

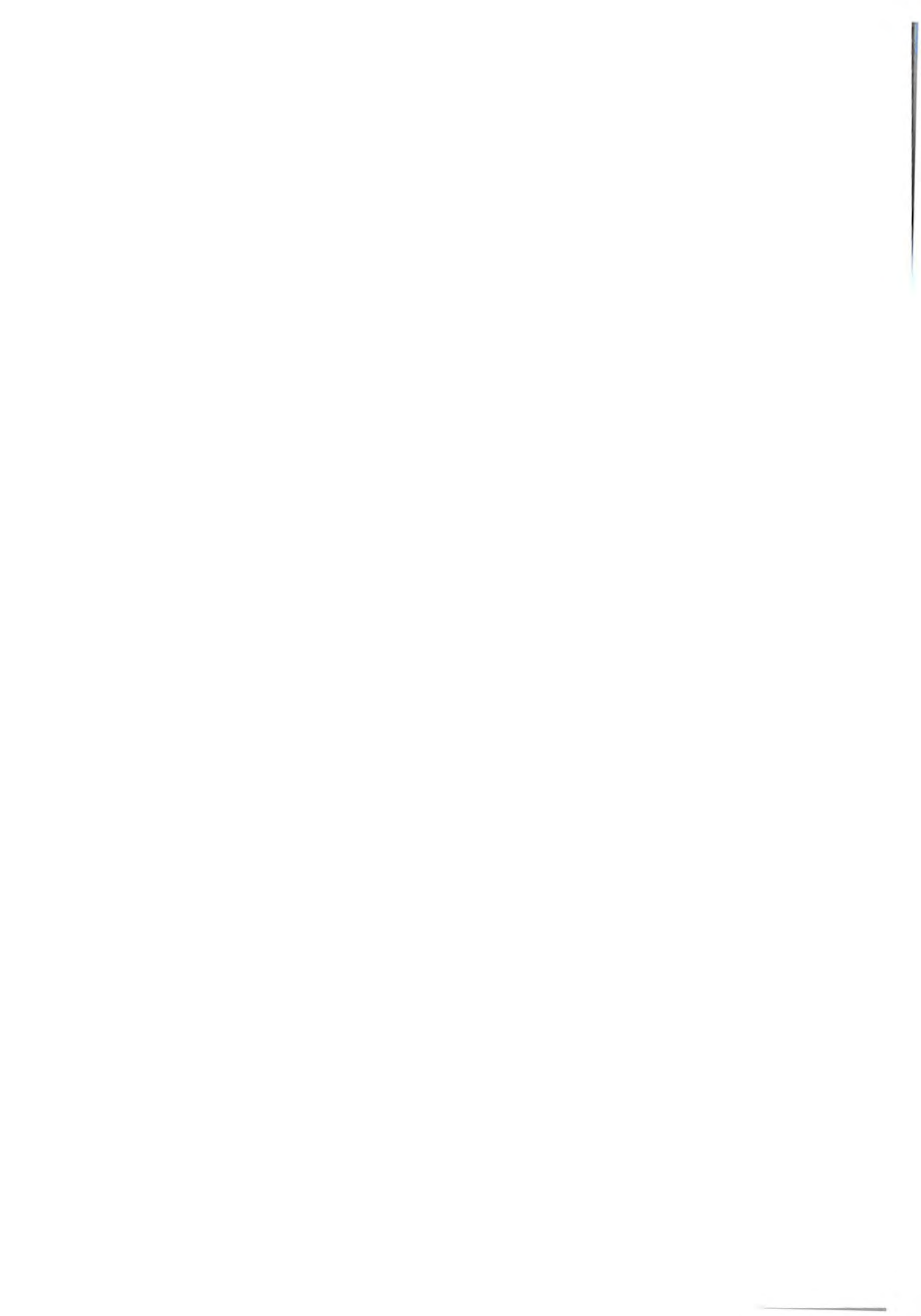
Ontwikkeling van een model voor jaaranalyses ten behoeve van het beleidsinformatiesysteem verkeersveiligheid (BIS-V)

R-92-68

F.D. Bijleveld en drs. S. Oppe

Leidschendam, 1992

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV



INHOUD

1. *Inleiding*
2. *Uitwerking van de basisconceptie van het prognosemodel 'Jaaranalyse nieuwe stijl'*
 - 2.1. Koppeling van de OVG/VOR-gegevens aan de macrotrends
 - 2.2. Globale werkwijze en resultaten
3. *Gegevens*
 - 3.1. Bevolkingsgegevens
 - 3.1.1. Bevolkingsgegevens uit 'SIDO'
 - 3.1.2. Bevolkingsgegevens anderszins van het CBS
 - 3.1.3. Bevolkingsprognoses van het CBS
 - 3.2. Mobiliteitsgegevens
 - 3.2.1. Macrogegevens
 - 3.2.2. Mobiliteitsgegevens uit het OVG
 - 3.3. Slachtoffergegevens
 - 3.3.1. Macrogegevens
 - 3.3.2. Slachtoffergegevens uit VOR-bestanden
 - 3.4. Samenvatting
4. *Modellen*
 - 4.1. Verkeersprognoses
 - 4.2. Model voor gedisaggregeerd model van risico
 - 4.3. Gedisaggregeerd model van risico
 - 4.4. Gedisaggregeerd model voor slachtoffers
5. *Betrouwbaarheid*
 - 5.1. Inleiding
 - 5.2. Implementatie
 - 5.2.1. Risico-ontwikkelingen
 - 5.2.2. Slachtofferontwikkelingen
6. *Conclusies*
7. *Samenvatting en conclusies met betrekking tot modellen voor tijdreeksanalyse*

Literatuur

Figuren

Bijlage A. Niet-lineaire Poissonregressie



SAMENVATTING

Ten behoeve van de SWOV-jaaranalyses en het Beleidsinformatiesysteem Verkeersveiligheid (BIS-V) is, op basis van reeds bij de SWOV gebruikte methoden, een nieuwe methode van tijdreeksanalyse ontwikkeld. Dit rapport legt verslag van de ontwikkeling van deze methode.

In de Inleiding wordt eerst een beschrijving gegeven van het analysemodel in relatie tot andere SWOV-modellen.

In Hoofdstuk 2 wordt een beschrijving gegeven van de algemene doelstelling en daarnaast een overzicht van de beoogde verbeteringen ten opzichte van de reeds bestaande modellen. De theoretische overwegingen die hierbij gelden komen eveneens aan de orde. Het betreft aannamen over de globale trends voor de ontwikkeling van het verkeer en van de veiligheid zoals deze in het SWOV-macromodel zijn uitgewerkt voor gegevens sinds 1950 en overwegingen over het modelleren van de zogenaamde 'economische golf'. Verder een discussie over het te kiezen niveau van disaggregatie, over de gegevens van de Dienst Verkeersongevallenregistratie (VOR) en gegevens uit het CBS-Onderzoek Verplaatsingsgedrag (OVG), gegevens die beschikbaar zijn vanaf 1978 en over de wijze waarop deze gegevens kunnen worden ingepast in de lange-termijntrends. Ook wordt een aantal praktische keuzes besproken die gemaakt zijn bij het samenstellen van het model.

In Hoofdstuk 3 worden de gebruikte gegevens in detail besproken. Het gaat behalve om de VOR- en OVG-bestanden voor diverse groepen van verkeersdeelnemers ook om bevolkingsgegevens en bevolkingsprognoses voor deze groepen.

In Hoofdstuk 4 worden de verschillende modellen voor de diverse trends besproken. Aan de orde komen voorspellingen voor de mobiliteit, al dan niet rekeninghoudend met de ontwikkeling en samenstelling van de bevolking. Vervolgens wordt besproken hoe de mobiliteitsgegevens die zijn opgesplitst naar wijze van vervoer en leeftijdklasse gemodelleerd worden om tot prognoses te komen. Voor de prognoses van de mobiliteit voor iedere deelgroep wordt naast de OVG-gegevens ook gebruik gemaakt van bevolkingsprognoses van het CBS voor bepaalde leeftijdsklassen. Vervolgens wordt beschreven hoe de risico- en slachtofferprognoses worden verkregen voor specifieke leeftijd- en vervoerclassen.

Hoofdstuk 5 gaat in op de betrouwbaarheid van de beschreven modellen. Er wordt een model voor de onbetrouwbaarheid voorgesteld op basis van eenvoudige statistische assumpties. Een meer gedetailleerde uitwerking daarvan is in een bijlage toegevoegd.

Hoofdstuk 6 en 7 geven de conclusies weer voor het model, de methode en het toepassingsbereik. Tevens wordt aangegeven hoe het model getoetst en verbeterd kan worden.



1 INLEIDING

De SWOV heeft een aantal projecten uitgevoerd en in uitvoering genomen die op de één of andere wijze gegevens beschrijven of verklaren die bestaan uit reeksen van waarnemingen in de tijd. Soms worden daaraan ook prognoses verbonden.

Bij de SWOV-kwartaalanalyses gaat het om een eerste signalering van afwijkingen die kwartaalcijfers in een nieuw kwartaal hebben ten opzichte van verwachte uitkomsten. Het betreft de ontwikkeling van de aantallen verkeersslachtoffers, uitgesplitst naar een aantal hoofdkenmerken ten opzichte van verwachtingen die zijn gebaseerd op dezelfde cijfers in de afgelopen vijf jaar.

Uitgegaan wordt van een lineaire trend over de afgelopen periode (die kan stijgen, dalen of gelijkblijven) en een kwartaaleffect. Op basis hiervan wordt een verwachting gegeven voor de nieuwe cijfers, met een betrouwbaarheidsgebied daaromheen. Er is dus niet echt sprake van een tijdreeksanalyse en daarop gebaseerde prognoses, maar van een eenvoudig instrument om cijfers van één nieuw kwartaal te vergelijken met oude cijfers en afwijkingen te detecteren die interessant kunnen zijn voor nadere beschouwing.

Om een effect te kunnen aantonen van de in de jaren zeventig ingevoerde verplichting tot het dragen van autogordels is een echte tijdreeksanalyse uitgevoerd op basis van lange reeksen maandgegevens, zie Bos & Bijleveld (1991). Hierbij is gebruik gemaakt van het zgn. Harvey-model, een model voor structurele tijdreeksanalyse, zie Harvey & Durbin (1986). Het model is zeer geschikt om interventies op te sporen en de ermee gepaard gaande niveauveranderingen in een reeks statistisch te toetsen. Het is een model om in de eerste plaats wetmatigheden in de reeks zelf op te sporen en deze voort te zetten over een wat langere periode. Maar ook exogene variabelen kunnen worden opgenomen ter verklaring van ontwikkelingen. Globaal gesproken is het een regressie-analyse-achtige aanpak, waarbij niet echt sprake is van een theoretische basis voor de verklaring van een reeksontwikkeling.

Bij het SWOV-macromodel is wel sprake van een strikt theoretisch model. Aangenomen wordt dat de ontwikkelingen van de mobiliteit en het verkeersrisico met behulp van specifieke (logistische en exponentiële) functies beschreven kunnen worden. Het model beschrijft de ontwikkeling van het totale aantal slachtoffers over een langere periode en maakt ook prognoses over een langere tijd. Het model beperkt zich echter tot de macro-trend.

Ook het SWOV-project 'Mobiliteit en veiligheid' (M&V) gaat het over het beschrijven van (jaar)reeksen van mobiliteits- en veiligheidsgegevens, hier echter uitgesplitst naar een aantal hoofdkenmerken (leeftijd, geslacht, wijze van verkeersdeelname, wegtype, tijdstip). Daarnaast is het de bedoeling om naast een beschrijving en verklaring in termen van reizigerskilometers, ook een beschrijving te geven per conflicttype van verkeersdeelnemers (het aantal slachtoffers bij voetgangers is zowel afhankelijk van het aantal voetgangers als van het aantal rijdende voertuigen). Praktisch gesproken is nog slechts het aantal reizigerskilometers gebruikt en heeft er alleen een uitsplitsing naar wijze van verkeersdeelname, leeftijd en geslacht plaatsgevonden. Basis voor deze beschrijvingen waren de gegevens uit het onderzoek Verplaatsingsgedrag (OVG) van 1985 t/m 1990 en de bijbehorende

gegevens uit de VOR-ongevallenbestanden. Deling van de slachtofferaantallen door de bijbehorende reizigerskilometers geeft de mate van risico aan. Evenals in het macromodel werden deze risico's geacht negatief-exponentieel te dalen, maar nu voor de cellen van de uitgesplitste tabel elk met een eigen parameter. Op basis hiervan zijn verwachte risico's berekend en met de geobserveerde waarden vergeleken. Deze gegevens zijn, samen met verwachtingen over de ontwikkelingen in de mobiliteit, gecombineerd tot verwachtingen over de aantallen slachtoffers. Gezien de zeer korte reeks beschikbare jaren is de periode waarvoor vooruit voorspeld kan worden natuurlijk zeer beperkt.

Tenslotte zijn tot en met 1990 door de SWOV ook jaaranalyses uitgevoerd. Hierbij ging het er vooral om verklaringen te vinden voor geconstateerde ontwikkelingen of afwijkingen ervan. Dit gebeurde meer kwalitatief. Er was geen strikte procedure volgens welke reeksen van jaarcijfers werden geanalyseerd om (statistisch getoetste) afwijkingen vast te stellen. Om hierin verandering te brengen is een nieuw project gestart met als doel om alsnog te proberen een procedure te ontwikkelen voor het systematisch beoordelen van ontwikkelingen in reeksen van gedisaggregeerde onveiligheidsgegevens. De uitsplitsingen overlaptten deels die welke in het project M&V waren gebruikt, en waren deels gebaseerd op ontwikkelingen die interessant zijn vanuit het beleid gezien (bijvoorbeeld de speerpunten uit het MPV).

Bij de overheid is er, vooral sinds de kwantitatieve taakstelling binnen het MPV, eveneens een toenemende belangstelling te constateren voor kwantitatieve beschrijvingen van ontwikkelingen op het gebied van de verkeersveiligheid en daaraan gekoppelde prognoses. Deze prognoses hebben vooral tot doel beleidsmaatregelen te kunnen toetsen en verder om nieuwe speerpunten te kunnen formuleren. De behoeften van de overheid hebben zich toegespitst op de realisering van een Beleidsinformatiesysteem Verkeersveiligheid (BIS-V), waarin deze gegevens voor de beleidsmedewerkers toegankelijk gemaakt dienen te worden. Twee tijdstippen zijn daarbij voor het beleid vooral van belang: tussentijdse toetsing van beleidsmaatregelen zal plaatsvinden in 1996, terwijl het voorlopige einddoel van het geformuleerde beleid is bepaald op het jaar 2010. Het lange tijdsperspectief tot 2010 maakt het noodzakelijk om prognoses te geven op een lange termijn. Het macromodel is hiertoe wel in staat, maar slechts voor het totaal. Dit is echter onvoldoende. De OVG/VOR-bestanden geven voldoende detail, maar zijn te beperkt in tijdsperspectief om een lange-termijnprognose op te baseren. In het project 'Jaaranalyse nieuwe stijl' is de oplossing gezocht in het koppelen van de korte-termijnbeschrijvingen en prognoses gebaseerd op de OVG/VOR-gegevens aan de lange-termijnbeschrijving en prognose van het macromodel. Vervolgens is besloten de activiteiten in het project 'Jaaranalyse nieuwe stijl' en de ontwikkeling van een BIS-V te koppelen. De twee projecten werden geïntegreerd tot één project, vooral gericht op de beschrijving en prognose van de belangrijkste autonome ontwikkelingen en die verkeersveiligheidsontwikkelingen, die beleidsrelevant zijn. Daarnaast zal in het project M&V de oorspronkelijke doelstelling verder worden uitgewerkt. De uiteindelijke output van dit project kan dan worden gezien als basis voor onderzoek naar de ontwikkelingen die meer algemeen (voor diverse doelen) van belang zijn. Hieraan kunnen ontwikkelingen voor meer specifieke beleidsgerichte uitsplitsingen gekoppeld worden. Bij beide projecten zal er naar worden gestreefd om zoveel mogelijk uit te gaan van eenzelfde basisconceptie bij de modelontwikkeling.

Hieraanvolgend wordt beschreven hoe binnen 'het oorspronkelijke project 'Jaaranalyse nieuwe stijl' en later dus in BIS- V een eerste aanzet is gemaakt om tot een dergelijke basisconceptie te komen. Daarnaast, welke problemen zich bij de toepassing hebben voorgedaan en tenslotte welke procedure voor de volgende fase van modelontwikkeling (met diverse keuze-alternatieven) wordt voorgesteld. Argumenten voor een definitieve keuze voor die basisconceptie dienen zowel vanuit het project M&V als vanuit BIS-V aangedragen te worden. Ook dienen de grenzen tussen beide projecten te worden vastgesteld.

Een belangrijke tekortkoming op dit moment is het ontbreken van goede cijfers voor de beschrijving van ontwikkelingen naar wegtype. In het SWOV-project 'Kencijfers' binnen het SWOV-Meerjaren Onderzoekplan Project 5 (MOP-5) zijn voor enkele tijdsmomenten gedetailleerde cijfers aanwezig. Voor het aangeven van ontwikkelingen in de tijd zijn deze echter te beperkt. Bekeken zal moeten worden hoe de gegevens uit het project 'Kencijfers' in het model kunnen worden ingebracht en op welke wijze (met behulp van welke eenvoudige aannamen) prognoses kunnen worden gegeven voor ontwikkelingen per wegtype. Dergelijke prognoses zijn met name van belang voor onderzoek en beleid gekoppeld aan de ontwikkeling van verkeers- en vervoerregio's.

2 UITWERKING BASISCONCEPTIE PROGNOSEMODEL 'JAARANALYSE NIEUWE STIJL'

Hiervoor zijn als uitgangspunt genomen de jaarreeksen zoals die bij eerder uitgevoerde jaaranalyses naar voren waren gekomen en die eveneens voor andere doelen (met name BIS-V) nodig zijn. Doel was om de gedisaggregeerde reeksen verkeersslachtoffers (overledenen en ziekenhuisgewonden) te beschrijven, uitgaande van de VOR- en OVG-gegevens over een zo lang mogelijke periode (liefst tien jaar). De nadruk moet daarbij liggen op de beschrijving van de trends naar de mobiliteitscomponent en de risicocomponent. Een afgeleid doel is de ontwikkelingen in deze reeksen op een zodanige manier te koppelen aan de macrotrends dat prognoses:

- stabiel(er) zouden worden;
- over een wat langere periode konden plaatsvinden;
- zouden optellen tot de macrotrend.

Bij de uitwerking is geprobeerd om met zo min mogelijk parameters toch modellen te genereren die goed beschrijven. Dat betekent bijvoorbeeld dat disaggregatie naar twee of meer kenmerken zoveel mogelijk gezien wordt als een onafhankelijke toepassing van trendontwikkelingen van elk kenmerk afzonderlijk (bijvoorbeeld: de verdeling van voertuigkilometers naar leeftijd is gelijk voor mannen en vrouwen). Voor een aantal kenmerken zal dit niet opgaan. Bijvoorbeeld zeker niet voor de combinatie leeftijd en wijze van verkeersdeelname. Juist vanwege deze complicaties is begonnen met die combinatie.

2.1 Koppeling van de OVG/VOR-gegevens aan de macrotrends

Hierbij ging het er vooral om ervoor te zorgen dat de som van alle deelprognoses samen niet een irrealistisch beeld zouden geven van het totaal. Daarnaast is ervoor gekozen het totaal als uitgangspunt te nemen, dus als norm te hanteren voor de delen. De bijdrage van elke deeltrend zou dan per jaar evenredig moeten worden verlaagd of verhoogd met een bepaalde factor om op de macrotrend uit te komen. [N.B. meer stabiele trends zouden eventueel minder moeten worden aangepast dan instabiele trends, maar daarmee is nog geen rekening gehouden]. Aanvankelijk was het de bedoeling deze procedure zowel op de macrotrend voor het verkeer als op die voor de risico's toe te passen. Later is daarvan om allerlei redenen voorlopig afgezien en is alleen genormeerd op de verkeersprestaties. De totale risicotrend is dus op dit moment nog geen norm voor het cumulatieve effect van de deelrisico's. Bij de modellering van de macrotrend voor de voertuigkilometers diende rekening te worden gehouden met zowel met de globale trend (logistische curve) als de systematische afwijking ervan (de 'economische golf'). De verdere uitwerking van het macromodel door Koornstra (1988) maakt dit mogelijk, maar dat programma bevindt zich nog in het stadium van ontwikkeling. Gekozen is voorlopig voor de (symmetrische) logistische curve.

Oorspronkelijk is, met behulp van een tijdreeksanalyse op de afwijkingen van de bestpassende logistische curve, getracht alsnog rekening te houden met het effect van de economische golf. Gekeken werd of de afwijking in een bepaald jaar afhangt van de meting van het er direct aan voorafgaande jaar. Dit voegt aan de drie logistische parameters nog een parameter toe voor de economische golf. Deze beschrijvingen voldeden goed. Voor prognoses werd gesteld dat op de wat langere termijn vooral de logistische parameters de doorslag zouden moeten geven en op de kortere termijn de trendcorrectie een belangrijke aanvulling zou moeten geven. Gedacht werd daarom aan een voorspelling waarin van een exponentiële uitdoving van de trendparameter sprake was. Complicaties bij de realisering op gedisaggregeerd niveau hebben later doen besluiten om hiervan voorlopig af te zien en uitsluitend te normeren op de logistische curve, te meer daar het logistische model zelf al uitstekend voldeed aan de voorwaarden voor normering. In een later stadium moet daar opnieuw naar worden gekeken. Wel zal daarbij blijvend onderscheid gemaakt worden tussen de globale trend, waaraan het verkeersveiligheidsbeleid op de langere termijn gekoppeld is en de golf die aangeeft in welke mate extra beleid op korte termijn gewenst is. Het is namelijk bekend dat afwijkingen van de lange-termijnontwikkeling direct gevolgen hebben voor de veiligheid. Ook wanneer later in het model mobiliteitsontwikkelingen uit externe bron zullen worden gebruikt (zoals zowel in het project M&V als bij BIS-V de bedoeling is), dient dit onderscheid tussen fluctuaties op de kortere en de langere termijn om deze reden gehandhaafd te blijven.

2.2 Globale werkwijze en resultaten

De hieronder beschreven analyse moet vooral gezien worden als een haalbaarheidsstudie: is het mogelijk om voor gedisaggregeerde gegevens een model te ontwikkelen op basis van koppeling van de macrotrend aan uitgesplitste gegevens? Om toch enige zekerheid te hebben over de toepassing bij meer complexe gegevens is evenals in het eerdere project M&V gekozen voor de combinatie van leeftijd en wijze van verkeersdeelname. Voor de analyses kon gebruik worden gemaakt van de OVG/VOR-gegevens vanaf 1979. Deze zijn inmiddels beschikbaar gekomen in een vorm overeenkomstig aan die van de periode 1985 t/m 1990. Hierdoor konden ook meer stabiele analyses worden uitgevoerd op de gedisaggregeerde gegevens. Basis voor de trendontwikkelingen op gedisaggregeerd niveau waren dan ook de $7 \times 5 (\text{min } 2) = 33$ combinaties van leeftijdklassen en wijzen van verkeersdeelname. De bedoeling was om op basis van de OVG- en VOR-cijfers voor elke combinatie een trend te beschrijven; voor de voertuigkilometers een logistische trend en voor de risico's een exponentiële trend. Elke klasse zou dus unieke logistische parameters krijgen, maar zonder tijdreeksparameter voor de voertuigkilometers en verder twee exponentiële parameters voor het risico. Dit bleek al spoedig onrealistisch. Vervolgens werd ingedikt tot $3 \times 3 = 9$ combinaties. Hierbij bleken er problemen te ontstaan bij de beschrijving van de verkeersprestatie van de jongste leeftijdsgroep als logistische trend, terwijl dit voor de oudere leeftijdsgroepen dit niet gold. Na onderzoek bleek dat de ontwikkelingen voor alle leeftijdklassen (ook bij de vijf- in plaats van driedeling) een logistische trend over de tijd in combinatie met de bevolkingsomvang van de groep voor elk jaar redelijk voldeed. Op basis van deze uitkomsten is vervolgens weer voor de 33 oorspronkelijke combinaties een beschrijvingsmodel ontwikkeld voor de verkeersprestatie. Op basis van de prognoses van

de bevolkingsomvang naar leeftijd zijn vervolgens prognoses geleverd voor de ontwikkeling van de verkeersprestaties voor elke groep. Koppeling van de gedisaggregeerde trends aan de macrotrend zoals boven omschreven, leverde uiteindelijke prognoses en beschrijvingen. De overeenkomst tussen de geobserveerde aantallen en de voorspelde aantallen bleek voor de meeste deelcategorieën uiteindelijk bevredigend te zijn.

Het expliciet maken van de bevolkingsomvang bij het leveren van mobiliteitsprognoses zou ook al op het niveau van de macrotrend toegepast kunnen worden. Een alternatief model hiervoor is elders beschreven. Een probleem bij het beschrijven van de macrotrend is verder dat de totale verkeersprestatie meer is dan die van het gemotoriseerde verkeer in het macromodel. Fietsers-, bromfietzers- en voetgangerskilometers behoren hier ook toe. Voor deze wijzen van verkeersdeelname is voorlopig uitgegaan van een over de tijd constante hoeveelheid verkeer. Het niveau is bepaald op grond van de OVG-gegevens. Ook hierbij zou op zijn minst rekening gehouden dienen te worden met de bevolkingsomvang. Deze aanname levert een gesimplificeerd beeld op van de beschrijving van de gerealiseerde kilometers. Een voordeel van deze eenvoudige aanname is dat veranderingen ten opzichte van een voorgaande periode duidelijk zichtbaar worden, hetgeen gebruikt kan worden voor de bepaling van aandachtsgebieden voor het beleid. Zoals eerder gezegd, wordt het lange-termijnbeleid niet bepaald door incidentele schommelingen in de verkeersdeelname. Als basiscurve (naast de logistische voor het gemotoriseerde verkeer) dienen ook hier 'gladde' trends gebruikt te worden voor de langere termijn. Tijdreeksanalyse kan dan aanvullend een beschrijving geven voor de kortere termijn. Er dient overigens een duidelijk onderscheid gemaakt te worden tussen beschrijvingen en prognoses op de langere en op de kortere termijn, omdat beide een andere functie vervullen en ander beleid vragen.

De toepassing van de exponentiële curve op de gedisaggregeerde risico's leverde geen problemen op. In een paar gevallen (bij groepen met geringe omvang) leek toename van risico plaats te vinden. Daarna is de eis aan het model gesteld dat de parameter niet positief mocht worden, maar hooguit gelijk aan nul. Een dergelijke beslissing was ook al eerder in het project M&V genomen. Dit leidde uiteindelijk tot beschrijvingen van de aantallen slachtoffers per categorie en een prognose hiervan.

De eerste resultaten voor deze slachtofferprognoses zijn bevredigend. Voor een aantal deel categorieën is er echter een duidelijke afwijking te zien tussen de verwachte verkeersprestatie en de gegevens uit het OVG. Nadere analyse van significante afwijkingen moet nog plaats vinden. Toetssteen is nu vooral: zijn de gedisaggregeerde voorspellingen voor de komende jaren realistisch? Op basis van de eerste ervaringen en hierboven genoemde overwegingen zijn aanpassingen van het model gewenst. Bij de conclusies wordt hierop nader ingegaan.

3 GEGEVENS

3.1 Bevolkingsgegevens

3.1.1 Bevolkingsgegevens uit 'SIDO'

Voor ieder jaar en iedere leeftijdklasse wordt een bevolkingsaantal bepaald. Dit aantal wordt *niet* als aan toeval onderhevig verondersteld, zodat er geen standaardfout aan gekoppeld hoeft te worden.

3.1.2 Bevolkingsgegevens anderszins van het CBS

Voor de bevolkingsgegevens zijn ook CBS-gegevens beschikbaar. Deze betreffen de jaren sinds begin deze eeuw, met enige onderbrekingen. Voor zover dat betekenis heeft, komen de gegevens uit oudere CBS-bronnen (CBS, 1989). De leeftijdsindeling welke in deze bron is gebruikt wijkt af van de voor BIS-V gewenste indeling (0-14, 15-24, 25-50, 50-64, 65+) en is 0-19, 20-44, 45-64, 65-79, 80+. Uit deze bron zijn daardoor slechts de totalen per jaar bruikbaar. Eventueel kan later een opsplitsing naar geslacht gebruikt worden. Voor de jaren 1979 tot aan het begin van de gegevens uit SIDO zijn de naar leeftijd opgesplitste gegevens met de hand uit de statistische jaarboekjes overgenomen.

3.1.3 Bevolkingsprognoses van het CBS

Recentelijk zijn op elektronische basis gepresenteerde gedetailleerde bevolkingsprognoses beschikbaar gekomen (CBS, 1991). Voor ieder jaar is een lage, midden en hoge prognose van het bevolkingsaantal per leeftijd beschikbaar. Het is niet precies duidelijk hoe de verschillen tussen deze varianten tot stand zijn gekomen, maar het lijkt redelijk aan te nemen dat met een zekere waarschijnlijkheid het bevolkingsaantal in een bepaald jaar tussen de lage en de hoge variant zal liggen, met als verwacht aantal de middenvariant. Voorgesteld wordt de middenvariant te kiezen en uit het verschil hiertussen en de hoge en lage varianten een standaardfout te berekenen.

In de 'Bevolkingsprognose voor Nederland 1988-2010' (CBS, 1991) wordt vermeld dat de kans dat het bevolkingsaantal in het jaar 2000 binnen de hoge en lage variant ligt ongeveer 2 tegen 1 is. Een 66%-betrouwbaarheidsinterval dus. Er wordt echter een stok achter de deur gehouden en dat zullen wij ook moeten doen.

Uit bovengenoemde bronnen wordt een reeks van de totale bevolkingsomvang verkregen plus voor de te prognostiseren periode een standaardfout.

Het bovenstaande zal resulteren in een reeks cijfers:

$$b_t = \text{populatie}_t \quad sb_t = \text{stderr}(b_t) \quad (3.1)$$

3.2 Mobiliteitsgegevens

3.2.1 Macrogegevens

Uit verschillende CBS-bronnen is in het verleden een schatting gemaakt van de verkeersprestaties (meestal voertuigkilometers) van verschillende typen vervoer. Daarbij zijn opsplitsingen gemaakt naar zowel tijd van de dag, wegtype als provincie. Niet alle gegevens zijn voor ieder jaar aanwezig en de meeste zijn niet beschikbaar voor de jaren voor 1974.

3.2.2 Mobiliteitsgegevens uit het OVG

Vanaf 1979 zijn gedetailleerde mobiliteitsgegevens beschikbaar uit het CBS-Onderzoek Verplaatsingsgedrag (OVG). Deze gegevens zijn in tegenstelling tot de verkeersprestaties in par 4.1 personenkilometers, opgehoogd tot populatiegegevens¹. Per combinatie van leeftijd en wijze van verkeersdeelname en jaar worden deze gegevens gemiddeld over personen. Dit levert voor iedere leeftijd- en vervoerklasse en jaar een gemiddelde op met een daarbij behorende standaardfout. We zullen dit cijfer m_{tij} noemen, de fout sm_{tij} . Hierbij staat t voor het jaar, i voor vervoerwijze en j voor de leeftijdklasse. Er moet opgemerkt worden dat twee categoriën zijn verwijderd, namelijk zeer de jonge en de oudere vrachtwageninzittenden. Verder moet ook nog met zorg gekeken worden naar de vrachtwagengegevens over de leeftijdklassen en jaren. Waarschijnlijk zijn alleen gegevens over 1985 t/m 1988 bruikbaar. Daarnaast moet opgemerkt worden dat de mobiliteitgegevens slechts vanaf de leeftijd van 12 jaar beschikbaar zijn. Een correctie wordt uitgevoerd op basis van CBS (1992), zie onderstaande tabel waarin voor de jaren 1988 t/m 1991 schattingen worden gegeven voor de mobiliteit voor onder-12-jarigen. Deze is als volgt uit te splitsen:

Wijze verkeersdeelname	$\times 10^9$ km
Auto	10.0
Trein	0.2
Bus/tram/metro	0.7
Fiets	0.8
Lopen	0.9

3.3 Slachtoffergegevens

3.3.1 Macrogegevens

Analoog aan par 3.2.1 zijn ook voor slachtoffers reeksen beschikbaar. Niet alle definities zijn hier echter over de jaren gelijk gebleven, zodat bijvoorbeeld het cijfer voor de ziekenhuisgewonden niet geheel consistent is gebleven. Tot op heden is er binnen dit project geen praktische toepassing geweest van deze gegevens. In een later stadium kan wel gebruik worden gemaakt van deze gegevens, bijvoorbeeld om aantallen slachtoffers te normeren.

¹Dit moet nog steeds gecontroleerd worden!

3.3.2 Slachtoffergegevens uit VOR-bestanden

Uit de VOR-bestanden zijn de slachtoffergegevens betrokken. Deze betreffen overleden slachtoffers en ziekenhuisgewonden over de jaren 1979 t/m 1990 en als speciaal geval 1991. Zowel voor de overledenen als voor de ziekenhuisgewonden zijn per jaar per leeftijd en per wijze van verkeersdeelname de aantallen opgeteld. Er is rekening mee gehouden dat sommige cellen niet voorkomen onder de ongevallen, maar wel als mobiliteitscategorie en omgekeerd dat sommige categorieën ongevallen voorkomen, maar niet als zodanig in de analyses worden gebruikt. Slachtoffers onder zeer jonge of zeer oude vrachtauto-inzittenden zijn naar de betreffende leeftijdsklassen in de rest van het verkeer ondergebracht. Per jaar, wijze verkeersdeelname en leeftijdsklasse levert dit een getal o_{tij} , met foutmaat so_{tij} .

Voor de slachtoffercategorieën is een Poissonmodel gebruikt. Dit is niet de beste oplossing, maar voorlopig wel de best haalbare. Normaliter wordt het aantal ongevallen verondersteld Poisson-verdeeld te zijn in plaats van het aantal slachtoffers. In de toekomst moet onderzocht worden of een verfijning op dit punt nuttig of zelfs noodzakelijk blijkt.

Veel verschillen in voorspellingen zelf zal dit niet opleveren, zeker niet bij voorspellingen op korte termijn.

3.4 Samenvatting

Van de hiervoor genoemde gegevens zijn feitelijk gebruikt:

- b_t Bevolkingsgegevens per jaar vanaf 1900 tot 2050, de fout sb_t is nul in het verleden. Dit houdt in dat verondersteld wordt dat we (bijna) precies weten hoeveel inwoners Nederland had op 1 januari van een bepaald jaar. We veronderstellen dat dat aantal gelijk blijft over dat jaar. Dit zou beter kunnen via interpolatie, zodat bijvoorbeeld het aantal inwoners per 1 juli gebruikt kan worden. Dit is tot op heden niet toegepast. Deze gegevens zijn ook bekend voor leeftijdindelingen b_{tj} .
- v_{ti} De verkeersprestatie in voertuigkilometers voor vervoerwijze i per jaar vanaf 1950.
- m_{tij} Mobiliteitsconstanten per jaar vanaf 1979, voor vervoerwijze i en leeftijdsklasse j .
- o_{tij} Aantallen slachtoffers per jaar t vanaf 1979, vervoerwijze i en leeftijdsklasse j . Zowel voor overleden slachtoffers als voor ziekenhuisgewonden.

Daarnaast zijn er nog enige tabellen met slachtoffergegevens beschikbaar vanaf het jaar 1950.

4 MODELLEN

Voor het Beleidsinformatiesysteem Verkeersveiligheid (BIS-V) zijn prognoses gewenst voor:

- ontwikkelingen van de mobiliteit,
- ontwikkelingen van het verkeersrisico,
- ontwikkelingen in termen van aantallen slachtoffers (overledenen en gewonden).

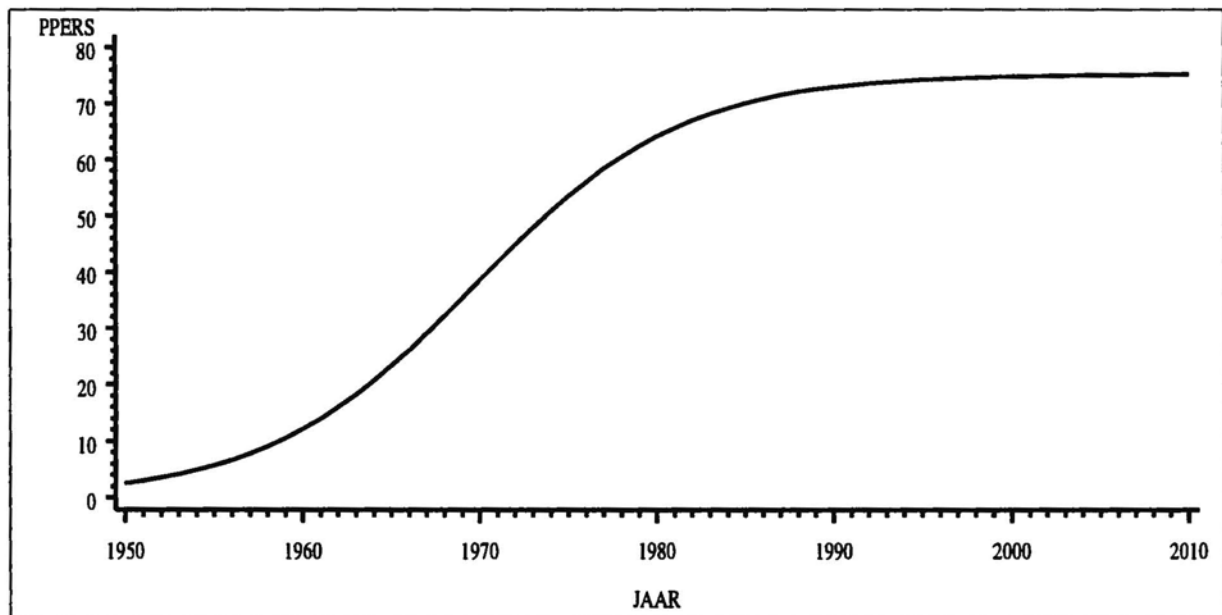
De mobiliteitsontwikkelingen worden geacht uiteindelijk extern aangeleverd te worden, eventueel met alternative scenario's. De risico- en slachtofferontwikkelingen zullen met name een inbreng van de SWOV betreffen.

4.1 Verkeersprognoses

Voor de verkeersprognoses wordt de ontwikkeling verondersteld volgens een logistiek model,

$$y_t = \frac{\psi}{1 + e^{\kappa + \rho t}} \quad (4.1)$$

Over het algemeen zal ρ een negatieve waarde hebben. Hierbij behoort een reeks sy_t van standaardfouten. De reeks y_t bestaat in de voorperiode uit verkeersprestaties uit par 3.2.1.



Figuur 1: Logistische kromme $y_t = \frac{\psi}{1 + e^{\kappa + \rho t}}$, zonder bevolkingseffect

Voorlopig is aangenomen dat deze ontwikkeling geldt per individu. Een alternatief zou zijn aan te nemen dat deze ontwikkeling geldt voor de gehele bevolking. Gegeven de ontwikkeling per individu resten twee schattingsmethoden:

- De mobiliteit per hoofd van de bevolking te berekenen en daarop een logistische ontwikkeling schatten:

$$\frac{y_t}{\text{populatie}_t} = \frac{\bar{\psi}}{1 + e^{\bar{\kappa} + \bar{\rho}t}} \quad (4.2)$$

- Een logistische ontwikkeling op de totale mobiliteit te schatten met expliciet gebruik van de populatiegrootte:

$$y_t = \frac{\text{populatie}_t \times \bar{\psi}}{1 + e^{\bar{\kappa} + \bar{\rho}t}} \quad (4.3)$$

De laatste methode (4.3) heeft een aantal voordelen:

- Door het verschil in behandeling van de residuen door de vermenigvuldiging met de populatiegrootte zijn de observaties met een grote populatie het belangrijkste. Praktisch gezien heeft dit ook nog het voordeel dat de meest recente observaties de meeste invloed hebben daar de populatie nog steeds toeneemt.
- Deze methode maakt verschil tussen de mobiliteitsgroei per persoon en de omvang van de bevolking. Beide effecten kunnen gemodelleerd worden.
- Deze opzet sluit op logische wijze aan bij de eveneens op de omvang van de populatie gebaseerde veronderstellingen over gedisaggregeerde ontwikkelingen.

Bij een ongeveer constante populatie maakt het natuurlijk weinig of niets uit. Bij grote veranderingen in omvang van deelpopulaties lijkt het tweede model niet realistisch. Op praktische gronden is voorlopig gekozen voor toepassing van de tweede methode (4.3).

In principe zijn schattingen van de totale aantallen motorvoertuigkilometers beschikbaar vanaf 1950. Gegevens over andere wijzen van verkeersdeelname zijn vanaf 1974 beschikbaar. Alleen voor personenautoverkeer en vrachtverkeer lijkt de logistische aanname toepasbaar. Daarom worden alleen logistische functies berekend voor personenauto- en vrachtverkeer in respectievelijk de variabelen $y_t(\text{pers}, .)$ en $y_t(\text{vracht}, .)$. De andere gegevens zijn onbetrouwbaar of worden geacht zich niet logistisch te ontwikkelen. Voor deze andere wijzen van verkeersdeelname (bijvoorbeeld fietsen, lopen) is een gelijk blijvende vervoersprestatie per inwoner aangenomen. De verkeersprognoses voor de totale populatie of groepen daarvan behoeven dus niet constant te zijn. In Figuur 2 tot en met Figuur 4 zijn de verschillende ontwikkelingen van de verkeersprestatie van personenauto's en vrachtauto's in grafieken weergegeven. Deze worden zowel voor het model zonder populatiegegevens weergegeven als voor de modellen met populatiegegevens. Deze laatste gegevens zijn uiteindelijk gebruikt.

4.2 Model voor gedisaggregeerde mobiliteit

Dit model is zoals gezegd toegepast op de mobiliteitsgegevens uit het OVG, welke vanaf 1979 beschikbaar zijn. Het eerste afgeleide cijfer dat in het model wordt gebruikt is de *relatieve mobiliteit* $rm_{i,j}$. De index i staat voor de wijze van verkeersdeelname, de index j de leeftijdsgroep waarvoor de mobiliteit m is uitgesplitst. Deze mobiliteit is relatief

in de zin dat dit de mobiliteit per hoofd van de bevolking voorstelt. Deze relatieve mobiliteit wordt zowel per leeftijdscategorie als per wijze van verkeersdeelname bepaald. Dit getal wordt berekend door de mobiliteitscijfers te sommeren voor de jaren waarvoor deze aanwezig zijn en te delen door de som van de bevolkingsaantallen (per leeftijdsklasse) van diezelfde jaren. De gegevens van het vrachtverkeer leveren echter problemen op. Het relatieve aantal wordt verondersteld constant te blijven over de periode 1979 t/m 1990 en met name in de jaren daarna, om prognoses mogelijk te maken. Onderzocht zou kunnen worden of er ook hier een trendmatige ontwikkelingen zijn. Dit zal echter problematisch blijven, onder meer vanwege het geringe aantal jaren welke beschikbaar zijn om de trend te schatten.

Daar de bevolkingsaantallen (bijna) exact bekend zijn, is voor de relatieve mobiliteit op eenvoudige wijze een fout term rm_{ij} vast te stellen. De relatieve mobiliteit wordt als volgt gedefiniëerd:

$$rm_{ij} = \frac{\sum_k m_{kij}}{\sum_k b_{kj}} \quad (4.4)$$

Waarin b_{kj} de bevolking van de leeftijdsklasse j in het jaar k voorstelt. Hierin volgt k de tijd. In het vervolg definiëren we voor elke leeftijdsklasse j :

$$\beta_j = \frac{1}{\sum_k b_{kj}} \quad (4.5)$$

Zodat (4.4) ook als volgt geschreven kan worden:

$$rm_{ij} = \beta_j \sum_k m_{kij} \quad (4.6)$$

De tweede afgeleide variabele is de *voorspelde* mobiliteit em_{tij} voor een leeftijdsklasse en wijze van verkeersdeelname. Dit getal wordt berekend door voor elk jaar de relatieve mobiliteit te vermenigvuldigen met het bevolkingsaantal van die leeftijdsklasse. Dit is dus *niet* het omgekeerde van de vorige stap, waarin de bevolking over de jaren is opgeteld. In de voorspellingsperiode houdt dit in dat twee grootheden welke aan het toeval onderhevig zijn, met elkaar worden vermenigvuldigd. Het berekenen van de foutterm in het produkt is niet eenvoudig en zal lineair benaderd worden. In de geobserveerde periode (de periode tot heden) is dit geen probleem.

$$em_{tij} = rm_{ij} b_{tj} = \beta_j b_{tj} \sum_k m_{kij} \quad (4.7)$$

Deze voorspelde mobiliteit wordt gebruikt om per vervoerwijze i de verhouding tussen de verkeersprestaties over de leeftijdsklassen j te bepalen. Dit levert een 'fractie' mobiliteit op: fm_{tij}

$$\begin{aligned} fm_{tij} &= \frac{em_{tij}}{\sum_r em_{tir}} = \\ &= \frac{\beta_j b_{tj} \sum_k m_{kij}}{\sum_r \beta_r b_{tr} \sum_k m_{kir}} \end{aligned}$$

Waarin r wordt gebruikt voor optelling over de leeftijdsklassen. Duidelijk is dat onder invloed van de ontwikkeling van de bevolkingssamenstelling deze fracties jaarlijks ten opzichte van elkaar veranderen.

Per vervoerwijze i en leeftijdklasse j wordt nu een jaarlijkse verkeersprestatie ev_{tij} berekend door het produkt te nemen van de totale, jaarlijkse verkeersprestatie voor die vervoerwijze en de 'fractie' mobiliteit voor die verkeerswijze in de betreffende leeftijdklasse in dat jaar:

$$ev_{tij} = v_{ti} f m_{tij} \quad (4.8)$$

Deze grootheid wordt de *voorspelde verkeersprestatie* genoemd.

Duidelijk is dat beide termen aan het toeval onderhevig zijn, zodat de foutafschatting ook hier niet triviaal meer is.

4.3 Gedisaggregeerd model van risico

Gebruikmakend van deze voorspelde verkeersprestatie en de geobserveerde slachtoffergegevens o_{tij} kunnen nu risico's berekend worden. Belangrijk is dat de risico's dus niet direct met behulp van de mobiliteitsgegevens uit het OVG berekend worden, doch dat gebruik gemaakt wordt van de enigzins egaliserende werking van de 'voorspelde' verkeersprestatie. De risico's zijn als volgt:

$$r_{tij} = \frac{o_{tij}}{ev_{tij}}$$

Vervolgens kan op analoge wijze als bij de verkeersprestatie de invloed van de populatieomvang op de risico's rechtstreeks geschat worden, waarbij wordt aangenomen dat de risico's een *negatief* exponentiële ontwikkeling volgen:

$$r_{tij} = e^{(at+b)}, \quad a \leq 0 \quad (4.9)$$

of door deze indirect te af te leiden uit:

$$o_{tij} = ev_{tij} \times e^{(\bar{a}t+\bar{b})}, \quad a \leq 0 \quad (4.10)$$

Methode twee (4.10) heeft het voordeel dat Poisson-aannames eenvoudig geïmplementeerd kunnen worden omdat de risico's zó geschat kunnen worden dat de O_{tij} 's meest aannemelijk zijn. Het heeft echter het nadeel ten opzichte van de eerste, klassieke methode (4.9) dat het risico slechts indirect wordt geschat, met mogelijke consequenties vandien. In het BIS-onderzoek is de laatste methode (4.10) toegepast. In Bijlage A wordt verder op technische aspecten, zoals implementatie binnen SAS ingegaan als ook de constructie van betrouwbaarheidsintervallen.

4.4 Gedisaggregeerd model voor slachtoffers

Indien gebruik gemaakt is van (4.10) wordt de voorspelling van het aantal slachtoffers direct verkregen. Na het schatten van de risico's uit (4.9) is het eenvoudig de slachtoffers te voorspellen voor de diverse categorieën:

$$eo_{tij} = ev_{tij} \times r_{tij}$$

Het verkrijgen van een betrouwbaarheidsmaat is weer niet triviaal. Hier geldt een gelijksoortig probleem als in voorgaande gevallen.

5 BETROUWBAARHEID

5.1 Inleiding

Deze paragraaf bevat een poging enig inzicht te geven in de (on)betrouwbaarheid van de schattingen welke het model moet voortbrengen. Dit zal worden aangegeven met behulp van betrouwbaarheidsintervallen. Een betrouwbaarheidsinterval is een interval waarbinnen een waarde (bijvoorbeeld een voorspelling) met een zekere kans (95% is een gebruikelijke waarde) komt te liggen.

De onbetrouwbaarheid in de voorspellingen is natuurlijk in de eerste plaats het gevolg van toevallige afwijkingen in de geobserveerde data en misspecificatie van het veronderstelde model. Met dat laatste wordt bedoeld dat het werkelijke model anders is dan zoals in het onderzoek wordt gebruikt. Bijvoorbeeld, het werkelijke model zal slechts bij benadering negatief exponentieel zijn. Deze onbetrouwbaarheid in de voorspellingen wordt in de praktijk geschat op basis van de (onbetrouwbaarheid in de) parameterschattingen en de modelafwijkingen. De onbetrouwbaarheid wordt vaak uitgedrukt in betrouwbaarheidsintervallen.

De betrouwbaarheidsintervallen zijn niet direct bedoeld als intervallen waarin met een bepaalde waarschijnlijkheid een aantal ongevallen moet vallen, doch wel noodzakelijk daarvoor in een later stadium. Deze intervallen geven een indicatie van de spreiding in de voorspellingen zelf. Meer precies, voor geven van een betrouwbaarheidsinterval voor het aantal slachtoffers is zowel een betrouwbaarheidsinterval voor de voorspelling noodzakelijk als de bepaling van de variatie in het aantal slachtoffers, gegeven de voorspelling. Dit laatste wordt gewoonlijk via de Poissonvariantie bepaald. Een combinatie van beide bovengenoemde aspecten zal een betrouwbaarheidsinterval voor het aantal slachtoffers opleveren.

5.2 Implementatie

5.2.1 Risico-ontwikkelingen

Daar het berekenen van (kans)verdelingen tot grote problemen kan leiden, is bij dit onderzoek besloten gebruik te maken van *benaderingen* van (kans)verdelingen. In het algemeen zal de verdeling analoog aan de normale verdeling worden uitgedrukt in een verwachte waarde en een variantie of standaarddeviatie. De verwachte waarde is normaliter al beschikbaar, in dat geval wordt ook de variantie geschat. Deze schattingen zullen in de meeste gevallen eveneens benaderingen zijn. Tot nog toe zijn dit eerste-orde-benaderingen geweest. Grofweg houdt dit in dat verondersteld wordt dat de fout in een parameter ondergeschikt is aan zijn waarde. In het algemeen is deze eis een voorwaarde voor het opnemen van een parameter in een model.

Het geval (4.9) heeft de eenvoudigste oplossing: van de termen r_{ti} moet de standaarddeviatie worden aangegeven. Deze standaarddeviatie zal leiden tot een (Fisher) informatiematrix welke, onder onze huidige assumpties, leidt tot een *asymptotische* schatting van

de variantie in de parameters. Deze methode levert dan ook via dezelfde weg schattingen voor de fout in de voorspellingen.

Als eerste, en moeilijkste, de fout in de termen r_{tij} :

$$\begin{aligned}
 r_{tij} &= \frac{O_{tij}}{ev_{tij}} \\
 &= \frac{O_{tij}}{v_{ti} f m_{tij}} \\
 &= \frac{O_{tij} \sum_r em_{tir}}{v_{ti} e m_{tij}} \\
 em_{tij} &= \beta_j b_{tj} \sum_k m_{kij} \\
 &= \frac{O_{tij} \sum_r \beta_r b_{tr} \sum_k m_{kir}}{v_{ti} \beta_j b_{tj} \sum_k m_{kij}} \\
 \beta_j &= \frac{1}{\sum_k b_{kj}} \\
 &= \frac{O_{tij} \sum_r b_{tr} / \sum_k b_{kr} \sum_k m_{kir}}{v_{ti} b_{tj} / \sum_k b_{kj} \sum_k m_{kij}}
 \end{aligned}$$

Duidelijk is dat het verstandiger is deze expressie in delen op te splitsen. De fout in r_{tij} is afhankelijk van zowel de fout in O_{tij} als die in ev_{tij} . Deze zouden samen, eventuele covariantie tussen de fouttermen verwaarlozend, afgeschat kunnen worden als:

$$sr_{tij}^2 = sO_{tij}^2 \times \left(\frac{\partial}{\partial O_{tij}} r_{tij} \right)_{O_{tij}=\hat{O}_{tij}}^2 + sev_{tij}^2 \times \left(\frac{\partial}{\partial ev_{tij}} r_{tij} \right)_{ev_{tij}=\hat{ev}_{tij}}^2 \quad (5.1)$$

Waarbij:

$$sev_{tij}^2 = sv_{ti}^2 \times \left(\frac{\partial}{\partial v_{tij}} ev_{tij} \right)_{v_{tij}=\hat{v}_{tij}}^2 + s f m_{tij}^2 \times \left(\frac{\partial}{\partial f m_{tij}} ev_{tij} \right)_{f m_{tij}=\hat{f} m_{tij}}^2 \quad (5.2)$$

De fout sv_{ti} komt óf uit de voorspellingen van de totale verkeersprestatie óf uit extern verkregen informatie. $s f m_{tij}$ hangt af van em_{tij} als gedefiniëerd in (4.7). Op analoge wijze kan een schatting gemaakt worden voor de fout hierin, welke afhangt van bevolkingsaantallen en de OVG-mobiliteit²

5.2.2 Slachtofferontwikkelingen

De slachtoffercijfers zijn in het eerste geval (4.9) minder ingewikkeld te berekenen dan in het tweede geval (4.10).

In het eerste geval (4.9) zal op analoge wijze als in (5.1) het risico berekend worden uit het produkt van beide termen.

In het tweede geval (4.10) komt de onbetrouwbaarheid uit de asymptotische variantie van het (maximale aannemelijkheids)model. De enige complicatie is dat er een 'errors in variables' model gedefiniëerd moet worden voor het niet-lineaire schattingsmodel. Er zit namelijk nog onzekerheid in de gebruikte mobiliteitsvoorspellingen ev_{tij} . Een oplossing voor dit probleem ligt niet voor de hand, maar er is wel literatuur beschikbaar, bijvoorbeeld Seber & Wild (1988). Zie verder Bijlage A voor de implementatie.

²Gegevens over de betrouwbaarheid van OVG-cijfers ontbreken. Daarom wordt voorlopig de 'standard error of the mean' gebruikt om daarin inzicht te krijgen, ook al omdat aangenomen wordt dat de jaarlijkse OVG-cijfers schatters zijn van dezelfde grootte.

6 CONCLUSIES

De belangrijkste conclusie tot op heden is dat het model nog niet is uitgekristalliseerd. Aan de kant van de technische implementatie moeten nog verschillende tests plaatsvinden en beslissingen worden genomen. Dit proces is nog gaande. Aan de kant van de gegevens zal ook nog enig werk verricht moeten worden. Zo zal moeten worden uitgemaakt of er gebruik gemaakt wordt van ziekenhuisgewonden uit VOR-bestanden zoals tot nog toe is gedaan of dat aanvullend gebruik gemaakt moet worden van gegevens uit de Landelijke Medische Registratie. Daarnaast moet nog gecontroleerd worden hoe de gegevens uit het OVG zijn opgehoogd tot landelijke cijfers.

Naast deze technische verfijningen zal in eerste plaats onderzocht moeten worden of het model voor praktische toepassing geschikt is of dat er nog aanvullingen noodzakelijk zijn. Hierbij zal vooral gedacht moeten worden aan manipulatiemogelijkheden met betrekking tot mobiliteitsontwikkelingen. Indien dit ook voor deeltrends noodzakelijk blijkt zal hier voor een grondige aanpassing nodig zijn.

Een mogelijke uitbreiding van het model is het gebruik maken van mobiliteit van botspartners in de niet-lineaire-regressie, zie ook Bijlage A.

Behalve deze inzichten heeft het model tot zoverre ook prognoses voortgebracht. In Figuur 2 en Figuur 3 worden de mobiliteitsontwikkelingen voor personenauto's en vrachtverkeer volgens het logistieke model opgetekend. In Figuur 4 en Figuur 5 wordt hetzelfde gedaan met daarin de bevolkingstoename apart gemodelleerd. Dit leidt tot een enigzins hogere prognose.

In Figuur 6 en Figuur 7 is de ontwikkeling voor zowel ziekenhuisgewonden (VOR) als overleden slachtoffers opgetekend. Deze trends zijn de som van de deeltrends over al het verkeer en leeftijdklassen.

Verder is, bij wijze van voorbeeld, in Figuur 8 t/m Figuur 12 een complete set gegevens bijgevoegd met betrekking tot fietsers van 65 jaar en ouder. Een belangrijk verschijnsel is de misspecificatie van het model voor de ontwikkeling in de mobiliteit. Dit zal het gevolg zijn van het feit dat de fietsmobiliteit (in elk geval voor deze groep) aan het toenemen is. Het is de vraag of deze ontwikkeling tijdelijk of blijvend is. In het laatste geval zal het model hierop aangepast moeten worden.

7 SAMENVATTING EN CONCLUSIES MET BETREKKING TOT MODELLEN VOOR TIJDREEKSANALYSE

Zoals aangegeven zijn de modellen oorspronkelijk ontwikkeld voor het uitvoeren van jaaranalyses en vervolgens aangepast aan de behoeften die voor BIS-V zijn geformuleerd. Behalve voor deze twee doelen zijn er echter meer projecten die op de één of andere manier een relatie hebben met deze wijze van modelleren. Verdere uitwerking van het model is voorzien in het project M&V. Het nader uit te werken model zal echter ook de basis moeten vormen voor diverse andere toepassingen.

7.1 Toepassingsbereik

In de doelstelling van het project M&V is duidelijk aangegeven aan welke voorwaarden het model moet voldoen binnen het project. Deze voorwaarden sluiten aan bij die welke voor de 'Jaaranalyse nieuwe stijl' en BIS-V waren geformuleerd. In feite komt dit neer op een verdere uitwerking van het hier ontwikkelde model. Hieronder zullen kort worden aangegeven welke stappen genomen dienen te worden bij die verdere uitwerking. Daarnaast is er een aantal aspecten die binnen dat project zelf niet aan de orde komen, maar die wel van belang zijn voor het gebruik in een aantal andere projecten. Daarop zal nu eerst worden ingegaan.

7.1.1 Kwartaalanalyse

Tot op dit moment is uitsluitend gewerkt met jaarcijfers. Zowel voor de nieuwe jaaranalyses als voor BIS-V en M&V is dit voldoende. Verdere detaillering naar kwartaal en eventueel zelfs naar maand is mogelijk. Het is een variant die aan het model kan worden toegevoegd ten behoeve van kwartaalanalyses. Deze variant dient in een apart project te worden uitgewerkt. Een nog fijnere tijndeling (dag, uur) is niet zinvol als tijdstap in het model, wel kan het als categorie-indeling worden opgenomen voor toepassing bij specifieke problemen (evenals bijv. dag versus nacht etc.).

7.1.2 Regionaal onderzoek

Hierin is voorlopig voorzien middels een zeer vereenvoudigde procedure: uitsplitsing van het totale aantal slachtoffers naar aandeel per provincie (over de gehele VOR-periode) geeft de verdeelsleutel voor de prognose. Verdere nuancering is mogelijk als bijvoorbeeld mobiliteitsontwikkelingen per provincie worden geleverd door het Nieuw Regionaal Model (NRM). Hierin is op termijn voorzien binnen het project M&V. Op de korte termijn zijn er echter varianten denkbaar, bijvoorbeeld gebaseerd op het bevolkingsaandeel of een op een andere manier geschat aandeel in de reizigerskilometers.

7.1.3 Verkeers- en vervoerregio's

Het probleem hierbij is de toedeling van de mobiliteit aan de diverse wegcategorieën binnen de verkeers- en vervoerregio's. Uiteindelijk hopen we ook over dit soort cijfers te kunnen beschikken in M&V uit het NRM, maar voor de korte termijn zouden schattingen uit andere bron moeten komen, bijvoorbeeld uit het SWOV-project 'Kencijfers'. Een uitsplitsing naar wegcategorie is overigens ook in het project M&V op korte termijn nodig voor de totale mobiliteit.

7.1.4 Interventie-analyse

Om het effect van (meer algemene) maatregelen vast te stellen is het nuttig om te kunnen beschikken over referentiewaarden (zie bijvoorbeeld de tijdreeksanalyse die is uitgevoerd om het effect van de gordelwet vast te stellen). De prognoses van het model kunnen worden gebruikt als referentiewaarden voor ongewijzigd beleid. Bij dit soort onderzoek is vaak sprake van een specifieke opsplitsing van het aantal slachtoffers (bijvoorbeeld dag vs. nacht, alcohol vs. geen alcohol etc.). Mogelijke hypothesen zijn dan dat de verhoudingsconstante eenmalig verandert door de interventie (incidentele werking van de maatregel), dan wel dat de ontwikkelingen uit elkaar zullen lopen (continue werking).

7.1.5 Ongevallenonderzoek

Tot nu toe is uitgegaan van de aantallen slachtoffers. Vaak wordt echter onderzoek gedaan naar de ongevallen waarbij die slachtoffers vallen. Dit heeft ook statistische voordelen: de assumptie dat ongevallen onafhankelijk van elkaar gebeuren (de ongevalsprocessen niet aan elkaar gerelateerd zijn) is een redelijke assumptie, maar dit geldt niet voor de slachtoffers. Als eigenlijke analyses zijn gebaseerd op ongevallen, waarvan de slachtofferaantallen worden afgeleid door vermenigvuldiging met constanten, dan is dit probleem oplosbaar. In feite zal bij de meeste toepassingen het aantal slachtoffers primair van belang zijn.

7.2 Voorstel voor nadere uitwerking van het model

7.2.1 Macrotrend voor de mobiliteit en het risico

De eerste disaggregatie van de macrotrend betreft de uitsplitsing naar wijze van verkeersdeelname. In het SWOV-macromodel wordt uitsluitend uitgegaan van de ontwikkeling van het gemotoriseerde verkeer. Er is uitsplitsing nodig naar deeltrends voor personenauto's, vrachtauto's, motoren, bromfietsen en overig (gemotoriseerd) verkeer. Daarnaast is aanvulling nodig voor fietsers en voetgangers. Uiteindelijk verwachten wij deze gegevens aangeleverd te krijgen uit externe bronnen. Het is de vraag of dat ook realiseerbaar is voor alle wijzen van verkeersdeelname. Het betreft hier niet alleen jaarlijkse cijfers, maar ook prognoses over langere perioden. Zo is door ons (voorlopig) aangenomen dat de voetgangers- en fietserskilometers per hoofd van de bevolking constant blijven in de tijd. Hiervoor zijn vast meer aannemelijke assumpties te geven. Vooralsnog (bij gebrek

aan beter) gaan we daar echter nog wel vanuit. Voorgesteld wordt om voorlopig de eerder gemaakte assumpties te handhaven, d.w.z. een logistische trend voor het gemotoriseerde verkeer, met daarin proportionele aantallen personenauto's en vrachtauto's en constante aantallen motoren, bromfietzers en overig verkeer uitgedrukt in reizigerkilometers. Aan deze logistische trend wordt een component toegevoegd voor de economische golf. Dit gebeurt hetzij door toevoeging van een tijdreeksparameter (deze optie is reeds operationeel) of door uitbreiding van het theoretische model met extra parameters (model van Koornstra). Aan deze prognose voor het gemotoriseerde verkeer worden constante hoeveelheden fietsers en voetgangers toegevoegd. Dit totaal zal gelden per hoofd van de bevolking en worden vermenigvuldigd met het (verwachte) aantal inwoners.

Het model wordt zodanig vormgegeven dat elke alternatieve ontwikkeling voor het totaal of delen daarvan kan worden opgenomen. Zo zijn aantrekkelijke varianten een ontwikkeling voor het gemotoriseerde verkeer volgens een logistisch model dat in 2010 uitkomt op het in het SVV geschatte niveau met en zonder gewijzigd beleid. Verder de varianten uit het LMS voor delen van de mobiliteitsontwikkeling, met name voor het gemotoriseerde verkeer. Voor de macrotrend van het risico wordt de (negatief) exponentiële functie als standaardvorm aangenomen voor elk deelrisico, maar met verschillende parameters. Hierbij geldt dat de risico's berekend worden op basis van de 'verwachte' en niet de feitelijk geobserveerde reizigerskilometers. Dit mag wat wonderlijk lijken, maar is theoretisch wel te verdedigen en past ons inziens ook beter in de filosofie van een prognosemodel. Voorgesteld wordt dit zo te laten, met dien verstande dat wel uiteindelijk normering plaatsvindt van alle deelrisico's op de macrotrends voor het risico (apart voor doden en gewonden).

Samengevat komt dit neer op de volgende stappen:

- 1) Keuze van de logistische functie, met economische golf, voor het totale aantal motorvoertuigkilometers.
- 2) Disaggregatie naar vrachtverkeer, motoren, bromfietzen en overig gemotoriseerd verkeer.
- 3) Bepalen van de (macro-) trend voor voetgangers en fietsers.
- 4) Op basis van de OVG-cijfers van 1979 t/m 1990 vaststellen van het niveau van de macrotrend voor elke wijze van verkeersdeelname.
- 5) Omrekenen naar reizigerskilometers.
- 6) Berekenen van de macrotrend voor het risico op basis van de totale aantallen slachtoffers en de totale aantallen reizigerskilometers.

Dit levert macrotrends voor de mobiliteit per wijze van verkeersdeelname van 1950 tot 2010.

7.2.2 Disaggregatie van de macrotrends

Hier worden de volgende stappen onderscheiden:

- 1) Op basis van de OVG-cijfers van 1979 t/m 1990 uitsplitsen naar leeftijd met behulp van een constante per leeftijdscategorie, geeft een macrotrend per leeftijdscategorie, per wijze van verkeersdeelname.
- 2) Delen door de omvang van de leeftijdscategorie geeft aantallen reizigerskilometers per persoon, per leeftijdscategorie, per wijze van verkeersdeelname, per jaar [check: som van aantal personen klopt met opgegeven bevolkingsaantal].
- 3) Vermenigvuldigen van de aantallen reizigerskilometers per hoofd met de (geobserveerde of verwachte) bevolkingsomvang per categorie per jaar levert verwachte aantallen reizigerskilometers per categorie van 1950 tot 2010.

Voor het risico geldt:

- 1) Per jaar worden de aantallen slachtoffers per categorie gedeeld door het verwachte aantallen reizigerskilometers voor die categorie.
- 2) De negatief-exponentiële functie wordt gefit op elke reeks risicocijfers van de $(5 \times 7 - 2)$ categorieën van wijze van verkeersdeelname \times leeftijd.
- 3) De som van de individuele trends van de categorieën wordt genormeerd op de macrotrend voor het totale risico.

Dit levert verwachte risico's per categorie van 1950 tot 2010.

Voor het aantal slachtoffers geldt dan: Vermenigvuldiging van geschatte aantal reizigerskilometers met het geschatte risico per reizigerskilometer geeft het geschatte aantal slachtoffers (overledenen en ziekenhuisgewonden) per categorie per jaar van 1950 tot 2010.

7.2.3 Slotopmerkingen

Toetsing van het model kan op twee manieren plaatsvinden. Ten eerste dient gekeken te worden of het model de bestaande ontwikkelingen bevredigend beschrijft. Daarnaast of de prognoses realistisch blijken te zijn. Daarbij dient onderscheid te worden gemaakt tussen de voorspelling op de langere en op de kortere termijn. Zo zal de tijdreekscomponent korte-termijnafwijkingen dienen te beschrijven, maar de globale macrotrends de richting van ontwikkelingen op een wat langere termijn. Hierbij speelt nog een bijzonder aspect een rol. Indien er sprake is van een monotone afwijking van de lange-termijntrend voor de mobiliteitsontwikkeling van een deelgroep over de gehele OVG-periode, dan zal deze (mede door de keus om risico's te berekenen op basis van geschatte reizigerskilometers in plaats van geobserveerde) gedeeltelijk worden gecompenseerd bij de schatting van de risicoparameters. De aantallen slachtoffers worden dan beter voorspeld dan de basiscomponenten zelf. Het is dan de vraag hoe een dergelijke uitkomst moet worden geïnterpreteerd: als een tijdelijke afwijking van een realistische trend of als een indicatie voor een misspecificatie van het model. Aanpassing van het model betekent dan echter wel dat dit ook consequenties krijgt voor de langere-termijnvoorspelling, bijvoorbeeld door

een constant niveau te laten stijgen tot een hogere waarde. Het heeft geen zin uit esthetische overwegingen ad-hoc verbeteringen aan te brengen voor korte-termijnontwikkelingen zonder daar ook een voorspellende waarde voor de toekomst aan te verblijden.

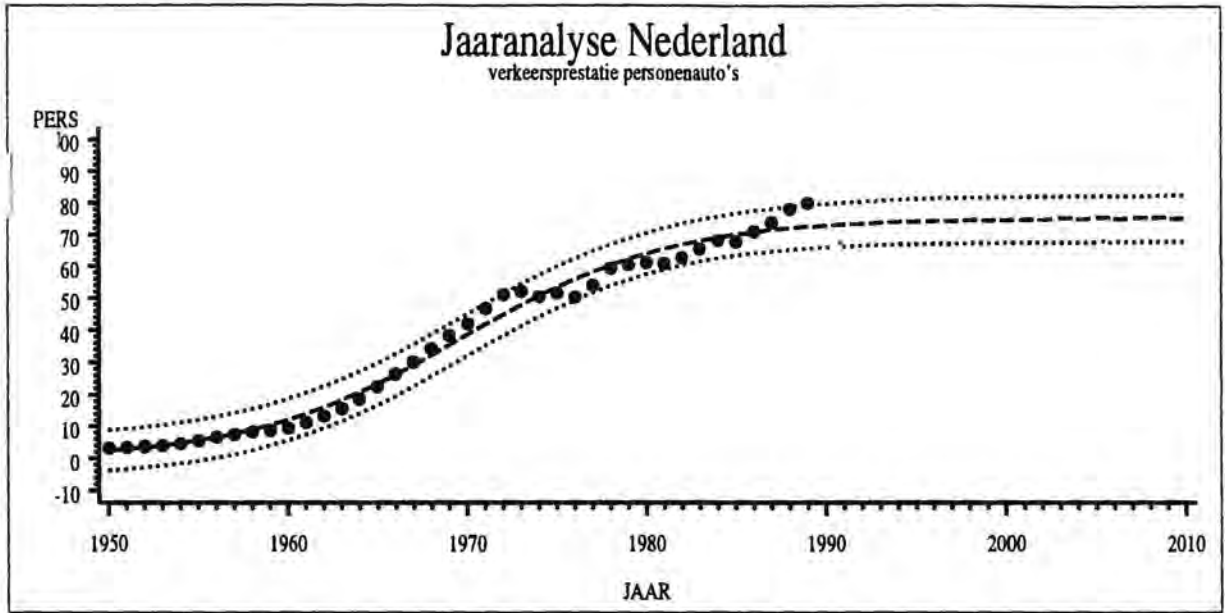
Tot zover is uitsluitend gesproken over een disaggregatie naar leeftijd en wijze van verkeersdeelname. Dit is gebeurd omdat wij aannemen dat uitsplitsing naar andere kenmerken geen extra complicaties zal opleveren. Zo zou bij de uitsplitsing naar geslacht bijvoorbeeld in eerste instantie kunnen worden volstaan met het berekenen van de totale proporties afgelegde kilometers naar geslacht op basis van de OVG-cijfers en het berekenen van algemene risicofactoren naar geslacht. Wanneer significant afwijkende resultaten gevonden worden voor deelgroepen (en naar wij aannemen niet voor het totaal), dan kan verder worden gedisaggregeerd, bijvoorbeeld naar leeftijd of naar wijze van verkeersdeelname, volgens de eerder aangegeven methode. Verdere uitsplitsing zal hoogst waarschijnlijk niet realistisch blijken te zijn.

LITERATUUR

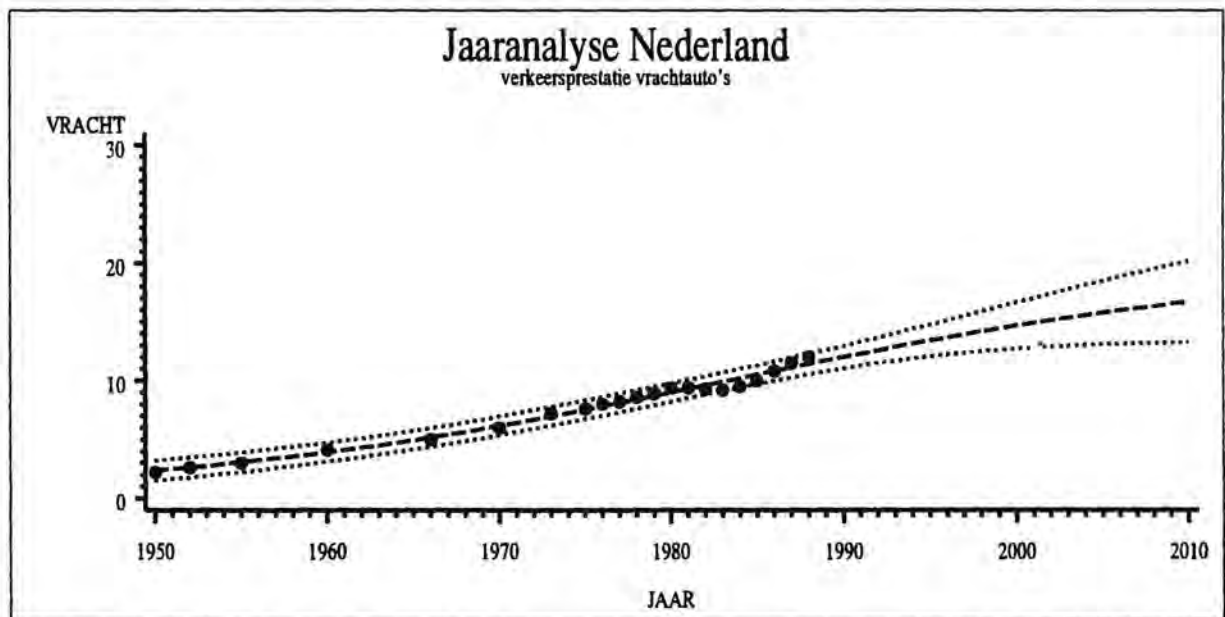
- Bos, J.M.J & Bijleveld, F.D. (1991). Tijdreeksanalyse van het gordeleffect. R-91-92. SWOV, Leidschendam.
- CBS (1989). 90 jaren statistiek in tijdreeksen. Floppy. Kengetal AZ15DB.
- CBS (1991). Veronderstellingen bevolkingsprognose 1991 + kerncijfers 1991-2010. Floppy. Kengetal B19D.
- CBS (1992). Totale vervoersprestatie 1986-1991. Vervoersprestatie onder 12 jaar, Int. ref. Steven Harris dd. 19 juni 1992.
- Harvey, A. C. & Durbin, J. (1986). The effects of seatbelt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society*, 149:187-227.
- Koornstra, M. J. (1988). Development of road safety in some European countries and USA. R-88-33. SWOV, Leidschendam.
- Seber, G. A. F. & Wild, C. J. (1988). *Nonlinear Regression*. John Wiley & Sons, New York.

FIGUREN

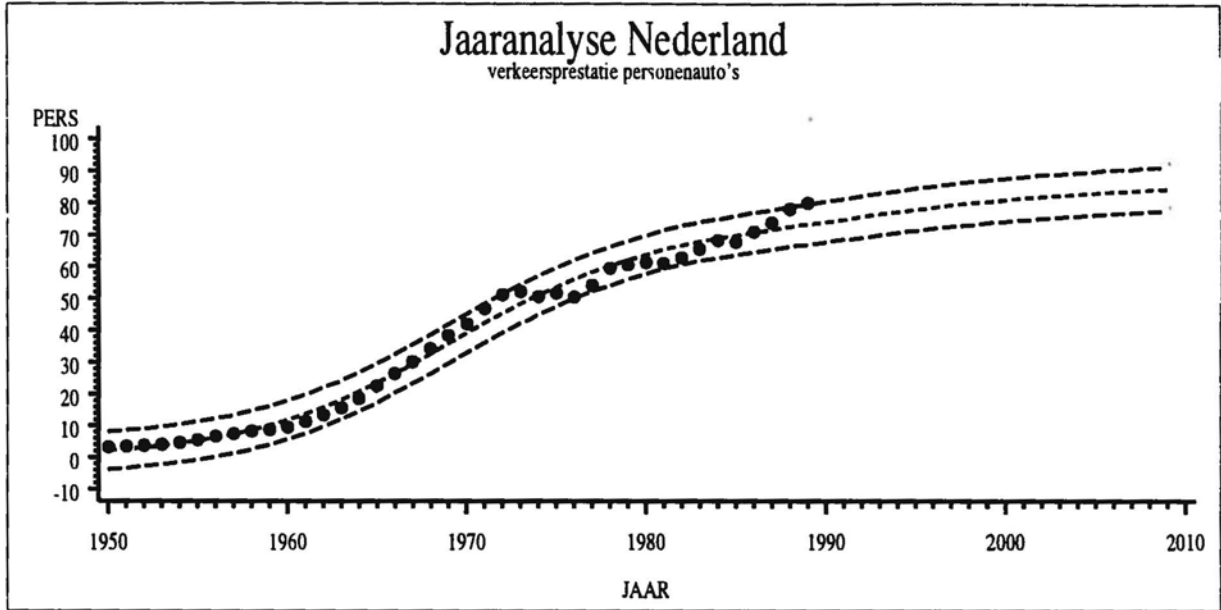
- 1 Logistische kromme $y_t = \frac{\psi}{1+e^{\kappa+\rho t}}$, zonder bevolkingseffect
- 2 Ontwikkelingen en prognoses van de verkeersprestatie van personenauto's
- 3 Ontwikkelingen en prognoses van de verkeersprestatie van vrachtauto's
- 4 Ontwikkelingen en prognoses van de verkeersprestatie van personenauto's met bevolkingsgegevens
- 5 Ontwikkelingen en prognoses van de verkeersprestatie van vrachtauto's met bevolkingsgegevens
- 6 Ontwikkelingen en prognoses bij ziekenhuisgewonden totaal
- 7 Ontwikkelingen en prognoses bij overleden slachtoffers totaal
- 8 Ontwikkelingen en prognoses van de mobiliteit van fietsers van 65 jaar en ouder
- 9 Ontwikkelingen en prognoses van ziekenhuisgewonden onder fietsers van 65 jaar en ouder
- 10 Ontwikkelingen en prognoses van overleden fietsersslachtoffers van 65 jaar en ouder
- 11 Ontwikkelingen en prognoses van het verkeersrisico van ziekenhuisgewonden onder fietsers van 65 jaar en ouder
- 12 Ontwikkelingen en prognoses van het verkeersrisico van bij overleden onder fietsers van 65 jaar en ouder



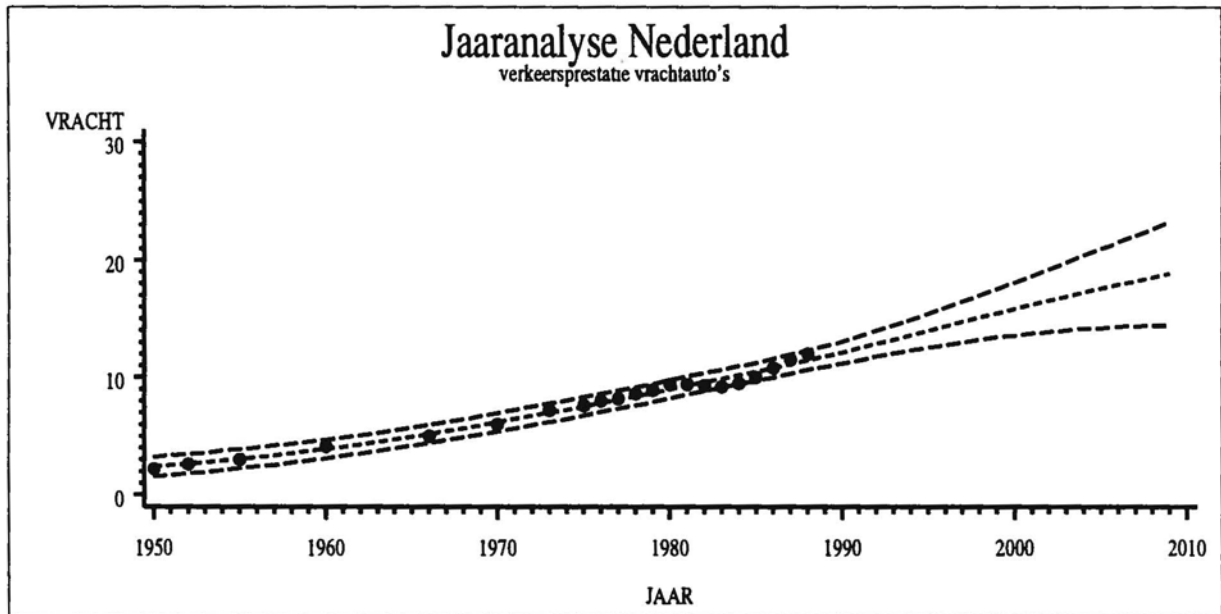
Figuur 2: Ontwikkelingen en prognoses van de verkeersprestatie van personenauto's



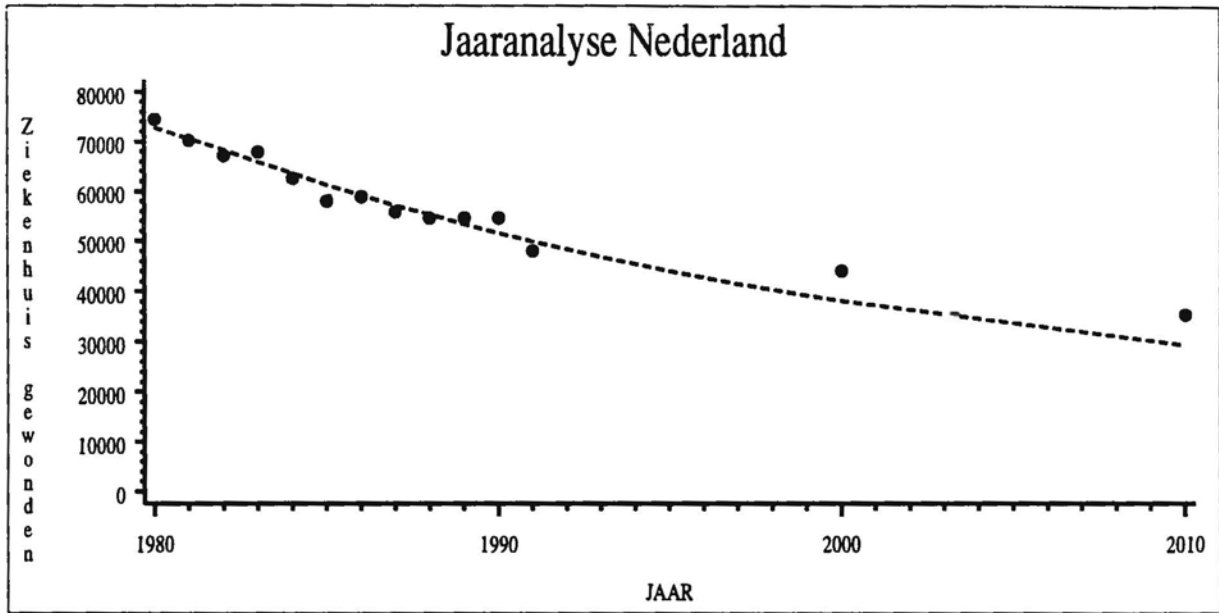
Figuur 3: Ontwikkelingen en prognoses van de verkeersprestatie van vrachtauto's



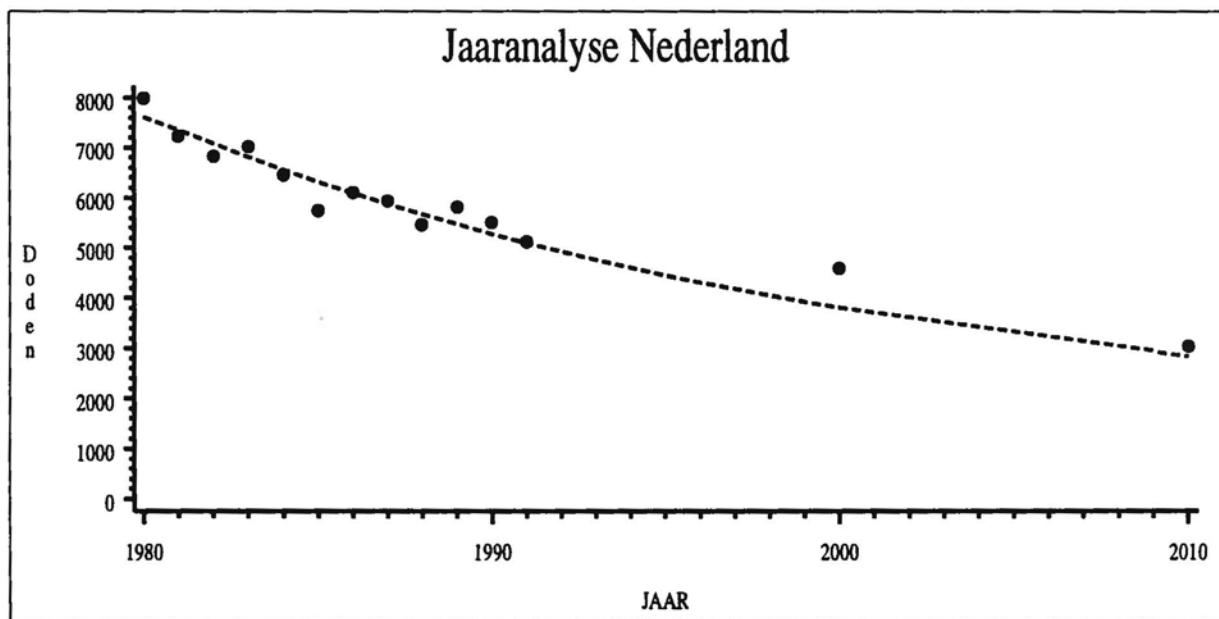
Figuur 4: Ontwikkelingen en prognoses van de verkeersprestatie van personenauto's met bevolkingsgegevens



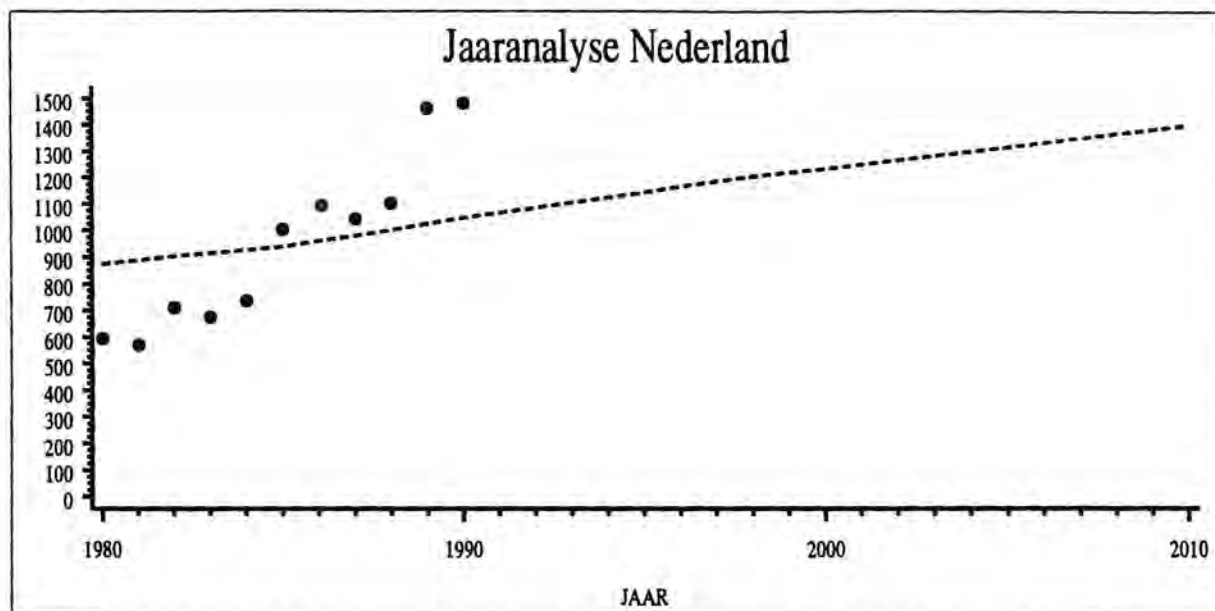
Figuur 5: Ontwikkelingen en prognoses van de verkeersprestatie van vrachtauto's met bevolkingsgegevens



Figuur 6: Ontwikkelingen en prognoses bij ziekenhuisgewonden totaal

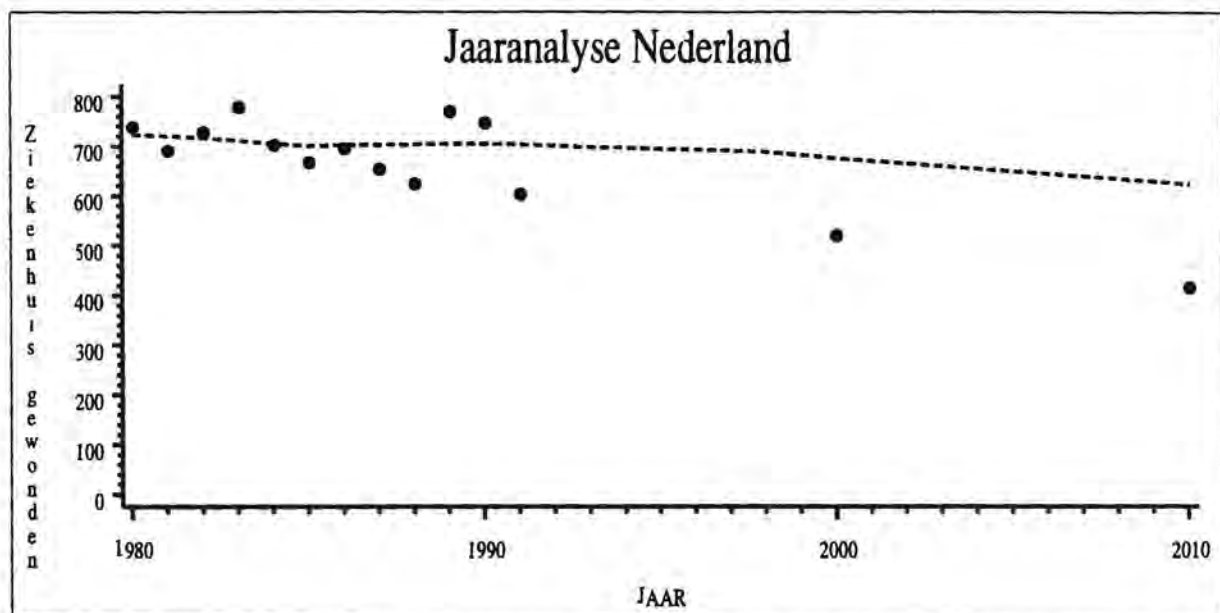


Figuur 7: Ontwikkelingen en prognoses bij overleden slachtoffers totaal

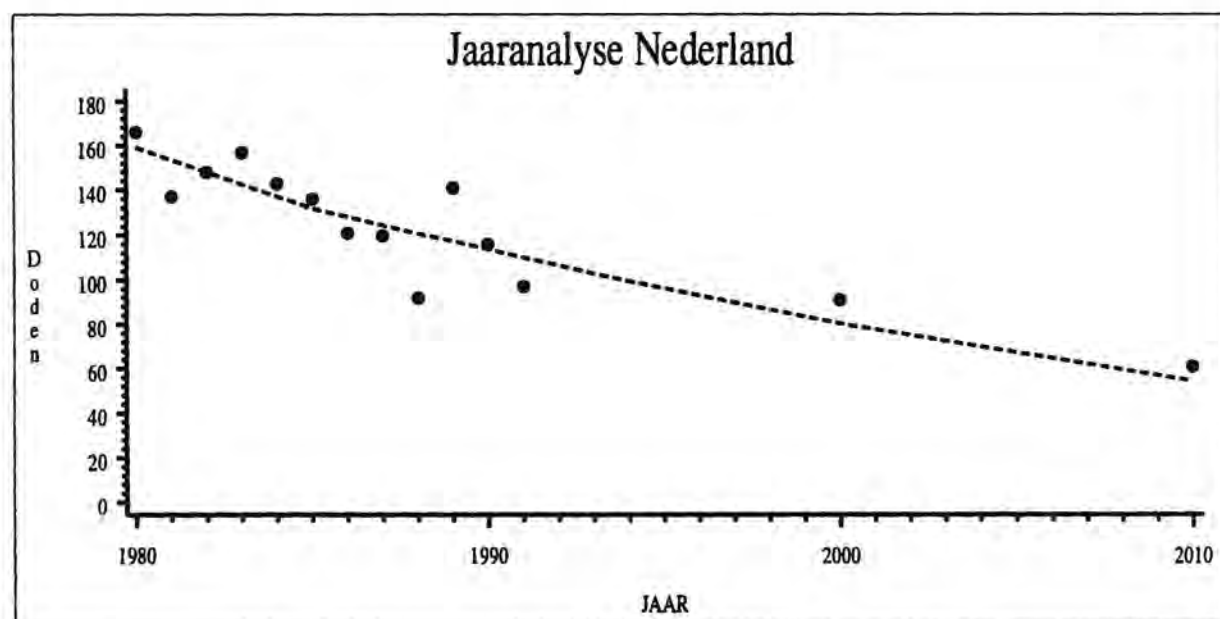


Figuur 8: Ontwikkelingen en prognoses van de mobiliteit van fietsers van 65 jaar en ouder.

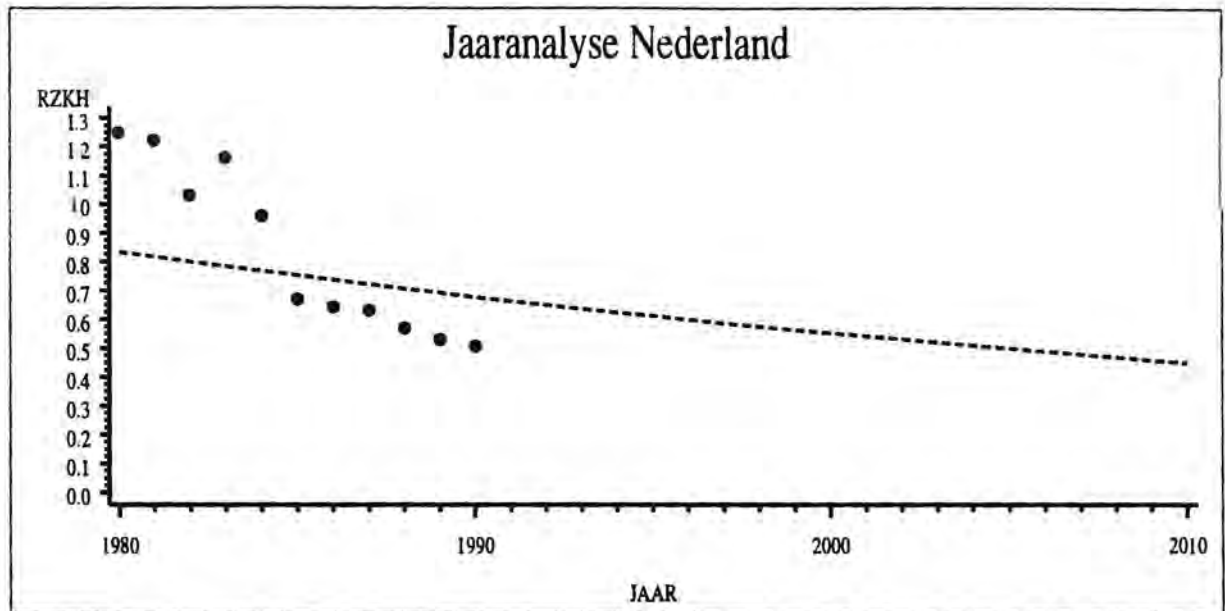
Duidelijk zichtbaar is de systematische afwijking van de prognose, welke veel conservatiever is dan de cijfers uit het OVG. Een verklaring is dat de totale ontwikkeling voor het fietsverkeer in het model zoals nu beschikbaar géén groei in de mobiliteit verondersteld. Op de lange termijn lijkt dit zinnig, doch op de korte termijn lijkt er sprake te zijn van een toename in de mobiliteit van fietsers. Of deze trend zich in dit tempo doorzet lijkt echter ook niet waarschijnlijk. Het lijkt nuttig nieuwe bronnen te zoeken om beter de ontwikkeling van de mobiliteit van fietsers te kunnen weergeven. Het model heeft als eigenschap dat het, via een eveneens incorrecte risico-afschatting, wel tot *redelijke* kortetermijnschattingen voor de slachtoffers komt. Op de langere termijn zullen deze echter wel afwijken van een optimale versie.



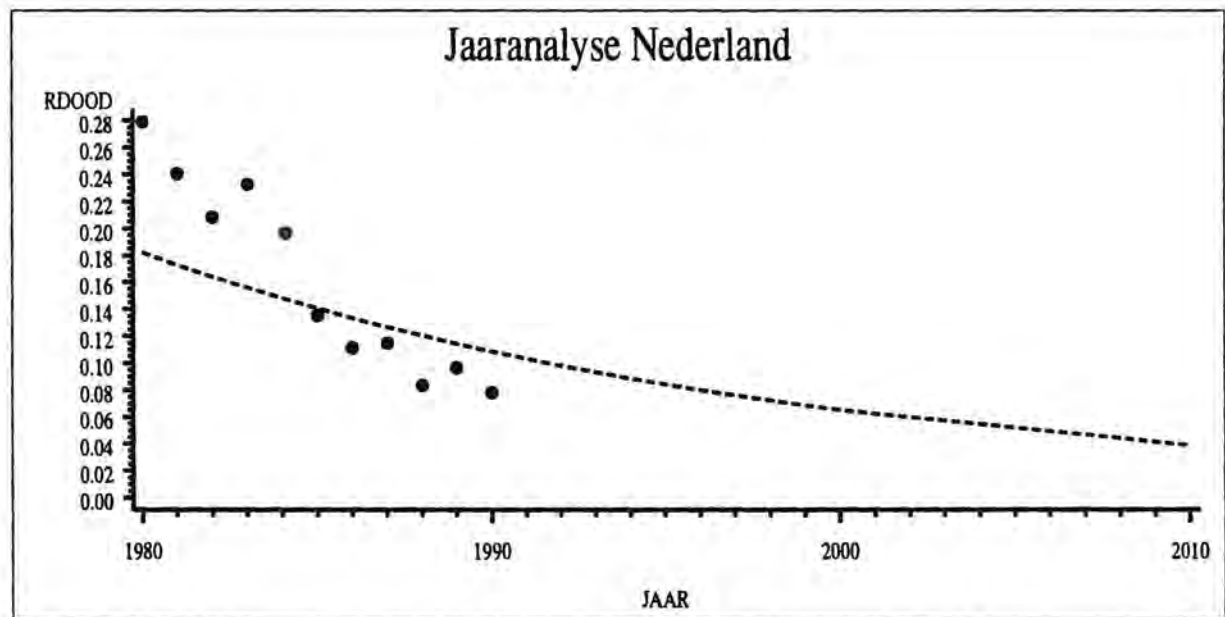
Figuur 9: Ontwikkelingen en prognoses van ziekenhuisgewonden onder fietsers van 65 jaar en ouder.



Figuur 10: Ontwikkelingen en prognoses van overleden fietsersslachtoffers van 65 jaar en ouder.



Figuur 11: Ontwikkelingen en prognoses van het verkeersrisico van ziekenhuisgewonden onder fietsers van 65 jaar en ouder.



Figuur 12: Ontwikkelingen en prognoses van het verkeersrisico van bij overleden onder fietsers van 65 jaar en ouder.



BIJLAGE A NIET-LINEAIRE POISSONREGRESSIE

In principe moet een volgend model worden geschat:

$$o_t = f(x_t, t) \quad (\text{A.1})$$

waarbij o_t verondersteld wordt een Poissonverdeling ($P(y|\mu) = e^{-\mu}\mu^y/y!$) te volgen. De vector x_t stelt een vector externe verklarende variabelen voor, welke overigens ook niet deterministisch hoeft te zijn. Stel we analyseren het aantal overleden fietsers, dan kan x_t bestaan uit een component voor het aantal fietserskilometers én een component voor het aantal motorvoertuigkilometers (op wegen waar fietsverkeer voorkomt). In het onderhavige model wordt aangenomen dat $f(x_t, t)$ in ieder geval een risicocomponent bevat, dit zou als volgt kunnen worden omschreven:

$$f(x_t, t) = g(x_t) \times e^{\alpha t + \beta} \quad (\text{A.2})$$

waarbij bijvoorbeeld $g(x_t) = e^{v_{t,j}}$ zoals in (4.10). Een logische suggestie is dan

$$o_t = e^{\alpha t + \beta + \log(g(x_t))} \quad (\text{A.3})$$

Of als lineaire regressie:

$$\log(o_t) = \alpha t + \beta + \log(g(x_t)) \quad (\text{A.4})$$

Voor de keuze van de schattingsmethode zijn een aantal overwegingen van belang:

- Gevoeligheid voor kleine aantallen; het moet mogelijk zijn het model voor deelpopulaties te gebruiken. Dit kan tot gevolg hebben dat een aantal cellen leeg blijken.
- Flexibiliteit van model specificatie; het moet eenvoudig modelijk zijn aanpassingen aan het model uit te voeren.

Het eerste punt heeft als gevolg dat niet zomaar een *logaritmische* transformatie op de aantallen ongevallen als in (A.4) kan worden uitgevoerd. Er zal in een dergelijk geval een kleine aanpassing aan de data moeten worden toegepast, daar $\log(0)$ niet gedefinieerd is.

Het tweede punt houdt in dat niet altijd een lineaire regressie uitgevoerd kan worden op logaritmisch-getransformeerde data.

Beide punten samen lijken een 'rechtdoorzee' specificatie te suggereren.

Gebruikt is theorie uit Seber & Wild (1988), p. 37, in het algemeen en pp. 46-47, met betrekking tot het Poisson verdeelde geval ten aanzien van de overgang van de LS-methode (Least Squares) via IRLS (Iterated Reweighted Least Squares) naar maximum likelihood.

In grote lijnen werkt de procedure als volgt:

- 0 Er wordt een gewone kleinste-kwadraten analyse uitgevoerd, of eigenlijk, een analyse met gewichten met waarde 1. Er wordt dus een vector gewichten gegenereerd met waarden van 1.
- 1 Met de bestaande gewichten wordt een analyse uitgevoerd.

2 Er wordt gekeken of de methode gereed is, zo niet dan worden nieuwe gewichten genereerd en naar 1 gesprongen.

Bij implementatie bleek het niet noodzakelijk te zijn de IRLS op de klassieke wijze toe te passen. Dit zou zijn uitgevoerd door een aantal malen een complete kleinste-kwadraten analyse uit te voeren. Dit bleek niet nodig. De gebruikte procedure in SAS (PROC NLIN) blijkt een optie te hebben waarmee een ander convergentie criterium gebruikt kan worden. Dit biedt de mogelijkheid het optimaliseringsproces te beëindigen wanneer de likelihood niet meer verhoogd wordt. Het probleem is echter dat de gebruikte procedure nog steeds zoekt alsof nog steeds de kleinste-kwadraten geminimaliseerd worden. Nu is het mogelijk door het herhaald herwegen van de observaties beide convergentie criteria op elkaar aan te laten sluiten. Per sub-iteratie kunnen de gewichten worden aangepast.

De procedure is als volgt: het aantal ongevallen o_t wordt voorspeld met $\hat{o}_t = \exp(\alpha t + \beta + \log(g(x_t)))$. De loglikelihood is:

$$\mathcal{L}(t) = o_t \log(\hat{o}_t) - \hat{o}_t - \log(o_t!) \quad (\text{A.5})$$

Hierbij wordt $\log(o_t!)$ berekend met $\log(\Gamma(o_t + 1))$ en $\log(\hat{o}_t) = \alpha t + \beta + \log(g(x_t))$. In de praktijk wordt meestal niet de likelihood gemaximaliseerd maar de minus likelihood geminimaliseerd. Dit komt uiteraard op het zelfde neer, maar leidt tot een andere notatie.

Zoals vermeld heeft SAS een systeem beschikbaar waarmee een gewicht kan worden gegeven aan een observatie. De bijdrage in de loglikelihood wordt hiermee vermenigvuldigd. De truc is nu in de specificatie van het model de bijdrage aan de loglikelihood met een getal (de variantie) te vermenigvuldigen en vervolgens het gewicht één gedeeld door de variantie te laten zijn:

$$-\mathcal{L}(t) = (\hat{o}_t + \log(o_t!) - o_t \log(\hat{o}_t)) \times \hat{o}_t$$

$$w_t = 1/\hat{o}_t$$

Dit levert waar het de loglikelihood aangaat netto het zelfde resultaat. Dat geldt echter niet voor de kleinste-kwadratenkant van de berekening. Numeriek berekend SAS nog steeds zijn minimum alsof het een kleinste-kwadratenprobleem is. SAS minimaliseert:

$$\sum w_t (o_t - \hat{o}_t)^2 \quad (\text{A.6})$$

waarbij w_t als constante beschouwd wordt. Dit betekent dat de gradient van deze functie gelijk is aan een (negatieve) constante maal de gradient van de loglikelihood. Dat betekent dat beide problemen gelijke oplossingen hebben, hetgeen het gebruik rechtvaardigt van de (asymptotische) betrouwbaarheidsintervallen, welke voor de kleinste-kwadratenanalyse zijn bedoeld.

De gradiënt van (A.5):

$$o_t \frac{\hat{o}_t'}{\hat{o}_t} - \hat{o}_t' = \frac{\hat{o}_t' (o_t - \hat{o}_t)}{\hat{o}_t}$$

De gradiënt van (A.6):

$$\frac{-2\hat{o}_t' (o_t - \hat{o}_t)}{\hat{o}_t}$$