



# GEBRUIK VAN SENSOREN OM MEETFOUTEN IN ENQUÊTES TE CORRIGEREN

Het CBS publiceert cijfers over het goederenvervoer over de weg. Deze cijfers geven inzicht in het gebruik van het Nederlandse wegennet en de doorvoer- en distributiefunctie van het Nederlandse wegtransport. Ook worden de resultaten gebruikt voor het bepalen van beleid op het gebied van milieu en verkeersveiligheid. Om het goederenvervoer over de weg in kaart te brengen voert het CBS een enquête uit onder de eigenaren van de Nederlandse bedrijfsvoertuigen. Door de uitkomsten van de enquêtes en administratieve gegevens te koppelen aan de data die een sensorsysteem van Rijkswaterstaat verzamelt, kan met vangst-hervangst-technieken in potentie onderrapportage worden gemeten.

Vrachtwagens op de A28, bij knooppunt Hattermerbroek. Foto: BakkeBaarend CC

JONAS KLINGWORT, BART BUELENS, JOEP BURGER, RAINER SCHNELL

Er is steeds meer belangstelling voor het gebruik van sensordata in het maken van officiële statistieken. Bij officiële statistieken wordt veel gebruik gemaakt van kanssteekproeven omdat daarmee efficiënt zuivere schattingen over de hele populatie gemaakt kunnen worden. Via een a-selecte steekproef worden mensen verzocht een vragenlijst in te vullen. Deze methode heeft wat nadelen: is potentieel belastend voor de respondent, niet iedereen doet mee en antwoorden zijn gevoelig voor fouten – al dan niet bewust gemaakt.

Sensoren kunnen de antwoorden op sommige vragen direct meten zonder tussenkomst van een respondent. Vaak hebben sensoren dan weer als nadeel dat niet elk element in de populatie een bekende en positieve kans heeft om gemeten te worden. In dit onderzoek zijn sen-

sordata geïntegreerd met enquête- en registerdata om met vangst-hervangst (VH) technieken de kwaliteit van de populatieschattingen te verbeteren. In de hier beschreven toepassing wordt gebruik gemaakt van de Enquête Wegvervoer, de bedrijfs- en kentekenregisters en Weigh-in-Motionwegin-sensoren. Dit onderzoek is de eerste keer dat het CBS sensorgegevens op microniveau koppelt aan enquêtes en administratieve gegevens om mogelijke vertekening door onderrapportage in dagboekenquêtes te beoordelen.

## Data

De Enquête Wegvervoer is een dagboekonderzoek naar het goederenvervoer over de weg. De gepubliceerde cij-

fers geven inzicht in het gebruik van het wegennet en de functie van het wegtransport. Verder worden ze gebruikt voor het bepalen van beleid op het gebied van milieu en verkeersveiligheid. Het is een verplichte enquête met een hoog responspercentage van ongeveer 90% op een steekproef van ongeveer 34 duizend kentekens. De bezitter van de vrachtwagen wordt verzocht alle ritten en zendingen per dag gedurende een toegewezen week op te geven. Onderrapportage door onterechte opgave dat de vrachtwagen niet is gebruikt zal leiden tot onderschatting van het goederenvervoer over de weg.

De sensordata worden verzameld door het Weigh-in-Motionsensorsysteem van Rijkswaterstaat om overbelading van vrachtwagens op te sporen. In 2015 werd op 18 locaties (figuur 1) het gewicht van elke pas-

serende vrachtwagen gemeten en werd het type van de vrachtwagen geïdentificeerd, resulterend in ongeveer 36 miljoen waarnemingen. Daarnaast is een foto van de kentekenplaat van de voor- en achterkant gemaakt, die een directe koppeling met de registers mogelijk maakt (figuur 2). Hierdoor kan het lege vrachtwagengewicht van de sensormeting worden afgetrokken. Dit verschil resulteert in het vervoersgewicht, dat gelijk is aan de definitie van het gevraagde gewicht in de enquête. De mogelijkheid om de sensorwaarnemingen op microniveau te koppelen aan de enquête maakt toepassing van VH-methoden mogelijk.

De bedrijfs- en kentekenregisters leveren aanvullende gegevens op met informatie over de technische kenmerken van de vrachtwagens en de specifieke ken-





Figuur 1. Locatie van de wegsensoren van het Weigh-in-Motionsensorsysteem; elk van de negen systemen heeft twee stations, één in iedere richting, aangeduid met een plus en een cirkel



Figuur 2. Weigh-in-Motionsoftware met onder andere het gewicht per as (het kenteken is om privacyredenen vervaagd)

merken van de vrachtwagenbezitters. Deze achtergrondkenmerken worden gebruikt in de VH-modellen. Fouten in de sensormetingen van de asgewichten, bijvoorbeeld door oneffenheden in de weg, kunnen leiden tot vertekening van het geschatte vervoersgewicht. Om dit risico te beperken zijn extreme waarden geïmputeerd met een conditioneel gemiddelde gegeven enkele achtergrondkenmerken.

### Methode

Vangst-hervangst-technieken zijn ontwikkeld in de ecologie om de (onbekende) omvang van dierenpopulaties te schatten. In de hier beschreven toepassing is de steekproefenquête de eerste vangst en de sensorwaarneming de tweede vangst. VH-technieken corrigeren voor zowel nonrespons als onderrapportage in de enquête. Indien de selectieve nonrespons in de enquête volledig wordt verklaard door het weegmodel, kan het relatieve verschil tussen de gewogen enquêteschatting en een VH-schatting duiden op onderrapportage in de enquête.

Voor de toepassing van dergelijke technieken zijn de aantallen in tabel 1 nodig, die zijn af te leiden door

de koppeling van beide bronnen op kenteken en dag. Er zijn elementen (kenteken-dagcombinaties) die alleen in de enquête gerapporteerd worden, elementen die alleen door de sensoren gedetecteerd worden, en elementen die in beide bronnen: de hervangst. Elementen die noch in de enquête worden gerapporteerd, noch door de sensoren worden gedetecteerd, worden geschat met behulp van VH-technieken. Op basis van de combinatie van beide bronnen kan dan een betere populatieschatting gemaakt worden dan op basis van een enkele bron, mits aan een aantal voorwaarden wordt voldaan (zie volgende paragraaf).

De vangkans in de enquête en de sensoren worden gemodelleerd met logistische regressie en loglineaire modellen met behulp van covariaten uit de twee registers. Als de covariaten de heterogeniteit in vangkans verklaren, maakt het niet uit hoeveel en waar de sensoren geïnstalleerd zijn. Er worden vijf schatters toegepast voor twee doelvariabelen: vrachtwagendagen (D) en vervoerd gewicht (W). Een vrachtwagendag (D) wordt gedefinieerd als een dag dat een vrachtwagen in Nederland op de weg is geweest.

De eerste schatter (ENQ) is gebaseerd op de antwoorden in de enquête gecorrigeerd voor selectieve

		SENSOREN		TOTAAL
		gedetecteerd	niet gedetecteerd	
ENQUÊTE	gerapporteerd	Hervangst	Alleen gerapporteerd	Eerste vangst
	niet gerapporteerd	Alleen gedetecteerd	?	?
TOTAAL		Tweede vangst	?	?

Tabel 1. Combinatie van de rapportages van enquêtes en van sensoren

nonrespons met behulp van poststratificatie-weging. Deze schatter wordt vergeleken met de VH-schatters. Aangezien VH-schattingen voor zowel de selectieve nonrespons als meetfouten corrigeren, kan het verschil tussen VH-schattingen en ENQ worden toegeschreven aan meetfouten, dat wil zeggen onderrapportage in de enquête. Voor een naïeve uitbreiding van de huidige schatter worden de waarnemingen van de enquête aangevuld met de sensorwaarnemingen (ENQX). De Huggins-schatter (HUG) modelleert de vangkans met een logistisch regressiemodel als functie van achtergrondkenmerken, gegeven dat een vrachtwagen door ten minste één bron is waargenomen. De Lincoln-Petersen-schatter (LP) gebruikt alleen de hoeveelheden uit bovenstaande tabel. De eerste vangst wordt opgehoogd met de verhouding tussen de tweede vangst en de hervangst. De loglineaire schatter (LL) modelleert de hoeveelheden in bovenstaande tabel, uitgesplitst naar achtergrondkenmerken, met een Poisson-regressiemodel. Een stapsgewijze selectieprocedure (gebaseerd op BIC) wordt toegepast om de balans te vinden tussen modelfit en modeleenvoud (parsimonie). De steekproefvariantie wordt voor alle schatters geschat door middel van bootstrapping.

### Aannames

De VH-schatting is gebaseerd op de volgende vijf aannames: onafhankelijke datasets, gesloten populatie, homogene vangkans, alle elementen behoren tot de doelpopulatie en perfecte koppeling van datasets. Het voldoen aan de vijf VH-aannames is een vereiste om onvertekende VH-schattingen te garanderen. Aan de eerste aanname (onafhankelijkheid) wordt voldaan. De tweede aanname (gesloten populatie) wordt niet ernstig geschonden, aangezien het eerste en laatste kwartaal van het voertuigregister meer dan 99% overlappen. Aan de derde aanname (homogene vangkans) wordt voldaan zolang de heterogeniteit in vangkans verklaard wordt door de achtergrondkenmerken in de modellen. Dezelfde aanname wordt gemaakt bij het corrigeren van de enquêterespons voor selectieve nonrespons. Aan de vierde aanname (elementen behoren tot de doelpopulatie) wordt voldaan omdat de enquêtesteekproef uit het populatieregister wordt getrokken en in de sensordata alleen wordt gezocht naar voertuigen in de steekproef.

De vijfde aanname (perfecte koppeling) kan soms worden geschonden in beide gegevensbronnen om verschillende redenen. Ten eerste kan het zijn dat voer-

tuigeigenaren te weinig of te veel dagen van de week rapporteren, of dat zij de verkeerde dagen van de week rapporteren. Ten tweede moeten de voertuigeigenaren de dag van laden melden, welke kan verschillen van de dag van rijden zoals gedetecteerd door de sensoren. Ten derde registreren de sensoren ritten waar niet over gerapporteerd hoeft te worden, zoals lege ritten en ritten voor onderhoud of tanken. Ten vierde kan de sensor een kentekenplaat soms niet herkennen. Om de gevoeligheid van de resultaten voor deze respons- en sensorfouten te beoordelen zijn simulatiestudies uitgevoerd.

### Resultaten en discussie

Alle VH-schattingen leveren grotere schattingen op voor het aantal vrachtwagendagen en het vervoersgewicht dan de schattingen op basis van de enquête. We raden het gebruik van de loglineaire schatter (LL) aan, omdat deze rekening houdt met de heterogeniteit in vangkansen (in tegenstelling tot LP) en is gebaseerd op het gekoppelde bestand, onvoorwaardelijk op waarneming in ten minste één bron (in tegenstelling HUG).

De simulatiestudies met betrekking tot de vijfde aanname (perfecte koppeling) toonden het volgende aan. In de eerste plaats is LL robuust voor overrapportage in de enquête, terwijl ENQ zou afnemen, waardoor de geschatte onderrapportage zou toenemen. Ten tweede zou LL toenemen met onderrapportage in de enquête, maar langzamer dan ENQ, waardoor de geschatte onderrapportage zou afnemen. Maar zelfs als elke gerapporteerde lading over de volgende dagen zou worden verdeeld, blijft ENQ lager dan LL. De geschatte onderrapportage kan dus niet volledig worden verklaard door een verschil tussen de dag van laden zoals gerapporteerd in de enquête en de dag van rijden zoals gedetecteerd door de sensor. Ten derde verdwijnt de geschatte onderschatting geleidelijk naarmate de fractie foutpositieven in de sensor toeneemt. Het is een open vraag hoe waarschijnlijk het is dat een vrachtwagen in de steekproef die door een sensor wordt gedetecteerd, niet gerapporteerd hoeft te worden. Ten vierde is de zuiverheid van de LL-schattingen robuust tegen het niet herkennen van het kenteken door de sensorsoftware. Alleen de precisie zal afnemen naarmate minder kentekens gekoppeld kunnen worden.

Toekomstig onderzoek zou de hypothese kunnen testen dat het relatieve verschil tussen de enquêteschatting en de VH-schatting prominenter is in handmatige

responsmodi (internet- of papieren vragenlijsten) dan in geautomatiseerde responsmodi (XML). Het is ook aan te raden om te onderzoeken of een deel van het verschil alsnog door nonrespons verklaard kan worden met een beter weegmodel. Verder is het van belang na te denken over het behandelen van de antwoordcategorie 'niet in bezit' als kaderfout of (technische) nonrespons. Tot slot is er behoefte aan een methode om de kans op foutpositieven te kunnen schatten.

De hier gepresenteerde methode is van toepassing op elke validatiestudie waarbij enquête-, register en sensor-data (of een andere externe bigdatabron) op microniveau kunnen worden gekoppeld. Het voldoen aan de vijf VH-aannames garandeert onvertekende VH-schattingen. Simulatiestudies kunnen inzicht geven in de gevoeligheid van de resultaten voor het schenden van een of meerdere aannames.

#### LITERATUUR

Klingwort, J., Buelens, B., & Schnell, R. (2019). Capture-recapture techniques for Transport Survey estimate adjustment using permanently installed highway sensors. *Social Science Computer Review*, OnlineFirst, doi: 10.1177/0894439319874684.

JONAS KLINGWORT is wetenschappelijk medewerker en PhD-kandidaat in de Onderzoeksgroep Methodologie van prof. dr. Rainer Schnell aan de Universiteit van Duisburg-Essen. Daarnaast is hij statistisch onderzoeker bij het *Center for Big Data Statistics* van het Centraal Bureau voor de Statistiek (CBS).

E-mail: jonas.klingwort@uni-due.de

BART BUELENS is senior data scientist bij de Vlaamse Instelling voor Technologisch Onderzoek (VITO). Eerder was hij senior methodoloog bij het CBS.

E-mail: bart.buelens@vito.be

JOEP BURGER is methodoloog bij het CBS. Eerder was hij evolutionair ecoloog, evolutionair geneticus en entomoloog.

E-mail: j.burger@cbs.nl

RAINER SCHNELL heeft de leerstoel Empirisch Sociaalwetenschappelijk Onderzoek aan de Universiteit van Duisburg-Essen. Eerder was hij directeur van het Centre for Comparative Surveys van de City, Universiteit van Londen en de oprichter-redacteur van het ESRA wetenschappelijk tijdschrift *Survey Research Methods* tijdens de eerste zeven jaar. Sinds 2017 is hij adjunct-professor aan de Faculteit Gezondheidswetenschappen van de Universiteit van Curtin, Perth.

E-mail: sekretariat.schnell@uni-due.de



Foto: Naj

Er wordt wat afgemeten dezer dagen. Zo ongeveer ieder bedrijf dat zichzelf serieus neemt overspoelt zijn klanten met vragen over hun ervaringen. Zelfs mijn apotheek heeft mij de afgelopen maand drie maal een vragenlijstje gestuurd over mijn tevredenheid, zo'n mailtje wordt kennelijk automatisch verstuurd na ieder bezoek. Maar wat is de waarde van al deze gegevens, afgezien van mijn twijfel of iemand er überhaupt wel iets mee doet?

Mochten mijn kinderen deze column onder ogen krijgen dan zouden ze zuchten 'daar heb je papa weer met zijn eeuwige obsessie'. Maar ik blijf volhouden, er worden veel te vaak gegevens verzameld die nutteloos zijn of op zijn minst minder geschikt voor het doel dat men er mee beoogde.

Een veel gemaakte fout is dat men niet de juiste doelgroep benadert. Neem bijvoorbeeld een bakker die een maand lang zijn klanten een vragenlijstje over hun tevredenheid meegeeft. Heeft zo'n bakker iets aan die antwoorden? Verspilde moeite en geld, ik kan hem zonder vragenlijst al direct zeggen dat de meerderheid van zijn klanten tevreden tot zeer tevreden zal zijn. De niet-tevreden klanten zullen immers allang geen klant meer zijn! Wat die bakker *eigenlijk* wil is zijn omzet vergroten. Dat is prima, maar dan moet hij andere vragen stellen aan andere mensen. Om meer klanten te trekken moet hij juist diegenen die géén klant zijn benaderen met vragen over de redenen daarvoor. Pas dan kan hij de juiste actie ondernemen om hen zijn winkel in te krijgen. En als hij tóch zijn bestaande klanten wil ondervragen: vraag dan niet naar hun tevredenheid, maar naar wat ze missen in zijn assortiment waardoor hij meer producten aan hen kan slijten en op die manier zijn omzet kan vergroten.

Mijn favoriete slachtoffer voor dit soort boutades is de NS. Regelmatig publiceren zij vol trots de laatste

punctualiteitscijfers zoals: afgelopen maand reed 90,8 % van de treinen op tijd. Nu weet ik heus wel dat achter dit simpele getal veel meer schuilt, maar alleen al het feit dat een trein die is uitgevallen niet meetelt kan mij doen stijgen. Maar dan nog zegt dit cijfer mij weinig, wordt een boemeltje met enkele tientallen passagiers net zo zwaar meegeteld als een intercity met vele honderden reizigers? En wat is op tijd: op tijd vertrokken of op tijd aangekomen, en worden vertragingen op tussenliggende overstapstations ook geteld? Het lijkt soms wel of NS denkt dat hun taak bestaat uit het laten rijden van treinen, maar dat is niet zo. Als ze dat zo graag willen gaan ze maar op hun zolder zitten met een Märklinbaan. Hun kernopdracht is namelijk het punctueel vervoeren van *reizigers*. Die moet je meten om te kijken of ze op tijd aankomen, niet de treinen. En kijk daarbij naar het totaaltraject, een vertraging van 3 of 4 minuten halverwege kan leiden tot een vertraging van minimaal 30 minuten omdat men een overstap mist. Kijk ook niet alleen naar de vertraging die de reiziger heeft, maar vooral naar de beleving daarvan. Iemand die slechts enkele keren per jaar een dagje naar familie reist zal een vertraging ongetwijfeld anders ervaren dan een dagelijkse forens. Al was het alleen maar omdat de meeste vertragingen in de spits voorkomen en forensen helaas meestal tot de spits veroordeeld zijn en vaker reizen. Ik heb ooit de voorzichtige schatting gemaakt dat als 90% van de treinen op tijd rijdt een dagelijkse spitsforens de kans loopt op twee tot drie vertragingen per week!

Meten is fantastisch, ik heb het veel gedaan en rekenen met de resultaten kan heel bevredigend zijn. Maar doe het wel met de juiste gegevens, anders is er weinig verschil met het analyseren van random gegenereerde data.

GERRIT STEMERDINK is eindredacteur van *STATOR*.  
E-mail: gjstemerding@hotmail.com