

10 AUG. 1966

Verslagen V-181

KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT

Smoothing and Filtering

door

Ir. Th. de Crook

De Bilt, februari 1966

Kon. Ned. Meteor. Inst.
De Bilt

Smoothing and Filtering

Inleiding.

Bij numerieke prognoses e.d. zijn de waarden van een atmosferische grootheid gegeven in de roosterpunten, terwijl de funktie $f(x, y)$, die deze grootheid weergeeft, vaak tamelijk grillig is. Het is de bedoeling dit min of meer grillige karakter te vervangen door een met de praktijk overeenstemmende verdeling, terwijl toch de essentiële hoofdtrekken aanwezig blijven.

Er zijn reeds vele methoden ontwikkeld om te trachten dit te bereiken. Deze berusten als regel op het toepassen, in de een of andere vorm, van een vereffening op de oorspronkelijke resultaten. Bij de meeste methoden echter gaan bij deze vereffening ook details verloren, die men op meteorologische gronden liever zou handhaven.

In "a technique for maximizing details in numerical weather Map analysis" door L.S. Barnes (Journ. of Appl. Met. 3 (1964) 396-409) wordt een methode behandeld, die het voordeel heeft dat details die verloren zijn gegaan door af te vlakken worden teruggewonnen. Bovendien heeft deze methode het voordeel dat er geen faseverschuivingen optreden.

De methode is gebaseerd op het feit dat een verdeling van een atmosferische grootheid $f(x, y)$ op elk tijdstip kan worden voorgesteld door de som van een oneindig aantal onafhankelijke harmonische golven (d.w.z. een oneindige Fourierreeks).

In dit verslag is een wijziging van de methode van Barnes beschreven, die het mogelijk maakt meer af te vlakken, ~~zonder dat evenwel meer details verloren gaan.~~

Analytische gedeelte.

We definiëren een afvlakfunctie $g(x, y)$:

$$g(x, y) = \int_0^{2\pi} \int_0^{\infty} f(x + r \cos \theta, y + r \sin \theta) w r dr d\theta \quad (1)$$

$$\text{waarbij de gewichtsfunktie } w = \frac{1}{\sqrt{\pi} k} e^{-r^2/4k} \quad (2)$$

Het punt (x, y) is de oorsprong en r en θ zijn poolcoördinaten.

De parameter k bepaalt de vorm van de gewichtsfunktie.

Substitutie van (2) in (1) geeft:

$$g(x, y) = \int_0^{2\pi} \int_0^{\infty} f(x + r \cos \theta, y + r \sin \theta) \frac{1}{\sqrt{\pi} k} e^{-r^2/4k} r dr d\theta \quad (3)$$

met de nieuwe gewichtsfunctie $\eta = e^{-r^2/4k}$ (4)

Het is gemakkelijk te verifiëren dat:

$$\int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} \eta d\left(\frac{r^2}{4k}\right) d\theta = 1$$

Hieruit volgt:

$$\int_0^R \eta d\left(\frac{r^2}{4k}\right) + \int_R^{\infty} \eta d\left(\frac{r^2}{4k}\right) = 1, \text{ met } 0 < R < \infty$$

We veronderstellen dat $\int_R^{\infty} \eta d\left(\frac{r^2}{4k}\right) = \varepsilon$ en vinden dan dat

$$\int_0^R e^{-r^2/4k} d\left(\frac{r^2}{4k}\right) = 1 - \varepsilon \Rightarrow e^{-R^2/4k} = \varepsilon \Rightarrow \frac{R^2}{4k} = -\ln \varepsilon \equiv E \quad (5)$$

Indien ε klein genoeg is kunnen we de invloed van de gewichtsfunctie in elk punt voldoende nauwkeurig bepalen. Bijvoorbeeld, $\varepsilon = 0.02$ dan is $E = 4$; dit betekent dat alle punten binnen een cirkel met straal R voor 98% de invloed van de gewichtsfunctie bepalen.

We noemen R de "radius of influence". De formule (5) zullen we later gebruiken om bij een gegeven verdeling de vorm van de gewichtsfunctie te bepalen, die afhankelijk van de parameter k is.

In de praktijk kunnen we (3) niet gebruiken omdat we de analytische vorm van $f(x,y)$ niet kennen, we weten slechts de functiewaarden in de roosterpunten. Daarom moeten we $g(x,y)$ benaderen door een eindige som.

$$g(x,y) = \frac{\sum_j^M \eta(r_j) f_j}{\sum_j^M \eta(r_j)} \quad (6)$$

met M het aantal punten binnen het invloedsgebied.

Voor (6) kunnen we schrijven:

$$g(x,y) = D f(x,y) \quad (7)$$

d.w.z. indien we een operator D toepassen op $f(x,y)$ verkrijgen we de afvlakfunctie $g(x,y)$. ($0 \leq D \leq 1$).

Rekenschema.

We passen de operator D toe op de gegeven functiewaarden van $f(x,y)$.

Dit geeft een eerste afvlakking $g_1(x,y) = D f(x,y)$

Vlak daarna $g_1(x,y)$ af, dit geeft $g_2(x,y) = D g_1(x,y)$

Ga zo door tot $g_n(x,y) = D g_{n-1}(x,y)$ (8)

Er is nu n keer afgevlakt, vervolgens gaan we p keer herstellen.

Dit gaat als volgt (Let op onderscheid in definitie van g_m voor $m \leq n$ en $m > n$):

$$g_{n+1}(x,y) = g_n(x,y) + D^n [f(x,y) - g_n(x,y)] \quad (9)$$

$$\begin{aligned} \varepsilon_{n+2}(x,y) &= \varepsilon_{n+1}(x,y) + D^n [f(x,y) - \varepsilon_{n+1}(x,y)] \\ \vdots \\ \varepsilon_{n+p}(x,y) &= \varepsilon_{n+p-1}(x,y) + D^n [f(x,y) - \varepsilon_{n+p-1}(x,y)] \end{aligned} \quad (10)$$

Convergentie van het rekenschema.

$$\varepsilon_n(x,y) = D\varepsilon_{n-1}(x,y) = D^2\varepsilon_{n-2}(x,y) \dots = D^n f(x,y) \quad (11)$$

Formule (9) is nu als volgt te schrijven:

$$\varepsilon_{n+1}(x,y) = D^n f(x,y) + D^n [f(x,y) - D^n f(x,y)]$$

Hieruit volgt:

$$\varepsilon_{n+1}(x,y) = D^n f(x,y) [1 + (1-D^n)] = D^n f(x,y) \sum_{i=0}^1 (1-D^n)^i \quad (12)$$

We kunnen gemakkelijk aantonen dat voor $\varepsilon_{n+p}(x,y)$ geldt:

$$\varepsilon_{n+p}(x,y) = D^n f(x,y) \sum_{i=0}^p (1-D^n)^i \quad (13)$$

$$\text{of: } \varepsilon_{n+p}(x,y) = D^n f(x,y) \quad \text{met } D' = D^n \sum_{i=0}^p (1-D^n)^i \quad (14)$$

Wat is $\lim_{p \rightarrow \infty} [f(x,y) - \varepsilon_{n+p}(x,y)] = ?$

Deze limiet bepalen we m.b.v. formule (13).

$$\begin{aligned} \lim_{p \rightarrow \infty} [f(x,y) - \varepsilon_{n+p}(x,y)] &= \lim_{p \rightarrow \infty} \left[f(x,y) - D^n f(x,y) \sum_{i=0}^p (1-D^n)^i \right] = \lim_{p \rightarrow \infty} \\ & f(x,y) \left[1 - D^n \sum_{i=0}^p (1-D^n)^i \right] = 0, \text{ omdat } \sum_{i=0}^p (1-D^n)^i \text{ een meetkundige reeks} \\ & \text{is, nl. } 0 \leq D^n \leq 1 \text{ dus ook } 0 \leq D^n \leq 1 \text{ en } \lim_{p \rightarrow \infty} \sum_{i=0}^p (1-D^n)^i = \frac{1}{1-(1-D^n)} = \frac{1}{D^n} \end{aligned}$$

Dit wil zeggen dat de oorspronkelijke functie langzamerhand weer te voorschijn komt en het rekenschema dus convergent is.

Een bijzonder geval van formule (14) is:

$$D' = D \sum_{i=0}^p (1-D)^i \quad \text{voor } n=1 \quad (15)$$

Deze formule (15) is door Barnes gebruikt in het in de inleiding genoemde artikel.

Grafieken.

Aangezien de verdeling $f(x,y)$ kan worden voorgesteld door een oneindige Fourierreeks, dus door een oneindig aantal onafhankelijke golven, zullen we in het nu volgende alleen één enkelvoudige harmonische golf beschouwen, terwijl we ons steeds dienen te realiseren dat het ook geldt voor elke andere golf van de functie $f(x,y)$.

Ter vereenvoudiging beschouwen we een eendimensionale golf nl.

$$f(x) = \sin ax \quad (15)$$

met $a = \pi/L$ en L de halve golflengte.

Substitutie van (15) in (1) geeft na veel rekenwerk (zie aanhangsel):
 $g(x) = D f(x)$ met $D = e^{-ka^2}$ (maat voor de amplitude) (16)

We weten dat $k = \frac{R^2}{4E}$ en $a = \frac{\pi}{L}$, dit gesubstituëerd in (16) geeft:

$$D = \exp\left[-\left(\frac{\pi^2}{4E}\right) \left(\frac{R}{L}\right)^2\right] \text{ of } D = \exp\left[-\left(\frac{\pi^2}{4E}\right) \left\{\left(\frac{R}{d}\right)/\left(\frac{L}{d}\right)\right\}^2\right] \quad (17)$$

waarbij d de roosterpuntsafstand is.

We bepalen nu D als functie van L/d en R/d , zie grafiek 1.

Uit deze grafiek is onmiddellijk af te lezen dat de golven met een grote golflengte praktisch niet worden afgevlakt, terwijl de golven met een kleine golflengte bijna geheel verdwijnen m.a.w. de kleine golven worden uitgefilterd, echter sterk afhankelijk van R/d .

Het is ook mogelijk deze grafiek met het rekenschema te bepalen.

Kies daartoe achtereenvolgens golven met verschillende golflengten en neem bijvoorbeeld $R/d = 1.6$. Omdat de roosterpuntsafstand d bekend is, kunnen we R berekenen. De parameter k bepalen we uit de formule $k = R^2/4E$, waarbij we $E = 4$ kiezen (d.w.z. ϵ klein).

We kennen nu de vorm van de gewichtsfunctie en daarmee de operator D . Pas deze operator D toe op $f(x)$ en bepaal ^{de} afgevlakte functie $g_1(x)$ en hiervan de amplitude.

In grafiek 2 zijn D en D' bepaald als functie van L/d met $R/d = 1.6$ en $E = 4$. Er is hier één keer afgevlakt en daarna hersteld. Na een oneindig aantal keren te herstellen verkrijgen we de oorspronkelijke functie terug.

Het herstellen gaat als volgt. Uit (13) volgt voor $n = 1$ dat $g_{1+p}(x) = D \times \sum_{i=0}^p (1-D)^i f(x)$ of: $g_{1+p}(x) = D' f(x)$ met $D' = D \sum_{i=0}^p (1-D)^i$

D is bekend, dus D' is gemakkelijk te berekenen.

Grafiek 2 kan ook met het rekenschema worden bepaald door verschillende golven te herstellen.

Door steeds meer te herstellen wordt de lijn (het filter) steeds steiler.

Na een keer herstellen wordt er het meeste hersteld in het middengebied met een verschuiving naar links (de kleinere golven) bij toenemend aantal keren herstellen.

Het is niet nodig vele keren te herstellen, reeds na enkele keren is het filter al aanzienlijk steiler geworden. Door nog verder te herstellen worden de kleinere golven meer hersteld, totdat tenslotte de oorspronkelijke

Een nader onderzoek bracht aan het licht dat de resultaten XXIV (resp. XXVI), zie tabel I, vergelijkbaar zijn met de resultaten V, waardoor de draaitijd tot ca 1 min. (resp. ca 2 min.) kon worden verminderd. In grafiek 4 zijn de bijbehorende filters getekend, die slechts een kleine verschuiving t.o.v. elkaar vertonen.

Met behulp van het plot-programma zijn stromingspatronen gemaakt, waarvan er hier slechts enkele zijn weergegeven, nl.:

- 1e $f(x,y)$, 24 hr. prognose (3) voor 12-12-1962.
- 2e $g(x,y)$, 24 hr. prognose (3) voor 12-12-1962, ns= 10 en nis= 10 (III)
- 3e $g(x,y)$, 24 hr. prognose (3) voor 12-12-1962, ns= 5 en nis= 10 (V)
- 4e $g(x,y)$, 24 hr. prognose (3) voor 12-12-1962, ns= 3 en nis= 4 (XXIV).

Het algolprogramma en de overige resultaten liggen ter inzage bij Th.de Crook.

funktie weer te voorschijn komt.

Door meerdere keren af te vlakken, zie grafiek 3, verschuift het filter naar rechts. Grafiek 3 is bepaald met de formule $g_n(x) = D^n f(x)$, maar kan ook bepaald worden met het rekenschema.

Grafiek 4 geeft $D' = D'(\frac{L}{d}, ns, nis)$ met $R/d = 1.6$ en $E = 4$ en is bepaald met de formules (11) en (14).

In de praktijk gaat het er om een filter te kiezen, dat voldoet aan de gestelde eisen bv. de kleine golven tot een zekere golflengte moeten verdwijnen, de grotere golven moeten zo goed mogelijk behouden worden, terwijl het tussengebied zo klein mogelijk moet blijven.

De bepaling van het filter hangt af van R/d en E (dus van k want $k = \frac{R^2}{4E}$) en bovendien van het aantal keren afvlakken (ns) en herstellen (nis).

Voor meerdere bijzonderheden verwijzen we naar het artikel van Barnes (zie inleiding).

Experimentele resultaten.

De funktiewaarden van een atmosferische grootheid $f(x,y)$, hier isohypsen in het 500 mb. vlak, zijn gegeven in een rooster van 17 bij 29 met een roosterpuntsafstand $d = 375$ km.

Voor de waarden k , ns en nis , die het filter bepalen, werden verschillende waarden genomen, waarbij k steeds bepaald werd door de formule $k = \frac{R^2}{4E}$ met $E = 4$.

De waarde $R/d = 1.6$ verdiende de voorkeur. Dit staat in verband met het aantal punten binnen het invloedsgebied (dit is het gebied binnen de cirkel met straal R). Dit bedroeg steeds negen punten.

De diverse berekeningen werden uitgevoerd op de TR4 te Delft.

Het algolprogramma is gebaseerd op het rekenschema (zie ook stroomdiagram), waaraan een plot-programma is toegevoegd.

De draaitijd van het programma is o.a. sterk afhankelijk van ns en nis .

ns = number of smoothing.

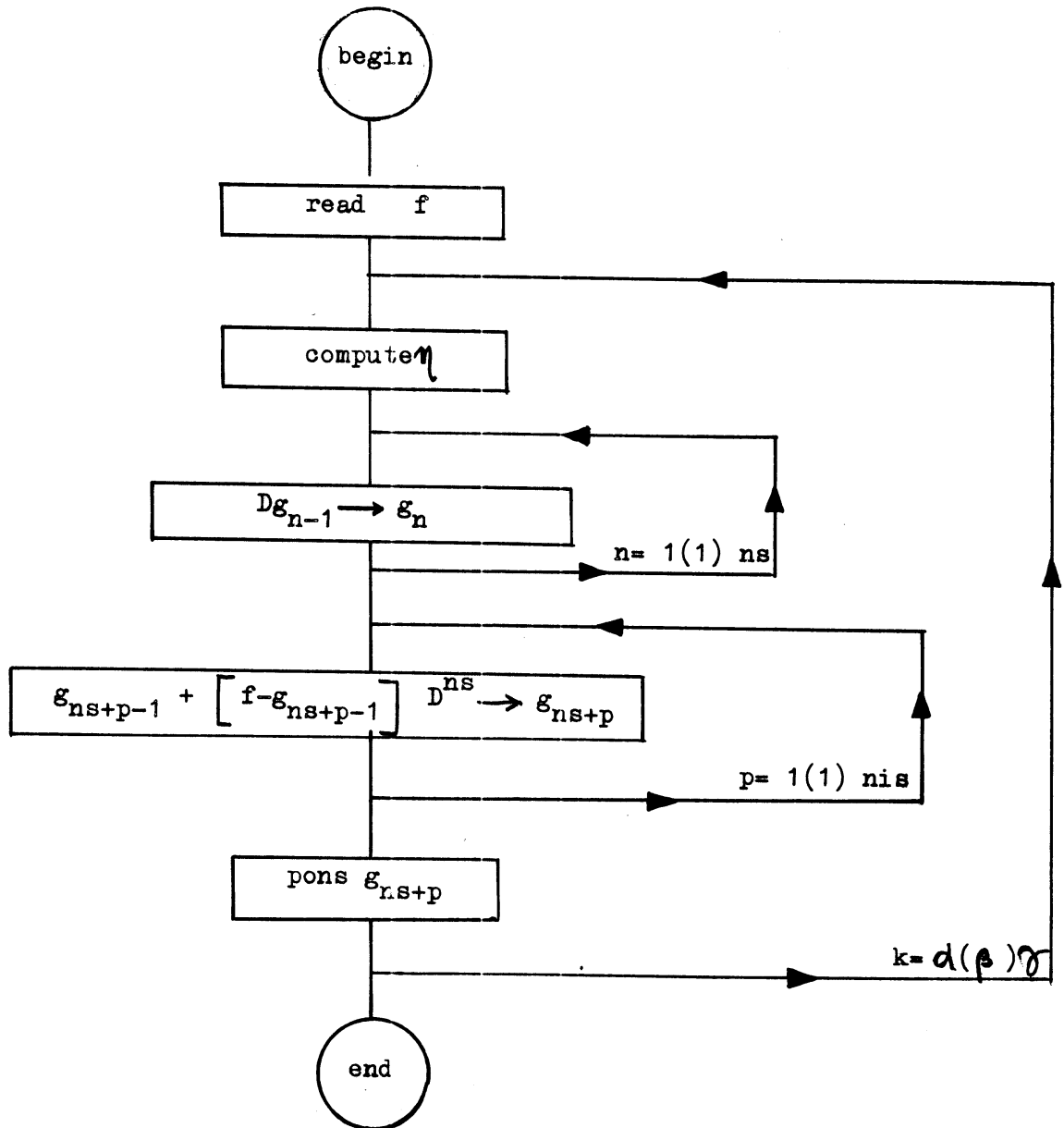
nis = number of inverse smoothing.

$sx = ty = d$ = roosterpuntsafstand.

De berekeningen die achtereenvolgens zijn verricht, zijn in tabel I genummerd van I t/m XXVI.

Het filter bepaald door $R/d = 1.6$, $E = 4$, $ns = 5$ en $nis = 10$ (zie tabel I en grafiek 4) toegepast op de funktie $f(x,y)$ gaf resultaten waarbij het grillige karakter verdween en de essentiële hoofdtekken aanwezig bleven. De bijbehorende draaitijd was ca 3 min.

Stroomdiagram:



Smoothing and Filtering

Tabel I

	R/d	ns	nis	f	draaitijd in min. TR4
I	1.6	40	10	24 hr.-prog.(3) voor 12-12-'62	21
II	1.6	20	10	"	11
III	1.6	10	10	"	6
IV	1.6	7	10	"	4
V	1.6	5	10	"	3 x
VI	1.6	3	10	"	2
VII	1.6	1	10	"	2
VIII	1.6	1	3	"	1
IX	1.6	5	6	"	3
X	1.6	1	0	"	1
XI	1.1	7	10	24 hr.-prog.(3) voor 12-12-'62	4
XII	1.1	5	10	"	3
XIII	1.1	3	10	"	2
XIV	1.1	1	10	"	2
XV	1.1	1	3	"	1
XVI	1.1	5	3	"	2
XVII	1.1	1	0	"	1
XVIII	1.6	5	10	24 hr.-prog.(3) voor 13-12-'62	3
XIX	1.6	5	10	24 hr.-prog.(3) voor 14-12-'62	3
XX	1.6	5	10	24 hr.-prog.(3) voor 15-12-'62	3
XXI	1.6	5	10	24 hr.-prog.(3) voor 16-12-'62	3
XXII	1.6	3	0	24 hr.-prog.(3) voor 12-12-'62	1
XXIII	1.6	3	2	"	1
XXIV	1.6	3	4	"	1 x
XXV	1.6	4	4	"	1
XXVI	1.6	4	6	"	2 x

AANHANGSEL.

In formule (1) hebben we een afvlakfunctie gedefinieerd, nl.:

$$g(x,y) = \int_0^{2\pi} \int_0^{\infty} f(x+r \cos \theta, y+r \sin \theta) w r dr d\theta / \int_0^{2\pi} \int_0^{\infty} w r dr d\theta$$

$$\text{met } w = \frac{1}{4\pi k} e^{-r^2/4k}$$

Het is gemakkelijk te verifiëren dat $\int_0^{2\pi} \int_0^{\infty} w r dr d\theta = 1$, zodat:

$$g(x,y) = \int_0^{2\pi} \int_0^{\infty} f(x+r \cos \theta, y+r \sin \theta) w r dr d\theta. \quad (18)$$

Voor het één-dimensionale geval definiëren we analoog een afvlakfunctie, nl.:

$$g(x) = \int_{-\infty}^{+\infty} f(x+t) w dt / \int_{-\infty}^{+\infty} w dt$$

$$\text{met } w = \frac{1}{\sqrt{4\pi k}} e^{-t^2/4k}$$

Ook nu geldt weer $\int_{-\infty}^{+\infty} w dt = 1$, zodat:

$$g(x) = \int_{-\infty}^{+\infty} f(x+t) w dt \quad (19)$$

Substitueren we vervolgens $f(x) = \sin ax$ (15) in formule (19) dan krijgen we:

$$g(x) = \int_{-\infty}^{+\infty} \sin a(x+t) \frac{1}{\sqrt{4\pi k}} e^{-t^2/4k} dt = \frac{1}{\sqrt{4\pi k}} * \int_{-\infty}^{+\infty} (\sin ax \cos at + \sin at \cos ax) e^{-t^2/4k} dt = \frac{1}{\sqrt{4\pi k}} \sin ax * \int_{-\infty}^{+\infty} \cos at \cdot e^{-t^2/4k} dt, \text{ omdat } \int_{-\infty}^{+\infty} \sin at \cdot e^{-t^2/4k} dt = 0$$

De integraal $\int_{-\infty}^{+\infty} \cos at \cdot e^{-t^2/4k} dt = \sqrt{4\pi k} \cdot e^{-ka^2}$, zodat

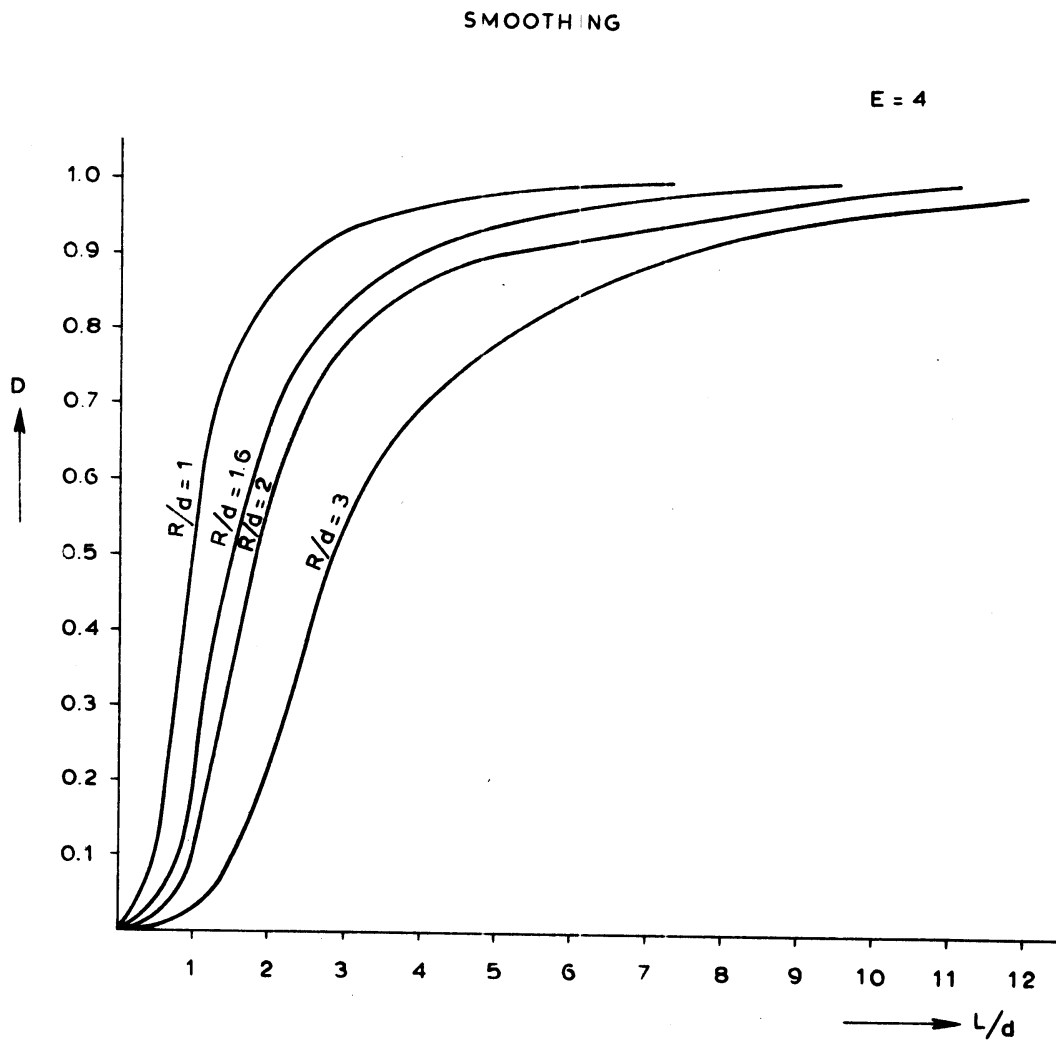
$$g(x) = \frac{1}{\sqrt{4\pi k}} \sin ax \sqrt{4\pi k} \cdot e^{-ka^2} \text{ dus:}$$

$$g(x) = e^{-ka^2} \sin ax$$

Stellen we $D = e^{-ka^2}$ dan verkrijgen we voor $g(x)$:

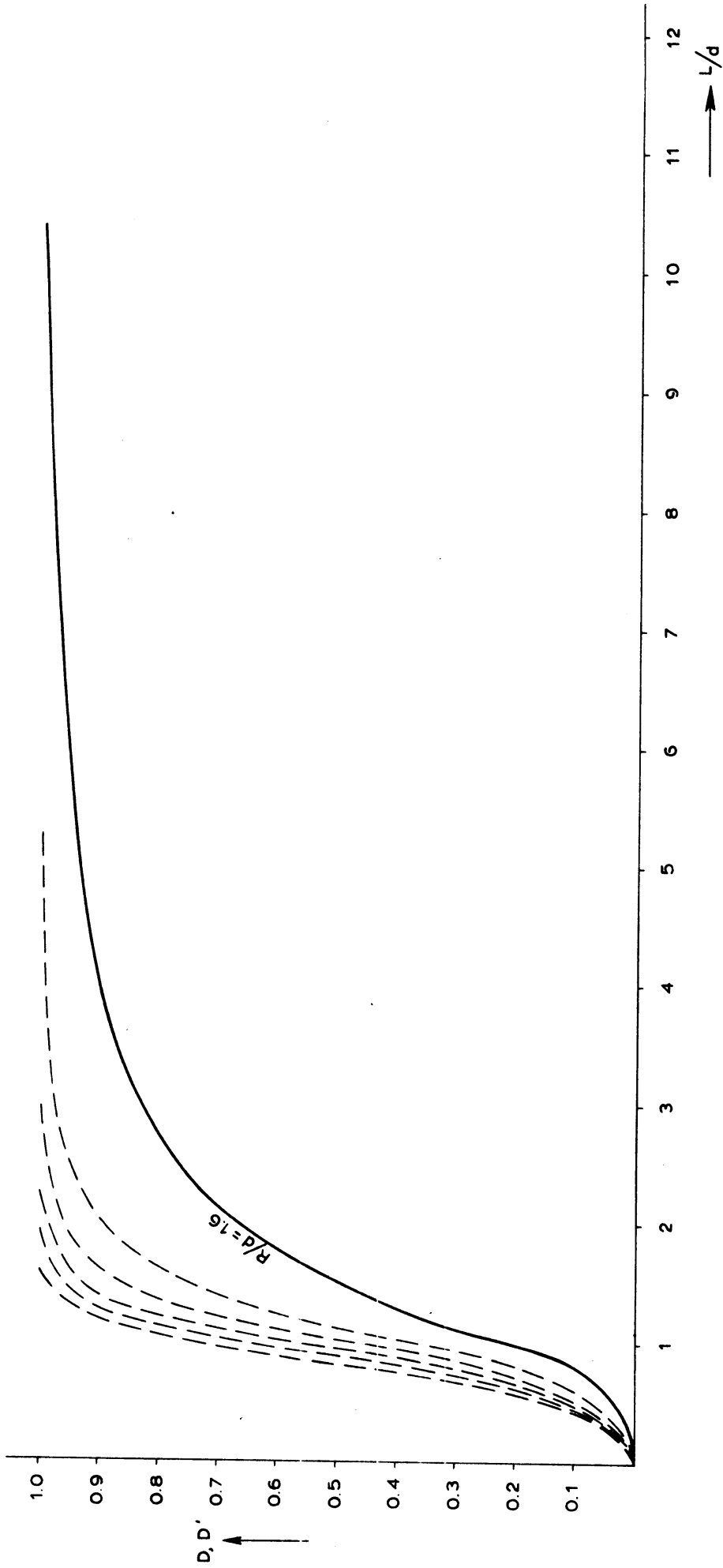
$$g(x) = D \sin ax = D f(x) \text{ (zie formule 16).}$$

Grafiek 1



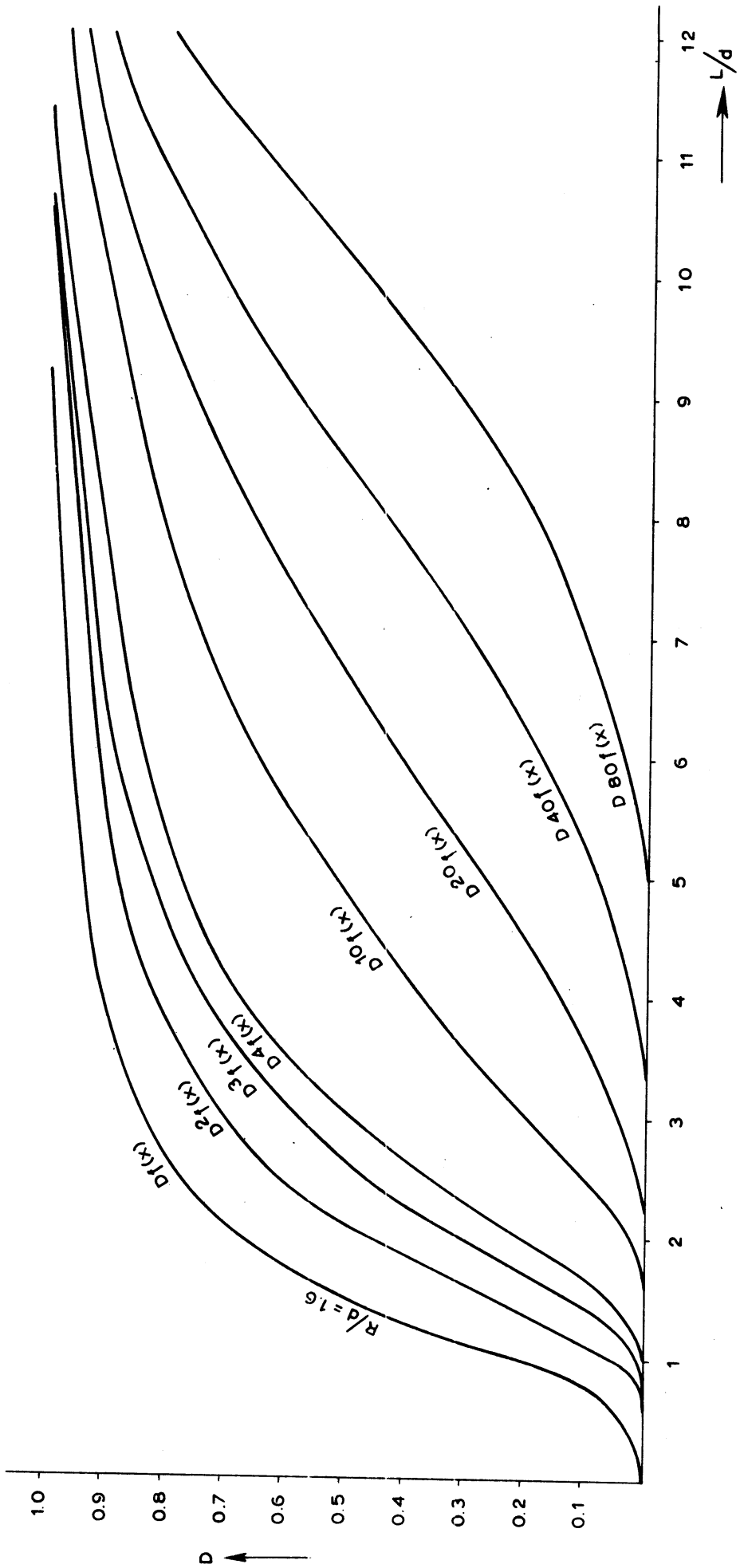
INVERSE SMOOTHING

$R/d = 1.6$ $E = 4$



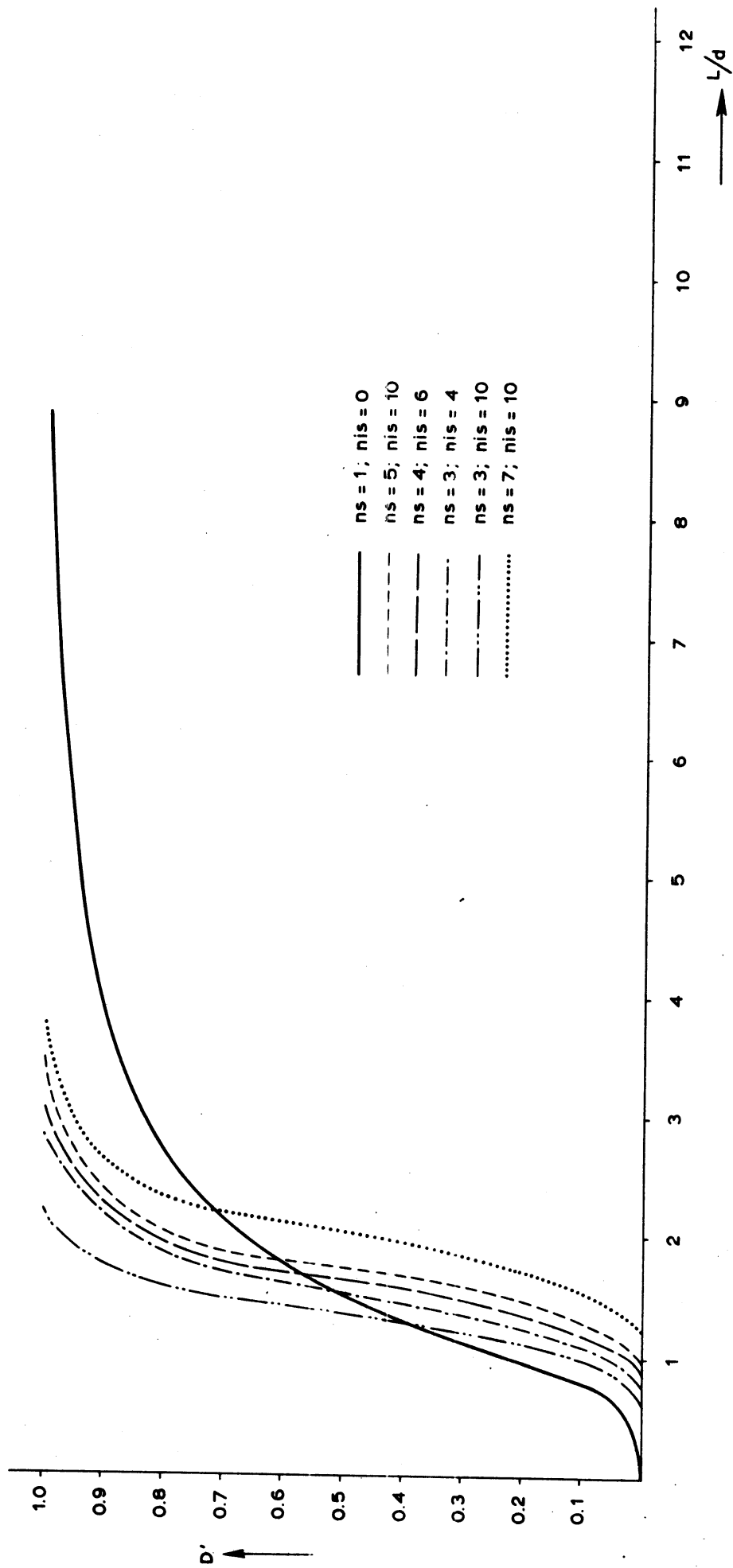
SMOOTHING

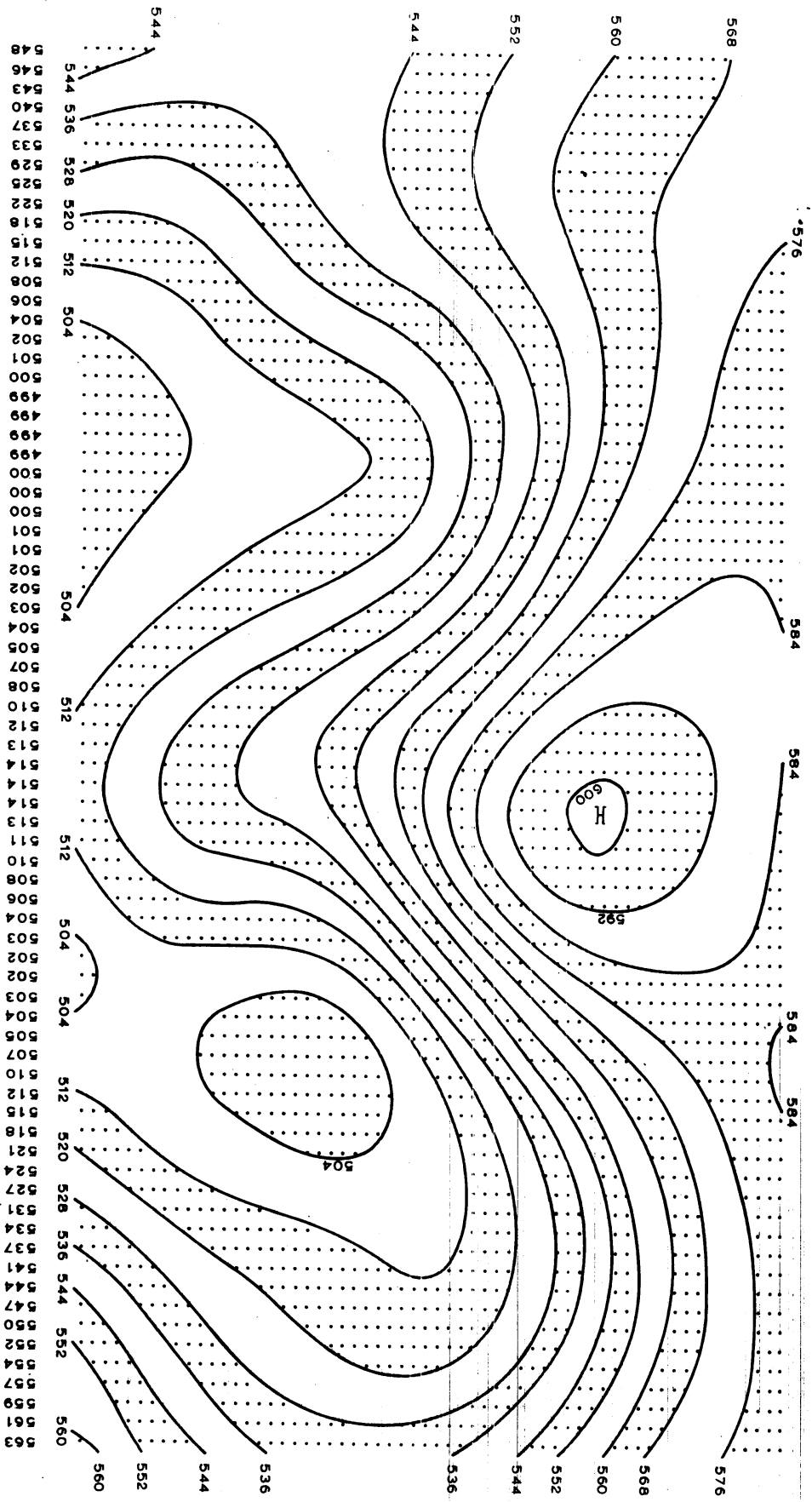
$R/d = 1.6$ $E = 4$



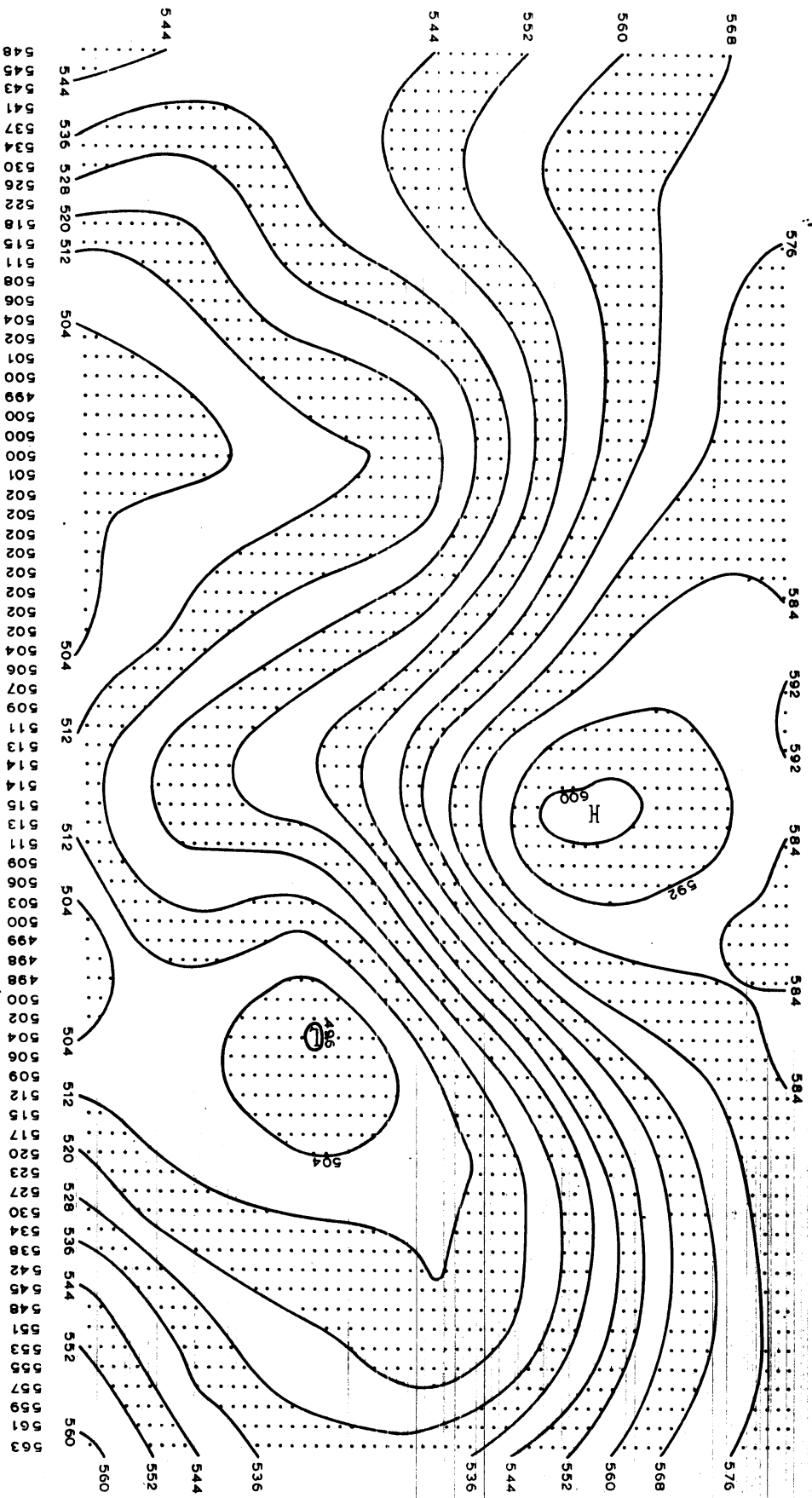
INVERSE SMOOTHING

$R/d = 1.6$ $E = 4$



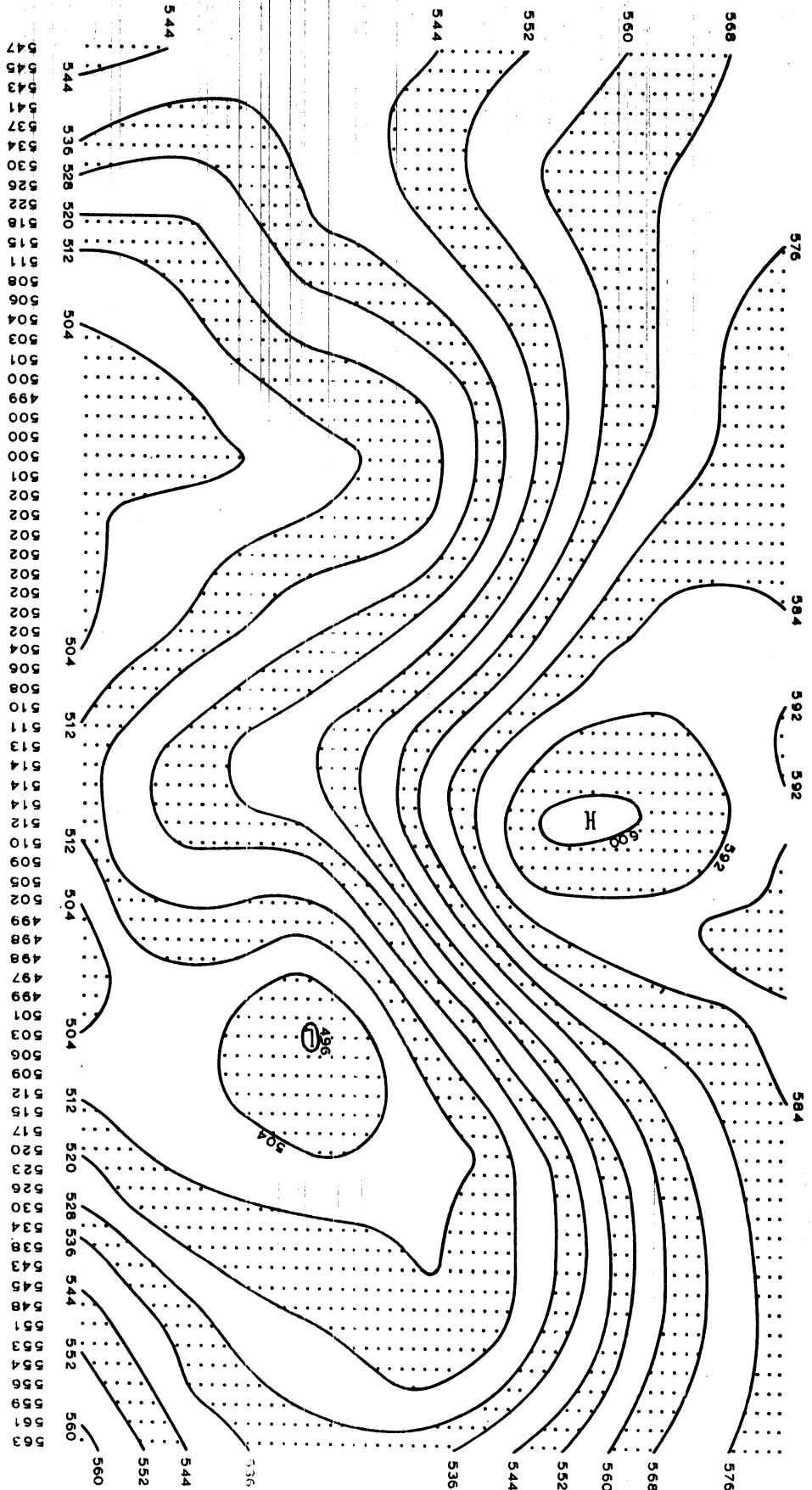


RESULTATEN III



RESULTATEN X

568 576 584 592 596 600 604 608 612 616 620 624 628 632 636 640 644 648 652 656 660 664 668 672 676 680 684 688 692 696 700 704 708 712 716 720 724 728 732 736 740 744 748 752 756 760 764 768 772 776 780 784 788 792 796 800 804 808 812 816 820 824 828 832 836 840 844 848 852 856 860 864 868 872 876 880 884 888 892 896 900 904 908 912 916 920 924 928 932 936 940 944 948 952 956 960 964 968 972 976 980 984 988 992 996 1000



RESULTATEN XXIV