

Rotterdams verkeersmodel door verrijking met GSM-data nu nog beter

ing. Stefan de Graaf Msc. – Goudappel Coffeng – sdgraaf@goudappel.nl
ir. Klaas Friso – DAT.Mobility – k.friso@dat.nl
Jeroen Rijdsdijk – Metropoolregio Rotterdam Den Haag – j.rijdsdijk@rotterdam.nl

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

Samenvatting

De herkomst-bestemmingsmatrix (HB-matrix) is de kern van ieder strategisch verkeersmodel, maar in de basis is en blijft het een *wiskundige schatting* van verplaatsingspatronen. De kwaliteit ervan is, ondanks wijdverbreid, beperkt vast te stellen. Het toevoegen van, of confronteren met, meer data over feitelijke verplaatsingen pakt die beperking aan. Dit paper beschrijft een onderzoek naar het verbeteren van de (HB-matrix) van verkeersmodellen door het toepassen van waargenomen verplaatsingen vanuit GSM-data.

Het moet ook beter. De omissies in de HB-matrixschatting worden prangender nu de toepassingsfocus verschuift van de strategische lange termijn (het plannen van nieuwe grootschalige infrastructuur) naar de kortere termijn (met lokale aanpassingen en gedragsinterventies knelpunten aanpakken of voorkomen). En dat is volop aan de hand. Nederland is áf, we zitten in het tijdperk van Beter Benutten. Een goede 'ijking' op de huidige situatie wordt steeds belangrijker.

Moderne mobiele telefoons maken enkele honderden keren per dag contact met GSM-masten. Hierdoor zijn locatie en beweging van mobiele telefoons waarneembaar waarmee we vervolgens –uit datasets met een enorme steekproef– mobiliteitspatronen kunnen afleiden. Dat met GSM-data een verkeersmodel is op te bouwen in gebieden zonder kennis over verplaatsingspatronen was al bekend, maar wat was de kwaliteit ervan? Voor het eerst confronteren wij deze data nu daarom aan één van de meest omvangrijke strategische, multimodale verkeersmodellen, die in Nederland ook daadwerkelijk dagelijks gebruikt worden: het Regionale Verkeersmodel Rotterdam (RVMK3.1).

De state-of-the-practice GSM-data die momenteel voor mobiliteitsonderzoek beschikbaar is blijkt (nog) geen ideale databron om nu al volledig data-driven de H/B-matrix af te leiden. Het detailniveau van de zonering is bijvoorbeeld grofmaziger dan in de meeste verkeersmodellen. Onderdeel van het onderzoek was een nadere verkenning van de kwaliteit, dekkingsgraad en zeggingskracht van de huidige databron, met als leitmotiv: wat er daarmee wél aan verbetering van een verkeersmodel mogelijk is en wat de meerwaarde daarvan is.

Het leidde tot een sterk verbeterd inzicht in verplaatsingspatronen in de regio Rotterdam. Vooral op regionaal niveau zijn bekende of vermoede 'modelfouten' inderdaad te corrigeren tot een beeld dat dichter bij de werkelijkheid ligt. Zo lijken steden 'die gericht zijn op' echt te bestaan, Zoetermeer trekt bijvoorbeeld meer dan modelmatig verwacht naar Den Haag. Indien de technische ontwikkeling de komende jaren voortzet, is de weg naar verbeterde en meer data gedreven verkeersmodellering definitief ingezet.

1. Big-data scherpt feitenbasis verkeersbeleid aan

1.1 Verkeersmodellen als onderzoekstool

Regeren is vooruitzien, zeker op het gebied van mobiliteit en daarmee samenhangende verkeersstromen. Infrastructuur aanleggen kost tijd en ligt er voor jaren. Nederland heeft daarom een rijke traditie op het gebied van verkeersmodellen: kwantitatieve 'voorspellers' van mobiliteitsstromen. De werking van een verkeersmodel is gebaseerd op feitelijkheden over (geplande) ruimtelijke ordening, wetenschappelijke inzichten over verplaatsingsgedrag en zo realistisch mogelijke schattingen van mobiliteitsbepalende factoren (bevolkingsgroei, geografische spreiding van arbeidsplaatsen etc.).

De kern van een verkeersmodel is een HB-matrix: een tabel waarin alle verplaatsingen tussen gebieden in het verkeersmodel zijn opgenomen. De HB-matrix wordt geschat op basis van de hiervoor genoemde bronnen. Het is een *wiskundige schatting* en waarvan de interne structuur feitelijk tot op heden nauwelijks toetsbaar is aan de realiteit. In die zin kunnen we stellen dat in verkeersmodellen, hoewel wijdverbreid toegepast, juist de kwaliteit van het belangrijkste onderdeel de zwakst toetsbare schakel is.

Nu omgaan met verkeer steeds meer een vraagstuk wordt voor de korte termijn en met een hoger detailniveau, wordt het des te belangrijker dat verkeersmodellen een goede 'ijking' hebben op de huidige situatie. Nederland is namelijk 'af'. We leven in het tijdperk van gedragsverandering en kleine lokale aanpassingen om knelpunten op te lossen en te voorkomen. Lange tijd is de relatief onbekende kwaliteit van de verkeersdistributie als een gegeven beschouwd. Budgettair en methodologisch zijn er namelijk grenzen aan wat er met enquêtes en steekproeven aan feitenbasis kan worden verzameld. Met de komst en ontsluiting van grootschalige registratie van individuele verplaatsingspatronen kan deze status-quo doorbroken worden. Schatten wordt meten [1,3] en dat kan een enorme kwaliteitsimpuls voor verkeersmodellering betekenen.

1.2 De eerste stap naar data-driven verkeersmodellen

Goudappel Coffeng bouwde ongeveer twee jaar geleden voor Ivoorkust al een HB-matrix met GSM-gegevens die de actuele mobiliteit weergeeft, zowel nationaal als binnen de hoofdstad Abijan [2]. Voor een land zonder betrouwbare statistieken is dat een enorme stap voorwaarts. Een recente prijswinnende analyse in Senegal borduurde voort op dit principe [4]. In deze situaties konden we echter niet de kwaliteit van de toepassing toetsen omdat daarvoor geen data beschikbaar was.

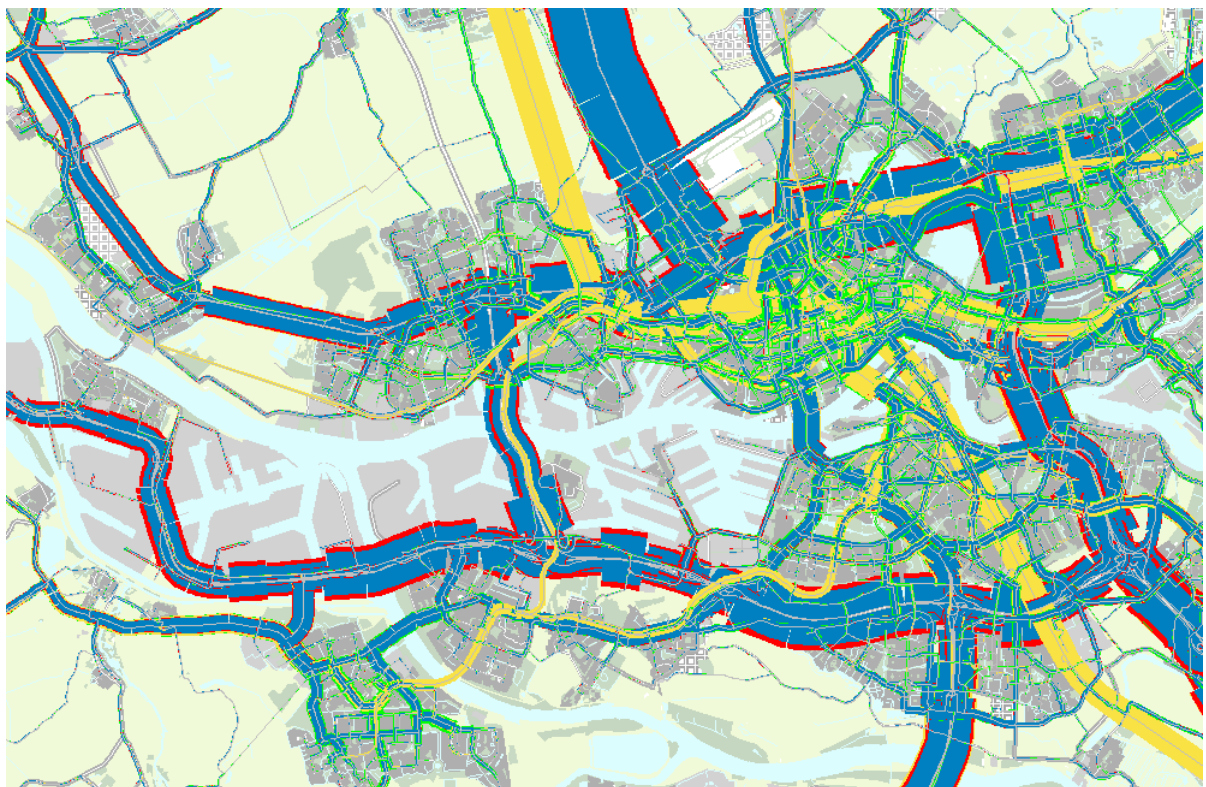
In Nederland is die data er wel. Onze huidige verkeersmodellen beschrijven al behoorlijk goed de daadwerkelijke verplaatsingspatronen. Die situatie maakt het mogelijk –anders dan in de Afrikaanse cases– gericht te kijken naar de kwaliteit van HB-matrices zoals die kunnen worden opgebouwd met Nederlandse GSM-data. We hebben immers legio verkeersmodellen operationeel met een uitgebreid toetsingskader aan enquête- en teldata. Hoewel we verwachten dat ook hier op termijn volledige data-driven verkeersmodellen mogelijk worden, is dat moment nu nog niet bereikt.

Met GSM-data is het al wel mogelijk een grote bron van onzekerheidsmarges in de huidige generatie modellen, de verkeersdistributie, te vervangen door gemeten data. De verkeersdistributie is de interne structuur van de HB-matrix die beschrijft waar de verplaatsingen hun herkomst en bestemmingen vinden. Een logische stap is dus om nu te kijken hoe we een operationeel verkeersmodel in Nederland kunnen verrijken met verplaatsingsgegevens uit GSM-data. Daarvoor hebben Goudappel Coffeng, DAT.Mobility en de Metropoolregio Rotterdam Den Haag een pilot uitgevoerd die we in dit paper beschrijven: om te laten zien dát het kan, wát er precies kan en wat daarvan de toegevoegde waarde is voor betere beleidsbeslissingen over verkeer en ruimte.

2. De casestudy

2.1 *Het verkeersmodel*

Het verkeersmodel dat voor deze case is gebruikt betreft de RVMK3.1 Rotterdam, het regionale verkeersmodel dat de voormalige stadsregio Rotterdam als studiegebied heeft. Het totale model heeft een omvang van bijna 6.000 zones en geldt daarmee als een van de meest omvangrijke strategische verkeersmodellen in Nederland. De zonering in het studiegebied is ongeveer op postcode-5 niveau. Het multimodale verkeersmodel wordt voor tal van zaken ingezet, zoals regionale infrastructuurstudies, openbaar vervoerstudies en bestemmingsplananalyses. Afbeelding 2.1 geeft een overzicht van de gemodelleerde vervoersstromen in een deel van de regio Rotterdam waarbij de kleuren verschillende modaliteiten aangeven (blauw = auto, rood = vracht, groen = fiets en geel = openbaar vervoer).



Afbeelding 2.1

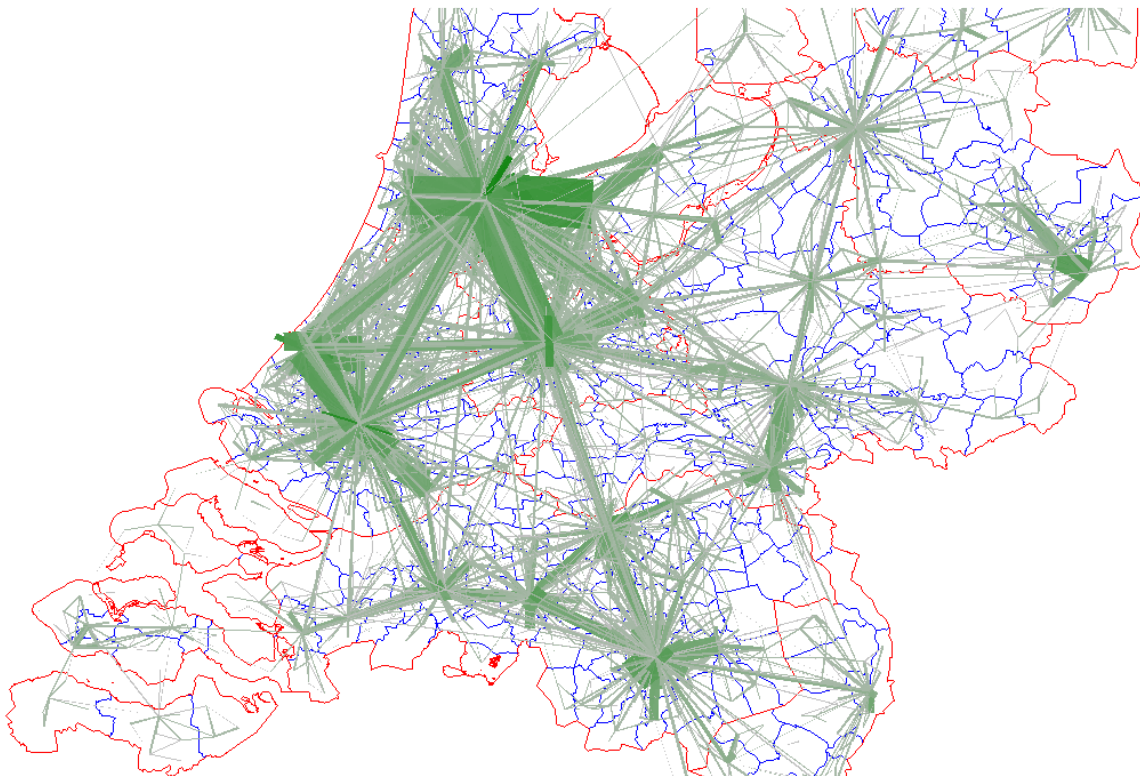
Vervoersstromen in verkeersmodel regio Rotterdam

Het model behoort tot de state-of-the-art van Nederland en is geijkt op een groot aantal gegevens. In een aantal projecten is desondanks gebleken dat de wiskundige distributie van verkeer niet altijd juist leek te zijn (dus niet de hoeveelheid verplaatsingen maar de richting waarin deze plaatsvinden). Een interessante vraag is dus of nieuwe data, zoals de GSM-data, dat onderbuikgevoel kon onderbouwen en wat er te doen is om de distributie te verbeteren?

2.2 De databron

De GSM-dataset is voor dit onderzoek beschikbaar gesteld door Mezero en betreft data van een van de drie grote netwerkproviders in Nederland in november 2014. De verplaatsingen zijn beschikbaar op HB-niveau in ongeveer 1.250 gebieden en bevatten ruim 100 miljoen reizen.

Figuur 2.2 is een weergave van de landelijke HB-matrix voor één dag. De verplaatsingen zijn afgeleid uit de operationele netwerkdata. Mobiele devices melden zich –ook als ze niet actief gebruikt worden– 20 tot 300 keer per dag aan bij zendmasten en laten zo weten in welke ‘zone’ ze dan zijn. De privacy is hierbij gegarandeerd via technische procedures waarbij de data wordt geanonimiseerd, gefilterd (< 16 verplaatsingen verwijderd) en maximaal een maand bewaard. Op geen enkele wijze mag het namelijk mogelijk zijn data te herleiden tot individuele verplaatsingen. Een plaats wordt als ‘bestemming’ aangemerkt als een GSM op die locatie meer dan 30 minuten is gesignaleerd. Op deze manier wordt voorkomen dat ‘onderweg’ (bij een verkeersopstopping bijvoorbeeld) als ‘bestemming’ wordt aangemerkt.



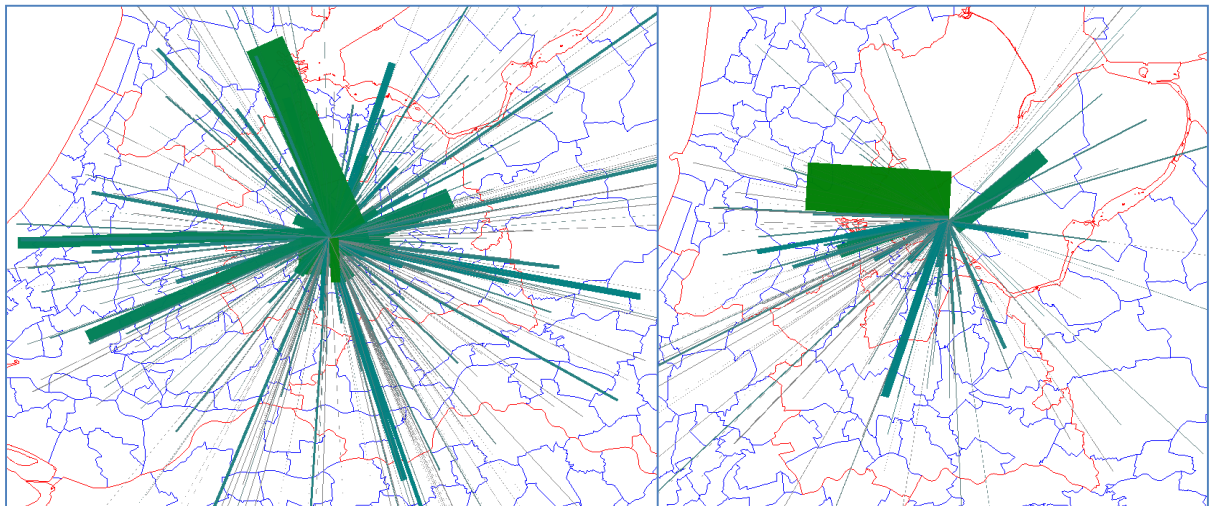
Figuur 2.2 Lineair geplote verplaatsingspatronen uit de GSM-data

De gebieden in de GSM-dataset zijn op enkele uitzonderingen na voor het grootste gedeelte woonplaats-georiënteerd. De grotere steden zijn opgesplitst in enkele stadsdelen en enkele kleinere plaatsen zijn juist samengenomen. De stad Rotterdam bestaat in deze dataset uit acht districten (of zones in verkeersmodeltermen).

3. Kwaliteit van de GSM-database

3.1 Herkenbare beelden?

Figuur 2.2 laat duidelijk zien hoe via GSM-waargenomen verplaatsingen zich concentreren tussen en rondom stedelijke gebieden. Naast deze eerste visuele plausibiliteitstoets zijn een aantal andere logische checks uitgevoerd om de kwaliteit van de GSM-data te controleren. Figuur 3.1 toont bijvoorbeeld de verplaatsingen vanuit Utrecht en vanuit Almere. Als centrale stad in Nederland is duidelijk zichtbaar dat hoewel de zwaarste relatie Utrecht-Amsterdam is, naar alle windrichtingen ritten plaatsvinden. Bij Almere is het patroon –conform onze verwachting– veel eenzijdiger. De sterke relatie met Amsterdam springt duidelijk in het oog. Door op deze wijze door de dataset heen te lopen werd een goed gevoel gekregen voor de kwaliteit van de gemeten verplaatsingen. De data laat heel herkenbare beelden zien. Er zijn ook opvallende zaken, zoals onder andere dat de Maasvlakte bij Rotterdam weinig lange-afstandsverplaatsingen kent volgens deze dataset. Dat lijkt erop te duiden dat het vrachtverkeer beperkt in de data aanwezig is. Dit kan mogelijk verklaard worden door het gebruik van het 30-minuten criterium en de aangebrachte privacy filtering. Overigens is vrachtverkeer nu niet iets waar we in deze pilot op focussen, aangezien het havengebonden vrachtverkeer reeds op basis van bluetooth informatie wordt ingevuld in verkeersmodel.



Figuur 3.1 Utrecht (links) en Almere (rechts) uitgelicht

3.2 *Vergelijking met OViN*

Een tweede controle was een vergelijking tussen de GSM-data en het Onderzoek Verplaatsingsgedrag In Nederland (OViN). Dit onderzoek vond plaats op het niveau van de voormalige stadsregio Rotterdam (voor de periode 2010-2013). Beide bronnen beschrijven immers de Nederlandse mobiliteit waardoor er een verband tussen de twee databronnen zichtbaar zou moeten worden. Op basis van diverse vergelijkingen tussen de GSM-data en het OViN kunnen de volgende conclusies worden getrokken:

- De GSM-data beschrijft op het niveau van het absoluut aantal ritten niet de Nederlandse Mobiliteit. Voor de data in deze vorm moet nog een accuraat ophoogmechanisme worden ontwikkeld, die rekening houdt met onder andere landelijke dekking en andere specifieke factoren in de GSM-data.
- Tussen de GSM-steekproef en OViN is sprake van enkele definitieverschillen. OViN betreft persoonsverplaatsingen, waarbij in de GSM-data verwacht mag worden dat ook vracht- en bestelverkeer deels vertegenwoordigd is. Daarnaast bevat de OViN-analyse voor het verkeersmodel niet de hoofdmodaliteit 'lopen', waarbij dat voor de GSM-data waarschijnlijk wel zo is.
- De ritlengteverdeling in de GSM-data is langer, wat duidt op een mogelijke ondervertegenwoordiging van korte verplaatsingen.
- Op geaggregeerd niveau naar enkele gemeenten vertonen de distributiepatronen uit OViN en GSM-data goede overeenkomsten.

Na de uitgevoerde kwaliteitscontroles is ondanks enkele optredende verschillen voldoende vertrouwen gekregen om de dataset te confronteren met het verkeersmodel. Daarbij is gekeken naar de mogelijkheden om een verkeersmodel te kunnen verrijken, gegeven de nieuw verkregen inzichten over de kwaliteit en aandachtspunten van de GSM-data.

4. **Verrijking van het verkeersmodel**

4.1 *Werkwijze*

Specifiek richtte de casestudy zich op het verbeteren van de distributie van het verkeersmodel. De reden hiervoor is dat over dit aspect de grootste twijfel was bij de huidige beheerders van het verkeersmodel en dat dit aspect het sterkst gemeten wordt door GSM-data. Dit in tegenstelling tot de totale hoeveelheid verplaatsingen (ritproductie) waar meer algemene kennis van is en dat juist nog niet goed uit de GSM-data is af te leiden.

We kozen voor een werkwijze die alle overige parameters constant hield om het effect op de distributie te kunnen isoleren. De bestaande ritproductie uit een bepaalde zone gebruikten we om de relatieve verdeling over de bestemmingen volgens de GSM-data te schalen. Ook de verdeling over de dagdelen en vervoerswijzen pasten we niet aan. De werkwijze was als volgt:

- We telden alle deelmatrices in het verkeersmodel op tot een etmaaltotaalmatrix persoonsverplaatsingen. Het betrof de dagdeelmatrices (ochtendspits, avondspits en restdag) en de vervoerswijzen auto, fiets en OV. We namen de matrices zoals

afgeleid uit de reguliere data vóór kalibratie op tellingen (de zogenaamde a-priori-matrices).

- De 6.000 zones van deze verkeersmodelmatrix aggregaerden we vervolgens naar de 1.200 zones van de HB-matrix van de GSM-data.
- De distributie van verplaatsingen tussen de zones in deze aggregaerde etmaaltotaalmatrix is relatief gelijk getrokken met die van de GSM-data. De intrazonale verplaatsingen binnen een GSM-gebied bleven ongecorrigeerd aangezien daar vanuit de dataset geen informatie van is.
- De op deze manier verrijkte aggregaerde etmaaltotaalmatrix is vervolgens weer gedesaggregeerd naar de 6.000 zone-indeling van het verkeersmodel en daarna weer terugvertaald naar dagdeelmatrices per vervoerswijze.

Het resultaat is dat elke dagdeelmatrix voor elke vervoerswijze nog steeds per zone dezelfde hoeveelheid verkeer bevat conform de schatting van het verkeersmodel, maar dat de verdeling over de bestemmingen is aangepast op basis van de GSM-waarnemingen. Voor deze analyse is alleen gekeken naar de autoverplaatsingen. De matrices zijn toegedeeld aan de infrastructuur (over welke routes worden deze verplaatsingen gemaakt?) waarna verdere analyses en vergelijkingen zijn gemaakt.

4.2 In eerste instantie: geen plausibel resultaat

De eerste poging leidde niet tot een plausibel resultaat. Analyse van de data toonde aan dat er grote aanpassingen in de matrix optraden. Dit werd veroorzaakt door een verschil in ritlengte. Eerder hadden we al gezien dat de ritlengte in de GSM-data substantieel langer was dan op basis van OViN verwacht werd. Dit werd nu zichtbaar in het verkeersmodel: grote overschattingen van de verkeersintensiteiten op vooral de hogere orde wegen. Dit wordt vooral veroorzaakt doordat de GSM-matrix een ondervertegenwoordiging van korte verplaatsingen bevat. De oorzaak hiervan is onderwerp van nader onderzoek. De veronderstelling is dat de GSM-activiteit gedurende korte verplaatsingen minder is en dat het 30-minuten criterium hier mogelijk beperkend werkt.

Duidelijk is dat dit beeld niet aan de verwachtingen voldeed. De hypothese vooraf was dat de intensiteitsveranderingen op wegvakniveau beperkt zouden zijn, en dat vooral de interne distributie zou worden aangepast. Bij het maken van een zogenaamde 'selected link' zou de intensiteit op de gekozen link dan ongeveer gelijk zijn, maar ziet het HB-patroon er anders uit.

4.3 Ritlengten corrigeren

Idealiter wordt de ritlengteverdeling in de GSM-data verbeterd door middel van datafusie met andere bronnen. Of beter gezegd: de lacune van ontbrekende korte verplaatsingen in de GSM-data dient te worden opgevangen. De wijze waarop dit het beste kan gebeuren vraagt echter nader onderzoek. Voor ons onderzoek is besloten tot pragmatiek. De verplaatsingen tot een afstand van 8 kilometer en vanaf 60 kilometer lieten we in het verkeersmodel ongewijzigd, omdat vergelijkingen aantoonde dat de GSM-data hier onvoldoende waarnemingen c.q. kwaliteit bevat. De ritten tussen 8 en 60 km volgens de GSM-data zijn gecorrigeerd naar de gemiddelde ritlengte van het OViN. Op de

afstandsklasse 8-13 kilometer is de GSM-data opgehoogd, tussen 14-40 kilometer is de GSM-data verlaagd en vanaf 40 kilometer ongewijzigd.

Op deze manier ontstond een ritlengteverdeling die past bij de kracht van beide databronnen. OViN-data is prima geschikt om te toetsen op ritlengteverdelingen en modal split in een groter gebied, maar bevat te weinig steekproefwaarnemingen om expliciet onderscheid te maken tussen individuele relaties tussen gebieden. GSM-data heeft die grote hoeveelheid waarnemingen wel, althans voor bepaalde afstandsklassen. Ter vergelijking: jaarlijks worden er ruim 40.000 personen geënquêteerd in het OViN, terwijl in de maand november alleen al ruim 1 miljoen verplaatsingen in de (gefilterde) GSM-data aanwezig zijn.

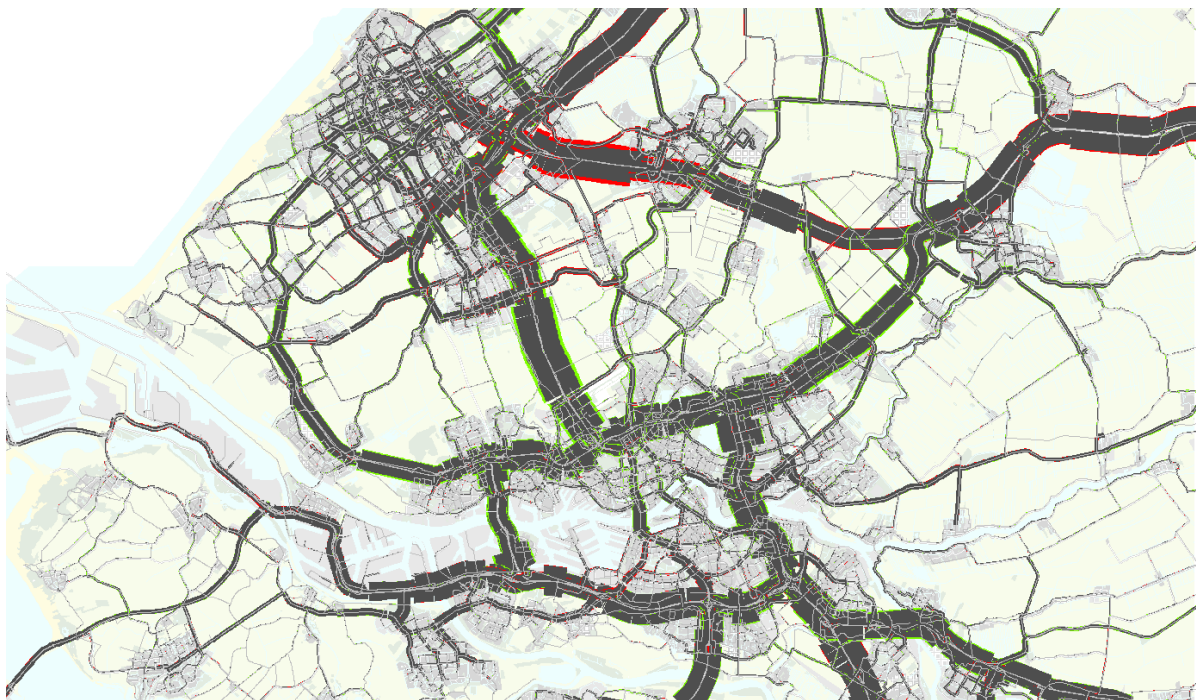
5. Resultaten

5.1 Op hoofdlijnen

De resultaten na de correctie op ritlengte zijn aanmerkelijk beter. De vooraf gestelde hypothese blijkt nu te kloppen: op verreweg de meeste wegvakken komen de verkeersintensiteiten voor en na de verrijkingsslag goed overeen. De gemiddelde ritlengte van het verkeersmodel bedraagt nu na correctie 16.8 km en ligt daarmee vrijwel op het niveau van het a-priori-model (16.4 km). Figuur 5.1 toont een verschilanalyse (grijs = gelijk, rood = toename na GSM-correctie, groen = afname na GSM-correctie) tussen de verrijkte en basissituatie.

Er zijn ook verschillen: op de relatie Zoetermeer-Den Haag laat bijvoorbeeld de verrijking meer verkeer zien dan het a-priori-model. We hoopten dat effect ook te zien. Net als Almere-Amsterdam is Zoetermeer-Den Haag een historisch gegroeide relatie die moeilijk in traditionele modellen is te modelleren. Verrijking met GSM-data zorgt dat deze relatie nu meer in lijn met het werkelijke verplaatsingspatroon in het model is terechtgekomen.

Verder zien we op de A13 –de zwaar overbelaste snelweg tussen Rotterdam en Den Haag- minder verkeer na verrijking en juist op de A16 richting Dordrecht een lichte toename. Ook op de relatie Rotterdam-Maasvlakte is meer verkeer zichtbaar na verrijking. Deze effecten zijn in lijn met observaties door specialisten van de Metropoolregio Rotterdam-Den Haag, die aangeven dat op de genoemde relaties het ongecorrigeerde verkeersmodel inderdaad moeite heeft om de daadwerkelijke patronen te modelleren.



Figuur 5.1 Vergelijking toedeling auto etmaal (matrix verrijkt met GSM-data versus a-priori-verkeersmodel)

5.2 Naar input voor een netwerkkalibratie

Naast de visuele check aan de toedeling is een tweede toets uitgevoerd om te controleren of de verrijking leidt tot een betere benadering van de werkelijkheid, een zogenoemde T-waarde. Deze toets vergelijkt modelwaarden met beschikbare tellingen en wordt gebruikt de toegedeelde matrix te valideren aan beschikbare verkeerstellingen. De maat kijkt naar de afwijking van modelwaarde en telwaarde waarbij rekening wordt gehouden met zowel de omvang van de relatieve als de absolute afwijking. Uit de resultaten volgt dat de matrices na verrijking ongeveer 5% beter scoren aan de T-waarde. De matrices zijn hierdoor gemakkelijker te kalibreren aan de verkeerstellingen (er zijn minder kalibratiecorrecties nodig), wat aangeeft dat de kwaliteit van het verkeersmodel is verbeterd. In deze casestudy heeft overigens nog geen kalibratie van de verrijkte matrices plaatsgevonden.

We hebben op dit punt dus een verrijkt model dat plausibele resultaten oplevert die op een aantal punten –zoals we beoogden– dichter bij de werkelijkheid liggen dan het huidige model.

5.3 Een stap dieper: op linkniveau

Als volgende toets zoomden we in op een tweetal wegen (A13, A20) en een specifiek gebied (Hellevoetsluis). We vergeleken de toedeling op deze selected links respectievelijk selected zone in het verrijkte model met die in het oorspronkelijke a-priori-model.

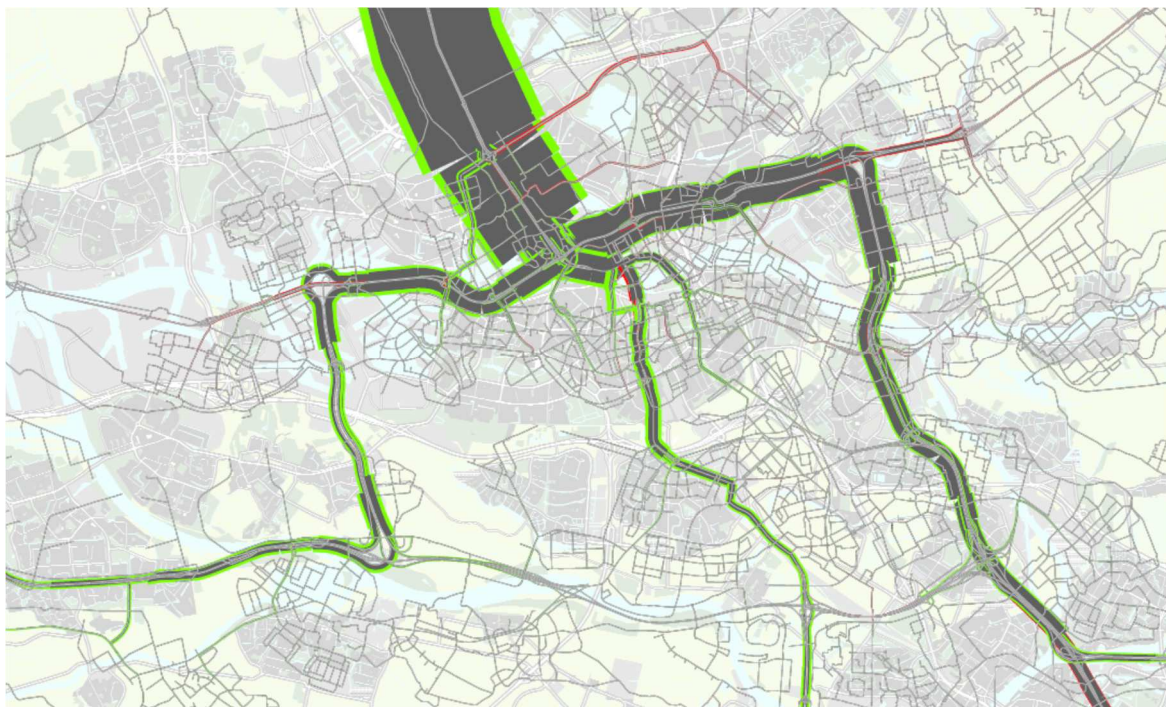
A13

De analyse op de A13 (figuur 5.2) toont in het verrijkte model veel minder verkeer dan in het a-priori-model. De door het verrijkte model geschatte intensiteit ligt dichterbij de getelde intensiteit. Dat is dus een verbetering ten opzichte van het a-priori-model.

Dat op de A13 in het huidige verkeersmodel sprake is van een overschatting was overigens bekend. De A13 kent namelijk een capaciteitsbeperkend effect en dat is moeilijk te modelleren. Met een iteratieve matrixschattingsprocedure waarbij congestie wordt meegenomen in de verkeersvraagberekening wordt dat in het huidige model gedeeltelijk ondervangen, maar dit is kennelijk niet accuraat genoeg. De berekende weerstanden zijn vermoedelijk nog te laag.

We zijn overigens benieuwd wat er op dit deel van het wegennetwerk gaat gebeuren als de A4 Delft-Schiedam wordt opengesteld. Treedt de nu met GSM-gecorrigeerde vervorming ook dan op? We doen dan graag een vervolganalyse met dan geldende GSM-gegevens.

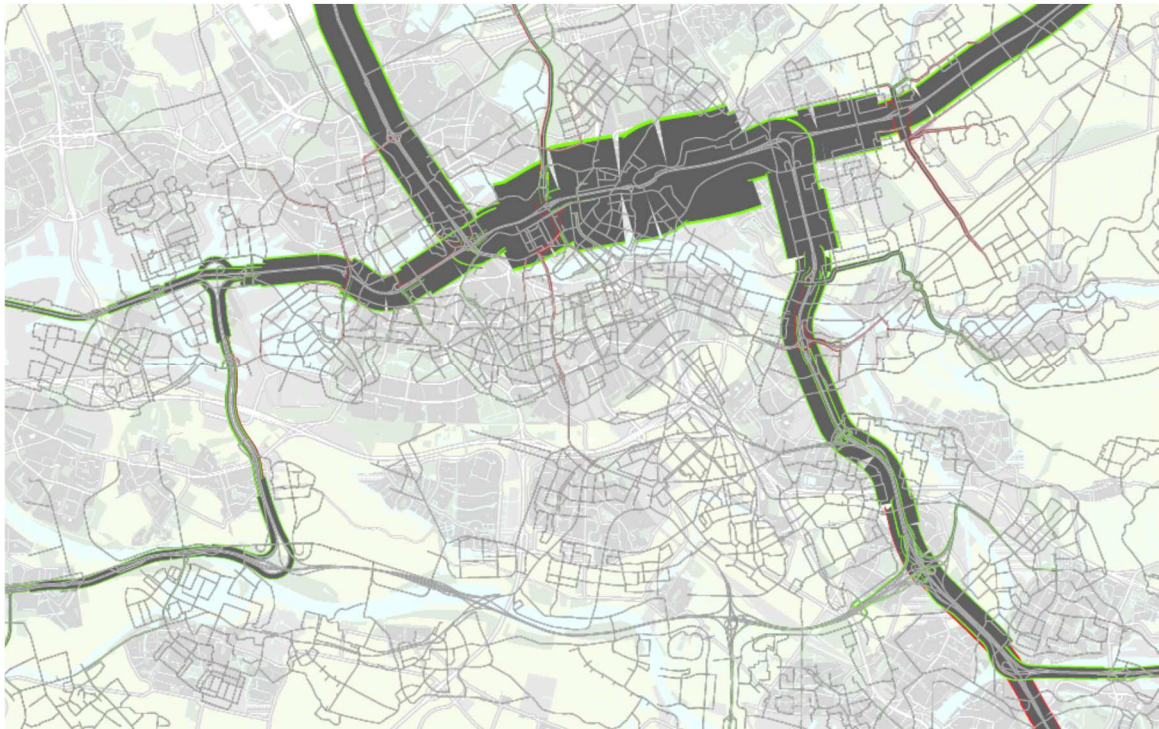
Op detailniveau is verder zichtbaar dat de A13 via de N209 in het verrijkte model meer verkeer aantrekt vanuit Rotterdam Noord dan de zuidelijker gelegen wijken.



Figuur 5.2 Verschilanalyse distributie A13

A20

Op de A20 (figuur 5.3) zagen we relatief weinig verschillen tussen de intensiteiten voor en na de verrijking. Dit is een gewenste constatering. Eerdere analyse van kentekenkamera's van De Verkeersonderneming toonde namelijk al aan dat de A20-A16 relatie behoorlijk goed in het model zit. De verrijking verstoort dat beeld dus niet.

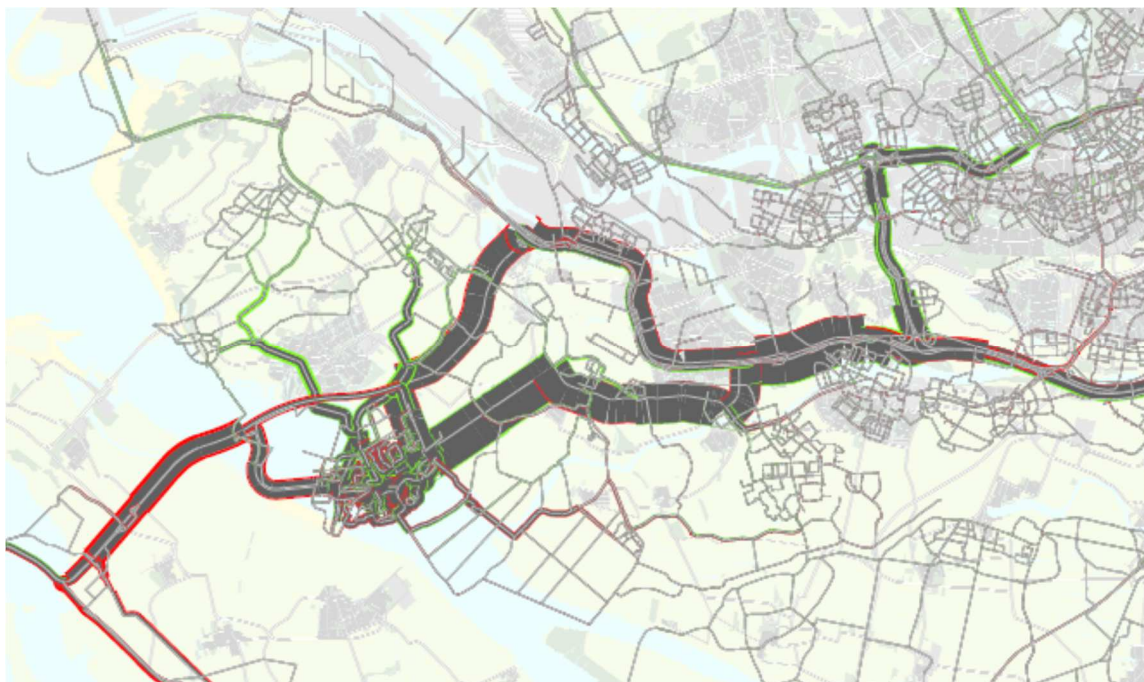


Figuur 5.3 Verschilanalyse distributie A20

Hellevoetsluis

De grootste effecten in de toedeling zijn zichtbaar wanneer een analyse van een heel gebied wordt gemaakt. In figuur 5.4 is het verrijkte distributiepatroon van Hellevoetsluis vergeleken met het a-priori-model. Zichtbaar is dat de verrijking meer verkeer laat zien op de relatie met Goeree Overflakkee en de relatie met Rotterdam-zuid. Afnames werden zichtbaar in de directe omgeving op Voorne Putten & de Maasvlakte en de langere afstand van/naar Den Haag en verder.

Vooraf was inderdaad onze hypothese dat met name voor Hellevoetsluis een verschuiving over de verschillende afstandsklassen zou optreden. Het a-priori-model berekent namelijk veel verplaatsingen naar de Maasvlakte vanwege de hoeveelheid nabij gelegen arbeidsplaatsen. De GSM-data laat zien dat de werkelijkheid anders is en we zien dat terug in de toedeling. Vanwege de ligging in het stedelijk gebied zijn er vooral veel ritten met een bovengemiddelde afstandsverdeling naar specifieke delen van het netwerk (Rotterdam, Zeeland).



Figuur 5.4 Verschilanalyse distributie Hellevoetsluis

6. Conclusies

De analyses hebben aangetoond dat het zeer goed mogelijk is met de geformuleerde werkwijze de distributie in de HB-matrices van een operationeel verkeersmodel te corrigeren aan de hand van GSM-data. Gebruik van de ruwe GSM-data in de huidige vorm leidt echter nog niet tot verbeteringen. Met enige vorm van bijstelling/ophoging van de GSM-data vertoont de resulterende toedeling wel een goede gelijkens met het oorspronkelijke a-priori-model. Er komt geen volledig ander mobiliteitspatroon uit. Dat was gezien de kwaliteit van het operationele model ook ongeloofwaardig geweest. Wel komt de toedeling op de onderzochte punten dicht bij de werkelijkheid. Dat is precies wat we wilden.

Door voor het eerst te werken met een full-scale operationeel verkeersmodel haalden we dit onderwerp uit een laboratorium-setting naar de operationele praktijk. Dit heeft ons veel kennis gegeven ten aanzien van de weg die nog moet worden bewandeld naar data-driven modellering en daarnaast zeer bruikbare informatie opgeleverd over de huidige verplaatsingspatronen in de regio Rotterdam.

Kanttekening is dat de kwaliteit van de GSM-data op dit moment nog niet op het niveau is dat deze zonder meer direct en automatisch tot plausibele resultaten leidt. Vooral de ophoging van de GSM-data naar een absoluut mobiliteitsniveau die een logische ritlengteverdeling bevat is de eerste stap die daarin nog gezet moet worden. Voor nu is dit opgelost door op relatief pragmatische wijze een filtering op HB-paren toe te passen en te corrigeren op basis van afstandsverdelingen. Verbetering van de GSM-data blijft dus nodig.

Een tweede kanttekening is het moment van correctie in het modelproces. We doen dat nu pragmatisch en vooral *na* de matrixschatting. Idealiter hebben we een methode waar de GSM-matrix ons helpt bij de correctie van de weerstandsmatrices die gebruikt worden *in* de matrixschatting. Een dergelijke werkwijze verbetert namelijk ook de verklarende waarde van het verkeersmodel.

Naar aanleiding van dit onderzoek zijn de volgende onderdelen benoemd waarop vervolg onderzoek zich gaat richten:

1. Verbetering begrip en kennis van de GSM-data, alsmede de data-inwinning ervan. Vooral op het gebied van de korte verplaatsingen ligt een uitdaging, waar datafusie met andere bronnen, zoals WiFi-data en bluetooth-metingen een mogelijkheid is.
2. Onderzoek naar vervolgstappen zoals de kalibratie en verwerking in prognosematrices om zodoende de beschrijvende werking van verkeersmodellen te verbeteren.
3. Onderzoek naar de verwerking van de GSM-data in de matrixschattingsmodule, bijvoorbeeld in de weerstandsmatrices, om de verklarende werking van verkeersmodellen te verbeteren.

De casestudy op het verkeersmodel van Rotterdam heeft veel bruikbare resultaten opgeleverd, zowel op het gebied van modeltechniek als inzicht in verplaatsingspatronen in de regio Rotterdam. Het grootste leerpunt is wellicht dat bij dit soort big-data toepassingen vaak wordt gefocust op imperfectie van de data en daardoor ontwikkeling wordt afgeremd. In deze pilot is juist sterk gefocust op wat wel bruikbaar is. Uiteindelijk heeft dit de eerste grootschalige praktijktoepassing van GSM-data in verkeersmodellering opgeleverd die zeker op regionaal niveau aantoonbare meerwaarde heeft. Indien deze ontwikkeling de komende jaren voortzet is de weg naar verbeterde en meer datagedreven verkeersmodellering definitief ingezet.

Referenties

- [1] P. van der Mede; 'GSM-data niet meer weg te denken', GIS Magazine, p 26-27, Maart 2013.
- [2] Nanni , M. et al; 'MP4-A Project: Mobility Planning For Africa' D4D 2013 – Data for Development – Special session of the Third International Conference on the Analysis of Mobile Phone Datasets (Cambridge, USA, 2-3 May 2013) Proceedings article n. 40. www.netmob.org 2013.
- [3] P. v.d. Mede, 'Over het meten van mobiliteit met gsm-data: mogelijkheden en onmogelijkheden', Colloquium Vervoersplanologisch Speurwerk, 20-21 November 2014.
- [4] E. de Romph, Gonçalo H.A. Correia, Yihong Wang; 'National and Regional Road Network Optimization for Senegal Using Mobile Phone Data'. Data for Development Challenge Senegal, 2015.