# Using multiobjective classification to model communities of soil microarthropods

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

[1] Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia,

[2] Department of Terrestrial Ecology, National Environmental Research Institute, Roskilde, Denmark

# outline

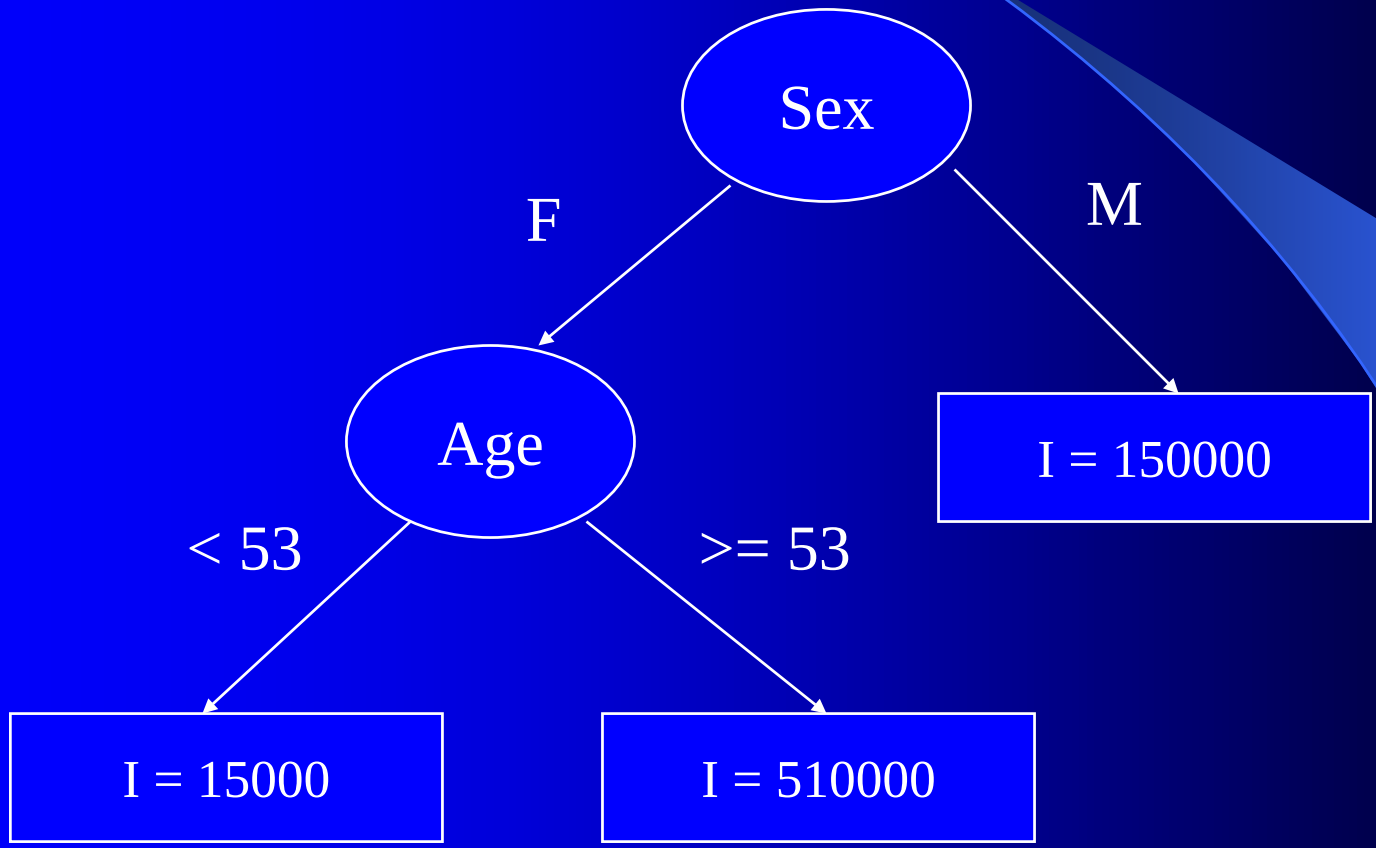
- multiobjective classification
- motivation
- datasets
- results
- conclusions

# multiobjective classification

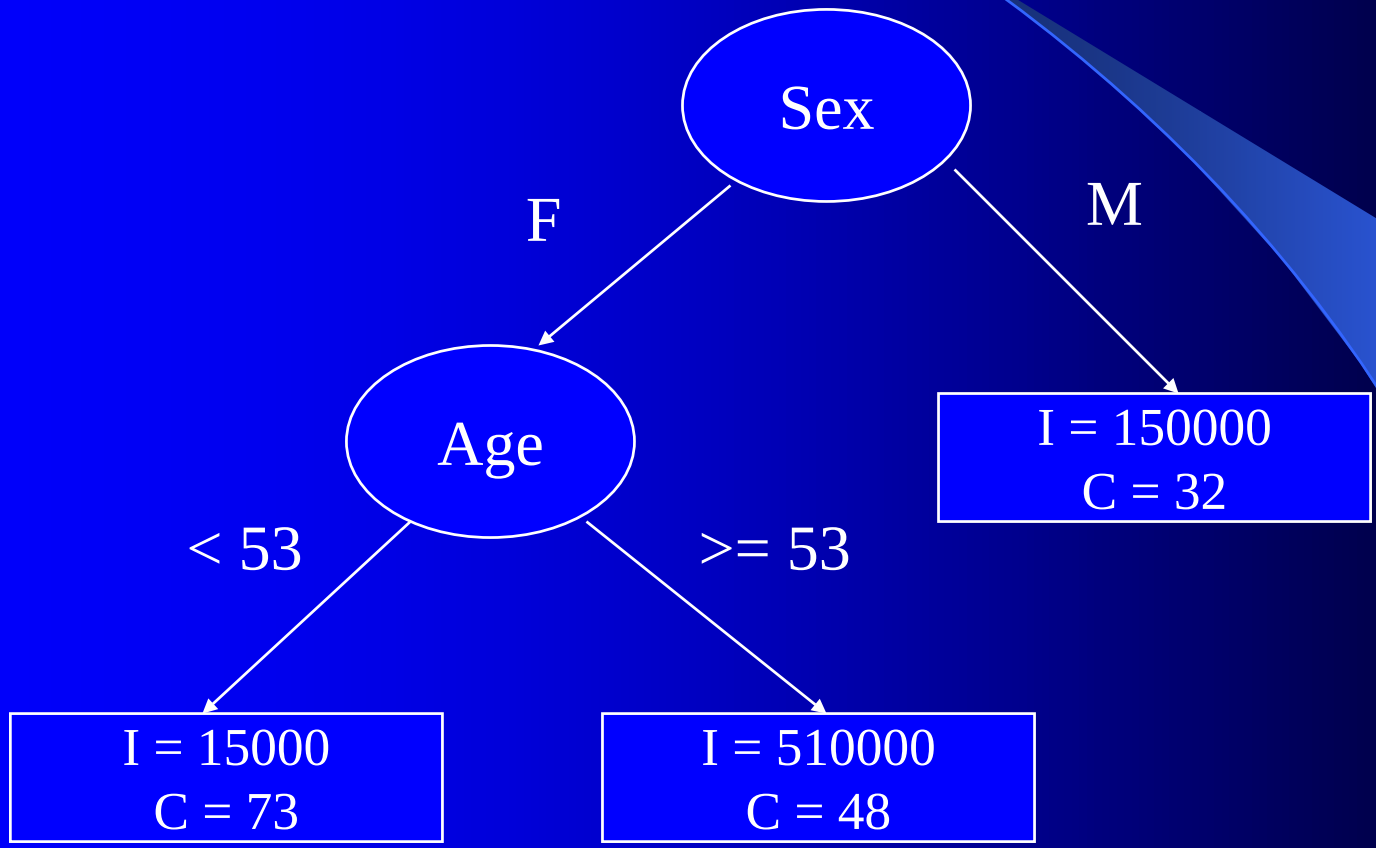
- use one model to classify several target variables
- can be used to indentify differences between target variables

- we used CLUS

# regression tree



# multiobjective regression tree



# ECOGEN: Soil ecological and economic evaluation of genetically modified crops oobjectives

- ecological and economical assessment and comparison of integrated farming systems using genetically modified higher plants with a conventional farming system.
- provide an ecological risk assessment of a GM cropping system and a conventional cropping system for the soil ecosystem.
- economic assessment of GM plant crops and conventional crops with respect to a quantification of the expected trade-offs between the two.
- incorporate ecological knowledge from single species tests, multispecies tests, field investigations and economic information from farming practices into a rule based model to be used for predictions of economic decision-making processes and ecosystem behaviour.

# ECOGEN WP6: data analysis and decision support

objective:

build an integrated rule-based model
for assessing the sustainability of farming
taking into account ecology and economy

the model would be used

for decision support
to choose a suitable farming system

# motivation

- extract new knowledge from the data
- extract old knowledge from the experts (use the models to construct questions for the experts)
  - identify the most important factors for the community of soil microarthropods
  - identify and try to explain the different reactions of different parts of the community

# the datasets

- foulum dataset
  - 4 experimental farming systems
  - conventional and organic
  - 1989-1993
  - 530 samples
- nk2 dataset
  - several organic farms
  - 2002 and 2003
  - 800 samples (430 from 2002, 370 from 2003)
- total of 1330 samples, 1138 without missing data

# the datasets

- agricultural measures
(packing, tillage, fertilizer and pesticide use)
- history of crops and grazing for the last 3 years
- environmental variables
- dependant (class) variables
  - biodiversity index (H)
  - total number of mites (acari)
  - total number of springtails (coll)
  - 45 different species / groups

$$H = \sum_{i=1}^{n} p_i * \log_2(p_i)$$

# the datasets

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

ECOGEN

# the datasets

| 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* | Apygm[2] | *Anurida pygmaea* |

[1] acari [2]collembola

ECOGEN

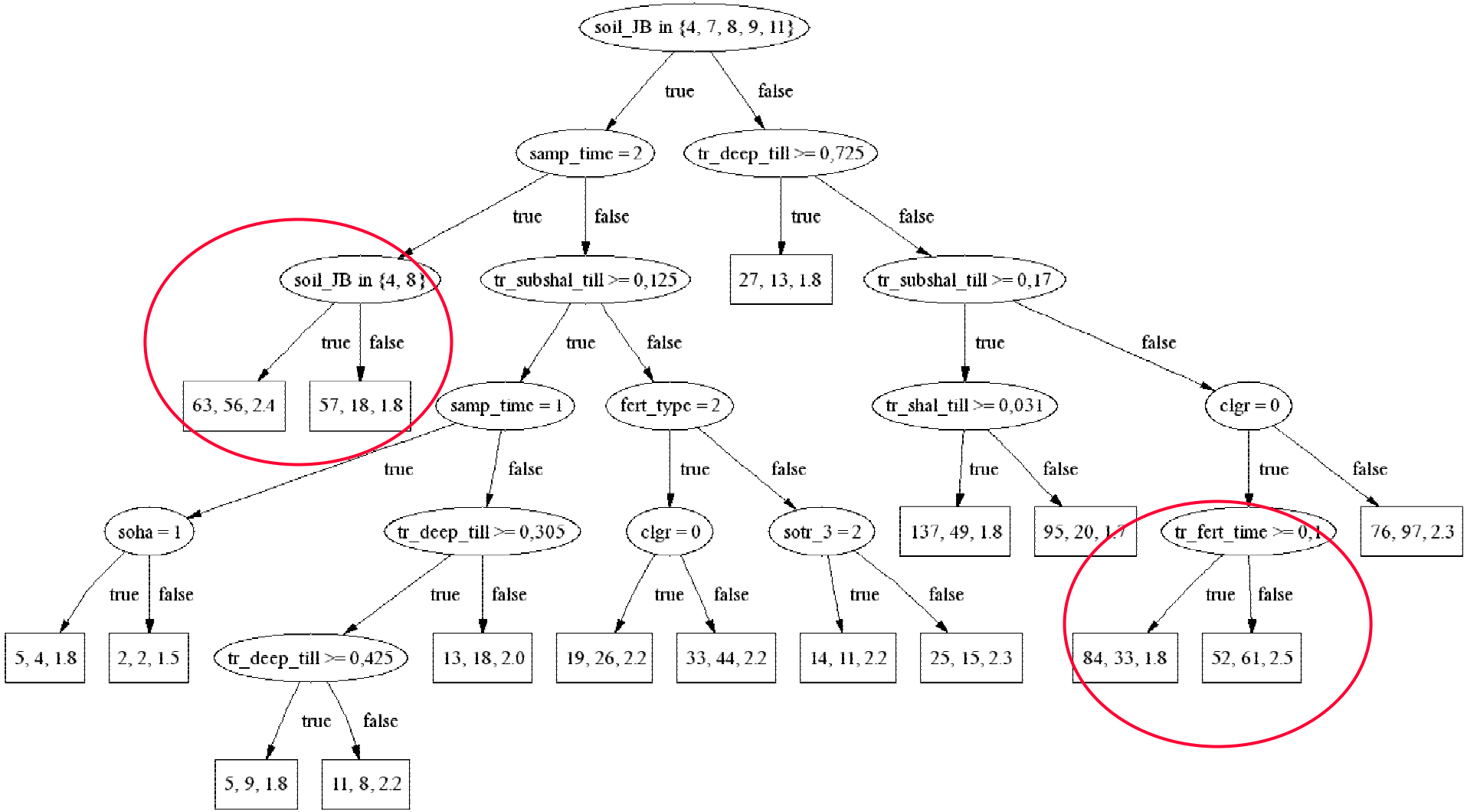
# multiobjective models

- used tool CLUS to predict several measures at once
- compared the accuracies with CLUS prediction of single measure
- acari / coll / biodiversity
- all species / groups (not yet evaluated)

# Multiobjective model

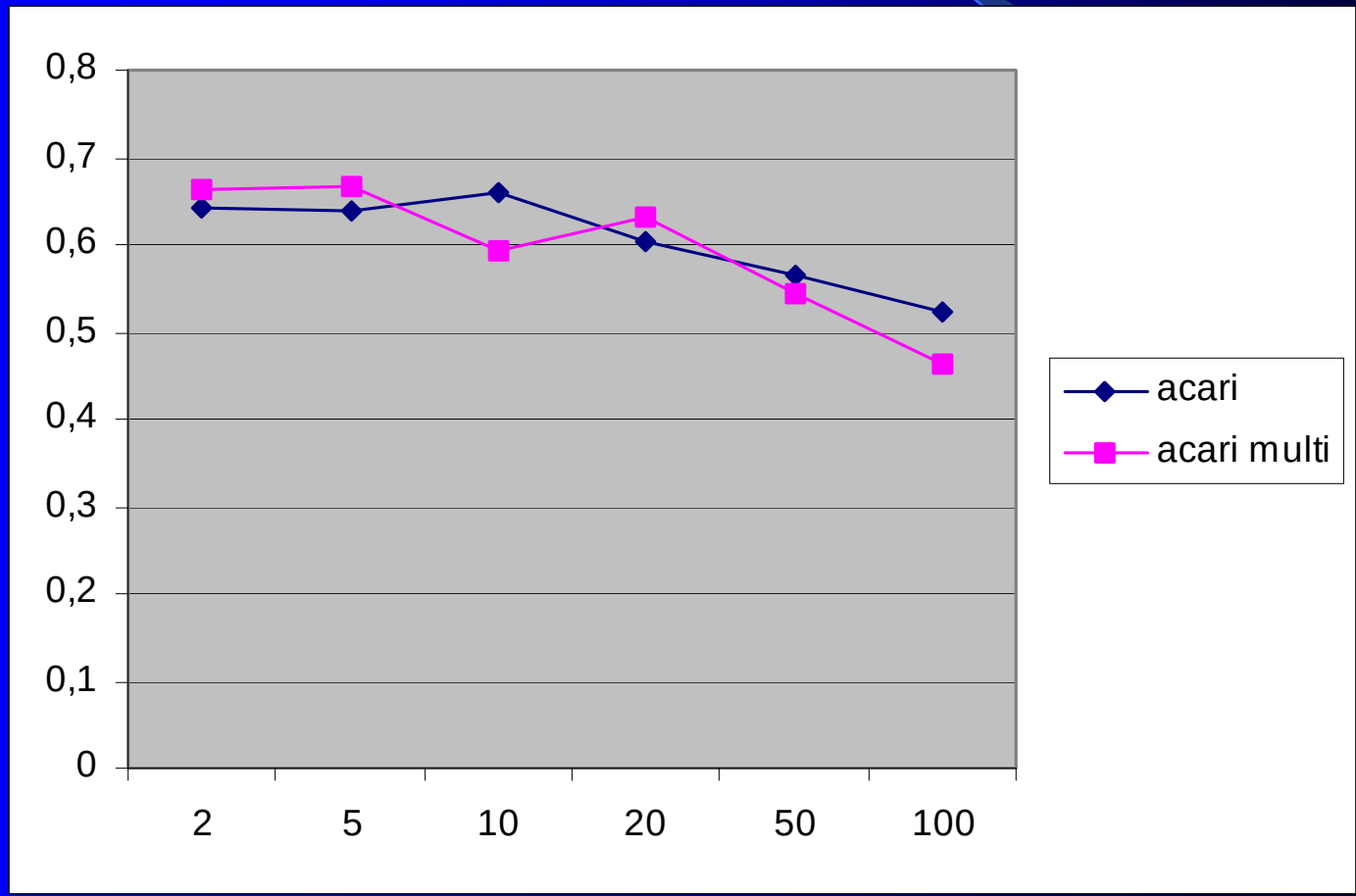
```
soil_JB in {11,4,7,8,9}
+--yes: samp_time = 2
|       +--yes: soil_JB in {4,8}
|       |       +--yes: [63227,575756 high mid ,56446,909174 high mid ,2,478269 high] : 81
|       |       +--no:  [57453,801291 high mid ,18890,521352 low mid ,1,890191 mid] : 76
|       +--no:  tr_subshal_till >= 0,125
|               +--yes: samp_time = 1
|               |       +--yes: soha = 1
|               |       |       +--yes: [5873,925643 low ,4511,21425 low ,1,816766 mid] : 56
|               |       |       +--no:  [2469,428036 low ,2519,517855 low ,1,573708 low] : 55
|               |       +--no:  tr_deep_till >= 0,305
|               |               +--yes: tr_deep_till >= 0,425
|               |               |       +--yes: [5620,645629 low ,9139,584454 low ,1,851626 mid] : 97
|               |               |       +--no:  [11061,344 low ,8914,57 low ,2,200479 high] : 53
|               |               +--no:  [13437,061506 low ,18326,515227 low mid ,2,047582 high mid] : 80
|               +--no:  fert_type = 2
|                       +--yes: clgr = 0
|                       |       +--yes: [19412,214744 low ,26461,953533 low mid ,2,215455 high] : 92
|                       |       +--no:  [33046,814499 low mid ,44941,613306 mid ,2,207047 high] : 58
|                       +--no:  sotr_3 = 2
|                               +--yes: [14888,327633 low ,11274,965902 low ,2,234485 high] : 54
|                               +--no:  [25493,512971 low mid ,15037,699693 low mid ,2,351412 high] : 74
+--no:  tr_deep_till >= 0,725
        +--yes: [27857,687196 low mid ,13808,983475 low ,1,868849 mid] : 66
        +--no:  tr_subshal_till >= 0,17
                +--yes: tr_shal_till >= 0,031
                |       +--yes: [137373,820398 HIGH ,49045,906321 high mid ,1,840351 mid] : 50
                |       +--no:  [95526,267516 high ,20688,654538 low mid ,1,783139 low mid] : 52
                +--no:  clgr = 0
                        +--yes: tr_fert_time >= 0,1
                        |       +--yes: [84701,949268 high ,33120,80018 mid ,1,820153 mid] : 56
                        |       +--no:  [52483,917114 mid ,61587,456419 high ,2,536311 HIGH] : 60
                        +--no:  [76746,840473 high mid ,97745,781348 HIGH ,2,385857 high] : 78
```

# Multiobjective model

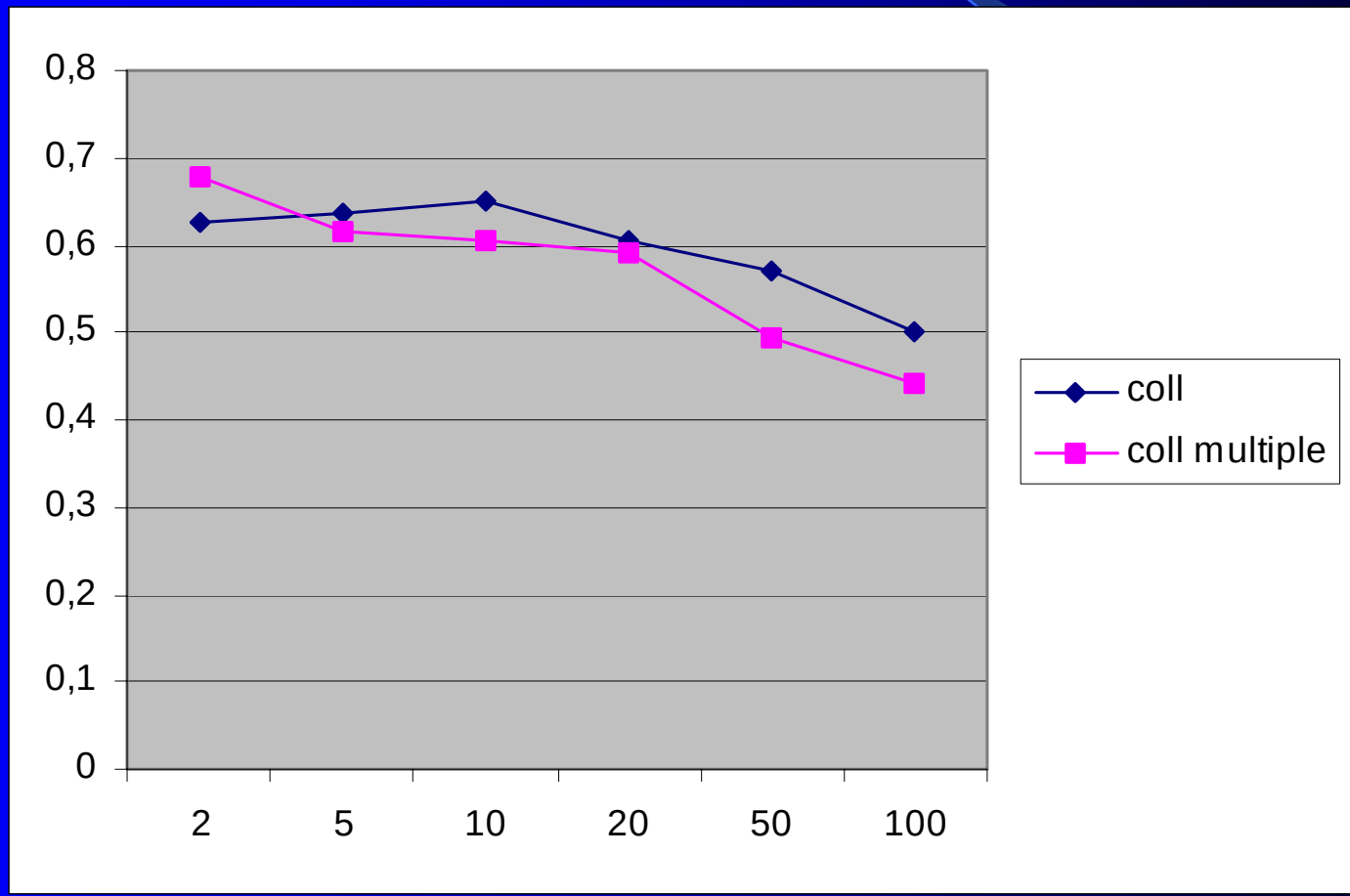



acari in 1000, coll in 1000, biodiversity

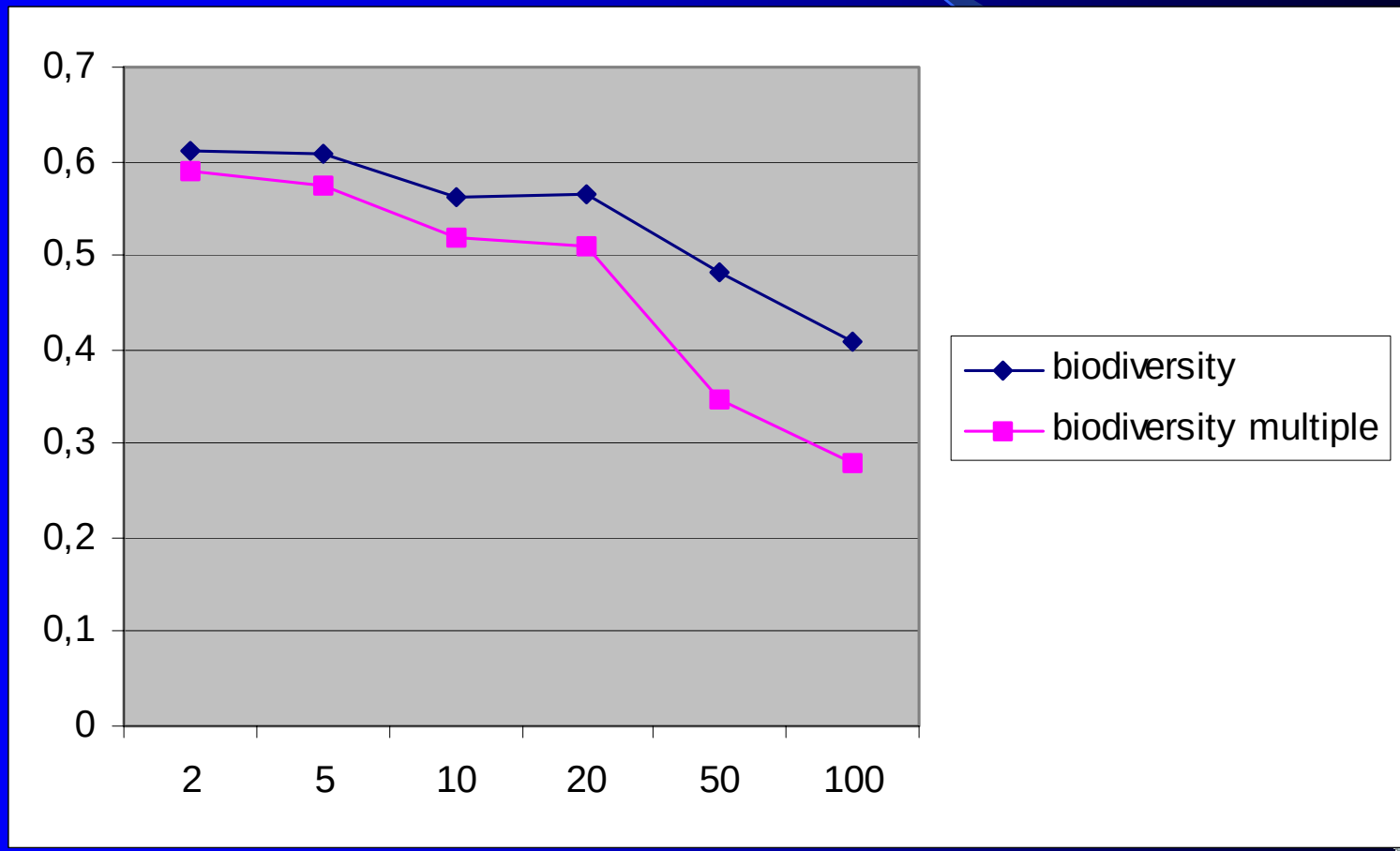
# multiobjective  model acari correlation



# multiobjective  model coll correlation



# multiobjective  model biodiversity correlation



# multiobjective model findings

- acari  and coll abundances depend most on soil type (worst are 7 9 11), while biodiversity is independent of the soil type
- subshallow tillage has longest big negative effect on biodiversity, while acari and coll abundancies are not as highly effected
- recent deep tillage has big negative effect on accari and medium negative effect on coll and biodiversity
- (not so recent) subshallow tillage (with no recent deep tillage) has positive effect on acari, small negative effect on coll and medium to big negative effect on biodiversity
- recent fertilization has positive effect on acari, while it has negative effect on coll and biodiversity

# conclusions

- only slight or even no performance loss when using multiobjective models
- can be used to identify the factors important for the whole community, not just one part
- simple identification of different reactions to (agricultural) actions

- 1 multiobjective model is simpler than several single objective  models

# to do

- experts evaluation of acari/coll/biodiv multiobjective model
- evaluation of multiobjective species model
- using single species models to predict compound measures

# thank you

ECOGEN