ECOLOGICAL MODELLING 195 (2006) 3–10

available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/ecolmodel

Fuzzy knowledge-based models for prediction of *Asellus* and *Gammarus* in watercourses in Flanders (Belgium)

Veronique Adriaenssens[a,b], *Peter L.M. Goethals*[a,*], *Niels De Pauw*[a]

[a] *Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, J. Plateaustraat 22, B-9000 Gent, Belgium*
[b] *Ecosystems Science Group, Environment Agency, Evenlode House, Howbery Park, Wallingford, Oxon, OX10 8BD, United Kingdom*

ARTICLE INFO

Article history:
Published on line 14 February 2006

Keywords:
River ecology
Macroinvertebrates
Predictive models
Fuzzy logic
Expert knowledge

ABSTRACT

Fuzzy logic has become an interesting technique in modelling ecosystem processes and ecological assessment. Aside its capacity to take the inherent uncertainty of ecological variables into account during inference processing, it can express non-linear relations between ecological variables in a transparent way. In the present study, fuzzy knowledge-based models are constructed for the prediction of abundance levels of the macroinvertebrate taxa *Asellus* and *Gammarus* in river basins in Flanders (Belgium) and the results are validated by means of empirical data from the Zwalm river basin. Although the fuzzy models are based on a small set of input variables and the inference system is relatively simple, their performance was comparable to that of other modelling techniques, such as classification trees. This research therefore illustrates the strength of simple and robust predictive fuzzy models, and can be a valuable contribution to the practical application of predictive models for river management purposes.

1. Introduction

The use of models for the prediction of the distribution of organisms from environmental data is widespread in ecology and conservation biology (Manel et al., 2001; Guisan and Zimmerman, 2000; Guisan et al., 2002; Jørgensen, 2005). In bio-assessment, the main objectives of predictive models are the identification of major influences on species distribution, as such revealing indicator values, and the discrimination between effects of the physical habitat and pollution on species distribution (Utzinger et al., 1998). Another application is to predict the effect of management actions on the composition of biological communities (Goethals and De Pauw, 2001; Olden et al., 2002). To fulfil these objectives, abiotic and biotic variables are used to predict the abundance and presence/absence of the target organism(s) (Jongman et al., 1995; Manel et al., 2001).

The science of ecological modelling to support river quality assessment has evolved substantially during recent years (Recknagel, 2002; Guisan and Zimmerman, 2000). In the development of decision support systems for river quality management, there is today a grown interest in modelling techniques such as artificial neural networks (Lek and Guégan, 1999), decision trees (Dzeroski, 2001), evolutionary algorithms (Caldarelli et al., 1998) and fuzzy logic (Silvert, 2000).

To construct models for use in river management, mainly ecological monitoring data are used. However, such data often bear a large uncertainty, which is mostly not only epistemic uncertainty (e.g. measurement error, natural variation, . . .), but also includes linguistic uncertainty (e.g. vagueness) (Regan, 2002). Sometimes the relations between the ecosystem components are not exactly known and analytical models for establishing these relationships are not available or the data are insufficient for statistical analysis. In such a case, a model can be build based on expert knowledge and a fuzzy

* *Corresponding author*. Tel.: +32 9 2643776; fax: +32 9 2644199.
E-mail address: Peter.Goethals@UGent.be (P.L.M. Goethals).
0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.11.043

logic approach used for solving uncertainty problems (Salski, 1992; Yen, 1999).

Fuzzy set theory (Zadeh, 1965) is an artificial intelligence technique that makes use of fuzzy sets and fuzzy 'linguistic' rules to incorporate this uncertainty into the model. Classical set theory can be extended to handle partial memberships, enabling to express vague human concepts using fuzzy sets and also describe the corresponding inference systems based on fuzzy rules (Berthold, 1999). 'Fuzzy set theory' is often replaced by the term 'fuzzy logic'. The central concept of fuzzy set theory is a membership function, which represents numerically to what degree an element belongs to a set. In fuzzy set theory, an element can be a member of a particular set to a certain degree and at the same time be a member of a different set to a certain degree. To what degree an element belongs to a certain set is called the membership degree. In fuzzy rule-based systems, knowledge is represented by *if–then* rules. Fuzzy rules consist of two parts: an antecedent part stating conditions on the input variable(s) and a consequent part describing the corresponding values of the output variable(s). Usually, the case of a single output variable is considered. In Mamdani–Assilian type models, both antecedent and consequent parts consist of fuzzy statements concerning the value of the variables involved (Mamdani, 1977), whereas in Takagi–Sugeno type models (Takagi and Sugeno, 1985) the consequent part expresses a (non-)linear relationship between the input variables and the output variable.

The aim of this study was the construction of fuzzy knowledge-based models for the prediction of the macroinvertebrate taxa *Gammarus* and *Asellus* in rivers based on the Mamdani–Assilian approach (reviewed by Adriaenssens et al., 2004). These fuzzy predictive models were based on an expert knowledge database and an ecological validation set with physical–chemical variables and macroinvertebrate monitoring data. Macroinvertebrate communities are important elements in river quality management and environmental impact assessment, as emphasized in the European Water Framework Directive (EU, 2000). Sampling data from the Zwalm river basin in Flanders were used to validate the fuzzy models. The Zwalm river basin is a typical Flemish river basin and it has a very wide range of different habitat features and states of degradation. *Gammarus* and *Asellus* were chosen as representative taxa because of their highly variable presence in these headwaters and their use as bio-indicators in river quality assessment (MacNeil et al., 2002).

2. Materials and methods

For implementation of fuzzy set theory into the models, the fuzzy logic toolbox from MATLAB 5.3 for MS Windows^{TM} was used.

For validation of the models, monitoring data from the Zwalm river basin were used. The Zwalm river basin is a part of the Upper-Scheldt basin and mainly consists of numerous small brooks. It has a total surface of 11.650 ha and the Zwalm river itself has a length of 22 km. The basin is mainly polluted by untreated urban wastewater and diffuse pollution originating from agricultural activities. Habitat degradation of the watercourses is caused mainly by erosion effects. Because of its specific geomorphology, the springs are located in small but valuable forests, it has a unique fauna in the headwaters (Goethals and De Pauw, 2001). During September and October of both 2000 and 2001, 60 sites of the Zwalm river basin in Flanders (Belgium) (Fig. 1) were monitored.

The macroinvertebrates were collected with a standard handnet consisting of a metal frame holding a conical net (mesh-size 350 μm) (IBN, 1984). The handnet is held in a vertical position on the river bottom. The bottom material located immediately upstream is turned over by foot. In this way, the dislodged animals are carried into the net by the current. The objective of the sampling consists in collecting the most representative diversity of macroinvertebrates at the station examined (De Pauw and Vanhooren, 1983). The sampling method is based on a multi-habitat design, where major habitats are sampled according to their proportional distribution within a sampling reach and consisted of 10 min sampling in a 10 m reach of the watercourse. At non-wadeable places (at six sites within the Zwalm river basin), artificial substrates (De Pauw and Vanhooren, 1983) were used (three replicates).

Validation of the model's predictive results was based on the number of correctly classified instances (CCIs) (=matching coefficient, cf. Buckland and Elston, 1993; Fielding, 1999) and Cohen's Kappa (Cohen, 1960). In this study, an instance was considered as correctly classified when the predicted output had a degree of membership of more than 0.5 in a range [0 1] to the measured output class. Cohen's Kappa (Cohen, 1960) measures the proportion of all possible cases of presence or absence that are predicted correctly by a model after accounting for chance (Manel et al., 2001). CCI and Cohen's Kappa are expressed on a scale between 0 and 1.

Cohen's Kappa (K) and the number of correctly classified instances (CCI) are measured as follows:

Fig. 1 – The Zwalm river basin, located in the Upper-Scheldt basin in Flanders (Belgium).

Measured	Model		
	Absent	Present	Total
Absent	A	B	A + B
Present	C	D	C + D
Total	A + C	B + D	A + B + C + D

$$\kappa = \frac{(A+D)(A+B+C+D)-(A+B)(A+C)-(C+D)(B+D)}{(A+B+C+D)^2-(A+B)(A+C)-(C+D)(B+D)} \quad CCI = \frac{A+D}{A+B+C+D}$$

For Cohen's Kappa in medical applications (Landis and Koch, 1977), values of $K < 0.40$ are considered to indicate slight to fair model performance, values of $0.40 < K < 0.60$ moderate, and values of $0.60 < K < 0.80$ and $K > 0.80$ substantial and excellent, though this is quite arbitrary and depends on the application (Manel et al., 2001).

3. Results

3.1. *Selection of input variables, construction of the membership functions and the fuzzy rule base*

Literature research allowed the formation of an ecological knowledge database that was used to select the (physical–chemical) input variables and to construct the fuzzy sets and rules of the model. For Gammaridae, only *Gammarus pulex* was present in the watercourses of the Zwalm river basin. This species prefers streams with high flow velocity (Bayerisches Landesamt für Wasserwirtschaft, 1996). Compared to Asellidae, Gammaridae generally colonize streams with a higher stream velocity because of their superior swimming abilities (Brehm and Meijering, 1990). In accordance, *G. pulex* is almost non-tolerant for low oxygen conditions (Wesenberg-Lund, 1982), but can tolerate low oxygen concentrations when water temperatures are low. It generally prefers well oxygenated localities and temperatures well below 20 °C, which could also be derived from the induced ANN models (Dedecker et al., 2005) on the same data set. *G. pulex* is suppressed by high organic conditions (Hawkes, 1979), though can stand moderate organic pollution (Gledhill et al., 1976, 1993). It prefers substrate-heterogeneity (Tolkamp, 1980), especially detritus substrates or detritus mixed with sand or gravel or leaf material (Tolkamp, 1982). Gammaridae are more sensitive to high conductivity values than Asellidae, but at conductivity values above 1000 µS/cm, both macroinvertebrate taxa experience adverse influences (Macrofauna-Atlas of North Holland, 1990). *G. pulex* appears in all kinds of water lakes, headwaters, river tributaries, canals, . . . (Holthuis, 1956; Karaman and Pinkster, 1977; Hawkes, 1979; Verdonschot, 1990). *G. pulex* is less tolerant than *A. aquaticus* to inorganic pollutants (Martin and Holdich, 1986) and organic sewage (Whitehurst, 1991a,b). The developed model will be named as the *Gammarus* model, but refers to the *G. pulex* species.

Two *Asellus* species (*A. aquaticus* and *A. meridianus*) were present in the samples of the Zwalm river basin. These species have almost no apparent differences in ecological preferences, although *A. aquaticus* appears to be somewhat more resistant to pollution than *A. meridianus* (Gledhill et al., 1976; Chambers, 1977; Cuppen, 1980; Gongrijp, 1981; Verdonschot, 1990). *A. aquaticus* is very resistant to low oxygen conditions (Hawkes, 1979; Verdonschot, 1990) and is tolerant against organic loads. It often replaces *Gammarus* species at high levels of organic pollution (Hawkes, 1979; Verdonschot, 1990). *A. aquaticus* lives preferentially in waters where a varied detritus layer is present. Asellidae are mentioned to behave as indifferent to water velocity (Bayerisches Landesamt für Wasserwirtschaft, 1996), though other sources report they have a preference for waters with a low flow velocity and also prefer a higher width within the headwaters (Macrofauna-Atlas of North Holland, 1990). Because of their close ecological preferences in rivers, a common predictive *Asellus* model was constructed for both *A. aquaticus* and *A. meridianus*.

MacNeil et al. (2002) revealed by means of both univariate and multivariate analysis that the *Gammarus*:*Asellus* ratio was sometimes responsive to changes in parameters linked to organic pollution, but also appeared correlated with variables such as conductivity and distance from source. Holland (1976) found that, although the severity of pollution tolerated by *G. pulex* and *A. aquaticus* was only little different, the levels at which these species were highly abundant differed radically (MacNeil et al., 2002). *G. pulex* tolerates dissolved oxygen down to 2.7 mg/L and is highly abundant at 7.4 mg/L or above (Macan, 1961). *A. aquaticus* on the other hand, tolerates levels as low as 1.5 mg/L and is highly abundant at 5.8 mg/L (Holland, 1976).

Using this knowledge base, an ecological data survey and information from experts, relevant and available input variables were fuzzificated into fuzzy sets. Conductivity (µS/cm), dissolved oxygen concentration (mg/L), width (m) and stream velocity (m/s) were selected as relevant input variables. Each input variable was divided into two fuzzy sets reflecting low and high values. The output variable reflects the abundance classes for each species, and is divided into three sets reflecting low, medium and high abundance of the modelled species. Boundaries for the fuzzy sets were determined by the knowledge database. The width of the overlap between the fuzzy sets of input and output variables was defined by means of the level of uncertainty of the classification process. Construction of the membership functions of input and output variables

Table 1 – Trapezoidal membership functions of the input and output variable(s)

Shape of membership function	
Trapmf	The trapezoidal curve is a function of a vector, $\boldsymbol{x}$, and depends on four scalar parameters, a, b, c, d, as given by $f(x;a,b,c,d) = \left\{ \begin{array}{ll} 0, x \le a & \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, d \le x & \end{array} \right\}$ The parameters a and d locate the "feet" of the trapezoid and the parameters b and c locate the "shoulders"

Table 2 – Membership functions used in the fuzzy models, defined by means of the trapmf (trapezoidal) function, are explained by means of the characterizing nodes (*a*, *b*, *c*, *d*)

Input variables (trapezoidal function)	Low [*a b c d*]	High [*a b c d*]
Conductivity (μS/cm)	trapmf[0 0 1200 1200]	trapmf[400 1400 2500 2500]
Dissolved oxygen concentration (mg/L)	trapmf[0 0 10 10]	trapmf[4 12 15 15]
Water velocity (m/s)	trapmf[0 0 2 2]	trapmf[0.4 1.2 2.5 2.5]
Width (cm)	trapmf[0 0 100 100]	trapmf[0 100 1000 1000]

Output variable (trapezoidal function)	Low [*a b c d*]	Intermediate [*a b c d*]	High [*a b c d*]
Abundance (number of individuals)	trapmf[0 0 65 65]	trapmf[25 40 60 85]	trapmf[65 100 5000 5000]

Table 3 – Rule base system for *Gammarus* species

Conductivity	D.O.	Water velocity	Width	*Gammarus*
Low	Low	Low	Low	Low
Low	Low	Low	High	Low
Low	Low	High	Low	Intermediate
Low	Low	High	High	Low
Low	High	Low	Low	Intermediate
Low	High	Low	High	Intermediate
Low	High	High	Low	High
Low	High	High	High	Intermediate
High	Low	Low	Low	Low
High	Low	Low	High	Low
High	Low	High	Low	Low
High	Low	High	High	Low
High	High	Low	Low	Intermediate
High	High	Low	High	Low
High	High	High	Low	Intermediate
High	High	High	High	Intermediate

Table 4 – Rule base system for *Asellus* species

Conductivity	D.O.	Water velocity	Width	*Asellus*
Low	Low	Low	Low	Low
Low	Low	Low	High	Intermediate
Low	Low	High	Low	Low
Low	Low	High	High	Low
Low	High	Low	Low	Intermediate
Low	High	Low	High	High
Low	High	High	Low	Intermediate
Low	High	High	High	Intermediate
High	Low	Low	Low	Low
High	Low	Low	High	Low
High	Low	High	Low	Low
High	Low	High	High	Low
High	High	Low	Low	Low
High	High	Low	High	Intermediate
High	High	High	Low	Low
High	High	High	High	Low

was the same for the *Gammarus* model as for the *Asellus* model. Membership functions for the input and output variables were based on trapezoidal (Table 1) functions and are described in Table 2.

Fig. 2 gives a schematic overview of the trapezoidal-based fuzzy sets for an input variable (conductivity) and output variable (abundance of a species) as applied for the prediction of *Asellus* and *Gammarus* in rivers.

A fuzzy rule base system (Tables 3 and 4) was constructed for each model that connects the input variables to the output by means of *if–then* rules. These rules were implemented in a fuzzy inference system of the Mamdani–Assilian type, which produces a crisp output. 'And' has been used as a conjunction operator in the fuzzy rule base.

Table 5 – Results of the fuzzy predictive models correctly classified instances (CCI) and Cohen's Kappa (*K*) on a scale of [0 1]

	CCI_{rare}	K_{rare}	CCI_{low}	K_{low}	CCI_{high}	K_{high}
Gammarus	0.69	0.500	0.85	0.394	0.52	0.519
Asellus	0.83	0.359	0.93	0.400	0.84	0.157

3.2. *Validation and optimization of the fuzzy models*

CCI and Cohen's Kappa (*K*) were used to evaluate the model based on the Zwalm river basin data set. *K* and CCI values for the *Gammarus and Asellus* models are given in Table 5.

By comparing the predicted with the measured results for the Zwalm river basin, matrices of confusion (Fielding and Bell,

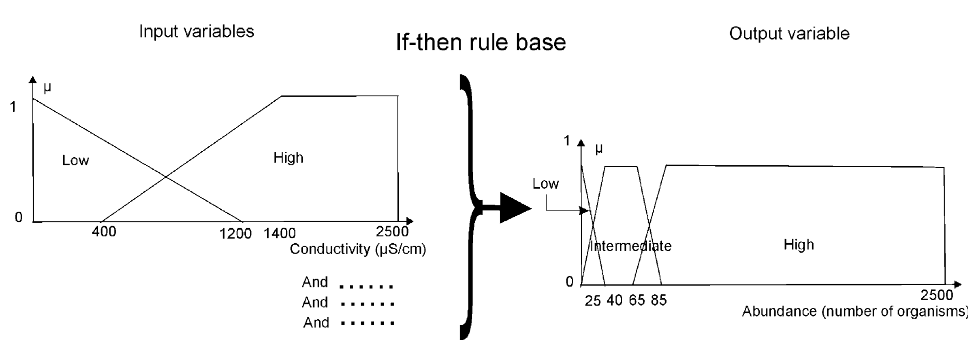

Fig. 2 – Fuzzy model for the prediction of *Asellus* and *Gammarus* in rivers with trapezoidal-based fuzzy sets of the input variables (e.g. conductivity) and the output variable (abundance), connected to each other via an *if–then* rule base.

Table 6 – Confusion matrices for the *Gammarus* model

Gammarus **low**, predicted as low true positive 54/120 = 0.45	*Gammarus* not low, predicted as not low true negative 29/120 = 0.24
Gammarus low, predicted as not low false negative 23/120 = 0.19	*Gammarus* not low, predicted as low false positive 14/120 = 0.12
Gammarus **intermediate**, predicted as intermediate true positive 8/120 = 0.07	*Gammarus* not intermediate, predicted as not intermediate true negative 94/120 = 0.78
Gammarus intermediate, predicted as not intermediate false negative 2/120 = 0.02	*Gammarus* not intermediate, predicted as intermediate false positive 16/120 = 0.13
Gammarus **high**, predicted as high true positive 15/120 = 0.13	*Gammarus* not high, predicted as not high true negative 62/120 = 0.52
Gammarus high, predicted as not high false negative 18/120 = 0.15	*Gammarus* not high, predicted as high false positive 25/120 = 0.21

Table 7 – Confusion matrices for the *Asellus* model

Asellus **low**, predicted as low true positive 85/120 = 0.71	*Asellus* not low, predicted as not low true negative 15/120 = 0.125
Asellus low, predicted as not low false negative 15/120 = 0.13	*Asellus* not low, predicted as low false positive 5/120 = 0.04
Asellus **intermediate**, predicted as intermediate true positive 3/120 = 0.03	*Asellus* intermediate, predicted as not intermediate true negative 109/120 = 0.91
Asellus intermediate, predicted as not intermediate false negative 3/120 = 0.025	*Asellus* not intermediate, predicted as intermediate false positive 5/120 = 0.04
Asellus **high**, predicted as high true positive 15/120 = 0.125	*Asellus* not high, predicted as not high true negative 86/120 = 0.072
Asellus high, predicted as not high false negative 4/120 = 0.03	*Asellus* not high, predicted as high false positive 15/120 = 0.13

1997) were constructed, identifying true positive, false positive, false negative and true negative cased for each model (Tables 6 and 7).

4. Discussion

Predictive models could be of practical use for decision support in river management. These techniques combine physical–chemical and biological data, and can assist in developing our understanding of the processes that influence aquatic organisms in running waters (Parasiewicz and Dunbar, 2001). For river managers, knowledge of the environmental factors that favour key biota can guide their planning of restoration actions and conservation management (Goethals and De Pauw, 2001; Manel et al., 2001).

At present, few applications in ecological modelling integrate knowledge-based prediction and simulation of ecological interactions in the aquatic ecosystems through fuzzy logic (Daunicht et al., 1996; Bock and Salski, 1998; Jorde et al., 2000; Kampichler et al., 2000; Mackinson, 2000). Rather, most of the predictive models used today, rely on preference function based approaches and only include hydraulic measurements (e.g. Giesecke et al., 1999; Mallet et al., 2000; Baptist et al., 2002; Parasiewicz and Dunbar, 2001; Lamouroux and Capra, 2002). These preference function models consider parameters separated from each other or in combination only with one or two other parameters. In contrast, fuzzy rules allow the inclusion of large numbers of combinations of parameters into habitat simulation tools and therefore it is easy to include more parameters, if these turn out to be relevant (Jorde et al., 2000).

Here, fuzzy models have been used for prediction of the occurrence of Crustacean species (*Asellus* and *Gammarus*) in rivers in Flanders. Prediction results evaluated by CCI and the Kappa's coefficient K reflect a moderate to good performance. The *Asellus* models seem to perform better than the *Gammarus* models based on the CCI, but when considering the Kappa's coefficient K, which allows a correction for the degree of random error in the predictions, the *Gammarus* models have the best performance. This is a consequence of the prevalence of the taxa in the river basin, an important aspect to consider within the evaluation of model results and also mentioned by Fielding and Bell (1997) and Manel et al. (2001). The abundance of *Asellus* organisms is less evenly distributed over the three constructed output classes than the one of *Gammarus*, and as such this influences the prediction results, incorporating a greater effect of chance (dominant portion of true negative predictions, see confusion matrices Tables 6 and 7).

The developed model, used in this context for the prediction of *Gammarus pulex* abundances, is not specific for any particular *Gammarus* species because there is a range of tolerances to organic pollution within this genus (Meijering, 1991; Cao et al., 1996; Walley and Hawkes, 1996). Likewise, ecological preferences for the *Gammarus* genus cannot be generalized when looking at specific geographical constraints, because different *Gammarus* species inhabit the range between upstream regions and more downstream regions with a different preference for stream velocity, water level, oxygen concentration and habitat diversity (Holthuis, 1956; Pinkster and Platvoet, 1986). Environmental preferences for both *Asellus aquaticus* and *Asellus meridianus* were generalized because of the little difference in ecological niche, except for a small difference in organic pollution tolerance (Gledhill et al., 1976; Chambers, 1977; Cuppen, 1980; Gongrijp, 1981; Verdonschot, 1990). For bio-assessment, this generalization could be of an important practical use, reducing the number of models for prediction of macroinvertebrate taxa in rivers.

Selection of the input variables comprised a combination of cost-efficiency of monitoring and relevancy of the input variables as reported in literature. This resulted in the selection of 'conductivity', 'dissolved oxygen concentration', 'width' and 'stream velocity'. The latter input variables are correlated with one another, although this was not significant in case of the Zwalm basin river data. A major part of the abundance of macroinvertebrate taxa in rivers can be explained by the input variable 'conductivity'. Due to the high values of this variable, most of the variance of this input variable has to be explained by pollution, most likely caused by agricultural activities and (treated and untreated) wastewater effluents. The same results were obvious when an ecological database for benthic macroinvertebrates, monitored in Flanders rivers, was used to validate a model based on decision trees and when input variables were selected by means of genetic algorithms. Conductivity was also one of the most relevant variables, besides dissolved oxygen, which indicates that the macroinvertebrate presence is mainly characterized by pollution-caused influences rather than natural and structural variability (D'heygere et al., 2003). The implementation of dissolved oxygen into the fuzzy models expresses an important notion of organic pollution present in the rivers. It is in that context that the Gammaridae:Asellidae ratio is used in running waters in the U.K. (Hawkes and Davies, 1971; Whitehurst, 1988). This ratio can detect subtle changes in organic pollution levels, because the change in organic load alters the relative abundance of Asellidae and Gammaridae species rather than total species composition (MacNeil et al., 2002). The overlap of the fuzzy sets of the output variable, demonstrated by the non-crisp (fuzzy) boundaries between abundance classes, reflected the uncertainty of the sampling method, which is semi-quantitative.

Although it is clear that fuzzy model techniques can be very useful in ecosystem management, still there is certainly a need for a more rigid basis for model construction and optimization. Applying genetic algorithms to adjust the shape of membership functions seems promising in this respect (Arslan and Kaya, 2001; Adriaenssens et al., 2004).

Performance of these fuzzy models is assessed by their predictive success and a whole set of validation measures is available each revealing different properties of the evaluated models (Guisan and Zimmerman, 2000; Olden et al., 2002; Guisan et al., 2002). Still, few studies perform such validation measures, as could be done by statistical validation exercises (Fielding and Bell, 1997; Manel et al., 2001; Manly, 1997), and even fewer perform field validation (Rykiel, 1996; Manel et al., 2001). In this study, CCI and Cohen's Kappa were used as performance measures of the models. The use of Cohen's Kappa (Cohen, 1960; Fielding, 1999) in combination with the CCI measure is important because it is possible to obtain high overall accuracy using trivial rules, when, for example prevalence is low (Fielding and Bell, 1997), and this can be reflected within the Cohen's Kappa statistic, although some criticism concerning overestimation exists (Foody, 1992). Matrices of confusion provided a general evaluation of the performance of the models (Manel et al., 2001), indicating if 'true positive' or 'true negative' hits seem to be the most important in the success of the predictions (Foody, 2002).

The validation set of the developed fuzzy models comprised monitoring data from the Zwalm river basin, but this data set can serve as a prototype for headwater river basins throughout whole Flanders, because a range of pollution sources (households, agriculture and small industry) and river types (large brooks, small brooks and source brooks) are included (Goethals and De Pauw, 2001).

Similar predictive results for macroinvertebrates were obtained when classification trees were used to predict the presence of Asellidae and Gammaridae validated by the Zwalm river basin data set (Goethals et al., 2001). A J48 algorithm (Witten and Frank, 2000) was used for inducing classification trees for macroinvertebrate taxa in the Zwalm river basin. Although, the application of classification trees was very useful to extract rules from a data set without prior knowledge, these rules are only based on correlations and do not reflect any kind of causality. Through this data-driven approach, only a certain preference of Asellidae for rivers characterized by a great width could be detected (Goethals et al., 2001), limiting the applications for simulating management scenarios.

In the fuzzy models, some limitations appear regarding to the scale of sampling. The spatial scale of the monitored validation data set is at the watercourse level, and encompasses a multi-habitat sampling. The models produced in this study were as such probably too robust, because collections from more than one habitat type may introduce variation that can potentially mask water quality differences among sites (Parsons and Norris, 1996). As such, the mesohabitat characteristics could be of great importance. In the future, a habitat-specific-sampling could possibly reveal this gap in knowledge.

5. Conclusion

In comparison to other predictive modelling techniques (ANN, multivariate analysis), fuzzy models have the advantage to be simple (relations between input and output variables can be explained in a linguistic-based rule base) and robust (performance is not depending on training and new input variables and rules can be easily added). The developed fuzzy models for the prediction of *Gammarus* and *Asellus* in rivers, as evaluated by CCI and Cohen's Kappa *K*, seem to perform well and can have practical application in the decision support related to water management. They can be improved, mainly through the implementation of habitat characteristics and by the hybridization of fuzzy logic with data-based modelling techniques, which ease the optimization of the models.

Acknowledgements

The first author received funding from IWT Vlaanderen, the Institute for the Promotion of Innovation by Science and Technology in Flanders.

REFERENCES

Adriaenssens, V., De Baets, B., Goethals, P.L.M., De Pauw, N., 2004. Fuzzy rule-based models for decision-support in ecosystem management. Sci. Total Environ. 319, 1–12.

Arslan, A., Kaya, M., 2001. Determination of fuzzy logic membership functions using genetic algorithms. Fuzzy Sets Syst. 118, 297–306.

Baptist, M.J., Van der Lee, G.E.M., Kerle, F., Mosselman, E., 2002. Modelling of morphodynamics, vegetation development and fish habitat in man-made secondary channels in the river Rhine, The Netherlands. In: Fourth International Ecohydraulics Symposium, Cape Town.

Bayerisches Landesamt für Wasserwirtschaft, 1996. Ökologische Typisierung der aquatischen Makrofauna.

Inforamtionsberichte des Bayerischen Landesamtes für Wasserwirtschaft. Heft 4/96.
Berthold, M., 1999. Fuzzy logic. In: Berthold, M., Hand, D.J. (Eds.), Intelligent Data Analysis. Springer, pp. 269–298.
Bock, W., Salski, A., 1998. A fuzzy knowledge-based model of population dynamics of the yellow-necked mouse (*Apodemus flavicollis*) in a beech forest. Ecol. Modell. 108 (1–3), 155–161.
Brehm, J., Meijering, M.P.D., 1990. Fließgewässerkunde. Einfürhrung in de Limnolie der Quellen, Bäche und Flüsse. Biologische Arbeitsbücher 36; 295 S. Quelle and Meyer Verlag, Heidelberg, Wiesbaden.
Buckland, S.T., Elston, D.A., 1993. Empirical models for the spatial distribution of wildlife. J. Appl. Ecol. 37, 789–805.
Caldarelli, G., Highes, P.G., McKane, A.J., 1998. Modelling coevolution in multispecies communities. J. Theor. Biol. 193, 345–358.
Cao, Y., Bark, A.W., Williams, P., 1996. Measuring the responses of macroinvertebrate communities to water pollution: a comparison of multivariate approaches, biotic and diversity indices. Hydrobiologia 341, 1–19.
Chambers, M.R., 1977. A comparison of the population ecology of *Asellus aquaticus* (L.) and *Asellus meridianus* Rac. in the reed beds of the Tjeukemeer. Hydrobiologia 53 (2), 147–154.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46.
Cuppen, H.P.J.J., 1980. De macrofauna in een aantal droogvallende en permanente stilstaande wateren in het ruilverkavelingsgebied Brummen-Voorst. Regionale milieu-raad Oost-Veluwe.
Daunicht, W., Salski, A., Nohr, P., Neubert, C., 1996. A fuzzy knowledge-based model of annual production of skylarks. Ecol. Modell. 85 (1), 65–73.
Dedecker, A.P., Goethals, P.L.M., D'Heygere, T., Gevrey, M., Lek, S., De Pauw, N., 2005. Application of artificial neural network models to analyse the relationships between *Gammarus pulex* L. (Crustacea, Amphipoda) and river characteristics. Environ. Monit. Assess. 111, 223–241.
De Pauw, N., Vanhooren, G., 1983. Method for biological quality assessment of watercourses in Belgium. Hydrobiologia 100, 153–168.
D'heygere, T., Goethals, P.L.M., De Pauw, N., 2003. Use of genetic algorithms to select input variables in artificial neural network models for the prediction of benthic macroinvertebrates. Ecol. Modell. 160, 291–300.
Dzeroski, S., 2001. Application of symbolic machine learning to ecological modelling. Ecol. Modell. 146, 263–273.
EU, 2000. Directive of the European Parliament and of the Council 20000/60/EC establishing a framework for community action in the field of water policy. European Union, The European Parliament, The Council, PE-CONS 3639/1/00 REV 1 EN.
Fielding, A.H., 1999. Machine Learning Methods for Ecological Applications. Kluwer Academic Publishers, Norwell, MA.
Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.
Foody, G.M., 1992. On the compensation for chance agreement in image classification accuracy assessment. Photogrammetric Eng. Remote Sensing 58 (10), 1459–1460.
Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sensing Environ. 80, 185–201.
Giesecke, J., Schneider, M., Jorde, K., 1999. Analysis of minimum flow stretches based on the simulation model CASIMIR. In: Proceedings 28th IAHR Congress, 22–27 August.
Gledhill, T., Sutcliffe, D.W., Williams, W.D., 1976. A Revised Key to the British Species of Crustacea:Malacostraca Occurring in Fresh Water, With Notes on Their Ecology and Distribution. Freshwater Biological Association, 71 pp. (Scientific Publication No. 32).
Gledhill, T., Sutcliffe, D.W., Williams, W.D., 1993. British Freshwater Crustacea Malacostraca: A Key with Ecological Notes. Freshwater Biological Association, 173 pp.
Goethals, P.L.M., De Pauw, N., 2001. Development of a concept for integrated ecological river assessment in Flanders, Belgium. J. Limnol. 60, 7–16.
Goethals, P.L.M., Gasparyan, K., De Pauw, N., 2001. River restoration simulations by ecosystem models predicting aquatic macroinvertebrate communities based on J48 classification trees. Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen 66 (4), 213–217.
Gongrijp, A., 1981. Biologische beoordeling van slootkwaliteit. Een literatuurstudie en een voorstel voor onderzoek in Zuid-Holland. -Hoogheemraadschap van Rijnland, technische dienst, Afd. Chemie en Technologie.
Guisan, A., Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–168.
Guisan, A., Edwards Jr., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Modell. 157 (2–3), 89–100.
Hawkes, H.A., Davies, L.J., 1971. Some effects of organic enrichment on benthic invertebrate communities in stream riffles. In: Duffey, E., Watt, A. (Eds.), The Scientific Management of Animal and Plant Communities for Conservation. Blackwell, Oxford, pp. 271–299.
Hawkes, H.A., 1979. Invertebrates as indicators of river water quality; Chapter 2. In: James, A., Evinson, L. (Eds.), Biological Indicators of Water Quality. John Wiley and Sons, Ltd.
Holland, D.G., 1976. The distribution of the freshwater Malacostraca in the area of the Mersey and Weaver River Authority. Freshwater Biol. 6, 265–276.
Holthuis, L.B., 1956. Notities betreffende limburgse Crustacea III. De Amphipoda (Vlokreeftjes) van Limburg. Nat. historisch maandblad 45 (7–8).
IBN, 1984. Biological water quality: determination of the biotic index based on aquatic macroinvertebrates, NBN T92-402. Institut Belge de Normalisation (IBN).
Jongman, R.H.G., ter Braak, C.J.F., van Tongeren, O.F.R., 1995. Data Analysis in Community and Landscape Ecology, second ed. Cambridge University Press, Cambridge, MA.
Jorde, K., Schneider, M., Zoellner, F., 2000. Analysis of instream habitat quality—preference functions and fuzzy models. In: Wang, Hu (Eds.), Stochastic Hydraulics 2000. Balkema, Rotterdam, pp. 671–680.
Jørgensen, S., 2005. Ecological modelling: editorial overview 2000–2005. Ecol. Modell. 188, 137–144.
Kampichler, C., Barthel, J., Wieland, R., 2000. Species density of foliage-dwelling spiders in field margins: a simple, fuzzy rule-based model. Ecol. Modell. 129, 87–99.
Karaman, G.S., Pinkster, S., 1977. Freshwater Gammarus species from Europe, North Africa and adjacent regions of Asia (Crustacea-Ampipoda). Part 1. Gammarus pulex-group and related species. Bijdrage tot de dierkunde 47 (1).
Lamouroux, N., Capra, H., 2002. Simple predictions of instream habitat model outputs for target fish populations. Freshwater Biol. 47, 1543–1556.
Landis, J.R., Koch, G.G., 1977. The measurements of observer agreement for categorical data. Biometrics 33, 159–174.
Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Modell. 120, 65–73.
Macan, T.T., 1961. Factors that limit the range of freshwater animal communities. Biol. Review 52, 45–70.

Mackinson, S., 2000. An adaptive fuzzy expert system for predicting structure, dynamics and distribution of herring shoals. Ecol. Modell. 126, 155–178.
MacNeil, C., Dick, J.T.A., Bigsby, E., Elwood, R.W., Montgomery, W.I., Gibbins, C.N., Kelly, D.W., 2002. The validity of the *Gammarus*:*Asellus* ratio as an index of organic pollution: abiotic and biotic influences. Water Res. 36 (2), 75–84.
Macrofauna-Atlas of North Holland, 1990. Distribution maps and responses to environmental factors of aquatic invertebrates/H.A. Steenbergen. -Haarlem, 650 pp., ISBN 90-72624-41-6.
Mallet, J.P., Lamouroux, N., Sagnes, P., Persal, H., 2000. Habitat preference of European grayling in a medium size stream, the Aim river, France. J. Fish Biol. 56 (6), 1312–1326.
Mamdani, E.H., 1977. Application of fuzzy logic to approximate reasoning using linguistic systems. Fuzzy Sets Syst. 26, 1182–1191.
Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921–931.
Manly, B.F.J., 1997. Randomization, Bootstrap and Monte Carlo Methodology Methods in Biology, second ed. Chapmann and Hall, New York.
Martin, T.R., Holdich, D.M., 1986. The acute lethal toxicity of heavy metals to peracarid crustaceans (with particular reference to freshwater Asellids and Gammarids). Water Res. 20, 1137–1147.
Meijering, M.P.D., 1991. Lack of oxygen and low pH as limiting factors for Gammarus in Hessian brooks and rivers. Hydrobiologia 223, 159–169.
Olden, J.D., Jackson, D.A., Peres-Neto, P.R., 2002. Predictive models of fish species distributions: a note on proper validation and chance predictions. Trans. Am. Fish. Soc. 131, 329–336.
Parasiewicz, P., Dunbar, M., 2001. Physical habitat modelling for fish: a developing approach. Arch. Hydrobiol. 135 (Suppl.), 239–268.
Parsons, M., Norris, R., 1996. The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. Freshwater Biol. 36, 419–434.
Pinkster, S., Platvoet, D., 1986. De Vlokreeften van het Nederlandse oppervlaktewater. Wet. Med. K.N.N.V. 172, 1–40.
Recknagel, F., 2002. Ecological Informatics. Understanding Ecology by Biological-Inspired Computation. Springer-Verlag, Berlin/Heidelberg/New York, 398 pp.
Regan, H.M., 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol. Appl. 12 (2), 618–628.
Rykiel Jr., E.J., 1996. Testing ecological models: the meaning of validation. Ecol. Modell. 90, 229–244.
Salski, A., 1992. Fuzzy knowledge-based models in ecological research. Ecol. Modell. 63, 103–112.
Silvert, W., 2000. Fuzzy indices of environmental conditions. Ecol. Modell. 130, 111–119.
Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its application to modeling and control. IEEE Trans. Syst. Man Cybernet. 15 (1), 116–132.
Tolkamp, H.H., 1980. Organism-substrate relationships in lowland streams. Ph.D. Thesis. Wageningen, The Netherlands, Pudoc.
Tolkamp, H.H., 1982. Microdistribution of macroinvertebrates in lowland streams. Hydrobiol. Bull. 16 (2–3), 133–148.
Utzinger, J., Roth, C., Peter, A., 1998. Effects of environmental parameters on the distribution of bullhead *Cottus gobio* with particular consideration of the effects of obstructions. J. Appl. Ecol. 35, 882–892.
Verdonschot, P.F.M., 1990. Ecological characterization of surface waters in the Province of Overijssel (the Netherlands). Ph.D. Thesis. Provinciale waterstaat van Overijssel/R.I.N, the Netherlands, 255 pp.
Walley, W.J., Hawkes, H.A., 1996. A computer-based reappraisal of the biological monitoring working party scores using data from the 1900 river quality survey of England and Wales. Water Res. 9, 2086–2094.
Wesenberg-Lund, C., 1939 (reprint 1982). Biologie der süsswassertiere. Wirbellose tiere. Julius Springer, Wien, Cramer, Braunschweig, Koeltz, Koenigstein.
Whitehurst, I.T., 1988. Factors effecting the *Gammarus* to *Asellus* ratio in unpolluted and polluted waters. Ph.D. Dissertation. Brighton Polytechnic, Brighton, England.
Whitehurst, I.T., 1991a. The *Gammarus:Asellus* ratio as an index of organic pollution. Water Res. 25 (3), 333–339.
Whitehurst, I.T., 1991b. The effects of sampling techniques on the *Gammarus*, *Asellus* ratio. Water Res. 25 (6), 745–748.
Witten, I.H., Frank, E., 2000. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers, San Francisco.
Yen, V.C., 1999. Rule selections in fuzzy expert systems. Expert Syst. Appl. 16, 79–84.
Zadeh, L.A., 1965. Fuzzy sets. Inform. Control 8, 338–353.