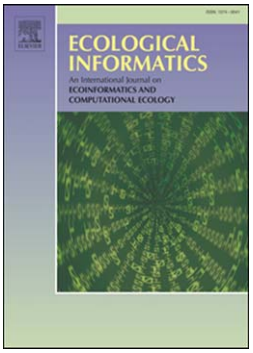
Accepted Manuscript



An applicability index for reliable and applicable decision trees in water quality modelling

G. Everaert, E. Bennetsen, P.L.M. Goethals

PII: S1574-9541(15)00200-9
DOI: doi: 10.1016/j.ecoinf.2015.12.004
Reference: ECOINF 646

To appear in: *Ecological Informatics*

Received date: 9 April 2015
Revised date: 30 November 2015
Accepted date: 6 December 2015

Please cite this article as: Everaert, G., Bennetsen, E., Goethals, P.L.M., An applicability index for reliable and applicable decision trees in water quality modelling, *Ecological Informatics* (2015), doi: 10.1016/j.ecoinf.2015.12.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

An applicability index for reliable and applicable decision trees in water quality modelling

G. Everaert[1,+] , E. Bennetsen[1,+,*], and P.L.M. Goethals[1]

[1] Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Ghent, Belgium

* Author for correspondence: Tel: +32(0)92643996; E-mail: Elina.Bennetsen@UGent.be

[+] These authors contributed equally to this work

Key words: decision trees, decision making, freshwater ecology, water quality management, low countries

Highlights:

- The applicability index (API) quantifies the applicability of decision trees
- To do so, the API integrates the simplicity submetric and the binary submetric
- The value of the API is illustrated on the basis of two water quality datasets
- The API is a complement to existing statistical criteria for model selection

Abstract

Data driven environmental models are mainly assessed on the basis of their model fit and only limited attention is given to their applicability for end-users. In this paper we present the applicability index (API) that scores decision trees in terms of their interpretability and applicability for end-users. The API integrates two criteria, viz. the simplicity of the model and its ability to predict the classes of the response variable. We developed 10,000 decision trees with different parameterizations and assessed the use of API for model selection with two different datasets. The API reduced the number of decision trees that were retained only based on statistical criteria from 2,806 to 173 and from 1,117 to 784 respectively. The models that were retained were more easily interpretable, equally statistically reliable but less complex. Conventional statistical criteria such as Cohen's Kappa and the number of Correctly Classified Instances were only moderately correlated with the API ($r = 0.26$ and $r = 0.49$, respectively). This indicates that the API is a useful complement to the existing statistical criteria available for model selection. The API was tested for two datasets consisting of water quality data in lowland rivers in Belgium and the Netherlands, hence its validity needs to be tested for other types of data and modelling domains.

Download English Version:

https://daneshyari.com/en/article/6295834

Download Persian Version:

https://daneshyari.com/article/6295834

Daneshyari.com