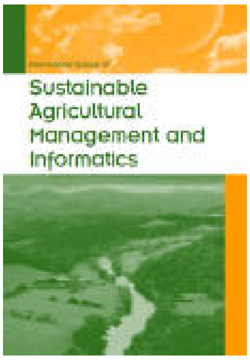


International Journal of Sustainable Agricultural Management and Informatics

ISSN online: 2054-5827 - ISSN print: 2054-5819
https://www.inderscience.com/ijsami

A standardised method for estimating environmental and agronomic covariates to discriminate the explanatory variables effects on bioindicators: a case study on soil fauna

Stefano Macchio, Monica Vercelli, Michela Gori, Luisa Nazzini, Valter Bellucci, Pietro Massimiliano Bianco, Carlo Jacomini, Enrico Rivella, Susanna D'Antoni

DOI: 10.1504/IJSAMI.2023.10058368

Article History:

Received:	23 February 2023
Last revised:	19 April 2023
Accepted:	20 April 2023
Published online:	08 December 2023

A standardised method for estimating environmental and agronomic covariates to discriminate the explanatory variables effects on bioindicators: a case study on soil fauna

Stefano Macchio

Italian Institute for Environmental Protection and Research (ISPRA),
Via Vitaliano Brancati 48, 00144 Rome, Italy
Email: stefano.macchio@isprambiente.it

Monica Vercelli*

Strada Comunale Val Pattonera 184,
10133, Turin, Italy
Email: monibee13@icloud.com
*Corresponding author

Michela Gori, Luisa Nazzini, Valter Bellucci, Pietro Massimiliano Bianco and Carlo Jacomini

Italian Institute for Environmental Protection and Research (ISPRA),
Via Vitaliano Brancati 48, 00144 Rome, Italy
Email: michela.gori@isprambiente.it
Email: luisa.nazzini@isprambiente.it
Email: valter.bellucci@isprambiente.it
Email: pietro.bianco@isprambiente.it
Email: carlo.jacomini@isprambiente.it

Enrico Rivella

Regional Agency for Environmental Protection of
Piedmont (APRA Piemonte),
Via Pio VII 9, 10135 Turin, Italy
Email: enrico.rivella@arpa.piemonte.it

Susanna D'Antoni

Italian Institute for Environmental Protection and Research (ISPRA),
Vitaliano Brancati 48,
00144 Rome, Italy
Email: susanna.dantoni@isprambiente.it

Abstract: A method was developed for the standardisation and analysis of environmental and agronomic covariates to discriminate the effects of specific explanatory variables on a given bioindicator. To test it, the effects of plant protection products (PPP) was assessed on soil fauna sampled in organic and conventional hazelnut orchards. More than 100 standardised covariates were numerically reduced, by Principal coordinates analysis (PCoA), to two derived covariates. Then, redundancy analysis (RDA) was applied using, as explanatory variables, two indexes referred to PPP input and the derived covariates. The results showed a marked differentiation of the soil fauna communities between the two groups of sampled sites and their clear response to the use of PPP. The procedure proved to be effective in reducing the 'background noise' determined by a great number of covariates. This method can be successfully applied in monitoring activities concerning the effects on biodiversity of several initiatives aimed at reducing PPP use.

Keywords: plant protection products impact; agricultural management; covariates; bioindicators; organic farming; principal coordinates analysis; PCoA; canonical correspondence analysis; CCA; soil fauna.

Reference to this paper should be made as follows: Macchio, S., Vercelli, M., Gori, M., Nazzini, L., Bellucci, V., Bianco, P.M., Jacomini, C., Rivella, E. and D'Antoni, S. (2024) 'A standardised method for estimating environmental and agronomic covariates to discriminate the explanatory variables effects on bioindicators: a case study on soil fauna', *Int. J. Sustainable Agricultural Management and Informatics*, Vol. 10, No. 1, pp.1–26.

Biographical notes: Stefano Macchio is a biologist and researcher. He worked at the National Wildlife Institute (INFS) on bird migration and habitat selection. He held numerous positions in protected areas dealing with wildlife and natural resources monitoring and management. Currently at the Italian Institute for Environmental Protection and Research (ISPRA) where his activities mainly deal with the biostatistical data analysis and the development of methods for environmental and zoological sampling for bioindication. He fills the role of national expert about the river fish fauna, having defined and brought to UE intercalibration the Italian method of rivers classification through the fish communities as ecological indicators under the WFD 2000/60.

Monica Vercelli is a Postdoctoral researcher in Apicultural Science and Professor of Biological Control and Biodiversity and Management of Pollinating Insects, Professional beekeeper. She earned a Bachelor's degree at the University of Studies of Turin in Organic Agriculture, a Master's degree in Agroecology, and PhD in Natural Sciences and Innovative Technologies. She is expert in the fields of: honeybee and wild bees, beekeeping, honey bee colony issues, bee flora, pollination, honey and bee products, biodiversity conservation, pollinators monitoring, and cooperation in international beekeeping programs. She is a member of International Honey Commission and National registers of "Experts of Sensory analysis of the Honey" and "Experts of Melissopalynology".

Michela Gori graduated in Biological Sciences with a specialisation in zoology. She has an MSc in Faunal Biodiversity Conservation. She has been working since 2004 in the Italian Institute for Environmental Protection and Research (ISPRA). She carries out technical and scientific activities related to ecological connectivity, biological indicators, environmental impact assessment and strategic impact assessment and she provides technical-scientific support activities to protected areas.

Luisa Nazzini graduated in Environmental Sciences, has over 20 years of experience in research and environmental protection. Since 2005, she has been working in the Italian Institute for Environmental Protection and Research (ISPRA), where she carries out technical and scientific activities related to eco-sustainable territorial planning, ecological connectivity, environmental governance, biodiversity conservation and biological indicators. She also provides technical-scientific support activities to protected areas. She carries out assessments on the effectiveness of measures implemented for the protection of biodiversity envisaged by national guidelines and strategies. She is an expert in soil biomonitoring and reporting.

Valter Bellucci has been working as a biologist since 1992 initially at the University of Rome 'Tor Vergata', then since 2006 at the Italian Institute for Environmental Protection and Research (ISPRA). He has specialisations and MSc in Science and Technology for the Environment and Nature. He is responsible for scientific activities in the sector of analysis and protection of ecosystems, ecosystem services provided by pollinating insects. His fields of activity and consultancy are conservation and management of Apoidea and associated flora. He published numerous articles on the role of pollinating insects for the protection of biodiversity and on the sustainable use of agroforestry resources.

Pietro Massimiliano Bianco graduated in Natural Sciences and holds a PhD in Territorial Sciences and Technologies. He has been working in the Italian Institute for Environmental Protection and Research (ISPRA) since 2005; dealing with cartography, identification of pressures and protection of species and habitats, organisation of environmental data in GIS environment, sustainable fight against harmful organisms, environmental education, environmental impact assessments and strategic environmental assessments. He is currently a member of the Scientific Committee of the Network for the study of Mycological Diversity and of the Technical-Scientific Committee for the Management and Coordination of the National Biodiversity Network.

Carlo Jacomini is a Soil Ecologist and Head of the Soil and Land Ecology Laboratory, Biology Area, in the National Centre for the national Laboratories' network of ISPRA, the Italian Institute for Environmental Protection and Research. He has over 30 years of experience in environmental monitoring, biological indicators and nature conservation, particularly on soil biodiversity and sustainable management. He has a long field and lab experience and produced cross-cutting research and reports of national and international scale. He is also lecturer in universities and PhD courses, and published several scientific reports and papers, book chapters and dissemination publications.

Enrico Rivella graduated in Biology and Natural Sciences and he works in the Environmental Agency of Piedmont Region, Italy (ARPA Piemonte). He is expert in environmental impact assessment, restoration ecology, rural landscapes and different monitoring schemes ranging from effects of climate change on alpine biodiversity, to ancient woods, to vegetation of cultivated crops.

Susanna D'Antoni is a naturalist and technologist at the Italian Institute for Environmental Protection and Research (ISPRA) where she has been Responsible of Protected Areas, Planning and Management of the territory and the landscape Section since the 2017. She coordinates projects regarding biodiversity conservation in Protected Natural Areas and in Natura 2000 Network, in particular. Her main activities focus on management, planning and monitoring of wildlife, natural resources, freshwater ecosystems and territory

and the publication and dissemination of technical-scientific indications, guidelines or regulatory instruments for prevent the human impact, especially caused by pesticides.

1 Introduction

'Bioindicator' is a term taken from environmental toxicology where it is defined as "an organism or biological response that reveals the presence of the pollutants by the occurrence of typical symptoms or measurable responses" (Karr, 1981). Over time, other disciplines have successfully experimented bioindication and currently a wider definition of bioindicators can be the following: ''a species or group of species that readily reflects the abiotic or biotic state of an environment, represents the impact of environmental change on a habitat, community, or ecosystem, or is indicative of the diversity of a subset of taxa, or of the wholesale diversity, within an area'' (McGeoch, 1998). In the European Union an impetus to the use of protocols for bioindication has been given by their introduction into regulatory instruments such as the Water Framework Directive (Dir. 2000/60/CE) (López-López and Sedeño-Díaz, 2014; Monteagudo and Moreno, 2016).

Although bioindicators are effective in recording environmental variations (Bispo et al., 2009; Botias et al., 2017; Cunningham et al., 2022; Edwards et al., 1996; Eijsackers, 1983; Gerlach et al., 2013; Girotti et. al, 2013; Li et al., 2005; van Straalen and Krivolutsky, 1996), assessing which variables are responsible for those variations could be extremely difficult. In fact, in observational studies, the effects of the target explanatory variables (object of the study) on the response variables (generally a population or a taxocenosis) are always, at least partially, masked by the effects of other not target independent variables that can be defined as covariates.

In order to eliminate this background noise, it is very important to draw up a list, as complete as possible, of the covariates that can have consistent effects on the response variables, associated with their specific measurement or classification method in order to fulfil two fundamental objectives:

1 to obtain detailed and (semi) quantitative descriptions of the environmental contexts that must be compared, which helps choose sampling sites with as little dissimilarities as possible except for the target variables

2 to possibly use the covariates together with the target variables in a multivariate analysis that is able to identify which ones have a solid effect on the bioindicator communities.

It is well established today that it is necessary to standardise the bioindicator sampling protocols (Smallshire and Beynon, 2010; Stahlschmidt and Brühl, 2012; Tourinho and Lo-Man-Hung, 2021). In a similar way, the organisation of a shared standard protocol for collecting data on environmental variables and pressures represents an objective of primary importance to gather comparable data and ensure replicable studies.

All of these topics have been addressed in a 5-year Italian project (2015–2019) financed by the Italian Ministry of the Environment and Energy Security (MASE), coordinated by Italian Institute for Environmental Protection and Research (ISPRA) and carried out in collaboration with Regional agencies for the protection of the environment

(ARPA) of Latium and Piedmont Regions, the University of Turin and the University of Rome Tor Vergata. The main aim of the project was to verify if organic farming and good practices of agroecology are more compatible with the conservation of biodiversity than conventional farming where PPP are used. This is in line with the provisions of Measures 13 and 16 provided in the 'Guidelines for the protection of the aquatic environment and drinking water and for the reduction of the use of plant protection products (PPP) and their relative risks in Natura 2000 sites and in natural protected areas' (Interministerial Decree 10/3/2015) for the application of the Italian National Action Plan for the sustainable use of PPP (NAP) (Interministerial Decree 22/01/2014) according to the European Directive 2009/128/EC. Moreover, the study was aimed at identifying bioindicators that are useful for evaluating the effects of PPP on biodiversity (D'Antoni et al., 2020).

The project focused on three different permanent crops: rice fields, vineyards, and hazelnut orchards. These crops have been selected because, among those grown in protected areas and in Natura 2000 sites, they are all subjected to a high number of treatments with PPP (Italian Ministry of Agricultural, Food and Forestry Policies, 2022). The study has been carried out in two Italian Regions: Piedmont for rice fields and vineyards, and Latium for hazelnut orchards. The data have been collected during 2015–2016 and 2018–2019 campaigns.

In order to compare organic and conventional farming and, therefore, to observe the effects of PPP on biodiversity at different spatial and temporal scales, a wide range of bioindicators have been selected and tested: flora and vegetation, soil fauna, soil arthropods and carabid beetles (only in hazelnut orchards), bees, butterflies, dragonflies (only in rice fields), amphibians (only in rice fields), reptiles (only in hazelnut orchards) and bats.

This paper focuses on one of the main outputs of the project which is the development of a rapid method for the identification and the analysis of environmental and agronomic covariates in order to appreciate the qualitative and quantitative effects of the explanatory variable on a given bioindicator. This method has been tested using the project dataset concerning soil fauna sampled in hazelnut orchards. A secondary objective of the paper is to assess the effects that the use of PPP has on this taxocenosis.

2 Methodology

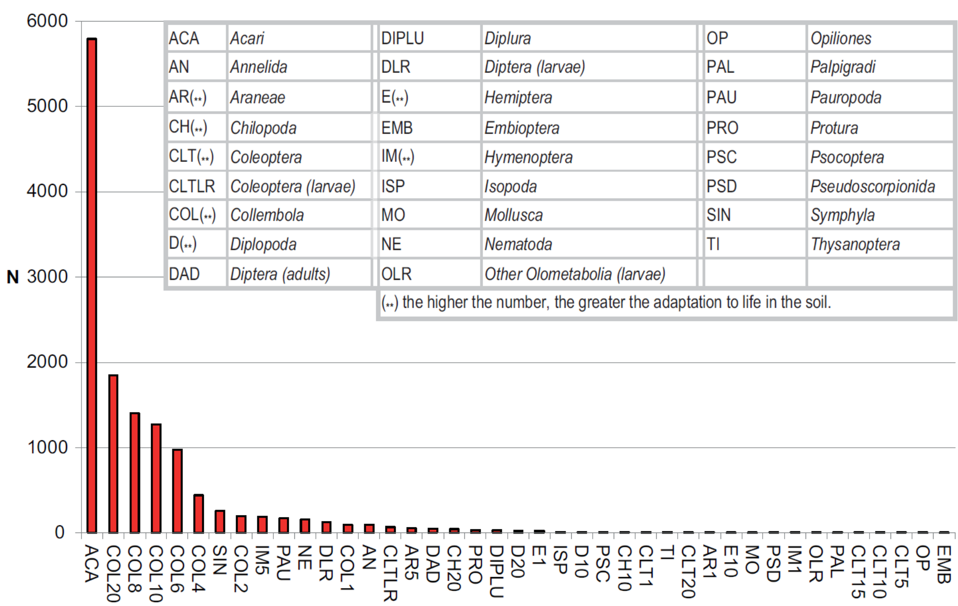
To achieve the above mentioned objectives, 6 hazelnut orchards cultivated in an organic regime (labelled with acronym 'OH') and 6 hazelnut orchards selected in conventional farms (labelled with acronym 'CH') have been selected and compared for the presence and abundance of a set of soil fauna taxa as a bioindicator case study. To minimise the covariates effects, study areas have been selected in pairs of fields (organic versus conventional) having geographic location, environmental characteristics and size as similar as possible. The selection was based on the covariates as described in §2.1.

First of all, a standardised list of environmental and agronomic covariates, suitable for agricultural contexts, has been created (Annex 1). In the list, methods of measurement and/or classification are given for each covariate. Then, these original covariates have been numerically reduced by Principal Coordinates Analysis (PCoA) in a smaller number of new derived covariates. The derived covariates, together with a variable referred to PPP input in neighbouring fields and with an index related to PPP used within the field have been used as target explanatory variables. Finally, the effectiveness of this approach was tested using the soil fauna as bioindicators in assessing the response to the use of PPP by redundancy analysis (RDA) carried out on data relating to organic fields and the corresponding conventional fields (ter Braak and Verdonschot, 1995; Legendre and Legendre, 1998).

2.1 Testing the approach on a bioindicator used as a case study

The sampling method used to monitor soil fauna as a bioindicator is based on Parisi (2001) for evaluating the biological quality of soil. Briefly, in each field, three cubic clods (10 cm side) were extracted, equidistant along the diagonal and both intra and inter row of hazelnut trees. The cubic samples were transported within 24 hours to the laboratory, closed in a hermetic container and protected from thermal shock or bumps. There, they were carefully placed in modified Berlese-Tullgren extractors (Górny and Grüm, 1993), with 2-mm sieve and 40 W lamps, for 14 days. The results of extraction were kept in hermetic canisters filled with preservative solution (3 parts 75% ethanol and 1 part glycerol).

Figure 1 Sample structure in terms of decreasing absolute abundance in taxa or subcategories of taxa as classified by Parisi (2001) where the numbers following the codes (along the abscissa axis) are associated with the adaptation to edaphic life (see online version for colours)



Specimens were observed under a stereomicroscope at low magnification (range 5×–100×) to identify individual micro-arthropod taxa, annelids, mollusks and nematodes. Micro-arthropods have been classified following Parisi (2001) which divides some taxa into smaller groups according to their different adaptation levels to the edaphic environment. Such classification verifies if, inside the same taxon (in some cases orders which count several families and species), there are different responses to the use of PPP depending on the different edaphic level occupied. Moreover, all individuals of each taxonomic group have been counted.

Since the number of valid sampling sessions was not identical for the different hazelnut orchards during the 2015–2016 and 2018–2019 campaigns, the average number of individuals counted per session was used for each taxon. Moreover, as natural populations of different taxa show great differences in numbers of individuals (up to two orders of magnitude) (Figure 1), (Galli et al., 2019; Galli et al., 2011; Hopkin, 1997; Mateos, 2016) the averages of the counts have been transformed in the natural logarithm [ln (n + 1)] in order to limit the effect on the results that this disparity can cause. In addition, the taxa or the subcategories of taxa (as classified by Parisi, 2001) detected only in a single station were not considered in the analysis in order to exclude the 'noise' of accidental taxa and to avoid putting too much weight on rare taxa in the analysis (Leps and Smilauer, 2003), so bringing to 41 subcategories of taxa used for the analyses.

Table 1 Description of the different topics into which the covariates have been classified according to their coherence of information

Topic	*Description*
1	Data relating to farms, crops and the agronomic practices
2	Information on soil tillage
3	Data regarding timing and quantities of treatments with plant protection products
4	Timing and quantities of treatments with fertilisers
5	Presence and size of agricultural annexes and the presence of any crops in a 10 metres buffer around the perimeter of the monitored field
6	Coverage of EUNIS land use categories in a 10 metres buffer around the perimeter of the monitored field. For the EUNIS categories relating to tree vegetation, the measurement of basic structural parameters (trunk diameter and height) and the assessment of the maturity of the formation in classes were also envisaged
7	Variables relating to the flowering and fruiting as well as the structure of the crop and of any natural vegetation within the monitored field and in the 10 metres buffer
8	Meteorological data recorded before and during the sampling events
9	Soil data and other variables directly detected at the sampling points. The surface considered for data collection is a circle centred at the sampling point and having 1m radius
10	Data of laboratory analysis of soil samples
11	Data from cartographic analysis (using Corine Land Cover inventory) carried out for field and field buffers of 10, 50, 100, 200 and 500 metres
12	Indicators of use of plant protection products, developed during the research

2.2 Statistical analysis

2.2.1 Original covariates and target variables and their reduction into a smaller number of derived covariates

A list of environmental and agronomic variables, suitable for agricultural contexts, has been created consulting a wide bibliography (citing only the most relevant: Colemana and Withman, 2005; Ferrari et al., 2008; Previati et al., 2007; Rismondo et al., 2011; Smeets and Wetering, 1999; Taffetani and Rismondo, 2009). In regard to the extreme heterogeneity of environmental variables and mechanical agronomic practices, the covariates were detected in different formats: measures, categorical and ordinal, with the latter by far more frequent. Covariates were organised into topics (from 1 to 12; see Table 1) based on their coherence of information. Further descriptors derived from collected data, such as ordinal versions (ranks) of measures or categorical data, or by calculating ratios between different covariates (Annex 1).

In order to select study areas having environmental characteristics and size as similar as possible, data concerning the variables listed in topics 1, 2, 5, 6 and 11 of Annex 1 have been collected since the first phase of the project for a large number of hazelnut orchards.

The list of environmental variables presented in Annex 1 has been developed to be used by those who undertake research studies on agroecosystems, considering those that may affect the status of the species and habitats considered in the study. For this reason, Annex 1 collects a great number of environmental and agronomic variables, and data from territorial analysis performed through GIS and from soil analysis. Therefore, depending on the type of study the researchers intend to undertake, it is possible to select the suitable covariates for that specific research within it. It should be underlined that for the study presented in the paper, all the variables listed in Annex 1 were considered.

Table 2 Criteria to assign a score to each plant protection product

Criteria	*Score*
Products candidate for replacement and banned	1
Environmental toxic products (N)	0,75
Products with precautionary phrases for the environment (SPe) defined by	
Directive 2003/82/EC which indicate possible impacts on non-target organisms + phrases H400 ‘highly toxic to aquatic organisms’ + H410 ‘highly toxic to aquatic organisms with long-term effects’	0,50
Products with risk phrases for aquatic ecosystems (H411, H412, H413)	0,25
Products authorised in organic farming without precautionary or risk phrases for the environment	0

Source: The highest score has been given to the most toxic products according to the Commission Regulation EU no. 547/2011, the Directive 2003/82/EC and the Interministerial Decree 10/3/2015

The use of PPP inside each study area was expressed through the plant protection products index (PPI), expressly developed to give information on the quantity and quality of products distributed inside each field. To calculate PPI, first of all, each PPP has been assigned a score plant protection products score (PPS) based on its degree of toxicity for plant and animal species and for habitats as reported on the product label in accordance with the Commission Regulation EU no. 547/2011 and the Directive 2003/82/EC (see Table 2). These criteria are the same considered in Measure 13 of the Guidelines for the application of the NAP (Interministerial Decree 10/3/2015) for the indication of PPP that must be eliminated/replaced/reduced in protected areas and Natura 2000 sites.

Then PPI has been developed considering, for each growing season, how many different products and how many times they have been used and the level of toxicity for the environment of each product used (PPS). Regarding the quantities dispensed on each field, the farmers confirmed they followed the indications given on the product labels.

The PPI for field ‘i’ and crop season ‘j’ was then calculated as

$$PPIij = \sum PPSij.$$

A second explanatory variable, PPP input in neighbouring fields (PNF), gives information on whether or not PPP are used in the neighbouring fields.

A large number of covariates is usually required to describe the environmental state of the sites. On the other hand, for statistical aims, the high number of original covariates and their wide-ranging multicollinearity requires their reduction and the removal of the correlations among them; in fact, such correlations can cause considerable distortions, capable of magnifying the effect of some variables while masking that of others.

First of all, the covariates that showed values too similar in both hazelnut orchard groups (due to the selection of the pairs of fields, organic and conventional, aimed to minimise differences except for PPP treatments), were excluded from the statistical analysis. The reduction of all the other original covariates was carried out by applying the analysis of the principal coordinates PCoA, a method that processes heterogeneous data together (Legendre and Legendre, 1998).

The reduction through PCoA was not applied to all the original covariates together, because, in order to maintain a level of traceability with the original data, it was applied for groups coherent for spatial scale:

a Covariates detected in the field describing the morphological and management characteristics and habitat availability in the field and in the 10m buffer. These covariates are listed in topics 1, 5, 6, 7 of Annex 1 (named 'X0107' in the figures and tables.

b Covariates described in topic 11 and derived from cartographic analysis (using Corine Land Cover classes) in progressive radius buffers from 10 to 500 metres from the perimeter of the sample fields (named 'X11' in the figures and tables). These covariates describe the territorial matrix in which the field is embedded.

c Covariates relating to soil grain size measured in the laboratory and listed in topic 10, adding the richness index (number of soil grain size classes) and the dominance index (1-Simpson index of the arcsine transformed coverage percentages of grain soil size classes) (named 'X10' in the figures and tables).

For each of the three groups of original covariates, a smaller number of principal coordinates (called derived covariates) was selected to explain about 70% of the observed group variability.

Since the data to be reduced consisted of environmental variables, the Spearman rho coefficient was chosen as a measure of monotonic association of symmetrical type (this coefficient also includes the cases in which the considered variable is absent in both sites under comparison, the so-called 'double zeros' in the calculations).

2.2.2 Multivariate methods

canonical correspondence analysis (CCA) and RDA are multivariate methods for elucidating the relationships between biological communities (composition and abundance of taxa) and their environment. Both methods are designed to extract synthetic environmental gradients from ecological datasets. Gradients are the basis for synthetically describing and visualising the different habitat preferences (niches) of taxa through an ordering diagram. While CCA assumes unimodal functions in the habitat preferences of taxa, RDA assumes linear function.

To decide whether the biological data are homogeneous or heterogeneous (and therefore more suitable for linear rather than unimodal sorting methods, respectively), the detrended correspondence analysis (DCA) was firstly performed to check the length of the first axis in units of standard deviation (SD) (Lepš and Šmilauer, 2003): if greater than 4 SD it is appropriate to apply the CCA, if less than 3 SD the RDA is more adequate while in the interval between 3 and 4 SD both techniques can be used. Since the length of the first axis obtained by DCA is less than 3 SD, the RDA has been applied.

Finally, the relationships between environmental variables and each ln-transformed taxonomic group abundance have been examined using multiple linear regression with stepwise backward selection by Akaike information criterion (Akaike, 1973).

3 Results and discussion

3.1 Covariates dimensionality reduction and its efficacy tested on the soil fauna

Original covariates were listed in Annex 1 that is a very important part of the study since it describes, for each covariate, its method of measurement, classification, standardisation and ranking. Following the reduction, as described in §2.2.1, a number of derived covariates has been selected to explain about 70% of the variability for each of the 3 groups of original covariates: one for the covariates detected in the field (‘X0107UP’); two for covariates derived from cartographic analysis (‘X11AP’ and ‘X11BP’), the first of the two being most important, explaining more variability of the group than the second; one for the group of covariates related to soil grain size measured in the laboratory (‘X10UP’) (Table 3). A statistically significant correlation remained (rho-Spearman $P < 0.05$) between the derived covariates representative of the soil particle sizes (X10UP) and the first of the two representing the set of covariates detected in progressive buffers (X11AP). Therefore, one of them had to be excluded from the analysis. Since it is reasonable to expect a greater influence from the covariates acting on a limited spatial scale for the soil fauna, such as the grain size of the soil, than those acting on a larger scale, the covariates derived from cartographic analysis (X11), referred to the territorial matrix, have been excluded. Therefore, starting from the analysis of more than 100 standardised covariates, they were numerically reduced to two derived covariates by PCoA.

Table 3 Final number of derived covariates obtained by applying the PCoA to the three groups of original covariates

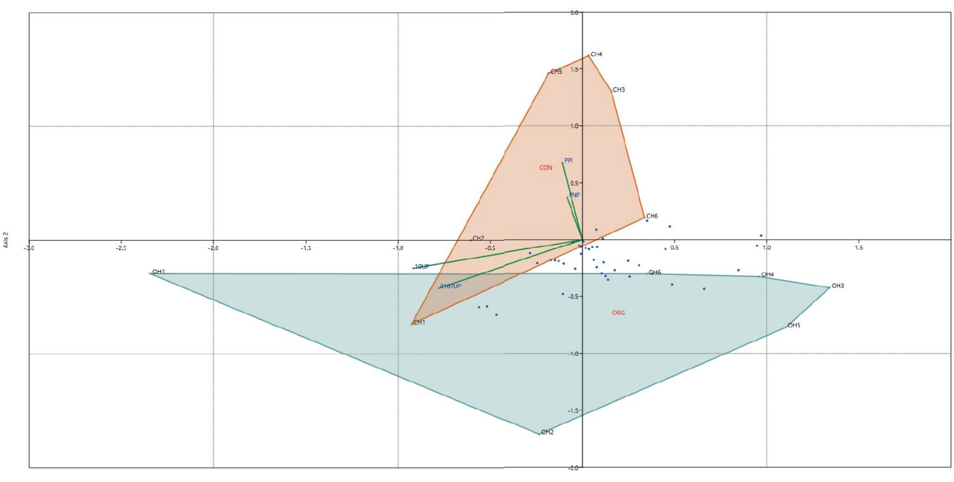
New covariates group derived from PCoA reduction and original topics associated	*Number and name of selected Axes*	*Eigen value*	*Variation explained*
X0107 (topics 01, 05, 06, 07)	1: X0107UP	0.7344	79.65%
X10 (topic 10)	1: X10UP	0.1010	72.55%
X11 (topic 11 for buffer of 500, 200, 100, 50, 10 metres)	1: X11AP	0.1927	54.39%
	2: X11BP	0.0530	14.95%

Notes: For derived covariates X0107 and X10 a single axis each is sufficient to explain more than 70% of the variability, while for X11 two axes are needed to reach the threshold of 70% of the explained variability.

3.2 Results of multivariate methods

The graph in Figure 2 is set up with scaling 'type 2', to be preferred when most of the explanatory variables are non-binary, and with fitted site scores, i.e., site scores are expressed as linear combinations of the environmental variables. The blue dots express the taxa and subcategories of taxa, the differently coloured dots flanked by the prefix OH or CH represent the organic and conventional hazelnut orchards, respectively, while the position of the environmental variables (X0107UP; X10UP; PPI index and PNF variable) can be identified in the ends of the vectors in green.

Figure 2 RDA tri-plot showing the distribution of taxa and subcategories of taxa (blue dots), sampled sites (CHx and OHx) and environmental variables (green vectors) (see online version for colours)



The points representing the sampled sites (hazelnut orchards) with the same type of management are graphically collected within their own-coloured convex hull (the smallest convex polygon containing a given set), allowing to appreciate their dispersion in the two-dimensional space formed by the first two main axes. The closer the sites are to each other (and the smaller the convex hull), the more similar the structure and composition of the soil fauna sampled therein. The graph shows a good separation of the convex hulls of organic and conventional hazelnut orchards (with only a modest overlap due to CH1), indicating a marked differentiation of the soil fauna communities between the two groups of sampled sites.

The separation between the two convex hulls is greater along the vertical axis, i.e., the second of the two main axes represented, with all biological hazelnut orchards having negative ordinates and all conventional hazelnut orchards, with the exception of CH1, having positive ordinates. The second main axis is also characterised by the highest correlation of PPI and PNF, whose absolute ordinate values are significantly higher than those of the derived covariates X10UP and X0107UP.

On the contrary, the derived covariates show a higher correlation towards the first principal axis (this is graphically expressed by the higher absolute values of the respective abscissae).

In addition, it can be observed that most of the taxa express a negative ordinate, thus being closer to the convex hull of biological hazelnut fields and, at the same time, they are more distributed along the environmental gradient, expressed by the first major axis, which is predominantly characterised by the covariates rather than by the PPI index and the PNF variable.

Table 4 RDA statistics and overall permutation tests to assess the significance of the constraints

Sum eigenvalue canonical axes = 10.0186						*Overall test*	
Axis	*Eigenvalue*	*Axis inertia explained %*	*Cumulative inertia explained %*	*taxa – environment correlations (R)*	*Axis p-permutation test significance*	*R^2*	*0.6326*
RDA1	5.3119	33.54	33.54	0.939	0.002 **	R^2_{adj}	0.4226
RDA2	2.9710	18.76	52.30	0.926	0.024 *	*F*	3.013
RDA3	1.2281	7.75	60.05	0.860	0.400 N.S.	p – permutation (n = 999)	0.001 (***)
RDA4	0.5077	3.21	63.26	0.849	0.804 N.S.		
Sum eigenvalue residual axes = 5.8193							

The variance explained by the canonical axes (‘constrained variance’) is clearly higher than that explained by the residual axes (‘unconstrained variance’, 63.26% vs. 36.74%). The explanatory variables defined for the analysis (factors associated with the use of PPP and covariates) were therefore able to explain most of the variation shown in the response variables (the abundances of the different taxa).

Table 5 Partial RDA statistics to show the effect of the covariates (‘conditional’) with respect to that of the environmental variables of interest (‘constrained’)

Partial RDA						
	Inertia	*Percentages*	*Rank*	p-permutation test significance		
Total	15.8378	100				
Conditional	6.1052	38.55	2			
Constrained	3.9134	24.71	2	0.017*		
Unconstrained	5.8193	36.74	7			
Eigenvalues for constrained axes						
RDA1	RDA2					
3.0807	0.8327					
Eigenvalues for unconstrained axes						
PC1	PC2	PC3	PC4	PC5	PC6	PC7
2.1731	1.5457	0.8542	0.4947	0.3386	0.2123	0.2006

Table 6 Results of the multiple linear regression analyses between environmental factors and abundance ln-transformed for each taxonomic group

Taxonomic group	*Environment variable*	*Statistical significance*			*Model*	R^2
		Env. variable Ln-transformed	*Intercept*	*Model*		
ACA	PPI	*	***	(*)	ln(ACA + 1) = 8.445–1.088PNF-0.261PPI–1.393X10UP	0.589
AR1	PPI	*		*	ln(AR + 1) = 0.03792–0.08545PPI + 1.24652X0107UP + 0.2532PNF–0.64326X10UP	0.738
	X0107UP	**				
AR5	PPI	**	***	***	ln(AR5 + 1) = 1.718–0.184PPI + 2.512X0107UP	0.824
	X0107UP	***				
OP	PPI	*	**	(*)	ln(OP + 1) = 0.153–0.031PPI–0.349X10UP + 0.434X0107UP	0.588
	X0107UP	*				
	X10UP	*				
PAL	PPI	*	*	*	ln(PAL + 1) = 0.1022 + 0.0288PPI–0.3158X0107UP + 0.2532PNF–0.64326X10UP	0.61
	X0107UP	*				
PAU	X10UP	**	**	*	ln(PAU + 1) = 4.922–1.251PNF–0.186PPI–3.036X10UP	0.66
SIN	X10UP	*	***	*	ln(SIN + 1) = 2.9531–1.292X10UP	0.478
PRO	PPI	(*)	**	*	ln(PRO + 1) = 4.293–1.603PNF–0.155PPI–1.156X0107UP	0.701
	X0107UP	(*)				
	PNF	*				
DIPLU	X10UP	*	***	*	ln(DIPLU + 1) = 1.285–0.113PPI–1.396X0107UP	0.523
COL1	PPI	(*)	***	(*)	ln(COL1 + 1) = 2.309–0.136PPI + 2.379X0107UP–1.373X10UP	0.592
	X0107UP	*				
COL2	PNF	*		*	ln(COL2 + 1) = 4.168–0.767PNF	0.362
COL4	X10UP	***	***	***	ln(COL4 + 1) = 3.471–1.634X10UP	0.684
COL6	PPI	**	***	**	ln(COL6 + 1) = 4.632–0.235PPI–4.113X10UP + 2.462X0107UP	0.832
	X0107UP	**				
	X10UP	***				
COL8	X10UP	*	***	**	ln(COL8+1)=4.061-3.231X10UP	0.612
COL10	PPI	*	***	**	ln(COL10 + 1) = 6.586–0.994PNF–0.192PPI + 0.9357X0107UP–3.577X10UP	0.890
	PNF	*				
	X10UP	***				

Note: For the legend of the “Taxonomic group” column, see Figure 1.

Table 6 Results of the multiple linear regression analyses between environmental factors and abundance ln-transformed for each taxonomic group (continued)

Taxonomic group	*Environment variable*	*Statistical significance*			*Model*	R^2
		Env. variable Ln-transformed	*Intercept*	*Model*		
COL20	PPI	***	***	**	ln(COL20 + 1) = 5.435 + 1.232X0107UP–0.209PPI–1.824X10UP	0.811
	X0107UP	**				
	X10UP	**				
EMB	PNF	*		**	ln(EMB + 1) = –0.096 + 0.121X0107UP + 0.084PNF-0.008PPI + 0.082PPI	0.833
	X0107UP	(*)				
PSC	X0107UP	*	***	*	ln(PSC + 1) = 0.525 + 0.824X0107UP	0.377
E10	PPI	(*)		**	ln(E10 + 1) = –0.074–0.054PPI + 1.132X0107UP + 0.244PNF	0.837
	X0107UP	***				
TI	PNF	*	*	**	ln(TI + 1) =1.368 + 0.543X0107UP + 0.041PPI + 0.587X10UP–0.595PNF	0.837
CLT1	PPI	*	***	(*)	ln(CLT1+1) = 0.711 + 0.802X0107UP–0.088PPI–0.819X10UP	0.588
	X0107UP	*				
	X10UP	*				
CLT20	PPI	*	***	*	ln(CLT20 + 1) = 0.591–0.105PPI–1.171X10UP + 0.996X0107UP	0.644
	X0107UP	*				
	X10UP	*				
IM1	PPI	*	*	(*)	ln(IM1 + 1) = 0.584–0.036PPI–0.154PNF	0.460
DLR	PPI	**	***	***	ln(DLR + 1) = 4.069–0.768PNF–0.210PPI + 1.432X0107UP + 0.724X10UP	0.917
	PNF	*				
	X0107UP	*				
AN	PNF	*	***	**	ln(AN + 1) = 4.683–1.477PNF	0.544
NE	PPI	**	***	***	ln(NE + 1) = 4.267–0.172PPI–0.849PNF + 1.370X10UP + 0.862X0107UP	0.907
	PNF	*				
	X10UP	*				

Note: For the legend of the "Taxonomic group" column, see Figure 1.

The goodness of the solution produced by the analysis tested through the overall permutation test was statistically significant. This indicates the effectiveness of the environmental variables chosen in explaining most of the variability in the bioindicator community (Table 4). The two first canonical axes are statistically significant at the permutation test, indicating that the corresponding environmental gradients fit very well with the abundance distribution of the taxa categories used as bioindicators.

The result of the partial RDA (Table 5) shows that the soil fauna of the sampled hazelnut orchards soils is currently distributed mainly along the gradient represented by the set of natural variables and mechanical agronomic practices (conditional inertia: 38.55%), mainly characterising axis 1. At the same time, it highlights that the effect of PPP, mainly characterising axis 2, assumed a considerable importance in differentiating it (constrained inertia 24.71%), with a statistically significant permutation test ($P < 0.05$), which is likely to increase with a PPP prolonged use over time.

The developed methodology has proven to be effective as it reduces the 'background noise' determined by agronomic management and by environmental and anthropogenic variables. Therefore, the qualitative and quantitative effects of the PPP on the soil fauna sampled in hazelnut orchards could be clearly appreciated.

The work has highlighted that the use of PPP has a considerable impact on the soil fauna of hazelnut orchards. PPI index and PNF variables have proven to be functional in detecting the impacts on biodiversity of PPP in hazelnut orchards, having noted for them an evident sensitivity by most of the soil fauna groups used as bioindicators.

Table 6 shows the 26 taxa (over the 41 taxa analysed) whose models were found to be statistically or almost significant (threshold at $p < 0.07$) by multiple linear regression analysis with stepwise backward selection of the explanatory variables. The models for Pauropoda (PAU), Symphyla (SIN), Diplura (DIPLU), Collembola4 (COL4), Collembola8 (COL8) and Psocoptera (PSC) do not incorporate variables related to the use of PPP but only the variable associated with the soil grain size (X10UP), except for the Psocoptera. Therefore, a poor capacity of bioindication could be attributed to these taxa relative to treatment with PPP. On the other hand, Acari (ACA), Collembola2 (COL2), Thysanoptera (TI), Hymenoptera1 (IM1) and Annelida (AN) incorporate variables related to the use of PPP in their models, showing a presumably good bioindication capacity. The remaining 15 taxa models require both the variables associated with the use of PPP and other covariates.

4 Conclusions

This work defines a wide range of covariates to be considered in monitoring the effects of the use of PPP as target explanatory variable, on species and habitats related to agroecosystems.

The system used for detecting and recording covariates in agricultural contexts represents the baseline for the development of standard protocols to collect comparable data and ensure the replicability of the study on biological communities linked to agroecosystems. This methodology has already been tested in the activities of insect pollinators monitoring in the National Parks funded by the Italian MASE (Ministry Directives 2020, 2021, 2022). In particular, it has been used in order to have comparable datasets at a national level and share them at the European level according to the EU

Pollinator Initiative (Italian Ministry of Environmental and Energy Security, 2023; European Union, 2023c).

Therefore, the sampling of the most significant covariates, selected based on the statistical analysis carried out, is strongly recommended in monitoring activities, especially in order to assess the impacts of PPP on non-target species or biological communities and to define appropriate management measures aimed at mitigating their effects on biodiversity. This type of monitoring will also be increasingly necessary to verify the effects on biodiversity of the PPP in order to reduce their use, especially in protected areas and Nature 2000 sites, according to the European Biodiversity Strategy for 2030 and of the Farm to Fork Strategy (European Union, 2023a, 2023b). In fact, both provide for the reduction of the use of PPP, in particular those most dangerous for human health, environment and biodiversity.

For example, the European Pollinators Initiative (European Union, 2023c) addresses the decline of pollinators emphasising the importance of increasing monitoring of certain taxa (e.g., bees, butterflies, hoverflies) in terms of richness and abundance. In this context, considering environmental and management covariates that may influence the status and trends of pollinator populations as well as the presence of PPP in the environment, could provide useful information for an integrated approach to define conservation actions and policies (Hermoso et al., 2022). Indeed, within the same project we highlighted that insect pollinators abundance was conspicuously higher in organic fields compared to conventional ones (Bonelli et al., 2020; D'Antoni et al., 2020).

The proposed method is easy to implement as data on covariates needs to be recorded in categories and classes and does not have to be measured precisely. However, we underline that one limitation is that the method is time consuming as it requires a long list of environmental and agronomic variables that have an influence on species and habitats related to agroecosystems.

Future research lines will concern the identification of the variables, through statistical analyses on covariates and bioindicators already sampled, that have the greatest effect on each taxonomic group. Thus, in further ecological studies, sampling only the set of covariates that has a significant effect on the selected bioindicator will be possible. Moreover, since different taxonomic groups of soil fauna have shown a different bioindication capacity concerning treatments, it may be possible in the future to optimise the bioindication capacity by appropriately selecting subsets of more sensitive taxa based on the type of crop and the explanatory variable of interest.

Acknowledgements

The authors would like to thank Laura Pettiti (Italian Ministry of the Environment and Energy Security) for her precious support in the definition of the project. Special thanks are given to: Simona Bonelli (University of Turin) for significantly contributing to the project and identifying the covariates; Francesca Floccia, Cecilia Lorusso, Stefania Mandrone (ISPRA) for the collection and identification of soil fauna; Giorgio Vizzini (ISPRA) for the collection and analyses of soil data; Roberto Crosti (ISPRA) for revising the manuscript.

References

Akaike, H. (1973) 'Information theory and an extension of the likelihood principle', in Petrov B.N. and Csaki F. (Eds.): *International Symposium on Information Theory*, pp.267–281, Akadémiai Kiadò, Budapest.

Bispo, A., Grand, C. and Galsomies, L. (2009) 'Le programme ADEME 'Bioindicateurs de qualité des sols'', *Étude Gest Sols*, Vol. 16, Nos. 3/4, pp.145–158.

Bonelli, S., Vercelli, M., Audisio, M., Ferri, V, Bellucci, V., D'Antoni, S., Gori, M., Lorusso, J. and Mattoccia, M. (2020) 'Experimentation of measures to implement the National Action Plan for the sustainable use of PPP in Natura 2000 sites and protected areas: preliminary results on pollinators', Poster presented at *Halting the loss of Pollinators - Role of EU Agricultural and Regional Development Policies, I.E.E.P., Committee of the Regions*, 21 February 2020, Brussels.

Botias, C., David, A., Hill, E.M. and Goulson, D. (2017) 'Quantifying exposure of wild bumblebees to mixtures of agrochemicals in agricultural and urban landscapes', *Environmental Pollutution*, Vol. 222, pp.73–82, Epub 2017 Jan 11. PMID: 28087090, doi: 10.1016/j.envpol.2017.01.001.

Coleman, D.C. and Whitman, W.B. (2005) 'Linking species richness, biodiversity and ecosystem function in soil systems', *Pedobiologia*, Vol. 49, No. 6, pp.479–497.

Cunningham, M.M., Tran, L., McKee, C.G., Ortega Polo, R., Newman, T., Lansing, L., Griffiths, J.S., Bilodeau, G.J., Rott, M. and Guarna, M.M. (2022) 'Honey bees as biomonitors of environmental contaminants, pathogens, and climate change', *Ecological Indicators*, Vol. 134, p.108457, http://dx.doi.org/10.1016/j.ecolind.2021.108457.

D'Antoni, S., Bonelli, S., Gori, M., Macchio, S., Maggi, C., Nazzini, L., Onorati F., Rivella, E. and Vercelli, M. (2020) *La sperimentazione dell'efficacia delle Misure del Piano d'Azione Nazionale per l'uso sostenibile dei prodotti fitosanitari (PAN) per la tutela della biodiversità*, ISPRA, 330/2020/REP.

Edwards, C.A., Subler, S., Chen, S.K. and Bogomolov, D.M. (1996) 'Essential criteria for selecting bioindicator species, processes or systems to assess the environmental impact of chemicals on soil ecosystems', in Van Straalen, N.M. and Krivolutsky, D.A. (Eds.): *Bioindicator Systems for Soil Pollution,* pp.67–84, Kluwer Academic Publishers, Dordrecht.

Eijsackers, H. (1983) 'Soil fauna and soil microflora as possible indicators of soil pollution', *Environmental Monitoring and Assessment*, Vol. 3, Nos. 3–4, pp.307–316.

European Union (2023a) *Biodiversity Strategy for 2030* [online] https://environment.ec.europa.eu/strategy/biodiversity-strategy-2030_en (accessed 1 February 2023).

European Union (2023b) *Farm to Fork Strategy* [online] https://food.ec.europa.eu/horizontal-topics/farm-fork-strategy_en (accessed 5 February).

European Union (2023c) 'Revision of the EU pollinators initiative', *A new deal for pollinators, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, *Communication - COM/2023/35 Final* [online] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2023%3A35%3AFIN&qid=1674555285177 (accessed 9 February 2023).

Galli, L., Capurro, M. and Torti, C. (2011) 'Protura of Italy, with a key to species and their distribution', *ZooKeys*, Vol. 146, pp.19–67, Epub 2011 Nov 9. PMID: 22207788; PMCID: PMC3233705, doi: 10.3897/zookeys.146.1885.

Galli, L., Capurro, M., Molyneux, T., Tortia, C. and Zinni M. (2019) 'Ecology of Italian Protura', *Pedobiologia*, Vol. 73, pp.20–28, ISSN 0031-4056, https://doi.org/10.1016/j.pedobi.2019.01.004

Gerlach, J., Samways, M. and Pryke, J. (2013) 'Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups', *Journal of Insect Conservation*, Vol. 17, No. 4, pp.831–850.

Girotti, S., Ghini, S., Maiolini, E., Bolelli, L. and Ferri, E.N. (2013) 'Trace analysis of pollutants by use of honeybees, immunoassays, and chemiluminescence detection', *Analytical and Bioanalytical Chemistry*, Vol. 405, Nos. 2–3, pp.555–571.

Górny, M. and Grüm, L. (1993) *Methods in Soil Zoology*, Elsevier Science, Amsterdam.

Hermoso, V., Carvalho, S.B., Giakoumi, S., Goldsborough, D., Katsanevakis, S., Leontiou, S., Markantonatou, V., Rumes, B., Vogiatzakis, I.N. and Yates, K.L. (2022) 'The EU Biodiversity strategy for 2030: opportunities and challenges on the path towards biodiversity recovery', *Environmental Science and Policy*, Vol. 127, pp.263–271, ISSN 1462-9011, https://doi.org/10.1016/j.envsci.2021.10.028.

Hopkin S.P. (1997) *Biology of the Springtails*: (Insecta: Collembola), Oxford University Press, Oxford.

Italian Ministry of Agricultural, Food and Forestry Policies (2022) *Database of Plant Protection Products* [online] https://www.sian.it/mimfFitoPub/home.get (accessed 27 November 2022).

Italian Ministry of Environmental and Energy Security (2023) *Biodiversity Directive* [online] https://www.mase.gov.it/sites/default/files/direttiva_biodiversita.pdf (accessed 13 January 2023).

Karr, J.R. (1981) 'Assessment of biotic integrity using fish communities', *Fisheries*, Vol. 6, No. 6, pp.21–27.

Legendre, P. and Legendre, L. (1998) *Numerical Ecology*, 2nd ed., Elsevier Science, Amsterdam.

Lepš, J. and Šmilauer, P. (2003) *Multivariate Analysis of Ecological Data using CANOCO*, Cambridge University Press, New York.

Li, Y., Wu, J., Huili, C. and Jiakuan, C. (2005) 'Nematodes as bioindicator of soil health: methods and applications', *Chin. J. Appl. Ecol*., Vol. 16, No. 8, pp.1541–1546.

López-López, E. and Sedeño-Díaz, J.E. (2014) 'biological indicators of water quality: the role of fish and macroinvertebrates as indicators of water quality', in Armon, R.H. and Hanninen, O (Eds.): *Environmental Indicators*, pp.643–661.

Mateos, E., (2016) 'La fauna del sòl i les xarxes tròfiques edàfiques', *L'Atzavara*, Vol. 26, pp.15–24 [online] https://raco.cat/index.php/Atzavara/article/view/307593.

McGeoch, M.A. (1998) 'The selection, testing and application of terrestrial insects as bioindicators', *Biological Reviews*, Vol. 73, pp.181–201, https://doi.org/10.1111/j.1469-185X.1997.tb00029.x.

Monteagudo, L. and Moreno, J.L. (2016) 'Benthic freshwater cyanobacteria as indicators of anthropogenic pressures', *Ecological Indicators*, Vol. 67, pp.693–702, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2016.03.035.

Parisi, V. (2001) 'La qualità biologica del suolo.Un metodo basato sui microartropodi / Parisi, Vittorio', in *Acta Naturalia De L'ateneo Parmense*, Vol. 37, pp.105–114, ISSN 0392-419X.

Previati, E., Fano, E.A. and Leis, M. (2007) 'Arthropods biodiversity in agricultural landscapes: effects of land use and anthropization', *Italian Journal of Agronomy*, Vol. 2, No. 2, pp.135–141.

Previati, E., Fano, E.A. and Leis, M. (2007) 'Arthropods biodiversity in agricultural landscapes: effects of land use and anthropization', *Italian Journal of Agronomy*, Vol. 2, No. 2, pp.135–141.

Rismondo, M., Lancioni, A. and Taffetani, F. (2011) 'Integrated tools and methods for the analysis of agro-ecosystem's functionality through vegetational investigations', *Fitosociologia*, Vol. 48, No. 1, pp.41–52.

Smallshire, D. and Beynon T. (2010) *Dragonfly Monitoring Scheme Manual*, British Dragonfly Society.

Smeets, E. and Weterings, R. (1999) *Environmental Indicators: Typology and Overview*, p.19, European Environment Agency, Technical report No. 25.

Stahlschmidt, P. and Brühl, C.A. (2012) 'Bats as bioindicators–the need of a standardized method for acoustic bat activity surveys', *Methods in Ecology and Evolution*, Vol. 3, No. 3, pp.503–508.

Taffetani, F. and Rismondo, M. (2009) 'Bioindicator system for the evaluation of the environmental quality of agro-ecosystems', *Fitosociologia*, Vol. 46, No. 2, pp.3–22.

ter Braak, C.J.F. and Verdonschot, P.F.M (1995) 'Canonical correspondence analysis and related multivariate methods in aquatic ecology', *Aquatic Science*, Vol. 57, pp.255–289.

Tourinho, A.L. and Lo-Man-Hung, N. (2021) 'Standardized Sampling Methods and Protocols for Harvestman and Spider Assemblages', in Santos, J.C. and Fernandes, G.W. (Eds.): *Measuring Arthropod Biodiversity*, Springer, Cham.

van Straalen, N.M. and Krivolutsky, D.A. (Eds.) (1996) *Bioindicator Systems for Soil Pollution*, Kluwer Academic Publishers, Dordrecht.

Annex 1

List of environmental and agronomic variables that have an influence on species and habitats related to agroecosystems. In order to be clearly described, they have been grouped into coherent topics. Further descriptors have also been calculated from the original variables. All methods of measurement, classification, standardisation and ranking are described for each variable in the following table. The annex is a tabular representation of all the field data sheets.

Annex 1 (continued)

Variable	*Description*
	Topic 1 records data relating to farms and crops and the agronomic practices
Agricultural system	Agricultural management system: ORG = organic; CONV = conventional
Field ID	Identification code of the monitored field
Farm surface	Farm surface (hectare, ha) ranked in 5 classes: 1) ≤ 4 ha; 2) 4–16 ha; 3) 16–64 ha; 4) 64–144 ha; 5) ≥ 144 ha
Crop surface	Crop surface (hectare, ha) ranked in 5 classes: 1) ≤ 0.25 ha; 2) 0.25–1 ha; 3) 1–4 ha; 4) 4–16 ha; 5) > 16 ha
Field surface	Field surface (hectare, ha) ranked in 5 classes: 1) ≤ 0.25 ha; 2) 0.25–1 ha; 3) 1–4 ha; 4) 4–16 ha; 5) > 16 ha
Ratio between crop and farm surface	Ratio between total crop surface and total farm surface
Ratio between field and farm surface	Ratio between monitored field surface and farm surface
Ratio between field and crop surface	Ratio between monitored field and total crop surface
Difference between crop and field surface	Difference between total crop surface and monitored field surface
Protected areas	The field is cultivated in a protected area
	1) yes 2) no 3) partially
Nature 2,000 sites	The field is cultivated in a Nature 2000 sites
	1) yes; 2) no; 3) partially
Altitude	Field altitude (metres, m) ranked in 5 classes:
	1) ≤ 100 m; 2) 100–300 m; 3) 300–500 m; 4) 500–700 m; 5) > 700 m
Slope	Field slope (degrees, °) ranked in 5 classes
	1: ≤ 3°; 2: 3–6°, 3: 6–10°; 4: 10–20°; 5: > 20°. The data can be also indicated in percentage (%)
Exposure	Field exposure to the cardinal points
	NN: North; NE: North-East; EE: East; SE: South-East; SS: South; SW: South-West; WW: West; NW: North-East
Geomorphological position	Position of the field in relation to the surrounding area, ranked in 4 classes
	1) the field is on a summit; 2) the field is on a slope 3) the field is on a plain 4) the field is in a depression
Beehives	Presence of beehives in the farm, 1) yes 2) no
Crop age	Crop age (years), ranked in 5 classes: 1) ≤ 2 years; 2) 2–5 years; 3) 5–10 years; 4) 10–20 years; 5) > 20 years
Rice field floods per year	Number of rice flooding, ranked in 5 classes
	1) 1 flood; 2) 2 floods; 3) 3 floods; 4) 4 floods; 5) 5 or more floods

Annex 1 (continued)

Variable	*Description*
Planting pattern	Scheme of planting pattern
	Q = quadrilateral system; T = triangular or quincunx system; M: mixed
Planting distance	Inter-row and intra-row distances (metres, m)
Planting pattern surface	Planting pattern surface (Inter-row distance x intra-row distance, square metres, m2)
	ranked in 5 classes: 1): ≤ 6 m^2; 2) 6–12 m^2; 3) 12–24 m^2; 4) 24–36 m^2; 5) > 36 m^2
Mowing/shredding	Number of mowing or shredding per year, ranked in 5 classes
	1) 1 operation; 2) 2 operations; 3) 3 operations; 4) 4 operations; 5) 5 or more operations
Mechanical machinery passages	Number of mechanical machinery passages, ranked in 5 classes
	1) ≤ 3; 2) 3–6, 3) 6–10; 4) 10–15; 5) > 15
Most frequently used machinery weight	Weight of the most frequently used machinery (tons, t), ranked in 5 classes
	1) ≤ 20; 2) 20–50, 3) 50–250; 4) 250–1,000; 5) > 1,000
Heaviest machinery weight	Weight of the heaviest used machinery (tons, t), ranked in 5 classes
	1) ≤ 20; 2) 20–50; 3) 50–250; 4) 250–1,000; 5) > 1,000
Tyres or caterpillar tracks	It indicates whether machineries used in the field move on tyres or on caterpillar tracks, adopting 4 classes
	1) no machinery is used; 2) use of tyres; 3) use of tyres and caterpillar tracks; 4) caterpillar tracks
Length of the tyres or caterpillar tracks	Length of tyres or caterpillar tracks in contact with the ground (metres, m)
Width of the tyres or caterpillar tracks	Width of tyres or caterpillar tracks in contact with the ground (metres, m)
Surface of the tyres or caterpillar tracks	Contact surface of tyres or caterpillar tracks on the ground (square metres, m^2)
Machinery compression on the ground	Weight, per unit of area, of the most frequently used machinery. It is calculated considering the contact surface of tyres or caterpillar track on the ground
Number of fertilisations	Number of field fertilisations per crop season, ranked in 5 classes
	1) 1 fertilisation; 2) 2 fertilisations; 3) 3 fertilisations; 4) 4 fertilisations; 5) 5 or more fertilisations
Type of fertilisers	Type of fertilisers used in the field, ranked in 5 classes
	1) no fertilisers used; 2) organic fertilisers; 3) mixed fertilisers (organic and mineral or organic and chemical); 4) mineral fertilisers; 5) chemical fertilisers
Use of plant protection products	1) yes; 2) no
Treatments	Number of treatments with plant protection products per crop season
Category of plant protection products used	To be selected among the following categories: nematocide, acaricide, algicide, bactericide, herbicide, insecticide, fungicide, slugicide, other
Most used plant protection products	Names of the 4 most used plant protection products in the monitored field
Type of spraying treatment system for herbaceous crops	To be selected among the following categories: atomiser, sprayer bars, sprayers of other types for herbaceous crops, sprayers for weeding and sowing, other
Type of spraying treatment system for tree crops	To be selected among the following categories: air-assisted atomisation, pneumatic atomisation, atomisation by pressure

Annex 1 (continued)

Variable	*Description*
Anti-drift systems used in spraying	1) yes; 2) no. If present, the type must be indicated (from a provided list)
Cover crop	Number of different crops used for cover cropping, ranked in 5 classes: 1) 1 crop; 2) 2 crops; 3) 3 crops; 4) 4 crops; 5) no crop
Minimum tillage	1) yes; 2) no
Irrigation system	1) yes; 2) no
Type of irrigation system	To be selected among the following categories: no irrigation; surface irrigation; furrow Irrigation; drip irrigation; sprinkler irrigation; centre pivot irrigation; sub-irrigation; other system
Topic 2 records information on soil tillage	
Tillage of the soil and vegetation	To be selected among the following categories: mechanical plowing or hoeing; soil topping; chemical weeding; slicing; soil harrowing; soil grubbing; fertilisation; milling; lasering; levelling; soil morganisation; minimum tillage; pruning; ripping; rolling; mowing by hand; mechanical mowing; flame weeding; steam mowing; shredding; water in the rice field; cover crop sowing 1; cover crop sowing 2; cover crop sowing 3; cover crop harvesting 1; cover crop harvesting 2; cover crop harvesting 3
Depth of tillage	Measure expressed in centimetres, cm
Topic 3 collects data regarding timing and quantities of treatments with plant protection products	
Plant protection products associated with date and concentrations	Date of treatment with plant protection products; name of commercial products (from a provided list) and relative concentration adopted in treatments
Pathogens and target parasites	Species for which treatments are carried out (from a provided list)
Topic 4 is about timing and quantities of treatments with fertilisers	
Fertilisers associated with date and concentrations	Date of treatment with fertilisers; name of commercial products (from a provided list) and relative concentration adopted
Topic 5 collects the presence and size of agricultural annexes and the presence of any crops in a 10 metres buffer around the perimeter of the monitored field	
Presence of dry stone walls	1) yes; 2) no
Dry stone wall measurements	Estimation of linear development, height and thickness (metres, m)
Presence of rows of trees	1) yes; 2) no
Rows of trees measurements	Estimation of linear development and height (metres, m)
Presence of isolated trees	1) yes; 2) no
Presence of hedges	1) yes; 2) no
Hedges measurement	Estimation of linear development, height and thickness (metres, m)
Neighbouring crops	List and number of the crops present in the 10 metres buffer
Field orientation	Orientation of each field side to the cardinal points
Plant protection products input in neighbouring fields (PNF)	Use of plant protection products in neighbouring fields 1) yes; 2) no

Annex 1 (continued)

Variable	*Description*
Topic 6 records the coverage of EUNIS land use categories in a 10 metres buffer around the perimeter of the monitored field. For the EUNIS categories relating to tree vegetation, the measurement of basic structural parameters (trunk diameter and height) and the assessment of the maturity of the formation in classes were also envisaged	
Presence of water bodies	1) yes; 2) no
Types of water bodies	ranked in 5 classes: 0: Unknown; 1) stream, river and lake > 1 ha; 2) pond and natural lake (0.5–1 ha); 3) artificial canal and artificial lake (0.5–1 ha); 4) artificial puddle < 0.5 ha; 5) no water body
Soil cover	Percentage of soil cover for each of the following categories: water bodies; agricultural outbuildings; roads and paths, houses, other urban areas; rocks; decorticated soil; natural herbaceous vegetation; shrubbery; arboreal vegetation; EUNIS-type vegetation (up to a maximum of 5), ranked in 6 classes: 0: 0%; 1: < 10%; 2: 10–30%; 3: 30–50%; 4: 50–70%; 5: ≥ 70%
Number of EUNIS categories	EUNIS diversity synthesis indicator obtained by counting how many EUNIS habitats are present (from 1 to 5. If more, only the 5 widest are counted)
Topic 7 gathers the variables relating to the flowering and fruiting as well as the structure of the crop and of any natural vegetation within the monitored field and in the 10 metres buffer	
Natural vegetation within the field	Percentage of soil cover for each of the following categories of natural vegetation: herbaceous vegetation, shrubbery, arboreal vegetation. Ranked in 6 classes: 0) 0%; 1) < 10%; 2) 10–30%; 3) 30–50%; 4) 50–70%; 5) ≥ 70%
Flowering and fruiting in the buffer	Soil cover percentage in the 10m buffer for flowering and fruiting, estimated for each vegetation layer (herbaceous, shrubbery and arboreal vegetation), ranked in 6 classes: 0) 0%; 1) < 10%; 2) 10–30%; 3) 30–50%; 4) 50–70%; 5) ≥ 70%
Most frequent types of fruit in the buffer	To be selected among the following categories: none; big and juicy fruit (> 2 cm); small pods (seed pods < 5 cm); big pods (seed pods > 5 cm); grain (cereal); small and dried fruit (< 1.5 cm, for example sunflower); samaras (small, with flattened wings of fibrous and papery tissue); big and hard-coated fruits (> 1,5 cm, for example hazelnut); acorn; chestnut; cones of deciduous trees; cones of conifers
Entomophilous plant species	Percentage of soil cover of entomophilous species in bloom, estimated for each vegetation layer (herbaceous, shrubbery and arboreal vegetation), both within the field and in the buffer. All ranked in 4 classes: 0) <10%; 1) 10–30%; 2) 30–50%; 3) ≥50%
Distribution of entomophilous plant species in bloom	It counts how many of the 6 vegetational layers considered the entomophilous plant species in bloom are present. The 6 layers considered are: grasses, shrubs and trees both in the field and in the buffer.
Crop and natural vegetation structure	Height (in centimetres, cm) of the crop within the field and height of the natural vegetation both within the field and in the 10m buffer (assessed in the herbaceous, shrub and tree layer). For tree vegetation the diameter of the trunks was also considered
	For the height the following ranks are attributed: Height – Herbaceous vegetation: 5: < 10 cm; 4: 10–30cm; 3: 30–50cm; 2: 50–100cm; 1: >100cm
	Height – Shrubs: 5: < 50cm, 4: 50–100 cm; 3: 100–200 cm; 2: 200–300 cm; 1: >300 cm
	Height – Trees: 5: < 150cm; 4: 150–300 cm; 3: 300–600 cm; 2: 600–1,000 cm; 1: > 1,000 cm
	For the diameter of the trunks of the arboreal vegetation the following ranks are attributed:
	Trunk diameter: 1: ≤ 15 cm; 2: 15–30 cm; 3: 30–50 cm, 4: 50–100 cm; 5: > 200 cm
Mowing in the buffer	1) no mowing or mowing by hand; 2) mechanical mowing; 3) steam mowing or other type (to be specified in notes); 4) flame weeding; 5) chemical weeding
Shredding and rolling in the buffer	Shredding (1: yes; 2: no), rolling (1: yes; 2: no)
Heterogeneity of natural vegetation in the buffer	The heterogeneity in the heights and in the maturity of the natural vegetation within the 3 vegetation layers (grasses, shrubs and trees) has been estimated. Regarding the maturity, it has been assessed measuring the diameter of the trunks of shrubs and trees. It is ranked as follow:
	For the herbaceous vegetation: 'heterogeneous heights' (rank = 1) when at least 20% of the grasses show different heights from the rest, otherwise it's 'uniform heights' (rank = 2)
	For the shrub and arboreal vegetation: 'heterogeneous' (rank = 1) when at least 20% show different heights and diameters from the rest; 'intermediate' (rank = 2) when heights are all uniform but at least 20% of the plants have a different diameter from the rest; 'uniform' (rank = 3) when more than 80% of the plants show uniform height and diameter

Annex 1 (continued)

Variable	*Description*
Topic 7 gathers the variables relating to the flowering and fruiting as well as the structure of the crop and of any natural vegetation within the monitored field and in the 10 metres buffer	
Maturity of arboreal vegetation in the buffer	Ranked in 6 classes: 1) high forest of different age; 2) high forest of the same age and mature coppice; 3) aged coppice of different age; 4) aged coppice of the same age; 5) young coppice of different age; 6) young coppice of the same age
Topic 8 collects the meteorological data recorded before and during the sampling events	
Time interval of weather monitoring	Time interval in which the weather has been monitored before sampling, expressed in classes. 0D: the same day of the sampling; 1D: one day before the sampling; 3D: 3 days before the sampling; 1W: one week before the sampling; 2W: weeks before the sampling; 1M: one month before the sampling; 2M: 2 months before the sampling; 3M: 3 months before the sampling
Last precipitation (days)	Number of days since the last precipitation
Last precipitation (hours)	If the precipitation occurred on the same day as the sampling, indicate from how many hours it has stopped
Sky cover	Percentage ranked in 6 classes: 1) 0%; 2) < 10%; 3) 10–30%; 4) 30–50%; 5) 50–70%; 6) ≥ 70%
Wind speed	Evaluation following the Beaufort scale or anemometric measurement
Wind origin	Referred to the cardinal points
Air temperature and humidity	Temperature expressed in degrees Celsius (°C) and humidity in percentage (%)
Topic 9 gathers the soil data and other variables directly detected at the sampling points. The surface considered for data collection is a circle centred at the sampling point and having 1m radius	
Slope at the sampling point	Expressed in degrees or percentage
Soil temperature and humidity	Temperature (in degrees Celsius) and humidity (in percentage) on the soil surface and at 10 cm depth
Consistency of any surface crust	Ranked in 5 classes: 0 = absent; 1 = weak: soft or slightly hard crust, thickness < 5 mm; 2 = soft moderate: soft or slightly hard crust, thickness > 5 mm; 3 = hard moderate: hardened crust with a thickness < 5 mm; 4 = strong: hardened crust with a thickness > 5 mm
Resistance to penetration	The measurement is carried out using a penetrometer or with an empirical estimate scale based on the effort made to penetrate the ground at a depth of 10 cm with a spade of standardised dimensions, ranked in 4 classes: 1= mild: sinks easily using one hand without effort; 2 = moderate: it sinks with only one hand making a certain effort but without the need to support the weight of the body; 3 = consistent: consistent effort is required, using 2 hands and/or leaning part of the body weight; 4 = strong: it is essential to use 2 hands and the weight of the whole body
Adhesion	Ranked in 4 classes: 1 = none: by applying pressure between thumb and forefinger and then separating the fingers: no soil particles adhere; 2= weak: the sample adheres to both the thumb and the forefinger in a clearly perceptible way, but when the fingers separate it tends to detach clearly from one or the other and does not extend appreciably; 3 = enough: the sample clearly adheres to both the thumb and forefinger and tends to extend until it comes off on one side instead of both; 4 = a lot: the sample adheres so strongly between the thumb and forefinger that when the fingers are separated, it definitely tends to stretch, until it breaks partly on the thumb and partly on the forefinger
Plasticity	Ranked in 4 classes: 1 = none: rolling the cylinder between thumb and forefinger it is not possible to form a cylinder 4 cm long and 6 mm thick; 2 = weak: it is possible to form a cylinder 4 cm long and 6 mm thick, which bears its own weight, but by decreasing the thickness to 4 mm, the cylinder cannot bear its own weight; 3 = enough: a 4 cm long and 4 mm thick cylinder can be formed, which bears its own weight, but a 2 mm thick cylinder cannot bear it; 4 = a lot: 4 cm long and 2 mm thick cylinder can be formed, which bears its own weight
Small stones	Percentage of surface covered with small stones, ranked in 5 classes: 1) < 5%; 2) 5–15%; 3) 16–35%; 4) 36–70%; 5) >70%
Medium-large stones	Percentage of surface covered with medium-large stones, ranked in 6 classes: 1)< 0.3%; 2) 0.3–1%; 3) 1.1–3%; 4) 3.1–15%; 5) 15.1–50%; 6) > 50%

Annex 1 (continued)

Variable	*Description*
Topic 9 gathers the soil data and other variables directly detected at the sampling points. The surface considered for data collection is a circle centred at the sampling point and having 1m radius	
Small rocks	Percentage of surface covered with small rocks, ranked in 6 classes: 1) 0; 2) 0–2%; 3) 2.1–10%; 4) 10.1–25%; 5) 25.1–50%; 6: >50%
Soil cover at the sampling point	Percentage of soil coverage by evaluating respectively the crop and the herbaceous, shrubby and tree layers of natural vegetation as well as areas without turf or with emerging rocks, each classified into 6 classes: 1) 0%; 2) <10%; 3) 10–30%; 4) 30–50%; 5) 50–70%; 6) ≥ 70%
Height and diameter of the crop at the sampling point	Average height of the crop and trunk diameter (major sucker) in the case of tree crops
Topic 10 records data of laboratory analysis of soil samples	
Granulometry by sieving	Percentage for each of the 10 dimensional classes (> 64 mm, 32–64 mm, 16–32 mm, 8–16 mm, 4–8 mm, 2–4 mm, 1–2 mm, 0.5–1 mm, 0.25–0.5 mm, 0.125–0.25 mm)
Granulometry by densitometry	Percentage for each of the 4 dimensional classes (> 0.063 mm, 0.05–0.063 mm, 0.02–0.05 mm, < 0.002 mm)
pH	soil pH
Salinity	Salinity of the soil, in millisiemens per centimetre (mS/cm), measured at the level of the upper horizon and below the surface, ranked in 4 classes: 1) < 2 mS/cm; 2) 2–4 mS/cm; 3) 4–8 mS/cm; 4) > 8 mS/cm)
Reaction to HCL 10%	Ranked in 6 classes: 0: none = no auditory or visual effect; 1: very weak = no visual manifestation, indistinct or moderately audible fizz on hearing; 2: weak(A) =a faint effervescence limited to single granules is barely visible, indistinct hearing or only moderately audible; 3: weak(B) = upon careful observation, a distinctly or only moderately audible faint general effervescence can be noted; 4: strong = a moderate effervescence is observed, easily audible and with bubbles up to 3 mm in diameter; 5: violent = strong effervescence is observed, easily visible and with bubbles up to 7 mm in diameter
Porosity	Percentage measured on a 10 x 10 x10 cm3 clod of soil
Topic 11 collects data from cartographic analysis (using Corine land cover inventory) carried out for field and field buffers of 10, 50, 100, 200 and 500 metres	
Number of water bodies	Main water bodies present in the buffer, up to a maximum of 4
Distance from the 5 main water bodies	Distance (metres, m) from the 5 main water bodies present in the 500 m buffer, ranked in 5 classes: 1) ≤ 50 m; 2) 50–150 m; 3) 150–250 m; 4) 250–500 m; 5) > 500 m
Types of water bodies	Ranked in 5 classes: 0) Unknown; 1) stream, river and lake > 1 ha; 2) pond and natural lake (0.5-1 ha); 3 artificial canal and artificial lake (0.5–1 ha); 4) artificial puddle < 0.5 ha; 5) no water body
Width of water bodies	Wet bed width of water bodies, ranked in 5 classes: 1) > 20 m; 2) 10–20 m; 3) 5–10 m; 4) 1–5 m; 5) ≤ 1 m
Length of intersection with water bodies	Measurement in metres (m) of the water bodies length in the buffer
Surface of intersection with water bodies	Surface (hectares, ha) of the water bodies intersected in each the buffer
Presence of bank vegetation	1) yes; 2) no
Number of different land use categories	Number of different categories of land use as defined by Corine Land Cover, up to a maximum of 4 (the largest ones by area)
Land use category	Main habitat types as defined by Corine Land Cover codes
Surface of land use category	Surface (hectares, ha) of each land use category as defined by Corine Land Cover

Annex 1 (continued)

Variable	*Description*
Topic 11 collects data from cartographic analysis (using Corine land cover inventory) carried out for field and field buffers of 10, 50, 100, 200 and 500 metres	
ercentage coefficient of variation of the surfaces of the Corine categories	Percentage coverage variability of the different land use categories in the buffer, up to a maximum of 4 (the largest ones by area). It gives an indication of the diversity
Ranking of the percentage coefficient of variation of the surfaces of the Corine categories	Rank of the percentage coverage variability of the different land use categories in the buffer, up to a maximum of 4 (the largest ones by area). Ranked in 5 classes: 1) ≤ 0.3; 2) 0.3–0.5; 3) 0.5–1.0; 4) 1.0–1.5; 5) > 1.5
Slope	Expressed in degrees (°), ranked in 5 classes: 1) ≤ 3°; 2) 3–6°; 3) 6–10°; 4) 10–20°; 5) > 20°. The data can be also indicated in percentage
Sides and perimeter	Length (metres, m) of each side of the field and its perimeter (metres, m)
Sides distance	Maximum distance (metres, m) between the sides of the field
Coefficient of variation of the lengths of the sides	Coefficient of percentage variation of the field side lengths, ranked in 5 classes: 1) ≤ 0.3; 2) 0.3–0.5; 3 0.5–1.0; 4) 1.0–1.5; 5) > 1.5
Form factor	Ratio between the area of the field and the area it would have if it had a perfectly circular shape (with the same perimeter)
Topic 12 collects indicators of use of plant protection products, developed during the research	
Plant protection products score (PPS)	The plant protection products score is calculated for each product on the basis of the toxic hazards to environment (N) and the safety precautions related to the environment (SPe) reported on the product label. Particular attention was paid to the impacts on aquatic ecosystems, aquatic organisms and birds. The products with the highest score are those candidates of substitution or banned, while those authorised by organic farming regulations and which have no precautionary or environmental hazard statements on the product label have the lowest score. Scores are given as follow: products candidate for replacement and banned: 1) environmental toxic products (N): 0,75; products with precautionary phrases for the environment (SPe) indicating 'possible impacts on aquatic ecosystems (Spe3 e Spe4) and for the birds (Spe7)' + phrases H400 'highly toxic to aquatic organisms' + H410 'highly toxic to aquatic organisms with long-term effects': 0.50, products with risk phrases for aquatic ecosystems (H411, H412, H413): 0,25; products authorised in the organic farming without precautionary or risk phrases for the environment: 0
Plant protection products index (PPI)	The index considers the number of different products used, the number of treatments and the level of toxicity for the environment of each product used, but not the quantities dispensed. The plant protection products index for field 'i' and crop season 'j' is calculated as $PPIij = \sum PPSij$