Available online at www.sciencedirect.com

**ScienceDirect**

# Pedo biologia

SOIL ECOLOGICAL AND ECONOMIC EVALUATION OF GENETICALLY MODIFIED CROPS – ECOGEN

# Hierarchical classification of environmental factors and agricultural practices affecting soil fauna under cropping systems using *Bt* maize

# Marko Debeljak[a,*], Jérôme Cortet[b,c], Damjan Demšar[a], Paul H. Krogh[d], Saso Džeroski[a]

[a]*Department of Knowledge Technologies (E8), Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*
[b]*ENSAIA-INPL/INRA, UMR 1120, Laboratoire Sols et Environnement, 2, avenue de la forêt de Haye, BP 172, 54505 Vandoeuvre-lès-Nancy, France*
[c]*Institut Mediterraneen d'Ecologie et Paleoecologie, Europole de l'Arbois, BP 80. Bâtiment Villemin, 13545 Aix-en-Provence Cedex 04, France*
[d]*Soil Fauna and Ecotoxicology Research Unit , Department of Terrestrial Ecology, National Environmental Research Institute, P.O. Box 314, Vejlsoevej 25, DK-8600, Silkeborg, Denmark*

## Summary

The population dynamics of soil organisms under agricultural field conditions are influenced by many factors, such as pedology and climate, but also farming practices such as crop type, tillage and the use of pesticides. To assess the real effects of farming practices on soil organisms it is necessary to rank the influence of all of these parameters. *Bt* maize (*Zea mays* L.), as a crop recently introduced into farming practices, is a genetically modified maize with the Cry1Ab gene which produces a protein toxic to specific lepidopteran insect pests. To assess the effects of *Bt* maize on non-target soil organisms, we conducted research at a field site in Foulum (Denmark) with a loamy sand soil containing 6.4% organic matter. The study focused on populations of springtails (Collembola) and earthworms (Oligochaeta) from samples taken at the beginning and at the end of the maize crop-growing season during 2 consecutive years. Farming practices, soil parameters, the biological structure of soil communities, and the type and age of the crop at the time of sampling, were used as attributes to predict the total abundance of springtails and biomass of earthworms in general and the abundance or biomass for specific functional groups (epigeic, endogeic and anecic groups for earthworms, and eu-, eu to hemi-, hemi-, hemi to epi- and epiedaphic groups for Collembola). Predictive models were built with data mining tools, such as regression trees that predict the value of a dependent variable from a set of independent variables. Regression trees

*Corresponding author. Tel.: +38614773124.
*E-mail address:* marko.debeljak@ijs.si (M. Debeljak).

were constructed with the data mining system M5′. The models were evaluated by qualitative and quantitative measures of performance and two models were selected for further interpretation: anecic worms and hemi-epiedaphic Collembola. The anecic worms ($r^2 = 0.83$) showed preferences for less clay and more silt soil with medium pH but were not influenced directly by farming practices. The biomass of earthworms was greater in early autumn than in spring or late autumn. Biomass of hemi-epiedaphic Collembola ($r^2 = 0.59$) increased at the end of the maize growing season, while higher organic matter content and pH tended to increase their biomass in spring. Greater abundance of Collembola was also noted in early autumn if the crop was non-*Bt* maize. The models assessed by this research did not find any effects of the *Bt* maize cropping system on functional groups of soil fauna.

## Introduction

Genetically modified (GM) crops can be engineered to be resistant to insect pests and are grown to reduce producer inputs and operating costs rather than to achieve higher yields (Gordon, 2004). Environmental concerns about the ecological impacts of GM crops have focused on the invasion of GM crops into wild habitats, gene flow to wild relatives of GM crops (hybridization) and toxic effects on non-target organisms. These concerns need to be evaluated using science-based ecological risk assessment (Andow and Hilbeck, 2004). Soil biota play a vital role in key soil processes (nutrient cycling, and the physical and biochemical degradation of organic matter), and can be used to evaluate the potential impacts of GM crops on soil processes, which is one of the purposes of the ECOGEN project (www.ecogen.dk).

*Bt* maize (*Zea mays* L.) is GM to express the Cry1Ab gene from *Bacillus thuringiensis* and produces a protein considered to be toxic to specific lepidopteran insect pests (Schnepf et al., 1998). Within the ECOGEN project most of the studies concerning the effects of *Bt* maize on soil biota have shown no or very few significant effects, and it has been stated that the effects of the *Bt* trait on soil organisms are negligible compared to the natural fluctuation of populations or consequences of other farming practices (Griffiths et al., 2005; Cortet et al. 2006; other papers appearing in this issue). Indeed, populations of soil organisms under agricultural field conditions may be influenced by many factors, including abiotic parameters such as climate and soil texture, biotic parameters such as crop type and interactions with other soil populations, and farming practices such as tillage and application of various pesticides. For example, studies about the impacts of soil compaction on soil biota (Hendrix et al., 1986; Neher and Barbercheck, 1999; Sabatini et al., 1997; Schrader and Lingnau, 1997) and the effects of different agricultural production systems, such as integrated, organic or conventional systems (Jagers op Akkerhuis et al., 1988; El Titi and Ipach, 1989; Brussaard et al., 1990; Dekkers et al., 1994; Cortet et al., 2002) have shown effects on the populations of soil organisms due to induced physical and chemical changes of their habitats.

This paper used data mining in order to classify some of the environmental factors, including soil parameters and agricultural practices, which are thought to affect soil fauna. This approach has been previously and successfully used to predict microarthropod populations from biological and physicochemical parameters or from agricultural practices (Kampichler et al., 2000; Debeljak et al., 2005; Demsar et al., 2006). Here we propose to focus on earthworms and Collembola, which can be classified into functional groups representing ecological categories. For Collembola we followed the classification of Gisin (1943), which is based on the vertical distribution of life forms and reflects the preferential habitat of animals in the soil. For earthworms, we used the classification proposed by Bouché (1972), based on the habitat and activity of worms. Many positive results of applying machine learning methods in ecological modeling, in general, and data mining, in particular, have been reported (Stankovski et al., 1998; Debeljak et al., 2001; Džeroski, 2001; Jerina et al., 2003), Thus, the machine learning technique of regression tree induction was applied in developing models about the effects of environmental factors and agricultural practices on earthworms and Collembola functional groups in a cropping system context using GM *Bt* maize.

## Materials and methods

### The data set

The experiment to assess the effects of *Bt* maize on the abundance of Collembola and biomass of

earthworms under real cropping systems was conducted in Foulum (Jutland, Denmark, 56°30′N, 9°35′E) during 2 consecutive years. Three types of varieties were sown (MEB 307, Monumental and DK 242 in 2002, MEB 307 and Monumental in 2003) and different farming practices were applied (Andersen et al., 2007; Table 1). The field experiment was designed using a completely randomized four-block system.

The collection of Collembola and earthworms is described by Cortet et al. (2007) and Krogh et al. (2007), whose papers also appear in this issue. Briefly, for Collembola, six intact soil cores (5 cm depth, 6 cm diameter) were collected from each plot (in the middle of the plot within the rows between plants) at each sampling time (Spring and Autumn 2002 and 2003) from each of four replicate blocks. Collembola were extracted using Berlese-Tullgren extractors (Berlese, 1905) and the McFadyen methodology (Peterson, 1978). Species were identified and counted using a binocular microscope, then classified into the following five functional groups (Gisin, 1943):

- *euedaphic* (*eu*): they live in the soil and are thus, true soil living animals; they are usually blind and white colored and are adapted to quite stable conditions;
- *eu-hemiedaphic* (*eu-hemi*): intermediate between the hemiedaphic and euedaphic functional groups;
- *hemiedaphic* (*hemi*): they live in the litter layer or in the upper few cm of the soil;

- *hemi-epiedaphic* (*hemi-epi*): intermediate between the epidaphic and hemiepidaphic functional groups;
- *epiedaphic* (*epi*): they live on the soil and in habitats on top of the soil; they are physiologically adapted to climatic variations occuring at the surface of soils.

For earthworms, two soil monoliths were collected in each plot at each sampling occasion (Autumn 2002, Spring and autumn 2003). To maximize efficacious sampling, a relatively large volume of soil ($0.5 \times 0.5 \times 0.3\,m^3$) was hand sorted and earthworms were identified to the species level. Both earthworm density and the total biomass (determined as fresh weight of worms with emptied gut) of each species were measured for each soil monolith. Earthworms were classified into three functional groups (Bouché, 1977; Lavelle, 1981):

- *epigeic*: they live and feed on plant litter; they are efficient compost makers but have no impact on the soil structure; because of the low amounts of organic matter; they are usually very scarce in agricultural soils;
- *endogeic*: they live in the soil and are geophagous; they usually do not form permanent galleries and are major agents of soil aggregation;
- *anecic*: they have a mixed regime, living in the soil and feeding on the soil organic matter and plant litter which they collect at the soil surface;

**Table 1.** Farming practices and sampling dates for the selected site (Foulum, see Andersen et al., 2007 and Krogh et al., 2007, for more details)

| Taxa | Sampling dates | Varieties | Soil treatment |
|---|---|---|---|
| Earthworms | Autumn 2002 | MEB 307 (*Bt* trait) (4 plots) Monumental (near-isogenic non-*Bt* trait) (4 plots) | Conventional tillage (CT) (12 plots) |
| | Spring and autumn 2003 | MEB 307 (*Bt* trait) (4 plots) Monumental (near-isogenic non-*Bt* trait) (4 plots) | Conventional tillage (CT) (8 plots) Reduced tillage (RT) (8 plots) |
| Collembola | Spring and autumn 2002 | MEB 307 (*Bt* trait) (4 plots) Monumental (near-isogenic non-*Bt* trait) (4 plots) DK 242 (non-*Bt* trait) (4 plots) | Conventional tillage (CT) (12 plots) |
| | Spring and autumn 2003 | MEB 307 (*Bt* trait) (8 plots) Monumental (near-isogenic non-*Bt* trait) (8 plots) | Conventional tillage (CT) (8 plots) Reduced tillage (RT) (8 plots) |

**Table 2.** Abundance of collembolan functional groups and anecic earthworm biomass (mean±SE) at Foulum, according to sampling dates, tillage (CT – conventional tillage, RT – reduced tillage) and maize varieties

| Date | Tillage | Crop | Collembola (no./kg soil) | | | | | Earthworms (g/m²) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Eu | Eu-hemi | Hemi | Hemi-epi | Epi | Endogeic | Anecic |
| July 2002 | CT | DK242 | 36.69±4.29 | 6.25±1.24 | 0.12±0.12 | 1.85±0.85 | 5.21±1.03 | – | – |
| | CT | MEB307 | 43.40±5.35 | 9.49±1.86 | 0.35±0.20 | 0.46±0.22 | 5.09±1.92 | – | – |
| | CT | Monu | 48.38±5.24 | 11.23±3.43 | 0.12±0.12 | 1.04±0.46 | 7.18±3.64 | – | – |
| October 2002 | CT | DK242 | 70.02±8.12 | 9.38±1.99 | 5.67±1.31 | 2.20±0.79 | 9.49±1.89 | – | – |
| | CT | MEB307 | 90.16±12.58 | 12.38±2.71 | 11.23±4.10 | 1.85±0.56 | 17.01±4.12 | 2.71±0.25 | 16.78±1.55 |
| | CT | Monu | 89.00±11.38 | 10.88±2.34 | 5.32±1.17 | 5.90±1.71 | 6.94±1.21 | 2.34±0.11 | 12.60±0.79 |
| June 2003 | RT | MEB307 | 43.75±15.26 | 4.63±1.41 | 1.85±1.14 | 0.93±0.55 | 16.67±3.15 | 3.03±0.37 | 4.36±0.60 |
| | CT | MEB307 | 40.97±8.60 | 1.16±0.47 | 0.23±0.23 | 0.23±0.23 | 31.94±10.90 | 1.89±0.25 | 7.23±1.10 |
| | RT | Monu | 29.86±4.37 | 7.64±2.43 | 0.23±0.23 | 0.93±0.55 | 23.15±8.65 | 2.24±0.40 | 3.00±0.45 |
| | CT | Monu | 54.86±19.40 | 1.62±0.97 | 0.00±0.00 | 0.00±0.00 | 18.98±4.69 | 2.05±0.23 | 2.75±0.37 |
| November 2003 | RT | MEB307 | 79.49±23.80 | 2.59±1.27 | 9.06±2.99 | 52.13±10.16 | 22.20±7.89 | 1.35±0.075 | 9.76±0.78 |
| | CT | MEB307 | 38.76±7.94 | 2.10±0.82 | 2.72±1.35 | 37.98±8.56 | 12.39±3.61 | 1.78±0.28 | 7.94±1.47 |
| | RT | Monu | 39.53±5.20 | 6.13±3.53 | 4.57±1.86 | 53.57±15.60 | 9.37±3.62 | 1.85±0.13 | 9.60±0.35 |
| | CT | Monu | 29.79±5.37 | 1.62±0.81 | 1.02±0.57 | 52.68±13.95 | 18.79±7.02 | 2.54±0.31 | 9.08±1.05 |

they build dense semi-permanent gallery networks and translocate large amounts of organic matter into the soil.

A total of 480 soil cores were collected for Collembola and 100 soil monoliths for earthworms. Mean abundance of Collembola and biomass of earthworm functional groups is presented in the Table 2. Each sampling unit was described by the set of attributes shown in Table 3, which provides information about the soil environment and farming operations.

## Data mining and statistical analysis

Regression trees are a representation for piece-wise constant or piece-wise linear functions and models are given in a form of hierarchical structures of their elements. Like classical regression equations, they predict the value of a dependent variable (e.g., attributes of biological structure of the soil community from Table 3) from the values of

**Table 3.** Attributes used for the description of soil samples

| Groups of attributes | Attributes |
|---|---|
| Type of the crop | MEB307 (*Bt* variety) Monumental (near-isogenic non-*Bt*-trait) DK 242 (conventional variety, non-*Bt*-trait) |
| Soil properties | Soil texture (clay, silt, sand): [%] Organic matter [%] pH ($H_2O$): [$-\log_{10}H^+$] Cation-exchange capacities ($Ca^{2+}$, $K^+$, $Mg^{2+}$): [meq/100 g soil] |
| Farming practices | Seed treatment: [no, yes] Tillage: [yes, no] |
| Temporal aspects of farming practices at sampling dates | Time since tillage: [# days] Time since sowing: [# days] Time since harvest: [# days] |
| Biological structure of soil communities | Collembola: [# individuals/ kg dry soil/functional group] Earthworms: [fresh weight in mg/m$^{-2}$ soil /functional group] |

a set of independent variables (e.g., first four groups of attributes in Table 3).

Data represented in the form of a table can be used to construct a regression tree. In the table, each row (sample) has the form $(x_1, x_2, ..., x_N, y)$, where $x_i$ are values of the number of attributes (e.g., texture, organic matter content, pH, etc.) and $y$ is the value of the dependent variable (e.g., number of Collembola/kg dry soil/functional group). Unlike classical regression approaches, which find one single equation for a given set of data, regression trees partition the space of examples into axis-parallel rectangles and fit a model to each of these partitions. A regression tree, which has an inverse hierarchical structure, has a test in each inner node (junction from were two links go to the lower hierarchical levels) that tests the value of a certain independent variable, and in each leaf (the lowest level of hieratical tree) can be a linear equation or just a constant for predicting the value of dependent variable.

A number of systems exist for inducing regression trees, such as CART (Breiman et al., 1984) and M5 (Quinlan, 1992). The system M5′ (Wang and Witten, 1997), a reimplementation of M5 within the software package WEKA (Witten and Frank, 1999) was used here, the parameters of M5′ being set to their default values. The quality of the model (i.e., predictive performance) was evaluated with the Pearson correlation coefficient and several other error measures (i.e., mean average error, root mean square error, and relative average error and root relative square error) using ten-fold cross validation (Quinlan, 1986).

## Evaluation of the models

The quality assessment and the final selection of the models were based on quantitative, qualitative and contextual criteria. The quantitative aspects of the models were evaluated by several measures: correlation coefficient, mean absolute error, root mean squared error, relative absolute error and root relative squared error. The correlation coefficient of cross validation was used as the most important measure and the models with correlation coefficients higher then 0.50 were considered for further evaluation. The complexity of the model presents the qualitative aspect of the model quality. The model complexity is described by the number of hierarchical levels of the model structure, and it influences the complexity and difficulties for the interpretation of the model. Models with up to three hierarchical levels were selected for further interpretation. The third evaluation criterion applied was the validation of the internal structure of the models, where the expected and logical relationships between the elements of the model were assessed according to the existing knowledge about the soil ecology of microarthropods and earthworms, and the expected possible and logic interrelationships between the attributes of the model.

## Results

Data mining was applied under different combinations of model making parameters (e.g., number of instances in the leaves of the model (4, 10, 20, 40, 60, 80, 100), smoothed and non-smoothed models, regression and model trees). A total of 540 models were produced.

Based on the qualitative criterion, 20 out of the total 540 models were selected for further evaluation. Regarding the previously described selection principles of the second and the third evaluation criterion, the model for the prediction abundance of hemi-epidaphic Collembola and the model for the prediction of biomass of anecic earthworms were selected for final interpretation.

### Hemi-epedaphic Collembola model

The species of Collembola identified at Foulum in 2002 and 2003 are listed in Table 4. The model selected for predicting the abundance of hemi-epidaphic Collembola had the highest correlation coefficient for a cross-validated model (0.59) among the models for the prediction of Collembola functional groups. The model had two hierarchical levels (Fig. 1) and the relationships between the model's elements have been shown to be logical, and are extensively explained in the discussion section of this paper. The model shows that the abundance of hemi-epidaphic Collembola increased at the end of the maize growing season (time since sowing criteria), while higher organic matter content and higher pH tended to increase their abundance at this time as well as in the spring. A greater abundance of Collembola was also noted in early autumn if the crop was located near the isogenic non-*Bt* maize variety Monumental. A statistical description of the model attributes is given in Fig. 2.

### Anecic earthworms

At the Foulum site four species were identified during the sampling period: *Aporrectodea caliginosa*,

*Aporrectodea longa, Allobophora rosea* and *Lumbricus terrestris. Aporrectodea caliginosa* adults and *L. terrestris* were classified as anecic earthworms and represented around 80% of the earthworm biomass at this site (Krogh et al., 2007 in this issue). The model selected had the highest correlation coefficient among the earthworm functional groups ($r^2 = 0.83$) and it had three hierarchical levels

**Tabel 4.** List of Collembola species

| Functional group | Species |
| --- | --- |
| Euedaphic | *Pseudosinella alba* |
| | *Parisotoma terricola* |
| | *Willemia intermedia* |
| | *Mesaphorura* sp. |
| | *Protaphorura meridiata* |
| | *Neelus minimus* |
| | |
| | *Arrhopalites caecus* |
| Eu-hemiedaphic | *Folsomia bisettosa* |
| | *Folsomia fimetaria* |
| | |
| Hemiedaphic | *Heteromurus nitidus* |
| | *Parisotoma notabilis* |
| | *Sminthurinus aureus* |
| | *Frisea mirabilis* |
| | |
| Hemi-epedaphic | *Entomobrya lanuginosa* |
| | *Lepidocyrtus cyaneus* |
| | *Isotoma viridis* |
| | *Isotoma nivalis* |
| | *Ceratophysella succinea* |
| | *Ceratophysella denticulata* |
| | |
| Epiedaphic | *Sminthurinus elegans* |
| | *Bourletiella hortensis* |

(Fig. 3). The pattern of attributes included in the model is extensively explained in the discussion section of this paper. The model shows that the observed anecic earthworm species were influenced by the proportion of clay and silt in the soil and soil pH (i.e., preferring a near neutral pH). The farming practices seemed to have had no effect on their biomass. The biomass of the anecic functional group was greater in early autumn compared to spring or late autumn. This model consists of attributes which belong to the same groups of attributes as in the Collembolan model. A brief analysis of the statistical properties of attributes included in the earthworm model is given in Fig. 2.

## Discussion

### Hemi-epiedaphic Collembola

The primary factors influencing Collembola abundance were the sampling date, soil properties (such as organic matter content, texture or pH) and maize variety. The main predictor was sampling date within year with higher abundance observed in autumn ($>89$ days since sowing) compared to spring ($\leqslant89$ days since sowing). This can be explained by the greater vegetation biomass present in autumn and the consequently greater resources for microarthropods, combined with favorable pedo-climatic conditions (Hopkin, 1997). Time since sowing is thus mostly an indirect parameter measuring the available resources for Collembola. This hypothesis is supported by the influence of organic matter content of the soil, which appears only in spring, when the amount of
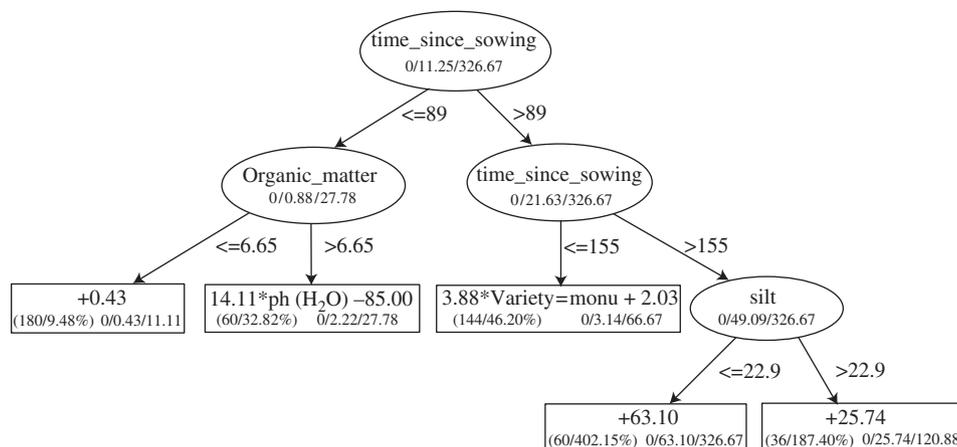


**Figure 1.** Regression model for predicting the abundance (number kg$^{-1}$ soil) of hemi-epiedaphic Collembola. The additional information given in nodes is min/mean/max of Collembolan abundance, while this information is extended with number of examples and root mean square error in the leaf divided by absolute deviation of all examples in leaves. Monu is an abbreviation for the maize variety *Monumental*.
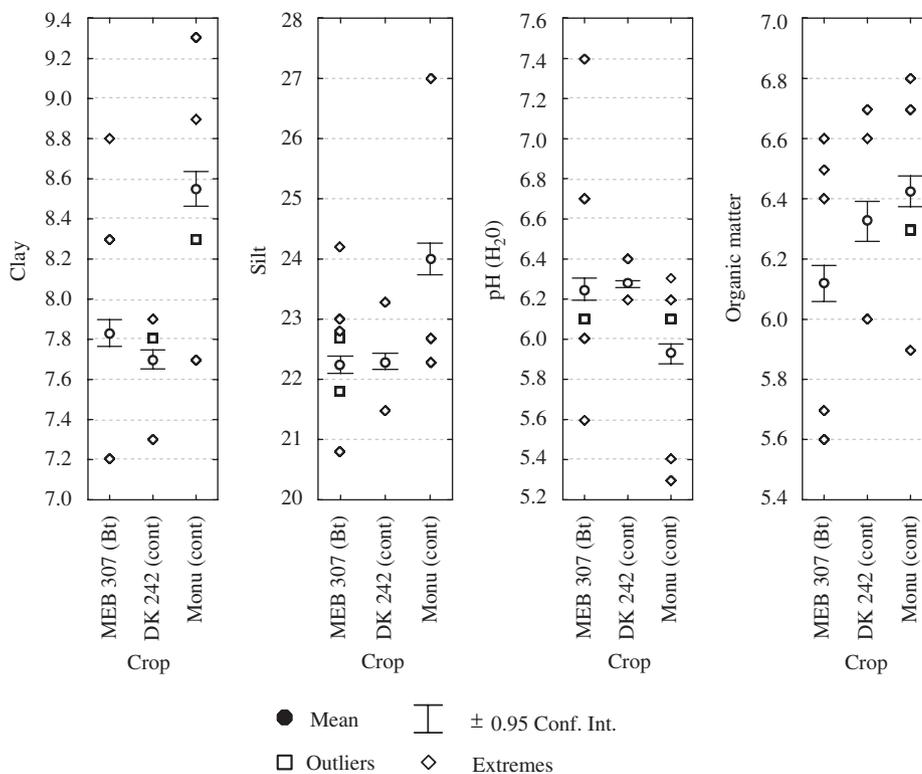
**Figure 2.** Box–Whisker diagrams for attributes included in the hemi-epiedaphic Collembola regression model (Fig. 1).
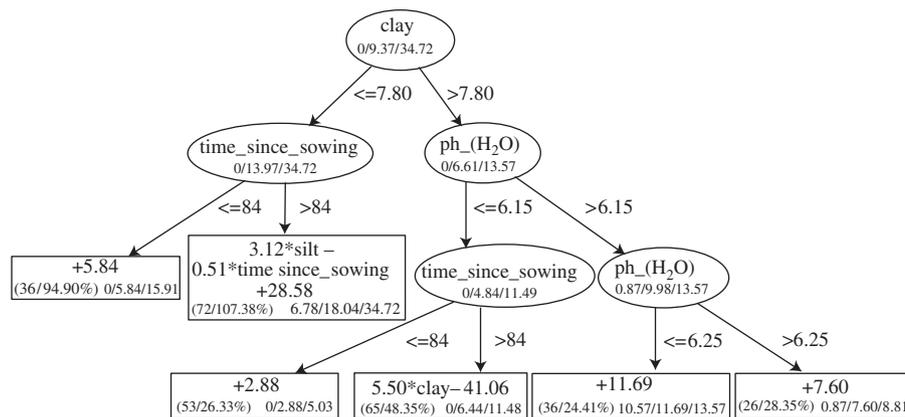


**Figure 3.** Regression model for predicting the biomass of anecic earthworms (g m$^{-2}$). The additional information given in nodes is min/mean/max of earthworm biomass, while this information is extended with number of examples and root mean square error in the leaf divided by absolute deviation of all examples in leaves.

detritus coming from the growing maize is small. These findings are in good agreement with the literature, which reports that food quality and the quantity of organic matter from the vegetation are key factors influencing soil microarthropods in agroecosystems (Garrett et al., 2001; Renaud et al., 2004). In this context, it is not surprising that maize variety appeared as a predicting attribute in the classification tree model for collembolan abundance (Fig. 1), with an increase in abundance in 2002 when the variety Monumental was grown. Indeed, variety effects are possible in the context of *Bt* maize introduction where significant differences have been noticed between crop yields due to pleitropy, with a slightly higher nitrogen content for MEB 307 compared to Monumental (Vercesi et al., 2006). This kind of effect, particularly differences in maize lignin content, has also been previously mentioned by several authors working on *Bt* effects on non-target soil organisms (Escher et al. 2000; Saxena and Stotzky, 2001; Zwahlen et al., 2003; Poerschmann et al., 2005).

Normally, one would like to test whether the influence of variety on abundance is significant by estimating the predictive power of induced models. We have used the cross-validation procedure for this purpose. With this procedure the data is split into 10 parts. All parts except one are combined, a model is developed from these data and subsequently tested on the remaining part. Once this procedure is repeated leaving out all 10 parts in turn, the average correlation coefficient is computed and can be used as a measure of model's predictive power.

## Anecic earthworms

Regarding the models, three driving forces seemed to influence *L. terrestris* biomass in Foulum: texture, sampling date and pH (Fig. 3). Clay content, which ranged from 7.2% to 9.3% in this soil, was selected as the most important attribute for the prediction of the biomass of anecic earthworms (i.e., it was on the highest hierarchical level in the model). Texture has been previously shown to influence earthworm biomass because of effects on other critical parameters for earthworms, mostly soil water content (Khalaf El-Duweini and Ghabbour, 1965). Earthworms seem unable to survive with prolonged periods of water potential less than $-0.10$ MPa (Lavelle, 1971) and it has been shown that *L. terrestris*, in particular, prefers a soil moisture around 30% under laboratory conditions (Lowe and Butt, 2005). Many studies have concluded that increasing the percentage of fine particles, like clay, would increase the biomass of earthworms (Nordstrom and Rundgren, 1974; Hendrix et al., 1992; Baker et al., 1998). However, in our case, even though the clay content was a key parameter influencing anecic biomass, it was clearly not enough to predict biomass. Indeed, there was no direct linear correlation between the clay content and the biomass. Thus, low clay content did not necessarily lead to low anecic earthworm biomass. In this study, if the percentage of clay is low ($\leqslant 7.8$), the sampling date should also be taken into account in order to predict biomass. For samples collected in spring ($< = 84$ days since sowing), the biomass will be moderate ($5.84$ g m$^{-2}$), while for autumn samples ($> 84$ days since sowing), the biomass can increase to reach its maximum ($18.04$ g m$^{-2}$ with the regression tree model) or vary depending on the silt content and the sampled year (greater biomass in autumn 2002 compared to autumn 2003). It is interesting to note that the seasonal effect (autumn/spring sampling), appears higher in the model, and thus has a heavier influence on anecic biomass compared to the inter-annual effect (autumn 2002/autumn 2003). Indeed, it is well known that in temperate arable ecosystems, anecic earthworms reach their minimum in winter, due to low temperature, and their maximum in autumn, after spring and summer reproduction and development. On the other hand pH, which varied between 5.3 and 7.4 in this study, surfaced as a good parameter to predict the biomass of earthworms. In Foulum, the optimal pH appears to be between 6.15 and 6.25, which is not really surprising, as anecic earthworms are known to be more abundant at neutral pH (Salmon, 2001). But here again, the pH parameter can be moderated by other parameters, such as the sampling date and clay content. Agricultural practices attributes, such as tillage or maize variety were not found to be good predictors for anecic biomass and thus appeared to have no influence on earthworms. With respect to *Bt* maize, these findings are in agreement with previous studies (Zwahlen et al., 2003).

## Conclusions

Predictions concerning anecic worms were based on basic soil properties (texture, pH) and on the biological cycles of earthworms, whereas the predictions concerning hemi-epiedaphic Collembola were instead based on the available organic resources. It appears that earthworms are influenced by (or influence) environmental parameters at larger spatio-temporal scales compared to smaller detritivores like Collembola, as suggested by Lavelle et al. (1993). However the results presented in this paper are valid only for this particular Foulum situation. To make a more general statement about the results and conclusions, this type of analysis needs to be applied to different data sets varying with climate, soil properties, and other environmental attributes.

## Acknowledgments

## References

Andersen, M.N., Sausse, C., Lacroix, B., Caul, S., Messéan, A., 2007. Agricultural studies of GM maize

and the field experimental infrastructure of ECOGEN. Pedobiologia 51, 175–184.

Andow, D.A., Hilbeck, A., 2004. Science-based risk assessment for non-target effects of transgenic crops. Bioscience 54, 637–649.

Baker, G.H., Carter, P.J., Curry, J.P., Cultreri, O., Beck, A., 1998. Clay content of soil and its influence on the abundance of *Aporrectodea trapezoides* Dugès (Lumbricidae). Appl. Soil Ecol. 9, 333–337.

Berlese, A., 1905. Apparicchio per raccogliere presto ed in gran numero di piccoli artropodi. Redia 2, 85–89.

Bouché, M.B., 1972. Lombriciens de France, Ecologie et systématique, INRA, 72–2. INRA, Paris.

Bouché, M.B., 1977. Stratégies lombriciennes. In: Lohm, U., Persson, T. (Eds.), Soil Organism as Components of Ecosystems. Ecology Bulletin/NFR, Stockholm, pp. 122–132.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont.

Brussaard, L., Bouwman, L.A., Geurs, M., Hassink, J., Zwart, K.B., 1990. Biomass, composition and temporal dynamics of soil organisms of a silt loam soil under conventional and integrated management. Neth. J. Agr. Sci. 38, 283–302.

Cortet, J., Poinsot-Balaguer, N., Viaux, P., Chabert, A., Beaufreton, Ch., Cancela da Fonseca, J.P., 2002. Impact of agricultural practices on the biodiversity of soil microarthropods: the example of French arable crops. Eur. J. Soil Biol. 38, 239–244.

Cortet, J., Andersen, M.N., Caul, S., Griffiths, B.S., Joffre, R., Lacroix, B., Sausse, C., Thompson, J., Krogh, P.H., 2006. Decomposition processes under *Bt* (*Bacillus thuringiensis*) maize: results of a multi-species experiment. Soil Biol. Biochem. 38, 195–199.

Cortet, J., Griffiths, B.S., Bohanec, M., Demšar, D., Andersen, M.N., Caul, S., Birch, A.N.E., Pernin, C., Tabone, E., de Vaufleury, A., Ke, X., Krogh, P.H., 2007. Evaluation of effects of transgenic *Bt* maize on microarthropods in a European multi-site experiment. Pedobiologia 51, 207–218.

Debeljak, M., Džeroski, S., Jerina, K., Kobler, A., Adamic, M., 2001. Habitat suitability modeling for red deer (*Cervus elaphus* L.) in South-Western Slovenia with classification trees. Ecol. Model. 138, 321–330.

Debeljak, M., Cortet, J., Demšar, D., Džeroski, S., 2005. Using data mining to assess the effects of *Bt* maize on soil microarthropods. In: Hrebicek, J., and Racek, J., (Eds.), Networking enviromental information: EnviroInfo Brno 2005: Proceedings of the 19th International Conference Informatics for Environmental Protection, 7–9 September 2005, Brno, Czech Republic, pp. 615–620.

Dekkers, B.M., Van der Werff, P.A., Van Amelsvoort, P.A.M., 1994. Soil collembola and acari related to farming systems and crop rotations in organic farming. Acta Zool. Fennica 195, 28–31.

Demsar, D., Dzeroski, S., Larsen, T., Struyf, J., Axelsen, J.A., Bruus Pedersen, M., Krogh, P.H., 2006. Using multi-objective classification to model communities of soil microarthropods. Ecol. Model. 191, 131–143.

Džeroski, S., 2001. Applications of symbolic machine learning to ecological modelling. Ecol. Model. 146, 263–273.

El Titi, A., Ipach, U., 1989. Soil fauna in sustainable agriculture : results of an integrated farming system at Lautenbach, F. F.R.G. Agric. Ecosyst. Environ. 27, 561–572.

Escher, N., Käch, B., Nentwig, W., 2000. Decomposition of transgenic *Bacillus thuringiensis* maize by microorganisms and woodlice *Porcellio scaber* (Crustacea: Isopoda). Basic Appl. Ecol. 1, 161–169.

Garrett, C.J., Crossley Jr, D.A., Coleman, D.C., Hendrix, P.F., Kisselle, K.W., Potter, R.L., 2001. Impact of the rhizosphere on soil microarthropods in agroecosystems on the Georgia piedmont. Appl. Soil Ecol. 16, 141–148.

Gisin, H., 1943. Oekologie und Lebensgemeinschaften der Collembolen im Schweizerischen Excursionsgebiet Basels. Rev. Suisse Zool. 50, 131–224.

Gordon, B., 2004. GM crops – an international perspective on the economic and environmental benefits. In: van Emden, H.F., Gray, A.J., (Eds.), GM Crops – Ecological Dimensions, Aspects in Applied Biology, vol. 74, pp. 3–24.

Griffiths, B.S., Caul, S., Thompson, J., Birch, N., Scrimgeour, C., Andersen, M.N., Cortet, J., Messean, A., Sausse, C., Lacroix, B., Krogh, P.H., 2005. Microbial community structure, protozoa and nematodes in soil from field plots of genetically modified maize expressing the *Bacillus thuringiensis* toxin. Plant Soil 275, 135–146.

Hendrix, P.F., Parmelee, R.W., Crossley Jr., D.A., Coleman, D.C., Odum, E.P., Goffman, P.M., 1986. Detritus food webs in conventional and no-tillage agroecosystems. Bioscience 36, 374–380.

Hendrix, P.J., Mueller, B.R., Bruce, R.R., Langdale, G.W., Parmelee, R.W., 1992. Abundance and distribution of earthworms in relation to landscape factors on the Georgia Piedmont, USA. Soil Biol. Biochem. 24, 1357–1361.

Hopkin, S.P., 1997. Biology of the springtails, Insecta: Collembola. Oxford University Press, Oxford.

Jagers op Akkerhuis, G.A.J.M., De Ley, F., Zwetsloot, H.J.C., Ponge, J.F., Brussaard, L., 1988. Soil microarthropods (Acari and Collembola) in two crop rotations on a heavy marine clay soil. Rev. Ecol. Biol. Soil 25, 175–202.

Jerina, K., Debeljak, M., Dzeroski, S., Kobler, A., Adamic, M., 2003. Modeling the brown bear population in Slovenia: A tool in the conservation management of a threatened species. Ecol. Model. 170, 453–469.

Kampichler, C., Dzeroski, S., Wieland, R., 2000. Application of machine learning techniques to the analysis of soil ecological data bases: relationships between habitat features and collembolan community characteristics. Soil Biol. Biochem. 32, 197–209.

Khalaf El-Duweini, A., Ghabbour, S.I., 1965. Population density and biomass of earthworms in different types of Egyptian soils. J. Appl. Ecol. 2, 271–287.

Krogh, P.H., Griffiths, B., Demšar, D., Bohanec, M., Debeljak, M., Andersen, M.N., Sausse, C., Birch, A.N.E., Caul, S., Holmstrup, M., Heckmann, L.H., Cortet, J., 2007. Responses by earthworms to reduced tillage in herbicide tolerant maize and *Bt* maize cropping systems. Pedobiologia 51, 219–227.

Lavelle, P., 1971. Etude préliminaire de la nutrition d'un ver de terre africain *Millsonia anomala* (Acanthodrilidae, Oligochètes). Ann. Zool. Ecol. Anim., Hors Sér., 133–145.

Lavelle, P., 1981. Stratégies de reproduction chez les vers de terre. Acta Oecol. Oec. Gen. 2, 117–133.

Lavelle, P., Blanchart, E., Martin, A., Martin, S., Barois, I., Toutain, F., Spain, A.V., Schaefer, R., 1993. A hierarchical model for decomposition in terrestrial ecosystems. Application to soils of the humid tropics. Biotropica 25, 130–150.

Lowe, C.N., Butt, K.R., 2005. Culture techniques for soil dwelling earthworms: a review. Pedobiologia 49, 401–413.

Neher, D.A., Barbercheck, M.E., 1999. Diversity and function of soil mesofauna. In: Collins, W.W., Qualset, C.O. (Eds.), Biodiversity in Agroecosystems. CRC Press, New York, pp. 27–47.

Nordstrom, S., Rundgren, S., 1974. Environmental factors and lumbricid associations in southern Sweden. Pedobiologia 9, 188–214.

Peterson, H., 1978. Some properties of two high-gradient extractors for soil microarthropods, and an attempt to evaluate their extraction efficiency. Nat. Jutlandica 4, 95–121.

Poerschmann, J., Gathmann, A., Augustin, J., Langer, U., Gorecki, T., 2005. Molecular composition of leaves and stems of genetically modified *Bt* and near-isogenic non-*Bt* maize-Characterization of lignin patterns. J. Environ. Qual. 34, 1508–1518.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 5, 239–266.

Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence. World Scientific, Singapore, pp. 343–348.

Renaud, A., Poinsot-Balaguer, N., Cortet, J., Le Petit, J., 2004. Influence of four soil maintenance practices on Collembola communities in a Mediterranean vineyard. Pedobiologia 48, 623–630.

Sabatini, M.A., Rebecchi, L., Cappi, C., Bertolani, R., Fratello, B., 1997. Long-term effects of three different continuous tillage practices on collembola populations. Pedobiologia 41, 185–193.

Salmon, S., 2001. Earthworm excreta (mucus and urine) affect the distribution of springtails in forest soils. Biol. Fertil. Soils 34, 304–310.

Saxena, D., Stotzky, G., 2001. *Bt* corn has a higher lignin content than non-*Bt* corn. Am. J. Bot. 88, 1704–1706.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., Dean, D.H., 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. R. 62, 775–806.

Schrader, S., Lingnau, M., 1997. Influence of soil tillage and soil compaction on microarthropods in agricultural land. Pedobiologia 41, 202–209.

Stankovski, V., Debeljak, M., Bratko, I., Adamič, M., 1998. Modelling the population dynamics of red deer (*Cervus elaphus* L.) with regard to forest development. Ecol. Model. 108, 145–153.

Vercesi, M.L., Krogh, P.H., Holmstrup, M., 2006. Can *Bacillus thuringiensis* (*Bt*) corn residues and *Bt*-corn plants affect life-history traits in the earthworm *Aporrectodea caliginosa*? Appl. Soil Ecol. 32, 180–187.

Wang, Y., Witten, I.H., 1997. Induction of model trees for predicting continuous classes. Proceedings of the poster papers of the ECML 97. University of Economics, Faculty of Informatics and Statistics, Prague.

Witten, I.H., Frank, E., 1999. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco.

Zwahlen, C., Hilbeck, A., Howald, R., Nentwig, W., 2003. Effects of transgenic *Bt* corn litter on the earthworm *Lumbricus terrestris*. Mol. Ecol. 12, 1077–1086.