C. R. Biologies 334 (2011) 695–704

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Biological modelling/Biomodélisation

Performance comparison among multivariate and data mining approaches to model presence/absence of *Austropotamobius pallipes* complex in Piedmont (North Western Italy)

Comparaison des prestations entre des techniques de statistique multivariée et data mining pour prévoir la présence/absence de Austropotamobius pallipes *complex au Piémont (Italie nord-occidentale)*

Tina Tirelli *, Livio Favaro, Marco Gamba, Daniela Pessani

Dipartimento di Biologia Animale e dell'Uomo, Università degli Studi di Torino, Via Accademia Albertina, 13, 10123 Torino, Italy

ARTICLE INFO

Article history:
Received 19 October 2010
Accepted after revision 10 July 2011
Available online 23 August 2011

Keywords:
Freshwater ecosystem
Management
Logistic regression
Decision trees
Artificial neural network

ABSTRACT

Freshwater inhabitants in Piedmont (Italy) have been deeply disadvantaged by environmental changes caused by human disturbance. Hence there are engendered species that need human intervention of an entirely different kind – better management through the development of innovative practical tools. The most ecologically important of the river-dwelling invertebrates is a threatened species, the native white-clawed crayfish *Austropotamobius pallipes*. This is the species that we focused on in our effort to contribute to species conservation. Specifically we contrasted three different techniques of managing data relating to the presence/absence of this species: logistic regression, decision-tree models and artificial neural networks (ANN). Logistic regression and decision tree models (unpruned and pruned) performed worse than ANN. In this case, tree-pruning techniques did not make these models significantly more reliable, but did make the trees less complex and therefore did make the models clearer. ANN performed the best. Therefore we have judged them to be the most effective techniques.

© 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Freshwaters, which are rapidly deteriorating all around the world, have been the focus of more and more attention [1–3]. This attention has inspired many studies analyzing the ecological, environmental and habitat factors that affect the distribution of freshwater organisms at different spatial scales. However, one kind of freshwater organism that has been relatively neglected is the crustacean [4–9]. In relation to crustaceans, we endeavored to analyze the relationship between species distribution and ecological factors, a fundamental step towards increasing our knowledge of freshwater ecosystems, of the communities associated with them, and of information important for management and conservation. Worldwide, freshwater habitats are being subjected to such marked human disturbance that the extinction rate of freshwater species is predicted to be five times that of terrestrial species and three times that of coastal marine mammals [10]. All this hastens us to foster habitat and species preservation by developing practical tools for assessing running waters and species conditions ecologically.

* Corresponding author.
E-mail address: santina.tirelli@unito.it (T. Tirelli).

1631-0691/$ – see front matter © 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crvi.2011.07.002

The biological model we used in this research project is the white-clawed crayfish *Austropotamobius pallipes* complex, the biggest indigenous freshwater invertebrate in Western and Central Europe [11,12]. Over the last few decades, European populations of native crayfish have been fragmented and have declined all over the continent [13]. Human disturbance has provoked habitat fragmentation, deforestation and water deterioration. Larger, more aggressive, and quicker-growing non-native crayfish [14–17] have been introduced. On top of this, human disturbance is liable to become even more severe in the future while non-indigenous species are transmitting the crayfish plague due to *Aphanomyces astaci* (Schikora, 1906) [18].

Obviously, *A. pallipes* has been in need of special protective measures and so was listed as "vulnerable" on the Red List of threatened animal species compiled by the International Union for the Conservation of Nature and Natural Resources [19] and in annexes II and V of the Habitat Directive (Council of the European Communities, 1992, 1997). In Piedmont (NW Italy), *A. pallipes* is protected locally by a Regional Law (L.R. number 37 dated 29/12/06), which lays down new regulations for the management of aquatic fauna, habitat, and fishing. In particular, it provides policies aimed at re-establishing consistent populations of native species.

A. pallipes, like other native crayfish, is considered a keystone species [20], an important component of many food webs in freshwater ecosystems [21–24]. Crayfish are involved in the food chain: they are prey for vertebrate predators [25] and, in turn, are omnivorous feeders with a significant impact on community structures [26–31]. They play an important role in the well-being of running water ecosystems [32] and take part in the cycling of matter and the flow of energy [33]. Although *A. pallipes* have long been considered valid bioindicators of water quality [34–36], they also inhabit moderately polluted waters [8,9,37]. These were the factors that have led us to investigate the relationship between the environment and the presence/absence of *A. pallipes*.

In our research project, we have used modeling, a tool being considered more and more important for defining management and conservation policies. Ecosystems have highly complex nonlinear relationships among their input variables, and so researchers have been applying machine-learning methods to ecology in the last decade [38–46]. One reason is that machine-learning techniques introduce fewer prior assumptions about the relationships among the variables and hence are better than traditional statistical analysis in many ways. There are many machine learning techniques. However, decision trees [47], artificial neural networks [48], fuzzy logic [49], and Bayesian belief networks [50] are the techniques that seem to model habitat suitability the best [41,51].

Our research project evaluates the reliability of various current classification techniques in modeling *A. pallipes* presence/absence and ranks their performances. We used two types of approaches. Firstly, we used the multivariate-statistics approach, where we applied logistic regressions (LRs). Secondly, we used the machine-learning approach, where we applied decision trees (DTs) and artificial neural networks (ANNs). These types of machine-learning techniques have been used at various rates – ANNs quite often from mid-1990s [44–46,48,52–62], DTs sporadically [41,45,46], and LRs most frequently [56].

2. Material and methods

2.1. Study area and data collection

We chose sites for sampling *A. pallipes* distribution on the basis of both recent information and historical records – by examining the literature, by collecting information from museums, and by contacting local town administrators, natural-park and wildlife-reserves personnel, and local people. The 175 sites we chose covered a total area of 25,399 km^2 and were located along brooks and small tributaries flowing into the Po River. They mostly were characterized by running waters inhabited by native crayfish in the past. We performed samplings from late spring to early autumn 2005–2009 in all 8 provinces of Piedmont Region: Alessandria, Asti, Biella, Cuneo, Novara, Verbania, Vercelli, and Torino.

The sites had geological conditions typical for Piedmont, ranging from the siliceous to the calcareous, and therefore widely varied in their physical and chemical characteristics. Species presence was assessed both during the day performing manual surveys (2 people for 1 hour) and at night using traps ($50 \times 25 \times 25$ cm with a 3 mm mesh size, baited with pig or chicken liver, left overnight). Each site was sampled three times before considering it not inhabited by the crayfish.

2.2. The choice of input variables

The more the parameters used, the more complex the models are, the greater the calculation times, the greater the field data collection efforts, and – unfortunately – the more obfuscated the models. Accordingly, we chose only a few variables, those most important for detecting *A. pallipes* presence, as reported [8,9,15].

2.2.1. Environmental variables

Some stream characteristics were considered *in situ*: altitude; width at moderate flow; width at high flow; percentages (0–100%, not classes) of the sampled area classified according to granulometry-bedrock, boulders and pebbles, medium gravel (≥ 1 cm), little gravel (1 cm $<$ dimension $\leq$ 2 mm), sand and silt (dimension $<$ 2 mm); water velocity; and amount of shade (classes 0–5; the larger the shade, the larger the value).

2.2.2. Physical-chemical variables

In each site we measured pH, conductivity (C) and percentage of dissolved oxygen (DO) through a multiparameter probe (mod. Hydrolab Quanta). To avoid floating materials, we set a 15 cm depth for collecting two 100 mL water samples from each site. We stored the samples in sterile polythene test tubes and froze them until they were analyzed chemically. We measured the concentrations of the following inorganic ions that are commonly used to assess water quality: ammonium (NH_4^+), nitrates (NO_3^-), ortho-phosphate (PO_4^{3-}), chlorides (Cl^-), sulfates (SO_4^{2-}),

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات