

## Management of forest vegetation data series: the role of database in the frame of Quality Assurance procedure

Roberto CANULLO\*, Giandiego CAMPETELLA, Maria-Cristina ALLEGRINI and Vincenzo SMARGIASSI<sup>1)</sup>

Department of Botany and Ecology, Laboratory of Plant Population Ecology, University of Camerino, Via Pontoni 5, 62032 Camerino (MC), Italy

<sup>1)</sup>New Information Services, Via Tiburtina 1110, 00100 Roma, Italy

\*e-mail corresponding author: roberto.canullo@unicam.it

---

### ABSTRACT

If data from diachronic records on permanent areas are to be made available, the quality of the historic sequences must be standardised, preserved, organised and checked in such a way as to permit continuous input and comparison. The "Ground Vegetation Assessment" group of the CONECOFOR programme designed a database with extended search capability to ensure rapid and precise access to data. The vegetation is analysed within a network of permanent plots, based on field surveys conducted at community and population level. Assessments include specific, stratified and overall cover estimates as well as detailed cover scores and density of aboveground shoots (respectively on 24 100 m<sup>2</sup> and 100 0.25 m<sup>2</sup> sampling units). In addition to archiving data, the database runs functions to check their validity. The integrity of the dataset and its conformation to the user defined range can be assessed, and the entire sequence can be validated before the new data is saved in the database. Subsequent cross-checks among attributes allow further tests of validity and precision. These functions are an integral part of the overall Quality Assurance Control system. The data are organised into seasonal surveys, plots and sampling units. Each species has a field code, with reference to a second archive of coded nomenclature established at a European level. A section for addition and deletion of data makes output available according to the appropriate EC regulations. The system guarantees the visualisation of a certain number of simple statistics, and also permits export of analytic data to external statistical tools.

*Key words:* Ground Vegetation Assessment, CONECOFOR, Permanent Plots, Data Validation, long-term surveys

---

### 1. INTRODUCTION

A systematic programme of long-term ecological study involves a number of procedural problems, such as those of standardisation and data management. Developing some standard experimental designs with standardised protocols and utilising the best current technology and techniques in measurement programmes is crucial (Franklin 1989; Corona *et al.* 2002). Long-term monitoring studies and, in particular, diachronic records on permanent areas are conditioned by the overwhelming body of data that can be stored (Pickett 1991), since the historical sequences have to be kept and their quality checked, and the data have to be stored and organised using proper standards. Data must also be available for continuous use and comparison. Data accumulation is always a two-sided affair: its positive aspects are greater knowledge and its storage, but this entails the risk of increased entropy, especially for complex natural objects.

Statistical packages are not the best solution for problems related to data entry and management, due to a number of limitations (Elzinga *et al.* 2001). The information must be organised in a kind of database which allows rapid and easy access to data, precise queries and availability of the data in various formats and for different purposes. In addition, considering the increasing need to make this data available to national and regional monitoring programmes (Risser *et al.* 1991; Christian *et al.* 1999), such a database should be capable of pro-

viding access to information which may be used for studies in other sectors and by other research groups.

Data quality assurance is another very important factor (Franklin 1989). In fact, reliability and accuracy are essential to the interpretation and use of long-term data sets (Strayer *et al.* 1986; Schniederjans & Karuppan 1995; Walterbeek *et al.* 1996). Documentation of initial conditions and measurement programmes has been a chronic problem (Michener 1986; Hellawell 1991). Data need to be assessed and archived on a systematic basis involving uniform procedures.

Collecting data and stuffing them into electronic closets is clearly inadequate: good practice demands that data should be periodically analysed and reported.

All these considerations are of importance in ensuring data comparability, which involves statistical and non-statistical aspects.

In the context of the ICP-Forests Intensive Monitoring Programme (Reg. EC 1091/94), the National Focal Centre of the Italian Network (CONECOFOR programme) utilises, as a coordinating strategy, an interdisciplinary programme of integrated and combined evaluation of the data emerging from the intensive monitoring of forest ecosystems (Ferretti *et al.* 2000; Allavena *et al.* 2000). In this framework, the "Ground Vegetation Assessment" group, led by the Department of Botany and Ecology of Camerino University, is implementing an internal Quality Assurance (QA) system, which is vital in addressing the challenges of data reli-

ability and, in particular, the non-statistical aspects which can affect comparability (i.e. bias due to the observer's uncertainty and to the errors generated during the step-wise management of data) in long-term observations. A database has been designed to meet the requirements described above, and in addition, to be rapid enough to deal efficiently with new problems or improvements and additions as the need may arise.

This work describes the ways in which a Quality Assurance procedure contributes to the reliability of stored data, and seeks to show how a database, thanks to specific dedicated functions, can be an essential part of that procedure.

## 2. MATERIALS AND METHODS

### 2.1. Data collection

Vegetation analysis in the CONECOFOR network provides a series of parameters, among which the specific coverage assessments are considered a priority (Dupouey 1998).

The vegetation analysis is performed in permanent plots (50×50 m), and involves field surveys at frequencies which vary according to the level of the study. Field surveys are planned:

- at the community level (scale of 10×10 m) to assess the total vegetation cover and the stratified cover per layer, as well as the specific cover per vertical layer;
- at the population level to survey the understorey (up to 1.30 m), recording the specific percentage cover and density (in 50×50 cm detail).

Data matrices differ qualitatively and quantitatively. At the community level, species richness is reliable for the entire phytocoenosis; abundance is estimated less accurately. At the population level, the number of species recorded is limited to the understorey, and ascertained with a different sampling design, while abundance is assessed both by detailed cover scores and by counting aboveground shoots in different categories.

Data collection in the field considers all the plant species, including ferns, bryophytes and lichens. Given the value of the project at the European level, it is crucial to identify the species using the correct nomenclature. Field data are based on national referenced floras following Pignatti (1982), Corley *et al.* (1981), Corley & Crundwell (1991), Grolle & Long (2000) and local floras for lichens. The information must be ultimately managed using the Flora Europaea coded list, adopted by the Pan-European programme as a common base.

At the community level, 12 out of 25 sampling units (10×10 m) were surveyed in each fenced permanent plot. As in a chessboard, every other unit was chosen for survey. A further 12-unit system located in the surrounding area was sampled for comparative purposes. Data were collected following the Braun-Blanquet scale (1932, 1964), based on visual assessment of cover using coded percentage intervals: r = rare; + = <1%; 1 =

1-5%; 2 = 5-25%; 3 = 25-50%; 4 = 50-75%; 5 = 75-100%. The total vegetation cover per each fixed layer (tree, shrub, herb and moss layers) was also assessed, including percentages of litter and bare soil cover.

The population level approach entails data collection in 100 sampling units (50×50 cm) systematically distributed inside the permanent plot. In this case only the low layer of vegetation (up to 1.30 m) was considered. Each species was characterised by its percentage cover (computed by estimating its surface in cm<sup>2</sup> relatively to the sample unit area of 2500 cm<sup>2</sup>), and the number of functional individuals (aboveground shoots) was counted for vascular plants; for woody species individual height was also measured.

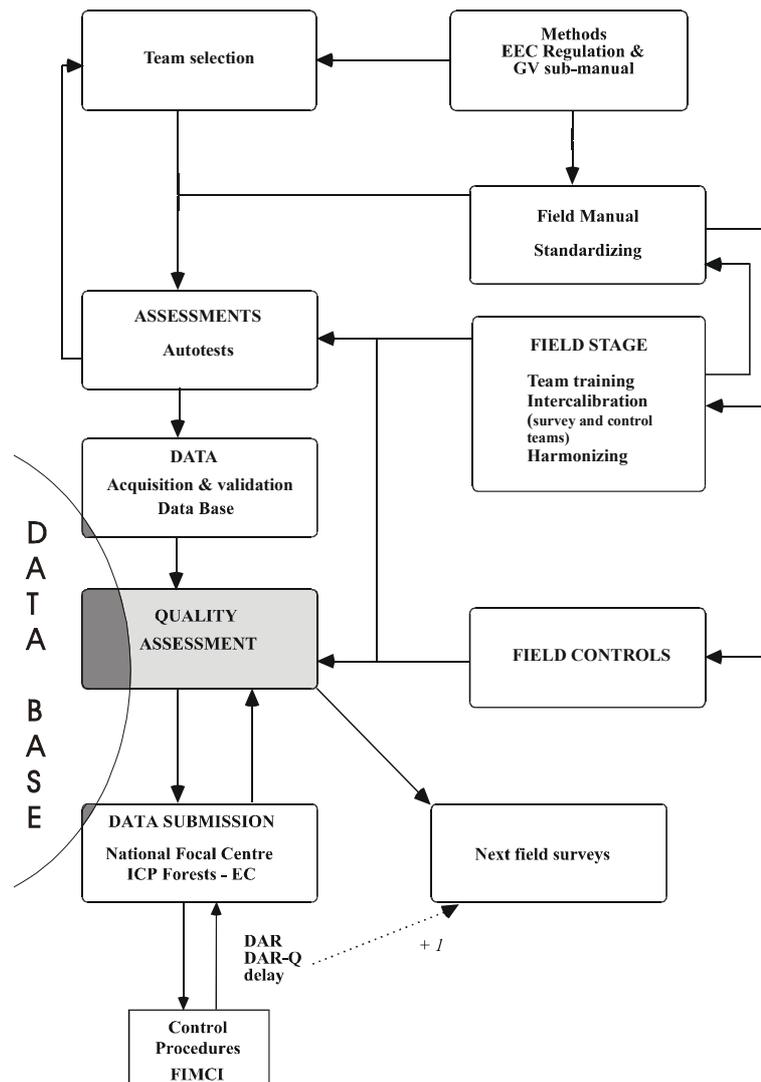
### 2.2. Quality assessment and database

Quality control is the best tool for achieving consistency of the data collected in the permanent plots. In fact, in this kind of investigation, all possible sources of error must be considered, including those deriving from the subjectivity of the surveyor and from the contradictory results which can appear during the various steps of data collection. The Italian "Ground Vegetation Assessment" group has developed an illustrative flowchart (Fig. 1) to highlight the critical points of the Quality Assurance (QA) procedure.

This paper discusses three phases of the QA procedure described in the flowchart, regarding the intercalibration and harmonisation of survey methods and the performance of the surveyors: field-training courses, field controls and autotests. The database software is also a very important quality assessment tool. In particular, it must be effective in avoiding data input errors, automatic format swaps, misspelling of species, invalid or inconsistent data, etc. The "Ground Vegetation Assessment" database was conceived for use in a LAN (Local Area Network) by multiple users with different access levels. The database Administrator can specify what kind of action each user may perform.

In addition to archiving data, the database can run functions to check their validity and consistency. The records must be imported from Excel files but the entire sequence undergoes a validation process before the new data are actually saved. In particular, the integrity of the datasets and their conformity to the user-defined range can be assessed. Moreover, the database allows subsequent crosschecks among attributes for further tests of validity and precision. Once these functions are accomplished, the database arranges summary tables according to the amended Annex VII of EU-Regulation n. 1091/94. These functions are an integral part of the overall systematic procedure which must be performed to provide adequate quality assurance (Fig. 1).

The system should guarantee the visualisation of a certain number of simple statistics, and permit export functions of analytic data so that the entire mass of information can be transferred to external statistical tools.



**Fig. 1.** A dynamic flowchart highlights the critical points where the "ground vegetation assessment" programme includes particular tasks to be performed to achieve Quality Assurance. Topics discussed in the text are shown in capital letters.

Obviously, the entire database is structured to allow future queries through a Web browser, within the limitations established to protect its contents and regulate its use.

### 3. RESULTS

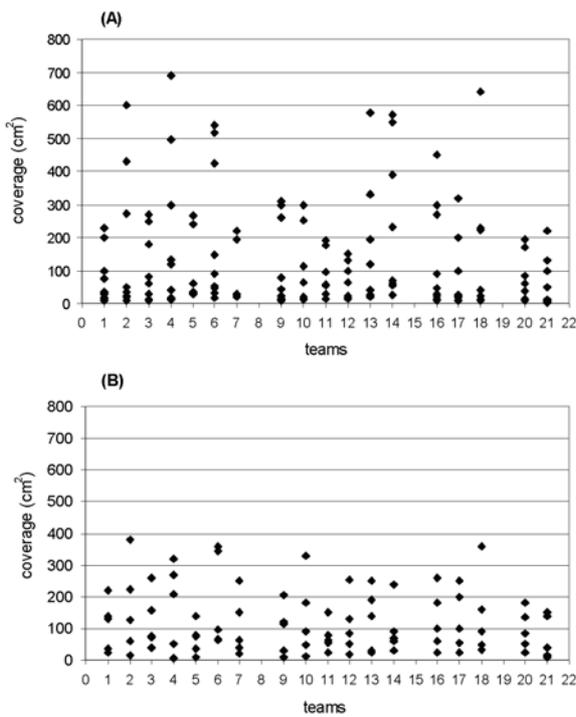
The result of many brainstorming sessions with all the participants (field teams included), trials, adjustments and repeated process analyses, was the creation of a procedural sequence which took into account the aims, requirements and considerations described above.

As explained above, this paper discusses only the operational phases related to the management of data series relating to forest ground vegetation. Through these steps (field-training courses, field controls, autotests) and the database functions, certain definitive applications are already functional.

#### 3.1. Field-training

The annual team-training course comprises a series of lectures followed by guided field work practice. The objective here is to assure conformity with the field manual (Canullo *et al.* 2000) and intercalibration among all the teams composed of two experienced members; a control team has also been used since the manual was written.

An important result emerging from this stage of the procedure was the evaluation of the expected distance between the performance of the control and the survey teams. Comparisons during field-training, especially in terms of extreme values, and repeated practice exercises on the same sampling units, are designed to progressively reduce dispersion (variance) of both the assessments and the counts (Fig. 2).



**Fig. 2.** Results of harmonisation and intercalibration exercises for a given species. A, first and B, second field exercise replications.

### 3.2. Field controls

After training, the teams perform the actual surveys in the permanent plots network. At the same time the control team carries out field controls on randomly chosen plots and a related fraction of sampling units. The data from the two sources can be used to observe the distance in the values recorded between the control and the survey teams, compared with the expected relative distance which emerged during the training course.

### 3.3. Autotests

The next stage of screening involves an "autotest": each surveyor separately assesses a 10% fraction of the sampling units at both observed levels. In this way, each team can be assessed to see whether the in-pair records are affected (unbalanced) by one of the surveyors.

To assess the performance stability of the teams, relative distances (from field controls) and individual scores (from autotests) can both be used as a warning system. When the distances between the survey and the control teams exceed and/or shift from the expected values by 10%, and when one of the two surveyors strongly influences the cover assessments, the quality protocol initiates immediate discussion and revision. If no agreement is reached, additional in situ field training is required, while special care is given to future team selection and the next field-training course. In the case of team unreliability, further analytic criteria can be used to define validity thresholds (i.e. more than one

degree of the Braun-Blanquet scale, more than 10% for total cover estimates and more than 20% for detailed population assessments) and part of the data can be invalidated.

Intercalibration and harmonisation of survey methods should be enhanced by these three phases of field-training, field controls and autotests.

### 3.4. Database

The QA/QC flowchart (Fig. 1) shows three critical points in database management (data acquisition and validation, quality assessment, data submission) in which the database is the essential tool.

### 3.5. Acquisition and validation

This procedure is characterised by a set of controls whose implications are illustrated below.

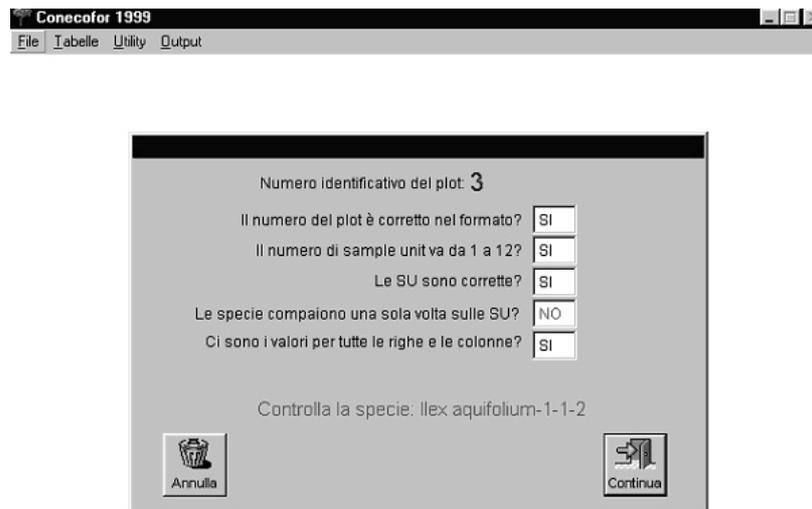
Each team inputs the field data into formatted spreadsheets without any other changes, therefore the species names must be pasted from a supplied list. Only problematic cases and provisional species attribution can be entered as further specified. Some immediate, simple controls are performed by the Excel filter functions. Afterwards, the files are directly importable by the database software.

The first part of the import procedure is the validation of all species names according to Pignatti's Italian flora (1982) using a dedicated archive. The software highlights each species name which does not correspond to the archive list; the authorised user must decide if it is a case of a misspelled word, or if a new species has to be changed according to the archive list, or added to the list.

Afterwards, the database performs some controls for each attribute, verifying its correspondence with the upper and lower limits previously set by the system Administrator. A simple warning allows the user to confirm the admission of out-of-range data: the ranges are conceived as security and plausibility data checks, according to criteria based on procedure, biology, the literature and experience.

The full matrix is then checked for sequence consistency, i.e. that the sampling unit numbers correspond to the different surveys (spring-summer, in plot - out plot combinations). Further controls on data integrity are also performed for both community and population levels of observations, by means of about twelve different crosschecks (Fig. 3). This also avoids misunderstanding of "null" values (no data vs zero value). While some control functions can be selected by the user, the import function executes special checking functions as routine, in order to separate "legal data" from incorrectly typed numbers.

The last tool of data acquisition and validation is the automatic association between the Italian flora archive and the coded archive of Flora Europaea (derived from the PANDORA taxonomic database system of the



**Fig. 3.** An example of check function warnings. The sequence identifies: plot number, sampling unit integrity, survey sequence integrity, species name duplication, consistency of relative attributes, and species identification with its position in the surveys and layers.

Royal Botanic Garden, Edinburgh). The relationship between the two archives is dynamic: it is activated when a previously unrelated species is added and must be explicitly confirmed by the Administrator.

When full nomenclature correspondence is not achieved, the software highlights the incorrect species name and an expert must find the correct attribution in the coded list.

The association is automatically suggested when a name fits both archives (but needs user confirmation as well). Presumed new species or those with a provisional attribution are coded with a provisional code and, if confirmed, can take on the new official code without all the stored data being lost. At the end of the control procedure, the full dataset is imported.

### 3.6. Quality Assessment

This process provides a key contribution to Quality Assessment for data storage and management (Fig. 1).

### 3.7. Data Submission

The result of the import process is an easily consulted data set organised by surveys, plots and observation levels. Simple summary tables can be obtained and full thematic matrices can be exported. A software section for addition and deletion of data makes the output available in text format according to EC Regulations, when the Administrator selects a particular year.

## 4. CONCLUSIONS

The whole process of data gathering and management for forest ground vegetation analysis in the framework of the CONECOFOR programme in Italy has provided an opportunity to address the problems of

protocol standardisation and data management. A Quality Assurance process has been established to allow continuous data comparability and facilitate its continuous improvement. A standard protocol for training, harmonising and intercalibrating the survey teams has been established as an essential prerequisite for obtaining sets of reliable data. In addition, field checks of expected relative distances between control and survey teams have been performed. A dedicated database was designed not simply as an "information manager" but as an active part of data management QA procedure; this is seen particularly in a set of import functions which process the standard files compiled by survey teams.

The validation of species names through an active association with specific archives avoids storing misspelled or unaccepted species, and relates field records to a coded list adopted at a European level. Attributes plausibility and data consistency and integrity are also assured by means of a number of crosschecks. While the data acquisition process makes a key contribution to QA procedure for data storage and management, the database comprises a fully organised data source for queries and thematic matrix exports. The hierarchic access to the database within a Local Area Network guarantees security in the management of the system. The procedure described has been operational since the 1999 surveys, so that the correct sequences of data sets are fully comparable. Obviously, subjectivity in forest ground vegetation assessment cannot be entirely avoided: apart from accidental errors and the natural fluctuation of attributes, visual estimates and the techniques used, however highly standardised, can lead to a degree of variability that is difficult to quantify.

This may have less importance when we are dealing with protocols for long-term surveys such as the

CONECOFOR programme, in which the major trends are expected to be well expressed by this quality of data.

The high number of errors due to manipulation and inconsistent data dropped significantly when the import procedure was activated in the database.

The procedure described here and the whole process of Quality Assurance are important methods for further error analysis leading to an operative, statistical level of comparability which can meet the requirements of a variety of objectives and research groups.

#### REFERENCES

- Allavena, S., R. Isopi, B. Petriccione & E. Pompei. 2000. *Programma Nazionale Integrato per il Controllo degli Ecosistemi Forestali – Secondo rapporto*. Ministero delle Politiche Agricole e Forestali, Corpo Forestale dello Stato.
- Braun-Blanquet, J. 1932. *Plant Sociology*. New York.
- Braun-Blanquet, J. 1964. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Springer, New York.
- Canullo, R., M-C. Allegrini, G. Nicoletti, G. Campetella, S. Allavena, R. Isopi, B. Petriccione & E. Pompei. 2000. *Analisi della vegetazione nelle aree permanenti della Rete Nazionale Integrata CONECOFOR. Manuale Nazionale di Riferimento per il coordinamento, i rilevatori ed i responsabili di zona*. Università degli Studi di Camerino.
- Christian, R.R., C. French, J. Gosz & R. Waide. 1999. Perspective on International Long term Ecological research. In: A. Farina (Ed.), *Perspectives in ecology: A glance from the VII International Congress of Ecology (Florence 19-25 July 1998)*. INTECOL, Backyus Publishers, Leiden, NL: 99-105 pp.
- Corley, M.F.V., A.C. Crundwell, M.O. Mull & A.J.E. Smith. 1981. Mosses of Europe and the Azores: an annotated list of species, with synonyms from recent literature. *J. Briol.*, 11: 609-689.
- Corley, M.F.V. & A.C. Crundwell. 1991. Additions and amendments to the mosses of Europe and the Azores. *J. Briol.*, 16: 337-356.
- Corona, P., G. Chirici & M. Marchetti. 2002. Forest ecosystem inventory and monitoring as a framework for terrestrial natural renewable resource survey programmes. *Plant Biosystems*, 136(1): 69-82.
- Dupouey, J.L. 1998. *Indicators for forest biodiversity in Europe: proposal for terms and definitions*. Technical Report, 4 (version 1.0), Birmensdorf.
- Elzinga, C.L., D.W. Salzer, J.W. Willoughby & J.P. Gibbs. 2001. *Monitoring Plant and Animal populations*. Blackwell Science.
- Ferretti, M., F. Alianiello, S. Allavena, T. Amoriello, E. Amorini, F. Biondi, A. Buffoni, F. Bussotti, G. Campetella, R. Canullo, A. Costantini, A. Cutini, G. Fabbio, C. Ferrari, P. Giordano, E. Magnani, A. Marchetto, G. Matteucci, C. Mazzali, G. Mecella, R. Mosello, B. Nibbi, B. Petriccione, E. Pompei, F. Riguzzi, G. Scarascia Mugnozza & M. Tita. 2000. The Integrated and Combined (I&C) evaluation system – Achievements, problems and perspectives. *Annali Istituto Sperimentale per la Selvicoltura*, 30: 151-156.
- Franklin, J.F. 1989. Importance and justification of Long-Term Studies in Ecology. In: G.E. Likens (Ed.), *Long-Term Studies in Ecology: Approaches and Alternatives*. Springer-Verlag, New York: 4-19.
- Grolle, R. & D.G. Long. 2000. An annotated check-list of Hepaticae and Anthocerotae of Europe and Macaronesia. *J. Briol.*, 22: 103-140.
- Hellawell, J.M. 1991. Development of a rationale for monitoring. In: F.B. Goldsmith (Ed.), *Monitoring for Conservation Ecology*. Chapman and Hall: 1-14.
- Michener, W.K. (Ed.). 1986. *Research data management in the ecological sciences*. Belle W. Baruch Library in Marine Science No. 16, University of South Carolina Press, Columbia, Sc.
- Pickett, S.T.A. 1991. Long-term Studies: Past Experience and Recommendation for the Future. In: P.G. Risser (Ed.), *Long-Term Ecological Research: an international perspective*. Scope 47: 71-88.
- Pignatti, S. 1982. *Flora d'Italia*. 1-3. Edagricole, Bologna.
- Risser, G.R., J.R. Melillo & J.R. Gosz. 1991. Current Status and Future of Long-term Ecological Research. In: P.G. Risser (Ed.), *Long-Term Ecological Research: an international perspective*. Scope 47: 275-285.
- Schniederjans, M.J. & M.C. Karuppan. 1995. Designing a quality control system in a service organisation: A goal programming case study. *European Journal of Operational Research*, 81: 249-258.
- Strayer, D.J., C.G. Glitzenstein, C.G. Jones, J. Kolasa, G.E. Likens, M.J. McDonnell, G.G. Parker & S.T.A. Pickett. 1986. *Long-term ecological studies: an illustrated account of their design, operation, and importance to ecology*. Institute of Ecosystem Studies Occasional Publications 1.
- Walterbeek, H.Th., P. Bode & T.G. Verburg. 1996. Assessing the quality of biodiversity via signal-to-noise ratio analysis. *The Science of the total Environment*, 18: 107-116.