



Temporal transferability of stream fish distribution models: can uncalibrated SDMs predict distribution shifts over time?

Jian Huang, Emmanuel A. Frimpong* and Donald J. Orth

Department of Fish and Wildlife
Conservation, Virginia Polytechnic Institute
and State University, Blacksburg, VA, USA

ABSTRACT

Aim We aim to assess the temporary transferability of species distribution models (SDMs) for stream fish species in terms of discrimination power and calibration.

Location New River basin, eastern United States.

Methods In this study, we used Lasso-regularized logistic regression (LLR), boosted regression trees (BRT), MaxEnt, and ensemble model (ENS) to evaluate the habitat suitability of 16 fish species with different rarity and temperature preference based on historical species occurrences obtained during 1950–1990. These SDMs were used to make probabilistic predictions of species presence in the independent datasets sampled during 2012–2014. We evaluated the temporal transferability of these SDMs in terms of discrimination power and calibration with the temporarily independent datasets.

Results The area under the receiver-operating-characteristic curve (AUC) was over 0.6 for 13 (81%) of the species in the evaluations of ENS models with the independent datasets. Cool-water species and species with small local population size traits tended to have good temporal transferability. With observed species prevalence as the discrimination cut-off, LLR had the highest overall accuracy for 13 of the 16 species and highest specificity for 10 species, whereas MaxEnt had the highest sensitivity for 14 species. Biases, under- or over-fitting problems were common in the temporal model transfers, among the modelling approaches used here.

Main conclusions SDMs developed with historical data generally had moderate to good discrimination power but they tended to systematically underestimate current probability of species presence. To predict species distribution shifts over time, SDMs should be well-calibrated with high discrimination power. We suggest reclassifying predicted probability of species occurrence to ordinal ranks to deal with (under- and over-estimation) bias, and fine-tuning variable selection with regularization or cross validation to remedy under- and over-fitting.

Keywords

climate change, discrimination and calibration, discrimination power, fish, independent testing, model calibration, occupancy model, species distribution models, temporal transferability.

*Correspondence: Emmanuel A. Frimpong,
Department of Fish and Wildlife
Conservation, Virginia Polytechnic Institute
and State University, Blacksburg, VA 24061,
USA.
E-mail:frimp@vt.edu

INTRODUCTION

Species distribution models (SDMs) are commonly used quantitative approaches to predict spatial distribution shifts

over time (Steen *et al.*, 2010; Brown *et al.*, 2011; Bouska *et al.*, 2015). Climate change has been recognized as one of the most influential disturbances to alter species distributions (Thuiller, 2007; Sharma & Jackson, 2008; Brown *et al.*,

2011). Global warming would profoundly affect ectothermic taxa whose growth, recruitment, and dispersal are influenced by temperature (Buisson *et al.*, 2008). Successful predictive SDMs could provide critical information on delineating suitable habitats for species at risk (Elith *et al.*, 2006; Sindt *et al.*, 2012) and adjusting stocking strategies for game species (Flebbe *et al.*, 2006). However, one step researchers usually neglect before predicting future spatial distribution under climate and land cover change is to assess the temporal transferability of SDMs developed using historical or current observations. Assessing transferability is the process to evaluate how accurately and reliably SDMs developed based on training datasets predict species distribution in a different region or time frame (Randin *et al.*, 2006). We assessed temporal transferability in this study by transferring SDMs among temporally independent datasets.

Transferring SDMs over time is widely recognized as risky because these applications inevitably involve extrapolation into new environmental space (Elith & Leathwick, 2009; Schibalski *et al.*, 2014). The predictions of species occurrence could be unreliable if the environmental gradients and species–habitat relationships vary unpredictably among the examined time frames (Strauss & Biedermann, 2007; Elith & Leathwick, 2009). In our previous study (Huang & Frimpong, 2016, in press), we found that spatial transfer of models can be constrained by mismatches in the range and location of the environmental gradients in the training region and prediction region. These same constraints might also hinder temporal transference of SDMs. Additionally, transferability depends on model type (Meynard & Quinn, 2007; Wenger & Olden, 2012), the quality of training data (Strauss & Biedermann, 2007; Wang & Jackson, 2014), and species traits (Bulluck *et al.*, 2006; Randin *et al.*, 2006).

Most previous studies assessed the temporal or spatial transferability of SDMs by means of discrimination power, i.e., a model's ability to discriminate occupied and unoccupied sites (Pearce & Ferrier, 2000). Sensitivity (true positive or presence rate), specificity (true negative or absence rate), accuracy (correct classification rate) and AUC (area under the receiver-operating-characteristic curve) are commonly used measures of discrimination power. Calibration statistics (also known as model estimation in Franklin & Miller, 2009) that describe the bias and spread of predicted presence probabilities should also be included in the assessment of temporal model transferability (Strauss & Biedermann, 2007). In this study, bias is defined as a measure of systematic under- or over-estimation of predicted probabilities and spread is used to describe the distribution pattern of predicted probability, e.g., too extreme or clumped, or wrong trend (Pearce & Ferrier, 2000). In a well-calibrated model, if the predicted probabilities of species presence was between 0.5 and 0.7 for 100 sites, then approximately 60 sites would actually be occupied by the species. Even a model with large discrimination power could systematically under- or over-estimate probabilities of species presence.

Depending on the source of testing data, evaluations of model discrimination power and calibration have several forms with varied efficacies. In cross validation, the testing datasets are iteratively sampled from the original data and predicted by the models developed based on training datasets. Performance measures then can be obtained by relating the observed and predicted values of the testing datasets. In external evaluations, the testing datasets are obtained independent of the training data, for example, from different sampling campaign, time frames or regions. Most studies used independent testing data from single-visit sampling (Randin *et al.*, 2006; Tuanmu *et al.*, 2011; Wang & Jackson, 2014). Alternatively, and perhaps more robustly, independent testing data from occupancy-based sampling can be used in the assessment of model transferability. In occupancy-based sampling, temporally replicated observations of species presence/absence at each site are obtained, which allows separation of the probability that a species is present (occupancy) from the conditional probability of detecting the species (MacKenzie *et al.*, 2002). However, occupancy-based design requires more investment in time and labour. Thus, it is worth comparing model evaluation based on occupancy-based and single-visit independent datasets because if these two approaches yield equivalent results then we can have more confidence in the single-visit evaluations which are less costly in time and labour.

In this study, we develop SDMs for 16 fish species in the New River using historical data collected during 1950–1990, and evaluate the temporal transferability of these SDMs by means of discrimination power (AUC, sensitivity, specificity and accuracy) and calibration statistics (bias and spread of predicted probabilities of presence) in the independent field survey data collected during 2012–2014. The effects of modelling approach, species rarity and temperature preference on AUC in the model transfers are evaluated. Additionally, we compare calibration statistics between single-visit independent samples and occupancy-based independent samples.

METHODS

Study system and species

The New River (Fig. 1) originates in North Carolina, heads north and drains through southwestern Virginia and West Virginia, USA, for a total drainage area of 21,700 km² (Jenkins & Burkhead, 1994). The New River is heterogeneous geologically as it spans four types of major eastern U.S. physiographic provinces: Blue Ridge, Piedmont, Valley-and-Ridge Appalachia and Appalachian Plateaus (Fenneman & Johnson, 1946). At the Kanawha Falls, the New River and Gauley River merge to form the Kanawha River, which then flows into the Ohio River and Mississippi River systems. Owing to isolation and habitat heterogeneity, the New River has maintained a particularly high proportion of endemic fish species.

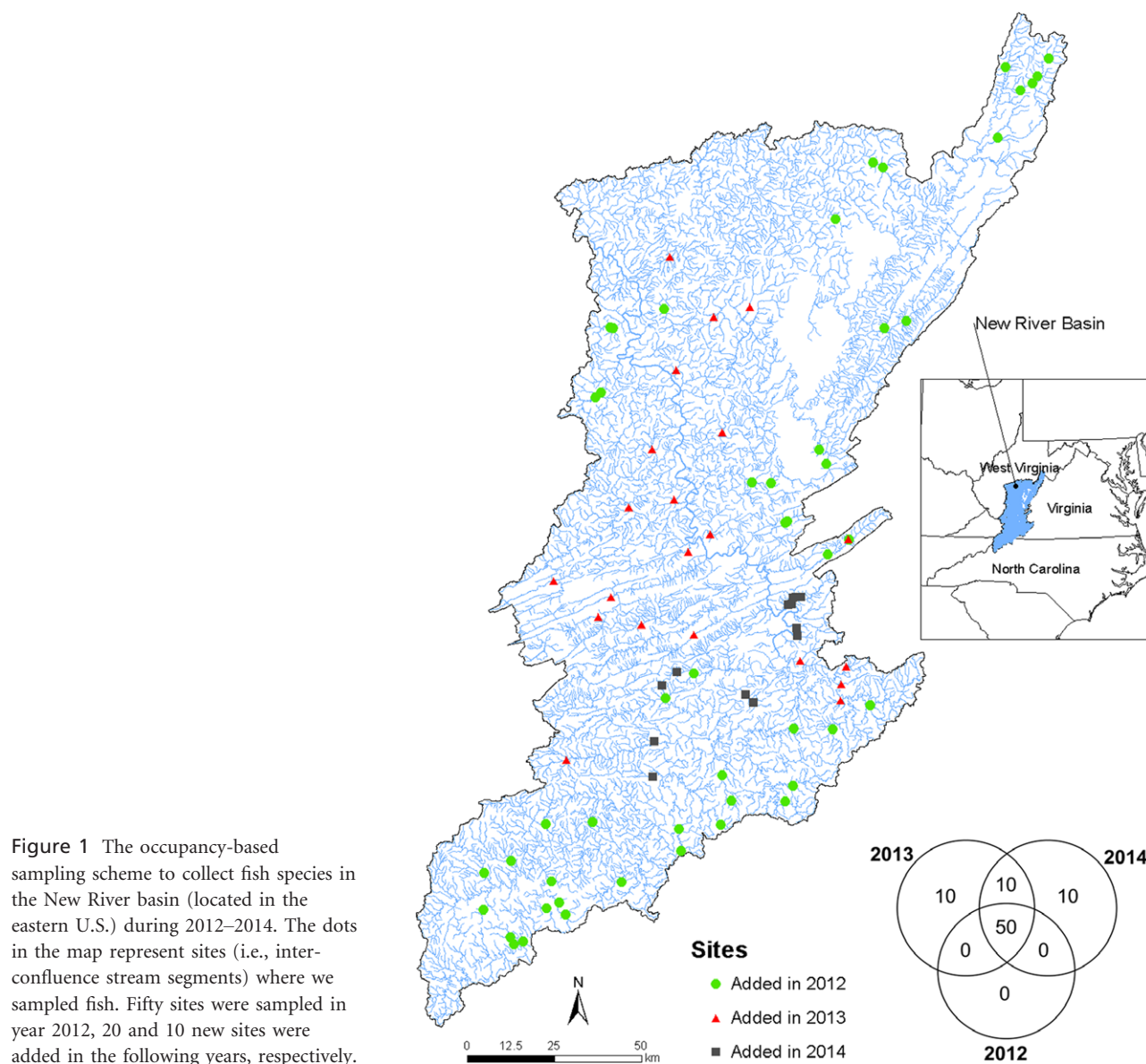


Figure 1 The occupancy-based sampling scheme to collect fish species in the New River basin (located in the eastern U.S.) during 2012–2014. The dots in the map represent sites (i.e., inter-confluence stream segments) where we sampled fish. Fifty sites were sampled in year 2012, 20 and 10 new sites were added in the following years, respectively.

We modelled the spatial distribution of 16 fish species (Table 1) found in the wadeable streams in the New River Basin. These 16 species are diverse in traits (e.g., body size, habitat preference) and rarity (Pritt & Frimpong, 2010); six of them are New River endemic species (Table 1). *Micropterus salmoides*, *Micropterus dolomieu* and *Ambloplites rupestris* are warm-water species, *Oncorhynchus mykiss*, *Salmo trutta* and *Salvelinus fontinalis* are cold-water species, while the other 10 species are cool-water species (Cherry *et al.*, 1975; Shingleton *et al.*, 1981; Beitinger *et al.*, 2000; Lyons *et al.*, 2009). We determined the thermal classification of three endemic species, *Phenacobius teretulus*, *Etheostoma kanawhae* and *Percina gymnocephala*, based on their habitat preferences (Menhinick, 1991; Jenkins & Burkhead, 1994; Stauffer *et al.*, 1995) and our field observations since their temperature tolerances have not been well-documented in the literature. The 16 species examined in our study, except for warm-water species, were hypothesized to move

northward or into higher elevation streams given only water warming, and this may be compounded by disruptions in landscape connectivity (e.g., waterfalls and dams) and future landscape transformations (e.g., urbanization, deforestation).

Developing SDMs

Historical species occurrence data (2269 presence records) used in the models were extracted from the *IchthyMaps*, a metacommunity-based database that contains fish presence records sampled by seining and electrofishing primarily during 1950–1990 in the United States (Frimpong *et al.* 2015, <https://www.sciencebase.gov/catalog/item/5446a5a1e4b0f888a81b816d>). These species presences were spatially joined to 1035 National Hydrography Dataset (NHD) inter-confluence stream segments at 1:100,000 scale. For a game species, those sampled segments without any presence record of the focal species were designated as absence; for a non-game species,

Table 1 The list of 16 fish species examined in this study. ‘Rare-Extent’, ‘Rare-Habitat’ and ‘Rare-Local’ represented rare in terms range size, habitat breadth and local population size, respectively (Rabinowitz, 1981; Pritt & Frimpong, 2010). Family number of these species were retrieved from Nelson (2006).

Species	Common name	Rare-Extent	Rare-Habitat	Rare-Local	Temperature preference	Family number
<i>Notropis scabriceps</i> *	New River shiner	1	0	1	Cool	102
<i>Phenacobius teretulus</i> *	Kanawha minnow	1	1	1	Cool	102
<i>Etheostoma kanawhae</i> *	Kanawha darter	1	0	1	Cool	350
<i>Etheostoma osburni</i> *	Candy darter	1	0	1	Cool	350
<i>Percina gymnocephala</i> *	Appalachia darter	1	1	1	Cool	350
<i>Nocomis platyrhynchus</i> *	Bigmouth chub	1	0	0	Cool	102
<i>Cottus kanawhae</i>	Kanawha sculpin	0	0	0	Cool	320
<i>Rhinichthys atratulus</i>	Blacknose dace	0	0	0	Cool	102
<i>Chrosomus oreas</i>	Mountain redbelly dace	1	0	0	Cool	102
<i>Noturus insignis</i>	Margined madtom	0	0	0	Cool	143
<i>Ambloplites rupestris</i>	Rock bass	0	0	0	Warm	349
<i>Micropterus dolomieu</i>	Smallmouth bass	0	0	0	Warm	349
<i>Micropterus salmoides</i>	Largemouth bass	0	0	0	Warm	349
<i>Salvelinus fontinalis</i>	Brook trout	0	0	0	Cold	175
<i>Salmo trutta</i>	Brown trout	0	0	0	Cold	175
<i>Oncorhynchus mykiss</i>	Rainbow trout	0	0	0	Cold	175

*New River endemic species.

those sampled segments without any presence record of the focal species but with at least one presence of another non-game species were designated as absences for that species (Huang & Frimpong, 2015).

Fourteen habitat variables constantly used in studies of spatial distribution of stream fish (e.g., Steen *et al.*, 2010; Wenger & Olden, 2012; Bouska *et al.*, 2015) are selected as predictors in our SDMs (Table 2). The climate data (e.g., temperature, precipitation) during 1950–2014 were retrieved from the PRISM Climate Group (2004). The land cover data

in the inter-confluence catchment of each stream segment in the 1980s (Price *et al.*, 2006) and 2006 (Fry *et al.*, 2011), respectively, corresponding to historical and current fish occurrence data, were obtained from the United States Geological Survey (USGS) Land Cover Institute. The habitat condition index (HCI) was retrieved from the National Fish Habitat Partnership (NFHP) databases (National Fish Habitat Board, 2012), and other variables were retrieved from NHDplusV1 and NHDplusV2 (EPA, USGS and Horizon Systems Corporations, 2010, 2012). The values of habitat vari-

Table 2 The list of habitat variables used to develop species distribution models for 16 stream fish species in the New River. Only one measure was obtained for ending variables (i.e., ELE, SLP, SO, SINU) and other variables (e.g., BFI, RL and HCI).

Variable	Source	Descriptions	Mean \pm SD (year 1961–1980)	Mean \pm SD (year 1995–2014)
BFI	NHDplusV1	The ratio of base flow (i.e., the component of stream flow attributed to ground–water discharge) to total flow (%) in the watershed.	44.49 \pm 9.03	
ELE	NHDplusV2	Mean elevation in the watershed (m)	756.39 \pm 168.74	
SLP	NHDplusV2	Mean slope in the watershed (°)	12.39 \pm 4.78	
SO	NHDplusV2	(Strahler) Stream order	1.90 \pm 1.25	
SINU	NHDplusV2	Sinuosity (i.e., reach length divided by length of straight line between two stream nodes)	1.16 \pm 0.27	
HCI	NFHP	An index of cumulative disturbance in the watershed based on 15 disturbance variables (Esselman <i>et al.</i> , 2011). A HCI of 0 indicates pristine condition.	3.20 \pm 0.62	
RL	NFHP	Total road length in the watershed (km)	3576.18 \pm 7150.27	
L_UB	Price <i>et al.</i> (2006)	Percentage of agriculture in the watershed	2.28 \pm 9.40	2.94 \pm 9.58
L_AG	Price <i>et al.</i> (2006)	Percentage of forest in the watershed	25.67 \pm 30.90	26.43 \pm 29.49
L_FR	Price <i>et al.</i> (2006)	Percentage of urban in the watershed	69.07 \pm 32.60	70.08 \pm 25.14
PPT	PRISM	20-year mean annual precipitation (mm)	1101.02 \pm 152.94	1130.27 \pm 156.92
TMIN	PRISM	20-year mean annual minimum temperature (°C)	−6.77 \pm 1.15	−5.24 \pm 1.13
TMAX	PRISM	20-year mean annual maximum temperature (°C)	27.13 \pm 1.07	26.19 \pm 1.18
TMEAN	PRISM	20-year mean annual temperature (°C)	10.45 \pm 0.849	10.93 \pm 0.99

ables in year 1961–1980 were used with historical fish occurrence data to develop SDMs, whereas the habitat data in year 1995–2014 were used with current fish occurrence data in model evaluations. Environmental conditions were assumed constant during 1961–1980. This assumption was likely valid because (1) enduring variables such as elevation and stream order usually remain constant over decades; (2) land use and land cover data are not available annually but there is evidence that the landscape in the New River basin had not changed drastically during that time period (e.g., Drummond & Loveland, 2010); (3) our SDMs in the pilot analysis showed good performance in the model training given this assumption. Additionally, we took the 20-year mean for the annual minimum temperature, maximum temperature, mean temperature and annual total precipitation to account for annual fluctuations of these variables.

Spatial autocorrelation in the habitat variables was filtered by the principal coordinates of neighbour matrices (PCNM; Borcard & Legendre, 2002) with package *VEGAN* (Oksanen *et al.*, 2015) and *SPDEP* (Bivand & Piras, 2015) in R (R Core Team, 2014). Spatialized habitat variables (i.e., predicted habitat variables by PCNM spatial eigenvectors in a multivariate regression model) were used as predictors in the LLR, BRT and MaxEnt.

Based on the historical fish occurrence data and habitat variables, we developed SDMs for each of the 16 species using Lasso-regularized logistic regression, boosted regression trees (BRT; Friedman, 2001), and maximum entropy presence-only model (MaxEnt; Phillips *et al.*, 2006). An ensemble model was formed for each species by averaging the predicted presence probabilities of the other three models (Marmion *et al.*, 2009; Bouska *et al.*, 2015). The BRT and MaxEnt are widely used in studies of SDMs (e.g., Elith *et al.*, 2006). The Lasso (least absolute shrinkage and selection operator) is used to regularize logistic regressions through adding a constraint of absolute sum of coefficients to the likelihood optimization (Friedman *et al.*, 2010). The penalty parameter can be tuned in the cross validation to balance model accuracy and simplicity (Tibshirani, 1996). We implemented LLR with the package *GLMNET* (Friedman *et al.*, 2010), and BRT and MaxEnt models with the package *DISMO* (Hijmans *et al.*, 2013). Final set of predictor variables were determined by maximizing AUC in the fivefold cross-validation of BRT, LLR and MaxEnt.

Collecting temporally independent datasets

We conducted field surveys to collect fish occurrence data during 2012 and 2014 to evaluate temporal transferability of SDMs. A total of 80 Wadeable stream segments (sites) were sampled in an occupancy-based survey. Fifty, seventy and seventy sites were sampled in 2012, 2013 and 2014, respectively (see Fig. 1 for details about sample scheme). These 80 sites were evenly distributed across the New River basin, and approximately 1/3 of them were located in each of the three major physiographic provinces. These segments represented a

wide range of stream size, geomorphology, landscape characteristics and disturbance. Each site, with length of 120–300 m, contained at least four riffle-run-pool sequences to capture fish species with different microhabitat preference. All sites were sampled with single-pass electrofishing during May to August. The sampling crew included four to five members equipped with two backpack electrofishers (Smith-Root LR-24) and four nets (14 × 17 × 8 inch trapezoid, 0.25 inch Mesh).

From these field surveys, we obtained occupancy-based independent samples (OBS) because 60 out of 80 sites were sampled at least two times during 2012–2014. We randomly subsampled the OBS to form single-visit independent samples (SVS) that contained only one observation of species occurrence at each of 80 sites. We used SVS to evaluate discrimination power and calibration, whereas OBS was used only for calibration in model transfers.

Evaluating temporal transferability

Discrimination power

The observed prevalence (proportion of presences in all observations) for each species was used as the threshold probability to calculate sensitivity, specificity and accuracy. Sensitivity and specificity tend to be equal when the threshold probability of species presence is equal to the observed prevalence (Franklin & Miller, 2009). The AUC was calculated in the model training and external evaluations with the OBS and SVS datasets. The AUC between 0.7 and 0.9 indicates moderate discrimination power and between 0.9 and 1 is considered excellent in the model training (Swets, 1988). However, a lower threshold, for example AUC of 0.6–0.7 (Randin *et al.*, 2006; Strauss & Biedermann, 2007), is often used to distinguish transferable from non-transferable SDMs.

Model calibration

We measured bias and spread using Cox (1958)'s approach for model calibration (Fig. 2). The observed probabilities of presence (π_i) in the independent testing data were regressed against the logit of the predicted probability of presence (\hat{p}_i) in a calibration model (Fig. S1):

$$\text{Logit}(\pi_i) = a + b * \text{Logit}(\hat{p}_i) \quad (1)$$

If the probability of species presence at site i (φ_i) in the occupancy model is assumed to be true and be a substitute of π_i , then the calibration model takes the linear form:

$$\varphi_i = a + b * \hat{p}_i \quad (2)$$

We tested whether intercept a and slope b of the calibration line significantly deviated from 0 and 1 respectively using three likelihood ratio tests (Miller *et al.*, 1991). A significance level of 0.01 was used to control the type I error in the multiple hypothesis tests (Fig. 2). If bias in \hat{p}_i occurs, a

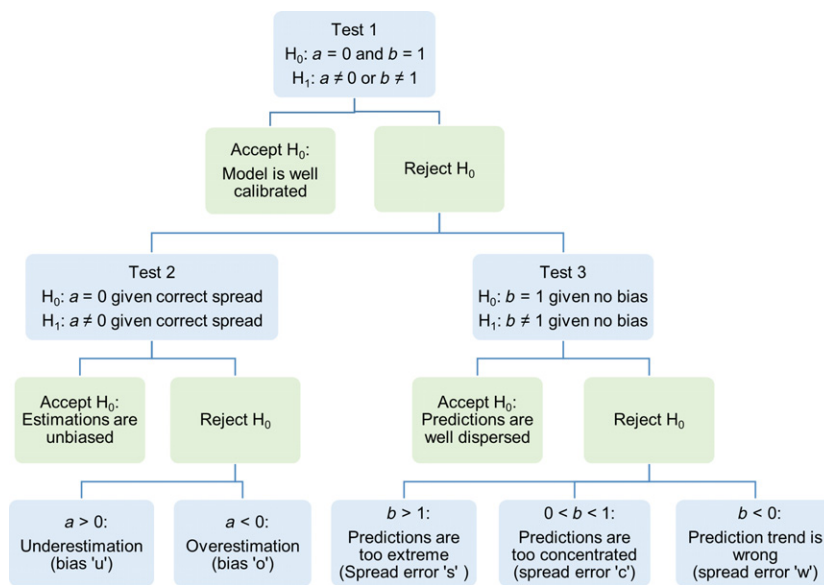


Figure 2 The species distributions models for 16 New River species were calibrated in terms of bias and spread (Miller *et al.*, 1991; Pearce & Ferrier, 2000). Test 1 is the likelihood ratio test to examine whether the intercept (a) and slope (b) of the calibration line significantly deviate from 0 and 1, respectively. Test 2 and Test 3 test the intercept and slope separately if the null hypothesis of good calibration is rejected in the Test 1. If the null hypothesis is rejected in the Test 2 or Test 3, fitted values of a or b in the alternative hypothesis would be evaluated to make a decision ('u', 'o', 's', 'c', and 'w').

would significantly deviate from 0. In the case of over- or under-fitting, b would significantly deviate from 1.

Evaluations with the single-visit independent samples (SVS)

The SDMs of 16 fish species were independently tested by SVS. Temporal transferability was evaluated with three steps: (1) predicting the presence probabilities of 16 species at 80 sites using SDMs built based on historical data, (2) calculating discrimination power (AUC, sensitivity, specificity and accuracy) based on predicted and observed fish occurrence, and (3) measuring bias and spread of predicted presence probabilities \hat{p}_i using calibration model 1.

We then used an ANCOVA (i.e., analysis of covariance) in R to evaluate the effects of biotic factors: (1) species rarity, with three dimensions: range size, habitat breadth and local population size, and (2) temperature tolerance of species (with three levels: cold, cool and warm) on temporal transferability in terms of AUC. Family numbers of 16 species (Nelson, 2006) was the covariate in the ANCOVA. AUC was log transformed to satisfy the normality and constant variance assumptions. Modelling approaches were compared in terms of AUC with species as a block factor in the Friedman Test (Friedman, 1937).

Evaluations with the occupancy-based independent samples (OBS)

We further evaluated the temporal transferability with OBS in a close-population occupancy model (MacKenzie *et al.*, 2002). We used the habitat variables (Table 2) as covariates to estimate site-specific probability of presence, and used water temperature ($^{\circ}\text{C}$), discharge ($\text{m}^3 \text{s}^{-1}$), and sampling effort (s) as sampling covariates to estimate site- and year-specific probability of detection in the R package UNMARKED (Fiske & Chandler, 2011). Water temperature and sampling effort were recorded after sampling at each site. The

discharge at a site was estimated as (real-time discharge at US Geological Survey (USGS) gage) \times (watershed area of site)/(watershed area of USGS gage). The number of candidate occupancy models for each species was 61,208, i.e., $\left[\sum_{k=0}^5 \binom{14}{k} \right]$, where $\binom{14}{k}$ means the combination of selecting k of 14 habitat covariates. We set the maximum of k to five to limit the number and complexity of candidate occupancy models. For each species, we compared and ranked 61,208 models using AIC (Akaike Information Criterion). The site-specific presence probabilities in the five occupancy models were averaged with Akaike weights for each species. The bias and spread of \hat{p}_i in the model transfers were evaluated using the calibration model (equation 2).

RESULTS

Evaluations with the training datasets

The spatial distribution of 16 fish species were well-described (with moderate to high AUC) in the model development using historical training data. AUC was higher than 0.75 for all species in the training of ENS, BRT and MaxEnt models (Fig. 3). Annual average (air) temperature in the New River basin has increased by 0.48°C (i.e., 10.45 ± 0.85 to 10.93 ± 0.99) during the past four decades (Table 2). Temperature measures were key predictors and were negatively related to the occurrence of 10 species in the partial dependence plots (Table S1). Other important habitat variables included BFI (base flow index), stream order and elevation.

Evaluations with single-visit independent samples (SVS)

The Friedman Test showed that the AUC of BRT was significantly higher than MaxEnt in the evaluation with SVS,

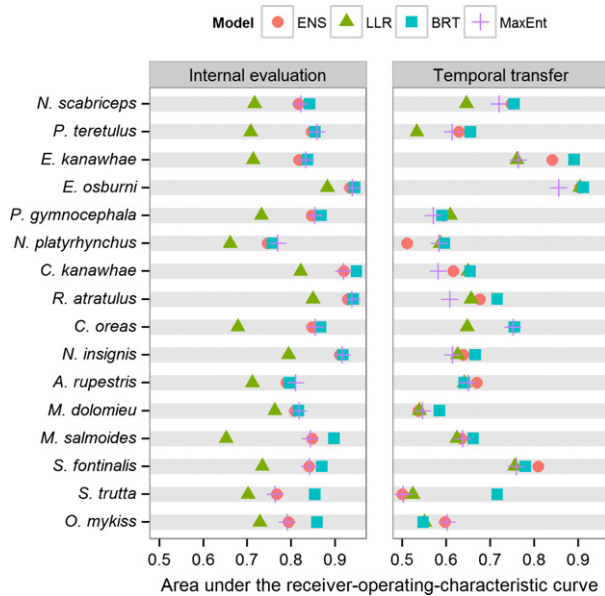


Figure 3 AUC (area under the Receiver-Operating-Characteristic curve) in the model training and temporal transfers to single-visit independent samples. The performance of ensemble model (ENS), Lasso-regularized logistic regression (LLR), boosted regression trees (BRT) and MaxEnt were evaluated.

with P -value = 0.002 in the *post hoc* comparison. The LLR had highest accuracy and specificity while the MaxEnt had the highest sensitivity for most species (Fig. S2). The sum of sensitivity and specificity of BRT was the highest of the three model types for 10 of the 16 species. The ensemble of SDMs did not improve discrimination power in the temporal transfers (Fig. 3). For brevity, the default modelling approach was BRT if not specified in the following paragraphs.

The temporal transferability of SDMs varied among species (Figs 3 & S2) in the evaluation with SVS. The ANCOVA (Table 3) indicated that temporal transferability was significantly affected by temperature preference and rarity (in terms of local population size) of fish species (P -value = 0.048 and 0.001, respectively). *Post hoc* multiple comparison showed cool-water species tended to have higher temporal transferability than cold-water and warm-water species (Fig. S5). AUC in the model transfers was high for *E. osburni* and *C. kanawhae*, but was low for two trout species, *S. trutta* and *O. mykiss*. According to the partial dependence plots (Fig. S3) and histogram (Fig. S4), we found that *E. osburni* showed negative response to the increase of temperature and that pattern was consistent between 1961–1980 data and 1995–2014 data.

In the evaluation with SVS, the number of well-calibrated ENS, LLR, BRT and MaxEnt models were 7, 6, 5 and 3, respectively (Table 4). Underestimating \hat{p}_i was more frequent than overestimating. Generally, ENS and BRT were well calibrated for the six New River endemic species (Table 4).

Table 3 A table summarizing the analysis of covariance (ANCOVA) that evaluated the sources of effects on the area under the receiver-operating-characteristic curve in the temporal transfers of species distribution models. ‘Rare-Extent’, ‘Rare-Habitat’ and ‘Rare-Local’ are three dimensions of rarity, i.e., rare in terms range size, habitat breadth and local population size, respectively (Rabinowitz, 1981; Pritt & Frimpong, 2010). We blocked the effect of modelling approach to test effects of biotic factors (i.e., rarity and temperature preference) solely.

Source	Type	Degree of freedom	Mean square	F -value	P -value
Temperature preference	Factor	2	0.026	3.285	0.048
Rare-Extent	Factor	1	0.014	1.682	0.202
Rare-Habitat	Factor	1	0.031	3.881	0.056
Rare-Local	Factor	1	0.112	13.957	0.001
Family number	Covariate	1	0.002	0.233	0.632
Model	Block	2	0.012	1.522	0.231
Residuals	Residual	39	0.008		

Evaluations with occupancy-based independent samples (OBS)

According to the occupancy model, temperature (maximum and mean), stream order, BFI, % urban and % agriculture in the watershed were the important habitat covariates for the 16 fish species (Table S2). Discharge was the key factor that affected the detection of nine species (e.g., *N. scabriceps*, *E. kanawhae* and *E. osburni*). Cool- and cold-water species such as *S. fontinalis*, *N. scabriceps*, *P. teretulus*, *E. osburni* and *P. gymnocephala* favoured elevated mountain streams with low maximum July temperature, which was consistent with the patterns revealed in the partial dependence plot based on historical data (Table S1).

Calibration with OBS revealed that bias and spread error in \hat{p}_i were common in the SDMs (Table 4). Eleven of sixteen ENS models showed systematic underestimation of \hat{p}_i . The calibration of SDMs for *R. atratulus*, *C. oreas*, *N. insignis*, *A. rupestris*, *M. dolomieu* and *O. mykiss* were consistent between the SVS and OBS datasets.

DISCUSSION

In addition to a general pole-ward shift, cold-water species and those cool-water species (e.g., *N. scabriceps* and *E. osburni*) that showed negative relationships between temperature and presence are predicted to move upstream under global warming. Flebbe *et al.* (2006) also predicted that more than 50% of total trout habitat area in the southern Appalachians would be lost by the year 2100 based on two global climate circulation models, and only small refuges in elevated headwater streams would remain. We recommend evaluating temporal transferability of SDMs in the study of effects of climate changes on spatial distribution of species.

Generally, the discrimination power of SDMs for the New River fish species were moderate to good in the temporal

Table 4 Calibrating species distribution models (ENS, ensemble model; LLR, Lasso-regularized logistic regression; BRT, boosted regression model; MaxEnt, maximum entropy) for New River fish species in the single-visit independent sample (SVS) and occupancy-based independent sample (OBS). 'P.t' and 'P.i' are the observed species prevalence in the training data and single-visit validation sample, respectively. 'C.b' and 'C.s' are respectively the calibrations on bias (o, overestimation; u, underestimation) and spread (s, too dispersed; c, too clumped) of predicted presence probabilities. If a model is well-calibrated in terms of both bias and spread, then the cells of 'C.b' and 'C.s' are null, otherwise at least one cell for the species would not be null.

Species	Prevalence		Calibration with SVS dataset								Calibration with OBS dataset							
			ENS		LLR		BRT		MaxEnt		ENS		LLR		BRT		MaxEnt	
	P.t	P.i	C.b	C.s	C.b	C.s	C.b	C.s	C.b	C.s	C.b	C.s	C.b	C.s	C.b	C.s	C.b	C.s
<i>N. scabriceps</i>	0.21	0.14									u		u		u		u	
<i>P. teretulus</i>	0.15	0.20							o	s		s				s	o	s
<i>E. kanawhae</i>	0.19	0.25		c	o			c								c		c
<i>E. osburni</i>	0.12	0.13								s				c				
<i>P. gymnocephala</i>	0.14	0.22								c	u		u				u	
<i>N. platyrhynchus</i>	0.26	0.33							o	c	u	c	u		u			c
<i>C. kanawhae</i>	0.11	0.27		s	o	s		s	o			c		c	u	c		
<i>R. atratulus</i>	0.14	0.81	u		u		u		u	s	u	s	u	s	u		u	s
<i>C. oreas</i>	0.20	0.62	u		u		u		u		u		u		u		u	
<i>N. insignis</i>	0.11	0.32	u		u			c	u		u		u	c	u	s	u	
<i>A. rupestris</i>	0.19	0.66	u		u		u		u		u		u		u		u	
<i>M. dolomieu</i>	0.17	0.47		s	o	s	u			s	u	s		s	u	s		c
<i>M. salmoides</i>	0.03	0.11					o			c					o			
<i>S. fontinalis</i>	0.08	0.22			u	c	u		o	s	u		u	c				
<i>S. trutta</i>	0.04	0.34	u		u		u				u			c		c	u	
<i>O. mykiss</i>	0.06	0.33	u		u		u		u	s	u		u		u		u	s

model transfers. Temporal transferability was species-specific: the AUC and accuracy were high for cool-water endemic species with low local population, such as *N. scabriceps* and *E. osburni*. A weak negative effect of climate change can be a serious threat to these endemic species. The fish-habitat relationships for the endemic species were more accurately described, likely because the whole range of the environmental gradients for the species were evaluated, free of the problem of mismatches in the range and location of variables in the model transfers. Additionally, the endemic species in the New River basin show different physiographic preference: *N. scabriceps*, *P. teretulus* and *P. gymnocephala* are often common in the upper section of the New River Basin within the Blue Ridge province but usually rare or uncommon in the Valley and Ridge province; *E. kanawhae* is restricted to the Blue Ridge Province, while *E. osburni* is restricted to Valley and Ridge and Appalachian Plateau provinces (Jenkins & Burkhead, 1994). Being restricted by geographical extent, habitat and local population size, these endemic fish species would be strongly threatened by even a weak negative effect of climate change.

The AUC and accuracy were low for trout and bass species. For more wide-spread species (e.g., *M. salmoides*, and *M. dolomieu*), the New River covers a mere fraction of the range of these two species (Brewer & Orth, 2015), thus the fish-habitat relationships described in the SDMs might be unreliable (Jackson *et al.*, 2001). It has been widely recognized that distribution of generalist species are particularly

difficult to predict (Brotons *et al.*, 2004) and to transfer over time and space (Randin *et al.*, 2006; Strauss & Biedermann, 2007; Schibalski *et al.*, 2014). Additionally, the spatial distribution and abundance of the game species in this study, particularly the three trout species in the New River drainage, might be reflecting management interventions (e.g., stocking strategies, capture-return policy) of the fish and wildlife agencies in the three states (Virginia stocked trout management plan; <http://www.dgif.virginia.gov/fishing/trout/management-plan/stocked-trout-management-plan-draft.pdf>).

Bias and misspecified spread in the predicted probability of occurrence of the New River fish species were common in the external evaluation with both OBS and SVS datasets (Table 4). Without calibration or refinement of probabilistic predictions of presence, SDMs could not adequately predict distribution shifts over time. Pearce & Ferrier (2000) attributed the bias in the estimations to the inconsistent species prevalence between the training and testing data. We found that, if the prevalence in the training dataset was higher than independent testing samples, then the model transferred would likely suffer from overestimation bias; otherwise underestimation tended to occur in model transfers (Table 4). There might be some other reasons causing poor calibration when a model is applied to new data. The observed prevalence might not be a good estimator of the 'true' species prevalence. Another major explanation for limited transferability is the inherent difficulty in extrapolating SDMs to novel environments (Elith & Leathwick, 2009;

Schibalski *et al.*, 2014). After all, conventional SDMs are static and they may sensitively capture the changes in the predictors, but they may not explicitly account for population dynamics or dispersal (Guisan & Zimmermann, 2000; Bullock *et al.*, 2006), interactions of habitat variables, and dynamic species-habitat relationships.

If overestimation occurs, predicted probabilities of species presences need to be adjusted to lower values or reclassified to ordinal scale (e.g., probability of species presence: low, slight, moderate, high, and extremely high); otherwise, for example, an actual probability of occurrence of 0.4 might be represented by an estimated probability of 0.8. Using ordinal scale for habitat suitability was implemented in Ottaviani *et al.* (2004) and Vaughan & Ormerod (2005), and also recommended by Strauss & Biedermann (2007). In the case of underestimation, predicted probabilities of species presence need to be adjusted to larger values or to ordinal values. The major cause of misspecified spread of the predictions is over-fitting and under-fitting (Pearce & Ferrier, 2000). Over-fitting creates the extreme spread of predictions of presence probability, either close to 0 or 1. Models with too many predictors or complex high-order terms tend to have problems of over-fitting. Over-fitting was not very common in the LLR and BRT, likely because these two approaches have built-in cross validation and the complexity was further controlled by Lasso-regularization and simple classifiers, respectively in this study. However, we found that many models, particularly the MaxEnt models, had the problem of lack-of-fit. Under-fitting resulted in the clumped spread of predictions of presence probabilities, namely, underestimating when the actual probability is > 0.5 , and overestimating when the actual probability is < 0.5 . Under-fitting can be remedied by loosening the penalization (i.e., choose smaller penalty parameter in the tuning process) in the Lasso, increasing the complexity of trees or decreasing learning rate in the BRT, and using higher-order and non-linear terms in the MaxEnt. Parameters in a model can be tuned by optimizing discrimination power in the cross validation or independent validation, given that the model is well calibrated. It is unlikely that SDMs, constrained by the availability of testing data, can be calibrated to suit each climate and landform scenarios for each species. A more efficient approach might be generalizing the associations of model transferability and biological or ecological traits (Strauss & Biedermann, 2007), rarity, and management strategies of species.

The assessment of model generality in terms of discrimination power and calibration would be affected by the source of training data and testing data. Using a training dataset with large sample size is one the most efficient ways to develop accurate and reliable SDMs to be transferred over time or space. Our SDMs based on training data of 1035 samples had moderate to good discrimination power. The assessment of model performance has been predominantly based on internal evaluations such as model fitting and cross validation. However, we found that conclusions on the 'best'

model in the model training and external evaluations might be inconsistent (Fig. 3). This emphasizes the need for assessing temporal transferability of SDMs under climate and landscape change because good performance in the internal evaluation does not guarantee transferable models. The calibration of SDMs evaluated based on single-visit samples and occupancy-based independent samples were consistent for most species. Some inconsistency in the model evaluations are expected because sample size, species prevalence, sampling approach, and environmental gradients covered might be different among the testing datasets. The occupancy-based testing data would be more desirable in the case of low detection rate. However, a trade-off of such decomposition of probabilities is that intensive time and labour are required in the occupancy-based design. Occupancy-based designs need temporally replicated sampling over seasons or years, rendering such designs easily constrained by funding and logistical challenges, such as maintaining the same crew. Under most logistical situations, the single-visit models appear more practical than the occupancy models. A comparative study (Welsh *et al.*, 2013) indicated that adjusting non-detection with occupancy models would not necessarily improve accuracy and precision in estimation, compared to ignoring non-detection completely, particularly when the data are sparse.

Discrimination power and calibration could be improved by fine-tuning the set of variables in the model. The main cause of lower accuracy from SDM development to validation is the unavailability or exclusion of local habitat predictors (Onikura *et al.*, 2012; Sindt *et al.*, 2012; Gies *et al.*, 2014). Franklin *et al.* (2013) suggested that local habitat variables need to be included in predictive SDMs under climate change scenarios. Performance of SDMs also relies on how well we could account for the effects of landscape configuration (e.g., connectivity and heterogeneity) and spatial autocorrelation (Elith *et al.*, 2006; Zimmermann *et al.*, 2010; Huang & Frimpong, 2015). Several studies have indicated that landscape composition and configuration are important predictors of species spatial distribution in the climate change context (Steen *et al.*, 2010; Franklin *et al.*, 2013; Kuemmerlen *et al.*, 2015). Comparatively, studies that addressed single types of disturbance would likely estimate current distribution range with bias and misspecified spread, due to the exclusions of other constraints in the fundamental niche of the focal species, and would be even more misleading when the study objective was predicting future ecological patterns (Onikura *et al.*, 2012; Sindt *et al.*, 2012). The commonly used approach that evaluates species' responses purely along environmental gradients (i.e., without considering spatial autocorrelation) in the SDMs could only capture a portion of factors and mechanisms that drive current species distributions and future shifts (Borcard & Legendre, 2002). The SDM frameworks that suit hierarchical structure and stream connectivity and incorporate cutting-edge geospatial methodologies are needed to fine-tune the predictions of distribution changes (e.g., fragmentation of

population) and shifts (e.g., upstream-ward, north-ward) of different species and of different causes.

ACKNOWLEDGEMENTS

This work was funded by the US Geological Survey Aquatic gap analysis program. A special thanks to Brandon Peoples, Stephen Floyd, Joe Buckwalter, Steve Watkins, Caitlin Worsham and numerous summer interns for help with fish sampling in the New River during 2012–2014. We are also very grateful to Dr. Janet Franklin and three anonymous reviewers for their constructive reviews and suggestions on an earlier version of the paper.

DATA ACCESSIBILITY

Fish presence data can be accessed from: <https://www.science.gov/catalog/item/5446a5a1e4b0f888a81b816d> Please refer to Table 2 for the sources of abiotic data.

REFERENCES

- Beitinger, T., Bennett, W. & Mccauley, R. (2000) Temperature tolerances of North American freshwater fishes exposed to dynamic changes in temperature. *Environmental Biology of Fishes*, **58**, 237–275.
- Bivand, R. & Piras, G. (2015) Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, **6318**, 1–36.
- Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- Bouska, K., Whitledge, G. & Lant, C. (2015) Development and evaluation of species distribution models for fourteen native central U.S. fish species. *Hydrobiologia*, **747**, 159–176.
- Brewer, S.K. & Orth, D.J. (2015) Smallmouth Bass *Micropterus dolomieu* Lacepède, 1802. *American Fisheries Society Symposium*, **82**, 9–26.
- Brotons, L., Thuiller, W., Araujo, M. & Hirzel, A. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.
- Brown, C.J., Schoeman, D.S., Sydeman, W.J., Brander, K., Buckley, L.B., Burrows, M., Duarte, C.M., Moore, P.J., Pandolfi, J.M., Poloczanska, E., Venable, W. & Richardson, A.J. (2011) Quantitative approaches in climate change ecology. *Global Change Biology*, **17**, 3697–3713.
- Buisson, L., Thuiller, W., Lek, S., Lim, P. & Grenouillet, G. (2008) Climate change hastens the turnover of stream fish assemblages. *Global Change Biology*, **14**, 2232–2248.
- Bulluck, L., Fleishman, E., Betrus, C. & Blair, R. (2006) Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. *Global Ecology and Biogeography*, **15**, 27–38.
- Cherry, D.S., Dickson, K.L. & Cairns, J. Jr (1975) Temperatures selected and avoided by fish at various acclimation temperatures. *Journal of the Fisheries Research Board of Canada*, **32**, 485–491.
- Cox, D.R. (1958) Two further applications of a model for binary regression. *Biometrika*, **45**, 562–565.
- Drummond, M.A. & Loveland, T.R. (2010) Land-use pressure and a transition to forest-cover loss in the eastern United States. *BioScience*, **60**, 286–298.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Jin, L., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Peterson, A.T. & Phillips, S.J. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- EPA (Environmental Protection Agency), USGS (United States Geological Survey), Horizon Systems Corporations. (2010) NHDplusV1 Data. Available at: <http://www.horizon-systems.com/nhdplus/data.php>
- EPA (Environmental Protection Agency), USGS (United States Geological Survey), Horizon Systems Corporations. (2012) NHDplusV2 Data. Available at: http://www.horizon-systems.com/nhdplus/NHDplusV2_data.php
- Esselman, P.C., Infante, D.M., Wang, L., Wu, D., Cooper, A.R. & Taylor, W.W. (2011) An index of cumulative disturbance to river fish habitats of the conterminous United States from landscape anthropogenic activities. *Ecological Restoration*, **29**, 133–151.
- Fenneman, N.M. & Johnson, D.W. (1946) *Physiographic divisions of the conterminous U. S.* U.S. Geological Survey, Reston, VA.
- Fiske, I. & Chandler, R. (2011) Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, **43**, 1–23.
- Flebbe, P.A., Roghair, L.D. & Bruggink, J.L. (2006) Spatial modelling to project southern Appalachian trout distribution in a warmer climate. *Transactions of the American Fisheries Society*, **135**, 1371–1382.
- Franklin, J. & Miller, J.A. (2009) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge; New York.
- Franklin, J., Davis, F.W., Ikegami, M., Syphard, A.D., Flint, L.E., Flint, A.L. & Hannah, L. (2013) Modelling plant species distributions under future climates: how fine scale do climate projections need to be? *Global Change Biology*, **19**, 473–483.
- Friedman, M. (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, **32**, 675–701.
- Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, **29**, 1189–1232.

- Friedman, J., Jerome, F., Trevor, H. & Rob, T. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1–22.
- Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N. & Wickham, J. (2011) Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing*, **77**, 858–864.
- Gies, M., Sondermann, M., Hering, D. & Feld, C.K. (2014) Are species distribution models based on broad-scale environmental variables transferable across adjacent watersheds? A case study with eleven macroinvertebrate species. *Fundamental and Applied Limnology*, **185**, 3–4.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. (2013) *dismo*: Species distribution modelling. R package version 0.9-3. Available at: <http://CRAN.R-project.org/package=dismo>.
- Huang, J. & Frimpong, E.A. (2015) Using historical atlas data to develop high-resolution distribution models of freshwater fishes. *PLoS ONE*, **10**, e0129995.
- Huang, J. & Frimpong, E.A. (2016) Limited transferability of stream-fish distribution models among river catchments: reasons and implications. *Freshwater Biology*, (in press).
- Frimpong, E.A., Huang, J. & Liang, Y. (2016) IchthyMaps: A Database of Historical Distributions of Freshwater Fishes of the United States. *Fisheries*, (in press).
- Jackson, D.A., Peres-Neto, P.R. & Olden, J.D. (2001) What controls who is where in freshwater fish communities – the roles of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 157–170.
- Jenkins, R.E. & Burkhead, N.M. (1994) *Freshwater fishes of Virginia*. American Fisheries Society, Bethesda, Maryland.
- Kuemmerlen, M., Schmalz, B., Cai, Q., Haase, P., Fohrer, N. & Jähnig, S.C. (2015) An attack on two fronts: predicting how changes in land use and climate affect the distribution of stream macroinvertebrates. *Freshwater Biology*, **60**, 1443–1458.
- Lyons, J., Zorn, T., Stewart, J., Seelbach, P., Wehrly, K. & Wang, L. (2009) Defining and characterizing coolwater streams and their fish assemblages in Michigan and Wisconsin, USA. *North American Journal of Fisheries Management*, **29**, 1130–1151.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, A.J. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.
- Marmion, M., Parvianinen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- Menhinick, E.F. (1991) *The freshwater fishes of North Carolina*. North Carolina Wildlife Resources Commission; Distributed by Larkin Distributors, Raleigh, North Carolina.
- Meynard, C.N. & Quinn, J.F. (2007) Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, **34**, 1455–1469.
- Miller, M.E., Hui, S.L. & Tierney, W.M. (1991) Validation techniques for logistic regression models. *Statistics in Medicine*, **10**, 1213–1226.
- National Fish Habitat Board (2012) *National Fish Habitat Action Plan (NFHAP)*, 2nd ed. Association of Fish and Wildlife Agencies, Washington, DC.
- Nelson, J.S. (2006) *Fishes of the world*, 4th edn. Wiley, New York.
- Oksanen, J., Blanchet, G.F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2015) *vegan*: Community Ecology Package. R package version 2.3-0. Available at: <http://CRAN.R-project.org/package=vegan>.
- Onikura, N., Nakajima, J., Miyake, T., Kawamura, K. & Fukuda, S. (2012) Predicting distributions of seven bitterling fishes in northern Kyushu, Japan. *Ichthyological Research*, **59**, 124–133.
- Ottaviani, D.G., Boitani, J. & Boitani, L. (2004) Two statistical methods to validate habitat suitability models using presence-only data. *Ecological Modelling*, **179**, 417–443.
- Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Price, C.V., Nakagaki, N., Hitt, K.J. & Clawges, R.C. (2006) Enhanced historical land-use and land-cover data sets of the U.S. Geological Survey, USGS Digital Data Series 240. Available at: <http://pubs.usgs.gov/ds/2006/240>.
- PRISM Climate Group. (2004) PRISM climate data. Oregon State University, Corvallis, Oregon. Available at: <http://prism.oregonstate.edu>.
- Pritt, J.J. & Frimpong, E.A. (2010) Quantitative determination of rarity of freshwater fishes and implications for imperiled-species designations. *Conservation Biology*, **24**, 1249–1258.
- R Core Team. (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org/>.
- Rabinowitz, D. (1981) Seven forms of rarity. *The biological aspects of rare plant conservation* (ed. by H. Synge), pp. 205–217. John Wiley and Sons, Chichester.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- Schibalski, A., Lehtonen, A. & Schröder, B. (2014) Climate change shifts environmental space and limits transferability of treeline models. *Ecography*, **37**, 321–335.

- Sharma, S. & Jackson, D.A. (2008) Predicting smallmouth bass (*Micropterus dolomieu*) occurrence across North America under climate change: a comparison of statistical approaches. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 471–481.
- Shingleton, M.V., Hocutt, C.H. & Stauffer, J.R. (1981) Temperature preference of the New River shiner. *Transactions of the American Fisheries Society*, **110**, 660–661.
- Sindt, A.R., Pierce, C.L. & Quist, M.C. (2012) Fish species of greatest conservation need in wadeable Iowa streams: current status and effectiveness of aquatic gap program distribution models. *North American Journal of Fisheries Management*, **32**, 135–146.
- Stauffer, J.R. Jr, Boltz, J.M. & White, L.R. (1995) The fishes of West Virginia. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **146**, 1–389.
- Steen, P.J., Wiley, M.J. & Schaeffer, J.S. (2010) Predicting future changes in Muskegon River watershed game fish distributions under future land cover alteration and climate change scenarios. *Transactions of the American Fisheries Society*, **139**, 396–412.
- Strauss, B. & Biedermann, R. (2007) Evaluating temporal and spatial generality: how valid are species–habitat relationship models? *Ecological Modelling*, **204**, 104–114.
- Swets, J. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Thuiller, W. (2007) Biodiversity: climate change and the ecologist. *Nature*, **448**, 550–552.
- Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267–288.
- Tuanmu, M.-N., Viña, A., Roloff, G.J., Liu, W., Ouyang, Z., Zhang, H. & Liu, J. (2011) Temporal transferability of wildlife habitat models: implications for habitat monitoring. *Journal of Biogeography*, **38**, 1510–1523.
- Vaughan, I.P. & Ormerod, S.J. (2005) The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, **42**, 720–730.
- Wang, L. & Jackson, D. (2014) Shaping up model transferability and generality of species distribution modelling for predicting invasions: implications from a study on *Bythotrephes longimanus*. *Biological Invasions*, **16**, 2079–2103.
- Welsh, A.H., Lindenmayer, D.B. & Donnelly, C.F. (2013) Fitting and interpreting occupancy models. *PLoS ONE*, **8**, e52015.
- Wenger, S.J. & Olden, J.D. (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, **3**, 260–267.
- Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B. & Svenning, J.-C. (2010) New trends in species distribution modelling. *Ecography*, **33**, 985–989.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Key predictors and trends of the partial dependence plots in the boosted regression trees.

Table S2 Results of occupancy models used to estimate the probability of presence of 16 fish species at 80 inter-confluence segments.

Figure S1 Examples of calibration curves drawn by regressing logit of the predicted probability (\hat{p}_i) against the observed occurrence (π_i) at evaluation sites in a logistic (calibration) model.

Figure S2 Discrimination power (accuracy, sensitivity and specificity) of species distribution models in the independent testing.

Figure S3 Partial dependent plots showing that *Etheostoma osburni* (Candy darter) and minimum January temperature are negatively related.

Figure S4 Histograms showing that *Etheostoma osburni* (Candy darter) and minimum January temperature were negatively correlated.

Figure S5 *Post hoc* multiple-group comparisons for the effect of temperature preference of 16 species on the area under the receiver-operating-characteristic curve (AUC) in the model transfers.

BIOSKETCHES

Jian Huang is a post-doctoral scientist working with EAF. He developed a PhD dissertation on assessing the predictive performance and transferability of stream fish distribution models, out of which this manuscript was created.

Emmanuel A. Frimpong is Associate Professor in the department of Fish and Wildlife Conservation whose research focuses on life history, anthropogenic mediation of the distribution of fishes, and quantitative ecology.

Donald J. Orth is Professor in the department of Fish and Wildlife Conservation whose research focuses on stream fish ecology, habitat assessment and management.

Author contributions: E.A.F. and J.H. conceived the ideas; E.A.F. and J.H. collected the data; all authors analysed and interpreted the data with J.H. leading; J.H. led the writing and E.A.F. and D.J.O. co-wrote and edited the manuscript.

Editor: Janet Franklin