



Andersson
Eiffers
Felix

Methodiek bodem- en streefwaarden KPI's NS

Validatierapport

Datum

28 november 2022

Opdrachtgever

Ministerie van Infrastructuur en Waterstaat (IenW)

Onderzoeksteam

██████████
████████████████████

Referentie

GI182/Eindrapportage

bezoekadres Maliebaan 16 3581 CN Utrecht
postadres Postbus 85198 3508 AD Utrecht

telefoon ████████████████████
website www.aef.nl

/ Managementsamenvatting

De concessie voor het hoofdrailnet is voor de periode 2015-2025 door het ministerie van Infrastructuur en Waterstaat (IenW) verleent aan de Nederlandse Spoorwegen (NS). Het ministerie stuurt de NS aan met behulp van kritische prestatie-indicatoren (KPI's). Voor deze KPI's worden bodem- en streefwaarden afgesproken. De bodemwaarde van een KPI is de minimale waarde die de NS ieder jaar dient te realiseren. De streefwaarde is de waarde die een KPI aan het eind van de concessieperiode dient te hebben.

De methodiek die de NS hanteert om te komen tot een voorstel voor bodem- en streefwaarden is tijdens de midterm review in 2020 door een extern bureau gevalideerd. De conclusie destijds was dat de gehanteerde methode kenmerken had van een valide methodiek, maar beter (kwalitatief) onderbouwd diende te worden. Naar aanleiding van dit rapport heeft het ministerie van IenW Andersson Elffers Felix (AEF) gevraagd om samen met NS de benodigde verbeteringen door te voeren en zo te komen tot een valide methodiek.

Als uitgangspunt voor dit onderzoek zijn de eerder geconstateerde tekortkomingen geanalyseerd en is een afwegingskader ontwikkeld om vast te leggen wat een methode 'valide' maakt. In een tweede stap is de huidige methodiek in beeld gebracht door middel van documentstudie en gesprekken met medewerkers van NS. Ten opzichte van de methodiek die tijdens de validatie bij de midterm review werd gehanteerd, heeft NS een aantal dingen aangepast in haar werkwijze. In de derde stap van dit onderzoek is de beschrijving van de methodiek door AEF en NS samen doorontwikkeld. Dit heeft uiteindelijk geleid tot een zelfstandig document van NS, waarin de methodiek wordt beschreven.

Uit de analyse van de huidige methodiek die in dit rapport wordt gedaan, blijkt dat alle tekortkomingen die waren geconstateerd tijdens de midterm review van 2020, weg zijn genomen. Ten slotte heeft AEF aan de hand van het eerder vastgestelde afwegingskader de methodiek (en vooral de beschrijving daarvan) beoordeeld. De beschrijving van de huidige methodiek voldoet aan alle eisen uit dit afwegingskader. Omdat deze eisen gezamenlijk de definitie van 'valide methodiek' vormen die in dit onderzoek wordt gehanteerd, kan AEF stellen dat NS momenteel een valide methodiek hanteert om te komen tot een (bandbreedte voor) bodem- en streefwaarden.

/ Inhoud

/	Managementsamenvatting	2
/ 1	Inleiding	4
1.1	Aanleiding	4
1.2	Vraagstelling en afbakening	4
1.3	Onderzoeksaanpak	4
1.4	Leeswijzer	5
/ 2	Validatierapport Rebel 2020	6
2.1	Inleiding	6
2.2	Beschrijving methodiek	6
2.3	Conclusies validatierapport	7
/ 3	Verbeterde methodiek	9
3.1	Eisen aan methodiekbeschrijving	9
3.2	Beschrijving verbeterde methodiek	9
3.3	Verbetering eerdere tekortkomingen	12
3.4	Conclusie	15
/ A	De huidige KPI's	17

/ 1 Inleiding

1.1 Aanleiding

Als concessieverlener stuurt het ministerie van Infrastructuur en Waterstaat (IenW) de Nederlandse Spoorwegen (NS) aan door middel van de vervoerconcessie 2015-2025 van het hoofdrailnet. Onderdeel van het hierin vastgelegde sturingsinstrumentarium is een tiental kritische prestatie-indicatoren (KPI's), welke ter volledigheid in bijlage A zijn opgenomen. Voor deze KPI's gelden bodem- en streefwaarden als onderdeel van een bonus-malus regeling.

Tijdens de midterm review van de vervoerconcessie in 2020¹ is de methodiek waarmee de NS de inzet voor de nieuwe bodem- en streefwaarden bepaalt door een extern bureau gevalideerd.² Uit dit onderzoek bleek dat de methode kenmerken heeft van een valide methodiek, maar dat deze beter (kwalitatief) onderbouwd dient te worden. NS en het ministerie van IenW hebben hierop afgesproken om de methodiek te verbeteren op de punten waarop tekortkomingen zijn geconstateerd.

1.2 Vraagstelling en afbakening

Het ministerie van IenW heeft Andersson Elffers Felix (AEF) gevraagd om samen met NS de benodigde verbeteringen door te voeren en zo te komen tot een valide methodiek. De scope van de opdracht is hierbij gericht op tekortkomingen in de processtappen om tot (een bandbreedte voor) bodem- en streefwaarden te komen. Daarbij ligt de focus op de KPI's die onder de huidige concessie vallen, maar is het ook van belang dat de methodiek voor de volgende concessieperiode gebruikt kan worden.

1.3 Onderzoeksaanpak

Om te komen tot een valide methodiek en voorliggend rapport heeft AEF de volgende stappen doorlopen:

- ▶ In **stap 1** is de basis gelegd voor het verdere onderzoek. De eerder geconstateerde tekortkomingen zijn geïnventariseerd en geanalyseerd. Daarnaast is (kwalitatief) bepaald waaraan het proces om te komen tot een bandbreedte voor bodem- en streefwaarde zou moeten voldoen.
- ▶ In **stap 2** heeft AEF de beschikbare documentatie bestudeerd en gesprekken gevoerd met medewerkers van de NS om een beter beeld te krijgen van de huidige methodiek van de NS om tot een bandbreedte voor bodem- en streefwaarden te komen. Op basis hiervan zijn de belangrijkste witte vlekken c.q. verbeterpunten voor het proces in beeld gebracht.
- ▶ **Stap 3** stond in het teken van het verbeteren van de documentatie van de methodiek. Hiervoor is een werkgroep met experts van NS samengesteld waarmee AEF een iteratief proces heeft

¹ De huidige vervoerconcessie loopt van 2016 tot en met 2025. Conform de afspraken in de vervoerconcessie is halverwege deze looptijd een tussentijdse evaluatie uitgevoerd. Zie ook Kamerbrief over midterm reviews concessies van NS en ProRail d.d. 11 juni 2020.

² [Validatierapport - Methodieken NS voor totstandkoming prognose reizigersgroei en inzet bodem- en streefwaarden vervoerconcessie HRN | Rapport | Rijksoverheid.nl](#)

doorlopen waarin de beschrijving van de methodiek steeds verder is doorontwikkeld tot op het punt waarop deze als compleet, transparant, navolgbaar en controleerbaar, en dus als valide, kon worden aangemerkt.

- ▶ Ten slotte is de opdracht in **stap 4** afgerond met het opstellen van dit validatierapport.

1.4 Leeswijzer

Het volgende **hoofdstuk 2** gaat kort in op de belangrijkste bevindingen in het validatierapport van Rebel uit 2020. In **hoofdstuk 3** wordt de verbeterde methodiek van NS beschreven, waarbij expliciet stilgestaan wordt bij de wijze waarop de eerder geconstateerde tekortkomingen zijn geadresseerd. Dit rapport besluit met een korte conclusie aan het eind van hoofdstuk 3.

/ 2 Validatierapport Rebel 2020

2.1 Inleiding

In mei 2020 heeft onderzoeksbureau Rebel het validatierapport 'Methodieken NS voor totstandkoming prognose reizigersgroei en inzet bodem- en streefwaarden vervoerconcessie HRN' (hierna ook: 'het validatierapport 2020') opgeleverd. Doel van het rapport was om te "controleren of en in hoeverre NS een duidelijke en herleidbare methodiek heeft toegepast om tot een realistische inzet van bodem- en streefwaarden voor de KPI's te komen."³ Omdat onderhavig onderzoek gericht is op het wegnemen van de destijds geconstateerde tekortkomingen in het proces om te komen tot bodem- en streefwaarden, worden de belangrijkste conclusies van Rebel samengevat.⁴ Dit hoofdstuk is een objectieve weergave van het eerdere validatierapport, en bevat geen beschouwingen en conclusies van AEF.

2.2 Beschrijving methodiek

Alvorens stilgestaan wordt bij de conclusies uit het validatierapport 2020, wordt eerst kort de methodiek voor het bepalen van de bodem- en streefwaarden weergegeven zoals destijds door Rebel beschreven op basis van de van NS ontvangen informatie.⁵

De verschillende KPI's worden eerst ingedeeld in vier clusters. De vier clusters die in het rapport worden benoemd zijn: Algemeen klantoordeel, Zitplaatskans, Betrouwbaarheid en Reisinformatie. Op elk van deze clusters wordt vervolgens de methodiek toegepast die schematisch is weergegeven in onderstaande figuur.



Figuur 1: Omschrijving in validatierapport van methodiek NS voor het bepalen van de inzet voor bodem- en streefwaarden.

Deze stappen worden onderstaand kort beschreven:

- ▶ In de **eerste stap** wordt bij interne experts van NS informatie opgehaald over externe en interne ontwikkelingen in de spoorsector. Deze ontwikkelingen worden opgedeeld in vijf onderwerpen: assets, dienstregeling, commercie, stations en overig. In deze context wordt een ontwikkeling gedefinieerd als elk effect dat mogelijk een invloed zou kunnen hebben op de dienstverlening van de NS, of (direct of indirect) op de KPI's die zijn vastgesteld. Daarbij kan het gaan om effecten waar NS geen, of weinig, invloed op heeft, zoals bijvoorbeeld werkzaamheden aan het spoor of

³ Validatierapport Rebel d.d. 25 mei 2020, p. 4.

⁴ Alleen de onderdelen uit het validatierapport die van belang zijn voor dit voorliggende onderzoek zijn opgenomen. Zo is er bijvoorbeeld geen samenvatting opgenomen over de conclusies omtrent de methodiek om te komen tot een prognose van de reizigersgroei.

⁵ Validatierapport Rebel d.d. 25 mei 2020, p. 9-11.

reizigersgroei. Daarnaast omvatten deze ontwikkelingen ook alle (deels) beheersbare effecten, zoals bijvoorbeeld het invoeren van nieuw materieel.

- ▶ In **stap twee** wordt gekeken welke van de geïdentificeerde ontwikkelingen op welk(e) cluster(s) invloed hebben.
- ▶ In de **derde stap** wordt voor iedere ontwikkeling de kans dat deze plaatsvindt, de kans dat dit impact heeft op de KPI, en de orde grootte van de impact bepaald. Dit gebeurt aan de hand van een expert opinion en data-analyse door NS zelf. Voor elk van de vijf onderwerpen wordt een totale impact bepaald aan de hand van de impact van de onderliggende ontwikkelingen. Het rapport beschrijft expliciet dat er in de destijds gehanteerde methodiek geen overzicht van de *drivers* van een KPI werd gehanteerd. Drivers zijn de parameters die de dynamiek van een KPI bepalen. Het zijn de bouwstenen die onderdeel uitmaken van de berekening die gedaan wordt om tot een waarde van een KPI te komen. Voor de KPI zitplaatskans is bijvoorbeeld het aantal aangeboden zitplaatsen een driver. De ontwikkelingen die zijn geïdentificeerd kunnen invloed uitoefenen op de drivers, welke op hun beurt weer zorgen voor een verandering in de KPI waartoe zij behoren.
- ▶ In **stap vier** worden de inzichten op de KPI's samengebracht en gevalideerd door senior management en KPI-eigenaren van NS.

2.3 Conclusies validatierapport

De **hoofdconclusie** van het validatierapport 2020 is dat *“uit de door de NS beschikbaar gestelde documentatie onvoldoende blijkt dat NS een valide methodiek heeft toegepast om tot een reële inzet voor bodem- en streefwaarden te komen.”* Hierbij wordt gesteld dat de methode om te komen tot een complete set aan ontwikkelingen in de basis valide is, maar dat de aanpak voor het toewijzen hiervan aan KPI-clusters onvoldoende herleidbaar is. Daarnaast is onvoldoende duidelijk op basis van welke aspecten, aannames en uitgangspunten NS de prognose voor de KPI-prestatie heeft vertaald in een inzet voor bodem- en streefwaarden.

Deze hoofdconclusie is gebaseerd op vijf onderliggende bevindingen, die hieronder beknopt zijn weergegeven:

- ▶ **NS heeft relevante ontwikkelingen op een valide wijze in kaart gebracht.** Meerdere interne experts brengen interne en externe ontwikkelingen in kaart. Het is aannemelijk dat deze experts in staat zijn om tot een goed inzicht van ontwikkelingen te komen.
- ▶ **NS gebruikt geen raamwerk van invloedsfactoren om ontwikkelingen aan KPI's te koppelen.** Bij het koppelen van ontwikkelingen aan KPI's maakt NS geen gebruik van een vooraf opgesteld overzicht van invloedsfactoren (belangrijke drivers) van de KPI. Hierdoor is niet vast te stellen of met alle belangrijke factoren rekening is gehouden.
- ▶ **De geschatte impact van ontwikkelingen op KPI's is onvoldoende onderbouwd.** De inzet van eigen medewerkers voor deze inschatting is valide. Daarnaast wordt er gebruik gemaakt van data-analyses, maar wordt er niet onderbouwd hoe deze een bijdrage leveren aan de totstandkoming van de geschatte impact. Er wordt niet onderbouwd waarom er voor het merendeel van de ontwikkelingen slechts één impact voor een cluster als geheel wordt gegeven. De opgenomen afwegingen bieden onvoldoende onderbouwing voor de vastgestelde orde grootte. Een paar specifieke punten die worden aangedragen zijn:
 - ▶ Voor de KPI's 'sociale veiligheid' is niet omschreven wat de verwachting is van niet door de NS te beïnvloeden aspecten.
 - ▶ Voor de KPI 'aantal drukke treinen' zijn de verwachte ontwikkelingen niet voorzien van een impact. Ook is NS hiervoor wel in staat om het aantal drukke treinen per dag op specifieke

trajecten vast te stellen, maar is niet navolgbaar hoe dit wordt gebruikt in het komen tot een prognose.

Naast het feit dat niet inzichtelijk is of en hoe de mogelijkheden die de NS heeft om geavanceerde data-analyses uit te voeren worden gebruikt om tot prognoses te komen, is ook onduidelijk hoe NS omgaat met onzekerheden.

- ▶ **NS heeft niet inzichtelijk gemaakt hoe er tot een prognose per KPI is gekomen.** Er wordt gesteld dat de impact van individuele ontwikkelingen niet opgeteld kan worden tot een totale KPI-prognose, maar er is geen zicht op welke wijze dit dan wel gebeurt. De enige beschrijving is dat dit gebeurt op basis van een expert opinion van eigen medewerkers.
- ▶ **NS toont niet hoe zij tot een inzet voor bodem- en streefwaarden is gekomen.** Het is niet inzichtelijk hoe er op basis van een prognose tot bodem- en streefwaarden wordt gekomen. Het NS-management is hiervoor bepalend, maar de informatie waarop zij dit bepalen is niet duidelijk.

/ 3 Verbeterde methodiek

3.1 Eisen aan methodiekbeschrijving

Voorafgaand aan het doorlopen van het verbetertraject is – mede ingegeven door de eerder geconstateerde tekortkomingen – bepaald waaraan de methodiekbeschrijving om te komen tot (een bandbreedte voor) bodem- en streefwaarden moet voldoen. De methodiek moet transparant, navolgbaar, uitlegbaar en compleet zijn om te kwalificeren als een valide methodiek. In de onderstaande kolom zijn deze begrippen nader geoperationaliseerd. Deze begrippen waren richtinggevend in het verbetertraject.

Onderwerp	Criteria
Transparantie	<ul style="list-style-type: none"> ▶ Er is expliciet vastgelegd wat de functie is van de verschillende onderzoeksmethoden in het komen tot een finale prognose. ▶ Er is vastgelegd welke werknemers op welke manier betrokken zijn bij welke stappen.
Navolgbaarheid	<ul style="list-style-type: none"> ▶ Met behulp van de documentatie zou een derde partij het volledige proces in principe zelf kunnen doorlopen. ▶ Indien de prognose een tweede keer, volgens dezelfde methodiek, zou worden gemaakt, zou dit hetzelfde resultaat geven.
Uitlegbaarheid	<ul style="list-style-type: none"> ▶ De methode is consistent en bevat geen interne tegenstrijdigheden.
Compleetheid	<ul style="list-style-type: none"> ▶ De methode neemt alle aspecten die van belang zijn voor het komen tot bodem- en streefwaarden voor de KPI's in ogenschouw.

In de volgende paragraaf staat de verbeterde methodiek beschreven. Vervolgens wordt inzichtelijk gemaakt hoe de eerder geconstateerde tekortkomingen zijn weggenomen. In de concluderende paragraaf wordt ten slotte teruggekomen op bovenstaande eisen.

3.2 Beschrijving verbeterde methodiek

Methodiek op hoofdlijnen

Voor het bepalen van bodem- en streefwaarden wordt voor iedere KPI eenzelfde proces doorlopen. Gedurende dit onderzoek heeft NS in een zelfstandig leesbaar document een beschrijving van dit proces ontwikkeld. Onderstaand is een samenvatting van deze beschrijving opgenomen. Op hoofdlijnen ziet de verbeterde methodiek er als volgt uit:



Hieronder wordt per stap toelichten wat er wordt gedaan. Vanuit het oogpunt van (bedrijfs)vertrouwelijkheid wordt dit beknopt beschreven en wordt niet op details ingegaan. Om de lezer een concreet beeld te geven van de betekenis van bepaalde stappen, zijn er een aantal voorbeelden opgenomen in blauwe kaders. Deze voorbeelden betreffen fictieve situaties.

Stap 1. Bepalen relevante ontwikkelingen voor NS

In de eerste stap wordt er een overzicht gemaakt van alle ontwikkelingen die gedurende de concessieperiode naar verwachting plaats zullen vinden. Het gaat daarbij zowel om interne als externe ontwikkelingen. Bij het opstellen van de lijst met ontwikkelingen zijn verschillende afdelingen binnen NS betrokken, zodat alle expertise en kennis over ontwikkelingen meegenomen kan worden. Zoals eerder beschreven zijn omvatten deze ontwikkelingen alle effecten die mogelijk een invloed kunnen hebben op de dienstverlening van de NS, of (direct of indirect) op de KPI's. Dit omvat zowel effecten waar NS invloed op heeft als effecten waar NS geen of weinig invloed op heeft. Per ontwikkeling wordt bepaald wanneer deze zich naar verwachting zal voordoen, en hoe zeker dit is. Naast de lijst met ontwikkelingen hanteert NS ook een overzicht van alle *drivers* per KPI. Zoals eerder beschreven kunnen de drivers worden gedefinieerd als de parameters die de dynamiek van een KPI bepalen. Zij kunnen worden gezien als de bouwstenen die nodig zijn om de waarde van een KPI te kunnen berekenen. Als men de drivers kent, kent men dus in feite ook de KPI. Hoewel de drivers dus direct een KPI beïnvloeden, beïnvloeden ontwikkelingen de drivers, en daarmee dus indirect de een KPI. Het overzicht van alle drivers (per KPI) is een resultaat van analyses die in de loop van de jaren zijn gedaan en de kennis die daarmee is opgedaan.

Voorbeeld

De waarde van de KPI zitplaatskans wordt berekend aan de hand van een groot aantal bouwblokken of *drivers*. De zitplaatskans hangt af van het aantal reizigers enerzijds, en het aantal zitplekken anderzijds. Enkele drivers van deze KPI zijn bijvoorbeeld de ritprijs, autonome groei, en de perronlengte. Een voorbeeld van een *ontwikkeling* die invloed kan hebben op deze laatste driver, zijn geplande werkzaamheden die erop gericht zijn om een perron te verlengen.

Stap 2. Berekenen impact op KPI's

Vervolgens wordt van alle ontwikkelingen die met redelijke zekerheid zullen plaatsvinden bepaald wat de impact op de verschillende KPI's is. Dit gebeurt in drie stappen:

- A. Voor elke KPI wordt voor elke ontwikkeling bepaald of deze ontwikkeling wel of geen invloed kan hebben op de KPI.

Voorbeeld

De ontwikkeling "snelheidsverhoging op een bepaald traject" heeft geen invloed op (een van de drivers) van de KPI Zitplaatskans, maar heeft potentieel wel invloed op (een van de drivers van) de KPI Reizigerspunctualiteit.

- B. Voor ontwikkelingen die invloed hebben op de KPI wordt vervolgens de orde grootte van de impact bepaald. Hiervoor wordt gebruik gemaakt van vier categorieën van impact: geen, klein, middel en groot. Deze inschatting wordt gemaakt door de ontwikkeling te vergelijken met soortgelijke ontwikkelingen uit het verleden. Er wordt zowel een optimistische als een pessimistische inschatting gemaakt.

Voorbeeld

Onderstaand is een lijst van fictieve ontwikkelingen op een KPI weergegeven, waarbij een fictieve impact in beide scenario's is opgenomen.

Ontwikkeling	Impact optimistisch scenario	Impact pessimistisch scenario
Ontwikkeling 1	Geen	Klein (negatief)
Ontwikkeling 2	Middel (negatief)	Groot (negatief)
Ontwikkeling 3	Klein (positief)	Klein (negatief)

- C. Ten slotte worden de categorieën vertaald naar kwantitatieve bandbreedtes. Hierbij wordt gekeken naar de werkelijke impact van vergelijkbare ontwikkelingen uit het verleden. Indien een verwachte ontwikkeling exact gelijk is aan een ontwikkeling uit het verleden, wordt de exacte impact uit het verleden gehanteerd.

Voorbeeld

Onderstaand is een fictief overzicht opgenomen van ordegroottes van impact op een fictieve KPI.

Impact ordegrootte	Impact definitie	Onderbouwing / voorbeeld
Geen	0%	Verandering van omroepstem op stations
Klein	0% - 0,1%	Introductie van nieuw sprintermaterieel
Middel	0,1% - 0,5%	Kortere treinen tussen Zwolle en Groningen
Groot	> 0,5%	Grootschalige werkzaamheden op Utrecht CS

Voor elke KPI bestaat er na deze stap een overzicht van alle ontwikkelingen met de impact (veelal in een bandbreedte) die zij naar verwachting zullen hebben op de KPI in zowel een optimistisch als een pessimistisch scenario.

Ontwikkelingen die onzeker zijn worden niet op dezelfde manier meegenomen, omdat dit leidt tot zeer grote onzekerheden bij de inschattingen van de impact. Het ministerie van IenW en NS zijn met elkaar in gesprek over hoe er met dit soort onzekere ontwikkelingen moet worden omgegaan.

Stap 3. Uitwerken impact tot bandbreedte

De impact die in stap 2 is gekoppeld aan de ontwikkelingen is uitgedrukt in een bandbreedte. Om al deze ontwikkelingen op te tellen tot een uiteindelijke impact op een KPI, moet elke ontwikkeling een enkele kwantitatieve waarde hebben. Om tot deze waarde te komen wordt steeds het middelpunt genomen van de ordegrootte-categorie van een ontwikkeling. Voor ontwikkelingen die sterk lijken op ontwikkelingen uit het verleden, wordt de exacte impact uit het verleden genomen. Nadat zo aan elke ontwikkeling een kwantitatieve impact is gekoppeld, worden alle ontwikkelingen opgeteld in zowel het optimistische als het pessimistische scenario. Dit geeft een totale impact op de KPI in beide scenario's. Het verschil tussen de scenario's is de bandbreedte van de impact op de KPI.

Ten slotte wordt de impact op een KPI vertaald naar een verwachte waarde van deze KPI. Het uitgangspunt hiervoor is de waarde van de betreffende KPI uit het laatste representatieve jaar, gecorrigeerd met een prognose voor de latere jaren tot aan de start van de desbetreffende concessieperiode. Door de (bandbreedte van de) verwachte impact op te tellen bij dit uitgangspunt, ontstaat een bandbreedte voor de verwachte waarde van de KPI.

Voorbeeld

Voor dit voorbeeld worden de fictieve ontwikkelingen uit het voorbeeld bij stap 2B gebruikt. De definities van ordegrootten die voor dit voorbeeld gehanteerd worden, zijn die uit het voorbeeld bij stap 2C. Voor de ordegrootten 'klein' en 'middel' wordt het middelpunt van de bandbreedte gebruikt (respectievelijk 0,05% en 0,3%). Voor de ordegrootte 'groot' wordt de waarde 0,75% genomen. De totale impact van de drie ontwikkelingen uit het eerdere voorbeeld is dan als volgt:

Optimistisch scenario: 0% (geen) – 0,3% (middel) + 0,05% (klein) = –0,25%

Pessimistisch scenario: – 0,05% (klein) – 0,75% (groot) – 0,05% (klein) = –0,85%

Als fictieve uitgangswaarde van de KPI wordt 80% aangenomen voor dit voorbeeld. Dat betekent dat de verwachte waarde voor deze KPI over de periode van de concessie tussen de 79,15% en de 79,75% ligt. Dit is de bandbreedte van deze KPI.

Stap 4. Intern valideren

In stap 4 wordt een intern validatieproces doorlopen waarbij er voor elke KPI op meerdere manieren kritisch wordt gekeken naar de uitwerking en invulling. In het datateam van NS ligt de focus op het challengen van de cijfers en de berekeningen. In sessies met stakeholders die onderdeel zijn van de KPI-keten (zowel interne stakeholders van NS, zoals andere afdelingen, alsook externe stakeholders, zoals ProRail) worden met name de onderliggende aannames besproken en wordt getoetst of deze consistent zijn. Vervolgens wordt er afgestemd met de verantwoordelijke directeur en het MT binnen NS. De finale validatie wordt uiteindelijk gedaan door de directeuren en de RvB van NS. Tijdens deze stap is er veel aandacht voor het meenemen van de directeuren en de RvB van NS in de aannames en inschattingen die zijn gedaan, met ruimte voor kritische vragen. Hier vinden geen grote wijzigingen meer plaats. Eén van de weinige wijzigingen die in deze fase nog plaats zouden kunnen vinden, is de beslissing of een bepaalde (relatief onzekere) exogene ontwikkeling binnen een bandbreedte meegenomen dient te worden, of dat het de voorkeur heeft om hier met het ministerie van IenW over in gesprek te gaan.

Stap 5. Onderhandelen

Na deze vier stappen is voor elke KPI een bandbreedte bekend. De manier om van deze bandbreedte te komen tot bodem- en streefwaarden hangt af van de afspraken die er met het ministerie van IenW zijn gemaakt. In de huidige concessieperiode zijn de bodem- en streefwaarden op basis van onderhandeling tot stand gekomen. In dit geval zijn de bandbreedtes van de KPI input waarmee NS de onderhandelingen in gaat.

3.3 Verbetering eerdere tekortkomingen

In de procesbeschrijving die gedurende dit traject is opgesteld (waarvan de beschrijving hierboven een samenvatting is) is voldoende herleidbaar op welke wijze NS tot een reële bandbreedte voor de bodem- en streefwaarden is gekomen. In het bijzonder zijn de eerder geconstateerde tekortkomingen in het validatierapport 2020 weggenomen. Dit wordt hieronder specifiek toegelicht. Hierbij wordt eerst beknopt de geconstateerde tekortkoming herhaald, waarna vervolgens (in de tekstvakken) aan wordt gegeven hoe dit is geadresseerd in de huidige methodiek.

1. NS gebruikt geen raamwerk van invloedsfactoren om ontwikkelingen aan KPI's te koppelen.

Bij het koppelen van ontwikkelingen aan KPI's maakt NS geen gebruik van een vooraf opgesteld overzicht van invloedsfactoren (belangrijke drivers) van de KPI. Hierdoor is niet vast te stellen of met alle belangrijke factoren rekening is gehouden.

Voor iedere KPI hanteert NS een overzicht van alle relevante drivers. Of dit overzicht ten tijde van het validatierapport van Rebel al bestond, maar niet werd omschreven, is niet duidelijk. Feit is in ieder geval dat dit overzicht momenteel bestaat, wordt gebruikt, en ook als dusdanig is opgenomen in de documentatie van het proces. Daarnaast bestaat er een lijst van alle verwachte ontwikkelingen, waarvan wordt bezien op welke drivers zij effect hebben, en daarmee ook op welke KPI's zij invloed hebben. Voor de totstandkoming van deze ontwikkelingen (en het bepalen op welke drivers zij invloed hebben) worden medewerkers vanuit verschillende afdelingen van NS betrokken. Er is in de huidige opzet van het proces dus sprake van een vooraf vastgesteld overzicht van ontwikkelingen en een vooraf inzichtelijk overzicht van drivers per KPI. Bij dit proces worden veel verschillende afdelingen van NS betrokken, zoals bijvoorbeeld dienstregeling, infrastructuur, en IT / innovatie. Doordat alle relevante afdelingen hierbij zijn betrokken, wordt er vanuit verschillende perspectieven gekeken en is het aannemelijk dat met alle belangrijke factoren rekening wordt gehouden. Deze methodiek is oorspronkelijk ontwikkeld voor twee KPI's, en heeft historisch bewezen om behulpzaam te zijn. Op basis van deze ervaringen is in het verleden besloten om de methode uit te breiden naar alle KPI's.

2. De geschatte impact van ontwikkelingen op KPI's is onvoldoende onderbouwd.

De inzet van eigen medewerkers voor deze inschatting is valide. Daarnaast wordt er gebruik gemaakt van data-analyses, maar wordt er niet onderbouwd hoe deze een bijdrage leveren aan de totstandkoming van de geschatte impact.

De documentatie van het proces bevat een beschrijving van de rol die data speelt in het tot stand komen van de geschatte impact op een KPI. Medewerkers van NS bepalen of een ontwikkeling naar verwachting impact heeft op een KPI, en koppelen vervolgens een verwachte ordegraad aan elke ontwikkeling. Dit laatste doen zij op basis van ervaringen met vergelijkbare ontwikkelingen uit het verleden. Ten slotte wordt aan elke ordegraad een kwantitatieve bandbreedte gekoppeld. Hiervoor wordt gebruik gemaakt van eigen data-analyses. De wijze waarop data-analyses leiden tot de geschatte impact, en de rol die medewerkers daarin hebben, is dus helder vastgelegd in de huidige documentatie.

Er wordt niet onderbouwd waarom er voor het merendeel van de ontwikkelingen slechts één impact voor een cluster als geheel wordt gegeven. De opgenomen afwegingen bieden onvoldoende onderbouwing voor de vastgestelde ordegraad.

In het huidige proces wordt er voor elke ontwikkeling gekeken op welke drivers deze invloed heeft, en daarmee op welke KPI er een effect wordt verwacht. Er wordt dus niet langer gewerkt met een impact op clusters, maar louter met impact op individuele KPI's. Dit is ook vastgelegd in de documentatie.

Een paar specifieke punten die worden aangedragen zijn:

- ▶ Voor de KPI's 'sociale veiligheid' is niet omschreven wat de verwachting is van niet door de NS te beïnvloeden aspecten.
- ▶ Voor de KPI 'aantal drukke treinen' zijn de verwachte ontwikkelingen niet voorzien van een impact. Ook is NS hiervoor wel in staat om het aantal drukke treinen per dag op specifieke trajecten vast te stellen, maar is niet navolgbaar hoe dit wordt gebruikt in het komen tot een prognose.

In dit onderzoek is gekeken naar de generieke methodiek die op alle KPI's van toepassing is, en is niet ingegaan op individuele KPI's, behalve als voorbeeld om een begrip te vormen van de wijze waarop de methodiek in de praktijk werkt. Alle verwachte ontwikkelingen met een minimaal redelijke zekerheid zijn voorzien van een impact. De wijze waarop men tot prognoses komt (bijvoorbeeld voor de genoemde KPI 'aantal drukke treinen') is door de effecten van ontwikkelingen op drivers te bepalen. Dit gebeurt in het genoemde geval door middel van geijkte modellen en berekeningen. Doordat de drivers op een vastgestelde manier gekoppeld zijn aan KPI's, volgt hieruit ook de ontwikkeling van de KPI.

Naast het feit dat niet inzichtelijk is of en hoe de mogelijkheden die de NS heeft om geavanceerde data-analyses uit te voeren worden gebruikt om tot prognoses te komen, is ook onduidelijk hoe NS omgaat met onzekerheden.

Voor een procesbeschrijving op hoofdlijnen is het niet noodzakelijk om de technische details van de data-analyses inzichtelijk te hebben. Voor wat betreft onzekerheden is in de huidige documentatie helder vastgelegd hoe NS hiermee om gaat. Onzekerheden van ontwikkelingen komen tot stand door de gehanteerde categorieën van ordegroottes te koppelen aan ervaringen met soortgelijke ontwikkelingen uit het verleden. Hiervoor wordt voor elke ontwikkeling een optimistisch en pessimistisch scenario gehanteerd. In beide scenario's wordt vervolgens voor elke ontwikkeling het middelpunt van de bandbreedte van de categorie van ordegrootte genomen. Door al deze waarden (sommige negatief, sommige positief) op te tellen, ontstaat er voor beide scenario's een totale verwachte impact. Het verschil tussen beide scenario's is ten slotte de totale bandbreedte van de impact op de KPI.

3. NS heeft niet inzichtelijk gemaakt hoe er tot een prognose per KPI is gekomen. Er wordt gesteld dat de impact van individuele ontwikkelingen niet opgeteld kan worden tot een totale KPI-prognose, maar er is geen zicht op welke wijze dit dan wel gebeurt. De enige beschrijving is dat dit gebeurt op basis van een expert opinion van eigen medewerkers.

In de momenteel gehanteerde methodiek worden de individuele ontwikkelingen (in beide scenario's) opgeteld tot een totale impact. Zie hiervoor ook de eerdere beschrijving. De wijze waarop NS van impact op een KPI naar de verwachte waarde van deze KPI komt is ook helder beschreven in de documentatie. Hiervoor wordt uitgegaan van de historische data uit het laatste representatieve jaar, gecorrigeerd voor een prognose voor de jaren tot aan de start van de desbetreffende concessieperiode. Door de verwachte impact op te tellen bij deze 'basis', ontstaat de verwachte waarde van elke KPI. Het betreft dus een vaste methodiek, en gebeurt niet op basis van expert opinions van medewerkers van NS.

4. NS toont niet hoe zij tot een inzet voor bodem- en streefwaarden is gekomen. Het is niet inzichtelijk hoe er op basis van een prognose tot bodem- en streefwaarden wordt gekomen. Het NS-management is hiervoor bepalend, maar de informatie waarop zij dit bepalen is niet duidelijk.

Het NS-management is niet bepalend in het komen tot bodem- en streefwaarden. Wel voeren zij de finale validatie uit op de bandbreedte voor de verschillende KPI's. Hierbij worden ze meegenomen in de verschillende afwegingen, inschattingen en aannames die zijn gedaan en kunnen ze kritische vragen stellen. Bij deze validatie vinden er geen grote wijzigingen meer plaats. Een gesprek dat in deze fase nog plaats kan vinden, is of een exogene ontwikkeling binnen een bandbreedte meegenomen moet worden, of dat NS hier aanstuurt op aparte afspraken met het ministerie van IenW.

Na de validatie is er een set van KPI's ontstaan, waarbij voor elke KPI een bandbreedte geldt waarbinnen deze zich naar verwachting zal ontwikkelen tijdens de nieuwe concessieperiode. De exacte wijze waarop men tot bodem- en streefwaarden komt hangt af van de afspraken die er met het ministerie van IenW worden gemaakt. In de huidige concessieperiode zijn deze definitieve bodem- en streefwaarden op basis van onderhandelingen tot stand gekomen.

3.4 Conclusie

In het eerdere validatierapport uit 2020 werden aan de methodiek van de NS om te komen tot bodem- en streefwaarden een aantal tekortkomingen geconstateerd. Deze tekortkomingen zagen met name op de herleidbaarheid van het proces dat NS volgde om tot een inzet voor bodem- en streefwaarden voor KPI's te komen. In de onderliggende documentatie waren niet alle stappen even duidelijk en miste bepaalde onderdelen van de methodiek.

In samenwerking met NS heeft AEF gewerkt aan een verbeterde vastlegging van het proces. Zoals beschreven in de vorige paragraaf ligt er nu een procesbeschrijving waarin de alle geconstateerde tekortkomingen zijn weggenomen.

Om daadwerkelijk te kunnen concluderen of bij de huidige (beschrijving van de) methodiek sprake is van een 'valide methodiek', wordt hier teruggerepen op de eerder geformuleerde eisen in paragraaf 3.1. Onderstaand wordt voor elk van de eisen kort beschreven in welke mate de huidige methodiek hieraan voldoet:

- ▶ **Transparantie.** De documentatie omvat het volledige proces om te komen tot bodem- en streefwaarden. Hierin is een generieke beschrijving opgenomen, waarbij ter verduidelijking concrete voorbeelden worden genoemd. Bij elke stap is opgenomen welke afdelingen of werknemers betrokken zijn, en welke onderzoeksmethoden worden gehanteerd.
- ▶ **Navolgbaarheid.** Indien een derde partij zou beschikken over de huidige documentatie en de achterliggende data van de NS, zou deze partij exact hetzelfde proces kunnen volgen. In principe zou deze partij dan uitkomen op dezelfde prognoses, met de kanttekening dat er dan ook gebruik gemaakt moet worden van eenzelfde expertise als waar de medewerkers van de NS over beschikken.
- ▶ **Uitlegbaarheid.** Er wordt gewerkt met een logisch systeem van drivers, ontwikkeling en uiteindelijk impact op KPI's. De impact van ontwikkelingen wordt (deels) ingeschat op basis van resultaten uit het verleden en sluit dus inhoudelijk zo goed mogelijk aan bij de werkelijkheid. De

interne validatie die onderdeel uitmaakt van het proces zorgt ervoor dat aannames consistent worden gedaan en er geen interne tegenstrijdigheden ontstaan.

- ▶ **Compleetheid.** De beschrijving van de methodiek gaat uit van het gebruikte framework van drivers en KPI's. Het proces begint vervolgens met het in kaart brengen van alle mogelijke ontwikkelingen en doorloopt alle stappen tot het komen tot een bandbreedte van de verwachte waarden van een KPI. De laatste stap is om van deze verwachting te komen tot bodem- en streefwaarden. Momenteel zijn deze bodem- en streefwaarden uitkomst van een onderhandelingsproces.

Naast het feit dat met de beschrijving van de huidige methodiek de eerder geconstateerde tekortkomingen worden weggenomen, kan AEF dus ook concluderen dat de methodiek die de NS hanteert aan alle bovenstaande eisen voldoet. Aangezien deze eisen gezamenlijk de definitie vormen van een 'valide methodiek' (zie paragraaf 3.1), kan AEF concluderen dat NS een valide methodiek hanteert om te komen tot een (bandbreedte voor) bodem- en streefwaarden.

/ A De huidige KPI's

In de huidige vervoerconcessie 2015-2025 wordt gewerkt met de tien onderstaande kritische prestatie-indicatoren (KPI's). Deze KPI's zijn geordend naar vijf prestatiegebieden.

Prestatiegebieden	KPI's
Algemeen	<ul style="list-style-type: none">▶ Algemeen klantoordeel▶ Algemeen klantoordeel sociale veiligheid
Deur-tot-deur reis	<ul style="list-style-type: none">▶ Kwaliteit van aansluitingen NS op andere vervoerders
Reisgemak	<ul style="list-style-type: none">▶ Zitplaatskans in de spits HRN▶ Zitplaatskans in de spits HSL▶ Aantal drukke treinen per werkweek in de spits HRN
Reisinformatie	<ul style="list-style-type: none">▶ Gebruiksvriendelijke reisinformatie
Betrouwbaarheid	<ul style="list-style-type: none">▶ Reizigerspunctualiteit HRN '5 min.▶ Reizigerspunctualiteit HRN '15 min.▶ Reizigerspunctualiteit HSL '5 min.