

K O N I N K L I J K N E D E R L A N D S
M E T E O R O L O G I S C H I N S T I T U U T

D e B i l t

WETENSCHAPPELIJK RAPPORT

W.R. 76-4

S. Kruizinga en G.J. Yperlaan

Ruimtelijke interpolatie van de dag-
totalen van de neerslag.
Een experimentele benadering

De Bilt, 1976

Publikationsnummer: K.N.M.I. W.R. 76-4 (Stat.Bur.)

U.D.C.: 551.577.21 :
551.501.777 :
519.2

Inhoud

0. Summary.
1. Inleiding.
2. De neerslagmetingen.
 - 2.1 Netwerk en gegevens.
 - 2.2 Ruimtelijke samenhang.
 - 2.3 Hoofdcomponentenanalyse.
 - 2.4 Conditionele kansen.
 - 2.5 Ruimtelijke variabiliteit.
3. Onderzoek interpolatiemethoden.
 - 3.1 Inleiding.
 - 3.2 Eénstationsschatter.
 - 3.3 Meerstationsschatters met gekozen gewichten.
 - 3.4 Meerstationsschatters met optimale gewichten.
 - 3.5 Schatting door aanpassing van kwadratische functies.
 - 3.6 Samenvatting resultaten.
4. Conclusies.

Samenvatting.

In dit verslag wordt een evaluatie gegeven van een aantal schattingstechnieken van de dagelijkse neerslaghoeveelheid. De onderzochte technieken zijn niet nieuw, meestal zelfs algemeen bekend. Met behulp van de metingen van omliggende stations is met een bepaalde techniek de neerslag van een refentiestation geschat. Door vergelijking van de geschatte en de gemeten neerslag is een indruk verkregen van de werking van de onderzochte technieken.

De eenvoudigste schatting - de waarde van het dichtstbijgelegen station - is duidelijk slechter dan de andere methoden. De kwaliteitsverschillen tussen de overige schattingsmethoden zijn vrij klein.

Uit het onderzoek zijn de volgende conclusies getrokken:

- a. De schattingsfout is afhankelijk van de gemiddelde neerslaghoeveelheid.
- b. De één-stationschatter is afhankelijk van de afstand, vooral in de zomer.
- c. Schattingen met gewogen gemiddelden zijn duidelijk beter dan schattingen met het dichtstbijgelegen station en zijn minder afhankelijk van de afstand.
- d. Bij gebruik van gewogen gemiddelden heeft de keuze van de gewichten weinig invloed op de schattingen; alleen bij sterke variaties in de onderlinge afstanden verdienen de afstandafhankelijke gewichten de voorkeur.
- e. Optimale schatters zijn niet essentieel beter dan gewogen gemiddelden.
- f. Ook aangepaste tweede-graads vlakken geven geen betere uitkomsten dan gewogen gemiddelden.

"Spatial interpolation of daily rainfall totals"

0. Summary.

In this report the evaluation of several interpolation techniques is presented. The element studied was the daily total amount of rainfall. The studied interpolation techniques were not new but mostly widely known. The evaluation was performed by using a given technique for the estimation of the rainfall at a site where the rainfall was also measured, without using this measurement. By comparing estimated and measured rainfall on many days an impression is obtained of the performance of a given technique. The studied techniques were all of the type "weighted average of surrounding stations", which type includes the fitted planes techniques, but not for instance spline-fitting.

Chapter 1 gives a general introduction on the subject of this report.

In chapter 2 some general rainfall statistics of the network used in this study are presented. The data set used consisted of twelve years of daily rainfall measurementd at thirty-five sites. The data set was split up into four seasonal data sets. The presented general statistics are:

- Ch. 2.1 Some geometric information of the network and the mean and root mean square of the daily rainfall of each station of the four seasons (table 2.1.1).
- Ch. 2.2 The coherency between a central station and the other stations, defined in formula 2.2.1, is presented in the figures 2.2.1 - 2.2.4.
- Ch. 2.3 The coherency-matrix has been decomposed in principal components. Some examples are shown in the figures 2.3.1 - 2.3.3.

- Ch. 2.4 The figures 2.4.1 - 2.4.4 give some examples of conditional probability patterns. The figures give the probability that the rainfall at a given station will exceed a certain level, provided that the rainfall at the central station exceeds that level. The given probabilities are much higher than the according climatological probabilities.
- Ch. 2.5 The spatial variability will influence the performance of the interpolation. This spatial variability varies from day to day. In the figures 2.5.1 - 2.5.4 an estimate of the spatial variability (formula 2.5.1) of each day is plotted against the mean rainfall of that day for each season. From these figures it is concluded that the spatial variability depends on season as well as on mean rainfall. Digits and letters in these figures indicate the number of times that the given combination occurred. (A indicated 10-times and so on).

In chapter 3 the results of the interpolation experiments are reported. As a measure for the interpolation error we have used the mean absolute error (expressed in millimeters); this measure will be less sensitive to outliers. In some cases the interpolation errors are grouped on the basis of the rainfall at the reference station because of the expected dependence of interpolation error on total rainfall. For each technique several tests were performed. The techniques tested are:

Ch. 3.1 Nearest neighbour estimation.

This is the simplest estimation technique. The performance was considerably lower than that of the following techniques. The results are shown in the tables 3.2.1 and 3.2.2 and in the figures 3.2.1 - 3.2.11. The figures 3.2.2 and 3.2.3 show that there is a fairly strong and simple relationship between the mean error within the groups and the mean overall error.

Ch. 3.2 Weighted mean of surrounding stations with chosen weights.

The interpolation was performed according to formula (3.1.1). The weights were chosen in several ways viz. [a] constant, [b] proportional to the inverse of the distance, [c] proportional to the inverse of the squared distance, [d, r₀] proportional to an exponential decay with parameter r₀. In table 3.1.1 the total mean absolute error is given for a number of experiments. In this table also results of techniques described in the following paragraphs are given. In the figures 3.3.1 - 3.3.4 some plots of "measured against estimated" are given,

Ch. 3.3 Optimal estimation with seven surrounding stations.

In this case the estimate of the rainfall at the reference station was written in the form of a multiple linear regression forced through zero. The coefficients were optimized on the basis of some minimal squared error. In the case indicated with opt. 1

$$D^2 = (h_{\text{estimated}} - h_{\text{measured}})^2 \text{ was}$$

minimized. In the case indicated with opt. 2 the minimized term was

$$D^2 = (h_{\text{estimated}} - h_{\text{measured}})^2 / h_{\text{mean}}$$

in order to take into account the earlier mentioned dependence of interpolation error on h_{mean}. In this case h_{mean} was defined as the mean rainfall at the seven stations used. The regression coefficients were derived from selected data. Only days with h_{mean} > 1 mm were taken into account. the regressions were evaluated on the data from which they were derived including the days that did not fulfil the condition h_{mean} > 1 mm. This means that the absolute errors found are bad indicators for the performance of the technique itself but good reference material for other techniques.

In the tables 3.4.2 en 3.4.3 some evaluation results are given. In the figures 3.4.1 and 3.4.2 the regression coefficients of one of the tested cases are plotted versus minus the inverse distance. The figures seem to favour the choice of weights proportional to the inverse distance (see previous paragraph). In figure 3.4.3 the relation between overall error and weighted mean distance is depicted.

Ch. 3.5 Interpolation by fitting functions.

In this paragraph interpolation by fitting quadratic surfaces is tested. First a proof is given that this method can also be described as a weighted mean method (see also De Bruin and Edwards). The results of the interpolation evaluation are given in table 3.5.1. (This interpolation technique can lead to negative estimated rainfall; in that case these estimates are taken equal to zero). Overall results are also given in table 3.3.1.

Ch. 4. Conclusions.

The experiments described above lead to the following conclusions:

- a. The interpolation error depends on areal mean rainfall.
- b. The one station estimator shows a distance-dependence especially in summer.
- c. Weighted mean estimators are markedly better than nearest neighbour estimators and show less distance-dependence.
- d. The choice of the weights for the weighted mean estimators is not critical; only when the distanced are strongly varying distance-dependence of weights is needed.
- e. Optimal estimators are not essentially better than weighted means.
- f. Fitting quadratic surfaces are not better than weighted means.

1. Inleiding.

Bij de inrichting van een meteorologisch netwerk zal de gewenste stationsdichtheid één van de kernproblemen zijn. De vraag die men zich daarbij stelt is "Hoe dicht dient het netwerk te zijn opdat voldoende nauwkeurig de meteorologische parameter als functie van de plaats vastgelegd kan worden". Wat hierbij als voldoende nauwkeurig moet gelden wordt voor het grootste deel door het probleem van de gebruiker bepaald. Hieruit volgt dat het niet mogelijk is om à priori de werkdichtheid vast te leggen. Wel is het mogelijk om een relatie tussen de nauwkeurigheid en de dichtheid op te stellen, op basis waarvan dan bij een gegeven probleem de gewenste netwerkdichtheid vastgesteld kan worden. Om het begrip nauwkeurigheid te quantificeren wordt algemeen de interpolatie-fout gehanteerd. Hiermee wordt bedoeld de fout waarmee een grootte op een gegeven plaats geschat kan worden, op basis van de metingen in het netwerk. Indien deze fout overal klein genoeg is, stelt men dat de netwerkdichtheid voldoende is. Naast het interpolatie criterium hanteert men bij neerslag ook wel de nauwkeurigheid van het gebiedsgemiddelde als criterium. Dit laatste criterium zal in dit rapport niet worden behandeld (zie ook appendix B).

Over het probleem hoe moet worden geïnterpoleerd en welke fout daarbij wordt gemaakt, bestaat zeer veel literatuur, in de meteorologische literatuur vooral met betrekking tot de neerslag. De reden voor deze extra aandacht is de enorme grilligheid van het verschijnsel neerslag. Men kan ten aanzien van het interpoleren ruwweg twee hoofdrichtingen aangeven.

I. Interpoleren met behulp van **gewogen gemiddelden**.

II. Interpoleren met behulp van **aangepaste functies**.

Onder de methoden in groep I worden verstaan, die methoden waarbij gesteld wordt

$$h(\text{geschat}) = \sum_{h=1}^K g_k h_k$$

waarin h_k de neerslagsommen op de stations en g_k de gewichtsfactoren. De gewichten zijn in het algemeen afhankelijk van de afstand tussen het interpolatiepunt en het neerslagstation.

In sommige gevallen hebben de gewichten een gekozen afhankelijkheid van de afstand [Salter (1)] en in andere gevallen worden ze geoptimaliseerd met behulp van de correlatiefunctie of de structuurfunctie [Gandin (2)]. In het laatste geval wordt de interpolatiefout afgeleid uit deze zelfde functies, in het eerste geval is daarover weinig bekend.

Bij de methoden uit groep II kan men nog meerdere mogelijkheden onderscheiden, bijvoorbeeld:

Men stelt dat het neerslagpatroon goed wordt weergegeven door de functie $h(x, y) = f(x, y, p, \dots, p_k)$, waarin $h(x, y)$ een schatting van de neerslag en f een analytische functie met parameters p_1 t/m p_k . Deze parameters worden voor elk neerslagpatroon opnieuw afgeleid uit de gemeten neerslagwaarden op basis van minimum χ^2 - methoden [Edwards (3)]. Door De Bruin (5) is bewezen dat deze methode ook in een vorm als onder I gebracht kan worden.

Vele andere methoden zijn nog in gebruik, evenals mengvormen van I en II [Shaw (4)]. Niet alle methoden zijn zonder meer in een vorm als onder I te brengen. Over de interpolatiefout is bij deze methoden weinig bekend.

In dit rapport zal uitsluitend aandacht worden besteed aan de interpolatie via gewogen gemiddelden. Deze technieken kunnen gemakkelijk op routinebasis worden toegepast. Andere methoden zijn daarvoor vaak te rekenintensief. Er zal een aantal mogelijkheden langs experimentele weg worden onderzocht. Dat wil zeggen dat in een groot aantal gevallen, de neerslag ter plaatse van een station zal worden geschat op basis van een gegeven methode, met behulp van de neerslag op andere stations. Door vergelijking van geschatte neerslag zal een indruk kunnen worden verkregen van de prestaties van de gegeven methode.

In dit rapport zal alleen met dagsommen worden gewerkt en niet met meerdaagse sommen (2 tot 10).

Ten eerste zal namelijk een netwerk dat geschikt is voor ééndaagse sommen, in de meeste gevallen ook voor meerdaagse sommen geschikt zijn. Ten tweede mag men niet verwachten, dat de benodigde netwerkdichtheid erg sterk zal dalen voor de andere situaties.

Het is dan handiger om eerst een netwerk-keus te maken en dan gericht uit te zoeken wat dit netwerk presteert voor de meerdaagse sommen.

Gaat men over naar sommen over 30 dagen of langer, dan kan men de experimentele benadering beter verlaten en een meer theoretische statistische methodiek gebruiken.

2. De neerslagmetingen,

2.1 Netwerk en gegevens.

Voor het onderzoek van de diverse schattingsmethoden is gebruik gemaakt van de neerslaggegevens (dagsommen) die verkregen zijn in het Nederlandse klimatologische netwerk. Deze gegevens zijn door de Klimatologische Dienst van het K.N.M.I. verzameld en gecontroleerd en onder toezicht van deze dienst op magneetband gezet.

Uit de meetposten, gelegen in Vak VII (Zuid-Holland) werd een aantal stations (35) geselecteerd die gedurende een gemeenschappelijk tijdvak (1958 t/m 1969), een ononderbroken reeks leverden. Dit bestand werd verder niet gecontroleerd, alleen de neerslagsom op station Lynden van 31 juli 1968 (777,7 mm) werd gecorrigeerd tot 0 mm. In figuur 2.1.1 is de ligging van het netwerk (in Nederland) aangegeven, terwijl figuur 2.1.2 een nauwkeurig beeld geeft van de onderlinge ligging der stations. De overige figuren in dit verslag, met betrekking tot het netwerk, zijn op dezelfde schaal als figuur 2.1.2. In tabel 2.1.1 zijn de namen van de gebruikte stations gegeven samen met het nummer in het Nederlandse net en het in het onderhavige verslag toegewezen nummer.

Op basis van het twaalfjarige tijdvak werd ook een aantal eenvoudige statistische gegevens bepaald. Daar de eigenschappen van de neerslag sterk afhankelijk zijn van het seizoen, werd het materiaal gesplitst naar seizoenen. De seizoenen werden als volgt gedefinieerd: de lente bevat de maanden maart, april en mei; de zomer de maanden juni, juli en augustus; de herfst bevat de maanden september, oktober en november; de winter bevat de resterende maanden. In tabel 2.1.1 is ook de gemiddelde dagsom en de standaardafwijking van de dagsommen per seizoen gegeven.

N.B. Alle resultaten en studies in dit rapport hebben betrekking op dagsommen. De neerslagsommen worden genoteerd als millimeter-neerslaghoogte.

Uit de tabel blijkt dat de gemiddelde dagsommen enigszins variëren, het was echter moeilijk hierin een duidelijk patroon te herkennen. Daar verder ook niet bekend is of de gegeven verschillen verband houden met reële klimatologische verschillen of met eventuele verschillen in de opstellingen, werd besloten om in dit rapport (uitgezonderd par. 2.4), een gestandaardiseerde neerslag in te voeren. De gestandaardiseerde neerslag $h_{n,k}$ van dag n op station k werd als volgt gedefinieerd,

$$h_{n,k} = h'_{n,k} \frac{\overline{h'_{n,k}}^{N,K}}{\overline{h'_{n,k}}^N}$$

waarin:

$h'_{n,k}$: de gemeten neerslag op dag n op station k.

$\overline{h'_{n,k}}^N$: gemiddelde dagsom op station k, gemiddeld over alle dagen van dat seizoen in de gegeven 12 jaren.

$\overline{h'_{n,k}}^{N,K}$: de gemiddelde dagsom gemiddeld over alle stations.

Tabel 2.1.1 : Lijst van gebruikte stations met coördinaten en enkele statistische gegevens.

nr.	KNMI nr.	Naam	Coördinaten		Gemiddelde				Standaardafwijking			
			N	O	Lente	Zomer	Herfst	Winter	Lente	Zomer	Herfst	Winter
1	225	Overveen	79,5	36,9	1,7	2,4	2,9	2,4	3,4	4,8	5,5	4,1
2	229	Zandvoort	74,0	30,0	1,6	2,4	2,8	2,3	3,2	4,7	5,4	4,1
3	435	Heemstede	72,0	39,2	1,7	2,5	2,9	2,3	3,4	5,0	5,7	4,2
4	436	Sassenheim	59,0	33,4	1,6	2,5	2,7	2,4	3,4	4,9	5,1	4,1
5	437	Lijnden	74,0	48,4	1,7	2,5	2,8	2,4	3,5	5,1	5,5	4,3
6	438	Hoofddorp	68,4	42,6	1,6	2,5	2,8	2,3	3,4	5,0	5,5	4,1
7	439	Oude - Wetering	59,0	39,2	1,7	2,6	2,8	2,5	3,4	5,2	5,3	4,2
8	440	Scheveningen	48,0	16,2	1,7	2,5	2,9	2,5	3,4	5,3	5,4	4,4
9	441	Amsterdam	76,0	57,5	1,8	2,9	2,9	2,6	3,5	5,7	5,8	4,7
10	442	Boskoop	44,3	41,5	1,7	2,7	2,7	2,4	3,3	5,4	5,2	4,3
11	443	Gouda	35,0	44,9	1,7	2,9	2,6	2,5	3,2	5,8	5,1	4,5
12	444	Katwijk a/d Rijn	55,4	24,3	1,6	2,3	2,7	2,2	3,1	4,7	5,3	4,0
13	445	Rotterdam Westerkade	23,9	28,8	1,6	2,8	2,5	2,3	3,3	6,1	5,2	4,3
14	448	Naaldwijk	33,2	9,4	1,5	2,5	2,6	2,3	3,1	5,2	4,9	4,0
15	449	Delft	35,0	20,8	1,6	2,7	2,7	2,4	3,3	5,3	5,1	4,2
16	450	Numansdorp	3,5	24,3	1,5	2,6	2,4	2,1	3,0	5,7	4,7	3,7
17	451	IJsselmonde	22,0	33,4	1,7	2,7	2,6	2,4	3,3	5,7	4,9	4,3
18	453	Bergschen- hoek	33,2	30,0	1,7	2,9	2,7	2,4	3,5	6,0	5,3	4,3
19	454	Lisse	62,8	33,4	1,6	2,6	2,7	2,4	3,3	5,3	5,5	4,2
20	455	Mookhoek	7,2	36,9	1,7	2,7	2,4	2,3	3,3	5,5	4,8	4,3

Vervolg en verklaring zie volgende pagina.

nr.	KNMI nr.	Naam	Coördinaten		Gemiddelde				Standaardafwijking			
			N	O	Lente	Zomer	Herfst	Winter	Lente	Zomer	Herfst	Winter
21	456	Oostvoorne	25,7	2,5	1,6	2,5	2,7	2,4	3,1	5,4	4,9	4,2
22	457	Rotterdam Keilehaven	23,9	25,4	1,7	2,9	2,6	2,5	3,4	6,4	5,3	4,5
23	458	Aalsmeer	64,7	48,3	1,7	2,6	2,8	2,3	3,4	5,2	5,2	4,2
24	461	Barendrecht	16,5	33,4	1,7	2,8	2,5	2,3	3,5	5,9	5,1	4,2
25	463	Nieuw- Helvoet	16,5	3,6	1,5	2,5	2,5	2,2	2,9	5,5	4,9	3,8
26	464	Brielle	23,9	7,1	1,5	2,6	2,7	2,4	3,0	5,3	5,2	4,2
27	466	Wassenaar	49,8	23,1	1,7	2,7	2,8	2,5	3,4	5,5	5,4	4,4
28	467	Poortugaal	18,3	23,1	1,6	2,9	2,5	2,3	3,2	6,3	5,0	4,1
29	469	Leiden	55,4	30,0	1,6	2,6	2,8	2,3	3,3	5,1	5,4	4,2
30	470	Zegveld	49,8	55,2	1,8	2,6	2,7	2,4	3,5	5,4	5,2	4,4
31	548	Loenen a/d Vecht	59,1	66,6	1,9	2,9	2,6	2,4	3,6	6,2	5,0	4,3
32	559	Vleuten	44,3	64,4	1,8	2,8	2,5	2,4	3,6	6,0	5,0	4,4
33	561	Benschop	36,8	63,2	1,8	2,8	2,4	2,3	3,3	5,6	4,9	4,2
34	563	Weesp	68,4	66,6	1,8	2,8	2,5	2,4	3,4	5,6	5,1	4,6
35	572	Abcoude	64,7	63,2	1,7	2,7	2,5	2,4	3,3	5,3	5,1	4,5

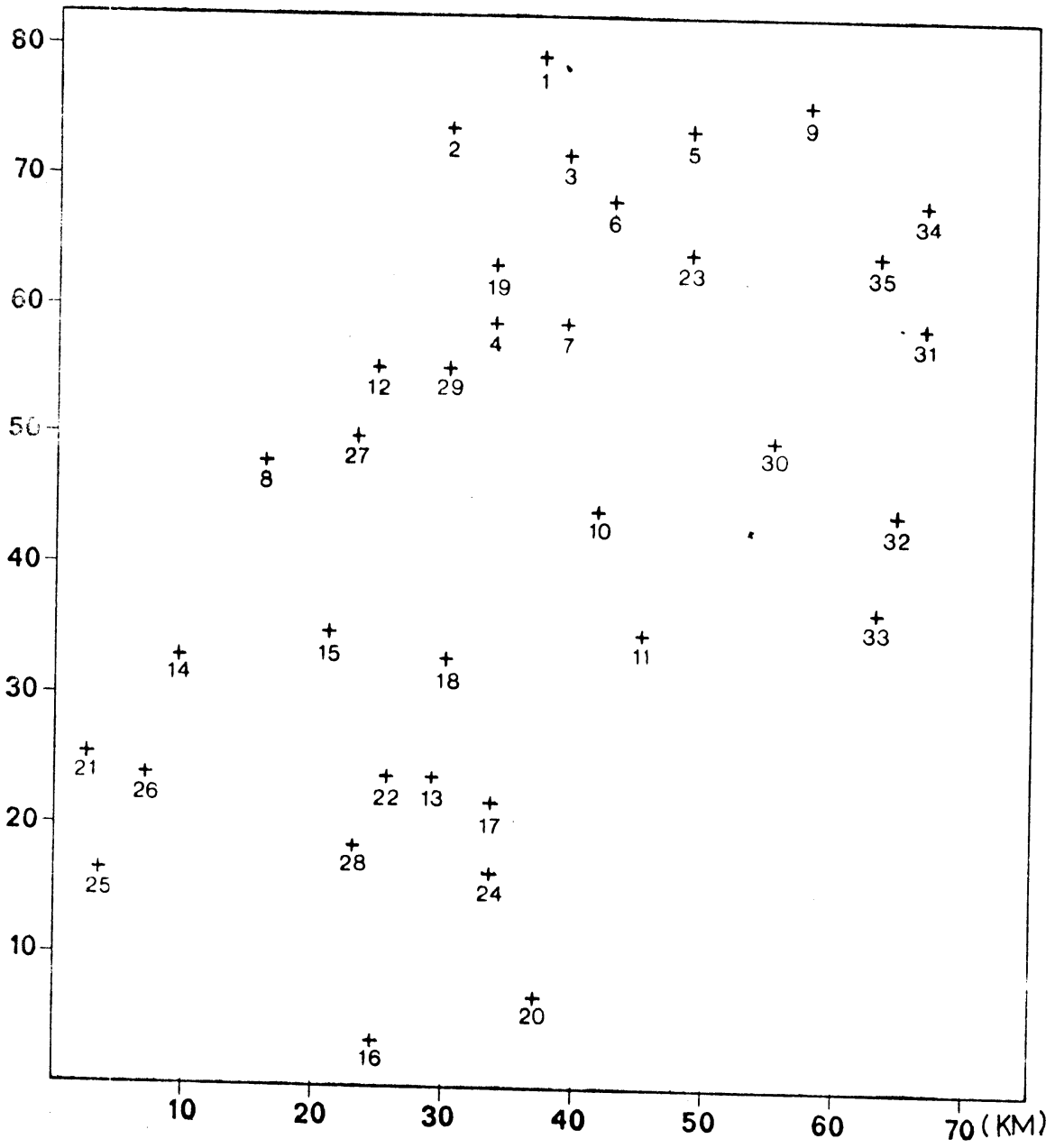
Verklaring:

Coördinaten: Afstand in kilometers respectievelijk in Noordelijke
en Oostelijke richting tot referentiepunt (51°41'N, 4°04'O).

Gemiddelde en standaard afwijking van de dagsommen zijn berekend over alle
dagen en gegeven in millimeter neerslaghoogte.



Figuur 2.1.1. Ligging van het netwerk in Nederland.



Figuur 2.1.2 Het proefnetwerk.

2.2. Ruimtelijke samenhang van de neerslag.

Naast het gemiddelde en de standaardafwijking is ook de parameter, die de samenhang tussen de neerslag gemeten op de verschillende stations vastlegt, van belang. In het geval dat met normaal verdeelde grootheden wordt gewerkt is de correlatie-coëfficiënt hiervoor zeer geschikt. Bij een variabele die zo scheef verdeeld is als de neerslag, en bovendien eenzijdig begrensd, wordt de waarde van de correlatie-coëfficiënt discutabel. Vooral daar veelal niet duidelijk is wat men met het grote aantal droge dagen moet doen. Door sommige auteurs wordt dan ook voorgesteld om de correlatie-coëfficiënt alleen te berekenen over die dagen waarvoor geldt dat op tenminste één station de neerslag een gegeven grens overschrijdt. Hiermee wordt uiteraard een zekere willekeur geïntroduceerd. Door enkele auteurs [Zawadzki (6)] wordt ook voorgesteld de samenhang (die we zullen noteren als S) als volgt vast te leggen,

$$S = \frac{\overline{h_{n,k} \cdot h_{n,k}}^N}{\sqrt{\overline{h_{n,k}^2} \cdot \overline{h_{n,k}^2}}} \quad (2.2.1)$$

waarin $h_{n,k}$: neerslag op station k op dag n

en $h_{n,k'}$: " " " k' " .

De parameter S wordt, evenals de correlatie-coëfficiënt, gelijk aan 1 in het geval van perfecte samenhang. Een nadeel is echter dat S niet gelijk nul wordt bij volstrekte on-afhankelijkheid.

(11.1) De ondergrens van S hangt af van de verdeling van de neerslagsommen en is moeilijk te schatten.

Deze ondergrens is experimenteel geschat door een reeks van dagsommen van eenzelfde seizoen, over 12 jaar achter elkaar gezet, de samenhang te berekenen met de reeks van over 100 dagen verschoven dagsommen. Op deze manier werd voor de lente $S = .20$ gevonden, voor de zomer $S = .22$, voor de herfst $S = .19$ en voor de winter $S = .28$. Deze resultaten zijn belangrijk lager dan de samenhang die wordt gevonden tussen de stations.

Deze parameter vertoont enige overeenkomst met de beste schatter voor a in het lineaire regressiemodel dat door de oorsprong geforceerd wordt namelijk

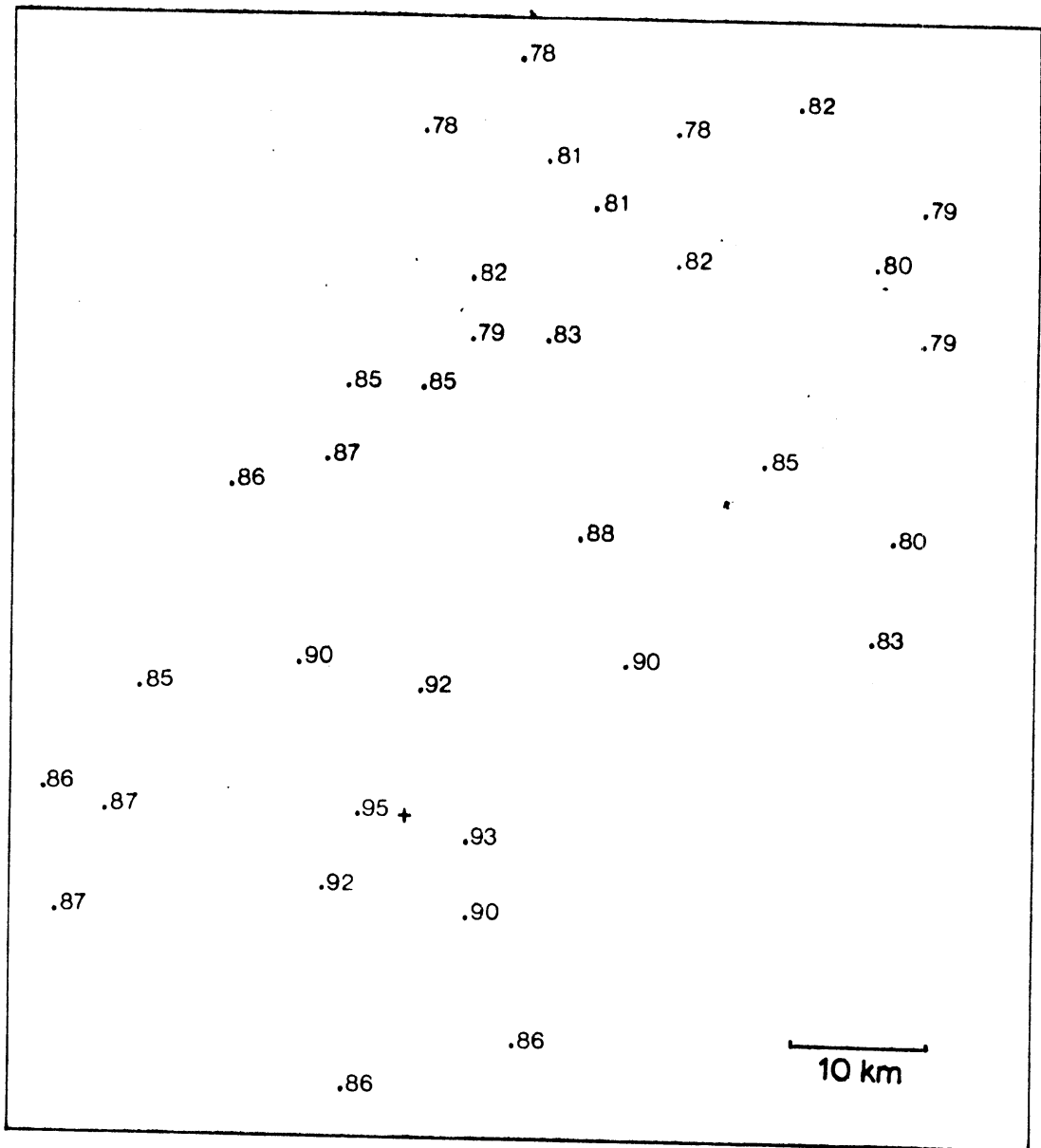
$$y = a \cdot x \quad (2.2.2)$$

met als oplossing

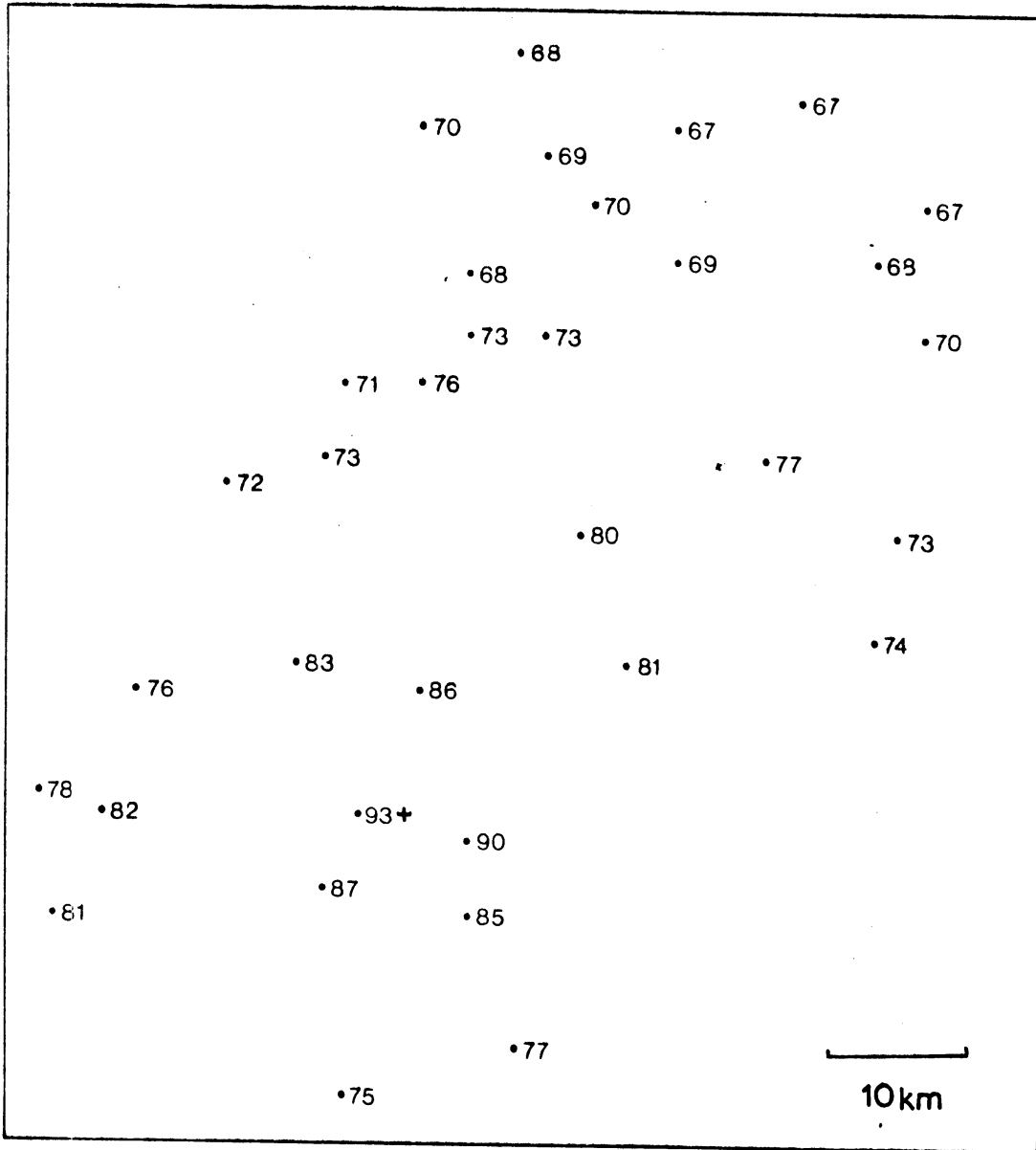
$$a = \frac{\overline{x \cdot y}}{\overline{x^2}} \quad (2.2.3)$$

(zie ook appendix A)

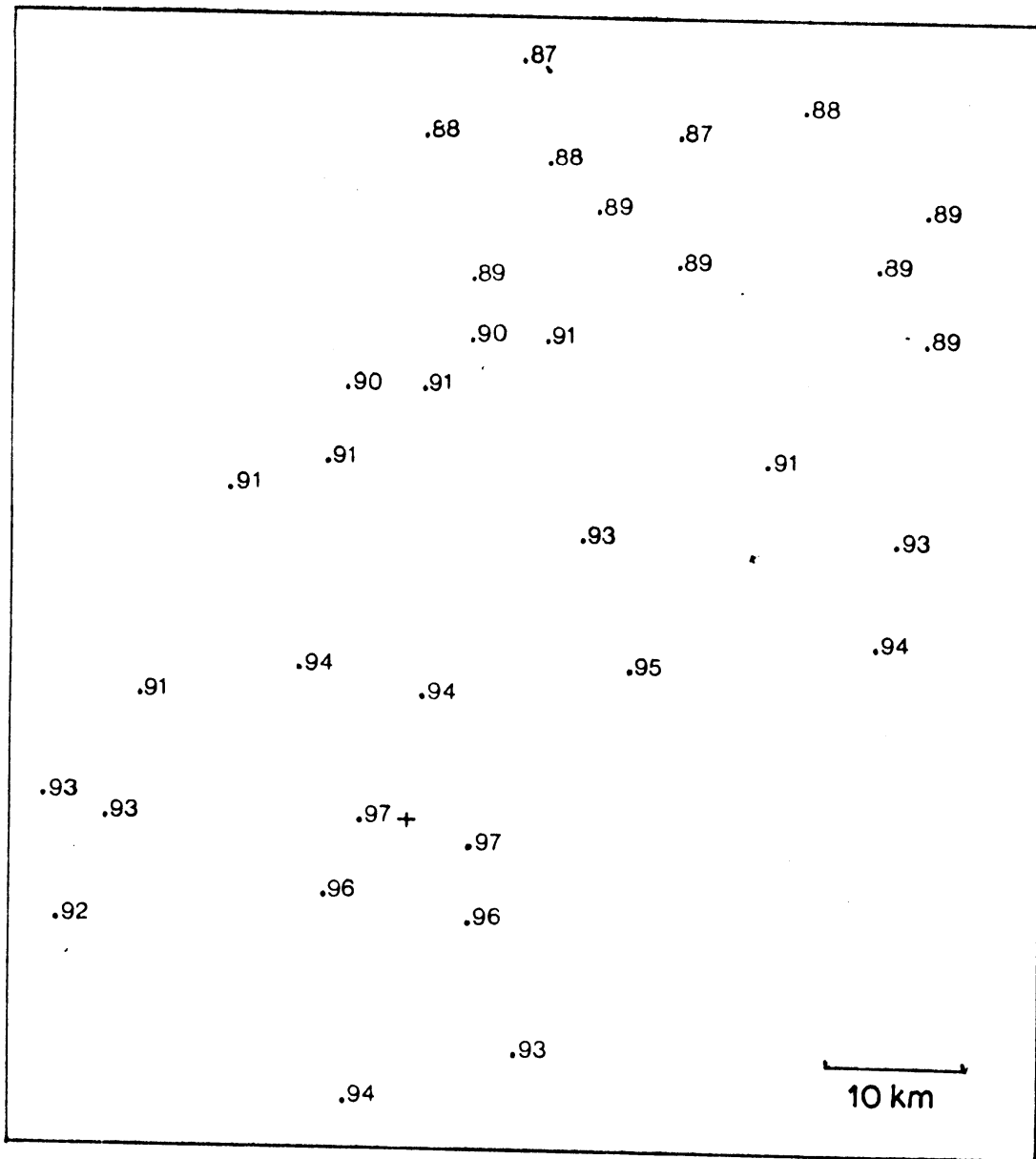
In de figuren 2.2.1 t/m 2.2.4 is voor de verschillende seizoenen de samenhang van de andere stations met station 13 gegeven. In de figuur 2.2.5 zijn deze zelfde S -waarden uitgezet tegen de afstand. De parameter S zou kunnen dienen om andere gebieden of daarbuiten te vergelijken met het proefgebied.



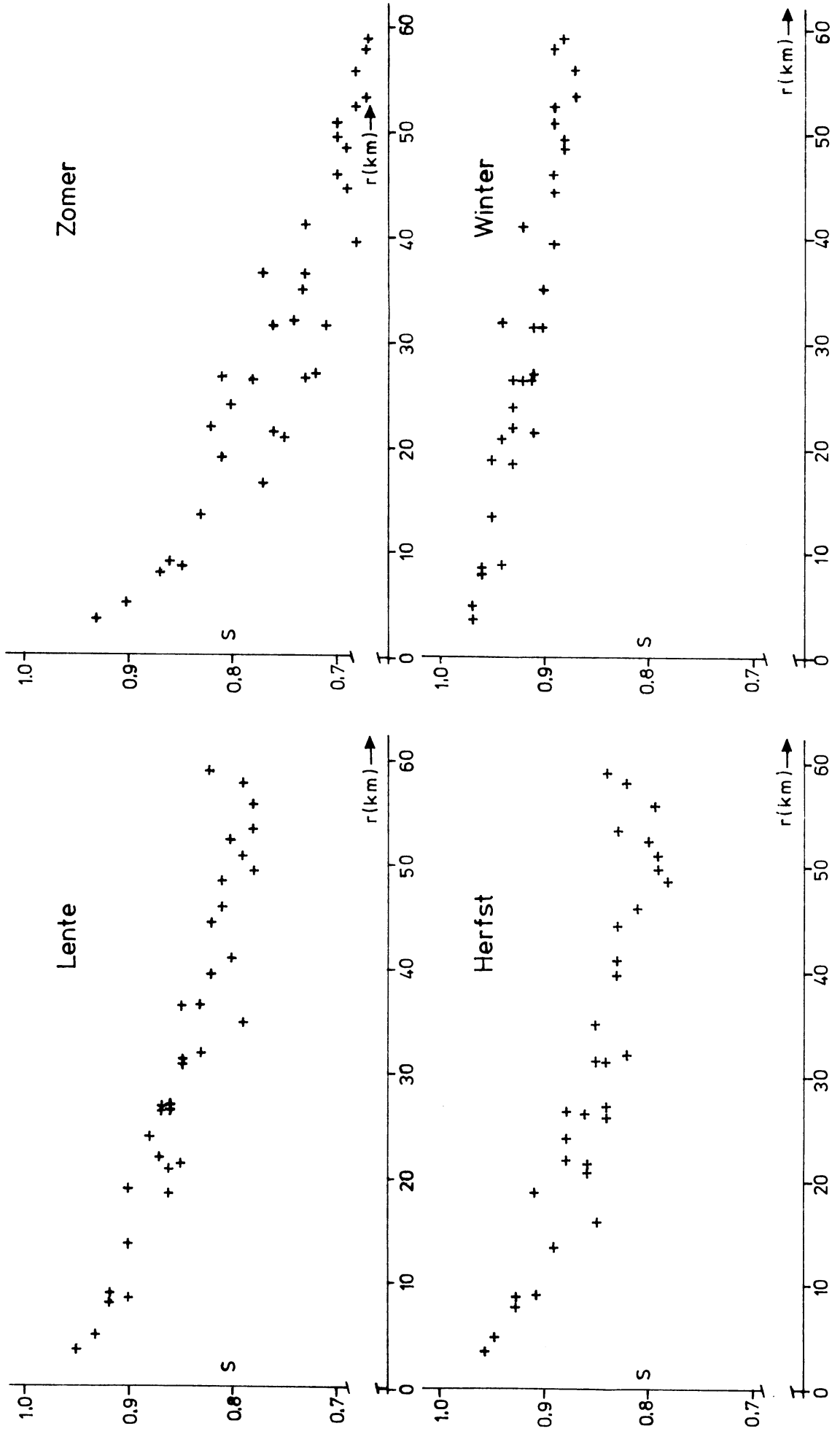
Figuur 2.2.1 De samenhang van de neerslag op de verschillende stations met de neerslag op station 13 (+) in de lente.



Figuur 2.2.2 Samenhang met station 13 (+) in de zomer.



Figuur 2.2.4 Samenhang S met station 13 (+) in de winter.



Figuur 2.2.5 Samenhang S als functie van de afstand per seizoen

2.3. Hoofdcomponentenanalyse.

Hoofdcomponentenanalyse is een statistische techniek uit de multi-variabelen analyse die de mogelijkheid geeft om ingewikkelde correlatiestructuren uiteen te rafelen. bij deze techniek wordt de covariantie-matrix ontbonden in "hoofdcomponenten" en "hoofdvariabelen" respectievelijk de eigenwaarden en eigenvectoren van de covariantie-matrix. De interpretatie van de resultaten van deze techniek is niet eenvoudig, maar in het algemeen kan worden gesteld dat de eigenvectoren behorend bij de grootste eigenwaarden correlatiestructuren in het materiaal weergeven, terwijl de overige eigenvectoren steeds meer de steekproef-effecten weergeven. Waar de scheiding zal liggen is niet van te voren aan te geven.

Normaal wordt de covariantie-matrix berekend ten opzichte van het gemiddelde van ieder der variabelen. Om dezelfde redenen als in voorgaande paragraaf is dit voor de neerslag niet zo'n gelukkige keuze. Daarom is ook hier ten opzichte van het nulpunt gerekend. De matrix \vec{H} , welke met de hoofdcomponentenanalyse is bewerkt, is als volgt gedefinieerd:

$$H = \begin{pmatrix} \overline{h_1 \cdot h_1}^N, & \overline{h_1 \cdot h_2}^N, & & \overline{h_1 \cdot h_K}^N \\ \overline{h_2 \cdot h_1}^N & & & \\ & & & \\ & & & \\ \overline{h_K \cdot h_1}^N, & & & \overline{h_K \cdot h_K}^N \end{pmatrix}$$

waarin h_k de neerslag op station k is, terwijl het symbool $\overline{\quad}^N$ aangeeft dat gemiddeld is over N dagen.

In tabel 2.3.1 zijn per seizoen de eerste zeven eigenwaarden λ_i gegeven. Wel blijkt, dat de eerste eigenwaarde in alle seizoenen de volgende eigenwaarden verre overtreft.

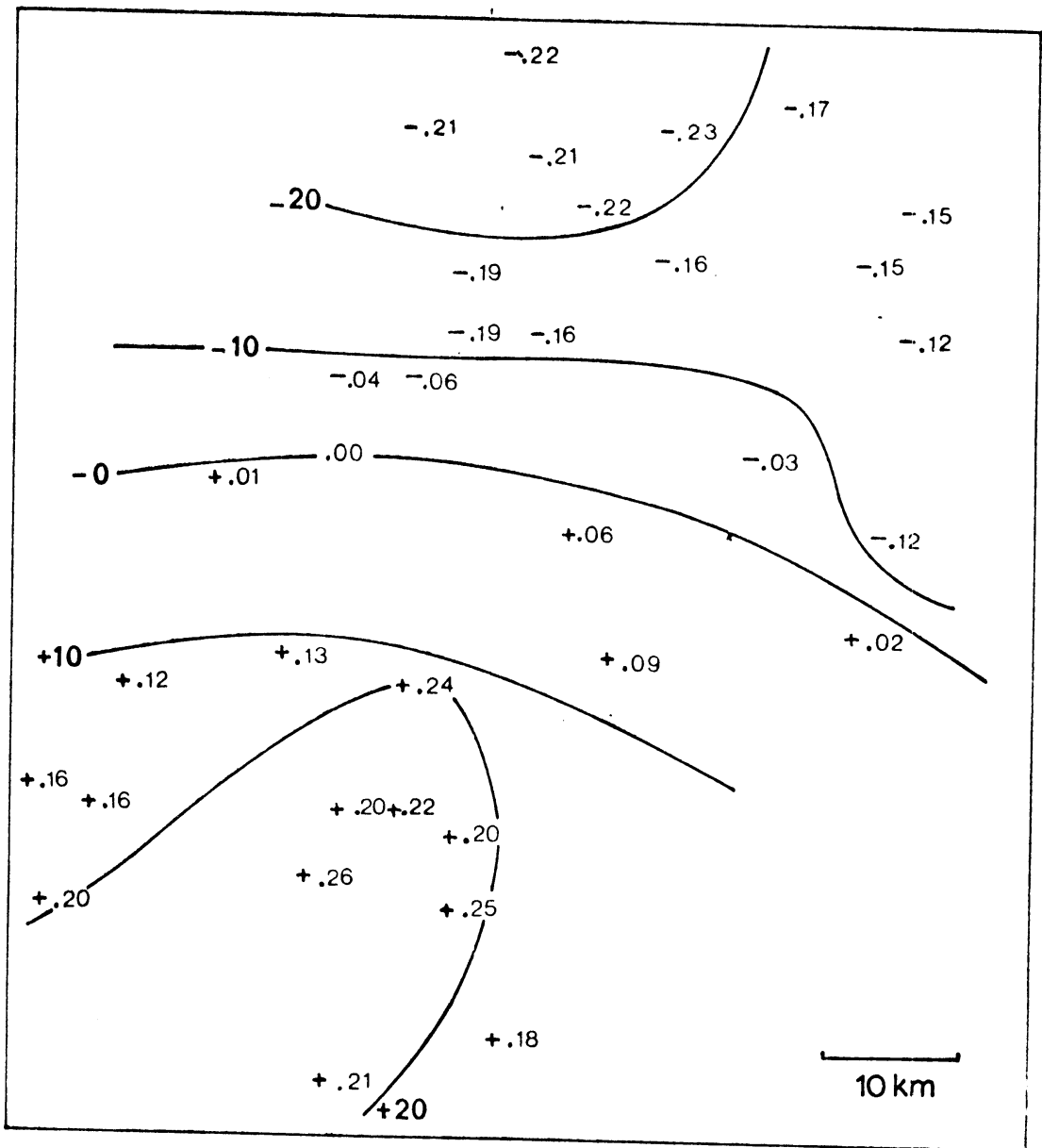
Tabel 2.3.1 De eerste zeven eigenwaarden van de samenhang-matrix S.

i	Lente	Zomer	Herfst	Winter
1	417,3	988,3	1035,0	759,6
2	18,7	69,0	37,8	19,2
3	9,7	43,9	23,6	9,6
4	4,6	21,9	13,2	4,7
5	3,6	19,4	9,5	3,2
6	3,3	17,0	8,6	2,7
7	3,0	12,8	6,6	2,4

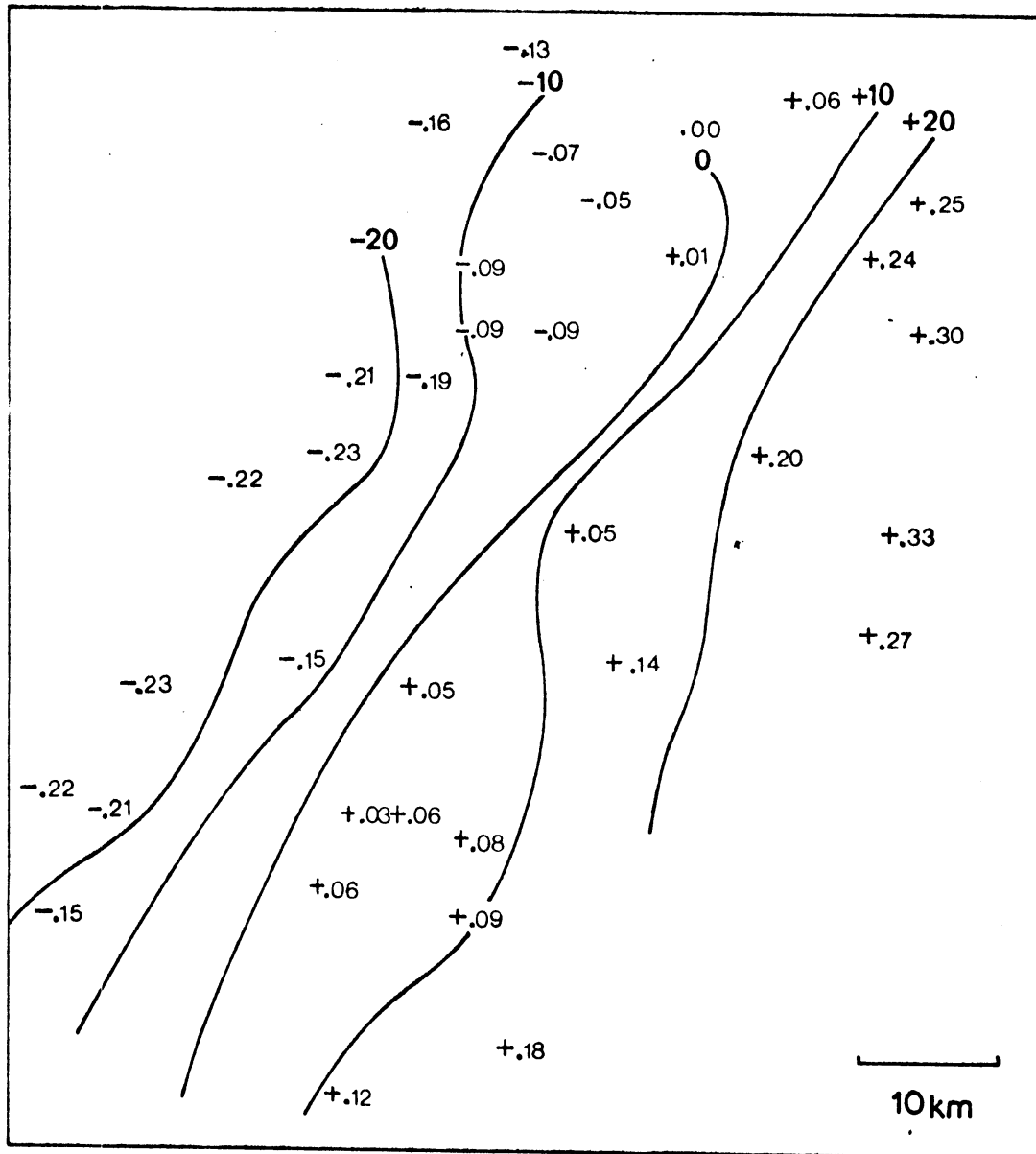
De bij de eerste eigenwaarde behorende eigenvector geeft het verwachte beeld, namelijk dat in eerste aanleg de neerslag totalen op alle stations gelijk op en neer zullen gaan, de componenten van de vector zijn onderling vrijwel gelijk. De volgende eigenvectoren vertonen meer structuur, in de figuren 2.3.1 t/m 2.3.3 zijn de eigenvectoren 2 t/m 4 van de lente in kaart gebracht. Vergelijkt men nu de eigenvectoren 2 en 3 van de lente met de overeenkomstige eigenvectoren van de andere seizoenen dan blijken zeer grote overeenkomsten te bestaan. (Tekonomkeer heeft bij eigenvectoren geen betekenis). Op grond hiervan kan gesteld worden dat dit geen steekproef-effecten zijn doch reële effecten. Bij eigenvector 4 is geen duidelijke overeenstemming meer, zodat men kan aannemen dat deze en hoger genummerde eigenvectoren steekproef-effecten beschrijven. In hoeverre deze aanname juist is dient nader onderzocht te worden. We willen in dit rapport echter alleen attenderen op het bestaan van deze techniek en de mogelijkheid om hiermee bijzondere effecten in de neerslag te detecteren.

De tweede **eigenvector** duidt erop dat er in alle seizoenen een anticorrelatie bestaat tussen het noordelijk en het zuidelijk deel van het gebied, dit effect is het belangrijkste in de zomer. De derde **eigenvector** geeft aan dat er tevens een anticorrelatie bestaat tussen kust en binnenland; ook dit effect is het belangrijkste in de zomer.

* (De isolijnen in de kaarten zijn met de hand getrokken).

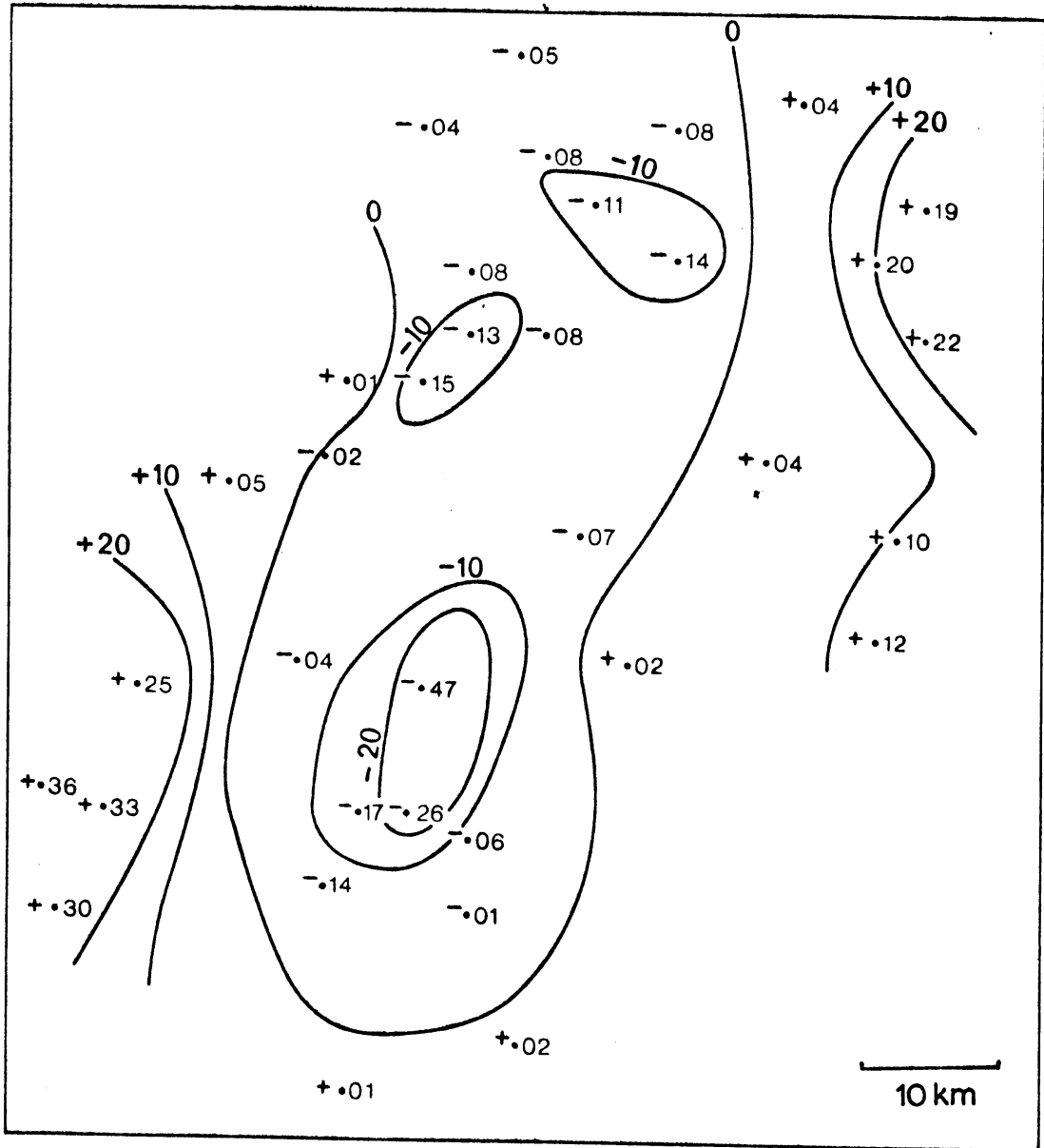


Figuur 2.3.1. Tweede eigenvector in de lente
 $\lambda_2 / \lambda_1 = 0,045$.



Figuur 2.3.2. Derde eigenvector in de lente

$$\lambda_3 / \lambda_1 = 0,028.$$



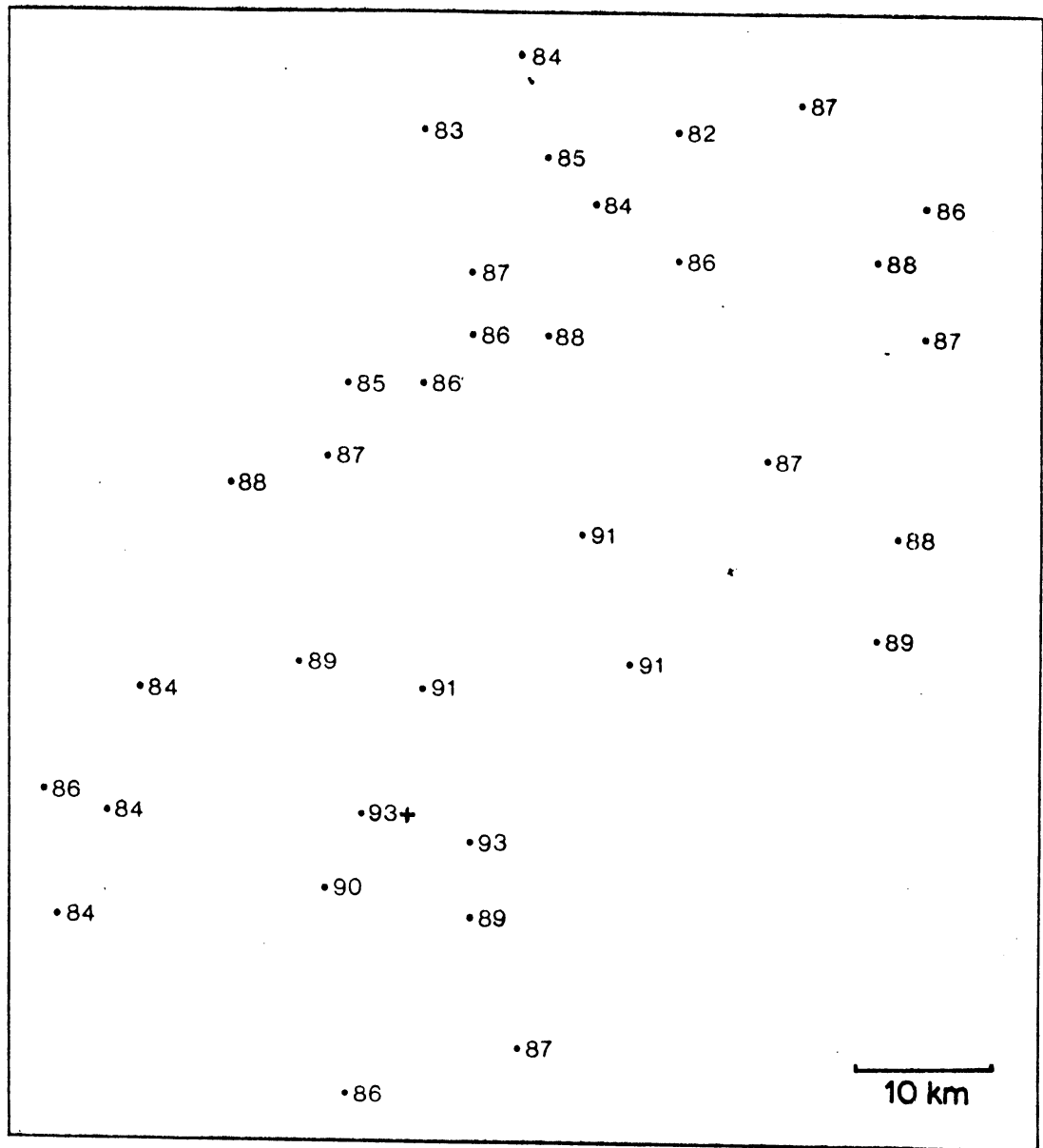
Figuur 2.3.3. Vierde eigenvector in de lente

$$\lambda_4 / \lambda_1 = 0,011.$$

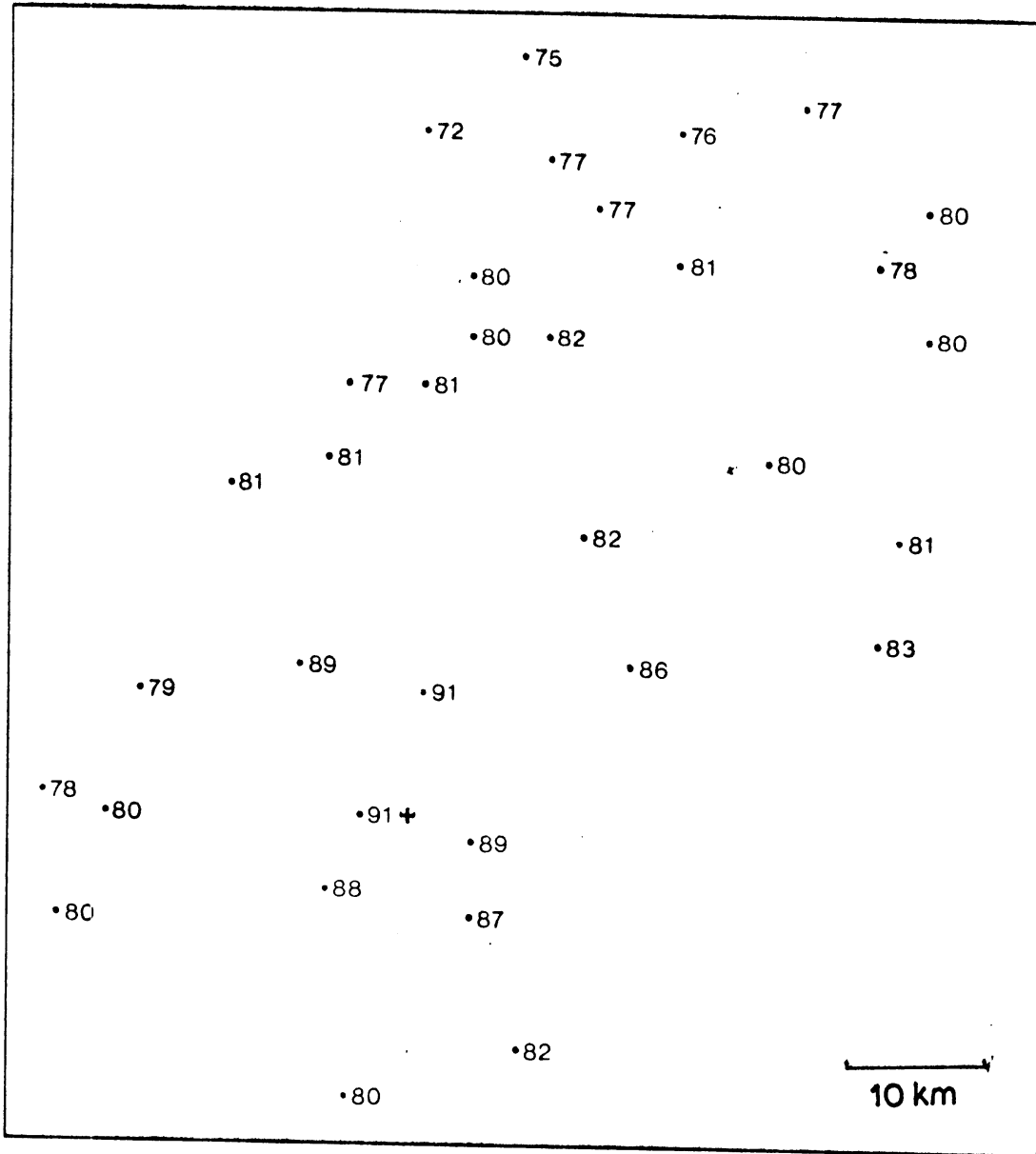
2.4. Conditionele kansen.

Lang niet altijd zal er behoefte bestaan aan een schatting van de hoeveelheid neerslag. In een aantal gevallen zal ook volstaan kunnen worden met aan te geven of een bepaalde grens op een bepaalde plaats al of niet is overschreden. Dit kan echter slechts met een bepaalde onzekerheid worden aangegeven. Voor deze probleemstelling is het interessant om na te gaan hoe groot de kans is dat op een willekeurig gekozen plaats een gegeven grens wordt bereikt of overschreden, onder de voorwaarde dat op een andere gegeven plaats die grens is bereikt of overschreden, de zogenaamde conditionele kans. In de figuren 2.4.1 t/m 2.4.4. zijn voor de grenzen 0,3 mm en 2,0 mm de conditionele kansen voor de zomer en de winter gegeven voor alle overige stations onder de voorwaarde dat op station 13 de grens is bereikt of overschreden. De klimatologische kans dat zonder voorwaarde de grens van 0,3 mm wordt bereikt of overschreden is voor seizoen 1 gelijk aan 0,43 en respectievelijk 0,44, 0,46 en 0,53 voor de overige seizoenen. Voor de grens van 2,0 mm zijn deze kansen 0,23, 0,29, 0,29 en 0,31. Uit de figuren 2.4.1 t/m 2.4.4. blijkt duidelijk dat voor afstanden variërend van 10 tot 60 km de conditionele kans nauwelijks afhangt van de afstand. Alleen op zeer korte afstand stijgt deze kans aanmerkelijk. Wil men deze kans dus ingrijpend opvoeren dan impliceert dat een zeer dicht netwerk.

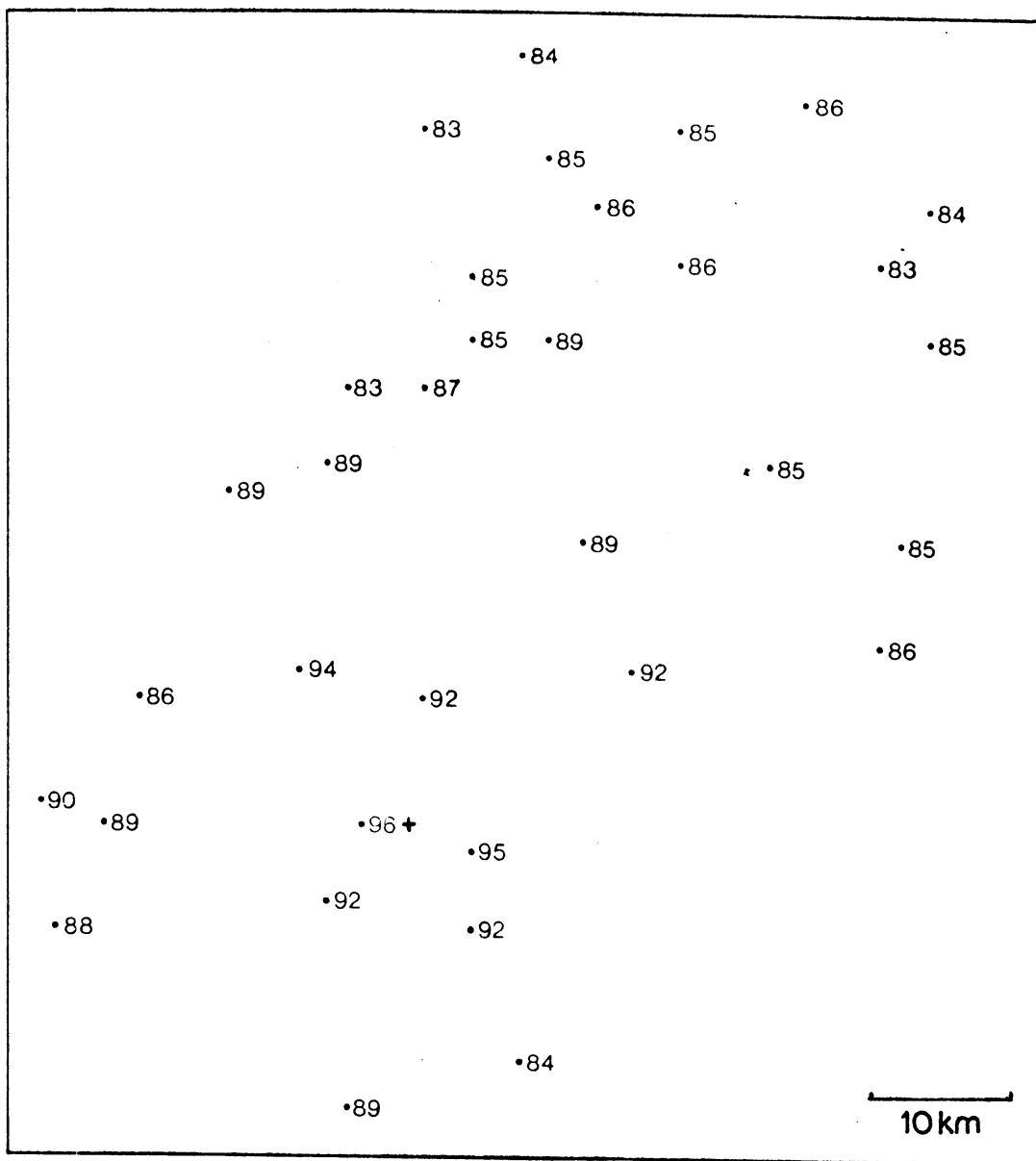
N.B. De conditionele kansen werden berekend op basis van ongestandaardiseerde neerslagsommen.



Figuur 2.4.1. Conditionele kansen in de zomer
Grenshoogte 0,3 mm.



Figuur 2.4.2. Conditionele kansen in de zomer
Grenshoogte 2.0 mm.



Figuur 2.4.4. Conditionele kansen in de winter
Grenshoogte 2,0 mm.,

2.5. Ruimtelijke variabiliteit.

Bij stochastische functies heeft de grootte van de ruimtelijke verschillen invloed op de nauwkeurigheid waarmee kan worden geïnterpoleerd. Een getal waarmee de ruimtelijke variabiliteit kan worden geparametriseerd in de deelruimte R voor een veld $a(\vec{r})$ is de ruimtelijke variantie

$$T_R^2 = \frac{1}{R} \int (a(r) - \frac{1}{R} \int a(r) dr)^2 dr \quad (2.5.1)$$

Voor de meeste velden van meteorologische grootheden zal deze parameter niet sterk variëren met de tijd zodat deze parameter niet expliciet in de interpolatiefout naar voren komt. Bij de neerslag varieert deze parameter juist heel sterk van dag op dag en in buiige situaties ook nog heel sterk van plaats tot plaats, zodat het noodzakelijk is hiermee rekening te houden. Deze ruimtelijke variabiliteit is voor het gegeven gebied onderzocht.

Uitdrukking (2.5.1) zou men kunnen benaderen voor een gegeven dag n met

$$T_R^2 \approx \frac{1}{K} \sum (h_{n,k} - \bar{h}_{n,k})^2 \quad (2.5.2)$$

Het nadeel van deze uitdrukking is dat een "uitschieter" (met uitschieter wordt hier bedoeld een sterke afwijkende waarde op één station, dat is in koufront situaties niet abnormaal) sterk bepalend is voor het resultaat, terwijl bij neerslag "uitschieters" tamelijk frequent optreden. Daarom is een andere maat genomen die niet zo gevoelig is voor "uitschieters" en wel

$$D_n = \frac{1}{K} \sum |h_{n,k} - \bar{h}_{n,k}| \quad (2.5.3)$$

Deze grootte zal ongeveer evenredig zijn met de wortel uit de variantie.

In de figuren 2.5.1 t/m 2.5.4 is per seizoen de waarde D_n , geplot tegen $\overline{h_{n,k}}^K$ (gebiedsgemiddelde) voor iedere dag waarvoor geldt $\overline{h_{n,k}}^K > 0.5$ mm.

N.B. Voor het lezen der figuren geldt de volgende handleiding. Het tweedimensionale veld is verdeeld in vakjes waarvan de breedte en de hoogte overeenkomt met respectievelijk 0,2 mm neerslag en 0,2 mm van de ruimtelijke variabiliteit. In ieder vakje is het aantal malen dat de gegeven situatie optrad in code genoteerd. De code wordt als volgt vertaald

Blank	:	Niet opgetreden	A	:	Tienmaal opgetreden
1	:	Eenmaal opgetreden			
			Z	:	35-maal opgetreden
9	:	Negen maal opgetreden	*	:	Meer dan 35-maal,

Combinaties van waarden die buiten het veld vallen worden genoteerd op de dichtstbijzijnde grens van de figuur.

Uit deze figuren blijkt duidelijk dat de waarde van D_n sterk kan variëren. Grote waarden van D_n komen, zoals verwacht, voor in de zomer in welk seizoen interpolatie dan ook het moeilijkst zal zijn. Verder blijkt duidelijk uit de figuren dat er een samenhang is tussen het gebiedsgemiddelde en de mogelijke waarden voor D_n . Daarom is in dit rapport de interpolatie-fout in vele gevallen geselecteerd naar de neerslaghoeveelheid.

3. Onderzoek van enige schattingsmethoden.

3.1. Inleiding.

Zoals reeds gesteld in de inleiding zullen in dit rapport alleen schattingsmethoden worden onderzocht die kunnen worden weergegeven in de vorm,

$$\hat{h} = \sum_{k=1}^K (g_k \cdot h_k) \quad (3.1.1)$$

$$\text{met als eis } \sum_{k=1}^K g_k = 1 \quad (\hat{h} = h \text{ geschat})$$

De diverse schatters onderscheiden zich alleen door de waarden van de gewichten g_k . Door de gegeven eis en het gegeven dat met de gestandaardiseerde neerslag wordt gewerkt, wordt bereikt dat het gemiddelde van \hat{h} gelijk is aan het "gemiddelde" van de werkelijke h . Dit betekent dat alleen schatters in aanmerking worden genomen die niet systematisch fout zijn (zie ook appendix A).

De in (3.1.1) gegeven uitdrukking lijkt veel op de formules die gegeven worden voor de multiple lineaire regressie. Het ligt dan ook voor de hand om voor de keuze van g_k gebruik te maken van deze techniek, daarom is een van de onderzochte schatters een variant uit deze techniek. In de termen van de lineaire regressie is (3.1.1) een regressie welke door nul gedwongen wordt, de reden hiervoor is dat het onnatuurlijk is om iets anders dan nul te schatten indien op alle betrokken stations neerslag nul wordt gemeten.

De methode waarop de schatters worden onderzocht is zeer eenvoudig. Met iedere schattingsmethode wordt in een groot aantal gevallen de neerslag ter plaatse van een station geschat op basis van de gemeten neerslagsommen op een aantal andere stations. Door vergelijking van de geschatte neerslag met de gemeten neerslag kan de gemiddelde schattingsfout worden vastgesteld.

Door onderlinge vergelijking kan worden vastgesteld welke methode van schatten de beste is. Bovendien kan door het variëren van de keuze van stations waaruit geschat wordt, een indruk worden verkregen van de relatie tussen gemiddelde schattingsfout en netwerkdichtheid. In een aantal gevallen werden tevens figuren vervaardigd van geschat tegen gemeten zodat visueel de waarde van een gegeven schattingsfout kon worden beoordeeld.

Een nadeel van de hier geschetste methode is dat geen onderscheid gemaakt kan worden tussen meetfouten en schattingsfouten. Voor onderlinge vergelijking van de schattingsmethoden is dit niet zo belangrijk. Voor een eventueel advies of hetzij de netwerkdichtheid moet worden opgevoerd of de kwaliteit van de stations moet worden verbeterd is deze scheiding echter onmisbaar.

In het voorgaande zijn twee begrippen geïntroduceerd welke nog niet scherp zijn gedefinieerd en wel het begrip schattingsfout en het begrip netwerkdichtheid.

3.1.4 De gemiddelde schattingsfout Δ .

De meest gebruikelijke methode om de gemiddelde schattingsfout in een getal uit te drukken is via de standaardafwijking S van de schattingsfout. Indien men in een groot aantal gevallen N de neerslag zowel geschat als gemeten heeft kan men s berekenen uit

$$s = \sqrt{\frac{\sum (h - h_{\text{gemeten}})^2}{N}} \quad (3.1.2)$$

Een nadeel van deze parameter is dat het berekende resultaat sterk wordt beïnvloed door eventuele uitschieters. Om deze reden werd door ons een andere maat Δ gebruikt, welke als volgt werd gedefinieerd

$$\Delta = \sqrt{\frac{\sum \text{abs}(h - h_{\text{gemeten}})}{N}} \quad (3.1.3)$$

De hier gedefinieerde grootte zullen we aanduiden met Δ tot.

Op grond van paragraaf 2.5 werd echter verwacht dat de schattingsfout samen zou hangen met de neerslag hoeveelheid, vandaar dat de schattingsfouten veelal werden ingedeeld in klassen welke werden vastgelegd door de neerslaghoogte. De gemiddelde schattingsfout binnen een klasse duiden we aan met Δ_h . Deze Δ_h werd op overeenkomstige manier berekend als Δ_{tot} .

3.1.b De netwerkdichtheid.

Alhoewel dit begrip veelvuldig wordt gehanteerd bestaat er geen algemeen aanvaarde en e nduidige definitie waarmee dit begrip kan worden gekwantificeerd. Over het algemeen maakt men gebruik van het gemiddelde aantal stations per vierkante kilometer, berekend over een groot gebied. Deze definitie voldoet niet, daar in veel gevallen (in het bijzonder bij onregelmatige netwerken) de netwerkdichtheid een lokale grootte is die bovendien geen rechtstreeks verband kan houden met de schattingsfout daar hierbij het verdelingspatroon van de stations en de schattingsmethode ook een rol spelen. Het duidelijkst wordt dit gedemonstreerd in het geval dat de neerslag op een gegeven plaats, geschat wordt met de neerslag van het dichtstbijzijnde station. In dit geval is de eenvoudigste maat voor de bruikbaarheid op die gegeven plaats, de afstand tot het dichtstbijzijnde station. Hoeveel andere stations er nog zijn en op welke afstand, is voor de genoemde schatting niet van belang. Door ons zal als maat voor de bruikbaarheid van netwerk plus schattingsmethode, het gewogen gemiddelde van de afstanden van de schattende stations worden gebruikt,

$$R = \sum_{k=1}^K g_k * r_k \quad (3.1.4)$$

waarin g_k : de gewichten identiek aan die in (3.1.1)

en r_k : de afstand tot het k^e station.

3.2. De éénstationsschatting.

De eenvoudigste schatting is wel die waarbij men als schatting voor de neerslag op een gegeven plaats, de neerslag op het dichtstbijzijnde station neemt. Zowel formule 3.1.1 als formule 3.1.4 worden dan zeer eenvoudig. De prestaties van deze schatting zijn onderzocht door de neerslag op een gegeven station met behulp van andere stations op verschillende afstanden te schatten. In tabel 3.2.1 zijn een aantal resultaten samengevat. Hierbij werd onderscheid gemaakt tussen de verschillende seizoenen daar zeker verwacht mag worden, dat de schattingsfout af zal hangen van het seizoen. Verder is het materiaal ook gesplitst naar een aantal neerslagklassen daar tevens verwacht werd dat de schattingsfout hier mee samen zal hangen. De indeling naar neerslagklassen werd gemaakt op basis van de neerslag op het station waarvan de schatting werd gemaakt (in dit geval steeds station 13). Onderzoek van tabel 3.2.1 leert dat zowel het seizoenseffect als het neerslaghoogte-effect aanwezig is. Verder kan opgemerkt worden dat er ook een relatie is met de afstand. In figuur 3.2.1 is de relatie tussen neerslaghoogte en interpolatie-fout in beeld gebracht voor station 22. Men ziet hieruit dat de fout Δ (h) ruwweg evenredig is met de wortel uit de neerslaghoogte. In het praktijkgeval waarbij de neerslaghoogte niet bekend is kan men de fout schatten door voor de neerslaghoogte de schatting daarvan te gebruiken.

Voor nader onderzoek van de relatie tussen de interpolatie-fout en de afstand is het nuttig om eerste na te gaan of deze relatie nog afhangt van de neerslaghoogte. In de figuren 3.2.2 en 3.2.3 is daarom voor de stations uit tabel 3.2.1 voor een aantal neerslagklassen (1.0-2.0, 4.0-6.0 en 10.0-20.0) de interpolatie-fout Δ (h) tegen de totale interpolatie-fout Δ_{tot} , uitgezet op dubbel logaritmische schaal. De rechte lijnen in deze figuren zijn getrokken met de helling één. Uit deze figuren kan men concluderen dat er geen grote verschillen zullen bestaan tussen de afstandsrelaties zodat het voldoende is om de afstandsrelatie te onderzoeken aan de hand van Δ_{tot} .

In tabel 3.2.2 zijn alle combinaties vermeld die gebruikt zijn voor het onderzoek van de afstandsrelatie. In de figuren 3.2.4 t/m 3.2.7 is de afstandsrelatie voor de verschillende seizoenen in beeld gebracht. Er is in deze figuren onderscheid gemaakt tussen situaties met station 13 als referentie en situaties met andere stations als referentie. Uit de figuren 3.2.4 en 3.2.7 blijkt dat een eventueel plaatseffect niet erg belangrijk kan zijn. Om tevens visueel te kunnen beoordelen wat een gegeven gemiddelde absolute fout in feite betekend, zijn voor een aantal situaties de gemeten waarden h tegen de geschatte waarde \hat{h} uitgezet. Als voorbeelden worden hier de figuren 3.2.8 t/m 3.2.9 gegeven, waarin de prestatie in zomer en winter van station 22 als schatter van station 13 wordt weergegeven. In de figuren 3.2.10 en 3.2.11 is dit gedaan voor station 10. Voor de interpretatie van de figuren wordt verwezen naar de nota bene in paragraaf 2.5. Bij deze figuren is de vakjesbreedte en hoogte anders gekozen en wel verticaal 0,5 mm en horizontaal 0.3 mm.

Tabel 3.2.1 Resultaten éénstationsschatter.
 Schatting van de neerslag op station 13
 met stations op verschillende afstand.

Lente

Gemiddelde absolute fout

Neerslag klasse	Aantal dagen	Station						
		22	17	18	15	10	4	5
0	550	.035	.045	.044	.042	.141	.153	.191
0.1 - 0.5	131	.19	.16	.30	.35	.40	.47	.54
0.5 - 1.0	83	.40	.43	.62	.75	.82	1.00	1.13
1.0 - 2.0	82	.61	.82	.79	.75	.93	1.11	1.20
2.0 - 4.0	88	.78	1.11	1.20	1.26	1.28	2.08	2.04
4.0 - 6.0	71	1.00	1.09	1.34	1.40	1.79	2.40	2.79
6.0 -10.0	54	1.12	1.59	2.02	2.46	2.31	3.38	3.09
10.0-20.0	39	2.34	2.42	2.97	3.39	2.95	4.10	4.23
Totaal	1104	.41	.50	.59	.65	.73	.97	1.03
Afstand in km.		3.5	5.0	9.0	13.5	23.5	35.0	53.5

Zomer

Neerslag klasse	Aantal dagen	Station						
		22	17	18	15	10	4	5
0	549	.016	.092	.091	.139	.170	.197	.246
.1 - .5	122	.25	.48	.50	.65	.60	1.02	1.12
.5 - 1.0	44	.46	.45	1.05	1.00	1.42	1.38	1.53
1.0 - 2.0	76	.94	.88	1.33	1.54	1.23	1.86	2.92
2.0 - 4.0	103	1.02	1.18	1.71	1.74	1.67	2.39	2.63
4.0 - 6.0	52	1.30	1.76	2.21	3.17	2.31	4.20	4.64
6.0 -10.0	70	1.85	2.20	3.08	2.92	4.11	4.01	4.18
10.0-20.0	62	3.22	4.87	4.65	5.28	5.40	6.78	6.11
Totaal	1104	.75	.99	1.21	1.37	1.46	1.84	1.98

N.B. In de rij totaal zijn ook de resultaten voor $h > 20.0$ mm verwerkt.

Tabel 3.2.1 vervolg.

Herfst

Neerslag klasse	Aantal dagen	Station						
		22	17	18	15	10	4	5
0	496	.030	.048	.063	.077	.084	.151	.118
.1 - .5	134	.18	.19	.26	.30	.30	.41	.47
.5 - 1.0	59	.48	.38	.81	.87	.81	1.06	1.06
1.0 - 2.0	88	.65	.79	.91	1.09	1.12	1.63	1.85
2.0 - 4.0	89	.87	.98	1.24	1.44	1.72	1.96	2.04
4.0 - 6.0	65	1.33	1.41	1.95	2.32	2.24	2.65	2.99
6.0 -10.0	73	1.43	1.54	2.59	2.85	3.02	3.26	3.86
10.0-20.0	61	2.55	2.31	3.78	4.19	4.21	4.89	5.51
Totaal	1092	.60	.66	.95	1.09	1.16	1.36	1.49

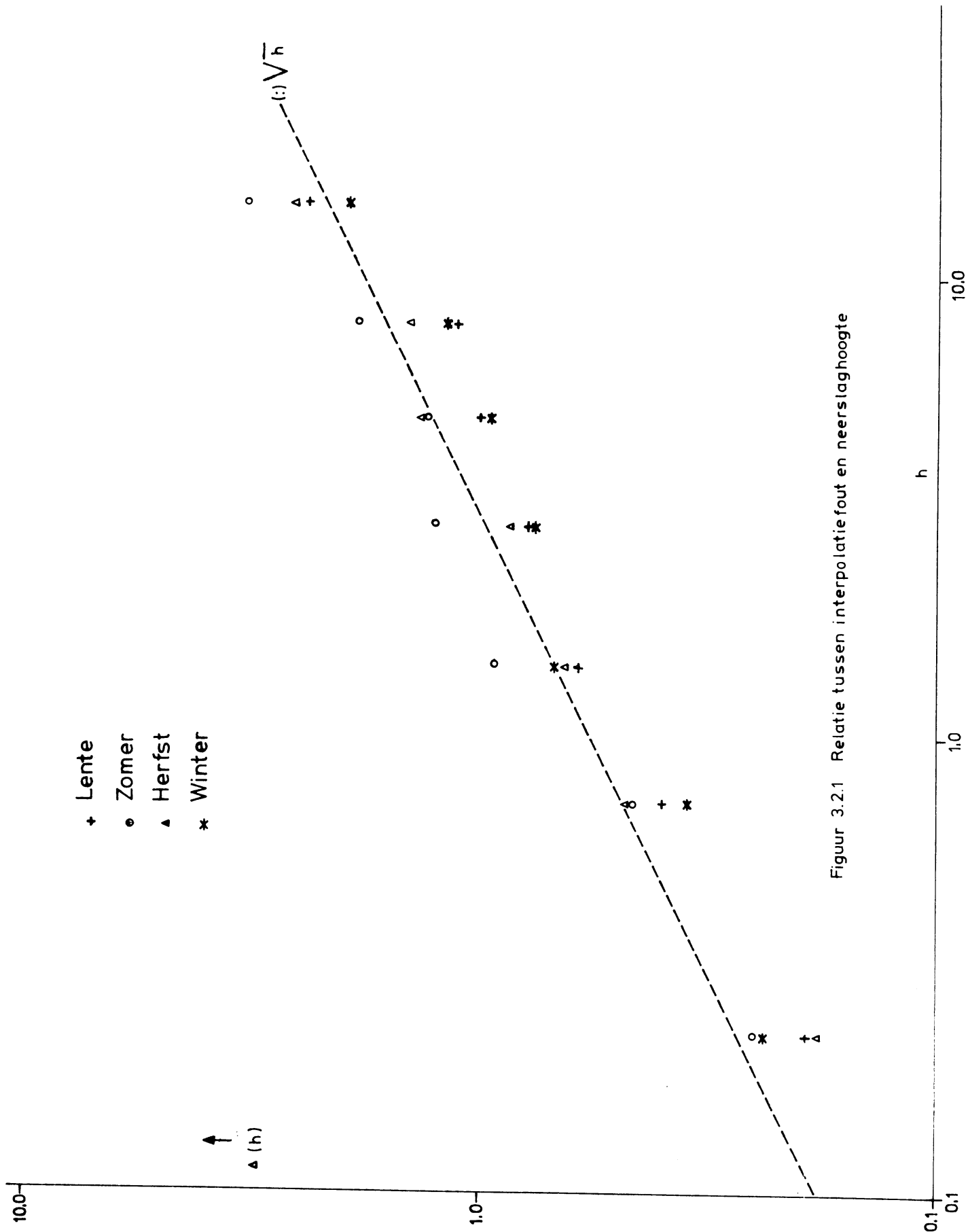
Winter

Neerslag klasse	Aantal dagen	Station						
		22	17	18	15	10	4	5
0	405	.037	.048	.046	.081	.084	.121	.124
.1 - .5	151	.24	.22	.25	.26	.33	.48	.45
.5 - 1.0	91	.35	.45	.76	.56	.68	.89	.92
1.0 - 2.0	103	.69	.54	.85	.84	1.01	1.24	1.32
2.0 - 4.0	115	.76	.68	.98	.89	1.06	1.47	1.61
4.0 - 6.0	80	.95	1.12	1.43	1.46	1.50	1.99	2.35
6.0 -10.0	76	1.18	1.16	1.93	1.52	1.99	2.73	2.93
10.0-20.0	51	1.95	2.22	3.46	3.11	3.65	4.57	4.95
Totaal	1083	.50	.50	.75	.69	.81	1.08	1.17

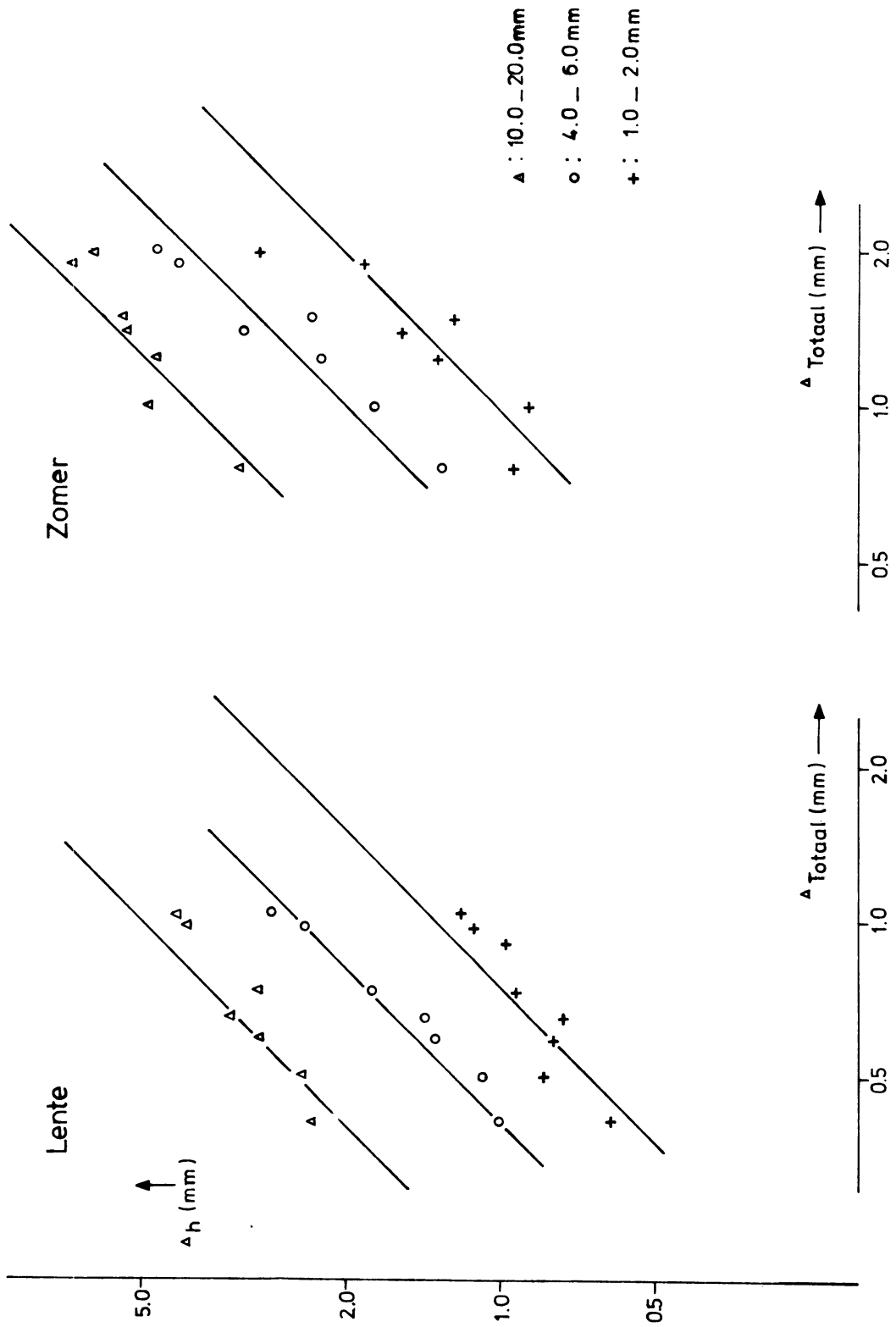
Tabel 3.2.2 Relatie tussen afstand en totale schattingsfout bij de éénstationsschatter.

Δ_{tot} = Gemiddelde absolute fout.

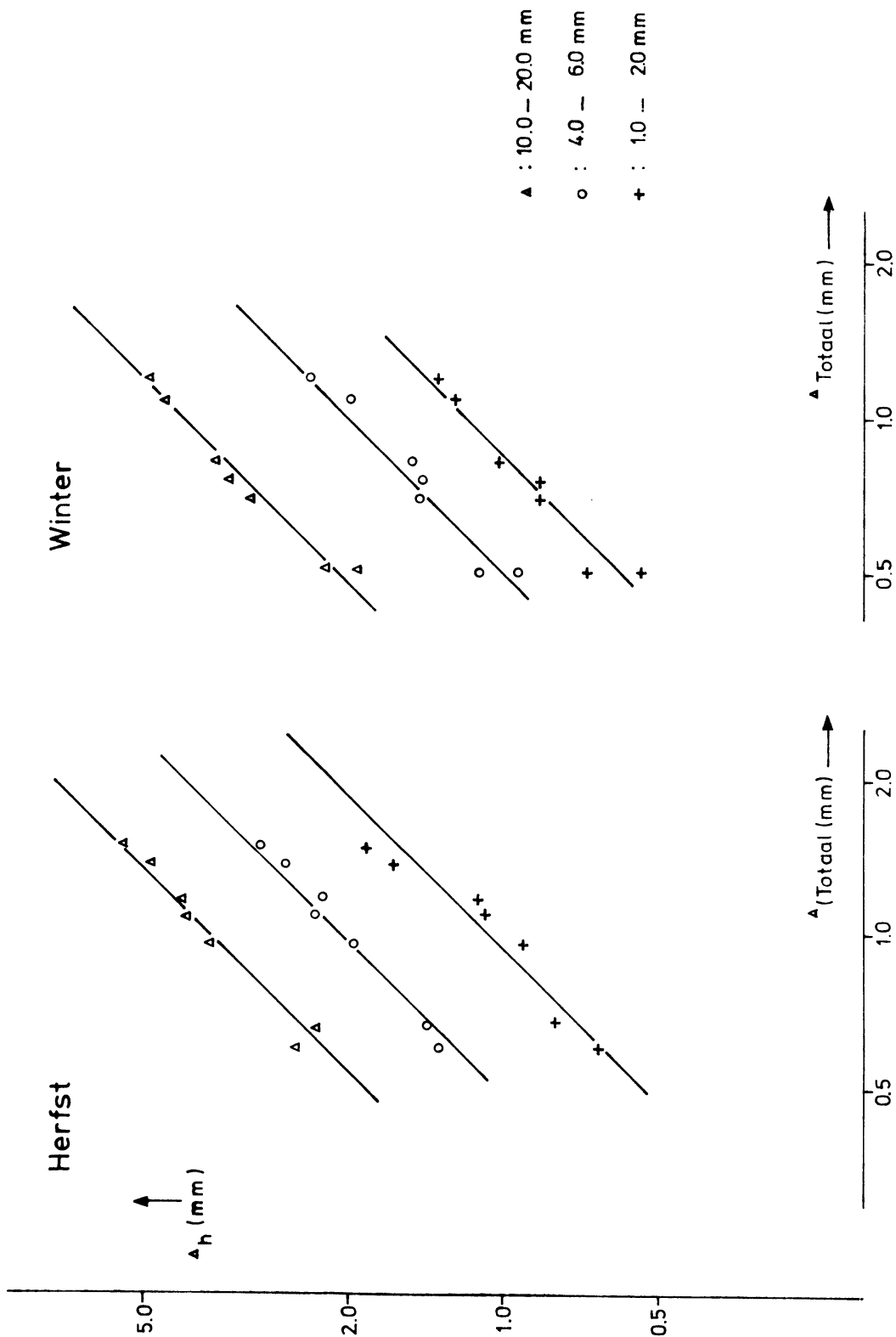
Stations	Afstand km.	Lente Δ	Zomer Δ	Herfst Δ	Winter Δ
13 - 22	3,5	0,41	0,75	0,60	0,50
13 - 17	5,0	0,50	0,99	0,66	0,50
13 - 18	9,0	0,59	1,21	0,95	0,75
13 - 10	23,5	0,73	1,46	1,16	0,81
13 - 28	8,0	0,53	1,03	0,80	0,64
13 - 24	8,5	0,58	1,24	0,89	0,63
13 - 26	22,0	0,78	1,50	1,24	0,86
13 - 15	13,5	0,65	1,46	1,09	0,69
13 - 5	53,5	1,03	1,98	1,49	1,17
13 - 4	35,0	0,97	1,84	1,36	1,08
12 - 29	5,5	0,51	0,99	0,71	0,52
12 - 27	5,5	0,45	0,92	0,84	0,58
12 - 4	9,5	0,61	1,10	0,86	0,65
12 - 7	15,0	0,64	1,39	0,94	0,70
12 - 8	11,5	0,51	1,00	0,88	0,64
4 - 19	4,5	0,41	0,92	0,75	0,56
4 - 28	41,5	1,01	1,91	1,41	1,10
4 - 31	33,0	0,90	1,58	1,21	0,90
9 - 33	39,0	0,92	1,76	1,48	1,07
11 - 19	30,5	0,86	1,94	1,22	1,05
21 - 24	32,0	0,80	1,74	1,38	1,00



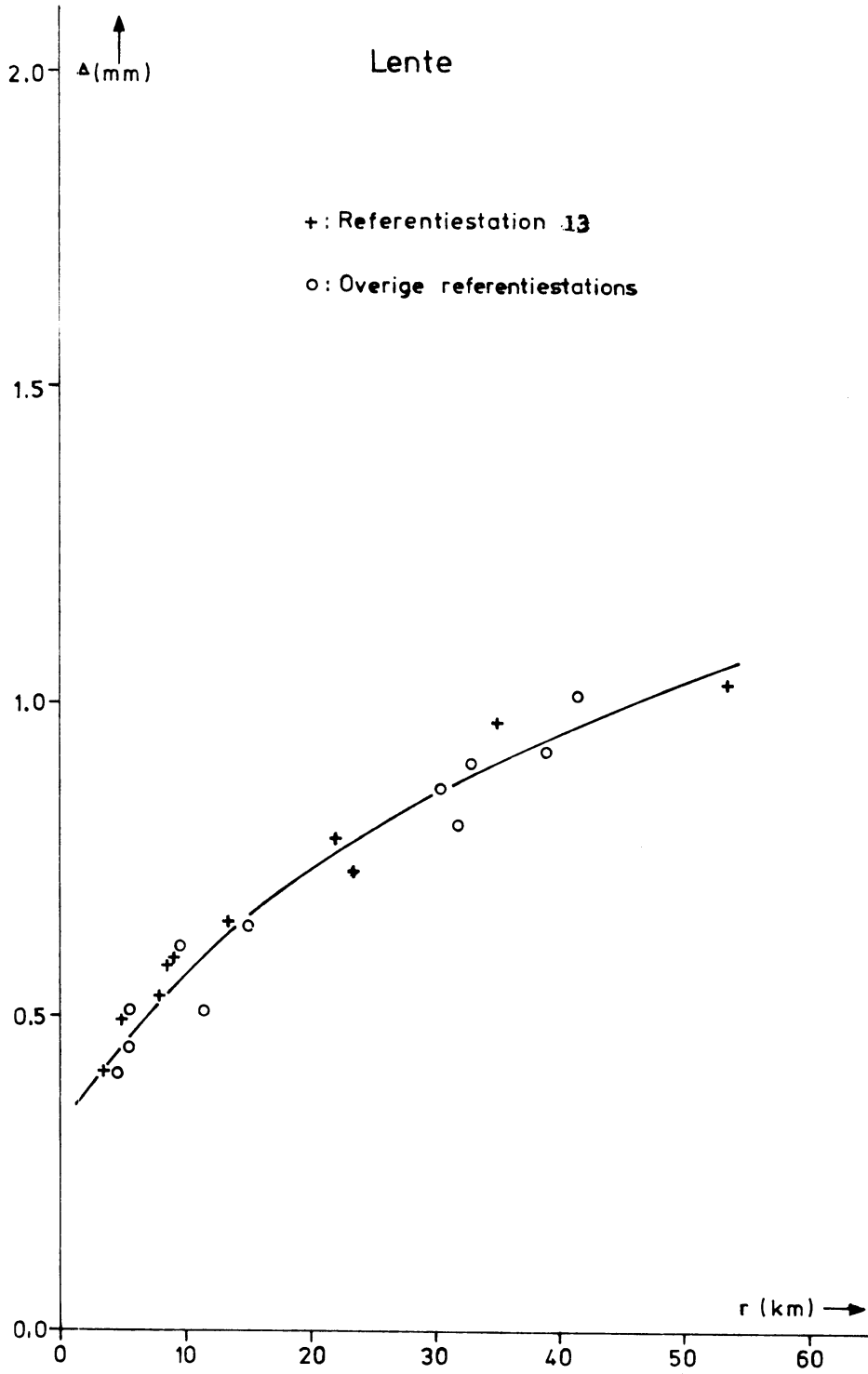
Figuur 3.2.1 Relatie tussen interpolatiefout en neerslaghoogte



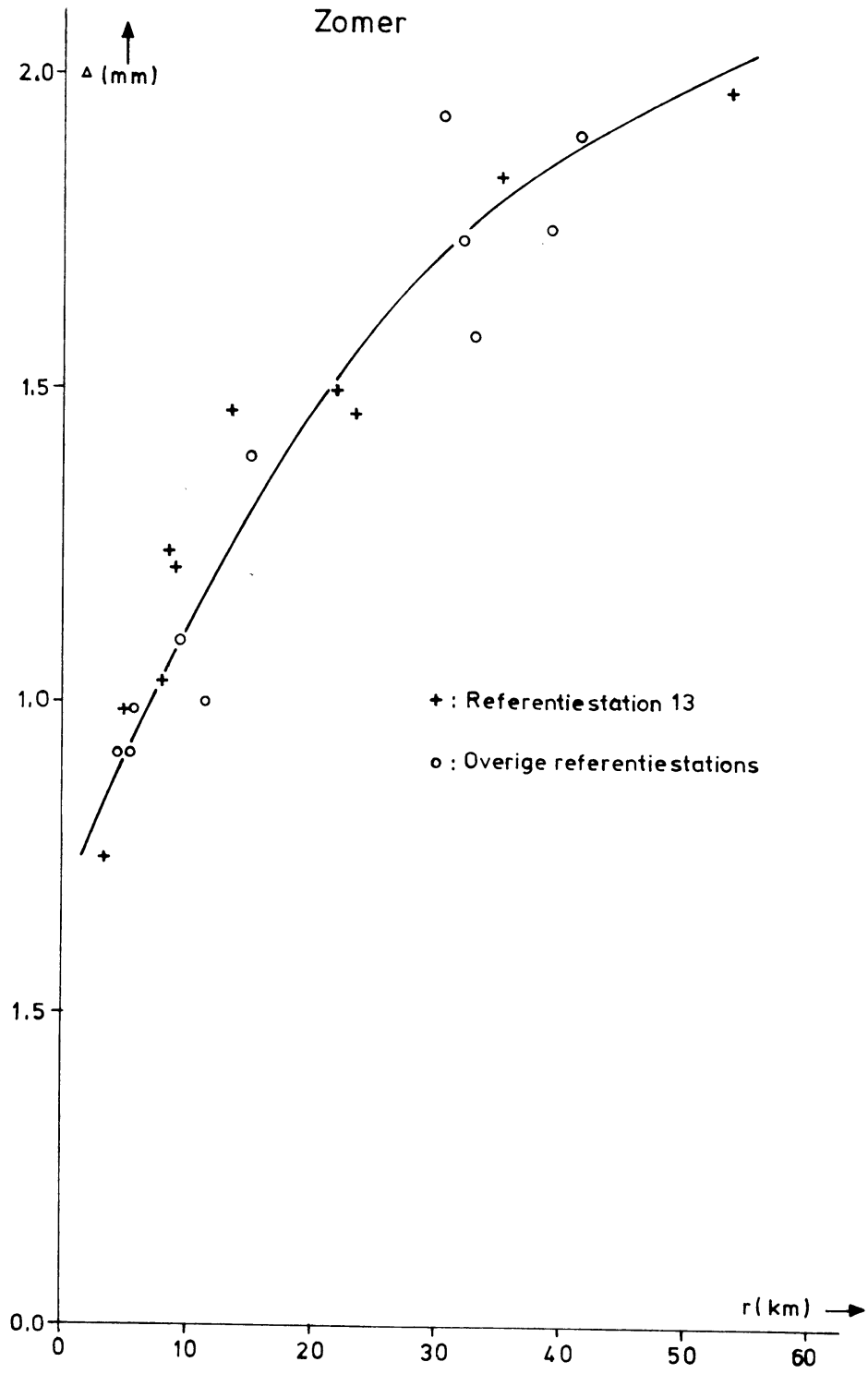
Figuur 3.2.2 Relatie tussen Δ tot. en Δh



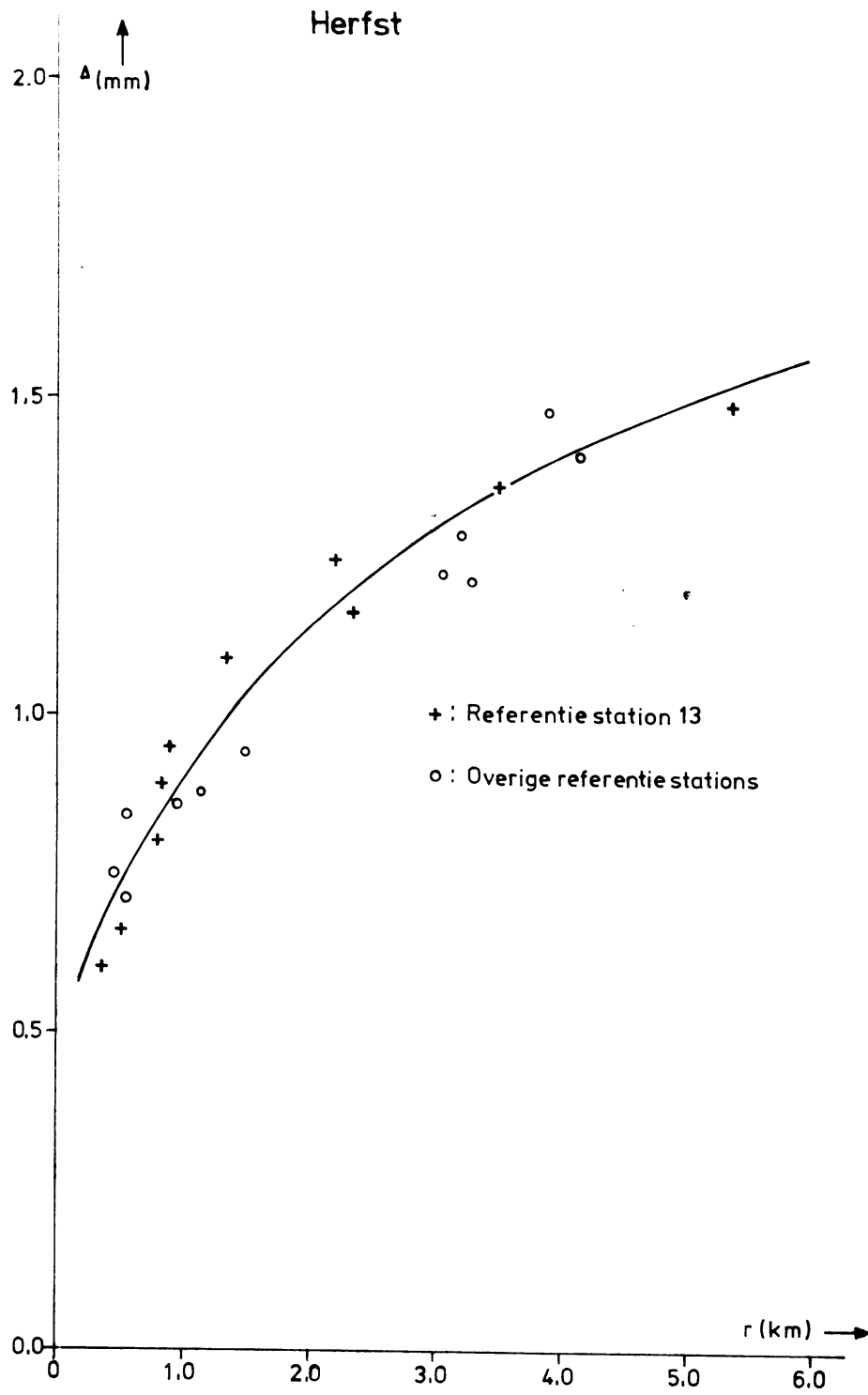
Figuur 3.2.3 Relatie tussen Δ_{tot} Δ_h



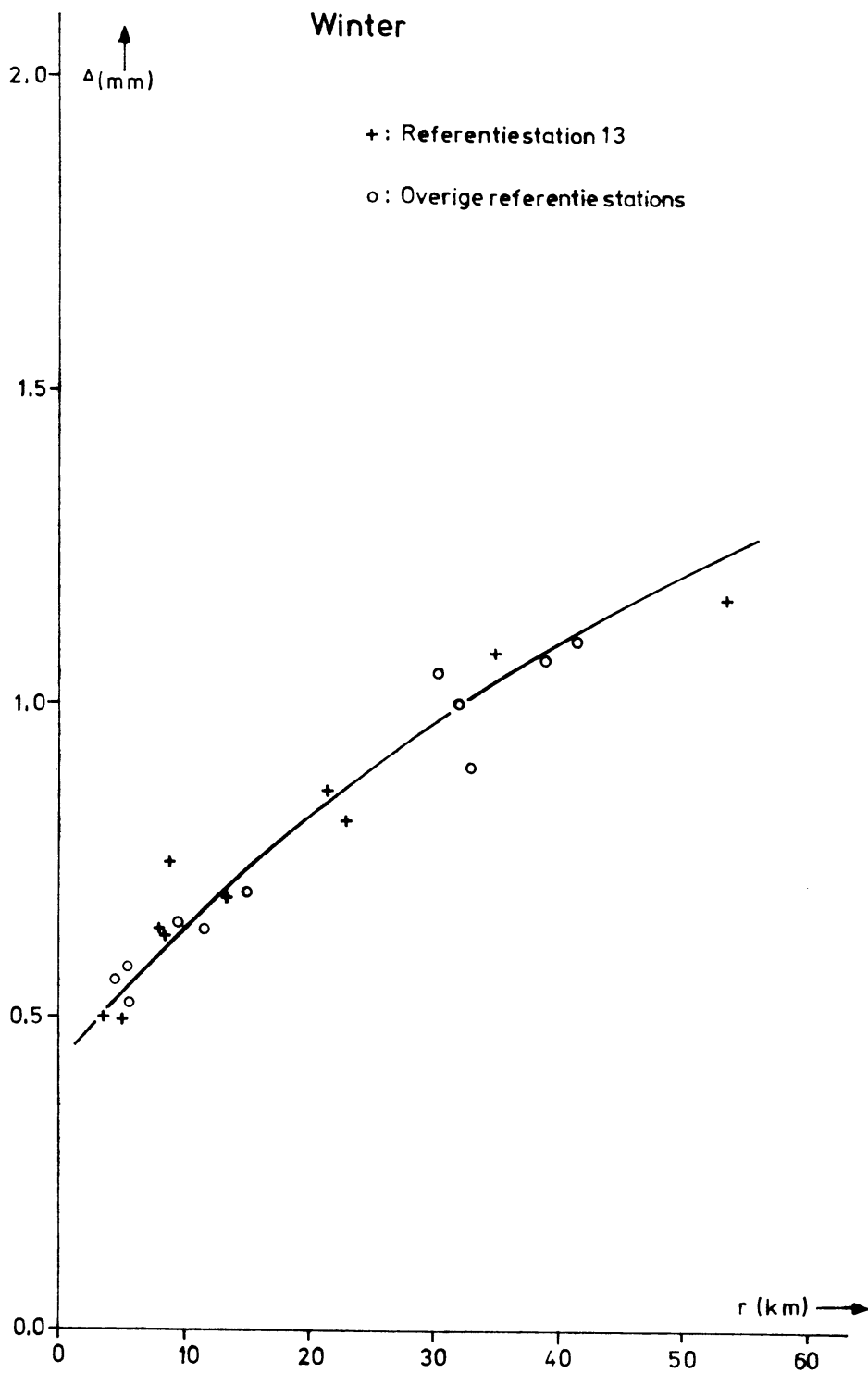
Figuur 3.2.4 Relatie Δ met afstand r
Eenstationsschatter



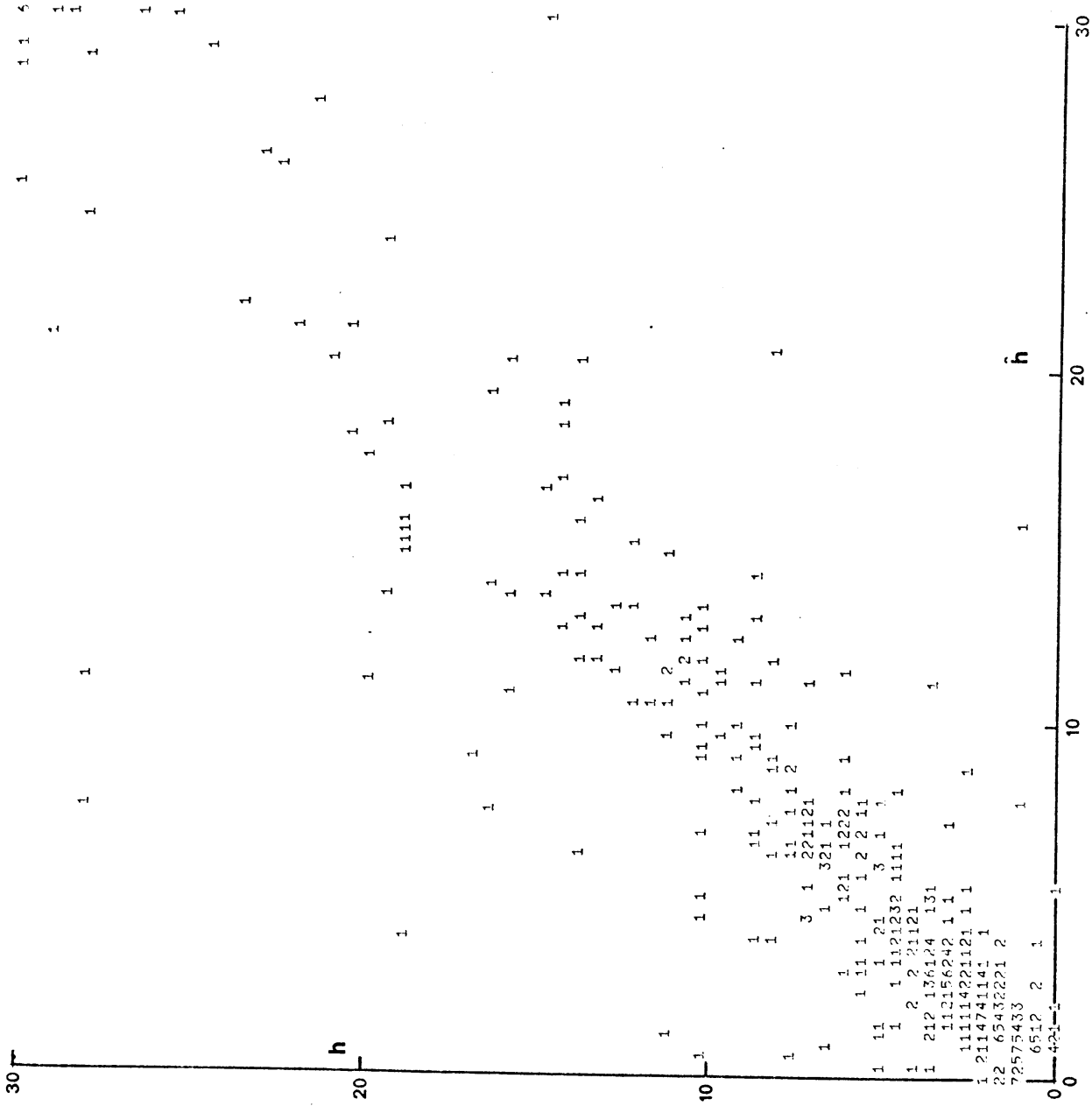
Figuur 3.2.5 Relatie Δ met afstand r
Eénstationsschatter



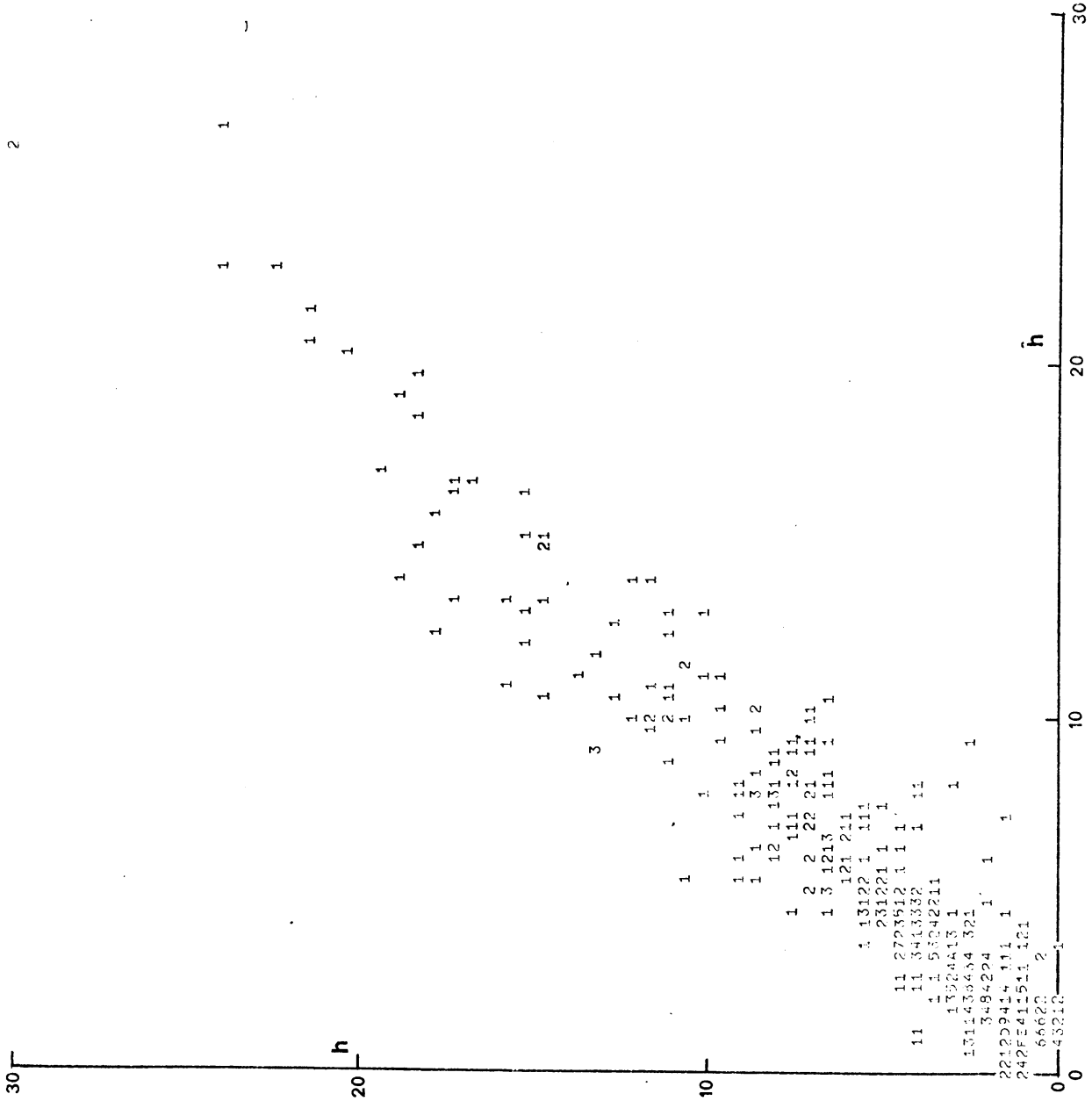
Figuur 3.2.6 Relatie Δ met afstand r
Eénstationschatter



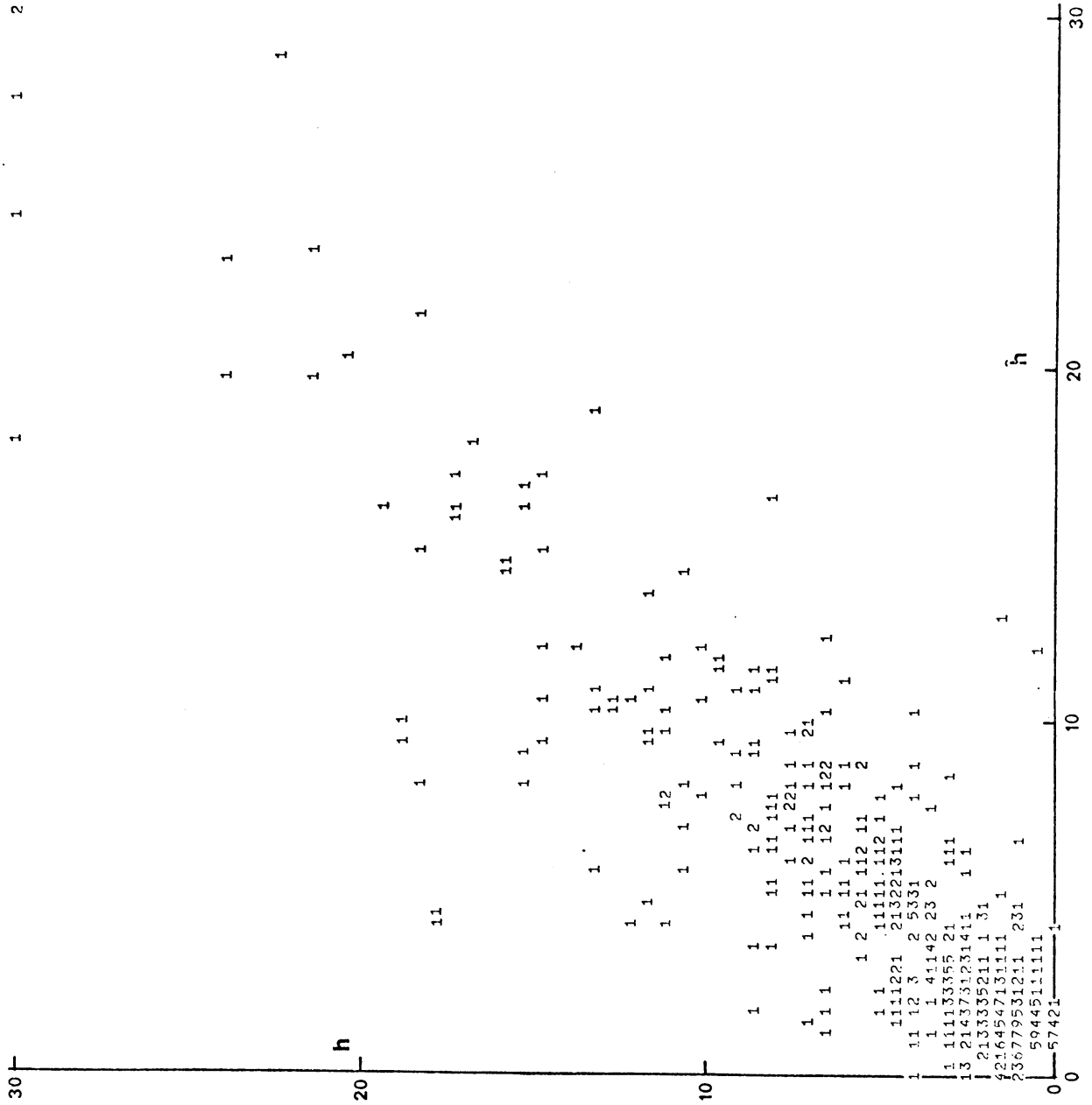
Figuur 3.2.7 Relatie Δ met afstand r
Eénstationschatter



Figuur 3.2.8 Verband tussen gemeten en geschatte neerslag van station 13 in de zomer (geschat uit station 22)



Figuur 3.2.9. Verband tussen gemeten en schatte neerslag van station 13 in de winter (geschat uit station 22)



Figuur 3.2.11 Verband tussen gemeten en geschatte neerslag van station 13 in de winter (geschat uit station 10)

3.3. Meerstationsschatters met gekozen gewichten.

De meeste schattingsmethoden maken gebruik van een beperkt aantal (K') omringende stations. De methoden verschillen onderling slechts door de wijze waarop de gewichten worden bepaald. In deze paragraaf zullen alleen methoden besproken worden waarbij de gewichten zijn vastgelegd op basis van een gekozen afstandsrelatie. De volgende afstandsrelaties zijn onderzocht,

a. g_k is onafhankelijk van de afstand, dus $g_k = 1/K'$.

b. g_k is omgekeerd evenredig met de afstand,

$$g_k = (1/r_k) / \left(\sum_{k=1}^{K'} 1/r_k \right).$$

c. g_k is omgekeerd evenredig met het kwadraat van de afstand,

$$g_k = (1/r_k^2) / \left(\sum_{k=1}^K 1/r_k^2 \right).$$

d) g_k neemt exponentieel af met de afstand,

$$g_k = \exp(-r_k/r_0) / \sum_{k=1}^K \exp(-r_k/r_0).$$

We kunnen bij alle methoden K' nog variëren terwijl bovendien bij methode d) r_0 nog gevarieerd kan worden. Bij de methoden a, b en c werd K' gekozen, de aanduiding voor deze methoden is a, b of c [K'], door keuze van de stations te variëren kon worden geëxperimenteerd met de netwerkdichtheid. Bij methode d werd r_0 gekozen terwijl K' werd bepaald door stations met een gewicht kleiner dan 0,001 te schrappen; deze methode wordt aangeduid met d [r_0, K']. De dichtheid kon worden gevarieerd door de dichtstbijzijnde stations uit te sluiten. We zullen van deze experimenten alleen Δ_{tot} presenteren, de samenhang met de neerslaghoogte vertoont hetzelfde gedrag als die van de éénstations-schatter. In tabel 3.3.1 zijn de resultaten van een aantal experimenten samengevat.

In de figuren 3.3.1 en 3.3.2 is voor de zomer en de winter, voor het experiment C [7] (4), aangegeven met een * in de tabel, een figuur gegeven van geschatte tegen gemeten waarden. Wat betreft Δ_{tot} is de situatie vrijwel identiek met de éénstationsschatter gebaseerd op station 22 versus 13. (figuren 3.2.8 en 3.2.9). Onderlinge vergelijking van de figuren geeft aan dat Δ_{tot} in beide gevallen ook min of meer hetzelfde betekent.

Op basis van tabel 3.3.1 kan worden geconcludeerd dat de keuze van de relatie gewicht-afstand vooral belangrijk is bij grote variatie in de afstanden, in het andere geval is dit nauwelijks merkbaar. De keuze van de soort afhankelijkheid is nauwelijks belangrijk. De prestaties van deze schatters steken verre uit boven de van de éénstationsschatter vooral bij wijde netten. Met een groep van 7 stations op gemiddeld 30 km afstand wordt ongeveer hetzelfde resultaat bereikt als met één station op tien kilometer afstand. Bij een dichtere netwerk is het effect niet zo belangrijk. Ook van de situatie C [7] (6) zijn een aantal figuren van geschatte tegen gemeten waarden gegeven (fig. 3.3.3 en 3.3.4).

Tabel 3.3.1 Enige resultaten meerstationsschatters met gekozen gewichten en met optimale gewichten.

Methode	R	Refer. station	Schattingsfout Δ_{tot}				zie onder
			Lente	Zomer	Herfst	Winter	
a [3]	7.3	13	.39	.83	.60	.48	1
a [3]	19.9	13	.55	1.22	.91	.64	2
a [34]	32.2	13	.61	1.25	.96	.71	3
a [7]	10.3	13	.39	.83	.60	.46	} 4
b [7]	7.6	13	.35	.72	.53	.43	
c [7]*	5.8	13	.34	.67	.51	.43	
opt.1	4.5	13	.35	.66	.51	.42	
opt.2	4.2	13	.34	.65	.50	.43	
a [7]	20.0	13	.53	1.15	.87	.61	} 5
b [7]	19.6	13	.52	1.15	.86	.60	
c [7]	19.0	13	.51	1.15	.86	.59	
opt.1	20.0	13	.51	1.11	.86	.57	
opt.2	19.6	13	.52	1.11	.84	.56	
a [7]	16.2	10	.52	.96	.72	.45	} 6
b [7]	15.2	10	.50	.93	.68	.43	
c [7]*	14.2	10	.47	.91	.66	.41	
opt.1	11.8	10	.44	.89	.63	.40	
opt.2	11.8	10	.44	.90	.63	.40	
a [7]	32.1	10	.60	1.19	.85	.58	} 7
b [7]	30.6	10	.60	1.17	.84	.54	
c [7]	29.3	10	.60	1.17	.83	.53	
opt.1	30.4	10	.55	1.14	.80	.52	
opt.2	30.0	10	.55	1.14	.80	.52	

Vervolg tabel 3.3.1.

Methode	R	Refer. station	Schattingsfout Δ_{tot}				zie onder
			Lente	Zomer	Herfst	Winter	
d [3,12]	5.4	13	.34	.67	.50	.42	8
d [5,22]	7.3	13	.35	.71	.52	.43	9
d [3*,14]	9.3	13	.41	.90	.64	.51	10
Functie 1	4.7	4	.41	.74	.57	.43	
Functie 2	8.8	10	.52	.97	.73	.50	
Functie 3	7.1	27	.42	.88	.78	.49	

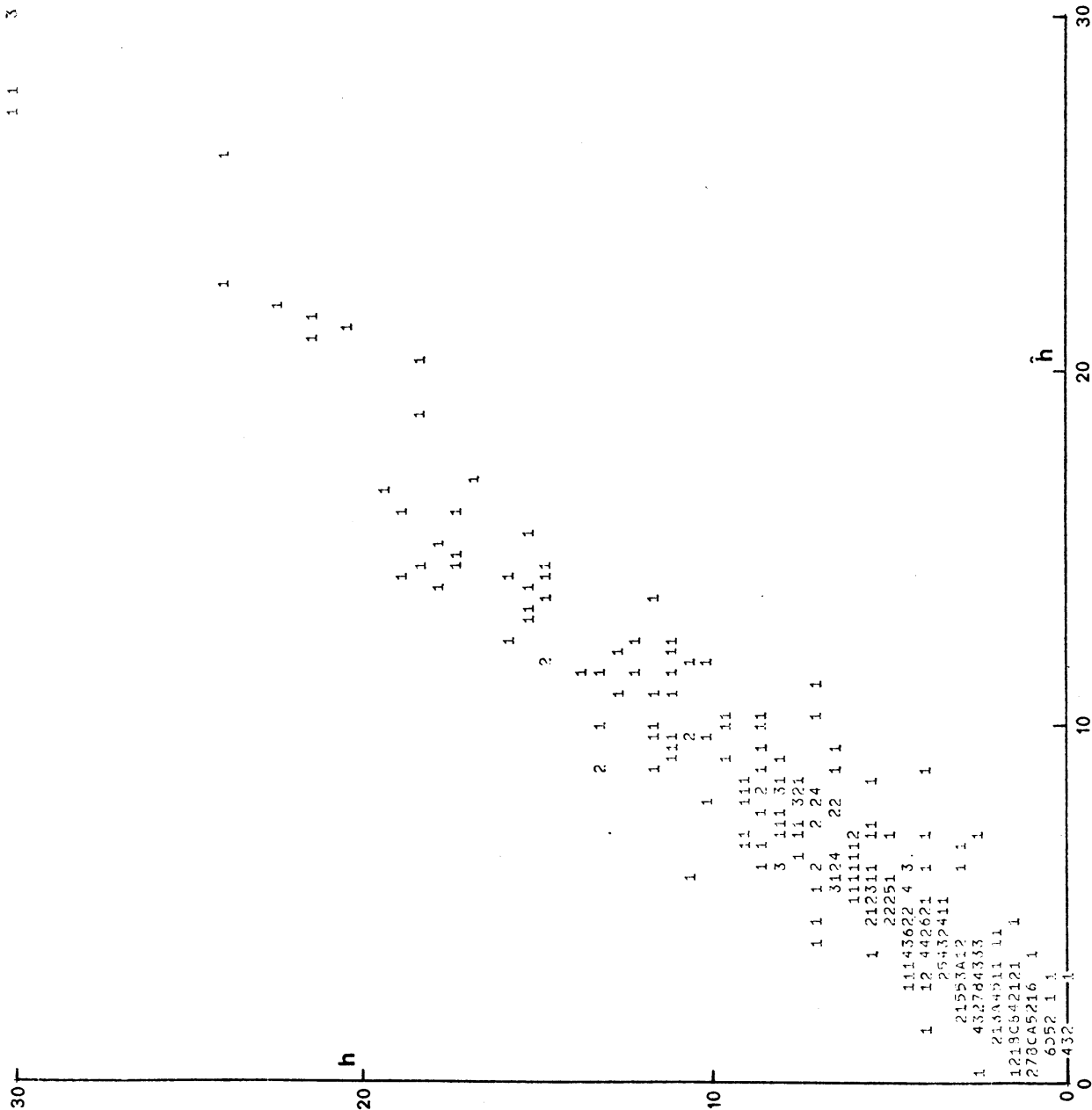
Stations waarmee geschat werd,

- 1 : Stations 17, 18, 28.
- 2 : Stations 11, 14, 20.
- 3 : Alle stations behalve station 13.
- 4 : Stations 15, 16, 17, 18, 22, 24, 28.
- 5 : Stations 10, 11, 14, 15, 16, 20, 25.
- 6 : Stations 7, 11, 15, 20, 27, 29, 30.
- 7 : Stations 2, 5, 8, 10, 25, 31, 33.
- 8 : Stations 11, 14, 15, 16, 17, 18, 20, 22, 24, 26, 28.
- 9 : Stations 4, 8, 10, 11, 12, 14, 15, 16, 17, 18, 20,
21, 22, 24, 25, 26, 27, 28, 29.
- 10 : Stations 8, 10, 11, 14, 15, 16, 18, 20, 21, 24, 25,
26, 27, 28.

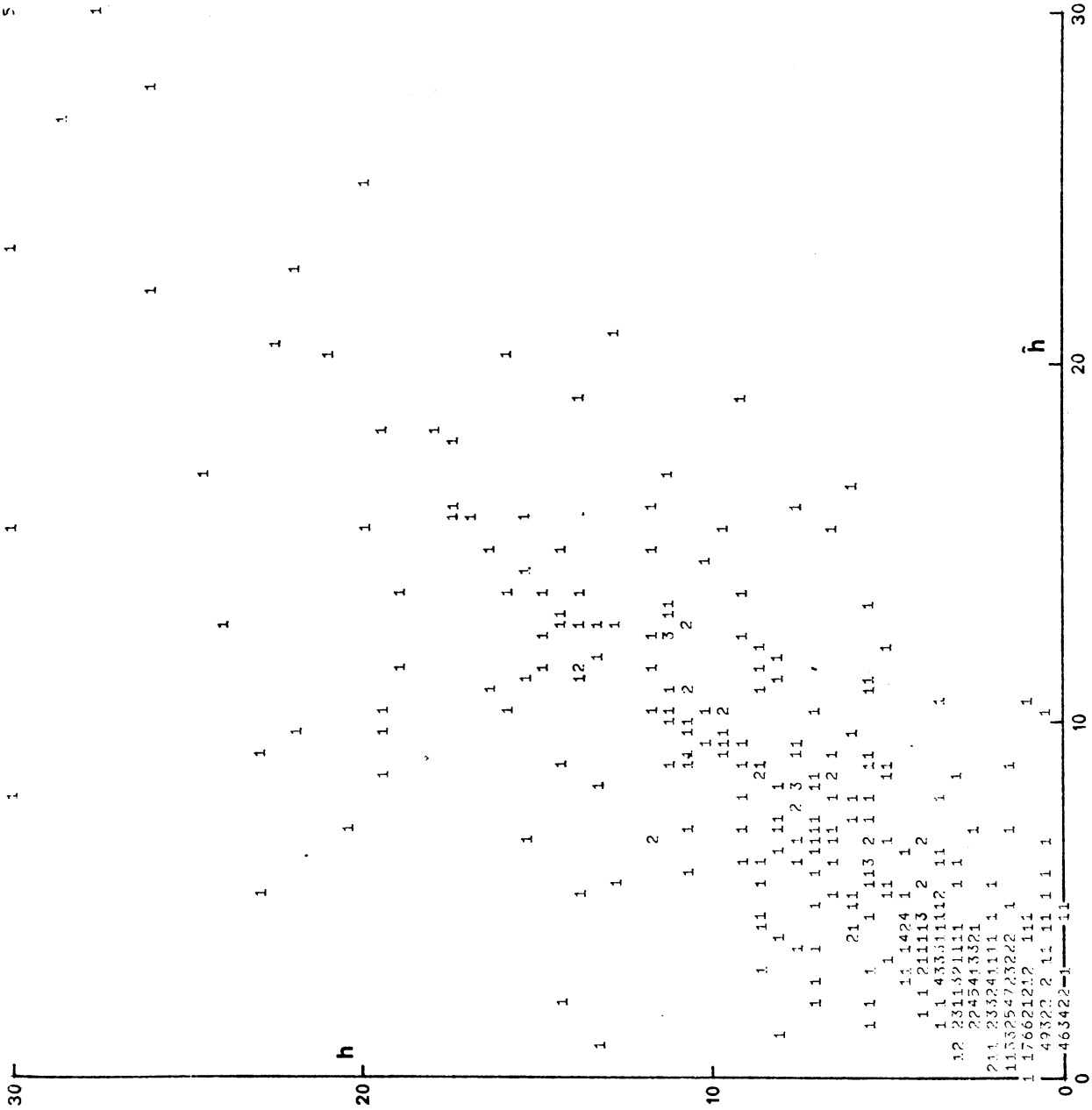
De stations 17 en 22 werden hierbij vooraf
weggelaten.

Voor opt. 1 en opt. 2 zie volgende paragraaf.

Voor Functie 1, 2 en 3 zie paragraaf 3.5.



Figuur 3.3.2 Verband tussen de gemeten neerslag van



Figuur 3.3.3 Verband tussen de gemeten neerslag van station 13 in de zomer en een schatting daarvan uit metingen van zeven stations.

3.4. Optimale lineaire schatters.

3.4.1. Inleiding.

Onder een optimale lineaire schatter wordt hier verstaan een schattingsmethode volgens (3.1.1), waarvan de gewichten g_k zo gekozen zijn, dat zo goed mogelijk voldaan is aan een bepaalde gestelde eis (hierbij is de eis $\sum g_k = 1$ weggelaten). Indien men de methode op zichzelf zou willen onderzoeken, zou het materiaal in tweeën gesplitst moeten worden om uit het ene deel de gewichten af te leiden en daarna met behulp van het andere deel de prestaties te onderzoeken. Men kan echter ook aannemen, dat de methode van de optimale schattingen de beste schatters geeft, daarbij buiten beschouwing latend of de berekende gemiddelde schattingsfout door steekproef-effecten misschien wat te gunstig uit valt. In dat geval kan deze gemiddelde schattingsfout beschouwd worden als de beste haalbare van alle mogelijke lineaire schatters. De optimale gewichten en de schattingsfout zijn daarom beide uit het gehele materiaal berekend en de laatste is als referentie voor de andere schattingsmethoden bedoeld.

Het aantal stations, waaruit geschat wordt, zou verder in feite nog als een variabele geïntroduceerd moeten worden. Om echter vergelijking onderling en met andere methoden mogelijk te maken, werd voor een vast aantal stations, nl. zeven gekozen. Bovendien werd ook van te voren gekozen, welke stations gebruikt zouden worden.

Om een goede indruk van de invloed van de netwerkdichtheid op de schattingen te krijgen, zijn de schattende stations telkens zoveel mogelijk op ongeveer dezelfde afstand van het te schatten station en bovendien op onderling ongeveer gelijke afstanden van elkaar gekozen. Alleen in de twee gevallen met de kleinste gemiddelde afstanden kon aan deze voorwaarde niet worden voldaan, omdat er slechts incidenteel enkele stations zeer dicht bij elkaar liggen.

Aangezien de variabiliteit van de neerslag sterk afhankelijk is van het neerslagtype en het voorkomen daarvan op haar beurt weer sterk afhankelijk is van het seizoen, zijn de schatters voor elk seizoen afzonderlijk bepaald.

3.4.2. De berekening van de optimale schatters.

De gewichten g_k van de optimale schatters zijn bepaald volgens de methode van de kleinste kwadraten. Daartoe is geëist, dat het kwadraat van de schattingsfout $(h_o - \hat{h}_o)^2$, gesommeerd over alle N neerslagsommen van een seizoen zo klein mogelijk is, dus

$$\sum_{n=1}^N (h_{n,o} - \sum_{k=1}^7 g_k h_{n,k})^2 = \min. \quad (3.4.2)$$

Hieruit volgt

$$\frac{\partial}{\partial g_k} \sum_{n=1}^N (h_{n,o} - \sum_{k=1}^7 g_k h_{n,k})^2 = 0 \quad (3.4.3)$$

Door differentiatie naar de zeven gewichten ontstaan de volgende zeven vergelijkingen, waaruit de zeven gewichten berekend kunnen worden.

$$\begin{pmatrix} \sum_{n=1}^N (h_{n,1} \cdot h_{n,1}), & \sum_{n=1}^N (h_{n,1} \cdot h_{n,2}), & \sum_{n=1}^N (h_{n,1} \cdot h_{n,7}) \\ \sum_{n=1}^N (h_{n,2} \cdot h_{n,1}), & \sum_{n=1}^N (h_{n,2} \cdot h_{n,2}), & \\ \sum_{n=1}^N (h_{n,3} \cdot h_{n,1}), & & \\ & & \\ & & \\ \sum_{n=1}^N (h_{n,7} \cdot h_{n,1}), & \sum_{n=1}^N (h_{n,7} \cdot h_{n,7}) & \end{pmatrix} \begin{pmatrix} G_1 \\ G_2 \\ G_3 \\ \\ \\ G_7 \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^N (h_{n,o} \cdot h_{n,1}) \\ \sum_{n=1}^N (h_{n,o} \cdot h_{n,2}) \\ \\ \\ \sum_{n=1}^N (h_{n,o} \cdot h_{n,7}) \end{pmatrix} \quad (3.4.4)$$

Bij de bepaling van de gewichten is geen rekening gehouden met het feit, dat grote dagsommen sterker kunnen spreiden dan kleinere, zoals in par. 2.2 reeds is vermeld. Daarom zijn ook gewichten bepaald, waarbij geëist is:

$$\sum_{n=1}^N \frac{(h_{n,o} - \hat{h}_{n,o})^2}{1/7 \sum_{k=1}^7 h_{n,k}} = \min. \quad (3.4.5)$$

Analoog aan het voorgaande ontstaan hier weer zeven vergelijkingen, waaruit de gewichten bepaald kunnen worden.

$$\begin{pmatrix} \sum_{n=1}^N (h_{n,1} \cdot h_{n,1}) \bar{h}_n^{-1}, \dots, \sum_{n=1}^N (h_{n,1} \cdot h_{n,7}) \bar{h}_n^{-1} \\ \sum_{n=1}^N (h_{n,2} \cdot h_{n,1}) \bar{h}_n^{-1} \\ \vdots \\ \sum_{n=1}^N (h_{n,7} \cdot h_{n,1}) \bar{h}_n^{-1}, \dots, \sum_{n=1}^N (h_{n,7} \cdot h_{n,7}) \bar{h}_n^{-1} \end{pmatrix} \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_7 \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^N (h_{n,0} \cdot h_{n,1}) \bar{h}_n^{-1} \\ \vdots \\ \sum_{n=1}^N (h_{n,0} \cdot h_{n,7}) \bar{h}_n^{-1} \end{pmatrix}$$

(3.4.6)

In het vervolg zullen deze beide methoden aangeduid worden als optimale schatters met constante (opt. 1) resp. toenemende schattingsfout (opt.2).

De hier beschreven bewerkingen vertonen sterke overeenkomst met de multiële regressieberekening, echter met dit verschil, dat hier de waarden van h vooraf niet verminderd zijn met hun gemiddelden. Hiermee is bereikt, dat, als er op geen der schattende stations neerslag is gemeten, de geschatte neerslag op het referentiestation ook gelijk nul is.

3.4.4. Uitkomsten van de optimale schatters.

In tabel 3.4.1. zijn een aantal groepen van stations gegeven, waarvoor optimale schatters zijn bepaald. De gewichten van de dichtstbijzijnde stations zijn bij de methode met constante schattingsfout vaak iets hoger dan die van de andere methode en de gewichten van ver afgelegen stations hebben wat vaker een, weliswaar kleine, negatieve waarde, maar over het algemeen zijn de verschillen tussen de gewichten van beide methoden niet groot.

Zoals reeds eerder is vermeld, konden bij twee bewerkingen met de kleinste gemiddelde afstanden de stations niet zo gekozen worden, dat alle afstanden tussen de schattende stations en het te schatten station ongeveer gelijk zijn.

Hierdoor was het mogelijk, het verband tussen de afstand van de schattende stations tot het referentie-station en de gewichten nader te onderzoeken. Dit is alleen uitgevoerd voor de zomer en de winter. Uit de figuren 3.4.1 en 3.4.2 blijkt, dat er een lineair verband bestaat tussen de gewichten en de inverse van de afstand tussen de stations. In de zomer is de curve in het algemeen wat steiler dan in de winter, wat veroorzaakt wordt door de grotere ruimtelijke variabiliteit van de neerslag in de zomer. Ook hier zijn er tussen de beide methoden geen grote verschillen te vinden.

In de tabellen 3.4.2 en 3.4.3 zijn de gemiddelde absolute schattingsfouten voor een aantal neerslagklassen van enkele schatters gegeven. Bovendien is daarin vermeld de totale gemiddelde absolute schattingsfout en het gewogen gemiddelde van de afstand. In figuur 3.4.3 zijn deze beide nog eens tegen elkaar uitgezet. Hieruit blijkt dat de gemiddelde absolute schattingsfout althans bij kleinere afstanden tussen de stations toeneemt met de gemiddelde afstand. Bij grote gemiddelde afstand is het verband met de schattingsfout nauwelijks aantoonbaar.

Tabel 3.4.1. Overzicht van referentiestations en schattende stations.

schattend station	refentiestations						
	13	6	18	10	30	13	10
1	15	2	10	7	7	10	2
2	16	3	11	11	10	11	5
3	17	4	13	15	11	14	8
4	18	5	15	18	23	15	20
5	22	7	17	27	31	16	25
6	24	19	22	29	33	20	31
7	28	23	27	30	35	25	33

Tabel 3.4.2. Gem. absolute schattingsfouten van schatters met constante fout.

		<u>Lente</u>						
Ref.st.		13	6	18	10	30	13	10
kl.gr.								
0.0 mm		.038	.040	.067	.053	.103	.088	.090
< 0.5		.16	.26	.27	.22	.32	.30	.32
0.5 - 1.0		.35	.36	.43	.34	.64	.62	.46
1.0 - 2.0		.53	.56	.76	.80	.74	.68	1.03
2.0 - 4.0		.68	.63	.97	.75	1.00	.82	.94
4.0 - 6.0		.78	1.21	1.31	.94	1.35	1.12	1.13
6.0 -10.0		1.17	1.13	1.69	1.48	1.54	1.57	1.63
10.0-20.0		1.77	1.46	2.61	2.56	2.49	2.28	2.82
Totaal		.35	.36	.53	.43	.54	.51	.55
gem. afstand		4.5	8.0	10.0	11.8	17.0	20.0	30.4

		<u>Zomer</u>						
Ref.st.		13	6	18	10	30	13	10
kl.gr.								
0.0 mm		.048	.061	.121	.075	.126	.133	.148
< 0.5		.29	.57	.31	.61	.52	.63	.74
0.5 - 1.0		.43	.70	.64	.84	1.09	.88	1.02
1.0 - 2.0		.81	.67	.81	1.08	0.93	1.15	1.40
2.0 - 4.0		.87	1.07	1.10	1.09	1.49	1.32	1.27
4.0 - 6.0		1.27	1.50	1.92	1.43	1.70	1.87	1.92
6.0 -10.0		1.56	1.73	2.20	2.05	2.46	2.86	2.61
10.0-20.0		2.78	2.68	3.09	3.43	2.88	3.67	4.19
Totaal		.65	.72	.86	.88	.95	1.11	1.13
gem. afstand		5.5	7.0	12.3	14.6	16.6	20.0	31.3

Vervolg tabel 3.4.2.

Herfst

Ref.st.	13	6	18	10	30	13	10
kl.gr.							
0.0 mm	.039	.048	.081	.038	.127	.085	.083
< 0.5	.15	.22	.27	.22	.29	.26	.31
0.5 - 1.0	.37	.60	.51	.60	.72	.81	.65
1.0 - 2.0	.59	.83	.64	.72	.87	.90	.92
2.0 - 4.0	.75	1.00	.97	.80	1.03	1.12	1.20
4.0 - 6.0	1.07	1.06	1.59	1.25	1.32	1.85	1.47
6.0 -10.0	1.11	1.26	1.74	1.72	1.67	1.87	1.90
10.0-20.0	1.90	2.24	2.05	2.40	2.61	3.28	2.81
Totaal	.51	.60	.66	.63	.77	.86	.80
gem. afstand	3.5	6.6	11.8	13.3	18.3	20.0	30.0

Winter

Ref.st.	13	6	18	10	30	13	10
kl.gr.							
0.0 mm	.053	.052	.098	.049	.116	.085	.063
< 0.5	.20	.15	.25	.16	.27	.25	.23
0.5 - 1.0	.35	.39	.41	.28	.53	.48	.45
1.0 - 2.0	.50	.46	.50	.41	.67	.68	.57
2.0 - 4.0	.59	.79	.80	.54	1.10	.77	.75
4.0 - 6.0	.87	.89	1.06	.84	1.11	1.15	.97
6.0 -10.0	.98	.94	1.15	1.04	1.43	1.47	1.16
10.0-20.0	1.60	1.65	2.09	1.54	1.94	2.18	2.17
Totaal	.42	.44	.53	.40	.63	.56	.52
gem. afstand.	6.6	7.9	11.9	13.8	15.3	16.7	28.2

Tabel 3.4.3. Absolute schattingsfouten van schatters met toenemende fout.

Lente

Ref.st. kl.gr.								
		13	6	18	10	30	13	10
0.0 mm		0.038	.037	.065	.053	.098	.090	.092
< 0.5		.15	.24	.24	.22	.30	.29	.32
0.5 - 1.0		.33	.36	.36	.34	.58	.63	.47
1.0 - 2.0		.51	.54	.64	.82	.72	.65	1.02
2.0 - 4.0		.64	.59	.91	.75	.95	.81	.96
4.0 - 6.0		.77	1.13	1.15	1.01	1.28	1.10	1.09
6.0 -10.0		1.06	1.12	1.54	1.48	1.63	1.66	1.64
10.0-20.0		1.85	1.52	2.43	2.55	2.54	2.33	2.91
Totaal		.34	.35	.49	.44	.54	.51	.55
gem. afstand		4.2	7.8	11.3	11.8	16.4	19.6	30.0

Zomer

Ref.st. kl.gr.								
		13	6	18	10	30	13	10
0.0 mm		.046	.061	.117	.074	.123	.130	.151
< 0.5		.28	.55	.32	.60	.51	.59	.77
0.5 - 1.0		.39	.70	.60	.91	1.03	.84	1.05
1.0 - 2.0		.77	.67	.78	1.10	.95	1.11	1.46
2.0 - 4.0		.82	1.03	1.08	1.11	1.46	1.31	1.30
4.0 - 6.0		1.15	1.48	1.86	1.52	1.61	1.83	2.05
6.0 -10.0		1.49	1.71	2.17	2.13	2.39	2.77	2.68
10.0-20.0		2.77	2.63	2.98	3.44	2.99	3.64	3.90
Totaal		.64	.71	.86	.90	.95	1.11	1.14
gem. afstand		5.1	7.6	12.7	14.8	16.1	20.2	30.5

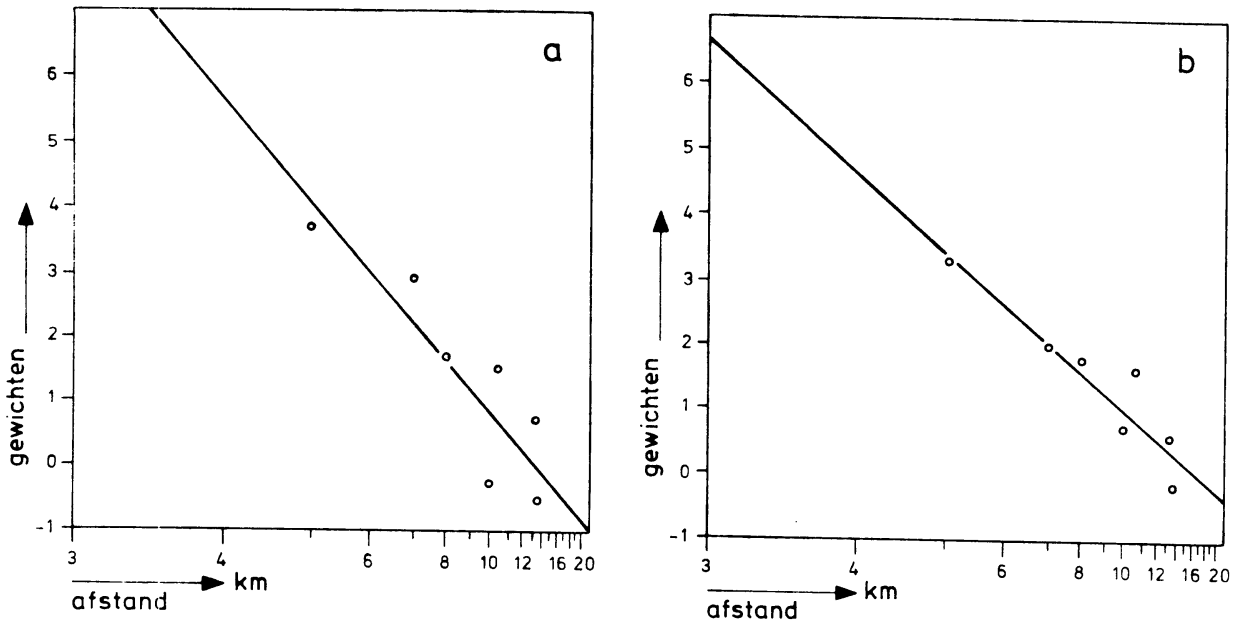
Vervolg tabel 3.4.3.

Herfst

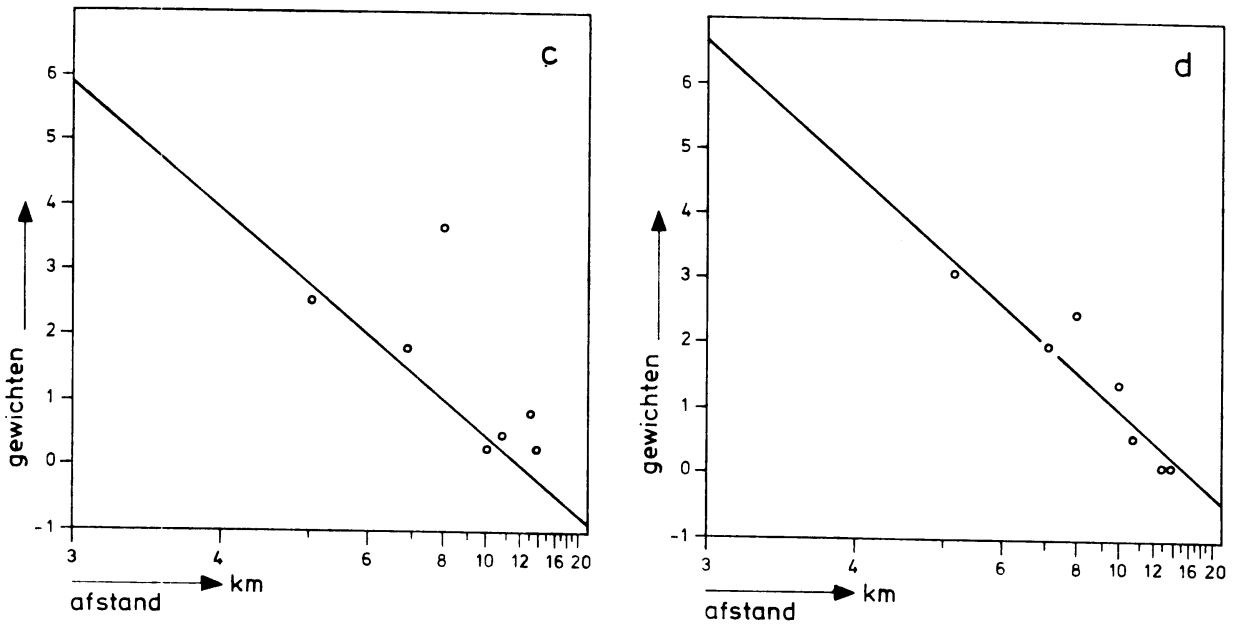
Ref.st.	13	6	18	10	30	13	10
kl.gr.							
0.0 mm	.040	.050	.079	.041	.126	.086	.076
< 0.5	.15	.19	.26	.20	.27	.25	.30
0.5 - 1.0	.35	.55	.48	.52	.64	.77	.66
1.0 - 2.0	.56	.75	.63	.69	.83	.86	.89
2.0 - 4.0	.71	.90	.95	.80	.98	1.03	1.20
4.0 - 6.0	1.01	.95	1.62	1.21	1.28	1.75	1.47
6.0 -10.0	1.09	1.18	1.69	1.70	1.74	1.84	1.85
10.0-20.0	1.78	2.20	1.99	2.44	2.48	3.13	2.90
Totaal	.50	.58	.65	.63	.76	.34	.80
gem. afstand	4.1	7.2	11.7	14.5	15.3	19.6	28.5

Winter

Ref.st.	13	6	18	10	30	13	10
kl.gr.							
0.0 mm	.051	.050	.101	.047	.113	.081	.064
< 0.5	.19	.15	.24	.15	.26	.24	.23
0.5 - 1.0	.33	.37	.40	.27	.51	.48	.44
1.0 - 2.0	.48	.44	.47	.38	.65	.65	.54
2.0 - 4.0	.57	.75	.77	.53	1.06	.72	.75
4.0 - 6.0	.85	.89	1.04	.86	1.04	1.09	.98
6.0 -10.0	1.06	.95	1.17	1.08	1.48	1.44	1.17
10.0-20.0	1.77	1.67	2.13	1.51	2.03	2.32	2.18
Totaal	.42	.44	.52	.39	.63	.56	.52
gem. afstand	6.7	7.4	12.7	13.1	15.7	19.0	28.1



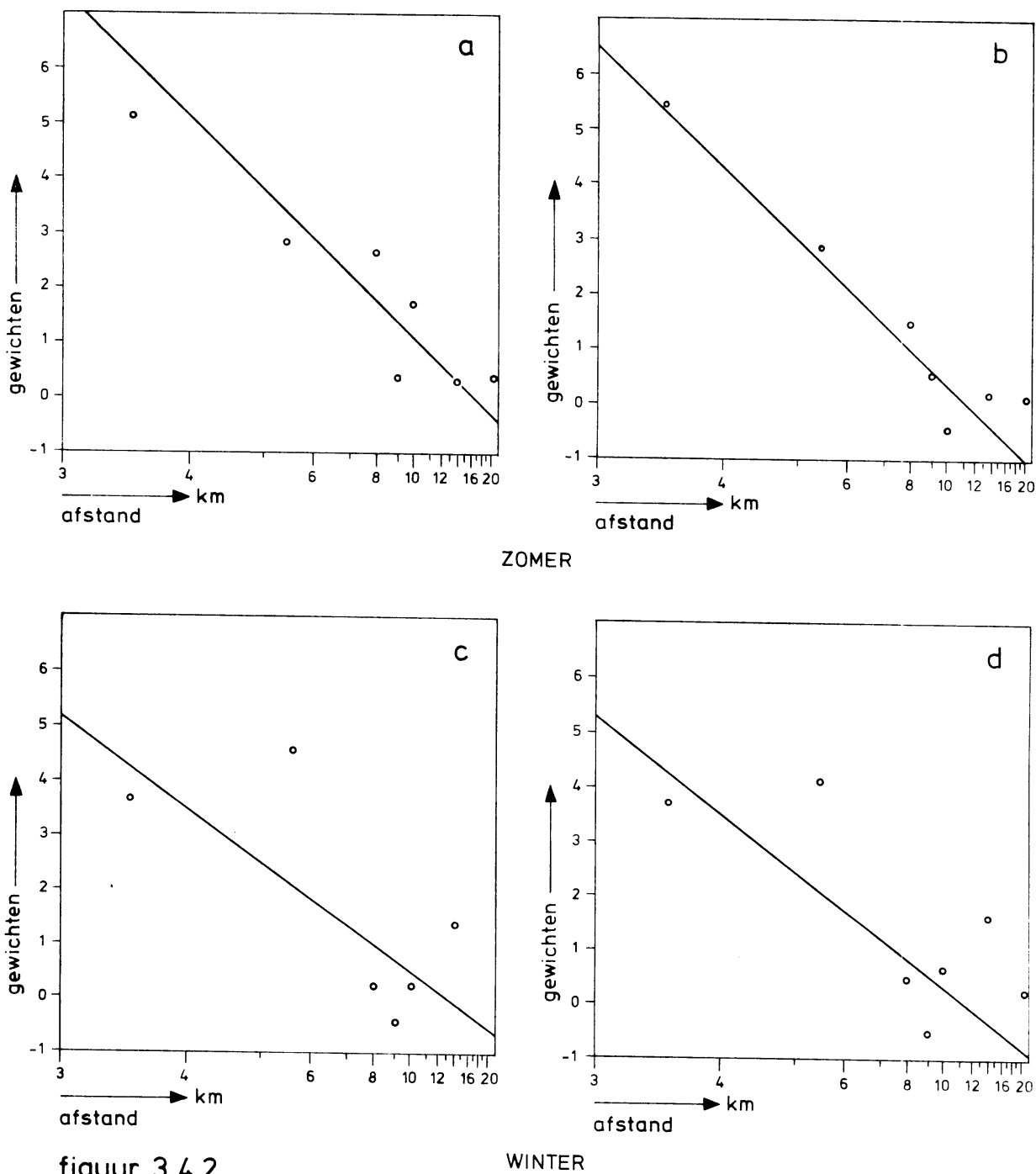
ZOMER



WINTER

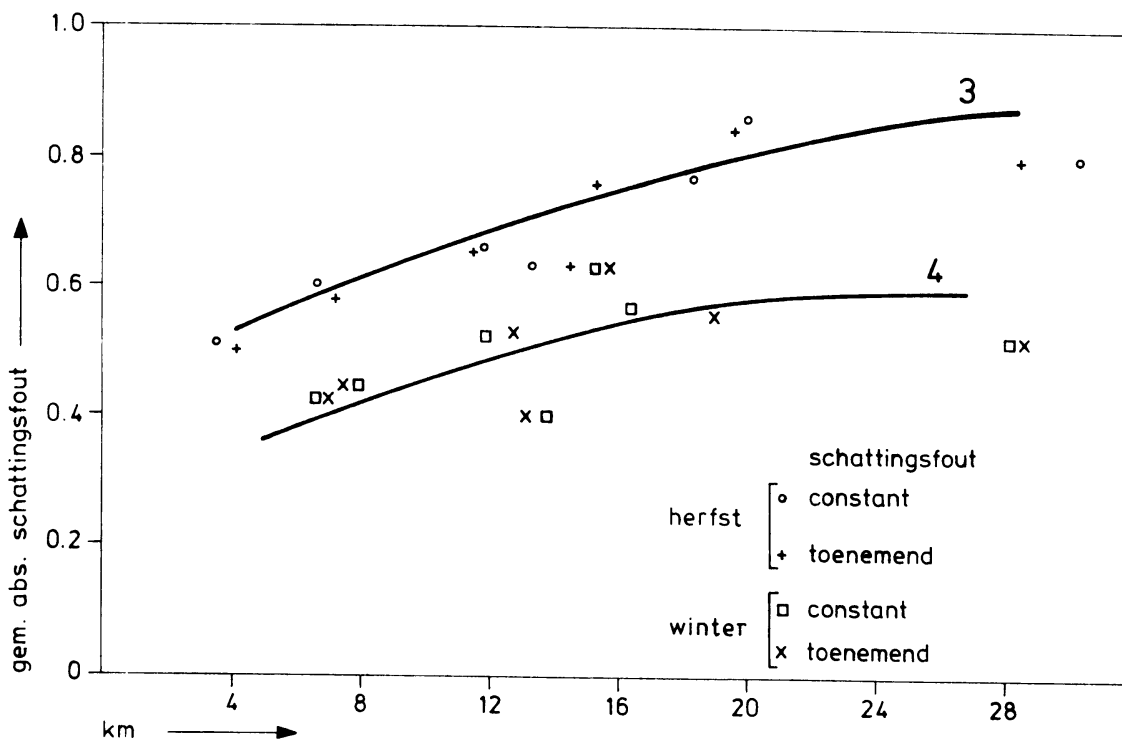
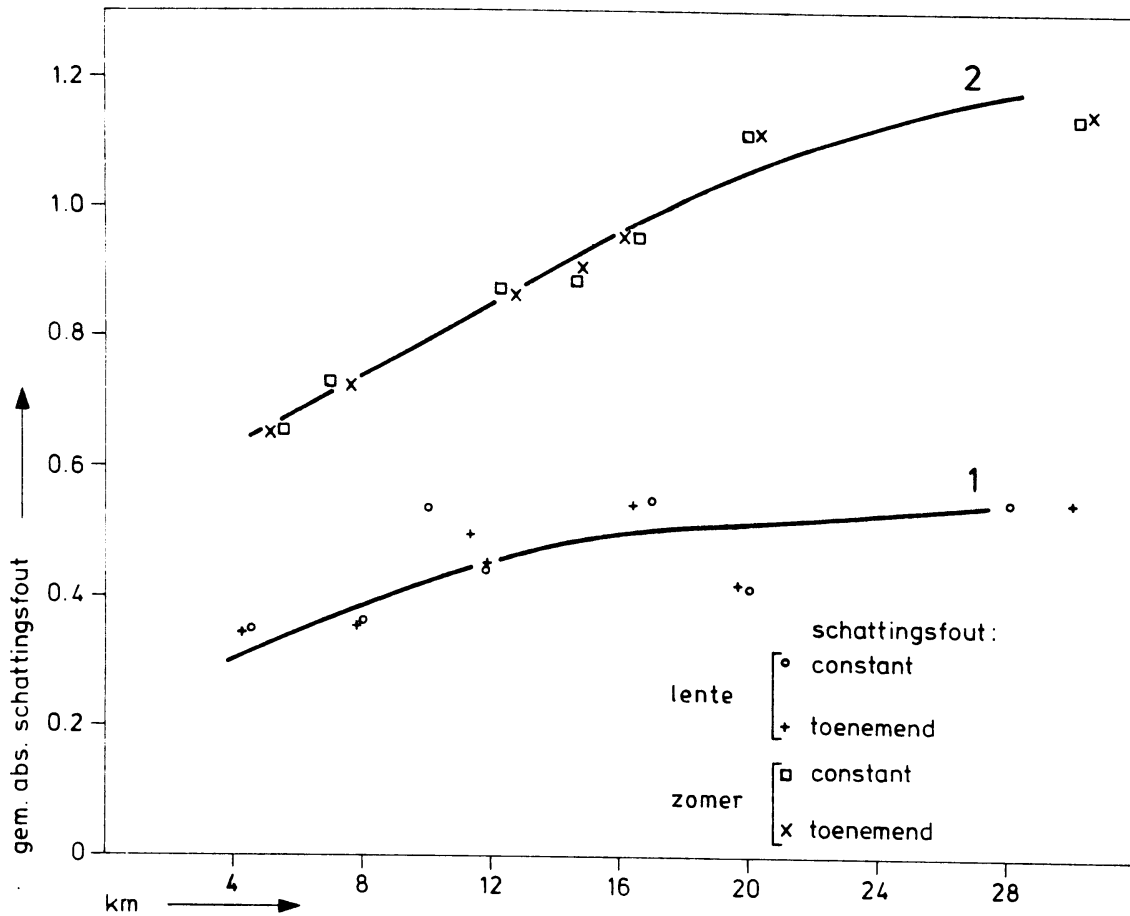
figuur 3.4.1

Verband tussen de afstand van de schattende stations tot het referentiestations 6 en de gewichten van de optimale schatters opt 1 (a en c) en opt 2 (b en d) in zomer en winter.



figuur 3.4.2

Verband tussen de afstand van de schattende stations tot het referentiestations 13 en de gewichten van de optimale schatters opt 1 (a en c) en opt 2 (b en d) in zomer en winter.



figuur 3.4.3

Verband tussen de gemiddelde afstand van een netwerk en de gemiddelde absolute schattingsfout per seizoen.

3.5. Schatting door aanpassing van polynomen.

3.5.1. Inleiding.

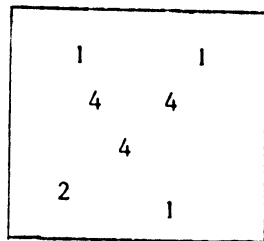
Stel we beschikken in een gebied R over K regengegevens h_k op de plaatsen r_k met coördinaten x_k, y_k .

We willen hier nu een functie van de vorm

$$f(x, y) = \sum_{n=1}^N a_n g_n(x, y) \tag{3.5.1}$$

op aanpassen, waarbij de functies $g_n(x, y)$ simpele polynomen in x en y voorstellen. Dus $g_1(x, y) = 1, g_2(x, y) = x, g_3(x, y) = y, g_4(x, y) = x^2$ enzovoorts.

Door De Bruin [5] is bewezen dat deze methode van interpolatie ook in de vorm van (3.1.1) kan worden geschreven. De methode onderscheidt zich dus alleen van de in vorige paragrafen behandelde methoden door de manier waarop de gewichten worden vastgesteld. Toch heeft de methode iets voor op de andere. Stel men heeft een neerslagpatroon



Figuur 3.5.1

zoals in figuur 3.5.1 is weergegeven. Intuïtief zal men ergens in het centrum een maximum willen aangeven, dat hoger is dan 4. Bij de voorgaande methoden echter werkt men in het algemeen met positieve gewichten (tussen 0 en 1) zodat als interpolatie-waarden altijd een waarde tussen 1 en 4 gevonden zal

worden. Passen we echter op dit patroon een tweedegraads functie aan dan zal waarschijnlijk ergens tussen de drie viereen een maximum gevonden worden.

Een nadeel van deze methode van schatten is dat ook negatieve schattingen mogelijk zijn. Bij het hiernavolgende onderzoek is dit verholpen door negatieve schattingen te vervangen door nullen. Verder is de keuze van de soort polynomen of functies nog een probleem. Welke functies de beste prestaties leveren is niet theoretisch vast te stellen doch kan alleen aan de hand van experimenten worden uitgezocht.

Over de verdere eigenschappen van de interpolatie-techniek is weinig bekend.

3.5.2. Theorie.

* De volgende afleiding is een variant van de afleiding gegeven door De Bruin (5). Zie ook Edwards (3).

Voor de schatting van de coëfficiënten a_n in (3.5.1) uit $K \geq N$ neerslaggegevens wordt verondersteld, dat de volgende uitdrukking

$$D^2 = \sum_{k=1}^K \left(h_k - \sum_{n=0}^N a_n g_n (r_k) \right)^2 \quad (3.5.2)$$

minimaal moet zijn. De waarden van de coëfficiënten a_n worden gevonden door de volgende N vergelijkingen op te lossen.

$$\frac{\partial}{\partial a_n} D^2 = -2 \sum_{k=1}^K \left[\left(h_k - \sum_{n'=0}^N a_{n'} g_{n'} (r_k) \right) g_n (r_k) \right] = 0 \quad (3.5.3)$$

voor $n = 1, 2, \dots, N,$

Verder worden de volgende vectoren en matrices geïntroduceerd:

- \vec{h} : is de kolomvector (h_1, h_2, \dots, h_k) ;
- \vec{g}_k : zijn de rij vectoren $(g_1(x_k, y_k), g_2(x_k, y_k), \dots, g_N(x_k, y_k))$;
- G : is een matrix met N kolommen en K rijen met in de rijen de vectoren \vec{g}_n .
- \vec{a} : is een kolomvector $(a_1, a_2, a_3, \dots, a_N)$

dan kan (3.5.3) ook geschreven worden als

$$(\vec{h} - G\vec{a})^T \otimes G = 0 \quad (3.5.4)$$

(\otimes Matrix produkt)

waarbij we de factor (-2) hebben weggelaten.

(T : betekent de gespiegelde). Uitgewerkt levert dit

$$\vec{a}^T \otimes G^T \otimes G = \vec{h}^T \otimes G$$

$$\text{of } \vec{a}^T = \vec{h}^T \otimes G \otimes (G^T \otimes G)^{-1} \quad (3.5.4)$$

$$\text{of } \vec{a} = ((G^T \otimes G)^{-1})^T \otimes G^T \otimes \vec{h}$$

Stel nu dat men wil interpoleren op de plaats (x_0, y_0) . Daartoe wordt de daarbij behorende rij vector $\vec{g}_0 = (g_0(x_0, y_0), g_1(x_0, y_0), \dots, g_N(x_0, y_0))$ geïntroduceerd. Dan kan de waarde van de functie $f(x_0, y_0)$ berekend worden uit

$$\hat{h} = f(x_0, y_0) = \vec{g}_0 \otimes \vec{a} \quad (3.5.5)$$

of met (3.5.4)

$$h = \vec{g}_0 \otimes (G^T \otimes G)^{-1})^T \otimes G^T \otimes \vec{h} \quad (3.5.6)$$

Definiëren we nu de K dimensionale rij vector

$$\vec{g} = \vec{g}_0 \otimes ((G^T \otimes G)^{-1})^T \otimes G^T \quad (3.5.7)$$

dan wordt (3.5.7)

$$\hat{h} = \vec{g} \otimes \vec{h} \quad (3.5.8).$$

Deze uitdrukking is mathematisch equivalent met de uitdrukking (3.1.1). De vector \vec{g} representeert de vector gewichten. Uit formule (3.5.7) blijkt bovendien dat deze vector niet afhankelijk is van de neerslagsommen doch alleen van de keuze van de polynomen.

3.5.3. Experiment.

De prestaties van deze methode werden op dezelfde manier onderzocht als de voorgaande methoden. Als referentie station werd in het eerste experiment station 4 gebruikt. De waarde op dat station werd geschat uit de gegevens van de stations 2, 3, 5, 6, 7, 10, 12, 19, 23, 27 en 29. Als functieset werd gekozen $g_1(x, y) = 1$, $g_2(x, y) = x$, $g_3(x, y) = y$, $g_4(x, y) = x^2$, $g_5(x, y) = x y$ en $g_6(x, y) = y^2$. Voor dit geval geldt dus $K = 11$ en $N = 6$. Voor de eenvoud werden bovendien de coördinaten x en y vastgelegd ten opzichte van station 4. Waardoor dit station automatisch de coördinaten $x = 0$ en $y = 0$ kreeg. De gewichten vector werd berekend met behulp van formule 3.5.7:

$$\vec{g} = (-.0578, +.1389, -.1432, +.1593, +.2943, -.0038, +.0732, +.2861, -.0223, +.0085, +.2670).$$

Bij het tweede experiment werd de functie aangepast aan de neerslag-sommen van de stations 6, 7, 11, 12, 15, 17, 18, 23, 30, 33 en 35. Op basis hiervan werd de neerslag van de stations 10 en 27 geschat. De gewichtsvector waren

$$\vec{g}(10) = (-.1088, +.2945, +.4012, -.0396, +.0193, -.1342, +.2758, +.1237, +.4067, -.1009, -.1376)$$

en

$$\vec{g}(27) = (+.0257, +.1441, -.0449, +.5581, +.4008, -.1769, +.1176, -.0310, +.0124, +.0763, -.0825).$$

Voor alle drie de gewichtensets geldt $\sum_1^{11} g_i = 1$. Brengen we de gewichten in kaart dan zien we een duidelijke afstandsafhankelijkheid zodat we niet veel verschil met voorgaande methoden verwachten. Het mogelijke voordeel ten opzichte van voorgaande methoden is dat ook een patroon afhankelijkheid is geïntroduceerd. De resultaten van de experimenten zijn samengevat in de tabel 3.5.1.

De resultaten zijn ongeveer gelijk aan die van de meerstationsschat-
ters met een eenvoudiger gekozen gewichten (zie tabel 3.5.1 resp. Functie 1, 2 en 3). Voor het verkrijgen van betere schattingen levert deze methode dus geen duidelijke voordelen.

Tabel 3.5.1. Resultaten aanpassing kwadratische functies.

Neerslag klasse	Lente			Zomer		
	4	10	27	4	10	27
0	.059	.063	.063	.060	.068	.080
.1 - .5	.25	.28	.26	.35	.58	.43
.5 - 1.0	.39	.38	.40	.55	.90	.75
1.0 - 2.0	.55	.96	.68	.90	1.12	1.08
2.0 - 4.0	.72	.85	.72	1.10	1.38	1.61
4.0 - 6.0	.89	1.35	.83	1.85	1.83	1.46
6.0 -10.0	1.00	1.71	1.17	1.75	2.67	2.02
10.0-20.0	1.68	2.64	2.08	2.88	3.63	3.24
Totaal	.41	.52	.42	.74	.97	.88

Neerslag klasse	Herfst			Winter		
	4	10	27	4	10	27
0	.046	.039	.047	.062	.047	.044
.1 - .5	.22	.20	.28	.18	.21	.20
.5 - 1.0	.42	.63	.56	.30	.34	.42
1.0 - 2.0	.81	.77	.95	.53	.51	.51
2.0 - 4.0	.87	.93	.91	.59	.71	.83
4.0 - 6.0	.97	1.37	1.69	.75	1.08	.95
6.0 -10.0	1.58	2.06	2.22	1.15	1.34	1.06
10.0-20.0	1.82	2.93	2.63	1.23	1.63	1.76
Totaal	.57	.73	.78	.43	.50	.49

De aantallen per neerslagklasse variëren van seizoen tot seizoen en van station tot station. De aantallen zijn van dezelfde grootte als in voorgaande tabellen.

3.6. Samenvatting.

Uit de voorgaande experimenten kunnen in het kort de volgende conclusies worden getrokken.

- a. De interpolatiefout hangt samen met de gemiddelde neerslaghoogte.
- b. De éénstationsschatter vertoont vooral in de zomer een sterke afstandsafhankelijkheid.
- c. De meerstationsschatter is altijd beter dan de éénstationsschatter, gebaseerd op het dichtstbijzijnde station van de groep.
- d. De keuze van de afstandsafhankelijkheid van de gewichten bij de meerstationsschatter is niet kritiek. Bij sterk variërende afstanden is een afstandsafhankelijkheid wel belangrijk.
- e. Schatters gebaseerd op zeven stations met gekozen gewichten zijn praktisch even goed als de schatters gebaseerd op optimale gewichten.
- f. Aanpassing van kwadratische functies aan neerslaghoogten levert even goede schatters als meerstationsschatters met gekozen gewichten. Een voordeel van deze methode is dat maximale waarden teruggevonden worden. Een nadeel van deze methode is dat het voorbereidende rekenwerk nogal moeilijk is.

4. Conclusie.

De doelstelling van dit rapport was om een aantal praktische gegevens bij elkaar te zetten die het mogelijk maken om een verantwoorde keuze te doen ten aanzien van de opbouw en de dichtheid van een regennetwerk. Het probleem is hierbij toegespitst op de vraag in hoeverre het mogelijk is een schatting te geven van de neerslag op plaatsen waar niet gemeten is bij een gegeven netwerk. Als nevenprobleem komt hierbij: "hoe moet men schatten".

In paragraaf 2 wordt aandacht besteed aan meer algemene statistische eigenschappen van de neerslag in het proefgebied. Het doel van deze paragraaf is, later vergelijkingen te kunnen maken met eventuele andere gebieden.

In paragraaf 3 wordt een aantal experimenten beschreven met het bestaande netwerk. Hierbij is zowel de schattingsmethode als de netwerkdichtheid gevarieerd. Uit deze experimenten is naar voren gekomen, dat een gewogen gemiddelde van een klein aantal omringende stations goed voldoet als schatter van de neerslag. De keuze van de gewichten is daarbij niet zo kritisch en het is zeker niet nodig om daarbij geoptimaliseerde gewichten te introduceren. Een simpele afstandsafhankelijkheid van de gewichten is voldoende.

Ten aanzien van de schattingsfout kan gesteld worden, dat de dichtheid van het net geen grote invloed heeft op de numerieke waarde van de fout. De visuele indruk geeft echter wel behoorlijke verschillen. Op basis van de hier gepresenteerde resultaten is het mogelijk een schatting te maken van de interpolatie-fout van een nieuw netwerk.

Hoe het regennetwerk er uit dient te zien kan niet op grond van dit rapport worden vastgelegd. Hierbij zijn nog te veel externe factoren van belang. De belangrijkste factor is wel van welke grootte de regenstructuren zijn die men nog wenst te detecteren. De afstanden tussen de stations dienen of kleiner of hier ongeveer aan gelijk te zijn. Geen enkele schattingsmethode is namelijk in staat om structuren weer te geven op kleinere schaal dan de gemiddelde stationsafstand. Als tweede factor komt dan de vraag hoe nauwkeurig men de structuur wenst te kunnen beschrijven.

Appendix A.

Interpolatie-technieken gebaseerd op lineaire regressie en minimale χ^2 -technieken zijn nogal populair in de literatuur (zie bijvoorbeeld Gandin). Uit het hierna volgende voorbeeld kan blijken dat men deze werkwijze niet altijd kritiekloos moet aanvaarden. Stel dat de waarde voor een gegeven grootte op twee plaatsen als volgt wordt gegenereerd,

$$h_1 = h (1 + \epsilon_1)$$

en $h_2 = h (1 + \epsilon_2)$

waarin h een exponentieel verdeelde grootte met $h \geq 0$ is en ϵ_1 en ϵ_2 twee normaal verdeelde grootheden zijn met gemiddelde nul ^{en een} (gelijke) variantie die veel kleiner dan 1 is. Verder veronderstellen we h , ϵ_1 en ϵ_2 onderling onafhankelijk.

We zien nu eenvoudig:

$$\overline{h_2} = \overline{h_1} = \overline{h (1 + \epsilon_1)} = \overline{h} + \overline{\epsilon_1 h} = \overline{h}$$

dit laatste wegens $\overline{\epsilon_1 h} \equiv 0$ (\overline{x} : x gemiddeld)

en verder

$$\overline{h_2^2} = \overline{h_1^2} = \overline{h^2 (1 + \epsilon_1)^2} = \overline{h^2} + \overline{h^2 \epsilon_1^2} = \overline{h^2} + \overline{h^2} \cdot \overline{\epsilon_1^2}$$

Stel nu dat men h_2 uit h_1 wil schatten. Met onze kennis van dit proces zeggen we intuïtief: Schat h_2 gelijk aan h_1 of

$$\hat{h}_2 = h_1 \quad (\hat{x} : \text{schatting van } x)$$

het is nu trivaal dat

$$\overline{\hat{h}_2} = \overline{h_2} \quad \text{en} \quad \overline{\hat{h}_2^2} = \overline{h_2^2}$$

dus deze schatter geeft behoud van de meest simpele statistieken van h_2 .

Verder geldt voor de variantie $\overline{\Delta^2}$ van de schatter ten opzichte van de werkelijke waarde

$$\overline{\Delta^2} = \overline{(\hat{h}_2 - h_2)^2} = \overline{(h(1 + \epsilon_1) - h(1 + \epsilon_2))^2} =$$

$$h^2 \cdot \overline{(\epsilon_1 - \epsilon_2)^2} = 2 h^2 \cdot \overline{\epsilon_1^2}$$

Stel dat men in plaats van deze schatter een lineaire regressie-schatter wil gebruiken met minimale variantie. Het ligt voor de hand om in dit geval een regressie te kiezen die door nul wordt geforceerd oftewel

$$\hat{h}_2 = a h_1$$

en de parameter a kiezen we nu zodanig dat

$$\overline{\Delta^2} = \overline{(\hat{h}_2 - h_2)^2} = \text{minimaal}$$

of
$$\overline{(a h_1 - h_2)^2} = \text{minimaal}$$

of
$$\frac{\partial}{\partial a} \overline{(a h_1 - h_2)^2} = 0$$

deze laatste vergelijking geeft als oplossing

$$a_{\text{opt}} = \overline{(h_1 \cdot h_2)} / \overline{h_1^2}$$

of
$$a_{\text{opt}} = \overline{h^2 (1 + \epsilon_1) (1 + \epsilon_2)} / (\overline{h^2} + \overline{h^2 \cdot \epsilon^2}) =$$

$$= \overline{h^2} / (\overline{h^2} + \overline{h^2 \cdot \epsilon^2}) = \frac{1}{1 + \overline{\epsilon^2}}$$

De variantie van de schatter ten opzichte van de werkelijke waarde wordt in dit geval

$$\Delta^2_{opt} = (1 - a^2_{opt}) \cdot (1 + \epsilon^2) \cdot \overline{h^2}$$

Stel nu bijvoorbeeld $\overline{h^2} = 1$, $\epsilon^2 = 0,1$ dan vindt men voor de fout bij de eerste schatter

$$\Delta^2 = 2 \times 1 \times 0,1 = 0,2$$

en voor de coëfficiënt a bij de tweede schatter

$$a = 1 / (1 + 0,1) = 0,91$$

en voor de fout in de tweede schatter

$$\Delta^2_{opt} = 0,19.$$

In dit geval dus een zeer kleine verbetering wat betreft de fout in de schatting. Echter voor de tweede schatting geldt eveneens

$$\hat{h} = a \cdot h_1 = a \cdot h_1 = 0,9 \cdot h_2$$

en
$$\overline{\hat{h}^2} = a^2 \cdot \overline{h_1^2} = 0,81 \cdot \overline{h_2^2}$$

Zodat we als nadeel zouden moeten accepteren dat we systematisch te laag schatten. In veel gevallen is dit niet acceptabel. Het verdient dan ook aanbeveling om de op deze manier verkregen schatters op deze eigenschappen te onderzoeken en niet zonder meer als de best mogelijke te accepteren.

Appendix B.

In dit rapport is niet ingegaan op de vraag "Hoe goed kan men het gebiedsgemiddelde van de neerslag schatten". Deze vraag kan op dezelfde manier aangepakt worden als in dit rapport gebeurd is ten aanzien van de schatting van de puntneerslag. Het proefnetwerk is echter voor dit soort experimenten niet zo geschikt daar het netwerk nogal onregelmatig is. Er zijn toch enkele experimenten op dit gebied uitgevoerd.

Uit de voorgaande experimenten kunnen ook enige conclusies worden getrokken met betrekking tot het gebiedsgemiddelde. Indien men aanneemt dat het gemiddelde van de zeven stations die gebruikt worden bij de schatting van de neerslag op één station representatief is voor het gebiedsgemiddelde rond het geschatte station dan kan men stellen dat de fout waarmee het station geschat wordt vermeld onder a [7] (zie par. 3.3) ongeveer gelijk is aan de fout waarmee door dit station het gebiedsgemiddelde wordt geschat. De representatieve gebiedsgrootte kan in dit geval dan ongeveer gesteld worden op het kwadraat van de gemiddelde afstand. Verder kan uit a [34] (tabel 3.3.1) de fout, waarmee station 13 het gemiddelde over het gehele proefgebied (ong. 4000 km²) schat, worden afgeleid.

Er zijn ook enkele meer op gebiedsgemiddelde gerichte experimenten uitgevoerd. Bij de eerste set experimenten werd aangenomen dat het gemiddelde van de stations 2, 3, 4, 5, 6, 7, 19, 23 en 29 representatief is voor het gebiedsgemiddelde van een gebied van 400 km² rond deze stations. Dit gebiedsgemiddelde werd nu achtereenvolgens geschat met één stations (6 [I]), met drie stations (2, 23 en 29 [II]) en met vier stations (1, 6, 27 en 30 [III]). De gewichtenkeuze was in deze gevallen respectievelijk (1), (.33, .33, .33) en (.25, .25, .25, .25). De resultaten zijn samengevat in de tabel B.1. De indeling naar neerslagklassen geschiedde nu op grond van het gebiedsgemiddelde.

Bij het laatste experiment werd het gemiddelde van alle stations geschat uit de neerslag op de stations 3, 24, 26 en 30 met gelijke gewichten. De resultaten van dit experiment zijn samengevat in tabel B.2. Uit deze tabellen blijkt dat ook bij het gebiedsgemiddelde de schattingsfout afhankelijk is van seizoen en neerslaghoogte. Uit tabel B.1 (Δ II) blijkt verder dat het zeer goed mogelijk is om met drie stations het gebiedsgemiddelde van een gebied van 400 km^2 te schatten. De cijfers hier gegeven zijn echter wat optimistisch daar dezelfde drie stations ook gebruikt zijn bij het bepalen van het werkelijke gebiedsgemiddelde. De situatie van Δ III is wat eerlijker en geeft ook nog zeer goede resultaten.

Uit tabel B.2 blijkt dat met vier stations een zeer redelijke schatting van het gebiedsgemiddelde over het totale gebied mogelijk is. Opvallend is wel dat deze vier stations betere prestaties leveren voor de schatting van het totale gebiedsgemiddelde dan de vier stations onder Δ III voor het kleine gebied.

De resultaten van deze studie moeten slechts als een indicatie worden gezien. Meer gefundeerde cijfers vereisen een groter aantal experimenten, bovendien is het mogelijk dat het invoeren van Thiessen gewichten nog een duidelijke verbetering geeft.

Tabel B.1 Resultaten schatting gebiedsgemiddelde.
Gebiedsgrootte 400 km².

Δ : Gemiddelde absolute schattingsfout in mm.

Neerslag klasse	Lente				Zomer			
	N	Δ I	Δ II	Δ III	N	Δ I	Δ II	Δ III
0	326	0	0	.01	383	0	0	.01
.1 - .5	328	.10	.06	.12	226	.12	.08	.19
.5 - 1.0	88	.34	.21	.25	59	.48	.36	.41
1.0 - 2.0	101	.61	.37	.59	91	.76	.45	.57
2.0 - 4.0	114	.67	.41	.55	100	1.01	.76	.90
4.0 - 6.0	50	.95	.59	.81	72	1.43	.72	1.04
6.0 -10.0	55	1.01	.63	.60	88	1.84	1.39	1.60
10.0-20.0	38	1.43	1.02	1.35	71	2.75	1.67	2.14
Totaal	1104	.34	.21	.29	1104	.70	.45	.57

Neerslag klasse	Herfst				Winter			
	N	Δ I	Δ II	Δ III	N	Δ I	Δ II	Δ III
0	215	0	0	.01	136	0	0	.01
.1 - .5	370	.08	.05	.08	385	.09	.05	.09
.5 - 1.0	78	.32	.22	.30	81	.28	.18	.25
1.0 - 2.0	92	.67	.34	.51	122	.48	.23	.34
2.0 - 4.0	96	.72	.50	.70	137	.65	.38	.47
4.0 - 6.0	76	1.21	.79	.97	86	.77	.52	.56
6.0 -10.0	79	1.47	.78	.91	88	1.25	.61	.83
10.0-20.0	67	2.47	1.04	1.21	40	1.65	.82	1.21
Totaal	1092	.56	.31	.40	1083	.43	.24	.32

N is aantal in betreffende neerslagklasse.

Tabel B.2 Resultaten schatting gebiedsgemiddelde.

Gebiedsgrootte 4000 km², 4 stations.

Δ : Gemiddelde absolute schattingsfout in mm.

Neerslag klasse	Lente		Zomer		Herfst		Winter	
	N	Δ	N	Δ	N	Δ	N	Δ
0	170	0	220	0	93	0	51	0
.1 - .5	453	.06	359	.06	488	.06	467	.06
.5 - 1.0	102	.21	71	.34	75	.25	80	.25
1.0 - 2.0	103	.40	87	.46	85	.40	129	.33
2.0 - 4.0	125	.52	119	.70	119	.61	137	.39
4.0 - 6.0	51	.74	73	1.19	68	.80	82	.47
6.0 -10.0	68	.72	86	1.26	82	1.00	86	.60
10.0-20.0	29	.82	78	1.93	65	1.68	40	1.03
Totaal	1104	.25	1104	.49	1092	.40	1083	.27

