

Gemeentetypen en lineariteit

Diskussiestuk SISWO-werkgroep  
Longitudinaal School- en beroepsloopbaan  
onderzoek.

Jan de Leeuw  
Datatheorie FSW/RUL  
Breestraat 70  
2311 CS Leiden

## 0: Inleiding

Deze notitie bevat een heranalyse van gegevens verzameld, en eerder geanalyseerd, door het KASKI in het kader van het onderzoek "Regionale factoren en individuele schoolloopbanen" (Bun-Siersma en Spruit, 1982). Voor 792 gemeenten zijn acht gemeente-kenmerken verzameld. Dit zijn, respectievelijk,

- 1: SOC: sociale achterstand,
- 2: LBO: percentage leerlingen LBO,
- 3: WRK: percentage werklozen,
- 4: BIB: aanwezigheid bibliotheek,
- 5: TYP: aantal schooltypen,
- 6: ONG: percentage ongeschoolde werklozen,
- 7: JGD: percentage jeugdwerklozen,
- 8: ABO: aantal krantenabbonementen per gezin.

Informatie over de definitie en meting van deze acht variabelen vindt men in het KASKI-rapport, daarin wordt ook vermeld hoe de 792 gemeenten geselecteerd zijn.

Het KASKI heeft op deze acht variabelen een principale componenten analyse met VARIMAX-rotatie uitgevoerd. In tabel 1 vindt men de produkt moment korrelaties tussen de acht variabelen (uit Bun-Siersma en Spruit, 1982, p. 28). De eerste drie eigenwaarden van deze matrix zijn 2.97, 2.17, en 0.91. Mede op grond van VARIMAX-rotatie hebben de KASKI-onderzoekers besloten drie componenten te gebruiken om de variabelen te beschrijven. De komponentladingen, na rotatie, staan in tabel 2 (uit Bun-Siersma en Spruit, 1982, p. 29).

In deze notitie willen we uitzoeken of en in hoeverre niet-lineaire principale componenten analyse, zoals ontwikkeld door Gifi (1981) en anderen, tot afwijkende of additionele resultaten leidt. We gebruiken hiervoor gediskretiseerde versies van de acht variabelen, in feite is dit niet nodig maar de analyses voor gediskretiseerde variabelen zijn eenvoudiger, zowel om uit te leggen als om uit te voeren.

## 1: Principale componenten analyse

We veronderstellen dat de lezers bekend zijn met 'gewone' principale componenten analyse, zoals bijvoorbeeld gebruikt wordt in het KASKI rapport. Stel nu, dat we niet precies weten welke transformatie van de variabelen het best gebruikt kan worden in gewone PKA (principale componenten analyse dus). We gebruiken bijvoorbeeld een categorische variabele als beroepsnivo, waarvan we alleen ruwweg de ordening van de categorieën kennen, of een variabele als intelligentiequotient waarvan misschien even goed de wortel of het kwadraat gebruikt had kunnen worden. Kiezen van verschillende *kwantifikaties* van categorische variabelen of van verschillende *transformaties* van numerieke variabelen leidt vanzelfsprekend tot verschillende korrelatiematriksen en daardoor tot verschillende uitkomsten van de PKA's die men op die korrelatiematriksen kan toepassen.

Dit feit nu kunnen we gebruiken als definitie van niet-lineaire principale componenten analyse: we bestuderen de korrelatiematriks en de PKA (dat wil zeggen: de eigenwaarden en eigenvektoren van de korrelatiematriks) als functie van de diverse toegestane transformaties en kwantifikaties. In het algemeen zijn niet alle transformaties en kwantifikaties toegestaan. Het kan bijvoorbeeld heel goed zijn, dat we redenen hebben om ons te beperken tot monotone (stijgende en/of dalende) transformaties, of tot polynomen van een lage graad. Vooral bij kontinu variërende grootheden (zoals lengte, maar ook zoals de variabelen SOC, LBO, WRK, ONG, JGD, ABO in dit rapport) is toestaan van alle mogelijke transformaties bepaald niet gewenst: er zijn teveel parameters en we kunnen daardoor alle mogelijke korrelatiematriksen en alle mogelijke PKA-uitkomsten vinden. Polynomen en ook diskretisering en ook diskretisering zoals wij gebruiken dienen ervoor het aantal vrije parameters te beperken, en daardoor interpreteerbare en stabiele oplossingen mogelijk te maken.

De eerste beperking is dat we niet alle transformaties toestaan, we bekijken alleen een welomschreven klasse. De tweede beperking is dat we niet alle transformaties in die klasse uitproberen, met bijbehorende PKA's, eenvoudigweg omdat dat er in het algemeen

oneindig veel zijn. We bekijken daarom in de klasse van toegestane transformaties een beperkt aantal mogelijkheden. In de eerste plaats dikwijls de oorspronkelijke variabelen, zoals ze gekodeerd zijn, oftewel zoals een gewone PKA ze zou analyseren. En vervolgens één of meer speciale keuzen, die ons om diverse redenen interessant lijken.

Een doelstelling van PKA in veel toepassingen is bijvoorbeeld om na te gaan in hoeverre de variabelen tot één enkele variabele (een zogenaamde schaal) gereduceerd kunnen worden. Men gaat dit gewoonlijk na door te kijken of de korrelatiematriks van rang één is, dat wil zeggen of de eerste eigenwaarde van de korrelatiematriks een groot percentage van de variantie 'verklaart'. Een speciale oplossing, die ons daarom dikwijls interesseert, is die keuze van de transformaties of kwantifikaties waarvoor de korrelatiematriks zo één dimensionaal mogelijk is, oftewel waarvoor de eerste eigenwaarde zo groot mogelijk is. Voor diskrete (of gediskretiseerde) variabelen wordt een dergelijke analyse uitgevoerd door het programma PRIMALS, ontwikkeld door Albert Gifi (1981). Dit is de tweede speciale transformatie die we in dit verhaaltje bekijken.

Het is genoegzaam bekend dat korrelatiecoëfficiënten alleen dan de bivariate samenhang goed weergeven als de beide regressies lineair zijn. Het is tevens duidelijk dat als de regressies voor een bepaalde keus van transformaties of kwantifikaties lineair zijn, ze dat dan voor een andere keus in het algemeen niet zullen zijn. Als derde speciale transformatie bekijken we daarom de transformatie die zo goed mogelijk de regressies lineariseert. Of, wat precieser, die transformatie waarvoor het verschil tussen de korrelatieratios en de gekwadrateerde korrelatiecoëfficiënten zo klein mogelijk is. We berekenen deze optimale transformatie met een speciaal voor dit doel geschreven APL-programma.

Deze drie keuzen putten de mogelijkheden niet uit. In Gifi (1981) worden nog aanzienlijk meer mogelijkheden besproken. Wel merken we op dat keuze van een techniek in deze klasse impliceert dat men veronderstelt de multivariate verbanden

goed te kunnen beschrijven met behulp van produkt-moment korrelaties. Dit betekent dat men ervan uitgaat dat het deel van de regressie dat niet gelineariseerd kan worden hetzij niet interessant is hetzij apart bekeken gaat worden. Hetzelfde geldt voor interacties tussen meer dan twee variabelen. Anders gezegd: we handelen alsof transformaties op ieder van de variabelen apart ervoor kunnen zorgen dat de getransformeerde variabelen bij benadering multinormaal verdeeld zijn.

## 2: Marginalen

Het is uiterst zinvol om voor we de analyses uitvoeren eerst eens naar de univariate en bivariate marginalen te kijken. We hanteren daarbij de volgende diskretisering (uitgevoerd door Judy Knip, afdeling Datatheorie).

Variabele 1: Sociale achterstand. Deze variabele neemt integer-waarden aan tussen -3000 en +2000. We geven intervallen en marginale frekwenties.

	<u>benedengrens</u>	<u>bovengrens</u>	<u>marginiaal</u>
1.	-3000	-1500	14
2.	-1500	-1000	24
3.	-1000	- 500	71
4.	- 500	- 0	160
5.	+ 0	+ 500	261
6.	+ 500	+1000	213
7.	+1000	+2000	49

Variabele 2: Percentage leerlingen LBO. Integerwaarden tussen 0 en 1000.

	<u>benedengrens</u>	<u>bovengrens</u>	<u>marginiaal</u>
1.	0	100	4
2.	100	150	4
3.	150	200	17
4.	200	250	38
5.	250	300	64
6.	300	350	108
7.	350	400	162
8.	400	450	171
9.	450	500	123
10.	500	550	59
11.	550	600	20
12.	600	650	13
13.	650	800	9

Variabele 3: Percentage werklozen. Integerwaarden tussen 0 en 100.

	<u>benedengrens</u>	<u>bovengrens</u>	<u>marginaal</u>
1.	0	10	209
2.	10	15	172
3.	15	20	137
4.	20	25	104
5.	25	30	83
6.	30	35	38
7.	35	40	20
8.	40	45	9
9.	45	60	20

Variabele 4: Aanwezigheid bibliotheek. Binair.

	<u>label</u>	<u>marginaal</u>
1.	Nee	344
2.	Ja	448

Variabele 5: Aantal schooltypen. Neemt vijf waarden aan.

	<u>marginaal</u>
1.	309
2.	131
3.	156
4.	21
5.	175

Variabele 6: Percentage ongeschoolde werklozen. Tussen 0 en 16%.

	<u>percentage</u>	<u>marginaal</u>
1.	0	98
2.	1	227
3.	2	156
4.	3	105
5.	4	69
6.	5	36
7.	6	31
8.	7	20
9.	8	19
10.	9	10
11.	10	8
12.	11	7
13.	12 - 16	6

Variabele 7: Percentage jeugdwerklozen. Tussen 0 en 10%.

	<u>percentage</u>	<u>marginaal</u>
1.	0	51
2.	1	194
3.	2	220
4.	3	145
5.	4	99
6.	5	43
7.	6	22
8.	7-10	18

Variabele 8: Aantal krantenabbonementen per gezin. Integer tussen 0 en 1700 (varieert tussen 0 en 1.700).

	<u>benedengrens</u>	<u>bovengrens</u>	<u>marginaal</u>
1.	0	500	13
2.	500	600	11
3.	600	700	39
4.	700	800	80
5.	800	900	152
6.	900	1000	195
7.	1000	1100	172
8.	1100	1200	89
9.	1200	1300	22
10.	1300	1400	13
11.	1400	1700	6

In de eerste appendix van dit rapport staan bovendien nog alle bivariate marginalen (de kruistabellen). We gaan daar verder niet op in, maar het loont de moeite er lang en uitvoerig naar te kijken om een indruk te krijgen van de vorm van de regressies en de structuur van de bivariate verdelingen in het algemeen.

### 3: Analyses

De korrelatiematriks van de gediskretiseerde variabelen staat in tabel 3, deze verschilt van tabel 1 door de diskretisatie. De eerste drie eigenwaarden van deze tabel zijn 2.96, 2.15, en 0.92. De eerste dimensie 'verklaart' 37% van de variantie, de eerste twee samen 'verklaren' 64%, voor tabel 1 is dit respectievelijk 38% en 63%, geen dramatisch verschil. De ongeroteerde komponentladingen op de eerste twee componenten van tabel 3 staan in tabel 4, we bekijken maar twee componenten omdat de derde eigenwaarde kleiner dan één is en omdat er een duidelijke 'knik' in de eigenwaarde-verdeling zit na de tweede dimensie. We gebruiken geen VARIMAX of een andere rotatie-techniek, omdat we niet ieder van de dimensies afzonderlijk willen interpreteren, maar het vlak van de eerste twee dimensies als geheel willen gaan bekijken. Het is daarom belangrijk zoveel mogelijk variantie in dit vlak te konsentrieren. VARIMAX en dergelijke hebben de neiging variantie over de dimensies te verdelen, dit blijkt ook uit het KASKI-rapport. Na rotatie 'verklaart' de eerste dimensie nog slechts 32%, de eerste twee samen 'verklaren' 56%. De derde dimensie is door VARIMAX van ongeveer 11% ongeroteerd

naar 20% geroteerd gebracht.

In tabel 5 staan de wortels uit de korrelatie-ratio's voor de gediskretiseerde ruwe data. Tabel 5 is dus niet noodzakelijk symmetrisch, de elementen van tabel 5 zijn altijd ten minste even groot als de korresponderende korrelatiekoefficient uit tabel 3 (absoluut genomen). Als elementen van tabel 3 en tabel 5 op plaats (i,j) ongeveer gelijk zijn, dan is de regressie van variabele i op variabele j ongeveer lineair, wat niet betekent dat de regressie van j op i ook ongeveer lineair is. Zo zijn bijvoorbeeld de elementen in de kolom behorend bij BIB in tabel 3 en tabel 5 gelijk, omdat de regressie op een binaire variabele altijd lineair is. Dit geldt echter niet voor de elementen in de rij BIB. We zien uit vergelijking van tabel 3 en tabel 5 dat de regressies in veel gevallen aardig lineair zijn (uit plots gebruik makend van de bivariate marginalen in de appendix zou dit natuurlijk fraaier blijken). Er zijn evenwel een aantal afwijkingen, die mogelijkerwijs verbeterd kunnen worden. De som van de korrelatie-ratio's min de gekwadrateerde korrelaties is voor deze ruwe skores 0.92.

Voor onze tweede analyse passen we PRIMALS toe, een programma dat de grootste eigenwaarde van de korrelatiematriks zo groot mogelijk maakt. De transformaties of kwantifikaties die we vinden staan in figuur 1, de bijbehorende korrelatiematriks in tabel 6, de komponentladingen in tabel 7, de wortels uit de korrelatieratios in tabel 8. De eerste drie eigenwaarden zijn nu 3.18, 1.90, en 0.95, dat wil zeggen dat de eerste dimensie nu 40% 'verklaart', en de eerste twee samen 64%. De bijdrage van de eerste dimensie is omhoog gegaan, logisch natuurlijk omdat PRIMALS deze bijdrage zo groot mogelijk maakt, maar tezamen 'verklaren' de eerste twee dimensies evenveel als voor PRIMALS. De komponent-ladingen geven een zelfde beeld: op de eerste dimensie worden ze gemiddeld een beetje groter, op de tweede dimensie gemiddeld een beetje kleiner. Vergelijken van tabel 6 en tabel 8 laat zien dat PRIMALS ook behoorlijk de regressies gelineariseerd heeft, de globale maat van niet-lineariteit is nu 0.68 (in plaats van 0.92). De verschuivingen in de resultaten van de PKA zijn



dus miniem, de schalingen wijken echter nogal aanzienlijk van lineariteit af.

Wat zien we aan de transformaties in figuur 1 ? In de eerste plaats zijn er een aantal duidelijke onregelmatigheden, zoals categorie van variabele 2 en categorie 4 van variabele 5. In het laatste geval heeft dit ongetwijfeld te maken met de lage marginaal, de positie van de categorie in de transformatie is daardoor uiterst instabiel. Men zou dit kunnen verduidelijken door een stabiliteitsanalyse door middel van bootstrap of delta methode (uiteengezet in Gifi, 1981). Bij variabele 2 (LBO) lijken de eerste drie categorieën niet stabiel, dit blijkt ook al uit de volgende analyse die we een stukje verderop rapporteren. Verder is opmerkelijk dat de transformaties van de drie werkloosheidsvariabelen (3:WRK, 6:ONG, 7:JGD) glad en konkaaf zijn, waarschijnlijk omdat alle drie de variabelen scheef naar links zijn. Dit wijst erop dat percentage variabelen in dit soort analyses dikwijls goed logaritmisch getransformeerd kunnen worden. Er is enige onregelmatigheid in de toppen van de transformaties, maar die kunnen we weer verklaren uit lage marginalen en instabiliteit. Meest opmerkelijk is natuurlijk de U-vorm van de transformatie van variabele 8:ABO, die erop wijst dat de regressie van de eerste principale komponent op ABO U-vormig is. Dat wil zeggen: als die eerste komponent vooral samenhangt met werkloosheid, zie verderop, dan blijkt uit deze analyse dat er weinig werkloosheid is in gemeenten waarin veel kranten gelezen worden, en ook relatief weinig werkloosheid in gemeenten waarin heel weinig kranten gelezen worden. Het eerste lijkt vooral te slaan op voorsteden en forensengemeenten, het tweede op typische plattelandsgemeenten. Zie verderop. Er komt in ieder geval interessante ekstra informatie uit een niet-lineaire PKA, de korrelatiekoefficienten en eigen-oplossingen zijn echter opmerkelijk stabiel (vergelijk het vorige diskussiestuk De Leeuw, Van der Burg, 1981).

De derde analyse heeft geen naam, we noemen het programma daarom maar even LINEALS, zoals uiteengezet probeert het alle bivariante regressies te lineariseren, en de globale maat van niet-lineariteit te minimaliseren. De transformaties uit LINEALS staan nu in figuur

2, de korrelaties, komponent-ladingen, korrelatie-ratios, in tabellen 9,10,11. De eigenwaarden, grootste drie, zijn nu 3.14, 2.06, en 0.89, dat wil zeggen 39% voor de eerste dimensie en 65% voor de eerste twee samen. Alweer geen grote verschillen in de korrelaties, evenals bij PRIMALS wordt ABO zo geschaald dat de korrelaties met WRK, ONG, JGD toenemen ten koste van die met BIB en TYP als we vergelijken met de ruwe data. De veranderingen zijn echter niet aanzienlijk. Transformaties van LINEALS in figuur 2 gedragen zich 'netter' dan die voor PRIMALS, het lijkt alsof LINEALS minder gevoelig is voor lage marginalen. Met name 5:TYP wordt keurig monotoon, 2:LBO ziet er ook een stuk netter uit. De konkaafheid van de percentage variabelen is nog steeds overduidelijk (ook LBO is nu konkaaf, en 1:SOC en 5:TYP gaan ook die kant op. De U van 8:ABO verandert nauwelijks. De globale maat van niet-lineariteit, die door LINEALS zo klein mogelijk gemaakt wordt, is nu 0.54. Dus: voor de ruwe gediskretiseerde data 0.92, voor de PRIMALS oplossing 0.68, en voor de LINEALS oplossing 0.54. Deze maat is aanzienlijk gevoeliger voor keuze van kwantifikatie dan bijvoorbeeld de grootste eigenwaarde, dit kan verklaren waarom LINEALS een over het algemeen wat bevredigender PKA oplossing geeft. De afwijkingen van lineariteit die overblijven zitten vrijwel geheel in de rij overeenkomend met 5:TYP, dat wil zeggen de regressies van 5:TYP op de overige variabelen, met name op 1:SOC, 2:LBO, 8:ABO zijn nog steeds niet lineair. Het blijft zo dat 5:TYP het slechts bij de anderen past, in termen van regressie althans. Ook 4:BIB is niet best.

In figuur 3 staan de komponent-ladingen uit alle drie de analyses op de eerste twee dimensies geploteerd. Dat er drie klusters van variabelen zijn is duidelijk. De eerste kluster bestaat uit 3:WRK, 6:ONG, 7:JGD, drie werkloosheidsvariabelen die onderling hoog correleren. De tweede kluster zijn 1:SOC, 2:LBO, en 8:ABO. Het KASKI noemt dat sociaal-educatieve achterstand (Bun-Siersma & Spruit, 1982, p. 29). De derde kluster bestaat uit 4:BIB en 5:TYP, dat noemt het KASKI culturele gelegenheid. De conclusies uit de PKA zijn hetzelfde als die van het KASKI, alleen vinden wij twee componenten genoeg voor een volledige beschrijving. De

drie verschillende PKA's lijken zeer veel op elkaar, de enige systematische verschuiving lijkt dat 1:SOC, 2:LBO en 8:ABO voor de ruwe data een grotere hoek maken met (lager korreleren met) 3:WRK, 6:ONG, 7:JGD dan voor de PRIMALS en LINEALS oplossingen. De verschuiving is echter niet aanzienlijk. De derde faktor van het KASKI, kulturele gelegenheid, onze kluster 4:BIB en 5:TYP, zou best eens kunnen samenhangen met gemeentegrootte. Het lijkt ons dat deze twee variabelen de enigen zijn die daar expliciet mee samen hangen.

Onze voorlopige konklusie is, dat de KASKI analyse niet misleidend is, maar de interessantste bivariate effecten wel min of meer verdoezeld. De keuze van de variabelen lijkt niet erg gelukkig, de werkloosheidsvariabelen zijn oververtegenwoordigd, twee variabelen hangen met gemeentegrootte samen, verschillende regressies zijn niet lineair en moeilijk lineair te maken. Ook in dit voorbeeld bevestigt niet-lineaire PKA de resultaten van lineaire PKA, met dien verstande dat er ook aanvullende informatie berekend wordt.

#### 4: Schalings

De PKA's in de vorige paragraaf transformeren/kwantificeren de variabelen, berekenen korrelaties, en eigenwaarden/eigenvektoren. Dit gaat allemaal op het nivo van variabelen, de individuele gemeenten spelen geen rol, komponentscores worden niet berekend. In deze laatste paragraaf bekijken we de komponentscores uit twee andere technieken: PRINCALS en HOMALS. PRINCALS uitleggen is relatief eenvoudig met de kennis die we nu hebben: PRINCALS kiest zijn transformaties op zo'n manier dat de som van de  $p$  grootste eigenwaarden van de korrelatiematriks zo groot mogelijk wordt. Hierbij kan  $p$  vrij gekozen worden, wij gebruiken  $p = 2$ . Voor  $p = 1$  is PRINCALS gelijk aan PRIMALS, dat wil zeggen PRIMALS is een speciaal geval. In PRINCALS is het ook mogelijk te eisen dat de transformaties monotoon zijn, we hebben dat hier inderdaad gedaan. Dit heeft alleen konsekwenties voor de transformatie van 8:ABO, die plat wordt aan de beide uiteinden. De tweede

PKA-dimensie wordt wat belangrijker om dit effect op te kunnen vangen, maar het effect op korrelaties en eigen-oplossing is verder gering. De komponenteskores staan in figuur 4, de 792 punten zijn gelabeld met het gemeentenummer zoals dat gehanteerd wordt door het CBS (en in navolging daarvan door het KASKI). Het beste kan men een representatie als in figuur 4 opvatten als een meetkundige representatie, waarin gemeenten met een overeenkomstig antwoordpatroon op de acht variabelen dicht bij elkaar liggen. In PRINCALS is het daarbij zo dat iedere variabele een aantal evenwijdige lijnen in de plot definieert, het aantal lijnen is het aantal categorieën min één, de lijnen staan loodrecht op de richting gedefinieerd door de komponentladingen van die variabele. Alle gemeenten in de eerste categorie liggen tussen oneindig en de eerste evenwijdige lijn, alle gemeenten in de tweede categorie liggen tussen de eerste en de tweede evenwijdige lijn, enzovoorts. Dat kan nooit precies, PRINCALS plaatst de gemeenten op zo'n manier dat het zo goed mogelijk kan. Categorieën van de variabele komen dus overeen met evenwijdige strips in de plot, voor verschillende variabelen is de stand van de strips verschillend (voor meer details zie Gifi, 1981).

Zonder figuur 4 in detail te analyseren valt het volgende op. Links boven liggen gemeenten zoals Bloemendaal (377), Heemstede (397), Warmond (628), Voorschoten (626), Haren (017), Voorburg (624), Rijswijk (603), Amstelveen (362), De Bilt (310), Leiderdorp (547), Oegstgeest (579). De betere standen dus. Onderin vinden we Mijdrecht (336), Est en Opijnen (234), Everdingen (508), Zoelen (300), Renswoude (339), Ameide (485), Groot-Ammers (520), Benschop (309). Deze liggen aan de linkerkant, het zijn plattelandsgemeenten in Utrecht en Zuid-Holland. Meer naar rechts onderin liggen plattelandsgemeenten in het oosten, zuiden, en noorden. Uiterst rechts vinden we Beerta (006), Finsterwolde (013), Midwolda (028), Emmen (114), Hoensbroek (921), Kerkrade (928), Vaals (981), Brunssum (899). Dit zijn redelijk grote kernen, met een zeer hoge werkloosheid. Bovenaan in het midden Vught (865), Enschede (153), Arnhem (202), Alkmaar (361), Den Haag (518), Hengelo (164), dus grote

gemeenten, met naar rechts toenemende werkloosheid. De grote steden Amsterdam (363), Utrecht (344), Den Haag (518), Rotterdam (599), Groningen (014), Eindhoven (772), Leeuwarden (080), Haarlem (392) liggen allemaal bovenaan, Haarlem het meest links, Leeuwarden, Groningen, Rotterdam, Eindhoven, en Maastricht (935) meer naar rechts. Uiterst links, maar op de tweede dimensie rond het midden, vinden we Abcoude (305), Rhoon (596), Haarlemmerliede (393), Ouder-Amstel (437), Moerkapelle (560), Nootdorp (577), Diemen (384), Blaricum (376), Loosdrecht (330), Voorhout (625), ook allemaal forensengemeenten, vooral in het westen, maar niet zo welvarend als links boven.

In figuur 5 hebben we een soortgelijke oplossing, maar nu uit HOMALS. HOMALS probeert meer direkt gemeenten met een overeenkomstig antwoordpatroon dicht bij elkaar te leggen. Als er een dominante eerste dimensie is (hier: werkloosheid) dan maakt HOMALS vaak gebruik van het feit dat een niet-lineaire transformatie van die eerste dimensie vaak nog beter is dan een onafhankelijke tweede dimensie. We krijgen dan twee ongekorreleerde maar duidelijk funktioneel afhankelijke dimensies, vaak in de vorm van een 'hoefijzer' of 'dikke V'. Links op het hoefijzer zien we weer Bloemendaal en Heemstede, met Wassenaar (629), enzovoorts. Rechts op het hoefijzer staan Heerlen (917), Roermond (957), Venlo (983), Sittard (968), Schaesberg (960), 's Hertogenbosch (796), Maastricht (935), Nijmegen (268), Hoensbroek (921), Kerkrade (928), het diepe zuiden dus. Onderaan, in de punt van de V, vinden we Boekel (755), Heukelum (533), Leerdam (545), Lith (808), Sprang-Kapelle (849), Herwijnen (250), die weinig gemeen hebben, maar 'gemiddeld' zijn. Bovenaan in de punt van de V vinden we Beverwijk (375), Amsterdam (363), Alkmaar (361), Zaanstand (479). Grotere gemeenten komen bovenaan in de V terecht.

In figuur 6 en 7 hebben we voor de PRINCALS en HOMALS oplossing steeds de zwaartepunten (gemiddelden) uitgerekend van de gemeenten in de categorieën van ieder van de variabelen. Het punt gelabeld 1 in de eerste plot van figuur 6 is dus de gemiddelde positie van alle gemeente met een SOC-waarde van

één, waarbij de positie dus de positie is in figuur 4. We zien 3:WRK, 6:ONG, 7:JGD naar rechts toenemen, 4:BIB en 5:TYP nemen naar boven toe. De kluster 1:SOC, 2:LBO, 8:ABO rangschikt van links boven naar rechts onder, waarbij zowel 2:LBO als 8:ABO aan de uiteinden vreemd doen. In figuur 7 blijkt dat de werkloosheidsdimensie keurig langs het hoefijzer loopt. We zien dat de categorieën van 1:SOC, 2:LBO, 8:ABO in de linkertak van het hoefijzer liggen (negatieve kant van de X-as), en dat ABO de neiging heeft een nieuw hoefijzertje binnen de linkertak te vormen. Gemeentes aan het uiteinde van dit hoefijzer (waar uiterst weinig gelezen wordt, maar niet zoveel werkloosheid is) zijn bijvoorbeeld Beusichem (210), Ilpendam (407), Poortugaal (592), Angerlo (199), Polsbroek (338), Obdam (429).

Uit deze analyse is duidelijk dat gemeentegrootte een rol speelt via BIB en TYP, dat werkloosheid domineert, en dat sociaal-kulturele gelegenheid weliswaar gekorreleerd is met werkloosheid, maar toch nog voldoende onafhankelijk varieert. Ook in HOMALS en PRINCALS komen de niet-lineaire regressies duidelijk terug, het is zelfs zo dat bijvoorbeeld HOMALS prima diagnostisch gebruikt kan worden om te laten zien wat voor afwijkende dingen er in de gegevens verborgen zitten.

SOC	1.00								
LBO	0.60	1.00							
WRK	0.28	0.09	1.00						
BIB	-0.14	-0.17	0.26	1.00					
TYP	-0.28	-0.16	0.23	0.54	1.00				
ONG	0.26	0.14	0.82	0.26	0.29	1.00			
JGD	0.30	0.15	0.83	0.20	0.15	0.65	1.00		
ABO	0.50	0.28	0.18	-0.09	-0.17	-0.13	0.18	1.00	

TABEL 1: Korrelaties uit KASKI-rapport

SOC	0.24	0.84	-0.19
LBO	0.02	0.81	-0.06
WRK	0.95	0.10	0.14
BIB	0.15	-0.07	0.86
TYP	0.15	-0.19	0.85
ONG	0.86	0.13	0.21
JGD	0.89	0.14	0.06
ABO	0.11	0.70	-0.06

TABEL 2: VARIMAX-geroteerde componenten uit KASKI-rapport

SOC	1.00							
LBO	0.59	1.00						
WRK	0.26	0.10	1.00					
BIB	-0.15	-0.15	0.24	1.00				
TYP	-0.28	-0.15	0.22	0.54	1.00			
ONG	0.25	0.15	0.82	0.26	0.29	1.00		
JGD	0.30	0.17	0.82	0.20	0.15	0.65	1.00	
ABO	0.47	0.29	0.19	-0.10	-0.17	0.14	0.18	1.00

TABEL 3: Korrelaties ruwe data gediskretiseerd

SOC	0.50	-0.69
LBO	0.35	-0.62
WRK	0.90	0.17
BIB	0.30	0.64
TYP	0.25	0.72
ONG	0.86	0.20
JGD	0.86	0.08
ABO	0.37	-0.54

TABEL 4: Komponent-ladingen uit tabel 3

SOC	1.00	0.62	0.34	0.15	0.32	0.33	0.37	0.51
LBO	0.60	1.00	0.14	0.16	0.17	0.19	0.21	0.33
WRK	0.29	0.16	1.00	0.24	0.23	0.83	0.82	0.24
BIB	0.16	0.20	0.28	1.00	0.55	0.28	0.25	0.14
TYP	0.32	0.22	0.25	0.54	1.00	0.34	0.18	0.21
ONG	0.30	0.21	0.83	0.26	0.29	1.00	0.66	0.22
JGD	0.32	0.21	0.82	0.20	0.16	0.67	1.00	0.30
ABO	0.48	0.30	0.24	0.10	0.18	0.19	0.22	1.00

TABEL 5: Korrelatie-ratios ruwe data gediskretiseerd



SOC	1.00							
LBO	0.62	1.00						
WRK	0.32	0.17	1.00					
BIB	-0.14	-0.07	0.26	1.00				
TYP	-0.15	-0.03	0.20	0.53	1.00			
ONG	0.33	0.23	0.78	0.26	0.29	1.00		
JGD	0.36	0.24	0.81	0.22	0.17	0.68	1.00	
ABO	0.44	0.32	0.26	-0.07	-0.04	0.24	0.32	1.00

TABEL 6: Korrelaties na PRIMALS

SOC	0.60	-0.61
LBO	0.48	-0.53
WRK	0.87	0.20
BIB	0.29	0.71
TYP	0.27	0.69
ONG	0.84	0.21
JGD	0.86	0.11
ABO	0.50	-0.42

TABEL 7: Komponent-ladingen na PRIMALS

SOC	1.00	0.65	0.34	0.14	0.25	0.35	0.37	0.47
LBO	0.65	1.00	0.18	0.07	0.11	0.25	0.27	0.34
WRK	0.33	0.17	1.00	0.26	0.20	0.78	0.82	0.27
BIB	0.16	0.20	0.28	1.00	0.55	0.28	0.25	0.14
TYP	0.24	0.21	0.23	0.52	1.00	0.33	0.18	0.19
ONG	0.33	0.24	0.79	0.26	0.30	1.00	0.68	0.25
JGD	0.37	0.25	0.82	0.22	0.17	0.68	1.00	0.32
ABO	0.48	0.33	0.29	0.07	0.10	0.25	0.32	1.00

TABEL 8: Korrelatie-ratios na PRIMALS

SOC	1.00							
LBO	0.65	1.00						
WRK	0.32	0.15	1.00					
BIB	-0.15	-0.12	0.26	1.00				
TYP	-0.22	-0.12	0.21	0.55	1.00			
ONG	0.32	0.21	0.76	0.26	0.29	1.00		
JGD	0.36	0.23	0.81	0.23	0.17	0.68	1.00	
ABO	0.49	0.34	0.26	-0.09	-0.10	0.23	0.30	1.00

Tabel 9: Korrelaties na LINEALS

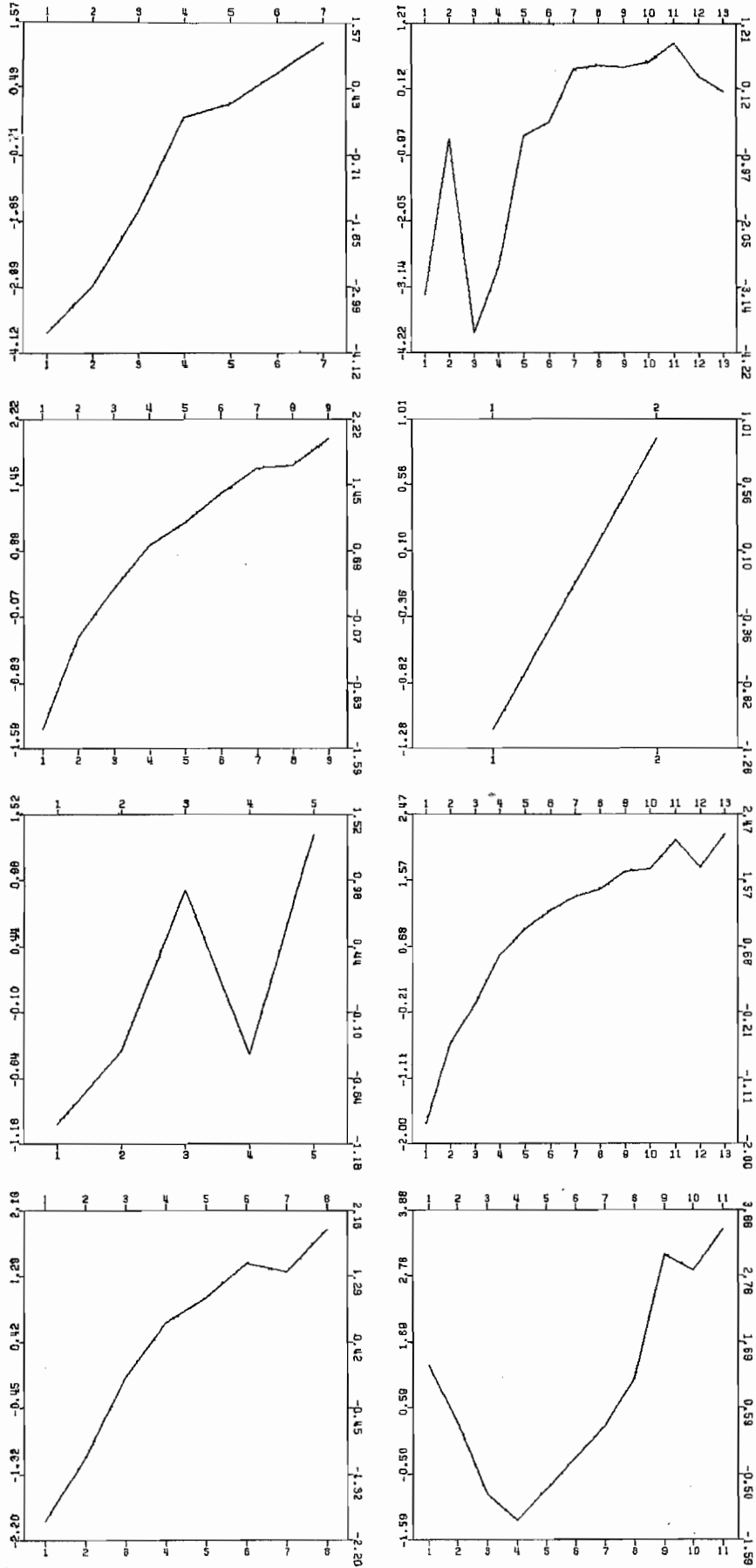
SOC	0.62	-0.61
LBO	0.48	-0.57
WRK	0.86	0.23
BIB	0.26	0.71
TYP	0.22	0.73
ONG	0.83	0.24
JGD	0.86	0.15
ABO	0.51	-0.44

Tabel 10: Komponent-ladingen na LINEALS

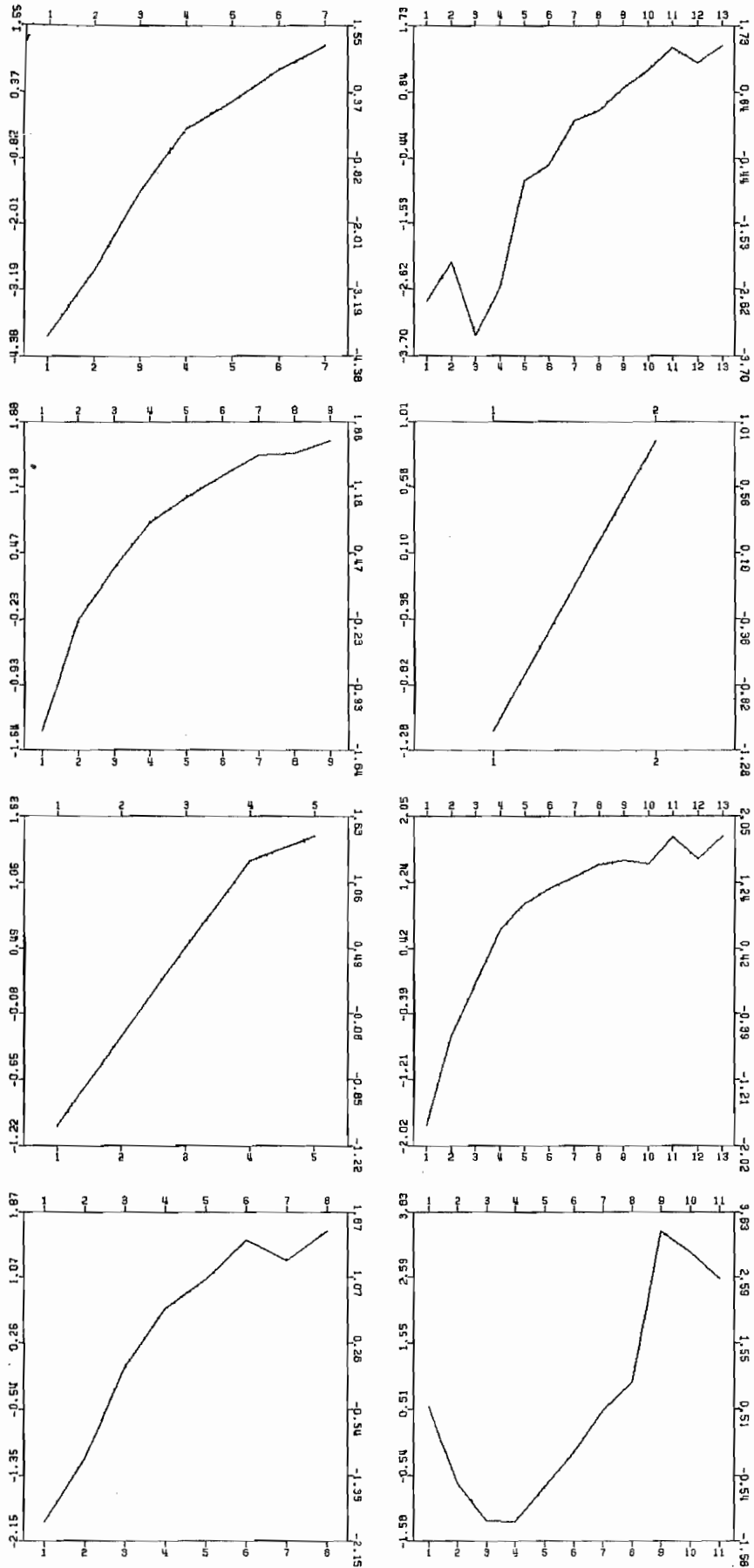
SOC	1.00	0.66	0.34	0.15	0.28	0.35	0.37	0.49
LBO	0.66	1.00	0.17	0.12	0.15	0.23	0.26	0.35
WRK	0.33	0.17	1.00	0.26	0.21	0.77	0.82	0.27
BIB	0.16	0.20	0.28	1.00	0.55	0.28	0.25	0.14
TYP	0.30	0.22	0.24	0.55	1.00	0.34	0.19	0.20
ONG	0.33	0.25	0.78	0.26	0.30	1.00	0.68	0.25
JGD	0.37	0.25	0.81	0.23	0.17	0.68	1.00	0.32
ABO	0.50	0.35	0.28	0.09	0.12	0.24	0.30	1.00

Tabel 11: Korrelatie-ratios na LINEALS

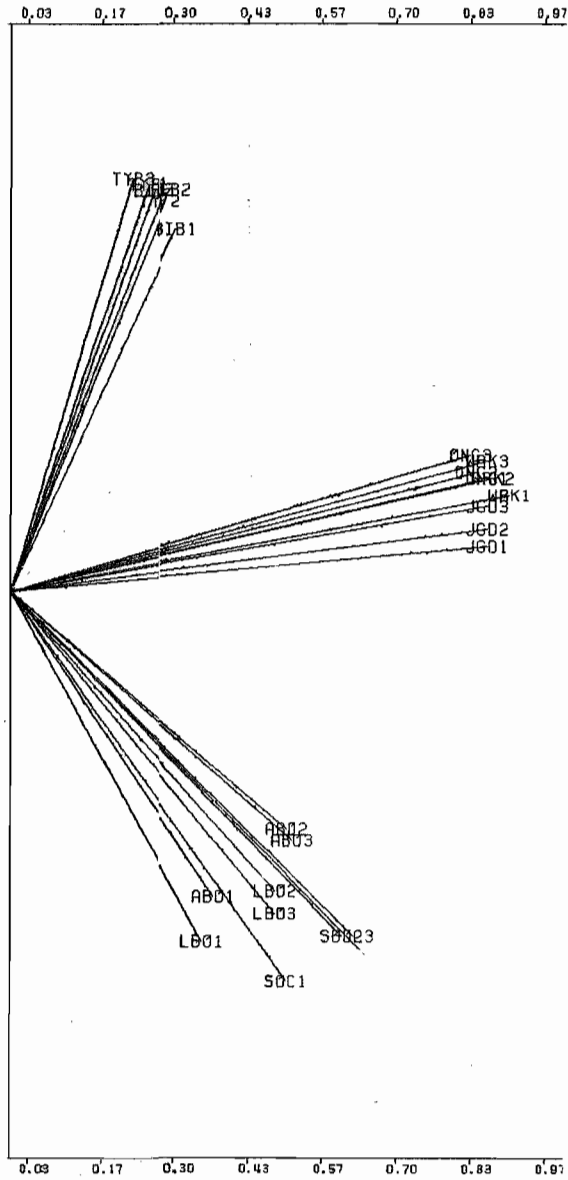
F1



Figuur 1: Transformaties/kwantifikaties uit PRIMALS

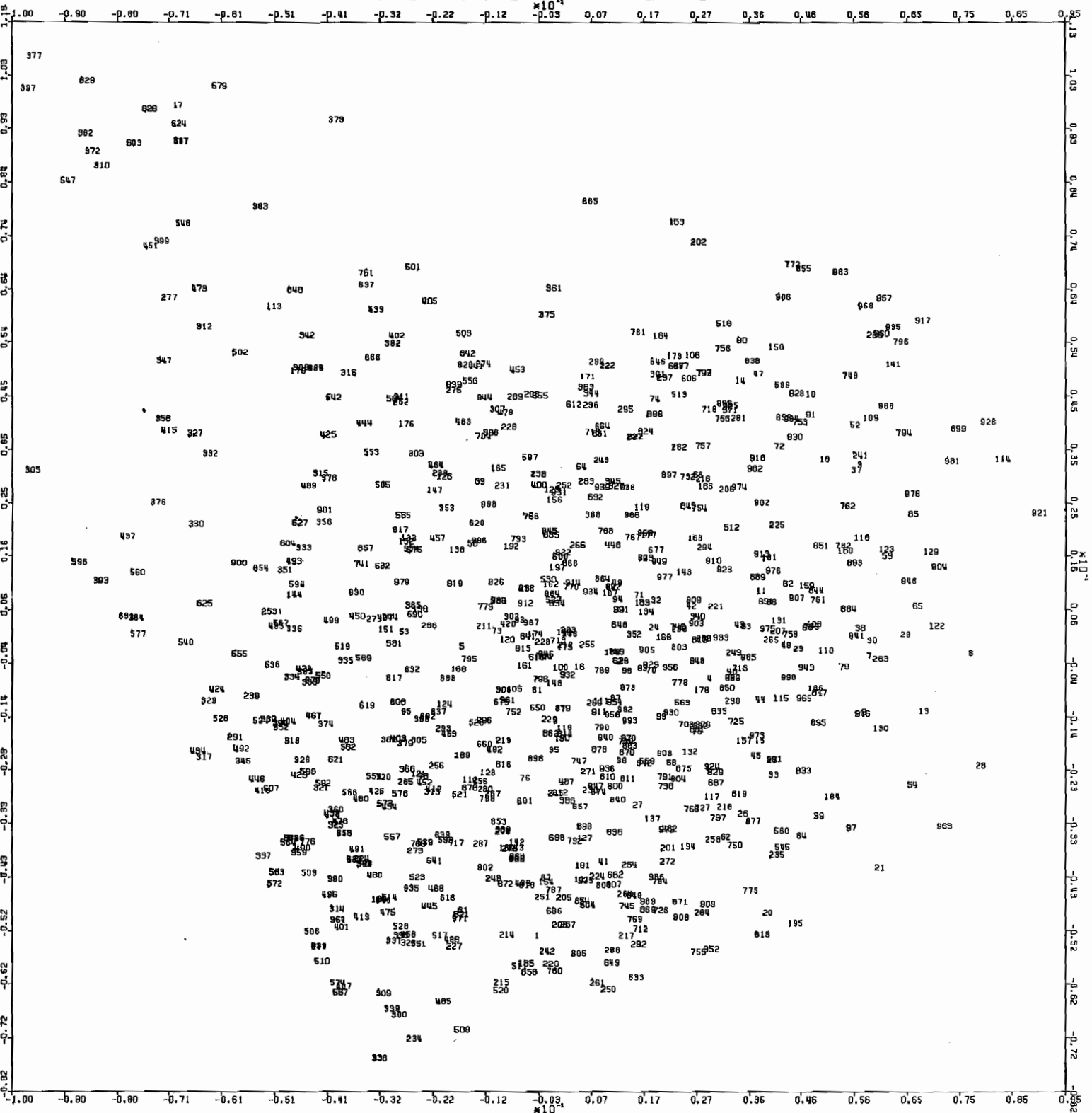


Figur 2: Transformaties/kwantifikaties uit LINEALS



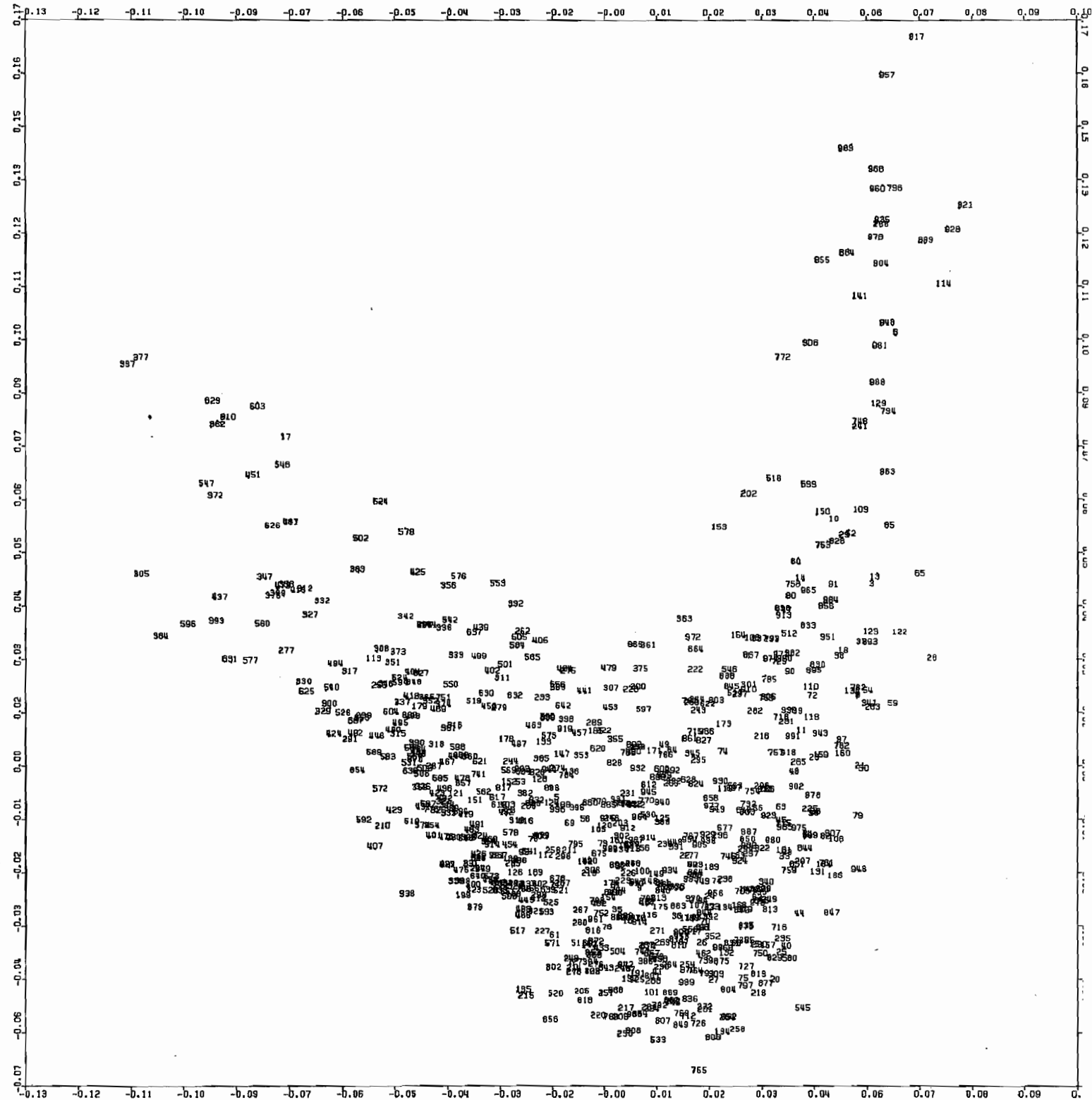
Figuur 3: Komponent-ladingen  
 1: ruwe data, gediskretiseerd  
 2: na PRIMALS  
 3: na LINEALS

# PRINCALS KASKI

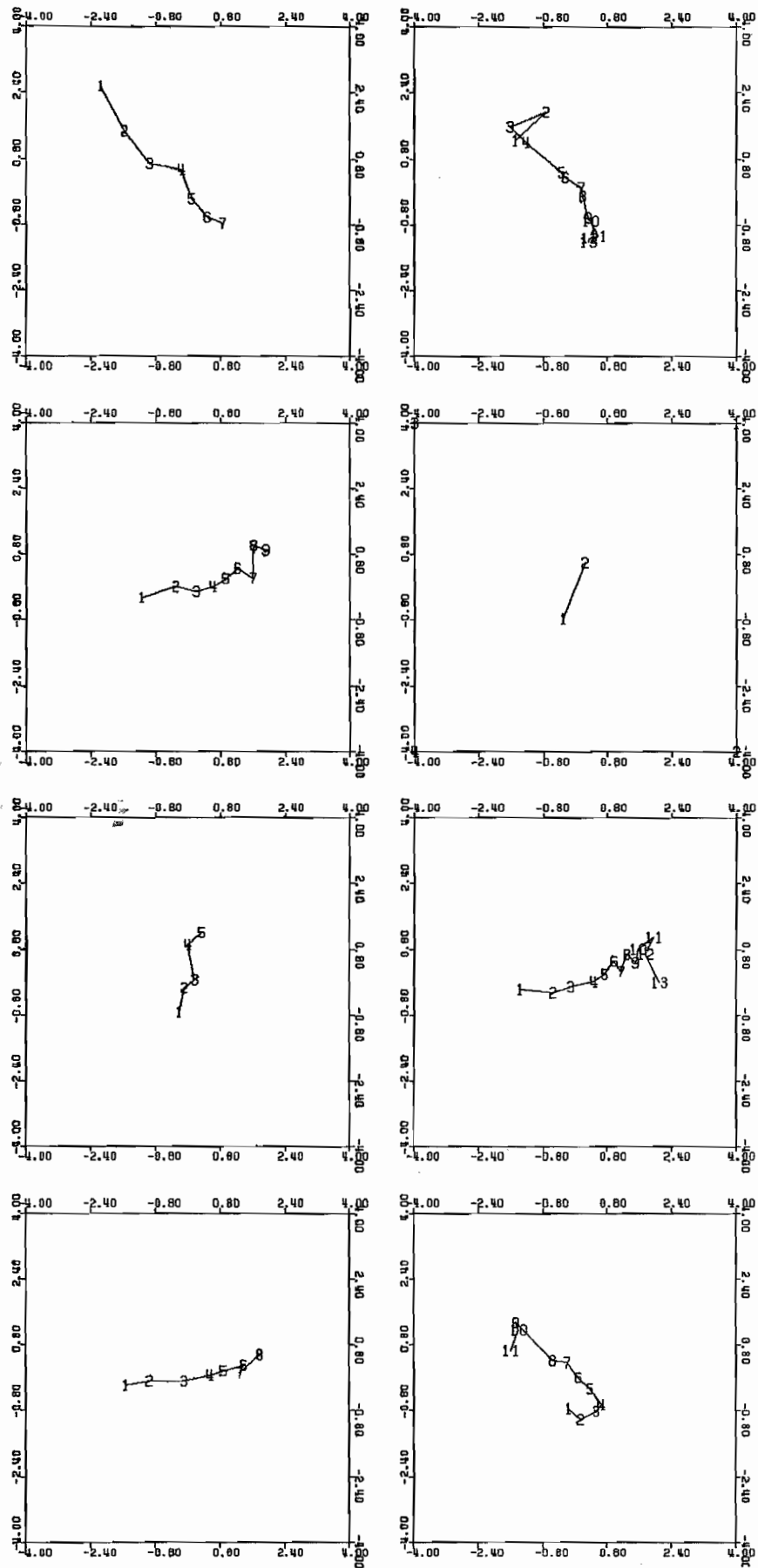


Figuur 4: PRINCALS komponent scores voor gemeenten

# HOMALS KASKI

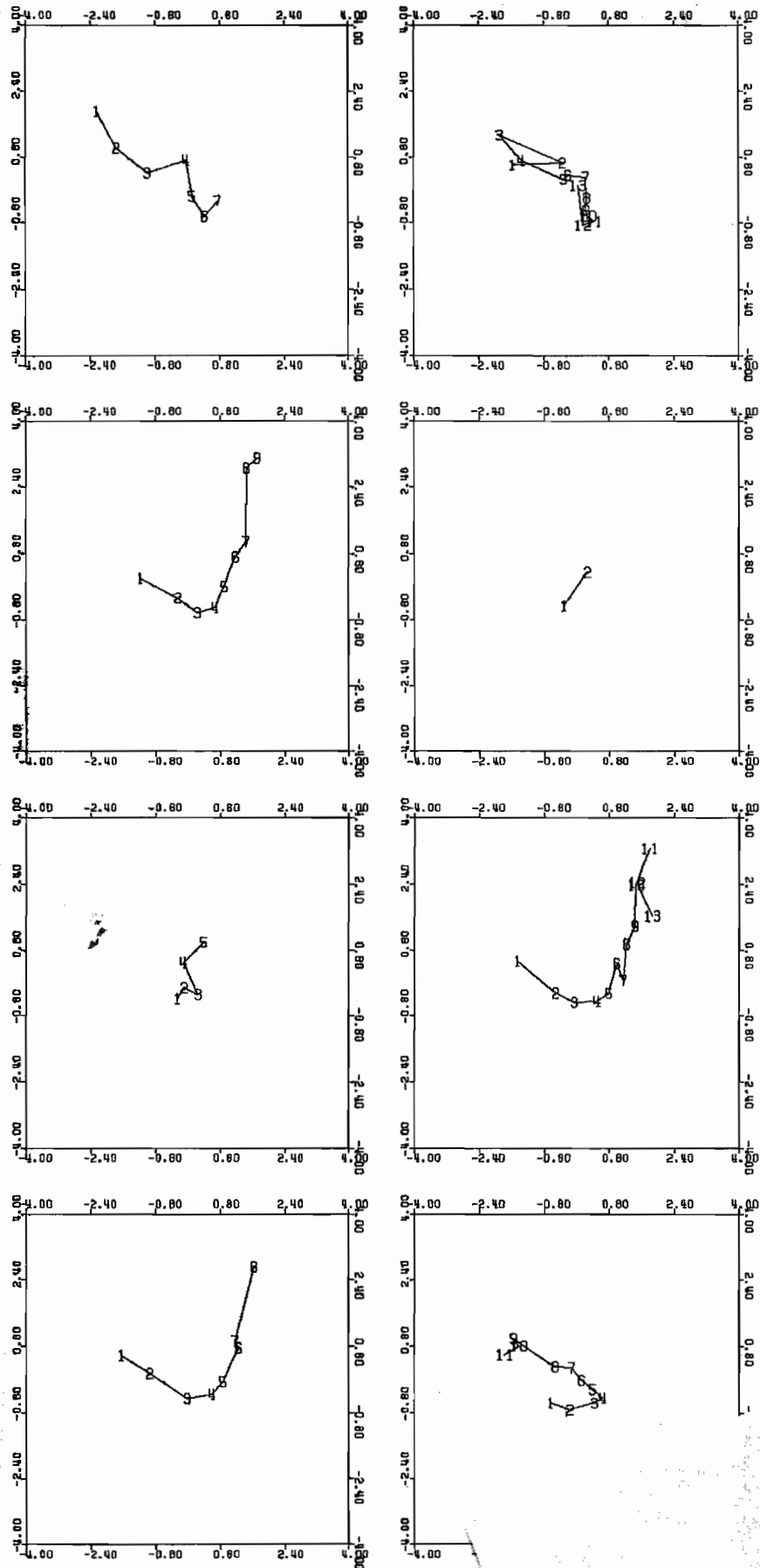


Figuur 5: HOMALS component scores voor gemeenten



Figuur 6: PRINCALS categorie kwantifikaties





Figuur 7: HOMALS categorie kwan

RR-05-