# The potential of soil habitat features for the modelling of numerical abundance of Collembola: a neural network approach

Christian Kampichler

## Synopsis

Since the early days of soil zoology, researchers have repeatedly discussed the importance of edaphic environmental characteristics for the distribution and abundance of soil animals. An extensive soil-biological survey at the experimental farm of the FAM Munich Research Association for Agricultural Ecosystems provided the opportunity of testing the predictive potential of soil habitat features for the density of microarthropods. In April 1991 cores were taken at 400 points distributed over a 50x50 m sampling grid to measure the abundance of soil microarthropods, microbial biomass, microbial respiration, soil moisture, pH, carbon content, nitrogen content and soil texture. The uniquely large sample size made the application of a novel approach in ecological data analysis possible: this paper reports on an attempt to model the abundance of the dominant Collembolan species *Folsomia quadrioculata* by the use of artificial neural networks (NNs). NNs do not require an *a priori* assumption of the functional form of the relationship between independent and dependent variables; governed by certain learning algorithms they are able to let the data define the form by themselves.

NN models performed better than models based on multiple linear regression and were capable of identifying sampling points with the potential of housing large numbers of *F. quadrioculata*. This reflects the non-linear nature of the relationship between abundance and habitat characteristics. It was, however, not possible to obtain a finer resolution of the abundance than merely distinguishing between points with low density and points with the potential for high density. Thus, this study also sheds a light on the limits of predicting local abundance by a pure habitat quality model. The environmental conditions may well set the limits of where a species is able to exist or not, but its actual density is determined by various processes acting at different spatial and temporal scales (e.g. aggregation behaviour, immigration from neighbouring habitats).

*Collembola, Folsomia quadrioculata, soil microarthropods, abundance, habitat features, artificial neural networks, modelling, agroecosystem, microbial biomass, carbon content*

## 1 Introduction

Soil microarthropods – to the largest part Acarina and Collembola – are exceedingly numerous in natural as well as in cultivated ecosystems (PETERSEN & LUXTON 1982; LARINK 1997); they are said to have the potential of influencing major soil-based ecosystem processes (VERHOEF & BRUSSAARD 1990; LUSSENHOP 1992). For decades, soil ecologists have regarded habitat conditions to be of major importance for determining distribution and abundance of soil microarthropods. Early workers meticulously recorded information on soil type, litter characteristics and other components at the sampling and collecting localities (e.g. FRANZ & BEIER 1948; FRANZ & SERTL-BUTSCHEK 1954), and modern textbooks on soil zoological methods still emphasize the importance of a thorough habitat description (DUNGER & FIEDLER 1989). Since it is difficult to gain experimental access to the interactions of microarthropods with their habitat, the analyses of whether or not and to what extent habitat features influence the abundance and distribution of soil microarthropods, are exclusively based on a correlative approach. Among the environmental variables that have been shown to exert a noticeable influence on abundance and distribution are temperature, soil type, soil moisture, soil acidity and microbial characteristics (cf. reviews by EKSCHMITT 1993 and HOPKIN 1997 and references therein). Typically these variables only explain part of the total variance in the microarthropod data (e.g. 35% in KLIRONOMOS & KENDRICK 1995). The remaining variance could either be put down to the fact that some variables have not yet been measured or to the fact that spatial distribution of soil microarthropods is highly aggregated. As far as we know, aggregations of Collembola typically appear to have a size of 5 to 35 cm (EKSCHMITT 1993). Thus, when analysing habitat feature:density relationships the soil zoologist is always confronted with the difficulty of having to filter these relationships from the often overriding data on aggregation. The patchy distribution patterns of soil microarthropods and the difficulties imposed by assessing even their basic population parameters, explains the large amount of published material available on statistics of microarthropod sampling which have been accumulated from the beginning of soil zoology as a science until now (e.g. HEALY 1962; EKSCHMITT 1993). However, in many cases the choice of the sample size is a matter of feasibility (man-power, financial supply, etc.) and not one of statistical requirements. The fact that soil zoologists are forced to reduce the number of sampling units means that many of the assumptions about relationships between habitat features and microarthropod abundance are based on weak statistical inference. Sometimes even the number of study sites is kept to a minimum, giving rise to pseudo-replicative study designs (i.e. each site is the only replication of a treatment and therefore the true sample size n = 1, regardless of how many »pseudoreplicated« cores are taken within a site (HURLBERT 1984; HEFFNER & al. 1996).

The following work is based on a large data-set from the experimental farm of the FAM Munich Research Association for Agricultural Ecosystems. It consists of a sample of almost 400 soil cores taken across a landscape section of more than 100 ha with a large pedological variability. It thus yields a unique opportunity to answer the question as to whether the number of individuals at a point in space can be inferred from the habitat features prevailing at that point. The unusually large size of the data-set makes the application of a novel means of ecological modelling possible, namely artificial neural networks (NNs). NNs are universal approximators and, thus, act as a type of nonparametric statistic enabling us to model complex functional forms (WARNER & MISRA 1996). When a regression analysis is applied some functional form has to be imposed on the data. This is, however, not necessary in the case of a NN. Its basic idea is to let the data define the functional form by themselves. A regression model may be used when either the functional relationships between independent and dependent variables are known or when there are at least plausible assumptions about these relationships. A NN, in contrast, is useful when the functional relationships are unknown. Recently ecologists recognized NNs as a potential tool for data analysis and modelling and use them in various contexts (LEK & al. 1996; PARUELO & TOMASEL 1997; RECKNAGEL & al. 1997). These studies have shown that NNs can fit the complexity and nonlinearity of ecological phenomena to a high degree.

Thus, the aim of this paper is (1) to create a NN model that explains local density of microarthropod in terms of habitat characteristics and (2) to test the model's ability to predict microarthropod density on sites where the habitat characteristics are known. This report concentrates on an attempt to model numerical abundance of the dominant Collembolan species *Folsomia quadrioculata* (TULLBERG, 1871).

## 2 Material and Methods

### 2.1 Study site and data origin

The FAM runs a 153 ha experimental farm in Scheyern approx. 40 km N of Munich. It is situated at an elevation of 450 to 490 m above sea level; mean annual temperature and mean annual precipitation amount to 7.5° C and 833 mm, respectively. The farm was conventionally managed until 1992 and was then divided into a part with integrated farming (46 ha), a part with organic farming (68 ha) and a part that is reserved for plot trials (39 ha). In April 1991 the corner points of a 50x50 m grid were sampled with one soil core (7.8 cm diameter, 5 cm depth) each and yielded a total of 396 cores. The majority of points were situated in arable fields (n = 302), the remainder in pastures, meadows and arable fields on former hop fields. All arable land (except grassland) was uniformly grown with winter barley to create similar conditions before the change of management systems in 1993. Microarthropods were counted and Collembola identified by species (FROMM 1998). For the measurement of the following environmental factors cores were taken from the same grid points at a distance of approximately 25 cm: microbial biomass, microbial respiration, soil moisture, soil acidity, carbon content (Ct) and nitrogen content (Nt) (WINTER 1998). Soil texture at the grid points was determined by SINOWSKI (1994). Table 1 sums up methods and observed values for all variables. All data were drawn from the FAM data base at the GSF in Neuherberg, Germany (URL: http://www.gsf.de/FAM/famdaten.html).

*F. quadrioculata* exhibits moderate to low correlation coefficients with single environmental factors (Table 1). Fig 1, for example, shows the relationship between density of *F. quadrioculata* and carbon content in soil. Clearly visible are (1) the skewed frequency distribution of abundance of *F. quadrioculata* as well as of carbon content, and (2) the tendency of cores with a high abundance to appear more frequently at a higher soil carbon content. Due to small scale aggregations of Collembola, even on sites with high levels of carbon content some cores contain few or no individuals of *F. quadrioculata* at all. The functional form of the relationship is not clearly visible, which speaks in favour of a NN modelling approach.

### 2.2 Neural networks

A NN consists of a set of computational units, termed *cells*. These cells are joined by a set of one-way *connections*. At certain times a cell examines its inputs, computes an output and passes it on along the connections leading to other units. The most popular

Table 1

**Analytical methods, data range and product moment correlation coefficient between environmental variables and abundance of *F. quadrioculata* at Scheyern experimental farm ($n = 195$).**

| Variable | Method | Range | | r |
|---|---|---|---|---|
| Microbial respiration | automatic IRGA[a] | 0.76 – 10.03 | µg $CO_2$ $g^{-1}$ soil d.m. $h^{-1}$ | 0.36 |
| Microbial biomass | automatic IRGA[a] | 4.97 – 80.98 | µg $CO_2$ $g^{-1}$ soil d.m. $h^{-1}$ | 0.43 |
| Soil acidity | $CaCl_2$ | 4.72 – 7.04 | $-\log [H^+]$ | –0.28 |
| Soil moisture | dried at 105° C | 5.1 – 48.13 | % weight of soil d.m. | 0.48 |
| $C_t$ | elemental analyser[b] | 0.96 – 5.86 | % weight of soil d.m. | 0.43 |
| $N_t$ | elemental analyser[b] | 0.079 – 0.833 | % weight of soil d.m. | 0.39 |
| Median particle size | sieving[c], sedimentation[d], laser diffraction[e] | 6.3 – 630 | µm[f] | –0.04 |
| *F. quadrioculata* | soil core (7.82 x 5 $cm^3$) | 0 – 95 | individuals $core^{-1}$ | – |

[a] after Heinemeyer & al. (1989)

[b] oxidization at 1020° C in a Carlo Erba NA 1500 and subsequent determination of gas concentration of $CO_2$ and $N_2$ by gas chromatography and heat conductivimetry

[c] after Hartge & Horn (1989)

[d] after Köhn (1928)

[e] after Heuer & Leschonski (1985)

[f] class limits at 12.5, 20, 28, 40, 50, 63, 100 and 200 µm

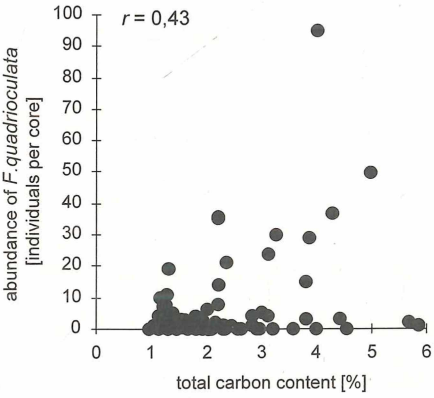

Fig. 1

**Relationship between carbon content and abundance of *F. quadrioculata* at the Scheyern experimental farm ($n = 195$ points of a 50 x 50 m sample grid).**

type of NN is the *backpropagation network*. In this type of network the cells are typically organized in three layers – an input layer, an intermediate (or hidden) layer and an output layer – with connections between all cells of one layer to all cells of the next higher layer. There are no connections within layers and no connections to pass back information to a lower layer (Fig. 2).

Each connection to a cell $u_i$ has a numerical weight $w_{i,j}$ that specifies the influence of cell $u_j$ on $u_i$. The total input into a cell, thus, is a weighted sum S of the outputs from all cells $u_j$ connected to it and can be expressed as

$$S_i = \underset{i}{\phantom{\sum}} w_{ij}u_j \tag{1}$$

The output of a cell $u_i$ – its activation – is then calculated by a non-linear and differentiable activation function $f(S_i)$. Most commonly used is the logistic or sigmoid function

$$u_i = \frac{1}{1+e^{-S_i}} \tag{2}$$

The process of passing on activations from the input to the output layer is called *forward propagation.*

Similar to coefficients in regression analysis, the connection weights are adjusted to solve the problem presented to the NN. This is achieved in a *training* (or *learning)* stage: the NN is provided with training patterns with known outputs for a given input. The NN first calculates its output based on a randomly chosen initial set of connection weights. It then com-

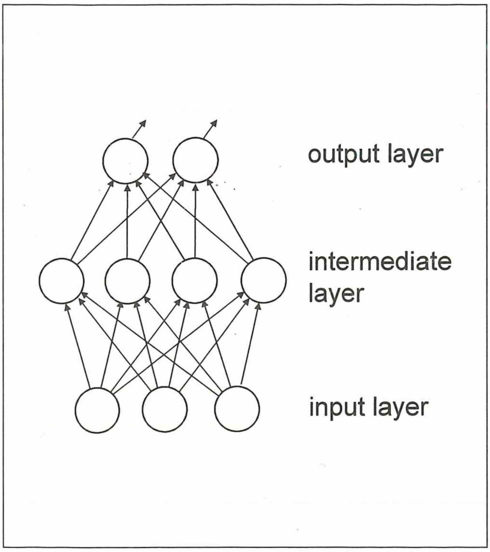

Fig. 2

**Topology of a backpropagation network.**

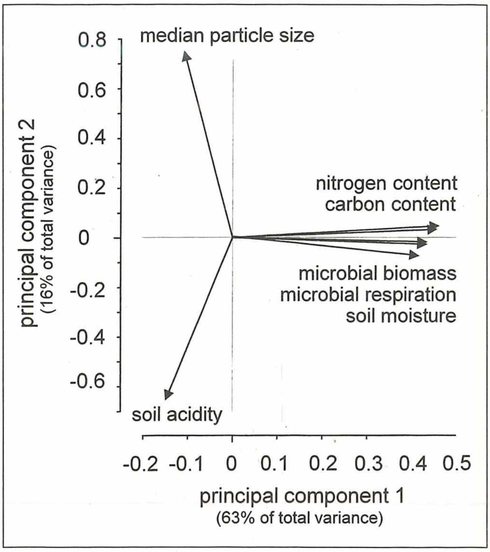

Fig. 3

**Plot of component weights of the environmental variables on the first and second principal component ($n = 195$).**

pares the computed output with the desired output. Governed by a gradient descent technique the error is propagated top-down through the NN (= *backpropagation).* By slightly altering the connection weights the NN in a small step approaches the desired output. In an iterative process one training pattern after the other is presented to the NN and the weight matrix successively adjusted.

GALLANT (1993) and WARNER & MISRA (1996) provide introductions to backpropagation learning; the former also presents a formal derivation of the calculations and a simple numerical example.

## 2.3 Modelling

Grid points with missing data for one or more of the environmental variables were omitted from the data set with leaving 195 points in the analysis (155 of which to be found in arable fields). (The set of values of environmental factors and of abundance of *F. quadrioculata* at one point will be called a *pattern* due to the usual nomenclature in NN modelling.) 150 patterns were randomly chosen as training patterns to train a NN with 7 cells in the input layer – one for each environmental variable – , 7 cells in the intermediate layer and 1 cell in the output layer (abundance of *F. quadrioculata*). The data were transformed to the interval [0 1] by

$$x_i' = (x_i - x_{min}) / (x_{max} - x_{min}) \qquad (3)$$

with $x_{min}$ being the smallest and $x_{max}$ being the largest value of a variable in the training patterns. The frequency distribution of the abundance of *F. quadrioculata* is heavily skewed to the right with very few cores containing very high numbers of individuals. Thus, before applying (3) an upper threshold value at the 95%-quantile was defined and all values larger than 16 individuals per core were denoted as 16.

The number of learning cycles amounted to 250; a larger number showed effects of *overlearning.* In that case the NN gradually loses its ability to model other patterns than the training patterns. Subsequently the 45 test patterns were transformed according to eq. (3) and the environmental variables were used as input to the trained NN. Its output is thus the prediction of the NN model for the abundance of *F. quadrioculata* at a given constellation of environmental factors.

Assuming that for the estimation of a function at least 5 data points are necessary a NN with 7 input variables would thus require $7^5 = 16\,807$ patterns or more for training (R. Wieland, pers. comm.). It therefore needs a parsimonious approach to design the NN and to keep the number of input variables as small as possible. A principal component analysis demonstrated a high correlation between several of the environmental variables (Fig. 3). Therefore, in addition to the NN design described above, I also trained a NN with only 4 input cells (carbon content, acidity, median size particle and a »management« variable distinguishing between arable field, pasture, meadow and former hop field), 4 intermediate cells and 1 output cell. The same training and test patterns as above were used. The two networks will be named NN7 and NN4, respectively, throughout the paper.

All designing, training and testing of NNs was performed with the SNNS Stuttgart Neural Network Simulator v4.0 and v4.1 (ZELL & al. 1995).

The predictive power of both NN models was compared to the power of a multiple regression analysis (MR) assuming the simplest possible functional form, a linear relationship, between independent and dependent variables. The same 150 training patterns as described above were used to calculate the regression equation; the same 45 test patterns were used to compare the computed output of the regression model (= prediction of abundance of *F. quadrioculata*) with the observed values. In analogy to NN modelling one MR was performed with 7 independent variables (MR7), and an additional MR with only 4 independent variables (MR4).

## 3 Results

$NN_7$ is as successful in predicting low abundance (0–4 individuals per core) as it was in predicting high abundance (> 10 individuals per core) (Fig. 4). However, it misclassifies some points distinctly and predicts medium abundance where low abundance was observed and predicts low abundance where medium abundance was observed. MR7 performs worse in distinguishing between points with low and high abundance. In a scatterplot of observation vs. prediction the data points of different observed abundance are hardly separable along the axis designating predicted abundance (Fig. 4). This is reflected by the product-moment correlation coefficients between observation and prediction with $r = 0.76$ in $NN_7$ and $r = 0.64$ in $MR_7$. The mean error, i.e. the average of the difference between predicted and observed values, is 0.0 in $NN_7$ and -0.3 in $MR_7$; its standard deviation is larger in $MR_7$ due to some points where abundance is drastically underestimated by 8 to 10 individuals (Fig. 5).

The predictions of $NN_4$ are not as good as those of $NN_7$: points of high observed abundance again have the highest predicted values but are less well set off against points with low observed abundance (Fig. 4). $MR_4$ definitely fails in identifying points with high abundance predicting maximum values for points with low observed abundance (Fig. 4). Thus, the correlation coefficient between observation and prediction is distinctly higher in $NN_4$ ($r = 0.66$) than in $MR_4$ ($r = 0.36$). The mean error of $NN_4$ is close to 0 and equals -0.4 in $MR_4$. Due to some points that are under- and overestimated, respectively, by 10 to 12 individuals the standard deviation of the absolute error again is higher in $MR_4$. In both $NN_4$ and $MR_4$ the standard deviation is greater than in $NN_7$ and $MR_7$, respectively, reflecting the lower efficiency of the models with fewer independent variables (Fig. 5).

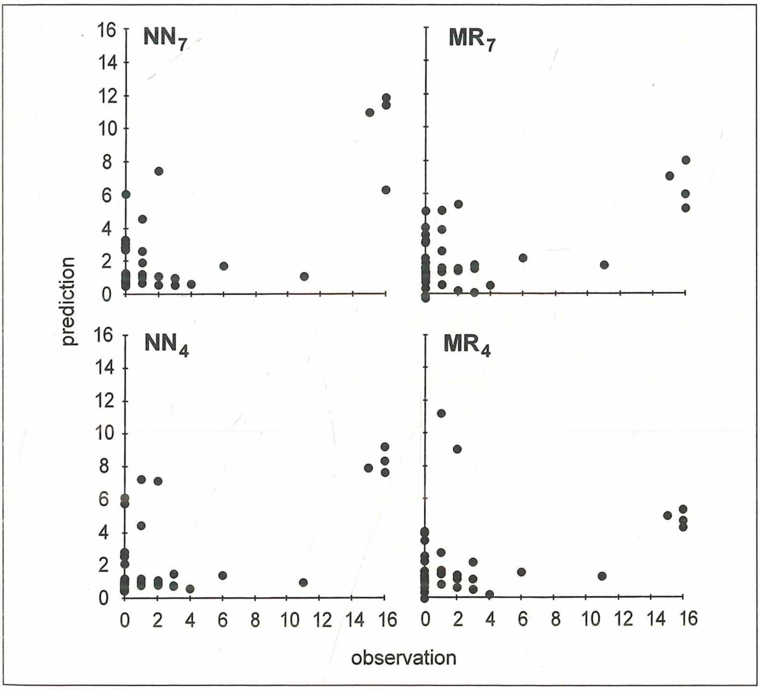

Fig. 4
**Scatterplots of observed vs. predicted abundance of *F. quadrioculata* [individuals per core] as calculated by the models NN7, MR7, NN4 and MR4 (n = 45). See Material and Methods for description of models.**

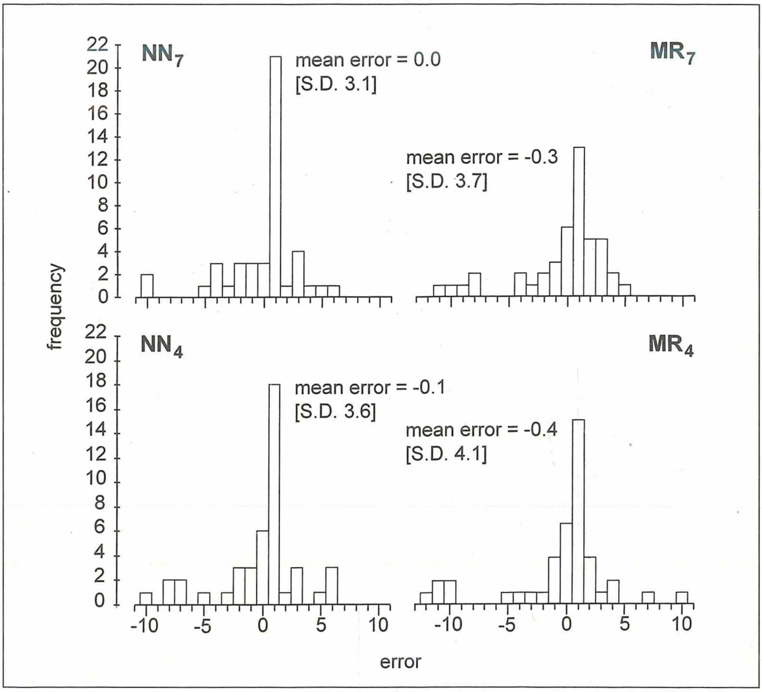

Fig. 5

**Frequency distribution of the error of the predicted abundance of *F. quadrioculata* [individuals per core] by the models $NN_7$, $MR_7$, $NN_4$ and $MR_4$ ($n = 45$). See Material and Methods for description of models.**

**S.D., standard deviation**

In all models, predictions of maximum abundance are too low compared to the observed data (Fig. 4). This is particularly true for the MR models with maximum estimates of 8.0 in $MR_7$ and 5.3 in $MR_4$ – 11.4 and 9.0 are misclassified points – compared to 11.8 in $NN_7$ and 9.2 in $NN_4$.

## 4 Discussion

The NN models perform better than the models based on MR; this reflects non-linearities in the relationship between habitat features and abundance of *F. quadrioculata*. Similar results were found by KAMPICHLER (1997) who compared the efficiency of NNs and MRs in modelling the total number of individuals of endogeic and hemiedaphic Collembola at the Scheyern experimental farm. In his study correlation coefficients between prediction and observation were significantly higher in NN models than in MR models. Their lack of transparency, however, partly outweighs the better performance of NN models. In a MR model a coefficient is applied to each independent variable, explicitly reflecting its weight on the dependent variable. The analysis of the response of a trained NN to different values of the input variables, however, is a laborious and time-consuming task (cf. LEK 1996) and has not yet been carried out for the *F. quadrioculata* model. Such an analysis normally is postponed until the final model with optimum predictive power is developed. Until that stage a NN is used as a black box.

As can be seen in bivariate plots (cf. Fig. 1) we cannot expect a simple deterministic relationship between habitat features and abundance. The environmental conditions set the limits of where a species is able to exist, and $NN_7$ and $NN_4$ are well able to characterize the potential of a site to house individuals of that species. The actual density at a certain point in space and time, however, is dependent on a variety of processes at different spatial and temporal scales, e.g. on the aggregative behaviour of the species (EKSCHMITT 1993), on the immigration from adjacent habitats or on stochastic disturbances (RICKLEFF 1987). EKSCHMITT (1993) points out that differences in habitat quality are possibly only reflected in distribution patterns when animal density is high and the carrying capacity of the habitat is approached. Thus, the data points lined up along the y-axes of $NN_7$ and $NN_4$ in Fig. 4 characterize a number of micro-habitats where these models expect *F. quadrioculata* to find suitable conditions for existence. Whether or not this potential of the habitat actually is utilized by the species, however, cannot be answered by a model based exclusively on habitat features. This illustrates the limits of environmental characteristics for the modelling of numerical abundance: $NN_7$ and $NN_4$ cannot predict fine modulations of abundance, but can only roughly identify the constellations of habitat factors where low, medium or maximum abundance may be established. The question arises as to whether we might achieve this much simpler goal by a modelling approach that (1) is able to represent the non-linearities of the relationships between habitat features and abundance and, at the same time, (2) exhibits more transparency than a NN. Possibly the large data set is suited for the application of machine learning techniques that automatically extract explicit rules for the relationship between variables. This approach has already been successfully used in ecological modeling (DZEROSKI & al. 1997).

Consequently, NN models may be a tool to identify the potential of a site to house large numbers of individuals of *F. quadrioculata.* Since at any given constellation of habitat conditions at least some soil cores will contain only few or no individuals at all, the predicted maximum abundance attainable at those conditions is always lower than the observed maximum abundance. However, if the model is to characterize the potential of a site rather than estimate a kind of average abundance at a given constellation of habitat features, the output has to be corrected so as to give realistic values. The most simple approach would be to add another transformation and to calculate

$$x_i' = 16\ (x_i - x_{min})\ /\ (x_{max} - x_{min}) \qquad (4)$$

with $x_{min}$ being the smallest and $x_{max}$ being the largest predicted value. This transformation expands the predicted date to the interval [0 16] with the value 16 denoting $\geq 16$ individuals (cf. section 2.3. Modeling). This is justified by the assumption, that the range of abundance of *F. quadrioculata* is sufficiently well characterized by the number of 150 training patterns and that a similar range is to be expected in the test patterns. Thus, minimum and maximum of predicted abundance may be transformed into minimum and maximum abundance of the training patterns. This transformation does not affect the efficiency of the NN models reflected by slightly higher mean errors and similar standard deviations ($NN_7$: mean error = 0.2 [S.D. 3.1]; NN4: mean error = 0.8 [S.D. 3.8]. The mean error of $MR_7$ predictions, however, increases and deviates distinctly from 0 ($MR_7$: mean error = 2.2 [S.D. 3.7.]). (The transformation of course cannot bring about an improvement in the case of $MR_4$ where the »wrong« points were predicted to have high abundance values in the first place.)

KAMPICHLER (1997) has shown, that repeated random-dividing of the data-set into 150 training and 45 test patterns may yield extremely differing results in predicting total abundance of endogeic and hemiedaphic Collembola. Correlation coefficients between prediction and observation ranging from -0.06 to 0.70 with an average at 0.50 ($n = 10$) demonstrate this observation. This is most possibly due to the skewed frequency distribution of abundance, with a large number of cores containing only few individuals and only few cores containing many individuals. A successful model output can only be expected when the cores with a large number of individuals are equally represented in the training and the test patterns. If there are no (or too few) cores with a high abundance among the training patterns, a NN cannot identify any functional relationship between habitat features and abundance; if there are no cores with high abundance among the test patterns, the model output consists only of the data points with low to medium abundance – the points that appear in the left lower corner of the prediction vs. observation plots in Fig. 4 – and these points exhibit no internal structure. This critique also applies to modelling the abundance of *F. quadrioculata.* Preliminary analyses have shown, that the correlation coefficient between prediction and observation may be as low as 0.24 in $NN_7$ and as low as 0.15 in $NN_4$. Thus, even a data set based on 195 cores – which is bigger in size than a typical soil zoological sample – may not be sufficient for unequivocally modeling relationships between habitat characteristics and abundance. Taking a number of cores at each sampling point and thus determining local average density would certainly help in reducing the skewness of the frequency distribution. However, the resulting number of cores certainly surpasses any limit of feasibility. Possibly the application of local bulk samples could overcome this difficulty by minimizing the time for counting individuals in single cores (BRUCKNER & BARTH 1997).

Due to ecological processes that superimpose the pattern of habitat quality and due to difficulties arising by the skewed frequency distribution of abundance even the large data-set from the Scheyern experimental farm could not convincingly show the predictive power of habitat features for the local abundance of a dominant Collembolan species. Therefore, soil zoologists should be even more cautious in interpreting data from even smaller samples.

## Acknowledgements

I am grateful to R.Wieland (Müncheberg) for valuable information about theory and application possibilities of backpropagation networks and to M.Pilgram (Weihenstephan) for discussion and computing support. I also owe thanks to M.Scherf (Neuherberg) for putting computing facilities at my disposal, and to M.Herberstein (Melbourne) and E. Tadsen-Duch (Bonn) for language editing.

The scientific activities of the research network »Forschungsverbund Agrarökosysteme München« (FAM) are financially supported by the Federal Ministry of Culture and Science, Research and Technology (BMBF 0339370). Rent and operating expenses are paid by the Bavarian State Ministry for Education and Culture, Science and Art.

## Literature

BRUCKNER, A. & G. BARTH, 1997: Eine Mischprobenmethode zur Schätzung der Abundanz von Bodenmikroarthropoden. – In: W. DUNGER & K. VOIGTLÄNDER (ed) Bedeutung, Stand und aktuelle Entwicklung der Systematik von Bodentieren. Abh. Ber. Naturkunde Mus. Görlitz 69, 2: 245–249.

DUNGER, W. & J. FIEDLER, 1989: Methoden der Bodenbiologie. – Gustav Fischer, Stuttgart: 432 pp.

DZEROSKI, S., J. GRBOVIC, W.J. WALLEY & B. KOMPARE, 1997: Using machine learning techniques in the construction of models. II. Data analysis with rule induction. – Ecol. Model. 95: 95–111.

EKSCHMITT, K., 1993: Über die räumliche Verteilung von Bodentieren. – Dissertation, Universität Bremen. 173 pp.

FRANZ, H. & M. BEIER, 1948: Zur Kenntnis der Bodenfauna im pannonischen Klimagebiet Österreichs. II. Die Arthropoden. – Ann. Nat.–hist. Mus. Wien 56: 440–549.

FRANZ, H. & E. SERTL-BUTSCHEK, 1954: 25.Ordnung: Collembola. – In: H. FRANZ (ed) Die Nordostalpen im Spiegel ihrer Landtierwelt. – Univ.-Verlag Wagner, Innsbruck: 579–641.

FROMM, H., K. WINTER, J. FILSER, R. HANTSCHEL & F. BEESE, 1993: The influence of soil type and cultivation system on the spatial distributions of the soil fauna and microorganisms and their interactions. – Geoderma 60: 109–118.

FROMM, H., 1998: Räumliche und zeitliche Variabilität der Collembolenfauna und ihre Bedeutung für C- und N-Umsatz in einer Agrarlandschaft. – FAM-Bericht 26. – Shaker Verlag, Aachen. 190 pp.

GALLANT, S.I., 1993: Neural Network Learning and Expert Systems. – MIT Press, Cambridge MA: 365 pp.

HARTGE, K.H. & R. HORN, 1989: Die physikalische Untersuchung von Böden. 2. Auflage. – Enke, Stuttgart: 175 pp.

HEALY, M.J.R., 1962: Some basic statistical techniques in soil zoology. – In: P.W. MURPHY (ed) Progress in Soil Zoology. – Butterworths, London: 3–9.

HEFFNER, R.A., M.J. BUTLER & C.K. REILLY, 1996: Pseudoreplication revisited. – Ecology 77: 2558–2562.

HEINEMEYER, O., H. INSAM, E.-A. KAISER & G. WALENZIK, 1989: Soil microbial biomass and respiration measurements: An automated technique based on infra-red gas analysis. – Plant Soil 116: 191–195.

HEUER, M. & K. LESCHONSKI, 1985: Erfahrungen mit einem neuen Gerät zur Messung von Partikelgrößenverteilungen aus Beugungsspektren. 3. Europäisches Symposium »Partikelmeßtechnik«, Nürnberg: 515–538.

HOPKIN, S.P., 1997: Biology of the Springtails (Insecta: Collembola). – Oxford University Press, Oxford: 330 pp.

HURLBERT, S.H., 1984: Pseudoreplication and the design of ecological field experiments. – Ecol. Monogr. 54: 187–211.

KAMPICHLER, C., 1998: Modelling abundance of edaphic Collembola in an agricultural landscape by means of artificial neural networks. – In: V. PIZL & K. TAJOVSKY (eds) Soil Zoological Problems in Central. Proceedings of the 4th Central European Workshop on Soil Zoology, Ceske Budejovice, April 23–24, 1997. ISB ASCR, Ceske Budejovice (in press)

KLIRONOMOS, J.N. & B. KENDRICK, 1995: Relationships among microarthropods, fungi, and their environment. – Plant Soil 170: 183–197.

KÖHN, M., 1928: Bemerkungen zur mechanischen Bodenanalyse. III. Ein neuer Pipettapparat. – Z. Pflanzenern. Düngung Bodenk. 11: 50–54.

LARINK, O., 1997: Springtails and mites: important knots in the food web of soils. – In: G. BENCKISER (ed) Fauna in Soil Ecosystems. – Marcel Dekker, New York: 225–264.

LEK, S., M. DELACOSTE, P. BARAN, I. DIMOPOULOS, J. LAUGA & S. AULAGNIER, 1996: Application of neural networks to modelling nonlinear relationships in ecology. – Ecol. Model. 90: 39–52.

LUSSENHOP, J., 1992: Mechanisms of microarthropod-microbial interactions in soil. – Adv. Ecol. Res. 23: 1–33.

PARUELO, J.M. & F. TOMASEL, 1997: Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. – Ecol. Model. 98: 173–186.

PETERSEN, H. & M. LUXTON, 1982: A comparative analysis of soil fauna populations and their role in decomposition processes. – Oikos 39: 287–388.

RECKNAGEL, F., M. FRENCH, P. HARKONEN & K.-I. YABUNAKA, 1997: Artificial neural network approach for modelling and prediction of algal blooms. – Ecol Model 96: 11–28.

RICKLEFF, R.E., 1987: Community diversity: relative roles of local and regional processes. – Science 235: 167–171.

SINOWSKI, W., 1994: Die dreidimensionale Variabilität von Bodeneigenschaften – Ausmass, Ursachen und Interpolation. – GSF – National Research Centre for Environment and Health, Oberschleissheim (Germany): 159 pp.

VERHOEF, H.A. & L. BRUSSAARD, 1990: Decomposition and nitrogen mineralization in natural and agro-ecosystems: the contribution of soil animals. – Biogeochemistry 11: 175–211.

WARNER, B. & M. MISRA, 1996: Understanding neural networks as statistical tools. – Am. Stat. 50: 284–293.

WINTER, K., 1998: Räumliche und zeitliche Variabilität der mikrobiellen Biomasse und ihrer Aktivität in einer heterogenen Agrarlandschaft. – FAM-Bericht 20. – Shaker Verlag, Aachen: 129 pp.

ZELL, A., G. MAMIER, M. VOGT, N. MACHE, R. HÜBNER, S. DÖRING, K.-U. HERRMANN, T. SOYEZ, M. SCHMALZL, T. SOMMER & al., 1995: SNNS Stuttgart Neural Network Simulator, User Manual, Version 4.1. Report 6/95 – University of Stuttgart, Institute for Parallel and Distributed High Performance Systems (IPVR), Stuttgart: 312 pp.

## Address

Dr. Christian Kampichler
GSF Forschungszentrum für Umwelt und Gesundheit
Institut für Bodenökologie
D-85764 Neuherberg

present address:
Freie Universität Berlin
Institut für Zoologie
Grunewaldstraße 34
D-12165 Berlin
email: kampichl@zedat.fu-berlin.de

Zeitschrift/Journal: Verhandlungen der Gesellschaft für Ökologie

Jahr/Year: 1997

Band/Volume: 28_1997

Autor(en)/Author(s): Kampichler Christian

Artikel/Article: The potential of soil habitat features for the modelling of numerical abundance of Collembola: a neural network approach 151-159