

Nota 103 d.d. 29 november 1961

Statistische analyse van de werkelijke verdamping

door A.H. Ryhiner en L. Kamil

1. Doel van het onderzoek:

Voor het sluitend krijgen van de waterbalans zal het veelal nodig zijn één van de twee onbekenden uit de waterbalansvergelijking te leren kennen. Het lag voor de hand om met de gegevens van de weegbare lysimeterinstallatie te Wageningen, waar alle termen uit de waterbalansvergelijking - hetzij direct of indirect - meetbaar zijn aan een statistische analyse te onderwerpen, met de veronderstelling, dat er steeds een functie te vinden zal zijn die zich zo goed mogelijk aan de waarnemingen aanpast.

Het doel van dit onderzoek is dan ook om uitgaande van de waterbalansvergelijking

$$E_w = N + I - A \pm \Delta V \quad (1)$$

met de meetbare grootheden uit deze vergelijking de werkelijke verdamping te verklaren.

2. Bespreking variabelen:

De termen neerslag (N) en de infiltratie min afvoer (I-A) zijn algemeen bekend zodat hierop niet nader behoeft worden ingegaan. Het feit dat de waarnemingen voor dit onderzoek betrekking hebben op zandlysimeter 4 van de weegbare lysimeterinstallatie kon door middel van wegingen zowel de vochtverandering in het profiel (ΔV) als de werkelijke verdamping (E_w) worden bepaald.

In (1) is de E_w lineair afhankelijk van de N, I-A en ΔV en zou dus de werkelijke verdamping reeds met deze factoren volledig verklaard kunnen worden. Teneinde dit te ondervangen werd voor de ΔV , de gesommeerde vochtonttrekking ($\Sigma \Delta V$) in de vergelijking gesubstitueerd.

Aangezien de werkelijke verdamping voor een groot deel verloopt volgens meteorologische gradiënten was het noodzakelijk de open waterverdamp-



1787069

ping (E_0) - als maat voor het verdampend vermogen van de atmosfeer - in het onderzoek te betrekken. Bovendien zal het gewas de werkelijke verdamping mede beïnvloeden en werd de graslengte (G.L) aan de reeks variabelen toegevoegd.

Met uitzondering van $\Sigma \Delta V$ werden de overige variabelen in mm/etm. en de graslengte in cm omgerekend. De $\Sigma \Delta V$ daarentegen geeft aan de vochtonttrekking in het midden van de balansperiode, uitgedrukt in mm water.

Teneinde een beter inzicht te verkrijgen over de uitdroging van het profiel beneden veldcapaciteit en het verloop van de werkelijke verdamping onder minimale vochtomstandigheden werd voor dit onderzoek de uitzonderlijke droge zomer van 1959 in beschouwing genomen. De beschouwde periode loopt van 1 april - het tijdstip waarop de afvoer overgaat in de infiltratie - tot eind oktober; in totaal 49 balansperioden van gemiddeld vier dagen.

3. Verwerkingsmethodiek:

Beschikt men over een aantal paren waarnemingsuitkomsten (x_k, y_k) $k = 1, \dots, n$ welke voldoen aan het model: $y_k = \alpha + \beta x_k + \varepsilon_k$ dan zal volgens de methode der kleinste kwadraten de beste schatting van α en β verkregen worden, wanneer $\Sigma(y - \alpha - \beta x)^2$ minimaal is.

Dit is zo indien:

$$\left. \begin{aligned} \frac{\partial \Sigma(y - a - bx)^2}{\partial a} &= 0 \\ \text{en} \\ \frac{\partial \Sigma(y - a - bx)^2}{\partial b} &= 0 \end{aligned} \right\}$$

De normaal vergelijkingen luiden dan:

$$\begin{aligned} an + b \Sigma x &= \Sigma y \\ a \Sigma x + b \Sigma x^2 &= \Sigma xy \end{aligned} \tag{2}$$

Hieruit zijn dan a en b op te lossen.

De restvariantie wordt geschat door:

$$S(\sigma^2) = \frac{\Sigma(y - a - bx)^2}{n-2} \tag{3}$$

Een criterium voor lineaire afhankelijkheid wordt gegeven door de schatting van de correlatiecoëfficiënt:

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2 \Sigma (y-\bar{y})^2}} \quad (4)$$

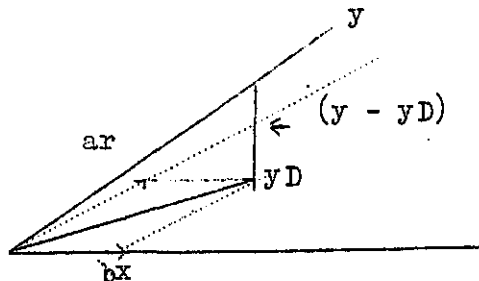
Deze methode leidt tot zogenaamde "beste lineaire schattingen" en vormt als zodanig de basis van de methodiek zoals die op het lysimeteronderzoek wordt toegepast. In het volgende zal deze methode met een tweedimensionaal voorbeeld in het kort worden toegelicht.

De n-componenten van \underline{y} en \underline{x} zijn te beschouwen als coördinaten van de punten \underline{y} en \underline{Y} in een n-dimensionale ruimte. Het model wordt dan:

$$\underline{y} \cong \alpha r + \beta \underline{x} + \sigma \underline{x}_n,$$

waarin r de niveauvector $(1,1, \dots, 1)$ is en \underline{x}_n de n-dimensionale normale stochastiek. Is y_D de projectie van \underline{y} op D dan luidt in meetkundige termen: " \underline{y}_D ligt in de 2-dimensionale deelruimte, die door de vectoren \underline{x} en r wordt opgespannen".

Deze deelruimte wordt D genoemd.



De beste lineaire schattingen worden verkregen wanneer $(\underline{y} - \underline{y}_D)^2 = (\underline{y} - \alpha r - \beta \underline{x})^2$ minimaal is. Het is duidelijk, dat dit het geval is indien $(\underline{y} - \underline{y}_D) \perp D$ dus als $(\underline{y} - \underline{y}_D) \perp \underline{x}$ en $(\underline{y} - \underline{y}_D) \perp r$. Hieruit volgt:

$$\left. \begin{aligned} (\underline{y} - \alpha r - \beta \underline{x}) \underline{x} &= 0 \\ (\underline{y} - \alpha r - \beta \underline{x}) r &= 0 \end{aligned} \right\}, \quad (5)$$

hetgeen leidt tot de normaalvergelijkingen (2).

Uit de hoofdstelling voor normale stochastieken volgt:

$$(\underline{y} - \underline{y}_D)^2 \cong \sigma^2 \chi^2_{n-2},$$

zodat een schatting van σ^2 gegeven wordt door:

$$s(\sigma^2) = \frac{(y-y_D)^2}{n-2} = \frac{(y-ar-bx)^2}{n-2} = \frac{\Sigma(y-a-bx)^2}{n-2}$$

In woorden: De schatting van de restvariantie is het kwadraat van de lengte van de projecterende $(y-y_D)$, gedeeld door de dimensie van de restruimte.

De correlatiecoëfficiënt is de cosinus van de hoek tussen de vooberren $(y-\bar{y})$ en $(x-\bar{x})$:

$$\cos \varphi = r = \frac{(x-\bar{x})(y-\bar{y})}{\sqrt{(x-\bar{x})^2(y-\bar{y})^2}}$$

Bij meervoudige correlatie is de correlatiecoëfficiënt de cosinus van de hoek tussen de vector $(y-\bar{y})$ en zijn projectie op de deelruimte opgespannen door de contrast vectoren $(x_1-\bar{x}_1)$, $(x_2-\bar{x}_2)$, ..., $(x_k-\bar{x}_k)$.

Past men eerst schaalverdeling toe zodat met de vectoren y , x , bedoeld wordt

$$y = \frac{y-\bar{y}}{\sqrt{(y-\bar{y})^2}}, \quad x = \frac{x-\bar{x}}{\sqrt{(x-\bar{x})^2}},$$

dan is de correlatiecoëfficiënt direct te vinden aangezien

$$r = y y_D = y_D.$$

Bij dit onderzoek werd uitgegaan van de veronderstelling, dat het stelsel bestaat uit functies die lineair zijn in de onbekende parameters en de gevraagde functie voldoende benaderd kan worden door een tweede graads polynomium.

Daartoe werd het aantal variabelen volgens onderstaand schema uitgebreid:

	x_1	x_2	x_3	x_4	x_5	
$x_1 = N$	x_6	x_{11}	x_{12}	x_{13}	x_{14}	
$x_2 = I-A$		x_7	x_{15}	x_{16}	x_{17}	
$x_3 = E_0$			x_8	x_{18}	x_{19}	
$x_4 = G.L$				x_9	x_{20}	
$x_5 = \Sigma \Delta V$					x_{10}	$x_{21} = r(1,1,\dots,1)$

Bekend zijn 49 waarden voor $\underline{y} = E_w$, met evenveel waarden van de "verklarende" variabelen (x_i).

De bewerking wordt nu geheel analoog aan gegeven voorbeeld uitgevoerd, doch slechts met dit verschil, dat nu geprojecteerd wordt op een 21-dimensionale deelruimte opgespannen door de vectoren x_1 tot en met x_{21} .

Is X de matrix van 21 verklarende variabelen = kolommen, dan luidt de onderstelling

$$\underline{y} \cong X\beta + \sigma \epsilon \quad 49$$

Schattingen van β worden opgelost uit de normaalvergelijkingen

$$X'\underline{y} = X'X\beta$$

Indien de vectoren \underline{y} , x_1 , x_{20} geprojecteerd worden op de deelruimte D , en bovendien hun lengten op 1 wordt herleidt, waarbij onder x_i wordt verstaan

$$x_i = \frac{x_i - \bar{x}_i}{\sqrt{(x_i - \bar{x}_i)^2}},$$

dan verkrijgt men bij projectie van y op D , opgespannen op de vectoren x_1 tot en met x_{20} , direct de correlatiecoëfficiënten tezamen met de regressiecoëfficiënten, die daarna in hun oorspronkelijke schalen moeten worden teruggerekend.

Als het model goed is gekozen, dan zal de correlatiecoëfficiënt (y_D) dicht bij 1 liggen.

Een schatting van σ^2 volgt uit

$$\underline{s}(\sigma^2) = \frac{(\underline{y} - X\beta)^2}{28} = \frac{\underline{y}^2 - (X\beta)^2}{28} = \frac{1 - y_D^2}{28},$$

hierin is y_D^2 het kwadraat van de projectie van y op de ruimte D , opgespannen op de vectoren x_1 tot en met x_{21} .

4. Resultaat

De aldus berekende multiple correlatiecoëfficiënt bedraagt 0.85. Dit betekent dat slechts 72,6% van de werkelijke verdamping kan worden ver-

klaard. Met de differenties van de gemeten en de berekende verdamping werd vervolgens een grafische analyse uitgevoerd met het doel na te gaan welke complementaire factor(en) mogelijk nog de restvariantie in E_w veroorzaken. Een vrijwel lineair verband werd gevonden met de ΔN , dit is het verschil van de neerslag juist voor het begin en aan het eind van de balansperiode, omgerekend op een gemiddelde regenval in mm/etm. Bovendien werden enkele verdampingswaarden, die extreem afwijken ten opzichte van de verdampingswaarden van de andere lysimeters en kennelijk een gevolg zijn van weegfouten, gecorrigeerd. Op grond hiervan werd de bewerking herhaald met de gecorrigeerde E-waarden, terwijl de ΔN lineair in de bewerking werd meegenomen.

Met betrekking tot bovenstaande is het wel gewenst in te gaan op het resultaat van de grafische bewerking welke voor de winterperiode werd uitgevoerd. Het blijkt namelijk in deze periode mogelijk te zijn - door middel van lineaire vereffening op de afvoer - de vochtverandering in het profiel (ΔV) uit het neerslagverschil (ΔN) te verklaren:

$$\Delta V = \alpha_1 \Delta N_1 + \alpha_2 \Delta N_2 + \alpha_3 \Delta N_3$$

Voor zandlysimeter 4 werden de volgende constanten gevonden:

$$\begin{aligned} \alpha_1 &= 0.619 \pm 0.063 \\ \alpha_2 &= 0.388 \pm 0.082 \\ \alpha_3 &= 0.022 \pm 0.191 \quad (r = 0.87) \end{aligned}$$

Daar de regressiefunctie als zodanig fysisch weinig betekenis heeft, zal dit onderzoek pas volledig zijn gerealiseerd indien de gevonden functie zowel statistisch verantwoord als fysisch aanvaardbaar is. Teneinde dit te verwezenlijken werd - gezien het flexibel karakter van dergelijke type functies - de bewerking eveneens uitgevoerd met de reciproke waarden.

De berekende resultaten worden onderstaand weergegeven:

	TABEL I	r^2
"lineair" polynomium	ongecorrigeerd	72.6
	gecorrigeerd + ΔN	92.3
"reciproke" polynomium	ongecorrigeerd zonder ΔN	63.0
	gecorrigeerd zonder ΔN	79.0

5. Selectie:

In 3 werd aangenomen dat de gevraagde functie benaderd kon worden door een tweede graads polynomium en wel volgens de regressievergelijking:

$$E_w = b_1 E_0 + b_2 N + b_3 \Delta N + b_4 G.L. + b_5 1-A + b_6 \Sigma \Delta V + b_7 E_0^2 + \dots + b_{11} \Sigma \Delta V^2 + b_{12} E_0 \cdot N + \dots + b_{21} 1-A \cdot \Sigma \Delta V + a(\text{constant})$$

en

$$E_w = b_1 \left(\frac{1}{E_0}\right) + b_2 \left(\frac{1}{N+1}\right) + b_3 \left(\frac{1}{G.L.}\right) + b_4 \left(\frac{1}{1-A}\right) + b_5 \left(\frac{1}{\Sigma \Delta V}\right) + b_6 \left(\frac{1}{E_0}\right)^2 + \dots + b_{10} \left(\frac{1}{\Sigma \Delta V}\right)^2 + b_{11} \left(\frac{1}{E_0} \cdot \frac{1}{N+1}\right) + \dots + b_{20} \left(\frac{1}{1-A} \cdot \frac{1}{\Sigma \Delta V}\right) + a$$

Men kan zich nu evenwel afvragen welke van in de bewerking opgenomen variabelen, tweede graads - en/of produkttermen overbodig zijn en is het wenselijk de overbodige veranderlijken uit te schakelen.

Welke van hen het meest voor verwijdering in aanmerking komt hangt af van hun partiële dispersie. Hoewel in de literatuur verschillende middelen worden aangegeven om de keuze van de uitgeschakelde veranderlijken zo doeltreffend mogelijk te maken werd voorlopig de volgende selectiemethode toegepast.

y_0 wordt achtereenvolgens op elk der vectoren x_1 tot en met x_{21} geprojecteerd. Is x_k de vector, die de hoogste projectie levert, dan wordt y_0 geprojecteerd op de deelruimten, welke door de paren vectoren x_k en x_i ($i = 1, 2, \dots, 21$) wordt opgespannen. Zo voortgaande volgt een opeenvolging van de hoogste projectie op de deelruimte R^m , opgespannen op een combinatie van m vectoren en de hoogste projectie op R^{m+1} , opgespannen op een combinatie van $(m+1)$ vectoren, dusdanig dat het kwadraat van de projectie op R^{m+1} minder verschilt van het kwadraat van de projectie op R^m dan de schatting van σ^2 .

Elke volgende verbetering ligt dan binnen de grootte van het toeval. Alle vectoren, die niet in de combinatie van m voorkomen zijn uitgeleerd als overbodig.

Selectie volgens deze methode met het beste resultaat leverde het volgende resultaat op.

TABEL II

variabele	D	S_b	t	gekozen door selectie
E_o	1.7594	0.376	4.68	*
N	-0.3393	0.088	3.87	
GL	0.0171	0.015	1.13	
I-A	-0.1613	0.141	1.14	
$\Sigma\Delta V$	-0.0444	0.261	0.17	*
$E_o.N$	0.0784	0.060	1.30	*
$E_o.GL$	0.0302	0.092	0.33	*
$E_o.I-A$	-0.0597	0.092	0.65	*
$E_o.\Sigma\Delta V$	0.0193	0.048	0.40	*
N.GL	-0.0466	0.024	1.98	*
N. $\Sigma\Delta V$	-0.0095	0.013	0.74	*
$r = 1,1, \dots, 1$	-2.2648			

N.b. De termen neerslag, graslengte en infiltratie min afvoer werden in de combinatie opgenomen omdat zij als interactie voorkomen.

De berekende correlatiecoëfficiënt bedraagt 0.95, terwijl de variantie van E_w zonder regressie, werd verlaagd tot 13.3%. Uit de uitkomsten van deze selectiemethode wordt wel heel sterk de indruk verkregen dat de open waterverdamping het grootste aandeel heeft in de beschrijving van de functie en de andere factoren als correctietermen kunnen worden opgevat. Een grafische voorstelling van de te bereiken aanpassing met deze variabelen (standaard afwijking 0.50) wordt gegeven in figuur 1, waar de berekende en gemeten verdamping tegen elkaar werden uitgezet.

Verder worden in bijgaande grafieken de betrekkingen tussen E_w en enkele variabelen voor verschillende niveau's van E_o , N, GL, I-A en $\Sigma\Delta V$ schematisch weergegeven. De gemiddelde waarden, waarop deze figuren betrekking hebben, gelden voor:

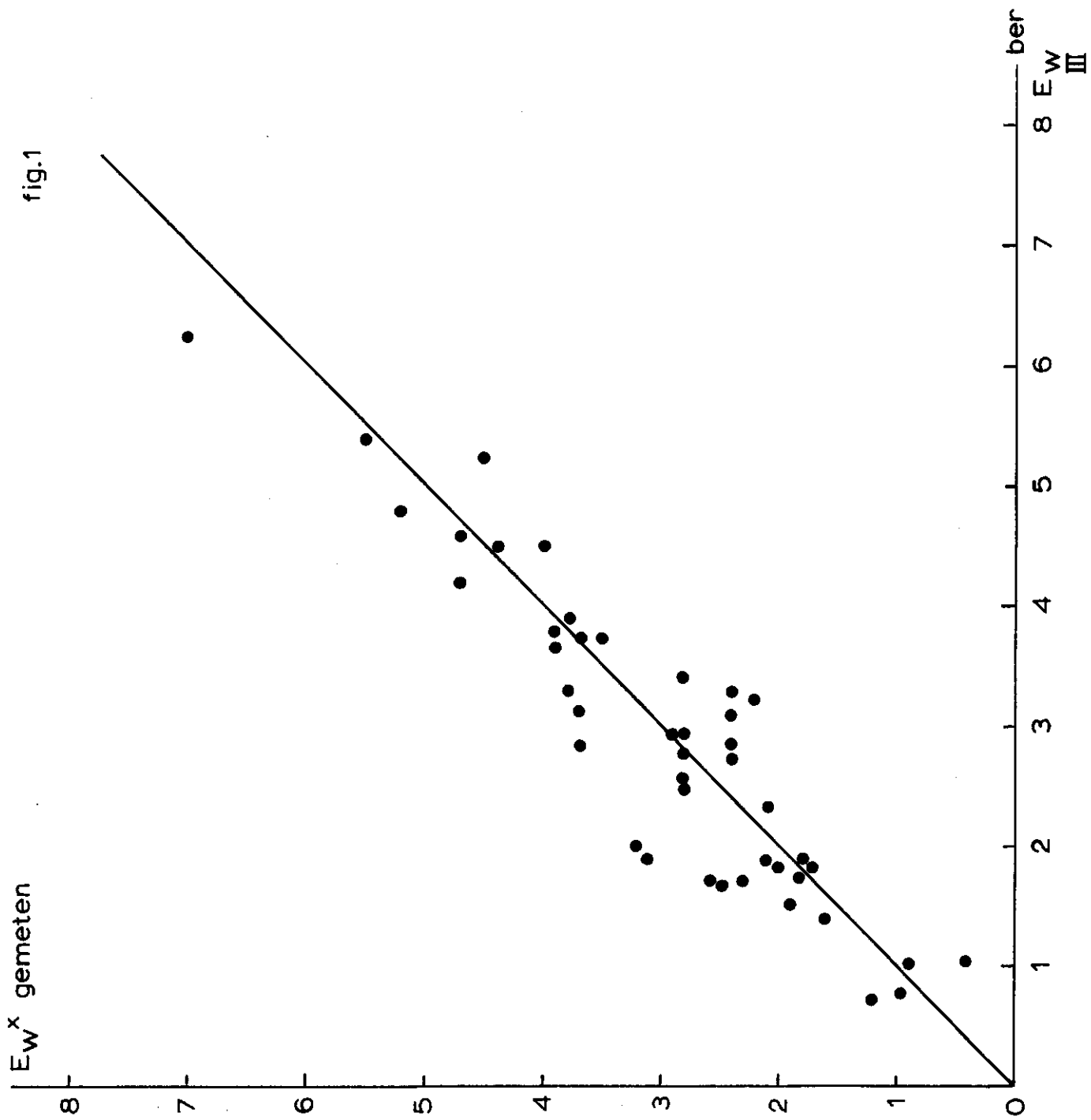
$$\begin{aligned} \overline{E_o} &= 3.7 \text{ mm} \\ \overline{N} &= 0.6 \text{ mm} \\ \overline{I-A} &= 1.8 \text{ mm} \\ \overline{\Sigma\Delta V} &= -48.8 \text{ mm} \\ \overline{GL} &= 6.18 \text{ cm} \end{aligned}$$

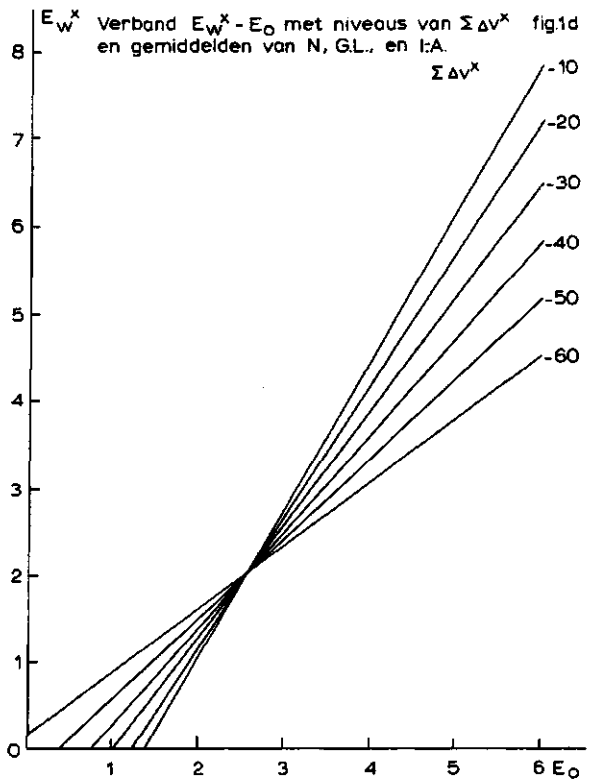
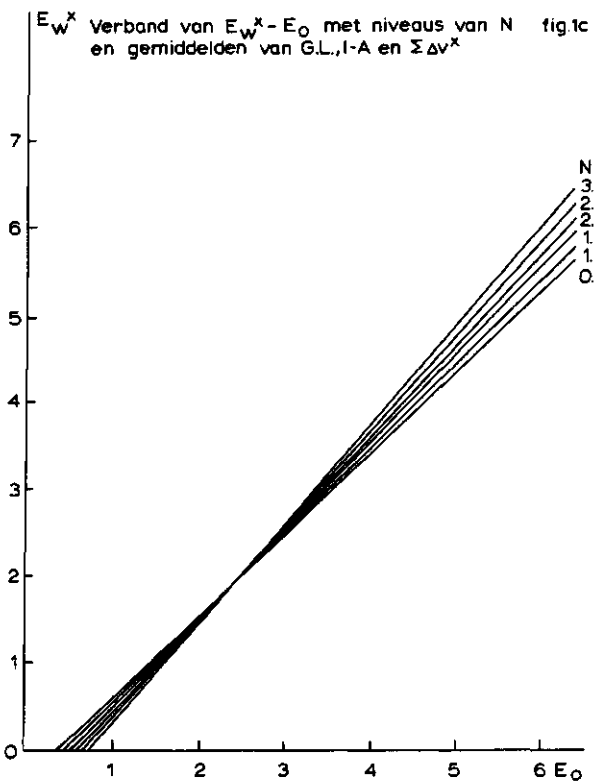
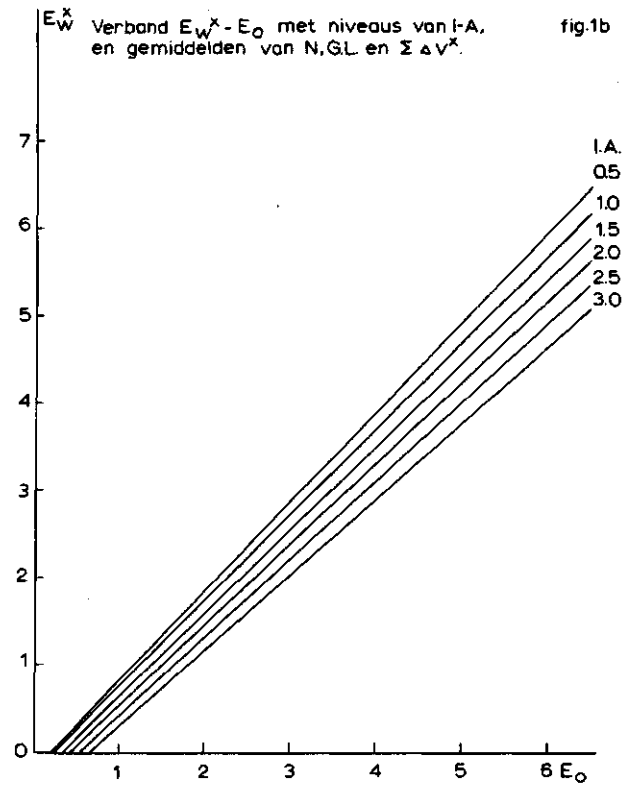
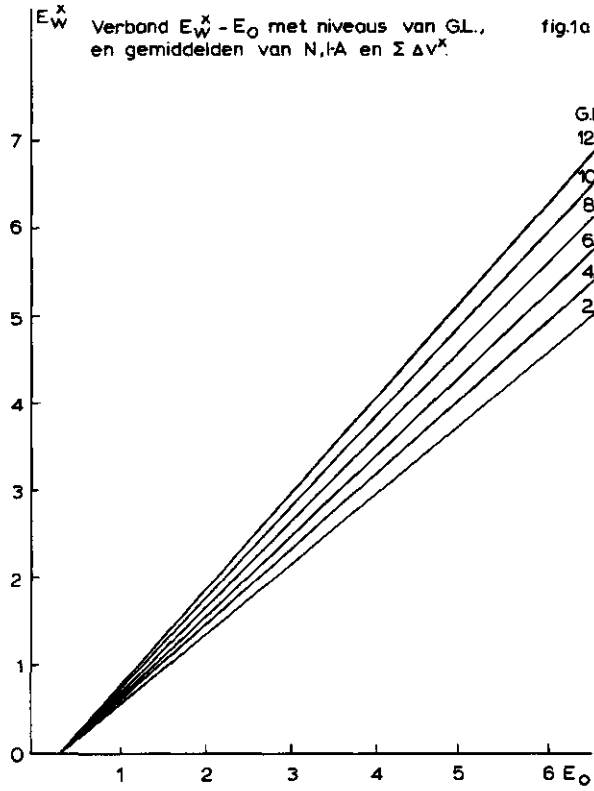
De gevonden verbanden zijn rechtlijnig en dit moet hoofdzakelijk worden toegeschreven aan het feit, dat het beschouwde interval te kort is, waardoor eventuele kromming niet tot uiting komt. Daar deze functies slechts gelden voor het gekozen interval en extrapolatie naar het "natte" traject uiteraard niet geoorloofd is, werd, gezien het resultaat, besloten dezelfde bewerking voor 2 opeenvolgende jaren uit te voeren. In een volgende nota zal hierop nader worden ingegaan, terwijl bovendien de fysische aspecten zullen worden belicht.

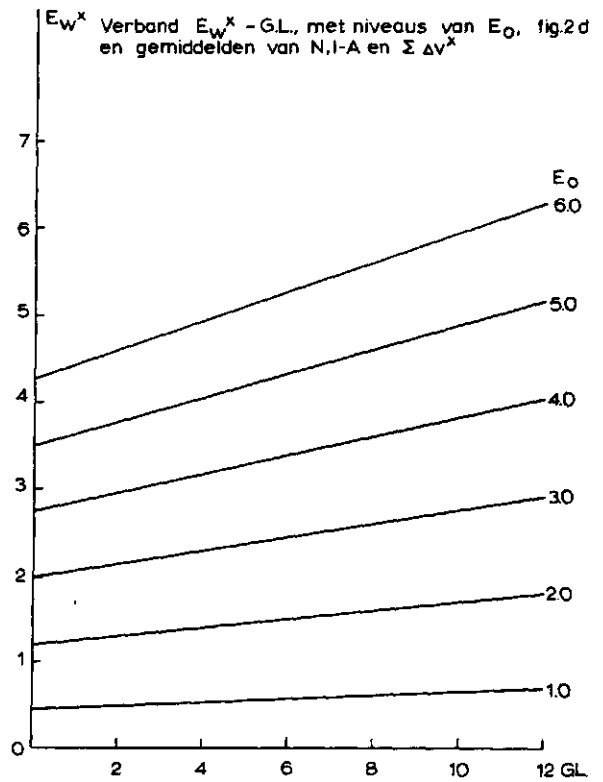
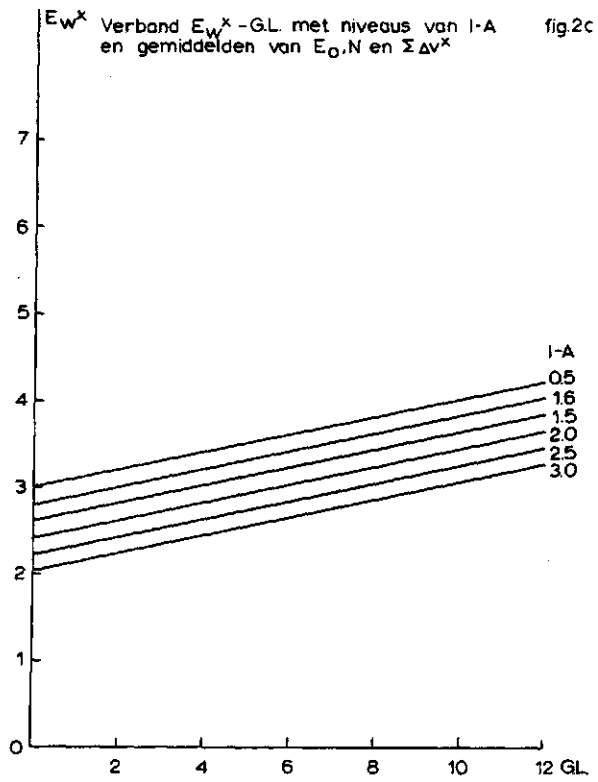
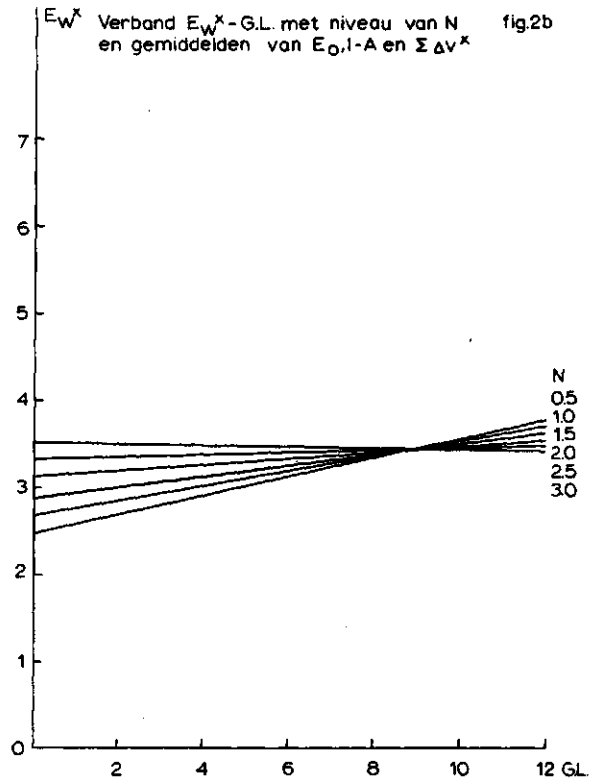
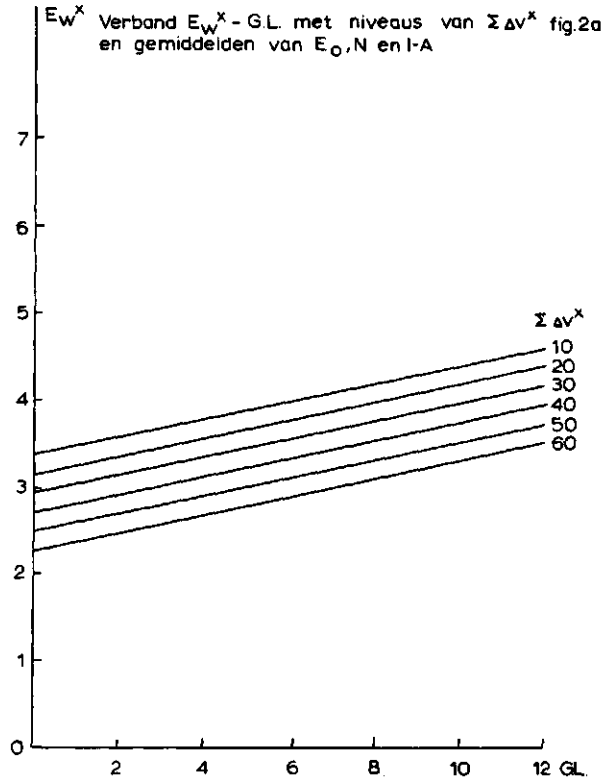
Samenvatting:

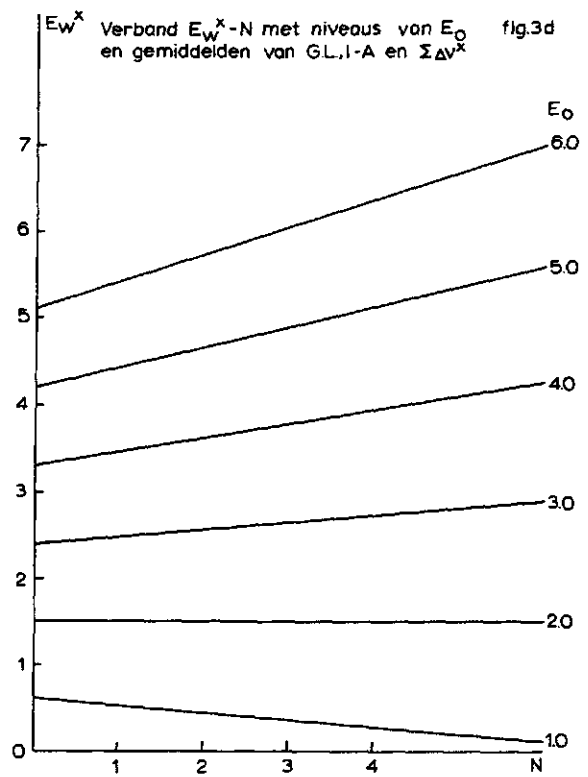
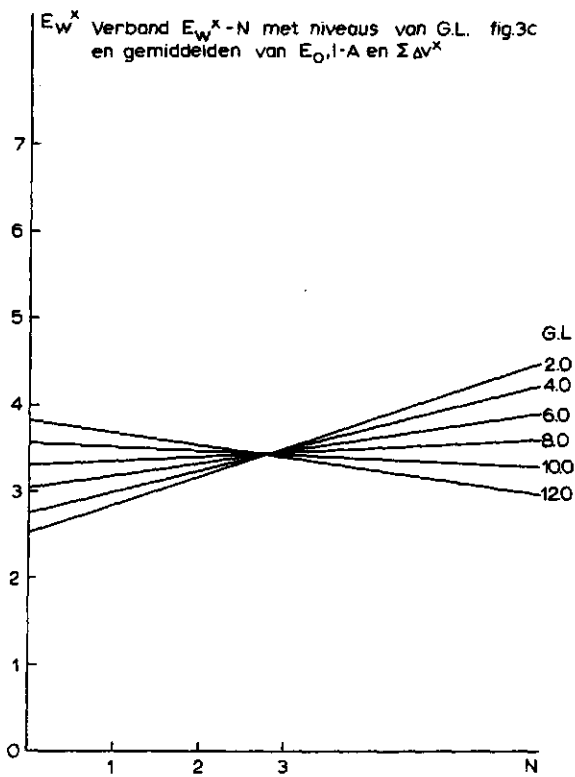
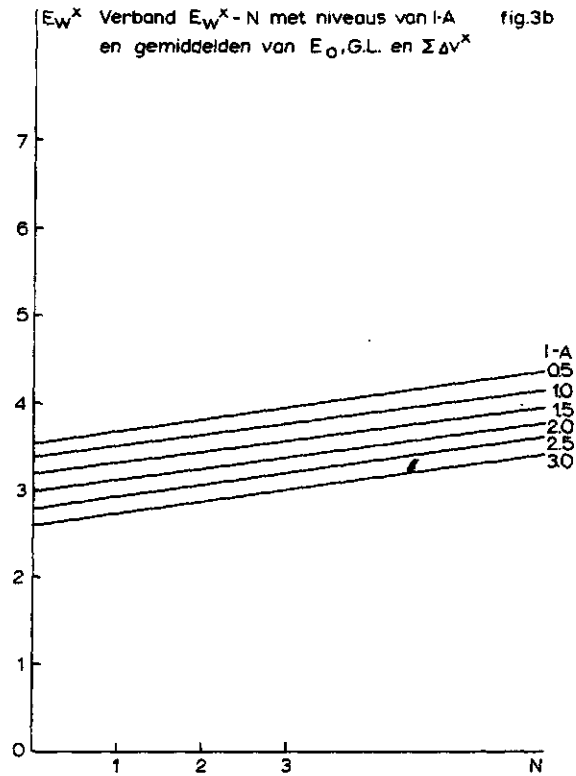
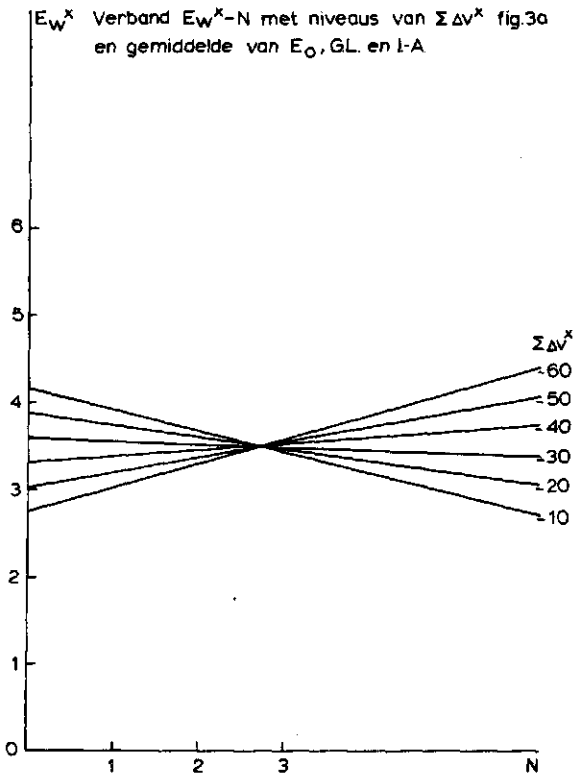
Met het oog op de overdraagbaarheid van weegbare lysimetergegevens naar de niet-weegbare lysimeter- en polderbalansen zal het gewenst zijn de werkelijke verdamping af te leiden uit de neerslagcijfers, afvoergegevens en open waterverdamping. Het vinden van een functie die het verband tussen de verdamping en het gewas, de bodemvochtigheid en het klimaat legt is een eerste vereiste. Voor dit doel werd, uitgaande van de waterbalansvergelijking $E_w = N + I - A \pm V$, met de meetbare grootheden uit deze vergelijking een statistische analyse uitgevoerd.

Bij dit onderzoek werd aangenomen dat de gevraagde functie, $E_w = \Phi$ (verklarende variabelen), benaderd kan worden door een tweede graads polynoom. De uit dit onderzoek verkregen resultaten waren aanleiding dezelfde bewerking, die in deze nota vrij uitvoerig wordt beschreven, eveneens toe te passen op twee opeenvolgende jaren, teneinde het type functie beter te kunnen benaderen.



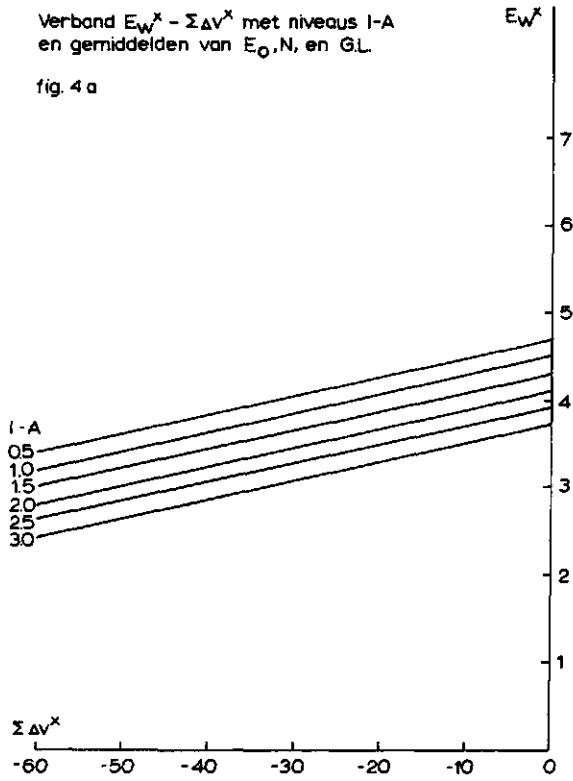






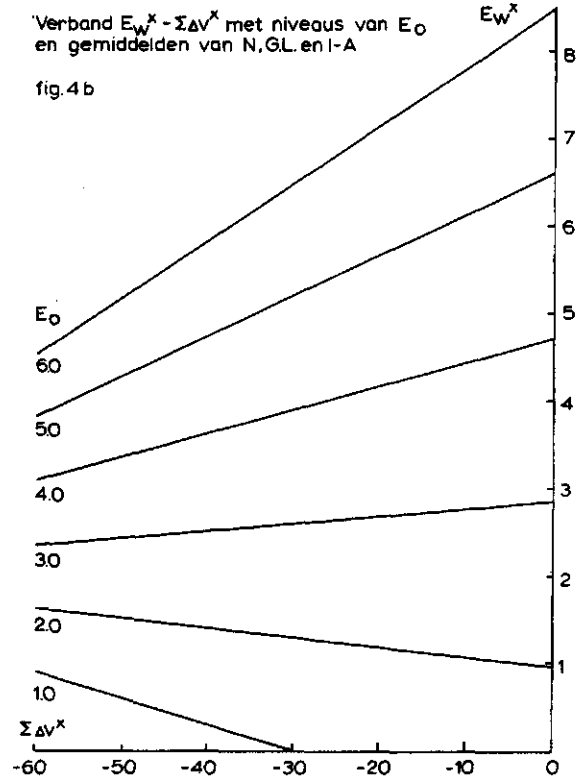
Verband $E_W^x - \Sigma \Delta v^x$ met niveaus I-A en gemiddelden van E_0 , N, en GL.

fig. 4 a



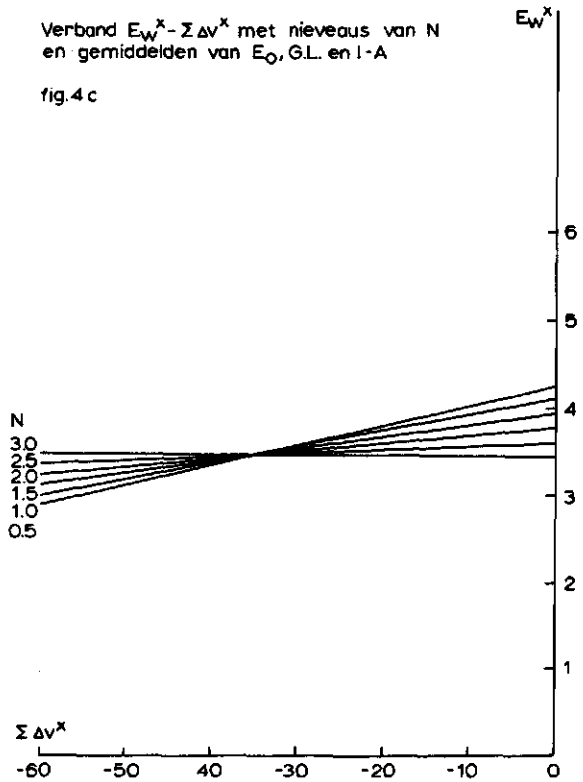
Verband $E_W^x - \Sigma \Delta v^x$ met niveaus van E_0 en gemiddelden van N, GL en I-A.

fig. 4 b



Verband $E_W^x - \Sigma \Delta v^x$ met niveaus van N en gemiddelden van E_0 , GL, en I-A.

fig. 4 c



Verband $E_W^x - \Sigma \Delta v^x$ met niveaus van GL en gemiddelden van E_0 , N en I-A.

fig. 4 d

