

**Over het meten van mobiliteit met gsm-data:  
mogelijkheden en onmogelijkheden**

Peter van der Mede  
DAT.Mobility (onderdeel van Goudappel Groep)  
pvdmede@dat.nl

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

## Samenvatting

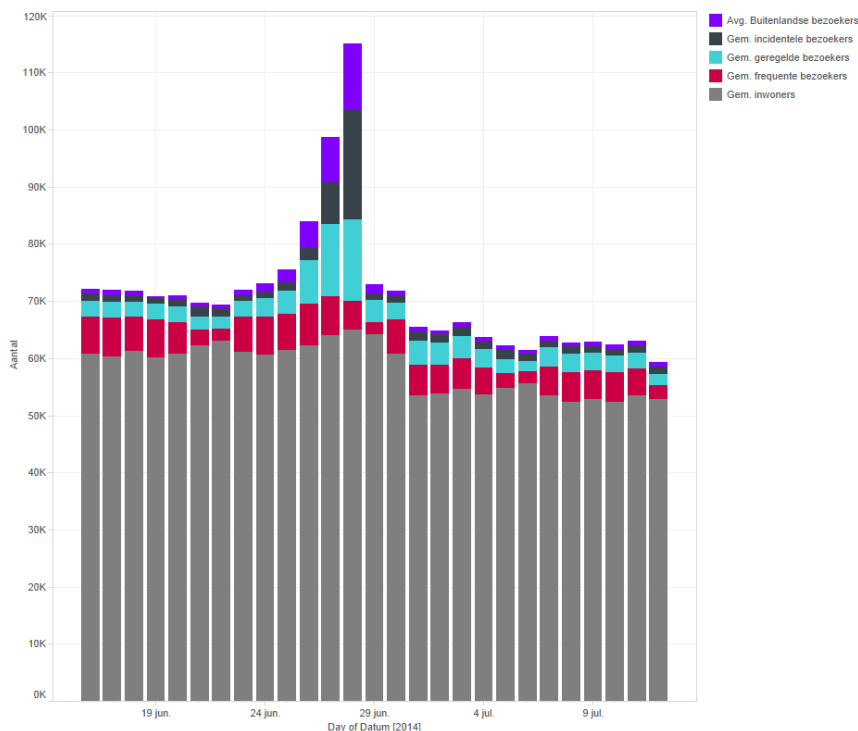
### *Over het meten van mobiliteit met gsm-data: mogelijkheden en onmogelijkheden*

Data uit mobiele telefoons zijn een nieuwe, ongekend rijke bron om allerlei mobiliteitsgerelateerde informatie uit af te leiden. De toepassing van deze bron in Nederland is nu een feit en de informatie eruit is in principe voor iedereen toegankelijk. De mogelijkheden van gsm-data zijn legio: druktebeelden, bezoekfrequenties, buitenlandse bezoekers, herkomstenkaarten, vervoerstromen en HB-matrices kunnen lokaal, regionaal en landelijk en 24/7 worden gegenereerd. De toepassingsvelden zijn veel breder dan verkeer en vervoer alleen: retail, citymarketing, events en festivals, toerisme, economie en veiligheid (crowd control) zijn andere disciplines waarin deze mobiliteitsgerelateerde gegevens nieuwe inzichten kunnen genereren. Gsm-data zijn behulpzaam en kosteneffectief bij monitoring, evaluatie, beleidvorming en planning.

### Ophoogmethodiek

De vertaling van gsm-waarnemingen naar werkelijke aantallen mensen heeft flinke stappen gemaakt. Er is recentelijk door DAT.Mobility een ophoogmethodiek ontwikkeld waardoor inzicht in het gedrag van werkelijke aantallen mensen in geheel Nederland op continue basis met grote betrouwbaarheid nu mogelijk is.

Mobiele telefoongegevens hebben ook hun beperkingen. Eén van de beperkingen is de ruimtelijke (on)nauwkeurigheid van de gegevens, die voorlopig tot PC-4-niveau beperkt is. In Nederland mogen de ruwe gegevens vanuit privacyoverwegingen bovendien niet direct gebruikt worden. De gegevens zijn daarom alleen in geaggregeerde vorm beschikbaar, waardoor kleine aantallen (15 of minder) verplaatsingen niet waarneembaar zijn. Een andere beperking is dat uit de momenteel beschikbare gegevens maar in beperkte mate modaliteitinformatie is af te leiden en dan met name op lange afstanden. Dat neemt niet weg dat de landelijke dekking en de voortdurende, en hierna historische beschikbaarheid van de gegevens de mogelijkheden van toepassingen enorm vergroot.



**Figuur A: Het bezoekerseffect van de TT in Assen**

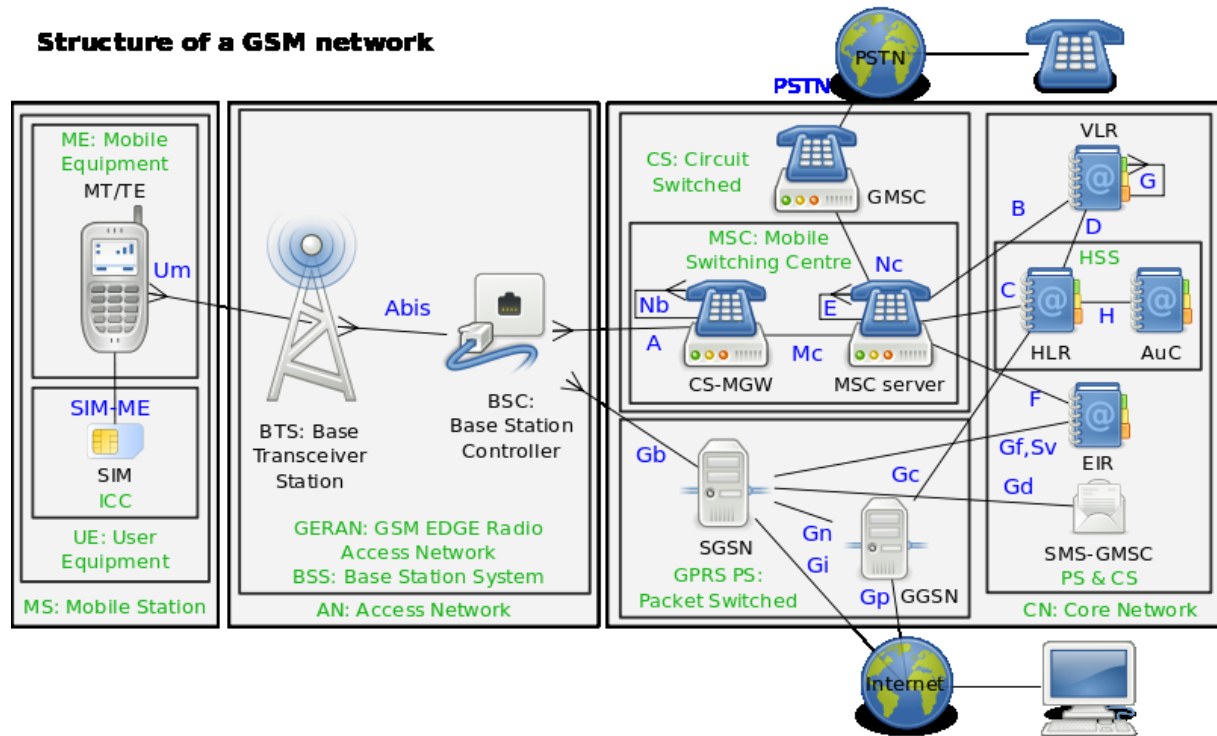
## 1. Over het meten van mobiliteit met gsm-data

Deze paper gaat in op mogelijkheden en beperkingen van gsm-data als proxy voor mobiliteitsinformatie en biedt een eerste overzicht van stand van zaken in Nederland. Mobiliteit dient – zoals uit het vervolg blijkt - in deze context meeromvattend dan alleen verkeerskundig te worden opgevat, zoals uit het vervolg blijkt.

### *Hoe gsm-telefoons<sup>1</sup> mobiliteit meten*

Mobiele telefoons, en in het bijzonder smartphones meten op veel manieren de locatie en de beweging van het apparaat. Met ten minste tien (combinaties van) sensoren voor GPS, WiFi, inertie, luchtdruk, Bluetooth, ultrasoon, en gsm-netwerkcontacten, worden aspecten van mobiliteit met verschillende frequenties en nauwkeurigheden gemeten. Allerlei locatiegerelateerde apps en (navigatie)systemen maken van deze sensoren gebruik.

In dit paper gaan we in op één van deze sensoren, namelijk gsm (Global System for Mobile Communications), dat sinds 1991 operationeel is en wereldwijd inmiddels door meer dan 3 miljard mensen in 210 landen wordt gebruikt. Specifiek kijken we naar de mogelijkheden en beperkingen die de gegevens uit dit systeem bieden voor het genereren van mobiliteitsgerelateerde informatie.



Figuur 1: structuur van het GSM netwerk

In Nederland is Vodafone het eerste, en tot nu toe enige telecombedrijf dat het aandurfde om zijn mobiele telefoondata commercieel te gebruiken en vermarkten. Samen met Mezuro, dat de ruwe gegevens van Vodafone tot basisgegevens verwerkt, is DAT.Mobility (voor zover bekend) het enige bedrijf dat informatiediensten op basis van deze mobiele telefoongegevens aanbiedt (onder de productnaam 'View.dat'). Binnen de academische wereld wordt al jaren onderzoek gedaan naar de toepassingsmogelijkheden van gsm-data. Waarschijnlijk omdat daar de privacyregels

<sup>1</sup> Ik schrijf in deze paper steeds over 'telefoons'. Hiermee worden ook andere devices met simkaarten zoals tablets bedoeld. Andere apparaten met simkaarten vallen hier niet onder.

minder strikt zijn is in de Verenigde Staten is de vermarkting van telefoondata al eerder ingezet. Op de Noord-Amerikaanse markt brengt het bedrijf AirSage gsm-data van de telecombedrijven Verizon en Sprint commercieel op de markt. In het Verenigd Koninkrijk levert INRIX op basis van telefoongegevens van telecomprovider O2 herkomst- en bestemmingsmatrices. In Estland past Positium gegevens uit mobiele telefoons toe in het domein van toerisme. Ongetwijfeld zijn er hiernaast meerdere partijen bezig om diensten op dit gebied te ontwikkelen<sup>2</sup>.

#### *De kenmerken van CDR-data (call-detail records)*

Wanneer een mobiele telefoon door bel-, sms- of andere data-activiteit contact maakt met het gsm-netwerk, c.q. een mast (een zogenaamde Base Transceiver Station) in dit netwerk, dan wordt deze activiteit door de telecomoperator vastgelegd in een zogenaamd *call detail record* of CDR. De nauwkeurigheid waarmee de locatie van de telefoon, c.q. de telefoonactiviteit kan worden vastgelegd, is onder meer afhankelijk van de technische infrastructuur van het netwerk van het telecombedrijf. Als het telecombedrijf een systeem heeft waarmee de signaalsterkte van meerdere masten in de omgeving tegelijk met de telefoonactiviteit kan worden vastgelegd, is het mogelijk om door middel van triangulatie de locatie van mobiele telefoons tot op zo'n 50-100 meter nauwkeurig vast te stellen. Als alleen informatie over de specifieke mast waarop de activiteit wordt afgehandeld bekend is, wordt de nauwkeurigheid waarmee de locatie van de telefoon kan worden bepaald aanzienlijk kleiner. Hoe nauwkeurig de plaatsbepaling is op basis van communicatie met één mast, is afhankelijk van de mastlocatie en de functie van de mast. In dichtbevolkte gebieden staan meer masten dan in dunbevolkte. Het geografische bereik van masten in dichtbevolkte gebieden is uit capaciteitsoverwegingen kleiner dan in dunbevolkte gebieden. Daardoor neemt de nauwkeurigheid van locatiebepaling in dichtbevolkte gebieden toe. Maar er zijn ook masten die een regiofunctie hebben en telefoonverkeer uit een groot gebied – dus ook een dichtbevolkt gebied – (kunnen) afhandelen. Het is bovendien niet per se voorspelbaar op welke mast de activiteit van een telefoon, die zich op een bepaalde locatie bevindt, wordt afgehandeld. Onderdeel van de geruchtmakende Deventer moordzaak is bijvoorbeeld een discussie over de bewijskracht van een opgevangen telefoonsignaal over de daadwerkelijke geografische aanwezigheid van de verdachte op de locus en tempus delict.

De maximale nauwkeurigheid van de locatiebepaling van een telefoon op basis van informatie van één mast is waarschijnlijk niet groter dan een postcode-4-gebied. Daarbij moet wel worden opgemerkt dat er speciale masten zijn die een bereik hebben in een klein, gespecificeerd gebied, bijvoorbeeld op stations of winkelcentra (small cells). Deze masten zijn vanzelfsprekend niet altijd en overall aanwezig omdat masten in principe geplaatst worden om het telefoonverkeer in een bepaald gebied te kunnen afhandelen. De verwachting is wel dat er op termijn meer small cells zullen verschijnen of dat omwille van de monitoring specifieke small cells al dan niet tijdelijk geplaatst gaan worden.

## **2. Mogelijkheden van gsm-data als mobiliteitsproxy**

---

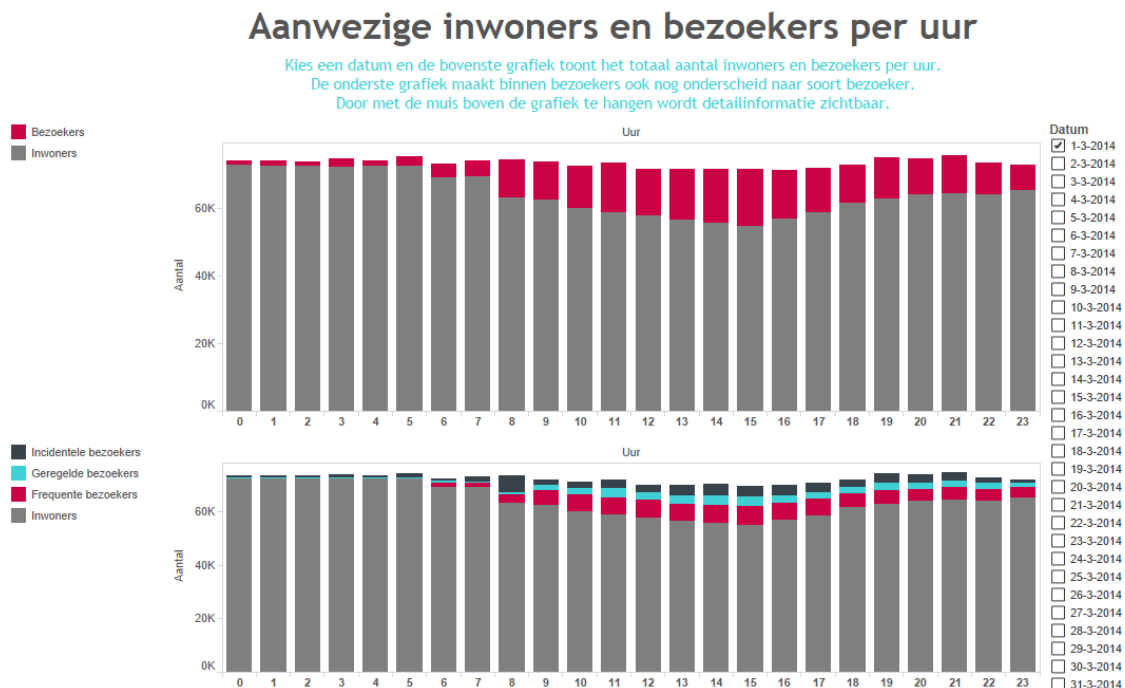
<sup>2</sup> Eerder gaf bijvoorbeeld het Franse Orange een dataset van Ivoorkust vrij waarmee een HB-matrix en eerste verkeersmodel van Ivoorkust is gemaakt (zie [http://www.goudappel.nl/media/files/uploads/Goudappel\\_-\\_KDD\\_Lab\\_Isit\\_-\\_D4D-book\\_Netmob\\_2013\\_MIT\\_Boston.pdf](http://www.goudappel.nl/media/files/uploads/Goudappel_-_KDD_Lab_Isit_-_D4D-book_Netmob_2013_MIT_Boston.pdf))

## 2.1 Drukbeelden (figuur 2)

Drukbeelden geven aan hoeveel personen aanwezig zijn of waren in een bepaald gebied. Druk is voor meerdere toepassingen interessant. Denk aan monitoring, evaluatie en beleidsvorming op gebieden als veiligheid, toerisme, citymarketing, grote venets en vervoers- en faciliteitsplanning.

Een eerste toepassing van CDR gegevens is het in beeld brengen van drukte. Naarmate er meer telefoons in een bepaald gebied worden waargenomen, is het er (waarschijnlijk) drukker, zo is het uitgangspunt. Bij de kroning van Willem-Alexander en Maxima zijn telefoongegevens gebruikt om de actuele drukte in het centrum van Amsterdam in kaart te brengen waarbij crowd control één van de achterliggende doelen was. Real-time druktebeelden maken overigens van een andere bron dan CDR gebruik, namelijk informatie over de actuele drukte op verschillende masten.

Drukbeelden zijn met name voor evenementen zeer inzichtelijk, zeker waar het gaat om gratis evenementen. Bij betaalde evenementen is uit de kaartverkoop of telineformatie bij de poort meestal goed te bepalen hoeveel bezoekers er waren, c.q. hoe druk het is geweest. Bij gratis evenementen zijn de drukteschattingen vaak gebaseerd op indrukken waarbij onder- en overschatting op de loer liggen, zoals bij demonstraties vaak blijkt als de schattingen van de politie en de schattingen van de organisator naast elkaar worden gelegd. Het grote voordeel van druktebeelden op basis van CDR data is dat deze in principe 24/7 beschikbaar zijn en op een consistente wijze gemeten worden, waarbij ook druktepatronen gedurende het evenement zichtbaar worden.



Figuur 2. Aanwezige inwoners en bezoekers per uur op 1 maart 2014 in Deventer

## 2.2 Herkomsten en bestemmingen (figuur 3)

Telefoons voegen een extra dimensie toe aan drukte: namelijk verplaatsing. Dat is de tweede toepassing van CDR-data. Telefoons verplaatsen zich gewoonlijk samen met hun eigenaar verplaatsen, en dus wordt met de telefoonactiviteit een 'gebruikersspoor'

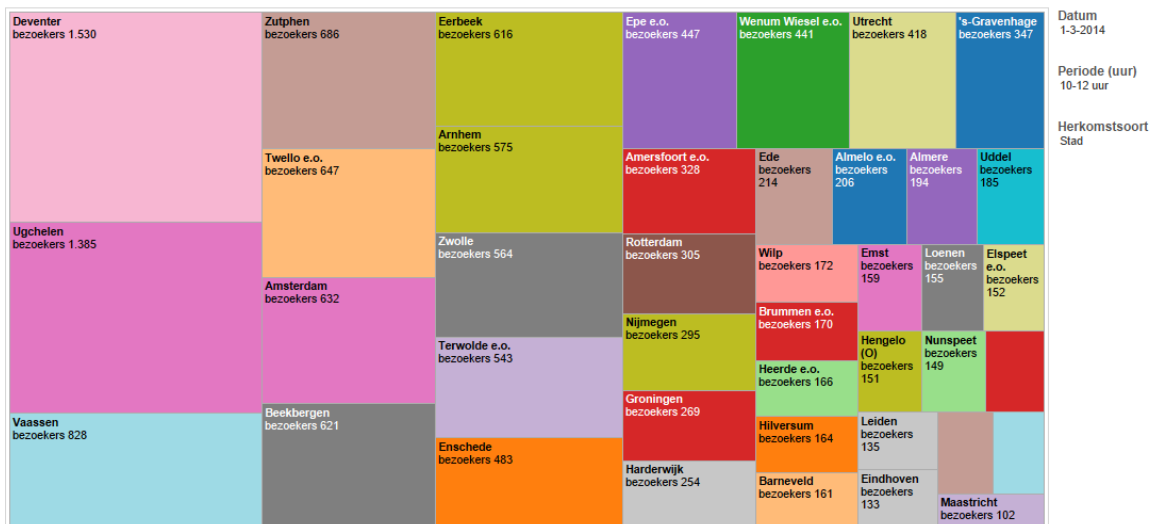
achtergelaten. Dit spoor toont de momenten en de plaatsen waarop de telefoon contact had met een mast.

Het aantal plaatsen waarop het gebruikersspoor zichtbaar is, is onder meer afhankelijk van het type telefoon. Een klassieke gsm die niet gebruikt wordt maar wel aanstaat, zal zich enkele keren per dag melden in het netwerk. Smartphones laten een veel rijker spoor (150-200 datapunten per dag) na dan klassieke gsm-telefoons, omdat deze niet alleen bij bellen en sms'en contact maken met het netwerk, maar ook bij synchronisatie van email, internetgebruik en wanneer met apps via het netwerk wordt gecommuniceerd. Uit de gsm-data kunnen de herkomsten en de bezoekplaatsen van gebruikers gedestilleerd worden, en daarmee biedt gsm een goede bron voor dynamische en actuele herkomst en bestemmingsmatrices.

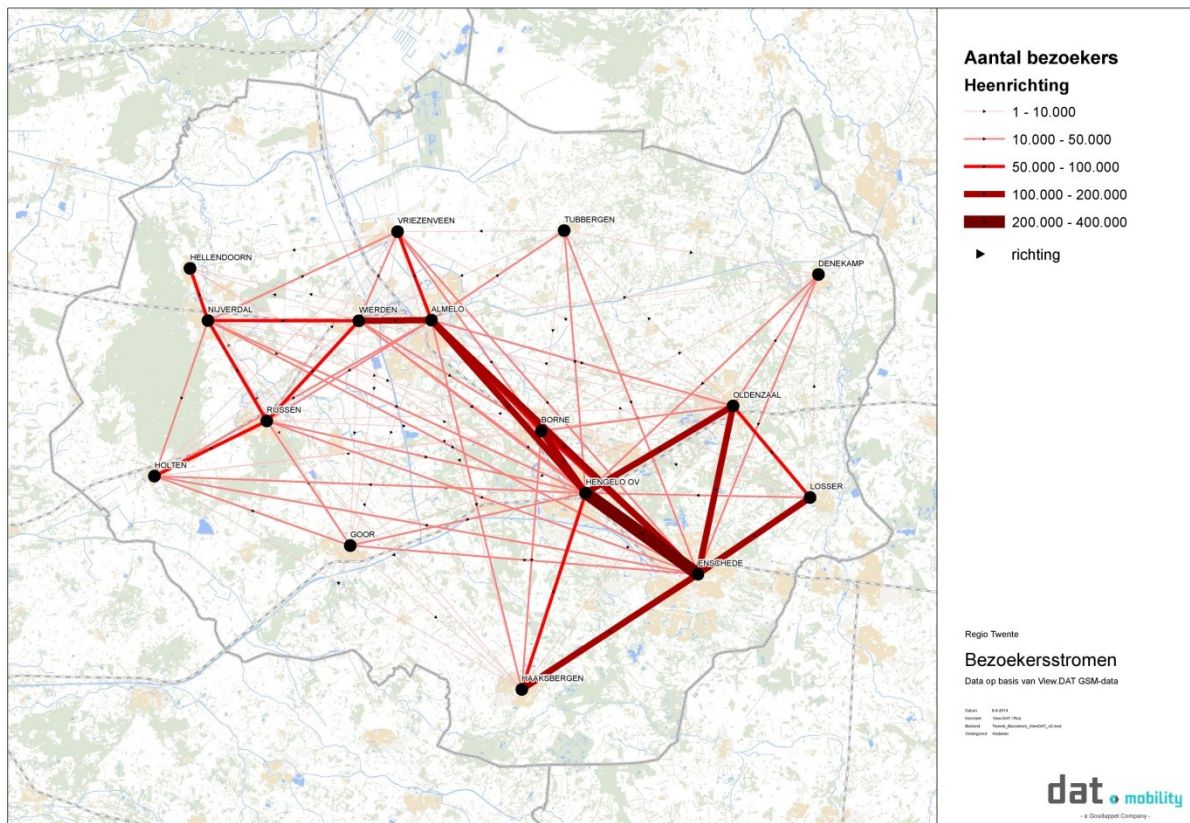
GSM data zijn daarmee een uitstekend antwoord op de roep naar meer gegevens over de actuele situatie in verkeersmodellen. De synthetische matrices die nu voor verkeersmodellen gebruikt worden, zijn veelal geschat op relatief oude socio-economische gegevens. Een schaa sprong is dus mogelijk, waarbij ook aannames over de drukteverdeling over het etmaal, over de week en de verschillende dagen van de week feitelijk getoetst of afgeleid kunnen worden. Een bijkomend voordeel van is de uniformiteit van de meetmethodiek, waardoor landsdekkende HB-matrices gemaakt kunnen worden, en waardoor verschillen aan de randen van verschillende regionale matrices eenduidig kunnen worden opgelost.

## Aantal bezoekers obv herkomst (treemap)

Kies een datum en herkomstsoort. De 'treemap' toont het aantal bezoekers.  
 Hoe groter de herkomstlocatie is, hoe meer bezoekers van deze locatie afkomstig zijn.  
 Door met de muis boven de grafiek te hangen wordt detailinformatie zichtbaar.



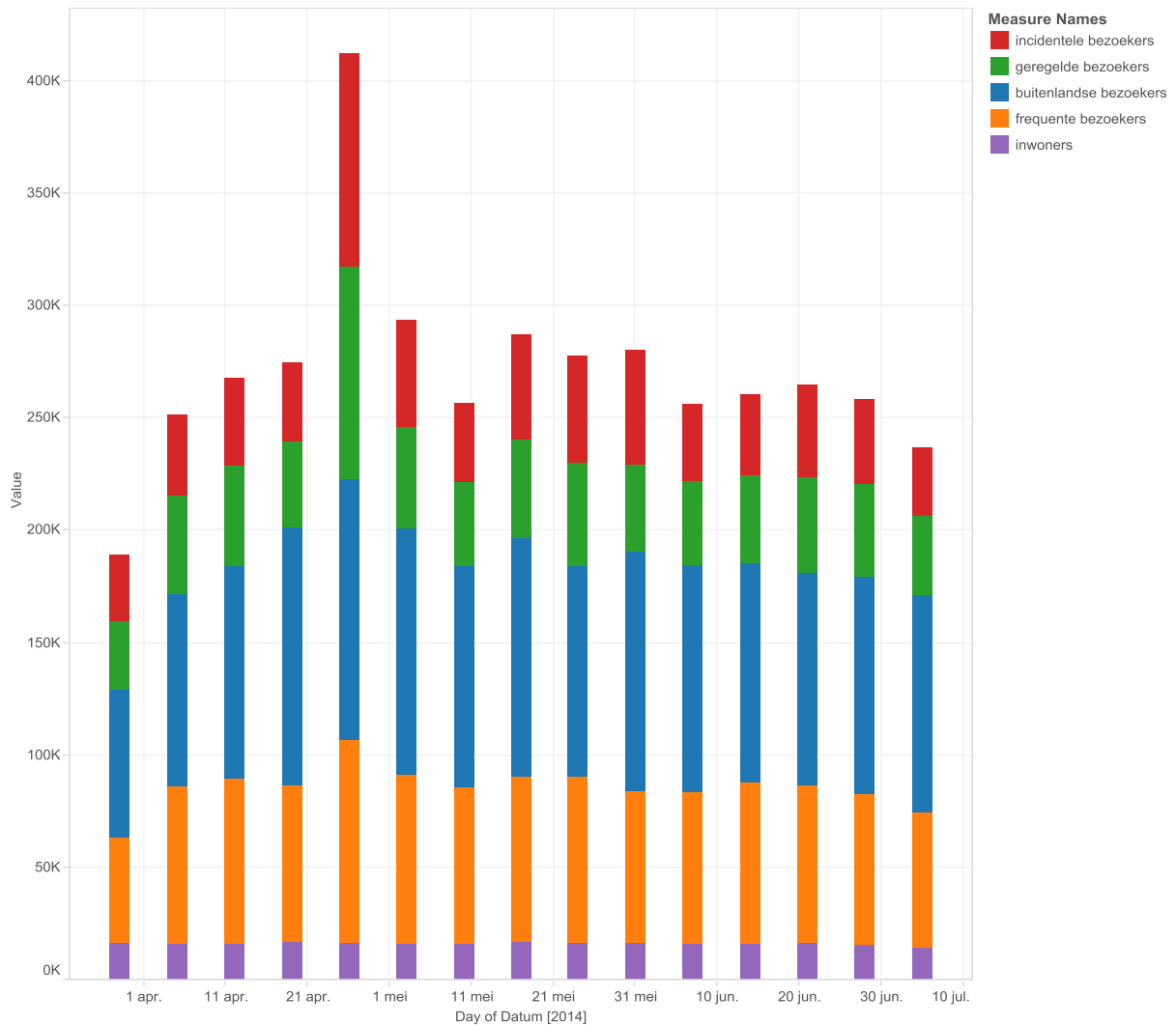
Figuur 3. Herkomsten van bezoekers aan Apeldoorn op 1 maart 2014 tussen 10-12 uur.



Figuur 4: Voorbeeld van bezoekersstromen in Twente gebaseerd op gsm-data

### 2.3 Bezoekersprofielen (figuur 5)

Nog meer inzicht ontstaat als CDR-data over een langere periode beschikbaar zijn. Dan wordt het mogelijk om hieruit kenmerken van de gebruiker van de telefoon af te leiden. Als een telefoon in een periode van een maand 's nachts stevast in plaats A wordt waargenomen, dan mogen we met een hoge mate van zekerheid aannemen dat A de woonplaats is van deze gebruiker. Met dat gegeven kunnen we ook kijken waarheen en hoe vaak deze gebruiker zich naar plaatsen buiten A verplaatst. Plaatsen die met een hoge frequentie (>10 x per maand) worden bezocht zijn met een behoorlijke waarschijnlijkheid werk- of schoolgerelateerde locaties. Als we de verschillende frequenties waarmee bezoeken aan locaties gebracht worden van elkaar onderscheiden, dan kunnen we een beeld creëren van het type bezoeker van bepaalde locaties, c.q. van bepaalde gebeurtenissen en evenementen.

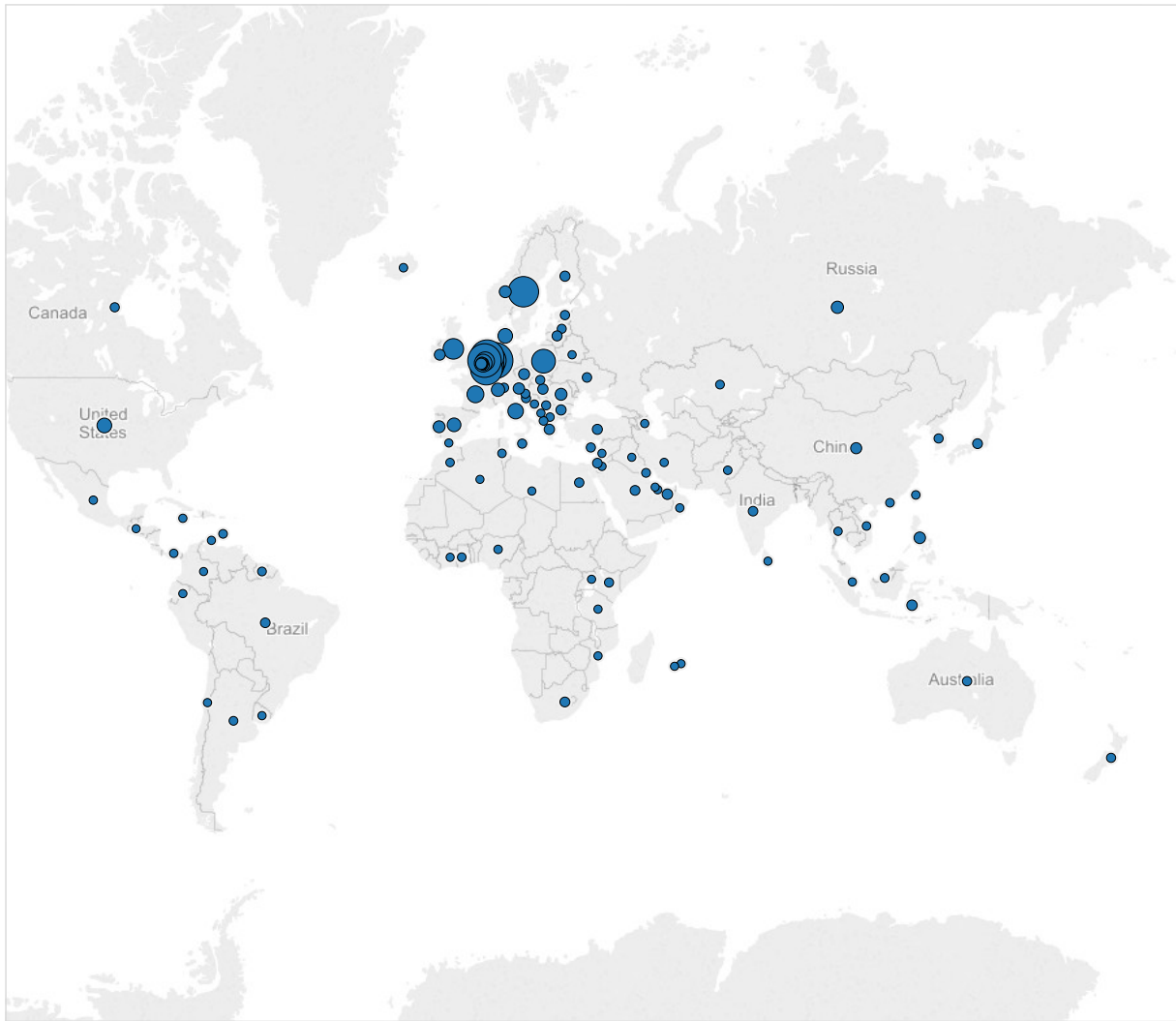


Figuur 5. Bezoekers van Amsterdam-Centrum op zaterdagen naar frequentie (inclusief buitenlanders).

#### 2.4 Buitenlanders (figuur 6)

Het gedrag en de interesses van buitenlandse bezoekers voor verschillende plaatsen en evenementen in Nederland is onder meer vanuit het oogpunt van toerisme en evenementorganisaties interessant. CDR leveren ook gegevens over de locatie van buitenlandse telefoons in Nederland, en bieden daarmee een goede benadering voor de locaties van bezoekers uit het buitenland. Over een langere periode gemeten wordt duidelijk dat uit bijna alle landen van de wereld telefoons worden waargenomen in de dataset.

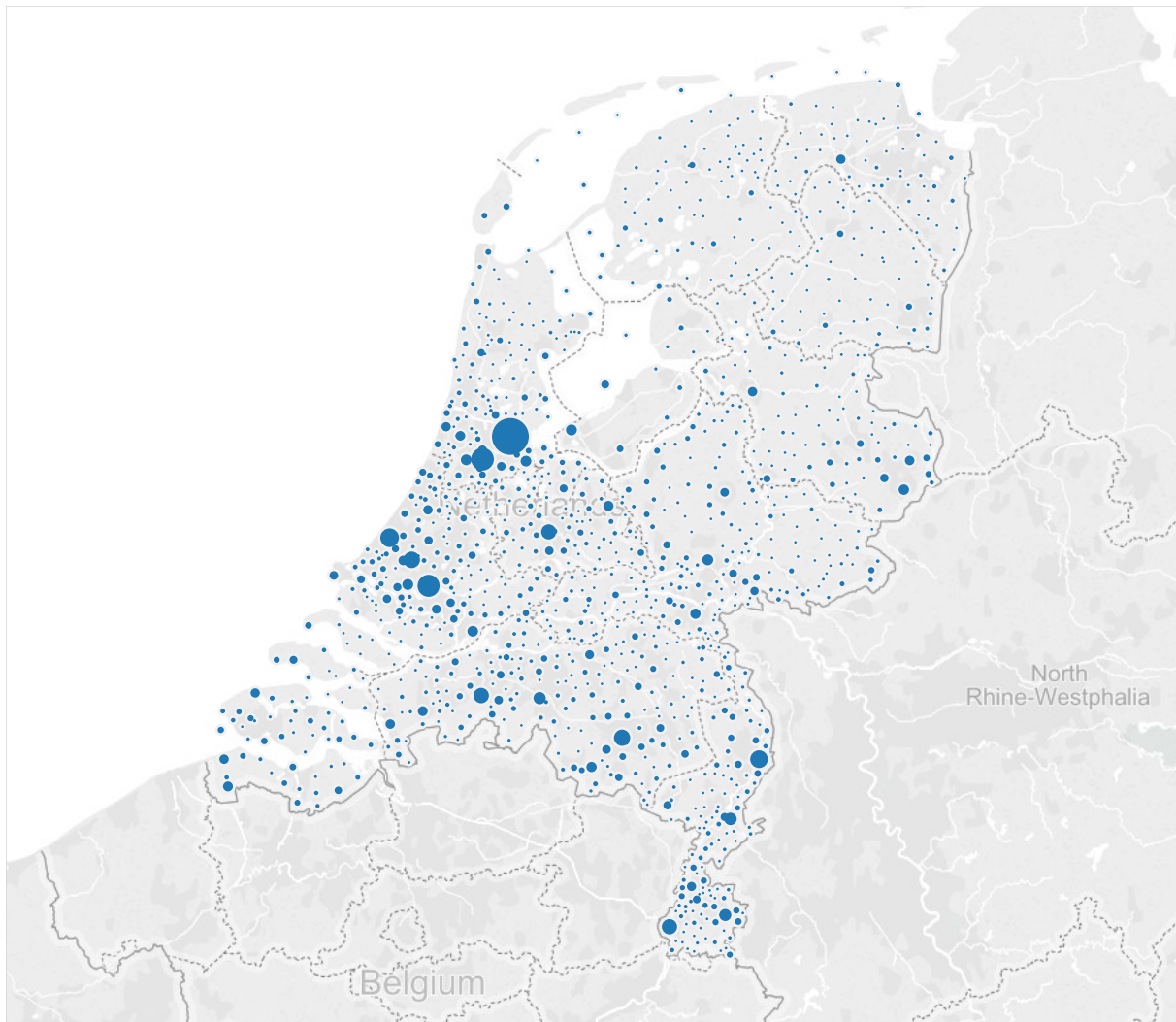




Figuur 6: Herkomstlanden van buitenlandse telefoons in de Nederlandse gsm/data.

### 2.5 Vervoerstromen

Met gsm-data kunnen we integrale vervoersstromen in beeld brengen zoals eerder duidelijk werd. Door voortdurend de locaties van telefoons vast te leggen en hieruit op maandbasis de 'woonplaats' van gebruikers te bepalen, is het mogelijk pendelstromen tussen plaatsen te bepalen. Het grote voordeel van deze stromen is dat ze niet onderscheiden zijn naar modaliteit en dus een compleet beeld geven van het aantal verplaatsingen tussen verschillende plaatsen. En dat kan dus voor elke onderscheiden plaats in Nederland. Voor buitenlanders is alleen het land van herkomst bekend, maar kan wel gekeken worden welke plaatsen in Nederland ze bezoeken (figuur 7).



Figuur 7: De bezoeklocaties van alle buitenlandse gasten in Nederland

### 2.6 Snelheid en reistijd

Een minder gebruikte toepassing van gsm-data is het berekenen van snelheid en reistijd. Dit is in principe wel mogelijk omdat van elke telefoonactiviteit de tijd en locatie bekend zijn. Van deze mogelijkheid wordt geen of weinig gebruik gemaakt omdat via andere ruimschoots voorhanden bronnen (HERE, TomTom, Be-Mobile) op GPS gebaseerde snelheden en reistijden aanzienlijk nauwkeuriger te bepalen zijn.

### 2.7 Verblijfstijden en activiteitenpatronen

Als het moment en de duur van verplaatsingen (globaal) in kaart kunnen worden gebracht met gsm-data, dan is het ook mogelijk de tussenliggende periodes, waarin sprake is van activiteiten op locaties, vast te leggen. Daarmee is het dus mogelijk om verblijfstijden en activiteitenpatronen van mensen te bepalen. In deze gebieden vinden momenteel nieuwe ontwikkelingen plaats die de bruikbaarheid van de data verder zullen vergroten.

### 2.8 Gebruikersprofielen

Als we meer weten over de kenmerken van reisgedrag, bezoekfrequentie en de kenmerken van bezoeklocaties, is het ook mogelijk om gebruikersprofielen op te stellen. Denk aan gebruikers die regelmatig op zaterdag doe-het-zelfcentra bezoeken of

gebruikers die sportevenementen bezoeken. Voor marketingdoeleinden zijn dit waardevolle gegevens. De in Nederland gebruikte gsm-gegevens laten dergelijke analyses niet toe, maar in de VS laat telecomprovider Verizon zien dat het mogelijk is. In Nederland is dergelijke profilering vanuit privacy-overwegingen niet mogelijk.

### *2.9 Vervoerwijze*

Omdat het in principe mogelijk is om modaliteiten uit gsm data af te leiden, wordt hier relatief veel energie in gestoken om dit voor elkaar te krijgen. Vooral op langere afstanden lijkt het mogelijk om auto van trein en vliegtuig te onderscheiden. Vanwege de grofmazigheid van de in Nederland beschikbare CDR-data is het niet mogelijk om langzaam verkeer (voetgangers en fietsen), openbaarvervoer en autoverkeer over kleinere afstanden (stedelijk niveau) van elkaar te scheiden.

## **Kenmerken van CDR data**

### *Actualiteit*

Een groot voordeel van gsm-data is hun actualiteit. Eigenlijk wordt er voortdurend en actueel gemeten waardoor altijd en in principe zeer snel over gegevens beschikt kan worden. Drukbeelden zijn indien gewenst actueel te leveren, maar om een interpretatie van gegevens naar woonplaats, bezoekfrequentie, en bestemmingen te kunnen maken, zijn gegevens over een langere periode noodzakelijk, bijvoorbeeld een maand. Hierdoor kunnen dergelijke gegevens pas op zijn vroegst een maand later geleverd worden.

### *Representativiteit*

De omvang van de gemeten steekproef is zo groot dat in eerste instantie wel eens gedacht wordt dat de statistiek over boord kan. Eindelijk weten we hoe het echt in elkaar zit. Immers, een steekproef van miljoenen gebruikers trek je niet dagelijks en waargenomen verschillen zijn altijd 'statistisch significant'. Onder betrouwbaarheid en validiteit van de data zullen we evenwel zien dat het zo eenvoudig helaas niet is.

### *Methodische vergelijkbaarheid/continuïteit*

Een ander groot voordeel is dat de meetmethode vastligt, het meetinstrument verandert maar in beperkte mate en is conceptueel eenduidig waardoor een voortdurende stroom van goed vergelijkbare data vrijkomt.

### *Historie*

Door de data vast te leggen in databases ontstaat een rijke bron waarmee ook historische analyses gedaan kunnen worden. Het komt vaak voor dat we de effecten van een maatregel of gebeurtenis achteraf willen weten, maar geen goede gegevens voor een voor- of nameting beschikbaar hebben. De gsm-data zijn 24/7 beschikbaar en lossen dit probleem (zeker op termijn) op.

### *Betrouwbaarheid en validiteit*

Door het enorme aantal meetpunten, zowel qua gebruikers als qua spreiding in de tijd en in de ruimte, is de betrouwbaarheid van de data zeer groot. De onbetrouwbaarheid die de data bezit is voornamelijk een gevolg van de bewerkingen op en filteringen van de data, en technische storingen in en wijzigingen van het gsm-systeem, denk aan verplaatsingen van masten. De bewerkingen en filters zijn in eerste instantie noodzakelijk om anonimiteit en privacy van de gegevens te waarborgen.

Of de data een valide meting van mobiliteit vertegenwoordigen staat buiten kijf, zolang we aannemen dat mobiele telefoons zich met hun gebruikers verplaatsen.

#### *Anonimiteit & Privacy*

De anonimiteit, c.q. de privacy van mobiele telefoongegevens is - zeker in de Europese context - een heet hangijzer. De deskundigen zijn het er over eens dat de definitie van het begrip privacy en de wet- en regelgeving hierover, als het gaat om het gebruik van mobiele telefoondata, onvoldoende zijn. Dit betekent niet dat er geen definitie is of dat er geen wet- en regelgeving is. In Nederland is er een Wet Bescherming Persoonsgegevens. Het College Bescherming Persoonsgegevens (CBP) ziet erop toe dat de persoonsgegevens zorgvuldig worden gebruikt en de privacy ook in de toekomst gewaarborgd blijft. Binnen dat kader is er door Vodafone voor gekozen de ruwe gegevens (CDR) niet aan derden beschikbaar te stellen. Om gebruik te kunnen maken van de gegevens worden daarom binnen Vodafone bewerkingen op de data uitgevoerd.

De privacyborging verloopt via een stringent protocol. Ten eerste wordt de identificerende informatie (telefoonnummer) versleuteld. De versleutelde informatie mag daarnaast maximaal een maand worden bewaard. In de volgende maand wordt een nieuwe versleuteling toegepast. Hierdoor kunnen de bewegingen van eenzelfde telefoon nooit langer dan een maand herkend worden. Een belangrijkere actie in het proces is de 'minimaal 16 regel'. Deze houdt in dat op iedere vraag die aan de data gesteld wordt, het antwoord minimaal '16' moet zijn. Dus als gevraagd wordt hoeveel telefoons op locatie A worden waargenomen tussen 6-7 's ochtend, dan moet het antwoord ten minste 16 zijn, anders wordt geen informatie geleverd. Op deze manier zijn nooit gegevens van individuele telefoons te achterhalen.

### **3. Beperkingen van mobiele telefoondata**

#### *Gsm-data zijn niet open en niet gratis*

Soms worden gsm data verward met open data. Het kenmerk van open data is dat hierbij gestreefd wordt naar een zo groot en goedkoop mogelijke ontsluiting, liefst gratis. Veelal gaat het daarbij om data die zijn verzameld door overheden (en waarvoor dus al met belastinggeld is betaald), of data die door vrijwilligers of belangenorganisaties worden verzameld en die gezien worden als gegevens die van algemeen nut kunnen zijn. Over het eigenaarschap van gsm-data lopen nog wel discussies. Het is te eenvoudig om te stellen dat de telecomoperator de enige en wettige eigenaar is van de data. Immers, je kunt net zo goed stellen dat de individuele gebruikers eigenaar zijn van 'hun' data. Hoe dan ook, voor het verzamelen en opslaan van de data zijn dure systemen nodig. Tevens zijn de data privacygevoelig. Dit zijn twee duidelijke redenen waarom gsm-data - zeker in ruwe vorm - geen open data zijn. In Zuid-Korea wordt overigens wel onderzocht of de overheid in de omvorming van ruwe data een rol zou kunnen spelen, waardoor de informatie uit deze gegevens wél als open data beschikbaar zou kunnen worden gesteld.

#### *Granulariteit*

De dekking van het gsm-netwerk is wereldwijd en in Nederland in het bijzonder uitstekend waardoor over bijna alle locaties informatie uit de data verkregen kan worden. Maar de granulariteit is relatief grof, waardoor vragen die over kleine locaties (kleiner dan een PC4-gebied) moeilijk of niet mee te beantwoorden zijn. Door de uitrol van 4G neemt de granulariteit overigens nog wel toe de komende tijd en ook met small-cells zijn grotere ruimtelijke nauwkeurigheden lokaal haalbaar. Een wezenlijke stap naar

nauwkeuriger data kan plaatsvinden indien met triangulatie gewerkt zou kunnen worden, maar dit brengt grote investeringen voor de telecomoperator met zich mee.

### *Routes*

*Vanwege de ruimtelijke grofheid van gsm/gegevens is het niet mogelijk om uit de gegevens direct routes af te leiden. Een inschatting van gebruikte routes kan natuurlijk gemaakt worden met verkeersmodellen en routezoekalgoritmes. Werkelijke routes kunnen alleen betrouwbaar met GPS metingen worden bepaald. Voor OV-reizigers kunnen OV-chipcarddata gebruikt worden om routes vast te stellen.*

### *Ophoging*

Een belangrijke beperking is dat we met gsm/data in principe nog steeds geen mensen tellen, maar telefoons waarnemen. CDR-data zijn een afspiegeling van telefoonactiviteit gespreid in de ruimte en in de tijd. Dit is niet vanzelfsprekend en één-op-één te vertalen naar werkelijke aantallen mensen. Niet iedereen heeft of gebruikt een (smart)Phone en er zijn meer vertekeningen in de data die het gevolg zijn van heel verschillende fenomenen. Met ophogingsmethodieken moet die vertaling naar aantallen mensen opgelost moeten worden. DAT.Mobility heeft met name hier haar research en ontwikkeling in het afgelopen jaar op geconcentreerd.

*Verschillen in telefoonbezit:* Hoewel in Nederland 21,7 miljoen mobiele telefoons zijn (gemiddeld 1,3 mobiele telefoon per persoon) en Vodafone 5,3 miljoen abonnees heeft, is het mobiele telefoonbezit en -gebruik niet evenredig over de bevolking verdeeld. Zo hebben kleine kinderen meestal nog geen mobiele telefoon en maakt het oudste deel van de bevolking aanzienlijk minder gebruik van mobiele telefoons. Dan is er nog een groep jongeren die hun smartphone alleen nog maar via WiFi gebruiken. Dergelijke telefoons worden niet waargenomen op het gsm-netwerk. Daarnaast is het nog zo dat een smartphone andere datasporen achterlaat dan een klassieke gsm-telefoon waardoor verschillende nauwkeurigheden in de mobiliteitsbeelden ontstaan.

- *De Vodafone gebruikers- of abonneedichtheid is niet identiek voor verschillende plaatsen.*
- *Niet iedereen is in Nederland:* Soms zijn mensen op reis in het buitenland. Vooral in de zomervakantieperiode zijn er daardoor beduidend minder telefoons in Nederland.
- *Niet iedereen heeft zijn telefoon aan staan:* als een telefoon uit staat wordt deze niet (meer) waargenomen.
- *Niet iedereen gebruikt zijn telefoon (evenveel):* als een telefoon wel aanstaat wordt deze niet altijd in dezelfde mate gebruikt en wordt daarna niet per se op dezelfde manier waargenomen.
- *Het gebruik van telefoons is tijdsafhankelijk:* 's nachts wordt er over het algemeen minder gebeld dan overdag.
- *Het gebruik van mobiele telefoons is plaatsafhankelijk:* mensen die buitenshuis zijn gebruiken hun telefoon vaker dan mensen die binnenshuis zijn.
- *Van ruwe data naar tellingen;* niet de ruwe data (CDR) maar alleen tellingen van meer dan 15 mensen worden geregistreerd. Hierdoor worden gegevens op rustige plaatsen en tijdens rustige perioden in absolute zin onderschat.

### *Anonimiteit & Privacy (2)*

We hebben al eerder aangegeven welke stappen worden ondernomen om de data te anonimiseren. Omdat de versleutelde identiteit van dezelfde telefoon maximaal een

maand mag worden bewaard, moeten profielen van mensen gebaseerd worden op data van maximaal die duur. Het gevolg daarvan is dat het vaststellen van een woonplaats in vakantieperiodes, wanneer mensen vaak langer dan twee weken niet thuis zijn, vertekend wordt in de richting van vakantieadressen.

#### *Buitenlanders (2)*

De versleutelde identiteit van buitenlandse telefoons mogen niet langer dan één dag bewaard worden. Hierdoor is het niet mogelijk om bezoekfrequentie of verblijfsduur van buitenlanders in Nederland uit de data af te leiden.

#### **4. Conclusies**

Mobiele telefoondata zijn een nieuwe en veelbelovende databron als mobiliteitsproxy. De voortdurende en historische beschikbaarheid en landsdekkendheid van de data en de eenduidige dataverzamelingstechniek bieden grote voordelen voor de toepassing van de informatie die hierop gebaseerd is als beslissingsondersteunend en monitoringsinstrument. Gsm-data elimineert potentieel discussies die het gevolg kunnen zijn van methodische verschillen in meetinstrumenten.

Tegelijk moet worden gewaakt voor een euforie rondom het gebruik van deze data als panacee voor allerlei meetproblemen. Het is goed te blijven beseffen dat de telefoondata een *proxy* zijn voor mobiliteit en geen directe meting van mobiliteit.

Verwacht mag worden dat het gebruik van dergelijke data in de komende jaren een vlucht zal nemen, dat de kwaliteit en de nauwkeurigheid ervan verder zullen toenemen en dat als gevolg van de beschikbaarheid van deze data nieuwe toepassingen zullen ontstaan.