**UNIVERSITÀ DEGLI STUDI**
**DI MODENA E REGGIO EMILIA**

**Dottorato di ricerca in Scienze, tecnologie e biotecnologie agro-alimentari**

Ciclo XXXIII

# Quantitative models for biodiversity analysis in agroecosystems

**Candidato ANNA SIMONETTO**

Relatore (Tutor): Prof. GIANNI GILIOLI

Correlatore (Co-Tutor): Prof. LARA MAISTRELLO

Coordinatore del Corso di Dottorato: Prof. ALESSANDRO ULRICI

# INDEX

# Aims and structure of the thesis

This thesis proposes an exploration of some methodological and quantitative issues related to the assessment of biodiversity and its responses to environmental conditions and agronomic management processes in the context of agroecosystems study and management.

**Chapter one** introduces the topic of biodiversity assessment in the broader context of the scientific and applicative interest of linking biodiversity and ecosystem services. Then, this chapter explores the key elements for a methodology supporting the link between structural biodiversity to the functional properties of ecological networks. Emphasis is given to mathematical framework suitable for representing complex functional networks and their dynamics properties.

**Chapter two** is devoted to the analysis species' niche space with emphasis to the responses of individual taxa to abiotic environmental variables and to land use. In particular, quantitative models are applied to study the link between taxon characteristics and environmental determinants at landscape level. The choice of the quantitative approach had to consider and integrate two important objectives: the development of interpretative models of the phenomenon (allowing to understand the factors underlying the process) and obtaining a high predictive capacity of the model. The constraints were given by the peculiarities of the system: strongly unbalanced dataset, since only about 10% of the observations (on average) have the characteristic of interest, collinearity of the predictor variables and possible high nonlinearity of the relationships between taxon characteristic and environmental determinants. The case study concerns the estimation of the model of habitat suitability for *Popillia japonica*. The study (in the publication phase) is particularly relevant for the agri-food sector of Northern Italy. *P. japonica* is an invasive species recently introduced in Italy that seriously threat crop production and to some typical agri-food productions of Northern Italy (e.g., flower farming). The habitat suitability of the Japanese beetle is little known and has never been investigated in the Italian environmental conditions.

The **third chapter** presents models allowing the analysis of the impact of environmental determinants and management strategies on species pattern of presence and co-occurrence at farmscape level. Case studies refer to the analysis of biodiversity of arthropods in the vineyard agroecosystem. In the first case presented, we investigated how environmental conditions and farming systems (organic compared to conventional farming) influence the composition of the soil arthropod community in different Italian pedoclimatic contexts, including the major wine-growing areas of our country. The second case study is particularly relevant given the number of samples collected in a fairly limited wine-growing area. The combination of quantitative methodologies used made it possible to characterize the edaphic biocenoses in terms of co-occurrence presence pattern and to assess niches preferences of the microarthropod in terms of environmental variables. The peculiarity of the approach used was to apply techniques known in the literature for the reduction of the dimensionality of the data and the ranking of

observations in hyperspaces in order to characterize the relationship between environmental determinants and species pattern of co-occurrence. Furthermore, these methodologies have been integrated with data mining techniques in order to increase the interpretability of the results.

The **fourth** chapter explores quantitative tools to support the network analysis approach to biodiversity assessment. The chapter begins with a brief excursus on the conditions that made possible the development and development of quantitative tools to support network analysis. The fundamental characteristics common to all quantitative models for the network approach are then presented. In the second part of the chapter the two fundamental modelling categories for the assessment of biodiversity are described: machine learning and models for latent variables. For both the categories, the most widespread models offering the best prospects for investigating ecological systems are presented. For each model, the essential characteristics, advantages and possible limitations in their use and the main research lines in which they are currently used are described.

Key words: Biodiversity, functional traits, multidimensionality, arthropods, ecosystem services

# Obiettivi e struttura della tesi

Questa tesi propone un'esplorazione di alcune questioni metodologiche e quantitative relative alla valutazione della biodiversità e delle sue risposte alle condizioni ambientali e ai processi di gestione agronomica nel contesto dello studio e della gestione degli agroecosistemi

Il **primo capitolo** introduce il tema della valutazione della biodiversità nel contesto più ampio dell'interesse scientifico e applicativo della relazione esistente tra biodiversità e servizi ecosistemici. In questo capitolo vengono esplorati gli elementi chiave per lo sviluppo e l'adozione di metodologie a supporto dell'analisi del legame tra la biodiversità strutturale e le proprietà funzionali delle reti ecologiche. In particolare, l'attenzione è posta sulla struttura matematica adatta a rappresentare reti funzionali complesse e le loro proprietà dinamiche.

Il **secondo capitolo** è dedicato all'analisi dello spazio di nicchia delle specie, focalizzando l'attenzione sulle risposte dei singoli taxa alle variabili ambientali abiotiche e all'uso del suolo. In particolare, vengono applicati modelli quantitativi per studiare il legame tra le caratteristiche del taxon considerato e alcuni determinanti ambientali a livello di paesaggio (landscape). Due importanti obiettivi hanno determinato la scelta dell'approccio quantitativo utilizzato nell'analisi: lo sviluppo di modelli interpretativi del fenomeno (che permettano di comprendere i fattori alla base del processo) e l'esigenza di avere un'elevata capacità predittiva del modello. I vincoli sono stati determinati dalle peculiarità del sistema: un dataset fortemente 'unbalanced', poiché mediamente solo il 10% circa delle osservazioni presentava la caratteristica di interesse, la collinearità delle variabili predittive e una potenzialmente elevata non linearità delle relazioni tra la caratteristica del taxon e le determinati ambientali. Il caso studio riguarda la stima del modello di idoneità dell'habitat per *Popillia japonica*. Lo studio (in fase di pubblicazione) è particolarmente rilevante per il settore agroalimentare del Nord Italia. *P. japonica* è una specie invasiva recentemente introdotta in Italia che minaccia gravemente la produzione agricola e alcune produzioni agroalimentari tipiche del Nord Italia (es. floricoltura). L'idoneità all'habitat del coleottero giapponese è poco conosciuta e non è mai stata indagata nelle condizioni ambientali italiane.

Nel **terzo capitolo** vengono presentati alcuni modelli che consentono l'analisi dell'impatto dei determinanti ambientali e delle strategie di gestione sui pattern di presenza e co-occorrenza delle specie a livello di paesaggio agricolo. I casi studio si riferiscono all'analisi della biodiversità degli artropodi nell'agroecosistema del vigneto. Nel primo caso presentato, abbiamo indagato come le condizioni ambientali ed i sistemi di allevamento (biologico rispetto all'agricoltura convenzionale) influenzano la composizione della comunità degli artropodi del suolo in diversi contesti pedoclimatici italiani, comprese le principali aree vitivinicole del nostro Paese. Il secondo caso studio è particolarmente rilevante dato il numero di campioni raccolti in un'area viticola abbastanza limitata. La combinazione delle metodologie quantitative utilizzate ha permesso di caratterizzare le biocenosi edafiche in termini di pattern di presenza di co-occorrenza e di valutare le preferenze di nicchia dei microartropodi in termini di variabili

ambientali. La particolarità dell'approccio utilizzato è stata quella di applicare tecniche note in letteratura per la riduzione della dimensionalità dei dati e la classificazione delle osservazioni in iperspazi al fine di caratterizzare la relazione tra determinanti ambientali e pattern di co-occorrenza di specie. Inoltre, queste metodologie sono state integrate con tecniche di data mining al fine di aumentare l'interpretabilità dei risultati.

Il **quarto** capitolo esplora gli strumenti quantitativi per supportare l'approccio dell'analisi di rete alla valutazione della biodiversità. Il capitolo inizia con un breve excursus sulle condizioni che hanno reso possibile lo sviluppo e lo sviluppo di strumenti quantitativi a supporto dell'analisi di rete. Vengono quindi presentate le caratteristiche fondamentali comuni a tutti i modelli quantitativi per l'approccio di rete. Nella seconda parte del capitolo vengono descritte le due categorie di modellazione fondamentali per la valutazione della biodiversità: machine learning e modelli per variabili latenti. Per entrambe le categorie vengono presentati i modelli più diffusi che offrono le migliori prospettive di indagine sui sistemi ecologici. Per ogni modello vengono descritte le caratteristiche essenziali, i vantaggi e le possibili limitazioni al loro utilizzo e le principali linee di ricerca in cui vengono attualmente utilizzati.

Key words: Biodiversità, tratti funzionali, multidimensionalità, artropodi, servizi ecosistemici.

# 1. Introduction

> Biodiversity is '*the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems*' (United Nations, 1992).

In 2005, the United Nations promulgated the results of the Millennium Ecosystem Assessment (MA) project. The MA had the objective of analysing, with robust multidisciplinary scientific bases, the evolution of the planet's ecosystems mainly due to human activities, the related impacts on well-being conditions and identifying intervention strategies for sustainable development. Ecosystem services, according to the definition given by the MA project (2005), are the multiple benefits provided by ecosystems to mankind. The explicit reference to utility for the human community is an important defining aspect that allows to distinguish between 'environmental function', generically referred to a function provided by the environment (OECD, 2005), and ecosystem service (ES), that need to have a close relationship with human well-being.(Boyd & Banzhaf, 2007).

ESs are attributes at the ecosystem or community level; however the functions from which the services are generated are often linked to habitat types, food webs, guilds (functional groups), species, or individuals that collectively produce them (Kremen, 2005). Although the ecosystem is the level of biological organization generally considered optimal for the study of environmental problems, including those of habitat alterations and the decrease in biodiversity, biomes or ecoregions have also been used for this purpose (Odum & Barrett, 2005).

According to the interpretation proposed by the MA, biodiversity is clearly a fundamental constituent element of life on Earth and ecosystems. Therefore, biodiversity, as a fundamental actor in the provision of ES, is the key for ES analysis, understanding and, subsequently, management.

Since food production is an ES, the protection and enhancement of biodiversity represent one of the pillars of a scientific approach to sustainability in agriculture. Knowledge of the structure of biodiversity, its functional traits and the interaction between functional aspects and ecosystem processes must become the foundation for the definition and implementation of sustainable agronomic strategies. These strategies operate to guarantee the regulation and regeneration of ESs while preserving natural capital. Therefore, to conceive and implement a transformation of agricultural systems in the direction of sustainability, alongside the strategies that minimize the impacts on environmental matrices (water, air and soil) and the demand for resources, it is necessary to develop scientific paradigms and agronomic and zootechnical practices that minimize impacts on biodiversity.

## 1.1. International regulations, directives and guidelines on biodiversity management and protection

Starting from the reflections proposed by MA project, the key role played by biodiversity in the sustainable development strategies has been increasingly recognized by the actions and directives promulgated at the international level.

In 2011, FAO proposed the innovative paradigm of sustainable intensification. This paradigm, which originates from conservative agriculture, aims to increase agricultural production to meet the increase in nutritional needs linked to the increase in the world population, while keeping the impacts below the natural recovery threshold. Over the last decade this paradigm has been refined and, at the beginning of 2020, FAO issued the Strategy on Mainstreaming Biodiversity across Agricultural Sectors' which aims to promote sustainable agricultural practices and protect, enhance, preserve and restore biodiversity.

Another fundamental step was made by the assembly of the United Nations which, in 2015, promulgated the 17 Sustainable Development Goals (SDGs). Among these, Goal 2 – Zero hunger defines the guidelines for solving, on a global level, the problem of food safety and the sustainability of primary production, Goal 14 – Life below water aimed to promote the protection of marine and coastal ecosystems and Goal 15 – Life on land specifically aimed at safeguarding life on Earth which outlines a clear path to halt soil degradation and biodiversity loss by 2030.

Also at the European level there has been a path towards the construction of an increasingly clear legislative framework for protecting biodiversity, starting from the Water Framework Directive (Directive 2000/60/EC), the Birds and Habitats Directives (European Commission, 2014); the Environmental Liability Directive (Directive 2004/35/CE), and the Framework for achieving sustainable use of pesticides (Directive 2009/128/EC).

In 2020, the European Union promulgated the "Biodiversity Strategy for 2030" to strengthen the protection of ecosystems and to restore them to achieve “good environmental status”. It stresses the need for an ecosystem-based approach to the management of human activities. Biodiversity is recognised as crucial for safeguarding global food security, underpinning healthy and nutritious diets, and improving agricultural productivity. Furthermore, with the "From farm to fork strategy" (Winkler et al., 2016), European union stated the key role of biodiversity to support and promote healthier and more sustainable food systems guaranteeing fair economic compensations along the entire food chain.

## 1.2. From structural to functional biodiversity

The growing attention placed on the role of biodiversity in the genesis and regulation of ESs has led the scientific world to focus its efforts on the analysis of biodiversity. Research based on traditional and on innovative molecular methods is generating a considerable amount of data on the richness and diversity of species in natural environments as well as in agroecosystems. In the face of this considerable accumulation of data, some important methodological questions on how to approach biodiversity assessment are still open. In particular, it is emerging the need to define scientific-based methodological framework for interpreting, at various levels of complexity, the responses of individual taxa and biotic communities to abiotic environmental variables, to the management practices of ecosystems, broadly defined, and to the type, intensity and frequency of perturbations.

An even more complex scientific challenge is to relate the structural aspects of biodiversity (i.e., aspects referred to biocoenoses composition) to the functional aspects of biodiversity. Indeed, this issue involves understanding the role that the different taxa and the interactions among them have in the genesis and regulation of ecosystem processes and in the provisioning of the services by ecosystems.

The investigation of the link between the structure of ecological communities (i.e., species present in the ecosystem and their abundance) and the ecological processes they promote and regulates requires a systemic approach and adequate modelling tools that can support the level of complexity of the analysis. From a systemic perspective, biotic communities are considered as networks whose nodes (components of biodiversity, such as specific taxon or functional groups) are characterized by attributes (such as, presence/absence, abundance, degree of activation of specific functional traits) and by specific interaction patterns with other nodes.

The basic idea of systemic approaches is that ecosystem processes, as emergent properties of the interacting components in an ecological community, are generated and regulated by the functioning of these networks. The study of the responses of these networks to different management schemes or perturbations should consider the adaptive properties of the communities and the way in which they manifest resistance and resilience responses to perturbations.

In relation to the study objectives, a systemic approach to biodiversity must adequately consider three fundamental methodological options, briefly introduced and discussed in the following sections.

### 1.2.1. Choice of the proper spatial-temporal scale for the analysis

Decisions on the of spatial and temporal scales (both in terms of extension and resolution) are preliminary to any further methodological consideration for systemic analysis. The extension of the study should be chosen taking into account the most appropriate area and time period for the investigated processes and the biological mechanisms involved (Altermatt et al., 2020).

The spatial scale for biodiversity assessment in agroecosystems has a large range of variation. Dimensions of $10^{-1}$ meters or lower is the appropriate scale to investigate processes with high spatial resolution, such as analyses on microbial biodiversity in the rhizosphere. A larger extension, to $10^{1}-10^{2}$ meters, is needed for investigations involving single environmental units or sets of neighbouring environmental units involved, for example, in the movements of arthropods of interest to understand the interactions between host plants, parasites and natural control agents (Dal Cortivo et al., 2020; Hanaka et al., 2019; Machado et al., 2019). Greater extensions, in the order of $10^2 - 10^3$meters, are considered in relation to the need to understand the effects on biodiversity of habitat composition and agronomic practices at the farmscape level (Bàrberi et al., 2010; Smukler et al., 2010; Wezel et al., 2014). Extensions of the order of $10^3 - 10^4$ meters are typically used for the analysis at the landscape level, the hypotheses is that landscape management processes influence the dynamics of meta-population and meta-community and therefore, in a multifunctional perspective, the ecosystem services of large areas (Bukvareva, 2018; Martin et al., 2019; Reddy et al., 2020).

### 1.2.2. Bottom-up and top-down approaches to traits-ecosystem services analysis

Several authors underlined the need for a trait-based approach for a proper understanding of the functional dynamics related to the genesis and regulation of ecosystem services (Aguirre-Gutiérrez et al., 2016; Faucon et al., 2017; Goodness et al., 2016; Wong et al., 2019). The justification and strength of this approach relies on the fact that it constitutes a fully mechanistic approach to functional ecology and it is in agreement with the cascade model explaining the genesis of ecosystem services proposed by Haines-Young and Potschin (2010).

The complexity that a fully mechanistic approach (hence bottom-up approach) generates is evident in all attempts to interpret the interactions between ecological traits and services in terms of clusters traits (D'Andrea et al., 2019; Hanisch et al., 2020; Hevia et al., 2017; Pomerleau et al., 2015). Each cluster is undoubtedly only a simplifying abstraction of the networks of causal relationships that connect, according to a bottom-up logic, properties of ecological units (individual, population or species) and services (Water Research Commission et al., 2014). Possible limitation of trait-based approach in ecology are not restricted to the simplification introduces by the trait-services clusters approach. Another and probably even more important drawback of bottom-up approach is related to the considerable complexity of the analysis of traits even at the level of the single species or population (since the intra-specific variability in traits expression should also be considered) (Degen et al., 2018; Hamilton et al., 2020; Leroux & Loreau, 2015; Wong et al., 2019).

The complexity of bottom-up or reductionist approaches is one of the main reason justifying that most of the studies on the relationship between biodiversity and ESs is not strictly causal but more properly referable as correlative (Degen et al., 2018; Krause et al., 2014; Schmitz et al., 2015). These correlative approaches (hence top-down approaches) try to interpret ecosystem functionality in terms of relationships that ecosystem functionality has with the ecosystem structure. According to this perspective, the structure of ecosystems (usually it is represented by components of biodiversity) is considered as a proxy of functional properties of biodiversity itself and it is correlated with functional properties manifested by ecosystems. Structural properties and attributes of biodiversity are often investigated to explain the quality and efficiency of provisioning of specific ES. For example, aspects of the landscape or farmscape structure and organization are used as proxies to explain landscape or farmscape functional properties such as biological control or pollination. The simplification introduced is quite evident, however the top-down approaches still have the advantages of offering a preliminary picture of how structural aspects affect the functional ones and, even if with some limitations, it can provide useful indications on how to design and manage the agricultural landscape and farmscape (Hendrickx et al., 2007; Nkurunziza et al., 2020; Therond et al., 2017)[REF].

### 1.2.3. Types of modelling approaches

To improve feasibility and quality of systemic analysis in biodiversity assessment, two main requirements should be considered. First, the complexity of the biocenoses, in terms of species and the relationships that link them in functional networks, requires an appropriate level of simplification. Second, adequate quantitative tools should be chosen to study networks of considerable size (number of nodes) and complexity in the pattern of connections.

The complexity of functional networks poses serious limitations to the traditional approaches in mathematical ecology focussed on the mathematical representation of the growth dynamics of populations and the interactions between them. The traditional approaches to mathematical ecology constraint the analyses to networks with a limited number of nodes and relatively simplified connection patterns (Ulanowicz, 2004). Hence the need to explore new analytical tools to make the analysis of networks of interacting species more effective from both the scientific and the applicative perspective.

Network analysis models have been proposed as the most powerful modelling approach to investigate species - trait - function relationships underpinning the provision of ESs (Farage et al., 2020). Thanks to the network approach, it is possible to study the large-scale statistical properties of the whole network, the properties of individual species and interactions, and how they jointly contribute to the structure, function, and stability of ecosystems.

A key characteristic of network approach is that species interactions are determined by traits, i.e. species can interact each other only if they have matching traits (e.g. the right body size for the right size of mouth, live at the same depth in the water column). The degree of trait matching is also a component determining the strength of the interactions between species. Trait matching is easy to detect and to model if a single trait determines the interaction between two species (a classical example reported is the passiflora species and their very long flower tubes that exclude most flower visitors and may have coevolved with bill length in the sword-billed hummingbird *Ensifera ensifera*). Typically, however, the situation is more complex because multiple traits and combinations determine interactions. When interactions are defined by multiple traits, these traits define a ‘niche-space’ (Dormann et al., 2017) and require additional attention in terms of network and model definition.

# 2. Habitat suitability

Habitat Suitability (HS) models investigate the relationship between the presence of a species, both in qualitative (presence/absence) and quantitative (population abundance or density) terms, and the characteristics of the habitat in which it lives (Guisan et al., 2017; Guisan & Zimmermann, 2000; Morrison et al., 2006). HS indicator (or index of suitability) expresses the potential presence and abundance of the species in terms of the suitability of the environment, identifying suitable habitat at different spatial scales (Edenius & Mikusiński, 2006).

Suitability distributions represent a significant step towards the use of landscape analysis in biodiversity management: the successful management of species highly depends on the knowledge of where they occur and why they are there . Habitat suitability models allow to know which factors can be manipulated to favour or hinder the presence of the species in a territory to enhance the genesis and regulation of ecosystem services (Elith & Leathwick, 2009; Jarvie & Svenning, 2018). HS models do not allow to determine where, in a given area, a species is actually present or, in the case of new entry of alien species, where and when it will spread, but where the species can find favourable environmental conditions to establish (presence/absence model) or the potential level of the species population (abundance models) in a given area.

The results of a HS model can be summarized in a map of habitat suitability. These maps show which level each area meet the ecological needs of the species according to the explanatory variables considered. Depending on the HS indicator, the maps can indicate the probability of the presence or the potential density of the species (Antúnez et al., 2018; Bede-Fazekas et al., 2014).

The three component of a HS model are i) variables describing species presence/abundance, ii) explanatory variables (e.g. weather, land use, land cover, pedological characteristics of the soil, altitude), and iii) the classification function.

The explanatory parameters that can be included in the model are many and obviously vary according to the species. They concern all those morphological, vegetational, climatic/meteorological, trophic, anthropic factors determining, or influencing, the availability of food and shelters, the possibility of reproducing, the interaction with other species, or the human disturbance.

The classification function expresses the relationship that links the indices of suitability to explanatory variables. It can assume different degrees of complexity: from a simple logical relation to a complex function. With the aim of generating knowledge that can be used in network analysis, we focus on quantitative approaches to HS and we consider only univariate o multivariate statistical/mathematical models (Guisan et al., 2017; Hatziiordanou et al., 2019; Miller, 2010).

## 2.1. Quantitative approach to habitat suitability models

Mapping species distribution has already been addressed in the past. In fact, since the 19th century or earlier, numerous field surveys have been developed to investigate the distribution of both animal and plant species. However, the worldwide distributions of many species are still unknown. The causes are manifold. First of all, in most field inventories ecological theory is rarely explicitly considered (M. Austin, 2007). Surveys that do not provide an understanding of the mechanisms underlying the distributions, but a simple description of results of these mechanisms (i.e. the species presence or abundance) may become non-informative outside the geographical area and the season where they were performed or if change in environmental factors occur (e.g. changes in land use). Furthermore, often studies failed to test for unimodal or skewed responses species responses to environmental variables (M. P. Austin, 2013; M. P. Austin & Smith, 1989) and a discussion of relationships is lacking (M. Austin, 2007).

In the last two decades, the recent advances in quantitative modelling have provided tools capable of covering some gaps highlighted by previous studies, in particular for comprehensive mapping of species habitat (Franklin, 2010; Guisan et al., 2017; Ovaskainen & Abrego, 2020; Pecchi et al., 2019). These models go beyond the classic approach of describing the suitability of a habitat through the definition of homogeneous and standardized areas compliant with the sites under investigation in the field. The objective of the quantitative approaches is the understanding of the factors that determine the suitability level of the habitat. This knowledge therefore makes it possible to infer the acquired information to different contexts, e.g., habitats not directly observed, or to develop scenario analyses testing the effects of one or more environmental perturbations on the suitability of a habitat.

The drawback of quantitative approaches can be represented by data collection system, that is generally rather long and expensive to guarantee the flexibility of the model and the high informative potential of the results (in terms of inference). Furthermore, these data require a complex and demanding process of analysis for the development and interpretation of the model.

### 2.1.1. Data required for the development of an HSI

Two types of data are necessary to develop HS models:

- data referred to the presence and/or the abundance of the species in a given area;
- data describing the environmental variables in that area.

First of all, it is essential that all data be spatially explicit or referred to well-defined geographical entities (e.g. areas or points).

Species presence and/or abundance can be obtained through direct observations of the animals or estimated through indirect indices of the presence of the species (e.g. traces, excrements, nests) (Cavender-Bares et al., 2020). The choice to collect information on presence or abundance clearly influences the type of result that can be obtained. The abundance allows to obtain decidedly more informative results, but it requires a greater effort in the collection of the data. Absence/presence data and presence-only data should be differently treated. In the first case, where the species is absent it is plausible to think that the environmental considerations are not favourable. When only presence data are collected, it is not possible to infer that outside the points where the species was identified, the area has unfavourable conditions for the species (Dettmers & Bart, 1999).

Data on environmental variables can be collected simultaneously with data on species presence, extrapolated from available knowledge (e.g. thematic maps, satellite images) or estimated by modelling tools (e.g. soil temperature from air temperature data). Data extrapolated from maps or open data repository are generally cheaper than those collected in field and they allow to calculate environmental variables otherwise difficult to measure (e.g. ecosystem diversity indices, distances from edges, land use distribution) in very large areas. The choice of environmental variables to be considered strongly depend on the species under study. Selection of explanatory variables is a trade-off between the need to consider all the aspects that are potentially of interest and the cost (in terms of time and economic effort) of collecting this information. Furthermore, constraints of the effective availability of information (e.g. types, temporal and spatial resolution and extension of open data or thematic maps) must be taken into consideration. Finally, in terms of network analysis, it is not always better to collect many variables, which can create what is called ‘background noise’. That is, they can cause confusion regarding the identification of a relationship between the dynamics of the presence of the species and some environmental variables.

Reliability of the results greatly depends on the degree of precision of these data. The main sources of errors are may relate to the spatial (georeferencing) and temporal information (the date of detection), the estimate of the number of individuals and also the characteristics of each individual animal (sex, age, activity that the animal is carrying out) (Cavender-Bares et al., 2020).

### 2.1.2. Estimation of the classification function

We can distinguish the quantitative approaches to habitat suitability models in two macro categories: explanatory models and predictive models.

In the category of predictive models we include all the techniques which, while having the aim of maximizing the ability to correctly classify sites in terms of species presence/abundance, do not disregard the need to understand and make evident the causal or association mechanisms that can explain the habitat suitability. In these models, therefore, the process generating the suitability index is made explicit, so that it is possible to interpret the coefficients of the model in terms of importance and weight of the factors considered on the estimate of the probability of presence or the level of abundance of a species.

In the HS literature, the most used explanatory models are:

- Analysis of univariate variance (ANOVA) or multivariate (MANOVA): it is generally used to select, among the many environmental variables available, those that explain the variability of the indicator is better and, therefore, will be subjected to subsequent more in-depth analyzes (Brennan et al. 1994; Dueser & Shugart 1978; Meriggi et al. 1992a; Meriggi et al. 1992b; Milne et al. 1989; Smith 1977 );

- Multiple regression: used when a continuous variable is considered as an indicator, generally density (Lindenmayer et al. 1997; Meriggi et al. 1992a);
- Logistic regression: used when we want to consider as an indicator not the density of the species but its presence / absence (Buckland et al. 1996; Lindenmayer et al. 1997). Logistic regression can also be used to discriminate, among those available, the variables that best explain the presence of the species in a territory (Smith & Connors 1994). This method was not widely used until a few years ago for computational problems, now it is more and more widespread;
- Poisson regression: used when the independent variable is an integer (for example: number of individuals) (Lindenmayer et al. 1997);
- Generalized Linear Models: it is a generalization of linear regression that allows to consider also non-linear effects and continuous or categorical dependent variables with normal distribution or belonging to the exponential family (gamma, Possion, binomial, etc.) (Guisan et al., 2002, 2017)(Lindenmayer et al. 1997) ;
- Discriminant analysis: used when the presence / absence of the species or density classes is to be considered as an indicator, or in the case of a categorical dependent variable and continuous independent variables (Meriggi et al. 1992a);
- Bayesian method: based on the Bayes formula of conditional probability. A characteristic of this method is that the conditional probabilities, as well as being experimentally determined, can be estimated a priori by an expert and then updated on the basis of the new acquired data. This method can be used to analyse both continuous and categorical independent variables (Aspinall & Veitch 1993; Holl 1982; Milne et al. 1989; Morrison et al. 1992);
- Principal component analysis (PCA): used to synthesize the most significant environmental variables for the characterization of the habitat then relying on other analyses (typically the discriminant analysis) for the study of the species-habitat relationship. The application of PCA in studying the relationship between a species and the habitat in which it lives is particularly interesting when highly correlated environmental variables are present (Buckland et al. 1996; Collins 1983);
- Canonical correspondence analysis: able to analyse the relationship between several environmental variables and several species (Hill 1991). Each of these statistical techniques has its own strengths and weaknesses and hardly alone provides good predictive models. However, the cascade use of some of these analyses can provide satisfactory results. In literature there are many different combinations of the methodologies described above used in studies of the relationship between a species and its habitat;

- Structural equation models: these models are widely described in chapter 4 (section 4.2.1).

In the category of predictive models, we include models in which the main goal is to maximize the accuracy and precision of the estimate obtained, without explicating the estimation process. The accuracy of a model indicates the ability to obtain unbiased estimates, i.e., as close as possible to the real value of the parameter. Precision measures the degree of agreement between measurements (refers to the variability of estimates). The "black box" models belong to predictive models. In a black-box model, the process of estimating the probability of presence or the level of abundance is hidden and unintelligible. The models developed according to ensemble learning approach fall into this category. A detailed explanation of these models is presented in chapter 4 (section 4.3).

Many studies compared HS approaches and model accuracies, without reaching a conclusive result. The choice heavily depends on the objective of the study and the types of data available. Furthermore, there is a lack of knowledge with respect to transferability of HS model results across disjunctive geographic (Duque-Lazo et al., 2016; Heikkinen et al., 2012; Latif et al., 2016; Manzoor et al., 2018; Wenger & Olden, 2012).

### 2.1.3. Utility and limitations of MVA

Habitat suitability maps can be effectively used in the management of agroecosystems, of natural environmental as well as in urban and regional planning to identify areas of environmental importance that need to be protected, requalified and adequately managed. HS models also find application in the preliminary study for the reintroduction of locally extinct species, in the environmental impact assessment, in the assessment of biodiversity loss or to assess impact of climate change on ecosystem services provisioning. (Brambilla et al., 2010).

HS models should fail to differentiate between suitable and unsuitable areas for the species when i) the explanatory variables selected are insufficient to describe species preferences, ii) explanatory variables are detected at a spatial-temporal resolution inadequate to describe species preferences, ii) habitat preferences differ across population stages and the survey is not sufficiently diversified to take them into account (Johnson, 2005; Pecchi et al., 2019; Wiens & Rotenberry, 1981).

## 2.2. Case study: Habitat suitability of *Popillia japonica*

The Japanese beetle (*Popillia japonica* Newman) is a highly polyphagous invasive species originated from north-eastern Asia. Since its first detection in the US in 1916, the species was able to invade vast areas of North America and some areas in southern Canada. In the 1970s, the Japanese beetle (JB) was introduced in the Azores and since 2014, the species has established in mainland Europe (Italy). Since 2017, few occurrences of the species were reported in Switzerland. In Italy *P. japonica* is currently distributed along the Ticino Valley on the border between Lombardy and Piedmont Regions and, since its first detection, the infested area has increased over time. The species is considered a quarantine insect pest and thus the Italian National and the two Regional Phytosanitary Services are implementing a set of measures aimed at monitoring and implementing the phytosanitary measures for containing the spread of the Japanese beetle.

In the literature, knowledge on habitat suitability for *P. japonica* is limited, although the first studies date back to the middle of the last century (Hawley & Dobbins, 1945). Influence of temperature on the development of immature stages and the impact of soil moisture and texture on survival and development of eggs and first instar have been studied by Régnière et al. (1981a, 1981b) through laboratory tests. Dalthorp et al. (2000) describe the grub patch dynamics of the JB on a central New York golf course describing size, shape, and diffuseness of patches according to a spatial covariance approach. Szendrei and Isaacs investigated the influence of ground covers on abundance and behaviour of the JB (2005, 2006). Zhu et al. (2017) considered 1048 sampling points (45 in Japan and 1003 in North America), Kistner-Thomas et al. (2019) adopted a CLIMEX approach (Sutherst & Maywald, 1985) to estimate HS based on the analysis of 1028 occurrence sites (40 in Asia, 6 in Europe, 981 in North America and 1 in Central America).

In 2018, the Lombardy Region has funded the GESPO Project[1], which aims at developing rational (cost-efficient) and sustainable (low impacts) solutions for the integrated management of *P. japonica*. One of the objectives of GESPO project is the development of quantitative tools aimed at supporting the definition and the implementation of management strategies against the species.

The habitat suitability model investigates the presence of the larval stage of *P. japonica* taking into account land-use variables, soil characteristics (e.g., soil organic carbon, soil texture etc.) and weather variables (i.e. air and soil temperature). The model parameters have been estimated considering the rich and accurate dataset on *P. japonica* distribution and abundance collected by the Lombardy Regional Phytosanitary Service from 2017 to 2018. Data on larvae are collected by soil coring.

### 2.2.1. Data collection

Following the first reports the *Japanese beetle* (JB) in Lombardy (Italy), from 2015 the Lombard regional phytosanitary service has activated a monitoring and control plan, envisaging actions both with respect to adult and larval stages.

In this study we use data referred to three seasons of the larvae monitoring plan: 2016/17, 2017/18, 2018/19. For the first two seasons (2016/17 and 2017/18) samplings were planned in spring, investigating the presence of larvae at the end of the larval stage (hereinafter they will be defined as late larvae). The third season, 2018/19, sampling was conducted in autumn, investigating the presence of the larvae at the beginning of the larval stage, hereinafter called early larvae.

---

[1] The Project GESPO is funded by Direzione Generale Agricoltura - Regione Lombardia (D.d.s. 28 marzo 2018 - n. 4403D.g.r. n. X/7353 14 novembre 2017, project number E86C18002720002).

Only for the purposes of this study, Bellinzago Novarese was hypothesized to be a possible entry point for the pest in Italy as it is the centroid of the first municipalities where the presence of the *P. japonica* was initially detected. The observation area is eastern Lombardy a circular area with a radius of 32 kilometres and centre in Bellinzago Novarese (Novara, Italy).

The monitoring plan for *JB* larvae provides for coring in stable meadows and arable crops. Coring points were randomly sampled in stable meadows and arable land, according to a regular hexagonal grid of Lombardy. The coring was performed with a special core drill (15 cm diameter x 20 cm depth). The number of larvae present in the sample was counted at each core coring point. For the purposes of the HS model, for each sampling point only the presence/absence of *JB* larvae is considered.

### 2.2.2. Explanatory variables

Each coring point was characterized by a set of variables related to the biological cycle (late larvae or early larvae), land use, soil geopedological characteristics, meteorology, and infestation dynamics (Gengping Zhu et al., 2017).

Land use was defined through the DUSAF 6 (Use of Agricultural and Forest Soils) classification system (Regione Lombardia, 2018).

The geopedological characteristics of the soil have been extracted from the map of soils Lombardy 1: 250,000 (Regione Lombardia, 2013). In the suitability habitat model, we included the average organic carbon content on the first meter of soil, classified in low [<1%], medium [1%-2%] and high percentage content [>2 %], the useful depth (cm), the texture of the first meter of soil, the particle size of the first meter of soil, and the average of pH of the first meter of soil.

To evaluate the influence of meteorological factors on the presence of JB larvae, both air and soil related variables were considered. The details on the variables included in the model are reported in Table 1. According to the sampling period (autumn or spring, which corresponds to early or late larvae), the periods in which the meteorological variables are evaluated are different. Information of reference period are included in Table 1.

As proxy of the infestation dynamics, we considered two variables. The distance from the point assumed as the entry point of the species (for the sole purposes of the article, Bellinzago Novarese was hypothesized as entry point) and the number of years from which the area where the sampling takes place was classified as occupied by the Japanese beetle.

*Table 1 – Meteorological variables tested in the Habitat suitability model*

| Acronym | Description | Autumn (early larvae) | Spring (late larvae) | Unit of measure |
|---|---|---|---|---|
| Rain | Cumulative rain | May-October | May-October | mm |
| Relative humidity | Average relative humidity | May-October | May-October | Number (proportion) |
| Average air temperature | Average air temperature at 2 meters | May-October | May-October | Celsius degree |
| Average soil water | Average soil water [depth 0-10 cm] | June-October | June-May | Number (proportion) |
| Length of soil dry stress | Number of hours of soil water < 0.15 | June-October | June-May | Number |
| Average soil temperature | Average soil temperature [depth 0-10 cm] | June-October | September-May | Celsius degree |
| Minimum soil temperature | Average of minimum soil temperature | June-October | September-May | Celsius degree |
| Maximum soil temperature | Average of maximum soil temperature | June-October | September-May | Celsius degree |
| Soil Cold Degree Days | Soil degree days lower than 10°C | June-October | September-May | Celsius degree*day$^{-1}$ |
| Soil Warm Degree Days | Soil degree days between 15°C and 28°C | June-October | September-May | Celsius degree*day$^{-1}$ |
| Length soil cold stress | Number of hours of soil temperature < 10°C | June-October | September-May | Number |
| Length soil hot stress | Number of hours of soil temperature > 28°C | June-October | September-May | Number |

### 2.2.3. Statistical models

The aim of the project is to understand the factors that can influence the presence of larvae od *P. japonica* in the soil. For this reason, we choose a statistical model that would allow the results interpretation from a mechanistic perspective, instead of models, perhaps more performing in terms of estimation accuracy, but which do not allow assessments on the specific effect of the factors considered, 'black boxes' models.

We performed a logistic regression model using a full stepwise approach, i.e. explanatory variables are included or excluded into the model according to their statistical significance measured by Akaike information criteria (Atkinson, 1980; Harrell, 2015). To validate the results of the logistic regression model, we performed a K-fold cross-validation procedure (`cv.glm` function of ‘boot’ R package version 1.3-20). This procedure computes the prediction error for generalized linear models. Dataset is divided randomly into *K* groups; a generalized linear model is fitted on the dataset omitting the *k*-th group (training set) that is used as test set. Prediction are made by fitting the model with the test set. The overall cost function sums up all the prediction errors made on the K test sets.

The results of the logistic regression model will be the starting point to investigate possible nonlinear relationships between explanatory and dependent variables through a generalized additive model (GAM) (Hastie & Tibshirani, 1990; Heikkinen et al., 2007). This nonparametric regression technique allow to fit both linear and complex additive response curves. GAM were developed with gam function of `mgcv` R package, using data on presence-absence of JB larvae as response variable and a logit link function (family binomial). The results of the GAM will be used for the joint exploration of two dimensions of habitat suitability (e.g., pH and soil temperature).

To evaluate the performance of the logistic model, we used accuracy, sensitivity, specificity, and balanced accuracy. These metrics are regularly used to assess habitat modelling performance in ecology (Allouche et al., 2006; Duque-Lazo et al., 2016; Freeman & Moisen, 2008; García et al., 2009; Giannini et al., 2013; Miller, 2010). To evaluate the estimation results, we used a confusion matrix showing true positive (samples observed as positive and correctly classified), false positive (samples observed as negative and wrongly classified), false negative (samples observed as positive and wrongly classified), and true negative (samples observed as negative and correctly classified) (Table 2).

*Table 2 - Confusion matrix for the evaluation of our presence–absence model*

| | | Observed value | |
|---|---|---|---|
| | | Presence | Absence |
| Predicted value | Presence | True positive (a)<br>Correctly predicted | False positive (b)<br>wrongly predicted |
| | Absence | False negative (c)<br>wrongly predicted | True negative (d)<br>Correctly predicted |

Accuracy [eq. Accuracy = $\frac{a+d}{n}$ [1] measures the percentage of correctly predicted observations on the total sample size. Sensitivity Sensitivity = $\frac{a}{a+c}$ [2] represents the percentage of observations correctly classified on the total of observations in which the species was present. Specificity [eq. Specificity = $\frac{d}{b+c}$ [3] represents the percentage of observations correctly classified on the total of observations in which the species was absent. Balanced accuracy [eq. Balanced accuracy = (Sensitivity + Specificity) / 2 [4] normalizes overall accuracy of the model and it is especially useful when the classes are imbalanced.

$$\text{Accuracy} = \frac{a+d}{n} \qquad [1]$$

$$\text{Sensitivity} = \frac{a}{a+c} \qquad [2]$$

$$\text{Specificity} = \frac{d}{b+c} \qquad [3]$$

$$\text{Balanced accuracy} = (\text{Sensitivity} + \text{Specificity}) / 2 \qquad [4]$$

To compute the classification matrix, it is customary to set the classification cut-off equal to 0.5, i.e., every observation with a predicted probability from the model equal to or greater than 0.5 is classified as positive. In our study data matrix is sparse, i.e., the percentage of sites where *P. japonica* larvae were found is much lower than the percentage of sites where *P. japonica* larvae were not found. Considering the unbalanced dataset, we choose the classification cut-off for the confusion matrix based on the ROC (Receiver Operating Characteristic) curve (estimated with `roc` function of ‘`pROC`’ R package) and the best trade-off between specificity and sensitivity (Freeman & Moisen, 2008; Weiss & Provost, 2001). The best cut-off is chosen as trade-off between specificity and sensitivity.

The goodness of fit of the GAM is evaluated with QQ-plot (quantile-quantile plot) of residuals. The Q-Q Plot allows to graphically compare the cumulative distribution of the observed variable with the cumulative distribution of the normal. If the observed variable has a normal distribution, the points of this joint distribution gather on the diagonal from bottom to top and from left to right (Wilk & Gnanadesikan, 1968).

### 2.2.4. Results

In the three coring seasons included in the study, 8904 sites were sampled. Table 3 shows, per sampling season, the total number of sites, the number of positive samples, where at least one larva of the Japanese beetle was found, and the number of negative samples, where JB larvae were absent.

Each sampling site was characterized by the variables presented in Section 2.2.2. Table 5 summarizes the distributions of the explanatory variables in terms of mean and standard deviation (for quantitative variables) or frequency distribution (for categorical variables), according to presence or absence of larvae of *P. japonica*. The correlation matrix (graphically reported in Table 5) shows clusters of variables that present a very high correlation (greater than 85%). For the estimation of the logistic regression model, only one variable was considered for each correlation cluster. For this reason, average of minimum soil temperature, average of maximum soil temperature, length of soil cold stress, and length of soil hot stress were excluded.

From the descriptive analysis of the variables (Figure 1), it emerges that levels of organic carbon and pH are influenced by the texture and particle size (respectively) of the soil. for this reason, the corresponding interaction factors are also included in the regression model.

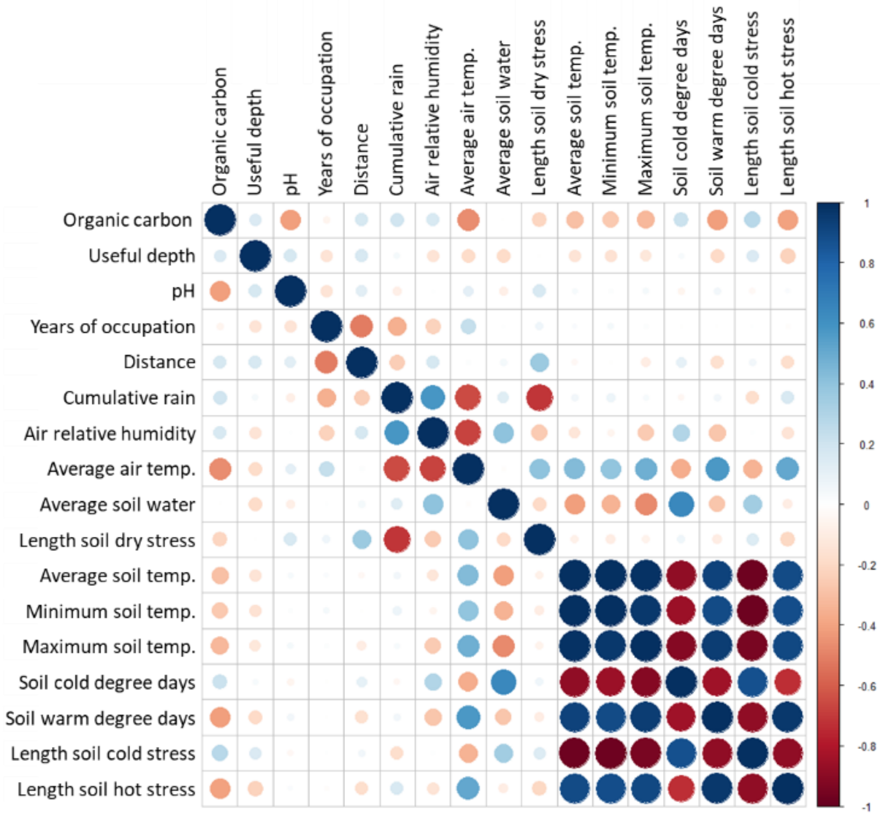
*Table 3 – Number of coring sites included in the study, by season and presence/absence of P. japonica larvae*

| Season | Sampling period | Total sites | Positive sites (presence of JB larvae) | Negative sites (absence of JB larvae) |
|---|---|---|---|---|
| 2016/17 | Spring | 5728 | 485 (8.47 %) | 5243 (91.53 %) |
| 2017/18 | Spring | 2760 | 397 (14.38 %) | 2363 (85.62 %) |
| 2018/19 | Autumn | 420 | 270 (64.29 %) | 150 (35.71 %) |
| Total | | 8908 | 1152 (12.93 %) | 7756 (87.07 %) |

*Table 4 - List of explanatory variables calculated for each sampling site and used as predictors of the presence of P. japonica larvae*

| **Category** | **Explanatory variables** | **[Range] Frequency distribution (%)** | **Mean (± SD)** |
|---|---|---|---|
| Land use | Arable crops stable | 64.13 % | |
| | Meadows | 30.13 % | |
| | Others | 5.74 % | |
| Geopedological characteristics | Organic carbon content | | |
| | Low (<1 %) | 60.00 % | |
| | Medium (1%-2%) | 29.84 % | |
| | High (>2%) | 10.16 % | |
| | Useful depth (cm) | [20 - 201] | 93.65 (±52.74) |
| | Texture | | |
| | Sandy loam | 69.19 % | |
| | Loamy sand | 20.87 % | |
| | Silt loam | 4.65 % | |
| | Loam | 4.43 % | |
| | Sand | 0.86 % | |
| | Particle size | | |
| | Coarse-loamy | 50.65 % | |
| | Loamy-skeletal | 22.59 % | |
| | Sandy | 10.96 % | |
| | Sandy-skeletal | 10.78 % | |
| | Others | 5.02 % | |
| | pH | [4.9 - 7.3] | 6.16 (±0.59) |
| Meteorological factors | Rain | [22.06 – 547.0] | 322.66 (±147.59) |
| | Air Relative humidity | [0.42 – 0.62] | 0.52 (±0.04) |
| | Average air temperature | [20.01 – 24.62] | 22.28 (±0.86) |
| | Average soil water | [0.26 - 0.95] | 0.37 (±0.07) |
| | Length of soil dry stress | [2.0 – 2391.5] | 293.59 (±383.31) |
| | Average soil temperature | [0.79 – 23.41] | 11.9 (±2.55) |
| | Minimum soil temperature | [0.60 – 19.56] | 8.92 (±2.14) |
| | Maximum soil temperature | [1.01 – 28.42] | 15.73 (±3.14) |
| | Soil cold degree days | [58.6 – 2648.9] | 765.96 (±254.30) |
| | Soil warm degree days | [23.9 – 1365.5] | 646.11 (±191.78) |
| | Length of soil cold stress | [151.4 – 6302.0] | 3043.6 (±704.51) |
| | Length of soil hot stress | [5.45 – 1024.8] | 310.67 (±162.74) |
| Invasion dynamics | Distance from first entry | [2402 – 31978] | 17111.4 (±6116.7) |
| | Years of occupation | [0 – 3] | 1.34 (±0.66) |

*Table 5 – Correlation matrix of quantitative explanatory variables*



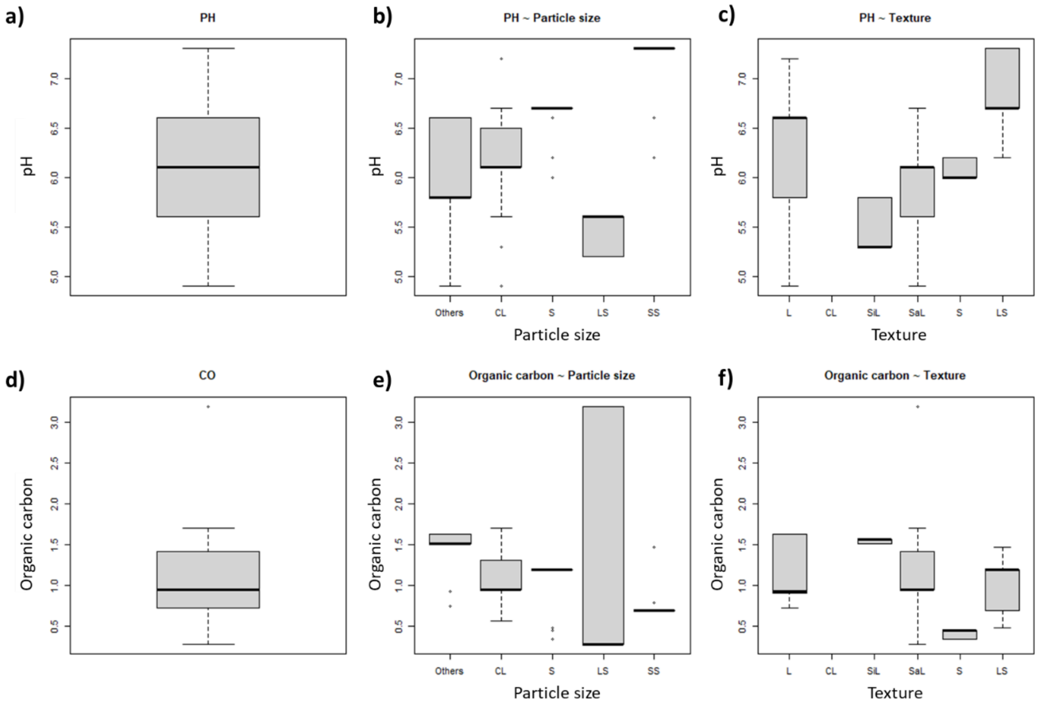


*Figure 1 – Distribution of pH and organic carbon in the whole dataset (a ,d) according particle size classification (b, e) or according texture classification (c, f)*

*Particle size : CL – coarse loamy, S – sandy, LS – loamy skeletal, SS – sandy skeletal;*
*Texture: L – loam, CL – clay loam, SiL – silt loam, SaL – sandy loam, S – sand – LS – loamy sand*

## Logistic regression model for habitat suitability of *P. japonica* larvae

Based on the stepwise procedure, ten explanatory variables and four interaction terms have been selected as significative in the logistic model. Results of the final logistic model are reported in computed based on this cut-off: 72.6 % of the observations are correctly classified (accuracy), 70.1 % of the positive sites are correctly classified (sensitivity), and 89.3 % of the negative sites are correctly classified (specificity). Balanced accuracy is equal to 0.80. According to the evaluation metrics reported, model shows a good fitting to the observed data.

Table 6.

Positive sites (where the larvae of JB were found) are 12.93 % of the total of sites considered in the analysis, therefore the dataset is strongly unbalanced. To select the optimal classification cut-off, we computed the ROC curve (represented in Figure 2). The best trade-off between sensitivity and specificity (highlighted with a red star in Figure 2) is obtained with a cut-off equal to 0.09. Confusion matrix ( Table 7) and the evaluation metrics are computed based on this cut-off: 72.6 % of the observations are correctly classified (accuracy), 70.1 % of the positive sites are correctly classified (sensitivity), and 89.3 % of the negative sites are correctly classified (specificity). Balanced accuracy is equal to 0.80. According to the evaluation metrics reported, model shows a good fitting to the observed data.

| **Explanatory variables** | **Coefficient estimated** | **Standard error** | $Pr(> \lvert z \rvert)$ |
|---|---|---|---|
| Intercept | 6.221 | 2.408 | 0.010 |
| Sampling period = autumn | 2.693 | 0.247 | <0.001 |
| Texture=LS | - 1.937 | 0.279 | <0.001 |
| Soil organic carbon - medium | 0.836 | 0.132 | <0.001 |
| Soil organic carbon - high | - 0.632 | 0.216 | 0.003 |
| pH | - 0.437 | 0.095 | <0.001 |
| pH : Particle size=S | - 0.145 | 0.025 | <0.001 |
| pH : Particle size=CL | - 0.162 | 0.015 | <0.001 |
| Distance | - 0.0001 | 0.00001 | <0.001 |
| Years of infestation | 0.882 | 0.112 | <0.001 |
| Cumulative rain | - 0.003 | 0.001 | <0.001 |
| Average air temperature | - 0.399 | 0.105 | <0.001 |
| Length of soil dry stress | 0.001 | 0.0002 | <0.001 |
| Soil cold degree days | 0.002 | 0.0006 | <0.001 |
| Soil warm degree days | 0.004 | 0.001 | <0.001 |

Table 6 – Logistic regression model for presence/absence of P. japonica larvae in coring

| **Explanatory variables** | **Coefficient estimated** | **Standard error** | $Pr(> \lvert z \rvert)$ |
|---|---|---|---|
| Intercept | 6.221 | 2.408 | 0.010 |
| Sampling period = autumn | 2.693 | 0.247 | <0.001 |
| Texture=LS | - 1.937 | 0.279 | <0.001 |
| Soil organic carbon - medium | 0.836 | 0.132 | <0.001 |

| | | | |
|---|---|---|---|
| Soil organic carbon - high | - 0.632 | 0.216 | 0.003 |
| pH | - 0.437 | 0.095 | <0.001 |
| pH : Particle size=S | - 0.145 | 0.025 | <0.001 |
| pH : Particle size=CL | - 0.162 | 0.015 | <0.001 |
| Distance | - 0.0001 | 0.00001 | <0.001 |
| Years of infestation | 0.882 | 0.112 | <0.001 |
| Cumulative rain | - 0.003 | 0.001 | <0.001 |
| Average air temperature | - 0.399 | 0.105 | <0.001 |
| Length of soil dry stress | 0.001 | 0.0002 | <0.001 |
| Soil cold degree days | 0.002 | 0.0006 | <0.001 |
| Soil warm degree days | 0.004 | 0.001 | <0.001 |

*Table 7 – Confusion matrix of the logistic regression model*

| | | Observed value | | |
|---|---|---|---|---|
| | | Presence | Absence | Total |
| Predicted value | Presence | 1029 | 2317 | 3346 |
| | Absence | 123 | 5439 | 5562 |
| | Total | 1152 | 7756 | 8908 |

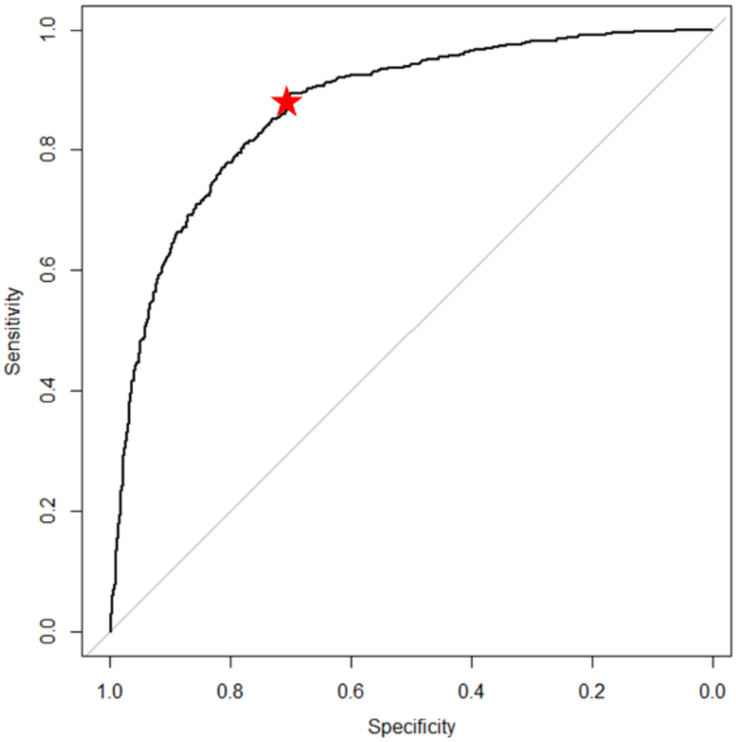


*Figure 2 – ROC curve of the logistic model for habitat suitability of P. japonica larvae. The red star represents the best trade-off between sensitivity and specificity*

Results of logistic regression models should be interpreted in terms of odds [eq. $Odds = \frac{P(Y=1)}{P(Y=0)} = \frac{P(Y=1)}{1-P(Y=1)}$ [5], i.e. the ratio of the probability of success over the probability of failure. In logistic regression model, the linear relationship between the dichotomous outcome variable $Y$ (in this case larval presence) and the k explanatory variables $X_i$ is based on a logit transformation of the outcome variable [eq. $logit\big(P(Y=1)\big) = \log\frac{P(Y=1)}{P(Y=0)} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$ [6]. For this reason, the exponentiated regression coefficient can be interpreted as the increase/decrease of odds as one-unit increase on the corresponding explanatory variable, being fixed the values of the other explanatory variables.

$$Odds = \frac{P(Y=1)}{P(Y=0)} = \frac{P(Y=1)}{1-P(Y=1)} \qquad [5]$$

$$logit\big(P(Y=1)\big) = \log\frac{P(Y=1)}{P(Y=0)} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \qquad [6]$$

### Generalized additive model for habitat suitability of P. japonica larvae

The explanatory variables of the logistic regression model are used as inputs to the GAM model, interaction terms were not included. GAM confirmed the statistical significance for all explanatory variables, although for organic carbon only the high category results as significant, and detected six non-linearity: pH, cumulative rain, average air temperature, length of soil dry stress, soil cold degree days and distance. Detailed results are presented in Table 8. The coefficients of the GAM linear terms can be interpreted as usual in the linear regression model. For the variables that are included in the model through nonlinear link functions (smooth terms), the effective degrees of freedom (EDF) provide an indication of the complexity of the smooth function. If EDF is equal to 1, the link function is linear. When EDF is equal to 2, the link function is a quadratic curve, and so on. The estimated smooth functions are presented in Figure 4. The statistical significance of the exploratory variables, both linear and nonlinear link, is indicated with asterisks in the p-values column.

The qq-plot of the model (Figure 3) shows a good fit of the model to the data.

*Table 8 – GAM results for the nonlinear prediction of habitat suitability for P. japonica larvae*

| **Linear link function** | | | | **Nonlinear link function** | | |
|---|---|---|---|---|---|---|
| Explanatory variables | Coefficient estimates | p-value | | Explanatory variables | EDF | p-value |
| Intercept | - 6.934 | <0.001 | | pH | 6.710 | 0.004 |
| Sampling period = autumn | 1.394 | 0.002 | | Distance | 5.601 | <0.001 |
| Texture=LS | 0.956 | <0.001 | | Cumulative Rain | 8.011 | <0.001 |
| Soil organic carbon - high | 0.815 | 0.077 | | Average air temperature | 6.668 | <0.001 |
| Years of occupation | 0.687 | <0.001 | | Length of soil dry stress | 5.503 | <0.001 |
| Soil warm degree days | 0.004 | <0.001 | | Soil cold degree days | 4.308 | <0.001 |

EDF = Effective degrees of freedom

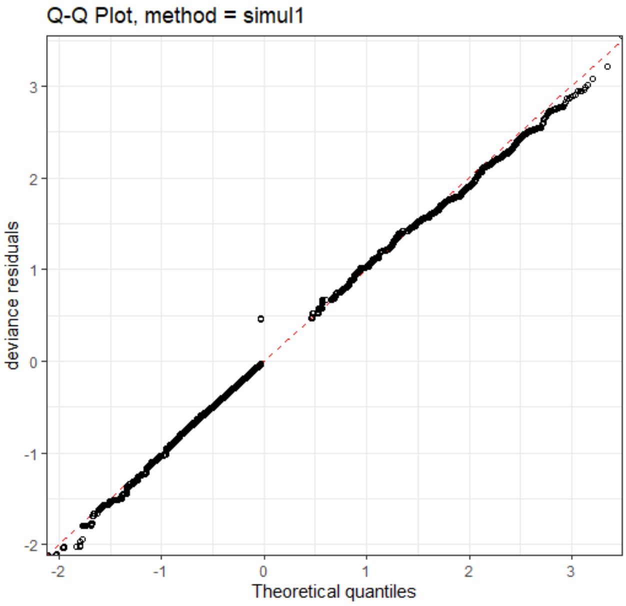


*Figure 3 – QQ-plot of residuals of GAM*

Results of the GAM have been used to estimate the probability of the presence of the P. japonica larvae according the values assumed by couples of explanatory variables, being fixed the values of others explanatory variables. The analysis of all possible two-dimensional niches according to an ecological perspective led to the identification of two interesting niches, defined by 'length of the soil dry stress' in relation with 'soil warm degree days' (Figure 5 a,b) and 'soil cold degree days' (Figure 5 c,d). The probability of presence of JB larvae in the multivariate model according to the joint distribution of the two explanatory variables is represented in the 2-dimension plot by a colour scale, while in the 3-dimesion plot is represented by the blue plane.

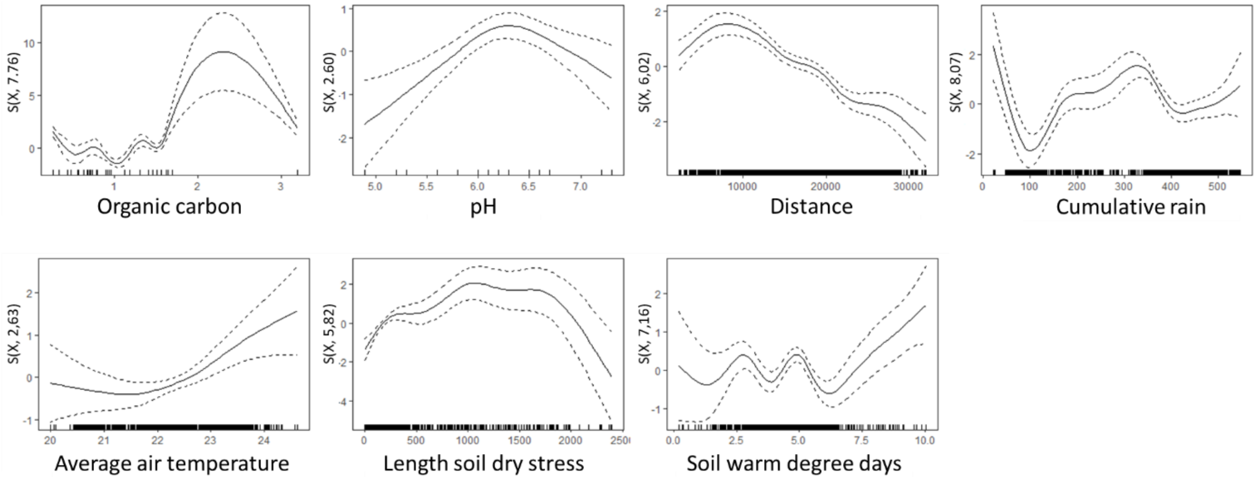


*Figure 4 – Smooth function of the GAM model. Plot shows the individual effects of nonlinear terms, where only the default layers are added.*

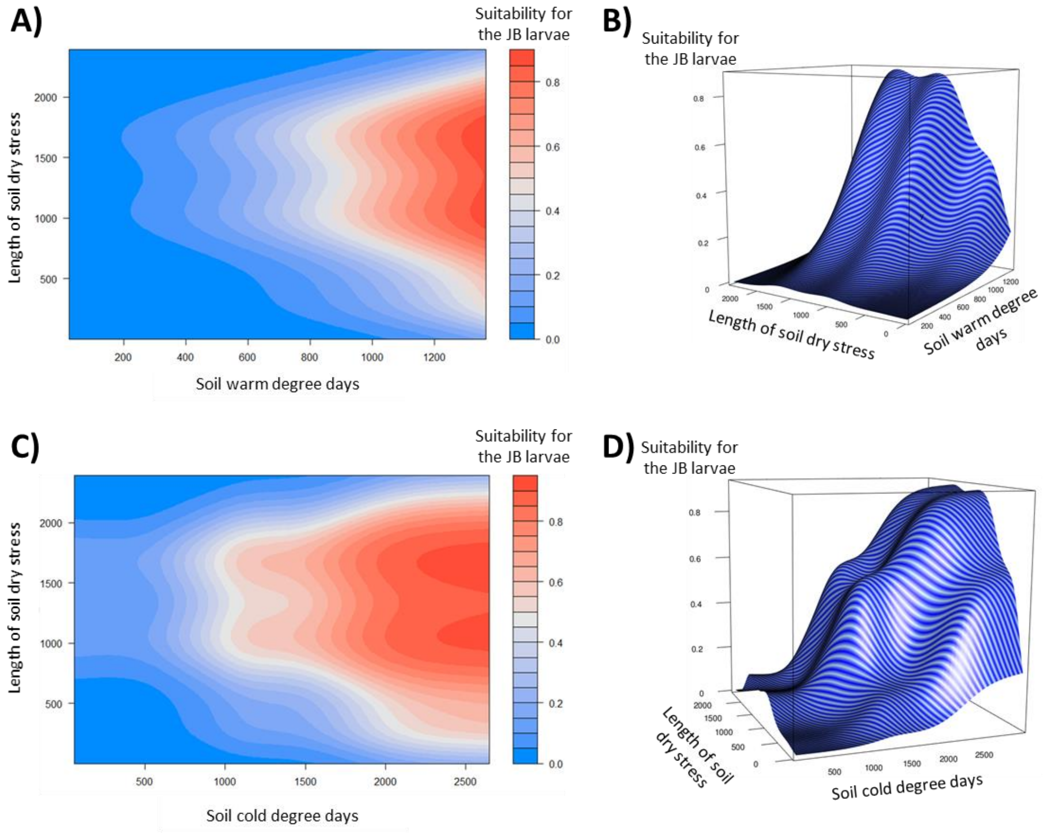


*Figure 5 - Probability of the presence of the Japanese beetle (JB) larvae in the soil, based on the values assumed by 'length of the soil dry stress' and 'soil warm degree days' (A and B) or by 'length of the soil dry stress' and 'soil cold degree days' (C and D). The left plots (A and C) are the 2D representation of the niches, the probability of the occurrence of the species is represented by a colour scale (blue represents low suitability areas and red are the favourable areas). In the 3-dimesion plots (B and D), the blue planes represent the probability of the species presence in the multivariate model according to the joint distribution of the two explanatory variables represented on the x-axis and y-axis., being fixed the values of others explanatory variables.*

### 2.2.5. Discussion

The results of the logistics model confirm the few results available in literature on the distribution of P. japonica and, in particular, they extend knowledge to the larval stage ad to pedo-climatic condition in Italy, neglected in currently published studies (Allsopp et al., 1992; Kistner-Thomas, 2019; Régnière et al., 1981a, 1981b; Shanovich et al., 2019).

The sampling period (spring or autumn) is considered a very important factor for predicting the presence of P. japonica larvae. In particular, the estimated regression coefficient, equal to 2.69, indicates that the odds (i.e. defined as the probability that the event will occur divided by the probability that the event will not occur) of coring carried out in the autumn is over 14 times the odds of coring carried out in the spring. This result, consistent with what observed in Table

3, suggests that there may be a survival issue of whole larval clusters during the winter period. This aspect requires further in-depth studies.

According to logistic regression model, P. japonica shows a preference for less acidic soils. This preference is influenced by the particle size of the soils, in particular for coarse loamy and sandy soils. Focusing on texture, loamy sand soils are unfavourable environment for the JB, in agreement with previous scientific evidences (Allsopp et al., 1992; Hawley & Dobbins, 1945). The probability of finding P. japonica larvae in sites with a loam sandy texture is considerably lower than that calculated in soils with other types of textures, all other explanatory variables being equal.

The coefficients related to soil organic carbon showed that the JB prefers soils with reduced organic carbon content, consistently with Dalthorp et al. (2000). The logistic model shows that the probability of finding P. japonica larvae in soils with an organic carbon content between 1 and 2% is double that of those with a lower organic carbon content, while in soils with an organic carbon content greater than 2% the probability of finding the JB larvae is 0.5 times that of soils with low organic carbon content.

The probability of finding clusters of larvae is closely linked to the distance from the site of first entry and the number of years since the species occupied the sampling area. Both these results agree with expectations based on a diffusion process. In fact, in the sites of more recent occupation, the diffusion process is still in progress, therefore is strongly influenced by the dynamics of population growth.

In agreement with Kistner-Thomas (2019), the larvae of the JB showed high sensitivity to soil water content, evidenced by the negative effect of cumulative rain (that is a proxy of the water input of the soil) and the positive effect of the duration of soil dry stress.

The negative effect of the average air temperature (logistic regression coefficient = - 0.231) could be interpreted in terms of a temperature-mediated effect on the fertility of adults of the Japanese beetle (Régnière et al., 1981a).

The availability of degree days in the comfort range of 15-28°C accumulated in the soil is instead a factor that positively affects the probability of the presence of larvae. This could be explained in terms of the survival of larval clusters of *P. japonica* under more favourable thermal conditions but it could also be the product of the effect of temperature on the whole of the components of the life cycle of the Japanese beetle, in particular the survival and development rate of all stages of the insect (Kistner-Thomas, 2019; Régnière et al., 1981a; Shanovich et al., 2019).

The results of the GAM analysis are used to investigate the ecological niches determined by the interaction of two variables and in terms of the interaction of pairs of variables explaining in terms of suitability for P. japonica larvae. In particular, the nonlinearity included in the GAM allows to identify optimal ranges of explanatory variables. The results of the model combined with an ecological interpretation of bi-dimensional niche spaces, led to the identification of a very interesting relationship. In fact, Figure 5 shows that there is a comfort zone for JB larvae, defined by a duration of the soil water stress period between 800 and 1800 hours and soil cold degree days greater than 1000. In this area the probability of the presence of P. japonica larvae is high, outside the zone identified by these intervals, the probability drastically decreases.

### 2.2.6. Concluding remarks on habitat suitability models

Quantitative models used in this study have provided important information for understanding the habitat suitability of larvae of *P. japonica*. The descriptive analysis of the dataset was essential in defining the subset of variables to be used in the model. The collinearity between explanatory variables has been analysed to guarantee the robustness of the models considered. Furthermore, the presence of possible interactions between the explanatory variables to be included in the classification model were evaluated, for example the relationship between texture, particle size, pH and organic carbon.

The two-phase modelling approach used to investigate the relationship between explanatory variables and presence of larvae of the Japanese beetle revealed to be very useful. The logistic regression model allowed a deep investigation of the effect of each explanatory variable, in terms of direction (favourable or unfavourable) and of the intensity of the relationship with the probability of presence of larvae in the coring site. This knowledge has been integrated with the results of the GAM, allowing the identification two-dimensional suitability niches.

# 3. Soil biodiversity assessment

*Disclaimer: the text of this chapter is a reworking (Section 3.1) or a faithful quotation (Sections 3.2 and 3.3) of two published papers to which I contributed significantly for the conceptualization, the quantitative methodology and the interpretation of the results:*

- *Ghiglieno, I., Simonetto, A., Donna, P., Tonni, M., Valenti, L., Bedussi, F., & Gilioli, G. (2019).* ***Soil Biological Quality Assessment to Improve Decision Support in the Wine Sector****. Agronomy, 9(10), 593. https://doi.org/10.3390/agronomy9100593 (introduction and section 3.2)*
- *Ghiglieno, I., Simonetto, A., Orlando, F., Donna, P., Tonni, M., Valenti, L., & Gilioli, G. (2020).* ***Response of the Arthropod Community to Soil Characteristics and Management in the Franciacorta Viticultural Area (Lombardy, Italy)****. Agronomy, 10(5), 740. https://doi.org/10.3390/agronomy10050740 (introduction and section 3.3)*

Soil is a fundamental natural resource that forms the basis of terrestrial ecosystems and provides essential services to humanity. The provision of nutrients for plant growth, the recycling of these nutrients, the regulation of the climate are among the fundamental services that the soil provides. An healthy soil is necessary to maintain the correct level of provisioning of these ecosystem services and is therefore fundamental to guarantee sustainable production in the agricultural sector (FAO, 2015). The health and fertility of soils are seriously threatened by various stressors of natural and anthropogenic origin, e.g., effects of erosion and loss of organic carbon stored in the soil become increasingly evident. It therefore becomes urgent to develop strategies for sustainable use of agricultural land, in association with the management of issues related to agricultural production that impact on the soil.

## 3.1. Soil biodiversity

Soil has been described as the most complex and diverse ecosystem in the world (Decaëns et al., 2006; Kopittke et al., 2019; Wolters, 2001). It represents an important pool of biodiversity, the European Commission (2010) estimates that about a quarter of living species on our planet live in the soil. The importance of soil biodiversity is mainly linked to the key role that soil biota plays in regulating ecosystem processes (Costantini et al., 2015; Geisen et al., 2019; Karlen et al., 1997; Powlson et al., 2011).

Soil biological diversity is one of the most sensitive component of agroecosystem biodiversity to environmental stress and farming practices (Juan-Ovejero et al., 2019; Marasas et al., 2001; Ruf et al., 2003; Wallwork, 1972). The influence of abiotic and biotic variables and their interactions varies according to the climate, type of soil and agricultural practices (Migliorini et al., 2004; Xu et al., 2012). Despite the increasing number of studies on soil biodiversity, many structural and functional aspects of this biodiversity remain largely unexplored (Büchs et al., 2003; Cameron et al., 2018; van Straalen, 1998).

Soil moisture and soil temperature have emerged as important factors determining arthropod distribution, although the extent of the impact highly vary between taxa (Choi et al., 2002, 2006; Clapperton et al., 2002; Frampton et al., 2001; Ikemoto, 2003, 2005; J.-L. Liu et al., 2017; O'Lear & Blair, 1999; Tsiafouli et al., 2005). Generally, soil moisture has a positive effect on the abundance of soil arthropod communities (Frampton et al., 2001; Rapoport & Tschapek, 1967). The optimal temperature ranges between 5 °C and 10 °C for soil arthropod species active in winter, while it ranges between 10 °C and 18 °C for species active in summer (Eisenbeis & Wichard, 1987).

Soil chemical and physical characteristics have been identified as important drivers in soil arthropod distribution and abundance (André et al., 1994; Andrés et al., 2016; Bagyaraj et al., 2016; Culliney, 2013; Holmstrup et al., 2007; Lavelle et al., 1995; Migliorini et al., 2004; Petersen & Luxton, 1982; Ruf et al., 2003; van Straalen, 1998; van Straalen & Verhoef, 1997; Wu et al., 2014). Soil texture (Andrés et al., 2016), soil organic matter content (Potapov et al., 2017; Shakir & Ahmed, 2015), pH (Lavelle et al., 1995; van Straalen, 1998; van Straalen & Verhoef, 1997) and heavy metal contamination (Holmstrup et al., 2007; Migliorini et al., 2004) have been shown to have a great influence on soil biota. In particular, soil pH and soil organic matter represent the most significant drivers of arthropods role in soil organic matter degradation (Bagyaraj et al., 2016; Culliney, 2013). Soil arthropods contribute, in fact, to nutrient cycling as secondary decomposers, conditioning litter through comminution and passage through the gut, for further breakdown by the microflora and stimulating microbial mineralisation of nutrients through grazing activity (Bagyaraj et al., 2016; Culliney, 2013).

To investigate soil biological quality, Parisi et al. (2001) proposed the index 'Qualità Biologica del Suolo' (Soil Biological Quality) or QBS. We focused on the version of the index based on microarthropods: QBS-ar Index. The QBS-ar methodology is based on the principle that the importance of a soil arthropod taxon as indicator of soil quality is directly proportional to the sensitivity of the taxon to variability and perturbation of soil conditions. QBS-ar applies the criterion of "biological forms" to edaphic microarthropods: different species are grouped according to their morphological characteristics. The Eco-Morphologic Index (EMI) is then associated to each taxon based on the level of adaptation to environmental conditions of the taxon. QBS-ar Index does not provide any information about the abundance of taxa (Yan et al., 2012). Currently, QBS-ar Index is widely used in the agricultural sector (Mazzoncini et al., 2010; C. Menta et al., 2014; Cristina Menta, Conti, Pinto, et al., 2018; Tabaglio et al., 2009) and it has also been applied to different contexts, such as forests, urban and degraded ecosystems (Blasi et al., 2013; Galli et al., 2014; Hartley et al., 2008; Madej et al., 2011; C. Menta et al., 2014; Santorufo et al., 2012).

In the following sections of this chapter, attention will be focused on the vineyard agro-ecosystem. Several studies have been carried out on soil microbial diversity, earthworm, nematode and microarthropod communities as soil quality bioindicators in vineyard (Costantini et al., 2015; Fusaro et al., 2018; Priori et al., 2015; Zarraonaindia et al., 2015). A positive effect of organic management, age of planting and meteorological conditions on soil arthropod abundance and distribution have been demonstrated (Caprio et al., 2015; Gagnarli et al., 2015), however, the results greatly varied according to the taxon investigated (Puig-Montserrat et al., 2017; Seniczak et al., 2018). Furthermore, the effect on soil arthropod biodiversity of length of the period of organic management has been little investigated (Costantini et al., 2015). Further research is therefore needed to assess the medium and long-term effects of organic agriculture on soil biodiversity (Döring et al., 2019).

In Section 3.2, we investigated the relationship between biological soil quality (represented by the QBS-ar index), environmental conditions and farming strategy (organic or conventional) in 70 soil samples surveyed in different Italian wine contexts is reported (Ghiglieno, Simonetto, et al., 2019). The study is based on a Decision Support Systems (DSSs) developed for wine sector biodiversity management assessment (BIOPASS®).

In Section3.3, the study carried out to investigate soil arthropod taxa responses in vineyards in Franciacorta (Lombardy, Italy) is presented (Ghiglieno et al., 2020).

## 3.2. Soil Biological Quality Assessment to Improve Decision Support in the Wine Sector

Decision support systems (DSSs) are useful tools to assess biodiversity in agroecosystems. The importance of DSSs in agriculture is widely recognised (Cancela et al., 2019; Matthews et al., 2008)]. For the wine sector a large number of DSSs have been developed in the last few years, although these are mostly dedicated to pest and disease management (Caffarra et al., 2012; Calonnec et al., 2008; Kuflik et al., 2009; Pérez-Expósito et al., 2017), or to vinification management (Merwe et al., 2011; Rinaldi et al., 2006; The Wine Institute & Kennedy/Jenks Consultants, 2008). Despite their importance, DSSs are poorly considered as easy and affordable tools for assessing biodiversity in the field of wine, meeting the needs of the agricultural sector, as underlined by Doran and Zeiss (Doran & Zeiss, 2000).

In this study we carried out in-depth analysis of edaphic biodiversity to provide useful information for the design and interpretation of soil biodiversity assessment, to improve the DSS. We investigated how environmental conditions (meteorological and pedological) and farming systems (organic compared to conventional farming) influence the composition of the soil arthropod community. With this aim a total of 70 QBS-ar indices were obtained in 2014, 2015 and 2016 from different organic and conventional Italian winemaking contexts. Experimental sites have been chosen among wine farming where the DSS were tested. Chemical soil analyses were carried out in each sampling point, to identify the possible relationship between the pedological and biological characteristics of the soils. Data from different meteorological stations close to the sampling points were collected.

### 3.2.1. The DSS to assess the sustainability of the wine production chain

In this study we referred to the BIOPASS® DSS (*Agronomisata*; Ghiglieno, Donna, et al., 2019), developed to assess the sustainability of the wine production chain. The DSS defines a system in which three main compartments are described: soil, vines and wine, as shown in Figure 6. Specific input and output are identified for each compartment, as well as external drivers influencing the overall system. The input concerned management of the compartment and it was divided into two groups: i) chemical, physical, mechanical, and human related inputs; ii) actions (e.g., agronomical practices). As regards the output, we considered i) production outputs, such as agronomical characteristics, chemical characteristics of grapes and wine, ii) sensory properties, and iii) chemical-biological outputs, if considered to have an impact on the specific environmental context.

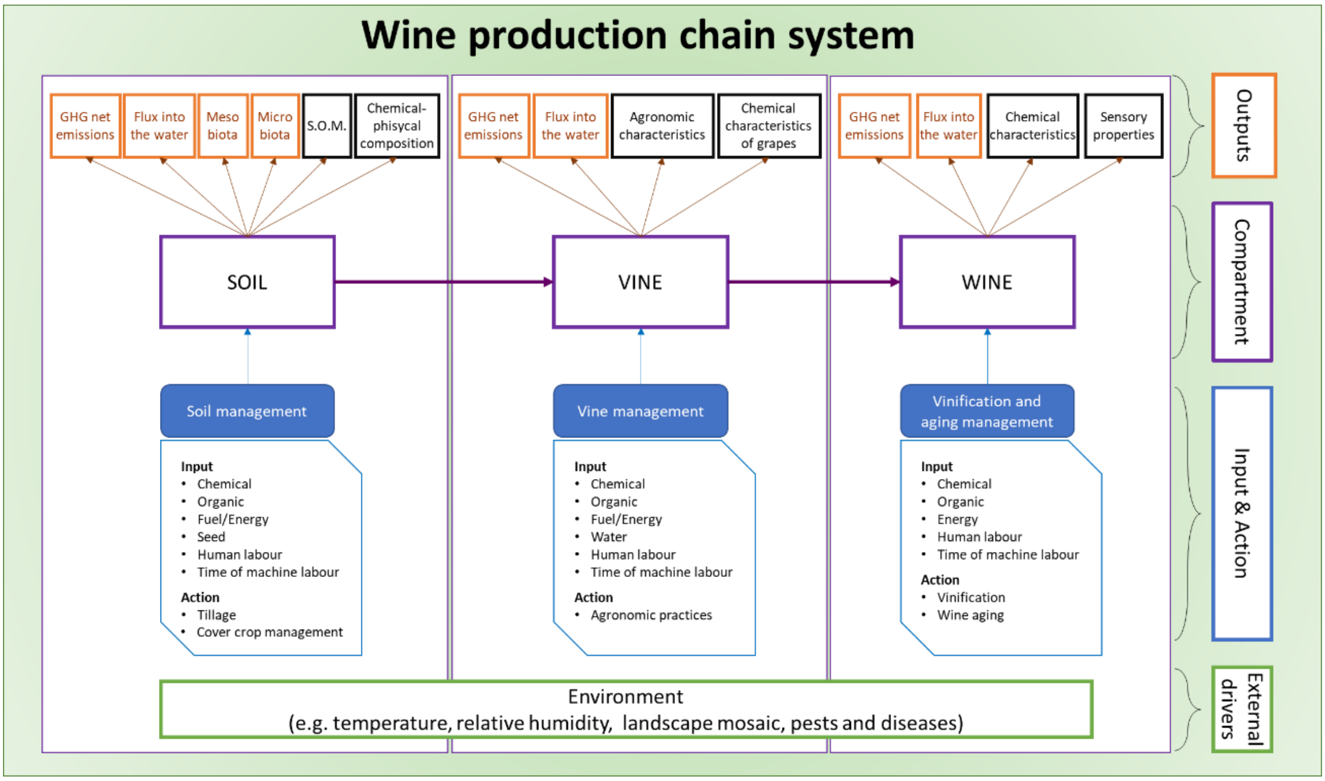


*Figure 6 - Diagram of the wine production chain. Input and actions (blue boxes), and production output (black boxes) are shown for each component, in addition to chemical and biological output (orange boxes).*

The DSS is made up of five modules for comprehensive evaluation of the wine company system:

- Module 1 addresses the evaluation of structural biodiversity of the farm considered, using indices proposed by ISPRA, an Italian public research institute for environmental studies and protection (Caporali, 2009);
- Module 2 assesses the sustainability of the company. It evaluates how the production processes fulfil the requirements of good practice for wine company sustainability. The evaluation system is based on application of Section 8 of the GEAvite® protocol, as defined in Valenti et al. (Valenti et al., 2013).
- Module 3 evaluates structural and chemical soil quality. Visual soil assessment is based on the protocol proposed by FAO (FAO, 2008), and chemical soil analysis considers the most commonly measured chemical components (e.g., organic matter content,

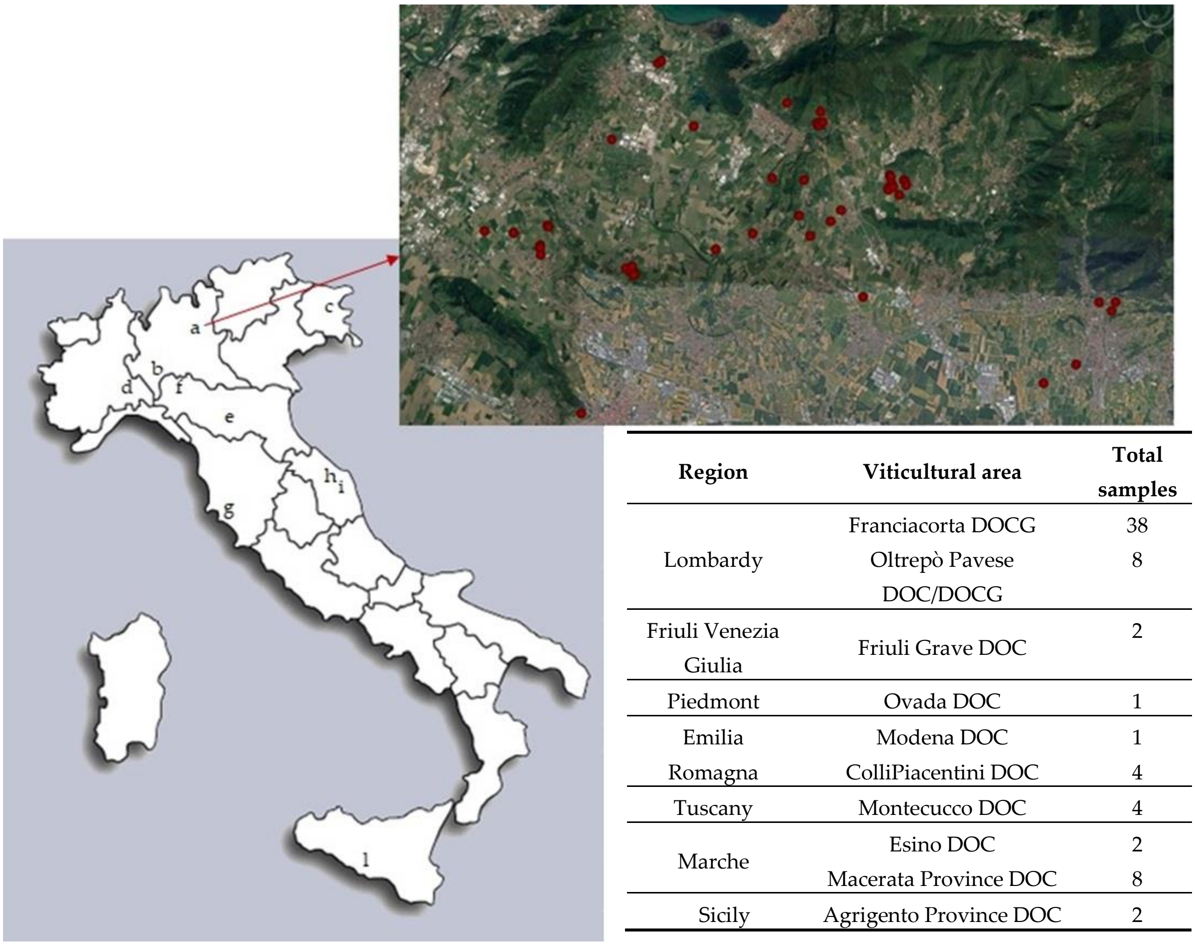
available phosphorous, potassium and magnesium, for further details see the Materials and Methods section).

- Module 4 considers soil penetration capacity, which is assessed with static penetrometric measurements (Eslami et al., 2020; Eslami & Fellenius, 2004).
- Module 5 addresses soil biodiversity, evaluating three major components of soil biota: earthworm presence and demographic structure, mycorrhizae presence, and soil arthropod biodiversity, through application of the QBS-ar index (V. Parisi, 2001). This paper focus on the latter component of module 5 with the aim of better understanding the environmental and agronomic influences on the QBS-ar index and then obtaining indication for improving the use of the DSS.

### 3.2.2. Study Sites

Soil data were collected in ten viticultural areas in Italy. A total of 70 soil samples were collected from 70 different vineyards (without replicates) over three years in 2014, 2015 and 2016. Sampling were performed in May (17%), June (74%), July (6%), and August (3%). All samples were analysed for QBS-ar and chemical characteristics of soil. Vineyards were in 10 different Italian wine growing areas. In Figure 7 locations and number of samples collected in each wine-growing area are presented.

The soil farming practices were defined by each wine company, however there were criteria allowing their classification into two categories: organic farming, if they are compliant to European Regulation on Organic Farming (reg (CE) n° 834/2007; reg (CE) n° 889/2008 and subsequent modifications and integrations), and conventional farming (for all other cases without any specific environmental certification). In total, 56 samples were from organic vineyards and 14 from conventional vineyards. Organic vineyards were divided into different groups, based on how long ago they had converted to organic production from conventional farming: 3 years or less, between 4 and 9 years, and 10 years or more.



| Region | Viticultural area | Total samples |
|---|---|---|
| Lombardy | Franciacorta DOCG | 38 |
| | Oltrepò Pavese DOC/DOCG | 8 |
| Friuli Venezia Giulia | Friuli Grave DOC | 2 |
| Piedmont | Ovada DOC | 1 |
| Emilia Romagna | Modena DOC | 1 |
| | ColliPiacentini DOC | 4 |
| Tuscany | Montecucco DOC | 4 |
| Marche | Esino DOC | 2 |
| | Macerata Province DOC | 8 |
| Sicily | Agrigento Province DOC | 2 |

*Figure 7 - Map of locations of different wine growing areas where samples has been collected. For Franciacorta DOCG a deeper level of detail is shown in relation to the high number of vineyards investigated.*

### 3.2.3. Explanatory Variables

To investigate the influence of temperature and rainfall on the QBS-ar the following variables were calculated for each soil sample, considering a period of 30 days before sampling:

- *Tmax_l:* number of days in which the daily maximum temperature was below 20 °C;
- *Tmax_m:* number of days in which the daily maximum temperature was between 20 °C and 30 °C;
- *Tmax_h:* number of days in which the daily maximum temperature was above 30 °C;
- *Prec_t:* total cumulative precipitation (mm);
- *Prec_l:* low precipitation period, if cumulative rainfall was 13.50 mm or below;
- *Prec_m:* medium precipitation period, if cumulative rainfall was between 13.50 mm and 186.51 mm;
- *Prec_h:* high precipitation period, if cumulative rainfall was 186.51 mm or above.

The thresholds used to define the levels of precipitation (low, medium or high) were calculated based on the range defined by the mean ± 1.5*standard deviation of total cumulative precipitation. Values of total cumulative precipitation inside this range were classified as "Prec_m", values lower than the lower bound of the range where classified as "Prec_l", while values greater than the upper bound of the range were classified as "Prec_h". Data were gather from 13 different meteorological stations (see Table 9). The soil sampling dates have been defined in order to avoid or minimize the possibility of overlapping the 30-day observation period of the meteorological variables for sites that refer to the same weather station.

*Table 9 -Location of meteorological stations.*

| **Location** | **Region** | **Latitude (Geographical Coordinates)** | **Longitude (Geographical Coordinates)** | **Altitude- m a.s.l** |
|---|---|---|---|---|
| Ancona | Marche | 43.617 | 13.517 | 103 |
| Udine/Campoformido | Friuli Venezia Giulia | 46.033 | 13.183 | 94 |
| Cuneo Levaldigi | Piedmont | 44.533 | 7.617 | 386 |
| Novi Ligure | | 44.767 | 8.783 | 187 |
| Piacenza | Emilia Romagna | 44.913 | 9.723 | 139 |
| Modena | | 44.65 | 10.95 | 33 |
| Arezzo | Tuscany | 43.467 | 11.85 | 249 |
| Sciacca | Sicily | 37.517 | 13.083 | 125 |
| Gela | | 37.083 | 14.217 | 33 |
| Monticelli | Lombardy | 45.622 | 10.091 | 230 |
| Corte Franca | | 45.633 | 10.021 | 220 |
| Erbusco | | 45.592 | 9.972 | 215 |
| Rodengo Saiano | | 45.596 | 10.124 | 160 |

Meteorological data were obtained from the meteorological service of the Brescia province for the Lombardy region and from the American NOAA Global Surface Summary of the Day (GSOD) dataset for all the other regions.

Characterisation of the soil chemistry involved consideration of: soil texture (sand, clay and loam content expressed in g/kg of soil), pH, cation exchange capacity (CSC expressed in cmoli$_{(+)}$/kg of soil), total and active limestone (expressed in g $CaCO_3$/kg of soil), organic matter content (expressed in g/kg of soil), available phosphorus (mg $P_2O_5$/kg of soil), available potassium (mg $K_2O$/kg of soil) and magnesium (mg MgO/kg of soil).

Samples were taken at a depth of 0–20 cm (excluding the first cm of turfgrass) and mixed uniformly. The collected soil samples were air-dried, homogenized and passed through a 2 mm sieve for chemical analysis. Chemical analysis was performed by the Chemical Laboratory of the Fondazione Edmund Mach (San Michele all'Adige, Trento, Italy) according to the Italian regulation (DM 13/09/1999).

### Soil biological Quality Evaluation (QBS-ar)

For this survey, a cubic sample of soil (with a side length of 12 cm) was collected at each site, at the same depth described for chemical soil analyses. The sample was placed in a Berlese–Tullgren funnel and under a 100 W incandescent bulb, the soil was warmed until complete dehydration. Small invertebrates tend to migrate away from the light and take refuge in the damp part of the soil sample (the bottom), then abandoning the soil and dropping into the cavity of the tunnel, from where they slip into a preserving solution.

Division into biological forms was carried out in relation to the characteristics of adaptation to the soil, which makes it possible to associate each systematic group with a numerical value defined as the "Ecomorphological Index" (EMI): the higher the EMI value, the higher the number of morphological characteristics linked to adaptation to the soil. The EMI value ranges from 1 to 20 (V. Parisi, 2001).

For some systematic groups there is a uniform level of adaptation to edaphic life for various species; in this case assignment of a single EMI value is envisaged. Vice versa, for species with different adaptation to soil, increasing EMI values are assigned according to increased adaptation (e.g. no wings, no eyes, etc.). If several biological forms are recognised in a group and therefore different EMI values are attributed, only the highest EMI value is considered for calculation of the QBS-ar, which represents the maximum degree of adaptation to life in the soil shown by the group under examination. Calculation of the QBS-ar index value is obtained from the sum of the EMI values attributed to each systematic group. QBS values can vary from a minimum of 0 to a maximum of 349.

## 3.2.4. Statistical Analysis

To analyse the linear relationships between the selected explanatory variables (regressors) and the QBS-ar, a multiple linear regression model (MLR) with a stepwise approach (Venables & Ripley, 2002) was applied. We adopted a bidirectional stepwise method, i.e., an automatic procedure to select the best set of explanatory variables in a large set of potential regressors that could explain the variance of the dependent variable. At each iteration of the algorithm, the decision to insert or to delete a regressor was based on minimisation of the Akaike Information Criteria (AIC) (Akaike, 1974). The full model to be tested with a stepwise regression model considered the following variables as regressors: type of farming (organic or conventional), years of organic farming, pH, soil texture, total limestone, active limestone, soil organic matter, available phosphorus, available potassium, available magnesium, Tmax_l, Tmax_m, Tmax_h, Prec_t, Prec_l, Prec_m and Prec_h. The final model adopted was made up of the subset of regressors resulting statistically significant, based on the stepwise procedure. Statistical analysis was performed using R software (version 3.5.3), `MASS` package (Venables & Ripley, 2002).

### 3.2.5. Results

The analysed dataset is made up of 70 observations. Surveys were performed from May 2014 to June 2016. The descriptive statistics for explained variables included in the full model are shown in Table 10.

*Table 10 - Descriptive statistics for quantitative variables in the dataset.*

| | **Unit of Measure** | **Mean ± SD*** | **Median (Q25–Q75)** | **Min** | **Max** |
|---|---|---|---|---|---|
| QBS-ar | | 92.29 ± 40.32 | 84.00 (59.25–127.75) | 28.00 | 193.00 |
| Years of organic farming | | 4.11 ± 5.02 | 2 (1.00–7.00) | 0.00 | 20.00 |
| pH | | 7.33 ± 0.89 | 7.75 (6.63–8.00) | 5.30 | 8.40 |
| Active limestone | (g $CaCO_3$/kg) | 31.70 ± 41.96 | 12.50 (0.00–57.00) | 0.00 | 130.00 |
| Soil organic matter | (g/kg) | 21.66 ± 9.31 | 21.00 (15.00–27.75) | 5.00 | 42.00 |
| Assimilable phosphorus | (mg $P_2O_5$/kg) | 34.40 ± 23.99 | 27.00 (17.00–47.50) | 5.00 | 94.00 |
| Exchangeable potassium | (mg $K_2O$/kg) | 191.10 ± 118.11 | 156.00 (114.20–219.50) | 60 | 747.00 |
| Exchangeable magnesium | (mg MgO/kg) | 362.50 ± 310.82 | 259.50 (159.00–433.50) | 72 | 1585.00 |
| Tmax_l | | 3.00 ± 2.63 | 2.00 (1.00–6.00) | 0.00 | 8.00 |
| Tmax_m | | 24.51 ± 3.17 | 24.00 (22.25–27.00) | 16.00 | 30.00 |
| Tmax_h | | 2.49 ± 3.46 | 1 (0–4) | 0.00 | 14.00 |
| Prec_t | | 100.01 ± 57.67 | 79.50 (51.00–153.10) | 0.00 | 190.60 |

*SD: standard deviation, Q25: first quantile of distribution, Q75: third quantile of distribution.

The distribution of all soil parameters showed high variability, due to the extensive heterogeneity characterising the study sites. Considering the reference period (30 days before each soil sampling), the mean number of days in which the daily maximum temperature was between 20 and 30°C was 24.51 ± 3.17 per site. Total precipitation ranged between 0 and 190.6 mm in the reference period. Only 1 observation recorded cumulative rainfall of less than 13.50 mm (Prec_l), in 59 sites cumulative rainfall was between 13.5 and 186.51 mm (Prec_m) and 10 observations recorded cumulative rainfall of over 186.51 mm (Prec_h).

The mean QBS-ar value in the dataset was 92.29 ± 40.32 (QBS-ar values ranging between 28 and 193). The different arthropod biological groups identified, and their EMI values are shown in Table 11. The distribution of groups identified ranges between 3 and 14 (the mean value is 7).

The results of the bidirectional stepwise linear regression model are shown in Table 12. All the estimated coefficients are statistically significant (p-value < 0.1), except the estimated coefficient of Tmax_l, though this regressor was selected by the stepwise procedure. The adjusted R-squared value was 0.1945.

*Table 11 - Groups identified in all the analysed samples. The EMI scores has been attributed following values proposed by Parisi et al. (2001).*

| Group | EMI Value | EMI Attribution |
|---|---|---|
| Acari | 20 | Default value |
| Araneae | 1 | Forms >5 mm EMI 1 |
| | 5 | Small forms scarcely pigmented EMI 5 |
| Chilopoda | 10 | Forms >5 mm, well-developed legs EMI 10 |
| | 20 | Other forms EMI 20 |
| Coleoptera | 1 | Clearly epigeous forms EMI 1 |
| | 5 | Clearly epigeous forms (EMI1), dimensions smaller than 2 mm (additional points 4) |
| | 10 | Clearly epigeous forms (EMI1), dimensions smaller than 2 mm (additional points 4) and the occurrence of two of the following conditions:<br>• thin integument, often testaceous (tan-brown) colour (additional points 5)<br>• hind wings highly reduced or absent (additional points 5)<br>• microphtalmy or anophtalmy (additional points 5) |
| | 20 | Clearly epigeous forms (EMI1), dimensions smaller than 2 mm (additional points 4), thin integument, often testaceous (tan-brown) colour (additional points 5), hind wings highly reduced or absent (additional points 5), microphtalmy or anophtalmy (additional points 5). |
| Collembola | 1 | Clearly epigeous forms: middle to large size, complex pigmentation present, long, well-developed appendages, well developed visual apparatus (eye spot and eyes) |
| | 4 | Small size—though not necessarily—forms, usually limited to litter, with modest pigmentation, average length of appendages, developed visual apparatus |
| | 8 | Hemi-edaphic forms with reduced number of ommatidia, scarcely developed appendages, often short or absent furca, pigmentation present |
| | 10 | Eu-edaphic forms with no pigmentation, reduction or absence of ommatidia, furca present—but reduced |
| | 20 | Clearly eu-edaphic forms: no pigmentation, absent furca, short appendages, presence of typical structures such as pseudo-oculi, developed postantennal organs (character not necessarily present), apomorphic sensorial structures |
| Diplopoda | 20 | Forms <5 mm EMI 20 |
| Diplura | 20 | Default value |
| Diptera (larvae) | 10 | Default value |
| Hemiptera | 1 | Mostly epigeous (above-ground) or root feeding forms |
| Hymenoptera | 1 | Default value without Formicidae |
| | 5 | Formicidae |
| Isopoda | 10 | Default value |
| Opiliones | 10 | Default value |
| Palpigradi | 20 | Default value |
| Pauropoda | 20 | Default value |
| Protura | 20 | Default value |
| Pseudoscorpiones | 20 | Default value |
| Psocotteri | 1 | Default value |
| Symphyla | 20 | Default value |
| Thysanoptera | 1 | Default value |
| Other holometaboulos insects (larvae) | 10 | Default value |
| Other holometaboulos insects (adults) | 1 | Default value |

*Table 12 - Results of the bidirectional stepwise linear regression model.*

| Coefficient. | Estimate | Standard Error | *p*-value |
|---|---|---|---|
| Intercept | 64.82 | 20.36 | 0.0023 |
| Organic farming | 40.21 | 13.08 | 0.0031 |
| Tmax_l | -3.21 | 2.26 | 0.1614 |
| Tmax_h | -4.89 | 1.57 | 0.0028 |
| Prec-t | 0.23 | 0.13 | 0.0777 |
| Prec-h | -40.78 | 21.87 | 0.0668 |

As shown in Table 12, the model applied identified a positive relationship between QBS-ar and organic farming systems. In Figure 8, we have plotted the distribution of QBS-ar according to farming systems. The QBS-ar median value is higher for organic farming than for conventional farming, although its distribution is characterised by greater variability compared to conventional systems.

The stepwise procedure did not select the number of years of organic farming as significant in describing the variability of QBS-ar.

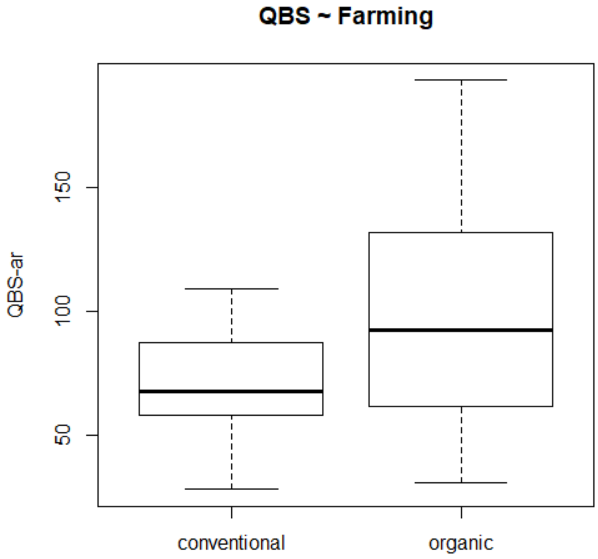


*Figure 8 - Boxplot of QBS-ar distribution according to farming system.*

The statistically significant relationships between QBS-ar and meteorological variables identified by the stepwise regression model can be summarised as follows:

- High daily maximum temperatures (Temp_h) were negatively related to the QBS-ar value, i.e., the number of days when the maximum daily temperature exceeded 30°C in the30 days before sample collection increased the QBS-ar value;
- Total cumulative precipitation (Prec_t) was positive related to QBS-ar, i.e., an increase in total precipitation increased the estimated QBS-ar value.

The model also shows that high precipitation (Prec_h) was negatively related to QBS-ar, i.e., each mm of rainfall accumulated in the 30 days preceding the soil sampling decreased the estimated QBS-ar value by 40.78 points. The relationships highlighted in these two areas are represented graphically in Figure 9.

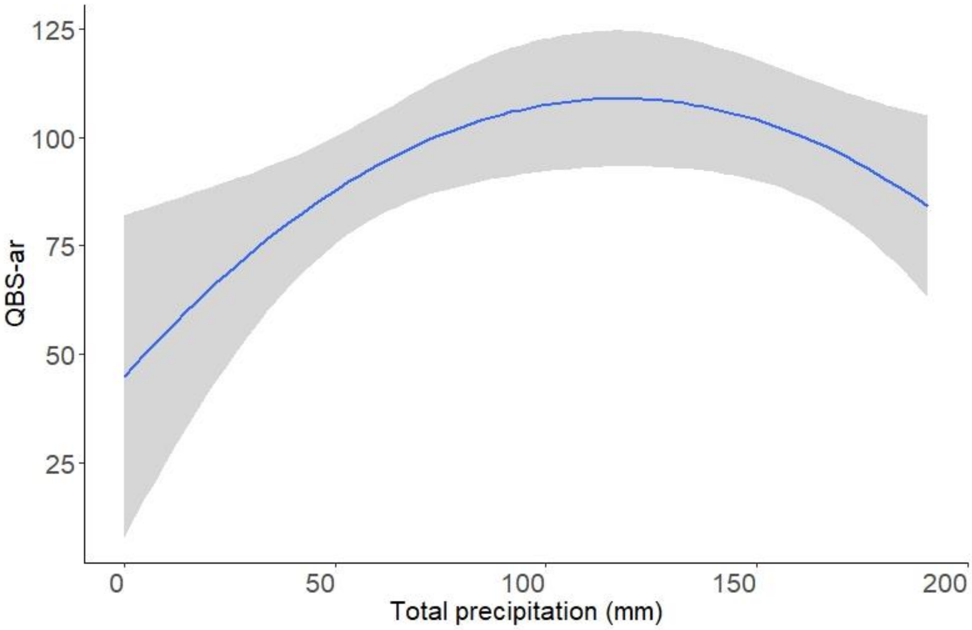


*Figure 9 - QBS-ar value distribution in relation to total cumulative precipitation values.*

The stepwise procedure selected low daily maximum temperatures (lower than 20°C) as a significant regressor explaining QBS-ar variance, but estimation of the regression coefficient of the bidirectional stepwise linear regression model was not statistically significant.

The model applied did not identify any significant relationship between QBS-ar and pedological variables. This may be related to the fact that the nexus between soil characteristics and QBS-ar was not linear, or mean that some dynamics explaining this nexus were captured by meteorological and management variables.

### 3.2.6. Discussion and conclusions

The mean QBS-ar index value in our dataset was 92.29±40.32, in agreement with previous studies observing a mean QBS-ar value of 91 for agricultural land (Cristina Menta, Conti, & Pinto, 2018). As regards the range of values in the index, in the literature a range of between 40 and 204 has been identified, depending on the year of sampling and the age of vineyards (Costantini et al., 2015), while other authors have recorded a range of between 98 and 203 depending on farming systems (Gagnarli et al., 2015). In our dataset, the QBS-ar values ranged from 28 to 193.

The results of the linear regression model showed a positive relationship between QBS-ar and organic farming, which is supported by previous studies identifying the same behaviour in different contexts (Hansen et al., 2001; Vittorio Parisi et al., 2005).

Important relationships between meteorological conditions in the 30-day period before soil sampling and the QBS-ar index were observed. A positive linear relationship with cumulative precipitation and a negative relationship with high temperature (≥ 30 °C) and high precipitation (>186.51 mm) were highlighted.

The effect of precipitation on the QBS-ar index is however known, as Parisi et al. (Vittorio Parisi et al., 2005) suggests collecting QBS-ar samples when soil moisture ranges between 40% and 80% of field capacity. Other studies reported a positive effect of rainfall on microarthropod abundance (Costantini et al., 2015) always in relation to soil moisture conditions. Moreover, the seasonal variability of soil fauna has already been observed by other authors (Galli et al., 2014; Cristina Menta et al., 2011; Neave & Fox, 1998; Tabaglio et al., 2009), although Neave and Fox (1998) related this behaviour to mechanical and physical soil modifications rather than meteorological effects. Effect of temperature have been less investigated even if Costantini et al. (2015) highlight a positive relationship between biological diversity and temperature.

Microarthropods are very sensitive to the variations of edaphic microhabitat in which they are living (soil pores). Thermal variations and water content of this microhabitat are then factors that generate avoiding behaviours because this creates stress in the transpiration and leads to anoxic conditions due to the saturation of soil pores. In the same way, in presence of limited water contents of the soil, a stress due to water loss occurs. Vertical migrations take then place in the direction of higher depth where variations are more mitigated. This behaviour may be related to the ability of soil organisms to migrate along the soil profile, to avoid stress conditions, and then not being found in the analysed soil sample (Clark et al., 2007; Stillman, 2003). This suggests that the sampling design must considering the depth of the different samplings collection. This is also aimed to verify the hypothesis of “vertical migration” as adaptive mechanism to stress conditions. Besides the migratory behaviour, variations in the abundance of microarthropods may also be attributed to the mortality increase, that stress conditions can generate (Villani et al., 1999).

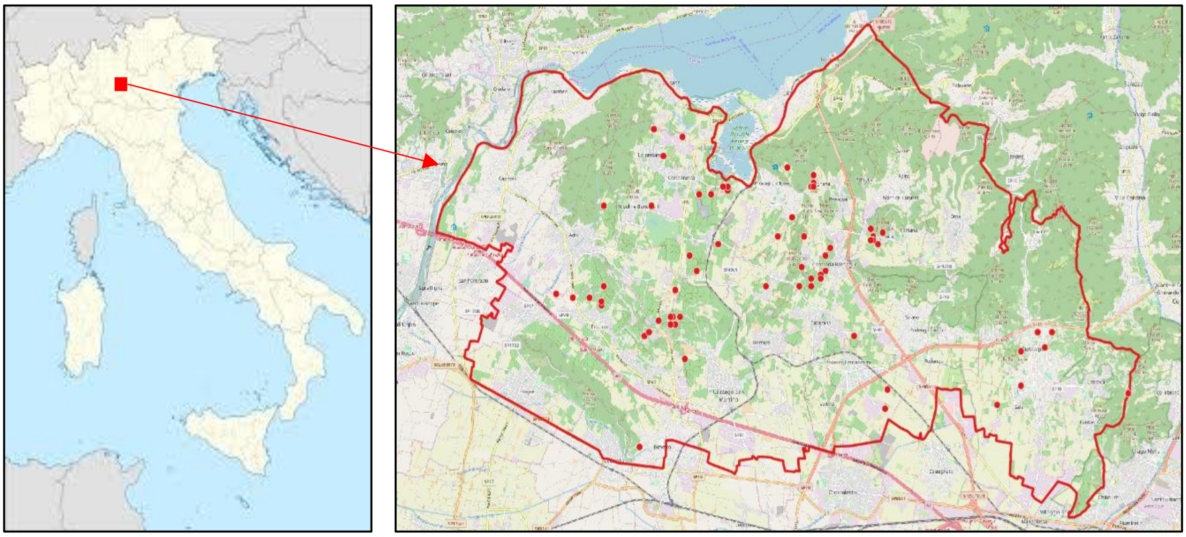
The relationship between meteorological conditions and QBS-ar identified in this paper helps to go beyond the seasonality effect. The identification of a clear link between QBS-ar and the level of precipitation and temperature makes it possible to improve the accuracy of the analysis, highlighting differences within the same season, or between the different years of observations.

The results described give useful indications for wine sector operators, providing them with some fundamental rules regarding arthropod dynamics in vineyard soil. In relation to the large number of observations considered and the variability of the geographical context observed, the DSS applied to wine sector biodiversity can be improved with indicators increasing knowledge about the timing of sampling and viticulture farming management assessment. Further research is needed to increase understanding of the meteorological relationship observed and the relationship between vineyard management and the biological quality of soil.

## 3.3. Response of the Arthropod Community to Soil Characteristics and Management in the Franciacorta Viticultural Area (Lombardy, Italy)

This study analysed diversity and co-occurrence patterns of different taxa of soil arthropods in relation to abiotic factors, such as soil temperature, soil moisture and soil chemical properties. Moreover, the influence of vineyard management (conventional vs. organic) and the time of conversion from conventional to organic on arthropod biodiversity was investigated.

This study was carried out in a major Italian winemaking area: Franciacorta (Lombardy, Italy). Franciacorta is the most famous Italian wine region to produce sparkling wine using the champenoise method and is located in the Lombardy Region . The zone covers a total area of 2615 ha (as of 2018) and hosts 117 wineries (as of 2019). This research collected a total of 100 soil samples from 100 different vineyards over the period 2014–2018. Eighty-five per cent of samples were collected in spring (May or June) and 15% of samples in autumn (September, October or November). All the samples were characterised by presence of arthropods and the chemical characteristics of soil.



In Figure 10, the location of each sampling site is shown.

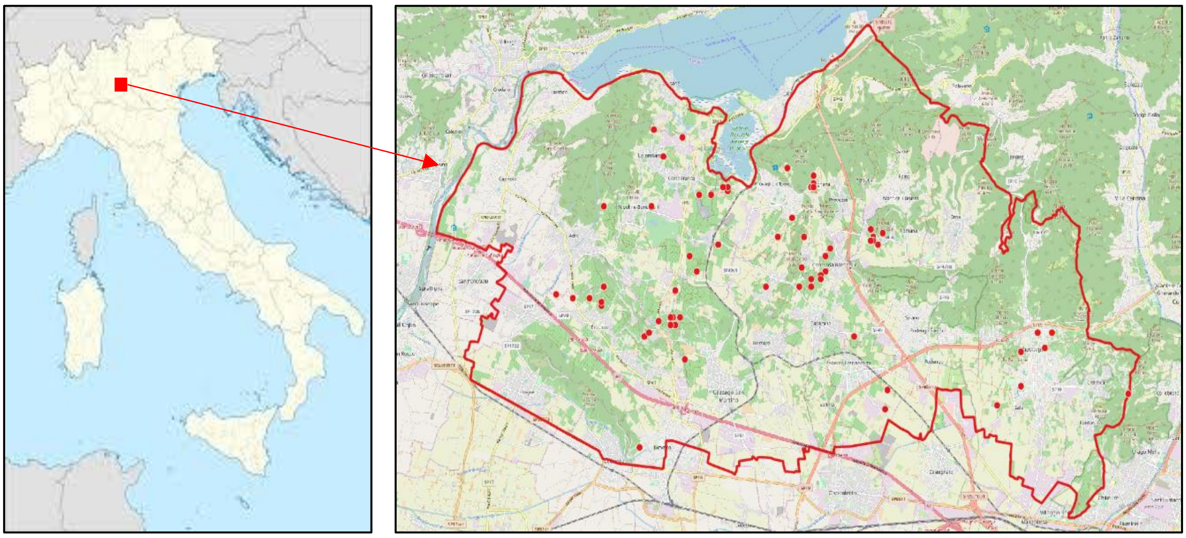


*Figure 10 - Map of the Franciacorta DOCG winegrowing area. The locations of vineyards where samples were collected are indicated with red dots.*

Vineyard management systems were classified in two main groups: conventionally managed vineyards without any specific environmental certification (conventional) and organic vineyards managed in compliance with the European Regulation on organic farming (reg EC n. 2018/848 and subsequent amendments and additions) (organic). For organic vineyards, we refer to the presence of this certification that implies compliance with the provisions of the law.

In addition to this, we have verified a minimal set of conditions occurred in each farm monitored in organic farming. These actions refer to: no use of synthetic chemicals for plant protection and for fertilizing the vineyard; the integration of organic matter into the soil through the supply of organic matrices; the total absence of use of herbicides and the management of the sub-row through mechanical intervention; the preservation of the herbaceous covering on the ground; and the minimum tillage adoption. Organic vineyards were then further divided into three subgroups, based on how long ago they had been converted from conventional to organic farming: 3 years or less (organic ≤ 3), between 4 and 9 years (4 ≤ organic ≤ 9), and 10 years or more (organic ≥ 10).

### 3.3.1. Explanatory Variables

Soil moisture (SM) and soil temperature (ST) data for the Franciacorta area from 2014 to 2018 were obtained from the National Centers for Environmental Predictions (Commerce, 2015). These data were then re-analysed using the Weather Research and Forecasting (WRF) simulations (Powers et al., 2017). The WRF model (version 4.02) was applied to a high spatial resolution grid (each cell of the grid representing a 2 × 2 km area) to generate hourly data. In particular, the Noah scheme (Ek et al., 2003) has been used as land surface model (LSM) scheme (i.e., Noah, Noah-MP, and CLM4) to assess detailed multi-layer soil moisture and soil temperature. We focused on a depth of 0–15 cm below land surface. Each vineyard was associated with the nearest grid node to allow extraction of the specific soil temperature and moisture values.

To assess the influence of environmental variables on the presence of soil arthropods, SM and ST were evaluated for each vineyard in a 30-days reference period prior to the sampling date. Two thermal thresholds ( $\mathrm{ST_{low}}$ and $\mathrm{ST_{up}}$) were considered to define two intervals of temperature. These intervals characterise organisms that prefer lower temperature features (taxa occurring more frequently in the interval $[\mathrm{ST_{low}}, \mathrm{ST_{up}}]$) or higher temperature features (taxa occurring more frequently when soil temperature is higher than $\mathrm{ST_{up}}$). The lower threshold ($\mathrm{ST_{low}}$) was set at 10 °C and the upper threshold ($\mathrm{ST_{up}}$) was set at 20 °C (Eisenbeis & Wichard, 1987). Soil temperatures lower than 10 °C were not included since they can be considered to be below the lower development threshold for most taxa. Considering $\mathrm{ST_{low}}$ and $\mathrm{ST_{up}}$, two variables related to soil temperature were calculated for the reference period (720 h):

$$\mathrm{TL} = \frac{1}{24}\sum_{\mathrm{i}=1}^{720}(\mathrm{ST_i} - \mathrm{ST_{low}}), \qquad \mathrm{ST_{low}} \leq \mathrm{ST_i} < \mathrm{ST_{up}}$$

$$\mathrm{TH} = \frac{1}{24}\sum_{\mathrm{i}=1}^{720}\left(\mathrm{ST_i} - \mathrm{ST_{up}}\right), \qquad \mathrm{ST_i} \geq \mathrm{ST_{up}}$$

where $\mathrm{ST_i}$ is the hourly soil temperature. TL is the daily cumulative soil temperature degrees exceeding 10 °C when $\mathrm{ST_i}$ is between 10 °C and 20 °C; TH is the daily cumulative soil temperature degrees exceeding 20 °C when $\mathrm{ST_i}$ is greater than 20 °C.

A soil moisture threshold ($\mathrm{SM}$) was defined to discriminate between organisms that prefer drier conditions, i.e., taxa occurring more frequently when soil moisture ranges in the interval $[0, \mathrm{SM}]$, or wetter conditions, i.e., taxa that more frequently occur when soil moisture is in the range $[\mathrm{SM},1]$. SM was set equal 0.35 (corresponding to 35%), which represents a reference value that can be associated, in different ecological contexts, to a status that satisfies the requirements in terms of humidity of soil arthropods (Tsiafouli et al., 2005; Vikram Reddy & Venkataiah, 1990). Two variables associated with soil moisture were calculated for the reference period (720 h):

$$\mathrm{MD} = \frac{1}{24}\sum_{\mathrm{i}=1}^{720} |\mathrm{SM_i} - \mathrm{SM}|, \qquad 0 \leq \mathrm{SM_i} \leq \mathrm{SM}$$

$$\mathrm{MH} = \frac{1}{24}\sum_{\mathrm{i}=1}^{720} (\mathrm{SM_i} - \mathrm{SM}), \qquad \mathrm{SM_i} > \mathrm{SM}$$

where $\mathrm{SM_i}$ is the hourly soil moisture. MD is the daily sum of absolute deviations in soil moisture values from the threshold value when $\mathrm{SM_i}$ is lower than 0.35; MH is the daily cumulative soil moisture exceeding 0.35, when $\mathrm{SM_i}$ is higher than 0.35.

Chemical analysis of soils was performed according to the Italian regulation (DM 13/09/1999). Soil samples were taken at a depth of 0–15 cm and mixed homogeneously. Leaf litter was excluded, as it is not part of the soil itself. The collected soil samples were air-dried, homogenized, and passed through a 2 mm sieve for chemical analysis.

Characterisation of the soil chemistry involved measuring soil texture (TXT), pH, active limestone (expressed in g $CaCO_3$/kg of soil) (AL), organic matter content (expressed in g/kg of soil) (SOM), available phosphorus (mg $P_2O_5$/kg of soil) (P), available potassium (mg $K_2O$/kg of soil) (K), available magnesium (mg MgO/kg of soil) (Mg) and copper content (mg/kg) (Cu). Soil texture was classified following the USDA soil texture triangle classification(Soil Science Division Staff, 2017).

A cubic sample of soil (with a dimension of about 30 $cm^3$) was collected at the same depth described for chemical soil analysis, at each vineyard. Arthropods were extracted by placing the soil sample in a Berlese–Tüllgren funnel under a 60 W incandescence bulb, leading soil arthropods to migrate towards the damp part of the soil sample (away from the light). The soil arthropods fell through the cavity, into a preserving solution (2/3 alcohol and 1/3 glycerol). Determination of biological forms was carried out according to the QBS-ar method as proposed by (Cristina Menta, Conti, Pinto, et al., 2018), and the definition of the taxonomic entities and the biological stages is in agreement with the one reported in the same paper.

### 3.3.2. Statistical Analysis

To measure soil arthropods biodiversity a taxa co-occurrence approach was used. For each vineyard, a taxa presence profile was defined, i.e., a vector indicating the presence or absence of the taxa in each vineyard. The presence profile did not consider population abundance. Based on the presence profiles, vineyards and taxa were described in a J dimensional space (J is the number of taxa considered), allowing taxa to be ordered by their vineyard presence profiles. Two taxa are close to each other if they share a similar pattern of co-occurrence in the vineyards, they are far from each other if one is present in the vineyards where the other is absent and vice versa.

To allow easy visualisation and interpretation of dissimilarity in soil biodiversity and taxa co-occurrence, it is useful to represent these profiles in a two-dimensional space, called an ordination plane. Non-metric multidimensional scaling (NMDS) can be used to summarise information and reduce the dimensionality of profiles (Kenkel & Orloci, 1986). By applying NMDS, vineyards and taxa can be ordered by the dissimilarity of the presence profiles. Bray–Curtis dissimilarity (Clarke, 1993), used extensively in the ecological field, was adopted. NMDS analysis was performed using the metaMDS function of the vegan package in R (Oksanen et al., 2019). Loss of information due to a reduction in dimensionality is assessed by the stress value, which refers to the disagreement between 2-D representation and original positions of taxa in multidimensional space.

To test which environmental drivers (Cu, pH, AL, SOM, P, K, Mg, TL, TH, MD and MH) are significantly correlated to the first two axes of the NMDS ordination plane, we applied the `envfit` function of the `vegan` R package (Oksanen et al., 2019). Each variable was correlated independently and plotted on the plane as a vector. The direction of the vector represents the gradient direction of the environmental driver, while the length of the vector is proportional to the correlation of the ordination system and the environmental driver.

Taxa were grouped into clusters as homogeneous as possible in terms of co-occurrence patterns, based on taxa ordination results (Clarke, 1993). To perform hierarchical cluster analysis, the `hclust` function of R software (R Core Team, 2019) was applied.

To assess the impact of vineyard management on the biodiversity of soil biota, decision tree analysis was performed. The number of taxa present in each soil sample was considered as a measurement of edaphic biodiversity, and three categories of soil biodiversity were defined: ‘low’ when the number of taxa was lower or equal to 4, ‘medium’ when the number of taxa in the soil sample was between 5 and 8, and ‘high’ when the number of taxa was greater than 8. A classification decision tree allowed to split the soil samples into homogeneous groups according to edaphic biodiversity based on the different vineyard management classes. Recursive partitioning and regression tree (RPART) analysis were performed by applying the `rpart` package of R software (Therneau & Atkinson, 2018). The fitting of the model was investigated using the accuracy index that corresponds to the percentage of cases correctly classified.

### 3.3.3. Results

The descriptive statistics for environmental variables included in the full model are shown in Table 13. Seven types of soil texture were considered: clay, clay loam, silty clay loam, sandy clay loam, loam, silt loam and sandy loam.

Vineyard management was categorised into four classes: conventional management (7% of the sample), vineyards converted to organic farming in the last three years (45% of the sample), vineyards converted between 4 and 9 years ago (31% of the sample), and vineyards converted at least 10 years ago (17% of the sample).

A total of 19 taxa were identified in the soil samples. In case of Diptera and Coleoptera, the biological stage of larvae was also detected Table 14. Collembola, Acari and Hymenoptera recorded the highest frequency of presence in the soil samples analysed. Collembola and Acari were reported in 89 of the 100 vineyards, Hymenoptera in 80 vineyards. The lowest frequency of occurrence was recorded for Psocoptera, Thysanoptera and Isopoda (8/100, 7/100, 6/100 respectively).

*Table 13 - Descriptive statistics of continuous variables (soil characteristics and environmental drivers) included in the analysis.*

| | **Unit of Measure** | **Mean ± Standard Deviation** | **Median (Q25-Q75)** | **Min** | **Max** |
|---|---|---|---|---|---|
| **pH** | | 7.10 ± 0.87 | 7.30 (6.35–7.9) | 5.30 | 8.20 |
| **SOM** | (g/kg) | 21.94 ± 9.08 | 23.00 (15.00–25.00) | 5.00 | 42.00 |
| **Cu** | (mg/kg) | 58.68 ± 32.81 | 55.40 (36.9–72.2) | 4.20 | 170.00 |
| **P** | (mg $P_2O_5$/kg) | 54.47 ± 40.20 | 51.00 (26.00–64.00) | 9.00 | 222.00 |
| **K** | (mg $K_2O$/kg) | 148.52 ± 67.53 | 145.00 (94.00–178.00) | 60.00 | 354.00 |
| **Mg** | (mg MgO/kg) | 165.75 ± 75.37 | 138.00 (117.00–210.00) | 66.00 | 383.00 |
| **TL** | °C | 68.76 ± 44.90 | 69.75 (33,67–104.39) | 0.00 | 161.46 |
| **TH** | °C | 106.05 ± 59.06 | 121.76 (55.73–153.38) | 14.20 | 241.77 |
| **MD** | Pure number | 10.39 ± 26.03 | 5.16 (1.51–7.88) | 0.00 | 135.53 |
| **MH** | Pure number | 0.11 ± 0.20 | 0.00 (0.00–0.156) | 0.00 | 0.63 |

*Table 14 - Distribution of taxa according to stages considered in the analysis and presence (i.e., number of soil samples in which the taxon has been identified).*

| Taxa | Larvae | N° of Presences | Other Stages * | N° of Presences |
|---|---|---|---|---|
| *Acari* | | | *x* | 89 |
| *Myriapoda—Diplopoda* | | | *x* | 12 |
| *Myriapoda—Chilopoda* | | | *x* | 17 |
| *Myriapoda—Symphyla* | | | *x* | 56 |
| *Myriapoda—Pauropoda* | | | *x* | 32 |
| *Hymenoptera* | | | *x* | 80 |
| *Thysanoptera* | | | *x* | 7 |
| *Pseudoscorpionida* | | | *x* | 11 |
| *Psocoptera* | | | *x* | 8 |
| *Coleoptera* | | | *x* | 31 |
| *Coleoptera larvae* | *x* | 39 | | |
| *Collembola* | | | *x* | 89 |
| *Diptera* | | | *x* | 31 |
| *Diptera larvae* | *x* | 32 | | |
| *Protura* | | | *x* | 25 |
| *Diplura* | | | *x* | 27 |
| *Hemiptera* | | | *x* | 12 |
| *Isopoda* | | | *x* | 6 |
| *Other_holometabolous* [2] | | | *x* | 20 |

* Other stages include all forms that produce active participation in soil cycles (e.g., pupae are excluded). In the case of the 'Other_holometabolous' taxon, the pupal stage is also included. Two Other_holometabolous taxa include Mecoptera, Neuroptera and Raphidioptera orders in agreement with QBS-ar method (Cristina Menta, Conti, Pinto, et al., 2018).

## Co-Occurrence Pattern Identification

Taxa dispersion in the non-metric multidimensional scaling plane is shown in Figure 11. Taxa were ordered according to their co-occurrence profiles. Neighbouring taxa in the plane were characterised by the presence in the same vineyards (e.g., Collembola and Coleoptera larvae, Psocoptera and Pseudoscorpionida); the more distant are two taxa, greater is the difference in terms of their presence in the vineyards (e.g., Diptera and Psocoptera, Acari and Pauropoda). The stress value estimated for the model was equal to 0.2, indicating the model has good ability to predict data in the reduced space.

The results of analysis of the correlation between environmental drivers and the NMDS plane are shown in Table 15.

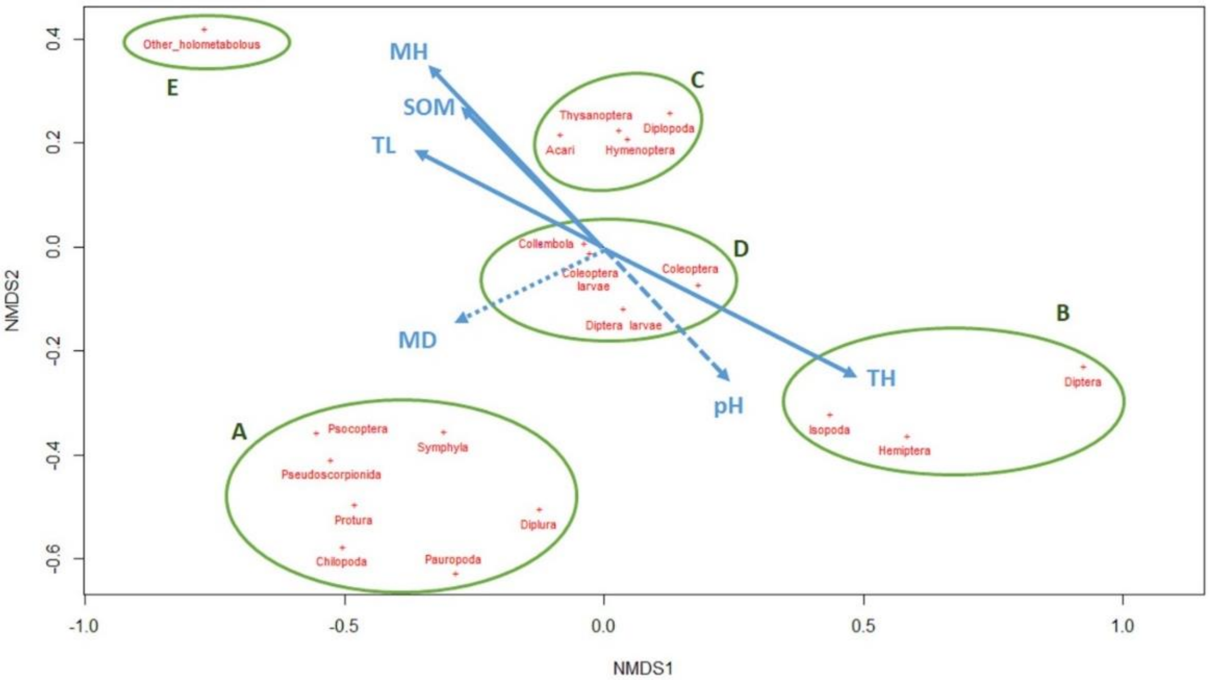


*Figure 11 - Results of non-metric multidimensional (NMDS) analysis: dispersion of taxa (red points) according to their co-occurrence profiles (NMDS1 and NMDS2 are the two axes of the ordination plane). Blue arrows refer to the correlation of environmental drivers and soil characteristics with NMDS ordination pattern (solid line —— p-value < 0.5, dashed line – – . p-value < 0.1, dotted line ........ p-value < 0.15). The five clusters of taxa according their presence pattern are highlighted with the green circles.*

*P*-values of the correlation coefficients were used to discriminate the intensity of the relationship between environmental drivers and the taxa ordering system (Table 3): strong correlation for SOM, TL, TH and MH (*p*-value < 0.05); medium intensity correlation for pH (*p*-value < 0.1); low intensity correlation for MD (*p*-value < 0.15). The other environmental drivers were not significantly correlated with the first two axes of the NMDS system.

*Table 15 - Correlation analysis of environmental drivers and soil characteristics with NMDS ordination pattern.*

| Variable | Squared Correlation Coefficient | *p*-Value [5] of Correlation Coefficient | |
|---|---|---|---|
| **Cu** | 0.05 | 0.17 | |
| **pH** | 0.06 | 0.09 | ** |
| **AL** | 0.02 | 0.48 | |
| **SOM** | 0.08 | 0.05 | *** |
| **P** | 0.02 | 0.43 | |
| **K** | 0.01 | 0.66 | |
| **Mg** | 0.01 | 0.64 | |
| **TL** | 0.08 | 0.04 | *** |
| **TH** | 0.15 | 0.01 | *** |
| **MD** | 0.05 | 0.15 | * |
| **MH** | 0.12 | 0.01 | *** |
| **TXT** | 0.04 | 0.83 | |

**p*-value < 0.15, ** *p*-value < 0.1, *** *p*-value < 0.05.

The results obtained from NMDS and cluster analysis (Figure 11) allowed the taxa to be divided into five groups according to their co-occurrence pattern. The five clusters shown in the cluster dendrogram correspond to the clusters identified by the green circles in the NMDS plane (Figure 12).

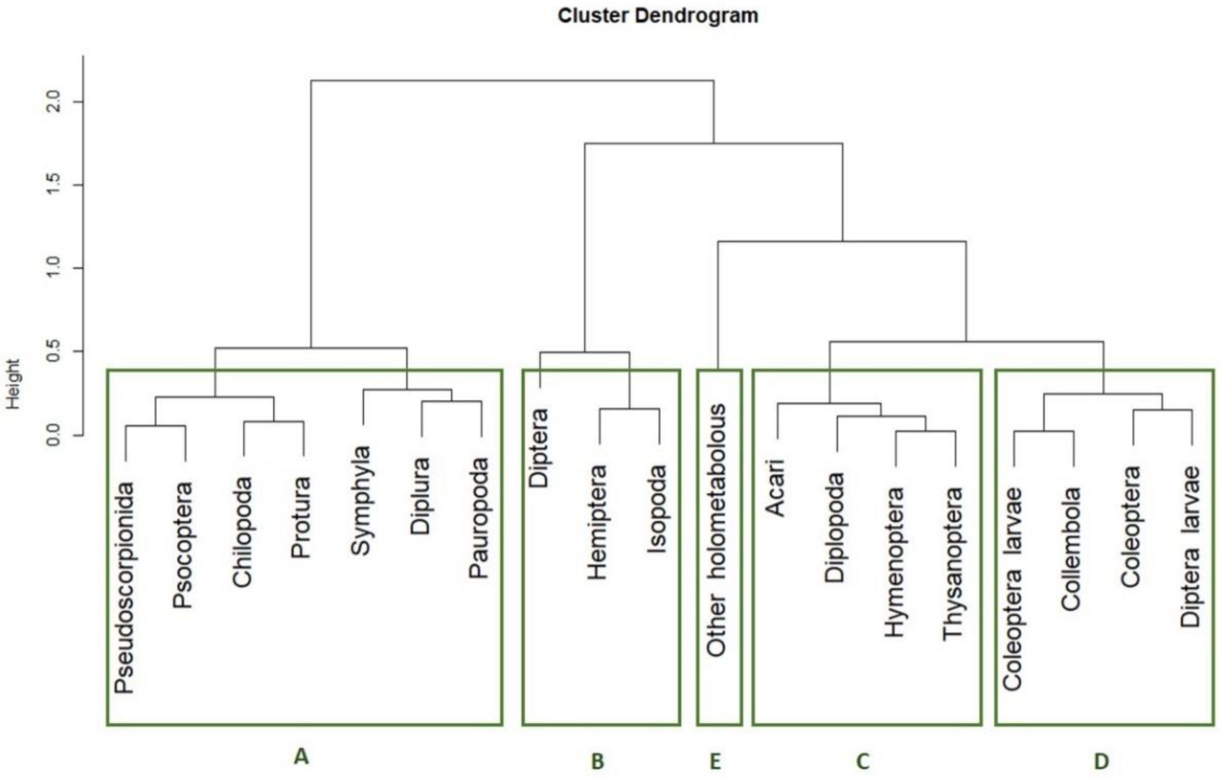


*Figure 12 - Dendrogram of hierarchical cluster analysis of taxa based on NMDS results. The five clusters are highlighted in green.*

Group A included the largest number of taxa and specifically the Pseudoscorpionida, Psocoptera, Protura, Diplura Chilopoda, Symphyla and Pauropoda. Group B was made up of

Diptera, Hemiptera and Isopoda taxa, while the larval form of Diptera was located in group D, together with Coleoptera, both as larvae and other biologic forms, and Collembola. The Acari, Hymenoptera, Thysanoptera and Diplopoda taxa made up group C. Group E is only represented by the taxa defined as 'Other_holometabolous'.

### Vineyard Management

The results of the classification tree showed that variable vineyard management could be useful for discriminating different categories of soil biodiversity. In particular, the analysis of the tree shown in Figure 13 showed that conventionally managed vineyards were associated with a low level of biodiversity, vineyards that had adopted organic management for a maximum of three years were associated with a medium level of biodiversity, and vineyards that had adopted organic management for at least four years were associated with a high level of biodiversity. The accuracy index showed a good fit of the model as 57% of cases was correctly classified.

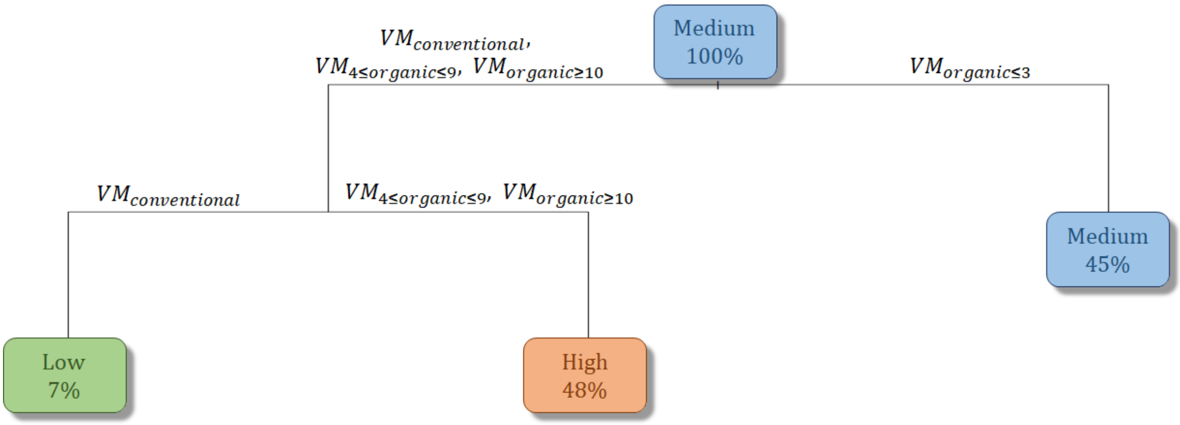


*Figure 13 - Classification decision tree of soil samples to predict soil biodiversity according to vineyard management. The predicted level of soil biodiversity (low, medium, high) is reported in the squared box, together with the percentage of soil samples included in that node. The paths from the initial box (with 100% of cases) to the final boxes represent the classification rules.*

## 3.3.4. Discussion and conclusions

The results obtained in this study allowed to identify the co-occurrence pattern for 19 taxa of soil arthropods based on a 5-year investigation carried out in the Franciacorta viticultural area (Lombardy, Italy). The NMDS showed significant relationships between investigated soil arthropod taxa and soil moisture (MD and MH), soil temperature (TL, TH), soil organic matter (SOM) and pH. The decision tree showed an increased taxa diversity in relation to organic vineyard management and to the increase of time period of conversion from conventional to organic management.

In line with the expectations, Collembola and Acari were the most frequent of the 19 taxa identified, confirming that they are the most present groups of arthropods in soil (André et al., 1994). The high level of presence of Hymenoptera recorded in our analysis is in agreement with other studies carried out in different agricultural contexts and reporting a significant presence of this taxon, mostly represented by Formicidae, in vineyard soils (Pérez Bote & Romero, 2012).

Based on taxa co-occurrence patterns, identified through NMDS analysis, five groups were found. Moreover, NMDS analysis made it possible to explore the relationship between soil abiotic variables and the aggregation of arthropod taxa in groups. In particular, the results obtained from our study pointed out that presence patterns characterising group A showed only one significant correlation ($p$-value < 0.15) with low soil moisture (MD). This result is compatible with the hypothesis that the taxa included in group A were relatively less dependent on high humidity values. The taxa in group B and the Coleoptera and Diptera larvae taxa (group D) were associated with higher pH ($p$-value < 0.1) and higher soil temperatures (TH) ($p$-value < 0.05), in line with the possible thermophilic habit of some representatives of these taxa (Eisenbeis & Wichard, 1987, 1987; Guodong Zhu et al., 2018). The relationship with a higher pH level is more evident for the Isopoda and this is in agreement with van Straalen (van Straalen, 1998), who underlined weakly alkaliphilous or sub-neutral behaviour for some species of Isopoda. The detected ubiquitous presence of Collembola (group D) could be partially explained by the significant variability of responses to soil temperature, moisture, and chemical properties of the different species of this taxon. In particular, the effect of soil moisture on Collembola has been documented by different authors (Choi et al., 2006; Sjursen & Holmstrup, 2004), while species-specific responses have been reported (Tsiafouli et al., 2005). Furthermore, Heiniger et al. (Heiniger et al., 2015) highlight that the role of microclimate for Collembola could be less important for their distribution than the role of trophic resources and competition. The presence of taxa in groups C and E is mostly determined by soil organic matter (SOM), soil moisture value higher than threshold level of 0.35 (MH) and lower temperature (TL) ($p$-values < 0.05). The relationship with SOM can be related to the involvement of these taxa in the soil food webs that starts from decomposition of dead organic matter generated by the activity of bacteria and fungi (Petersen & Luxton, 1982). Diplopoda (Millipedes) are involved in SOM degradation, as their feeding activity is focused on dead organic matter (Bagyaraj et al., 2016; Bertrand & Lumaret, 1992). A significant influence of soil nitrogen on species richness and biodiversity has been observed for this taxon (Stašiov et al., 2012), while Hymenoptera are involved in the decomposition of organic substances (Bagyaraj et al., 2016). In relation to the positive response of group C to soil moisture increase, some authors have underlined that soil water availability is an important factor controlling presence of mites (Acari) (Badejo, 1990). Other authors have showed that Oribatid mites (Acari: Oribatida) are positively influenced by soil temperature (J.-L. Liu et al., 2017) and that their distribution is dependent on soil moisture (Bagyaraj et al., 2016). The relationship observed between group E and soil moisture can be related for Mecoptera (included in Other holometabolous taxon) with data reported for pre-imaginal stages of this order which develop in the soil and showed preference for high soil moisture (Byers & Thornhill, 1983).

The co-occurrence pattern of the taxa identified in our study is in line with similar pattern reported in the literature. Taxa co-occurrence in group B agrees with the results in (Varga, 2003) that confirmed Diptera and Isopoda co-existence in some specific habitats. Acari and Hymenoptera (group C) have also been grouped together by other authors (Shakir & Ahmed, 2015). The composition of groups A and C suggests that the co-occurrence pattern can also be influenced by biotic relationships among taxa. According to Eisenbeis and Wichard (Eisenbeis & Wichard, 1987) (p. 192), the trophic niche of Diplura includes Symphyla, while Weygoldt (Weygoldt, 1969) noted that Pseudoscorpionida feed on different orders of small soil arthropods, including Psocoptera. All these associations support the taxa co-occurrence in group A. Similarly, Coleoptera contain taxa (e.g., Carabid beetle) that have been described as predators of Collembola. This association is in line with the co-occurrence of these two taxa in group D.

The results obtained analysing the role of vineyard management on soil arthropods diversity allows to identify an increase of taxa diversity in relation to organic vineyard management. This is consistent with previous studies, which reported a general increase of arthropod biodiversity (Caprio et al., 2015) and arthropod abundance (Gagnarli et al., 2015) associated to organic vineyard management. This effect was evident even before a 3-year period after conversion. The effect on arthropod biodiversity markedly increases with the length of the period since organic farming adoption.

The results obtained in this study provide additional knowledge supporting the interpretation of diversity and co-occurrence patterns in soil Arthropoda in vineyard. The importance of abiotic variables together with the interpretation of the possible role of biotic relationship among taxa have been explored in the specific geographic context of the Franciacorta viticultural area. Furthermore, our study confirmed the effect of organic vineyard management in increasing arthropod taxa diversity and, most importantly, it showed the critical role of the time of conversion from conventional to organic farming in increasing arthropod biodiversity. Further experiments are needed to extend these results to other viticultural contexts.

# 4. Network analysis in biodiversity assessment

In the first chapter the need to consider ecosystems as a network of interconnected and interdependent functional elements was introduced. Currently, the relationships between species, traits and ecosystem functions are still too poorly understood, particularly in complex biocenoses, either due to a high number of species in the same community or multiple ecosystem services generated by the same community. Overcoming this lack of knowledge is essential to provide informed and scientifically based support to decision makers. This is critical given the accelerating decline in global biodiversity (IPBES, 2019) because species - trait - function relationships underpin the provision of ecosystem services (Duncan et al., 2015).

Quantitative network analysis appears to be the most promising approach to investigate how the structures of ecosystem networks are related to the functional traits of the species. Thanks to the network approach, it is possible to study the large-scale statistical properties of the whole network, the properties of individual species and interactions, and how they jointly contribute to the structure, function and stability of ecosystems.

## 4.1.Modelling approaches

The biodiversity assessment is based on web of networks, each network is characterized by nodes and link/interactions. Nodes corresponds to the entities under investigation, e.g. individuals of the same species, different species characterized by having the same traits or sharing a niche, species at different levels of the trophic web, or the traits to be investigated. The interaction processes represented by the links between the nodes greatly differ according to the aim of the analysis. Network analysis allows to investigate negative - negative interactions (such as competition), negative - positive interactions (such as nutrition) and positive - positive interactions (such as pollination and seed dispersal). The level of complexity of the network can vary according to the degree of complexity of the represented system. Simpler networks can investigate the distribution of a species (hence the species-environment relationship), more complex relationships can be prey- plants-pollinators, hosts-parasitoids, soil arthropod communities - nutrient cycle. Even complex relationships can be managed with the network approach, for example pest species - natural enemies - ecological structures (Dehling et al., 2014; Eklöf et al., 2013; Ohlmann et al., 2019; Stang et al., 2009). For instance, non-trophic edges between coexisting species have been investigated using network approaches, as well as the dynamic and the functional structure of trophic chains (Legras et al., 2019).

### 4.1.1. Main features of quantitative models for network analysis

In this section, we briefly present the basic features that a model for the assessment of biodiversity according to a network approach must satisfy (Di Biase et al., 2018; Dietze et al., 2018; Evans, 2012).

The analysis of the traits of the species and of the functions requires the ability to manage very different types of data. We can have continuous quantitative variables (eg height, weight, CO2 emissions) or discrete (eg number of eggs, number of pollinated flowers), ordinal categorical variables (eg low, medium, high) or nominal (eg color of leaves, variety of a species), and boolean or dichotomous variables (eg true / false, present / absent, male / female). The measurement scales used in the model and the link functions must know how to combine evaluations made with one or more of these types.

Models must adequately deal with objects at different levels of aggregation. ecosystems can be considered as web of networks, based on the level of investigation considered we focus on one or more of these networks, but the information could have been collected at different resolution levels, e.g. there may be characteristics measured at the level of individual, species, cluster of traits and / or ecosystem. A good model for the assessment of sustainability must be able to disaggregate information when the level of analysis is at the level of a single individual or, on the contrary, generate summary information when the data is collected at the individual level and the analysis is carried out at the community or cluster level.

The processing of large amounts of data is a key feature for a networked model. a problem that is often encountered is the management of a matrix in which the number of variables is greater than the number of observations (fat matrix). In order to correctly manage this information, the models often have a phase of selection of the variables of interest based on the estimation objective. Another problem, not always correctly managed by the models, are the sparse matrices (i.e. number of non-null elements is significantly lower than the total number of elements in the matrix). It is a case that occurs frequently when the object of the analysis is a trait or a function that is possessed by a limited number of individuals / species in the biocenosis. Models that know how to handle this data correctly are called zero-inflated models.

The exploration of the existing relationships between traits-functions and ecosystem services requires the adoption of an adaptive approach, as the model is not fully specified. Adaptive management is defined as ‘systematic acquisition and application of reliable information to improve management over time’ (Wilhere, 2002). In essence, the model must allow for adjusting its behaviour in response to meaningful changes in the operational context. Adaptive approach to manage uncertainty and to improve long-run investigation of the system

A final essential feature of the models for network analysis is the modularity of the model, based on an incremental approach to the knowledge of the system. Once the system has been designed, it may be that not all its components (nodes and / or interactions) can be known at the same time in the same way. There may be components that, for various reasons, must be hypothesized, that is, to which estimates not directly generated by data are associated. If it is possible to support the estimation of that component through data and new information, the model must allow it, without changing the architecture of the whole network. Furthermore, there may be nodes that at the time of the estimate are considered as unique entities, but which in a second moment need to be "exploded" as knowledge has been acquired that is capable of treating the entity itself as a network. The module approach allows to develop a general analysis of the relationships between modules and to deepen each module based on the information and data available.

### 4.1.2. The times are ripe

The complexity underlying the biodiversity assessment according to network approach was the primary cause that in the past slowed the development of quantitative models for this purpose. Quantitative ecology, which has ancient roots, has developed over time simplified approaches in terms of relations considered, such as dynamic approaches to trophic network analysis through the extension of Lotka-volterra equations. In recent decades, particularly in the new millennium, the conditions for an important development of quantitative tools from a network analysis point of view have been created. Factors that contributed to proliferation of models to achieve precise and unbiased ecological prediction are: great increase in the availability of ecological data, increased understanding of ecological systems, increase in computing power analysis, diffusion of a quantitative culture in many natural sciences courses.

The increase in data availability has several origins. The diffusion of remote sensing tools allowed the collection of big masses of data, in (almost) real time, at very low costs. International funding programs for ecological research have led to the development of large projects for data collection and processing. This factor was instrumental towards the definition of international data collection standards. In fact, a systemic approach requires that the data populating the models be as homogenous in the information collection and data processing systems as possible. Finally, there has been a wide diffusion of citizen science systems. This allows an extremely widespread survey of the data in the territory. The quality of this tool in generating reliable data is clearly closely related to the ability to adequately involve citizens in the acquisition and transmission of information. The spread of digital tools (especially smartphones) is considerably facilitating the acquisition of domesticity by individuals with computerized data collection systems and allows them to increase their engagement as there is the perception of greater direct involvement in projects and in the results achieved. (Ershadi et al., 2014; Jordan et al., 2015; Kennedy et al., 2014; Moura et al., 2018; Peters et al., 2014);

The increased ability to interface the growing amounts of data with our understanding of ecological systems is another important factor for the development of innovative network models. Information gained on ecosystem functioning through traditional systems has created critical mass of knowledge that allows researchers to address network relationships. To understand the functioning of systems with multiple relationships it is in fact necessary to have acquired sufficient information on the basic structures (Evans, 2012; Williams & Jackson, 2007).

Technological innovations led to an exponential increase in the computing capacity of computers, enabling the processing of billions of information per second. This growth was accompanied by a significant reduction in the cost of highly performing infrastructure for scientific computing. The combination of these factors has allowed a widespread diffusion of highly performing tools for the elaboration of data.

Finally, the last but precious factor for the development and dissemination of innovative tools for network analysis is the professionals training in the natural science studies. Many degree courses have introduced quantitative teachings, offering to the students the tools to understand and manage even complex models. The increasing introduction of quantitative models into the ecological scientific literature has given further impetus to the quantitative literacy of researchers in this field.

### 4.1.3. Classification of modelling approaches

Over the past two decades, quantitative models for network analysis have proliferated given the favourable conditions that have arisen in terms of the availability of technological and information infrastructure. The investigated system determines the characteristics of the model that must investigate it, the approaches of network analysis have diversified to be able to offer an adequate solution to the many problems existing in the evaluation of biodiversity.

An important distinction between different models for biodiversity assessment is the research question that informs the study. There are two broad categories of research questions: the need to obtain a description as clear and accurate as possible of the dynamics of the system, the optimization of predictive capacity. Among the models that have mostly focused attention on the explicit description of the relationships between the components of the system we find the models with latent variables. When the main focus is on the accuracy and precision of the estimates, models belonging to the machine learning family are certainly the favoured solution. This class of models is based on algorithms specifically designed to minimize the estimation error, thus allowing those who use them to have a high degree of confidence on the results obtained.

**Latent variable models** powerful multivariate analysis allowing to describe how ecological interactions change over space and time and how they are shaped by environmental conditions. This approach allows to investigate the system at different levels of spatial resolution(e.g. locally, regionally) and across different species aggregation levels (e.g. from species to trophic groups) to get a better understanding of network structure. Latent variable models, among which the most well-known and widespread exponents are certainly the models with structural equations (SEMs), allow to directly test the hypothesis of the researcher, i.e. models based on deductive-scientific stance, prior knowledge or related theory. In fact, it is possible to build the network structure (nodes and iterations) and check whether the data validate this structure or not (Shipley, 2016). Also this class of models is useful to test and clarify differences between correlation and causation relationships (Grace, 2006). The possibility of defining causal links in latent variable models, however, requires that researchers are able to adequately identify or motivate theoretical foundations undermine the causal relationship in the hypotheses (Shipley, 2016). A possible critical issue of this class of models is the variable selection phase and the model identification. It is quite difficult to find a completely correct model, where each latent variable has an assigned scale and at least two observable indicators.

According to estimation methods, in many latent variable models maximum likelihood estimation is set as default. This assumed that indicators are continuous, normally distributed and with few missing data. In addition, many models require a large sample size (typically greater than 100 observations, some authors estimate over 200)

**Machine learning** techniques are highly conserved over spatial scales and across different species. They are based on algorithms that well manage different types of data . Machine learning models guarantee high global efficiency of information transfer. They show high performance in clustering (associated with robustness to random error). Machine learning is useful to quantify spatial connectivity, to monitor the behaviour of hundreds, or even thousands of individuals concurrently for long time (from days to months) (Hussey et al., 2015; Jacoby & Freeman, 2016; Ropert-Coudert & Wilson, 2005; Rutz & Hays, 2009). Some family of models belonging to machine learning could allow to test, at least indirectly, ecological hypotheses. The algorithms underlying these models are very efficient in estimating species distributions, although they generally do not provide clues to the mechanisms that explain these distributions (Ferrier & Guisan, 2006; Park & Chon, 2007). Furthermore they are very performing in the case of non-linear relationships. Machine learning models usually don't require specific assumptions concerning the distributional characteristics of the independent variables (i.e., nonparametric). A possible obstacle to the spread of machine learning models is the need for users to own strong computational skills. The optimization phase of the network structure is developed in an iterative way and for complex systems it can be extremely time consuming. Furthermore, the parameterization of the network can be strongly influenced by the definition of the initial conditions of the system, thus requiring a long calibration phase(Cao et al., 2019; Görgens et al., 2015; Olden et al., 2008).

## 4.2.Latent variable models

Network approach can be applied to measure abstract concept, such as ecosystem resistance and resilience or provisioning of ecosystem service. While direct measurements for these abstract concepts (hence latent variables) may not exist, statistical methods can derive these values from other related variables. The latent variable is not included in the data, it is derived from the other variables and could indicate a model's cause or effect (Grace, 2006; Hoyle & Isherwood, 2013). We then use observable variables, usually two or more, to approximate the construct of interest. The transition from observed to latent variables is never a trivial process and requires special attention, considering the fact that observable indicators are only approximations of latent constructs. Characteristic of latent variable models is local independence, i.e. the manifest variables have nothing in common after controlling for the latent variable.

### 4.2.1. Structural equation model (SEM)

Structural equation modelling (SEM) is a family of mathematical models, computer algorithms, and statistical methods that fit networks of constructs to data. SEM combines two tools within a single framework: confirmatory factor analysis and path analysis. The CFA (confirmatory factor analysis) aims to build a model suitable for studying the relationships between the observed variables and the latent variables. Path analysis, on the other hand, aims to explain the random relationships between constructs. It allow to find the causal relationship among variables by creating a path diagram. SEM merges these two objectives in an unified quantitative approach, allowing to test for the nature and magnitude of direct and indirect effects of multiple interacting factors

SEM results are represented by i) a set of parameters that quantify the relationships specified in the model, ii) one or more indices that measure the goodness of the fit of the hypothesized model to the observed data, that is the correspondence (adequacy) of the model with respect to the data collected (e.g. chi-square, CFI, RMSEA), and iii) a variance / covariance matrix estimated through the model parameters.

Correlations among the traits of organisms or groups of organisms in order to evaluate complex causal relationships can be analysed by SEM approach. In ecological studies, SEM widespread, thanks to its flexibility (Fan et al., 2016). Furthermore, SEMs developed with reflexive approaches have a favourable property. If the structure of the measurement model is correctly defined, it is possible to modify one or more indicator variables without disturbing or invalidating the validity of the defined network. This can be useful in the case of application of a model in different agroecosystems.

A full SEM is composed by two sub-models: the structural model representing the relationships between latent variables, and the measurement model linking indicators with their latent variables.

SEM are useful tool for the biodiversity assessment of ecosystems. In particular the separation between the structural model and the measurement model has proved particularly effective in modelling the interactions between traits and other ecosystem components. Sonnier et al. (2010) clearly states how SEM can be applied to quantify interactions between 10 functional traits (e.g. plant lifespan, leaf dry matter content, allocation to structure ) and environmental variables (such as litter, mineralizable nitrogen and humidity).

SEM can be used to detect assembly rules along different spatial scales or to investigate the impact of ecological and climatic variables on population regulation (Figure 14). For instance, Legras et al. (2019) applied a SEM to assess the assembly rules structuring stream invertebrate communities before and after a severe drought in southeast Arizona. Regulatory role of abiotic factors and biological interactions on lake phytoplankton dynamics and water clarity during the summer stratification period has been investigated by SEM (Arhonditsis et al., 2006). Ecological structures can be assed also based on time-series data (Almaraz, 2005).

Furthermore SEM can be applied to assess resilience properties of ecosystem (Grace, 2006; Grace et al., 2010). By way of example, these models have been effectively used to estimate the influence of some environmental factors and the age of plants on the regeneration rate of a forest.

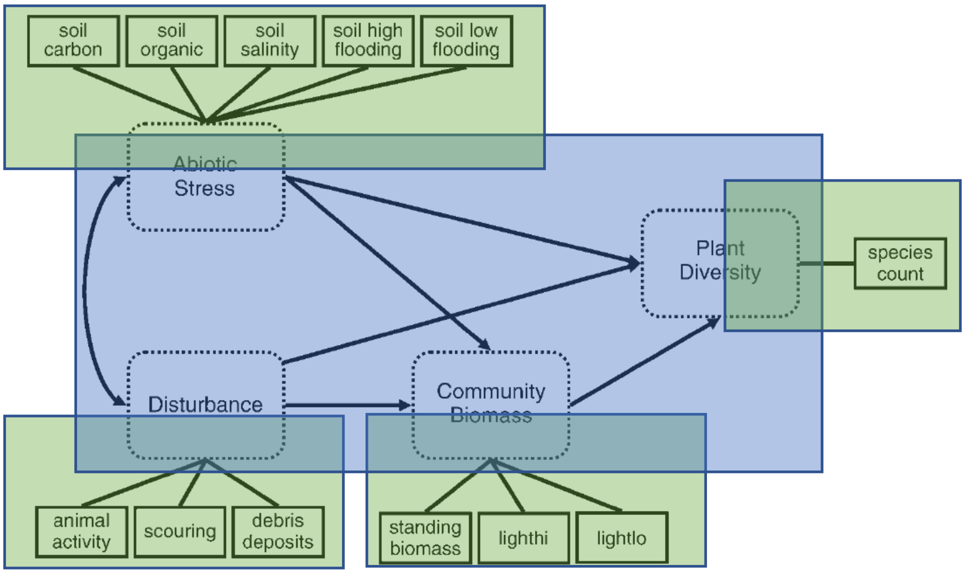


*Figure 14 – SEM investigating the biodiversity in coastal grassland, according to abiotic stress, disturbance, community biomass and plant diversity (Grace et al., 2010). Blue shadow box is the structural model, green shadow boxes are the measurement models(one for each latent variable)*

### 4.2.2. Partial Least Squares Path Modelling (PLS-PM)

Partial Least Squares Path Modelling (PLS-PM) approach allows the investigation of complex cause-effect relationships existing between attributes, measured by sets of indicators (Tenenhaus et al., 2005; Vinzi et al., 2010). The attributes are estimated as a linear combination (a weighted sum) of their indicators. Each composite indicator is obtained in such a way as to be the most representative of its underlying indicators and the most correlated with the other composite indicators to which it is linked. The PLS-PM is an iterative algorithm that allows to estimate the relationships existing between the different blocks of variables through a system of linear equations. PLS-PM is a component-based method, i.e. it is based on the search for latent components.

PLS-SEM is the preferred method when the theoretical basis of the research is not yet well developed, particularly when there is little or no prior knowledge on the causal relationship under investigation. It is a particularly useful model for the exploration phase rather than the confirmation phase (in which the SEMs are more efficient). For this reason, it is recommended that users in the initial phase of research apply PLS-PM to generate the necessary evidence for causal relationship and variable selections. This will allow users to continue collecting long-term data by updating their assumptions (Monecke & Leisch, 2012).

PLS-SEM, which is based on a different algorithm from SEM, requires neither a large sample size nor a specific hypothesis on the distribution of the data, or even missing data. Unlike SEM, PLS-PM is never under-identified (it follows the basic rules of a regression model) and does not require any hypothesis on the shape of the distribution of the manifest variables, nor on that of the latent variables. The iterative algorithm directly provides an estimate of the latent variables (latent variable scores). The convergence of the PLS-PM is empirically ensured but there is no mathematical proof (except in the case of only two blocks of manifest variables). Furthermore, in the PLS-PM it is not possible to identify a single function to be optimized; however, specific cases of the PLS-PM optimize specific criteria (Hair et al., 2012; Jr. Hair et al., 2009, 2016).

The flexibility in the definition of causal links in PLS-PM allows to test which relationships (and consequently which drivers) are significant with respect to the target variable. For this reason it is a tool suitable to investigate the causes of biodiversity loss in ecosystems. Sanches Fernandes et al. (2018) applies PLS-PM to understand the main drivers of biodiversity loss in rural and urban watersheds in Portugal .

PLS-PM can be used to explore the relationships between management practices and species traits. Serrano et al. (2014) performed a PLS-PM analysis to explore whether farm type, swine management and pathogens directly or indirectly influenced post-weaning mortality (Figure 15). In the context of farmer strategies, the impact of ecological intensification and conventional management practices on maize yield, N losses, N2O emissions, greenhouse gas (GHG) emissions, and nitrogen (N) cycling microbial populations associated with nitrification and denitrification in fluvo-aquic soil and black soil have been investigated by a PLS-PM approach (Ullah et al., 2020). Honvault et al. (2020) investigated the role of multiple belowground morphological and physiological traits on plant P acquisition strategies (Figure 16).

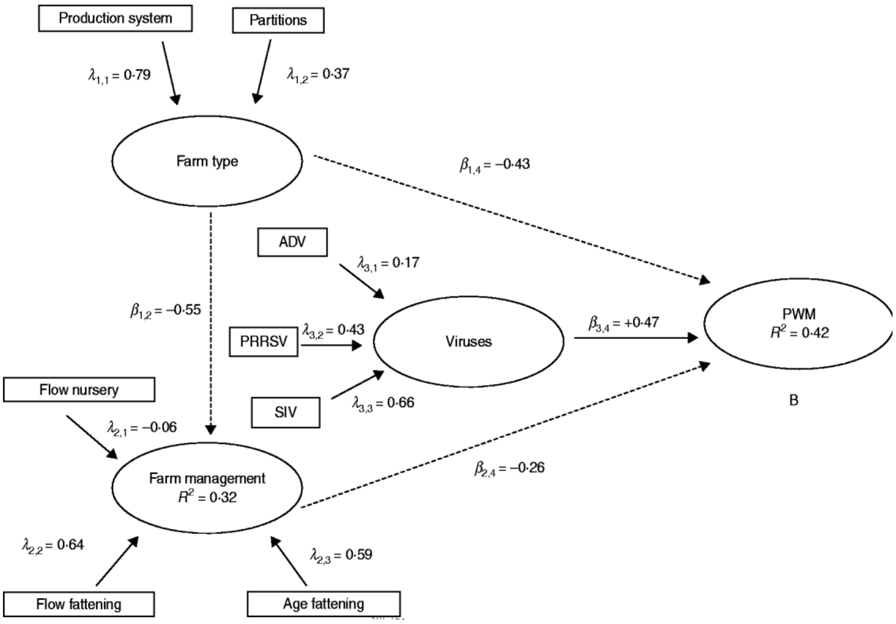


*Figure 15 – PLS-PM describing causes of post-weaning mortality. Ellipses represent the latent variables, the rectangles the indicators (measurable variables) used to estimate the latent variables. The arrows between the ellipses are the interactions studied in the model (Serrano et al., 2014).*

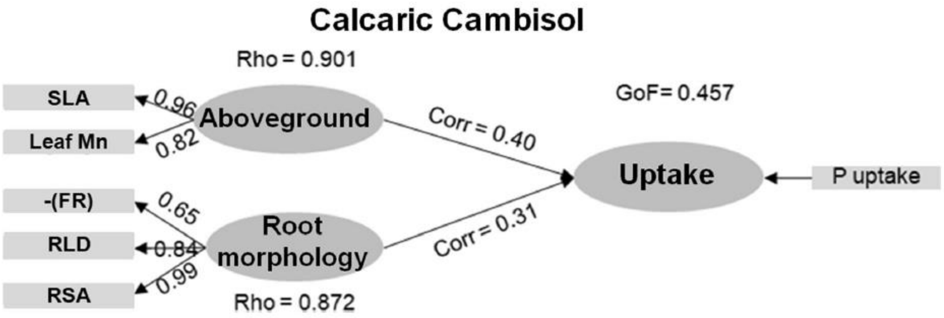


*Figure 16 – PLS-PM for the predicting phosphorus uptake, based on root morphology and aboveground traits (Honvault et al., 2020).*

## 4.3. Machine learning

Basically, Machine Learning (ML) algorithms use mathematical-computational methods to learn information directly from data, without mathematical models and predetermined equations. Machine Learning algorithms improve their performance in an adaptive way as the "examples" to learn from increase. In this context, the concept of learning (experience) consists in improving the performance of the software after carrying out a task or completing an action (even incorrect, starting from the assumption that the principle "failing to learn") (Samuel, 1959). ML is a family of computational algorithms aimed at the identification of structure in complex, often nonlinear, data. ML models emphasize the accurate prediction based on such data (Breiman, 2001; Fielding, 1999; Goodfellow et al., 2016; Olden et al., 2008).

Four types of learning methodologies can be used:

- **Supervised learning** is based on pairs of data containing the original data and the expected result. The task of the algorithm is to find the rule (function or model) with which to create a relationship between the two data so that, when a previously unknown example occurs, it can obtain the correct result. The Supervised learning is mainly used for classification problems. and range from testing biogeographical, ecological, and evolutionary hypotheses to modelling species distributions for conservation and management planning.
- **Unsupervised learning** does not use previously classified and labelled data, so it is not known a priori to which categories the data belong. The algorithm (machine) is then asked to extract a rule that groups the cases according to characteristics it can derive only from the data itself. One of the main applications is clustering, that is the grouping of data into homogeneous groups called clusters. For this reason, this type of model could be useful to define, for example, cluster traits.
- **Reinforcement learning** is a machine learning technique that aims to create systems capable of learning and adapting to changes in the environment in which they are immersed through the distribution of a ‘reward’ called reinforcement, given by performance evaluation. Its application in the ecological field is not currently widespread

- **Semi-supervised learning** is a hybrid model where the computer is provided with an incomplete set of data for training / learning; some of these inputs are equipped with the respective output examples (as in supervised learning), while others lack them (as in unsupervised learning). The basic objective is always the same: to identify rules and functions for solving problems, as well as models and data structures useful for achieving certain objectives.

### 4.3.1. Artificial Neural network (ANN)

Artificial neural networks (ANN) are mathematical models inspired by biological neural networks (hence the name) and are used to solve engineering problems of Artificial Intelligence. ANN are models made up of interconnections of information deriving from artificial neurons and computational processes based on the cognitive science model called "connectionism".

A neural network actually looks like an adaptive system capable of modifying its structure (nodes and interconnections) based on both external data and internal information that connect and pass through the neural network during the learning phase and reasoning.

ANN are non-linear structures of statistical data organized as modelling tools: they receive external signals on a layer of nodes (which represents the processing unit, the processor); each of these "input nodes" is connected to various internal nodes of the network which, typically, are organized at several levels so that each single node can process the signals received by transmitting the result of its processing to subsequent levels (therefore information more advanced, detailed).

In general, neural networks are made up of three layers, which, however, can involve thousands of neurons and tens of thousands of connections (Figure 17):

1) the input layer has the task of receiving and processing the input signals, adapting them to the demands of the neurons of the network;

2) the hidden layer is in charge of the actual processing process (and can also be structured with multiple columns-levels of neurons);

3) the output layer collects the results of the processing of the hidden layer and adapted them to the requests of the next level-block of the neural network.

In order for this process to be efficient, it is necessary to "train" the neural networks, that is to make them learn how to behave when an engineering problem has to be solved, such as the recognition of a human being from the analysis of images (through for example facial recognition technology).

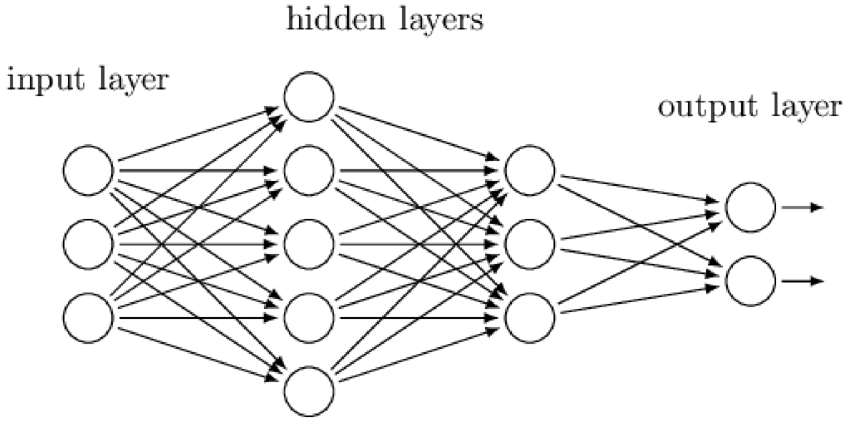


*Figure 17 – Basic scheme of an artificial neural network*

ANNs provide a much more flexible way of modelling ecological data with respect to latent variable models. To improve data fit, the complexity of the model can be varied modifying the internal architecture of the network by changing the transfer function, the number of neurons, hidden layers or output neurons. It is this flexibility that has probably led to the increased popularity of neural networks in ecology.The major drawback of neural networks is the fact that their computation process cannot be fully analyzed. By this we mean that they are able to supply correct or sufficiently correct outputs, but they do not allow to examine the single processing stages that determine them (black box). This allows for very accurate predictions, but not being able to describe by which process they are generated.

The high flexibility of ANN allows the modelling of multiple ecological response variables, as in the case of multiple species analyses (Özesmi & Özesmi, 1999) or of entire communities (Olden et al., 2006). Görgen et al. (2015) evaluate the performance of three machine learning tools for predicting stand volume of fast-growing forest plantations, based on statistical vegetation metrics extracted from an Airborne Laser Scanning (ALS) survey. Larsen et al. (2012) applied neural networks to predict microbial community structure as a function of environmental parameters and microbial interactions.

ANN have been effectively used to model species distributions because it can the model abundance in the environmental feature space that present very irregular shape. Pattern of species distribution can be difficult to capture with classical models in these cases (Botella et al., 2018; Li, 2020; Rew et al., 2020). Heikkinen et al. (2007) ANN to predict the distribution of the threatened clouded apollo butterfly (*Parnassius mnemosyne*) in south-west Finland.

ANN have also been used for taxonomic classifications (Figure 18). The high precision and accuracy that characterizes the algorithm in identifying groups makes it an ideal tool for this purpose (Khawaldeh et al., 2017).

In the field of research on the impact of climate change on biodiversity and the provision of ecosystem services, ANN have been widely used as they allow for the processing of big data in a relatively short time. Furthermore, since it is not necessary to define causal links, they can find patterns of behaviour that would hardly be identified by researchers (Abdullahi & Elkiran, 2017; Juan et al., 2017, pag.; Z. Liu et al., 2010). This aspect, which is certainly an advantage of the ANNs, also represents an important limitation. In fact, since there is no possibility of intervening on the causal links identified by the network, the interpretation of the results is very complex, the more complex the data matrix is. Furthermore, the network could identify relationships that are meaningless from a biological / ecological point of view.

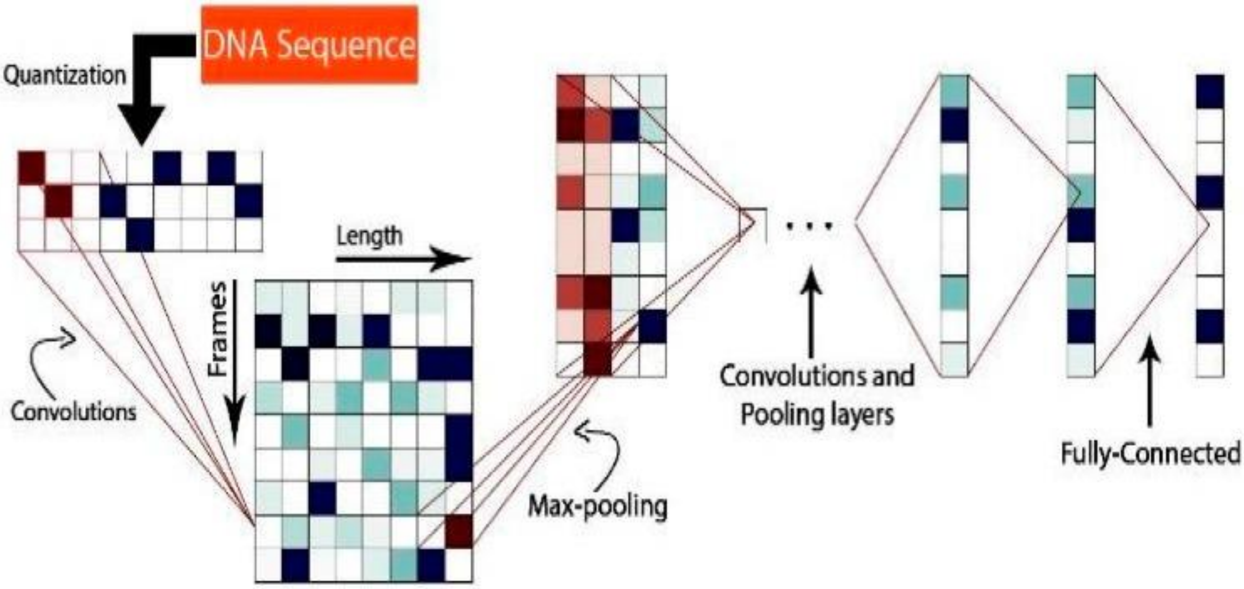


*Figure 18 - Convoluted neural network model used to analyse DNA sequences to classify living organisms (Khawaldeh et al., 2017)*

### 4.3.2. Support Vector machine (SVM)

Support Vector machines are of recent introduction in the field of ecology, in particular for the recognition of patterns in complex data structures and in classification problems in remote sensing. Formally, a support vector machine constructs a hyperplane or set of hyperplanes in a multi-dimensional or infinite-dimensional space, which can be used for classification, regression, and other purposes such as anomaly detection. Intuitively, a good separation can be obtained when the hyperplane has a great distance from the closest point (of the training set) of each of the classes; in general, the greater the margin between these points, the smaller the generalization error made by the classifier.

A SVM model is a representation of data as points in space, mapped in such a way that the data belonging to the two different categories are clearly separated by as large a space as possible (Figure 19). The new examples are then mapped in the same space and the prediction of the category to which they belong is made on the basis of the side in which it falls. In addition to linear classification, it is possible to make use of SVMs to effectively perform non-linear classification using the kernel method, implicitly mapping their inputs into a multi-dimensional characteristics space.

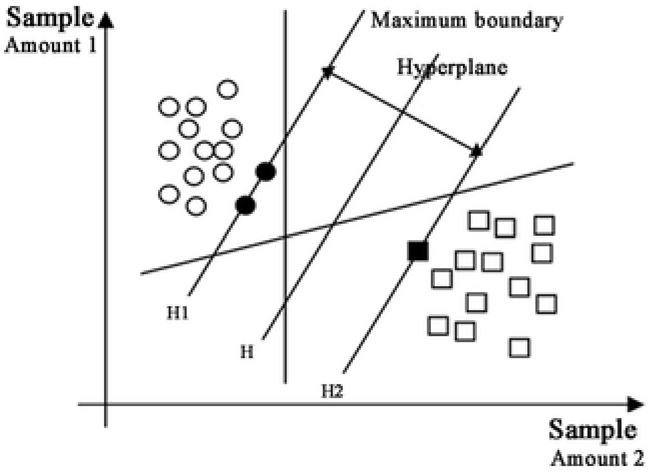


*Figure 19 – Graphical display of hyperplane defined applying support vector machine algorithm (Cao et al., 2019)*

SVMs has been widely used in species identification (Londhe & Kanade, 2015), species mapping (Abeysinghe et al., 2019; Shiferaw et al., 2019) and disease distribution (Rumpf et al., 2010). An important feature of this class of models is the ability to manage high-dimensional data with a corresponding lack of knowledge of the underlying distribution (as well as possibly a relatively small sample size). In fact, SVMs have no distributive assumptions, other than the fact that data is independent and identically distributed.

In the context of functional biodiversity assessment, SVMs could be useful quantitative tool for identifying clusters of traits. Being able to process large amounts of data, they allow to identify affinities between species in terms of sharing traits or ecological niches. They are black box models, so they do not explain the processes through which results are obtained. The clusters obtained must therefore be characterized in a second step based on their common characteristics (Cervantes et al., 2020; Drake et al., 2006; Hu et al., 2012; Priya et al., 2012).

### 4.3.3. Bayesian network (BN)

A Bayesian network (BN) is a probabilistic graph model that represents a set of stochastic variables with their conditional dependencies through the use of a direct acyclic graph (Pearl, 1988). In statistics, the Bayesian network is used to more easily identify the absolute and conditional dependence relationships between variables, in order to reduce the number of combinations of the variables to be analysed. From the graphic point of view, the Bayesian network is an acyclic oriented graph (without cycles) in which each node indicates a certain system variable while the oriented edges represent the possible dependence relationships (parent → child) between the variables. The directed arc (arrow) is equivalent to saying that the parent node directly affects the child node (dependent variable). Nodes that are not connected represent variables that are conditionally independent of each other. Each node is associated with a probability function which takes as input a particular set of values for the variables of the parent node and returns the probability of the variable represented by the node.

The acyclic oriented graph is also known by the acronym DAG (Direct Acyclic Graph). The set of arcs and nodes is called the network topology. An example of a Bayesian network is the reported in Figure 20. Each node has a conditional probability distribution that lists the effects on the node in question, based on the various states that its parent nodes can assume.

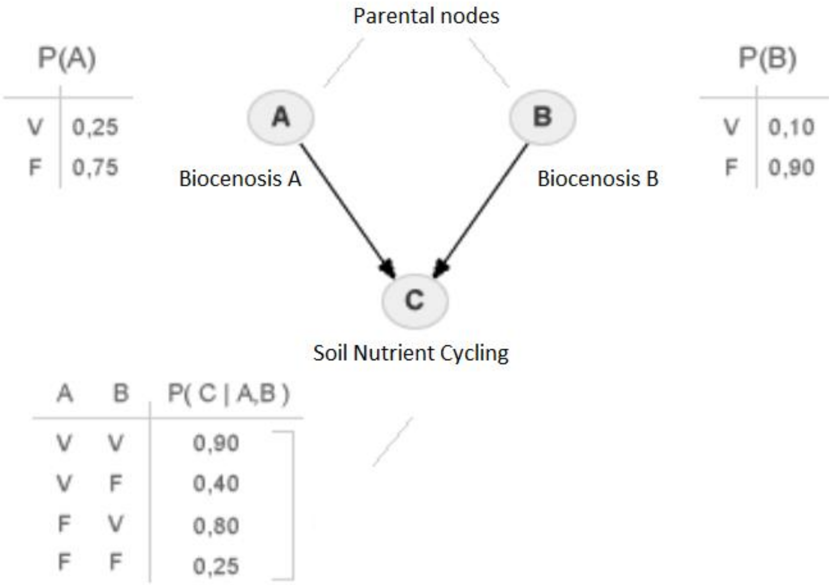


*Figure 20 – Example of a Bayesian network, of marginal and conditional probability distributions*

Bayesian networks allow inference on unobserved variables. BN can be used to update knowledge of the state of a subset of variables when other variables (the test variables) are observed. Furthermore, using machine learning techniques, it is possible to automatic learn the structure of the graph and the parameters of the local distributions of a BN starting from the data.

Bayesian networks have several advantages over traditional methods (such as topological approaches and dynamic models): species can have different probability of establishment or development 'a priori', disturbances can affect several species simultaneously, and the risk of extinction of predators increases with the loss of resources.

Bayesian networks can be used to model complex networks of ecological interactions. BN could represent the probabilistic relationship existing between ecosystem traits and services. Given the traits, the network can be used to calculate the level of delivery of one or more services. Liber et al. (2020) developed a BN approach to investigate plant traits, in particular they studied the importance of biological and agri-environmental parameters in the ability of plants to bioaccumulate chlordecone (Figure 21).

Models based on Bayesian statistics combine probabilities of observing species with their probabilities of occurrence conditional to the value (or class of values) of each environmental predictor (Montesinos-Navarro et al., 2018; Sander et al., 2017). Furthermore, BN copes well with sparse matrices which are a frequent problem when it comes to modelling aspects related to rare species (Maldonado et al., 2016; Thompson et al., 2019). Cirtwill et al. (2019) investigated interspecific interactions as well as the uncertainty around each interaction by BN, combining data on observed co-occurrences with prior knowledge. Influence on fauna distributions in fire-affected landscape have been studied through a BN approach (Figure 22) by Hradsky et al. (2017).

BN are widespread in the genomic field. Santos et al. (Santos et al., 2020) used BN to phenotype a diversity panel of 869 biomass sorghum (*Sorghum bicolor* (L.) Moench) lines.

In vegetation mapping, BN allow to define a posteriori probabilities for each vegetation unit and the unit with the highest probability is predicted at every candidate site (Dlamini, 2011).

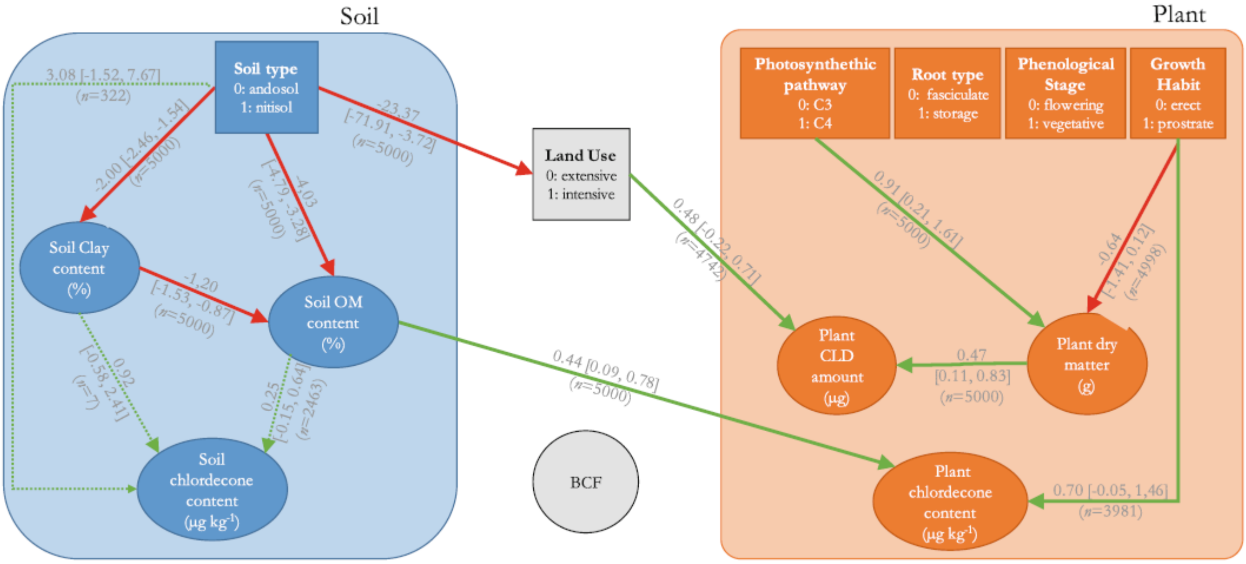


*Figure 21 - Bayesian network evaluating factors linked with the accumulation of chlordecone in the plant in Guadeloupe (Liber et al., 2020)*

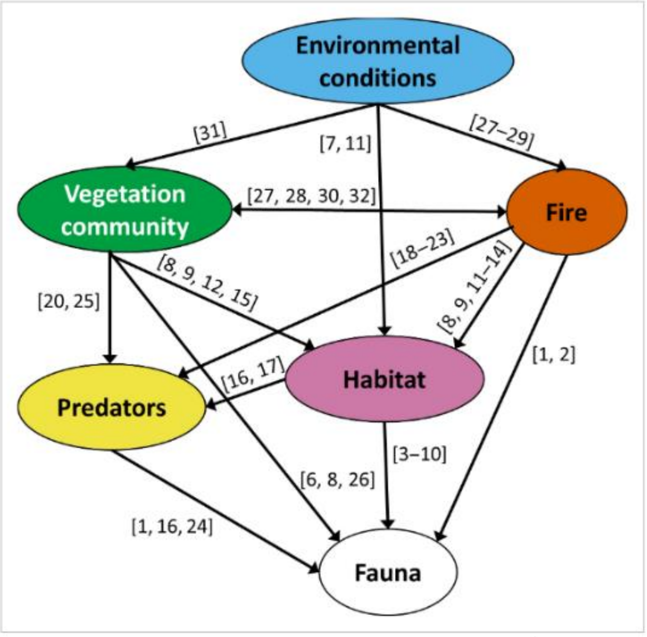


*Figure 22 - Factors that influence fauna distributions in fire-affected landscapes (Hradsky et al., 2017)*

## 4.4. Conclusion

The application of network approaches to biodiversity assessment is to the dawn, but is proving extremely promising. The renewed interest on the part of the scientific world and international bodies and organizations for biodiversity are the best omen to give new life to research in the direction of a more precise and scientifically based approach to sustainability and to the biological and ecological mechanisms that guarantee the genesis and regulation of ecosystem services it.

# 5. References

Abdullahi, J., & Elkiran, G. (2017). Prediction of the future impact of climate change on reference evapotranspiration in Cyprus using artificial neural network. *Procedia Computer Science*, *120*, 276–283. https://doi.org/10.1016/j.procs.2017.11.239

Abeysinghe, T., Simic Milas, A., Arend, K., Hohman, B., Reil, P., Gregory, A., & Vázquez-Ortega, A. (2019). Mapping Invasive Phragmites australis in the Old Woman Creek Estuary Using UAV Remote Sensing and Machine Learning Classifiers. *Remote Sensing*, *11*(11), 1380. https://doi.org/10.3390/rs11111380

Aguirre-Gutiérrez, J., Kissling, W. D., Carvalheiro, L. G., WallisDeVries, M. F., Franzén, M., & Biesmeijer, J. C. (2016). Functional traits help to explain half-century long shifts in pollinator distributions. *Scientific Reports*, *6*(1), 24451. https://doi.org/10.1038/srep24451

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Allsopp, P. G., Klein, M. G., & McCoy, E. L. (1992). Effect of Soil Moisture and Soil Texture on Oviposition by Japanese Beetle and Rose Chafer (Coleoptera: Scarabaeidae). *Journal of Economic Entomology*, *85*(6), 2194–2200. https://doi.org/10.1093/jee/85.6.2194

Almaraz, P. (2005). Ecological time-series analysis through structural modelling with latent constructs: Concepts, methods and applications. *Comptes Rendus Biologies*, *328*(4), 301–316. https://doi.org/10.1016/j.crvi.2004.12.005

Altermatt, F., Little, C. J., Mächler, E., Wang, S., Zhang, X., & Blackman, R. C. (2020). Uncovering the complete biodiversity structure in spatial networks: The example of riverine systems. *Oikos*, *129*(5), 607–618. https://doi.org/10.1111/oik.06806

André, H. M., Noti, M.-I., & Lebrun, P. (1994). The soil fauna: The other last biotic frontier. *Biodiversity and Conservation*, *3*(1), 45–56. https://doi.org/10.1007/BF00115332

Andrés, P., Moore, J. C., Simpson, R. T., Selby, G., Cotrufo, F., Denef, K., Haddix, M. L., Shaw, E. A., de Tomasel, C. M., Molowny-Horas, R., & Wall, D. H. (2016). Soil food web stability in response to grazing in a semi-arid prairie: The importance of soil textural heterogeneity. *Soil Biology and Biochemistry*, *97*, 131–143. https://doi.org/10.1016/j.soilbio.2016.02.014

Antúnez, P., Suárez-Mota, M., Valenzuela-Encinas, C., & Ruiz-Aquino, F. (2018). The Potential Distribution of Tree Species in Three Periods of Time under a Climate Change Scenario. *Forests*, *9*(10), 628. https://doi.org/10.3390/f9100628

Arhonditsis, G. B., Stow, C. A., Steinberg, L. J., Kenney, M. A., Lathrop, R. C., McBride, S. J., & Reckhow, K. H. (2006). Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecological Modelling*, *192*(3), 385–409. https://doi.org/10.1016/j.ecolmodel.2005.07.028

Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model. *Biometrika*, *67*(2), 413–418. https://doi.org/10.1093/biomet/67.2.413

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, *200*(1), 1–19. https://doi.org/10.1016/j.ecolmodel.2006.07.005

Austin, M. P. (2013). Vegetation and Environment: Discontinuities and Continuities. In *Vegetation Ecology*. van der Maarel E., Franklin J. (Eds), Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/9781118452592.ch3

Austin, M. P., & Smith, T. M. (1989). A new model for the continuum concept. *Vegetatio, 83*, 35–47.

Badejo, M. A. (1990). Seasonal Abundance of Soil Mites (Acarina) in Two Contrasting Environments. *Biotropica*, *22*(4), 382–390. JSTOR. https://doi.org/10.2307/2388555

Bagyaraj, D. J., Nethravathi, C. J., & Nitin, K. S. (2016). Soil Biodiversity and Arthropods: Role in Soil Fertility. *Economic and Ecological Significance of Arthropods in Diversified Ecosystems*, 17–51. https://doi.org/10.1007/978-981-10-1524-3_2

Bàrberi, P., Burgio, G., Dinelli, G., Moonen, A. C., Otto, S., Vazzana, C., & Zanin, G. (2010). Functional biodiversity in the agricultural landscape: Relationships between weeds and arthropod fauna. *Weed Research*, *50*(5), 388–401. https://doi.org/10.1111/j.1365-3180.2010.00798.x

Bede-Fazekas, Á., Horváth, L., & Kocsis, M. (2014). Impact of climate change on the potential distribution of Mediterranean pines. *Quarterly Journal of the Hungarian Meteorological Service*, *118*(1), 41–52.

Bertrand, M., & Lumaret, J. P. (1992). The role of Diplopoda litter grazing activity on recycling processes in a Mediterranean climate. *Vegetatio*, *99*(1), 289–297. https://doi.org/10.1007/BF00118236

Blasi, S., Menta, C., Balducci, L., Conti, F. D., Petrini, E., & Piovesan, G. (2013). Soil microarthropod communities from Mediterranean forest ecosystems in Central Italy under different disturbances. *Environmental Monitoring and Assessment*, *185*(2), 1637–1655. https://doi.org/10.1007/s10661-012-2657-2

Botella, C., Joly, A., Bonnet, P., Monestiez, P., & Munoz, F. (2018). A deep learning approach to Species Distribution Modelling. In A. Joly, S. Vrochidis, K. Karatzas, A. Karppinen, & P. Bonnet (A c. Di), *Multimedia Tools and Applications for Environmental & Biodiversity Informatics* (pagg. 169–199). Springer. https://doi.org/10.1007/978-3-319-76445-0_10

Boyd, J., & Banzhaf, S. (2007). What are ecosystem services? The need for standardized environmental accounting units. *Ecological Economics*, *63*(2), 616–626. https://doi.org/10.1016/j.ecolecon.2007.01.002

Brambilla, M., Casale, F., Bergero, V., Bogliani, G., Crovetto, G. M., Falco, R., Roati, M., & Negri, I. (2010). Glorious past, uncertain present, bad future? Assessing effects of land-use changes on habitat suitability for a threatened farmland bird species. *Biological Conservation*, *143*(11), 2770–2778. https://doi.org/10.1016/j.biocon.2010.07.025

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Büchs, W., Harenberg, A., Zimmermann, J., & Weiß, B. (2003). Biodiversity, the ultimate agri-environmental indicator?: Potential and limits for the application of faunistic elements as gradual indicators in agroecosystems. *Agriculture, Ecosystems & Environment*, *98*(1), 99–123. https://doi.org/10.1016/S0167-8809(03)00073-2

Bukvareva, E. (2018). The optimal biodiversity–A new dimension of landscape assessment. *Ecological Indicators*, *94*, 6–11. https://doi.org/10.1016/j.ecolind.2017.04.041

Byers, G. W., & Thornhill, R. (1983). Biology of the Mecoptera. *Annual Review of Entomology*, *28*(1), 203–228. https://doi.org/10.1146/annurev.en.28.010183.001223

Caffarra, A., Rinaldi, M., Eccel, E., Rossi, V., & Pertot, I. (2012). Modelling the impact of climate change on the interaction between grapevine and its pests and pathogens: European grapevine moth and powdery mildew. *Agriculture, Ecosystems & Environment*, *148*, 89–101. https://doi.org/10.1016/j.agee.2011.11.017

Calonnec, A., Cartolaro, P., Naulin, J.-M., Bailey, D., & Langlais, M. (2008). A host-pathogen simulation model: Powdery mildew of grapevine. *Plant Pathology*, *57*(3), 493–508. https://doi.org/10.1111/j.1365-3059.2007.01783.x

Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., Guerra, C. A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Delgado-Baquerizo, M., Ferlian, O., Fierer, N., Kreft, H., Lovejoy, T. E., Montanarella, L., Orgiazzi, A., … Eisenhauer, N. (2018). Global gaps in soil biodiversity data. *Nature ecology & evolution*, *2*(7), 1042–1043. https://doi.org/10.1038/s41559-018-0573-8

Cancela, H., Higgins, A., Pagès-Bernaus, A., & Plà-Aragonès, L. M. (2019). Prologue – BigData and DSS in agriculture. *Computers and Electronics in Agriculture*, *161*, 1–3. https://doi.org/10.1016/j.compag.2019.05.004

Cao, J., Wang, M., Li, Y., & Zhang, Q. (2019). Improved support vector machine classification algorithm based on adaptive feature weight updating in the Hadoop cluster environment. *PLOS ONE*, *14*(4), e0215136. https://doi.org/10.1371/journal.pone.0215136

Caporali, F. (2009). *Indicatori di biodiversità per la sostenibilità in agricoltura: Linee guida, strumenti e metodi per la valutazione della qualità degli agroecosistemi*. ISPRA.

Caprio, E., Nervo, B., Isaia, M., Allegro, G., & Rolando, A. (2015). Organic versus conventional systems in viticulture: Comparative effects on spiders and carabids in vineyards and adjacent forests. *Agricultural Systems*, *136*, 61–69. https://doi.org/10.1016/j.agsy.2015.02.009

Cavender-Bares, J., Gamon, J., & Townsend, P. (A c. Di). (2020). *Remote Sensing of Plant Biodiversity*. Springer International Publishing. https://doi.org/10.1007/978-3-030-33157-3

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Choi, W. I., Moorhead, D. L., Neher, D. A., & Ryoo, M. I. (2006). A modeling study of soil temperature and moisture effects on population dynamics of Paronychiurus kimi

(Collembola: Onychiuridae). *Biology and Fertility of Soils*, *43*(1), 69–75. https://doi.org/10.1007/s00374-005-0062-5

Choi, W. I., Ryoo, M. I., & Kim, J.-G. (2002). Biology of Paronychiurus kimi(Collembola: Onychiuridae) under the influence of temperature, humidity and nutrition. *Pedobiologia*, *46*(6), 548–557. https://doi.org/10.1078/0031-4056-00159

Cirtwill, A. R., Eklöf, A., Roslin, T., Wootton, K., & Gravel, D. (2019). A quantitative framework for investigating the reliability of empirical network construction. *Methods in Ecology and Evolution*, *10*(6), 902–911. https://doi.org/10.1111/2041-210X.13180

Clapperton, M. J., Kanashiro, D. A., & Behan-Pelletier, V. M. (2002). Changes in abundance and diversity of microarthropods associated with Fescue Prairie grazing regimes. *Pedobiologia*, *46*(5), 496–511. https://doi.org/10.1078/0031-4056-00155

Clark, M. S., Thorne, M. A., Purać, J., Grubor-Lajšić, G., Kube, M., Reinhardt, R., & Worland, M. R. (2007). Surviving extreme polar winters by desiccation: Clues from Arctic springtail (Onychiurus arcticus) EST libraries. *BMC Genomics*, *8*(1), 475. https://doi.org/10.1186/1471-2164-8-475

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, *18*(1), 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x

Commerce, N. C. for E. P. W. S. S. D. of. (2015). *NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids* [Data set]. UCAR/NCAR - Research Data Archive. https://doi.org/10.5065/D65Q4T4Z

Costantini, E. a. C., Agnelli, A. E., Fabiani, A., Gagnarli, E., Mocali, S., Priori, S., Simoni, S., & Valboa, G. (2015). Short-term recovery of soil physical, chemical, micro- and mesobiological functions in a new vineyard under organic farming. *SOIL*, *1*(1), 443–457. https://doi.org/10.5194/soil-1-443-2015

Culliney, T. (2013). Role of arthropods in maintaining soil fertility. In *Agriculture* (Vol. 3).

Dal Cortivo, C., Ferrari, M., Visioli, G., Lauro, M., Fornasier, F., Barion, G., Panozzo, A., & Vamerali, T. (2020). Effects of Seed-Applied Biofertilizers on Rhizosphere Biodiversity and Growth of Common Wheat (Triticum aestivum L.) in the Field. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00072

Dalthorp, D., Nyrop, J., & Villani, M. G. (2000). Spatial ecology of the Japanese beetle, Popillia japonica. *Entomologia Experimentalis et Applicata*, *96*(2), 129–139. https://doi.org/10.1046/j.1570-7458.2000.00688.x

D’Andrea, R., Riolo, M., & Ostling, A. M. (2019). Generalizing clusters of similar species as a signature of coexistence under competition. *PLoS Computational Biology*, *15*(1). https://doi.org/10.1371/journal.pcbi.1006688

Decaëns, T., Jiménez, J. J., Gioia, C., Measey, G. J., & Lavelle, P. (2006). The values of soil animals for conservation biology. *European Journal of Soil Biology*, *42*, S23–S38. https://doi.org/10.1016/j.ejsobi.2006.07.001

Degen, R., Aune, M., Bluhm, B. A., Cassidy, C., Kędra, M., Kraan, C., Vandepitte, L., Włodarska-Kowalczuk, M., Zhulay, I., Albano, P. G., Bremner, J., Grebmeier, J. M., Link, H., Morata, N., Nordström, M. C., Shojaei, M. G., Sutton, L., & Zuschin, M. (2018). Trait-based approaches in rapidly changing ecosystems: A roadmap to the future polar oceans. *Ecological Indicators*, *91*, 722–736. https://doi.org/10.1016/j.ecolind.2018.04.050

Dehling, D. M., Fritz, S. A., Töpfer, T., Päckert, M., Estler, P., Böhning-Gaese, K., & Schleuning, M. (2014). Functional and phylogenetic diversity and assemblage structure of frugivorous birds along an elevational gradient in the tropical Andes. *Ecography*, *37*(11), 1047–1055. https://doi.org/10.1111/ecog.00623

Dettmers, R., & Bart, J. (1999). A Gis Modeling Method Applied to Predicting Forest Songbird Habitat. *Ecological Applications*, *9*(1), 152–163. https://doi.org/10.1890/1051-0761(1999)009[0152:AGMMAT]2.0.CO;2

Di Biase, R. M., Fattorini, L., & Marchi, M. (2018). Statistical inferential techniques for approaching forest mapping. A review of methods. *Annals of Silvicultural Research*, *42*(2), 46–58.

Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., Keitt, T. H., Kenney, M. A., Laney, C. M., Larsen, L. G., Loescher, H. W., Lunch, C. K., Pijanowski, B. C., Randerson, J. T., Read, E. K., Tredennick, A. T., Vargas, R., Weathers, K. C., & White, E. P. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*, *115*(7), 1424–1432. https://doi.org/10.1073/pnas.1710231115

Dlamini, W. M. (2011). A data mining approach to predictive vegetation mapping using probabilistic graphical models. *Ecological Informatics*, *6*(2), 111–124. https://doi.org/10.1016/j.ecoinf.2010.12.005

Doran, J. W., & Zeiss, M. R. (2000). Soil health and sustainability: Managing the biotic component of soil quality. *Applied Soil Ecology*, *15*(1), 3–11. https://doi.org/10.1016/S0929-1393(00)00067-6

Döring, J., Collins, C., Frisch, M., & Kauer, R. (2019). Organic and Biodynamic Viticulture Affect Biodiversity and Vine and Wine Properties: A Systematic Quantitative Review. *American Journal of Enology and Viticulture*. https://doi.org/10.5344/ajev.2019.18047

Dormann, C. F., Fründ, J., & Schaefer, H. M. (2017). Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 559–584. https://doi.org/10.1146/annurev-ecolsys-110316-022928

Drake, J. M., Randin, C., & Guisan, A. (2006). Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, *43*(3), 424–432. https://doi.org/10.1111/j.1365-2664.2006.01141.x

Duncan, C., Thompson, J. R., & Pettorelli, N. (2015). The quest for a mechanistic understanding of biodiversity–ecosystem services relationships. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1817). https://doi.org/10.1098/rspb.2015.1348

Duque-Lazo, J., van Gils, H., Groen, T. A., & Navarro-Cerrillo, R. M. (2016). Transferability of species distribution models: The case of Phytophthora cinnamomi in Southwest Spain and Southwest Australia. *Ecological Modelling*, *320*, 62–70. https://doi.org/10.1016/j.ecolmodel.2015.09.019

Edenius, L., & Mikusiński, G. (2006). Utility of habitat suitability models as biodiversity assessment tools in forest management. *Scandinavian Journal of Forest Research*, *21*(S7), 62–72. https://doi.org/10.1080/14004080500486989

Eisenbeis, G., & Wichard, W. (1987). *Atlas on the Biology of Soil Arthropods*. Springer-Verlag. https://doi.org/10.1007/978-3-642-72634-7

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(D22). https://doi.org/10.1029/2002JD003296

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., Sassi, C. de, Galetti, M., Guimarães, P. R., Lomáscolo, S. B., González, A. M. M., Pizo, M. A., Rader, R., Rodrigo, A., Tylianakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks. *Ecology Letters*, *16*(5), 577–583. https://doi.org/10.1111/ele.12081

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., & Wood, E. F. (2014). Multi-site evaluation of terrestrial evaporation models using FLUXNET data. *Agricultural and Forest Meteorology*, *187*, 46–61. https://doi.org/10.1016/j.agrformet.2013.11.008

Eslami, A., & Fellenius, B. (2004). CPT and CPTu data for soil profile interpretation: Review of methods and a proposed new approach. *Iranian Journal of Science & Technology, Transaction B*, *28*.

Eslami, A., Moshfeghi, S., MolaAbasi, H., & Eslami, M. M. (2020). *Piezocone and Cone Penetration Test (CPTu and CPT) Applications in Foundation Engineering*. Elsevier. https://doi.org/10.1016/C2017-0-04195-2

European Commission. (2020). *EU biodiversity strategy for 2030. Bringing nature back into our lives*. COM(2020) 380 final. https://ec.europa.eu/info/sites/info/files/communication-annex-eu-biodiversity-strategy-2030_en.pdf

European Commission. Directorate General for the Environment. (2010). *Soil biodiversity: Functions, threats and tools for policy makers. Final report.* Publications Office. https://data.europa.eu/doi/10.2779/14571

European Commission, E. D. G. (2014). *The EU Birds and Habitat Directives*. Luxembourg: Office for Official Publications of the European Union.

Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, 32000L0060, EP, CONSIL, OJ L 327 (2000). http://data.europa.eu/eli/dir/2000/60/oj/eng

Directive 2004/35/CE of the European Parliament and of the Council of 21 April 2004 on environmental liability with regard to the prevention and remedying of environmental damage, 32004L0035, CONSIL, EP, OJ L 143 (2004). http://data.europa.eu/eli/dir/2004/35/oj/eng

*Directive 2009/128/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for Community action to achieve the sustainable use of pesticides*, (2009) (testimony of European Parliament and of the Council). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A02009L0128-20091125

Evans, M. R. (2012). Modelling ecological systems in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1586), 181–190. https://doi.org/10.1098/rstb.2011.0172

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes, 5*(1), 19. https://doi.org/10.1186/s13717-016-0063-3

FAO. (2008). *Visual Soil Assessment (VSA) Field Guides*. Rome. http://www.fao.org/3/i0007e/i0007e00.htm

FAO. (2011). *Save and grow: A policymaker's guide to sustainable intensification of smallholder crop production* (M. Rai, T. G. Reeves, S. Pandey, & L. Collette, A c. Di). Food and Agriculture Organization of the United Nations.

FAO. (2015). *Status of the World's Soil Resources: Main Report*. Rome. http://www.fao.org/documents/card/en/c/c6814873-efc3-41db-b7d3-2081a10ede50/

FAO. (2020). *FAO Strategy on Mainstreaming Biodiversity across Agricultural Sectors*. https://doi.org/10.4060/ca7722en

Farage, C., Edler, D., Eklof, A., & Rosvall, M. (2020). *A dynamical perspective to community detection in ecological networks*. 31.

Faucon, M.-P., Houben, D., & Lambers, H. (2017). Plant Functional Traits: Soil and Ecosystem Services. *Trends in Plant Science, 22*(5), 385–394. https://doi.org/10.1016/j.tplants.2017.01.005

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology, 43*(3), 393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Fielding, A. H. (A c. Di). (1999). *Machine Learning Methods for Ecological Applications*. Springer US. https://doi.org/10.1007/978-1-4615-5289-5

Frampton, G., Van den Brink, P., & Gould, P. (2001). Effect of spring drought and irrigation on farmland arthropods in southern Britain. *Journal of Applied Ecology, 37*, 865–883. https://doi.org/10.1046/j.1365-2664.2000.00541.x

Franklin, J. (2010). *Mapping Species Distributions Spatial Inference and Prediction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810602

Freeman, E. A., & Moisen, G. G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling, 217*(1), 48–58. https://doi.org/10.1016/j.ecolmodel.2008.05.015

Fusaro, S., Gavinelli, F., Lazzarini, F., & Paoletti, M. G. (2018). Soil Biological Quality Index based on earthworms (QBS-e). A new way to use earthworms as bioindicators in agroecosystems. *Ecological Indicators, 93*, 1276–1292. https://doi.org/10.1016/j.ecolind.2018.06.007

Gagnarli, E., Goggioli, D., Tarchi, F., Guidi, S., Nannelli, R., Vignozzi, N., Valboa, G., Lottero, M. R., Corino, L., & Simoni, S. (2015). Case study of microarthropod communities to assess soil quality in different managed vineyards. *SOIL, 1*(2), 527–536. https://doi.org/10.5194/soil-1-527-2015

Galli, L., Capurro, M., Menta, C., & Rellini, I. (2014). Is the QBS-ar index a good tool to detect the soil quality in Mediterranean areas? A cork tree Quercus suber L. (Fagaceae) wood as a case of study. *Italian Journal of Zoology*, *81*(1), 126–135. https://doi.org/10.1080/11250003.2013.875601

García, V., Mollineda, R. A., & Sánchez, J. S. (2009). Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. In H. Araujo, A. M. Mendonça, A. J. Pinho, & M. I. Torres (A c. Di), *Pattern Recognition and Image Analysis* (pagg. 441–448). Springer.

Geisen, S., Briones, M. J. I., Gan, H., Behan-Pelletier, V. M., Friman, V.-P., de Groot, G. A., Hannula, S. E., Lindo, Z., Philippot, L., Tiunov, A. V., & Wall, D. H. (2019). A methodological framework to embrace soil biodiversity. *Soil Biology and Biochemistry*, *136*, 107536. https://doi.org/10.1016/j.soilbio.2019.107536

Ghiglieno, I., Donna, P., Tonni, M., & Valenti, L. (2019). Valutare la biodiversità nell'azienda viticola. *Millevigne*, *2*, 18–20.

Ghiglieno, I., Simonetto, A., Donna, P., Tonni, M., Valenti, L., Bedussi, F., & Gilioli, G. (2019). Soil Biological Quality Assessment to Improve Decision Support in the Wine Sector. *Agronomy*, *9*(10), 593. https://doi.org/10.3390/agronomy9100593

Ghiglieno, I., Simonetto, A., Orlando, F., Donna, P., Tonni, M., Valenti, L., & Gilioli, G. (2020). Response of the Arthropod Community to Soil Characteristics and Management in the Franciacorta Viticultural Area (Lombardy, Italy). *Agronomy*, *10*(5), 740. https://doi.org/10.3390/agronomy10050740

Giannini, T. C., Chapman, D. S., Saraiva, A. M., Alves-dos-Santos, I., & Biesmeijer, J. C. (2013). Improving species distribution models using biotic interactions: A case study of parasites, pollinators and plants. *Ecography*, *36*(6), 649–656. https://doi.org/10.1111/j.1600-0587.2012.07191.x

Goodfellow, I., Bengio, Y., & Courville, Aa. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org/

Goodness, J., Andersson, E., Anderson, P. M. L., & Elmqvist, T. (2016). Exploring the links between functional traits and cultural ecosystem services to enhance urban ecosystem management. *Ecological Indicators*, *70*, 597–605. https://doi.org/10.1016/j.ecolind.2016.02.031

Görgens, E. B., Montaghi, A., & Rodriguez, L. C. E. (2015). A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Computers and Electronics in Agriculture*, *116*, 221–227. https://doi.org/10.1016/j.compag.2015.07.004

Grace, J. B. (2006). *Structural Equation Modeling and Natural Systems*. Cambridge University Press. https://doi.org/10.1017/CBO9780511617799

Grace, J. B., Anderson, T. M., Olff, H., & Scheiner, S. M. (2010). On the specification of structural equation models for ecological systems. *Ecological Monographs*, *80*(1), 67–87. https://doi.org/10.1890/09-0464.1

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*(2), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models with application in R*. Cambridge University press.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling, 135*(2), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Haines-Young, R., & Potschin, M. (2010). The links between biodiversity, ecosystem services and human well-being. In D. G. Raffaelli & C. L. J. Frid (A c. Di), *Ecosystem Ecology* (pagg. 110–139). Cambridge University Press. https://doi.org/10.1017/CBO9780511750458.007

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science, 40*(3), 414–433. https://doi.org/10.1007/s11747-011-0261-6

Hamilton, A. T., Schäfer, R. B., Pyne, M. I., Chessman, B., Kakouei, K., Boersma, K. S., Verdonschot, P. F. M., Verdonschot, R. C. M., Mims, M., Khamis, K., Bierwagen, B., & Stamp, J. (2020). Limitations of trait-based approaches for stressor assessment: The case of freshwater invertebrates and climate drivers. *Global Change Biology, 26*(2), 364–379. https://doi.org/10.1111/gcb.14846

Hanaka, A., Plak, A., Zagórski, P., Ozimek, E., Rysiak, A., Majewska, M., & Jaroszuk-Ściseł, J. (2019). Relationships between the properties of Spitsbergen soil, number and biodiversity of rhizosphere microorganisms, and heavy metal concentration in selected plant species. *Plant and Soil*, *436*(1), 49–69. https://doi.org/10.1007/s11104-018-3871-7

Hanisch, M., Schweiger, O., Cord, A. F., Volk, M., & Knapp, S. (2020). Plant functional traits shape multiple ecosystem services, their trade-offs and synergies in grasslands. *Journal of Applied Ecology*, *57*(8), 1535–1550. https://doi.org/10.1111/1365-2664.13644

Hansen, B., Alrøe, H. F., & Kristensen, E. S. (2001). Approaches to assess the environmental impact of organic farming with particular regard to Denmark. *Agriculture, Ecosystems & Environment, 83*(1), 11–26. https://doi.org/10.1016/S0167-8809(00)00257-7

Harrell, F. E. (2015). *Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. https://link.springer.com/book/10.1007/978-3-319-19425-7

Hartley, W., Uffindell, L., Plumb, A., Rawlinson, H. A., Putwain, P., & Dickinson, N. M. (2008). Assessing biological indicators for remediated anthropogenic urban soils. *Science of The Total Environment*, *405*(1), 358–369. https://doi.org/10.1016/j.scitotenv.2008.06.004

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall/CRC. https://www.routledge.com/Generalized-Additive-Models/Hastie-Tibshirani/p/book/9780412343902

Hatziiordanou, L., Fitoka, E., Hadjicharalampous, E., Votsi, N., Palaskas, D., & Malak, D. (2019). Indicators for mapping and assessment of ecosystem condition and of the ecosystem service habitat maintenance in support of the EU Biodiversity Strategy to 2020. *One Ecosystem*, *4*.

Hawley, I. M., & Dobbins, T. N. (1945). The Distribution and Abundance of the Japanese Beetle from 1935 through 1943, with a Discussion of Some of the Known Factors That Influence Its Behavior. *Journal of the New York Entomological Society, 53*(1), 1–20.

Heikkinen, R. K., Luoto, M., Kuussaari, M., & Toivonen, T. (2007). Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landscape and Urban Planning*, *79*(3), 347–357. https://doi.org/10.1016/j.landurbplan.2006.04.002

Heikkinen, R. K., Marmion, M., & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, *35*(3), 276–288. https://doi.org/10.1111/j.1600-0587.2011.06999.x

Heiniger, C., Barot, S., Ponge, J.-F., Salmon, S., Meriguet, J., Carmignac, D., Suillerot, M., & Dubs, F. (2015). Collembolan preferences for soil and microclimate in forest and pasture communities. *Soil Biology and Biochemistry*, *86*, 181–192. https://doi.org/10.1016/j.soilbio.2015.04.003

Hendrickx, F., Maelfait, J.-P., Wingerden, W. V., Schweiger, O., Speelmans, M., Aviron, S., Augenstein, I., Billeter, R., Bailey, D., Bukacek, R., Burel, F., Diekötter, T., Dirksen, J., Herzog, F., Liira, J., Roubalova, M., Vandomme, V., & Bugter, R. (2007). How landscape structure, land-use intensity and habitat diversity affect components of total arthropod diversity in agricultural landscapes. *Journal of Applied Ecology*, *44*(2), 340–351. https://doi.org/10.1111/j.1365-2664.2006.01270.x

Hevia, V., Martín-López, B., Palomo, S., García-Llorente, M., Bello, F. de, & González, J. A. (2017). Trait-based approaches to analyze links between the drivers of change and ecosystem services: Synthesizing existing evidence and future challenges. *Ecology and Evolution*, *7*(3), 831–844. https://doi.org/10.1002/ece3.2692

Holmstrup, M., Maraldo, K., & Krogh, P. H. (2007). Combined effect of copper and prolonged summer drought on soil Microarthropods in the field. *Environmental Pollution*, *146*(2), 525–533. https://doi.org/10.1016/j.envpol.2006.07.013

Honvault, N., Houben, D., Nobile, C., Firmin, S., Lambers, H., & Faucon, M.-P. (2020). Tradeoffs among phosphorus-acquisition root traits of crop species for agroecological intensification. *Plant and soil*. https://link.springer.com/article/10.1007/s11104-020-04584-3

Hoyle, R. H., & Isherwood, J. C. (2013). Reporting Results from Structural Equation Modeling Analyses in Archives of Scientific Psychology. *Archives of Scientific Psychology*, *1*(1), 14–22. https://doi.org/10.1037/arc0000004

Hradsky, B. A., Penman, T. D., Ababei, D., Hanea, A., Ritchie, E. G., York, A., & Stefano, J. D. (2017). Bayesian networks elucidate interactions between fire and other drivers of terrestrial fauna distributions. *Ecosphere*, *8*(8), e01926. https://doi.org/10.1002/ecs2.1926

Hu, J., Li, D., Duan, Q., Han, Y., Chen, G., & Si, X. (2012). Fish species classification by color, texture and multi-class support vector machine using computer vision. *Computers and Electronics in Agriculture*, *88*, 133–140. https://doi.org/10.1016/j.compag.2012.07.008

Hussey, N. E., Kessel, S. T., Aarestrup, K., Cooke, S. J., Cowley, P. D., Fisk, A. T., Harcourt, R. G., Holland, K. N., Iverson, S. J., Kocik, J. F., Flemming, J. E. M., & Whoriskey, F. G. (2015). Aquatic animal telemetry: A panoramic window into the underwater world. *Science*, *348*(6240). https://doi.org/10.1126/science.1255642

Ikemoto, T. (2003). Possible existence of a common temperature and a common duration of development among members of a taxonomic group of arthropods that underwent speciational adaptation to temperature. *Applied Entomology and Zoology*, *38*(4), 487–492. https://doi.org/10.1303/aez.2003.487

Ikemoto, T. (2005). Intrinsic Optimum Temperature for Development of Insects and Mites. *Environmental Entomology*, *34*(6), 1377–1387. https://doi.org/10.1603/0046-225X-34.6.1377

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services* (summary for policy makers). Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, IPBES. https://doi.org/10.5281/ZENODO.3553579

Jacoby, D. M. P., & Freeman, R. (2016). Emerging Network-Based Tools in Movement Ecology. *Trends in Ecology & Evolution*, *31*(4), 301–314. https://doi.org/10.1016/j.tree.2016.01.011

Jarvie, S., & Svenning, J.-C. (2018). Using species distribution modelling to determine opportunities for trophic rewilding under future scenarios of climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1761), 20170446. https://doi.org/10.1098/rstb.2017.0446

Johnson, M. D. (2005). Habitat quality: A brief review for wildlife biologists. *Transactions of the Western Section of the Wildlife Society*, *41*, 11.

Jordan, R., Crall, A., Gray, S., Phillips, T., & Mellor, D. (2015). Citizen Science as a Distinct Field of Inquiry. *BioScience*, *65*(2), 208–211. https://doi.org/10.1093/biosci/biu217

Jr. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis* (7° edizione). Pearson College Div.

Jr. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications. https://us.sagepub.com/en-us/nam/a-primer-on-partial-least-squares-structural-equation-modeling-pls-sem/book244583

Juan, C., Genxu, W., Tianxu, M., & Xiangyang, S. (2017). ANN Model-Based Simulation of the Runoff Variation in Response to Climate Change on the Qinghai-Tibet Plateau, China. *Advances in Meteorology*, *2017*, e9451802. https://doi.org/10.1155/2017/9451802

Juan-Ovejero, R., Benito, E., Barreal, M. E., Rodeiro, J., & Briones, M. J. I. (2019). Tolerance to fluctuating water regimes drives changes in mesofauna community structure and vertical stratification in peatlands. *Pedobiologia*, *76*, 150571. https://doi.org/10.1016/j.pedobi.2019.150571

Karlen, D. L., Mausbach, M. J., Doran, J. W., Cline, R. G., Harris, R. F., & Schuman, G. E. (1997). Soil Quality: A Concept, Definition, and Framework for Evaluation (A Guest Editorial). *Soil Science Society of America Journal*, *61*(1), 4–10. https://doi.org/10.2136/sssaj1997.03615995006100010001x

Kenkel, N. C., & Orloci, L. (1986). Applying Metric and Nonmetric Multidimensional Scaling to Ecological Studies: Some New Results. *Ecology*, *67*(4), 919–928. https://doi.org/10.2307/1939814

Kennedy, R. E., Andréfouët, S., Cohen, W. B., Gómez, C., Griffiths, P., Hais, M., Healey, S. P., Helmer, E. H., Hostert, P., Lyons, M. B., Meigs, G. W., Pflugmacher, D., Phinn, S. R., Powell, S. L., Scarth, P., Sen, S., Schroeder, T. A., Schneider, A., Sonnenschein, R., … Zhu, Z. (2014). Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, *12*(6), 339–346. https://doi.org/10.1890/130066

Khawaldeh, S., Pervaiz, U., Elsharnoby, M., Alchalabi, A. E., & Al-Zubi, N. (2017). Taxonomic Classification for Living Organisms Using Convolutional Neural Networks. *Genes*, *8*(11). https://doi.org/10.3390/genes8110326

Kistner-Thomas, E. J. (2019). The Potential Global Distribution and Voltinism of the Japanese Beetle (Coleoptera: Scarabaeidae) Under Current and Future Climates. *Journal of Insect Science*, *19*(2), 16. https://doi.org/10.1093/jisesa/iez023

Kopittke, P. M., Menzies, N. W., Wang, P., McKenna, B. A., & Lombi, E. (2019). Soil and the intensification of agriculture for global food security. *Environment International*, *132*, 105078. https://doi.org/10.1016/j.envint.2019.105078

Krause, S., Le Roux, X., Niklaus, P. A., Van Bodegom, P. M., Lennon, J. T., Bertilsson, S., Grossart, H.-P., Philippot, L., & Bodelier, P. L. E. (2014). Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00251

Kremen, C. (2005). Managing ecosystem services: What do we need to know about their ecology? *Ecology Letters*, *8*(5), 468–479. https://doi.org/10.1111/j.1461-0248.2005.00751.x

Kuflik, T., Prodorutti, D., Frizzi, A., Gafni, Y., Simon, S., & Pertot, I. (2009). Optimization of copper treatments in organic viticulture by using a web-based decision support system. *Computers and Electronics in Agriculture*, *68*(1), 36–43. https://doi.org/10.1016/j.compag.2009.04.008

Larsen, P. E., Field, D., & Gilbert, J. A. (2012). Predicting bacterial community assemblages using an artificial neural network approach. *Nature Methods*, *9*(6), 621–625. https://doi.org/10.1038/nmeth.1975

Latif, Q. S., Saab, V. A., Hollenbeck, J. P., & Dudley, J. G. (2016). Transferability of habitat suitability models for nesting woodpeckers associated with wildfire. *The Condor*, *118*(4), 766–790. https://doi.org/10.1650/CONDOR-16-86.1

Lavelle, P., Chauvel, A., & Fragoso, C. (1995). Faunal activity in acid soils. *Plant-Soil Interactions at Low PH: Principles and Management*, 201–211. https://doi.org/10.1007/978-94-011-0221-6_29

Legras, G., Loiseau, N., Gaertner, J.-C., Poggiale, J.-C., Ienco, D., Mazouni, N., & Mérigot, B. (2019). Assessment of congruence between co-occurrence and functional networks: A new framework for revealing community assembly rules. *Scientific Reports*, *9*(1), 19996. https://doi.org/10.1038/s41598-019-56515-7

Leroux, S. J., & Loreau, M. (2015). Theoretical perspectives on bottom-up and top-down interactions across ecosystems. In T. C. Hanley & K. J. La Pierre (A c. Di), *Trophic Ecology* (pagg. 3–28). Cambridge University Press. https://doi.org/10.1017/CBO9781139924856.002

Li, C. (2020). Biodiversity assessment based on artificial intelligence and neural network algorithms. *Microprocessors and Microsystems*, *79*, 103321. https://doi.org/10.1016/j.micpro.2020.103321

Liber, Y., Cornet, D., Tournebize, R., Feidt, C., Mahieu, M., Laurent, F., & Bedell, J.-P. (2020). A Bayesian network approach for the identification of relationships between drivers of chlordecone bioaccumulation in plants. *Environmental Science and Pollution Research*, *27*(33), 41046–41051. https://doi.org/10.1007/s11356-019-07449-0

Liu, J.-L., Li, F.-R., Liu, L.-L., & Yang, K. (2017). Responses of different Collembola and mite taxa to experimental rain pulses in an arid ecosystem. *CATENA*, *155*, 53–61. https://doi.org/10.1016/j.catena.2017.03.008

Liu, Z., Peng, C., Xiang, W., Tian, D., Deng, X., & Zhao, M. (2010). Application of artificial neural networks in global climate change and ecological research: An overview. *Chinese Science Bulletin*, *55*, 3853–3863. https://doi.org/10.1007/s11434-010-4183-3

Londhe, S. S., & Kanade, D. S. S. (2015). *Bird Species Identification Using Support Vector Machine*. 11.

Machado, I., Costa, J. L., Leal, M. C., Pasquaud, S., & Cabral, H. (2019). Assessment level and time scales of biodiversity indicators in the scope of the Marine Strategy Framework Directive – A case study for the NE Atlantic. *Ecological Indicators*, *105*, 242–253. https://doi.org/10.1016/j.ecolind.2019.05.067

Madej, G., Barczyk, G., & Gdawiec, M. (2011). Evaluation of Soil Biological Quality Index (QBS-ar): Its Sensitivity and Usefulness in the Post-Mining Chronosequence – Preliminary Research. *Polish Journal of Environmental Studies*, *20*(5), 1367–1372.

Maldonado, A. D., Aguilera, P. A., & Salmerón, A. (2016). Modeling zero-inflated explanatory variables in hybrid Bayesian network classifiers for species occurrence prediction. *Environmental Modelling & Software*, *82*, 31–43. https://doi.org/10.1016/j.envsoft.2016.04.003

Manzoor, S. A., Griffiths, G., & Lukac, M. (2018). Species distribution model transferability and model grain size – finer may not always be better. *Scientific Reports*, *8*(1), 7168. https://doi.org/10.1038/s41598-018-25437-1

Marasas, M. E., Sarandón, S. J., & Cicchino, A. C. (2001). Changes in soil arthropod functional group in a wheat crop under conventional and no tillage systems in Argentina. *Applied Soil Ecology*, *18*(1), 61–68. https://doi.org/10.1016/S0929-1393(01)00148-2

Martin, E. A., Dainese, M., Clough, Y., Báldi, A., Bommarco, R., Gagic, V., Garratt, M. P. D., Holzschuh, A., Kleijn, D., Kovács-Hostyánszki, A., Marini, L., Potts, S. G., Smith, H. G., Hassan, D. A., Albrecht, M., Andersson, G. K. S., Asís, J. D., Aviron, S., Balzan, M. V., … Steffan-Dewenter, I. (2019). The interplay of landscape composition and configuration: New pathways to manage functional biodiversity and agroecosystem services across Europe. *Ecology Letters*, *22*(7), 1083–1094. https://doi.org/10.1111/ele.13265

Matthews, K. B., Schwarz, G., Buchan, K., Rivington, M., & Miller, D. (2008). Wither agricultural DSS? *Computers and Electronics in Agriculture*, *61*(2), 149–159. https://doi.org/10.1016/j.compag.2007.11.001

Mazzoncini, M., Canali, S., Giovannetti, M., Castagnoli, M., Tittarelli, F., Antichi, D., Nannelli, R., Cristani, C., & Bàrberi, P. (2010). Comparison of organic and conventional stockless arable systems: A multidisciplinary approach to soil quality evaluation. *Applied Soil Ecology*, *44*(2), 124–132. https://doi.org/10.1016/j.apsoil.2009.11.001

Menta, C., Conti, F. D., Pinto, S., Leoni, A., & Lozano-Fondón, C. (2014). Monitoring soil restoration in an open-pit mine in northern Italy. *Applied Soil Ecology*, *83*, 22–29. https://doi.org/10.1016/j.apsoil.2013.07.013

Menta, Cristina, Conti, F. D., & Pinto, S. (2018). Microarthropods biodiversity in natural, seminatural and cultivated soils—QBS-ar approach. *Applied Soil Ecology*, *123*, 740–743. https://doi.org/10.1016/j.apsoil.2017.05.020

Menta, Cristina, Conti, F. D., Pinto, S., & Bodini, A. (2018). Soil Biological Quality index (QBS-ar): 15 years of application at global scale. *Ecological Indicators*, *85*, 773–780. https://doi.org/10.1016/j.ecolind.2017.11.030

Menta, Cristina, Leoni, A., Gardi, C., & Conti, F. D. (2011). Are grasslands important habitats for soil microarthropod conservation? *Biodiversity and Conservation*, *20*(5), 1073–1087. https://doi.org/10.1007/s10531-011-0017-0

Merwe, A. van der, Dyk, E. van, & Vuuren, J. van. (2011). Decision support for grape harvesting at a South African winery. *ORiON*, *27*(2), 83–100. https://doi.org/10.5784/27-2-99

Migliorini, M., Pigino, G., Bianchi, N., Bernini, F., & Leonzio, C. (2004). The effects of heavy metal contamination on the soil arthropod community of a shooting range. *Environmental Pollution*, *129*(2), 331–340. https://doi.org/10.1016/j.envpol.2003.09.025

Millennium Ecosystem Assessment (A c. Di). (2005). *Ecosystems and human well-being: Synthesis*. Island Press.

Miller, J. (2010). Species Distribution Modeling. *Geography Compass*, *4*(6), 490–509. https://doi.org/10.1111/j.1749-8198.2010.00351.x

Monecke, A., & Leisch, F. (2012). semPLS: Structural equation modeling using partial least squares. *Faculty of Commerce - Papers (Archive)*. https://ro.uow.edu.au/commpapers/3138

Montesinos-Navarro, A., Estrada, A., Font, X., Matias, M. G., Meireles, C., Mendoza, M., Honrado, J. P., Prasad, H. D., Vicente, J. R., & Early, R. (2018). Community structure informs species geographic distributions. *PLOS ONE*, *13*(5), e0197877. https://doi.org/10.1371/journal.pone.0197877

Morrison, M. L., Marcot, B. G., & Mannan, R. W. (2006). *Wildlife-Habitat relationships—Concepts and applications* (Third edition). Island press.

Moura, A. C. M., Rocha, N. A. da, Sena, Í. S. de, Kechagioglou, X., & Campagna, M. (2018). Vegetation cover indexes in urban areas: Landscape Ecology studies with Sentinel2 images. *Revista Brasileira de Cartografia*, *70*, 1686–1714.

Neave, P., & Fox, C. A. (1998). Response of soil invertebrates to reduced tillage systems established on a clay loam soil. *Applied Soil Ecology*, *9*(1), 423–428. https://doi.org/10.1016/S0929-1393(98)00100-0

Nkurunziza, L., Watson, C. A., Öborn, I., Smith, H. G., Bergkvist, G., & Bengtsson, J. (2020). Socio-ecological factors determine crop performance in agricultural systems. *Scientific Reports*, *10*(1), 4232. https://doi.org/10.1038/s41598-020-60927-1

Odum, E. P., & Barrett, G. W. (2005). *Fundamentals of ecology* (5th ed.). Thomson Brooks/Cole.

OECD. (2005). *OECD Glossary of Statistical Terms—Environmental functions Definition*. https://stats.oecd.org/glossary/detail.asp?ID=6424

Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O’Connor, L., & Thuiller, W. (2019). Diversity indices for ecological networks: A unifying framework using Hill numbers. *Ecology Letters*, *22*(4), 737–747. https://doi.org/10.1111/ele.13221

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package* (2.5-6) [Computer software]. https://CRAN.R-project.org/package=vegan

Olden, J. D., Joy, M. K., & Death, R. G. (2006). Rediscovering the species in community-wide predictive modeling. *Ecological Applications: A Publication of the Ecological Society of America*, *16*(4), 1449–1460. https://doi.org/10.1890/1051-0761(2006)016[1449:rtsicp]2.0.co;2

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine Learning Methods Without Tears: A Primer for Ecologists. *The Quarterly Review of Biology*, *83*(2), 171–193. https://doi.org/10.1086/587826

O'Lear, H. A., & Blair, J. M. (1999). Responses of soil microarthropods to changes in soil water availability in tallgrass prairie. *Biology and Fertility of Soils*, *29*(2), 207–217. https://doi.org/10.1007/s003740050546

Ovaskainen, O., & Abrego, N. (2020). *Joint Species Distribution Modelling: With Applications in R*. Cambridge University Press. https://doi.org/10.1017/9781108591720

Özesmi, S. L., & Özesmi, U. (1999). An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, *116*(1), 15–31. https://doi.org/10.1016/S0304-3800(98)00149-5

Parisi, V. (2001). La qualità biologica del suolo. Un metodo basato sui microartropodi. *Acta Naturalia de L'Ateneo Parmense*, *37*, 105–114.

Parisi, Vittorio, Menta, C., Gardi, C., Jacomini, C., & Mozzanica, E. (2005). Microarthropod communities as a tool to assess soil quality and biodiversity: A new approach in Italy. *Agriculture, Ecosystems & Environment*, *105*(1), 323–333. https://doi.org/10.1016/j.agee.2004.02.002

Park, Y.-S., & Chon, T.-S. (2007). Biologically-inspired machine learning implemented to ecological informatics. *Ecological Modelling*, *203*(1), 1–7. https://doi.org/10.1016/j.ecolmodel.2006.05.039

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Elsevier. https://doi.org/10.1016/C2009-0-27609-4

Pecchi, M., Marchi, M., Burton, V., Giannetti, F., Moriondo, M., Bernetti, I., Bindi, M., & Chirici, G. (2019). Species distribution modelling to support forest management. A literature review. *Ecological Modelling*, *411*, 108817. https://doi.org/10.1016/j.ecolmodel.2019.108817

Pérez Bote, J., & Romero, A. (2012). Epigeic soil arthropod abundance under different agricultural land uses. *Spanish Journal of Agricultural Research*, *10*. https://doi.org/10.5424/sjar/2012101-202-11

Pérez-Expósito, J. P., Fernández-Caramés, T. M., Fraga-Lamas, P., & Castedo, L. (2017). VineSens: An Eco-Smart Decision-Support Viticulture System. *Sensors*, *17*(3), 465. https://doi.org/10.3390/s17030465

Peters, D. P. C., Havstad, K. M., Cushing, J., Tweedie, C., Fuentes, O., & Villanueva-Rosales, N. (2014). Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology. *Ecosphere*, *5*(6), art67. https://doi.org/10.1890/ES13-00359.1

Petersen, H., & Luxton, M. (1982). A Comparative Analysis of Soil Fauna Populations and Their Role in Decomposition Processes. *Oikos*, *39*(3), 288–388. JSTOR. https://doi.org/10.2307/3544689

Pomerleau, C., Sastri, A. R., & Beisner, B. E. (2015). Evaluation of functional trait diversity for marine zooplankton communities in the Northeast subarctic Pacific Ocean. *Journal of Plankton Research*, *37*(4), 712–726. https://doi.org/10.1093/plankt/fbv045

Potapov, A., Goncharov, A., Semenina, E., Korotkevich, A., Tsurikov, S., Rozanova, O., Anichkin, A., Zuev, A., Samoylova, E., Semenyuk, I., Yevdokimov, I., & Tiunov, A.

(2017). Arthropods in the subsoil: Abundance and vertical distribution as related to soil organic matter, microbial biomass and plant roots. *European Journal of Soil Biology*, *82*, 88–97. https://doi.org/10.1016/j.ejsobi.2017.09.001

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., … Duda, M. G. (2017). The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bulletin of the American Meteorological Society*, *98*(8), 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Powlson, D. S., Gregory, P. J., Whalley, W. R., Quinton, J. N., Hopkins, D. W., Whitmore, A. P., Hirsch, P. R., & Goulding, K. W. T. (2011). Soil management in relation to sustainable agriculture and ecosystem services. *Food Policy*, *36*, S72–S87. https://doi.org/10.1016/j.foodpol.2010.11.025

Priori, S., Lagomarsino, A., Agnelli, A. E., Valboa, G., Castaldini, M., Pellegrini, S., Simoni, S., D'Avino, L., Gagnarli, E., Guidi, S., Goggioli, D., Landi, S., D'Errico, G., Perria, R., Puccioni, S., Leprini, M., Zombardo, A., Storchi, P., & Costantini, E. A. C. (2015). Soil functionality assessment in degraded plots of vineyards. *Suoli Di Qualità per Una Vita Di Qualità*, *1*, 190–199. https://orgprints.org/30955/

Priya, C. A., Balasaravanan, T., & Thanamani, A. S. (2012). An efficient leaf recognition algorithm for plant classification using support vector machine. *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, 428–432. https://doi.org/10.1109/ICPRIME.2012.6208384

Puig-Montserrat, X., Stefanescu, C., Torre, I., Palet, J., Fàbregas, E., Dantart, J., Arrizabalaga, A., & Flaquer, C. (2017). Effects of organic and conventional crop management on vineyard biodiversity. *Agriculture, Ecosystems & Environment*, *243*. https://doi.org/10.1016/j.agee.2017.04.005

R Core Team. (2019). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing;* https://www.R-project.org/

Rapoport, E. H., & Tschapek, M. (1967). Soil water and soil fauna. *Rev. Ecol. Biol. Sol.*, *4*, 1–58.

Reddy, C. S., Kurian, A., Srivastava, G., Singhal, J., Varghese, A. O., Padalia, H., Ayyappan, N., Rajashekar, G., Jha, C. S., & Rao, P. V. N. (2020). Remote sensing enabled essential biodiversity variables for biodiversity assessment and monitoring: Technological advancement and potentials. *Biodiversity and Conservation*. https://doi.org/10.1007/s10531-020-02073-8

Regione Lombardia. (2013). *Basi informative dei suoli*. http://www.geoportale.regione.lombardia.it/metadati?p_p_id=PublishedMetadata_WAR_geoportalemetadataportlet&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_PublishedMetadata_WAR_geoportalemetadataportlet_view=editPublishedMetadata&_PublishedMetadata_WAR_geoportalemetadataportlet_uuid={A7138B8A-9025-4802-82BC-52267B60A3D7}&_PublishedMetadata_WAR_geoportalemetadataportlet_editType=view&_PublishedMetadata_WAR_geoportalemetadataportlet_fromAsset=true&rid=local

Regione Lombardia. (2018). *Uso e copertura del suolo 2018 (DUSAF 6.0)*. http://www.geoportale.regione.lombardia.it/metadati?p_p_id=PublishedMetadata_WAR_geoportalemetadataportlet&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=vi

ew&_PublishedMetadata_WAR_geoportalemetadataportlet_view=editPublishedMetadata&_PublishedMetadata_WAR_geoportalemetadataportlet_uuid={18EE7CDC-E51B-4DFB-99F8-3CF416FC3C70}&_PublishedMetadata_WAR_geoportalemetadataportlet_editType=view&_PublishedMetadata_WAR_geoportalemetadataportlet_fromAsset=true&rid=local

Régnière, J., Rabb, R. L., & Stinner, R. E. (1981a). Popillia japonica: Simulation of Temperature-Dependent Development of the Immatures, and Prediction of Adult Emergence. *Environmental Entomology*, *10*(3), 290–296. https://doi.org/10.1093/ee/10.3.290

Régnière, J., Rabb, R. L., & Stinner, R. E. (1981b). Popillia japonica: Effect of Soil Moisture and Texture on Survival and Development of Eggs and First Instar Grubs2. *Environmental Entomology*, *10*(5), 654–660. https://doi.org/10.1093/ee/10.5.654

Rew, J., Cho, Y., Moon, J., & Hwang, E. (2020). Habitat Suitability Estimation Using a Two-Stage Ensemble Approach. *Remote Sensing*, *12*(9), 1475. https://doi.org/10.3390/rs12091475

Rinaldi, S., Tiano, A., Şerban, S., Pittson, R., Lajic, Z., Politi, H., Murr, N., Armani, A., & Cavazza, A. (2006). Monitoring wine quality and fermentation kinetics with innovative technologies. *XXIX Congreso Mundial de La Viña y El Vino: 4a Asamblea General de La O.I.V.; Ministerio de Agricultura, Pesca y Alimentaciòn*. /paper/Monitoring-wine-quality-and-fermentation-kinetics-Rinaldi-Tiano/d2aa7c6fdebe8fe7943052daf1577a49533e8040

Ropert-Coudert, Y., & Wilson, R. P. (2005). Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, *3*(8), 437–444. https://doi.org/10.1890/1540-9295(2005)003[0437:TAPIAR]2.0.CO;2

Ruf, A., Beck, L., Dreher, P., Hund-Rinke, K., Römbke, J., & Spelda, J. (2003). A biological classification concept for the assessment of soil quality: “Biological soil classification scheme” (BBSK). *Agriculture, Ecosystems & Environment*, *98*(1), 263–271. https://doi.org/10.1016/S0167-8809(03)00086-0

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., & Plümer, L. (2010). Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, *74*(1), 91–99. https://doi.org/10.1016/j.compag.2010.06.009

Rutz, C., & Hays, G. C. (2009). New frontiers in biologging science. *Biology Letters*, *5*(3), 289–292. https://doi.org/10.1098/rsbl.2009.0089

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sanches Fernandes, L. F., Fernandes, A. C. P., Ferreira, A. R. L., Cortes, R. M. V., & Pacheco, F. A. L. (2018). A partial least squares – Path modeling analysis for the understanding of biodiversity loss in rural and urban watersheds in Portugal. *Science of The Total Environment*, *626*, 1069–1085. https://doi.org/10.1016/j.scitotenv.2018.01.127

Sander, E. L., Wootton, J. T., & Allesina, S. (2017). Ecological Network Inference From Long-Term Presence-Absence Data. *Scientific Reports*, *7*(1), 7154. https://doi.org/10.1038/s41598-017-07009-x

Santorufo, L., Van Gestel, C. A. M., Rocco, A., & Maisto, G. (2012). Soil invertebrates as bioindicators of urban soil quality. *Environmental Pollution*, *161*, 57–63. https://doi.org/10.1016/j.envpol.2011.09.042

Santos, J. P. R. dos, Fernandes, S. B., McCoy, S., Lozano, R., Brown, P. J., Leakey, A. D. B., Buckler, E. S., Garcia, A. A. F., & Gore, M. A. (2020). Novel Bayesian Networks for Genomic Prediction of Developmental Traits in Biomass Sorghum. *G3: Genes, Genomes, Genetics*, *10*(2), 769–781. https://doi.org/10.1534/g3.119.400759

SATA Studio agronomico. (2020). *Agronomisata*. https://www.agronomisata.it/progetto-biopass/

Schmitz, O. J., Buchkowski, R. W., Burghardt, K. T., & Donihue, C. M. (2015). Functional Traits and Trait-Mediated Interactions: Connecting Community-Level Interactions with Ecosystem Functioning. In S. Pawar, G. Woodward, & A. I. Dell (A c. Di), *Advances in Ecological Research* (Vol. 52, pagg. 319–343). Academic Press. https://doi.org/10.1016/bs.aecr.2015.01.003

Seniczak, A., Seniczak, S., García-Parra, I., Ferragut, F., Xamaní, P., Graczyk, R., Messeguer, E., Laborda, R., & Rodrigo, E. (2018). Oribatid mites of conventional and organic vineyards in the Valencian Community, Spain. *Acarologia*, *58*(Suppl), 119–133. https://doi.org/10.24349/acarologia/20184281

Serrano, E., López-Soria, S., Trinchera, L., & Segalés, J. (2014). The use of null models and partial least squares approach path modelling (PLS-PM) for investigating risk factors influencing post-weaning mortality in indoor pig farms. *Epidemiology & Infection*, *142*(3), 530–539. https://doi.org/10.1017/S0950268813001295

Shakir, M. M., & Ahmed, S. (2015). Seasonal abundance of soil arthropods in relation to meteorological and edaphic factors in the agroecosystems of Faisalabad, Punjab, Pakistan. *International Journal of Biometeorology*, *59*(5), 605–616. https://doi.org/10.1007/s00484-014-0874-9

Shanovich, H. N., Dean, A. N., Koch, R. L., & Hodgson, E. W. (2019). Biology and Management of Japanese Beetle (Coleoptera: Scarabaeidae) in Corn and Soybean. *Journal of Integrated Pest Management*, *10*(9). https://doi.org/10.1093/jipm/pmz009

Shiferaw, H., Bewket, W., & Eckert, S. (2019). Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem. *Ecology and Evolution*, *9*(5), 2562–2574. https://doi.org/10.1002/ece3.4919

Shipley, B. (2016). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R* ($2^a$ ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139979573

Sjursen, H., & Holmstrup, M. (2004). Cold and drought stress in combination with pyrene exposure: Studies with Protaphorura armata (Collembola: Onychiuridae). *Ecotoxicology and Environmental Safety*, *57*(2), 145–152. https://doi.org/10.1016/S0147-6513(02)00144-6

Smukler, S. M., Sánchez-Moreno, S., Fonte, S. J., Ferris, H., Klonsky, K., O'Geen, A. T., Scow, K. M., Steenwerth, K. L., & Jackson, L. E. (2010). Biodiversity and multiple ecosystem functions in an organic farmscape. *Agriculture, Ecosystems & Environment*, *139*(1), 80–97. https://doi.org/10.1016/j.agee.2010.07.004

Soil Science Division Staff. (2017). *Soil survey manual* (USDA Handbook 18). C. Ditzler, K. Scheffe, and H.C. Monger (eds.).

Sonnier, G., Shipley, B., & Navas, M.-L. (2010). Quantifying relationships between traits and explicitly measured gradients of stress and disturbance in early successional plant communities. *Journal of Vegetation Science*, *21*(6), 1014–1024. https://doi.org/10.1111/j.1654-1103.2010.01210.x

Stang, M., Klinkhamer, P. G. L., Waser, N. M., Stang, I., & van der Meijden, E. (2009). Size-specific interaction patterns and size matching in a plant–pollinator interaction web. *Annals of Botany*, *103*(9), 1459–1469. https://doi.org/10.1093/aob/mcp027

Stašiov, S., Stašiová, A., Svitok, M., Michalková, E., Slobodník, B., & Lukáčik, I. (2012). Millipede (Diplopoda) communities in an arboretum: Influence of tree species and soil properties. *Biologia*, *67*(5). https://doi.org/10.2478/s11756-012-0097-7

Stillman, J. H. (2003). Acclimation Capacity Underlies Susceptibility to Climate Change. *Science*, *301*(5629), 65–65. https://doi.org/10.1126/science.1083073

Sutherst, R. W., & Maywald, G. F. (1985). A computerised system for matching climates in ecology. *Agriculture, Ecosystems & Environment*, *13*(3), 281–299. https://doi.org/10.1016/0167-8809(85)90016-7

Szendrei, Z., & Isaacs, R. (2005). Do plant cues influence the oviposition behavior of Japanese beetles? *Entomologia Experimentalis et Applicata*, *117*(2), 165–174. https://doi.org/10.1111/j.1570-7458.2005.00346.x

Szendrei, Z., & Isaacs, R. (2006). Ground Covers Influence the Abundance and Behavior of Japanese Beetles. *Environmental Entomology*, *35*(3), 789–796. https://doi.org/10.1603/0046-225X-35.3.789

Tabaglio, V., Gavazzi, C., & Menta, C. (2009). Physico-chemical indicators and microarthropod communities as influenced by no-till, conventional tillage and nitrogen fertilisation after four years of continuous maize. *Soil and Tillage Research*, *105*(1), 135–142. https://doi.org/10.1016/j.still.2009.06.006

Tenenhaus, M., Vinzi, V. E., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, *48*(1), 159–205. https://doi.org/10.1016/j.csda.2004.03.005

The Wine Institute, & Kennedy/Jenks Consultants. (2008). *Comprehensive Guide to Sustainable Management of Winery Water and Associated Energy*. 316.

Therneau, T., & Atkinson, B. (2018). *Recursive Partitioning and Regression Trees. R package version 4.1-13*. https://CRAN.R-project.org/package=rpart

Therond, O., Duru, M., Roger-Estrade, J., & Richard, G. (2017). A new analytical framework of farming system and agriculture model diversities. A review. *Agronomy for Sustainable Development*, *37*(3), 21. https://doi.org/10.1007/s13593-017-0429-7

Thompson, P. R., Fagan, W. F., & Staniczenko, P. P. A. (2019). *Predictor species: Improving assessments of rare species occurrence by modelling environmental co-responses* [Preprint]. Ecology. https://doi.org/10.1101/546879

Tsiafouli, M. A., Kallimanis, A. S., Katana, E., Stamou, G. P., & Sgardelis, S. P. (2005). Responses of soil microarthropods to experimental short-term manipulations of soil moisture. *Applied Soil Ecology*, *29*(1), 17–26. https://doi.org/10.1016/j.apsoil.2004.10.002

Ulanowicz, R. E. (2004). Quantitative methods for ecological network analysis. *Computational Biology and Chemistry*, *28*(5), 321–339. https://doi.org/10.1016/j.compbiolchem.2004.09.001

Ullah, S., Ai, C., Huang, S., Song, D., Abbas, T., Zhang, J., Zhou, W., & He, P. (2020). Substituting ecological intensification of agriculture for conventional agricultural practices increased yield and decreased nitrogen losses in North China. *Applied Soil Ecology*, *147*, 103395. https://doi.org/10.1016/j.apsoil.2019.103395

United Nations. (1992). *Convention on biological diversity*. https://www.cbd.int/doc/legal/cbd-en.pdf

United Nations, G. A. (A c. Di). (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. https://doi.org/10.1891/9780826190123.ap02

Valenti, L., Ghiglieno, I., Tonni, M., Donna, P., Dell'Orto, M., & Carletti, F. (2013). GEA.vite, an Example of Assesment Protocol to Evaluate the Efficiency and Sustainability of Viticultural Italian Companies. *American Journal of Industrial and Business Management*, *3*(3), 322–329. https://doi.org/10.4236/ajibm.2013.33037

van Straalen, N. M. (1998). Evaluation of bioindicator systems derived from soil arthropod communities. *Applied Soil Ecology*, *9*(1), 429–437. https://doi.org/10.1016/S0929-1393(98)00101-2

van Straalen, N. M., & Verhoef, H. A. (1997). The Development of a Bioindicator System for Soil Acidity Based on Arthropod pH Preferences. *Journal of Applied Ecology*, *34*(1), 217–232. JSTOR. https://doi.org/10.2307/2404860

Varga, I. (2003). Structure and changes of macroinvertebrate community colonizing decomposing rhizome litter of common reed at Lake Fertő/Neusiedler See (Hungary). *Hydrobiologia*, *506*(1), 413–420. https://doi.org/10.1023/B:HYDR.0000008619.84133.9f

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer. https://www.springer.com/gp/book/9780387954578

Vikram Reddy, M., & Venkataiah, B. (1990). Seasonal abundance of soil-surface arthropods in relation to some meteorological and edaphic variables of the grassland and tree-planted areas in a tropical semi-arid savanna. *International Journal of Biometeorology*, *34*(1), 49–59. https://doi.org/10.1007/BF01045820

Villani, M. G., Allee, L. L., Díaz, A., & Robbins, P. S. (1999). Adaptive strategies of edaphic arthropods. *Annual Review of Entomology*, *44*(1), 233–256. https://doi.org/10.1146/annurev.ento.44.1.233

Vinzi, V. E., Trinchera, L., & Amato, S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (A c. Di), *Handbook of Partial Least Squares* (pagg. 47–82). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-32827-8_3

Wallwork, J. A. (1972). Distribution Patterns and Population Dynamics of the Micro-Arthropods of a Desert Soil in Southern California. *Journal of Animal Ecology*, *41*(2), 291–310. JSTOR. https://doi.org/10.2307/3470

Water Research Commission, Sieben, E. J. J., Mtshali, H., Janks, M., & South Africa. (2014). *National wetland vegetation database: Classification and analysis of wetland vegetation types for conservation planning and monitoring* (WRC Report No. 1980/1/14).

Weiss, G. M., & Provost, F. (2001). *The Effect of Class Distribution on Classifier Learning: An Empirical Study* (Technical Report ML-TR-44; pag. 26). Department of Computer Science, Rutgers University.

Wenger, S., & Olden, J. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*, 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Weygoldt, P. (1969). *The biology of pseudoscorpions*. Harvard University Press, Cambridge. https://trove.nla.gov.au/work/15737764?q&versionId=45255718

Wezel, A., Casagrande, M., Celette, F., Vian, J.-F., Ferrer, A., & Peigné, J. (2014). Agroecological practices for sustainable agriculture. A review. *Agronomy for Sustainable Development*, *34*(1), 1–20. https://doi.org/10.1007/s13593-013-0180-7

Wiens, J. A., & Rotenberry, J. T. (1981). Habitat Associations and Community Structure of Birds in Shrubsteppe Environments. *Ecological Monographs*, *51*(1), 21–42. https://doi.org/10.2307/2937305

Wilhere, G. F. (2002). Adaptive Management in Habitat Conservation Plans. *Conservation Biology*, *16*(1), 20–29. https://doi.org/10.1046/j.1523-1739.2002.00350.x

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis for the analysis of data. *Biometrika*, *55*(1), 1–17. https://doi.org/10.1093/biomet/55.1.1

Williams, J. W., & Jackson, S. T. (2007). Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, *5*(9), 475–482. https://doi.org/10.1890/070037

Winkler, T., Schopf, K., Aschemann, R., & Winiwarter, W. (2016). From farm to fork – A life cycle assessment of fresh Austrian pork. *Journal of Cleaner Production*, *116*, 80–89. https://doi.org/10.1016/j.jclepro.2016.01.005

Wolters, V. (2001). Biodiversity of soil animals and its function. *European Journal of Soil Biology*, *37*(4), 221–227. https://doi.org/10.1016/S1164-5563(01)01088-3

Wong, M. K. L., Guénard, B., & Lewis, O. T. (2019). Trait-based ecology of terrestrial arthropods. *Biological Reviews*, *94*(3), 999–1022. https://doi.org/10.1111/brv.12488

Wu, P., Liu, X., Liu, S., Wang, J., & Wang, Y. (2014). Composition and spatio-temporal variation of soil microarthropods in the biodiversity hotspot of northern Hengduan Mountains, China. *European Journal of Soil Biology*, *62*, 30–38. https://doi.org/10.1016/j.ejsobi.2014.02.013

Xu, G.-L., Kuster, T. M., Günthardt-Goerg, M. S., Dobbertin, M., & Li, M.-H. (2012). Seasonal exposure to drought and air warming affects soil Collembola and mites. *PloS One*, *7*(8), e43102. https://doi.org/10.1371/journal.pone.0043102

Yan, S., Singh, A. N., Fu, S., Liao, C., Wang, S., Li, Y., Cui, Y., & Hu, L. (2012). A soil fauna index for assessing soil quality. *Soil Biology and Biochemistry*, *47*, 158–165. https://doi.org/10.1016/j.soilbio.2011.11.014

Zarraonaindia, I., Owens, S. M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N. A., Mills, D. A., Martin, G., Taghavi, S., Lelie, D. van der, & Gilbert, J. A.

(2015). The Soil Microbiome Influences Grapevine-Associated Microbiota. *MBio*, *6*(2). https://doi.org/10.1128/mBio.02527-14

Zhu, Gengping, Li, H., & Zhao, L. (2017). Incorporating anthropogenic variables into ecological niche modeling to predict areas of invasion of Popillia japonica. *Journal of Pest Science*, *90*(1), 151–160. https://doi.org/10.1007/s10340-016-0780-5

Zhu, Guodong, Luo, Y., Xue, M., Zhao, H., Xia, N., & Wang, X. (2018). Effects of high-temperature stress and heat shock on two root maggots, Bradysia odoriphaga and Bradysia difformis (Diptera: Sciaridae). *Journal of Asia-Pacific Entomology*, *21*(1), 106–114. https://doi.org/10.1016/j.aspen.2017.11.001