

Rijkswaterstaat
Dienst Getijdewateren

Hoofdcomponenten-analyse van meetreeksen

H 535
juli 1988

AFGEHANDELD



waterloopkundig laboratorium | WL

Hoofdcomponenten-analyse van meetreeksen

C. de Valk

	BIBLIOTHEEK POSTBUS 177 3420 MA DELFT Waterloepkundig Laboratorium
3799	
H 0535	
	

R0004280

VER

INHOUD

	blz.
1. <u>Hoofdcomponenten-analyse</u>	1
1.1 Principe.....	1
1.2 Betekenis van hoofdcomponenten.....	4
1.3 Berekening.....	6
1.4 Voorspelling, stabiliteit.....	9
1.5 Significantie, selectie van hoofdcomponenten.....	11
2. <u>Toepassing op jaargemiddelde zeestanden en tijverschillen</u>	12
2.1 Jaargemiddelde zeestanden.....	12
2.2 Jaargemiddelde tijverschillen.....	13
2.3 Stabiliteit van de hoofdcomponenten-ontbinding.....	16
3. <u>Summary and conclusions</u>	18

LITERATUUR

VERKLARING VAN DE FIGUREN

FIGUREN

1. Hoofddcomponenten-analyse

1.1 Principe

Hoofddcomponenten-analyse (principal component analysis, empirical orthogonal functions analysis; verder afgekort als HCA) is een data-reductie methode: een methode om uit een grote data-matrix de meest essentiële informatie in compacte vorm weer te geven. Zo'n data-matrix bestaat bijvoorbeeld uit N metingen van M variabelen, of uit N samples van M functies, bijvoorbeeld tijdreeksen. Deze gegevens kunnen weergegeven worden als y_{ij} ; $i=1,\dots,M$; $j=1,\dots,N$, ofwel als de matrix Y met M kolommen en N rijen. In dit onderzoek komt met name de toepassing van HCA op M tijdreeksen van jaargemiddelde zeestand of tijverschil op N achtereenvolgende jaren aan de orde. Daarom zullen we aannemen dat we te maken hebben met de M functies $y_1,\dots,y_M$. en van iedere functie y_i zijn N samples $y_{i1},\dots,y_{iN}$ aanwezig.

Neem aan $N \geq M$. We kunnen deze functies ieder apart bekijken of paarsgewijs vergelijken, maar met name als N en M groot zijn is het moeilijk om een totaalbeeld te krijgen. Het idee is nu om een functie h_1 te zoeken en M weegfactoren $u_{11},\dots,u_{M1}$, zodat

$$\sum_{i=1}^M \sum_{j=1}^N [y_{ij} - u_{i1}h_{1j}]^2 \quad (1)$$

minimaal is. De functies $y_1,\dots,y_M$ worden dus, afgezien van vermenigvuldiging met een weegfactor, benaderd met dezelfde functie h_1 , de eerste hoofddcomponent: zowel h_1 als de weegfactoren $u_{11},\dots,u_{M1}$ worden zo gekozen dat de kwadraten som van de afwijkingen

$$d_{ij} = y_{ij} - u_{i1}h_{1j} \quad (2)$$

over alle functies en samples minimaal is. Overigens kunnen de weegfactoren zowel positief als negatief zijn. Voor het gemak zullen we aannemen dat

$$\sum_{i=1}^M u_{i1}^2 = 1 \quad (3)$$

(h_1 kan altijd zo gekozen worden dat dit het geval is). Het is verhelderend om een geometrische voorstelling te maken van het probleem: $y_{.j}$, $j=1,\dots,N$ kunnen worden voorgesteld als N punten (vectoren) in de M-dimensionale ruimte R^M .

De benaderingen $u_{.1}h_{1j}$, $j=1,\dots,N$ van deze punten liggen alle op een enkele lijn, want $u_{.1}h_{1j}$ is de vector $u_{.1}$ vermenigvuldigd met het getal h_{1j} . Dit is het gemakkelijkst in te zien voor het geval dat $M = 2$, zie figuur 0. Een belangrijke eigenschap van de gevonden benadering is dat de residuen $d_{.j}$ loodrecht staan op de benaderingen $u_{.1}h_{1j}$ van de punten $y_{.j}$: stel dat $u_{.1}$ al is gegeven, dan moet h_{1j} zo gekozen worden dat de lengte van $d_{.j}$ minimaal is, d.w.z. $u_{.1}h_{1j}$ is de loodrechte projectie van $y_{.j}$ op de rechte lijn L_1 door $u_{.1}$ (zie fig.1). Het feit dat $d_{.j}$ loodrecht staat op $u_{.1}$, ofwel

$$\sum_{i=1}^M d_{ij} u_{i1} = 0 \quad (4)$$

betekent dat (gebruik makend van (3))

$$\sum_{j=1}^N \sum_{i=1}^M y_{ij}^2 = \sum_{j=1}^N h_{1j}^2 + \sum_{j=1}^N \sum_{i=1}^M d_{ij}^2 \quad (5)$$

ofwel de totale kwadraten som is de kwadraten som van de benadering (de middelste term in (5)) plus de kwadraten som van de residuen (de laatste term in (5)). Minimaliseren van de kwadraten som van de residuen ((1) of de laatste term in (5)) komt dus overeen met maximaliseren van de middelste term in (5). We kunnen nu een stap verder gaan: ieder van de punten (data) $y_{.j}$, $j=1,\dots,N$ is benaderd d.m.v. $u_{.1}h_{1j}$; we gaan deze benadering verbeteren door ieder van de residuen $d_{.j}$, $j=1,\dots,N$ op hun beurt te benaderen d.m.v. $u_{.2}h_{2j}$, een punt op een rechte lijn L_2 . Omdat de residuen $d_{.j}$, $j=1,\dots,N$ loodrecht staan op de lijn L_1 door $u_{.1}$, geldt dat L_2 loodrecht moet staan op L_1 , ofwel:

$$\sum_{i=1}^M u_{i1} u_{i2} = 0 \quad (6)$$

In het geval dat $M = 2$ zijn we nu klaar: we hebben ieder punt $y_{.j}$ exact uitgedrukt in twee nieuwe coördinaten, nl. h_{1j} , de lengte van de projectie van $y_{.j}$ op L_1 , en h_{2j} , de lengte van de projectie van $y_{.j}$ op L_2 (zie fig. 0). Wanneer M groter is dan 2 kunnen we nog verder gaan, en $u_{.3}$ en $h_{3j}, \dots, u_{.M}$ en h_{Mj} zoeken (in het geval dat het aantal lineair onafhankelijke kolommen r van Y kleiner is dan M , dan liggen alle punten $y_{.j}$ in een r -dimensionale deelruimte van R^M ; in dat geval zullen $h_{r+1}, \dots, h_M$ gelijk aan nul zijn). Uiteindelijk is de data-matrix Y (zie boven) dus weergegeven als

$$y_{ij} = \sum_{p=1}^M u_{ip} h_{pj} \quad (7)$$

met de vectoren u_p ; $p=1, \dots, M$ van lengte 1 en onderling loodrecht (ze vormen een orthonormaal stelsel) (cf. 6)):

$$\sum_{i=1}^M u_{ip} u_{im} = \delta_{pm} \quad (8)$$

en waarin h_m , de m -de hoofdcomponent, en u_m zo zijn gekozen dat de data zo goed mogelijk benaderd worden met de eerste m hoofdcomponenten, dus zo dat

$$\sum_{j=1}^N \sum_{i=1}^M [y_{ij} - \sum_{p=1}^m u_{pi} h_{pj}]^2 = \sum_{j=1}^N \sum_{i=1}^M y_{ij}^2 - \sum_{j=1}^N \sum_{p=1}^m h_{pj}^2 \quad (10)$$

(cf. (5)) minimaal is gegeven $u_1, h_1, \dots, u_{m-1}, h_{m-1}$; Er geldt uiteindelijk:

$$\sigma^2 \equiv \sum_{j=1}^N \sum_{i=1}^M y_{ij}^2 = \sum_{j=1}^N \sum_{p=1}^M h_{pj}^2 \quad (11)$$

Vgl. (11) geeft aan dat de totale kwadratensom van de data σ^2 is verdeeld over de functies $h_1, \dots, h_M$. (d.w.z. de kwadratensom van de hoofdcomponenten gezamenlijk is gelijk aan de kwadratensom van de oorspronkelijke data).

Terugkerend naar figuur 1, dan is duidelijk dat de kwaliteit van een benadering van een puntenwolk d.m.v. punten op een of meerdere rechte lijnen door de oorsprong sterk afhangt van de positie van de puntenwolk t.o.v. de oorsprong. De benadering is altijd het best (de kwadratensom van de residuen het kleinst) wanneer de oorsprong in het zwaartepunt van de puntenwolk ligt: d.w.z.

$$\sum_{j=1}^N y_{ij} = 0 \text{ voor } i=1, \dots, M \quad (12)$$

Daarom wordt HCA altijd uitgevoerd op functies $y_1, \dots, y_M$ die vooraf genormeerd zijn op gemiddelde nul. Het gevolg is dat ook $h_1, \dots, h_M$ ieder gemiddeld nul zijn.

Dikwijls worden de hoofdcomponenten genormeerd op kwadratensom 1: Laat σ_p^2 de kwadratensom zijn van de p -de hoofdcomponent:

$$\sum_{j=1}^N h_{pj}^2 = \sigma_p^2 \quad (13)$$

Laat v_p de hoofdcomponent h_p zijn, genormeerd op kwadratensom 1:

$$h_{p.} = \sigma_p v_{p.} \text{ met } \sigma_p = \sqrt{\sigma_p^2} \text{ en } \sum_{j=1}^N v_{pj}^2 = 1 \quad (14)$$

dan wordt (7):

$$y_{ij} = \sum_{p=1}^N v_{pj} \sigma_p u_{ip} \quad (15)$$

en (11) wordt

$$\sigma^2 = \sum_{j=1}^N \sum_{i=1}^M y_{ij}^2 = \sum_{p=1}^M \sigma_p^2 \quad (16)$$

Zowel de functies $h_{1.}, \dots, h_{M.}$ als de genormeerde functies $v_{1.}, \dots, v_{M.}$ worden dikwijls aangeduid als de hoofdcomponenten. We zullen hier "hoofd component" gebruiken voor $h_{p.}$ en "genormeerde hoofdcomponent" voor $v_{p.}$, teneinde verwarring te voorkomen. De getallen $\sigma_1, \dots, \sigma_M$ worden aangeduid als de singuliere waarden (zie ook 1.3).

1.2 Betekenis van hoofdcomponenten

Hoofdcomponenten-analyse is een methode om de meest essentiële kenmerken van een aantal bemonsterde functies (we zullen verder veronderstellen: tijdreeksen) weer te geven. Er gelden een aantal voorwaarden opdat de toepassing van HCA zin heeft. In de eerste plaats moeten de tijdreeksen vergelijkbaar zijn, in de zin dat een som of verschil van twee tijdreeksen geen onzin oplevert; de hoofdcomponenten zijn namelijk lineaire combinaties van de oorspronkelijke reeksen (zie (7): de reeksen zelf zijn lineaire combinaties van de hoofdcomponenten, dus het omgekeerde is ook het geval; dit feit is expliciet uitgedrukt in (18) van 1.3). Bijvoorbeeld: een analyse van een set reeksen die zowel zee-standen als tijverschillen omvat heeft geen zin. Overigens wordt deze voorwaarde in de praktijk niet altijd even strikt genomen; in het geval van reeksen van niet-vergelijkbare grootheden worden de reeksen vaak niet alleen genormeerd op gemiddelde nul, maar ook op kwadratensom 1 (of een andere constante). Dit voorkomt dat bepaalde reeksen met een grote kwadratensom het resultaat van de analyse gaan domineren. Een tweede voorwaarde is dat ook de samples vergelijkbaar moeten zijn: in het geval van tijdreeksen houdt dit in dat het j-de sample in iedere reeks betrekking heeft op (ongeveer) hetzelfde tijdstip of tijd-interval. Verder moeten de gemeten waarden y_{ij} , $j=1, \dots, N$, $i=1, \dots, M$ beschouwd kunnen worden als getallen op de reële lijn; reeksen met alleen enen en nullen (wel/niet) komen bijvoorbeeld niet in aanmerking. Of een hoofdcompo-

nenten-analyse zinvolle informatie verschaft kan pas achteraf beoordeeld worden. In het algemeen geldt dat een HCA succesvol is als middel tot data-reductie wanneer de bijdrage van een klein aantal hoofdcomponenten tot de totale kwadratensom van de data zo groot mogelijk is (men zegt ook wel: wanneer de fractie van de totale kwadratensom "verklaard" door een klein aantal hoofdcomponenten zo groot mogelijk is). Dit betekent namelijk dat in de oorspronkelijke data een sterke structuur aanwezig is: de puntenwolk uit 1.1 is sterk geconcentreerd in een bepaalde deelruimte, dus er is een duidelijke (empirische) relatie tussen de reeksen. Deze relatie kan zowel een overeenkomst zijn (d.w.z. de weegfactoren uit 1.1 hebben hetzelfde teken) als een contrast (het teken van de weegfactoren varieert). Data-reductie is in het algemeen geen doel op zich: het doel is meestal interpretatie van de gegevens. Wel geldt dat hoe groter de bijdrage van een klein aantal hoofdcomponenten tot de totale kwadratensom is, hoe meer reden er is om een interpretatie van deze hoofdcomponenten te zoeken: de structuur die blijkbaar in de gegevens aanwezig is vraagt dan om een verklaring. In het algemeen hebben we dus vooral belangstelling voor de hoofdcomponenten met de grootste bijdrage tot de totale kwadratensom.

De hoofdcomponenten die ieder maar weinig tot de totale kwadratensom bijdragen zijn ofwel niet te interpreteren, ofwel weerspiegelen de specifieke kenmerken van een enkele reeks (dit laatste verschijnsel doet zich vanzelfsprekend sterker voor naarmate een kleiner aantal reeksen geanalyseerd wordt). In analogie met regressie zou men een criterium kunnen wensen om te bepalen welke hoofdcomponenten significant zijn. Het begrip significantie zoals dat in regressie gehanteerd wordt blijkt m.b.t. HCA niet relevant (hoofdcomponenten zijn functies van de waarnemingen zelf, i.t.t. de onafhankelijke variabelen in regressie). Er bestaat wel een ander concept dat (onder bepaalde vrij strikte voorwaarden) bruikbaar is om het minimale aantal hoofdcomponenten vast te stellen dat een adequate beschrijving van de data oplevert, namelijk model orde bepaling. Dit is verder uitgewerkt in paragraaf 1.5. Feitelijk is het nut van een techniek om hoofdcomponenten te selecteren niet zo groot: HCA is een techniek om een patroon in data te ontdekken, het hoeft geen model voor de data op te leveren. In de praktijk is het vooral van belang of een hoofdcomponent geïnterpreteerd kan worden. Het nut van een HCA bij het interpreteren van gegevens hangt met name af van de mate waarin verschillende kenmerken van de tijdreeksen "uit elkaar worden getrokken". Bijvoorbeeld: in de reeksen van jaargemiddeld tijverskil zit zowel een trend als een sinusvormige component. Wanneer

trend en sinus in twee verschillende hoofdcomponenten zouden verschijnen, dan zijn die hoofdcomponenten elk apart gemakkelijk te interpreteren. Helaas is het meestal niet zo fraai (dit blijkt ook in hoofdstuk 2.); je mag alleen verwachten dat de aspecten die van belang zijn (in termen van kwadratensom) in de eerste hoofdcomponenten opduiken. Een poging om tot hoofdcomponenten te komen die ieder op zichzelf een eenduidige interpretatie hebben is de factoranalyse, zoals ontwikkeld met name voor het psychologisch onderzoek. Dit houdt in dat nieuwe lineaire combinaties van hoofdcomponenten gezocht worden, die duidelijker te interpreteren zijn (opm: onder "factoranalyse" wordt ook wel iets anders verstaan). Een dergelijke procedure leidt soms tot aardige resultaten, maar is vanzelfsprekend subjectief, en zeker niet gemakkelijk toe te passen.

Tot slot kan nog worden opgemerkt dat er een sterke overeenkomst bestaat tussen spectrale analyse en HCA. Zowel in spectrale analyse als in HCA hebben we een aantal reeksen die beschouwd kunnen worden als realisaties van een proces (in spectrale analyse zijn dit vaak deelreeksen van een enkele reeks). In beide gevallen worden de reeksen geprojecteerd op (ontbonden in) onderling loodrechte functies: in spectrale analyse zijn dit sinus- en cosinus functies met bepaalde frequenties; in HCA zijn deze functies niet vooraf gegeven, maar zijn bepaald door de data zelf. Het spectrum, de verdeling van de totale kwadratensom over de verschillende frequenties, is te vergelijken met de waarden $\sigma^2_1, \dots, \sigma^2_M$, de verdeling van de kwadratensom over de verschillende hoofdcomponenten. Een verschil is dat de interpretatie van een spectrum min of meer voor zich spreekt; dit geldt zeker niet voor de interpretatie van hoofdcomponenten. Daar tegenover staat dat spectrale analyse alleen bruikbaar is voor (in gemiddelde en variantie) stationaire signalen; deze beperking geldt niet voor HCA.

1.3 Berekening

We gaan hier uit van de data-matrix Y , met als kolommen de functies $y_1, \dots, y_M$; rij j komt overeen met sample j . Gezocht is een weergave van Y van de vorm (7), dus

$$y_{ij} = \sum_{p=1}^M h_{pj} u_{ip} \quad (17)$$

vanwege (8) geldt nu

$$\sum_{i=1}^M y_{ij} u_{im} = \sum_{i=1}^M \sum_{p=1}^M h_{pj} u_{ip} u_{im} = h_{mj}$$

Ofwel voor $m=1, \dots, M$: $h_{mj} = \sum_{i=1}^M y_{ij} u_{im}$ (18)

$h_{1.}$ en $u_{.1}$, moeten zo worden gekozen dat (1) minimaal is, dus (zie (11)) zo dat de bijdrage van de eerste hoofdcomponent aan de totale kwadratensom,

$$\sum_{j=1}^N h_{ij}^2$$
 (19)

maximaal is, en met (18) volgt dat $u_{.1}$ dus de vector met lengte 1 is die

$$\sum_{j=1}^N \left[\sum_{i=1}^M y_{ij} u_{i1} \right]^2 = \sum_{j=1}^N h_{ij}^2$$
 (20)

maximaliseert; ofwel in een meer compacte schrijfwijze: $u_{.1}$ is de vector met lengte 1 die

$$u_{.1}' Y' Y u_{.1}$$

maximaliseert (Y' is getransponeerde van Y , $u_{.1}'$ is de rij-vector overeenkomend met de kolom-vector $u_{.1}$). Uit appendix 1 (zie volgende pagina) volgt dat $u_{.1}$ dus de eigenvector van de matrix $Y'Y$ moet zijn die behoort bij de grootste eigenwaarde σ_1^2 van $Y'Y$. Voor $u_{.m}$ en $h_{m.}$, $1 < m \leq M$ geldt ongeveer hetzelfde: dan moet een $h_{m.}$ gekozen worden en een $u_{.m}$ loodrecht op $u_{.1}, \dots, u_{.m-1}$, zodat

$$\sum_{j=1}^N h_{mj}^2 = \sum_{j=1}^N \left[\sum_{i=1}^M y_{ij} u_{im} \right]^2 = u_{.m}' Y' Y u_{.m}$$
 (21)

zo groot mogelijk. Uit appendix 1 volgt dan dat $u_{.m}$ de eigenvector is van de matrix $Y'Y$ die correspondeert met de m -de (in grootte) eigenwaarde σ_m^2 van $Y'Y$. Nu $u_{.1}, \dots, u_{.M}$ bekend zijn, kunnen $h_{1.}, \dots, h_{M.}$ direct gevonden worden uit (18). Nu blijkt eveneens dat $h_{1.}, \dots, h_{M.}$ onderling loodrecht zijn:

$$\sum_{j=1}^N h_{mj} h_{pj} = \sum_{j=1}^N \left[\sum_{i=1}^M y_{ij} u_{im} \right] \left[\sum_{i=1}^M y_{ij} u_{ip} \right] =$$

$$u_{.m}' Y' Y u_{.p} = u_{.m}' u_{.p} \sigma_m^2 = \sigma_m^2 \delta_{mp}$$
 (22)

Laat v_p , de functie h_p , zijn genormeerd op lengte 1, volgens (13), (14):

$$\sigma_p v_p = h_p \quad (23)$$

(merk op dat σ_p nul kan zijn); dan kunnen we de ontbinding (15) van Y nu ook schrijven als

$$Y = V A U' \quad (24)$$

waarin de M kolommen van V onderling orthogonaal zijn en van lengte 1; hetzelfde geldt voor de kolommen van U, en A is een MXM diagonaalmatrix met op de diagonaal achtereenvolgens $\sigma_1, \dots, \sigma_M$, met $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M \geq 0$. De ontbinding van Y als in (18) wordt de singuliere waarden ontbinding genoemd. De berekening van de u_p, v_p , en σ_p , $p=1, \dots, M$ kan gebeuren via de eigenwaarden en eigenvectoren van $Y'Y$ (uit publicaties blijkt dat deze methode nog veel wordt toegepast). Een andere methode is het direct berekenen van de singuliere waarden ontbinding (18) (zie Golub and Reinsch, 1970); deze methode is veel beter uit oogpunt van numerieke stabiliteit. De singuliere waarde ontbinding (singular value decomposition) is opgenomen in FORTRAN-libraries zoals IMSL en NAG.

Appendix 1: Rayleigh's principe.

Stel we hebben een NXN symmetrische matrix Z (bijvoorbeeld $Z = Y'Y$); deze heeft dan M eigenwaarden $\sigma_1^2 \geq \dots \geq \sigma_M^2 \geq 0$ en een stelsel van M onderling loodrechte eigenvectoren $u_1, \dots, u_M$, elk genormeerd op lengte 1 (zie b.v. Noble and Daniel, theorem 9.3). Gevraagd is de vector b_1 van lengte 1 zodat $b_1' Z b_1$ maximaal. Stel U is $[u_1, \dots, u_M]$, de matrix met als kolommen $u_1, \dots, u_M$. Dan is er een vector c zodat $b_1 = Uc$ (nl. $c = U' b_1$) en

$$b_1' Z b_1 = c' U' Z U c = c' U' \sum_{j=1}^M \sigma_j^2 u_j u_j' c = c' U' U A c = c' A c$$

met A een diagonaalmatrix met op de diagonaal van boven naar beneden $\sigma_1, \dots, \sigma_M$. $c' A c$ is maximaal als $c = [1 \ 0 \ \dots \ 0]'$, en dan geldt

$$b_1' Z b_1 = c' A c = \sigma_1^2 \quad (A.2)$$

In dat geval is $b_1 = Uc = u_1$, de eigenvector van Z die correspondeert met de grootste eigenwaarde van Z. Op analoge wijze volgt dat de vector b_2 met lengte

1 loodrecht op b_1 , waarvoor $b_2'Zb_2$ maximaal is, gelijk is aan u_2 , de eigenvector van Z behorend bij de op een na grootste eigenwaarde σ^2_2 van Z , etc. etc.

1.4 Voorspelling, stabiliteit

In het geval dat de data, de functies $y_1, \dots, y_M$ tijdreeksen zijn, dan zouden de hoofdcomponenten gebruikt kunnen worden om deze reeksen te voorspellen; de eerste m hoofdcomponenten worden voorspeld, en hieruit wordt vervolgens het toekomstige verloop van de oorspronkelijke reeksen gereconstrueerd. Een mogelijk voordeel van deze methode is dat het gemakkelijker is een klein aantal hoofdcomponenten te voorspellen dan om alle oorspronkelijke reeksen apart te voorspellen, of dat de hoofdcomponenten beter voorspelbaar zijn dan de oorspronkelijke reeksen. Wil de methode werken, dan moet aan de volgende voorwaarden voldaan zijn:

- [1] $y_1, \dots, y_M$ adequaat kunnen worden weergegeven met de eerste $m < M$ hoofdcomponenten $h_1, \dots, h_m$.
- [2] Deze eerste m hoofdcomponenten kunnen voldoende nauwkeurig worden voorspeld.
- [3] de bijdragen $u_{1p}, \dots, u_{Mp}$ van de hoofdcomponent h_p tot de tijdreeksen $y_1, \dots, y_M$ zijn voldoende constant in de tijd (dit wordt aangeduid als stabiliteit van de hoofdcomponenten-ontbinding). Is dit het geval, dan kan een voorspelling van de reeksen $y_1, \dots, y_M$ worden geconstrueerd uit een voorspelling van de hoofdcomponenten.

Een andere toepassing van hoofdcomponenten-analyse in voorspelling is het terugbrengen van het aantal inputsignalen van een regressiemodel dat wordt gebruikt voor voorspelling. Bijvoorbeeld: een regressiemodel dat wordt gebruikt voor voorspelling van rivierafvoeren kan een groot aantal input-signalen omvatten (gemeten afvoeren bovenstrooms, neerslaggegevens) die onderling sterk gecorrelleerd kunnen zijn. In zo'n geval kan de informatie in de inputsignalen effectief worden weergegeven in een klein aantal hoofdcomponenten. Deze reductie van het aantal ingangssignalen leidt dan tot een eenvoudiger model. In Nyberg (1981) worden zowel de te voorspellen signalen (zeestanden bij verschillende stations) als de inputsignalen (luchtdruk) ontbonden in hoofdcom-

ponenten. Ook bij ontbinding van inputsignalen in hoofdcomponenten is stabiliteit (zie [3] boven) van belang, omdat de weegfactoren u niet op ieder tijdstip opnieuw berekend worden.

Stabiliteit van een hoofdcomponenten-ontbinding wordt onderzocht door te kijken of de weegfactoren u berekend uit de gegeven set tijdreeksen over de samples $1, \dots, N$ gelijk zijn aan de weegfactoren berekend uit dezelfde signalen maar over de samples $1+L, \dots, N+L$, met L de verschuiving in de tijd. Gegeven een set reeksen van vaste lengte, dan kunnen de weegfactoren berekend over de eerste helft van de reeksen, en de weegfactoren berekend over de tweede helft van de reeksen vergeleken worden; deze kunnen ook weer vergeleken worden met de weegfactoren u berekend uit de volledige reeksen.

De stabiliteits-analyse zoals geïmplementeerd (zie 2.3) is als volgt. Een data-matrix Y (zie 1.1) met M kolommen (M tijdreeksen) en N rijen (N samples, neem aan N even) is gesplitst in 2 matrices van M kolommen en $N/2$ rijen Y_1 en Y_2 (dus iedere tijdreeks is gesplitst in 2 deelreeksen). Een HCA wordt uitgevoerd op Y_1 en Y_2 apart (na normeren van de kolommen (tijdreeksen) van Y_1 en Y_2 op gemiddelde nul): dit leidt tot de hoofdcomponenten-ontbindingen (zie 1.3):

$$y^1_{ij} = \sum_{p=1}^M h^1_{pj} u^1_{ip} \quad y^2_{ij} = \sum_{p=1}^M h^2_{pj} u^2_{ip} \quad (25)$$

Stabiliteit houdt in dat $u^1_{.p}$ en $u^2_{.p}$ overeenkomen, dus dit kan onderzocht worden door $u^1_{.p}$ en $u^2_{.p}$ samen in 1 figuur te plotten en te vergelijken voor $p=1, \dots, M$. Bovendien geldt (zie (18)):

$$h^1_{pj} = \sum_{i=1}^M y_{ij} u^1_{ip} \quad h^2_{pj} = \sum_{i=1}^M y_{ij} u^2_{ip} \quad (26)$$

Definieer

$$w^1_{pj} = \sum_{i=1}^M y_{ij} u^2_{ip} \quad w^2_{pj} = \sum_{i=1}^M y_{ij} u^1_{ip} \quad (27)$$

Wanneer $u^1_{.i} = u^2_{.i}$, dan zou dus gelden dat $h^1_{pj} = w^1_{pj}$ en $h^2_{pj} = w^2_{pj}$. We kunnen nu het effect van verschillen tussen $u^1_{.i}$ en $u^2_{.i}$ op de hoofdcomponenten nagaan door $h^1_{.p}$ te vergelijken met $w^1_{.p}$ en $h^2_{.p}$ te vergelijken met $w^2_{.p}$; d.w.z. we vergelijken de hoofdcomponenten van de eerste deelreeks met de hoofdcomponenten die gevonden zouden worden als de weegfactoren die van de tweede deelreeks waren, en omgekeerd. In het geïmplementeerde algoritme worden indien nodig $u^2_{.p}$ en $h^2_{.p}$ beiden met -1 vermenigvuldigd wanneer

$$\sum_{i=1}^M [u1_{ip} + u2_{ip}]^2 < \sum_{i=1}^M [u1_{ip} - u2_{ip}]^2 \quad (28)$$

(deze transformatie laat de hoofdcomponenten-ontbinding (7) intact, maar zorgt ervoor dat de afstand tussen $u1_{.p}$ en $u2_{.p}$ zo klein mogelijk is).

1.5 Significantie, selectie van hoofdcomponenten

Zowel in HCA als in regressie worden de data (de meetreeksen $y_{i.}$, $i=1, \dots, M$) ieder uitgedrukt als een lineaire functie van andere functies. In HCA zijn dit de hoofdcomponenten, in regressie zijn dit de vooraf gegeven zgn. onafhankelijke variabelen, bijvoorbeeld een constante en een functie die lineair toeneemt in de tijd. In het laatste geval is de trend in iedere functie $y_{i.}$ de parameter die de bijdrage weergeeft van de lineair stijgende functie tot $y_{i.}$. In regressie-analyse is een trend significant wanneer onder de aanname dat de trend nul is (de nulhypothese), de kans dat de trend groter of gelijk zou kunnen zijn aan de gemeten trend beneden een bepaalde waarde ligt (gegeven dat de verstoringen onderling onafhankelijk en normaal verdeeld zijn). In HCA is dit concept is niet toepasbaar, omdat we in dit geval geen nulhypothese hebben: de hoofdcomponenten zijn vooraf niet vastgelegd, omdat het functies zijn van de data zelf. Het komt er op neer dat significantie van hoofdcomponenten alleen relevant is wanneer we een model hebben voor de data. Een voorbeeld van een techniek om het aantal "significante" hoofdcomponenten te schatten gegeven een bepaald model voor de data is te vinden in Wac and Kailath (1975).

De veronderstellingen waarop deze techniek gebaseerd is beperken de toepasbaarheid voor veel problemen in de praktijk, waar hoofdcomponenten-analyse gebruikt wordt: met name de aannamen dat de signalen $s_{k.}$, $k=1, \dots, m$ stationair zijn en dat de verstoringen v_{ij} , $i=1, \dots, M$, $j=1, \dots, N$ dezelfde variantie hebben zullen in veel gevallen vrij onrealistisch zijn. Daarbij moet worden opgemerkt dat men zich in zulke gevallen moet afvragen of het gebruik van een techniek voor het scheiden van hoofdcomponenten in "signaal" en "ruis" wel zin heeft: meestal gaat het er om of een hoofdcomponent al dan niet geïnterpreerd kan worden; regeltjes zoals het MDL-criterium gaan altijd uit van een of ander voorgebakken model voor de data dat meestal niet relevant is. Wanneer hoofdcomponenten worden gebruikt als data-reductie methode, dan kan men vooraf kiezen welk en dan het kleinste aantal hoofdcomponenten selecteren waarmee aan dit criterium wordt voldaan.

2. Toepassing op jaargemiddelde zeestanden en tijverschillen

2.1 Jaargemiddelde zeestanden

Een HCA is uitgevoerd op jaargemiddelden van de zeestand gemeten op 10 verschillende plaatsen langs de kust (Vlissingen, Zierikzee, Brouwershaven, Hellevoetsluis, Hoek van Holland, IJmuiden, Den Helder, Terschelling, Harlingen en Delfzijl). Omdat niet van alle stations over dezelfde jaren gegevens beschikbaar zijn, zijn 3 runs gemaakt:

- [1] met alle 10 stations over 1887-1969.
- [2] zonder de stations Brouwershaven en Hellevoetsluis, over 1887-1985.
- [3] zonder Terschelling, over 1872-1969.

Voorafgaand aan iedere analyse zijn de reeksen genormeerd op gemiddelde nul. Normering van de reeksen op kwadraten som heeft niet plaatsgevonden: dit wordt in dit geval als een onnodige vervorming van de oorspronkelijke gegevens beschouwd (ze zijn namelijk goed vergelijkbaar), en kan alleen de interpretatie van de resultaten compliceren. Resultaten van de verschillende runs zullen hieronder worden besproken. De nummering van de figuren komt overeen met de nummering van de runs. Hoofdcomponent wordt afgekort als HC. De figuren dienen als volgt geïnterpreteerd te worden: neem bijvoorbeeld fig. 1, de eerste hoofdcomponent.

De onderste figuur toont het verloop in de tijd van de genormeerde eerste hoofdcomponent v_1 . (zie 1.1 vgl. (15)), en de bovenste figuur toont de bijdragen $u_{.1}$ van de verschillende meetreeksen tot de eerste hoofdcomponent (zie (15)). De singuliere waarde σ_1 behorend bij de eerste hoofdcomponent is rechts onder de figuur aangegeven. Van de eerste hoofdcomponent tot de totale kwadraten som van de data σ^2 is σ_1^2 . De fractie van σ^2 verklaard door de eerste hoofdcomponent, σ_1^2/σ^2 , is eveneens onder de figuur aangegeven (zie (11)). Om er nu achter te komen welk deel van het oorspronkelijke signaal 6 op het jaar 1933 wordt beschreven door de eerste hoofdcomponent, wordt de waarde van de genormeerde eerste hoofdcomponent in het jaar 1933 ($v_1, 1933$) vermenigvuldigd met σ_1 en met de waarde voor signaal 6 in de bovenste figuur ($u_{6,1}$). De waarden van signaal 6 in het jaar 1933 kan worden gereconstrueerd door hetzelfde te doen voor de overige hoofdcomponenten 2,...,12, en het resultaat te sommeren als in (15).

[1]

De bijdrage van HC1 tot de totale kwadratensom is 87%; voor de volgende vier hoofdcomponenten is dit percentage achtereenvolgens 3, 3, 3, 1. Het verloop in de tijd van HC1 vertoont een positieve lineaire trend, naast nog wat relatief kleine schommelingen (van dezelfde relatieve orde van grootte als in de oorspronkelijke reeksen). In HC1 blijkt dus met name de algemene zeespiegelstijging. De bijdrage u_{1i} van de verschillende reeksen y_i tot HC1 varieert tussen de +0.4 en +0.6; er blijkt geen verloop langs de kust. HC2 vertoont een maximum tussen 1900 en 1940; de bijdrage van de meetreeksen tot HC2 heeft een grillig verloop langs de kust. Voor HC3 geldt precies het omgekeerde: het verloop in de tijd is een zwakke slingering, met vrij sterke hoogfrequente fluctuaties; het verloop langs de kust echter is zeer regelmatig: van positief in Zeeland naar negatief in het Noorden. Dit wijst dus in de richting van een negatieve correlatie tussen noord en zuid in de snelle fluctuaties in de jaargemiddelde zeestand (tijdschaal 1 of enkele jaren).

[2]

In dit geval zijn de stations Brouwershaven en Hellevoetsluis weggelaten om een analyse over een langere periode te kunnen maken. De resultaten zijn niettemin vrijwel gelijk aan die onder [1]; De bijdragen van de eerste vijf hoofdcomponenten tot de totale kwadratensom zijn resp. 90, 3, 3, 1 en 1%.

[3]

In deze run is Terschelling weggelaten om tot 1872 terug te kunnen gaan. In grote lijnen zijn de resultaten weer als onder [1] en [2]; echter de positieve trend vanaf 1887 is niet terug te vinden in de periode 1872-1886. In een aantal meetreeksen blijkt in die periode een relatief hoge zeestand, met name in Vlissingen en IJmuiden. Dit verschijnsel blijkt ook sterk in HC2. De hoofdcomponenten dragen achtereenvolgens 83, 8, 3, 2, 2 en 1% van de totale kwadratensom bij.

2.2 Jaargemiddelde tijverschillen

Meetreeksen van het jaargemiddelde tijverschil zijn geanalyseerd op dezelfde manier als de zeestanden; data zijn afkomstig van 8 van de in 1.1 genoemde 10 stations: weggelaten zijn Den Helder en Harlingen, waar het tijverschil sterk is beïnvloed door de afsluiting van de Zuiderzee. Met de oorspronke-

lijke, op gemiddelde nul genormeerde meetreeksen is de analyse in drievoud uitgevoerd:

- [4] over 1887-1985 met de reeksen van Vlissingen, Zierikzee, Hoek van Holland, IJmuiden, Terschelling en Delfzijl;
- [5] over 1887-1964 met dezelfde reeksen als onder 1 plus Hellevoetsluis en Brouwershaven;
- [6] over 1872-1964 met dezelfde reeksen als onder 2 maar zonder Terschelling.

Resultaten:

[4]

In deze run zijn de meest recente data verwerkt; gegevens van Brouwershaven en Hellevoetsluis kunnen daarom (t.g.v. effect deltawerken) niet gebruikt worden. De bijdragen van de hoofdcomponenten tot de totale kwadratensom zijn achtereenvolgens 81, 10, 4, 3 en 2%. In het verloop in de tijd van HCl blijkt de periodiciteit van 18.6 jaar die in een aantal meetreeksen vrij dominant aanwezig is, en verder een positieve trend in de tweede helft van de reeks. Langs de kust gezien is de bijdrage aan HCl overal positief, maar in Hoek van Holland en IJmuiden aanmerkelijk kleiner dan in de overige stations. HC2 is vrij glad en vertoont een minimum rond 1960. De bijdrage van de meetreeksen tot HC2 heeft een sterk wisselend en onregelmatig verloop langs de kust. HC3 bestaat voornamelijk uit de sinusvorige component met periode 18.6 jaar; deze is voornamelijk bepaald door Vlissingen, waar het tijverschil (en waarschijnlijk daarom ook de sinusvormige component) veel groter is dan bij andere stations.

[5]

Deze run is over de relatief korte periode waarover van alle 8 stations data beschikbaar zijn. Over deze periode vertoont het tijverschil op de verschillende stations een vrij lineair verloop, in vergelijking met de perioden ervoor en erna. De bijdragen van de hoofdcomponenten tot de totale kwadratensom zijn achtereenvolgens 65, 20, 5, 3 en 2%. Wat onder [4] al opviel blijkt nu veel sterker: van de tijverschil gegevens is de totale kwadratensom veel sterker uitgesmeerd over de hoofdcomponenten dan het geval was bij de zeestanden. Dit is met name het gevolg van de sinusvormige component met een periode van 18.6 jaar. HCl vertoont deze periodieke component, en daarnaast een rede-

lijk lineaire trend. De bijdrage van de meetreeksen tot HC1 verloopt van Vlissingen tot Hoek van Holland van sterk positief tot nul; verder naar het Noorden is de bijdrage ook positief, maar in Terschelling aanzienlijk groter dan in Delfzijl. Ook het tijdsverloop van HC2 vertoont een positieve trend naast nog wat fluctuaties. De bijdrage van Terschelling tot HC2 is positief (daar is de trend in het getijverschil dan ook verreweg het grootst) en de bijdrage tot HC2 van de overige stations is negatief of vrijwel nul. HC3 en HC4 tonen een contrast van Vlissingen, resp. Hellevoetsluis en Delfzijl met de overige stations. Deze lijken moeilijk te interpreteren.

[6]

Deze run is vanaf 1872 tot 1964, dus wat verder terug in de tijd dan [4] en [5]. De bijdragen van de hoofdcomponenten tot de totale kwadratensom zijn achtereenvolgens 62, 18, 8, 4 en 3%. HC1 en HC2 lijken op die van run [5]. Het afwijkende verloop van een aantal meetreeksen voor 1880 is met name in HC3 en HC4 terug te vinden.

Wat betreft de tijverschillen blijkt dat de kwadratensom nogal wordt uitgesmeerd over een aantal hoofdcomponenten, wat interpretatie niet gemakkelijker maakt. Met name de periodieke component en de lineaire trend blijken niet uit elkaar gehaald te worden. Omdat de periodieke component vrij goed te beschrijven is, is de bovenstaande analyse nog eens herhaald op data waaruit deze component vooraf verwijderd is. De gebruikte parameters (periode, fase, amplitude) van de sinus in iedere meetreeks zijn overgenomen uit gegevens van RWS. De met [4],[5] en [6] overeenkomende runs zijn aangegeven met [4'],[5'] en [6'].

[4']

De bijdragen van de hoofdcomponenten tot de totale kwadratensom zijn achtereenvolgens 80, 12, 3, 2 en 2%. Van de periodieke component is in de hoofdcomponenten nauwelijks meer iets terug te vinden. Verder lijken de trend in HC1 en het verloop van de bijdrage aan HC1 langs de kust sprekend op [4]. HC2 van [4'] is zelfs in alle opzichten gelijk aan HC2 van [4], en HC3 van [4'] is vrijwel gelijk aan HC4 van [4] (HC3 van [4] was volledig bepaald door de periodieke component, die is er nu uit).

[5']

De bijdragen van de hoofdcomponenten tot de totale kwadratensom zijn achtereenvolgens 59, 20, 6, 5 en 4%. In vergelijking met [5] lijkt de trend nu

sterker HC1 te bepalen: met name wordt de trend bij Hoek van Holland nu negatief, wat ook is waargenomen. In HC2 blijkt nu minder van de trend dan in [5] het geval was. Verwijderen van de periodieke component heeft dus duidelijk een gunstige invloed op de analyse van de trends in de hoofdcomponenten.

[6']

De bijdragen van de hoofdcomponenten tot de totale kwadraten-som zijn achtereenvolgens 42, 28, 11, 7 en 5%. Ook in dit geval is het effect van de trend duidelijker in HC1 geconcentreerd. Over deze periode (1872-1964) blijkt zowel bij Hoek van Holland als bij IJmuiden een negatieve trend.

2.3 Stabiliteit van de hoofdcomponenten-ontbinding

Stabiliteit van hoofdcomponenten uit jaargemiddelde zeestanden en jaargemiddelde tijverschillen is onderzocht volgens de methode beschreven in 1.4.

[7]

Een test is uitgevoerd op jaargemiddelde zeestanden, namelijk op alle tien in 2.1 genoemde stations, over de twee deelperioden 1887-1928 (deelperiode 1) en 1929-1969 (deelperiode 2). De weegfactoren (bijdrage van de verschillende stations tot HC1) $u_{1.1}$ en $u_{2.1}$ van resp. deelperiode 1 en 2 vallen vrijwel samen; ook HC1 van de periode 1 en 2, $h_{1.1}$ en $h_{2.1}$, vallen vrijwel samen met resp. $w_{1.1}$ en $w_{2.1}$, de eerste "hoofdcomponenten" van deelperiode 1 en 2 berekend na verwisseling van $u_{1.1}$ en $u_{2.1}$ (zie 1.4) (in de figuren zijn overigens de genormeerde hoofdcomponenten geplot). De overeenkomst neemt aanzienlijk af wanneer we HC2 en volgenden bekijken.

Met de tijverschillen zijn twee tests uitgevoerd met de 8 stations Vlissingen, Zierikzee, Brouwershaven, Hellevoetsluis, Hoek van Holland, IJmuiden, Terschelling, en Delfzijl, over de twee deelperioden 1887-1925 en 1926-1964, zowel [8] op de oorspronkelijke meetgegevens; als [9] op de gegevens waaruit de sinusvormige component met periode 18.6 aar is verwijderd.

Alleen de weegfactoren $u_{1.p}$ en $u_{2.p}$ (bijdrage van de verschillende stations aan HCp) zullen worden besproken; het effect van op de hoofdcomponenten is min of meer hetzelfde.

[8]

Test met de oorspronkelijke gegevens: $u1.1$ en $u2.1$ (bijdrage van de verschillende stations tot HCl) blijken vrijwel gelijk, behalve voor Terschelling; de overeenkomst neemt sterk af bij volgende hoofdcomponenten. Interessant is deze analyse over twee aparte deelperioden de resultaten te vergelijken met de analyse over de volledige periode 1887-1964 (zie 2.2 [5]): bij opsplitsing in 2 aparte deelperioden blijkt de periodieke component (18.6 jaar) vooral in HCl terecht te komen, en de trend in HC2. Deze scheiding tussen periodieke component en trend bleek niet op te treden in een analyse over de hele periode 1887-1964: daar was HCl sterk bepaald door zowel de periodieke component als de trend. Het feit dat in de korte deelperioden HCl sterker bepaald wordt door de periodieke component en minder sterk door de trend komt doordat de kwadraten van de data t.g.v. een trend relatief sterk toeneemt bij toenemende reekslengte N (nl. als de derde macht van N), in vergelijking met de kwadraten van de data t.g.v. een sinusvormige component (deze neemt ongeveer lineair toe met N). Dit betekent dat de trend sterker HCl bepaalt naarmate de reeksen langer zijn. Het gevonden verschil in de resultaten van een HCA van een volledige reeks en van 2 deelreeksen wijst op het belang van het analyseren van deelreeksen.

[9]

Test met de gegevens waaruit sinusvormige component is verwijderd: $u1.1$ en $u2.1$ verschillen aanzienlijk, met name voor Terschelling. De periodieke component is dus de "stabiele" component in de oorspronkelijke data (over langere tijd ongeveer dezelfde bijdrage in de reeksen van de afzonderlijke stations); uit de resultaten van de HCA van gegevens waaruit de sinusvormige component verwijderd is over de periode 1887-1964 blijkt al dat de eerste hoofdcomponent een trend is met in de tijd toenemende helling, dus instabiliteit is in dit geval te verwachten. Ook bij volgende hoofdcomponenten HC p ; $p=2, \dots, 4$ is het verschil tussen $u1.p$ en $u2.p$ aanzienlijk.

3. Summary and conclusions

Principal component analysis (PCA, also called "eigen analysis") of a set of M time-series, sampled at the same N instants or over the same N intervals, is a representation of these data of the form

$$y_{ij} = \sum_{p=1}^M h_{pj} u_{ip} \quad (S.1)$$

in which y_{ij} is the j-th sample of the i-th time-series y_i .; h_{pj} is the j-th sample of the p-th principal component h_p .; and u_{ip} is the contribution of the p-th principal component to the i-th time-series y_i ., normalized to

$$\sum_{i=1}^M u_{ip}^2 = 1 \quad \text{for } p=1, \dots, M. \quad (S.2)$$

The decomposition (S.1) is chosen in such a way that the data are approximated as close as possible (in the least-squares sense) by the first m terms of the summation on the right-hand side of (S.1); the first terms are the most important in terms of percentage of the total variance of the data described.

Usually, the series average is subtracted from each of the time-series before a PCA is performed; sometimes they are also scaled to have unit variance (in this context, variance will simply mean sum of squares of the deviations from the series average divided by the number of samples). The purposes of PCA are extraction of information from a set of time-series in an efficient way, and data-reduction, by truncation of the summation in (S.1).

PCA has been applied to series of yearly mean sea level and tidal range, measured at up to 10 stations along the Dutch coast, over a period of up to 99 year. From South-West to North-East, the stations are (the numbers correspond to the numbers on the "signal" axis in the figures):

1. Vlissingen.
2. Zierikzee.
3. Brouwershaven.
4. Hellevoetsluis.
5. Hoek van Holland.
6. IJmuiden.
7. Den Helder.

8. West-Terschelling.
9. Harlingen.
10. Delfzijl.

A number of trials have been made with different subsets of the stations over different periods of time, and comparisons have been made. Prior to decomposition, the original series have been adjusted to obtain a zero average; scaling to obtain time-series of equal standard deviation has not been applied because this would cause amplification of noise (fluctuations specific to individual stations) in the series of small total variance (e.g. small trend).

The figures in this report show some results; the first plot on each page displays the contributions u_{ip} of the p -th principle component to the i -th time series (signal) (see (s.1)), the second plot shows the p -th principle component h_p , normed to unit sum of squares (to obtain h_p , the values in this figure should be multiplied with the number σ_p written under the figure).

From the analyses of yearly mean sea level data (see figures 1-3), an approximately linear positive trend appears as the most significant feature of the first principal component, which accounts for about 90 percent of the total variance of the data; its contribution to the original series is fairly uniform and positive. Another interesting component is a high-frequent ("noisy") signal; its contribution to the series of stations along the Dutch coast decreases smoothly from positive in the South to negative in the North.

In the analyses of records of yearly mean tidal range (see figures 4-6), the first principal component shows a positive trend as well as a sinusoidal fluctuation (period 18.6 years); it accounts for 65-80 percent of the total variance of the data (depending on stations and selected period); its contribution to the original series is positive or zero, but varies considerably. Both trend and sinusoidal fluctuation appear also in other principal components; this illustrates the fact that principal components are determined by the data itself, and consequently do not necessarily correspond to signal modes considered by the analyst. A better separation of trend and sinusoidal fluctuation is obtained when the series are split into two subseries, which are analyzed separately. The same analyses of yearly mean tidal range have been performed after removal of the sinusoidal fluctuation from the original series (see figures 7-9). Now the first principal component shows a positive trend with slope increasing in time.

In general, the rather small number of stations (ten or less) has the effect that some of the principal components are dominated by specific features of one of the time-series.

Stability of the principal component decomposition has been checked by comparison of the results of PCA applied on subseries. The first principal component of yearly mean sea level and tidal range appear to be quite stable, as opposed to the other principal components; if the sinusoidal fluctuation has been removed from the records of yearly mean tidal range before analysis, all components appear unstable.

LITERATUUR

Wax, M. and Kailath, T.

Detection of signals by information theoretic criteria

IEEE ASSP-33, no2., 1985.

Nyberg, L.

Sea level forecasts with an EOF-model

Symposium on North Sea dynamics, Hamburg, 31-8-1981 - 04-09-1981.

Golub, G.H. and Reinsch, C.

Singular value decomposition and least squares solutions.

Nume. Math. 14, 403-420 (1970) (ook in Wilkinson, J.H. and Reinsch, C.,

Handbook of automatic computation, volume II: Linear algebra

Springer-Verlag, 1971.

Noble, B. and J.W. Daniel

Applied linear algebra

Prentice-Hall Inc. 1977.

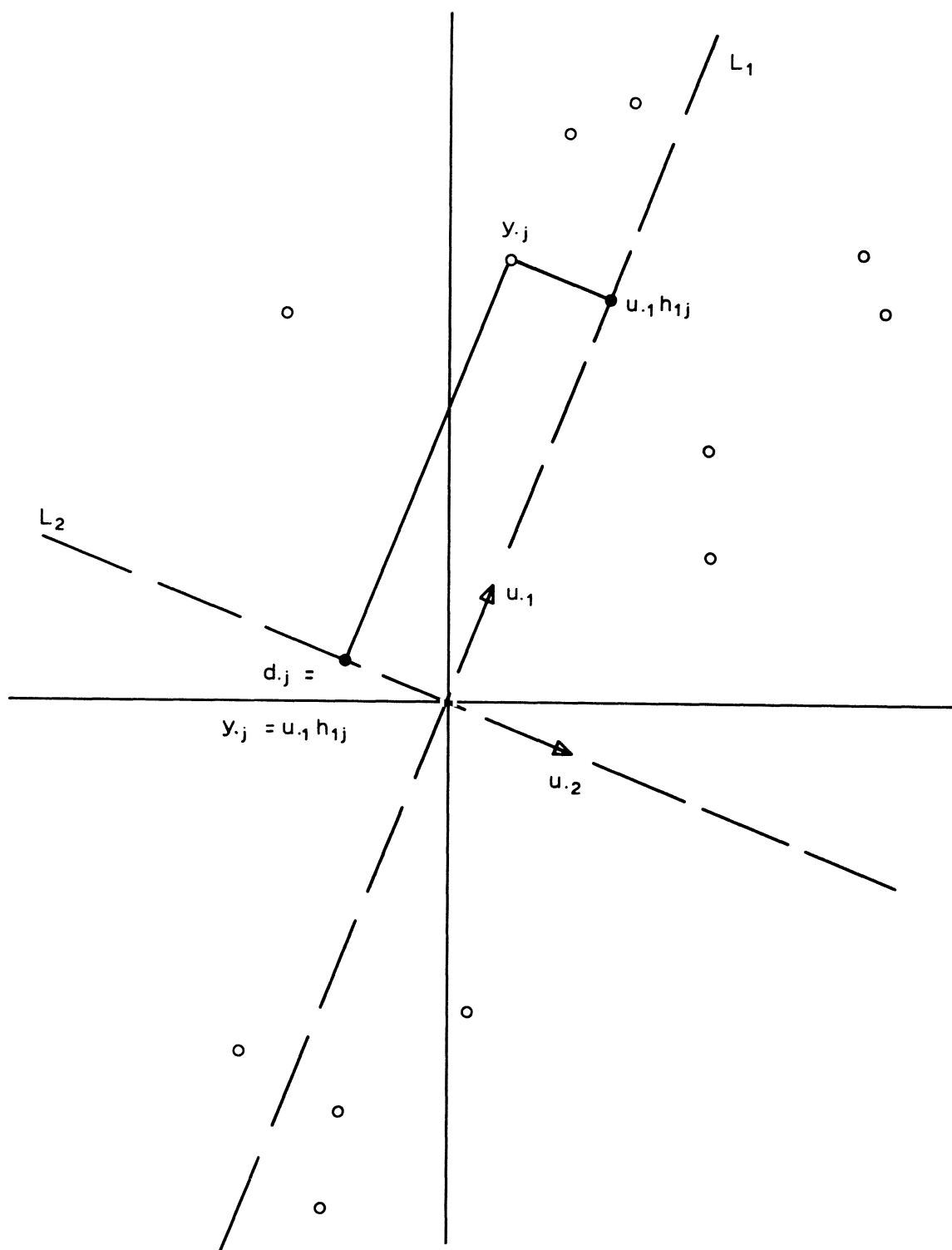
VERKLARING VAN DE FIGUREN

Code voor verwijzing naar figuren:

[3] verwijst naar run [3] in de text "eigenfunction 2" duidt de tweede hoofdcomponent aan, genormeerd op kwadratensom 1. (ofwel $v_{2.}$ in 1.1); "contribution of EF 2" (eigenfunction 2) duidt de bijdragen $u_{.2}$ van van de tweede hoofdcomponent $h_{2.} = \sigma_2 v_{2.}$ tot de verschillende signalen aan.

De nummers op de "signal"-as komen als volgt met de stations overeen:

- 1 = Vlissingen
- 2 = Zierikzee
- 3 = Brouwershaven
- 4 = Hellevoetsluis
- 5 = Hoek van Holland
- 6 = IJmuiden
- 7 = Den Helder
- 8 = Terschelling
- 9 = Harlingen
- 10 = Delfzijl.

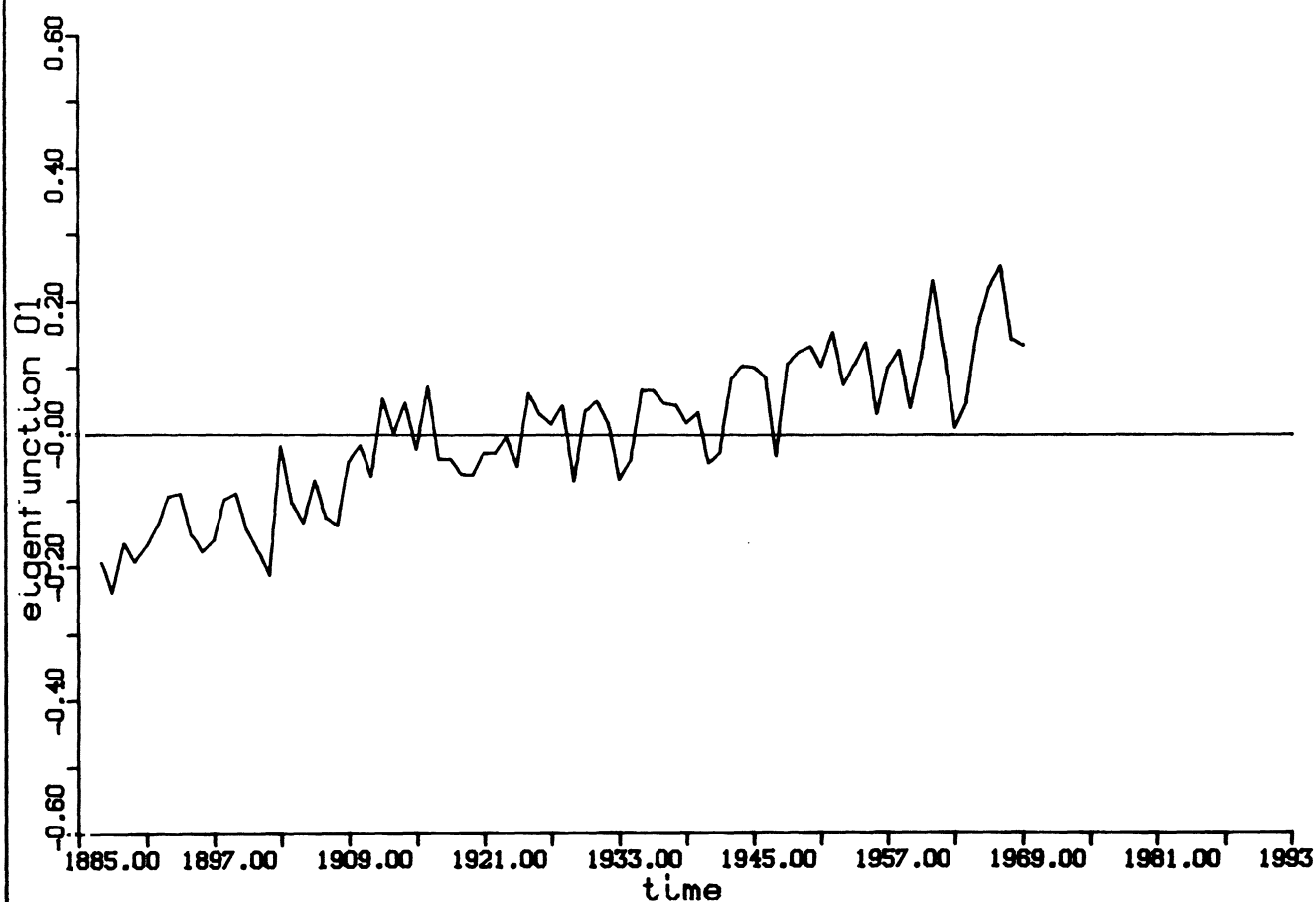
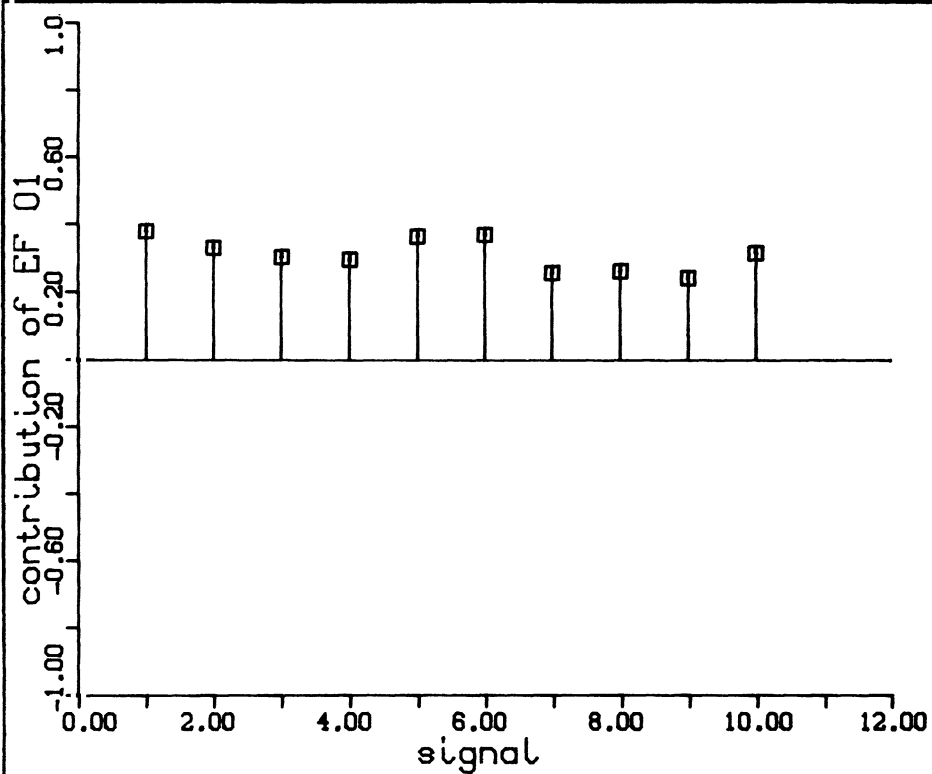


GEOMETRISCHE VOORSTELLING VAN HOOFD-
COMPONENTEN - ANALYSE (ZIE 1.1)

DELFT HYDRAULICS

H 535

FIG. 0



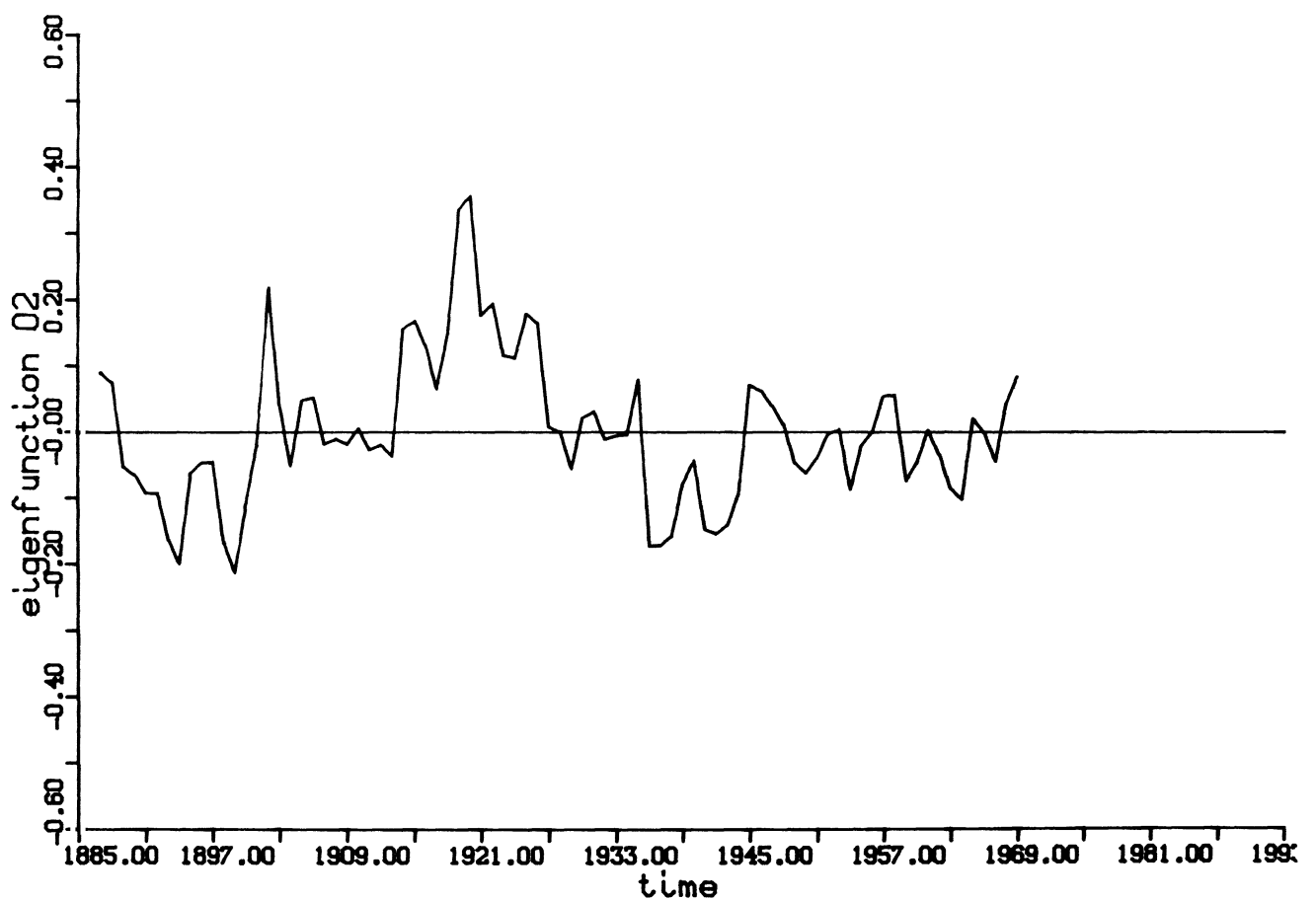
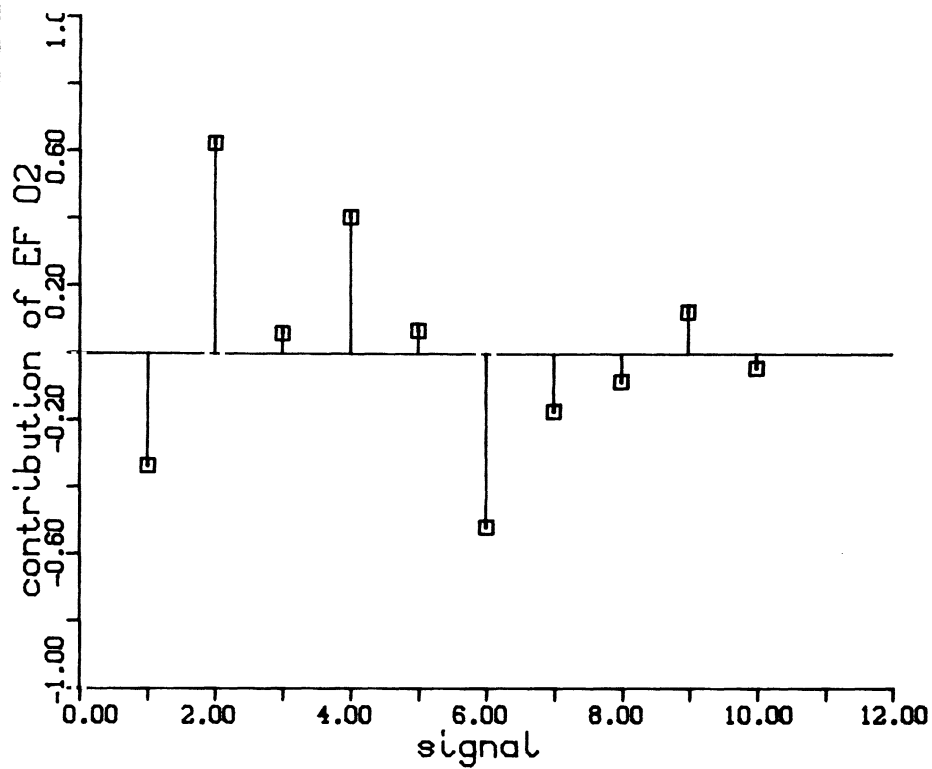
HOOFDCOMPONENT 1

$\sigma_1 = 153.8$

$\sigma_1^2 / \sigma^2 = 0.872$

Delft Hydraulics

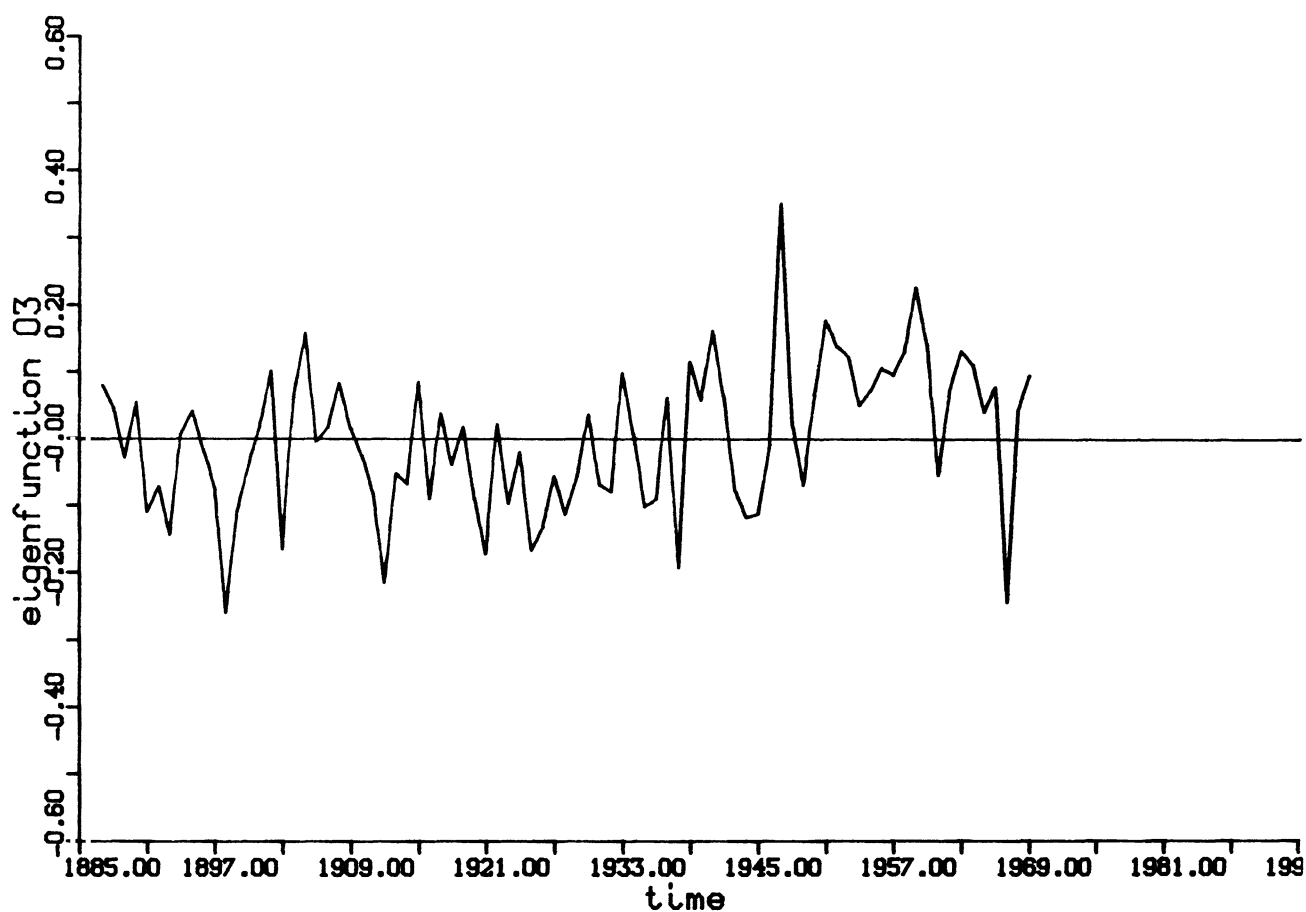
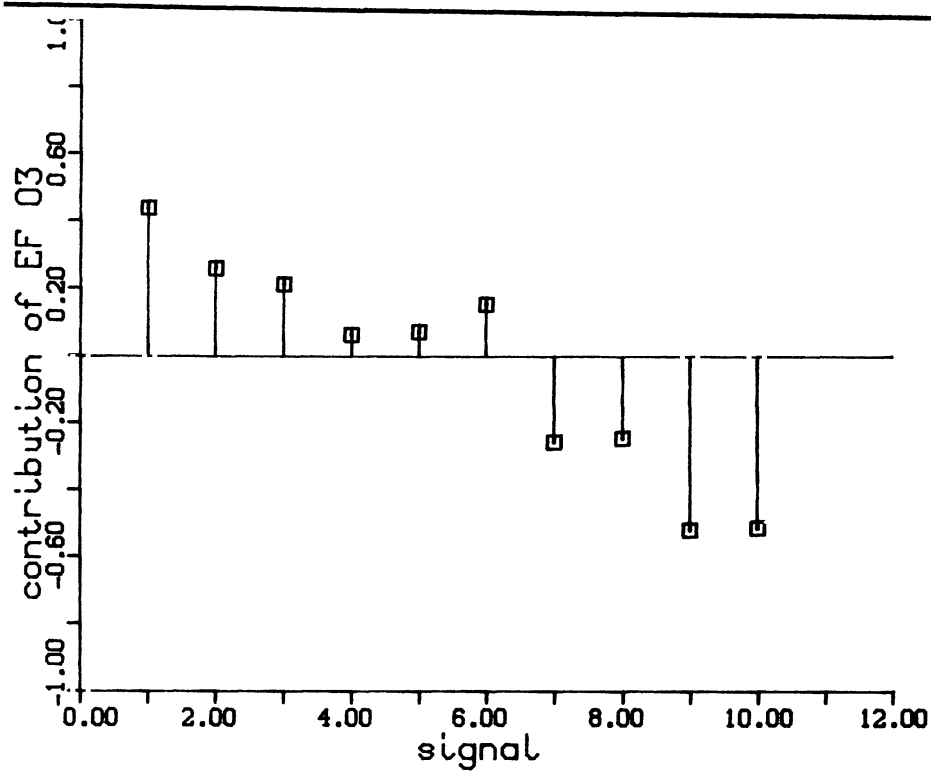
Fig. [1]



HOOFDCOMPONENT 2

$\sigma_2 = 30.0$

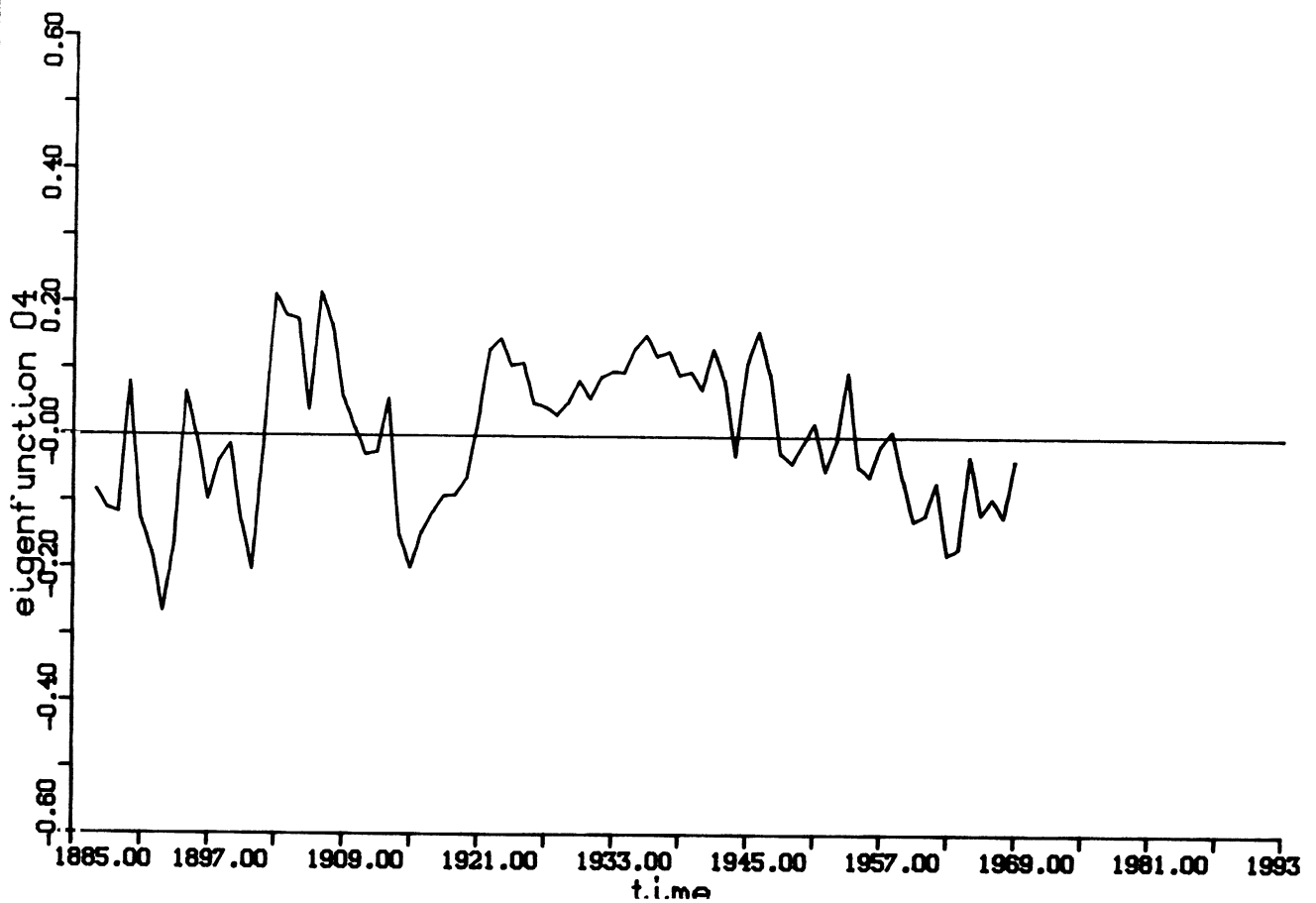
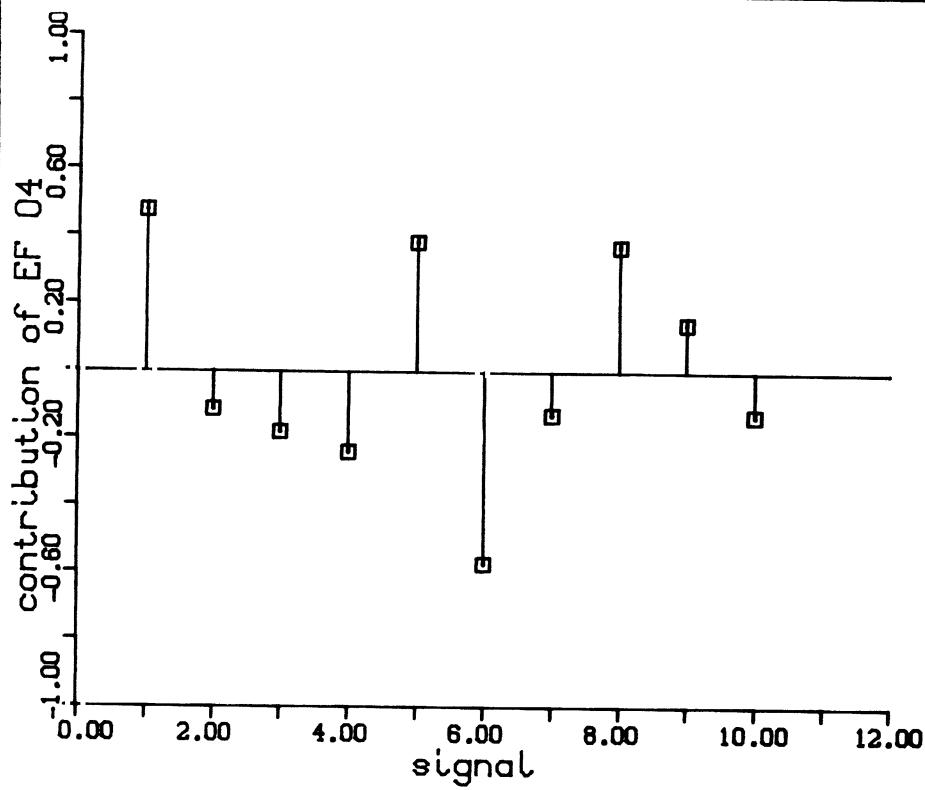
$\sigma_2^2 / \sigma^2 = 0.033$



HOOFDCOMPONENT 3

$$\sigma_3 = 26.9$$

$$\sigma_3^2 / \sigma^2 = 0.027$$



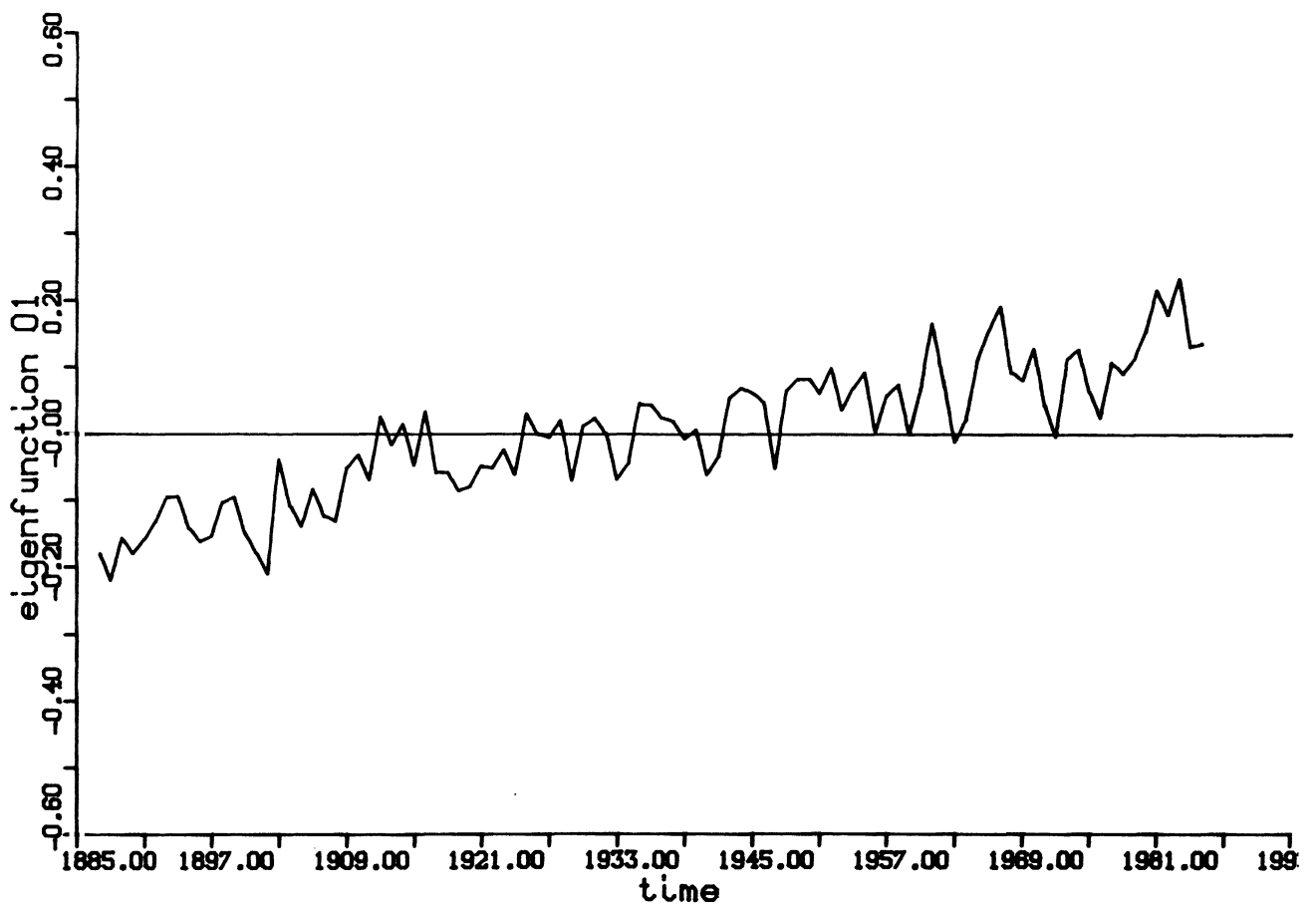
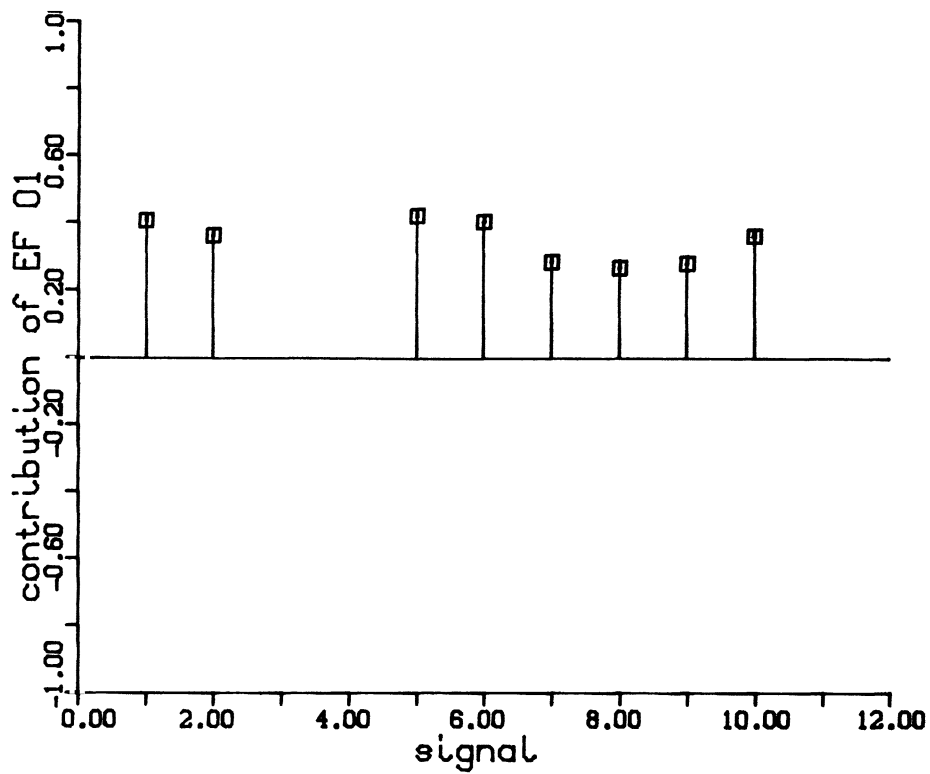
HOOFDCOMPONENT 4

$$\sigma_4 = 21.9$$

$$\sigma_4^2 / \sigma^2 = 0.018$$

Delft Hydraulics

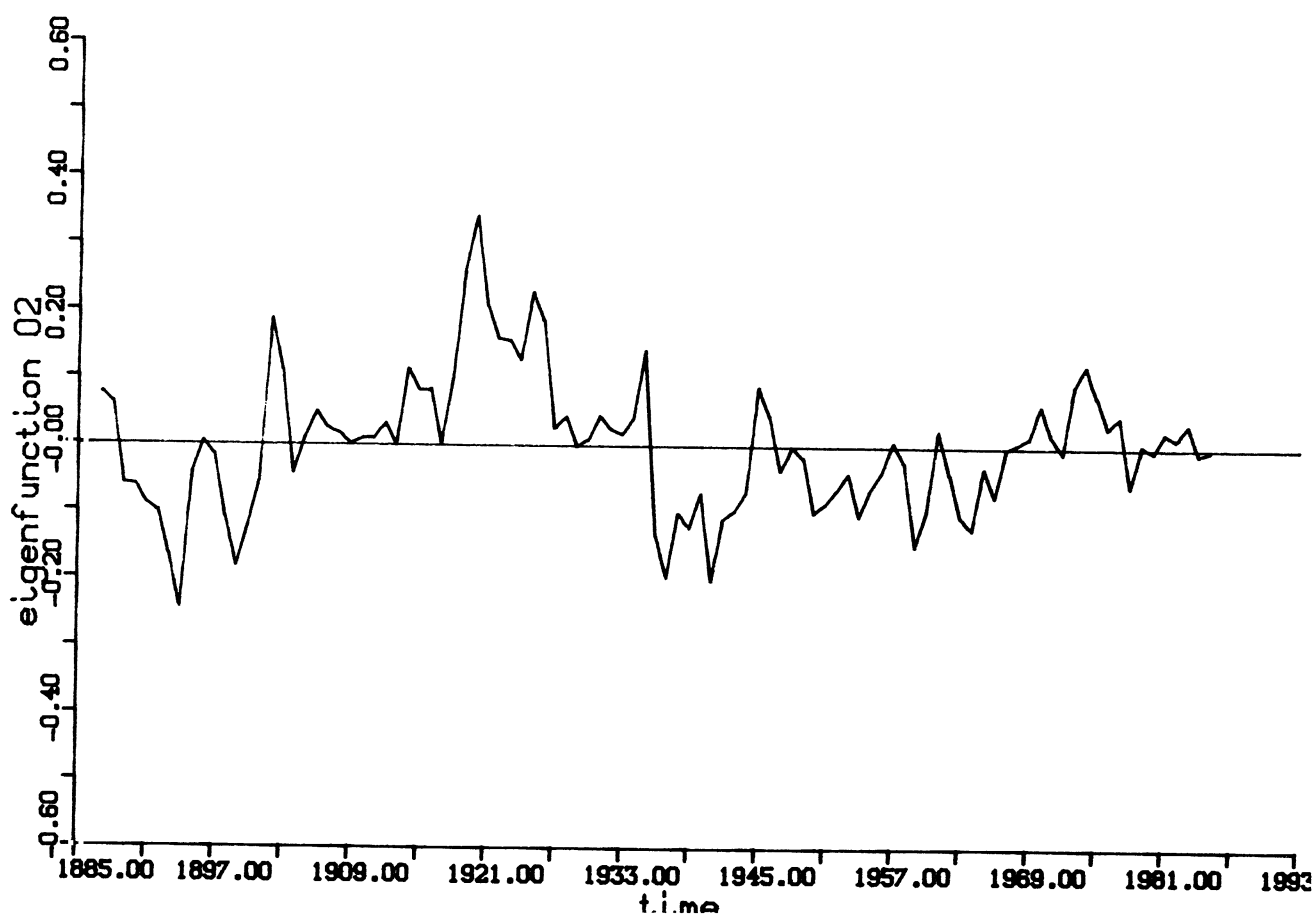
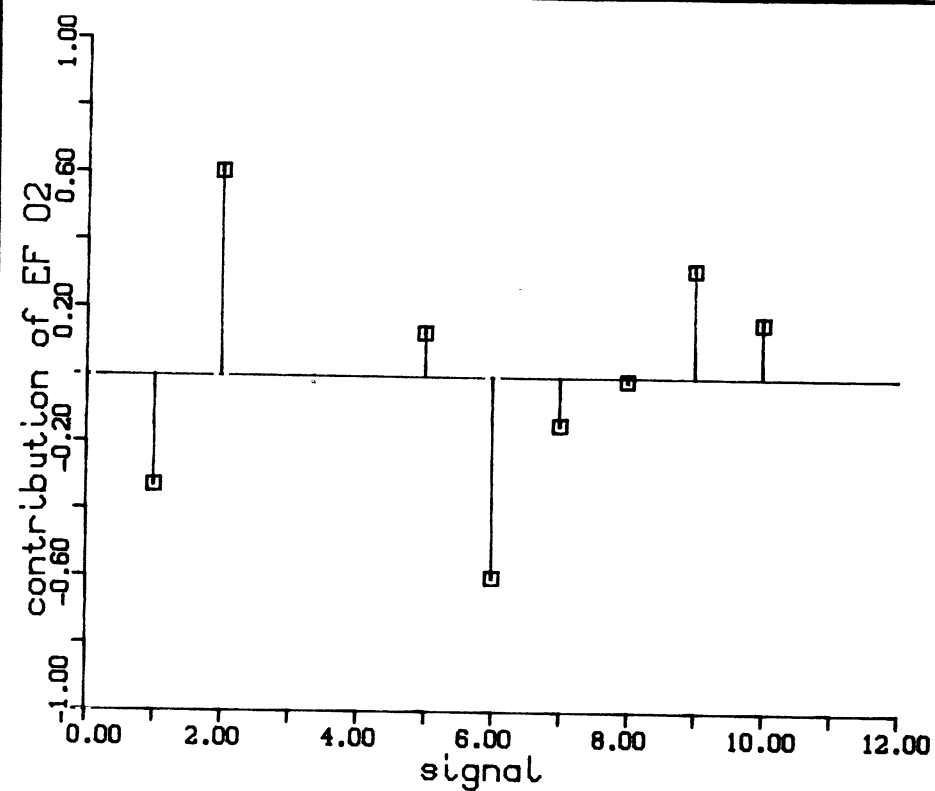
Fig. [1]



HOOFDCOMPONENT 1

$$\sigma_1 = 169.0$$

$$\sigma_1^2 / \sigma^2 = 0.899$$



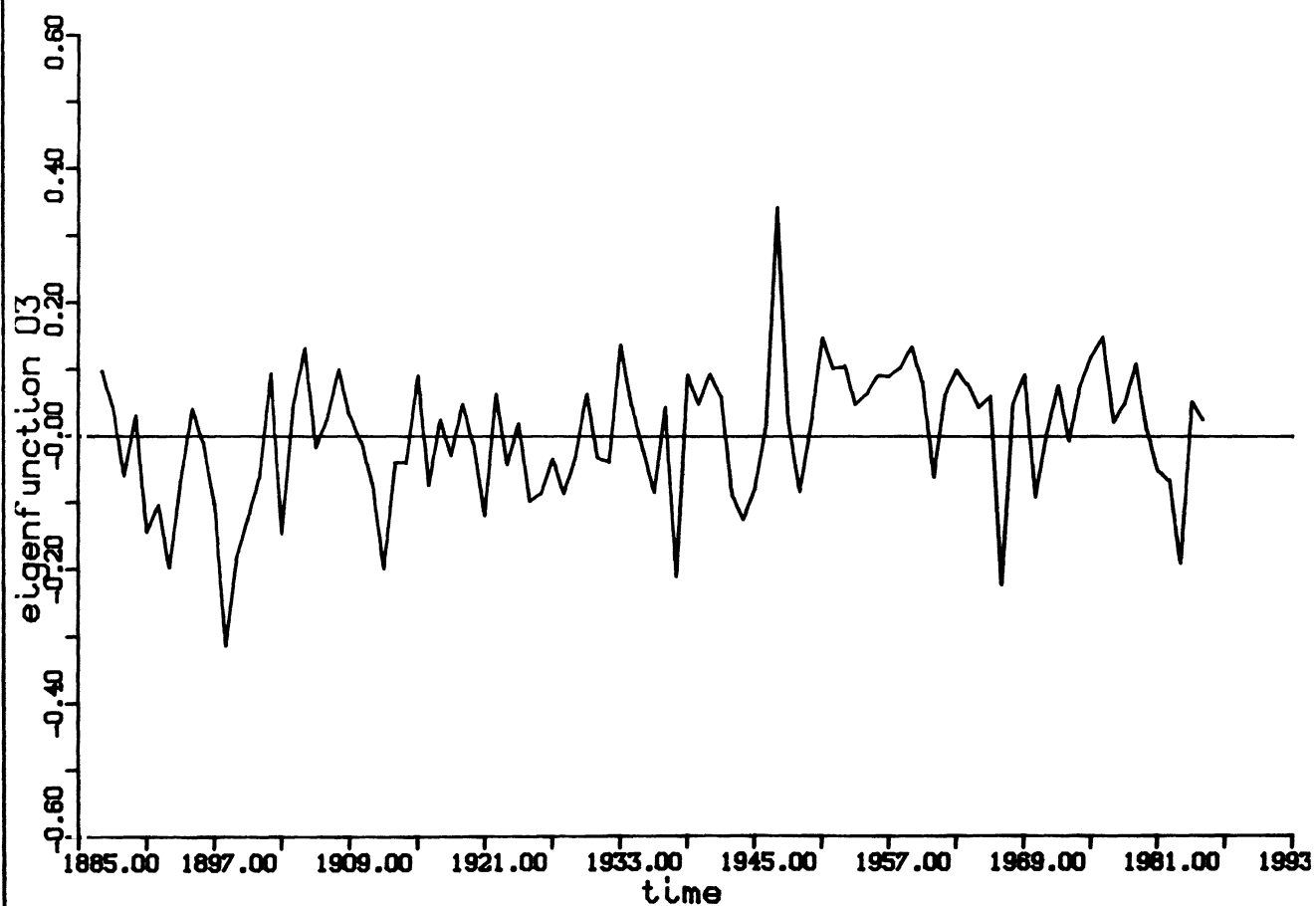
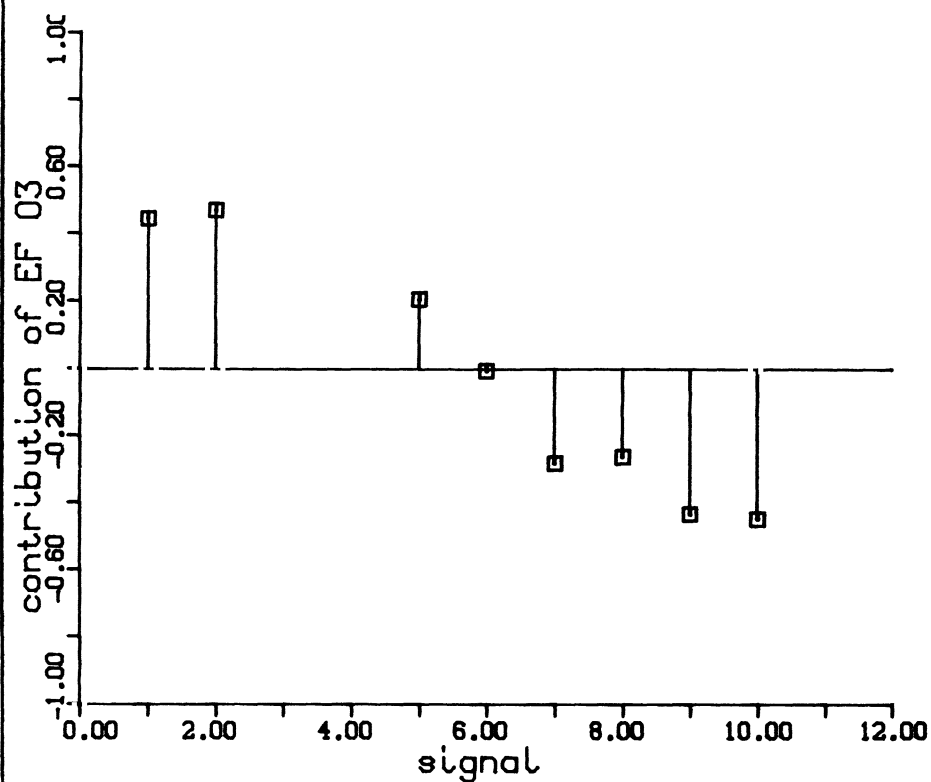
HOOFDCOMPONENT 2

$\sigma_2 = 28.7$

$\sigma_2^2 / \sigma^2 = 0.026$

Delft Hydraulics

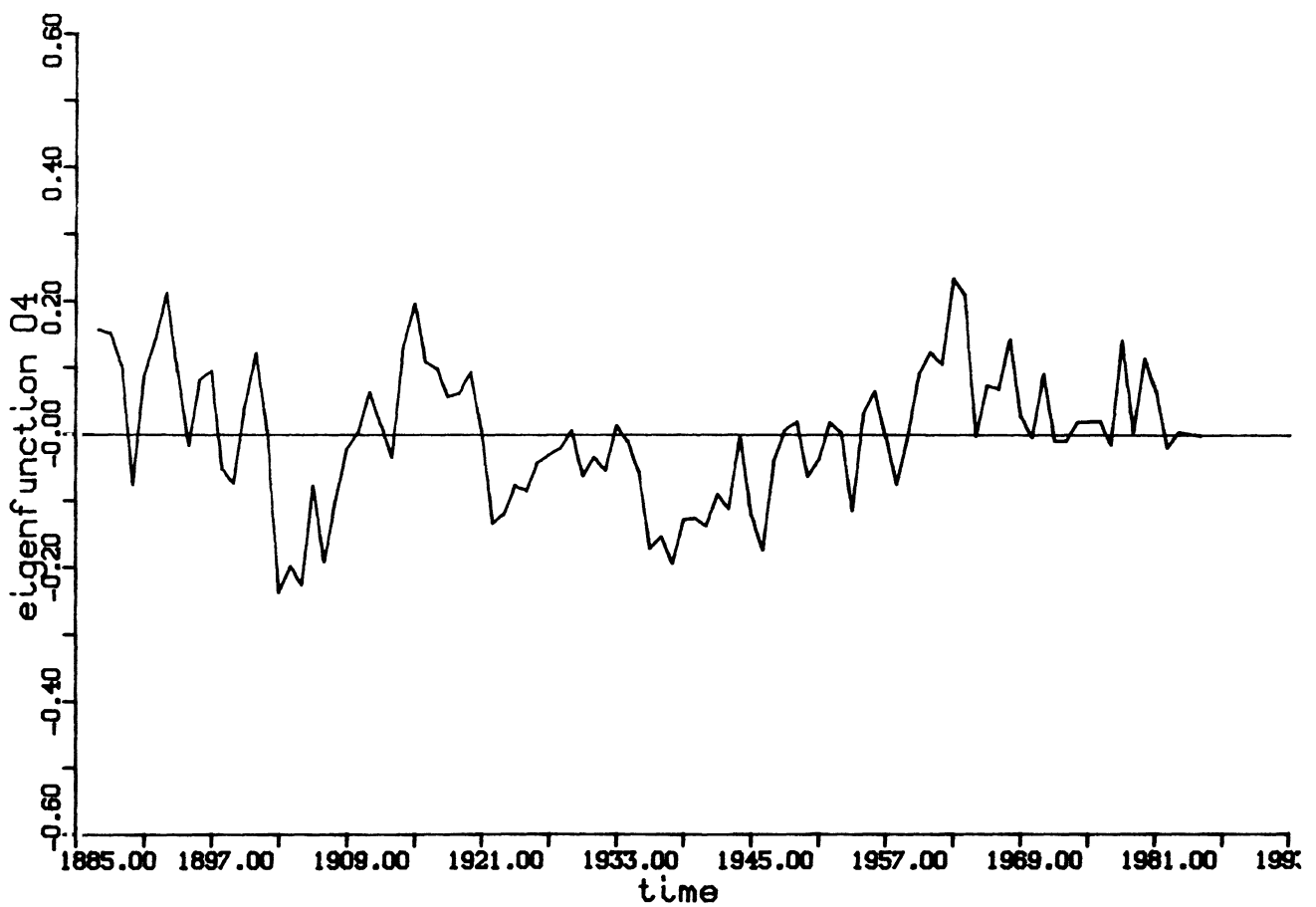
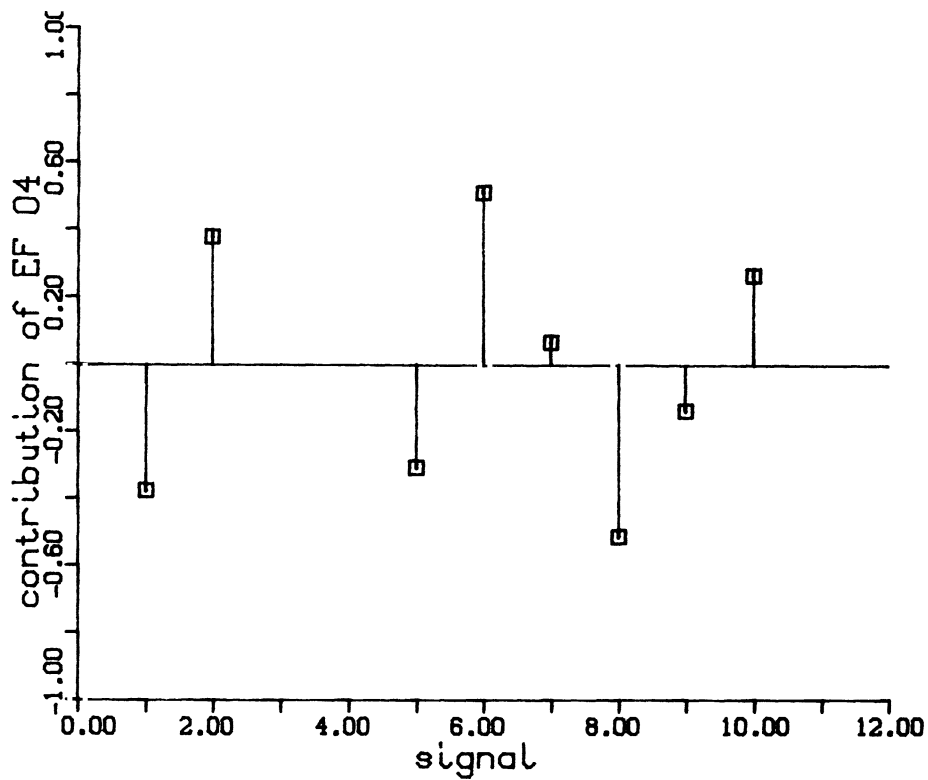
Fig. [2]



HOOFDCOMPONENT 3

$$\sigma_3 = 28.1$$

$$\sigma_3^2 / \sigma^2 = 0.025$$



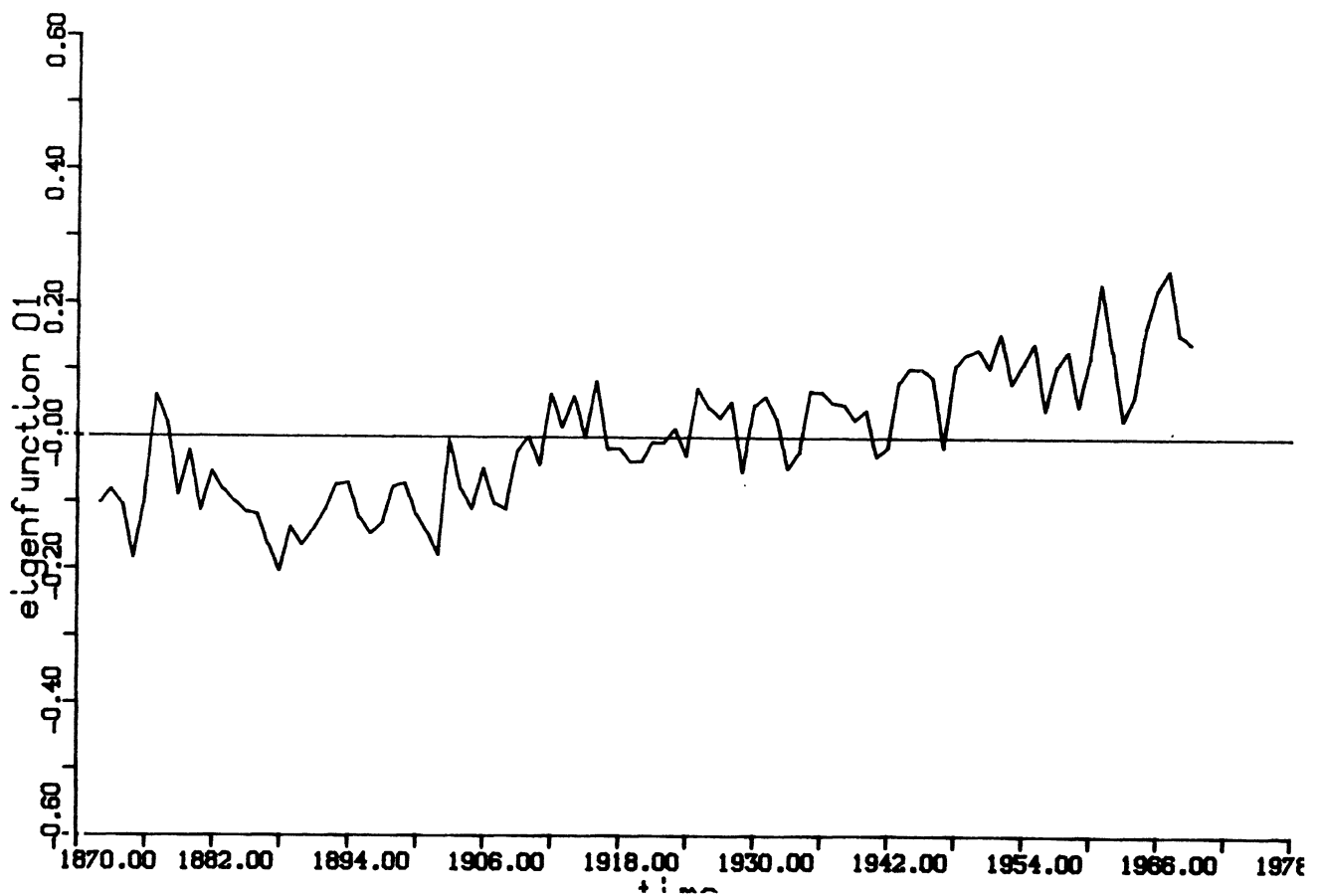
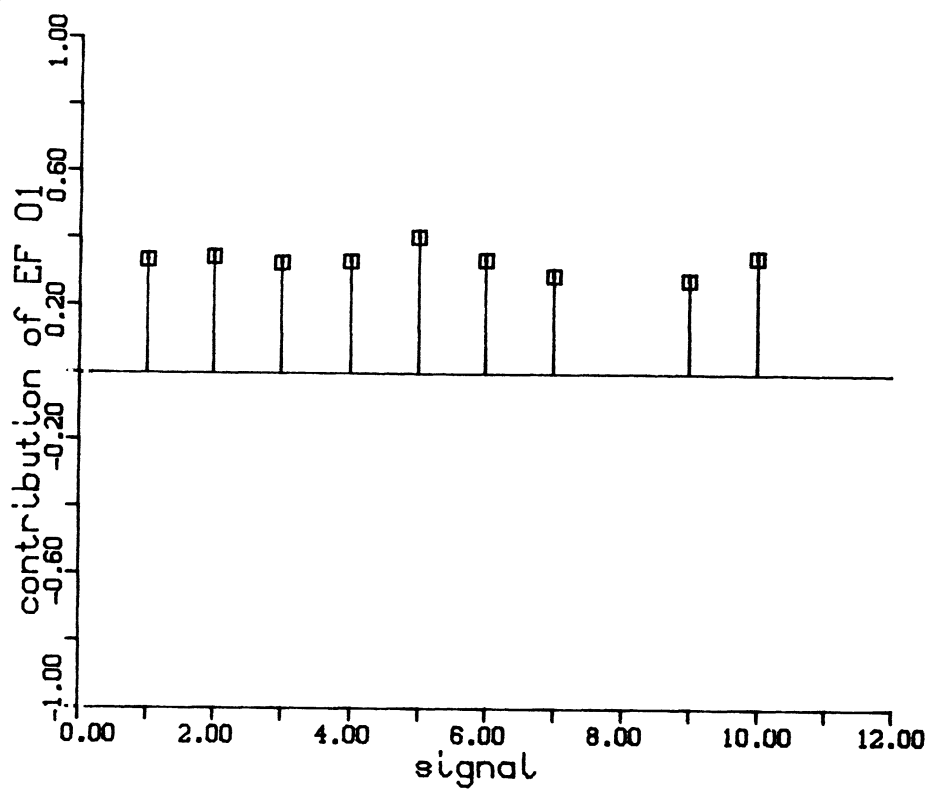
HOOFDCOMPONENT 4

$$\sigma_4 = 21.7$$

$$\sigma_4^2 / \sigma^2 = 0.015$$

Delft Hydraulics

Fig. [2]



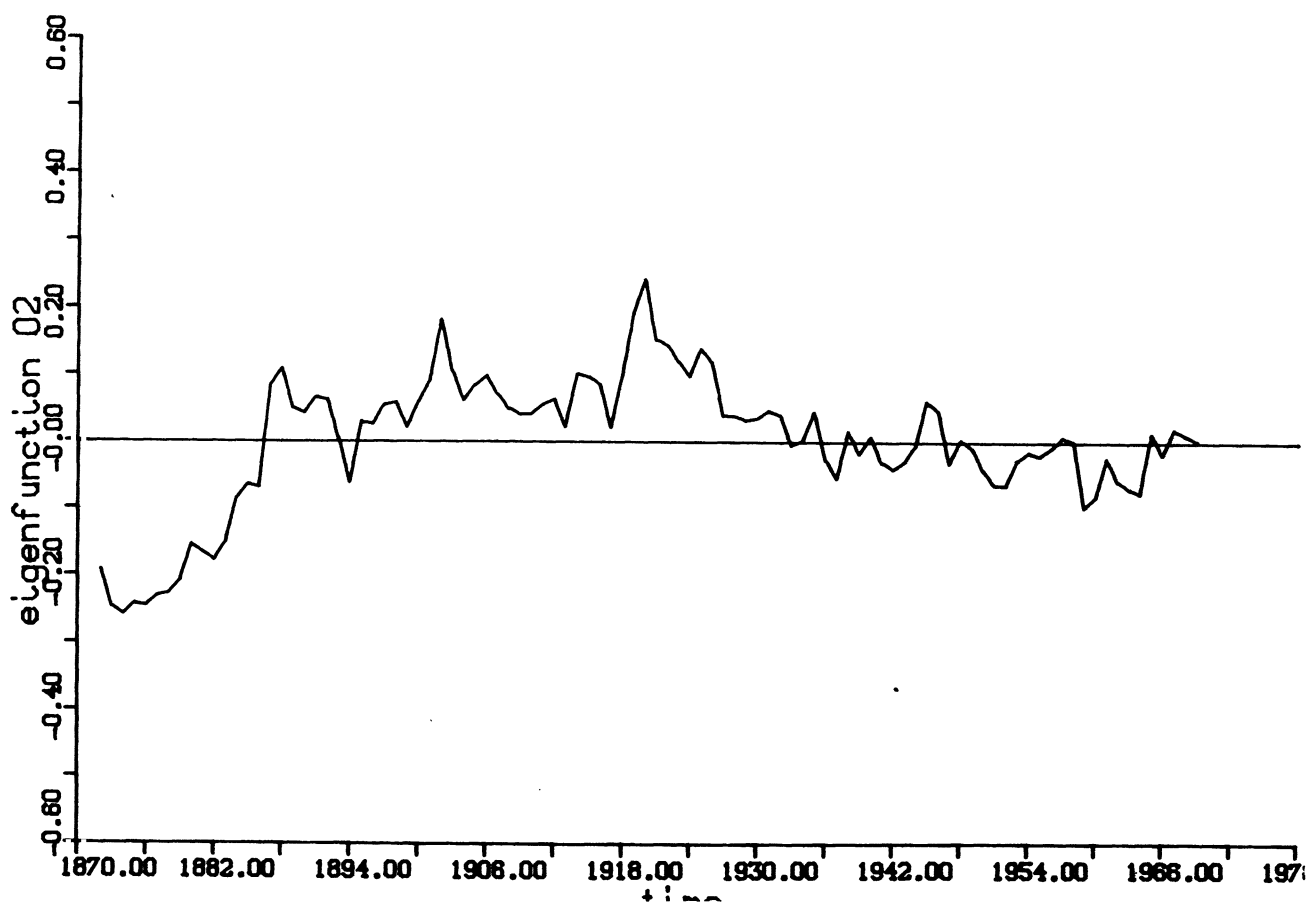
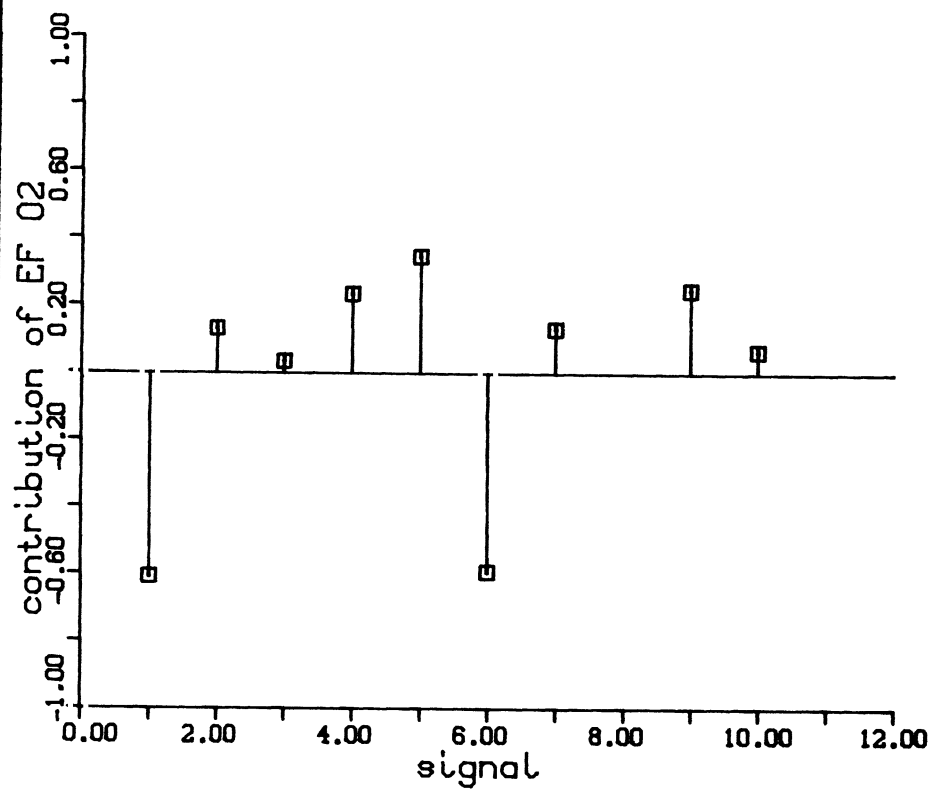
HOOFDCOMPONENT 1

$\sigma_1 = 161.1$

$\sigma_1^2/\sigma^2 = 0.825$

Delft Hydraulics

Fig. [3]



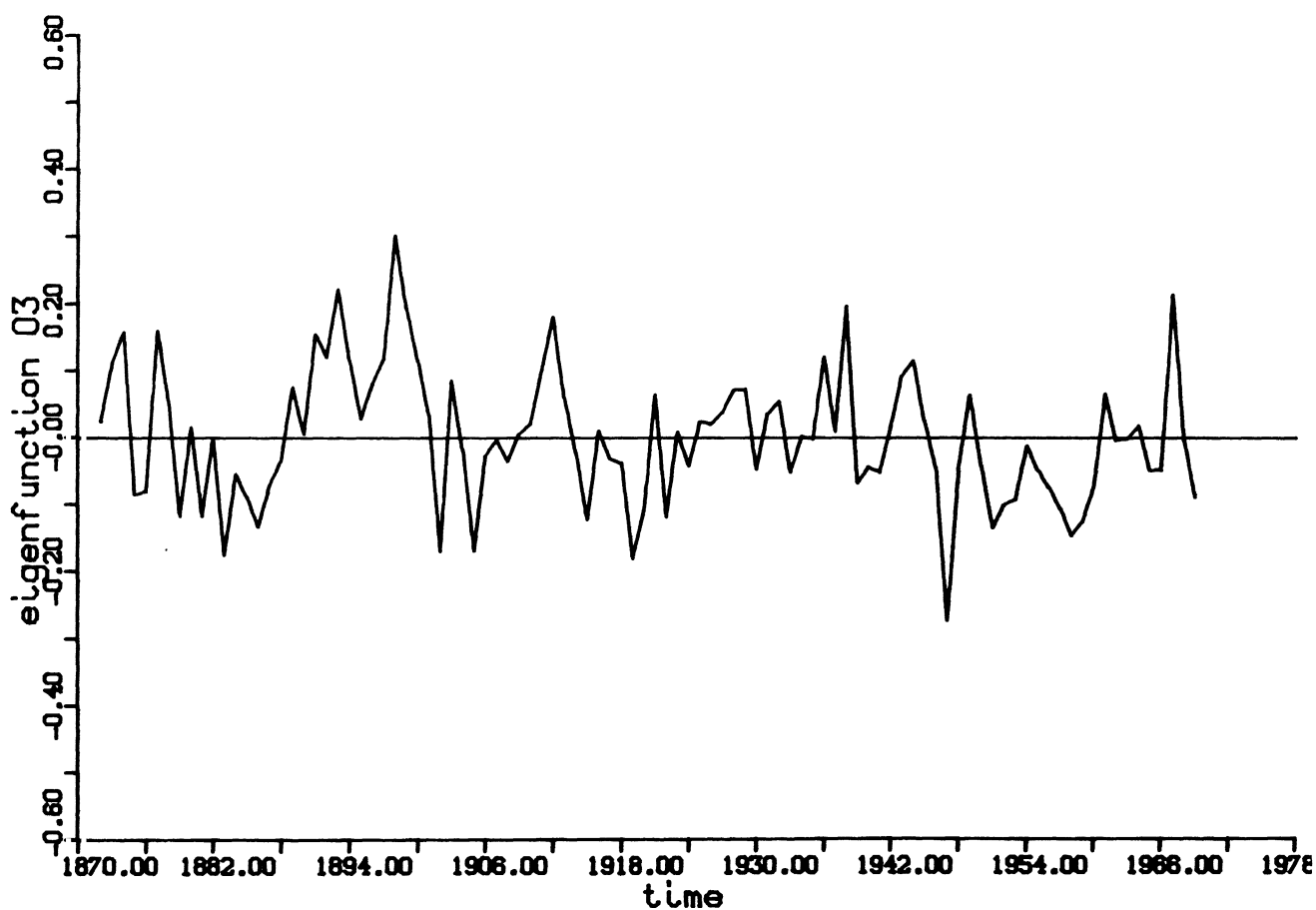
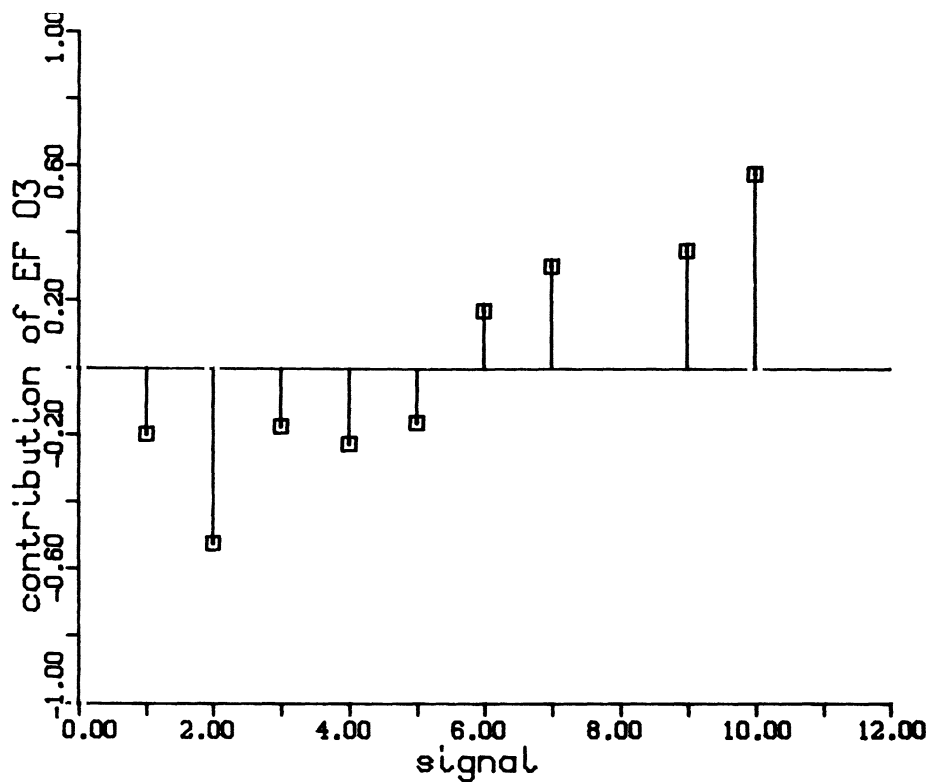
HOOFDCOMPONENT 2

$$\sigma_2 = 49.7$$

$$\sigma_2^2 / \sigma^2 = 0.079$$

Delft Hydraulics

Fig. [3]



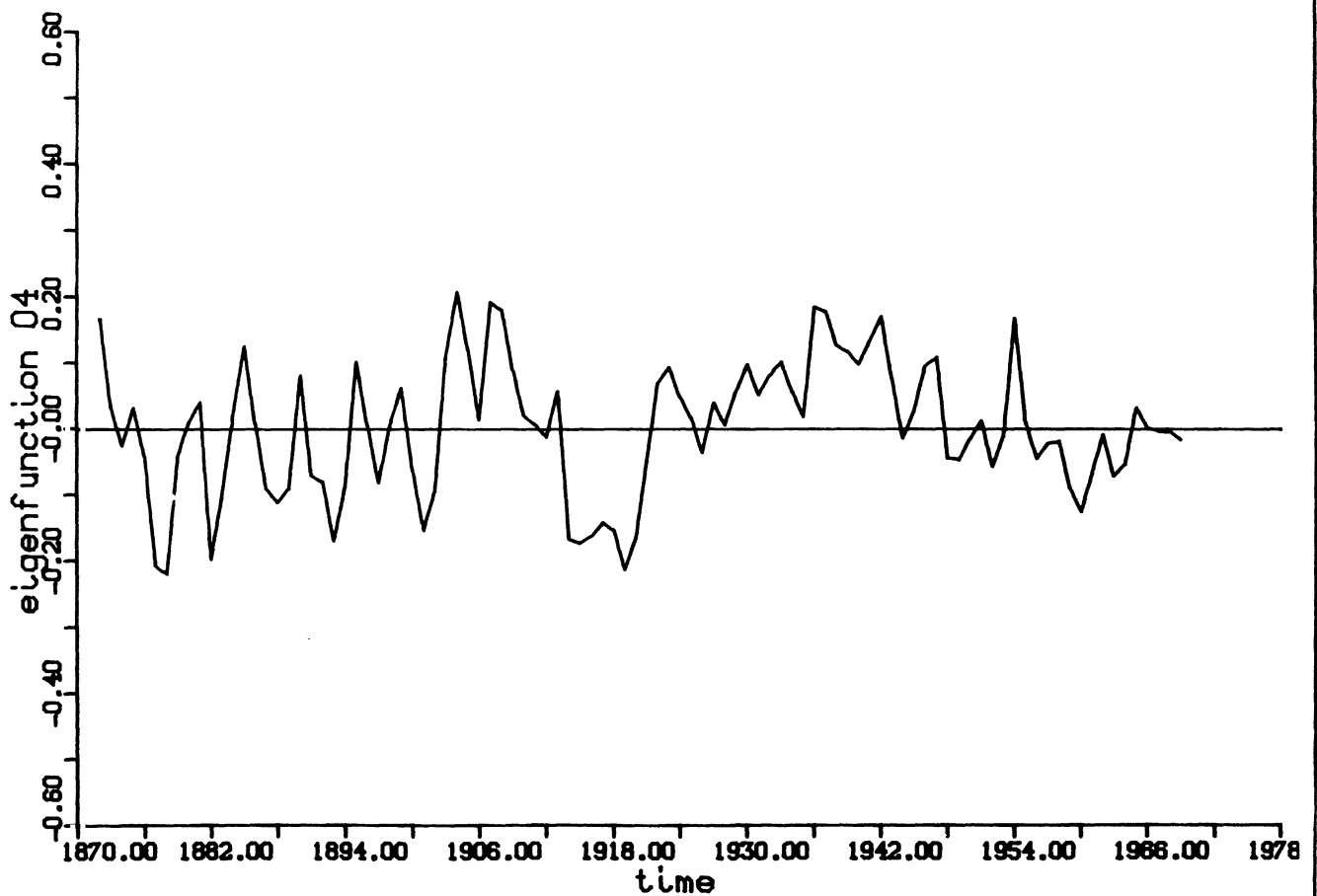
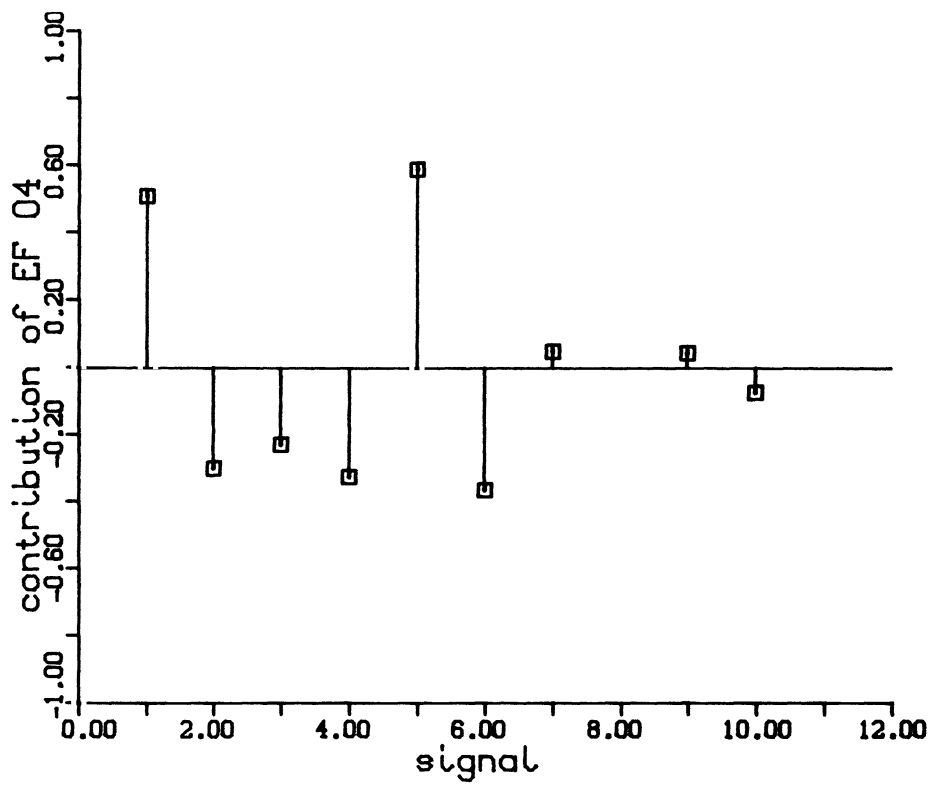
HOOFDCOMPONENT 3

$$\sigma_3 = 29.9$$

$$\sigma_3^2 / \sigma^2 = 0.028$$

Delft Hydraulics

Fig. [3]



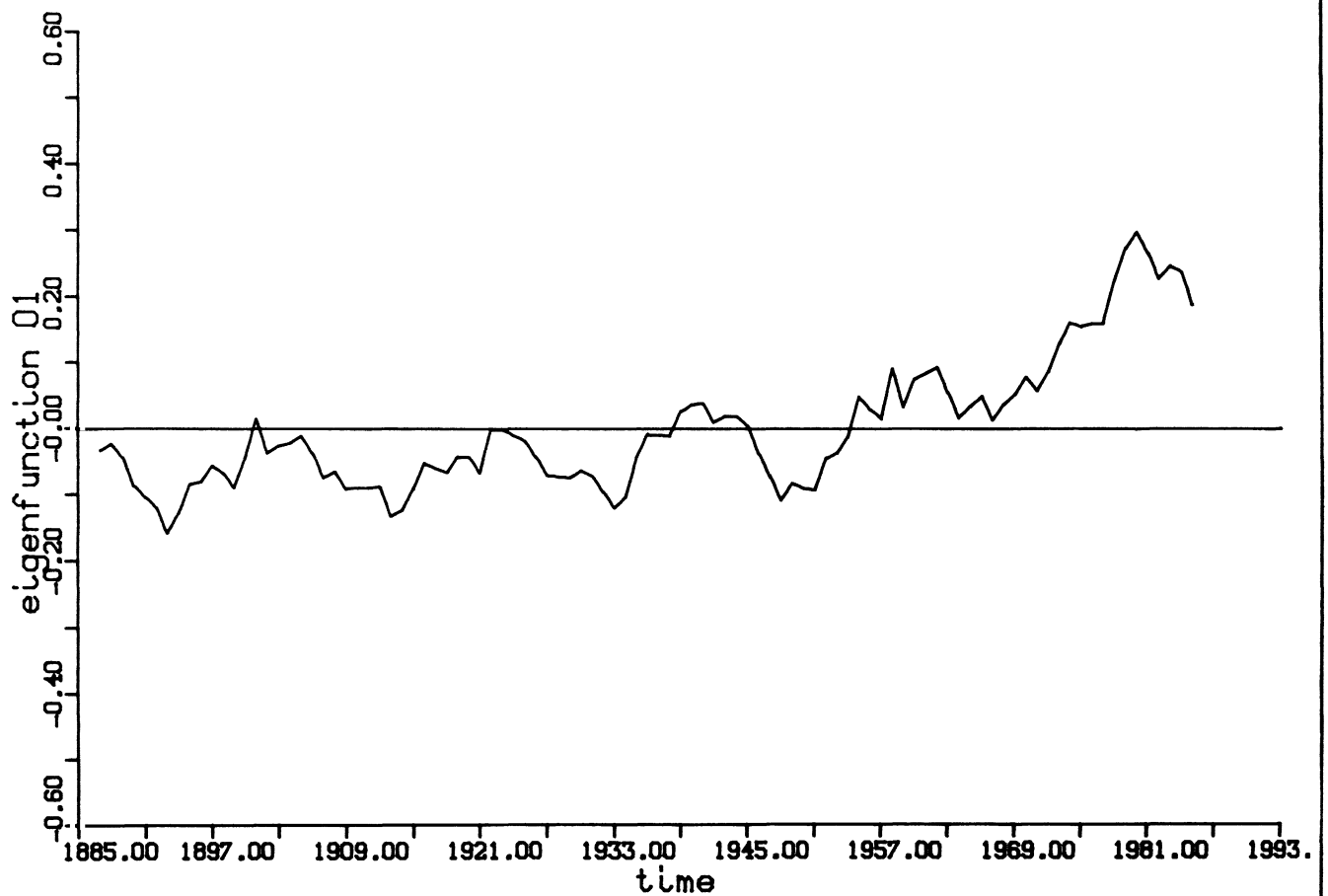
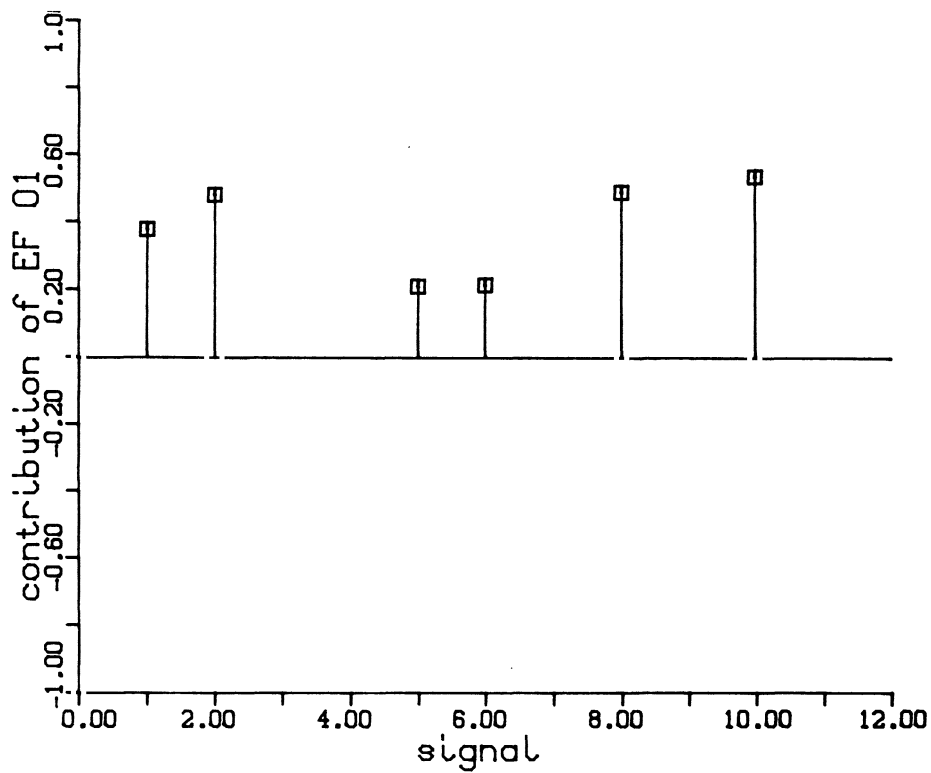
HOOFDCOMPONENT 4

$$\sigma_4 = 23.7$$

$$\sigma_4^2 / \sigma^2 = 0.018$$

Delft Hydraulics

Fig. [3]



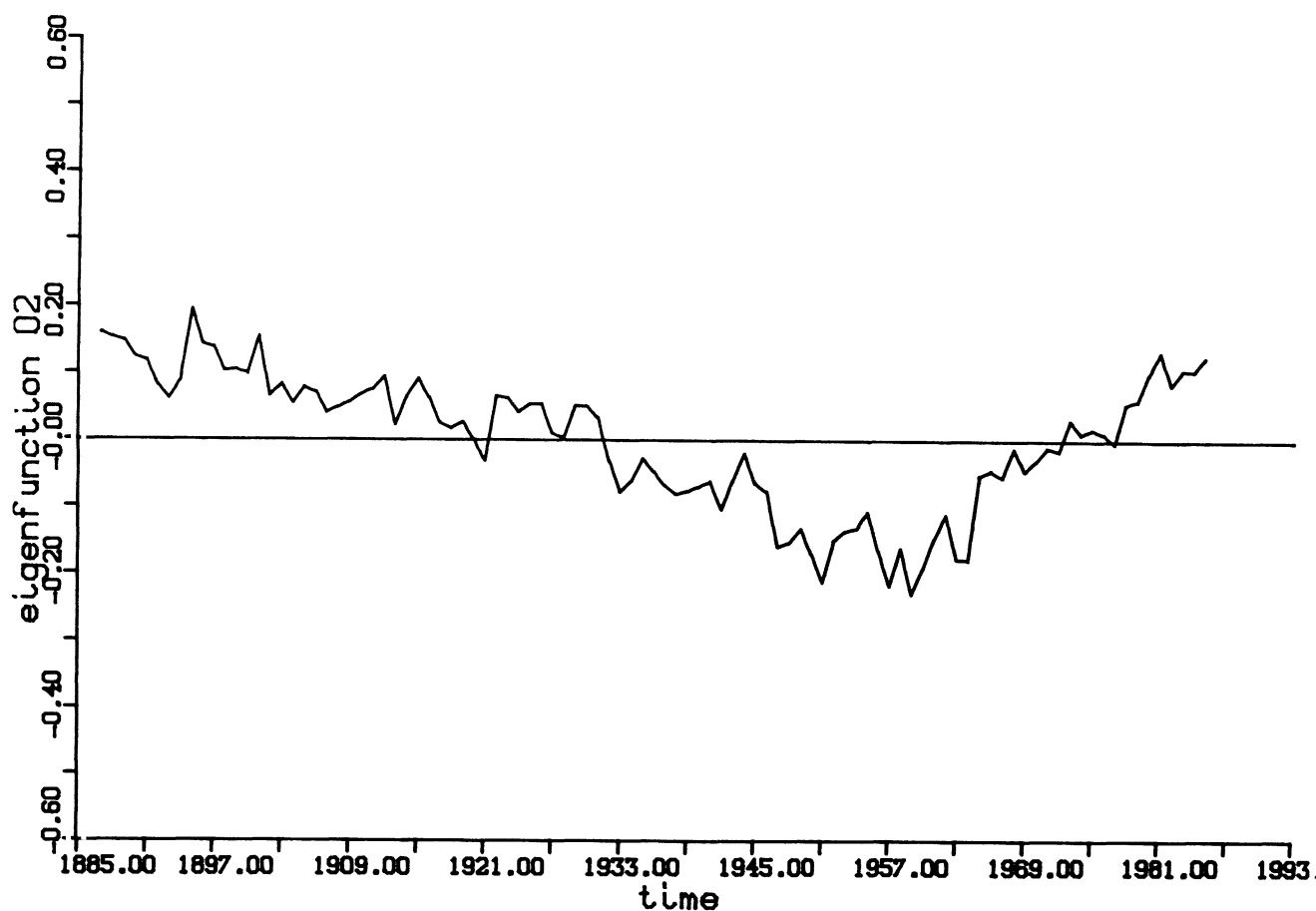
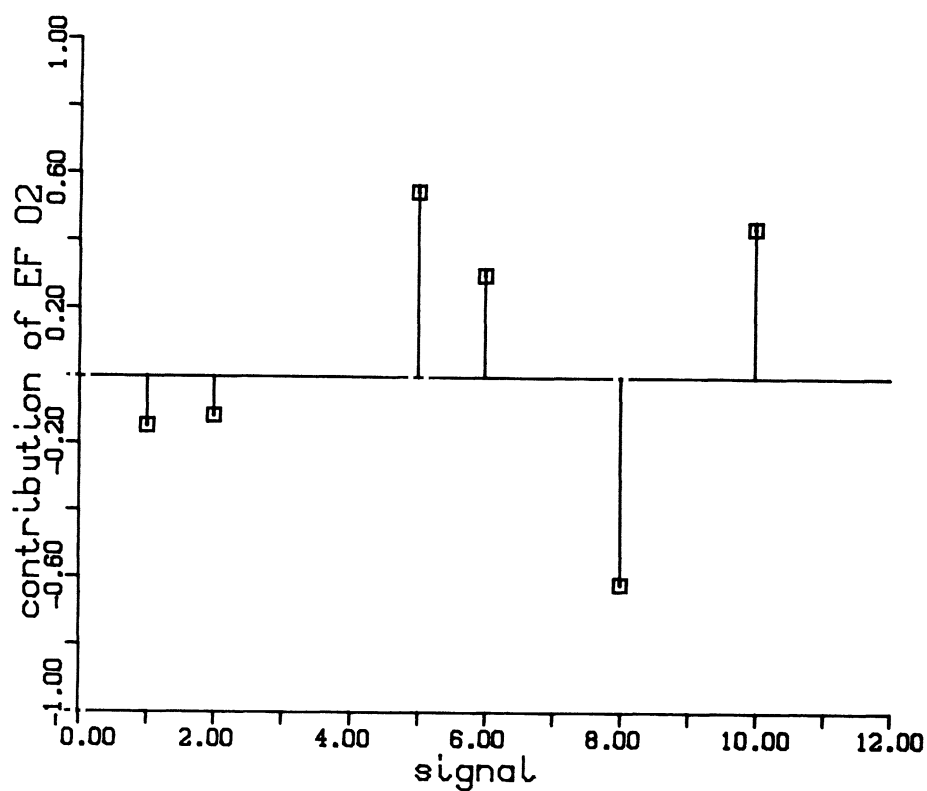
HOOFDCOMPONENT 1

$$\sigma_1 = 157.6$$

$$\sigma_1^2 / \sigma^2 = 0.807$$

Delft Hydraulics

Fig. [4]



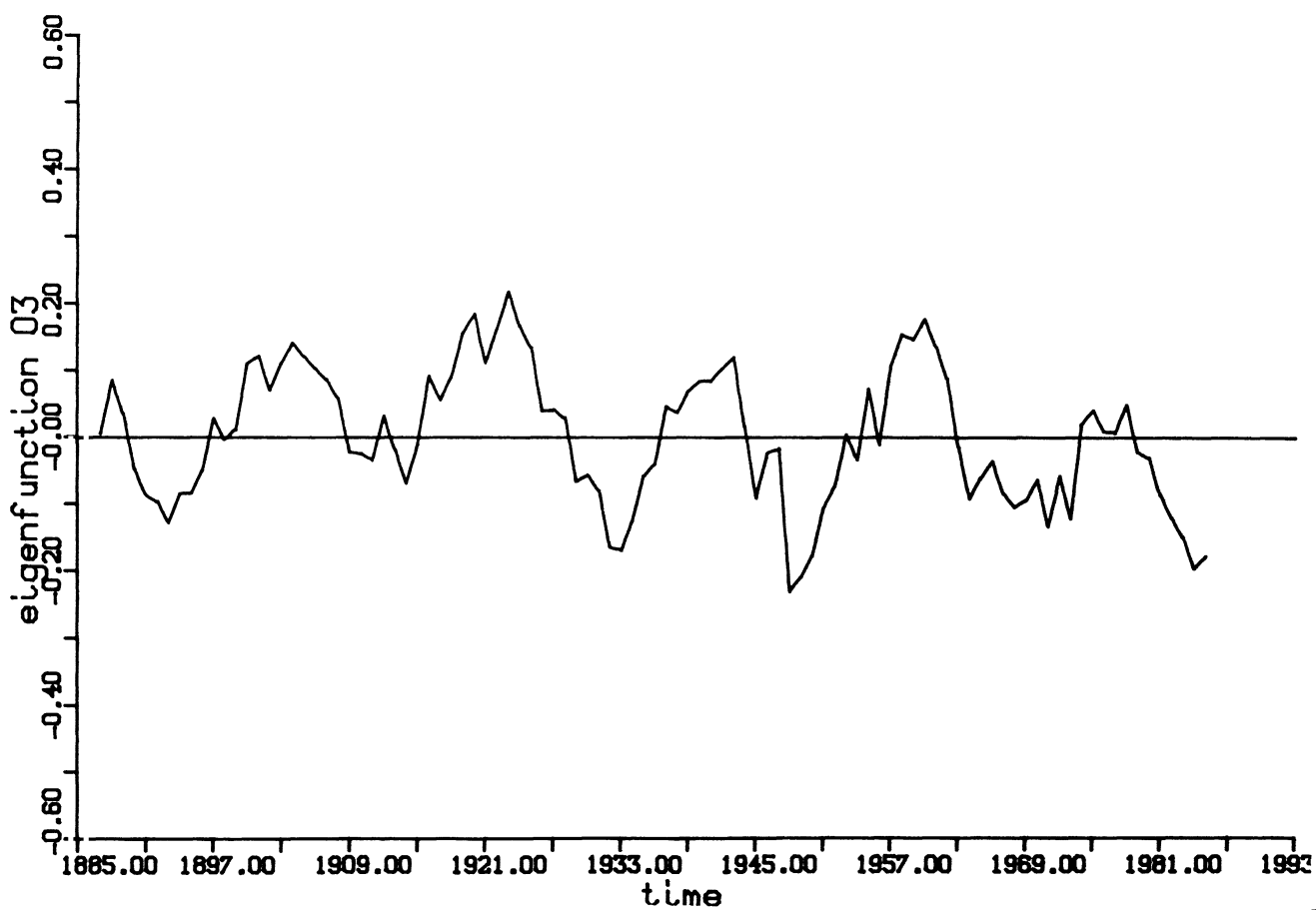
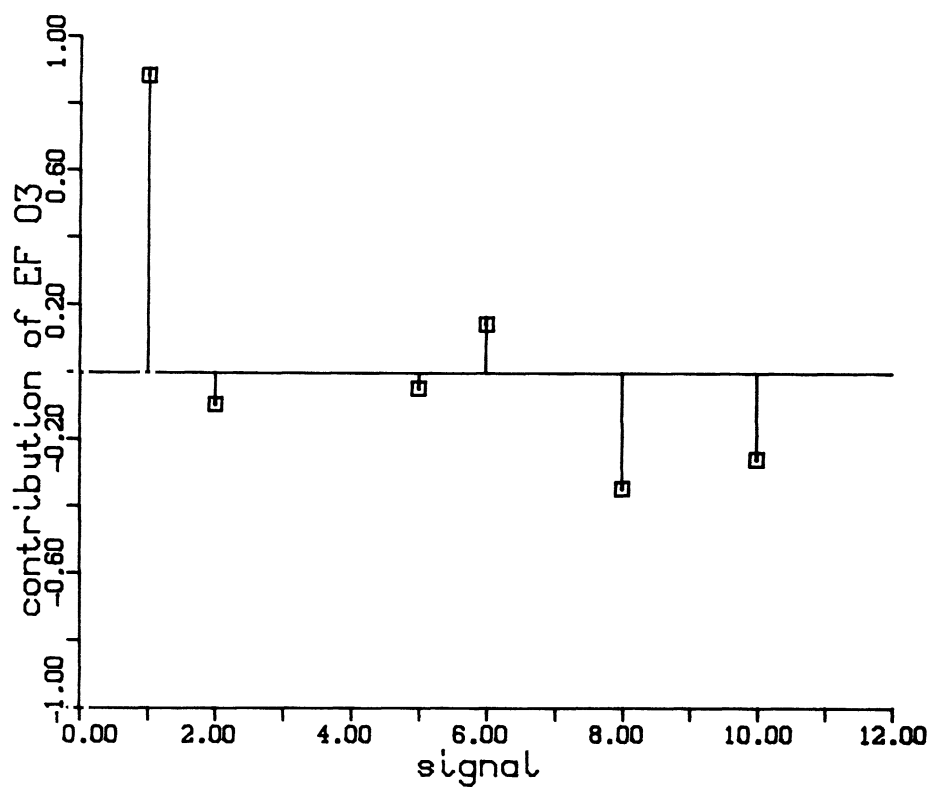
HOOFDCOMPONENT 2

$$\sigma_2 = 54.2$$

$$\sigma_2^2 / \sigma^2 = 0.096$$

Delft Hydraulics

Fig. [4]



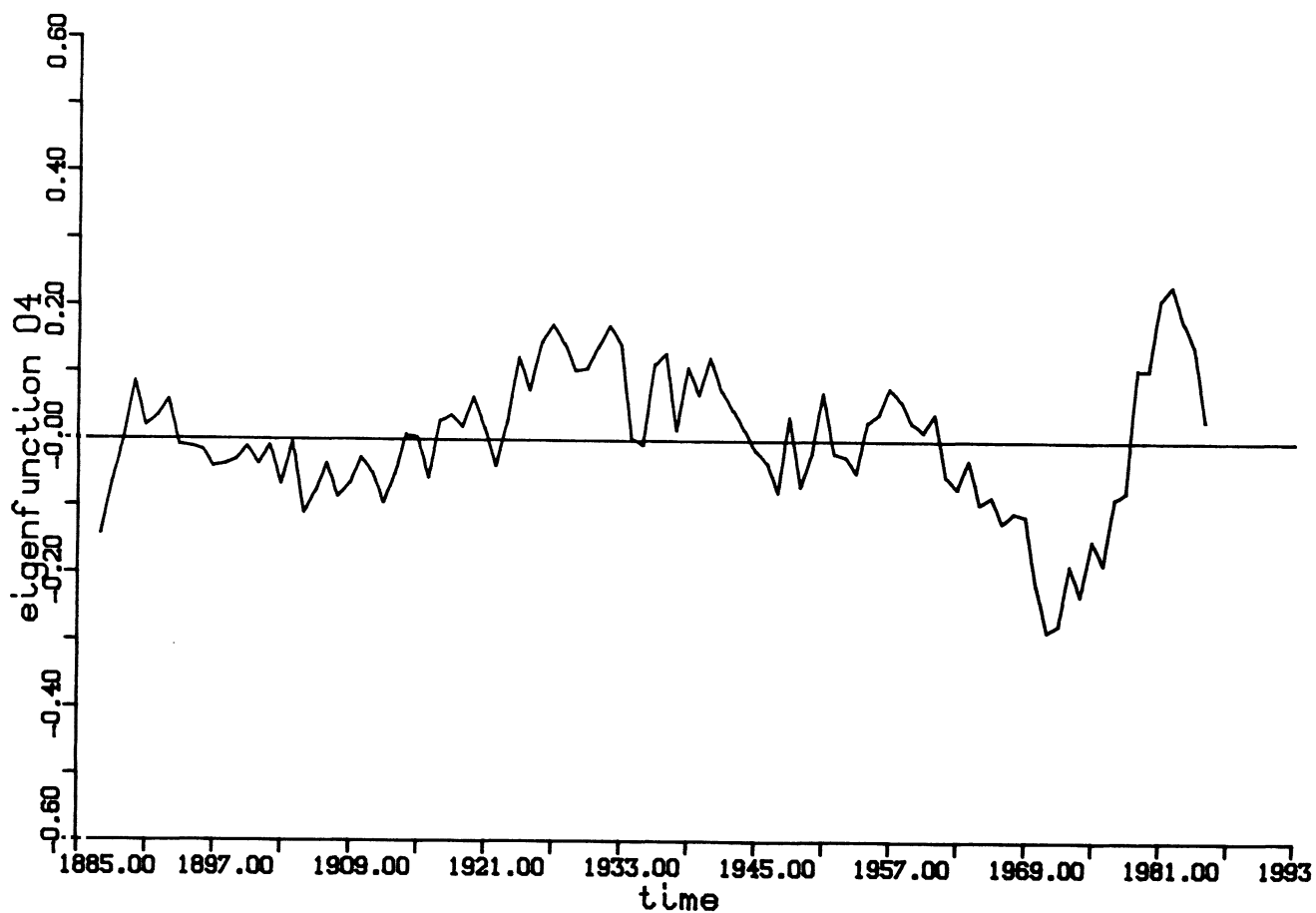
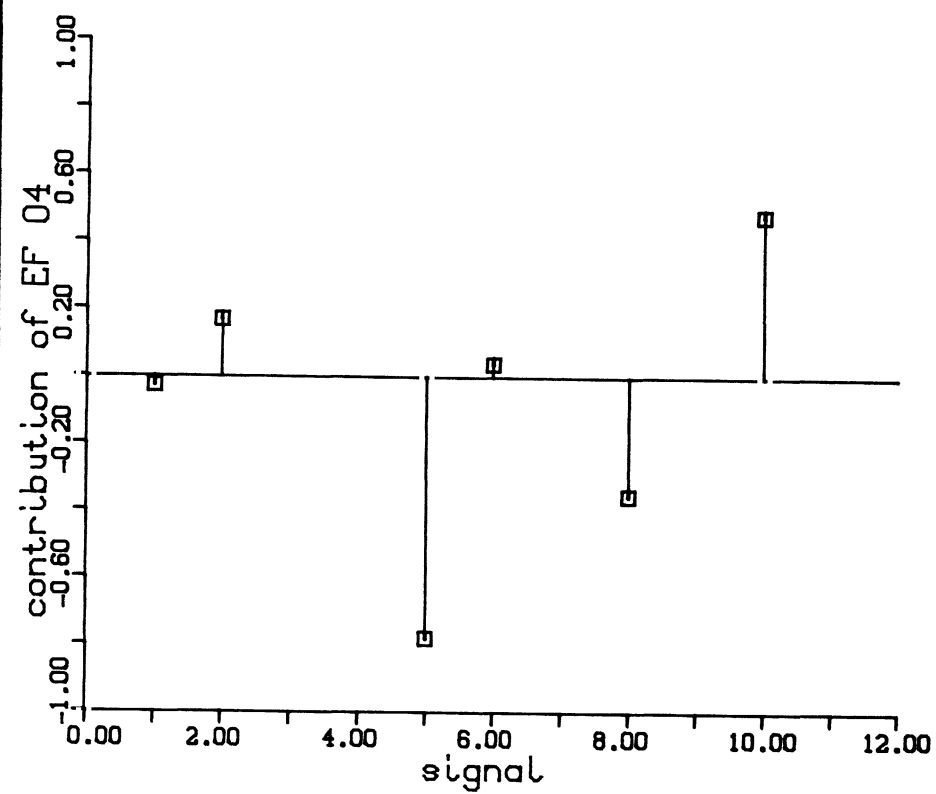
HOOFDCOMPONENT 3

$\sigma_3 = 36.1$

$\sigma_3^2 / \sigma^2 = 0.042$

Delft Hydraulics

Fig. [4]



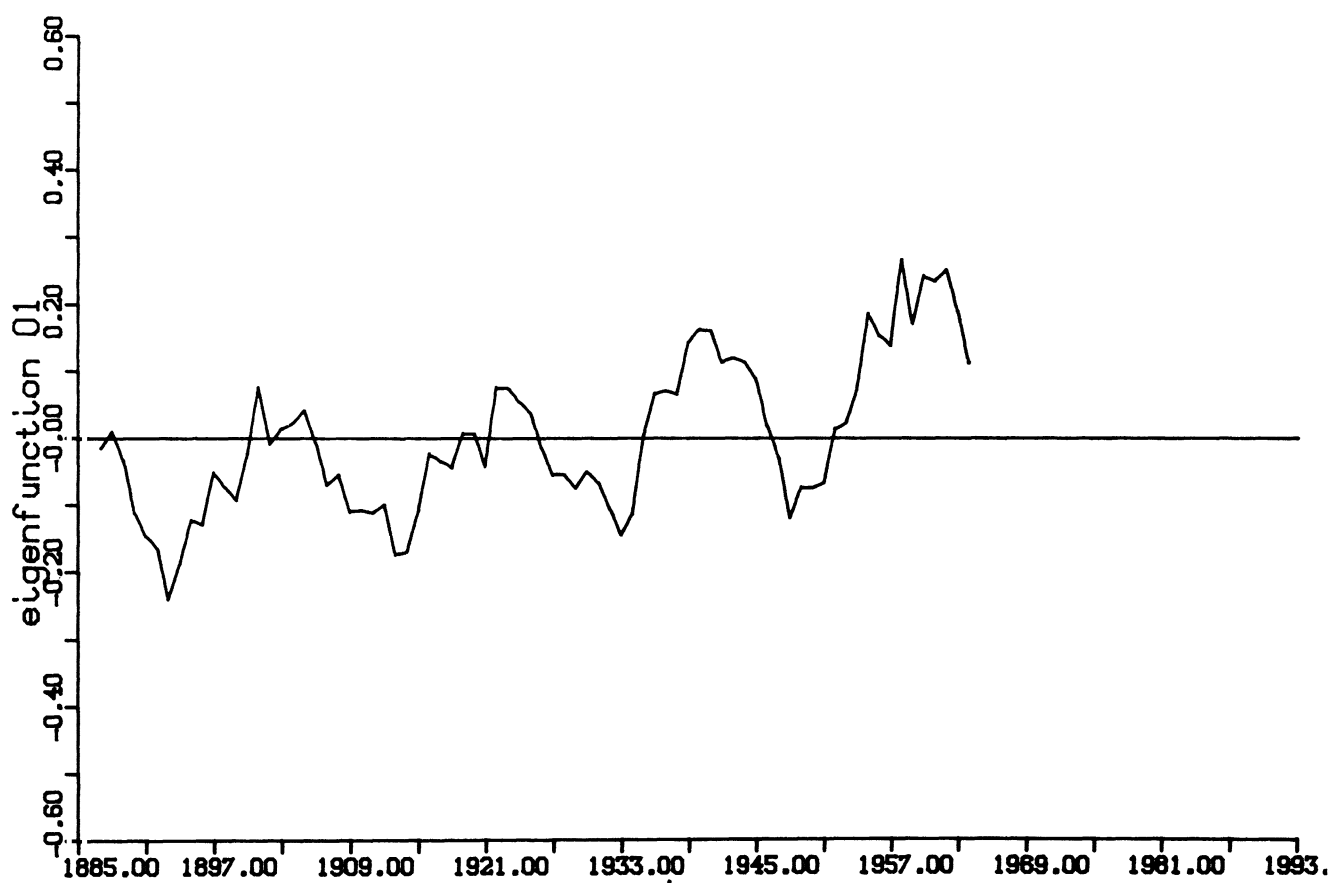
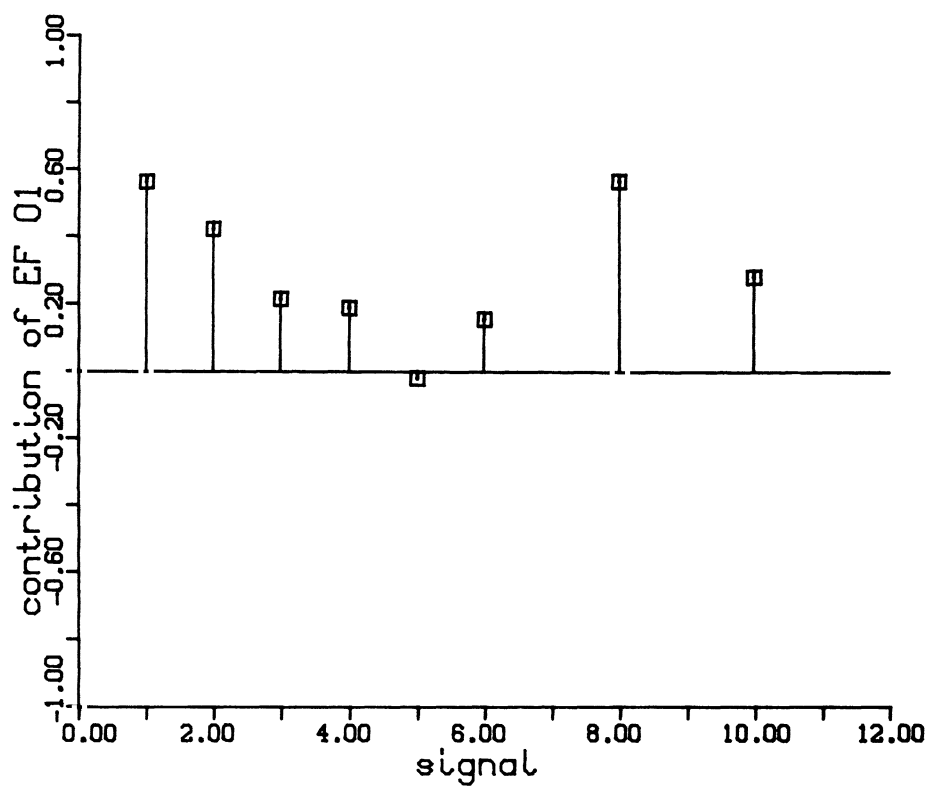
HOOFDCOMPONENT 4

$$\sigma_4 = 27.7$$

$$\sigma_4^2 / \sigma^2 = 0.025$$

Delft Hydraulics

Fig. [4]



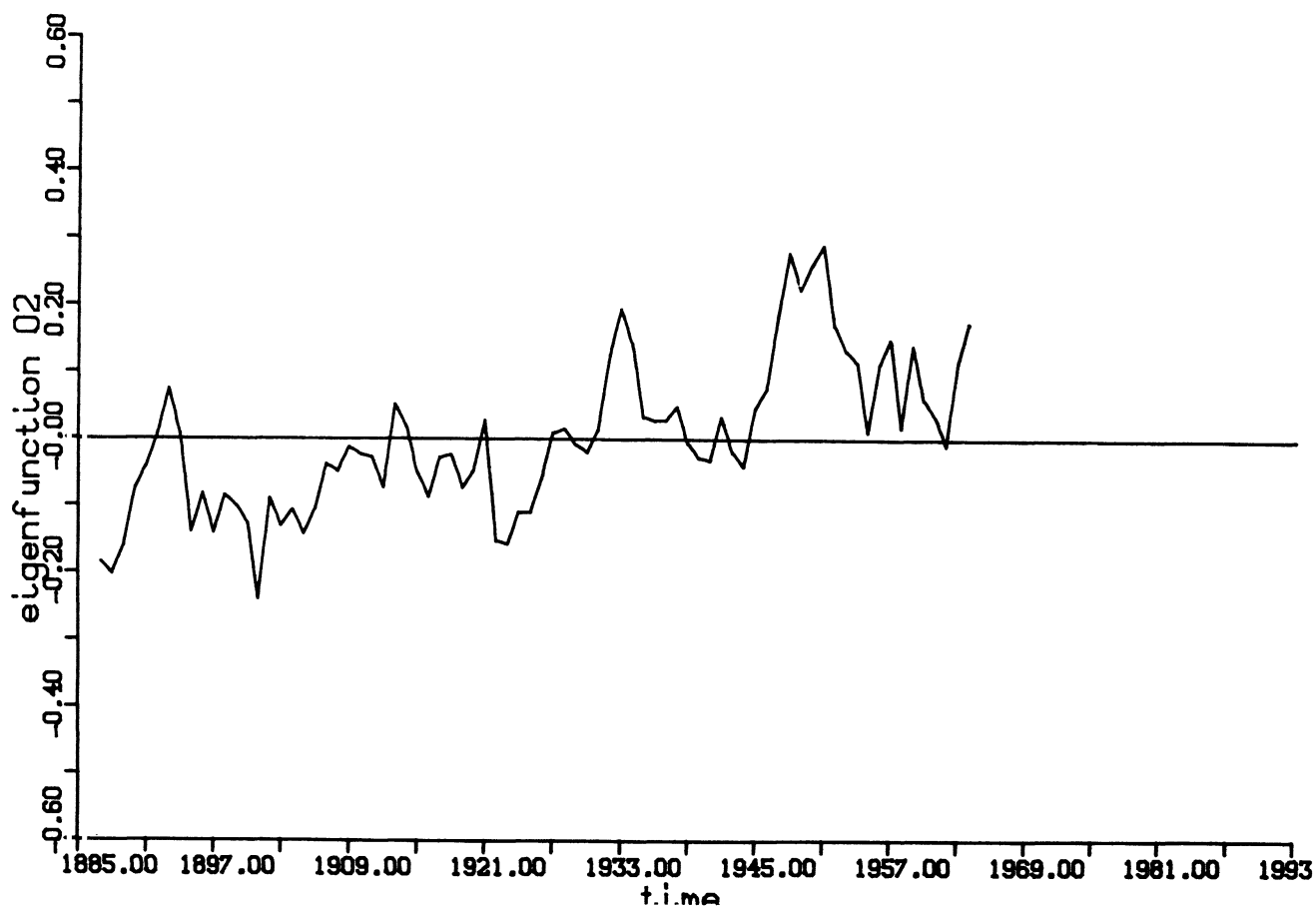
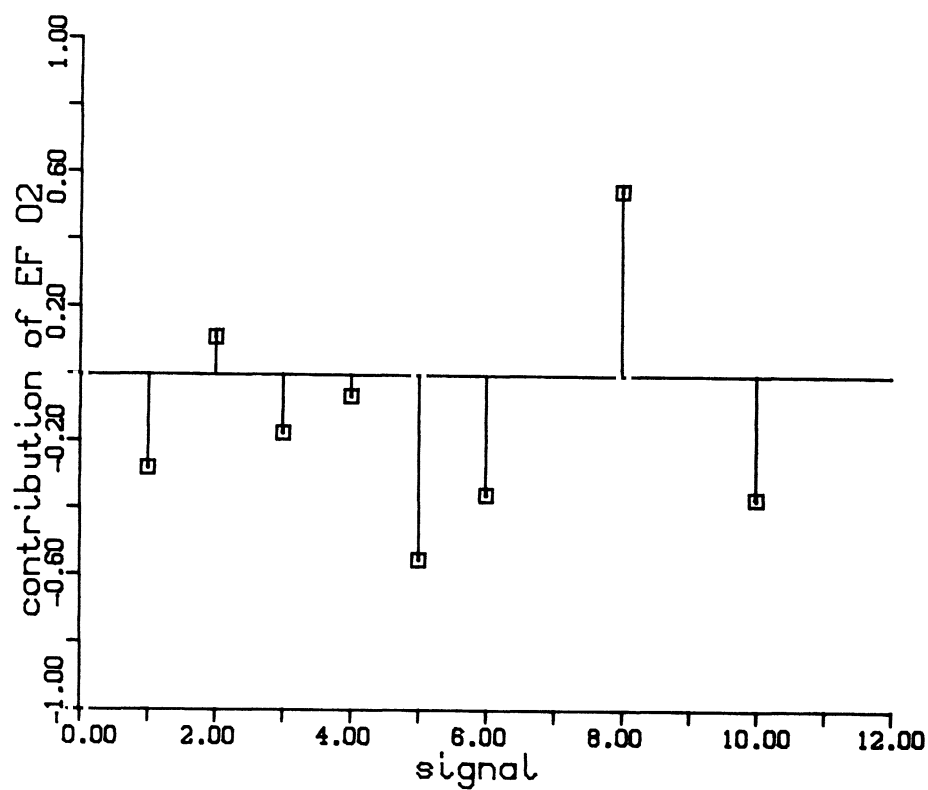
HOOFDCOMPONENT 1

$\sigma_1 = 87.6$

$\sigma_1^2 / \sigma^2 = 0.654$

Delft Hydraulics

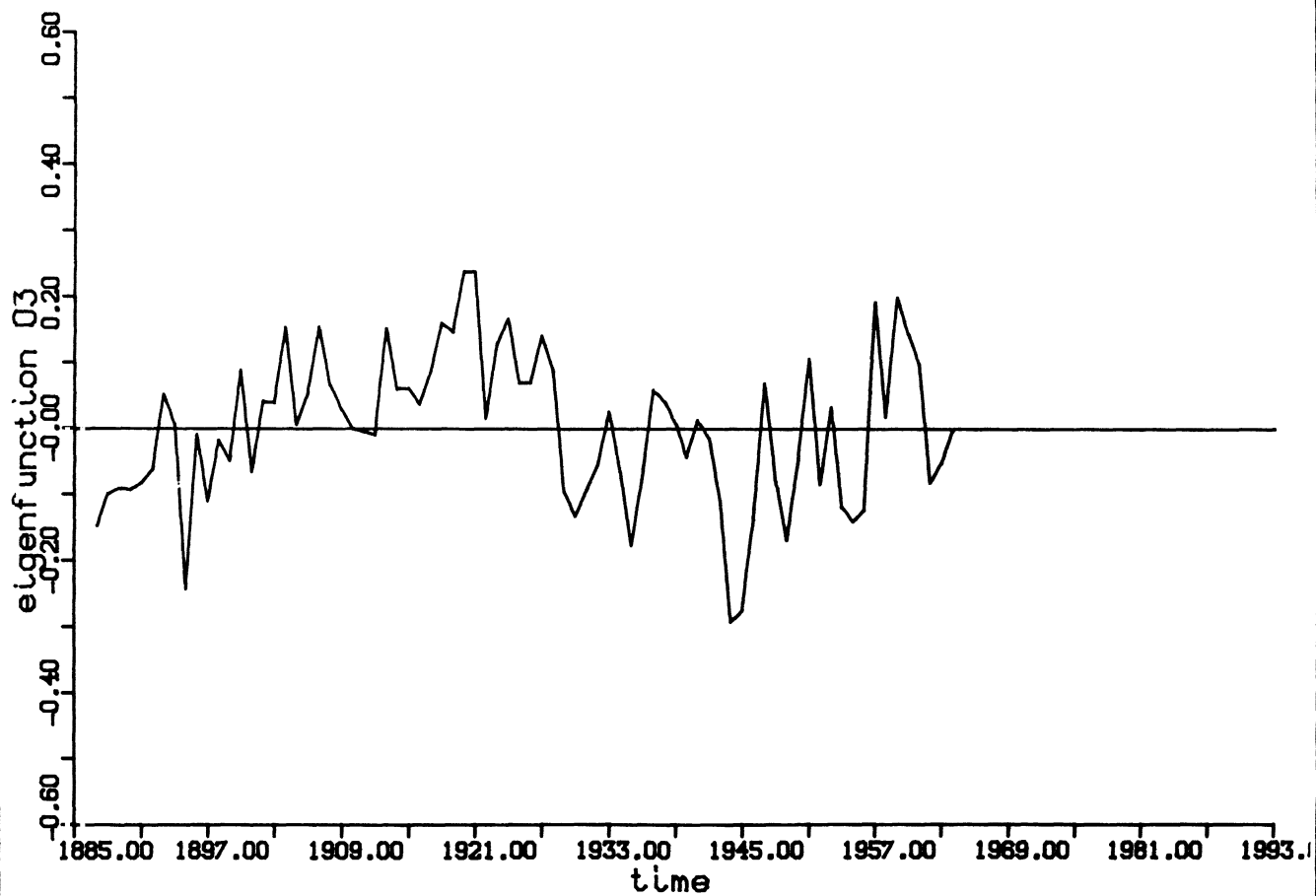
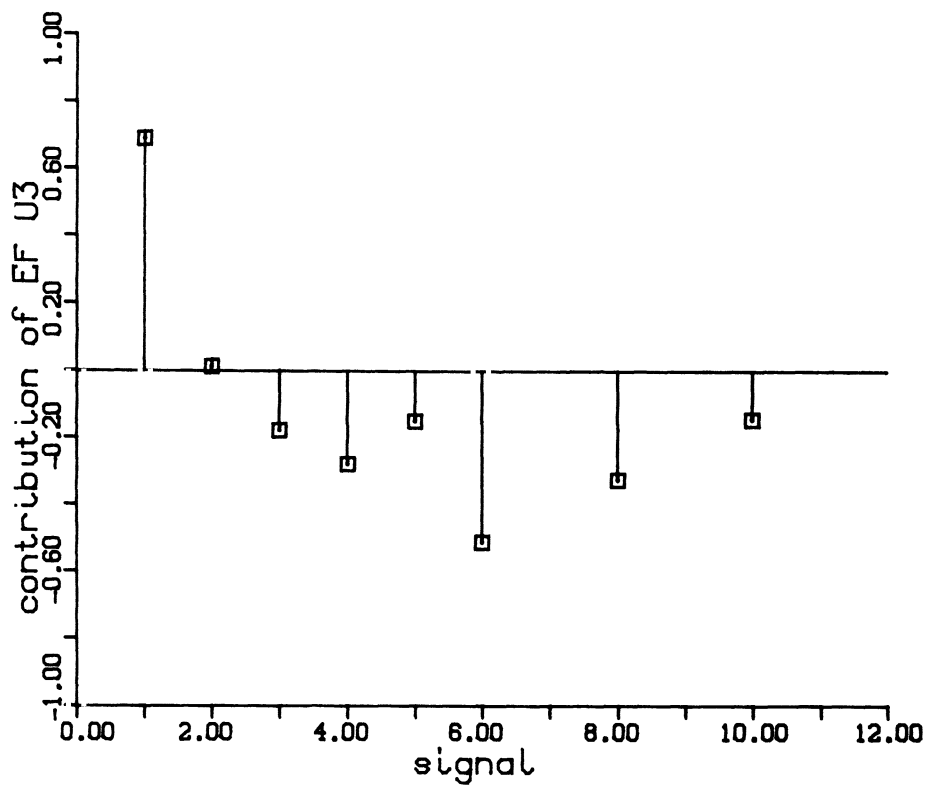
Fig. [5]



HOOFDCOMPONENT 2

$\sigma_2 = 48.2$

$\frac{\sigma_2^2}{\sigma^2} = 0.198$



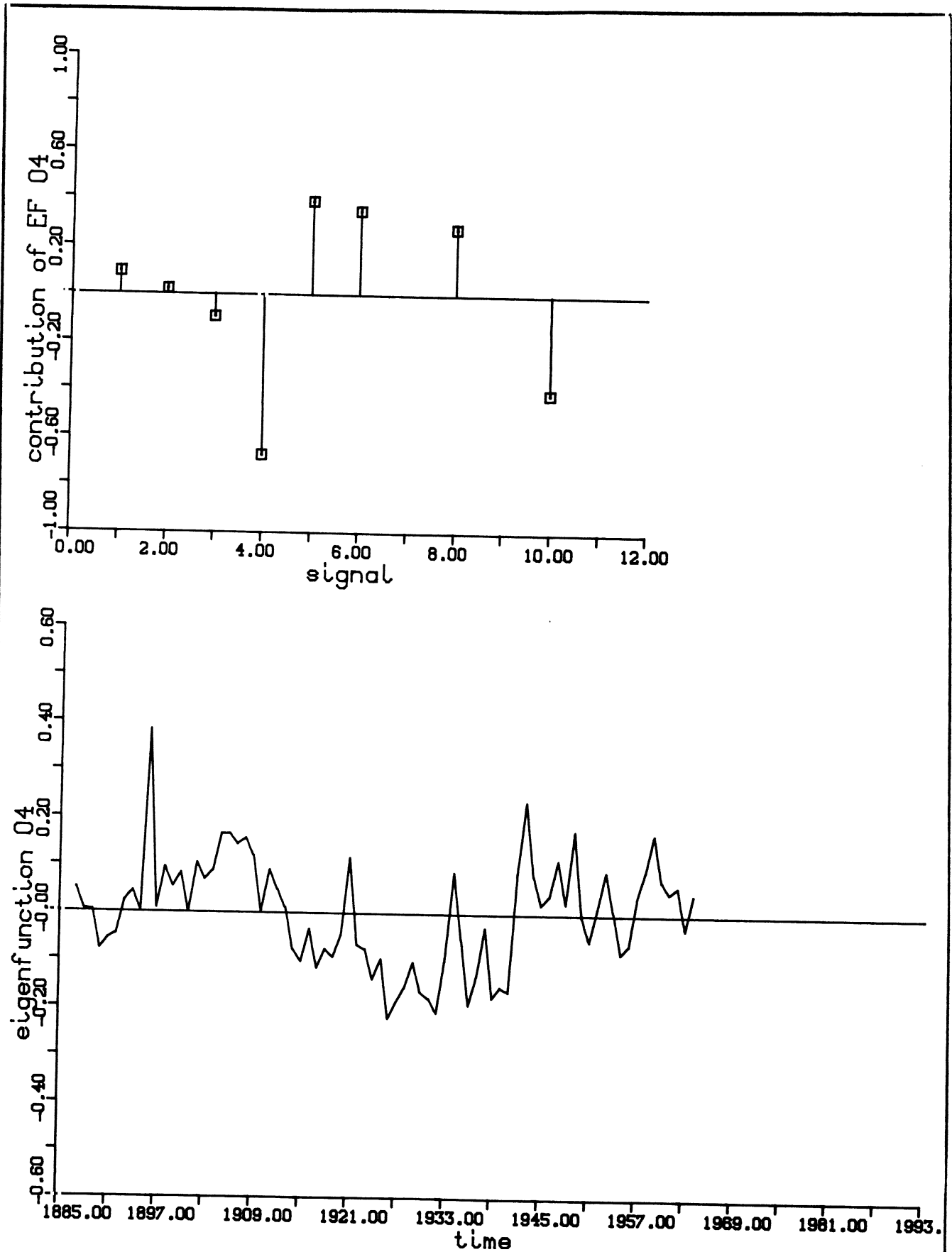
HOOFDCOMPONENT 3

$$\sigma_3 = 23.6$$

$$\sigma_3^2 / \sigma^2 = 0.048$$

Delft Hydraulics

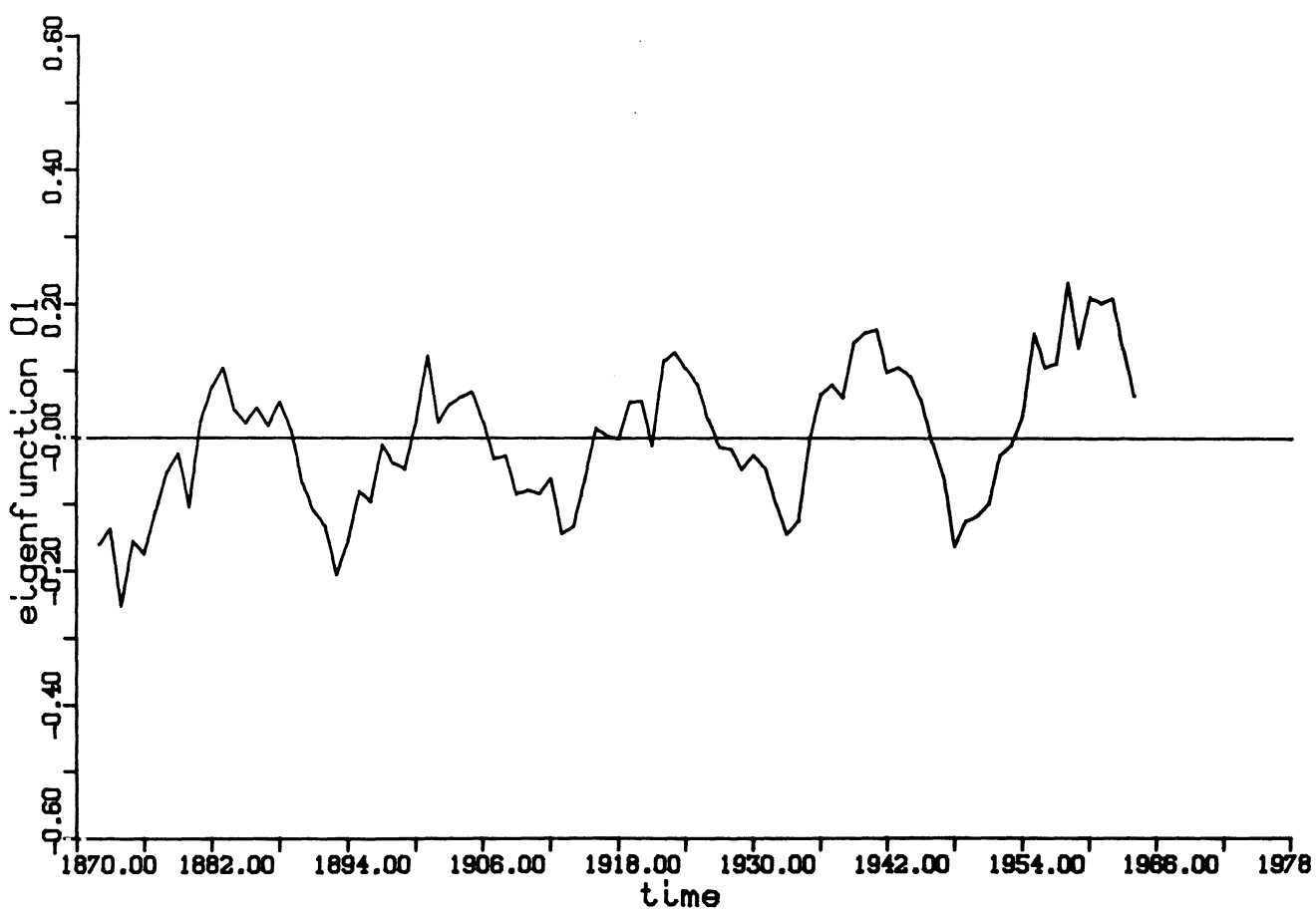
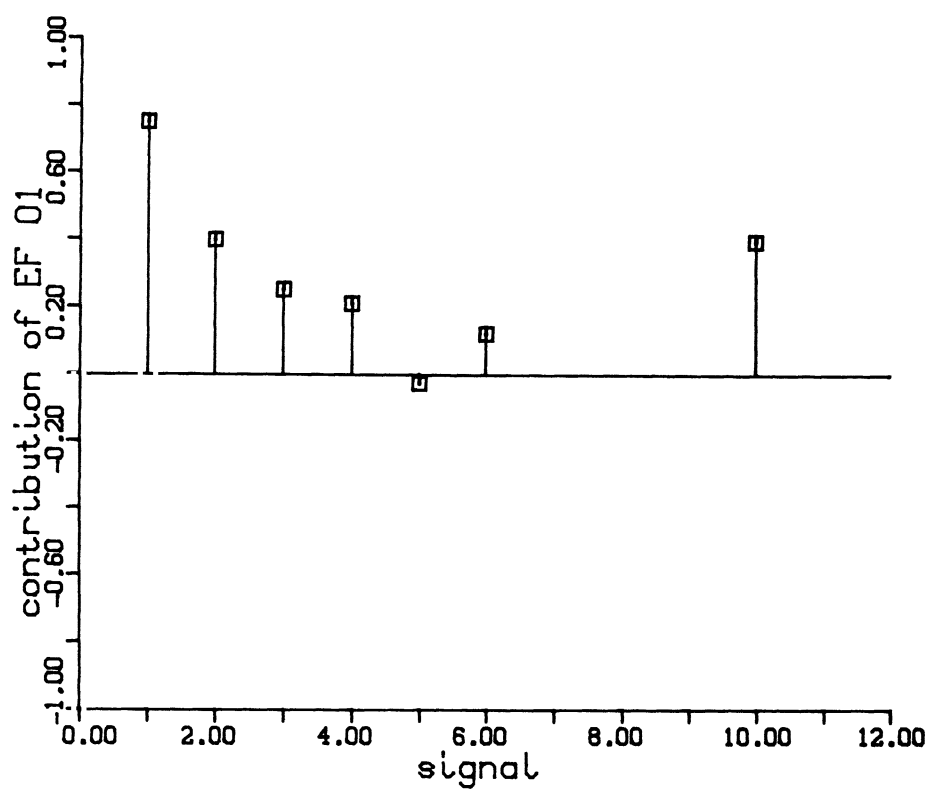
Fig. [5]



HOOFDCOMPONENT 4

$$\sigma_4 = 19.4$$

$$\sigma_4^2 / \sigma^2 = 0.032$$



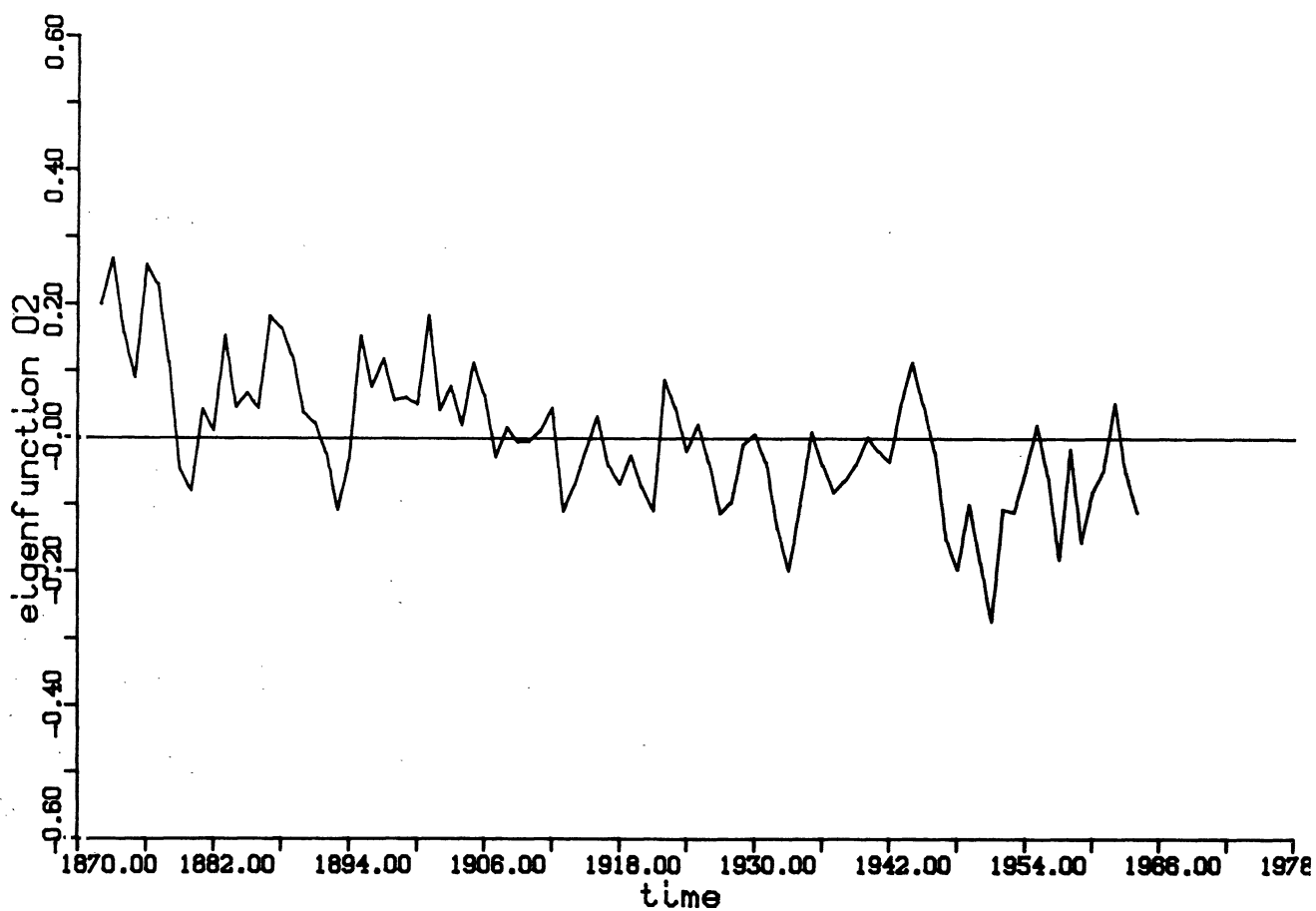
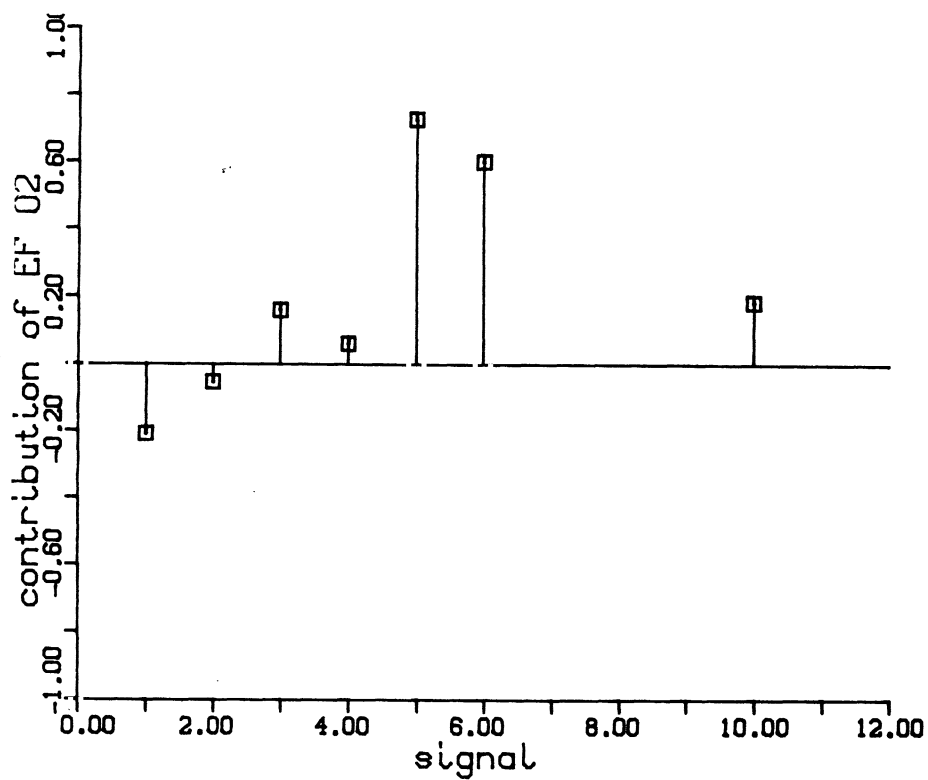
HOOFDCOMPONENT 1

$$\sigma_1 = 84.3$$

$$\sigma_1^2 / \sigma^2 = 0.624$$

Delft Hydraulics

Fig. [6]



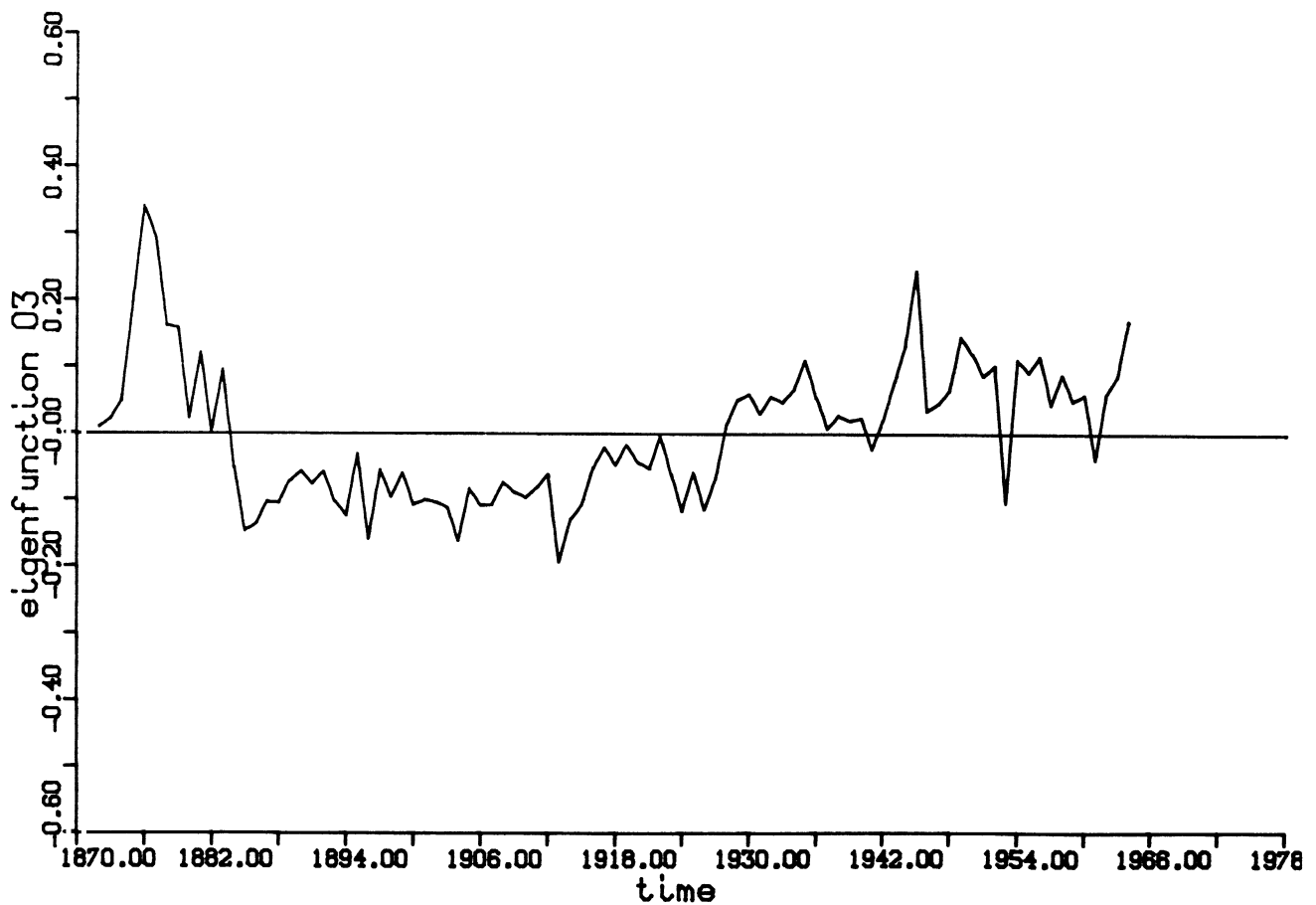
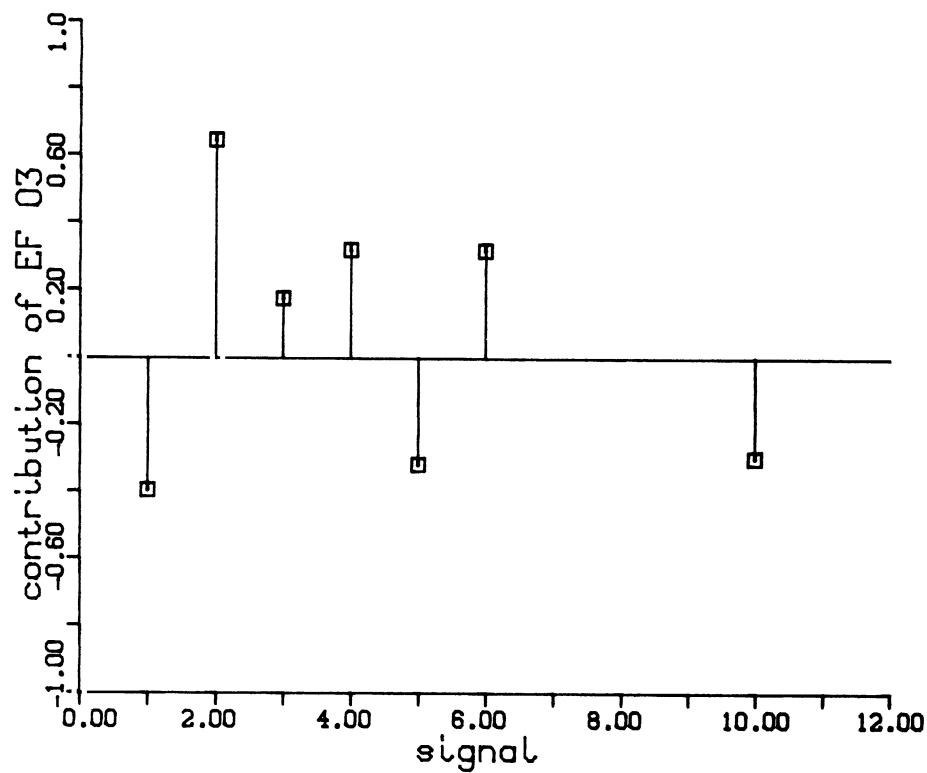
HOOFDCOMPONENT 2

$$\sigma_2 = 45.1$$

$$\sigma_2^2 / \sigma^2 = 0.178$$

Delft Hydraulics

Fig. [6]



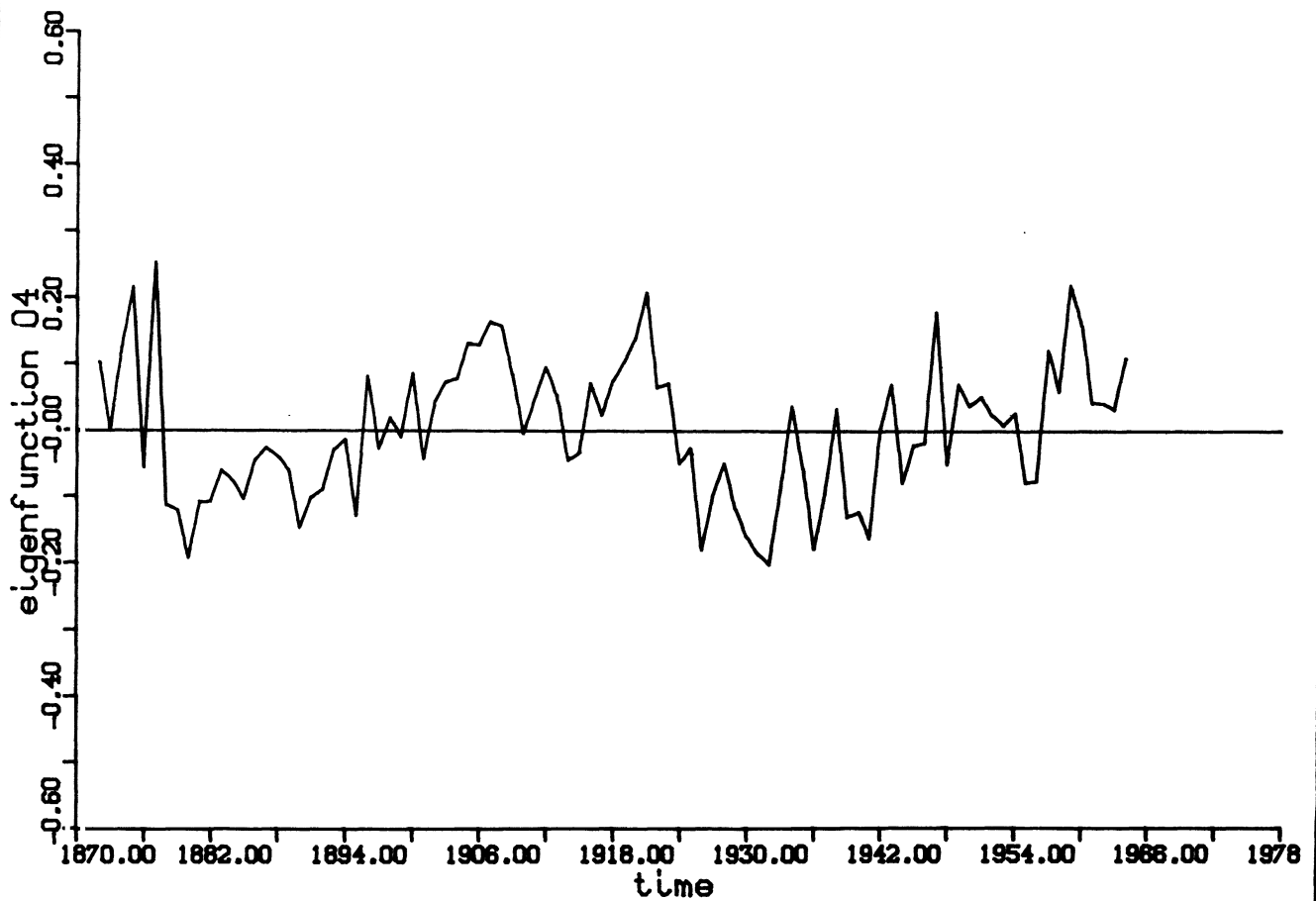
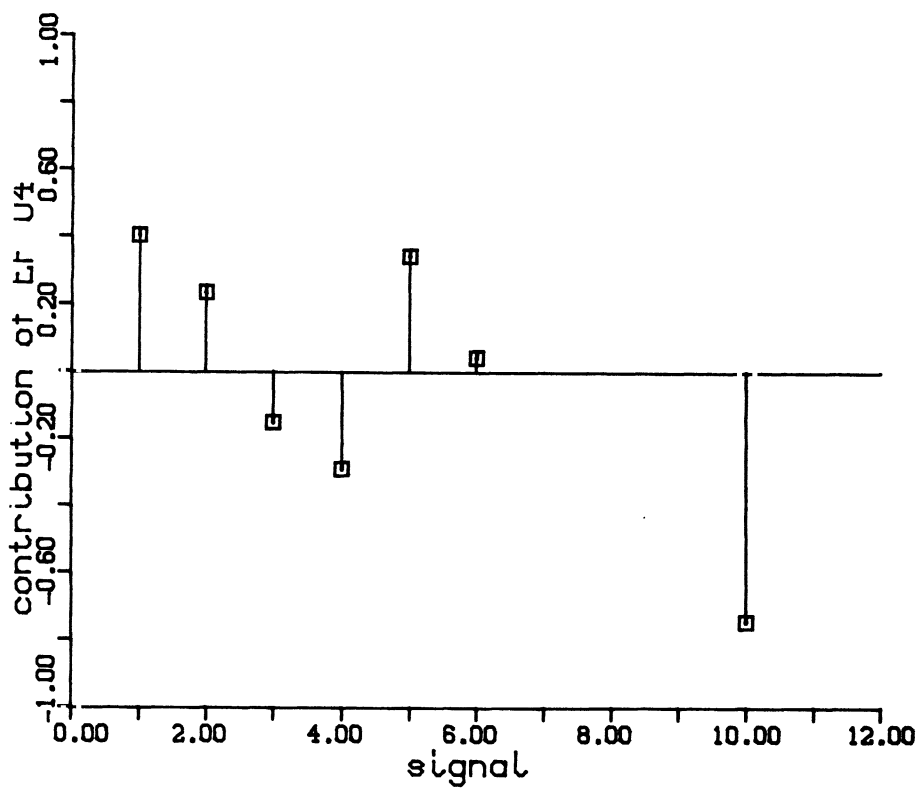
HOOFDCOMPONENT 3

$\sigma_3 = 30.1$

$\sigma_3^2 / \sigma^2 = 0.080$

Delft Hydraulics

Fig. [6]



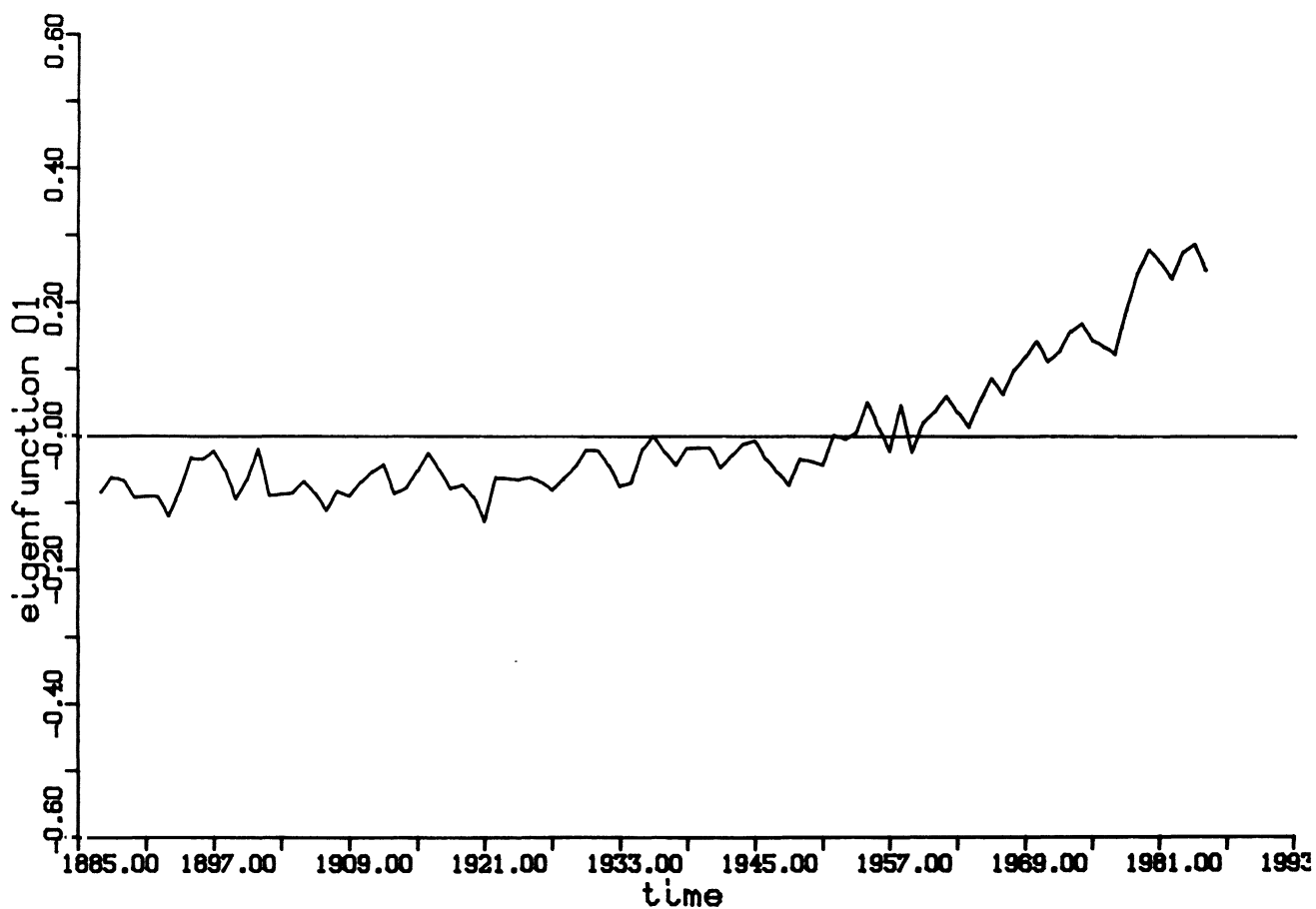
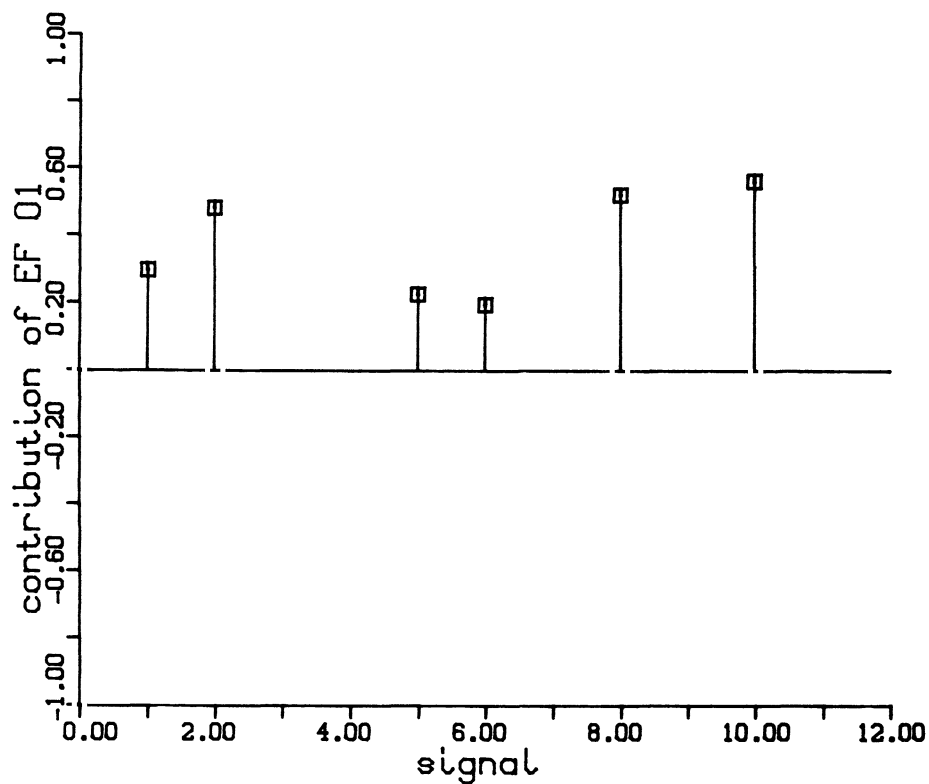
HOOFDCOMPONENT 4

$$\sigma_4 = 22.4$$

$$\sigma_4^2 / \sigma^2 = 0.044$$

Delft Hydraulics

Fig. [6]



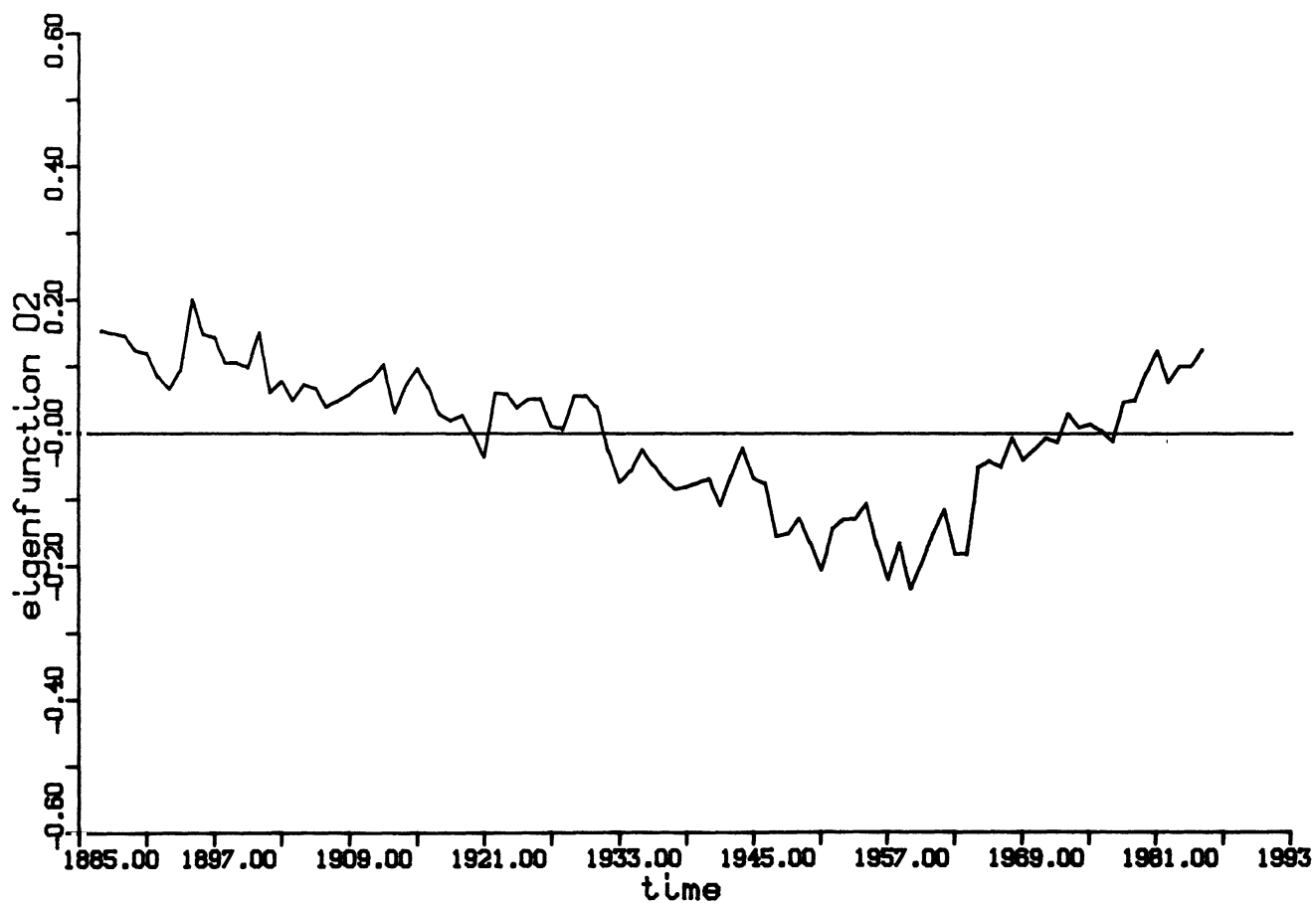
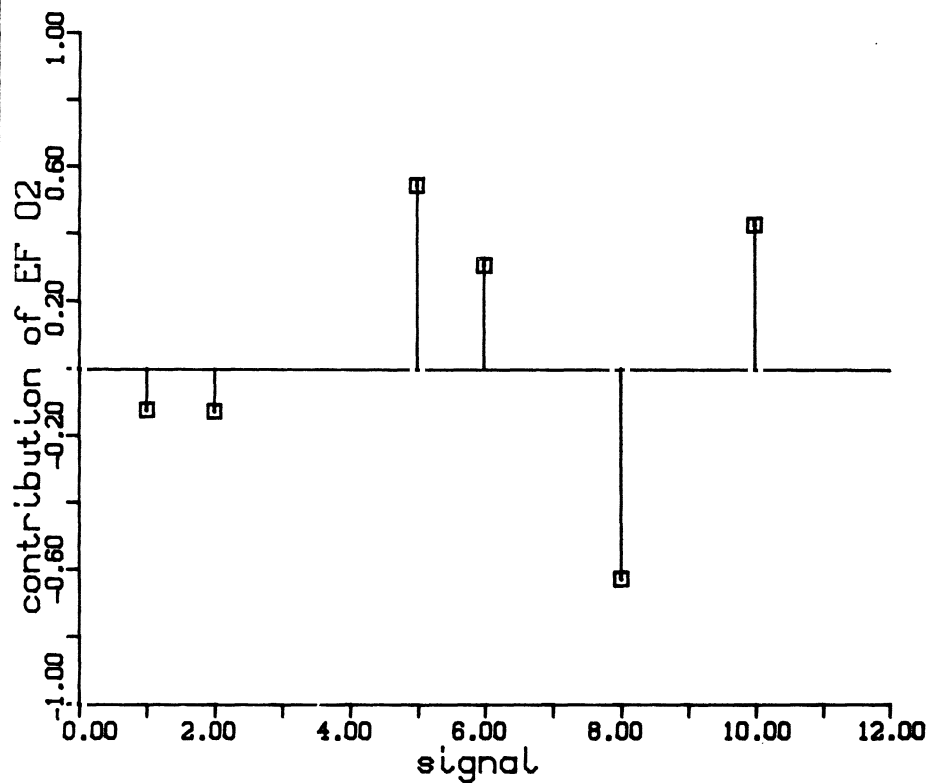
HOOFDCOMPONENT 1

$$\sigma_1 = 142.8$$

$$\sigma_1^2 / \sigma^2 = 0.800$$

Delft Hydraulics

Fig. [4']



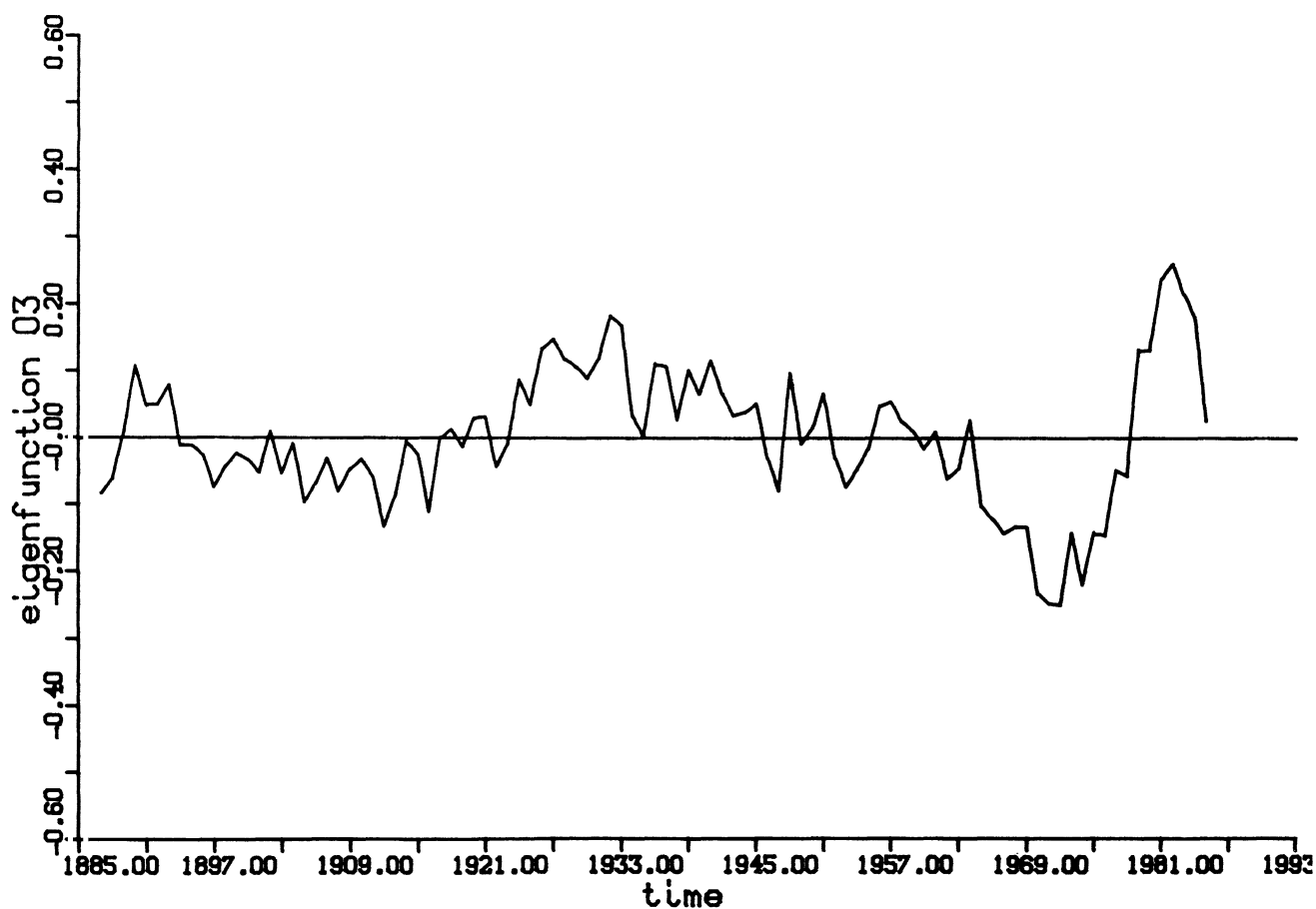
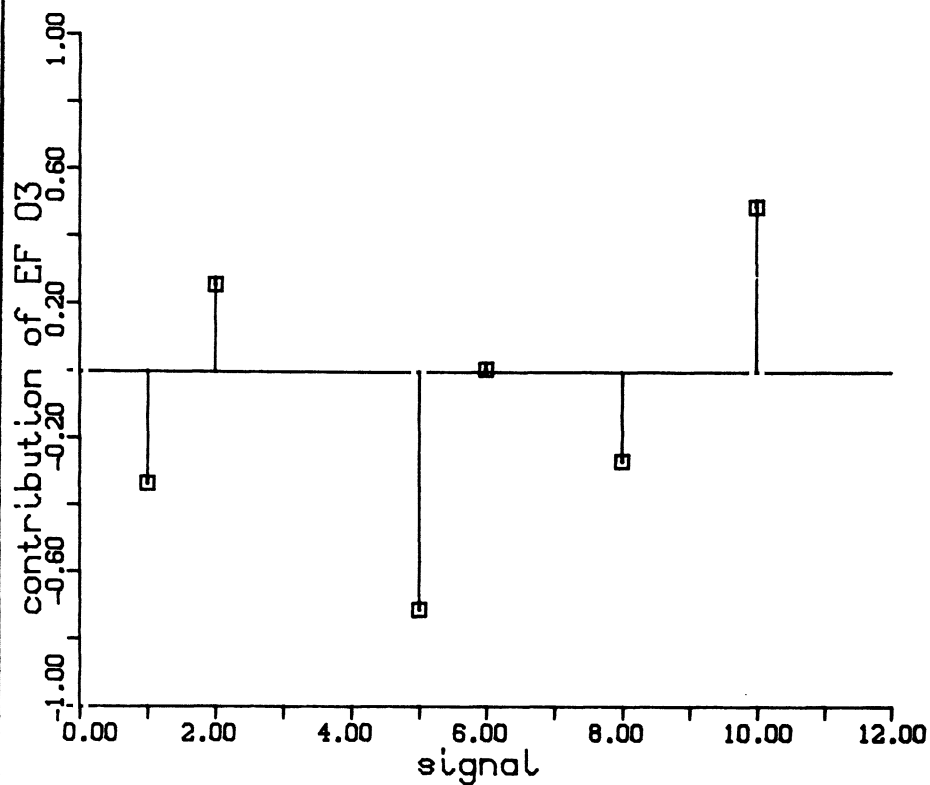
HOOFDCOMPONENT 2

$\sigma_2 = 54.4$

$\sigma_2^2 / \sigma^2 = 0.116$

Delft Hydraulics

Fig. [4']



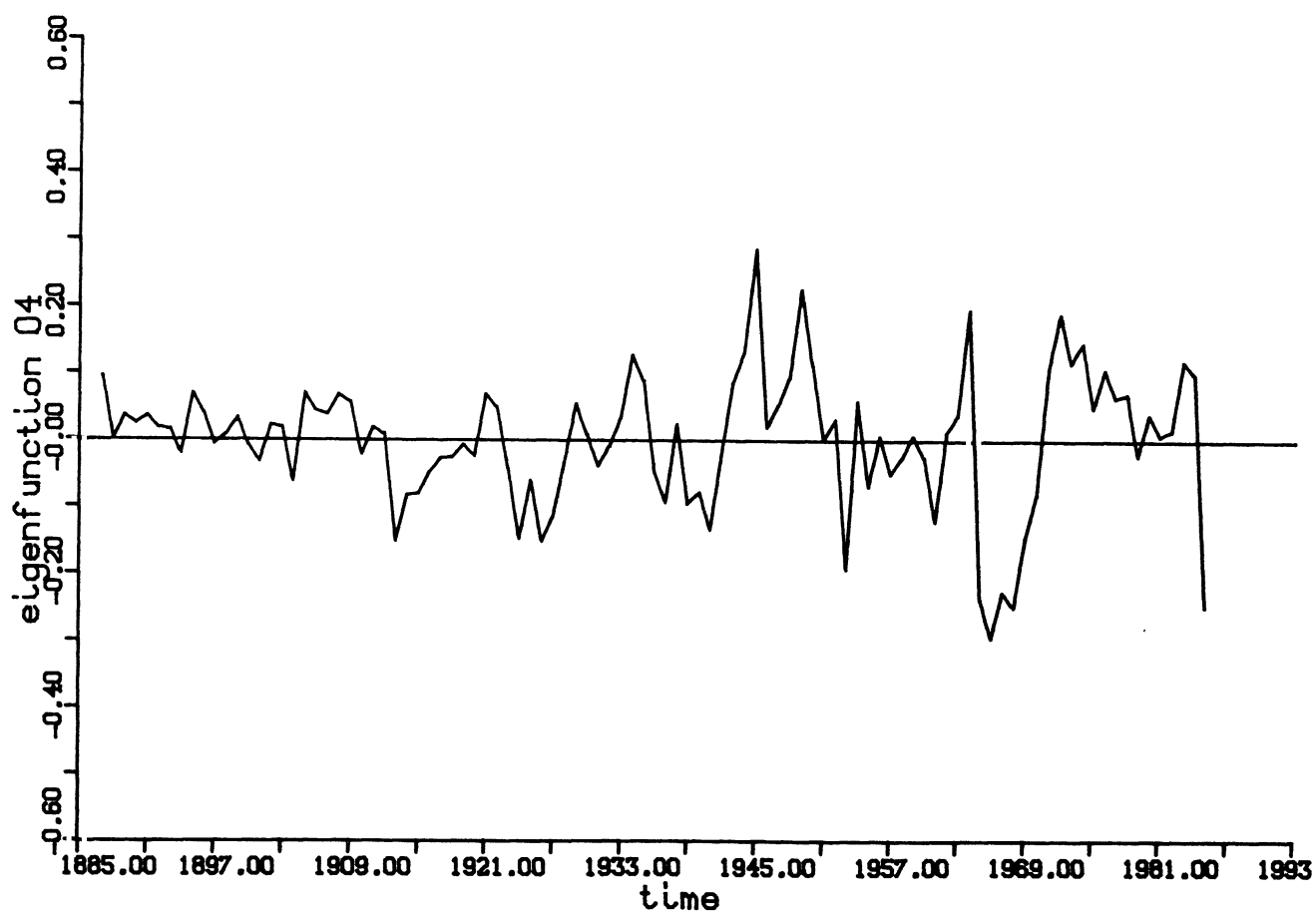
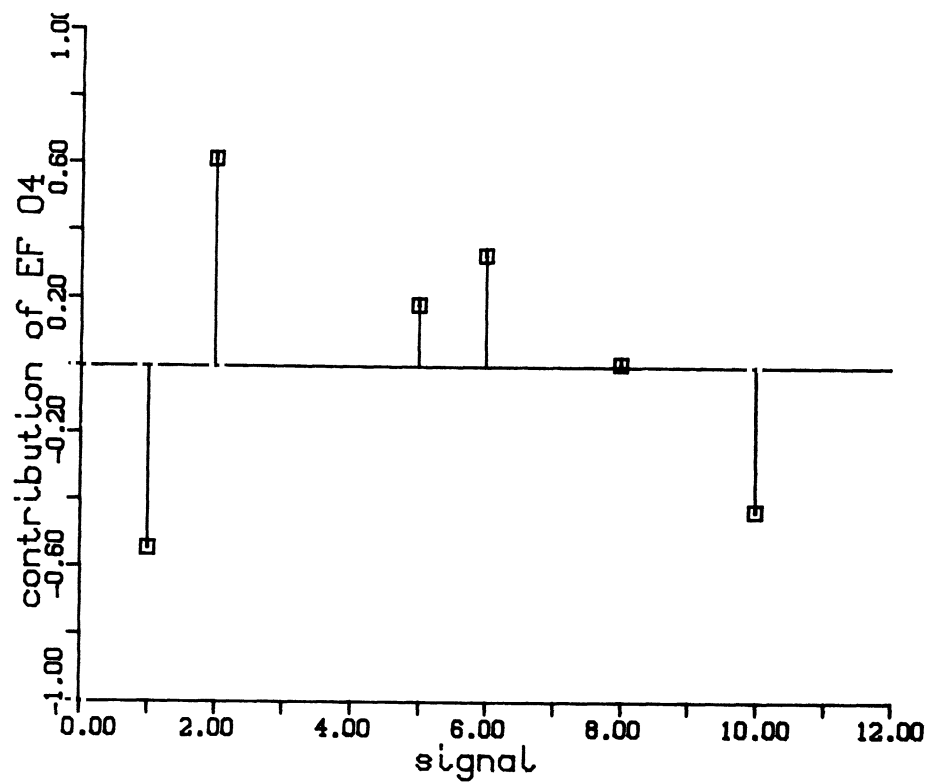
HOOFDCOMPONENT 3

$$\sigma_3 = 28.4$$

$$\sigma_3^2 / \sigma^2 = 0.032$$

Delft Hydraulics

Fig. [4']



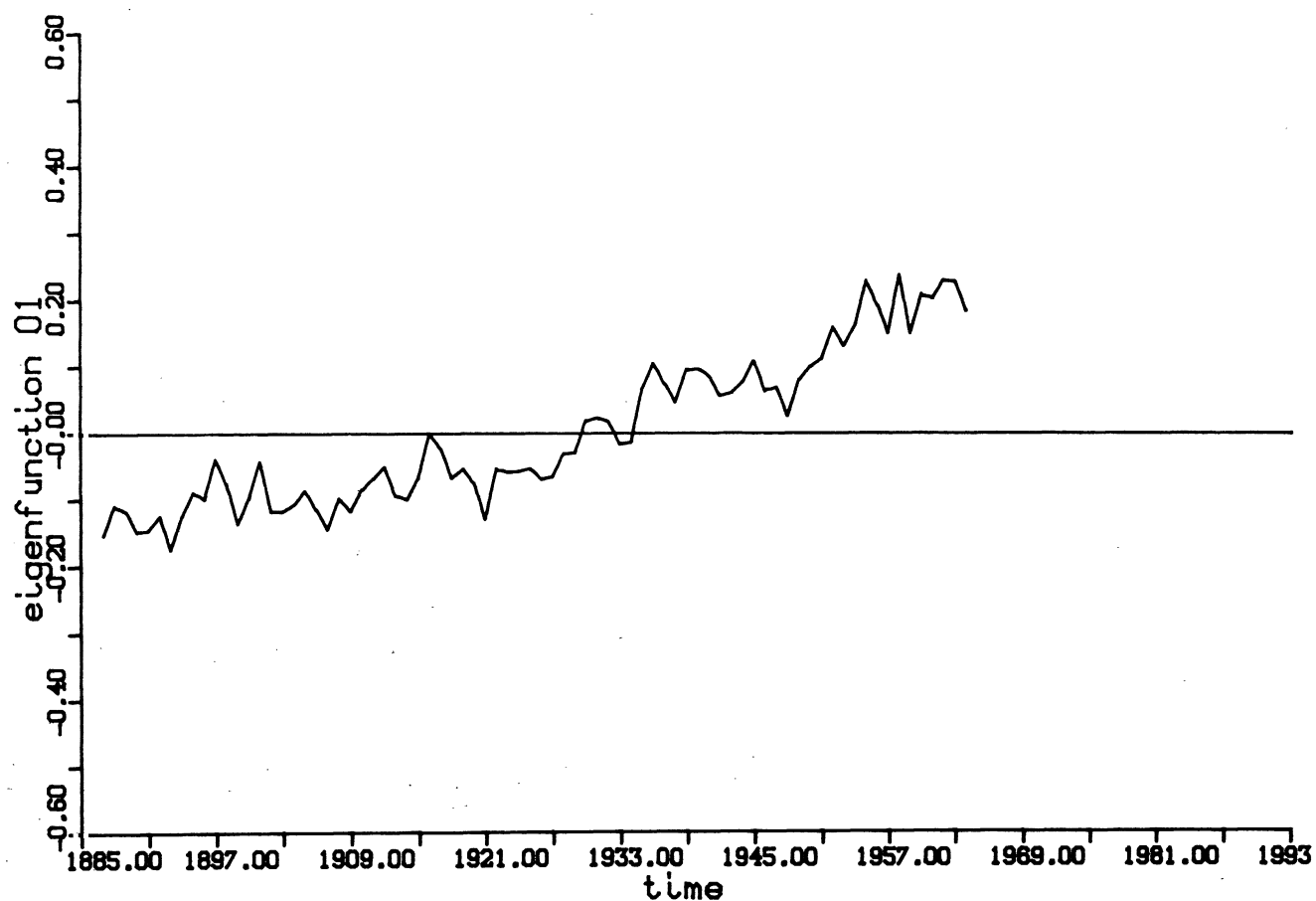
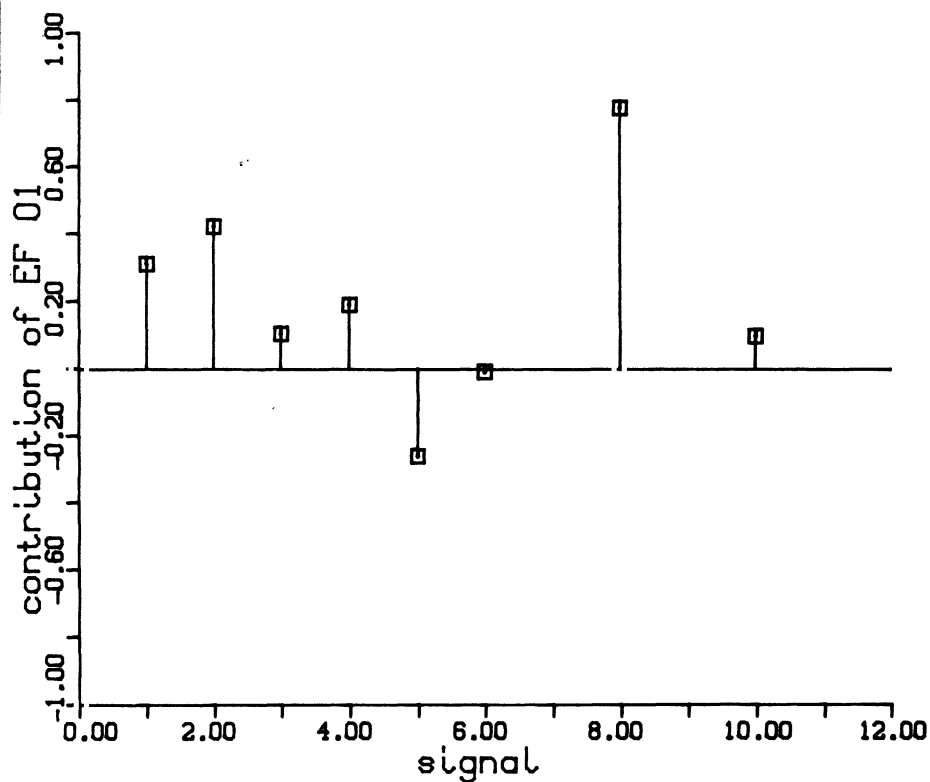
HOOFDCOMPONENT 4

$\sigma_4 = 23.7$

$\sigma_4^2 / \sigma^2 = 0.022$

Delft Hydraulics

Fig. [4']



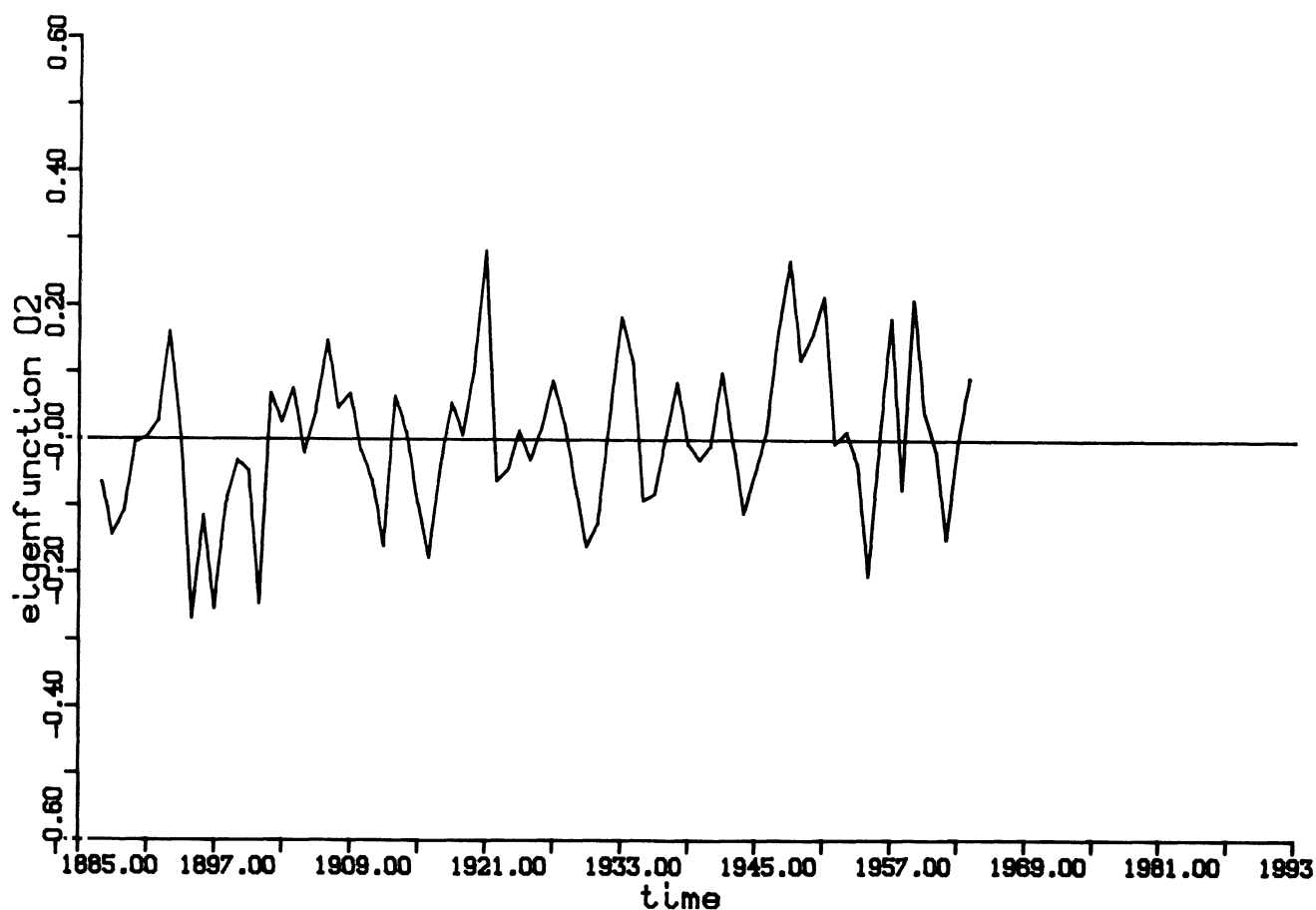
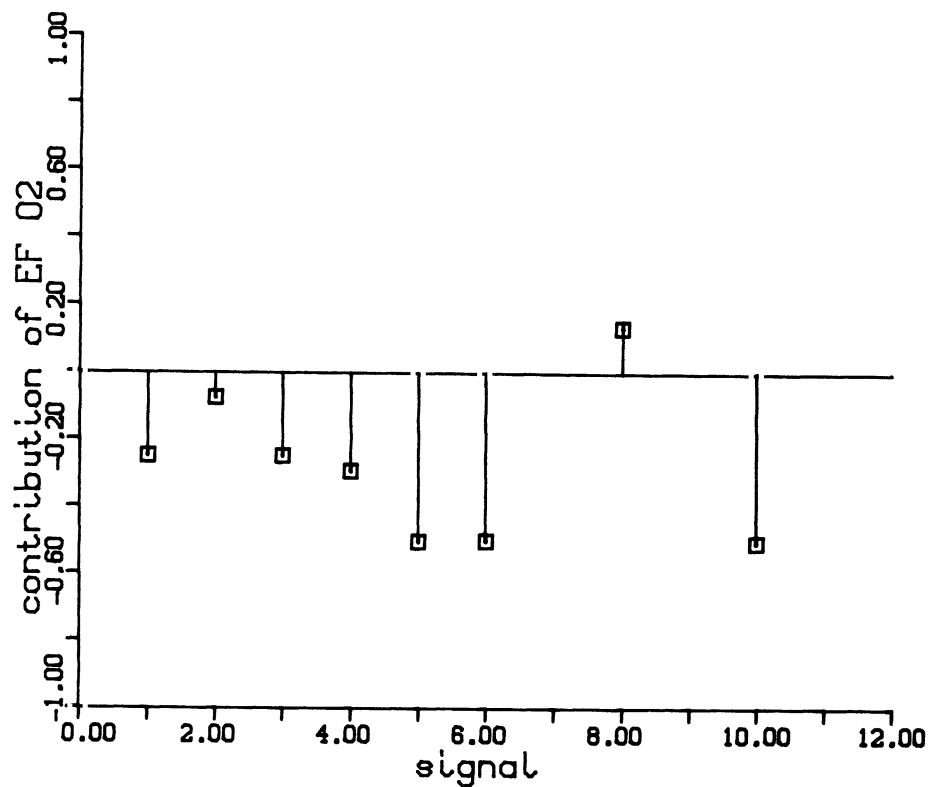
HOOFDCOMPONENT 1

$\sigma_1 = 65.9$

$\sigma_1^2 / \sigma^2 = 0.593$

Delft Hydraulics

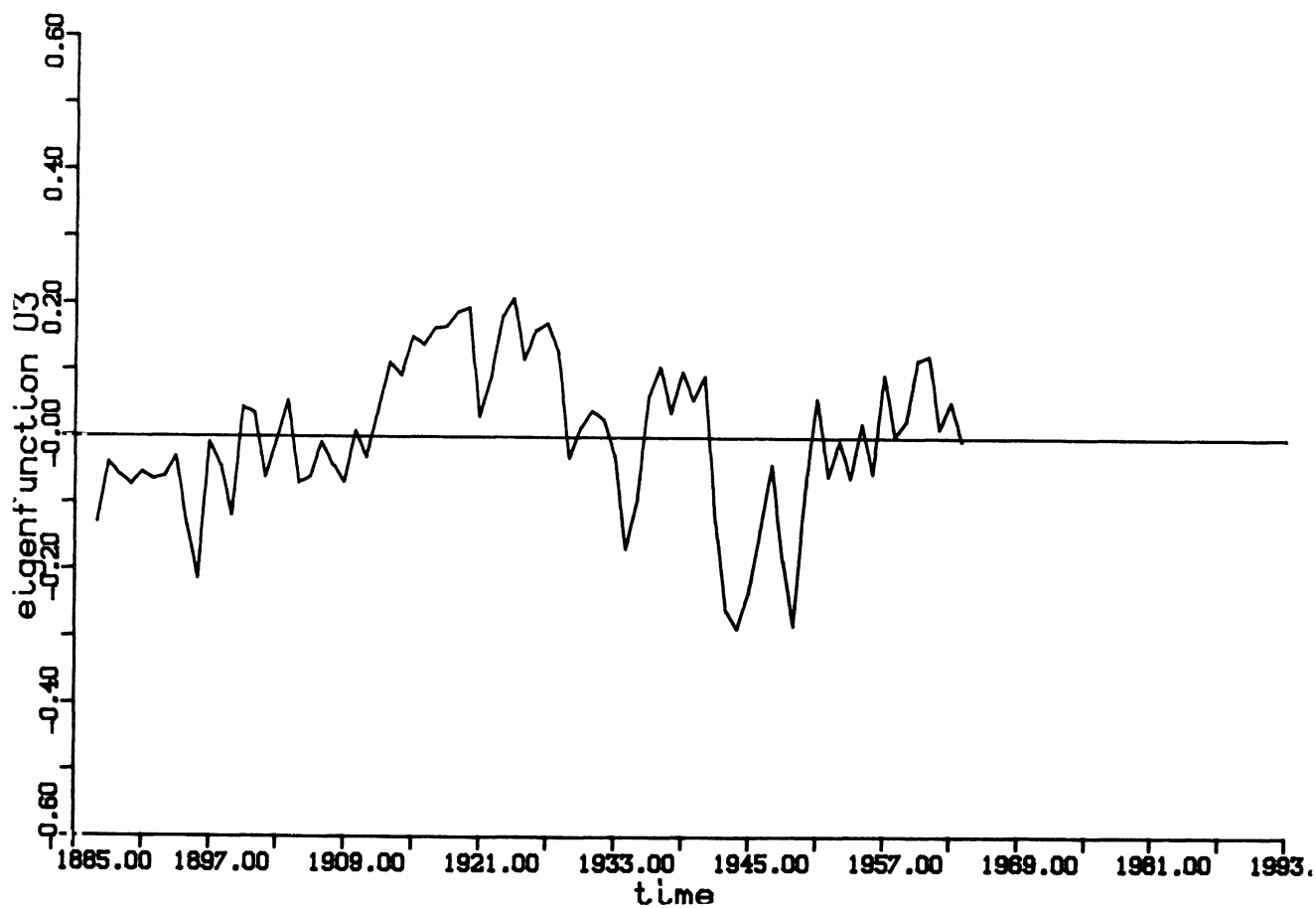
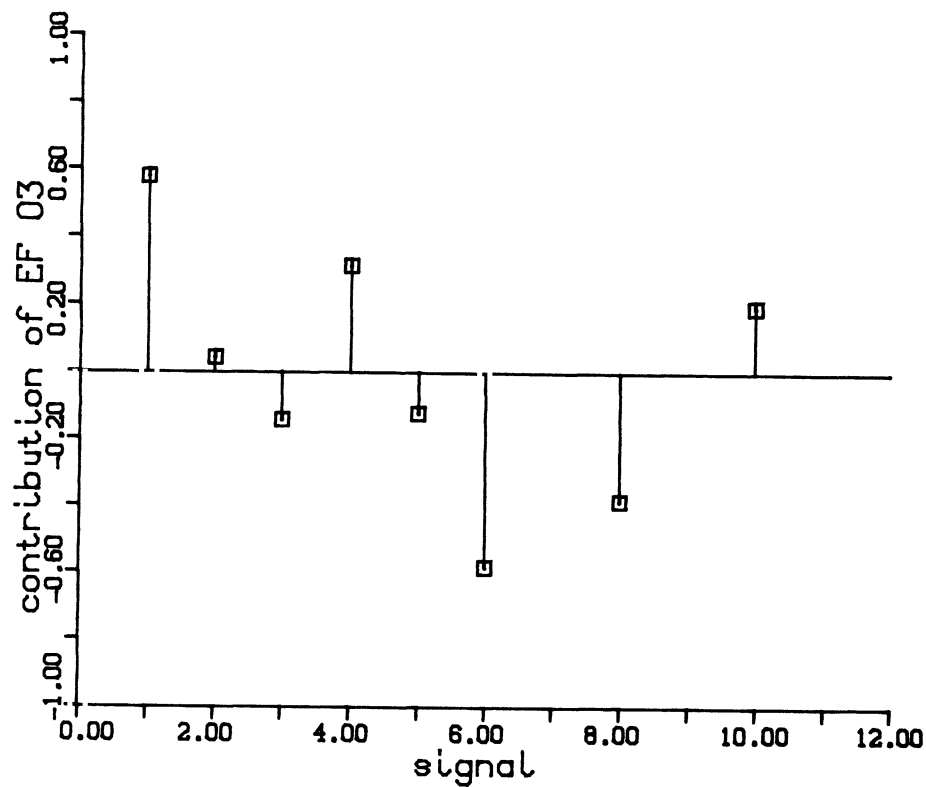
Fig. [5']



HOOFDCOMPONENT 2

$\sigma_2 = 37.8$

$\sigma_2^2 / \sigma^2 = 0.195$



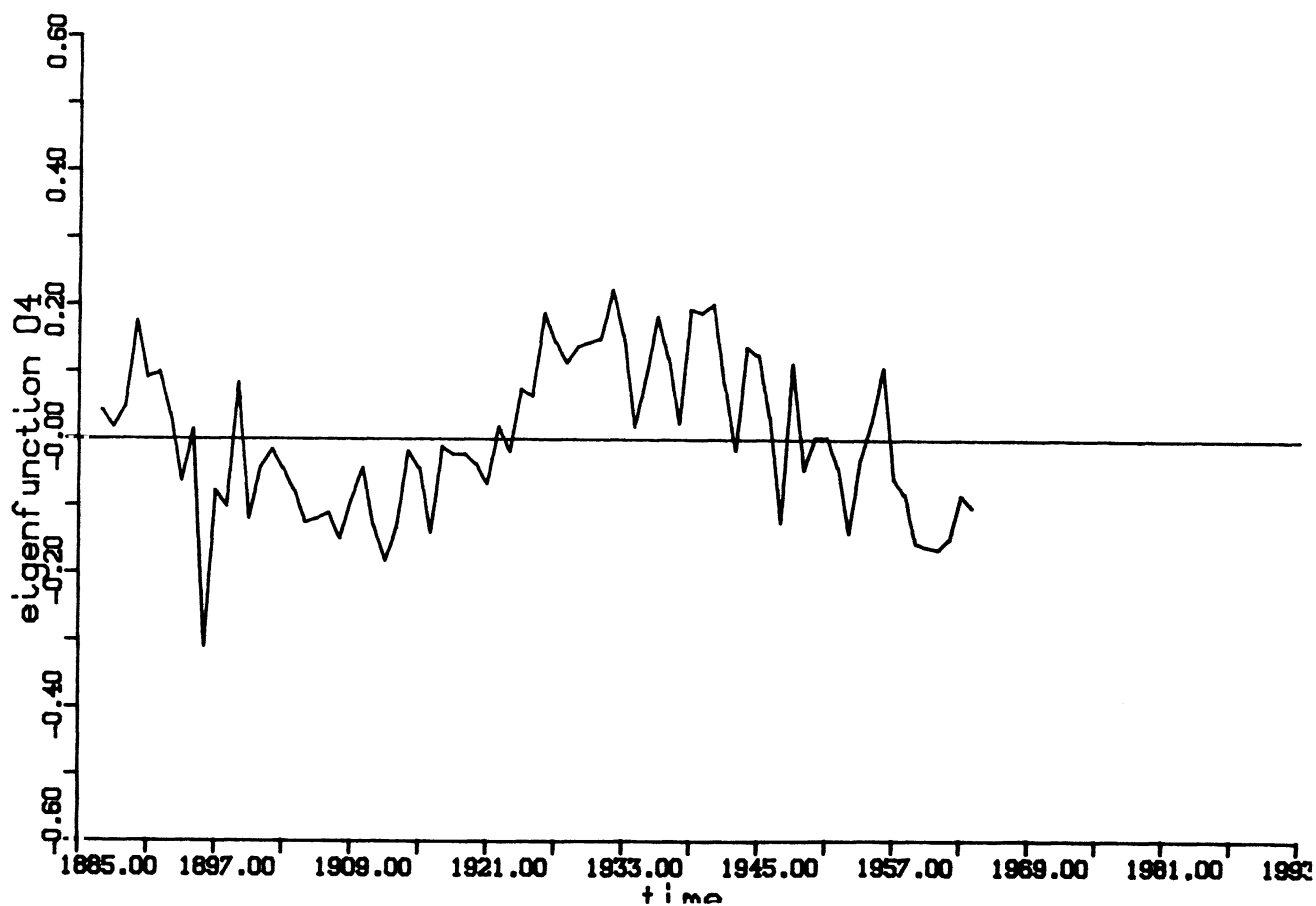
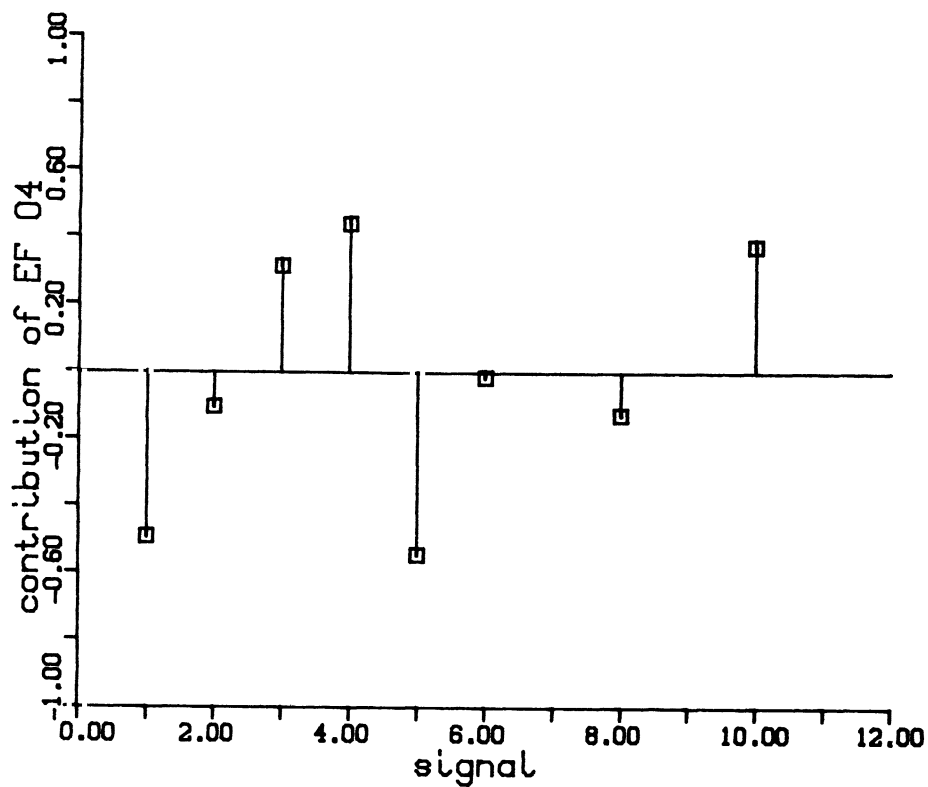
HOOFDCOMPONENT 3

$\sigma_3 = 21.2$

$\sigma_3^2/\sigma^2 = 0.061$

Delft Hydraulics

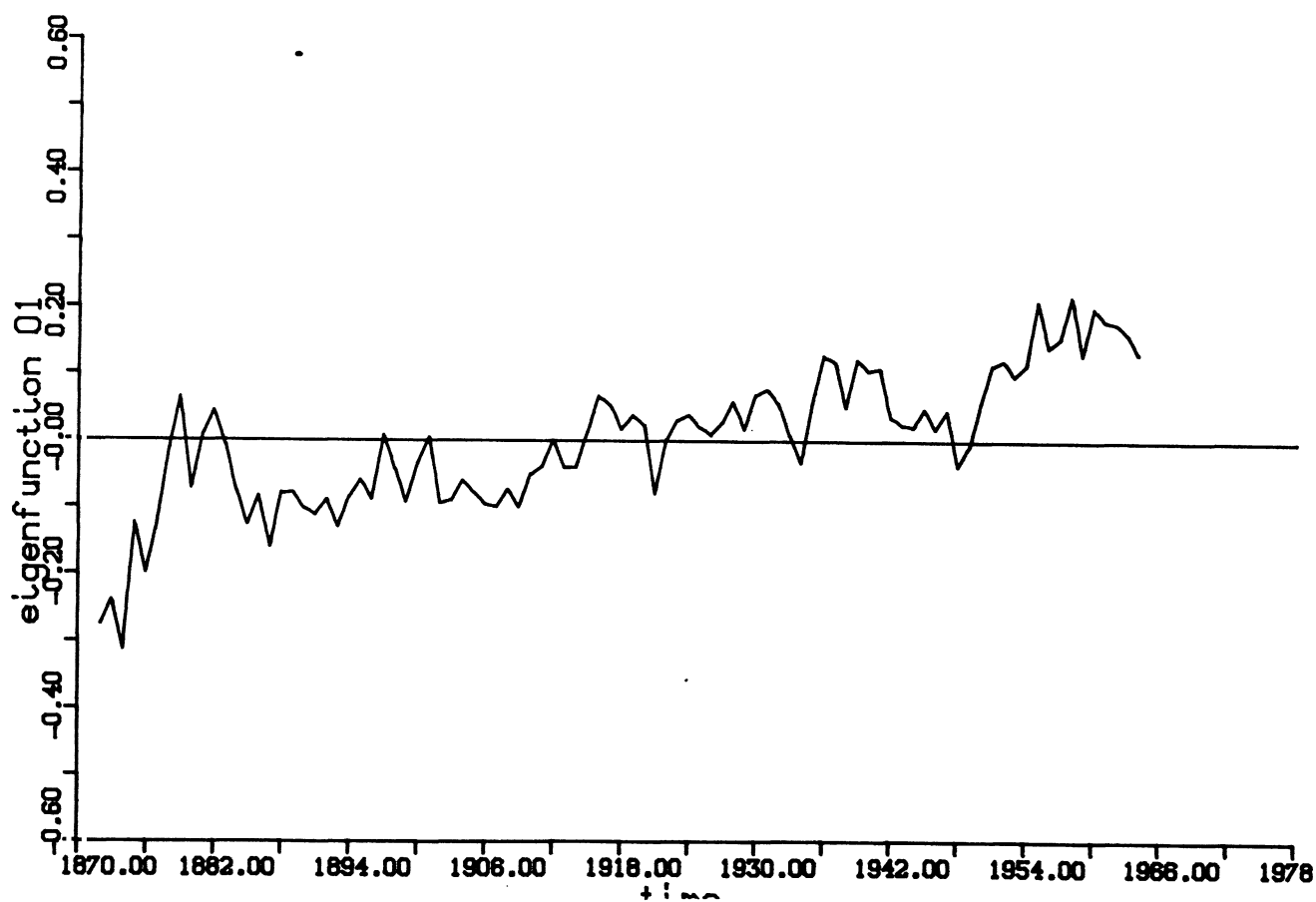
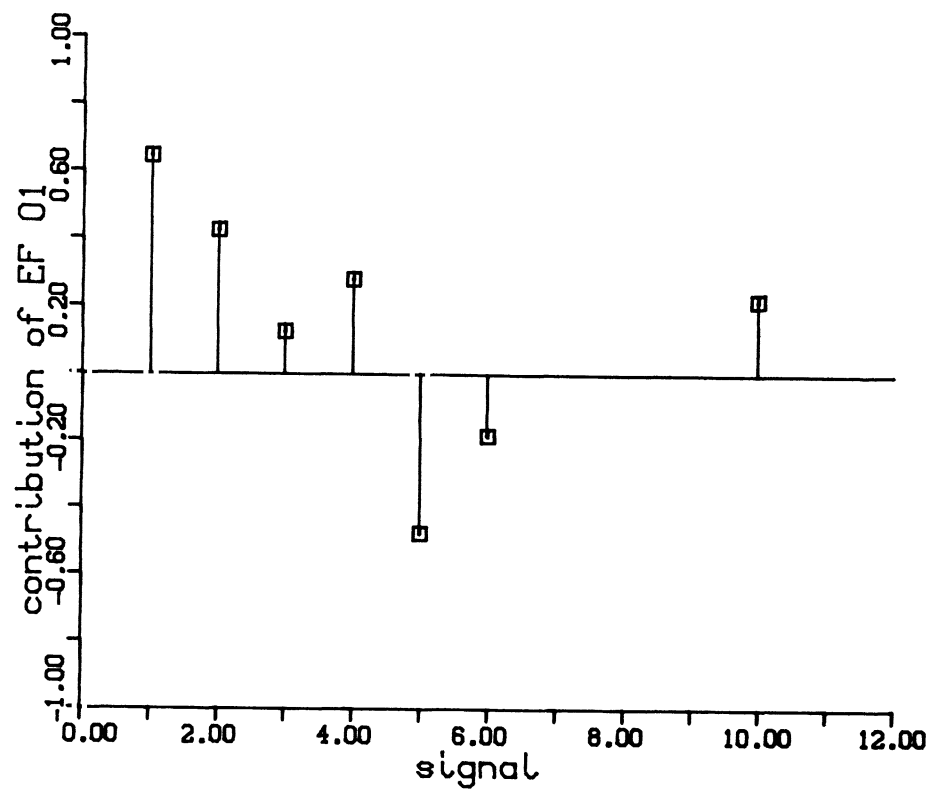
Fig. [5']



HOOFDCOMPONENT 4

$\sigma_4 = 18.3$

$\sigma_4^2 / \sigma^2 = 0.046$



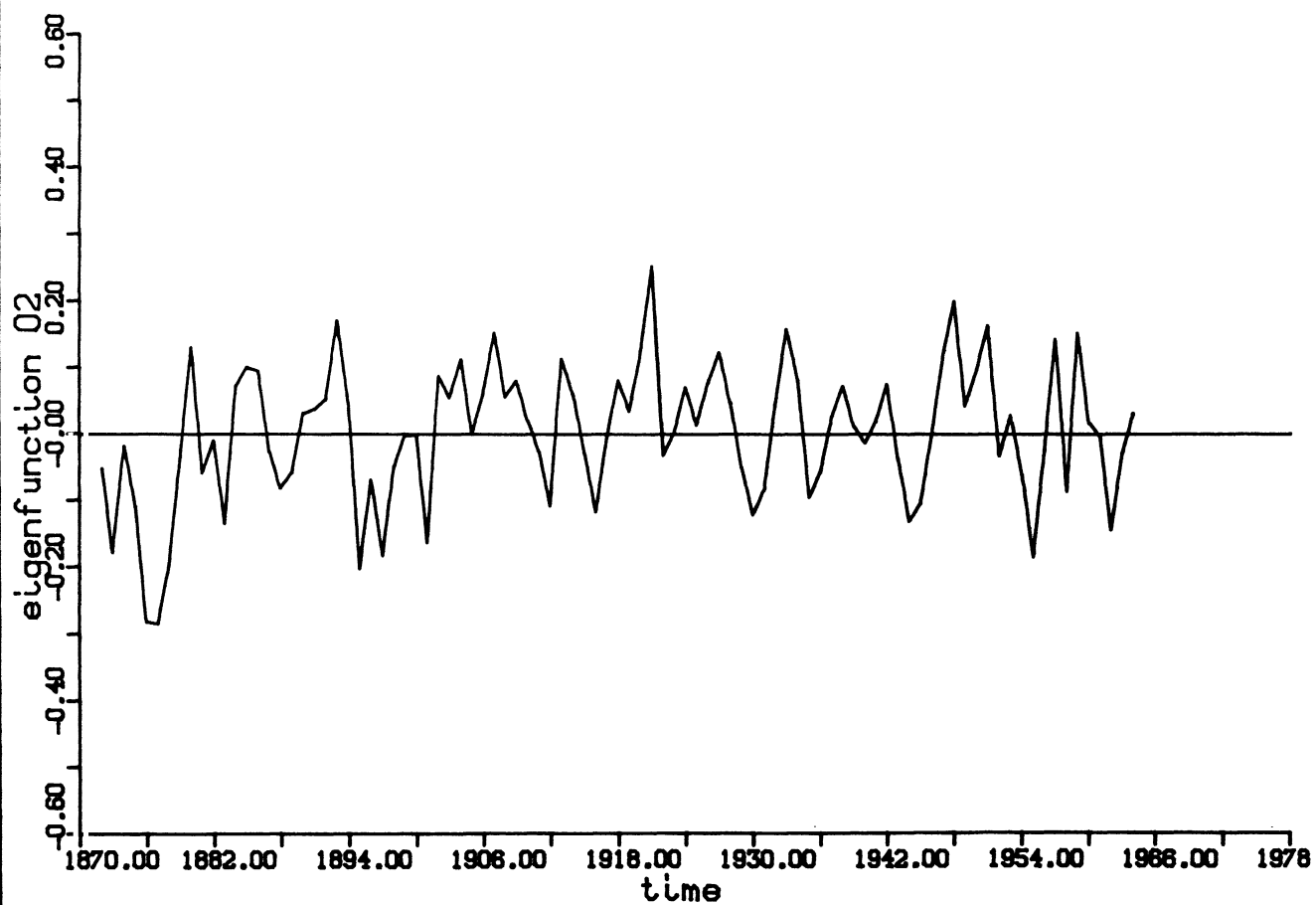
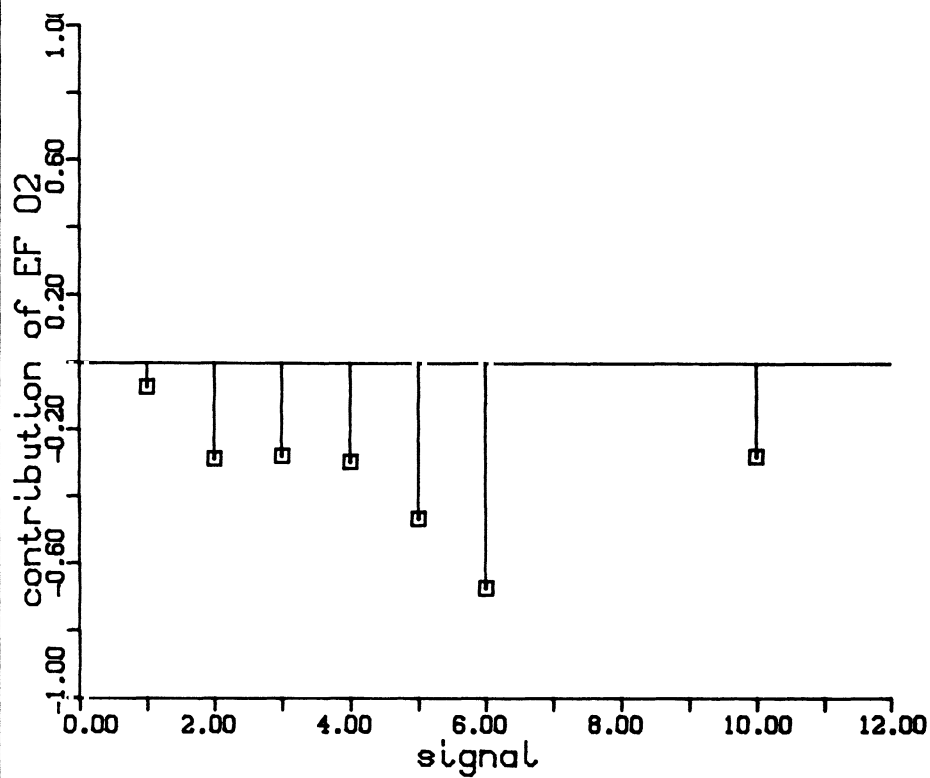
HOOFDCOMPONENT 1

$\sigma_1 = 53.2$

$\sigma_1^2 / \sigma^2 = 0.415$

Delft Hydraulics

Fig. [6']



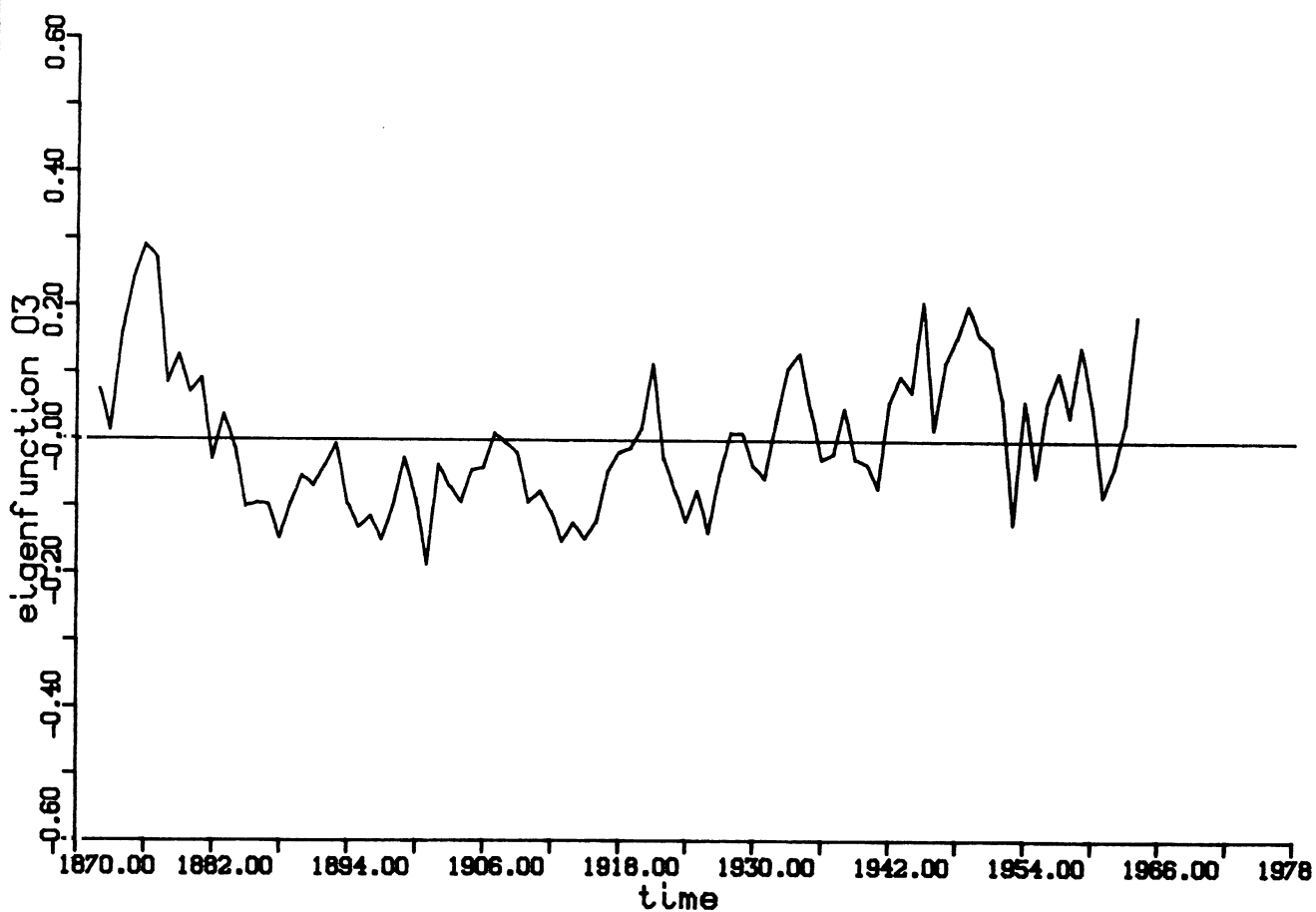
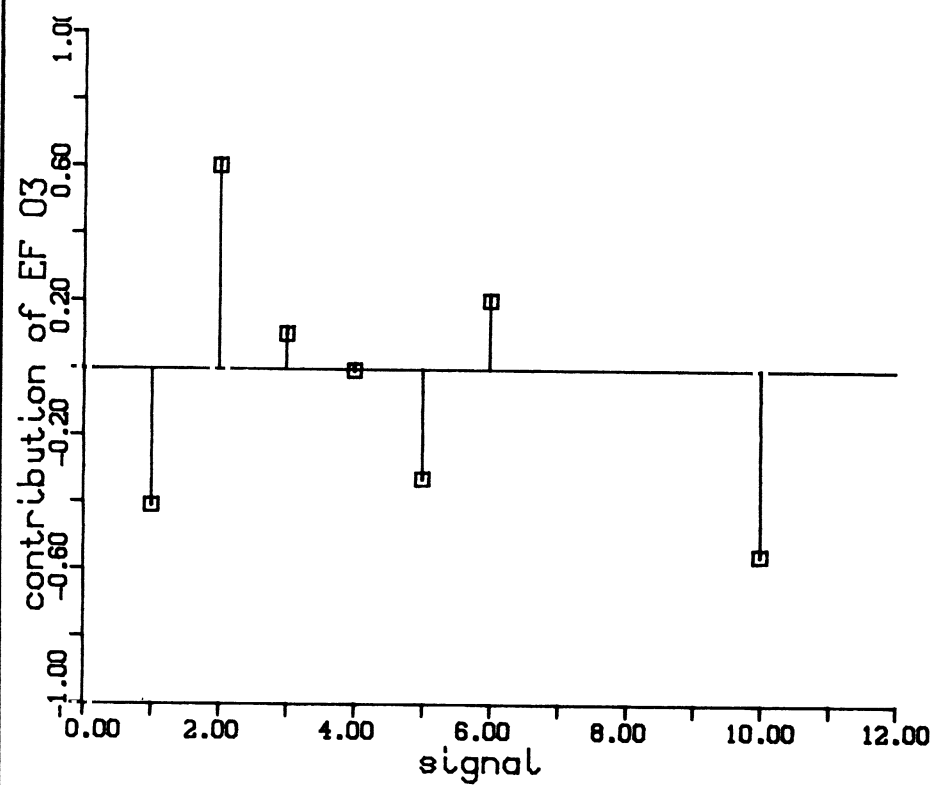
HOOFDCOMPONENT 2

$$\sigma_2 = 44.0$$

$$\sigma_2^2 / \sigma^2 = 0.283$$

Delft Hydraulics

Fig. [6']



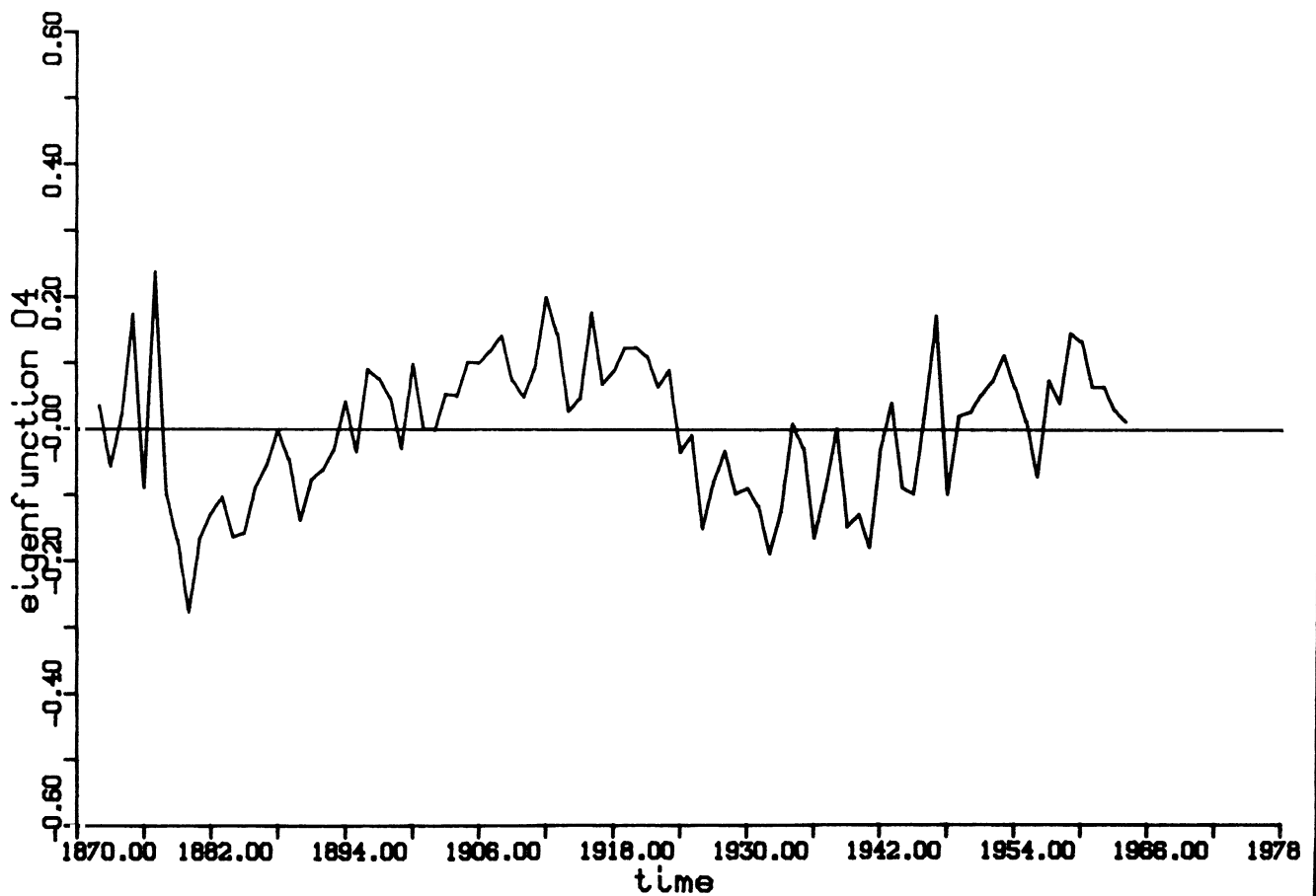
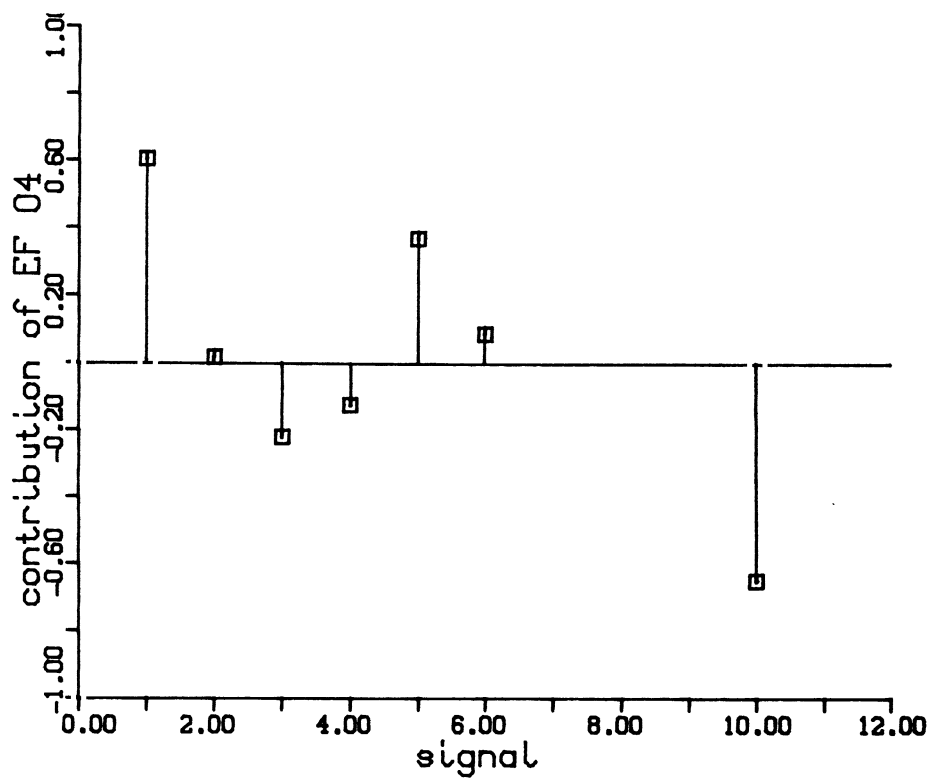
HOOFDCOMPONENT 3

$$\sigma_3 = 27.8$$

$$\sigma_3^2 / \sigma^2 = 0.113$$

Delft Hydraulics

Fig. [6']



HOOFDCOMPONENT 4

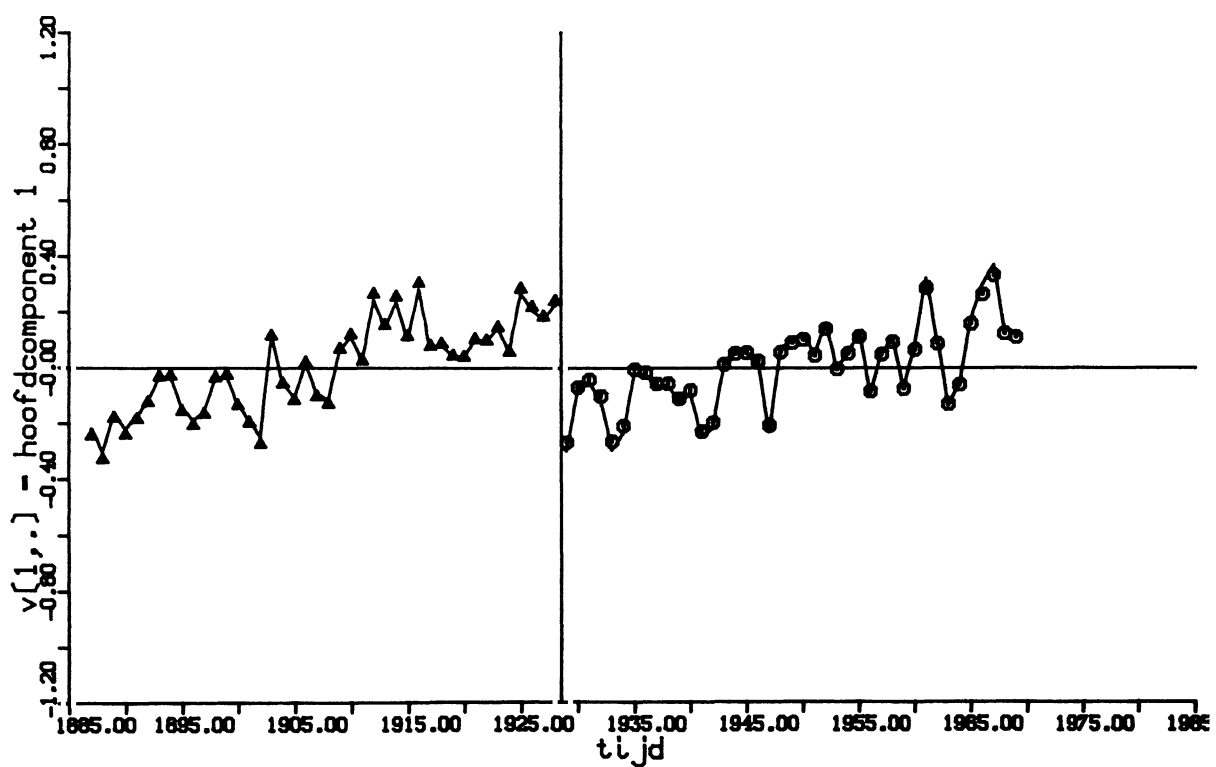
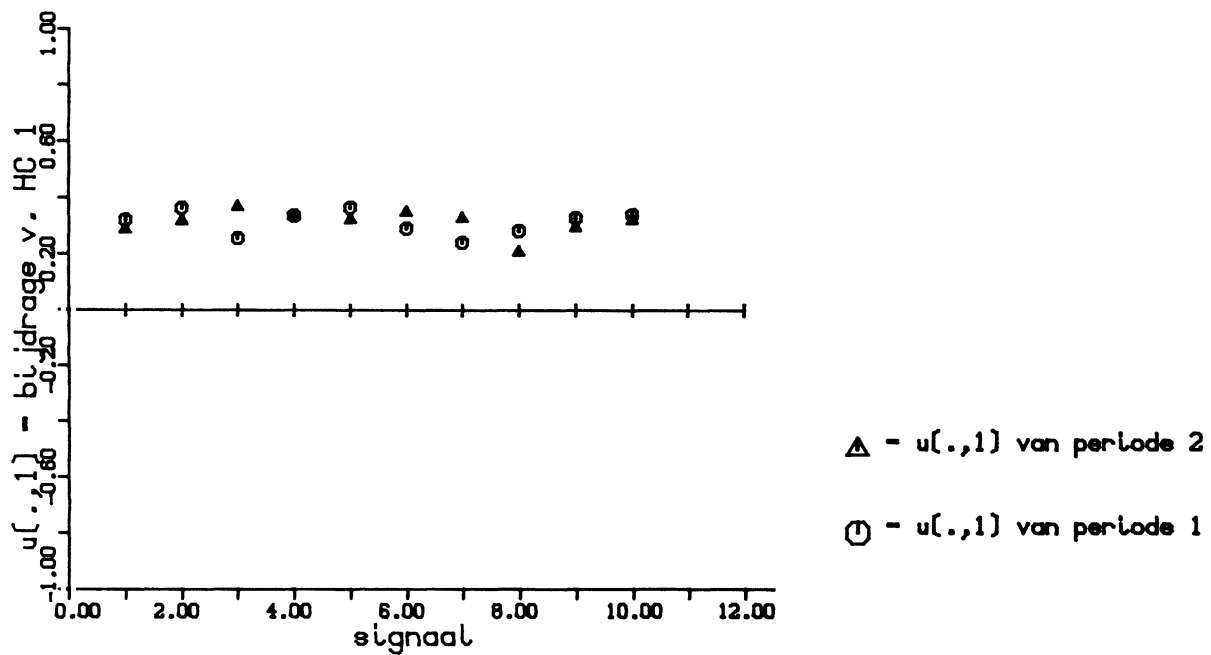
$$\sigma_4 = 21.7$$

$$\sigma_4^2 / \sigma^2 = 0.069$$

Delft Hydraulics

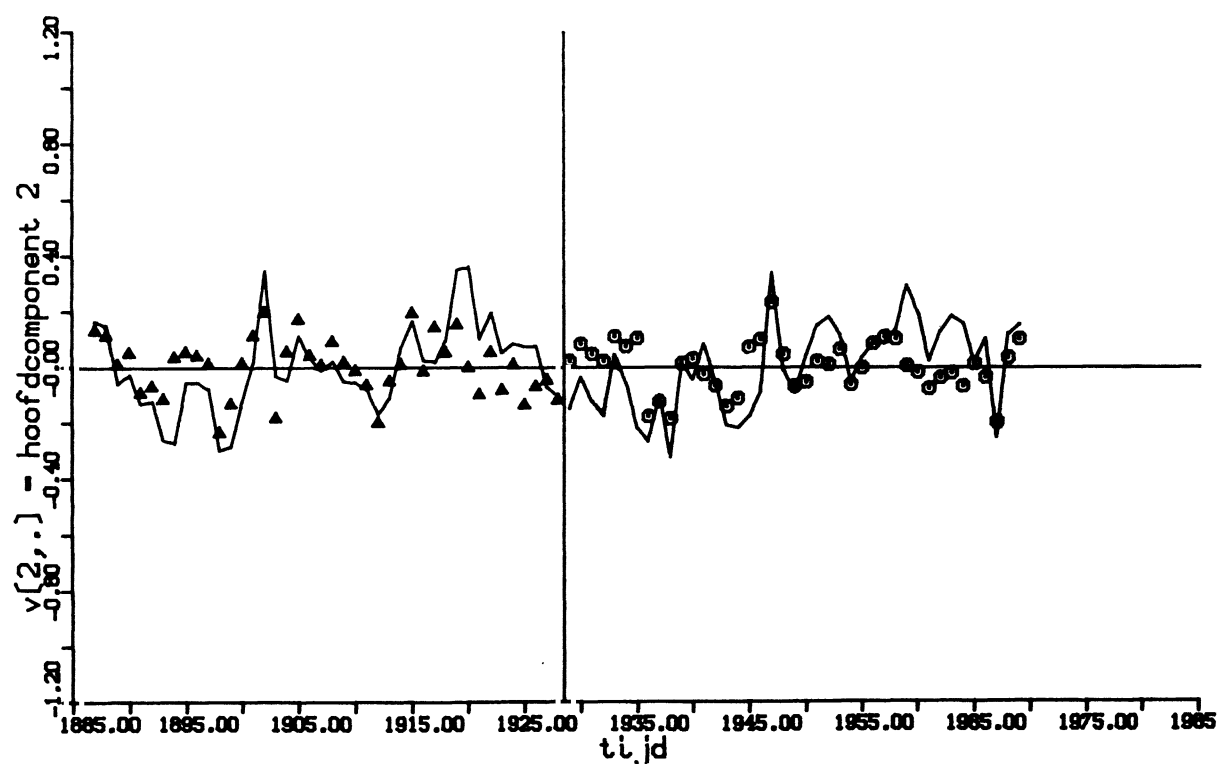
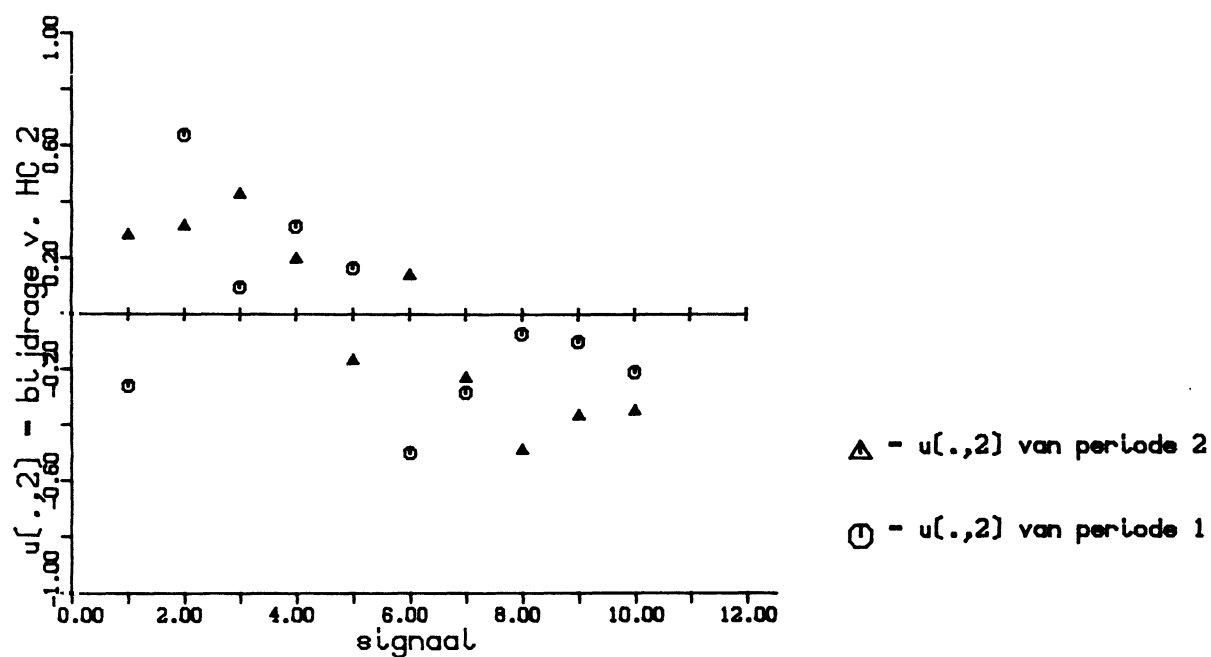
Fig. [6']

stabiliteit van hoofdcomponenten-ontbinding



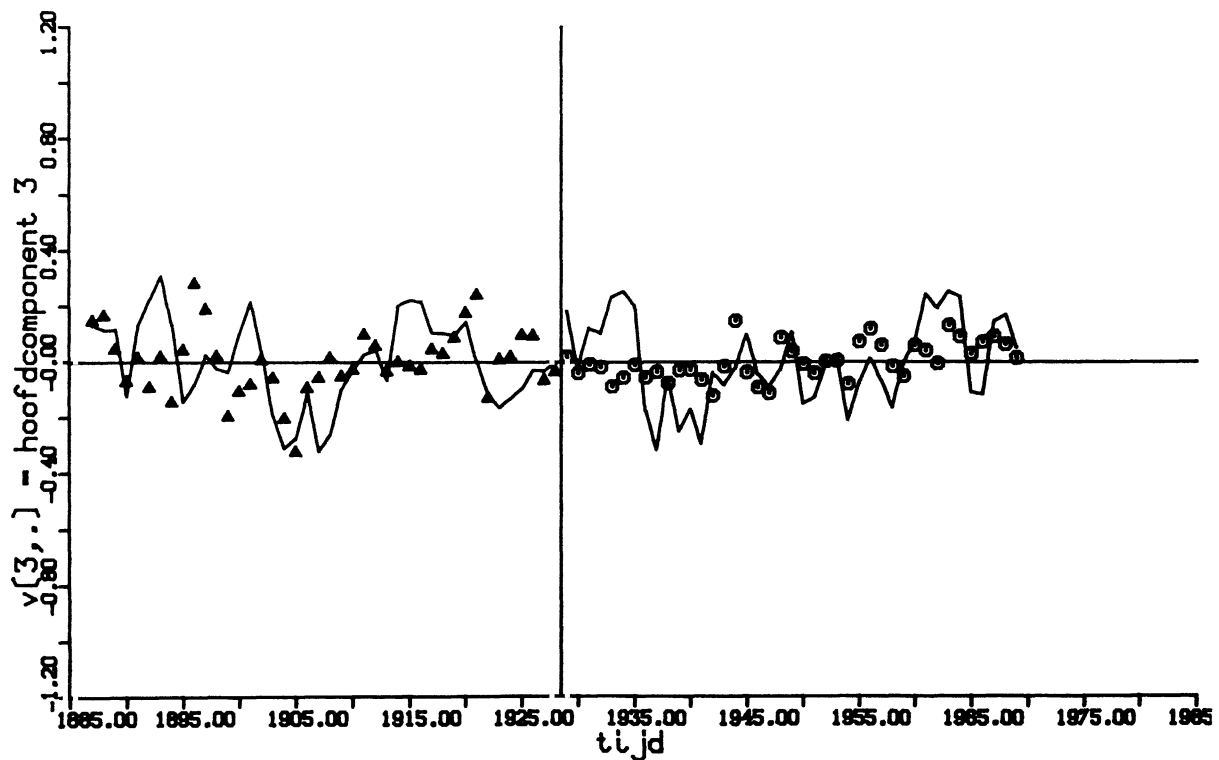
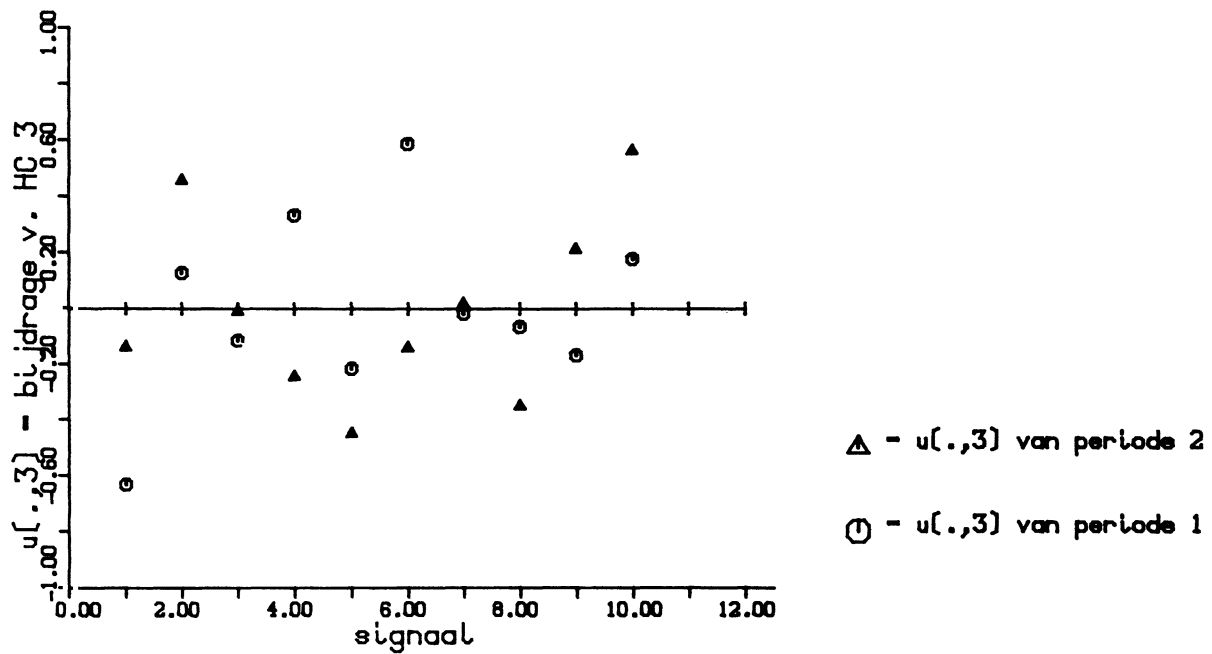
HOOFDCOMPONENT 1

stabiliteit van hoofdcomponenten-ontbinding



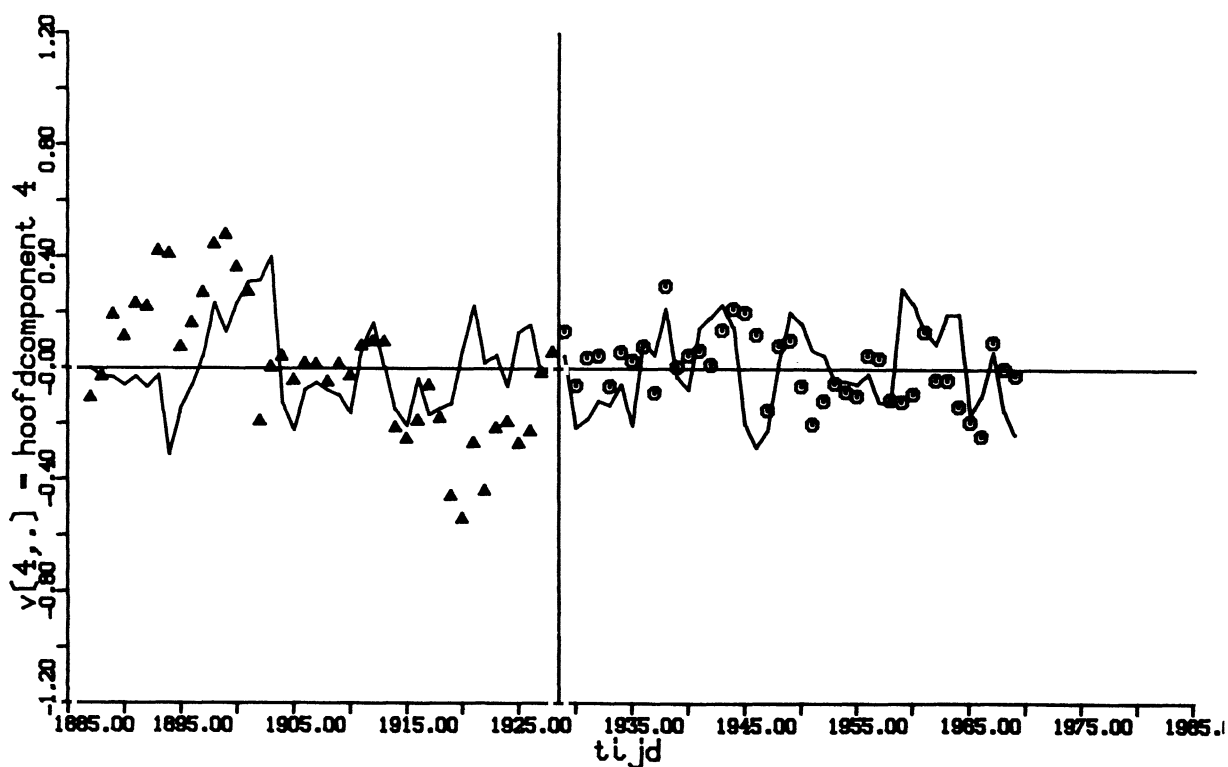
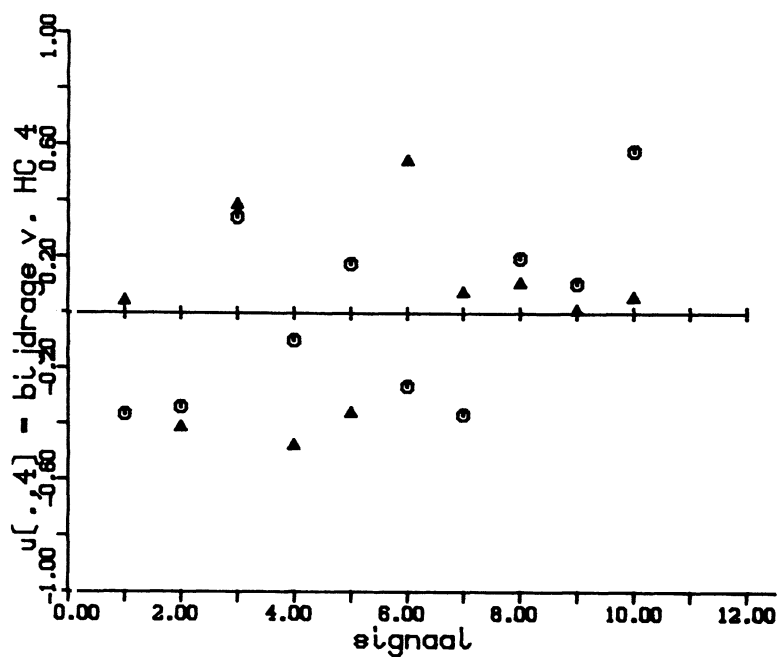
HOOFDCOMPONENT 2

stabiliteit van hoofdcomponenten-ontbinding



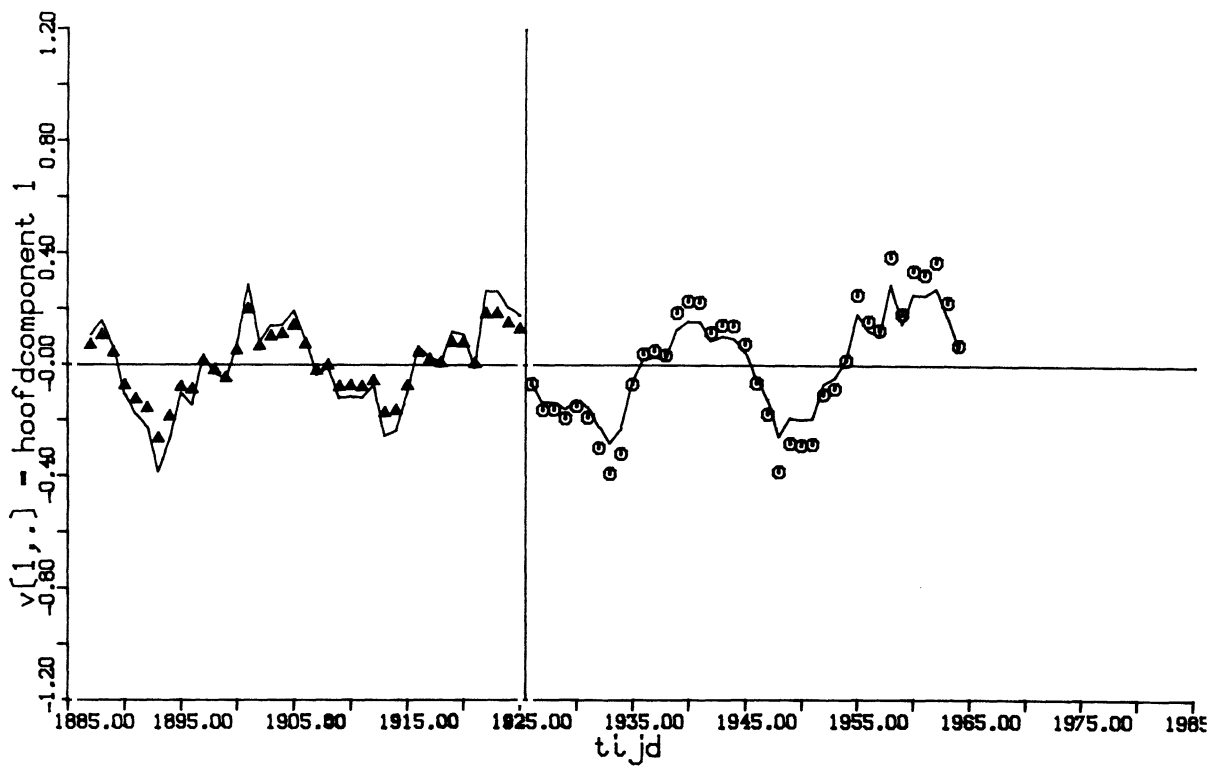
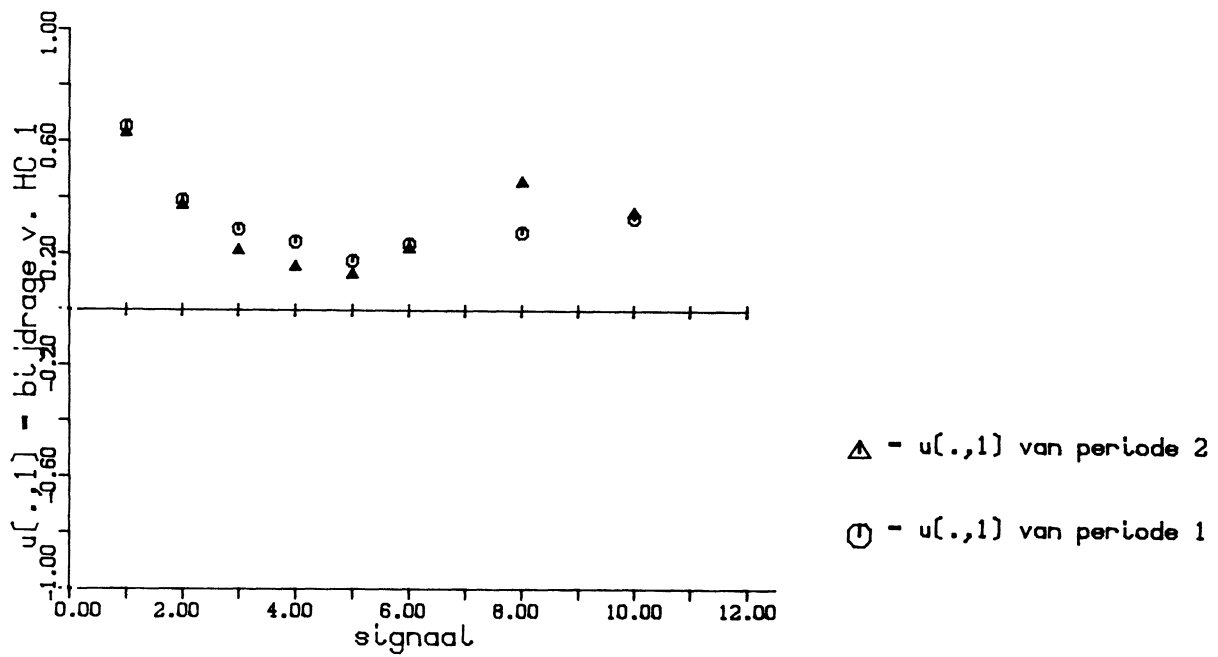
HOOFDCOMPONENT 3

stabiliteit van hoofdcomponenten-ontbinding



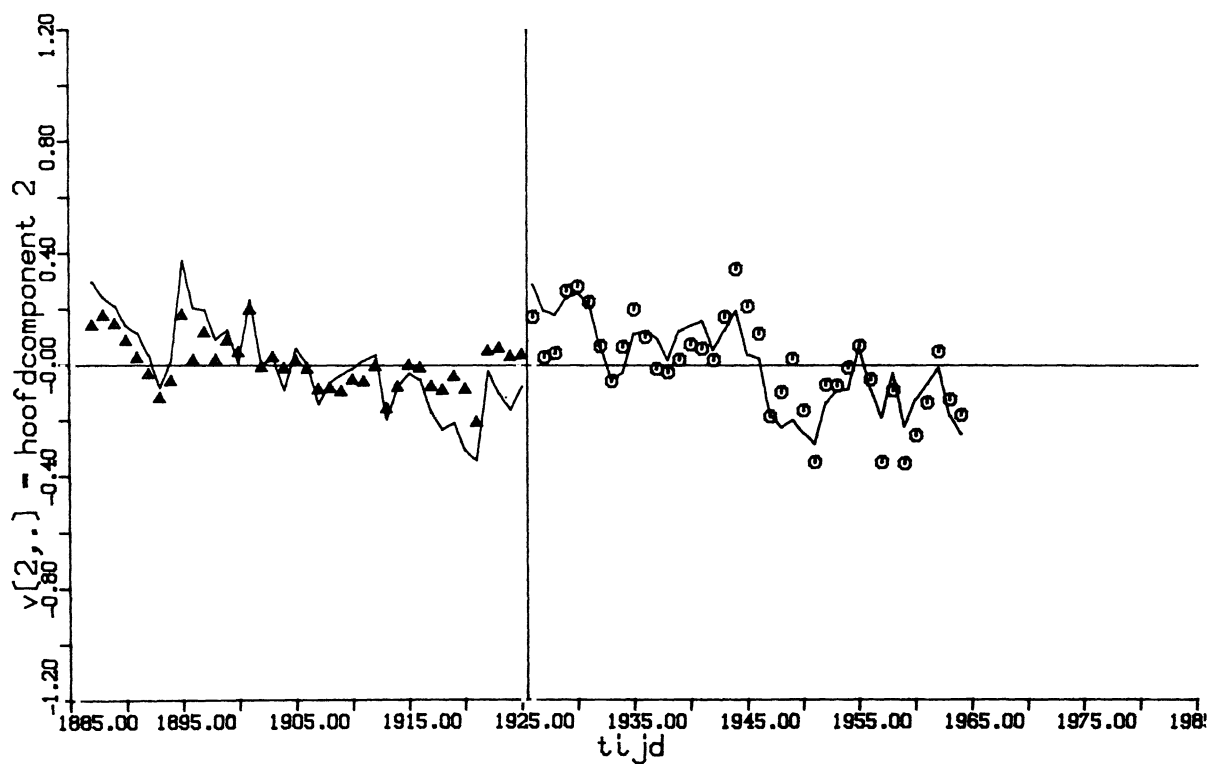
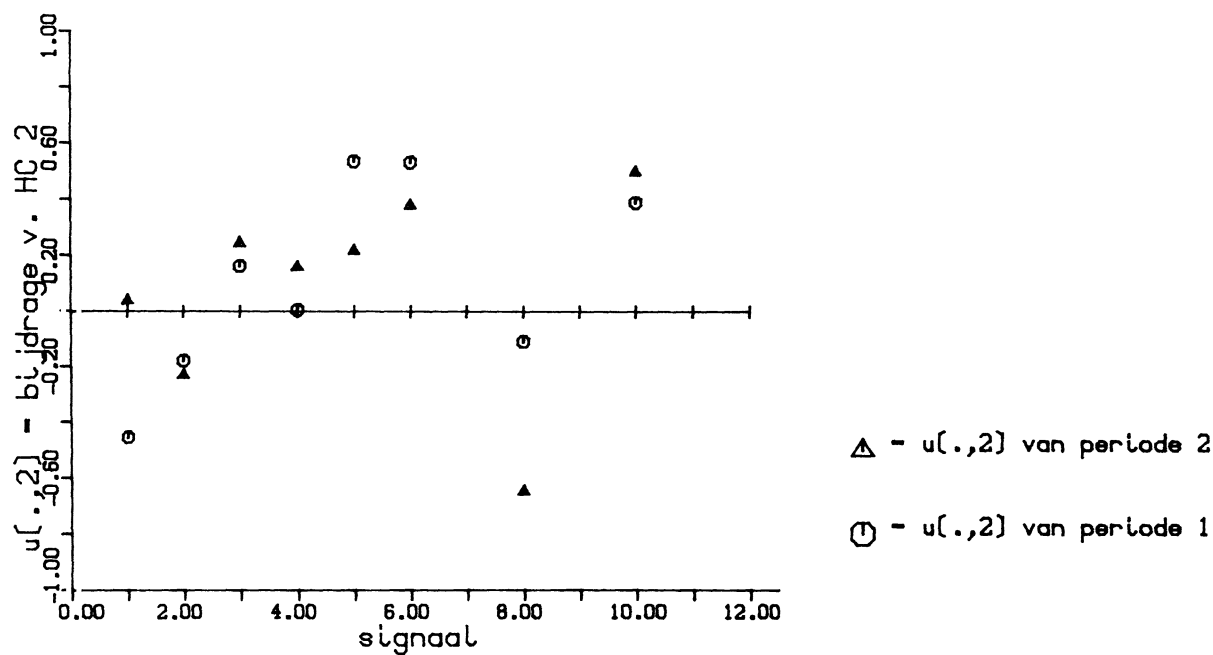
HOOFDCOMPONENT 4

stabiliteit van hoofdcomponenten-ontbinding



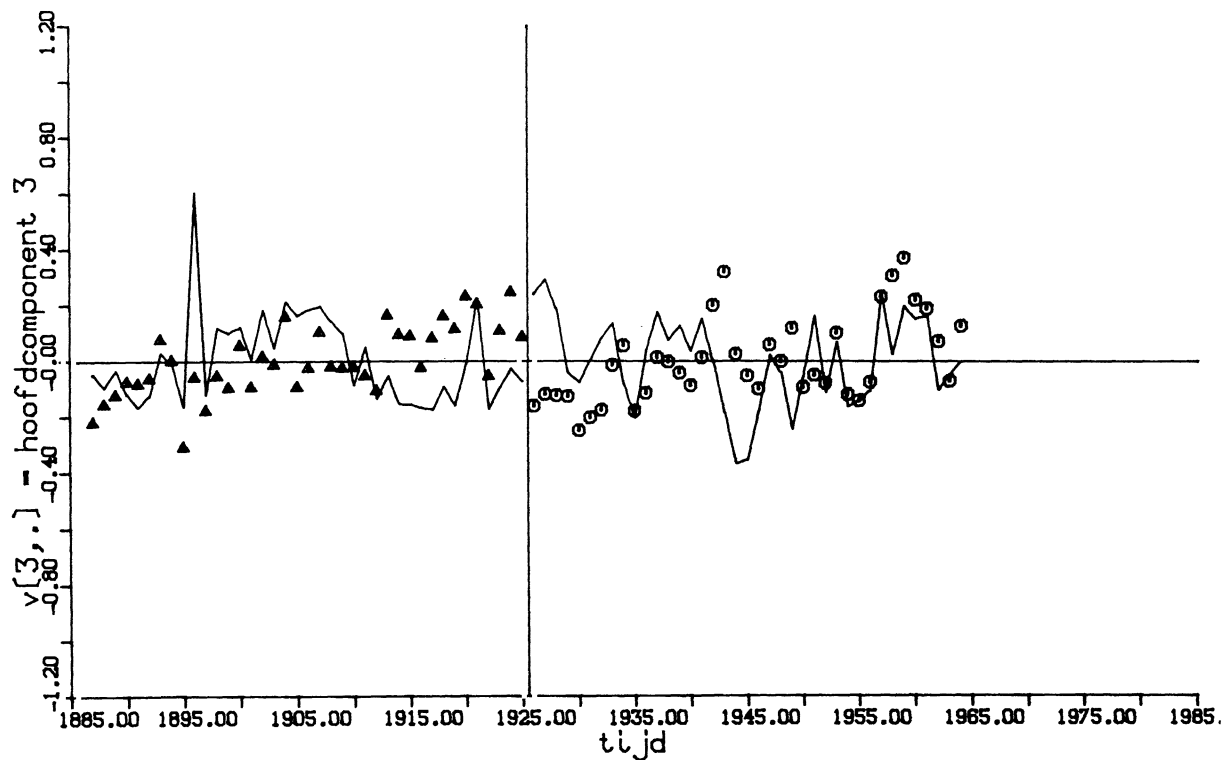
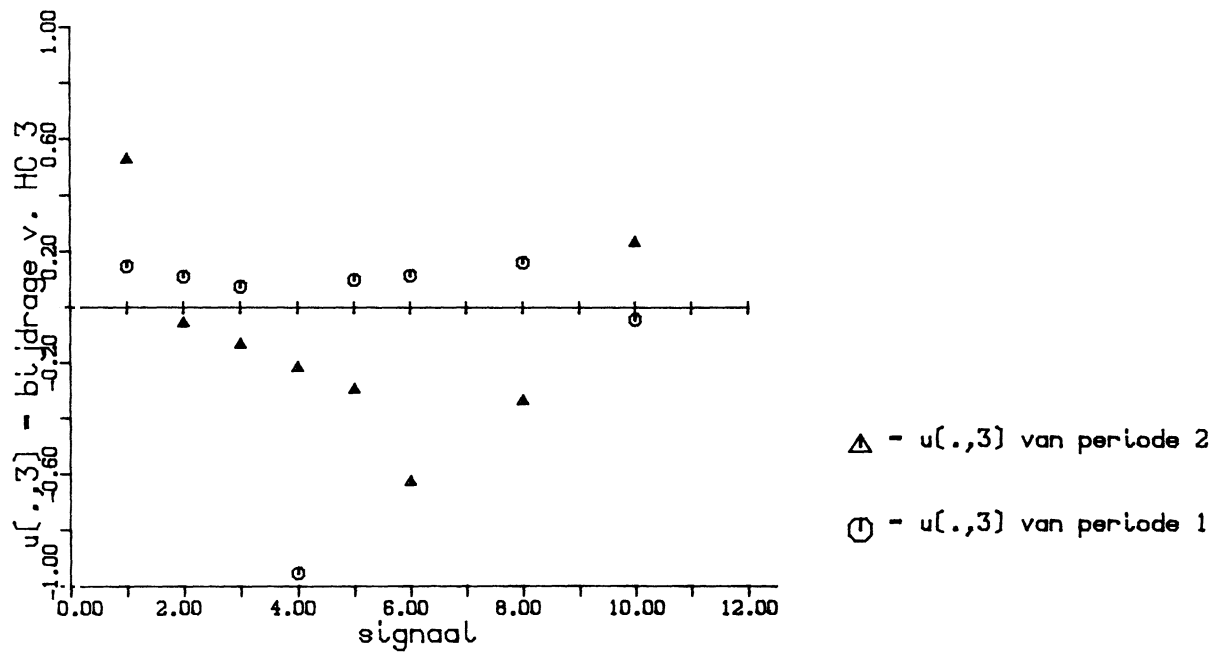
HOOFDCOMPONENT 1

stabiliteit van hoofcomponenten-ontbinding



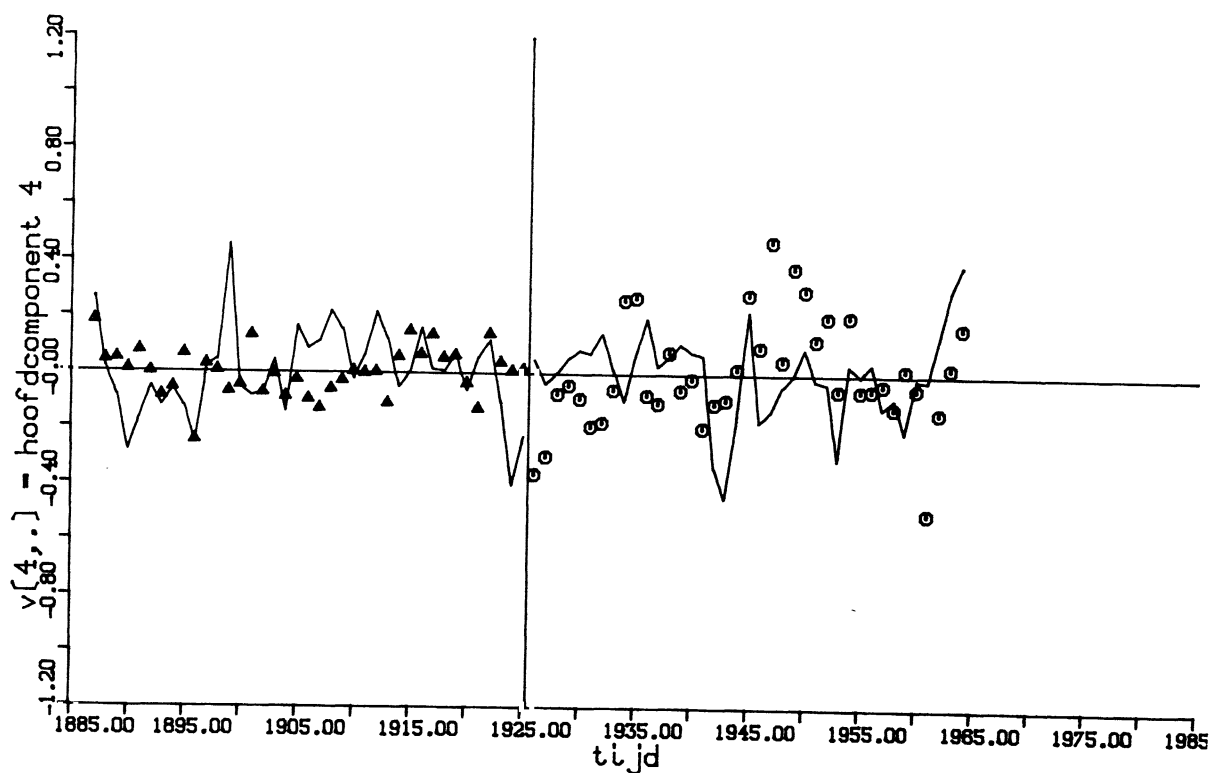
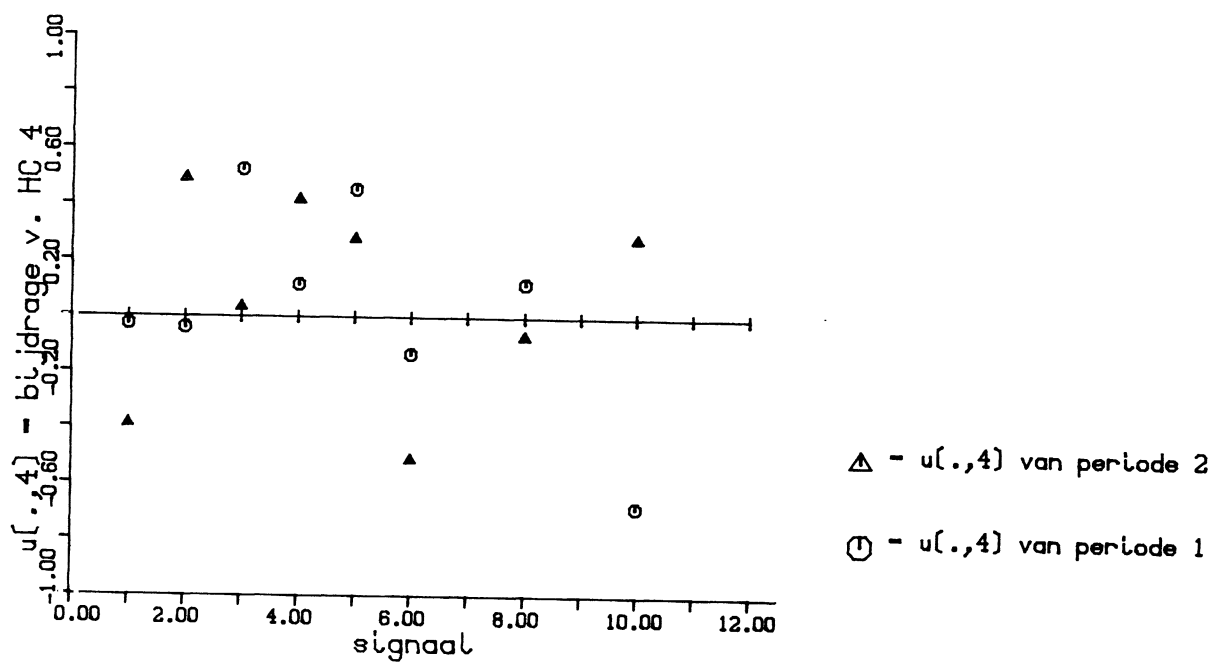
HOOFDCOMPONENT 2

stabiliteit van hoofcomponenten-ontbinding



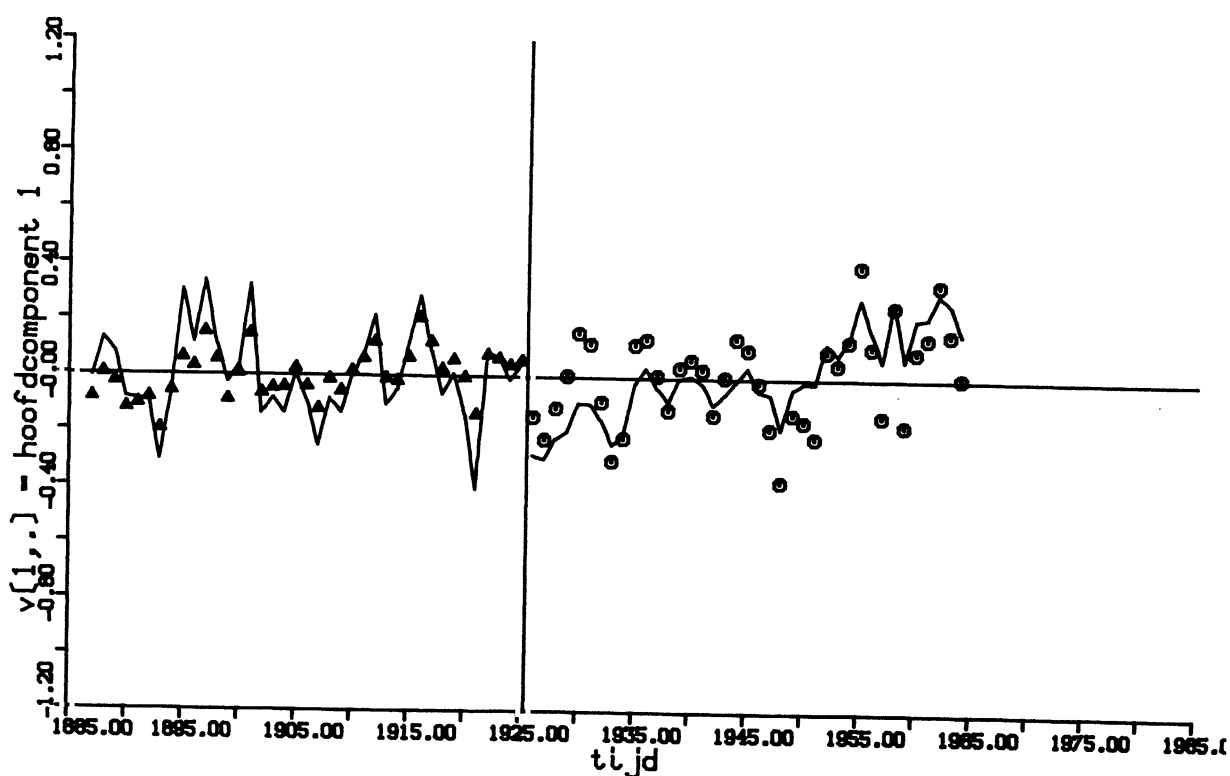
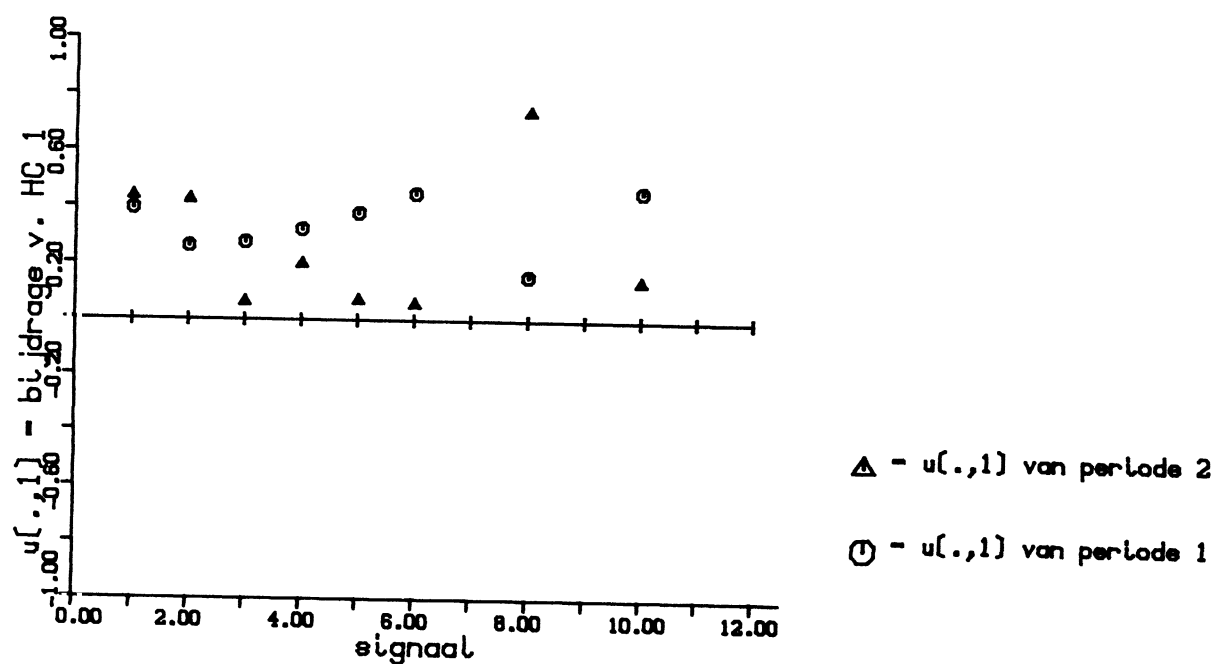
HOOFDCOMPONENT 3

stabiliteit van hoofcomponenten-ontbinding



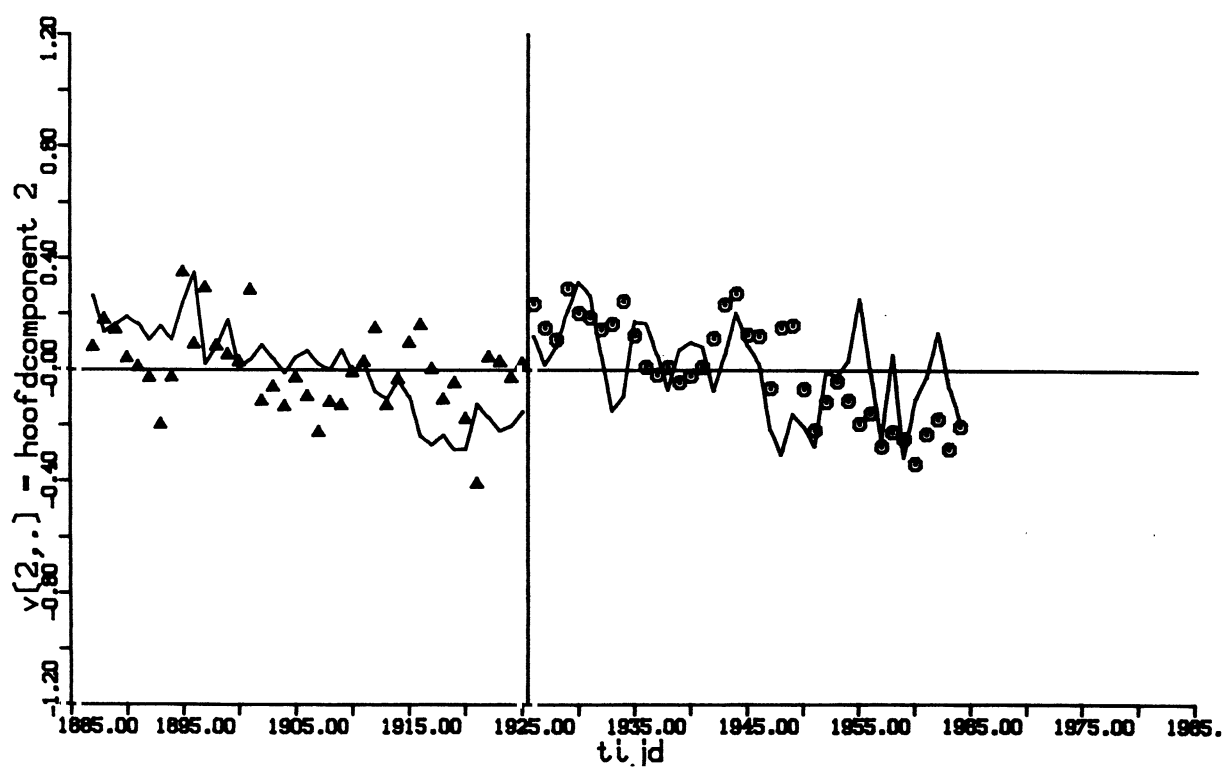
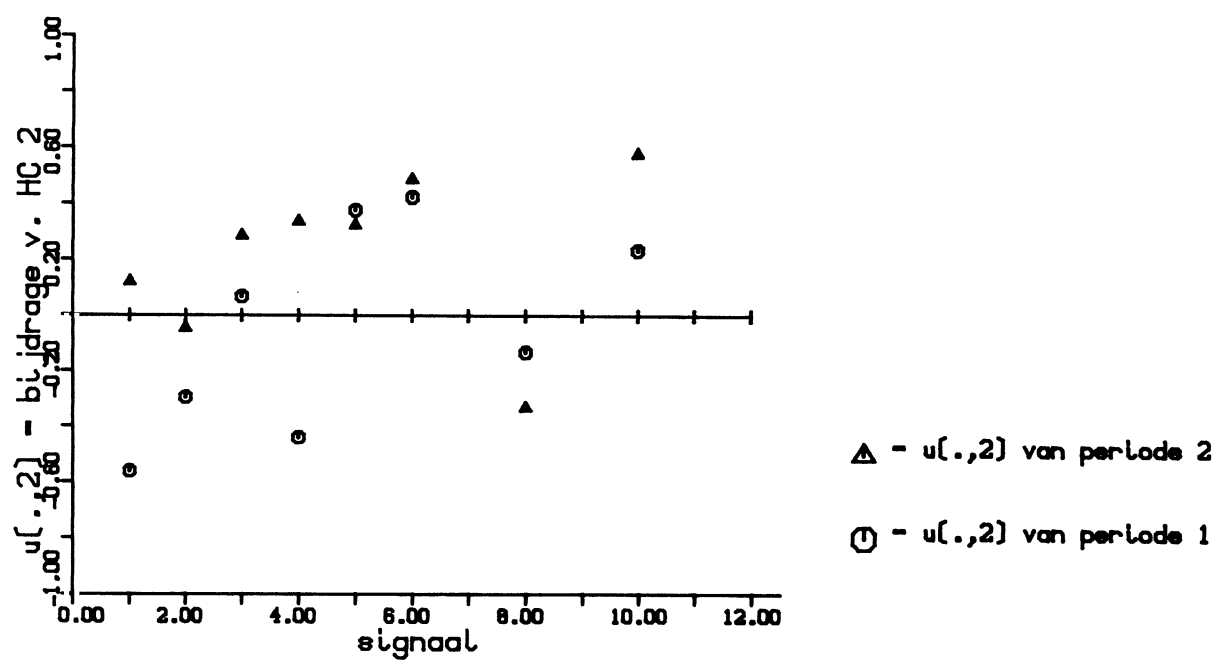
HOOFDCOMPONENT 4

stabiliteit van hoofdcomponenten-ontbinding



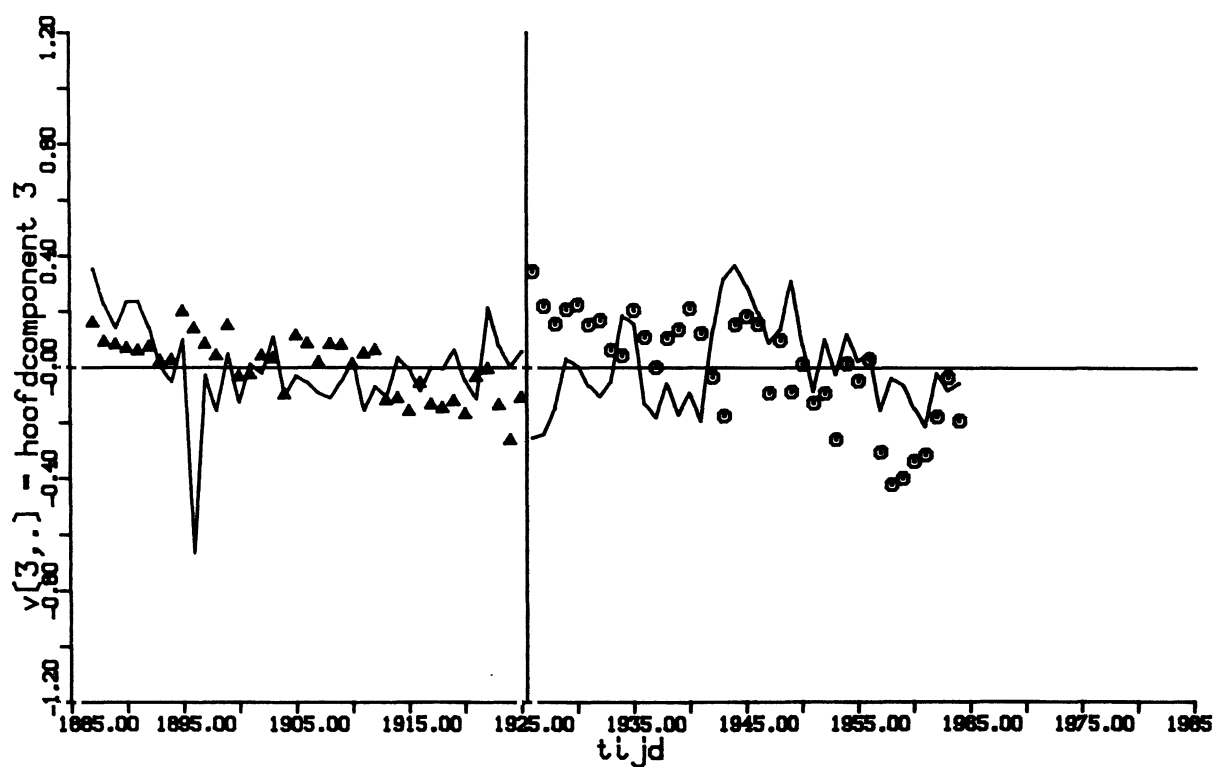
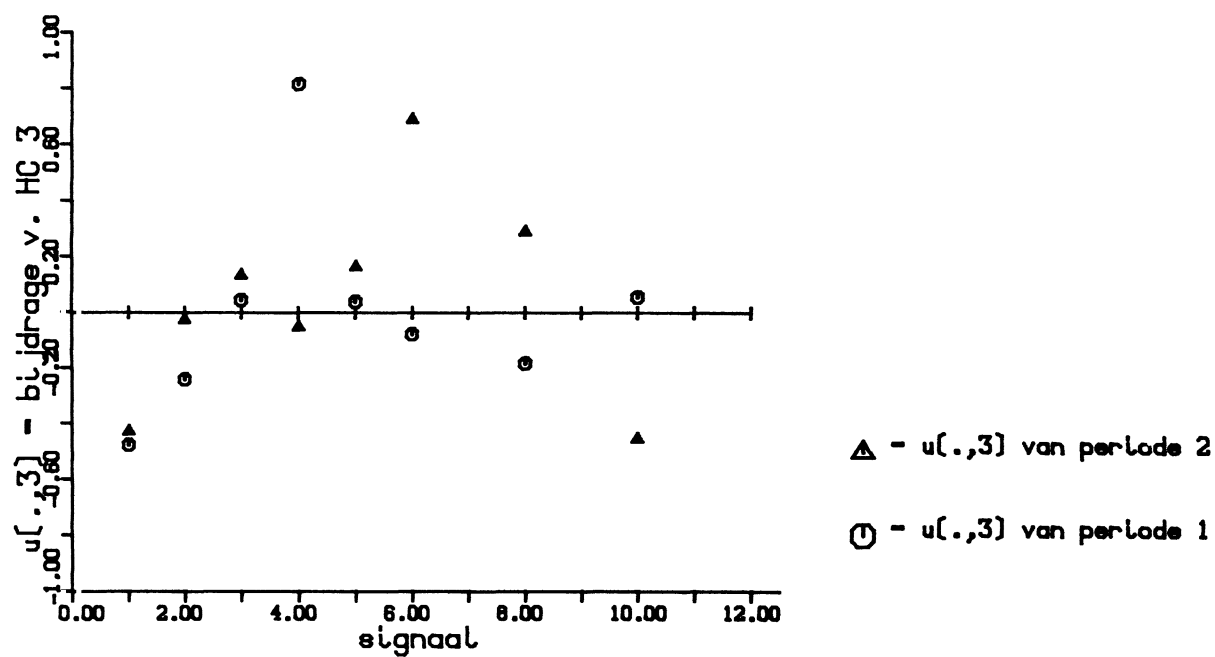
HOOFDCOMPONENT 1

stabiliteit van hoofdcomponenten-ontbinding



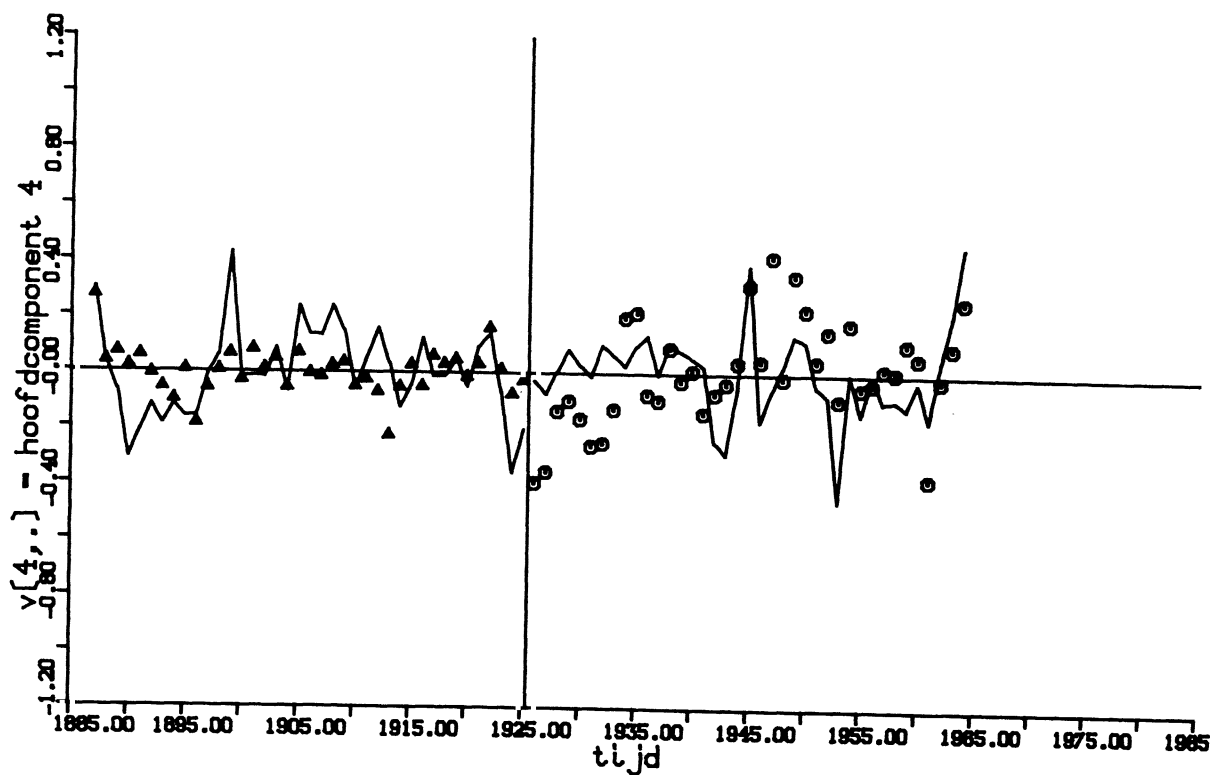
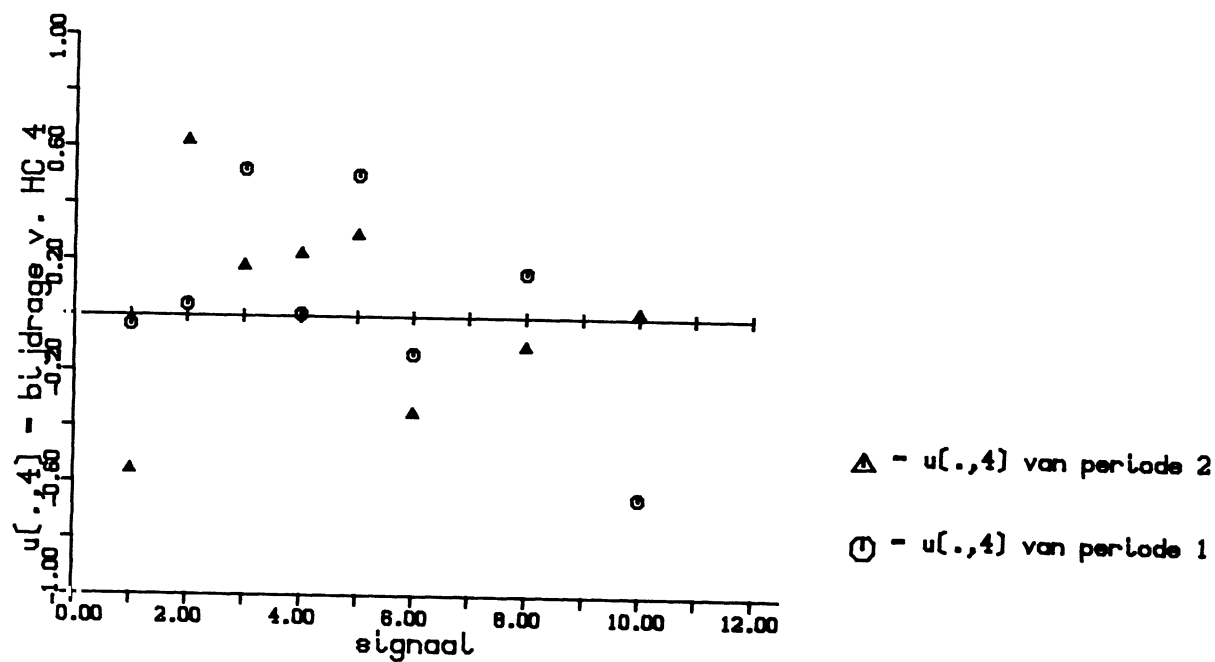
HOOFDCOMPONENT 2

stabiliteit van hoofdcomponenten-ontbinding



HOOFDCOMPONENT 3

stabiliteit van hoofdcomponenten-ontbinding



HOOFDCOMPONENT 4