



Application of Burnaby's and Goodall's similarity indexes for local soil classification



Gilberto Bragato^{a,*}, Paola Ganis^b, Enrico Feoli^b

^a CREA – Centro di ricerca Viticoltura ed Enologia, Via Trieste 23, 34170 Gorizia, Italy

^b Università degli Studi di Trieste, Dipartimento di Scienze della Vita, Via Giorgieri, 10, 34127 Trieste, Italy

ARTICLE INFO

Keywords:

Burnaby's coefficient
Goodall's index
Local soil classification
Mixed data
Pairwise similarity measures
Tuber magnatum environment

ABSTRACT

Land use assessment is among the practical purposes of soil classification. Several researches has been specifically focused on the use of conventional surveys in the evaluation of soil suitability for agriculture products whose quality is influenced by the interactions between soil, plants and the biological stock of the rhizosphere. Our aim was to expand the application of the soil suitability protocol by relating soil types with biodiversity and the ecological equilibria of biological communities. This goal can be pursued by combining the approaches to soil type delineation with the techniques of quantitative ecology that are based on the similarity theory. Given the qualitative scale of several field-recorded attributes make their numerical processing difficult, we focused on numerical techniques for multivariate sets of data to combine with geostatistics. We thought these techniques would originate local soil classes meaningful in terms of both soil processes and soil suitability evaluation. Since auger boring data are formally comparable to vegetation data, we tested Goodall's and Burnaby's pairwise similarity indexes: the former assumes that pairs of observations sharing an infrequent value are more similar than pairs which share more frequent values; the latter considers associations among attributes, giving higher weights to independent ones.

We did an intensive soil survey in a 1200-ha flood plain of the Istria region, Croatia. The morphological characteristics of soil cores were recorded and analysed to produce pairwise similarities that were partitioned by hierarchical clustering into similarity vectors. Such vectors were in the end submitted to geostatistical analysis for the drawing up of similarity maps. Both similarity measures originated five partially overlapping clusters that were consistent with the main soil forming processes present in the investigated area. Goodall's index gave the most meaningful results, fulfilling three compulsory requirements for soil mapping: i) similarity vectors were meaningful in terms of fluvial dynamics; ii) similarities displayed a structured spatial variability; and, iii) similarity maps were consistent with the soil forming factors acting in the investigated area. The results obtained indicate that Goodall's similarity index could be currently used in soil suitability evaluation, allowing to better exploit field-recorded data and to extend its application to the relations existing between soil types and the ecological equilibria of biological communities.

1. Introduction

From its beginnings, land use assessment was among the practical purposes of soil classification: Whitney (1909) explicitly defined soil types – the lower level of his three-tiered classification system – as homogeneous units in terms of agricultural production. This interest was formalized in the FAO framework for land evaluation (FAO, 1976), after which several researches focused on the suitability of soil for agriculture products whose quality is influenced by the interactions between soil, plants and the biological stock of the rhizosphere. The most part of them adopted the following investigation sequence: i) soil

mapping at a semi-detailed to detailed spatial scale; ii) field trials to compare product quality from different soil types; iii) partition of soil types into suitability classes; and, iv) organization of map units into a hierarchy system related to the production than can be achieved (Vaudour, 2003; Bodin and Morlat, 2006; Costantini et al., 2012; Vaudour et al., 2015).

This approach was partially extended to non-wood forestry products, like truffles, that are related to the suitability of soil to host edible ectomycorrhizal fungi (Bragato et al., 2010; Bragato and Marjanović, 2016). A step forward might be that of relating soil types to biodiversity and the ecological equilibria of biological communities.

* Corresponding author.

E-mail address: gilberto.bragato@crea.gov.it (G. Bragato).

However, the extension of the soil suitability protocol to biological communities interacting with soil is limited by the second step of the procedure, which requires experimental conditions often fulfilled in cultivated fields only.

Adjustments to FAO protocol could be pursued by considering the specific characteristics of soil and biological community data, and by combining the approaches to soil type delineation with the techniques of quantitative ecology capable to classify observation data and compare classification matrices (Orlóci, 1978; Podani, 2000). The comparison of classifications obtained with variables used to describe independently the same set of observations can be done using similarity theory as a common basis (Feoli and Orlóci, 2011). Starting from the paradigm of plant community studies that similar communities tend to converge to similar physical-chemical environments, Feoli and Orlóci (2011) suggested to test the variation within a set of plant communities by comparing with the test of Mantel (Mantel, 1967; Mantel and Valand, 1970) the similarity matrices of vegetation relevés separately classified according to plant communities and environmental attributes, an approach that Feoli et al. (2017) used to analyse the area-diversity patterns among soils in South-East Spain.

Starting from these researches, we were aimed at testing if similarity theory was applicable to the subset of environmental factors acting in the pedosphere and that are considered in detailed soil surveys for soil suitability evaluation. Since the comparison between similarity matrices concerns the feature space, but soil classification must be also coherent with the spatial pattern of soil forming processes, we sought to ascertain whether it was possible to translate a local soil classification in terms of pairwise similarity values that had significance both in the feature space and the geographical space. Only when both requirements are fulfilled, the comparison between similarity matrices might be used correlating biological communities and soil environment.

Assuming soil as a complex natural body organized in components of increasing complexity that are recorded in terms of the hierarchical organization of soil categories (Hoosbeek et al., 1999), local relationships between soil and biological communities may be investigated by focusing on discrete, lower-level classes like soil series and phases (Soil Science Division Staff, 2017) that delineate mutually exclusive soil categories and map their spatial distribution as a set of non-overlapping polygons. However, the partition of the soil continuum into lower-level classes has to deal with the transition of classification criteria from clear-cut genetic relationships to rules assessing the influence of local factors and processes on soil characteristics and behaviour (Butler, 1980), and the shift from clear discontinuities in landforms to diffuse vertical and lateral subsurface variations in the feature space (Burrough et al., 1997). A further question is the lack of knowledge about local relationships between soil classes and biological communities, which can be addressed by delineating soil classes with no predefined diagnostic characteristics, thus avoiding to neglect meaningful soil-biological community relationships.

All these questions had been addressed by partitioning soil into continuous classes. Starting from fuzzy logic, McBratney and de Grujter (1992) proposed fuzzy k-means (FKM) to partition soil into classes displaying a gradual variability in the feature space. Combined with geostatistical techniques, FKM was effective in characterizing continuous soil classes in the geographical space (McBratney et al., 1992; Odeh et al., 1992; Triantafyllis et al., 2001). When applied to quantitative morphological attributes of auger boring data, the consistent increase in the number of observations helped to formulate spatial distribution models also useful for a locally oriented land use planning (Verheyen et al., 2001; Bragato, 2004). The selection of horizon sequences representative of continuous soil classes was also the goal of Carré and Girard (2002) and Carré and Jacobson (2009). Focusing on the feature space, they implemented a dynamic fuzzy clustering procedure that, after calculating a distance metrics over horizons, allocates profiles to existing classes, or creates a new classification of the profiles. These two lines of investigation were



Fig. 1. Location of the investigated area.

incorporated in the bottom-up analytical sequence proposed by Odgers et al. (2011a, 2011b) to partition the soil continuum into soil series classes.

Since fuzzy clustering can be considered a logical extension of the similarity theory (Feoli and Orlóci, 2011), the membership matrices produced with fuzzy clustering could be used to compare soil types and biological communities. However, since these approaches can only process quantitative and binary attributes, their application narrows the investigation to sets of data almost only obtained from soil profiles that, due to high laboratory costs (see for instance Kempen et al., 2012), consistently limit the sample size, resulting in a lower detail in the geographical space and a loss of information on local soil forming processes.

According to Bragato (2004), sample density can be increased by exploiting the information obtained from less expensive auger boring campaigns, which field-recorded attributes are expression of soil forming processes like those listed by Bockheim and Gennadiyev (2000). To be effective, this approach have to deal with the question that auger boring recordings are composed by quantitative, nominal and ordinal data – i.e. mixed data – that should be processed together.

Since auger boring data are formally comparable to vegetation data, we focused our attention on pairwise similarity measures for mixed data, specifically considering the similarity index of Goodall (1966) and the similarity coefficient of Burnaby (1970). Goodall built a probabilistic index based on the assumption that a pair of observations sharing an infrequent value is more similar than two which share a more frequent one. Burnaby instead considered the associations among attributes, giving higher weights to independent ones and introducing probability in order to account for the frequency of occurrence of different attribute states.

At the time they were proposed, both methods were ignored because of the lack of a theoretical framework on multiple class membership and spatial variability. Their application was also limited by the remarks of Gower (1970) who criticized their choice to give more

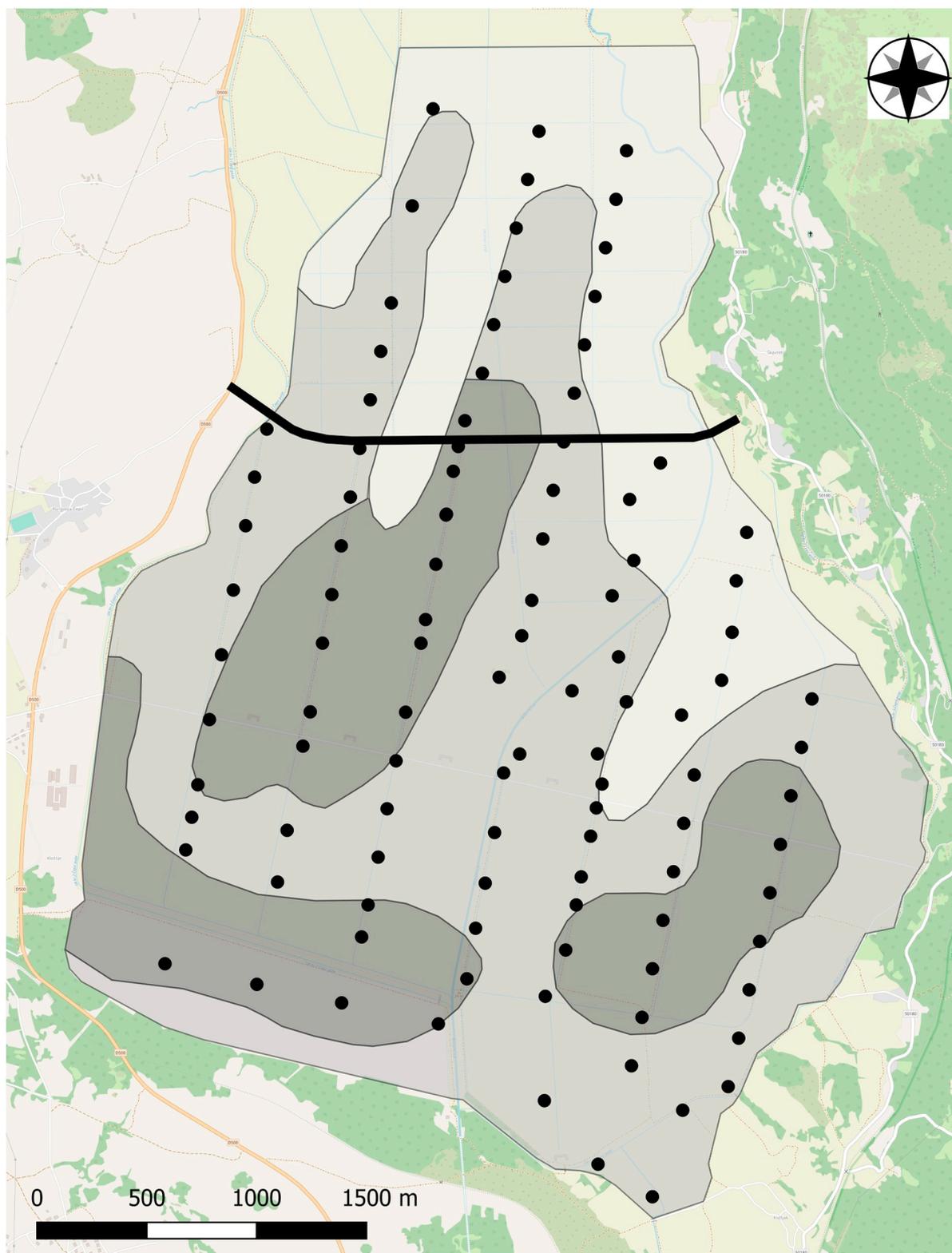


Fig. 2. Soil mapping units (SMUs) (grey areas) of the 1:50,000 soil map of Croatia, and sampling locations (black circles) in Čepić Polje. SMUs were classified in relation to the drainage class: the darker the color, the slower the drainage. The bold line marks a cobbled road already present in 1820.

weight to rare events and the use of uncorrelated (Goodall, 1966) or correlated but automatically weighted attributes (Burnaby, 1970). Taking account of Gower's observations and the progress of fuzzy logic, Goodall and Feoli (1991), and Carranza et al. (1998) set up software tools that successfully tested both indexes in quantitative ecology investigations.

Our work started from the hypothesis that the theory of similarity and the techniques of numerical analysis connected to it could form a framework in which to insert local soil classifications that take into account locally significant attributes. Our specific purposes were: i) to test the usefulness of Burnaby's and Goodall's similarity measures when processing mixed data from auger boring recordings; and, ii) to find

Table 1

Descriptive statistics of the selected attributes ($n = 105$). For discrete attributes, the frequency of occurrence f_i of each discrete state is reported.

A. Discrete attributes									
Attribute	Type	State	f_i %	State	f_i %	State	f_i %	State	f_i %
Prevailing hue	Binary	2.5Y	27	10YR	73				
Chroma, C horizon	Binary	/3	20	/4	80				
2nd horizon	Nominal	No	24	AC	43	B	15	CI	18
Value, A horizon	Ordinal	3/	3	4/	52	4.5/	30	/5	15
Chroma, A horizon	Ordinal	/2	18	/3	55	/4	27		
Value, C horizon	Ordinal	2/	6	3/	74	4/	20		
Texture class ^a	Ordinal	SIL	44	SiCL	47	SiC	9		

B. Continuous attributes				
Attribute	Mean	Std. dev.	Min	Max
Thickness 2nd horizon, cm	23	17	0	65
Depth of RMFs, cm	62	26	10	100
Redox concentrations, %	6	10	0	60
Redox depletions, %	5	6	0	30

^a SIL: silty loam; SiCL: silt clay loam; SiC: silty clay.

which of the two measures was best suited for local soil classifications and the assessment of their relationship with biodiversity and the local ecological equilibria of biological communities. The ultimate goal was to test an analytical sequence capable of emphasizing the effect of the main soil forming processes, and delineating meaningful local soil classes possibly characterized by gradual boundary transitions.

We did the investigation in a polje that was characterized by the presence of the typical landforms of a fluvial landscape. The area was chosen to fulfil the following criteria: i) the prevalent effect of few, well-known soil forming processes; ii) the presence of historically recorded environmental modifications that the survey should outline; and, iii) a gradual soil variability in space. The research took advantage of the presence in the area of the subterranean ectomycorrhizal fungus *Tuber magnatum*, a renown truffle species. Since this species is highly selective for specific fluvial landforms and soils (Bragato and Marjanović, 2016) and a previous research in the same polje outlined such environmental connections (Bragato et al., 2010), we exploited the environmental selectivity of *T. magnatum* to further examine the usefulness of the tested approach.

2. Materials and methods

2.1. Study area and soil survey

We did the investigation in the 1200-ha area of Čepić Polje (55°19'06"N lat, 14°13'10"E long) (Fig. 1). It is a depression flood plain located in the eastern part of Istria (Croatia) characterized by a north-east to south-west geological sequence of Cretaceous limestones and Eocene turbidites. Čepić Polje is actually characterized by Holocene depositions of soil materials eroded from turbidites and deposited by River Boljunščica. According to 1:50,000 soil map of Croatia – Sheet 16 (Škorić et al., 1987) and the World Reference Base (IUSS Working Group WRB, 2014), the main soil types in the slopes surrounding Čepić Polje are Lithic Leptosols/Eutric Cambisols on limestones and Eutric Regosols on turbidites. The fluvial plain is instead characterized by three soil mapping units belonging to Eutric Regosols that can be placed in a Fluvisol-Cambisol sequence (Fig. 2). Škorić et al. (1987) mainly distinguished them on the basis of the depth of the water table because the polje was partly occupied by a shallow lake until 1932, when it was reclaimed thanks to an artificial tunnel that flowed water down into the Adriatic Sea. The lake originated about 7000 years B.P. from a massive colluvial discharge or, alternatively, a slope failure associated to a

major tectonic event (Balbo et al., 2006) that disconnected River Boljunščica from the neighbouring River Raša, which flows into the Adriatic Sea about 20 km south-east of Čepić Polje.

According to the Köppen-Geiger classification (Kottek et al., 2006), the climate of the area can be classified of type Cfb, i.e. warm temperate, fully humid with warm summer. In the 1961–2016 time interval, the nearest weather station of Pazin recorded a mean annual temperature of 11.4 °C – with mean minimum and maximum values of 2.8 °C and 21.0 °C in January and July, respectively – and a mean annual rainfall of 1132 mm, ranging from 69 mm in July to 143 mm in November (Državni Hidrometeorološki Zavod, 2018).

The present study was based on auger boring recordings of an intensive soil survey done in August 2005 to compare soils of areas characterized by the presence or absence of *T. magnatum* (Bragato et al., 2010). In that survey, 82 locations were located at the nodes of a 225 × 450 m rectangular grid that evenly covered the polje and further 50 locations were randomly selected in five transects. We specifically used the data recordings of the systematic sampling and 23 locations on transects, selected to investigate the spatial variability at distances shorter than the grid mesh while limiting the uneven spatial distribution of locations in transects (Fig. 2). Each location was augered to a depth of 100 cm, recording the sequence of horizons and, for each horizon, thickness, matrix color, percentage of redoximorphic features (RMFs) and texture class by the feel method. Eleven independent attributes were selected to allow the comparison between similarity indexes (Table 1).

2.2. Similarity measures

The similarity measures we considered were determined in two steps: pairwise similarities/dissimilarities s_{jk}/d_{jk} between observations j and k were independently calculated for the i th attribute, then they were combined to produce an overall pairwise similarity measure S_{jk} .

Goodall (1966) defined the dissimilarity d_{jk} as the probability that a pair of observations (j,k) would be as different as they are if the two observations simply constituted a random sample of attribute values from the whole set. In practice, given a sample of n observations, d_{jk} is estimated from the frequency distribution of the various states v of the attribute in the set of n observations. When observations j and k have the same state q , the computation of d_{jk} only depends on the proportion p_l of occurrence of q in the population.

$$d_{jk} = \sum_{q \in Q} p_l^2 \quad (1)$$

where $1 \leq q \leq v$ and Q is the set

$$Q = \{q: (p_l \leq p_j)\} \quad (2)$$

The computation of d_{jk} changes with the attribute scale when pairs of observations with differing states are considered (see the Appendix of Goodall, 1966). In nominal and binary attributes, they are all regarded as equally dissimilar:

$$d_{jk} = 1 \quad (3)$$

When attributes are ordinal, pairwise dissimilarities are ranked by their probability of occurrence, and d_{jk} is calculated according to the number of states lying between each pair of observations: the fewer they are, the less dissimilar are observations. In case of quantitative attributes, dissimilarities are ordered by the magnitude of the difference between states and dissimilarity depends simply on the difference between the two values.

The separate probabilistic dissimilarities are then combined by Fisher's transformation for continuous probabilities according to the assumption that the states taken by m different attributes in the same observation are independent

$$\chi^2 = -2 \sum_{i=1}^m \ln(d_{jk})_i \quad (4)$$

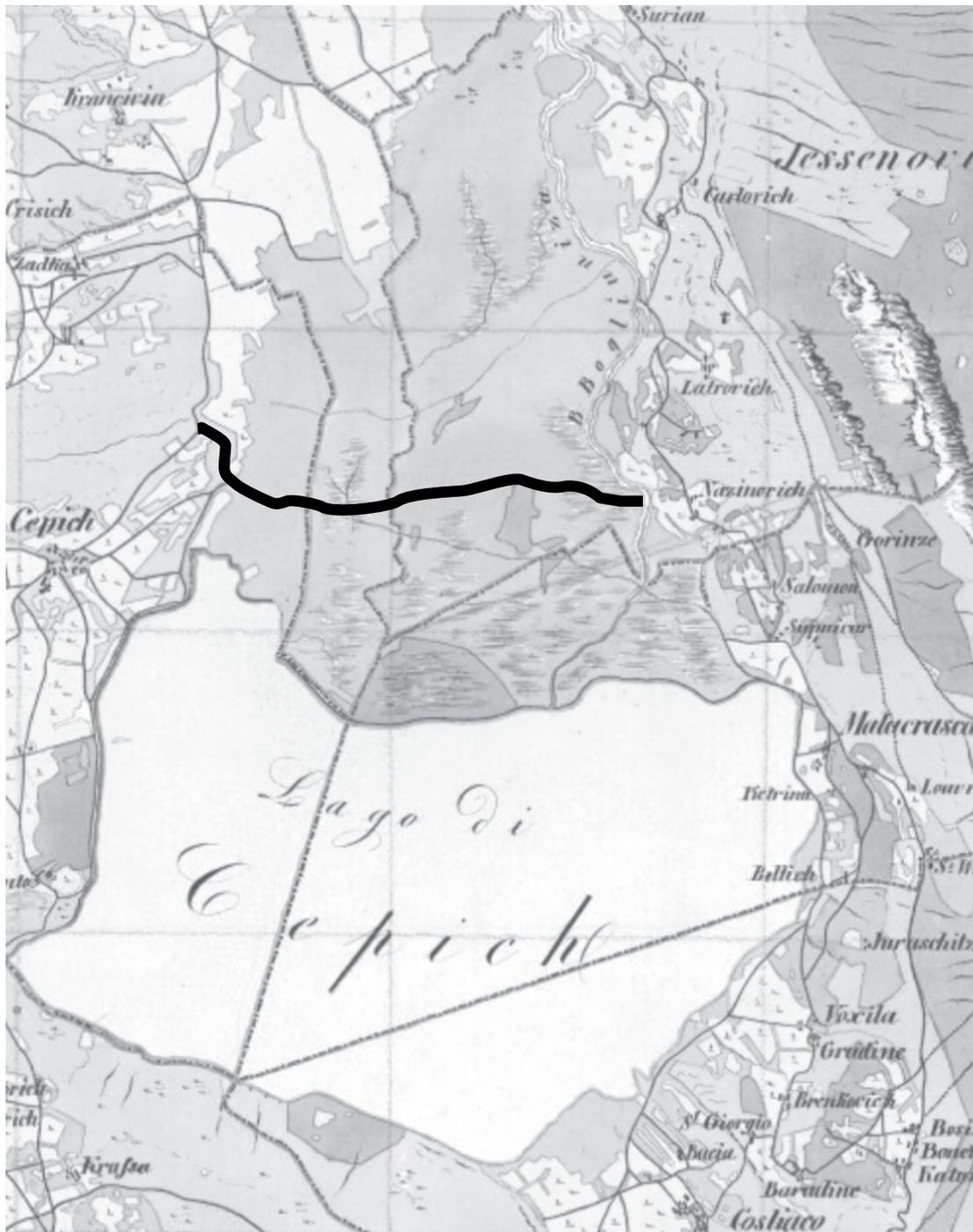


Fig. 3. Part of the 1:28800-scale chorography map of Istria and Dalmatia published in 1820. The bold line marks the cobble road also reported in Fig. 1, at that time located upstream of a marshy area at the mouth of River Boljunščica.

where the quantity χ^2 is distributed as χ^2 with $2m$ degrees of freedom (Fischer, 1948).

When applied to discrete probabilities, Fisher's transformation is affected by a positive bias that can be corrected following the suggestions of Lancaster (1949).

$$\chi^2 = 2 \sum_{i=1}^m \{1 - [(d_{jk})_i \ln(d_{jk})_i - (d'_{jk})_i \ln(d'_{jk})_i] / [(d_{jk})_i + (d'_{jk})_i]\} \quad (5)$$

where $(d'_{jk})_i$ is the first smaller dissimilarity value next to $(d_{jk})_i$. The sum of continuous and discrete probabilities is still distributed as χ^2 with $2m$ degrees of freedom, and the probability of the χ^2 value combines

dissimilarities. The overall pairwise similarity S_{jk} is in the end calculated as the complement to 1 of the χ^2 value probability.

Burnaby's similarity coefficient (Burnaby, 1970) was reconsidered and improved by Carranza et al. (1998). It is defined by the equation

$$K_{jk} = \left[\sum_{i=1}^m w_i (s_{ij})_i (I_{jk})_i \right] / \left[\sum_{i=1}^m w_i (I_{jk})_i \right] \quad (6)$$

where w_i is the weight assigned to the i th attribute on the basis of its independence from all other attributes, and $(I_{jk})_i$ is the information weight of the states of the i th attribute in the couple of observations under comparison. For ordinal and quantitative attributes, the w_i is null since the probabilities of quantiles are equal.

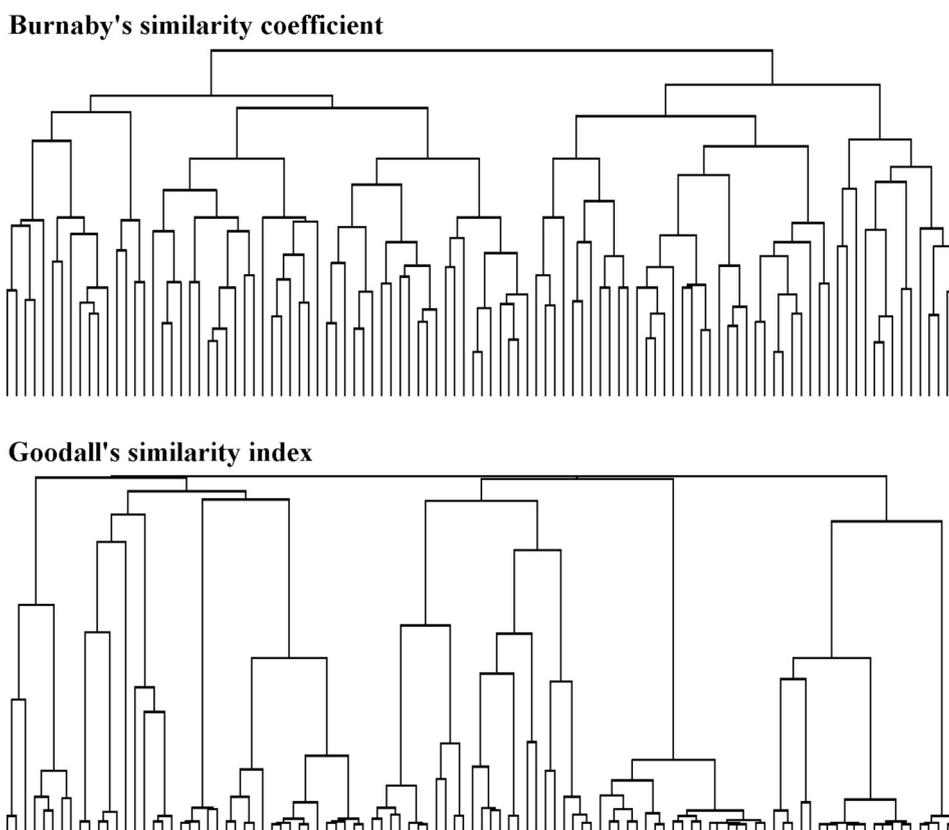


Fig. 4. Dendrograms of Burnaby's and Goodall's pairwise similarities obtained with the complete linkage clustering criterion.

Table 2

Contingency table of Burnaby's vs. Goodall's similarity index. The χ^2 value is 186.

	Bu1	Bu2	Bu3	Bu4	Bu5
Go1	8	0	0	0	0
Go2	1	5	16	10	0
Go3	1	14	7	3	0
Go4	0	0	1	18	0
Go5	6	0	0	1	1

Table 3

Descriptive statistics of the five similarity vectors.

Cluster	Mean	Std. dev.	Min	Max	Skewness	Kurtosis
A. Burnaby's coefficient						
Bu1	0.36	0.08	0.19	0.55	0.22	2.66
Bu2	0.43	0.09	0.24	0.62	0.24	2.43
Bu3	0.43	0.09	0.24	0.62	0.28	2.13
Bu4	0.38	0.10	0.18	0.55	-0.18	2.01
Bu5	0.36	0.09	0.19	0.57	0.43	2.65
B. Goodall's index						
Go1	0.41	0.17	0.12	0.81	0.67	2.62
Go2	0.49	0.14	0.14	0.75	-0.32	2.84
Go3	0.48	0.17	0.15	0.77	-0.07	1.97
Go4	0.51	0.28	0.12	0.97	0.51	1.89
Go5	0.53	0.23	0.07	0.96	0.33	2.18

The similarity $(s_{jk})_i$ assumes different states according to the scale of the i th attribute. For nominal attributes, it is 1 if the attribute states agree, 0 otherwise. For ordinal and quantitative attributes, it is calculated on the basis of

$$(s_{jk})_i = 1 - (x'_j - x'_k)/(x'_{max} - x'_{min}) \tag{7}$$

where x'_j , x'_k , x'_{max} and x'_{min} are quantile class marks calculated using

ranks of attribute states.

Burnaby suggested to always compute w_i on 2×2 contingency tables, irrespective of the attribute scale. This result is achieved transforming nominal attributes with v_i states into v_i binary attributes describing the presence-absence of each state. Ordinal and quantitative attributes are instead transformed into quintiles and the χ^2 statistics is computed on 2×2 contingency tables obtained by removing median classes and grouping the remaining cells by tetrads. This suggestion was strongly criticized by Gower (1970) but, when compared to other ways of obtaining contingency tables from quantitative attributes, it gives results closest to the product-moment correlation coefficient (Carranza et al., 1998). The use of quantiles instead of other types of classes (i.e. the classical equal range frequency classes) guarantees the equidistribution of observations between classes, avoiding empty classes (for instance in the case of bi- or multimodal distributions) or classes with few elements (i.e. in the tails of the normal distribution).

Pairwise similarities were calculated with the programs *Simil* (Goodall et al., 1991) and *Burnaby* (Carranza et al., 1998), disregarding double zeros in the computation of Burnaby's coefficient. The resulting similarity matrices were partitioned by hierarchical cluster analysis using several clustering criteria (Podani, 2000). The classification suggested by dendrograms was tested according to the permutation technique proposed by Feoli et al. (2009), which consists in calculating the evenness of the eigenvalues of the within/between similarity matrix of the selected clusters, and finding the probability that the observed evenness would be lower than that calculated by random allocation of observations to clusters after n permutations – in this case n was set to 100,000. The sharpest and more significant classification resulting in the lowest number of clusters was used to define fuzzy sets with the method of Feoli and Zuccarello (1991). According to this method, the degrees of membership to sets may be calculated by averaging the similarity values within and between clusters. It is based on the idea that a similarity matrix is a fuzzy set matrix in which similarity values are all

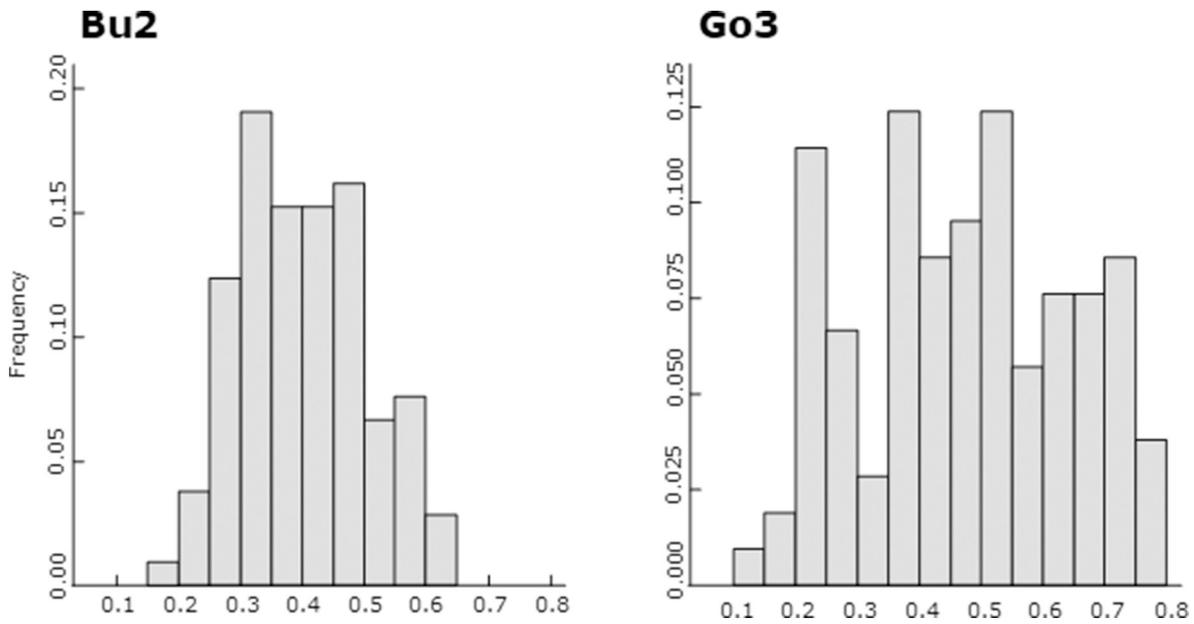


Fig. 5. Histograms of a pair of similarity vectors.

Table 4

Within/between similarity matrices. The W/B similarity ratios are obtained by dividing the within cluster similarity on matrix diagonal by the average of between cluster similarities on matrix columns.

A. Burnaby's coefficient						
	Bu1	Bu2	Bu3	Bu4	Bu5	W/B similarity ratio per cluster
Bu1	0.47	0.36	0.38	0.31	0.31	1.39
Bu2	0.36	0.55	0.42	0.33	0.38	1.47
Bu3	0.38	0.42	0.55	0.39	0.37	1.41
Bu4	0.31	0.33	0.39	0.48	0.30	1.45
Bu5	0.31	0.38	0.37	0.30	0.37	1.30
Average W/B similarity ratio						1.44

B. Goodall's index						
	Go1	Go2	Go3	Go4	Go5	W/B similarity ratio per cluster
Go1	0.72	0.36	0.41	0.41	0.37	1.86
Go2	0.36	0.60	0.40	0.51	0.45	1.38
Go3	0.41	0.40	0.68	0.33	0.50	1.64
Go4	0.41	0.51	0.33	0.96	0.36	2.37
Go5	0.37	0.45	0.50	0.36	0.87	2.06
Average W/B similarity ratio						1.86

degrees of belonging to the sets represented by the single objects to be classified (Zhao, 1986; Marsili-Libelli, 1991; Feoli et al., 2009). All procedures on similarity matrices were carried out with the package *MATEDIT* (Burba et al., 2008).

2.3. Geostatistical analysis

The similarity values of the clusters obtained from similarity matrices were interpolated with the geostatistical approach, which was aimed at predicting the unknown value of a random variable $Z(x)$ at an unobserved location x_0 using the values recorded at x_α surrounding sampling locations.

Interpolations were based on the intrinsic hypothesis (Journel and Huijbregts, 1978):

$$E[Z(x_2) - Z(x_1)] = 0 \tag{8}$$

and

Table 5

Centroidal values of the five clusters extracted from similarity matrices.

Attribute	Cluster				
A. Burnaby's coefficient					
	Bu1	Bu2	Bu3	Bu4	Bu5
Color, A horizon	10YR 4/3	10YR 4.5/3	10YR 4/3	10YR 4/4	2.5Y 4.5/2
Color, C horizon	10YR 5/4	10YR 5/4	10YR 5/4	10YR 5/4	2.5Y 6/3
Intermediate horizon	–	AC, 22 cm	C, 34 cm	–	AC, 34 cm
Depth of RMFs, cm	34	42	64	88	55
RMF concentrations, %	12	4	9	1	3
RMF depletions, %	16	8	5	1	10
Texture class ^a	SiL	SiCL	SiCL	SiL	SiL

Attribute	Cluster				
B. Goodall's index					
	Go1	Go2	Go3	Go4	Go5
Color, A horizon	10YR 4/3	10YR 4/3	10YR 4.5/3	10YR 4/3	2.5Y 4.5/3
Color, C horizon	10YR 5/4	10YR 5/4	10YR 5/4	10YR 5/4	2.5Y 6/3
Intermediate horizon	–	B, 32 cm	AC, 22 cm	–	AC, 28 cm
Depth of RMFs, cm	30	67	46	100	51
RMF concentrations, %	7	6	4	0	7
RMF depletions, %	10	1	10	0	13
Texture class ^a	SiCL	SiL	SiCL	SiCL	SiL

^a SiL: silty loam; SiCL: silt clay loam.

$$E[Z(x_2) - Z(x_1)]^2 = 2\gamma(h) \tag{9}$$

where $Z(x_1)$ and $Z(x_2)$ are random variables at any two locations; h is the separation vector between locations; and $\gamma(h)$ is the semivariance. The model applied to the scattergram of semivariances against distance – i.e. to the experimental variogram – is used for the following interpolation step.

Most similarity vectors fulfilled assumption (8) and originated variograms with a sill. Predictions were then obtained with the Ordinary Kriging (OK) predictor (Goovaerts, 1997; Oliver and Webster, 2014):

$$Z(x_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha) \tag{10}$$

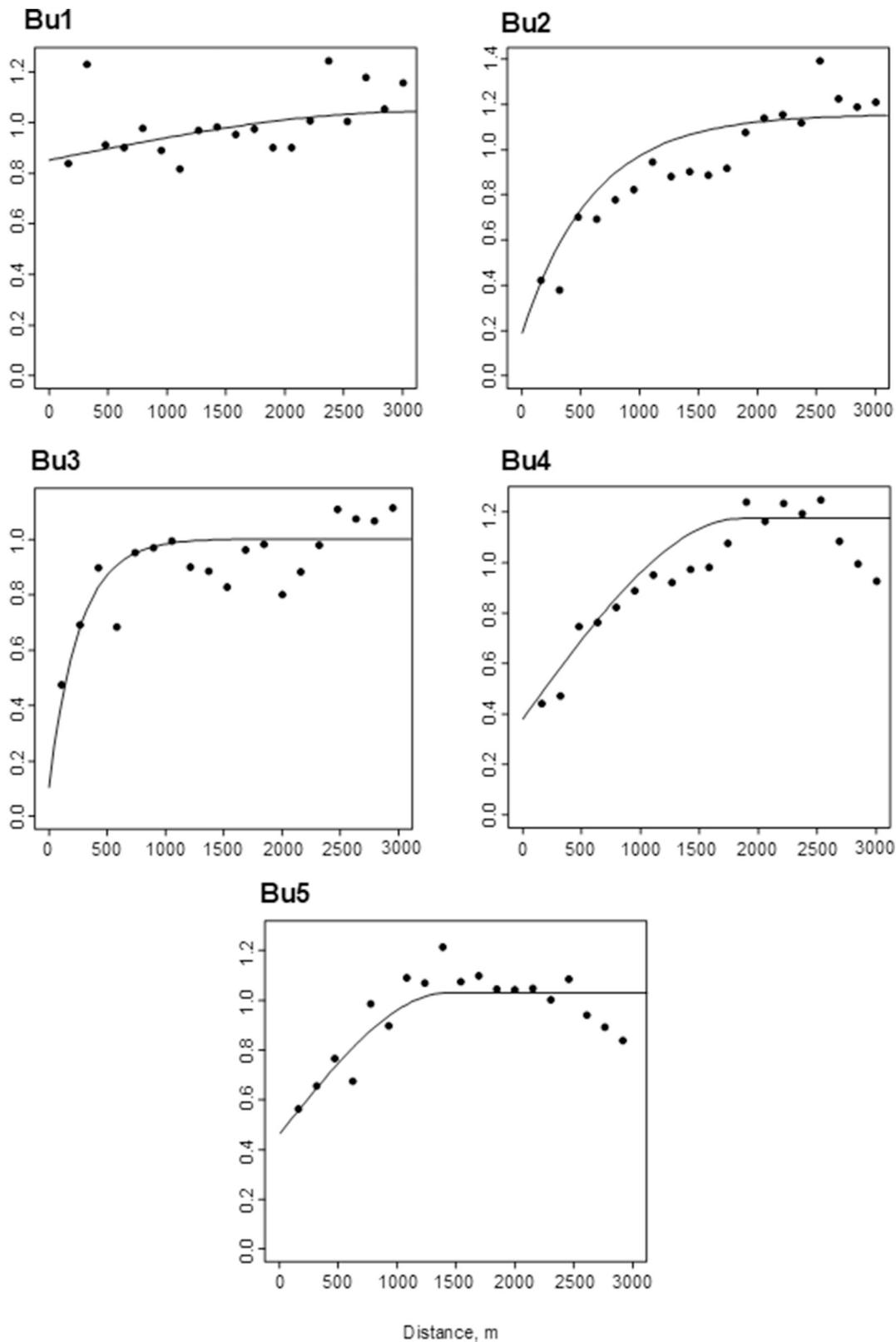


Fig. 6. Variograms of Burnaby's similarity coefficient. Data were previously transformed by Gaussian anamorphosis.

where λ_α are kriging weights which sum is 1.

Clusters that showed a non-constant mean were interpolated with the Universal Kriging (UK) predictor, which models the local trend with a polynomial regression of the Cartesian coordinates (Diggle and Ribeiro, 2007) and convert observations to residuals before calculating

the variogram:

$$Z(x_0) = \sum_{l=0}^2 \beta_l f_l(x_0) + \sum_{\alpha=1}^n \lambda_\alpha \left[Z(x_\alpha) - \sum_{l=0}^2 \beta_l f_l(x_\alpha) \right] \quad (11)$$

with the constraint

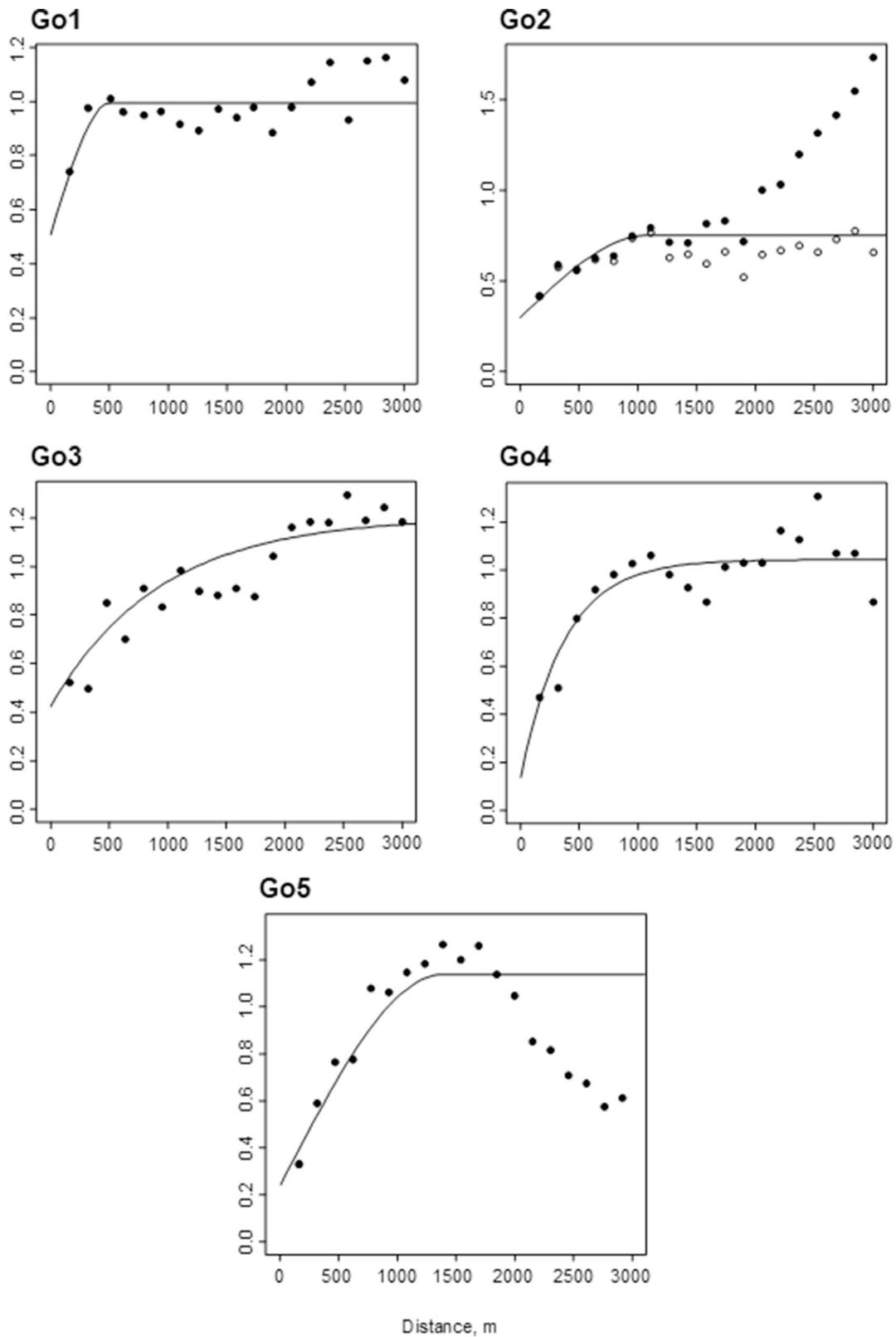


Fig. 7. Variograms of Goodall's similarity index. Data were previously transformed by Gaussian anamorphosis.

$$\sum_{\alpha=1}^n \lambda_{\alpha} f_{\alpha}(x_{\alpha}) = f_i(x_0) \tag{12}$$

where X_i are covariables; β_i is the regression coefficient associated to each covariable. The covariable X_0 has the constant value 1, whereas X_1 and X_2 are easting and northing coordinates, respectively.

Experimental variograms were modelled using the Restricted

Maximum Likelihood Estimation (REML) method, which yields the best parameter estimates when the sample size is less than 150 observations and/or sampling locations are unevenly distributed in space (Marchant and Lark, 2007). Since values in similarity vectors are not normally distributed, they were normalized through Gaussian anamorphosis, a mathematical function $Z = \phi(Y)$ that relates a raw variable Z with any

Table 6
Variogram models and parameters.

Cluster	Nugget variance	Model	Practical range, m	Sill variance
Burnaby's coefficient				
<i>Bu1</i>	0.852	Spherical	3375	0.192
<i>Bu2</i>	0.188	Exponential	1950	0.970
<i>Bu3</i>	0.106	Exponential	840	0.897
<i>Bu4</i>	0.379	Spherical	2000	0.796
<i>Bu5^a</i>	0.468	Spherical	1540	0.551
Goodall's index				
<i>Go1</i>	0.507	Spherical	545	0.488
<i>Go2^a</i>	0.299	Spherical	1175	0.408
<i>Go3</i>	0.417	Exponential	2995	0.786
<i>Go4</i>	0.137	Exponential	1204	0.907
<i>Go5^a</i>	0.250	Spherical	1500	0.812

^a Model fitted to the variogram of residuals.

distribution to a Gaussian random function Y with zero mean and unit variance by means of a transformation function ϕ that gives the correspondence between each one of the sorted raw data and the corresponding frequency quantile in the standardised Gaussian scale (Chilés and Delfiner, 1999). Transformation was done using the inverse of ϕ expanded into Hermite polynomials $H_i(Y)$ monotonically increasing within the interval defined by the minimum and the maximum of sample values (Wackernagel, 2003).

Data were interpolated at the nodes of a 5 m-mesh grid using a circular neighbourhood with a radius of 1500 m. Finally, interpolated values were back-transformed to raw values through the Gaussian anamorphosis functions.

Gaussian anamorphosis and geostatistical analysis were carried out in the R software environment (R Core Team, 2016). The former was done with the *RGeostats* package (Renard et al., 2017), the latter with the *geoR* package (Ribeiro and Diggle, 2016). Vector and raster files for maps were managed with the software *qGis* (Quantum GIS Development Team, 2016).

2.4. Validation with auxiliary information

Internal validation procedures are not applicable to Burnaby's and Goodall's similarity measures. Therefore we tested the results by comparing interpolation maps with auxiliary sources of information. The first of them was the 1:28800-scale chorography map of Istria and Dalmatia published in 1820 (Archivio di Stato di Trieste, 1998) and based on 1:2880-scale cadastral maps. The part regarding Čepić Polje is shown in Fig. 3, which reports the road network and several environmental features of the area. In our investigation, we paid particular attention to a still existing cobbled road – marked in bold in both Figs. 2 and 3 – that was located upstream of a marshy area at the mouth of River Boljunščica. We assumed that the road was built on a position rarely affected by floods and that it marked the transition between present and recent alluvia, i.e. between alluvial soils characterized by different stages of development.

The second information is related to the portions of Čepić Polje where *Tuber magnatum* truffles are collected. This subterranean ectomycorrhizal fungus is highly selective for the soil environment where it grows. In fluvial plains, it is found in lime-rich, coarse loamy to fine silty, continuously rejuvenated soils (Bragato and Marjanović, 2016). In Čepić Polje, truffle-producing sites are specifically located on natural levees of River Boljunščica, while lacking in the fine-textured soil on lacustrine sediments (Bragato et al., 2010).

3. Results

Hierarchical clustering of similarity values produced the dendrograms of Fig. 4, which were obtained with the complete linkage criterion. The evenness of permutations resulted in five most significant

clusters for both measures.

The relationship between the two classifications is summarized in the contingency table of Table 2, which was drawn up by assigning observations to the cluster with which they displayed the greatest similarity. The overall correlation between the two classifications was strong, as indicated by the resulting χ^2 value of 186. The main combinations were *Bu2-Go3*, *Bu3-Go2*, and *Bu5-Go5*, whereas *Bu4* displayed affinities both with *Go2* and *Go4*.

A first assessment of which of the two similarity measures was most effective can be made on the basis of Table 3, which reports the descriptive statistics of similarity vectors calculated by averaging similarity values within and between clusters. The clusters obtained with Goodall's index displayed higher maxima and larger ranges of variation than those produced with Burnaby's coefficient. Maxima, in particular, ranged between 0.75 and 0.96 for the former similarity measure, and between 0.55 and 0.62 for the latter. As far as the probability distribution function is concerned, a Gaussian distribution is expected for Burnaby's coefficient while the uniform distribution is characteristics of Goodall's index when the set of observations is homogeneous. In the study case, none of the clusters was skewed (Table 3), but clusters obtained with Burnaby's method showed histograms with a slightly bimodal behaviour, and those of Goodall's similarities deviated from the uniform distribution (Fig. 5), both results suggesting heterogeneity in the variance of clusters.

A more detailed analysis of the relationships between pairs of clusters was done on the basis of the within/between clusters (W/B) similarity matrices reported in Table 4. The within-cluster similarities in the matrix diagonals increased from 0.37 to 0.55 of Burnaby's coefficient to 0.60–0.96 of Goodall's index. The off-diagonal, between-cluster similarities also suggested a relationship between clusters *Bu2* and *Bu3*, *Go2* and *Go4*, *Go3* and *Go5*. The effectiveness of partitions can be better assessed with the average W/B ratio: the higher it is, the better the partition. In Table 4, this parameter equals 1.44 and 1.86 for Burnaby's and Goodall's similarity measures, respectively.

The results of Tables 2 and 4 help to analyse cluster centroids in Table 5. Using the depth of RMFs as a sorting criterion related to soil drainage, clusters *Bu1* to *Bu4* can be ordered in terms of increasing drainage capacity, while *Bu5* is distinguished by its more yellow hue. Applying the same criterion to Goodall's measure, clusters can be ordered in the sequence *Go1-Go3-Go2-Go4*. Also in this case the fifth cluster – *Go5* – is characterized by a more yellow hue. It worth also noting the presence of a subsurface B horizon in *Go2* that is absent in the related cluster *Bu3*.

Differences and connections between clusters concerned not only centroids, but also the spatial variability of similarity values. After data transformation by Gaussian anamorphosis, similarity vectors generated the omnidirectional variograms of Figs. 6 and 7, which were fitted by the models summarized in Table 6. Starting from the results of Table 2, variograms can be analysed by pairs of resembling clusters. The *Bu1* and *Go1* variograms displayed an almost pure nugget behaviour. Since they indicated a random variation in space, the corresponding clusters were not considered in the interpolation step.

Distinctly different variogram shapes characterized the *Bu3-Go2* pair. While the former cluster produced a bounded variogram with a range of 840 m, the latter showed a non-stationary behaviour, with semivariances gradually increasing from a distance of 1000 m. In this case, we modelled the variogram of residuals (empty dots in Fig. 7) and applied the UK interpolation.

Also the *Bu5* and *Go5* variograms showed a non-constant mean in form of a concavity at 1400–1600 m, suggesting the presence of several similar observations in a relatively central part of the investigated area. Following the suggestions of Myers (1988), who showed that splines and UK are equivalent, also in this case we modelled and interpolated them with the UK approach.

The remaining pairs of clusters originated bounded variograms that, while allowing the application of OK interpolation, were characterized

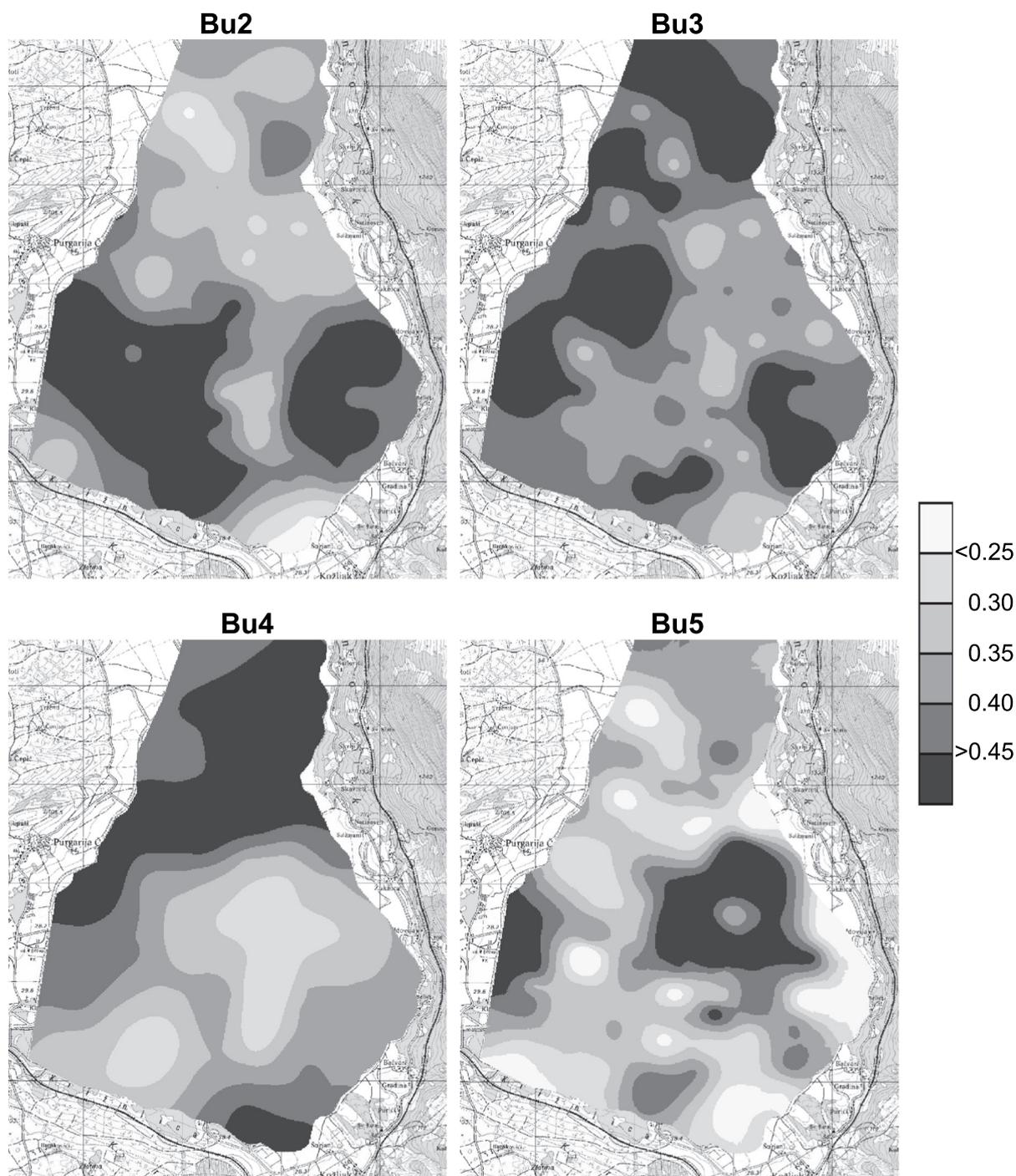


Fig. 8. Interpolation maps of Burnaby's similarity vectors.

by different values of the model parameters. The *Bu2* variogram shows a range of 2000 m and a nugget variance of 0.189 that are much lower than the values of 3000 m and 0.417 of the model fitted to cluster *Go3*. The relationship is reversed for *Bu4/Go4* pair, with the latter showing a shorter range and a smaller nugget variance than the former.

After the interpolation, the data were back-transformed to draw the similarity maps shown in Figs. 8 and 9. The legends of the figures show the different range of variation already reported in Table 3. Taking as threshold values 0.40 for Figs. 8 and 0.60 for Fig. 9, high similarity values characterize *Bu5/Go5* clusters in the central portion, and *Bu2/Go3* in the area of the polje once occupied by the lake. The remaining maps are characterized by the constant presence of high similarity values in the northern portion of the polje, with *Bu4* and *Go4* similarity

vectors showing a second small area of high values in the very south of the investigated area.

4. Discussion

According to the results of Table 2, the two similarity measures are strongly correlated. However, the average W/B ratio reported in Table 4 indicates that the clusters of Goodall's similarity values were better partitioned than those obtained with Burnaby's method. A better performance of Goodall's index is also suggested by the wider range of variation of values in the [0,1] interval, a result also related to the fully probabilistic nature of this index.

Goodall's index gives better results also in terms of soil types

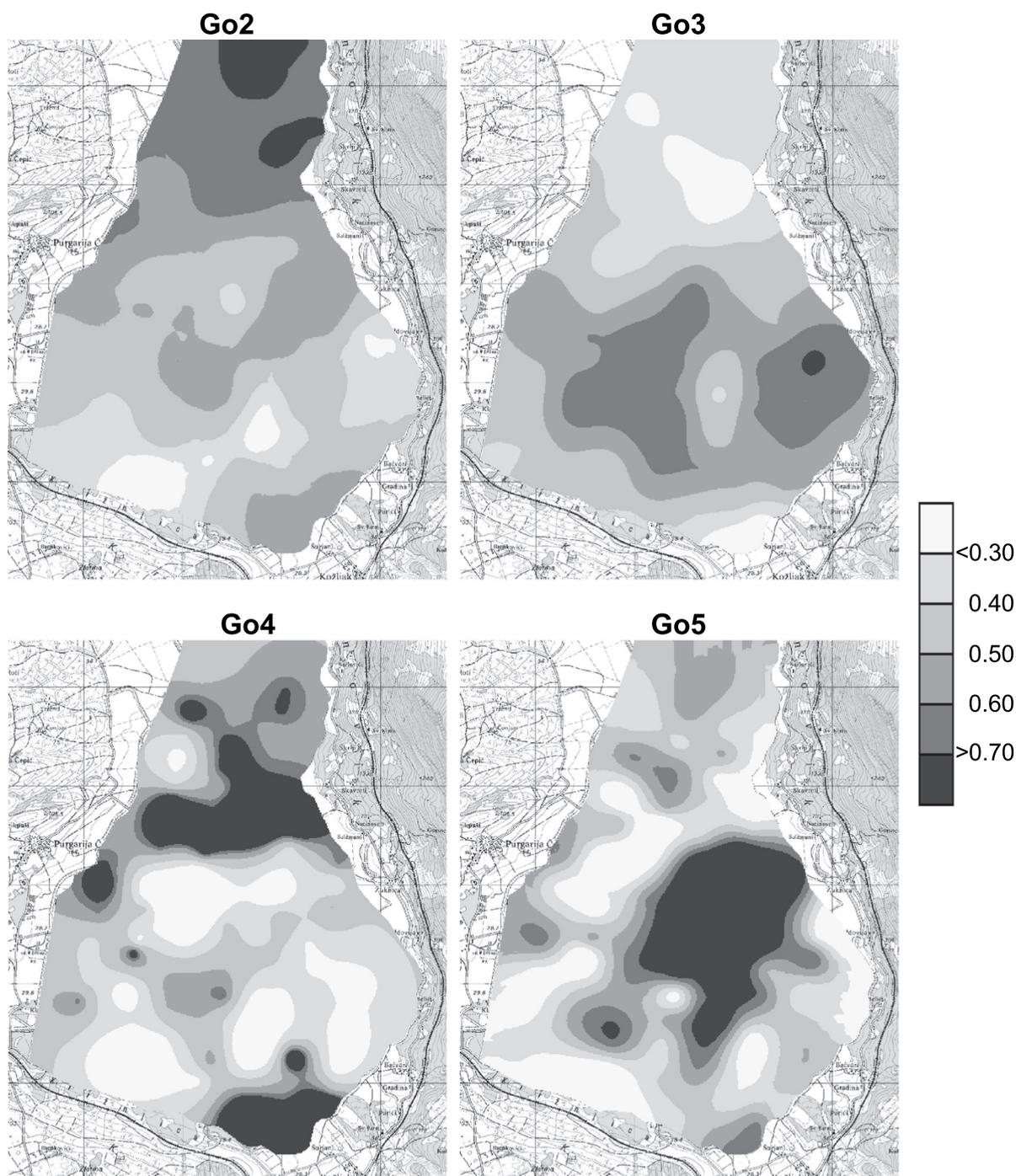


Fig. 9. Interpolation maps of Goodall's similarity vectors.

indicated by similarity vectors. The centroidal values reported in Table 5 are explainable from the standpoint of the soil transition from Fluvisols to Cambisols if we assign clusters *Go3* and *Go5* to Fluvisols, cluster *Go2* to Cambisols, and consider *Go4* as a young Cambisol rarely affected by floods. A further link to fluvial dynamics is the change of texture between *Go5* and *Go3* that sets them in the frame of a decreasing flood energy related to an increasing distance from the riverbed.

The combination of Table 5 and variogram analysis provides further evidence of the link between soil forming processes and the activity of River Boljunščica.

The linear trend shown by the *Go2* variogram can be easily explained with a gradual evolution of soils from Fluvisols of the reclaimed

area to Cambisols of the upper surfaces located in the northern part of the polje. Furthermore, the spatial correspondence between the convexity of variograms *Bu5* and *Go5*, and the slight concavity displayed by the other variograms looks connected to the halfway position of Boljunščica mouth with respect to the former lake of Čepič. A similar convexity was observed by Bragato (2004) in a crevasse splay that, after a man-made partial deviation of the River Piave in the 17th century, originated a lens of sandy Fluvisols in the middle of the present day fluvial plain.

Clusters showing higher similarities values in the central and southern part of Čepič Polje are related to the area of maximal seasonal expansion of the ancient lake in Čepič. High similarity values in clusters *Bu2* and *Go3* concern the portion of flood plain almost always covered

by water, where slow sedimentation in calm water originated fine-textured deposits. Clusters *Bu5* and *Go5* are instead related to the bottom part of River Boljunščica, and similarity values larger than 0.3 (Burnaby) and 0.8 (Goodall) concerned almost all places where *T. magnatum* truffles are collected (Bragato et al., 2010), i.e. the natural levees still affected by river floodings. Unlike cluster *Bu5*, the map of *Go5* marks the last part of the ancient riverbed that has left a shallow, elongated depression in the field. These four maps suggest a connection between high similarity values and the average extension of autumn floods that every few years occurred in Čepić Polje before its reclamation. Floods were strong enough to originate a preferential deposition of coarser particles in the vicinity of Boljunščica riverbed. This explanation is supported by the large percentage of redox depletions recorded in their centroids (Table 5), relating clusters to the presence of a hyporheic flow fed by the river, and the pattern of large-similarity values in the southern part of the map, that coincide with the Boljunščica riverbed present in 19th century chorography map of Fig. 3.

High similarity areas overlap in the transition areas between *Bu2/Go3* and *Bu5/Go5* clusters. Gradual lateral changes in soil are the norm in fluvial plains where floods of different energy originate vertical layers that are heterogeneous in the average size of mineral particles. In Čepić Polje, this process was reinforced by the periodical variation of the lake surface – typical of polje lakes – between the maximum size after autumn floods and the minimum size in summer.

The remaining maps were unrelated to the recent activity of River Boljunščica. The southern boundaries of *Go2* and, to a less extent, *Bu3* high-similarity areas cross the 19th century road already highlighted in Fig. 3. This detail proves that the northern portion of Čepić Polje has rarely been flooded in the last centuries, allowing soil to develop a Cambic B horizon. The same road crosses the area where similarity values of clusters *Bu4* and *Go4* are the highest. This area is placed between soils with a well-developed B horizon and those related to River Boljunščica and the former lake of Čepić. The results of Table 5 are again useful in explaining the relationships between clusters. The characteristics of the *Go4* centroid are for instance intermediate between those of clusters *Go2* and *Go3*, suggesting the presence of a transitional type of soil in an area where occasional floods were frequent enough to slow down the development of a B horizon. The same explanation may be applied to the high-similarity spot located to the very south of maps *Bu4* and *Go4*. This spot overlaps the fan of a small stream rooted on Eocene turbidites standing alone in the south-eastern slopes of Čepić Polje. Its upper location with respect to the water level of the ancient lake was probably responsible for the same soil development recorded along the cobbled road.

In general, the wider range of variation characterizing Goodall's similarity index allowed to draw less smoothed and more varied interpolation maps that were capable to show in more detail the effect of environmental processes in Čepić Polje.

5. Conclusions

The similarity maps of Čepić Polje displayed a soil pattern that is closer to field reality of local soil processes than that usually shown in discrete soil maps. They may help surveyors to separately consider the distribution of soil types, to assess their relationships with biodiversity and biological communities, and therefore to manage their use. The two similarity measures produced comparable information about soil characteristics and the spatial distribution of soil types in Čepić Polje, but our investigation suggests that Goodall's similarity index originates better partitions between soil observations in areas where soil forming processes are mainly related to recent and current fluvial dynamics. The spatial patterns of clusters *Go3*, *Go4* and *Go5* are, for example, very similar, but differences related to fluvial dynamics change their suitability for *T. magnatum* and should suggest the surveyor how to manage the different portions of Čepić Polje. The tested approach also helps to explain and delineate areas showing high similarities for more than one

cluster. In discrete maps the surveyor should decide whether to draw an arbitrary boundary between mapping units or create intergrade units difficult to be dealt with from the taxonomical point of view. In Čepić Polje, on the contrary, we can explain some of them as transition areas originated by spatially continuous environmental processes.

Generally speaking, the analytical sequence we investigated met three compulsory requirements to move from field observations to soil processes and to soil mapping: clusters were meaningful from the fluvial dynamics point of view; similarities displayed a structured spatial variability; similarity maps were consistent with the soil forming factors acting in the area. These results indicate that Goodall's similarity index could be currently used in soil suitability evaluation procedures, allowing to better exploit the data collected in the auger boring campaign and to extend its application to relate soil types with the ecological equilibria of biological communities. Investigations on Eocene to Pliocene outcrops suitable for other truffle species are now close to completion and they will help to validate the proposed procedure in other environmental situations of the Mediterranean area.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank the township of Kršan (HR) for the access permission to truffle locations in Čepić Polje.

References

- Archivio di Stato di Trieste, 1998. Mappe del catasto franceschino. Inventario n. 89A. (Trieste).
- Balbo, A.L., Andrić, A., Rubinić, A., Moscarillo, A., Miracle, P.T., 2006. Palaeoenvironmental and archaeological implications of a sediment core from Polje Čepić, Istria, Croatia. *Geol. Croat.* 59, 109–124.
- Bockheim, J.G., Gennadiyev, A.N., 2000. The role of soil-forming processes in the definition of taxa in Soil Taxonomy and the World Soil Reference Base. *Geoderma* 95, 53–72. [https://doi.org/10.1016/S0016-7061\(99\)00083-x](https://doi.org/10.1016/S0016-7061(99)00083-x).
- Bodin, F., Morlat, R., 2006. Characterization of viticulture terroirs using a simple field model based on soil depth. I. Validation of the water supply regime, phenology and vine vigour in the Anjou vineyard (France). *Plant Soil* 281, 37–54. <https://doi.org/10.1007/s11104-005-3768-0>.
- Bragato, G., 2004. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. *Geoderma* 118, 1–16. [https://doi.org/10.1016/S0016-7061\(03\)00166-6](https://doi.org/10.1016/S0016-7061(03)00166-6).
- Bragato, G., Marjanović, Z., 2016. Soil characteristics for tuber magnatum. In: Zambonelli, A., Iotti, M., Murat, C. (Eds.), *Truffle (Tuber spp.) in the World. Soil Ecology, Systematics and Biochemistry*. Springer, New York, pp. 191–209. https://doi.org/10.1007/978-3-319-31436-5_12.
- Bragato, G., Vignozzi, N., Pellegrini, S., Sladonja, B., 2010. Physical characteristics of the soil environment suitable for Tuber magnatum production in fluvial landscapes. *Plant Soil* 329, 51–63. <https://doi.org/10.1007/s11104-009-0133-8>.
- Burba, N., Feoli, E., Malaroda, M., 2008. MATEDIT: A software tool to integrate information in decision making processes. In: Neves, R., Baretta, J.W., Mateus, M. (Eds.), *Perspectives on Integrated Coastal Zone Management in South America*. IST PRESS, Lisbon, pp. 123–127.
- Burnaby, T.P., 1970. On a method for character weighting a similarity coefficient, employing the concept of information. *Math. Geol.* 2, 25–38. <https://doi.org/10.1007/BF02332078>.
- Burrough, P.A., van Gaans, P.F.M., Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma* 77, 115–135. [https://doi.org/10.1016/S0016-7061\(97\)00018-9](https://doi.org/10.1016/S0016-7061(97)00018-9).
- Butler, B.E., 1980. *Soil Classification for Soil Survey*. Oxford University Press, Oxford.
- Carranza, M.L., Feoli, E., Ganis, P., 1998. Analysis of vegetation structural diversity by Burnaby's similarity index. *Plant Ecol.* 138, 77–87. <https://doi.org/10.1023/A:1009760808505>.
- Carré, F., Girard, M.C., 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. *Geoderma* 110, 241–263. [https://doi.org/10.1016/S0016-7061\(02\)00233-1](https://doi.org/10.1016/S0016-7061(02)00233-1).
- Carré, F., Jacobson, M., 2009. Numerical classification of soil profile data using distance metrics. *Geoderma* 148, 336–345. <https://doi.org/10.1016/j.geoderma.2008.11.008>.
- Chilés, J.P., Delfiner, P., 1999. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
- Costantini, E.A.C., Bucelli, P., Priori, S., 2012. Quaternary landscape history determines the soil functional characters of terroir. *Quat. Int.* 265, 63–73. <https://doi.org/10.1016/j.quaint.2011.08.021>.
- Diggle, P.J., Ribeiro Jr., P.J., 2007. *Model-based Geostatistics*. Springer, New York.
- Državni Hidrometeorološki Zavod, 2018. Monthly values for Pazin in 1961–2016 period. http://klima.hr/klima_e.php?id=mjes&Grad=pazin, Accessed date: 1 February

- 2018.
- FAO, 1976. *A Framework for Land Evaluation*. Soil Bulletin no. 32 FAO, Rome.
- Feoli, E., Orlóci, L., 2011. Can similarity theory contribute to the development of a general theory of the plant community? *Community Ecol* 12, 135–141. <https://doi.org/10.1556/ComEc.12.2011.1.16>.
- Feoli, E., Zuccarello, V., 1991. Ordination based on classification: Yet another solution?!. In: Feoli, E., Orlóci, L. (Eds.), *Computer Assisted Vegetation Analysis*. Handbook of Vegetation Science. vol. 11. Springer, Dordrecht, pp. 255–263. https://doi.org/10.1007/978-94-011-3418-7_22.
- Feoli, E., Gallizia Vuerich, L., Ganis, P., Woldu, Z., 2009. A classification approach integrating fuzzy set theory and permutation techniques for land cover analysis: a case study on a degrading area of the Rift Valley (Ethiopia). *Community Ecol.* 10, 53–64. <https://doi.org/10.1556/ComEc.10.2009.1.7>.
- Feoli, E., Pérez-Gomez, R., Oyonarte, C., Ibanez, J.J., 2017. Using spatial data mining to analyze area-diversity patterns among soil, vegetation, and climate: a case study from Almería, Spain. *Geoderma* 287, 164–169. <https://doi.org/10.1016/j.geoderma.2016.09.011>.
- Fischer, H.R., 1948. *Statistical Methods for Research Workers*, tenth ed. Oliver and Boyd, Edinburgh.
- Goodall, D.W., 1966. A new similarity index based on probability. *Biometrics* 22, 882–907.
- Goodall, D.W., Feoli, E., 1991. Application of probabilistic methods in the analysis of phytosociological data. In: Feoli, E., Orlóci, L. (Eds.), *Computer Assisted Vegetation Analysis*. Handbook of Vegetation Science. vol. 11. Springer, Dordrecht, pp. 137–146. https://doi.org/10.1007/978-94-011-3418-7_13.
- Goodall, D.W., Ganis, P., Feoli, E., 1991. Probabilistic methods in classification: A manual for seven computer programs. In: Feoli, E., Orlóci, L. (Eds.), *Computer Assisted Vegetation Analysis*. Handbook of Vegetation Science. vol. 11. Springer, Dordrecht, pp. 453–467. https://doi.org/10.1007/978-94-011-3418-7_40.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Gower, J.C., 1970. A note on Burnaby's character-weighted similarity coefficient. *Math. Geol.* 2, 39–45. <https://doi.org/10.1007/BF02332079>.
- Hoosbeek, M.R., Amundson, R.G., Bryant, R.B., 1999. Pedological Modeling. In: Sumner, M.E. (Ed.), *Handbook of Soil Science*. CRC Press, Boca Raton, pp. E77–E116.
- IUSS Working Group WRB, 2014. World reference base for soil resources 2014. In: *International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*. World Soil Resources Reports No. 106 FAO, Rome.
- Journal, A., Huijbregts, C., 1978. *Mining Geostatistics*. Academic Press, London.
- Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M., de Vries, F., 2012. Efficiency comparison of conventional and digital soil mapping for updating soil maps. *Soil Sci. Soc. Am. J.* 76, 2097–2115. <https://doi.org/10.2136/sssaj2011.0424>.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* 15, 259–263. <https://doi.org/10.1127/0941-2948/2006/0130>.
- Lancaster, H.O., 1949. The derivation and partition of χ^2 in certain discrete distributions. *Biometrika* 36, 117–129. <https://doi.org/10.1093/biomet/36.1-2.117>.
- Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.
- Mantel, N., Valand, R.S., 1970. A technique of nonparametric multivariate analysis. *Biometrics* 26, 547–558. <https://doi.org/10.2307/2529108>.
- Marchant, B.P., Lark, R.M., 2007. Robust estimation of the variogram by residual maximum likelihood. *Geoderma* 140, 62–72. <https://doi.org/10.1016/j.geoderma.2007.03.005>.
- Marsili-Libelli, S., 1991. Fuzzy clustering of ecological data. In: Feoli, E., Orlóci, L. (Eds.), *Computer Assisted Vegetation Analysis*. Handbook of Vegetation Science. vol. 11. Springer, Dordrecht, pp. 173–184. https://doi.org/10.1007/978-94-011-3418-7_15.
- McBratney, A.B., de Gruijter, J.J., 1992. A continuum approach to soil classification and mapping: classification by modified fuzzy k-means with extragrades. *J. Soil Sci.* 43, 159–175. <https://doi.org/10.1111/j.1365-2389.1992.tb00127.x>.
- McBratney, A.B., de Gruijter, J.J., Brus, D.J., 1992. Spatial prediction and mapping of continuous soil classes. *Geoderma* 54, 39–64. [https://doi.org/10.1016/0016-7061\(92\)90097-Q](https://doi.org/10.1016/0016-7061(92)90097-Q).
- Myers, D.E., 1988. Interpolation with positive definite functions. *Sci. Terre* 28, 251–265.
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Soil pattern recognition with fuzzy c-means: application to classification and soil-landform interrelationship. *Soil Sci. Soc. Am. J.* 56, 505–516. <https://doi.org/10.2136/sssaj1992.03615995005600020027x>.
- Odgers, N.P., McBratney, A.B., Minasny, B., 2011a. Bottom-up digital soil mapping. II. Soil series classes. *Geoderma* 163, 30–37. <https://doi.org/10.1016/j.geoderma.2011.03.013>.
- Odgers, N.P., McBratney, A.B., Minasny, B., 2011b. Bottom-up digital soil mapping. I. Soil layer classes. *Geoderma* 163, 38–44. <https://doi.org/10.1016/j.geoderma.2011.03.014>.
- Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *Catena* 113, 56–69. <https://doi.org/10.1016/j.catena.2013.09.006>.
- Orlóci, L., 1978. *Multivariate Analysis in Vegetation Research*, second ed. Junk, The Hague.
- Podani, J., 2000. *Introduction to the Exploration of Multivariate Biological Data*. Backhuys Publishers, Leiden.
- Quantum GIS Development Team, 2016. Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org/>, Accessed date: 25 June 2018.
- R Core Team, 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL. <https://www.R-project.org/>, Accessed date: 15 March 2018.
- Renard, D., Bez, N., Desassis, N., Beucher, H., Ors, F., Freulon, X., 2017. RGeostats: The Geostatistical R Package. Version:11.1.2. MINES ParisTech/ARMINES, Paris.
- Ribeiro Jr., P.J., Diggle, P.J., 2016. geoR: Analysis of Geostatistical Data. R Package Version 1.7–5.2.
- Škorić, A., Mayer, B., Vranković, A., Bašić, F., 1987. *Pedološka karta Istre*. Institute for Pedology at the Faculty of Agriculture. University of Zagreb, Zagreb.
- Soil Science Division Staff, 2017. *Soil survey manual*. In: Ditzler, C., Scheffe, K., Monger, H.C. (Eds.), *USDA Handbook 18*. Government Printing Office, Washington, DC.
- Triantafyllis, J., Ward, W.T., Odeh, I.O.A., McBratney, A.B., 2001. Creation and interpolation of continuous soil layer classes in the lower Namoi valley. *Soil Sci. Soc. Am. J.* 65, 403–413. <https://doi.org/10.2136/sssaj2001.652403x>.
- Vaudour, E., 2003. *Les Terroirs Viticoles*. Dunod, Paris.
- Vaudour, E., Costantini, E., Jones, G.V., Mocali, S., 2015. An overview of the recent approaches to terroir functional modelling, footprinting and zoning. *Soil* 1, 287–312. <https://doi.org/10.5194/soil-1-287-2015>.
- Verheyen, K., Dries, A., Hermly, M., Deckers, S., 2001. High-resolution continuous soil classification using morphological soil profile descriptions. *Geoderma* 101, 1–48. [https://doi.org/10.1016/S0016-7061\(00\)00088-4](https://doi.org/10.1016/S0016-7061(00)00088-4).
- Wackernagel, H., 2003. *Multivariate Geostatistics: An Introduction with Applications*, third ed. Springer-Verlag, Berlin.
- Whitney, M., 1909. *Soils of the United States*. USDA Bur. of Soils Bull, vol. 55 U.S. Gov. Print. Office, Washington, DC.
- Zhao, S.X., 1986. Discussion on fuzzy clustering. In: *Eighth Int. Conf. On Pattern Recognition*. IEEE Press, New York, pp. 612–614.