RR – 85-21

RECOVERY AND STABILITY IN NON LINEAR PRINCIPAL COMPONENTS ANALYSIS

J.L.A. van Rijckevorsel
DEPARTMENT AOA & DATATHEORY
FACULTY OF SOCIAL SCIENCES
ERASMUS UNIVERSITY
ROTTERDAM

B. Bettonvil
J. de Leeuw
DEPARTMENT OF DATATHEORY
FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF LEIDEN

Paper presented at the European meeting of the Psychometric Society.
Groningen, 1980, The Netherlands.

Recovery and Stability in Non-Linear Principal Components Analysis

by Jan van Rijckevorsel
Bert Bettonvil
Jan de Leeuw

Summary

Non-linear optimal one-dimensional transformations of stochastic variables are given by the eigenvector belonging to the dominant eigenvalue of the correlation matrix of the transformed variables. The corresponding loss-functions and interpretations are presented together with other work in this context. Two special cases are derived where we respectively know the distribution of the stochastic variables beforehand and thus the type of the optimal transformations and the case where we can approximate the optimal continuous non-linear transformations with discrete stepfunctions. A random study with bootstraps (Efron,1979) using three sample sizes and five types of distortions of discretized continuous multinormal random variables is presented to illustrate the discrete approximation of the continuous transformations. Eigenvalues and optimal transformations are nicely recovered by the discrete technique. The underlying distribution is recovered for all sample sizes and all distortions. The algorithm and computerprogram of the discrete approximation technique are presented and previous interpretations and alternatives are discussed.

keywords: NON-LINEAR WEIGHTS; PRINCIPAL COMPONENTS ANALYSIS;
DISCRETIZATION; BOOTSTRAP

) This research was supported in part by grant nr. 56-97 of the Netherlands organization for the advancement of pure research (Z.W.O.)

1. INTRODUCTION AND NOTATION

1.1. INTRODUCTION

Computing the first principal component of the correlation matrix to define linear weights for the variables has interesting optimal properties. This applies both to the population matrix and the sample matrix. The explicit use of the first principal component as linear weights to construct a one - dimensional scale originates with Horst (1936) and Edgerton & Kolbe (1936), although it is implicit in Pearson (1901) and Hotelling (1933).

In this paper we define non-linear weights as optimal if the dominant eigenvalue of the correlation matrix of the weigthed variables is as large as possible. This generalizes optimal linear weighting (Gifi, 1980). The approximation of continuous non-linear transformations with discrete stepfunctions from a finite basis is a special case of applying non-linear weights. This technique is called non-linear principal components analysis of categorical data a.o. by Guttman (1941), Johnson (1950), Lord (1958), Bock (1960), De Leeuw (1973), Nishisato (1978), and Gifi (1980). The corresponding computerprogram used in this study is called HOMALS (Van Rijckevorsel and De Leeuw, 1980).

Another special case of applying non-linear optimal weights uses orthogonal polynomials as an infinite basis. These functions are the Hermite-Chebyshev polynomials in the case of the multinormal distribution. All bivariate distributions are diagonalized here, we know the form of the optimal transformations and linear- and non-linear weighting amounts to the same thing. Another relevant fact for this study is that the results of non-linear weighting are invariant under one-to-one non-linear transformations of the variables. The recovery of the original distribution for three different sample sizes and five different non-linear transformations is presented here; the results are compared with theoretical population values for all transformations. To get an idea of the variation and stability of the parameters we used a simulation method, because analytical methods are too expensive.

The simulation method we used is called the *bootstrap* method (Efron, 1979-1). It is a convenient and simplified version of the Quenouille-Tukey jackknife.
The mapping of a continuous interval into a point by means of a non-linear transformation is called *discretization*. We can think of several discretizations that either match types of frequency distributions that occur with real data or that follow the normal curve " as good as possible ". We have chosen for three discretizations of the first kind: skew, equal and U-shape, and two of the latter: optimal and pseudo-optimal.
What we want to know is: what happens if one uses this non-linear weighting technique (i.e. HOMALS) on some discretizations of a continuous distribution ? We want answers to questions like: does HOMALS transform to multinormality and does HOMALS symmetrize skew distributions etc. We employ a kind of sensitivity analysis that shows how the non-linear weighting procedure behaves in an ideal situation and to what extent the results change if we change the data. The ideal situation here is the multinormal distribution, because this simplifies the theory of non-linear weighting. It makes it possible to approximate the population parameters, which on their turn are approximated by our HOMALS solutions. We are well aware of the fact that in many situations the assumption of multinormality is both not very realistic and difficult to test.

1.2 NOTATION

We have a finite set N of *individuals* and a finite set of *variables*. There are n numbered individuals $N = \{\upsilon_1, \upsilon_2, \ldots, \upsilon_n\}$ and m variables with index j. A variable is a function η_j that maps the set N into the set K_j. The k_j elements of K_j are the *categories* of a variable: $K_j = \{\kappa_1^j, \kappa_2^j, \ldots, \kappa_{k_j}^j\}$.

The *datamatrix* H is an n x m matrix with $h_{ij} = \eta_j(\upsilon_i)$. Every h_{ij} is an element of K_j, but it is not necessarily a number.
If $K_j = \{1, 2, \ldots, k_j\}$ then the matrix H will contain the *category numbers*.

The variable $\eta_j: N \to K_j$ defines an $n \times k_j$ indicator matrix. This is a (0,1)-matrix G_j with

$$g^j_{ir} = 1 \text{ if } \eta_j(\upsilon_i) = k^j_r,$$

and

$$g^j_{ir} = 0 \text{ if } \eta_j(\upsilon_i) \neq k^j_r.$$

The index of the category numbers is $r = 1,\ldots,k_j$. Every row of G_j contains exactly one element equal to one so all rows of G_j add up to one. If we introduce an additional n-element vector u equal to one, then we can say that the column totals of G_j are the elements of $d_j \triangleq G_j'u$. If D_j is diagonal with $D_ju = d_j$ then $G_j'G_j = D_j$; the columns of G_j are thus orthogonal.

Define also $C_{j\ell} \triangleq G_j'G_\ell$. This is the contingency table of variables j and ℓ. It is sometimes convenient to combine all G_j matrices in one single supermatrix G with dimensions $n \times (\Sigma^m_j k_j)$. The $(\Sigma^m_j k_j) \times (\Sigma^m_j k_j)$ matrix $C = G'G$ has m^2 submatrices C_j with dimensions $k_j \times k$. We can collect the block diagonal submatrices $C_{jj} = D_j$ in a $(\Sigma^m_j k_j) \times (\Sigma^m_j k_j)$ diagonal supermatrix D. The matrix C contains the bivariate marginals and the matrix D the univariate marginals. The matrices C,D,G and H are respectively shown in the tables 1.1 , 1.2 , 1.3 and 1.4.

	a	b	c	p	q	r	u	v	w
a	6	0	0	5	0	1	3	3	0
b	0	2	0	1	1	0	0	2	0
c	0	0	2	2	0	0	0	2	0
p	5	1	2	8	0	0	3	5	0
q	0	1	0	0	1	0	0	1	0
r	1	0	0	0	0	1	0	1	0
u	3	0	0	3	0	0	3	0	0
v	3	2	2	5	1	1	0	7	0
w	0	0	0	0	0	0	0	0	0

Table 1.1 The bivariate marginals (matrix C).

	a	b	c	p	q	r	u	v	w
a	6	0	0	0	0	0	0	0	0
b	0	2	0	0	0	0	0	0	0
c	0	0	2	0	0	0	0	0	0
p	0	0	0	8	0	0	0	0	0
q	0	0	0	0	1	0	0	0	0
r	0	0	0	0	0	1	0	0	0
u	0	0	0	0	0	0	3	0	0
v	0	0	0	0	0	0	0	7	0
w	0	0	0	0	0	0	0	0	0

Table 1.2 The univariate marginals (matrix D).

a	b	c	p	q	r	u	v	w
1	0	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0
1	0	0	0	0	1	0	1	0
1	0	0	1	0	0	1	0	0
0	1	0	1	0	0	0	1	0
0	0	1	1	0	0	0	1	0
1	0	0	1	0	0	1	0	0
1	0	0	1	0	0	0	1	0
0	0	1	1	0	0	0	1	0
1	0	0	1	0	0	0	1	0

Table 1.3 <u>The indicator matrix (matrix G).</u>

a	p	u
b	g	v
a	r	v
a	p	u
b	p	v
c	p	v
a	p	u
a	p	v
c	p	v
a	p	v

Table 1.4 <u>The data matrix H</u>

We also want to use this notation for stochastic variables for which N is infinite. The variable η_j then defines a stochastic variable $\underline{h}_j$ whose values are in K_j, which is a subset of R. $E(\underline{h}_j)$ denotes the expected value of $\underline{h}_j$, $V(\underline{h}_j)$ is the variance. $C(\underline{h}_j,\underline{h}_\ell)$ is the covariance and $R(\underline{h}_j,\underline{h}_\ell)$ is the correlation of $\underline{h}_j$ and $\underline{h}_\ell$.

2. OPTIMAL WEIGHTS

2.1. LINEAR WEIGHTS

We want to find linear weights a_j for the stochastic variables $\underline{h}_1,\ldots,\underline{h}_m$ with $E(\underline{h}_j) = 0$. This is to be done in such a manner that a new stochastic variable $\underline{z}$ with $E(\underline{z}) = 0$ resembles the weigthed variables $a_j\underline{h}_j$, $j=1,\ldots,m$, as much as possible. This "resembling" as much as possible can be stated more formally as the minimization over a and $\underline{z}$ of the following quadratic loss function

$$\sigma(\underline{z};a) = \frac{1}{m}\Sigma_j^m V(\underline{z} - a_j\underline{h}_j),$$

with normalizations either on a : $a'Da = m$, or on $\underline{z}$: $V(\underline{z}) = 1$.

These normalizations are needed because we want to get rid of the solution with $\sigma(\underline{z};a) = 0$ and $\underline{z} = 0$ and $a = 0$. We will omit the proof that both normalizations lead to the same minimum. The plausability of this result is enhanced by the following two interpretations of $\sigma(\underline{z};a)$. Depending on the particular normalization one can describe the minimization of $\sigma(\underline{z};a)$ either as choosing a vector $\underline{z}$ such that the average squared correlation of $\underline{z}$ with $\underline{h}_j$ is as large as possible with the normalization $V(\underline{z}) = 1$, or as the maximizing of the sum of covariances $B = \Sigma_j^m \Sigma_\ell^m C(a_j\underline{h}_j, a_\ell\underline{h}_\ell)$ while the sum of variances $T = \Sigma_j^m V(a_j\underline{h}_j)$ is held constant. The final loss irrespective of normalization is

$$\sigma(*;*) = 1 - \lambda_+,$$

where λ_+ equals the dominant eigenvalue of $\frac{1}{m} R(a_j\underline{h}_j, a_\ell\underline{h}_\ell)$ and where, in terms of the other normalization $V(\underline{z}) = 1$, λ_+ equals the average squared correlation. Both normalizations lead to the same results but differently scaled. The corresponding normalized eigenvector enables us to find the optimal a and $\underline{z}$.

2.2. NON-LINEAR WEIGHTS

We can formulate a loss function for non-linear weighting, that is to be minimized over $\underline{z}$ and ϕ, analogously to the linear weights loss function

$$\sigma(\underline{z};\phi) = \frac{1}{m} \Sigma_j^m V(\underline{z} - \phi_j(\underline{h}_j)),$$

with the normalizations either $\Sigma_j^m V(\phi_j(\underline{h}_j)) = m$ or $V(\underline{z}) = 1$. The non-linear transformations $\phi_j(\underline{h}_j)$ have to resemble the scale $\underline{z}$ as much as possible. The analogy with the linear case applies not only to loss function and normalizations but also to their interpretations, although the transformations are now non-linear and the minimization of $\sigma(\underline{z};\phi)$ is rather complicated. In both the linear and the non-linear case, we are looking for a stochastic variable $\underline{z}$, such that the average correlation ratio of $\underline{z}$ with h_j is as large as possible. We define the correlation ratio as the conditional variance of $\underline{z}$ given $\underline{h}_j$, divided by the variance of $\underline{z}$. This ratio equals the squared correlation coefficient in the linear case. The maximum averaged correlation ratio is equal to the dominant eigenvalue λ_+ of $\frac{1}{m} R(\phi(\underline{h}_j),\phi(\underline{h}_\ell))$, with R as the correlation matrix of the non-linear transformed variables. The final result is again

$$\sigma(*;*) = 1 - \lambda_+.$$

The derivation of this result is due to Gifi (1980). We will have a more detailed look at this, not only to show that the minimal loss can be found but also because we need some of these results in the sequel.

We want to minimize $\sigma(\underline{z};\phi)$ over $\underline{z}$ and ϕ with the normalization $\Sigma_j^m V(\phi_j(\underline{h}_j)) = m$.

Define

$$\sigma(*;\phi) = \min\{\sigma(\underline{z};\phi):\underline{z}\}.$$

The minimum is attained for

$$\underline{z} = \frac{1}{m}\Sigma_j^m\phi_j(\underline{h}_j),$$

and if $\Sigma_j^m V(\phi_j(\underline{h}_j)) = m$, then

$$\sigma(*,\phi) = 1 - m^{-2}\ \Sigma_j^m\Sigma_\ell^m C(\phi_j(\underline{h}_j),\phi_\ell(\underline{h}_\ell)).$$

Thus the original problem is equivalent to minimizing the sum of covariances, while keeping the sum of the variances equal to a constant.

Define for each variable space

$$\mathcal{L}_j = \{\phi_j | E(\phi_j(\underline{h}_j)) = 0\ \&\ V(\phi_j(\underline{h}_j)) < \infty\},$$

a complete orthonormal basis g_{js}, $s = 1,2,\ldots$, such that for every s and t

$$C(g_{js}(h_j),g_{jt}(h_j)) = \delta_{st}\ ,$$

with δ_{st} the Kronecker delta. This entails that the elements on the diagonal of the m x m covariance submatrices C_{st} $(s \neq t)$ between dimensions s and t are zero, the diagonal elements of C_{ss} are equal to one. We can express the transformations of every variable as

$$\phi_j(\underline{h}_j) = \Sigma_s^\infty a_{js} g_{js}(\underline{h}_j)$$

and hence

$$C(\phi_j(\underline{h}_j), \phi_\ell(\underline{h}_\ell)) = \Sigma_s^\infty \Sigma_t^\infty a_{js} a_{\ell t} C(g_{js}(\underline{h}_j), g_{\ell t}(\underline{h}_\ell)).$$

The sum of covariances in this context is

$$B = \Sigma_s^\infty \Sigma_t^\infty a_s' C_{st} a_t,$$

and the sum of variances is

$$T = \Sigma_s^\infty a_s' a_s.$$

The characteristic equation is

$$\Sigma_t^\infty C_{st} a_t = \lambda a_s.$$

The difficulty is, although this is an eigenproblem, that we are dealing with eigenvalues of an infinitely large supermatrix that consists of an infinite number of m x m correlation matrices. It is not particularly easy to compute such eigenvalues except for special cases where additional assumptions can be made. And yet we can use this derivation very well for our purpose, because on one hand we will deal with that special infinite case where we can compute those eigenvalues and on the other hand we will discuss the situation where we can approximate those eigenvalues with finite bases.

2.3 THE SPECIAL INFINITE CASE

With one extra restriction on the infinite correlation matrix of paragraph 2.2 we obtain a new vista of simple results and interpretations. The extra restriction is that we chose our bases in such a way that

$$R(g_{js}(\underline{h}_j),g_{\ell t}(\underline{h}_\ell)) = \delta_{st}r_{sj} .$$

This implies that in terms of the covariance matrix C from the last paragraph we now also assume that all correlations and thus all covariances between dimensions are zero. And this again means that all bivariate distributions are diagonalized. If we use Hermite-Chebyshev polynomials as an infinite basis such diagonalization is allowed for the bivariate normal distribution and for this distribution we may say

$$B = \Sigma_s^\infty a_s' R_s a_s ,$$

and the total minimal loss is equal to

$$\sigma(*;*) = 1 - \frac{1}{m} \max \lambda_+(R_s).$$

We can inspect all R_s separately because the covariances between dimensions are zero. Meanwhile we can always choose a_s as an eigenvector of one of the R_s so that $a_t = 0$ for all $s \neq t$. For a given s this reduces the number of possible vectors of weights to one: a_s, and if g_{js} are orthogonal polynomials all optimal transformations, $a_{js}g_{js}(\underline{h}_j)$, are of order s. Mehler's formula proves that $r_{sj} = (r_j)^s$ in case Hermite-Chebyshev polynomials are used with the bivariate normal distribution. The index s has turned into the power s (Tricomi, 1955 page 254). Another result discussed by Styan (1973) is that, if $R^{(s)}$ denotes the matrix with s-powered correlations, $\lambda_+(R^{(1)}) > \lambda_+(R^{(2)}) > \ldots$. The final loss for the normal distribution amounts to

$$\sigma(*;*) = 1 - \frac{1}{m} \lambda_+(R^{(1)}).$$

This result is exactly the same as the one in paragraph 2.1 for linear weights where the same minimum was found for $\phi_j(\underline{h}_j) = a_j\underline{h}_j$.

Or in other words: applying linear or non-linear weights in this one dimensional case makes no difference. We explicitly fit a linear function in paragraph 2.1 and in the non-linear case the linear function happens to be the best fitting non-linear function. It is clear what we mean by non-linear: not necessarily linear. We will return to this subject later on. It is important to realize that only in this particular case we know that linear functions are the best fitting ones. If for for instance we had mixed two multinormal distributions with identical correlation parameters but with opposite sign, then quadratic functions would have shown the best fit. And only a non-linear technique would recover these functions. Future research is needed to show the behaviour of other distributions in this context. Another rather important result is that, if

$$\underline{h}_j = \psi_j(\underline{e}_j),$$

with $\underline{e}_1,\ldots,\underline{e}_m$ multinormally distributed and ψ_j one-to-one transformations, then our non-linear weighting procedure will find the following optimal transformations

$$\phi_j(\underline{h}_j) = a_j\psi_j^{-1}(\underline{h}_j) = a_j\underline{e}_j.$$

So if we distort the multinormal distribution by using a one-to-one non-linear transformation the original underlying distribution will be recovered by our technique. We will show in paragraph 3 how this recovery is achieved for several serious distortions of the normal distribution. The work in this paragraph is based on De Leeuw (1973),and Gifi (1980). See also Hill (1974).

2.4.HOMALS OR THE APPROXIMATION OF CONTINUOUS NON-LINEAR TRANSFORMATIONS WITH FINITE BASES

The minimization of a non-linear loss leads to an eigenproblem that is not easy to solve, see paragraph 2.2. An alternative with more manageable matrices becomes feasible if we use orthonormal bases for finite dimensional subspaces of

$\mathcal{L}_j$. The space of functions $\mathcal{L}_j$ is a separable Hilbert space which has a denumerably infinite dimension, because we confine ourselves to functions of $\underline{h}_j$ with finite variance. Another restriction that we need is that we partition the space induced by $\underline{h}_j$ into a finite number of measurable sets and define the g_{js} as indicator functions of those sets. This means that for every j a basis of indicator functions is used with a finite number of elements in that basis. In this new basis C and a are partitioned into variables in stead of dimensions. The indicator functions define stepfunctions which are used to approximate the non-linear transformations $\phi_j(\underline{h}_j)$. The matrix C becomes now the supermatrix with bivariate marginals and the matrix D the diagonal supermatrix with the univariate marginals. The indexes are changed because of the new partitioning. Note the fact that the indicator functions are orthogonal and not necessarily orthonormal. So the respective sums of covariances and variances are

$$B = \quad \Sigma_j^m \Sigma_\ell^m a_j' C_{j\ell} a_\ell ,$$

and

$$T = \quad \Sigma_j^m a_j' D_j a_j .$$

The characteristic equation is

$$\Sigma_\ell^m C_{j\ell} a_\ell \; = \; m\lambda D_j a_j .$$

The stationary equation that we find if we require $V(\underline{z}) = 1$ is

$$\frac{1}{m} \;\; \Sigma_j^m \; P_j(\underline{z}) \; = \; \lambda \underline{z}$$

with P_j the orthogonal projector on the subspace spanned by the $g_{js}(\underline{h}_j)$. These two eigenvalue problems have the same eigenvalues because they are both derived from the singular value problem

$$\lambda^{\frac{1}{2}}\underline{z} \; = \; \frac{1}{m} \; \Sigma_j^m \phi_j(\underline{h}_j) \quad \text{and} \quad \lambda^{\frac{1}{2}}\phi_j(h_j) \; = \; \Sigma_s^\infty a_{js} g_{js}(\underline{h}_j) ,$$

with $\quad a_{js} \; = \; C(g_{js}(h_j), z)/V(g_{js}(h_j)).$

This discrete finite approach can be interpreted in two ways. First as a method to <u>find</u> the optimal transformation $\phi_j(h_j)$ in case h_j is a discrete stochastic variable with k_j different values. This is the interpretation of Horst (1936), Guttman (1941), Johnson (1950), Lord (1958), Bock (1960), De Leeuw (1973) and Nishisato (1978). And secondly as an <u>approximation</u> of the optimal transformation $\phi_j(h_j)$ with a step-function with k_j steps. This approach is more common to the French school like Dauxois & Pousse (1976) and Lafaye de Micheaux (1978) and it is the approach advocated in this paper. Both interpretations are treated by Gifi (1980) and Stoop (1980).

The computerprogram that solves the stationary equations from the previous section is called HOMALS. Before we discuss the algorithm it is necessary to realize that z is now a vector of length n and the sum of squares of z is equal to n, z contains the values of $\underline{z}$ and SSQ(z) correspond with $V(\underline{z})$. The optimal transformation is $G_j a_j$, where G_j is an n x k_j indicator matrix and $a_j = D_j^{-1}G_j'z$. The orthogonal projector P_j is equal to $G_jD_j^{-1}G_j'$. The algorithm is iterative and one iteration consists of the two steps (1) and (2). The values of a preceding iteration have a high index ^{0} and the updates in the next iteration are indexed with $^\times$. The algorithm starts with standardized random numbers between 0 and 1 as an estimate of z. Steps (1) and (2) are alternated until the difference in stress between two successive iterations is smaller than a previously decided small number.

(1) $$a_j^0 = D_j^{-1}G_j'z^0$$

(2) $$z^\times = \tilde{z}(\tilde{z}'\tilde{z})^{-\frac{1}{2}}$$

with $$\tilde{z} = \Sigma_j^m G_ja_j^0.$$

Step (1) is executed succesively for all variables within one iteration. The stress is equal to $1 - \frac{1}{m}\Sigma_j^m a_j'D_ja_j$ and the loss is minimized over z and a with restrictions SSQ(z) = 1 and u'z = 0. The computerprogram and algorithm are actually p-dimensional with matrices Z (n x p) and A_j (k_j x p). The value of p is chosen by the user. In this paper we only discuss the situation where p = 1. The algorithm is fast and efficient because the sparseness of the indicator matrices reduces multiplication to addition while each indicator matrix is stored as a vector, and because the iterative approach of singular vectors is used in stead of a complete eigenvalue decomposition. Although Richardson suggested this algorithm in a rudimentary form in 1935, the nonlinear weighting problem was always solved by the complete eigenvalue decomposition of the scaled C matrix of bivariate marginals, which made the analysis prohibitive if the number of categories was very large, cf. Lingoes (1968). The ALS algorithm here was suggested by De Leeuw in 1976.

3. HOMALS IN MONTE CARLO

3.1 DISCRETIZATION

We want to approximate continuous non-linear transformations with a discrete technique. This entails that continuous variables are discretized by mapping continuous intervals into points with stepfunctions. This problem is known as quantization in communication research, the communication theory people have also studied optimal discretizations. But here we are interested in non-optimal discretizations as well. The idea is that we deliberately distort the multinormal distribution by our discretizations up to a certain degree. The range varies from optimal via pseudo-optimal, U-shape, equal to skew discretization. The number of intervals is arbitrarily fixed at five. Two discretizations follow the original distribution rather closely, the other three on the contrary ignore the original distribution and are modelled after some common types of marginal distributions in data analysis. There exists a loss function that actually describes the loss caused by a certain discretization for a fixed number of categories with respect to the normal distribution. This loss can be computed for all our discretizations and we assume that it will be minimal for the optimal approach, that is why we call it optimal, and larger for the other discretizations. The loss , which is to be minimized over all stepfunctions $\phi(\underline{h})$ with five steps is

$$\varepsilon^2 = V(\underline{h} - \phi(\underline{h}))^2,$$

where $\underline{h}$ is univariate standard normal.
Max (1960) computed optimal discretization points for several numbers of categories and the corresponding minimal ε^2 values. For further details and many references see Gifi (1980).

Tables 3.1 to 3.5 illustrate the discretizations we used with the discretization points and corresponding histograms.

The optimal discretization. This approach is optimal in the sense that it minimizes Max's ε^2 parameter and compared with our other discretizations with the same number of categories, it will produce the greatest eigenvalue. The distribution of the surface under the curve over intervals is: .1067 .2444 .2978 .2444 and .1067 . The total surface under the curve is equal to one. The corresponding histogram is rather flat. See figure 3.1

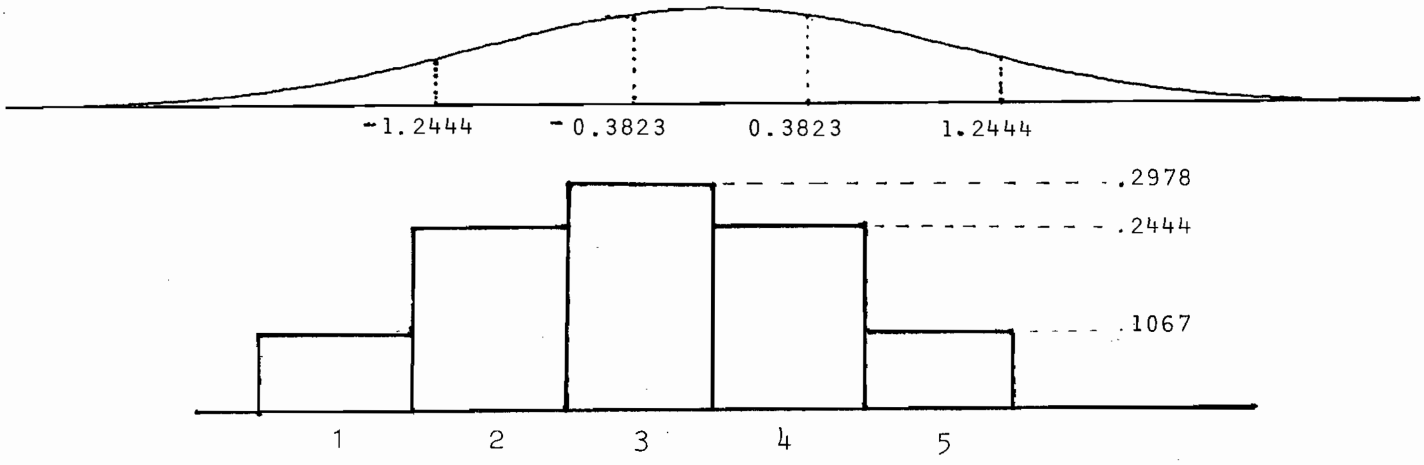

Figure 3.1 Histogram and discretization points of the optimal discretization.

The pseudo-optimal discretization. We have also computed a " quick " approximation to the optimal discretization. The surfaces under the curve corresponding to the intervals are: .1 .2 .4 .2 .1 . The curve and histogram are shown below.

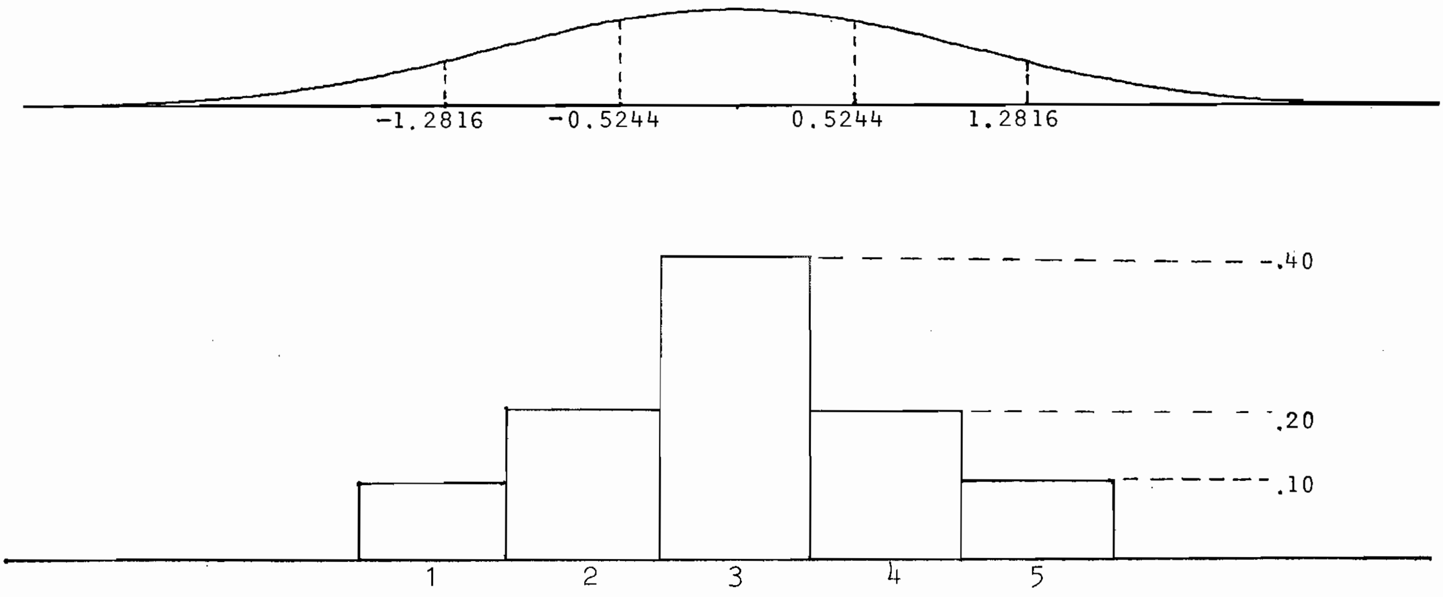

Figure 3.2 Histogram and discretization points of the pseudo optimal discretization.

The U-shape discretization. The shape of the histogram in figure 3.3 needs no explanation. The distribution of the surfaces under the curve is .3 .15 .1 .15 .3 . This U-shape pattern is often found in questionnaires with rating scales of the type: strongly agree-agree-neutral-disagree-strongly disagree. The curve and histogram are shown below.

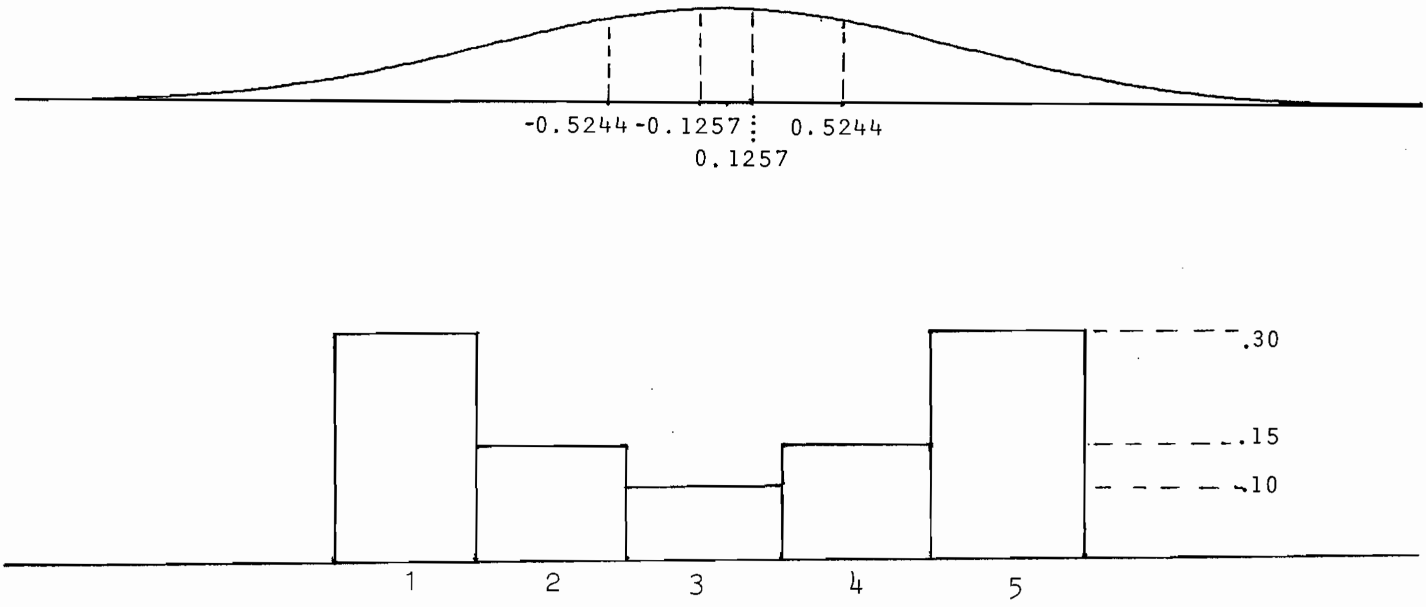

Figure 3.3 Histogram and discretization points of the U-shape discretization.

The equal discretization. Again the picture in fig. 3.4 is self-explanatory. The distribution of the surface under the curve is obviously .2 .2 .2 .2 .2 .

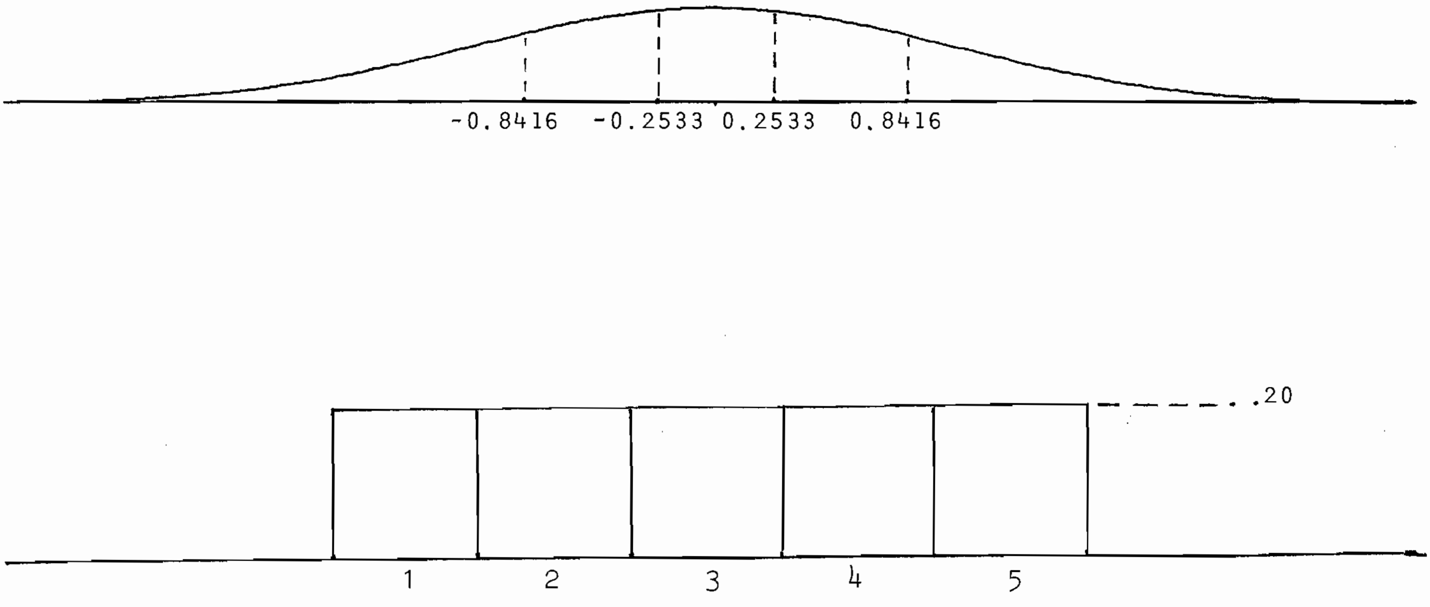

Figure 3.4 Histogram and discretization points of the equal discretization.

The skew discretization. This discretization together with the U-shape is the least like the normal distribution. The surface distribution values are: .45 .25 .15 .10 .05 . See fig. 3.5 .

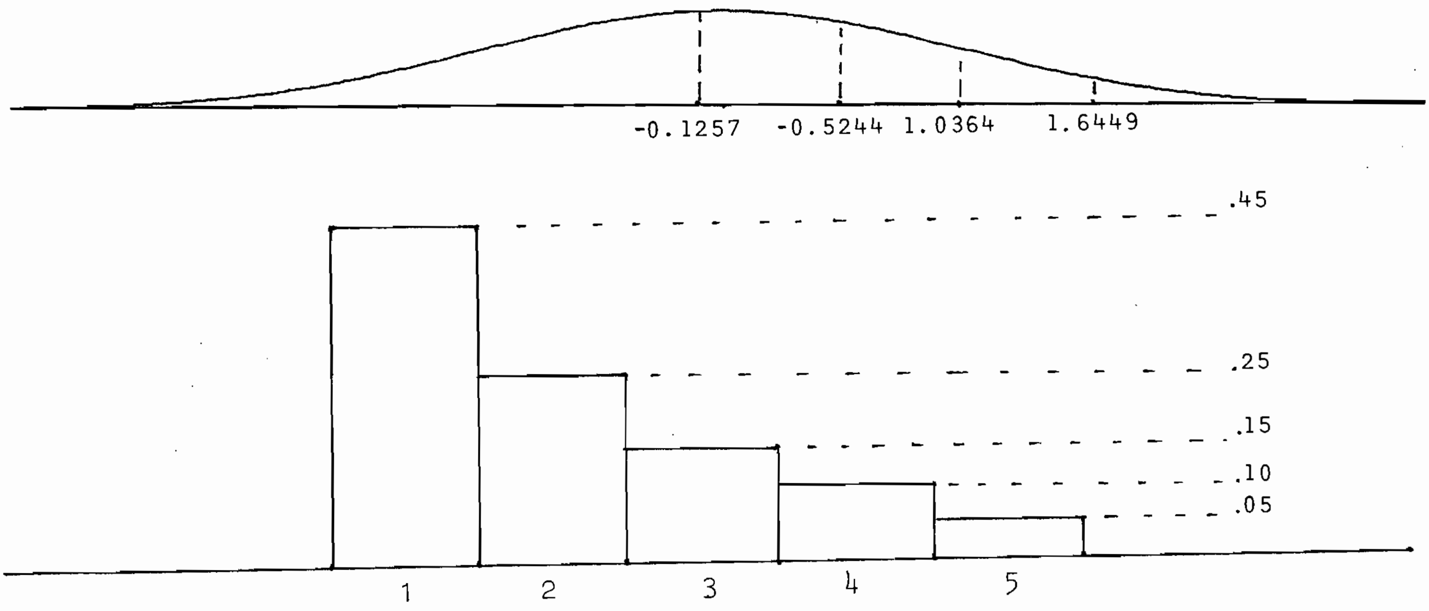

Figure 3.5 Histogram and discretization points of the skew discretization.

One has to keep in mind that all discretizations here are results of non-linear monotonic transformations.

3.2 THE POPULATION VALUES OF THE HOMALS PARAMETERS.

As in paragraph 2.3, in the one-dimensional case with multinormally distributed variables, the optimal ϕ_j equals identity, and λ_+ is the greatest eigenvalue of $\frac{1}{m}$ R, in which R is the correlation matrix of the variables. In this Monte Carlo study we have 9 variables, all with variance equal to 1, and all mutual correlations are equal to .5, so for the continuous distribution the dominant eigenvalue is

$$\lambda_+ = 5/9 = .5556$$

(For an m x m matrix with all diagonal elements equal to 1 and all off-diagonal elements equal to ρ, the greatest eigenvalue is of course $1 + (m - 1)\rho$ and the corresponding eigenvector has equal elements.)

After the discretization in one of the five ways from par. 3.1, we can compute the bivariate probability distributions for each pair of variables. The probability that variable j is in category i and variable ℓ is in category k is equal to the probability that $\underline{h}_j$ is in the i-th interval and $\underline{h}_\ell$ is in the k-th interval, where $\underline{h}_j$ and $\underline{h}_\ell$ have variance 1 and covariance .5. The intervals are defined by the different types of discretizations.[1]
Once we have these probabilities, we can compute the optimal ϕ_j and λ_+ by means of HOMALS.
Though we are dealing with nine variables , we can compute the population values by using only two variables. Owing to symmetry we can then translate our results to the case of nine variables. The results are given in Table 3.1. All nine variables have by symmetry identical optimal transformations for each kind of discretization.

		category number	OPTIMAL	PSEUDO-OPTIMAL	U-SHAPE	EQUAL	SKEW
LINEAR	EIGENVALUE		0.5216	0.5183	0.4873	0.5135	0.4954
	OPTIMAL TRANSFORMATIONS	1	-1.7225	-1.8212	-1.2156	-1.4114	-0.8729
		2	-0.8612	-0.9106	-0.6078	-0.7057	-0.0430
		3	0.0	0.0	0.0	0.0	0.7868
		4	0.8612	0.9106	0.6078	0.7057	1.6167
		5	1.7225	1.8212	1.2156	1.4114	2.4465
NON-LINEAR	EIGENVALUE		.5222	.5183	.4938	.5160	.4981
	OPTIMAL TRANSFORMATIONS	1	-1.7714	-1.8150	-1.2612	-1.4630	-0.9186
		2	-0.8165	-0.9169	-0.3763	-0.5901	0.0942
		3	0.0	0.0	0.0	0.0	0.7823
		4	0.8165	0.9169	0.3763	0.5901	1.4776
		5	1.7714	1.8150	1.2612	1.4630	2.4703

Table 3.1 The approximated theoretical population parameters.

[1] These probabilities are actually computed for an approximation of the normal distribution: the sum of 12 uniform (½,½ variates. This is done, because all our samples are taken from this distribution using the SSP Fortran subroutine GAUSS.

3.3 THE BOOTSTRAP.

3.3.1 DESCRIPTION.

As a method for estimating the variance of the various HOMALS parameters we use the Efron-bootstrap (Efron, 1979-1). This method amounts to the following:
We have n observations, from which we compute the HOMALS parameters. Next we take a bootstrap sample. This is a sample with replacement of size n taken from the original set of observations, where all observations have a probability of n^{-1} to be taken. We apply HOMALS on this bootstrap sample. This procedure is repeated several times, each sample taken from the original set of observations. For all HOMALS parameters this gives us a number of replications.

3.3.2 STATISTICAL APPROACH.

In HOMALS we are dealing with categorical data: we have a finite number of variables, each of which can assume a finite number of values. Suppose the n observations we start with consist of a random sample from some unknown population. In this population only a finite number, say q, of sets exists in such a manner that all members in one set have the same score on all variables, and members of two different sets have different scores on at least one of the variables. We can represent the probabilities, that we sample a member of a certain set by a vector $\pi \in R^q$ consisting of non-negative numbers adding up to 1.
A sample of size n can be represented as a vector in R^q of observed frequencies, as well as a vector $\underline{x}^{(n)} \in R^q$ of observed relative frequencies. The representation $\underline{x}^{(n)}$ ignores the information about the total number of observations, but the advantage is that $\underline{x}^{(n)}$ is an estimator of π. And the HOMALS parameters do not change anyway if all frequencies are multiplied with a constant (Gifi ,1980). This is why we can define a set $D := \{x \in R^q; x_i > 0 \ (i = 1,\ldots,q), \Sigma_i^q x_i = 1\}$ and we can write every single HOMALS parameter as a function $f : D \to R$.

If $\underline{x}^{(n)}$ is an estimate of π,then $f(\underline{x}^{(n)})$ is an estimate of $f(\pi)$. Now we can use the bootstrap method to estimate the bias and variance of $f(\underline{x}^{(n)})$. We can represent a bootstrap sample originating from $\underline{x}^{(n)}$ as a vector $\underline{y}(\underline{x}^{(n)})$. Once again this is a vector of non-negative numbers adding to one. We construct $\underline{y}(\underline{x}^{(n)})$ by taking a sample of size n from a multinomial distribution defined by $\underline{x}^{(n)}$, and by considering the corresponding relative frequencies. Note the fact that going from $x^{(n)}$ to $\underline{y}(\underline{x}^{(n)})$ is essentialy the same as going from π to $\underline{x}^{(\bar{n})}$. This leads us to the conclusions that,with weak restrictions on f (Bettonvil & De Leeuw, forthcoming)

$$E\ f(\underline{x}^{(n)}) - f(\pi);\ E((E\ f(\underline{y}(\underline{x}^{(n)})) - f(\underline{x}^{(n)}))|\underline{x}^{(n)};$$
$$V\ f(\underline{x}^{(n)});E(V\ f(\underline{y}(\underline{x}^{(n)}))|\underline{x}^{(\bar{n})})$$

are of order n^{-1} and that

$$(E\ f(\underline{x}^{(n)}) - f(\pi)) - (E(E\ f(\underline{y}(\underline{x}^{(n)})) - f(\underline{x}^{(n)}))|\underline{x}^{(n)})$$

and

$$(V\ f(\underline{x}^{(n)})) - (E(V\ f(\underline{y}(\underline{x}^{(n)}))|\underline{x}^{(n)}))$$

are of the order n^{-2}.

Thus we can use the bias of $f(\underline{y}(\underline{x}^{(n)}))$ with respect to $f(\underline{x}^{(n)})$ as an estimate of the bias of $f(\underline{x}^{(n)})$ with respect to $f(\pi)$; and analogously we can use the variance of $f(\underline{y}(\underline{x}^{(n)})|\underline{x}^{(n)})$ as an estimate of the variance of $f(\underline{x}^{(n)})$.

We now define a new estimate of $f(\pi)$, which we call the bootstrap pseudovalue

$$2\ f(\underline{x}^{(n)}) - f(\underline{y}(\underline{x}^{(n)})).$$

This new estimate has a bias of order n^{-2}. In this paper we only refer to these pseudo values. In the Efron-paper the bias reducing version of the bootstrap was not given, and consequently bias reduction seemed to be an advantage for the Quenouille-Tukey jackknife (Miller, 1974).

Other reasons why we prefer the bootstrap over the jackknife in this application are the independence of the bootstrap replications, the simplicity of the method and the possibility to take as many bootstrap samples as desired.

3.4 THE RESULTS.

3.4.1 THE TRANSFORMATIONS.

The main result is evidently the recovery of the multinormal distribution for all discretizations and nearly all sample-sizes. This conclusion is based on the monotone transformations of the discretized intervals. We have seen in par. 2.3 that the optimal transformations were linear and thus monotonic. The discretization disturbed this linearity somewhat but the weaker constraint of monotonicity should hold at all events. This is the case, apart from six violations, i.e. wrongly placed categories, in the analyses with n = 100. See fig. 3.6 3.15and 3.18.The larger samples of 1000 and 10000 observations have all monotone transformations. This means that even for the worst distortitions, like skew and U-shape, the original distribution is recovered for nearly all samples.
However we are not only interested in recovery but also in the stability of the transformations. The bootstrap points in figures 3.6 → 3.20 are very illustrative in this respect. The more they are spread in a vertical direction, the more unstable the transformations are. The S-points in these plots are the sample transformations and as such the estimates of the P-points, population transformations. The *-points, the Bootstrap transformations are also estimates of the P-points but on the other hand the dispersion of the *-points is an estimate of the dispersion of the S-points. Really unstable transformations occur only when both the sample-size is small (=100) and the original distribution is heavily distorted like it is the case with the skew and U-shape discretization and to a lesser extent with the equal discretization. The most obstinate in this respect is the U-shape discretization. Not surprisingly this is the discretization with the smallest population eigenvalue. The direction of the deviations in itself are also interesting (Bettonvil & De Leeuw, forthcoming).

3.4.2 THE EIGENVALUES.

The discretized population eigenvalues from par. 3.2 are nicely approximated by all our samples. In section 2 we extensively discussed the properties of these eigenvalues. The population eigenvalues, the sample eigenvalues, the means of the pseudo bootstrap eigenvalues, and their variances for all discretizations are collected in table 3.2 for the linear case and in table 3.4 for the non-linear case. The relation between eigenvalue and type of discretization is already discussed in par. 3.2.
The approximation of the population eigenvalue is better as the sample size increases and there is no discretization whose eigen value is better approximated than those of the other discretizations. Also the variances seem to be independent of discretization.
The bias reduction of the bootstrap values is only effective for the discretization with the smallest eigenvalue. The eigenvalues of the non-linear approach tend to be somewhat larger than the linear ones, especially for smaller sample sizes. The non-linear approach has more freedom to correct for the discretization bias and for the small sample bias, which explains this phenomenom.

	POPULATION	N	SAMPLE	BOOTSTRAP	
				MEAN	VARIANCE
OPTIMAL	.5216	100	.4665	.4754	$7.8*10^{-4}$
		1000	.5313	.5355	$1.0*10^{-4}$
		10000	.5192	.5206	$1.2*10^{-5}$
PSEUDO- OPTIMAL	.5183	100	.5140	.5135	$7.4*10^{-4}$
		1000	.5261	.5247	$9.6*10^{-5}$
		10000	.5189	.5197	$1.3*10^{-5}$
U-SHAPE	.4873	100	.4534	.4519	$4.7*10^{-4}$
		1000	.4950	.4947	$1.8*10^{-4}$
		10000	.4897	.4899	$3.5*10^{-6}$
EQUAL	.5135	100	.4438	.4464	$1.6*10^{-3}$
		1000	.5038	.4969	$1.1*10^{-4}$
		10000	.5161	.5151	$8.9*10^{-6}$
SKEW	.4954	100	.5244	.5252	$1.1*10^{-3}$
		1000	.5153	.5156	$2.1*10^{-4}$
		10000	.4951	.4954	$2.1*10^{-5}$

Table 3.2 Eigenvalues of the linear transformations.

	POPULATION	N	SAMPLE	BOOTSTRAP	
				MEAN	VARIANCE
OPTIMAL	.5222	100	.4836	.4762	$9.6*10^{-4}$
		1000	.5339	.5369	$9.6*10^{-5}$
		10000	.5197	.5208	$1.2*10^{-5}$
PSEUDO-OPTIMAL	.5183	100	.5360	.5229	$9.0*10^{-4}$
		1000	.5293	.5270	$9.3*10^{-5}$
		10000	.5192	.5198	$1.3*10^{-5}$
U-SHAPE	.4938	100	.4876	.4582	$1.1*10^{-3}$
		1000	.5045	.5013	$1.6*10^{-4}$
		10000	.4964	.4964	$4.6*10^{-6}$
EQUAL	.5160	100	.4606	.4359	$1.0*10^{-3}$
		1000	.5069	.4988	$1.0*10^{-4}$
		10000	.5185	.5174	$8.9*10^{-6}$
SKEW	.4981	100	.5535	.5327	$1.3*10^{-3}$
		1000	.5187	.5169	$1.8*10^{-4}$
		10000	.4985	.4984	$2.0*10^{-5}$

Table 3. 3 Eigenvalues of the non-linear transformations.

3.4.3 PERMUTED CATEGORY SCORES.

The fact that the linear transformations are the best fitting transformations for the discretized normal distribution requires that the category numbers are monotone with the order of the intervals of the continuous distribution and that they are equally spaced. These restrictions are not relevant for non-linear transformations. We have permuted the category numbers and the tables 3.4 to 3.7 show clearly that linear fit has become very poor while the non-linear fit is as good as ever. Unfortunately we failed to use only one sample for this experiment. The non permuted linear analysis is performed on another sample as the permuted run. But because of large sample size we may assume that the parameters are in the same order generalized over samples over samples. It is worth noticing that the linear analysis used approximately 12% more CPU in the permuted case than in the non permuted case. We used the correlation matrices because they illustrate the effect nicely. These are the correlations computed with the optimally, linear or non-linear, transformed variables. The according transformations are shown in figures 3.21 and 3.22.

λ = .2163 (Sample 1)

0.080							
0.017	-0.021						
-0.217	-0.104	0.025					
-0.290	-0.099	0.010	0.188				
-0.343	-0.089	-0.006	0.215	0.260			
0.247	0.036	0.018	-0.106	-0.133	-0.181		
-0.015	0.058	-0.033	-0.020	-0.020	-0.012	-0.010	
0.072	0.013	-0.003	-0.041	-0.053	-0.054	0.024	-0.006

Table 3.4 The correlation matrix based on linear transformations of permuted category numbers; n=10000, Skew.

λ = .4949 (Sample 1)

0.428							
0.430	0.434						
0.432	0.434	0.431					
0.441	0.437	0.442	0.432				
0.430	0.428	0.438	0.437	0.436			
0.432	0.427	0.442	0.422	0.427	0.432		
0.446	0.434	0.426	0.436	0.435	0.428	0.419	
0.423	0.414	0.428	0.426	0.446	0.418	0.427	0.438

Table 3.5 The correlation matrix based on non-linear transformations of permuted category numbers; n=10000, Skew.

λ = .4951 (Sample 2)

0.443							
0.438	0.437						
0.436	0.444	0.428					
0.426	0.434	0.435	0.434				
0.435	0.441	0.432	0.431	0.430			
0.430	0.427	0.436	0.435	0.434	0.421		
0.415	0.423	0.425	0.427	0.432	0.429	0.431	
0.425	0.431	0.437	0.433	0.432	0.426	0.445	0.422

Table 3.6 The correlation matrix based on linear transformations of non-permuted category numbers; n=10000; Skew.

$\lambda = .4985$ (Sample 2)

0.446							
0.441	0.440						
0.441	0.448	0.432					
0.429	0.438	0.438	0.438				
0.440	0.445	0.434	0.434	0.432			
0.434	0.434	0.441	0.441	0.437	0.429		
0.419	0.428	0.428	0.431	0.434	0.434	0.435	
0.427	0.436	0.441	0.435	0.434	0.429	0.449	0.425

Table 3.7 The correlation matrix based on non-linear transformations of non-permuted category numbers; n=10000, Skew.

To show the permutation effect we arbitrarily chose the skew discretization.

3.5 THE PROCEDURE.

We generated a random sample of nine continuous multinormally distributed centered variables with variance equal to 1 and an overall correlation coefficient equal to .5.
For each discretization we partitioned such a sample into five continuous connected non-overlapping intervals. All elements in such an interval are categorized into a discrete category with a corresponding category number.
We computed the optimal *linear* transformations of the discretized sample with a *singular value decomposition* and used these optimal values as an initial configuration for the computation of the optimal *non-linear* transformations with *HOMALS*.
10 random bootstrap samples with replacement and with the same size as the earlier mentioned discretized sample were taken from this discretized sample. Linear and non-linear optimal transformations were computed of each of these samples.
We repeated this procedure for five different discretizations and three different sample-sizes. For all these 15 combinations we used different samples. This means 15 different samples and 150 different bootstrap samples taken from these 15 discretized samples.
The computer program that computes the non-linear transformations is called HOMALS and the computer program that embodies both the optimal linear transformations and the optimal non-linear transformations in one analysis is called PRINCALS (De Leeuw & van Rijckevorsel, 1979). Both programs are fully documented and they are available from the Department of Datatheory, Breestraat 70, University of Leiden.

LITERATURE.

B. Bettonvil & J. de Leeuw: The Efron bootstrap as a method for biasreduction. Forthcoming.

R.D. Bock: Methods ans applications of optimal scaling. University of North Carolina, L.L. Thurstone Lab. Report 25, 1960.

J. Dauxois & A. Pousse: Les analyses factorielles en cacul des probabilités et en statistique: essai d'étude synthétique. Dissertation, Université Paul Sabatier, Toulouse, 1976.

B. Efron: Bootstrap methods: another look at the jackknife. Ann. Statist., 7, 1979, 1-26.

B. Efron: Computers and the theory of statistics: thinking the unthinkable. Siam Review, 21, 1979, 460-480.

A. Gifi: Niet-lineaire multivariate analyse. Dep. of Datatheorie, Univ. Leiden, 1980.

L. Guttman: The quantification of a class of attributes: a theory and method of scale construction. In: P. Horst (ed): The prediction of personal adjustment. New York, SSRC, 1941.

M.O. Hill: Correspondence analysis: a neglected multivariate method. Journal of the royal statistical society, series C, 23, 1974, 340-354.

P. Horst: Obtaining a composite measure from a number of different measures of the same attribute. Psychometrika, 1, 1936, 53-60.

H. Hotelling: Analysis of a complex of statistical variables into principal components. J. Educ. Psychol, 24, 1933, 417-441, 498-520.

P.O. Johnson: The quantification of qualitative data in discriminant analysis. J. Am. Statist. Ass., 45, 1950, 65-76.

D. Lafaye de Michaux: Approximation d'analyse canoniques nonlinéaires de variables aléatoires. Dissertation, Université de Nice, 1978.

J. de Leeuw: Canonical analysis of categorical data. Doct. Dissertation University of Leiden, 1973

J. de Leeuw: HOMALS. Paper read at the Psychometric Society Meeting, Murray Hill N.Y.,1976.

J. de Leeuw & J.L.A. van Rijckevorsel: HOMALS & PRINCALS. Paper presented at the Second International Symposium on Data Analysis and Informatics, Versailles, 1979.

J.C. Lingoes: The multivariate analysis of qualitative data. Mult. Behavioural Res. , 3, 1968, 61-94.

F.M. Lord: Some relations between Guttman's principal components of scale analysis and other psychometric theory. Psychometrika, 23, 1958, 291-296.

J. Max: Quantizing for minimum distortion. Proceedings IEEE, Information theory, 6, 1960, 7-12.

R.G. Miller: The jackknife: a review. Biometrika, 61, 1974, 1-15.

S. Nishisato: Analysis of categorical data: dual scaling and its applications. Unpublished manuscript.

K. Pearson: On lines and planes of closest fit to points in space. Phil. Mag., 2, 1901, 559-572.

J.L.A. van Rijckevorsel & J. de Leeuw: An outline of HOMALS. Dep. of Datatheorie, University of Leiden, 1980.

I. Stoop: Secundaire analyse van de tot jaar tot jaar data. Unpublished master's thesis. Dep. of Datatheory, University of Leiden, 1980.

G.P. Styan: Hadamard products and multivariate statistical analysis. Linear Algebra Appl., 6, 1973, 217-240.

F.G. Tricomi: Vorlesungen uber Orthogonalreihen. Berlin, Springer, 1955.

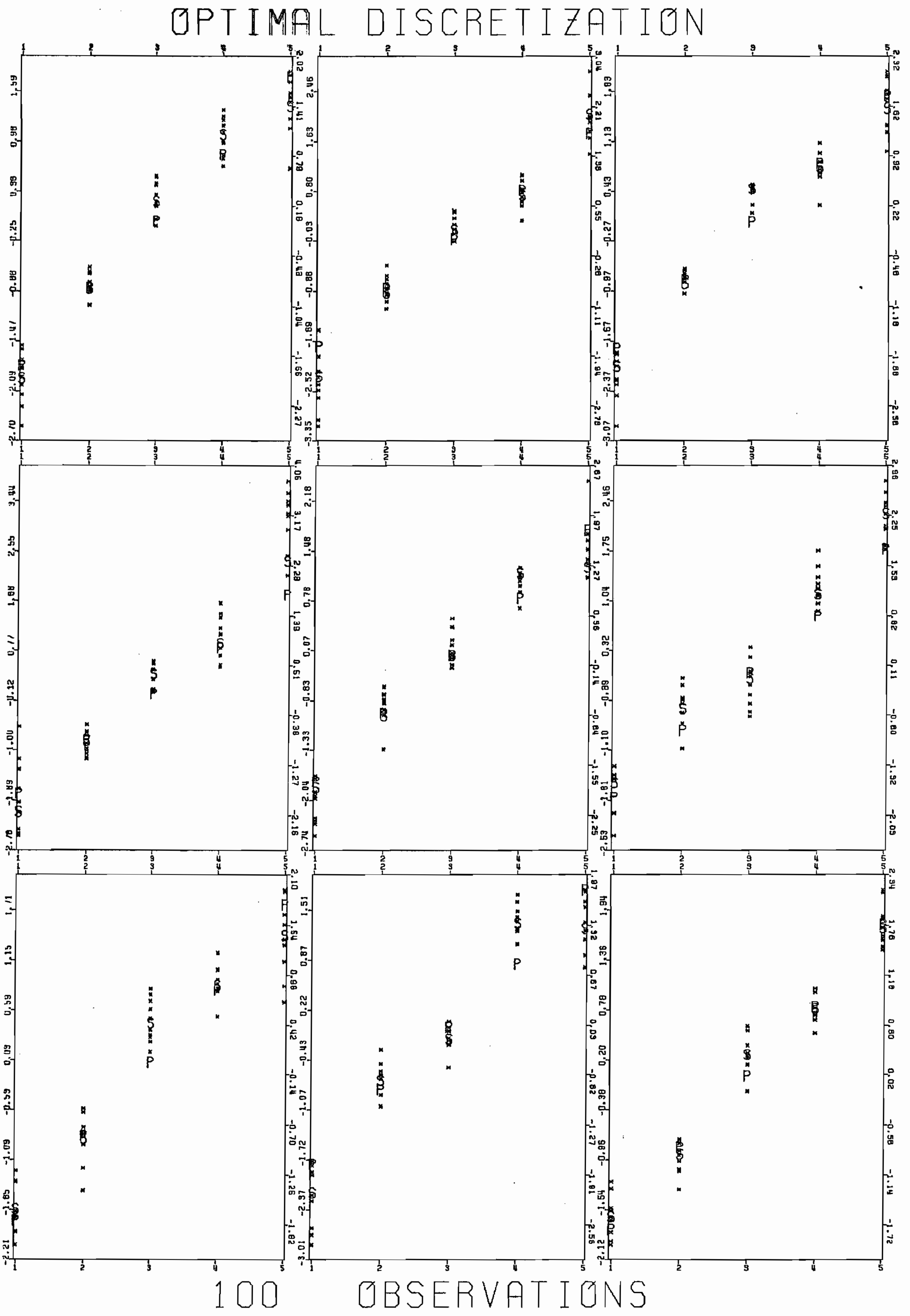

Figure 3.6

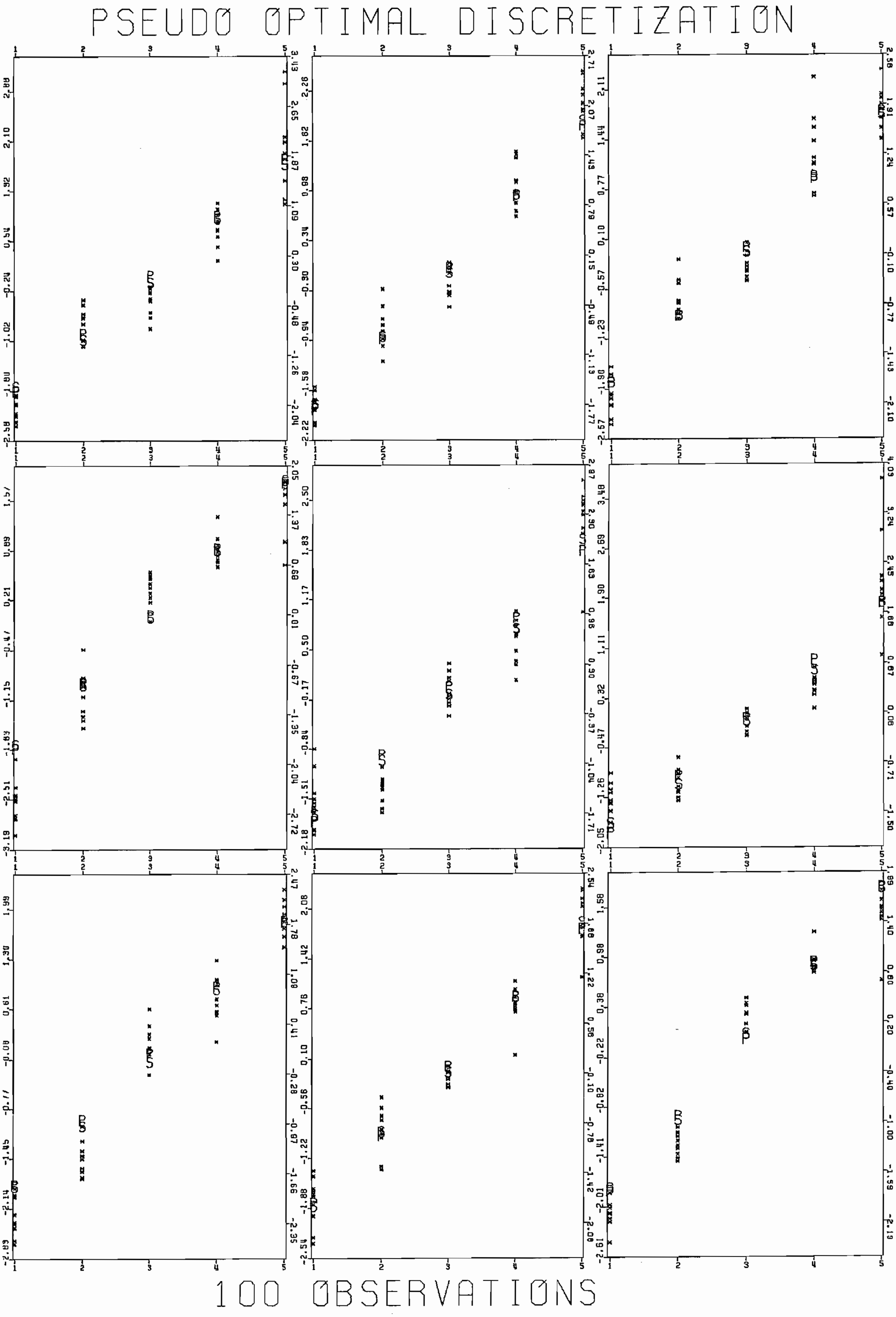
PSEUDO OPTIMAL DISCRETIZATION
1
2
3
4
5
-2.58
-1.80
-1.02
-0.24
0,54
1,32
2,10
2,88
-2.04
-1.26
-0.48
0,30
1,09
1,87
2,65
3,43
-2.22
-1.58
-0.94
-0.30
0,34
0,98
1,62
2,26
-1.77
-1.13
-0.49
0,15
0,79
1,43
2,07
2,71
-2.57
-1.90
-1.23
-0.57
0,10
0,77
1,44
2,11
-2.10
-1.43
-0.77
-0.10
0,57
1,24
1,91
2,58
-3.19
-2.51
-1.83
-1.15
-0.47
0,21
0,89
1,57
-2.72
-2.04
-1.35
-0.67
-0.01
0,68
1,37
2,05
-2.18
-1.51
-0.84
-0.17
0,50
1,17
1,83
2,50
-1.71
-1.04
-0.37
0,90
0,98
1,63
2,90
2,87
-2.05
-1.26
-0.47
0,32
1,11
1,90
2,69
3,48
-1.50
-0.71
0,08
0,87
1,66
2,45
3,24
4,03
-2.83
-2.14
-1.45
-0.77
-0.08
0,61
1,30
1,99
-2.35
-1.66
-0.97
-0.28
0,41
1,08
1,78
2,47
-2.54
-1.88
-1.22
-0.56
0,10
0,76
1,42
2,08
-2.08
-1.42
-0.76
-0.10
0,56
1,22
1,88
2,54
-2.61
-2.01
-1.41
-0.82
-0.22
0,38
0,98
1,58
-2.19
-1.59
-1.00
-0.40
0,20
0,80
1,40
1,99
100 OBSERVATIONS

Figure 3.7

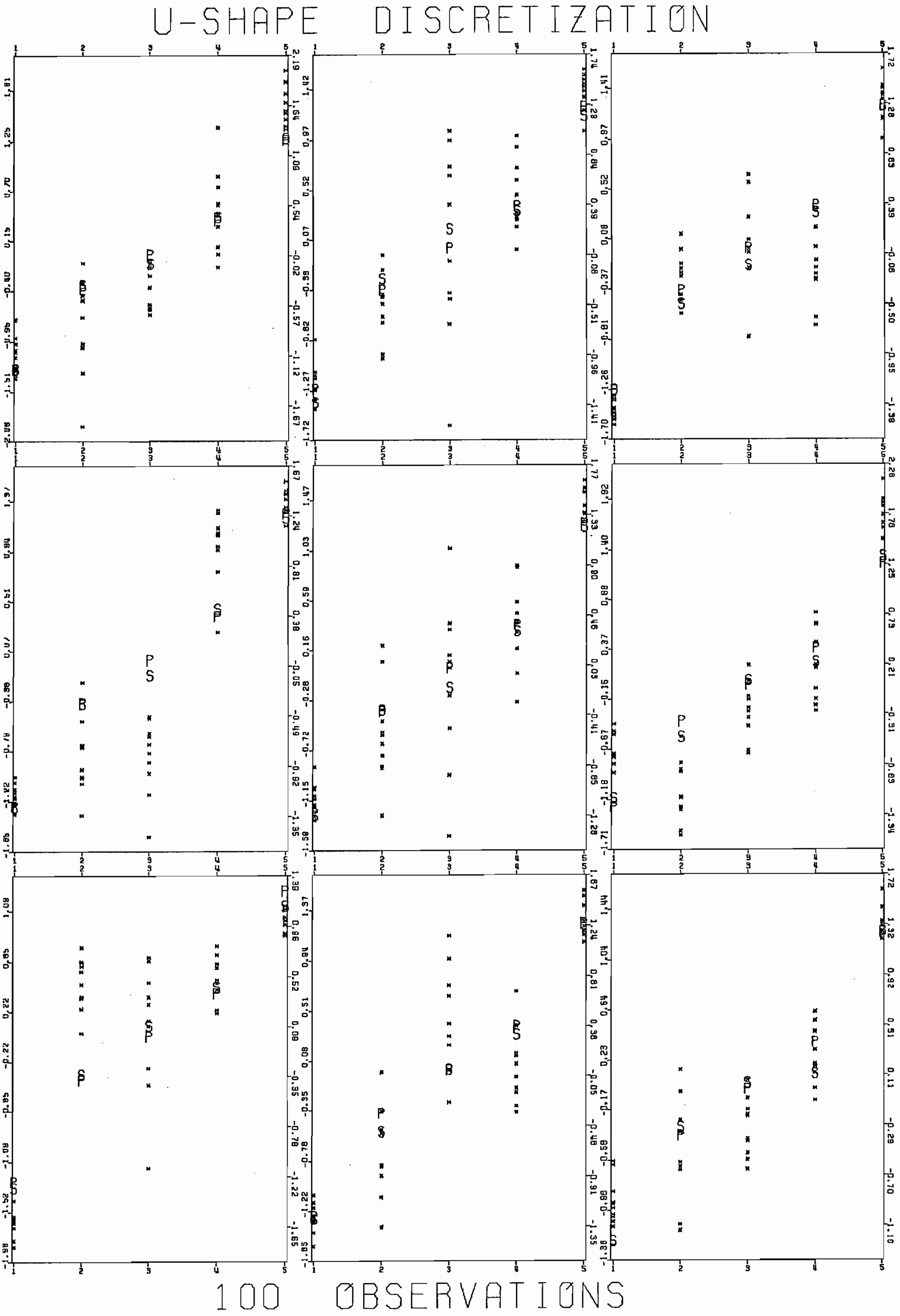

Figure 3.8

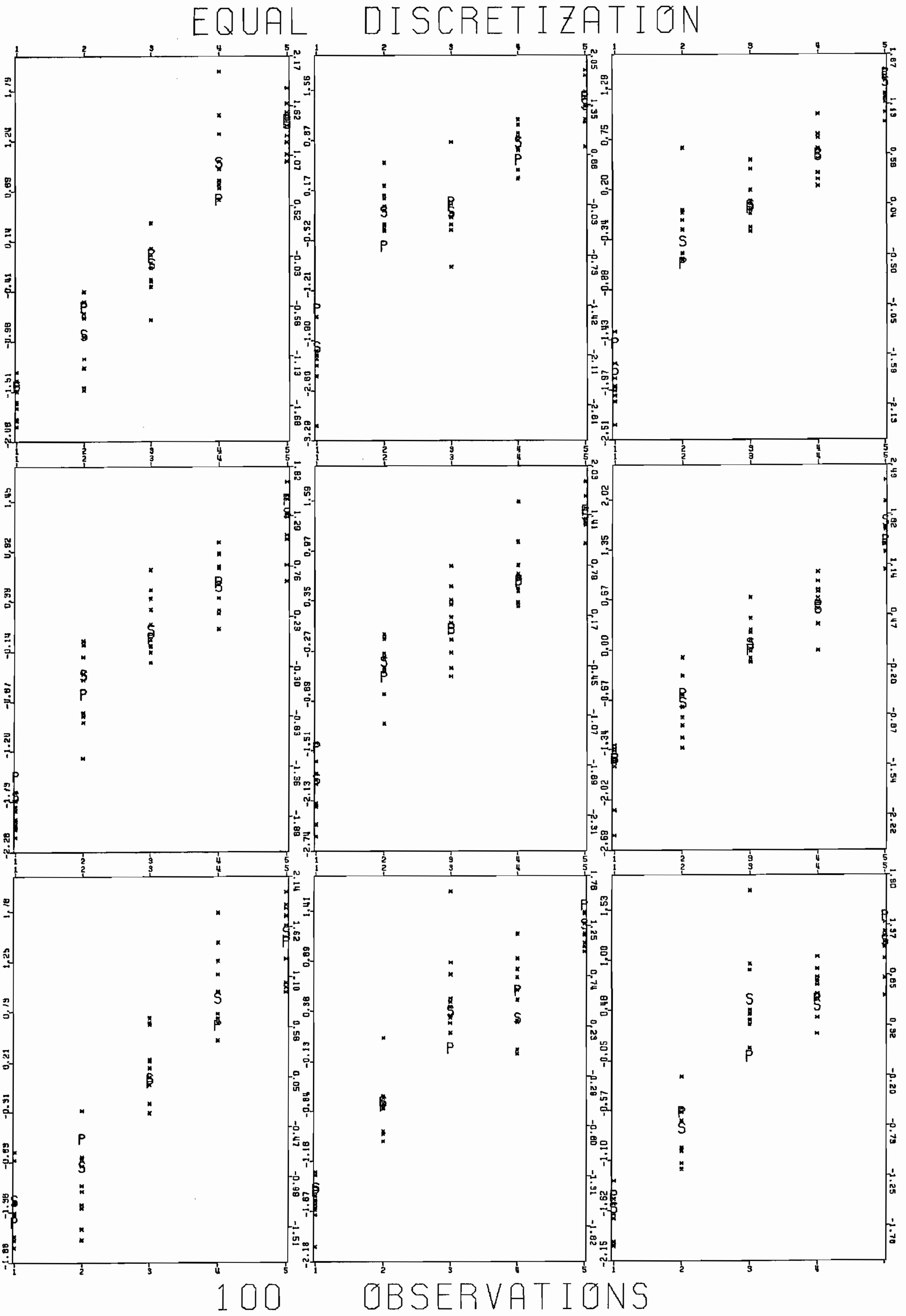

Figure 3.9

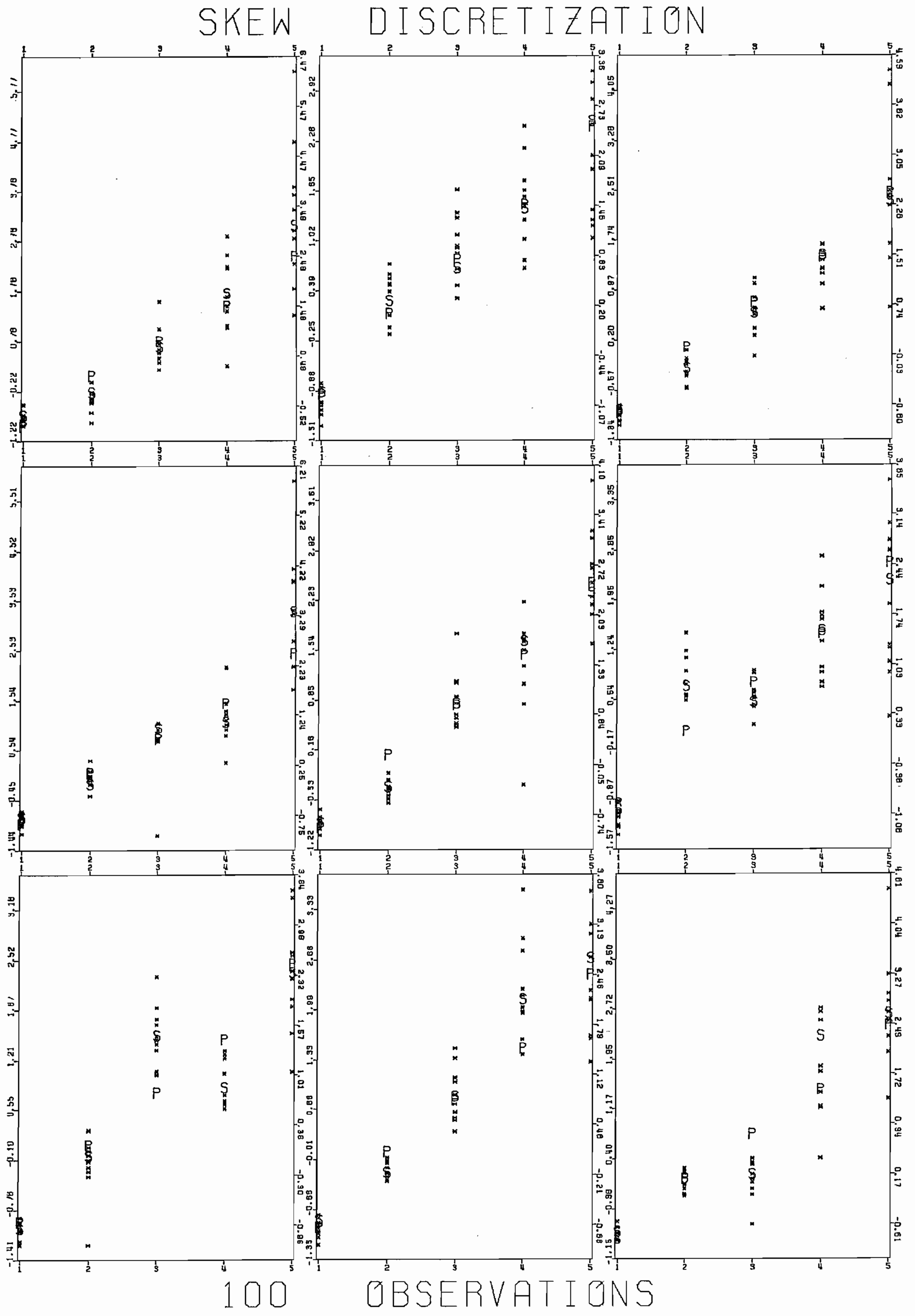

Figure 3.10

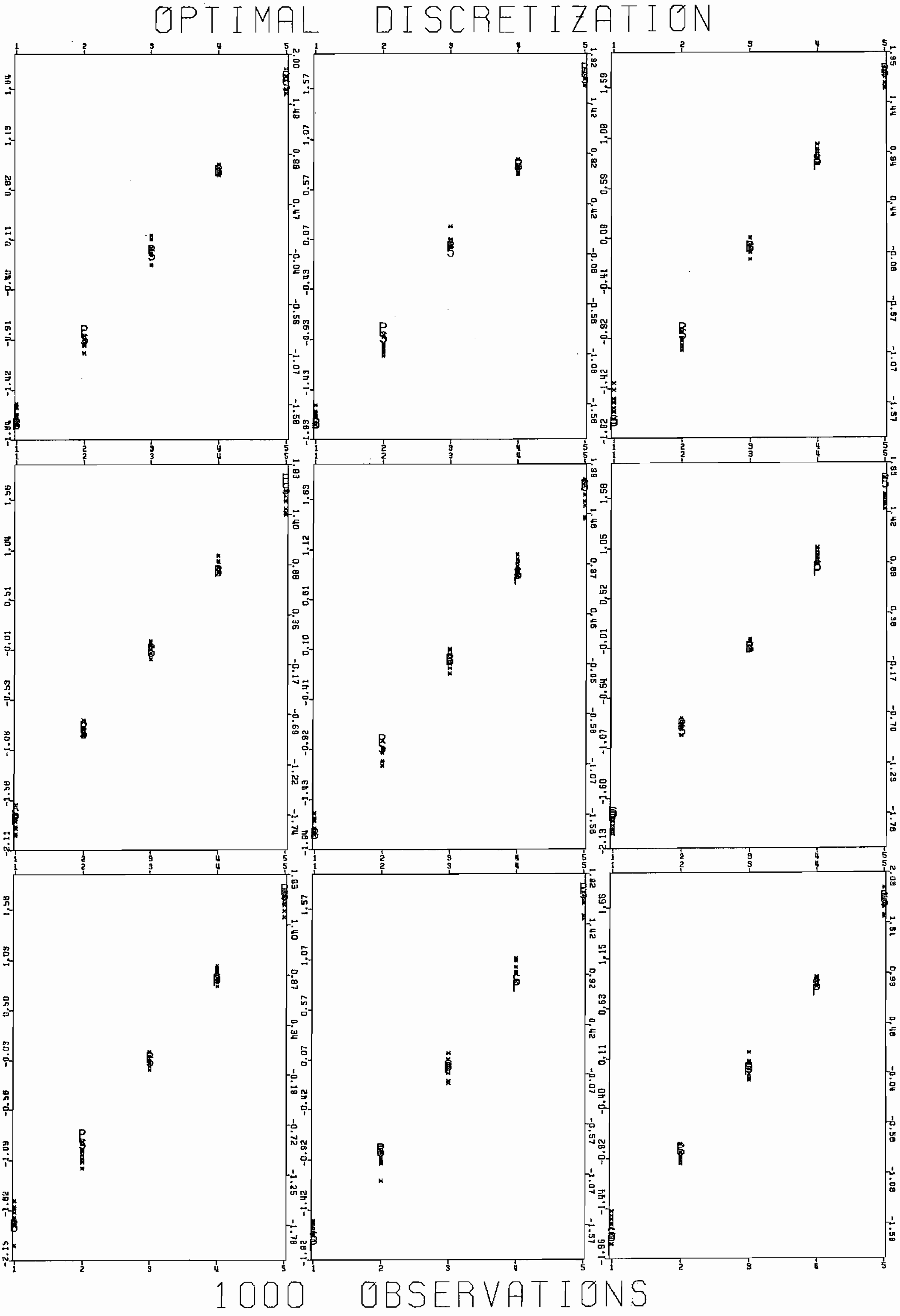
ØPTIMAL DISCRETIZATIØN
1 2 3 4 5
1.84 1.13 0.62 0.11 -0.40 -0.91 -1.42 -1.94
2.00 1.48 0.98 0.47 -0.04 -0.56 -1.07 -1.58
1.57 1.07 0.57 0.07 -0.43 -0.93 -1.43 -1.83
1.82 1.42 0.82 0.42 -0.08 -0.58 -1.08 -1.58
1.59 1.08 0.59 0.09 -0.41 -0.92 -1.42 -1.82
1.95 1.44 0.94 0.44 -0.06 -0.57 -1.07 -1.57
1.56 1.04 0.51 -0.01 -0.53 -1.06 -1.58 -2.11
1.93 1.40 0.88 0.36 -0.17 -0.69 -1.22 -1.74
1.63 1.12 0.61 0.10 -0.41 -0.92 -1.43 -1.94
1.99 1.48 0.97 0.46 -0.05 -0.56 -1.07 -1.58
1.58 1.05 0.52 -0.01 -0.54 -1.07 -1.60 -2.13
1.95 1.42 0.88 0.38 -0.17 -0.70 -1.23 -1.76
1.56 1.03 0.50 -0.03 -0.56 -1.09 -1.62 -2.15
1.93 1.40 0.87 0.34 -0.18 -0.72 -1.25 -1.78
1.57 1.07 0.57 0.07 -0.42 -0.92 -1.42 -1.92
1.82 1.42 0.92 0.42 -0.07 -0.57 -1.07 -1.57
1.66 1.15 0.63 0.11 -0.40 -0.92 -1.44 -1.96
2.03 1.51 0.99 0.48 -0.04 -0.56 -1.08 -1.59
1000 ØBSERVATIØNS

Figure 3.11

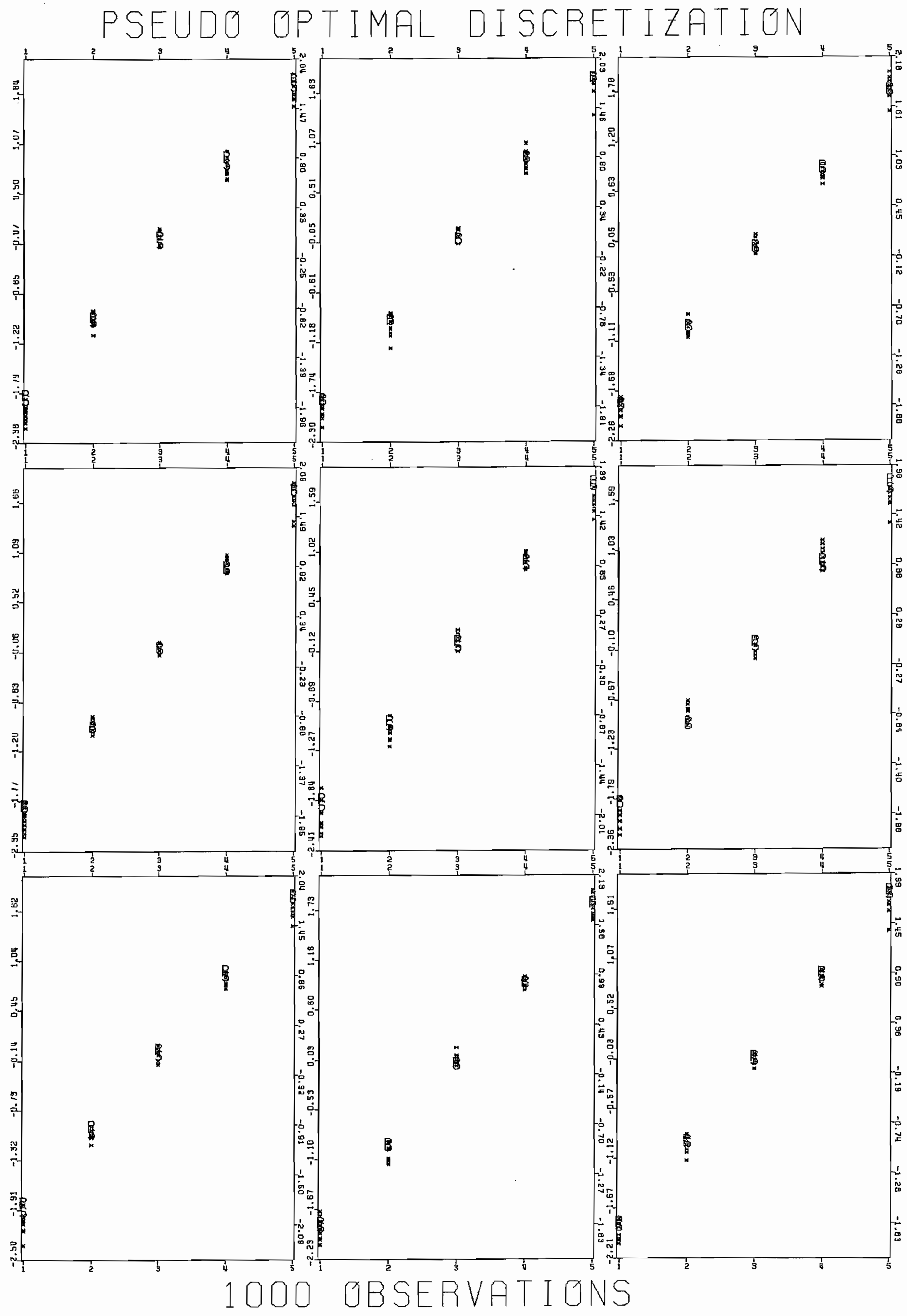

Figure 3.12

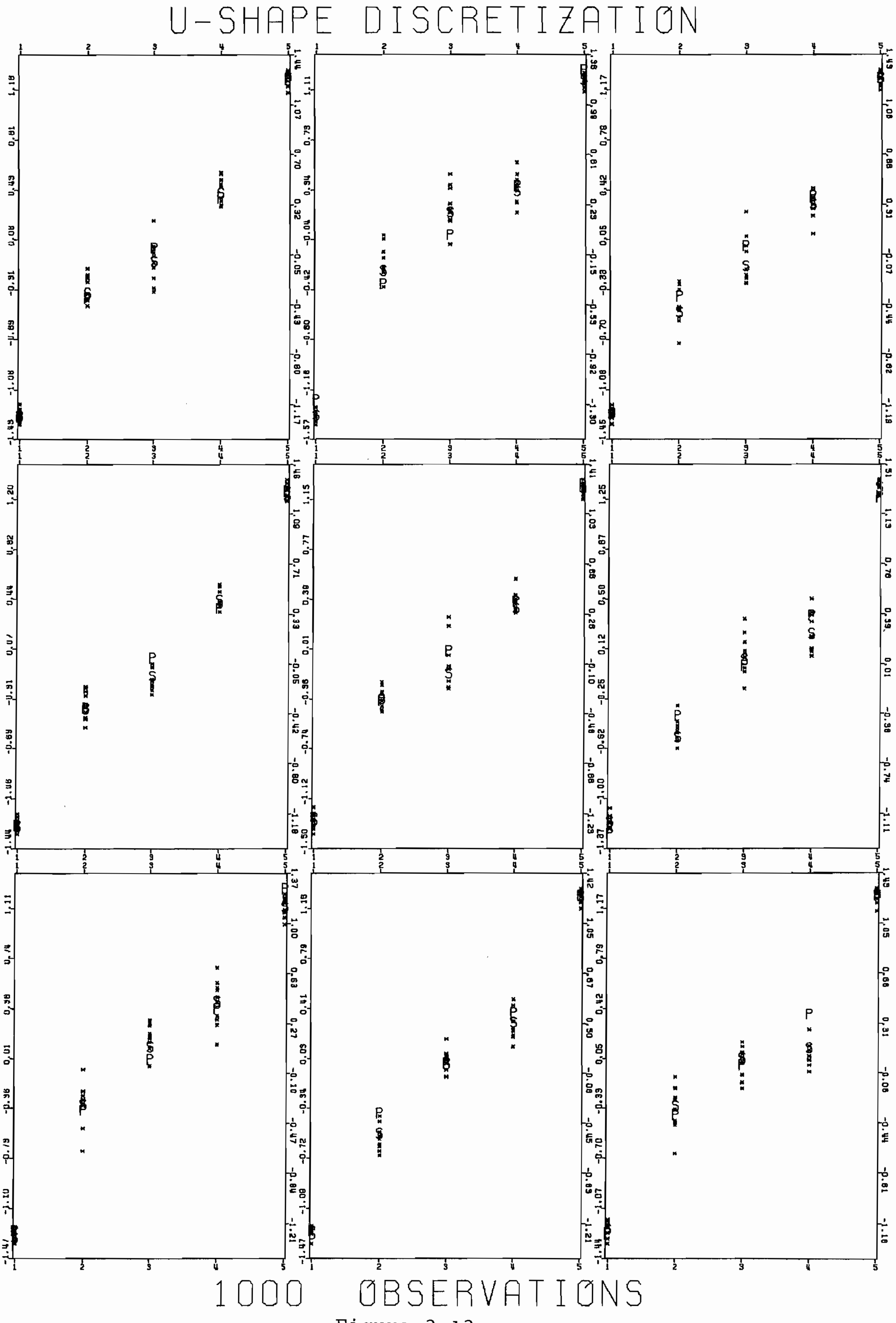
U-SHAPE DISCRETIZATIØN
1 2 3 4 5
-1.43 -1.08 -0.69 -0.31 0.08 0.43 0.81 1.18
-1.17 -0.80 -0.43 -0.05 0.32 0.70 1.07 1.44
-1.57 -1.18 -0.80 -0.42 -0.04 0.34 0.73 1.11
-1.30 -0.92 -0.59 -0.15 0.23 0.61 0.99 1.38
-1.45 -1.08 -0.70 -0.33 0.05 0.42 0.79 1.17
-1.18 -0.82 -0.44 -0.07 0.31 0.68 1.06 1.43
-1.44 -1.06 -0.69 -0.31 0.07 0.44 0.82 1.20
-1.18 -0.80 -0.42 -0.05 0.33 0.71 1.09 1.48
-1.50 -1.12 -0.74 -0.36 0.01 0.39 0.77 1.15
-1.23 -0.86 -0.48 -0.10 0.28 0.65 1.03 1.41
-1.37 -1.00 -0.62 -0.25 0.12 0.50 0.87 1.25
-1.11 -0.74 -0.38 0.01 0.39 0.76 1.13 1.51
-1.47 -1.10 -0.73 -0.36 0.01 0.38 0.74 1.11
-1.21 -0.84 -0.47 -0.10 0.27 0.63 1.00 1.37
-1.47 -1.09 -0.72 -0.34 0.03 0.41 0.79 1.16
-1.21 -0.83 -0.45 -0.08 0.30 0.67 1.05 1.42
-1.44 -1.07 -0.70 -0.33 0.05 0.42 0.79 1.17
-1.18 -0.81 -0.44 -0.06 0.31 0.68 1.05 1.43
1000 ØBSERVATIØNS

Figure 3.13

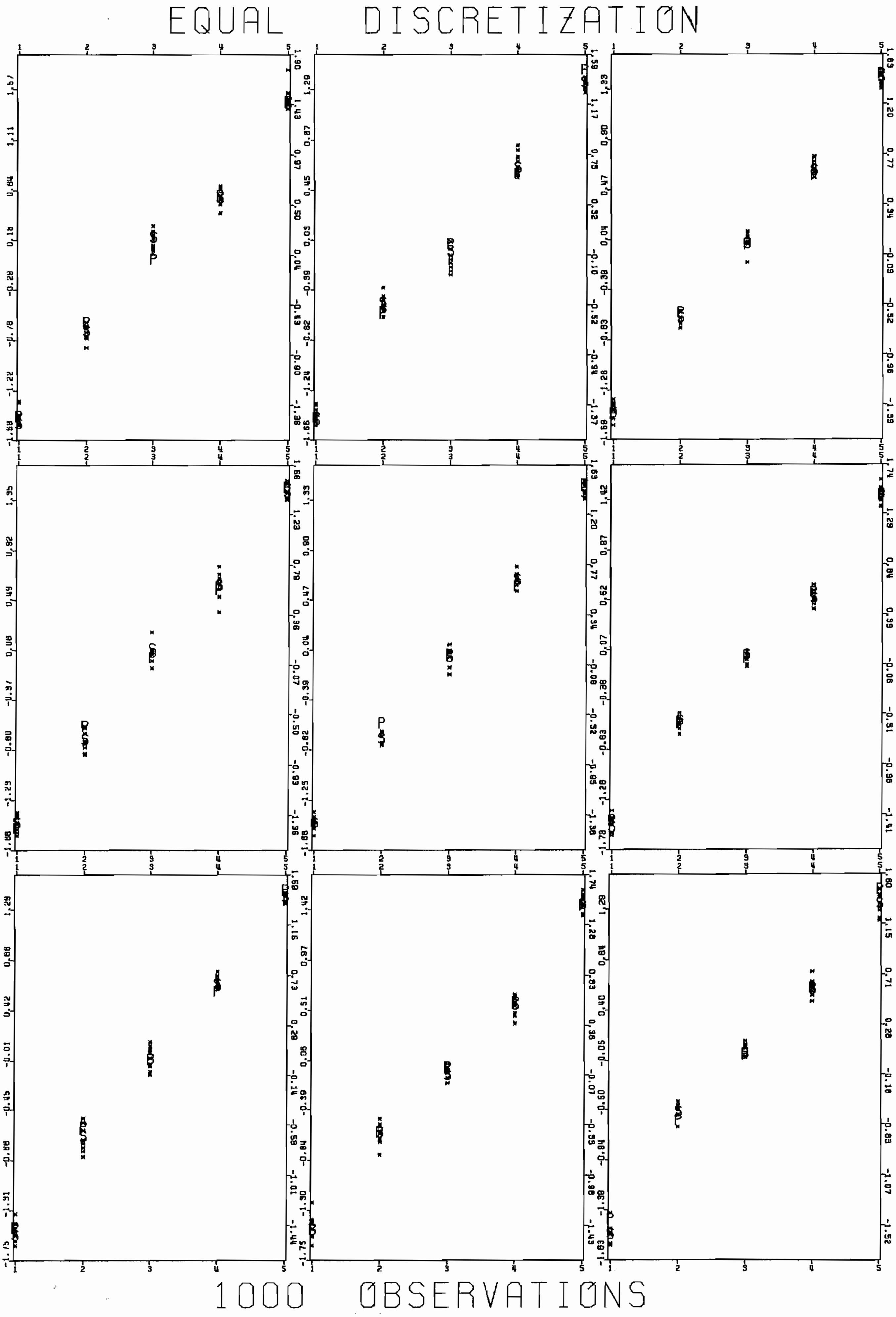

Figure 3.14

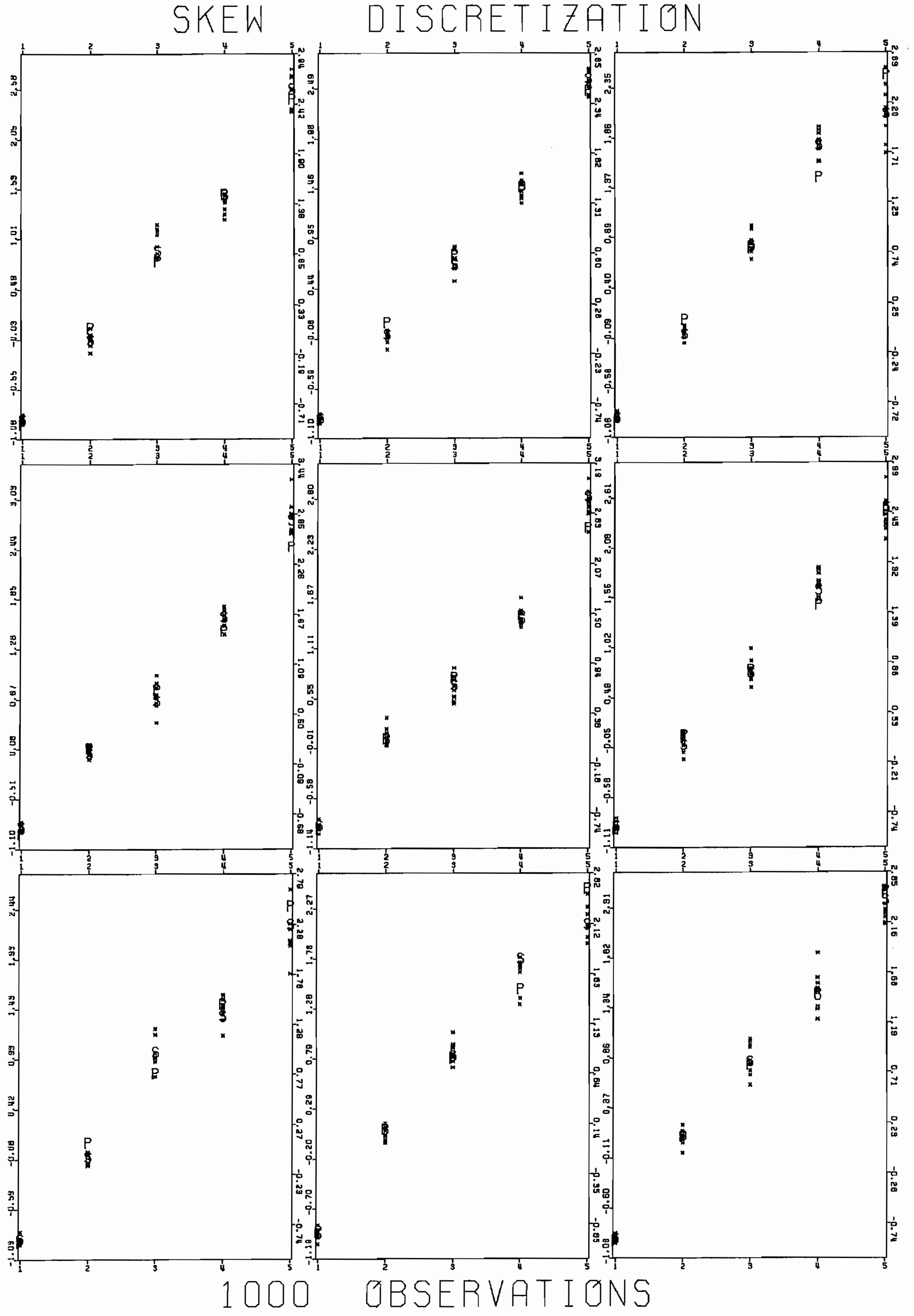

Figure 3.15

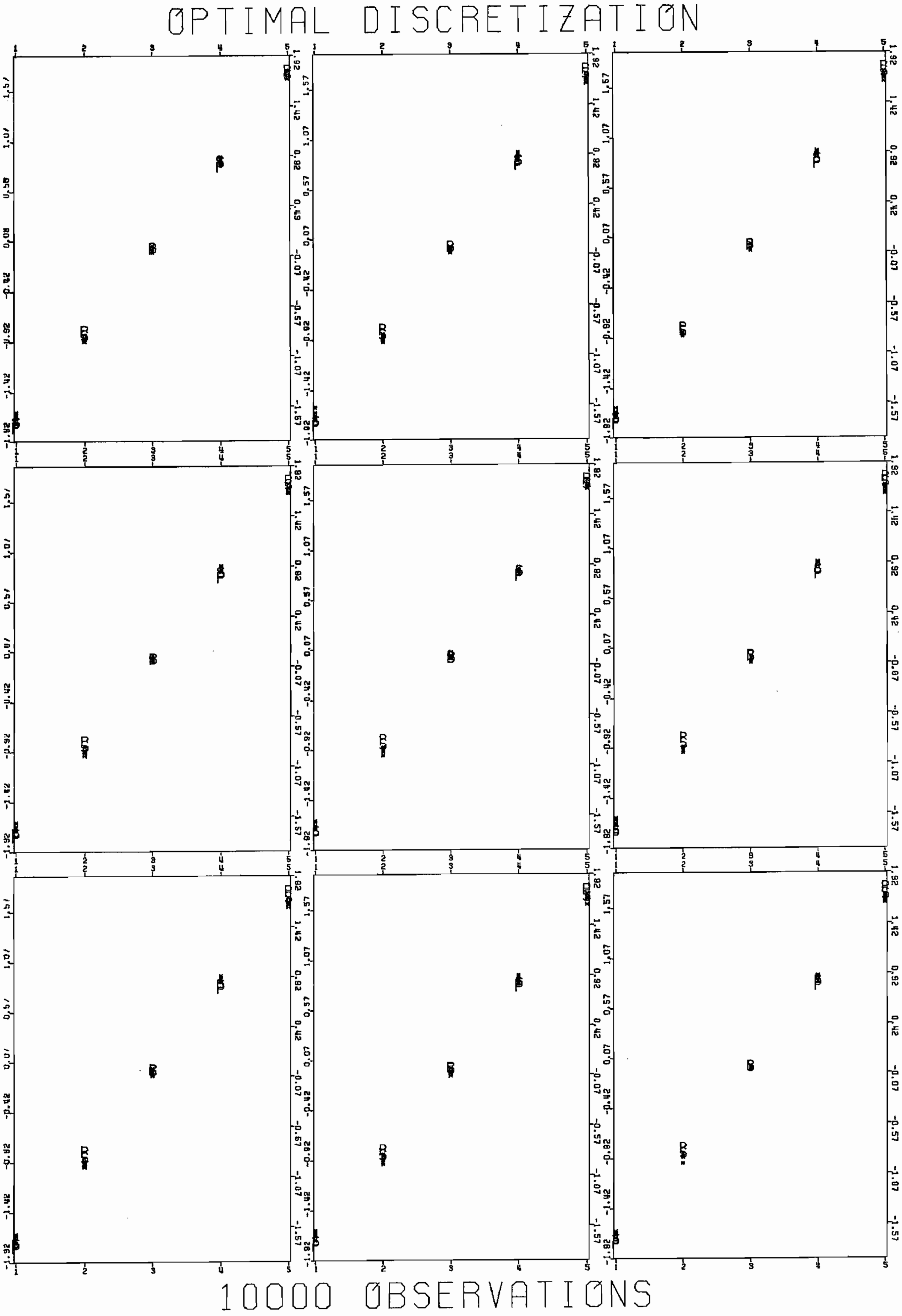
ØPTIMAL DISCRETIZATIØN
1
2
3
4
5
1.92
1.57
1.42
1.07
0.92
0.57
0.42
0.07
-0.07
-0.42
-0.57
-0.92
-1.07
-1.42
-1.57
-1.92
10000 ØBSERVATIØNS

Figure 3.16

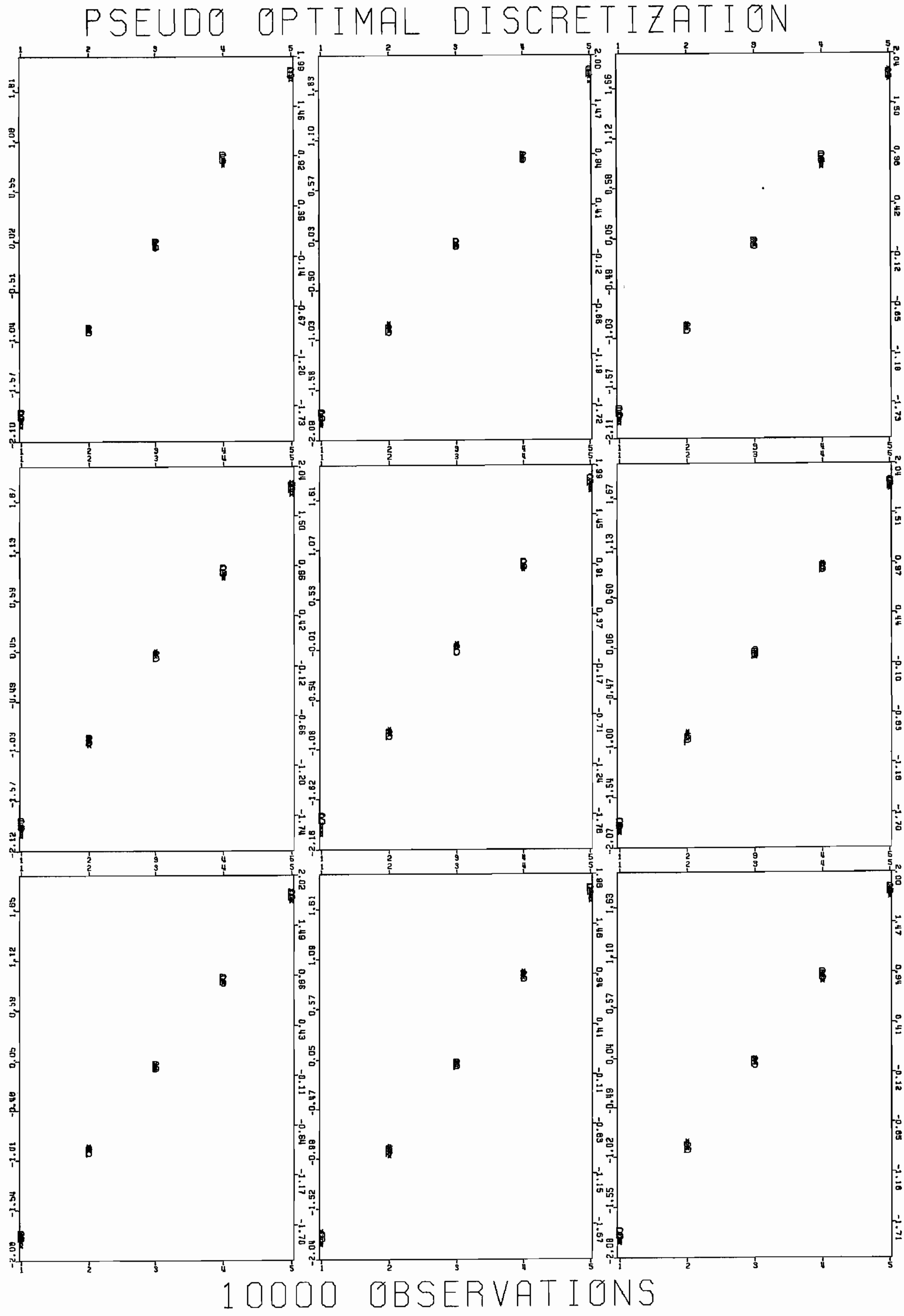
PSEUDØ ØPTIMAL DISCRETIZATIØN
1 2 3 4 5
1.61 1.08 0.55 0.02 -0.51 -1.04 -1.57 -2.10
1.99 1.46 0.92 0.38 -0.14 -0.67 -1.20 -1.73
1.83 1.10 0.57 0.03 -0.50 -1.03 -1.56 -2.09
2.00 1.47 0.94 0.41 -0.12 -0.66 -1.18 -1.72
1.66 1.12 0.58 0.05 -0.49 -1.03 -1.57 -2.11
2.04 1.50 0.96 0.42 -0.12 -0.65 -1.19 -1.73
1.67 1.13 0.59 0.05 -0.49 -1.03 -1.57 -2.12
2.04 1.50 0.96 0.42 -0.12 -0.66 -1.20 -1.74
1.61 1.07 0.53 -0.01 -0.54 -1.08 -1.62 -2.16
1.99 1.45 0.91 0.37 -0.17 -0.71 -1.24 -1.78
1.67 1.13 0.60 0.06 -0.47 -1.00 -1.54 -2.07
2.04 1.51 0.97 0.44 -0.10 -0.63 -1.16 -1.70
1.65 1.12 0.58 0.05 -0.48 -1.01 -1.54 -2.08
2.02 1.49 0.96 0.43 -0.11 -0.64 -1.17 -1.70
1.61 1.08 0.57 0.05 -0.47 -0.98 -1.52 -2.04
1.88 1.46 0.94 0.41 -0.11 -0.63 -1.15 -1.67
1.63 1.10 0.57 0.04 -0.49 -1.02 -1.55 -2.08
2.00 1.47 0.94 0.41 -0.12 -0.65 -1.18 -1.71
10000 ØBSERVATIØNS

Figure 3.17

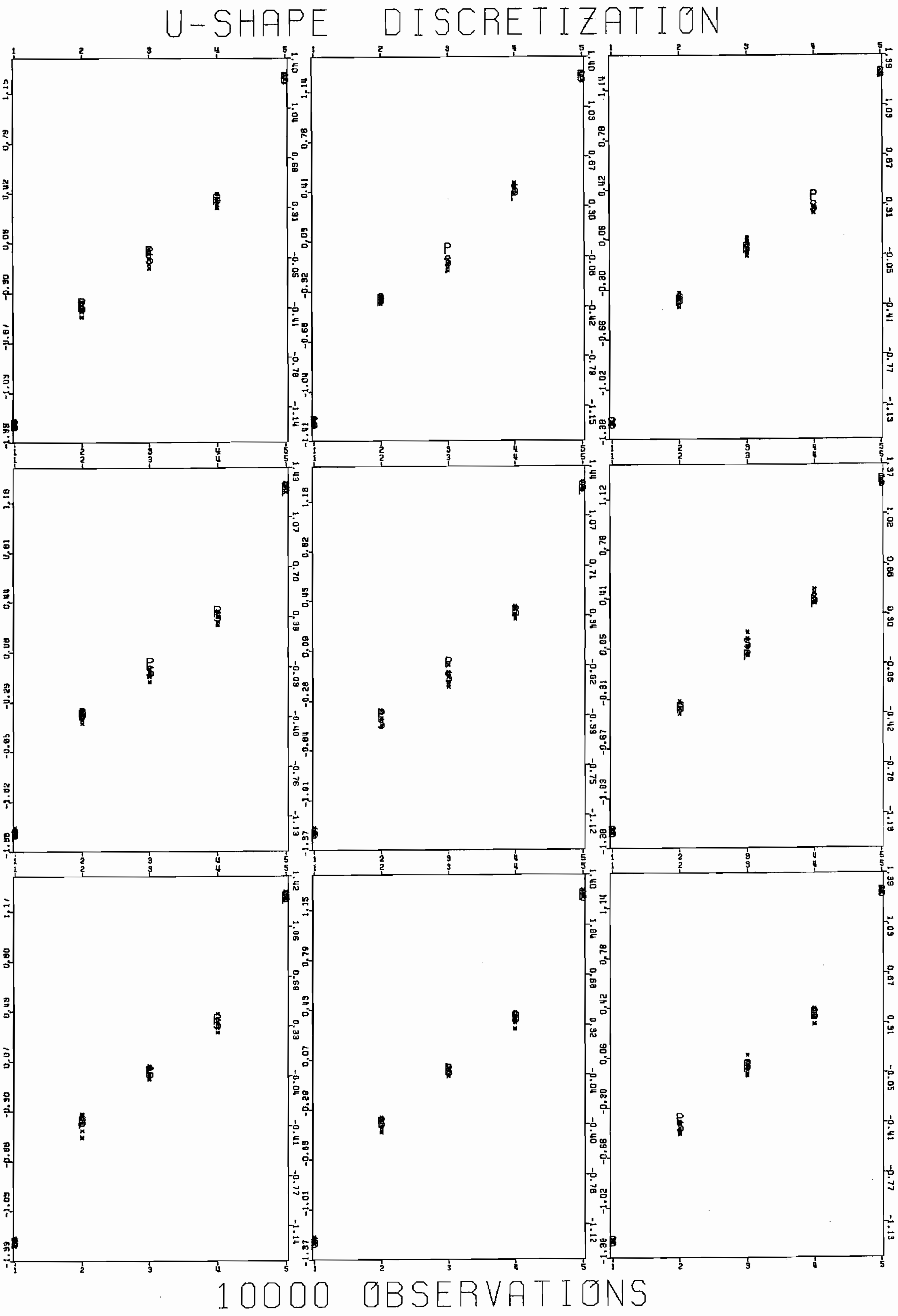

Figure 3.18

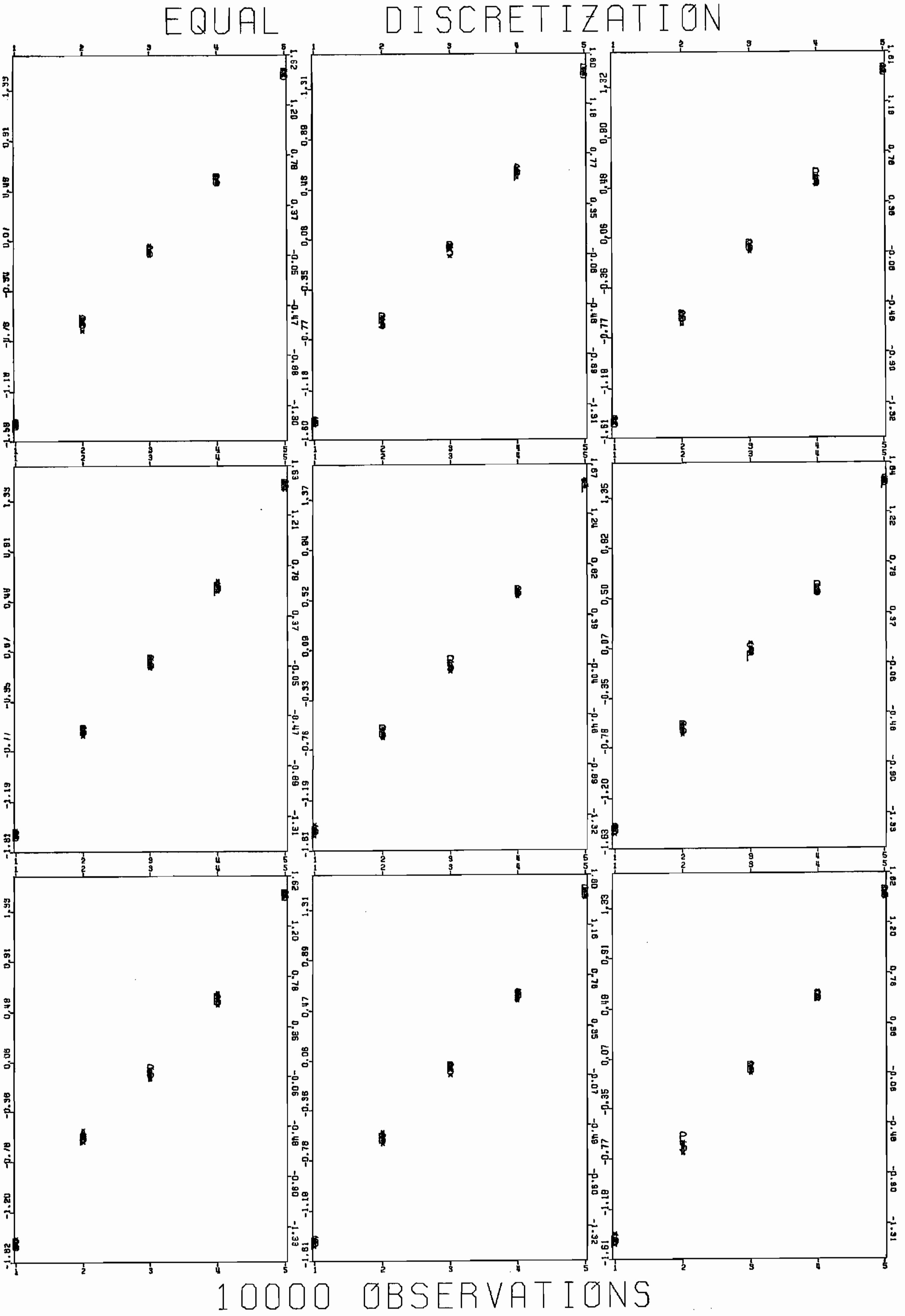

Figure 3.19

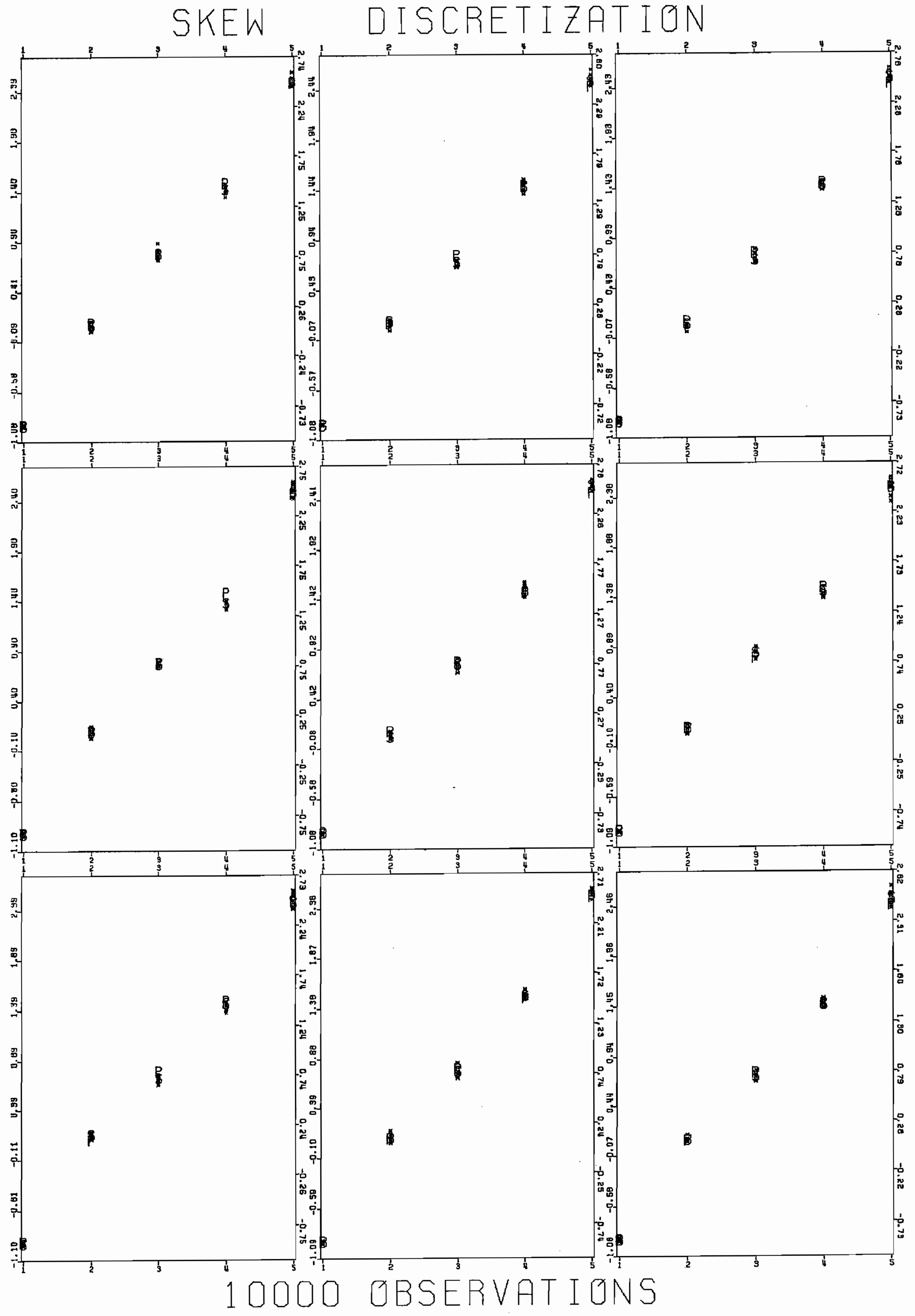

Figure 3.20

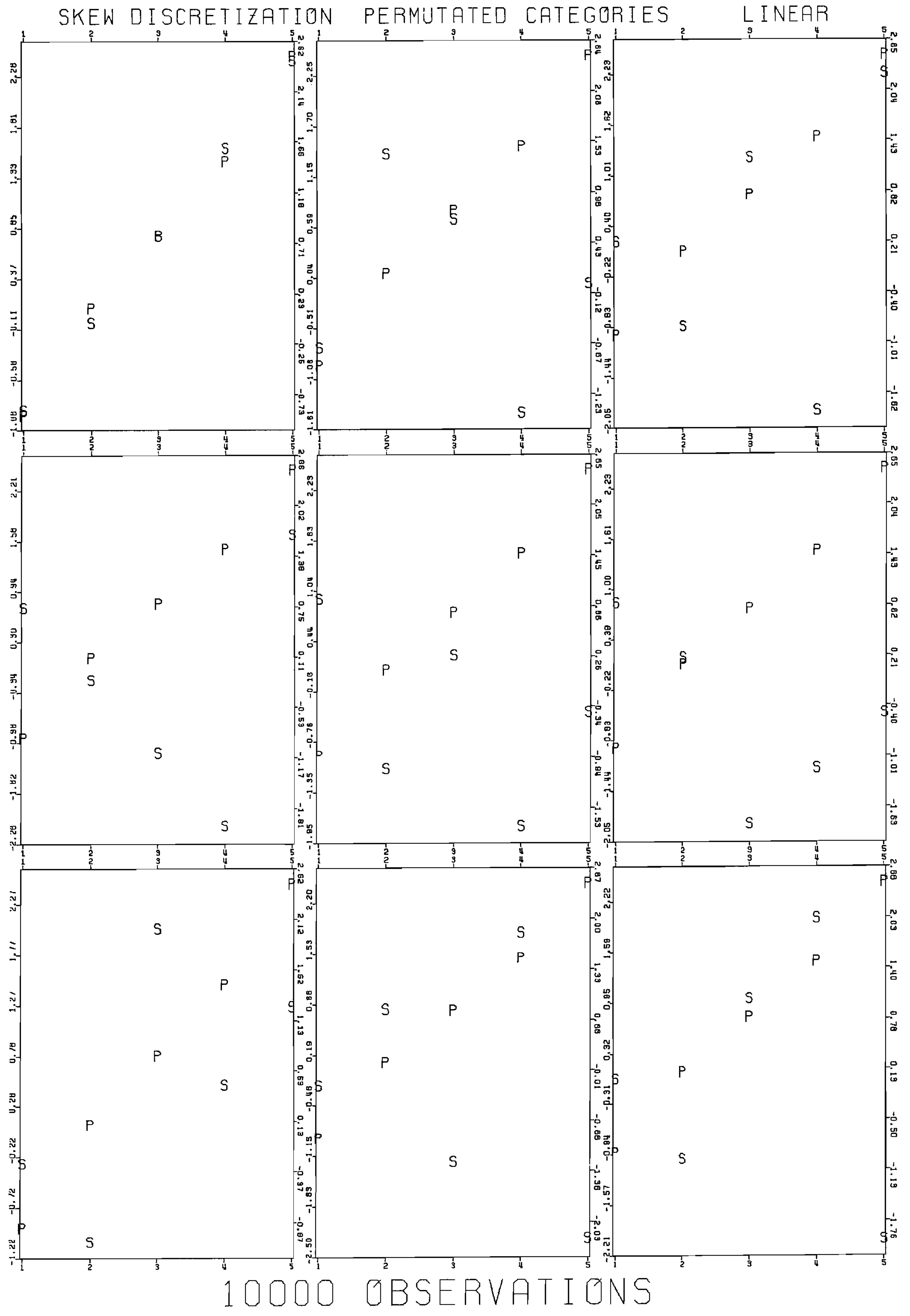
SKEW DISCRETIZATION PERMUTATED CATEGORIES LINEAR
10000 OBSERVATIONS

Figure 3.21

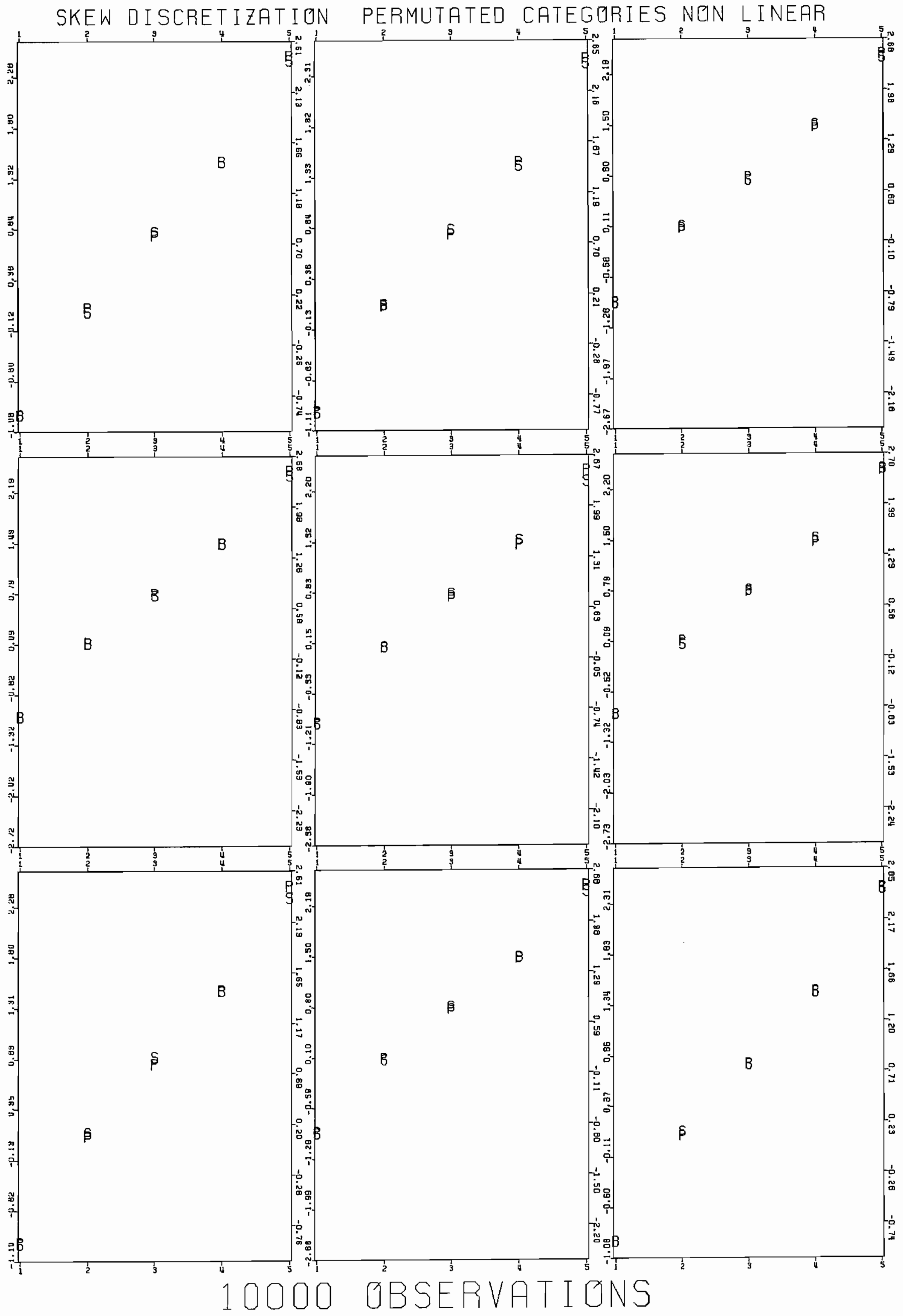
SKEW DISCRETIZATION PERMUTATED CATEGORIES NON LINEAR
1
2
3
4
5
10000 OBSERVATIONS

Figure 3.22