Ecological Modelling 342 (2016) 1–18

Contents lists available at ScienceDirect

Ecological Modelling

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

# Landscape and flow metrics affecting the distribution of a federally-threatened fish: Improving management, model fit, and model transferability

Thomas A. Worthington[a,1], Tianjiao Zhang[b], Daniel R. Logue[a], Aaron R. Mittelstet[c,d], Shannon K. Brewer[e,*]

[a] Oklahoma Cooperative Fish and Wildlife Research Unit, Oklahoma State University, 007 Agriculture Hall, Stillwater, OK 74078, USA
[b] Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, PR China
[c] Department of Biosystems and Agriculture Engineering, Oklahoma State University, 117 Agriculture Hall, Stillwater, OK, 74078, USA
[d] Biological Systems Engineering Department, University of Nebraska-Lincoln, 245 L.W. Chase Hall, Lincoln, NE 68583, USA
[e] U.S. Geological Survey, Oklahoma Cooperative Fish and Wildlife Research Unit, 007 Agriculture Hall, Oklahoma State University, Stillwater, OK 74078, USA

ARTICLE INFO

*Article history:*
Received 7 January 2016
Received in revised form 27 August 2016
Accepted 15 September 2016

*Keywords:*
Natural flow regime
High flow pulses
Pelagophils
MaxEnt
Temporal transferability

ABSTRACT

Truncated distributions of pelagophilic fishes have been observed across the Great Plains of North America, with water use and landscape fragmentation implicated as contributing factors. Developing conservation strategies for these species is hindered by the existence of multiple competing flow regime hypotheses related to species persistence. Our primary study objective was to compare the predicted distributions of one pelagophil, the Arkansas River Shiner *Notropis girardi*, constructed using different flow regime metrics. Further, we investigated different approaches for improving temporal transferability of the species distribution model (SDM). We compared four hypotheses: mean annual flow (a baseline), the 75th percentile of daily flow, the number of zero-flow days, and the number of days above 55th percentile flows, to examine the relative importance of flows during the spawning period. Building on an earlier SDM, we added covariates that quantified wells in each catchment, point source discharges, and non-native species presence to a structured variable framework. We assessed the effects on model transferability and fit by reducing multicollinearity using Spearman's rank correlations, variance inflation factors, and principal component analysis, as well as altering the regularization coefficient (β) within MaxEnt. The 75th percentile of daily flow was the most important flow metric related to structuring the species distribution. The number of wells and point source discharges were also highly ranked. At the default level of β, model transferability was improved using all methods to reduce collinearity; however, at higher levels of β, the correlation method performed best. Using β = 5 provided the best model transferability, while retaining the majority of variables that contributed 95% to the model. This study provides a workflow for improving model transferability and also presents water-management options that may be considered to improve the conservation status of pelagophils.

Published by Elsevier B.V.

## 1. Introduction

The distribution of aquatic organisms has been vastly affected by human alteration of the landscape (Dudgeon et al., 2006; Vörösmarty et al., 2010; Perkin and Gido, 2011). Coarse-scale changes affect aquatic biota through numerous pathways including altered hydrology, water quality and temperature, sediment dynamics, channel geomorphology, and habitat complexity. The multifaceted relationships between these pathways (Brown et al., 2005; Hughes et al., 2006) are sometimes difficult to relate to distributional changes due to data availability or the complexity of the interactions. However, land-use changes such as urbanization and agricultural expansion have received substantial attention (Marshall et al., 2008; Utz et al., 2010; Yates and Bailey, 2010), partly facilitated by the development of readily-available geospatial data. Further, the availability of such data has allowed the examination of

* Corresponding author.
*E-mail address:* shannon.brewer@okstate.edu (S.K. Brewer).
[1] Current address: Cardiff School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff, Wales, CF10 3AX, UK.

http://dx.doi.org/10.1016/j.ecolmodel.2016.09.016
0304-3800/Published by Elsevier B.V.

interactions between structuring variables (e.g., geology, soils) and land-use alteration (e.g., urbanization) on species distributional changes (Brewer and Rabeni, 2011).

Improving our ability to make predictions about the drivers of distributional shifts is underpinned by the development of more meaningful metrics of environmental change that can benefit either the conservation or management of species. Some landscape features have been difficult to quantify and are often lacking from distributional analyses. For example, irrigation has caused significant declines in groundwater levels and baseflow (Robson and Banta, 1995) and leads to stream habitat fragmentation and loss (Falke et al., 2011), but can be difficult to quantify despite its ecological importance. Simply-derived flow metrics (e.g., mean annual flow) are often included in analyses, but are challenging to translate into meaningful conservation and management strategies (e.g., how and when to allocate water, Worthington et al., 2014a). Finally, biotic interactions are often lacking from distributional analyses and represent a significant area of criticism (Cassini, 2011). Despite these challenges, species distribution models (SDM) remain a very powerful tool that can be used to illustrate distributional changes both spatially and temporally, and are particularly useful for the conservation and management of threatened and endangered species.

Pelagophils represent a group of imperilled fishes (Jelks et al., 2008) where SDMs have been developed to assess temporal distributional changes (e.g. Labay et al., 2011). This reproductive guild represents approximately 20 diminutive minnows that release semi-buoyant eggs that absorb water following fertilization and then require at least minimum velocities to remain in suspension (Mueller et al., 2013) and a substantial length of river for the ichthyoplankton (i.e., eggs and larvae) to drift during ontogeny (Williams and Bonner, 2006; Hoagstrom et al., 2011; Perkin and Gido, 2011). Several environmental factors structure their historic and current distributions, but those that primarily relate to human-induced landscape changes are drift distance and mean annual flow (Worthington et al., 2014a). Drift distance, representing landscape fragmentation, can be caused by a variety of human changes but was limited in Worthington et al. (2014a) to anthropogenic barriers. The importance of mean annual flow is supported by others (Bonner and Wilde, 2000; Hughes, 2005), but creates a problem for managers because it is not obvious what portion of the flow regime should be manipulated with the expectation of improving conditions for these species.

Several aspects of the flow regime can be predicted to be important to pelagophils based on the current knowledge of their life history. Dewatering, combined with other forms of fragmentation (e.g., large reservoirs), act synergistically to exacerbate the decline of pelagophils (Perkin et al., 2015a). These minnows are able to rapidly recolonize habitats following rewetting but repeated drying has been related to extinctions (Perkin et al., 2015b). Pelagophils are also thought to migrate long distances during the spawning season (e.g., Pecos Bluntnose Shiner *Notropis pecosensis*, Chase et al., 2015), suggesting stream connectivity to be important during this period. Further, spawning by pelagophils appears to coincide with increasing discharge (Bestgen et al., 1989; Robertson, 1997; Propst, 1999; Dudley, 2004), although some individuals also spawn regardless of environmental conditions (e.g., smalleye shiner *Notropis buccula*, Durham and Wilde, 2008; plains minnow *Hybognathus placitus*, Urbanczyk, 2012; sharpnose shiner *Notropis oxyrhynchus*, Durham and Wilde, 2014). Therefore, it is likely that several aspects of the flow regime are important to these species and determining which metrics relate to overall population persistence would be beneficial. Incorporating ecological-relevant environmental metrics in SDMs may improve our ability to be confident in model results, predict future distribution changes, conserve remaining populations, and highlight restoration opportunities.

Improving our ability to transfer model results to different spatial or temporal periods is a major challenge for SDMs (Wenger and Olden, 2012) and represents a significant area of research. SDMs are commonly used to make predictions about species distributions in different spatial or temporal locations. For example, SDMs are used to examine the spread of non-native species (Sorte et al., 2013) and examine the effects of climate change or other altered weather patterns on species distributions (e.g., Dyer et al., 2013). Whereas the importance of model validation for real-time models is emphasized (Olden et al., 2002), it is not possible to completely assess the accuracy of predictions that are made in locations that have not yet been colonized or in alternative time periods (Rastetter, 1996), but we can improve our predictions. Backcasting and forecasting model results to different time periods (Worthington et al., 2014a) or over range shifts (Araújo et al., 2005) suggest our ability to transfer model results may be hindered by several factors. Factors that may prevent adequate model transfer may be biological (e.g., species mobility, Pöyry et al., 2008; metapopulation responses, Sinclair et al., 2010; changes between species and the environment over time, Worthington et al., 2014a), or related to statistical considerations and modelling techniques (e.g., interactions between predictor variables, Guisan et al., 1999; Thuiller et al., 2003; model overfitting, Guisan et al., 1999; Thuiller, 2004). Model transferability may be improved by selecting appropriate techniques (e.g., consensus models, Araújo et al., 2005). Evaluating and improving model transferability will greatly improve the confidence in SDMs that are used to make predictions in different spatial or temporal periods.

The objective of this study was to test competing hypotheses related to the importance of different ecologically-relevant flow metrics to the persistence of Arkansas River Shiner (ARS, *Notropis girardi*), a federally-threatened pelagophil, while also assessing different approaches for improving model fit and temporal transferability. Building on the original models described in Worthington et al. (2014a), we calculated additional functionally-relevant predictors including groundwater withdrawals, point source discharge, and the distribution of a perceived non-native competitor. Our approach was to develop competing hypotheses to evaluate the relative importance of different aspects of the flow regime. We then used three common methods for reducing the number of multicollinear variables in the model to improve overall model transferability. Finally, to simplify the models, we evaluated the use of different regularization coefficients to produce less complex models.

## 2. Methods

### 2.1. Study area

The Arkansas River basin extends from the Southwestern Tablelands ecoregion of New Mexico and Colorado to the eastern portion of the Arkansas Valley in Arkansas, with the majority of our study area located within the Great Plains Ecoregion (Fig. 1). The Great Plains occupies a major west-east climate gradient with precipitation ranging from 40 to 145 cm $yr^{-1}$ (O'Neill et al., 1997; Woods et al., 2005). Land use in the basin is dominated by cropland or pasture with transition to an oak-savanna mix in the east. The major rivers of the basin (Arkansas River, Canadian River, Cimarron River, and North Canadian River) are generally wide, shallow, and dominated by sand or mud substrates (Matthews et al., 2005). Historically, the rivers of the Great Plains were characterized by large fluctuations in discharge, temperature and turbidity (Matthews, 1988; Dodds et al., 2004). This highly-variable landscape resulted in aquatic organisms specifically adapted to tolerate these chang-