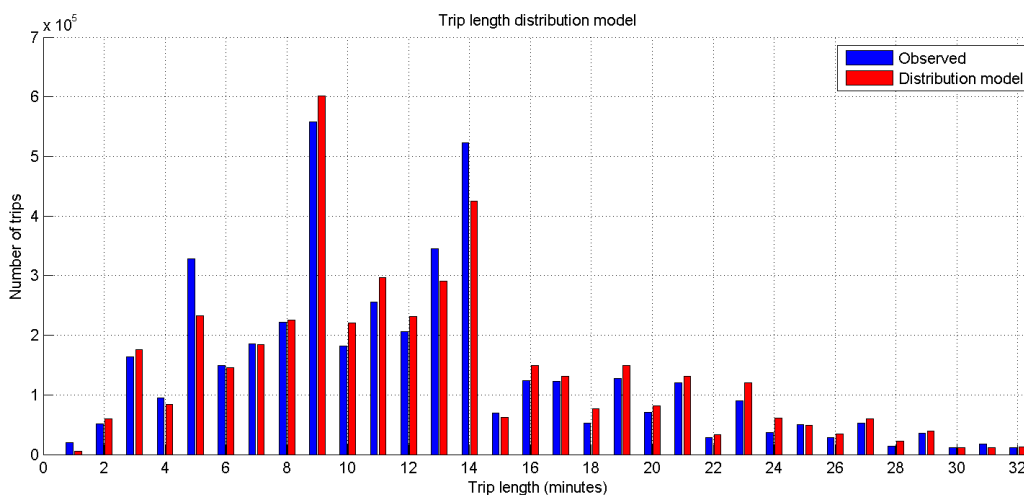


*Figuur 5: Bushaltes in Tilburg met aantal instappers volgens TLS-data (blauw) en gemodelleerd (rood). Wegens overlappende cirkels worden niet alle haltes weergegeven.*

De figuur laat zien dat er veel haltes zijn die volledig of vrijwel volledig blauw zijn. Dit zijn haltes waar de ritproductie volgens het model negatief is. Het gebied ten noorden van het Centraal Station kent vrijwel geen instappers volgens het model. Bij de haltes in het noordwesten komt het aantal instappers juist vrij goed overeen met de werkelijkheid. Het model is dus wel toe te passen in de ene wijk in Tilburg, maar niet in de andere. Het resultaat bij de drukste haltes is goed.

#### 4.2 Distributie

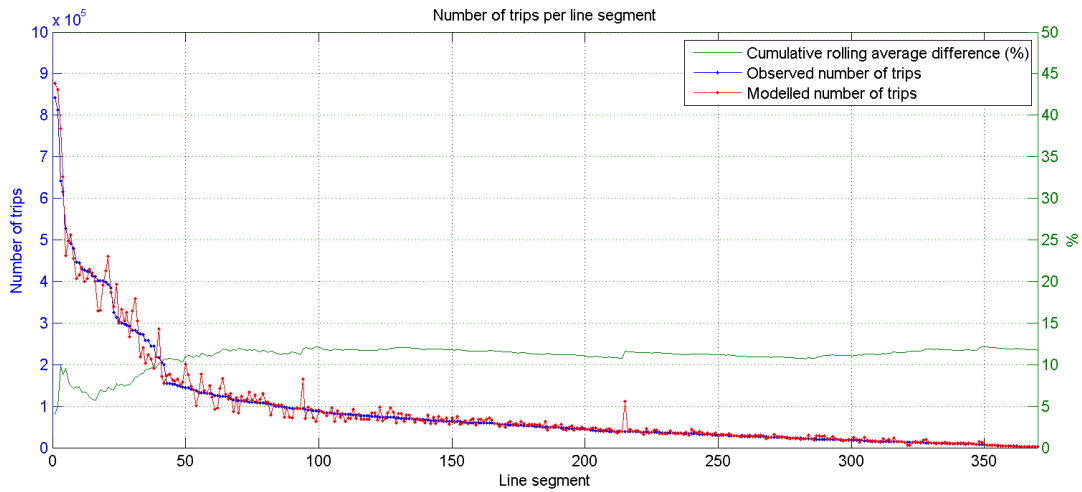
Toepassing van het top lognormale distributiefunctie van Breda in Tilburg geeft een goed resultaat (zie figuur 6).



*Figuur 6: Ritlengtefrequentieverdelingen zoals gemeten en volgens de gekalibreerde top log-normale distributiefunctie*

### 4.3 Toedelen

Ook de toedeling van verplaatsingen aan het netwerk geeft een goed resultaat (zie figuur 7). De toedeling verschilt gemiddeld rond de 12% van de waarneming, waar dit in Breda ongeveer 5% is.



Figuur 7: Gemeten en gemodelleerde aantallen reizigers per lijnsegment in Tilburg

## Conclusies en aanbevelingen

Onderzocht is in hoeverre het mogelijk is om op basis van OV-chip data zoals een vervoerder (Veolia) heeft, een verkeersmodel voor een stad (Breda) te schatten en vervolgens de coëfficiënten van dat model te gebruiken om voor een andere stad (Tilburg) de verkeersbehoefte te prognotiseren.

Het OV-model Breda beschreef de situatie behoorlijk goed. De vertrekken en aankomsten voor een gemiddelde werkdag konden goed beschreven worden ( $R^2$  van 0,981 productie en een  $R^2$  van 0,997 voor de attractie). Met de top log-normale distributiefunctie kon de distributie nauwkeurig beschreven worden. De uiteindelijke toedeling week zo'n 5% van de waarneming af.

Bij toepassing van de coëfficiënten van het Breda model in Tilburg kon het niet anders dan dat het model voor Tilburg slechter zou presteren. Het is echter zo dat het model voor Tilburg behoorlijk goede resultaten laat zien.

De vertrekken en aankomsten voor een gemiddelde werkdag in Tilburg konden goed beschreven worden ( $R^2$  van 0,974 productie en een  $R^2$  van 0,973 voor de attractie). Met de top log-normale distributiefunctie en de coëfficiënten uit Breda kon de distributie heel goed beschreven worden. De uiteindelijke toedeling week zo'n 12% van de waarneming af.

Overall zijn de resultaten van het onderzoek bemoedigend. Met de OV-chipkaart data, zoals een vervoerder nu al ter beschikking heeft, is het mogelijk om een behoorlijk nauwkeurig unimodaal OV-model te maken. Met name distributie en toedeling geven goede resultaten

Met slechts twee onderzochte steden is het nog te vroeg om vast te stellen dat er een universele distributiefunctie in het busvervoer bestaat, maar dit is zeker de moeite van verder onderzoek waard.

Op dit moment beschikken de vervoerders niet over de overstappen van de reiziger, waardoor het niet mogelijk is om het verplaatsingsgedrag volledig uit waarnemingen op te bouwen. Model veronderstellingen zijn daarom nodig om de leemte in informatie in te vullen. Zodra ook de overstap tussen lijnen beschikbaar is zal in principe alleen voor de relatie tussen de sociaaleconomische gegevens en de uiteindelijke berekening een modelveronderstelling nodig zijn.

Uiteindelijk doel is om een *tool* te ontwikkelen waarmee het mogelijk is om een OV-systeem te ontwikkelen op basis van het waargenomen verplaatsingsgedrag. Met het onderzoek in Breda en Tilburg is aangetoond dat het goed mogelijk is om het verplaatsingsgedrag te beschrijven. Daarmee is er voldoende grond om de ontwikkeling in de richting van die ontwerp *tool* voort te zetten. Evaluatie van bestaande OV-systemen is hierbij een essentiële stap. De uitdaging zal ook zijn om ook het potentieel van OV-verplaatsingen inzichtelijk te maken en te integreren in het *tool*.