Aspects and application of an integrated environmental model with a satellite design (E 85/4)

Brouwer, F.; Nijkamp, P.

1986

SERIE RESEARCH MEMORANDA

ASPECTS AND APPLICATION OF AN
INTEGRATED ENVIRONMENTAL MODEL
WITH A SATELLITE DESIGN
(E 85/4)

Floor Brouwer
Peter Nijkamp

Researchmemorandum 1986-2 februari 1986

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
A M S T E R D A M

ASPECTS AND APPLICATION OF AN INTEGRATED ENVIRONMENTAL MODEL WITH A SATELLITE DESIGN

(E 85/4)

Floor Brouwer[1])
Peter Nijkamp[2])

Paper prepared for the MAB-Arbeitstagung "Integration von Modellen Sozialer, Ökonomischer und Ökologischer Systeme", Osnabrück, 13 - 15 March, 1985.

1) Free University
Institute for Environmental Studies
P.O. Box 7161
1007 MC Amsterdam

2) Free University
Department of Economics
P.O. Box 7161
1007 MC Amsterdam

The study is supported by the Netherlands Organization for the Advancement of Pure Research (ZWO), project-number 46-125.

TABLE OF CONTENTS

1. INTRODUCTION

The development of methodological tools to analyse, in a qualitative or a quantitative way, phenomena from different disciplines is increasingly becoming necessary because of the rise in complexity and the dynamics of societies.
A uniform and coherent way of analyzing the integration of, inter alia, economics, demography, recreation, and natural environment is necessary in view of the different natures of phenomena from these disciplines. Systems analysis will be discussed in the paper as a methodological tool to conceptualize an integrated environmental model (abbreviated as IEM) and to analyse the structure of strongly interrelated components. Another advantage of the use of systems analysis is that it also has the ability to decompose a system into subsystems (called modules in the paper).
In section 2 of this paper, a so-called horizontal design and a vertical design of an IEM will be discussed by a systems analytical approach. The aspects and main characteristics of both model designs will be described briefly. Next, a satellite design of an integrated environmental model will be proposed in section 2.3, which is characterized as a mixture of the horizontal and vertical model structure. The aspects of a satellite design and a conceptual application for the Biesbosch area in the Netherlands will be presented in this section as well.
Some empirical aspects which deal with an IEM at a regional level by means of a satellite design are presented in section 3. Here, special emphasis will be placed upon the following points:

(i) The spatial aggregation and spatial scale level of variables, which are both relevant for modelling results at the regional level(see section 3.2). It deals with the measurement unit of variables. The main reason to put emphasis upon the spatial aggregation and scale level is that the zoning systems are not uniquely meaningful fixed entities. However, we need to mention that nearly all applications depend on the assumption that the spatial units are a priori given.

(ii) The difficulties in empirical practice to obtain precise information because of measurement problems, lack of time or simply lack of money to collect reliable quantitative data, and so forth (see section 3.3). In such cases a qualitative approach about the impacts between parameters may be very appealing. In the context of this paper, a set of linear equations $Ax = b$ will be solved for vector x by means of sign-solvability analysis. This approach belongs to the analysis of qualitative relations, when the sign (positive, negative or zero) is the only information about the impacts between variables.

(iii) High-quality information in natural sciences is obtained under experimental and well-defined conditions and is normally metric in nature. However, the information in regional-economic research, geography and demography etc. will frequently have been obtained from imprecise data sources or from a wide range of sample surveys, and is often non-metric (qualitative, categorical or discrete) in nature. A family of interrelated statistical models have been developed in recent years, called Generalized Linear Models (GLMs), which are able to deal with such information (see section 3.4). The main tools of qualitative statistical modelling are log-linear models and logistic-linear models. Both model approaches are useful for the analysis of nominal data and belong to the family of GLMs. It is possible to relate, inter alia, ordinary least squares regression with metric data to a wide range of statistical models with qualitative data in terms of GLMs.

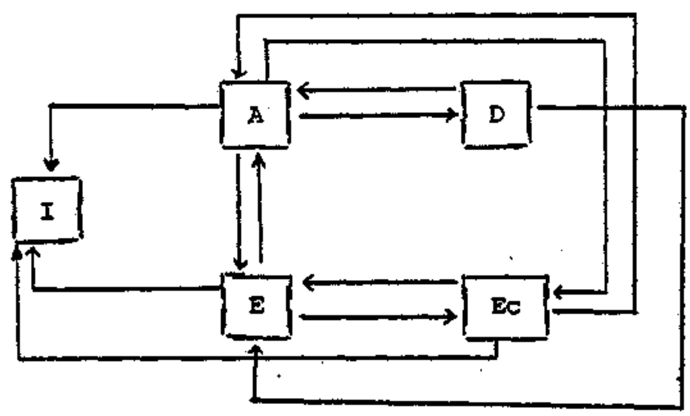
(iv) A wide range of scaling procedures are available to represent phenomena which are described by different values of points in space. An exploratory way of data analysis like a scaling procedure aims at deriving a quantitative representation of qualitative information (viz. measured at the nominal or ordinal scale). We will discuss briefly a recently developed tool for exploratory analysis of nominal data, viz. HOMALS (Homogeneity Analysis by Alternative Least Squares) (see section 3.5).

The aspects of an IEM with a satellite design will be presented by means of an application to the Dutch Biesbosch area. A central part of the analysis is made up by recreational activities with impacts on the natural environment and the regional economy. Information about recreational types and recreational activities is obtained by a survey analysis during the summer of 1983 with about 400 respondents and counts of recreational activities in the Biesbosch area in the same period.

2. SYSTEMS ANALYTICAL APPROACHES OF AN IEM.

2.1. Introduction.

An integrated environmental model consists of a set of modules which are interrelated to each other, while the background of such a model consists of many different disciplines, like economics, geography, natural environment, recreation, transportation, etc. The phenomena from these disciplines are represented by so called modules. An economic module for example, may include variables like production, consumption and employment, while an ecological module includes variables which reflect the association and diversity of ecosystems. However, in spite of the different phenomena of these disciplines, they interact with each other in an IEM.
A modeling approach to analyze the structure of strongly interrelated phenomena is in general necessary, and system analysis can be thought of as a research strategy whereby complex problems can be structured and decomposed into subsystems (see also Morgan, 1981). Consider for example a correlative representation of the links between modules from an IEM, represented in Figure 1 (see also Brouwer et al., 1983a).



D = demographic module
E = economic module
Ec = ecological module
A = artificial environment module
I = intermediate module

Figure 1. A simple correlative system of an IEM.

The correlative system im Figure 1 only presents the links between the modules, which denote the key-factors and are the links between the modules.

However, also links exist between variables within a module which are not represented in Figure 1.
A system simply consists in general terms of three elements, viz. an input, say x, and an output, say y, related to each other by a transfer function f, or: $y = f(x)$. The structure properties of a system like in Figure 1 can be characterized by the following points (see also Bennett and Chorley, 1978):

(i) the properties of each component (represented by modules) influence a system as a whole;

(ii)the properties of each component influence at least one other component.

We can conclude from these two points that a system cannot be partitioned into sets of mutual independent components. A system as a whole, just like in Figure 1, has properties which are quite different from the separate components so that "the behavior of the 'whole' system is usually much richer than the sum of the parts" (Wilson, 1981; p. 3).
A simple illustration of an IEM as the one presented in Figure 1 gives a description and summary of a series of interrelationships. The interrelationships between the modules are the links of a probable relationship and may be represented in either a quantitative or a qualitative way (see also Jones, 1983).
Some approaches of an IEM will be discussed hereafter. A so called horizontal model approach and a vertical model approach of an IEM will be discussed in section 2.2. A satellite model design of an IEM will be presented in section 2.3.

2.2. A Horizontal and a Vertical Model Approach

We will distinguish between two model approaches which deal with integrated modelling in this section, viz. a horizontal model approach and a vertical model approach.
A distinction between a horizontal model approach and a vertical model approach is interpreted in this section as alternative ways to link modules. Both approaches are illustrated by means of an example of an IEM.
A horizontal model approach is mainly characterized by interactions between monodisciplinary modules. All relevant modules have an equal contribution in the phase of model development and model operationalization, and the system is determined by interactions between monodisciplinary modules.
An example of a horizontal model approach with links between economy and natural environment at a regional level is developed by Duckstein et al. (1980, 1982). A mutual dependence exists in their analysis between economy and natural environment, which is represented by agricultural activities and

phosphorus loadings reduction, with the Lake Balaton area in Hungary as a case study example.

The conflicting goals of the analysis, viz. an increase of agricultural benefits (in monetary terms) and a reduction of phosphorus loadings (in terms of percentages) are related to each other in a multiobjective framework, and is presented in Figure 2.

Agricultural activities

monetary terms

↓

dimensionless figures

COMPROMISE PROGRAMMING

dimensionless figures

↑

percentage reduction

Phosphorus loadings

Figure 2. Trade-off between agricultural activities and phosphorus reduction.

The two variables in Figure 2 are transformed into dimensionless figures to be able to relate them to each other in a compromise programming approach and no distinction has been between the different phenomena of the variables.

A hierarchy of modules is considered in a vertical model approach, with one or more of them being superior to all others. The vertical model approach places special emphasis on the relationship between the dominant module and the other modules. The other possible relationships receive less attention and are assumed to be of only minor importance. A hierarchy of modules will be relevant in empirical situations when the level of one factor is the major point of study which is input for all other modules. A major advantage of the selection of a vertical approach is the considered hierarchy of modules which may be of practical use in case of a system being modelized by means of a decomposition into sub-systems. A hierarchy of modules is mainly selected because "initial attempts to cover all topics in a similar degree of detail have proved to be overambituous, in terms of staff time and data availabili-

ty, and more recently it has become almost standard practice to adopt an approach focusing upon selected topics with major implications for policy or for short-term investment programs" (Batey, 1984; p. 65).
An application of the hierarchy of modules can be illustrated by means of a policy analysis of watermanagement for the Netherlands (or, shortly PAWN-study) (see also Goeller et al., 1983). The aim of the PAWN-study is the management of alternative water policies of the Netherlands. The links between modules are presented in Figure 3.

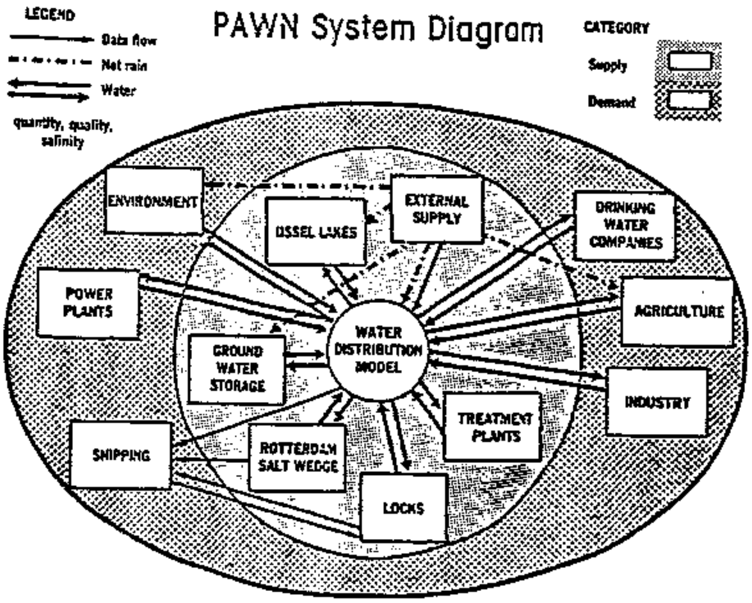


Figure 3. PAWN-system diagram. (Source: Goeller et al., 1983, p. 57)

Figure 3 shows that the system diagram consists of both a supply and a demand category. The central part of the analysis is the water distribution module which reflects the hierarchical nature of the model approach. The water distribution module simulates the distribution of water between the supply and demand categories. Only indirect links exist between the other modules.
An indirect link exists for example, between the agriculture module and the environment module with the water distribution module which acts as an intermediate module.
The vertical model approach is reflected in the PAWN-study by the hierarchical nature of the analysis with a set of modules which are linked to the water distribution module. The other possible relationships are treated to be of minor importance or neglected at all.

2.3. A satellite model approach of an IEM.

A model approach with characteristics of both the horizontal and the vertical model approach will be discussed now and is called a <u>satellite model design</u>.

The satellite design of an IEM will be presented by means of an empirical illustration of the Dutch Biesbosch area (see also Van der Ploeg et al., 1984). A recreational module is the core of the analysis and the other modules are a (regional) economic module, a natural environment module and a demographic module.
The satellite design is presented in Figure 4 below.

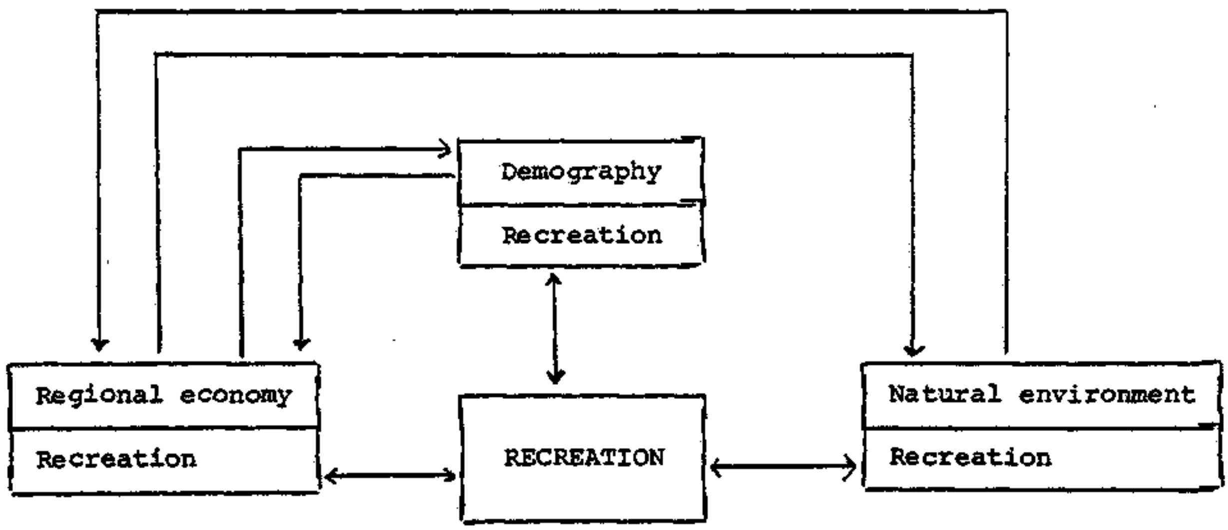


Figure 4. A correlative system of an IEM with a satellite design.

The satellite design in Figure 4 consists of three steps:

(i) the recreational aspects which relate to successively, a demographic module, a regional economic module and a natural environment module are determined;

(ii) the recreational aspects from the separate modules are linked to each other, to obtain a recreation module. The recreation module is the core of the analysis in this case;

(iii) the other aspects of the modules are modelled and linked to each other.

The relations which deal with steps (i) and (ii) are related to the recreation module in Figure 4 and are monodisciplinary relationships. The relationships dealing with step (iii) denote the links between modules and are for that reason multidisciplinary in nature.
A vertical modelling approach is followed in step (i) where the recreational aspects of the modules are determined. The hierarchy of the analysis, which

is a main characteristic of the vertical model approach, is denoted in the satellite design too. The recreational aspects of the modules denote the hierarchy in Figure 4.

A horizontal modelling approach is also used in the satellite design, viz. in step (iii) where the links are given between the non-hierarchical modules. In this third step of the satellite design for Figure 4 are given the links between the demographic, regional-economic and the natural environment modules.

A clear distinction will be made here between mondisciplinary and multidisciplinary relationships. Step (i) and step (ii) denote monodisciplinary relationships and the multidisciplinary relationships are represented by the last step of the satellite approach.

The main emphasis of the satellite design, like the example in Figure 4, is given to the recreational aspects of the modules. It is denoted at an aggregated level, because Figure 4 only shows in what way the links between modules are represented. No emphasis will be placed on the links between variables in this phase of the conceptualization of an IEM.

3. FOUR ASPECTS OF AN IEM.

3.1. Introduction.

Four aspects of an IEM will be discussed in this section, because of its relevance to link modules from different disciplines, with impacts at a regional level.

First,the distinction between the spatial aggregation level and the spatial scale level as well as their consequences for modelling results will be discussed in section 3.2. This geographical approach is relevant for models operationalized at a regional level. However, we note that the size of a region depends on the country: in the USA the size of a state is a regional level with size corresponding to a small European country. The size of a regional level is for that reason defined here between the national level and the local level; no further distinction will be made to characterize a region.

Secondly,the analysis of qualitative relations, especially the determination of sign-solvability conditions from a system of linear equations will be discussed in section 3.3. The relationships between variables are analysed in a qualitative way, i.e., when the sign of the impact between two variables is obtained from prior knowledge concerning the signs of the structural parameters.

The third point will be that different levels of measurement are available for variables in an IEM, viz. either metric(cardinal) or non-metric (ordinal, nominal, categorical). A family of qualitative statistical model approaches which belong to the Generalized Linear Models (GLMs) will be discussed in section 3.4.

Finally, an exploratory data analysis for nominal data which makes use of scaling procedures will be discussed briefly in section 3.5. The original nominal information is transformed into cardinal measured information by means of a scaling procedure. This data transformation is available by means of a loss of dimensions. An originally K-dimensional set of observations is transformed into a one- or two-dimensional quantitative (or metric) representation.

3.2. The spatial aggregation level and spatial scale level of an IEM.

The validity of nearly all applications of quantitative techniques with spatial data depend on the assumption that the spatial units are considered to be a priori given. This approach can be doubted by modellers to be a satisfactory geographical assumption for regional applications. The relevance of the zoning systems will be discussed here for that reason.

It can be seen easily that there will be a tremendously large number of different ways by which any study area can be subdivided into non-overlapping areal units and this is the essential point of the modifiable areal unit problem (abbreviated as MAUP) (see also Openshaw, 1983). The relevance of the zoning system was already mentioned by Kendall and Yule when they wrote that "our correlations will accordingly measure the relationship between the variates for the specified units chosen for the work. They have no absolute validity independently of these units, but are relative to them. They measure, as it were, not only the variations of the quantities under consideration, but the properties of the unit-mesh which we have imposed on the system in order to measure it" (Kendall and Yule, 1950, p. 312).

The MAUP can be subdivided into two separate but closely related problems, viz. the scale problem and the aggregation problem (see also Openshaw and Taylor, 1979; 1981).

The scale problem is the variation in model results that may be obtained when increasingly larger spatial units of analysis will be selected and is related to the micro and macro approaches in economic analysis. There are also many different possibilities to partition an area into a fixed number of spatial units, also called the aggregation level. The aggregation level will also affect model results.

The information obtained from some census for example, may be collected at the local, regional or national level. The different zoning systems provide alternative results when quantitative measures (i.e., mean values, correlations) are determined for each zone. Such alternative results are due to the scale problem because of larger spatial units of analysis.

It is noted that the MAUP is also closely related to what is called the ecological fallacy problem. An ecological fallacy here occurs when results that are based on aggregated data are applied to the individual who form the zones being studied. Some simple examples of the scale and aggregation levels are presented in Figure 5. The counts in this Figure denote nominal values.

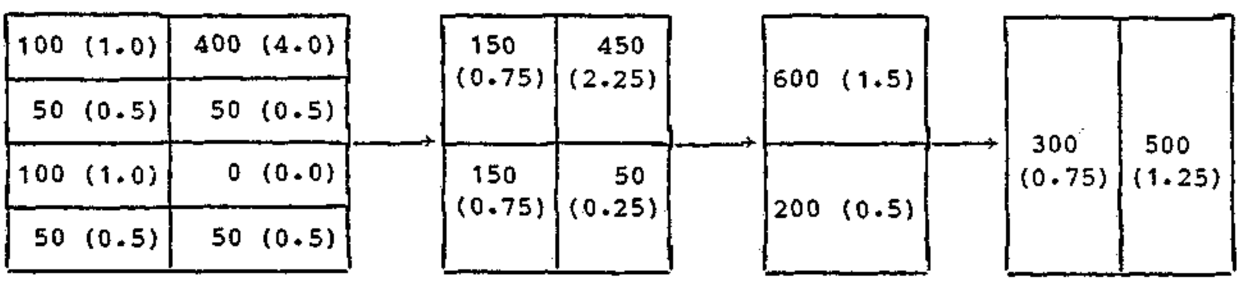


Figure 5a Figure 5b Figure 5c Figure 5d

Figure 5. Illustrative example of scale problem and aggregation problem.

Each count in Figure (5a) is assumed to be related to the number of people in that zone, which is considered to be equal to 100. Mean values for each zone are presented between brackets in Figure 5. The range of mean values in Figure 5 is successively $[0; 4.0]$, $[0.25; 2.25]$, $[0.5; 1.5]$ and $[0.75; 1,25]$. The scale problem is represented in Figures (5a), (5b) and (5c) because of the alternative aggregations into 8, 4 and 2 regions. The aggregation problem is illustrated by Figures (5c) and (5d) where two spatial units of the analysis have been selected and ordered in different ways.

Openshaw and Taylor (1979) discussed some illustrations of the scale and aggregation problems for correlation coefficients. The correlation coefficients they presented for their data set had values in the range between -0.999 and 0.999 depending on the scale level and aggregation level. They discussed by this example that a million or so correlation coefficients could be obtained by changes of the scale level and aggregation level.

Empirical examples of the sensitivity of model outcomes in regional-economic applications for the scale and aggregation level are given by Lohmoeller et al. (1985) and Nijkamp et al. (1984).

Because of the geographical nature of the MAUP Openshaw (1983) proposed to solve this problem in a geographical way. He therefore suggested an algorithm - the Automatic Zoning Procedure - which consists of a series of steps:

(i) selection of the scale level, viz. the number of regions required in the aggregation;

(ii) selection of a criterium function which will be optimized like, for example, optimization of the correlation coefficient between variables;

(iii) a level of aggregation will be computed in an iterative procedure while the criterium function is optimized.

The regional level has been subdivided into the scale level and the aggregation level and both of them are relevant for areal measures like density, pattern and average number of events per unit area. The essence of both parts should be notified when the concepts of an IEM are operationalized.

3.3. The analysis of model specification of an IEM

The relationships between variables are analysed in this section in a qualitative way, viz. a linear system $Ax = b$, where both the non-singular matrix A and the vector b contain qualitative information, in contrast with the traditionally used cardinally or metric defined systems. The qualitative information concerning the parameters a_{ij} and b_j, $i, j = 1, \ldots, n$ is either positive, negative or zero in nature. A zero impact denotes then absence of a theoretical relationship between some pair of variables.

The analysis of qualitative relations, or qualitative calculus, concerns with the solution of a linear system $Ax = b$ for vector x(called sign-solvability analysis) as well as the analysis of the qualitative stability conditions of a set of linear differential equations

$$\dot{y} = \frac{dy}{dt} = A\,y$$

(see also Brouwer and Nijkamp (1985) where qualitative stability is applied to some ecological predator-prey models).

Both elements of qualitative calculus are formulated in graph theoretical terms.

In the context of this paper we will only discuss the conditions of sign-solvability.

The analysis of sign-solvability of $Ax = b$ deals with the identification of the changes in vector x in a qualitative way (positive, negative or zero) due to changes in the vector of exogenous variables b (also either positive, negative or zero).

The main reasons for the development of sign-solvability and the treatment of qualitative information in economic modelling are twofold:

(i) "Ordinarily, the economist is not in possession of exact quantitative knowledge of the partial derivatives of his equilibrium conditions" (Samuelson, 1947, p. 26), because of the limited availability of suitable quantitative data.

(ii) The qualitative information about the various impacts may have a more solid empirical basis than the functional model structure (or model specification).

The major part of developments in the field of sign-solvability analysis took place in mathematics, economics and ecology (see also Greenberg and Maybee, 1981, where a wide range of theoretical developments with empirical evidence are discussed).

The elements of the analysis of qualitative relations will be discussed by means of an illustration. Consider for that reason the following set of linear equations with qualitative information:

$$\begin{bmatrix} - & + & - \\ - & - & - \\ + & 0 & - \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ - \\ 0 \end{bmatrix}$$

This set of equations is represented by means of graphs with elements (A, -b) in Figure 6.

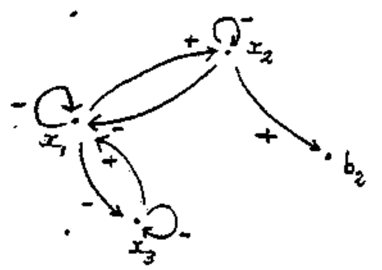


Figure 6. A graph representation of a qualitative model by signed digraphs.

Necessary and sufficient conditions for sign-solvability of Ax =b have been formulated by Bassett et al. (1968). These conditions make use of graph theoretical methods by signed digraphs (formulated as directed graphs with either a positive or a negative sign). The conditions with A considered to be a non-singular matrix are:

(i) the diagonal elements of A are all negative, i.e. $a_{ii} < 0$ for all i;

(ii) all cycles in the matrix A of length at least two are nonpositive or: $a_{ii_1} a_{i_1 i_2} \ldots\ldots a_{i_k j} < 0$ if $i_1 \neq i_2 \neq \ldots\ldots \neq i_k$ $(k > 1)$;

(iii) $b < 0$;

(iv) if $b_k < 0$, then every path $i \to k$ is non-negative for $i \neq k$.

All conditions for sign-solvability are fulfilled in the example from Figure 2 and the final solution $x = A^{-1}b$ becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} - & - & + \\ + & - & 0 \\ - & - & - \end{bmatrix} \begin{bmatrix} 0 \\ - \\ 0 \end{bmatrix} = \begin{bmatrix} + \\ + \\ + \end{bmatrix}$$

The changes in the variables x_1, x_2 and x_3 are all positive due to a decrease of the second exogenous variable.

The solution of the reduced form $x = A^{-1}b$ is obtained into two steps, viz. the inversion of the matrix A and the matrix A^{-1} multiplied by vector b.

Both steps need to have unique solutions up to their signs in order to be able to solve a linear set of equations uniquely.

The sign-solvability analysis can be interpreted as a kind of sensitivity analysis. The change of the endogenous variables(positive, negative or zero) is defined uniquely when the above mentioned conditions of sign-solvability hold,and in such cases it holds for all cardinal values of the matrix A and vector b up to their signs.

The relevance of sign-solvability can be found in case of not exactly quantified and not estimated equations because of a lack of (sufficiently reliable) data. The necessary and sufficient conditions for sign-solvability are formulated in terms of graphs to visualize the qualitative impacts between variables.

A disadvantage of qualitative calculus is that the conditions for sign-solvability are very severe, especially the second one about the non-positive cycles in a graph. See for example Brouwer et al. (1985) for a discussion of sign-solvability of the well-known and simple economic model of the U.S.A. developed by L.R. Klein. The conditions of sign-solvability are not fulfilled in that empirical example.

Therefore, a less restrictive qualitative approach may be very appealing. The conditions of sign-solvability can be supplemented with plausible information which is based on logical, empirical or theoretical evidence. For example, the share of consumption in national income is expected to be smaller than one because of logical and empirical evidence.

Another research direction in the field of sign-solvability is partial sign-solvability when the conditions (i) to (iv) hold for only a part of the set of equations,especially when the matrix A is reduced into a set of non- zero matrices and a zero matrix.

We may conclude from the above brief discussion that the analysis of qualitative relations can be regarded as a method to solve static and dynamic models with qualitative information. It is a major issue in integrated environmental modelling in case of lack of data or unreliability about the data.

3.4. Qualitative statistical models of an IEM.

Traditionally the statistical modelling approaches in geographic, regional-economic and recreation research dealt with metric or quantitative data measured at the interval or ratio scale. This research tradition followed the practice in natural science where information is collected under well defined conditions from controlled experiments.

However, the information is frequently non-metric (qualitative, discrete or categorical) in nature, when variables are either not well defined or information is obtained from questionaires, panel-studies or survey-sampling. For example, surveys may contain categorical data classified in a dichotomous or polytomous way such as sex (male/female), income (low level, medium-sized, high level) or recreational activities (land recreation, water recreation). Such variables can only be distinguished by their names or phenomena and are for that reason measured at a nominal scale.

The categorical variables can be represented in contingency tables, with cross-classified variables as cell-elements when two or more variables are classified according to their categories (see also Bishop et al., 1975).

Significant methodological progress has taken place in qualitative statistical modelling during the past twenty years. Particular attention will be placed here upon log-linear modelling and logistic-linear models for the analysis of contingency table. Such models will be specified within Nelder and Wedderburn's family of "generalized linear models" (GLMs).

A "generalized linear model" can be expressed in the form (see also McCullagh and Nelder, 1983; Nelder, 1985):

$$y_i = \mu_i + \varepsilon_i \quad , i = 1, \ldots, N \qquad (1)$$

where y_i is a response variable which is assumed to come from the exponential family of probability distributions;

μ_i is the mean of y_i, i.e. $\mu_i = E(y_i)$, also called the systematic component;

ε_i is a randomly distributed error term, also called the random component.

A GLM can be concisely characterized by the following three points:

(i) a response variable y is independently distributed with mean μ and a variance-covariance structure for the response variables determined by the underlying probability distribution function;

(ii) a linear relationship, say $\eta = X\beta$, is assumed to exist between η and the set of explanatory (or stimulus) variables; X is a matrix of explanatory variables and β a vector of parameters to be estimated. η is called the linear predictor;

(iii) a relationship is defined between η and the mean μ, viz. $\eta = g(\mu)$, where g is called a link-function between the linear-predictor and the mean.

Different types of generalized linear models may be obtained by varying either the distribution function of the response variable and/or the link-function g.

The ordinary least squares regression model with metric data can be interpreted in terms of GLMs, if the observations are considered to be independent and to follow a normal distribution with mean μ and constant variance σ^2, with a linear link-function ($\eta = \mu$). Some examples of GLMs are presented in Table 1.

Table 1. Examples of generalized linear models.

Model	Link function	Error distribution
Linear regression	Identity	Normal
ANOVA (fixed effects)	Identity	Normal
ANOVA (random effects)	Identity	Gamma
Logistic/logit regression	Logit	Binomial or multinomial
Log-linear model	Logarithmic	Poisson

Log-linear models and logistic regression models will be discussed briefly in this section because of its relevance for the analysis of qualitative and categorical data (see also among others Brouwer et al., 1983b; O'Brien and Wrigley, 1984 for a more detailed discussion of categorical data analysis). Log-linear models for contingency table analysis is an exploratory tool to decompose the structure of a contingency table into main effects and interaction effects. The so called maximum likelihood estimates of cell frequencies and model parameters are obtained from a log-linear model interpreted in GLM terms when a Poisson error distribution and a logarithmic link function is considered. Consider for example an I x J contingency table with two variables (A and B) classified successively into I and J categories, and cell-elements n_{ij} (i = 1, ..., I; j = 1, ..., J) (see also Table 2).

Table 2. I x J contingency table for variables A and B.

	variable B					Total
	n_{11}	...	n_{1j}	....	n_{1J}	n_{1+}
	.		.		.	
	.		.		.	
	.		.		.	
Variable A	n_{i1}	...	n_{ij}	....	n_{iJ}	n_{i+}
	.		.		.	
	.		.		.	
	.		.		.	
	n_{I1}	...	n_{Ij}	....	n_{IJ}	n_{I+}
Total	n_{+1}		n_{+j}		n_{+J}	

The expected cell-frequencies of Table 2, denoted by m_{ij}, can be expressed into the following log-linear models:

$$\ln m_{ij} = \lambda \qquad (2a)$$

$$\ln m_{ij} = \lambda + \lambda_i^A \qquad (2b)$$

$$\ln m_{ij} = \lambda + \lambda_j^B \qquad (2c)$$

$$\ln m_{ij} = \lambda + \lambda_i^A + \lambda_j^B \qquad (2d)$$

$$\ln m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \qquad (2e)$$

where λ is the grand-mean effect, λ_i^A and λ_j^B are the main effect parameters for variables A and B successively and λ_{ij}^{AB} is the first-order interaction effect between variables A and B. The parameters λ_i^A, λ_j^B and λ_{ij}^{AB} are defined as deviations from the grand-mean effect λ.

Model (2a) is called the basic model and is a simple model with expected frequencies of cell-elements equal to the average mean value (all cell-elements are equal probable and equal to the sample size N divided by the number of cell-elements). Model (2b) and model (2c) consist of the main effects for variables A and B successively. The main effects of variables A and B denote the variation which exist between successively the row sums and column sums of the contingency table. Model (2d) is the one where mutual independence is considered between the variables A and B. Model (2e) is called the saturated model with as many independent parameters as the number of cell-elements in the table and expected cell-frequencies which are equal to the observed ones by definition.

A selection of one of the mentioned log-linear models is based on the following two considerations:

(i) the model should be as simple as possible in terms of the number of parameters included;

(ii) the model should fit the data well in a statistical way.

The fit of a log-linear model for a two-way contingency table can be determined by a goodness-of-fit statistic (also called the <u>deviance</u>) like the likelihood-ratio value defined by:

$$\text{Deviance} = -2 \ln \lambda = \sum_{i,j} n_{ij} \ln \left(\frac{n_{ij}}{m_{ij}}\right). \qquad (3)$$

with degrees of freedom equal to the number of independent terms. The goodness-of-fit statistic follows a chi-square distribution asymptotically for a Poisson error function and a logarithmic link function.

The log-linear models in Formulae (2a) - (2e) can be generalized easily for higher order contingency tables when three or more variables are classified according to their categories.

Table 1 showed that the so-called logistic regression model is interpreted in the GLM-approach when a logit link function and a binomial or a multinomial error distribution function is considered.

In a logistic regression model, the logit transformation, i.e. the natural logarithm of the ratio of two odds is considered to be a linear function from a set of variables. The general form of the linear logistic model is given by:

$$\ln \left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_{k=1}^{K} \beta_k \cdot x_{ik} \qquad (4)$$

where:

p_i = probability of success of observation i, i = 1, ..., N;

X_{ik} = value of the k-th covariate or explanatory variable for observation i (k = 1, ..., K);

α and β_k are the (K+1) parameters to be estimated.

The left-hand side of (4) is called the logit transformation of p_i with values in the range between $-\infty$ and ∞ when p_i is in the range between 0 and 1 (see also Brouwer and Nijkamp, 1984).

Explanatory analysis of a categorical response variable cannot be employed within the ordinary least squares regression approach because of the following problems (see also Wrigley, 1979):

(i) heteroscedasticity or non-constant variance of the error terms. When regression analysis is applied to binary data with probability of occurrence of some event y_i being p_i, the variance differs for each event and is equal to $p_i(1-p_i)$, i = 1, ..., N.
Homoscedasticity or constant variance is a basic assumption in OLS-regression analysis. However, an effect of heteroscedasticity is that it will not result in consistent parameter estimates;

(ii) predictions may have values outside the meaningful range of its interpretation when the response variable is defined in terms of probabilities with values in the range between 0 and 1.

An interactive computer package for GLMs like the log-linear models in Formula (2) and the linear logistic model in Formula (4) is GLIM (Generalized Linear Interactive Modelling) developed by NAG (Numerical Algorithms Group) in Oxford (see also Baker and Nelder, 1978 for a manual of the GLIM 3 version). This package produces maximum likelihood values of parameter estimates and expected cell-frequencies and makes use of an iterative weighted least squares regression analysis.

This section showed that a wide range of statistical models either with qualitative or quantitative data can be interpreted in terms of GLMs by variation of the link function or the error distribution function. Applications of the linear-logistic models and the log-linear models are presented in section 4.

3.5. Homogeneous scaling for categorical and qualitative data in an IEM

An exploratory analysis with qualitative and categorical data by means of homogeneous scaling will be discussed in this section. The scaling procedure HOMALS (Homogeneity Analysis by Alternative Least Squares) is an algorithm developed in the field of psychometrics at the Department of Data Theory from

the University of Leiden (see also Gifi, 1981).
Homogeneous scaling procedures are based on the assumption that the variables included in the analysis are classified into a number of categories, and measured at the nominal scale. The main structure of the HOMALS program is given in Figure 7.

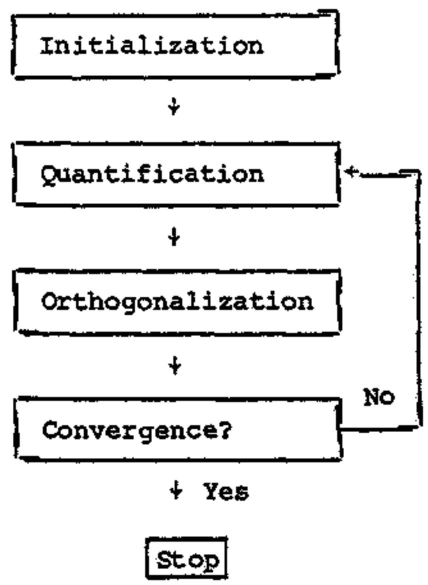


Figure 7. Main structure of the HOMALS program.

A quantitative configuration is obtained from a qualitatively measured system by a scaling procedure like HOMALS, which is besed on a goodness-of-fit criterion in the convergence step of Figure 7 between the categorical classified data information and the quantitative output.
The HOMALS output can be plotted in a N dimensional Euclidean space, with N equal to one or two (see also Nijkamp and Voogd, 1984, for an application of the HOMALS procedure to provincial infrastructure information).

4. APPLICATION OF AN IEM.

In this section, we will present some preliminary modelling results of integrated analysis for the Biesbosch area which makes use of the satellite principle proposed in Figure 4 in Section 2.3. The relevance of the tools discussed in sections 3.2 - 3.5 will be presented in this section too.
Different types of information is available:

(i) a survey sample of about 400 respondents during the summer of 1983. The respondents (recreationers of the Biesbosch area) gave their opinion about the environmental key factors of the ares, were classified according to their demographic natures (age, education), and answered questions about the way they spent their money during their stay in the area.

(ii) counts of recreational levels in 12 sub-areas of the Biesbosch during 8 days in the summer of 1983.

Table 2 shows the classification of the recreationers who participated in the survey to the variables 'areas' and 'days'. The areas of visit with five categories are classified into six days during the summer of 1983 (see also Van der Linden and Van Eijk, 1984).

Table 2. Characterization by area and date of recreationers.

day / area	June 11	June 25	July 13	July 16	August 10	August 13	Total	Name of area
1	35	19	15	14	35	0	118	Rietplaat
2	1	10	3	1	2	2	19	Honderdendertig
3	40	23	25	28	26	0	142	Keesjes Killeke
4	16	24	4	28	6	0	78	Merwelanden
5	8	12	0	1	5	2	28	Zuidhollandse Biesbosch
Total	100	88	47	72	74	4	385	

The demographic characteristics of the recreationers which is the core of the integrated model analysis, will be discussed first (see also Figure 4). Figure 8 shows the distance between the Biesbosch area and the area the recreationers live. This Figure shows that the major part of recreationers in the Biesbosch area have their home-address at a distance of less than 20 km from the Biesbosch. For the area Merwelanden even more than 50% of the respondents stay less than 10 kilometers from their home-address.

Figure 8. Distance between Biesbosch area and home adress for recreationers.

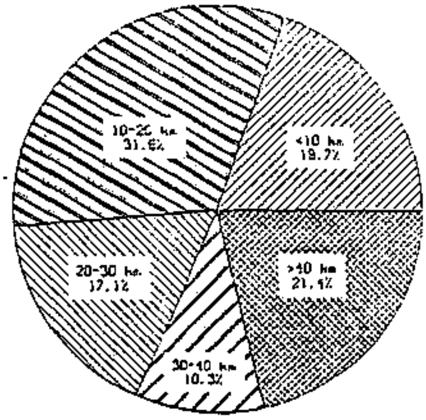


Figure 8a. Distance between Rietplaat and home-address.

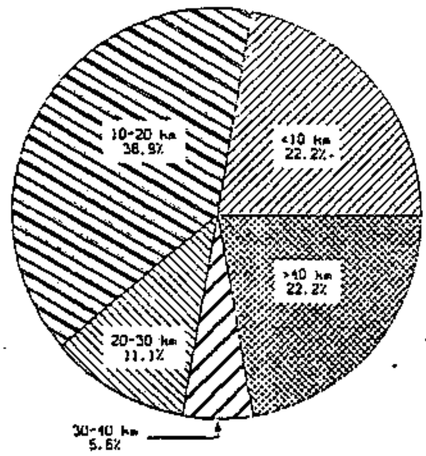


Figure 8b. Distance between Honderd-endertig and home-address.

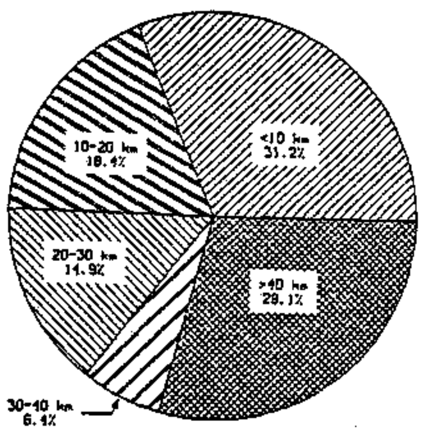


Figure 8c. Distance between Keesjes Killeke and home-address.

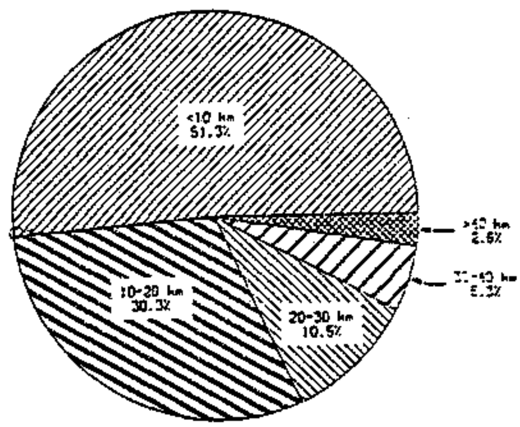


Figure 8d. Distance between Merwe-landen and home-address.

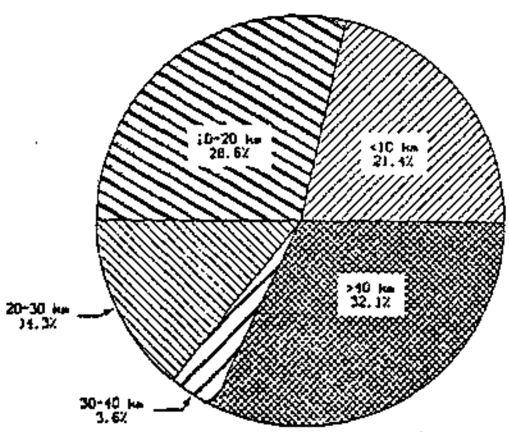


Figure 8e. Distance between Zuid-holl. Biesbosch and home-address.

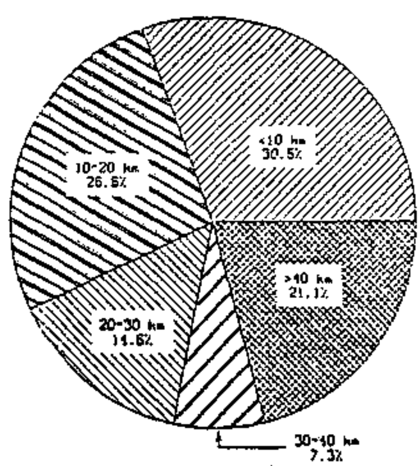


Figure 8f.Distance between Biesbosch and home-address.

The demographic characteristics of the recreationers will be discussed by means of qualitative statistical models. The aim of the analysis is to describe the difference in demographic characteristics of the recreationers for the five regions of the Biesbosch area.
A classification of recreationers to area, age and education is presented in Table 3 by means of a 5 x 3 x 3 contingency table.

Table 3. A three-way contingency table to cross-classify the variables area, age and education.

age	education = 1			education = 2			education = 3			
area	< 29	30-50	> 50	< 29	30-50	> 50	< 29	30-50	> 50	Total
1	9	29	18	10	23	3	4	12	1	109
2	0	3	1	4	1	0	1	7	0	17
3	10	26	19	17	21	12	13	13	6	137
4	7	15	28	5	8	5	3	2	3	76
5	2	4	5	4	2	4	1	3	1	26
Total	28	77	71	40	55	24	22	37	11	365

area: 1 = Rietplaat
2 = Honderdendertig
3 = Keesjes Killeke
4 = Merwelanden
5 = Zuidholl. Biesbosch

(missing values = 20)

education: 1 = primary school
2 = secondary school
3 = higher education

The interactions between the variables 'area','age' and 'education' are quantified and tested by means of a set of log-linear models. The variables 'area','age' and 'education' are represented here by successively A,B and C. Results will be presented for the following log-linear models for contingency table analysis:

$$\log_e m_{ijk} = \lambda \tag{5-a}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C \tag{5-b}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} \tag{5-c}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} \tag{5-d}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC} \tag{5-e}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} \tag{5-f}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} \tag{5-g}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \tag{5-h}$$

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \tag{5-i}$$

Models (5-a)- (5-i) are a three-way generalization of the log-linear models in (2-a)- (2-e). Model (5-b) is the one where all main effects are included for the variables A,B and C and mutual independence is considered for each pair of variables. Models (5-c) -(5-i) are log-linear models with one or more interaction effects between pairs of variables. Multiple independence has been considered in (5-c) to (5-e): a joint pair of variables is independent of the third variable. Conditional independence has been considered in (5-f) to (5-h): two variables are independent given the level of the third variable. All models are so-called hierarchical in nature because inclusion of some interaction effect also considers the inclusion of all possible lower-order interaction effects.

The deviance values of the log-linear models (5-a) - (5-i) for the contingency table in Table 3 are presented in Table 4. The results are obtained by the computer package GLIM (Generalized Linear Interactive Modeling), developed by the NAG (Numerical Algorithms Group).

It should be noted that the 5 x 3 x 3 table has some zero cell-frequencies, also called sampling zeros which become non-zero, at least in a theoretical way, when the sample size is sufficiently large. The logarithms of such zero cell-frequencies are not defined and for that reason all cell-frequencies are increased with a small value. The sensitivity of values in the range between 0.01 and 0.5 are presented also in Table 4.

Table 4. Deviance values of log-linear models with a 5 x 3 x 3 contingency table.

Model	df	crit.value	0.01	0.05	0.1	0.2	0.3	0.4	0.5
(5-a)	44	63.32	326.7	324.0	321.0	315.5	310.5	305.7	301.2
(5-b)	36	50.96	89.37	88.09	86.80	84.65	82.81	81.20	79.73
(5-c)	28	41.34	61.74	60.77	59.83	58.26	57.04	55.92	54.91
(5-d)	28	41.34	66.63	65.48	64.33	62.44	60.87	59.50	58.27
(5-e)	32	47.17	61.96	60.87	59.80	58.08	56.66	55.44	54.37
(5-f)	20	31.41	39.01	38.16	37.36	36.11	35.10	34.22	33.44
(5-g)	24	36.42	34.75	33.92	32.83	31.74	30.89	30.17	29.54
(5-h)	24	36.42	39.87	38.64	37.33	35.88	34.72	33.75	32.90
(5-i)	16	26.30	17.03	16.28	15.61	14.61	13.85	13.21	12.67

The deviance values in Table 4 are asymptotically distributed with a chi-square distribution with number of degrees of freedom (or the number of independent parameters) and corresponding critical value at the 95% level given in column 2 and column 3 successively.

The aim of the analysis is to select a log-linear model from the set (5-a) - (5-i) which fits the data well (in terms of a deviance value which is less than the critical value) and is also parsimonious in the number of parameters. Models (5-g) and (5-i) both fit the data. Model (5-g) will be selected from these two, because of the parsimonity criterium and the interaction effect between the variables 'area' and 'education' is for that reason not included.
The variation in cell-frequencies of a simultaneous analysis of the variables 'area' and 'education' do not give a significant decrease of deviance value, which follows from model (5-e) and model (5-b).

Analogous results are presented for the contingency table with the variables 'area classified' with 'age of the respondents' and the 'distance between home address'. The contingency table is given in Table 5.

Table 5. 5 x 3 x 3 contingency table to cross-classify the variables area, distance and age

distance	Age: < 29 year			30 - 50 year			> 50 year		
area	< 20 km	20-40 km	> 40 km	< 20 km	20-40 km	> 40 km	< 20 km	20-40 km	> 40 km
1	15	6	3	32	17	20	12	9	2
2	3	1	0	6	2	4	2	0	0
3	24	6	10	28	14	21	18	10	10
4	15	0	0	20	3	1	27	9	1
5	6	1	0	2	2	6	6	2	3
Total	63	14	13	88	38	52	65	30	16

The deviance values of the log-linear models (5-a) - (5-i) for this three-way table are presented in Table 6.

Table 6. Deviance values of 9 log-linear models with a 5 x 3 x 3 contingency table.

Model	df	crit.value	0.01	0.05	0.1	0.2	0.3	0.4	0.5
(5-a)	44	63.32	386.8	383.1	379.0	371.8	365.3	359.3	353.7
(5-b)	36	50.96	99.50	97.30	95.14	91.65	88.76	86.26	84.04
(5-c)	28	41.34	61.56	59.79	58.15	55.60	53.59	51.90	50.43
(5-d)	28	41.34	71.17	69.20	67.30	64.28	61.84	59.74	57.90
(5-e)	32	47.17	82.33	80.34	78.42	75.39	72.95	70.87	69.05
(5-f)	20	31.41	33.24	31.69	30.30	28.24	26.67	25.38	24.29
(5-g)	24	36.42	44.40	42.83	41.42	39.35	37.78	36.51	35.45
(5-h)	24	36.42	54.01	52.23	50.12	48.03	46.02	44.35	42.91
(5-i)	16	26.30	16.05	14.72	13.59	12.04	10.94	10.10	9.430

Model (5-f) fits the data well when at least a value of 0.1 is added to the cell-frequencies (because of the occurrence of sampling zeros).In model (5-f) are included the interaction effects between the variables 'area' and 'distance' and the variables 'area' and 'age'. The possible inclusion of the interaction effect between the variables 'distance' and 'age' does not improve the log-linear model in a significant way (in terms of a significant decrease of the deviance value).

The deviance values in table 4 shows that the interaction effects between the variables 'area' and 'age' and the variables 'age' and 'education' both give a significant decrease of the deviance values. This means that the variation of the variable 'area' in Table 3 is related to the variation in the variable 'age'. The same conclusion holds for the pair-wise variation between the variables 'age' and 'education'.

The log-linear models from Tabel 5 - with deviance values presented in Table 6 - show that model (5-f) is the parsimonious one which fits the data well in a statistical way. The pairwise variation between the variables 'area' and 'distance' as well as 'area' and 'age' both give a significant decrease of the deviance values and fit the data well.

The distinction between models (5-f) and (5-i) is the inclusion of the interaction effect between the variables 'distance' and 'age'. This interaction effect is included due to a loss of 4 degrees of freedom (the number of parameters which denote the interaction between 'distance' and 'age').

The regional economic impact of recreational activities by means of log-linear models again classified according to the area of visit is presented in

Table 7. The contingency table shows the cross-classification of the variables 'area', 'upkeep by boat' and 'amount of money' (expenditures of upkeep by boat).

Table 7. A three-way contingency table to cross-classify the variables area, upkeep and amount of money.

upkeep	amount = 1		amount = 2		amount = 3		amount = 4		amount = 5		amount = 6	
area	1	2	1	2	1	2	1	2	1	2	1	2
1	4	0	4	2	11	6	29	16	11	7	8	4
2	0	0	3	1	3	2	3	2	1	0	0	0
3	4	3	9	2	9	6	34	8	19	5	9	4
4	3	1	14	0	17	1	18	2	9	2	4	0
5	1	0	1	0	2	1	4	2	3	1	0	5
Total	12	4	31	5	42	16	88	30	43	15	21	13

(N = 320)

Upkeep: 1 = money spent in the Biesbosch area;
2 = money spent outside the Biesbosch area.

Amount: 1 = nil
2 = hfl. 10 - 100
3 = hfl. 100 - 200
4 = hfl. 200 - 500
5 = hfl. 500 - 1000
6 = > 1000

The results of the log-linear modelling analysis of the 5 x 2 x 6 contingency table are presented in Table 8.

Table 8. Deviance values of log-linear models with a 5 x 2 x 6 table.

Model	df	crit.value	0.01	0.05	0.1	0.2	0.3	0.4	0.5
(5-a)	59	77.92	408.8	403.2	397.3	387.0	378.0	369.8	362.3
(5-b)	49	66.33	77.72	74.97	72.55	69.08	66.55	64.56	62.94
(5-c)	45	64.49	56.34	53.75	51.54	48.44	46.24	44.56	43.21
(5-d)	29	42.56	46.75	45.05	43.56	41.42	39.84	38.58	37.54
(5-e)	44	63.32	72.08	69.43	67.13	63.84	61.46	59.59	58.06
(5-f)	25	37.65	25.36	23.83	22.55	20.78	19.54	18.58	17.82
(5-g)	40	55.76	50.70	48.21	46.11	43.20	41.15	39.59	38.33
(5-h)	24	36.42	41.11	39.51	38.13	36.18	34.75	33.61	32.66
(5-i)	20	31.41	21.92	20.41	19.15	17.41	16.18	15.23	14.46

The results of Table 8 show that the parsimonious model which fits the data well is model (5-c) with 45 degrees of freedom. This is the hierarchical model with the interaction-effect between 'area' and 'upkeep' (whether money is spent inside or outside the Biesbosch area). The interaction-effect between the second and third variable does not lead to a significant improvement of the log-linear model (see also the difference in deviance values - with a loss of 5 degrees of freedom - between models (5-b) and (5-e)).
Table 9 shows the cross-classification of the variables 'night' (with values 1 or 2 depending on whether the night is spent at home or in the Biesbosch area) distance between home address and the Biesbosch area and the duration of the stay of recreationers.

Table 9. A three-way contingency table to cross-classify three variables.

distance	days = 1				days = 2				days = 3			
night	<10 km	10-20 km	20-40 km	>40 km	<10 km	10-20 km	20-40 km	>40 km	<10 km	10-20 km	20-40 km	>40 km
1	35	11	7	14	3	5	4	2	1	2	1	1
2	2	4	1	4	53	60	43	34	19	17	28	23
Totaal	37	15	8	18	56	65	47	36	20	19	29	24

night: 1 = night spent at home
2 = night spent in the Biesbosch area

days: 1 = duration of stay is one day
2 = duration of stay is two or three days
3 = duration of stay is at least four days

The linear logistic models which will be analysed with this contingency table are:

$$\log_e\left(\frac{P_{1i}}{1-P_{1i}}\right) = \beta_0 \qquad i = 1,\ldots 12 \qquad (6\text{-a})$$

$$\log_e\left(\frac{P_{1i}}{1-P_{1i}}\right) = \beta_0 + \beta_1 X_{1i} \qquad i = 1,\ldots,12 \qquad (6\text{-b})$$

$$\log_e\left(\frac{P_{1i}}{1-P_{1i}}\right) = \beta_0 \qquad + \beta_2 X_{2i} \qquad i = 1,\ldots,12 \qquad (6\text{-c})$$

$$\log_e\left(\frac{P_{1i}}{1-P_{1i}}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \qquad i = 1,\ldots 12 \qquad (6\text{-d})$$

The explanatory variables X_1 and X_2 (both with levels i, i=1,..,12) denote successively the distance between home-address and Biesbosch area and the duration of stay.
The models (6-a)- (6-d) denote whether the variation which exists in Table 9 for the variable 'night' is explained by the variation of the variables 'distance' or 'days' or by both of them.
Results of the analysis are presented in Table 10.

Table 10. Deviance values by means of linear-logistic models.

Model	Deviance	df	critical value
(6-a)	206.1	11	19.68
(6-b)	192.9	8	15.51
(6-c)	7.084	9	16.92
(6-d)	5.567	6	12.59

A parsimonious model which fits the data well with this 2 x 4 x 3 table by means of a logistic-linear model is equation (6-c), with deviance value 7.084 and 9 degrees of freedom.

The variables 'days' will give a significant decrease in the deviance value to the variation of the variable 'night'. However, the 'distance' variable will not give a significant decrease to the deviance value.

An exploratory analysis for the respondents of the survey-sample makes use of the scaling procedure HOMALS for nominal data.

Four variables, with 24 categories in total are represented in a two-dimentional space. The variables are:

(i) type of boat, in Figure 9 denoted by a dotted line with categories 1 to 6;

(ii) distance between home address and the Biesbosch area, in Figure 9 denoted by a dashed line with categories 1 to 7 which represent increased distance;

(iii) level of education, in Figure 9 denoted by a chain-dotted line, with categories 1 to 6 for increasing level of education (primary school to university);

(iv) age, in Figure 9 denoted by a chain-dashed line, with categories 1 to 6 for higher ages.

Figure 9 gives a graphical representation of the category quantification for each of the 5 sub-areas of the Biesbosch region.

The discrimination measures per variable and per dimension as well as the eigenvalues for factor 1 and factor 2 are given in Table 11 for the 5 Biesbosch areas.

Table 11. Discrimination measures and eigenvalues for factors 1 and 2.

area	variables								eigenvalues	
	1		2		3		4		1	2
Rietplaat	0.412	0.437	0.093	0.396	0.682	0.458	0.560	0.326	0.4368	0.4043
Honderdendertig	0.527	0.807	0.424	0.522	0.855	0.623	0.843	0.445	0.6623	0.5991
Keesjes Killeke	0.682	0.427	0.246	0.338	0.370	0.447	0.705	0.435	0.5004	0.4115
Merwelanden	0.552	0.179	0.601	0.374	0.710	0.657	0.408	0.657	0.5676	0.4669
Zuidholl. Biesbosch	0.817	0.393	0.800	0.595	0.669	0.520	0.789	0.846	0.7687	0.5886

The bivariate interactions between the levels of the four variables are presented in Figures (9a) -(9e) for the five areas.

An erratic movement of some line shows that a wide variation exists for the nature of the corresponding variable. Consider for example the variation of the variable 'distance' in the areas Rietplaat and Merwelanden. The respondents in the area Merwelanden with a distance of more than fifty kilometers from their home-address have a quite different nature than the other respondents in this area.

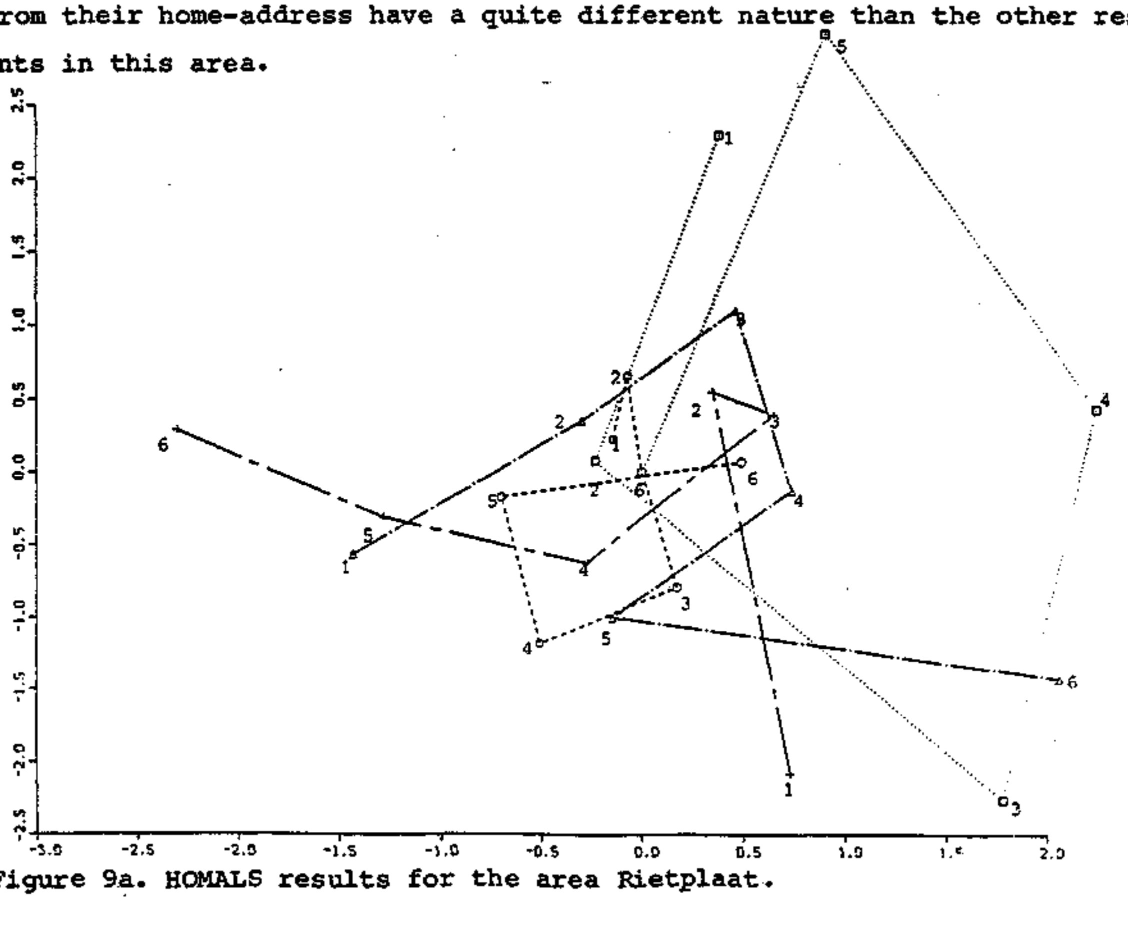


Figure 9a. HOMALS results for the area Rietplaat.

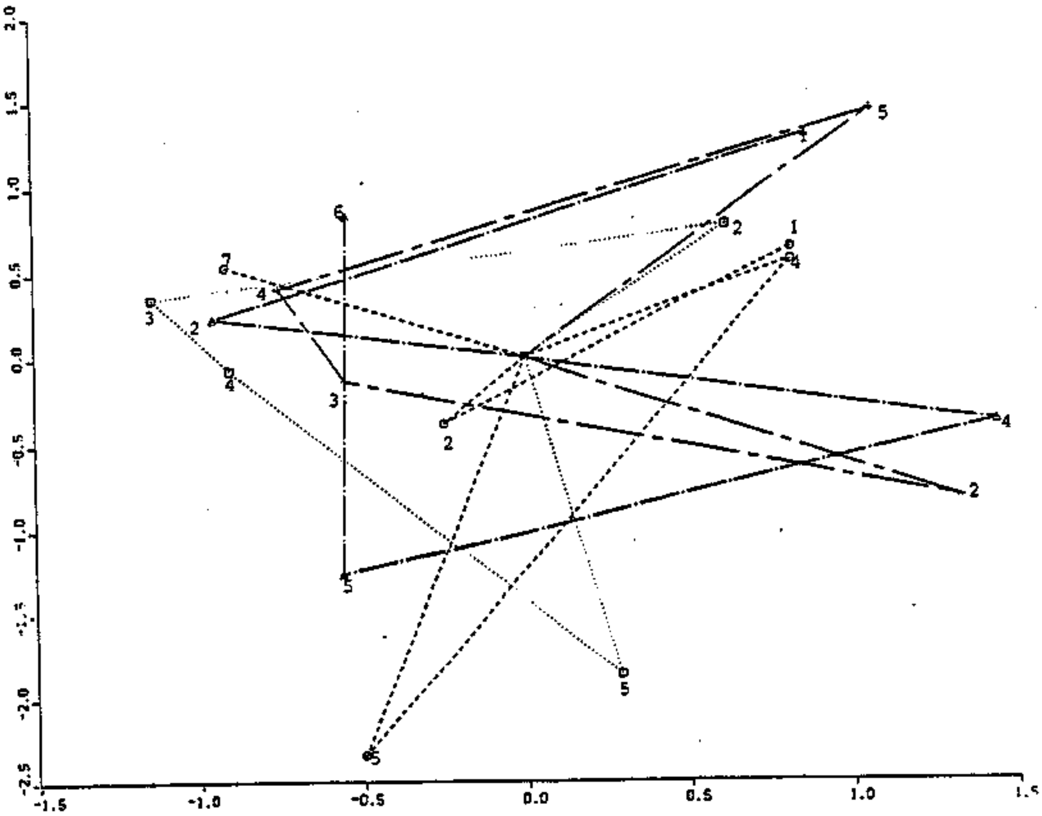


Figure 9b. HOMALS results for the area Honderdendertig.

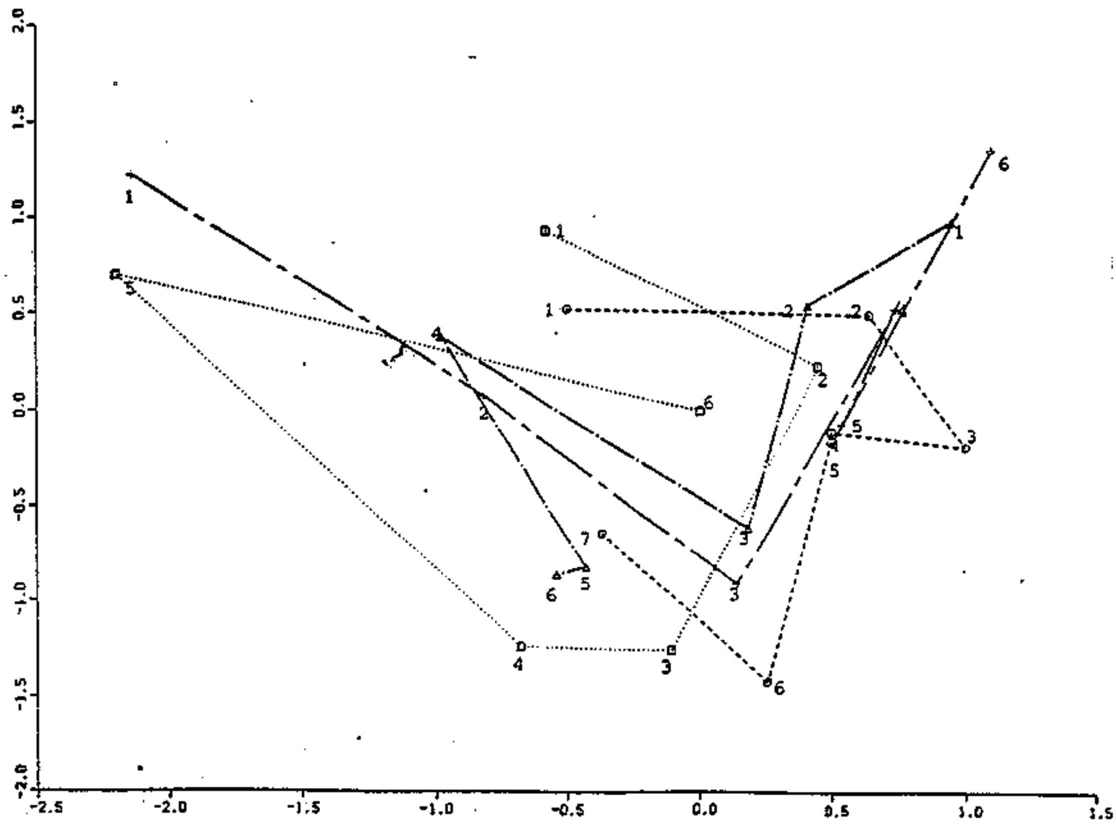


Figure 9c. HOMALS results for the area Keesjes Killeke.

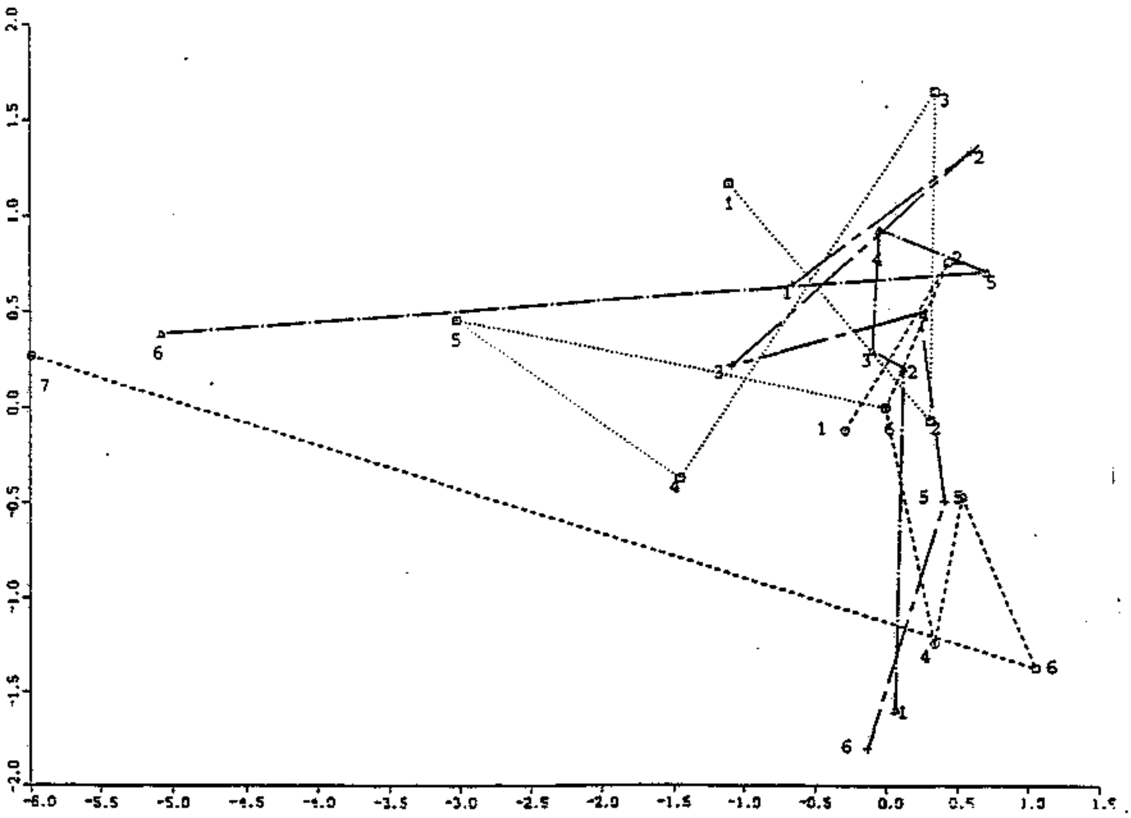


Figure 9d. HOMALS results for the area Merwelanden.

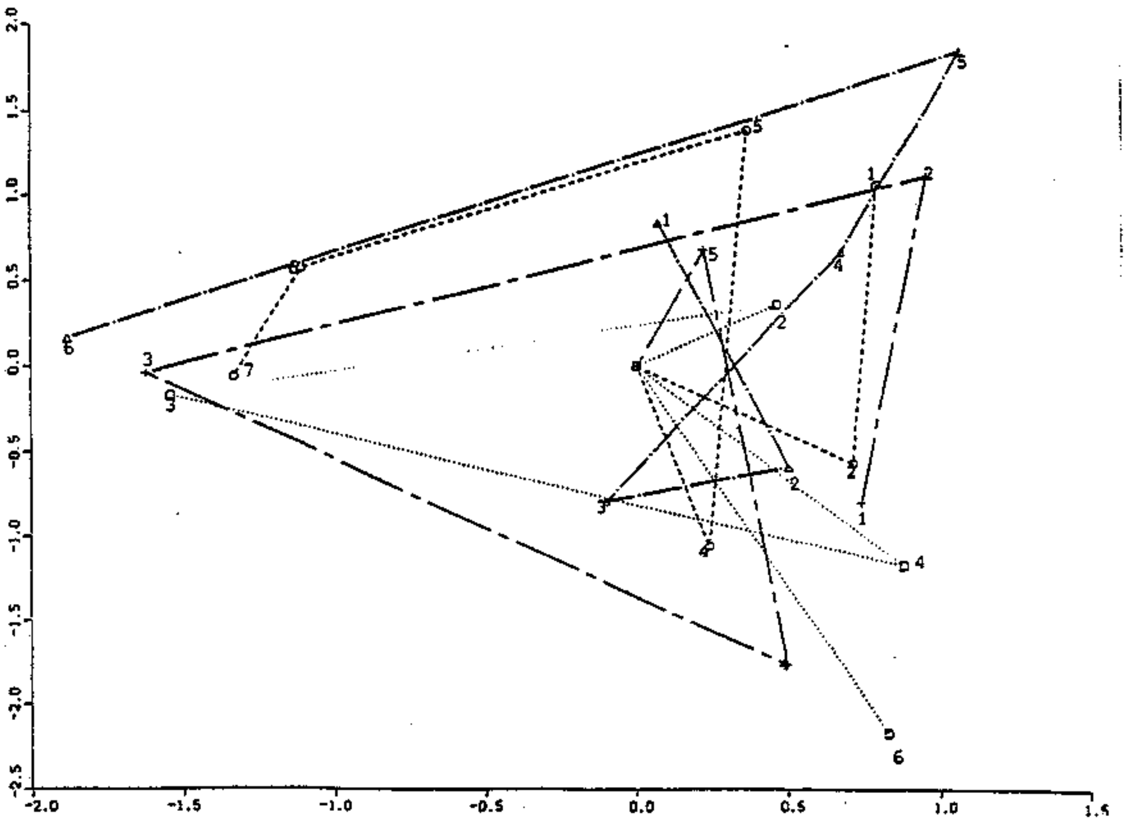


Figure 9e. HOMALS result for the area Zuid-hollandse Biesbosch.

Sign-solvability analysis will be discussed now for the analysis of qualitative relations. Consider for example the following set of linear equations:

$$y_1 = f(y_1, y_3, d)$$
$$y_2 = g(y_1, y_3, d)$$
$$y_3 = h(y_1, y_2)$$

with: y_1 = recreational activities;
y_2 = employment related to recreational activities;
y_3 = population level around the Biesbosch area;
d = recreational facilities (exogenous variable).

The model structure is represented in terms of signed digraphs in Figure 10,and the analogous matrix formulation becomes:

$$\begin{bmatrix} + & 0 & - \\ - & 0 & - \\ + & + & 0 \end{bmatrix} \begin{bmatrix} dy_1 \\ dy_2 \\ dy_3 \end{bmatrix} = \begin{bmatrix} + \\ + \\ 0 \end{bmatrix}$$

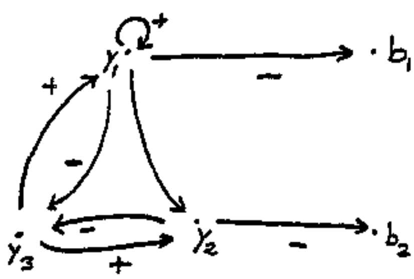


Figure 10. Model structure with signed digraphs.

The inversion of the matrix A is defined uniquely up to its sign because the conditions (i) and (ii) of sign-solvability analysis hold after row and column operations. The row and column operations (viz. permutation of any two rows and columns, reversement of signs in any row and any column) do not affect the solution vector. The multiplication of the matrix A^{-1} with vector b is not defined uniquely up to its sign and becomes:

$$\begin{bmatrix} dy_1 \\ dy_2 \\ dy_3 \end{bmatrix} = \begin{bmatrix} + & - & 0 \\ - & + & + \\ - & - & 0 \end{bmatrix} \begin{bmatrix} + \\ + \\ 0 \end{bmatrix} = \begin{bmatrix} * \\ * \\ - \end{bmatrix}$$

The result of the analysis will be that the third endogenous variable will decrease due to an increase of two exogenous variables given the structure of the qualitative impacts between the endogenous variables. The signs for the other endogenous variables are undefined because the conditions for sign-solvability do not hold. Additional information(for example, a priori or plausible parameter restrictions) is needed to determine the sign of the first and second endogenous variables.

5. CONCLUSION.

The aims of the paper were threefold.

First, the concept of an IEM is specified and discussed and this concept makes use of the so-called satellite concept, with one or more modules as the core of analysis. The concept of the satellite design deals with both the horizontal and the vertical model design. The main advantage of the satellite design is that it can be applied in situations where one or more modules are the key-factors in the analysis and when they affect the other modules, while the interelationships between the other modules are included as well.

The second aim of the paper was the discussion of four aspects of an IEM, viz. (i) the spatial aggregation level and spatial scale level, (ii) the use of qualitative information in integrated modelling, (iii) the use of metric and non-metric information in statistical modelling and (iv) exploratory data analysis to represent nominal information in a cardinal metric.

Finally, preliminary modelling results are presented for the Biesbosch area in the Netherlands to apply the above mentioned concept and four aspects. The Biesbosch area is a region with a large potential of recreational activities and impacts on both the regional-economy and the natural environment.

REFERENCES

Baker, R.J., and J.A. Nelder, The GLIM Manual: Release 3, Numerical Algorithm Group, Oxford, 1978.

Bassettt, L., J.S. Maybee and J. Quirk, 1968, Qualitative Economics and the Scope of the Correspondence Principle, Econometrica vol. 36, pp. 544-563.

Batey, P.W.J., 1984, Information for long-term Planning of Regional Development, in: P. Nijkamp and P. Rietveld (eds.), Information systems for integrated Regional Planning, North-Holland Publishing Company, Amsterdam, pp. 63-79.

Bennett, R.J., and R.J. Chorley, 1978, Environmental Systems: Philosophy, Analysis and Control, Methuen, London.

Bishop, Y.M., Fienberg, S.E., and P.W. Holland, 1975, Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge Massachusetts.

Brouwer, F., and P. Nijkamp, 1984, Linear logit Models for Categorical Data in Spatial Mobility Analysis, Economic Geography, vol 60 (2), pp. 102-110.

Brouwer, F., and P. Nijkamp, 1985, Qualitative Structure Analysis of Complex Systems, in: P. Nijkamp, H. Leitner and N. Wrigley (eds.), Measuring the Unmeasurable, Martinus Nijhoff Publishers, Dordrecht, The Netherlands, p. 509-530.

Brouwer, F., J-P. Hettelingh, and L. Hordijk, 1983a, an Integrated Regional Model for Economic-Ecological-Demographic-Facility Interactions, Papers of the Regional Science Association, vol. 52, pp. 87-103.

Brouwer, F., P. Nijkamp, and H. Scholten, 1983b, Hybrid log-linear Models for Spatial Interaction and Stability Analysis, Research Memorandum 1983-17, Department of Economics, Free University, Amsterdam.

Brouwer, F., P. Nijkamp, and H. Voogd, 1985, Mixed Qualitative Calculus and Economic Modeling, Research Memorandum 1985-8, Department of Economics, Free University, Amsterdam.

Duckstein, L., Bogardi, I., and L. David, 1980, Multiobjective Control of Nutrient Loading into a Lake, in: Y. Haimes, J. Kindler (eds.), Water and Related Land Resource Systems, pp. 413-418.

Duckstein, L., Bogardi, I., and L. David, 1982, Dual Objective Control of Nutrient Loading into a Lake, Water Resources Bulletin vol. 18 (1), pp. 21-26.

Gifi, A., 1981, HOMALS Users Guide, Department of Datatheory, Leiden, the Netherlands.

Goeller, B.F. et al., 1983, Policy Analysis of Water Management for the Netherlands, Report R-2500/1, Rand Corporation, Santa Monica.

Greenberg, H.J., and J.S. Maybee (eds.), 1981, Computer Assisted Analysis and Model Simplification, Academic Press, New York.

Jones, G.E., 1983, The Usefulness of System Theory in Ecosystem Studies, Area, vol. 15 (2), pp. 111-116.

Kendall, M.G., and G.U. Yule, 1950, An Introduction to the Theory of Statistics, Charles Griffin, London.

Linden, J.W. van der, and T. van Eijk, 1984, Recreatie gedrag en recreatiepatroon, Instituut voor Milieuvraagstukke, Amsterdam.

Lohmoeller, J-B. et al., 1985, Unemployment and the Rise of National Socialism: Contradicting Results from different Regional Aggregations, in: P. Nijkamp, H. Leitner and N. Wrighley (eds.), Measuring the Unmeasurable, Martinus Nijhoff Publishers, Dordrecht, the Netherlands, pp. 357-370.

McCullagh, P., and J.A. Nelder, 1983, Generalized Linear Models, Chapman and Hall, London.

Morgan, R.K., 1981, Systems analysis: a problem of methodology?, Area vol. 13(3), pp.219-230.

Nelder, J.A., 1985, Statistical Models for Qualitative Data, in: P. Nijkamp, H. Leitner, and N. Wrighley (eds.), Measuring the Unmeasurable, Martinus Nijhoff Publishers, Dordrecht, the Netherlands, pp. 31-38.

Nijkamp, P., and H. Voogd, 1984, Multidimensional and homogeneous scaling in spatial analysis, in: G. Bahrenberg, M.M. Fischer and P. Nijkamp (eds.), Recent Developments in Spatial Data Analysis: Methodology, Measurement, Models, Gower, Aldershot, pp. 201-214.

Nijkamp, P., Rietveld, P. and A. Rima, 1984, Information Content of Data from different Spatial Aggregation Levels, in: P. Nijkamp and P. Rietveld (eds.), Information Systems for Integrated Regional Planning, North-Holland Publishing Company, Amsterdam, pp. 215-228.

O'Brien, L.G., and N. Wrigley, 1984, A Generalised Linear Models Approach to Categorical Data Analysis: Theory and Applications in Geography and Regional Science, in: G. Bahrenberg, M.M. Fischer, and P. Nijkamp (eds.), Recent Development in Spatial Data Analysis: Methodology, Measurement, Models, Gower Publishing Company, Aldershot, pp. 231-251.

Openshaw, S., 1983, The modifiable areal unit problem, Series: Concepts and Techniques in Modern Geography, vol. 38 (Norwich, Geo Books).

Openshaw, S., and P.J. Taylor, 1979, A million or so correlation coefficients; three experiments on the modifiable areal unit problem, in: N. Wrigley and R.J. Bennett (eds.), Quantitative Geography: a British View, Routledge and Kegan Paul, London, pp. 60-69.

Ploeg, S.W.F. van der,et al., 1984, Openluchtrecreatie en natuurlijk milieu, Instituut voor Milieuvraagstukken, Amsterdam.

Samuelson, P., 1947, Foundations of Economic Analysis, Harvard University Press.

Wilson, A.G., 1981, Geography and the Environment, John Wiley & Sons, New York.

Wrigley, N., 1979, Developments in the Statistical Analysis of Categorical Data, Progress in Human Geography, vol. 3, pp. 315-355.