S. Herrando-Pérez, M. Baratti, and G. Messana – Subterranean ecological research and multivariate statistics: a review (1945–2006). *Journal of Cave and Karst Studies,* v. 70, no. 2, p. 120–128.

SUBTERRANEAN ECOLOGICAL RESEARCH AND MULTIVARIATE STATISTICS: A REVIEW (1945–2006)

Salvador Herrando-Pérez[1,*], Mariella Baratti[2], and Giuseppe Messana[2]

Abstract: Subterranean ecosystem studies using multivariate ordination and/or agglomerative classification statistical methods were reviewed in the Science Citation Index (SCI) between 1945 and 2006. Nearly 57,000 publications cited subterranean habitats or their associated biota in the SCI abstracts, however, multivariate statistics applied to strictly hypogean taxa occurred in only 65 papers from 1990 onwards. Over 90% of the multivariate applications were devoted to morphometric or genetic studies of single species and to relationships between the environment and species assemblages. In terms of taxa and ecosystem types, stygobite and waterless cave studies featuring multivariate applications predominated, respectively. Only six different methods (Agglomerative Clustering, Canonical Correspondence Analysis, Correspondence Analysis, Discriminant Analysis, non-metric Multidimensional Scaling, Principal Component Analysis) were used among the >30 multivariate techniques available within the biostatistical toolbox. The retrieved set of publications was sorted in a simple table by keyword according to type of biota, habitat, research topic and multivariate method, while online biostatistical resources are appended. Further comments are made on the use of statistics in the biological sciences in general.

Introduction

Multivariate statistics (i.e. ordination, classification and hybrid methods) are useful in studies where one of the main goals may be to quantify relationships among a set of samples in which a large number of variables has been measured. A matrix of abundance of species assemblages and a matrix of measurements of environmental variables from an array of samples spread in space and/or time is a classical data set to deploy multivariate statistics in ecological research. For example, samples may be different caves surveyed in a single region, a single cave monitored several times per year, or a combination of the latter, while the relative abundances of troglobite taxa found in one baited trap per cave and physical variables measured around each trap (e.g. humidity, temperature, etc.) may constitute the raw data. Common research questions asked of such matrices are whether there are patterns of species assemblages among caves and what environmental factors are underlying those patterns (*e.g.* Christman and Culver, 2001).

Multivariate statistical methods originated from the pioneering work by Petrie (1899) and Spearman (1904). Since then, over thirty different multivariate techniques have been described (see Gauch, 1982; Jongman et al., 1995; Legendre and Legendre, 1998; Clarke and Warwick, 2001; McCune and Grace, 2002; Leps and Smilauer, 2003; Zuur et al., 2007), mostly within the last two decades. These techniques have a number of advantages over more traditional, univariate techniques, notably in their graphical and numerical outputs. In ordination methods, complex multivariate information is summarized in a low-dimensional scatter diagram where points commonly represent samples and distances among them are proportional to their degree of likeliness measured by a given resemblance measure, e.g. Euclidean distances or Bray-Curtis similarities (see Clarke et al., 2006). As to classification techniques (Legendre and Legendre, 1998), dendrograms or trees are the display options, where branches link and cluster groups of sample units in a sequential fashion based on resemblance. Second, as part of the numerical result of a multivariate analysis, a series of coefficients are produced, which quantify the importance of each of the variables in the ordination patterns observed; hence principal coefficients in eigen-analysis-based techniques like Principal Component Analysis (PCA; Hotelling, 1933) or Correspondence Analysis (CA; Hirschfeld, 1935; Hill and Gauch, 1980; ter Braak and Schaffers, 2004), or similarity contributions (Clarke, 1993) from distance-based methods like non-Metric Multi-Dimensional Scaling (nMDS; Shepard, 1962; Belbin, 1991), Cluster Analysis (CLUSTER; Sneath and Sokal, 1973) or Polar Ordination (Bray and Curtis, 1957).

Ordination techniques may be indirect (unconstrained) or direct (constrained, also known as canonical) depending on whether one or two data matrices, respectively, are input into the algorithm. The whole family of ordination methods is sometimes termed as gradient analysis following ter Braak and Prentice (1988). Within the suite of canonical methods, ordinations can be constrained to extract a

* Present address: School of Earth & Environmental Sciences, University of Adelaide, SA 5005, Australia (Salvador.herrando-Perez@adelaide.edu.au)

[1] BIOESTUDIOS SAGANTA, c/ San Pedro 6, 12200 Onda - Spain

[2] Institute for the Study of Ecosystems (ISE), Via Madonna del Piano, 10, 50019 Sesto Fiorentino (Fi) – Italy

specific fraction of information from a species x sample data matrix, for instance, the amount of assemblage variation that can be explained by a set of environmental predictors (one type of constraint, i.e., the second data matrix). This is the case in Redundancy Analysis (RDA; de Wollenberg, 1977), Canonical Correspondence Analysis (CCA; ter Braak, 1986; Makarenkof & Legendre, 2004) and Canonical Correlation Analysis (CCoA; Hotelling, 1936; Anderson & Willis, 2005). On the other hand, ANOVA-like tests for differences among a priori groups of samples (a further type of constraint) in Discriminant Analysis (DA; Fisher, 1936; Legendre & Anderson, 1999; McArdle & Anderson, 2001; Anderson and Robinson, 2005; Millar et al., 2005) represent the second major kind of canonical analyses. Ordination methods have been employed traditionally in a descriptive fashion, but recent advances have fostered ordination-based frameworks to carry out MANOVA inferential tests (Legendre and Anderson, 1999; McArdle and Anderson, 1999) and predictive modeling (e.g. Guisan et al., 1999). Comparative reviews over the properties and relative merits and drawbacks of the various techniques can be found in Minchin (1987), Clarke and Green (1988), Økland (1996) and Anderson and Willis (2003). Moreover, advanced classification methods can also quantify relationships among two sets of variables in the so-called Multivariate Regression Trees (Breiman et al., 1984; Moore et al., 1991; Vayssieres et al., 2000; De'ath, 2002), or integrate classification and ordination techniques into two-way ordered tables known as TWINSPAN (Hill, 1979; Gauch, 1982) which have been particularly popular among botanists.

Although multivariate techniques are now firmly established in many fields of ecology (particularly so for marine benthos, freshwater, and botany), their use seems restricted in subterranean ecosystem studies. To quantify this apparent lack of implementation, we describe trends in the utilization of multivariate ordination methods and agglomerative classification for the study of subterranean fauna and microbes in ecological research through a literature review in the Science Citation Index (1945–2006). Records of genetic studies (which do not subscribe strictly to ecology) and of submarine cave papers (which do not deal with subterranean ecosystems proper) were also included as sample examples of other type of multivariate applications from a common raw data entry. The main fields of utilization are highlighted, a synthesis of on-line resources for multivariate statistics is provided, and the full set of selected publications (n=65) is appended after sorting them by type of biota, habitat, research topic and multivariate method.

Methods

The literature review was undertaken in the Science Citation Index (SCI, 1945–2006, inclusive) on-line at the Department of Analytical Chemistry, University Jaume I (Castellón de la Plana, Spain). A total of 358 key words were browsed in the database in three compound sets linked with the standard commands AND, OR, " " and * as appropriate. These sets consisted of groups of words that could retrieve from SCI any publication incorporating multivariate methods (117 words), subterranean biota (69 words) and subterranean habitats (172 words). Subterranean habitats included caves, the soil vadose and phreatic zones, lava tubes, cenotes and blue holes, while nomenclature for subterranean biota followed Malard et al. (2003). Searches were done by abstract, after first undertaking a trial by title which only found three publications. The review was restricted to research articles published in English.

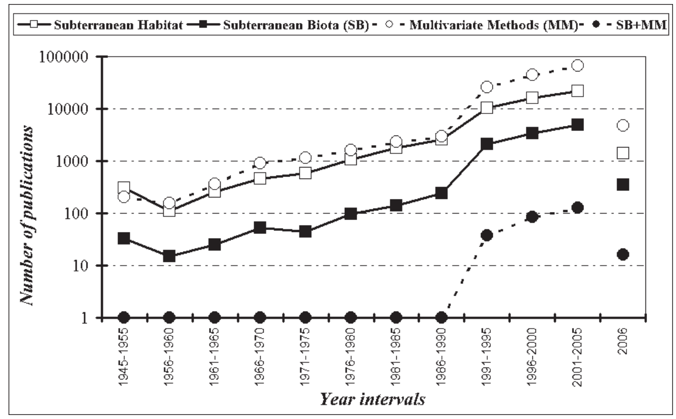

Figure 1. Number of publications (in a log-scale) cited in the *Science Citation Index* (1945–2006) after a search by abstract including key words standing for Subterranean Habitat, Subterranean Biota (SB), Multivariate ordination and classification Methods (MM), and SB+MM.

Results

Throughout the period of 61 years surveyed, a total of 56,559 publications cited subterranean ecosystems or their associated biota in their abstracts. Of this total, approximately 20% and 80% dealt with subterranean biota and habitats, respectively, and only 266 papers (0.47%) had used multivariate ordination and/or classification methods. The number of publications on subterranean biota and habitats, whether multivariate statistics were included or not, increased with time, particularly from 1985 to 1995 (Fig. 1). The first paper to use multivariate methods in SCI-cited subterranean research was published in 1990.

Many of the papers extracted from the SCI were devoted to flora and fauna that do not inhabit the subterranean realm, for instance surface vegetation in karstic regions or aquatic invertebrates in rivers receiving water from a subterranean course. As such, all abstracts were manually scrutinized. Out of the 266 publications using multivariate methods, only 65 papers were specifically related to hypogean ecosystems and they form the literature compilation used hereafter (see Appendix).

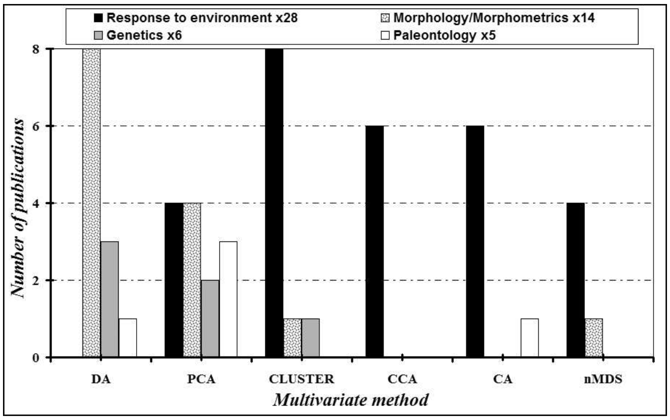

Figure 2. Number of publications employing a total of 6 multivariate methods in research papers on subterranean biota. Publication records (n=65) were obtained from the *Science Citation Index* (1945–2006) after a search by abstract including key words standing for the subterranean biota. See legend on top with overall subjects times the number of publications/subject (records without citation of multivariate methods in abstracts are excluded). Methods are: DA= Discriminant Analysis PCA= Principal Component Analysis CLUSTER= Agglomerative Clustering CCA= Canonical Correspondence Analysis CA= Correspondence Analysis nMDS= non-Metric Multi-Dimensional Scaling (nMDS).

Approximately two thirds (41 papers) of the multivariate applications focused on hypogean species assemblages, mainly dealing with invertebrates (hyporheos and freshwater stygobites; 15 papers) and with bacteria from aquifers and hyporheic environments (12 papers). On the other hand, the majority of population studies (i.e. addressing a single species) were carried out in waterless caves where the target fauna were troglobite arthropods (insects and arachnids, eight papers) and trogophile vertebrates (bats and rodents, 10 papers). There were no multivariate records for cenotes, lava tubes or blue holes.

Subject-wise, studies using multivariate methods could be allocated to four categories (Fig. 2). A total of 46% (31 papers) of the multivariate applications centered upon the relationship between species assemblages and environmental variables, 27 papers were on morphology/morphometrics or genetics, or a combination of both (three papers), of individual species, whereas seven papers were published in the area of paleontology, occasionally focusing on fossil hominids (two papers). A total of six different multivariate methods were recorded in this review, namely DA, PCA, CLUSTER, CCA, CA and nMDS, in that order of popularity (Fig. 2), whereas in 11 abstracts the methods used were not mentioned. All the multivariate methods except DA were employed in describing environment-community relationships, while DA was utilized in half of the morphological and genetic studies on single species.

DISCUSSION

Classification and ordination multivariate statistics have only been employed in the study of species assemblages and population structure of subterranean biota from the very beginning of the 1990's onwards. Both the total number of papers published and those that report multivariate methods follow an exponential increase, a popular trend across scientific literature often named after Price's model (Fernández-Cano et al., 2004). The assimilation of multivariate statistics in ecological subterranean research responds to a two-fold development of this field that basically enhanced the collection of quantitative data amenable to these chiefly quantitative or semiquantitative statistics. First, major subterranean research teams were established in Europe and the USA in the 1980s (Gibert et al., 1994) and this doubtless brought about an improved development of more rigorous sampling techniques, as well as a need to explore different avenues of research. Secondly, in the last fifteen years, the explosion of molecular analyses supplied an enormous amount of quantitative information that was central to progressing the systematics of subterranean fauna (Culver, 1982; Sbordoni et al., 2000; Finston et al., 2007).

The application of ordination/classification methods is larger for taxa that are most conspicuous and/or frequent in subterranean habitats, basically insects and arachnids in waterless caves, and crustaceans in hyporheos and freshwater caves. Up to two thirds of the multivariate cases were applied to data from aquatic species which are generally easier to sample quantitatively than their terrestrial counterparts. Moreover, assemblages of aquatic taxa, both invertebrates and bacteria, were often described by means of ordination methods whereby indicator taxa were identified in order to assess water quality of aquifers and of phreatic and vadose environments. The six multivariate methods reported in this literature review are among the classical ones in ecological research. In fact, CLUSTER, DA and PCA totalled just 50% of the case studies featuring multivariate applications. Similar trends were found by James and McCulloch (1990). The preferred use of DA in genetic and morphometric studies, where differences in subsets of individuals from populations and communities are looked for, clearly takes advantage of the ANOVA-like approach of this method.

However, despite the power of multivariate methods to analyze a wide variety of data in biodiversity studies, the intensity of their use can be regarded as low for research into subterranean biota, with an average of only four publications per year surveyed in SCI. Furthermore, less than 20% of the total number of the different methods available seems to have been employed, with no records of those techniques described from the middle of the 1990s onwards (see Introduction). Such a meagre utilization may be for different reasons. On one side, despite there being ample literature of quantitative studies focusing on hypogean biota, the relatively low diversity, often mainly recorded at small spatial scales, can turn into simplified data sets that may be deemed unsuitable for multivariate analysis. This can be exacerbated by the fact that many subterranean habitats are fragmented, so causing confined distributions of hypogean fauna (hence simple species assemblages) with distinct taxa not overlapping but occurring at sites geographically very close (Marmonier et al., 1993; Sket, 1999; Gibert and Deharveng, 2002). On the other side, survey efforts for unravelling diversity patterns of subterranean biota are largely concentrated in a group of localities in North America, Europe and Australia. Proof of this resides in the fact that total diversity of cave-dwelling fauna at global, landscape, regional and even local scales is largely ignored (Culver and Holsinger, 1992; Culver and Sket, 2001). Thus, less research investment necessarily entails less publication output, regardless the nature of the analytical approaches carried out.

Cave and, principally, ground water ecosystems are severely endangered in many parts of the world, and precise management alternatives are urgently required (Danielopol et al., 2003) whereby subterranean biodiversity can represent a key indicator of ecosystem health (Sket, 1999). However, there is a general absence of keystone or flag hypogean species, unlike other terrestrial ecosystems where the distribution of species and assemblages of commonly encountered vertebrates have proven useful for management and conservation practices, including identification of priority areas and regional patterns of subterranean biodiversity. Despite considerable effort being invested into describing broad biodiversity patterns in Europe (Sket, 1999; Ferreira et al., 2003) and North America (Peck, 1998; Christman and Culver, 2001), the paucity of quantitative data elsewhere means that data are reported in the form of species lists and graphical representation of distributions is often poor, which limit their impact on managers and decision-making groups. On the premise that ordinations are good communication tools (ter Braak, 1994), bio-regionalization of subterranean habitats should benefit from ordination and classification of sites for which abundance, or presence-absence, data of hypogean taxa have been measured. This practice should be stimulated in the future.

Statistics and Biology: Complementary?

Overall, a number of academic, sometimes practical, aspects hamper the access of biologists to multivariate statistics. Some multivariate methods like PCA, and its canonical analogous RDA, were described outside the realm of biology, and their adoption by biologists lagged behind for years. Others are relatively new, so they have not yet joined up to most textbooks and are mainly taught in (generally very expensive) specific training courses. Furthermore, those multivariate methods that resort to ordination solutions by permutations, like nMDS (Kruskal, 1964), together with extremely complex algorithms, like CCA (ter Braak, 1986), are computer-dependant and could not be widely implemented until the 1970's when the second-generation supercomputers became openly available to the scientific community. Finally, from their initial graduate training and thereafter, biologists are reluctant to get into complex mathematics, which are too often taught by non-biologists as an isolated subject not integrated in the many-fold disciplines of biology and unrelated to lab and field studies (see Johnson et al., 2001 for recommendations to achieve sound statistical training and background in environmental research). Additionally, statistical instruction and utilization may sometimes be misleading. Thus, many authors have documented an ill-use of statistics in ecology and wildlife research regarding their application or the reporting and interpretation of results (James and McCulloch, 1990; Johnson et al., 1999; Stephens et al., 2005). Sometimes mistakes may linger on unnoticed for both naïve and experienced statistical users and conceptual confusion prevails over the years, even for well-established concepts. For instance, Hubbard and Bayarri (2003) made a poignant observation on the pervasive confusion between p confidence values and α error rates in hypothesis testing, yet it is striking that Ellison (2006) confused both concepts, (ironically) after citing Hubbard and Bayarri (2003), in her high-ranked journal review on Bayesian inference. Back to multivariate analysis, McCune and Grace (2002) were unable to cite appropriate or adequately justified applications of CA in the literature of community ecology, while James and McCulloch (1990) "could not find a single application of multiple regression to recommend as a good example" after reviewing in the highly ranked journals of ecology and systematics.

In summary, the statistical tools are widely accessible, but they are often misused, badly-reported and (not surprisingly!) generally feared by biologists. See Table 1 for a synthetic list of online resources for multivariate statistics covering mailing lists, free and major commercial software, biostatistical experts' web pages, training courses, and research journals within the fields of biology and ecology. An illustrative starting point for those interested in improving the standards of statistical reporting should read the papers by David Anderson, Douglas Johnson and collaborators (see Anderson et al., 2001, Johnson et al., 2002, and references therein) for inferential tests, and by James and McCulloch (1990) and Paukert and Wittig (2002) for multivariate statistics. Lastly, universities hosting programs on environmental sciences should be encouraged to improve (multivariate) statistical training by tailoring it to the needs and disciplines of students. By facing real world problems (instead of only card games, lottery probabilities and the like) biologists' attitudes towards statistics may improve, while their future academic and professional profiles would be enhanced further by attaining an early environment-focused statistical background.

Table 1. On-line resources addressing multivariate statistical methods in environmental sciences.

Link	Brief Description
http://www.okstate.edu/artsci/botany/ordinate/	A bible of ordination methods by Mike Palmer at Oklahoma State University (USA).
http://www.okstate.edu/artsci/botany/ordinate/ordnews.htm	Mailing list on ordination methods, though a wide range of statistical topics is discussed.
http://www.classification-society.org	International Society of Classification, including links with branches world-wide.
http://www.classification-society.org/csna/lists.html#class-l	Mailing list on classification methods.
http://cc.oulu.fi/~jarioksa/	Personal page by Jari Oksanen at Oulu University (Finland). Free software available.
http://www.stat.auckland.ac.nz/~mja	Personal page by Marti Anderson at Auckland University (New Zealand). Free software available.
http://www.bio.umontreal.ca/legendre/	Personal page by Pierre Legendre at Montreal University (Canada). Free software available.
http://www.r-project.org/	Free software environment for statistical computing and graphics.
http://www.highstat.com	Training courses and BRODGAR software by Alain Zuur (Scotland).
http://regent.bf.jcu.cz	Training courses and CANOCO software by Jan Leps and Petr Smilauer (Check Republic).
http://home.centurytel.net/~mjm/	Training courses and PCOrd software by MjM Software Design (USA).
http://www.primer-e.com	Training courses and PRIMER software by Bob Clarke (UK).
http://aerg.canberra.edu.au/envirostats/	On-line training courses at Canberra University (Australia).
http://www.statistics.com/	On-line training courses (USA). Not specifically environmental.
http://biomet.oxfordjournals.org/	Biometrika (Research journal, Oxford University Press).
http://www.springer.com/life+sci/ecology/journal/10651	Environmental and Ecological Statistics (Research journal, Springer).
http://www.elsevier.com/wps/find/journaldescription.cws_home/622892/description#description	Environmetrics (Research journal, Wiley Inter-Science).
http://www.springer.com/statistics/statistical+theory+and+methods/journal/357	Journal of Classification (Research journal, Springer).
http://www.sciencedirect.com/science/journal/0047259X	Journal of Multivariate Analysis (Research journal, Elsevier).

ACKNOWLEDGMENTS

This paper was a contribution to the 1st Congress of Speleology of the Valencian Community (Spain) held in Alcoy in April–May 2006. Thanks go to the GECEN and the Conselleria de Territorio y Vivienda which supported my participation (SH-P) in the latter event. We particularly thank Professor Félix Hernández, Head of the Department of Analytical Chemistry, for offering so generously the use of the bibliographic resources at the Universitat Jaume I. We are especially grateful to Chris Glasby, John Holsinger, Sammy de Grave and Sam Panno for a thorough revision that notably improved the final versions of this paper. A grant from the Ayuntamiento de Onda (Spain, SH-P) supported the completion of the manuscript.

REFERENCES

Anderson, M.J., and Robinson, J., 2003, Generalized discriminant analysis based on distances: Australian and New Zealand Journal of Statistics, v. 45, p. 301–318.

Anderson, M.J., and Willis, T.J., 2003, Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology: Ecology, v. 84, p. 511–525.

Belbin, L., 1991, Semi-strong Hybrid Scaling, a new ordination algorithm: Journal of Vegetation Science, v. 2, p. 491–496.

Bray, J.R., and Curtis, J.T., 1957, An ordination of the upland forest communities in southern Wisconsin: Ecological Monographs, v. 27, p. 325–349.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, Classification and Regression Trees, Chapman and Hall, New York, USA.

Christman, M.C., and Culver, D.C., 2001, The relationship between cave biodiversity and available habitat: Journal of Biogeography, v. 28, p. 367–380.

Clarke, K.R., 1993, Non-parametric multivariate analyses of changes in community structure: Australian Journal of Ecology, v. 18, p. 117–143.

Clarke, K.R., and Green, R.H., 1988, Statistical design and analysis for a "biological effects" study: Marine Ecology Progress Series, v. 46, p. 213–226.

Clarke, K.R., and Warwick, R.M., 2001, Change in marine communities – An approach to statistical analysis and interpretation: 2nd Edition, Primer-E, Plymouth, UK, 176 p.

Clarke, K.R., Somerfield, P.J., and Chapman, M.G., 2006, On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded

assemblages: Journal of Experimental Marine Biology and Ecology, v. 330, p. 55–80.
Culver, D.C., 1982, Cave life: evolution and ecology, Harvard University Press, Cambridge, 189 p.
Culver, D.C., and Holsinger, J.R., 1992, How many species of troblobites are there?: NSS Bulletin, v. 54, p. 79–80.
Culver, D.C., and Sket, B., 2000, Hotspots of subterranean biodiversity in caves and wells: Journal of Cave and Karst Studies, v. 62, p. 11–17.
Danielopol, D.L., Griebler, C., Gunatilaka, A., and Notenboom, J., 2003, Present state and future prospects for groundwater ecosystems: Environmental Conservation, v. 30(2), p. 104–130.
Death, G., 2002, Multivariate Regression Trees: a new technique for modeling species-evironment relationships: Ecology, v. 83, p. 1105–1117.
De Wollenberg, A.L., 1977, Redundancy Analysis – an alternative for Canonical Correlation Analysis: Psychometrika, v. 42, p. 207–219.
Ellison, A.M., 2004, Bayesian inference in ecology: Ecology letters, v. 7, p. 509–520.
Fernández-Cano, A., Torralbo, M., and Vallejo, M., 2004, Reconsidering Price's model of scientific growth: an overview: Scientometrics, v. 61 (3), p. 301–321.
Ferreira, D., Dole-Olivier, M.J., Malard, F., Deharveng, L., and Gibert, J., 2003, Faune aquatique souterraine de France: base de données et elements of de biogeography: Karstologia, v. 42(2), p. 15–22.
Finston, T.L., Johnson, M.S., Humphreys, W.F., Eberhard, S.M., and Halse, S.A., 2007, Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape: Molecular Ecology, v. 16(2), p. 355–65.
Fisher, R.A., 1936, The use of multiple measurements in taxonomic problems: Annuals of Eugenics, v. 7, p. 179–188.
Gauch, H.G., 1982, Multivariate analysis in community ecology, New York, Cambridge University Press, 298 p.
Gibert, J., Danielopol, D., and Stanford, J., 1994, Groundwater Ecology, Academic Press, San Diego, USA, 571 p.
Gibert, J., and Deharveng, L., 2002, Subterranean ecosystems: a truncated functional biodiversity: BioScience, v. 52, p. 473–481.
Guisan, A., Weiss, S.B., and Weiss, A.D., 1999, GLM versus CCA spatial modelling of plant species distribution: Plant Ecology, v. 143, p. 107–122.
Hill, M.O., 1979, TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes, Ithaca, New York, Ecology and Systematics, Cornell University.
Hill, M.O., and Gauch, H.G., 1980, Detrended correspondence analysis, an improved ordination technique: Vegetation, v. 42, p. 47–58.
Hirschfeld, H.O., 1935, A connection between correlation and contingency: Proceedings of the Cambridge Philosofical Society, v. 31, p. 520–524.
Hotelling, H., 1933, Analysis of a complex of statistical variables into principal components: Journal of Educational Psychology, v. 24, p. 417–441.
Hotelling, H., 1936, Relations between two sets of variates: Biometrika, v. 28, p. 321–377.
Hubbard, R., and Bayarri, M.J., 2003, Confusion over measures of evidence (p's) versus error (α's) in classical statistical testing (with discussion): American Statistician, v. 57, p. 171–182.
James, F., and McCulloch, C.E., 1990, Multivariate analysis in ecology and systematics: panacea or pandora's box? Annual Review of Ecology and Systematics, v. 21, p. 129–166.
Johnson, D.H., 2002, The role of hypothesis testing in wildlife science: Journal of Wildlife Management, v. 63(2), p. 272–276.
Johnson, D.H., 1999, The insignificance of statistical significance testing: Journal of Wildlife Management, v. 63(3), p. 763–772.
Johnson, D.H., Shaffer, T.L., and Newton, W.E., 2001, Statistics for wildlifers? How much and what kind: Wildlife Society Bulletin, v. 29, p. 1055–1060.
Jongman, R.H.G., Ter Braak, C.J.F., and van Tongeren, O.F.R., 1995, Data Analysis in Community and Landscape Ecology, Cambridge University Press, New York, USA, 299 p.
Legendre, P., and Anderson, M.J., 1999, Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments: Ecological Monographs, v. 69, p. 1–24.
Legendre, P., and Legendre, L., 1998, Numerical Ecology. Developments in environmental modeling 20: Elsevier, 853 p.
Leps, J., and Smilauer, P., 2003, Multivariate Analysis of Ecological Data using CANOCO, Cambridge University Press, 269 p.
Makarenkov, V., and Legendre, P., 2002, Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression: Ecology, v. 83, p. 1146–1161.
Malard, F., Dole-Oliver, M.J., Mathieu, J., and Stock, J.H., 2003, Sampling manual for the assessment of regional groundwater biodiversity. PASCALIS European Project: Protocols for the assessment and conservation of aquatic life in the subsurface. UMR CNRS 5023. Ecologie des Hydrosystèmes Fluviaux. Université Claude Bernard, Lyon, France, 74 p.
Marmonier, P., Vervier, P., Gibert, J., and Dole-Oliveir, M.J., 1993, Biodiversity in ground waters: Trends in Ecology and Evolution, v. 8, p. 392–395.
McArdle, B.H., and Anderson, M.J., 2001, Fitting multivariate models to community data: a comment on distance-based redundancy analysis: Ecology, v. 82, p. 290–297.
McCune, B., and Grace, J.B., 2002, Analysis of Ecological Communities, MjM Software Design, USA, 300 p.
Millar, R.B., Anderson, M.J., and Zunun, G., 2005, Fitting non-linear environmental gradients to community data: a general distance-based approach: Ecology, v. 86, p. 2245–2251.
Minchin, P.R., 1987, An evaluation of the relative robustness of techniques for ecological ordination: Vegetatio, v. 69, p. 89–107.
Moore, D., Lee, B.G., and Davey, S.M., 1991, A new method for predicting vegetation distributions using decision tree analysis in a geographic information system: Environmental Management, v. 15, p. 59–71.
Økland, R.H., 1996, Are ordination and constrained ordination alternative or complementary strategies in general ecological studies: Journal of Vegetation Science, v. 7, p. 289–292.
Paukert, C.P., and Wittig, T.A., 2002, Applications of multivariate statistical methods in fisheries: Fisheries (Bethesda), v. 27(9), p. 16–22.
Peck, S.B., 1998, A summary of diversity and distribution of the obligate cave-inhabiting fauna of the United States and Canada: Journal of Cave and Karst Studies, v. 60(1), p. 18–26.
Petrie, W.M.F., 1899, Sequences in prehistoric remains: Journal of the Anthropological Institute, v. 29, p. 295–301.
Sbordoni, V., Allegrucci, G., and Cesaroni, D., 2000, Population genetic structure, speciation and evolutionary rates in cave–dwelling organisms: *In:* Wilkens, H., Culver, D.C., and Humphreys, W.F., eds., Ecosystems of the world - 30, Subterranean Ecosystems, Elsevier, Amsterdam, p. 453–477.
Shepard, R.N., 1962, The analysis of proximities: multidimensional scaling with an unknown distance function: Psychometrika, v. 27, p. 125–139.
Sket, B., 1999, The nature of biodiversity in hypogean waters and how it is endangered: Biodiversity Conservation, v. 8, p. 1319–1338.
Sneath,, P.H.A., and Sokal, R.R., 1973, Numerical Taxonomy – The principles and practice of numerical classification, Ed. W. H. Freeman, San Francisco, USA, 573 p.
Spearman, C., 1904, "General Intelligence", objectively determined and measured: American Journal of Psychology, v. 15, p. 201–293.
Stephens, P.A., Buskirk, S.W., Hayward, G.D., and Martínez del Río, C., 2005, Information theory and hypothesis testing: a call for pluralism: Journal of Applied Ecology, v. 42, p. 4–12.
ter Braak, C.J.F., 1986, Canonical Correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis: Ecology, v. 67, p. 1167–1179.
ter Braak, C.J.F., 1994, Canonical community ordination. Part I.: Basic theory and linear methods: Ecoscience, v. 1(2), p. 127–140.
ter Braak, C.J.F., and Prentice, I.C., 1988, A theory of gradient analysis: Advances in Ecological Research, v. 18, p. 271–317.
ter Braak, C.J.F., and Schaffers, A.P., 2004, Co-correspondence analysis: a new ordination method to relate two community compositions: Ecology, v. 85, p. 834–846.
Vayssiers, M., Plant, R. and Allen-Díaz, B., 2000, Classification trees: an alternative non-parametric approach for predicting species distributions: Journal of Vegetation Science, v. 11, p. 679–694.
Zuur, A.F., Ieno, E.N., and Smith, G.M., 2007, The Analysis of Ecological Data: Springer-Verlag, 672 p.

APPENDIX

Papers using multivariate classification and ordination methods for analyzing data from subterranean biota as retrieved from the *Science Citation Index (1945–2006)*.

CLASSIFICATION (numbers correspond with the list of publications further below):

By biota:

Marine benthos (sponges)
9,30,64

Microfauna, mainly bacteria
2,6,8,11,15,21,28,32,40,42,46,53,58

Stygobite invertebrates
1,13,16,17,20,22,25,26,33,36,37,43,44,45,47,54,56,57,60,61,63

Troglobite/Troglophile invertebrates
4,5,10,14,18,23,38,49,62

Troglophile/Trogloxene vertebrates
3,12,19,24,27,29,31,34,35,39,48,50,51,52,55,59,65

Vegetation
7,41

By habitat:

Subterranean waters, mainly aquifers
2,8,15,28,32,40,42,46

Hyporheic environment
3,18,21,25,27,28,32,41,44,48,52,54,56–59

Freshwater caves
1,20,35,36,37,45,61,63

Marine caves
3,9,17,30,64

Waterless caves
4,5,7,10,11,12,14,16,18,19,23,24,27,29,31,34,38,39,41,48–52,55,59,62,65

By multivariate method:

Canonical Correspondence Analysis (CCA)
1,9,22,26,36,54,57

Classification Analysis (CLUSTER)
9,11,13,15,17,20,31,42,64

Correspondence analysis (CA)
31,52,64

Discriminant Analysis (DA)
5,12,18,20,27,29,35,38,39,45,49,51,65

nonmetric Multi-Dimensional Scaling (nMDS)
3,6,9,14,30

Principal Component Analysis (PCA)
7,8,21,23,27,32,37,40,42,46,48–50,52,55

By subject:

Biota response to environment
1,2,3,6,8,9,15,17,21,22,25,26,28,30,31,32,33,36,42,43,44,46,47,53,54,56–58,60,64

Genetics
5,10,11,12,24,37,39,62,63

Morphology/Morphometrics
4,14,16,18,20,23,24,29,35,38,45,49–51,59,61,63,65

Paleontology
7,19,27,41,48,52,55

Sampling methods
13,40

PUBLICATION LIST (ordered by year of publication):

1. Pipan, T; Blejec, A; Brancelj, A, 2006. Multivariate analysis of copepod assemblages in epikarstic waters of some Slovenian caves. Hydrobiologia 559, 213–223.
2. Schryver, JC; Brandt, CC; Pfiffner, SM; Palumbo, AV; Peacock, AD; White, DC; McKinley, JP; Long, PE, 2006. Application of nonlinear analysis methods for identifying relationships between microbial community structure and groundwater geochemistry. Microbial Ecology 51, 177–188.
3. Carpentieri, P; Colloca, F; Ardizzone, GD, 2005. Day-night variations in the demersal nekton assemblage on the Mediterranean shelf-break. Estuarine Coastal and Shelf Science 63, 577–588.
4. Catala, S; Sachetto, C; Moreno, M; Rosales, R; Salazar-Schettino, PM; Gorla, D, 2005. Antennal phenotype of *Triatoma dimidiata* populations and its relationship with species of *Phyllosoma* and *Protracta* complexes. Journal of Medical Entomology 42, 719–725.
5. Fernández, GC; Juárez, MP; Monroy, MC; Menes, M; Bustamante, DM; Mijailovsky, S, 2005. Intraspecific variability in *Triatoma dimidiata* (Hemiptera: Reduviidae) populations from Guatemala based on chemical and morphometric analyses. Journal of Medical Entomology 42, 29–35.
6. Jaatinen, K; Tuittila, ES; Laine, J; Yrjala, K; Fritze, H, 2005. Methane-oxidizing bacteria in a Finnish raised mire complex: Effects of site fertility and drainage. Microbial Ecology 50, 429–439.
7. Kaniewski, D; Renault-Miskovsky, J; de Lumley, H, 2005.Palaeovegetation from a *Homo neanderthalensis* occupation in Western Liguria: archaeopalynology of Madonna dell'Arma (San Remo, Italy). Journal of Archaeological Science 32, 827–840.
8. Mouser, PJ; Rizzo, DM; Roling, WFM; Van Breukelen, BM, 2005. A multivariate statistical approach to spatial representation of groundwater contamination using hydrochemistry and microbial community profiles. Environmental Science & Technology 39, 7551–7559.
9. Preciado, I; Maldonado, M, 2005. Reassessing the spatial relationship between sponges and macroalgae in sublittoral rocky bottoms: a descriptive approach. Helgoland Marine Research 59, 141–150.
10. Schilthuizen, M; Cabanban, AS; Haase, M, 2005. Possible speciation with gene flow in tropical cave snails. Journal of Zoological Systematics and Evolutionary Research 43, 133–138.
11. Taylor, ML; Chavez-Tapia, CB; Rojas-Martinez, A; Reyes-Montes, MD; del Valle, MB; Zuniga, G. 2005. Geographical distribution of genetic polymorphism of the pathogen *Histoplasma capsulatum* isolated from infected bats, captured in a central zone of Mexico. Fems Immunology and Medical Microbiology 45, 451–458.
12. Tiranti, SI; Dyzenchauz, FJ; Hasson, ER; Massarini, A, 2005. Evolutionary and systematic relationships among tuco-tucos of the *Ctenomys pundti* complex (Rodentia : Octodontidae): a cytogenetic and morphological approach. Mammalia 69, 69–80.
13. Boulton, AJ; Dole-Olivier, MJ; Marmonier, P, 2004. Effects of sample volume and taxonomic resolution on assessment of hyporheic assemblage composition sampled using a Bou-Rouch pump. Archiv fur Hydrobiologie 159, 327–355.
14. Bustamante, DM; Monroy, C; Menes, M; Rodas, A; Salazar-Schettino, PM; Rojas, G; Pinto, N; Guhl, F; Dujardin, JP, 2004. Metric variation among geographic populations of the chagas vector *Triatoma dimidiata* (Hemiptera: Reduviidae: Triatominae) and related species. Journal of Medical Entomology 41, 296–301.
15. de Lipthay, JR; Johnsen, K; Albrechtsen, HJ; Rosenberg, P; Aamand, J, 2004. Bacterial diversity and community structure of a sub-surface aquifer exposed to realistic low herbicide concentrations. Fems Microbiology Ecology 49, 59–69.
16. Lozouet, P, 2004. The European Tertiary Neritiliidae (Mollusca, Gastropoda, Neritopsina): indicators of tropical submarine cave environments and freshwater faunas. Zoological Journal of the Linnean Society 140, 447–467.
17. Marti, R; Uriz, MJ; Ballesteros, E; Turon, X, 2004. Benthic assemblages in two Mediterranean caves: species diversity and coverage as a function of abiotic parameters and geographic distance. Journal of the Marine Biological Association of the United Kingdom 84, 557–572.
18. Muster, C; Schmarda, T; Blick, T, 2004. Vicariance in a cryptic species pair of European pseudoscorpions (Arachnida, Pseudoscorpiones, Chthoniidae). Zoologischer Anzeiger 242, 299–311.
19. Postawa, T, 2004. Changes in bat fauna during the Middle and Late Holocene as exemplified by thanatocoenoses dated with C-14 AMS from Krakow-Czestochowa Upland caves, Poland. Acta Chiropterologica 6, 269–292.
20. Prevorcnik, S; Blejec, A; Sket, B, 2004. Racial differentiation in *Asellus aquaticus* (L.) (Crustacea: lsopoda: Asellidae). Archiv fur Hydrobiologie 160, 193–214.
21. Findlay, SEG; Sinsabaugh, RL; Sobczak, WV; Hoostal, M, 2003. Metabolic and structural response of hyporheic microbial communities to variations in supply of dissolved organic matter. Limnology and Oceanography 48, 1608–1617.
22. Hunt, GW; Stanley, EH, 2003. Environmental factors influencing the composition and distribution of the hyporheic fauna in Oklahoma streams: Variation across ecoregions. Archiv fur Hydrobiologie 158, 1–23.
23. Machado, SF; Ferreira, RL; Martins, RP, 2003. Aspects of the population ecology of *Goniosoma* sp (Arachnida Opiliones Gonyleptidae) in limestone caves in southeastern Brazil. Tropical Zoology 16, 13–31.
24. Maharadatunkamsi; Hisheh, S; Kitchener, DJ; Schmitt, LH, 2003. Relationships between morphology, genetics and geography in the cave fruit bat *Eonycteris spelaea* (Dobson, 1871) from Indonesia. Biological Journal of the Linnean Society 79, 511–522.
25. Malard, F; Galassi, D; Lafont, M; Doledec, S; Ward, JV, 2003. Longitudinal patterns of invertebrates in the hyporheic zone of a glacial river. Freshwater Biology 48, 1709–1725.
26. Malard, F; Ferreira, D; Dolédec, S; Ward, JV, 2003. Influence of groundwater upwelling on the distribution of the hyporheos in a headwater river flood plain. Archiv fur Hydrobiologie 157, 89–116.
27. Neves, WA; Prous, A; González-José, R; Kipnis, R; Powell, J, 2003. Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. Journal of Human Evolution 45, 19–42.
28. Yan, TF; Fields, MW; Wu, LY; Zu, YG; Tiedje, JM; Zhou, JZ, 2003. Molecular diversity and characterization of nitrite reductase gene fragments (nirK and nirS) from nitrate- and uranium-contaminated groundwater. Environmental Microbiology 5, 13–24.
29. Armstrong, KN, 2002. Morphometric divergence among populations of *Rhinonicteris aurantius* (Chiroptera: Hipposideridae) in northern Australia. Australian Journal of Zoology 50, 649–669.
30. Bell, JJ, 2002. The sponge community in a semi-submerged temperate sea cave: Density, diversity and richness. Marine Ecology (Napoli) 23, 297–311.
31. Furman, A; Ozgul, A, 2002. Distribution of cave-dwelling bats and conservation status of underground habitats in the Istanbul area. Ecological Research 17, 69–77.
32. Haveman, SA; Pedersen, K, 2002. Distribution of culturable microorganisms in Fennoscandian Shield groundwater. Fems Microbiology Ecology 39, 129–137.
33. Scarsbrook, MR; Halliday, J, 2002. Detecting patterns in hyporheic community structure: does sampling method alter the story? New Zealand Journal of Marine and Freshwater Research 36, 443–453.
34. Wilkinson, GS; South, JM, 2002. Life history, ecology and longevity in bats. Aging Cell 1, 124–131.
35. Burr, BM; Adams, GL; Krejca, JK; Paul, RJ; Warren, ML, 2001. Troglomorphic sculpins of the *Cottus carolinae* species group in Perry County, Missouri: Distribution, external morphology, and conservation status. Environmental Biology of Fishes 31–32, 279–296.
36. Jersabek, CD; Brancelj, A; Stoch, F; Schabetsberger, R, 2001. Distribution and ecology of copepods in mountainous regions of the Eastern Alps. Hydrobiologia 541, 309–324.
37. Baltanás, A; Namiotko, T; Danielopol, DL, 2000. Biogeography and disparity within the genus *Cryptocandona* (Crustacea, Ostracoda). Vie et Milieu-Life and Environment 50, 297–310.
38. Gracco, M; Catala, S, 2000. Inter-specific and developmental differences on the array of antennal chemoreceptors in four species of Triatominae (Hemiptera: Reduviidae). Memorias do Instituto Oswaldo Cruz 95, 67–73.
39. Nevo, E; Ivanitskaya, E; Filippucci, MG; Beiles, A, 2000. Speciation and adaptive radiation of subterranean mole rats, *Spalax ehrenbergi* superspecies, in Jordan. Biological Journal of the Linnean Society 69, 263–281.

40. Choi, KH; Dobbs, FC, 1999. Comparison of two kinds of Biolog microplates (GN and ECO) in their ability to distinguish among aquatic microbial communities. Journal of Microbiological Methods 36, 203–213.
41. Cowling, RM; Cartwright, CR; Parkington, JE; Allsopp, JC, 1999. Fossil wood charcoal assemblages from Elands Bay Cave, South Africa: implications for Late Quaternary vegetation and climates in the winter-rainfall Fynbos biome. Journal of Biogeography 26, 367–378.
42. Franklin, RB; Taylor, DR; Mills, AL, 1999. The distribution of microbial communities in anaerobic and aerobic zones of a shallow coastal plain aquifer. Microbial Ecology 38, 377–386.
43. Nelson, SM; Roline, RA, 1999. Relationships between metals and hyporheic invertebrate community structure in a river recovering from metals contamination. Hydrobiologia 397, 211–226.
44. Strayer, DL; Reid, JW, 1999. Distribution of hyporheic cyclopoids (Crustacea: Copepoda) in the eastern United States. Archiv fur Hydrobiologie 145, 79–92.
45. Turk-Prevorcnik, S; Blejec, A, 1998. *Asellus aquaticus infernus*, new subspecies (Isopoda: Asellota: Asellidae), from Romanian hypogean waters. Journal of Crustacean Biology 18, 763–773.
46. Colwell, FS; Lehman, RM, 1997. Carbon source utilization profiles for microbial communities from hydrologically distinct zones in a basalt aquifer. Microbial Ecology 33, 240–251.
47. Dole-Oivier, MJ; Marmonier, P; Beffy, JL, 1997. Response of invertebrates to lotic disturbance: Is the hyporheic zone a patchy refugium? Freshwater Biology 37, 257–276.
48. Serrano F.; Guerra-Merchán, A.; Lozano-Francisco, C; Vera-Peláez, JL, 1997. Multivariate analysis of remains of molluscan foods consumed by latest Pleistocene and Holocene humans in Nerja Cave, Malaga, Spain. Quaternary Research 48, 125–227.
49. Tizado, J; Núñez-Pérez, E, Salgado, JM; Salgado, EJ, 1997. Morphometric relationships among populations of the cave-dwelling beetle *Quaestus nietoi* (Salgado, 1988) (Coleoptera:Leiodidae). Annales de la Societe Entomologique de France 33, 215–231.
50. Arita, HT, 1997. Species composition and morphological structure of the bat fauna of Yucatan, Mexico. Journal of Animal Ecology 66, 83–97.
51. Guzvica, G; Kuber, D; Modric, S, Radanovic-Guzvica, B, 1996. Use of craniometry in discrimination of brown and cave bears. Veterinarski Archiv 66, 251–257.
52. Chaline, J; Brunet-Lecompte, P; Campy, M, 1995. The last glacial interglacial record of rodent remains from the Gigny karst sequence in the French Jura used for paleoclimatic and paleoecological reconstructions. Palaeogeography, Palaeoclimatology and Palaeoecology 117, 229–252.
53. Lehman, RM; Colwel, FS; Ringelberg, DB; White, DC, 1995. Combined microbial community-level analyses for quality assurance of terrestrial subsurface cores. Journal of Microbiological Methods 22, 263–281.
54. Plenet, S; Gibert, J, 1995. Invertebrate community responses to physical and chemical factors at the river aquifer interaction zone .2. Downstream from the city of Lyon. Archiv fur Hydrobiologie 65–88.
55. Woodman, N, 1995. Morphological variation between Pleistocene and recent samples of *Cryptotis* (Insectivora, Soricidae) from the Yucatan peninsula, Mexico. Journal of Mammalogy 76, 223–231.
56. Plenet, S, Hugueny, H; Gibert, J, 1994. Invertebrate community responses to physical and chemical factors at the river aquifer interaction zone .1. Upstream from the city of Lyon. Archiv fur Hydrobiologie 136, 165–189.
57. Richards, C; Bacon, KL, 1994. Influence of fine sediment on macroinvertebrate colonization of surface and hyporheic stream substrates. Great Basin Naturalist 54, 106–113.
58. Zheng, M; Kellogg, ST, 1994. Analysis of bacterial-populations in a basalt aquifer. Canadian Journal of Microbiology 40, 944–954.
59. Simson, S; Lavie, B; Nevo, E, 1993. Penial differentiation in speciation of subterranean mole rats *Spalax ehrenbergi* in Israel. Journal of Zoology 229, 493–503.
60. Williams, DD, 1993. Changes in fresh-water meiofauna communities along the groundwater-hyporheic water ecotone. Transactions of the American Microscopical Society 112, 181–194.
61. Allegrucci, G; Baldari, F; Cesaroni, D; Thorpe, RS; Sbordoni, V, 1992. Morphometric analysis of interspecific and microgeographic variation of crayfish from a Mexican cave. Biological Journal of the Linnean Society 47, 455–468.
62. Allegrucci, G; Caccone, A; Cesaroni, D; Sbordoni, V, 1992. Evolutionary divergence in *Dolichopoda* cave crickets - a comparison of single copy DNA hybridization data with allozymes and morphometric distances. Journal of Evolutionary Biology 5, 121–148.
63. Cesaroni, D; Allegrucci, G; Sbordoni, V, 1992. A narrow hybrid zone between 2 crayfish species from a Mexican cave. Journal of Evolutionary Biology 5, 643–659.
64. Uriz, MJ; Rosell, D; Martín, D, 1992. The sponge population of the Cabrera archipelago (Balearic Islands) - Characteristics, distribution, and abundance of the most representative species. Marine Ecology 13, 101–117.
65. Gardner, SL; Duszynski, DW, 1990. Polymorphism of *Eimerian* oocysts can be a problem in naturally infected hosts - An example from subterranean rodents in Bolivia. Journal of Parasitology 76, 805–811.