KONINKLIJK NEDERLANDS
METEOROLOGISCH INSTITUUT

Verslagen V-70
(R III-255-1960)

Een statistisch onderzoek naar het verband tussen het weer en de opbrengst van het aardappelras Eigenheimer op de proefboerderij Borgercompagnie

door

Ir. R.F. Fisscher en H. de Hart

Inhoud.

1. Inleiding.

Uit ervaring en in bepaalde gevallen uit vrij uitvoerige studies is bekend, dat de invloed van het weer op de opbrengst van land- en tuinbouwgewassen vaak zeer belangrijk is. Soms is de invloed zo evident, dat deze zonder meer is aan te wijzen; dit is b.v. het geval bij intensieve en langdurige droogte of regen, die zelfs tot een totale mislukking van het gewas kan leiden.

Wil men zich echter een indruk vormen van de kwantitatieve invloed van het weer, dan kan men het bezwaarlijk zonder statistisch onderzoek stellen.

Wanneer men b.v. vraagt hoe groot de schade is die b.v. door de droogte werd veroorzaakt ofwel hoeveel mm regen nodig zouden zijn geweest om de schade te voorkomen, dan vereist de beantwoording een intensieve studie, waarbij een doelmatige statistische methode onmisbaar is.

Het was daarom te verwachten, dat men reeds betrekkelijk spoedig nadat statistisch onderzoek in de land- en tuinbouw ingang vond, langs verschillende wegen heeft getracht de invloed van weerfaktoren op de opbrengst exakt en kwantitatief te bepalen.

Van de bestaande statistische methoden die op dit doel zijn gericht, heeft de methode van de multipele normale correlatie- en lineaire regressie de meeste toepassing gevonden. In ons land kunnen in dit verband de onderzoekingen worden genoemd van Frankena (2), Woudenberg (3), Ignatius en de Wit (4), Post (5, 6), Kramer, Post en Wilten (7), Fisscher (10) en Woudenberg en Poelstra (11).

Opvallend is evenwel, dat vele van de onderzoekers op dit gebied, zowel in binnen- als buitenland, nimmer hebben getracht de betrouwbaarheid van de resultaten van hun werk aan de praktijk te toetsen; blijkbaar menen zij hun taak als geëindigd te moeten beschouwen, wanneer de regressievergelijking, zo zij deze al hebben opgesteld, op papier staat.

Dat op deze wijze de algemene waarde der uitkomsten van het onderzoek niet bepaald hoog mag worden aangeslagen, ligt min of meer voor de hand.

Terecht formuleert v.d. Bijl (8) zijn oordeel over deze nalatigheid in de volgende scherpe bewoordingen:

"Iedereen, die een regressievergelijking heeft opgesteld, zou verplicht moeten worden elk jaar het al of niet slagen bekend te maken. Is er een redelijk succes, dan kunnen degenen voor wie de vergelijkingen zijn opgesteld (Overheid, enz.) daarvan ten volle profiteren. Blijkt de regressievergelijking niet te voldoen, dan moet dit ruiterlijk bekend worden, opdat geen argeloze collega's in vol vertrouwen de dupe worden van de ongeschikt gebleken regressies.

Dat de meeste der onderzoekers niet jaarlijks hun verwachtingen en het resultaat daarvan publiceren, moet wel als een bewijs worden opgevat, dat zij zelf in hun eigen methoden geen vertrouwen stellen." -

Tot de weinigen die hun als resultaat van het onderzoek gevonden lineaire regressievergelijking aan de praktijk hebben getoetst, behoren Woudenberg en Poelstra. Zij hebben dit voor een tiental jaren buiten de basisreeks voor de provincie Groningen gedaan. De verschillen tussen de berekende en de werkelijke zomertarwe-opbrengsten bleken te variëren van + 220 kg/ha tot -1100 kg/ha. Voor een tweetal jaren waren de berekende opbrengsten hoger (100 resp. 220 kg/ha), voor de overige jaren lager (100 - 300 kg/ha) tot veel lager (500, 950 en 1100 kg/ha) dan de werkelijke.

Ook Fisscher heeft enkele jaren na de voltooiing van zijn onderzoek nagegaan of en in hoeverre de door hem voor enkele havergebieden van Nederland opgestelde lineaire regressie-vergelijkingen bruikbaar zouden zijn voor een oogstprognose o.a. in het zuidelijk zandgebied en in de Veenkoloniën. De afwijkingen van de berekende - t.o.v. de werkelijke opbrengsten bleken eveneens zodanig groot, dat aan een dergelijke oogstverwachting weinig of geen waarde voor de praktijk mag worden toegekend.

De beide aangehaalde voorbeelden van de teleurstellende resultaten van een met behulp van de normale correlatie- en regressie-methode verricht statistisch onderzoek naar de samenhang tussen het weer en de opbrengst van gewassen waren aanleiding tot de vraag of het betreffende probleem wellicht niet beter langs een andere statistische weg zou zijn te benaderen.

Het in 1954 verschenen boek van Sanderson (9), vestigde de aandacht op de methode van de regressie-integralen, een werkwijze die voor zover onzerzijds bekend, tot dusver op het Europese vasteland nog niet in gebruik is, vermoedelijk vanwege de, althans voor de niet statistisch georiënteerde landbouwkundige, ingewikkelde mathematiek (hoewel Sanderson zelf terzake van oordeel is dat "the mathematical solution of the problem is fundamentally simple"!).

De betreffende methode is door R.A. Fisher (1) uitgedacht en in 1924 gepubliceerd.

In het voormalige Nederlands-Indië is deze methode, zoals een onzer bekend, in de dertiger jaren toegepast door S.H. Justesen, destijds hoofd van de wiskundige afdeling van het Algemeen Proefstation voor de Landbouw te Buitenzorg, thans Directeur van het Centrum voor Landbouwwiskunde te Wageningen.

In ons land heeft de Hart (12) in een kortgeleden gepubliceerd rapport een volledige uiteenzetting daarvan gegeven.

Bij het onderhavige onderzoek naar de invloed van het weer op de opbrengst van aardappelen is nu voor het eerst h.t.l. van deze methode gebruik gemaakt en daarmede afgestapt van de tot nog toe gebruikelijke "klassieke" correlatie- en regressie-methode.

Alvorens echter de resultaten van het onderzoek te vermelden, is het wenselijk een, zij het vluchtige, indruk te geven van wat de methode-Fisher te bieden heeft en van de wijze waarop zij is opgebouwd.

2. De methode van de regressie-integralen.

Deze methode onderscheidt zich fundamenteel van de klassieke correlatie- en regressie-methode door de grondgedachte, dat de invloed van een weerfaktor, b.v. van de hoeveelheid regen, gedurende een groeiperiode op de opbrengst van een gewas een continue functie van de tijd is, d.w.z. de invloed van de regen zal op twee opeenvolgende tijdstippen relatief weinig verschillen.

Nemen wij b.v. aan dat voor de invloed van een weerfaktor in mei 0 wordt gevonden, doch voor die in juni een belangrijk bedrag, dan is het niet moeilijk ons voor te stellen dat deze invloed continu toeneemt gedurende mei en juni. Hiermede wil natuurlijk niet gezegd zijn, dat er geen bepaalde kritieke momenten kunnen zijn waarop de invloed maximaal is; er wordt slechts mee bedoeld, dat deze geleidelijk, zij het meer of minder snel, bereikt en overschreden zullen worden. Indien men de tijdsintervallen maar voldoende klein neemt, zal dit moeten blijken.

De correlatie- en regressie-methode houdt met de genoemde gedachtengang geen rekening. Zoals bekend geeft zij een lineaire regressie-vergelijking van de gedaante:

y' (=berekende opbrengst) $= b_1x_1 + b_2x_2 + \ldots\ldots + b_nx_n + c$, waarin b_1 , b_2 enz. de regressie-coëfficiënten voorstellen en x_1 , x_2 enz. de weerfaktor in de n intervallen, waarin de groeiperiode is onderverdeeld (eenvoudigheidshalve is hier slechts één faktor beschouwd).

Op grond van de vorenvermelde hypothese behoort het ware verband te zijn van de vorm:

$$y' = c + \int_a^b \beta \, x \, dt \qquad 1)$$

Hierin is x dt de weerfaktor in een klein tijdsinterval dt en β een continue functie van de tijd, of juister van het ontwikkelingsstadium van het gewas. De integratie geschiedt over de gehele groeiperiode a - b .

Aangezien zowel β als x een functie van de tijd is, kunnen wij daarvoor schrijven f(t) resp. g(t); de regressie-integraal wordt dan:

$$y' = c + \int_a^b f(t).g(t) \, dt \qquad 2)$$

Hoe moeten wij ons nu de functie β voorstellen? Beschouwen wij de weerfaktor regenval, dan is β niet anders dan de invloed op de opbrengst (in kg/ha) van elke mm regen die op een bepaald tijdstip valt. De bedoeling is deze functie van t uit onze gegevens te berekenen.

Het is à priori niet te verwachten dat deze functie erg ingewikkeld of aan snelle veranderingen onderhevig is, evenmin dat zij gedurende de vegetatieperiode een aantal maxima en minima vertoont, aangezien de ontwikkeling van het gewas zelf immers een geleidelijk verlopend proces is.

Fisher betrok bij zijn onderzoek uitsluitend de weerfaktor regenval, waarbij hij het jaar verdeelde in 61 perioden van 6 dagen. Het doel nu om de regenhoeveelheid en haar verloop over de tijdsintervallen in enkele constanten uit te drukken, wordt bereikt door het neerslagverloop gedurende het gehele jaar, d.i. dus op grond van 61 waarnemingen tot een vloeiende lijn te vereffenen.

Het is niet de bedoeling in dit bestek deze procedure uitvoerig uiteen te zetten. In het kort komt het er op neer, dat de mathematische bewerking van de regressie-integralenmethode in twee etappes wordt uitgevoerd:

1. de berekening van een bepaalde uitdrukking voor $x = g(t)$, voor elk van de jaren waaruit de basisreeks bestaat.
2. de berekening van $\beta = f(t)$ uit de gegevens ad 1.

De eerste stap die wij dus hebben te doen is voor $x = g(t)$ een analytische uitdrukking te geven, die een zo goed mogelijke benadering vormt van de wijze waarop de uit de weertabellen af te lezen gegevens met de tijd verlopen (b.v. de dekade-gemiddelden). Daarbij wordt $g(t)$ ontwikkeld als een machtreeks in t , waarbij gebruik gemaakt wordt van orthogonale veeltermen. Op deze wijze ontstaat een functie, die in de meest eenvoudige vorm er als volgt uitziet:

$$x' = g(t) \approx a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \dots \; a_k t^k$$

waarin x' de vereffende weerfaktor, b.v. de regenval en t de tijdsintervallen voorstelt, terwijl a_0, a_1 enz. de coëfficiënten van het regenverloop zijn.

Het is duidelijk dat hoe groter k, hoe nauwkeuriger de benadering van dit verloop is.

Bij zijn onderzoek heeft Fisher gemeend voor een voldoende nauwkeurigheid te kunnen volstaan met een 5e.graads benadering, zodat een veelterm met 6 termen in het rechterlid werd verkregen.

Door deze aanpassing nu voor alle jaren van de basisreeks uit te voeren, wordt de variabele "weerfaktor" dus herleid tot 6 series grootheden, elk bestaande uit evenveel termen als er jaren bij het onderzoek betrokken zijn (in het geval van n jaren, komen er dus n a_o-waarden, n a_1-waarden, enz.). Dit betekent dat, waar wij volgens de normale correlatie-methode met 61 variabelen zouden moeten werken, nu met slechts 6 kan worden volstaan.

In meer algemene vorm kunnen wij de vorengenoemde functie-uitdrukking als volgt schrijven:

$$x' = g(t) \approx p_o \varphi_o(t) + p_1 \varphi_1(t) + p_2 \varphi_2(t) + \ldots\ldots p_k \varphi_k(t) \qquad 3)$$

Hierin stelt $\varphi_i(t)$ voor een orthogonale veelterm van de graad i in t. De p's zijn de bijbehorende coëfficiënten; zij karakteriseren de weerfaktor en hebben uiteraard voor elk jaar doorgaans een andere waarde.

Worden voor de berekeningen dekadegemiddelden gebruikt gedurende een tijdvak van b.v. 15 dekaden (groeiseizoen april t/m augustus), dan kan t de waarden 1 t/m 15 aannemen. Voor t = j vinden wij dan bij benadering het gemiddelde van de j^{de} dekade uit dat tijdvak.

De functie $\beta = f(t)$ is eveneens door middel van een orthogonale veelterm als een machtreeks in t voor te stellen volgens:

$$\beta = f(t) \approx q_o \varphi_o(t) + q_1 \varphi_1(t) + q_2 \varphi_2(t) + \ldots\ldots q_k \varphi_k(t) \qquad 4)$$

waarin de q's wederom de coëfficiënten zijn welke β karakteriseren. Deze q's zijn elk jaar hetzelfde; β is een functie van de tijd.

Uit deze vergelijking vinden wij voor t = j de grootte van de regressie-coëfficiënt in de j^{de} dekade.

Nu is $y' = c + \int_a^b \beta\, x\, dt = c + \int_a^b f(t).g(t)$ bij benadering voor te stellen door:

$$y' \approx c + \int_a^b \left\{ p_o \varphi_o(t) + \ldots\ldots p_k \varphi_k(t) \right\} \left\{ q_o \varphi_o(t) + \ldots\ldots q_k \varphi_k(t) \right\} dt \qquad 5)$$

Bij toepassing van de orthogonale eigenschap

$$\int_a^b \varphi_i(t)\, \varphi_j(t)\, dt \begin{cases} = 0, \text{ voor } i \neq j \\ = 1, \text{ voor } i = j \end{cases} \qquad 6)$$

gaat 5) na uitwerking over in:

$$y' \approx c + q_o p_o + q_1 p_1 + \ldots\ldots q_k p_k \qquad 7)$$

Het blijkt dan dat de coëfficiënten q_o t/m q_k uit 4) de regressie-coëfficiënten zijn in het regressieverband tussen i j en p_o t/m p_k.

Uit deze coëfficiënten is tenslotte de waarde van β te berekenen, waarmee in principe het vraagstuk betreffende het verloop van de invloed van een bepaalde weerfaktor op het gewas is opgelost.

Bij het onderhavige onderzoek, ondersteld eenvoudigheidshalve dat met één weerfaktor wordt gewerkt, is voor de regressie-analyse gebruik gemaakt van matrices. Het voordeel daarvan is, dat men het resultaat van de analyse in een vorm verkrijgt, waarin het direkt mogelijk is de signifikantie van de diverse coëfficiënten van $\beta = f(t)$ te toetsen.

Voor elk der k+2 coëfficiënten (zie 7)) wordt dan n.l. een grootheid verkregen welke, gekwadrateerd, dat deel van de variantie van de opbrengst voorstelt, dat door de betreffende coëfficiënt wordt "verklaard".

Worden deze grootheden in mindering gebracht van de oorspronkelijke variantie van de opbrengst, dan houdt men het deel van deze variantie over, dat niet door de grootheden uit de regressievergelijking wordt verklaard, de z.g. restvariantie.

Bij de toetsing wordt nu als eis voor een signifikantie van een bepaalde coëfficiënt gesteld dat de daardoor verklaarde variantie statistisch groter dan de restvariantie moet zijn. Is dit niet het geval, dan heeft de betreffende coëfficiënt blijkbaar geen invloed op de opbrengst.

Door deze signifikantie-toetsing is de graad van de functie $\beta = f(t)$ te bepalen.

Een ander voordeel van het gebruikte rekenschema is dat men de totale correlatie-coëfficiënten op eenvoudige wijze uit de rechtstreeks uit het schema afgeleide restvariantie en de oorspronkelijke variantie te berekenen zijn.

Bij het onderzoek is de methode voor elke weerfaktor afzonderlijk toegepast. Uiteraard is het mogelijk twee of meer weerfaktoren tegelijk te beschouwen. Dit kan echter tot onnodig veel rekenwerk leiden, wanneer er achteraf van een of meer van deze faktoren geen invloed op de opbrengst aanwezig mocht blijken te zijn.

Eerst wanneer dus op de vorengeschetste wijze van elk der daarvoor in aanmerking komende weerfaktoren afzonderlijk de invloed is onderzocht, wordt tenslotte van de gevonden signifikante faktoren de gezamenlijke invloed nagegaan, waarna eventueel een regressievergelijking, als een oogstverwachtingsformule, kan worden opgesteld.

Voor een volledig inzicht in de methode kan kortheidshalve worden verwezen naar de eerder vermelde uiteenzetting van de Hart.

3. De landbouwkundige gegevens.

Aangezien het onderzoek slechts ten doel had na te gaan of de daarbij thans in ons land voor het eerst toegepaste methode van de regressie-integralen tot betere (meer betrouwbare) resultaten zou leiden dan de tot dusver gebruikelijke correlatie- en regressiemethode, werd het beperkt tot een enkel bedrijf, n.l. de Veenkoloniale proefboerderij Borgercompagnie bij Veendam. Dit biedt overigens in vergelijking met een onderzoek over een groot gebied o.m. de volgende voordelen:

1. De opbrengst-gegevens zijn betrouwbaar; zij zijn n.l. niet gebaseerd op

schattingen zoals deze voor de grote landbouwgebieden plegen te worden verricht en als zodanig in de Verslagen en Mededelingen van de Directie van de Landbouw jaarlijks worden gepubliceerd.

2. Men is gevrijwaard van de invloed van rasverschillen, aangezien men zich tot één enkel ras kan bepalen.

3. Bodemverschillen treden praktisch niet op.

4. Men heeft evenmin te maken met verschillen in cultuurmaatregelen. Een behoorlijke verzorging van het gewas is steeds gewaarborgd.

Het onderzoek is verricht aan het (standaard)ras Eigenheimer en omvat de periode 1920 t/m 1958, met uitzondering van de jaren 1949 en 1953, waarin dit ras door omstandigheden niet in de rassenproeven werd opgenomen.

Fig.1 geeft het verloop weer van de opbrengstcurve in de betreffende periode. De gemiddelde ha-opbrengst bedraagt 37.2 ton, variërend van 51.0 ton in 1939 tot 26.1 ton in 1956.

4. De meteorologische gegevens.

Het zou ideaal geweest zijn de opbrengstgegevens van de proefboerderij Borgercompagnie in verband te kunnen brengen met de weersomstandigheden ter plaatse, waardoor men over nauwkeurig vergelijkbaar materiaal zou beschikken.

Aangezien dit echter niet mogelijk was, moesten de weergegevens worden ontleend aan stations in de naaste omgeving, waarbij stilzwijgend werd aangenomen dat zij ook voor het betreffende bedrijf representatief zijn.

Bij het onderzoek zijn slechts drie weerfaktoren betrokken, welke als verreweg de belangrijkste zijn te beschouwen, t.w. neerslag, temperatuur en zonneschijn.

Met betrekking tot de neerslag zijn de gegevens gebruikt van het meest nabij gelegen regenstation Veendam; die van de beide andere weerselementen zijn afkomstig van het klimatologische hoofdstation Eelde (Groningen).

De gegevens betreffende het aantal regendagen (met minstens 1,0 mm regen), waarvan eveneens de invloed op de opbrengst werd onderzocht, zijn gemiddelden van de waarnemingen van de stations Eelde en Witteveen; die van het regenstation Veendam zelf waren n.l. niet in bruikbare vorm beschikbaar.

De betreffende weerfaktoren zijn nader als volgt onderscheiden:

Neerslag : sommen per dekade
: sommen per pentade
: aantal regendagen per maand
: aantal regendagen per 2 dekaden

Temperatuur	: overdaggemiddelde	
	: dagelijkse maximum	) per dekade
	: dagelijkse amplitudo	)
Zonneschijn	: percentage per dekade	

De figuren 2 en 3 geven een overzicht van de dekadegemiddelden van de regenval en de zonneschijn resp. van de overdagtemperatuur, de dagelijkse maximum-temperatuur en de dagelijkse temperatuur-amplitudo van de reeks 1920-1958 (excl. de jaren 1949 en 1953) gedurende de vegetatie-periode van de aardappel (april t/m augustus) voor het klimaat-gebied, waartoe de proefboerderij geacht wordt te behoren.

In fig. 2 is de regenval over het tijdvak maart - augustus weergegeven, omdat ook de periode vóór het begin van de teelt (gemiddeld medio april) in het onderzoek werd betrokken. In de fig. is voorts het aantal uren zonneschijn opgenomen, zulks slechts ter vergelijking met het percentage.

5. De resultaten van het onderzoek.

5.1 De gevolgde werkwijze.

Aangezien rekening moest worden gehouden met de mogelijkheid dat de variatie in de opbrengsten (zie fig. 1) mede het gevolg is van andere dan weerfaktoren, die derhalve storend zouden werken, werd het eigenlijke onderzoek voorafgegaan door een toetsing op de aanwezigheid van een z.g. trend en van een invloed van de pootdatum.

Overeenkomstig de verwachting is er van een trend geen sprake; een correlatie met de jaren geeft r = -0.01 .

Van de pootdatum is evenmin een aanwijsbare invloed aanwezig. De gemiddelde pootdatum van de 37-jarige reeks is 14 april, met een spreiding van 12 dagen. De vroegste is 14 maart, de laatste 13 mei. De correlatie tussen opbrengst en pootdatum bedraagt 0.14, en is dus verre van significant. Zij is het ook niet, als de na eliminatie van de invloedrijke faktoren verkregen opbrengsten met de pootdata worden gecorreleerd.

Het gemis aan ervaring met de nieuwe, althans in ons land tot dusver onbekende methode was oorzaak, dat aanvankelijk bewerkingen werden uitgevoerd, die achterwege hadden kunnen worden gelaten. Er kon n.l. achteraf worden aangetoond, dat het voor een eerste oriëntatie voldoende zou zijn geweest de weerfaktor als een eerste-graadsfunctie van de tijd uit te drukken. Eerst wanneer op grond hiervan aanwijzingen worden gevonden omtrent de invloed, heeft het zin tot een meer gedetailleerde bewerking over te gaan en dienovereenkomstig $x = f(t)$ verder te ontwikkelen tot een hogere graad in t.

Op deze wijze werden de weerfaktoren: zonneschijnpercentage en maximumtemperatuur per dekade en de neerslagsom per pentade behandeld.

De neerslaghoeveelheid per dekade werd onmiddellijk voorgesteld door een functie van de 5e graad; de overdagtemperatuur en de dagelijkse temperatuuramplitudo door een van de 4e graad.

Ten aanzien van de aantallen regendagen, die door eenvoudige hele getallen zijn weergegeven, werd aanvankelijk ter oriëntering slechts een meervoudige regressie berekend, n.l. met de aantallen per maand. Op grond van de daarmee verkregen uitkomsten werd er naderhand toch toe overgegaan gebruik te maken van regressie-integralen, teneinde de signifikantie van de invloed na te kunnen gaan.

De weerfaktor die het eerst in het onderzoek werd betrokken, was de regenval per dekade, waarvan een signifikante invloed werd gevonden. Voor het onderzoek van de overige faktoren werd gebruik gemaakt van opbrengstcijfers waaruit de neerslaginvloed was geëlimineerd, d.w.z. van de afwijkingen van de oorspronkelijke (werkelijke) opbrengsten t.o.v. de uit de regressie neerslag-opbrengst berekende waarden.

Naderhand werd onderzocht of de neerslaginvloed beter uit pentade - dan uit dekadesommen zou zijn te berekenen. Dit bleek evenwel niet het geval.

Tenslotte werd nagegaan of de gevonden invloed van de neerslagsom en van de dagelijkse temperatuuramplitudo ook kwadratisch zou zijn aan te tonen. Dit is te doen door in de regressie-berekening van de opbrengst met de coëfficiënten (p_0, p_1 .... p_k) van de weerfunctie $x = g(t)$ als extra variabele het kwadraat van de nulde-graadscoëfficiënt (p_0^2) in te voeren.

5.2 De invloed der weerfaktoren.

Van de onderzochte weerfaktoren is slechts van een drietal signifikant of bij benadering signifikant invloed geconstateerd, t.w.:

1. de neerslaghoeveelheid per dekade
2. het aantal regendagen per maand
3. de dagelijkse temperatuuramplitudo per dekade.

De invloed van de eerstgenoemde faktor op de opbrengst is werkelijk signifikant; die van de overige faktoren slechts bij benadering.

1. <u>De neerslaghoeveelheid.</u>

Van de functie van de dekadesom geeft de coëfficiënt van de eerste graad signifikantie. M.a.w. de regressiecoëfficiënt is een lineaire functie van de tijd en wel in die zin dat de invloed van de regenval gedurende het groeiseizoen toenemend negatief is. De correlatiecoëfficiënt $R = 0.40$, de 5 % betrouwbaarheidsdrempel = 0.32 .

Het gebruik van pentadesommen geeft geen aanleiding tot verbetering. Wèl is ook hierbij de eerste-graadscoëfficiënt signifikant, doch $R = 0.36$.

De invloed van de regenval zou uit praktijk overwegingen geacht moeten worden niet lineair te zijn, d.w.z. als hij b.v. kwadratisch zou zijn, zou de regressie-coëfficiënt behalve van de tijd tevens een lineaire functie van de grootte van de neerslaghoeveelheid moeten zijn. Het kwadratisch karakter van de invloed is echter niet statistisch signifikant aantoonbaar.

2. Het aantal neerslagdagen.

De neerslagdagen zijn discrete grootheden, zodat het theoretisch niet juist is deze voor een regressieberekening te gebruiken. Ten behoeve van een oriëntatie omtrent een eventuele invloed van deze faktor is deze bewerking nochthans uitgevoerd.

Een meervoudige regressie-berekening van het aantal regendagen in elke maand van het groeiseizoen met de opbrengsten toont aan dat deze faktor in de maanden april en juli in het geheel geen invloed heeft, in de overige maanden slechts een geringe. In het laatste geval is n.l. R = 0.39 (5 % drempel = 0.33) .

Bij gebruikmaking van de opbrengstgegevens waaruit de neerslaginvloed is geëlimineerd blijkt evenwel, dat ook de invloed in augustus te verwaarlozen is, terwijl voor mei en juni zelfs een overschrijdingskans van juist iets meer dan 5 % wordt gevonden. Dit zou dus in feite betekenen, dat de faktor regendagen op zichzelf geen, doch in combinatie met de neerslaghoeveelheid wèl invloed op de opbrengst heeft.

Aangezien mogelijkerwijze de invloed toch niet geheel te verwaarlozen is, is deze nader onderzocht met behulp van regressie-integralen. Daarbij is gebruik gemaakt van het aantal regendagen per twee dekaden, telkens over een dekade verschoven.

Zonder eliminatie van de invloed van de neerslagsom uit de opbrengstcijfers is signifikantie voor de coëfficiënt van de 3e-graad gevonden. Daarbij is R = 0.43, een resultaat dat vrijwel overeenkomt met dat voor het aantal regendagen per maand, terwijl in dit geval 4 variabelen (0^e t/m 3^e graad) nodig zijn.

De invloed, zo men hieraan betekenis wil hechten, is evenals die van de neerslagsom negatief.

3. De dagelijkse temperatuuramplitudo.

De invloed van deze faktor is eveneens gering. Wanneer direkt met de oorspronkelijke opbrengsten wordt gecorreleerd, vinden wij geen signifikante invloed. Is evenwel de neerslaginvloed uit de opbrengsten verwijderd, dan is de 1^e-graads-coëfficiënt zwak signifikant.

Een dergelijke invloed is nagenoeg te verwaarlozen, temeer daar de variantie, die op rekening van de coëfficiënt van de 0^e-graad komt, zeer gering is. Dat laatste is oorzaak dat de variantie op rekening van de termen van 0e- en 1e-graad tezamen nog onder de signifikantie-drempel blijft.

Zo men m.b.t. deze weerfaktor van invloed mag spreken, is deze gemiddeld over de gehele vegetatieperiode zeer zwak negatief met een tendens tot toeneming in negatieve zin in de loop van het groeiseizoen.

Voor het bestaan van een kwadratische invloed is geen enkele aanwijzing gevonden.

In onderstaand schema is een overzicht gegeven van de vorenbesproken faktoren die tenslotte gezamenlijk voor een regressievergelijking zijn gebruikt, met vermelding van de variantie die voor elk dier faktoren wordt "verklaard". Men vindt hieruit R = 0.62 , bij 27 vrijheidsgraden.

Faktor	Graad van $x = g(t)$	Variantie	Aantal vrijheids graden	Toetsingsgrootheid $Z(=\frac{1}{2}\lg\frac{var}{restvar})$	$Z(5\%)$*
Neerslagsom per dekade	1	104,48	2	0,895	0,605
" per 2 dekaden	3	44,25	4	0,474	0,502
Dag.temp.ampl.per dekade	1	46,51	2	0,491	0,605
Neerslagsom kwadratisch	-	60,65	1	0,624	0,719
Rest		17,43	27		
Totaal		28.06	36		

* Z(5%) is de signifikantiedrempel. Een faktor wordt geacht invloed te hebben, als de bijbehorende variantie signifikant groter is dan de restvariantie. Dit is het geval als $Z > Z(5\%)$.

Resumerende moet worden geconcludeerd, dat de neerslagsom per dekade de enige faktor is die een signifikante invloed op de opbrengst heeft. Van de overige in het schema opgenomen faktoren is de invloed twijfelachtig te noemen.

Op grond van de verkregen uitkomsten is het derhalve niet mogelijk een regressievergelijking op te stellen met behulp waarvan met enige nauwkeurigheid een oogstverwachting is uit te spreken. Dit kan als volgt worden toegelicht:

Het verband tussen de spreiding van de oorspronkelijke variabele (Si j) en de spreiding van het verschil tussen deze variabele en de uit de regressievergelijking verkregen schatting (Si j-rest) is voor te stellen door:

$$S_{ij}\,rest = \sqrt{1 - R^2}\, . \,S_{ij}$$

Een regressievergelijking heeft eerst enige waarde voor de praktijk, wanneer de restvariantie slechts een kleine fractie van de oorspronkelijke bedraagt. Het zal echter duidelijk zijn dat in het onderhavige geval daarvan geen sprake is, zelfs al zouden alle variabelen uit het schema worden gebruikt. Het verband wordt dan immers:

$$Sij\ rest = \sqrt{1 - R^2}\ .\ Sij \approx 0{,}8\ Sij$$

De spreiding in de aardappelopbrengsten van het proefbedrijf Borgercompagnie bedraagt 5,3 bij een gemiddelde opbrengst van 37,2 ; de restvariantie is 4,2 (een en ander uitgedrukt in ton/ha).

Zou nu met de genoemde faktoren een regressievergelijking worden opgesteld, dan is met 95 % waarschijnlijkheid te zeggen dat de daarmee berekende opbrengsten t.o.v. de werkelijke een afwijking zullen geven van ± 2.4,2 = ± 8,4 ton/ha.

Met dezelfde waarschijnlijkheid geldt, dat de afwijkingen van de werkelijke opbrengsten t.o.v. het gemiddelde van de basisreeks binnen ± 2.5,3 = ± 10,6 ton/ha liggen.

Dit betekent dus, dat in doorsnee een met de regressievergelijking berekende waarde niet belangrijk minder van de werkelijke zal afwijken dan het verschil bedraagt tussen de werkelijke- en de gemiddelde opbrengst van de reeks.

Dit blijkt inderdaad wanneer wij de regressievergelijking toepassen voor een schatting van de oogst 1959. De vergelijking is:

$$y' = -13{,}8+4{,}407a_0-0{,}919a_1-0{,}1008a_0^2-0{,}825b_0+12{,}75b_1+3{,}47b_2-73{,}39b_3+1{,}623c_0-$$
$$-94c_1$$

waarin de a's, de b's en de c's de coëfficiënten voorstellen van resp. het neerslagverloop per dekade, de neerslagdagen per 2 dekaden en de dagelijkse temperatuuramplitudo per dekade, terwijl a_0^2 de term is voor de kwadratische regressie met de neerslagsom per dekade.

Voor 1959 hebben de coëfficiënten de volgende waarden: $a_0 = 11{,}5$; $a_1 = -0{,}13$; $a_0^2 = 132{,}25$; $b_0 = 4{,}5$; $b_1 = 0{,}09$; $b_2 = 0{,}11$; $b_3 = -0{,}02$; $c_0 = 10{,}7$; $c_1 = 0{,}16$.

Substitueren wij deze waarden in de vergelijking, dan krijgen wij als uitkomst $y' = 39{,}7$ ton/ha. De werkelijke opbrengst bedraagt evenwel 11,6 ton/ha.

Het blijkt dat 39,7 en 11,6 veel meer van elkaar verschillen dan het dubbele van de restvariantie 4,2

6. Slotbeschouwing.

Een vergelijking van de methode van de regressieintegralen met de "klassieke" correlatie- en regressiemethode leert, dat in beide gevallen de berekening tenslotte neerkomt op een correlatieberekening. Tegenover deze overeenkomst staat evenwel het wezenlijke verschil, dat de eerstgenoemde methode gebruik maakt van de wetenschap dat de invloed van de weerfaktoren een vloeiend verloop moet vertonen.

Beide methoden hebben overigens hun voor- en nadelen:

A. De lineaire correlatiemethode is niet zo bewerkelijk en is vrij eenvoudig uit te voeren. De nadelen zijn echter de volgende:

1. Aangezien veelal met dekadegemiddelden wordt gewerkt, houdt dit in dat men per weerfaktor voor elk groeiseizoen (van 5 maanden) met 15 variabelen te maken heeft. Bij een dergelijk aantal is een toevallig optredende hoge correlatie niet denkbeeldig.
2. De aanwezigheid van een invloed die zich over de gehele groeiperiode uitstrekt, doch per dekade slechts zwak is, wordt er niet mee onderkend. Ook wanneer alle dekadegemiddelden tegelijk voor een meervoudige regressie zouden worden gebruikt, zou dit toch zoveel vrijheidsgraden kosten, dat de restvariantie nauwelijks signifikant vermindert en dientengevolge de invloed niet signifikant kan worden vastgesteld.
3. Zij is niet te gebruiken voor discreet verdeelde variabelen.

B. De methode van de regressie-integralen heeft de volgende voordelen:

1. De kans op een toevallig hoge correlatie wordt aanzienlijk verminderd, aangezien elke weerfaktor immers door slechts een gering aantal variabelen wordt voorgesteld i.p.v. door vijftien (dekadegemiddelden).
2. Het werkelijk aanwezige verband wordt op juistere wijze benaderd.
3. Er wordt een veel intensiever gebruik gemaakt van de beschikbare gegevens, zodat ook minder lange reeksen nog betrouwbare resultaten kunnen geven.
4. Zij is ook te gebruiken voor discreet verdeelde variabelen.

Tegenover de genoemde voordelen zijn de volgende nadelen te stellen:

1. De toepassing vereist vrij veel rekenwerk en meer wiskundig-statistische kennis.
2. Zij is slechts te gebruiken wanneer men te maken heeft met invloeden die geleidelijk over een langere periode verlopen. Een kortstondige invloed kan er niet mee worden ontdekt.

Ten aanzien van de toepassing van de methode der regressie-integralen moet de aandacht nog op het volgende worden gevestigd:

Bij gewassen met een kortere vegetatieduur, zoals b.v. bij tabak in de tropen, zullen de veranderingen in de waarde van β waarschijnlijk sneller plaatsvinden. Hieraan zal tegemoet dienen te worden gekomen door de tijdsintervallen te verkleinen, dus door b.v. met tijdseenheden van 5 dagen te werken. Een verdere onderverdeling zal doorgaans op berekeningstechnische bezwaren stuiten.

Het verdient aanbeveling voor elk te onderzoeken geval na te gaan waar het natuurlijke begin van de teelt ligt. Immers de voorgeschiedenis, zoals b.v. de toestand waarin de grond verkeert voordat met de teelt wordt begonnen, moet evenseens als een functie van de regenval worden beschouwd. Deze overweging heeft er toe geleid dat in het onderhavige onderzoek de invloed van de weerfaktor neerslag over de periode maart t/m augustus werd onderzocht i.p.v. over het eigenlijke groeiseizoen april t/m augustus.

Tenslotte moet worden opgemerkt, dat de thans bij ons onderzoek verkregen teleurstellende resultaten geen aanleiding behoeven te zijn de methode van de regressie-integralen te verwerpen. Deze maakt immers onder de gegeven omstandigheden een veel intensiever gebruik van de beschikbare gegevens dan het met de methode der lineaire correlatie mogelijk zou zijn. Het ligt meer voor de hand om op grond van de grote restvariantie te concluderen dat er nog andere dan de onderzochte faktoren zijn, die het betreffende probleem beheersen, doch waarvan wij niet over gegevens beschikken.

Zeer belangrijk is ongetwijfeld het feit dat de regressie-coëfficiënten als functie van de tijd worden beschreven, terwijl het reëler zou zijn ze te beschouwen als functies van de opeenvolgende fenologische stadia. Het gebrek aan fenologische gegevens betekent voor het onderzoek uiteraard een beperking, die een ons onbekend verlies van informatie tengevolge heeft.

De Bilt, juni 1960.

Literatuuropgave.

1. 1924 R.A. Fisher : The influence of rainfall on the yield of wheat at Rothamsted. Phil.Trans.Royal Soc. of London. Series B. Vol. 213.

2. 1932 H.J. Frankena: Een statistisch onderzoek naar de invloed van het weer op de opbrengst en het gehalte van suikerbieten in Nederland. Diss.Wageningen.

3. 1946 J.P.M. Woudenberg: Verband tussen het weer en de opbrengst van wintertarwe. Med. en Verh. K.N.M.I. No.102. 50.

4. 1949 J.C. Ignatius en W. de Wit: Onderzoek naar de invloed van het weer op de appel- en perenoogst. Landbouwk.Tijdschrift 3, 153-167.

5. — J.J. Post: Statistisch onderzoek naar de samenhang tussen het weer, de grasproductie en de melkaanvoer. Diss. Wageningen. Med. en Verh. K.N.M.I. (A) 55.

6. 1952 J.J. Post: Het verband tussen het weer en de opbrengst van tuinbouwzaden (wortelzaad). Med. Dir. v.d. Tuinb. 15, 218-222.

7. — C. Kramer, J.J. Post en W. Wilten: Klimaat en brouwgerst in Nederland. Med. en Verh. K.N.M.I., 102, (A) 57.

8. 1954 W. v.d. Bijl: Rapport over de statistische methoden in gebruik bij het onderzoek naar het verband tussen meteorologische factoren en opbrengsten van land- en tuinbouwgewassen en melk. Rapporten K.N.M.I. R III-140

9. — F.H. Sanderson: Methods of crop forecasting.

10. 1955 R.F. Fisscher: Bijdrage tot de kennis omtrent de samenhang tussen het weer en de opbrengst van haver in Nederland (onderzoek verricht i.h.kader v.h. z.g. Plan Tendeloo).

11. 1957 J.P.M. Woudenberg en P. Poelstra: Samenhang tussen weerfactoren en de opbrengst van zomertarwe. Verslagen K.N.M.I. V.19 R.III-210.

12. 1960 H. de Hart: De methode der regressie-integralen ter bepaling van de invloed van weerfaktoren op de opbrengst van gewassen. V-66 R.III-252.

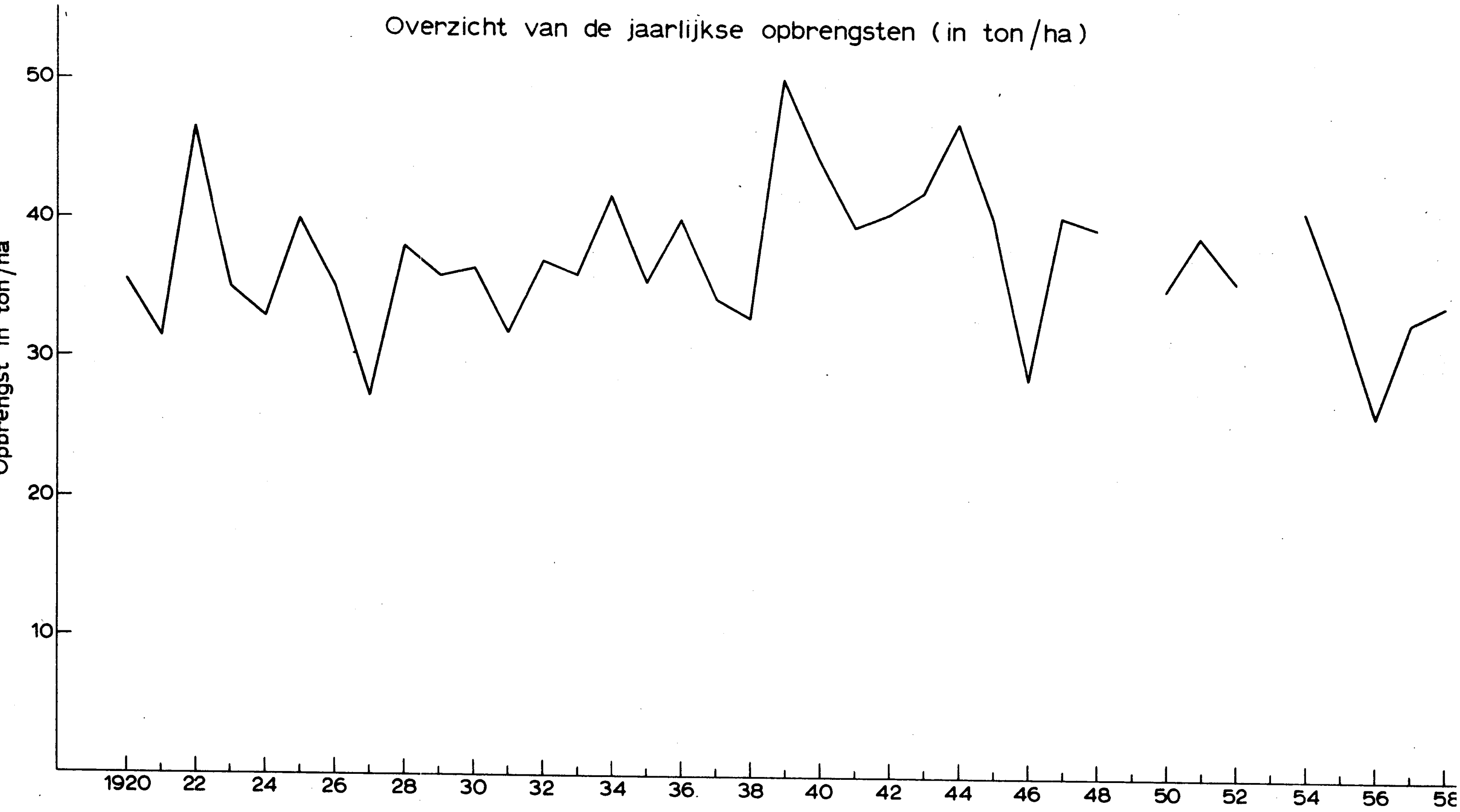
Overzicht van de jaarlijkse opbrengsten (in ton/ha)
Opbrengst in ton/ha
50
40
30
20
10
1920
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56

fig. 1

Dekadegemiddelden (1920 - 1958) van de regenval, het aantal uren - en het percentage zonneschijn gedurende de vegetatieperiode

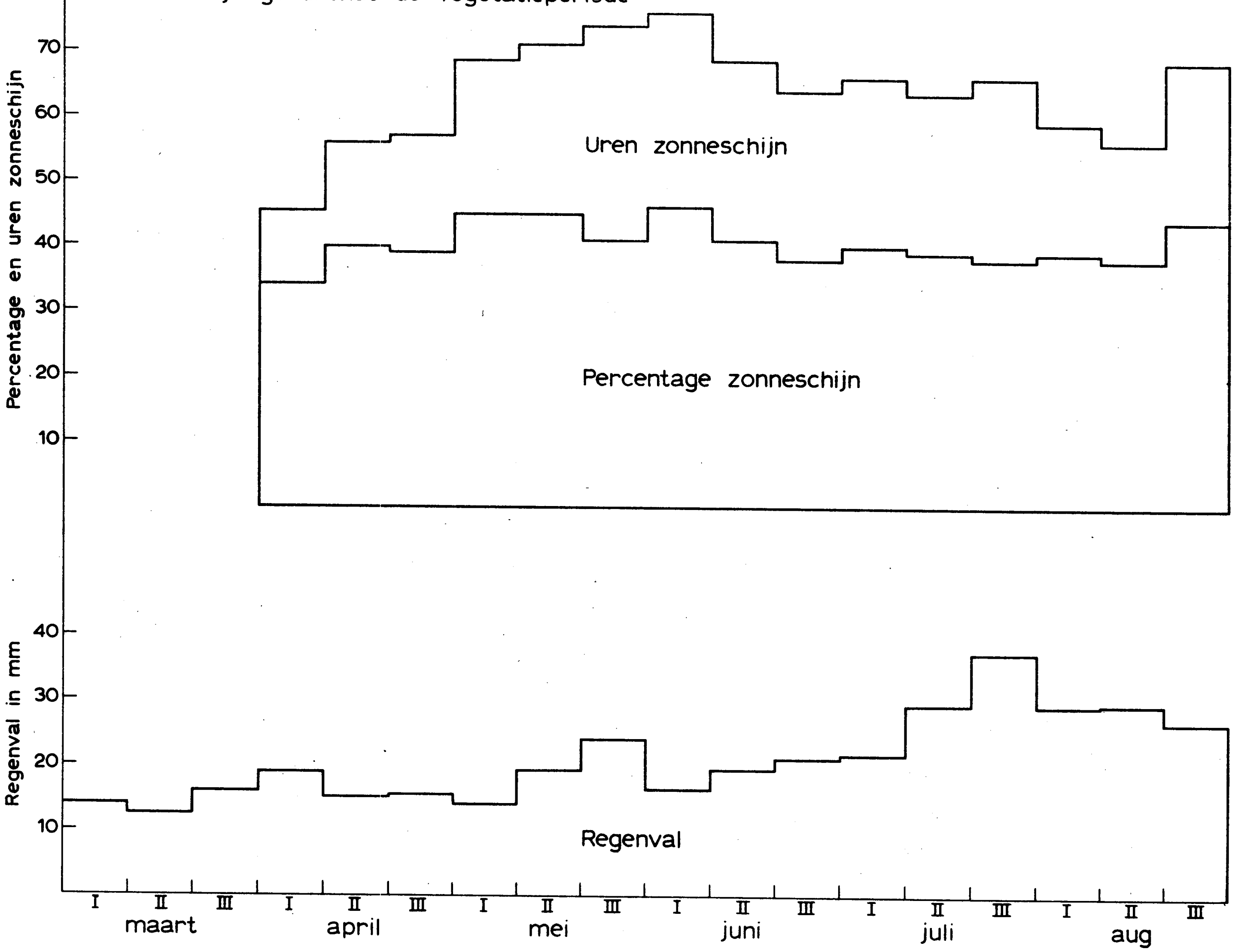

fig. 2

Dekadegemiddelden (1920-1958) van de overdagtemperatuur (——), maximum temperatuur (– – –) en de dagel. amplitudo (–·–·–) gedurende de vegetatieperiode

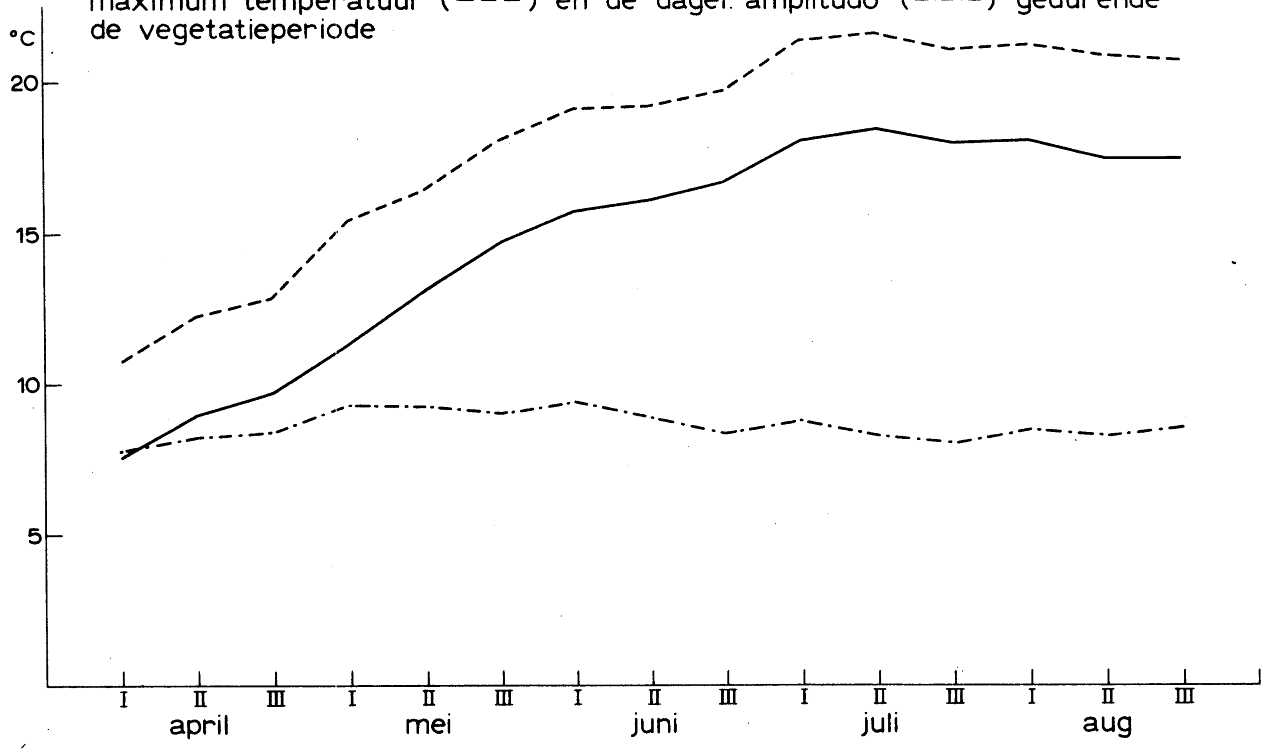

fig. 3