COENOSES 4 (3): 169-173, 1989



# FUZZY LINGUISTICS CONCEPT IN REDESCRIPTION OF VEGETATION DATA

Zbigniew R. Boryslawski, Institute of Botany, University of Wroclaw, Kanonia 6/8, 50-328 Wroclaw, Poland
and
Ewa Krusińska, Institute of Computer Science, University of Wroclaw, Przesmyckiego 20, 51-151 Wroclaw, Poland

**Keywords.** Fuzzy linguistics, Redescription, Vegetation space.

**Abstract.** The concept of fuzzy linguistic variables in redescription of vegetation data is proposed. Species frequency based original data table is redefined on the basis of species life/growth-form category. The example presents analysis of 22 types of Pinus forests described by 300 species. The method may be used in transformation of sample space into other spaces defined on various species attributes.

## Introduction

Species and their performance are the most common attributes in studying vegetation processes both in small and large scale. But in many cases the responses to ecological factors described in terms of species scores are insufficient in detection of underlying factors or vegetation changes. Indeed, the real vegetation response may be hidden when entries in a table are merely species presence or abundance.

Many criteria may be adopted to accomplish analysis using other attributes rather than species to describe stands (see Feoli 1984). For instance Dale and Clifford (1976) recoded species presence data using higher taxonomic units to accomplish reduction of original data set. Feoli and Scimone (1984) used a set of species features in redefining vegetation table for textural analysis. There are also other approaches (e.g. Rejmanek 1977, Dale, Clifford, Ross 1984).

In this paper we present a method of redefinition of species based vegetation data using the concept of fuzzy linguistic variables. The fuzzy set theory has been already introduced in vegetation ecology (Dale 1977, Roberts 1986, Feoli and Zuccarello 1986, Dale 1988). The meaning of linguistics for ecosystem studies was discussed by Dale (1979).

## The method of fuzzy linguistic variables

The idea of linguistic variables in a fuzzy sense was introduced by Zadeh as an extension of fuzzy set theory (Zadeh 1965, 1975). Linguistic variables take as their value several fuzzy variables. Let us consider as example variable "development phase":

development phase = $\{F_1, F_2, F_3, F_4\}$ =
= {juvenile, prematured, matured, senile}.

Each of variables $F_i$ ($i=1, 2, 3, 4$) has a fuzzy character and a membership function may be defined for each.

The method of obtaining linguistic variables used in our paper was originally developed by Saitta and Torasso (1981) for medical purpose. It consists in transforming groups of several features with fuzzy character but coded as discrete numbers into linguistic variables. In our case the features correspond to a single species and variables to a group of species representing the same biological or ecological category. The characterization of vegetation may be now viewed as the problem of how much each category contributes to description of a vegetation type, with each single species as a specific question.

Assume that each linguistic variable $L_i$ ($i=1, 2, \dots r$) is described by a set of features $Q^{(i)}$. From the mathematical point of view the features $q_j^{(i)} \epsilon Q^{(i)}$ have a discrete character (are coded as discrete) and may be observed at $A_{ij}$ states. A weight $\gamma jk^{(i)}$ is associated with each state $s_k^{(i)}$ of the feature $q_j^{(i)} \epsilon Q^{(i)}$ ($j=1, 2, \dots M_i$) where $M_i$ is the number of features in $Q^{(i)}$, ($k=1, 2, \dots, A_{ij}$), where $A_{ij}$ is the number of states of the feature $q_j^{(i)}$. The weights range from $-1$ to 1, indicating the degree of agrement between $s_k^{(i)}$ and $L_i$. When $\gamma_k^{(i)}$ equals 0 there is no information, positive or negative about the agreement. The weights $\gamma_k^{(i)}$ are determined a priori.

Next we define the vector of weights $\alpha^{(i)}=(\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_{M_i}^{(i)})$ for the features in $Q^{(i)}$.

These weights can be set to 1 (no differential weighting), associated according to investigator's experience or calculated by a suitable numerical method. If the aim of study is identification the values of Student t statistic computed for differentiated groups may be used Saita and Torasso 1981). On can also use various weighting methods based on scalar products or information quantities (Orloci 1978, Feoli, Lagonegro and Orloci 1984).

We applied the following procedure to determine weights $\alpha_j^{(i)}$: First within each group we calculated column totals

$$T_{il} = \sum_{j=1}^{M_i} s_{jl}^{(i)},$$

$i=1, 2, \ldots, r$; $l=1, 2, \ldots, n$, where n is the number of columns, the state $s_{jl}^{(i)}$ is observed for the lth object.

Then we computed Spearman rank correlation coefficient between feature $q_j^{(i)}$ and the totals. Since for vegetation data some ranks are repeated the formula may be given as follows (Siegel 1956)

$$r_s = \frac{A + B - \sum_{l=1}^{n} d_l^2}{2\sqrt{AB}}$$

where $d_l = x_l - y_l$ ($x_l$, $y_l$, $l = 1, 2, \ldots, n$) are two sequences of ranks - for the feature $q_j^{(i)}$ and for the totals),

$$A = \frac{n^3 - n}{12} - \sum_x T_x,$$

$$B = \frac{n^3 - n}{12} - \sum_y T_y,$$

$$T_x = \frac{h_x^3 - h_x}{12}, \qquad T_y = \frac{h_y^3 - h_y}{12}$$

$h_x$ is the observed number tied at the same rank in the sequence x, $\Sigma T_x$ means summing over all different values x which repeat in the sequence x.

$h_y$ and $\Sigma T_y$ have the analogous meaning in reference to y.

For large n ($n > 10$) the transformed coefficient has the Student t distribution:

$$t = r_s \frac{n-2}{1-r_s^2}$$

with n—2 degrees of freedom.

This was a basis for weights $\alpha_j^{(i)}$. We put $\alpha_j^{(i)} = t_j^{(i)}$ where $t_j$ is the computed value of the t statistic for the features $q_j^{(i)} \in Q^{(i)}$. We set $\alpha_j^{(i)} = 0$ if $t_j^{(i)}$ was less than $t_{n-2,0.1} \cdot 10^{-1}$ where $t_{n-2,0.1}$ is the critical value of the Student t statistic with n—2 degrees of freedom at the significance level equal to 0.1). In the case of a single feature in linguistic variable we assume the weight to equal $\max \alpha_j^{(i)}$ ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, M_i$) the maximum over all values of $\alpha_j^{(i)}$).

Now, the total weight for each state $s_{jk}^{(i)}$ is given as:

$$w_{jk}^{(i)} = \alpha_j^{(i)} \cdot \gamma_{jk}^{(i)}.$$

Finally for each vegetation type we define variable $m_i$ ($i = 1, 2, \ldots, r$) associated with the linguistic variable $L_i$ as follows:

$$m_{il} = \sum_{j=1}^{M_i} w_{jl}^{(i)}$$

**Table 1. List of vegetation types and data source used in the example.**

| No | Type |
|---|---|
| 1 | Pinus forest on serpentine soils, SE Alps; Orno-Pinetum nigrae (Poldini 1984); |
| 2 | Pinus forest on alluvial deposits, SE Alps; Alno-Pinetum, (Poldini 1984); |
| 3 | Pinus forest on dry calcareous soils, N Alps; Dorycnio-Pinetum, (Poldini 1984); |
| 4 | Pinus forest on dry substrates, N Alps; Salici-Pinetum, Poldini (1984); |
| 5 | Pinus forest on alluvial deposits, SW Alps, W Italy; Calamagrostio pseudophragmitae-Pinetum, (Poldini 1984); |
| 6 | Cladonia rich coastal Pinus forest, Baltic region, Poland; Empetro-Pinetum (Matuszkiewicz and Matuszkiewicz 1973); |
| 7 | coastal Pinus forest, Baltic region, Poland; Empetro-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973); |
| 8 | suboceanic form of lichen rich Pinus forest, Poland; Cladonio-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973); |
| 9 | Subcontinental form of lichen rich Pinus forest, Poland; Cladonio-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 10 | Sarmatic form of subcontinental Pinus forest, Poland; Peucedano-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 11 | subboreal form of subcontinental Pinus forest, NE Poland, Peucedano-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 12 | continental form of fresh Pinus forest, SW Poland; Leucobrio-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 13 | suboceanic form of fresh Pinus forest, SW Poland; Leucobrio-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 14 | inland Pinus forest on wet podsol soils, Poland; Molinio-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 15 | boreal form of continental marsh Pinus forest, Poland; Vaccinio uliginosi-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 16 | typical form of inland marsh Pinus forest, Poland; Vaccinio uliginosi-Pinetum, (Matuszkiewicz and Matuszkiewicz 1973; |
| 17 | relict type of fresh Pinus forest, Germany; Peucedano-Pinetum, (Zeidler and Straub 1967); |
| 18 | Cladonia rich, suboceanic Pinus forest, Germany; |
| 19 | typical form of suboceanic Pinus forest, Germany; Leucobrio-Pinetum, (Zeidler and Straub 1967); |
| 20 | Scandinavian, boreal Pinus forest, SE Norway; Vaccinio-Pinetum boreale, (Kielland-Lund 1981); |
| 21 | Lichen rich, Scandinavian Pinus forest, SE Norway; Cladonio-Pinetum boreale (Kielland-Lund 1981); |
| 22 | subalpine type, Scandinavian Pinus forest, SE Norway; Barbilophozio-Pinetum lapponicae, (Kielland-Lund 1981). |

where $w_{jl}^{(i)}$ is the total weight associated with the state $s_{jl}^{(i)}$ observed for the lth object (l = 1, 2, ..., n).

Using variables $m_i$ the membership function $\mu_i$ can also be calculated. The following function was used in the study:

$$\mu_i (m_{il}) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{m_{il} - m_i}{s_i} \cdot \frac{1}{b_i} .$$

$m_i$ is the mean value for variable $m_i$ over all types of vegetation and $S_i$ is its standard deviation. Calculating $b_i$ we followed Saita and Torasso (1981) by computing this value in such a way that

$$\mu_i (m_i') \leqslant 10^{-1},$$

$$\mu_i(m_i'') \geqslant 1-10^{-1},$$

where

$$m_i' = \sum_{j=1}^{M_i} (\min_{1 \leqslant k \leqslant A_{ij}} w_{jk}^{(i)}),$$

$$m_i'' = \sum_{j=1}^{M_i} (\max_{1 \leqslant k \leqslant A_{ij}} w_{jk}^{(i)}).$$

This gives the value for membership function within interval [0, 1].

## Examples

The phytosociological material used for the example forms a subset of data bank prepared for large study on dynamics of forest vegetation. The detailed description of the data and its structure will be given later in another paper (Feoli, Boryslawski, in prep.) In our present study the initial data table contained 22 columns and 300 rows. The columns of the table represented vegetation types of Pinus forests, the rows corresponded to species. Brief description of the types is given in Table 1. Entries in the data table corresponded to the species frequencies derived from the table considered and coded with values 0, 1, 2, 3, 5, 7, 9. Each species was assigned to one of the life/growth-form category listed in Table 2.

**Table 2. Number of species assigned to each type of growth life category used in example.**

| | |
|---|---|
| A Deciduous phanerophytes* | 22 |
| B Evergreen phanerophytes | 6 |
| C Deciduous nanophanerophytes | 25 |
| D Evergreen nanophanerophytes | 2 |
| E Woody chamaephytes | 22 |
| F Herbaceous chamaephytes | 11 |
| G Hemicryptophytes | 95 |
| I Geophytes | 28 |
| J Therophytes | 2 |
| K Therophytes/hemicryptophytes | 1 |
| L Liverworts | 10 |
| M Lichens | 39 |
| N Mosses | 37 |

* Nomenclature and definitions according to Landolt (1977).

Now each life/growth-form category of species corresponds to a linguistic variable $L_i$ with species as features. The following initial weights $\gamma_{jk}^{(i)}$ were assigned to species scores:

−1, −0.9, −0.6, −0.4, 0.7, 0.9, 1.

Weights $\alpha_j^{(i)}$ were calculated as indicated earlier, that is we computed totals from scores for each vegetation type within each species category, the totals and the species scores were then ranked and Spearman rank correlation coefficient was calculated between them. Recoding was completed by calculating values for each category according to the function previously given. The final transformed table consisted of 22 columns and 13 rows.

Having transformed the data we can now examine the redefined vegetation types. This could be done in many ways but for the sake of example simplicity we limited ourselves to unsophisticated methods. We have used a minimum spanning tree calculated for Euclidean distance and also principal components analysis based on the covariance matrix. The results are summarised in Fig. 1.

The general pattern on the graph may be explained as follows:

As it is seen from Table 1 the first axis reflects S-N and the second W-E geographical variation. The first axis separates the types into two groups that differ in life/growth-form category diversity. The types to the right comprise low number of different categories. Group on the left is composed of Central Europe forests where almost all categories are presented with relatively high abundance. The second axis represents physiognomic trend from types abundant in trees and shrubs towards decrease of higher vegetation strata and increase in lower plants.

Individual MST links provide additional information. Connection 20-6 represents syngenetic affinity of type 6 to boreal forests. Indeed, this type is regarded as a relict on the south coast of Baltic (Matuszkiewicz and Matuszkiewicz 1973). Suboceanic Pinus forest 18 is linked with azonal forest from SE Europe through its characteristic combination of species and with boreal forests through its physiognomy. The types 19 and 8 although assigned to different syntaxa show affinity caused by their suboceanic character.

We do not intend to give a detailed discussion here as it will be the object of separate study. But we can

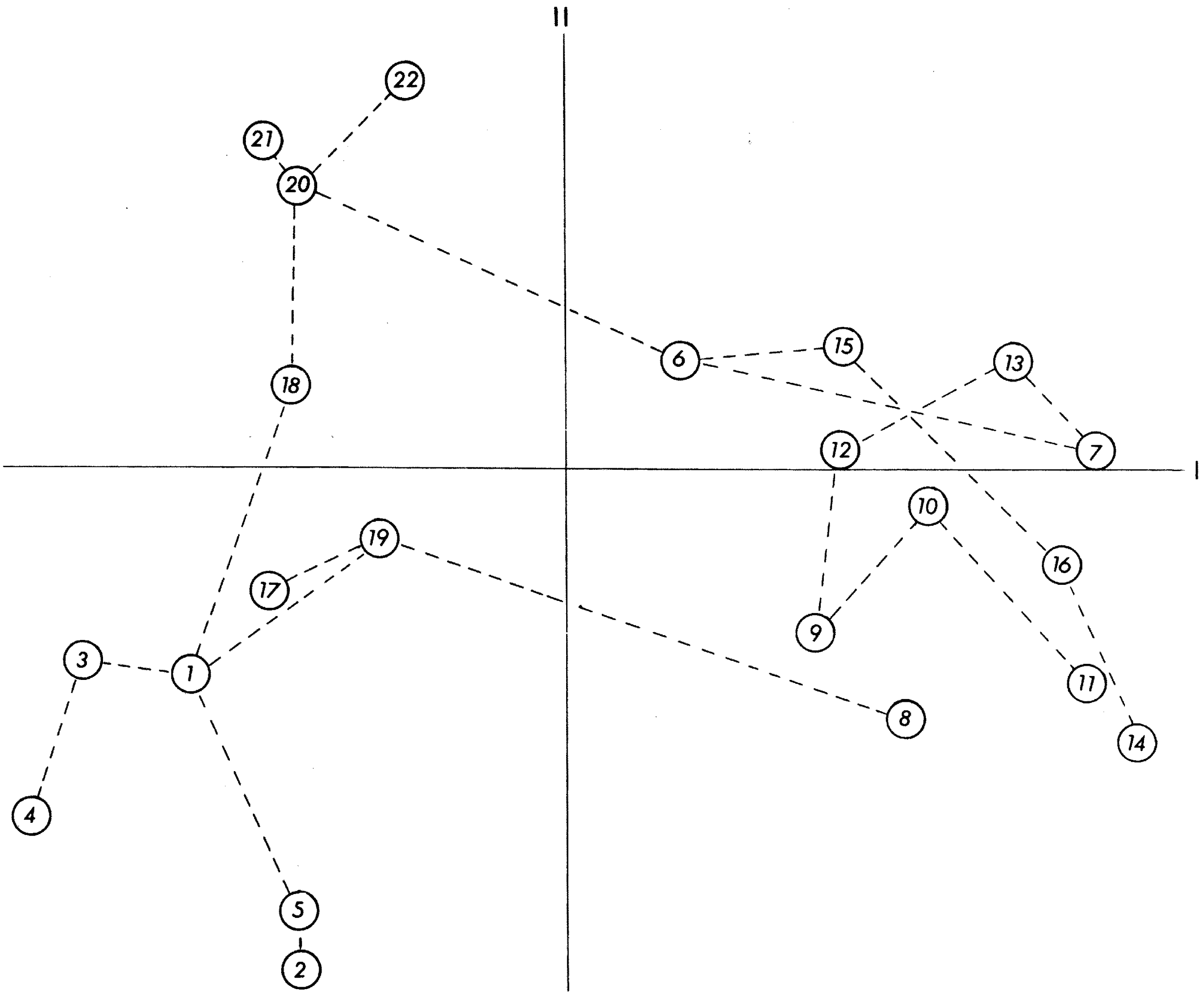


**Fig. 1. Ordination of 22 vegetation types. Dashed lines represent connections from MST. Note that length of the links is irrelevant.**

conclude that results presented seem to be promising. The type of analysis used by us provides a parsimonious description of both geographical distribution of the forests and their common features in terms of structural and physiognomic characteristics.

## Conclusions

Our example suggests that this type of data transformation is at least as efficient and sensitive as recoding by means of logical matrix multiplication. But the procedure we proposed for redefinition seems to be a logical concept rather than a new method itself. It might be profitable to regard a vegetation table as a set of specific questions while species occurrence and their attributes form answers. Such a view, which could also be taken for other vegetation data, implies the use of fuzzy linguistic approach in vegetation analysis.

The method while being free of restrictive assumption and allowing for elements of subjectivity is however based on formal mathematical theory. This flexibility might be desirable especially at the level of hypothesis generation. Thus by introducing this procedure we hope to propose a complement to a group of methods that are useful at the initial phase of data elaboration.

**Acknowledgement.** We would like to thank E. Feoli and to M.B. Dale for encouragement and for reading and commenting on the text. This work was partially supported by the Italian CNR within a grant to E. Feoli and has the University of Trieste under the terms of an Inter-University Scientific Cooperation Agreement.

## REFERENCES

DALE, M.B. 1977. Planning an adaptive numerical classification. Vegetatio 35: 131-136.

DALE, M.B. 1979. On linguistic approaches to ecosystems and their classification. In: Orlóci L., Rao C.R. and Stiteler W.M.

(eds.) Multivariate Methods in Ecological Work. International Co-operative Publishing House. Fairland. Maryland.

DALE, M.B. 1988. Some fuzzy approaches to phytosociology: ideals and instances. Folia Geobotanica Praha (in press).

DALE, M.B. and H.T. CLIFFORD. 1976. On the effectiveness of higher taxonomic ranks for vegetation analysis. Australian Journal of Ecology. 1: 37-62.

DALE, M.B., H.T. CLIFFORD and D.R. ROSS. 1984. Species equivalence and morphological redescription: a Stradbroke Island vegetation study. In: Coleman J., Covacevich J. and Davie P. (eds.) Focus on Strandbroke: New Information on North Stradbroke Island and surrounding areas 1974-1984. Boolarong Publ. Brisbane and Stradbroke Island Managament Organization, Amity Point.

FEOLI, E. 1984. Some aspects of classification and ordination of vegetation data in perspective. Studia Geobotanica 4: 7-21.

FEOLI, E. and M. SCIMONE. 1984. A quantitative view on textural analysis of vegetation and example of application of some methods. Archivio Botanico e Biogeografico Italiano 60: 71-94.

FEOLI, E., M. LAGONEGRO and L. ORLOCI. 1984. Information analysis of vegetation data. Tasks for vegetation science, vol. 10, Junk, The Hague.

FEOLI, E. and V. ZUCCARELLO. 1986. Ordination based on classification: jet another solution?! Abstracta Botanica 10: 203-219.

KIELLAND-LUND, J. 1981. Die Waldgesellschaften SO-Norwegens. Phytocoenologia 9: 53-250.

LANDOLT, E. 1977. Oekologische Zeigewerte zur Schweizer Flora. Ber. Geobot. Inst. ETH, 64: 64-207.

MATUSZKIEWICS, W. and J. MATUSZKIEWICZ. 1973. Pflanzensoziologische Uebersicht der Waldgesellschaften von Polen. 2. Die Kiefernwalder (in Polish, German summary). Phytocoenosis 2.4: 273-356.

ORLÓCI, L. 1978. Multivariate ana lysis in vegetation research. Junk, The Hague.

POLDINI, L. 1984. Eine neue Waldkieferngesellschaft auf Flussgeschiebe der Suedostalpen. Acta Bot. Croat. 43: 235-242.

REJMANEK, M. 1977. The concept of structure in phytosociology with references to classification on plant communities. Vegetatio 35: 55-61.

ROBERTS, D.W. 1986. Ordination on the basis of fuzzy set theory. Vegetatio 66: 123-131.

SAITTA, L. and P. TORASSO. 1981. Fuzzy characterization of coronary disease. Fuzzy Sets and Systems 5: 245-258.

SIEGEL, S. 1956. Nonparametric statistics for the behavioral sciences. Mc Graw-Hill, New York.

ZADEH, L.A. 1965. Fuzzy sets. Information and Control. 8: 338-353.

ZADEH, L.A. 1975. The concept of linguistic variable and its application toapproximate reasoning. Information Sciences. 8: 199-249 and 301-357.

ZEIDLER, H. and R. STRAUB. 1967. Waldgesellschaften mit Kiefer in der heutigen potentiellen naturlichen Vegetation des mittleren Maingebietes. Mitt. flor-soz. Arbeitsgem. N.F. 11/12: 88-126.

*Manuscript received: November 1988*