

## **MODELING ENVIRONMENTAL- ECONOMIC PHENOMENA AT A REGIONAL LEVEL: METHODOLOGICAL PROGRESS**

**FLOOR BROUWER**

*Agricultural Economics Research Institute  
The Hague, The Netherlands*

**PETER NIJKAMP**

*Free University, Amsterdam, The Netherlands*

### **ABSTRACT**

The availability and nature of information are important issues in modeling environmental-economic phenomena. This article addresses three types of limitation that may characterize such issues, including 1) insufficient precision of information for drawing reliable quantitative conclusions regarding policy aspects, 2) the wide variety of means of data gathering with their corresponding varying levels of measurement, and 3) the multivariate nature of environmental-economic phenomena. Various appropriate methodological tools have been developed over the past years to deal with such limitations. The relevance of the approaches will be illustrated for a regional-economic and environmental analysis of recreation.

### **INTRODUCTION**

Since the end of the sixties and the beginning of the seventies regional economic models have gained increasing popularity, and are now widely used in the context of regional planning and policy making in order to investigate the consequences of alternative policy instruments or conflicting policy options. These regional economic models allow various types of economic variables and sociodemographic trends to be taken into account (such as production, consumption, demand for production factors, as well as labor force development). The main progress and state-of-the-art of multiregional economic modeling

(e.g., the inclusion of multiregional phenomena and interregional interactions, such as flows of commodities or migration of people) is discussed in [1].

Environmental issues in regional economic studies have come to the fore since the early seventies, mainly because of environmental degradation from pollution of air, water and soil, which resulted in a growing concern among scientists and politicians about limited natural resources. The conflict between environmental degradation and economic development became a field of interest in regional-economic and environmental policy making. The fact that "pollution was widely recognized by economists and others as a serious threat to human well-being" [2, p. 3] has also led to the development of environmental sciences.

Environmental issues are typically multidimensional because the phenomena under study emerge from various disciplines (economics, ecology, demography and natural science, etc.) and may even cross the boundary of these single disciplines. Consider for example the assessment of urban quality of life, which involves a multidimensional profile with elements representing the quality and location of dwellings, the availability of natural parks and recreational facilities, and the quality of the transportation network and educational and cultural facilities. The interdisciplinary needs come to the fore in case of the analysis of, for example, the health effects of environmental degradation from air pollution. One of the first attempts to include environmental issues in the framework of regional modeling was the input-output approach [3]. However, the principal disadvantage of using input-output models is that they suffer from severe constraints such as 1) the static equilibrium assumption, 2) the huge effort required to collect all relevant information, and 3) the lack of opportunity to cover intraregional capabilities [4].

One of the major flaws in operationalizing environmental-economic phenomena, mentioned in Brouwer [5], concerns the availability of appropriate information. His study presents a state-of-the-art survey of environmental-economic models pertaining to the classical environmental compartments air, water and land. The major concerns about the availability of information include 1) the poor quality of information (or even the lack of information) needed to operationalize processes in an appropriate way, a characteristic which holds especially for ecological processes, and 2) the various data sources and the corresponding different levels of precision of variables within the framework of modeling environmental-economic phenomena. However, in recent years many research efforts have led to improvements with respect to information and data availability, and many sciences have contributed to the rapidly developing field of research on the quality of information retrieval and processing, viz. geography, regional economics, sociology, psychology and biology. A state-of-the-art overview of methods which deal with various kinds of limitations with respect to the available information and with "qualitative" statistics and modeling is presented [6]. An assessment of environmental-economic phenomena at a regional level will therefore benefit from the methodological progress achieved so far in various related disciplines.

In this article we discuss recent developments regarding methodological tools which are vital for modeling environmental-economic phenomena at the regional level. The limitations of data availability for modeling this kind of phenomena will be discussed in Section 2. This investigation considers 1) the insufficient precision of information, which precludes drawing reliable quantitative conclusions concerning policy aspects, 2) the different data sources of information, which imply that the information may be measured at different scales, with a distinction between a metric (or cardinal) scale and a non-metric scale (either nominal or ordinal), and 3) the multivariate nature of information. The methodological tools that are useful in such cases will be discussed in Section 3. Special attention will be devoted to graph theory, qualitative calculus, generalized linear modeling (GLM), log-linear and linear-logit modeling, and multidimensional scaling (MDS). By using such appropriate tools in the case of imprecise data, human behavior regarding environmental commodities, for instance in the field of recreation can be properly analyzed. Recreational activities in natural regions (such as in rivers and streams) exhibit a mutual dependence between ecological processes and economic activities at a regional scale. The different tools discussed in Section 3 will be elaborated and applied in Section 4 in the framework of an analysis of outdoor recreation in the Biesbosch area in The Netherlands.

## **2. AVAILABILITY OF INFORMATION FOR MODELING ENVIRONMENTAL-ECONOMIC PHENOMENA**

A wide variety of models has been developed over the past ten years for the analysis of environmental policy issues such as land use planning and water management planning [5, 7]. Most models differ in their design and area of application and they may incorporate land, air and water (see [8-10] for an exposition on modeling efforts to integrate economic activities and environmental processes).

Three types of limitations may exist with respect to the availability of information, each of them having specific methodological requirements to be taken into account:

1. First, the available information in modeling environmental-economic phenomena may be measured with insufficient precision for drawing reliable quantitative conclusions. For instance, in a study on water management for The Netherlands which was prepared to determine the spatial distribution of water as a result of supply and demand factors (covering also the economic and environmental effects on agriculture and shipping in case of a shortage of water), the environmental aspects of water management were clearly included, but they were simply dealt with by imposing global water quality standards because "it is simply too

difficult to treat the environment more directly” [11, p. 157], due to lack of reliable information. Absence of sufficient reliable information is mainly the result of the *complexity* in modeling environmental-economic phenomena and the *uncertainty* with respect to the nature of cause-effect relationships, and the reliability of the required information.

Fortunately, in the past decade various methodological advances have been made in analyzing phenomena when only limited or imprecise information is available (for instance, when only binary or qualitative information exists with respect to the causal model structure). A binary relationship between variables indicates whether or not a certain variable has an impact on another variable irrespective of its metric value, and a qualitative relationship indicates that, in addition to the binary information about the directional relationship between two variables, information concerning the sign of the relationship is also available. Research efforts in the field of binary and qualitative information treatment in the social sciences are needed since 1) “ordinarily, the economist is not in the position of having exact quantitative knowledge of the partial derivatives of his equilibrium conditions” [12, p. 26], because of the limited amount of suitable empirical data, and 2) difficulties may arise in practice in obtaining precise information, because of measurement problems or because of pragmatic reasons (such as time and money constraints in collecting the relevant data) [6]. The use of binary or qualitative approaches may provide tools for dealing with the kinds of problems mentioned, as will be elaborated in Section 3.1.

2. The research tradition in modeling regional-economic, environmental and geographic phenomena followed the standard practice in natural sciences of dealing with information as if it were obtained under well defined conditions from controlled experiments. Such a research tradition, however, often ignored the realities of regional-economic, social and environmental research (which is frequently based on data obtained from various unconventional sources such as survey-sampling, panel studies or expert judgments) and did not take into account that “measurement of properties of human behavior poses many problems such as instability, bias by reactivity, inherent discreteness and sometimes nominal or ordinal measurement level” [13, p. 171]. The consequence of this was that “for too long social scientists had to manage with methods designed for use in the natural sciences in which the variables are well defined and readily capable of measurement. In the last decade or two the balance has been somewhat restored by the development of such things as the log-linear model” [14, p. 27].

Means of information gathering in analyzing environmental-economic phenomena may vary from controlled experiments under properly defined conditions in natural sciences to methods of survey-sampling to assess properties

of human behavior in the social sciences. The corresponding measurements vary from a cardinal (or metric) scale to a nominal scale [15]:

- (i) the nominal scale considers a set of objects, attributes or properties from variables, which have been classified into distinct groups, without restrictions on their numerical representations;
- (ii) the ordinal scale meets the requirements of the nominal scale with an additional requirement that a logical ordering of magnitude exists concerning the events measured. Such observations can be ranked with a hierarchical nature from “low” to “high” with numbers (1, 2, 3, ...). However, the difference between two ordinal figures has no numerical meaning;
- (iii) the interval scale meets the requirements of the ordinal scale, and also allows the determination of relative Euclidean distances between figures. However, the figures themselves have no absolute meaning;
- (iv) the ratio (or cardinal) scale has an absolute numerical interpretation, and its values can be represented in a metric system.

Various statistical models with either qualitative or quantitative information have been developed over the past decade, and these methodological advances will be discussed in Section 3.2 in the framework of modeling environmental-economic phenomena at a regional level.

3. Thirdly, a simultaneous analysis of economic and environmental phenomena results in variables that are multivariate in nature, because the components to be included (such as regions), can be characterized by various interrelated items. A region may for example be characterized by the existence of a series of infrastructure items such as transportation, energy supply, water management, recreation and nature. The major purpose in such cases is to identify a spatial structure in a set of multivariate observations concerning the various items. Multidimensional procedures are then employed to assign numbers to each of these items and to extract cardinal information from qualitative data. The methodological tools on the analysis of multivariate information that will be discussed in Section 3.3 have originally been developed in psychometrics, but they have also shown their relevance in other areas of research over the past years, including environmental-economic applications (see [16] for a multivariate analysis of qualitative relations on natural resources).

### **3. APPROPRIATE METHODOLOGICAL TOOLS FOR ENVIRONMENTAL-ECONOMIC PHENOMENA**

The aim of this section is to review the major methodological progress that has been achieved over the past decade, and to discuss the current uses and future potential of these tools in modeling environmental-economic phenomena.

### 3.1. Qualitative Information in Modeling Environmental-Economic Phenomena

In this subsection various methodological tools will be discussed for modeling environmental-economic phenomena when only a limited amount of information is available concerning the impacts between variables, i.e. when impacts are either binary (or zero/one) or qualitative in nature (and depicted in terms of positive or negative signs). A zero impact denotes the absence of a prior theoretical relationship between a pair of variables.

A major methodological advance can be found in the field of structure analysis of models when only binary information or cause-effect relationships is available, which provides information on the structure of stimulus-response relationships, such as 1) the level of interwovenness of a system and its subsystems, 2) the key driving forces, and 3) the hierarchical ordering of phenomena. Graph theory appeared to be particularly important here. Besides, in recent years the analysis of qualitative relations has become an important tool in impact analysis for policy modeling, mainly because of the unsatisfactory data base for policy impact assessments. This problem became particularly urgent in complex multidisciplinary systems, notably in modeling environmental-economic phenomena, and resulted into a renewed interest in qualitative calculus after some first developments in the area of a qualitative analysis of comparative statics by Samuelson in the 1940s. Both graph theory and qualitative calculus will briefly be described here.

Graph theory serves to provide relevant tools to identify the structure of a system of causal relationships represented in terms of zero/one information.

A graph  $G$  is identified by a set of vertices  $V = (v_1, \dots, v_M)$  and a set of edges  $E$ , with  $E = (e_1, \dots, e_N)$ . A so-called adjacency matrix  $A(G)$  with elements  $a_{ij}$ , which expresses the model structure, can be defined in terms of zero/one information by

$$a_{ij} = \begin{cases} 1 & \text{if there is an edge vertex } j \text{ into vertex } i (i, j = 1, \dots, M) \\ 0 & \text{if otherwise} \end{cases}$$

The causal relationships between variables—depicted in terms of a graph  $G$ , or in a similar way in terms of the adjacency matrix  $A(G)$  of order  $M \times M$ —can be subdivided into three groups, viz. 1) a direct causal link (called an edge) between vertices  $v_i$  and  $v_j$ , 2) an indirect causal link (called a path) from vertex  $v_i$  to vertex  $v_j$  in two or more steps, and 3) an interdependent causal link (called a cycle) from vertex  $v_i$  back to vertex  $v_i$  [17].

Graph theory is a proper tool for an analysis of the nature and magnitude of the causal structure among strongly linked phenomena. This is especially relevant in the assessment of the linkages between environmental and economic phenomena, because of their complex nature.

Next, we consider a set of linear equations in matrix form  $Ay + b = 0$ , where the matrix  $A$  of impact parameters only has qualitative information about the signs (positive, negative or zero) on the impacts of a stimulus (exogenous) variable on a response (endogenous) variable. The vector  $b$  of exogenous variables also contains qualitative information. Given the assumption that the matrix  $A$  of order  $M \times M$  is non-singular, the solution of this set of equations is then given in the reduced form by  $y = -A^{-1}b$ . Now the question is whether it is possible to predict the signs of  $y$ , given the information incorporated in  $A$  and  $b$ . This procedure is also called sign-solvability analysis. So, the analysis of sign-solvability focuses on the identification of changes in endogenous variables in a qualitative way due to changes in exogenous variables.

The conditions of sign-solvability of the set of equations have been formulated by Bassett et al. [18] and are based on the principles of graph theory. Necessary and sufficient conditions such that  $Ay + b = 0$  can be reformulated in its reduced form are:

1. the diagonal elements of matrix  $A$  are all negative, i.e.,  $a_{ii} < 0$  for all  $i = 1, \dots, M$ ;
2. all cycles of the matrix  $A$  with length at least two need to be non-positive;
3.  $b \leq 0$ , i.e., the elements of vector  $b$  need to be non-positive;
4. if some  $b_k < 0$ , then every path from vertex  $k$  to vertex  $i$  is non-positive for  $i \neq k$ .

These conditions are now discussed by using the following illustration of a set of four linear equations with qualitative information about the impacts between variables and the qualitative changes of the exogenous variables:

$$\begin{bmatrix} - & + & - & 0 \\ - & - & 0 & 0 \\ 0 & 0 & - & + \\ 0 & 0 & - & - \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} + \begin{bmatrix} 0 \\ - \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

All main diagonal elements of the matrix  $A$  are negative, and the first condition of sign-solvability holds. The second condition is based on the cycles of matrix  $A$ . The length of a cycle is defined as the number of terms from such an interdependent link. The graph related to matrix  $A$  has two cycles of length at least two, viz.  $y_1 \rightarrow y_2 \rightarrow y_1$  and  $y_3 \rightarrow y_4 \rightarrow y_3$ , which are both negative so that the second condition of sign-solvability also holds. The elements of vector  $b$  are non-positive, because vertex  $b_2$  has an outgoing edge which is negative in sign (condition three). The path  $b_2 \rightarrow y_2 \rightarrow y_1$  is negative in sign, and therefore the final condition of sign-solvability also holds. The qualitative solution of this set of equations, which can be interpreted as the changes from the equilibrium position then becomes:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} - & + & - & 0 \\ - & - & 0 & 0 \\ 0 & 0 & - & + \\ 0 & 0 & - & - \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ + \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} - & - & + & + \\ + & - & - & - \\ 0 & 0 & - & - \\ 0 & 0 & + & - \end{bmatrix} \begin{bmatrix} 0 \\ + \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} - \\ - \\ 0 \\ 0 \end{bmatrix}$$

The changes from the equilibrium position in the variables  $y_1$  and  $y_2$  become negative, while the other two variables do not change from their equilibrium position, given the model structure that is represented by matrix A, and the signs of the exogenous variables in the vector b.

A qualitative approach is a useful mathematical tool to understand the global functioning and coherence of a model [19], especially when the data are of poor quality or when they are not available at all. Then only qualitative conclusions concerning the impacts of policy instruments are reasonable and reliable. This qualitative approach is an appropriate complementary tool to conventional econometric estimation techniques and related simulation procedures.

The main developments in the field of sign-solvability analysis took place in mathematics (see [20], a symposium on Computer Assisted Analysis and Model Simplification, and [21], a review of recent developments in sign-solvability analysis). An example of qualitative approaches in urban economic planning is found in [22], where a dynamic simulation model for urban decline of the city of The Hague in The Netherlands is analyzed on the basis of qualitative information concerning impact parameters.

### 3.2. Nominal and Cardinal Information in Modeling Environmental-Economic Phenomena

The level of measurement of information of variables may vary from a nominal and ordinal scale (also called categorical variables) to an interval and ratio scale (also called continuous variables). Various statistical models concerning qualitative and quantitative information have been developed over the past two decades. A comprehensive statistical framework for this kind of information was established by Nelder and Wedderburn by means of a family of generalized linear models (GLMs). The main advantage of the GLM-development is that a link can be formalized between statistical models for categorical information and for information that is measured in a conventional quantitative way; a GLM can be expressed in the following way [23]:

$$y_i = \mu_i + \epsilon_i, i = 1, \dots, I, \tag{1}$$

where

- $y_i$  is a response variable that is considered to originate from the exponential family of probability distributions;
- $\mu_i$  is the expected value of  $y_i$ , that is,  $\mu_i = E(y_i)$ ;
- $\epsilon_i$  is a randomly distributed error term.



The explanatory variables  $X_{ik}$  which influence the variation in the response variable  $y_i$  can be summarized in the structure of the so-called “linear predictor”  $\eta_i$  which takes the form:

$$\eta_i = \sum_{k=1}^K \beta_k X_{ik}, i = 1, \dots, I, \tag{2}$$

where  $\beta_k$  are the parameters to be estimated. The linear predictor can then be related to the expected value of  $y_i$  by means of the so-called “link function”  $g$ :

$$\eta_i = g(\mu_i), i = 1, \dots, I. \tag{3}$$

By substituting the linear predictor and the inverse link function into (1), a generalized linear model can be written in the following form:

$$y_i = g^{-1}(\eta_i) + \epsilon_i. \tag{4}$$

Any particular member of the class of generalized linear models can then be derived from a combination of the three elements of a GLM, that is, from the 1) link function, 2) linear predictor, and 3) probability (“error”) distribution. The GLM approach therefore allows one to formally link statistical models based on information originating from various sources and measured at different scales.

A very powerful exploratory method for the analysis of the structural relationships among categorical variables is log-linear modeling, in which main effects and interaction effects serve to identify patterns of association between cross-classified variables. Contingency tables are appropriate ways to represent nominal information with two or more categorical variables cross-classified by means of two-way tables (with two categorical variables) and multi-dimensional tables (with three or more categorical variables). The most general log-linear model for the representation of information from two-way contingency tables in linear additive components is presented in equation (5). This model is linear in the (natural) logarithms of the expected cell-frequencies (with an  $I \times J$  table of  $I$  rows and  $J$  columns concerning the cross-classified variables  $A$  and  $B$ , with sample size  $N$ , and with  $n_{ij}$  representing the frequencies for element  $(i, j)$ ) and appears [24] to be equal to:

$$\log_e m_{ij} = \log_e E(n_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}; i = 1, \dots, I; j = 1, \dots, J. \tag{5}$$

All information about the structure of the contingency table, in terms of association and interaction effects, is contained in equation (5). The parameter  $\lambda$  is the grand-mean effect,  $\lambda_i^A$  and  $\lambda_j^B$  are the main effects of variables  $A$  (for the  $i$ -th category) and  $B$  (for the  $j$ -th category), and  $\lambda_{ij}^{AB}$  is the first-order interaction effect between variables  $A$  and  $B$ .

The number of parameters in (5) is equal to  $IJ + I + J + 1$  which is larger than the number of cell elements. Therefore, in order to identify the parameters of this model, in total  $I + J + 1$  constraints must be imposed; a widely used system of parameter constraints is the “cornered effect” system, with:

$$\lambda_1^A = \lambda_1^B = \lambda_{11}^{AB} = \lambda_{1j}^{AB} = 0; i = 1, \dots, I; j = 1, \dots, J. \quad (6)$$

By putting different parameters in (5) equal to zero, it is possible to specify a family of log-linear models. Each member of this family has a totally different interpretation, and is associated with a particular hypothesis about the nature of the structural relationships between the variables A and B in the two-dimensional contingency table. The models presented below in (7)-(10) define what is termed the hierarchical set of log-linear models for a two-dimensional contingency table, and this set has the property that higher-order parameters are only included if all related lower-order parameters are also included.

$$\log_e m_{ij} = \lambda + \lambda_i^A + \lambda_j^B, i = 1, \dots, I; j = 1, \dots, J. \quad (7)$$

$$\log_e m_{ij} = \lambda + \lambda_i^A, i = 1, \dots, I; j = 1, \dots, J. \quad (8)$$

$$\log_e m_{ij} = \lambda + \lambda_j^B, i = 1, \dots, I; j = 1, \dots, J. \quad (9)$$

$$\log_e m_{ij} = \lambda_j, i = 1, \dots, I; j = 1, \dots, J. \quad (10)$$

The linear model (5) is referred to as the saturated log-linear model for a two-dimensional contingency table, since it is linear in the logarithms of the expected cell-frequencies, and it has as many independent parameters as there are cells in the contingency table (in case the parameter constraints from (6) are included). Model (7) is a representation of the hypothesis of independence between variables A and B. Similarly, models (8) and (9) are representations of the hypotheses that the categories of B (or A) variables are equally probable, and finally model (10) is a representation of the hypothesis that all categories of both variables are equally probable.

The log-linear modeling approach, as described above in case of a two-dimensional table, includes the elimination, in a hierarchical fashion, of the parameters of the saturated model which are not essential to a description of the structural relationships between the variables in the table. The model is reduced to the most parsimonious form that is consistent with the relationships between the variables from the contingency table.

The real advantage of log-linear modeling can be seen most clearly in the case of multidimensional contingency tables, which include three or more cross-classified variables. In case of a three-dimensional table of order  $I \times J \times K$ , the saturated log-linear model, analogous to the two-dimensional representative in (5), can be represented as follows:

$$\log_e m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad (11)$$

A hierarchical family of log-linear models for this three-dimensional case can be specified by putting different parameters in the saturated model (11) equal to zero (see [25], which presents an overview of various estimation procedures appropriate for qualitative statistical models).

The log-linear models discussed above are most suitable for a description of structural relationships among classified variables. An analogue of the log-linear model is the linear-logit model where one of the variables is a response variable, and which can be written in the case of a dichotomous (or 2-category) response variable in the form:

$$\log_e \left[ \frac{p_i}{1 - p_i} \right] = \beta_0 + \sum_{k=1}^K \beta_k X_{ik}, \quad i = 1, \dots, N. \tag{12}$$

$p_i$  represents the probability that the first category will be selected by the  $i$ -th individual or at location  $i$ , given the values of the  $K$  explanatory variables. The left-hand side of the model in (12) is known as the logit transformation and its value increases from  $-\infty$  to  $+\infty$  as  $p_i$  increases from 0 to 1. The equivalence between log-linear models and linear logit models can be demonstrated, since a saturated log-linear model can be expressed in terms of a linear logit model [25].

The qualitative statistical models will be evaluated in terms of their goodness-of-fit statistics, which is also known as the deviance in case of the GLM approach. The deviance is based on the maximization of the log-likelihood of the fitted model, which is—in case of a log-linear model that is based on an  $I \times J \times K$  table—equal to

$$D = 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K n_{ijk} \left[ \log_e \frac{n_{ijk}}{\hat{m}_{ijk}} \right] \tag{13}$$

A model selection procedure for log-linear and linear logit models is essentially based upon the goodness-of-fit statistic described above. Especially in case of three or more dimensional tables, a systematic and efficient model selection procedure is required, because of the great many possible models (for example, a four dimensional contingency table has 167 hierarchical log-linear models). A simultaneous test procedure was developed by Aitkin according to the GLM framework [26, 27], which is currently known as the computer package GLIM (Generalized Linear Interactive Modeling) for all kinds of GLM-based models. This procedure is a statistically elegant one, since it includes a compensation for multivariate testing, so as to avoid that parameter significance will be attributed to what is merely random variation. Aitkin determined an overall type-one error rate  $\gamma$  for the null hypothesis that a set of parameters is simultaneously equal to zero [27]. Such a rate  $\gamma$  can be interpreted as the chance of at least one incorrect rejection of a true null hypothesis. A level of insignificance  $\alpha$  is recommended such that  $\gamma$  is in the range between 0.25 and 0.50.

Finally, it is noteworthy that the analysis of qualitative data in regional-economic and environmental studies appears to be closely linked to advances in statistical methodology; the present subsection provided only a brief introduction to this field (see [25] for a fuller analysis of the use of qualitative statistical models in regional economic studies). Altogether, the conclusion can

be drawn that log-linear models and linear-logit models are powerful statistical tools in analyzing the functional qualitative structure of and the interactions among economic, social and environmental phenomena, as will be shown in Section 4.

### 3.3. The Multivariate Nature of Environmental-Economic Phenomena

Multivariate phenomena can be represented by spatially disaggregated profiles, such as socio-economic profiles (e.g., traffic and transport, telecommunication, energy supply) and environmental profiles (e.g., attractiveness of nature). The spatial structure of such a multivariate phenomena can be described by a multivariate analysis, with a variate defined to be a quantity that may take any value within a specified set (such as categories or rank numbers). The appropriate scaling procedure for qualitative data that will be discussed here is multidimensional scaling (MDS).

The rationale behind the use of MDS is to transform ordinal information into cardinal units. It takes for granted that a set of  $N$  objects or indicators in a  $K$ -dimensional space is transformed into an  $S$ -dimensional Euclidean space, with  $S \leq K$ . The aspects of the phenomena concerned are represented in an  $S$ -dimensional space in such a way that it should be consistent with the observed ordinal rankings. The MDS-approach then uses the ensuing degrees of freedom to obtain a metric representation of the information that is originally ordinal in nature.

Suppose a symmetric  $N \times N$  paired comparison table  $\Delta$  with (dis)similarities between  $N$  objects being expressed by ordinal numbers. The elements  $\delta_{nm}$  of  $\Delta$  represent rank numbers for the (dis)similarity between points  $n$  and  $m$  ( $n, m = 1, \dots, N$ ). Each object  $n$  (either an indicator or a region) is characterized in an  $S$ -dimensional space with the coordinates  $x_{ns}$  ( $s = 1, \dots, S$ ). The Euclidean distance  $d_{nm}$  between each pair of points ( $x_{ns}, x_{ms}$ ) is then:

$$d_{nm} = \left[ \sum_{s=1}^S (x_{ns} - x_{ms})^2 \right]^{1/2}, \quad n, m = 1, \dots, N. \quad (14)$$

The distance metric from (14) is symmetric and the triangle inequality holds. Besides, it is invariant for translation and is homogeneous in nature.

The purpose of MDS is now to assess the coordinates  $x_{ns}$  and  $x_{ms}$  in such a way that the distance  $d_{nm}$  is in agreement with the original rankings  $\delta_{nm}$ , which means that such a distance has a maximum goodness-of-fit according to the original data. This condition means that a badness-of-fit or stress function has to be defined, which minimizes the residual variance between all distances  $d_{nn}$  and  $\delta_{nm}$ . However, as was already mentioned in Section 2, no arithmetic operations are allowed on the dissimilarity measure  $\delta_{nm}$ , since it is ordinal in nature. A metric auxiliary variable  $\hat{d}_{nm}$  is therefore calculated such that  $\hat{d}_{nm} \leq \hat{d}_{n1}$ ,

whenever  $\delta_{nm} \leq \delta_{nl}$  (which is called the condition of order-isomorphism). The goodness-of-fit statistic can then be specified in terms of the following simple Minkowski metric:

$$\Phi = - \left[ \frac{\sum_{n=1}^N \sum_{m=1}^N (d_{nm} - \hat{d}_{nm})^2}{\sum_{n=1}^N \sum_{m=1}^N d_{nm}^2} \right]^{1/2}, n \neq m, \tag{15}$$

in which the sum of squared distances representing the normalization condition is constant. The smaller the value of the loss function, the better the order-isomorphism (or monotonicity) between the original rankings as well as the interpoint distances defined in (14). A set of coordinates  $x_{nm}$  can then be estimated in an iterative procedure.

The MDS-approach has been used over the past decade for the analysis of various economic management and policy phenomena, e.g. [16] for environmental-economic profiles, [28] for the analysis of water quality in Japan, and [29] for an analysis of Dutch regional infrastructure profiles.

This multivariate approach for the analysis of ordinal information is relevant to represent the originally non-metric information in a quantitative configuration. The next section will be devoted to an empirical illustration of the relevance of the abovementioned approaches for regional-economic and environmental analysis.

#### 4. AN ENVIRONMENTAL-ECONOMIC ANALYSIS OF RECREATION

##### 4.1. Introduction

The case study in this section concerns the Biesbosch area. This is a region in the south-western part of The Netherlands with many attractive facilities for outdoor recreation related to rivers and streams (such as swimming, fishing, surfing and sailing). Such recreational activities in natural regions exhibit a mutual dependence between ecological processes and economic activities. This mutual dependence will be analyzed in the sequel of this section, by making use of the various methodological tools, reviewed in the previous section.

A survey was organized during the summer of 1983 concerning the spatial pattern of recreational activities in this area. Its objective was to explore the preferences and behavior of recreationers with respect to outdoor recreation (such as kind of activities), related to ecological characteristics (such as relevance of landscape characteristics) and socio-economic phenomena (such as the availability of recreational facilities). The total sample size conducted about 400 respondents, and the area was subdivided into five regions to enable a spatial

characterization of the recreational activities. In addition to this survey, a counting was organized during the same period at a more detailed spatial level (with in total 11 subregions), in order to obtain a deeper insight into the nature of recreational activities.

A regional environmental-economic analysis of recreational activities is presented in Section 4.2 on the basis of qualitative information. A more detailed analysis of the interaction between socio-economic activities and recreational patterns is presented in Section 4.3, where socio-economic (notably shopping) behavior of recreationers visiting the Biesbosch area is assessed based on a spatial disaggregated cross-classification of 1) the duration of stay in the area, and 2) the distance between the home address and the Biesbosch area. Finally, a multivariate analysis of recreational activities is presented in Section 4.4 which is based on a spatially disaggregated analysis of recreation (both its pattern and behavior) and natural environment.

#### **4.2. A Structure Analysis of Recreational Activities Based on Qualitative Information**

In the context of long-run regional-economic and environmental analysis for the area under consideration, a dynamic simulation model of recreational activities has been designed in order to assess qualitative impacts, owing to the scarcity of reliable information and of adequate experiments enabling parameter estimation in a conventional econometric way. The phenomena which are considered to be vital for recreational activities (including patterns and behavior) are regional-economic activities, demographic trends and environmental nature of the area. The relationships between the ten most important and appropriate variables for the analysis of recreational activities will be described below.

- RA: recreational attractiveness, which is a supply variable reflecting the availability of recreational facilities;
- P: population size;
- SA: stimulating regional incentives;
- EA: environmental attractiveness;
- REC: demand for recreational facilities;
- EMP: supply of labor;
- CR: expenditures on consumption goods by recreationers;
- CO: non-consumptive expenditures by recreationers;
- ER: demand for labor driven by the recreational activities in the Biesbosch area;
- EO: demand for labor not directly driven by the recreational activities (such as for example, agriculture).

The model, as a system of second-order difference equations, is represented in matrix forms as:

$$Ay_t = By_{t-1} + Cy_{t-2} + Dx_t, \tag{16}$$

with:

$$A = \begin{bmatrix} -1 & \alpha_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_1 & -1 & -\beta_2 & 0 & 0 & 0 & 0 & 0 \\ \gamma_1 & \gamma_2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon_1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & \eta_1 & 0 & \eta_2 & -1 & 0 & 0 \\ 0 & 0 & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & -1 & \lambda_5 \\ 0 & 0 & 0 & \mu_1 & 0 & 0 & \mu_2 & -1 \end{bmatrix}$$

$$B = \begin{bmatrix} -1 & 0 & \alpha_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -\beta_2 + \beta_3 & 0 & 0 & 0 & 0 & 0 \\ \gamma_1 & \gamma_2 + \gamma_3 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon_1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & \eta_1 & 0 & \eta_2 & -1 & 0 & 0 \\ 0 & 0 & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & -1 & \lambda_5 \\ 0 & 0 & 0 & \mu_1 & 0 & 0 & \mu_2 & -1 \end{bmatrix}$$

$$C = \begin{bmatrix} 0 & 0 & -\alpha_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\beta_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} -\alpha_3 & 0 \\ -\beta_4 & 0 \\ -\gamma_4 & -\gamma_5 \\ 0 & -\delta_1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -\lambda_6 & 0 \\ 0 & 0 \end{bmatrix}, \quad y_t = \begin{bmatrix} SA_t \\ EA_t \\ REC_t \\ EMP_t \\ CR_t \\ CO_t \\ ER_t \\ EO_t \end{bmatrix}, \quad x_t = \begin{bmatrix} RA_t \\ P_t \end{bmatrix}$$

Consider for example the first equation, where the rate of change in the development of recreational activities is stimulated by means of an increasing level of environmental attractiveness as well as by the recreational attractiveness, while it will be discouraged in case of an increase in demand for recreational activities.

As a first step of our analysis, we will investigate the sign-solvability of the simulation model presented in (16). In case the conditions for sign-solvability hold for this simulation model, the qualitative impacts of the exogenous variables and the lagged endogenous variables on the endogenous variables can then also be expressed. Sign-solvability of the model in (16) means that the vector  $y_t$  can be expressed in analytical terms as:

$$y_t = A^{-1} B y_{t-1} + A^{-1} C y_{t-2} + A^{-1} D x_t, \quad (17)$$

with  $y_t$  expressed in qualitative terms with positive, negative or zero entries, and based on the qualitative information from the stimulus-response relationships reflected in matrices  $A$ ,  $B$ ,  $C$ , and  $D$ . The conditions (1) and (2) for sign-solvability, which were described in Section 3.1, have to be fulfilled for the inversion of matrix  $A$ . The first condition is fulfilled since all diagonal elements of matrix  $A$  are negative. However, the second condition that the cycles of matrix  $A$  with length at least two should be negative is not fulfilled. The cycles  $SA \rightarrow ES \rightarrow SA$  and  $ER \rightarrow EO \rightarrow ER$ , which are of length two, are both positive, while the cycle  $EA \rightarrow SA \rightarrow REC \rightarrow EA$ , which is of length three, is negative. The inverse of matrix  $A$  is therefore not defined up to the signs of the cell-entries, and thus the vector  $y_t$  cannot be expressed in qualitative terms like the reduced form in (17). The set of stimulus-response relations are therefore not sign-solvable, and the quantitative impacts on policy instruments cannot be fully described in qualitative terms when only qualitative information is available.

However, the sign-solvability conditions can be further analyzed by decomposing matrix  $A$  into two parts, i.e., one matrix incorporating impact parameters between the variables  $SA$ ,  $EA$ ,  $REC$  and  $EMP$  (called  $y_{1,t}$ ), and another matrix based on the impact parameters of vector  $y_{2,t}$ , which includes the variables  $CR$ ,  $CO$ ,  $ER$  and  $EO$ . Vector  $y_{1,t}$  is independent from vector  $y_{2,t}$ , and the sign-solvability conditions of that vector can therefore be analyzed independently from the other elements. Next, it is noteworthy that a systems model like the one in (16), which was not entirely sign-solvable in a purely qualitative sense, may become at least partially sign-solvable when a mixture of qualitative and quantitative information is available. This would imply that certain parameters have to be given cardinal numerical values. Parameter experiments of the simulation model are described in [8]. The sign-solvability solution can then be determined when these selected parameter values of the simulation model are included. Then the qualitative solution of (16) when reformulated in terms of two parts of (16) becomes:

$$y_{1,t} = M_1 y_{1,t-1} + M_2 y_{1,t-2} + M_3 x_t, \quad (18)$$

and

$$y_{2,t} = K_1 y_{2,t-1} + K_2 y_{1,t-1} + K_3 y_{1,t-2} + K_4 x_t, \quad (19)$$

with:



$$\begin{aligned}
 M_1 &= \begin{bmatrix} + & + & - & 0 \\ + & + & - & 0 \\ + & - & + & 0 \\ 0 & 0 & 0 & + \end{bmatrix}, M_2 = \begin{bmatrix} 0 & - & + & 0 \\ 0 & - & + & 0 \\ 0 & + & + & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, M_3 = \begin{bmatrix} + & - \\ + & - \\ + & + \\ 0 & + \end{bmatrix}, \text{ and } x_t = \begin{bmatrix} RA_t \\ P_t \end{bmatrix} \\
 K_1 &= \begin{bmatrix} + & 0 & 0 & 0 \\ 0 & + & 0 & 0 \\ 0 & 0 & + & 0 \\ 0 & 0 & 0 & + \end{bmatrix}, K_2 = \begin{bmatrix} + & - & - & 0 \\ + & - & - & 0 \\ + & - & - & 0 \\ + & - & - & 0 \end{bmatrix}, K_3 = \begin{bmatrix} 0 & + & + & 0 \\ 0 & + & + & 0 \\ 0 & + & + & 0 \\ 0 & + & + & 0 \end{bmatrix}, \text{ and } K_4 = \begin{bmatrix} + & + \\ + & + \\ + & + \\ + & + \end{bmatrix}.
 \end{aligned}$$

The reduced form equations in (18) and (19) show that an increase in environmental attractiveness (EA) during period  $t-1$  will, with a lag of one period, result into an increase in the following variables: EA, REC, CR, CO, ER and EO. Similarly, an increase in the size of population (P) will lead to a decrease in the variables SA and EA, and to an increase in all remaining variables. Thus on the basis of limited quantitative information, the remaining qualitative part of a model may become sign-solvable.

Table 1 shows the results of a sequential introduction of quantitative information, and it shows whether or not the matrices  $K_1$  to  $K_4$  are determined in a qualitative sense.

An originally non sign-solvable set of equations (such as in step 1 from Table 1) may thus become sign-solvable when additional information is included with respect to the parameters of ER and EO. It also shows that the use of more quantitative information does not necessarily improve sign-solvability. This can be seen by comparison of step 1 with steps 2-7. Qualitative information on the signs of the equations concerning EMP, CR and CO is sufficient for sign-solvability as reflected by step (8) in Table 1. A mixture of qualitative and a sequential introduction of quantitative information is very helpful to elaborate sign-solvability of a simulation model.

Table 1. Sign-solvability and Sequential Introduction of Quantitative Information

Step	Information	$K_1$	$K_2$	$K_3$	$K_4$
(1)	SA, EA, REC	known	unknown	unknown	unknown
(2)	(1) + REC	known	unknown	unknown	unknown
(3)	(1) + CR	known	unknown	unknown	unknown
(4)	(1) + CO	known	unknown	unknown	unknown
(5)	(3) + CO	known	unknown	unknown	unknown
(6)	(1) + ER	known	unknown	unknown	unknown
(7)	(1) + EO	known	unknown	unknown	unknown
(8)	(6) + EO	known	known	known	known

### 4.3. An Analysis of Recreational Behavior with Nominal Information

It was already mentioned before that the assessment of the mutual dependence between environmental and economic phenomena includes information at different levels of measurement. An assessment of recreational activities with non-metric (e.g., nominal) information will be presented in this section, and this assessment is based on some behavioral characteristics of recreationers.

Table 2 shows a four-way contingency table of size  $5 \times 2 \times 3 \times 2$  to analyze the interaction-effects among the response variable, i.e., the place where daily purchases are being made, to be classified with three explanatory variables, i.e., 1) the area of the Biesbosch region (i.e., variable A with 5 regions), 2) the distance to the home address (variable B with 2 classes: B = 1 when the distance is less than 30 kilometers, and B = 2 when the distance is over 30 kilometers), and 3) the duration of stay in the Biesbosch region (variable C with 3 classes: C = 1 when staying for one day; C = 2 when staying for less than three days; C = 3 when staying for at least four days).

A linear-logit model will be assessed with variable X from Table 2 as the response variable; this variable is also categorical in nature, with two classes, i.e., X = 1 when the daily purchases are bought in the Biesbosch region, and X = 2 when otherwise.

Table 3 shows the goodness-of-fit statistics of six sets of linear logit models, each of them starting from the model including all main effects, and adding additional interaction effects each of them starting from the model including all main effects, and stepwise adding additional interaction effects.

The linear logit model with all main effects included, which is analogous to Equation (12), becomes:

$$\log_e \left[ \frac{P_{ijk}}{1 - P_{ijk}} \right] = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C, \quad i = 1, \dots, 5; j = 1, 2; k = 1, 2, 3. \quad (20)$$

Table 2. Four Dimensional Contingency Table of Order  $5 \times 2 \times 3 \times 2$

	X = 1						X = 2					
	C=1		C=2		C=3		C=1		C=2		C=3	
	B=1	B=2	B=1	B=2	B=1	B=2	B=1	B=2	B=1	B=2	B=1	B=2
A=1	6	1	39	4	20	12	1	7	6	10	3	2
A=2	1	0	7	1	1	1	1	1	0	2	0	0
A=3	9	0	34	7	12	11	12	4	10	21	8	5
A=4	13	0	34	2	6	2	0	1	5	2	0	0
A=5	1	0	7	3	5	2	0	4	2	1	0	0
Total	30	1	121	17	44	28	14	17	23	36	11	7

Table 3. Stepwise Selection Applied to the 5x2x3x2 Contingency Table from Table 2

<i>Model</i>	$G^2$	<i>d.f.</i>	<i>Parameter</i>	<i>Deviance</i>	<i>d.f.</i>
A+B+C	40.02	22			
A+B+C+AxB	37.93	18	AxB	2.09	4
A+B+C+AxB+AxC	24.30	10	AxC	13.63	8
A+B+C+AxB+AxC+BxC	3.68	8	BxC	20.62	2
A+B+C+AxB+AxC+BxC+AxBxC	0.00	0	AxBxC	3.68	8
<hr/>					
A+B+C	40.02	22			
A+B+C+AxB	37.93	18	AxB	2.09	4
A+B+C+AxB+BxC	18.76	16	BxC	19.17	2
A+B+C+AxB+BxC+AxC	3.68	8	AxC	15.09	8
A+B+C+AxB+BxC+AxC+AxBxC	0.00	0	AxBxC	3.68	8
<hr/>					
A+B+C	40.02	22			
A+B+C+AxC	29.12	14	AxC	10.90	8
A+B+C+AxC+AxB	24.30	10	AxB	4.82	4
A+B+C+AxC+AxB+BxC	3.68	8	BxC	20.62	8
A+B+C+AxC+AxB+BxC+AxBxC	0.00	0	AxBxC		8
<hr/>					
A+B+C	40.02	22			
A+B+C+AxC	29.12	14	AxC	10.90	8
A+B+C+AxC+BxC	7.18	12	BxC	21.94	2
A+B+C+AxC+BxC+AxB	3.67	8	AxB	3.50	4
A+B+C+AxC+BxC+AxB+AxBxC	0.00	0	AxBxC	3.67	8
<hr/>					
A+B+C	40.02	22			
A+B+C+BxC	20.20	20	BxC	19.82	2
A+B+C+BxC+AxB	18.76	16	AxB	1.44	4
A+B+C+BxC+AxB+AxC	3.67	8	AxC	15.09	8
A+B+C+BxC+AxB+AxC+AxBxC	0.00	0	AxBxC	3.67	8
<hr/>					
A+B+C	40.02	22			
A+B+C+BxC	20.20	20	BxC	19.82	2
A+B+C+BxC+AxC	7.18	12	AxC	13.02	8
A+B+C+BxC+AxC+AxB	3.67	8	AxB	3.50	4
A+B+C+BxC+AxC+AxB+AxBxC	0.00	0	AxBxC	3.67	8

The advantage of making use of a simultaneous test procedure was briefly indicated in Section 3.2 to enable a statistical elegant selection of the most appropriate model from the results presented in Table 3. Consider the linear logit model in Equation (20), with the null hypothesis that all three first-order interaction effects and the one second-order interaction effect are zero. When a type one error rate  $\alpha$  of 0.10 is established, the overall type one error rate  $\gamma$  for that model with all main effects included is given by  $\gamma = 1 - (1 - \alpha)^4 = .34$ . This means that the probability of at least one incorrect rejection of the null

Table 4. Parameter Assessments of the Logit Model Fitted to Table 2

<i>Parameter</i>	<i>Estimate</i>	<i>Standard Error</i>
$\lambda$	1.11	0.43
$\lambda_2^A$	0.21	0.75
$\lambda_3^A$	-0.98	0.33
$\lambda_4^A$	0.67	0.50
$\lambda_5^A$	0.56	0.67
$\lambda_2^B$	-4.02	1.12
$\lambda_2^C$	0.77	0.42
$\lambda_3^C$	0.61	0.49
$\lambda_{22}^{BC}$	1.74	1.17
$\lambda_{23}^{BC}$	4.16	1.25

hypothesis (i.e., all first- and second-order interaction effects zero) is 34%. The overall type one error rate  $\gamma$  of 0.34 is within the range of (.25; .50), as recommended by Aitkin.

With the level of significance  $\alpha$  equal to 0.10, the critical values of the pooled effects are the following:  $\gamma_2 = 1 - (1 - \alpha)^3 = 1 - .73 = .27$ , and  $\gamma_3 = 1 - (1 - \alpha) = .10$ , where  $\gamma_3$  is the level of significance to test the second-order interaction effect, and  $\gamma_2$  is the level of significance to test the first-order interaction effects.

The deviance value of the second-order interaction effect  $A \times B \times C$  is 3.67, which is not significant, because the critical chi-squared value with 8 degrees of freedom and a 0.10 level of significance, is equal to 13.4. However, the significance level of the pooled first-order and second-order interaction effects may be determined by  $\gamma_{2,3} = 1 - (1 - \gamma_2)(1 - \gamma_3) = 0.34$ .

The chi-square critical value with 22 degrees of freedom, and a .34 level of significance equals 24.35, and the pooled first-order and second-order interaction effects are significant because the  $G^2$  value equals to 40.02. The purpose of the selection procedure is to eliminate those terms of the saturated model until the critical chi-squared value (which is 24.35 in this case) is passed. The model selected by the simultaneous selection procedure is  $A + B + C + B \times C$ , with a goodness-of-fit statistic equal to 20.20 and 20 degrees of freedom. Table 4 shows the parameter values of the logit model selected; they are obtained by making use of the GLIM-package.

The main effect parameters that are related to variable **A** show the spatial diversity concerning the place where the daily purchases will be bought. The recreationers visiting area 4 show a strong tendency to buy their daily purchases in the Biesbosch area, while the opposite tendency was found for those visiting area 3. The main effect parameter based on variable **B** reflects the importance of the distance categories with respect to whether or not the daily purchases are bought in the Biesbosch region. The negative value of this

parameter in Table 4 strongly indicates that the recreationers who live more than 30 kilometers from the Biesbosch area buy their daily purchases outside the Biesbosch. The first-order interaction effects between variables B and C, which is included to the base model from equation (17), reflect the diversity with respect to 1) the distance to the home address, and 2) the number of days of staying in the Biesbosch area. The estimated interaction effects in Table 4 show that the recreationers who live more than 30 kilometers from the home address buy their daily purchases in the Biesbosch area in case they stay for a longer period.

The main effects and the first-order interaction effects between variables B and C are vital components in describing the variation in shopping behavior of recreationers. Other components such as the spatial variation with respect to the duration of stay, or with respect to the distance factor, show statistically insignificant improvements of the model on shopping behavior of recreationers.

#### 4.4. A Multivariate Analysis of Recreational Activities

Environmental features (for example, the character of landscape, or the availability of recreational facilities) are major determinants of the actual use of an area for outdoor recreation. Ten environmental characteristics of recreation in the Biesbosch region will be analyzed simultaneously in this section. This MDS multivariate analysis employs the following variables:

- Variable 1: type of boat: motorized boats;
- Variable 2: type of boat: unmotorized boats;
- Variable 3: type of bank: willow-trees and holm;
- Variable 4: type of bank: reed vegetation and holm;
- Variable 5: type of bank: landing-stages;
- Variable 6: type of bank: beaches;
- Variable 7: type of activity: from the boat, such as fishing;
- Variable 8: type of activity: in water, such as swimming;
- Variable 9: type of activity: "quiet," on the beach (e.g., walking); and
- Variable 10: type of activity: "unquiet," on the beach (e.g., camp out).

The MDS-procedure has been applied here to four regions in the Biesbosch area, i.e.,

- |   |                              |
|---|------------------------------|
| 1. western part of the Biesbosch;       | 3. Heenplaat area; and       |
| 2. south-western part of the Biesbosch; | 4. Dordrecht Biesbosch area. |

Table 5 shows the two-dimensional metric representation of the ten variables. The spatial pattern of the variables can be displayed on a map, with similarly trending variables clustered together.

The *Western part* of the Biesbosch area shows two clusters, i.e., one including variables 2, 6, 7, and 8, and the other including variables 3, 4, and 9. The *South-western part* of the Biesbosch area also has two clusters (variables 1, 3, 4, 7, and

Table 5. Two Dimensional Metric Representation of the 10 Variables by Means of MDS

	Area = 1		Area = 2		Area = 3		Area = 4	
v1	-.26	-.17	-.04	.02	.11	-.26	.43	.02
v2	-.03	.10	.23	-.14	-.33	-.10	-.76	-.02
v3	-.65	-1.00	-.22	.12	.19	.35	.43	.03
v4	-.48	-.54	-.17	.09	.10	.65	-.13	-.01
v5	.46	-.44	.22	-.16	.49	-.04	-.27	-1.00
v6	-.01	.40	.28	.01	-.35	-.09	-.23	-.76
v7	-.15	.05	-.19	.10	.15	.34	.61	-.01
v8	-.06	.21	-.13	.07	.10	-.37	-.51	-.01
v9	-.72	-.30	.23	-.15	.19	-.28	-.38	-.41
v10	.63	.15	.23	-.16	-.52	-.01	.01	-.19

8, and variables 2, 5, 9, and 10). Recreational activities for *Heenplaat* are characterized by two clusters (1, 3, 4, 7, and 8, and 2, 5, 9, and 10). Finally, the *Dordrecht Biesbosch* has two clusters (1, 3, and 7, and 2, 6, 8, and 9).

The spatial characterization of the areas, which is based on the investigation of recreational behavior in different parts of the Biesbosch area, show a wide range of patterns on recreational behavior and related items.

## 5. CONCLUDING REMARKS

A number of methodological problems in modeling environmental-economic phenomena were identified in this article:

1. variables may be measured on a non-metric scale so that conventional statistical procedures based on assumptions such as the presence of a normal distribution are inadequate;
2. insufficient and unreliable data may constrain the conventional econometric ways of estimating parameters; and
3. phenomena may be multidimensional in nature, requiring multivariate analysis for proper investigation.

This article has shown that methodological progress has been achieved in various ways over the past decade and that several new tools are available to deal with environmental-economic phenomena.

## REFERENCES

1. P. Nijkamp, P. Rietveld, and F. Snickers, The Use of Regional and Multiregional Economic Models, in *Handbook of Regional and Urban Economics*, Volume 1: Regional Economics, P. Nijkamp (ed.), North Holland Publishing Company, Amsterdam, pp. 257-294, 1986.

2. A. C. Fisher and F. M. Peterson, The Environment in Economics: A Survey, *Journal of Economic Literature*, 14, pp. 1-33, 1976.
3. W. Isard, On the Linkage of Socioeconomic and Ecologic Systems, *Papers of the Regional Science Association*, 21, pp. 79-99, 1968.
4. B. D. Solomon and B. M. Rubin, Environmental Linkages in Regional Econometric Models: An Analysis of Coal Development in Western Kentucky, *Land Economics*, 61:1, pp. 43-57, 1985.
5. F. Brouwer, *Integrated Environmental Modelling: Design and Tools*, Martinus Nijhoff, Dordrecht/Boston, 1987.
6. P. Nijkamp, H. Leitner, and N. Wrigley (eds.), *Measuring the Unmeasurable*, Martinus Nijhoff, Dordrecht/Boston, 1985.
7. L. C. Braat and W. F. J. van Lierop (eds.), *Economic-Ecological Modeling*, Studies in Regional Science and Urban Economics, Volume 16, North-Holland, Amsterdam, 1987.
8. F. Brouwer and P. Nijkamp, A Satellite Design for Integrated Regional Environmental Modelling, *Ecological Modelling*, 35, pp. 137-148, 1987.
9. K. Fedra, Environmental Modeling under Uncertainty: Monte Carlo Simulation, Research Report RR-83-28, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1983.
10. A-M. Jansson (ed.), *Integration of Economy and Ecology: An Outlook for the Eighties*, Proceedings from the Wallenberg Symposia, Stockholm, 1984.
11. B. F. Goeller, Policy Analysis of Water Management for The Netherlands, Volume 1: Summary Report, Rand Corporation, Santa Monica, 1983.
12. P. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, 1947.
13. I. W. Molenaar, Statistics in the Social and Behavioral Sciences, *Statistica Neerlandica*, 39:2, pp. 169-179, 1985.
14. D. J. Bartholomew, Discussion of the Paper: Regression Models for Ordinal Data by P. McCullagh, *Journal of the Royal Statistical Society, Series B*, 42, pp. 127-129, 1980.
15. I. Adelman and C. T. Morris, The Derivation of Cardinal Scales from Ordinal Data: An Application of Multidimensional Scaling to Measure Levels of National Development, in *Economic Development and Planning*, W. Sellekaerts (ed.), Macmillan, London, pp. 1-39, 1974.
16. H. Blommestein and P. Nijkamp, Multivariate Methods for Soft Data in Development Planning, in *Spatial, Environmental and Resource Planning in Developing Countries*, M. Chatterji, P. Nijkamp, T. R. Lakshmanan and C. R. Pathak (eds.), Gower, Aldershot, pp. 343-360, 1984.
17. K. J. Tinkler, Introduction to Graph Theoretical Methods in Geography, *Series: Concepts and Techniques in Modern Geography*, 14, Institute of British Geographers, London, 1977.
18. L. Bassett, J. S. Maybee, and J. Quirk, Qualitative Economics and the Scope of the Correspondence Principle, *Econometrica*, 36, pp. 544-563, 1968.
19. D. Royer and G. Ritschard, Qualitative Structural Analysis: Game or Science?, in *Analysing the Structure of Econometric Models*, Series: Advanced Studies in Theoretical and Applied Econometrics, Volume 2, J. P. Ancot (ed.), Martinus Nijhoff, The Hague, pp. 3-13, 1984.

20. H. J. Greenberg and J. S. Maybee (eds.), *Computer Assisted Analysis and Model Simplification*, Academic Press, New York, 1984.
21. J. S. Maybee and H. Voogd, Qualitative Impact Analysis through Sign-Sign-Solvability: A Review, *Environment and Planning B: Planning and Design*, 11, pp. 365-376, 1984.
22. F. Brouwer and P. Nijkamp, Mixed Qualitative Calculus as a Tool in Policy Modeling: A Dynamic Simulation Model of Urban Decline, *Journal of Policy Modeling*, 8:1, pp. 69-88, 1986.
23. J. A. Nelder, Statistical Models for Qualitative Data, in *Measuring the Unmeasurable*, P. Nijkamp, H. Leitner, and N. Wrigley (eds.), Martinus Nijhoff, Dordrecht, pp. 31-38, 1985.
24. Y. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Massachusetts, 1975.
25. N. Wrigley and F. Brouwer, Qualitative Statistical Models for Regional Economic Analysis, in *Handbook of Regional and Urban Economics*, Volume 1: Regional Economics, P. Nijkamp (ed.), North Holland Publishing Company, Amsterdam, pp. 443-490, 1986.
26. M. Aitkin, A Simultaneous Test Procedure for Contingency Table Models, *Applied Statistics*, 28, pp. 233-242, 1979.
27. ———, A Note on the Selection of Log-Linear Models, *Biometrics*, 36, pp. 173-178, 1980.
28. S. Ikeda and H. Itakura, Multidimensional Scaling Approach to Clustering Multivariate Data for Water Quality Modeling, in *Uncertainty and Forecasting of Water Quality*, M. B. Beck and G. van Straten (eds.), Springer Verlag, Berlin, pp. 205-223, 1983.
29. P. Nijkamp and H. Voogd, Multidimensional and Homogeneous Scaling in Spatial Analysis, in *Recent Developments in Spatial Data Analysis: Methodology, Methods, Models*, G. Bahrenberg, M. M. Fischer, and P. Nijkamp (eds.), Gower, Aldershot, pp. 201-214, 1984.

Direct reprint requests to:

Floor Brouwer  
Agricultural Economics Research Institute  
P.O. Box 29703  
NL-2502 LS The Hague  
The Netherlands