Ecological Modelling 251 (2013) 187–198

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

# Consistency of fuzzy rules in an ecological context

Juan C. Gutiérrez-Estrada[a,*], Inmaculada Pulido-Calvo[a], David T. Bilton[b]

[a] Dpto. Ciencias Agroforestales, Escuela Técnica Superior de Ingeniería, Campus de La Rábida, Universidad de Huelva, 21819 Palos de la Frontera, Huelva, Spain
[b] Marine Biology and Ecology Research Centre, School of Marine Science and Engineering, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK

ARTICLE INFO

*Article history:*
Received 16 July 2012
Received in revised form
12 December 2012
Accepted 19 December 2012
Available online 23 January 2013

*Keywords:*
Evolutionary algorithm
Water beetle
Shannon diversity index
Contingency
Generalized additive models

ABSTRACT

In this paper, we assess the performance of fuzzy inference systems (FISs) and the consistency of fuzzy rules generated from a meta-analysis exploring diversity–environment relationships, in a system of temporary and fluctuating ponds located in two regions of southern England. The analyses focus on aquatic coleopteran assemblages, which act as excellent surrogates of wider freshwater macroinvertebrate diversity. Evaluated FISs were calibrated using evolutionary algorithms and the consistency of the rules examined using a consistency index specifically developed in this work. The best fit accounted for 76% of observed variability in the Shannon diversity index across ponds in the validation phase, which was 56 points better than the benchmark value established by a generalized additive model (GAM). The analysis of fuzzy rules indicated that the basic dynamics of this system are controlled by 8 rules. Another 10 complementary rules were detected, suggesting that more than a single dimension controlled the dynamics of the system. Therefore, water beetle diversity appears to be driven by a relatively short set of rules which relate diversity and environmental factors in a non-linear manner. These rules can be grouped according to their consistency levels, which reflect differences in coleopteran community composition.


## 1. Introduction

Ever since Lotfi Zadeh developed his theory of fuzzy sets, there are many fields of science and engineering in which this theory has been implemented efficiently to solve complex problems (Zadeh, 1973; Pappis et al., 1977; Yu et al., 2004). Basically, Zadeh formalized mathematically the way in which humans interact with the environment. The human mind tends to filter out fuzziness so that decisions and actions are more easily made and, therefore, tends to perceive discrete objects and events, distinct boundaries and definite classes of things (Bosserman and Ragade, 1982). These features of fuzzy sets theory have clearly been of interest to ecologists, which explains the significant increase of applications of this heuristic to ecological problems (Meesters et al., 1998; Kampichler et al., 2000; Addriaenssens et al., 2004; Cheung et al., 2005; Mouton et al., 2009).

There are several specific properties that have popularized the use of fuzzy sets theory among ecologists. The first is that it is an intelligent computational method which does not require a very detailed mathematical description of the process to be analyzed. Generally, these models are easily integrated-implemented in fuzzy logic controllers (FLCs) or fuzzy inference systems (FISs) and utilize a form of many-valued logic. This means that a fuzzy set can be divided in different regions by mean geometric partitions associated with linguistic concepts which allow us to describe a discrete point as a function of its membership to different sets (Zeldis and Prescott, 2000). Other very interesting property of fuzzy sets theory and FISs is that the rules that control the system dynamic can be grouped and modelled into a defined rule-base as a set of identifiable and comprehensible linguistic labels. In this rule-base (also known as fuzzy associative memory or FAM), a set of antecedents or premises which also could be interpreted as independent variables ($x_i$), are related with consequents or dependent variables ($y$) in the following form: $R_i$ = IF $x_1$ is $A_{i1}(x_1)$ and $x_2$ is $A_{i2}(x_2)$ and. . .and $x_n$ is $A_{in}(x_n)$, THEN $y$ is $B_i(y)$, where $A_{i1}(x_1)$, $A_{i2}(x_2)$, $A_{in}(x_n)$ and $B_i(y)$ are linguistic concepts, and $R_i$ is the $i$th rule of the FAM. Therefore, from an ecological point of view, an important advantage of analyzing an ecosystem with fuzzy sets theory is that the information or knowledge is structured in the form of rules.

However, the knowledge extraction via rules can be an inconvenient in some circumstances. For example, let us suppose that, using an FIS, we are trying to explain the effects of five environmental variables (chemical or physical factors) on the species richness of a faunistic group. Also, let us suppose that each environmental variable, that is, each independent variable is made up of three fuzzy partitions. This implies that, theoretically, the FAM could have a total of $3^5$ rules (243 rules). In this FAM we could include rules that apparently provide contradictory information, which would reduce our ability to find an ecological meaning of the variables implied in the system. This is particularly common when the FISs are calibrated from real environmental data. Therefore, when we

* Corresponding author. Tel.: +34 959217528; fax: +34 959217528.
*E-mail address:* juanc@uhu.es (J.C. Gutiérrez-Estrada).

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2012.12.013

meta-analyze an ecosystem with FISs, it is not enough to provide an FAM. It is also necessary to analyze the consistency of rules, that is, the absence of contradictory rules in the FAM.

In the literature there are very few works that analyze the consistency of FAMs and all are focused to engineering applications (Alonso et al., 2008). For example, Jin et al. (1999) proposed a methodology based on evolutionary algorithms for generating consistent and compact FAMs for the automatic control of safety distance in cars. They provided a consistency index calculated as a function of the similarities of rule premises and rule consequents. On the other hand, Cheong and Lai (2000) used genetic algorithms (GAs) to design FISs for controlling three different industrial plans. These authors used a genetic strategy to obtain an FIS with restrictions in the FAM. In this way, the number of rules fired at the same time was minimized to improving the consistency of the FAM. In both these works it is assumed that the consistency of the FAM results in the absence of contradictory rules, in the sense that rules with similar premises or antecedents should have similar consequent parts (Gacto et al., 2011). In other words, two inconsistent rules will have similar premise parts, but will have rather different consequents.

This initial assumption on the consistency of rules could generate erroneous conclusions when we carry out a meta-analysis in an ecological context because unlike engineering problems, the environmental data base could be made up of contingent records. Contingency arises when the nature and composition of ecosystems in different locations are different realizations of the same underlying processes (Schmitz, 2010). Therefore, any consistency analysis of rules generated from environmental data must explicitly confront the issue of contingency. To cope with the above mentioned problem, a new approach of consistency based on the contingent nature of environmental data is proposed in this paper.

In summary, we use an FIS optimized with evolutionary algorithms to extract the set of rules that controls the composition of a local assemblage. We selected a system which has been previously analyzed, and therefore relatively well studied, providing a set of reference results. The system selected was composed of a series of temporary and fluctuating ponds in two regions of southern England (Lizard and New Forest). In both regions, we focused our attention on the different diversity patterns of the community of aquatic Coleopera and their relationships with environmental factors. This macroinvertebrate assemblage was selected because this group is relatively diverse, ecologically well understood and occurs across a wide spectrum of pond types (Bilton et al., 2006; Sánchez-Fernández et al., 2006). Evaluation of the accuracy and validity of each FIS model was carried out via the comparison of error levels with those obtained in a reference model, which in our case was a generalized additive model (GAM). Finally, the consistency of the extracted rules from the best FISs were analyzed with our new approach and compared with existing methodologies.

## 2. Methods

### 2.1. Study area and sampling

The dataset used in this study was obtained from 76 temporary ponds located in the New Forest (Hampshire) and the Lizard Peninsula (Cornwall), both in southern England. These regions contain a high density of temporary and fluctuating ponds, differing widely in biological, physical and chemical characteristics. A detailed description of these ecosystems can be found in Bilton et al. (2001), Rundle et al. (2002), Bilton et al. (2006) and Bilton et al. (2009).

Each pond was sampled during February and March 2000, a time when the spatial extent and the presence of ponds were at their maximum (Bilton et al., 2006). Ponds were sampled using a hand net (1 mm mesh, dimensions 20 cm × 25 cm), taking semi-quantitative 1 m sweeps amongst aquatic vegetation. Each 1 m sweep involved approximately 10 s of back and forth netting over the same area of habitat. This sampling protocol has been favourably evaluated in several investigations on invertebrate assemblages (Rundle et al., 2002; Foggo et al., 2003; Bilton et al., 2006). Two or three such samples were taken from the largest sites according to their area. Sweeps were pooled and samples preserved in 95% ethanol in the field.

In addition, a wide range of environmental variables was recorded. Before Coleoptera were sampled pH, temperature compensated conductivity and turbidity readings were taken on-site using a Solomat 520C probe (Zellweger Analytics, Poole, UK). Water depth in the area sampled was estimated using a 1 m rule (mean of five measurements). For analysis of metal cations (calcium, magnesium, aluminium, nickel, chromium, cobalt, iron, zinc and copper) and nutrient concentrations (organic nitrate and soluble reactive phosphorus), two water samples from each pond were also collected. However, Gutiérrez-Estrada and Bilton (2010) have previously demonstrated that only four environmental variables (conductivity, turbidity, magnesium concentration and depth) were enough to explain more than 82% of the variation of water beetles diversity in New Forest and Lizard regions. Therefore, in this study only these four variables were considered as inputs to the FIS.

Later, in the laboratory beetles were counted and determined to species level. Shannon's index ($H'$) was calculated for Coleoptera from each pond following Brower et al. (1998). This diversity measure (the output of the FIS) was selected because it is easy to calculate and reflects both species richness and the relative abundance of species within assemblages. $H'$ normally varies between 1.5 and 3.5, with values higher than 3 being seen as representing diverse communities whilst those below 2 are relatively uniform (Cowell et al., 2004).

### 2.2. Fuzzy inference system (FIS)

A fuzzy inference system is a model in which the theory of fuzzy sets has been implemented. In these models, three types of parameters must be optimized in function of input and outputs variables: (1) the shape of the fuzzy sets or geometrical partitions; (2) the degree of overlap between fuzzy sets; and (3) the definition of the IF-THEN rules. In our case, both the independent variable or input variable (the diversity of aquatic Coleoptera community measure as Shannon's index) and dependent variables (environmental factors: conductivity, turbidity, magnesium concentration and depth) were divided in three geometrical partitions labelled as 'Low Level', 'Normal Level' and 'High Level' for which four types of normal and convexes membership functions were tested: (a) symmetric and non-symmetric triangular forms; (b) symmetric and non-symmetric trapezoidal forms; (c) symmetric and non-symmetric non-continuous multipoint forms (with seven points); and (d) symmetric and non-symmetric continuous PI forms (Fig. 1).

In this work, an evolutionary algorithm was used to find the optimal values of the model parameters. Evolutionary algorithms are non-linear search and optimization methods inspired by the biological processes of natural selection and survival of the fittest (Holland, 1975; Goldberg, 1989). Generally, in an evolutionary algorithm the basic unit is the gene. Various genes contain the information required to define a chromosome whose decoding is interpreted as an individual. Thus, the parameters of the model were coded as genes in the chromosome. Once the initial information has been coded, four classical types of operators (reproduction, crossover, mutation and abort) were used in order to evolve towards a suboptimal fuzzy configuration. The steps in the