

BTO 2017.045 | November 2017

BTO rapport

Voorspellen
storingsfrequenties:
haalbaarheid en
methodiek

BTO

Voorspellen storingsfrequenties: haalbaarheid en methodiek

BTO 2017.045 | November 2017

Opdrachtnummer

400554-194

Projectmanager

Drs. P.G.G. (Nellie) Slaats

Opdrachtgever

BTO - Thematisch onderzoek - Assetmanagement

Kwaliteitsborger(s)

dr. ir. E.J.M. (Mirjam) Blokker

Auteur(s)

dr. J. (Jozanneke) van Vossen en dr. ir. K. (Karel) van Laarhoven

Verzonden aan

Jaar van publicatie
2017

Meer informatie

dr. Jozanneke van Vossen
T 030 60 69 598

E

Jozanneke.van.Vossen@kwrwater.nl

Keywords

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511

F +31 (0)30 60 61 165

E info@kwrwater.nl

I www.kwrwater.nl



BTO 2017.045 | November 2017 © KWR

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

BTO Managementsamenvatting

Storingsfrequenties voorspellen met fysische gegevens en statistische analyses

Auteur(s) dr. J. (Jozanneke) van Vossen en dr. ir. K. (Karel) van Laarhoven

Het voorspellen van storingsfrequenties is van groot belang voor de ondersteuning van saneringsbeslissingen. De onzekerheden zijn echter groot en de onderbouwing voor een keuze in aanpak is mager. Daarom is in dit project een methodiek uitgewerkt, waarin verschillende gangbare statistische modellen worden vergeleken en aangevuld met kennis uit een fysisch model (Comsima¹). De methodiek blijkt goed toepasbaar; er zijn voldoende data beschikbaar en de uitkomsten van het fysische model suggereren dat de methodiek klaar is om succesvol te worden geïmplementeerd. Met de huidige resultaten kunnen de drinkwaterbedrijven op een bewustere manier omgaan met de voorspellingen van beslissingsondersteunende systemen voor leidingsanering. Dat betekent rekening houden met optredende onzekerheden en terughoudend zijn bij het splitsen van cohorten.



Storingen geven een indicatie van de conditie van de leiding en geven daardoor belangrijke informatie voor saneringsbeslissingen. Foto Beuken en Moerman (2017), rondbreuk in een grijs gietijzeren leiding.

¹ Comsima berekent belastingen op een buis op basis van leidingkenmerken en liggingomstandigheden. Comsima houdt ook rekening met veroudering.

Belang: onderbouwing van investeringen

De toekomstige ontwikkeling van de storingsfrequentie is een belangrijke indicator voor de toestand van leidingen. Diverse softwarepakketten die drinkwaterbedrijven gebruiken ter ondersteuning van hun saneringsbeslissingen doen voorspellingen voor toekomstige storingsfrequenties. De aanpak voor het genereren van deze voorspellingen verschilt per pakket en dit heeft grote invloed op het rekenresultaat van deze software en daarmee op saneringsbeslissingen en investeringen. Een onderbouwde keuze voor één van die aanpakken of andere internationaal gangbare aanpak is momenteel feitelijk niet mogelijk. Daarom is het van belang om een methodiek te ontwikkelen waarmee de voorspelling goed kan worden onderbouwd en waarvan de onzekerheid kan worden gekwantificeerd.

Aanpak: ontwikkelen methodiek en onderzoeken haalbaarheid

In dit onderzoek is een methodiek ontwikkeld en getoetst op haalbaarheid voor het voorspellen van toekomstige storingsfrequenties. Daarbij is uitgegaan van de UKNOW-aanpak, namelijk onzekerheden reduceren door informatiebronnen slim te combineren. De meest voor de hand liggende informatiebronnen zijn USTORE voor historische storingsdata en Comsima voor kennis van degradatiemechanismen.

Resultaten: combineren van statistiek en fysische kennis van leidingen is haalbaar en toepasbaar

De methodiek voor het voorspellen van toekomstige storingsfrequenties combineert een vergelijking van verschillende statistische

voorspelmodellen met een onderbouwing vanuit een voorspelling op basis van resultaten van een fysisch model (Comsima). De toepasbaarheid van deze methodiek is afhankelijk van de hoeveelheid en kwaliteit van storingsregistraties, homogeniteit van de leidingcohorten en de toepasbaarheid van de resultaten van Comsima op storingsfrequenties. De conclusie van de haalbaarheidsanalyse is dat er (1) voldoende data zijn om de methodiek toe te passen op het grootste deel van het leidingnet en dat (2) de voorgestelde methodiek voldoende potentie heeft om deze te implementeren in een casusstudie.

Implementatie: bewust omgaan met onzekerheden

Aanbeveling is om de ontwikkelde methodiek verder uit te werken en te implementeren in een casusstudie. Op basis van de huidige resultaten kunnen de drinkwaterbedrijven op een bewustere manier omgaan met de voorspellingen van beslissingsondersteunende systemen voor leidingsanering. Dat betekent rekening houden met optredende onzekerheden en terughoudend zijn bij het splitsen van cohorten. Daarnaast kunnen de bedrijven bijdragen aan een succesvolle implementatie door gegevens van lokale omstandigheden en kennis van conditie van gesaneerde leidingen systematisch bij te houden.

Rapport

Dit onderzoek is beschreven in rapport *Voorspellen storingsfrequenties: haalbaarheid en methodiek* (BTO-2017.045).

Inhoud

Inhoud	1
1 Inleiding	2
1.1 Motivatie	2
1.2 Doel	2
1.3 Leeswijzer	3
2 Kader en aanpak	4
2.1 Inleiding	4
2.2 Achtergrond	5
2.3 Van storingsdata naar degradatiecurve	8
Kansdichtheidfunctie	10
De cumulatieve en complementair cumulatieve distributiefuncties	10
De momentane storingsfrequentie	11
2.4 Voorgestelde methodiek voorspellen storingsfrequenties	13
3 Toepasbaarheid methodiek op bestaande cohortindelingen	17
3.1 Introductie	17
3.2 Algemene criteria voor indeling cohorten	17
3.3 Statistische evaluatie van cohorten	20
3.4 Huidige cohortindeling	21
3.5 Beoordeling huidige cohortindeling	22
3.6 Discussie en conclusies	24
4 Haalbaarheidsanalyse kiezen van een kansverdeling	27
4.1 Inleiding	27
4.2 Datafit statistische modellen	27
4.3 Op basis van Comsima	28
4.4 Discussie	36
5 Conclusies en aanbevelingen	38
5.1 Conclusies	38
5.2 Aanbevelingen	38
• Referenties	40
Bijlage I Resultaten toetsing cohorten	42

1 Inleiding

1.1 Motivatie

De ontwikkeling van de storingsfrequentie is een belangrijke indicator voor de toestand van leidingen. De voorspellingen voor toekomstige storingsfrequenties zijn de basis voor langetermijnprognoses waarbij investeringen worden vergeleken met prestaties. Diverse softwarepakketten die drinkwaterbedrijven gebruiken voor het ondersteunen van saneringsbeslissingen, gebruiken degradatiecurven voor het voorspellen van toekomstige storingsfrequenties. Een degradatiecurve is een wiskundige functie die gefit is door de storingsfrequenties als functie van leeftijd, waarna deze functie wordt gebruikt om de storingsfrequenties in de toekomst te schatten.

De vorm van de degradatiecurve (type curve: lineair, kwadratisch, exponentieel, samengesteld, etc.) is van invloed op het rekenresultaat van deze software². Om die reden is het van belang om de vorm van deze degradatiecurven zo goed mogelijk vast te stellen. Dit is tevens één van de bevindingen uit het SPO Kwantitatieve vergelijking restlevensduurmodellen dat KWR voor Vitens heeft uitgevoerd en waar ook Brabant Water, Dunea, PWN en Waternet aan hebben deelgenomen (Beuken et al., 2016; Beuken and Vossen, 2017). Er blijkt tot nu toe maar zeer beperkte onderbouwing te zijn om voor een specifieke vorm te kiezen (Scheidegger et al., 2015).

De drinkwaterbedrijven willen graag een onderbouwde voorspelling van toekomstige storingsfrequenties met een algemeen aanvaardbare methodiek, zie Figuur 1.1 (wens drinkwaterbedrijven). Het voorliggende rapport beschrijft de uitkomsten van een haalbaarheidsstudie naar de mogelijkheid om op basis van de huidig beschikbare kennis en data tot een dergelijke onderbouwing te komen. Dit is de eerste stap richting het uiteindelijke doel, zie Figuur 1.1 (stap 2017).



FIGUUR 1.1 SCHETS VAN HET EINDDOEL VOOR HET VOORSPELLEN VAN TOEKOMSTIGE STORINGSFREQUENTIES MET DE WEG DAAR NAAR TOE.

1.2 Doel

Doel is het onderzoeken van de haalbaarheid om op basis van bestaande storingsregistraties een realistische degradatiecurve op te stellen, en indien mogelijk, hiervoor een betrouwbare methodiek te ontwikkelen.

² Storingsvoorspellingen zijn van invloed op saneringsbeslissingen, maar zijn niet de enige factor in een saneringsbeslissing.

1.3 Leeswijzer

In Hoofdstuk 2 wordt de voorgestelde methodiek beschreven en toegelicht welke stappen nodig zijn voor de haalbaarheidsanalyse. In Hoofdstuk 3 en 4 worden de resultaten van deze stappen beschreven. In Hoofdstuk 3 staat de analyse van de storingsregistraties in de huidige cohortindeling met aanbevelingen voor het indelen van het net in cohorten. In Hoofdstuk 4 staat de analyse van de fysische onderbouwing voor de degradatiecurve met Comsima. In Hoofdstuk 5 staan conclusies en aanbevelingen.

2 Kader en aanpak

2.1 Inleiding

Het doel van dit project is een methodiek voor het voorspellen van storingsfrequenties voor een cohort op basis van het extrapoleren van historische storingsfrequenties. Om te laten zien wat een dergelijke methodiek inhoudt, wordt ter introductie een voorbeeld voor een fictief cohort van een dergelijke voorspelling van storingsfrequenties uitgewerkt. Welke storingsfrequentie nog acceptabel is voor drinkwaterbedrijven is beleid en valt buiten de scope van dit rapport.

Figuur 2.1 laat voor dit fictieve voorbeeldcohort de storingsfrequenties zien. Dit fictieve cohort bestaat uit 1300 km in leeftijd variërend van 10 tot 100 jaar. Het aantal kilometers per leeftijd staat weergegeven in Figuur 2.1 (oranje). In Figuur 2.1 staat in blauw ook de storingsfrequentie als functie van leeftijd weergegeven. Voor het voorbeeld wordt aangenomen dat het aantal kilometers binnen de groep niet verandert. Storende leidingen worden in dit voorbeeld gerepareerd, maar niet vervangen. Er komen geen nieuwe leidingen meer bij, reparatie van een leiding heeft in dit fictieve voorbeeld geen invloed op de ontwikkeling van de storingsfrequentie van die leiding. De lengte van een reparatiesectie is dusdanig klein t.o.v. de totale leidinglengte per leeftijd dat het verwaarloosd mag worden. Door de berekende storingsfrequenties te combineren met de groepsopbouw qua leidinglengte kan de verwachte storingsfrequentie van het cohort worden bepaald met de volgende formule:

$$SF(jaartal)_{cohort} = \frac{\sum_t h_T(t) * L(t)}{\sum_t L(t)},$$

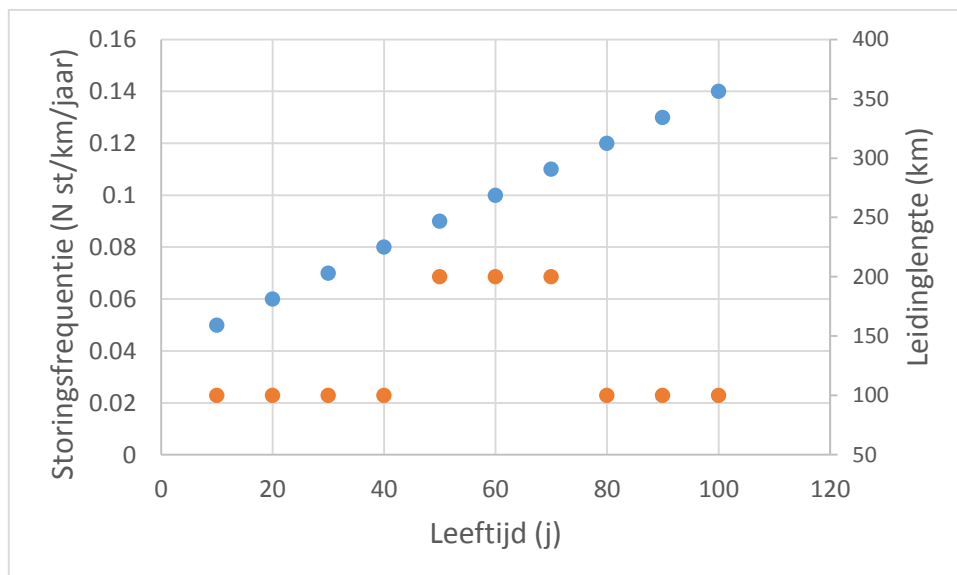
Waarin $SF(jaartal)$ de gemiddelde storingsfrequentie in aantal storings per kilometer per jaar is van de groep leidingen, $h_T(t)$ de gemiddelde storingsfrequentie per leeftijdscategorie en $L(t)$ de kilometers leidinglengte per leeftijdscategorie. Door de formule in te vullen met de gegevens in Figuur 2.1 kan worden berekend dat de huidige storingsfrequentie voor de totale groep 0.10 storings/km/jaar is.

Als de storingsfrequentie voor de toekomst moet worden voorspeld, is er in dit voorbeeld het probleem dat voor leidingen ouder dan 100 jaar geen storingsfrequenties beschikbaar zijn door het ontbreken van data voor dusdanig oude leidingen. Er wordt daarom een trend door de data heen gefit (een degradatiecurve), met de aanname dat de huidige lineaire trend zich doorzet in de toekomst. Deze degradatiecurve verandert niet als het cohort verouderd, maar is een intrinsieke eigenschap van een cohort (zolang het cohort homogeen is, zie verder paragraaf 2.3). De verwachte storingsfrequentie van het cohort op een bepaald moment is niet alleen afhankelijk van de degradatiecurve, maar ook van de leeftijdsopbouw en die verandert dus wel. Dat betekent dat de verwachte gemiddelde storingsfrequentie van het cohort over 10 jaar 0.11 storings/km/jaar zal zijn en over 20 jaar 0.12 storings/km/jaar. De toename in storingsfrequentie komt doordat het cohort verouderd en de storingsfrequentie voor oudere leidingen hoger is dan voor jongere leidingen. Ook speelt dat er in het cohort een piek is in kilometers leidinglengte bij een leeftijd van 50-70 jaar (huidige situatie). De bijdrage van die leidingen is in de formule het grootst.

Een methodiek voor het voorspellen van storingsfrequenties bevat dan ook:

- Een indeling van het net in groepen met een bekende leeftijdsopbouw qua leidinglengte, Hoofdstuk 2.2.2;
- Storingsfrequenties op basis van storingsregistraties, Hoofdstuk 2.2.2;
- Een degradatiecurve, Hoofdstuk 2.3;

In dit hoofdstuk worden de achtergrond van deze begrippen toegelicht en een methodiek voor het voorspellen van storingsfrequenties voorgesteld, Hoofdstuk 2.4.



FIGUUR 2.1 VOORBEELD ONTWIKKELING STORINGSFREQUENTIE ALS FUNCTIE VAN DE LEEFTIJD VAN EEN LEIDING (IN BLAUW STAAN DE STORINGSFREQUENTIES WEERGEGEVEN, IN ORANJE DE LEIDINGLENGTES ALS FUNCTIE VAN LEEFTIJD).

2.2 Achtergrond

2.2.1 Storingen en degradatie

Een storing is een lek in het leidingnet, dat wordt geregistreerd en hersteld. Herstel kan plaatsvinden door reparatie ter plekke dan wel vervanging van het leidingdeel. Een lekkage kan optreden in verschillende onderdelen van het net (zoals een leidingsegment of verbinding) en kan verschillende vormen hebben (zoals een puntlek, scheur, of uitschuiving), zie de documentatie van USTORE voor de meest voorkomende vormen (Vloerbergh and van Thienen, 2011).

Een storing in een onderdeel treedt op wanneer het zwaarder belast wordt dan het kan weerstaan. Dit kan komen door een (onvoorziene) toename in uitwendige of inwendige belasting en/of door een afname van de weerstand tegen (reguliere) belasting door degradatie van het onderdeel. Zowel bronnen van belasting als degradatiemechanismen zijn divers en zijn afhankelijk van het onderdeel, de omgeving en menselijk ingrijpen (zoals operationeel gebruik of schade door derden).

Degradatie verschilt sterk per materiaal, bijvoorbeeld:

- Het belangrijkste degradatiemechanisme bij AC leidingen is uitloging van calciumhydroxide. Hierdoor vermindert de sterkte van de wand tot deze te klein is om de belastingen te kunnen weerstaan. Uitloging kan aan de binnenzijde plaatsvinden en aan de buitenzijde. De uitloging aan de buitenzijde is gerelateerd aan omgevingskenmerken, zoals de agressiviteit van de bodem en het grondwater (Slaats and Mesman, 2003). Uitloging aan de binnenzijde is gerelateerd aan de kwaliteit van het drinkwater. Daarnaast is een veel voorkomende oorzaak van storingen bij ouder AC lekkende natuurrubberringen in verbindingen (Mesman et al., 2016).
- Voor PVC leidingen is langzame scheurgroei rond imperfecties in het materiaal de belangrijkste oorzaak van falen (Mesman, 2015).
- Voor gietijzeren leidingen is het belangrijkste degradatiemechanisme corrosie (Mesman and Slaats, 2003).

2.2.2 Storingsfrequentie als functie van leeftijd

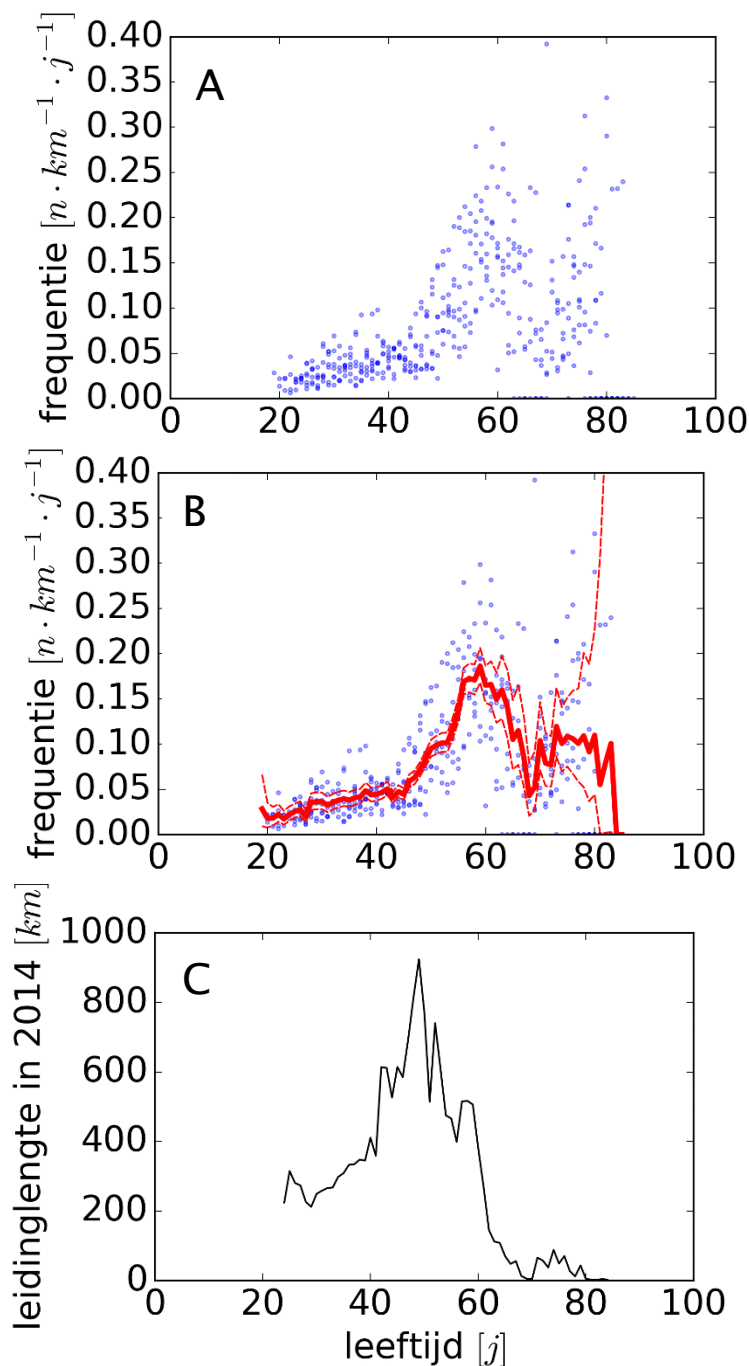
De complete storingsgeschiedenis van individuele leidingen met een lange levensduur is in de praktijk niet beschikbaar. Bovendien worden leidingen na een beperkt aantal storingen meestal vervangen. Daarom wordt de te verwachten storingsfrequentie als functie van leeftijd in de praktijk afgeleid uit het storingsgedrag van groepen leidingen i.p.v. individuele leidingen. Om de storingsfrequenties van verschillende leidingen onderling te kunnen vergelijken en trends te analyseren als functie van tijd, wordt de volgende aanname gedaan: leidingen met vergelijkbare materiaaleigenschappen in vergelijkbare liggingomstandigheden kennen een vergelijkbare ontwikkeling van storingsfrequentie als functie van leeftijd. Een groep leidingen waarvan wordt aangenomen dat de storingsfrequenties op eenzelfde manier samenhangen met leeftijd wordt aangeduid als "cohort". De storingsfrequentie van een cohort is het aantal storingen dat per tijdseenheid in die groep optreedt (N storingen/km/jaar).

Een cohort bestaat meestal uit leidingen en/of verbindingen met verschillende aanlegjaren waarvan de gecombineerde storingsdata een zo goed mogelijk beeld van de storingsfrequentie tijdens de gehele (potentiele) levensduur van een leiding laten zien. Een cruciale aanname daarbij is dat de ontwikkeling in de storingsfrequentie als functie van leeftijd vergelijkbaar is voor verschillende aanlegjaren, of een leiding van een cohort nu of 80 jaar geleden werd aangelegd. In de praktijk is er waarschijnlijk wel variatie met het aanlegjaar te verwachten door bijvoorbeeld aanpassing van productieprocessen, veranderde materiaalsamenstelling of veranderende eisen door wet- en regelgeving. Dit zorgt voor variaties in de trends. Als deze variaties heel groot zijn, passen leidingen niet meer bij elkaar en moeten ze in aparte cohorten worden ondergebracht. Dit type variaties is niet altijd gemakkelijk te duiden en mede daarom is het samenstellen van cohorten niet triviaal, zie Hoofdstuk 3.

Figuur 2.2 toont als voorbeeld de storingsfrequentie per leeftijd voor alle AC leidingen (USTORE data tussen 2009 en 2015). De individuele datapunten representeren de frequentie van één leeftijdscategorie (in dit geval 1 jaar) in één registratiejaar (Figuur 2.2A). De lijn door de datapunten is de gemiddelde storingsfrequentie als functie van leeftijd op basis van alle storingsregistraties tussen 2009 en 2015 (Figuur 2.2B). Het is duidelijk dat er van jaar tot jaar een grote spreiding is rond het gemiddelde. In Hoofdstuk 3 worden zowel AC leidingen als de andere materialen nader beschouwd.

Ter illustratie is ook het 95% betrouwbaarheidscontour van het gemiddelde weergegeven op basis van een Poisson distributie. Dit geeft een indicatie of er voldoende storingsdata

aanwezig zijn om de berekende storingsfrequentie te vertrouwen, zie ook Hoofdstuk 3. Merk op dat zowel de storingsfrequentie als de lengte van een set leidingen invloed hebben op de hoeveelheid beschikbare storingsdata: leidingen met een lage frequentie leveren minder storingsdata, en meer kilometers leidingen leveren meer storingsdata. Het betrouwbaarheidsinterval is voor de meeste leeftijden veel kleiner dan de spreiding in de storingsfrequenties van jaar tot jaar. Het is belangrijk om te benadrukken dat de betrouwbaarheid van het gemiddelde niet hetzelfde is als de spreiding van de data rond het gemiddelde. Beide zijn wel in te schatten met statistische analyses.



In het voorbeeld in Figuur 2.2 wordt het betrouwbaarheidsinterval wijder bij leidingen jonger dan 20 jaar en nog veel duidelijker, bij leidingen ouder dan 70 jaar (aangelegd vóór 1950). Dit is een indicatie dat er voor die leeftijd te weinig storingsdata is verzameld om betrouwbaar statistische analyses te kunnen uitvoeren. AC leidingen jonger dan 1993 zijn niet in USTORE geregistreerd (verbod op nieuw aanleg AC sinds 1993). Het onderste figuur (Figuur 2.2C) geeft het aantal leidingkilometers in het cohort weer als functie van leeftijd. Daaruit volgt dat er maar zeer beperkt AC leidingen ouder dan 70 jaar aanwezig zijn. Niet alleen is dit de beginperiode van AC, waardoor er al beperkt leidingen uit die periode aanwezig zijn, een groot aantal is naar alle waarschijnlijkheid al eerder gesaneerd.

De gemiddelde storingsfrequentie laat een trend zien van oplopende storingsfrequenties tot een leeftijd van ongeveer 60 jaar, waarna het gemiddelde daalt tot een lager niveau. Vanaf de leeftijd dat deze verlaging inzet, vermindert de betrouwbaarheid van het gemiddelde. Daarom kunnen we op basis van deze data niet met zekerheid zeggen of de trend vanaf 75 jaar weer zal stijgen, zal blijven dalen, of af zal vlakken. Het kennen van die trend is wel cruciaal voor het voorspellen van storingsfrequenties voor de komende decennia, omdat steeds meer leidingen in die leeftijdscategorie terecht zullen komen. Daarom is het belangrijk om de juiste trend te kiezen.

2.3 Van storingsdata naar degradatiecurve

Storingsdata zoals die geïllustreerd in Figuur 2.2 kunnen de basis vormen voor een statistische voorspelling van toekomstige storingsfrequenties. De kern daarbij is het opstellen van een degradatiecurve. De degradatiecurve is een wiskundige uitdrukking voor de trend in storingsfrequenties van een leiding of cohort als functie van leeftijd. Degradatie is hier een brede term die niet alleen degradatie van materialen omvat, maar de degradatie van algehele conditie van de leiding met meenemen van liggingomstandigheden en omgevingskenmerken. Door deze degradatiecurve te extrapoleren naar leeftijden waarop geen storingsdata meer beschikbaar zijn, kan voorspeld worden hoe de storingsfrequentie van de leiding of het cohort zich in de toekomst zal ontwikkelen. Het gemiddelde van de gemeten storingsfrequenties per leeftijd is in feite een eerste aanzet voor een degradatiecurve, Figuur 2.2B. Dit gemiddelde is niet bruikbaar bij onvoldoende data en is niet voor alle leeftijden beschikbaar. Door een wiskundige uitdrukking te fitten door de data wordt dit wél mogelijk.

Het is belangrijk om opnieuw het onderscheid te benadrukken tussen de degradatiecurve en de storingscurve van een cohort. De degradatiecurve representeert (bij voldoende en betrouwbare data) het generieke storingsgedrag met leeftijd voor een willekeurige buis in het cohort en dit verandert niet bij een veranderende leeftijdsopbouw van de verzameling buizen in het cohort. De toekomstige storingsvoorspelling is niet generiek voor een cohort, maar afhankelijk van de opbouw van het cohort in kilometers leiding per leeftijdscategorie. Doordat cohorten verouderen en worden gesaneerd, verandert deze leeftijdsopbouw en daarmee de toekomstige storingsvoorspelling. Een voorbeeld van een dergelijke fit met een wiskundige functie is een statistische methode zoals regressie. Dit heeft enkele voordelen:

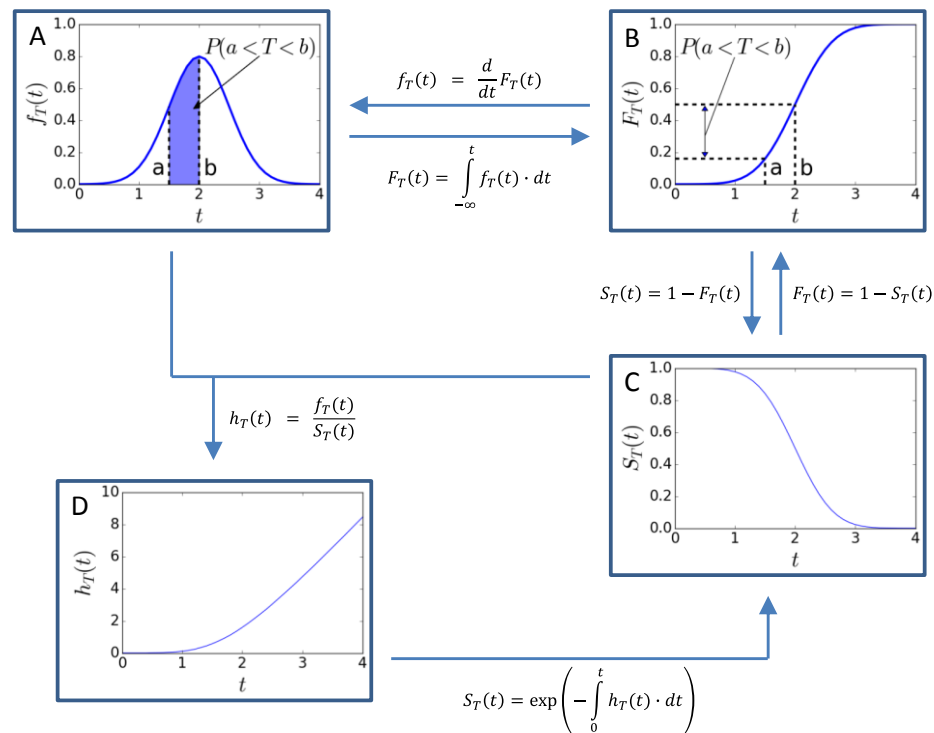
- De degradatiecurve is gebaseerd op de hele dataset tegelijk. De resulterende gemiddelden voor leeftijden met weinig beschikbare data worden daardoor 'betrouwbaarder'. Deze interpolatie kan ook nadelen hebben, zoals het wegfilteren van reële variaties van groepen leidingen die net iets afwijken van het cohort.
- De curve kan worden geëxtrapoleerd naar tijdsintervallen waarvoor nog helemaal geen data beschikbaar zijn.
- Deze methoden benaderen niet alleen de gemiddelde storingsfrequentie, maar geven ook uitdrukkingen voor de onzekerheid in het gemiddelde en voor de te verwachten

mate van spreiding van de data rond het gemiddelde. Nogmaals wordt er op gewezen dat de onzekerheid in een gemiddelde en de spreiding rond een gemiddelde niet hetzelfde zijn: op basis van een groot aantal datapunten met een grote spreiding kan nauwkeurig een gemiddelde met een kleine onzekerheid worden bepaald.

Een belangrijke consequentie van het gebruik van statistische methoden is dat de resulterende curve niet meer alleen op de data gebaseerd is, maar ook op aannames over de trends in de data. Eén van de centrale aannames is dat het gemodelleerde proces (storen van een leiding) een toevalsproces is, waarbij de kans P op iedere mogelijke uitkomst van het toevalsproces wordt bepaald door een bepaalde waarschijnlijkheidsverdeling.

Een waarschijnlijkheidsverdeling is een model voor het gedrag van een verschijnsel. Dit model kan op verschillende manieren weergegeven worden om verschillende aspecten van dit gedrag te beschrijven. We bespreken hier vier veelgebruikte gedaanten van waarschijnlijkheidsverdelingen. Als voorbeeld toont Figuur 2.3 de normale verdeling in ieder van deze vier vormen: de kansdichtheidsfunctie $f_T(t)$ in Figuur 2.3A, de cumulatieve distributiefunctie $F_T(t)$ in Figuur 2.3B, de complementaire cumulatieve distributiefunctie, of survival functie, $S_T(t)$ in Figuur 2.3C en de momentane storingsfrequentie $h_T(t)$ in Figuur 2.3D. De toevalsvariabele die door de kansverdeling wordt beschreven, het moment van storen wordt aangeduid als T . Een bepaalde uitkomst van T , dat wil zeggen de daadwerkelijke leeftijd bij falen, wordt daarbij steeds aangeduid als t .

Om de onderlinge relaties te illustreren geven de pijlen met formules weer hoe ieder van de functies omgeschreven wordt in de anderen. Dit benadrukt dat deze vormen slechts verschillende representaties zijn van hetzelfde onderliggende model, de waarschijnlijkheidsverdeling. Ieder bevat precies dezelfde informatie.



FIGUUR 2.3 SAMENHANG VAN DE VERSCHILLENDE VORMEN VAN EEN KANSVERDELING, MET A DE KANSDICHTHEIDSFUNCTIE, B DE CUMULATIEVE DISTRIBUTIEFUNCTIE, C DE COMPLEMENTAIRE CUMULATIEVE DISTRIBUTIEFUNCTIE EN D DE MOMENTANE STORINGSFREQUENTIE.

Kansdichtheidfunctie

De kansdichtheidfunctie $f_T(t)$ is de wiskundige uitdrukking voor de zogenaamde 'relatieve waarschijnlijkheid' van alle uitkomsten. De kansdichtheidfunctie maakt in één oogopslag zichtbaar welke uitkomsten waarschijnlijker zijn dan anderen. De kansdichtheidfunctie van de normale verdeling staat weergegeven in Figuur 2.3A.

Een belangrijke eigenschap van de kansdichtheidfunctie is dat de totale oppervlakte onder een kansdichtheidsverdeling (de integraal over een interval dat alle uitkomsten bevat) gelijk is aan 1: niet alle uitkomsten zijn even waarschijnlijk, maar het staat vast dat één van de mogelijke uitkomsten zal optreden. Met de kansdichtheidfunctie van de levensduur kan worden berekend wat de kans is dat een leiding faalt op een moment tussen twee bepaalde leeftijden a en b :

$$P(a < T < b) = \int_a^b f_T(t) \cdot dt$$

De cumulatieve en complementaire cumulatieve distributiefuncties

De tweede vorm is de cumulatieve distributiefunctie, ook wel de cumulative distribution function, CFD, of verdelingsfunctie genoemd, Figuur 2.3B. In termen van de hierboven beschreven kansdichtheidsfunctie: $F_T(t)$ is de kans dat de uitkomst op het interval van alle uitkomsten tot en met t ligt, oftewel:

$$F_T(t) = P(T \leq t) = \int_{-\infty}^t f_T(t) \cdot dt^3$$

³ Wiskundig wordt een kansverdeling nooit 0, waardoor een integraal begint bij $-\infty$. In praktijk zijn er geen leeftijden van leidingen kleiner dan 0, waardoor de kansverdelingen wordt afgekapt bij 0.

$$P(t_1 < T < t_2) = F_T(t_2) - F_T(t_1)$$

$$f_T(t) = \frac{d}{dt} F_T(t)$$

In de context van levensduurvoorspelling kan $F_T(t)$ geïnterpreteerd worden als de kans dat een object op leeftijd t al gestoord heeft. Voor levensduurvoorspelling wordt ook gewerkt met een nauwverwante functie: de complementaire cumulatieve distributiefunctie, $S_T(t)$. Deze geeft de kans dat een object op leeftijd t nog niet gefaald heeft. Om deze reden wordt $S_T(t)$ ook wel de *survival function* genoemd. $S_T(t)$ kan ook weer worden uitgedrukt in termen van $f_T(t)$ of $F_T(t)$:

$$S_T(t) = P(T > t) = 1 - F_T(t) = \int_t^{\infty} f_T(t) \cdot dt$$

De momentane storingsfrequentie

De momentane storingsfrequentie $h_T(t)$ (instantaneous failure rate, ook wel 'hazard function') staat weergegeven in Figuur 2.3D. Deze vorm representeert de relatieve waarschijnlijkheid dat een storing optreedt op moment t , gegeven de voorwaarde dat de storing niet optrad in het voorgaande tijdsinterval. Met andere woorden: $h_T(t)$ beschrijft welk percentage van alle storingen die vanaf t nog zullen optreden op precies moment t . Deze functie representeert dus de storingsfrequentie als functie van leeftijd: het percentage van kilometers bus dat stoort op leeftijd t ten opzichte van alle kilometers bus die die leeftijd bereikt hebben. De hazard function is als volgt gerelateerd aan de andere vormen van de kansverdeling:

$$h_T(t) = \frac{f_T(t)}{S_T(t)}$$

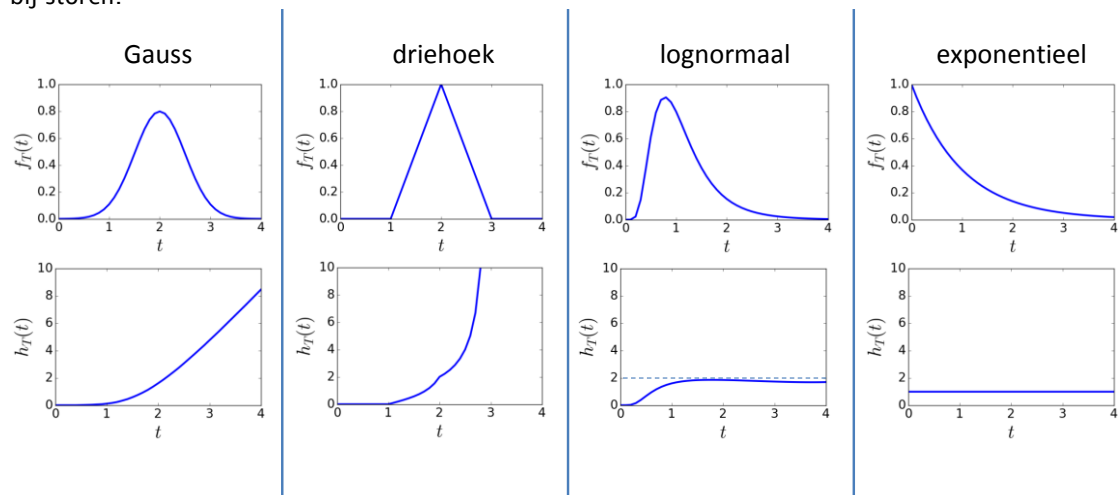
$$S_T(t) = \exp\left(-\int_0^t h_T(t) \cdot dt\right).$$

Deze vormen van de waarschijnlijkheidsverdeling gelden zowel voor een individuele leiding als voor een cohort (zolang voldaan wordt aan de aanname van een homogene groep leidingen). Voor een individuele leiding wordt de kans op storen beschreven, voor een cohort de verhouding tussen het storende en niet storende deel van het cohort. Let weer op het verschil tussen een degradatiecurve en storingsfrequentie van een cohort, de degradatiecurve is generiek, de storingsfrequentie zal zich ontwikkelen op basis van de leeftijdsopbouw van het cohort.

De momentane storingsfrequentie of hazard function komt overeen met de storingsfrequenties, zoals die berekend wordt vanuit geregistreerde storingen. Figuur 2.4 illustreert de samenhang tussen kansverdeling en de storingsfrequentie als functie van leeftijd voor een aantal veelgebruikte kansverdelingen, zie ook Vloerbergh en Blokker (2007). Per kansverdeling worden steeds de kansdichtheidsfunctie (boven) en de momentane storingsfrequentie (onder) getoond. Hoewel de eerste drie kansverdelingen kansdichtheidsfuncties met ogenschijnlijk vergelijkbare vormen hebben (een maximale kans op een specifieke leeftijd met afnemende kansen daarvoor en daarna), zijn de storingsfrequenties verschillend. Waar een Gaussische verdeling bijvoorbeeld een toename in storingsfrequenties als functie van leeftijd laat zien, laat de lognormale verdeling na de eerste toename een afvlakking of zelfs een daling zien. Dit is omdat in tegenstelling tot een Gaussische verdeling de lognormale en de exponentiële verdeling met toenemende leeftijd steeds langzamer naar nul lopen. Dat betekent dat als een cohort volgens een lognormale kansverdeling stoort, de oudste leidingen in dat cohort relatief steeds minder gaan storen (de beste blijven over), wat de afvlakking in de hazard curve geeft. Dit illustreert het belang van een onderbouwde keuze voor de kansverdeling waarmee de storingsdata worden

beschreven. Het eerder genoemde voordeel van een maat krijgen voor de onzekerheid zegt niets als de verkeerde curve wordt gekozen.

De veel benoemde “badkuip curve” met een hoge storingsfrequentie in het deel met lage leeftijden, gevolgd door een vlakke periode en een periode met oplopende storingsfrequenties met leeftijd is een samenstelling van verschillende degradatiecurves, waarbij gedurende de levensduur van een leiding verschillende mechanismen een rol spelen bij storen.



FIGUUR 2.4 VOORBEELDEN VAN KEUZES VOOR DE KANSDICHTHEIDSFUNCTIE EN DE DAARUIT VOLGENDE MOMENTANE STORINGSFREQUENTIE.

Er zijn veel verschillende kansverdelingen en statistische methoden waarop een degradatiecurve gebaseerd kan worden. Iedere aanpak heeft een eigen set van bijbehorende aannames over het geanalyseerde systeem en een set van eisen aan de geanalyseerde data. Een uitgebreid overzicht van aanpakken die in de literatuur zijn gebruikt is gegeven door Kleiner and Rajani (2001) en een update van dit overzicht is gegeven door Liu et al. (2012). Daarnaast geven Scheidegger et al. (2015) een overzicht van de aannames en data-eisen van de meest gebruikte aanpakken, met aandacht voor de wiskundige principes. Een kanttekening van Scheidegger et al. (2015) is echter dat de beschikbare onderbouwing om uit alle mogelijkheden één degradatiecurve te kiezen niet sterk is, terwijl de resultaten onderling sterk kunnen verschillen. Dit komt gedeeltelijk doordat het moment waarop de curves van elkaar gaan verschillen vaak overeenkomt met het deel van de dataset waar de trends onbetrouwbaar worden.

Een storingsvoorspelling kan ook gebaseerd worden op een beschrijving van de leiding en de omgeving door gebruik van fysische modellen, zie ook Moerman et al. (2016a). Comsima (Wols et al., 2015b) en het verouderingsmodel in het KSLB (Beuken and Mesman, 2011a) zijn voorbeelden van fysische modellen. Rajani and Kleiner (2001) bespreken ook een groot aantal fysische modellen. Het voordeel van fysische modellen is dat voorspellingen gebaseerd zijn op kennis van degradatiemechanismen en de kenmerken van specifieke leidingen en hun omgeving en dat er dus ook een uitspraak voor een specifieke leiding wordt gedaan. Het belangrijkste nadeel is dat een voldoende betrouwbare beschrijving van de degradatieprocessen veel lokale gegevens vraagt, zoals kenmerken van leidingen en de omgeving. Deze zijn lang niet altijd beschikbaar.

2.4 Voorgestelde methodiek voorspellen storingsfrequenties

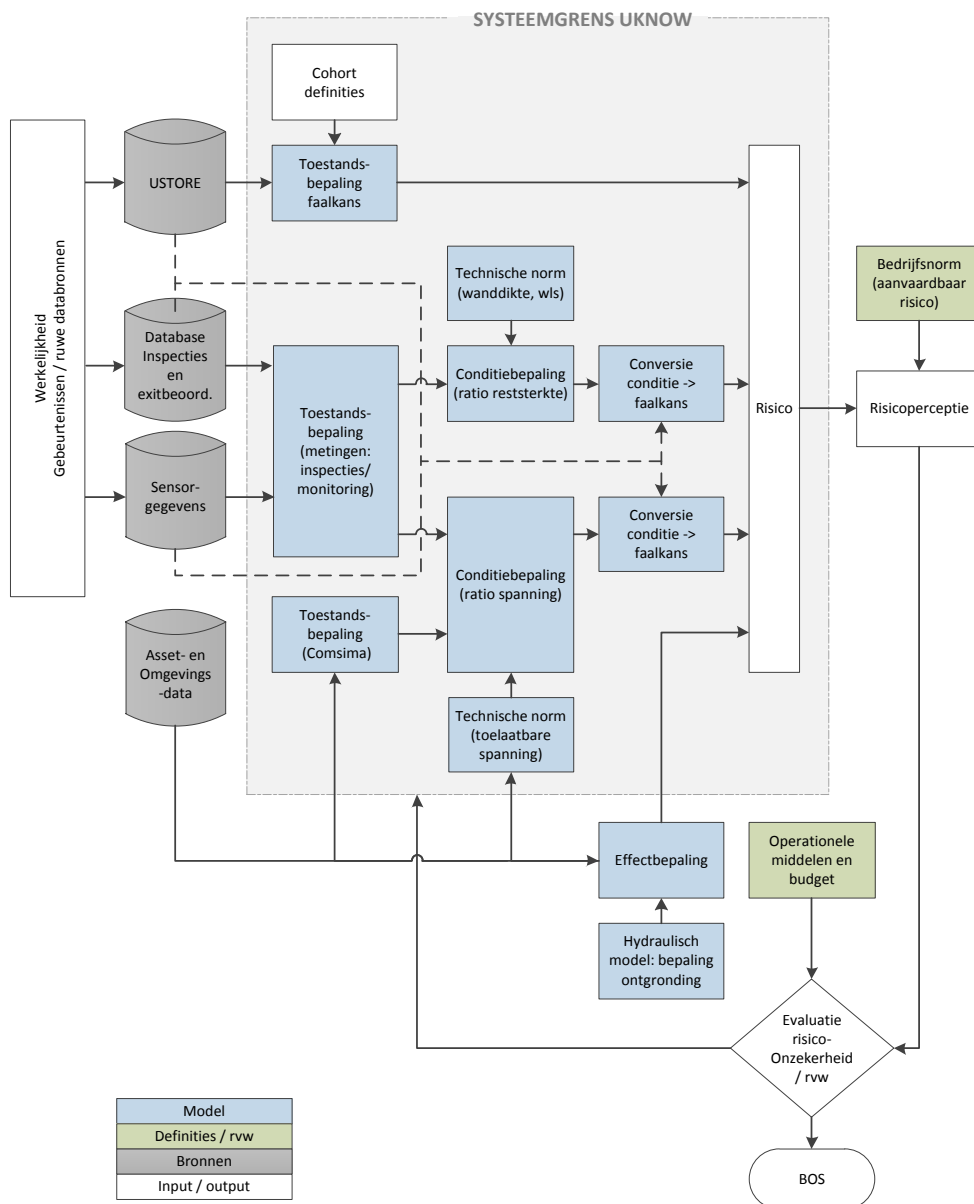
2.4.1 Uitgangspunt: UKNOW

UKNOW is gericht op het optimaal combineren van informatie en het verkleinen van onzekerheden (Moerman et al., 2016a). Figuur 2.5 laat de architectuur van UKNOW zien voor het combineren van informatie vanuit storingsgegevens, informatie vanuit veldmetingen en fysische modellen in een faalkans. Er zijn verschillende mogelijkheden om de onzekerheid in de faalkans te verkleinen:

- Op basis van alleen storingsgegevens kunnen verschillende statistische modellen worden vergeleken. Door alle statistische modellen tegelijk op dezelfde data toe te passen, geeft de mate waarin de voorspellingen uiteenlopen een maat voor de onzekerheid.
- Toevoegen van aanvullende informatie uit andere bronnen, namelijk faalkansen op basis van degradatiemetingen en faalkansen op basis van de beschikbare kennis van degradatie van leidingsystemen (en hun componenten) en kennis over ontoelaatbare interne en externe belastingen op leidingsystemen.

De voorgestelde methodiek voor het voorspellen van storingsfrequenties van cohorten is daarom gebaseerd op deze beide elementen:

- Het combineren van statische trendanalyse op basis van storingsgegevens met een onderbouwing op basis van een fysisch degradatiemodel (in dit geval Comsima). Dit biedt meerwaarde, niet alleen omdat op deze manier een betere degradatiecurve kan worden opgesteld, maar ook omdat de invloed van verschillende processen op de degradatiecurve expliciet kan worden onderzocht.
- Het vergelijken van de resultaten van de degradatiecurve op basis van Comsima met de resultaten van verschillende statistische modellen. Dit heeft tevens als voordeel dat de methodiek ook kan worden toegepast voor cohorten waar toepassing van Comsima (vooralsnog) geen optie is.



FIGUUR 2.5 CONCEPT ARCHITECTUUR UKNOW GEDETAILLEERD INGEVULD BIJ TOEPASSING VAN DRIE VERSCHILLENDE METHODEN VOOR TOESTANDBEPALING (STORINGSFREQUENTIE VOOR COHORT, COMSIMA, BEPALING RESTSTERKTE MET METINGEN).

Het eindplaatje van de methodiek is een verzameling degradatiecurves per cohort (waarvan één op basis van Comsima resultaten voor cohorten waar dit mogelijk is). Deze verzameling degradatiecurves geeft de toekomstige storingsfrequentie van een cohort met de daarbij horende onzekerheidsmarges die nodig zijn voor het inschatten van de betrouwbaarheid van de voorspelling.

Voorspellen van storingsfrequenties is een iteratief proces. Door toepassing op een dataset zal duidelijk worden wat de huidige mogelijkheden en onzekerheden van zowel de statistische modellen als Comsima zijn. Op basis hiervan kunnen verbeteringen worden aangebracht. Parallel met het verbeteren van Comsima groeit het aantal storingsgegevens, waardoor cohorten beter kunnen worden geanalyseerd. Eventuele aanpassingen van Comsima zijn geen onderdeel van de methodiekbeschrijving.

Om de haalbaarheid en toepasbaarheid van de methodiek te toetsen zijn de belangrijkste elementen:

- Toetsing van de cohorten op voldoende gegevens en homogeniteit (is er sprake van één degradatiecurve?). De resultaten van deze analyse staan in Hoofdstuk 3.
- Inschatting of de uitkomsten van Comsima voldoende toepasbaar zijn voor vergelijking met storingsgegevens. De resultaten van deze analyse staan in Hoofdstuk 4.

2.4.2 Stappen methodiek

De stappen worden dan als volgt:

1. Het kiezen van een cohort met de daarbij behorende controle op voldoende data voor de gewenste betrouwbaarheid van de berekende storingsfrequenties.
2. Voor de cohorten waar dit mogelijk is, het vertalen van de kennis over de conditie van leidingen naar een vorm voor de degradatiecurve. Dat betekent dat op basis van de uitkomsten van Comsima berekening de meest geschikte wiskundige verdeling (of samenstelling van wiskundige verdelingen) wordt gekozen (bv. een normale verdeling, een Weibull verdeling, een exponentiële verdeling etc.).
3. Het fitten van de statistisch meest gangbare modellen en de op basis van Comsima bepaalde degradatiecurve op de berekende storingsfrequenties in de cohorten inclusief extrapolatie naar leeftijden waar geen betrouwbare storingsfrequenties van zijn.
4. Het omrekenen van de gefitte storingsfrequenties naar een toekomstige storingsfrequentie van het betreffende cohort door rekening te houden met de verdeling van kilometers leidinglengte over de leeftijd van de leidingen.
5. Vergelijken van de resultaten van de verschillende degradatiemodellen.

Deze stappen worden in de volgende paragrafen kort toegelicht.

2.4.3 Kiezen van een cohort

De cohorten moeten zo klein mogelijk zijn om de homogeniteit zo groot mogelijk te maken, maar voldoende data bevatten voor een statistische analyse, zie Hoofdstuk 3. Idealiter representeren de data in een cohort slechts één degradatiecurve met een zo gering mogelijke spreiding van storingsfrequenties rond die curve.

2.4.4 Vertaling kennis leidingen naar vorm degradatiecurve

Deze stap bestaat uit een aantal onderdelen:

- Analyse van de leidingen in het cohort met omgevingskenmerken en liggingsomstandigheden om de spreiding in de invoerparameters voor Comsima te bepalen.
- Simulatie met Comsima inclusief gevoeligheidsanalyse om inzicht te krijgen in de robuustheid van de vorm van de degradatiecurve.
- Vergelijking van de vorm van de Comsima resultaten met verschillende statistische modellen, waarna een keuze wordt gemaakt voor de meest passende. Dit kan ook een samengestelde set zijn, denk bijvoorbeeld aan de 'badkuipcurve', waarbij afhankelijk van de leeftijd van de leidingen een ander storingsgedrag wordt verondersteld.
- Toetsen van de gekozen vorm met een statistische indicator, zie Hoofdstukken 3.3 en 3.5.

2.4.5 Fitten van de algemene vorm van de degradatiecurve op de berekende storingsfrequenties

Deze stap bestaat uit een aantal onderdelen:

- Fitten van de wiskundige degradatiecurve zoals bepaald met Comsim op de berekende storingsfrequenties.
- Fitten van de set gangbare statistische modellen op de berekende storingsfrequenties.
- Statistische toetsing van de volledige set modellen (gangbare statistische modellen aangevuld met de curve op basis van Comsim resultaten indien beschikbaar).
- Vergelijking van de verschillende modellen en indicatie van onzekerheden.
- Voorspellen van storingsfrequenties.

Voor het toetsen van de fit kunnen verschillende methoden worden gebruikt. Toetsen van de gekozen vorm met een statistische maat, bijvoorbeeld met cross-validatie waarbij de data worden opgedeeld in een training- en een validatieset. De vorm van de curve wordt bepaald met de trainingset en getoetst met de validatieset. Door dit te herhalen met meerdere opdelingen wordt voorkomen dat niet alle data kunnen worden gebruikt voor training en wordt tegelijk "overfitting" op de dataset voorkomen.

2.4.6 Omrekenen van de degradatiecurve naar de huidige en toekomstige storingsfrequenties van een cohort

De gemiddelde storingsfrequentie van een cohort kan worden berekend als het aantal leidingkilometers per leeftijd in het cohort bekend is, zie ook het rekenvoorbeeld in Hoofdstuk 2.1.

2.4.7 Vergelijken van de resultaten voor de verschillende degradatiemodellen

Het vergelijken van de resultaten van de verschillende degradatiemodellen geeft inzicht in de meest waarschijnlijke voorspelling en in de onzekerheden.

3 Toepasbaarheid methodiek op bestaande cohortindelingen

3.1 Introductie

In dit hoofdstuk wordt nader ingegaan op methodieken voor het opdelen van de storingsdata in cohorten. Ook wordt de huidige cohortindeling getoetst op toepasbaarheid van de voorgestelde methodiek.

3.2 Algemene criteria voor indeling cohorten

In deze paragraaf worden criteria voor het opsplitsen van cohorten toegelicht.

3.2.1 Groepen opdelen op storingsmechanismen en verklarende variabelen

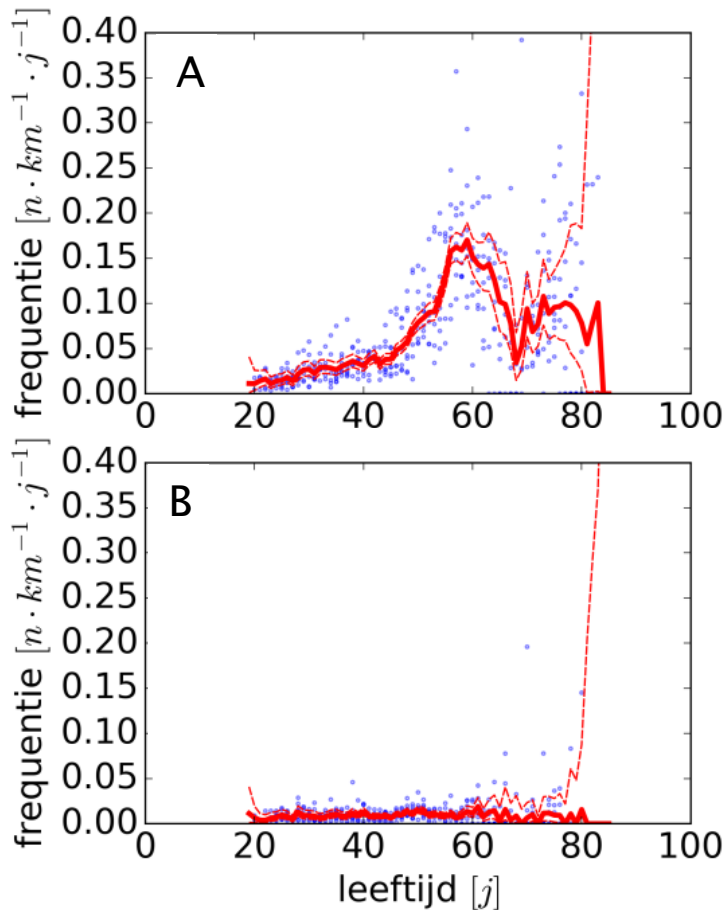
Spreiding binnen een groep kan verkleind worden door de storingsdata op te splitsen in storingsmechanismen. Leidingen storen op basis van verschillende storingsmechanismen, zoals

- Degradatie van één of meer componenten van het leidingsysteem.
 - Verschillende componenten kennen andere vormen van degradatie, zoals buizen en verbindingen. Een onderscheid tussen buis en verbindingen ligt daarom voor de hand.
 - Verschillende materialen kunnen verschillende vormen van degradatie vertonen, met verschillend storingsgedrag tot gevolg. Dit is de eerste groepering in de huidige cohortindeling, omdat er relatief veel bekend is over verschillen tussen degradatieprocessen van de verschillende leidingmaterialen.
- Externe belasting, bijvoorbeeld:
 - Zettingsverschillen kunnen leiden tot hoekverdraaiing in de verbindingen of hoge spanningen in de buisdelen.
 - Derden, zoals graafwerkzaamheden. Het onderscheid tussen spontane storingen en storingen door derden is niet triviaal; schade door derden treedt mogelijk eerder op in buizen die verder gedegradeerd zijn. De mogelijkheid bestaat dat storingen door derden en spontane storingen niet compleet onafhankelijk zijn. Daarnaast is het onderscheid tussen 'derden' en 'spontaan' niet altijd makkelijk te maken en is het dus de vraag of dit altijd goed geregistreerd staat, zie ook van Eijk and van Daal (2013).
 - Andere bronnen van externe belasting, zoals verkeersbelasting of stormschade. Zo is uit onderzoek gebleken dat er een relatie is tussen het optreden van hoge windsnelheden en het optreden van storingen (Wols et al., 2015a) en lokale verkeersbelastingen (Moerman et al., 2016b). Ook bij deze mechanismen is het vaak moeilijk om te bepalen bij welke storingen welke mechanismen optreden.

Verschiede mechanismen kunnen gelijktijdig optreden (bv. leidingdegradatie in combinatie met zettingsverschillen die hoekverdraaiing geven in de verbindingen).

Als voorbeeld laat Figuur 3.1 de storingsfrequenties zien voor spontane storingen versus storingen door derden bij AC leidingen. De spontane storingen laten duidelijk andere kenmerken zien dan de schade door derden:

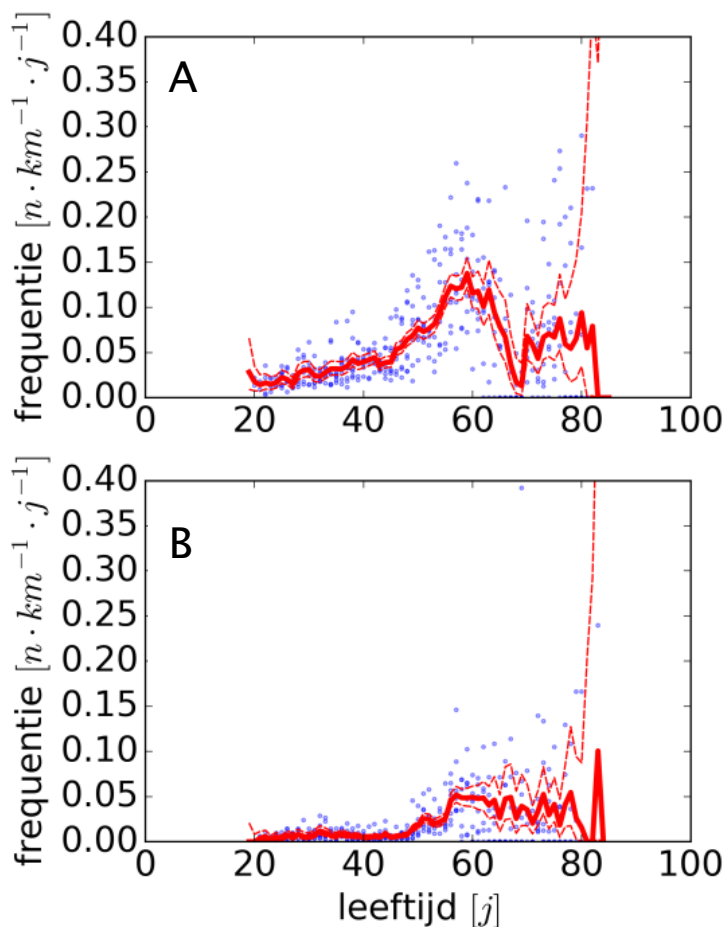
- Gemiddeld zijn de storingsfrequenties veroorzaakt door spontane storingen in dit cohort groter dan de storingsfrequenties veroorzaakt door schade door derden.
- De spreiding in storingsfrequenties is voor spontane storingen veel groter dan bij schade door derden.
- Er is sprake van een heel andere trend.



FIGUUR 3.1 DE STORINGSDATA (·) EN DE GEMIDDELTE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDSINTERVAL (---) VOOR ALLE STORINGEN IN AC LEIDINGEN DIE ZIJN GEREGISTREERD TUSSEN 2009 EN 2015. DE DATA IS OPGESPLITS IN SPONTANE STORINGEN (A) EN STORINGEN DOOR DERDEN (B).

Een ander voorbeeld is gepresenteerd in Figuur 3.2 namelijk de storingsfrequenties van buizen en verbindingen voor AC leidingen. Ook hier zijn een aantal verschillen te zien:

- De storingsfrequenties op verbindingen zijn gemiddeld wat lager dan op leidingen;
- De spreiding van storingsfrequenties bij verbindingen is kleiner dan bij leidingen;
- Er is sprake van een andere trend, maar dit verschil is niet zo duidelijk als in Figuur 3.1. Waar bij schade door derden eigenlijk geen trend te zien is, is bij verbindingen op het oog wel een trend zichtbaar (na een leeftijd van ongeveer 50 jaar lopen de storingsfrequenties op) en daarmee degradatie. Dit is conform verwachting.



FIGUUR 3.2 DE STORINGSDATA (•) EN DE GEMIDDELDE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDINTERVAL (---) VOOR ALLE STORINGEN IN ASBESTCEMENTLEIDINGEN DIE ZIJN GEREGISTREERD TUSSEN 2009 EN 2015. DE DATA IS OPGESPLITS IN STORINGEN OP BUIZEN (A) EN STORINGEN OP VERBINDINGEN (B).

Naast rechtstreekse opdeling op basis van storingsmechanismen (indien mogelijk, waarvan de opdeling in materialen er één van is) kan er verder worden opgedeeld op basis van verklarende variabelen of indicatoren. Dit zijn eigenschappen van de leiding of de omgeving die veel invloed hebben op storingsmechanismen. Voorbeelden zijn:

- Voor AC leidingen is wanddikte sterk bepalend voor de impact van het degradatiemechanisme uitloging, waardoor verschillend storingsgedrag verwacht mag worden voor leidingen met verschillende wanddikte. Helaas is wanddikte meestal slecht bekend.
- Materiaalkwaliteit (te vertalen naar aanlegjaar): er zijn verschillen bekend in materiaaleigenschappen door bijvoorbeeld in de loop der tijd veranderende eisen vanuit wet- en regelgeving (zie bijvoorbeeld de huidige opdeling in cohorten in Tabel 3.1).
- Zettingsgevoeligheid van een gebied: lokale zettingsverschillen kunnen hoekverdraaiing veroorzaken en/of spanningen in de leidingdelen.
- Chemische samenstelling bodem: bij AC en GGJ is het bekend dat dit invloed heeft op de (snelheid van) degradatieprocessen. Een moeilijk aspect is dat deze gegevens lang niet altijd beschikbaar zijn.

- Nabijheid van bronnen van kortstondige spanningspieken, zoals verkeerskruisingen of bomen als indicator voor gevoeligheid voor verkeersbelasting of stormschade.

3.2.2 Groepen opdelen op basis van data eigenschappen

Cohorten kunnen niet oneindig uitgesplitst worden op mechanismen en verklarende variabelen. Eigenschappen van de data limiteren de mogelijkheden tot verdere uitsplitsing:

- Er moeten voldoende storingsen en kilometers leiding in cohorten zitten om met voldoende betrouwbaarheid statistische analyses mee te kunnen doen. Paragraaf 3.3 gaat in meer detail hierop in.
- Verklarende variabelen kunnen alleen gebruikt worden om cohorten verder op te splitsen wanneer ze voor iedere storing geregistreerd zijn. Binnen USTORE is bijvoorbeeld maar beperkte beschikking over data van de chemische samenstelling van de bodem. Daarom moet de bodemsamenstelling in de analyse van paragraaf 3.4 buiten beschouwing worden gelaten, hoewel dit wel een criterium was voor de originele cohorten in tabel 3.1. Buiten USTORE zijn de data mogelijk wel beschikbaar, maar in dit project lag de focus op de USTORE data.

Er zijn alternatieven om data te groeperen puur gebaseerd op statistiek van de data, vanuit de clusteranalyse, zie voor een review van deze methoden o.a. (Fraley and Raftery, 1998; Jain et al., 1999). Er zijn heel veel mogelijkheden, gebaseerd op bijvoorbeeld het optimaliseren van zo min mogelijk variatie in een groep met zoveel mogelijk data (bv. natural breaks, kwantielen of standaard deviatie), het minimaliseren van afstanden tussen data in een groep, geografisch clusteren of het groeperen op datadichtheid. Dat betekent dat voor een goede toepassing inzicht moet bestaan in de oorzaken van verschillen in groepen.

3.3 Statistische evaluatie van cohorten

Cohorten kunnen worden geëvalueerd op significantie. Het rijkt te ver om een uitvoerig overzicht van statistische testen op te stellen. Enkele voorbeelden zijn:

- Poisson confidence interval (Thienen, 2012). Op basis hiervan kan bepaald worden of de gemiddelden van verschillende groepen daadwerkelijk verschillend zijn. In Figuur 3.1 kan op basis van dit betrouwbaarheidsinterval bijvoorbeeld bepaald worden dat storingsen door derden een statistisch significant andere gemiddelde frequentie hebben dan spontane storingsen.
- Hopkins statistiek (zie o.a. Banerjee and Dave (2004)). Hiermee wordt berekend in hoeverre de data multimodaal zijn; uniforme random data geven waarden rond 0.5, data die bestaan uit meerdere groepen hebben waarden dicht bij 1. De aanpak geeft daarmee een indicatie voor homogeniteit van een cohort.
- Analysis Of Variance (ANOVA) met toepassingen door o.a. Rajani (1999). Bij deze methode wordt op basis van een a-priori aangenomen kansverdeling een indicator uitgerekend die zegt met hoeveel zekerheid de twee subgroepen significant verschillen. Daarnaast wordt aangenomen dat data normaal verdeeld is rond de gemiddelde waarden. Hoewel deze aannames belangrijk zijn voor de uitkomst, is dit een goede manier om het verschil tussen groepen kwantitatief te maken.
- Mean Likelihood Estimation (MLE) (Scheidegger et al, 2015). Bij deze methode wordt op basis van een a-priori aangenomen kansverdeling onder andere de onzekerheid in de fitparameters van de kansverdeling op de data berekend. Hoewel deze aannames belangrijk zijn voor de uitkomst, kan deze methode niet alleen gebruikt worden om het verschil tussen groepen te bepalen, maar ook om te bepalen of het extrapoleren van de data leidt tot significant verschillende voorspellingen.

3.4 Huidige cohortindeling

In BTO onderzoek uit 2010 met een actualisatie in 2015 is op basis van literatuuronderzoek, een analyse van storingsgegevens uit USTORE, expertkennis en praktische beschikbaarheid van data de volgende cohortindeling voorgesteld (Beuken and Mesman, 2011b; Beuken and Mesman, 2015):

TABEL 3.1 GEACTUALISEERDE COHORTEN (BEUKEN AND MESMAN, 2015).

Cohort	Getoetst in 3.5	Materiaal	Aanlegperiode	Diameter	Omgeving
1	*	AC	<1965	<250 mm	Kalkarme grond
2		AC	<1965	<250 mm	Kalkhoudende grond
3	*	AC	<1965	≥250 mm	Kalkarme grond
4		AC	<1965	≥250 mm	Kalkhoudende grond
5	*	AC	≥1965	<250 mm	Kalkarme grond
6		AC	≥1965	<250 mm	Kalkhoudende grond
7	*	AC	≥1965	≥250 mm	Kalkarme grond
8		AC	≥1965	≥250 mm	Kalkhoudende grond
9	*	PVC	<1980	<160 mm	Kans op verschilzettingen
10		PVC	<1980	<160 mm	Beperkte kans op verschilzettingen
11	*	PVC	<1980	≥160 mm	Kans op verschilzettingen
12		PVC	<1980	≥160 mm	Beperkte kans op verschilzettingen
13	*	PVC	≥1980	<160 mm	Kans op verschilzettingen
14		PVC	≥1980	<160 mm	Beperkte kans op verschilzettingen
15	*	PVC	≥1980	≥160 mm	Kans op verschilzettingen
16		PVC	≥1980	≥160 mm	Beperkte kans op verschilzettingen
17	*	PE	<1980		
18	*	PE	≥1980	<100 mm	
19	*	PE	≥1980	≥100 mm	
20	*	GGIJ		<250 mm	Agressieve bodem
21		GGIJ		<250mm	Niet agressieve bodem
22	*	GGIJ		≥250 mm	Agressieve bodem
23		GGIJ		≥250 mm	Niet agressieve bodem
24	*	NGIJ	<1990		
25	*	NGIJ	≥1990		
26	*	Staal zonder KB			
27		Staal met KB			
28		Beton			
29		GVK			

De huidige cohortindeling is gebaseerd op een combinatie van opdeling op basis van storingsmechanismen, verklarende variabelen en data-eigenschappen. Zo komt de opdeling in materialen voort uit kennis over verschillende storingsmechanismen en veroudering van materialen. De opdeling in jaartallen komt o.a. uit observaties in de storingsdata, kennis van veranderingen in wet- en regelgeving en kennis over veranderingen in materiaalsamenstelling of productieprocessen. De opdeling in diameters komt vooral uit observaties in de storingsdata gecombineerd met kennis van storingsmechanismen en verklarende variabelen.

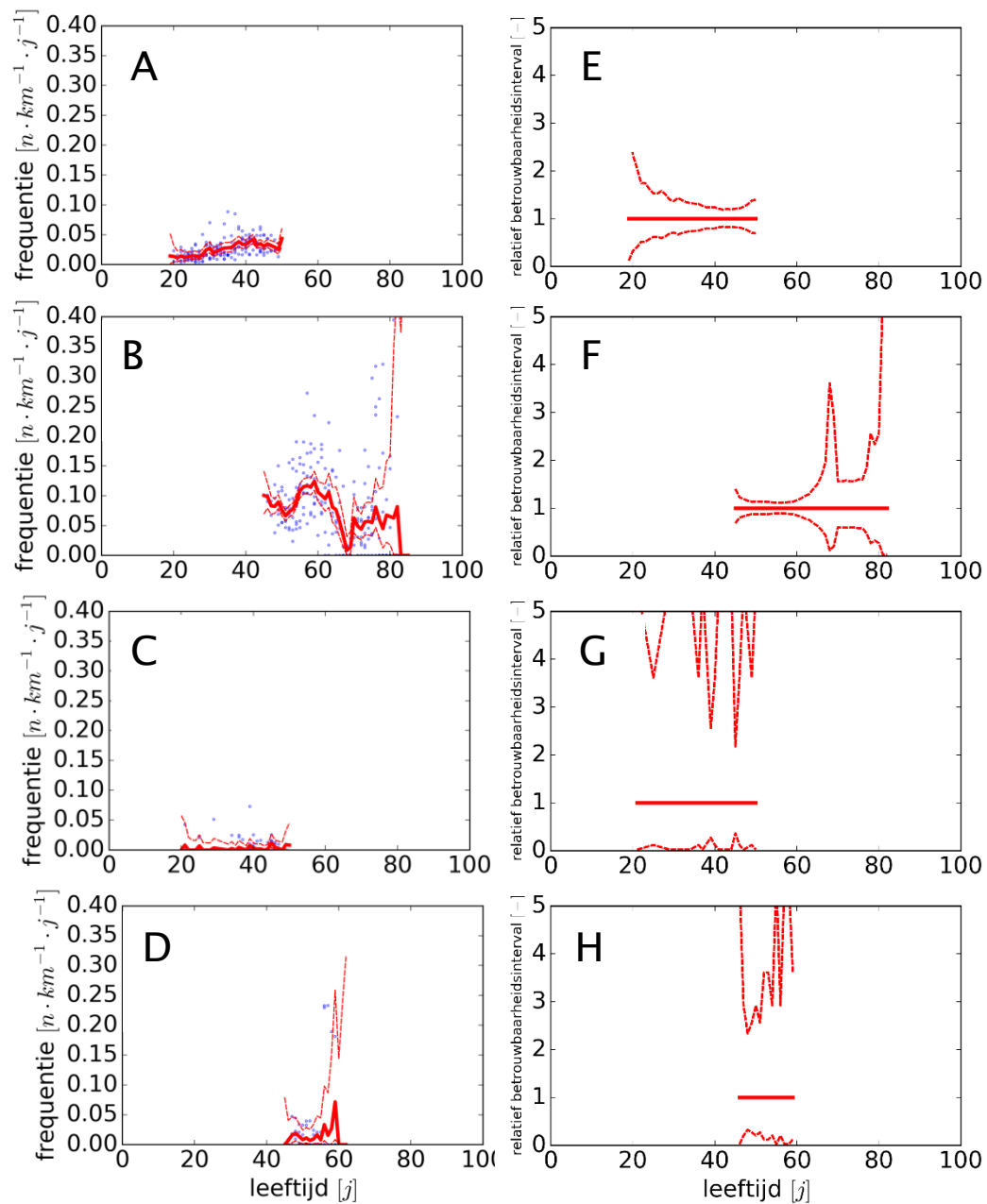
3.5 Beoordeling huidige cohortindeling

In deze paragraaf wordt een indicatie gegeven of cohorten voldoende storingsgegevens en leidingkilometers bevatten voor een zinvolle statistische analyse (trendbepaling en extrapolatie).

Het huidige onderzoek is beperkt tot informatie beschikbaar in USTORE. Van de cohorten in Tabel 3.1 is in USTORE geen informatie beschikbaar over de ligging in kalkarme of agressieve bodem of zettingsgevoelige grond. Hierop is dus in dit onderzoek geen onderscheid te maken. Ook is er geen informatie beschikbaar over de aanwezigheid van kathodische bescherming (KB). Daarnaast zijn er geen storingen op GVK aanwezig in de dataset. Op basis daarvan blijven er 17 cohorten over van de 29 die statistisch te toetsen zijn. Van beton waren maar 16 storingen beschikbaar, dus die is als statistisch onbetrouwbaar benoemd. Er zijn dus uiteindelijk 16 cohorten getoetst (zie de kolom met * in Tabel 3.1), de resultaten daarvan worden in onderstaande paragraaf besproken.

Van Thienen (2012) heeft op basis van een Poisson distributie een indicatie gegeven hoeveel data nodig zijn voor voldoende statistische betrouwbaarheid. Op basis van die analyse is voor ieder cohort een berekening gemaakt bij de storingsfrequenties voor het 95% betrouwbaarheidsinterval van deze berekende storingsfrequentie. Naast het absolute betrouwbaarheidsinterval bekijken we ook de relatieve betrouwbaarheid (95% marge gedeeld door de gemiddelde storingsfrequentie). Het gebruik van beide geeft niet alleen zicht op de betrouwbaarheid van de storingsfrequenties als in de invloed van variaties in de storingsfrequentie zelf. Als eerste indicatie beschouwen we in deze rapportage een factor 2 relatieve onzekerheid als grens voor statistische betrouwbaarheid (hier gedefinieerd als voldoende data voor trendanalyse).

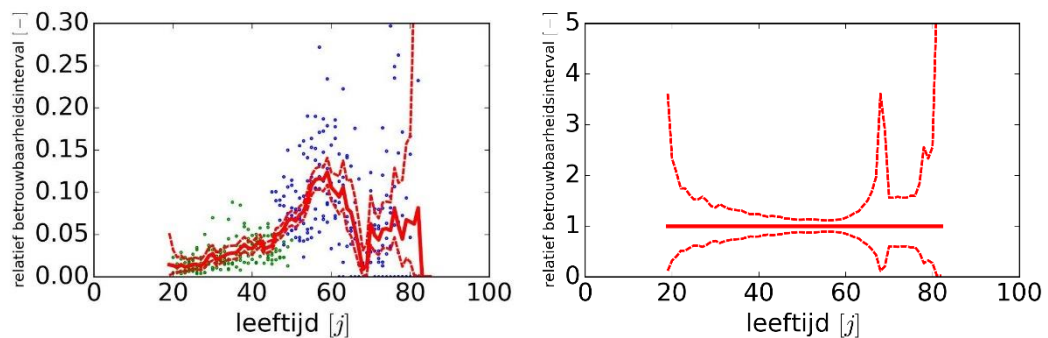
Een Poisson distributie gaat uit van een proces met discrete uitkomsten, d.w.z. in een specifieke periode treedt een gebeurtenis 0, 1, 2 ... keer op, maar niet 1,3 keer. Verder gaat de distributie uit van het vóórkomen van de gebeurtenissen met een in principe vaste frequentie en onafhankelijk van eerder opgetreden gebeurtenissen. De aanname van een vaste frequentie gaat in feite niet op. Als cohorten homogeen zijn, dan is de verwachting dat op de schaal van een enkel cohort bij een specifieke leeftijd van een leiding het aantal storingen zal zijn verdeeld volgens een Poisson distributie. Echter, in praktijk zullen omgevingsfactoren waarop de huidige cohorten niet zijn ingedeeld voor afwijkingen van deze Poisson verdeling zorgen. Dit zorgt voor een in praktijk grotere spreiding dan mag worden verwacht op basis van de Poisson verdeling. Dit betekent dat de resultaten alleen een maat zijn voor statistische betrouwbaarheid, geen exacte bepaling.



FIGUUR 3.3 DE STORINGSDATA (·) EN DE GEMIDDELTE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDINTERVAL (---) VOOR SPONTANE STORINGEN IN AC LEIDINGEN (2009 -2015) (A-D). DEZELFDE FIGUREN, MAAR DAN GENORMALISEERD MET DE GEMIDDELTE STORINGSFREQUENTIES STAAN IN (E-H). COHORT AANGELEGD NA 1964 MET DIAMETERS TUSSEN 0-249 MM (A EN E), COHORT AANGELEGD TOT 1964 MET DIAMETERS TUSSEN 0-249 MM (B EN F), COHORT AANGELEGD NA 1964 MET DIAMETERS >250 MM (C EN G) EN COHORT AANGELEGD VOOR 1964 MET DIAMETERS >250 MM (D EN H).

Figuur 3.3 laat de statistische betrouwbaarheid zien van de cohorten met AC leidingen (jonger of ouder dan 1964 en met diameters kleiner of groter dan 250 mm). Een aantal zaken springen in het oog:

- De statistische betrouwbaarheid is niet constant, maar leeftijdsafhankelijk. Dit komt doordat niet iedere leeftijd gelijk vertegenwoordigd is in het leidingbestand waarvoor storingen werden geregistreerd.
- De statistische betrouwbaarheid van de cohorten met diameters groter dan 250 mm is aanzienlijk lager dan de cohorten met diameters kleiner dan 250 mm en voldoen niet meer aan de norm van 2. Deze cohortsplitsing kunnen daarom momenteel als niet zinvol voor statistische analyses worden beschouwd.
- De statistische betrouwbaarheid van de cohorten met diameters kleiner dan 250 mm is voldoende voor statistische analyses voor het grootste deel van de aanwezige leidingleeftijden. Voor leidingen jonger dan 25 jaar en voor leidingen ouder dan 75 jaar neemt de betrouwbaarheid sterk af, dit deel van de dataset moet dus ofwel niet worden meegenomen in een trendbepaling ofwel gewogen met de betrouwbaarheid.
- Er is een opvallende sprong in relatieve statistische betrouwbaarheid rond een leeftijd van 70 jaar. De absolute afwijking is hier niet heel groot, omdat de storingsfrequentie erg laag is. Ook de data rond 70 jaar moeten dus gewogen worden meegenomen, net als de data ouder dan 75 jaar en jonger dan 25 jaar.
- Bij het punt waarop de cohorten zijn opgeknipt, is de statistische betrouwbaarheid iets slechter. Dit heeft te maken met de omrekening van aanlegjaar naar leeftijd. De degradatiecurve is een functie van leeftijd van leidingen. Omdat er 7 jaren met storingsregistraties zijn, zijn er voor iedere leeftijd 7 aanlegjaren waarmee de storingsfrequentie wordt bepaald. Rond de “cohort-knip” zijn er steeds minder aanlegjaren met data voor de bepaling van de storingsfrequentie beschikbaar. (De cohorten hebben wat leeftijd betreft een kleine overlap, de aanlegjaren zijn dus verdeeld over die overlap). Dit illustreert de invloed van meer data op de betrouwbaarheid van statistische analyses.
- De jongere AC leidingen (<50 jaar) hebben een aanzienlijk lagere storingsfrequentie dan oudere leidingen (>50 jaar). Het is echter niet duidelijk of er een duidelijke splitsing is in trend, waardoor een opdeling in twee cohorten noodzakelijk is. Hierop wordt in Hoofdstuk 4 dieper ingegaan, maar ter illustratie is ook het figuur weergegeven van AC leidingen zonder verdere cohortindeling (Figuur 3.4).



De resultaten van de overige cohorten staan in Bijlage I.

3.6 Discussie en conclusies

De analyses laten een aantal zaken zien:

- De materialen AC, PVC, GGJ en staal zijn geschikt voor toepassing van de methodiek voor het voorspellen van toekomstige storingsfrequenties zoals geformuleerd in Hoofdstuk 2.5.
- De betrouwbaarheid van storingsfrequenties is niet gelijk voor alle leeftijdscategorieën binnen cohorten. Een trend kan alleen worden bepaald op de statistisch betrouwbare leeftijdscategorieën.
- De cohorten met grotere diameters zijn bij alle materialen statistisch onvoldoende betrouwbaar voor statistische analyses.
- De storingsfrequenties van PVC en PE zijn in het algemeen lager dan de storingsfrequenties van de overige materialen.
- Vanuit het oogpunt van trendanalyse en extrapolatie van storingsfrequenties lijkt er geen duidelijke reden voor de meeste indelingen buiten het onderscheid in materialen. Bij AC ligt de storingsfrequentie van oudere leidingen weliswaar hoger dan voor jongere, maar dit is in lijn met de verwachting van een degradatiecurve.

De analyses in dit hoofdstuk zijn gericht op toepasbaarheid van de cohorten voor trendanalyse en het voorspellen van toekomstige storingsfrequenties. De conclusies en aanbevelingen zijn dus ook geschreven vanuit het oogpunt van trendanalyse. Er kunnen andere, praktische redenen zijn om voor andere doelen dan trendbepaling een andere opdeling te hanteren. Bijvoorbeeld:

- Kortdurende variaties in de storingsfrequenties die niet relevant zijn voor het bepalen van de degradatiecurve, maar voor saneringsprioritering wel van belang zijn. Voorbeelden zijn bijvoorbeeld de aanlegjaren 1940-1945 als daar bekend van is dat de leidingkenmerken toen verschilden van de omliggende periode.
- Zeer lokale aspecten kunnen niet worden meegenomen, bv. een aantal leidingen in zeer zettingsgevoelige bodem.
- De storingsvoorspelling voor een lokale, specifieke leiding zal altijd afwijken van de statistiek van een heel cohort ("de gemiddelde leiding" bestaat niet).

De huidige cohortindeling is gebaseerd op een combinatie van expertkennis en data-eigenschappen. De data-eigenschappen zijn echter op een kwalitatieve manier gebruikt, wat tot interpretatieverschillen kan leiden en daarmee tot discussie. Voor een kwantitatieve onderbouwing is het interessant om te onderzoeken of het geautomatiseerd opdelen van de storingsdata in cohorten meerwaarde biedt.

In het huidige onderzoek is de homogeniteit van de huidige groepen en het bepalen van een realistisch aantal groepen niet onderzocht. Het is aan te bevelen om bij toepassing van de methodiek voor het voorspellen van storingsvoorspellingen in elk geval een indicatie voor homogeniteit te berekenen bij het genereren van cohorten.

Informatie die niet in USTORE aanwezig is, is mogelijk wel bij de individuele drinkwaterbedrijven beschikbaar. Het is aan te bevelen om dit nader te onderzoeken en te analyseren of het zinvol en gewenst is om de informatie die aan USTORE wordt aangeleverd op basis hiervan aan te passen. Vanaf 2018 wordt de storingsregistratie in USTORE aangepast met geografische gegevens, het is de moeite waard om de cohorten dan opnieuw tegen het licht te houden.

Zo staan in USTORE gegevens die nu niet zijn geanalyseerd, maar mogelijk wel relevant zijn voor verschillen in berekende storingsfrequenties. Bijvoorbeeld:

- Verkeersbelasting;
- Nabijheid bomen;
- Ligging in grondwater;
- Gronddekking.

Zo kunnen leidingen onder drukke wegen structureel andere storingsfrequenties laten zien dan leidingen onder trottoirs, zie ook Moerman et al. (2016b). Op het moment dat in USTORE geografische gegevens beschikbaar zijn, waardoor de storingsdata kunnen worden gecombineerd met ruimtelijke gegevens, zoals een basisbestand bomen of verkeerswegen, biedt dit mogelijkheden voor het opnieuw indelen van cohorten.

De resultaten in dit rapport zijn gegenereerd met de aanname dat er geen fouten in de data aanwezig zijn bij het bepalen van het betrouwbaarheidsinterval. Fouten in de registraties beïnvloeden berekende storingsfrequenties en daarmee ook de trends. Voorbeelden zijn fouten in:

- Aanlegjaren en daarmee leeftijdsbepaling van leidingen;
- Materialen, bv. onderscheid tussen NGIJ en GGJ;
- Storingsoorzaak, zoals spontane versus niet spontane storingen of buizen versus verbindingen.

De dataset is niet op dit soort fouten gecontroleerd of gecorrigeerd. Het verbeteren en monitoren van de datakwaliteit van de storingsregistraties is een doorlopend traject binnen USTORE (Beuken and Moerman, 2017).

4 Haalbaarheidsanalyse kiezen van een kansverdeling

4.1 Inleiding

De methodiek voor het voorspellen van toekomstige storingsfrequenties van cohorten gebruikt zowel degradatiecurves vanuit statistische modellen als een degradatiecurve gebaseerd op resultaten van Comsima. In dit hoofdstuk wordt de haalbaarheid van beide stappen beschouwd.

4.2 Datafit statistische modellen

Als voorbeeld is in deze paragraaf voor alle AC leidingen (exclusief verbindingen) kwalitatief een fit bepaald voor degradatiecurves van een aantal verschillende vormen:

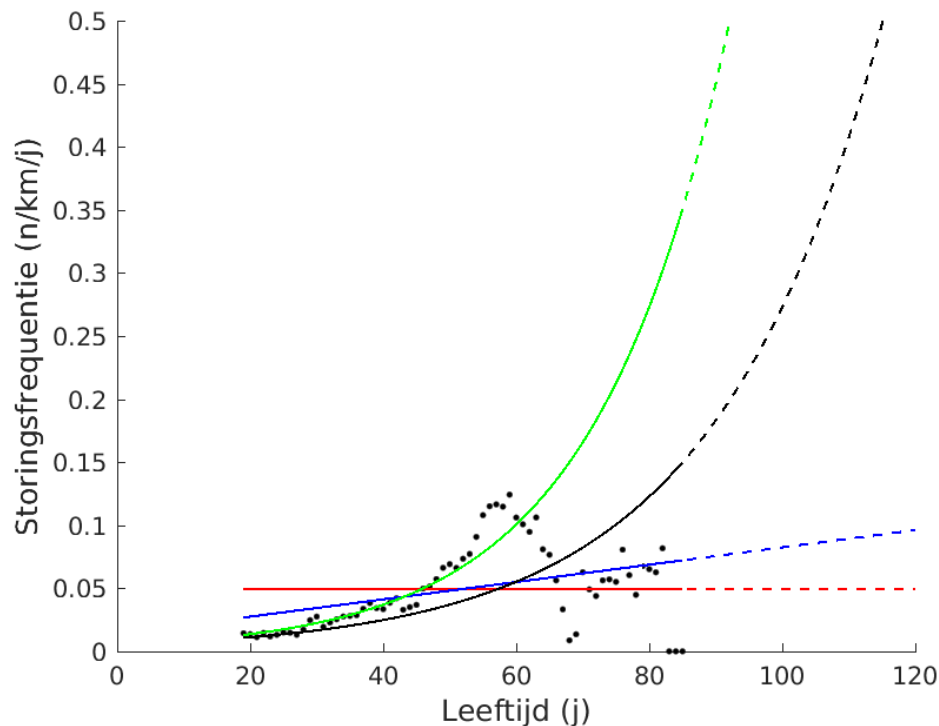
- twee verschillende exponentiële degradatiecurves (de exponentiële vorm komt overeen met de momentane storingsfrequentie van de Gumbel verdeling);
- een lineaire degradatiecurve (deze vorm komt overeen met de momentane storingsfrequentie van de Rayleigh verdeling);
- een constante degradatiecurve (deze vorm komt overeen met de momentane storingsfrequentie van de exponentiële verdeling).

Er zijn nog veel meer vormen mogelijk, waaronder ook curves die zijn samengesteld uit meerdere vormen. De bovenstaande vormen zijn ter illustratie.

De resultaten staan weergegeven in Figuur 4.1. De exponentiële degradatiecurves zijn zo gekozen dat ze een groter gewicht geven aan een ander deel van de data. Voor beide exponentiële modellen komt de curve beter overeen met de jongere storingsfrequenties en slechter met de oudere storingsfrequenties. Voor de lineaire degradatiecurve komen zowel jongere als oudere storingsfrequenties redelijk overeen, maar wordt de piek tussen 40 en 70 jaar slecht gerepresenteerd. De constante curve lijkt een slechte fit met vrijwel alle storingsfrequenties te geven.

De onderlinge verschillen tussen de vier degradatiecurves zijn zeer groot, er is sprake van factoren verschil in storingsfrequenties voor leeftijden met storingsfrequenties en nog veel grotere verschillen voor de extrapolatie richting de periode zonder storingsfrequenties. Kwalitatief is bijvoorbeeld te argumenteren dat de lineaire trend richting de toekomst mogelijk van de vier modellen de meest betrouwbare voorspelling geeft, maar een keuze voor dit model betekent tegelijk dat er een grote onderschatting van de storingsfrequentie van het totale cohort wordt gemaakt door de verwaarlozing van de piek tussen 40 en 80 jaar.

Het is duidelijk dat zonder expertkennis van de lokale situatie geen duidelijke onderbouwing mogelijk is. Dit illustreert de meerwaarde van het gebruik van Comsima om het beste degradatiemodel te kiezen, in Comsima kan expertkennis worden ingebracht en het geeft een fysieke onderbouwing. Voor de cohorten waar dit niet mogelijk is, kan een alternatief zijn om te kiezen voor meerdere modellen voor verschillende periodes, of de modellen alleen te gebruiken als indicatie voor de onzekerheid richting de toekomst.



FIGUUR 4.1 KWALITATIEVE TOETSING VAN VERSCHILLENDE DEGRADATIECURVES. DE ZWARTE PUNTEN GEVEN DE STORINGSFREQUENTIES WEER, DE GROENE EN ZWARTE LIJN EEN EXPONENTIËLE TREND, DE BLAUWE LIJN EEN LINEAIRE TREND EN DE RODE LIJN EEN CONSTATE TREND. DE GESTREEPTE LIJN GEEFT DE EXTRAPOLATIE VAN DE DEGRADATIECURVES NAAR PERIODES ZONDER STORINGSDATA WEER.

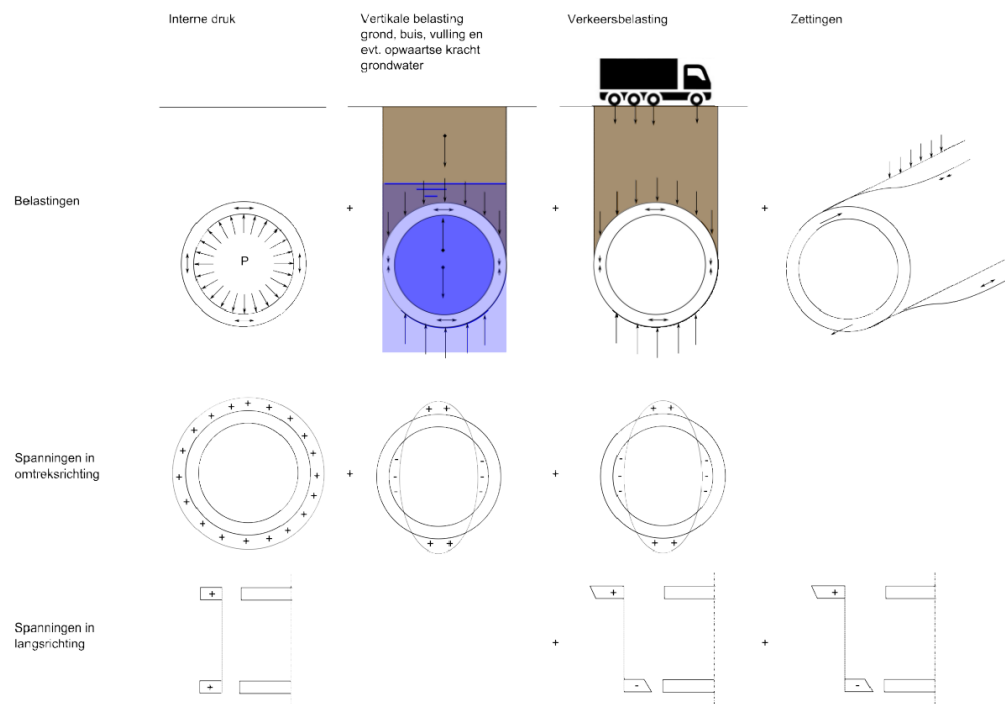
4.3 Op basis van Comsima

4.3.1 Inleiding Comsima

Comsima berekent de spanningen op een leiding, zie Figuur 4.2. Daarnaast bevat Comsima (degradatie)modellen voor buiseigenschappen als functie van leeftijd en de normen voor materiaalbelasting. Uit spanning, degradatie en norm wordt het moment bepaald waarop de conditie van de buis onvoldoende is. In deze rapportage wordt aangenomen dat als de berekende conditie van een doorgerekende, werkelijke leiding niet meer voldoet, dit ook het moment is dat de leiding stopt volgens de definitie in Hoofdstuk 2.

De buiseigenschappen als functie van leeftijd worden geëxtrapoleerd op basis van statische eigenschappen, namelijk een toestand bij aanvang van de levensduur, en een model voor de degradatie van de eigenschappen. Het degradatiemodel wordt in 2017 in Comsima ingebouwd voor de materialen AC, PVC en grijs gietijzer op basis van het KSLB 2.0 (Beuken and Mesman, 2011a). Op basis van de geschatte onzekerheden in de invoer genereert Comsima via een Monte Carlo simulatie de kans op falen als functie van leeftijd.

Comsima maakt in de berekening van zettingen onderscheid in buisdelen en verbindingen. Comsima berekent twee effecten van de zetting parallel: de hoek in de verbinding en de spanning in de buis door vervorming (de spanning in de verbinding wordt niet meegenomen in de berekening). De spanning in de buis wordt meegenomen in de reguliere spanningsbalans. De leiding stopt niet alleen door een negatieve spanningsbalans, maar ook wanneer de verbinding over zijn maximale hoekverdraaiing heen gaat.



FIGUUR 4.2 FIGUUR 2.1 UIT BTO 2015.082, OVERZICHT VAN DE BEREKENDE SPANNINGEN IN COMSIMA.

4.3.2 Relatie Comsima resultaten en degradatiecurves van cohorten

Comsima berekent de spanningen voor een leiding met eigenschappen en onder condities zoals voorgeschreven in de invoerparameters (of zoals in de Monte Carlo simulatie een spreiding spanningen voor dezelfde leiding onder een spreiding van leidingeigenschappen en condities). Dit is anders dan een trend door de storingsfrequenties over de tijd van een cohort met leidingen. Voor de vertaling van de rekenresultaten van de Monte Carlo simulatie van Comsima naar de degradatiecurve van een cohort zijn de volgende stappen toegepast:

1. De spreiding in de leiding- en omgevingskenmerken binnen de cohorten wordt binnen Comsima als onzekerheid meegenomen in de Monte Carlo simulatie. Hierbij wordt aangenomen dat de leidingen in een cohort omschreven kunnen worden als een gemiddelde leiding met een spreiding in leidingkenmerken, liggingssomstandigheden en omgevingskenmerken:
 - De gesimuleerde leiding in Comsima komt dan overeen met de gemiddelde leiding uit het cohort.
 - De onzekerheden in de invoer van de Monte Carlo simulatie komen overeen met de spreiding in de leidingkenmerken, liggingssomstandigheden en omgevingskenmerken in het cohort.

Met deze stap genereert Comsima een numerieke benadering van de waarschijnlijkheidsverdeling van storen binnen het cohort (naar keuze weergegeven als kansdichtheidsfunctie of als momentane storingsfrequentie). Deze waarschijnlijkheidsverdeling wordt niet 1 op 1 vertaald in de hazard functie voor storingsfrequenties vanuit USTORE, maar wordt in de voorgestelde methodiek alleen gebruikt om de juiste vorm van de waarschijnlijkheidsverdeling te kiezen (bv. lognormaal of Gausisch).

2. In de praktijk bevat een cohort meer dan één "gemiddelde leiding". Bijvoorbeeld, als er 100 kilometer PVC 63 mm ligt, 100 km 75 mm en 100 km 90 mm PVC, dan is er geen sprake van een gemiddelde leiding van $\varnothing 76$ mm, maar van 3 gemiddelde leidingen met elk hun eigen spreiding in leidingkenmerken, liggingomstandigheden en omgevingskenmerken. De interface van Comsima staat momenteel de invoer van slechts één set onzekerheden rond één gemiddelde leiding toe (dit is eenvoudig in te bouwen). Daarom rekenen we in dit project alleen resultaten uit voor specifieke diameters en worden de resultaten alleen kwalitatief beschouwd.
3. De uitkomsten van Comsima zijn zo goed als de kennis van leidingdegradatie en omgevingskenmerken die als invoer wordt gebruikt. De onzekerheden zijn in praktijk slecht bekend. Daarom worden de resultaten van Comsima in dit project alleen gebruikt als haalbaarheidsindicator voor de methodiek. Meer inzicht in spreiding in invoerparameters zal uit casestudies moeten volgen.

4.3.3 Invoer en onzekerheden in Comsima

Als input gebruikt Comsima de volgende parameters voor de verschillende materialen:

Parameter	Relevant voor materiaal
Nominale diameter	Alle
Wanddikte	Alle
Zettingsnelheidsverschil	Alle
Zettingsafstand	Alle
Interne druk	Alle
Afstand tussen verbindingen	Alle
Afname wanddikte	AC, GGIJ
Weerstand langzame scheurgroei	PVC
Diepteligging	Alle
Dichtheid leidingmateriaal	Alle
Indicatie verkeersdruk	Alle
Young's modulus	Alle
Grondwaterstand	
Elasticiteit bodem	
Dichtheid bodem (droog/nat)	

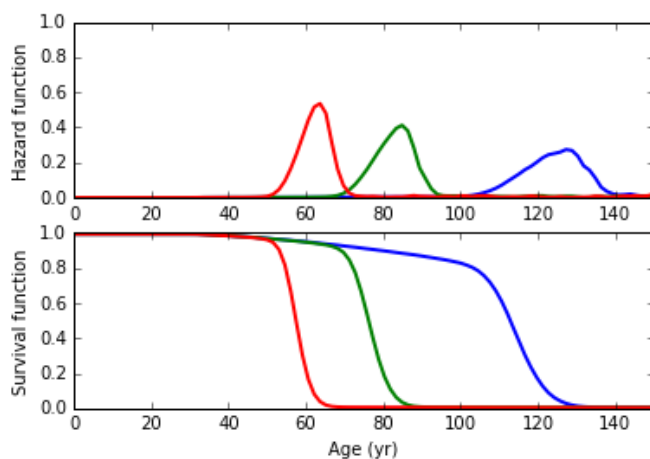
De waarden voor de input in combinatie met de spreiding hierin bepalen gezamenlijk de vorm van de waarschijnlijkheidsverdeling. De uitdaging is het kiezen van de beste waarden voor de parameters behorend bij het cohort. Andersom kunnen de storingsregistraties ook inzicht geven in de kennis van leidingmaterialen en effect van belastingen. Als op basis van de beste kennis een curve ontstaat die niet lijkt op de berekende storingsfrequenties, dan is dit een duidelijke indicatie dat de beschikbare kennis onvoldoende is. In de methodiek voor het voorspellen van storingsfrequenties moet dus een stap worden ingebouwd die controleert of de gevonden fit voldoende is voor een betrouwbare schatting.

De onzekerheid (samengevat als spreiding) in iedere parameter heeft invloed op de vorm van de resultaten. De spreiding kan zowel een norm zijn als een uitdrukking van onze kennis. Een willekeurig voorbeeld; een spreiding in de wanddikte kan worden ingevuld als een maximaal toegestane tolerantie van de norm of als de kennis dat in praktijk gevonden afwijkingen veel groter kunnen zijn, waarbij de norm een minimum is. Hoe beter de daadwerkelijke spreiding wordt beschreven en gebruikt, hoe dichter de resultaten bij de werkelijkheid zullen liggen. Een te klein gekozen spreiding suggereert onrealistisch zekere resultaten, een te ruim gekozen spreiding resulteert in een slecht te interpreteren resultaat.

Inzicht in de belangrijke parameters en de spreiding van de waarde van deze parameters binnen de cohorten is dus essentieel.

Voorbeelden van belangrijke parameters voor AC met bekend aanzienlijke onzekerheden zijn:

- De initiële wanddikte van AC buizen. Het is bekend dat de wanddikte sterk verschilt tussen en binnen leidingen, soms over zeer korte afstanden. Dit komt doordat het productieproces van AC buizen kan leiden tot variërende wanddikten: de cataloguswaarde is een minimum, de werkelijke wanddikte is afhankelijk van het aantal wikkels.
- De uitlogingssnelheid, ofwel de afname in wanddikte bij AC. Het is waargenomen dat de uitlogingssnelheid kan variëren over de tijd. Dit geldt met name voor inwendige uitloging bij in de tijd variërende waterkwaliteit. Daarnaast is waargenomen dat met name uitwendige uitlogingssnelheid kan variëren over de plaats op de leiding, vermoedelijk vanwege de variërende concentraties van aanvallende stoffen in de bodem. Beschikbare gegevens komen voornamelijk uit thymolphaleinetesten op uitgenomen ringen. Deze lokale testen kunnen geen volledig beeld geven omdat uitlogingssnelheid significant kan variëren over de lengte van een leiding. Figuur 4.3 laat de invloed van verschillen in uitlogingssnelheid zien op de momentane storingsfrequentie. Figuur 4.3 laat een voorbeeld zien van momentane storingsfrequenties berekend door Comsima bij een aantal uitlogingssnelheden.
- De diepteligging van leidingen is meestal slecht bekend. Bijvoorbeeld, in woonwijken wordt standaard een aanlegdiepte van 1 meter gehanteerd, maar in praktijk kan deze lokaal afwijken. Ook zijn er zettingen in de loop van de tijd en wegreconstructies waarbij de weg wordt opgehoogd. Een leidingdiepte van 2 meter kan voor komen. (Afhankelijk van het beleid van de drinkwaterbedrijven en de bereikbaarheid van leidingen zullen hele diepe leidingen in het algemeen worden vervangen bij reconstructies).



FIGUUR 4.3 REKENRESULTATEN COMSIMA VOOR DE BEREKENDE MOMENTANE STORINGSFREQUENTIE VOOR AC MET WANDDIKTES VAN 12 MM EN UITLOGINGSSNELHEDEN VAN 0.1 MM/JAAR (BLAUW), 0.15 MM/JAAR (GROEN) EN 0.2 MM/JAAR (ROOD).

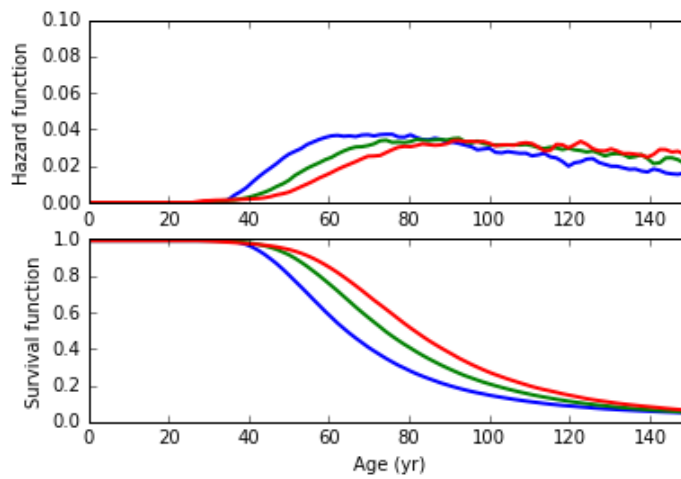
4.3.4 Voorbeelden voor AC en PVC met een vergelijking met USTORE storingsgegevens

Voorbeeld: AC

Figuur 4.4 laat voor drie wanddiktes de momentane storingsfrequentie als functie van leeftijd zien zoals berekend door Comsima met de invoerparameters in Tabel 4.1. Deze figuren zijn puur indicatief, de invoerparameters zijn niet afgestemd op een bestaand cohort. Voor deze invoerparameters is te zien dat een leiding gedurende de eerste 40 jaar vrijwel niet zal storen, waarna de storingsfrequentie oploopt tot een maximum, waarna de storingsfrequentie stabiliseert. De precieze leeftijd is afhankelijk van de gekozen waarden voor de invoerparameters en kan niet uit deze indicatieve figuren worden afgelezen. Wel is duidelijk dat de storingsfrequentie van AC niet op hoeft te blijven lopen als functie van leeftijd en dat er verschillen bestaan in de ontwikkeling van storingsfrequenties voor verschillende wanddiktes. Aangezien de uitloging (gemiddelde en spreiding) en daarmee de wanddikte-afname voor alle wanddiktes gelijk is gekozen, bepaalt de gekozen aanvangswanddikte het moment waarop de leiding voor het eerst begint te storen en deze is groter voor grotere diameters.

In Figuur 4.5 staat de storingsfrequentie als functie van leeftijd voor AC (alle diameters), zoals berekend vanuit de storingsregistraties in USTORE (alleen spontane storingen en alleen storingen op buisdelen). Te zien is dat de storingen toenemen met leeftijd tot een leeftijd van zo'n 60 jaar, waarna er een dip volgt rond 68-70 jaar (aanlegperiode vlak na de Tweede Wereldoorlog), waarna de storingsfrequenties weer stabiliseren. De oudst berekende storingsfrequenties (vanaf 77 jaar) zijn onbetrouwbaar door een zeer beperkte leidinglengte in het cohort. Uit deze curve mag worden geconcludeerd dat de storingsregistraties niet per definitie een voortdurende stijging met leeftijd laten zien. De onzekerheden bij de hoogste leeftijden zijn te groot om een definitieve uitspraak te doen over de trend aan het eind van het data bereik, maar de data heeft voldoende betrouwbaarheid om de daling rond 60 jaar te bevestigen. Vergelijking met de stabiliserende trend die wordt voorspeld met Comsima versterkt die conclusie. Dit demonstreert de meerwaarde van het ondersteunen van de trendanalyse met systeemkennis.

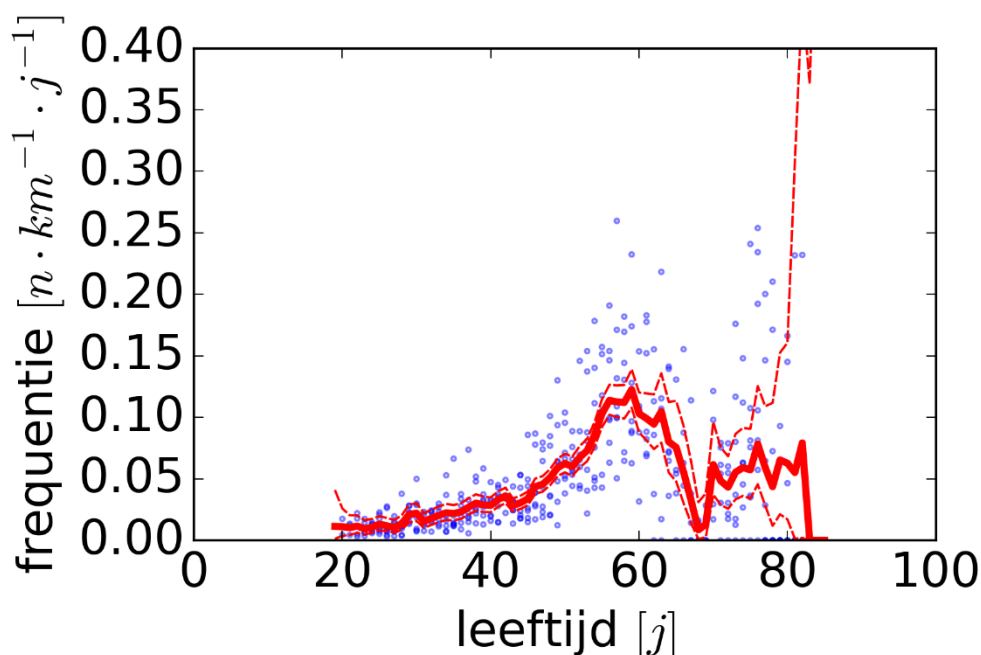
De stabilisatie en zelfs afname van storingsfrequenties met leeftijd komt door de eigenschappen van het deel van de leidingen dat nog voldoende conditie heeft op een bepaalde leeftijd. Bij voortdurende degradatie van leidingen is een oplopende storingsfrequentie met leeftijd te verwachten; leidingen met een specifieke leeftijd storen dan steeds vaker als de leidingen verouderen. Als een deel van de leidingen in een cohort minder snel degradeert dan een ander deel (bv. door minder zetting of minder aantasting door verschillen in lokale liggingomstandigheden), dan zal de storingsfrequentie afvlakken of zelfs afnemen als functie van leeftijd. De afvlakkende curves in Figuur 4.4 laten zien dat met de gekozen spreiding in de invoerparameters er leidingen zijn die een steeds grotere belasting ondervinden door zetting en steeds minder belasting kunnen hanteren door uitloging, maar dat er in dezelfde Monte Carlo simulatie ook leidingen zijn waar vrijwel niets mee gebeurt in termen van conditie.



FIGUUR 4.4 REKENRESULTATEN COMSIMA VOOR DE BEREKENDE MOMENTANE STORINGSFREQUENTIE VOOR AC MET WANDDIKTES VAN 10 MM (BLAUW), 12 MM (GROEN) EN 14 MM (ROOD). GRAFIEKEN ZIJN INDICATIEF.

TABEL 4.1 INVOERPARAMETERS COMSIMA VOOR REKENRESULTATEN AC.

Parameter	Waarde	Onzekerheid
Zetting	5 mm/jaar	±5 mm/jaar
Uitloging	0,15 mm/jaar	±0,05 mm/jaar
Waterdruk	0,3 MPa	±0,1 MPa
Diepte leiding	1 m	±0,1 m
Grondwaterstand	1,5 mmv	±0,25 m

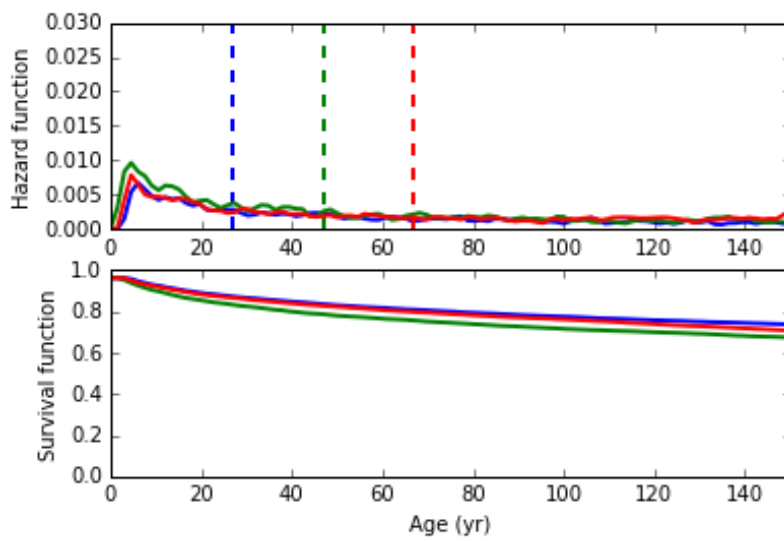


FIGUUR 4.5 STORINGSFREQUENTIES VOOR AC (ALLE DIAMETERS) BEREKEND MET USTORE REGISTRATIES (MAART 2017).

Voorbeeld: PVC

Figuur 4.6 laat voor drie nominale diameters de momentane storingsfrequentie als functie van leeftijd zien zoals berekend door Comsima met de invoerparameters in Tabel 4.2. Deze figuren zijn puur indicatief, de invoerparameters zijn niet afgestemd op een bestaand cohort. Te zien is dat er weinig onderscheid bestaat tussen de verschillende diameters. Ook is te zien dat er vrijwel geen trend in de storingsfrequenties als functie van leeftijd te zien is, de frequenties blijven vrijwel constant.

Figuur 4.7 laat de storingsfrequenties als functie van leeftijd zien zoals berekend met USTORE registraties. Ook hier is niet direct een trend zichtbaar in de periode waarin de statistische betrouwbaarheid voldoende is. Wel is er sprake van een korte periode van oplopende storingsfrequenties te zien rond een leeftijd van 40-45 jaar, overeenkomend met leidingen uit het begin van de jaren zeventig.

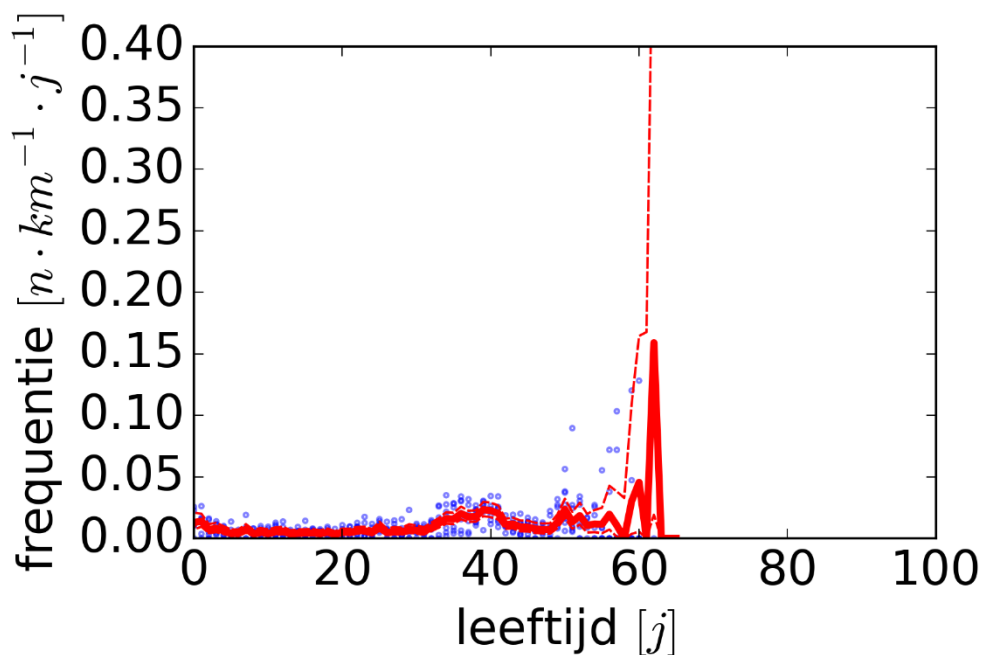


FIGUUR 4.6 REKENRESULTATEN COMSIMA VOOR DE BEREKENDE MOMENTANE STORINGSFREQUENTIE VOOR PVC. GRAFIEK IS INDICATIEF.

TABEL 4.2 INVOERPARAMETERS COMSIMA VOOR REKENRESULTATEN PVC.

Parameter	Waarde	Onzekerheid
Zetting	5 mm/jaar	± 5 mm/jaar*
Waterdruk	0,3 MPa	$\pm 0,1$ MPa
Diepte leiding	1 m	$\pm 0,1$ m
Grondwaterstand	1,5 mmv	$\pm 0,25$ m
WLS	-3,45 MPa/log(s)	$\pm 0,05$ MPa/log(s)

*10 mm/jaar is bijvoorbeeld normaal voor het gebied rond Gouda (*comm. R. Geers*)



FIGUUR 4.7 MOMENTANE STORINGSFREQUENTIE VOOR PVC BEREKEND MET USTORE REGISTRATIES (MAART 2017).

4.4 Discussie

- Om de rekenresultaten van Comsima toe te kunnen passen voor het bepalen van een fysische degradatiecurve is het belangrijk is de vertaling gemaakt van storingsfrequenties van een cohort leidingen naar een kansverdeling van het moment van bezwijken van een individuele leiding.
- Comsima berekent spanningen in een leiding op basis van leidingdegradatie en relevante belastingsfactoren, zoals zetting, bodemdruk en verkeer. Dat betekent dat er in de toekomst een onderverdeling in cohorten mogelijk wordt op basis van omgevingskenmerken. Dit sluit aan op de ontwikkelingen rond registratie van geografische informatie in USTORE. Vanaf 2018 gaan drinkwaterbedrijven in USTORE x y-coördinaten aanleveren van leidingen, hiermee kunnen storingen en storingsfrequenties aan specifieke leidingen worden toegewezen. Deze leidingen zijn weer te relateren aan omgevingskenmerken. Hoe snel cohorten op basis van omgevingskenmerken mogelijk zijn, hangt af van de beschikbaarheid van omgevingsdata. Informatie over de ligging van bomen is al beschikbaar tot op hoge resolutie, voor een parameter als zetting is het ook al mogelijk om hoge resolutie data te vinden vanuit satellietdata. De diepteligging van leidingen moet volgens de WION en INPSIRE rond 2020 beschikbaar zijn. Oftewel, het verschilt per omgevingskenmerk, maar de mogelijkheid om cohorten op basis van omgevingskenmerken te toetsen is iets wat op korte termijn interessant wordt.
- Voor de cohorten waar binnen Comsima nog geen degradatieproces is gemodelleerd (PE en staal) kan de fysische degradatiecurve alleen worden bepaald op basis van effecten van ligging en omgeving op belasting. Dit is voor de huidige cohorten onvoldoende voor het bepalen van de fysische degradatiecurve. Voor deze cohorten kan dan ook alleen de vergelijking tussen de gangbare statistische modellen worden

uitgevoerd. Echter, alleen voor staal zijn voldoende data beschikbaar voor een statistische analyse.

- De resultaten van een testsimulatie met een eerste schatting van invoerparameters voor de cohorten AC en PVC heeft laten zien dat de gesimuleerde vorm van de fysische degradatiecurve overeenkomsten laten zien met de storingsfrequenties op basis van USTORE-data.
- Voor de toepassing van de methodiek is de benodigde kennis aanwezig in gangbare softwarepakketten (bv. R, Matlab, SPSS).

5 Conclusies en aanbevelingen

5.1 Conclusies

Er is een methodiek opgesteld op basis van een gecombineerde aanpak van gangbare statistische degradatiecurves en een degradatiecurve op basis van kennis van de fysische processen die betrokken zijn bij leidingdegradatie en effecten van belasting door liggingomstandigheden en omgevingskenmerken. Deze “fysische degradatiecurve” wordt verkregen met behulp van Comsima. Toepassing van deze methodiek is afhankelijk van de hoeveelheid storingsregistraties, homogeniteit van de cohorten en de vergelijkbaarheid van de resultaten van Comsima met storingsfrequenties. De conclusie van de haalbaarheidsanalyse is dat er (1) voldoende data zijn voor een aanzienlijk deel van het leidingnet en dat (2) de voorgestelde methodiek voldoende potentie heeft om deze nader uit te werken en toe te gaan passen.

Door de vergelijking van de toetsing van verschillende statistische modellen voor het fitten op de storingsfrequenties en de aanvulling met de kennis over degradatie en belasting vanuit Comsima ontstaat inzicht in de onzekerheden in de toekomstige storingsvoorspellingen. Per cohort wordt het beste model bepaald. Dat maakt de methodiek veel betrouwbaarder dan als er slechts één statistisch model wordt gebruikt voor alle cohorten.

De huidige cohortindeling is getoetst op het bevatten van voldoende gegevens voor statistische analyses en homogeniteit:

- De betrouwbaarheid van de storingsfrequenties is niet constant met leeftijd binnen een cohort. Dit betekent dat een deel van de trend die zichtbaar lijkt te zijn in data onbetrouwbaar is. Dit deel moet ofwel niet worden gebruikt voor trendanalyse of in elk geval gewogen worden met de betrouwbaarheid. In het algemeen neemt de betrouwbaarheid toe als er meer data beschikbaar zijn (en komen).
- De storingsfrequenties in de cohorten AC, PVC, GGJ en staal (zonder verdere onderverdeling naar bijvoorbeeld aanlegjaar, diameter, omgevingskenmerken e.d.) zijn op voldoende data gebaseerd (voor voldoende leidinglengte en over een voldoende lange periode) om een trend door de storingsfrequenties te bepalen.
 - Voor het bepalen van trends in storingsen lijkt een onderverdeling gerechtvaardigd tussen spontane en niet spontane storingsen, alsmede buizen en verbindingen, op basis van kennis van storingsmechanisme en zichtbare trends in de data.
 - Er zijn per materiaal op basis van de huidige analyse van alleen de USTORE data geen redenen of om een verdere onderverdeling in cohorten te rechtvaardigen voor een statistische analyse.

5.2 Aanbevelingen

Algemeen

Op basis van de resultaten in dit onderzoek worden een aantal aanbevelingen voor het operationeel gebruik van storingsgegevens en storingsfrequenties gedaan:

- Gegeven de onzekerheden die in de huidige voorspelmodellen voor toekomstige storingsfrequenties optreden, is het aan te bevelen om een gevoeligheidsanalyse uit

te voeren op de statistische modellen die in de beslissingsondersteunende software worden gebruikt. Hieruit wordt duidelijk hoe robuust uitkomsten zijn, zowel wat betreft saneringsprioritering als aanbevolen totale saneringsvolumes. Bij deze gevoeligheidsanalyse is het aan te bevelen niet alleen een gevoeligheidsanalyse te doen naar de fit met het standaard gebruikte statistische model, maar ook een vergelijking te doen met één of meer andere statistische modellen (bv. lineair, exponentieel en afvlakkend).

- Verder is het aan te bevelen bij het gebruik van voorspellingen van toekomstige storingsfrequenties standaard informatie te geven over de robuustheid van de fit op de data en een onzekerheidsmarge mee te geven aan de resultaten (niet alleen in de storingsfrequenties, maar ook hoe deze doorwerken in andere parameters).
- Het blijft belangrijk om zo nauwkeurig mogelijk alle gevraagde informatie over storingen te registreren in USTORE.
- Naast het registreren van informatie is het belangrijk om bij het optreden van storingen zoveel mogelijk lokale informatie te registreren (en aan de storing te linken) die nu nog niet in USTORE past, zoals exit-beoordelingen van conditie en specifiek opvallende omstandigheden. Dit is niet alleen belangrijk voor de implementatie van de in dit rapport ontwikkelde methodiek, maar ook voor mogelijke toekomstige herzieningen van de clusterindeling en voor eventuele aanbevelingen voor aanvullende gegevens in USTORE.
- Het is aan te bevelen om bij statistische analyses de cohorten niet verder op te splitsen dan strikt noodzakelijk, dat wil zeggen dat cohorten zoveel mogelijk data moeten bevatten bij een vergelijkbare degradatie in leeftijd. Cohorten hoeven dus niet een constante storingsfrequentie bevatten, de storingsfrequentie moet alleen een voorspelbaar verloop kennen. Met de toevoeging van meer data (zowel binnen als buiten USTORE) kunnen er aanleidingen bestaan om nieuwe cohorten te definiëren. Aanbevolen wordt om de aanpak van het genereren van cohorten objectief te onderbouwen met een statistische clusteranalyse. Een dergelijke analyse geeft specifiek inzicht in:
 - Het aantal significant verschillende groepen ('clusters') in de dataset;
 - Een goede begrenzing van de verschillende groepen.

Implementatie methodiek

De volgende stap richting het eindplaatje van een rekentool voor het voorspellen van toekomstige storingsfrequenties is een voorbeeldimplementatie van de methodiek in een casusstudie. Dit houdt in:

- Nadere uitwerking van de benodigde stappen om de methodiek te implementeren.
- Inzicht in de mogelijkheden van USTORE data aangevuld met alle beschikbare informatiebronnen bij een drinkwaterbedrijf, zoals exit-beoordelingen en GIS-informatie. Hier hoort ook een kritische blik op de dataset in USTORE bij en gevolgen van beperkingen in de dataset (bv. kennis van samenstelling van de cohorten in het verleden, incomplete set storingsregistraties vanaf de aanleg van de leidingen etc.).
- Inzicht in de onzekerheden in gangbare statistische modellen en in de uitkomsten van Comsima door het analyseren van de spreiding in de invoerparameters.
- Inzicht in de benodigde stappen voor implementatie in de bedrijfspraktijk.

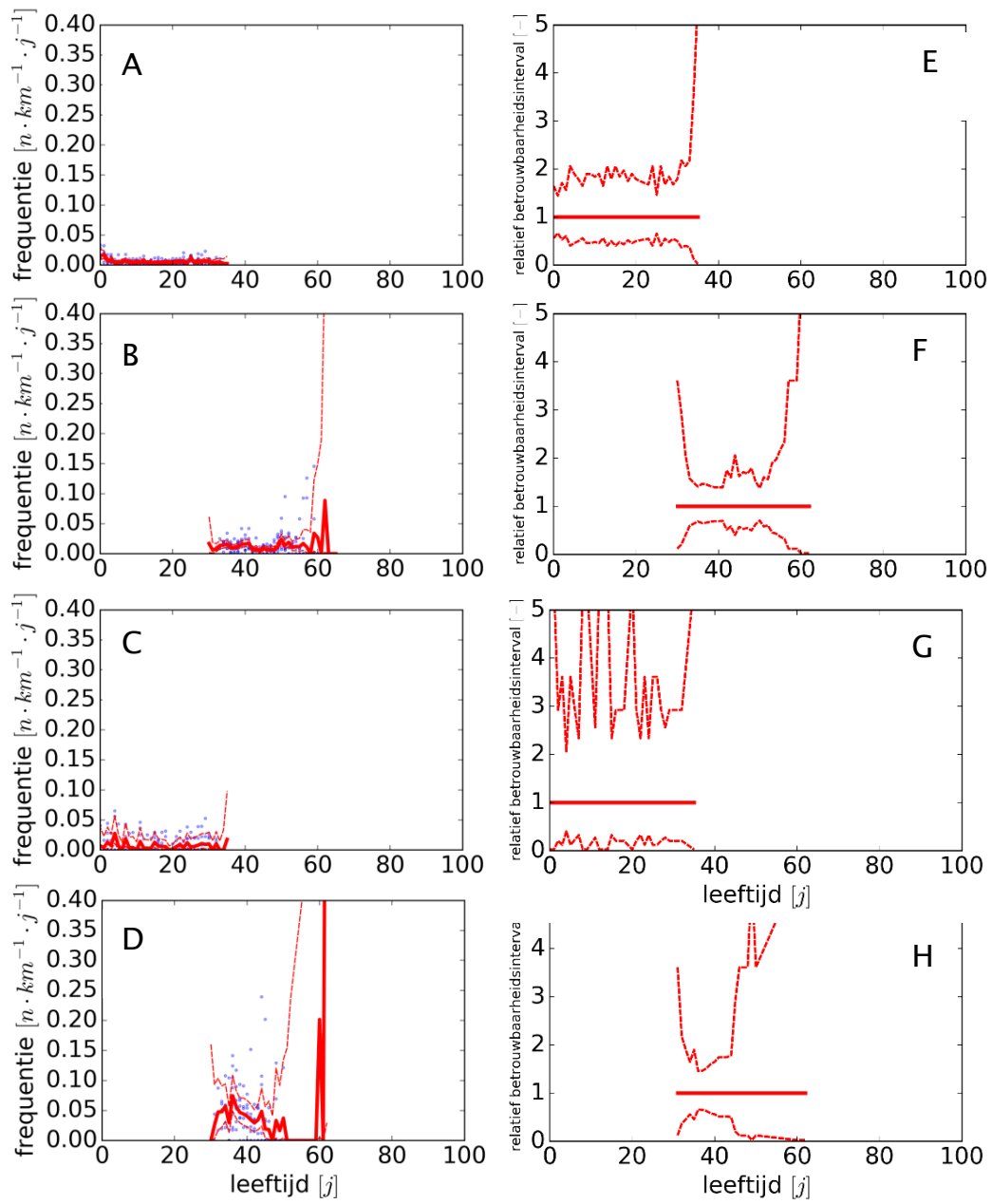
Referenties

- Banerjee, A., and Dave, R. N., 2004, Validating clusters using the Hopkins statistic, in: *Fuzzy systems, 2004. Proceedings. 2004 IEEE international conference on*, IEEE, pp. 149-153.
- Beuken, R., van Vossen, J., Trietsch, E., Jacobs, B., Geers, R., Horst, P., Louter, J., and Diemel, R., 2016, Comparing results of four decision support software tools on mains replacement, in: *Computation and Control for the Water Industry*, Amsterdam.
- Beuken, R. H. S., and Mesman, G. A. M., 2011a, Kennissysteem Levensduurbepaling Versie 2.0, KWR, Nieuwegein, BTO 2011.113(s), pp. 40.
- Beuken, R. H. S., and Mesman, G. A. M., 2011b, Technische levensduur voor groepen leidingen; naar een onderbouwing van het investeringsbeleid, KWR, Nieuwegein, BTO 2011.038, pp. 57.
- Beuken, R. H. S., and Mesman, G. A. M., 2015, Actualisatie driehoeksverdelingen voor onderbouwing investeringsprognoses leidingen, KWR, Nieuwegein, BTO 2015.223(s), pp. 38.
- Beuken, R. H. S., and Moerman, A., 2017, Uniforme storingsregistratie (USTORE): Praktijkcode voor het beheer van storingsregistratie leidingen (concept), KWR Watercycle Research Institute, Nieuwegein.
- Beuken, R. H. S., and Vossen, J. v., 2017, Kwantitatieve vergelijking van beslissingsondersteunende software voor leidingсанering, KWR Watercycle Research Institute, BTO 2017.066, Nieuwegein.
- Fraley, C., and Raftery, A. E., 1998, How many clusters? Which clustering method? Answers via model-based cluster analysis, *The computer journal* 41(8):578-588.
- I.N. Vloerbergh, and Blokker, E. J. M., 2007, Statistische storingsanalyse; mogelijkheden en beperkingen van de huidige storingsregistratie, Kiwa Water Research, BTO 2007.043, Nieuwegein.
- Jain, A. K., Murty, M. N., and Flynn, P. J., 1999, Data clustering: a review, *ACM computing surveys (CSUR)* 31(3):264-323.
- Kleiner, Y., and Rajani, B., 2001, Comprehensive review of structural deterioration of water mains: statistical models, *Urban Water* 3:131-150.
- Liu, Z., Y. Kleiner, B. Rajani, L. Wang, and Condit, W., 2012, Condition assessment technologies for water transmission and distribution systems, US EPA EPA/600/R-12/017, Cincinnati, USA.
- Mesman, G. A., 2015, Kennisregels PVC leidingen, KWR, Nieuwegein, BTO 2015.054, pp. 37.
- Mesman, G. A., Beuken, R., and Meerkerk, J., 2016, PCD 6 Conditiebepaling voor drinkwaterleidingen (2016), KWR, Nieuwegein, PCD 6 (2016), pp. 91.
- Mesman, G. A. M., and Slaats, P. G. G., 2003, Conditiebepaling gietijzeren waterleidingen - wanddikte, belastingen -, Kiwa Water Research, BTO 2003.038, Nieuwegein.
- Moerman, A., van Vossen, J., and Beuken, R. H. S., 2016a, UKNOW, KWR, Nieuwegein, BTO 2016.031, pp. 43.
- Moerman, A., Wols, B. A., and Diemel, R., 2016b, The effects of traffic loads on drinking water main failure frequencies in the Netherlands, *Water Practice and Technology* 11(3):524-530.
- Scheidegger, A., Leitao, J. P., and Scholten, L., 2015, Statistical failure models for water distribution pipes- a review from a unified perspective, *Water Research* 83:237-247.
- Slaats, P. G. G., and Mesman, G. A. M., 2003, Conditiebepaling asbestcement waterleidingen - wanddikte, belastingen -, Kiwa Water Research, BTO 2003.039, Nieuwegein.
- Thienen, P. v., 2012, Storingenregistratie en -analyse: hoeveel storingen hebben we eigenlijk nodig?, KWR Watercycle Research, Nieuwegein.
- van Eijk, R., and van Daal, K., 2013, Storingen als gevolg van graafwerkzaamheden, KWR, Nieuwegein, KWR 2013.093, pp. 39.
- Vloerbergh, I. N., and van Thienen, P., 2011, Registratie geeft storingen waarde. Implementatie en evaluatie van USTORE, KWR, Nieuwegein, BTO 2011.058, pp. 50.
- Wols, B. A., Moerman, A., and van Thienen, P., 2015a, Effecten van klimaatverandering op drinkwaterdistributie in kaart gebracht in: *H2O-Online*, pp. 9 juli.

Wols, B. A., Moerman, A., and Vertommen, I., 2015b, Comsima: model voor spanningen op ondergrondse leidingen, KWR, Nieuwegein, BTO 2015.082.

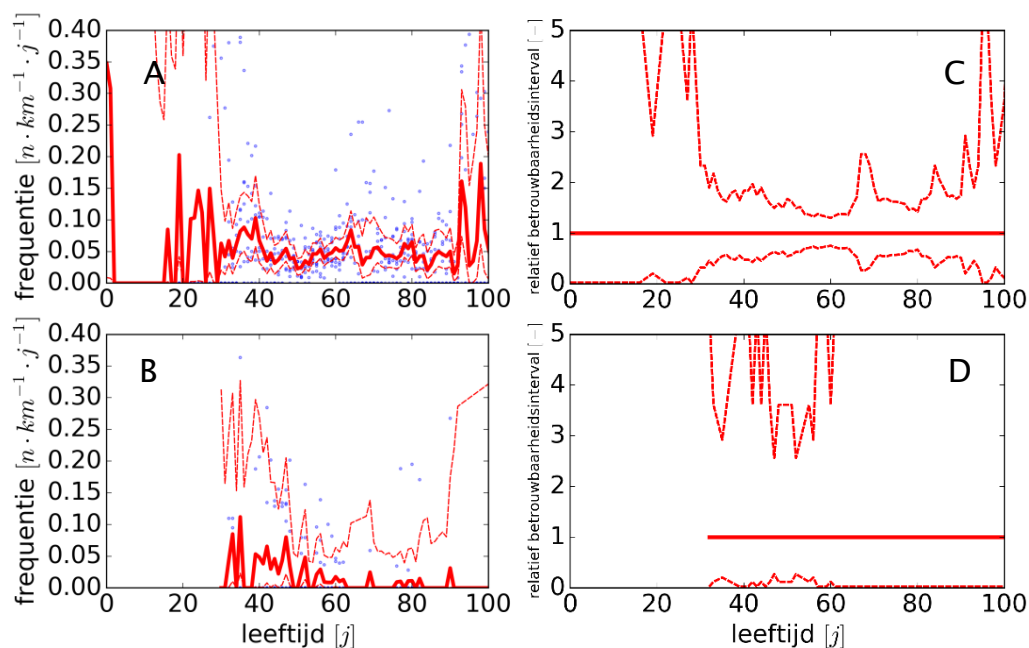
Bijlage I Resultaten toetsing cohorten

In deze bijlage staan de resultaten van de beoordeling van de cohorten van PVC, GGJ, NGJ, staal en PE.



Figuur I.1 laat de statistische betrouwbaarheid zien voor de cohorten met PVC (jonger of ouder dan 1979 en diameters kleiner of groter dan 160 mm). Een aantal zaken valt op:

- De statistische betrouwbaarheid van PVC is in het algemeen minder dan AC, omdat er veel minder storingen zijn. Dit is met name goed te zien aan de jaren bij de opsplitsing van het cohort, waar doordat er minder data zijn voor de leeftijd de onzekerheid snel toeneemt. Dit illustreert het belang van een langere registratieperiode. Juist voor PVC is de verwachting dat de betrouwbaarheid de komende jaren zal verbeteren.
- De statistische betrouwbaarheid van de grotere diameters is veel lager dan die van de kleinere diameters. Het cohort met jongere PVC leidingen is daarom statistisch niet zinvol, het cohort met oudere PVC leidingen slechts voor een zeer beperkte periode.
- De statistische betrouwbaarheid voor kleine diameters is in het algemeen goed, behalve voor leidingen ouder dan 50-55 jaar. Een trendanalyse moet daarom niet op dit deel van de dataset te baseren of moet worden gewogen met de betrouwbaarheid.
- Er is geen duidelijke trend te signaleren in de storingsfrequenties als functie van leeftijd.
- De storingsfrequenties van PVC zijn in het algemeen aanzienlijk lager dan de storingsfrequenties van AC.

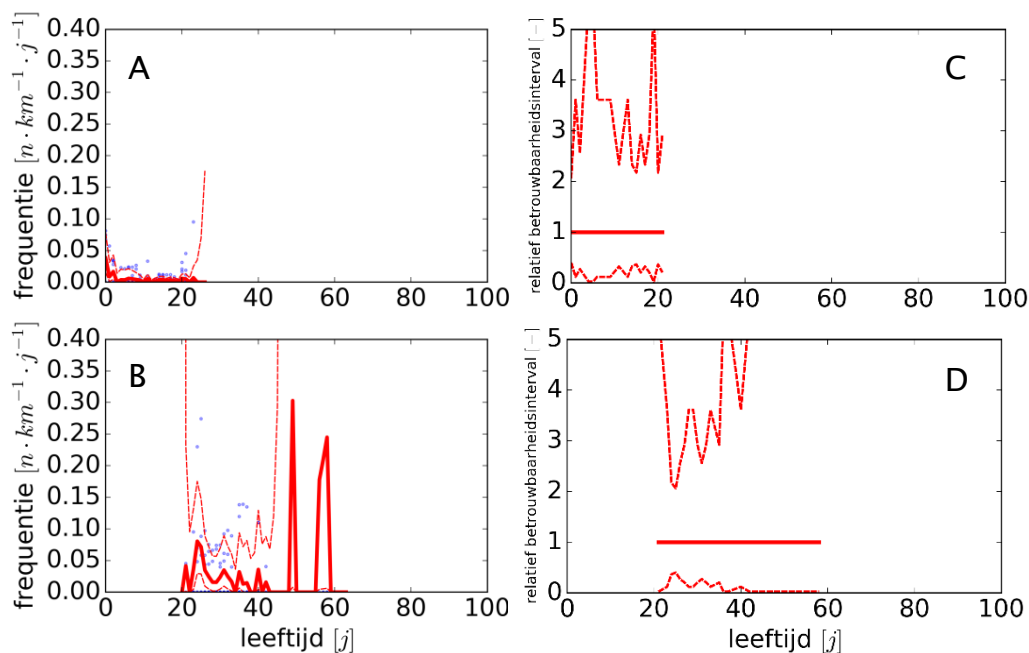


FIGUUR I.2 DE STORINGSDATA IN GGJ LEIDINGEN (·) EN DE GEMIDDELDE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDSINTERVAL (---) VOOR SPONTANE STORINGEN (2009-2015) (A-B). DEZELFDE FIGUREN, MAAR DAN GENORMALISEERD MET DE GEMIDDELDE STORINGSFREQUENTIES (C-D). DIAMETERS VAN 250MM OF KLEINER (A EN C), DIAMETERS VAN GROTER DAN 250MM (B EN D).

Figuur I.2 laat de statistische betrouwbaarheid zien van de cohorten GGJ (diameters kleiner en groter dan 250 mm). Een aantal zaken valt op:

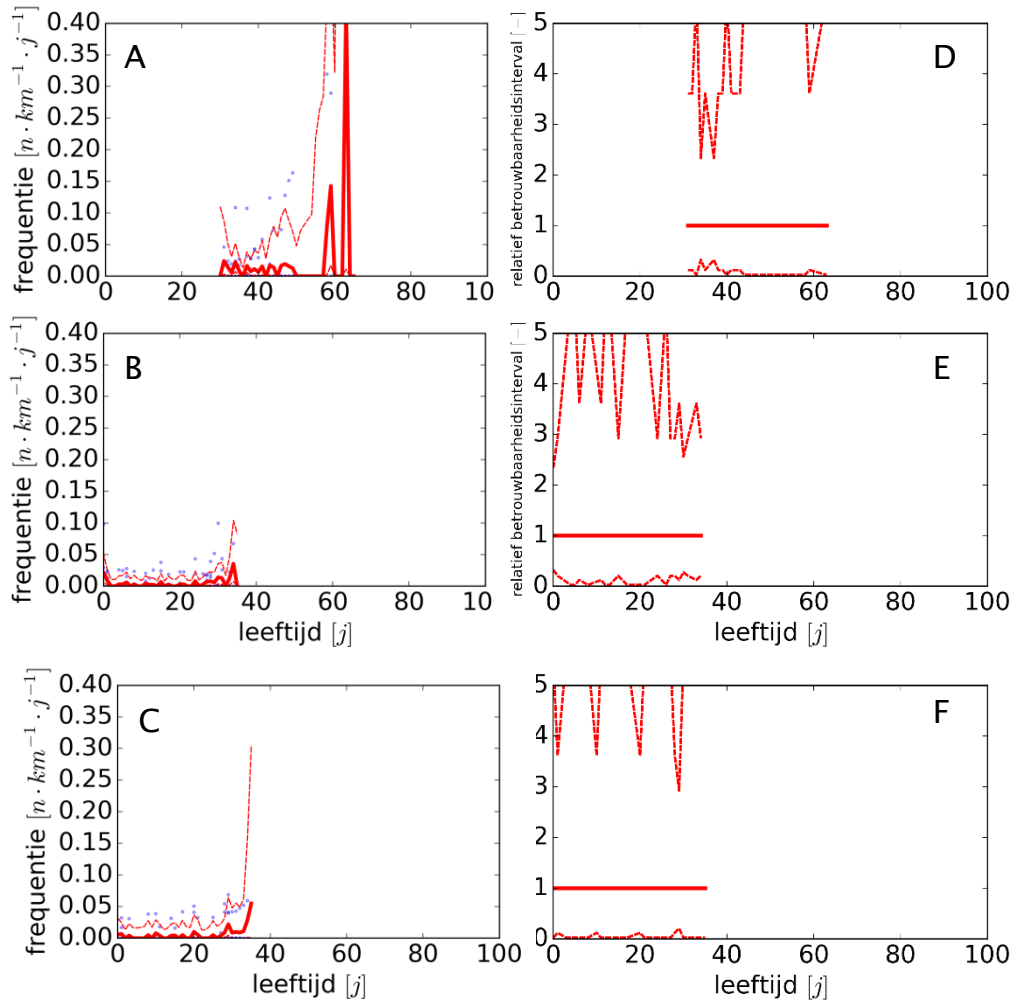
- Net als bij de overige cohorten is de statistische betrouwbaarheid van het cohort met grotere diameters onvoldoende voor alle leeftijden voor statistische analyses.

- Wat betreft de kleinere diameters is de statistische betrouwbaarheid beter, maar nog altijd beperkt voor een aanzienlijk deel van de leeftijden. Het beste deel van het cohort betreft leidingen tussen 40 en 80 jaar oud. Net als bij AC is ook in deze dataset een piekje in onzekerheid te zien rond de periode van de Tweede Wereldoorlog.
- De spreiding in storingsfrequenties is het laagst in dezelfde periode als waar de statistische betrouwbaarheid het hoogst is.
- Het deel van de dataset met voldoende data laat geen duidelijke trend zien.



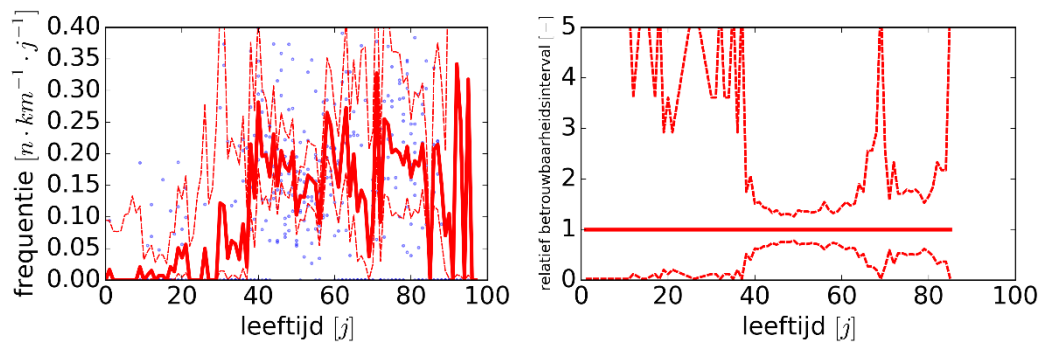
FIGUUR 1.3 DE STORINGSDATA IN NGIJ LEIDINGEN (•) EN DE GEMIDDELDE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDSINTERVAL (---) VOOR SPONTANE STORINGEN (2009-2015) (A-B). DEZELFDE FIGUREN, MAAR DAN GENORMALISEERD MET DE GEMIDDELDE STORINGSFREQUENTIES (C-D). AANGELEGD IN 1990 OF LATER (A EN C), AANGELEGD VOOR 1990 (B EN D).

Figuur 1.3 laat de statistische betrouwbaarheid zijn van de NGIJ cohorten. De statistische betrouwbaarheid van de NGIJ leidingen is onvoldoende voor statistische analyses.



FIGUUR 1.4 DE STORINGSDATA IN PE LEIDINGEN (•) EN DE GEMIDDELDE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDSINTERVAL (---) VOOR SPONTANE STORINGEN (2009-2015) (A-C). DEZELFDE FIGUREN, MAAR DAN GENORMALISEERD MET DE GEMIDDELDE STORINGSFREQUENTIES (D-F). ALLE DIAMETERS AANGELEGD VOOR 1980 (A EN D), DIAMETERS KLEINER DAN 100MM AANGELEGD IN 1980 OF LATER (B EN E), DIAMETERS GROTER DAN 100MM AANGELEGD IN 1980 OF LATER (C EN F).

Figuur 1.4 laat de statistische betrouwbaarheid zien voor de PE cohorten (jonger of ouder dan 1980, diameters groter of kleiner dan 100 mm). Geen van de cohorten bevat voldoende data voor een betrouwbare analyse. Dit komt zowel door het beperkte aantal kilometers als door de lage storingsfrequenties.



FIGUUR I.5 DE STORINGSDATA IN STALEN LEIDINGEN (•) EN DE GEMIDDELTE STORINGSFREQUENTIE (—) ALS FUNCTIE VAN LEEFTIJD MET HET BIJBEHORENDE 95% BETROUWBAARHEIDSINTERVAL (---) VOOR SPONTANE STORINGEN (2009-2015). RECHTS HETZELFDE FIGUUR, MAAR DAN GENORMALISEERD MET DE GEMIDDELTE STORINGSFREQUENTIES.

Figuur I.5 laat de statistische betrouwbaarheid zien voor het cohort staal. Hier is een aantal zaken opvallend:

- de statistische betrouwbaarheid is voldoende voor de leeftijden tussen 40 en 80 jaar met een uitzondering in de leeftijd rond 70 jaar (Tweede Wereldoorlog). Het aantal kilometers en/of storingen in deze periode is zo klein dat de onzekerheidspiek veel groter is dan bij bijvoorbeeld AC of GGJ.
- Voor de statistisch betrouwbare periode is geen visuele trend in de storingsfrequenties waarneembaar.