Ecological Modelling 257 (2013) 1–10

Contents lists available at SciVerse ScienceDirect

Ecological Modelling

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

Choice of study area and predictors affect habitat suitability projections, but not the performance of species distribution models of stream biota

Sami Domisch[a,b,*], Mathias Kuemmerlen[a,b], Sonja C. Jähnig[a,b,1], Peter Haase[a,b,1]

[a] Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt am Main, Germany
[b] Senckenberg Research Institute and Natural History Museum Frankfurt, Department of River Ecology and Conservation, Clamecystrasse 12, D-63571 Gelnhausen, Germany

ARTICLE INFO

Article history:
Received 14 August 2012
Received in revised form 15 January 2013
Accepted 18 February 2013
Available online 21 March 2013

Keywords:
Model accuracy
Commission error
Landscape
Pseudo-absences
River
TSS

ABSTRACT

Species distribution models (SDMs) that provide extrapolations of species habitat suitability are increasingly being used in stream ecosystems, however the effects of different modelling techniques on model projections remain unknown. We tested how different study areas and predictors affect SDMs by using consensus projections of a fixed set of 224 stream macroinvertebrate species and five algorithms implemented in BIOMOD/R. Four modelling designs were applied: (1) a landscape as a continuous study area without any discrimination between terrestrial and aquatic realms, (2) a stream network masked a posteriori from the previous design, (3) a stream network as the study area during the model-building stage, and (4) same as (3) but with a hydrologically corrected set of predictors. The true skill statistic (TSS) and accuracy of the consensus projections were not influenced by the different designs (TSS ranged from 0.80 to 1.00, accuracy ranged from 0.70 to 0.96). The projections of design (4) yielded a strong reduction in false positive predictions compared to (1) (on average by 56%), (2) (11%) or (3) (8%). Our results show how SDMs with equally high accuracy may differ widely in habitat suitability projections for benthic macroinvertebrates. As model performance and output are not necessarily congruent, habitat suitability projections of stream biota need to be carefully assessed.

1. Introduction

Estimations of the potential effects of climate change on species' ranges are important for understanding species' habitat suitability patterns under changing climatic conditions, and for mitigation and possible conservation efforts (Elith and Leathwick, 2009; Araújo et al., 2011). Species distribution models (SDMs) are promising and increasingly used tools for this task. One great challenge when using SDMs is their optimisation regarding study area, predictors, and presence–absence data to avoid false positive and negative predictions. In stream ecosystems, this optimisation is challenging, as false positive and negative predictions may be projected in terrestrial areas and not in the stream network, depending on the study area and thus on the spatial scale (Strayer et al., 2003). The effects of building a model on a continuous landscape as a study area (see e.g., Dominguez-Dominguez et al., 2006; Loo et al., 2007; Labay et al., 2011 for fish; Cordellier and Pfenninger, 2008; Bálint et al., 2011; Sauer et al., 2011; Shah et al., 2012 for aquatic invertebrates), without taking into consideration the stream network in which species were recorded are unknown and constitutes a major disadvantage for the further development of models, e.g., in terms of sensitivity analyses. Landscape-based models clearly provide a useful first approximation, e.g., in terms of climate-change-related vulnerability analyses, such as the means by which warming and changes in precipitation patterns might affect species' distributions (*sensu* Pearson and Dawson, 2003). However, the distribution and abundance of freshwater biodiversity also depend on other factors, considered in so-called catchment-related variables (Frissell et al., 1986; Poff, 1997; Allan, 2004). In the case of benthic macroinvertebrates or fish, structural (e.g., depth, velocity and substrate) and functional properties (e.g., flow regime, thermal regime and energy sources), both of very dynamic nature, influence community composition (Clausen and Biggs, 1997; Noss, 1990; Lancaster and Hildrew, 1993; Statzner et al., 1988; Brooks et al., 2005; Effenberger et al., 2006). When modelling at a catchment scale (i.e. large or landscape scale without taking the stream network into account), variables regarding the microhabitat in terms of stream flow

* Corresponding author at: Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt am Main, Germany.
Tel.: +49 6051 61954 3125; fax: +49 6051 61954 3118.
E-mail addresses: sami.domisch@senckenberg.de (S. Domisch), mathias.kuemmerlen@senckenberg.de (M. Kuemmerlen), sonja.jaehnig@senckenberg.de (S.C. Jähnig), peter.haase@senckenberg.de (P. Haase).

[1] These authors contributed equally.

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.02.019

conditions are inevitably ignored as their resolution tend to be too coarse to portray them accurately (Strayer et al., 2003; Allan, 2004; Newton et al., 2008).

The choice of whether the continuous landscape or stream network is used as the study area for predicting the distributions of stream macroinvertebrates has several relevant aspects, but the issues of species' presence–absence data and the choice of predictors used for delineating species ranges should be considered as the most important. In general, SDMs combine species' presence data with environmental predictors that yield species' habitat suitability after being extrapolated in space or time. SDMs can be roughly divided into two groups depending on the origin of the species records: presence–absence and presence-only SDMs (Elith and Leathwick, 2009 and references therein). SDMs of the former type use species' recorded absences and are thus based on species' true environmental envelopes, whereas those of the latter use background data or pseudo-absences for generating probabilities of species' habitat suitability. Because data sets with recorded absences of species are scarce, pseudo-absences are widely used (Lobo and Tognelli, 2011; Stokland et al., 2011). Obviously, the properties of pseudo-absences are highly dependent on the study area, and they are dependent on the sampling scheme used for them (i.e., the entire study area, or only a part of it; sampled by means of a specific procedure). In stream ecosystems, pseudo-absences can be allocated either distant (i.e., on the continuous landscape for instance on the catchment scale) or near (within the stream network on the reach scale, Allan, 2004) to species' environmental envelopes, likely affecting model performance (VanDerWal et al., 2009; Lobo et al., 2010; Barbet-Massin et al., 2012). In general, Lobo et al. (2010) define three types of species absences, which may be applied to stream ecosystems: contingent absences (i.e., the habitat is suitable but the species is absent due to restrictive forces such as dispersal limitations or biological interactions); environmental absences (the species is absent due to lack of suitable environmental or climatic conditions, Poff, 1997), and methodological absences (species presence is not detected due to biased sampling design or scarcity of survey information, Haase et al., 2004, 2006). The possibility that absences may be contingent or methodological illustrate that true absence data of stream organisms are difficult to record. A suitable work-around for calibrating and fitting SDMs is to use background data, i.e. pseudo-absences *sensu* Lobo et al. (2010). In the case of stream ecosystem modelling, the choice of study area is likely to affect the environmental absences, which can be allocated either on the entire landscape area or exclusively within the stream network. Thus, the model accuracy and the quantity of species' false positive and negative predictions are likely to be affected by the choice of study area because these absences differ in their distances (either near of far) to species' recorded presence records (Sowa et al., 2005; Hopkins, 2009; Mynsberge et al., 2009), especially considering that these predictions could be allocated beyond freshwater habitats, where evidently no suitable habitat is found.

Second, the choice of the study area inevitably influences the choice of predictors used in SDMs through the medium itself but also through scale, resolution, and availability of the data. On a continuous landscape, coarse-scale predictors, such as air temperature and precipitation, are useful to describe the climatic envelopes in which species occur, whereas predictors describing stream-specific conditions (e.g., stream type, flow accumulation) play a larger role at fine scales (hierarchical modelling framework *sensu* Pearson and Dawson, 2003; Allan, 2004). In contrast, when moving into finer scales, SDMs based on a stream network may include more specific predictors (landscape filter hypothesis, Poff, 1997) that allow simple hydrological predictors, such as stream type, flow accumulation or stream order, to be included, which are of relevance for characterising the habitat suitability of stream assemblages and communities (e.g., McNyset, 2005; Domisch et al., 2011). At even finer scales, entirely watershed-based models have been set up based on single stream segments (e.g., Steen et al., 2008; Dauwalter and Rahel, 2008; Wilson et al., 2011; Kuemmerlen et al., 2012).

However, working at reach scales also means dealing with additional uncertainties. For instance, small-scale variations of the stream topography are important to take into account, and predictors may need to be corrected because of spatial differences between the underlying digital elevation model (DEM) and the digitised stream network layer. The correction of relevant predictors based on the DEM, e.g., by filling artificial sinks derived from inaccurate remote sensing data, can therefore have a significant effect on model performance and thus on the projections of species' habitat suitability (Adriaenssens et al., 2004). The choice of which environmental predictors to apply in a model, as well as their extent, scale and resolution, both spatial and temporal, is frequently based more on their technical properties than on the biological interactions they have with the studied organisms. Here, freshwater ecosystems are a good example of how abiotic factors exert their influence on organisms at distinct levels (Malmqvist, 2002).

In this study, we analyse the effects of the extent of the modelled area and the choice of predictors on SDMs of stream macroinvertebrates, which are important ecologically and as indicators of stream condition. Using a fixed set of species, we vary the choice of the study area before and after the model-building stage using a fixed set of predictors. Here, models are either built on a continuous area and projections are masked a posteriori to the extent of a stream network, or they are built directly on a stream network. Moreover, we vary the choice of predictors from a non-corrected to a hydrologically corrected set within a fixed study area. We hypothesise that (1) a continuous landscape design will have the most false positive predictions because the terrestrial and aquatic realms are not differentiated at the model-building stage, (2) using a stream network as the study area at the model-building stage will increase model accuracy and reduce the number of false positive predictions because pseudo-absences will not include those ranging into terrestrial areas and (3) an inclusion of a hydrologically corrected set of predictors during the model-building stage will enhance the model accuracy and reduce the number of false positive predictions, as species' environmental envelopes are more accurately delineated.

2. Materials and methods

2.1. Modelling designs

Four different modelling designs were applied (see Fig. 1A–D). In the most basic approach, we modelled species' distributions on a continuous landscape area (hereafter referred to as a 'landscape' design, Fig. 1B), without any discrimination between streams and the terrestrial area.

In the second design, a stream network mask was applied to the 'landscape' projections, as the species are supposed to inhabit the streams and rivers (hereafter 'landscape masked', Fig. 1C). This design is identical to the previous 'landscape' design except that it is restricted to the grid cells of the river network (i.e., the projection for the subsequent analyses).

In the third design, the stream network area was masked prior to fitting the models; thus, only the stream network was considered at the model-building stage (Fig. 1D). For this design, we used an identical set of predictors as in the 'landscape' and 'landscape masked' designs (hereafter referred to as the 'stream network' design).

The last design also modelled species' distributions on the stream network, but used a partly different set of predictors to test for effects derived from using hydrologically corrected