PATH ANALYSIS WITH OPTIMAL SCALING

Jan de Leeuw
Department of Data Theory FSW, University of Leiden
Middelstegracht 4
2312 TW Leiden, The Netherlands

Abstract - In this paper we discuss the technique of path analysis, its extension to structural models with latent variables, and various generalizations using optimal scaling techniques. In these generalizations nonlinear transformations of the variables are possible, and consequently the techniques can also deal with nonlinear relationships. The precise role of causal hypotheses in this context is discussed. Some applications to community ecology are treated briefly, and indicate that the method is a promising one.

INTRODUCTION

In this paper we shall discuss the method of *path analysis*, with a number of extensions that have been proposed in recent years. The first part discusses path analysis in general, because the method is not very familiar to ecologists. In fact we have been able to find only very few papers using path analysis in the literature of community ecology. With the help of Pierre and Louis Legendre we located Harris and Charleston (1977), Chang (1981), Schwinghamer (1983), Gosselin *et al.* (1986), and Troussellier *et al.* (1986).

In this paper we combine classical path analysis models, first proposed by Wright (1921, 1934), with the notion of latent variables, due to psychometricians such as Spearman (1904) and to econometricians such as Frisch (1934). This produces a very general class of models. If we combine these models with the notion of least squares optimal scaling (or quantification, or transformation), explained in De Leeuw (1987), we obtain a very general class of techniques.

Now in many disciplines, for example in sociology, these path analysis techniques are often discussed under the name *causal analysis*. It is suggested, thereby, that such techniques are able to discover causal relationships that exist between the variables in the study. This is a rather unfortunate state of affairs (De Leeuw 1985). In order to discuss it more properly, we must start the paper with some elementary methodological discussion.

One of the major purposes of data analysis, in any of the sciences, is to arrive at a convenient description of the data in the study. By 'convenient' we mean that the data are described parsimoneously, in terms of a relatively small number of

NATO ASI Series, Vol. G 14
Developments in Numerical Ecology
Edited by P. and L. Legendre

parameters. If possible this description should be linked as tightly as possible to existing scientific theory, and consequently the parameters should not be merely descriptive, but they must preferably be part of a model for the phenomenon that is studied. This makes it possible to communicate efficiently, and to fit the results into an existing body of theory. Fitting data into existing theory, or creating new theory to incorporate the data, is called *explanation*. If the theory is formulated in terms of if-then relationships, or more generally in terms in functional relationships, then we can call this explanation *causal*.

Thus causality is interpreted by us as a way of formulating theories, a way of speaking about the world. Whether everything, or almost everything, moves or develops deterministically according to causal laws is, from a scientific point of view, not an interesting question. It is an undeniable fact that everybody, including scientists, uses causal language all the time. It is also true, that in most contexts the word causality suggests a *necessary* connection, a notion of the cause *producing* the effect, and the idea that it must be possible to change the effect by *manipulating* the cause. This does not imply, as we sometimes hear, that causal connections can only be established by experimental methods. Causal connections, if they are necessary connections, cannot be established at all, in the same way as natural laws cannot be proven inductively. Causality is a figure of speech, and there is no need to 'establish' a figure of speech.

This does not mean, of course, that persons engaged in scientific discourse can afford to choose their terminology in a misleading and careless way. The word 'causality' has all the connotations we have mentioned above (necessity, productivity, manipulation), and if social scientists, for instance, want to use the word, they must realize that it has these connotations. If social scientists set out to prove that 'social economic status' causes 'school achievement', and 'school achievement' causes 'income', then they will have a hard time convincing others that they are using the word 'cause' in the same sense as somebody who says that putting a kettle of water on the fire causes it to boil.

We briefly mention some other points that are important in this connection. There has been a justifiable tendency in statistical methodology either to avoid the word 'cause' altogether, or to give it a precise meaning which does not necessarily have much to do any more with the common sense notion. Simon (1953) and Wold (1954), for instance, define 'causality' as a property of systems of linear regressions, some are causal and some are not. This is not very objectionable, although of course not without its dangers. A very important point of view, defended for example by Pearson (1911), is that causation is merely the limiting case of perfect correlation. This resulted from a conscious attempt, started by the Belgian astronomer Quetelet, to bring the laws of the social and life sciences on an equal footing with the laws of the physical sciences. Pearson eloquently argued that correlation is the more fundamental scientific category, because causality is merely a degenerate special case, which does not really occur in practice. Again this point of view is not inherently wrong, provided we broaden the definition of correlation sufficiently.

This is related to the fact that lawlike relationships in the social sciences and the life sciences are usually described as *probabilistic* in stead of *deterministic*. If we have ten kettles, and we put them on the fire, then the water will boil in six or seven of them. But this difference is mainly a question of choosing the appropriate unit. A probabilistic relationship between individual units is a deterministic relationship, in fact a functional relationship, between the random variables defined on these units. A linear regression between status and income is a deterministic relationship between averages, even though it does not make it possible to predict each individual income precisely from a known status-value. If we call a law-like relationship between the parameters of multivariate probability distributions a *correlation*, then Pearson's point of view about causality makes sense. Of course we must again be careful, because another far more specific meaning of the word 'correlation', also connected with the name of Pearson, is around too. Compare Tukey (1954) for more discussion on this point.

Up to now we have concentrated on data analysis as a method of description. We summarize our data, preferably in the context of a known or conjectured model which incorporates the prior information we have. At the same time we also investigate if the model we use describes the data sufficiently well. But science does not only consist of descriptions, we also need to make *predictions*. It is not enough to describe the data at hand, we must also make statements about similar or related data sets, or about the behaviour of the system we study in the future. In fact it is perfectly possible that we have a model which provides us with a very good description, for example because it has many parameters, but which is useless for prediction. If there are too many parameters they cannot be estimated in a stable way, and we have to extrapolate on a very uncertain basis. Or, to put it differently, we must try to separate the stable components of the situation, which can be used for prediction, from the unstable disturbances which are typical for the specific data set we happen to have.

We end this brief methodological discussion with a short summary. The words 'correlation' and 'causality' have been used rather loosely by statisticians, certainly in the past. Causal terminology has sometimes been used by social scientists as a means of making their results sound more impressive than they really are, and this is seriously misleading. It is impossible, by any form of scientific reasoning or activity, to prove that a causal connection exists, if we interpret 'causal' as 'necessary'. What we are really looking for is invariant functional relationships between variables, or between the parameters of multivariate probability distributions. These invariant relations can be used for prediction. The method of path analysis, that we shall discuss in detail below, has the specific advantage over other data analysis techniques that it makes causal hypotheses explicit by translating them into regression equations. Thus it becomes possible to integrate prior 'causal' knowledge in the data analysis, and to test 'causal' hypotheses. These important positive aspects of the technique are important in so far as this prior knowledge is relatively well-established, and in so far the hypotheses really make sense. Incorporating prior knowledge which is just conjectural means that we are treating prejudice as certainty, and this can lead to very undesirable consequences (as the nature-nurture debate about the genetics of intelligence amply shows; compare for instance Jaspars and De Leeuw 1980).

PATH MODELS IN GENERAL

We shall now define formally what we mean by a path model. In the first place such a model has a qualitative component, presented mathematically by a *graph* or *arrow diagram.* In such a graph the variables in our study are the *corners*, the relationships between these variables are the *edges*. In the path diagrams the variables are drawn as boxes, if there is an arrow from variable V_1 to variable V_2 then we say that V_1 is a *direct cause* of V_2 (and V_2 is a *direct effect* of V_1).

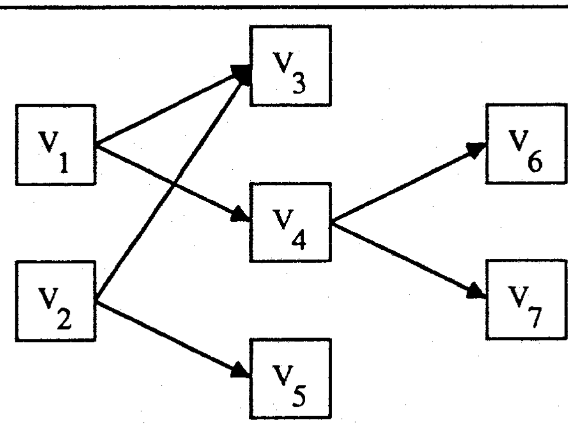


Figure 1.
Path diagram.

Compare Figure 1, for example. Observe that we use causal terminology without hesitation, but we follow the Simon-Wold example and give a precise definition of causes and effects in terms of graph theory. If there is a path from a variable V_1 to another variable V_2, then we say that V_1 is a *cause* of V_2 (and V_2 is an *effect* of V_1). In Figure 1, for instance, V_1 is a cause of V_6 and V_7, although not a direct cause.

Table 1.
Causal relations in Figure 1.

	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}

An important class of graphs is *transitive*, by which we mean that no path starting in a corner ever returns to that corner. Figure 1 would *not* be transitive any more with an arrow from V_7 to V_1, because of the path $V_1 \Rightarrow V_4 \Rightarrow V_7 \Rightarrow V_1$, but it would still be transitive with an arrow from V_7 to V_2. There have been heated discussions about the question whether or not non-transitive models can still be called causal. With our definition of causality they obviously can.

In transitive models we can define an interesting level assignment to the variables. This concept is due to De Leeuw (1984). Variables at which no arrows arrive are often called *exogenous* variables. They get level 0. The level of an *endogenous* (i.e. not exogenous) variable is one larger than the maximum level of its direct causes. We call V_1 a *predecessor* of V_2 (and V_2 a *successor* of V_1) if the level of V_1 is less than that of V_2. In the Table 1 we give causes, direct causes, and predecessors for the variables in Figure 1. Clearly the direct causes are a subset of the causes, and the causes are a subset of the predecessors. If x is any variable, we write this symbolically as $\mathbf{pred}(x) \supseteq \mathbf{cause}(x) \supseteq \mathbf{dcause}(x)$. By using $\mathbf{lev}(x)$ for the level, we can now say $\mathbf{dcause}(x) = \varnothing \Rightarrow \mathbf{lev}(x) = 0$, and $\mathbf{lev}(x) = 1 + \max \{\mathbf{lev}(y) \mid y \in \mathbf{dcause}(x)\}$. A model is transitive if $(\forall x)\{x \notin \mathbf{cause}(x)\}$. These qualitative concepts make it possible to explain what the general idea of path analysis is. We have defined our notion of causality in terms of the path diagram. Other notions which are important in path analysis will be discussed below.

TRANSITIVE PATH MODELS

We know make the path diagram quantitative, by embedding the qualitative notions in a numerical model for the variables. We restrict ourselves to *linear structural* models. There exist nonlinear path analysis techniques, developed in the framework of log-linear analysis (Goodman, 1978, Kiiveri and Speed, 1982), but these are outside our scope. They are discussed and compared with our approach in De Leeuw (1984). The only nonlinearity we allow for, at a later stage, is that connected with the transformation or quantification of variables. We assume, for the moment, that all variables are completely known, and, moreover, standardized to zero mean and unit variance. Thus VAR(x) = 1 for all variables x, and AVE(x) = 0.

The model in Figure 1 can be made numerical in the following way. We take all the endogenous variables in turn, and we suppose that they are a linear function of their direct causes, plus a disturbance term. The linear model corresponding with Figure 1 becomes

$$x_3 = \beta_{31}x_1 + \beta_{32}x_2 + \varepsilon_3, \tag{1a}$$
$$x_4 = \beta_{41}x_1 + \varepsilon_4, \tag{1b}$$
$$x_5 = \beta_{52}x_2 + \varepsilon_5, \tag{1c}$$
$$x_6 = \beta_{64}x_4 + \varepsilon_6, \tag{1d}$$
$$x_7 = \beta_{74}x_4 + \varepsilon_7. \tag{1e}$$

The assumptions we make about the disturbance terms ε_j are critical. These assumptions are in terms of uncorrelatedness, for which we use the symbol $\perp$. First assume for each j that the ε_j are uncorrelated with **dcause**(x_j). Thus

$$\varepsilon_3 \perp \{x_1, x_2\}, \tag{2a}$$
$$\varepsilon_4 \perp \{x_1\}, \tag{2b}$$
$$\varepsilon_5 \perp \{x_2\}, \tag{2c}$$
$$\varepsilon_6 \perp \{x_4\}, \tag{2d}$$
$$\varepsilon_7 \perp \{x_4\}. \tag{2e}$$

Now model (1)(2) describes any data set of seven variables perfectly. To see this it suffices to project each x_j on the space spanned by its direct causes, i.e. to perform a multiple regression with x_j as the dependent variable and $\mathbf{dcause}(x_j)$ as the independent ones, and to take ε_j equal to the residual. Then the disturbance is, per definition, uncorrelated with the direct causes in the same equation, and description is perfect. We can also say that the model is *saturated*, or *just identified*. It does not impose any restrictions, it merely provides us with an alternative description which is perhaps preferable to the original one because it links the data with some existing theory. But although description is, in a trivial sense, perfect, the performance of (1)(2) as a predictive model may still be very bad. The predictive power of the model is measured by the variances of the disturbances or residuals. If this is large, then we do not predict the corresponding variable efficiently. Thus we can have models which are good descriptors but poor predictors.

Path models can also be poor descriptors. But in that case we clearly must make stronger assumptions about the distribution of the disturbances. Let us call for any path model the assumption that for each j we have $\varepsilon_j \perp \mathbf{dcause}(x_j)$ the *weak orthogonality assumptions* . The *strong orthogonality assumptions* are defined for transitive models only. They are (i) that the disturbances are uncorrelated with the exogenous variables, and (ii) that disturbances of variables of different levels are uncorrelated with each other. In symbols this reads $\varepsilon_j \perp \{x \mid \mathbf{lev}(x) = 0\}$ and $\varepsilon_j \perp \{\varepsilon_k \mid \mathbf{lev}(x_k) \neq \mathbf{lev}(x_j)\}$.Thus, in a convenient compact notation, in our Figure 1,

$$\{\varepsilon_3,\varepsilon_4,\varepsilon_5,\varepsilon_6,\varepsilon_7\} \perp \{x_1,x_2\}, \tag{3a}$$
$$\{\varepsilon_3,\varepsilon_4,\varepsilon_5\} \perp \{\varepsilon_6,\varepsilon_7\}. \tag{3b}$$

Assumption (3) is much stronger than (2), and not all sets of seven variables satisfy (1) and (3). Because $\varepsilon_4 \perp \{x_1,x_2\}$, for example, regression of x_4 on x_1 and x_2 will give $\beta_{42} = 0$ if (1)(3) is true, and this is clearly restrictive. Thus model (1)(3) can be a poor descriptor as well as a poor predictor. It is clear, by the way, that a model which is a good predictor is automatically a good descriptor.

For the causal interpretation the following argument is useful. It extends to all transitive models. We have $\varepsilon_6 \perp \{x_1,x_2\}$ and $\varepsilon_6 \perp \varepsilon_3$. Thus, from (1a), $\varepsilon_6 \perp x_3$. In the same way $\varepsilon_6 \perp x_4$ and $\varepsilon_6 \perp x_5$. Thus $\varepsilon_6 \perp \{x_1,x_2,x_3,x_4,x_5\}$, which implies that $\mathbf{proj}(x_6|x_1,x_2,x_3,x_4,x_5) = \mathbf{proj}(x_6|x_4)$, with $\mathbf{proj}(y|x_1,\ldots,x_m)$ denoting least squares projection of y on the space spanned by $x_1,\ldots,x_m$. In words this says that the projection of x_6 on the space spanned by its predecessors is the projection of x_6 on the space spanned by its direct causes. The interpretation is that, given the direct causes, a variable is independent of its other predecessors. Thus the strong orthogonality assumptions in transitive models imply a (weak) form of *conditional independence* .

We shall now treat some more or less familiar models in which description is perfect. These models are consequently saturated. The structural equations defining the model can be solved uniquely, and the model describes the data exactly. The first, and perhaps simplest, example is the *multiple regression model.* An example is given in Figure 2.

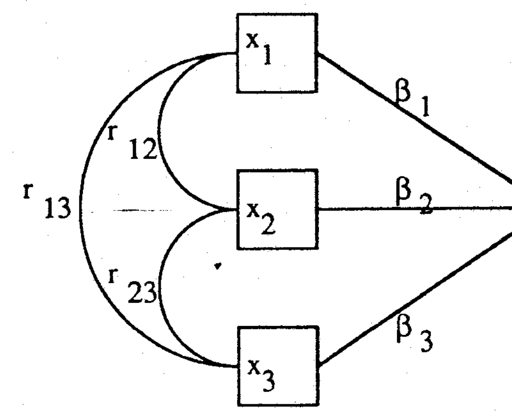


Figure 2.

Multiple regression model.

If we compare this with Figure 1 we see some differences which are due to the fact that we have made the model quantitative. In the first place the arrows now have values, the regression coefficients. In the second place it is convenient to use curved loops indicating the correlations between the exogenous variables. The curved loops can also be used to represent correlated disturbances. This becomes more clear perhaps if we add dummy equations like $x_j = \varepsilon_j$ for each of the exogenous variables, which is consistent with the idea that exogenous variables have no causes; exogenous variables are, in this sense, identical with disturbances. The strong orthogonality assumptions on disturbances can now be stated more briefly, because they reduce to the single statement $\varepsilon_j \perp \{ \varepsilon_k \mid \mathbf{lev}(x_k) \neq \mathbf{lev}(x_j) \}$. Arrows are also drawn in Figure 2 to represent uncorrelated disturbance terms.

In Figure 2 , and in multiple regression in general, there is only one endogenous variable, often called the *dependent* variable. There are several exogenous variables, often called *predictors* or *independent* variables. The linear structural model is

$$y = \beta_1 x_1 + \dots \beta_m x_m + \varepsilon. \tag{4}$$

The orthogonality assumptions on the disturbances are $\varepsilon \perp \mathbf{dcause}(y) = \{x_1,\dots,x_m\}$. In this case the strong assumptions are identical with the weak assumptions, because **dcause(y)** are exactly the exogenous variables. Thus (4) is a saturated model. If we project the dependent variable on the space spanned by the predictors, then the residual is automatically uncorrelated with each of the predictors. The description is perfect, although the prediction may be lousy. We measure quality of prediction by the multiple correlation coefficient $R^2 = 1 - VAR(\varepsilon)$, in this context also known as the *coefficient of determination* .

Figure 3 shows a somewhat less familiar model. Its linear structure is

$$x_2 = \beta_{21}x_1 + \varepsilon_2, \tag{5a}$$
$$x_3 = \beta_{31}x_1 + \beta_{32}x_2 + \varepsilon_3. \tag{5b}$$

The weak orthogonality assumptions, which make (5) a saturated model, are $\varepsilon_2 \perp \{x_1\}$ and $\varepsilon_3 \perp \{x_1,x_2\}$. It follows from this that ε_2 is the residual after projection of x_2 on x_1. Thus β_{21} is equal to the correlation between x_1 and x_2, and $\varepsilon_2 = x_2 - \beta_{21}x_1$ is a linear combination of x_1 and x_2. This implies that $\varepsilon_3 \perp \varepsilon_2$, and consequently the strong orthogonality assumptions are true as well. Although we did not require it, we automatically get uncorrelatedness of the disturbance terms.

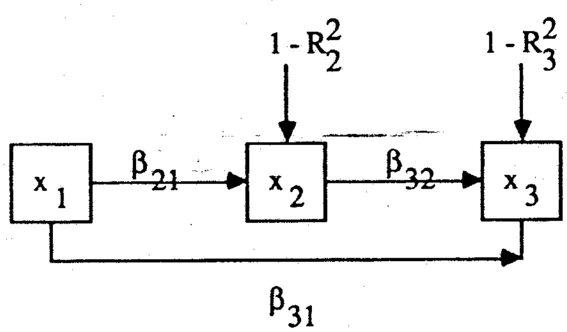


Figure 3.

A simple saturated recursive model.

If we try to generalize the structure in Figures 2 and 3 we find something like Figure 4. Variables are partitioned into sets, and variables in the same set have the same level. In *saturated block-transitive* models **dcause**(x) = **pred**(x) for all variables x. Thus there are arrows from each variables to all variables of a higher level. There are no arrows within sets. The arrows indicating errors in Figure 4 actually indicate correlated errors. *Saturated simple transitive models* (also called *causal chains*) have only one variable in each set, and thus all variables have a different level. For both block transitive models and simple transitive models the weak orthogonality assumptions , together with the structure, imply the strong orthogonality assumptions. And, consequently, imposing the strong orthogonality assumptions leaves the model saturated and the description perfect. Residuals of variables of different levels are uncorrelated, and residuals are uncorrelated with variables of a lower level. There can be correlation between the residuals of variables of the same level, or between residuals and variables of a higher level. We can find path coefficients by regressing each endogenous variable on the set of its predecessors. We have seen that transitive models are path models corresponding with transitive graphs having no 'causal loops'. Saturated transitive models, of which the block transitive models and simple transitive models are special cases, describe the dispersion matrix of the variables precisely. Non-saturated or restrictive transitive models, of which the model in Figure 1 is a special case, arise from saturated models by leaving out certain arrows. It is still the case that an unambiguous level assignment is possible, and the terminology of predecessors and successors still applies.

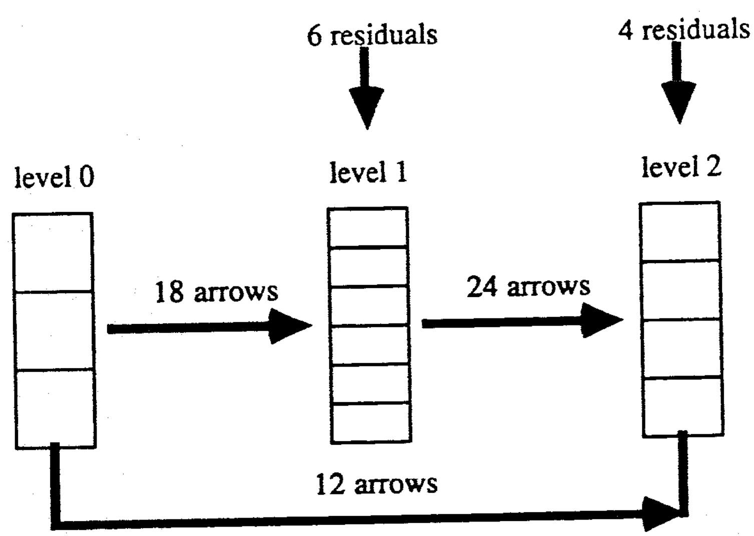


Figure 4.

General recursive saturated model.

In quantifying any path model we can simply use the path diagram to write down the linear structural equations. We also have to assume something about the disturbances in terms of their correlation with each other and with the x_j. The weak orthogonality assumptions can be applied in all cases. They make the model saturated, and have as a consequence that consistent estimation of the regression coefficients is possible by projecting a variable on the space spanned by its direct causes. In all transitive models, saturated or not, the strong orthogonality conditions follow from the weak orthogonality conditions and the linear structure. Thus the causal interpretation in terms of conditional independence is available.

The notion of a linear structural model is more general than the notion of a transitive model, of course. If we assume a structural model, such as (1), then we can make alternative assumptions about the residuals, for instance that they are all uncorrelated. In fact we can easily build linear structural models which are not transitive at all. Simply write down the model from the path diagram, one equation for each endogenous variable, and make some sort of assumption about the disturbances. By allowing for correlations between the disturbances we can create saturated nontransitive models, and we can also get into problems with *identifiability*. For these identification problems we refer to the econometric literature, for instance to Hsiao (1983) or Bekker (1986). Observe that nontransitive models can not be translated into conditional independence statements, which has caused some authors to say that nontransitive models are not causal.

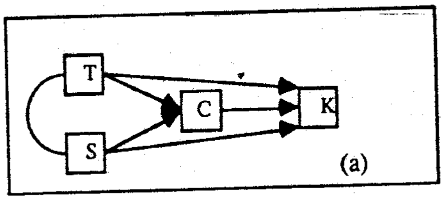
For a small ecological example we use a part of the correlation matrix given by Legendre and Legendre (1983, Table 5.6). The data have to do with primary production, and were collected in 1967 in the Baie des Chaleurs (Québec). There are 40 measurements on four variables. These are:

K: the biological attenuation coefficient which represents the relative primary production,
C: the concentration of chlorophyll a,
S: the degree of salinity,
T: the temperature.

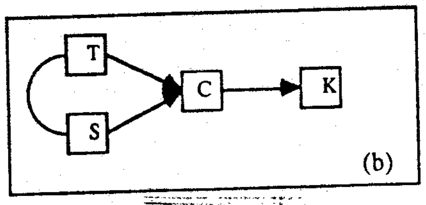
The correlation matrix, and some simple path models, are given in Table 2. Model (a) is the saturated model which has T and S as exogenous variables (level 0), has C as a variable of level 1, and K as the innermost variable of level 2. Model (b) is not saturated, because the paths from T and S directly to K are eliminated. All effects of T and S on K go through C, or, to put it differently, K is independent of T and S, given C. Model (c) is also saturated, but no choice is made about the causal priority of C or K. Thus C and K have correlated errors, because they both have level 1. In the part of Table 2 that gives the fitted coefficients we see that the covariance of the errors in (c) is .721. Because of this covariance variable K has a much larger error variance in model (c).

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

Correlations
Baie des Chaleurs

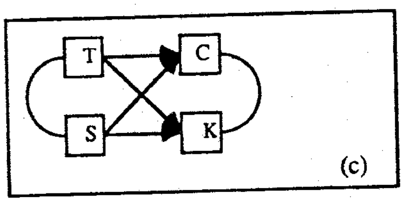


(a)



(b)

Three recursive models



(c)

Table 2.
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	*****
VAR ERR C	0.787	0.787	0.787
VAR ERR K	0.260	0.291	0.920
COV ERR C,K			0.721

Models (a) and (c) give a perfect description of the correlations, so the choice between them must be made purely on the basis of prior notions the investigator has. We are not familiar with the problems in question, so we cannot make a sensible choice. Model (b) is restrictive. If we compare it with (a) we still see that its description is relatively good. If we want to decide whether to prefer it to (a) we can either use statistics, and see if the description is 'significantly' worse. But we can also use (a) and (b) predictively, and see which one is better. Our guess is that on both counts (b) is the more satisfactory model.

DIRECT AND INDIRECT EFFECTS

In this paragraph we discuss the calculus of path coefficients explained by Wright (1921, 1934). We do not present the general theorems here, but we illustrate the calculus by using our examples. First consider the model in Figure 3. Let us use equations (5) to compute the correlations between x_1, x_2 and x_3. We find $r_{21} = \beta_{21}$ and $r_{31} = \beta_{31} + \beta_{32}r_{21} = \beta_{31} + \beta_{32}\beta_{21}$. In terms of Figure 3 the equation for r_{31} can be interpreted as follows: there is a *direct effect* of x_1 on x_3 with size β_{31}, and an *indirect effect* (via x_2) of $\beta_{32}\beta_{21}$. The indirect effect comes about because there is a path from x_1 to x_3, which passes by x_2. Coefficients along the path are multiplied to quantify the indirect effect. In the same way we find $r_{32} = \beta_{31}r_{12} + \beta_{32} = \beta_{31}\beta_{21} + \beta_{32}$. Again a direct and an indirect effect, but now the indirect effect does not correspond with a path in the directed graph but with the path in the corresponding undirected graph.

An even clearer example can be obtained from Table 2. In model (a), for instance, we have $K = \beta_{KC}C + \beta_{KT}T + \beta_{KS}S + \varepsilon_K$ and $C = \beta_{CT}T + \beta_{CS}S + \varepsilon_C$. Thus $r_{KC} = \beta_{KC} + \beta_{KT}r_{CT} + \beta_{KS}r_{CS} = \beta_{KC} + \beta_{KT}(\beta_{CT} + \beta_{CS}r_{ST}) + \beta_{KS}(\beta_{CT}r_{ST} + \beta_{CS}) = \beta_{KC} + \beta_{KT}\beta_{CT} + \beta_{KS}\beta_{CS} + \beta_{KT}\beta_{CS}r_{ST} + \beta_{KS}\beta_{CT}r_{ST}$. Thus r_{KC} is the sum of a direct effect, an indirect effect via T and another indirect effect via S. The two remaining contributions to r_{KC} come from the (undirected) paths from K to T to S to C and from K to S to T to C. In model (b) we have $r_{KC} = \beta_{KC}$, because this direct effect is the only path. In model (c) $K = \beta_{KT}T + \beta_{KS}S + \varepsilon_K$ and $C = \beta_{CT}T + \beta_{CS}S + \varepsilon_C$. Thus $r_{KC} = \beta_{KT}\beta_{CT} + \beta_{KS}\beta_{CS} + \beta_{KT}\beta_{CS}r_{ST} + \beta_{KS}\beta_{CT}r_{ST} + r(\varepsilon_K,\varepsilon_C)$ and there is no direct effect.

The terminology of direct and indirect effects is causal, of course, and our earlier warnings against taking this terminology too literally apply. For model (a) in Table 2 we find, for instance, for the direct effect from C on K +.916, the indirect effect via T is -.023, the indirect effect via S is -.230, and the two effects 'K to T to S to C' and 'K to S to T to C' are +.030 and +.149. The sum of these effects is +.842, which is indeed the correlation between C and K. It is difficult, and risky, to give a causal interpretation, because the values depend strongly on the model that we have chosen. In model (c), for instance, the indirect effect via T is +.466 and the indirect effect via S is +.768. The equation for r_{KC} in (c) becomes .842 = .466 + .768 - .616 - .497 + .721. The model also fits perfectly, but presumably the causal interpretation would be quite different.

Although the calculus of path coefficients in transitive models is an interesting and perfectly legitimate way to decompose correlation coefficients, causal interpretation in terms of direct and indirect effects seems valuable only if there are strong reasons to prefer the particular model in the study over other competing models. And this happens only if we already have a pretty good idea about the mechanisms that are at work in the situation we are studying. If the sociologist says that fathers's income only has an indirect effect on the career of the child, this is either just a figure of speech, or a statement that a particular partial correlation coefficient is small. In Chang (1981), and Troussellier *et al.* (in press), it is shown that the decomposition of the correlation coefficients in direct and indirect contributions (with respect to a particular path model) can lead to useful interpretations in community ecology.

LATENT VARIABLES

Now consider the path models in Figures 5 and 6. They are different from the ones we have seen before, because they involve *latent* or *unobserved* variables. In the diagrams we indicate these latent variables by using circles instead of squares. First we give the causal interpretation of Figure 5. If we project the observed variables on the space spanned by the unobserved variables then the residuals are uncorrelated. Thus the observed variables are independent given the unobserved variable. All relationships between the observed variables can be 'explained' by the latent variable, which is their *common factor*. In somewhat more intuitive terms a good fit of this common factor model to the data means that the variables all measure essentially the same property. A good fit, and small residuals, means that they all measure this property in a precise way. Again we see that the model can be a good description of the data without being a good predictor. Uncorrelated variables, for instance, are described perfectly by the model, but cannot be predicted at all.

The structural equations describing the model are

$$x_j = \alpha_j\zeta + \varepsilon_j. \tag{6}$$

The ε_j are assumed to be uncorrelated with ζ. Model (6) is saturated and transitive, but it has the peculiar property that the exogenous variable is not measured. In De Leeuw (1984) it was suggested that latent variables are just another example of variables about which not everything is known. We have nominal variables, ordinal variables, polynomial variables, splinical variables, and we also have latent variables. About latent variables absolutely nothing is known, except for their place in the model. Thus the basic optimal scaling idea that transformations and quantifications must be chosen to optimize prediction also applies to latent variables.

Consequently latent variables fit very naturally into the optimal scaling approach to path analysis.

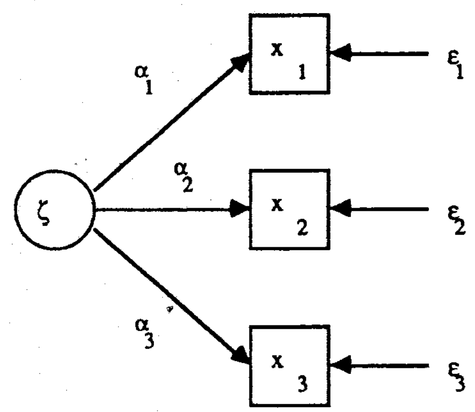


Figure 5.

One-factor model

The model in Figure 6 is a special case of the MIMIC model proposed by Jöreskog and Goldberger (1975). In MIMIC models there are two sets of variables. The exogenous variables influence the observable endogenous variables through the mediation of one or more latent variables.

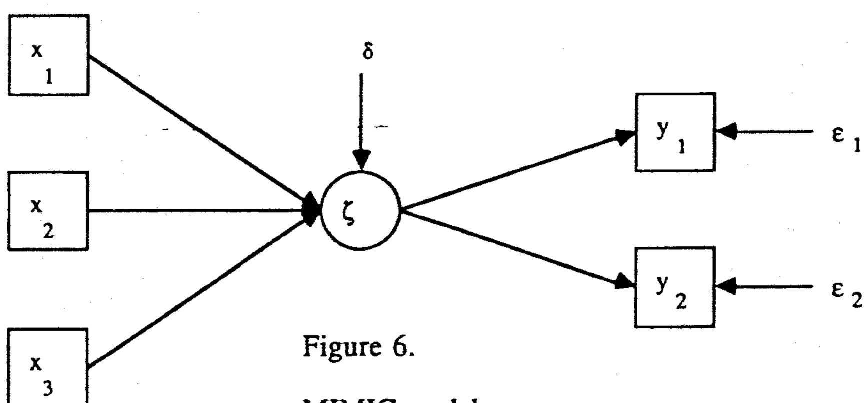


Figure 6.

MIMIC model.

The MIMIC model combines aspects of psychometrical modelling with aspects of econometric modelling. It follows from the MIMIC equations, that the observable endogenous variables satisfy a factor analysis model, while the joint distribution of exogenous and endogenous variables is a reduced rank regression model. For Figure 6 these equations are

$$\zeta = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \delta, \tag{7a}$$
$$y_1 = \alpha_1 \zeta + \varepsilon_1, \tag{7b}$$
$$y_2 = \alpha_2 \zeta + \varepsilon_2. \tag{7c}$$

The MIMIC model is closely related to canonical correlation analysis (Bagozzi, Fornell, and Larker, 1981) and to redundancy analysis (Gittins, 1985, section 3.3.1).

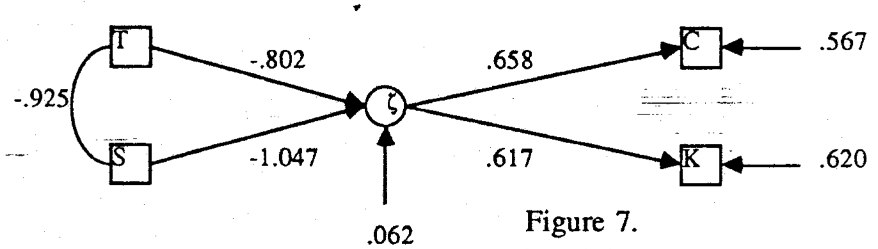


Figure 7.
MIMIC model,Legendre data.

Figure 7 illustrates an application of the MIMIC model to the Baie des Chaleurs data of Legendre and Legendre. The values of the path coefficients and the error variances are given in the diagram. The model provides a reasonably good description, compared with the transitive models in Table 2. The causal interpretation of Figure 7 is that temperature and salinity determine the unmeasured variable ζ, which in its turn determines primary production and chlorophyll concentration.

In our experience some people find it difficult to accept the concept of a latent variable. But there are several reasons why we still think that such a concept is useful. In the first place in many of the sciences measurement errors can not be neglected. This means that the observed variable is an *indicator* of the latent 'true' variable. The concept of an indicator can be generalized considerably, and this has happened mainly in psychometrics and in sociological methodology. It is not possible to measure 'intelligence' directly, but it is possible to measure a large number of indicators for intelligence. If the common factor model is acceptable, then we have found a way to measure intelligence as a linear combination of indicators (it is still possible, under these circumstances, that measurement of intelligence is poor in a predictive sense). The situation can be compared with determining the weight of a number of objects if we have a number of spring balances with unknown characteristics. This can be done quite well by common factor analysis. Social scientists happen to use a large number of concepts such as intelligence (or attitude, or status, or power), which can not be measured directly but for which indicators are available. It seems to us that the situation in ecology is not really different. This means that the path models in terms of the observed variables are theoretically not very statisfactory, because the theory says something about the relationships between conctructs or concepts, which should not be confused with their indicators. And finally, we have already used latent variables in classical path analysis as well. The errors or disturbances in the equations are also unobserved, and measurable only by making linear combinations of observed variables. If we allow for 'errors in equations', we may as well allow for 'errors in variables'.

OPTIMAL SCALING OF VARIABLES

We now briefly indicate where the theory of optimal scaling comes in. We have seen in De Leeuw (1987) that optimal scaling (or transformation, or quantification) can be used to optimize criteria defined in terms of the correlation matrix of the variables. In path analysis the obvious criteria are the coefficients of determination, i.e. the multiple correlation coefficients. In De Leeuw (1987) we already analyzed an example in which the multiple correlation between predictors SPECIES and NITRO and dependent variable YIELD was optimized. In path analysis we deal with nested multiple regressions, and we can choose which one (or which combination) of the multiple correlations we want to optimize. If there is no prior knowledge dictating otherwise, then it seems to make most sense to maximize the sum of the coefficients or determination of all the endogenous variables. But in other cases we may prefer to maximize the sum computed only over all variables of the highest level.

In general nontransitive models the methods of optimal scaling can be used exactly as in transitive models. We have one coefficient of determination for each endogenous variable, and we can scale the variables in such a way that the sum of these coefficients is optimized. This amounts to finding transformations or quantifications optimizing the predictive power of the model. Moreover it is irrelevant for our approach if the model contains latent variables or not. We have seen that latent variables are simply variables with a very low measurement level, and that they can be scaled in exactly the same way as ordinal or nominal variables. This point of view, due to De Leeuw (1984), makes our approach quite general. It is quite similar to the NIPALS approach of Wold , described most fully in Jöreskog and Wold (1982) and Lohmöller (1986).

It is of some interest that we do not necessary optimize the descriptive efficiency at the same time. Optimizing predictive power is directed towards the weak orthogonality assumptions. It is possible, at least in principle, that a model with optimized coefficients of determination has a worse fit to the strong orthogonality assumptions. Scaling to optimize predictability does not guarantee an improved fit in this respect. This has as a consequence that there is a discrepancy between the least squares and the maximum likelihood approach to fitting nontransitive path models. We do not go into these problems, but refer the interested reader to Dijkstra (1981), Jöreskog and Wold (1982), and De Leeuw (1984) for extensive discussions.

We now outline the algorithm that we use in nonlinear path analysis somewhat more in detail. We minimize the sum

$$\Sigma_j \|x_j - \Sigma_l \beta_{jl} x_l\|^2, \tag{8}$$

over both the regression coefficients β_{jl} and the quantifications (or transformations) of the variables. The outer summation, over j, is over all endogenous variables, the inner summation, over l, is over all variables that are direct causes of variable j. The algorithm we use of is the *alternating least squares* type (Young, 1981). This means that the parameters of the problem are partitioned into sets, and that each stage of the algorithm minimizes the loss function over one of the sets, while keeping the other sets fixed at their current values. By cycling through the sets of parameters we obtain a convergent algorithm. In this particular application of the general alternating least squares principle each variable defines a set of parameters, and the regression coefficients define another set.

We give an ecological illustration of this nonlinear PATHALS algorithm. The data are taken from Van der Aart and Smeenk-Enserink (1975), who reported abundance data for 12 species of hunting spiders in a dune area in the Netherlands. A total of 28 sites was studied, and the sites were also described in terms of a number of environmental variables. We have used a selection and coding from these data made by Ter Braak (1986a). He used the six environmental variables:

WC Water content, percentage dry weight,
BS Percentage bare sand,
CM Percentage covered by moss layer,

LR Reflection of soil surface at cloudless sky,
FT Percentage covered by fallen leaves or twigs,
CH Percentage covered by herbs layer.

Ter Braak categorized all variables into 10 discrete categories, in order to present them succinctly. We have taken over his categorization, and used it in our analysis.

The results of a MIMIC analysis with two latent variables (factors) are given in Table 3. Analyses with only a single latent variable were not very successful. We first performed a linear analysis, using the category scores from the coding by Ter Braak, and we then computed optimal monotone transformations. As an illustration the optimal transformations for the environmental variables are given in Figure 8. We see a large variety of shapes. It would carry us too far astray to give a detailed analysis of these nonlinearities. Of course these transformations are only optimal given the path model, in this case given the number of latent variables, for instance.

Table 3. Hunting spider data. Metric and nonmetric MIMIC analysis.

	weights metric		weights nonmetric		residual variances metric	residual variances nonmetric
WC	-.77	.20	-.97	.24		
BS	-.02	.11	-.30	.53		
CM	.20	.17	.09	.27		
LR	.17	.52	-.07	.54		
FT	.62	-.32	.29	-.02		
CH	-.26	.13	-.52	.41		
S1	-.77	.21	-.87	.19	.39	.21
S2	-.10	-.79	.21	-.86	.37	.21
S3	-.89	-.04	-.89	-.20	.21	.16
S4	-.91	.23	-.97	.18	.12	.04
S5	-.92	.26	-.96	.18	.08	.04
S6	-.88	.16	-.93	.16	.20	.10
S7	-.95	-.15	-.97	-.08	.07	.04
S8	-.75	.15	-.85	-.01	.42	.27
S9	-.25	.63	-.40	.73	.54	.31
S10	.30	.83	.15	.90	.22	.16
S11	.59	.53	.54	.72	.37	.18
S12	.57	.35	.60	.58	.56	.31

For a more detailed discussion and interpretation of the data we refer to Van der Aart and Smeenk-Enserink (1975) and to Ter Braak (1986a), who both performed forms of canonical analysis. Actually Ter Braak used canonical correspondence analysis, a form of nonlinear canonical analysis, also discussed in Ter Braak (1986b). We merely point out some 'technical' aspects of our analysis, and we

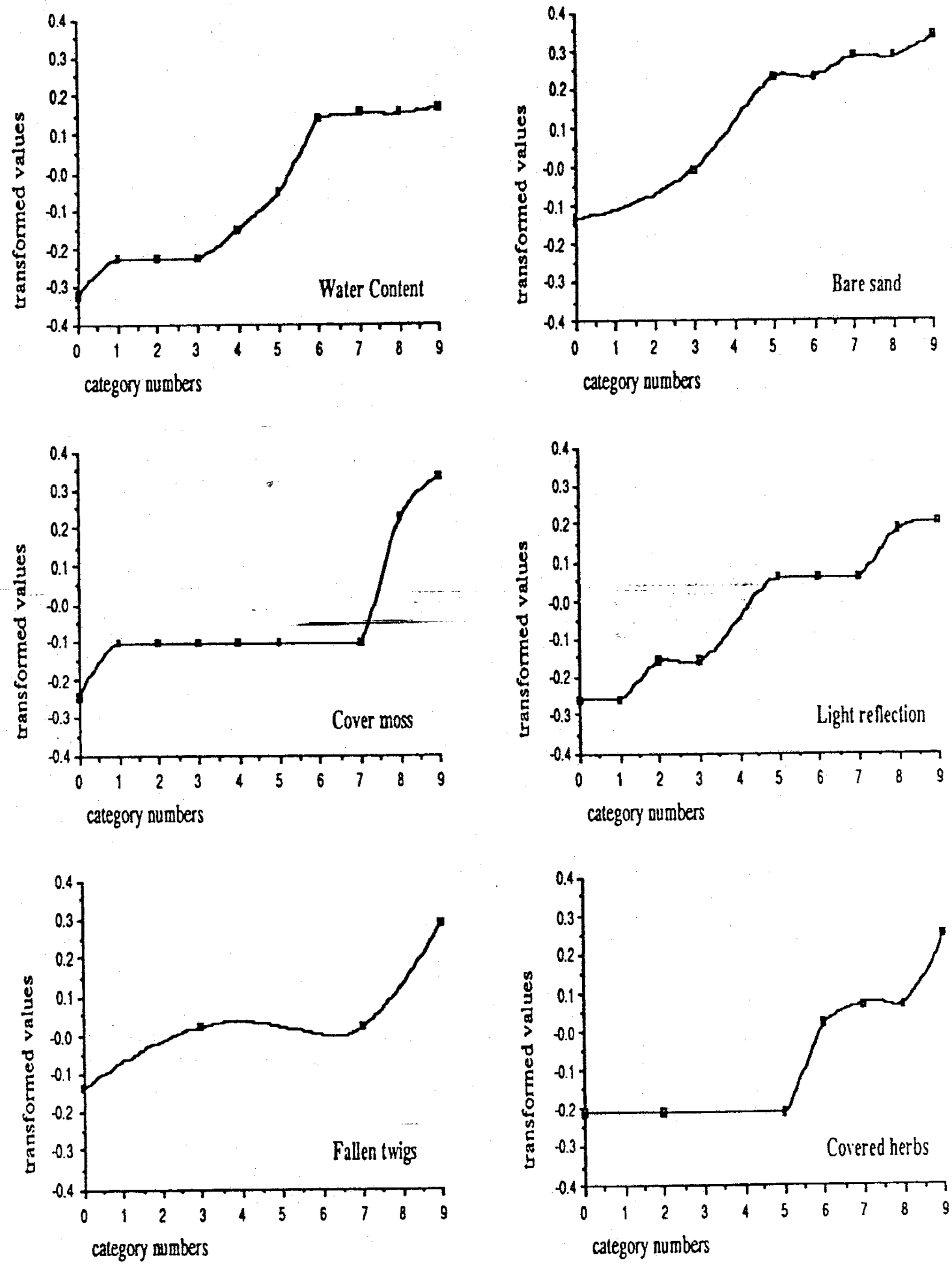


Figure 8. Optimal monotone transformations,environmental variables.

compare the linear and nonlinear solutions. It is clear that the 'explained' variances of the transformed abundance variables increase considerably. The table does not give the 'explained' variance of the two latent variables. For the metric analysis the residuals are .06 and .14, for the nonmetric analysis they are .01 and .01. Thus the latent variables in the nonmetric analysis are almost completely in the space of the transformed environmental variables, which implies that our method is very close to a nonmetric redundancy analysis.

The interpretation of the latent variables is facilitated, as is usual in forms of canonical analysis, by correlation the latent variables with the transformed variables. This gives *canonical loadings* . If we do this we find, for example, that the first latent variable correlates -.75 with both Water Content and Cover Herbs, while the second one correlates +.80 with Light Reflection and -.80 with Fallen Twigs. The analysis clearly shows some of the advantages of nonlinear multivariate analysis. By allowing for transformations of the variables we need fewer dimensions to account for a large proportion of the variance. Much of the remaining variation after a linear analysis is taken care of by the transformations, and in stead of interpreting high-dimensional linear solutions we can interprete low-dimensional nonlinear solutions, together with the transformations computed by the technique. Using transformations allows for simple nonlinear relationships in the data, and the optimal transformations often give additional useful information about the data.

CONCLUSIONS

Discussions of multivariate analysis , also in the ecological literature, often limit themselves to various standard situations, and the associated techniques. Thus multiple regression, principal component analysis, and canonical correlation analysis are usually discussed, for situation in which we want to predict one variables from a number of others, in which we want to investigate the structure of a single set of variables, or in which we want to relate two sets of variables. The path analysis techniques, with latent variables, discussed in this paper, make it possible to use a far greater variety of models, and even to design a model which may be especially suited for the data or the problem at hand. Usually the choice of the path model will be based on prior knowledge the investigator has about the causal relationships of the variables in the study. Although this far greater flexibility may have its dangers, it is clearly a very important step ahead because incorporating prior information into the analysis can enhance both the stability and the interpretability of the results.

The nonlinear extensions of path analysis discussed in his paper allow for even more flexibility. Not only can we choose the overall structure of the analysis by choosing a suitable path model, but within the model we can also choose the measurement level of each of the variables separately. Or, if one prefers this terminology, we can define a suitable class of transformations for each variable from which an optimal one must be chosen. The use of transformations can greatly increase the explanatory power of path models, at least for the data set in question. If the transformations we obtain are indeed stable, and also increase the quality of the predictions, is quite another matter. This must be investigated by a detailed analysis of the stability and the cross-validation properties of the estimates, which is a very important component of any serious data analysis.

Thus we can say that this paper adds a number of very powerful and flexible tools to the toolbox of the ecologist, with the logical and inevitable consequence that these new tools can lead to more serious forms of misuse than the standard tools, which are more rigid and less powerful. The major hazard is *chance capitalization*, i.e. instability, and the user of these tools must take precautions against this danger. But if suitable precautions are taken, the path analysis methods and the generalizations discussed in this paper provide us with a convenient and useful way to formalize scientific theories in situations, in which there is no precise knowledge of the detailed mechanisms, or in which there are too many factors influencing the system to make a precise deterministic description possible.

REFERENCES

BAGOZZI, R.P., C. FORNELL, AND D.F. LARKER. 1981. Canonical correlation analysis as a special case of a structural relations model. Multivariate Behavioural Research 16: 437-454.

BEKKER, P. 1986. Essays on the identification problem in linear models with latent variables. Doctoral Dissertation. Department of Econometrics, Tilburg University, Tilburg, The Netherlands.

CHANG, W.Y.B. 1981. Path analysis and factors affecting primary productivity. Journal of Freshwater Ecology 1: 113-120.

DE LEEUW, J. 1984. Least squares and maximum likelihood for causal models with discrete variables. Report RR-84-09, Department of Data Theory, University of Leiden, The Netherlands.

DE LEEUW, J. 1985. Review of four books on causal analysis. Psychometrika 50: 371-375.

DE LEEUW, J. 1987. Nonlinear multivariate analysis with optimal scaling. This Volume.

DIJKSTRA, T.K. 1981. Latent variables in linear stochastic models. Doctoral Dissertation. Department of Econometrics, University of Groningen, The

Netherlands.
FRISCH, R. 1934. Statistical confluence analysis by means of complete regression systems. Economic Institute, University of Oslo, Norway.
GITTINS, R. 1985. Canonical analysis. Springer, Berlin, BRD.
GOODMAN, L.A. 1978. Analyzing qualitative categorical data. Abt, Cambridge, Ma.
GOSSELIN, M., L. LEGENDRE, J.-C. THERRIAULT, S. DEMERS, AND M. ROCHET. 1986. Physical control of the horizontal patchiness of sea-ice microalgae. Marine Ecology Progress Series 29: 289-298.
HARRIS, R.E., AND W.A.G. CHARLESTON. 1977. An examination of the marsh microhabitats of *Lymnaea tomentosa* and *L. columella* (Mollusca: Gastropoda) by path analysis. New Zealand Journal of Zoology 4: 395-399.
HSIAO, C. 1983. Identification. *In* Z. Griliches, and M.T. Intriligator [eds.] Handbook of Econometrics I. North Holland Publishing Co., Amsterdam, The Netherlands
JASPARS, J.M.F., AND J. DE LEEUW. 1980. Genetic-environment covariation in human behaviour genetics. *In* L.J.Th. van der Kamp *et al.* (eds.) Psychometrics for Educational Debates. John Wiley and Sons, New York, NY.
JØRESKOG, K.G., AND A.S. GOLDBERGER. 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association 70: 631-639.
JØRESKOG, K.G., AND H. WOLD. 1982. Systems under indirect observation. North Holland Publishing Co., Amsterdam, The Netherlands.
KIIVERI, H., AND T.P.SPEED. 1982. Structural analysis of multivariate data. *In* S. Leinhardt (ed.) Sociological Methodology. Jossey-Bass, San Francisco, CA.
LEGENDRE, L., AND P. LEGENDRE. 1983. Numerical ecology. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands.
LOHMÖLLER , J.B. 1986. ~~Die Partialkleinst~~quadratmeth~~ode für~~ Pfadmodelle mit latenten Variablen und das Programm LVPLS. *In* L. Hildebrand *et al.* (eds.) Kausalanalyse in der Umweltforschung. Campus, Frankfurt, BRD.
PEARSON, K. 1911. The grammar of science. Third Edition.
SCHWINGHAMER, P. 1983. Generating ecological hypotheses from biomass spectra using causal analysis: a benthic example. Marine Ecology Progress Series 13: 151-166.
SIMON, H.A. 1953. Causal ordering and identifiability. *In* W.C. Hood, and T.C. Koopmans (eds.) Studies in Econometric Method. John Wiley and Sons, New York, NY.
SPEARMAN, C. 1904. General intelligence objectively measured and defined. American Journal of Psychology 15: 201-299.
TER BRAAK, C.L.F. 1986a. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology, in press.
TER BRAAK, C.L.F. 1986b. The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio, in press.
TROUSSELIER, M., P. LEGENDRE, AND B. BALEUX. 1986. Modeling of the evolution of bacterial densities in an eutrophic ecosystem (sewage lagoons). Microbial Ecology 12: 355-379.
TUKEY, J.W. 1954. Causation, regression, and path analysis. *In* O. Kempthorne

(ed.) Statistical Methods in Biology. Iowa State University Press, Ames, Iowa.
VAN DER AART, P.J.M., AND N. SMEEK-ENSERINK. 1975. Correlation between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. Netherlands Journal of Zoology 25: 1-45.
WOLD, H. 1954. Causality and econometrics. Econometrica 22: 162-177.
WRIGHT, S. 1921. Correlation and causation. Journal Agricultural Research 20: 557-585.
WRIGHT, S. 1934. The method of path coefficients. Annals of Mathematical Statistics 5: 161-215.
YOUNG, F.W. 1981. Quantitative analysis of qualitative data. Psychometrika 46: 347-388.