

De Fietsende Treinreiziger: Spits of Dal Reiziger?

Rik Schakenbos – NS Stations – rik.schakenbos@nsstations.nl

Danique Ton – NS Stations – danique.ton@nsstations.nl

Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 25 en 26 november 2021, Utrecht

Samenvatting

De combinatie fiets-trein is een groot succes in Nederland. In dit paper kijken we met een verkeerskundige blik naar de fietsende treinreiziger en onderzoeken we de reispatronen van deze groep reizigers binnen de totale populatie van treinreizigers in Nederland. Hiervoor vergelijken we fietsparkeerdata uit bewaakte stallingen met data over het totaal aantal treinreizigers op een treinstation. Het betreft fietsparkeerdata van 72 treinstations in Nederland over geheel 2019. Voor zover ons bekend is het de eerste keer dat de gerealiseerde reispatronen van de fietsende treinreizigers op basis van het daadwerkelijke stallingsgebruik grootschalig onderzocht worden.

Er zijn al cijfers bekend over het gemiddeld aandeel fietsende treinreizigers per station op een dag, maar in dit paper laten we zien dat het aandeel fietsende treinreizigers over een dag niet uniform verdeeld is. De fietsende treinreiziger is sterk oververtegenwoordigd in de spitsuren en ondervertegenwoordigd in de daluren. Hiermee is de 'typische' fietsende treinreiziger een spitsreiziger.

Deze bevinding heeft tot gevolg dat inschattingen van piekdruktes in fietsenstallingen gebaseerd op het aandeel fietsende treinreizigers op een dag tot grote onderschattingen kunnen leiden. Het nauwkeurig inschatten van piekdruktes in fietsenstallingen is van belang om drukte en lange wachtrijen te voorkomen. Daarom introduceren we een berekeningsmethode met een correctiefactor voor de ongelijke verdeling van fietsende treinreizigers. We tonen aan dat wanneer geen rekening wordt gehouden met deze correctiefactor, de piekdruktes in fietsenstallingen in werkelijkheid tot wel 60% hoger kan liggen op specifieke stations.

Op vrijwel alle stations ligt het aandeel fietsende treinreizigers in de spits hoger dan het gemiddelde aandeel fietsende treinreizigers over de hele dag, maar de verschillen in correctiefactoren tussen stations zijn groot. Om ook voor stations waar geen fietsparkeerdata beschikbaar is, een juiste correctiefactor te kunnen schatten, voeren we een hiërarchische clusteranalyse uit. We verdelen de stations in twee groepen met gelijksoortige correctiefactoren. De geografische ligging van het station, en daaruit volgend de gemiddelde reisafstand per trein lijkt een verklarende variabele te zijn voor het aandeel fietsende treinreizigers in een uur. Fietsende treinreizigers met een grotere reisafstand per trein blijken over het algemeen vroeger naar het station te komen.

Op basis van bovenstaande inzichten zijn we in staat om het verloop van de drukte in fietsenstallingen bij treinstations waarvoor nog geen data beschikbaar is nauwkeuriger in te schatten.

1. Introductie

De afgelopen jaren is NS Stations steeds meer bewaakte fietsenstallingen gaan exploiteren. Bij een deel van deze stallingen checken fietsenstallers bij de stalling in- en uit (bijvoorbeeld stallingen met een 1^e 24-uur gratis regime, of een zelfservice-stalling). Hierdoor is inmiddels veel data van het gebruik van deze fietsenstallingen beschikbaar is. Deze data geeft inzicht in reispatronen van de fietsende treinreiziger.

Met dit paper beogen we ten eerste bij te dragen aan de kennis van de fietsende treinreiziger in Nederland. Er zijn reeds verschillende onderzoeken gepubliceerd die dieper ingaan op wie de fietsende treinreiziger is, door te kijken naar het reizigersprofiel, bijvoorbeeld (Jonkeren et al., 2021, 2018; Shelat et al., 2018). Wij kiezen hier voor een verkeerskundige insteek, specifiek gericht op reispatronen. Dit onderzoeken we door een vergelijking te maken tussen de reispatronen van alle treinreizigers en van de fietsenstallers. Is de fietsende treinreiziger nou een echte spitsreiziger, een dalreiziger, of eigenlijk een hele gemiddelde reiziger?

Ten tweede gaan we in op de verkeerskundige consequenties in een fietsenstalling van een eventuele 'scheve' verdeling van fietsende treinreizigers over de dag en onderzoeken we verschillen tussen stations. Vanwege het succes van de fiets-trein combinatie in Nederland is het bij fietsenstallingen rondom het station niet alleen van belang dat er genoeg stallingen zijn (kan iedere fietsende treinreiziger zijn fiets stallen), maar moet ook de doorstroomcapaciteit van de stalling voldoende zijn (kan iedere fietsende treinreiziger snel en comfortabel zijn fiets stallen). Een scheve verdeling van fietsende treinreizigers leidt tot een onder- of onderschatting van piekdruktes in de fietsenstallingen. Wanneer er veel fietsers tegelijkertijd aankomen kunnen er lange wachtrijen ontstaan wanneer de doorstroomcapaciteit van de fietsenstalling hier niet op uitgerust is. Om een fietsenstalling op de juiste doorstroomcapaciteit te dimensioneren is het van belang een goede inschatting te hebben van de intensiteit op de drukste momenten.

Ten derde introduceren we een methode om de maatgevende intensiteit voor fietsenstallingen te berekenen bij een ongelijke verdeling van het aandeel fietsende treinreizigers. Bij een gebrek aan meer gedetailleerde data werd de piekdruktes in een fietsenstalling in het verleden ingeschat door uit te gaan van de verdeling van alle treinreizigers over de dag en hier een station-specifiek percentage van te nemen voor het aandeel fietsende treinreizigers van en naar een station. De impliciete aanname hierbij is dat het aandeel fietsende treinreizigers over de dag heen op een station uniform verdeeld is. In dit paper betogen wij, op basis van de stallingsdata en de treinreizigersdata, dat het aandeel fietsende treinreizigers over de dag niet constant is. Daarom introduceren wij een stationspecifieke correctiefactor per uur.

Tot slot clusteren we de verschillende stations in onze dataset die vergelijkbare correctiefactoren hebben. We verkennen of we een verklarende variabele kunnen vinden om te bepalen in welk station een cluster valt, zodat ook stations waarvoor we nog geen fietsparkeerdata hebben in een cluster ingedeeld kunnen worden.

2. Methode

De belangrijkste databronnen, stallingstransacties van fietsparkeerders en check-ins/outs van treinreizigers worden in 2.1 toegelicht. Vervolgens lichten we in 2.2 de dataverwerking, de berekening van de piekdrukke en een clustering van gelijksoortige stations toe.

2.1 Data

De belangrijkste databronnen voor deze analyse zijn de check-ins/outs van treinreizigers per station en de stallingstransacties van fietsparkeerders in fietsenstallingen in en rondom de stations.

Wat betreft fietsenstallingen is in dit onderzoek data beschikbaar van bewaakte fietsenstallingen die door NS Stations worden geëxploiteerd. Dit kunnen bemenste fietsenstallingen zijn, maar ook zelfservice fietsenstallingen. Voor al deze fietsenstallingen is een dataset samengesteld met het aantal fietsparkeerders per dag per uur per fietsenstalling. Met fietsparkeerder wordt in dit paper zowel iemand bedoeld die een fiets komt stallen als iemand die een gestalde fiets op komt halen. Wanneer een station meerdere fietsenstallingen heeft (bijvoorbeeld Utrecht Centraal) zijn de aantallen van deze stallingen geaggregeerd voor dit station. De resulterende dataset bestaat uit 72 unieke treinstations, zowel grote stations (Utrecht Centraal, Amsterdam Zuid) als kleinere stations (Zaltbommel, Voorburg). De stations bevinden zich verspreid over Nederland.

Voor de 72 treinstations waarvoor stallingsdata beschikbaar is, is eenzelfde dataset gemaakt met het aantal in-/uitstappers per dag per uur op deze treinstations. De aantallen treinreizigers zijn gebaseerd op de OV-chipkaart transacties op deze treinstations.

Ondanks dat de stallingstransactie van de fietsparkeerders en de check-in/out van treinreizigers op het station veelal met dezelfde OV-chipkaart gedaan wordt, zijn deze gegevens niet aan elkaar gekoppeld omdat hiervoor (mogelijk) persoonsgegevens verwerkt zouden moeten worden. We werken enkel met aantallen transacties per uur per station zonder enige andere kenmerken. Dit betekent dat we in de dataset van treinreizigers niet kunnen zien welke reizigers ook hun fiets gestald hebben en andersom kunnen we in de dataset van de fietsparkeerders niet zien of deze vervolgens ook een treinreis zijn gaan maken.

De dataset bestaat uit alle weekdays van 1 januari 2019 – 31 december 2019 (pre-corona).

2.2 Analysemethode

De analyse van de data bestaat uit drie verschillende elementen. Allereerst het verwerken van de data tot informatie over fietsende treinreizigers waar analyses op kunnen worden uitgevoerd. Vervolgens het berekenen van de piekdrukke, op basis waarvan onder andere de dimensionering van nieuwe fietsenstallingen wordt gedaan. Hier belichten we de oorspronkelijke manier van rekenen en introduceren we een nieuwe manier van rekenen op basis van een correctiefactor voor de scheve verdeling van fietsparkeerders ten opzichte van in-/uitstappers over de dag. Deze correctiefactor wordt berekend voor elk station, zijn die allemaal gelijk of zien we hier grote verschillen tussen stations? Vervolgens bespreken we hoe we middels hiërarchisch clusteren vergelijkbare patronen zoeken in de

correctiefactoren tussen de verschillende stations. Dit laatste is relevant voor extra begrip van fietsparkeerpatronen of voor extrapolatie/generalisatie naar andere stations.

Dataverwerking

Om de reispatronen van de fietsende treinreizigers te kunnen destilleren gebruiken we data van de in-/uitstappers per station en van de fietsparkeerders per station. Hierbij worden zowel de in- als uitstappers en de in- als uitcheckende fietsers meegenomen. Niet alle fietsparkeerders bij stations worden gezien in de bewaakte fietsenstallingen, aangezien er ook onbewaakte maaiveld stallingen zijn. Daarom bevat de fietsparkeerdata (vaak) maar een deel van alle fietsparkeerders op een station. Onze aanname is dat het patroon (verdeling over de dag) van deze fietsparkeerders representatief is voor alle fietsparkeerders op hetzelfde specifieke station. Om toch een vergelijking te kunnen maken in reispatronen tussen de in-/uitstappers en fietsparkeerders stellen we het totaal aantal observaties op doordeweekse dagen per station in het jaar 2019 gelijk aan 100%, voor zowel de in-/uitstappers als de fietsparkeerders waarvan we data tot onze beschikking hebben. Op basis van deze totalen, kunnen we het aandeel van in-/uitstappers en fietsparkeerders berekenen per uur per gemiddelde werkdag, zodat we kunnen gaan kijken naar verdeling over spits- en daluren. Vergelijking 1 en 2 tonen respectievelijk de berekening voor en het aandeel in-/uitstappers en het aandeel fietsparkeerders op een specifiek uur (u) op een station (s). Hierbij is $InUit$ het aantal in-/uitstappers en $Fiets$ het aantal fietsparkeerders

$$q_{InUit_{su}} = \frac{InUit_{su}}{\sum_{u=0}^{23} InUit_{su}} \quad (1)$$

$$q_{Fiets_{su}} = \frac{Fiets_{su}}{\sum_{u=0}^{23} Fiets_{su}} \quad (2)$$

In hoofdstuk 3 plotten we de aandelen per station (s) per uur (u) voor fietsparkeerders en in-/uitstappers in een grafiek zodat er visueel een beeld wordt geschetst van de verschillen en overeenkomsten tussen de beide patronen.

Berekening van de piekdrukke

Om de piekdrukke van de fietsenstalling te bepalen waarvoor nog geen specifieke data van fietsparkeerders beschikbaar is, kijken we normaal gesproken naar het aandeel in-/uitstappers in een tijdsperiode (bijvoorbeeld uur) ten opzichte van de dag (vergelijking 1) en vermenigvuldigen we dit aandeel met het aandeel treinreizigers dat met de fiets van of naar het station komt. In het verleden was niet meer gedetailleerdere data beschikbaar over het verschil gedurende de dag in het aandeel fietsparkeerders, dus veronderstelden we deze constant per station (Vergelijking 3). Deze methode noemen we in dit paper FTO (Fietsende Treinreizigers Oorspronkelijk)

$$FTO_{su} = q_{InUit_{su}} \cdot Aandeel_{Fiets_s} \quad (3)$$

$$Aandeel_{Fiets_s} = VTR_{sf} \cdot Productie_s + NTR_{sf} \cdot Attractie_s \quad (4)$$

Hierbij is $Aandeel_Fiets_s$ voor een station opgebouwd uit twee elementen, namelijk voortransport (VTR) en natransport (NTR)¹. De reizigers naar een station kunnen zowel met de fiets reizen tussen woning en station (voortransport), als met de fiets reizen tussen station en bestemming (natriansport). Daarnaast heeft elk station een productie- en attractiewaarde. Het aandeel productie is het aandeel reizigers op een station dat dagelijks tussen woning en station reist (zowel heen als terug). Het aandeel attractie is het aandeel reizigers dat dagelijks tussen station en bestemming reist (zowel heen als terug). Zo hebben stations gelegen in een woonwijk over het algemeen een hoog productieaandeel, en is het attractieaandeel hoog bij bijvoorbeeld een station in de buurt van een bedrijventerrein. Het totale aandeel fietsbewegingen (zowel stallen als ophalen) van alle treinreizigers is berekend door rekening te houden met de productie- en attractiewaarde van een station, in combinatie met het aandeel voor- en natransport fiets voor dit station (vergelijking 4).

De aanname die voorheen werd gedaan is dat de fietsparkeerders eenzelfde patroon volgen als de in-/uitstappers. Onze hypothese is dat dit niet het geval is en dat er andere piekpatronen zijn voor fietsende treinreizigers dan voor het totaal treinreizigers. Daarom introduceren we een correctiefactor op FTO, waarbij wordt gecorrigeerd voor de afwijking van de fiets-in/uit verhouding. Deze correctiefactor (CF) wordt berekend per station (s) per uur (u) in vergelijking 5:

$$CF_{su} = \frac{q_Fiets_{su}}{q_InUit_{su}} \quad (5)$$

Als de patronen van fietsparkeerders en in-/uitstappers gelijk zijn aan elkaar (en onze hypothese wordt verworpen) is de CF (nagenoeg) gelijk aan 1 voor alle uren en zal er geen correctie hoeven worden uitgevoerd. In dat geval voldoet de oorspronkelijke berekening van piekdrukke in de fietsenstallingen. Als dit niet het geval is, dan moet er een nieuwe berekening worden gemaakt waarbij de CF wordt opgenomen in een nieuwe berekeningsmethode FTN (Fietsende Treinreiziger Nieuw). Daarin wordt voor de fietsende treinreizigers per station (s) per uur (u) de correctiefactor op de afwijkende patronen van fietsparkeerders ten op zichte van treinreizigers meegenomen (vergelijking 6).

$$FTN_{su} = q_InUit_{su} \cdot Aandeel_Fiets_s \cdot CF_{su} \quad (6)$$

Het resultaat van zowel FTO als FTN is het verwachte aandeel fietsbewegingen in een specifiek uur ten opzichte van het totaal aantal in-/uitstappers over de gehele dag. Deze twee maten (FTO en FTN) van fietsende treinreizigers kunnen we naast elkaar leggen om te kijken waar de verschillen zitten. Het relevante verschil is dan het verschil tussen de maatgevende piek (hoogste waarde) volgens FTN en de maatgevende piek volgens FTO:

$$\Delta\text{Maatgevende Piek}_s = \max_{u \in \{0:23\}} (FTN_s) - \max_{u \in \{0:23\}} (FTO_s) \quad (7)$$

¹ Op <https://dashboards.nsjaarverslag.nl/reizigersgedrag> zijn per treinstation waar NS als vervoerder actief is de aandelen voor- en natransport per modaliteit te vinden. Deze gegevens zijn gebaseerd op onderzoek onder treinreizigers.

Door de introductie van de CF kan er een verandering zijn in welk uur het maximum wordt bereikt.

Clusteren van correctiefactoren

We hebben informatie over 72 stations waar een fietsenstalling door NS Stations wordt geëxploiteerd. Ieder van deze stations heeft een ander patroon rondom in-/uitstappers en fietsparkeerders, afhankelijk van bijvoorbeeld dienstregeling, aanbod op het station en ligging in het land. De vraag is allereerst of er overeenkomsten zijn tussen stations of dat deze echt anders zijn als het gaat om de reizigerspatronen en fietsparkeerpatronen. Als er overeenkomende patronen zijn tussen stations, die we kunnen verklaren aan de hand van kenmerken van de stations, dan zouden we kunnen extrapoleren naar stations waar we geen stallingsdata beschikbaar hebben. Dit betekent dat als daar bijvoorbeeld nieuwe fietsenstallingen worden gebouwd, we een betere inschatting kunnen krijgen van de maatgevende piek dan dat we voorheen hadden gekund.

Om de overeenkomende patronen te vinden voeren we een hiërarchische clustering uit op de data. Hierbij starten we met de observaties per uur over de stations en op basis van de Euclidische afstand tussen de CF per uur per station (vergelijking 5) gaan we clusteren, waarbij de gedachte is dat de afstand tussen stations in een cluster is geminimaliseerd en de afstand tussen stations in andere clusters is gemaximaliseerd. Hiërarchisch clusteren toont dan aan wat de afstand is voor 1 tot en met n clusters, waarbij n het totaal aantal stations is. Ergens tussen 1 en n ligt het optimum aantal clusters, wat wordt bepaald door de within-sum-of-squares (WSS) te minimaliseren en de gap statistic te maximaliseren. De WSS is ook bekend als de elleboogmethode, daar waar de knik het grootst is, kun je het optimaal aantal clusters vinden. De gap statistic vergelijkt intra-cluster variatie voor verschillende aantallen clusters met hun verwachte waarden onder de referentiewaarde van de verdeling van de data (Tibshirani et al., 2001). Het optimale aantal clusters is dat aantal waar de gap statistic het grootst is (het grootste gat tussen de geobserveerde variatie en de verwachte waarde). Ofwel, dit aantal clusters ligt het verst weg van een random uniforme verdeling. Door de combinatie van deze twee optimalisatie criteria kunnen we bepalen hoeveel clusters passend is voor onze data.

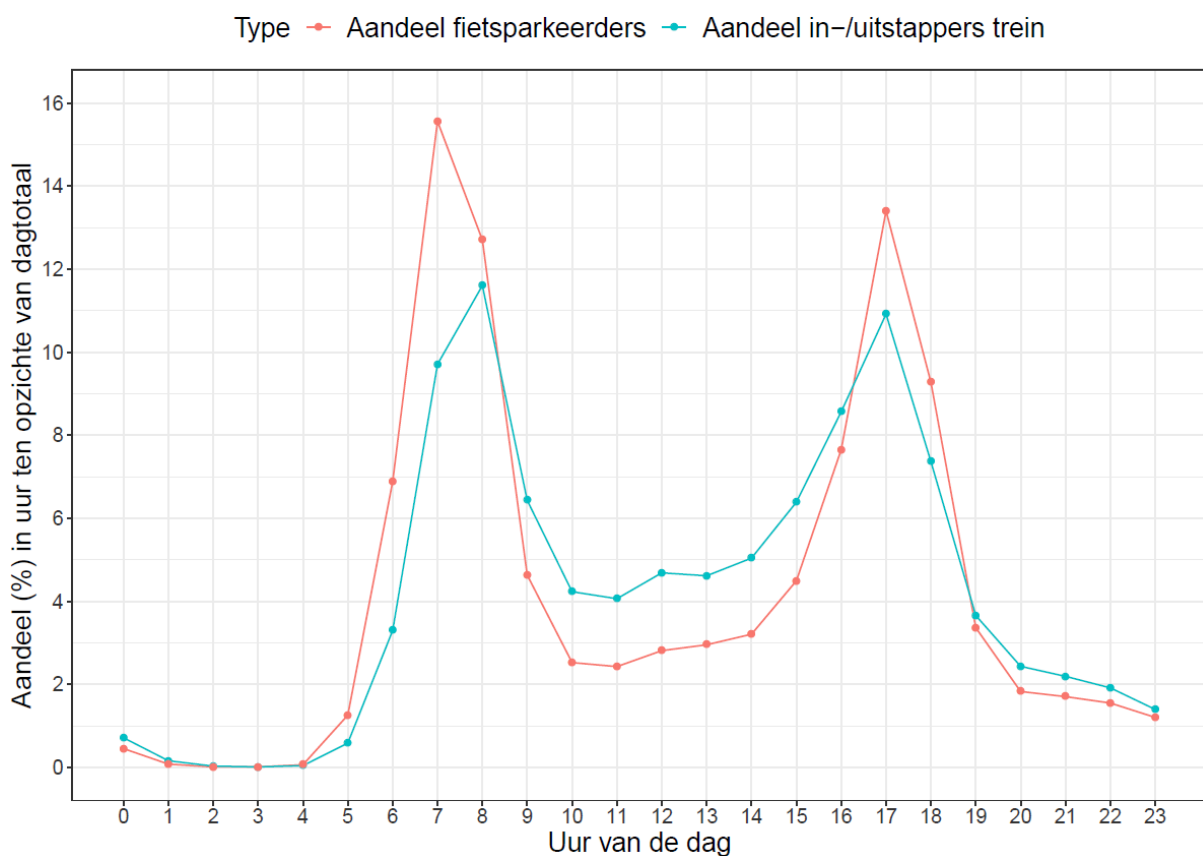
De aantallen clusters met hun eigen patronen van de CF kunnen we vervolgens gaan duiden door stations aan deze clusters toe te voegen. Door naar geografische spreiding en treinkenmerken per station te kijken kunnen we inschatten wat maakt dat een station in een bepaald cluster valt. Dit helpt vervolgens om te extrapoleren naar stations waar we nog geen informatie hebben over fietsparkeerders.

3. Reispatronen van fietsende treinreizigers

Op basis van de datasets van in-/uitstappers en fietsparkeerders zoals beschreven in 2.1 is voor de betreffende 72 stations een landelijk gemiddeld beeld gemaakt van de verdeling over de dag voor zowel alle in-/uitstappers als de fietsparkeerders volgens vergelijking 1 en 2. Om het gemiddelde beeld te maken is eerst de verdeling per uur per station berekend. Vervolgens is het gemiddelde aandeel per uur van deze 72 stations bepaald, waarbij ieder station (groot of klein) even zwaar woog. Er is gekozen om niet te wegen naar in-/uitstappers of fietsparkeerders, omdat dan enkele grote stations een zeer zwaar gewicht krijgen en aangezien niet alle treinstations in de data zitten deze stations dan oververtegenwoordigd zouden zijn. In Figuur 1 is voor zowel de in-/uitstappers als de fietsparkeerders de gemiddelde verdeling over de uren van de dag te zien.

Voor zowel in-/uitstappers als fietsparkeerders is een duidelijke ochtendspits en avondspits met daartussen een dalperiode zichtbaar. Echter zijn de pieken van de fietsparkeerders aanzienlijk hoger dan van de treinreizigers. Op deze 72 stations checkt gemiddeld <10% van de treinreizigers in of uit in uur 7, terwijl >15% van de fietsen gehaald of gebracht worden in dit uur. Dit betekent dat in uur 7 het aantal fietsparkeerders meer dan 50% hoger is dan op basis van de aantallen in-/uitstappers verwacht zou worden. Het omgekeerde effect is zichtbaar tijdens de daluren: in uur 11 zien we 4% van het totaal van de in-/uitstappers, terwijl we ongeveer 2.5% van het totaal aantal fietsparkeerders zien in dit uur.

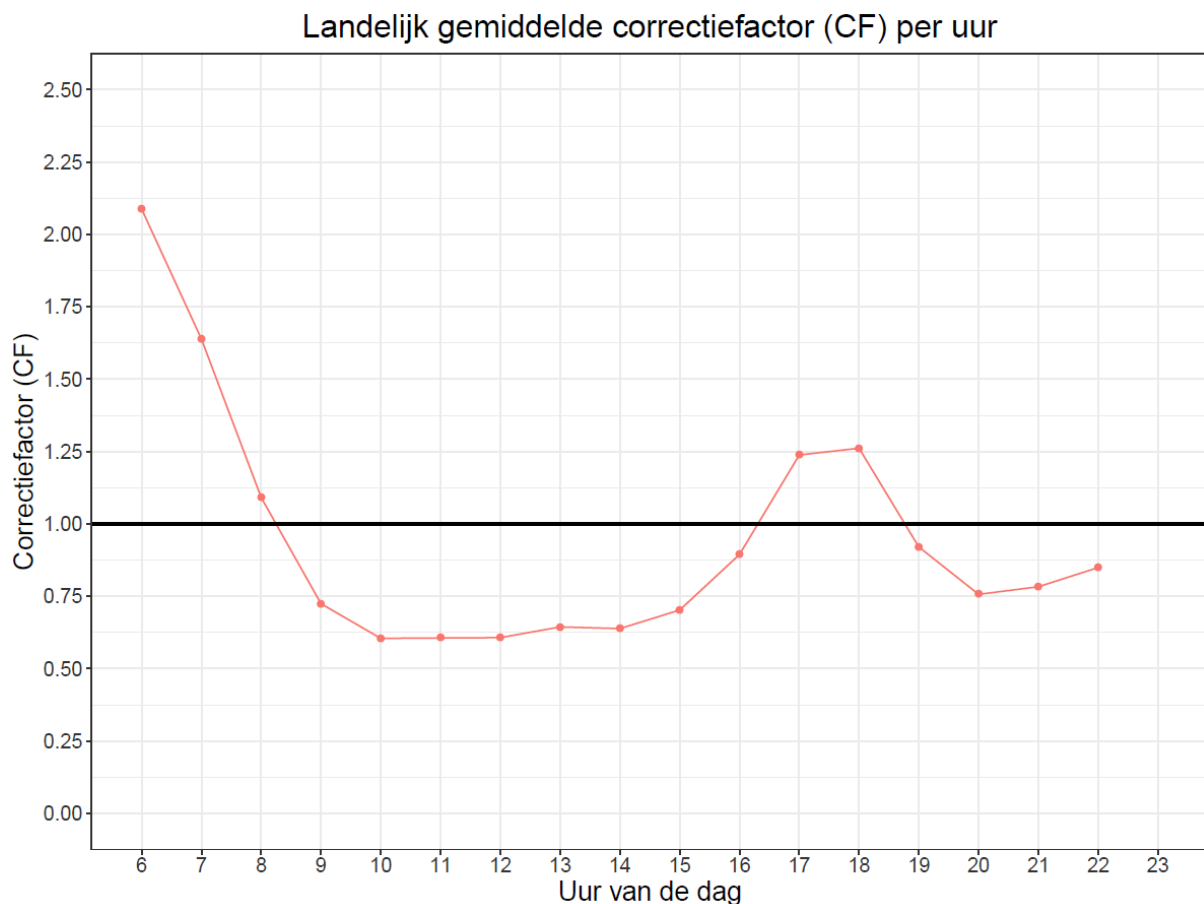
Landelijk gemiddeld patroon in-/uitstappers trein en fietsparkeerders



Figuur 1: Verhouding in-/uitstappers en fietsparkeerders over de dag

De verhouding tussen het aandeel fietsparkeerders en het aandeel treinreizigers is in vergelijking 5 gedefinieerd als de correctiefactor CF. De gemiddelde correctiefactor per uur voor de betreffende 72 stations is weergegeven in Figuur 2. Hierin is nogmaals duidelijk te zien in welke mate de fietsparkeerders in de ochtendspits zijn oververtegenwoordigd ten opzichte van het aandeel fietsparkeerders gedurende de rest van de dag. Maar ook in de avondspits (uren 17 en 18) zijn er 25% meer fietsparkeerders dan we ten opzichte van het aandeel treinreizigers zouden verwachten. Logischerwijs is daarmee het aandeel fietsparkeerders in de daluren juist laag (een $CF < 1$).

Deze verhoudingen laten zien dat landelijk het aandeel fietsende treinreizigers niet evenredig over de dag verdeeld is. Het relatieve aandeel fietsers is in de spits groter dan buiten de spits. De fietsende treinreiziger is daarmee voornamelijk een spitsreiziger. Dit komt overeen met bevindingen van onder andere het KIM (Jonkeren et al., 2021, 2018) dat de fietsende treinreiziger vaker een woon-werk motief heeft dan de treinreiziger die niet fietst. Woon-werk reizen worden namelijk grotendeels in de spits gemaakt.

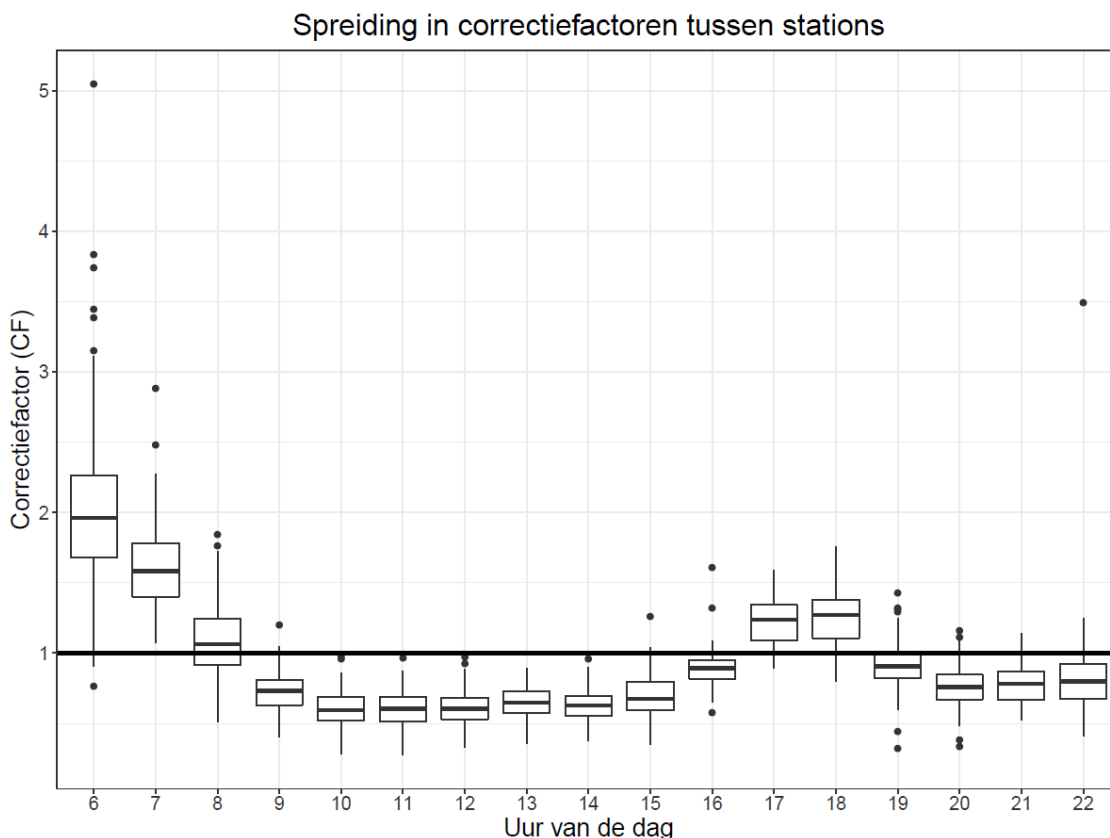


Figuur 2: Landelijke correctiefactor (CF)

4. Piekdrukke fietsparkeerders per station

Voor een goede doorstroming in de fietsenstalling is het van belang dat de fietsenstalling piekdruktes aan kan. Zoals in 2.2 is aangegeven werd de piekdrukke eerder ingeschat op basis van een directe afgeleide van de piek van het aantal in-/uitstappers (FTO, vergelijking 3) wanneer er geen fietsparkeerdata beschikbaar was. In hoofdstuk 3 hebben we echter laten zien dat deze methode een over- of onderschatting zal geven van het aantal fietsers in het betreffende uur. Daarom hebben we FTN (vergelijking 6) geïntroduceerd: met een specifieke correctiefactor CF wordt gecorrigeerd voor de scheve verdeling van het aandeel fietsende treinreizigers over de dag.

Deze correctiefactor (CF) is stationspecifiek. Figuur 3 laat in boxplots de betekenis dat het aandeel fietsparkeerders in verhouding is met het aandeel in-/uitstappers. Een correctiefactor boven de 1 betekent dat de fietsende treinreiziger is oververtegenwoordigd ten opzichte van de in-/uitstappers. Met name in de ochtendspits (07:00 – 09:00) is de spreiding tussen stations groot. Hier komt de correctiefactor bij veel stations ver boven de 1 uit. Tegelijkertijd gaat het in de ochtendspits in absolute zin ook om grote reizigersaantallen dus is deze correctiefactor zeer relevant: op de uren met het grootste aantal in-/uitstappers is het aandeel fietsende treinreizigers ook nog eens groter dan gemiddeld over de dag.

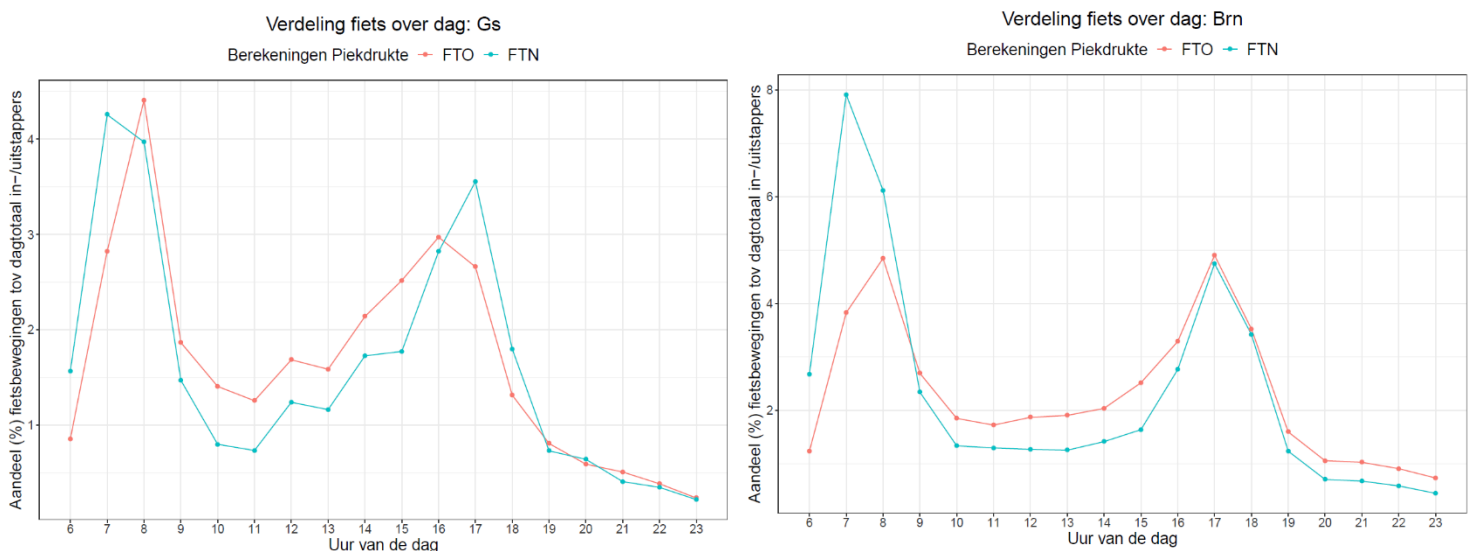


Figuur 3: Boxplots van stationspecifieke correctiefactoren (CF) per station

Een oververtegenwoordiging van het aandeel fietsende treinreizigers in een specifiek uur hoeft echter nog niet altijd te betekenen dat de maatgevende piekdrukke onderschat wordt. Hiervoor moet nog een combinatie gemaakt worden met het totale aantal treinreizigers in het specifieke uur. Zo kan het zijn dat in het uur waarin het aandeel fietsende treinreizigers het hoogste is, het in absolute zin nog om kleine aantallen gaat.

Om dit te onderzoeken hebben we voor ieder station in de dataset per uur het gemiddeld aantal fietsparkeerders berekend volgens de oorspronkelijke methode (FTO) en de nieuwe methode (FTN). Vervolgens hebben we het verschil in maatgevende pieken (Vergelijking 7) voor ieder station berekend. Voor een aantal stations was het verschil in maatgevende pieken erg klein en heeft de het gebruik van een correctiefactor weinig impact. Op een aantal andere stations kwamen wel grote verschillen voor.

Het kleinste verschil in maatgevende piek kwam voor op station Goes, het grootste verschil kwam voor op station Baarn. In Figuur 4 zijn de fietsparkeerders voor deze twee stations weergegeven volgens zowel methode FTO als FTN. Voor station Goes is te zien dat het verschil in maatgevende pieken tussen beide methodes erg klein is. De maatgevende piek lag op zo een 4.4% (aandeel fietsbewegingen in specifiek uur ten opzichte van het totaal aantal in-/uitstappers over de dag) en is in de nieuwe methode 4.2%. In de nieuwe methode nam de maatgevende piekdrukte dus licht af. Bij station Baarn zien we wel grote verschillen: De maatgevende piek lag iets onder de 5% volgens FTO, maar met de introductie van een correctiefactor ligt de piek op 8%. Dit betekent dat er in het drukste uur 60% meer fietsparkeerders zijn dan we op basis van de oude methode zouden verwachten.



Figuur 4: De kleinste en grootste verschillen tussen FTO (rood) en FTN (blauw), links station Goes, rechts station Baarn

Zowel op station Goes als station Baarn zien we dat het maatgevende uur voor fietsparkeerders verschoven is in FTN ten opzichte van het maatgevende uur volgens FTO. In Tabel 1 is het aantal keer dat een uur maatgevend is in FTO uitgezet tegen het aantal keer dat een uur in FTN maatgevend is. Te zien is dat in FTO uur 8 voornamelijk maatgevend was (op 40 van de 72 stations), terwijl dit in FTN verschoven is naar uur 7 (50 van de 72 stations). Dit is te verklaren doordat fietsen voornamelijk als voortransportmiddel wordt gebruikt en in mindere mate als natransport middel. Daardoor ligt de piek voor fietsparkeerders gemiddeld eerder dan de piek van in- en uitstappende treinreizigers.

Tabel 1: Aantal keer dat een specifiek uur maatgevend is voor fietsparkeerders en treinreizigers

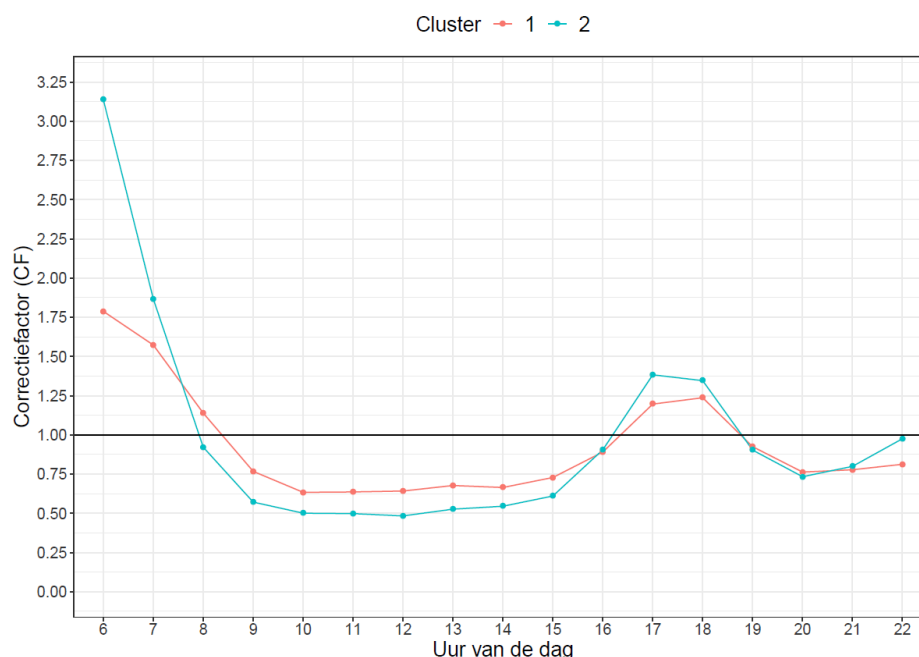
		FTN			
		07:00	08:00	17:00	Totaal
FTO	07:00	18	0	0	18
	08:00	22	16	2	40
	17:00	10	3	1	14
	Totaal	50	19	3	72

5. Clustering van zelfde type stations

Figuur 3 liet zien dat de correctiefactor tussen stations behoorlijk kan verschillen. Voor elk station is op elk uur van een gemiddelde werkdag een correctiefactor (CF) berekend voor de fietsende treinreiziger (Vergelijking 3). In dit hoofdstuk onderzoeken we of het mogelijk is om vergelijkbare stations te kunnen groeperen. Wanneer dit mogelijk is, en we verklarende variabelen voor deze groepering kunnen vinden, is het ook mogelijk een inschatting te maken in welke groep een station zou vallen waarvan we nog geen fietsparkeerdata beschikbaar hebben.

Om dit te onderzoeken passen we hiërarchische clustering toe met Euclidische afstand (op basis van verschillen (dissimilarities)). Middels de gap statistic en WSS komen we uit op twee clusters als optimaal aantal. Figuur 5 toont de gemiddelde profielen van de beide clusters, waarbij cluster 1 wat gematigder is dan cluster 2. Cluster 1 bevat 56 stations en cluster 2 bevat 14 stations. Bij het eerste cluster zien we duidelijk dat er een onderschatting is tijdens de ochtend- en avondspits van de fietsparkeerders ten opzichte van de in-/uitstappers, maar die onderschatting is geringer dan bij cluster 2. Cluster 2 toont zeker in de ochtendspits (vooral in uur 6) dat het aandeel fietsparkeerders erg groot is ten opzichte van wat we op basis van de aantallen in-/uitstappers zouden verwachten. De fietsende treinreizigers lijken op deze stations dus extra vroeg te vertrekken.

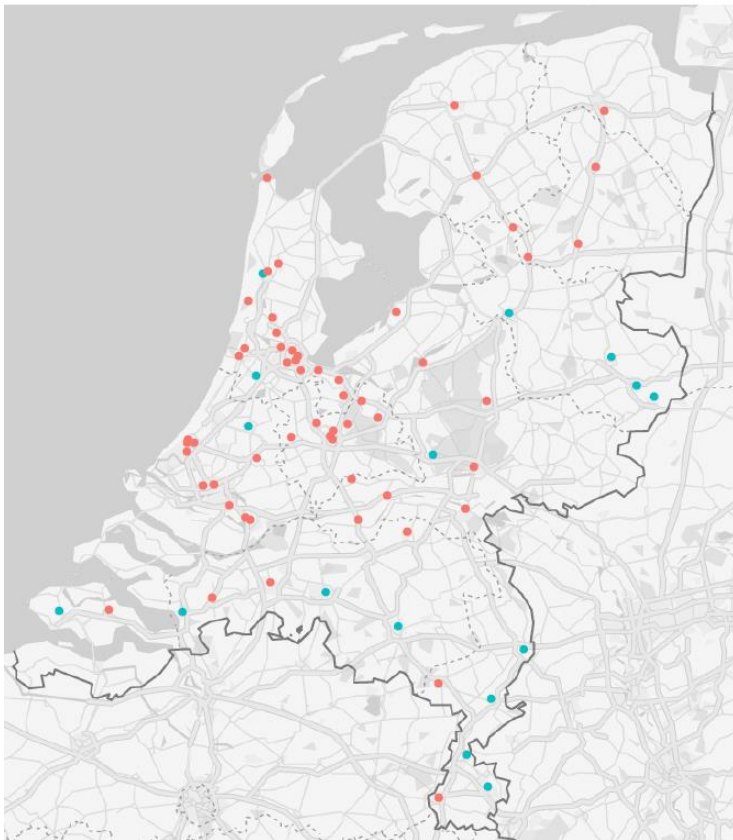
Gemiddelde correctiefactor (CF) per cluster



Figuur 5: Twee clusters met gemiddelde correctiefactorpatronen voor fiets/in-uitstapper

Indeling stations in clusters

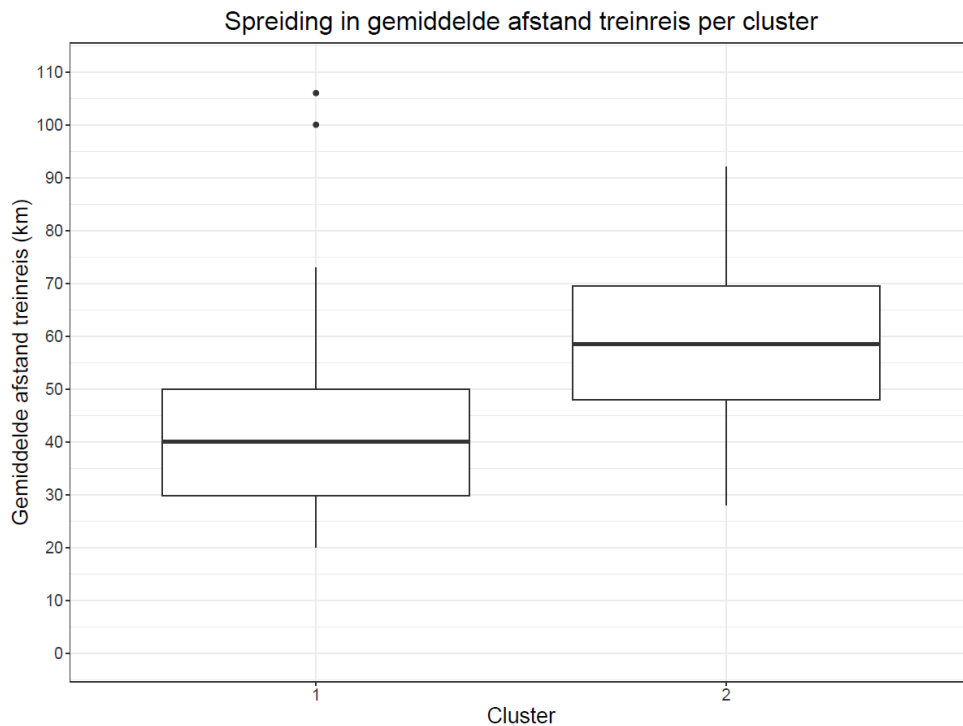
cluster • 1 • 2



Figuur 6: stations per correctiefactorcluster

Figuur 6 geeft de geografische spreiding van stations per cluster weer. Hierin zien we dat de stations in cluster 2 over het algemeen buiten de randstad liggen. Er zijn wat uitzonderingen aan te merken (zoals Alkmaar, Alpen aan den Rijn en Hoofddorp in cluster 2 en Maastricht en Goes in cluster 1).

Gegeven deze geografische spreiding tussen cluster 1 en 2 lijkt het waarschijnlijk dat de gemiddelde reisafstand een variabele is die kan verklaren of een station in cluster 1 of cluster 2 valt. In Figuur 7 is in boxplots de spreiding in reisafstand per trein in cluster 1 en cluster 2 weergegeven. Voor cluster 1 ligt de mediaan van de gemiddelde reisafstand over de stations op 40 kilometer, terwijl dit in cluster 2 op 58 kilometer ligt. Het verschil tussen de twee clusters en de gemiddelde reisafstanden per station is vrij groot. Het derde kwartiel van cluster 1 en het eerste kwartiel van cluster 2 liggen allebei rond de 50 kilometer. Dit betekent dat wanneer de gemiddelde reisafstand met de trein voor een specifiek station boven de 50 kilometer is, dit station 75% kans heeft dat deze in cluster 2 valt, bij minder dan 50 kilometer is de kans 75% dat deze in cluster 1 valt. Naarmate de gemiddelde reisafstand vanaf een treinstation groter wordt vertrekt de fietsende treinreiziger vroeger.



Figuur 7: clusters versus gemiddelde afstand treinreis

De gemiddelde reisafstand is een van de variabelen die we kunnen gebruiken om stations waarvoor we geen fietsparkeerdata beschikbaar hebben te kunnen classificeren. Een lage reisafstand van de trein duidt op cluster 1, een hogere reisafstand op cluster 2. Voor een station waarvoor nog geen fietsparkeerdata beschikbaar is, hebben we wel informatie beschikbaar over de gemiddelde reisafstand per trein. Op basis hiervan kunnen we dit station in het meest waarschijnlijke cluster indelen en de gemiddelde correctiefactor (CF) van dit cluster toepassen. Zo kunnen we een goede inschatting maken van FTN voor deze stations, zonder dat de specifieke correctiefactor bekend is.

6. Discussie

In dit hoofdstuk reflecteren we op een aantal aannames en simplificaties op de werkelijkheid die invloed zouden kunnen hebben op het resultaat. De belangrijkste aanname betreft de representativiteit van de data, zowel tussen treinstations als binnen treinstations.

Op een treinstation zijn vaak meerdere stallingsmogelijkheden voor fietsers aanwezig. Bij een treinstation met een bewaakte fietsenstalling is in veel gevallen ook nog een mogelijkheid tot onbewaakte stallen (bijvoorbeeld maaiveldstallingen). Omdat we alleen data beschikbaar hebben van de bewaakte stallingen is de aanname dat het patroon over de dag voor een specifiek station voor alle fietsende treinreizigers op dit station gelijk is. In werkelijkheid kan het zo zijn dat de bewaakte fietsenstallingen op andere uren gebruikt worden dan de onbewaakte stallingen. Bijvoorbeeld omdat het ene type stalling vol is en reizigers vanaf dan gebruik maken van het andere type stalling.

Daarnaast hebben we van een groot deel van de treinstations helemaal geen data van fietsparkeerders beschikbaar. Ondanks dat de gebruikte dataset met 72 stations relatief groot is, kan niet met zekerheid gezegd worden dat deze representatief is voor heel Nederland. Wanneer andere stations toegevoegd worden kan het landelijk beeld veranderen en zouden ook de gedefinieerde clusters anders uit kunnen vallen. Tegelijkertijd beslaat de dataset wel zo een groot aantal stations, met daarbinnen voldoende variatie in type stations (grootte, ligging, etc.), dat de verwachting niet is dat de conclusies van dit onderzoek wezenlijk zullen veranderen.

Een simplificatie die we in dit onderzoek hebben doorgevoerd is om geen onderscheid te maken tussen instappers en uitstappers en tussen fietsparkeerders die een fiets komen halen of brengen. Aangezien de fiets voornamelijk als voortransport wordt gebruikt ontstaat mogelijk een ander beeld wanneer voor- en natransport gescheiden zouden worden. Echter is uit de data niet eenduidig op te maken of een fietsparkering (of check-in/out van een treinreiziger) onderdeel is van het voortransport of het natransport. Een indicatie hiervan is het tijdstip: een fiets die 's ochtends gestald wordt is over het algemeen onderdeel van het voortransport en een fiets die 's ochtends gehaald wordt is over het algemeen onderdeel van het natransport. Dit is echter ook een aanname, en maakt het weer onmogelijk om naar het patroon van in-/uitstappers en fietsparkeerders over de hele dag te kijken. Van bijvoorbeeld een fiets die om 12 uur gehaald wordt is niet vast te stellen of deze onderdeel was van het voortransport of het natransport.

Dit onderzoek heeft aangetoond dat bij het vaststellen van de piekdrukke in de fietsenstalling een correctiefactor gebruikt moet worden om onderschatting van de piekdrukke te voorkomen. Voor bestaande stallingen met fietsparkeerdata kan de piekdrukke direct uit de fietsparkeerdata berekend worden. Voor bijvoorbeeld nog niet bestaande stallingen zullen correctiefactoren geschat moeten worden. Door middel van hiërarchische clustering hebben we twee groepen stations gedefinieerd waarbinnen de correctiefactoren op elkaar lijken. De gemiddelde reisafstand per trein blijkt een belangrijke verklarende variabele voor de indeling naar het cluster. Behalve reisafstand zullen er meer verklarende factoren zijn die de correctiefactoren bepalen. Meer onderzoek naar deze verklarende factoren kan bijdragen aan het nauwkeuriger schatten van correctiefactoren voor nog niet bestaande stallingen, of voor stallingen waarvoor geen data beschikbaar is.

7. Conclusies

Door het reispatroon over de dag van fietsparkeerders en in-/uitstappers met elkaar te vergelijken is zichtbaar geworden dat het aandeel fietsparkeerders binnen de treinreizigers niet uniform over de dag verdeeld is, maar sterk varieert. In zowel de ochtend- als de avondspits is het aandeel fietsparkeerders hoger dan gemiddeld over de dag, terwijl het aandeel fietsparkeerders in het dal juist lager ligt. Hiermee is de 'typische' fietsende treinreiziger een spitsreiziger.

Door een correctiefactor te introduceren voor deze ongelijke verdeling van fietsende treinreizigers over de dag is het mogelijk exacter piekdruktes in fietsenstallingen te bepalen. Wanneer geen rekening wordt gehouden met deze correctiefactor en het aandeel fietsende treinreizigers over de dag constant wordt verondersteld wordt de piekdrukke in fietsenstallingen sterk onderschat. Op specifieke stations kan de werkelijke piekdrukke tot wel 60% hoger liggen.

Ondanks dat er landelijk een duidelijk trend zichtbaar is in het aandeel fietsparkeerders per uur, ten opzichte van het totaal in-/uitstappers, zijn er grote verschillen tussen stations te zien. Wel is op bijna alle stations het aandeel fietsparkeerders in de spits hoger dan gemiddeld over de dag. Op basis van een clusteranalyse hebben we de 72 stations uit onze dataset kunnen verdelen in 2 clusters met min of meer gelijksoortige stations. Hier lijkt een verband te bestaan tussen de correctiefactoren per uur en de gemiddelde reisafstand die op een station per trein wordt afgelegd. Fietsende treinreizigers met een grotere reisafstand per trein lijken eerder naar het station te komen. Op basis van de gemiddelde reisafstand van treinreizigers kan daarmee ook een goede inschatting gemaakt worden in welk cluster een station valt waarvan nog geen fietsparkeerdata beschikbaar is.

Ook generiek zien we vroeger drukte in de fietsenstalling dan op het treinstation. Bij het merendeel van de stations komt de piekdrukke in de fietsenstalling voor tussen 07:00 – 08:00 terwijl de piekdrukke in het station over het algemeen tussen 08:00 – 09:00 voorkomt.

Op basis van bovenstaande inzichten zijn we in staat om het verloop van de drukte in fietsenstallingen waarvoor nog geen data beschikbaar is (of stallingen die nog gebouwd moeten worden) nauwkeuriger in te schatten. Een inschatting enkel op basis van het verloop van in-/uitstappers over de dag leidt tot grote onderschattingen van de piekdruktes in de fietsenstallingen.

Referenties

- Jonkeren, O., Harms, L., Jorritsma, P., Huibregtse, O., Bakker, P., 2018. Waar zouden we zijn zonder de fiets en de trein?
- Jonkeren, O., Kager, R., Harms, L., Te Brömmelstroet, M., 2021. The bicycle-train travellers in the Netherlands: personal profiles and travel choices. *Transportation (Amst)*. 48, 455–476.
- Shelat, S., Huisman, R., van Oort, N., 2018. Analysing the trip and user characteristics of the combined bicycle and transit mode. *Res. Transp. Econ.* 69, 68–76.
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B (Statistical Methodol)*. 63, 411–423.