DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Defining ecoregions based on soil invertebrates for defining pesticide exposure scenarios

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Ecología, realizada sob a orientação científica do Professor Doutor José Paulo Sousa, Professor Auxiliar do Departamento da Ciências da Vida da Universidade de Coimbra e do Doutor Jörg Römbke, Managing Director da ECT, Oekotoxikologie GmbH, Frankfurt

Liliana del Carmen Murillo Contreras

2011

Acknowledgments

I would like to thank all the people that have shared with me throughout my masters experience, professionally and personally.

To professor José Paulo Sousa, for his patience and guidance to make sure I was learning, and for keeping me motivated in this effort. To Dr. Jörg Römbke for his insight and ideas developing the concept. To professors João Cabral and Mario Santos and Rita Bastos from Universidade de Trás-os-Montes e Alto Douro for their help with the modeling and statistical analysis of the data. To JRC and EFSA for authorizing the use of their maps and databases.

To my friends and family for their support and love, that made this masters program an unforgettable experience.

Defining ecoregions based on soil invertebrates for defining pesticide exposure scenarios

Abstract

Environmental Risk Assessment (ERA) is a process of identifying and evaluating the adverse effects on the environment caused by a chemical substance. Modeling environmental relevant concentrations in soil (ERCsoil) requires a different approach than the standard exposure scenario. Ecologically relevant scenarios must calculate exposure according to the habitats of soil organisms' communities, their role in supporting soil functions and allow modeling ERC in different soil layers all around Europe. The aim of this study is to contribute in the definition of a EU-wide ecoregion-based map to improve the ecological relevance of soil exposure scenarios for collembola and isopods. These organisms were selected based on their importance ecological role in European soils, presence in a wide geographical scale, different morphological and ecological characteristics and data availability. Finland, Germany and Portugal were selected as model countries. The European Food Safety Authority (EFSA) databases used for this study compile information from published and some unpublished articles, species catalogs, and regional inventories. European Joint Research Center (JRC) maps provided the missing environmental variables for the spatial analysis. Soil organisms groups were classified by life form: euedaphic, hemiedaphic and epigeic for collembola; soil dwellers and litter dwellers for isopods; and then classified by dominance classes. Life form raw richness was used to create a generalized linear model (GLM) to describe the soil organisms' distribution and class dominance. The software STELLA was employed to design a Stochastic Dynamic Methodology (StDM) model to predict distribution of the target soil groups. The results of the GLM and StDM model simulations were incorporated in ArcView 9.2 using the spatial analyst and geostatistical analysis extensions. The raster calculator and Ordinary Kriging were chosen to produce raw richness distribution maps for all life forms of collembola and isopods and to map class dominance. The models were not very successful at predicting low frequencies of dominance classes. Regardless, they were in line with ecological and biogeographic information for the considered groups. For collembola, Finland was dominated by epigeic species, while Portugal showed a dominance of epigeic and hemiedaphic species. In the case of Germany, the analysis methods reached different conclusions and patterns, the raster calculator analysis showed clear epigeic dominance while the ordinary kriging map displayed epigeic and hemiedaphic dominance. For isopods, both methodologies produced similar values for the two life forms in all countries, on average from 0 to 50% for soil dwellers and from 50 to 100% for litter dwellers. The only worst-case scenario predicted for pesticide assessment in all three countries was litter to 1 cm. Overall, the results obtained from the spatial and the geostatistical analysts were not helpful to define ecoregions for pesticide risk assessment given the available data and the selected GLM variables, as they do not provide enough discrimination between worst-case scenarios. Future studies should consider including only site data with complete environmental variables information and a specified geographical location. Abundance would also be a welcome improvement to the model.

Keywords: ecoregions, collembola, isopods, risk assessment, geostatistical analysis, GLM, StDM

Index

Index of Figures

Index of tables

Index of Annexes

1 Introduction

Background considerations

Environmental Risk Assessment (ERA) is a process of identifying and evaluating the adverse effects on the environment caused by a chemical substance. From the perspective of risk assessment, environmental exposure to a chemical is predicted and compared to a predicted no-effect concentration, supplying risk ratios for different media.

An ecotoxicological risk assessment has to start with the question 'what has to be protected?' and include a protection aim with spatial and temporal components. Risk assessments of hazardous chemicals like plant production products (PPPs) are traditionally conducted by comparing a generically derived effect concentration with a generically derived exposure concentration (Toxicity-Exposure Ratio or TER). The endpoint of the exposure assessment is the Predicted Environmental Concentration (PEC).

Since the 1980's, predicted concentrations of pesticides in soil in Europe are calculated by using simple assumptions: the amount of the test substance per hectare is evenly distributed on the top 5 cm of a soil with a density of 1.5 g/cm3 dry weight ("standard" scenario; e.g. BBA 1986). Later modifications addressed the question of how much of the applied amount will reach the soil, by introducing vegetation interception factors or by modelingspray drift (Ganzelmeier, 2000). But consensus is building regarding the differences between soils across Europe and the general lack of knowledge on the soil organism communities that regulation should be protecting that is challenging this calculation (Boesten *et al.*, 2007). The Ecotoxicologically Relevant Concentration (ERC) represents the interface between effect assessment and exposure assessment defined as the type of concentration that gives the best correlation to ecotoxicological effects (Boesten *et al.*, 2007).

In the currently used Guidance Documents (for example, EC 2002) the protection goals are only described in a general way, but it seems that the protection of the structure and functions of the soil organism communities is the ultimate goal of the ERA of pesticides (EFSA 2007). Nevertheless, it seems that the discussion on pesticide ERA is moving in the direction already laid down in the draft Soil

Framework Directive (SFD; EC 2006) towards the protection of soil and its functions. One important, potentially far-reaching issue in this context is whether the exposure of soil organism communities towards pesticides has to be described on the species level or, probably more practical, on the level of ecologically defined life form types (e.g. for earthworms (Lee 1959, cited in Lee 1985; Bouché 1977).

Exposure estimations can provide an approximation but a pesticide active ingredient can show different behaviour in soils, depending on interactions between physical and chemical properties of the compound and soil characteristics. Adsorption or leaching of a chemical will result in different exposure risks to soil organisms, as communities will be more affected according to their life form types, particularly according to their preferred depth.

Modeling environmental relevant concentrations in soil (ERCsoil) requires a different approach than the standard exposure scenario. Ecologically relevant scenarios must calculate exposure according to the habitats of soil organisms' communities, their role in supporting soil functions and allow modeling ERC in different soil layers all around Europe. Therefore, abiotic differences of soil properties, as well as ecological differences of soil organism communities, have to be included into the process of defining exposure (EFSA, 2009). However, one must be aware that not only exposure has to be discussed as the topic is strongly influenced by the more general question of which are the protection goals of pesticide registration (Van der Linden, 2008).

The European Union has developed guides for exposure assessment in soil, with the FOrum for the Coordination of pesticide fate models and their USe (FOCUS). The organization is an initiative of the European Commission to harmonize the calculation of predicted environmental concentrations (PEC) of active substances of plant protection products (PPP) in the framework of the EU Directive 91/414/EEC and is based on cooperation between scientists of regulatory agencies, academia and industry. It started in 1993 via the FOCUS Leaching Modeling Workgroup and the installation of the FOCUS Steering Committee. In 1997, they developed a simple approach for estimating PECsoil but did not include first-tier scenarios, which were eventually created by FOCUS workgroups on surface water and groundwater.

FOCUS (1997) concluded that scenarios of crop, soil and weather data are needed not just for estimatingconcentrations of pesticides in soil, but also for leaching and other fate andexposure assessments. These scenarios should be accessible to all and should cover the whole EU. Soil - climate scenarios were constructed which can be used in the first step of the registration evaluation of plant protection products inEurope. To obtain predicted environmental concentrations (PEC) for realistic worst-case conditions, data has tobe analyzed further, including volatilization, interception by crop canopy, temperature and leaching. In further steps of the evaluation more refined scenariosshould be used in order not to overestimate or underestimate the concentrationsthat might occur in reality.

In 2006, detail guidance was achieved on estimating degradation rate parameters for laboratory and field studies, the emphasis of the work group was on analyzing data sets from existing regulatory studies rather than on developing strategies for conducting these regulatory studies, and no exposure scenarios were created (FOCUS, 2006).

The European Food Safety Authority'sPanel on Plant Protection Products and their Residues (PPR) has written multiplescientific opinions regarding pesticide risk assessment. One of the most recent papers focus on the assessment of exposure of organisms to substances in soil, taking into account crop type, soil tillage system, crop management and application techniques within the EU agriculture and incorporation of dissipation rates of PPP as well as wash-off. They also propose tiered approaches for exposure assessment based on information of crops planted within a regulatory zone under conventional and reduced tillage:

- Tier 1 is proposed to be based on a simple analytical model.
- Tier 2 is to be based on simulations with numerical models.
- Tier 3 is proposed to be again a simple analytical model but in this Tier specific crops and/or plant protection products with specific properties may be considered.
- Tier 4 is to be based on simulations with numerical models but, as in Tier 3, specific crops and/or plant protection products with specific properties can be considered.

To keep the approach as simple as possible, the Panel recommends having within Tier 1 and Tier 2 only one scenario for concentration in total soil and only one scenario for concentration in pore water. These scenarios are used for all annual crops and for all plant protection products in each regulatory zone. The development of soil exposure scenarios in the proposed Tier 4 is affected by limitations of existing soil databases at EU level, a problem that can only be overcome with a considerable amount of expert judgment for the selection of the soil profiles of the scenarios. These models can only be reliable if access to high-quality databases of soils, crop areas and weather with 100% coverage of the EU-27 is easily available to the stakeholders (EFSA, 2010a).

2 Objectives of the study

The aim of this study is to contribute in the definition of a EU-wide Ecoregion-based map to improve the ecological relevance of soil exposure scenarios for selected soil organism communities. After characterizing each group identified according to the life-form types and looking at the proportion of species from each life-form type, multivariate methods will be used to establish a link between community structure and soil properties, climatic factors and land use. The endpoint of this study is to develop a model to predict worst-case scenarios of pesticide exposure according to community composition, by using a holistic stochastic dynamic methodology (StDM) to improve the ecologically relevant exposure scenarios for collembolan and isopods.

2.1 The Ecoregion concept

The European Food Safety Authority’s Panel on Plant Protection Products and their Residues (PPR)has suggested an ecoregion approach to predicting effects of PPPs on non-target species and communities (2010b).

Ecoregions contain characteristic and geographically distinct assemblages of natural communities associated to specific soil and climate conditions. Ecoregions based on plant cover only (or on Potential Natural Vegetation) cannot predict the exact distribution of soil organisms, as they are also strongly influenced by physical and chemical soil properties; and similar community assemblages (in functional terms) can be found in different land-uses with different plant covers. For this reason, vegetation-based typologies are not suitable to define Ecologically Relevant Exposure Scenarios(ERES) for soil organisms(EFSA, 2010ab).

The exposure assessment of plant protection products in soil can be refined based on a new underlying concept using ecoregion maps to define ecologically relevant exposure profiles. The EFSA-developed concept is based on the following principles:

- Europe can be divided into a number of regions defined by soil properties, land-use and climate.
- Each region supports specific soil organism communities that may play different roles in supporting relevant soil services.
- The different species within each community could be subdivided into groups based on similar traits ("trait groups") that are related in the way they are exposed to chemicals.
- The combination of soil properties, land-use, climate and the potential soil community (based on a unique assemblage of "trait groups") defines an ecoregion.
- Each ecoregion is characterized by a different set of exposure scenarios, e.g. depth profiles that are defined by the trait groups present for which homogeneous ERC values can be modeled.

Within the soil community, it is the species traits that determine the way they are exposed to the pesticides and are the key to define ecoregions and their exposure profiles. Nonetheless, the actual exposure/availability may differ with respect to environmental conditions, since the degradation and/or metabolisation of PPPs as well as their availability also depend on soil properties and climate. Depending on the region, a combination of its abiotic properties and soil communities should be considered when modeling the actual exposure to a plant protection product.

When modeling the ERC of plant protection product at a specific site, the result is not only relevant for that specific set of profiles and that specific site but for all sites belonging to the same ecoregion (with comparable combinations of specific abiotic and biotic factors).

The compilation of environmental data such as soil, land-use and climate on a geographical basis are priorities, but the collection of ecological and geographical distribution data for soil fauna is essential to define their relative importance within each ecoregion and to define the relevant soil layers where organisms are exposed.

2.1.1 Fauna group selection

The classification of soil organism communities at agricultural sites regarding their exposure towards pesticides can, in theory, be performed according to taxonomical criteria. However, this approach does not cover the ecological similarities of communities consisting of different species within the same organism group. In other words, for example different earthworm species living in different regions of the EU can share the same morphological, physiological and ecological properties or traits, meaning that despite that they are taxonomically different; they fulfill the same ecological role. Traits can be morphological (e.g. size, permeability of exoskeleton, lipid content, complexity of the nervous system), physiological (e.g. mode of respiration, detoxifying enzymes or digestive strategy), and ecological (e.g. mobility, feeding behaviour, trophic level, place in the food web). In fact, the main constraint of the focus on individual species is that knowledge on the biology of many soil organism species is still in its early stages. Soil communities are very diverse and the richness of soil organisms in a certain location can easily overcome several hundreds of species (Lavelle and Spain, 2001) it is simply impossible to include all soil organism groups.

In contrast, it is proposed to put the focus on life form types, consisting of several species (in the case of nematodes even families). These life form types can then be used for the derivation and classification of exposure scenarios.

It was necessary to identify a small number of organism groups representing the most important guilds in European soils (Sousa *et al.*, 2009). The following selection criteria were used, listed in order of importance:

1. Important ecological role in European soils, in terms of biomass, soil structuring activity, and place in the food web.
2. Presence across a wide geographical scale.
3. Different morphological and ecological characteristics influencing exposure:
    a. Different size classes
    b. Soft-bodied versus hard-bodied species
4. Availability of information regarding their distribution, preferably in databases, maps or review papers.

5. Availability of trait data on the selected groups, particularly life-form traits indicating at which soil depth they are mainly active.
6. Groups including species being regularly used in ecotoxicological testing (for combining information from exposure modeling and effect testing).

From these criteria, the following combination of four groups fulfills the requirements for ecoregion classification (EFSA, 2010bb):

1. Collembola (springtails): Mesofauna, hard-bodied, important microbial regulators during the decomposition process, widely distributed with many species all over Europe;
2. Isopoda (woodlice): Macrofauna, hard-bodied, most species prefer warmer regions; important detritivores in the early stages of organic matter decomposition (usually called "litter transformers")
3. Lumbricidae (earthworms): Macrofauna, soft-bodied, important microbial regulators often with very high biomass, key group for soil structure formation and maintenance, widely distributed in Europe.
4. Enchytraeidae (potworms): Mesofauna, soft-bodied, important microbial regulators often in very high numbers, prefer cool, acid soils.

This selection of groups fits with recommendations recently made for biological soil monitoring in the EU. Sampling of earthworms (plus enchytraeids), springtails and soil microorganisms was recommended by the EU funded FP6 ENVASSO10 project for a first tier, while other organism groups (like nematodes) could be used to address specific biodiversity monitoring questions (Bispo *et al.*, 2009; EFSA, 2010c).

The criteria for the selection of biodiversity indicators adopted by ENVASSO use ecological relevance as the utmost condition for selecting an organism group. Nematodes, soil mites, diplopods and slugs are examples of ecological relevant groups and well-established functional classification, however the existing biogeographical information is scarce and limited to a few countries within the EU. For microorganisms, despite their dominance and fundamental relevance for the processes in soil, there are problems in classifying a functional endpoint (e.g., microbial respiration).

In particular, the different life forms of the organism groups assessed in the opinion paper are exposed in different soil depth profiles, as shown in Table 1.

Table 1: Soil depth profiles where the life form groups are exposed to pesticides (EFSA, 2010b).

	Depth profile where the organisms are exposed					
	Litter layer	0 – 1 cm	0 - 2.5 cm	0 – 5 cm	0 – 20 cm	burrows
Enchytraeids	litter dweller	litter dweller	intermediate	mineral dweller		
Earthworms	epigeic + anecic	epigeic + anecic			endogeic	anecic
Isopoda	litter dweller	litter dweller		soil dweller		
Collembola	epigeic	Epigeic	hemiedaphic	euedaphic		

The PPR Panel on their 2010 paper on ecoregion definition succeeded in using earthworms and entrichaids for worst-case scenario prediction, but the model used did not accomplish a good fit for collembolan and isopods. For this reason, this study is focusing on developing a new methodology to create ecoregion maps for only these two groups.

2.1.2 Classification of collembolan communities

Collembola is a very diverse taxon with about 7,000 species currently described, although the total number of existing species is expected to be as much higher (Deharveng, 2004). They are apterous hexapods close to the true insects, small and elongate with a characteristic springing organ (furca) that allows rapid jumping movements. Their body lengths range from a few tenths of a mm to 1-2 cm with individual biomasses between 1-20 ug dry weight. They live in the litter or in the pore space of the upper 5—10 cm of soil and are mainly saprophagous, feeding mainly on fungi, bacteria or algae growing on decomposing plant litter (Christiansen, 1964; Ponge, 1991; Lavelle and Spain, 2001). There is, however, considerable variation between species, especially between litter and soil dwellers.

Their role in soil processes is important, acting mainly as catalysts of the organic matter decomposition process (Petersen, 2002). Feeding on plant material and excreting it partially decomposed as fecal pellets, they contribute to increase the surface area for microbial attack (Hasegawa and Takeda, 1995); by doing so, they also act as dispersal agents of fungal spores and bacteria. Moreover, acting as selective grazers, Collembola may promote fungal succession in decomposing plant material (Faber *et al.*, 1992). This aspect makes them, together with nematodes, important bio-control agents in soil.

Due to their large specific and functional diversity, Collembola are known indicators of soil biodiversity (Bispo *et al.*, 2009) and changes in community composition and structure are used as ecological indicators of habitat quality both in crop and forest areas (Bonnet *et al.*, 1976; Filser, 1995; Heisler and Kaiser, 1995; Lavelle and Spain, 2001; Loranger and Bandyopadhyaya, 2001; Frampton and van den Brink, 2002; Van den Brink, 2002).

The vertical niche differentiation of collembolans is correlated along with species-specific morphological traits. According to the “life form concept” (after (Gisin, 1943) and (Christiansen, 1964)) springtails can be categorized based on the size of furca (springing organ) and antennae, the number of ocellae and their pigmentation into epigeic, hemiedaphic and euedaphic species. Although some species are strictly confined to a certain soil layer, many species have a broader vertical niche. Since they do not have the ability to create burrows, springtails depend on the existing pore system and burrows made by other organisms. The highest density of collembolans in open land habitats of central Europe can be expected in the upper 5 to 10 cm soil layer. Vertical migration regularly exists and is mainly induced by climatic factors.

In the following, the three ecological classes of Collembola are defined in table 2.

Table 2: Ecological classes of Collembola

Life form class	Characteristics	Example species
Euedaphic: Species with very low dispersal ability, living down to 5cm layer (in some case down to 10cm)	Blind species; very short antennae; furca absent or not well developed	*Protaphorura armata, Mesaphorura krausbaueri*

Hemiedaphic: Medium dispersal species, living down to 2.5cm layer	Variable number of ocelli; short antennae; furca reduced or short	*Megalothorax minimus, Micranurida pygmaea, Isotomiella minor, Folsomia quadrioculata, Folsomia candida*
Epigeic: Fast dispersal species, living in soil surface	Most species with more than 5+5 ocelli; long to very long antennae; furca fully developed	*Parisotoma notabilis, Entomobrya multifasciata, Pogonognathellus flavescens*

Community structure will vary not only between regions (being affected by factors like climate and major soil type), but also within regions, being influenced by the crop-type and management strategy adopted. This creates difficulties when trying to predict the community composition (or "focal communities") based on abiotic factors (e.g., soil parameters, climate) aiming at defining expose scenarios for this group.

Contrary to what happens with some representative earthworm species, the association between particular Collembola species and certain soil parameters is difficult to make. Although some information is available about the particular relations of some species with soil pH (Vikamaa and Huhta, 1986; Van Straalen and Verhoef, 1997; Ponge, 2000; Loranger and Bandyopadhyaya, 2001) and soil humus types (Ponge and Prat, 1982; Chagnon, 2000), to derive soil pedo-transfer functions for representative species of each life-form class is difficult. The strategy to adopt with this group would be to work at life-form level, classifying the community according to the relative richness of each life-form group, and to relate each community type with a range of soil and climate parameters.

2.1.3 Classification of isopod communities

With approximately 650 species described for central and southern Europe (Schmoelzer, 1965; Paoletti and Hassall, 1999), isopods are key macro-detritivores in litter systems of several terrestrial environments (Sutton *et al.*, 1972; Lavelle and Spain, 2001). Their size ranges from a few millimeters to 1-2 cm and their fresh weight is of the order of a few milligrams. Most species are highly susceptible to water loss and are thus restricted to moist, sheltered habitats although a few drought-resistant species colonize desert habitats. They also have little resistance to cold temperatures. In temperate environments, they have long periods of quiescence during winter.

Also named litter transformers“, terrestrial isopods are important players in the decomposition processes in terrestrial systems (Van Wensem *et al.*, 1993; Szlavecz, 1995). Participating in an early phase of this process, they promote litter fragmentation, by eating it, and influence microbial dynamics by altering substrate quality, when excreting the vegetal material as faeces. They contribute to an increase of substrate surface area accessible to microbial attack and to an increase of substrate pore volume and aeration, thus enhancing the overall microbial resource exploitation (Hassall *et al.*, 1987; Kayang *et al.*, 1996), and, ultimately, influencing nutrient mobilization rates in the system. As detritivores their diet consists mostly of decaying organic materials such as leaf litter, decayed wood, fungi, and bacterial mats. Much research has been devoted to consumption in woodlands and grasslands, and has shown that weathering of litter with conditioning by microorganisms improves its palatability to isopods (Hassall and Rushton, 1984; Hassall *et al.*, 1987). They can eat some animals and occasionally predate insect larvae. Coprophagy is used to improve their nutrient uptake especially in juveniles (Paoletti and Martinelli, 1981; Hassall and Rushton, 1985).

Most isopod species can be grouped in several life-forms, or eco-morphotypes, dictated mainly by body shape traits that are related to the type of habitat colonized (Fig. 1) (Schmalfuss, 1984, 1989):

a) Runners, which have large eyes, long legs, and sometimes mimetic colors. Known representatives of this group are species from the Ligidae family and also *Philoscia muscorum*.
b) Rollers, with a semi-circular body section, thus capable of rolling into a tight ball when disturbed. Known representatives of this group are most of the species of the Armadilidae family (e.g. *Armadilium vulgare*).
c) Clingers, less mobile than the preceding forms and with depressed

margins of the body that they press down on flat surfaces. Species from the genus *Porcellio* (e.g., *P. Spinicornis and P. Scaber*) and also *Porcellionides pruinosus*, are known representatives of this life-form group.

d) Creepers, which have developed tergal ribs and live in narrow interstices, caves, etc. Many species from the Trichoniscidae family belong to this life form.

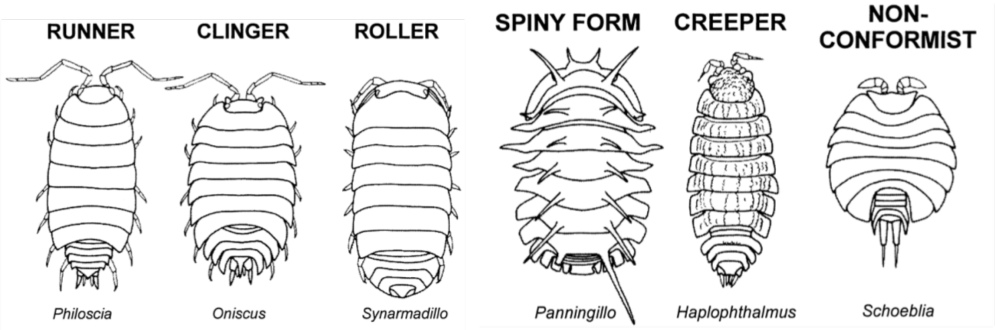

Figure 1: Isopod life forms according to (Schmalfuss, 1984)

Although isopod species can be divided into these ecomorphotypes, some life-traits like water loss rate, body size and vertical stratification are phylogenetically conserved and common across different life form groups (Berg, personal communication). Most species from Trichoniscidae family (mainly "creepers") and also *Porcellionides pruinosus* (a "clinger" species) are known to live also in soil layer (mostly down to 5cm), whereas most of all species from other families are typical litter species.

Isopod diversity and abundance is highly conditioned by vegetation composition, not only because of food quality (Souty-Grosset *et al.*, 2005) but also because of habitat structure, since soil coverage helps to retain moisture and to maintain a favourable habitat (Szlavecz, 1995; David *et al.*, 1999; Souty-Grosset *et al.*, 2005). Although more common in forests, they can reach moderate to high abundance and diversity in grasslands (up to 1000 individuals /m2 in calcareous grasslands), permanent crops and abandoned agricultural fields. In general, they are beneficial because of their role in enhancing nutrient cycling, by comminution of organic debris and transporting it to moister microsites in the soil. They also transport propagules of bacteria, fungi and vescicular arbuscolar mycorrhizae through soils (Paoletti and Hassall, 1999).

It has been observed that the specific diversity and abundance of terrestrial isopods decrease in intensive agricultural systems, with particularly marked differences between organically managed and more conventional plots. Herbicide application leads both to increased mortality and lowered fecundity (Paoletti and Hassall, 1999). These products reduce available food and can change soil pH (Van Straalen and Verhoef, 1997), also an important soil parameter influencing isopod distribution. Soil pH determines the distribution of many species on a micro as well as a macro scale. In addition to natural limits caused by low soil pH, anthropogenic soil acidification changes the community structure of the soil fauna (Abrahamsen, 1983; Lohm *et al.*, 1984; Kopeszki, 1992; Dmowska, 1993; Zimmer, 2000). Species-specific data on pH preferences or ranges of suitability are rare, but several authors reported ecophysiological (Zimmer and Topp, 1997a; Zimmer and Topp, 1997b; Kautz *et al.*, 2000) and behavioural (Sastrodihardjo and Van Straalen, 1993)responses of isopods to different pH levels of food or soil.

Closely related to soil pH is the soil calcium content. At low soil pH the availability of Ca and Mg will strongly diminish, even across small differences in pH (Berg *et al.*, 1997; Berg and Hemerik, 2004). The need of isopods and millipedes to accumulate Ca and Mg for their exoskeleton makes them vulnerable to low availability of these minerals (Hopkin and Read, 1992; Berg and Hemerik, 2004). This is the reason why isopods are usually more abundant in calcareous than acid soils.

But calcium is sufficiently available to meet the requirements of isopods and diplopods in many soils. The correlation with calcium could then be due to preferences for other site characteristics, like temperature or moisture (Thiele, 1959; Dunger, 1983) which are also known to affect isopod and diplopod distribution strongly (Haacker, 1968; Warburg *et al.*, 1984).

2.1.4 Selection criteria for the model countries

Model countries were selected based on the coverage of different biogeographical regions in Europe. Selected countries should maximize the differences in climate and soil properties, therefore containing different soil fauna communities, while also having available data for the selected soil organism groups in published papers and databases.

EFSA (2010b) made the following assumptions for country selection:

- Ecoregions differ with respect to presence and abundance of characteristic species of the selected taxa.
- Ecoregions have different relative number of species in different life forms within one taxon.
- Important ecological functions in different ecoregions are carried out by different taxa.
- The vertical distribution of life forms differs between ecoregions.

The countries selected were Finland (representative of the Boreal region), Germany (representative of the Continental region), and Portugal (representative of the Mediterranean region), to achieve the best coverage of the North-South gradient in Europe.

2.1.5 Ecologically Relevant Exposure Scenarios (ERES) assumptions

For the construction of the ERES, the following assumptions were made:

- Species are exposed to PPPs according to their traits.
- The exposure pathway of species having the same combination of traits ("trait group") is similar, BUT the actual exposure is different as it is mainly influenced by soil properties and climate.
- Ecological data (dominance of organisms and trait groups) will help define exposure scenarios taking into account:
    - If exposure in the litter layer needs to be modeled
    - In which soil depth exposure is to be determined (0- 1 cm, 0-20 cm or in-between).

    - If burrowing activity of important soil organisms has to be considered.

3 Methodology

3.1 Soil organisms' databases

In 2008, EFSA started to setup databaseswith information about the occurrence of collembolan, earthworms, entrichaeds and isopods in the three model countries in order to prepare a base for the definition of ecoregions all over Europe (EFSA, 2009; Sousa *et al.*, 2009).For each organism group considered, one database per country was built. Each database shares a similar structure composed of four sections:

- Section 1 – Site information, containing data on site location (geographical coordinates), land-use type and dominant vegetation;
- Section 2 – Soil type information, containing data on major soil properties;
- Section 3 – Species information, containing the taxonomic data for each species, abundance or density, and the sampling method used to collect the data. In this section the information related to the life form type is also included. This was defined using morphological and ecological traits available in published material and/or in trait databases.
- Section 4 – Bibliographic references, containing the complete information for all references included in the database.

A detailed description of the database can be found in Annex 1. Permission from EFSA was granted to use the databases for this study.

The following life-form groups were defined:

a. Collembola

Based on the 3 morphological traits, five life-form classes were defined in the database. However, for data analysis these were grouped into the 3 classes described.

- Life form 1: Euedaphic species with very low dispersal ability, living down to 5 cm

- Life form 2: Hemiedaphic (medium dispersal) species, living down to 2.5 cm
- Life form 3: Epigeic (fast dispersal) species, living at the soil surface

b. Isopods

- Life form 1: Soil dwellers, species living mostly in the soil surface, but that are able to burrow down to 2.5 cm depth
- Life form 2: Litter dwellers, species living mainly on the soil surface, particularly in the litter layer

The databases compile information gathered from published and some unpublished articles, species catalogs, and regional inventories. Different scientific papers reporting the same type of taxonomic information for the same sites were not considered. For this study, new papers were reviewed and included to the original EFSA (2010) ecoregions opinion databases.

For Germany a good coverage was achieved for collembolans and isopods. For Finland and Portugal a generally good coverage of the published literature was achieved for the organism groups. There were data sets not included due to doubts on their quality.

It was not possible to complete all information for every data entry in the database, mostly related to the nature of papers analyzed. Many were taxonomic papers containing no precise information on the geographical location, soil properties or on the land-use where the biological material was collected. Also, this limits the database analysis to a presence-absence, as information on the population abundance is not available for most of the papers collected.

To fill-in the data gaps, EU maps from the European Commission Joint Research Center (JRC) were used. The maps contain important environmental variables to predict soil community occurrence in the model countries. The variables considered are:

- Average annual temperature (in Centigrade)
- Total average precipitation (mm/year)

- Texture (coded by percentage of silt, clay and sand)
- Organic matter (g/g)
- pH
- Land use (coded with CORINE system)
- Bulk density of topsoil (kg/m^3)
- Water content at field capacity (m^3 m^{-3})

The codes used can be seen in Annex 2.

These data describe Europe on a 1-km2 scale and were linked through the site UTM (Universal Transverse of Mercator) coordinates to the biogeographical database. Missing coordinates in the biogeographical database were filled-in by deriving coordinates from the given name of the site (region, village/town, name/place, and additional site info).

It's important to point out that for most records the local scale of biological sampling is much smaller than the extent of the site. Therefore, the records in the biogeographical database that correspond to the same set of UTM coordinates and equal land use were assumed to originate from one site. This sometimes combines samples from several locations within one site. Some UTM coordinates specify a grid cell in databases where parameters were not available. Sites with incomplete environmental variables were not included in the statistical analysis.

3.2 Dominant life form class classification

For each site the number of different species per life form group was counted. The percentage of a life form group in relation to the total number of species defines the raw relative richness of that life form group on this site.

Dominant communities per site were calculated according to Figure 2:

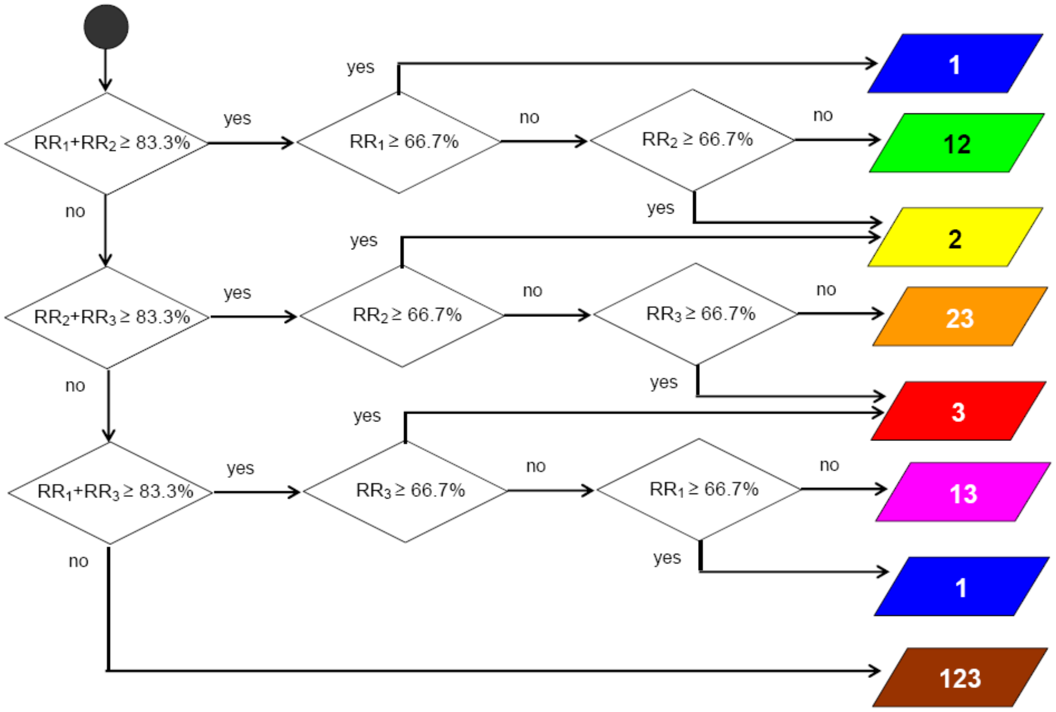

Figure 2: Categorization rule of the relative richness (RR) into dominance classes of three different life forms for collembola (called 1 for euedaphic, 2 for hemiedaphic, and 3 for epigeic in this graph) or their respective combinations (12, 23, 13, and 123).

The class dominance can be visualized in a triangle in figure 3:

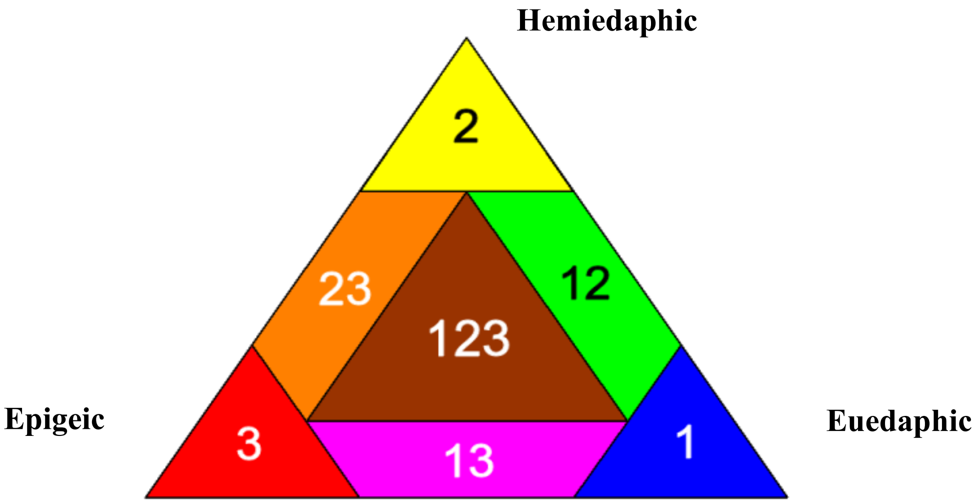

Figure 3: Categorization of the relative richness of three different life forms of Collembola into dominance classes (called 1, 2, and 3) or their respective combinations (for example, 12 for a euedaphic and hemiedaphic dominated community, 23 for hemiedaphic and epigeic, and 123 for codominance).

3.3 Statistical Analysis

The use of Generalized Linear Models (GLM) and spatial prediction for species prediction is well documented. Many other authors (Guisan, 2000; Higgins and Richardson, 2001; Guisan and Edwards, 2002; Miller, 2002; Hengl *et al.*, 2009) are using GLMs and GIS spatial prediction tools to improve the accuracy of point-measurement spatial predictions. Lehmann and Overton (2002) conceptualized this approach and named it generalized regression analysis and spatial prediction (GRASP). But the process of analysis is similar: Regression modeling is used to establish relationships between a response variable and a set of spatial predictors, the regression relationships are then used to make spatial predictions of the response. The process requires point measurements of the response, as well as regional coverage of predictor variables that are statistically (and preferably causally) important in determining the patterns of the response. This approach to spatial prediction is becoming more commonplace, and it is useful to define it as a general concept(Lehmann and Overton, 2002).

Since the ultimate goal is to produce simulations that permit the creation of more realistic scenarios, the applicability of a Stochastic Dynamic Methodology (StDM) was tested. The StDM proposed is a sequential modeling process initiated by a multi- variate conventional procedure. However, the fact that the data we considered consisted of n independent variables, does not automatically imply that all variables have a significant effect on the magnitude of the dependent variable. Therefore, the regression model with the maximum likelihood was selected using the Akaike Information Criteria (AIC) (Akaike, 1974). The AIC measures a trade-off between a small residual sum of squares (goodness- of-fit) and model complexity (number of parameters). The software GenStat was used to run the regressions for all life forms except isopods LF1 (analyzed with Statistica 7), using a Poisson distribution to fit the big number of zeros in the raw life-form type numbers per site.

For the development of the model, the software STELLA 9.0.3 was used. The information in the databases is static in time, but the flexibility of the software allows for dynamic variables to be added (such as agricultural management changes, mean temperature per month) and generate predictions for time series. In this case, the environmental variable information for each site was used as “time” (site1 = t1, site2

= t2…) to allow all data to be analyzed at the same time, simplifying the data analysis process.

The results of the model simulations were later incorporated in ArcView 9.2 using the spatial analyst and geostatistical analysis extensions. The spatial analyst extension performs cell-based raster data calculations, the map algebra fuction gives the flexibility to build complex expressions and execute them as a single command. This allows for all significant environmental variable maps to be taken into account in one single operation and relies on a good GLM for optimal results. The geostatistical analyst extension examines spatially dependent data and predicts values where no information is known, creating a continuous surface with Ordinary kriging and using a spherical semivariogram model, which adds weights to the measured data and assumes a progressive decrease of spatial autocorrelation between the observations.

Kriging has been used as a synonym for geostatistical interpolation for many decades; it originated in the mining industry in the early 1950's as a means of improving ore reserve estimation. The idea came from the mining engineers D. G. Krige and the statistician H. S. Sichel. The technique was first published in Krige in 1951, but it took almost a decade until a French mathematician G. Matheron derived the formulas and basically established the whole field of linear geostatistics (Cressie, 1990; Webster and Oliver, 2001; Zhou et al., 2007 quoted by Hengl, 2007).

Ordinary kriging predictions are based on the model:

$Z(s) = \mu + \varepsilon'(s)$

where μ is the constant stationary function (global mean) and $\varepsilon'(s)$ is the spatiallycorrelated stochastic part of variation. The predictions are made as in the equation:

$$\hat{z}_{\mathtt{OK}}(\mathbf{s}_0) = \sum_{i=1}^{n} w_i(\mathbf{s}_0) \cdot z(\mathbf{s}_i) = \lambda_{\mathbf{0}}^{\mathbf{T}} \cdot \mathbf{z}$$

where λ0 is the vector of kriging weights (wi), z is the vector of n observations at primary locations(Hengl, 2007).

4 Results

A total number of 156 points were used for the collembola model, and 600 for the isopod model.

The sites' location is displayed in Figure 4.

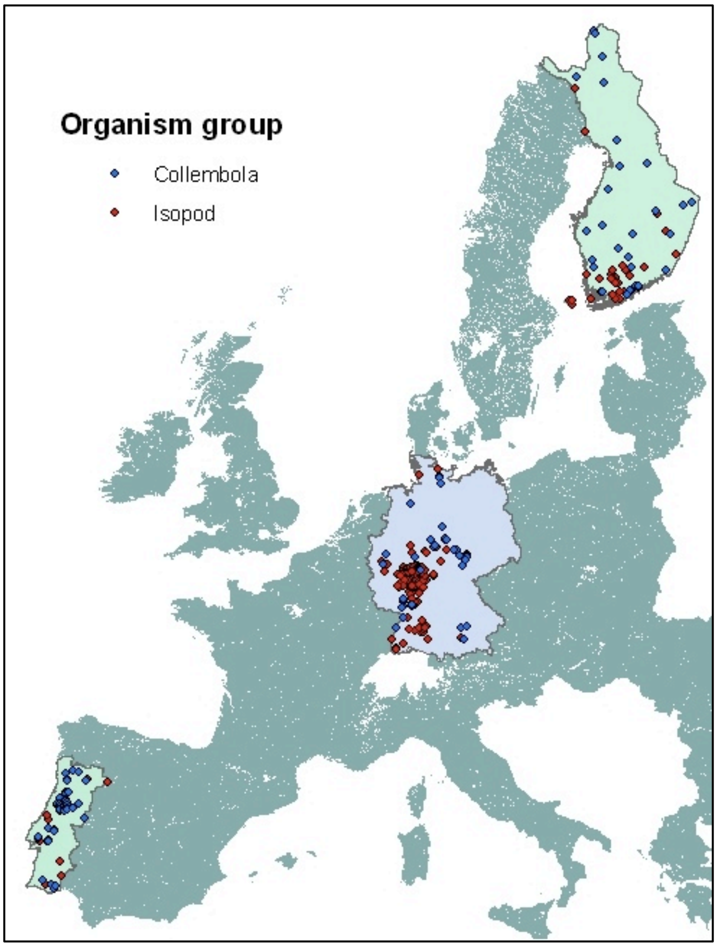

Figure 4: Site location maps

Points that were missing one or more environmental variables were removed from the analysis. Raw richness, richness-weighted per country and raw-richness percentage numbers were used in the prediction models for both organism groups, although the richness-weighted per country had a roughly higher R^2 number in the case of collembola, the resulting GLM model increased the number of expected eudaphic species and only one dominance class was predicted. The same was observed with the percentage numbers. Therefore, raw richness was selected for the analysis.

The proportion of land use per point by country is summarized on table 3 and 4:

Table 3: Total points per land use by country in collembola database

Country	**Land use**				**Total**
	Crop area	**Grassland**	**Shrub**	**Forest**	
Portugal	9	0	3	53	65
%	13.85%	0%	4.61%	81.54%	41.67%
Germany	15	6	0	28	49
%	30.61%	12.25%	0%	57.14%	31.41%
Finland	9	3	2	28	42
%	21.43%	7.14%	4.76%	66.67%	26.92%
Total	33	9	5	109	156
%	21.15%	5.77%	3.21%	69.87%	

Table 4: Total points per land use by country in isopod database

Country	**Land use**				**Total**
	Crop area	**Grassland**	**Shrub**	**Forest**	
Portugal	4	0	2	3	9
%	44.44%	0%	22.22%	33.33%	1.50%
Germany	169	73	0	314	556
%	30.40%	13.13%	0%	56.47%	92.67%
Finland	17	1	1	16	35
%	48.57%	2.86%	2.86%	45.71%	5.83%
Total	190	74	3	333	600
%	31.67%	12.33%	0.5%	55.5%	

The selected environmental variables per organism group and life form are summarized in table 5. The combinations of variables with the lowest AIC were selected to perform a Generalized Linear Model. A Poisson distribution was used and values were converted to log.

Table 5: Selected variables by organism group and life form

Organism **Life form**	**Selected variables**	**AIC**	**R2**
Collembola LF1 (Euedaphic)	Average annual temperature, Latitude, Texture	156.54	14.43
Collembola LF2 (Hemiedaphic)	Average temperature, Latitude	154.86	17.17
Collembola LF3 (Epigeic)	Latitude, Texture	157.43	7.25
Isopod LF1 (Soil dweller)	Country, Land use, Organic matter, Texture, Water content at field capacity	1553.426	
Isopod LF2 (Litter dweller)	Average temperature, Land use, Organic matter, Country	603.61	1.26

The results of the GLM are shown on table 6. The regression for isopod life form 1 was ran in different statistical software

Table 6: GLM results for soil organisms' life forms (Poisson/Log)

Organism lifeform	**Regression DF**	**Regression Deviance**	**Residual Deviance**	**Deviance ratio**	**Chi probability**
Collembola LF1	3	70.3	366.5	23.42	<0.001
Collembola LF2	2	169.9	735.4	84.97	<0.001
Collembola LF3	2	84.7	918.4	42.36	<0.001
Isopod LF1*	5	829.74 (null)	762.21	67.53	NA
Isopod LF2	4	11.7	599.8	2.93	0.02

The next step was to include all relevant variables and construct the models in the STELLA software. The main advantage of the StDM models is that they take into account all interactions between variables with all their combined contributions, so there is no danger of overfitting. The more significant variables included, the more successful the models predictions will be. Figures 5 and 6 represent the STELLA models. A sample dynamic variable was included in both models to demonstrate the software capability to include changes in time, although it is not relevant for the scope of this report. The code is included in Annex 3.

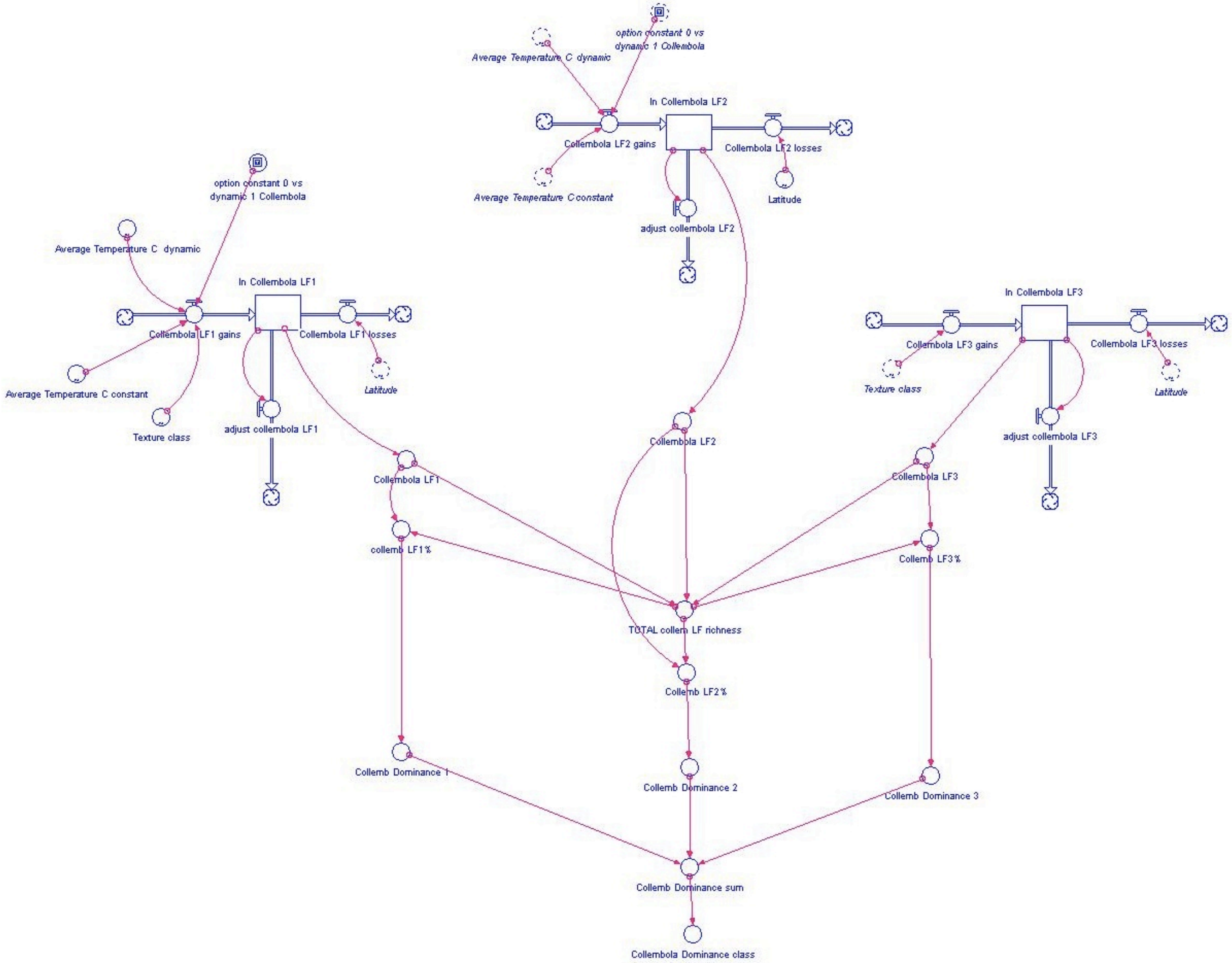

Figure 5: STELLA collembola model

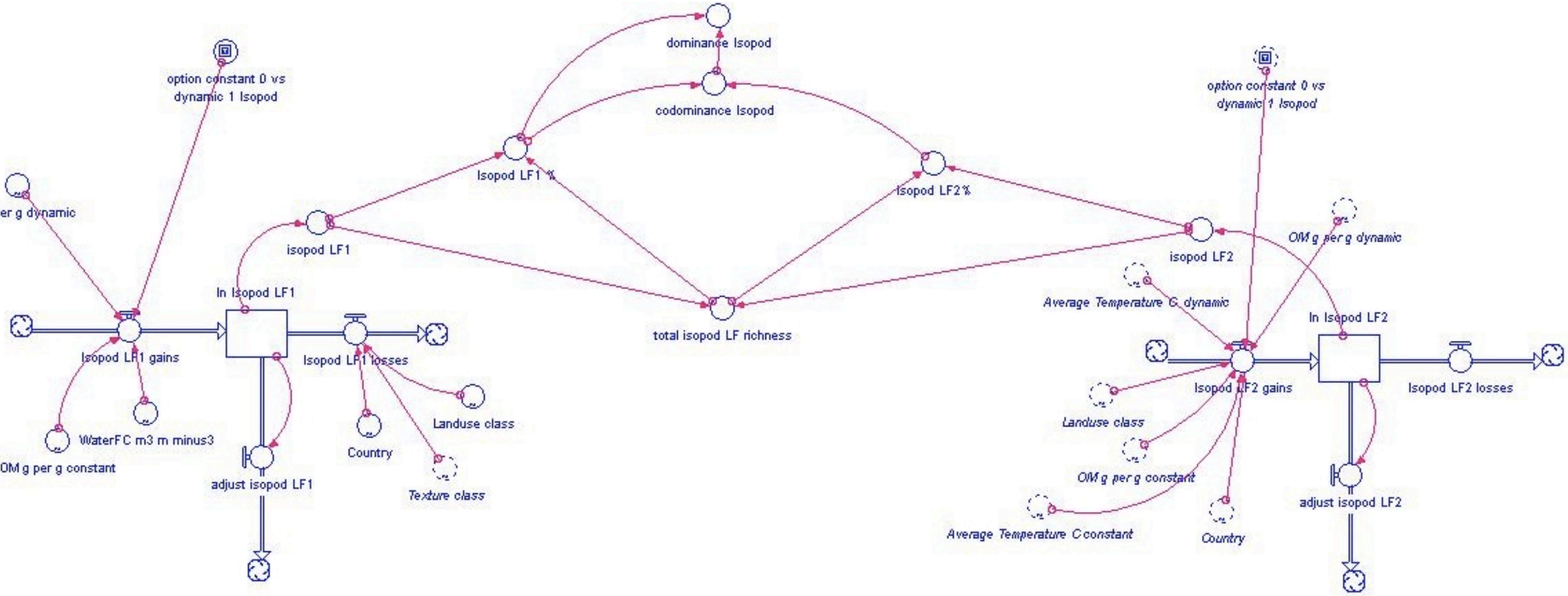

Figure 6: STELLA isopod model

The models were not more successful than the previous analysis by EFSA (2010b), since collembola were still predicted to be dominated by hemiedaphic and epigeic species (class 23), or by all three groups (class 123). The observed versus predicted are depicted in table 7, and table 8 shows a comparison by country.

Table 1: Collembola class dominance Observed vs. Predicted comparison

Dominance class	Observed	Predicted
1	1	0
12	5	0
123	46	29
13	3	0
2	3	0
23	84	127
3	14	0
Total	156	156

Table 2: Collembola class dominance Observed vs. Predicted by country

Country	Dominance class	Observed	Predicted
Portugal	12	1	0
	123	13	0
	2	1	0
	23	48	65
	3	2	0
Subtotal Portugal		65	65
Germany	1	1	0
	12	1	0
	123	21	19
	13	2	0
	23	14	30
	3	10	0
Subtotal Germany		49	49

Finland	12	3	0
	123	12	10
	13	1	0
	2	2	0
	23	22	32
	3	2	0
Subtotal Finland		42	42
Grand Total		156	156

For the isopods, the model was ill fitted and was not able to predict any co-dominance, giving a heavy preference to litter dweller species., the comparison is summarized on table 9.

Table 3: Isopod class dominance Observed vs. Predicted comparison

Dominance Class	**Observed**	**Predicted**
1	117	21
2	408	579
12	75	0
Total	600	600

Despite the higher number of points included in the analysis, factors that could explain the inferior quality of the model predictions include non-considered variables like evapotranspiration (which was significant in the EFSA 2010b study, but the map was not available for this study), the consideration of texture as a categorical value instead of a continuous one or other environmental variables not tested. As the PPR panel stated in their report, the 1-km^{2}spatial resolution can also be an explanation for poor fit, given the organisms's group dispersal potential and different community composition in the same space unit.

Another key consideration for both methodological approaches is the use of presence/absence data and more importantly, the inclusion of taxonomical studies that provide the location of a single specie/life form.

Regardless of the models shortcomings, an advantage of using STELLA is its ability to predict collembola and isopods frequencies simultaneously with both databases environmental variables. This opens the possibility to obtain a higher number of sites to run geostatistical methods in ArcGIS software in order to spatially predict occurrence of soil organism communities for risk assessment. However, the predictions are only based on GLM variables.

Country predictions were obtained with two methodologies in ArcGIS 9.2:

- Spatial Analyst: Countries were extracted from the JRC Europe maps with extract by mask. Cell-based analysis was then performed with Raster calculator using the GLM equations only (no site information) and the single output map algebra was executed to generate the dominance maps. The codes are included in detail in Annex 5.
- Geostatistical Analyst: Geostatistical wizard, ordinary kriging model with a spherical semivariogram,using the predicted STELLA percentage raw numbers and the EFSA database percentage raw numbers for validation. When more than one point were available in one location (coincidental samples), the setting ‘use maximum’ was chosen to highlight the dominant community. An alternate analysis using the mean value was performed, and the results for the dominant class maps were the same on country scale.

4.1 Soil organisms' maps

4.1.1 Collembola maps

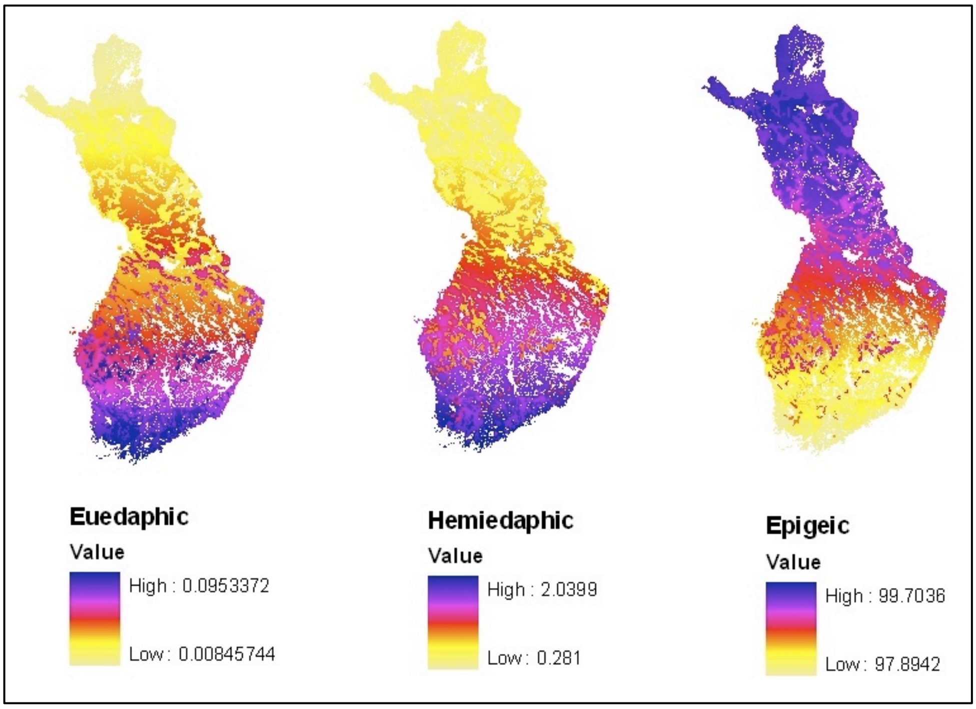

Figure 7: Predicted collembola life forms distribution for Finland in percentage (Raster calculator)

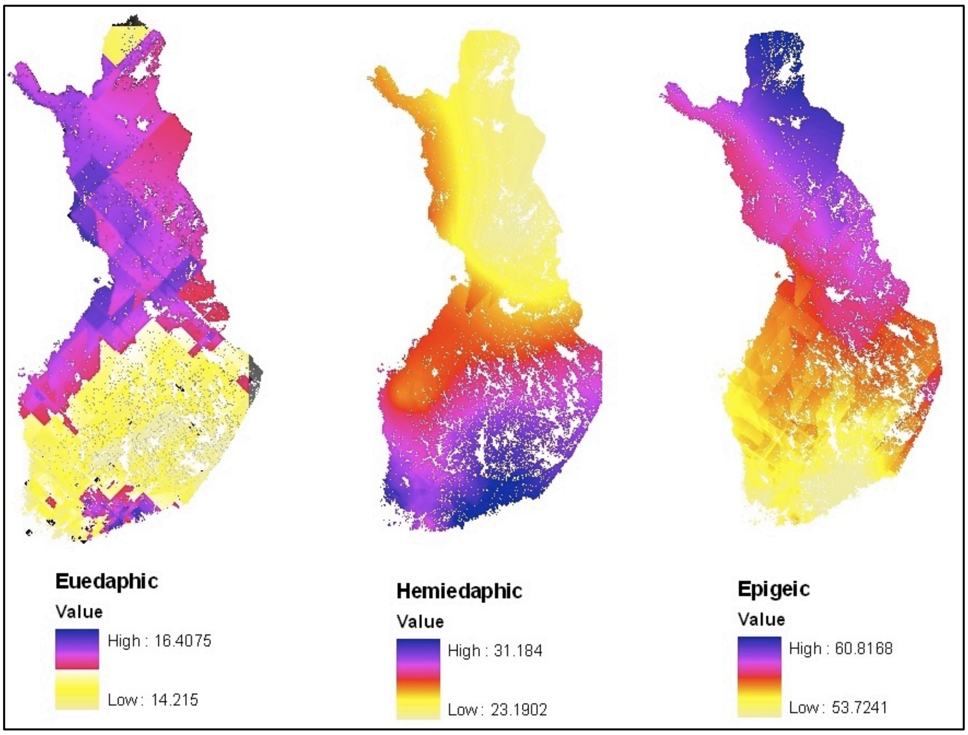

Figure 8: Predicted collembola life forms distribution for Finland in percentage (Geostatistical Analyst - Kriging)

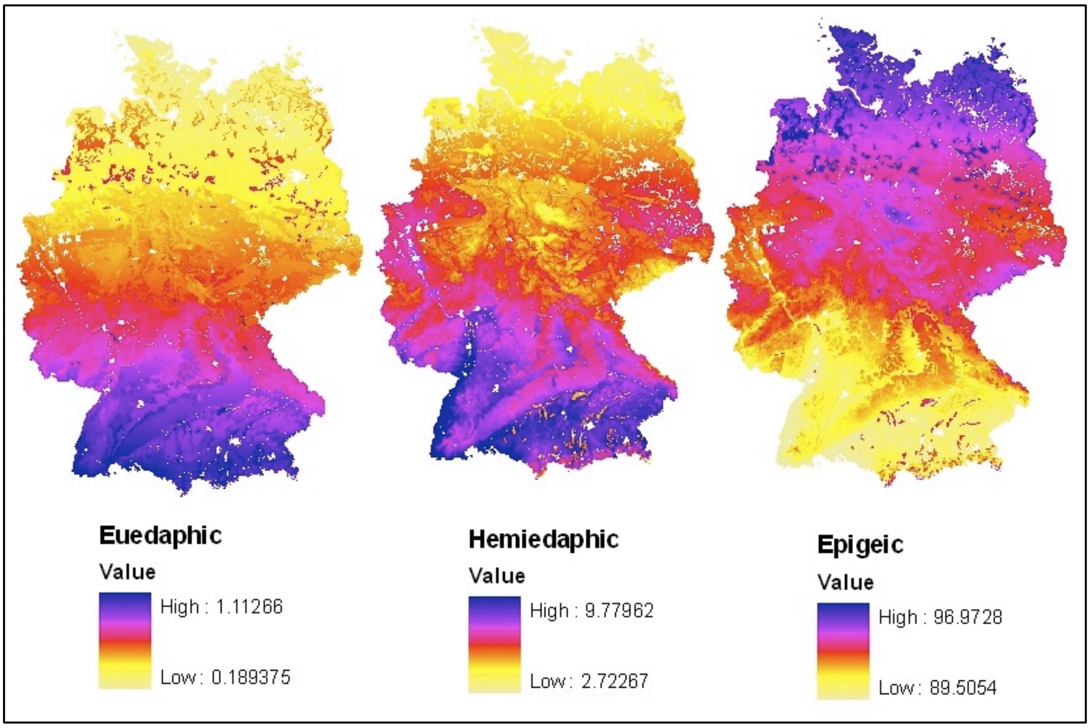

Figure 9: Predicted collembola life forms distribution for Germany in percentage (Raster calculator)

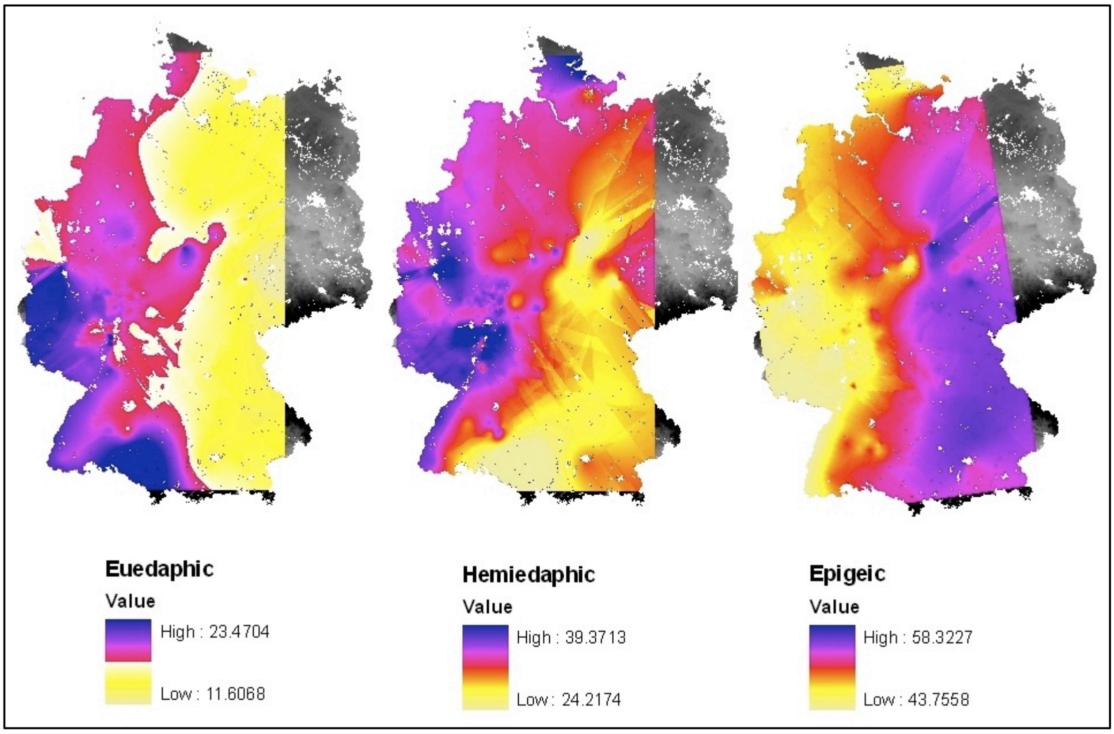

Figure 10: Predicted collembola life forms distribution for Germany in percentage (Geostatistical Analyst - Kriging). Non-predicted surface in gray.

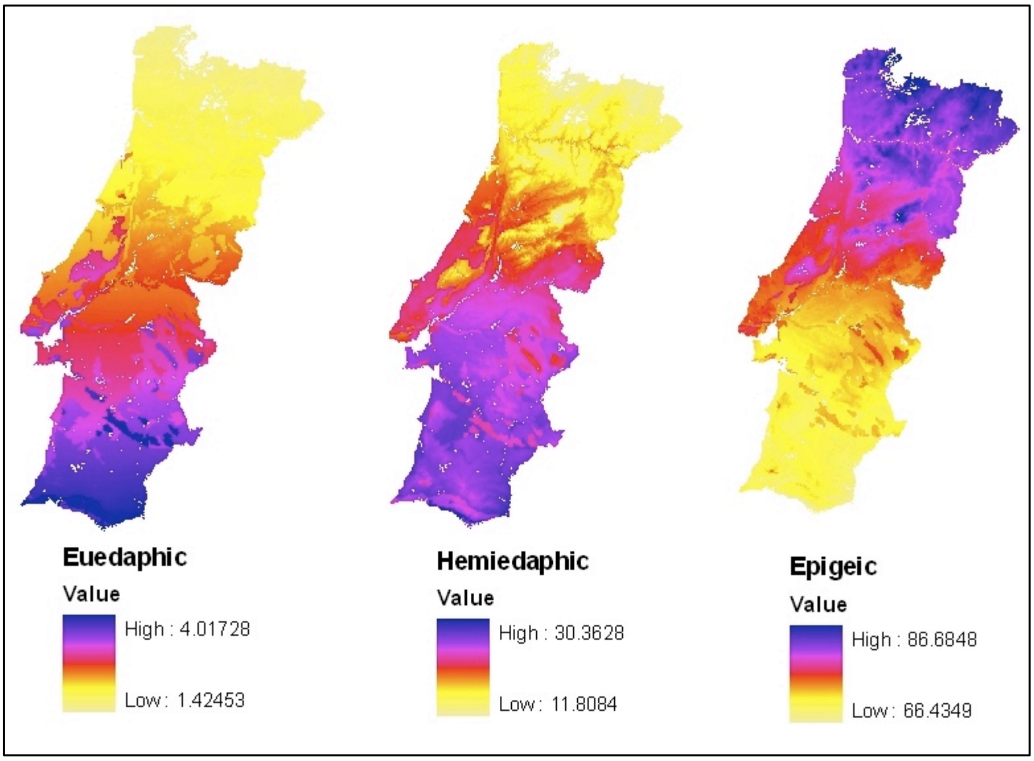

Figure 11: Predicted collembola life forms distribution for Portugal in percentage (Raster calculator)

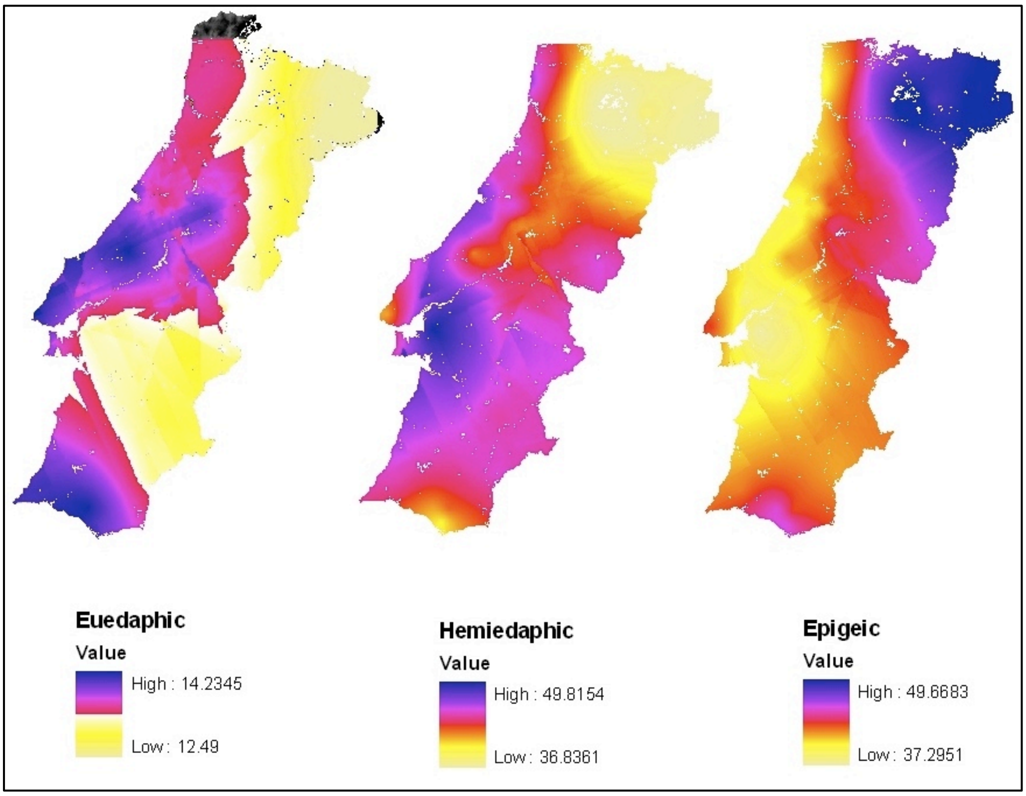

Figure 12: Predicted isopod life forms distribution for Portugal in percentage (Geostatistical Analyst - Kriging)

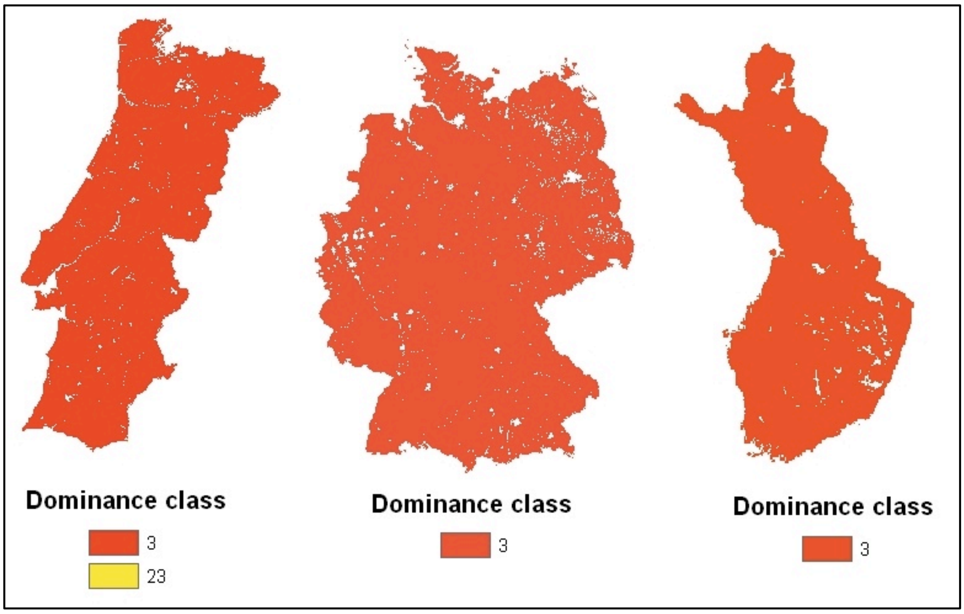

Figure 13: Collembola dominance class distribution by country (Raster calculator)

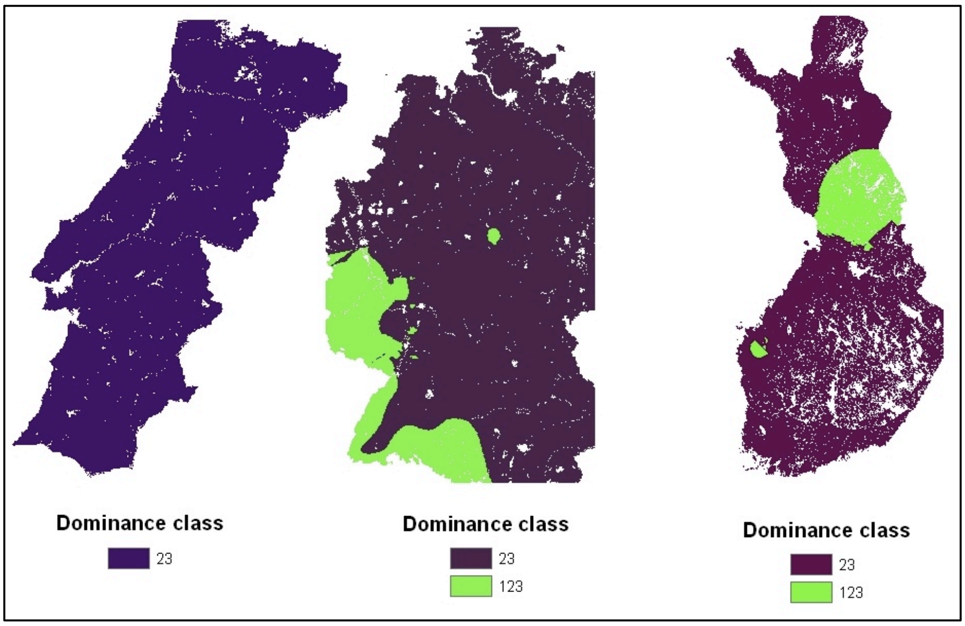

Figure 14: Collembola dominance class distribution by country (Geostatistical Analyst – Kriging)

The maps show the expected north-south gradient except for the Germany geostatistical analyst map, and the prediction for each life form percentages varies with the method used (raster calculator and kriging). They are based on the significant GLM variables average temperature, texture and latitude. There is a risk that the raster calculator method might have given too much weight to latitude when predicting distribution as the analysis in itself is not built to consider interactions between variables (not holistic) and the poor GLM fit and map resolution represent important limitations for predicting accurate life form distributions. The kriging-generated maps had different levels of goodness of fit, with hemiedaphic species adjusting best to the spherical semivariogram distribution, which means they have a stronger spatial dependency than the other life forms with the data considered.

For Finland, there was clear epigeic dominance in both models with the raster calculator predicting over 97% values throughout the country and kriging giving values between 53 and 60%. However, they differ greatly in the prediction of euedaphic and hemiedaphic species, the raster calculator estimating less than 1% of euedaphic versus 14-16% predicted with kriging. The same was true for hemiedaphic species, the predicted percentages increasing from 2% with the raster calculator to 23-31% with the geostatistical analyst.

In the case of Germany, the predictions were radically different. The spatial analyst method suggests a north-south gradient while the geostatistical analyst predicts an east-west gradient. This could be a result of the methods'calculating approaches, the influence of latitude in the raster calculator method and the position of the sites with kriging (concentration of sites in West Germany). The life form abundances were also very different between methods: the spatial analyst showed clear dominance of epigeic species (89-96%) with very low occurance of euedaphic and hemiedaphic species mostly in the south. Conversely, the geostatistical analyst predicted epigeic species percentages from 43 to 58% with the highest ones in the east, and hemiedaphic ones from 24% to 39% with higher values in the west. Euedaphic species' predicted percentages were higher in the kriging map, from 11 to 23% versus 1% with the raster calculator.

Portugal spatial analyst's maps indicated an epigeic species' dominance (values from 66% to 86%), low values for euedaphic species (1.4-4%) and medium values for hemiedaphic (11.8 – 30%). The geostatistical analyst map predicted a co-dominance of hemiedaphic and epigeic species with a north-south gradient, with hemiedaphic dominance in the south and epigeic in the north. Euedaphic species predicted values varied between 12 and 14%, being more abundant in the south.

The dominance class distribution maps resulting from both methodologies were not useful and even contradictory. They were obtained by adding the values of the raster cells (1 km^2) from the life form distribution maps using the single output map algebra. The previous decision of using maximum values when more than one point was found at a specific location for kriging might have an influence in the class dominance prediction in smaller resolution scales, but on the country level the predictions were the same. Only classes 23 (hemiedaphic + epigeic) and 123 (codominance) predicted, therefore, the resulting worst-case scenario is Litter to 1 for the modeled countries.

In general, the low number of predicted euedaphic species could be related to differences in land use (as they are more important in crop areas) and the higher number of forest sites. Euedaphic species also have low tolerance to drought, but the best GLM fit for the collembola data did not include land use or any water-related variables (like precipitation, evapotranspiration or water content at field capacity).

4.1.2 Isopod maps

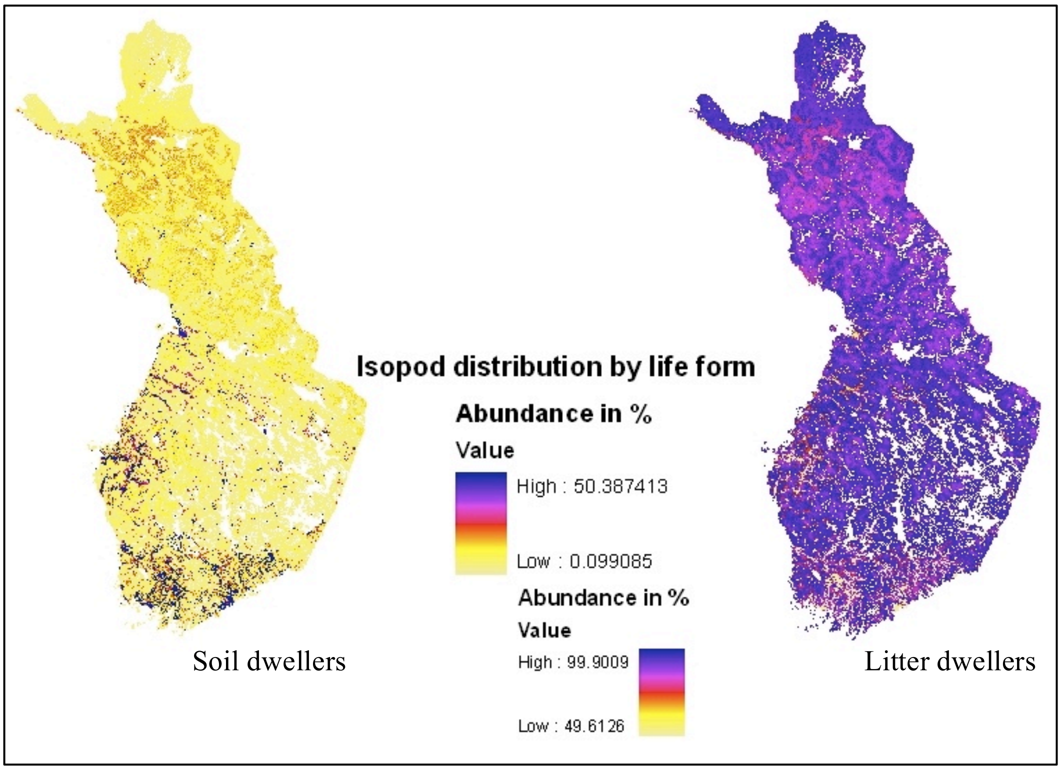

Figure 15: Predicted isopod life forms distribution for Finlandin percentage Raster calculator)

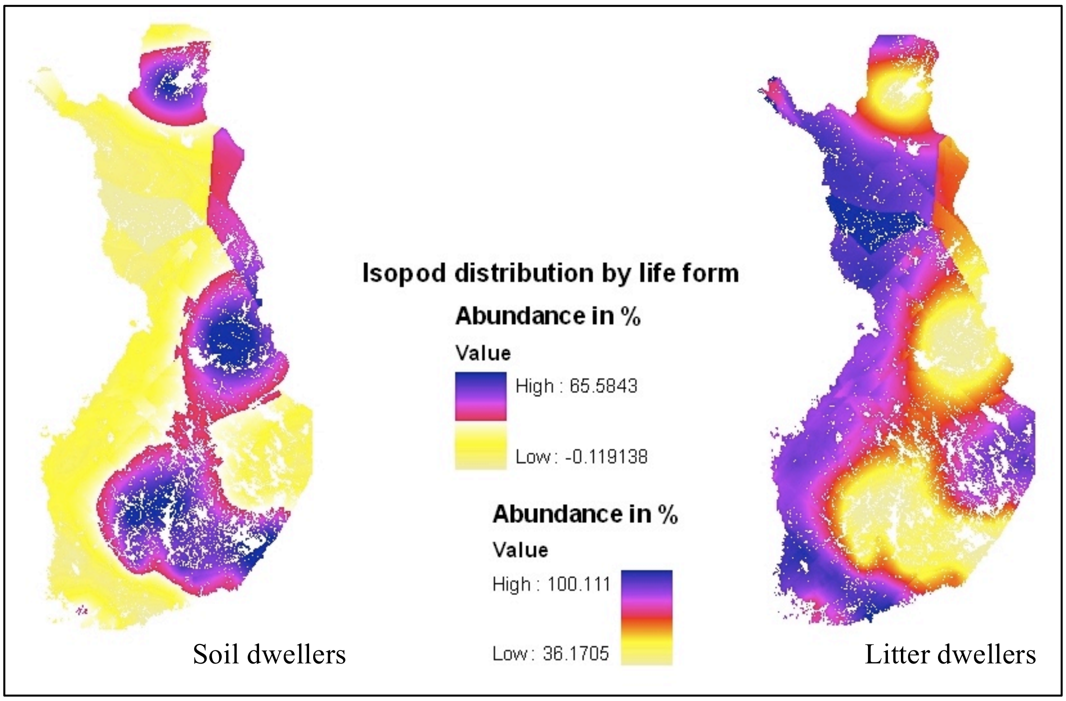

Figure 16: Predicted isopod life forms distribution for Finland in percentage (Geostatistical Analyst - Kriging)

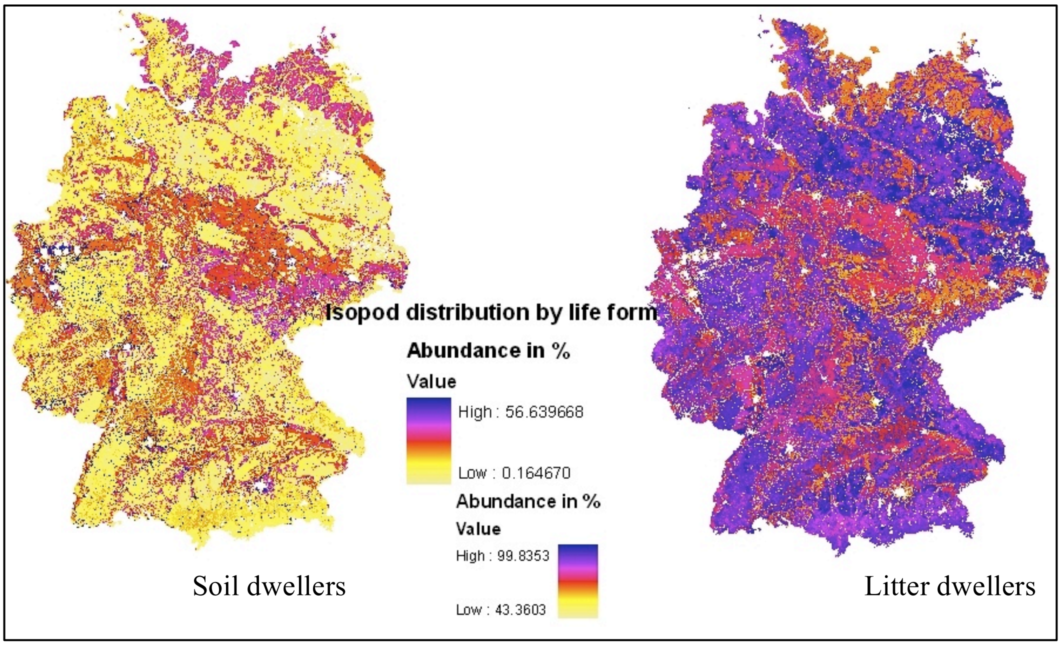

Figure 17: Predicted isopod life forms distribution for Germanyin percentage (Raster calculator)

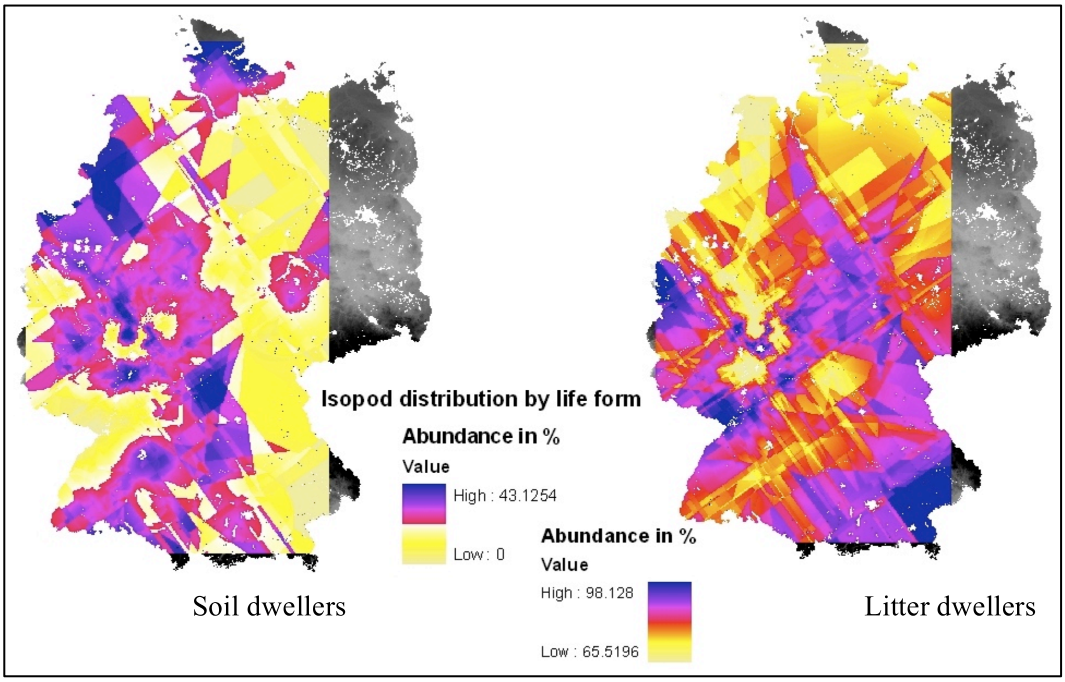

Figure 18: Predicted isopod life forms distribution for Germany in percentage (Geostatistical Analyst - Kriging). Non-predicted surface in gray.

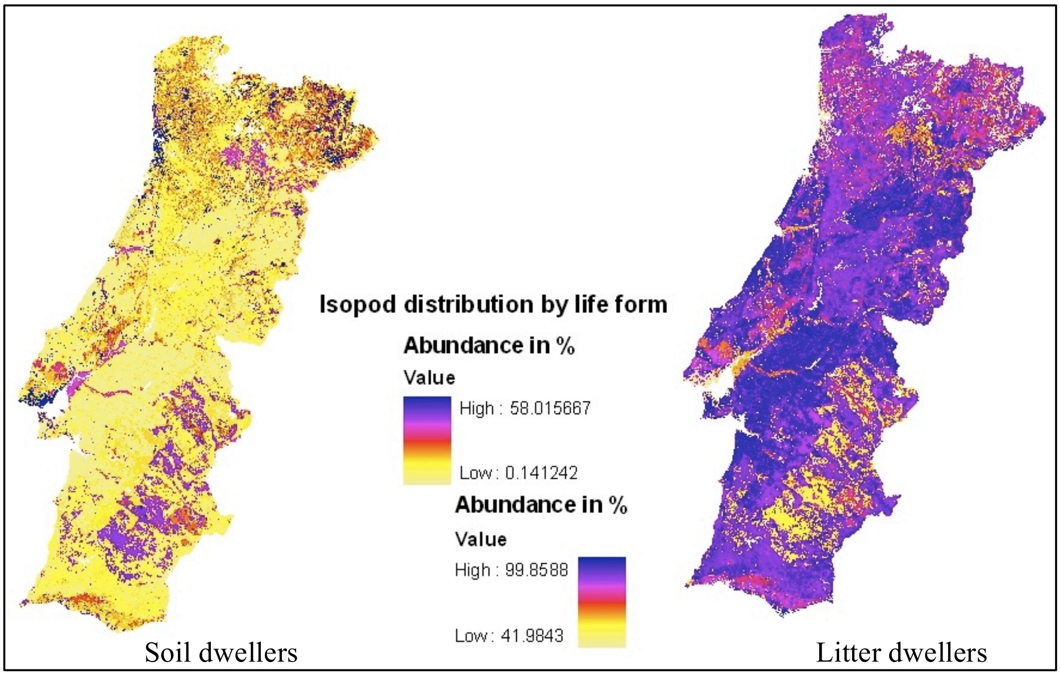

Figure 19: Predicted isopod life forms distribution for Portugal in percentage (Raster calculator)

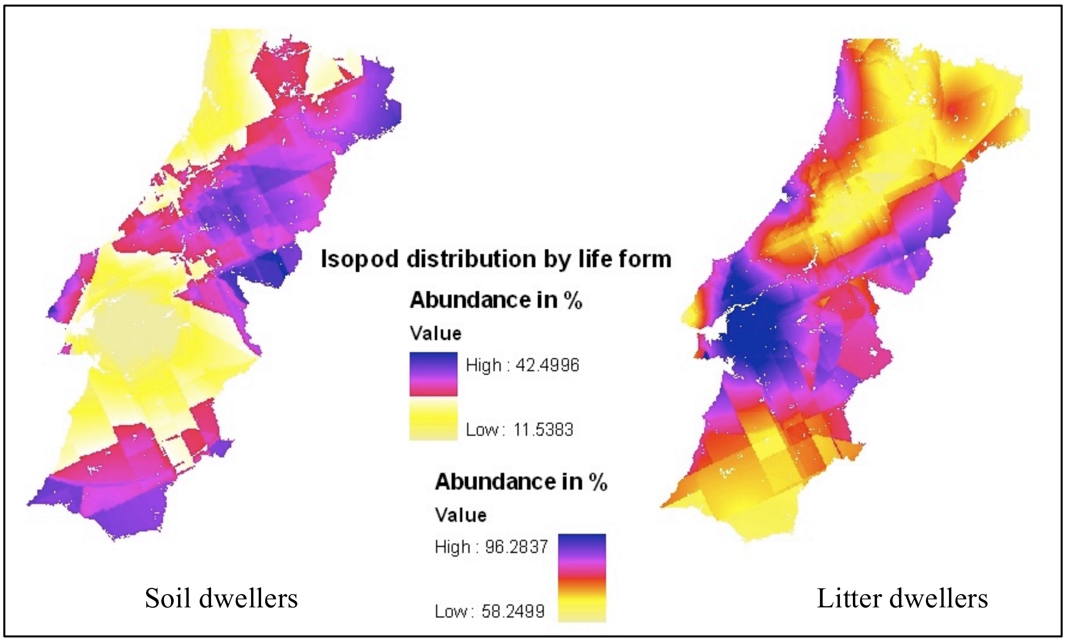

Figure 20: Predicted isopod life forms distribution for Portugal in percentage (Geostatistical Analyst - Kriging)

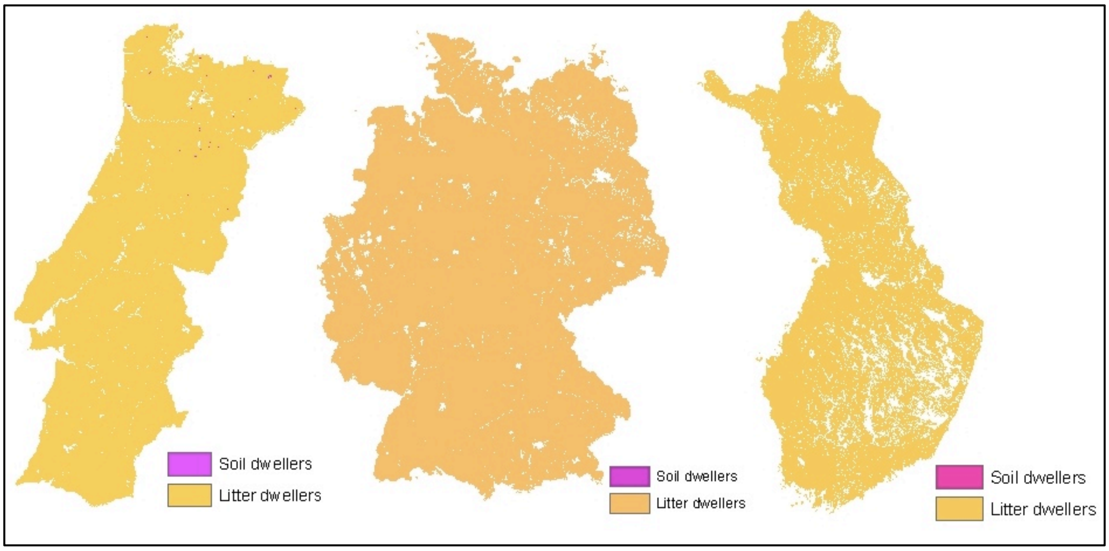

Figure 21: Isopod dominance class distribution by country (Raster calculator)

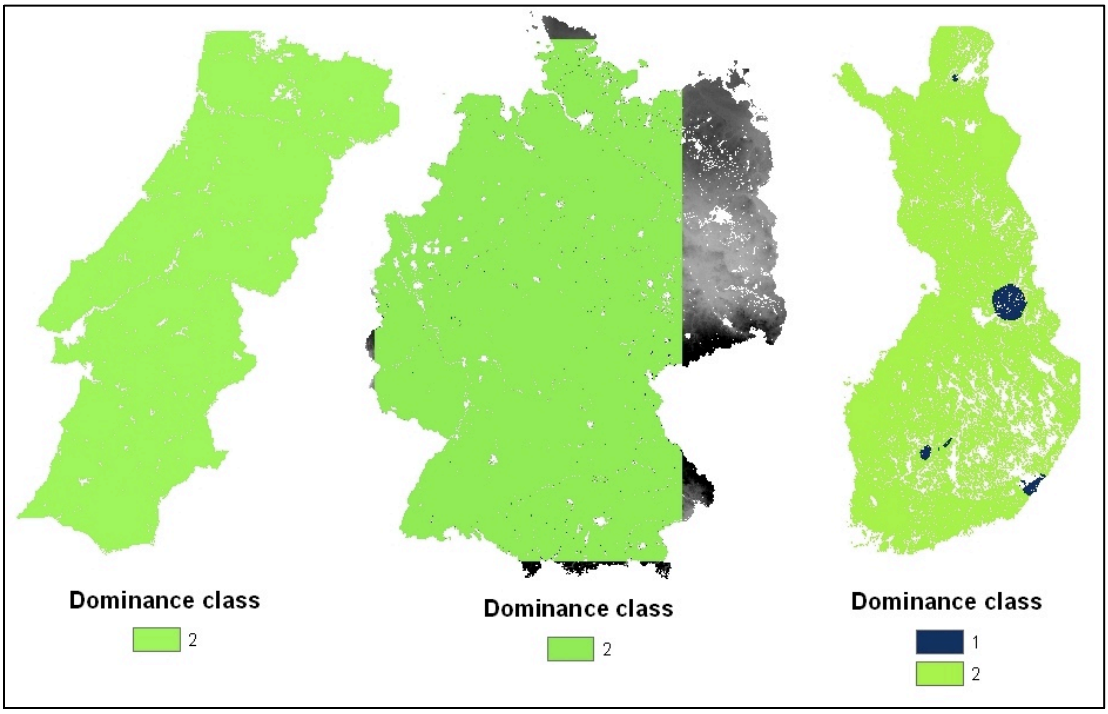

Figure 22: Isopod dominance class distribution by country (Geostatistical Analyst - Kriging). Non-predicted surface in gray.

The isopod distribution maps were based on the GLM significant variables Country, Land use, Organic matter, Texture, andWater content at field capacity. They did not show any particular dominance pattern, with all 3 countries dominated by litter dweller species. Kriging prediction maps showed medium to low spatial dependence, affecting the resulting distribution predictions.

Both methodologies produced similar values for the two life forms in all countries, on average from 0 to 50% for soil dwellers and from 50 to 100% for litter dwellers. Kriging prediction maps calculated an unexplained soil dweller dominance spot in Finland, maybe due to an abundance of sites with favorable conditions. Although the variables analyzed are consistent with ecological expectations of the isopod group, the inclusion of pH and calcium content could contribute improving the model fit. The resulting worst-case scenario is litter to 1 in all countries.

Overall, the results obtained from the spatial and the geostatistical analysts were not helpful to define ecoregions for pesticide risk assessment given the available data and the selected GLM variables, as they do not provide enough discrimination between worst-case scenarios.

5 Conclusions

The dominance maps obtained are not helpful for risk assessment use, as no discrimination was possible in the 3 model countries. The worst-case scenario for isopods and collembola is litter to 1 cm. for the entire area tested. The life form's distribution maps provide a glimpse of the probability of finding collembola and isopods in the 3 model countries. Nevertheless local scale predictions are not reliable because of the scale of the maps (1 km^2). For future studies, only sites with complete, locally measured environmental variables should be taken in consideration for model design to avoid mismatches of biological and environmental data, resulting from the extraction of environmental variables from European-scale maps. Data depuration is necessary to create a better-fitting generalized linear model and improve the prediction power of the GIS methods. Regardless, the models were in line with ecological and biogeographic information for the considered soil organism groups.

Albeit the possibilities that geostatistical analysis offers, a serious limitation is the analysis of categorical response variables. In this case the main objective is to display worst-case scenarios based on soil groups' depth distribution in soil, represented by dominance classes, but kriging prediction does not work well on categorical variables. Hengl (2007) advises against using indicator kriging as it leads to many computational problems, which probably explains why there are not many operational applications of geostatistical mapping of categorical variables in the world (Hession *et al.*, 2006; Hengl, 2007) and recommends the use of regression-kriging to develop map predictions. This analysis cannot be performed in ArcGIS and requires expertise in other GIS software and/or R.

Future studies looking to build up on this effort should consider including only site data with complete environmental variables information (such as soil texture, pH, land use, annual average precipitation, average temperature) and a specified geographical location. Abundance information would also be a welcome improvement to the model.

6 Bibliography

Abrahamsen, G., 1983. Effects of lime and artificial rain on the enchytraeid (Oligochaeta) fauna in coniferous forest. Holarctic Ecology 6, 247–254.

Akaike, H., 1974. A new look at the statistical model identification. Autom. Control, IEEE Trans. 19, 716–723.

BBA (Biologische Bundesanstalt für Land- und Forstwirtschaft) (1986): Richtlinie für die amtliche Prüfung von Pflanzenschutzmitteln, Teil IV, 4-1. Verbleib von Pflanzenschutz-mitteln im Boden - Abbau, Umwandlung und Metabolismus.

Berg, M., Hemerik, L., 2004. Secondary succession of terrestrial isopod, centipede, and millipede communities in grasslands under restoration. Biology And Fertility Of Soils 40, 163-170.

Berg, M.P., Verhoef, H.A., Bolger, T., Anderson, J.M., Beese, F., Coûteaux, M.M., Ineson, P., McCarthy, F., Palka, L., Raubuch, M., Splatt, P., Willison, T., 1997. Effects of air pollutant-temperature interac- tions on mineral-N dynamics and cation leaching in reciprocal forest soil transplantation experiments. Biogeochemistry 39, 295-326.

Bispo, A., Cluzeau, D., Creamer, R., 2009. Indicators for Monitoring Soil Biodiversity. Integr Environ Assess Manag 5, 717-719.

Boesten, J.J., Kopp, H., Adriaanse, P.I., Brock, T.C., Forbes, V.E., 2007. Conceptual model for improving the link between exposure and effects in the aquatic risk assessment of pesticides. Ecotoxicology and Environmental Safety 66, 291-308.

Bonnet, L., Cassagnau, P., Deharveng, L., 1976. Example Of Biocenotic Imbalance Due To Clearing - Soil Collembola Aopulations Of Piau-Dengaly (Hautes-Pyrenees). Revue Ecologie et Biologie Du Sol 13, 337-351.

Bouché, M.B. (1977): Strategies lombriciennes. Ecological Bulletin 25: 122-132.

Chagnon, M., 2000. Community structures of Collembola in sugar maple forests: relations to humus type and seasonal trends. Pedobiologia 44, 148-174.

Christiansen, K., 1964. Bionomics Of Collembola. Annual Review Of Entomology 9, 147.

David, J.-F., Devernay, S., Loucougaray, G., Le Floch, E., 1999. Below-ground biodiversity in a Mediterranean landscape: relationships between saprophagous macroarthropod communities and vegetation structure. Biodivers. Conserv, 753–767.

Deharveng, L., 2004. Recent advances in Collembola systematics. Pedobiologia 48, 415-433.

Dmowska, E., 1993. Effects of long-term artificial acid rain on species range and diversity of soil nematodes. European Journal of Soil Biology 29, 398–403.

Dunger, W., 1983. Tiere im Boden. A. Ziemsen Verlag, Wittenberg.

EC (European Commission) (2002): Guidance Document on Terrestrial Ecotoxicology Under Council Directive 91/414/EEC. SANCO/10329/2002 rev 2 final. http://ec.europa.eu/food/plant/protection/evaluation/guidance/wrkdoc09_en.pdf

EC (European Commission) (2006): Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Thematic strategy for soil protection. COM(2006)231 final.

EFSA (European Food Safety Authority) (2007): Opinion of the Scientific Panel on Plant Protection Products and their Residues on a request from the Commission related to the revision of Annexes II and III to Council Directive 91/414/EEC concerning the placing of plant protection products on the market - Ecotoxicological studies. The EFSA Journal: 461: 1-44

EFSA, 2009. Scientific Opinion of the Panel on Plant Protection Products and their Residues on a request from EFSA on the usefulness of total concentrations and pore water concentrations of pesticides in soil as metrics for the assessment of ecotoxicological effects. . The EFSA Journal, 1-90.

EFSA, 2010a. Scientific Opinion on outline proposals for assessment of exposure of organisms to substances in soil. EFSA Journal 8(1), 1-38.

EFSA, 2010b. Scientific Opinion on the development of a soil ecoregions concept using distribution data on invertebrates. EFSA Journal 8(10), 1-77.

EFSA, 2010c. Scientific Opinion on the importance of the soil litter layer in agricultural areas. EFSA Journal 8.

Faber, J.H., Teuben, A., Berg, M.P., Doelman, P., 1992. Microbial biomass and activity in pine litter in the presence of tomocerus-minor (insecta, collembola). Biology and Fertility Of Soils 12, 233-240.

Filser, J., 1995. The effect of green manure on the distribution of collembola in a permanent row crop. Biology And Fertility Of Soils 19, 303-308.

FOCUS, 1997. Soil persistence models and EU registration.

FOCUS, 2006. Guidance Document on Estimating Persistence and Degradation Kinetics from Environmental Fate Studies on Pesticides in EU Registration. Report of the FOCUS Work Group on Degradation Kinetics, p. 434 pp.

Frampton, G., van den Brink, P., 2002. Influence of cropping on the species composition of epigeic Collembola in arable fields. Pedobiologia 46, 328-337.

Ganzelmeier, H.a.R., D. , 2000. Drift, drift reducing sprayers and sprayer testing. Aspects of Applied Biology 57, 1-10.

Gisin, H., 1943. Ökologie und Lebensgemeinschaften der Collembolen im Schweizer Exkursionsgebiet Basel. Rev. Ecol. Biol. Sol 50, 131-224.

Guisan, A., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147-186.

Guisan, A., Edwards, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157, 89-100.

Haacker, U., 1968. Deskriptive, experimentelle und vergleichende Untersuchungen zur Aut-ökologie rhein-mainischer Diplopoden. . Oecologia 1, 87-129.

Hasegawa, M., Takeda, H., 1995. Changes in feeding attributes of 4 collembolan populations during the decomposition process of pine needles. Pedobiologia 39, 155-169.

Hassall, M., Rushton, S.P., 1984. Feeding behaviour of terrestrial isopods in relation to plant defence and microbial activity. In: Sutton, S.L., Holdich, D.M. (Eds.), The Biology of Terrestrial Isopods. The Zoological Society of London. Clarendon Press, Oxford, pp. 487–505.

Hassall, M., Rushton, S.P., 1985. The adaptative significance of coprophagous behaviour in the terrestrial isopod Porcellio scaber. Pedobiologia 28, 169-175.

Hassall, M.P., Turner, T.G., Rands, M.R.W., 1987. Effects of terrestrial isopods on the decomposition of woodland leaf litter. Oecologia 597–604.

Heisler, C., Kaiser, E.A., 1995. Influence of agricultural traffic and crop management on Collembola and microbial biomass in arable soil. Biology And Fertility Of Soils 19, 159-165.

Hengl, T., 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables. JRC Scientific and Technical Reports. Joint Research Center, Ispra, p. 165.

Hengl, T., Sierdsema, H., Radović, A., Dilo, A., 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling 220, 3499-3511.

Hession, S.L., Shortridge, A.M., Torbick, M.N., 2006. Categorical models for spatial data uncertainty. In: Caetano, M., Painho, M. (Eds.), Proceedings of the 7th

International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Instituto Geographico Portugues Lisbon, pp. 386–395.

Higgins, S., Richardson, D., 2001. Validation of a spatial simulation model of a spreading alien plant population. Journal of Applied Ecology 38, 571–584.

Hopkin, S.P., Read, H.J., 1992. The biology of millipedes. Oxford Science, Oxford.

Kautz, G., Zimmer, M., Topp, W., 2000. Responses of the parthenogenetic isopod, Trichoniscus pusillus, to changes in food quality. Pedobiologia 44, 75-85.

Kayang, H., Sharma, G.D., Mishra, R.R., 1996. The influence of isopod grazing on microbial dynamics in decomposing leaf litter of Alnus nepalensis D. Don. . European Journal of Soil Biology 32, 35-39.

Kopeszki, H., 1992. Veränderungen der Mesofauna eines Buchenwaldes bei Säurebelastung. Pedobiologia 36, 295–305.

Lavelle, P., Spain, A.V., 2001. Soil Ecology. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Lee, K.E. (1985): Earthworms: Their ecology and relationships with soils and land use. Sydney, Australia: Academic Press. 411 pp.

Lehmann, A., Overton, J., 2002. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling 157, 189-207.

Lohm, U., Larsson, K., Nömmik, H., 1984. Acidification and liming of coniferous forest soil: long-term effects on turnover rates of carbon and nitrogen during an incubation experiment. Soil Biology and Biochemistry 16, 343–346.

Loranger, G., Bandyopadhyaya, I., 2001. Does soil acidity explain altitudinal sequences in collembolan communities? Soil Biology & Biochemistry 33, 381-393.

Miller, J., 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. Ecological Modelling 157, 227-247.

Paoletti, M., Hassall, M., 1999. Woodlice (Isopoda: Oniscidea): their potential for assessing sustainability and use as bioindicators. Agric. Ecosyst. Environ. 174, 157– 165.

Paoletti, M.G., Martinelli, N., 1981. Comparazione delle preferenze alimentari di tre Androniscus su substrati vegetali naturali (Isopoda Trichoniscidae). Mem. de Biospeologie Moulis 8, 33– 37.

Petersen, H., 2002. Effects of non-inverting deep tillage vs. conventional ploughing on collembolan populations in an organic wheat field. European Journal Of Soil Biology 38, 177-180.

Ponge, J., 1991. Food Resources And Diets Of Soil Animals In A Small Area Of Scots Pine Litter. GEODERMA 49, {33-62}.

Ponge, J.-F., 2000. Acidophilic Collembola: living fossils? Contributions from the Biological Laboratory, Kyoto University 29, 65-74.

Ponge, J.-F., Prat, B., 1982. Les Collemboles, indicateurs du mode d'humification dans les peuplements résineux, feuillus et málangés: résultats obtenus en forêt d'Orléans. Revue d'Écologie et de Biologie du Sol 19, 237-250.

Sastrodihardjo, F.X.S., Van Straalen, N.M., 1993. Behaviour of five isopod species in standardized test for pH preference. Eur. J. Soil Biol 29, 127– 131.

Schmalfuss, H., 1984. Eco-morphological strategies in terrestrial isopods. In: Sutton, S.L., Holdich, D.M. (Eds.), The Biology of Terrestrial Isopods. The Zoological Society of London. Clarendon Press, Oxford, pp. 49–63.

Schmalfuss, H., 1989. Phylogenetics in Oniscoidea. In: Ferrara, F., Argano, R., Manicastri, C., Schmalfus, H., Taiti, S. (Eds.), Proceedings of the Second Symposium on the Biology of Terrestrial Isopods. Monitore Zoologico Italiano, pp. 49-63.

Schmoelzer, K., 1965. Ordnung Isopoda. Akademie-Verlag Berlin.

Sousa, J.P., Römbke, J., Gardi, C., Jeffrey, S., 2009. Development of a biogeographical data base and of a draft European ecoregion map (EU-27), based on Lumbricidae, Collembola, Isopoda and Enchytraeidae biogeographical and taxonomical literature. Report of EFSA Contract ref.: CT/EFSA/PPR/2008/01, p. 48.

Souty-Grosset, C., Badenhausser, I., Reynolds, J., Morel, A., 2005. Investigations on the potential of woodlice as bioindicators of grassland habitat quality. European Journal Of Soil Biology 41, 109-116.

Sutton, S., Harding, P., Burn, D., 1972. Key to British Woodlice. Ginn & Company, London, UK,.

Szlavecz, K., 1995. Diversity and spatial community structure of ter- restrial isopods (Isopoda, Oniscidea) in a mosaic of plant assemblages. In: Schram, F.R. (Ed.), Crustacean Issues 9, Terrestrial Isopod Biology, pp. 97-106.

Thiele, H.-U., 1959. Experimentelle Untersuchungen über die Abhängigkeit bodenbewohnender Tierarten vom Kalkgehalt des Standorts. Zeitschrift für Angewandte Entomologie 44, 1-21.

Van den Brink, P., 2002. Influence of cropping on the species composition of epigeic Collembola in arable fields. Proceedings Of The Xth International Colloquium On Apterygota, Ceske Budejovice 2000: Apterygota At The Beginning Of The Third Millennium. Pedobiologia, pp. 328-337.

Van der Linden, A.M.A., Boesten, J.J.T.I., Brock, T.C.M., Van Eekelen, G.M.A., Ter Horst M.M.S., De Jong, F.M.W., Montforts, M.H.M.M. & Pol, J.W. , 2008. Revised proposal for the risk assessment of persistence of plant protection products in soil. RIVM report 601712003/2008.

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, 1175-1183.

Van Wensem, J., Verhoef, H.A., Van Straalen, N.M., 1993. Litter degradation stage as a prime factor for isopod interaction with mineralization processes. Soil Biol. Biochem., 1175-1183.

Vikamaa, P., Huhta, V., 1986. Effects of fertilization and pH on communities of Collembola in pine forest soil. Annales Zoologici Fennici. 23, 167-174.

Warburg, M.R., Linsenmair, K.E., Bercovitz, K., 1984. The effect of climate on the distribution and abundance of isopods. Symposia of the Zoological Society of London pp. 339-367.

Zimmer, M., 2000. Correspondence analytical evaluation of factors that influence soil macro-arthropod distribution in abandoned grassland. Pedobiologia 44, 695-704.

Zimmer, M., Topp, W., 1997a. Does leaf litter quality influence population parameters of the common woodlouse, Porcellio scaber (Crustacea: Isopoda)? . Biology and Fertility of Soils 24, 435-441.

Zimmer, M., Topp, W., 1997b. Homeostatic responses in the gut of Porcellio scaber (Isopoda: Oniscidea) optimize litter degradation. Journal of Comparative Physiology B 167, 582–585.

Annex 1: EFSA Database structure

Section 1 – Site information (location, land-use)	
ID Entry	Entry in the database (one per each registry)
ID Site	Identification number of each site in the database
Country	Country name (in this case Finland, Germany, Portugal)
Region	Administrative region within the country.
Village/town	Name of the nearest village or town
Name/place	Name of the site
Coordinate (Long) Coordinates (Lat) Coordinates (format) Coordinates (datum)	Longitude Latitude Geographic system used (in this case UTM) Geodetic system used (in this case: WGS84) *NOTE: In the cases where no coordinates were mentioned in the literature searched, the coordinates were obtained using the approximate location of the sampling site (nearest village or town). This was done using the Google Earth search engine. In these cases the coordinates may fall within urban limits.*
Land-use	Land-use type (e.g., forest, pasture, crop area)
Dominant vegetation	Dominant vegetation at the site
Observations	Any relevant information can be placed in this field

Section 2 – Information on soil type and soil properties of that particular site	
ID Soil	Identification number of each soil type (usually at each site)
Class Class (typology used)	Soil class type and typology used
Texture Texture (typology used)	Soil texture and typology used
Sand (%) Silt (%) Clay (%)	Percentage of Sand, Silt and Clay
pH pH_SD pH_Min pH_Max	pH values (measure of variation and range if several pH values are reported for the same site)
Org. matter Org. matter_SD Org. matter_Min Org. matter_Max Org. matter_Unit	Soil organic matter content (measure of variation and range if several organic matter values are reported for the same site); Unit used (in most cases %)

Corg Corg_SD Corg_Min Corg_Max Corg_Unit	Soil organic carbon content (measure of variation and range if several soil organic carbon values are reported for the same site); Unit used (in most cases %)
Ntot Ntot_SD Ntot_Min Ntot_Max Ntot_Unit	Soil total nitrogen content (measure of variation and range if several Nitrogen values are reported for the same site); Unit used (in most cases %)
C/N C/N_SD C/N_Min C/N_Max	Soil C/N ratio (measure of variation and range if several C/N values are reported for the same site)
WHCmax WHC_Unit	Soil maximum water holding capacity and unit of expression
Humus type Reference Humus type	Humus type (if mentioned) typology used
Observations	Any relevant information can be placed in this field

Section 3 – Information on the species	
ID Sp	Identification number of each species in the database
Order, Family, Species Author, Year	Taxonomic information (including author and year of description)
Life-form typus	Information on life-form (dependent of the organism group). For Collembola 3 morphological traits were used to define the life form: ocelli, antenna and furca. Each one was coded between 1 and 5 as follows: Ocelli: 1= (0+0) ocelli; 2= (1+1)-(2+2) ocelli; 3=(3+3)-(4+4) ocelli; 4=(5+5)-(6+6) ocelli; 5=(7+7)-(8+8) ocelli Antenna: 1= <0.25 of body length; 2= 0.25-0.5 of body length; 3= 0.5-0.75 of body length; 4= 0.75-1 of body length; 5= >1 of body length Furca: 1= absent; 3= reduced/short; 5= fully developed These three traits were combined to create the following life-form typology (an higher score indicates a life–form adapted to upper soil layers and with a high dispersal capability): Life-form classes: class 1= score 1-3 (euedaphic; very low dispersal); class 2= score 4-6 (euedaphic-hemiedaphic; low dispersal); class 3= score 7-9 (hemiedaphic; medium dispersal); class 4= score 10-12 (hemiedaphic-epigeic; medium-fast dispersal); class 5= score 13-15 (epigeic-fast dispersal)

Depth	Soil depth at which the species was collected
Horizon	Horizon (litter layer or soil) at which the species was collected
Abundance Abundance_Min Abundance_Max Abundance_basis	Abundance of the species in the set of samples or sampling date. For Collembola this value can vary: total nº of individuals in the sample (the default measure); nº individuals/m2; nº individuals/trap (in case of the use of pitfall traps). For earthworms and enchytraeids this value is usually given in nº individuals/m2. For isopods this value (when available) is usually given in nº individuals/trap (in case of the use of pitfall traps)
Sampling method	Sampling method used to collect samples
Observations	Any relevant information can be placed in this field. For earthworms and enchytraeids dominance data is given in this field.

Section 4 – Information on the source of the data (database, publication, report)	
ID Ref	Identification number of each data source
First author	Name of the first author
Journal / Source	Name of the data source (usually a journal)
Year	Year of publication
Volume (issue)	Volume & issue (when applicable)
Pages	Page numbers (when applicable)
Observations	Any relevant information can be placed in this field

Annex 2: Codes for JRC Maps

a. Coded country

1. Portugal
2. Germany
3. Finland

b. Coded Land use

1. Crop area
2. Grassland
3. Shrub
4. Forest

c. Coded Texture

JRC Code	Description	Clay [%]	Silt [%]	Sand [%]
9	Full organic	0.0	0.0	0.0
5	Very fine	73.3	13.3	13.3
4	Fine	46.5	26.7	26.7
3	Medium fine	17.5	75.0	7.5
2	Medium	18.0	39.4	42.6
1	Coarse	7.6	13.7	78.7

Annex 3: STELLA codes

```
ln_Collembola_LF1(t) = ln_Collembola_LF1(t - dt) + (Collembola_LF1_gains - adjust_collembola_LF1 - Collembola_LF1_losses) * dt
INIT ln_Collembola_LF1 = 0
INFLOWS:
Collembola_LF1_gains = IF option_constant_0_vs_dynamic_1_Collembola=0 THEN (2.667+0.1125*Texture_class+0.02747*Average_Temperature_C_constant) ELSE (2.667+0.1125*Texture_class+0.02747*Average_Temperature_C__dynamic)
OUTFLOWS:
adjust_collembola_LF1 = ln_Collembola_LF1
Collembola_LF1_losses = 0.000003944*Latitude
ln_Collembola_LF2(t) = ln_Collembola_LF2(t - dt) + (Collembola_LF2_gains - Collembola_LF2_losses - adjust_collembola_LF2) * dt
INIT ln_Collembola_LF2 = 0
INFLOWS:
Collembola_LF2_gains = IF option_constant_0_vs_dynamic_1_Collembola=0 THEN (3.002+0.07160*Average_Temperature_C_constant) ELSE (3.002+0.07160*Average_Temperature_C__dynamic)
OUTFLOWS:
Collembola_LF2_losses = 0.000003228*Latitude
adjust_collembola_LF2 = ln_Collembola_LF2
ln_Collembola_LF3(t) = ln_Collembola_LF3(t - dt) + (Collembola_LF3_gains - Collembola_LF3_losses - adjust_collembola_LF3) * dt
INIT ln_Collembola_LF3 = 0
INFLOWS:
Collembola_LF3_gains = 3.384+0.0559*Texture_class
OUTFLOWS:
Collembola_LF3_losses = 0.000002308*Latitude
adjust_collembola_LF3 = ln_Collembola_LF3
```

ln_Isopod_LF1(t) = ln_Isopod_LF1(t - dt) + (Isopod_LF1_gains - Isopod_LF1_losses - adjust_isopod_LF1) * dt

INIT ln_Isopod_LF1 = 0

INFLOWS:

Isopod_LF1_gains = IF option_constant_0_vs_dynamic_1_Isopod=0 THEN (1.14413*OM_g_per_g_constant+20.41991*WaterFC_m3_m_minus3) ELSE (1.14413*OM_g_per_g_dynamic+20.41991*WaterFC_m3_m_minus3)

OUTFLOWS:

Isopod_LF1_losses = 4.09247+(0.37849*Country)+(0.0924*Landuse_class)+(1.00156*Texture_class)

adjust_isopod_LF1 = ln_Isopod_LF1

ln_Isopod_LF2(t) = ln_Isopod_LF2(t - dt) + (Isopod_LF2_gains - Isopod_LF2_losses - adjust_isopod_LF2) * dt

INIT ln_Isopod_LF2 = 0

INFLOWS:

Isopod_LF2_gains = IF option_constant_0_vs_dynamic_1_Isopod=0 THEN (0.1066*Average_Temperature_C_constant+0.0459*Landuse_class+0.748*OM_g_per_g_constant+0.509*Country) ELSE (0.1066*Average_Temperature_C__dynamic+0.0459*Landuse_class+0.748*OM_g_per_g_dynamic+0.509*Country)

OUTFLOWS:

Isopod_LF2_losses = 1.510

adjust_isopod_LF2 = ln_Isopod_LF2

codominance_Isopod = IF Isopod_LF1_%=Isopod_LF2% THEN 12 ELSE 0

Collembola_Dominance_class = IF Collemb_Dominance_sum=100 THEN 1 ELSE

IF Collemb_Dominance_sum=103 THEN 13 ELSE

IF Collemb_Dominance_sum=120 THEN 12 ELSE

IF Collemb_Dominance_sum=20 THEN 2 ELSE

Collemb_Dominance_sum

Collembola_LF1 = IF ln_Collembola_LF1=0 THEN 0 ELSE EXP(ln_Collembola_LF1)

Collembola_LF2 = IF ln_Collembola_LF2=0 THEN 0 ELSE EXP(ln_Collembola_LF2)

Collembola_LF3 = IF ln_Collembola_LF3=0 THEN 0 ELSE EXP(ln_Collembola_LF3)

Collemb_Dominance_1 = IF collemb_LF1%>=16.7 THEN 100 ELSE 0

Collemb_Dominance_2 = IF Collemb_LF2%>=16.7 THEN 20 ELSE 0

Collemb_Dominance_3 = IF Collemb_LF3%>=16.7 THEN 3 ELSE 0

Collemb_Dominance_sum = SUM(Collemb_Dominance_1,Collemb_Dominance_2,Collemb_Dominance_3)

collemb_LF1% = if TOTAL_collem_LF_richness=0 then 0 else (Collembola_LF1/TOTAL_collem_LF_richness*100)

Collemb_LF2% = if TOTAL_collem_LF_richness=0 then 0 else (Collembola_LF2/TOTAL_collem_LF_richness*100)

Collemb_LF3% = if TOTAL_collem_LF_richness=0 then 0 else (Collembola_LF3/TOTAL_collem_LF_richness*100)

dominance_Isopod = IF codominance_Isopod=0 AND Isopod_LF1_%>50 THEN 1 ELSE 2

isopod_LF1 = IF ln_Isopod_LF1=0 THEN 0 ELSE EXP(ln_Isopod_LF1)

Isopod_LF1_% = if total_isopod_LF_richness=0 then 0 else isopod_LF1/total_isopod_LF_richness*100

isopod_LF2 = IF ln_Isopod_LF2=0 THEN 0 ELSE EXP(ln_Isopod_LF2)

Isopod_LF2% = if total_isopod_LF_richness=0 then 0 else isopod_LF2/total_isopod_LF_richness*100

option_constant_0_vs_dynamic_1_Collembola = 0

option_constant_0_vs_dynamic_1_Isopod = 0

TOTAL_collem_LF_richness = Collembola_LF1+Collembola_LF2+Collembola_LF3

total_isopod_LF_richness = isopod_LF1+isopod_LF2

Average_Temperature_C_constant = GRAPH(TIME)

Average_Temperature_C__dynamic = GRAPH(TIME)

Country = GRAPH(TIME)

Landuse_class = GRAPH(TIME)

Latitude = GRAPH(TIME)

OM_g_per_g_constant = GRAPH(TIME)

OM_g_per_g_dynamic = GRAPH(TIME)

pH_constant = GRAPH(TIME)

pH_dynamic = GRAPH(TIME)

Texture_class = GRAPH(TIME)

* The GRAPH(TIME) values are taken from using each point environmental variable as time.

Annex 4: ArcGIS codes

a. Raster Calculator

Collembola

```
lf1cg = 2.667 + (0.1125 * [Extract_Tex_G]) + (0.02747 * [Extract_TM_G]) + (- 0.000003944 * $$ymap)
lf2cg = 3.002 + (0.07160 * [Extract_TM_G]) + (- 0.000003228 * $$ymap)
lf3cg = 3.384 + (0.0559 * [Extract_Tex_G]) + (- 0.000002308 * $$ymap)

lf1cf = 2.667 + (0.1125 * [Texture.asc]) + (0.02747 * [TMean.asc]) + (- 0.000003944 * $$ymap)
lf2cf = 3.002 + (0.07160 * [TMean.asc]) + (- 0.000003228 * $$ymap)
lf3cf = 3.384 + (0.0559 * [Texture.asc]) + (- 0.000002308 * $$ymap)

lf1cp = 2.667 + (0.1125 * [Texture.asc]) + (0.02747 * [TMean.asc]) + (- 0.000003944 * $$ymap)
lf2cp = 3.002 + (0.07160 * [TMean.asc]) + (- 0.000003228 * $$ymap)
lf3cp = 3.384 + (0.0559 * [Texture.asc]) + (- 0.000002308 * $$ymap)

Eup = Exp([lf1cp])
Eug = Exp([lf1cg])
Euf = Exp([lf1cf])
Hp = Exp([lf2cp])
Hg = Exp([lf2cg])
Hf = Exp([lf2cf])
Epip = Exp([lf3cp])
Epig = Exp([lf3cg])
Epif = Exp([lf3cf])
```

```
TRcf = [Epif] + [Euf] + [Hf]
TRcg = [Epig] + [Eug] + [Hg]
TRcp = [Epip] + [Eup] + [Hp]

pc_Eup = ([Eup] / [TRcp]) * 100
pc_Eug = ([Eug] / [TRcg]) * 100
pc_Euf = ([Euf] / [TRcf]) * 100
pc_Hp = ([Hp] / [TRcp]) * 100
pc_Hg = ([Hg] / [TRcg]) * 100
pc_Hf = ([Hf] / [TRcf]) * 100
pc_Epip = ([Epip] / [TRcp]) * 100
pc_Epig = ([Epig] / [TRcg]) * 100
pc_Epif = ([Epif] / [TRcf]) * 100
```

Single output map algebra

```
con (
pc_Eup + pc_Hp >= 83.3 & pc_Eup <= 66.7 & pc_Hp <= 66.7, 12,
pc_Eup + pc_Epip >= 83.3 & pc_Eup <= 66.7 & pc_Epip <= 66.7, 13,
pc_Hp + pc_Epip >= 83.3 & pc_Hp <= 66.7 & pc_Epip <= 66.7, 23,
pc_Eup >= 66.7, 1,
pc_Hp >= 66.7, 2,
pc_Epip >= 66.7, 3,
pc_Eup + pc_Hp <= 83.3 & pc_Hp + pc_Epip <= 83.3 & pc_Eup + pc_Epip <= 83.3, 123
)

con (
pc_Euf + pc_Hf >= 83.3 & pc_Euf <= 66.7 & pc_Hf <= 66.7, 12,
```

```
pc_Euf + pc_Epif >= 83.3 & pc_Euf <= 66.7 & pc_Epif <= 66.7, 13,
pc_Hf + pc_Epif >= 83.3 & pc_Hf <= 66.7 & pc_Epif <= 66.7, 23,
pc_Euf >= 66.7, 1,
pc_Hf >= 66.7, 2,
pc_Epif >= 66.7, 3,
pc_Euf + pc_Hf <= 83.3 & pc_Hf + pc_Epif <= 83.3 & pc_Euf + pc_Epif <= 83.3, 123
)
```

```
con (
pc_Eug + pc_Hg >= 83.3 & pc_Eug <= 66.7 & pc_Hg <= 66.7, 12,
pc_Eug + pc_Epig >= 83.3 & pc_Eug <= 66.7 & pc_Epig <= 66.7, 13,
pc_Hg + pc_Epig >= 83.3 & pc_Hg <= 66.7 & pc_Epig <= 66.7, 23,
pc_Eug >= 66.7, 1,
pc_Hg >= 66.7, 2,
pc_Epig >= 66.7, 3,
pc_Eug + pc_Hg <= 83.3 & pc_Hg + pc_Epig <= 83.3 & pc_Eug + pc_Epig <= 83.3, 123
)
```

Isopods

```
lf1ig = -4.09247 + (-0.37849 * 2) + (-0.0924 * [Extract_LU_G]) + (1.14413 * [Extract_OM_G]) + (-1.00156 * [extract_tex_g]) + (20.41991 * [Extract_Wfc_G])

lf2ig = -1.510 + (0.1066 * [extract_tm_g]) + (0.0459 * [Extract_LU_G]) + (0.748 * [Extract_OM_G]) + (0.509 * 2)
```

```
lf1if = -4.09247 + (-0.37849 * 3) + (-0.0924 * [Extract_LU_F]) + (1.14413 *
[Extract_OM_F]) + (-1.00156 * [extract_tex_f]) + (20.41991 * [Extract_Wfc_F])

lf2if = -1.510 + (0.1066 * [extract_tm_f]) + (0.0459 * [Extract_LU_F]) + (0.748 *
[Extract_OM_F]) + (0.509 * 3)


lf1ip = -4.09247 + (-0.37849 * 1) + (-0.0924 * [Extract_LU_P]) + (1.14413 *
[Extract_OM_P]) + (-1.00156 * [extract_tex_p]) + (20.41991 * [Extract_Wfc_P])

lf2ip = -1.510 + (0.1066 * [extract_tm_p]) + (0.0459 * [Extract_LU_P]) + (0.748 *
[Extract_OM_P]) + (0.509 * 1)


SDf = Exp([lf1if])

SDg = Exp([lf1ig])

SDp = Exp([lf1ip])

LDf = Exp([lf2if])

LDg = Exp([lf2ig])

LDp = Exp([lf2ip])


TRf = [LDf] + [SDf]

TRg = [LDg] + [SDg]

TRp = [LDp] + [SDp]


pc_SDf = ([SDf] / [TRf]) * 100

pc_SDg = ([SDg] / [TRg]) * 100

pc_SDp = ([SDp] / [TRp]) * 100

pc_LDf = ([LDf] / [TRf]) * 100

pc_LDg = ([LDg] / [TRg]) * 100

pc_LDp = ([LDp] / [TRp]) * 100


con (

pc_SDf == pc_LDf, 12,
```

```
pc_SDf > 50, 1,
pc_LDf > 50, 2
)

con (
pc_SDg == pc_LDg, 12,
pc_SDg > 50, 1,
pc_LDg > 50, 2
)

con (
pc_SDp == pc_LDp, 12,
pc_SDp > 50, 1,
pc_LDp > 50, 2
)
```

Single Output Map Algebra for Geostatistical Wizard – Ordinary Kriging (dominance class map)

```
con (
e_k_lf1cp_val + e_k_lf2cp_v >= 83.3 & e_k_lf1cp_val <= 66.7 & e_k_lf2cp_v <= 66.7, 12,
e_k_lf1cp_val + e_k_lf3cp_v >= 83.3 & e_k_lf1cp_val <= 66.7 & e_k_lf3cp_v <= 66.7, 13,
e_k_lf2cp_v + e_k_lf3cp_v >= 83.3 & e_k_lf2cp_v <= 66.7 & e_k_lf3cp_v <= 66.7, 23,
e_k_lf1cp_val >= 66.7, 1,
e_k_lf2cp_v >= 66.7, 2,
e_k_lf3cp_v >= 66.7, 3,
e_k_lf1cp_val + e_k_lf2cp_v <= 83.3 & e_k_lf2cp_v + e_k_lf3cp_v <= 83.3 & e_k_lf1cp_val + e_k_lf3cp_v <= 83.3, 123
```

```
)

con (
e_k_lf1cg_val + e_k_lf2cg_v >= 83.3 & e_k_lf1cg_val <= 66.7 & e_k_lf2cg_v <=
66.7, 12,
e_k_lf1cg_val + e_k_lf3cg_v >= 83.3 & e_k_lf1cg_val <= 66.7 & e_k_lf3cg_v <=
66.7, 13,
e_k_lf2cg_v + e_k_lf3cg_v >= 83.3 & e_k_lf2cg_v <= 66.7 & e_k_lf3cg_v <= 66.7,
23,
e_k_lf1cg_val >= 66.7, 1,
e_k_lf2cg_v >= 66.7, 2,
e_k_lf3cg_v >= 66.7, 3,
e_k_lf1cg_val + e_k_lf2cg_v <= 83.3 & e_k_lf2cg_v + e_k_lf3cg_v <= 83.3 &
e_k_lf1cg_val + e_k_lf3cg_v <= 83.3, 123
)

con (
e_k_lf1cf_v + e_k_lf2cf_v >= 83.3 & e_k_lf1cf_v <= 66.7 & e_k_lf2cf_v <= 66.7,
12,
e_k_lf1cf_v + e_k_lf3cf_v >= 83.3 & e_k_lf1cf_v <= 66.7 & e_k_lf3cf_v <= 66.7,
13,
e_k_lf2cf_v + e_k_lf3cf_v >= 83.3 & e_k_lf2cf_v <= 66.7 & e_k_lf3cf_v <= 66.7,
23,
e_k_lf1cf_v >= 66.7, 1,
e_k_lf2cf_v >= 66.7, 2,
e_k_lf3cf_v >= 66.7, 3,
e_k_lf1cf_v + e_k_lf2cf_v <= 83.3 & e_k_lf2cf_v + e_k_lf3cf_v <= 83.3 &
e_k_lf1cf_v + e_k_lf3cf_v <= 83.3, 123
)

con (
```

```
e_k_lf1ip_va == e_k_lf2ip_va, 12,
e_k_lf1ip_va > 50, 1,
e_k_lf2ip_va > 50, 2
)

con (
e_k_lf1ig_v == e_k_lf2ig_v, 12,
e_k_lf1ig_v > 50, 1,
e_k_lf2ig_v > 50, 2
)

con (
e_k_lf1if_v == e_k_lf2if_v, 12,
e_k_lf1if_v > 50, 1,
e_k_lf2if_v > 50, 2
)
```