

KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

DE BILT

VERSLAGEN

V-225

Experimentele weersvoorspelling op lange termijn
met de ADAM
(ADaptive Arithmetical Method)

door

J. van Galen

en

Th. de Crook

De Bilt, 1969

Publikatienummer: K.N.M.I. V 225-I
(tevens verslag van een colloquium
gehouden op 28 oktober 1969)

I N H O U D

0. Inleiding

1. Beschrijving van de ADAM
 - 1.1. Inleiding
 - 1.2. Bepaling van de klimatologie en de deviaties
 - 1.3. Eenmalige leerfase
 - 1.4. Voorspelling
 - 1.5. Suggesties ter verfijning en verbetering van de methode
 - 1.6. Referenties

2. Experimenten met de ADAM
 - 2.1. Korte omschrijving van de methode
 - 2.2. Verschillende versies van het ADAM-computerprogramma met resultaten
 - 2.3. Verdere ontwikkeling van de methode

0. Inleiding

Door Drs. J. van Galen is een methode ontwikkeld, waarmee voor de komende maand de gemiddelde etmaaltemperatuur voor de maand en de dekaden kan worden voorspeld.

Deze methode is beschreven in hoofdstuk 1.

Na het vertrek van Drs. J. van Galen, in september 1968, is de methode overgenomen door Ir. Th. de Crook, die de methode heeft beproefd en verbeterd. Hiervoor zij verwezen naar hoofdstuk 2.

Experimentele weersvoorspelling op lange termijn
met de ADAM (ADaptive Arithmetical Method)

1. Beschrijving van de ADAM

1.1. Inleiding

Problemen van statistische klassifikatie, "pattern recognition" e.d. worden in de laatste jaren aangevat met "lerende machines", waarvan de door Widrow (1962) zo genoemde ADALINE ("ADaptive LInear NEuron") het bekendste specimen is. De ADALINE is een conceptie, een model en kan uitgewerkt zijn als een machine (hardware) of een computerprogramma (software). Een ADALINE kan alleen besluiten tot "ja" of "nee" (binaire output). Wenst men meer detail, dan wordt gebruik gemaakt van een samenstel van meerdere ADALINES: een MADALINE ("Many ADALINES"). Een voorbeeld van de toepassing van een MADALINE op een probleem van meteorologische voorspelling is te vinden in een artikel van Glahn (1964). Voor de werking van een (M)ADALINE zij verwezen naar dit artikel.

Na een enkel experiment met een op het K.N.M.I. geprogrammeerde ADALINE kwam vast te staan dat de gehele eerste leer-fase, waarin het materiaal herhaalde malen wordt doorgevoerd, vrijwel equivalent is met het uitvoeren van enkele zeer eenvoudige arithmetische bewerkingen, welke bovendien slechts éénmalig behoeven te worden verricht. Uitwerking van dit procédé resulteerde in een, nu nog prille, ADAM (ADaptive Arithmetical Method) welke nog reminiscenties oproept aan de ADALINE. Te zijner tijd zou ADAM verder kunnen worden ontwikkeld tot een MADAM ("Many ADAMS").

1.2. Bepaling van de klimatologie en de deviaties

Om ADAM te beproeven is gedacht aan temperatuurvoorspelling op lange termijn op grond van deviaties van de klimatologische luchtdruk. Als te voorspellen grootte werd gekozen de maandgemiddelde etmaaltemperatuur te De Bilt. Kompakte informatie inzake de luchtdruk wordt verschaft door de "Grosswetterlagen" (GWL) volgens de nummering van Bijvoet en Schmidt (W.R. 58-4). Volgens de in dit rapport gegeven hoge druk - lage druk verdelingen kan aan elk van de 9 punten uit fig. 1 een waarde 1 of 0 worden toegekend, naargelang het punt in een gebied van relatief hoge resp. relatief lage druk gelegen is (zie Tabel 1).

In fig. 1a is een voorbeeld gegeven van een GWL met bijbehorende verdeling van enen en nullen.

Tabel 1

punt GWL	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	1	1	1
2	0	0	0	1	1	1	1	1	0
3	0	0	0	0	0	1	1	1	1
4	0	0	1	1	1	1	1	0	0
5	0	0	0	0	1	1	0	0	0
6	0	0	0	0	0	1	0	1	1
7	0	0	0	0	0	0	0	1	1
8	0	0	0	1	1	0	1	0	0
9	0	0	0	1	0	0	1	0	0
10	1	1	0	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	0
12	1	1	0	0	1	0	0	0	0
13	0	1	0	0	1	0	0	0	0
14	1	0	0	1	0	0	1	0	0
15	0	0	0	1	0	0	1	1	0
16	1	1	1	1	0	0	1	0	0
17	0	0	0	1	0	1	1	1	1
18	0	0	0	0	0	1	1	0	1
19	0	0	1	0	0	1	0	0	1
20	0	0	0	0	0	0	0	0	0
21	0	1	1	0	0	1	0	0	0
22	1	0	1	1	0	0	0	0	0
23	0	0	1	0	0	1	0	0	0
24	0	0	1	0	0	0	0	0	0
25	1	1	1	0	0	0	0	0	0
26	1	0	1	0	0	0	0	0	0
27	0	1	1	1	1	0	1	0	0
28	0	0	1	0	0	0	1	1	0
29 (ü)	0	0	0	0	0	0	0	0	0

Met een uitsluitend binaire output zou de temperatuur als "hoog" of "laag" voorspeld kunnen worden. In de praktijk doet zich echter de behoefte gevoelen ook de kwalifikatie "normaal" te hanteren.

Besloten werd om voor een eerste experiment de klasse-indeling B(elow),

N(ormal), A(bove) te gebruiken, mede omdat Dr. Schuurmans en Drs. Krijnen ook van deze indeling zich in hun "analoge gevallen" methode bedienen. Dit biedt het voordeel dat de uitkomsten van beide methoden vergeleken kunnen worden. De B, N, A-klassifikatie wordt maand voor maand vastgesteld op grond van de 75 jaren 1881 t/m 1955, waarbij de regel geldt dat in elk der klassen $33\frac{1}{3}\%$ van de gevallen is vertegenwoordigd. Zo geldt voor de maand juli de indeling $B \leq 16,2 \text{ }^\circ\text{C}$; $16,2 \text{ }^\circ\text{C} < N < 17,5 \text{ }^\circ\text{C}$; $A \geq 17,5 \text{ }^\circ\text{C}$. Elk der klassen bevat 25 gevallen. Is een éénduidige klasse-indeling aldus niet mogelijk, dan wordt alsnog de indeling éénduidig gemaakt door ervoor te zorgen dat:

- 1e. de som van de absolute waarde van de afwijkingen van de klasse-aantallen van de waarde 25 zo klein mogelijk is;
- 2e. indien er dan nog onzekerheid bestaat, klasse N zo groot mogelijk te nemen.

Laat er sprake zijn van 3 opeenvolgende maanden X, Y en Z, waarbij Z de maand is waarop de voorspelling betrekking heeft. Nu definiëren we drie perioden, t.w. (zie fig. 2):

periode III : 30 dagen te eindigen met de laatste dag van Y,
" II : als III, maar 7 dagen naar voren geschoven,
" I : " II, " " " " " " " .

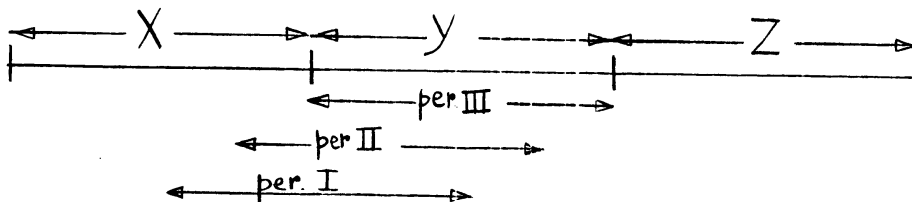


fig. 2

Het is duidelijk dat de 3 perioden elkaar gedeeltelijk overlappen. Perioden I en II bedekken gedeeltelijk X, gedeeltelijk Y. Periode III overdekt Y op de eerste dag na (Y 31 dagen) Y geheel (Y 30 dagen) of Y en bovendien 1 of 2 dagen van X (Y = februari).

Iedere dag vertegenwoordigt een GWL en dus een patroon van nullen en enen (tabel 1). Daarom wordt nu gedefiniëerd (per periode):

"drukscore": totale aantal malen 1 per punt;

"klimatologie": gemiddelde drukscore over de jaren 1881 t/m 1955;

"deviatie": drukscore minus klimatologie (alleen het teken wordt beschouwd).

Voor ieder der te voorspellen maanden Z is een afzonderlijke klimatologie per voorafgaande periode nodig (maximaal dus 12×3 klimatologieën) welke eenvoudig is te bepalen uit de GWL² van 1881 t/m 1955. De deviaties - m.b.t. iedere maand Z in totaal $3 \times 9 = 27$ deviaties - zullen nu de rol moeten vervullen van prediktors.

1.3. Eenmalige leerfase

Het ligt voor de hand dat, wanneer prediktor a positief is (a_+) en een belangrijke aanwijzing bevat t.a.v. op handen zijnde warmte in maand Z, een negatieve waarde van a (a_-) nog niet hoeft in te houden dat er kou op komst is. Daarom worden a_+ en a_- als afzonderlijke (mogelijke) prediktors beschouwd. Na een enkel experiment bleek dat van grotere indikatieve waarde meervoudige prediktors waren, zoals a_+b_+ , a_-c_+ , b_+c_- etc. Het aantal mogelijke prediktors komt daardoor (als we afzien van enkelvoudige prediktors zoals a_- , b_+ , c_- enz.) op $4 \times 13 \times 27 = 1404$. Hierbij is $a_+(a_-)$ een positieve (negatieve) deviatie in een willekeurig punt van een periode. Deze definitie geldt ook voor $b_-(b_-)$ en $c_+(c_-)$.

Verder zij opgemerkt dat a_+b_+ bijvoorbeeld zowel belangrijke aanwijzingen voor warmte als voor kou kan geven (extreme temperaturen). Zo zou a_+b_+ in de 75 jaren 1881 t/m 1955 n.b.t. maand Z 20 maal opgetreden kunnen zijn, terwijl Z 9 maal te warm (A) werd, 8 maal te koud (B) en slechts 3 maal normaal (N). Het is moeilijk om deze en dergelijke feitelijkheden onder te brengen in één prediktor, genaamd a_+b_+ . Derhalve konstrueren we hieruit 3 prediktors, nl. $a_+b_+ B$, $a_+b_+ N$ en $a_+b_+ A$. De eerste bevat de informatie omtrent de frequentie van het "B" worden van de maand Z, nadat a_+b_+ is gerealiseerd, de tweede bevat de informatie m.b.t. de frequentie waarmee de maand Z "N" wordt als a_+b_+ is gerealiseerd en de derde idem voor maand Z is "A". In totaal ontstaan hierdoor $3 \times 1404 = 4212$ mogelijke prediktors. Gesproken wordt van mogelijke prediktors, omdat zoals vanzelf spreekt een klein aantal zal moeten worden geselecteerd die een significante "inbreng" vertegenwoordigen. Dit houdt in dat aan alle mogelijke prediktors getallen ofwel gewichten zullen moeten worden toegekend. In bovenvermeld geval zal aan $a_+b_+ B$ zowel als aan $a_+b_+ A$ een groot positief gewicht worden toegekend en aan $a_+b_+ N$ een groot negatief gewicht, om aan te geven dat in geval van a_+b_+ de maand Z vermoedelijk "B" of "A" zal worden en niet "N". Om een vlot werkende procedure te verkrijgen, volgens welke aan elk der mogelijke prediktors een gewicht wordt toegekend, waarmee selectie van prediktors (en naderhand voorspelling) mogelijk wordt, beschouwen we een aan de werkelijkheid ontleend voorbeeld. We duiden een positieve (negatieve)

deviatie in punt 9 (zie fig. 1) voor periode II aan met a_+ (a_-) en een positieve (negatieve) deviatie in punt 9 voor periode III met b_+ (b_-). Voor $Z = \text{juli}$ kunnen de frequenties waarmee de maandtemperatuur in de klassen B, N of A viel, in afhankelijkheid van vooraf opgetreden combinaties a_+b_+ , a_+b_- , a_-b_+ en a_-b_- , worden afgelezen uit de volgende tabel (tabel 2).

Tabel 2

	B	N	A	gemiddeld
a_+b_+	5	14	9	9,33
a_+b_-	4	0	0	1,33
a_-b_+	2	0	4	2,00
a_-b_-	14	11	12	12,33

We zien dat a_+b_+ in totaal 28 maal voorkwam, waarbij 9 maal een te warme juli-maand volgde. Dit is niets bijzonders daar deze waarde vrijwel samenvalt met het gemiddelde aantal malen dat dit op grond van het toeval alléén, is te verwachten (9,33). De aanwijzingen voor N respektievelijk B zijn zwak positief respektievelijk zwak negatief. Ook bij a_-b_- treedt er niets bijzonders op. Bij a_+b_+ en a_-b_+ is de verdeling anders, hoewel hier het feit meespreekt dat het aantal gevallen zeer beperkt is. Toch moet aan het verschijnsel dat alle 4 malen a_+b_- gevolgd werden door een "B" aandacht worden besteed ofwel gewicht worden toegekend. Bij a_-b_+ 2 maal "B", precies gelijk aan het gemiddelde, hetgeen we alvast willen uitdrukken door te stellen: gewicht van a_-b_+ B = 0. Voor a_+b_+ N en a_-b_+ A moet op een of andere wijze een zwak negatief respektievelijk zwak positief gewicht worden vastgesteld. Bijvoorbeeld zou dit gewicht kunnen samenhangen met de "kans" waarmee 0 N- en 4 A-maanden op 6 gevallen onder zuiver toevallige omstandigheden optreden. Voorlopig volgen we een nog heuristisch aanpak.



fig. 3

We zagen (tabel 2) dat a_+b_+ , bij 28 keer voorkomen 5 maal "B" opleverde enz. In het algemeen zal bij a_+b_+ , op n maal voorkomen, k maal "B" volgen. Nu

onderscheiden we twee gevallen: $K > \frac{n}{3}$ (fig. 3a) en $K < \frac{n}{3}$ (fig. 3b). In deze beide gevallen willen we het gewicht laten samenhangen met de lengtes $(n-K)$ resp. (K) (fig. 3a en 3b). Deze moeten natuurlijk wel gezien worden in hun verhouding tot de totale lengte n . Voor het geval $K > \frac{n}{3}$ willen we dus het gewicht in eerste instantie positief laten zijn en laten afhangen van $\frac{n-K}{n}$. Voor het geval $K < \frac{n}{3}$ willen we het gewicht in eerste instantie negatief laten zijn en laten afhangen van $\frac{K}{n}$ ofwel van, wat hetzelfde is, $1 - \frac{n-K}{n}$. Na enig experimenteren werd voorlopig vastier gewerkt met

$$K > \frac{n}{3} : \text{gewicht} = -\ln \left\{ \frac{(n-K+1)(n-K+2)}{(n+1)(n+2)} * \frac{19 - \frac{3K}{n}}{8} \right\}.$$

$$K < \frac{n}{3} : \text{gewicht} = +\ln \left\{ \left[1 - \frac{(n-K)(n-K+1)}{(n+1)(n+2)} \right] * \frac{10 - \frac{3K}{n}}{5} \right\}.$$

In beide formules zijn de "bervormen" $\frac{n-K}{n}$ respectievelijk $1 - \frac{n-K}{n}$ nog te herkennen. In de buurt van $K = \frac{n}{3}$ is de aansluiting van beide formules niet exakt, maar wel is daar het gewicht ≈ 0 , hetgeen voor het beoogde doel - de selectie van predictors - voldoende is. Voor enige waarden van n en het limietgeval dat $n \rightarrow \infty$ zijn de gewichten als functie van de verhouding $\alpha = K/n$ weergegeven in fig. 4. Voor $\alpha = 1/3$ maakt het programma het gewicht gelijk aan nul.

Volgens bovenstaande formules*) werden de frequenties uit tabel 2 omgezet in de gewichten die vermeld staan in tabel 3.

Tabel 3

	B	N	A
a ₊ b ₊	-18	+25	+ 1
a ₊ b ₋	+101	-20	-20
a ₋ b ₊	0	-35	+39
a ₋ b ₋	+5	-2	+ 1

Dat a₊b₊ A en a₋b₋ A zwak positief zijn i.p.v. zwak negatief, is van geen betekenis.

Zouden nu de predictors worden geselecteerd met het grootste positieve gewicht, dan zou de informatie vervat in een aanzienlijk negatief gewicht niet worden gebruikt. Daarom wordt als volgt te werk gegaan. Geselecteerd worden de 20 predictors met de grootste positieve gewichten, welke betrek-

*) Alle gewichten werden bovendien met een faktor 50 vermenigvuldigd om met uitsluitend gehele getallen te kunnen werken.

king hebben op het optreden van "B", dus bijvoorbeeld $a_+ b_- B$ en andere prediktors van het type $x_{\pm} y_{\pm} B$. Los daarvan worden geselecteerd de 20 prediktors met het grootste gewicht, welke aanwijzing verschaffen voor het optreden van "N". Zo ook 20 "A"-prediktors. Bovendien worden voor elk der categorieën 20 prediktors geselecteerd met de grootste negatieve gewichten. In totaal dus, voor ieder der 12 maanden Z, 120 prediktors uit de groep van 4212 mogelijke prediktors (een kleine 3%). De eenmalige leerfase is hiermee voltooid, nu kan worden voorspeld.

1.4. Voorspelling

Voor de voorspelling worden de 27 deviaties bepaald en aan de hand hiervan wordt gekeken welke uiteindelijke prediktors "aktief" zijn en welke niet.

De 20 prediktors met positieve gewichten m.b.t. het optreden van "B", $p_1^{(B)}, p_2^{(B)}, \dots, p_{20}^{(B)}$ met gewichten $w_1^{(B)}, w_2^{(B)}, \dots, w_{20}^{(B)}$ leveren een getal (indikatie) $u^{(B)}$ op, gedefiniëerd door

$$\sum_{i=1}^{20} \delta_i^{(B)} \cdot w_i^{(B)} = u^{(B)}$$

Hierin is $\delta_i^{(B)} = 1$ of 0 , naargelang $p_i^{(B)}$ "Aktief" is of niet. Zo ook krijgen we een indicatie $v^{(B)}$ uit de prediktors die een negatief gewicht vertonen m.b.t. het optreden van "B". Op overeenkomstige wijze krijgen we de indicaties $u^{(N)}, v^{(N)}$ en $u^{(A)}, v^{(A)}$.

De negatieve indicatie $v^{(B)}$ betekent eigenlijk een positieve indicatie voor het optreden van "N" of "A" en zou dus, gelijk verdeeld, als positief bedrag bij $u^{(N)}$ en $u^{(A)}$ opgeteld kunnen worden. Zodoende vormen we de 3 indicaties $U^{(B)}, U^{(N)}$ en $U^{(A)}$:

$$U^{(B)} = u^{(B)} - \frac{1}{2} \left\{ v^{(N)} + v^{(A)} \right\}, \quad U^{(N)} = u^{(N)} - \frac{1}{2} \left\{ v^{(B)} + v^{(A)} \right\},$$

$$U^{(A)} = u^{(A)} - \frac{1}{2} \left\{ v^{(B)} + v^{(N)} \right\}.$$

Aangenomen dat we slechts letten op de onderlinge verschillen tussen $U^{(B)}, U^{(N)}$ en $U^{(A)}$, kunnen we overal de konstante $\left\{ v^{(B)} + v^{(N)} + v^{(A)} \right\} / 2$ bij-tellen, zodat het bovenstaande equivalent is met

$$U^{(B)} = u^{(B)} + \frac{1}{2} v^{(B)}, \quad U^{(N)} = u^{(N)} + \frac{1}{2} v^{(N)}, \quad U^{(A)} = u^{(A)} + \frac{1}{2} v^{(A)}$$

Op deze wijze komt bijvoorbeeld uit het terug-voorspellen van de gemiddelde etmaaltemperatuur in de maand juli 1881 (afhankelijk materiaal!):

$$u^{(B)} = +2, \quad v^{(B)} = -8; \quad u^{(N)} = +6, \quad v^{(N)} = -10; \quad u^{(A)} = +18, \quad v^{(A)} = 0$$

uit de bus, hetgeen resulteert in

$$U^{(B)} = -2; \quad U^{(N)} = +1; \quad U^{(A)} = +18$$

In dit geval zien we een zeer duidelijke indicatie voor het optreden van een warme (Above)-maand. Voor juli 1966 en juli 1967 waren de uitkomsten (onafhankelijk materiaal):

$$U^{(B)} = +23; \quad U^{(N)} = -9; \quad U^{(A)} = -6$$

respektievelijk

$$U^{(B)} = -7; \quad U^{(N)} = +1; \quad U^{(A)} = +14$$

Op grond hiervan kan men alleen maar "B" respektievelijk "A" voorspellen (hetgeen in deze gevallen ook uitkwam).

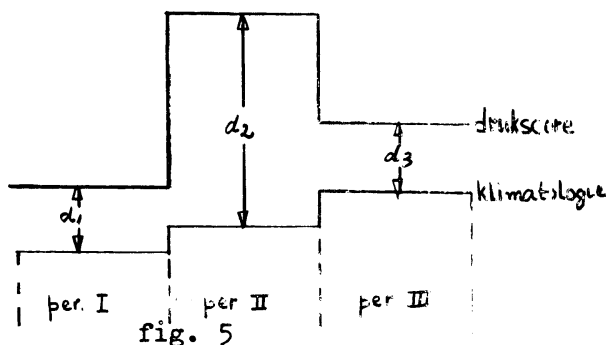
Niet altijd liggen de kaarten zo duidelijk. Wat bijvoorbeeld te zeggen van de uitspraak (voor juli 1935): $U^{(B)} = +8; U^{(N)} = +5; U^{(A)} = +9$? Eigenlijk wordt geen uitsluitsel gegeven en moet men zeggen: "ik weet het niet". De interpretatie van de 3 op objektieve wijze tot stand gekomen indicaties levert specifieke moeilijkheden op, welke nog bekeken moeten worden. Tot nu toe vond de interpretatie op subjektieve wijze plaats, maar het ligt natuurlijk in de lijn van de methode dit ook objektief (door de computer) te laten doen. Daarbij moet ruimte zijn voor het aangeven van 2 klassen of zelfs 3 (geen voorspelling mogelijk). Een mogelijkheid zou zijn de computer te laten "leren" te interpreteren, door uit een gegeven assortiment van mogelijke interpretatie-systemen, door automatische verandering van één of meer parameters e.d., de optimale te kiezen. Van 1881 tot heden is er immers oefenmateriaal genoeg. Dit impliceert wel dat het computerprogramma in staat moet zijn de voorspellingsuitspraken te verifiëren. Een eenvoudige evaluatie kan geschieden met de P.I. (prestatie-index) van Kuipers.

Tot nu toe zijn te weinig proefnemingen beschikbaar om een waarde-oordeel over de methode te bepalen. Wel kan voorzichtig gezegd worden dat het lijkt dat duidelijke en stellige uitspraken, zoals voor juli 1966 en juli 1967, enig houvast bieden en een bescheiden succes zouden kunnen opleveren. Dit kan worden onderzocht zodra objektieve interpretatie en verifikatie ge-programmeerd zijn. Wellicht is het goed bij het ontwikkelen van de objektieve interpretatie tevens een maat voor de stelligheid, waarmee de uitspraak wordt gedaan, te introduceren en deze ook bij de verifikatie te betrekken.

1.5. Suggesties ter verfijning en verbetering van de methode

Het onderzoek naar de mogelijkheden van weersvoorspelling op lange termijn met methoden zoals de ADAM is nog maar juist begonnen. Talloze suggesties zijn te geven voor verfijning (en verbetering?). De meest voor de hand liggende zijn wel:

1. de objectieve interpretatie, waarbij een maat van stelligheid wordt geïntroduceerd, welke kan worden omgezet in "klare taal": "vermoedelijk" A, "zeker" N, e.d.;
2. de objectieve verifikatie, waarbij de aangegeven mate van stelligheid van de uitspraak op de één of andere wijze wordt verdiskonteerd;
3. experimenten met de totstandkoming van de gewichten: dit is nu min of meer "op het oog" gedaan en zou op een zekere statistische basis geschoeid misschien meer opleveren(?);
4. nu is ADAM getraind op de jaren 188^t/m 1955, maar van de ervaring met de jaren 1956 tot heden is geen profijt getrokken - dit zou kunnen gebeuren in de verifikatiefase waarbij de gewichten aan de nieuwe ervaring worden aangepast, waardoor de methode eerst echt "adaptive" wordt;
5. een veel breder opzet door niet slechts drie perioden willekeurig te kiezen, maar alle mogelijke korte, lange, overlappende en niet-overlappende perioden aan de selectie van prediktors mee te laten doen - alleen zó komt men erachter welke perioden het zijn die relevante informatie bevatten;
6. niet alleen perioden, maar ook perioden-verschillen kunnen prediktors opleveren. In fig. 5 zijn de deviaties voor 3 opeenvolgende perioden



(I, II en III) alle positief (d_1, d_2, d_3 elk > 0). Het kan van belang zijn voor de voorspelling, dat de tweede deviatie groter is dan de beide andere. Dit kan worden medegedeeld door een soort "afgeleide deviatie", of "deviatie van de 2e orde", Δ_i te introduceren, door $\Delta_j = d_{j+1} - d_j$ bijvoorbeeld.

De nieuwe deviaties $\Delta_1 (> 0)$ en $\Delta_2 (< 0)$ maken bekend dat er een relatief maximum van de drukscore (verminderd met de klimatologische gang) valt in periode II, en kunnen als gelijkwaardige mogelijke prediktors

- meedoen aan de selectie van prediktors. Wellicht komt één van de afgeleide deviaties met een hoog gewicht uit de bus?
7. alle prediktors zijn tot nu toe afgeleid uit het verloop van de GWL ⁿ en dus uit de veranderende drukverdeling. Ook uit de temperatuur op een aantal plaatsen gedurende zekere perioden zou ADAM op de beschreven wijze de maandgemiddelde etmaaltemperatuur te De Bilt kunnen voorspellen (in principe althans). Zo zouden diverse ADAMS, elk zich baserend op een andere meteorologische parameter een uitspraak omtrent de te verwachten maandtemperatuur kunnen doen. Elk van deze uitspraken is dan weer op te vatten als een (mogelijke) prediktor, waaraan, op grond van de resultaten van de onderscheidene verifikaties, door een super-programma (MADAM) gewichten kunnen worden toegekend teneinde hiermee te komen tot een einduitspraak. De MADAM zal zo getraind moeten worden dat de verifikatie van de einduitspraak een hogere P.I. geeft dan de verifikaties van elk der afzonderlijke ADAM-uitspraken, anders heeft de MADAM geen zin!
 8. naast een MADAM die getraind is op maandgemiddelde etmaaltemperaturen kunnen andere MADAMS gespecialiseerd zijn in dekade-, seizoen- of jaargemiddelden van hetzij etmaal-, dag- of nachtgemiddelde temperaturen hetzij maximum- of minimumtemperaturen;
 9. behalve dergelijke temperatuur-MADAMS zijn ook zonnenschijn- en neerslag-MADAMS e.d. mogelijk;
 10. een gedetailleerdere klasse-indeling dan B, N, A, met elk 33-1/3% van alle gevallen, zoals M(uch) B, B, N, A, (M)uch A, met elk 20% van alle gevallen is op precies dezelfde wijze te realiseren als boven is beschreven. Nadeel is dat bij een gegeven leer-periode van bijvoorbeeld de 75 jaren 1881 t/m 1955 m.b.t. elke kategorie nu geen 25 doch slechts 15 jaren ervaring meetellen! Het heeft weinig zin om met fijnere klasse-indelingen te experimenteren zolang men er niet zeker van is dat de meest grove indeling meer succes biedt dan men kan verkrijgen met simpele andere verwachtingstechnieken welke gebruik maken van klimatologische waarden, persistentie e.d.

1.6. Referenties

- Glahn, H.R., 1964: An application of adaptive logic to meteorological prediction. J. appl. Meteor., 3, 718 - 725.
- Widrow, B., 1962: Generalization and information storage in networks of Adaline "Neurons". Self-organizing systems, Washington, D.C., Spartan Books.

- Hu, M.J.C. and Root, H.E., 1964: An adaptive data Processing System for Weather Forecasting. J. appl. Meteorol., 3, 513 - 523.
- Hu, M.J.C., 1963: A trainable weather-forecasting system.

2. Experimenten met de ADAM

2.1. Korte omschrijving van de methode

De ADAM bestaat uit drie delen:

- 1e. de voorbereiding tot het leerproces;
- 2e. het éénmalige leerproces;
- 3e. de voorspelling.

1e. De voorbereiding tot het leerproces

Dit omvat de bepaling van de deviaties van de "klimatologische luchtdruk" in drie nader te bepalen perioden (vanaf 1881).

Informatie inzake de "luchtdruk" wordt ons verschaft door de "Grosswetterlagen".

2e. Het éénmalige leerproces

Uit de deviaties (van drie perioden) en de maandgemiddelde etmaaltemperatuur (van de te voorspellen maand) van 1881 t/m 1955, worden volgens een bepaald procédé (zie hoofdstuk 1) 6 * 20 prediktors met bijbehorende gewichten bepaald.

3e. De voorspelling

Per jaar wordt voor de desbetreffende maand een voorspelling gemaakt m.b.v. de deviaties en de 6 * 20 geselecteerde prediktors.

2.2. Verschillende versies van het ADAM-computerprogramma met resultaten

Achtereenvolgens worden vier versies besproken van de ADAM.

2.2.1. Versie 0

Dit is de oorspronkelijke versie, zoals die beschreven is in hoofdstuk 1. Hierin worden de klimatologie en de deviaties bepaald in drie overlappende perioden van 30 dagen (overlapping 7 dagen).

2.2.2. Versie A

Hierin worden de klimatologie en de deviaties eveneens berekend in drie overlappende perioden van 30 dagen. Er worden nu echter bij de voorspelling extra gewichten ingevoerd afhankelijk van de grootte van de deviaties. Een actieve prediktor die bestaat uit grote deviaties krijgt naar verhouding een groter gewicht dan een actieve prediktor die bestaat uit kleinere deviaties.

Voor een grote deviatie G geldt $|G| \geq 10$ eenheden,
voor een kleine deviatie K geldt $|K| < 5$ en voor
een middelgrote deviatie M geldt $5 \leq |M| < 10$.

De toekenning van de extra gewichten is als volgt:

Tabel 4

prediktor	extra gewicht
G, G	2
G, M	1½
M, M	1
G, K	1
M, K	½
K, K	0

Dit wil zeggen dat indien bijvoorbeeld de actieve prediktor bestaat uit twee grote deviaties, het gewicht behorende bij deze prediktor met een factor twee vermenigvuldigd wordt.

De voorspelling geschiedt verder op analoge wijze als in 1.4.

2.2.3. Versie E

De klimatologie en de deviaties worden bepaald in drie aaneensluitende perioden van 15 dagen, waarbij de laatste dag van de derde periode samenvalt met de laatste dag van de maand voorafgaande aan de te voorspellen maand Z, dus:

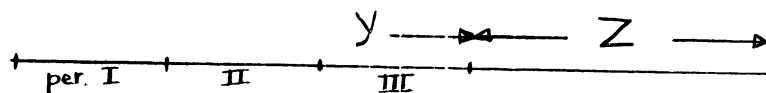


fig. 6

Voor het overige is de versie E identiek aan versie 0.

2.2.4. Versie T

Hierin wordt weer gewerkt met drie aansluitende perioden van 15 dagen. Bovendien treden de deviatieverschillen van de perioden op als prediktors. Uit de deviaties en de maandgemiddelde etmaaltemperatuur van 1881 t/m 1955 (75 jaar) worden 6 * 20 prediktors met bijbehorende gewichten geselecteerd.

De deviaties zijn voor ieder jaar bekend in drie perioden van elk negen punten a.v.:

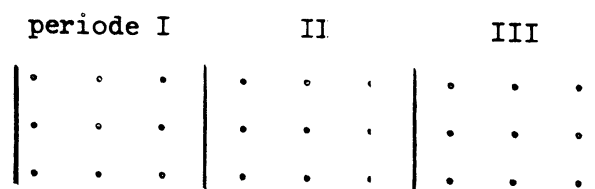


fig. 7

De deviatieverschillen worden nu als volgt berekend:

- 1e. Bepaal het verschil van de deviaties in periode II met de deviaties in de overeenkomstige punten van periode I. Zodoende verkrijgen we negen deviatieverschillen.
- 2e. Bepaal het verschil van de deviaties in periode III met de deviaties van periode II, dit geeft weer negen deviatieverschillen.
- 3e. Bepaal het verschil van de deviaties van periode III met de deviaties van periode I.

We hebben nu 27 deviatieverschillen verkregen. Op analoge wijze als bij de deviaties kunnen we weer 6×20 prediktors met bijbehorende gewichten opstellen en een voorspelling maken (zie 1.4.). De voorspelling die gebaseerd is op de deviaties levert de indicaties $U_1^{(B)}$, $U_1^{(N)}$ en $U_1^{(A)}$ en de voorspelling m.b.v. de deviatieverschillen de indicaties $U_2^{(B)}$, $U_2^{(N)}$ en $U_2^{(A)}$.

Uit deze zes indicaties worden drie nieuwe indicaties gevormd nl.:

$$U^{(B)} = \frac{U_1^{(B)} + U_2^{(B)}}{2}, \quad U^{(N)} = \frac{U_1^{(N)} + U_2^{(N)}}{2} \quad \text{en} \quad U^{(A)} = \frac{U_1^{(A)} + U_2^{(A)}}{2},$$

die ons een indicatie verschaffen voor het optreden van een resp. te koude, normale of te warme maand.

2.2.5. Resultaten

Ten einde de kwaliteit van de verschillende versies te kunnen beoordelen en de versies onderling met elkaar te kunnen vergelijken wordt voor elke versie een prestatie-index (P.I.) berekend.

De prestatie-index (in %) wordt bepaald met de formule:

$$P.I. = \frac{T - I}{N - I} \times 100$$

waarbij T = het aantal opgetreden treffers,

N = het totaal aantal mogelijke treffers

en I = het aantal treffers dat op grond van het toeval optreedt.

In tabel 5 zijn voor de maanden november t/m maart voor elke maand en voor elke dekade P.I.'s berekend van de versies O, A, E en T. Voor november en december zijn de P.I.'s berekend uit telkens 75 jaren (1881 t/m 1955) voor wat betreft het afhankelijke materiaal en uit 12 jaar (1956 t/m 1967) voor het onafhankelijke materiaal.

Voor januari zijn de P.I.'s van het afhankelijke materiaal bepaald uit de jaren 1882 t/m 1956 en van het onafhankelijke materiaal uit de jaren 1957 t/m 1968. Voor februari en maart zijn de P.I.'s van het afhankelijke materiaal berekend uit de jaren 1881 t/m 1955 en van het onafhankelijke materiaal uit de jaren 1956 t/m 1968.

Zo is voor maart de P.I. van de maand (voor het onafhankelijke materiaal) met versie E gelijk aan +19 (deze P.I. is berekend uit de maanden maart van 1956 t/m 1968).

Bij het onafhankelijke materiaal zien we een grote spreiding in de P.I.'s; voor het afhankelijk materiaal is de spreiding aanmerkelijk kleiner.

In tabel 6 zijn P.I.'s berekend voor de maand en de dekaden uit de maanden november t/m maart.

Hierbij zijn de P.I.'s van het onafhankelijke materiaal bepaald uit $3 \times 12 + 2 \times 13$ voorspellingen en de P.I.'s van het afhankelijk materiaal uit 5×75 voorspellingen.

Zo is bijvoorbeeld voor de 2e dekade met versie T de berekende P.I. gelijk aan +8 (onafhankelijk materiaal).

Ook hier is de spreiding van de P.I.'s voor het onafhankelijke materiaal het grootst.

Tabel 5

Ver- M en dek. sie	Onafh. materiaal				Afh.mat.		
	0	A	E	T	E	T	
11,0	-19	-25	+6	+13	+62	+64	november, maand november, 1e dekade november, 2e dekade november, 3e dekade
11,1	+13	+31	-6	+0	+58	+59	
11,2	-13	-25	+0	+6	+57	+62	
11,3	-13	-13	+6	+13	+51	+59	
12,0	+13	+13	+0	+0	+53	+61	december
12,1	+0	-13	+6	+13	+55	+67	
12,2	+19	+25	+0	-6	+59	+64	
12,3	+13	+0	+38	+38	+63	+67	
1,0	-38	-44	-25	-19	+59	+67	januari
1,1	-25	-13	+25	+25	+58	+63	
1,2	-13	+0	-13	-6	+51	+58	
1,3	-6	+0	+0	+25	+55	+52	
2,0	+25	+19	+25	+25	+61	+62	februari
2,1	+13	+8	-27	-38	+63	+73	
2,2	+13	+19	+19	+19	+57	+66	
2,3	-15	-27	-4	-4	+55	+66	
3,0	-15	+2	+19	+36	+57	+62	maart
3,1	+37	+31	+13	+42	+62	+69	
3,2	+25	+25	+25	+31	+62	+63	
3,3	-15	-21	+13	+19	+59	+77	

Tabel 6

	Onafh. materiaal				Afh. mat.	
	O	A	E	T	E	T
Maand	-7	-7	+5	+11	+58	+63
1e dek.	+8	+9	+2	+7	+59	+66
2e dek.	+6	+9	+6	+8	+57	+63
3e dek.	-7	-12	+11	+18	+57	+64
Gem.	+0	+0	+6	+11	+58	+64

Bovendien is voor elke versie (zie tabel 6) een gemiddelde P.I. berekend, zowel voor het afhankelijke als het onafhankelijke materiaal. Deze gemiddelde P.I.'s zijn bepaald uit de P.I.'s voor de maand en de dekaden, dus uit $4 * (3 * 12 + 2 * 13) = 248$ voorspellingen voor het onafhankelijke materiaal en uit $4 * 5 * 75 = 1500$ voorspellingen voor het afhankelijke materiaal.

We zien dat voor het onafhankelijke materiaal de O en de A versie een gemiddelde P.I. gelijk aan nul geven, d.w.z. dat de voorspelling niet beter is dan op grond van het toeval verkregen zou zijn. Voor de versie E is de gemiddelde P.I. gelijk aan +6 en voor de T versie +11.

Voor het afhankelijke materiaal zijn de gemiddelde P.I.'s voor de versies O, A, E en T resp. +52, +40, +58 en +64.

Beschouwen we de resultaten nader dan zien we dat zowel voor het onafhankelijke materiaal als voor het afhankelijke materiaal de T versie de hoogste gemiddelde P.I. heeft, hetgeen ons niet hoeft te verbazen, omdat bij deze versie meer informatie wordt gebruikt, naast deviaties ook deviatie-verschillen van de perioden.

Vergelijken we versie E met versie O dan zien we dat de gemiddelde P.I. van versie E het grootst is (zowel voor het afhankelijke als het onafhankelijke materiaal). Blijkbaar was de keuze van drie overlappende perioden van 30 dagen minder gelukkig.

Versie O wint het van versie A. De invoering van extra gewichten afhankelijk van de grootte van de deviaties heeft een nadelige invloed op de voorspelling, vooral bij het afhankelijke materiaal. Een verklaring hiervan zou kunnen zijn dat het geleerde m.b.v. het leerproces op een onjuiste manier wordt gebruikt bij de voorspelling.

2.2.6. De significantie van de verschillende versies

De significantie wordt getoetst met behulp van de t-toets. Voor de t-toets

geldt:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

waarbij $s = \sigma \sqrt{\frac{n}{n-1}}$

en $\sigma^2 = \sum (x - \bar{X})^2 / n = \sum x^2 / n - \bar{X}^2$

dus:

$$t = \frac{(\bar{X} - \mu)\sqrt{n-1}}{\sigma}$$

Hierbij is σ de standaardafwijking en μ de verwachtingswaarde. Nemen we voor x de 20 P.I.'s volgens tabel 5 - E versie, onafhankelijk materiaal - dan vinden we:

$n = 20$

$\bar{X} = 0.06$

$\sigma = 0.16$

dus $t = \frac{(0.06 - 0)\sqrt{19}}{0.16} = 1.64$ indien $\mu = 0$.

Volgens de tabel op blz. 251 van het boek "Standard Mathematical Tables" (1959) komt dit overeen met een grens van 13%. Stellen we de significantiegrens op 5% dan wil dat zeggen dat de gemiddelde P.I. van 0.06 niet significant van nul verschilt.

Voor de T-versie - zie tabel 5, onafhankelijk materiaal - vinden we:

$n = 20$

$\bar{X} = 0.11$

$\sigma = 0.20$

dus $t = \frac{(0.11 - 0)\sqrt{19}}{0.20} = 2.40$

Dit komt overeen met een grens van 3%, dus de gemiddelde P.I. van 0.11 verschilt significant van de verwachtingswaarde $\mu = 0$.

M.a.w. de T-versie is significant verschillend van de methode, waarbij de voorspelling op grond van het toeval wordt gedaan.

Omdat voor de O-versie en de A-versie de gemiddelde P.I. gelijk nul is (dus $t = 0$), is het duidelijk dat deze versies niet significant verschillen van de methode, die berust op het toeval ($\mu = 0$).

2.3. Verdere ontwikkeling van de methode

2.3.1. De nieuwe versie U

Het voorgaande heeft wel duidelijk gemaakt dat er mogelijkheden zijn de ADAM te verbeteren. Een mogelijke verbetering van de methode kan wellicht worden verkregen met de versie U, waarvan nu een beschrijving volgt.

De klimatologie en de deviaties worden weer bepaald in drie aaneensluitende perioden van 15 dagen. Voorts spelen de deviatieverschillen een rol bij het bepalen van de prediktors.

In hoofdstuk 1, tabel 2, staat een frequentietabel voor tweevoudige prediktors. We stellen nu een nieuwe tabel op die er schematisch als volgt uit ziet (tabel 7):

Tabel 7

	B	N	A
$a_+ b_+ c_+$	x	x	x
$a_+ b_+ c_-$	x	x	x
$a_+ b_- c_+$	x	x	x
$a_- b_+ c_-$	x	x	x
$a_- b_- c_+$	x	x	x
$a_- b_- c_-$	x	x	x

We hebben hier te maken met drievoudige prediktors, waarbij a_+ (a_-) een positieve (negatieve) deviatie voorstelt in een punt van een periode. Zo ook b_+ (b_-). De deviatieverschillen van a_+ en b_+ worden gedefinieerd door c_+ , waarbij c_+ een positief deviatieverschil voorstelt en c_- een negatief deviatieverschil. Het aantal prediktors komt bij versie U op $3 \times 6 \times 13 \times 27 = 6318$, dus $1\frac{1}{2}$ maal het aantal prediktors van versie O. Vergeleken met versie T spelen hier niet alleen de deviatieverschillen van de perioden mee, doch ook de deviatieverschillen van de punten binnen eenzelfde periode. We gebruiken dus meer informatie voor de voorspelling en hopen daarom op een verbetering van de methode.

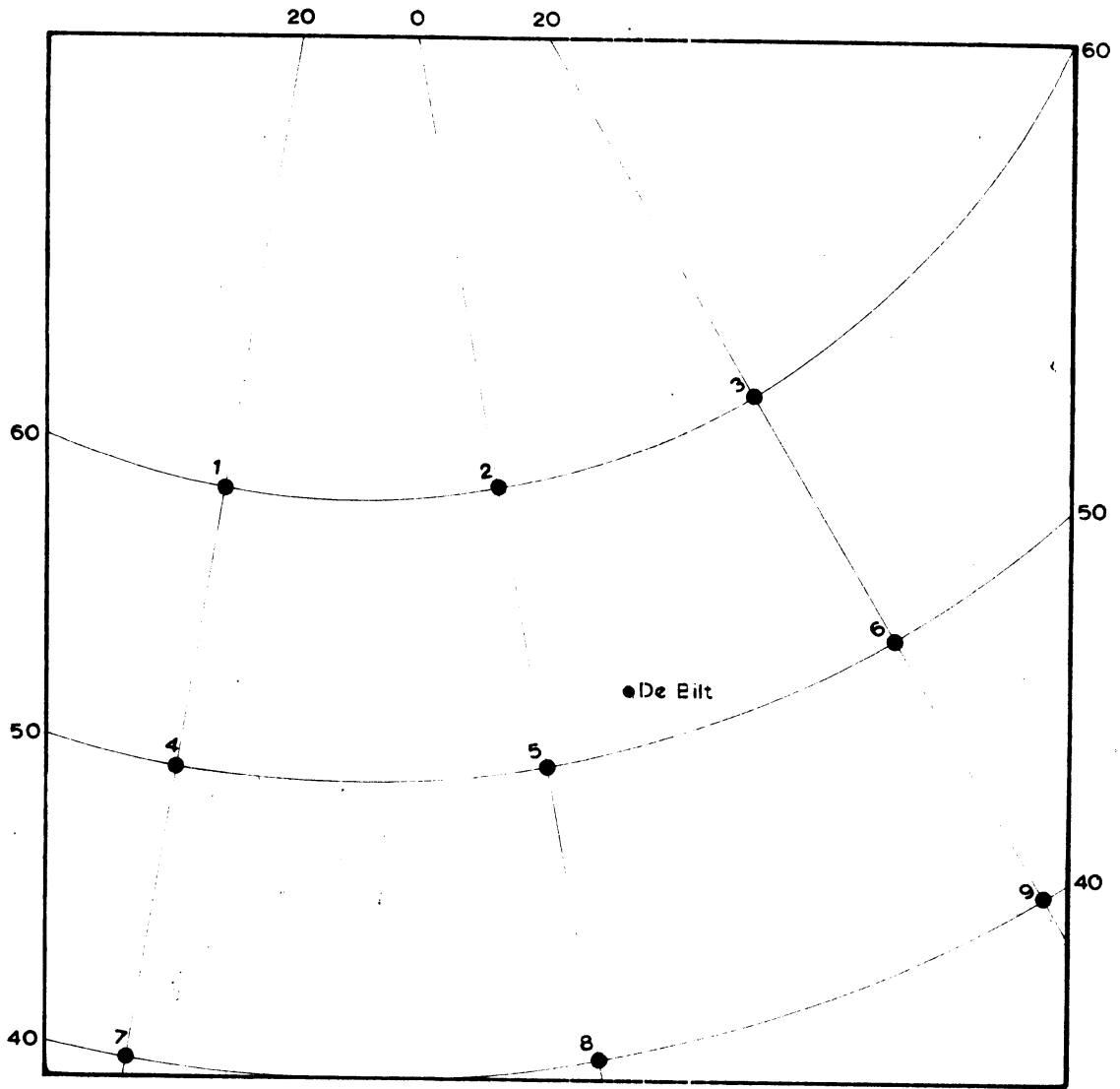
2.3.2. Interpretatie en verifikatie

Een ander punt dat stellig de aandacht verdient is de interpretatie van de drie op objectieve wijze tot stand gekomen indicaties voor het optreden van een warme, normale of koude maand (dekade). Deze interpretatie vindt thans nog op subjektieve wijze plaats. Ook de verifikatie - het bepalen van de prestatie-index - gebeurt nog manueel. Dit brengt veel werk met zich mee, omdat voor iedere nieuwe versie een groot aantal voorspellingen dient te worden gemaakt, teneinde een uitspraak te kunnen doen over de kwaliteit van de methode. Het verdient daarom aanbeveling de interpretatie objectief te maken, zodat zowel de interpretatie als de verifikatie door de computer kan worden verricht.

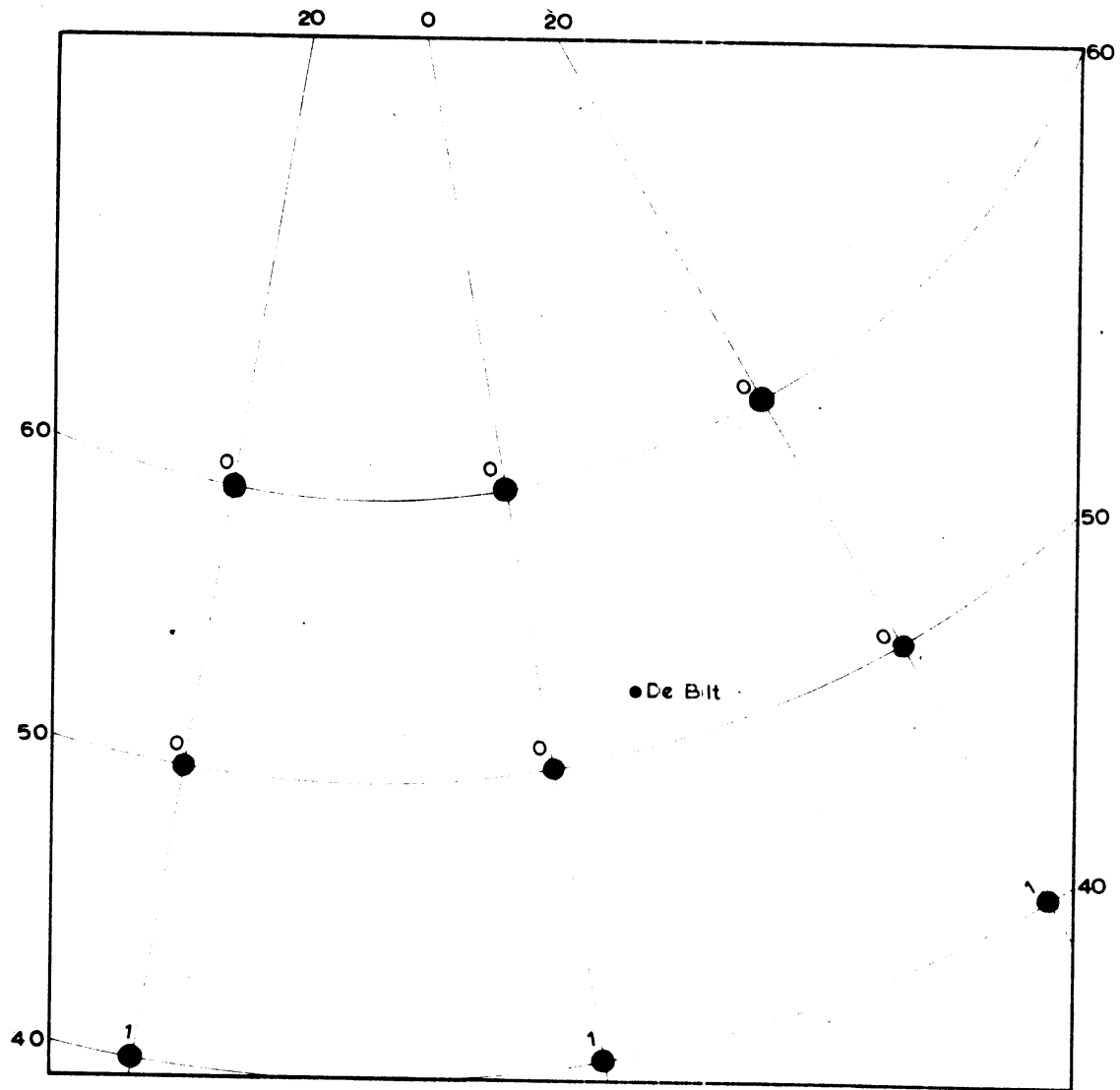
2.3.3. Voorspelling van neerslaghoeveelheden

Met de ADAM, die opgesteld is ter voorspelling van de gemiddelde etmaaltemperatuur van de maand en de dekaden, kan op analoge wijze voor de komende maand de neerslaghoeveelheid worden voorspeld van de maand en de dekaden.

Het ligt in de bedoeling hiermee te experimenteren en bij een gunstig resultaat, naast de gemiddelde etmaaltemperatuur, ook de neerslaghoeveelheid in de voorspelling op te nemen.

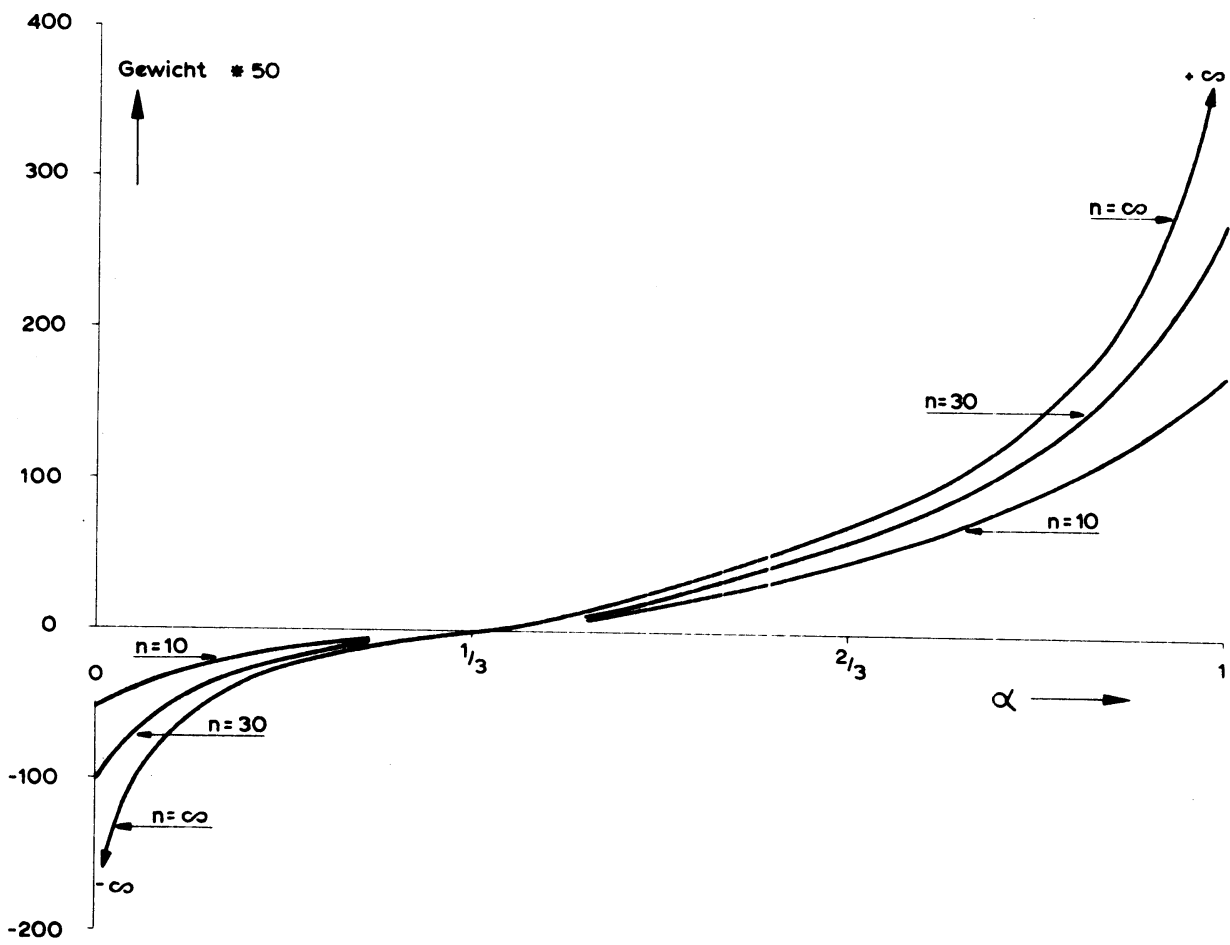


Plaats van de negen punten
Figuur 1



De GWL 1 (Ws) met bijbehorende verdeling van enen en nullen

Figuur 1^a



Gewichten als functie van $\alpha = k/n$

Figuur 4