Metodološki zvezki, Vol. 2, No. 1, 2005, 147-159

Using Machine Learning to Predict the Impact of Agricultural Factors on Communities of Soil Microarthropods

Damjan Demšar[1], Sašo Džeroski[1], Paul Henning Krogh[2], and Thomas Larsen[2]

Abstract

With the newly arisen ecological awareness in the agriculture the sustainable use and development of the land is getting more important. With the sustainable use of soil in mind, we are developing a decision support system that helps making decisions on managing agricultural systems and is able to handle both conventional and genetically modified crops as a part of the ECOGEN project. The decision support system considers economical and agricultural factors and actions including crop selection, crop sequence, pest and weed control actions etc. For such decision support system to work, it needs modules that predict results of different agricultural actions. One of the most important factors for sustainable use and fertility of soil is soil flora and fauna. Any change of that community can influence the short or long term soil fertility and soil usability.

With soil fauna being one of the most important factors we first need to model it. However, since the function of the individual species is not known, the only action we have is to try and model the community of soil fauna. We start by modelling the community soil microarthropods. For that goal we used machine learning methods - regression trees, model trees and linear equations. We identified previous crops and time since different kinds of tillage as the most important factors for the community of soil microarthropods.

1 Introduction

The possible use of genetically modified (GM) plants in agriculture needs in-depth investigations of ecological and economic consequences (Birch, 2003 and ECOGEN). The investigations are important for both the European Commission

[1] Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia, {damjan.demsar, saso.dzeroski}@ijs.si

[2] Department of Terrestrial Ecology, National Environmental Research Institute, Roskilde, Denmark, {phk, thl}@dmu.dk

(EC), who needs specifications for GM-plant risk assessment, and to farmers and the public who are concerned about the possible ecological and economic implications. Crop production involves complex decision-making processes, which require and justify the application of decision support systems.

The ECOGEN project (*Soil ecological and economic evaluation of genetically modified crops*) is an EC-funded project aimed at combining simple lab tests, studies of multi-species model mesocosms ecosystems, and field studies to acquire realistic knowledge about economic and ecological impacts of GM crops on the soil. Economic trade-offs are assessed and related to ecological effects. The economic and ecological knowledge gained in ECOGEN will be combined into a rule-based model for a decision support tool.

The goals of the ECOGEN project are to:

1. Provide ecological and economical assessment and comparison of integrated cropping systems using GM or conventional crops, respectively.
2. Provide an ecological risk assessment of a GM cropping system and a conventional cropping system for the soil ecosystem based on single species tests, multispecies tests and long-term field investigations.
3. Adapt existing ecotoxicity testing tools to GM plant material and validate their use.
4. Provide economic assessment of GM crops and conventional crops with respect to a quantification of the expected trade-offs between the two and the implications for the EU Agriculture Policy.

Finally, we wish to incorporate ecological knowledge from single species tests, multispecies tests, and field investigations, as well as economic information from farming practices into a *rule-based model* to be used for predictions of economic decision-making processes and ecosystem behaviour.

In this paper we present the current generation of the microarthropod models that are to be used as a part of decision support model. These models will be used to judge the results of the agricultural actions, and thereby act as an input to the upper levels of the decision support system.

2 Data

We combined the two available datasets: The first dataset describes four experimental farming systems (Foulum experimental station, Denmark) in the years from 1989 to 1993, allocated to 15 fields, with pesticide use in a conventional system and in two integrated farming systems and no pesticide use on the other (organic) fields, with 530 microarthropod samples collected (Krogh, 1994). The second dataset describes several organic farms (Foulum and Flakkebjerg experimental stations plus various farms in Jutland) in the year 2002.

Table 1: The available attributes.

Attribute	Explanation
soil_JB	soil classification number
samp_time	1 = March - April, 2 = May - June, 3 = July - August, 4 = September - November
ba	winter barley
be	beets/carrots
ca	cattle
cc	catch crop
ch	chicory
chgr	chicory+grass
clgr	clover+grass
fa	fallow
gr	grass
le	leeks
lu	lupin
oa	oates
pe	peas
po	potatoes
ra	rape
rd	radish
ry	rye
sba	spring barley
sf	stubble field
sh	Sheep
Si	silage/hay
Swh	spring wheat
Tc	Triticale
Wh	winter wheat
Wc	whole crop
O	seed bed (<1 mo)
Seha	seed bed harrowed
Sepl	seed bed plowed
Soha	soil treatment harrowed
Sopl	soil treatment plowed
Pesticide	Pesticide. 1=fields in a rotation where pesticides are used. 0=no pesticide
tr_packing	packing (months since) transformed using: $\frac{months - 10}{10}$
tr_shal_till	shallow tillage (weed harrowing etc) 0-5 cm layer (months since) transformed using: $\left(\frac{months - 10}{10}\right)^4$
tr_subshal_till	subshallow tillage 5-10 cm layer (months since) transformed using $\left(\frac{months - 10}{10}\right)^2$
tr_deep_till	deep tillage (plowing, rotovation) >10 cm layer (months since) transformed using $\left(\frac{months - 10}{10}\right)^2$
fert_lev	low=0, normal=1, high=2.
fert_type	no=0, solid=1, liquid=2
fert_time	fertilizaton time (mo)
crop_1	crop prev year
ca_1	no cattle=0, cattle=1
sotr_1	no treat=0, s or a=1, s and a=2
crop_2	crop prev 2nd year
ca_2	no cattle=0, cattle=1
sotr_2	no treat=0, s or a=1, s and a=2
crop_3	crop prev 3rd year
ca_3	no cattle=0, cattle=1
sotr_3	no treat=0, s or a=1, s and a=2

430 samples were collected. To those datasets we added newly available data from 2003, giving us a total of 1330 samples, of which 1192 were suitable for predicting Acari species, 1214 for prediction Collembolan species and 1138 for predicting biodiversity.

The sampling was replicated for each field. The distance between each sample was 5 m and all samples were collected within a 20x20 m area. The distance to hedges and ditches was at least 10 m. Sampling was performed in the upper 5.5 cm soil layer. The sampling containers measured 6 cm in diameter. Sampling was done using a split soil corer and extraction was performed using a MacFadyen high gradient heat extractor.

The datasets available for the study include the agricultural measures (attributes), for example, packing, tillage, fertilizer and pesticide use, crops planted and cattle grazing. The history of crops and grazing for the last 3 years is also available. The datasets also contain environmental variables describing the circumstances of the samples where community data on soil microarthropods have been produced. The variables used to model microarthropods are listed in Table 1, and were selected by domain experts. The transformations used on some attributes (different forms of tillage) were used to simulate the occasional non-linear reducing impact of tillage (different powers simulate differently steep curves of impact). The dataset also includes measured species (listed in Table 2). Some species were grouped into Acari group (mites), the rest of the measured species belong into Collembolan group (springtails) and all were used to calculate biodiversity using formula (2.1):

$$H = -\sum_{i=1}^{S} p_i \cdot \log_2 p_i \qquad (2.1)$$

Where p_i represents the proportion of abundance of species i of total sample abundance and S represents total number of species in sample.

3 Methodology

In this section we describe the machine learning methods we used to produce the models predicting the number of springtails, the number of mites and their biodiversity. We describe regression trees (as used in M5' (Wang and Witten, 1997) in Weka 3.2 (Witten and Frank ,1999). In parallel we describe model trees (also used in M5'), which are based on regression trees, by highlighting the differences between regression trees and model trees

Regression trees are used to represent piecewise constant functions. Model trees on the other hand represent piecewise linear functions (model trees are sometimes also called regression trees, however we use model trees, to avoid confusion between models). Both predict the value of a dependant variable (class) from values of independent variables (attributes).

Table 2: The observed species ([1]Collembolan groups – springtails [2]Acari groups – mites).

Abbreviation	Species	Abbreviation	Species
Iang[2]	*Isotoma anglicana*	Apygm[2]	*Anurida pygmaea*
Ipalu[2]	*Isotomurus palustris*	Iminor[2]	*isotomiella minor*
Hdent[2]	*Ceratophysella denticulata*	Hniti[2]	*Heteromurus nitidus*
Hsuc[2]	*Ceratophysella succinea*	Tquad[2]	*Stenaphorura quadrispina*
Xarma[2]	*Hypogastrua sp.*	Nmini[2]	*Neelus minimus*
Llanu[2]	*Lepidocyrtus lanunginosus*	Saure[2]	*Sminthurinus aureus*
Lcyan[2]	*Lepidocyrtus cyaneus*	Fspino[2]	*Folsomia spinosa*
Seleg[2]	*Sminthurinus elegans*	Cterm[2]	*Cryptopygus thermophilus*
Onych[2]	*Protaphorura sp.*	Will[2]	*Willemia sp.*
Sviri[2]	*Sminthurus viridis*	Ocinct[2]	*Orchesella cincta*
Sminsp[2]	*Smint. Sp.*	Owillo[2]	*Orchesella villosa*
crypt[1]	Cryptostigmata *(*Oribatida mite*)*	Nmusco[2]	*Neanura*
prost[1]	Prostigmata *(*Actinedida mite*)*	Psexoc[2]	*Pseudosinella sexoculata*
Tull[2]	*Mesaphorura sp.*	Iprod[2]	*Isotomodes productus*
Inot[2]	*Isotoma notabilis*	Iarma[2]	*Isotomodes armata*
Entosp[2]	*Entomobrya sp.*	IBiset[2]	*Isotomodes bistosus*
Fmirab[2]	*Friesea mirabilis*	Fquad[2]	*Folsomia quadrioculata*
ast[1]	Astigmata (Acaridida mite)	Icilia[2]	*Isotomurus sp.*
meso[1]	Mesostigmata (Gamasida mite)	Tomosp[2]	*Tomoserus sp.*
Ffim[2]	*Folsomia fimetaria*	Tflav[2]	*Tomocerus flavescens*
Palba[2]	*Pseudosinella alba*	Tminor[2]	*Tomocerus minor*
Bparv[2]	*Brachystomelle parvula*		

While usual regression methods (linear and non-linear regression) fit a single function to whole set of data, regression trees partition the data space into hyper-rectangles (multidimensional) and fit a model for each partition (in our case regression trees fit a constant and model trees fit a linear function). To achieve the partition the tree is build from inner nodes, that each include a test of particular attribute on it value. The terminal nodes, also called leaves, on the other hand include models.

In order to predict the value of class variable of new (or even test) example the evaluation of the tree starts from the root node. In each inner node the test is performed and according to the result of the test a particular branch is followed from that inner node. This process is repeated until a terminal node (leaf) is reached. In regression trees the predicted value of the class variable is the constant predicted by the leaf node, while in model trees the predicted value is the value of evaluated linear equation (which is a part of the leaf).

The regression trees are constructed from the top (root) down, starting with the all training examples and then continues recursively for each subtree. At each step the most discriminating attribute is selected, and subsets of the training examples are created according to the values of the selected attribute. If the selected attribute is continuous then a threshold value is selected and two branches (and two subtrees) are created. If the attribute is nominal (discrete) then either a branch is constructed for each possible value of attribute, or two subsets of the values are

created (again the most discriminate) and two branches are created (all in-between types are possible, but uncommon).

The most discriminating discrete attribute or continuous attribute test is the one that reduces most the variance of the values of the class variable. For continuous attributes, the values of the attribute that appear in the training set are considered as thresholds. For the subsets of training examples in each branch, the tree construction algorithm is called recursively. Tree construction stops when the variance of the class values of all examples in a node is small enough (or if some other stopping criterion is satisfied). Such nodes are called leaves and are labelled with a model (constant or linear equation) for predicting the class value.

An important mechanism used to prevent trees from over-fitting data is tree pruning. Pruning can be used during tree building (pre-pruning) or after the tree has been built (post-pruning). Usually, a minimum number of examples in branches can be set for pre-pruning and confidence level in error estimates in leaves for post-pruning.

A number of systems exist for inducing regression trees from examples, such as CART (Breiman, 1984) and M5 (Quinlan, 1993). M5 is one of the most well known programs for regression tree induction. We used the system M5' (Wand and Witten,1997), a re-implementation of M5 within the software package WEKA (Witten and Frank, 1999): simple model trees have simplified equations and are induced with the –U option, complex model trees are induced by M5' with default parameter settings. We also used regression trees and linear regression. The sizes of both model and regression trees were regulated using post-pruning methods. The nearest neighbour method IBk (Aha and Kibler, 1991) with 1, 5 or 10 neighbours was used as a benchmark for comparing accuracy. Each method was applied to each of the three regression problems. For measuring the predictive performance of the model, we evaluated the correlation coefficient and several error measures using ten-fold cross-validation. We evaluated mean average error, root mean square error, and relative average error and root relative square error.

4 Results

4.1 Acari models

First we produce the models describing the dependence of abundance of mites (Acari) to agricultural factors. Correlation factors and different error measures of different models can be seen in Table 3. As can be seen, the best correlation and are produced by nearest neighbour methods with 1 or 5 neighbours taken into consideration (depending on the chosen error measure). However nearest neighbours method does not create any model and thereby cannot be used to gain any new knowledge or even describe any already known knowledge. From the

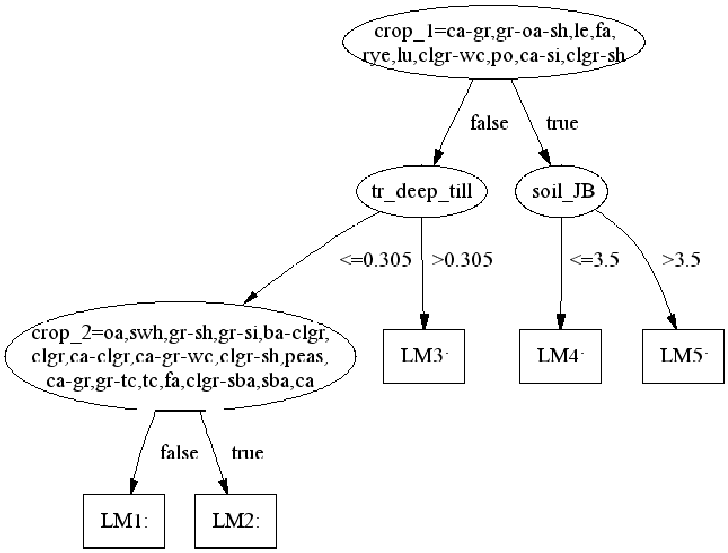
descriptive models the best of the produced models (when only accuracies are taken into account) is the model tree produced by M5 with default parameters. But the default model tree has rather complex equations attached to the leafs of the tree, and is thereby rather difficult to understand. It is even harder to judge which of the attributes used in the equations is most influential in each of the equations. To gain more understanding we prefer the simple model trees (or even regression trees), since the exacter models can be harder or even to hard to interpret. Here we can take the simple model tree (produced with default pruning) seen in Figure 1 and take only slight performance hit, but gain a lot in understandability. The complex model tree is the same size than the simple model tree, but has more complex linear equations in leafs. Usual equations in the complex model tree have the length of 20 or more, while the equations in the simple trees have the length of about 5.

Both the Acari simple model tree (Figure 1) and Acari regression tree (Figure 2) are very similar, however the model tree has much better correlation and much lower (10% or more) error measures. If we look at both models we can see that the most important factors for the community of soil mites are crops in previous years and tillage (especially deep tillage and subshallow tillage). And from the questions that we can ask when interpreting the models we can gain some new understanding of the problem that will help us with the construction of the final decision support model. For example, we found out (from the model and experts) that the previous years crops that were covered with grass or grazed by cattle are usually more undisturbed, and produce more food, which is better for mites. Also, such fields leave a lot of decomposing matter (food) in the soil, which helps mites in the following years and even speeds up the recovery of mites after tillage (which is the most important negative factor for mites). That one of the reasons that generally, it is very beneficial for the soil fauna and microbial life that a field is resting for a few years or is covered with clover where there is a minimum of tillage.

Table 3: Correlation coefficient and errors of Acari models (best models are in bold, shown models in italics). MAE- mean average error, RMSE – root mean square error, RAE – relative average error, RRSE – root relative square error.

name	size	corr	MAE	RMSE	RAE	RRSE
ibk 1	**0**	**0.668**	**20703.395**	**43540.345**	**57.189**	**75.112**
ibk 5	0	0.666	20953.490	43510.279	57.880	75.060
ibk 10	0	0.617	22337.137	45650.914	61.702	78.753
M5 linear equation	1	0.626	24773.783	45255.734	68.433	78.071
M5 model tree	5	0.650	22315.132	44137.612	61.641	76.142
M5 model tree simple	*5*	*0.643*	*22465.097*	*44560.573*	*62.056*	*76.872*
M5 model tree pruning 5	**5**	**0.610**	**23973.521**	**46070.750**	**66.222**	**79.477**
M5 model tree pruning 5 simple	5	0.606	24032.788	46284.588	66.386	79.846
M5 model tree pruning 15.5	1	0.579	25254.081	47268.335	69.760	81.543
M5 model tree pruning 15.5 simple	1	0.576	25300.044	47409.002	69.887	81.786
M5 regression tree	10	0.604	23434.691	46257.361	64.734	79.799
M5 regression tree pruning 15	*4*	*0.538*	*26215.648*	*48881.437*	*72.416*	*84.326*

While domain experts agree with the models and have mostly learned only the ranking of factors we have gained much knowledge that will help us in modelling mites in order to help decision makers choose the right decisions.




```
LM1:acari = 11300 + 4250samp_time - 5670wh=1 - 29300tr_subshal_till
          - 11200crop_1=wc,sba,ra,pe,wh,gr-sh,ba-clgr,ba-gr-sh,clgr,
                        ca-clgr,clgr-si,ca-gr,gr-oa-sh,le,fa,rye,lu,
                        clgr-wc,po,ca-si,clgr-sh
          + 8580crop_1=pe,wh,gr-sh,ba-clgr,ba-gr-sh,clgr,ca-clgr,
                        clgr-si,ca-gr,gr-oa-sh,le,fa,rye,lu,clgr-wc,
                        po,ca-si,clgr-sh
          + 9630crop_2=ba,pe,gr,ba-ra,be,ba-gr,oa,swh,gr-sh,gr-si,
                       ba-clgr,clgr,ca-clgr,ca-gr-wc,clgr-sh,peas,
                       ca-gr,gr-tc,tc,fa,clgr-sba,sba,ca


LM2:acari = 35300 + 14000clgr=1 + 36600wh=1 - 20400fert_lev
           + 11100sotr_3=2,0

LM3:acari = 17800 - 8070o=0 - 11200tr_subshal_till
           + 6340crop_3=ba-gr,pe,ba-clgr,ba,gr,be,gr-sh,ba-ra,oa,gr-si,
                        ca-clgr,clgr,wh,tc,ba-clgr-sh,sba,ca-gr,ba-pe,
                        fa,clgr-wc,lu,gr-oa-sh,ca - 5140sotr_3=2,0


LM4:acari = 158000 - 55700soil_JB + 156000tr_subshal_till
          + 66400sotr_1=0


LM5:acari = 15300 + 48000o=0 - 26100sotr_1=0
          + 23200crop_2=tc,fa,clgr-sba,sba,ca
```

Figure 1: Acari model tree.

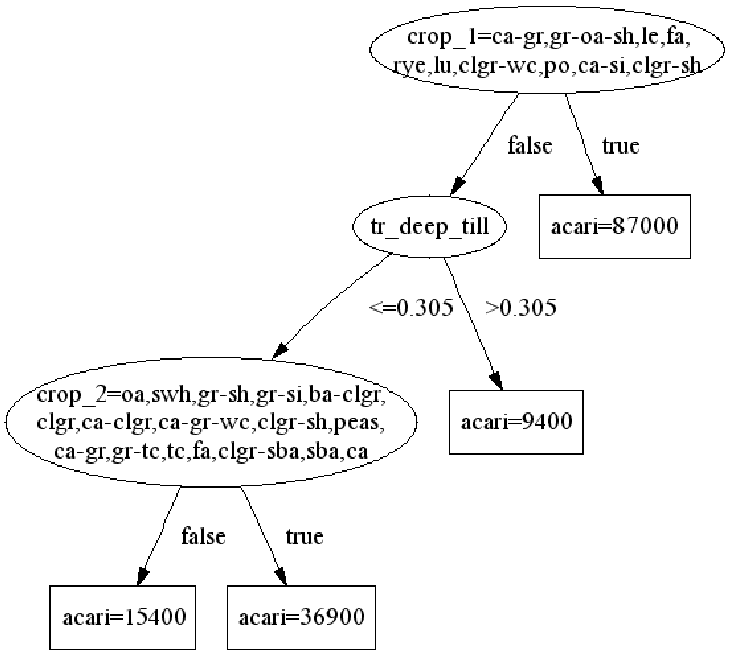


Figure 2: Acari regression tree.

4.2 Collembolan models

From the correlations and error measures of models describing collembolans (springtails) seen in Table 4, we can see that again the nearest neighbour method is the best, but only slightly outperforms the best of the descriptive models (in this case our preferred model – simple model tree). However since the simple model tree is too big and regression tree size can be better regulated with pruning we show only the regression tree (Figure 3), which is significantly worse than the simple model tree, but has similar main structure. From the shown regression tree and from the simple model tree we can recognize as the most important factors for Collembolan species are again previous crops (in descending order with time past) and tillage (especially subshallow tillage and deep tillage). However the experts expected tillage to be more important than previous crops. The questions that follow from the models lead us to knowledge like:

- Deep tillage has less impact with some crops. Crops that include grass/clover provide protection even if the field is deep tilled because the sods will still be intact and the clover residues add a lot of nitrogen to the soil (enhances microbial life and thus the food base).
- In the case that the crops are still there in the current year it means that there has been no tillage plus clover fertilizes the soil. Lupin also fertilizes the soil

- Tillage injures/kill Collembolan by physical disruption and destroys their habitat (pathways in the soil are broken and the soil structure is destroyed).
- Collembolans are considered to be more sensitive to tillage than mites because their cuticula are softer. The Oribatid mites (Cryptostigmata) differ from other microarthropods by having a calcareous exoskeleton that protect them
- If tillage was in the past then the biomass/growth of the standing crop is likely to be bigger than is the field was recently tilled. Higher biomass=more Collembolan food. Sometimes this is counteracted by high ammonia concentrations in the fertilizer that can kill the Collembolan if the animals are exposed to it directly.

The experts liked the produced models; the only surprise was the fact that some effects can be seen even 6 months after subshallow tillage.

Table 4: Correlation coefficient and errors of Collembolan models.

name	size	corr	MAE	RMSE	RAE	RRSE
ibk 1	**0**	**0.647**	**17124.997**	**33295.410**	**62.692**	**78.359**
ibk 5	0	0.621	18107.641	33616.717	66.289	79.115
ibk 10	0	0.603	18693.767	33939.280	68.435	79.874
m5 linear equation	1	0.562	21471.686	35331.697	78.604	83.151
m5 model tree	1	0.583	20445.514	34737.713	74.848	81.754
m5 model tree simple	1	0.592	20516.408	34471.645	75.107	81.127
m5 model tree pruning 0.57	18	0.631	18853.541	33059.075	69.020	77.803
m5 model tree pruning 0.57 simple	**18**	**0.636**	**18836.467**	**33094.360**	**68.957**	**77.886**
m5 regression tree	22	0.533	20000.864	36151.180	73.220	85.080
m5 regression tree pruning 10	*5*	*0.442*	*22836.821*	*38130.009*	*83.602*	*89.737*

4.3 Biodiversity models

When we model biodiversity, again the nearest neighbour method has the best correlation (Table 5), but in this case the difference to the best descriptive model is quite big with correlation, however the model tree has better root mean square error, and the same is true for the simple model tree. Again the simple model tree is too big to show in this paper, so we show only the regression tree (Figure 4), which has considerably lower correlation and higher error measures, but is similar in the main structure. From the models we can identify the most important factor for biodiversity is (the lack of) subshallow and shallow tillage, according to the models, the next most important factor is the crop from three years ago, which is surprising to the experts and the only explanation they could find was, that the crops in the past define the cropping sequence and thereby even the current crop.

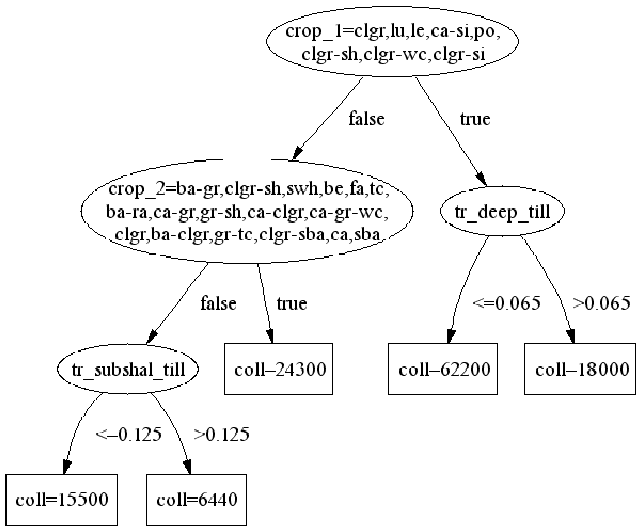


Figure 3: Collembolan regression tree.

Table 5. Correlation coefficient and errors of biodiversity models.

Name	size	corr	MAE	RMSE	RAE	RRSE
ibk 1	**0**	**0.623**	**0.361**	**0.505**	**77.485**	**79.936**
ibk 5	0	0.588	0.369	0.479	79.204	81.442
ibk 10	0	0.518	0.392	0.470	84.032	85.770
m5 linear equation	1	0.533	0.394	0.500	84.656	84.960
m5 model tree	**9**	**0.575**	**0.373**	**0.483**	**80.154**	**82.059**
m5 model tree simple	9	0.570	0.376	0.486	80.722	82.675
m5 model tree pruning 4	3	0.511	0.398	0.509	85.342	86.501
m5 model tree pruning 4 simple	3	0.506	0.400	0.512	85.782	86.945
m5 regression tree	21	0.495	0.401	0.515	86.160	87.570
m5 regression tree pruning 5	12	0.423	0.420	0.534	90.096	90.767
m5 regression tree pruning 7	*6*	*0.383*	*0.430*	*0.544*	*92.279*	*92.528*

The models helped us to get some additional knowledge from the domain experts, for example:

- The effects of tillage are long lasting - at least 5 months with subshallow tillage and 7 months with shallow tillage. A lot of species are sensitive to tillage and this will lower the biodiversity. More opportunistic and small species will dominate in intensive tilled soils. More sensitive species takes a longer time to recover.
- Fertilization increases the biomass of the plants (gives higher soil microbial activity = food) + the fertilizers stimulates soil microbial activity and creates new habitats (can live inside or around the organic matter)

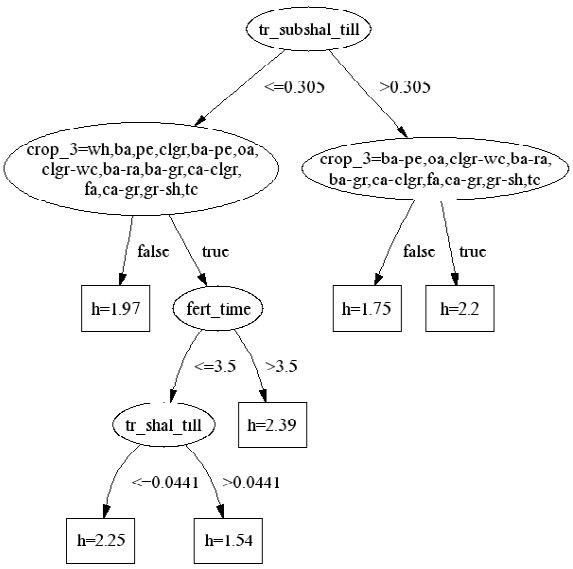


Figure 4: Biodiversity regression tree.

5 Conclusions

We tried to model community of soil microarthropods with machine learning methods from the data describing chemical, biological and mechanical actions on the fields. We then used so produced models to identify the most important parameters for soil mites, springtails and biodiversity of soil microarthropods. By preferring small and simple models to bigger and complex models. We discovered that the most important factor for community of soil microarthropods are previous crops grown in the observed field, and the different forms of tillage. Furthermore we used the models as a source of questions for the domain experts. We gained knowledge that will help us in further modelling and building decision support system for the management of farms. While the domain experts will mainly be relying on their knowledge in participating in decision support model building, they are somewhat guided by the models. With newly gained knowledge we also identified parts of the decision support model that need special care when building. We have shown that the machine learning models can be used in multiple ways from predicting new values, to gaining new knowledge about the relation between the attributes and the dependent variable, to extracting knowledge from the domain experts.

Acknowledgements

This work was supported by ECOGEN funded by the Fifth European Community Framework Programme: Quality of Life and management of living resources contract no QLK5-CT-2002-01666 and DARCOF, Nat. quality in organic farming.

References

[1] Aha, D. and Kibler, D. (1991): Instance-based learning algorithms, *Machine Learning*, **6**, 37-66.

[2] Birch, A.N.E., Krogh, P.H., Cortet, J., Tabone, E., Griffiths, B.S., Džeroski, S., Wesseler, J., Gomot de Vaufleury, A., Badot, P-M., Andersen, M.N., and Messéan, A. (2003): ECOGEN: Soil ecological and economic evaluation of genetically modified crops. Poster at *Biodiversity Implications of Genetically Modified Plant*s, September 7-12, 2003 Monte Verità, Ascona, Switzerland Centro Stefano Franscini, Swiss Federal Inst. of Technology Zürich.

[3] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984): *Classification and Regression Trees*. Belmont: Wadsworth.

[4] Demšar, D., Džeroski, S., Krogh, P.H., and Larsen, T. (2003): Identifying the most important agricultural factors for the soil community of microathropods, *Proceedings of the International Electrotechnical and Computer Science Conference*. Ljubljana, Slovenia

[5] ECOGEN: Soil ecological and economic evaluation of genetically modified crops. http://www.ecogen.dk

[6] Krogh, P.H. (1994): Microarthropods as bioindicators. A study of disturbed populations. PhD thesis Ministry of the Environment and Energy. National Environmental Research Institute, Silkeborg.

[7] Quinlan, J.R. (1993): Combining instance-based and model-based learning. In *Proceedings of the X. International Conference on Machine Learning.* 236–243.

[8] Morgan Kaufmann.Recio, B., Rubio, F., and Criado, J.A. (2002): A decision support system for farm planning using AgriSupport II. *Decision Support* Systems, **36**, 189–203.

[9] Steen, E. (1983): Soil animals in relation to agricultural practices and soil productivity. *Swedish J. agric. Res.*, **13**, 157-165.

[10] Wang, Y. and Witten, I.H. (1997): Induction of model trees for predicting continuous classes. Proceedings of the *Poster Papers of the ECML 97. University of Economics*. Prague: Faculty of Informatics and Statistics.

[11] Witten, I.H. and Frank, E. (1999): *Data Mining: Practical Machine Learning Tools with Java Implementations*. San Francisco: Morgan Kaufmann.