

MULTIPLICATIEVE ANALYSEMODELLEN

Beschrijving en toepassing bij de analyse van ongevallenratio's
als functie van verkeersintensiteit en stroefheid van het wegdek

R-77-10

Drs. S. Oppe

Voorburg, maart 1977

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

INHOUD

Inleiding

1. Gegevens
2. Analyse
3. Additive Conjoint Measurement
4. Stochastische interpretatie van het multiplicatieve model
 - 4.1. Log-lineaire modellen
 - 4.2. Gewogen Poisson-modellen

Literatuur

Tabellen en afbeeldingen

INLEIDING

In 1966 is door de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV in Nederland de Werkgroep Banden, wegdekken en slipongevallen opgericht. Subcommissie V van deze Werkgroep had tot taak de omvang van het aantal slipongevallen vast te stellen. Voorts zou de rol die wegdekstroefheid speelt bij het ontstaan van ongevallen in beschouwing worden genomen. In deze Subcommissie waren de volgende instanties vertegenwoordigd:

Het Rijkswegenbouwlaboratorium te Delft, de Dienst Verkeerskunde van de Rijkswaterstaat te Den Haag en de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV te Voorburg.

Om de omvang van het slip-probleem, te onderzoeken zijn ongevallen welke plaatsvonden op droog wegdek vergeleken met die welke op nat wegdek met en zonder regen plaatsvonden. De rol van de stroefheid is alleen onderzocht voor ongevallen tijdens regen. In dit laatste onderzoek zijn een aantal variabelen als snelheid en zicht om praktische redenen buiten beschouwing gelaten. Wel in het onderzoek zijn betrokken de uurintensiteiten van het verkeer, de verkeersprestatie, het wegtype en het voertuigtype.

Deze bijdrage is niet bedoeld om bovenstaand onderzoek weer te geven. Hiervoor wordt verwezen naar Schlösser (1977).

Het gaat hier slechts om het beschrijven van toegepaste analysetechnieken en uit die toepassing resulterende conclusies over de relatie tussen ongevallen, verkeersprestatie, uurintensiteiten en wegdekstroefheid. Bij de analyse wordt er vanuit gegaan dat het verkeer op twee manieren een rol kan spelen bij het tot stand komen van een ongeval. Enerzijds zal bij meer verkeer het verwachte aantal ongevallen toenemen ten gevolge van het grotere aantal verkeersdeelnemers dat vatbaar is voor een ongeval, met andere woorden: de expositie wordt groter. Ten aanzien van dit punt mag men verwachten dat er een evenredige verhoging van het aantal ongevallen is met de verkeersprestatie. Anderzijds zal bij hogere uurintensiteiten voor iedere individuele weggebruiker de kans om bij een ongeval betrokken te raken groter worden, met

andere woorden: de ongevalsvatbaarheid neemt toe.

Bij de analyse is een correctie toegepast voor de mate waarin de expositie een rol speelt. Hiertoe zijn de ongevallen welke plaatsvinden gedurende een bepaalde tijd op een bepaald weggedeelte gedeeld door de tijdens die periode op dat weggedeelte verreden voertuigkilometers. Deze ongevallenratio's worden geanalyseerd. Om de invloed van de verkeersintensiteit op de ongevalsvatbaarheid na te gaan is naast de toegepaste correctie op voertuigkilometers deze variabele gebruikt ter verklaring van het verschil in ongevallenratio's. Getracht wordt dus de ongevallenratio te beschrijven als een functie van zowel de stroefheid van het wegdek als van de verkeersintensiteit. Het is aannemelijk om te veronderstellen dat de toename van ongevalsvatbaarheid met de verkeersintensiteit niet voor alle wegen gelijk zal zijn. Dit heeft ertoe geleid dat de wegen zijn verdeeld in twee typen.

Tot wegtype I behoren de autosnelwegen: wegen met ongelijkvloerse kruisingen en gescheiden rijbanen, elk met minstens twee rijstroken en in het algemeen één vluchtstrook. Tot wegtype II behoren de overige primaire rijkswegen, voornamelijk wegen met één rijbaan, twee rijstroken, gelijkvloerse kruisingen en soms langzaam verkeer.

1. GEGEVENS

De voor het onderzoek benodigde ongevalgegevens zijn verkregen van Rijkswaterstaat.

Van de ongevallen is geregistreerd op welke locatie ze plaatsvonden, op welke datum en welk tijdstip en of het regende.

Voor wegtype I zijn de uurintensiteiten ingedeeld in 20 klassen met een breedte van 100 voertuigen per uur per rijrichting, voor wegtype II in 15 klassen met een breedte van 200 voertuigen per uur voor beide rijrichtingen. Voor elk weggedeelte is de langskrachtcoëfficiënt bij nat wegdek bepaald. Deze coëfficiënten zijn ingedeeld in negen stroefheidsklassen met een breedte van .05 meeteenheden vanaf $\leq 0,36 \text{ t/m} \geq 0,71$.

Uit de locatie, de datum en het tijdstip is van ieder ongeval bepaald welke wegdekstroefheids- en uurintensiteitsklassen erbij komen. Omdat de hoogste stroefheidsklasse ook ongevallen op natte wegdekken bij droog weer bevat is de gehele klasse uit het onderzoek verwijderd.

Uit de lengte van de weg, de verdeling der uurintensiteiten en de regenduur zijn, apart voor werk- en weekenddagen en gecorrigeerd voor maand en jaar, van elke combinatie van stroefheid en uurintensiteit de afgelegde voertuigkilometers berekend. Vervolgens is per stroefheids-intensiteitscombinatie de ongevalratio bepaald door het aantal ongevallen te delen door het aantal bijbehorende voertuigkilometers.

Dit resulteerde in twee tabellen met ongevalratio's corresponderend met de beide wegtypen.

2. ANALYSE

De bedoeling van de analyse is om na te gaan hoe de ongevalle-
ratio (A) afhangt van de verkeersintensiteit (F) en de stroefheid
van het wegdek (R).

Bij een dergelijk probleem ligt het voor de hand om op de data
multiple lineaire regressie (MLR) toe te passen. Dit komt in
het algemeen neer op een beschrijving van de afhankelijke variabele
als een lineaire combinatie van een aantal (hier twee) onafhanke-
lijke variabelen:

$$Z = aX + bY + c \quad (1)$$

Indien a, b, en c bekend zijn kan, door de waarden van X en Y
in te vullen, de waarde van Z worden gevonden.

Bij MLR wordt dan gezocht naar die waarden voor a, b en c die
de Z-waarden zo goed mogelijk voorspellen uit de X- en Y-waar-
den. Indien we echter de assumpties die bij dit regressiemodel
worden aangenomen, nader bekijken, dan blijkt toepassing ervan
op de gegeven data tot een aantal moeilijkheden te leiden.

Assumptie 1: de lineariteitsassumptie

Deze assumptie houdt in dat, indien de onafhankelijke variabele
X constant worden gehouden, de afhankelijke variabele (Z) lineair
samenhangt met de variabele Y en omgekeerd.

In zo'n geval spreken we van een MLR-model dat lineair is in de
onafhankelijke variabelen.

Deze lineariteitsassumptie brengt een aantal problemen met zich
mee.

Allereerst zal de manier waarop de stroefheidsklassen worden vast-
gesteld, bepalen hoe de ongevalle ratio samenhangt met de stroef-
heid van het wegdek. Op voorhand is er geen enkele reden om aan
te nemen dat deze relatie een lineaire relatie zal zijn. Ook wat
betreft de intensiteit is deze relatie niet zo duidelijk. Het
is inderdaad niet onaannemelijk om te stellen dat de ongevalle-
ratio binnen bepaalde grenzen toeneemt met de intensiteit van

het verkeer. Het is echter mogelijk dat bij zeer lage intensiteit de ongevallenratio toeneemt, terwijl deze bij hoge intensiteit en geringe wegcapaciteit juist afneemt. Een globale inspectie van de data doet vermoeden dat deze veronderstelling juist is. Voor het middengebied is het dan nog de vraag of de samenhang lineair is. Om aan deze problemen tegemoet te komen is het mogelijk het MLR-model uit te breiden met termen die kwadratisch zijn in de onafhankelijke variabelen of met termen van nog hogere orde.

Assumptie 2: de additiviteitsassumptie

Volgens deze assumptie zal de afhankelijke variabele geschreven kunnen worden als een (gewogen) som van onafhankelijke variabelen. Ten aanzien van assumptie 2 kan het volgende worden opgemerkt. Stel de kans dat een gegeven ongeval plaatsvindt op een wegdek behorend tot stroefheidsklasse j ($j=1, \dots, m$) wordt aangeduid met $p(R_j)$ en de kans dat dit ongeval plaatsvindt in intensiteitsklasse i ($i=1, \dots, n$) wordt aangeduid met $p(F_i)$. Indien we nu aannemen dat beide kansen onafhankelijk van elkaar zijn (hetgeen betekent dat de kansverdeling over de stroefheidsklassen gelijk is voor iedere intensiteitsklasse en omgekeerd) dan volgt daaruit dat de kans op een ongeval voor de combinatie van intensiteitsklasse i en stroefheidsklasse j geschreven kan worden als het produkt van de (marginale) kansen $p(F_i)$ en $p(R_j)$ dus:

$$p(F_i \cap R_j) = p(F_i) \cdot p(R_j) \quad (2)$$

Deze overweging zou moeten leiden tot de keuze van een multiplicatief model in plaats van een additief model.

We zouden deze hypothese kunnen toetsen door het eerder genoemde MLR-model (1) uit te breiden door toevoeging van een XY term, dus:

$$Z = aX + bY + cXY + d \quad (3)$$

Als de hypothesen omtrent lineariteit en multipliciteit waar zijn, dan verwachten we dat $a=b=0$ en $c \neq 0$.

Een volgende suggestie zou dan kunnen zijn om niet de data zelf te analyseren, maar de analyse uit te voeren over de logaritme van de data. Immers, als $Z = XY$, dan geldt $\log(Z) = \log(X) + \log(Y)$ en gaat multipliciteit over in additiviteit. Uit deze analyse zou de gevraagde informatie omtrent de bijdrage van F en R aan A dan kunnen worden afgeleid.

Vanuit formule (2) en de discussie omtrent de lineariteitsassumptie dringt zich echter de suggestie op om voor elke klasse van R en F een aparte parameter op te nemen in het model. Binnen het MLR-model is dit bijvoorbeeld mogelijk door een $m-1^{\text{de}}$ graads polynoom in R en een $n-1^{\text{de}}$ graads polynoom in F te gebruiken. Een model waarin op enigszins andere wijze aan deze eis tegemoet gekomen wordt, is het Additive Conjoint Measurement (ACM) model.

3. ADDITIVE CONJOINT MEASUREMENT

In het ACM-model wordt wel de additiviteitseis gesteld, maar ook hier kan door toepassing van een logaritmische-transformatie eventuele multipliciteit worden omgezet in additiviteit. De lineariteits-eis is vervangen door de eis dat er willekeurige functies f op X en g op Y zijn met behulp waarvan Z (of een logaritmische-transformatie van Z) beschreven kan worden als functie van X en Y .

Voor iedere waarde Z_{ij} , behorend bij de combinatie (X_i, Y_j) geldt dat:

$$Z_{ij} = \alpha_i + \beta_j + \gamma \quad (4)$$

Indien de n maal m Z -waarden worden opgevat als een vector Z en de $n + m + 1$ parameters als een vector θ , dan is model (4) te schrijven als $Z = V\theta$, waarin V wel de design matrix wordt genoemd. V is dan een matrix van enen en nullen zodanig dat aan iedere Z -waarde de parameters worden toegevoegd overeenkomstig de indexen i en j . Bij variantie-analyse ligt bij de specificatie van V de nadruk op toetsing van het "experimental design", bij ACM gaat het om het gezamenlijk meten van de variabelen X en Y met behulp van de parameters. Bij de genoemde MLR-modellen zou de matrix V worden vervangen door een matrix met als kolomvectoren de onafhankelijke variabelen of polynomen daarvan. Aan deze MLR-modellen zouden interactie-termen kunnen worden toegevoegd als XY , X^2Y , enz. Bij ACM-analyse wordt ervan uitgegaan dat deze effecten afwezig zijn.

Tot nu toe zijn er twee alternatieven genoemd voor toepassing van ACM. De eerste toepassingsmogelijkheid is de analyse direct op de ongevalratio's toe te passen, de tweede mogelijkheid is om de analyse toe te passen op de logaritmische van de ongevalratio's.

Een andere mogelijkheid is, om uitgaande van een door Kruskal (1965) ontwikkelde techniek, over te gaan op een analyse waarin

gezocht wordt naar die monotoon niet-dalende transformatie van Z , die, indien ingevuld voor Z , een oplossing geeft voor vergelijking (4). Als achteraf wordt nagegaan welke monotone transformatie tot een goede fit van het ACM-model leidt, kunnen bovenstaande argumenten omtrent additiviteit of multipliciteit alsnog worden geverifieerd. Bijvoorbeeld, is de monotone transformatie een lineaire transformatie, dan had ACM direct kunnen worden toegepast, is het een logaritme-transformatie dan pleit dit voor een multiplicatief model. Toegepast op MLR-modellen wordt meestal gesproken van niet-metrische MLR-analyse. Dit omdat in feite wordt aangenomen dat over Z slechts ordinale informatie bestaat.

Wat exacter geformuleerd komt deze methode op het volgende neer.

Stel dat f en g bekend zijn, dan is er voor elke Z_k en Z_l een $Z_k^* = f(X_k) + g(Y_k)$ en een $Z_l^* = f(X_l) + g(Y_l)$ waarvoor geldt dat indien $Z_k \geq Z_l$ dan ook $Z_k^* \geq Z_l^*$.

Samengevat betekent dit dus:

$$Z_k^* = f(X_k) + g(Y_k) \geq Z_l^* = f(X_l) + g(Y_l)$$

als en alleen als $Z_k \geq Z_l$

waarbij k en l indices zijn welke de stroefheids-intensiteitscombinaties $(1,1)$, \dots , $(1,n)$, \dots , (m,n) doorlopen.

In het algemeen zal een dergelijke transformatie maar tot zekere hoogte mogelijk zijn. Gezocht zal dan worden naar die transformatie waarvoor het model een zo goed mogelijke beschrijving van de data geeft. Als criterium voor een optimale beschrijving wordt hier evenals bij MLR gekozen voor een kleinste-kwadraten-criterium. Met andere woorden: laat Z_k^+ de waarde zijn behorend bij een bepaalde monotone transformatie. En Z_k^* de bijbehorende best bij model (4) passende predictie van Z_k^+ , dan wordt gezocht naar die monotone niet-dalende transformatie waarvoor de som van de discrepanties (S) tussen de Z_k^+ en Z_k^* waarden zo klein mogelijk is. Of, nauwkeuriger omschreven, waarvoor geldt dat

$$S = \min_{Z^+} \min_{\theta} \left[\frac{\sum_k (Z_k^+ - Z_k^*)^2}{\sum_k (Z_k^* - Z_k^*)^2} \right]$$

De noemer van deze expressie is hierbij slechts een schaalfactor. In een iteratief proces waarin gezocht wordt naar de best passende monotone transformatie worden de Z-waarden zelf als aanvangsconfiguratie gekozen.

Door de waarde van S , gevonden bij de aanvangsconfiguratie (S_d), te vergelijken met S_{mon} van de monotone transformatie is het mogelijk na te gaan hoezeer de oplossing kan worden verbeterd als we een monotone transformatie op Z toelaten. Indien we de analyse ook op de $\log Z$ -waarden toepassen, krijgen we weer een aanvangsoplossing met bijbehorende S_{\log} die vergeleken met S_d laat zien of er eerder van een additief dan van een multiplicatief model moet worden gesproken, terwijl S_{\log} vergeleken met S_{mon} (natuurlijk identiek voor beide uitgangssituaties) weer laat zien hoe deze oplossing kan worden verbeterd.

Indien de hypothese omtrent de multipliciteit juist is, dan verwachten we dat $S_d > S_{\log} = S_{\text{mon}}$

Resultaten^{**}

In Tabel 1A staan de ongevallen voor wegtype I weergegeven en in Tabel 1B de bijbehorende voertuigkilometers. In Tabel 2A en 2B dezelfde waarden voor wegtype II.

Figuur 1 is een weergave van de oplossing voor de acht waarden van de functie f uit de formule

$$A_{ij}^* = f(R_i) + g(F_j)$$

waarbij R_i de stroefheidsklasse i , F_j de intensiteitsklasse j , en A_{ij}^* de bijbehorende ongevallenratio na monotone transformatie weergeeft. Voor de klassen 2 t/m 7 geldt dat de grootte van de parameters ongeveer lineair afneemt met de klassewaarde,

^{**}) Bij de analyse van de gegevens is gebruik gemaakt van het computerprogramma ACM, geschreven in PLI door J. de Leeuw, Rijksuniversiteit Leiden, Sociale faculteit, afd. Datatheorie

Indien het multiplicatieve model juist is (de monotone transformatie dus een logaritmische-transformatie blijkt te zijn) en $f(R_i)$ inderdaad lineair is, dan betekent dit dat het verband tussen ongevalratio en stroefheidsklasse een exponentieel verband is.

Figuur 2 geeft de oplossing voor de waarden van de functie g . Het verband is niet zo duidelijk interpreteerbaar bij wegtype I. Er kan wel uit worden afgelezen dat de ongevalratio toeneemt met de intensiteit van het verkeer, behalve aan de uiteinden van de schaal. Bij zeer lage intensiteiten neemt de kans op een ongeval toe, bij zeer hoge intensiteiten af. Voor wegtype II zien we deze randeffecten niet.

In Figuur 3 en 4 zijn de transformaties van de ongevalratio's afgebeeld voor beide wegtypen. Voor wegtype I en in mindere mate voor wegtype II volgt uit deze figuren dat de transformatie inderdaad kan worden opgevat als een log-transformatie. Later zal blijken dat de extra kromming bij wegtype II niet veel bijdraagt aan het verbeteren van de oplossing.

Om een en ander nader te bekijken zijn de S-waarden voor elk van de fit-procedures van belang. Deze zijn gegeven in Tabel 3, respectievelijk voor de least-squaresoplossing van de oorspronkelijke data, de log-data en de uiteindelijke oplossing na monotone transformatie. Uit de tabel blijkt dat de stress na de log-transformatie over de data kleiner wordt, terwijl deze zelf natuurlijk hoger is dan de stress van de oplossing na monotone transformatie.

Om een idee te krijgen van de mate waarin de gevonden verschillen in stress significant zijn, heeft een Monte Carlo studie plaatsgevonden. De toegepaste procedure is als volgt:

Wijs de gevonden ongevalratio's at random toe aan de R- en F- klassen en pas een ACM-analyse toe op deze data en de bijbehorende log-data. Herhaal dit een groot aantal malen (uit economische overwegingen is dit hier slechts 40 maal gedaan). Bereken de gemiddelden en standaardscores. Met de zo verkregen gemiddelden kunnen de waarden uit Tabel 3 worden vergeleken. De Monte Carlo data

staan in Tabel 4. Uit Tabel 4 volgt (overigens onder de aanname dat de stresswaarden normaal verdeeld zijn):

a) De fit over de oorspronkelijke data en log-data is zeer significant beter dan random. Voor de log-data bijvoorbeeld vinden we een t-waarde van

$$t = \frac{.1334 - .6928}{.058} = -9.64 \text{ (df = 39).}$$

b) De stresswaarden voor de analyse over de Monte Carlo data en Monte Carlo log-data verschillen zoals te verwachten was niet van elkaar ($t = .183$); het verschil tussen de stresswaarden S_d en S_{\log} bij de oorspronkelijke analyse bedraagt .0515. Dit verschil, hoewel vrij hoog ($t = .0515/.038 = 1.35$), is niet significant. Voor wegtype II is het verschil absoluut gezien groter. Het is dus aannemenlijk om te kiezen voor het multiplicatieve model.

c) De gemiddelde (triviale) vermindering van de stress na monotone transformatie van de data voor de Monte Carlo data is .0603. Voor de oorspronkelijke analyse is deze waarde .0180 voor wegtype I en .0441 voor wegtype II, zodat er dus geen reden is om aan te nemen dat door monotone transformatie nog een extra verbetering plaatsvindt die niet triviaal is. Deze conclusie versterkt de opvatting, dat het multiplicatieve model juist is. Er blijkt tevens uit dat de eerder genoemde afbuiging van de curve uit Figuur 4 de oplossing nauwelijks verbetert.

4. STOCHASTISCHE INTERPRETATIE VAN HET MULTIPLICATIEVE MODEL

Als we aannemen dat het ongevalsgebeuren kan worden beschreven als een Poisson-proces met parameter λ en verder dat de ongevallen multinomiaal verdeeld zijn over de stroefheidsklassen en de intensiteitsklassen, terwijl de variabelen intensiteit en stroefheid een onafhankelijke invloed hebben op ongevallenkans, dan geldt:

1. Voor elke stroefheidsklasse R_i met multinomiaalkans p_i en elke intensiteitsklasse F_j met multinomiaalkans q_j kan het ongevalsgebeuren worden beschreven als een Poissonproces, met parameter λp_i en λq_j .

2. Voor iedere cel X_{ij} is de ongevallenverdeling een Poisson-verdeling met parameter $\mu_{ij} = \lambda \cdot p_i \cdot q_j$.

4.1. Log-lineaire modellen

De laatste jaren zijn er analyse-technieken ontwikkeld speciaal bestemd voor gegevens welke zijn verzameld in de vorm van kruistabellen. De boven omschreven indeling van de gegevens in intensiteits- en stroefheidsklassen is een voorbeeld van zo'n kruistabel. Als nu voor de waarden in de cellen van de kruistabel inderdaad kan worden aangenomen dat deze Poisson verdeeld zijn, dan kunnen deze technieken worden toegepast. Binnen deze Poisson-modellen wordt getracht de Poisson-parameters, die van cel tot cel kunnen verschillen, te beschrijven in termen van de variabelen van de kruistabel. Het boven omschreven multiplicatieve model is hiervan een specifiek voorbeeld. De Poisson-parameter van iedere cel is daar beschreven als opgebouwd uit drie deel-parameters: een algemene (voor iedere cel identieke) parameter λ , een (voor elke cel van één rij uit de kruistabel identieke) parameter p_i en een (voor elke cel van één kolom identieke) parameter q_j . Met andere woorden: aan de uiteindelijke Poisson-parameter van iedere cel worden restricties opgelegd welke te maken hebben met de plaats in de rij en kolom van die cel in de kruistabel. Echter,

het is één keuze uit een hoeveelheid mogelijke restricties. Als we bijvoorbeeld stellen dat de wegdekstroefheid geen enkele invloed heeft op de ongevallen, dus dat alle p_i aan elkaar gelijk zijn, dan zou het model kunnen worden vereenvoudigd. Voor iedere cel geldt dan dat zijn Poisson-parameter gelijk is aan λq_j (één algemene deel-parameter en één deel-parameter voor de plaats van de cel in een kolom).

De meest algemene vorm waarin de parameter kunnen worden beschreven is:

$$\begin{aligned} \mu_{ij} &= \lambda \cdot p_i \cdot q_j \cdot r_{ij} \\ \text{of, als we de logaritmie nemen:} \\ m_{ij} (= \log \mu_{ij}) &= \alpha + \beta_i + \gamma_j + \delta_{ij} \end{aligned} \quad (5)$$

waarin de termen na het 'is-gelijk-teken' de logaritmen van de vorige expressie aanduiden.

Modellen die een zodanige beschrijving trachten te geven van de Poisson-parameters μ_{ij} worden daarom log-lineaire modellen genoemd. Een uitvoerige beschrijving is te vinden bij Goodman (1970), Haberman (1974) en Bishop, Fienberg & Holland (1975).

Het multiplicatieve ACM-model is in feite ook een log-lineair model, maar dan zonder stochastische interpretatie. Bewezen kan worden dat het model ook van toepassing is, als in plaats van de Poisson-aanname wordt gesteld dat de aantallen in de cellen van de kruistabel een multinomiale verdeling volgen en de aantallen ongevallen op zich Poisson verdeeld zijn. Het met model (4) vergelijkbare multiplicatieve model legt de extra restrictie op dat $\delta_{ij} = 0$, voor alle combinaties (i, j). De gegevens van een kruistabel kunnen altijd perfect worden beschreven met behulp van (het oververzadigde) model (5). In feite wordt dan aangenomen dat elke cel een specifieke Poisson-parameter heeft.

Getoetst kan nu bijvoorbeeld worden of het (niet-verzadigde) multiplicatieve model $m_{ij} = \alpha + \beta_i + \gamma_j$ de gegevens significant slechter beschrijft dan het model (5).

Een voorbeeld van toepassing van een dergelijk type analyse op problemen uit het verkeer (onder een overigens afwijkende modelbeschrijving) vindt men bij Rasch (1973).

4.2. Gewogen Poisson-modellen

Toepassing van log-lineaire modellen op kruistabellen waarin ongevallen zijn gegeven lijkt gerechtvaardigd: de assumptie dat de aantallen ongevallen onafhankelijk Poisson verdeeld zijn, is een door velen aanvaardbaar geachte assumptie. Als we te maken hebben met ongevalratio's in plaats van ongevallen is een dergelijke analyse niet direct toepasbaar.

De Leeuw (1975) beschrijft een algemener model, toepasbaar op Poisson verdeelde variabelen welke worden gecorrigeerd door de variabelen te delen door een constante. Met andere woorden: de Poisson verdeelde variabelen worden eerst gewogen alvorens in de analyse te worden betrokken. De ongevalratio's kunnen worden opgevat als dergelijke gewogen variabelen. Een bezwaar hiertegen is dat de voertuigkilometers in strikte zin geen corrigerende constanten zijn, maar in feite stochastische variabelen. De variantie van deze variabelen is echter vele malen kleiner dan die van de ongevalsvariabelen, zodat het bezwaar niet veel praktische betekenis zal hebben.

Een tweede bezwaar, dat voor alle log-lineaire analyses geldt, is dat het model slechts asymptotisch toetsbaar is, hetgeen erop neerkomt dat het opgaat voorzover er voor de analyse een voldoende aantal ongevallen per cel zijn verzameld. In het onderhavige geval gaat deze voorwaarde beslist niet voor iedere cel op, hetgeen de toetsing van het model bemoeilijkt.

Een uitvoerige beschrijving en een voorbeeld van toepassing van gewogen Poisson-modellen is te vinden in de De Leeuw & Oppe (1976).

Resultaten*

In een eerste analyse van de gegevens van wegtype I zijn de gegevens voor de stroefheidsklassen 2 t/m 7 en intensiteitsklassen

*Bij de analyse van de gegevens is gebruik gemaakt van het computerprogramma WPM, geschreven in PLI.

1 t/m 16 en is de som van de klassen 17 t/m 20 als klasse toegevoegd. De resultaten van deze analyse zijn samen met die van de ACM-analyse over de log-data afgebeeld in Figuur 5 en 6.

In een tweede analyse zijn telkens twee intensiteitsklassen samengevoegd, zodat nu acht intensiteitsklassen en de restklasse werden bekeken. Ook het resultaat van deze analyse is, voor zover het de parameterschattingen voor de intensiteitsklassen betreft, in Figuur 5 afgebeeld. De overeenkomst tussen de ACM-log-oplossing en WPM-oplossing is over het algemeen groot. Uit de oplossing van de verdubbelde intensiteitsklassen is te zien dat instabiliteit van de curven sterk is gereduceerd, zodat het verloop van de maat voor ongevalsvatbaarheid over de intensiteitsklassen beter is te volgen. Uit de gegevens van Tabel 5 blijkt uit de grootte van de Chi-kwadraatwaarde dat vooral de stroefheid van het wegdek bepalend is voor het verschil in ongevalsvatbaarheid ($\chi^2 = 373.40$, $df = 5$).

Echter ook het verschil in intensiteit draagt in ruime mate bij ($\chi^2 = 72.323$, $df = 16$).

Er is geen significante interactie ($\chi^2 = 83.51$, $df = 80$) in de eerste analyse, daarentegen een lichte significante interactie als de intensiteitsklassen worden gecombineerd ($\chi^2 = 56.48$, $df = 40$). Dit alles pleit sterk voor aanvaarding van het multiplicatieve model als zodanig en verder voor het weglaten van de interactieterm.

Dit laatste betekent in feite dat het verband tussen ongevallenratio's en wegdekstroefheden dezelfde is voor elk der uurintensiteitsklassen en dat er slechts een niveau-verschil bestaat tussen de ongevallenratio's voor de intensiteitsklassen. In termen van maatregelen betekent dit dat overal op wegtype I dezelfde norm kan worden gehanteerd. De effectiviteit verschilt natuurlijk wel afhankelijk van de hoeveelheid verkeer.

Voor wegtype II zijn de stroefheidsklassen 1 t/m 7 geanalyseerd. Voor de intensiteitsklassen zijn de waarden van klasse 1 t/m 10 gebruikt en de waarden van de 11^e t/m 15^e klasse bij elkaar geteld. De resultaten staan afgebeeld in Figuur 7 en 8. Ook hier

is er grote overeenstemming tussen de ACM-log en WPM-oplossingen. Uit Tabel 5 blijkt dat de grootste bijdrage aan het verschil in ongevalsvatbaarheid wordt geleverd door de stroefheid ($X^2 = 331.41$, $df = 6$) en dat daarnaast de intensiteit ook hier een zeer significante bijdrage levert ($X^2 = 120.72$, $df = 10$).

Binnen het multiplicatieve model is er bij wegtype II echter wel sprake van een zeer significante interactie ($X^2 = 191.89$, $df = 60$), zodat gesteld kan worden dat het multiplicatieve model zonder interactieterm hier minder goed past dan bij wegtype I.

In een tweede analyse zijn de stroefheidsklassen 1 en 2 en de klassen 6 en 7 samengevoegd, terwijl de intensiteitsklassen 9 en 10 bij de restcategorie zijn ingedeeld. Hierdoor werd het aantal cellen met weinig observaties sterk gereduceerd. Ook nu bleek de interactie significant te zijn ($X^2 = 142.27$, $df = 32$) zodat het niet aannemelijk is, om de verklaring voor de gevonden interactie in de te kleine aantallen observaties in de cellen te zoeken. Een verklaring zou kunnen worden gezocht in de grote diversiteit van de wegen in wegtype II zoals blijkt uit de inleiding en het feit dat de rijbanen bij dit wegtype meestal niet gescheiden zijn. Verder kunnen hier de kruispuntongevallen verstorend werken.

Samenvattend kan worden geconcludeerd:

1. Het ACM-model geeft een goede descriptie van de log-data.
2. De op de uitkomst van deze analyse gebaseerde toepassing van het WPM-model voor de gegevens van wegtype I resulteert in het handhaven van de hypothese dat de wegdekstroefheid en uurintensiteit een onafhankelijke invloed hebben op het ontstaan van ongevallen.
3. Als gevolg hiervan kan een beschrijving van de ongevallenratio's worden gegeven in termen van slechts één van beide variabelen. De praktische consequenties hiervan voor te nemen maatregelen zijn uitgewerkt in Schlösser (1977).

LITERATUUR

1. Bishop, Y.M.M. Fienberg, S.E. & Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. MIT-Press, London, 1975.
2. De Leeuw, J. (1975). Weighted Poisson models with applications to accident data. Leyden State University, Leiden, 1975.
3. De Leeuw, J. & Oppe, S. (1976). Analyse van kruistabellen: log-lineaire Poisson-modellen voor gewogen aantallen. SWOV, Voorburg, 1976.
4. Goodman, L.A. (1970). The Multivariate Analysis of Qualitative Data. Interactions Among Multiple Classifications, J.A.S.A., 1970.
5. Haberman, S.J. (1974). The analysis of frequency data. University of Chicago Press, London, 1974.
6. Kruskal, J.B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. J. of the R., Sta. Soc., Serie B, 27, 1965.
7. Rasch, G. (1973). Two applications of the multiplicative Poisson models in road accidents statistics. In: Proc. of the 38th session of the ISI, Wien, 1973.
8. Schlösser, L.H.M. (1977). Traffic accidents and road surface skidding resistance. In: Proceedings Second International Skid Prevention Conference, Ohio, 1977.

F \ R	2	3	4	5	6	7	8
01		3.50	5.00	34.50	52.00	7.50	
02	3.00	1.00	6.50	23.75	33.25	8.75	.25
03	2.00		5.50	38.75	55.00	16.00	1.25
04	4.00	4.50	8.50	52.00	56.00	21.50	
05	3.00	2.00	6.25	58.75	67.50	11.50	
06	3.50	1.50	6.50	47.75	52.00	11.75	
07	15.00	2.50	8.50	52.50	81.25	10.25	
08	13.50	3.75	15.25	61.00	77.25	15.75	
09	8.00	2.50	13.25	81.50	66.50	7.75	
10	4.50	5.25	14.00	70.75	66.25	9.75	
11	4.00	3.75	18.50	85.75	85.50	7.00	1.00
12	2.50	1.50	26.25	66.75	58.25	7.25	1.00
13		.25	23.00	54.50	43.25	8.50	
14	3.50	1.50	11.75	59.50	34.50	7.25	
15		1.50	10.50	44.25	23.25	2.00	
16		.75	12.75	35.25	21.25	1.00	
17		.75	6.25	25.75	5.75		
18			6.75	27.75	3.00		
19			2.25	12.25	6.50		
20		.50	10.25	48.50	11.25	1.00	

Tabel 1A. Verdeling van het aantal ongevallen voor wegtype I naar de kenmerken wegdekstroefheid (R) en uurintensiteit (F). De gebroken getallen zijn het gevolg van opdeling van ongevallen over klassen als de klasse niet exact kan worden aangeduid.

F \ R	2	3	4	5	6	7	8
01	79	140	488	4145	9598	4204	610
02	116	165	711	6219	13662	5457	485
03	228	238	1273	11675	25830	11525	877
04	361	400	1699	15613	28172	11457	493
05	193	594	1610	14431	29096	10183	223
06	239	563	1586	14166	31019	8579	195
07	442	626	1539	13952	31060	7508	178
08	492	579	1729	14699	30537	6276	128
09	386	470	1585	14109	25306	4942	100
10	264	475	1518	12865	22231	3524	78
11	239	404	1715	11794	19160	2485	50
12	172	234	1667	10480	15703	2238	32
13	77	153	1029	8085	11093	1483	22
14	78	102	863	6132	8001	926	13
15	40	67	570	4453	5502	551	5
16	6	51	475	3129	3669	442	
17	2	45	495	2362	2782	284	
18		26	379	1702	1798	145	
19		19	236	1513	1454	115	
20	4	26	875	4148	3373	233	

Tabel 1B. Indeling van het aantal voertuigkilometers naar de kenmerken wegdekstroefheid (R) en uurintensiteit (F) voor wegtype I.

F \ R	1	2	3	4	5	6	7	8
01	8.00	24.00	20.00	49.00	189.00	369.50	93.50	6.00
02	14.00	50.00	57.00	92.50	290.25	487.75	130.00	9.00
03	16.00	35.00	40.00	78.50	323.75	439.25	83.00	1.00
04	11.00	21.00	38.00	70.00	309.50	357.50	43.00	1.00
05	4.00	21.00	33.00	63.50	197.00	168.00	19.00	
06	4.00	12.00	29.00	47.00	163.50	116.00	13.00	
07	1.00	9.00	13.00	36.00	83.00	58.00	6.00	2.00
08		7.00	4.00	13.00	66.00	41.00	3.00	
09	3.00	1.00	6.00	11.00	39.50	31.00	1.00	
10	2.00		1.00	4.00	29.00	17.00	1.00	
11	2.00			4.00	17.50	32.00		
12	2.00	1.00		6.00	13.00	8.00		
13	1.00	1.00		1.00	8.00	3.00		
14					3.00	1.00		
15						4.00		

Tabel 2A. Indeling van het aantal ongevallen voor wegtype II naar de kenmerken wegdekstroefheid (R) en uurintensiteit (F). De gebroken getallen zijn het gevolg van opdeling van ongevallen over klassen als de klasse niet exact kon worden aangegeven.

F \ R	1	2	3	4	5	6	7	8
01	620	2198	2919	4866	24512	60184	21175	1883
02	860	3470	5521	12544	53617	112993	36351	1583
03	781	2425	3886	11427	54940	86961	20241	583
04	387	1192	2543	7840	45977	55499	9680	491
05	145	589	2157	6800	28100	30695	4691	246
06	58	234	1182	4566	16184	19079	2570	129
07	30	53	436	2392	9722	12643	1676	46
08	38	54	294	1424	5525	7609	945	11
09	15	16	113	708	3070	4724	469	8
10	19	8	57	302	1946	3361	430	6
11	20	11	36	155	1594	2765	299	2
12	18	8	17	125	1109	1706	171	
13	7	3	7	66	337	622	47	
14		4	4	16	83	135	18	
15		2	3	1	182	301	12	

Tabel 2B. Indeling van het aantal voertuigkilometers naar de kenmerken wegdekstroefheid (R) en uurintensiteit (F) voor wegtype II.

	S_d	S_{log}	S_{mon}
wegtype I	.1849	.1334	.1154
wegtype II	.2355	.1230	.0789

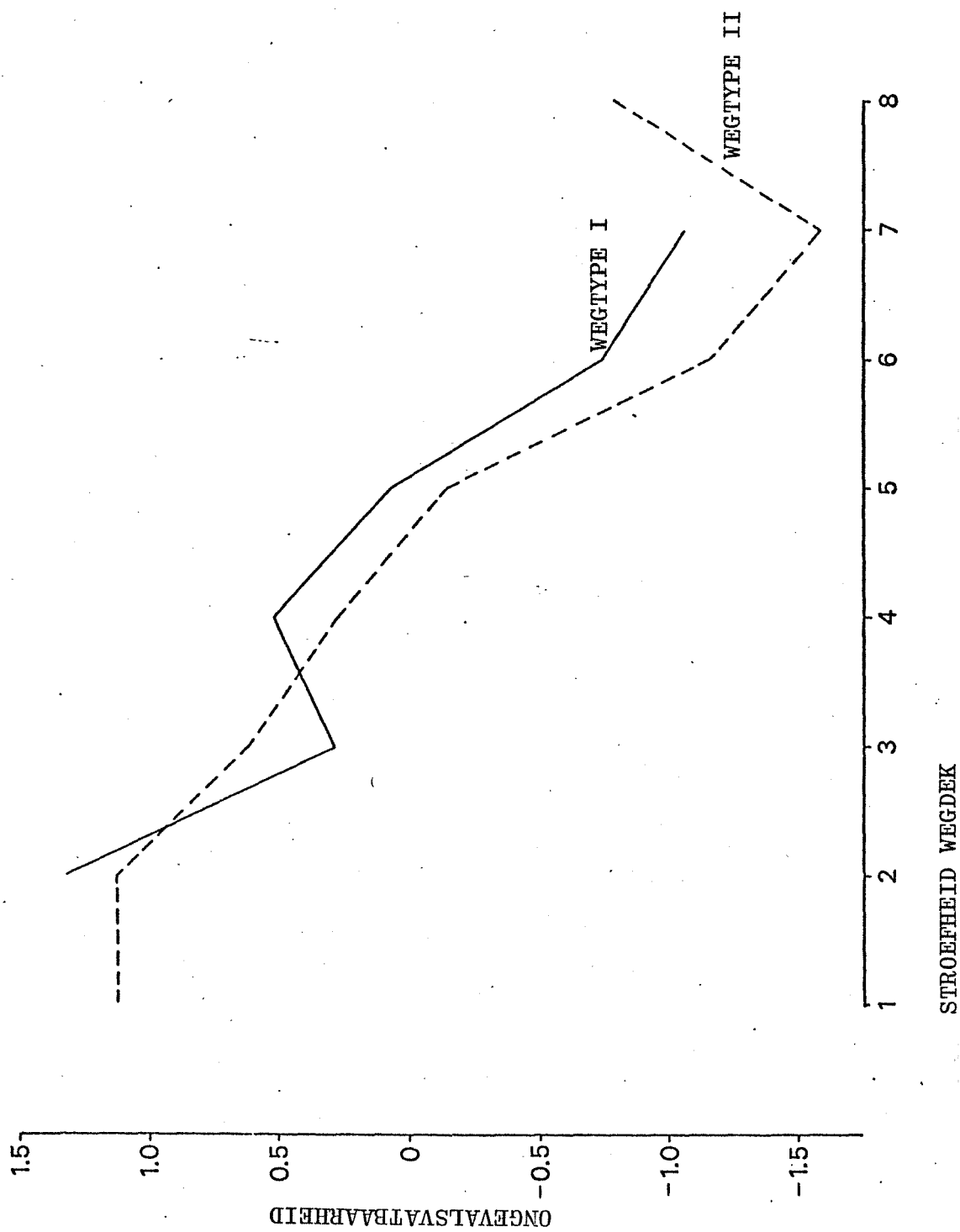
Tabel 3. Stresswaarden voor de oplossing over de data, de log-data en de uiteindelijke ACM-oplossing voor wegtype I en II.

	S_d	S_{log}	S_{mon}	$S_d - S_{log}$	$S_{log} - S_{mon}$
gemiddelde	.6939	.6928	.6325	.0011	.0603
s.d.	.054	.058	.054	.038	.038

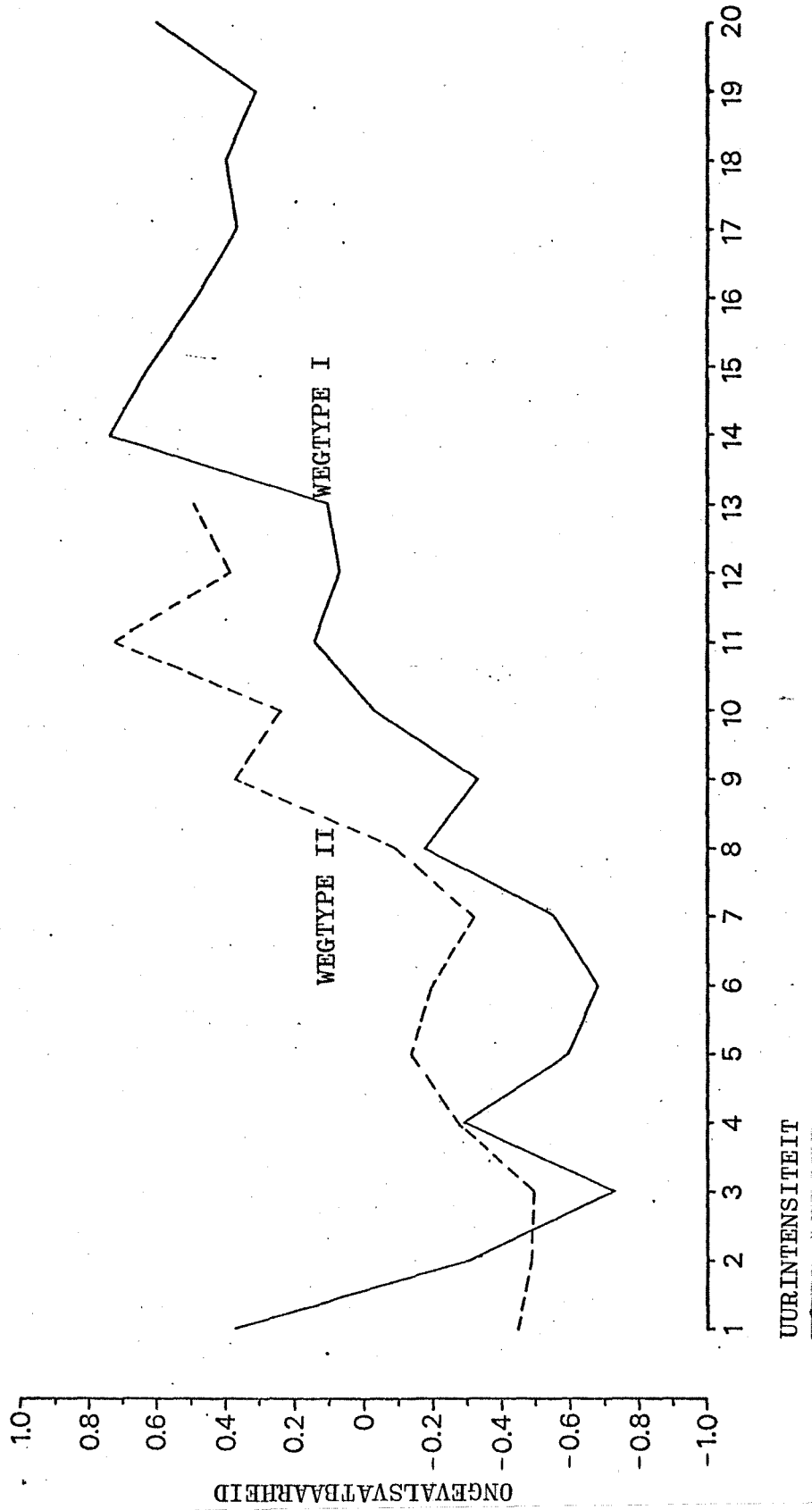
Tabel 4. Gemiddelde stresswaarden voor de oplossingen van de Monte Carlo data en de erbij behorende standaardafwijkingen voor de data van wegtype I. Aantal data sets is 40.

effect	X^2	DF	$X^2_{.95}$
<u>Wegtype I, volledige set:</u>			
F	72.3155	16	26.29
R	373.4048	5	11.07
F x R	83.5095 NS	80	101.88
<u>Wegtype I, opgeteld:</u>			
F	51.80	8	15.51
R	377.33	5	11.07
F x R	56.48	40	55.76
<u>Wegtype II, volledige set:</u>			
F	120.72	10	18.31
R	331.41	6	12.59
F x R	191.89	60	79.08
<u>Wegtype II, opgeteld:</u>			
F	215.92	8	15.51
R	652.44	4	9.49
F x R	142.27	32	46.19

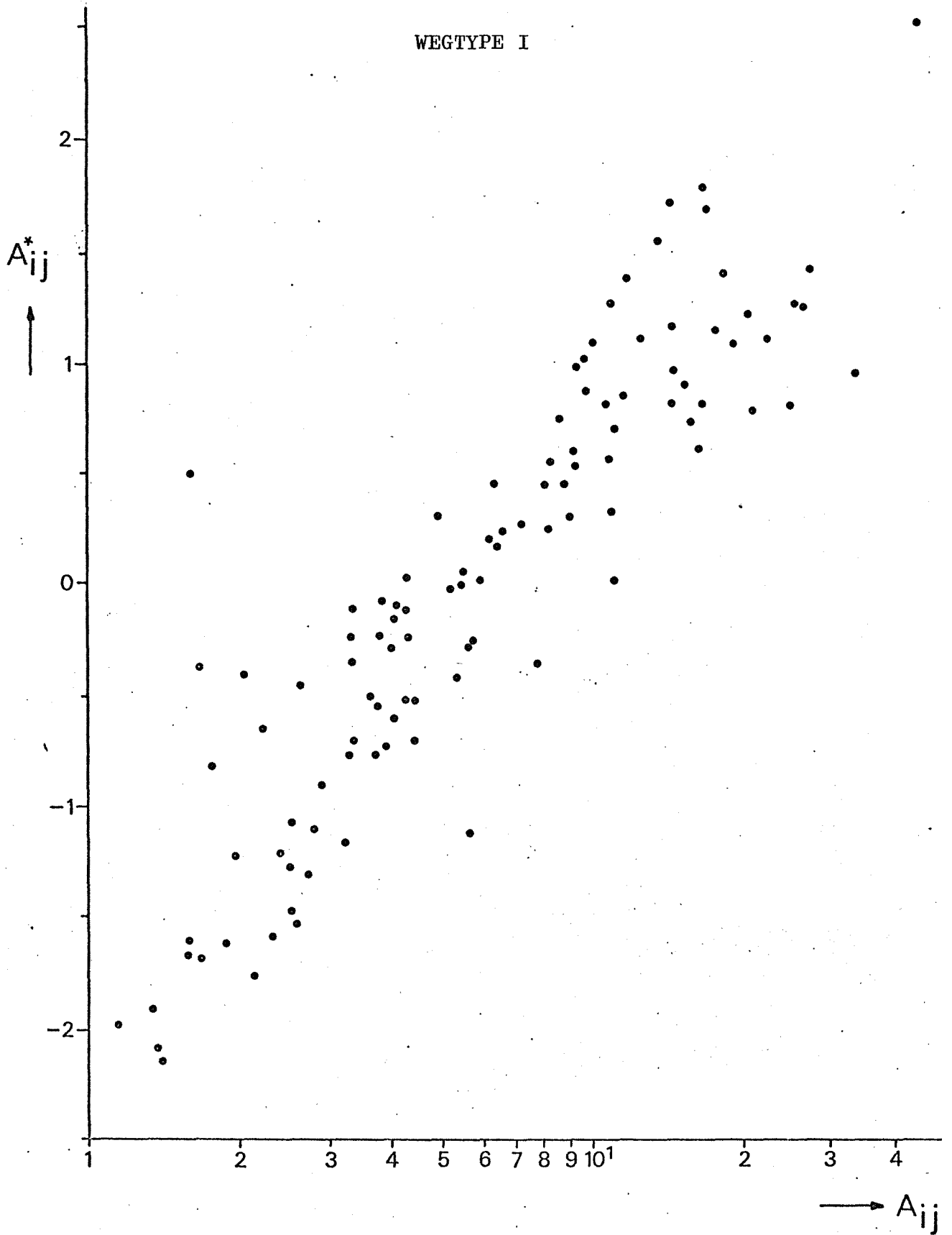
Tabel 5. Resultaten van de vier uitgevoerde WPM-analyses. Onder 'effect' staat de bron aangeduid. Daarnaast staan chi-kwadraat waarden (X^2) met de bijbehorende vrijheidsgraden (df) en de bij het 5%-niveau behorende grenswaarden voor chi-kwadraat ($X^2_{.95}$).



Figuur 1. ACM_{mon}-oplossing voor de stroefheidsklassen van wegtype I en II.

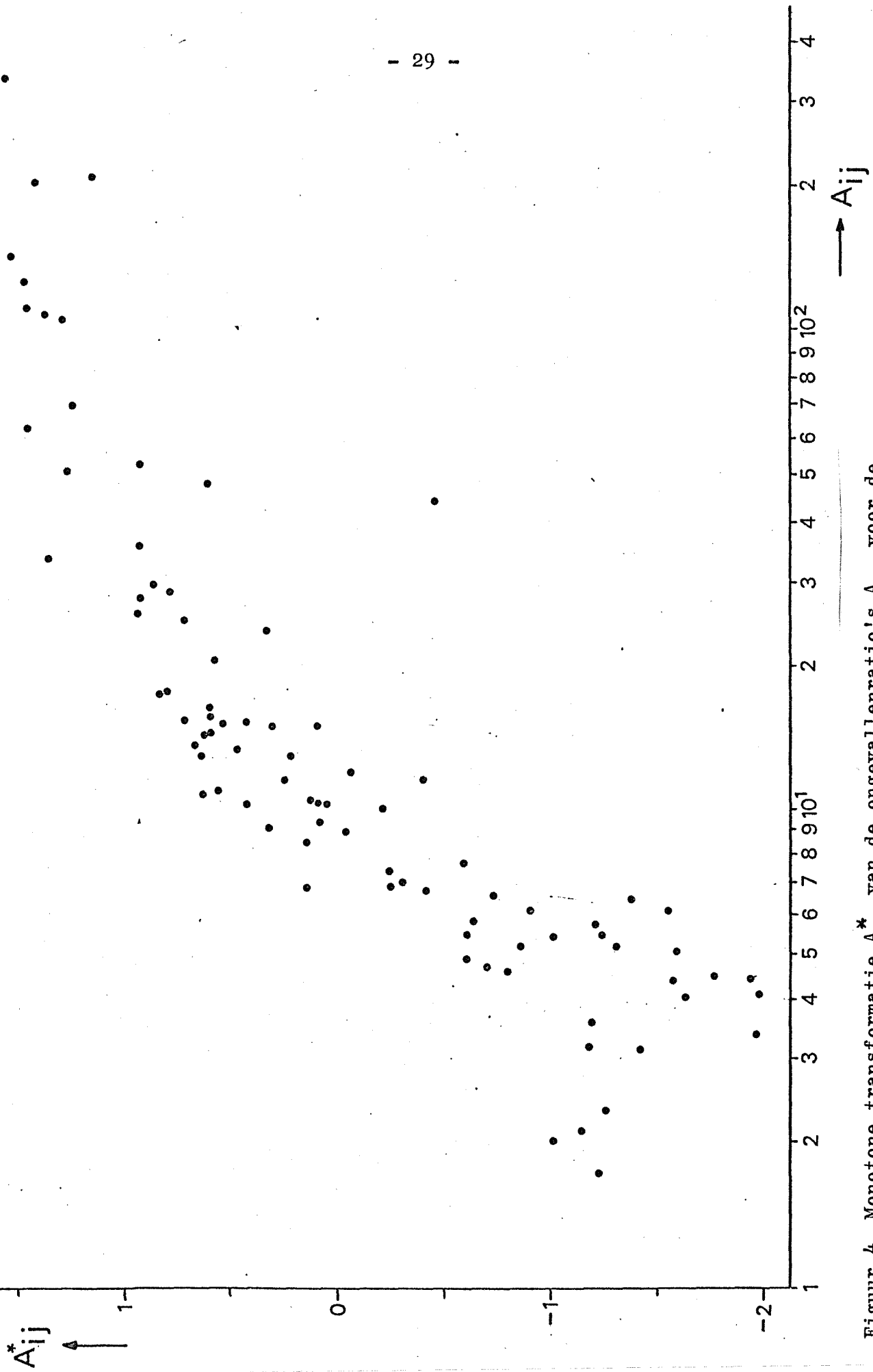


Figuur 2. ACM_{mon}-oplossing voor de uurintensiteitsklassen van wegtype I en II.



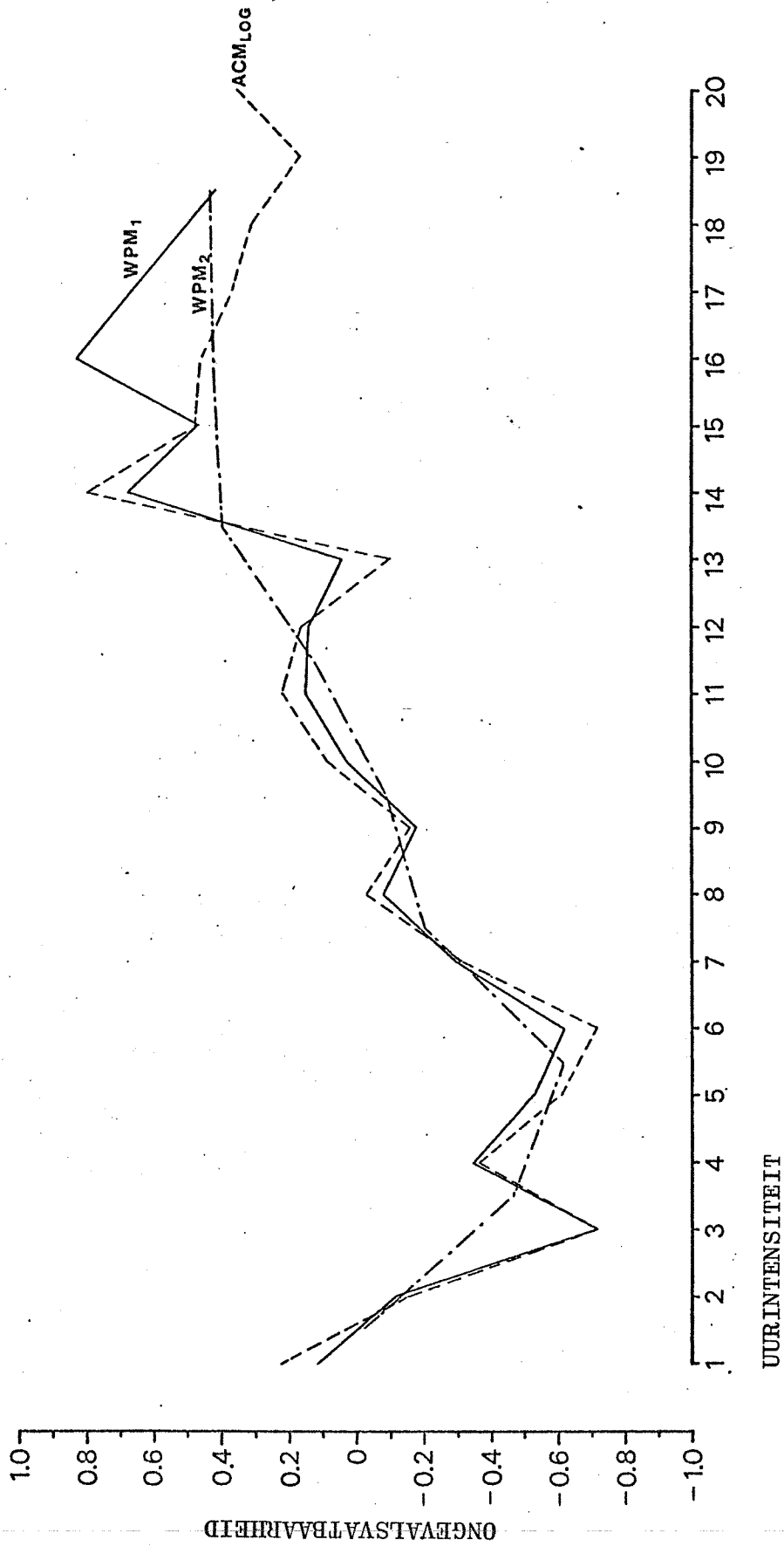
Figuur 3. Monotone transformatie A_{ij}^* van de ongevalleeratio's A_{ij} voor de data van wegtype I behorend bij de ACM-analyse.

WEGTYPE II



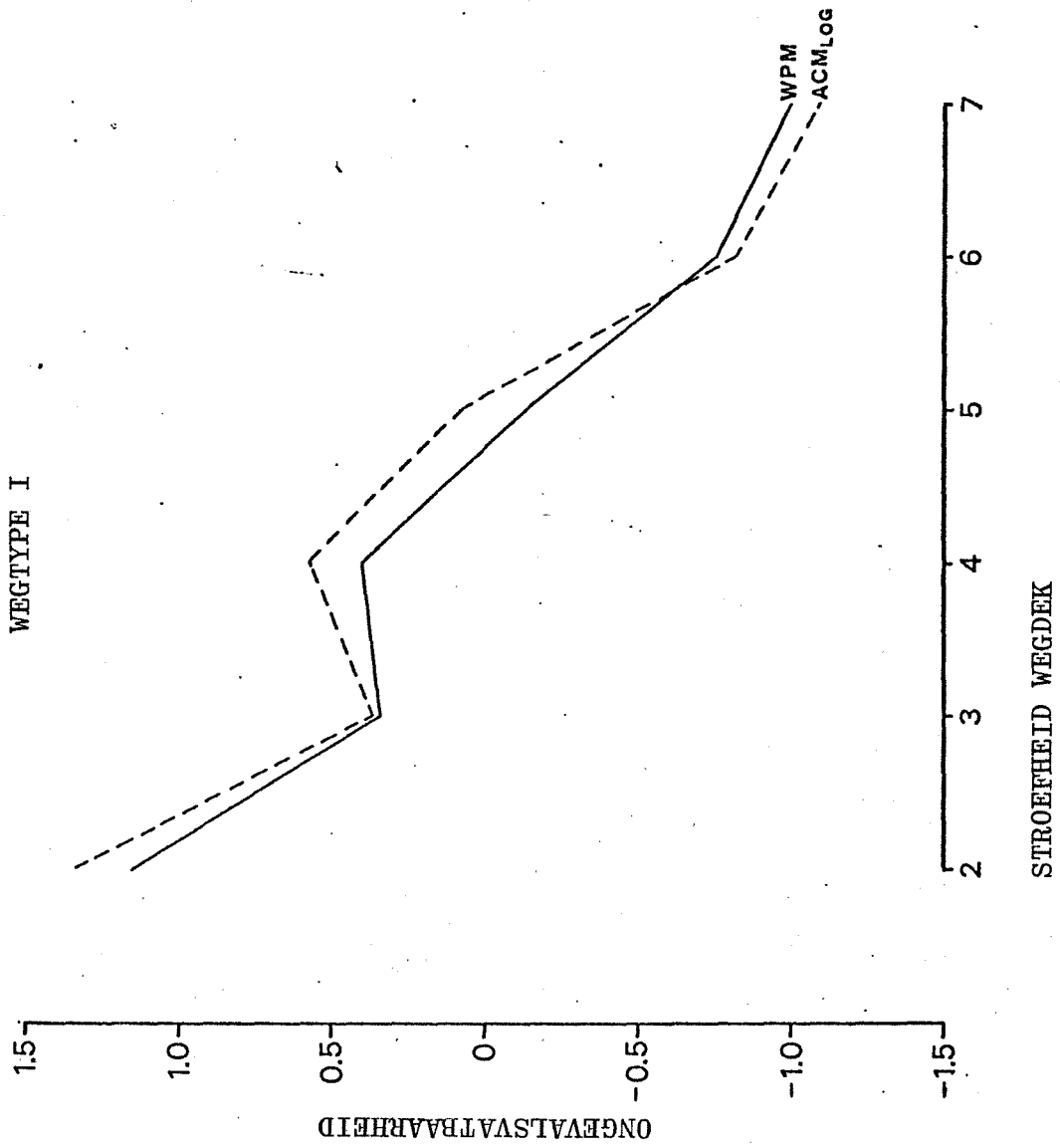
Figuur 4. Monotone transformatie A_{ij}^* van de ongevalratio's A_{ij} voor de data van wegtype II behorend bij de ACM-analyse.

WEGTYPE I



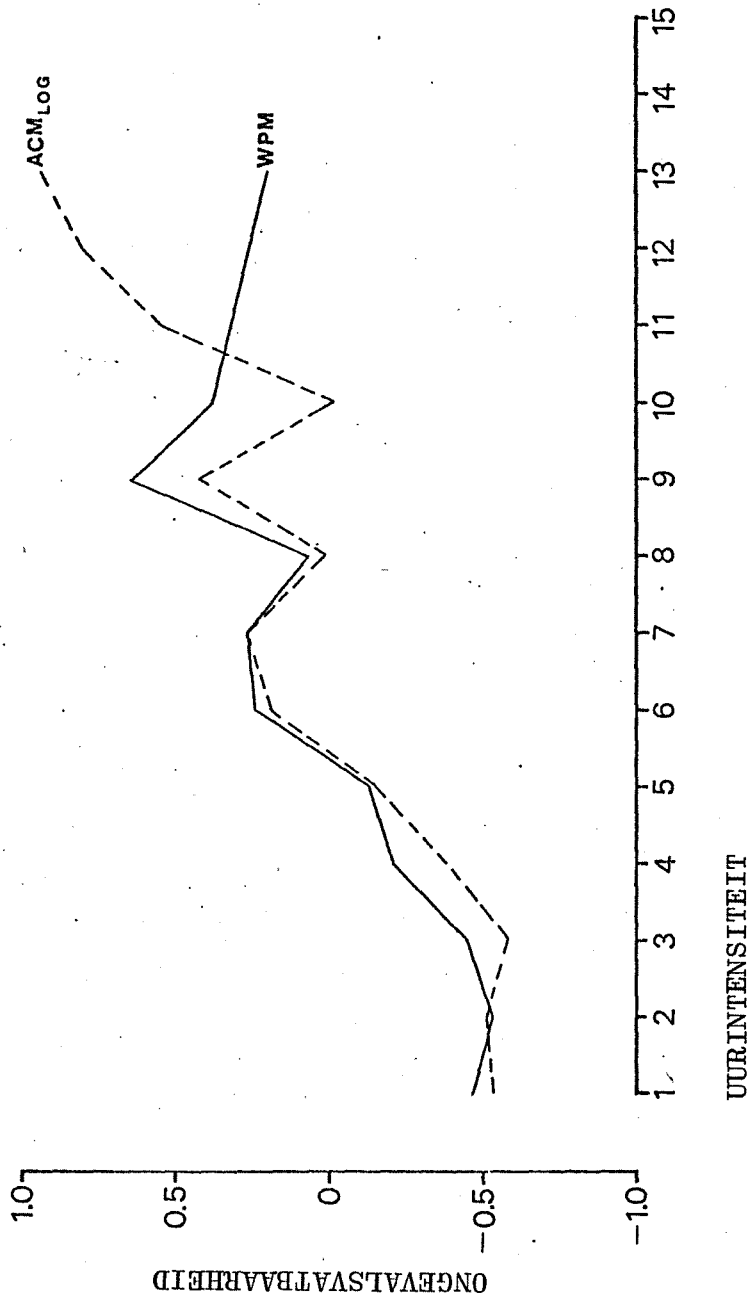
Figuur 5. ACM_{log}- en WPM-oplossing voor de uurintensiteitsklassen van wegtype I.

WPM₁ correspondeert met de volledige dataset, WPM₂ met de gesommeerde dataset.

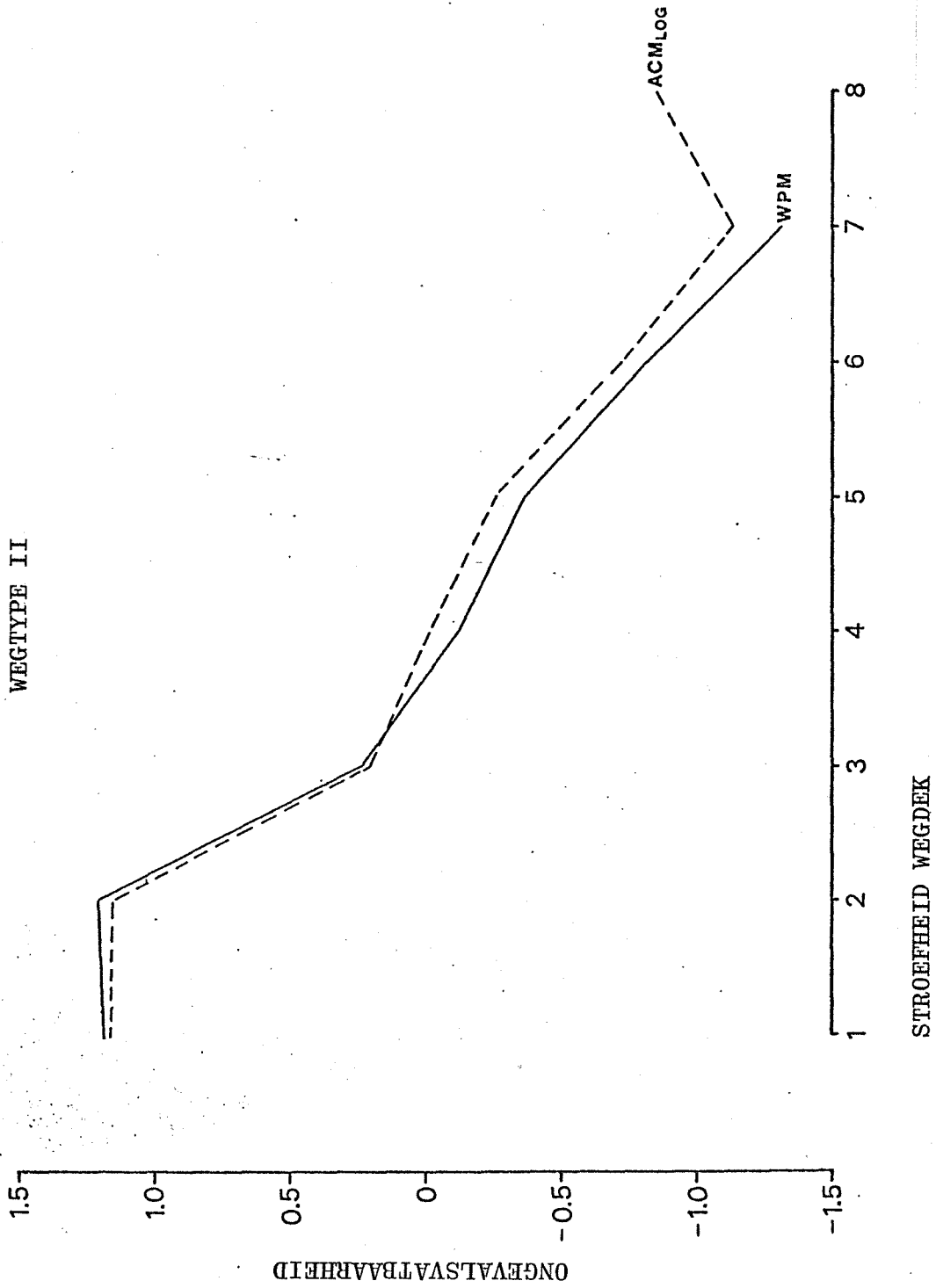


Figuur 6. ACM_{log} - en WPM-oplossing voor de stroefheidsklassen van wegtype I.

WEGTYPE II



Figuur 7. ACM_{log}- en WPM-oplossing voor de uurintensiteitsklassen van wegtype II.



Figuur 8. ACM_{log} - en WPM-oplossing voor de stroefheidsklassen van wegtype II.