

## **Een beter OV begint met een beter bestek**

dr. ir. Niels van Oort  
Goudappel Coffeng / TU Delft  
NvOort@Goudappel.nl

**Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk  
21 en 22 november 2013, Rotterdam**

## **Samenvatting**

### *Een beter OV begint met een beter bestek*

Om een hoogwaardige OV exploitatie te bereiken, stellen overheden diverse eisen aan vervoerders met betrekking tot de exploitatie van het OV. Dat gebeurt zowel vooraf (bij aanbestedingen via bestekken) als tijdens de exploitatie (via concessie-eisen). Deze eisen zijn van grote invloed op het gedrag van de vervoerder voor wat betreft planning en bijsturing en daarmee op de kwaliteit op straat. Hoewel via de toolbox "Beter Bestek" van KpVV afgelopen jaren grote sprongen in kwaliteit en uniformiteit van OV-bestekken zijn gemaakt, geldt dit nog niet voor het aspect "op tijd rijden". De aandacht is hiervoor weliswaar sterk toegenomen- de punctualiteit van NS is bijvoorbeeld een veelbesproken onderwerp- maar een uniforme aanpak en indicatoren ontbreken echter in Nederland.

In dit paper wordt onderzoek gepresenteerd dat inzicht biedt in de stand van zaken op dit moment voor wat betreft betrouwbaarheidseisen in bestekken (zowel in Nederland als in het buitenland). Hieruit blijkt dat er verschillende (suboptimale) aanpakken voorkomen. Op basis van de onderzoeken worden handvatten gepresenteerd om beter te kunnen sturen op een betrouwbaar OV, zowel bij aanbesteding als tijdens de exploitatie, met als resultaat beter OV op straat.

De belangrijkste aanbevelingen zijn om bij het vaststellen van eisen voor wat betreft het op tijd rijden een referentie- en gewenst kwaliteitsniveau vast te stellen, met daarbij een analyse wat haalbare verbeteringen zijn en wie daar invloed op heeft. Zo hebben OV autoriteit, gemeente en vervoerder elk hun rol en mogelijkheden. Opvallend is dat er momenteel weinig aandacht is voor de kwaliteit van de dienstregeling met betrekking tot punctualiteit, hoewel daar een sleutel voor succes ligt. Tot slot zijn de gebruikte indicatoren momenteel niet voldoende op de reiziger gericht, waardoor suboptimalisatie op de loer ligt. De veelgebruikte indicator punctualiteit houdt bijvoorbeeld niet direct rekening met fenomenen als te vroeg rijden, overstappen en regelmaat. In dit paper laten we rekenkundig zien wat de impact van verschillende, suboptimale indicatoren is. Om tegemoet te komen aan beperkingen van huidige indicatoren wordt in dit paper een indicator gepresenteerd, extra reistijd per reiziger, die wel focus op de reizigerseffecten biedt en daarmee een beter ontwerp en uitvoering van OV faciliteert.

Door in OV-bestekken te werken met goede indicatoren en ambitieuze, doch realistische eisen komt een beter OV snel dichterbij. Een beter OV begint immers met een beter bestek.

## 1. Inleiding

Het op tijd rijden van het OV krijgt steeds meer aandacht (Tahmasseby et. al 2007 en Van Oort 2011). Naast de snelheid van de reis speelt de onzekerheid over de aankomsttijd een belangrijke rol bij de keuze voor bijvoorbeeld auto of trein. Zowel in Nederland als in de rest van de wereld staat het OV bekend als (te) onbetrouwbaar. In Van Oort (2011) is een framework gepresenteerd waarin duidelijk wordt wat betrouwbaarheid vanuit reizigersperspectief behelst en hoe netwerken en dienstregelingen te ontwerpen die een betrouwbaardere reis faciliteren.

Betrouwbaarheid is de mate waarin de geleverde diensten overeenkomen met de verwachtingen van de reizigers. De mate van de dynamiek van de uitvoering bepaalt de betrouwbaarheid voor wat betreft reistijd en heeft voor reizigers grosso modo drie effecten:

- de gemiddelde reistijd wordt langer;
- de spreiding in de reistijd wordt groter en daarmee de aankomsttijd onzekerder;
- de kans op een zitplaats/mate van comfort neemt af.

Om te komen tot een betrouwbaar(der) OV is, naast het ontwerp van OV, ook van belang hoe het OV door de overheid wordt aangestuurd: welke eisen worden er gesteld en welk gedrag van vervoerders heeft dat tot gevolg? Om de aansturing van OV te verbeteren en te leren van ervaringen van elkaar is KpVV begonnen met de toolbox Beter Bestek. Met concrete voorbeeldteksten worden overheden geholpen om een bestek te maken dat leidt tot het OV dat ze graag wensen (KpVV 2013). Vele aspecten van OV-kwaliteit komen in dit project aan bod, maar OV-betrouwbaarheid is onlangs pas op de agenda gekomen. Om inzicht in deze aspecten te krijgen zijn er twee onderzoeken uitgevoerd, zowel internationaal via TU Delft als nationaal via KpVV. In dit paper worden de resultaten van deze onderzoeken gepresenteerd en aanbevelingen gedaan hoe betere aansturing en eisen leiden tot een beter OV. Een beter OV begint immers bij een beter bestek.

## 2. Onderzoek eisen aan "op tijd rijden"

Om inzicht te krijgen hoe OV autoriteiten omgaan met betrouwbaarheid hebben we twee onderzoeken uitgevoerd. De eerste betrof een internationale en de tweede een nationale. In de volgende paragrafen worden deze verder toegelicht.

### 2.1 Internationaal onderzoek betrouwbaarheid

In Van Oort (2011) zijn de resultaten gepresenteerd van een internationaal onderzoek naar betrouwbaarheid. Het doel was om te leren van overheden over de gehele wereld voor wat betreft het omgaan met onbetrouwbaarheid. Bijna 30 steden en regio's werkten mee aan dit onderzoek, welke zijn te zien in tabel 1.

Tabel 1: Deelnemende steden/regio's internationaal onderzoek betrouwbaarheid (Van Oort (2011))

Stad	OV Type	Stad	OV Type	Stad	OV Type
Amsterdam	<i>Metro, tram, bus</i>	Goteburg	<i>Tram</i>	Rouen	<i>Tram en bus</i>
Barcelona	<i>Metro en bus</i>	Halle	<i>Tram en bus</i>	Salt lake city	<i>Light rail</i>
Berlijn	<i>Rail</i>	Hong Kong	<i>Light rail</i>	Stockholm	<i>Metro en bus</i>
Brussel	<i>Tram</i>	Lolland	<i>Bus</i>	Stuttgart	<i>Rail</i>
Chicago	<i>Metro en bus</i>	London	<i>Tram en bus</i>	Tenerife	<i>Tram</i>
Den Haag	<i>Light rail, tram en bus</i>	Milaan	<i>Bus en tram</i>	Zürich	<i>S-Bahn</i>
Dresden	<i>Rail</i>	Minneapolis	<i>Bus</i>		
Dublin	<i>Tram</i>	Rotterdam	<i>Metro, tram, bus</i>		

## 2.2 Nederlands onderzoek betrouwbaarheid in bestekken

Naast het internationale onderzoek is er i.s.m. KPVV, Goudappel Coffeng en TU Delft een tweede onderzoek uitgevoerd. Dit betrof specifiek onderzoek naar hoe betrouwbaarheid in Nederlandse OV bestekken wordt meegenomen. Het onderzoek bestond uit twee delen, te weten een desk research, waarin 22 willekeurige bestekken zijn onderzocht (zie figuur 1 voor de bijbehorende regio's) en interviews met overheden.



Figuur 1: Gebieden waarvan OV bestek is onderzocht

De overheden die zijn geïnterviewd zijn:

- Provincie Flevoland;
- Provincie Friesland;
- Provincie Gelderland;
- Provincie Limburg;
- Provincie Noord-Holland;
- Provincie Overijssel;
- Provincie Utrecht;
- Provincie Zeeland;
- Regio Arnhem-Nijmegen;
- Regio Eindhoven;
- Regio Groningen-Drenthe;
- Regio Twente.

In de volgende hoofdstukken worden de resultaten en inzichten van beide onderzoeken gepresenteerd.

## 3. Meten van betrouwbaarheid

### 3.1 Gebruikte indicatoren

Het meten van betrouwbaarheid van openbaar vervoer kan op verschillende manieren. Hoewel betrouwbaarheid een belangrijk aspect is voor reizigers, ligt de focus bij het meten van betrouwbaarheid toch vaak op het aanbod in plaats van op de effecten voor de reiziger. De meest bekende methode voor het in kaart brengen van de betrouwbaarheid is de stiptheid (ook wel punctualiteit). Stiptheid is de mate waarin vertrek- of aankomsttijden overeenkomen met de dienstregeling. Stiptheid kan grofweg op twee manieren worden uitgedrukt.

Een veelgebruikte vorm is het percentage ritten dat binnen een vooraf vastgestelde bandbreedte voor de dienstregelingafwijking valt. Vergelijking 1 laat deze vorm zien voor een hele lijn. Input daarvoor is exploitatiedata die vergeleken wordt met dienstregelingsdata. In de volgende paragraaf komen praktijkwaarden aan bod voor de onder- en bovengrens,  $\delta^{\min}$  and  $\delta^{\max}$ . Het is duidelijk dat deze een grote rol spelen in de waarde van het uiteindelijke percentage.

$$P_l = \frac{\sum_{j=1}^{n_{l,j}} \sum_{i=1}^{n_{l,i}} P_{l,i,j} (\delta^{\min} < \tilde{D}_{l,i,j}^{act} - D_{l,i,j}^{sched} < \delta^{\max})}{n_{l,i} * n_{l,j}} \quad (1)$$

waarin:

- $P_l$  = Relatieve frequentie van voertuigen op lijn l die een dienstregelingafwijking hebben tussen  $\delta^{\min}$  en  $\delta^{\max}$   
 $P_{l,i,j}$  = Relatieve frequentie van voertuig i op lijn l die een dienstregelingafwijking heeft tussen  $\delta^{\min}$  en  $\delta^{\max}$  op halte j  
 $\tilde{D}_{l,i,j}^{act}$  = werkelijk vertrek van voertuig I op halte j op lijn l  
 $D_{l,i,j}^{sched}$  = gepland vertrek van voertuig I op halte j op lijn l  
 $\delta^{\min}$  = ondergrens bandbreedte dienstregelingafwijking  
 $\delta^{\max}$  = bovengrens bandbreedte dienstregelingafwijking  
 $n_{l,i}$  = aantal ritten op lijn l  
 $n_{l,j}$  = aantal haltes op lijn l

Een tweede manier om de stiptheid te beschrijven is de (absoluut) gemiddelde afwijking van de dienstregeling. De berekening hiervoor is gegeven in vergelijking 2. Te zien is dat het absolute verschil tussen uitvoering en planning genomen dient te worden om te voorkomen dat te vroege en te late voertuigen elkaar compenseren. Te vroeg of te laat wordt wel hetzelfde gewaardeerd, ondanks het feit dat de negatieve effecten van te vroeg rijden over het algemeen groter zijn voor reizigers (Van Oort et. al 2012). De gemiddelde afwijking van de dienstregeling kan voor een specifieke halte of voor alle haltes en lijnen berekend worden.

$$\bar{p}_j = \frac{\sum_i |t_{i,j}^{werkelijk} - t_{i,j}^{gepland}|}{n_j} \quad (2)$$

waarin:

- $\bar{p}_j$  = gemiddelde punctualiteit op halte j  
 $t_{i,j}^{werkelijk}$  = werkelijke vertrektijd van voertuig i op halte j  
 $t_{i,j}^{gepland}$  = geplande vertrektijd van voertuig i op halte j  
 $n_j$  = aantal voertuigen op halte j

Zoals beschreven geeft stiptheid de mate van overeenkomst tussen de uitvoering en de dienstregeling aan. Dit is voor reizigers relevant om het moment dat ze ook daadwerkelijk de dienstregeling gebruiken om hun aankomstmoment op de vertrekhalte te bepalen. Bij spoorwegen is dit heel gebruikelijk, terwijl bij hoogfrequente stadssystemen reizigers vaak aselekt naar de halte lopen. Door korte intervallen is de wachttijd immers toch niet lang. Onderzoek in Den Haag (Van Oort 2011) toont dat het omslagpunt tussen wel en niet de dienstregeling hanteren ligt op 10 min. voor stedelijk openbaar vervoer. Door toenemende frequenties van treinen is bij de spoorwegen ook te verwachten dat het aankomstgedrag van reizigers op stations gaat veranderen: de rol van het spoorboekje zal afnemen.

Op het moment dat reizigers aselekt naar een halte gaan, wordt de wachttijd niet meer bepaald door de stiptheid maar door de regelmaat: de mate waarin de intervallen tussen

twee opeenvolgende voertuigen constant zijn. Dit kan worden uitgedrukt in de relatieve onregelmaat van een lijn (genaamd PRDM volgens Hakkesteege en Muller (1981)). Vergelijking 3 laat de berekening zien van de PRDM. Hierin wordt de afwijking van de intervallen gerelateerd aan het geplande interval.

$$PRDM_j = \frac{\sum \left| \frac{TIT_{i,j} - TIA_{i,j}}{TIT_{i,j}} \right|}{n_j} \quad (3)$$

waarin:

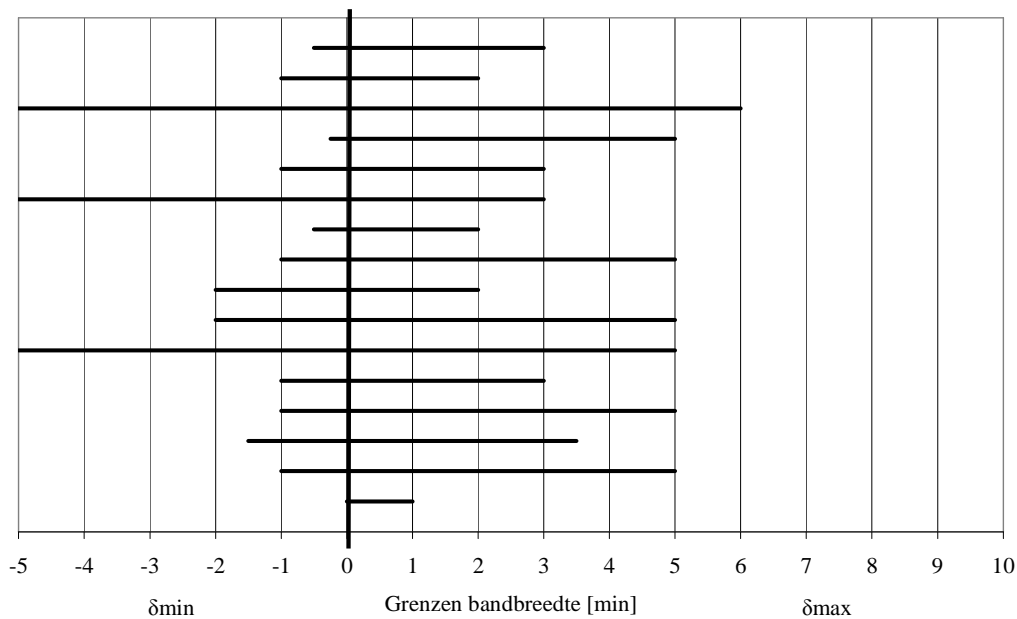
$PRDM_j$  = relatieve onregelmaat op halte  $j$   
 $TIA_{i,j}$  = werkelijk interval voertuig  $i$  op halte  $j$   
 $TIT_{i,j}$  = gepland interval voertuig  $i$  op halte  $j$   
 $n_j$  = aantal voertuigen op halte  $j$

Uit ons onderzoek blijkt dat punctualiteit een veelgebruikte indicator is om betrouwbaarheid in kaart te brengen: ca. 75% van de onderzochte steden/regio's gebruikt een bandbreedte om de punctualiteit te meten en ca. 20% de gemiddelde afwijking. Circa 5% meet betrouwbaarheid op een andere manier.

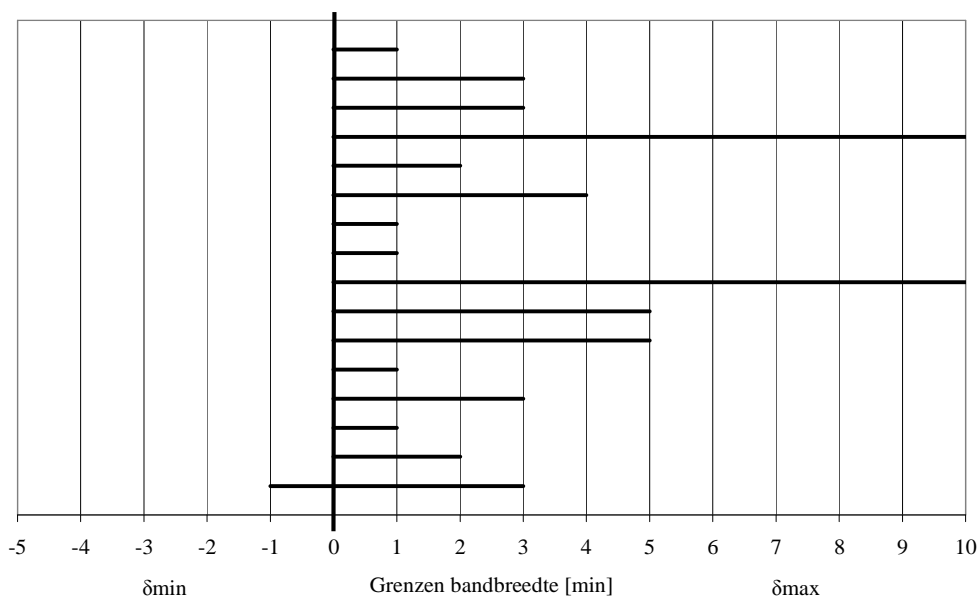
Het valt dus op dat punctualiteit (in twee vormen) veel gebruikt wordt, terwijl dat voor hoogfrequente systemen (met aselechte aankomsten van reizigers) een weinig relevante maat is. Reizigers kennen de dienstregeling immers niet. De regelmaat is dan veel belangrijker; een optimale regelmaat minimaliseert immers de wachttijd van reizigers (Van Oort 2011). Vaak is er bij het vaststellen van de punctualiteit ook geen sprake van weging naar aantal reizigers, maar naar aantal voertuigen. Volle of lege voertuigen tellen dus even zwaar. Tot slot blijft overstappen vaak buiten beschouwing. Een gemiste aansluiting kan echter een grote impact hebben op de reiziger. Het voorbeeld van de stipt vertrekkende trein die niet wacht op een licht vertraagde aansluitende trein is bekend. Verder op in dit paper wordt ingegaan op deze tekortkomingen.

### 3.2 Gebruikte drempelwaarden

Naast verscheidenheid in de indicatoren om betrouwbaarheid te meten is er ook nog een verschil in gebruikte drempelwaarden om een systeem als betrouwbaar te definiëren. Uit beide onderzoeken blijkt dat er ook grote verschillen zijn tussen gehanteerde drempelwaarden. Dit is te zien in figuur 2 (internationaal) en 3 (nationaal). In figuur 2 is te zien dat er grote verschillen in grenswaarden zijn en dat drie steden geen ondergrens voor stiptheid hanteren. Zes minuten is de hoogste bovengrens.



*Figuur 2: Gehanteerde bandbreedten voor meten van punctualiteit (internationaal)*



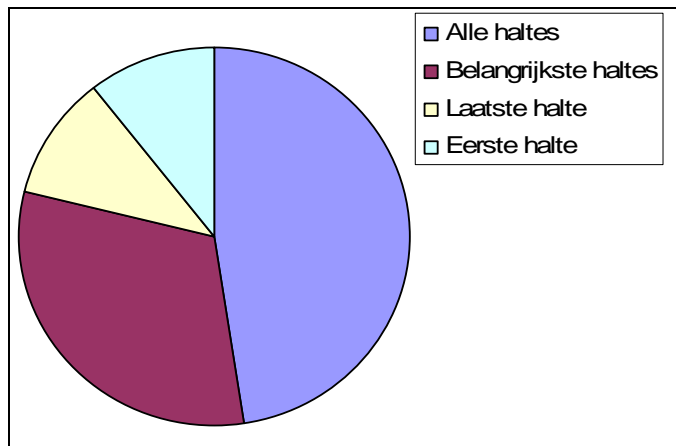
*Figuur 3: Gehanteerde bandbreedten voor meten van punctualiteit in Nederland*

Figuur 3 laat zien dat in Nederland de ondergrens scherp wordt benoemd (dus te vroeg vertrekken telt als onstipt), maar de maximum waarde varieert tussen de 1 en 10 minuten.

Zoals lange tijd onder andere bij de Nederlandse Spoorwegen gebruikelijk was, kan de stiptheid ook worden uitgedrukt als percentage treinen dat niet meer dan 3 minuten te laat is vertrokken. Bij veel andere Europese spoorbedrijven lag deze norm al langer wat ruimer: een voertuig telt pas als "te laat", wanneer het verschil tussen uitvoering en planning groter is dan 5 minuten (Landex en Kaas (2009), Schittenhelm en Landex (2009)). Inmiddels is NS ook over op deze grenswaarde. In de Verenigde Staten wordt zelfs met 30 minuten al bovengrens gewerkt (Bush (2007)).

Naast verschillende indicatoren en drempelwaarden is een ander verschil in het meten van de betrouwbaarheid de locatie van de metingen: de onderzoeken geven aan dat er

verschillend wordt omgegaan met deze factor. Soms wordt slechts de eerste halte gemeten of alleen de belangrijkste halte. Figuur 4 geeft het aandeel van de gebruikte locaties voor de steden uit de benchmark. Uiteraard bepaalt de locatie een groot deel van de waarde: gemiddeld presteren lijnen aan het begin immers beter voor wat betreft betrouwbaarheid dan aan het einde. Uit het nationale onderzoek bleek dat in ca. 20% van de onderzochte bestekken de punctualiteit op het beginpunt van belang is, in ca. 25% op het eindpunt en ook in ca. 25% op de belangrijkste knopen. Voor de overige bestekken was geen specificatie gegeven.



Figuur 4: Locaties voor het meten van de stiptheid

Bovenstaande analyse laat zien dat er geen eenduidigheid bestaat in het meten van betrouwbaarheid. Door verschillen in meetlocaties, indicatoren en drempelwaarden is het moeilijk objectief vast te stellen of een openbaar vervoer systeem ook echt betrouwbaar is. Daarnaast wordt slechts gekeken naar aanbodskenmerken (het voertuig) en worden de effecten op de reizigers niet expliciet meegenomen, zoals het verschil in (1 minuut) te vroeg of (1 minuut) te laat rijden. Daarnaast wordt ook niet altijd expliciet rekening gehouden met de reizigersverdeling op de lijn. Het verschil in reizigerseffect van stiptheid op een lijn met veel instappers aan het begin of juist aan het eind kan zeer groot zijn. Tot slot wordt het belang van regelmaat bij hoge frequenties vaak genegeerd, evenals het effect van stiptheid op overstappende reizigers. Eerder al is gepleit voor een goede betrouwbaarheidsindicator die expliciet de reizigerseffecten meeneemt (e.g. Landex en Nielsen (2006) en Van Oort en Van Nes (2009)). In paragraaf 3.3 wordt verder ingegaan op deze manier van betrouwbaarheid in kaart brengen.

### 3.2 Impact van dienstregelingontwerp

Zoals in de inleiding geschetst is, is de betrouwbaarheid de overeenkomst tussen het plan en de uitvoering. Er zijn dus per definitie twee mogelijkheden om deze match te verbeteren: zowel de uitvoering aanpassen als de planning. In Van Oort (2011) zijn verschillende mogelijkheden aangegeven om met een betere planning van zowel het netwerk als de dienstregeling een betere betrouwbaarheid te bereiken. Eén van de mogelijkheden is de rijtijdbepaling: door te kiezen voor een zeer ruime rijtijd (een rijtijd met veel speling) en nooit te vroeg vertrekken wordt de kans op vertraging geminimaliseerd. Er ontstaat zo dus een zeer betrouwbaar systeem. Echter, de gemiddelde snelheid neemt ook drastisch af. Een eenzijdige focus kan dus tot ongewenste neveneffecten leiden. In exploitatiecontracten is het dus onverstandig om alleen te focussen op betrouwbaarheid: dit kan leiden tot meer (en teveel speling) in de dienstregeling. Overigens leidt veel speling ook tot veel dienstregelingsuren (DRU's), wat ook vaak als belangrijke kwaliteitsindicator gebruikt wordt in aanbestedingstrajecten. De reiziger is echter duidelijk niet gebaat bij meer DRU's door langzamere voertuigen. In Van Oort et. al (2012) en Van Oort en Van Nes (2008) wordt de trade off berekend tussen betrouwbaarheid en snelheid. Opvallend is dat uit de twee onderzoeken blijkt dat de relatie met de dienstregeling en betrouwbaarheid niet expliciet gelegd wordt: vaak



ontbreken er naast eisen aan betrouwbaarheid bijpassende eisen aan de rijtijd en dienstregeling.

### *3.3 Nieuwe betrouwbaarheidsindicator: extra reistijd*

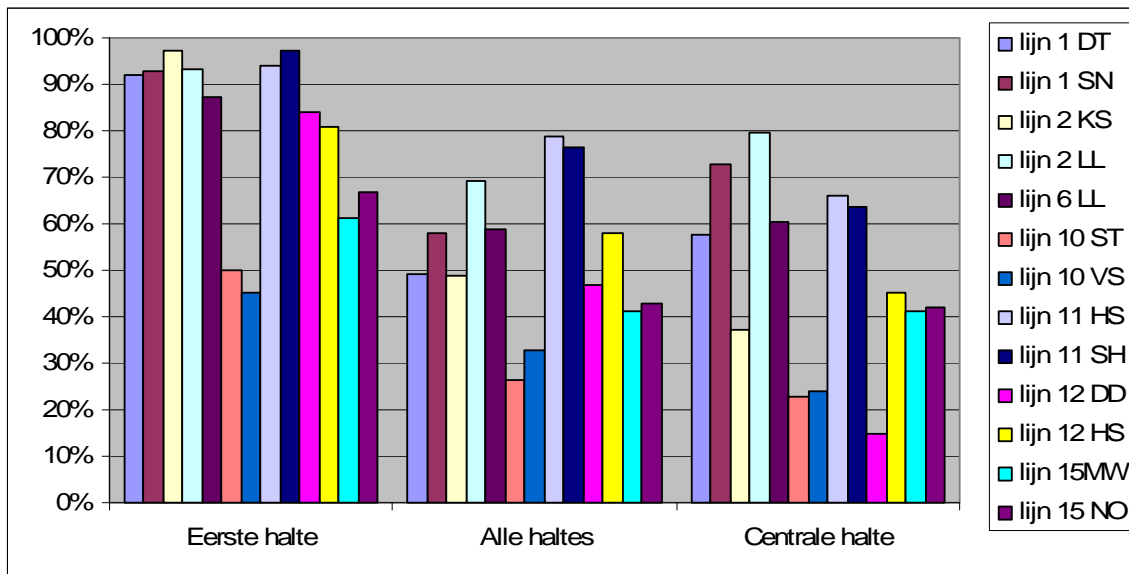
Zoals in paragraaf 3.1 is beschreven, is stiptheid een veel gebruikte kwaliteitsindicator, maar houdt deze niet voldoende rekening met reizigerseffecten. Van Oort (2011) ontwikkelde een nieuwe, betere indicator: de gemiddelde extra reistijd per reiziger. De stiptheid van een voertuig bepaalt de grootte van de wachttijd voor een reiziger aan het begin van een rit of tussen twee ritten in bij een overstap. Een onstipte dienstuitvoering leidt bij lage frequenties, als reizigers volgens dienstregeling naar de halte komen, direct tot extra reistijd. Hierbij is ook het (verschillende) effect van te vroeg of te laat vertrekken te zien. Bij hoge frequenties, als reizigers aselekt op een halte arriveren, bepaalt de regelmaat van opeenvolgende voertuigen, de extra reistijd. Met behulp van reizigerstellingen kan vervolgens door een weging de totale extra reistijd voor de gehele lijn worden berekend. In Van Oort (2011) wordt verder ingegaan op deze effecten en methode voor de berekening.

Met de introductie van de indicator extra reistijd verschuift de focus in betrouwbaarheidsmeting van de effecten op het voertuig naar de effecten voor de reiziger. Door dit in kaart te brengen wordt inzichtelijk hoeveel reistijdwinst of -verlies reizigers hebben en worden reizigerseffecten expliciet in kaart gebracht. Door te kijken naar de extra reistijd wordt ook de trade off tussen snelheid en betrouwbaarheid direct meegenomen. Dit helpt om beter inzichtelijk te maken wat het effect is van bepaalde maatregelen en voorkomt ongewenste neveneffecten. Ook kan deze indicator gebruikt worden betere inschattingen van toekomstige OV gebruik en maatschappelijke kosten en baten (zie Van Leusden en Van Oort 2012 en Van Oort et. al 2013).

In Nederland is extra reistijd een nieuwe indicator. Voor zover bekend wordt deze nog niet gebruikt om de betrouwbaarheid van openbaar vervoer te meten. Elders in de wereld is alleen Londen bekend, waar op een soortgelijke manier de betrouwbaarheid wordt gemeten en gepubliceerd. Frumin et al. (2009) en Uniman (2009) gaan verder in op het gebruik van betrouwbaarheidsindicatoren in Londen.

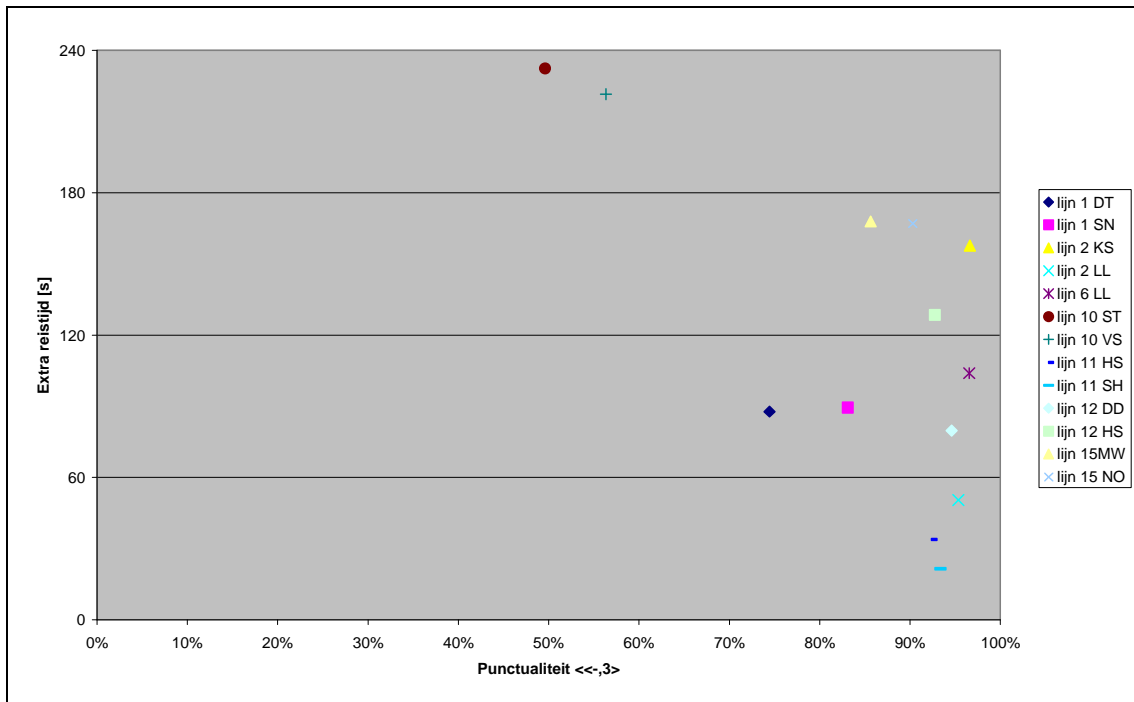
## **4. Beperkingen en effecten van gebruikte indicatoren**

Om inzicht te geven in mogelijke effecten van het gebruik van verschillende indicatoren om betrouwbaarheid uit te drukken is er een case studie gedaan in Den Haag. Alle tramlijnen zijn beoordeeld op betrouwbaarheid. Data is gebruikt van ochtendspitsuren in april in 2007. Figuur 5 laat het verschil zien indien er gemeten wordt op slechts de eerste halte, op een centrale halte of op alle haltes. Het percentage voertuigen dat vertrekt op deze locaties binnen een bandbreedte van 1 minuut te vroeg en 2 minuten te laat is gebruikt om punctualiteit uit te drukken. Deze figuur laat zien dat de rangorde van lijnen niet hetzelfde is, als de verschillende meetmethoden worden aangehouden. Een aantal lijnen scoort goed op een of twee typen, maar niet op het derde type. Dit laat zien dat de plaats van meten dus een groot stempel drukt op de betrouwbaarheid van een lijn en dat vergelijking van lijnen met meetdata met verschillende meetlocaties als meetpunt niet heel zinvol is.

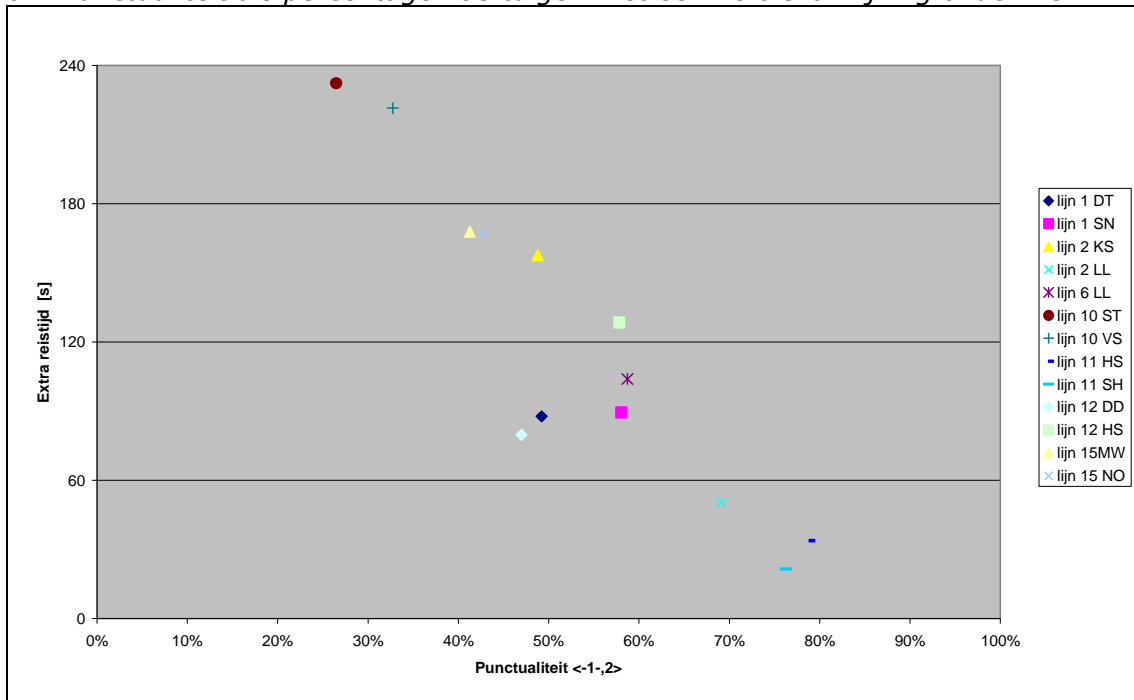


Figuur 5: Stiptheid <-1,+2> voor Haagse lijnen met verschillende meetmethoden.

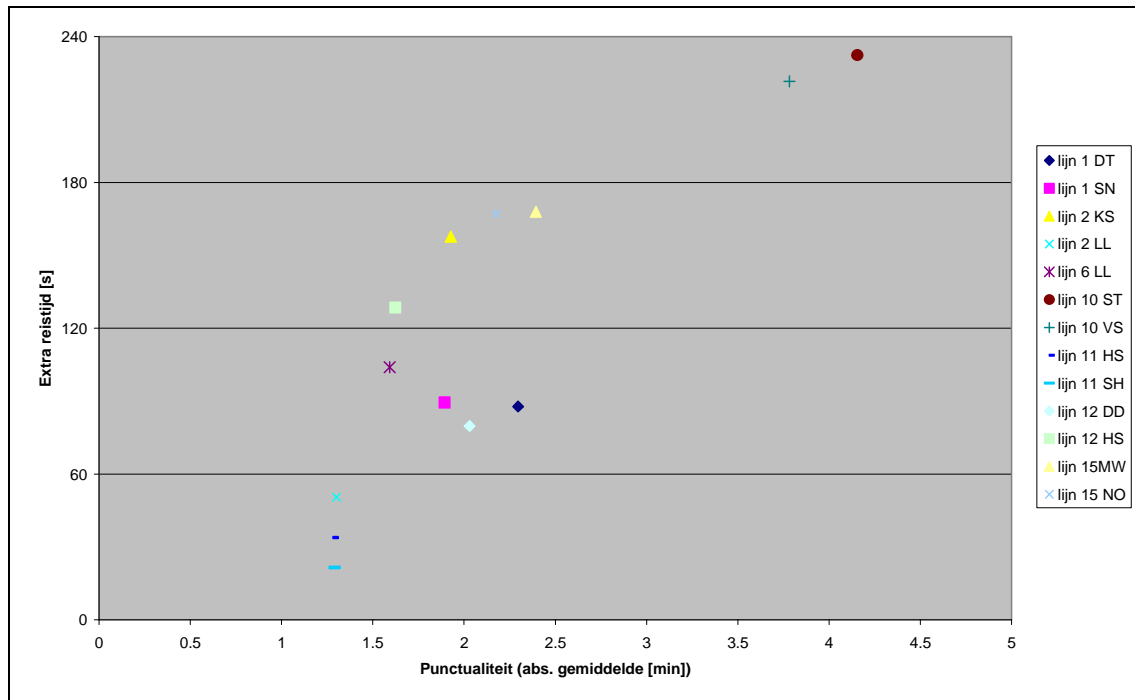
Naast de gebruikte meetlocatie is ook de gebruikte indicator van groot belang op de rangorde. Eerder al is geschetst dat ondanks dat stiptheid een sterk aanbodgerichte indicator is, deze veel gebruikt wordt. De definitie van stiptheid is daarnaast ook sterk uiteenlopend. De figuren 6A,B en C laten een vergelijking zien tussen drie definities en de nieuwe indicator "extra reistijd" voor de tramlijnen in Den Haag. De gebruikte definities voor stiptheid zijn het absoluut gemiddelde, het aantal vertrekken dat minder vertraagd was dan 3 minuten en het aantal vertrekken dat minder te laat was dan 2 minuten en niet meer te vroeg dan 1 minuut. Hoewel deze figuren globaal een lineair verband laten zien tussen deze indicatoren en de extra reistijd, is te zien dat er verschillen zijn in de rangorde van lijnen afhankelijk van de gekozen indicator. Zichtbaar is bijvoorbeeld dat lijn 15 MW slecht scoort op de indicator <-1,+2> (40% "op tijd") en heel goed scoort op <<-,+3> (85% "op tijd"). Lijnen waar dus veel te vroeg wordt vertrokken komen beter uit een indicator waarvoor geen ondergrens in opgegeven dan bij een indicator die zowel een onder- als bovengrens heeft. Een ander voorbeeld van inconsistentie is dat lijn 2 KS beter scoort op punctualiteit (tussen -1 en +2, figuur 6B) dan lijn 1 SN, maar wel een hogere extra reistijd per reiziger kent. De mate van betrouwbaarheid hangt dus af van de gekozen definitie. De extra reistijd heeft slechts één eenduidige definitie en is dus beter geschikt om een goede vergelijking en rangorde van verschillende lijnen te maken. Daarnaast wordt op deze manier ook het echte reizigerseffect duidelijk.



6A: Punctualiteit als percentage voertuigen met een vertrekafwijking onder +3 min



6B: Punctualiteit als percentage voertuigen met een vertrekafwijking tussen -1 en +2 min



6C: Punctualiteit als gemiddelde, absolute vertrekafwijking

Figuur 6: Relatie extra reistijd met drie verschillende typen van punctualiteit

## 5. Overige bevindingen

Naast kwantitatieve resultaten, zoals in vorige paragrafen gepresenteerd, zijn er uit de twee onderzoeken ook relevante kwalitatieve noties naar voren gekomen. Deze richten zicht met name op de aspecten reiziger, data en dienstregeling. Hieronder zijn ze benoemd.

De eerste conclusie is dat er weinig consistentie zit tussen de verschillende Nederlandse bestekken. Het soort eisen dat zich richt op betrouwbaarheid wisselt sterk, evenals de drempelwaarden. Er wordt erkend dat meer uniformering meerdere doelen dient: enerzijds helpt het bij het vinden van een optimum en anderzijds zijn vervoerders die over het hele land opereren gebaat bij uniformiteit. Dit verlaagt de drempel om in te schrijven. Een belangrijke valkuil is het stellen van te strenge eisen: het is immers simpel om eisen te stellen en te hopen dat daarmee het OV ook daadwerkelijk zo goed wordt. Aan de andere kant is het wel van belang een ambitieus niveau te vragen om de vervoerder uit te dagen het maximale te geven. Om hieraan tegemoet te komen is het van belang vast te stellen wat het huidige niveau van betrouwbaarheid is en wat op basis daarvan het gewenste en haalbare niveau is. Nieuwe databronnen, zoals GOVI (Van Oort et. al 2013) bieden hier grote hulp bij. Bij het vaststellen van het gewenste niveau is het van belang in kaart te brengen wat de oorzaken zijn van een (te) laag huidig niveau en wie er verantwoordelijk voor is. Er zijn immers vele oorzaken mogelijk (zie Weeda et. al 2006 en Van Oort 2011 voor een overzicht) en de vervoerder is niet voor alles verantwoordelijk. De gemeente en OV autoriteit spelen hierin ook een grote rol (bijv. infrastructuur).

In de interviews komt veel naar voren dat de reiziger belangrijk is. Zoals eerder gesteld zijn de gebruikte indicatoren echter voornamelijk voertuig georiënteerd: regelmaat wordt weliswaar genoemd als belangrijker dan stiptheid bij hoogfrequente systemen, maar wordt nog nergens gebruikt. Ook de weging naar de grootte van reizigersstromen in plaats van voertuigen is een verbetermogelijkheid die nog niet is toegepast. Tot slot is bij het stellen van eisen van belang differentiatie in type reizigers te maken: forensen stellen andere eisen dan scholieren of dagjesmensen.

Voor wat betreft monitoring liggen er grote kansen. Er komt steeds meer data vrij die helpt te monitoren en optimaliseren. Denk bijvoorbeeld aan GOVI en chipkaart. Deze inzichten kunnen ook gebruikt worden voor het stellen van betere eisen.

Tot slot de relatie met de dienstregeling: die is bijna alle gevallen beperkt of niet aanwezig. Er worden slechts niet-kwantitatieve eisen aan bijv. rijtijden gesteld (dat ze haalbaar moeten zijn bijvoorbeeld). Dat de keuze voor de rijtijd van grote invloed is op de mate van betrouwbaarheid (en snelheid) wordt vaak aan voorbij gegaan.

Tot zover de belangrijkste, gedeelde inzichten uit de interviews. Belangrijk nog te vermelden dat de aandacht voor betrouwbaarheid in alle gevallen sterk aan het toenemen is en dat er een belangrijk leerproces gaande is: de bestekken worden steeds beter.

## **6. Conclusies en aanbevelingen**

In toenemende mate worden er eisen aan betrouwbaarheid gesteld in bijv. OV bestekken. In veel steden en regio's wordt betrouwbaarheid gemeten met behulp van de indicator stiptheid. Uit een internationaal en nationaal onderzoek blijkt echter dat deze indicator niet eenduidig gehanteerd wordt. Er worden verschillende drempelwaarden gebruikt om een systeem als betrouwbaar aan te merken en niet overal wordt op dezelfde locaties gemeten. Dit maakt vergelijken lastig.

Tevens is alleen de indicator stiptheid niet voldoende om de kwaliteit van een systeem te meten, indien ook de reizigerseffecten inzichtelijk gemaakt dienen te worden. Door ruime rijtijden te hanteren wordt op tijd rijden vergemakkelijkt, maar daalt de operationele snelheid: een nadelig effect voor reizigers. Een eenzijdige focus kan dus tot ongewenste neveneffecten leiden.

De indicator stiptheid houdt geen rekening met alle effecten voor reizigers: alleen de gevolgen voor het voertuig worden meegenomen. Weging naar reizigeraantallen en effecten van overstappen worden genegeerd. In geval van hoge frequenties, waarbij reizigers de dienstregeling niet gebruiken, is sturen op punctualiteit niet optimaal. Sturen op regelmaat minimaliseert dan de reistijd, maar wordt nauwelijks genoemd in bestekken.

Een case studie in Den Haag laat zien dat verschillende definities van stiptheid ook tot verschillende uitkomsten leiden. De indicator extra wachttijd is geïntroduceerd om de focus naar de reiziger te verschuiven en is een betrouwbare manier om betrouwbaarheid te meten. De case studie in Den Haag onderbouwt de toegevoegde waarde van deze indicator ten opzichte van verschillende vormen van stiptheid, waarbij het effect op reizigers niet goed in beeld wordt gebracht.

Een belangrijke valkuil is het stellen van te strenge eisen: het is immers simpel om eisen te stellen en te hopen dat daarmee het OV ook daadwerkelijk zo goed wordt. Aan de andere kant is het wel van belang een ambitieus niveau te vragen om de vervoerder uit te dagen het maximale te geven. Om hieraan tegemoet te komen is het van belang vast te stellen wat het huidige niveau van betrouwbaarheid is en wat op basis daarvan het gewenste en haalbare niveau is. Nieuwe databronnen, zoals GOVI (Van Oort et. al 2013) bieden hier grote hulp bij. Bij het vaststellen van het gewenste niveau is het van belang in kaart te brengen wat de oorzaken zijn van een (te) laag huidig niveau en wie er verantwoordelijk voor is. Er zijn immers vele oorzaken mogelijk (zie Weeda et. al 2006 en Van Oort 2011 voor een overzicht) en de vervoerder is niet voor alles verantwoordelijk. De gemeente en OV autoriteit spelen hierin ook een grote rol (bijv. infrastructuur).

Door in OV-bestekken te werken met goede indicatoren en ambitieuze, doch realistische eisen komt een beter OV snel dichterbij. Een beter OV begint immers met een beter bestek.

## Acknowledgements

Dit onderzoek is uitgevoerd in samenwerking met KpVV, HTM Personenvervoer, Goudappel Coffeng en TU Delft. Dit onderzoek is verder mogelijk gemaakt door Transport Research Centre Delft en Railforum. Meer informatie en artikelen zijn te vinden op:

<http://www.goudappel.nl/projecten/betrouwbare-reistijd-als-kans-voor-het-openbaar-ve/>

## Referenties

Bush, R. (2007), *Does every trip need to be on time? Multimodal Scheduling Performance Parameters with an application to Amtrak Service in North Carolina*, Proceedings of 86<sup>th</sup> Annual Meeting of Transportation Research Board, Washington, D.C.

Frumin, M., D. Uniman, N.H.M. Wilson, R. Mishalani, J. Attanucci (2009), *Service Quality Measurement in Urban Rail Networks with Data from Automated Fare Collection Systems*, Proceedings of CASPT conference, Hong Kong.

Hakkesteegt, P., Th.H.J.Muller (1981), *Onderzoeksproject regelmaatsbevordering*, Verkeerskundige werkdagen, p.415-436.

KpVV (2013), Toolbox Beter Bestek, <http://www.kpVV.nl/Toolbox-Beter-Bestek>

Landex, A., A.H. Kaas (2009), *Examination of Operation Quality for High-Frequent Railway Operation*, Proceedings of 3<sup>rd</sup> International Seminar on Railway Operations Modelling and Analysis: RailZürich.

Landex, A., O. A. Nielsen (2006), *Simulation of disturbances and modelling of expected train passenger delays*, In: Computers in Railways X: Computer System Design and Operation in the Railway and Other Transit Systems.

Leusden, R. van en Oort, N. van (2011). *Excellent OV naar de Uithof: Capaciteits- en betrouwbaarheidsverbetering door tram in Utrecht*. In s.n. (Ed.), Bijdrage aan het colloquium vervoersplanologisch speurwerk (pp. 1-14).

Oort, N. van, Leusden, R. van en Ebben, M. (2013). *Goedkoper en beter openbaar vervoer door gebruik van GOVI-data*. Verkeerskunde: vaktijdschrift over verkeer en vervoer, 2013(1), 1-1

Oort, N. van, Boterman, J.W., Van Nes, R., 2012. The impact of scheduling on service reliability: trip-time determination and holding points in long-headway services. *Public Transport*, 4(1), 39-56.

Oort, N. van, 2011. *Service Reliability and Urban Public Transport Design*, T2011/2, TRAIL, PhD Thesis Series, Delft. ([http://www.goudappel.nl/media/files/uploads/2011\\_Proefschrift\\_Niels\\_van\\_Oort.pdf](http://www.goudappel.nl/media/files/uploads/2011_Proefschrift_Niels_van_Oort.pdf))

Oort, N. van, R. van Nes (2008), *Op tijd, dat telt: betrouwbaarder OV door een betere dienstregeling*, Verkeerskunde 5, pp. 48-53.

Schittenhelm, B., A. Landex (2009), *Quantitative Methods to Evaluate Timetable Attractiveness*, Proceedings of 3<sup>rd</sup> International Seminar on Railway Operations Modelling and Analysis: RailZürich.

Tahmasseby, S., R. van Nes., N. van Oort (2007), *Public transport network design and reliability*, Proceedings of the 3<sup>rd</sup> International Symposium on Transportation Network Reliability, The Hague.

Uniman, D. (2009), *Service quality measurement using afc smart card data - a framework for the London underground*, Master's thesis, MIT, Boston.

Weeda, V.A., P.B.L. Wiggeraad, K.S. Hofstra (2006), *Een treinvertraging zit in een klein hoekje*, Bundeling van bijdragen aan het Colloquium Vervoersplanologisch Speurwerk, Amsterdam.