Working Paper

Application of Multivariate Statistical Analysis for the Detection of Structural Changes in the Series of Monitoring Data

M.Ya. Antonovski,
V.M. Buchstaber,
L.S. Veksler

WP-91-37
October 1991

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Application of Multivariate Statistical Analysis for the Detection of Structural Changes in the Series of Monitoring Data

M.Ya. Antonovski,
V.M. Buchstaber,
L.S. Veksler

WP-91-37
October 1991

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Application of Multivariate Statistical Analysis for the Detection of Structural Changes in the Series of Monitoring Data

M.Ya. Antonovski, V.M. Buchstaber, and L.S. Veksler

Abstract

A new approach to the study of time series by the projection pursuit methods is described. The ideas are illustrated on the time series of the monitoring of the environment and climate:

(a) on time series of anomalies of global mean annual temperature – the main climatological parameter;

(b) on time series of atmospheric CO_2 concentrations – the main greenhouse gases;

(c) on time series of vegetation index (NDVI) – the main global characteristic of biota activity on the satellite data.

With the aid of the shift operator for time signal, we construct a curve in n-dimensional Euclidian space (shift operator and integer n are the parameters of method). So an analysis of a time series is reduced to the analysis of the most informative projections [for example, by the criterion of factor analysis or spectral analysis (discrete Fourie analysis)] of the corresponding n-dimensional curve. We show that the comparison of such projections for model-test time series with the projection of the time series under investigation gives an effective way of finding the structural changes of the monitoring time series. For example, the case of the Hansen-Lebedeff time series of anomalies of the global mean annual temperature (see *Trends '90*), shows that the structure of the series in the interval from 1920 until 1950 essentially differs from the structure on the intervals 1880–1920 and 1950–1987. For the series of CO_2 on the Mauna Loa and Barrow monitoring stations, we obtained dynamics of the amplitudes of the year and semi-year cycles. We give the construction of a nonparametric estimation of a model of the initial time series using k-dimensional projection of n-dimensional curve. As a consequence, for example, we found the main components of the CO_2 time series and obtained the models of the yearly behaviors of NDVI time series which permit one to carry out statistically stable classification of ecosystems by ecotypes and to describe dynamics of the separate ecosystems (see Appendices).

Thus it is proposed a tool for the creation of the statistical description of the current state of the given monitoring series in the form of geometrical image. These geometrical images permit us to analyze the anomalies in the monitoring series in the terms of deviation of these images. As it follows from the examples given below such a method of the analysis of monitoring data is an effective method. Between the theoretical let us stress the following: we show how the methods of the analysis of the time series widely used in statistical treatment of monitoring data could also be used in our approach as the tools of the projections pursuit for comparing the images of the curves $\varphi(\tau, f)$ of the signal under investigation with the curves of the corresponding signals; it is shown that the proposed approach permits us to join in a united method the achievement of the theory of operators of a generalized shift and exploratory analysis on the basis of the projection pursuit.

Introduction

The problem of the assessment and prediction of antropogenic impact on climate (and vice versa – influence of climatic changes on society) is one of the most important among the global problems. Taking decisions connected with this problem lean mainly on monitoring data. For example, the arguments for the impact of CO_2 emission from fossil fuel burning on changes of the radiative balance of the Earth (greenhouse effect) was published by Callendar in 1938 on the basis of the analysis of the trend of the series of mean annual temperature. But only a comparison of monitoring data of the trends in atmospheric CO_2 concentration time series with the trends of the series of antropogenic emission of the CO_2 in the atmosphere put a problem of the assessment and prediction of greenhouse effect on the line of the most known and actual problem. The efficacy of using monitoring data is defined mainly by the quality of the analysis of the given data. In this connection, the development of the methods of applied statistics that are oriented on the peculiarity of the monitoring time series and are directed on the solution of following questions of great importance:

1. A construction of a statistical model of a time series (basic time series);
2. Informative description of the dynamics of the anomalies (deviations from the basic time series);
3. Finding the structural changes of the time series

   (reconstructions of the basic time series);
4. Assessment of the uncertainty of the predictions of the time series (extrapolation of the basic time series with taking into account the dynamic of the anomalies and mathematical modeling of physical mechanisms of researched phenomenon.

In this paper we describe the approach to the solution of these questions on the basis of jointly using the results of the theory of generalized shift and of projection pursuit.

In §4 and 5 we give a more detailed description of the behaviors of NDVI time series.

§1. Description of method.

Let us start from the general construction for time signals as a function of continuous time. Let the signal $f(t)$ under investigation belongs to some linear space F of time signals. Let us fix a linear operator $A : F \to F$ and a set of marks (counts) $t_1, \ldots, t_n$. Let us construct for each $f(t) \in F$ a curve $\varphi_f(\tau)$, τ is an internal time on curve and f in the index shows that

φ_f is built using curve $f(t)$ [we will also use a notation $\varphi(\tau; f)$]. We construct $\varphi_f(\tau)$ as a piece-wise linear curve in R^n connected consequently with knots $x_1, \ldots, x_m$, where $x_i \in R^n$, and $x_1 = (f(t_1), \ldots, f(t_n))$, $x_2 = (Af(t_1), \ldots, Af(t_n))$, ..., $x_m = (A^{m-1}(f(t_1)), \ldots, A^{m-1}(f(t_n)))$ and $A^m = A(A^{m-1})$ – m-th iteration of operation A. Operator A and set of marks (counts) $\{t_1, \ldots, t_n\}$ are the parameters of the method selected from the following gesture.

Let $f(t)$ be a eigen function of operator A, i.e., $Af(t) = \lambda f(t)$, $\lambda = \lambda(f)$, λ independent from time t (dependent only from function f). Then by the construction $x_m = \lambda^{m-1} x_1$ for each set of marks. Thus, for the eigen functions of operator A the curve $\varphi(\tau, f)$ is disposed in one-dimensional subspace of the space R^n with the guiding vector in the form $x_f = \frac{x_1}{\|x_1\|}$. Moreover, if $|\lambda| < 1$, then a curve $\varphi(\tau; f)$ is on the interval $[-x_1, x_1]$. If $\lambda > 1$, then $\varphi(\tau; f) \subset R^1 \backslash [-x_1, x_1]$.

Now, let

$$f(t) = \sum_{q=1}^{k} c_q f_q(t) \ , \quad k < n \ \text{(hypothesis)} \ ,$$

where $f_q(t)$ is the eigen function of operator A with the eigen values λ_q correspondingly, ($Af_q = \lambda_q f_q$, $q = 1, \ldots, k < n$). Then a curve $\varphi(f; \tau)$ lies in k-dimensional subspace of the space R^n.

The operator A of general shift is also a parameter of the method. For an operator A of a given concrete form, it is possible to obtain a stronger assertion of the geometry of a curve $\varphi(t, f)$. Let us consider as an operator A, for example, the operator of classical shift A_1: $f(t) \to A_1 f(t) = f(t + \Delta t)$, where Δt is a step of a shift and also the parameter of the method. This operator confront to the function $f(t)$ a function $f(t + \Delta t)$. Then, m-th iteration of $f(t)$ is $f(t + m\Delta t)$: $A_1^m f(t) = A_1^{m-1} f(t + \Delta t) = f(t + m\Delta t)$. Hence, if $f(t)$ is a periodical function with the period $T = m\Delta t$ then the corresponding curve $\varphi(\tau; f)$ is a closed broken line at R^n with m-knots $(x_1, \ldots, x_m)$. So, as the eigen functions of operators A_1 are exponents, $\exp(pt)$, where p, generally speaking, is a complex number, then if $f(t)$ is a polyharmonical signal, $f(t) = \sum_{q=1}^{k} c_q \sin(w_q t + \varphi_q)$, $2k < n$ (hypothesis), then for any set of marks $\{t_1 \ldots, t_n\}$ the curve $\varphi(\tau; f)$ will be in subspace of the space R^n stretched on vectors $x_{1,q} = (\sin w_q t_l)$, $x_{2,q} = (\cos w_q t_l)$, $q = 1, \ldots, k; l = 1, \ldots, n$. A dimension of the subspace is equal to the range of the $(2k \times n)$-matrix – compounded from 2k vector column x_{1q}, x_{2q}, $q = 1, \ldots, k$. A range of matrix depends on selection of a set $\{t_1, \ldots, t_n\}$. It is easier to illustrate this remark on another example, that will play itself an important role. Let $f(t)$ be a polynomial of a degree s, $p(t) = \sum_{l=0}^{s} a_l t^l$, $s < n$. Then for any m

$$A_1^m p(t) = \sum_{l=0}^{s} a_l (t + m\Delta t)^l$$

is also a polynomial of a degree s. It is easy to see that in this case the curve $\varphi(\tau; p)$ is at the

subspace of the space R^n stretched on the vectors

$$y_l = p_l(t) \ , \quad l = 0, \ldots, s \ ,$$

where $p_0(t) = 1, p_1(t) = t, \ldots, p_s(t) = t^s$ is a basis in the space of the polynomial of the degree not exceeded by s. We have

$$y_l = p_l(t) = (t_1^l, \ldots, t_n^l) \ .$$

Dimension of the subspace in this case is equal to the range of $((s+1) \times n)$-matrix – compounded from $(s+1)$ of these vector columns y_l.

Let $J = (j_1 < j_2 < \cdots < j_{s+1})$ be a subset of $s+1$ elements of the set $\{1, 2, \ldots, n\}$. Then the minor Δ_J of the matrix corresponding to our set J as Vanderemond determinant is

$$\Delta_J = \prod_{j_\alpha > j_\beta} \left(t_{j_\alpha} - t_{j_\beta}\right) \ .$$

Hence the marks should stay one from another as far as possible. But if there are a lot of marks, then the product of the numbers each of which is less than 1 would be very close to zero. So the dimension of the subspace is $s+1$ for any set of n marks where $n > s$. At the same time, the stability of the results connected with the decomposition of the polynomials by a basis of y_l is defined by how far from zero the value $\max_J \Delta_J$ is [here J is running by the set $\{J\}$ of all subset $(j_1 < j_2 < \cdots < j_{s+1})$ of the set $\{1, 2, \ldots, n\}$].

Coming back to the general case, it is possible to formulate the following demands to the parameters of the methods:

1. A choice of the operator A is determined by the hypothesis of the time signal generator $f(t)$ under investigation. Namely, starting from the hypothesis, we choose a finite-dimensional subspace F_M (of the space F) of model signals, and an operator A is choosing under the condition $A : F_M \to F_M$. Above we have considered two different examples of subspace F_M: a subspace of polyharmonical signals for a given set of frequencies $w_1, \ldots, w_k$ and a subspace of the polynomials of the degree not exceeding s. In both examples, the operator of the classical shift A_1 maps the subspace $F_M \to F_M$.

2. The set of marks $\{t_1, \ldots, t_n\}$ was chosen by the way that $n > \dim F_M$ and restriction of the linear map $I : F \to R^n$, $f(t) \to (f(t_1), \ldots, f(t_n))$ onto subspace $F_M \subset F$ is an imbedding. It is clear that this condition defines the set of marks in a non-unique way. It is important to select such a set of marks, that if to identify F with the linear subspace $I(F)$ in R^n then the determinant of the map $I^T I : F \to F$ would be essentially separated

from zero. Here $I^T : R^n \to F$ is a map conjugated to I relatively Euclidian scalar product in R^n.

Furthermore, we will continue to suggest below, that in space F is picked out the subspace of model signals F_M. And operator A and a set of marks are satisfied to the conditions 1, and 2. For our method it is important to have the effective algorithms of comparison of the geometry of the curves $\varphi(\tau; f_M)$ in F_M for model signals with the geometry of the curves $\varphi(\tau, f)$ for the signals under investigation. In §3 will be described the results of the application of the algorithms of such a comparison to the time series of monitoring based on **factor analysis**. The selection of the form of the analysis is explained by the following:

1. The factor analysis of the set of n-dimensional vectors $x_1, \ldots, x_m, \ldots$ permits us to calculate the dimension of the subspace in which lies a curve $\varphi(\tau, f)$. Taking into account the fact that we know such a dimension for a model signal *a priori* (more precisely, we know an estimation from above of this dimension), and comparing with a dimension that we have obtained from the factor analysis we come to the first effective algorithm of the comparison.

2. The factor analysis permits us to obtain k-dimensional orthogonal projections of the examined geometrical image on k-dimensional subspaces of main factors which are characterized by the property that they explain, by the best possible way, the dispersion in the set of vectors $\{x_1, \ldots, x_m, \ldots\}$. Studying one- and two-dimensional projections of the curves $\varphi(\tau; f)$, we obtain the possibility to produce the tools for a description of the **types** of deviation of the projections of the curves under examination from the projections of model curves. Thus, we obtain algorithms based on the human possibility to differ effectively the visual images. The using of this ability of the human being lies in the foundation of the exploratory data analysis by the method of pursuit projection – a new perspective direction of multi-dimensional statistics. The results in this direction essentially based on the achievements of modern computer graphics.

In §4 we show how the methods of the analysis of the time series widely used in statistical treatment of monitoring data could be also used in our approach as tools of the projections pursuit for comparing the images of the curves $\varphi(\tau, f)$ of the signal under investigation with the curves for the corresponding model signals.

In §5, it will be shown that the proposed approach permits us to join, in a united method, the achievement of the theory of operators of a generalized shift and exploratory analysis on the basis of the projection pursuit.

§2. Description of algorithm.

The following algorithm that realizes the approach, given above, is used for the analysis of the time series: anomalies of mean annual temperature (see *Trends '90*), mean monthly concentration of the CO_2 in the atmosphere (see *Trends '90*) and NDVI (see Appendix 1), (see *Figure 1*, where also given the examples of model series 1a and series 1b that was obtained from the generator of random numbers and also three examples of real monitoring data series the research of which this paper is devoted.) Let $f = (f_1, \ldots, f_N)$ be a given time series, where $f_k = f(t_k)$, $k = 1, 2, \ldots, N$. Let us choose as an operator A the operator A_1 of the classical shift on **one** mark and as a set of marks $\{t_1, \ldots, t_n\}$ – the set of the first n marks. Then according to the general scheme of the method, we obtain the set $x_1, \ldots, x_m$ of n-dimensional vectors $x_1 = (f_1, \ldots, f_n)$, $x_2 = (f_2, f_3, \ldots, f_{n+1}), \ldots, x_{N-n+1} = (f_{N-n+1}, \ldots, f_N)$.

Or, in the terminology of applied statistics, we obtained a $(n \times (N-n+1))$-matrix of data $X = (x_{kl}, l \le k \le n, 1 \le l \le N-n+1)$, m-th row of which is a vector $x_m, m = 1, \ldots, N-n+1$. The selection of an operator A is corresponding to the selection of the model signals of the form

$$f_M(t) = f(t) = \sum_{q=1}^{k} p_q(t) \sin(w_q t + \varphi_q) \ ,$$

where $p_q(t) = \sum_{l=1}^{s_q} a_{ql} t^l$, $q = 1, \ldots, k$ are the polynomials of degree s_q.

The selection of the parameter n for seasonal time series is defined by the number of marks at season. For example, for atmospheric CO_2 concentration and vegetation index NDVI the natural **season** is a **year of observation** and hence $n = 12$, for the **mean monthly** concentration of CO_2 and the values of NDVI, each of which was obtained by the standard procedure of maximization by the set of observation data at the corresponding **month** of observations.

For the time series without a natural seasonal structure, the number n *a priori* is not fixed and is evaluated in the way of analysis. For example, for a series of the mean annual temperature anomalies, as described below, the analysis for $n = 21$, permits us to find the structural reconstruction of the series.

On the first stage of the algorithm, we conduct a complete factor analysis of the matrix X of the data, formed by the time series $f(t) = (f_1, \ldots, f_n)$. The result of the stage is a set of proper numbers $\lambda_1, \ldots, \lambda_n$ and the set of n-dimensional vectors $w_1, \ldots, w_{n_*}$, where $n_* \le n$, n_* is the number of the last non-zero proper numbers, i.e., $\lambda_q = 0, q > n_*$. Hence, the set of all non-zero n-dimensional vector $x_1, \ldots, x_{N-m+1}$ lies in n_*-dimensional subspace, stretched on the vectors $w_1, \ldots, w_{n_*}$.

The second stage of the algorithm consists of three steps: (1) Analysis of the projections of the set of the vectors $x_1, \ldots, x_{N-m+1}$ on the axes of the first two main factors. (2) Analysis of

the projections of the piece-wise linear curve $\varphi(\tau, f)$ with knots $x_1, \ldots, x_{N-m+1}$ on the plane of the pairs of the first main factors. (3) A construction of a non-parametric assessments of the initial series $f = (f_1, \ldots, f_n)$ by the projection of data matrix X on q-dimensional subspace of the first q main factors $q < n_*$.

Comments. The analysis of the set of vectors $x_1, \ldots, x_N$ essentially use that a data matrix $X = (x_{k,l}; 1 \le k \le n, 1 \le l \le N - n + 1)$ has highly special form, namely

$$x_{k,l} = f_{k+l-1} \ ,$$

i.e., x is a Hankel matrix (see Gantmacher, 1967). We will show that these projections have a natural interpretation in terms of the initial time series.

Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a fixed set of weights. The procedure of moving weighting (α-weighting) of the series $f = (f_1, f_2, \ldots, f_N)$ is an obtaining of a new series $g = (g_1, g_2, \ldots, g_{N-n+1})$ where

$$g_l = \sum_{k=1}^{n} \alpha_k f_{k+l-1} \ .$$

If, for example, $\alpha_1 = \alpha_2 = \ldots = 1/n$, then series g_l is a series of the moving average of the initial series. The procedure of moving weighting is often used for a smoothing of a given time series, picking up its trend and supressing the noise part. Using the Hankel matrix $X = (x_{k,l} = f_{k+l-1})$, it is possible to present series $g = (g_1, \ldots, g_{N-n+1})$ in the form

$$g = X\alpha \ .$$

It means that the set of the numbers

$$\frac{g}{\|\alpha\|} = \left(\frac{g_1}{\|\alpha\|}, \ldots, \frac{g_{N-n+1}}{\|\alpha\|}\right) \ ,$$

where

$$\|\alpha\|^2 = \sum_{k=1}^{n} \alpha_k^2$$

presents a projection of a set of vectors $x_1, \ldots, x_{N-n+1}$ on one-dimensional subspace, generated by the unit vector $\alpha/\|\alpha\|$.

The following lemma describes a criterion by which a projection of the set of vectors $x_1, \ldots,$ x_{N-n+1} on the axis of the first main factor of the data matrix X corresponds to the **best** procedure of α-weighting. Let us consider the functional $SS_T(f)$ putting in corresponding to time series $f = (f_1, \ldots, f_N)$ the dispersion of its value:

$$SS_T(f) = \sum_{l=1}^{N} (f_l - f)^2 \ , \quad \text{where } \overline{f} = \frac{1}{N} \sum_{l=1}^{N} f_l \ .$$

We use notation SS_T from the regression analysis (Afifi, Azen, 1979) in which the values of this functional are served as a scale under estimation of the qualitative property of the regression of time series $f(t)$.

Lemma. Let $g = (g_1, \ldots, g_{N-n+1})$ be the result of α-weighting of time series $f = (f_1, \ldots, f_N)$, then

$$SS_T(g) \leq \lambda_1 \|\alpha\|^2 \ ,$$

where λ_1 is the great eigen number of the matrix $X = (x_{lk} = f_{k+l-1})$. The equality is reached when $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a set of coordinates of the first main factor w_1. Thus, between all α-weighting of the initial series $f = (f_1, \ldots, f_q)$ the series obtained by the procedure of weighting with the aid of the first main factor w_1 have the biggest relative dispersion of the values.

The integral part of the method is a description of the projection of piece-wise curves $\varphi_M(\tau, f_M)$ for model signals – a creation of a bank of model images. Analysis of the projection of the curves $\varphi(\tau, f)$ for real signal is making in the terms of deviations (of anomalies) of its projections from the model images.

Further, under the demonstration of the method we shall use a description of the projections of the curves $\varphi_M(\tau, f_M)$ for the following model signal f_M:

1. polyharmonical signal;
2. polynomial signal;
3. noise signal generated by the different random numbers generators;
4. the combinations of signals of the first three classes.

For the model signals we have considered the series $f = (f_1, \ldots, f_N)$, where $N = 84$, $n = 12$. A selection of the values N and n links with a creation of data bank for model images for analysis of 7-year time series of NDVI (Appendix 1). After applying the shift operator for series f were formed (73×12)-matrix of data $X = X(f)$, and then complete factor analysis of matrix X was done according to the first stage of our algorithm.

Let us consider the results of the second stage of the algorithm on models signals. In *Figure 2* we present a harmonic $a_0 + a_1 \sin(2\pi k/12)$ and its projection in the space of principal component. The initial signals are shown in *Figure 2a*. The projection on the first pair (*Figure*

2b) is a circumference that corresponds to analytical result. A projection on the second part (*Figure 2c*) demonstrates some regular figure that does not correspond to analytical image of harmonic. The following analysis shows that the appearance of this figure links with the error of approximation at the calculation of the values of sin.

Increasing the preciseness of the calculation takes away this figure (see *Figure 2d*) and confirms the hypothesis that this figure connected with the effect reflected the rules of the approximation to (rounded off). Namely, in the first case a rounding off was done with the preciseness of the third valued number (and discovered secondary effects bear witness to the high sensitivity of the method). In *Figure 3* we present the results of polynomial series $a_0+a_1k+a_2k^2$. In *Figure 3a* we see the initial signal; in *Figure 3b* the projection on the principal component; on *Figure 3c* the projection on the second principal component; and on *Figure 3d* the projection on the plane of the first pair of the principal factors. As the initial series is clearly not a periodical one, then the curve (*Figure 3d*) is not closed. The assessment of the dimension of initial set of points completely corresponds to the analytical estimation. As we mentioned above, if we have a polynomial signal of degree k, then the dimension of the space where the points are plunged is $k+1$. For parabola, the dimension of the space of plungeness is equal to 3.

In *Figure 4* the results for signals of random generator numbers are given. In this figure we see that under the projections in factor space there do not appear any regular images. We analyzed the time series which were obtained by the generator of random numbers of different laws – uniform law, normal law with different initial parameters, namely with different dispersions and means. We obtained the images as chaotic clusters, so we could say that the projections of this series are the image of chaos.

Figure 5 shows harmonic (as in *Figure 2*) perturbed by noise, generated by normal law with dispersion $\sigma = 0.5$; *Figure 6* demonstrates the analogous situation with $\sigma = 1.0$. In the second case, *Figure 6*, the dispersion of a noise is equal to the amplitude of the initial harmonic. As a result visually the join signal completely loses its sinusoidal shape. Nevertheless, its projection on the plane of the first principal components *Figure 6* still remains an image that corresponds to the existence of a cycle. It confirms efficacy of the method for reconstruction of the initial structure that was exposed by random distortion. In *Figure 7* we can see the corresponding results for polyharmonical signal with linear trend. The projections on the principal components and the planes of the pairs of the main components show evident similarity with corresponding projections for structural compounds of the signals (see for comparison *Figures 2* and *3*). It permits us even in the case of a huge distortion to restore the structure of the initial signal. From *Figure 7* it is seen that polyharmonical and polynomial compounds of the signal are reflected in the principal components. Their order depends on the relative contribution of each component.

Let us remark that when we have a polyharmonical signal f, then the components of $\varphi(\tau, f)$ corresponding to the different harmonics are placed in orthogonal planes. If we add polynomial trend, then it is placed in some subspace not ortogonal to the subspace of polyharmonical (correctly, under an angle). This situation is shown in *Figure 8* where a polyharmonical signal is presented which is put on parabolic trend. In *Figure 8* pertubation of model images are clearly seen. But when we project into the space of the factors, we should know that, for example, the first projection corresponds to polynomial, or to harmonic, and so on. This is just an example of the explanation, how we can change the input of the different components. But as the subspaces corresponding to each component are not exactly orthogonal, then the projections are slightly distorted. Thus the study of model analytical curves permits us to estimate the stability and sensibility of the method in the frame of the models under consideration and to obtain model images.

The algorithm of the construction of non-parametric assessments of initial series $f = (f_1, \ldots, f_N)$ by the projections of data matrix X consists of the following.

Each vector $x_m = (f_m, \ldots, f_{m+n-1})$, $m = 1, \ldots, N - n + 1$ could be written in the form

$$x_m = \sum_{q=1}^{n_*} x_{mq} w_q \ ,$$

where $\{w_q \in R^n, q = 1, \ldots, n_*\}$ is the set of all principal components with non-zero eigen values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n_*}$. Let us put

$$x_m(r) = \sum_{q=1}^{r} x_{mq} w_q \ , \quad r \leq n_*$$

and let us denote by $X(r)$ the $(n \times (N - n + 1))$-matrix, m-th row of which is, by definition, n-dimensional vector $x_m(r)$.

We have $X(n_*) = X$, but for $r > n_*$ matrix $X(r)$ is not more Hankel-matrix.

Between all Hankel $(n \times (N - n + 1))$-matrixes, the nearest (in the Euclidean metric of the matrix space $R^{n(N-n+1)}$) to the matrix

$$X(r) = (x_{k,l}(r), \ 1 \leq k \leq n, \ 1 \leq l \leq N - n + 1)$$

is a matrix

$$HX(r) = (hx_{k,l}(r) = f_{k+l-1}(r)) \ ,$$

where

$$f_s(r) = \frac{1}{s_*} \sum_{k+l=s+1} x_{k,l}(r) \ , \quad s = 1, \ldots, N$$

and

$$S_* = \begin{cases} s \, , & \text{if } 1 \le s \le n \\ n \, , & \text{if } n \le s < N - n + 1 \\ N - s + 1 \, , & \text{if } N - n + 1 \le s \le N \end{cases}$$

To Hankel matrix $HX(r)$ corresponds time series $f(r) = (f_1(r), \ldots, f_N(r))$, where $r = 1, \ldots, n_*$ and $f(n_*) = f$.

Lemma. Non-parametric estimation of the initial time series $f = (f_1, \ldots, f_N)$ by the projection of data matrix $X = (x_{k,l} = f_{k+l-1})$ into the space of the first r main factors is a time series $f(r) = (f_1(r), \ldots, f_N(r))$.

§3. The results of analysis of some environmental monitoring time series.

3.1. Time series of the atmospheric CO_2 concentrations

For a demonstration of the method we selected the time series of CO_2 concentrations given in *Trends '90* (see also Elliot, ed., 1988).

In the present paper we use three time series:

- in Arctic zone – Barrow station (*Figure 9*);
- in Equatorial zone – Mauna Loa station (*Figure 10*);
- in Antarctic zone – South Pole station (*Figure 11*).

For each of the stations we present: (*a*) the initial time series of concentration; (*b*) the projections on the first, (*c*) the second, and (*d*) the fourth main components, and also on the planes of the four pairs of the principal components (*e*).

As it follows from the general theory (see §2) the projections (*b*), (*c*), and (*d*) gives non-parametric estimations of the trend, year and half-year cycles as the principal components of this series.

A visual analysis of the projections of these series into the factor space evidently demonstrate similar features in the structure of the series, and also their differences. Let us stress

some differences: the complex structure of the trend on Barrow that more strong distort of a harmonical component, and absence of the second half-year harmonic at the South Pole.

3.2. The temperature anomalies time series

The time series of the temperature anomalies are the longest monitoring time series. Global characteristics have a particular meaning – the anomalies of the temperatures averaged by the northern hemisphere, the southern hemisphere, and on the entire Earth – because they reflect the global tendencies in Climatic Changes. For analysis we used time series of Hansen-Lebedeff (*Trends '90*). For the series of mean year values during 107 years the notion of parameter n as a value proportional to the period of quasiharmonical oscillation is not completely correct, although it is not excluded the possibility of the correlations with the cycles of solar activity. Nevertheless the selection of this parameter has independent values as is noted in §2.

Figure 12 shows projections on the first principal components for a series of global temperature anomalies corresponding to the selection of the three values $n = 5$, 11 and 21. Here it is clearly seen that n has the sense of a smoothing parameter (analogously to the sense of parameter n in the algorithm of moving average). Further we use $n = 21$.

In *Figures 13, 14*, and *15* we present the results of the projections into the factor space of the three series of mean annual temperature anomalies averaged all over the globe Earth, T_{glob}, and also averaged by the northern hemisphere, T_N and by the southern hemisphere, T_S. From the comparison of these figures, it is seen that the structures of the two first series practically coincide. But the time series for the southern hemisphere is different from them. So the projection of T_S on the second principal component (*Figure 15c*) and on the plane of the second and the third main components (*Figure 15e*) points out the existing of a main cycle with $n = 21$, in spite of a strong distortion. At the same time, for the corresponding projections for the series T_{glob} and T_N (*Figure 13c* and *Figure 14e*), the main cycle is not observed. It could be interpreted as the main source of distortions of global temperature anomalies are placed in the Northern hemisphere. At the southern hemisphere these disturbances are coming in a smooth shape. And this is why the structure of the series T_S has a more natural character and follows to the cyclic law of 21 years.

The changes are seen even more exactly in the character of the projections on the second main factor for the time series T_{glob} and T_N.

From *Figure 13c* and *14c* it is easy to see that in the middle of the studying time interval there exists a fall in a comparing with the level in the beginning and the end of the time interval, i.e., the time series is decomposed on three intervals. At the second interval a restructurization takes place in comparison with the first interval, but on the third stage the initial structure of

the series is recovered. A testing of this hypothesis prompted by the analysis of the projections at factor space is presented in *Figure 16*. The analogous conclusion, we could make after appropriate investigation of the NDVI time series (see, for example, Appendix 1).

3.3. The time series of vegetation index NDVI

For the illustrative analysis we selected two characteristic series of observations. The first one (*Figure 17*) corresponds to the stable ecosystem (grass savanna) with a stable climatic and vegetation characteristics. The projections into the factor space show on the existing in the structure of the signal of two harmonics, and show a dynamics of its changes by years. A separate analysis of the picked-out structured components and comparing them with the real feature of the ecosystem, permit us to take into account the influence of the different parameters of the ecosystem on its vegetation activity.

Figure 18 corresponds to the ecosystem which in the process of dynamic reconstruction (transition) from one ecotype to the other. A projection on the first principal component shows two levels of the state of the system (complex trend). A projection on the second main component shows the main harmonic amplitude of which is changed sharply.

In spite of the length of the NDVI it is not long enough (only 7 seasons), the **proposed** method discovers not only the structure of time series but also the dynamics of its compounds. More detailed analysis of NDVI curves is given in the Appendices.

§4. The different methods of analysis of time series as the variants of a projection pursuit.

In this § we show that the approach to the analysis of time series on a base of *general shift* permits us to interpret as variants of a projection pursuit the following methods of the analysis of time series and widely used for statistical analysis of monitoring data: moving average, Karunen-Loev's decomposition, discrete Fourie transformation and others.

So we show that some set of methods appeared from the beginning *as* independent unlinked approaches as it turned out are the parts of a unified common approach. It permits us to make some ordering in this set of methods which means: when, why, and how to switch off from one method to another, for which goals, problems, hypotheses and so on a given method in the best (extremal in some sense).

It is shown in the example of the time series of El Niño.

1. The moving average as a procedure of projection pursuit.

Let $f = (f_1, \ldots, f_n)$ be a time series. Let us fix a number $n \ll N$ and some set of weight as vector $\alpha = (\alpha, \ldots, \alpha) \in R^n, \|\alpha\| \neq 0$. Let us construct *a set* of vectors $x_1, \ldots, x_{N-n+1}$, where $x_k = f_{k+l-1}, l = 1, \ldots, n$, *and* a data matrix $X = (x_{k,l}), x_{k,l} = f_{k+l-1}, k = 1, \ldots, N-n+1$, $l = 1, \ldots, n$, the row of which are vectors $x_1, \ldots, x_{N-n+1}$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-n+1} \end{pmatrix}$$

Then the series $g = g_\alpha = (g_1, \ldots, g_{N-n+1})$, where $g_k = \sum_{l=1}^{n} \alpha_l f_{k+l-1}$, $k = 1, \ldots, N-n+1$, as vector from R^{N-n+1} could be written in the form

$$g = X\alpha$$

It means that number series $g/\|\alpha\|$ is a series of projections of a set of n-dimensional vectors $x_1, \ldots, x_{N-n+1}$ on one dimensional subspace in R^n with directing vector $\alpha/\|\alpha\|$. Let us denote as $\alpha^* = (\alpha_1^*, \ldots, \alpha_n^*)$ directing vector of the first main component of the data matrix X.

Lemma. For any set at weights $\alpha = (\alpha_1, \ldots, \alpha_n)$ the following relation takes place:

$$\|g_\alpha - \tilde{g}_\alpha\|^2 \leq \|\alpha\|^2 \|g_{\alpha^*} - \tilde{g}_{\alpha^*}^*\|$$

where

$$\tilde{g}_\alpha = \bar{g} l(N-n+1) n \bar{g}_\alpha = \frac{1}{N-n+1} \sum_{k=1}^{N-n+1} g_k$$

mean value of α-weighted series g_α and $l(N-n+1) \in R^{N-n+1}$ is a vector each coordinate of which is equal to 1.

Proof. We have

$$\bar{g}_\alpha = \frac{1}{N-n+1}\ l'(N-n+1) g_\alpha = \frac{1}{N-n+1}\ l'(N-n+1) X\alpha$$

Hence, $g_\alpha = \hat{X}\alpha$, where

$$\hat{X} = \frac{1}{N-n+1}\ l'(N-n+1)\ l'(N-n+1) X\ .$$

Let us note now that $\hat{X}$ is a $(n \times (N-n+1))$-matrix each row of which coincide with the n-dimensional vector which was obtained from the initial series f by the procedure of moving average with a step equal to $(N-n+1)$.

Thus,

$$\|g_\alpha - \hat{g}_\alpha\|^2 = \|X_\alpha - \hat{X}\alpha\|^2 = \alpha^1(X-\hat{X})^1(X-\hat{X})\alpha = \alpha^1 V\alpha \ ,$$

where $V = (X-\hat{X})^1(X-\hat{X})$ is a covariation matrix of the set of n-dimensional vectors $x_1, \ldots, x_{N-n+1}$ as a transition from matrix X to matrix $X-\hat{X}$ corresponds to the centering of each column of the matrix X.

Now using the main theorem of the component analysis we obtain that the expression $\frac{\alpha^1 V\alpha}{\|\alpha\|}$ reaches a maximum if, and only if, when α is proportional to the vector α_* of the first main component of the data matrix X. So Lemma had proved.

Definition. Relative scattering of the moving α-weighted series g_α is called

$$\frac{\sum_{k=1}^{N-n+1}(g_{k,\alpha} - \bar{g}_\alpha)^2}{\|\alpha\|^2}$$

From the Lemma we get the following consequence. Between all moving α-weighted series $f = (f_1, \ldots, f_N)$ the largest relative scattering of the value has its moving α_*-weighted series, where α_1 is the first main component of the Hankel data matrix which corresponds to the series f.

Let us remark that the pursuit projection approach to the analysis of multidimensional data is contained in a description of the set of n-dimensional vectors under consideration in the terms of its extremal q-dimensional projection ($q \ll n$, frequently $q = 1$ or 2).

2. The notion of extremal projection.

The notion of extremal projection is introduced in the following way. Departing from the goal of analysis a criterion is selected that characterized projected data. This criterion defines a function on the manifold of all projections. The projection on which this function reaches the maximum is called extremal (the best) by this criterion.

From the preceding calculations it follows that moving weighting is a procedure of pursuit projection. The consequence shows that when as a criterion we select the dispersion, then the extremal is α_*-weighting. Here α_* is the first main component of the matrix $X = (f_{k+l-1})$, $k = 1, \ldots, N-n+1$, $l = 1, \ldots, n$. It is clear that for the other criteria, which characterize time series g_α the *extremal* will be the different vectors of weights.

In the conclusion of this point let us mark that the moving weighting of a time series $f = (f_1, \ldots, f_N)$ would be considered as a particular case of more general procedure corresponding to the following hypothesis: in the sequences of the intervals of length n of time series $f = (f_1, \ldots, f_N)$ each K-th interval x_K is a vector of n realization of K-th mark of the time signal under the investigation.

The method of treatment is concluded in construction of assessment of the value of each K-th mark by this realization. In particular, the moving average responds to assessment of the mark value as mean by n realization, and median smoothing as a median of n realizations. In such an approach median smoothing, for example, became a variant of nonlinear pursuit projecting, which corresponds to application to the rows of Hankel-matrix X of nonlinear operator of the projecting.

3. An analysis of time signals by the shift method and a Karunen-Loev's discrete decomposition.

First of all let us remark that the necessary facts about Karunen-Loev's decomposition are to be found, in more detail, in Fukunaga, 1972.

Let x be n-dimensional random vector. Let us choose an orthogonal and normed basis $\phi = [\phi_1, \ldots, \phi_n]$ in the space R^n. Then

$$x = \sum_{i=1}^{n} y_i \phi_i = \phi_y \ ,$$

where $y = (y_1, \ldots, y_n)$. Each coordinate y_i of the vector y is a feature of the initial vector x. Let us suppose that we would like to characterize vector x only by q features $y_1, \ldots, y_q$, $q < n$ fixing for the rest of the coordinates some values $y_{q+1} = b_{q+1}, \ldots, y_n = b_n$. Then instead of x we obtain its assessment

$$\hat{x}(q) = \sum_{i=1}^{q} y_i \phi_i + \sum_{i=q+1}^{n} b_i \phi_i \ .$$

Let us denote

$$\Delta x(q) = x - \hat{x}(q) = \sum_{i=q+1}^{n} (y_i - b_i) \phi_i \ .$$

Let us use a mean value of the square of norm at random vector $\Delta x(q)$ for measurement of the efficiency of subset of q-feature $y_1, \ldots, y_q$:

$$\bar{\varepsilon}^2(q) = E\left\{ \|\Delta x(q)\|^2 \right\} = \sum_{i=q+1}^{n} E\left\{ (y_i - b_i)^2 \right\} \ .$$

To each set of basis vectors $\phi_1, \ldots, \phi_n$ and the values of the constants $b_{q+1}, \ldots, b_n$ corresponds to some value of the formula $\bar{\varepsilon}^2(q)$.

Lemma (Karunen-Loev). Between *all* orthonormed basis $\phi = (\phi_1, \ldots, \phi_n)$ and *possible* constants $b_{q+1}, \ldots, b_n$ the *minimal* value of mean-square error is reached when $\phi_1, \ldots, \phi_n$ is a set of the proper vectors of covariance matrix Σ_x of random vector x, ordered by decreasing a proper numbers $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and $b_i = \phi_i' E\{x\}, i = q+1, \ldots, n$.

Definition. By discriminant Karunen-Loev's decomposition of random vector x is called decomposition by the basis of proper vectors of its covariance matrix.

In that case, when instead of random vector x we have data matrix X the row of which are a realization of the random vector x. Then, using a decomposition by the first q proper vectors of the covariance matrix of the sample X, we obtain sampling Karunen-Loev's discrete q-component decomposition. Turning to the time signal $f(t)$, to the shift operator A and to the set of marks $t_1, \ldots, t_n$ (see §1), let us consider our method from the point of view of hypotheses that the vectors $x_1 = (f(t_1), f(t_2), \ldots, f(t_n)),\ x_m = ((A^{m-1}f)(t_1), \ldots, (A^{m-1}f)(t_n))$, are realization of some random vector that completely characterized initial time signal.

Then projections of the set of the vectors, $x_1, \ldots, x_m$ on the space of the first q-main vectors of its covariance matrix will exactly correspond to the sampling Karunen-Loev's discrete q-component decomposition.

The specific feature of the method of the analysis of time signal with the aid of the shift operator A is that principal addition to the Karunen-Loev's decomposition: we analyze projections on q-main component of a curve $\varphi(\tau, f)$ as well as a set of vectors $x_1, \ldots, x_m, \ldots$

4. Analysis of the time series by shift method and discrete Fourie transformation.

Let us fix in the space R^π an orthogonal basis Fourie $W = [W_1, \ldots, W_n]$, where row

$$W_l = \left(\sin \frac{2\pi}{n} lt, t = 0, 1, \ldots, n-1\right) \in R^n$$

Let us remark that $(W_{l_1}, W_{l_2}) = \frac{n}{2}\delta_{l_1 l_2}$, where $\delta_{l_1 l_2}$ is a Kroneker's feature. In particular $\|W_l\|^2 = \frac{n}{2}$.

Definition. Discrete Fourie's transformation of the random vector x is called its decomposition by the Fourie basis.

Let $X = (x_{k,l}), x_{k,l} = f_{k+l-1}$ be a Hankel matrix of the time series $f = (f_1, \ldots, f_N)$.

For the testing of hypothesis that on each interval of the length n of the marks of the time series are formed all its main cycles. It is justified decomposition of the rows of the matrix X by the Fourie basis. Hence, we have a transition from the matrix X to the matrix $Y = (y_{ke})$, where $Y = WX^1$.

A matrix coefficient $y_{k,l}$ equal to amplitude of a harmonic signal $\sin \frac{2\pi}{n}\, lt$, $l = \overline{1, A}$, in the Fourie decomposition of k-th interval x_k, $k = 1, \ldots, N - n + 1$, of time series f. Thus coordinates $Y_{Pn+1,l}$, $P = 0, 1, \ldots$ are vectors of columns y^l of matrix Y gives information about the dynamics of the amplitude of harmonic component $\sin \frac{2\pi}{n}\, lt$ of initial series (in the conditions of the hypothesis that was given above).

Usually in the Fourie basis the vectors $W_1, \ldots, W_n$ are ordered by an increasing of the frequency of the harmonical signal. From the point of view of projection pursuit (for example, by criterion of dispersion, as in Karunen-Loev's decomposition) the basis vectors should be ordered by decreasing of the criterion

$$\Delta q = \sum_{k=1}^{N-n+1} \|W_q' X_k - W_q' \bar{X}\|^2 \ ,$$

where $\bar{x} \in R^n$ is a vector of means of vectors-columns of the matrix X, i.e., basis ϕ is taken in the form

$$\phi_{i1}, \phi_{i2}, \ldots, \phi_{in}$$

where $\Delta_{i1} \geq, \Delta_{i2}, \geq \cdots \geq \Delta_{in}$.

Under such ordering for the polyharmonial signals of the form

$$f(t) = \sum_{q=1}^{n} C_n \sin \frac{2\pi}{n} \alpha t$$

and the set of marks as natural numbers the shift method gives the same results as at Karunen-Loev's decomposition, so at discrete Fourie's transformation.

§5. A development of the method on the base of the theory of generalized shift.

Let us consider the following interpretation of the proposed approach. It could be accepted that from analysis of a time series of $f(t) = (f(t_1), \ldots, f(t_N))$ we pass to the analysis of a random process defined by its realizations in the time moments $t_1, \ldots, t_n$. Operator A is selected in such a way that the values $(A^{m-1} f)(t_k)$, $k = 1, \ldots, n$, could be considered as values of m-th realization of the process in these points.

As is shown in §4 in such interpretation an application of the factor analysis equal to an application Karunen-Loev's filter of this random process. A construction of the set of the important operator A for a generation of the realization of a random process is based on a general theory of the operators of a generalized shift that in its turn was created as a development of group approach to a description of physical phenomena.

In the case of continuous time such an operator of generalized shift T_t^τ corresponds to the time signal $f(t)$ a function of two variables $\Theta(t,\tau) = T_t^\tau f(t)$ that is fulfilled in the following conditions:

1. T_t^0 is an identity operator .
2. $T_t^{\tau_1}\Theta(t,\tau_2) = T_t^{\tau_1}\left(T_t^{\tau_2} f(t)\right) = T_{\tau_1}^{\tau_2}\left(T_t^{\tau_1} f(t)\right)$.

Let us explain in more detail the operation of composition of generalized shift operators. A function

$$\Theta(t,\tau) = T_t^\tau f(t)$$

could be considered as a function of an argument t and a parameter τ, and at the same time as a function of the argument τ and the parameter t. In notation of the operator T_t^τ the low index serves the notation of argument on functions of which this operator acts. Thus a function

$$T_t^{\tau_1} T_t^{\tau_2} f(t) = T_t^{\tau_1}\Theta(t,\tau_2)$$

is obtained by application of operator $T_t^{\tau_1}$ to the $\Theta(t_1\tau_2)$ as a function of variable t. A function

$$T_{\tau_1}^{\tau_2} T_t^{\tau_1} f(t) = T_{\tau_2}^{\tau_1}\Theta(t_1\tau_1)$$

is obtained by application of the operator $T_{\tau_1}^{\tau_2}$ to the $\Theta(t,\tau_1)$ as to a function of the argument τ_1.

The first examples of generalized shift operator are are linked with the group structure on R^1. Indeed, let * be some operation on R^1, for example:

$$t_1 * t_2 = t_1 + t_2 \ ,$$

or

$$t_1 * t_2 = t_1 + t_2 + t_1 \cdot t_2 \ ,$$

or

$$t_1 * t_2 = \frac{t_1 + t_2}{1 - t_1 t_2} \ ,$$

then formula $T_t^\tau f(t) = f(t * \tau)$ gives a shift on the set of the functions fulfilled to the axioms 1, 2. However, an operation of product $*$ does not cover all possible shifts. It is easy to verify that formula

$$T_t^\tau f(t) = \frac{1}{2}\left(f((\sqrt{t}+\sqrt{\tau})^2\right) + f\left((\sqrt{t}-\sqrt{\tau})^2\right)$$

gives a shift on the set of function also fulfills to the axioms.

Considering $\tau = \Delta t$ as a constant in the method described above it is possible to put: $T_t^{\Delta t} = A$. It explains why the operator A was called the operator of generalized shift.

It is clear that not each of the operators A has a shape $T_\tau^{\Delta t}$. But as our method is based on using eigen functions of operator A, then using the operators of the shape $T_t^{\Delta t}$ is preferable as ***exactly*** its eigen functions appear as model functions for the important physical processes. But a function $f(t)$ is an eigen function of the operator $T_t^{\Delta t}$ if when it is a solution of the equation

$$D_t f(t) = \lambda f(t) \ ,$$

where operator D_t acts on the functions by formula

$$D_t f(t) = \frac{\partial}{\partial \tau} T_t^\tau f(t)|_{\tau \geq 0} \ .$$

It is turned out that in series of the important cases D_t is a differential operator. For example, for a shift associated with usual addition on R^1, the operator D_t is the operator $\frac{d}{dt}$. For operator given by formula () this operator is

$$\frac{d}{dt} - 2t\frac{d^2}{dt^2} \ .$$

Let us stress that in reality the operators of the generalized shift are constructed by differential operators, the eigen functions of which model the chosen physical phenomena. Namely, there exists a general presentation for the operator of general shift as follows:

$$T_t^\tau f(t) = \psi(\tau \cdot D_t) f(t) \ ,$$

(Kolunogorof theorem of presentation of a function at two variables as a composition of the functions of one variable in the wide sense).

The $\psi(t)$ is an eigen function of the operator D_t with eigen values 1, i.e., a solution of the equation

$$D_t \psi(t) = \psi(t)$$

with the initial condition $\psi(0) = 1$.

In the case when $D_t = \frac{d}{dt}$ we obtain that $\psi(t) = l^t$ and formula $T_t^\tau f(t) = \psi(\tau D_t) f(t)$ transits to the formula:

$$T_t^\tau f(t) = f(t+\tau) = \exp\left(\tau \frac{d}{dt}\right) f(t)$$

i.e., to the classical form of Newton's series for the function $f(t)$.

In the case when

$$D_t = \frac{d}{dt} - 2t\frac{d^2}{dt^2}$$

we obtain that

$$\psi(t) = \frac{l^{\sqrt{t}} + l^{-\sqrt{t}}}{2} ch\sqrt{t} \ ,$$

and formula for corresponding generalized shift takes the form

$$T_t^\tau f(t) = \frac{f((\sqrt{t}+\sqrt{\tau})^2) + f((\sqrt{t}-\sqrt{\tau})^2)}{2} = ch\sqrt{aD_t} f(t) \ ,$$

i.e., presentation of two-values shift by its generator Buchstaber (1975).

For comparison of the results of the treatment of the signal $f(t)$ by our method, but corresponding to the one and two-valued shifts, we applied our method of treatment to the following signal

$$f(t) = 0.3t + (0.1 + 0.2t)\cos 1.555\pi\sqrt{t+1} \ .$$

As can be seen from this formula and *Figures 19* and *20* we are dealing with a **strong** nonstationary signal, because the distance between the picks increases as time increases.

Remark: For convenience of comparison with the results for operator A, given in *Figure 19*, the set of vectors rows of matrix X for operator A_2 is formed in the same way as that for operator A_1, i.e., the l-th row of matrix X is obtained as value of function

$$T_k^l f = \frac{1}{2}[f(\sqrt{k}+\sqrt{l})^2 + f(\sqrt{k}-\sqrt{l})^2] \ ,$$

where

$$f(t) = a_1 t + (a_2 + a_3 t)\cos(a_4\sqrt{t}) \ ,$$

i.e.,

$$f(K) = N_k \ .$$

References

Afifi, A.A., Azen, S.P., *Statistical Analysis. A Computer Oriented Approach*, Academic Press, New York, 1979.

Aivasian, S.I., Buchstaber, V.M., Yenyukov, I.S., Meshalkin, L.D., *Applied Statistics, Classification and Reduction of Dimension*, Moscow, Finance and Statistics, 1989.

Buchstaber, V.M., Two-values formal groups, Algebraic theory and application to cobordisms I. Proceedings of the USSR Academy of Sciences, Math. ser. 1975, V39, pp. 1044–1064.

Callendar, G.S., 1938, The artificial production of carbon dioxide and its influence on temperature, Q.Y. Roy, *Meteorol. Soc.* 64, 223–240.

Delsarte, J., Sur une extension de la formulae de Taylor, *Journ. Math. pures et appl.* 14 (1938), 213–230.

Delsarte, J., Hypergroupes et opérateurs de permutation et de transmutation, *Colloquer internat. Nancy* (9–15 avril)(1956), 29–44.

Elliott, W.P., ed., *Statistical Treatment of CO_2 Data Records*, NOAA Tech. Memorandum ERL ARL-173, Air Resources Laboratory, Silver Spring, Maryland, May 1989.

Fu, K.S., *Synthetic Methods in Pattern Recognition*, Academic Press, New York, London, 1974.

Gantmacher, F.R., Theory of matrix, *Nauka*, Moscow 1967, p.575.

Harman, G., Modern Factor Analysis.

Justice, C.O., Townshend, J.R.G., Holben, B.N., Tucker, C.J., Analysis of the phenology of global vegetation using meteorological satellite data, *Remote Sensing*, 1985, Vol. 6, pp. 1271-1318.

Levitan, B.M., Theory of operators of generalized shift, *Nauka*, Moscow, 1973.

Malingreau, J.P., Global vegetation dynamics: satellite observations over Asia, *Int. J. Remote Sensing*, 1986, Vol. 7, pp. 1121-1146.

Townshend, J.R.G., Goff, T.E., Tucker, C.J., Multitemporal dimensionality of images of normalized difference vegetation index at continental scales, *IEEE Trans. on Geosci. and Remote Sensing*, 1985, GE-23, N 6, pp. 888-895.

Townshend, J.R.G., Justice, C.O., Kalb, V., Characterization and classification of South American land cover types using satellite data, *Int. J. Remote Sensing*, 1987, Vol. 8, pp. 1189-1207.

TRENDS '90, CDIAC, T.A. Boden, P. Danciruk, M.P. Farrell.

Tucker, C.J., Townshend, J.R.G., Goff, T.E., African land cover classification using satellite data, *Science*, 1985, Vol. 227, pp. 369-375.

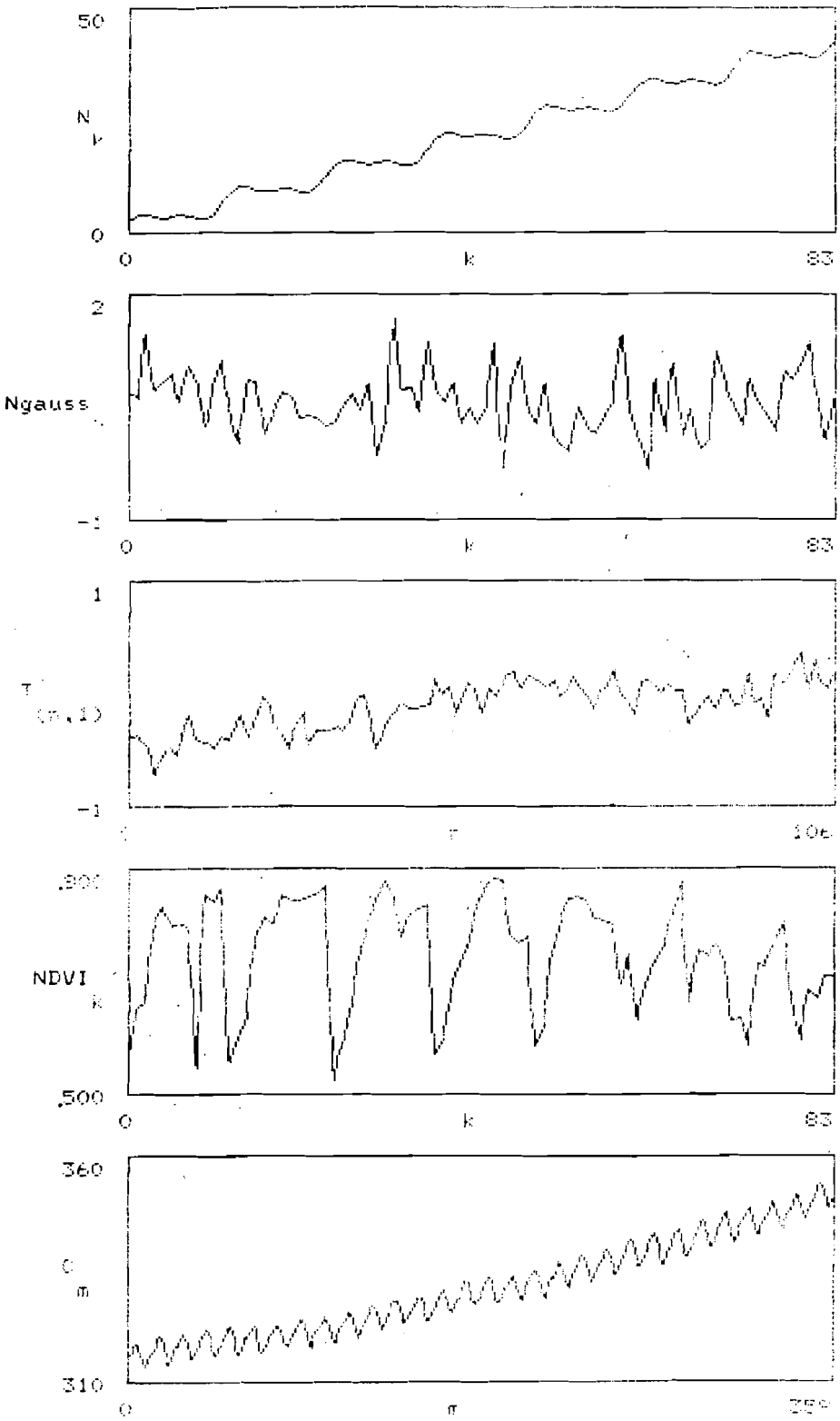


Figure 1. Characteristic curves considered: (a) polyharmonical curve with a trend; (b) time series obtained by random numbers generator; (c) yearly temperature anomalies for 107 years; (d) mean monthly values of vegetation index NDVI for 7 years; (e) mean monthly concentration of atmospheric CO_2 on Mauna Loa station for 30 years.

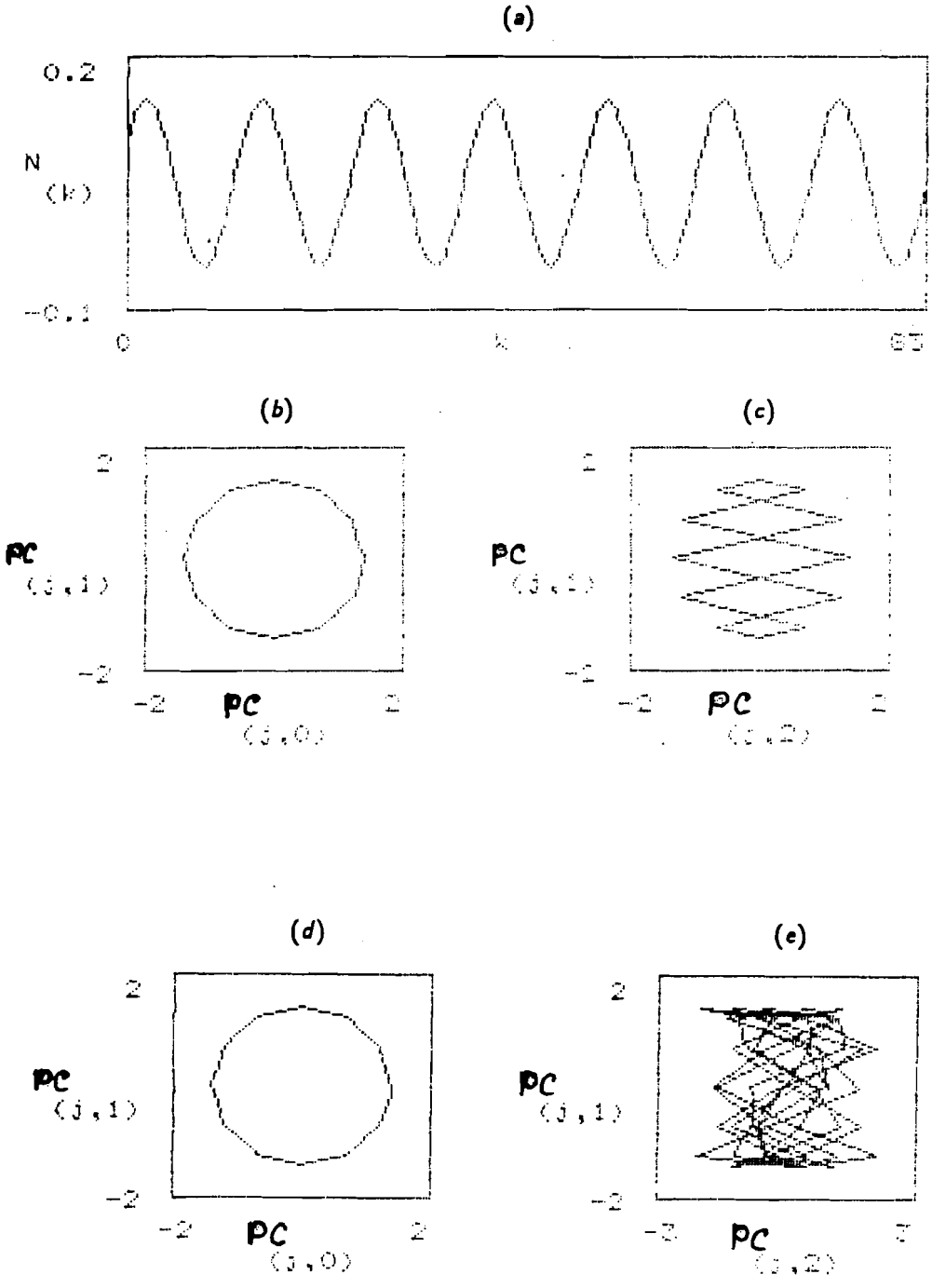


Figure 2. (a) Harmonics $N_k = a_0 + a_1 \sin(wk)$, $a_0 = 0.5, a_1 = 1$; and its projections on the planes of the factor space; (b) and (c) – under the calculation of N_k until the third valid figure; (d) and (e) – under the calculation of N_k until the fifth valid figure.

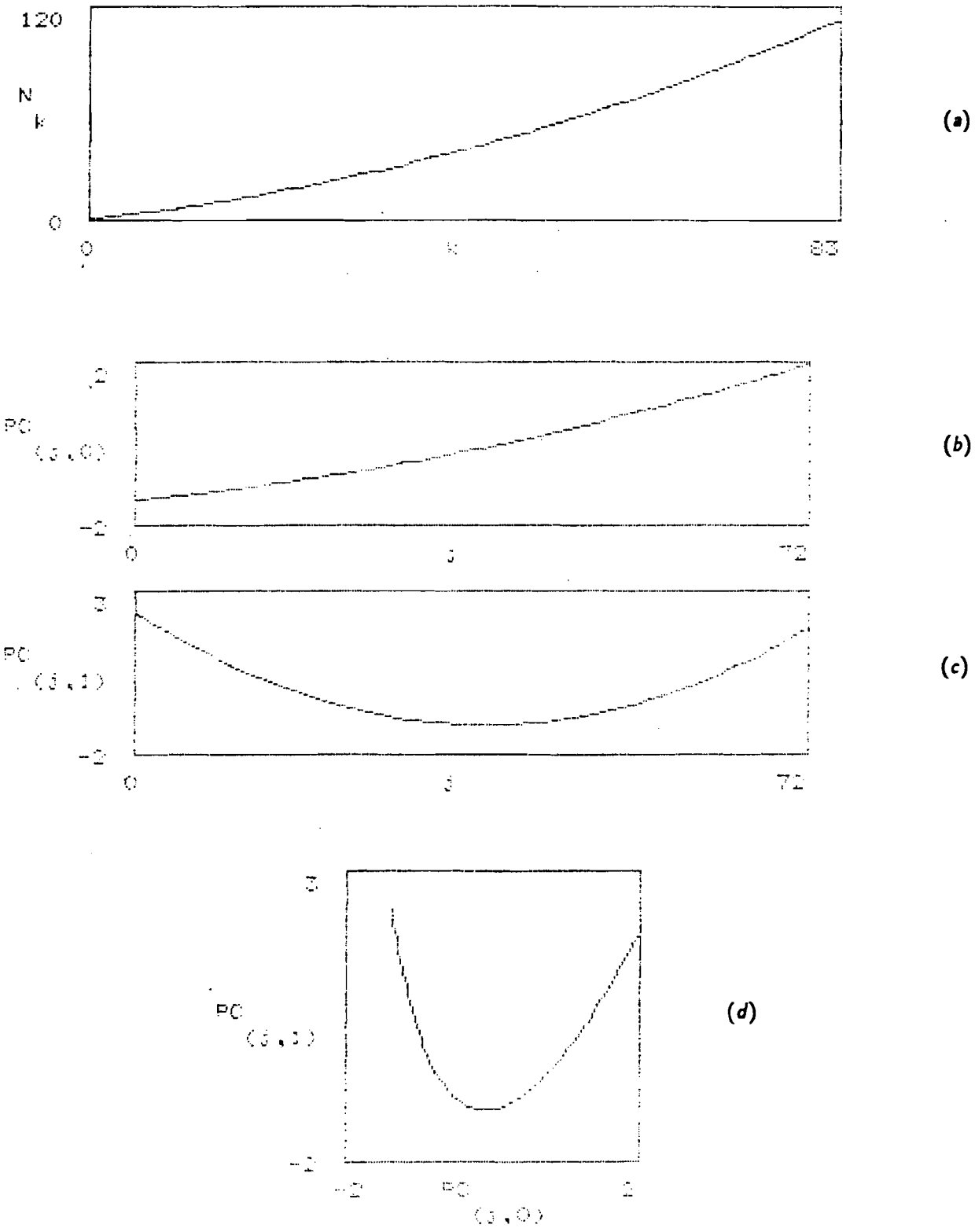
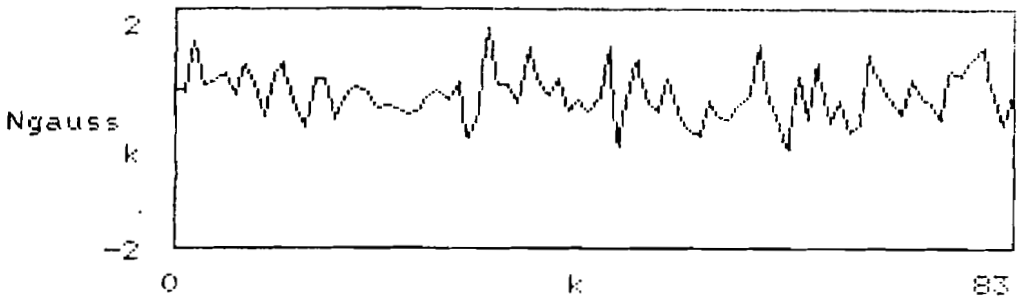
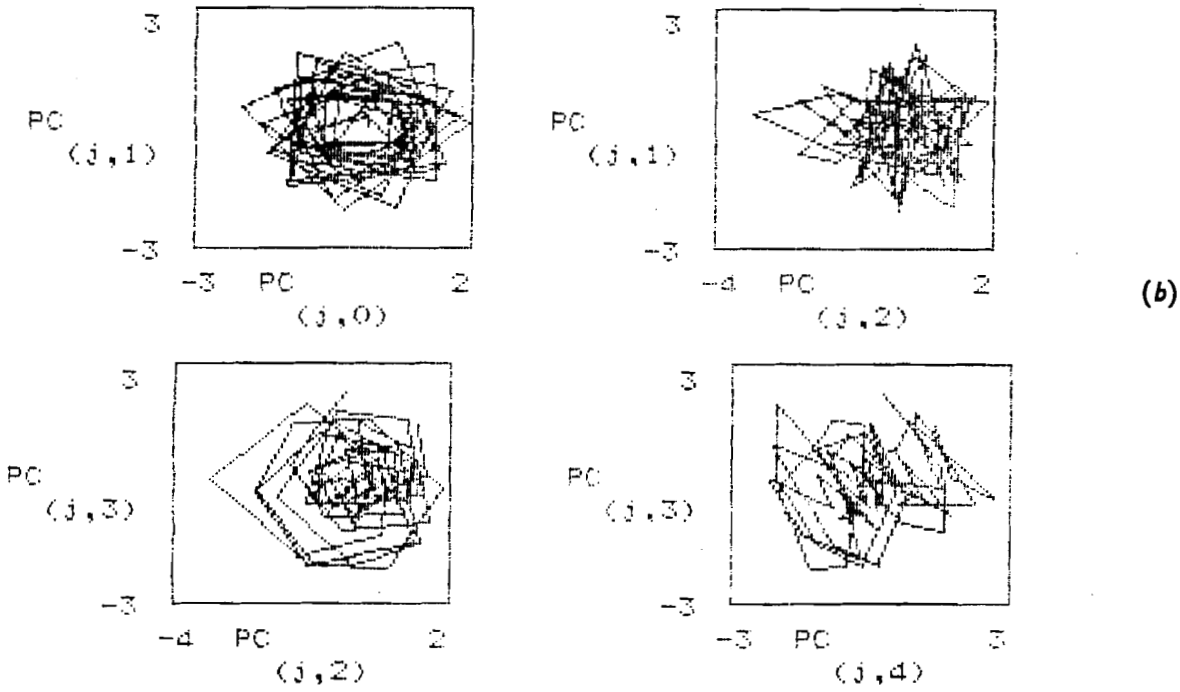


Figure 3. (a) Parabola $N_k = a_0 + a_1 k + a + 2k^2$; and its projection on the (b) first principal component; (c) second principal component; (d) plane of the first pair of principal components.



(a)



(b)

Figure 4. (*a*) Time series modeled by a random numbers generator; (*b*) its projections on the planes of the principal components.

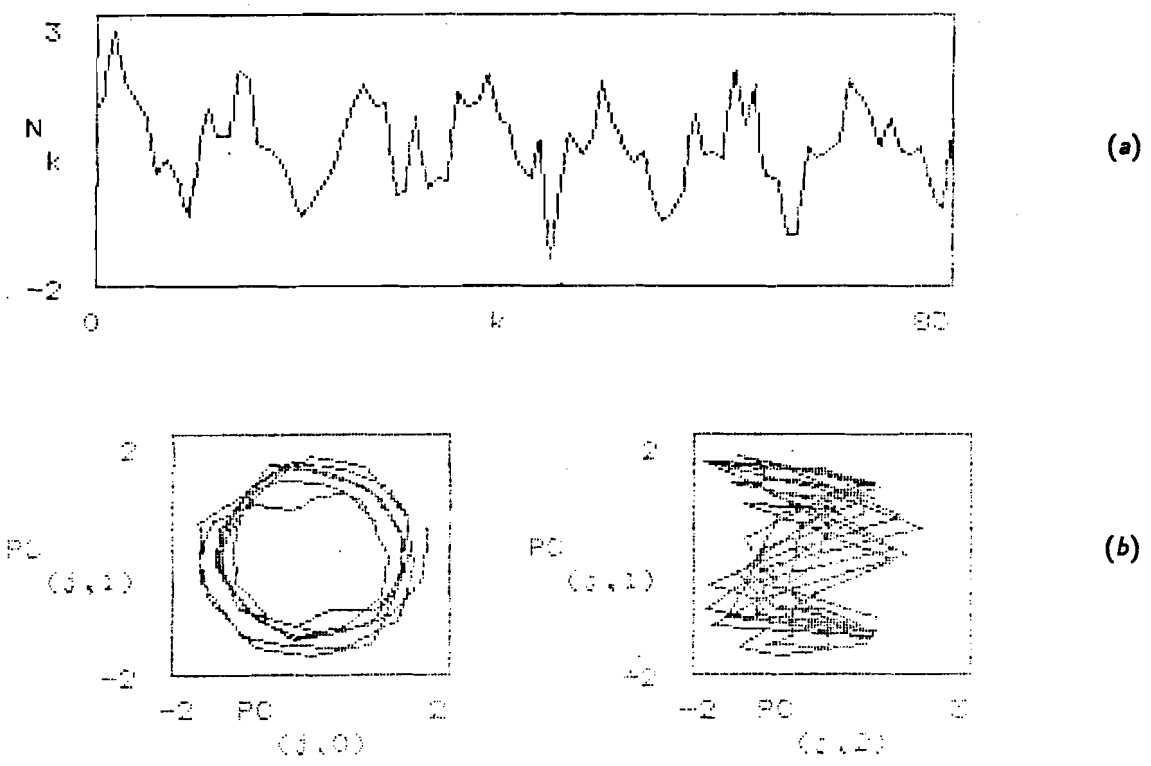
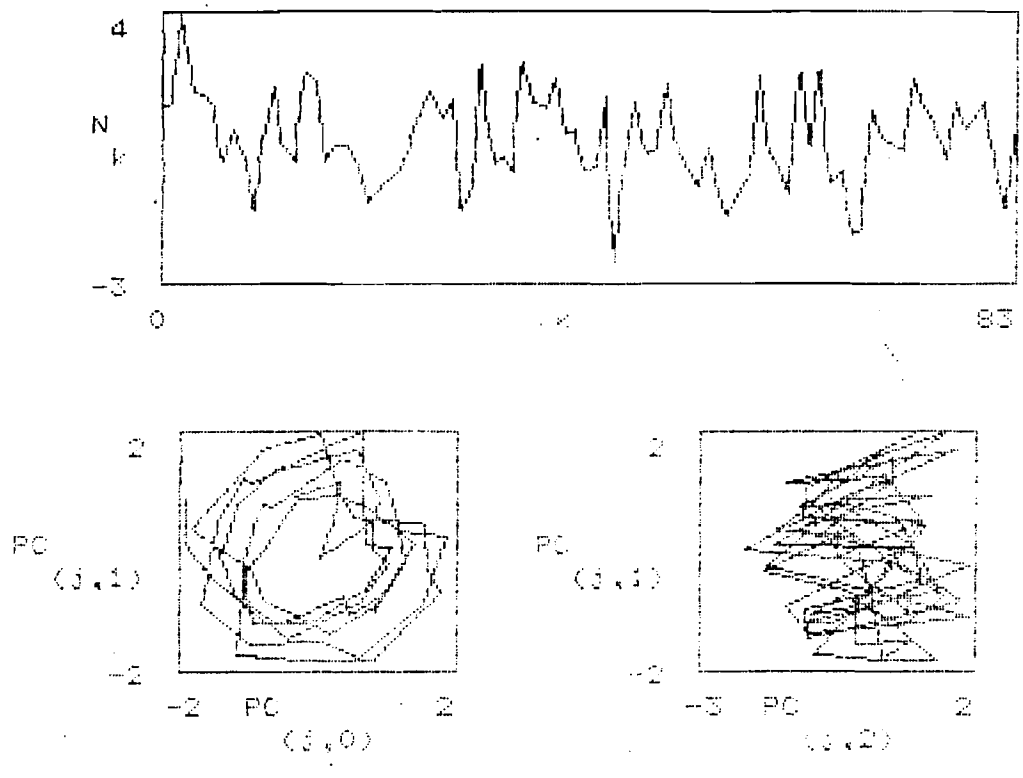


Figure 5. (*a*) Harmonics as in *Figure 2* but distorted by random noise, obtained by generator of random numbers by the law of normal distribution, $\sigma = 0.5$; (*b*) its projections on the planes of the principal components.



(a)

(b)

Figure 6. As on *Figure 6* but $\sigma = 1.0$.

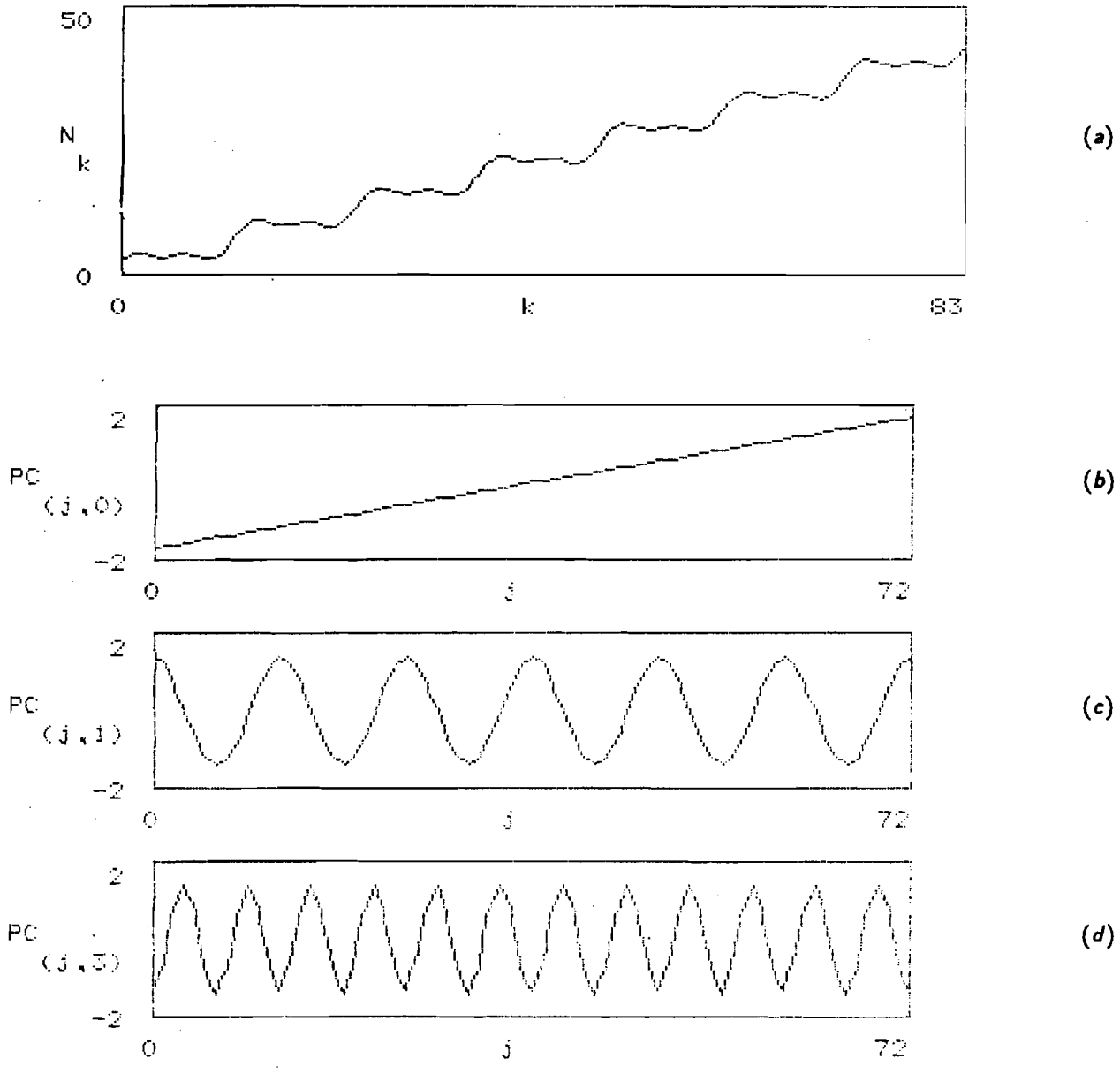


Figure 7. (a) Polyharmonical signal with linear trend

$$N_k = a_0 + a_1 k + a_2 \sin(2\pi k/12) + a\,,\sin(2\pi k/6)\ ;$$

and its projection on the (b) first principal component; (c) second principal component; (d) fourth principal component.

(e)

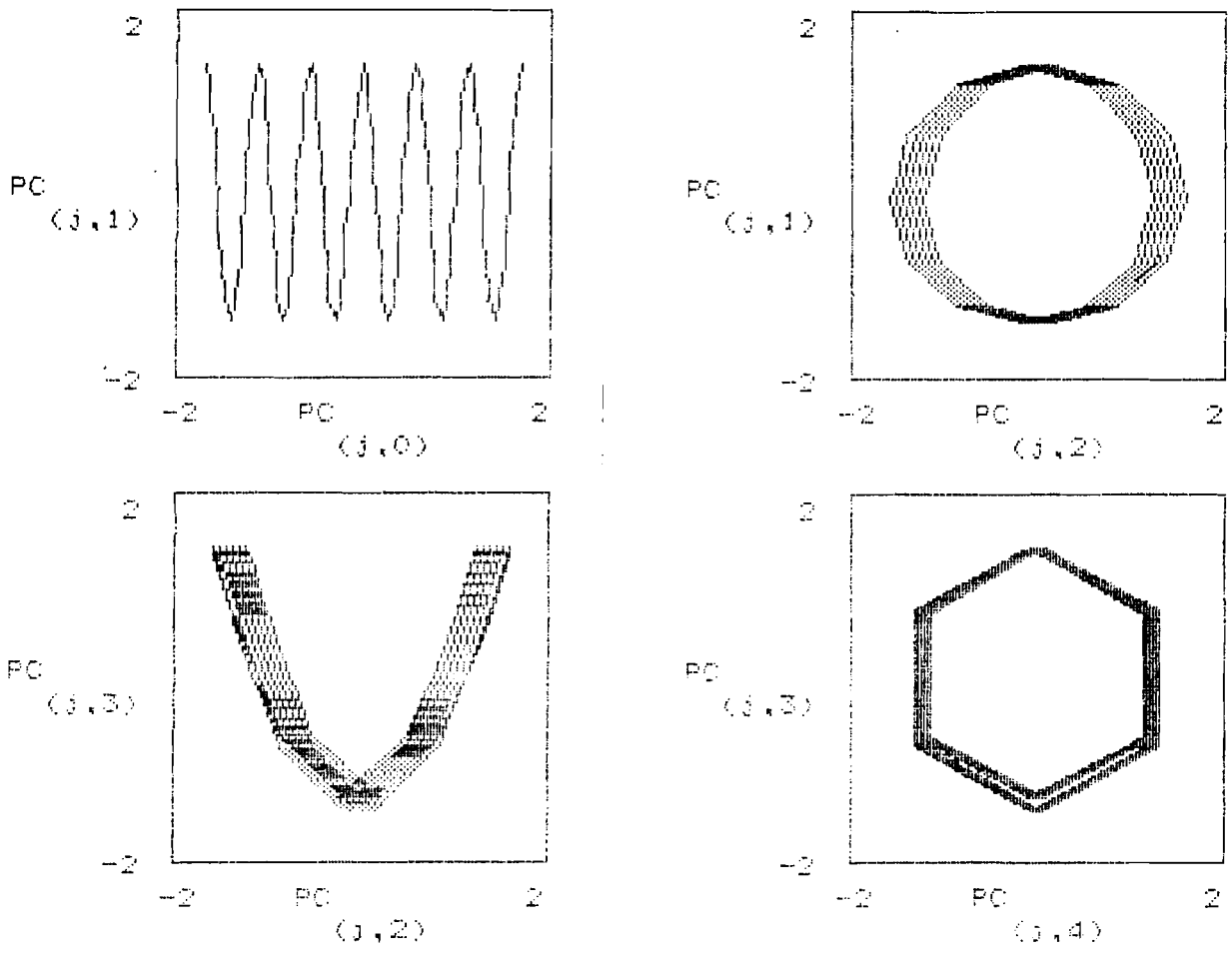


Figure 7. Continued. (*e*) projection on the planes of principal components pairs.

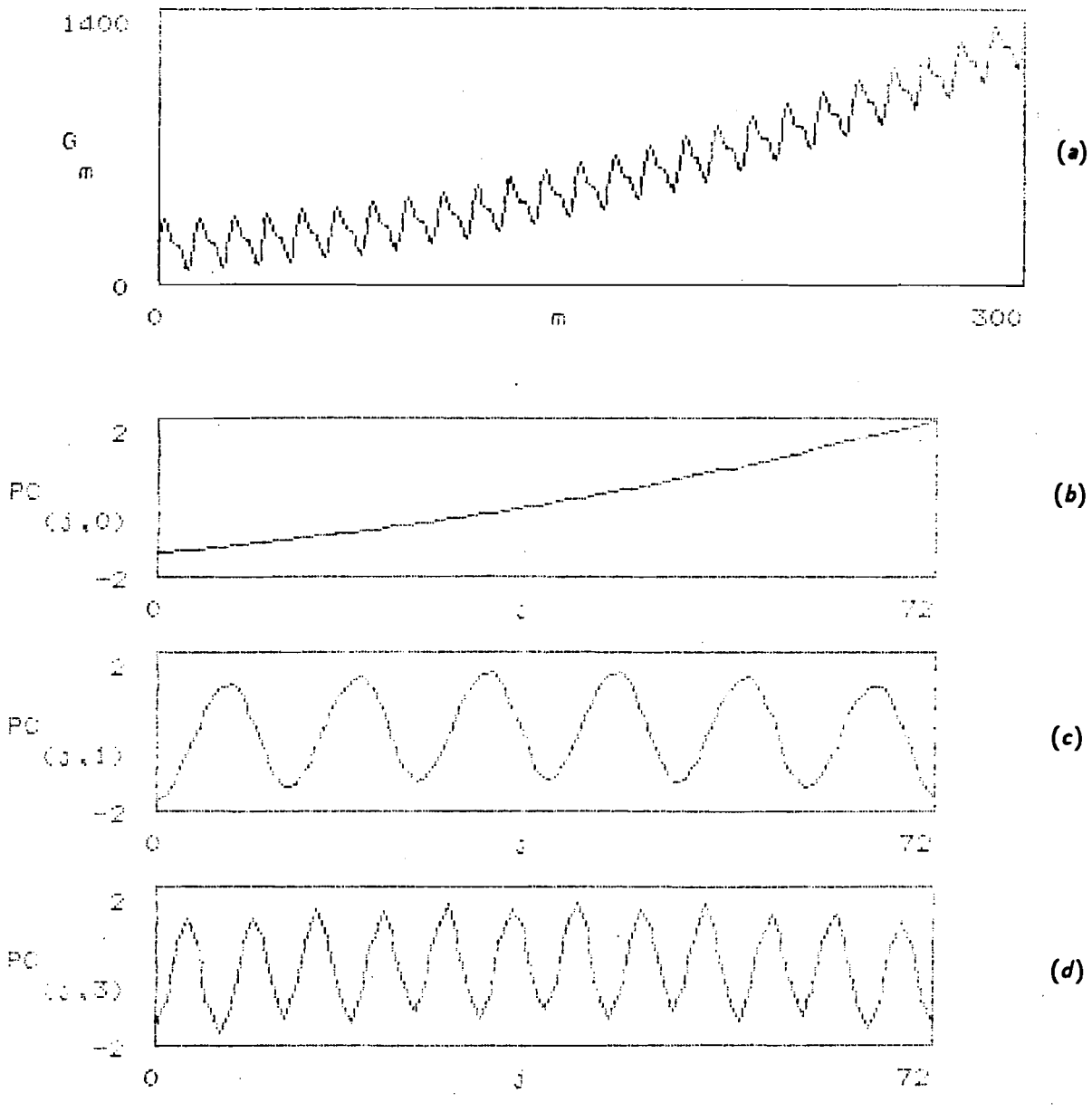


Figure 8. (a) Polyharmonical signal with parabolic trend:

$$N_k = a_0 + a_1 k + a_2 k^2 + a_3 \sin(2\pi k/12) + a_4 \sin(2\pi k/6) \ ;$$

(b), (c), (d), and (e) the same as in *Figure 7*.

(e)

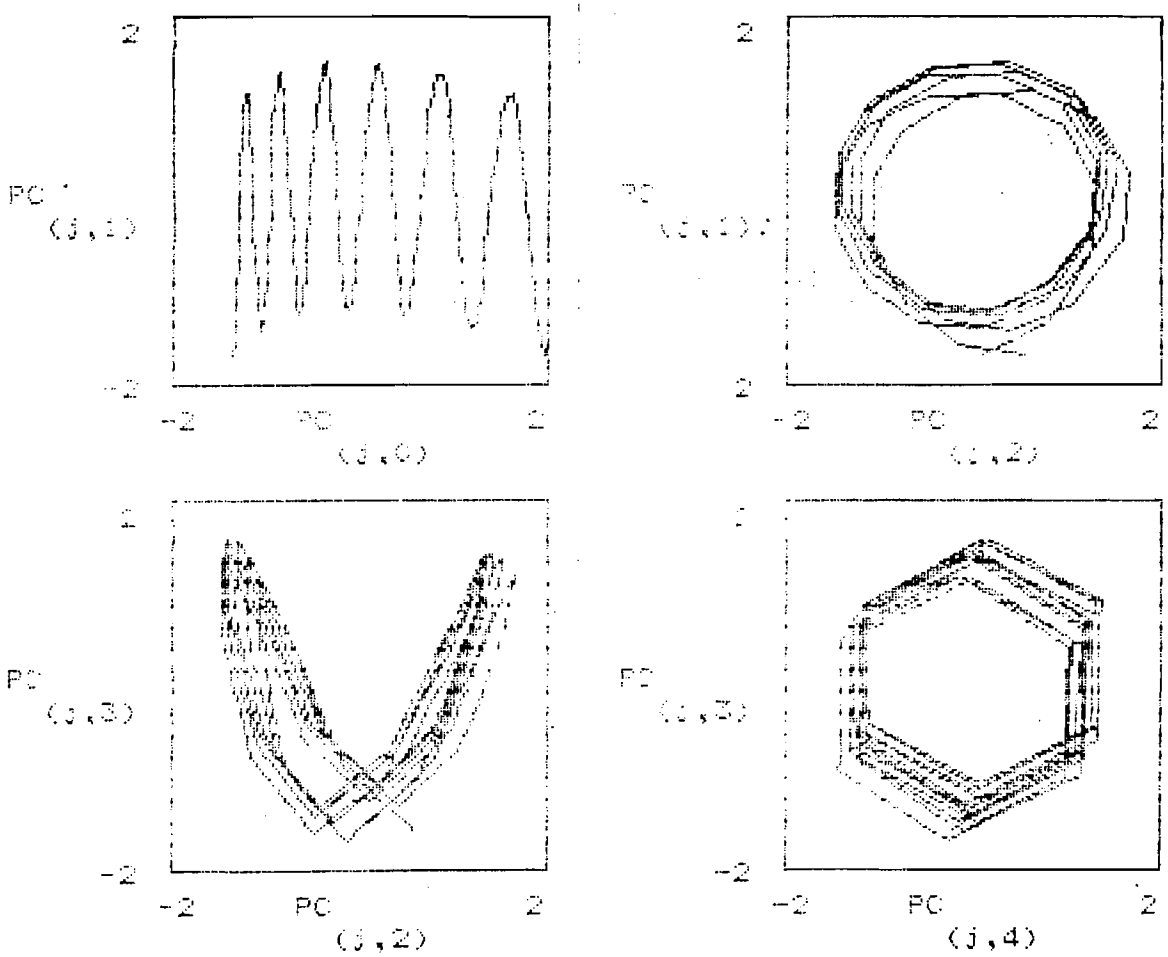


Figure 8. Continued.

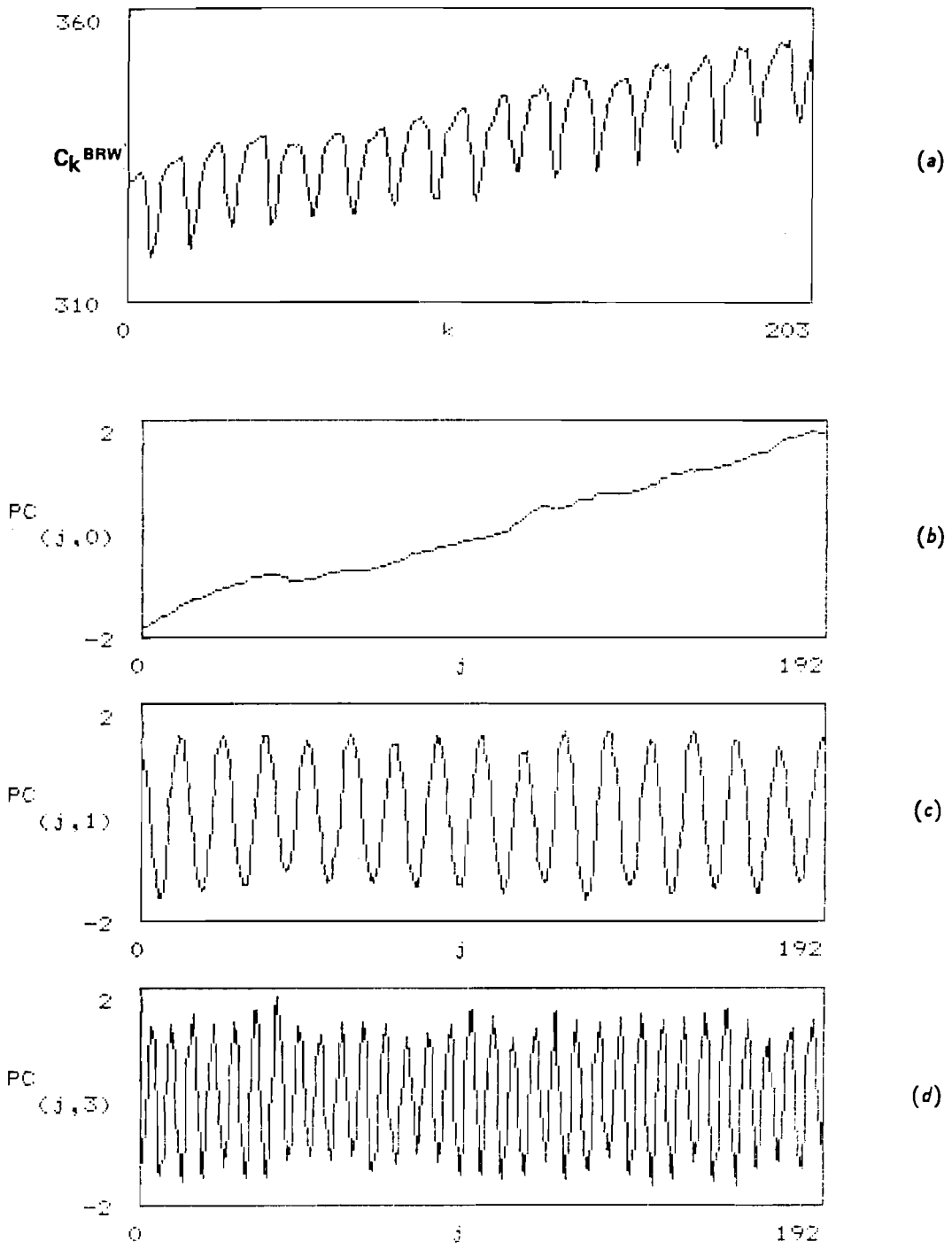


Figure 9. (a) Time series of atmospheric CO_2 concentrations on Barrow station; (b), (c), (d), and (e) as on *Figure 8*.

(e)

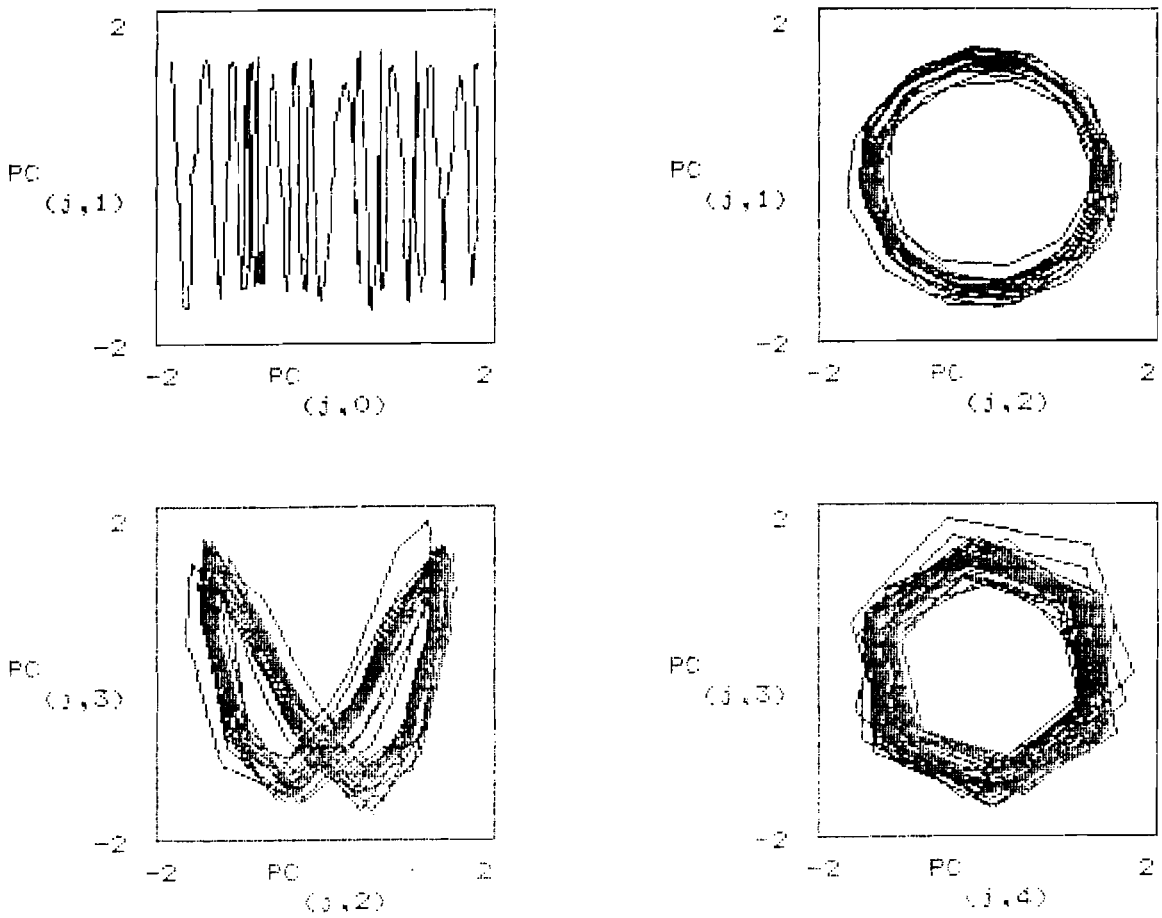


Figure 9. Continued.

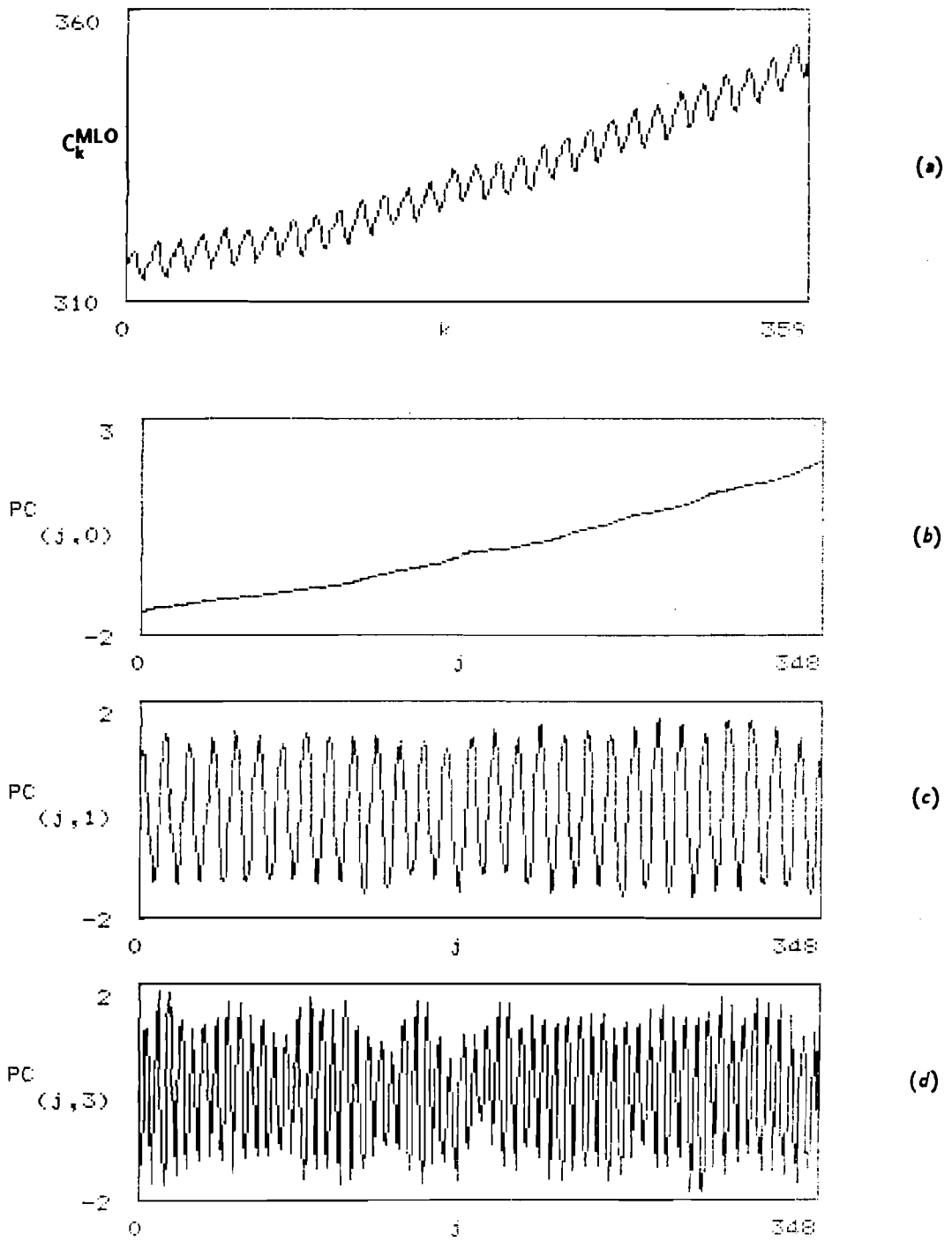


Figure 10. (a) Time series of atmospheric CO_2 concentration on Mauna Loa station; (b), (c), (d), and (e) as on *Figure 8*.

(e)

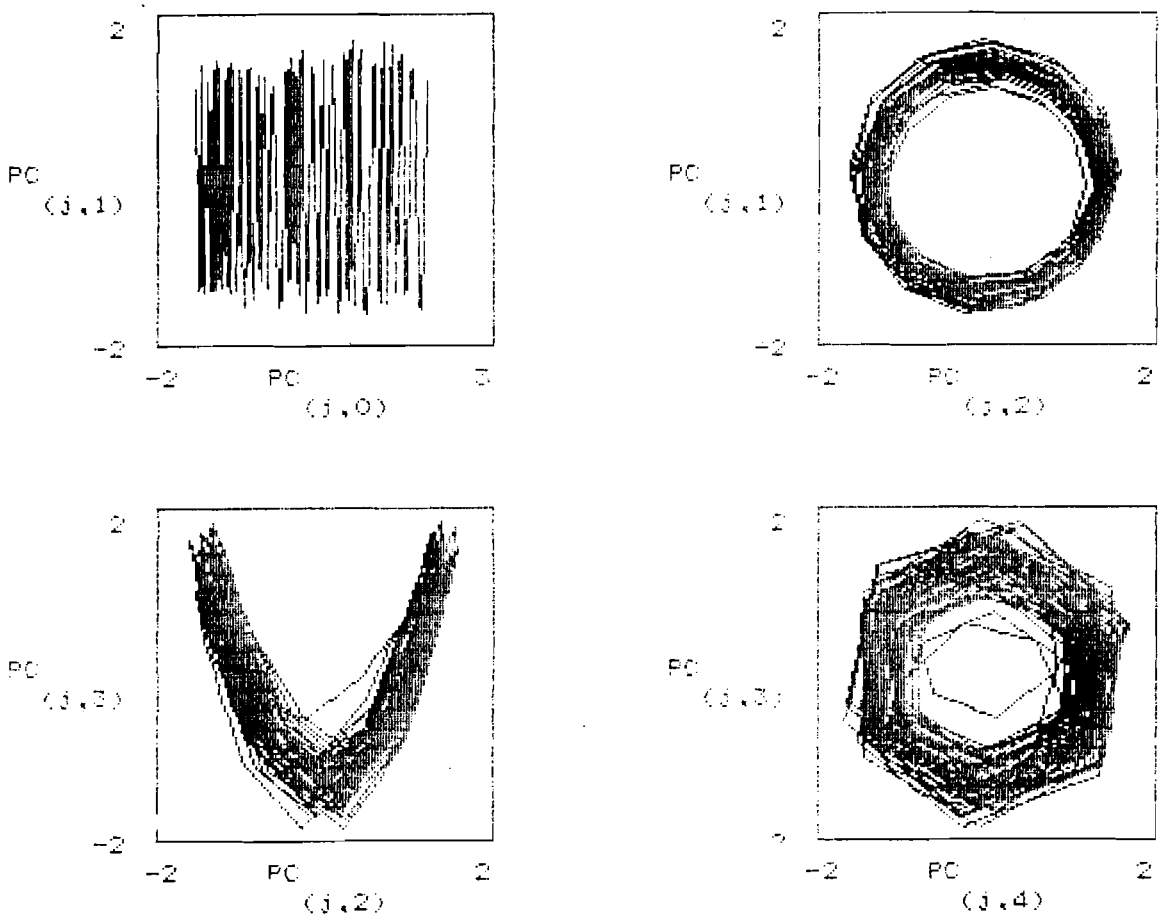


Figure 10. Continued.

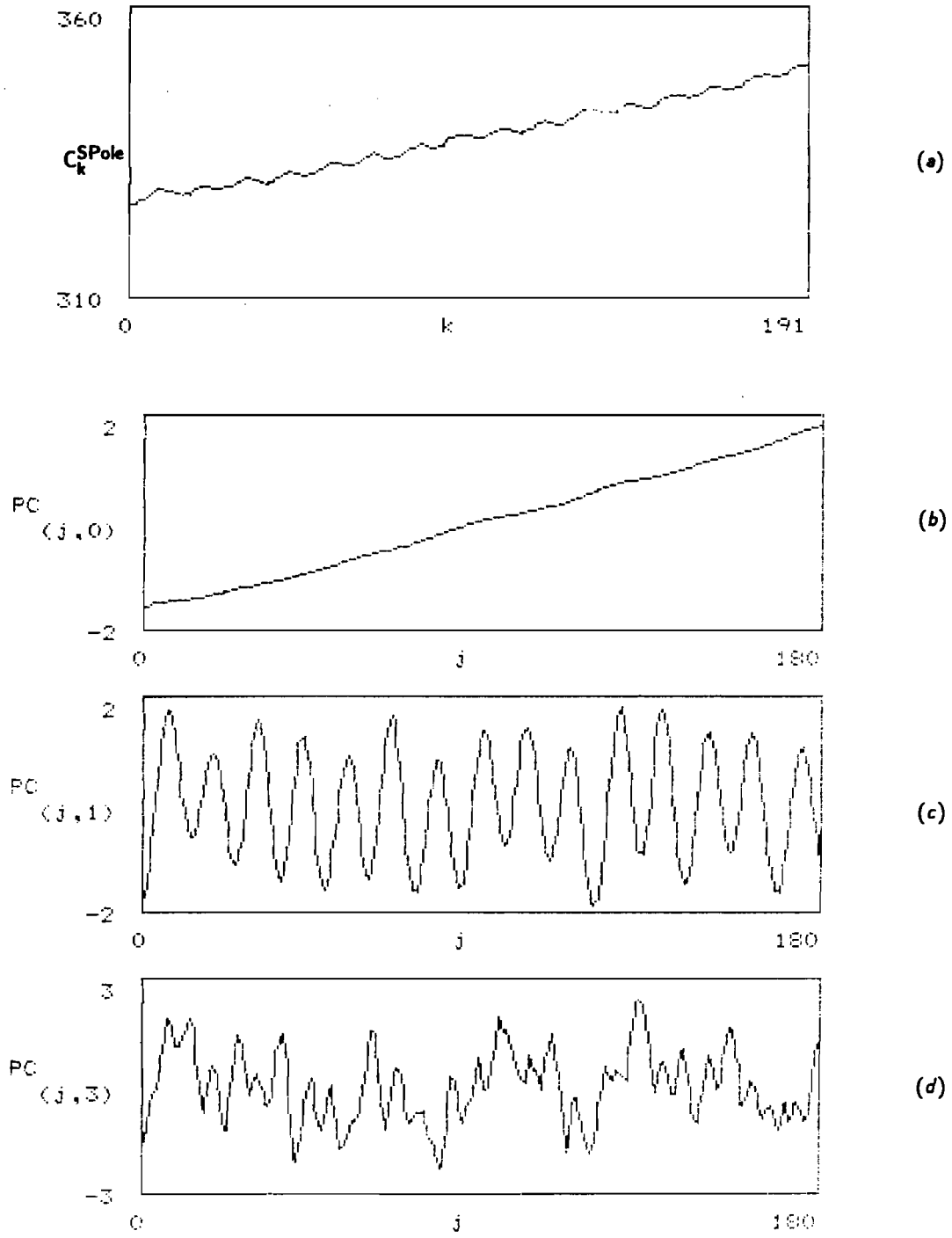


Figure 11. (a) Time series of atmospheric CO_2 concentration on the South Pole; (b), (c), (d), and (e) as on *Figure 8*.

(e)

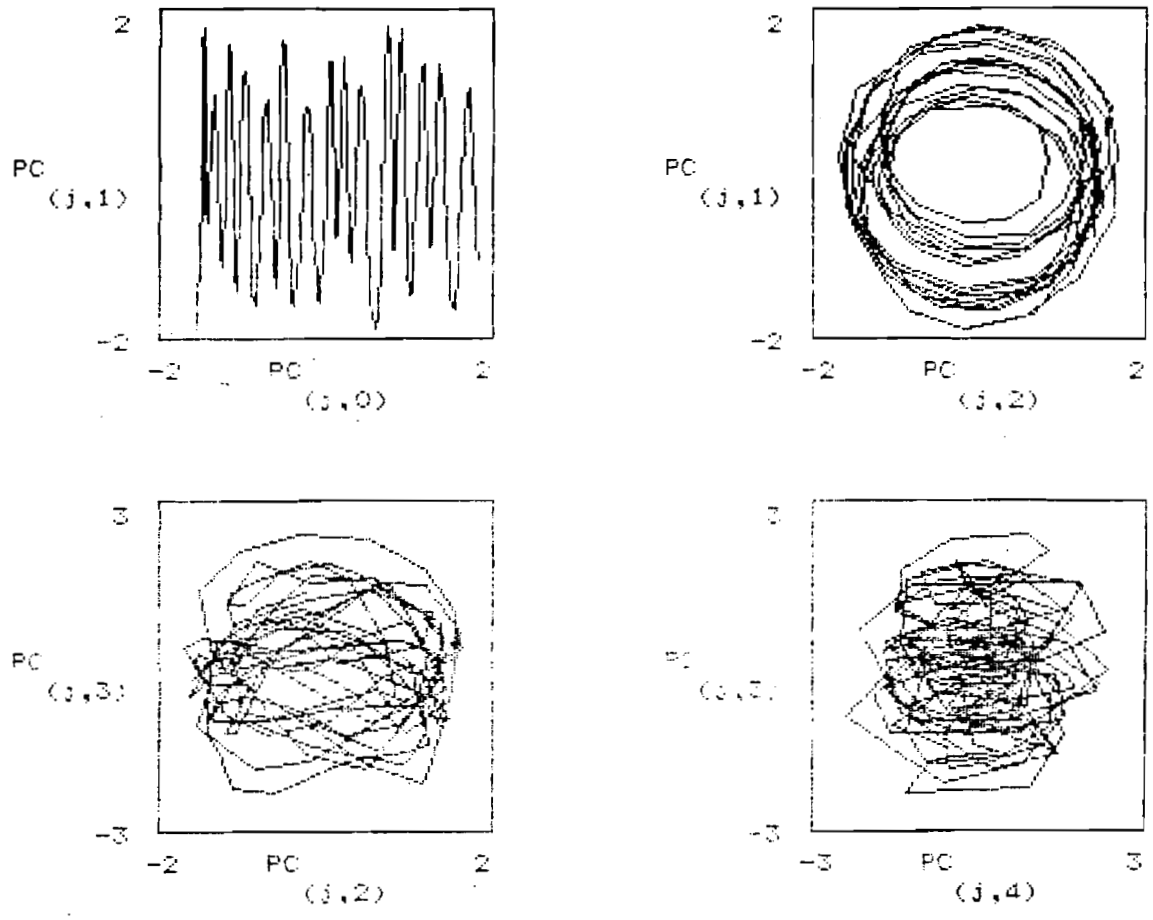


Figure 11. Continued.

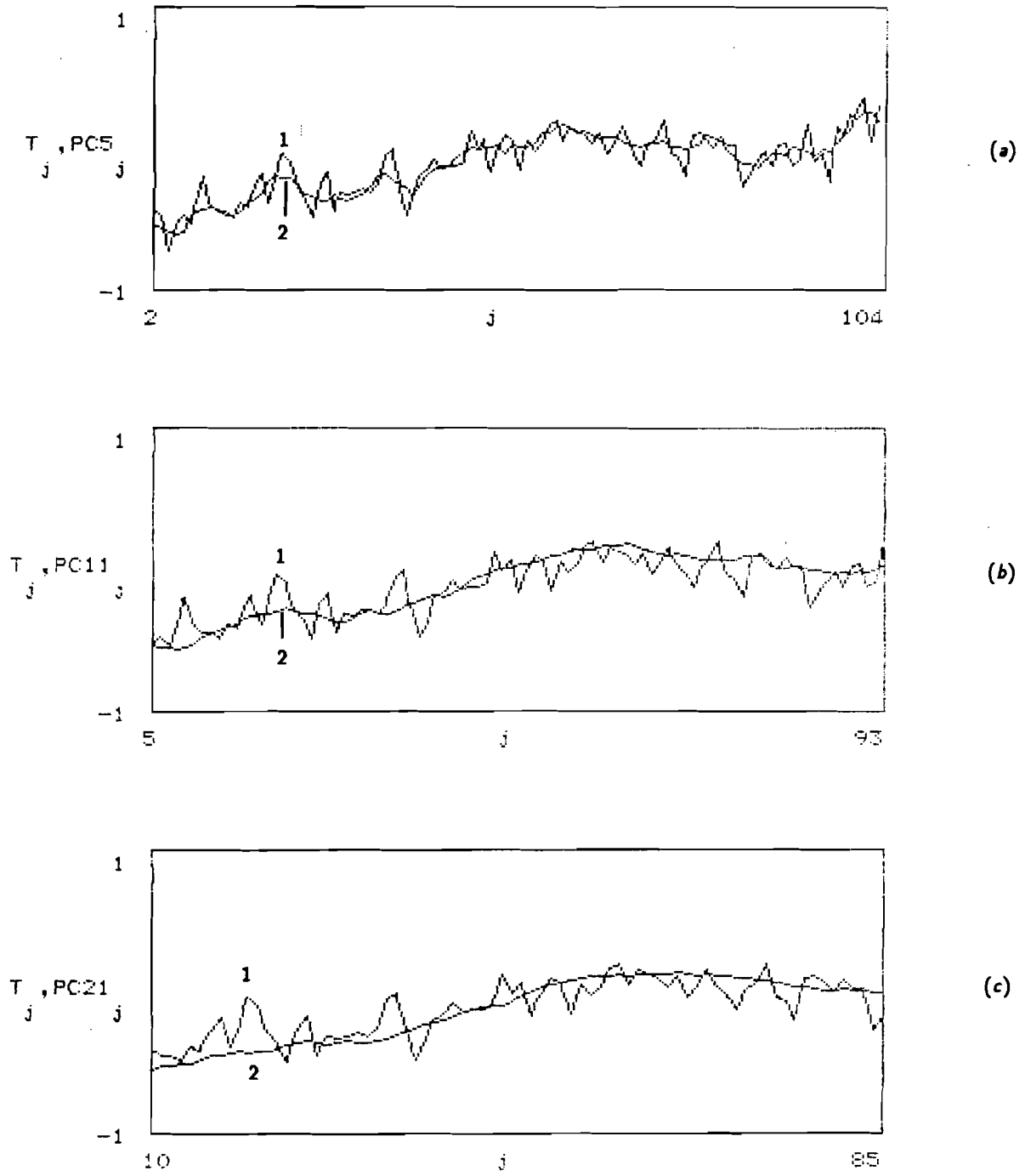


Figure 12. Comparison of the initial time series of global temperature anomaly (reduced to the normalized form) (1) with the projection on the first principal component (2); (*a*) $n = 5$; (*b*) $*n = 11$; (*c*) $n = 21$.

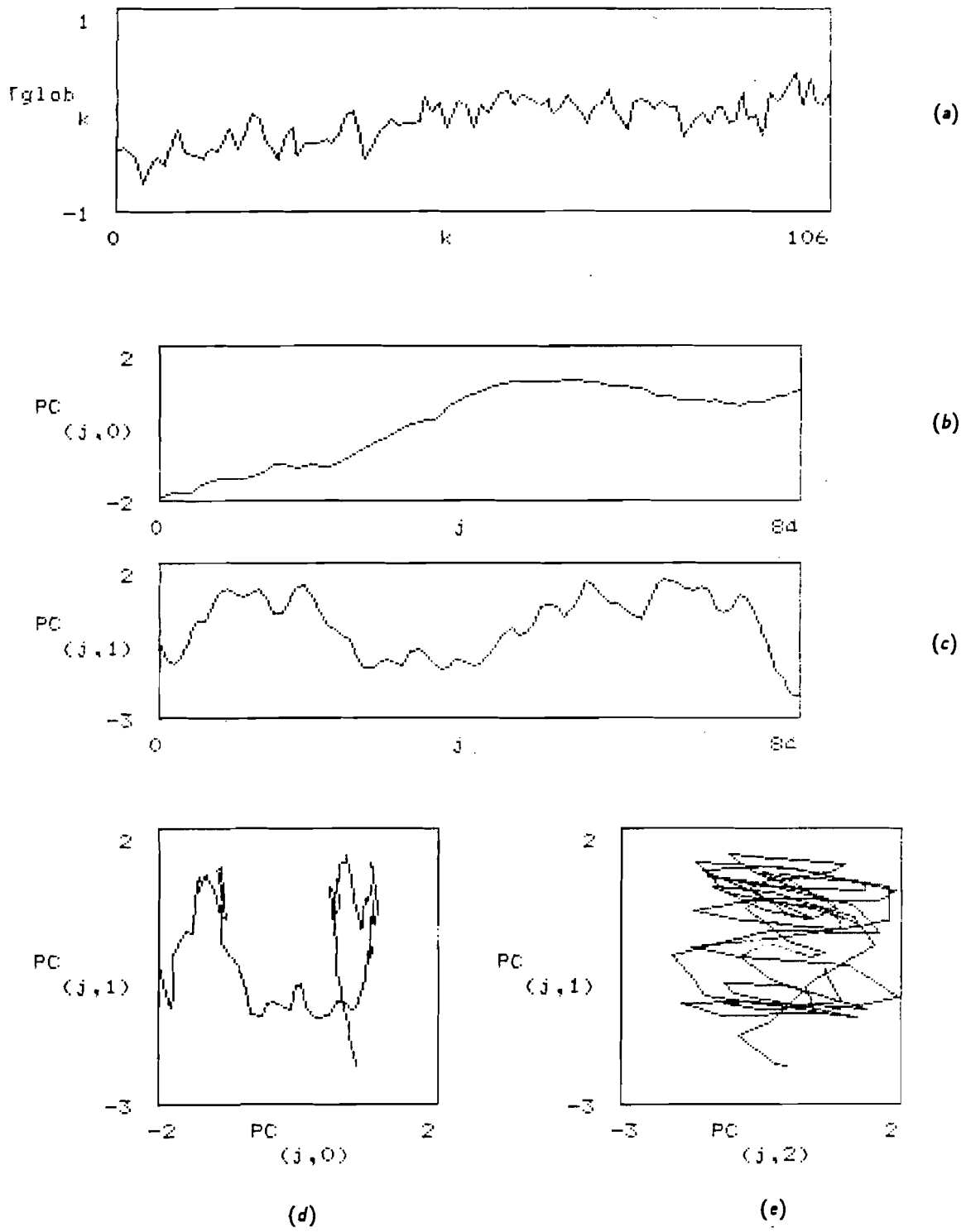


Figure 13. (a) Time series of the global temperature anomalies; and its projection on the (b) first principal component; (c) second principal component; (d) and (e) planes of the pairs of principal components.

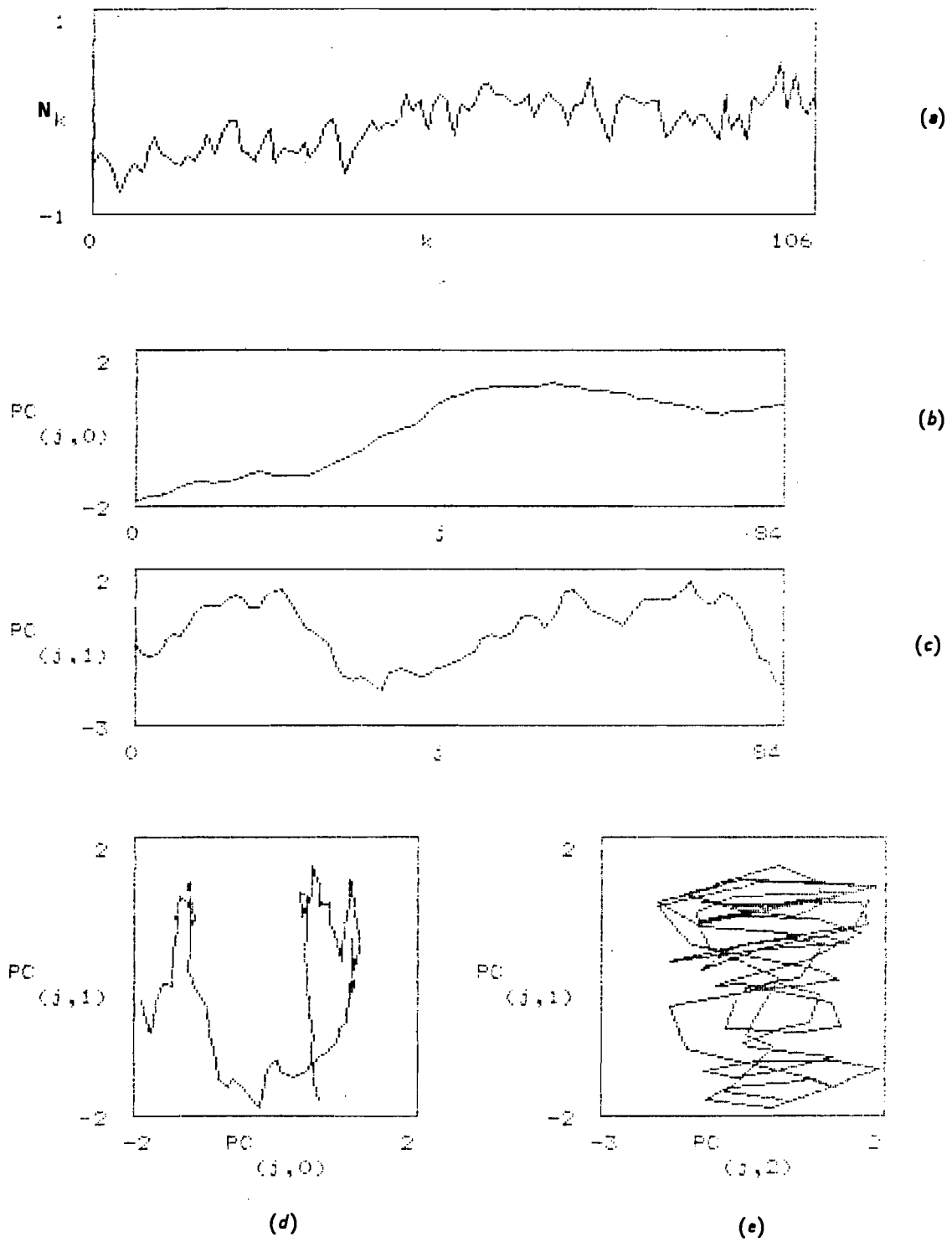


Figure 14. As on *Figure 13*, but for a series of the temperature anomalies in the Northern hemisphere.

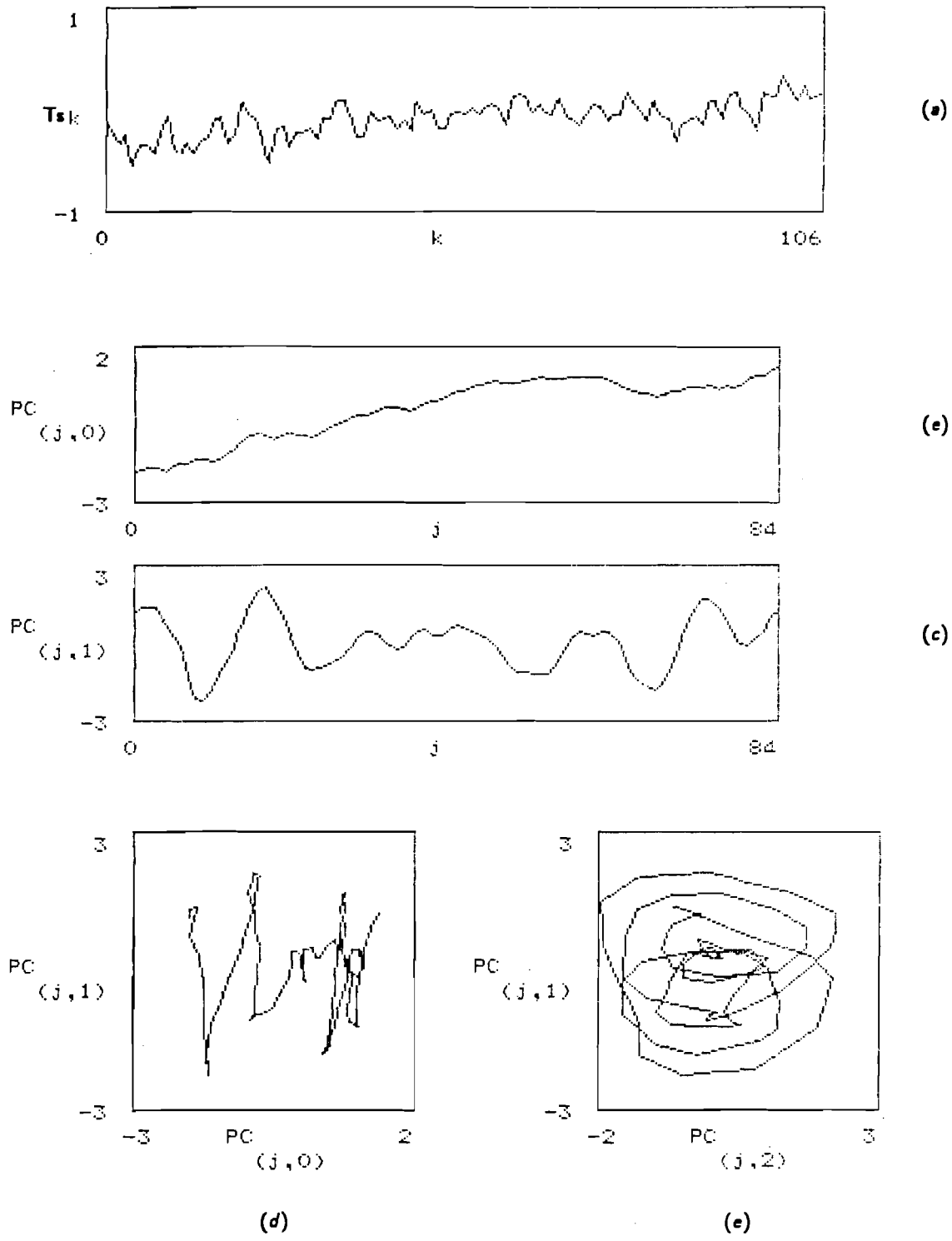


Figure 15. As on *Figure 13*, but for time series of the temperature anomalies in the Southern hemisphere.

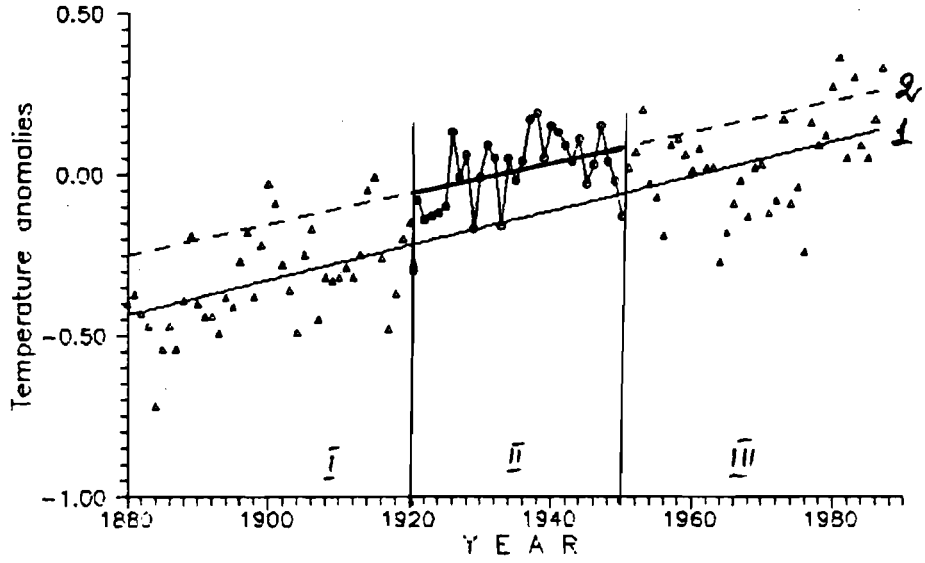


Figure 16. Comparison of the differences in the structure of time series of the global temperature anomalies: (1) the line of approximation for time intervals I and III; (2) the line of approximation for time interval II.

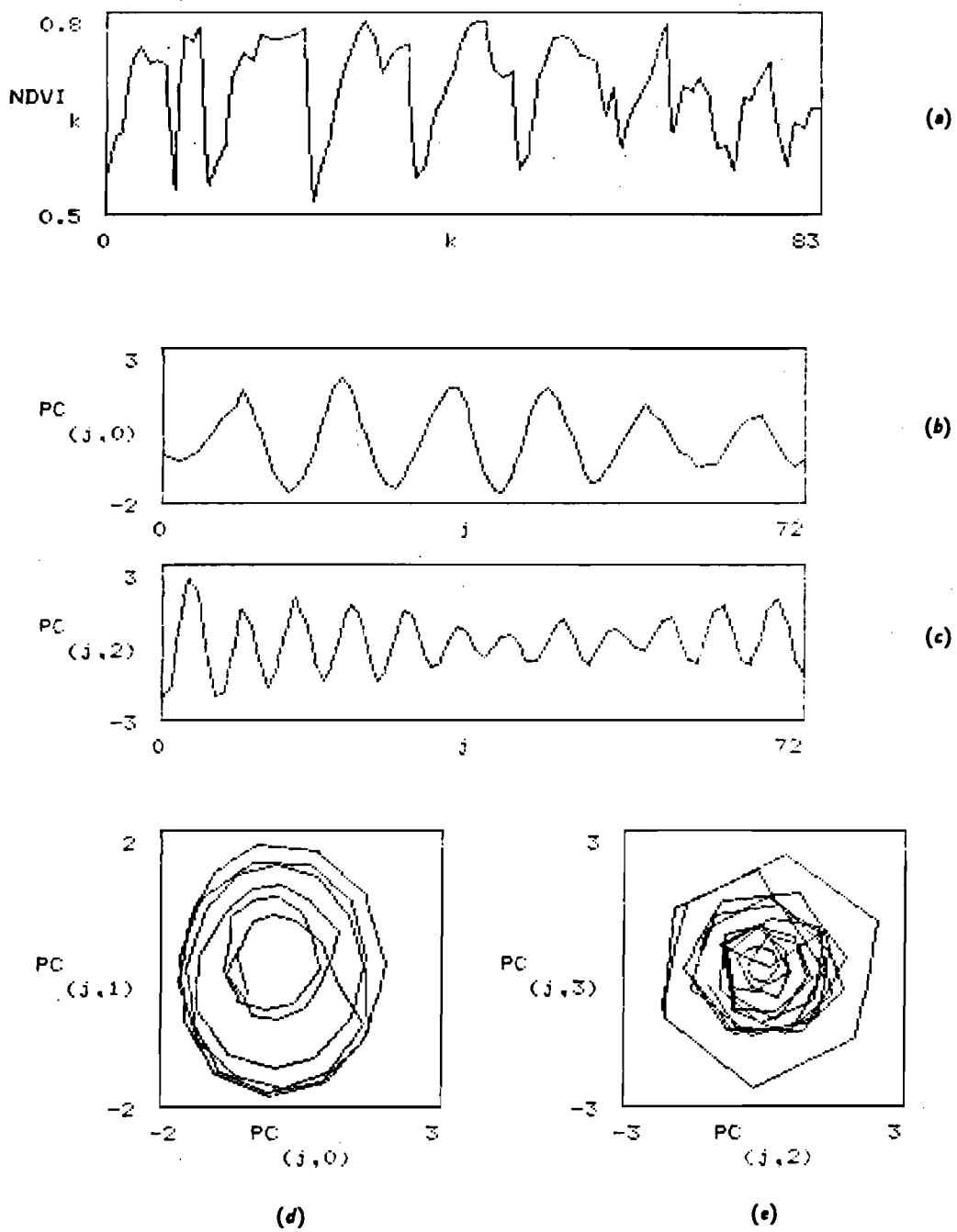


Figure 17. (a) NDVI curve for grass savanna (site 5); and its projection on the (b) first principal component; (c) third principal component; (d) and (e) planes of the first and the second pairs of the principal factors.

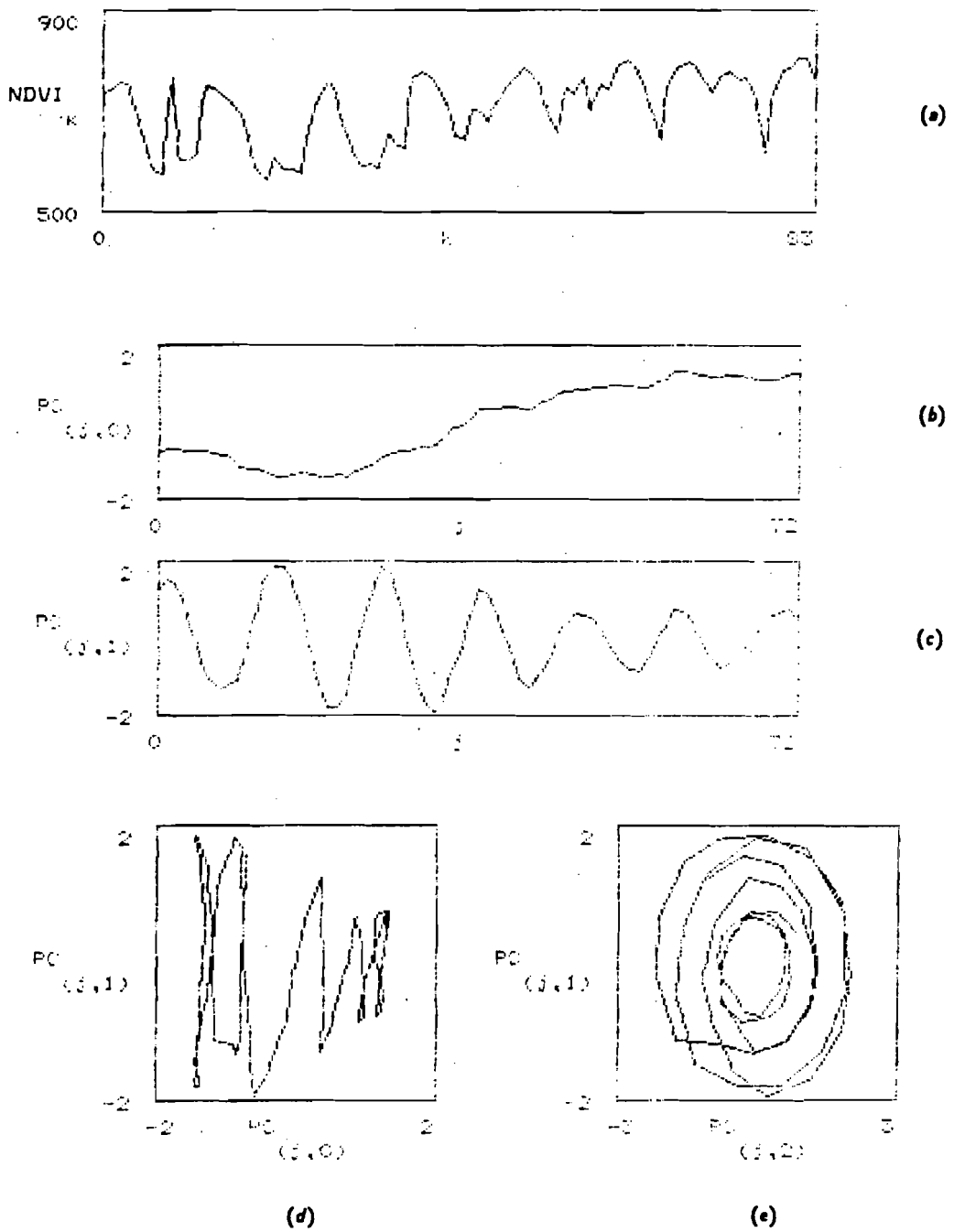


Figure 18. (a) NDVI curve for transition zone (Site 37); and its projection on the (b) first principal component; (c) second principal component; (d) plane of the first and second principal components; and (e) on the plane of the second and third principal components.

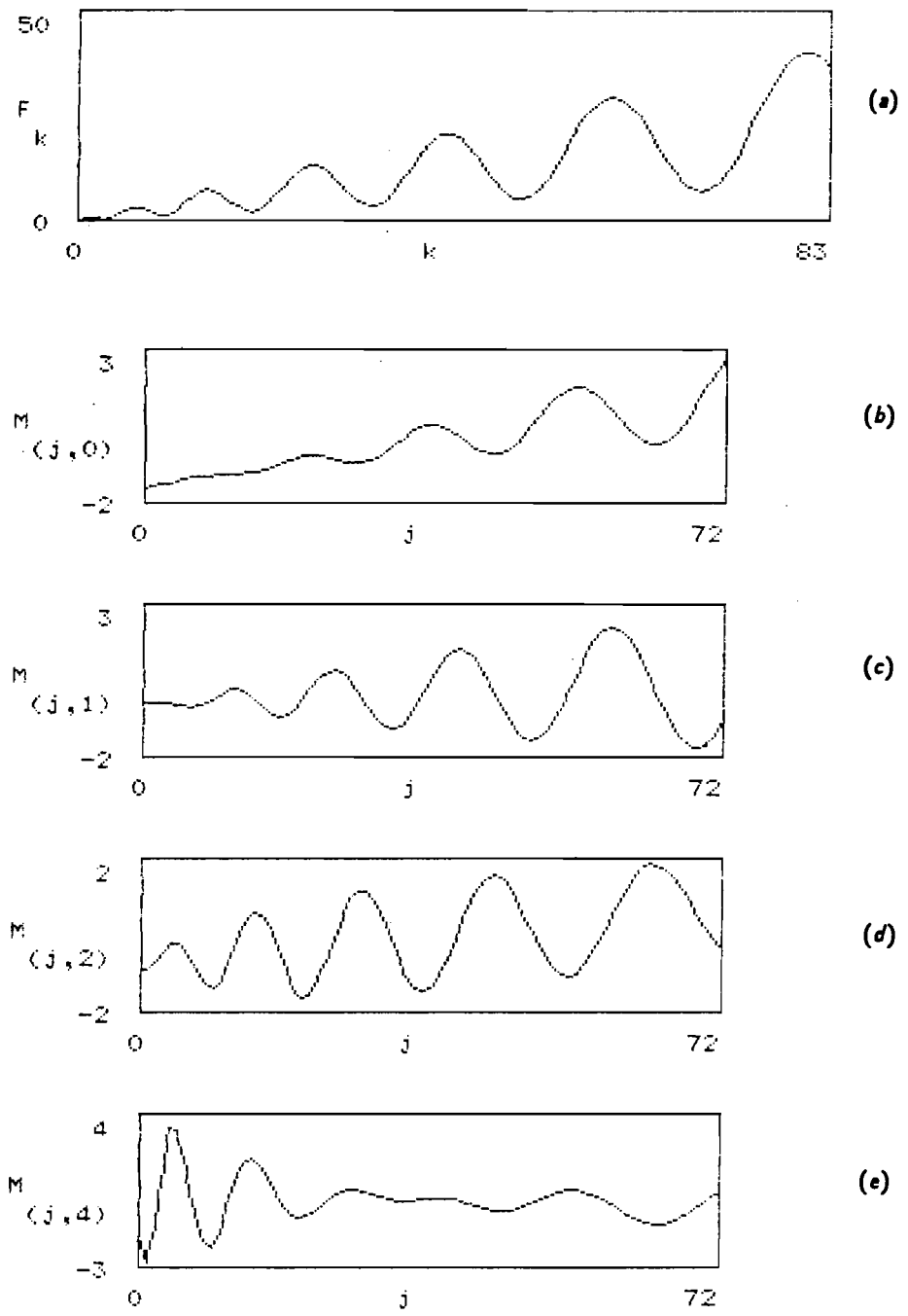


Figure 19. Application of classical shift operator $A_1 = T_i^{\tau} f(t) = f(t+\tau)$ to the function

$$f(k) = N_k = a_1 K + (a_2 + a_3 k) i \cos(ay\sqrt{k})$$

(a) initial signal (b), (c), (d) projection on the first, second, third and fifth principal components correspondingly, (f) projection on the plane of pair of the principal components.

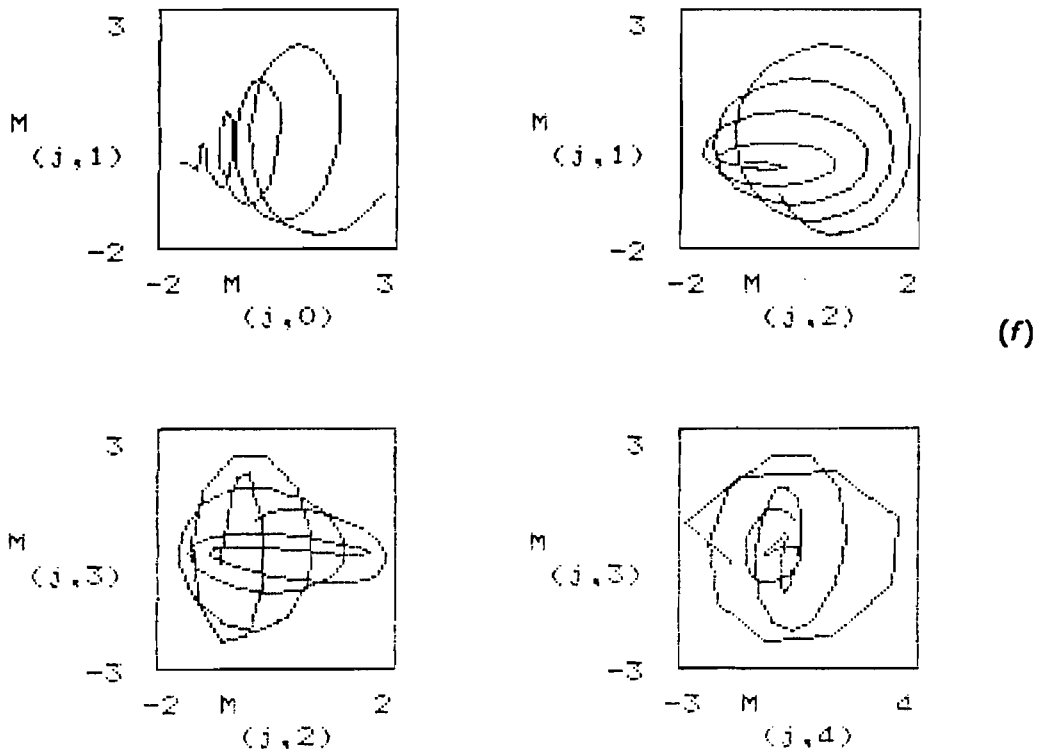


Figure 19. Continued.

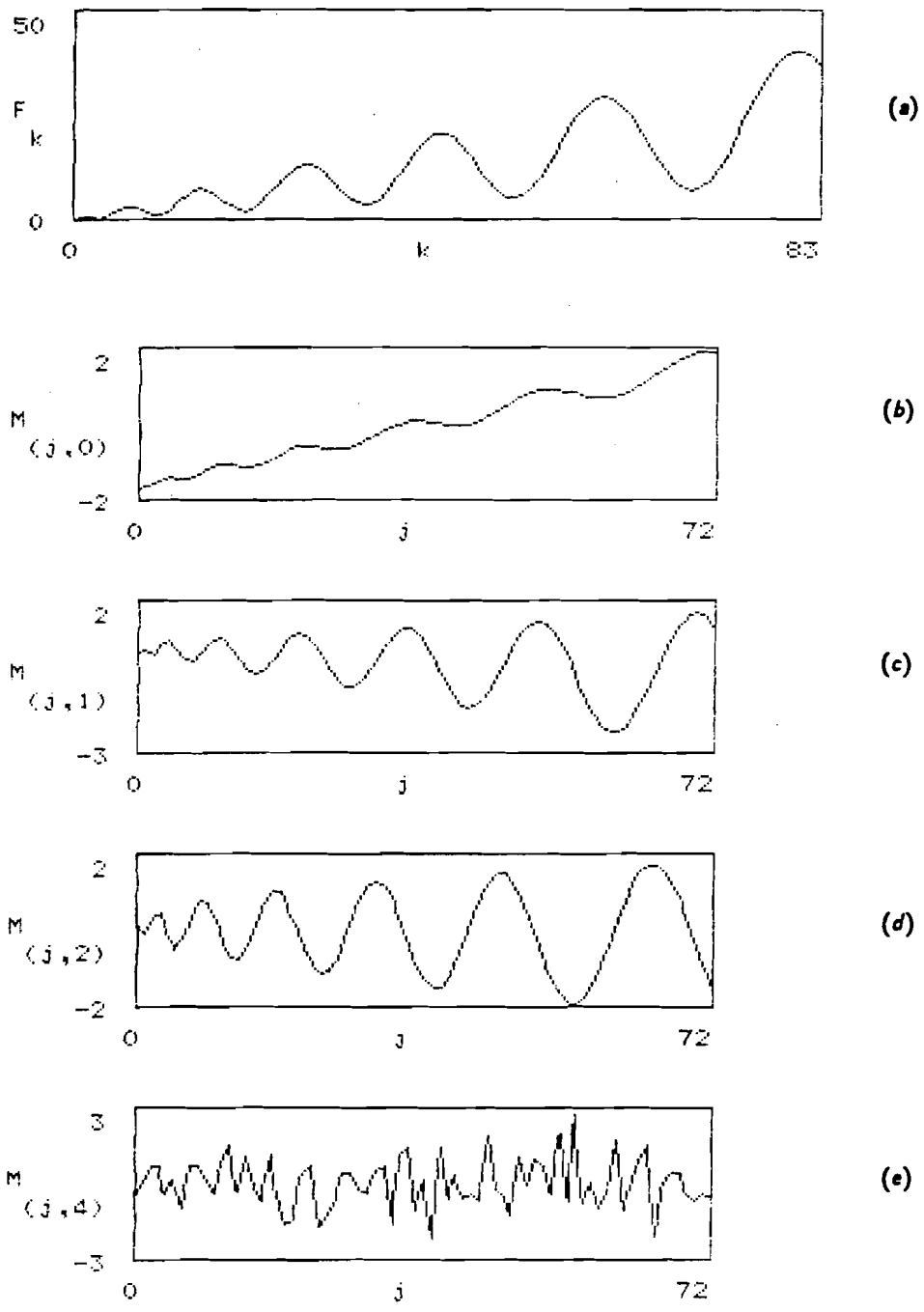
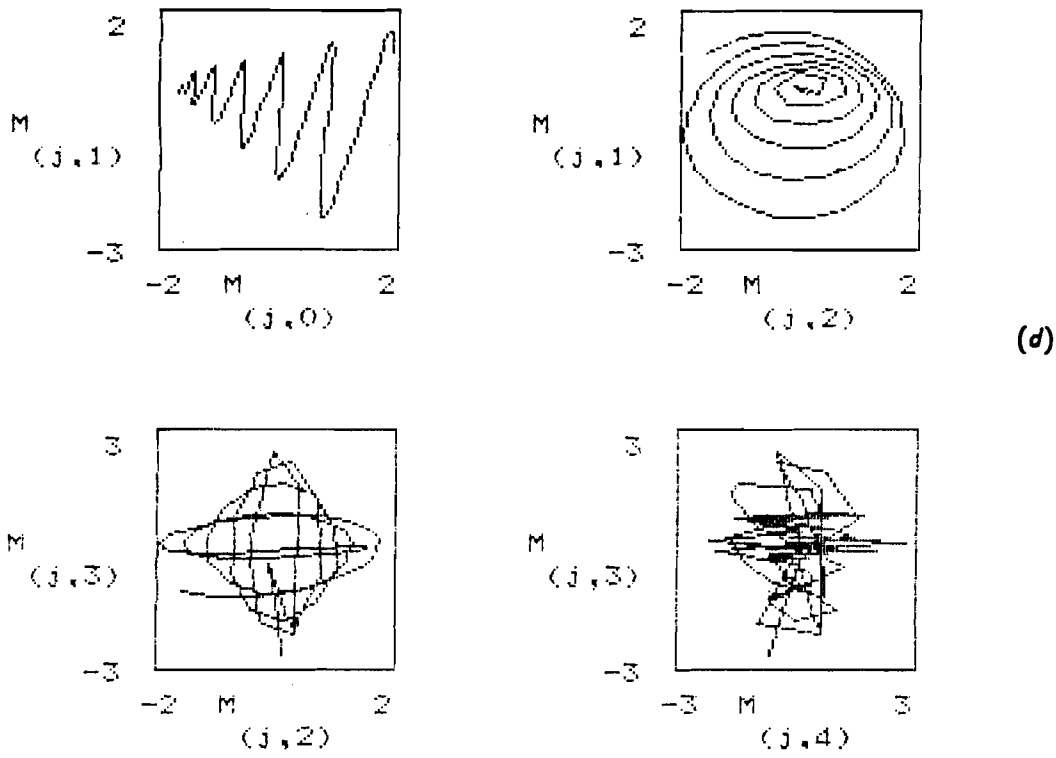


Figure 20. Application of the shift operator

$$A_2 = T_t^\tau f(t) = \frac{1}{2}\left(f((\sqrt{t}+\sqrt{\tau})^2) + f((\sqrt{t}-\sqrt{\tau})^2)\right) \ .$$

Initial function $f(k) = N_k$ and notations for (a) - (f) as on *Figure 19*.



(d)

Figure 20. Continued.

Appendix 1

In this part of the paper we considerably use the results of the research report by J.P. Malingreau, M. Antonovski, V. Buchstaber, and L. Veksler, *A Statistical Analysis of Time Series of Satellite Data Related to Tropical Vegetation*, compiled jointly with the Institute for Remote Sensing Applications of the Joint Research Center (JRC) of the CEC (Ispra, Italy), the IIASA Environmental Monitoring Project (Laxenburg, Austria), and the National Scientific and Research Institute for Physical–Technical and Radio–Technical Measurements (Mendeleevo, USSR) (forthcoming).

The data set used in this work is being developed within the framework of a NASA–JRC scientific agreement. It consists of a selected sample of the time series of the vegetation index data derived from the NOAA–AVHRR GAC data (so-called 8 km product) for 40 sites covering the tropical and subtropical ecosystems of Africa (see *Map 1*).

The data are monthly maximum vegetation index data values obtained using the standard maximization procedure. The period between January 1982 to January 1988 is covered (84 time-related measurements).

The vegetation index is taken as representing an aggregated measurement of green biomass activity in the selected ecosystems (*Table 1*). This research is based upon existing work on the significance of this index and is not specifically addressed to the problem itself.

Instead, work focuses on the temporal structure of the data upon differences/similarities between ecosystems represented in the sample.

Let us remember that vegetation indices are combinations of reflectance values in discrete spectral bands each of which exploit a particular characteristic of the plant canopy. The differential response of plant canopies to red (where the response is mainly determined by the absorption band of the leaf chlorophyll) and near-infrared NIR (where the response is the result of a multiple scattering determined by internal leaf structure and the structure of the canopy) illumination provides the base for calculation the "normalized difference vegetation index" which is simply a linear combination of the reflectances associated with each wave-length according to the following formula

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} - \mathrm{Red}} \ .$$

Instantaneous values of the vegetation index are therefore more related to the green biomass and its structural arrangement at the surface than to the type of vegetation. However, the temporal evolution of this index will closely follow the vegetation seasonal cycles. The analysis of the time

THE LOCATION OF THE SITES
10°N
0°
10°S
30°E
20°E
10°E
0°
10°W
River Niger
River Congo
○ Rain Forest
△ Savanna
● Transition zone and unknown ecosystems

Map 1.

Table 1: List of sites. (Coordinates and short description of the ecosystems for AVHRR time series).

No. of sites	Name of sites by Malingreau	Lat.	Long.	Short description
1	nzenth	8.18	− 8.16	Mixed secondary forest with plantations
2	nzesth	7.65	− 8.78	Idem
3	nzemix	7.99	− 8.61	Idem
4	libfor	6.89	− 9.79	Tropical rain forest
5	beyla	8.67	− 8.62	Grass savanna
6	burn	9.01	− 8.99	Shrub savanna often burned
7	kan	10.31	− 9.13	Mixed tree savanna – agriculture
8	for	6.57	− 2.16	Tropical rain forest
9	kissi	9.69	−10.42	Tree to shrub savanna
10	mac	8.40	− 9.30	Secondary forest with primary remnants
11	congo1	3.16	21.71	Primary rain forest between Congo and Ubangui
12	congo2	0.96	21.94	Primary rain forest, Central Congo basin
13	congo3	− 1.15	19.69	Primary rain forest (marsh) Central Congo
14	congo4	− 9.5	24.02	Primary rain forest Central Congo basin
15	congo5	1.92	27.47	Primary rain forest, eastern edge of basin
16	congo6	4.12	21.16	Primary rain forest northern edge of rain forest domain
17	congo7	4.15	19.94	Transition area – forest woodland savanna
18	congo8	4.67	19.24	Idem slightly further north toward savanna
19	congo9	3.98	16.62	Tropical forest savanna transition area. Edge of forest.
20	congo10	4.88	17.38	Woodland savanna ("Guinean savanna")
21	congo11	5.57	16.28	Woodland savanna
22	congo12	5.16	13.52	Woodland savanna
23	congo13	5.22	11.00	Tropical rain forest, western edge of Central Africa
24	congo14	4.95	14.83	Forest remnant in savanna region
25	congo15	7.87	19.86	Woodland savanna
26	congo16	7.28	22.14	Woodland savanna, northernmost example
27	congo17	6.26	21.38	Forest remnant in transition area (seasonal forest?)
28	congo18	5.29	23.09	Forest remnant in transition area (seasonal forest?)
29	congo19	5.85	25.46	?
30	congo20	− 3.49	15.04	Savanna or deforested area in Congo Republic (North)
31	congo21	− 1.38	14.27	Savanna or cleared areas in Congo Republic (South)
32	congo22	1.58	18.54	Rain forest – marsh? Between Congo and Ubangui rivers
33	congo23	0.07	18.59	Rain forest near Mbandaka Zaire
34	congo24	− 3.30	17.31	Forest (savanna?), west of Inongo Lake
35	congo25	− 4.73	24.67	Savanna at southern edge of Congo basin forest (South of Kindu)
36	congo26	− 3.28	24.17	Rain forest at its southern range in Basin (North of Kindu)
37	congo27	− 5.69	26.4	Forest east of Kindu
38	congo28	− 2.87	27.31	Kivu transition towards semi-tropical highlands
39	congo29	14.0	25.88	Forest gallery (secondary formation?) near Kisangani
40	congo30	1.00	36.61	Primary rain forest near Kisangani

series can, therefore, lead to the identification of the vegetation type and of events affecting its development (drought, stress, etc.).

It is easy to be sure, that NDVI time series is a characteristics of the researched ecosystem. On *Figure 1a* and *2a* are given NDVI curves for ecosystem displaced in the same geographical region. On *Figure 1a* are given curves for sites with a similar character of vegetation, and on *Figure 2a* – with sharply different characteristic. In the first case we have almost complete coincidences of the curves and in the second case transparent difference.

In a preliminary exploration, sample points have been selected over ecosystems of West and Central Africa on the base of an *a priori* knowledge of the nature of their ecosystems. The intention is indeed to assemble a collection of temporal curves representing as much as possible the range of ecosystems from the savanna to the tropical rain forest. The map gives the location of the samples. Time series were constructed by reading monthly vegetation index values (Malingreau, 1986). Multidimensional time-series analysis can now be combined with multidimensional statistical analysis. The problem in such an investigation is:

1. To separate noise from useful information.
2. To construct a "standard" time-dependent NDVI development curve for each ecosystem.
3. To develop criteria of comparison between the actual and the standard curves.
4. To develop mathematical and statistical models of NDVI curves.
5. To analyze the possibility of using the models to predict the dynamics of NDVI under different climate change scenarios.

The following steps have been taken into account in the present analysis:

1. drawing of a statistical picture,
2. component and factor analysis,
3. histogram analysis,
4. cluster analysis.

The sequence of the step has a principal value. It reflects the following logic of the research: discovering of the statistical valid (meaning) values regularities in the data; working out of informative description of these regularities; construction of the standards of the behavior of the year realization and finding the ecosystems, in which during the period of the observation was happening structural reconstruction; construction of the features of the year realization of the

NDVI curves for which an ordering by the scale of the values of these features according to the qualitative information of the biomass activity; hierarchical classification of the set of the NDVI curves for automatic recreation of the structure of the interrelation between the ecosystem; analyze the fact of how close this structure corresponds to the structure of the transition from savanna to tropical forest from the point of view of the changing of the integral characteristics of biomass activity.

Development of the "statistical picture" of the NDVI curves. For each sample, the seven 12-dimensional vectors of NDVI monthly development are drawn together. The visual examination of the set of curves gives a "picture" of the overall variability in the 12-month realizations during the seven years on record. Pattern recognition approaches (Fu, 1984) can be applied to derive the characteristics of the samples.

The analysis of statistical picture of 40 ecosystems (see, for example, Site 1-29 in Appendix 2) show that the majority of them like a statistical picture on *Figure 4*. It confirms the initial hypothesis about the existence of standard behavior of year realization of NDVI curves. A consideration of the character of deviations of concrete realizations from hypothetical standard permits us to see that there exists large deviations, as a rule, in the direction of smaller values of the NDVI. It is in accordance with physical information, that the factors caused the errors, for example, cloudiness, in the NDVI data, usually tend to the smaller values than the real values. The selection of a method of evaluation of standard behavior is independent task. For its solution it demands a large volume of information. In the present article we choose, as evaluation of standard behavior, median assessment (see *Figure 10*). It means that evaluation of standard behavior in a given month is a median of the series of 7 values of NDVI observing in this month during all periods of observation. The main advantage of this estimation compared with estimation by means is that median is more stable as related to large outlier. In support of median also says that using median in a subsequent analysis (see factor and hystogram analysis) leads to physical result.

Let us remark that construction of the standard of year behavior for each NDVI curve gives not only compact informative description of this curve in the case when initial hypothesis is confirmed, but also permits us to find the structural changes in it when the hypothesis is not confirmed clearly. For demonstration such potential possibilities of the method of statistical picture lets us consider more in detail typical examples of the statistical pictures.

Let us compare, for example, the NDVI development curve for the grass savanna (Site 5) and its statistical picture *Figure 3*. The advantage of the latter is clear as it underlines the closeness of the NDVI values obtained during the seven years for the January to June period while the July to December interval is characterized by a large variability and frequent "outliers" (see, for

example, August–September 1988); some of these outliers can be related to external factors such as cloudiness during the rainy season (i.e., September 1982, July 1987). The example indicates that there is less interannual variability in the greening-up period than during the period of maturation and senescence of vegetation. The statistical picture for the primary forest sample (Site 11, *Figure 4*) shows no clear seasonal differences between the patterns of interannual dispersion of NDVI. Of interest also are the transition ecosystem represented by the Guinean woodland savanna (Site 17, *Figure 5*) and the seasonal transition forest in the southern part of the Congo Basin. For these ecosystems, the semi-annual pattern is again visible with the first six months presenting little interannual differences. The pattern is, however, rather different for the woodland-transition forest sample selected in the southern hemisphere with the period of minimum interannual dispersion located in the yellowing down period of the vegetation cycle. The same sample shows that there is a large dispersion of vegetation index values during the second part of the year (greening up and maximum greenness periods).

Let us turn our attention to a large dispersion of the realization for ecosystem 37 at the second half of a year (*Figures 6* and *7*). It is seen, that for this ecosystem it is impossible to create presentation about the standard behavior of the NDVI curve in the second half year by seven realizations. It would mean that the hypothesis that all seven realizations for ecosystem 37 belong to the same process is not confirmed. Probably, during the period of observation it have happened a structural restructure of the year realization behavior. A result of a testing of this new hypothesis is seen on *Figure 7*. Seven realizations subdivided on three groups: the first three years of observation, the fourth year of observation, and the last three years of observation. Statistical pictures of the first and the third group on *Figure 7* show essentially more regular common behavior of realization of each from this group. Effect that we have detected put at an interpretator the question about the causes of structural restructure of the standard behavior of a year realization for ecosystem at the period from 1982 until 1988.

Joint analysis of statistical picture for ecosystems 5, 11, 17, and 37 tends to the conclusion that the following characteristic of a picture is important.

Let $f_i(m, y)$ $(1 \leq i \leq 40,\ 1 \leq m \leq 12,\ 1 \leq y \leq 7)$ be the value of NDVI for i-th ecosystem in m-th month of y-year of observation.

For each number i of ecosystem and number m of month we order the value $f_i(m, y)$ by increasing. We have

$$f_i(m, y_1) \leq f_i(m, y_2) \leq \cdots \leq f_i(m, y_7) \ .$$

So we construct a mapping

$$\varphi_i : M \to \sum(7) : \varphi_i(m) = (y_1, \ldots, y_7) \ ,$$

where $M = \{1, \ldots, 12\}, \sum(7)$ – the group of permutation of the set of 7 elements. For example from *Figures 4, 5, 6* and *7* we have:

$$\begin{aligned}
\varphi_5(1) &= (3,2,1,4,5,7,6), \ldots, \varphi_5(12) &= (6,7,5,4,2,1,3) \\
\varphi_{11}(1) &= (4,3,6,2,1,5,7), \ldots, \varphi_{11}(12) &= (6,5,7,2,4,1,3) \\
\varphi_{17}(1) &= (4,3,2,5,7,6,1), \ldots, \varphi_{17}(12) &= (6,5,7,4,1,2,3) \\
\varphi_{37}(1) &= (3,1,2,4,5,7,6), \ldots, \varphi_{37}(12) &= (2,1,3,5,6,4,7)
\end{aligned}$$

The function of permutation $\varphi_i(m)$ describes how much the values of a year realization have mixed one relative to others. And it gives tools for informative characteristics of statistical pictures. As one of the tools, it is naturally to take the decomposition of permutations on cycles. For example, for ecosystem 37 (*Figure 7*) we see that the first 3 years of observation gives a cycle of low values of NDVI and the following 4 years gives a cycle of high values. The NDVI curves for ecosystem 17 do not have this property.

Let us now describe the application of constructed function of permutation $\varphi_i(m)$. Let for some i-th ecosystem and the months $m_1 < m_2 < \cdots < m_k$, $k \geq 6$ in all permutations $\varphi_i(m_j)$, $1 \leq j \leq k$ there exists the same cycle of a length not less than 3. It permits us to form a hypothesis that in the years entered into this cycle, a peculiar structure of the year realization was formed. It is possible to test this hypothesis, constructing a statistical picture of realizations for these years. If in such a picture the joint behavior of realization entered in cycle differs from the behavior for all the intervals of observation, as in the example of ecosystem 37, then our hypothesis is not confirmed.

Component and Factor Analysis

Let us denote by $f_i(t)$, $t = 1, \ldots, 84$ the values of NDVI for the i-th ecosystem. In the previous notation $f_i(t) = f_i(y, m)$, where $t = 12(y-1) + m$, $y = 1, \ldots, 7$, $m = 1, \ldots, 12$. Let us put $f_i = [f_i(t)]$ and $f_{iy} = [f_i(y, m)]$.

The initial data set gives us sample $\mathcal{F}$ from 40 vectors in the Euclidian space R^{84} and sample F from 280 vectors in the space R^{12}. Such a description of NDVI curves permits us to involve methods of component and factor analysis for the research (Aivasian *et al.*, 1989; Harman, 1972).

For the convenience of the account, we give the main models of these methods, using the approach of the theory of projection pursuit (see citation on §1 of the main text).

Let O_n be the object under investigation, which is described by the vector $x_n \in R^p$, and hence the set of data about the objects $O_1, \ldots, O_n$ is described by the sample of vectors $X = \{x_1, \ldots, x_n\} \subset R^p$.

In our case, the O_n is or the state of ecosystem for all period of observation 1982–1988 either it state for one of the year of observation. The x_n is or f_i for $n = i$, either f_{iy} for $n = 7(i-1)+y$, i.e., $N = 40$ or 280, and $p = 84$ or 12.

Initial features of the objects O_n are coordinates of vector x_n. In our case these are the values of NDVI $f_i(t)$ or $f_i(y, m)$. First of all, we are interested only in such initial features (characteristics) that show the greatest changeability (variability) by transition from one object to another. For the other side, it is not necessary to use the initial features for the description of the state of object. For example, in the paper of Malingreau (1986), in particular, the following features of vector $f_i(t)$ of NDVI curve of the i-th ecosystem are used:

$$\int_{t_1}^{t_2} f_i(t) = \sum_{t=t_1}^{t_2} f_i(t) \ ; \quad tg f_i(t)_{t_1}^{t_2} = \frac{f_i(t_2) - f_i(t_1)}{t_2 - t_1} \ .$$

In the models of component analysis we are interested, first of all, in such initial features that show the most alteration (i.e., the most scatter), passing from one object to another. Each linear feature of object O_n, described by vectors of R^p, could be given by the formula

$$l(O_n) = \langle l, x_n \rangle \ ,$$

where

$$l = (l^1, l^2, \ldots, l^p) \in R^p \ \text{ and } \ \langle l, x_n \rangle = \sum_i l^i x_n^i$$

is a scalar product. Dispersion of the values of linear of linear indication l on the sample $X = [x_1, \ldots, x_n]$ is calculated by the formula:

$$S_l(x) = \frac{1}{\|l\|^2} \sum_{n=1}^{N} \langle l, x_n - \overline{x} \rangle^2 \ ,$$

where

$$\|l\|^2 = \sum_{i=1}^{p} (l^i)^2$$

is a norm of vector l, and

$$\overline{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

is a mean vector of the sample X.

Dispersion of the values of feature l on X becomes a measure of informativeness of the feature. Linear features with maximal dispersion on the sample X is called the first main component. It is clear from the definition that the first main component is given by the vector

$$l_1 = \arg\ \max \{S_l(X) : l \in R^p, \|l\| = 1\} \ .$$

The following main components are defined by induction. Let the first "k" main components $l_1, \ldots, l_k, 1 \leq k \leq p$ be constructed, then the $(k+1)$-th main component is the indication given by vector

$$l_{k+1} = \arg\ \max \{S_l(x) : l \in R^p, \|l\| = 1 \ , \ \langle l, l_j \rangle = 0 \ , \ j = 1, \ldots, k\} \ .$$

Thus the main components are the features that successively (in consecutive order) in the best way explain (reconstruct) the dispersion (variance) of a sample X. The model of classical factor analysis is analogous. The main factors are the features (characteristics) $\varphi_1, \ldots, \varphi_p$ that successively explain (reconstruct) correlation relationships (correlation matrix) of sample X. Namely, the informativeness of linear feature l on the sample $X = \{x_1, \ldots, x_n\}$ is calculated by the formula:

$$S_l(x) = \frac{1}{\|l\|^2} \sum_{n=1}^{N} \langle l, y_n \rangle^2$$

where

$$y_n = \left(y_n^1, \ldots, y_n^p\right) \quad \text{and} \quad y_n^k = \frac{x_n^k - \overline{x}^n}{\sigma^k} \ ,$$

and

$$\overline{x}^k = \frac{1}{N} \sum_{i=1}^{N} x_i^k \quad \text{and} \quad \left(\sigma^k\right)^2 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i^k - \overline{x}^k\right)^2$$

are a mean and dispersion, respectively. The transition from x^k to y^k is called standardization of the variables. Hence the factor analysis is a component analysis in standardized variables. Let us note that to variable x^k and to any linear transformation thereof, $x_L^k = A^k x^k + b^k$ corresponds to the same standardized variable y^k.

Hence the results of factor analysis are the same for the NDVI curves and PV curves that are connected by the formula

$$\text{NDVI} = 0.0019(\text{PV}) - 1 \ .$$

Let us now describe the results of applying the methods of analysis discussed above to the NDVI curves. The sample $F \subset R^{12}$ is the union of 40 subsample $\{f_{ij}, j = 1, \ldots, 7\}$. On the basis of

the analysis of a statistical picture of NDVI curves of the ecosystems we choose as standard year realization of NDVI for the i-th ecosystem the vector $M(f_i)(m) \in R^{12}$, where

$$M\left(f_i(m)\right) = \text{mediana } \{f_{i1}(m), f_{i2}(m), \ldots, f_{i7}(m)\} \quad .$$

So we obtain a sample M from 40 vectors $\{M(f_i) \in R^{12},\ i = 1, \ldots, 40\}$. Let us make a factor analysis of this sample project it on the plane of the first two main factors, i.e., let us describe each ecosystem by two main factors of standard year realization of its NDVI curve that is chosen by us. In *Figure 8* we see as very definitively separated open circle points (forest) and triangle points (savanna); between them exist closed circle points (transition zones and unknown ecosystems). It is possible to say that factor analysis we have passed gave the result which has physical interpretation.

It is spring up the following task: to reconstruct the meaning of the factor 1 and 2, i.e., to show what properties of the NDVI curves they characterize. As we have noted above, the factors φ_1 and φ_2 are vectors in the space R^{12}. The initial basis in this space are features characterizing values of NDVI by months; the first vector of basis corresponds to January, the second vector corresponds to February, and so on. Thus, we have (in the space R^{12}) two-dimensional plane stretched on two main factor and 12 points – the initial basis vectors, which we enumerate according to the order of numbers of the month. Let us project these 12 points on the plane main factors (see *Figure 9b*). From *Figure 9b* it is seen that the points-months are grouping: the first group (1,2,3,4), the second (6,7,8,10,11,12), and the third is (9,5). The projections of the first group on the axe of factor 2 and projection of the second group on the axe of factor 1 give approximate equal values. So we get that the first factor have a good approximation by mean value of NDVI for the months 6,7,8,10,11,12, and the second factor have a good approximation by mean value for the months 1,2,3, and 4.

Let us note that in a paper by Malingreau (1986) it was shown that the informative characteristics of the NDVI curves are the indications of sum activity by a half years (six months). Thus, the factor analysis, made by us, permit us to explain the result of Malingreau and to improve it by the criterion of pattern recognition of ecotype. This criterion has the following interpretation, the most information characteristics are the mean values for the periods of the biggest and the smallest vegetation activity:

$$l_1^*(f_{iy}) \approx \frac{1}{y} \sum_{m=1} y\ f_i(y, m)\ ,\ \ l_2^*(f_{iy}) \approx \frac{1}{6} \sum_{\substack{m=6 \\ m \neq y}} f_i(y_1, m)\ ,$$

where f_{iy} is a realization of NDVI curve in y-year of observation for i-th ecosystem, $y = 1, \ldots, 7$, $i = 1, \ldots, 40$. Relatively good informativeness of means by half year is explained that this is the appropriate period of time.

Hystogram Analysis

Visual analysis of NDVI curves and their statistical pictures permits us to pick up the characteristics that help to express common features for proxime ecosystems and particular not close. Such characteristics in our case are the following: the positions of minimum and maximum and their values for different evaluations of the standards of the year realizations, mean value of NDVI, and also various evaluations of a dispersion of process, presented by 7-year realizations. The list of such characteristics could be extended.

For each chosen characteristic were constructed hystograms. On the basis of their comparison there were selected the most informative characteristics (features).

Let us consider several examples, how informativeness of the features is evaluated by the hystograms. *Figure 11a* gives the hystogram of feature (max-min) of median evaluation of year behavior. This characteristic could be interpretated as a measure of vegetation activity during the year for each given ecosystem. It is seen from *Figure 11*, that in the domain of lesser values of the hystogram are grouped the ecosystem, corresponding to the tropical rain forest, and in the domain of greater values of scope – (max-min) – grass savanna. Moreover, let us look carefully at ecosystem 17 and 18. In the initial description it is said that both ecosystems are transient from forest to savanna, but ecosystem 18 is closer to savanna (slightly further north toward savanna). On hystogram these ecosystems disposed in the domain of intermediate values of indications, but ecosystem 18 moved to the side (domain) of greater values relative to ecosystem 17. So comparing mutual disposition of the ecosystems ordered sake for hystogram along the axis which correspond to (max-min) of the year evaluation of NDVI. With the initial description of the ecosystems, we have a result that agrees to a qualitative ordering by biomass activity, which is possible to estimate by means of description of ecosystems also.

For **another hystogram** (*Figure 11b*) as a classification feature was used mean value. This feature could be interpreted as an estimation of mean vegetation activity of the ecosystem during the whole period of observations. It is seen, that ecosystems of savanna type gravitate toward the domain of lesser value, and ecosystems of tropical rain forest type gravitate toward the domain of larger values. Comparing this ordering of ecosystems along the axis of mean values, we note, that it again, as a rule, coincides with initial description of ecosystem.

Let us stress that in the present research, we have dealt with well enough uniform sample contained a small (restricted) number of ecotypes with natural transitions between them. Analyzing the samples more extensively geographically and more diversely by ecotype from the article of Townshend *et al.* (1987), we see that ordering by each of these characteristics would not have **interpretation**, and then it is necessary to resort to the help of the method of

multi-dimensional classification on the basis of description of the ecosystem by the aggregate of indications (characteristics).

The Methods of Classification

The methods of classification is directed at automatically grouping the objects and features characterized by them. In our case, the target of such analysis could be the solution of the following problems:

1. To establish if gathered in one group the ecosystems close by the vegetation character and how far could be divided unlike systems (i.e., how much the description in NDVI terms reflect the real vegetation activity);

2. To restore vegetation type of the ecosystems for which this characteristic is not given in initial description or defined not enough reliably.

3. To single out the ecosystems which do not correspond to ecotype given in initial description and for the following analysis of the causes (reasons) for this. It should be also find processes that were or are going in these ecosystems.

We apply the methods of automatic classification to subsample of the objects for the it would be possible to account the median of 7 years realizations as a good evaluation of the standard of year realization.

The methods of classification include the different approaches, each of which have its own optimal domain of application. If we take vegetation activity as the main (internal) classification characteristic (feature), then in the case of our sample, it is natural to account that this characteristic runs a continuous scale. For such classification characteristics the more appropriate classification method is the hierarchical that defines the place of every object relative to the others in the hierarchy we obtained. The methods of hierarchical classification differs by the measure of closeness between the objects and the ways of assessment of distance between the classes (for example one-connected, two-connected and many-connected methods) on intermediate stages of classification (Aivazian *et al.*, 1989). The selection of measure of closeness (proximity) and of the agglomeration method are one of the more important elements of the tuning up of the classification method on peculiarity of the problem that we try to solve. Such a tuning was conducted on the stage of exploratory hierarchical classification. As a result it was shown that the most effective is one-connected method.

The results of classification by this method of standard year realization $M(f_i) \in R^{12}$, $i = 1,\ldots,40$ are given in *Figure 12*. From this figure it is seen that the (investigated) group of

ecosystem is described by the hierarchical structure in which as the ordering of the objects so the values of the measures of relations between objects and groups of objects have a good enough interpretation in the terms of initial description of ecosystems from grass savanna to the tropical forest with the transition zone as bush savanna, forest savanna and different types of mixing zone of savanna and forest (different mosaics).

In the concluding section, let us describe the result of classification of months of the observation as the features of standard year realization.

To each month m we put into the correspondence a vector the dimension of which is equal to the number of the ecosystems and i-th coordinate is equal to the characteristic values of NDVI in m-th month for the i-th ecosystems for all period of observation. Making the hierarchical classification of the 12 vectors, we obtain the description of the interrelations between the characteristic values of NDVI during the transition from month to month.

Figure 9a represents the result of such a mean-connected method for month-features described by 12 vectors $\varphi_m = \{M(f_i(m)), i = 1, \ldots, 40\} \in R^{40}$, where $M(f_i(m))$, as above, is a median of seven values of NDVI in m-th month of 7-year series of observations for i-th ecosystem. Comparing *Figure 9a* and *Figure 9b*, on which are represented the results of the classification of the month by the two methods, we obtain a more complete understanding of the feature of the standard year realization of the NDVI.

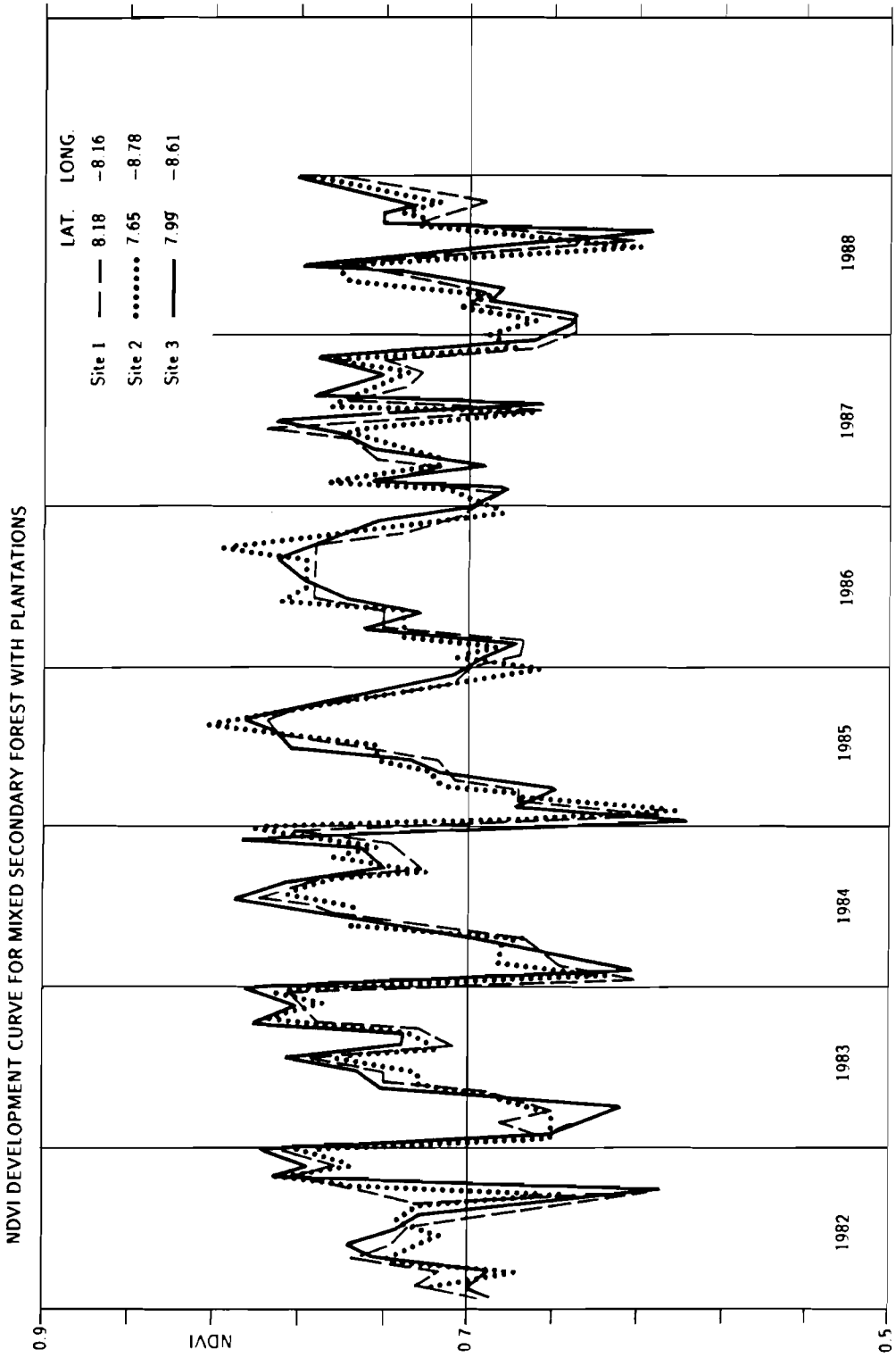

NDVI DEVELOPMENT CURVE FOR MIXED SECONDARY FOREST WITH PLANTATIONS
NDVI
0.9
0.7
0.5
LAT. LONG.
Site 1 — — 8.18 −8.16
Site 2 ••••••• 7.65 −8.78
Site 3 ——— 7.99 −8.61
1982
1983
1984
1985
1986
1987
1988

Figure 1.

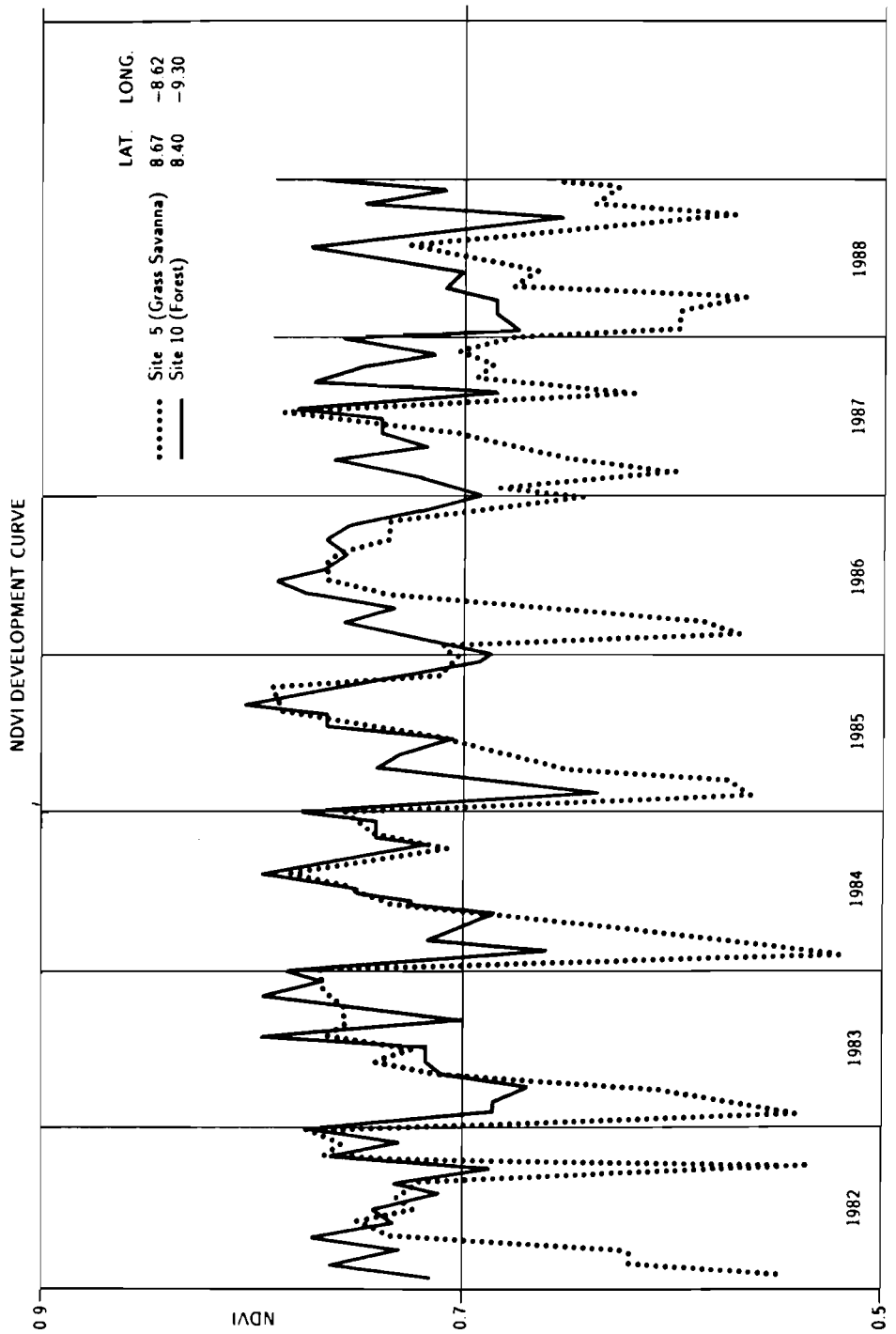

NDVI DEVELOPMENT CURVE
Site 5 (Grass Savanna)
Site 10 (Forest)
LAT. LONG.
8.67 −8.62
8.40 −9.30
NDVI
0.9
0.7
0.5
1982
1983
1984
1985
1986
1987
1988

Figure 2.

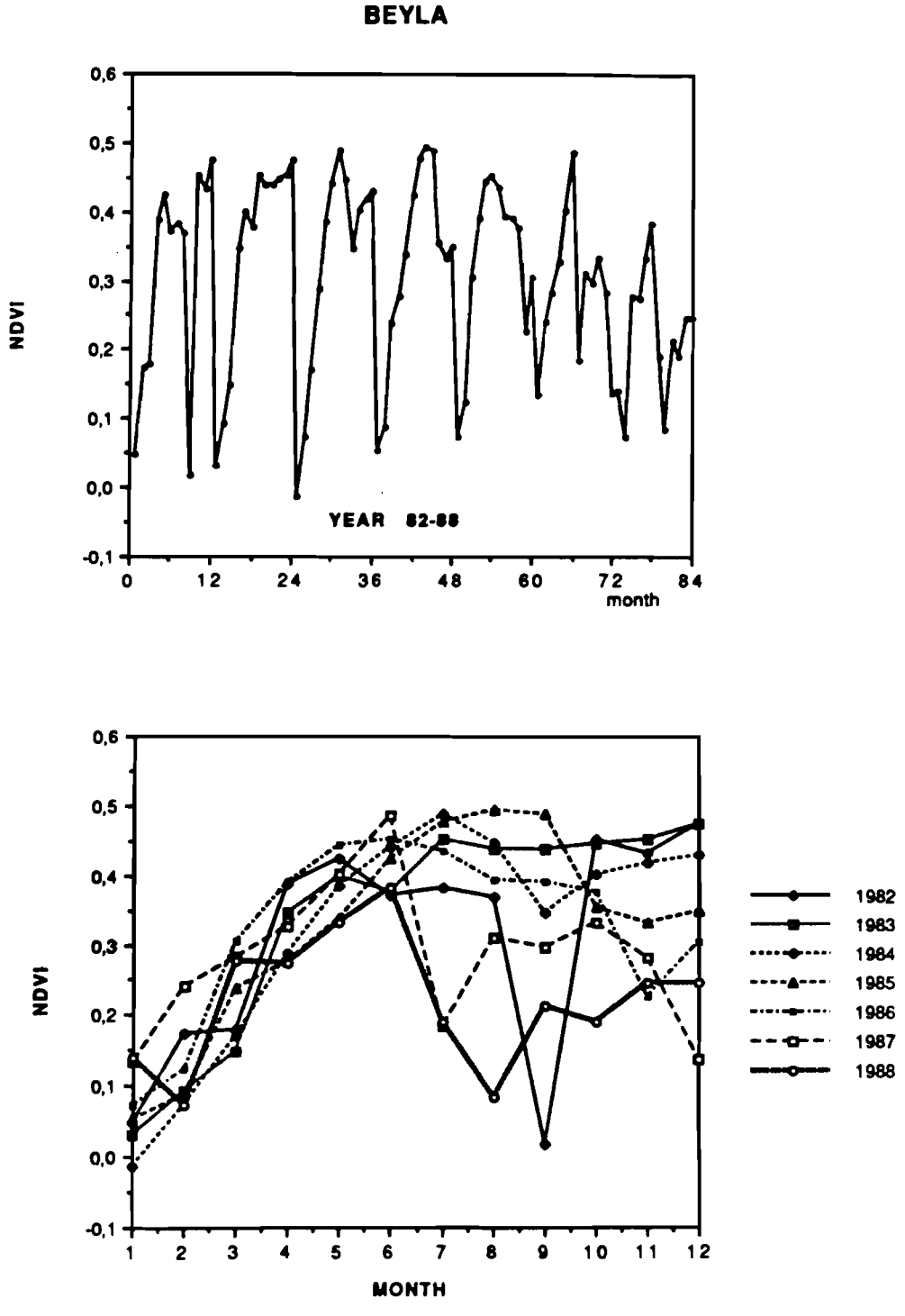

BEYLA
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Figure 3.

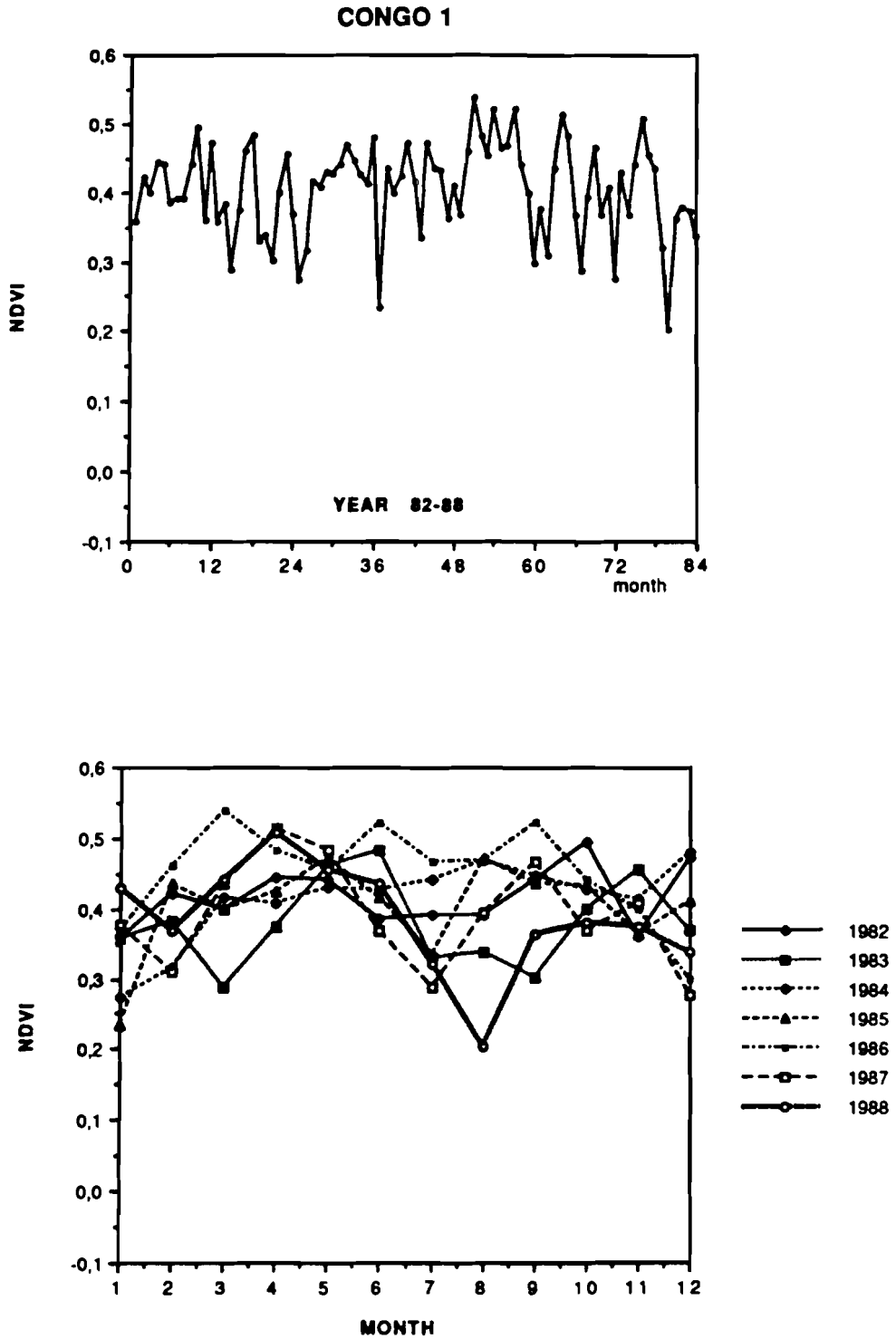

CONGO 1
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Figure 4.

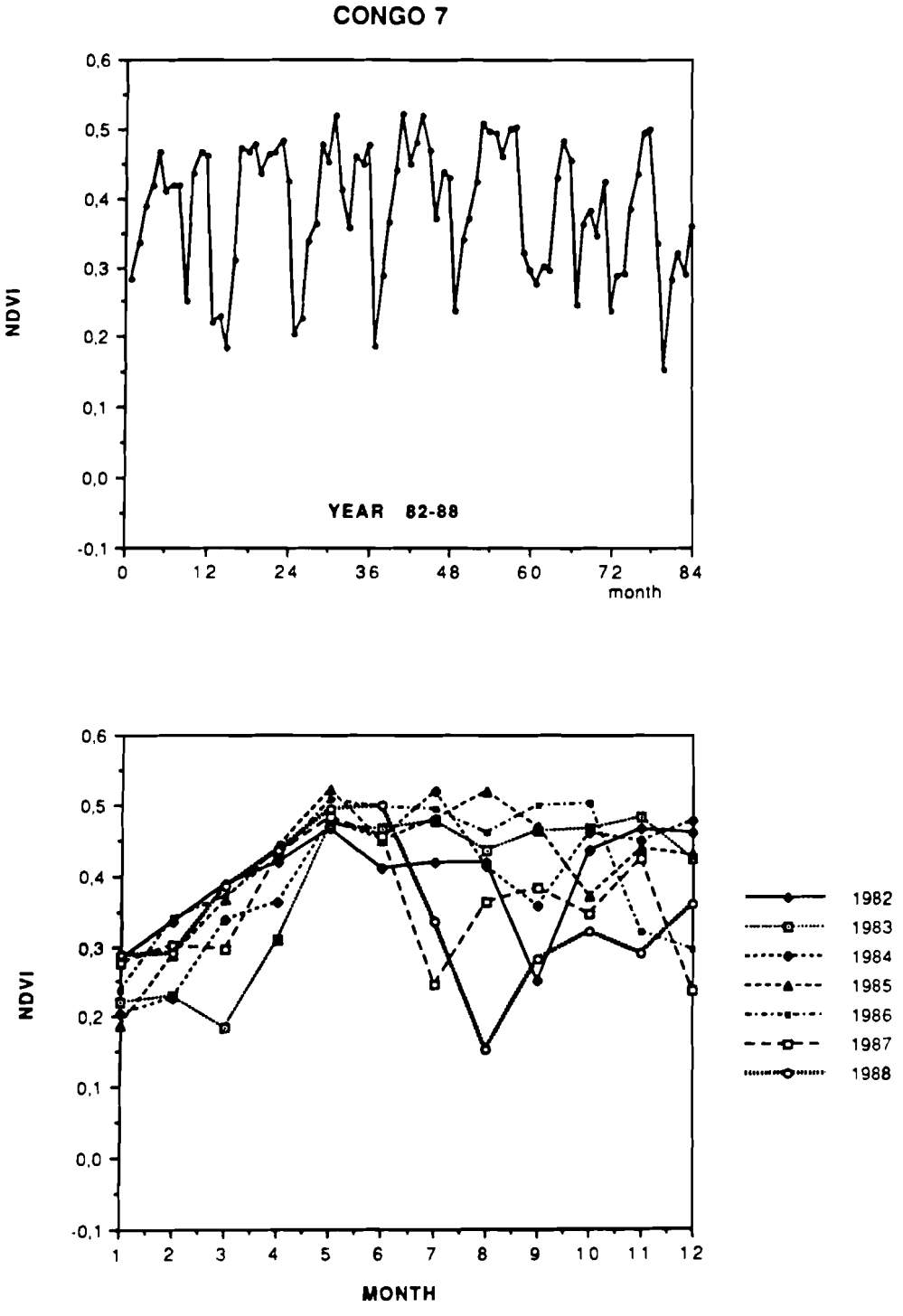

CONGO 7
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Figure 5.

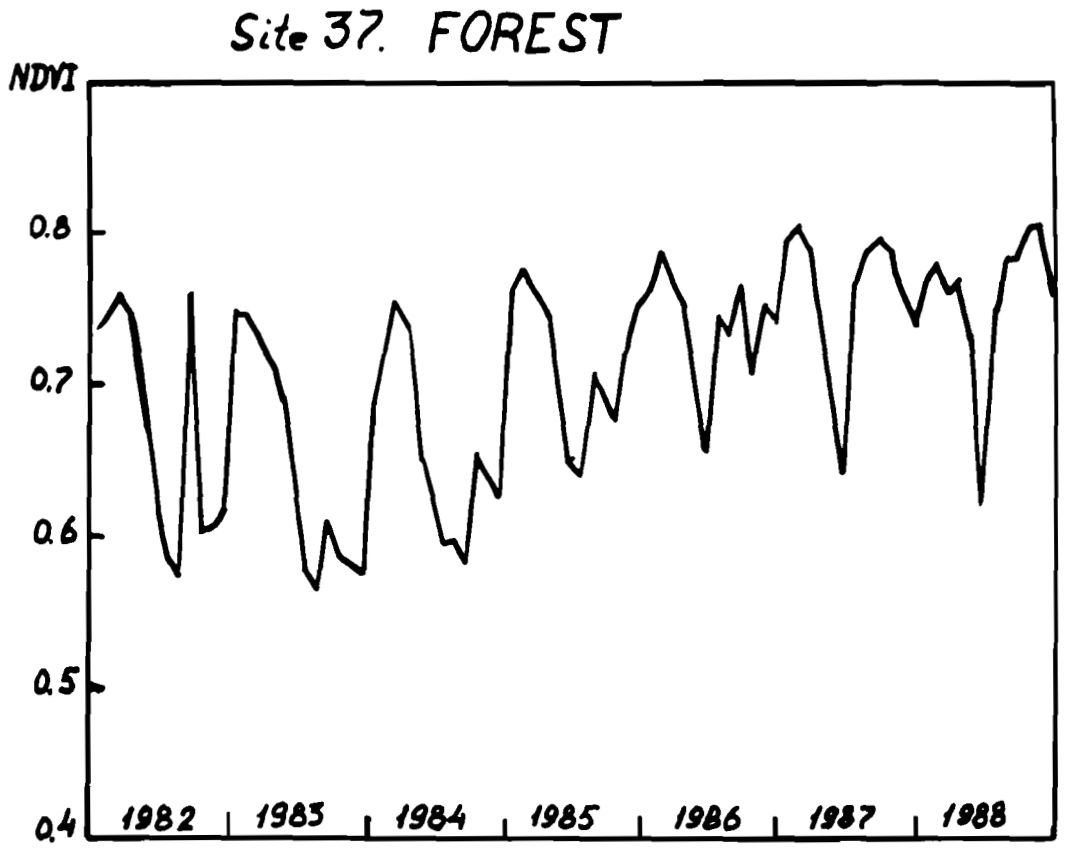
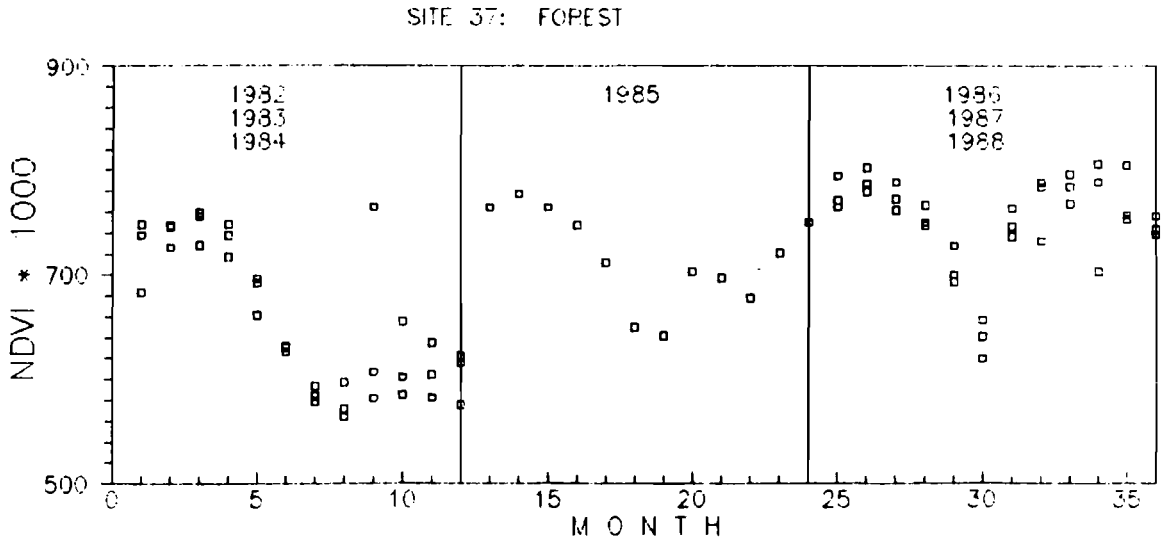

Site 37. FOREST
NDVI
0.8
0.7
0.6
0.5
0.4
1982
1983
1984
1985
1986
1987
1988

Figure 6.


SITE 37: FOREST
900
NDVI * 1000
700
500
1982
1983
1984
1985
1986
1987
1988
0
5
10
15
20
25
30
35
M O N T H

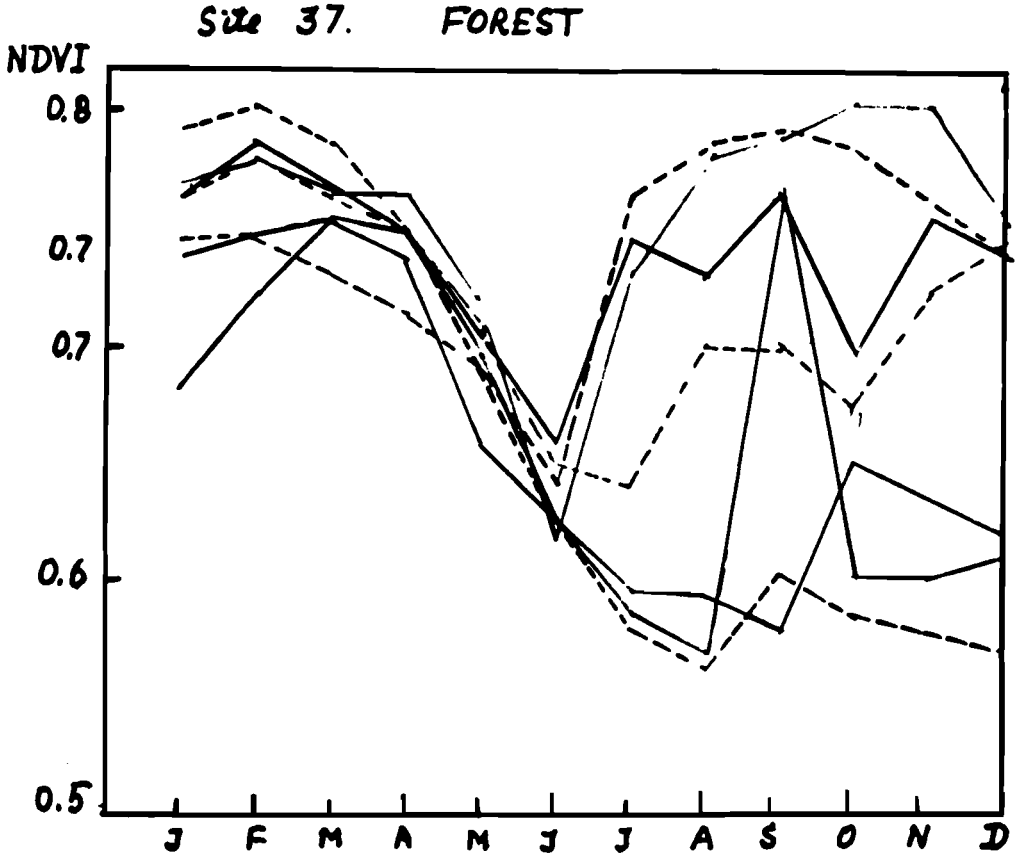

Site 37. FOREST
NDVI
0.8
0.7
0.7
0.6
0.5
J
F
M
A
M
J
J
A
S
O
N
D

Figure 7.

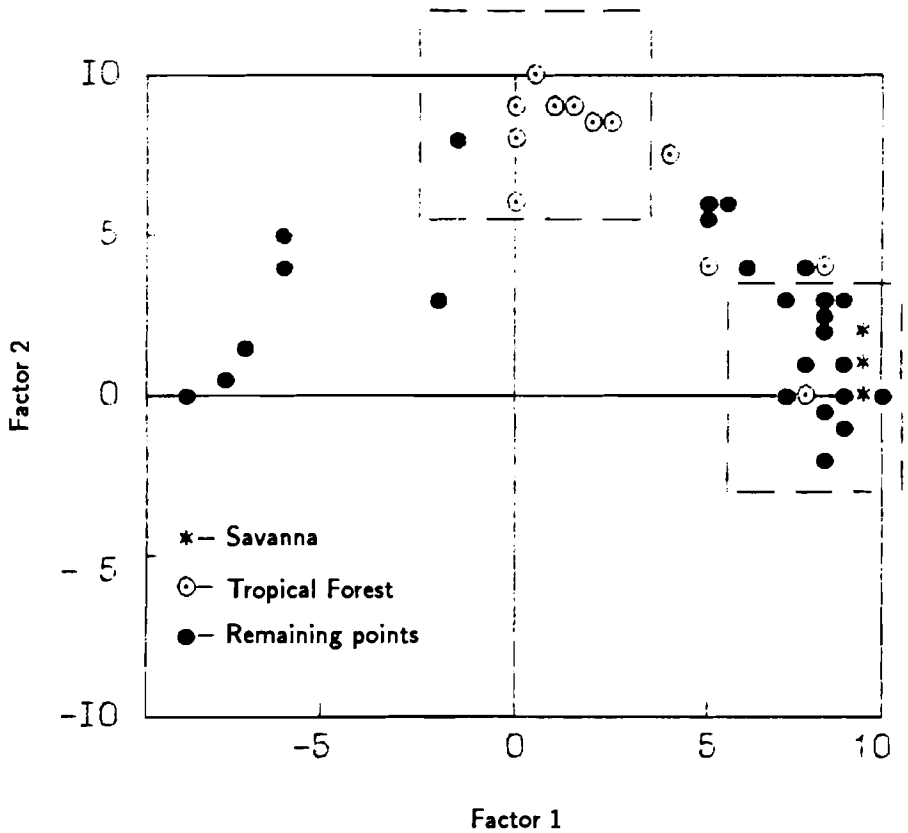
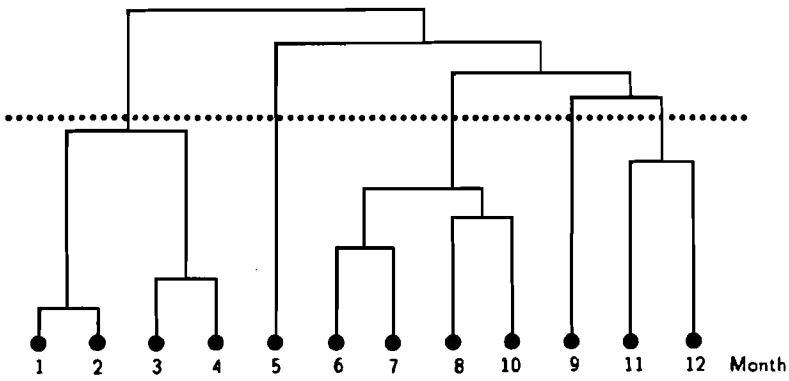


Figure 8. Projection of vectors, corresponding to median evaluation of year realization of ecosystem, on the plane of the first two main factors in R^{40}.

(a)



(b)

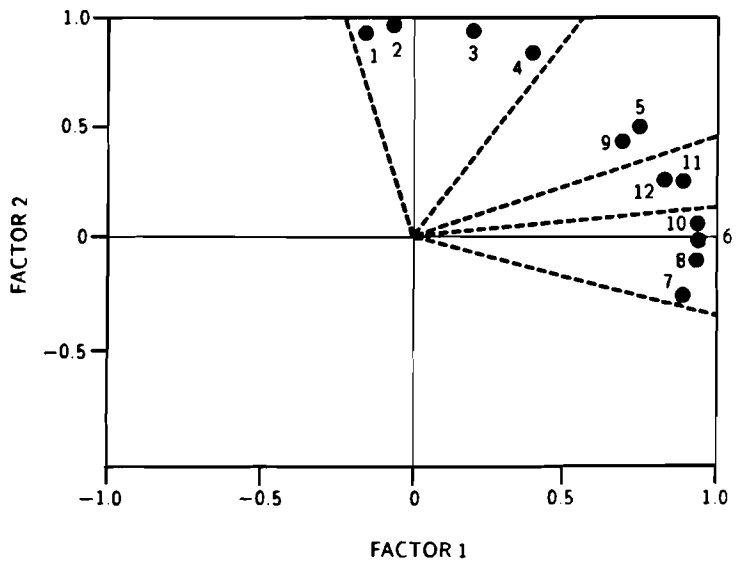


Figure 9. (*a*) Hierarchical tree of the month as features of standardized year realization of NDVI curves. (*b*) Projection of the vectors corresponding to the 12 points of marks (months), on the plane of the first main factors in R^{12}.

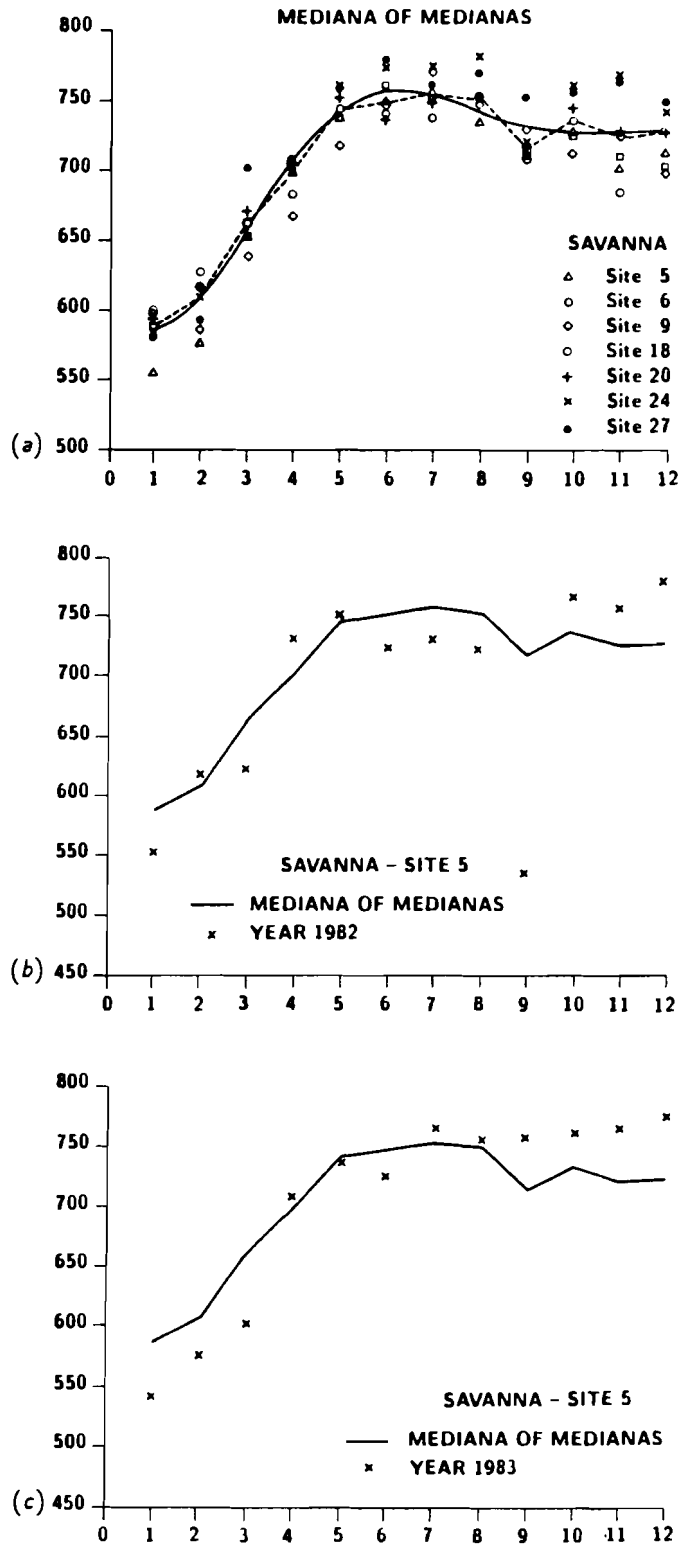

MEDIANA OF MEDIANAS
SAVANNA
Site 5
Site 6
Site 9
Site 18
Site 20
Site 24
Site 27
(a)
SAVANNA - SITE 5
MEDIANA OF MEDIANAS
YEAR 1982
(b)
SAVANNA - SITE 5
MEDIANA OF MEDIANAS
YEAR 1983
(c)

Figure 10.

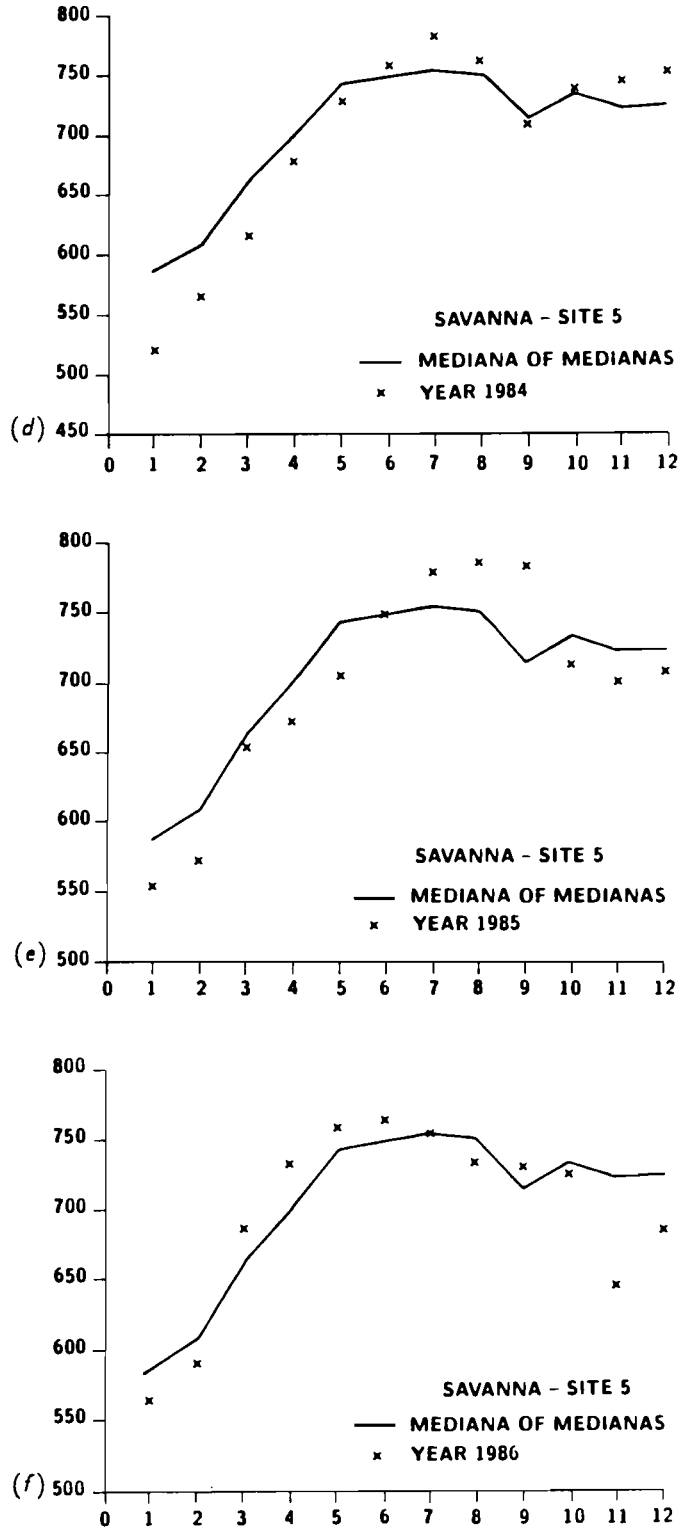


Figure 10. Continued.

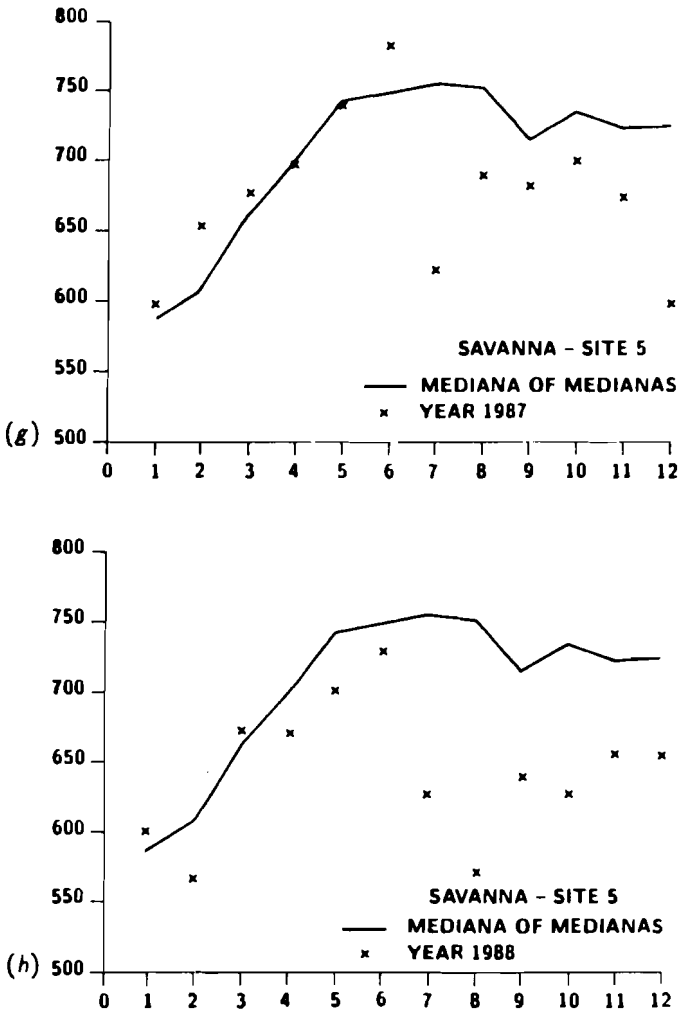


Figure 10. Continued.

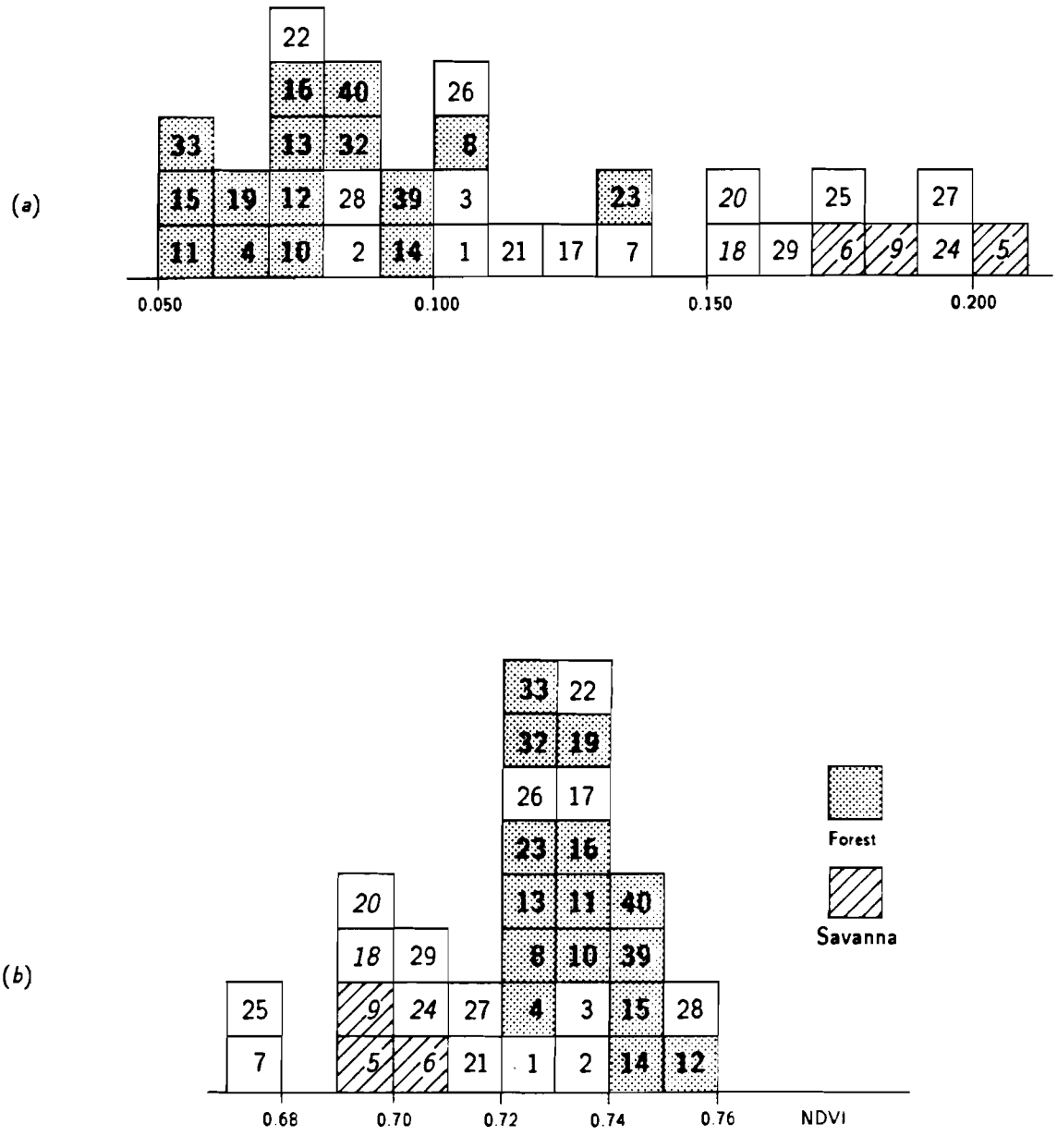


Figure 11. (*a*) Hystogram of difference (max-min) of median assessment. (*b*) Hystogram of mean value of seven-year series of observations.

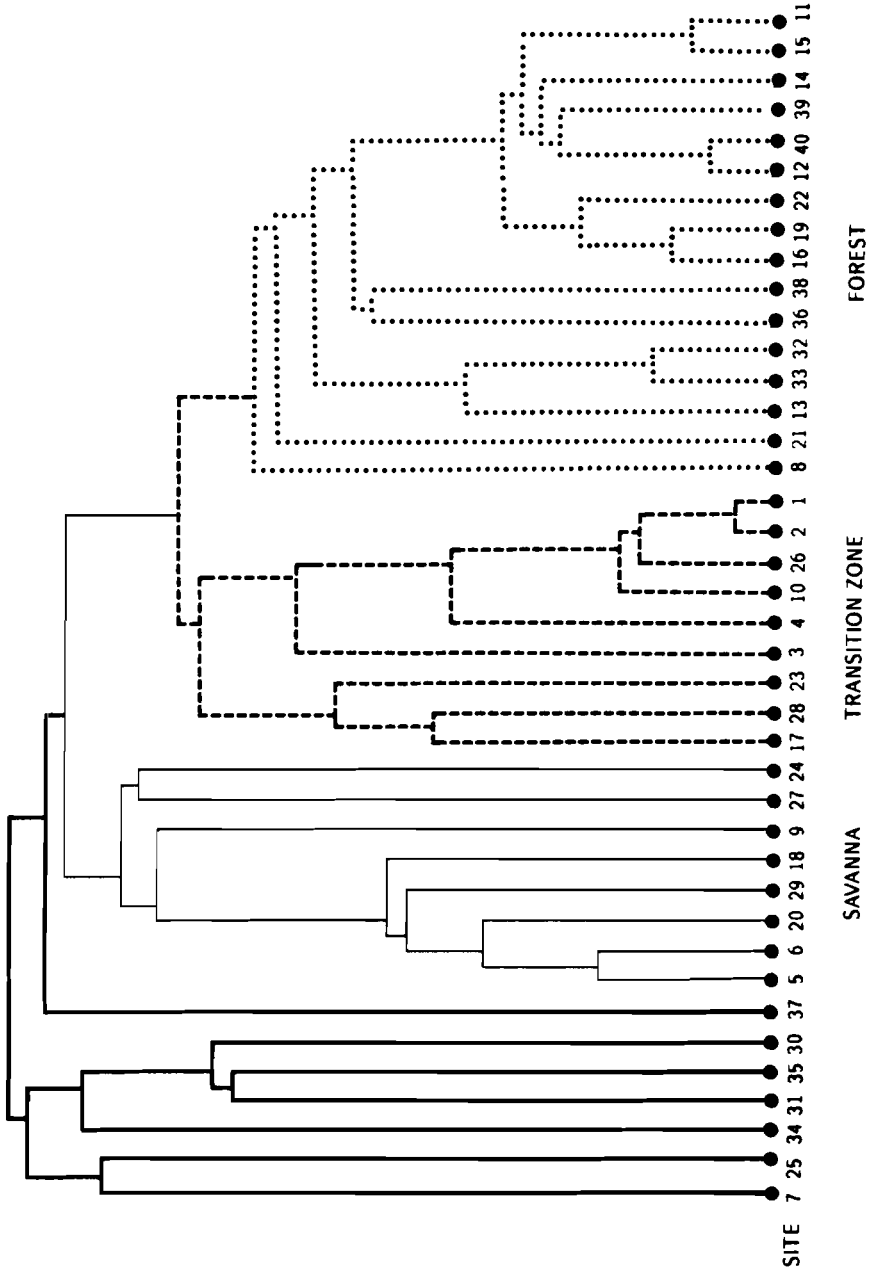
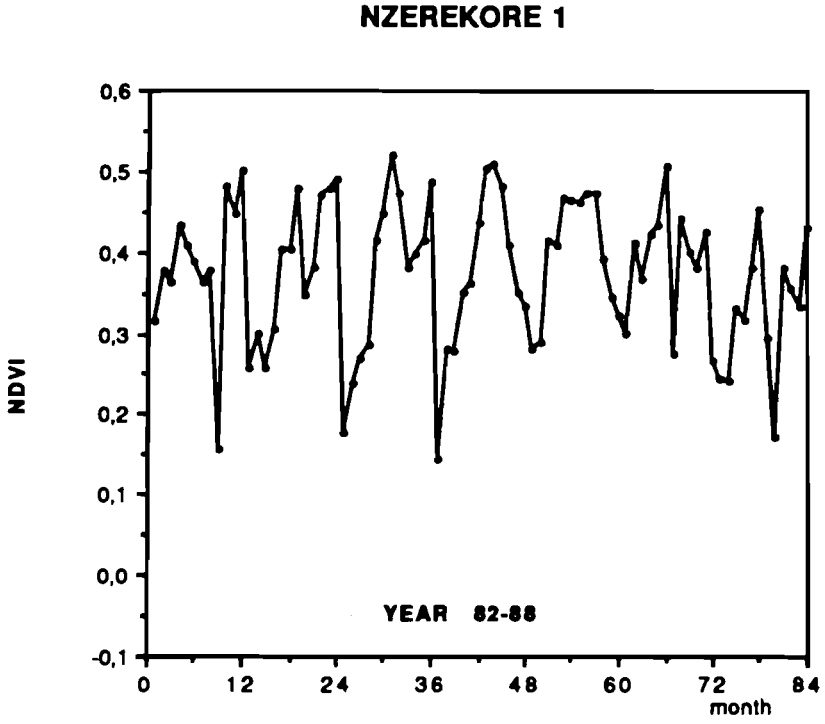
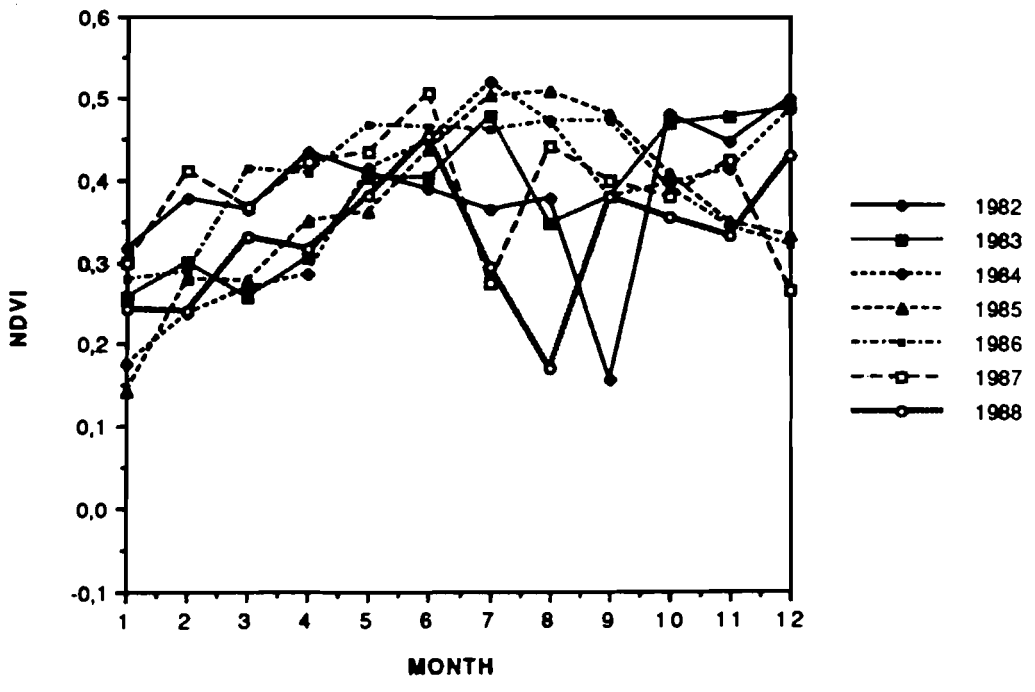


Figure 12. Hierarchical tree for ecosystems of West Africa at Equatorial zone.

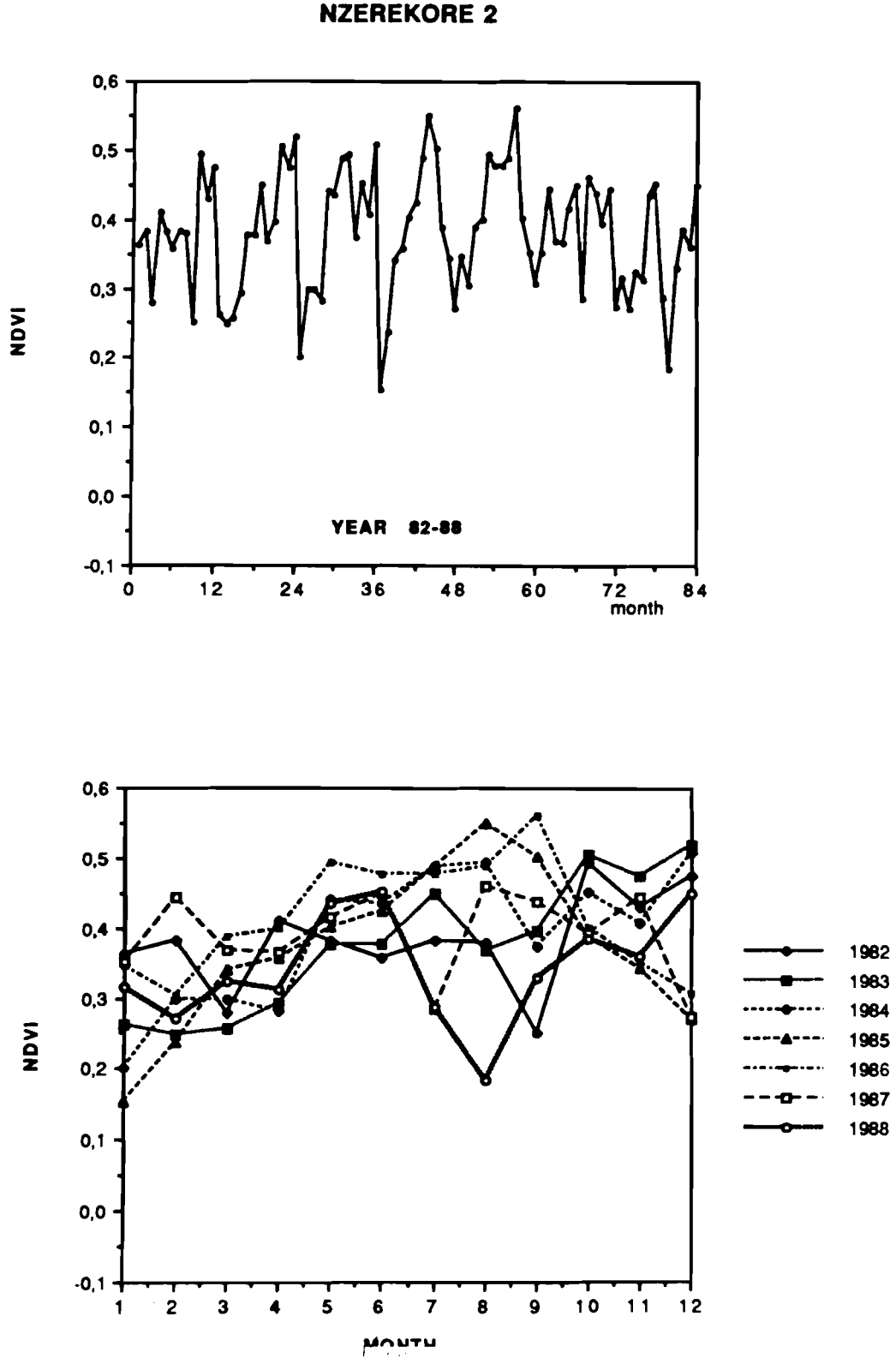
Appendix 2

Statistical Pictures

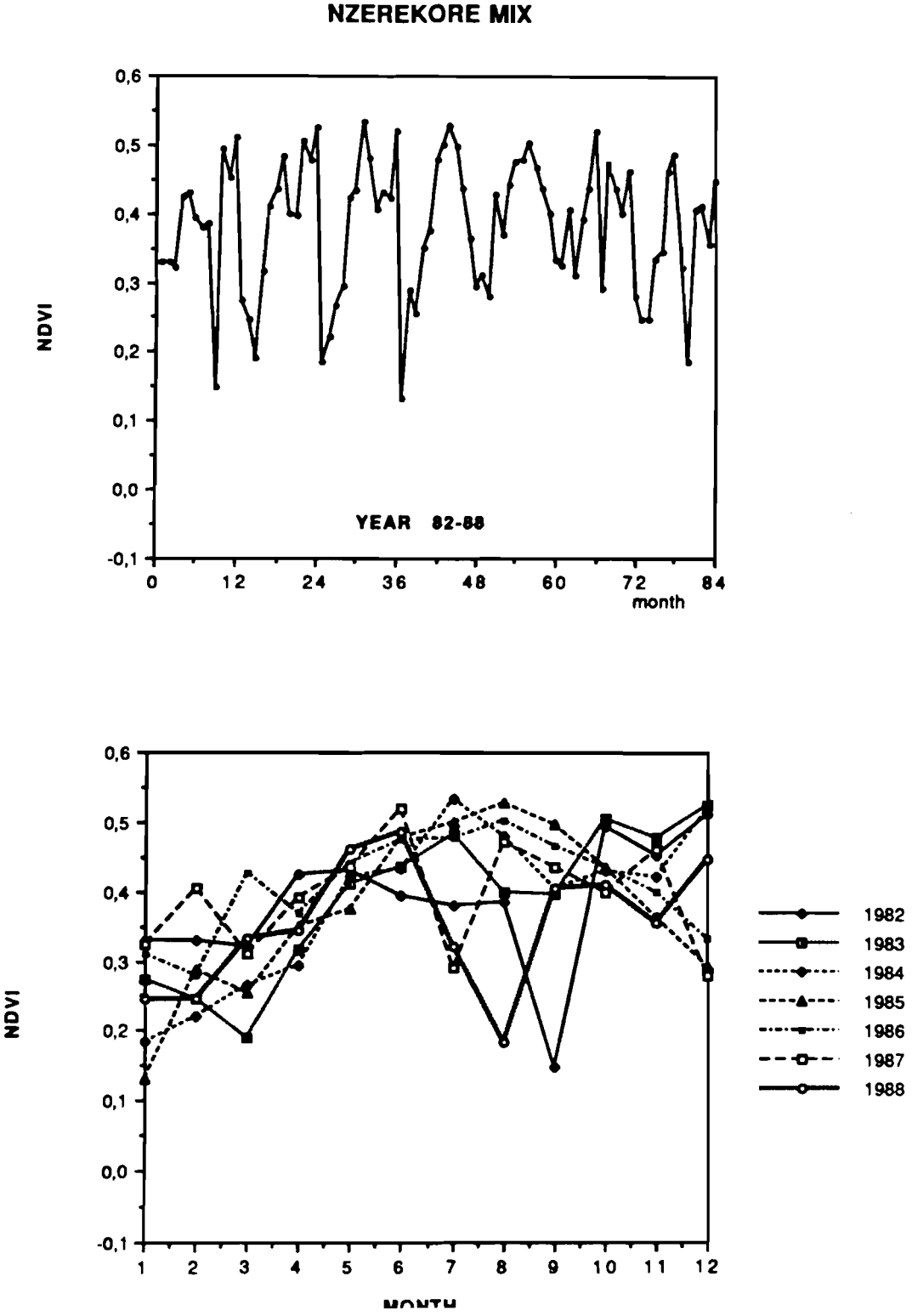

NZEREKORE 1
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month


NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

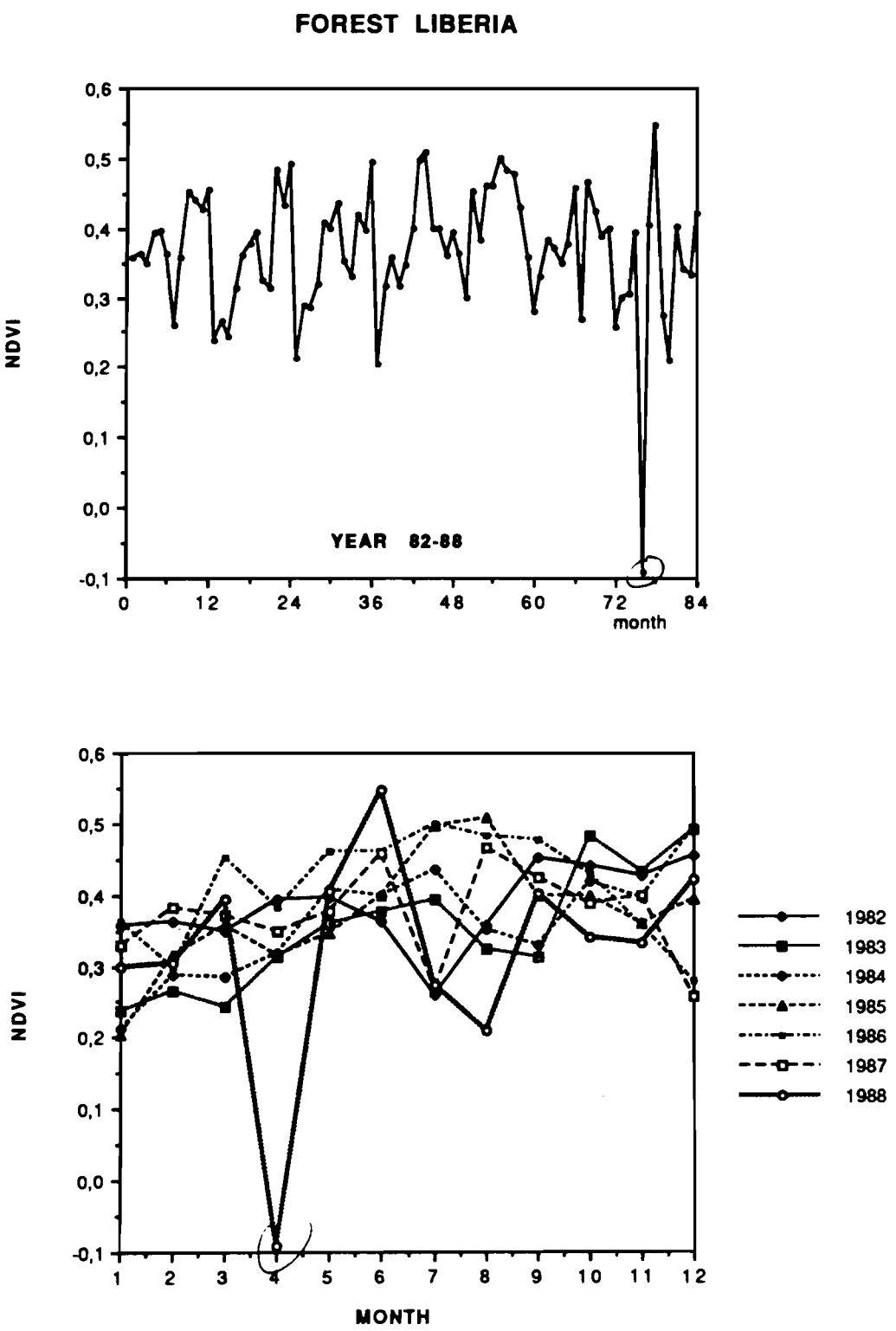
Site 1.


NZEREKORE 2
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

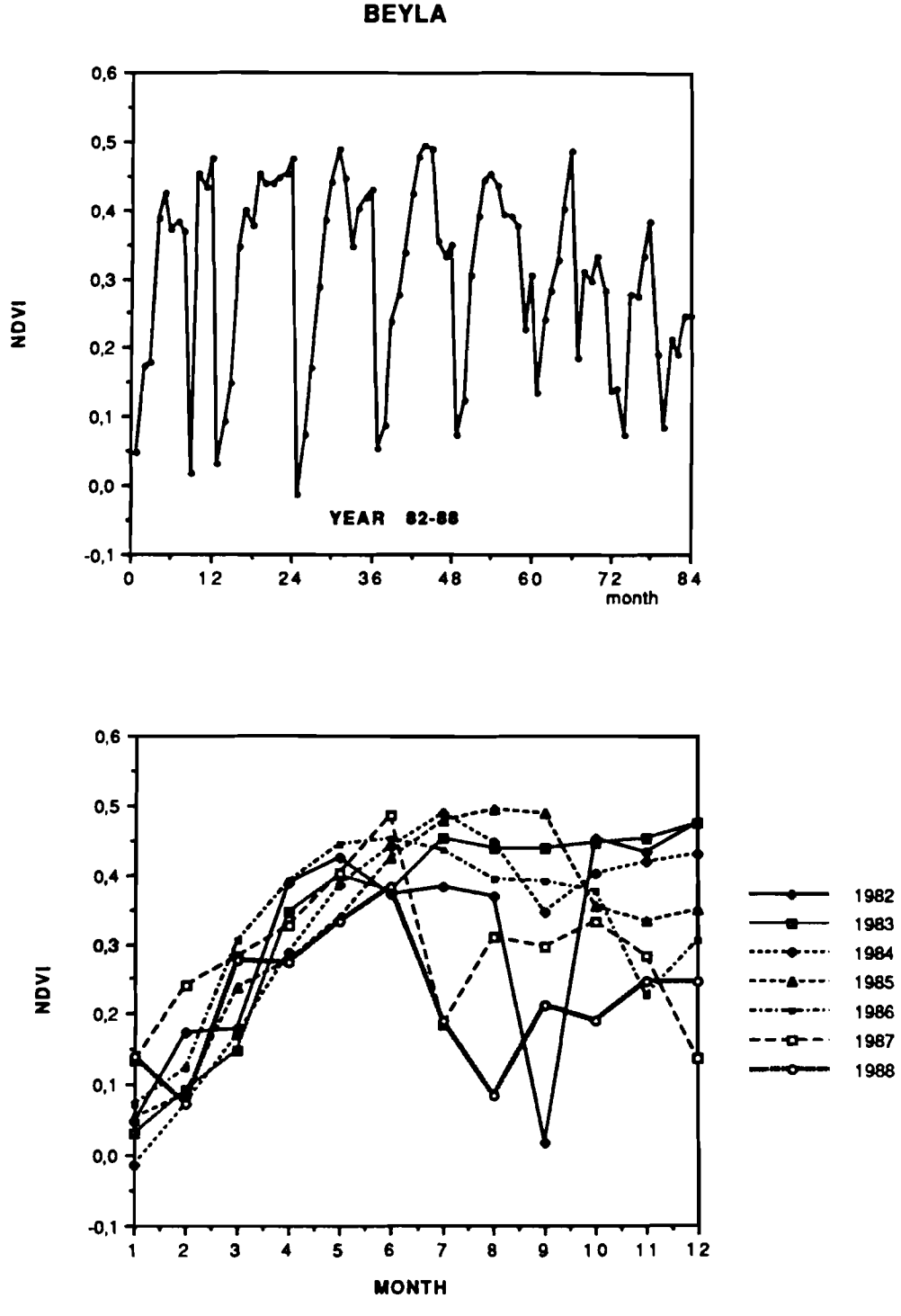
Site 2.


NZEREKORE MIX
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

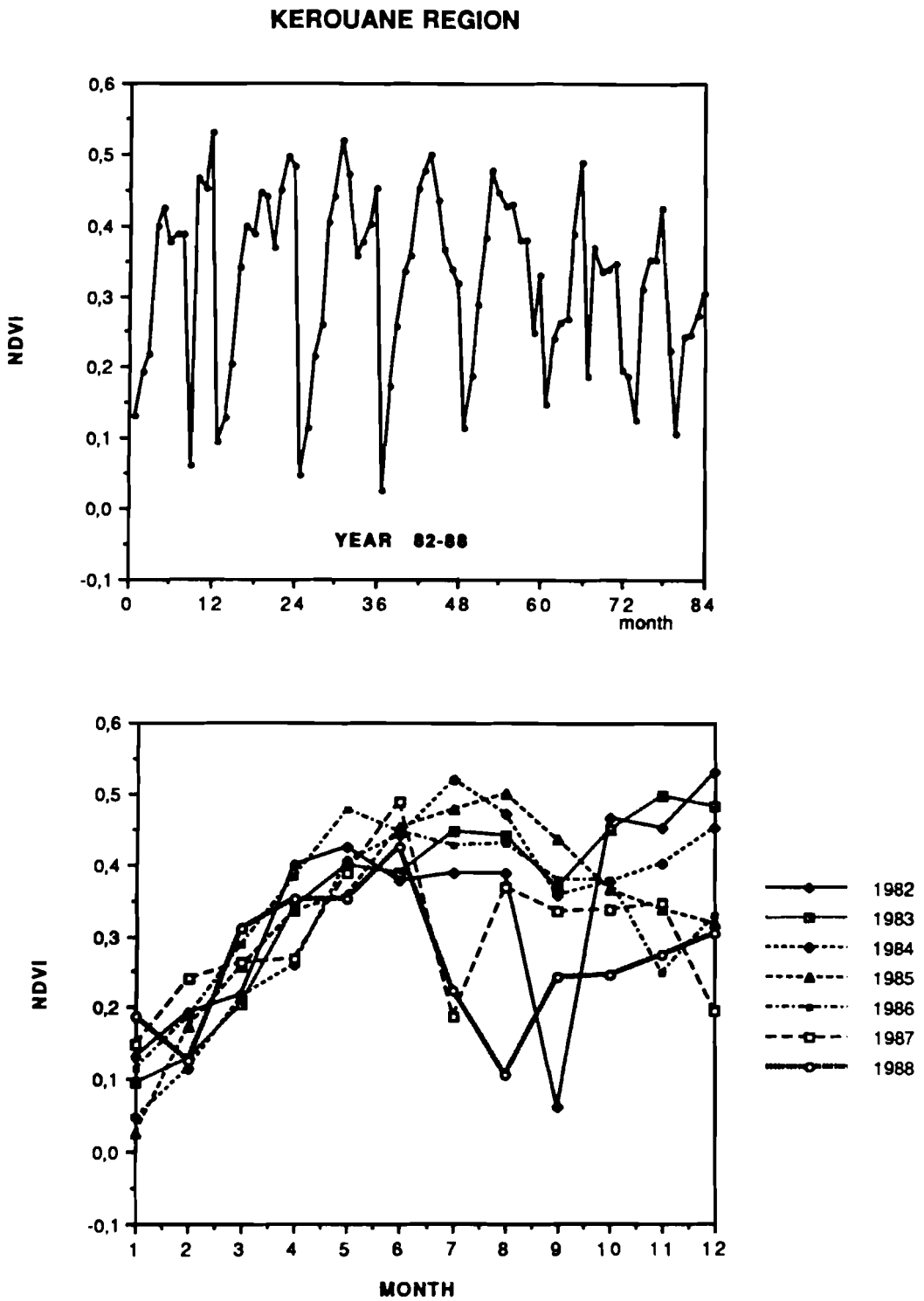
Site 3.


FOREST LIBERIA
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

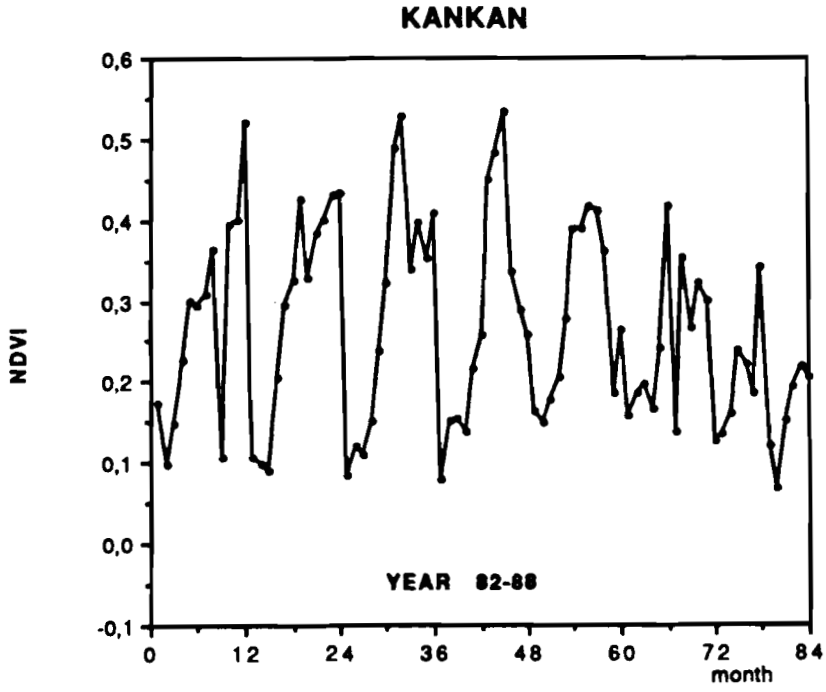
Site 4.

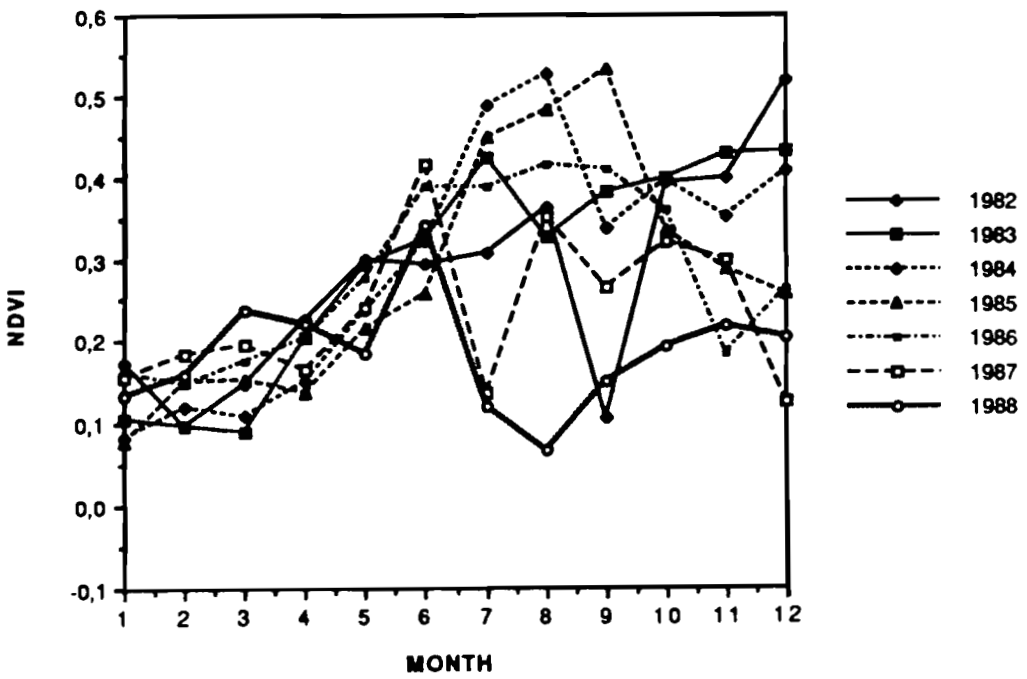

BEYLA
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Site 5.

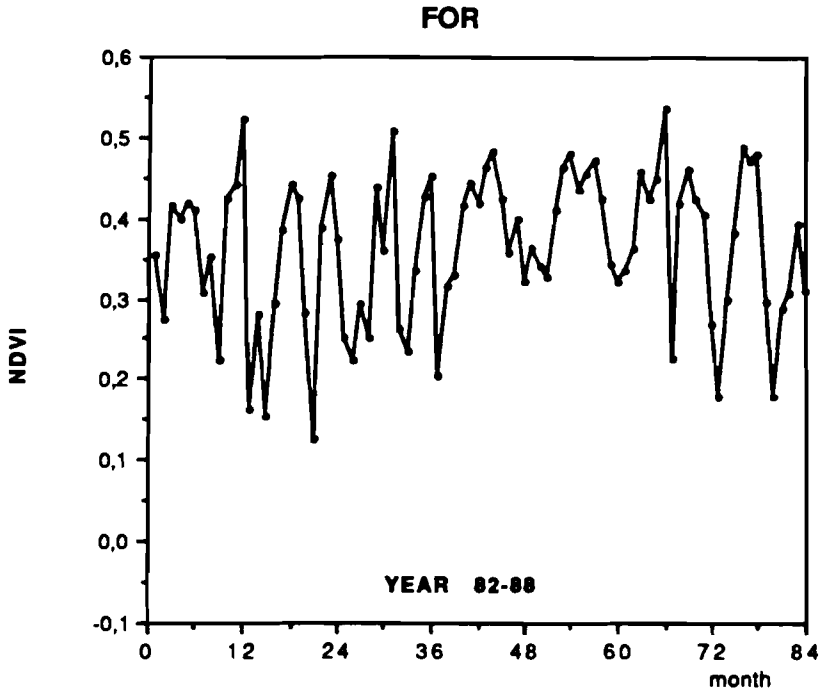

KEROUANE REGION
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

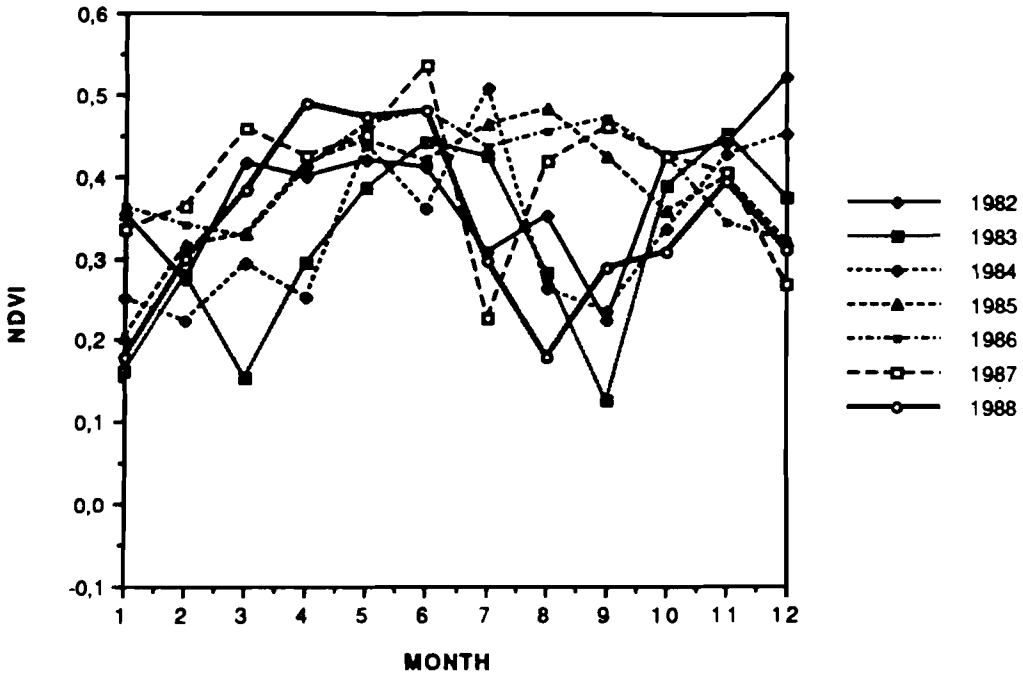
Site 6.


KANKAN
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

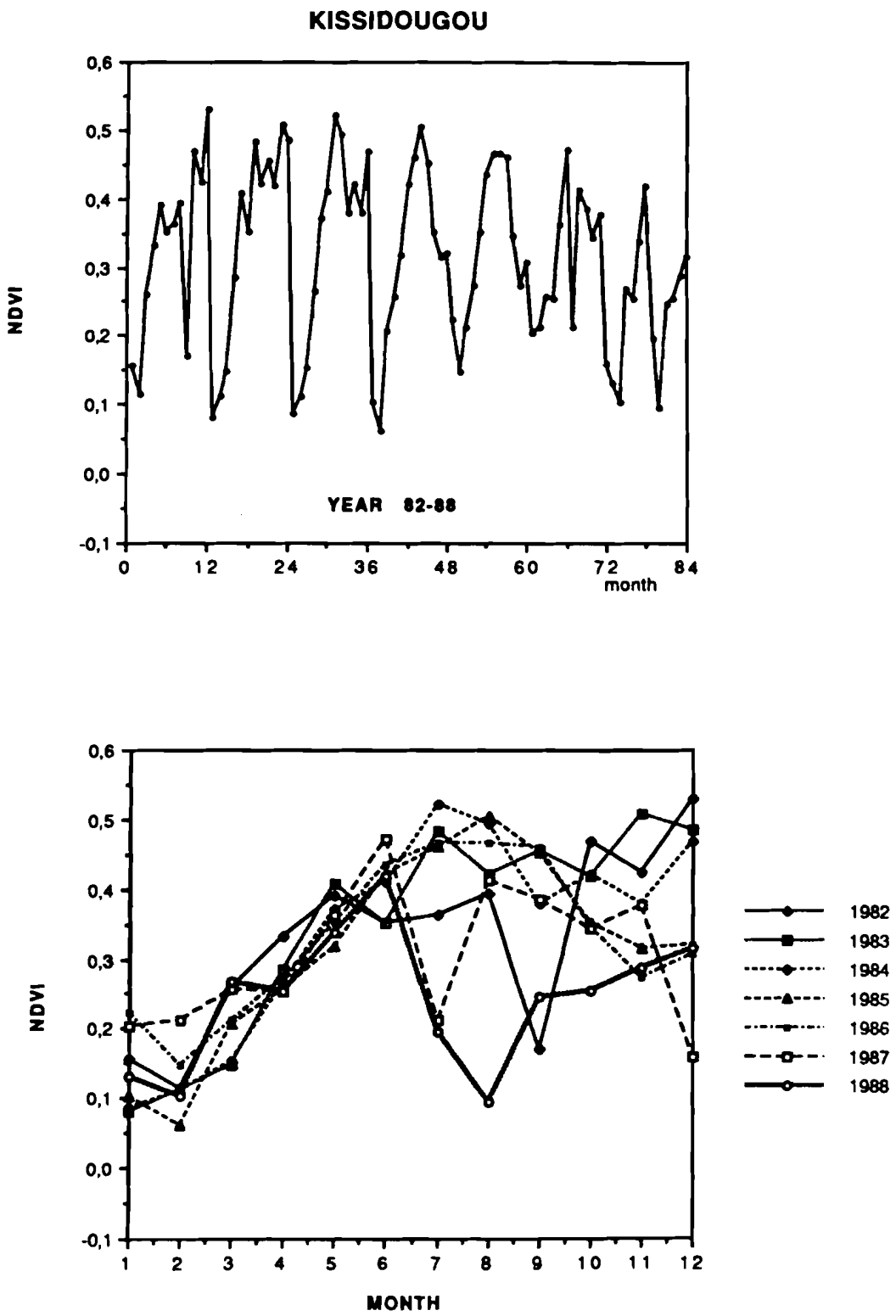

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

Site 7.

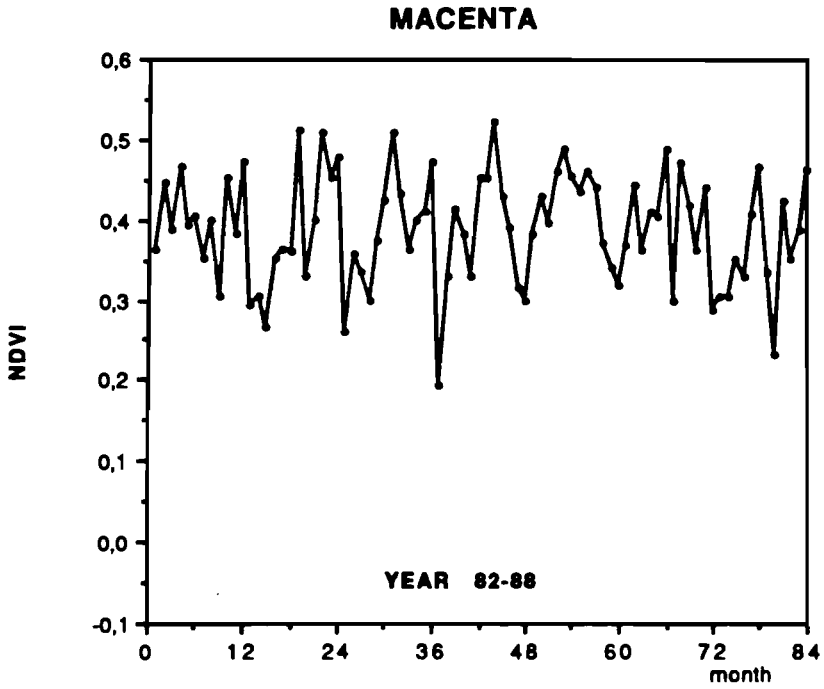

FOR
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

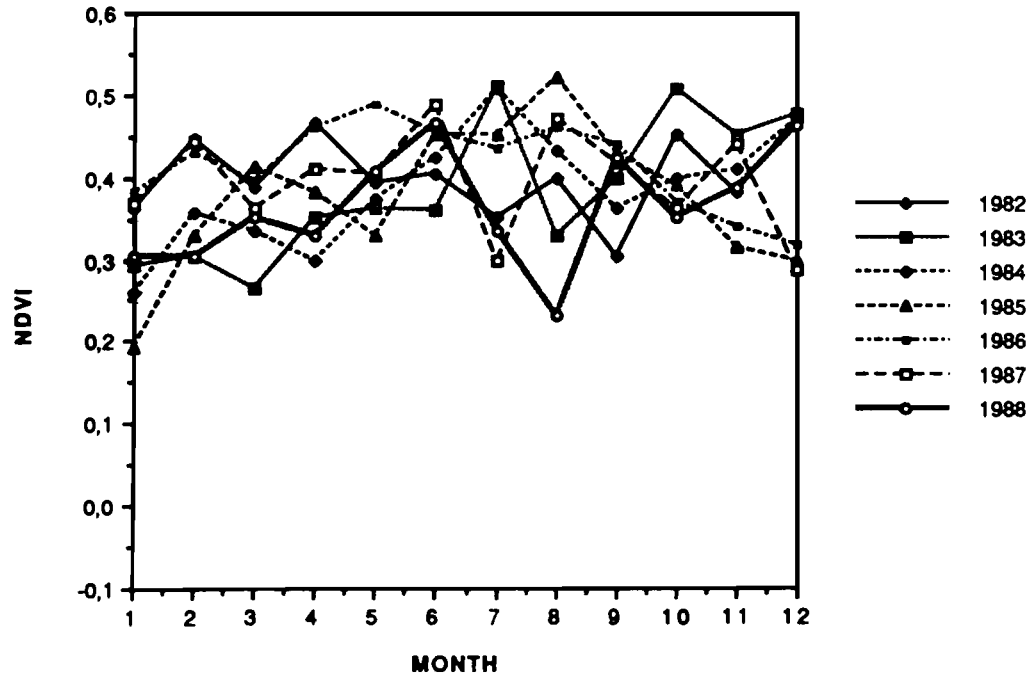

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

Site 8.

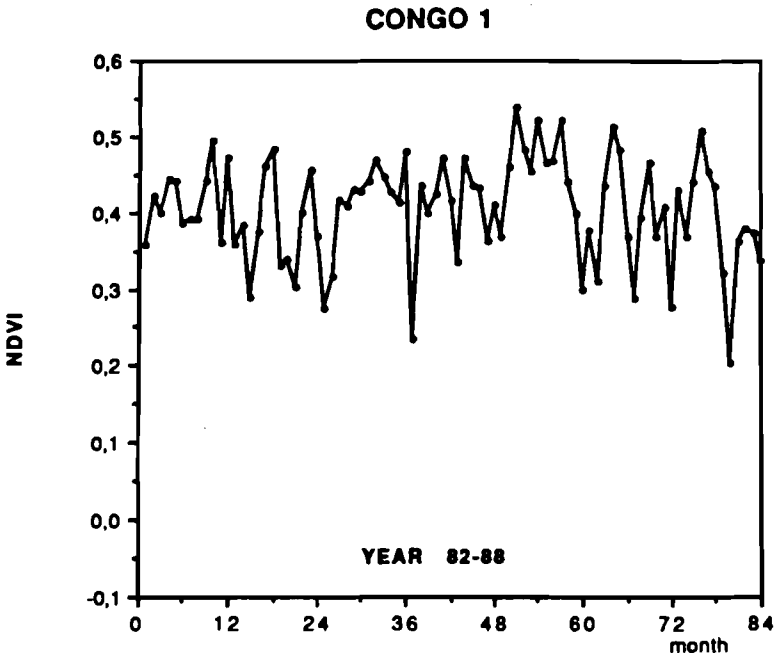

KISSIDOUGOU
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

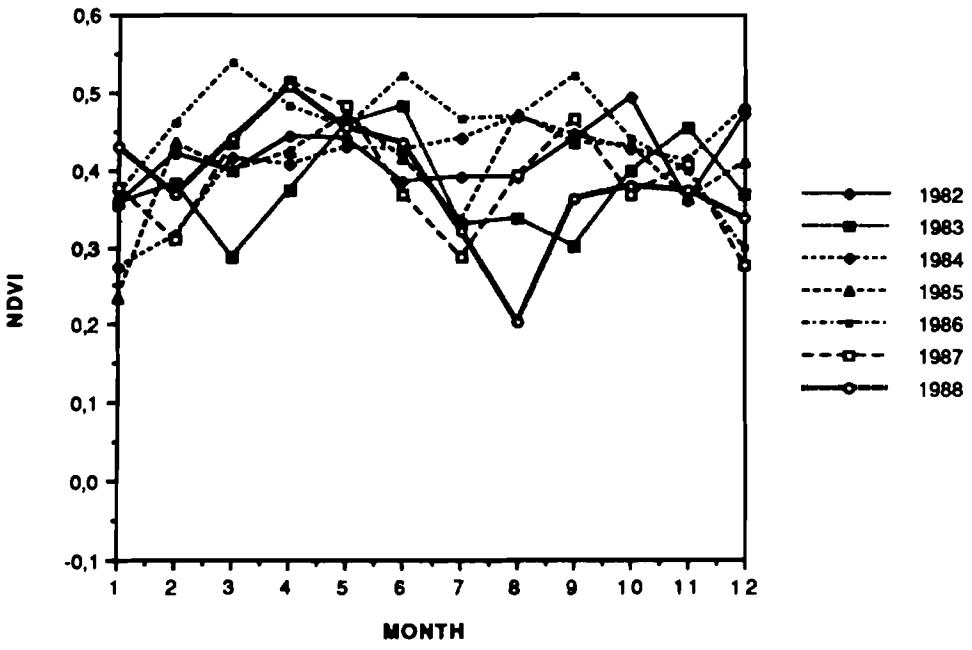
Site 9.


MACENTA
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

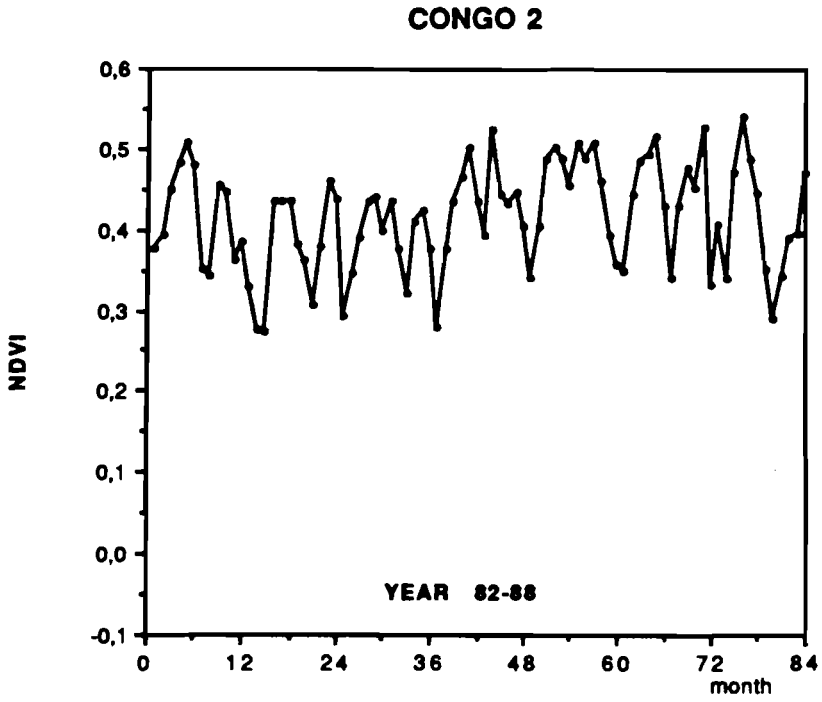

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

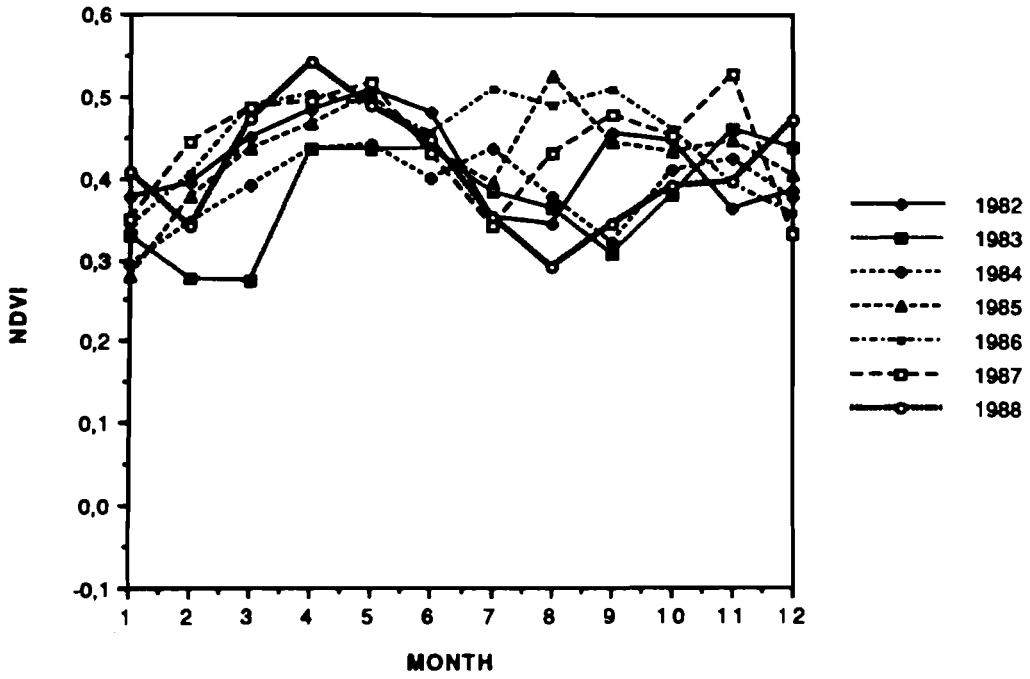
Site 10.


CONGO 1
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

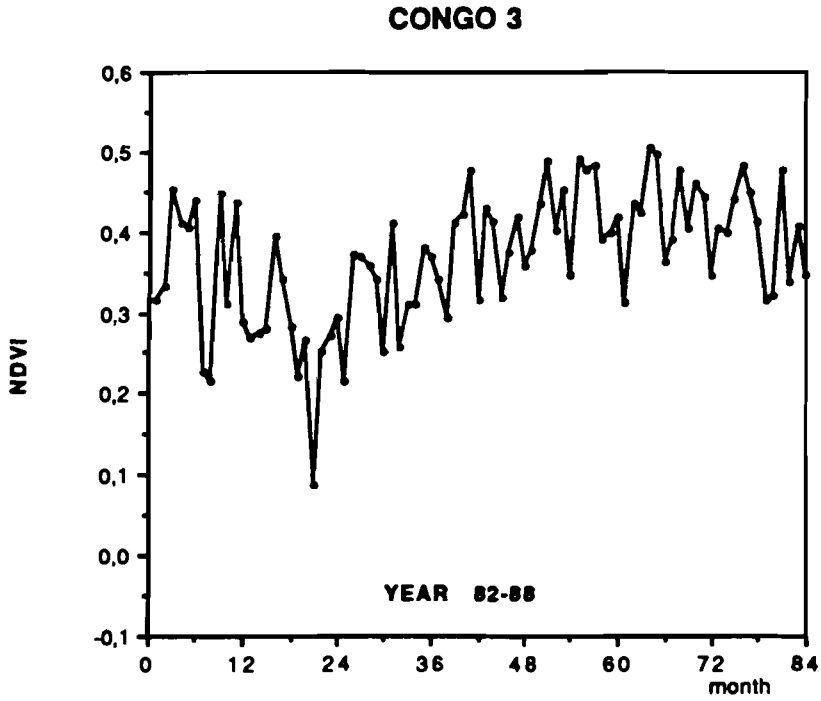

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

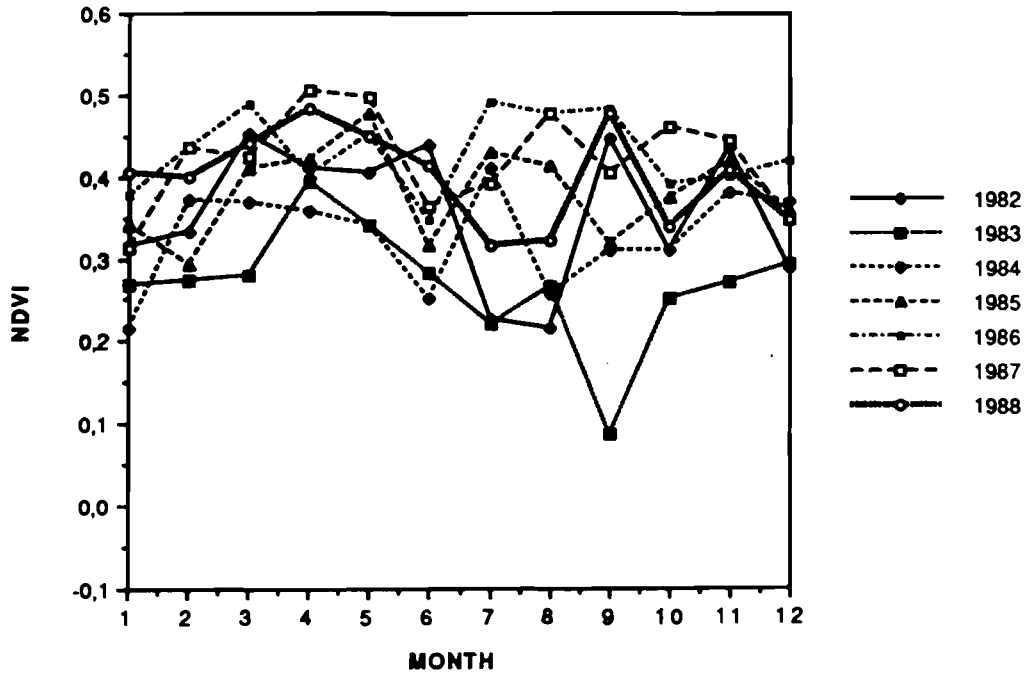
Site 11.


CONGO 2
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

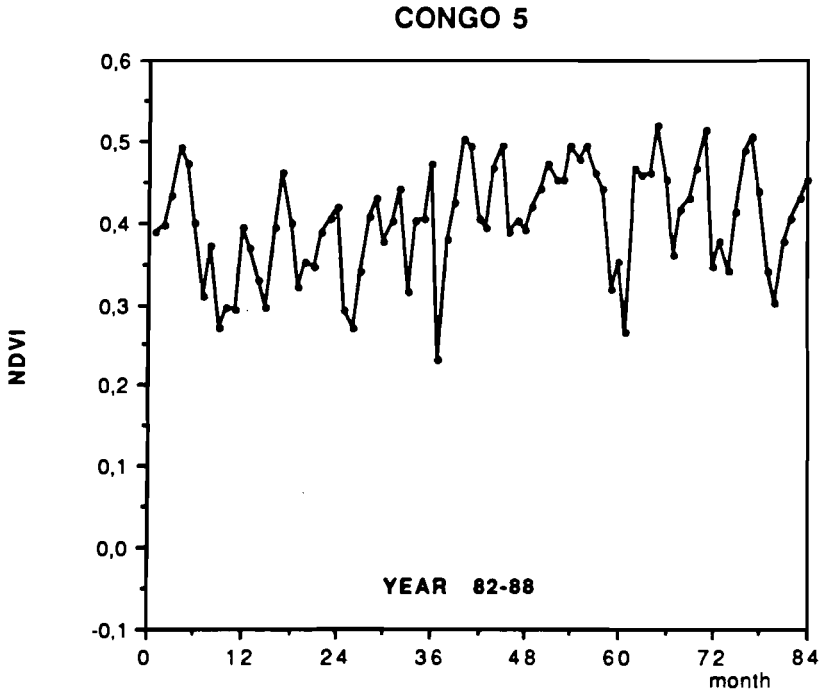

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

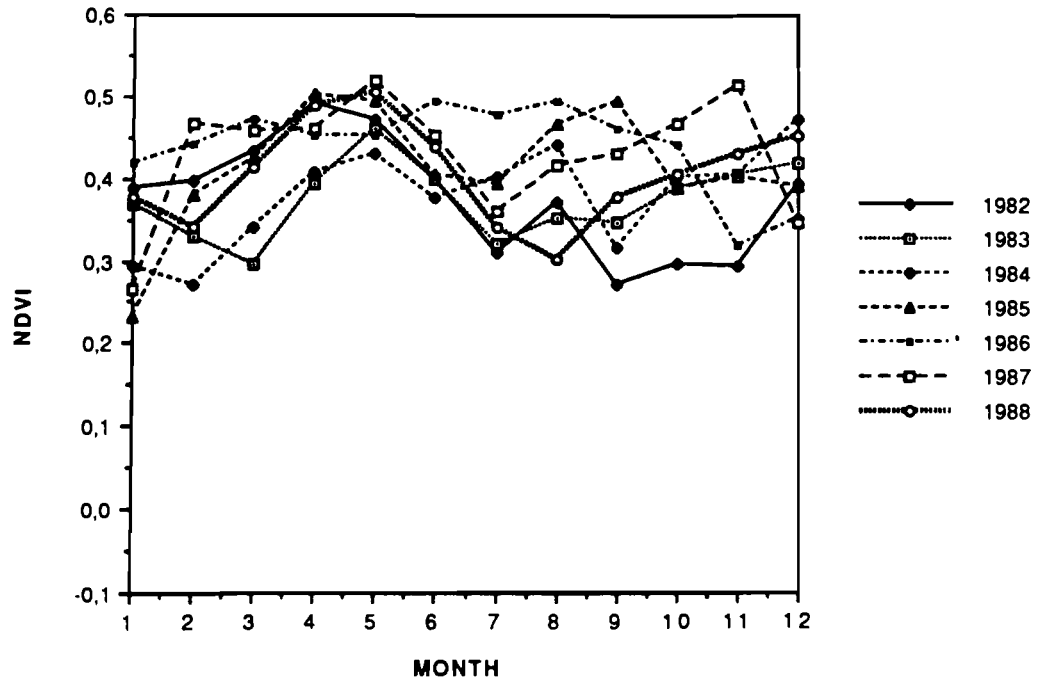
Site 12.


CONGO 3
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

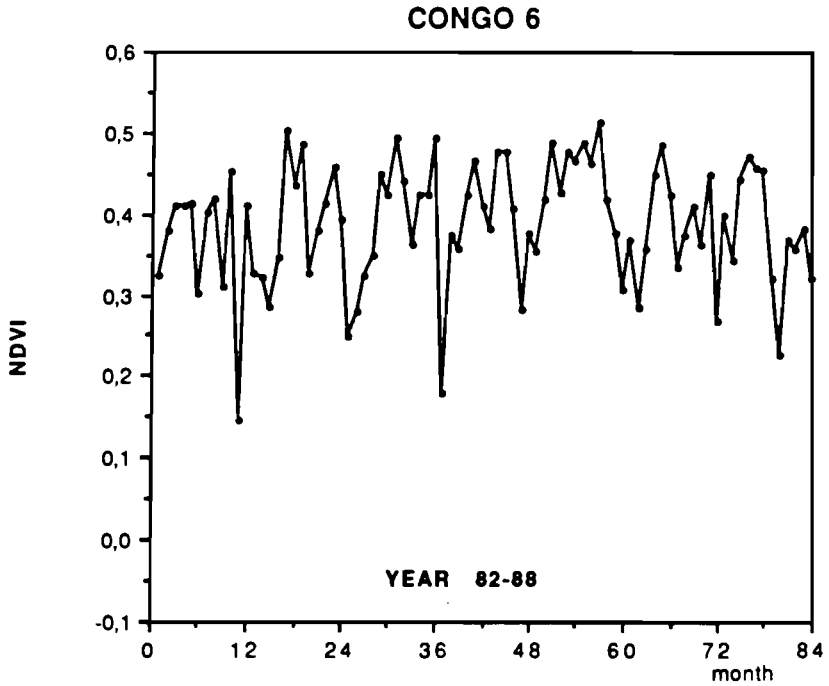

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

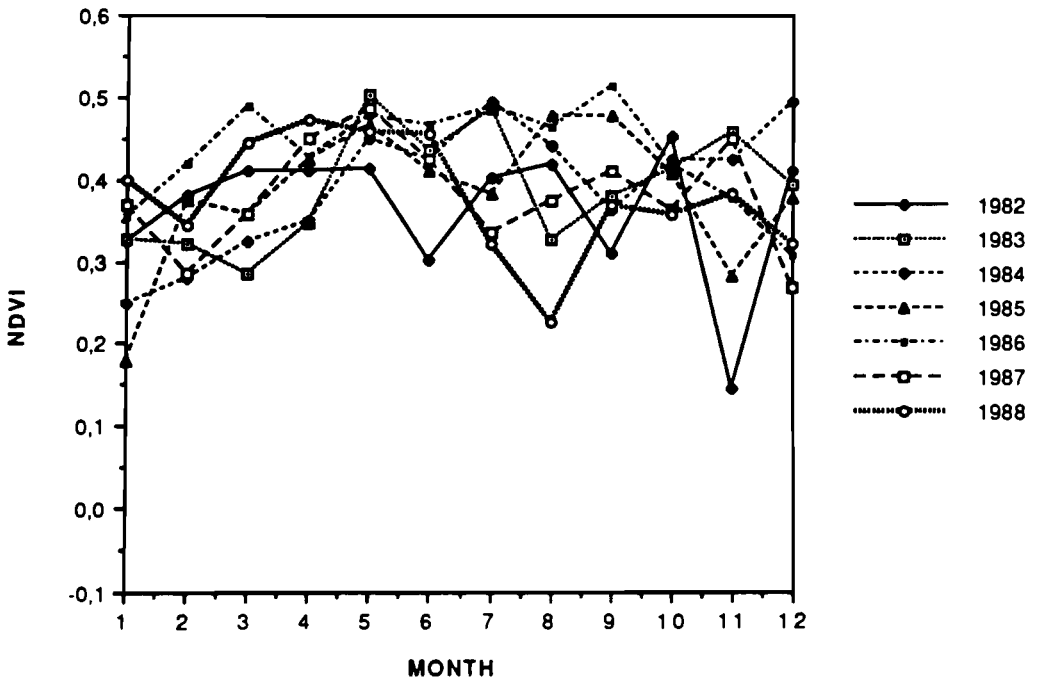
Site 13.


CONGO 5
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

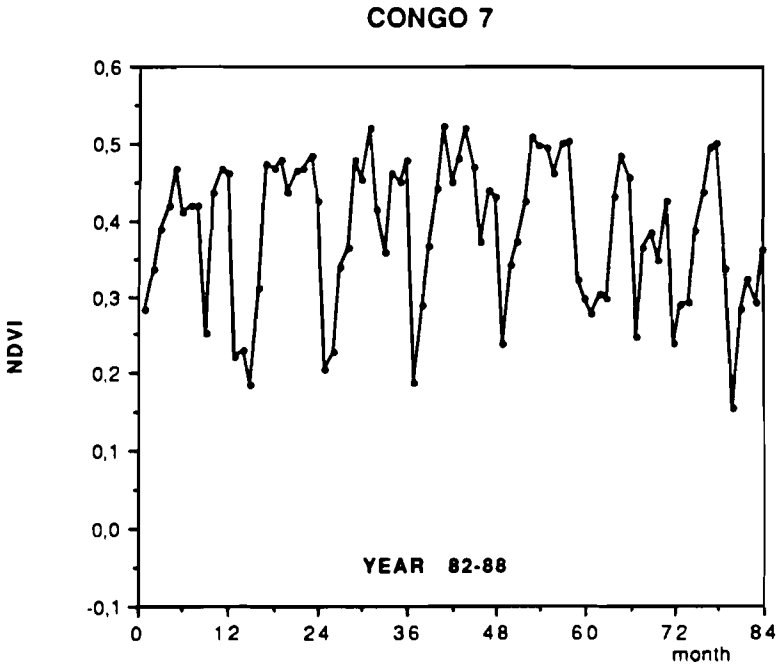

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

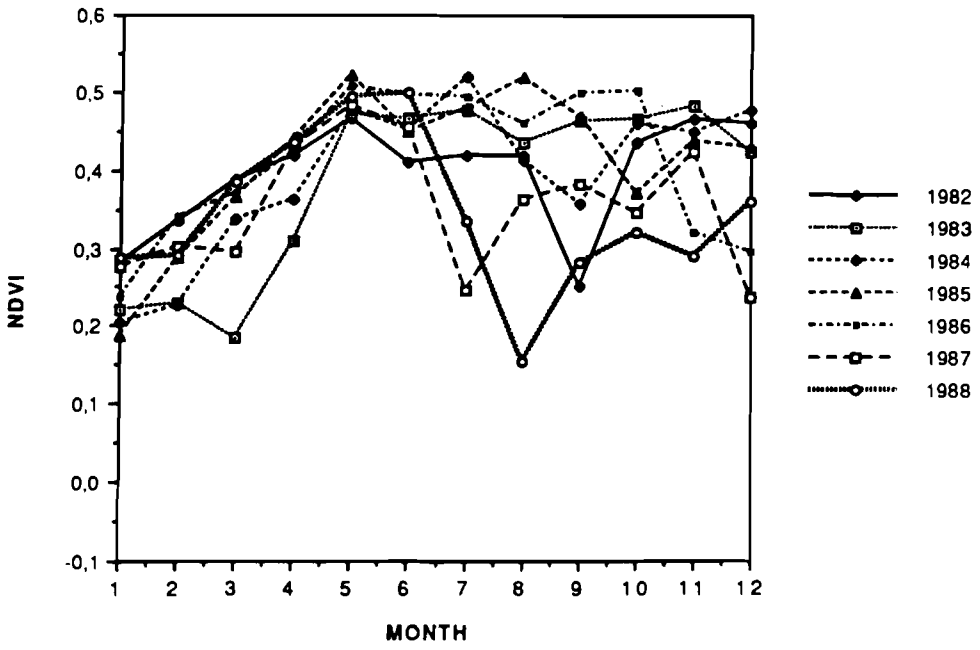
Site 15.


CONGO 6
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month

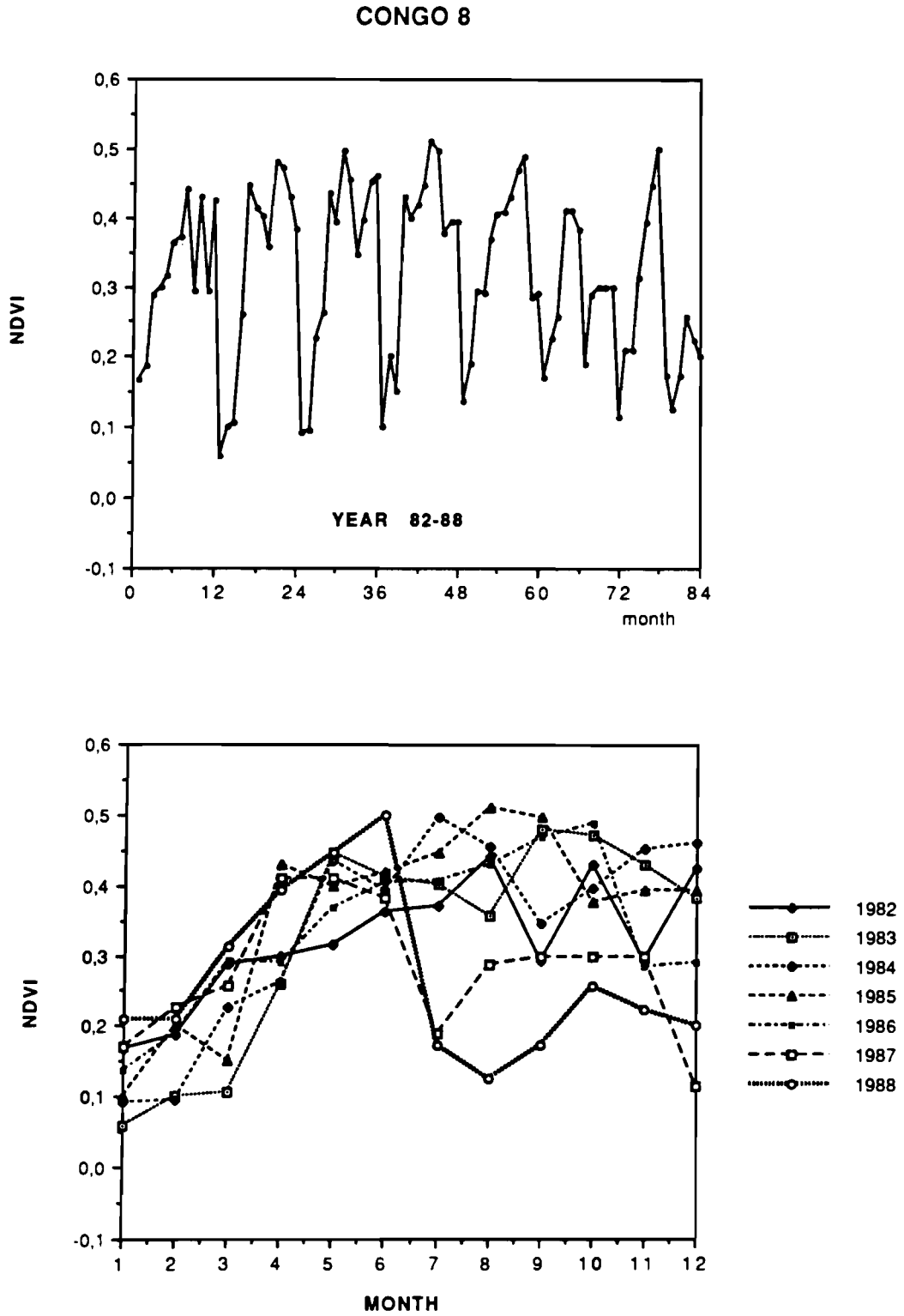

NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

Site 16.

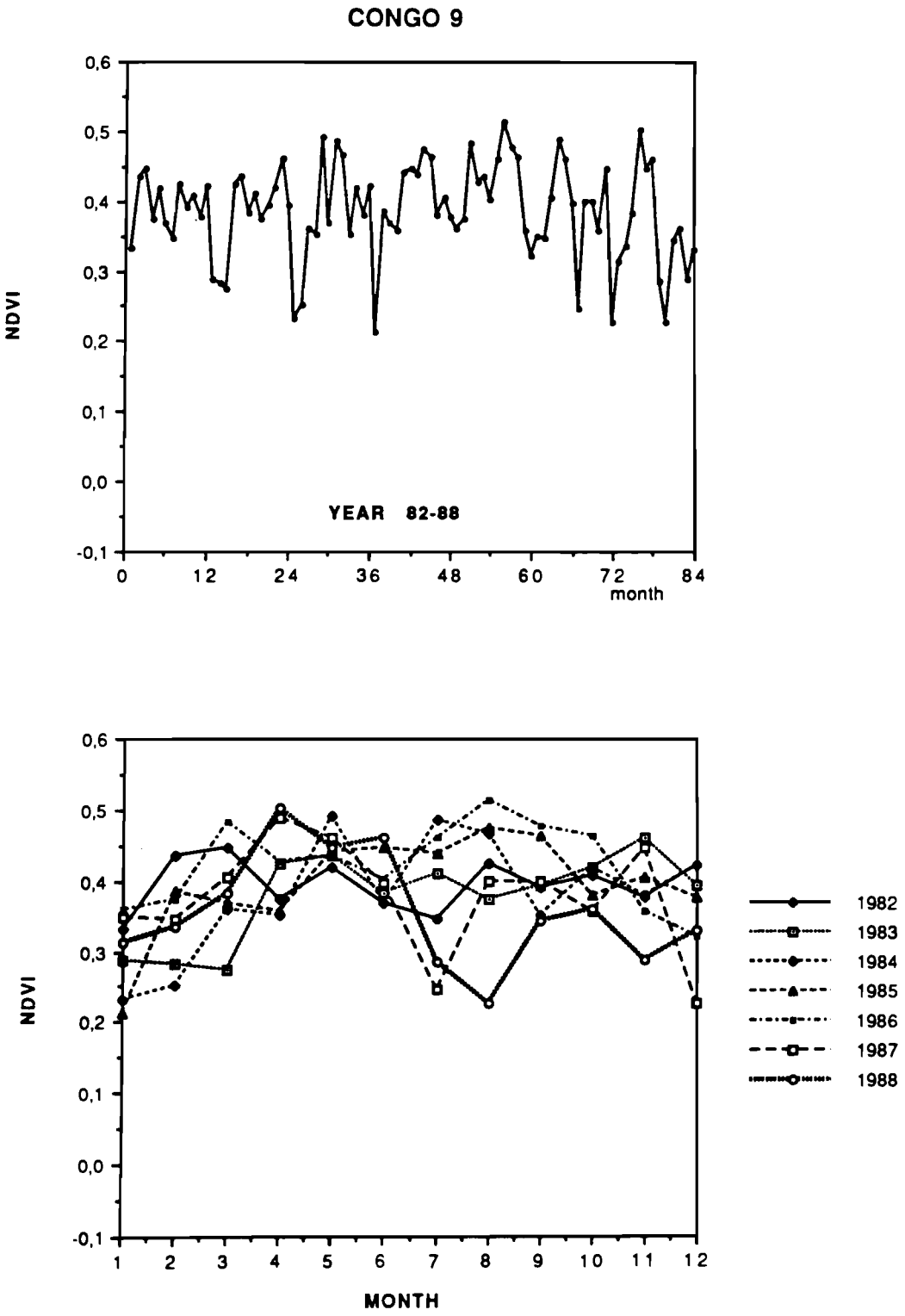

CONGO 7
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month


NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

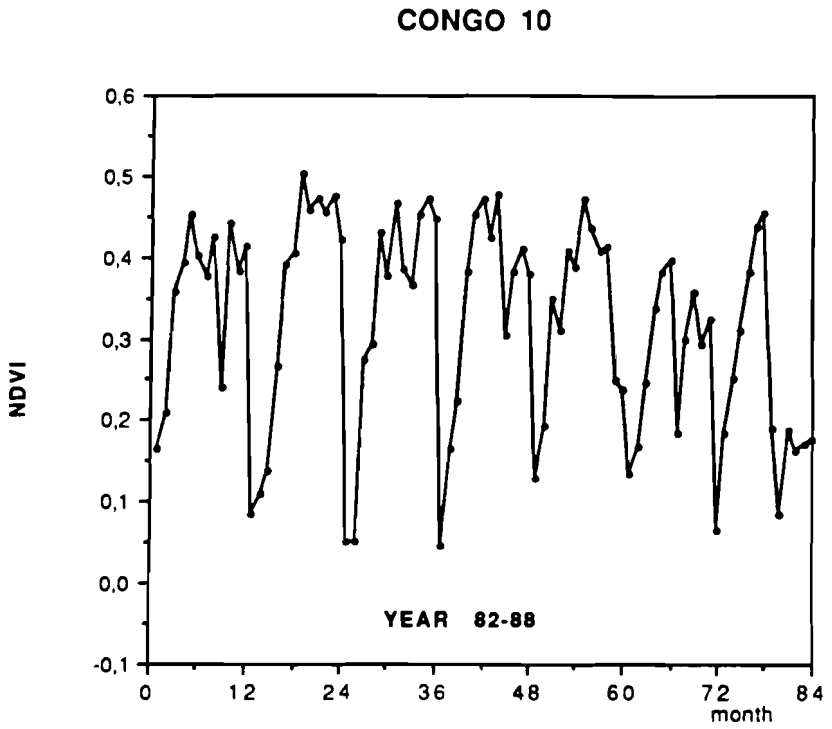
Site 17.

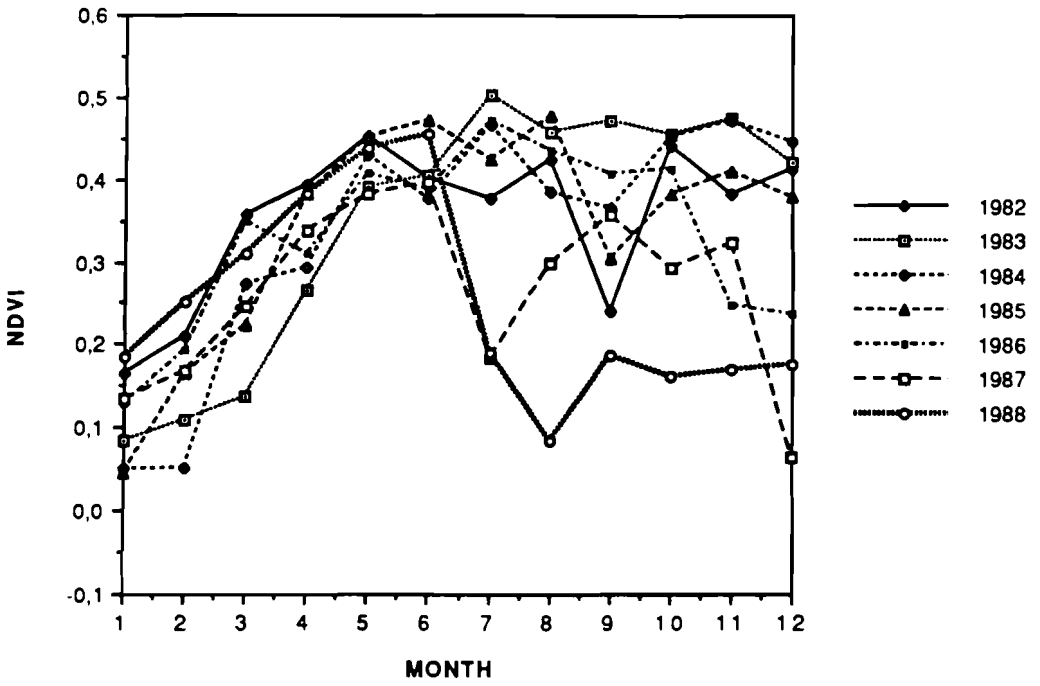

CONGO 8
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

Site 18.

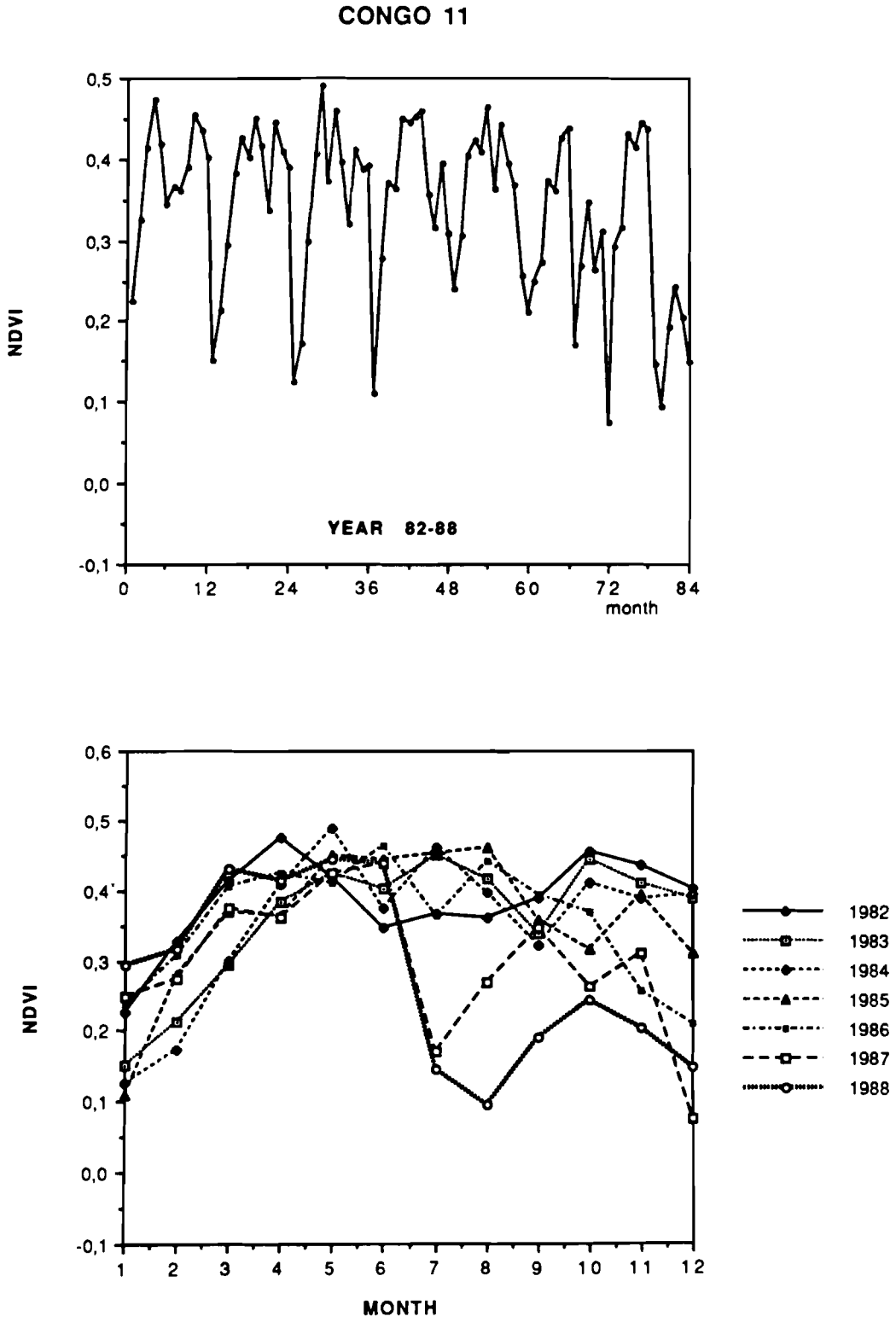

CONGO 9
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Site 19.

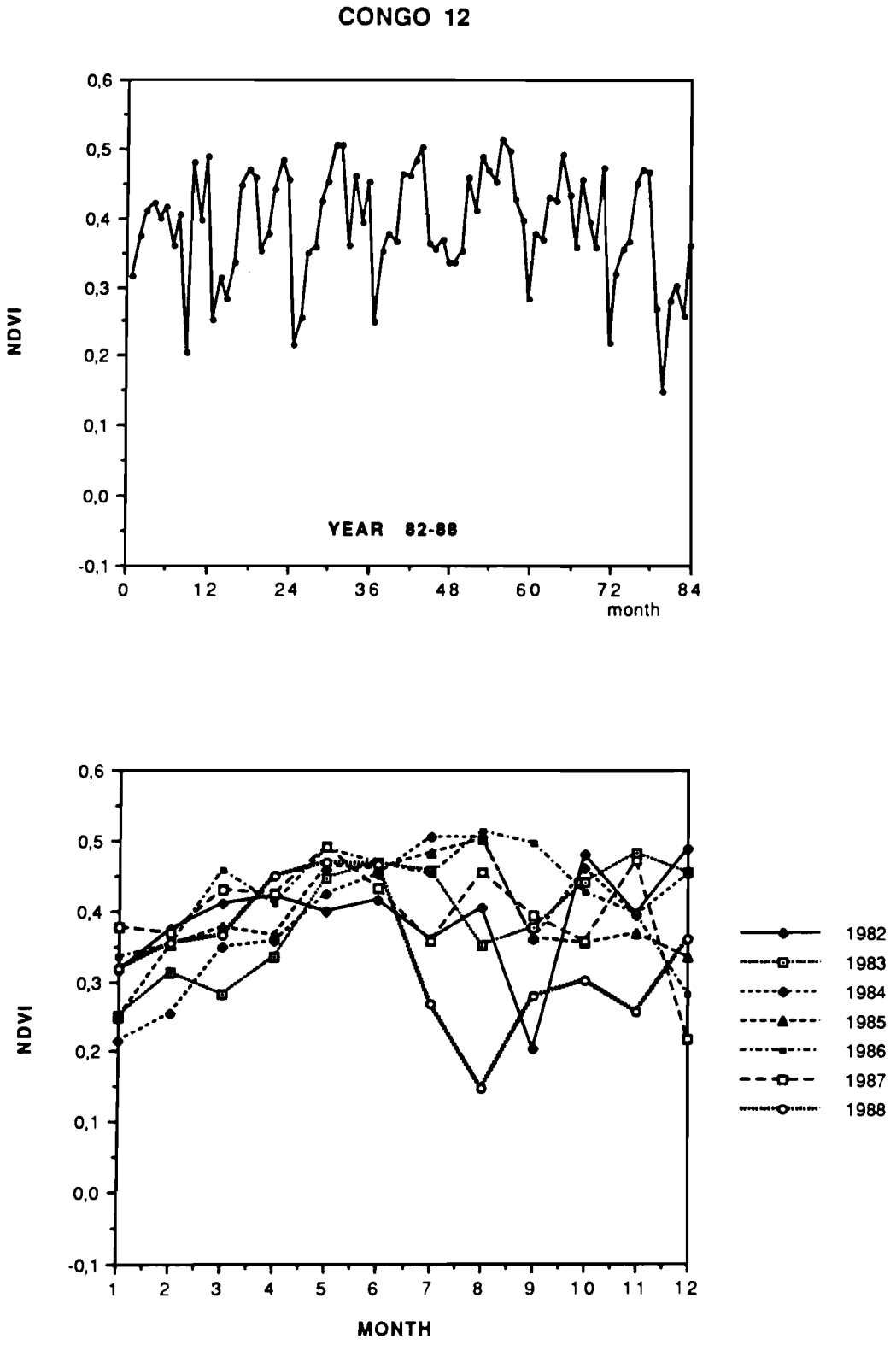

CONGO 10
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month


NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

Site 20.

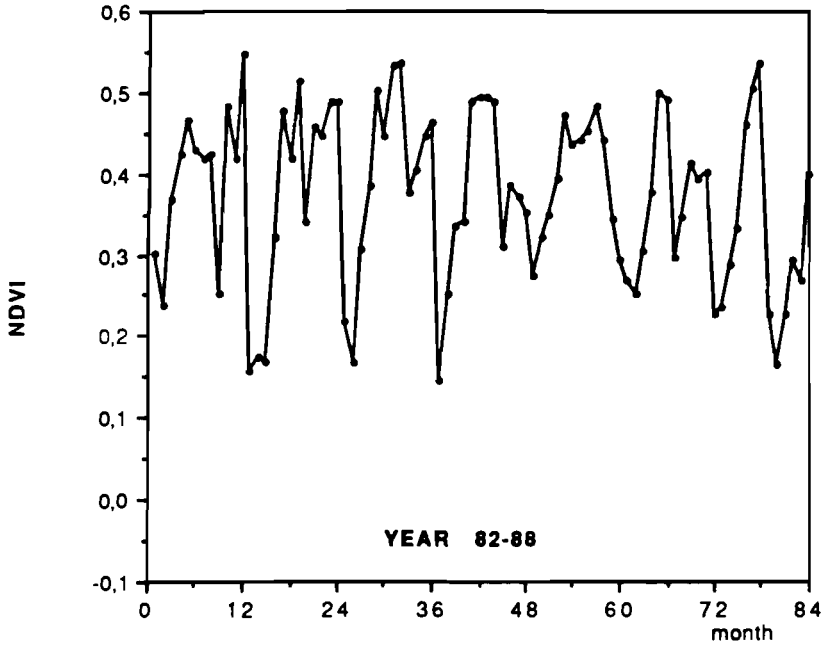

CONGO 11
NDVI
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
1
2
3
4
5
6
7
8
9
10
11
12
MONTH
1982
1983
1984
1985
1986
1987
1988

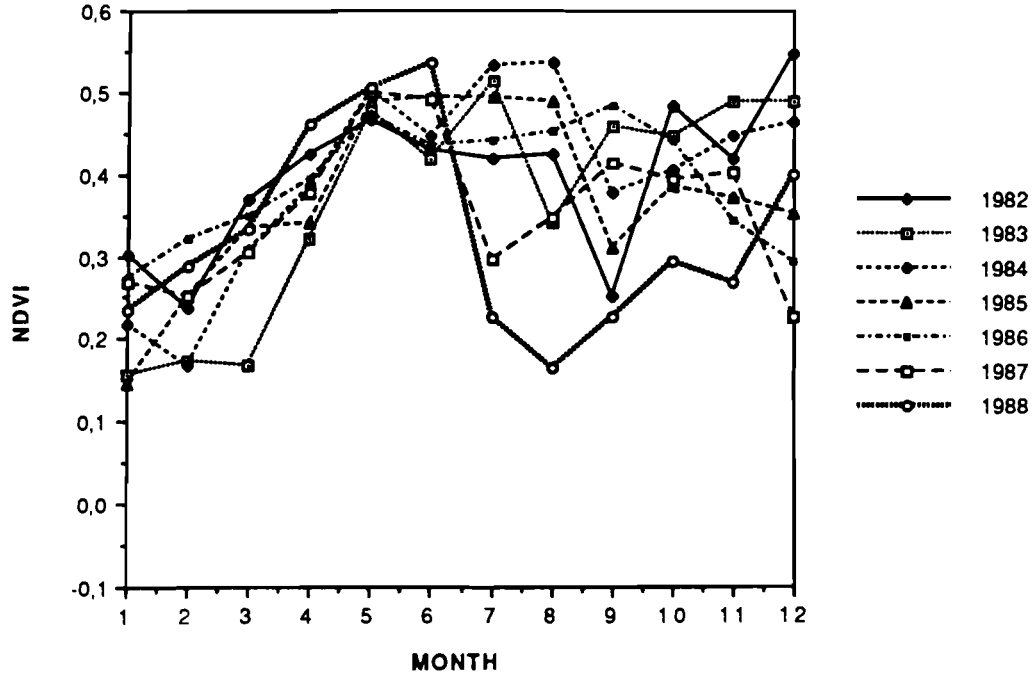
Site 21.


CONGO 12
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

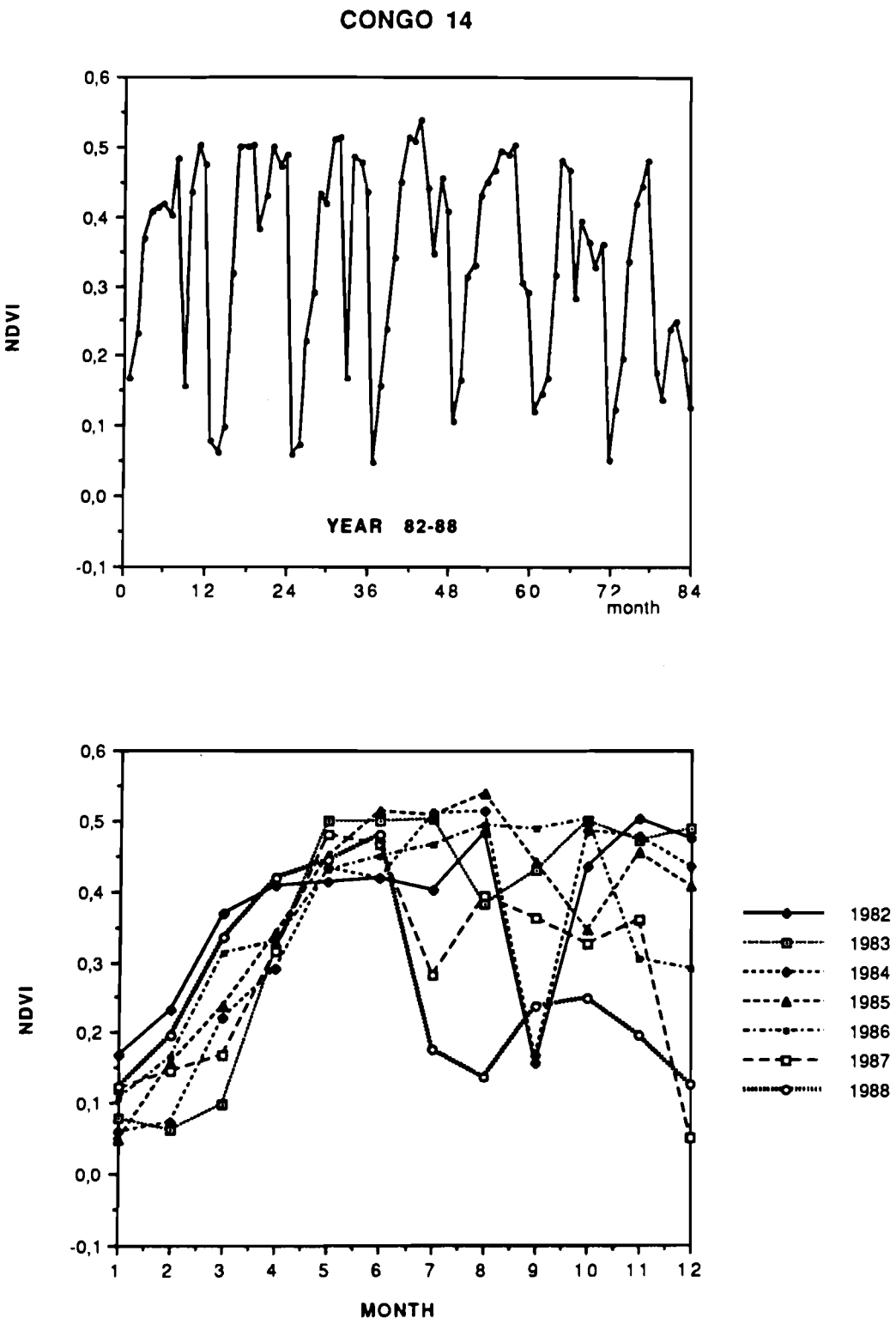
Site 22.

CONGO 13

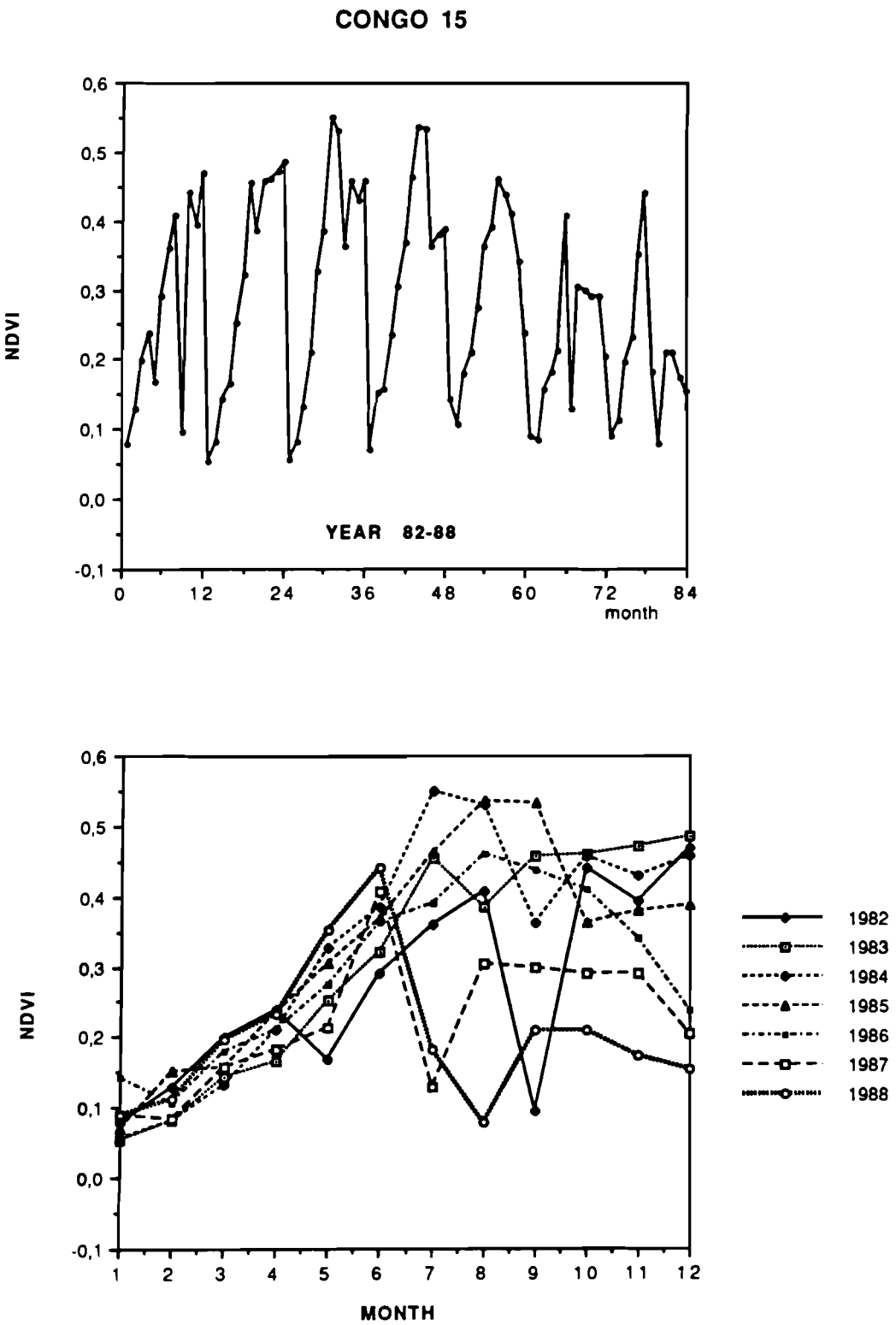




Site 23.

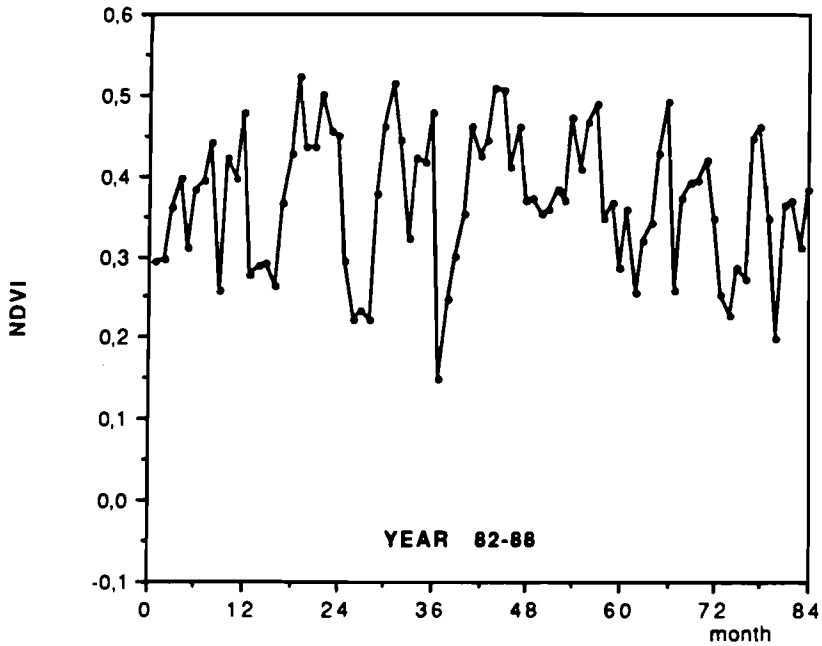

CONGO 14
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

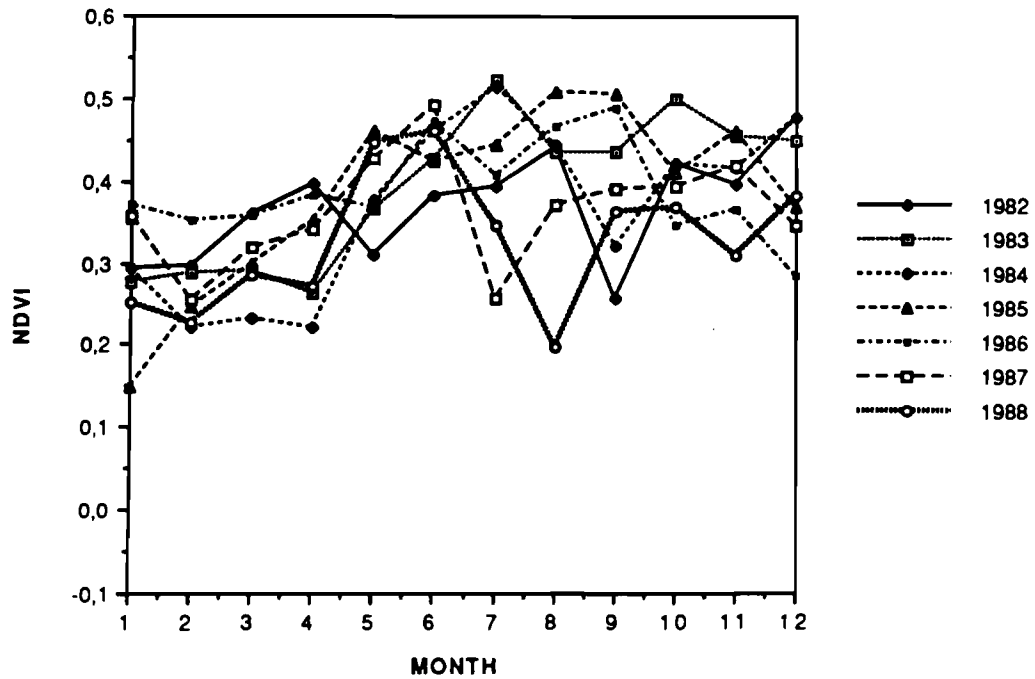
Site 24.


CONGO 15
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

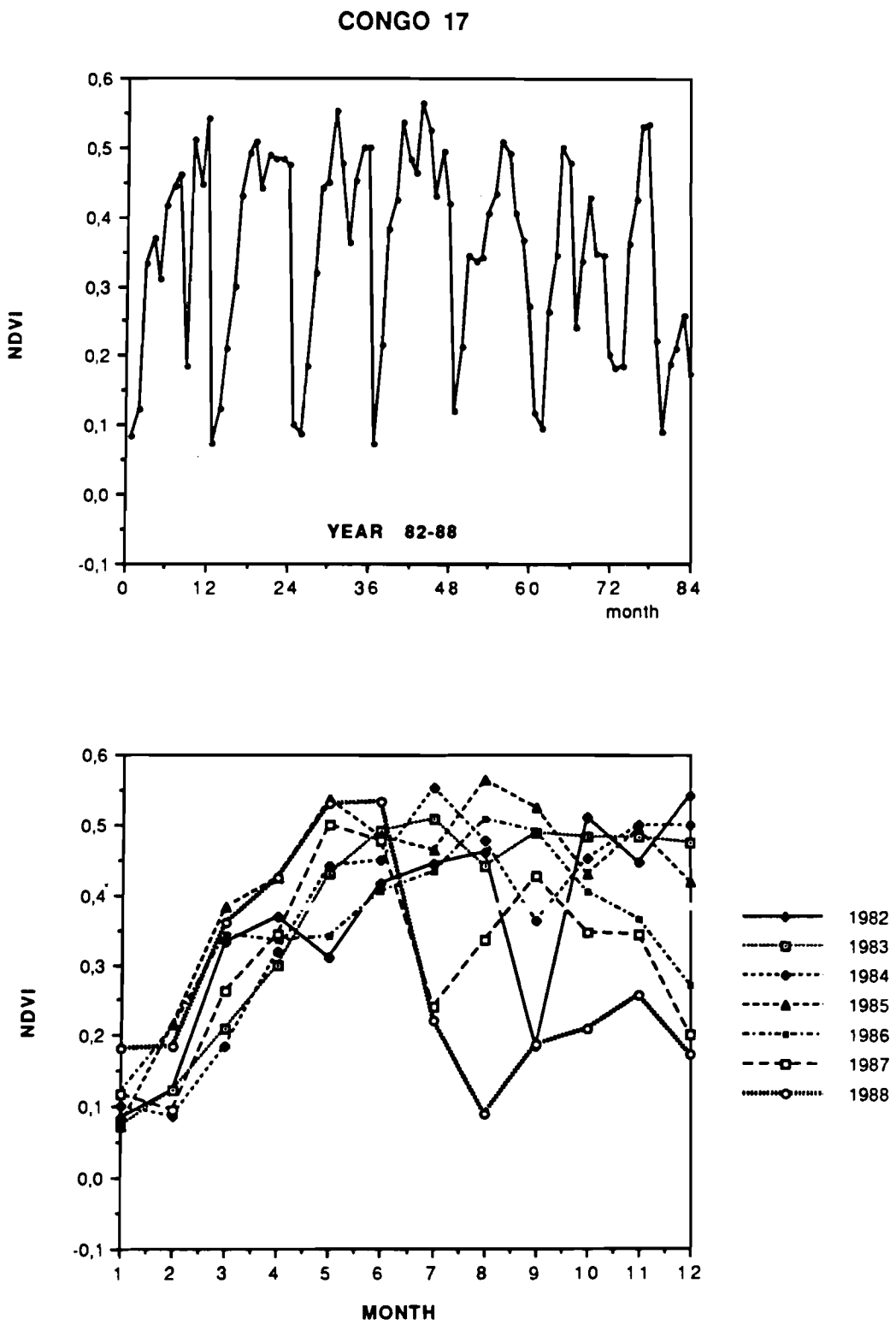
Site 25.

CONGO 16

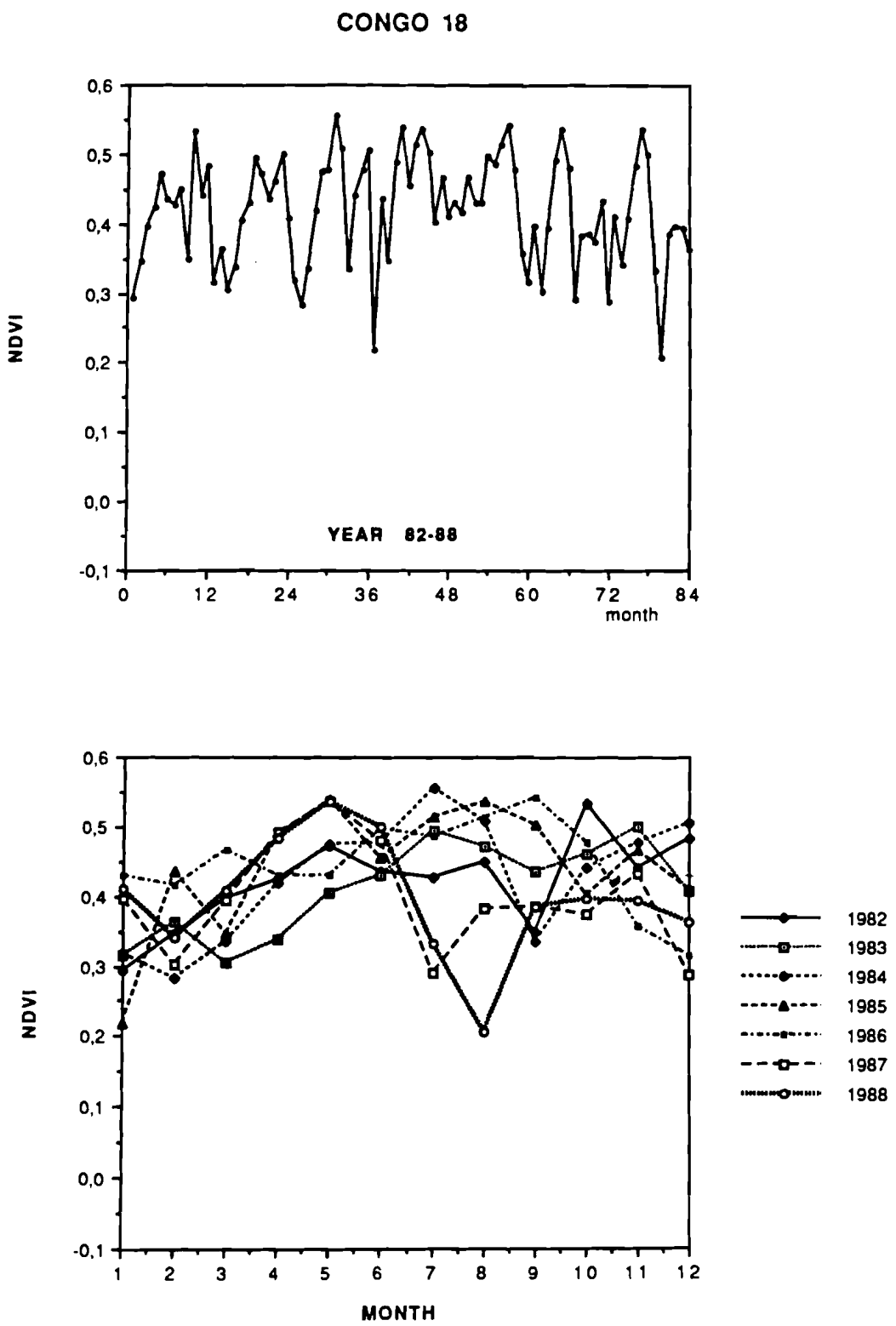




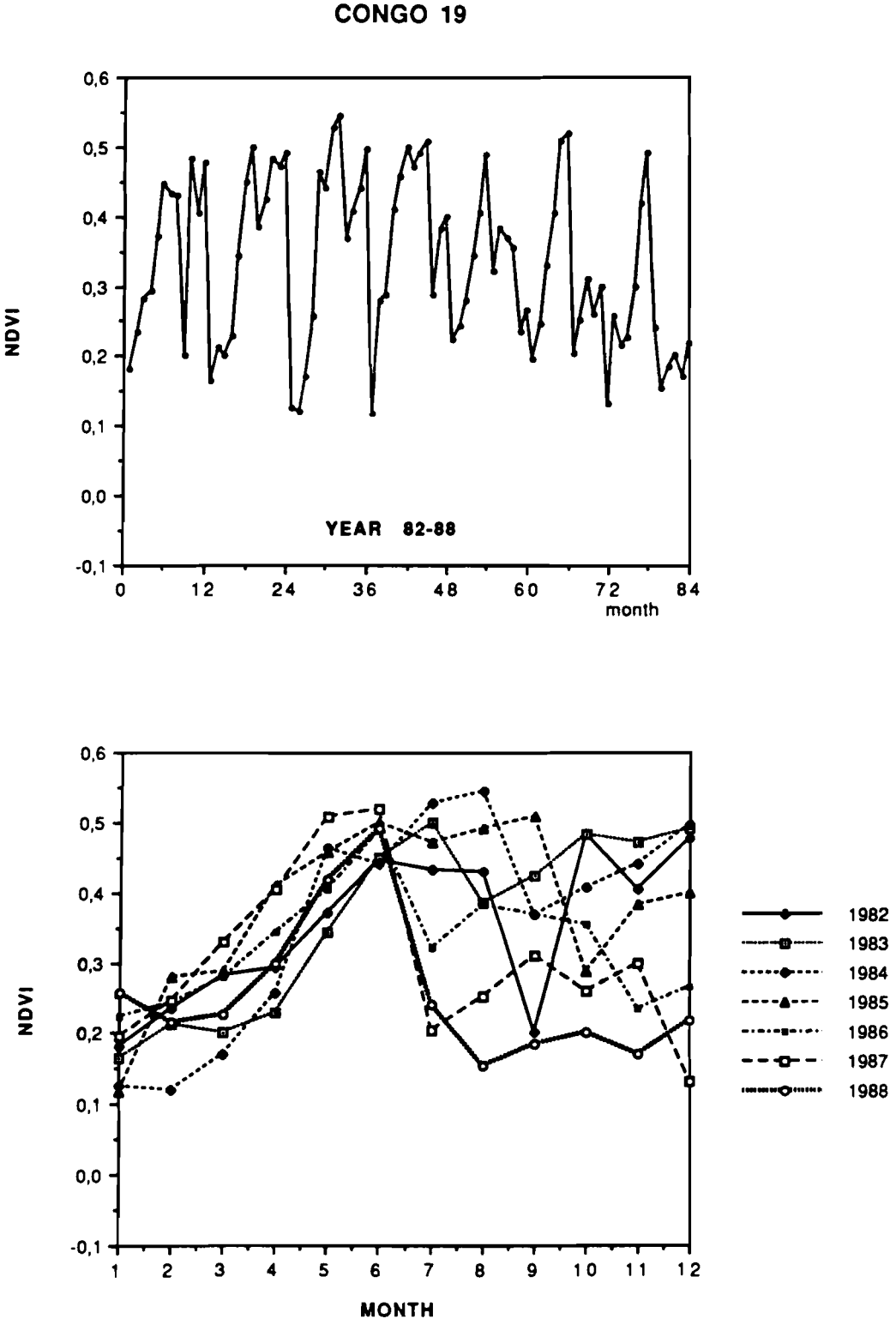
Site 26.


CONGO 17
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Site 27.


CONGO 18
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Site 28.


CONGO 19
NDVI
0,6
0,5
0,4
0,3
0,2
0,1
0,0
-0,1
YEAR 82-88
0
12
24
36
48
60
72
84
month
NDVI
MONTH
1
2
3
4
5
6
7
8
9
10
11
12
1982
1983
1984
1985
1986
1987
1988

Site 29.