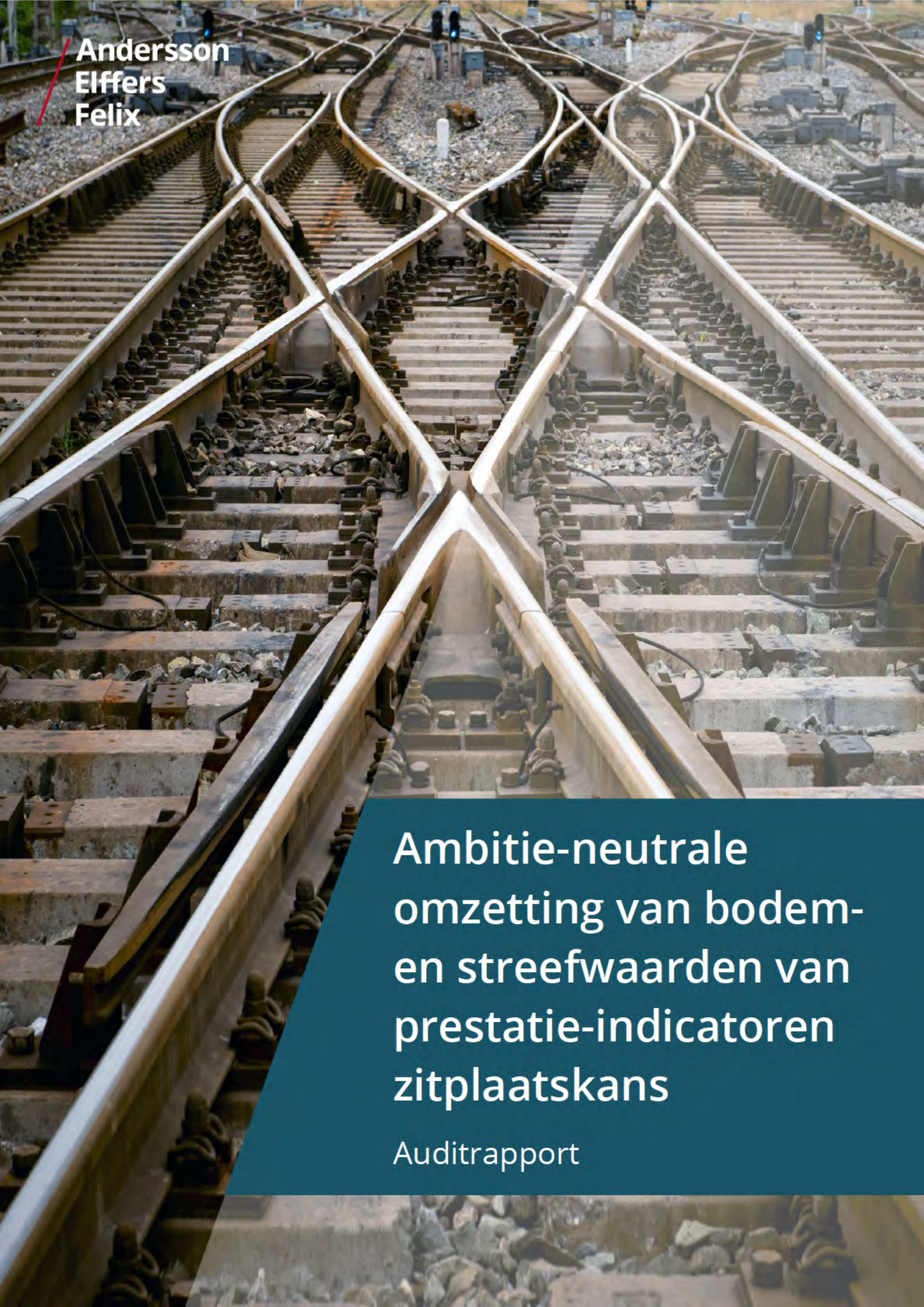


Andersson
Elffers
Felix

Ambitie-neutrale omzetting van bodem- en streefwaarden van prestatie-indicatoren zitplaatskans

Auditrapport

/Andersson Elffers Felix

Datum

23 november 2023

Opdrachtgever

Ministerie van Infrastructuur en Waterstaat (IenW)

Onderzoeksteam

Bescherming persoonlijke levenssfeer

Referentie

GI203/Eindrapportage

bezoekadres Maliebaan 16 3581 CN Utrecht
postadres Postbus 85198 3508 AD Utrecht
telefoon +31 30 236 30 30
website www.aef.nl

Managementsamenvatting

Om het aantal reizigers in treinen te bepalen, maakt NS momenteel gebruik van een systeem genaamd ROCKT, in combinatie met het zogeheten fractiemodel. Vanaf 2024 vervangt NS deze systemen door een nieuw systeem genaamd FLUX. De reizigersaantallen worden onder andere gebruikt bij het berekenen van een aantal prestatie-indicatoren (PI's) rondom de zitplaatskans en het aantal drukke treinen. Met het ministerie van Infrastructuur en Waterstaat (IenW) heeft NS voor elke PI een bodem- en streefwaarde afgesproken. De bodemwaarde is de waarde die NS voor de betreffende PI elk jaar minimaal dient te realiseren. De streefwaarde is de waarde die nagestreefd wordt voor het eind van de concessie (of tijdens de midterm-review).

Doordat het systeem om reizigersaantallen in kaart te brengen verandert, kunnen ook de waarden van de PI's waarvoor deze reizigersaantallen worden gebruikt veranderen. Omdat de bodem- en streefwaarden een bepaald ambitieniveau representeren, moeten deze mee veranderen met de verandering van de PI's. Deze verandering dient dusdanig te worden gedaan dat het ambitieniveau dat de bodem- en streefwaarden representeren, gelijk blijft. Met de overstap van ROCKT en het fractiemodel naar FLUX heeft NS een voorstel gedaan voor de aanpassing van de bodem- en streefwaarden. In dit rapport geven we ons oordeel over de mate waarin dit voorstel ambitie-neutraal te noemen is.

Voor deze audit maken we gebruik van een beoordelingskader met vijf criteria om het voorstel van NS met betrekking tot de omzetting van de bodem- en streefwaarden te beoordelen:

1. **Methodische correctheid.** Is de gekozen methodiek inhoudelijk correct? Sluit deze aan bij de daadwerkelijke verandering van de PI's.
2. **Optimale methodiek.** Is de gekozen methodiek beter dan beschikbare alternatieven?
3. **Objectiviteit methodiek.** Is de gekozen methodiek objectief? Heeft NS invloed gehad op de uitkomsten van de omzetting van bodem- en streefwaarden?
4. **Behoud van benodigde inspanning.** Blijft de inspanning die NS moet leveren om de bodem- en streefwaarden gelijk?
5. **Correcte toepassing.** Is de gekozen methodiek correct uitgevoerd?

Om ons oordeel op deze criteria te kunnen geven, hebben we gebruik gemaakt van verschillende bronnen. Hiervoor heeft NS onder andere de onderliggende waarden van de PI's op maandbasis (via zowel ROCKT en het fractiemodel als via FLUX) beschikbaar gesteld. Met behulp van deze gegevens hebben we de methodiek van NS kunnen valideren.

Onze conclusie is dat de gekozen methodiek de best passende methodiek is voor de ambitie-neutrale omzetting van bodem- en streefwaarden. Deze methodiek is op drie keuzes na volledig objectief. Van de drie keuzes heeft er één geen invloed op de uitkomsten. De andere twee keuzes zijn door NS vanuit inhoudelijk logische argumenten gemaakt. Hoewel deze keuzes een kleine invloed op de uitkomsten hebben, is met aan zekerheid grenzende waarschijnlijkheid te stellen dat een andere partij dezelfde keuzes zou hebben gemaakt. Tot slot is de beschreven methodiek cijfermatig correct uitgevoerd. Dit alles leidt ons tot de conclusie dat de omzetting van de bodem- en streefwaarden zo ambitieneutraal als mogelijk is gedaan.

/ Inhoud

/ Managementsamenvatting 2

/ 1 Inleiding 4
1.1 Aanleiding 4
1.2 Vraagstelling en afbakening 4
1.3 Onderzoeksaanpak 5
1.4 Leeswijzer 5

/ 2 De overgang naar FLUX 6
2.1 Het gebruik van reizigersaantallen 6
2.2 Reizigersstromen via ROCKT en het Fractiemodel 6
2.3 Reizigersstromen via FLUX 7
2.4 Mogelijke methoden voor omzetting van bodem- en streefwaarden 7
2.5 Voorstel van NS voor omzetting van bodem- en streefwaarden 10

/ 3 Beoordeling van de omzetting van bodem- en streefwaarden 12
3.1 Beoordelingskader 12
3.2 Oordeel per criterium 13
3.3 Conclusie 18

/ A Definities van PI's 19

/ B Aannames lineaire regressieanalyse 20

/ 1 Inleiding

1.1 Aanleiding

De Nederlandse Spoorwegen (NS) voert de treindiensten op het hoofdrailnet uit in opdracht van het ministerie van Infrastructuur en Waterstaat (IenW), dat hier door middel van een vervoerconcessie op stuurt. In deze concessie is een aantal prestatie-indicatoren (PI's) vastgelegd als onderdeel van het sturingsinstrumentarium.[1] Voor elk van deze PI's zijn bodem- en streefwaarden opgenomen. De bodemwaarde van een PI representeert de minimale waarde die NS jaarlijks dient te realiseren. De streefwaarde is de waarde van de PI die NS aan het eind van de concessieperiode (of tijdens de midterm-review) dient te bereiken. Gedurende de concessieperiode zorgt NS voor meting van de verschillende PI's. Naast PI's wordt er ook gebruik gemaakt van zogenaamde informatie-indicatoren (II's). Ook dit zijn maatstaven, maar in tegenstelling tot de PI's representeren de II's niet-bindende prestaties, waarvoor ook geen bodem- en streefwaarden zijn afgesproken.

Voor het berekenen van een aantal PI's en II's is het noodzakelijk om gebruik te maken van metingen van aantallen reizigers in treinen. Momenteel wordt dit gedaan via het toedelingsmodel ROCKT (Reizigers OV Chipkaart Koppelen aan Treinen), in combinatie met het fractiemodel.[2] Deze bron maakt voornamelijk gebruik van de check-in-check-out gegevens van reizigers met een OV-chipkaart. Voor ROCKT is het niet mogelijk om nieuwe betaalmethodes (zoals bankpassen) in kaart te brengen. Daarnaast is het systeem verouderd en is het afhankelijk van tellingen in de trein. Om deze redenen stapt NS per 1 januari 2024 over naar een nieuw toedelingsmodel, genaamd FLUX. Een dergelijke verandering van bron, betekent ook dat de waarden van de betreffende PI's en II's kunnen veranderen. De bodem- en streefwaarden die tussen NS en het ministerie van IenW zijn afgesproken representeren een bepaald ambitieniveau. Als de waarden van de PI's veranderen, is het van belang dat de waarden van de bodem- en streefwaarden dusdanig veranderen dat dit afgesproken ambitieniveau gelijk blijft.

1.2 Vraagstelling en afbakening

Op basis van de overstap van ROCKT naar FLUX, heeft NS een voorstel gedaan voor de ambitie-neutrale omzetting van de betreffende PI's en bijbehorende bodem- en streefwaarden. Om de kwaliteit van de PI's en II's en de (ambitie-neutrale omzetting van) bodem- en streefwaarden te borgen, dient hier een externe audit op uit te worden gevoerd. Het ministerie van IenW heeft AEF gevraagd om deze audit uit te voeren.

Om een goed voorstel te kunnen doen, heeft NS ter voorbereiding aan CQM gevraagd om haar te ondersteunen. CQM heeft in kaart gebracht hoe bepaald kan worden of de overstap naar FLUX een verbetering is, en welke methoden er zijn om de omzetting van bodem- en streefwaarden te doen. Deze methoden kennen verschillende voor- en nadelen die uiteindelijk door NS gewogen dienen te

[1] De prestatie-indicatoren (PI's) worden ook vaak 'kritische prestatie-indicatoren' (KPI's) genoemd. Omdat in de vervoerconcessie wordt gesproken over PI's, hebben we ervoor gekozen deze term ook in dit rapport aan te houden.
[2] Waar we in dit rapport spreken over 'ROCKT', bedoelen we altijd 'ROCKT en het fractiemodel'.

worden. Voor de audit van de omzetting hebben wij het rapport van CQM als uitgangspunt genomen. We hebben dus géén audit uitgevoerd naar de methoden die CQM in kaart heeft gebracht, maar wel naar de keuzes die NS heeft gemaakt, de argumenten die daarvoor zijn gegeven, en de mate waarin die keuzes aansluiten bij het advies van CQM.

1.3 Onderzoeksaanpak

Deze audit heeft plaatsgevonden in drie fasen. We zijn de eerste fase gestart met het objectief in beeld brengen van de verandering van ROCKT naar FLUX. Hierbij hebben we documentatie ontvangen van NS over de technische werking van beide systemen, de voor- en nadelen ervan, en de prestatie van FLUX ten opzichte van ROCKT en ten opzichte van de werkelijke metingen van de aantallen reizigers. In deze fase hebben we ook het onderzoeksrapport van CQM ter beschikking gekregen. Om te zorgen dat we de uitkomsten van dit rapport goed konden duiden, hebben we hier ook een toelichting op gehad van CQM.

In de tweede fase hebben we een beoordelingskader opgesteld om op een objectieve wijze te kunnen toetsen of het voorstel van NS ambitieneutraal is. Dit beoordelingskader hebben we getoetst bij zowel het ministerie van IenW als bij NS, zodat vooraf duidelijk was op welke criteria beoordeeld zou worden. Er zijn geen noemenswaardige wijzingen gedaan aan het beoordelingskader naar aanleiding van dit overleg.

In de derde en laatste fase van het onderzoek hebben we de door NS voorgestelde methodiek beoordeeld aan de hand van het beoordelingskader. Om dit goed te kunnen doen heeft NS alle benodigde gegevens ter beschikking gesteld. Op basis van deze gegevens hebben we eigen analyses uitgevoerd om uiteindelijk op alle criteria uit het beoordelingskader een oordeel te kunnen geven.

1.4 Leeswijzer

De rest van dit rapport is als volgt opgebouwd:

- In **hoofdstuk 2** beschrijven we de achtergrond van het vraagstuk. We gaan in op het gebruik van reizigersaantallen door NS, de manier waarop deze momenteel worden vastgesteld (ROCKT), en het systeem (FLUX) wat hier in de toekomst voor gebruikt zal worden. Hierna beschrijven we kort welke manieren CQM in kaart heeft gebracht om bodem- en streefwaarden van PI's van het ene naar het andere systeem te vertalen. Tot slot gaan we in op de methodiek die NS voorstelt om te gebruiken. In dit hoofdstuk worden alleen objectieve bevindingen gedaan en is er nog geen sprake van een beoordeling.
- In **hoofdstuk 3** starten we met het presenteren van ons beoordelingskader. Hierna gaan we op elk criterium individueel in. We starten steeds met onze bevindingen over de door NS beschreven methodiek, of vanuit onze eigen analyses. Elk criterium eindigen we met ons oordeel. We eindigen het hoofdstuk met de overkoepelende conclusie over de omzetting van bodem- en streefwaarden.
- In **bijlage A** hebben we de definities opgenomen van de drie PI's waar deze audit betrekking op heeft. In **bijlage B** geven we een technische uiteenzetting over het gebruik van de methode die NS voorstelt te gebruiken. Deze bijlage is ondersteunend bij ons oordeel over het tweede criterium uit het beoordelingskader.

/ 2 Overgang naar FLUX

2.1 Gebruik van reizigersaantallen

Goed zicht hebben op het aantal reizigers per traject op elk verschillend moment is voor NS om meerdere redenen belangrijk. Deze informatie is bijvoorbeeld noodzakelijk voor het optimaal inzetten van het beschikbare materieel. Naast het feit dat de reizigersstromen van belang zijn voor de bedrijfsvoering van NS, worden ze ook gebruikt voor de monitoring en verantwoording richting het ministerie van IenW. In de vervoersconcessie zijn verschillende PI's vastgelegd waarvan NS de waarde dient te rapporteren. Voor het berekenen van drie van deze PI's zijn de reizigersstromen nodig als input. Het gaat hierbij om de volgende PI's:

- Zitplaatskans in de spits op het hoofdrailnetwerk (HRN)
- Zitplaatskans in de spits op de hogesnelheidslijn (HSL)
- Aantal drukke treinen per werkweek in de spits op het HRN

Naast deze PI's, zijn er ook een aantal II's waarvoor de reizigersstromen nodig zijn. Deze II's worden door NS gerapporteerd aan het ministerie, maar hiervoor zijn geen waarden afgesproken waarop NS afgerekend of beloond kan worden bij het al dan niet behalen ervan. De betreffende II's zijn:

- Zitplaatskans in het dal op het HRN
- Zitplaatskans in het dal op de HSL
- Bezettingsgraad in de spits
- Reizigerskilometers in de spits
- Reizigerskilometers in het dal

In bijlage A is voor de volledigheid een definitie van deze PI's opgenomen.

2.2 Reizigersstromen via ROCKT en Fractiemodel

Voor de berekeningen van de eerder genoemde PI's en II's maakt NS gebruik van een model genaamd SOFA. De input van dit model bestaat (onder andere) uit de gegevens over de gerealiseerde reizigersstromen. Momenteel worden deze reizigersstromen bepaald met behulp van het systeem ROCKT.

Het systeem ROCKT deelt reizigers toe aan specifieke treinen. Hiervoor gebruikt het systeem alleen de gegevens van reizigers die met een OV-chipkaart zowel in- als uitchecken. ROCKT neemt dus geen reizigers mee die niet met een OV-chipkaart reizen, en ook geen reizigers die wel met een OV-chipkaart reizen, maar niet in- of uitchecken. Vóór de coronapandemie betekende dit dat ROCKT ongeveer 80% tot 85% van alle reizigers in kaart bracht. In de ochtendspits betrof het ongeveer 95% van alle reizigers.

Voor het berekenen van de PI's en II's met behulp van SOFA is het echter van belang om alle reizigers mee te nemen. Om te corrigeren voor de reizigers die niet via ROCKT worden geteld, wordt gebruik gemaakt van het zogenaamde Fractiemodel. Het Fractiemodel hoogt het aantal reizigers op elk traject (op elk moment) op met een bepaald percentage. Deze percentages zijn in principe verschillend, en zorgen ervoor dat de resulterende reizigersstromen 100% van de reizigers

meenemen. De percentages uit het Fractiemodel worden bepaald met behulp van de resultaten uit het veldwerk (de steekproeven waarbij werkelijke reizigers door IPSOS worden geteld).

2.3 Reizigersstromen via FLUX

Vanaf 1 januari 2024 zal het systeem FLUX worden ingevoerd ter vervanging van ROCKT. Naast het feit dat ROCKT volgens NS een verouderd systeem is en afhankelijk is van tellingen in de trein, maakt het ook alleen gebruik van de gegevens van reizigers die zowel in- als uitchecken met een OV-chipkaart. FLUX is niet afhankelijk van de tellingen in de trein, en kan ook andere betaalmethoden meenemen. Dit betreft zowel bestaande methoden (zoals bijvoorbeeld bankpassen), maar ook eventuele toekomstige betaalmethoden kunnen in FLUX worden geïncorporeerd.

Ten opzichte van het huidige systeem ROCKT kent FLUX een aantal voordelen:

- Alle betaalmethoden kunnen in principe meegenomen worden.
- Vanwege het gebruik van aanvullende datastromen is veldwerk niet langer noodzakelijk voor een betrouwbare bepaling van reizigersstromen.
- Doordat FLUX gebruikt maakt van kunstmatige intelligentie, leert het systeem van het reisgedrag van reizigers, waardoor de toedeling aan routes beter bij de werkelijkheid aansluit.
- FLUX kan werken met onvolledige reisgegevens, bijvoorbeeld gegevens van reizigers die met een OV-chipkaart wel inchecken, maar niet uitchecken.
- In de toekomst kunnen eventuele nieuwe betaalmethoden toegevoegd worden aan FLUX.

In de afgelopen maanden heeft NS onderzocht of de resultaten van FLUX dichter bij de werkelijkheid liggen dan de resultaten van ROCKT.[3] Uit dit onderzoek is gebleken dat FLUX gemiddeld beter aansluit bij de tellingen (het veldwerk) dan ROCKT. Gedurende de spits zijn de resultaten van beide methoden vergelijkbaar. In het weekend en in de dalperioden is het resultaat van FLUX duidelijk beter. Voor sprinters en de HSL komen de uitkomsten van FLUX ongeveer even dicht bij de werkelijke tellingen als de uitkomsten van ROCKT. Voor intercity's scoort FLUX echter veel beter. Daarnaast blijkt uit dit onderzoek ook dat de resultaten van FLUX stabieler zijn.

In vergelijkingen met de werkelijk getelde reizigers in september en oktober 2022 bleek dat FLUX in 45% van de gevallen minder dan 10 reizigers afwijkt, en in 90% van de gevallen minder dan 50 reizigers. FLUX ligt vooral dicht bij de werkelijke tellingen als het gaat om treinen met minstens 150 reizigers. Treinen met minder dan 150 reizigers zijn echter niet van belang voor de zitplaatskans PI's. Dit komt doordat het kleinste materieel 150 zitplaatsen heeft, en iedereen dus met zekerheid heeft kunnen zitten.

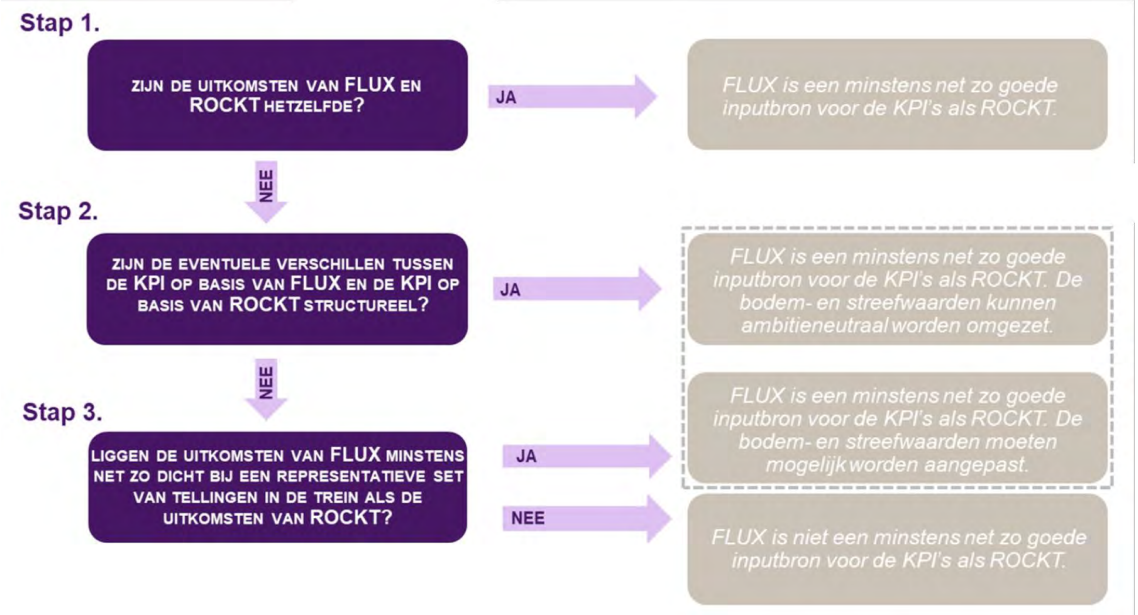
2.4 Mogelijke methoden voor omzetting van bodem- en streefwaarden

Met de overstap van ROCKT naar FLUX verandert de input die wordt gebruikt voor het berekenen van de betreffende PI's en II's. Hoewel de rekenmethode van de PI's en II's zelf niet verandert, betekent een verandering van input in principe ook dat de output van de berekening verandert. De overstap

[3] De informatie in deze sectie is overgenomen uit het bedrijfsvertrouwelijke rapport *Validatie FLUX* (NS, 2023).

van ROCKT naar FLUX zorgt er dus voor dat de waarden van PI's en II's veranderen. De bodem- en streefwaarden van de PI's reflecteren een bepaald ambitieniveau. Als de waarden van de PI's veranderen door de overstap naar FLUX, betekent dit dus ook dat de bodem- en streefwaarden moeten worden aangepast zodat het ambitieniveau gelijk blijft. NS heeft CQM gevraagd om haar te ondersteunen bij het in kaart brengen van de mogelijkheden voor deze aanpassingen.

Om te kunnen bepalen welke methodiek gebruikt kan worden voor de omzetting van bodem- en streefwaarden, heeft CQM eerst in kaart gebracht hoe onderzocht kan worden of de uitkomsten van FLUX minstens net zo goed zijn als de uitkomsten van ROCKT voor het berekenen van de zitplaatskans-PI's.[4] Voor deze opgave is een stappenplan ontwikkeld, waarbij niet alleen in kaart wordt gebracht óf FLUX minstens net zo goed is als ROCKT, maar ook welke implicaties dat heeft voor de omzetting van bodem- en streefwaarden. Onderstaand is het stappenplan uit de rapportage van CQM opgenomen.



Figuur 1. Het stappenplan dat CQM heeft ontwikkeld om te bepalen of de uitkomsten van FLUX minimaal even goed zijn als de uitkomsten van ROCKT. Deze figuur is overgenomen uit de rapportage van CQM.

Omdat de verschillende stappen van belang zijn voor het bepalen van de methodiek waarmee de bodem- en streefwaarden kunnen worden omgezet, beschrijven we ze onderstaand in meer detail.

Stap 1

In de eerste stap wordt simpelweg gecontroleerd of de uitkomsten exact gelijk zijn. Als dit zo is, is FLUX logischerwijs een minstens net zo goede bron voor de PI's als ROCKT.

Stap 2

In deze stap worden de PI's o.b.v. ROCKT enerzijds en FLUX anderzijds met elkaar vergelijken op maandelijks niveau. Dit wordt voor een geheel jaar gedaan, zodat jaarlijkse patronen worden meegenomen. Met behulp van een lineaire regressie wordt onderzocht of er sprake is van een lineair verband tussen beide sets van PI-uitkomsten. Het resultaat van deze lineaire regressie is een formule

[4] De teksten uit deze sectie zijn gebaseerd op het rapport *Een nieuw model voor reizigersaantallen: Voorbereiding op de audit op zitplaats-KPI's* (CQM, 2023).

waarmee de PI-waarden o.b.v. FLUX (voor elke maand) zo goed mogelijk uitgedrukt worden in de PI-waarden o.b.v. ROCKT. Deze formule ziet er als volgt uit:

$$PI_{FLUX} = a + b \times PI_{ROCKT} + e$$

Hierin zijn PI_{FLUX} en PI_{ROCKT} de waarden van de PI op basis van respectievelijk FLUX en ROCKT. Verder zijn *a* en *b* constante getallen, die voor elke maand hetzelfde zijn. Met behulp van deze waarden kan de waarde PI_{FLUX} zo goed mogelijk uitgedrukt worden in PI_{ROCKT}. De *e* in de formule is een foutterm. Deze term reflecteert de mate waarin de lineaire formule er níet in slaagt om de relatie tussen PI_{FLUX} en PI_{ROCKT} (per maand) te weerspiegelen.

Om te bepalen of de gevonden lineaire formule gebruikt kan worden voor de omzetting van bodem- en streefwaarden, moet worden bepaald of deze relatie tussen PI-waarden sterk genoeg is. Hiervoor moet de standaardafwijking op jaarniveau klein zijn. Hierbij is de vraag wanneer dit klein genoeg wordt bevonden. Hoewel dit oordeel uiteindelijk aan NS wordt gelaten, geeft CQM wel aan dat hiervoor de impactschaal kan worden gebruikt die NS ook hanteert bij het vaststellen van de impact van nieuwe ontwikkelingen op verschillende PI's.

Stap 3

Als blijkt dat de foutterm uit stap 2 niet klein te noemen is, kan gesteld worden dat er geen structureel verschil bestaat. Als dat het geval is, gaat men door naar deze derde stap. Hierin worden de reizigersaantallen o.b.v. FLUX en die o.b.v. ROCKT vergeleken met tellingen vanuit het veldwerk. Als de afwijkingen van FLUX kleiner zijn dan die van ROCKT, kan worden geconcludeerd dat FLUX dichter bij de werkelijkheid ligt.

Conclusies voor de bodem- en streefwaarden

De methodiek voor de omzetting van bodem- en streefwaarden die CQM voorstelt, is afhankelijk van de uitkomsten van stap 2 uit het hierboven beschreven stappenplan. Als blijkt dat de foutterm klein genoeg is, kunnen de bodem- en streefwaarden worden omgezet met de lineaire formule die is vastgesteld als relaties tussen de PI's op basis van ROCKT en op basis van FLUX. Specifiek zou dit betekenen dat de volgende formules worden gebruikt.

$$bodemwaarde_{FLUX} = a + b \times bodemwaarde_{ROCKT}$$
$$streefwaarde_{FLUX} = a + b \times streefwaarde_{ROCKT}$$

Waarbij de termen *a* en *b* zijn bepaald via de lineaire regressieanalyse op basis van de maandelijkse PI-resultaten.

Als echter blijkt dat de foutterm *e* uit stap 2 te groot wordt bevonden, dan betekent dit dat de verschillen tussen de PI's op basis van FLUX en op basis van ROCKT niet structureel zijn. In dat geval moet een andere methodiek worden gebruikt voor de omzetting van de bodem- en streefwaarden. Hiervoor geeft CQM een aantal methoden die overwogen kunnen worden:

- **Absolute omzetting.** Hierbij wordt het absolute verschil tussen de PI-waarden over een volledig jaar bepaald. Dit absolute verschil wordt vervolgens ook toegepast op de bodem- en streefwaarden.
- **Relatieve omzetting.** Hierbij wordt het relatieve verschil tussen de PI-waarden over een volledig jaar bepaald. Dit absolute verschil wordt vervolgens ook toegepast op de bodem- en streefwaarden.

- **Omzetting via regressie.** Hierbij wordt de eerder beschreven methodiek o.b.v. de lineaire regressie gebruikt, ondanks de grote foutterm.
- **Omzetting o.b.v. inzicht.** Als de inhoudelijke verschillen tussen de twee modellen inzicht geven in de mogelijke verbanden met de verschillen in PI-waarden, kan dit gebruikt worden voor de omzetting van bodem- en streefwaarden.

2.5 Voorstel van NS voor omzetting van bodem- en streefwaarden

Op basis van het onderzoek dat is uitgevoerd door CQM heeft NS een voorstel gedaan voor de omzetting van bodem- en streefwaarden van de betreffende PI's. Hiervoor heeft NS het stappenplan uit de vorige sectie gevolgd. In stap 1 heeft NS gecontroleerd of de uitkomsten van ROCKT en FLUX hetzelfde zijn. Dit is niet het geval, waardoor direct door wordt gegaan naar stap 2. Voor het doorlopen van deze stap zijn een aantal inhoudelijke keuzes gemaakt. Onderstaand beschrijven we om welke keuzes dit gaat.

Gebruikte dataset

In het advies van CQM is opgenomen dat er PI-waarden van minimaal 12 aaneengesloten maanden nodig zijn om stap 2 goed uit te kunnen voeren. De data vanuit FLUX is momenteel beschikbaar over de periode van juni 2022 t/m september 2023 (een totaal van 16 maanden). NS heeft onderzocht of de gekozen 12 maanden (vanuit deze set van 16 maanden) invloed heeft op de resultaten en geconcludeerd dat dit niet het geval is. Op basis hiervan is de keuze gemaakt om de periode van oktober 2022 t/m september 2023 te gebruiken als dataset. NS geeft aan dat deze keuze te maken heeft met het feit dat er nog 'na-ijlende corona-effecten' te zien zijn in het reizigersgedrag in de maanden juni t/m september 2022.

Weging van de PI-scores per maand

De verschillende PI's worden elke maand gemeten. Omdat er verschil zit tussen de maanden (bijv. in het aantal reizigers, of het aantal werkdagen), worden de maandelijkse resultaten ten opzichte van elkaar gewogen. Deze weging wordt gebruikt om een jaartotaal te bepalen, maar ook bij de lineaire regressie om de metingen per maand onderling vergelijkbaar te maken. In het advies van CQM is deze weging van de resultaten per maand ook opgenomen. Voor de PI 'Aantal drukke treinen per werkweek in de spits op het HRN' gebruikt NS het aantal werkdagen per maand als gewicht voor het PI-resultaat van de betreffende maand. Voor de PI's 'Zitplaatskans in de spits op het HRN' en 'Zitplaatskans in de spits op de HSL' gebruikt NS het aantal reizigers als gewicht voor het resultaat per maand. Dit aantal reizigers kan worden bepaald met behulp van ROCKT of FLUX. De keuze die NS hier heeft gemaakt is dit aantal reizigers te bepalen op basis van FLUX. Hierbij geeft NS aan dat de resultaten gelijk zouden zijn als hier gekozen zou worden voor ROCKT in plaats van FLUX.

Beoordeling van de grootte van de foutterm uit de lineaire regressie

Op basis van de vastgestelde dataset heeft NS een lineaire regressie uitgevoerd om te komen tot een formule waarmee de PI-resultaten van ROCKT vertaald kunnen worden in de PI-resultaten van FLUX. Zoals beschreven in de vorige sectie dient vervolgens te worden beoordeeld of de standaardafwijking op jaarniveau die bij deze formule hoort 'klein' te noemen is. Om de ordegrootte van deze standaardafwijking te beoordelen, maakt NS gebruik van de impactschaal die voor elke PI wordt gebruikt om de impact van ontwikkelingen te bepalen. Deze schaal kent vier ordegroottes: geen,

klein, middel, en groot. NS beoordeelt de standaardafwijking op jaarniveau als 'klein genoeg', als deze op basis van de impactschaal de ordegrootte 'geen' of 'klein' toegewezen zou krijgen.

Voor de drie betreffende PI's komt NS tot de conclusie dat de standaardafwijking klein genoeg is. Dit betekent volgens de methodiek die CQM ontwikkelde dat de gevonden lineaire relatie tussen de PI-waarden o.b.v. FLUX en die o.b.v. ROCKT gebruikt kan worden voor de omzetting van de bodem- en streefwaarden. NS heeft deze voorgestelde methodiek gevolgd om te komen tot een voorstel voor de omzetting van bodem- en streefwaarden. In onderstaande tabel zijn voor de drie betreffende PI's de huidige bodem- en streefwaarden (op basis van ROCKT) en de nieuw voorgestelde bodem- en streefwaarden opgenomen.

PI	BW ROCKT	BW FLUX	SW ROCKT	SW FLUX
Zitplaatskans in de spits HRN	94,3%	94,1%	95,5%	95,5%
Zitplaatskans in de spits HSL	91,2%	92,7%	94,9%	95,8%
Aantal drukke treinen per werkweek in de spits HRN	110	109	100	98

/ 3 Beoordeling van omzetting van bodem- en streefwaarden

3.1 Beoordelingskader

Om te kunnen beoordelen of de omzetting van de bodem- en streefwaarden ambitieneutraal is gedaan, maken we gebruik van een beoordelingskader. Dit kader richt zich op vijf criteria om de omzetting van de bodem- en streefwaarden te beoordelen. De criteria zijn (deels) volgordelijk. Dit houdt in dat een positief oordeel over eerdere factoren noodzakelijk is voor een positief oordeel over volgende factoren. Om uiteindelijk te kunnen concluderen dat de omzetting ambitieneutraal is uitgevoerd, moet er sprake zijn van een positieve beoordeling op alle factoren.

Met de overgang van ROCKT naar FLUX verandert de input die wordt gebruikt door SOFA om de waarden van de betreffende PI's en II's te berekenen. Doordat er een andere input-bron wordt gebruikt, kan ook de output (de waarden van de PI's en II's) veranderen. De rekenmethodiek zelf blijft echter hetzelfde. Dit betekent dat er voor de PI's en II's eigenlijk geen sprake is van een omzetting die beoordeeld kan worden. De waarden van de PI's en II's veranderen simpelweg door het veranderen van input-bron. De bodem- en streefwaarden die bij de PI's horen moeten echter wel worden aangepast om bij de nieuwe waarden van de PI's aan te sluiten. Het onderstaande beoordelingskader is daarom ook gericht op de omzetting van de bodem- en streefwaarden, en niet zozeer op de verandering van PI's en II's. Er wordt met dit kader dus feitelijk beoordeeld of de omzetting van bodem- en streefwaarden aansluit bij de verandering van de PI's.

	Criterium	Beschrijving
1	**Methodische correctheid**	Is de methodiek die is gekozen voor het omzetten van de bodem- en streefwaarden inhoudelijk correct? Met andere woorden, sluit de gekozen methodiek aan bij de daadwerkelijke verandering van PI-waarden door de overstap van ROCKT naar FLUX als input voor SOFA?
2	**Optimale methodiek**	Is de gekozen methodiek beter dan eventuele alternatieve methoden die beschikbaar zijn? Met 'beter' wordt in dit geval bedoelt dat er een grotere zekerheid is dat de nieuwe bodem- en streefwaarden inhoudelijk dezelfde betekenis hebben als de huidige waarden.
3	**Objectiviteit methodiek**	Is de gehanteerde methodiek voor het omzetten van bodem- en streefwaarden objectief? Met andere woorden, heeft NS buiten de keuze van de te hanteren methodiek geen invloed gehad op de omzetting van de waarden?
4	**Behoud van benodigde inspanning**	Blijft de haalbaarheid van de bodem- en streefwaarden behouden met de gekozen methodiek? Met andere woorden, blijft de inspanning die NS moet leveren om de bodem- en streefwaarden te behalen gelijk?
5	**Correcte toepassing**	Is de gekozen methodiek correct toegepast voor de omzetting van de PI's, II's en bodem- en streefwaarden? Met andere woorden, sluit de werkelijke omzetting aan bij de theoretische beschrijving van de gekozen methodiek?

3.2 Oordeel per criterium

Onderstaand beschrijven we voor elk van de vijf criteria uit hoe we het voorstel van NS voor de omzetting van de bodem- en streefwaarden beoordelen.

Methodische correctheid

De methodiek die NS heeft gekozen voor de omzetting van de bodem- en streefwaarden komt overeen met de methodiek die CQM heeft geadviseerd. Deze methodiek neemt de werkelijke metingen op basis van ROCKT en op basis van FLUX als uitgangspunt. Vervolgens wordt op een statistische wijze bepaald hoe via welke formule metingen op basis van ROCKT het beste kunnen worden vertaald naar de metingen op basis van FLUX. Tot slot wordt deze formule toegepast op de bodem- en streefwaarden op basis van ROCKT, om zo te komen tot de bodem- en streefwaarden op basis van FLUX.

De methode is statistisch van aard en gebruikt de werkelijke metingen op basis van ROCKT en FLUX om te komen tot een vertaling van de bodem- en streefwaarden. De methodiek voor de omzetting van bodem- en streefwaarden sluit dus inherent aan bij de daadwerkelijke verandering van PI-waarden door de overstap van ROCKT naar FLUX. **Ons oordeel is dan ook dat het voorstel van NS voldoet aan dit criterium**

Optimale methodiek

In haar advies heeft CQM een schema opgenomen om te bepalen welke methodiek het beste gebruikt kan worden. NS heeft dit schema gevolgd en heeft de methodiek die daaruit voortkwam gehanteerd bij de omzetting van bodem- en streefwaarden. Deze gekozen methodiek betreft een gewogen lineaire regressie. Een lineaire regressie is de geijkte methode om de waarde van een bepaalde variabele te voorspellen op basis van de waarde van een andere variabele. In dit geval is het doel om de waarden van een PI op basis van FLUX te voorspellen met behulp van de waarden van dezelfde PI op basis van ROCKT. Een lineaire regressie lijkt dus een methodiek te zijn die inhoudelijk goed past bij het vraagstuk, conform het advies van CQM. Om een lineaire regressie met vertrouwen te kunnen gebruiken, zijn er een aantal statistische eisen waaraan voldaan moet worden.

- **Lineaire relatie.** Er moet sprake zijn van een lineair verband tussen de variabelen (in dit geval de metingen van de PI's o.b.v. ROCKT en die o.b.v. FLUX. Ook mogen er geen grote uitbijters zijn.
- **Weinig tot geen multicollineariteit.** Als er meerdere variabelen gebruikt worden om één variabele te voorspellen, dan dienen deze onafhankelijk te zijn. Aangezien er in dit geval slechts één variabele wordt gebruikt, is deze eis niet van toepassing.
- **Geen auto-correlatie.** De (fouttermen van de) verschillende voorspelde waarden (de resultaten per maand) dienen onafhankelijk van elkaar te zijn.
- **Homoscedasticiteit.** De grootte van de fouttermen mogen niet toe- of afnemen als de voorspelde waarden groter of kleiner worden.
- **Normaliteit.** De verdeling van de fouttermen moet overeen komen met de normaalverdeling.

Om te toetsen of een lineaire regressieanalyse in deze situatie bruikbaar is, hebben we deze eisen één voor één getoetst. De uitgebreide resultaten hiervan zijn terug te vinden in bijlage B. Onze conclusie is dat een lineaire regressie in deze situatie een zeer goed passende methodiek is. Hoewel natuurlijk nooit met 100% zekerheid gesteld kan worden dat een methodiek 'optimaal' is, is er geen andere voor de hand liggende methodiek die zou leiden tot bodem- en streefwaarden die met grotere zekerheid dezelfde betekenis hebben als de huidige waarden. **Ons oordeel is dat de door NS voorgestelde methodiek zo optimaal is als redelijkerwijs kan worden gesteld.**

Objectiviteit methodiek

De methodiek die NS voorstelt (in lineaire regressieanalyse) is in principe volledig objectief. Er is een vaste manier om een lineaire regressie uit te voeren. Als verschillende personen of instanties dezelfde data hebben, komen ze tot hetzelfde eindresultaat. Voordat de lineaire regressie gedaan kan worden, moeten er over een aantal onderwerpen keuzes worden gemaakt. Bij het nagaan van de voorgestelde methodiek, hebben we drie onderwerpen gevonden waarop NS een keuze heeft gemaakt. Deze onderwerpen zijn ook al beschreven in paragraaf 2.5 en zijn:

- De weging van de PI-scores per maand
- De beoordeling van de grootte van de foutterm uit de lineaire regressie
- De gebruikte dataset

Onderstaand beschrijven we voor elk van deze drie onderwerpen kort ons oordeel over de mate waarin de keuze die NS heeft gemaakt invloed heeft gehad op de omzetting van de bodem- en streefwaarden. We richten ons daarbij vooral op de mate van objectiviteit van de keuzes. Met andere woorden, is de keuze die NS heeft gemaakt inhoudelijk logisch, onafhankelijk van de uitkomsten waar deze toe leidt.

De weging van de PI-scores per maand

Zoals eerder beschreven (en ook geadviseerd door CQM), is het van belang om de maandelijkse resultaten van de PI's te wegen. Voor de PI 'Aantal drukke treinen per werkweek in de spits op het HRN' is er slechts één logische factor beschikbaar voor deze weging, namelijk het aantal werkdagen in de betreffende maand. Voor de twee andere PI's is het logisch om als weegfactor het aantal reizigers te nemen dat ook in scope is voor deze PI's. Het aantal reizigers kan echter worden bepaald via ROCKT, of via FLUX. NS heeft gekozen voor het aantal reizigers op basis van FLUX, waarbij ze aangeven dat de keuze geen verschil maakt. Tijdens het onderzoek hebben we zelf een weging uitgevoerd op basis van FLUX, en een weging op basis van ROCKT. Hierbij hebben we kunnen verifiëren dat de resulterende bodem- en streefwaarden exact gelijk zijn, in ieder geval tot op de precisie van de afspraken met het ministerie. Er is dus inderdaad geen enkel verschil tussen beide weegfactoren.

De beoordeling van de grootte van de foutterm uit de lineaire regressie

De voorspellingen op basis van een lineaire regressie zullen nooit exact uitkomen op de werkelijk gemeten waarden. Om te kunnen beoordelen of de voorspellingen 'goed genoeg' zijn, moet worden bepaald wanneer een afwijking van de werkelijk gemeten waarden 'voldoende klein' is. In het advies van CQM is dit belang opgenomen, waarbij de exacte invulling van deze beoordeling aan NS wordt gelaten. Wel geeft CQM een mogelijkheid, namelijk de impactschaal van de betreffende PI die NS ook gebruikt voor andere doeleinden. In het voorstel van NS is deze impactschaal inderdaad gebruikt om te bepalen of afwijkingen 'voldoende klein' zijn. De impactschaal wordt door NS gebruikt voor ramingen richting de toekomst. De impact van elke verwachte ontwikkeling op de betreffende PI wordt ingedeeld in een van de categorieën 'geen', 'laag', 'middel', of 'hoog'. Voor elke PI zijn deze categorieën vervolgens vertaald naar kwantitatieve bandbreedtes.

De omzetting van bodem- en streefwaarden van ROCKT naar FLUX kan (net als bijv. een verbouwing van een station) als ontwikkeling worden gezien die impact heeft op een PI. Volgens die gedachte is het in ons oordeel dan ook logisch om de impactschaal te hanteren om te bepalen of de impact 'voldoende klein' is. De definitie voor 'voldoende klein' die dan voor de hand ligt, is een score op de impactschaal van maximaal 'klein' (de score 'geen' is numeriek gelijk aan exact 0, waardoor deze in de praktijk niet mogelijk is voor deze omzetting). NS heeft deze keuze dan ook gemaakt in het voorstel voor de omzetting van bodem- en streefwaarden.

De gebruikte dataset

In principe is er data van 16 maanden beschikbaar (juni 2022 t/m september 2023). In haar advies geeft CQM aan dat er in ieder geval 12 opvolgende maanden gebruikt dienen te worden. Daarnaast is het advies ook om, als er meer dan 12 maanden worden gebruikt, er met hele jaren gewerkt dient te worden (dus bijv. 24 maanden). De reden hiervoor is dat er in verschillende maanden sprake is van verschillend reizigersgedrag. Door slechts delen van jaren mee te nemen, zullen sommige reizigersbewegingen oververtegenwoordigd zijn, en andere ondervertegenwoordigd. Dit zou leiden tot een minder betrouwbare voorspelling. NS heeft de keuze gemaakt om de eerste vier maanden niet mee te nemen, en de analyse te baseren op het jaar lopend van oktober 2022 t/m september 2023. Het argument dat NS hiervoor geeft is dat er in de eerste vier maanden nog sprake is van enige corona-effecten. Hierbij geeft NS het volgende aan: "Onderzoek wijst uit dat het voor het eindresultaat niet of nauwelijks uitmaakt welke periode van 12 aaneengesloten maanden gekozen wordt."

Met 16 maanden aan data, en het advies dat er 12 aaneengesloten maanden gebruikt dienen te worden, zijn er in principe vijf opties (van maand 1 t/m 12, van maand 2 t/m 13, etc.). NS heeft voor elk van deze opties in kaart gebracht tot welke bodem- en streefwaarden die keuze zou leiden. Deze resultaten heeft NS ons ter beschikking gesteld. Door zelf eenzelfde lineaire regressieanalyse te doen (zie hiervoor ook het laatste criterium) hebben we kunnen verifiëren dat we voor alle opties tot dezelfde bodem- en streefwaarden kwamen. Als een andere keuze wordt gemaakt over welke 12 maanden gebruikt worden in de analyse, zijn er kleine verschillen in de resulterende bodem- en streefwaarden. In onderstaande tabel is weergegeven hoe de vier alternatieve keuzes voor 12 aaneengesloten maanden de bodem- en streefwaarden zouden beïnvloeden. Voor de zitplaatskans PI's is dit gedaan in termen van procentpunten (als voorbeeld: 55% ligt 1%-punt hoger dan 54%). Voor de PI betreffende drukke treinen is dit gedaan in termen van het aantal drukke treinen. De weergegeven verschillen van 2 en 3 drukke treinen komen ongeveer overeen met afwijkingen van 2% en 3% ten opzichte van de bodem- en streefwaarden uit het voorstel van NS.

	Alternatief 1	Alternatief 2	Alternatief 3	Alternatief 4
BW Zitplaatskans HRN	+ 0,1%-punt	+ 0,1%-punt	+ 0,1%-punt	+ 0,1%-punt
SW Zitplaatskans HRN	– 0,1%-punt	– 0,1%-punt	0	0
BW Zitplaatskans HSL	+ 0,2%-punt	+ 0,2%-punt	+ 0,2%-punt	+ 0,1%-punt
SW Zitplaatskans HSL	0	0	0	0
BW Drukke treinen	– 3 treinen	– 3 treinen	– 3 treinen	– 3 treinen
SW Drukke treinen	– 2 treinen	– 2 treinen	– 3 treinen	– 3 treinen

Voor de twee zitplaatskans PI's geldt dat een hogere bodem- of streefwaarde lastiger te realiseren is voor NS dan een lagere bodem- of streefwaarde. Voor de PI omtrent drukke treinen geldt dit precies omgekeerd: de bodemwaarde stelt een maximum aan drukke treinen dat behaald dient te worden, dus een lagere bodemwaarde is lastiger te realiseren voor NS. De bodemwaarden van beide zitplaatskans PI's, en de bodem- en streefwaarde van de PI rondom drukke treinen zouden voor NS lastiger te realiseren worden als gekozen was voor een andere set van 12 opeenvolgende maanden. De streefwaarden van de zitplaatskans PI's zouden ofwel niet veranderen, ofwel iets minder lastig worden om te realiseren.

De keuze van NS om de laatste 12 maanden te gebruiken als dataset leidt dus voor de meeste bodem- en streefwaarden tot iets eenvoudiger te realiseren waarden. Het gaat hierbij om kleine verschillen. Het is onmogelijk om op statistische wijze te bepalen welke set van 12 maanden tot 'de best passende' vertaling van bodem- en streefwaarden zou leiden. Bij afwezigheid van een dergelijke statistische methode, is het van belang om een goede inhoudelijke keuze te maken. De inhoudelijke onderbouwing bij de keuze van NS is dat er in de eerste vier maanden nog sprake was van enige naweeën van de coronapandemie. In onderstaande figuren zijn de reizigersaantallen in de vier maanden waarvan zowel uit 2022 als 2023 data beschikbaar was met elkaar vergeleken. In de linker figuur is dit gedaan voor het aantal reizigers dat binnen scope is voor de zitplaatskans PI die het HRN betreft, en in de rechter figuur voor de HSL.





In deze figuren is duidelijk te zien dat de reizigersaantallen in 2022 voor deze vier maanden fors lager liggen dan de reizigersaantallen in 2023. Ook is te zien dat het verschil in reizigersaantallen in september kleiner is dan in de maanden daarvoor. Dit is een indicatie dat de reizigersbewegingen in deze vier maanden inderdaad nog afwijkend waren (als gevolg van de naweeën van de coronapandemie), en dat deze afwijking in september begon af te nemen.

Voor de gegevens waarop de lineaire regressie is gebaseerd heeft NS naar ons oordeel een logische keuze gemaakt. In totaal beschikte NS over 16 maanden aan bruikbare data. Het advies van CQM dat er exact 12 opeenvolgende maanden meegenomen dienden te worden, is door NS opgevolgd. Hoewel andere keuzes van 12 opeenvolgende maanden tot bodem- en streefwaarden zouden hebben geleid die lastiger te realiseren waren geweest, is de gemaakte keuze gezien de naweeën van de coronapandemie, een logische. De effecten van de pandemie die NS noemt als argument voor deze keuze zijn ook terug te zien in de reizigersbewegingen.

Overkoepelend oordeel

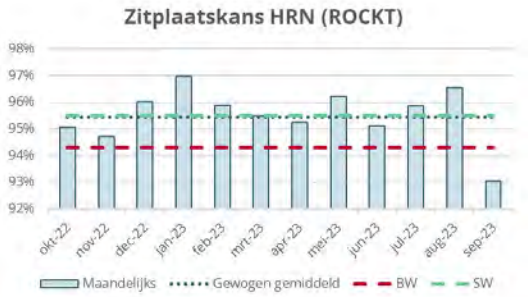
Wat dit criterium betreft waren er drie onderwerpen waarop NS een keuze heeft moeten maken. In het eerste geval heeft de keuze geen impact op de uitkomsten. In beide andere gevallen is de gemaakte keuze naar ons oordeel inhoudelijk de meest logische keuze. Hoewel er hiermee dus enige invloed van NS bestaat op de uitkomsten, zou een andere partij (met dezelfde beschikbare data) op basis van inhoudelijke argumenten dezelfde keuze hebben gemaakt. **Ons oordeel is dat de door NS voorgestelde methodiek zo objectief als mogelijk is.**

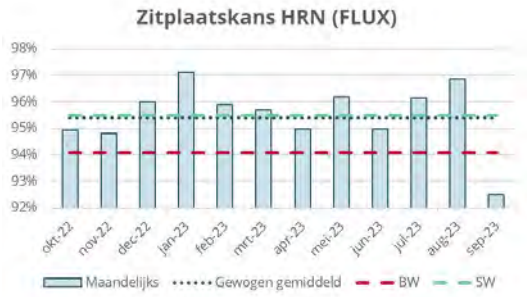
Behoud van benodigde inspanning

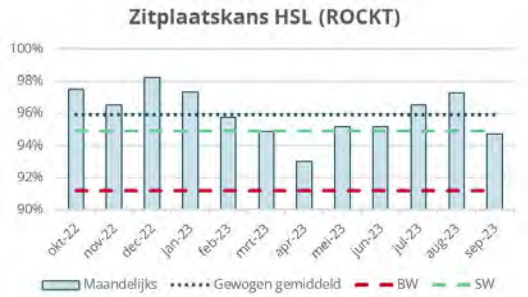
Door het gebruik van een lineaire regressieanalyse, worden de verschillen tussen de waarden van de PI's op basis van FLUX en de PI's op basis van ROCKT zo goed mogelijk voorspeld. De formule om de PI's op basis van ROCKT zo goed mogelijk om te zetten naar de PI's op basis van FLUX, wordt door NS ook gebruikt om de bodem- en streefwaarden om te zetten. Bij correcte toepassing van deze methodiek (waarover in andere eisen een oordeel wordt gegeven), leidt dit per definitie tot nieuwe

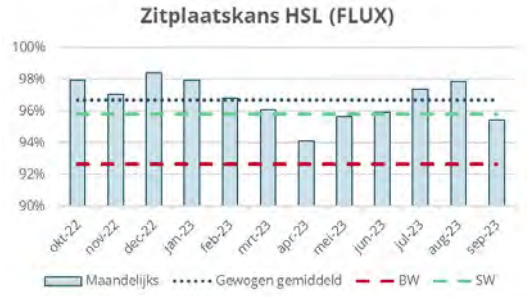
bodem- en streefwaarden die zich op dezelfde manier verhouden tot de nieuwe waarden van de PI's, als de huidige bodem- en streefwaarden zich verhouden tot de huidige waarden van de PI's.

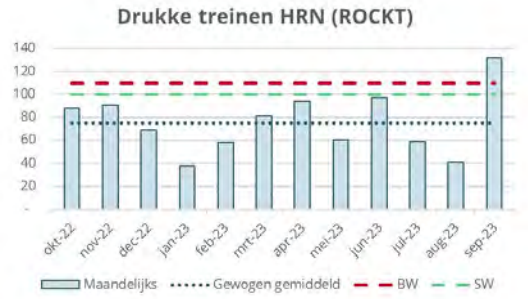
Om hier een extra onderbouwing bij te geven is in onderstaande figuren weergegeven hoe de gerealiseerde PI-waarden zich in beide situaties verhouden tot de bodem- en streefwaarden. De linker figuren betreffen de huidige situatie (op basis van ROCKT) en de rechter figuren de situatie waarin FLUX wordt gebruikt. De blauwe staven geven de maandelijkse PI-waarden aan en de donkerblauwe stippellijn het jaarlijks gemiddelde. De gestreepte rode lijn is de bodemwaarde en de gestreepte groene lijn de streefwaarde.

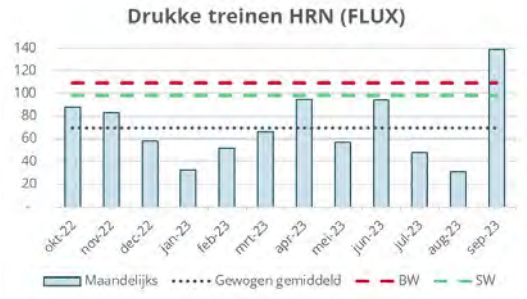












Door de figuren links te vergelijken met de figuren rechts ontstaat een beeld van het ambitieniveau van NS in beide situaties. De figuren zijn zeer vergelijkbaar te noemen. Hoewel er kleine verschillen zijn, kan over het algemeen gesteld worden dat de scores van PI's ten opzichte van de bodem- en streefwaarden relatief gelijk blijven. **Ons oordeel is dat de door NS voorgestelde methodiek leidt tot bodem- en streefwaarde die dezelfde inspanning vragen om gehaald te kunnen worden.**

Correcte toepassing

Ten behoeve van deze audit heeft NS alle gegevens overgedragen waarop zij de analyse hebben gebaseerd, en waaruit hun voorstel voor bodem- en streefwaarden voortkomt. Hierbij gaat het om de maandelijks gemeten waarden van de PI's op basis van zowel ROCKT als FLUX, en de gewichten

(aantal reizigers en aantal werkdagen per maand) op basis van beide systemen. Met behulp van deze gegevens hebben we zelf een volledige regressieanalyse uitgevoerd, conform de door NS voorgestelde methodiek. Buiten het volgen van de beschreven methodiek, is geen gebruik gemaakt van enige vorm van rekenmodellen of -middelen vanuit NS. De analyse is door ons dus compleet onafhankelijk gedaan van de analyse die NS zelf heeft uitgevoerd.

Alle resultaten die NS rapporteert zijn door ons geverifieerd. Hierbij gaat het uiteraard om de bodem- en streefwaarden, maar bijv. ook om de parameters van de voorspelmodellen die uit de regressieanalyse volgen. Ook de door NS gerapporteerde standaardafwijkingen komen exact overeen met de door ons gevonden waarden. **We oordelen dan ook dat NS de gekozen methodiek correct en volledig conform omschrijving heeft uitgevoerd.**

3.3 Conclusie

In de vorige paragraaf hebben we op elk van de criteria uit het beoordelingskader ons oordeel gegeven. Er kan redelijkerwijs gesteld worden dat de gekozen (en door CQM geadviseerde) methodiek de best passende methodiek is voor een dergelijk vraagstuk. Deze methodiek is op drie keuzes na volledig objectief. Van deze drie keuzes heeft er één geen invloed op de uitkomsten. De andere twee keuzes zijn door NS vanuit inhoudelijk logische argumenten gemaakt. Hoewel deze keuzes een kleine invloed op de uitkomsten hebben, is met aan zekerheid grenzende waarschijnlijkheid te stellen dat een andere partij dezelfde keuzes zou hebben gemaakt. Tot slot is de beschreven methodiek cijfermatig correct uitgevoerd. Dit alles leidt ons tot de conclusie dat de omzetting van de bodem- en streefwaarden zo ambitieneutraal als mogelijk is gedaan.

/ A Definities van PI's

In deze bijlage lichten we de PI's toe waarvan de waarde verandert door de overstap van ROCKT naar FLUX. De onderstaande beschrijvingen zijn beknopte versies van de meer uitgebreide definitie die NS hanteert. In deze definities is FLUX al aangehouden als input-bron.

Zitplaatskans in de spits op het HRN

Zitplaatskans in de spits HRN geeft een indicatie van de kans dat een reiziger die tijdens de spits in de trein stapt de volledige reis heeft kunnen zitten. Dat wil zeggen dat de reiziger bij zijn vertrekstation en bij ieder overstapstation op zijn reis in een trein binnen de HRN-concessie instapt, waarbij het aantal zitplaatsen ten minste gelijk is aan het totaal aantal reizigers. Hierbij wordt alleen gekeken naar de zitplaatskans van die delen van de treinreis die binnen de HRN-concessie worden afgelegd. In scope zijn reizigers die in- én uitchecken en waarbij de inchecktijd in de spits ligt. Als spitsperiode wordt gehanteerd: werkdagen (met uitzondering van officiële feestdagen) tussen 6:35:00 en 8:55:00 of tussen 16:05:00 en 18:25:00.

Zitplaatskans in de spits op de HSL

Zitplaatskans in de spits HSL geeft een indicatie van de kans dat een reiziger die tijdens de spits in de trein stapt de volledige reis heeft kunnen zitten voor de 12 geselecteerde reisrelaties. Deze indicator betreft een deelverzameling van de indicator Zitplaatskans in de spits HRN. In scope zijn reizigers waarbij de inchecktijd in de spits ligt en die in- én uitchecken op één van de onderstaande reisrelaties. Beide richtingen tellen mee:

- Amsterdam Centraal – Breda
- Amsterdam Centraal – Rotterdam Centraal
- Amsterdam Zuid – Rotterdam Centraal
- Rotterdam Centraal – Breda
- Schiphol Airport – Breda
- Schiphol Airport – Rotterdam Centraal

Reizigers op deze reisrelaties maken, door de snellere reistijd, hoofdzakelijk gebruik van de treinen over de HSL. Echter ook reizigers op deze reisrelaties die niet van de HSL treinen gebruik maken (bijvoorbeeld door een stremming op de HSL), zijn in scope van deze PI. Als spitsperiode wordt gehanteerd: werkdagen (met uitzondering van officiële feestdagen) tussen 6:35:00 en 8:55:00 of tussen 16:05:00 en 18:25:00.

Aantal drukke treinen per werkweek in de spits op het HRN

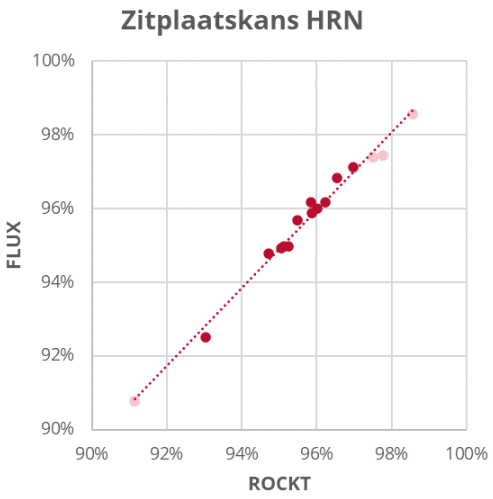
Aantal drukke treinen per werkweek in de spits HRN betreft het gemiddeld aantal treinen per werkweek, waar op enig moment tijdens de spits het aantal reizigers in de trein groter is geweest dan het aantal gerealiseerde vervoersplaatsen. In scope zijn treinen die binnen de HRN-concessie rijden en de stoptrajecten van deze treinen waarvan de gerealiseerde vertrektijd in de spits ligt. Als spitsperiode wordt gehanteerd: werkdagen (met uitzondering van officiële feestdagen) tussen 6:35:00 en 8:55:00 of tussen 16:05:00 en 18:25:00.

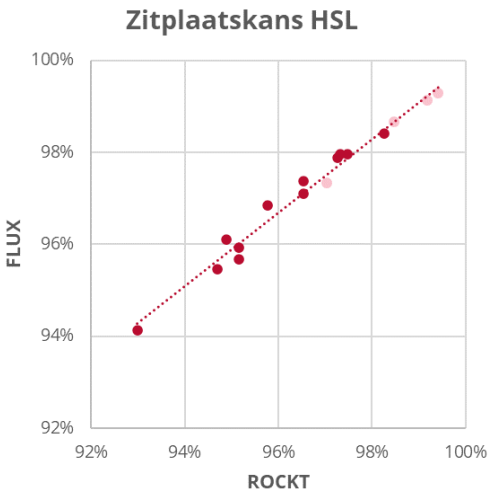
/ B Aannames lineaire regressieanalyse

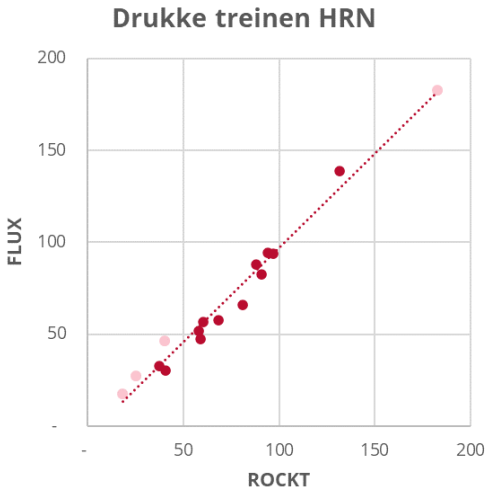
Aan het gebruik van een lineaire regressieanalyse liggen een aantal aannames ten grondslag. Als niet aan deze aannames is voldaan, zijn de resultaten van de lineaire regressie minder betrouwbaar. Onderstaand beschrijven we deze aannames en geven aan in hoeverre hieraan voldaan is in het geval van de omzetting van bodem- en streefwaarden voor de betreffende PI's.

Lineaire relatie

Om een lineaire regressieanalyse betrouwbaar te kunnen gebruiken om één variabele te voorspellen op basis van een andere variabele, moet er sprake zijn van een lineair verband tussen de twee variabelen. Daarnaast is het van belang dat er geen grote uitbijters zijn, aangezien die een grote invloed hebben op het resultaat van de regressieanalyse. In onderstaande figuren is voor de drie PI's de relatie tussen metingen op basis van ROCKT en metingen op basis van FLUX weergegeven. Elk meetpunt representeert een maand, waarbij de meting op basis van ROCKT de horizontale positie bepaalt, en de meting op basis van FLUX de verticale positie. De meetpunten uit de maanden juni t/m september 2022 zijn lichtrood weergegeven.







In elke figuur is de best passende lineaire lijn opgenomen. Met behulp van deze lijn is duidelijk zichtbaar dit er voor elk van de PI's sprake is van een lineair verband tussen beide meetmethoden (de meetpunten liggen vrij goed op een lijn). Daarnaast is er ook geen sprake van grote uitbijters (meetpunten die ver weg liggen van de lijn). We kunnen dus concluderen dat er aan deze eis is voldaan.

Weinig tot geen multicollineariteit

Als er meerdere variabelen worden gebruikt om samen één variabele te voorspellen, dan dienen deze variabelen onderling onafhankelijk te zijn. In dit geval wordt er slechts één variabele gebruikt, dus is deze aanname niet van toepassing.

Geen auto-correlatie

De verschillende voorspelde waarden (de PI-waarden per maand o.b.v. FLUX, voorspeld vanuit de PI-waarden o.b.v. ROCKT) dienen onderling onafhankelijk te zijn. Om dit te toetsen, kijken we naar de verschillen tussen de werkelijke PI-waarden o.b.v. FLUX en de voorspelde PI-waarden voor FLUX. Op deze fouttermen (of residuen) voeren we de Durbin-Watson test uit. Hierbij bepalen we eerst voor elke meting (elke maand, startend bij maand 2) het verschil tussen de foutterm van die maand en de foutterm van de maand ervoor. Dit verschil kwadrateren we, waarna we alle deze gekwadrateerde verschillen optellen. Dit resultaat delen we vervolgens door de som van de gekwadrateerde fouttermen (voor elk van de 12 maanden). In formulevorm ziet dat er als volgt uit

$$d = \frac{\sum_{n=2}^{N}(e_n - e_{n-1})^2}{\sum_{n=1}^{N} {e_n}^2}$$

Waarin n de maanden aanduidt, met N als laatste maand, en waarin e_n de foutterm van maand n is. Het resultaat van deze test (in bovenstaande formule d) zal altijd tussen de 0 en de 4 liggen. Een resultaat van exact 2 betekent dat er totaal geen auto-correlatie is. Over het algemeen wordt gesteld dat een resultaat tussen de 1,5 en 2,5 betekent dat er sprake is van een acceptabele hoeveelheid auto-correlatie. In onderstaande tabel zijn de resultaten voor deze test weergegeven voor de drie PI's.

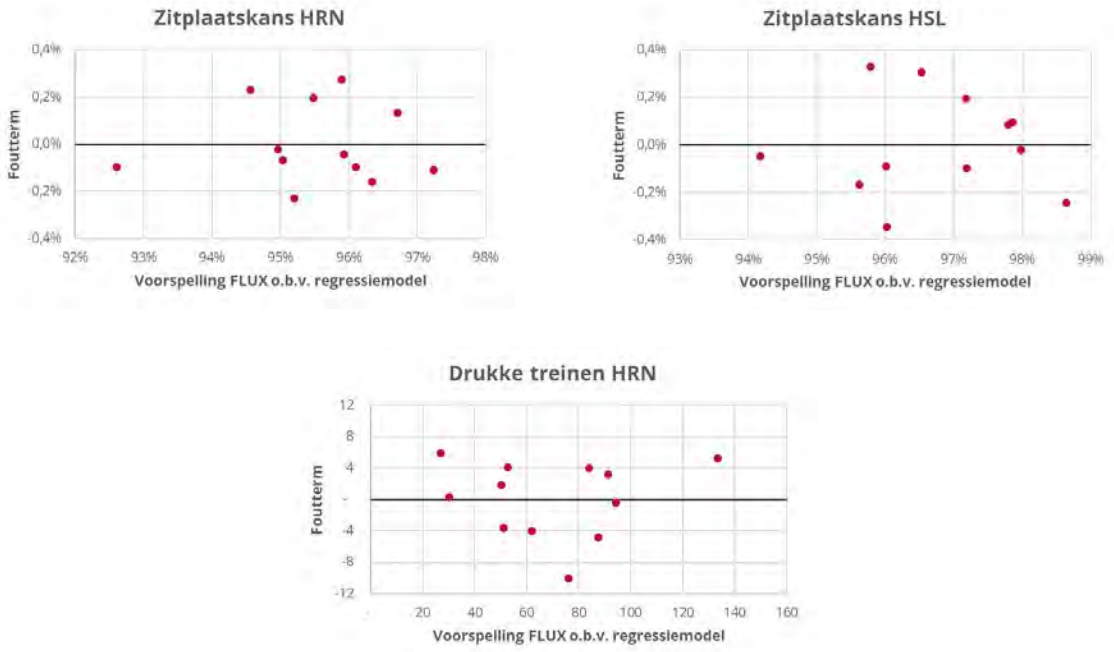
PI	Resultaat *d*
Zitplaatskans in de spits op het HRN	2,05
Zitplaatskans in de spits op de HSL	1,32
Aantal drukke treinen per werkweek in de spits op het HRN	2,22

In deze tabel is de zien dat er bij de eerste twee PI's geen sprake is van auto-correlatie. Bij de PI zitplaatskans in de spits op de HSL valt het resultaat d net onder de waarde 1,5. Hoewel dit over het algemeen kan betekenen dat er sprake is van een te grote hoeveelheid auto-correlatie, is het waarschijnlijk dat dit hier niet het geval is. De reden hiervoor is de seizoensdynamiek waar sprake van is. De Durbin-Watson test gebruikt resultaten van opvolgende maanden. In het geval van de metingen voor de PI kunnen de resultaten van opvolgende maanden fors verschillen. Dit kan leiden tot afwijkende waarden voor d. Wetende dat deze dynamiek speelt, stellen we dat er voldoende aan deze eis is voldaan om een lineaire regressie met vertrouwen uit te kunnen voeren.

Homoscedasticiteit

De voorspelling van de PI-waarden o.b.v. FLUX (voorspeld vanuit de PI-waarden o.b.v. ROCKT) zullen niet exact gelijk zijn aan de werkelijk gemeten PI-waarden o.b.v. FLUX. Voor een lineaire regressie is het van belang dat de varianties van de fouttermen of residuen (de afwijkingen tussen de

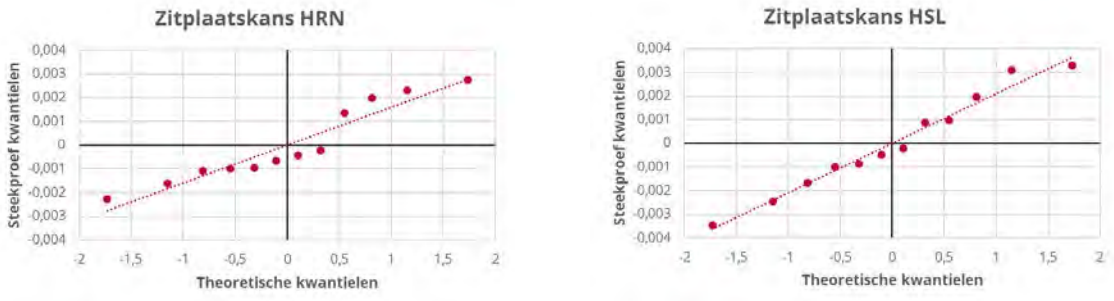
voorspellingen en de werkelijke waarden) ongeveer gelijk zijn. In onderstaande figuur zijn voorspellingen van FLUX (op basis van de formule uit de regressieanalyse) en de bijbehorende fouttermen weergegeven.

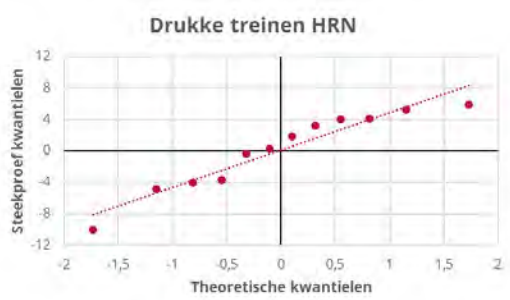


Als er sprake is van homoscedasticiteit, dan heeft de voorspelling van het regressiemodel (de waarde op de horizontale as) geen invloed op de grootte van de foutterm (de waarde op de verticale as). In alle drie de figuren is duidelijk te zien dat dit het geval is. De fouttermen worden niet significant groter of kleiner naar mate men verder naar links of rechts in de figuur kijkt. We kunnen dus concluderen dat aan deze eis wordt voldaan.

Normaliteit

De fouttermen dienen een normaalverdeling te volgen. Om dit te controleren maken we gebruik van een zogeheten Q-Q plot. Hiermee kan worden onderzocht of twee verschillende (kans)verdelingen op eenzelfde manier verdeeld zijn. Hiervoor beschouwen we de verdeling van de fouttermen als een kansverdeling, en vergelijken deze met de normaalverdeling (ook wel Gaussverdeling of klokverdeling genoemd). In onderstaande figuren zijn de resultaten weergegeven. Elk punt representeert de foutterm van één van de metingen (maanden). Hiervoor zijn de metingen eerst gesorteerd op grootte van de foutterm. De verticale positie van elk meetpunt representeert dus de plek die de foutterm van dit meetpunt inneemt binnen de verdeling van alle fouttermen. Op de horizontale as is de vertaling gemaakt naar de plek binnen de normaalverdeling waarmee dit correspondeert.





Als de verdeling van de fouttermen exact overeen zou komen met de normaalverdeling, dan zouden alle punten in de bovenstaande figuren op een rechte lijn liggen. In de figuren is de best passende rechte lijn weergegeven om te laten zien in hoeverre dit het geval is. Voor geen van de PI's is er sprake van een structurele afwijking (bijvoorbeeld afwijkingen van de normaalverdeling die steeds groter worden als men naar links of rechts in de figuur gaat). De afwijkingen die er dus zijn ten opzichte van de normaalverdeling berusten op toevalligheden. Hieruit kunnen we concluderen dat ook aan deze eis is voldaan.