

RITTYPEN GOEDERENVERVOER NAAR TOEKOMSTIGE LAADBEHOEFTE

J.W.A. van Beuningen, Centraal Bureau voor de Statistiek (Directie Verkeer en Vervoer)

M.J. Jacobs, Centraal Bureau voor de Statistiek (Directie Verkeer en Vervoer)

Samenvatting

De toekomstige (op)laadbehoefte inschatten is van belang voor de elektrificatie van het goederenvervoer vanwege de planning van de aankoop van elektrische vrachtvoertuigen en van de laadinfrastructuur. Om verschillen in laadbehoefte na elektrificatie in kaart te brengen, classificeert deze studie ritten op basis van drie ritkenmerken, namelijk: de afgelegde afstand, het vervoerd gewicht en het aantal deelritten. Er is gebruik gemaakt van gegevens over het goederenvervoer over de weg van het Centraal Bureau voor de Statistiek (CBS). De cluster analyses laten vier typen ritten zien die verschillen op de combinatie van de drie ritkenmerken, namelijk: 1) zwaar beladen ritten met gemiddeld een medium afstand meestal zonder tussenstops en met maximaal één tussenstop, 2) relatief korte lichte ritten meestal zonder tussenstops en met maximaal twee tussenstops, 3) ritten over lange afstand en medium gewicht met relatief veel tussenstops (tussen de vier en zes) en 4) ritten van medium afstand, medium gewicht met meestal enkele tussenstops (tussen de nul en drie). Deze clusters worden vergeleken op achtergrondkenmerken van de rit, het voertuig, de vervoerde goederensoort in het geval van beladen ritten en de meest voorkomende vertrek- en aankomstregio's.

1. Inleiding

Het inschatten van de toekomstige (op)laadbehoefte van het vrachtvervoer over de weg is relevant voor de aanleg van de zware laadinfrastructuur in Nederland. Het Ministerie van Infrastructuur en Waterstaat (IenW) heeft de Nationale Agenda Laadinfrastructuur (NAL) in het leven geroepen om de energietransitie in het verkeer en vervoer te faciliteren en op tijd een goede laadinfrastructuur aan te leggen.

Op dit moment is nog relatief weinig vrachtvervoer geëlektrificeerd, maar de verwachting is dat er snel geëlektrificeerd gaat worden bijvoorbeeld mede vanwege de Zero Emission milieuzones in steden waar voertuigen met uitstoot op termijn niet meer in mogen (Natuur en Milieu, 2020; Planbureau voor de Leefomgeving, 2022). In het Klimaatakkoord is afgesproken dat er steeds meer emissievrije zones in steden zullen komen en dat de logistieke sector gaat verduurzamen.

In Nederland gaat het meeste goederenvervoer over de weg, in 2020 werd 43,0 procent van het vervoerde gewicht met vrachtvoertuigen over de weg vervoerd (CBS Statline, 2021). Voor het binnenlands goederenvervoer was dit zelfs 81,9 procent. In 2020 bestond 1,1 procent van het wagenpark uit vrachtwagens en trekkers voor oplegger (CBS Statline, 2022a), terwijl deze zware bedrijfsvoertuigen in datzelfde jaar voor 25,9 procent van de CO₂-uitstoot van het wegverkeer zorgden (CBS Statline, 2022b). Hierbij moet worden opgemerkt dat in de emissiecijfers ook de buitenlandse voertuigen worden meegeteld, het gaat om emissies op het Nederlands grondgebied. De buitenlandse wegvoertuigen nemen een klein deel van het goederenvervoer in Nederland voor hun rekening. Met de elektrificatie van deze zware bedrijfsvoertuigen kan dus een relatief grote CO₂ emissiereductie behaald worden wat belangrijk is voor het behalen van de klimaatdoelen. De verwachting is dat vanaf 2023 elektrische vrachtauto's in bepaalde marktsegmenten kunnen concurreren tegen de vrachtauto's op diesel (Natuur en Milieu, 2020). Het is op dit moment nog onduidelijk of dit al het geval is.

Omdat er op dit moment nog vrij weinig informatie beschikbaar is over de toekomstige laadbehoefte van vrachtvoertuigen, wordt in deze studie gekeken of er met gegevens over ritten in het goederenvervoer van het Centraal Bureau voor de Statistiek (CBS) groepen ritten kunnen worden onderscheiden op basis van de verwachte laadvraag met behulp van cluster analyse en of deze ritten kunnen worden getypeerd aan de hand van hun achtergrondkenmerken.

2. Conceptuele achtergrond

In deze studie gaat het om de inschatting van het potentiële verbruik tijdens de totale rit; tussentijds bijladen is vanwege de extra benodigde tijd minder efficiënt voor de vervoerder. Aan het eind van de rit tijdens het lossen, tijdens verplichte rusttijden of aan het eind van de dag kunnen vrachtauto's mogelijk bijladen zonder dat dit (veel) extra tijd kost. Het is daarom van belang om inzicht te krijgen in

verschillen in laadbehoefte van type ritten. De verwachting is dat een groot deel van de voertuigen 's nachts op de standplaats of het distributiecentrum zal opladen, omdat dit financieel het meest aantrekkelijk is (Natuur en Milieu, 2020). Daarnaast zal bijladen onderweg of bij de klant tijdens het lossen voorkomen. In het geval van korte ritten van het bedrijf naar de klant is 's nachts entrepotladen op eigen terrein voor de hand liggend. Bij langere ritten of veel kilometers per dag kan tussentijds opladen op een loslocatie of onderweg nodig zijn (Natuur en Milieu, 2020). Onderzoek onder transportbedrijven die zich voornamelijk richten op stadslogistiek liet zien dat 73 procent verwacht hun vrachtauto's te gaan laden op het eigen bedrijf en dat men verwacht dat driekwart van de voertuigen 's nachts laadt (ElaadNL, 2021). De onderzoeksvraag in deze studie is als volgt: *Zijn er in het goederenvervoer met vrachtauto's (incl. trekker voor oplegger) typen ritten te onderscheiden op basis van verschillen in het verwachte energieverbruik tijdens de rit?*

Er zijn verschillende deelvragen:

- Welke ritkenmerken hangen samen met de (toekomstige) laadvraag van een rit? Het gaat om de laadvraag indien het voertuig geëlektrificeerd zou zijn. Wat zijn de belangrijkste kenmerken?
- Zijn er meerdere clusters van ritten te onderscheiden op basis van de geselecteerde ritkenmerken?
- Zijn deze clusters qua profiel informatief en onderscheidend op basis van soort vervoer (beroepsvervoer of eigen vervoer), stroom (binnenlands of internationaal vervoer), vervoerde goederensoorten en andere achtergrondkenmerken?

In de literatuur worden verschillende factoren genoemd die verband houden met het energieverbruik van zware vrachtvoertuigen. Het vervoerd gewicht en de lengte van de rit zijn gerelateerd aan het energieverbruik tijdens de rit; de energie-efficiëntie wordt gedefinieerd als de ratio van transport en energieverbruik, ofwel de tonkilometers per kilowattuur (Liimatainen en Pöllänen, 2010). De tonkilometers bestaan uit het vervoerd gewicht (in ton kilogram) gecombineerd met de afgelegde afstand (in kilometers) tijdens de rit.

Odhams et al. (2010) identificeren nog drie typen factoren die gerelateerd zijn aan het energie- of brandstofverbruik van zware bedrijfsvoertuigen, namelijk: 1) voertuigkenmerken (lengte, massa, volume, motorefficiëntie, rolweerstand, banden, aerodynamica en regeneratief remmen), 2) logistieke factoren (de ruimtelijke structuur van de logistieke keten, routing en voertuiggebruik) en 3) externe factoren (zoals snelheid, hoogteprofiel van de route, drukte op de weg, rijgedrag van de chauffeur en weersomstandigheden). Demir, Bektas en Laporte (2013) noemen ook voertuigkenmerken (waarbij ter aanvulling op voorgenoemde factoren o.a. onderhoud en bouwjaar genoemd worden) en logistieke factoren (waarbij aanvullend lege kilometers en aantal stops worden genoemd) en splitsen de externe factoren op in: omgevingsfactoren, verkeersomstandigheden en chauffeur factoren. Extra kenmerken die de auteurs als voorbeeld in deze categorieën noemen, zijn: het soort wegdek, de temperatuur, wind, optrekken en afremmen, keuze voor een versnelling en de tijd dat men stilstaat tijdens de rit.

Drie ritkenmerken zijn in deze studie gekozen als uitgangspunt voor de clusters, namelijk: de ritafstand, het vervoerd gewicht tijdens de rit en het aantal deelritten. Het aantal deelritten is meegenomen om het effect van extra starten en optrekken tijdens de rit mee te nemen bij het energieverbruik. Alleen ritten waarbij zowel afstand als vervoerd gewicht bekend is, zijn meegenomen in de analyse. Het aantal deelritten is altijd bekend. Er is in dit geval een focus op ritkenmerken die van ritplanning afhangen en niet van variabele omstandigheden als het weer, wegomstandigheden, de chauffeur of van het voertuig aangezien meestal ook voor een ander type voertuig kan worden gekozen. Voertuigkenmerken kunnen wel gebruikt worden om rittypen achteraf te beschrijven (profielen opstellen). Data over weer, wegomstandigheden en rijgedrag is op dit moment ook niet beschikbaar bij het CBS, maar zou met nieuwe databronnen in de toekomst wel beschikbaar kunnen komen.

Het vervoer over de weg in Nederland bestond in 2021 in totaal uit ongeveer 57 miljoen beladen ritten, waarvan 85 procent van de ritten binnen Nederland werd afgelegd, 12 procent was grensoverschrijdend waarbij een deel van de rit in Nederland werd afgelegd en 3 procent was volledig in het buitenland (CBS Statline, 2022c). De 3 procent geheel in het buitenland afgelegde ritten dragen niet bij aan de laadvraag in Nederland. Daar staat tegenover dat in deze cijfers alleen ritten van Nederlandse bedrijfsvoertuigen (met een maximum toegestaan gewicht van meer dan 3,5 ton) zijn meegenomen en niet de buitenlandse voertuigen die in Nederland rijden. Iets meer dan driekwart van de beladen ritten (76 procent) werd uitgevoerd door beroepsvervoerders, de rest was eigen vervoer. Het totale vervoerde ladinggewicht bedroeg in 2021 bijna 702 miljoen ton kilogram. In 2021 had 0,1 procent van de vrachtauto's en trekkers elektriciteit als hoofd- of nevenbrandstof (hybride) (CBS Statline, 2022d).

3. Methode

3.1. Data

Dit onderzoek is gebaseerd op het wegvervoer onderzoek van het CBS. Informatie over het goederenvervoer wordt door middel van een steekproef van bedrijfsvoertuigen uitgevraagd bij bedrijven. De steekproef wordt getrokken uit de registers van de Dienst Wegverkeer (RDW), de Kamer van Koophandel en de Belastingdienst. Jaarlijks worden ritgegevens van ongeveer 25 duizend voertuigen verzameld van in totaal rond de 140 duizend vrachtauto's en trekkers in het motorvoertuigenpark in recente jaren. Meer informatie staat in de onderzoeksbeschrijving (<https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/wegvervoer>).

De ritgegevens worden middels een internetvragenlijst of via XML (elektronisch aftappen van informatie uit het Transport Management Systeem van goederenwegvervoerders) door het CBS verzameld.

Alle ritten van vrachtauto's en trekkers voor oplegger in de respons van 2021 (en in 2019) zijn meegenomen in de analyses. Speciale voertuigen zijn uitgesloten van de studie omdat het hier om een kleine gemêleerde groep voertuigen met andere ritpatronen gaat. Zowel beladen als lege ritten zijn meegenomen. Buitenlandse voertuigen die in Nederland rijden, zijn niet meegenomen in deze analyses. In totaal zijn 409.753 ritten uit 2021 meegenomen in de analyses. Daarnaast is data uit 2019 gebruikt voor replicatie van de cluster analyses met wegvervoer gegevens van voor Covid-19.

3.2. Variabelen

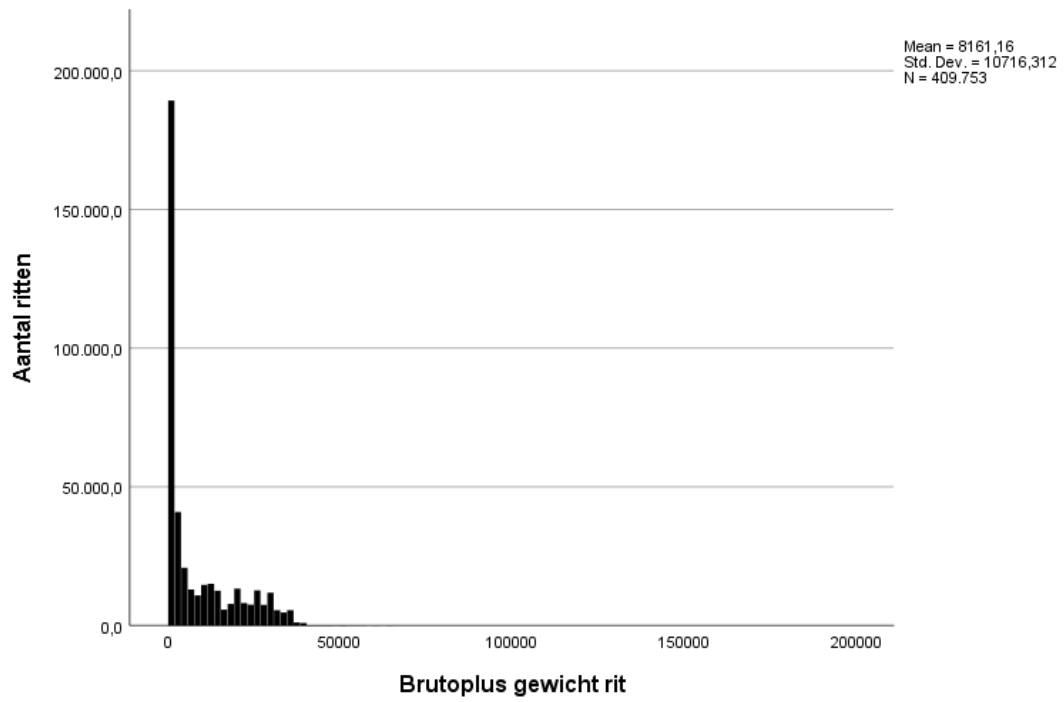
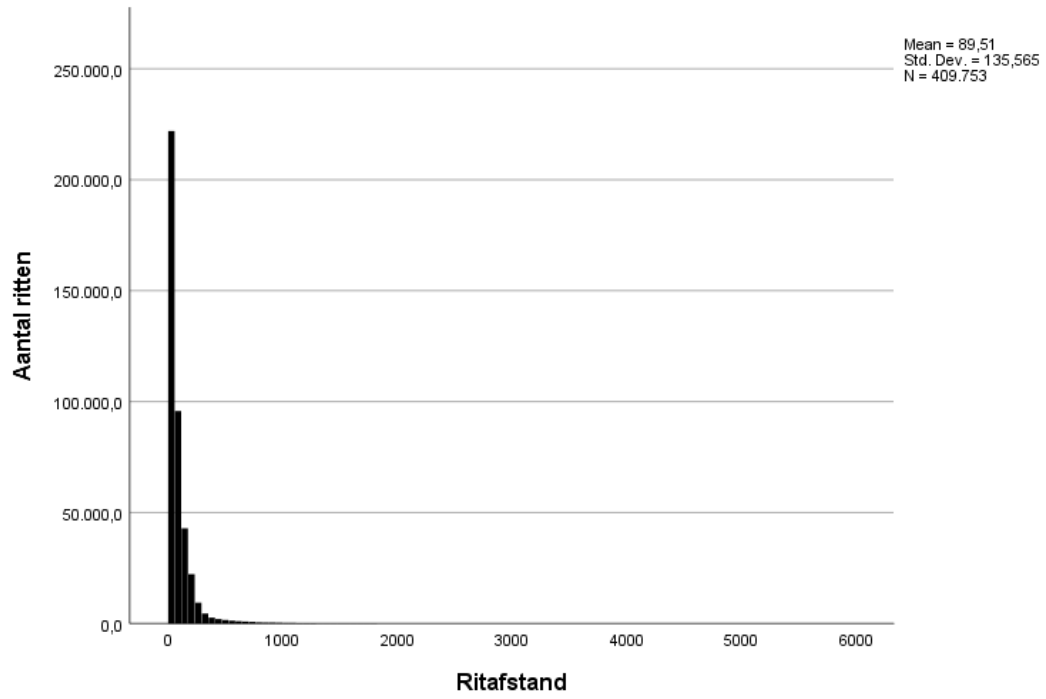
Ritafstand: de werkelijk gereden kilometers per rit, exclusief de kilometers die het motorvoertuig of de combinatie aflegt per trein, ferry, via binnenvaart of ander transport.

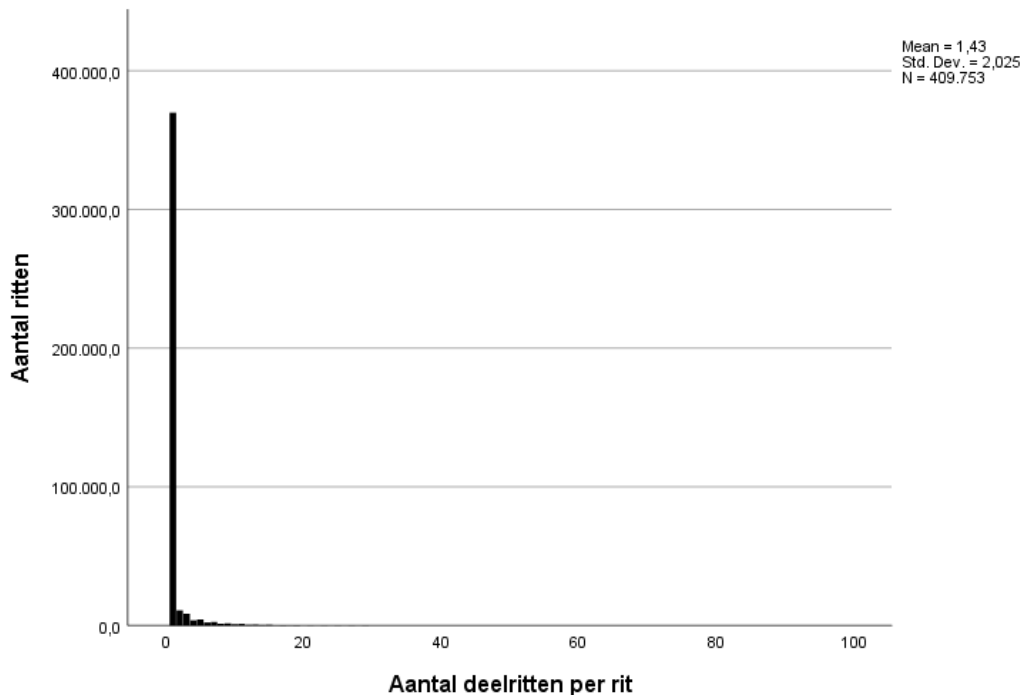
Vervoerd brutoplus gewicht: het brutogewicht van goederen vervoerd tijdens een rit inclusief verpakkingsmateriaal en transportmiddel uitgedrukt in kilogram.

Aantal deelritten: het aantal deelritten waaruit de rit bestaat, deelritten beschrijven de stukjes rit tussen laad- en loslocaties van de vervoerde zendingen. Er bestaan distributieritten, waarbij de zendingen op een laadlocatie worden ingeladen om vervolgens op verschillende locaties te worden uitgeladen, ieder stukje rit tussen het laden en iedere loslocatie is een aparte deelrit. Bij groupageritten is de loslocatie van alle zendingen hetzelfde, maar worden zendingen op verschillende plekken ingeladen. Ieder stukje rit tussen de laadlocaties en naar de loslocatie is een aparte deelrit. Verder bestaan er ritten die slechts een laad- en loslocatie hebben en dus ook één deelrit. En tot slot zijn er overige ritten die verschillende laad- en loslocaties combineren. Elk stukje rit waarbij gestart en gestopt wordt, is dan een aparte deelrit. Het aantal deelritten correspondeert met het aantal tussenstops tijdens de rit; bij een deelrit zijn er onderweg geen tussenstops geweest, bij twee deelritten is er één tussenstop geweest etc.

Alle drie de inputvariabelen zijn scheef verdeeld (skewness varieert van 1,3 voor vervoerd gewicht, 6,1 voor ritafstand en 9,9 voor aantal deelritten en kurtosis van 2,1 voor vervoerd gewicht, 69,7 voor ritafstand en 170,5 voor aantal deelritten) en niet sterk gecorreleerd (maximaal *Pearson r* = 0,20).

De helft van de ritten is maximaal 50 kilometer en de gemiddelde ritafstand in 2021 was 89,5 kilometer voor Nederlandse voertuigen (SD = 135,6). Daarnaast bestond 36 procent van alle ritten uit lege ritten, waarbij 0 kilogram gewicht werd vervoerd, het gemiddeld gewicht was 8.161 kilogram (SD = 10.716). En 90 procent van de ritten bestond uit één deelrit ofwel nul tussenstops met een gemiddelde van 1,4 deelritten (SD = 2,0). Figuur 1 laat de distributies met de drie inputvariabelen zien.





Figuur 1. Distributies van ritafstand, brutoplus vervoerd gewicht en aantal deelritten, 2021

3.2.1. Achtergrondkenmerken

Ladingtonkilometer: ladingtonkilometer is samengesteld uit de afgelegde ritafstand vermenigvuldigd met het vervoerde brutoplus gewicht.

Massa leeg voertuig: ledige massa van het voertuig.

Voertuigsoort: het soort voertuig waarmee de rit is uitgevoerd, in dit geval vrachtauto dan wel trekker voor oplegger.

Zendingen: het aantal zendingen tijdens de rit. Een zending bestaat uit een hoeveelheid goederen van één zelfde soort, die in één keer van één laadplaats naar één losplaats wordt vervoerd op één vervoermiddel. Als een zending over verschillende trajecten door verschillende voertuigen wordt vervoerd, dan moet dit gezien worden als een aparte zending per voertuig.

Type vervoer: eigen vervoer of beroepsvervoer.

Stroom rit: binnenlands vervoer met laad- en loslocatie in Nederland, grensoverschrijdend vervoer met laad- of loslocatie in Nederland (ook bilateraal vervoer genoemd) of ritten waarbij laad- en loslocatie in een ander land liggen (cabotage- of derdelandenvervoer). Deze laatste categorie ritten kunnen geheel in het buitenland plaatsvinden of deels door Nederland gaan (doorvoer).

Gecontaineriseerd vervoer: goederen worden wel of niet in een container vervoerd.

Euronorm klasse: emissiecode waarbij een hoger cijfer een lagere uitstoot inhoudt; de code voor de gemeten uitstoot van uitlaatgassen van een voertuig, mate van emissie uitgedrukt in euronormen. Codes variëren van 1 tot en met 6, hoewel code 1 in de praktijk niet meer voorkomt bij de vrachtvoertuigen in de wegvervoer data. Daarnaast zijn er enkele ZE voertuigen (Zero Emission), die bij de analyse worden meegerekend bij categorie 6.

Geconditioneerd voertuig: geconditioneerd voertuig met of zonder temperatuurregeling waarmee goederen met een constante temperatuur kunnen worden vervoerd. Ritten met koel- of vrieswagens kwamen niet voor in de clusters.

CO₂-uitstoot: totale CO₂-emissie in gram per kilometer zonder uitsplitsing naar type weg, exclusief CO₂ uit ureum. Dit wordt per voertuig berekend. Dit gegeven komt uit de Emissieregistratie en is door middel van een door TNO ontwikkelde methode berekend. Het gaat om voorlopige cijfers.

Dagkilometrage: gemiddeld aantal afgelegde kilometers per dag tijdens het goederenvervoer in de observatieperiode. Dit is inclusief lege ritten en exclusief ritten om te tanken of voor onderhoudsbeurten. Het dagkilometrage wordt per voertuig berekend.

Goederensoort: goederensoort die vervoerd wordt in deelritten volgens de NST 2007 indeling. De indeling in 20 goederengroepen is afgeleid van de 'Nomenclature uniforme des marchandises pour les Statistiques de Transport, 2007' (NST 2007), een uniforme goederennomenclatuur voor de vervoersstatistiek. De NST-2007-codering is vanaf het referentiejaar 2008 van kracht binnen de lidstaten van de Europese Unie.

Herkomst/ bestemming: startpunt en eindpunt van de rit in de vorm van NUTS3 regio of COROP gebied.

De achtergrondkenmerken met betrekking tot het voertuigen worden op voertuigniveau geanalyseerd, aangezien een voertuig meerdere ritten kan hebben gereden. Het gaat om 21.028 voertuigen die in een cluster vallen. Als voertuigen met hun ritten in meerdere clusters vallen, is de clustertoewijzing van de eerste rit als uitgangspunt genomen.

3.3. Aanpak cluster analyses

Cluster analyse is toegepast om verschillende typen ritten te kunnen onderscheiden. Dit is een exploratieve analysetechniek waarbij op verschillende manieren groepen in de data kunnen worden opgespoord. De resultaten kunnen gebruikt worden om een algemeen beeld te vormen van de soorten ritten die verschillen op laadvraag ofwel verwachte energieverbruik tijdens de rit. Precieze percentages zijn in dit geval minder informatief, omdat bij gebruik van een andere clustering methode deze percentages enigszins verschuiven door een iets andere samenstelling van de clusters.

Er is een aantal voorbereidingen op de dataset uitgevoerd, namelijk: de selectie van de inputvariabelen voor het maken van de clusters, het uitsluiten van extreme waarnemingen ofwel 'uitbijters' en het standaardiseren van de drie inputvariabelen. Daarna is het optimale aantal clusters bepaald. Hierbij zijn twee steekproeven getrokken uit de data (met ieder 7 000 ritten) en zijn de resultaten per steekproef op basis van verschillende afstandsmaten en methoden vergeleken. Op basis van 30 indices (ofwel evaluatiematen) is vervolgens het optimale aantal clusters gekozen. Hierbij is de uitkomst van de meerderheid van deze indices gebruikt. De indices gaven aan dat er twee, drie, vier of acht clusters in de data voorkomen, al naar gelang de gekozen afstandsmaten (voor bepalen van verschillen tussen ritten; kortste afstand tussen iedere twee punten in een rechte lijn of de afstand tussen iedere twee punten volgens de assen van de inputvariabelen) en methode (voor afbakening van de clusters; het minimaliseren van variantie binnen clusters of de gemiddelde afstand tussen ritten van verschillende clusters wordt gemaximaliseerd). Door meerdere steekproeven, afstandsmaten en afbakeningsmethoden te gebruiken, wordt het risico op een lokaal minimum bij het vaststellen van het optimaal aantal clusters verkleind.

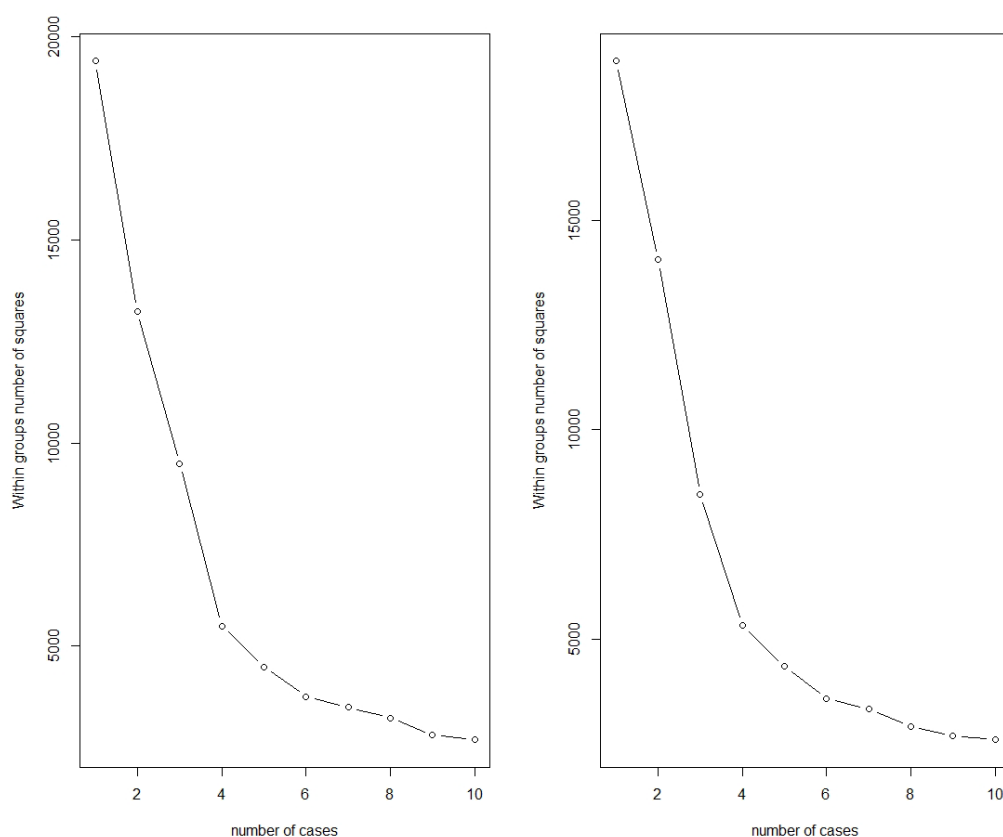
Tot slot zijn de clusters geschat en beschreven op basis van de achtergrondkenmerken van de ritten die bij ieder cluster horen. De schatting van de clusters is gebaseerd op alle ritten (met uitzondering van de uitbijters). Op basis van 2-staps clustering kwamen hierbij de 2- en 4-cluster oplossingen naar voren. Zowel de 2-, 3- 4- als 8-cluster oplossing zijn met k-means clustering geschat om de verschillen te bekijken (zie Bijlage tabel 2). Bij deze methode wordt de variantie binnen clusters geminimaliseerd. Als dezelfde analyses op data van ritten uit 2019 worden gedaan, dan blijken er in dit jaar 2 of 4 clusters ritten te bestaan. Deze clusters lijken qua grootte en samenstelling op basis van de inputvariabelen op de clusters uit 2021; dezelfde rittypen kunnen worden onderscheiden al verschuiven de gemiddelden op de inputvariabelen enigszins. Gedetailleerdere informatie over de cluster analyses is te vinden in de Bijlage.

4. Resultaten

4.1. Aantal clusters

Uit de exploratieve cluster analyses blijkt dat er op basis van de inputvariabelen 2 tot en met 4 clusters of 8 clusters ritten zijn in 2021 (zie de Bijlage voor details). Bij inspectie van de grootte van de clusters en de clustercentra (gemiddelde waarden van de inputvariabelen) blijkt dat er bij iedere cluster oplossing een klein cluster met beperkt aantal ritten te zijn. Bij de 8-cluster oplossing zijn er vier clusters met maximaal 2 procent van de ritten. In het geval van de 2-cluster oplossing op basis van k-means clustering is er een groot cluster met bijna alle ritten (96 procent) en een zeer klein cluster met de overige ritten (4 procent). Deze verdeling is weinig onderscheidend al zien we hier wel een kleine groep ritten met een relatief hoge laadbehoefte. Bij de 3-cluster oplossing is er ook een klein cluster met 4

procent van de ritten en wordt onderscheid gemaakt tussen een groot cluster met relatief 'lichte' ritten qua laadbehoefte (relatief korter, lichter en met weinig deelritten ten opzichte van de overige ritten), het hele kleine cluster met relatief 'zware' ritten met name door de vele deelritten en een medium groot cluster met relatief 'zware' ritten met name door het hoge vervoerd gewicht en weinig deelritten. De 4-cluster oplossing maakt hetzelfde inhoudelijke onderscheid tussen clusters als de 3-cluster oplossing, maar splitst de ritten met veel tussenstops in twee clusters in plaats van een. Dit is echter in totaal wel een grotere groep ritten dan bij de 3-cluster oplossing. Figuur 2 geeft het aantal clusters weer op basis van k-means clustering van de twee steekproeven, de 'knik' in dit elbow plot geeft het optimale aantal clusters weer, in dit geval 4. Bij meer dan 4 clusters voegen de additionele clusters qua verlaging van de intra-groep cluster variantie relatief weinig meer toe. Bij de 4-cluster oplossing is de 'within cluster sum of squares by cluster' 72 procent voor zowel steekproef 1 als 2. Deze waarde geeft aan hoe ver de waarnemingen per cluster van hun cluster centrum liggen, hoe lager deze waarde hoe compacter de clusters waardoor deze gemakkelijker van elkaar te onderscheiden zijn.



Figuur 2. Elbow plot met aantal clusters voor steekproef 1 (links) en steekproef 2 (rechts), 2021

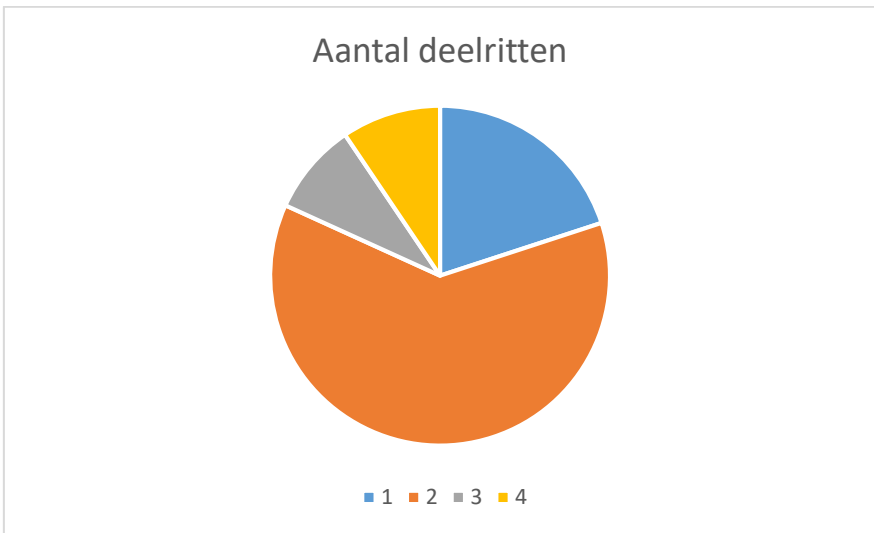
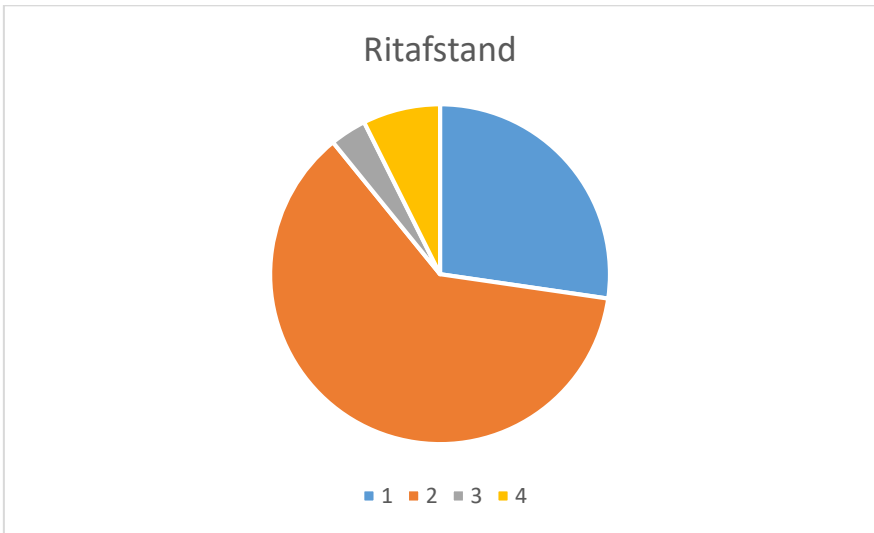
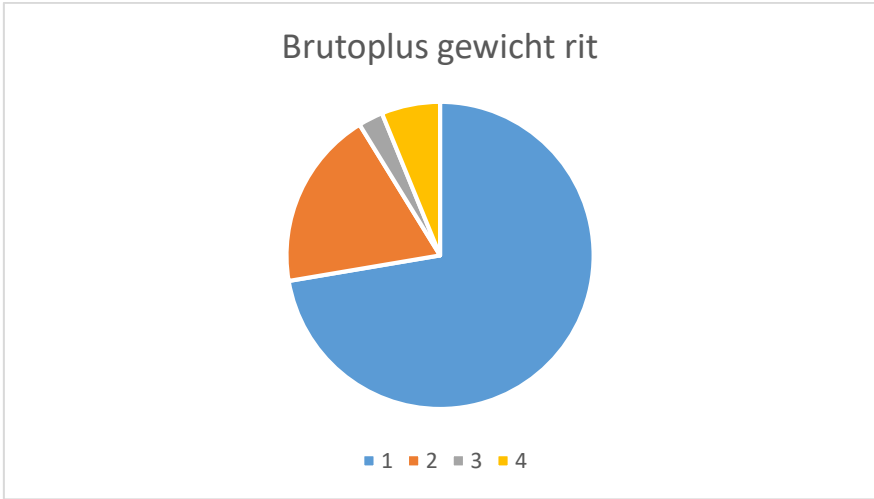
4.2. Beschrijving clusters

De meeste indices besproken in de voorgaande paragraaf geven de 4-cluster oplossing als het optimale aantal clusters. Daarnaast komt dit aantal zowel in 2019 als in 2021 als optimaal aantal clusters terug. De resultaten op basis van de 4-cluster oplossing worden daarom besproken in deze paragraaf. De ritten zijn verdeeld in een groot cluster 72 procent van de ritten, een kleiner cluster met 23 procent van de ritten en twee kleine clusters met de overige ritten van respectievelijk 4 en 2 procent van de ritten. Tabel 1 geeft een overzicht van de grootte van de clusters en de clustercentra (gemiddelde waarden van de inputvariabelen). Het grootste cluster (cluster 2) bestaat uit relatief korte ritten met een laag vervoerd gewicht zonder tussenstops. Dit is inclusief lege ritten. Het kleinste cluster (cluster 3) bestaat uit relatief lange ritten met een medium vervoerd gewicht dat tussen het cluster met gemiddeld het hoogste en gemiddeld het laagste vervoerd gewicht in zit en deze ritten bestaan uit veel deelritten en dus veel tussenstops. Het een na kleinste cluster (cluster 4) lijkt op het kleinste cluster qua afstand en vervoerd gewicht, maar heeft minder deelritten al worden gemiddeld wel enkele tussenstops gemaakt. Het laatste cluster (cluster 1) bestaat gemiddeld uit medium lange ritten met een hoog vervoerd gewicht zonder tussenstops.

Tabel 1. Overzicht van clusters ritten met aantal ritten en gemiddelden op de inputvariabelen, 2021

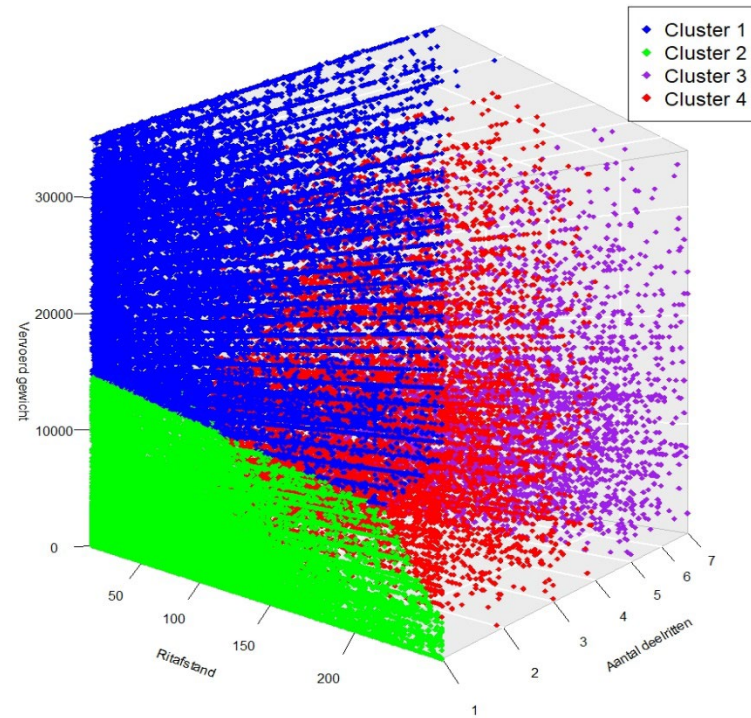
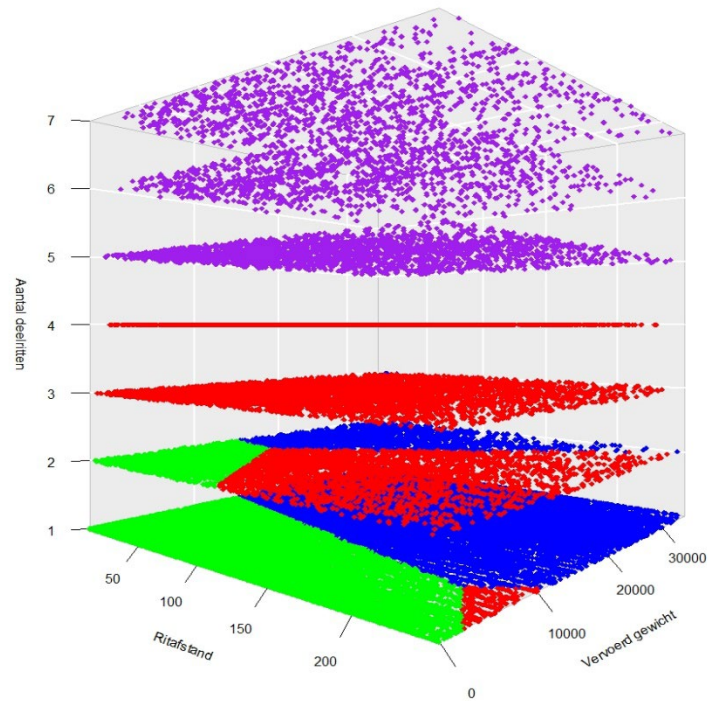
Clusternummer	Aantal ritten per cluster (%)	Gemiddelde ritafstand (SD)	Gemiddeld vervoerd gewicht (SD)	Gemiddeld aantal deelritten (SD)
1	85.403 (23)	75 (63)	23.518 (6.497)	1 (0,2)
2	267.506 (72)	54 (49)	1.960 (3.212)	1 (0,1)
3	6.639 (2)	123 (61)	10.893 (7.931)	6 (0,8)
4	14.584 (4)	119 (62)	11.766 (7.790)	3 (0,8)

Naast het percentage ritten dat in ieder cluster valt, kan gekeken worden naar het aandeel van ieder cluster in het totaal vervoerd gewicht, de afgelegde afstanden en de deelritten. Bij de deelritten worden in dit geval ook de lege ritten meegenomen. Uit Figuur 3 blijkt dat het grootste deel van het vervoerd gewicht in cluster 1 wordt vervoerd, terwijl de meeste ritafstand net als de meeste deelritten in totaal in cluster 2, qua ritten veruit het grootste cluster, wordt afgelegd.



Figuur 3. Aandeel van totaal per inputvariabele naar cluster, 2021

Figuur 4 laat het onderscheid tussen de vier clusters op basis van de inputvariabelen zien vanuit twee perspectieven waarbij de getoonde inputvariabelen op de assen variëren. Het derde cluster (paars) onderscheidt zich op basis van het aantal deelritten van de andere drie clusters, het eerste (blauw) en tweede cluster (groen) onderscheiden zich met name op basis van vervoerd gewicht van elkaar, respectievelijk zware versus lichte ritten, en hebben beiden weinig deelritten. Het vierde rode cluster heeft qua aantal deelritten de meeste spreiding maar niet het hoogste aantal deelritten; van een klein deel ritten met 1 deelrit tot en met 4 deelritten. Het heeft daarnaast net als het derde paarse cluster vrijwel geen korte ritten.



Figuur 4. Clusters met inputvariabelen, 2021

4.3. Achtergrondkenmerken

De cijfers in deze paragraaf zijn indicatief; vanwege het exploratieve karakter van cluster analyse kunnen clusters iets anders afgebakend worden als gebruik wordt gemaakt van een andere methode waardoor achtergrondkenmerken kunnen veranderen. Gemiddeld worden de meeste tonkilometers per rit in ritten van cluster 1 afgelegd. Daarnaast valt op dat de massa van het lege voertuig (bij de trekker wordt alleen de trekker en niet de oplegger meegenomen) hoger is bij de ritten in cluster 1. Dit komt overeen met het gemiddeld hoge gewicht dat tijdens ritten van cluster 1 vervoerd wordt. De medium lange zware ritten in cluster 1 worden relatief vaak met eigen vervoer uitgevoerd (laag aandeel beroepsvervoer) en het betreft relatief vaak grensoverschrijdend vervoer of ritten afgelegd in het buitenland (laag aandeel binnenlands vervoer). Dit cluster bevat het hoogste aandeel ritten met voertuigen met Euronorm 6, de laagste uitstoot. Dat aandeel ritten met Euronorm 6 voertuigen is het laagst in cluster 2 met de korte, lichte ritten. Uit de emissie en kilometrage gegevens komt naar voren dat de voertuigen in cluster 1 gemiddeld de hoogste CO₂-uitstoot per kilometer en het hoogste afgelegde kilometrage per dag hebben. Deze emissies en dagkilometrages per voertuig zijn het laagst in cluster 3. Tot slot komt geconditioneerd transport relatief het vaakst voor in het kleinste cluster 3 waarbij veel tussenstops worden gemaakt. Tabel 2 geeft per cluster informatie over de achtergrondkenmerken van ritten.

Tabel 2. Beschrijving van clusters aan de hand van achtergrondkenmerken van ritten, 2021

Achtergrondkenmerken	1	2	3	4
	(medium aantal ritten, medium afstand, zwaar)	(veel ritten, kort, licht)	(weinig ritten, lange afstand en medium gewicht, heel veel tussenstops)	(weinig ritten, medium afstand, medium gewicht, veel tussenstops)
Tonkilometers (gem. per rit)	1.660.769	98.201	944.625	1.121.526
Massa leeg voertuig (gem.)	9.327	8.640	8.390	8.152
Aandeel trekkers (%)	70	8 3	67	48
Aantal zendingen (gem.)	1	0,5	5	3
Aantal deelritten (min.-max.)	1-3	1-2	5-7	1-4
Aandeel 1 deelrit (%)	98	99	0	3
Aandeel beroepsvervoer (%)	16	22	32	24
Aandeel binnenlands vervoer (%)	83	91	93	88
Aandeel containervervoer (%)	9	6	0	1
Aandeel Euronorm 6 (%)	83	74	79	78

CO ₂ -uitstoot per vrachtauto (gram per km)	913	681	592	625
CO ₂ -uitstoot per trekker (gram per km)	901	843	822	826
Dagkilometrage voertuig (gem.)	385	342	267	352
Aandeel lege ritten (0 kilo vervoerd gewicht) (%)	0	54	0	0
Aandeel transport met geconditioneerd voertuig (%)	1	5	12	7

4.4. Vervoerde goederensoorten per cluster

Als we kijken naar de vervoerde goederen tijdens deelritten, dan moet rekening gehouden worden met het feit dat vooral bij cluster 3 en in mindere mate in cluster 4 ritten met veel deelritten voorkomen met mogelijk verschillende goederensoorten. De ritten in het eerste cluster bestaan in totaal uit 30 procent van alle deelritten met goederen terwijl het gaat om 23 procent van het totaal aantal ritten, voor het tweede cluster geldt dat het gaat om 43 procent van de deelritten met goederenvervoer en 72 procent van de ritten. Deze verschillen worden verklaard door het feit dat in de eerste twee clusters alleen ritten met weinig deelritten voorkomen en daarnaast in het tweede cluster lege ritten zitten. De 2 procent ritten uit het derde cluster nemen 13 procent van alle deelritten met goederen voor hun rekening en de 4 procent ritten uit het vierde cluster de overige 14 procent deelritten.

In cluster 1 worden relatief vaak goederen met NST 3 (metaalertsen en andere delfstoffen), 19 (niet-identificeerbare goederen), 9 (overige niet-metaalhoudende minerale producten) en 8 (chemische producten en synthetische of kunstmatige vezels) vervoerd. Bij de korte lichte ritten uit cluster 2 gaat het relatief vaak om goederen met NST 16 (uitrusting en materiaal voor het vervoer van goederen, dit is onder andere emballage, lege pallets of verpakkingsmaterialen), 17 (Vervoerde goederen in het kader van particuliere of bedrijfsverhuizingen), of 15 (brieven en pakketten). Terwijl bij de ritten uit cluster 3 met veel tussenstops relatief vaak NST 13 (meubelen en overige industrieproducten), 5 (textiel en textielproducten zoals kleding inclusief leder en lederwaren) en 7 (cokes en geraffineerde aardolieproducten) voorkomen. In cluster 4 tot slot komen vaak NST 4 (voedings- en genotmiddelen), NST 13 en NST 7 voor. NST 13 en 7 komen zowel in cluster 3 als cluster 4 relatief vaak voor. Het gaat hierbij om het aandeel deelritten in het totale aandeel bij de goederensoort waarbij het percentage wordt afgezet tegen het totaal aandeel van het cluster in de deelritten. Omdat cluster 3 en 4 klein zijn in het totale goederenvervoer over de weg, komt er altijd een laag percentage van de deelritten per goederensoort in deze clusters voor. Merk op dat het bij post- en pakketvervoer (NST 15 dat relatief vaak bij ritten uit cluster 2 wordt vervoerd) gaat om het vervoer door vrachtauto's en trekkers en niet

om bestelauto's, bijvoorbeeld het vervoer naar distributiecentra vanwaar bezorging aan huis vaak met bestelauto's wordt gedaan.

Tabel 3 geeft per cluster een overzicht van het percentage deelritten met een goederensoort, hierbij wordt uitgegaan van de goederensoort met het hoogste gewicht per deelrit aangezien meerdere goederensoorten tegelijkertijd vervoerd kunnen worden.

Tabel 3. Aandeel deelritten per cluster naar goederensoort (NST), 2021

Goederensoort (NST)	Clusternummer			
	1 (medium afstand, zwaar)	2 (kort, licht)	3 (lange afstand en medium gewicht, heel veel tussenstops)	4 (medium afstand, medium gewicht, veel tussenstops)
Totaal	30	43	13	14
01 Producten van de landbouw, jacht en bosbouw; vis en andere visserijproducten	35	36	12	16
02 Steenkool en bruinkool; ruwe aardolie en aardgas
03 Metaalertsen en andere delfstoffen; turf; uranium en thorium	82	14	2	2
04 Voedings- en genotmiddelen	32	24	20	24
05 Textiel en textielproducten; leder en lederwaren	10	39	34	18
06 Hout, hout- en kurkwaren (m.u.v. meubelen); vlecht- en mandenmakerswerk; pulp, papier en papierwaren; drukwerk en opgenomen media	25	34	21	20
07 Cokes en geraffineerde aardolieproducten	40	13	26	21
08 Chemische producten en synthetische of kunstmatige vezels; producten van rubber of kunststof; splijt- en kweekstoffen	49	21	15	15
09 Overige niet-metaalhoudende minerale producten	51	21	15	13
10 Metalen in primaire vorm; producten van metaal, andere dan machines en apparaten	30	37	18	15

Goederensoort (NST)	1 (medium afstand, zwaar)	2 (kort, licht)	3 (lange afstand en medium gewicht, heel veel tussenstops)	4 (medium afstand, medium gewicht, veel tussenstops)
Totaal	30	43	13	14
11 Machines, apparaten en werktuigen, n.e.g.; kantoormachines en computers; elektrische machines en apparaten, n.e.g.; radio-, televisie- en telecommunicatieapparatuur; medische apparatuur en instrumenten, precisie- en optische instrumenten, uurwerken	21	50	13	16
12 Transportmiddelen	17	52	16	15
13 Meubelen; overige industrieproducten, n.e.g.	7	36	35	22
14 Secundaire grondstoffen; gemeentelijk afval en overig afval	37	44	12	7
15 Brieven, pakketten	13	73	4	9
16 Uitrusting en materiaal voor het vervoer van goederen	4	91	2	4
17 Vervoerde goederen in het kader van particuliere of bedrijfsverhuizingen; door reizigers begeleide bagage en artikelen; voor reparatiedoeleinden vervoerde voertuigen; overige niet voor de markt bestemde goederen, n.e.g.	3	82	9	6
18 Gegroepeerde goederen: diverse soorten goederen die gezamenlijk worden vervoerd	27	33	19	20
19 Niet identificeerbare goederen: goederen die om de een of andere reden niet te identificeren zijn en daarom ook niet in de groepen 01 tot en met 16 kunnen worden opgenomen	64	27	3	6
20 Overige goederen, n.e.g.	0	0	0	0

Noot. Een '1' geeft aan dat er te weinig ritten zijn om betrouwbare, niet-onthullende informatie te kunnen publiceren.

Noot 2. Door afronding tellen rijpercentages soms niet op tot 100 procent.

4.5. Herkomst en bestemmingen per cluster

Tot slot is gekeken naar de herkomst en bestemmingen van de ritten per cluster. In cluster 1 heeft 90 procent van de ritten een vertrekplaats in Nederland en ook 90 procent een Nederlandse aankomstplaats, in cluster 2 is dat bij zowel vertrek- als aankomstplaats 94 procent, in cluster 3 97 en 94 procent en in cluster 4 94 en 92 procent. De meest voorkomende vertrek- en aankomstregio's in Nederland per cluster staan in tabel 4. Daarbij komen een paar regio's consequent terug die in het algemeen het vaakst herkomst- of bestemmingsregio van goederenvervoer zijn; Groot-Rijnmond met de Rotterdamse Haven, Utrecht, Groot-Amsterdam met de Amsterdamse haven en gebieden in Brabant. Er bestaan echter hierin wel nuanceverschillen binnen de clusters; Groot-Rijnmond is relatief minder belangrijk voor de ritten in cluster 3 en 4. Utrecht komt voor in de top-3 herkomst en bestemmingen van clusters 2, 3, en 4, maar niet in cluster 1. Regio's in Noord-Brabant komen niet voor als herkomst of bestemming in cluster 2. Groot-Amsterdam komt voor in de top-3 bestemmingen in cluster 3 en 4, maar niet in de top-3 herkomstregio's. Regio's in Noord-Brabant komen daarentegen voor in de top-3 herkomstregio's van cluster 3 en 4, maar niet in de top-3 bestemmingen. Het kan bij dit laatste bijvoorbeeld gaan om export waarbij goederen vanuit Brabant via zeeschip worden uitgevoerd.

Tabel 4. Meest voorkomende vertrek- en aankomstregio's in Nederland per cluster, 2021

	Clusternummer			
	1 (medium afstand, zwaar)	2 (kort, licht)	3 (lange afstand en medium gewicht, heel veel tussenstops)	4 (medium afstand, medium gewicht, veel tussenstops)
Meest voorkomende vertrekregio (% ritten in cluster)				
1.	Groot-Rijnmond (14)	Groot-Rijnmond (10)	Utrecht (8)	Groot-Rijnmond (7)
2.	West-Noord- Brabant (6)	Groot- Amsterdam (7)	Groot- Rijnmond (7)	Utrecht (7)
3.	Noordoost- Noord-Brabant (6)	Utrecht (6)	Zuidoost- Noord-Brabant (6)	Midden-Noord- Brabant (6)

**Meest voorkomende
aankomstregio
(% ritten in cluster)**

1.	Groot-Rijnmond (12)	Groot-Rijnmond (11)	Groot- Rijnmond (8)	Groot-Rijnmond (7)
2.	West-Noord- Brabant (6)	Groot- Amsterdam (7)	Groot- Amsterdam (7)	Utrecht (7)
3.	Groot- Amsterdam (5)	Utrecht (6)	Utrecht (6)	Groot- Amsterdam (6)

5. Conclusies

Cluster analyse is een exploratieve methode; bij een keuze voor een andere afstandsmaat of toewijzing van ritten aan clusters kunnen iets andere clusterafbakeningingen ontstaan omdat ritten aan een ander cluster worden toegewezen waardoor clustercentra op basis van de inputvariabelen en achtergrondkenmerken van clusters enigszins kunnen verschuiven. Daarnaast ligt de focus hier op de bulk reguliere ritten en worden extreme ritten buiten beschouwing gelaten (de zogenaamde uitbijters). Deze kunnen echter wel interessant zijn om extremen in laadbehoefte in kaart te brengen.

Op basis van de geselecteerde ritkenmerken kan geconcludeerd worden dat er verschillende typen ritten te onderscheiden zijn en dat deze ritten een ander verwacht energieverbruik hebben na elektrificatie van het voertuig. Er worden vier typen ritten onderscheiden, waarbij lichte korte ritten naar verwachting de laagste laadbehoefte hebben. In dit cluster vallen ook lege ritten en relatief veel emballageritten. Voertuigen die deze ritten uitvoeren, zullen mogelijk alleen op het eind van de dag op een depot hoeven te laden. Hierbij moet opgemerkt worden dat lege ritten meestal zijn gekoppeld aan beladen ritten die soms in een ander rittetype cluster vallen en dat deze lege ritten niet los gezien kunnen worden van de bijbehorende beladen ritten. Een voertuig rijdt bijvoorbeeld leeg terug naar de laadlocatie als de goederen gelost zijn. Hierdoor is de totale laadbehoefte waarschijnlijk hoger dan het cluster met lichte korte ritten suggereert, aangezien het desbetreffende voertuig per dag verschillende typen ritten uit kan voeren.

Bij het kleinste cluster ritten met heel veel tussenstops komt vaker beroepsvervoer voor en dit is ook relatief vaker binnenlands vervoer. Bij dit kleinste cluster vinden we voertuigen met gemiddeld de laagste CO₂-uitstoot per kilometer en het laagste dagkilometrage ook al gaat het in dit cluster om gemiddeld lange afstanden per rit. Voor ritten in dit cluster is er na elektrificatie de mogelijkheid tot tussentijds bijladen bij de klant tijdens het lossen op een tussenstop of, indien nodig, middels een snellader onderweg.

Daarnaast kan het type rit per vervoerde goederensoort verschillen; textiel en meubels worden bijvoorbeeld relatief vaak in ritten met veel tussenstops vervoerd. Bij brieven en pakketten vervoerd met vrachtauto's, emballage en goederen voor verhuizingen komen over het algemeen gelijksoortige relatief korte ritten voor met weinig of geen tussenstops. Bij de medium lange zware ritten is dit relatief vaak het vervoer van producten voor de bouw. Bij dit soort vervoer wordt wel al relatief vaak een voertuig met lage uitstoot ingezet. De vervoerde goederen of sector waarin het bedrijf werkzaam is, bepaalt vaak het type rit dat nodig is. Planningsmogelijkheden om rekening te houden met laadbehoefte, zijn daarom vaak beperkt en de laadstrategie aanpassen aan het ritpatroon zal vaak noodzakelijk zijn.

Bij vergelijking met de ritten uit 2019 valt op dat er ook vier rittypen van ongeveer dezelfde grootte als in 2021 worden gevonden, maar dat de samenstelling van de clusters iets anders is; in 2019 was er tussen cluster 1, 3 en 4 weinig verschil qua vervoerd gewicht terwijl in 2021 cluster 1 het hoogste vervoerd gewicht had. Daarnaast had cluster 1 met een kwart van de ritten gemiddeld de langste afstand per rit, terwijl dit in 2021 clusters 3 en 4 met veel deelritten zijn. Het veranderde consumptiegedrag vanwege lockdowns tijdens de coronacrisis zou een verklaring hiervoor kunnen zijn, aangezien in clusters 3 en 4 in 2021 relatief vaak ritten met textiel, meubels en voedings- en genotmiddelen zaten, maar dat is niet met zekerheid vast te stellen. CBS cijfers laten ten opzichte van 2019 in 2021 een stijging van het vervoerd gewicht van landbouwproducten en voeding, bouw en handelsgoederen (waaronder textiel, transportmiddelen, meubels, brieven en pakketten en groupage goederen) zien en een daling van meststoffen en afval en overige goederen (CBS Statline, 2022e). Toekomstig onderzoek zou de cluster analyses kunnen herhalen voor ritten in 2022 om vast te stellen hoe stabiel ritpatronen zijn over tijd.

Een belangrijke kanttekening bij dit onderzoek is dat er andere ritkenmerken zijn die invloed hebben op laadvraag ofwel energieverbruik en niet zijn meegenomen, namelijk het rijgedrag van de chauffeur, wegomstandigheden (o.a. drukte op de weg) en type wegen (binnen of buiten de bebouwde kom, snelwegen of niet). Ook voertuigkenmerken, zoals motorefficiëntie, zijn niet meegenomen in de bepaling van de clusters, maar achteraf gebruikt voor het opstellen van profielen van de typen ritten. Het gewicht van het voertuig zit in dit geval wel in het (brutoplus) vervoerd gewicht. Het ladinggewicht en snelheid van het voertuig zijn volgens Demir, Bektas en Laporte (2013) de meest bestudeerde ritkenmerken in eerdere studies naar verduurzaming van goederentransport.

Een segment dat een relatief hoog energieverbruik heeft tijdens de rit, is het geconditioneerd transport van bijvoorbeeld voedingsmiddelen. Het onderscheid tussen gekoeld (of geconditioneerd) en niet-gekoeld transport zou in toekomstig onderzoek meegenomen kunnen worden als inputvariabele om de toekomstige laadbehoefte meer gedetailleerd te kunnen inschatten.

Resultaten uit dit onderzoek kunnen mogelijk aanknopingspunten bieden voor branches om elektrificatie van voertuigen te prioriteren. Daarnaast kan vervolgonderzoek dieper ingaan op waar

bepaalde typen ritten worden gereden, bij welk type rit aan het eind van de dag het voertuig meestal terug naar het eigen bedrijf rijdt, ritpatronen per regio analyseren of rusttijden tijdens of tussen ritten in kaart brengen. Dit kan mogelijk helpen bij het plannen van de plaatsing van laadpalen voor elektrische vrachtvoertuigen. Tot slot kan vervolgonderzoek uitgaan van dagkilometrages om clusters te ontdekken om de totale dagelijkse laadbehoefte van een voertuig in kaart te brengen.

Referenties

Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A., 2015. NbClust: Determining the best number of clusters in a data set. Cran R-project. Beschikbaar op: <https://CRAN.R-project.org/package=NbClust>

Demir, E., Bektas, T. & Laporte, G., 2013. A Review of Recent Research on Green Road Freight Transportation. Research School for Operations Management and Logistics, *Beta working paper series 428*, blz. 1-40.

ElaadNL, 2021, Onderzoek onder transportondernemers: Veel uitdagingen en zorgen om laadbehoefte en netcapaciteit. Beschikbaar op: <https://elaad.nl/onderzoek-onder-transportondernemers-veel-uitdagingen-en-zorgen-om-laadbehoefte-en-netcapaciteit/>.

Hoaglin, D. C., Iglewicz, B. & Tukey, J.W., 1986. Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81 (396), blz. 991-999.

Hubert, L. & Arabie, P., 1985. Comparing partitions. *Journal of Classification*, 2 (1), blz. 193-218.

Liimatainen, H. & Pöllänen, M., 2010. Trends of energy efficiency in Finnish road freight transport 1995-2009 and forecast to 2016. *Energy Policy*, 38, blz. 7676-7686.

Natuur en Milieu, 2020. De elektrische vrachtwagen in opkomst. Uitdagingen en oplossingen voor laadinfrastructuur. Utrecht, Natuur en Milieu. Beschikbaar op: <https://natuurenmilieu.nl/app/uploads/De-elektrische-vrachtwagen-in-opkomst.pdf>

Odhams, A. M. C., Roebuck, R. L. , Lee, Y. J., Hunt, S. W. & Cebon, D., 2010. Factors influencing the energy consumption of road freight transport. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 224 (9), blz. 1995-2010.

Planbureau voor de Leefomgeving, 2022, Klimaat- en Energieverkenning 2022. Den Haag: PBL. Beschikbaar op: <https://www.pbl.nl/publicaties/klimaat-en-energieverkenning-2022>

Rousseeuw, P. J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, blz. 53-65.

Rousseeuw, P. J. & Hubert, M., 2011. Robust statistics for outlier detection. *WIREs Data Mining and Knowledge Discovery*, 1, blz. 73-79.

CBS Statline, 2021, Goederenvervoer; vervoerwijzen, vervoerstromen van en naar Nederland. Beschikbaar op: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83101NED/table?dl=74FA5>

CBS Statline, 2022a, Motorvoertuigen; voertuigtype, postcode en regio's, 1 januari, 2000-2022. Beschikbaar op: <https://opendata.cbs.nl/#/CBS/nl/dataset/37209hvv/table?dl=74F8C>

CBS Statline, 2022b, Emissies naar lucht op Nederlands grondgebied; wegverkeer. Beschikbaar op: <https://opendata.cbs.nl/#/CBS/nl/dataset/85347NED/table?dl=74F90>

CBS Statline, 2022c, Wegvervoer; kerncijfers. Beschikbaar op: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83077NED/table?dl=76A70>

CBS Statline, 2022d, Bedrijfsvoertuigen actief; voertuigkenmerken, regio's, 1 januari. <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85239NED/table?dl=76A7F>

CBS Statline, 2022e Wegvervoer; vervoerd gewicht naar goederensoort NST 2007. <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83079NED/table?dl=76F38>

Bijlage: Aanpak cluster analyses

Vorbewerkingen

De cluster analyse is door middel van een aantal stappen uitgevoerd. Eerst is gekeken welke aspecten van ritten gerelateerd zijn aan de verwachte laadvraag ofwel energieverbruik. Hierbij is gekozen om te kijken naar de combinatie van de afgelegde ritafstand, het brutoplus vervoerd gewicht tijdens de rit en het aantal deelritten tijdens de rit. Voor deze drie aspecten is bekeken of er ritten met extreme waarden (extreem lange afstanden, extreem hoog gewicht of extreem veel deelritten ten opzichte van de rest van de ritten) in de data voorkomen, de zogenoemde 'uitbijters'. Deze uitbijters zijn uitgesloten van de analyse omdat ze de uiteindelijke resultaten van de cluster analyses kunnen vertekenen. De uitbijters kunnen door hun afwijkende ritkenmerken bijvoorbeeld één cluster (groep) vormen, waardoor de ritpatronen van de reguliere ritten gemaskeerd worden en deze allemaal in een ander, groot cluster terecht komen. Hierdoor kan te weinig onderscheid gemaakt worden binnen de grote groep reguliere ritten. Alle ritten met scores van meer dan 1,5 keer de interkwartielafstand boven de grens van het derde kwartiel op ritafstand of bruto vervoerd gewicht zijn daarom uitgesloten. De ondergrens van 1,5 keer onder de grens van het eerste kwartiel is hier niet relevant, omdat dit niet-voorkomende waarden onder 0 zouden zijn. Deze veelgebruikte methode is in dit geval geschikter voor uitbijterdetectie dan selectie op basis van afwijkingen van standaarddeviaties, omdat de data niet normaal verdeeld zijn (Hoaglin, Iglewicz en Tukey, 1986; Rousseeuw en Hubert, 2011). Voor het aantal deelritten is deze methode niet toepasbaar, omdat de interkwartielafstand 0 is. Daarom is hier een afkappunt van 3 keer de standaarddeviatie boven het gemiddeld aantal deelritten gekozen (onder het gemiddelde is niet relevant aangezien het gemiddeld aantal deelritten 1,4 is en de standaarddeviatie 2,0 deelritten). Er is voor gekozen om alle univariate uitbijters uit te sluiten en niet een complexere methode voor detectie van multivariate uitbijters in te zetten, waardoor mogelijk door de strenge selectie in totaal iets teveel uitbijters zijn weggelaten. Hier is voor gekozen omdat we geïnteresseerd zijn in het beschrijven van de reguliere ritten en we het daarom minder erg vinden dat er mogelijk wat te veel uitzonderlijke ritten worden weggelaten. Ritten die 243 kilometer of langer zijn of waarbij 35.121 kilogram of meer is vervoerd of die meer dan 7 deelritten hadden, worden uitgesloten bij de cluster analyses. Bij deze afkappunten valt bij de afstand 6,2 procent van de ritten af, bij het vervoerd gewicht 1,3 procent en bij de deelritten 2,0 procent. In totaal worden op basis van de uitbijterdetectie 35.621 ritten ofwel 8,7 procent van de ritten niet meegenomen waardoor er voor de cluster analyses 374.132 ritten overblijven. Na het verwijderen van de uitbijters zijn de drie inputvariabelen gestandaardiseerd, zodat zij onderling vergelijkbaar zijn.

Bepalen aantal clusters

De tweede stap in de clusteranalyse betrof het bepalen van het optimale aantal clusters, ofwel het aantal duidelijk onderscheidbare groepen ritten. Vanwege de benodigde rekencapaciteit is dit op basis van 2 steekproeven van 7 000 ritten gedaan (1,7 procent van de ritten). Hiermee kan ook worden bekeken of het optimale aantal clusters varieert als een andere steekproef wordt gebruikt wat een idee geeft van de robuustheid van de resultaten, er wordt getest met minimaal 2 tot maximaal 15 clusters. Daarnaast zijn per steekproef een aantal schattingsmethoden (zowel verschillende afstandsmaten om afstanden tussen personen te bepalen als methoden om clusters af te bakenen) getest om de stabiliteit van de resultaten vast te stellen.

Op basis van agglomeratieve hiërarchische clustering zijn afstanden ofwel verschillen tussen ritten bepaald. Agglomeratieve clustering gaat uit van iedere rit als een apart cluster en gaat vervolgens ritten samenvoegen in grotere clusters. Clustering wordt dus bottom-up in plaats van top-down (één cluster met alle ritten dat vervolgens wordt opgesplitst) gedaan en is het meest gebruikte type hiërarchische clustering. Er is hierbij met twee verschillende afstandsmaten getest of het aantal clusters varieert, namelijk Euclidean (kortste afstand tussen iedere twee punten in een rechte lijn) en Manhattan (de afstand tussen iedere twee punten volgens de assen van de inputvariabelen) afstandsmaat. De keuze tussen deze twee afstandsmaten levert een andere clustering op. In combinatie met deze afstandsmaten zijn twee cluster methoden getest, namelijk Ward (het minimaliseren van variantie binnen clusters) en Average clustering (gemiddelde afstand tussen ritten van verschillende clusters wordt gemaximaliseerd). De methoden bepalen op verschillende manieren het aantal clusters en de grenzen tussen de clusters. Op basis van 30 indices (ofwel evaluatiematen) wordt vervolgens het optimale aantal clusters gekozen. Hierbij wordt de uitkomst van de meerderheid van deze indices gebruikt. Het R package NbClust is hiervoor gebruikt (Charrad, Ghazzali, Boiteau, en Niknafs, 2015). Onder andere de Silhouette index van Rousseeuw (1987) en de Hubert index (Hubert en Arabie, 1985) vallen onder deze 30 indices. Uit de analyse blijkt dat het optimale aantal clusters 2, 3 of 4 en in sommige gevallen 8 is al naar gelang de gekozen afstandsmaten en methoden. In 2019 varieert het aantal clusters dat met twee steekproeven van 7 000 ritten met dezelfde analyses gevonden wordt tussen de 2 of 4.

Tabel 1. Optimaal aantal clusters volgens 30 indices per steekproef (2) naar methode (2) en afstandsmaat (2), 2021

Steekproef en Methode	Afstandsmaat	
	Euclidean (aantal indices)	Manhattan (aantal indices)
1. Ward	4 (5)	8 (6)
1. Average	8 (8)	2 (7)
2. Ward	4 (6)	3 (6)
2. Average	2 (7)	2 (8)

Daarnaast kan door middel van 2-staps clustering naar alle ritten (met uitzondering van de uitbijters) worden gekeken. Als hierbij de loglikelihood afstandsmaat wordt toegepast, worden 4 clusters gevonden met een gemiddelde Silhouette index van 0,6 wat een goede kwaliteit model impliceert. Het nadeel van deze methode is de assumptie dat de data normaal verdeeld zijn, waar de data in dit geval niet aan voldoen. Bij een alternatieve 2-staps clustering met alle ritten op basis van de Euclidean afstandsmaat, worden 2 clusters gevonden met een goede modelkwaliteit (gemiddelde Silhouette index = 0,6). Bij deze 2 clusters, is er een cluster van 97 procent ritten en een cluster van 3 procent ritten. Dit is inhoudelijk weinig informatief aangezien slechts een zeer klein groepje ritten wordt onderscheiden van de rest.

Met de 2019 data zijn dezelfde analyses uitgevoerd, in de 2-staps clustering worden net als in 2021 4 clusters gevonden als de loglikelihood afstandsmaat wordt toegepast (gemiddelde Silhouette index = 0,6). En met de Euclidean afstandsmaat worden 2 clusters gevonden (gemiddelde Silhouette index = 0,5), eveneens zoals in 2021. Het kleinste cluster is met 4,7 procent van de ritten in 2019 relatief groter dan in 2021 (0,3 procent).

Schatting en beschrijving clusters

Als derde stap zijn de clusters met ritten geschat. Hierbij is gebruik gemaakt van k-means clustering (waarbij de totale variantie binnen clusters wordt geminimaliseerd) en zijn zowel de 2-cluster, 3-cluster, 4-cluster oplossing als de 8-cluster oplossing bekeken. Hierbij is naar de grootte van de clusters gekeken en de gemiddelde scores op de inputvariabele per cluster (zogenaamde clustercentra), tabel 2 geeft een overzicht van de vier cluster oplossingen. De grootte van de clusters is enigszins anders dan bij de 2-staps clustering.

Tabel 2. Aantal ritten per cluster voor 2-, 3-, 4- en 8-cluster oplossing met gemiddelde van inputvariabelen, 2021

Aantal clusters	Aantal ritten per cluster (%)	Gemiddelde ritafstand (SD)	Gemiddeld vervoerd gewicht (SD)	Gemiddeld aantal deelritten (SD)
2-cluster oplossing				
1	16.324 (4)	113 (62)	11.582 (8.368)	4 (1,4)
2	357.808 (96)	61 (54)	7.232 (10.127)	1 (0,2)
3-cluster oplossing				
1	269.693 (72)	55 (49)	1.990 (3.231)	1 (0,1)
2	15.597 (4)	116 (61)	11.248 (7.903)	4 (1,4)
3	88.842 (24)	78 (64)	23.239 (6.648)	1 (0,3)
4-cluster oplossing				
1	85.403 (23)	75 (63)	23.518 (6.497)	1 (0,2)
2	267.506 (72)	54 (49)	1.960 (3.212)	1 (0,1)
3	6.639 (2)	123 (61)	10.893 (7.931)	6 (0,8)
4	14.584 (4)	119 (62)	11.766 (7.790)	3 (0,8)
8-cluster oplossing				
1	1.930 (0,5)	172 (37)	11.576 (7.890)	7 (0,5)
2	57.526 (15)	36 (27)	24.633 (6.047)	1 (0,2)
3	27.315 (7)	158 (41)	21.754 (6.427)	1 (0,3)
4	197.599 (53)	30 (21)	1.987 (3.279)	1 (0,1)
5	74.363 (20)	124 (39)	2.305 (3.461)	1 (0,2)
6	7.305 (2)	69 (32)	11.380 (7.762)	3 (0,6)
7	3.260 (1)	71 (33)	9.445 (7.242)	6 (0,8)
8	4.834 (1)	175 (34)	13.461 (8.863)	4 (0,8)

Noot. Door afronding tellen de percentages niet altijd op tot 100 procent.

Uit de resultaten blijkt de 4-cluster oplossing het meest voor te komen als het optimaal aantal clusters. De beschrijving van de vier clusters wordt in de resultaten paragraaf gepresenteerd. Precieze percentages van de verdelingen naar achtergrondkenmerken per cluster kunnen licht afwijken als een andere schattingsmethode om clusters te maken, wordt gekozen. Het algemene beeld van de verschillen tussen de clusters, verandert echter niet.

In de data van 2019 tekent zich eenzelfde beeld af als in 2021 als k-means clustering voor 2, 3 en 4 clusters wordt toegepast, tabel 3 van de bijlage laat zien dat de cluster oplossingen in 2019 niet veel afwijken van die in 2021. Een 8-cluster oplossing wordt niet getoond, omdat dit aantal niet als optimaal aantal clusters uit de data van 2019 kwam. De relatieve grootte van de clusters is zeer vergelijkbaar, de gemiddelden wijken enigszins af maar laten ongeveer hetzelfde beeld van rittypen zien. Uitzondering

is de afgelegde afstand in de clusters met veel deelritten bij de 3- of 4-cluster oplossingen; deze clusters hadden in 2021 gemiddeld de langste afstanden, terwijl dit in 2019 voor het cluster met een kwart van de ritten met gemiddeld een deelrit gold. Daarentegen heeft dit cluster gemiddeld niet het zwaarste vervoerd gewicht, terwijl dit in 2021 wel het geval was, al is de standaarddeviatie groter dan bij de kleinere clusters met veel deelritten. De clusters zijn vanwege de vergelijkbaarheid op basis van aantal ritten op dezelfde volgorde als in tabel 2 weergegeven.

Tabel 3. Aantal ritten per cluster voor 2-, 3- en 4-cluster oplossing met gemiddelde van inputvariabelen, 2019

Aantal clusters	Aantal ritten per cluster (%)	Gemiddelde ritafstand (SD)	Gemiddeld vervoerd gewicht (SD)	Gemiddeld aantal deelritten (SD)
2-cluster oplossing				
1	16.870 (5)	113 (63)	11.590 (8.465)	4 (1,4)
2	329.720 (95)	61 (56)	7.938 (10.863)	1 (0,2)
3-cluster oplossing				
1	248.178 (72)	35 (27)	6.020 (9.580)	1 (0,3)
2	9.573 (3)	122 (63)	11.276 (8.098)	5 (1,1)
3	88.839 (26)	137 (48)	13.630 (12.117)	1 (0,5)
4-cluster oplossing				
1	85.062 (25)	138 (45)	12.303 (12.060)	1 (0,2)
2	243.443 (70)	34 (25)	6.336 (9.918)	1 (0,1)
3	6.518 (2)	123 (63)	11.280 (8.179)	6 (0,8)
4	11.567 (3)	103 (60)	12.986 (9.192)	3 (0,6)

Noot. Door afronding tellen de percentages niet altijd op tot 100 procent.