Global Ecology and Conservation 4 (2015) 589–601

Contents lists available at ScienceDirect

Global Ecology and Conservation

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

Original research article

Multiple factors and thresholds explaining fish species distributions in lowland streams

Cristina Trigal[a,*], Erik Degerman[b]

[a] Swedish Species Information Center, Swedish University of Agricultural Sciences, Uppsala, Sweden

[b] Department of Aquatic Resources, Swedish University of Agricultural Sciences, Drottningholm, Sweden

ARTICLE INFO

Article history:
Received 4 May 2015
Received in revised form 21 October 2015
Accepted 21 October 2015
Available online 7 November 2015

Keywords:
Threshold values
Boosted regression trees
Random forest
Sweden
Headwater streams

ABSTRACT

Appropriate restoration and conservation measures require a good understanding of the factors limiting the distribution of species, the presence of steep changes in the distribution along environmental gradients and the effect of environmental interactions on species distribution. We used 12 environmental variables describing connectivity, hydrology, climate and stream morphology, to model the distributions of 17 fish species from 2005 Swedish stream sites that were sampled between 2000 and 2011. Modeling was performed using boosted regression trees and random forest, two machine learning techniques to assess the relationship between species distributions and their environment. Temperature, width and connectivity (minimum distance to lake or the sea and water discharge), were the most important variables explaining changes in species distribution at large spatial scales. Response curves of fitted occurrence probabilities along predictors often showed abrupt changes, however, clear threshold effects were difficult to detect. Our results show also differences across species and even in the outcomes of the two algorithms, implying that a simultaneous assessment of multiple species may provide a better signal of ecosystem change than the use of surrogate species.

1. Introduction

Freshwater habitats, supporting ca 10% of all known species, are among the most threatened ecosystems in the world (Vorosmarty et al., 2010). Some of the most severe threats to freshwater biodiversity, such as habitat degradation, flow regulation, and species invasion, result in loss of taxa richness (Schinegger et al., 2012), declines in the distribution range and abundance of many species (Baxter et al., 2004; Byström et al., 2007), and eventually have negative effects on ecosystem functioning. Degradation of freshwater ecosystems will continue, as water demand and physical alterations will increase with human population density (Degerman et al., 2007; Schinegger et al., 2012), and as a result of anthropogenic induced climate change (Buisson et al., 2008; Griffiths et al., 2014). In an attempt to prevent European freshwater systems from further degradation, the Water Framework Directive (European Commission, 2000) and the EC Habitats Directive (Council of the European Communities, 1992) were developed, which aim at maintaining and restoring freshwater habitats to a favorable conservation status through the development of management and restoration strategies. This is, however, a difficult task because species are affected by multiple factors acting at different spatial scales (e.g. local and

* Correspondence to: Swedish Species Information Center, Swedish University of Agricultural Sciences, Box 7007, 750 07, Uppsala, Sweden. Tel.: +46 0 18672327.

E-mail address: cristina.trigal@slu.se (C. Trigal).

http://dx.doi.org/10.1016/j.gecco.2015.10.009

Table 1
Environmental predictors used in the models.

Predictor name	Type
Water temperature (°C)	Continuous (range, 5–27°C)
Substrate	Categorical (A, <0.02 cm; B, 0.02–0.2; C, 0.2–2; D, 2–10; E, 10–20; F, 20–30; G, 30–40; H, 40–200; $I > 200$ cm)
Annual discharge ($m^3\ s^{-1}$)	Continuous (range 0.002–966 $m^3\ s^{-1}$)
Width (m)	Continuous (range, 0.3–10 m)
CV discharge	Continuous (range, 3.8–242)
Minimum distance to lake or sea (km)	Continuous (range, 0.1–10 km)
Woody debris (number 100 m^{-2})	Continuous (range, 0–167 pieces 100m^{-2})
Mean depth (m)	Continuous (range, 0.02–1.3 m)
Shade (%)	Continuous (range, 0–100)
Barriers	Categorical (U, upstream; D, downstream; B, upstream and downstream the sampling point)
Sampling effort	Number of electrofishing passes (range, 1–3)
Flow velocity	Categorical (S, slow; F, fast)

catchment) (Degerman et al., 2007; Schinegger et al., 2012; Törnblom et al., 2011), and because they respond individually to environmental change (Olden et al., 2006; Parmesan and Yohe, 2003). To achieve a good conservation status and identify appropriate restoration and conservation measures, it is therefore necessary to first identify the factors limiting the distribution of individual species and to evaluate the effects of interactions among environmental drivers on species distribution (Guisan et al., 2013).

Conservation management also requires a good understanding of threshold effects along environmental gradients that may cause abrupt changes in species distribution (Roni et al., 2008), i.e. how much and what quality of habitat is required for different species in different environments? Thresholds are, however, difficult to predict, as they depend on a number of factors including landscape characteristics, species traits and non-linear relationships between species and the environment (Lindenmayer and Luck, 2005; Suding and Hobbs, 2008). In addition the interactions among environmental drivers may affect threshold values and produce complex responses in species distribution (Olden, 2007; Pittman and Brown, 2011), and complicate the outcomes of restoration. For example, many restoration programs in streams aim at increasing habitat heterogeneity through adding large woody debris or manipulating stream substrate; however, how fish species perceive environmental heterogeneity will depend on the interactions between the variable of interest and other local and regional variables (e.g. water level fluctuations, presence of barriers, etc.). Species distribution models are used to evaluate habitat suitability and the existence of thresholds in species occupancy over large spatial and temporal scales (Elith and Leathwick, 2009; Guisan et al., 2006, 2013). Those models often include non-linear relationships between species occurrence or abundance and habitat variability (Elith and Leathwick, 2009; Guisan et al., 2006). However, only a few studies have looked explicitly into the effect of interactions among drivers on threshold values and species occupancy.

In this study we use an extensive data set describing the distributions of 17 fish species across lowland streams in Sweden, sampled between 2000 and 2011. We use 12 environmental variables describing connectivity, hydrology, climate and stream morphology, which are important for fish (Degerman et al., 2004; Morin and Naiman, 1990; Rifflart et al., 2009). The aims of the study are to: (a) identify the drivers that contribute most to the distribution of individual species and community turnover; (b) identify changes in environmental drivers that result in abrupt changes in species occurrence (threshold effects); and (c) evaluate the consistency of the species–environment relationships over time. We use boosted regression trees (De'ath, 2007) and random forest (Hothorn et al., 2006), two machine learning techniques to assess the relationship between species distributions and their environment.

2. Methods

2.1. Study area

Fish and environmental data were drawn from the Swedish Electrofishing Register (SERS), a database containing more than 56 500 records from 17 500 sites sampled across Sweden from 1951 onwards. For this study we selected a subset of 2005 lowland sites sampled at least once between 2000 and 2011. The study sites were located at altitudes lower than 200 m a.s.l (see Fig. 1). This boarder coincides roughly with the Swedish highest coastline, which acts as a natural barrier and plays a role in limiting the dispersal of lowland fish species into streams at higher altitudes (Ekman, 1922). We selected sampling sites with a wetted width less than 10 m, due to the reduced effectiveness of electrofishing by wading in wide streams (Kennedy and Strange, 1981). Water temperature at the time of sampling ranged from 5 °C to 27 °C. The surrounding landscape consisted of forest, with coniferous species dominating, and agricultural lands, particularly in southern Sweden. Other environmental variables are described in Table 1.

2.2. Fish sampling

Sampling was performed in August, according to national standards. At each site a 20–50 m long transect (total area 200–300 m^2) was sampled by electric fishing, using a bank-based generator operated by a two-crew team wading and using