242 (1)

Nonlinear Path Analysis with Optimal Scaling

Jan de Leeuw
Department of Data Theory FSW
University of Leiden

Paper presented at the NATO Advanced Research Workshop on Numerical Ecology. Station Marine de Roscoff, Brittany, France. June 3-11, 1986.

Path analysis: general idea [History)

abundance — environmental gradients.

$$\begin{cases} y = \beta_1 x_1 + \cdots + \beta_m x_m + \epsilon \\ \epsilon \perp (x_1, \ldots, x_m). \end{cases}$$

Or:

abundance of species i

$$y_i = \beta_{i1} x_1 + \cdots + \beta_{im} x_m + \epsilon_i$$

$$\epsilon_i \perp (x_1, \ldots, x_m).$$

[perhaps some β_{ij} are known to be zero]

Or:

$$y = \beta_1 x_1 + \cdots + \beta_m x_m + \epsilon,$$

$$x_m = \alpha_{m1} x_1 + \cdots + \alpha_{m,m-1} x_{m-1} + \delta_m,$$

$$x_{m-1} = \alpha_{m1} x_1 + \cdots + \alpha_{m,m-2} x_{m-2} + \delta_{m-1}, \ldots$$

$$\ldots$$

$$\epsilon \perp (x_1, \ldots, x_m)$$

$$\delta_m \perp (x_1, \ldots, x_{m-1})$$

$$\ldots$$

perhaps some α_{ij} are known to be zero [or equal to each other].

example: time series

$$x_m = \alpha x_{m-1} + \delta_m$$

$$x_{m-1} = \alpha x_{m-2} + \delta_{m-1}$$

$$\ldots$$

First methodological point

Why introduce this structure (zero regression coeffs) (structure within the predictors)?

Answer.

The models are special, restricted case of models in which every variable in the system depends on all other variables. These very general models are not restrictive [M = X].

Reframed question

Why use restrictive models?

Answer.
- Trade-off of bias and precision
- Curse of dimensionality.

Second methodological point

Are these structural models $\boxed{\text{causal}}$ models?

Answer. ??

(a) you can formulate them in causal language
[X influences, explains, determines Y]

(b) you can interpret them in causal language
[the effect of X on Y is .74]

(c) but this has nothing to do with

$$\left.\begin{cases}\text{necessary connections}\\ \text{productivity}\\ \text{manipulation}\end{cases}\right\}\ \text{HUME}$$

(d) Thus the word causal is used in a colloquial, impr
sense, or in a statistical, precise sense. These two
uses must be kept strictly separate. [DANGER]

Third methodological question

What is non linear?

It is not

$$y = \phi(x_1, \dots, x_m) + \epsilon$$

but it is

$$\begin{cases} y = \phi_1(x_1) + \cdots + \phi_m(x_m) + \epsilon \\ \phi_j \in \Phi_j \end{cases}$$

as in the previous lecture.

Perhaps quasilinear, semilinear is better.

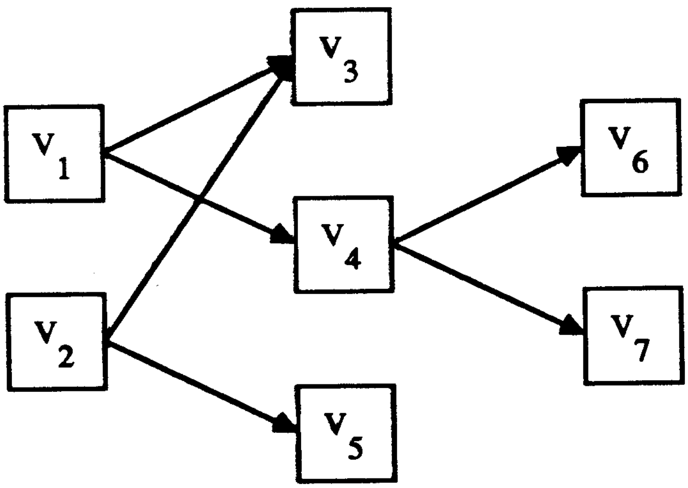


Diagram 1:
path diagram

	level	causes	direct causes	predecessors
Var 1	0	****	****	****
Var 2	0	****	****	****
Var 3	1	{1,2}	{1,2}	{1,2}
Var 4	1	{1}	{1}	{1,2}
Var 5	1	{2}	{2}	{1,2}
Var 6	2	{1,4}	{4}	{1,2,3,4,5}
Var 7	2	{1,4}	{4}	{1,2,3,4,5}

Table 1:

causal relations in Diagram 1

Example

(a) where do you get your arrows from?

- prior knowledge (theory)
- time and space

(b) colloquial interpretation.

- V_3 does not influence V_6
- V_1 influences V_6 through V_4

(c) ~~rigorous~~ - definitions [qualitative]

- graph theory

Table 1

exogeneous - endogeneous

transitive - recursive

Quantitative translation

$$\begin{cases} x_3 = \beta_{31} x_1 + \beta_{32} x_2 + \epsilon_3 \\ x_4 = \beta_{41} x_1 + \epsilon_4 \\ x_5 = \beta_{52} x_2 + \epsilon_5 \\ x_6 = \beta_{64} x_4 + \epsilon_6 \\ x_7 = \beta_{74} x_4 + \epsilon_7 \end{cases}$$

weak orthogonality

$$\left.\begin{array}{l} \epsilon_3 \perp (x_1, x_2) \\ \epsilon_4 \perp x_1 \\ \epsilon_5 \perp x_2 \\ \epsilon_6 \perp x_4 \\ \epsilon_7 \perp x_4 \end{array}\right\} \text{saturated model.}$$

strong orthogonality

(a) disturbances uncorrelated with exogeneous

$$\{\epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6, \epsilon_7\} \perp \{x_1, x_2\}$$

(b) disturbances different levels uncorrelated

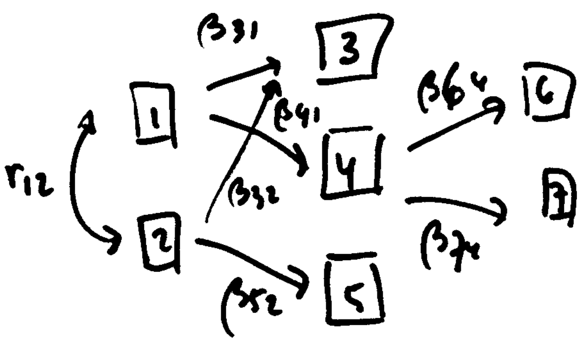
$$\{\epsilon_3, \epsilon_4, \epsilon_5\} \perp \{\epsilon_6, \epsilon_7\}.$$

Main results under strong orthogonality

(1) Causal interpretation

Given the direct causes a variable is independent of its other predecessors

(b) Calculus of path coefficients



$$r_{15} = r_{12}\beta_{52}$$

$$r_{26} = r_{12}\beta_{14}\beta_{64}$$

$$r_{34} = \beta_{41}\beta_{31} + \beta_{41}r_{12}\beta_{23}$$

useful decompositions

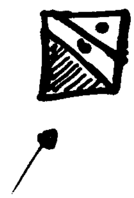
$$r_{13} = \underset{\text{direct}}{\beta_{31}} + \underset{\text{indirect}}{\beta_{32}r_{12}}$$

(c) ML = LS

Classes of models

1. **Recursive models** — path calculus; cond. independence

   (a) Saturated recursive models

   - block recursive models
   - causal chains

     $y = Ax + By + \varepsilon$

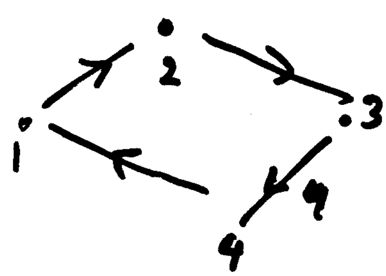
     exogeneous
   - multiple regression

   Strong orthogonality $\equiv$ weak orthogonality

   (b) nonsaturated recursive models

   [fig 1]. strong $\neq$ weak

2. **Nonrecursive models**

   weak $\not\Rightarrow$ saturated

   [no path calculus
   no cond. independence

   [no level assignment

$$r(x_1,y) = \beta_1 + \beta_2 r_{12} + \beta_3 r_{13}$$
$$r(x_2,y) = \beta_1 r_{12} + \beta_2 + \beta_3 r_{23}$$
$$\cdot \; - \; - \; - \; - \; - \; - \; -$$

$$\left(\begin{array}{l} r = R\beta \\ \beta = R^{-1} r. \end{array} \right)$$

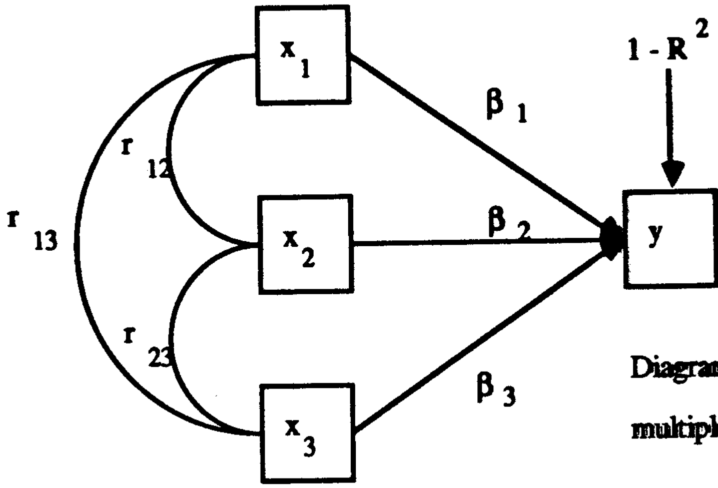


Diagram 2:
multiple regression model

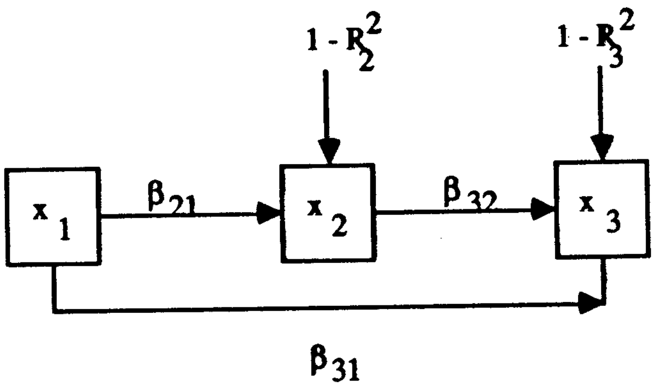


Diagram 3:
a simple
saturated recursive model

$$\beta_{31} = 0 \iff r_{13|2} = 0.$$

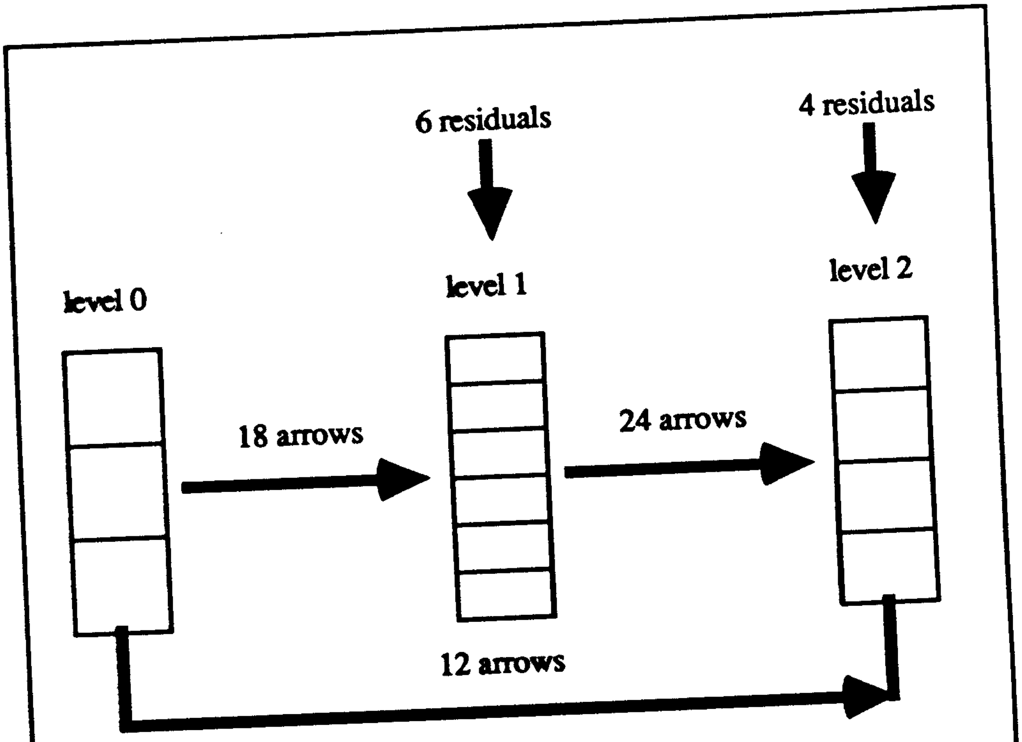


Diagram 4:

general recursive saturated model

Computation

Recursive models

LS [projection] [on direct causes]

Nonrecursive models

[Weak orth : LS]

strong or other : ??

Example from L.&L.

- K: primary production
- C: Chlorophyll
- S: Salinity
- T: Temperature.

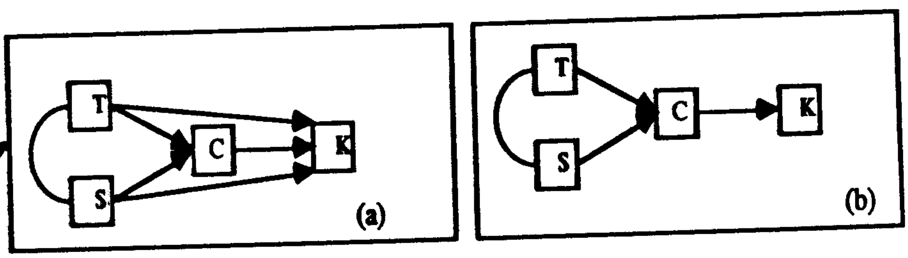
Alternative models : fit

	K	C	T
C	+.842		
T	+.043	+.236	
S	−.146	−.369	−.925

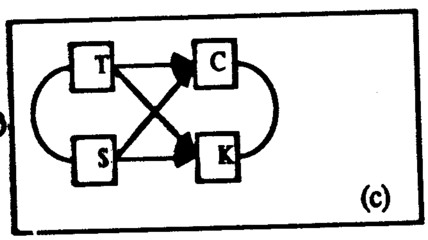
Correlations
Baie des Chaleurs

non sar

Sar



Sar



Three recursive models

Table 1:
Legendre and Legendre
Primary Production Data

	(a)	(b)	(c)
T ⇒ C	-0.730	-0.730	-0.730
S ⇒ C	-1.044	-1.044	-1.044
T ⇒ K	+0.031	*****	-0.638
S ⇒ K	+0.220	*****	-0.736
C ⇒ K	+0.916	+0.842	*****
V ERR C	0.787	0.787	0.787
V ERR K	0.260	0.291	0.920
C ERR C,K			0721

Seems fine

Latent variables

(a) to extent the notion of a path model (and make it cover other existing techniques)

(b) to model the idea of measurement error or of indicators

Examples in figure.

General example.

LISREL (Jöreskog)

EQS (Bentler)

COSAN (McDonald)

$$\chi = A\eta + B\chi + \xi$$ ← errors in equations

$$x = C\chi + \epsilon$$

$$y = D\eta + \delta$$

errors in variables.

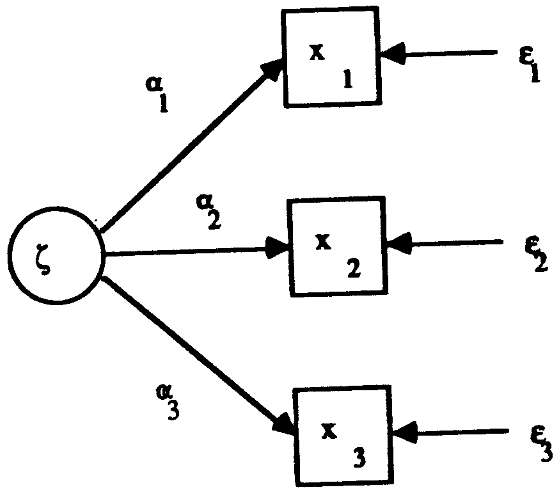


Diagram 5:
one-factor model

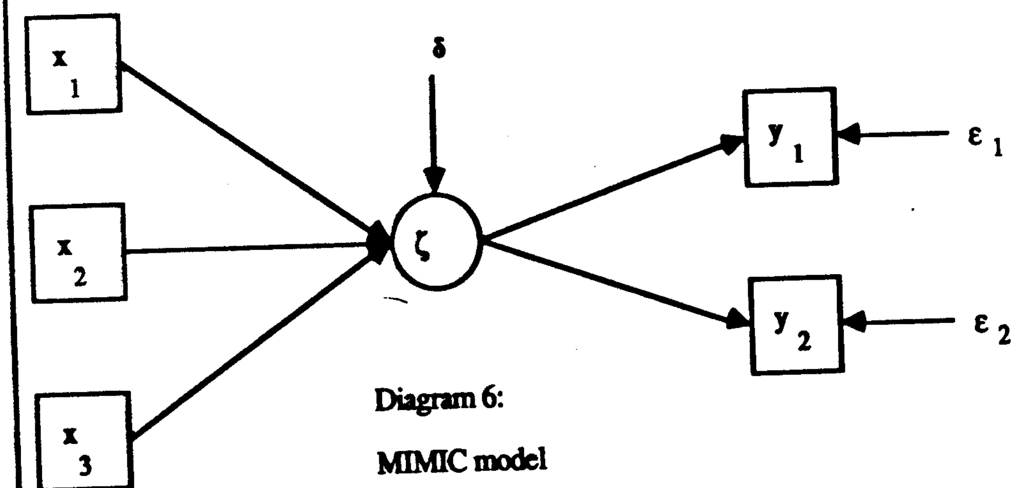


Diagram 6:
MIMIC model

Generalizes to more factors

Linear dynamic systems

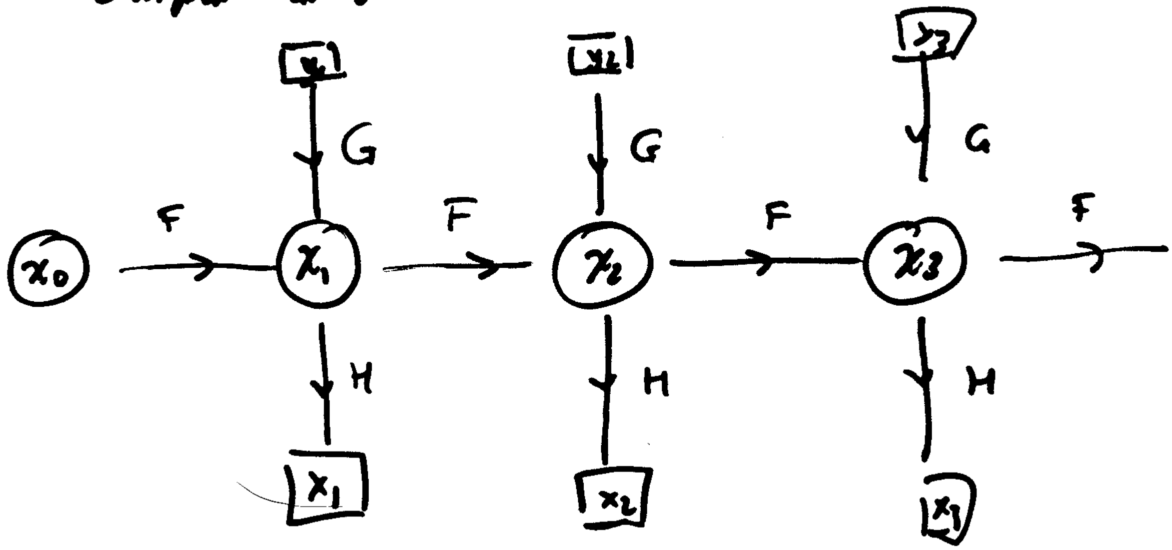
input or background at t

$$\chi_t = F\chi_{t-1} + G y_t + \epsilon_t$$

$$x_t = H\chi_t + \delta_t$$

state of system at t

output at t



- General AR-model
- Kalman filter
- Also space
- Also nonstationary

And now the most original part PATHALS

Step 1: Take an arrow diagram

Step 2: Translate it into regressions [linear structural equations]

Step 3: Translate these into a loss function

Step 4: Minimize over

(a) path coefficients

(b) latent variables, transformations, errors.

Example MIMIC from (6)

~~Loss~~ $\cancel{= \| \xi - \sum \alpha_j x_j \|^2 + \sum \|}$

$$\text{loss} = \| \sum \alpha_j x_j - \xi - \delta \|^2 + \sum \| y_k - \rho \xi - \epsilon_k \|^2$$

weak : $\delta \perp (x_1, x_2, x_3)$
$\xi \perp (\epsilon_1, \epsilon_2)$

Perhaps transform. ALS algorithm.

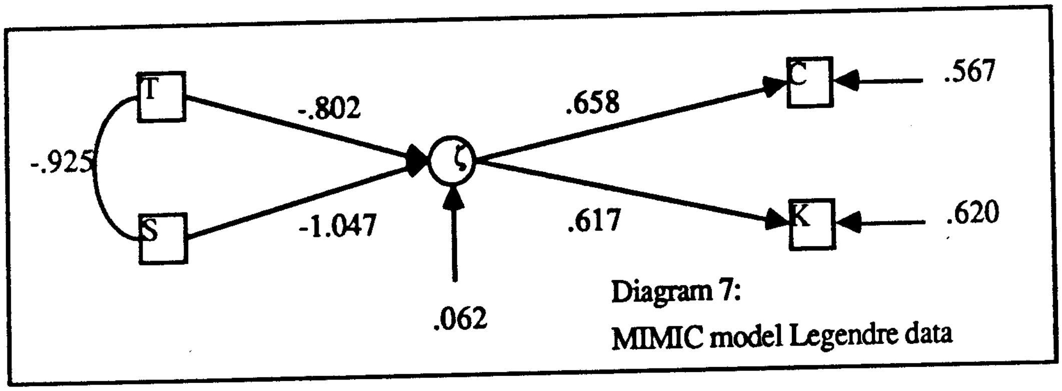


Diagram 7:
MIMIC model Legendre data

	weights metric		weights nonmetric		explained variances metric	explained variances nonmetric
WC	-.82	.10	-.96	-.20		
BS	-.06	.11	-.56	.39		
CM	.13	.21	-.14	-.33		
LR	-.02	.56	.28	.12		
FT	.72	-.24	.22	-.15		
CH	-.29	.10	-.71	.43		
S1	-.79	-.09	-.89	.22	.39	.21
S2	.04	-.79	.30	-.87	.36	.21
S3	-.85	-.35	-.88	-.16	.22	.16
S4	-.95	-.10	-.99	.21	.13	.04
S5	-.97	-.06	-.99	.22	.08	.04
S6	-.91	-.13	-.95	.19	.21	.10
S7	-.93	-.48	-.98	.01	.07	.04
S8	-.77	-.11	-.85	.00	.43	.27
S9	-.36	.52	.74	-.48	.53	.32
S10	.18	.88	.07	.90	.25	.16
S11	.52	.71	.48	.71	.36	.18
S12	.53	.53	.57	.54	.54	.31

Table 2: hunting spider data: metric and nonmetric MIMIC analysis

The important thing in PATHALS

- scaling of variables, including latent variables
- latent variables are like other variables, but with a very low measurement level

Instead of

$$\begin{pmatrix} \text{linear} \\ \text{latent} \end{pmatrix} \qquad \qquad \begin{pmatrix} \text{linear} \\ \text{monotonic} \\ \text{polynomial} \\ \vdots \\ \text{latent} \end{pmatrix}$$

<u>Example</u> Hunting spider data

[and earlier

$$\begin{tikzcd} \text{SPECIES} \arrow[dr] \arrow[d, leftrightarrow, bend right] & & \\ \text{NITRO} \arrow[r] & \text{YIELD} & E \arrow[l] \end{tikzcd}$$

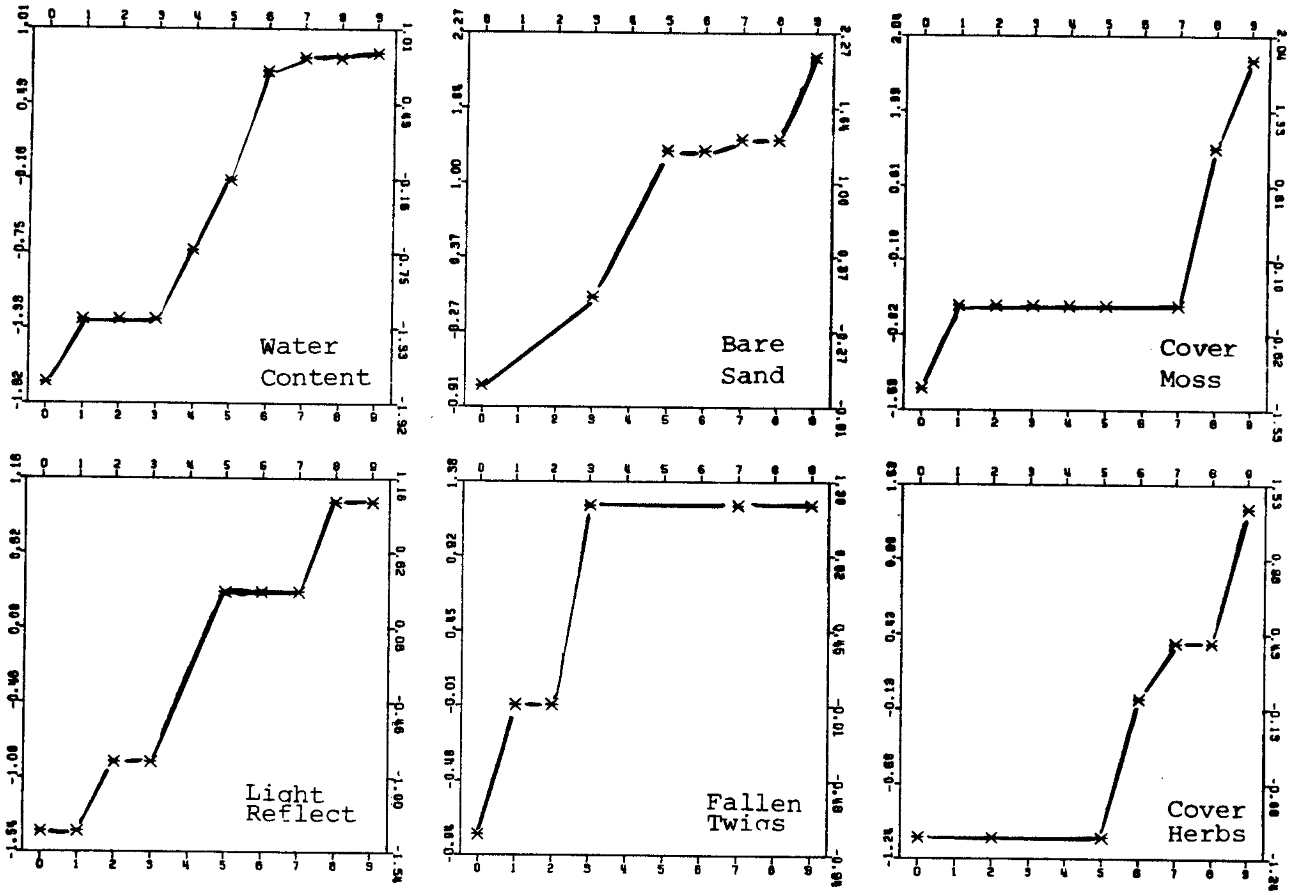
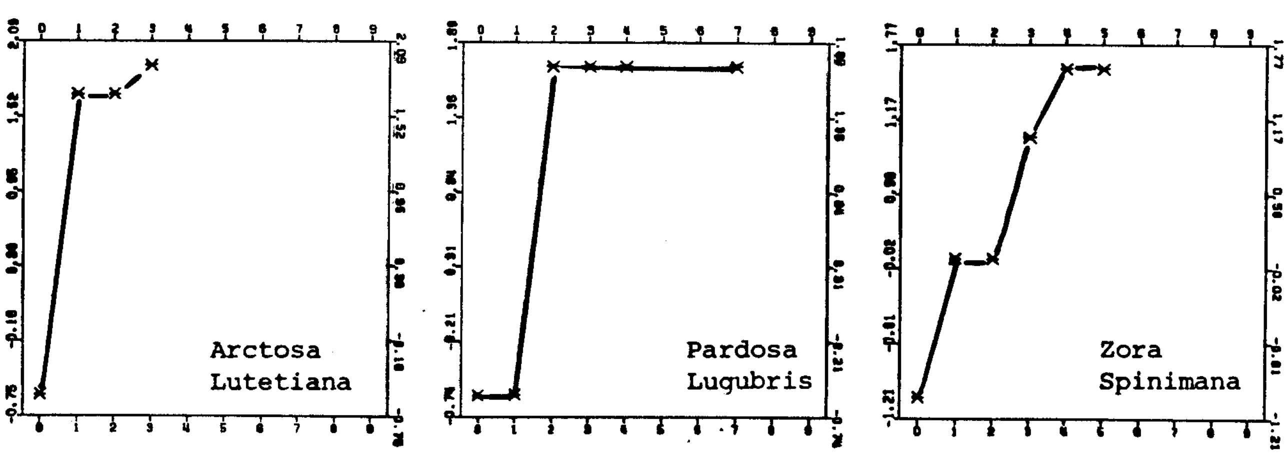


Diagram 8a:
hunting spider example
transformations of environmental variables



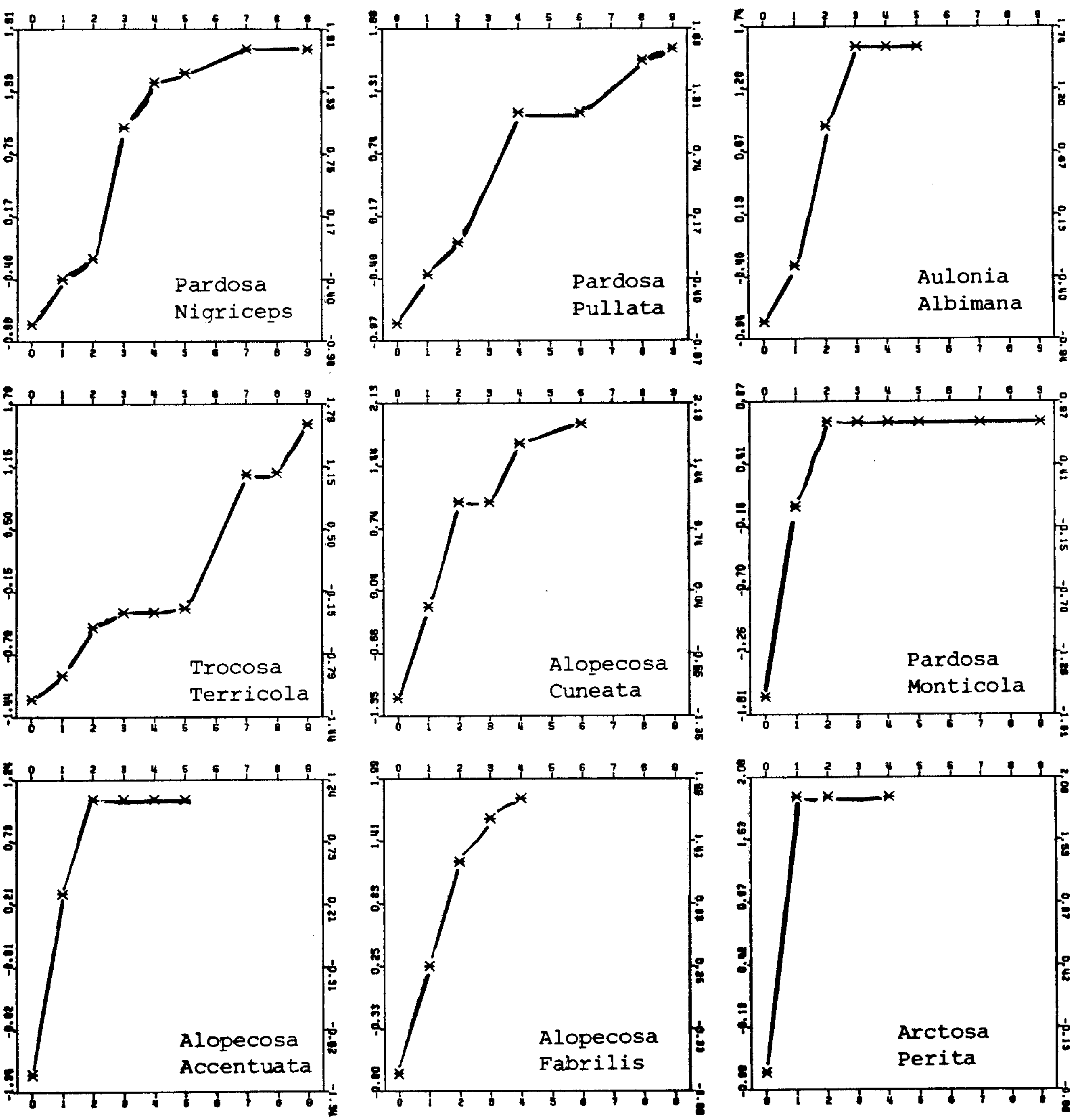


Diagram 8b:
hunting spider example
transformations of abundance variables