

DE CONTAINER IN DE LOGISTIEKE KETEN

Monique van den Berg	Rijkswaterstaat (WVL)
Mathijs Jacobs	Centraal Bureau voor de Statistiek (afdeling Verkeer & Vervoer)
Robbert Janssen	Ministerie van Infrastructuur en Waterstaat (DGMo)
Ralph Meijers	Centraal Bureau voor de Statistiek (afdeling Verkeer & Vervoer)
Marly Odekerken-Smeets	Centraal Bureau voor de Statistiek (afdeling Verkeer & Vervoer)
Maria Velema	Rijkswaterstaat (WVL)



Co-funded by the
European Union

Disclaimer: Het CBS-gedeelte wordt medegefinancierd door de Europese Unie. De geuite standpunten en meningen zijn echter uitsluitend die van de auteur(s) en geven niet noodzakelijkerwijs die van de Europese Unie of Eurostat weer. De Europese Unie noch de subsidieverstrekkende instantie kan hiervoor verantwoordelijk worden gehouden.

Samenvatting

Waar komen de zeecontainers vandaan die in de Rotterdamse haven op de kade wordt gezet? Welke producten zitten erin? Via welk transportmiddel worden ze verder vervoerd (zeeschip, binnenvaartschip, trein of vrachtauto)? Waar gaan de zeecontainers, via Nederlands grondgebied, uiteindelijk naar toe? En waarom is deze informatie van belang? In dit paper wordt antwoord gegeven op die vragen. Met behulp van nieuwe data (van terminals en vervoerders) en methodieken wordt de logistieke containerketen (statistisch) beschreven.

1. Inleiding

Het CBS publiceert verschillende unimodale goederenvervoer statistieken voor zeevaart, binnenvaart, spoorvervoer en wegvervoer. Hierbij wordt ook onderscheid gemaakt naar de verschijningsvorm van de goederen zoals: droge en natte bulk, stukgoederen en containers. Goederen gaan echter zelden van productielocatie naar consumptielocatie via slechts één modaliteit. Weten welke combinatie van modaliteiten gebruikt wordt is ook waardevol voor beleidsmakers en andere gebruikers van de data. Met name voor containers is het makkelijk te wisselen van modaliteit. Containers hebben ook een uniek identificatienummer waarmee ze in principe makkelijk gevolgd kunnen worden over verschillende modaliteiten heen. In 2018 presenteerden CBS, Rijkswaterstaat (RWS) en TNO met twee samenhangende papers op de VLW over het verzamelen en samenstellen van data over goederenvervoerketens (Jacobs et al., 2018, Fransen et al., 2018).

1.1. Wat ontbreekt er in de huidige (unimodale) statistische informatie?

De unimodale goederenvervoer statistieken en de in-, uit- en doorvoerstatistiek van het CBS worden al jaren gebruikt voor onder andere goederenvervoermodellering en -prognosticering. Echter zoals omschreven in het VLW2018 paper van TNO en RWS (Fransen et al., 2018): *"... is niet bekend hoe de vervoerbewegingen per modaliteit afzonderlijk tezamen vervoerketens vormen van productielocatie naar consumptielocatie (met onderweg overslag van de ene op de andere modaliteit, in zeehavens en op inland terminals). De basis voor de registratie is namelijk de modaliteit en niet de goederenstroom. ... Het gebrek aan transportketendata vormt in toenemende mate een belangrijke beperking, onder andere vanwege de steeds grotere behoefte om vraagstukken rond vervoer vanuit een integraal ketenperspectief te benaderen (niet per modaliteit apart). Zo heeft de Brexit bijvoorbeeld impact op de doorvoer door Nederland, maar zolang binnen het achterlandvervoer onbekend is welk deel daarvan onderdeel is van een doorvoerketen kan de impact op dit vervoer niet goed worden ingeschat."*

Andere gebeurtenissen die met ketendata beter onderzocht kunnen worden zijn bijvoorbeeld de effecten van de oorlog in de Oekraïne en de importheffingen van de nieuwe regering in de VS.

Een belangrijke aanbeveling van het VLW paper in 2018 was meer microdata te verzamelen om transportketendata te verbeteren. Allereerst worden onderstaand (paragraaf 1.2) nog een paar mogelijke toepassingen van een ketenbestand genoemd. In paragraaf twee volgt een beschrijving van een voorbeeldketen waaruit blijkt hoe een keten er in de praktijk uit kan zien (vanuit het perspectief van een vrachtwagenchauffeur) en welke (logistieke) data er zoal verzameld wordt bij het vervoer van containers. In paragraaf drie wordt de stand van zaken rondom het verzamelen van (nieuwe) databronnen gepresenteerd. Vervolgens in paragraaf vier worden de methoden beschreven voor het construeren van containerketens. Paragraaf vijf gaat in op de eerste resultaten uit het onderzoek, waarna in paragraaf zes wordt uitgelegd in hoeverre de nieuwe ketendata invloed hebben op de modellering in BasGoed en de prognoses die eruit volgen.

1.2. Toepassingen

De statistische informatie over goederenvervoer van het CBS wordt gebruikt in het BasGoed model van RWS dat dient als onderbouwing van het infrastructuurbeleid. Met informatie over containerketens zouden de prognoses van dit model verbeterd kunnen worden. We komen hier in paragraaf 6 op terug. Ander mogelijke toepassingen van een ketenbestand zijn:

- Monitoring modal shift op corridors;
- Kansen voor modal shift opsporen;
- De omvang en efficiency van het huidige vervoer op een bepaalde corridor bepalen;
- Monitoring van beleidsmaatregelen;
- Belasting van het wegennet inschatten;
- Beantwoorden van overige beleidsvragen gerelateerd aan duurzaamheid, robuustheid, veiligheid en leefbaarheid.

Door de data van de losse modaliteiten te combineren kun je bovendien de losse statistieken verbeteren. Om een voorbeeld te geven: goedereninformatie wordt vrij nauwkeurig opgegeven bij de zeevaart, maar niet bij de 'achterlandmodaliteiten'. Als je de goederen kunt volgen, kun je de goedereninformatie in het achterland verbeteren.

2. Een werkdag van chauffeur Janssen

04.00 uur in de ochtend. De wekker gaat. Chauffeur Janssen staat op. Om 05.00 moet hij zich melden bij het transportbedrijf in Rotterdam waar hij sinds kort werkzaam is. Als hij de parkeerplaats van de het transportbedrijf oprijdt, ziet hij de vrachtauto waar hij voornamelijk mee rijdt, al met een 'plat' chassis eraan gekoppeld, geparkeerd staan. Hij parkeert zijn luxewagen iets verder op de personenauto-parkeerplaats en loopt direct naar 'zijn' vrachtauto.

Janssen haalt zijn tachograafpas uit zijn portemonnee en stopt die in de boordcomputer. Direct licht de 'Touch screen' van de boordcomputer op. Zijn account is geactiveerd. Vanaf het scherm leest chauffeur Janssen zijn eerste rit van de dag op. Deze is door Kees van de planning gisterenavond nog vanuit het Transport Management Systeem in het Fleet Management Systeem gezet. Doorgaans weet Janssen wel wat er als eerste transport van de dag aan ritten te wachten staat. Het transportbedrijf heeft namelijk een langlopend contract met de verlader voor het haventransport van fruit in zeecontainers, maar vandaag is het anders; er staat:

'Volle 40 ft import-container (met containernummer CBSU 123456 5, referentienummer 121212 en zegelnummer 12345), met computeronderdelen, ophalen op Maasvlakte 2 bij terminal 5; afleveren bij distributiecentrum XLDC in Venlo'.

De deur van de bijrijders stoel wordt opengedaan. Chauffeur Jacobs stapt in en rijdt vandaag als mentor mee. Chauffeur Janssen zet de vrachtauto in zijn achteruit en rijdt de parkeerplaats af, de openbare weg op.

Net over zes uur in de ochtend, arriveren chauffeur Janssen en Jacobs bij de terminal op de Maasvlakte. Janssen rijdt de vrachtauto tot vlak bij de paal met de (self-service)-Kiosk. Hij stopt zijn 'cargocard' in de gleuf onder het beeldscherm van de kiosk en legt zijn wijsvinger op de vingerscan. Helaas krijgt hij nog niet te zien dat hij geautoriseerd is verder te rijden. 'De vingerscan komt niet overeen met de chauffeurspas', geeft het kiosk-scherm aan. Via de 'vrachtauto-problem-route' moet hij zich een kilometer verderop melden bij een securitykantoor op de terminal. Daar wordt alles nog een keer gecheckt en blijkt het gelukkig toch in orde. Na zich nogmaals bij de kiosk te hebben gemeld, voert hij weer de cargocard in en legt hij opnieuw zijn vinger, dit keer zijn middelvinger, onder de scan. Nu wordt hij door de kiosk toegestaan de 'TAR' code (Truck Appointment Registration) in te toetsen wat de planner van te voren voor hem heeft geboekt. Via de kiosk krijgt chauffeur Janssen te zien dat de container op blokkengroep 23 beschikbaar is en wordt het routeplan getoond. Álle signalen staan op groen, Janssen mag doorrijden.

Met deze signalen 'op groen' wordt onder meer bedoeld, nadat Kees van de planning vanuit de transporteur een *voormelding* heeft gedaan in Portbase, dat de rederij (dé klant van de terminal), via de cargadoor, groen licht heeft gegeven, oftewel de pick up rechten verleend, voor het ophalen van de container. Verder is er geen blokkade meer vanuit de douane / de vergunning voor Ruimte voor tijdelijke opslag (RTO) is verleend. Tenslotte is er dus ook een *exacte* match is met het transportbedrijf van Janssen, waaruit blijkt dat zijn bedrijf (digitaal) gemachtigd is, door de door de verlader ingeschakelde expediteur de container op te halen. Bij de terminal komen zowel de goederenstromen als de informatiestromen bij elkaar.

Met een achterwaartse boog, onder toeziend oog van zijn mentor, manoeuvreert chauffeur Janssen even later behendig zijn vrachtauto achteruit op een van de lege parkeerplaatsen van blok 23. Nadat hij de twistlocks open heeft gezet, meldt hij zich bij de kiosk van de blokkengroep, voert zijn cargocard in en bevestigt een aantal vragen: of het voertuig klaar voor ontvangst is en op de juiste plaats is gezet. Het volledig geautomatiseerde terminal operating systeem weet nu dat de container op de vrachtwagen kan worden gezet. Na een minuut of tien komt de volle container op hoogte aanvliegen. Janssen en Jacobs stappen uit de cabine en gaan bij de kiosk staan. Janssen drukt op de knop ten teken dat alles is in orde is en de container via de automatische kraan op het chassis kan worden gezet. Chauffeur Janssen houdt de knop ingedrukt, voor de kraan het bewijs dat de chauffeur, tijdens het gehele laadproces, niet in of nabij de vrachtauto is. De gerobotiseerde kraan plaatst zonder enige menselijke tussenkomst de container feilloos op het chassis. Chauffeur Janssen loopt meteen rondom het voertuig om de twistlocks vast te maken en de container te bevestigen aan het chassis.

Janssen maakt ondertussen wat foto's en controleert of het containernummer en zegelnummer overeenkomen met wat hij op zijn smartphone (die gekoppeld is met de boordcomputer) heeft staan.

De nummers kloppen. Dat stelt gerust dat hij de juiste container meeneemt en dat het onwaarschijnlijk is dat er met de lading is geknoeid. Ook constateert hij geen grote deuken of andere beschadigingen aan de container. Daar hoeft hij dit keer niets over te melden.

Janssen stapt de cabine weer in en start de motor van de vrachtwagen. Terwijl de dieselmotor 'lucht draait' om de remdruk op te voeren en de luchtvering op hoogte te brengen, kan hij ondertussen onder het genot van warme koffie uit zijn thermosfles de CMR-vrachtbrief invullen. Op basis van de ritgegevens in de boordcomputer vult hij gegevens in over de afzender, bestemming, ISO-type container en soort lading en gewicht in. Dat is allemaal snel voor elkaar en ze kunnen vertrekken. Bij het uitrijden van de terminal wordt de container via allerlei camera's en scanners gecontroleerd op schade, en ontvangt Janssen uiteindelijk nog het interchange bonnetje, als bewijs dat de container daadwerkelijk is opgehaald.

De rit naar Venlo duurt bijna 3 uur geeft het navigatiesysteem op de boordcomputer van de vrachtauto aan, maar chauffeur Janssen weet dat dit in de praktijk nog wel eens kan tegenvallen. Het systeem is nog niet dermate geavanceerd / up-to-date, dat alle files tijdig en correct worden gemeld. Een hele andere ervaring dan de moderne navigatieapps op zijn privé-smartphone. De boordcomputer meldt inmiddels een nieuwe rit met de volgende gegevens:

'Volle 20ft export-container (met containernummer RWSU 123456 1, referentienummer 898989 en zegelnummer 98765), met lederwaren ophalen bij LoodS in Venray; afleveren bij terminal op de Maasvlakte'.

Het verkeer zit vandaag niet mee vanwege het slechte weer. Een uur na de geplande aankomsttijd arriveert Janssen bij de terminal in Venlo. Dit is door de planningsafdeling al gemeld bij de terminal, waar de rijdende vrachtauto's via GPS (vanuit de boordcomputer) real-time worden gevolgd. Omdat het bedrijf ook gebruik maakt van het systeem van geofencing, is er een automatische melding naar de planning verstuurd op het moment dat de vrachtauto de eindbestemming naderde. De meldingen vanuit het geofencing systeem voorkomen dat chauffeurs tijdens het rijden handmatig berichten moeten sturen, terwijl de planners wel meteen op de hoogte zijn dat de aflevering aanstaande is.



Figuur 1: Tracé Corridor Zuid - Maasvlakte / haven Rotterdam / A16 / A58 / A2 / A67 (via de weg)

Ook in Venlo zit het tegen. Nadat de 40ft import-container nog wel snel en probleemloos wordt gelost, het is op dat moment 13.00 uur, blijkt er een storing bij de douane te zijn waardoor het douanedocument voor de lading in de 20ft export-container niet beschikbaar komt. Zonder het douanedocument mag de export-container niet worden vervoerd en geldt er een minimale wachttijd van drie uur. Voor chauffeur Janssen en mentor Jacobs zit er vervolgens niets anders op dan te wachten voor het dock. Iets voor 16.00 uur komt er dan toch groen licht en een paar minuten later zijn ze op weg naar de Rotterdamse Maasvlakte.

Op de A16 nabij Breda ontvangt Jacobs een telefoontje. Een zieke collega-chauffeur moet naar het ziekenhuis en zijn trekker-opleggercombinatie staat in de buurt van Oosterhout langs de A59. De planning instrueert de chauffeurs om de combinatie op te gaan halen door een stuk om te rijden. Chauffeur Jacobs neemt het stuur van de stilstaande vrachtwagencombinatie en chauffeur Janssen rijdt met de andere vrachtwagen alleen terug. Gelukkig is er op de container op de vrachtwagen geen strakke deadline van toepassing, de cargo closing time. Na deze tijd is het niet meer toegestaan om exportcontainers af te zetten op de terminal voor de specifieke afvaart van het diepzeeschip. Die is echter pas over een paar dagen. Dus hoewel de zeeterminal voor vrachtauto's 24/7 open is voor laden en lossen van containers, hoeft er vanavond niet meer naar de Maasvlakte te worden gereden. De combinatie kan vannacht worden geparkeerd op de standplaats van het transportbedrijf. Dit is ook noodzakelijk om te kunnen voldoen aan de verplichte maximaal rijtijden (10 uur) en arbeidstijden.

Na het afkoppelen van de trekker van de oplegger en het naar huis rijden in zijn luxewagen komt chauffeur Janssen weer thuis. Toch is hij nog wel even benieuwd naar resultaten van zijn rijgedrag en prestaties van de dag. Hij opent op zijn smartphone een app waar hij gegevens over onder meer brandstofverbruik, gebruik van cruise control, remgedrag en stationair draaien kan ophalen vanuit de boordcomputer in het voertuig. Chauffeur Janssen is niet ontevreden, al is er zeker nog ruimte voor verbetering constateert hij. Vooral het achteruit docken kost nog veel tijd en brandstof. Maar nu gaat hij snel naar bed. Morgen gaat de wekker om 04.00 uur en is er direct in de ochtend weer een container af te leveren.

2.1. Acht weken eerder, Sjanghai – China

De container met nummer: CBSU 123456 5, dezelfde container die chauffeur Janssen acht weken later van Maasvlakte 2 naar Venlo brengt, wordt geladen op een 'deep sea' schip in de zeehaven van Sjanghai. De container bevat airfryers die enkele weken daarvoor weer zijn geproduceerd in de Chinese regio Guangzhou. Het schip zal een extra lange weg moeten afleggen. De route via de Rode Zee naar het Suezkanaal, de kortste route naar Europa is ernstig verstoord door aanvallen van Houthi's op schepen. Daarom zal er nu om Afrika heen moeten worden gevaren. De onzekerheid maakt ook dat andere rederijen de keuze maken het vertrek van het zeeschip uit te stellen. Voor de kadeplanning van de Nederlandse terminal wordt het er niet makkelijker op. Verstoringen bij de aankomst liggen door deze

onzekerheid extra op de loer. Het schip doet, voordat hij aanmeert op Maasvlakte 2, achtereenvolgens nog de zeehavens aan in Singapore en Colombo (Sri Lanka).

Om het uiteindelijke datum/tijdstip van aankomst van het schip goed te kunnen schatten (Estimated time of arrival) monitort de kadeplanning via websites realtime de voortgang van de schepen. Uiteindelijk gaat de kadeplanning zo'n 8 dagen vooruit en dan nog is de planning continu aan verandering onderhevig. Nu was de omvaaroute van dit schip vanwege de aanvallen van Houthi's uiteraard vroeg bekend, maar allerlei andere verstoringen zoals wind en zoals dit keer: mist, maken dat de planning steeds moet worden aangepast. De mist was nu zo dicht en hardnekkig dat het schip dagenlang 'lay-by' lag alvorens het door de loodsdiens naar de aanmeerplek kon worden genavigeerd. Het schip zou na deels lossen van de bovenste lagen containers (waaronder de container die door chauffeur Janssen gaat worden meegenomen), waardoor de diepgang van het schip hoger kwam te liggen en waardoor het ook mogelijk werd verder te varen naar Hamburg, verder varen naar die haven.

3. Van mono- naar multimodaal in samenwerking met de transportsector

In Jacobs et al. (2018) worden de resultaten besproken van het construeren van een containerketenbestand door gebruik te maken van enerzijds registraties voor de zeevaart (inclusief goedereninformatie van de douane), binnenvaart en spoorvervoer en anderzijds de wegvervoerenquête. De beschikbare data bleek niet genoeg om voldoende ketens af te leiden: Het resultaat was op dat moment te selectief en moeilijk op te hogen.

3.1. Dataverzameling 2.0

Het CBS heeft standaard alleen toegang tot overheidsregistraties. Indien er een verordening is, kan het CBS aanvullend benodigde informatie uitvragen bij bedrijven. Zo'n verordening is er (nog) niet voor multimodaal containertransport. Het praktijkvoorbeeld in de vorige paragraaf laat zien welke (logistieke) data er zoal verzameld wordt over het vervoer van containers en dat deze vaak bij private partijen beschikbaar is.

In een pilot die het CBS samen met RWS heeft uitgevoerd, is aan bedrijven gevraagd *vrijwillig* logistieke data te delen uit hun administratieve systemen. Het doel is om de informatie voor beleidsbeslissingen te verbeteren, waarvan ook bedrijven zouden moeten profiteren. Daar is overwegend positief op gereageerd. Na deze (succesvolle) pilot is de dataverzameling verder uitgebreid en dit proces loopt nog. Het vergt meestal meerdere contactmomenten om een bedrijf te overtuigen deel te nemen aan het project en vaak duurt het dan nog een tijd voor een stabiele maandelijkse datastroom ingeregeld is. Bijzonder (voor het CBS) is nog dat bedrijven de data niet in een specifiek format hoeven aan te leveren. Voor de wegvervoerenquête (en andere onderzoeken van het CBS) wordt een vertaling gemaakt van de bedrijfsinformatie naar een gestandaardiseerd format waarin de data aangeleverd wordt. Dit kost tijd en geld en blijkt een belemmering. Een open format vergroot de bereidheid om data te delen aanzienlijk, maar is wel (in principe eenmalig) meer werk voor het CBS.

3.1.1. Databronnen

In de pilot is gekeken waar de belangrijkste lacunes zijn in de data die het CBS al tot zijn beschikking heeft en welke databronnen die kunnen aanvullen. De minste data is beschikbaar over het wegvervoer, omdat de enquête gebaseerd is op een steekproef en bedrijven maar gedurende 1 week hun ritten rapporteren. Uit data van terminals blijkt echter dat er meer dan 2.000 bedrijven actief zijn in het wegvervoer van containers. Dit zijn grotendeels bedrijven met maar 1 werkzame persoon, waarbij de digitalisering van de transportprocessen lager zal zijn dan bij grote wegvervoerders. De focus lag daardoor eerst op de andere modaliteiten. Als die goed in beeld zijn, wordt gekeken hoe het missende wegvervoertransport zo efficiënt mogelijk in kaart gebracht kan worden. De ontwikkelingen rond het Open Trip Model (OTM) kunnen hierin natuurlijk een grote rol spelen.

Uit de pilot bleek ook dat de beschikbare data voor het spoorvervoer onvoldoende is. ProRail levert het CBS informatie over alle treinen die op de ProRail infrastructuur rijden in Nederland, maar er is geen detailinformatie over de vervoerde containers. De verder beschikbare informatie over de per spoor vervoerde containers is te selectief om dit naar alle treinen te extrapoleren. Spoorvervoerders zijn vaak actief op bepaalde trajecten en daardoor is de data maar in beperkte mate vergelijkbaar. RWS en CBS proberen daarom zo veel mogelijk spoorvervoerders over te halen data te delen voor het construeren van containerketens.

De situatie voor de binnenvaart is anders. CBS maakt gebruik van IVS NEXT data van RWS en dat is in theorie een integrale registratie van het vervoer via de binnenvaart in NL. In de praktijk blijkt de dekking van opgegeven containers met containernummers op bepaalde trajecten en met name ook van lege containers niet volledig. Dit heeft deels te maken met de locatie van telpunten en deels met de vereisten vanuit BICS waarbij alleen voor gevaarlijke stoffen een containernummer vereist is. Voor bepaalde trajecten is het percentage bekende containernummers laag. In de pilot is data van bepaalde binnenlandterminals opgevraagd om deze lacunes te verhelpen. Daarbij bleek deze data ook andere toegevoegde waarde te hebben.

De binnenlandterminals hebben vaak informatie over het transport van de zeeterminals via de binnenlandterminal of direct naar de klant en vice versa. Die last en first mile vinden vaak plaats via de weg. Het is daarom een interessante bron om dit stuk van het wegvervoer in kaart te brengen. De Nederlandse binnenlandterminals behoren tot een beperkt aantal bedrijven en de logistieke data is goed beschikbaar uit administratieve systemen. Het is daarom efficiënt om de first en last mile van het wegvervoer via deze partijen te verzamelen.

Daarnaast volgt uit terminaldata, zowel in het binnenland als in de havens, de opslagtijd van de containers bij de terminal. Zonder deze opslagtijd ontstaat er onzekerheid of een container nog dezelfde inhoud heeft indien de container volgens vervoerder 1 van A naar B vervoerd wordt en bijvoorbeeld 30 dagen later via dezelfde of een andere vervoerder van B naar C. In de tussenliggende 30 dagen kan een container mogelijk meerdere keren tussen NL en GB vervoerd zijn en dan is de inhoud van de container zeer waarschijnlijk gewijzigd. Doordat de opslagtijd bij terminals bekend is, kunnen stukken

uit de containerketen verbonden worden met strengere eisen aan de koppeling en wordt het risico op koppelfouten gereduceerd.

In de zeeterminaldata is de herkomst van een binnenkomende container of de bestemming van een uitgaande container niet aanwezig. Wel is er informatie over de vorige / volgende modaliteit en vaak ook een id van het transportmiddel. Deze variabelen zijn nuttig bij het koppelen van aangrenzende legs in de keten en worden ook gebruikt bij de ophoging.

3.1.2. Variabelen

Niet alle datasets bevatten dezelfde variabelen. Terminals registreren met name de aankomst- en vertreklocatie en -momenten van containers bij de terminal. Voor de binnenlandterminals is soms ook nog de herkomst en bestemming als naam aanwezig, terwijl die variabelen bij zeeterminals altijd ontbreken. De bijbehorende datums en eventueel tijden worden ook nauwkeuriger bepaald bij de terminal zelf dan voor een locatie waar het naar toe gaat of vandaan komt. De modaliteit waarmee inkomende en uitgaande containers vervoerd worden is altijd aanwezig net als de belangrijkste variabele: het containernummer. Het containertype is vaak ook bekend. In veel gevallen is nog een id van het transportmiddel aanwezig zoals een kenteken, treinnummer, scheepsnaam of -nummer of een reisid. Voor binnenlandterminals is soms ook nog het (bruto)gewicht en een omschrijving van de goederen aanwezig. Afhankelijk van de terminal kan er dan ook nog zoiets als een boeking nummer aanwezig zijn.

Data van vervoerders (spoor en weg) zijn iets anders opgebouwd. Er is per containernummer een herkomst en bestemming met soms alleen een vertrekdatum en vaak geen exacte tijden. Het containertype is soms aanwezig, het gewicht vaak wel en minder vaak een goederenomschrijving en dan alleen bij het wegvervoer. De modaliteit is niet altijd aanwezig, maar volgt uit de vervoerder. Het containertype is soms aanwezig en dan uitsluitend bij het wegvervoer. Zowel voor spoor- als wegvervoer is het id van het transportmiddel aanwezig: kenteken bij wegvervoer en trein-, wagon- en/of reisnummer voor spoorvervoer. Afhankelijk van de dataset kunnen er nog extra, vaak administratieve, variabelen aanwezig zijn waar soms de bestemming beter uit bepaald kan worden.

4. Construeren van containerketens

Voor dit project wordt een groot aantal bronnen gebruikt afkomstig van verschillende dataleveranciers. Zoals besproken in de vorige paragraaf verschilt de inhoud per type bron: data van zeeterminals is bijvoorbeeld anders opgebouwd dan die van een spoorvervoerder. Daarnaast is er in dit project voor gekozen geen vast input format op te leggen. Er is wel een gewenste minimale hoeveelheid variabelen waarvan het containernummer één van de belangrijkste is, maar dataleveranciers leveren de data zoals dat voor hen het eenvoudigste is. Er wordt wel gepoogd de datalevering te standaardiseren, maar dit mag geen belemmering zijn voor het delen van data. De verzamelde data kunnen daardoor o.a. in de vorm van een Excel, json, XML, csv of Pdf's worden gedeeld. De data worden via beveiligde uploadkanalen naar het CBS gestuurd of opgehaald via een API.

In een aantal stappen worden de verschillende datasets geprepareerd voor statistische verwerking. Om snel in te kunnen spelen op veranderingen in de aangeleverde data is de verwerking modulair opgezet en worden verschillende type bestanden in eerste instantie apart van elkaar verwerkt. De eerste stap is het uniformeren van de datasets en standaardiseren van de variabelen waaronder de variabele namen en -typen. De ruwe data worden omgevormd naar tidy data; zie bijvoorbeeld (Wickham et al. 2023): In een tidy dataset is elke variabele een kolom, elke observatie een rij en elke waarde een cel. Specifiek voor datasets die verplaatsingen bevatten wordt de data, indien nodig, zo omgevormd dat elke rij een beweging van A naar B representeert.

Nadat de datasets technisch in orde zijn, worden ze inhoudelijk gecontroleerd en waar nodig aangepast. Een belangrijke stap is het coderen van locatieaanduidingen. Die zijn niet gestandaardiseerd in de ontvangen datasets. Vaak zijn het namen of afkortingen van bijvoorbeeld terminals of een plaatsnaam. Via koppellijsten en een geocodeerservice worden de geografische variabelen gecodeerd naar standaardindelingen, zoals gemeente, UN/LO code en waar mogelijk postcode. Ook de goederenomschrijvingen, die enerzijds van de douane komen en anderzijds van terminals en vervoerders, worden via een tekstclassificatie algoritme omgezet naar standaardindelingen, zoals NST2007, NSTR en ADR klasse. Verder worden inhoudelijke controles uitgevoerd zoals het maximaal toegestane brutogewicht voor een type container en consistentie tussen netto, bruto en brutoplus gewicht.

Nadat de losse stukken van de containerketens, de legs, per modaliteit inhoudelijk zijn gaafgemaakt, kunnen de verschillende delen gecombineerd worden om te proberen volledige ketens te construeren. Uit deze samengevoegde data kunnen verschillende outputtabellen gemaakt worden. Enkele voorbeelden zijn: een model split van en naar de zeehavens, een statistiek over de combinatie van verschillende modaliteiten of de trajecten die containers volgen uitgesplitst naar goederensoort.

Bij het samenvoegen kunnen nieuwe inconsistenties optreden; twee datasets kunnen in zichzelf consistent en valide zijn, maar door ze te combineren kan blijken dat bepaalde waardes niet valide zijn. Zo kan het gewicht van een container verschillen tussen twee datasets hoewel het om dezelfde container

met inhoud lijkt te gaan. Ook de omschrijving van de goederen kan anders zijn. Vaak is de omschrijving bij de douane een stuk uitgebreider dan in de data van vervoerders. Ook tijden blijken vaak niet consistent. Enerzijds kan dat komen door een definitieverschil: een terminal registreert het tijdstip dat het schip de kade verlaat of specifieker het tijdstip dat een container op het schip geladen wordt en IVS registreert het passeren van een telpunt. Dat laatste kan voor hetzelfde schip een aantal uur of zelfs meer dan 1 dag later zijn. De vertreklocatie van het schip in beide datasets kan in dit geval ook verschillen.

Voor het afleiden van containerketendata moet eerst bepaald worden welke legs van een container bij elkaar horen. Hiervoor worden de containerrecords per containernummer op starttijd van de leg gesorteerd. Dan kan bekeken worden of de tijden en locaties van opeenvolgende records compatibel zijn of dat er een stuk van de keten mist. Er mag ook niet te veel tijd tussen twee opeenvolgende records zitten, omdat anders de kans bestaat dat de container tussendoor nog andere verplaatsingen heeft meegemaakt en de inhoud gewijzigd is. Dan is het ook belangrijk te bepalen wat het begin en einde van een keten is. Als dat gevonden is, horen alle tussenliggende legs bij één containerketen. Het begin en einde van de ketens volgt soms uit de dataset, bijvoorbeeld als een container vol wordt afgeleverd bij een klant en leeg terugkomt naar een terminal. Een andere situatie is als een container naar het buitenland gaat en later weer terugkeert. We gaan er dan vanuit dat het om een nieuwe containerketen gaat.

Als laatste stap wordt gecorrigeerd voor missende data. Daarbij wordt eerst gekeken of er relaties in de data zitten waaruit een missende leg afgeleid kan worden. Als deze stap uitgevoerd is, worden missende legs geïmputeerd met een random imputatiemethode. Hiervoor kan bijvoorbeeld een K-Nearest Neighbour (knn) imputatie methode gebruikt worden. Het idee is dat de waarnemingen die het meest op elkaar lijken ook waarschijnlijk vergelijkbare waarden hebben voor de ontbrekende variabelen.

Een andere rekentechnisch minder intensieve manier om een waarde te imputeren op basis van vergelijkbare records is random hot deck (rhd) imputatie. Hierbij kan de data in groepen opgedeeld worden op basis van een aantal voorspellende variabelen. Daarna wordt een random waarde getrokken uit de records die tot dezelfde groep behoren. Als eerste model hebben we gekozen voor een rhd imputatie. Vanwege de grote hoeveelheid data, kost het veel tijd imputatiewaarden te berekenen en dit gaat sneller met de rhd imputatie dan met de knn imputatiemethode. Daarbij hebben we de volgende groeperingsvariabelen geselecteerd:

- Startlocatie (unlocode),
- Modaliteit (indien bekend),

om bijvoorbeeld een onbekende eindlocatie (unlocode) te imputeren.

We zien nog veel mogelijkheden en mogelijke methoden om de schattingen te verbeteren, bijvoorbeeld een ophoging voor de ketens waarvan helemaal niks bekend is. Hiervoor is nog meer onderzoek nodig.

5. Resultaten

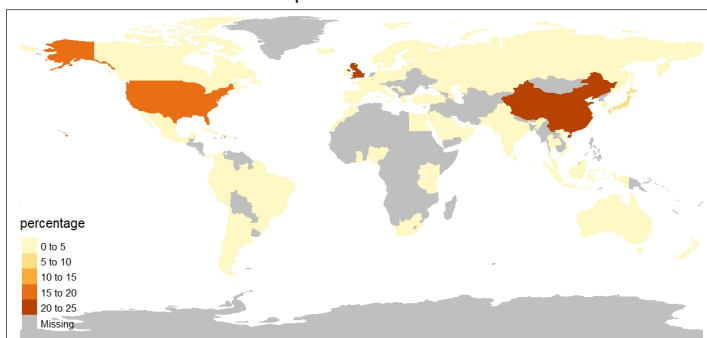
In paragraaf twee omschreven we een logistieke keten van in het Chinese binnenland geproduceerde airfryers, die in een zeecontainer vanaf de zeehaven in Sjanghai werden verscheept en met tussenstops in Singapore en Colombo in Rotterdam aankwamen op de Maasvlakte II om daar door chauffeur Janssen (na een melding op zijn boordcomputer) te worden opgehaald om vervolgens bij een distributiecentrum in Venlo te worden gelost. Dit is slechts één voorbeeld van de grote hoeveelheid zeecontainers die jaarlijks in een logistieke keten over Nederlands grondgebied worden getransporteerd.

In 2023 werden in totaal 4,2 miljoen zeecontainers op de kades van de Nederlandse zeehavens gelost. Ongeveer 36 procent (1,5 miljoen) van die containers gingen direct weer Nederland uit, per (feeder)zeeschip (transshipment). Ongeveer 2,7 miljoen van de containers werd naar het achterland vervoerd. Naast de container die chauffeur Janssen ophaalde, werden nog ongeveer 1,6 miljoen andere containers door vrachtwagens bij zeeterminals opgehaald (ongeveer 61 procent van alle containers die naar het achterland gaan). Per trein ging 7 procent verder (0,2 miljoen) en per binnenvaartschip 0,9 miljoen (32 procent). Dit kan worden gezien als de modal split van de overslag van zeecontainers bij zeeterminals in 2023.

Zeecontainers die per vrachtauto bij een zeeterminal worden opgehaald kunnen natuurlijk ook (enkele kilometers) verderop bij een binnenvaart- of spoorterminal weer op een binnenvaartschip of trein worden overgezet. Het aandeel wegvervoer in de modal split kan hierdoor, als bijvoorbeeld een modal split in een regio ruimer dan alleen bij de zeeterminals wordt bekeken, lager uitvallen.

Ongeveer 24 procent van de containers die bij Nederlandse zeeterminals in 2023 werden gelost hebben als eerste laadzeehaven een haven uit China. Twee van de belangrijkste productsoorten, (gemeten in aantal beladen containers), die naar Venlo worden gebracht vanuit Nederlandse zeeterminals zijn transportmiddelen (NST2007 = 12, dat is inclusief producten van de auto-industrie) en textiel(producten) en leder(waren) (NST2007 = 05). Onderstaand worden deze stromen visueel in kaarten getoond.

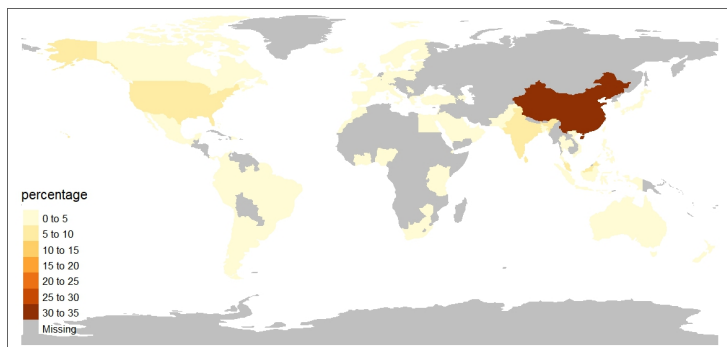
Aandeel containers met transportmiddelen naar land van herkomst



Figuur 2: Aandeel containers met transportmiddelen (NST 12) naar Venlo via alle modaliteiten uitgesplitst naar land van herkomst

De meeste naar Venlo vervoerde goederen met NST 12 komen uit Groot Brittannië en China, gevolgd door de VS.

Aandeel containers met textiel naar land van herkomst



Figuur 3: Aandeel containers met textiel (NST 05) naar Venlo via alle modaliteiten uitgesplitst naar land van herkomst

Verreweg het meeste textiel naar Venlo wordt per zeeschip vanuit China verscheept, gevolgd door Verenigde Staten en India.

Bij bovenstaande resultaten is in het achterland (Zeeterminal - regio Venlo) een relatief simpele weging toegepast. In de (nieuw) verzamelde data zijn er ruim 80 000 containers tussen Nederlandse zeeterminals en regio Venlo die gekoppeld kunnen worden aan douane data, waardoor ook gegevens over de inhoud van containers en vorige zeehavens bekend zijn. Hiervan zijn er ruim 11 000 containers uit China. Vanuit de niet aan de douane gekoppelde (mono) transportstatistieken wordt het aantal containers vanuit Nederlandse zeeterminals naar regio Venlo geschat op 100 duizend. Voor een geaggregeerd overzicht van huidige koppelpercentages zie Tabel 1. Deze koppelpercentages worden de komende periode verder verhoogd met extra waarneming. Ook zullen er geavanceerdere weeg- en imputatiemethoden worden ontwikkeld / toegepast (zie paragraaf 7. Conclusies en vervolg).

Tabel 1: Koppelpercentages zeeterminal – achterlandmodaliteiten

<i>Modaliteit</i>	<i>Geschatte dekkingsgraad van verkregen microdata</i>
Zeevaart	95%
Binnenvaart	70% IVS en 31% binnenlandterminals, gecombineerd 76%
Spoor	41%
Wegvervoer	8%

Het koppelpercentage voor binnenlandterminals lijkt vrij laag. Dit is echter het aandeel containers dat een zeeterminal via de binnenvaart verlaat en in een binnenlandterminal dataset gevonden wordt. Het CBS ontvangt echter niet van alle binnenlandterminals gegevens. Bovendien gaat een deel van de containers direct naar het buitenland zonder bij een binnenlandterminal overgeslagen te worden.

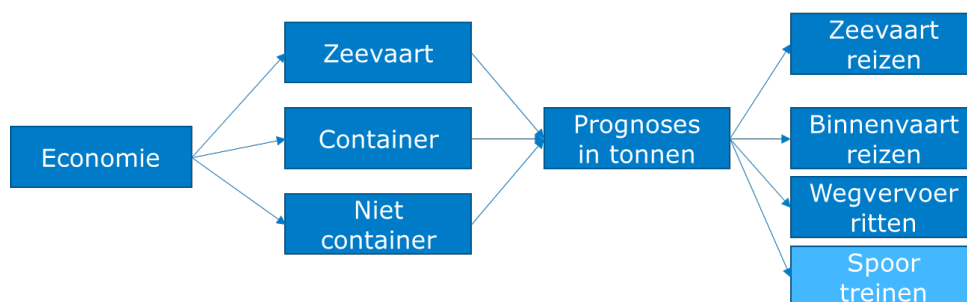
Het percentage containers dat een zeeterminal via de weg verlaat en dat gevonden wordt in beschikbare microdata is laag. Dit komt door het steekproefkarakter van de wegvervoerenquête en de beperkte

rapportageperiode. Via XML worden wat extra containerritten verzameld, maar het blijft in totaal steken op zo'n 8%. Echter bevat de data van binnenlandterminals ook informatie over de last en first mile tussen de binnenlandterminal en de klant. In totaal gaat het om bijna 870 duizend ritten via de weg wat neerkomt op ruim boven de 50% van het transport tussen terminal en klant (bestemming / herkomst). Een exact percentage is op dit moment niet te geven, omdat niet alle terminaldata beschikbaar is en de IVS en terminaldata niet helemaal vergelijkbaar lijken. Hier wordt momenteel verder onderzoek naar gedaan.

6. Discussie en vooruitblik

In deze paragraaf bespreken we wat de nieuwe ketendata voor invloed kunnen hebben op de modellering in BasGoed en de prognoses die eruit volgen. Wat verandert er met de nieuwe data aan het maken van prognoses?

Vanuit RWS gebruiken we het model BasGoed om prognoses te maken van het goederenvervoer in de toekomst. Deze worden onder andere gebruikt voor het bepalen van problemen op het (vaar-)wegennet, en bij het ontwerp van mogelijke oplossingen. Ook worden algemenere beleidsvragen rondom het goederenvervoer, zoals modal split vraagstukken, vrachtwagenheffing, en milieuvervuiling er mee geanalyseerd. Het model bestaat uit verschillende onderdelen, zie onderstaande figuur. Het model begint met het vertalen van economische verwachtingen naar hoeveelheid te vervoeren goederen. Vervolgens wordt deze hoeveelheid gesplitst in zeevaart, container en niet container vervoer. Voor elk van deze drie wordt bepaald hoeveel tonnen er van waar naar waar vervoerd worden, en met welke vervoerswijze. Dit geeft de prognoses in tonnen. Vervolgens wordt per vervoerswijze nog bepaald hoeveel voer-/vaartuigen hiervoor nodig zijn.



Figuur 4: Modules in het BasGoed model

Om de mogelijkheden van nieuwe ketendata te laten zien bekijken we ook hier de reis van één container met goederen vanuit China via Rotterdam naar Venlo. Dit keer niet rechtstreeks vanuit Rotterdam per vrachtauto naar Venlo (zoals de container die chauffeur Janssen vervoerde, beschreven in paragraaf twee, maar nu eerst per binnenvaartschip van Rotterdam naar de terminal in Wanssum, om vervolgens het laatste stukje per vrachtauto af te leggen naar Venlo. Dit keer worden er geen airfryers vervoerd, maar kantoormachines/computers. Zie de groene lijn in onderstaande figuur.



Figuur 5: Voorbeeld van een waargenomen containerketen vanuit China via Rotterdam en Wanssum naar Venlo (groen) vs een gemodelleerde reis met laatste leg van Wanssum naar Deurne

Het model beschrijft de reis in verschillende stappen, die elk beïnvloed worden door het gebruik van nieuwe ketendata. Onderstaand volgt een beschrijving van hoe nieuwe ketendata leidt tot verbetering binnen de containermodule binnen BASGOED. Voor een korte algemene beschrijving van de modules binnen BASGOED, zie (Rijkswaterstaat, 2021).

De container module is de eerste module waarop het verbeterde bestand effect heeft. De module is eerder toegelicht op de VLW (de Bok et al., 2016). Deze module beschrijft hoe het vervoer over het landzijdige deel van de reis (Rotterdam-Venlo) zal groeien.

Als startpunt van deze module is data nodig over de landzijdige ketens, eigenlijk de landzijdige kant van de nieuwe dataset. Deze data bestaat in de oude opzet niet, en daarom worden de uni-modale bestanden die er wel zijn gecombineerd om een benadering te maken, zie voor de volledige methode (Fransen et al., 2019).

Uitgangspunt is dat spoor- en binnenvaartreizen tussen terminals plaatsvinden, en dat er van en naar deze terminals transport over de weg plaatsvindt. Reizen uit het wegvervoer-bestand die voldoen aan de criteria om voor- en natransport te zijn, worden aan binnenvaart en spoorreizen gekoppeld. Omdat het wegvervoer bestand een steekproef is, zijn er vaak niet voldoende wegvervoerritten beschikbaar, waardoor wel waargenomen ritten gekopieerd moeten worden.

Het wegvervoer dat niet als voor- of natransport wordt geselecteerd maar wel containers vervoerd, wordt als direct wegvervoer in het ketenbestand opgenomen.

Op basis van dit ketenbestand wordt een productie-consumptie (PC) matrix samengesteld, waarbij de herkomst van de eerste link in de keten de P wordt, en de bestemming van de laatste link de C. Deze PC-matrix wordt gebruikt in de rest van de module.

In ons voorbeeld zou op die manier de keten Rotterdam-Wanssum-Deurne gemaakt kunnen worden. Deze keten komt niet overeen met de waargenomen keten Rotterdam-Wanssum-Venlo, maar aangezien er niet meer informatie bekend is in de huidige data, zou toch uitgegaan kunnen worden van Deurne. In de figuur is dit aangegeven met een rode pijl.

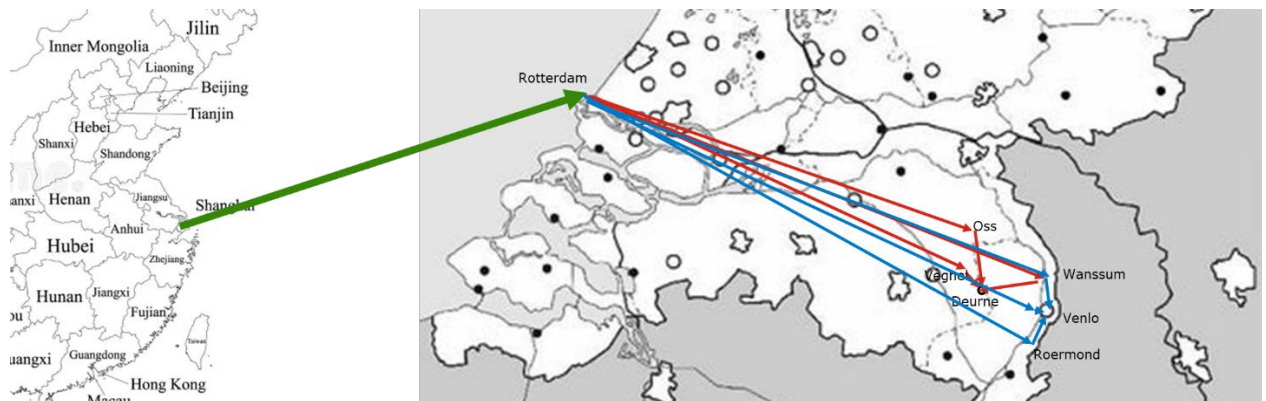
Als het nieuwe ketenbestand beschikbaar komt, is het wiskundig afleiden van de PC matrix niet meer nodig, deze kan dan direct op het nieuwe bestand gebaseerd worden. Hierdoor zijn de geselecteerde PC relaties daadwerkelijk in gebruik, en weten we zeker dat het voor- en natransport aan de juiste lange afstand stromen gekoppeld is. Het voorbeeld wordt in dit geval dus wel goed gevolgd: Rotterdam-Wanssum-Venlo.

Bij het maken van prognoses wordt de economische groei die in de economiemodule berekend is toegepast op de goederenstromen in het huidige jaar. De economiemodule geeft de groei in de productie en de attractie van zones, deze groei moet vertaald worden naar de groei in reizen. Om deze vertaling te maken wordt naar de groei van de herkomstzone gekeken, en naar die van de bestemmingszone. Via een furness methode wordt deze omgerekend naar de groei van de reizen tussen de herkomsten en bestemmingen.

In de huidige versie van het model zou voor ons voorbeeld naar de groei van Rotterdam en Deurne gekeken worden. De combinatie van de groei van die twee leidt dan tot de groei van de hele keten Rotterdam-Wanssum-Deurne. Met het nieuwe ketenbestand zou gekeken worden naar Rotterdam en Venlo, en zou de keten Rotterdam-Wanssum-Venlo opgehoogd worden. Voor het verkeer op de binnenvaart link Rotterdam-Wanssum zou dit een klein verschil kunnen maken, als de groei van Venlo anders is dan die van Deurne. Voor de verwachting van het wegverkeer maakt het wel veel verschil: óf Wanssum-Deurne groeit, óf Wanssum-Venlo. Dus óf het is druk op de N270, óf op de A73. Als op één van deze wegen een knelpunt wordt voorspeld, zou het kunnen dat maatregelen op de verkeerde plaats genomen worden.

Tot nu toe gingen we er van uit dat altijd dezelfde terminal gebruikt wordt voor een reis. Echter, in het model kunnen verschillende terminals gebruikt worden. Aan de hand van de afstanden en reistijden en de locatie van multimodale terminals worden mogelijke multimodale transportketens gegenereerd tussen herkomst en de eindbestemming van de goederenstromen. Deze ketens lopen via één of twee terminals. Alle gebruikte terminals moeten in het gebied van de herkomst of bestemming liggen, of in een aangrenzend gebied. Ook wordt er van uit gegaan dat er direct wegvervoer tussen de herkomst en bestemming mogelijk is.

Dit levert een set aan mogelijke ketens op. De totale goederenstroom wordt dan verdeeld over de mogelijke ketens, in verhouding tot de afstanden/kosten/reistijden op deze ketens. In ons voorbeeld is Rotterdam de herkomst haven, hier wordt geen voor- of natransport verondersteld, en wordt direct met binnenvaart vertrokken. In het huidige model wordt Deurne gezien als bestemming. Dan zijn de terminals in Wanssum, Veghel en Oss mogelijk. De stroom wordt dus verdeeld over deze terminals, zie de rode pijlen in onderstaande figuur. Met het nieuwe bestand is Venlo de eindbestemming, en zijn terminals Wanssum, Venlo en Roermond beschikbaar. Dan wordt de stroom verdeeld over deze terminals, volgens de blauwe pijlen in de figuur.



Figuur 6: Voorbeeld verdeling containerstroom op basis van modelaannames en bestemming Deurne (rode pijlen) vs. verdeling op basis van ketenbestand (blauwe pijlen)

Een andere mogelijkheid van het nieuwe bestand is het toetsen van de aanname over de mogelijke terminals. In het bestand kunnen we zien welke terminals gebruikt worden, en misschien andere aannames doen over mogelijke terminals. Als er aanleiding is om de aanname aan te passen kan het tot een andere routing van de stromen in de prognose leiden, en dus tot andere stromen per route.

Uiteindelijk worden de ketens weer opgeknipt in unimodale links. Per unimodale link wordt dan de groei berekend en een pivot-point methode gebruikt. Hiermee wordt de toekomstige stroom goederen op de links uitgerekend. Hierdoor groeit iedere link met zijn eigen factor (een combinatie van alle stromen die er overheen gaan), waarbij ketens niet meer herleidbaar zijn, en ook niet in z'n geheel opgehoogd worden. In ons voorbeeld wordt de keten Rotterdam-Wanssum-Deurne opgeknipt in links Rotterdam-Wanssum met binnenvaart en Wanssum-Deurne met wegvervoer. Deze links krijgen elk hun eigen groei, waarbij dus alle ketens die een link via Wanssum-Deurne hebben bijdragen aan de groei van die link. Die totale groei wordt dan toegepast op de waarneming op die specifieke link.

Met het nieuwe bestand zouden we eerst de groei per keten kunnen uitrekenen, omdat we nu een waarneming hebben voor de hele keten en zo een pivot-point op dat niveau kunnen toepassen. Daarna kunnen we de keten opsplitsen in unimodale links Rotterdam-Wanssum en Wanssum-Venlo. Zo blijft de samenhang tussen de verschillende links in de keten beter in stand, en is de ophoging van de links dus consistentier.

6.1. Wat betekent dit voor het gebruik van het model?

Uit de vorige paragraaf blijkt dat de nodige aannames gemaakt moeten worden in het model en dat verdeelfactoren geschat worden, Met het nieuwe ketenbestand hoeft minder verdeeld te worden en kunnen aannames beter getoetst worden.

De keuze voor Deurne vs. Venlo lijkt nu een redelijk groot verschil. Maar in werkelijkheid gaan er meerdere ketens van Rotterdam naar Wanssum, en worden deze relatief aan de waargenomen stromen verdeeld over Deurne en Venlo (en alle andere mogelijke bestemmingen). Daardoor is de koppeling niet direct goed, maar worden wel alle links vertegenwoordigd in het totaal. Doordat logische routes en

veelgebruikte verbindingen vaker voorkomen in de data, geeft dit op geaggregeerd niveau een goed beeld van de verkeersstromen. De parameters in de modellen worden zo vastgesteld dat deze daadwerkelijk waargenomen patronen ook worden gerealiseerd in de prognoses.

Hierdoor is het gebruik van het huidige model wel beperkt wat betreft het aggregatieniveau. Inzoomen op specifieke wegvakken, voertuigtypes of goederensoorten is niet realistisch, omdat de parameters zijn bepaald op basis van geaggregeerde waarden of gemiddelden. Zolang hier bij het toepassingsbereik van het model rekening mee gehouden wordt, geeft het model betrouwbare resultaten. Echter, voor veel nieuwe toepassingen, zoals vragen rondom zero-emissie zones, laadpunten, grotere voer-/vaartuigen en onderhoudsplanningen is gedetailleerdere informatie gewenst. Met het gebruik van het nieuwe integrale bestand zou inzoomen tot dat detailniveau wel mogelijk moeten worden.

7. Conclusies en vervolg

In dit paper wordt beschreven hoe een logistieke keten er in praktijk uit ziet en hoe RWS en CBS door middel van extra dataverzameling bij terminals en transporteurs en (statistische) methodieken deze logistieke ketens in beeld brengen.

Bij het construeren van een containerketenbestand zijn we in een vergelijkbare situatie beland als het land in de inleiding van De Waal et al. (2019): Het land deed hard haar best een zo goed mogelijke statistiek te maken op basis van één bron. Hun grootste wens was meer data zodat ze nog meer en betere statistieken konden maken gebaseerd op meerdere bronnen. Een goede fee / boze heks vervulde hun wens en gaf hun meer data dan waar ze van hadden kunnen dromen. Toen begonnen de problemen pas echt ...

Door de combinatie van bronnen kunnen inconsistenties ontstaan als gevolg van definitieverschillen, meetfouten, e.d. De data in de bronnen wordt niet primair verzameld om er statistieken over te maken in tegenstelling tot bijvoorbeeld een enquête. Er treden koppelfouten op en er zijn verschillen in dekking tussen de verschillende bronnen. Het zijn geen eenvoudige kanssteekproeven die makkelijk op te hogen zijn of integrale registers waar je de gewenste variabele alleen hoeft te tellen. Het probleem om goede (totaal)schattingen van de doelvariabelen te maken wordt een stuk lastiger. Dat geldt ook voor het bepalen van de kwaliteit van die schattingen.

7.1. Kwaliteit en schattingsmethoden

Paragraaf 6 geeft een voorbeeld van hoe een *ketenbestand* voor containers de prognoses van BasGoed kan helpen verbeteren. De kwaliteit van die prognoses wordt mede bepaald door de kwaliteit van de inputbestanden en daarom is het belangrijk die kwaliteit ook in beeld te hebben. Veel gebruikte kwaliteitsmaten zijn vertekening en variantie. Als de inputdata selectief is, moet de schattingsmethode daar rekening mee houden om de vertekening zo klein mogelijk te houden. De methode heeft ook invloed op de variantie. Er zijn ook nog andere afwegingen bij het kiezen van een goede schattingsmethode.

De dataverzameling zal naar verwachting nooit leiden tot 100 procent dekking van vervoerde containers. Het bijschatten van missende data kan met verschillende methoden. We hebben nu een vrij simpele imputatiemethode gebruikt, die mogelijk nog verfijnd kan worden. Voor de In-, Uit- en Doorvoerstatistiek wordt een macrointegratie-model gebruikt wat mogelijk ook een optie is voor de containerketens. Daarnaast zitten er ook tijdspatronen in de data waardoor tijdreeksmethoden wellicht interessant kunnen zijn. Het optimum is waarschijnlijk een combinatie van meerdere methoden.

Een ander punt dat meespeelt bij het kiezen van een geschikte schattingsmethode is de robuustheid tegen wisselende beschikbaarheid van data. Het komt regelmatig voor dat een bedrijf wordt overgenomen of dat er wijzigingen zijn in het gebruikte softwarepakket. Dit kan leiden tot een tijdelijke of permanente uitval van de datastroom. De schattingsmethode moet hiervoor kunnen compenseren (tot op zekere hoogte; als een bron als de goedereninformatie van de douane wegvalt dan zijn er andere maatregelen nodig).

7.2. Vervolg

Momenteel lopen er nog diverse trajecten om de dataverzameling verder op te schalen, waardoor de kwaliteit van de logistieke ketens verder verhoogd kan worden. Meer data is echter niet per definitie beter; als een nieuwe bron inconsistent is met een eerdere bron dan rijst de vraag wat de waarheid is. Belangrijk is de bestaande bronnen en eventuele inconsistenties goed te begrijpen en dan gericht te zoeken naar bronnen die de bestaande lacunes kunnen verhelpen als dat niet op een andere manier opgelost kan worden.

Het spoorvervoer is een voorbeeld waarbij aanvullende databronnen belangrijk zijn. Vervoerders zijn vaak actief op specifieke trajecten en daardoor niet makkelijk uitwisselbaar of op te hogen. Momenteel is zo'n 40-50% van de spoorvervoer data beschikbaar voor het afleiden van containerketens. Gesprekken met vier grote spoorvervoerders voor het delen van data lopen op dit moment. Als deze kunnen worden toegevoegd is zo'n 90% van het spoorvervoer bekend en kan de rest bijgeschat worden. Voor de binnenvaart is IVS NEXT data in de basis een goede en stabiele bron. In bepaalde gebieden is de registratie echter minder betrouwbaar. Als je de beperkingen begrijpt, bijvoorbeeld door de data te vergelijken met binnenlandterminal data, kun je de IVS bron zo goed mogelijk corrigeren. Daar waar IVS dan nog steeds onvoldoende informatie levert, kan gericht gezocht worden naar alternatieven.

Door het grote aantal bedrijven actief in het wegvervoer van containers lag hier tot nu toe de focus nog niet op bij de dataverzameling. Voor het first en last mile transport via de weg is echter informatie van binnenlandterminals erg nuttig: met weinig bronnen komt veel informatie beschikbaar. Daarnaast zou de implementatie van OTM erg kunnen helpen bij het delen van data. De strategie zal zijn de grootste wegvervoerders integraal te benaderen. Met de top 50 bedrijven kan 60-70% van de wegvervoerritten in kaart gebracht. Het resterende deel kan wellicht afgeleid worden uit de bekende ketens of er zou een kleine steekproef (< 100 bedrijven) getrokken kunnen worden om dit deel beter in kaart te brengen.

Het is belangrijk dat het ketenbestand ook gedeeld kan worden met RWS en andere stakeholders op een zo (regionaal) gedetailleerd mogelijk niveau. Door het beperkt aantal actieve bedrijven bij containervervoer zijn outputcellen echter vlug onthullend. Er wordt daarom gezocht naar methoden met betrekking tot statistische beveiliging om toch zo veel mogelijk te kunnen publiceren zonder gevoelige informatie te onthullen over individuele bedrijven.

Door de beschikbaarheid van een ketenbestand zijn er ook verbetermogelijkheden voor BasGoed. Nu beschouwt BasGoed het landzijdige verkeer eigenlijk los van het zeezijdige verkeer. Met de doorlopende ketens in het nieuwe bestand moet het mogelijk zijn om ook in het model de hele ketens te modelleren, dus het zeezijdige en landzijdige deel van de ketens niet meer van elkaar te scheiden maar ook als één geheel in het model te beschrijven.

De focus van het in kaart brengen van ketens lag tot nu toe bij containers, omdat een container met inhoud via het containernummer gevolgd kan worden. RWS en CBS breiden dit inmiddels ook uit naar andere verschijningsvormen zoals bulk. Dit traject is inmiddels opgestart met een verkenning bij een grote kolen- en ijzererts overslagterminal.

8. Referenties

De Blois, C. & Jacobs, M., 2018. Nieuwe ketendata, Top-down en Bottom-up, paper gepresenteerd op de Vervoerslogistieke werkdagen 2018.

De Bok, M., De Jong, G., Groot, N., Miete, O., Van den Berg, M., Van Meijeren, J.C. & Benjamins, M., 2016. Ontwikkeling van een multimodaal ketenkeuzemodel voor containertransport in het strategisch goederenvervoermodel BasGoed, paper gepresenteerd op de Vervoerslogistieke werkdagen 2016.

Fransen, R., Van Meijeren, J.C., Tilanus, P., Miete, O.M., Hazelhorst, W.O. & Spruijt, C.J., 2018. ONTWIKKELING MULTIMODALE TRANSPORTKETENDATA, paper gepresenteerd op de vervoerslogistieke werkdagen 2018.

Rijkswaterstaat, Water verkeer en leefomgeving, BasGoed 6 strategisch model voor goederenvervoer, 2021. <www.basgoed.nl>.

De Waal, T., Van Delden, A. & Scholtus, S., 2019. Quality measures for multisource statistics. *Statistical Journal of the IAOS*, 35, pp 179–192. Available at: <http://dx.doi.org/10.3233/SJI-180468>.

Wickham, H., Çetinkaya-Rundel, M. & Golemund, G., 2023. R for Data Science (2nd edition). <<https://r4ds.hadley.nz/data-tidy.html>> .