

**KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT**

Verslagen
V-307

S. Kruizinga

Logistische discriminant analyse

De Bilt, 1978

Publikationsnummer: K. N. M. I. V-307 (MO)

U. D. C. : 551.509.314 :
551.509.323

Logistische Discriminant Analyse.

S. Kruizinga

Inleiding

Naast multiple lineaire regressie vormt discriminant analyse een belangrijke weervertaaltechniek. De lineaire regressie wordt meestal gebruikt voor de vertaling naar continue grootheden zoals temperatuur en de discriminant analyse wordt veelal gebruikt voor de vertaling naar ja/nee verschijnselen zoals onweer/geen onweer. Ook vertaling naar grootheden met meerdere klassen is mogelijk. De werkwijze van de discriminant analyse is echter niet zo gestandaardiseerd als van de multiple regressie vandaar dat deze techniek vaak niet als zodanig wordt herkend. Voorbeelden van toepassingen op het KNMI zijn de regemethode van Hanssen en Kuipers [1] en de overgangskansen van Van den Dool en Nap [2].

Bij de hiervoor genoemde voorbeelden wordt veelal een grafische en heuristische of intuïtieve werkwijze gevolgd. Ook andere werkwijzen zijn mogelijk; bij de lineaire discriminant analyse wordt van een lineaire functie van de ingangsvariabelen gebruik gemaakt om tot een beslissing te komen omtrent de te voorspellen variabele (zie Hoel [3]). In sommige gevallen gaat men niet rechtstreeks op de ja/nee beslissing af doch wordt getracht de kans van optreden te specificeren in afhankelijkheid van de ingangsvariabelen. Op basis van deze kansen kunnen dan verdere beslissingen worden genomen. Een groot voordeel van de twee laatst genoemde werkwijzen is dat het meestal eenvoudiger is om meerdere ingangsvariabelen te combineren dan bij de eerstgenoemde voorbeelden. In dit verslag zal nader worden ingegaan op een methode welke tot de laatstgenoemde groep behoort, de logistische discriminant analyse. Bij deze techniek wordt het kansmodel gebaseerd op de zogenaamde logistische functie. In dit verslag zal kort op de theorie worden ingegaan en aansluitend zullen enige voorbeelden van de toepassing van deze techniek worden gegeven. Voor meer uitgebreide theorie wordt verwezen naar Anderson [4, 5].

2. Beschrijving van het probleem en het kansmodel.

Stel dat y een variabele is welke slechts K elkaar uitsluitende mogelijke uitkomsten heeft welke zijn genummerd van 1 t/m K . Deze uitkomsten kunnen van natuurlijk discrete aard zijn, bijvoorbeeld de uitkomst van een worp met een dobbelsteen ($K=6$) of regen/geen regen ($K=2$) of van kunstmatig discrete aard, bijvoorbeeld de temperatuur is lager dan 3°C of hoger dan 5°C of daar tussen in ($K=3$).

Verder zijn er nog M variabelen x_1 t/m x_M waarvan de uitkomsten zowel continu verdeeld als discreet verdeeld mogen zijn. De vragen die beantwoord dienen te worden zijn:

- a. Is er een relatie tussen de momentane waarden der x -en en de kans van optreden van de diverse uitkomsten van y ?
- b. Kan deze relatie in een mathematisch model worden vastgelegd?

De hierna te beschrijven methode zal deze vragen niet in hun algemene vorm beantwoorden. Wat in feite wordt gedaan is dat een bepaald mathematisch model wordt verondersteld waarvan daarna wordt onderzocht of het in voldoende mate bij de gegevens past en aanleiding geeft tot variatie in de kansen. Zo ja dan kunnen de vragen a. en b. met ja worden beantwoord. Is dit niet het geval dan kan alleen gesteld worden dat het met het gekozen model in ieder geval niet lukt.

Bij de logistische discriminant analyse wordt het zogenaamde logistische kans-model (Anderson [4, 5]) als basis gekozen. Dit kansmodel beschrijft de relatie tussen de predictoren x_m en de kansverdeling van y op de volgende manier:

$$P(y=k) = \frac{\exp(f_k)}{1 + \sum_{k=1}^{K-1} \exp(f_k)} \quad (k=1, K-1) \tag{2.1}$$

$$P(y=K) = 1 - \sum_{k=1}^{K-1} P(y=k)$$

Hierin zijn de f_k ; (K-1) functies van de variabelen $x_1, x_2 \dots x_m$ die nog nader omschreven dienen te worden. Het voordeel van een formulering volgens (2.1) is dat voor willekeurige waarden van de functies f_k de kansen $P(y=k)$ in het interval $[0,1]$ blijven zodat een realistisch model ontstaat. Voor de functies f_k en de constanten daarin hoeven dus geen restricties te worden ontworpen. Voor deze functies f_k wordt dan nog de volgende vorm gekozen:

$$f_k = \partial_{k,0} + \sum_{m=1}^M \partial_{k,m} \cdot x_m \quad (2.2)$$

De coëfficiënten $\partial_{k,m}$ ($k=1, K-1; m=0 \dots M$) zijn nog onbepaald en worden aangepast op het specifieke probleem. Het totaal aantal aan te passen parameters is dus $(K-1) * (M+1)$.

Indien er feitelijk geen samenhang is tussen y en x_1 t/m x_M dan is het model toch toepasbaar. In dat geval zullen de coëfficiënten $\partial_{k,m}$ zo gekozen worden dat de formules (2.1) voor alle waarden van $x_1 \dots x_M$ de klimatologische kansen opleveren. Dat wil zeggen alle $\partial_{k,1}$ t/m M worden zeer klein (in theorie nul).

Onder de variabelen x_m mogen zich nog dummy variabelen bevinden welke functies zijn van andere, al dan niet nieuwe, variabelen bijvoorbeeld:

$$x_2 = (x_1)^2 \quad \text{of} \quad x_3 = \arctg(T_x)$$

In deze functie mogen geen onbekende, nog aan te passen coëfficiënten voorkomen.

3. Aanpassen van de coëfficiënten.

Het model zoals beschreven in de voorgaande paragraaf dient op het specifieke probleem te worden aangepast door een gunstige keuze van de coëfficiënten. Voor het aanpassen van de coëfficiënten zal een steekproef van waargenomen x -waarden met bijbehorende opgetreden y -waarden als basis dienen. Dat wil zeggen dat de coëfficiënten θ zodanig worden gekozen dat het model zo goed mogelijk bij de steekproef past. Voor dit al of niet passen zal een norm dienen te worden vastgesteld.

Voor deze aanpassingsnorm wordt de voorwaardelijke likelihood L van de steekproef gebruikt. In het volgende voorbeeld zal dit begrip nader worden beschreven. Gegeven zij een steekproef van 4 dagen met waarden voor x_1 , x_2 en x_3 zoals gegeven in tabel 3.1. Tevens is de opgetreden klasse van y gegeven.

Tabel 3.1

x_1	x_2	x_3	y	$P(y=1)$	$P(y=2)$	$P(y=3)$
1	0	0	2	0,4	0,3	0,3
2	0	1	1	0,6	0,2	0,2
3	1	0	1	0,7	0,2	0,1
1	2	3	3	0,1	0,2	0,7

Op basis van een gegeven set coëfficiënten en het model zijn de kansen (in kolom 5 t/m 7) berekend behorend bij de gegeven x -waarden. Hieruit kan geconcludeerd worden dat voor de eerste dag de voorwaardelijke (bij gegeven x -waarden) kans op $y=2$ gelijk is aan 0,3. Voor de tweede dag is de voorwaardelijke kans op $y=1$ gelijk aan 0,6. De totale voorwaardelijke kans P_{tot} op de rij y -waarden in de steekproef met de gegeven x -waarden is dan:

$$P_{\text{tot}} = 0,3 \times 0,6 \times 0,7 \times 0,7 = 0,0882 \quad (3.1)$$

Deze waarde voor P_{tot} wordt dus gevonden voor een gegeven set coëfficiënten. Stel nu dat een andere set coëfficiënten een $P_{\text{tot}} = 0,200$ opleverd dan is het logisch om de tweede set van coëfficiënten een meer aannemelijke keuze te noemen dan de eerste set. De aangepaste coëfficiënten worden nu zo gekozen dat de voorwaardelijke kans P_{tot} maximaal is en worden maximaal aannemelijk (maximum likelihood) coëfficiënten genoemd. Hoe men te werk gaat bij het zoeken van deze maximaal aannemelijke coëfficiënten is een wiskundig probleem dat voor het model niet ter zake doet.

In het algemeen wordt niet direct met de kansen gewerkt maar met de natuurlijke logaritme ervan. Dit heeft als voordeel dat de berekening van (3.1) kan worden omgeschreven tot een som. Aangezien deze logaritmen allen negatief zijn wordt hier bovendien nog een minteken toegevoegd. De log likelihood LNL van de steekproef wordt dan:

$$\text{LNL} = (-\ln (P_1)) + (-\ln (P_2)) + (-\ln (P_3)) + (-\ln (P_4)) \quad (3.2)$$

Het maximum van P_{tot} valt samen met een minimum van LNL.

4. Toetsing van het model en verificatie van kansverwachtingen.

Nadat de coëfficiënten zijn aangepast is het van belang om na te gaan of het wel zinvol was om dit te doen en of het model goed past bij het werkelijke proces. Bij deze toetsing kunnen twee stadia worden onderscheiden en wel toetsing op afhankelijke en toetsing op onafhankelijke gegevens.

4.1 Toetsing op de afhankelijke gegevens.

In het in paragraaf 2 beschreven model wordt aangepast op een willekeurige steekproef dan zullen altijd waarden ongelijk nul gevonden worden voor de coëfficiënten. Dit is op zichzelf geen voldoende bewijs dat het model een zinvolle aanpassing geeft. Om te toetsen of het model een significante aanpassing geeft gaat men als volgt te werk. In eerste instantie worden alleen de coëfficiënten $\beta_{k,0}$ aangepast en de overige coëfficiënten worden nul gehouden. In dit geval beschrijft men dus uitsluitend de klimatologie van y in de steekproef. Het dan bereikte minimum van LNL wordt aangeduid met $LNL_{\text{ Klim}}$. Bij de volgende aanpassing worden dan alle coëfficiënten aangepast. Het dan bereikte minimum $LNL_{\text{ min}}$ dient zo laag te zijn dat geldt:

$$\Delta LNL = 2 (LNL_{\text{ Klim}} - LNL_{\text{ min}}) \geq \chi^2_{(K-1).M} (95\%) \quad (4.1)$$

wil er gesproken kunnen worden van een significante verbetering. Het uiterste rechterlid $\chi^2_{(K-1).M} (95\%)$ is een aanduiding voor het 95% percentiel van een χ^2 variabele met $(K-1).M$ vrijheidsgraden. Het voorgaande is een toepassing van de likelihood ratio test welke ook hier mag worden toegepast zoals is bewezen door Buishand.

4.2 Toetsing op onafhankelijke gegevens.

In 4.1 werd de toetsing van het model op de gegevens waaruit de coëfficiënten zijn afgeleid beschreven. Een werkelijke afdoende toetsing kan eigenlijk alleen plaatsvinden door het model toe te passen op nieuwe gegevens en de berekende kansen te verifiëren op de opgetreden waarde van y . Uiteraard zonder de coëfficiënten verder aan te passen. De verificatie van de kansverwachting tegen de opgetreden waarde van y zou bijvoorbeeld kunnen geschieden met behulp van een Brierscore. Echter ook de loglikelihood beschreven in paragraaf 3 is bruikbaar als verificatienorm. Deze verificatienorm zal ongeveer dezelfde eigenschappen hebben als de Brierscore echter grote missers worden zwaarder bestraft. Om te normeren voor de grootte van de steekproef zal bij de verificatie gewerkt worden met de gemiddelde loglikelihood LNL:

$$\overline{\text{LNL}} = \frac{1}{N} \sum_{n=1}^N (-\ln P'_n) \quad (4.2)$$

waarin P'_n : de voorspelde kans van de opgetreden waarde van y en N : het aantal steekproefelementen.

Deze LNL dient men dan te vergelijken met:

$$\overline{\text{LNL}}_{\text{KLIM}} = \frac{1}{N} \sum_{n=1}^N (-\ln P'_{n,\text{KLIM}}) \quad (4.3)$$

waarin $P'_{n,\text{KLIM}}$: de klimatologische kans van het opgetreden vak van y op dag n . Indien het model een zinvolle aanpassing geeft dan dient $\overline{\text{LNL}} < \overline{\text{LNL}}_{\text{KLIM}}$. Bij slechte aanpassing zal $\overline{\text{LNL}}$ groter worden dan $\overline{\text{LNL}}_{\text{KLIM}}$.

5. Toepassingen.

In deze paragraaf zal het hiervoor beschreven model worden toegepast op een aantal praktische situaties. In het eerste voorbeeld zal het onderzoek van Van den Dool en Nap [2] ook met behulp van dit model worden uitgevoerd. In het tweede voorbeeld zal getracht worden een relatie te leggen tussen het stromingspatroon op 500 mbar en de kans op neerslag te De Bilt in de periode 06-18 UT.

5.1 Voorspelling maandgemiddelde temperatuur.

In hun onderzoek onderscheiden Van den Dool en Nap drie klassen voor de maandgemiddelde temperatuur T_G . De grenzen van deze klassen zijn vermeld in tabel 5.1. De klassen worden aangeduid met B(elow), N(ormal) en A(bove). Daarna wordt getracht om de kans op ieder van die drie klassen te specificeren in afhankelijkheid van de maandgemiddelde temperatuur van de voorafgaande maand T_{G-1} .

In dit onderzoek zal zowel worden gezocht naar een relatie tussen T_G en T_{G-1} als naar een relatie tussen T_G , T_{G-1} en T_{Z-1} , de gemiddelde zeetemperatuur van de voorafgaande maand (Lichtschip Texel).

Deze relaties werden per kalendermaand onderzocht waarbij T_{G-1} steeds werd genoteerd als ΔT_{G-1} , de afwijking van de normaal.

De gebruikte normalen zijn eveneens vermeld in tabel 5.1. De basisgegevens werden beschikbaar gesteld door Van den Dool.

Er zijn dus twee modellen onderzocht welke uitsluitend verschillen in de functies f_k .

MODEL I

Als predictor werd hier ΔT_{G-1} gebruikt. Daar de kromme lijnen in figuur 4 van Van den Dool en Nap echter duiden op niet-lineair gedrag werd ook $(\Delta T_{G-1})^2$ ingevoerd als parameter. Daardoor ontstonden de volgende functies f_k :

$$f_B = \partial_{B,0} + \partial_{B,1} \Delta T_{G-1} + \partial_{B,2} (\Delta T_{G-1})^2 \quad (5.1)$$

$$f_N = \partial_{N,0} + \partial_{N,1} \Delta T_{G-1} + \partial_{N,2} (\Delta T_{G-1})^2$$

MODEL II

De predictoren van model II waren ΔT_{G-1} en T_{z-1} en dus de gebruikte functies:

$$f_B = a_{B,0} + a_{B,1} \Delta T_{G-1} + a_{B,2} T_{z-1} \quad (5.2)$$

$$f_N = a_{N,0} + a_{N,1} \Delta T_{G-1} + a_{N,2} T_{z-1}$$

Voor beide modellen dienden dus 6 parameters te worden aangepast. Voor model I werden de gegevens van de jaren 1882 t/m 1960 gebruikt voor de aanpassing van de coëfficiënten. Bij model II gegevens uit de jaren 1891 t/m 1960. De verificatie op onafhankelijke data vond bij beide plaats op gegevens uit de jaren 1961 t/m 1975. In tabel 5.1 is tevens nog de klimatologie van de jaren 1882 t/m 1960 van T_G vermeld. Deze klimatologie diende als basis voor de verificatie. De resultaten van beide modellen zijn vastgelegd in tabel 5.2. De waarden van de aangepaste coëfficiënten zijn vrij onbelangrijk en hier niet gegeven. Tabel 5.2 bevat uitsluitend toetsingsresultaten van de modellen. Daar bij beide modellen evenveel coëfficiënten worden aangepast ligt de kritieke grens voor loglikelihood verbetering bij beide modellen op het zelfde niveau. Uit de standaard statistische tabellen blijkt dat deze grens 9,5 is. Zoals te zien voldoet model I daar slechts aan voor vier maanden en model II slechts voor twee maanden. In het tweede deel van de tabel zijn de resultaten van de toetsing op onafhankelijk materiaal gegeven. Deze resultaten bevestigen de voorgaande conclusies over het algemeen.

In het geval van model I zijn de twee ingangsvariabelen rechtstreeks gekoppeld zodat in dit geval de berekende modelkansen kunnen worden geplotted als functie van ΔT_{G-1} . Deze plots zijn gegeven in de figuren 5.1 en 5.2. In ieder blokje wordt de totale kans van 100% (hoogte blokje) door twee gebogen lijnen verdeeld in drie delen. Het onderste deel geeft de kans op Below bij gegeven ΔT_{G-1} , het bovenste deel de kans op Above. Het tussenliggende deel geeft uiteraard de kans op Normaal. Uit deze figuren kan men afleiden dat model I en

de grafieken van Van den Dool en Nap in het algemeen in dezelfde kansen resulteren.

Tevens blijkt uit deze figuren dat voor de maanden die blijkens tabel 5.2 een significante aanpassing geven, de lineaire term van ΔT_{G-1} dominerend is zodat er sprake is van gewone persistentie. De maanden waarin de term $(\Delta T_{G-1})^2$ een duidelijke rol speelt (mei, juni) worden bij toetsing als niet significant beoordeeld.

Algemene conclusies:

1. Toevoegen van gemiddelde zeewatertemperatuur geeft geen verbetering in voorspelbaarheid van maandgemiddelde temperatuur.
2. Alleen in de zuiver persistente overgangen zijn de overgangsregels bruikbaar (februari, maart, augustus, september).
3. Voor de overige maanden mogen de overgangsregels feitelijk niet worden toegepast.

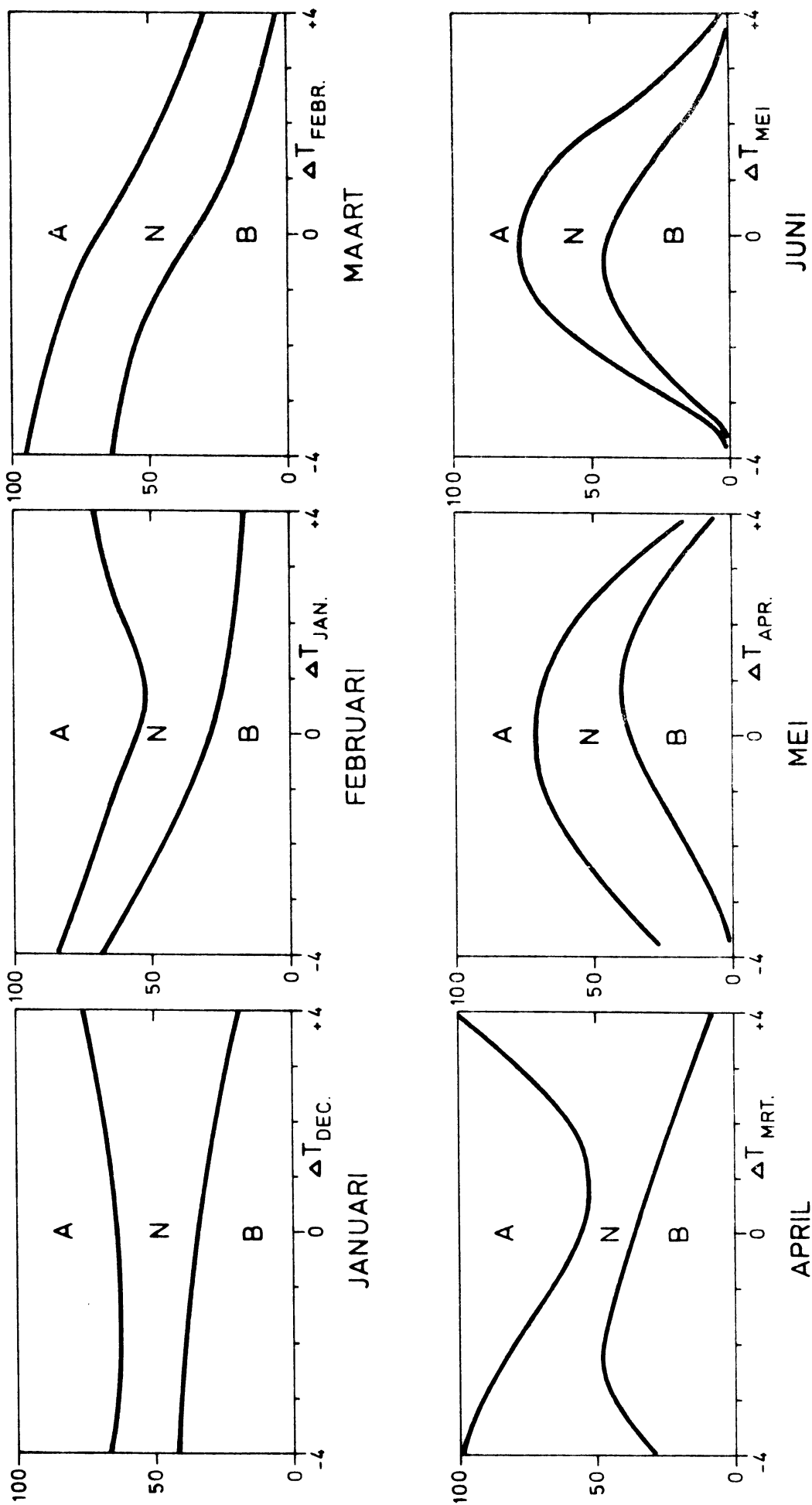
Tabel 5.1: Algemene gegevens.

Maand	klassegrenzen		Normaal	Klimatologie (%)	
	B ($T_G \leq$)	A ($T_G \geq$)		B	A
Jan.	0,7	3,1	1,8	33	34
Feb.	1,8	3,7	2,4	33	33
Maart	3,7	5,8	4,8	33	34
April	7,6	8,8	8,1	34	34
Mei	11,8	12,9	12,3	32	35
Juni	14,4	15,8	15,1	35	34
Juli	16,2	17,5	16,8	33	32
Aug.	15,7	16,8	16,4	30	35
Sep.	13,4	14,4	14,0	37	35
Okt.	9,2	10,6	9,7	33	28
Nov.	4,8	6,3	5,5	32	32
Dec.	2,1	3,7	2,8	32	35

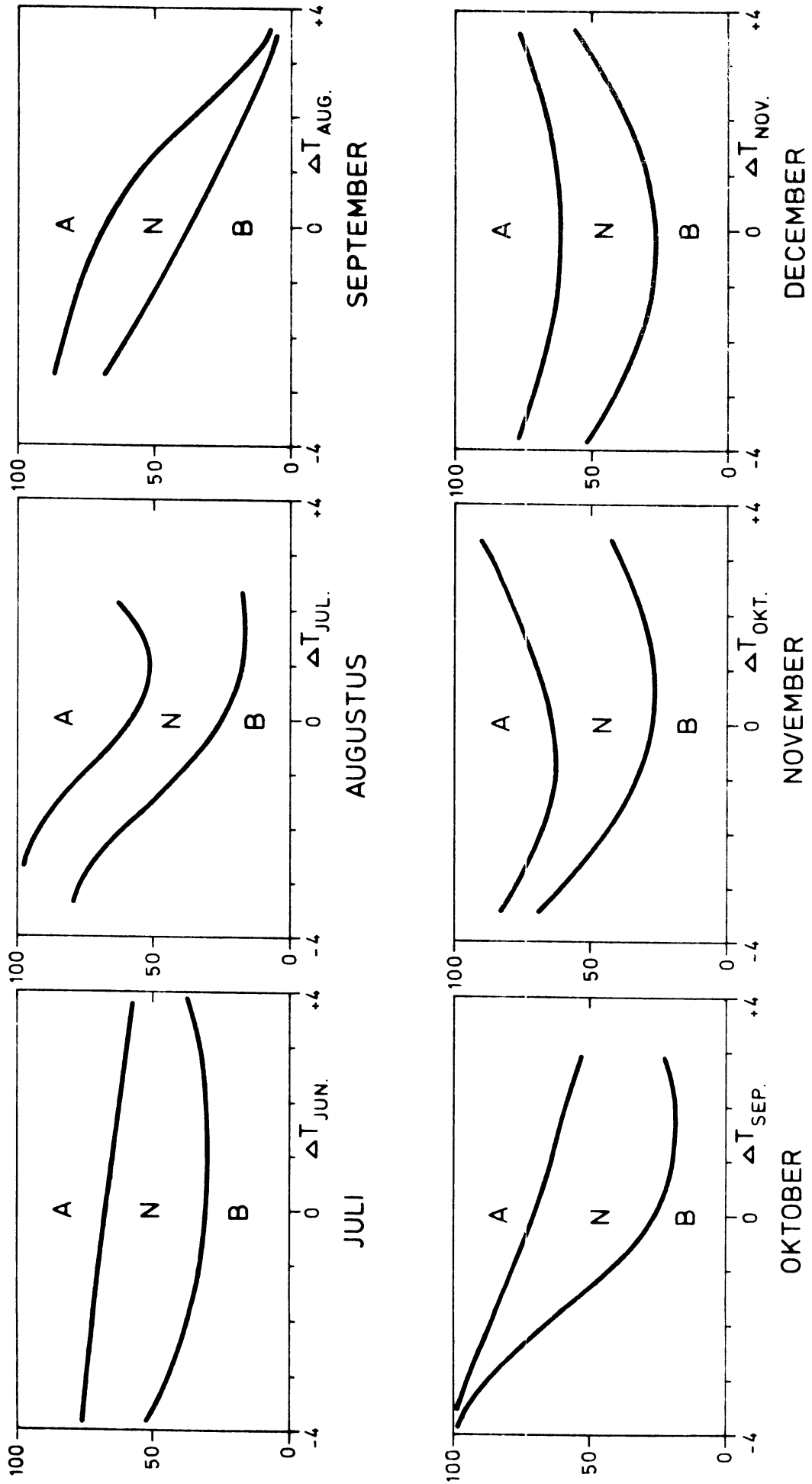
Tabel 5.2: Toetsingsresultaten model I en II.

Maand	Afhankelijke gegevens		Onafhankelijke gegevens		
	ΔLNL		\overline{LNL}	\overline{LNL}	\overline{LNL}
	I	II	KLIM	I	II
Jan.	2,0	3,2	1,10	1,12	1,09
Feb.	10,4*	15,2*	1,10	1,00	1,12
Maart	17,6*	20,2*	1,10	1,03	0,99
April	8,8	6,4	1,09	1,07	1,29
Mei	4,6	8,0	1,10	1,06	1,16
Juni	7,2	2,6	1,12	1,21	1,15
Juli	0,6	1,8	1,09	1,09	1,13
Aug.	12,0*	8,4	1,10	0,95	1,07
Sep.	10,2*	8,2	1,09	0,91	0,96
Okt.	8,8	5,0	1,15	1,12	1,10
Nov.	3,4	5,2	1,09	1,18	1,18
Dec.	1,6	0,8	1,10	1,10	1,15

Met name genoemde maand is de maand waarvoor de forecast geldig is.



Figuur 5.1: Kans op respectievelijk A, N of B in aangegeven maand als functie van ΔT in voorafgaande maand.



Figuur 5.2: Kans op respectievelijk A, N of B in aangegeven maand als functie van ΔT in voorafgaande maand.

5.2 Neerslag kans.

Bij het tweede voorbeeld is getracht de kans op neerslag te De Bilt tussen 06 en 18 UT te koppelen aan het stromingspatroon van het 500 mbar vlak. Het stromingspatroon werd hierbij beschreven met behulp van de eerste zes scores op de basispatronen (hoofdcomponenten) van de 500 mbar stroming.

Voor een nadere toelichting op deze variabelen wordt verwezen naar Kruizinga [6]. De y-variabele werd onderscheiden in twee klassen: D ($RR < 0,3$ mm) en N ($RR \geq 0,3$ mm). Daar er slechts twee klassen zijn, is er slechts één functie f_k

$$f_D = \partial_0 + \sum_{m=1}^6 \partial_m \cdot S_m \quad (5.3)$$

waaruit volgt

$$P_D = \exp(f_D) / (1 + \exp(f_D)) \quad (5.4)$$

Voor iedere kalendermaand werden eigen coëfficiënten afgeleid. De coëfficiënten ∂_0 t/m ∂_6 werden aangepast op de dagelijkse gegevens van de betreffende maand in de periode 1949 t/m 1970. Daarna werden de modellen getoetst op de gegevens uit de jaren 1971 t/m 1974. (voor iedere maand ongeveer 120 dagen). Ook nu is publicatie van de aangepaste coëfficiënten achterwege gelaten.

In tabel 5.3 zijn de toetsingsresultaten en enkele andere gegevens vermeld. Als eerste is vermeld de klimatologische kans op droog afgeleid uit de periode 1949 t/m 1970. In de tweede kolom is het toetsingsresultaat ΔLNL voor de afhankelijke gegevens vermeld. De kritieke grens is in dit geval 12,6. Zoals te zien liggen de resultaten daar ver boven, men dient echter wel te bedenken dat opeenvolgende dagen niet als onafhankelijk materiaal mogen worden beschouwd. Aansluitend zijn gegeven de bereikte \overline{LNL} op het onafhankelijke materiaal en de \overline{LNL}_{KLIM} van het onafhankelijke materiaal. Zoals te zien bevestigen deze resultaten de voorgaande toetsingen. Om de prestaties van het model vergelijkbaar te maken met andere

methodieken werd over de onafhankelijke proefset ook nog de prestatieindex PI bepaald. Hiertij werd de volgende regel gehanteerd "Voorspel droog als de berekende kans groter dan de klimatologische droogte kans anders nat".

Bij de bepaling van deze PI werd uitgegaan van de eerder in de tabel genoemde P_D . De laatste twee kolommen van 5.3 geven de model PI en de normale PI van de K-23 forecast welke werden ontleend aan de maandelijks overzichten van Daan. Uit de vergelijking van de laatste twee kolommen blijkt dat het gepresenteerde model zeker redelijke prestaties levert die mogelijk operationeel van nut kunnen zijn. Men dient echter nog te bedenken dat hier met 500 mbar analyses werd gewerkt. Voor een toets van de operationele bruikbaarheid zou toepassing op prognoses noodzakelijk zijn.

Tabel 5.3. Resultaten van het model voor de kans op een droge periode van twaalf uur.

Maand	Afhankelijke gegevens		Onafhankelijke gegevens			
	P_D %	ΔINL	\overline{INL}		PI	
			Model	KLIM	Model	Normaal K-23
Jan.	63,6	138	,54	,66	18	17
Feb.	64,3	136	,56	,60	13	15
Maart	68,4	142	,49	,57	14	12
April	67,1	170	,45	,61	17	24
Mei	70,9	64	,54	,59	12	14
Juni	73,3	128	,46	,60	20	16
Juli	65,4	117	,57	,62	10	14
Aug.	63,6	160	,44	,57	26	21
Sep.	66,3	211	,38	,64	32	23
Okt.	68,0	237	,42	,61	25	24
Nov.	60,8	126	,68	,73	15	18
Dec.	58,6	158	,54	,62	20	25
			Gemiddelde		18.5	18.6

6. Slot.

Doel van het onderzoek was om uit de theorie van de logistische discriminant analyse de benodigde programmatuur te ontwerpen om een dergelijke analyse ook te kunnen uitvoeren. Daarnaast diende aan de hand van enige praktische toepassingen de mogelijke waarde voor het statistisch-meteorologisch onderzoek te worden getoetst. Uit de hier beschreven voorbeelden blijkt dat deze analysetechniek zeker voor het element neerslag een nuttige vertaaltechniek kan zijn. Het is dan ook zinvol om de benodigde programmatuur verder te ontwikkelen en de operationele toepasbaarheid voor bijvoorbeeld het element neerslag meer diepgaand te onderzoeken. Een nevenresultaat van het hier beschreven onderzoek is dat de resultaten van Van den Dool en Nap voor de overgangsregels van de maandgemiddelde temperatuur voor enkele maanden kunnen worden bevestigd doch voor de overige maanden als discutabel gezien moeten worden.

Literatuur

1. Hanssen, A.W. and Kuipers, W.J.A.,
On the relationships between the frequency of rain and various meteorological parameters,
Meded. en verh. Kon. Ned. Meteor. Inst. Nr. 81. De Bilt, 1965.
2. Dool, H.M. van den en Nap, J.L.,
Praktisch gebruik van overgangsregels ten behoeve van de verwachting van de maandgemiddelde temperatuur te De Bilt, Verslagen; Kon. Ned. Meteor. Inst. V-200. De Bilt, 1977.
3. Hoel, P.G.,
Introduction to Mathematical Statistics Third edition p. 179, Wiley and Sons, New York.
4. Anderson, J.A.,
Logistic discrimination with medical applications,
Discriminant Analysis and Applications, Editor T. Cacoullos, Academic Press, New York, 1973.
5. Anderson, J.A.,
Separate sample logistic discrimination,
Biometrika, (1972), 59, p. 19.
6. Kruizinga, S.,
Objectieve classificatie van dagelijkse 500 mbar patronen,
Wet. rap. Kon. Ned. Meteor. Inst. W.R. 78-9. De Bilt, 1978.