

# Modeling spatially varying landscape change points in species occurrence thresholds

TYLER WAGNER<sup>1,†</sup> AND STEPHEN R. MIDWAY<sup>2,3</sup>

<sup>1</sup>U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, The Pennsylvania State University, University Park, Pennsylvania 16802 USA

<sup>2</sup>Department of Ecosystem Science and Management, The Pennsylvania State University, University Park, Pennsylvania 16802 USA

**Citation:** Wagner, T., and S. R. Midway. 2014. Modeling spatially varying landscape change points in species occurrence thresholds. *Ecosphere* 5(11):145. <http://dx.doi.org/10.1890/ES14-00288.1>

**Abstract.** Predicting species distributions at scales of regions to continents is often necessary, as large-scale phenomena influence the distributions of spatially structured populations. Land use and land cover are important large-scale drivers of species distributions, and landscapes are known to create species occurrence thresholds, where small changes in a landscape characteristic results in abrupt changes in occurrence. The value of the landscape characteristic at which this change occurs is referred to as a change point. We present a hierarchical Bayesian threshold model (HBTM) that allows for estimating spatially varying parameters, including change points. Our model also allows for modeling estimated parameters in an effort to understand large-scale drivers of variability in land use and land cover on species occurrence thresholds. We use range-wide detection/nondetection data for the eastern brook trout (*Salvelinus fontinalis*), a stream-dwelling salmonid, to illustrate our HBTM for estimating and modeling spatially varying threshold parameters in species occurrence. We parameterized the model for investigating thresholds in landscape predictor variables that are measured as proportions, and which are therefore restricted to values between 0 and 1. Our HBTM estimated spatially varying thresholds in brook trout occurrence for both the proportion agricultural and urban land uses. There was relatively little spatial variation in change point estimates, although there was spatial variability in the overall shape of the threshold response and associated uncertainty. In addition, regional mean stream water temperature was correlated to the change point parameters for the proportion of urban land use, with the change point value increasing with increasing mean stream water temperature. We present a framework for quantify macrosystem variability in spatially varying threshold model parameters in relation to important large-scale drivers such as land use and land cover. Although the model presented is a logistic HBTM, it can easily be extended to accommodate other statistical distributions for modeling species richness or abundance.

**Key words:** hierarchical Bayesian; landscape; *Salvelinus fontinalis*; species occurrence; thresholds.

**Received** 6 September 2014; **accepted** 9 September 2014; **published** 24 November 2014. Corresponding Editor: D. P. C. Peters.

**Copyright:** © 2014 Wagner and Midway. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. <http://creativecommons.org/licenses/by/3.0/>

<sup>3</sup> Present address: Department of Biology, Coastal Carolina University, Conway, South Carolina 29528 USA.

† **E-mail:** [txw19@psu.edu](mailto:txw19@psu.edu)

## INTRODUCTION

Ecological patterns and processes affecting species distributions and richness are increasing-

ly reported at larger scales than historically investigated. This partly reflects the fact that many stressors, such as land use and climate change, can affect a species over much, if not all,

of its range. Although small-scale studies (e.g., site level, such as a stream or forest catchment) provide the groundwork for ecological inference, appropriate scaling up of ecological findings to a macrosystems level (e.g., regions to continents; Heffernan et al. 2014) is becoming critical toward the development of a broad understanding of complex and often unseen ecological phenomena. Furthermore, as large-scale ecological conditions present differently to society—such as negatively through droughts and sea-level rise, or positively through ecosystem services—an increased value is being placed on the ability to predict large-scale biological distributions, ecological processes, and the outcomes of both (Levy et al. 2014).

Landscape characteristics, such as land use and land cover, are often a primary focus of large-scale investigations into species distributions and associated modeling efforts for several reasons: (1) many species distributions are tightly linked to local and regional landscape properties because landscape-scale processes can affect the survival and dispersal of populations (Allan 2004); (2) land use and land cover are affected by natural processes (e.g., insect infestation, wildfire) and anthropogenic activities (e.g., deforestation, climate change; Gevrey et al. 2009), and thus they have the potential to vary spatially and temporally; and (3) these data are readily available at multiple spatial and temporal scales. Because many individual species and the diversity of species in a region interact in complex ways with local and regional landscape properties (Mackey and Lindenmayer 2001), species distributions are not equally likely across landscapes. Often a threshold, an abrupt change in a state variable (e.g., occurrence or abundance), occurs along a narrow gradient of a landscape driver (Qian 2012). The mathematical value at which the response change takes place is referred to as a change point.

Threshold responses to landscapes have been identified for a variety of taxa and biodiversity. For example, thresholds have been identified for songbird and forest bird community occurrence in relation to landscape structure and amount of broadleaf forests, respectively (Betts et al. 2007, 2010). Reunanen et al. (2004) determined a landscape threshold for the occurrence of the Siberian flying squirrel (*Pteromys volans*), and

Estavillo et al. (2013) reported a biodiversity threshold of Atlantic forest small mammal biodiversity in response to the amount of forest cover. Thresholds in aquatic systems are also evident. Wagner et al. (2013) observed a steep decline in occurrence probability for eastern brook trout (*Salvelinus fontinalis*) at 5–10% impervious surface in upstream network catchments, and Thomson et al. (2010) detected temporal change points in abundance of pelagic estuarine species through examination of water quality over 40 years.

Because landscape thresholds likely exist for a variety of taxa, approaches for estimating landscape thresholds have recently evolved from conceptual models (Suding and Hobbs 2009) to more quantitative approaches (Brenden et al. 2008, Jones et al. 2011, Qian 2012). However, commonly used non-hierarchical threshold models (e.g., piecewise regression, nonparametric deviance reduction models; Brenden et al. 2008) estimate a single change point for any given dataset (i.e., a specific region of interest), which may not be appropriate if thresholds vary spatially in response to environmental gradients, such as climate or geological setting. One way to attempt to address this with traditional non-hierarchical methods would be to subset the data into subregions and estimate subregion-specific change points. This can be problematic, however, because subsetting the data may result in too few observations to estimate a change point, and if they are estimable the uncertainty around the estimate may be large. Thus, the use of traditional methods may be inappropriate for examining spatially varying thresholds across a species range. In addition, these traditional methods do not allow for the ability to model spatially varying parameters, such as slopes and change points, in an attempt to elucidate large-scale patterns and drivers of these parameters. This ability to model spatially varying parameters is an important component for both species distributional and macrosystem investigations, and may allow for the detection of cross-scale interactions that occur when the state variable (e.g., species occurrence) and drivers operate at different spatial scales (Soranno et al. 2014).

We present a newly developed hierarchical Bayesian threshold model that allows for the estimation and modeling of spatially varying

parameters within a macrosystem context. Although we illustrate the ability to model spatially varying parameters, the use of our hierarchical modeling framework is applicable regardless of the amount of spatial variability that may be present in parameters. We use brook trout, a coldwater fish species native to eastern North America, as a study organism to illustrate estimating spatially varying landscape thresholds in occurrence in relation to land use and land cover. However, our model can easily be extended to accommodate other state variables such as abundance or species richness. Simultaneously, we also model spatially varying parameter estimates (including change point estimates) as a function of mean regional stream water temperature to illustrate the integration of larger-scale predictor variables in an attempt to identify drivers of large-scale patterns and processes affecting species distributions.

## METHODS

### *Case study dataset*

We used brook trout occurrence data from streams located within the species native range across the eastern USA. Data were compiled through direct contact with state agencies and by downloading data directly from the Multistate Aquatic Resources Information System (MARIS) website (<http://www.marisdata.org>). Brook trout detection/nondetection data were compiled and sampling locations were linked to the National Hydrography Dataset Plus Version 1.0 (NHDPlus) dataset, which served as the base spatial unit for our analysis. The resulting dataset had  $n = 7798$  stream reaches with brook trout detection/nondetection observations. The data were subsetted to include only those network catchments (the entire upstream catchment) with areas smaller than  $125 \text{ km}^2$ , which corresponds to headwater streams and habitats suitable to brook trout. We chose to use Ecological Drainage Units (EDUs) as a regionalization framework based on Cheruvilil et al. (2013), who determined that EDUs outperformed other regionalizations for grouping similar aquatic ecosystems. EDUs are also a natural choice for a regionalization when examining stream ecosystems because they are watershed-based units that share common physiography, climate, and connectivity (Higgins et

al. 2005). Within our study region there were 41 EDUs (Fig. 1).

### *Model performance*

Although the primary purpose of our analysis was to gain an understanding of species-landscape relationships, and not to build a model that maximized predictive performance (Kuhn and Johnson 2013), we chose to randomly withhold 10% of observations from each EDU with  $>10$  observations for model validation. This resulted in a validation dataset ( $n = 777$  observations,  $n = 34$  EDUs) and a dataset used to fit the models ( $n = 7021$ ). To evaluate predictive ability we calculated the posterior distribution of receiver operating characteristics area under the curve (AUC). AUC provides a useful measure of performance relative to chance and ranges from 0 to 1. A value of 0.5 indicates that the model performs no better than a random guess, while a value of 1.0 implies perfect prediction. We calculated AUC to evaluate the model's ability to correctly predict locations that were occupied by brook trout for both the validation dataset and the dataset used to fit the models.

### *Predictor variables*

We chose to focus on two landscape predictors of brook trout occurrence that have been previously shown (Stranko et al. 2008, Wagner et al. 2013) to be important landscape characteristics influencing their distribution and are likely to show threshold effects on occurrence: the proportion of agricultural and urban land use in the upstream network catchment (correlation between agricultural and urban land use  $r = 0.07$ ). We chose not to report the effects of forested land cover in the network catchment on brook trout occurrence because the proportion of forest cover was correlated with both the proportion of agricultural and urban land use (correlation between agricultural land use and forest land cover  $r = -0.78$ ; correlation between urban land use and forest land cover  $r = -0.55$ ). Landscape predictors were obtained from the National Land Cover Database (U.S. Geological Survey 2008) and summarized within the upstream network catchment for each stream reach following Esselman et al. (2011).

Because water temperature is a key determinant of habitat suitability for brook trout

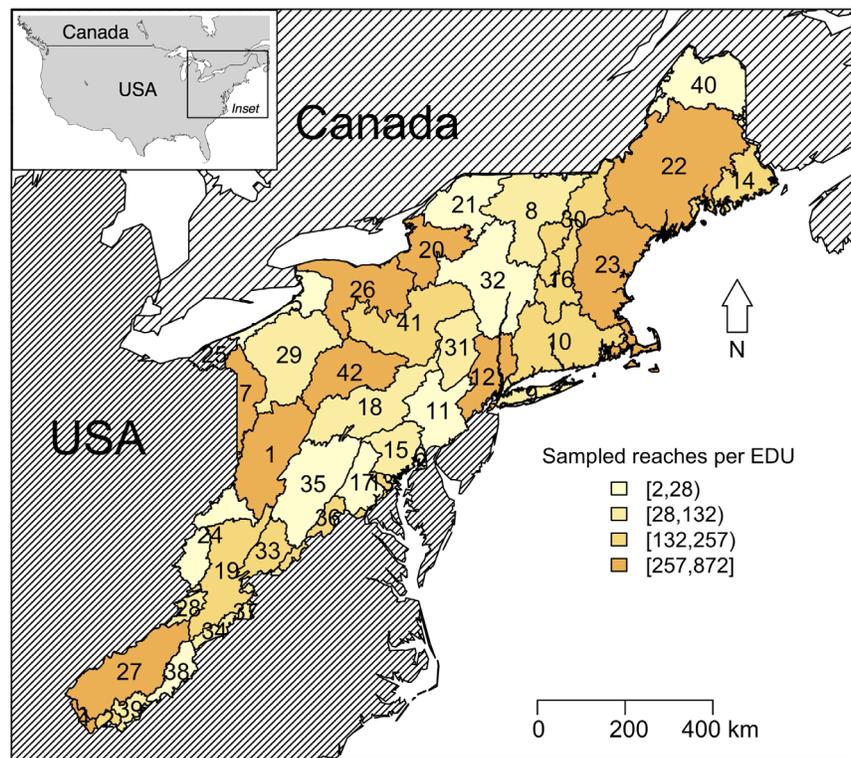


Fig. 1. Map of study region and 41 ecological drainage units (EDUs; labeled 1–41) which contained 7798 sampled stream reaches describing detection/nondetection of eastern brook trout. Shading of EDUs represents the number of streams sampled for brook trout within the EDU. EDU numbers presented here are consistent throughout the study. Note: EDU 25 is within the brook trout range, but contained no streams meeting our network catchment criteria, and was thus excluded from analyses.

regionally (MacCrimmon et al. 1971), we used mean stream water temperature as an EDU-level predictor to model estimated threshold model parameters in the context of identifying large-scale patterns. Specifically, we used predicted maximum 30-day mean stream water temperature from May through October (hereafter referred to as mean water temperature) from a neural network ensemble model (DeWeber and Wagner 2014) developed to predict mean daily water temperatures throughout the study region (DeWeber and Wagner, *in press*). EDU mean stream water temperature was summarized from reach-level predictions. Because brook trout is a cold-water species that is sensitive to landscape alteration, we predicted that change points and post-change point slopes (or the difference in slopes pre- and post-change point) would vary regionally capturing a moderating effect of stream water temperature at more northern

latitudes (correlation between predicted mean EDU stream water temperature and EDU mid-point latitude  $r = -0.73$ ). For example, we predicted an overall negative effect for both the proportion of agricultural and urban land uses, where brook trout occurrence would decrease rapidly at low levels of these land uses, and then at a change point the decline would be more gradual. We expected the change point to be at greater proportions of agricultural and urban land use and the post-change point slope to be less steep as you decreased mean EDU stream water temperatures, reflecting a moderating effect of temperature on anthropogenic influences to thermal habitat.

#### *Hierarchical Bayesian threshold model*

We extend a logistic threshold model (Jones et al. 2011) to a hierarchical form where all parameters are allowed to vary according to a

pre-specified spatial unit. For our case study we chose to use a hockey stick model based on brook trout ecology (i.e., they require high water quality habitats and are thermally sensitive) where we predicted a sudden change in the relationship between occurrence probability and land use types, and because other studies (e.g., Stranko et al. 2008) have illustrated a threshold relationship between brook trout populations and landscape characteristics such as developed land use. In addition to our prediction, and evidence from the literature, that we would expect a sudden change point (i.e., that the use of a hockey stick model would be appropriate), we also examined the posterior distributions of the change points to assess the existence of a change point (Qian 2014). We would expect relatively narrow posterior distributions if change points do in fact exist. However, depending on the species and landscape setting of interest, other statistical change point models (e.g., disjointed broken stick, step function) could be used and parameterized similarly to what is described below.

In this study, all parameters were allowed to vary among EDUs. Our modeling approach was first to fit a model that did not include an EDU-level predictor (i.e., was unconditional at level 2 of the model). This allowed for the quantification of among-EDU variation in threshold responses. We then fitted models that included the EDU-level predictor in an attempt to explain observed variation in model parameters. This model is parameterized for investigating thresholds in landscape predictor variables that are measured as proportions (i.e., catchment land use and land cover), and therefore restricted to values between 0 and 1. The model provides estimates of the population-average change point (a change point across all data and all EDUs) and EDU-specific change points. In addition, all EDU-specific parameters can be examined (simultaneously modeled) in an attempt to identify large-scale patterns and processes. The general form of the hierarchical threshold model is as follows:

$$y_i \sim \text{Bernoulli}(\psi_i), \text{ for } i = 1, \dots, n \quad (1)$$

$$\text{logit}(\psi_i) = \alpha_{j[i]} + \beta_{j[i]}x_i + \delta_{j[i]}(x_i - \phi_{j[i]})_+ \quad (2)$$

$$\begin{pmatrix} \log(\alpha_j) + C \\ \log(\beta_j) + C \\ \log(\delta_j) + C \\ \text{logit}(\phi_j) \end{pmatrix} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ for } j = 1, \dots, J \quad (3)$$

$$\boldsymbol{\mu} = (\bar{\alpha}, \bar{\beta}, \bar{\delta}, \bar{\phi}) \quad (4)$$

where the logit-probability of occurrence is modeled with group-specific, spatially varying intercepts ( $\alpha_j$ ), regression slopes prior to the change point ( $\beta_j$ ), change in regression slopes after the change point ( $\delta_j$ ), and change points ( $\phi_j$ ). The notation  $j[i]$  indexes EDU  $j$  for site (observation)  $i$ . The final term in Eq. 2  $(x_i - \phi_{j[i]})_+$  is equal to  $(x_i - \phi_{j[i]})$  if  $x_i > \phi_{j[i]}$ , and 0 otherwise.

We modeled the parameters  $\alpha_j$ ,  $\beta_j$ , and  $\delta_j$  on the log scale and  $\phi_j$  on the logit scale. Because of numerical instability during estimation a logarithmic-transformation was used to rescale parameters to similar magnitudes, which greatly improved convergence and efficiency of the MCMC sampling (Kimura 2008). As a result of the log-transformation  $\alpha_j$ ,  $\beta_j$ , and  $\delta_j$  were constrained to be positive during estimation; however, in some cases one or more of these parameters may be negative. Thus,  $\tilde{\alpha}_j = \alpha_j + C$ ,  $\tilde{\beta}_j = \beta_j + C$  and  $\tilde{\delta}_j = \delta_j + C$  were actually estimated. The bias of  $C$  was subtracted when converting parameters to the original scale such that, for example,  $\hat{\alpha}_j = e^{\tilde{\alpha}_j} - C$ . Thus, a constant  $C$  (usually 10) was added to  $\alpha_j$ ,  $\beta_j$ , and  $\delta_j$  to ensure that negative values were possible, as necessary (i.e., if possible parameter space did not include negative values then a constant was not necessary). The magnitude of the added constant was determined from preliminary model fits and the models were not sensitive to the value chosen for  $C$  as long as  $C$  was large enough to encompass the possible parameter space. Although this is somewhat an unconventional parameterization, it was done purely to improve numerical stability and results of this parameterization were nearly identical to a model fit where parameters were modeled on their original scale, which was possible in a preliminary analysis during model development. We also modeled the logit of  $\phi_j$  to ensure that estimates of change points for land use and land cover predictors were restricted to be between 0 and 1, and to ensure a plausible parameter space for coefficients estimated to

model variation in varying change points; spatially varying change points (which are proportions) were modeled as a linear function of predictors on the logit-scale; (see below). The spatially varying model parameters were assumed to come from a multivariate normal (MVN) distribution, where  $\mu$  and  $\Sigma$  are the population mean and variance-covariance matrix, respectively. As such,  $\bar{\alpha}$ ,  $\bar{\beta}$ ,  $\bar{\delta}$ ,  $\bar{\phi}$  describe the population-average threshold relationship across all EDUs, using all data. Prior probabilities on all parameters were diffuse. We used diffuse normal priors for  $\bar{\alpha}$ ,  $\bar{\beta}$ ,  $\bar{\delta}$ ,  $\bar{\phi}$  and modeled  $\Sigma$  using the scaled inverse-Wishart distribution (Gelman and Hill 2007).

To model the effects of EDU-specific predictors (mean stream water temperature in this analysis) on spatially varying parameters of interest, Eq. 4 can be modified as follows:

$$\begin{pmatrix} \log(\alpha_j) + C \\ \log(\beta_j) + C \\ \log(\delta_j) + C \\ \text{logit}(\phi_j) \end{pmatrix} = \begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \\ \gamma_{\delta 0} + \gamma_{\delta 1} \times \text{temperature}_j \\ \gamma_{\phi 0} + \gamma_{\phi 1} \times \text{temperature}_j \end{pmatrix}. \quad (5)$$

Eq. 5 shows the EDU-specific predictor of temperature modeling variation in the change in regression slopes after the change point ( $\delta_j$ ) and change points ( $\phi_j$ ), but predictors can be applied to any or all parameters depending on hypothesized spatial dynamics of any given system of interest. Diffuse normal priors were used for  $\gamma_x$ . We ran three parallel Markov chains beginning each chain with different values. From a total of 150,000 samples from the posterior distribution the first 100,000 samples of each chain were discarded then we retained every fourth sample for a total of 375,000 samples used to characterize the posterior distributions. We assessed convergence for all parameters both visually (trace plots and plots of posterior distributions), as well as with the Brooks-Gelman-Rubin statistic,  $\hat{R}$ , with values  $<1.1$  indicating convergence. Analyses were run using JAGS in the rjags package (Plummer 2013), run from within R (R Core Team 2014; see Supplement).

As parameterized, the occurrence model does not account for potentially imperfect detection when sampling brook trout (MacKenzie et al. 2002). If detection probability is  $<1$ , the effects of

landscape covariates on species occurrence could be biased (i.e., underestimated; Gu and Swihart 2004). However, range-wide data that would allow for accounting for imperfect detection do not exist for brook trout. In addition, the detection probability for brook trout is high, ranging from 0.87 (95% credible interval [CRI] = 0.79, 0.93) to 0.99 (95% credible interval = 0.98, 1.0; Wagner et al. 2013). However, if appropriate data exist, this modeling framework could be extended to an occupancy modeling framework, where imperfect detection is explicitly incorporated into the model.

## RESULTS

### Summary statistics

The number of stream reaches with brook trout detection/nondetection data within each EDU ranged from 2 to 872 (mean = 190, standard deviation [SD] = 231; Fig. 1). A little over half (51%) of the 7798 stream reaches had brook trout present. The proportion of agricultural land cover in the upstream network catchments ranged from 0 to 0.96 (mean = 0.13, median = 0.05, SD = 0.17), and the proportion of urban land use in the network catchments ranged from 0 to 1.0 (mean = 0.07, median = 0.03, SD = 0.11).

### Model performance

The predictive performance for both the agricultural and urban land use models, which included the EDU-level predictor, was similar. The models were able to predict brook trout occurrence much better than chance (agricultural land use model: mean AUC = 0.79, SD = 0.001; urban land use model: mean AUC = 0.78, SD = 0.001); however, they performed no better than chance when predicting the validation data set (agricultural land use model: mean AUC = 0.52, SD = 0.02; urban land use model: mean AUC = 0.51, SD = 0.02).

### Spatially varying thresholds

Spatially varying thresholds in brook trout occurrence were identified for both the proportion of agricultural and urban land use. As expected, there was, on average, a negative effect of agricultural and urban land use on brook trout occurrence with relatively little among-EDU variation in threshold estimates. The popula-

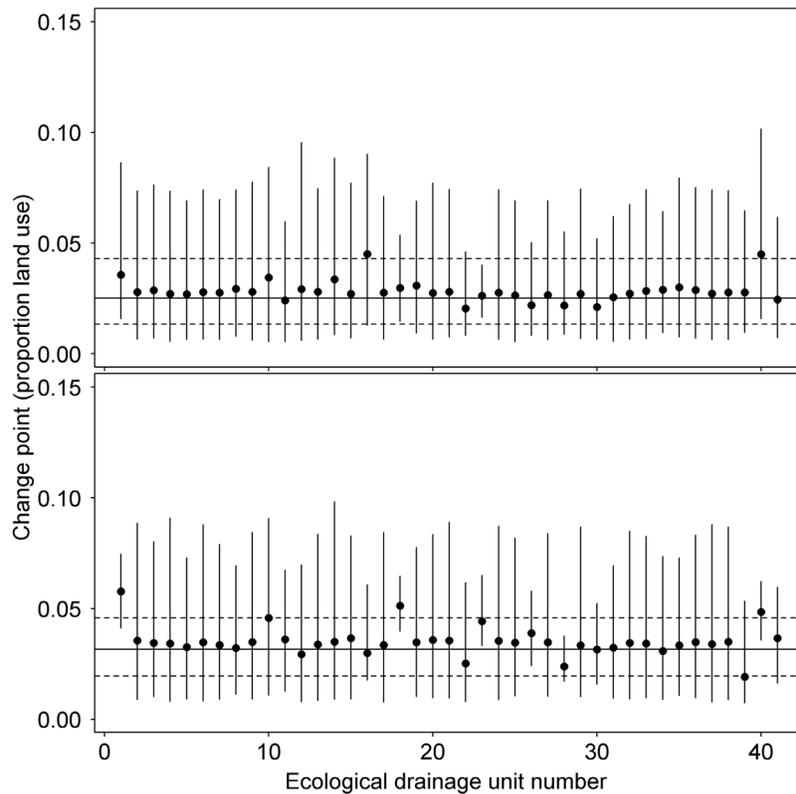


Fig. 2. EDU-specific change point estimates ( $\phi_j$ ) for brook trout occurrence in relation to proportion of network agricultural land use (upper panel) and proportion network urban land use (lower panel). Estimates are from models with no EDU-level predictors. The black dots indicate posterior means, the vertical lines indicate 90% credible intervals, with the population mean and 90% credible intervals indicated by the solid and dashed horizontal lines, respectively. EDU numbers correspond to those presented in Fig. 1.

tion-average change point ( $\bar{\phi}$ ) for proportion of agricultural land use suggested an abrupt decrease in the probability of brook trout occurrence when the proportion of agricultural land use exceeded 0.025 (90% CRI = 0.01, 0.04) in the network catchment. The change point was relatively similar across EDUs, with EDU-specific posterior means ( $\phi_j$ ) ranging from 0.02 to 0.05 (Fig. 2). The population-average change point for the proportion of urban land use in the network catchment indicated an abrupt decline in the probability of brook trout occurrence after the proportion of urban land use in the network catchment exceeded 0.03 (90% CRI = 0.02, 0.04). Similar to the findings for agricultural land use, there was relatively little variation among EDUs in the urban land use change point, ranging between 0.02–0.06 among EDUs (Fig. 2). The estimated change points for both analyses were

well concentrated, lending support for the existence of a threshold response of brook trout to both agricultural and urban land use. Although there was not substantial variation among EDUs in change points, there was variability in the predicted EDU-specific threshold relationships and associated uncertainty for both agricultural and urban land use (Figs. 3 and 4).

#### *Modeling spatially varying parameters*

To illustrate the ability to model large-scale variation in threshold model parameters, we modeled the change point ( $\phi_j$ ) and post-threshold change in slope ( $\delta_j$ ) as a function of EDU mean stream water temperature. For the agricultural land use model, mean stream water temperature was not an important predictor of either parameter (i.e., the 90% credible intervals

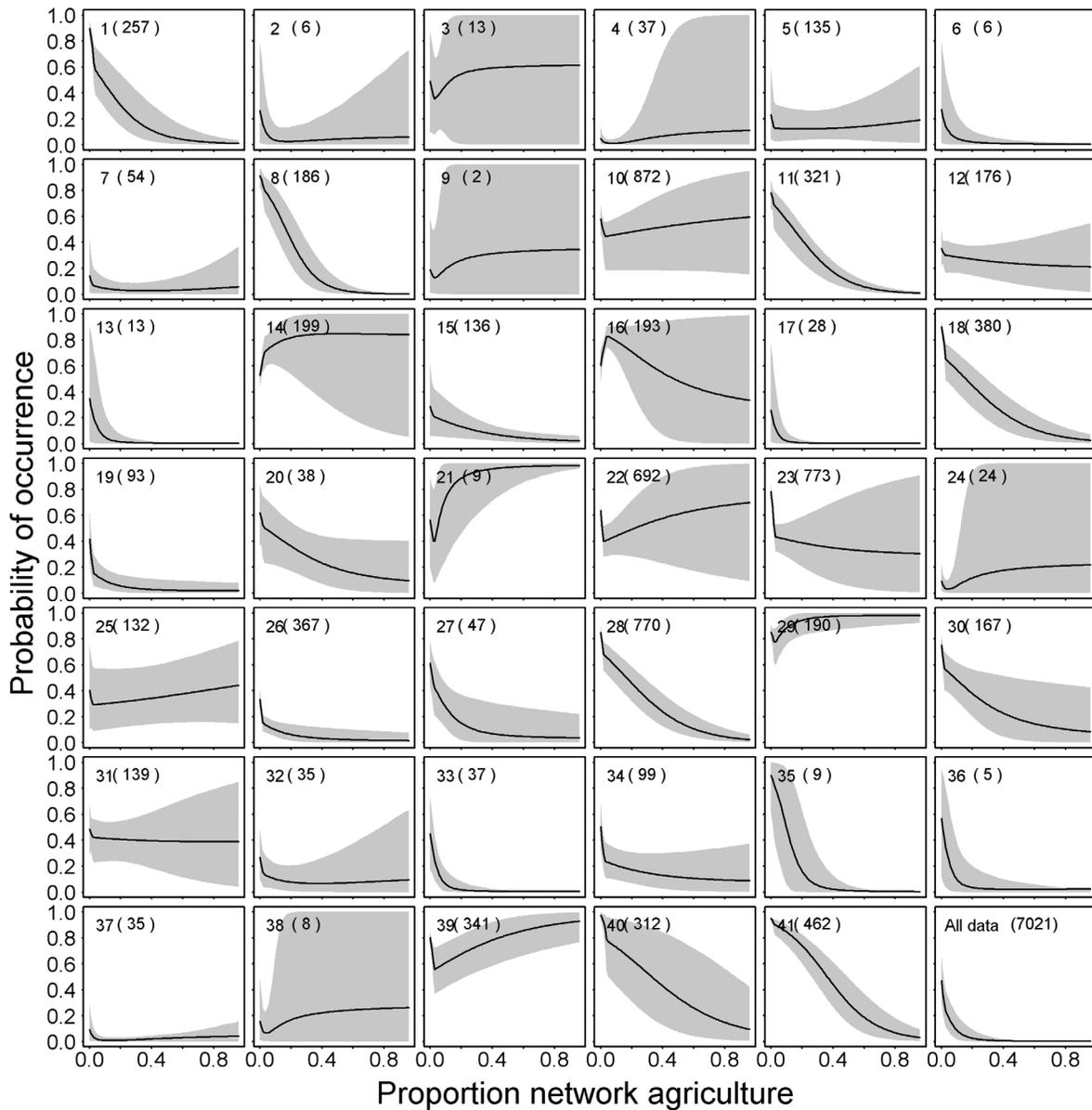


Fig. 3. EDU-specific threshold relationships (black lines) with 90% credible regions (shaded area) for brook trout occurrence probability in response to proportion of network catchment agricultural land use. Estimates are from a model with no EDU-level predictors. EDU numbers correspond to those presented in Fig. 1, and the lower right panel shows the population level (all data) threshold relationship. Number of observations are in parentheses.

for the effect of mean stream water temperature [ $\gamma_{\delta 1}$  and  $\gamma_{\phi 1}$ ] overlapped zero). The posterior mean effect of mean stream water temperature on the EDU-specific agricultural land use change points was 0.22 (90% CRI =  $-0.17, 0.64$ ), and the effect of mean stream water temperature on the

post-change point change in slope was  $-0.06$  (90% CRI =  $-0.13, 0.02$ ; Figs. 5 and 6). For the effect of urban land use, contrary to our predictions, mean stream water temperature was positively correlated with the change point parameter (posterior mean = 0.49, 90% CRI =

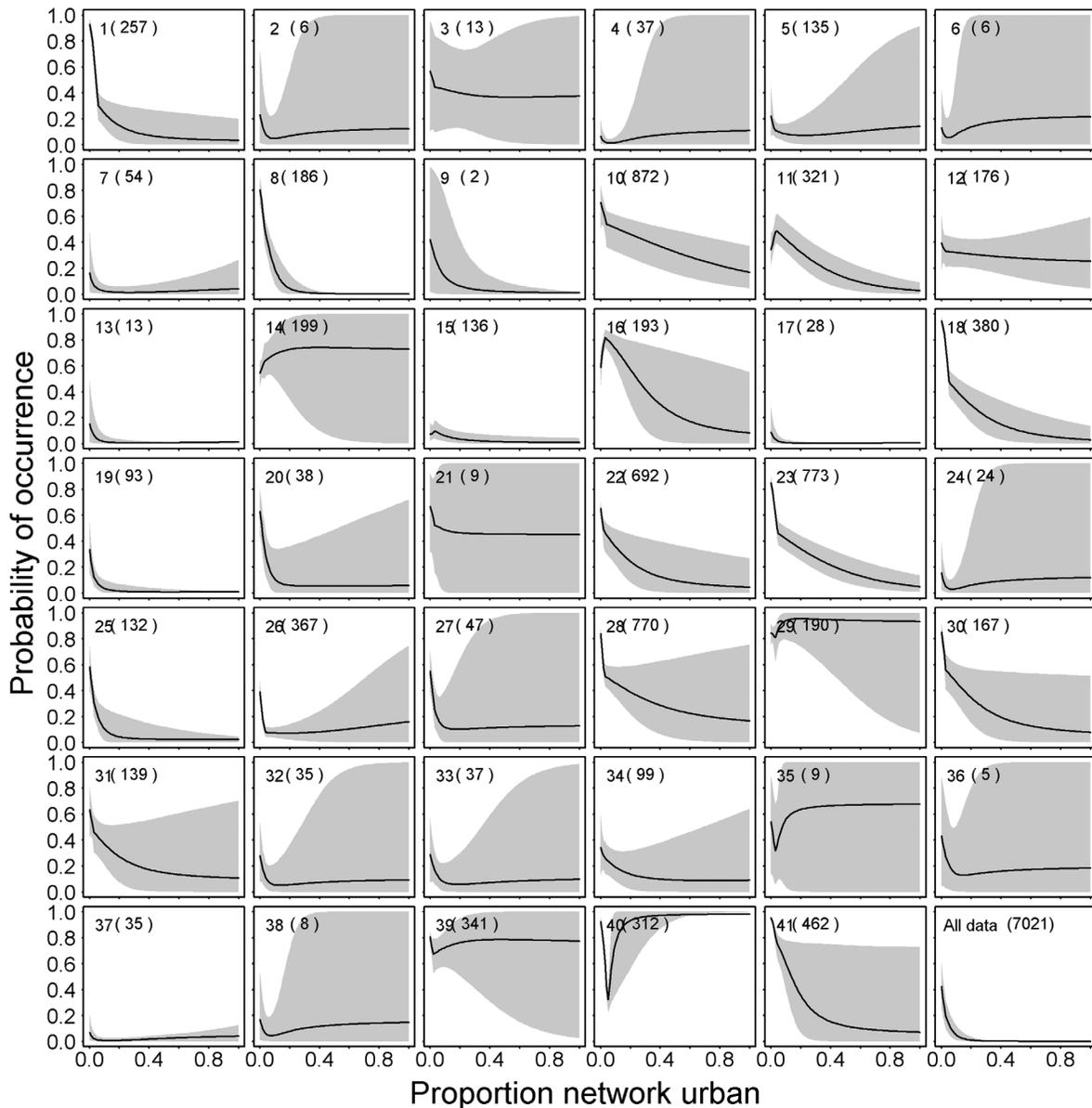


Fig. 4. EDU-specific threshold relationships (black lines) with 90% credible regions (shaded area) for brook trout occurrence probability in response to proportion of network catchment urban land use. Estimates are from a model with no EDU-level predictors. EDU numbers correspond to those presented in Fig. 1, and the lower right panel shows the population level (all data) threshold relationship. Number of observations are in parentheses.

0.18, 0.82; Fig. 7), with the proportion of urban land cover in the network catchment resulting in an increasing change point as EDU mean stream water temperature increased (Fig. 8). However, mean stream water temperature was not an important predictor of the post-change point change in slope parameter (posterior mean =

-0.05, 90% CRI = -0.14, 0.03; Fig. 7).

## DISCUSSION

Because processes affecting spatially structured populations act at multiple spatial scales, and because these processes are likely to vary

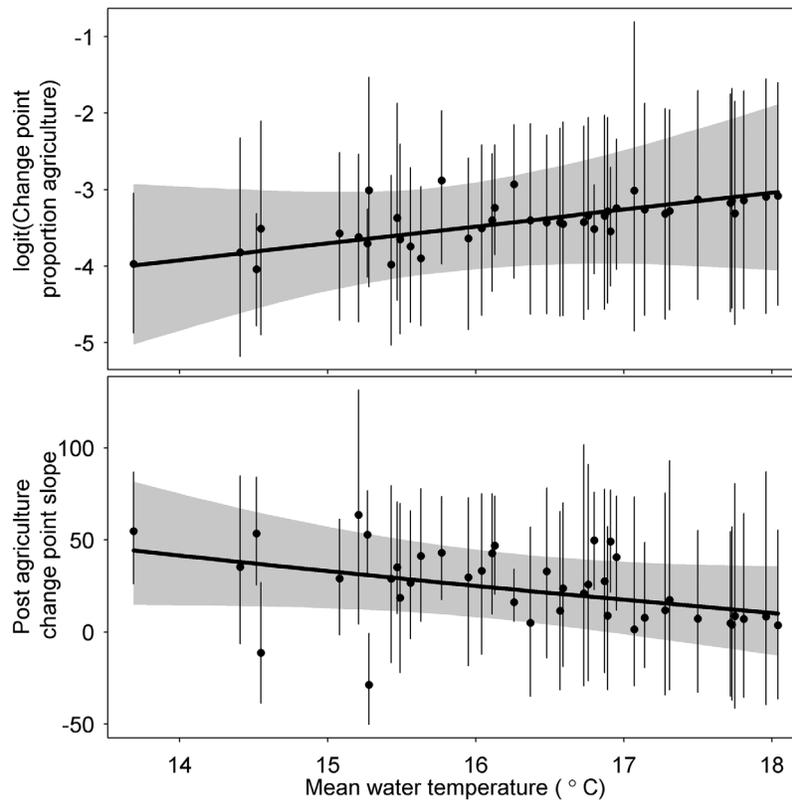


Fig. 5. Relationship between EDU-specific change points ( $\text{logit}(\phi_j)$ ; top panel) and post-change point change in slope ( $\delta_j$ ; bottom panel) from the network catchment agricultural land use threshold model with EDU mean stream water temperature. Solid points are posterior means and vertical lines are 90% credible intervals. The thick black line represents the hierarchical regression line and the grey shading is the 90% credible region. The 90% credible intervals for the effect of mean stream water temperature in both panels overlapped 0. The lower panel EDU estimates are displayed geographically in Fig. 6.

spatially as a function of region-specific ecological characteristics (e.g., climate, geology), there is a need for methods to quantify and describe macrosystem variability (Levy et al. 2014). We described a Bayesian hierarchical threshold model for investigating large-scale, spatially varying thresholds in species occurrence that can easily be extended to accommodate additional state variables such as abundance and species richness. Our model is specifically parameterized to accommodate landscape predictors that are measured as proportions, such as land use and land cover, data that are commonly used in species distribution models.

#### *Brook trout case study*

We identified spatially varying thresholds in

brook trout occurrence for both the proportion of agricultural and urban land use in the upstream network catchment. Brook trout are a cold-water fish species native to eastern North America; however, like many species (Wilcove et al. 1998, Pimm and Raven 2000), populations are declining over much of their native range due largely to habitat degradation and loss. As a result, many brook trout populations are isolated and restricted to headwater stream systems (Hudy et al. 2008). Thus, it is not unexpected that we observed an overall negative relationship between brook trout occurrence and agricultural and urban land uses. The estimated change points for both the proportion agricultural and urban land uses in the network catchment were low (0.025 and 0.03, respectively), and values of

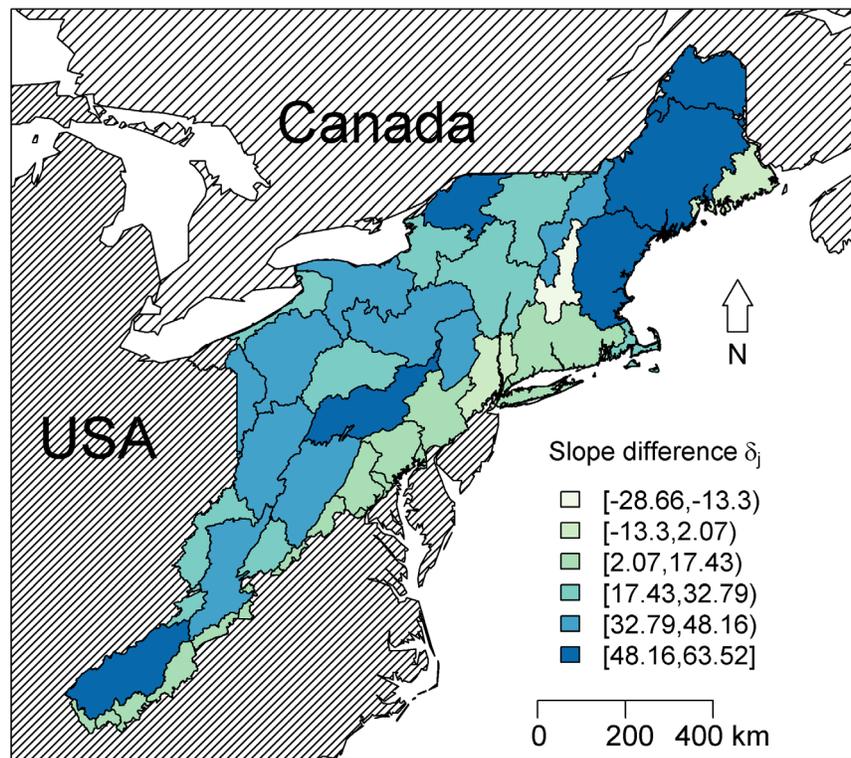


Fig. 6. EDU-specific changes in slope ( $\delta_j$ ) estimated from the network catchment agricultural land use model mapped by EDU. Data here are regressed against EDU mean stream water temperatures in the lower panel of Fig. 5.

these land uses in the network catchment above this change point resulted in a steep decline in brook trout occurrence probability. Wagner et al. (2013) also found a negative effect of the proportion of agricultural land use in the network catchment when evaluating brook trout occupancy dynamics in Pennsylvania, USA. Their study found brook trout occupancy was still  $>0.20$  even when the proportion of network agricultural land use exceeded 0.70, which, when looking at EDU-specific estimates, is similar to our findings. For example, the EDUs in our study that roughly encompass the study region by Wagner et al. (2013) are EDUs 28, 40 and 41, all of which show relatively higher occurrence probability at high levels of agricultural land use in the network catchment compared to the population-average relationship. In addition, our region-wide change point estimate for the effect of urban land use matches well with findings by Stranko et al. (2008), who found that brook trout almost never occurred in catchments where impervious

land cover exceeded 4%. Wagner et al. (2013) also found a steep decline in brook trout occupancy as the percentage of impervious surfaces in the network catchment increased.

For both landscape thresholds, there was relatively little spatial variability in change point estimates. We expect, however, that for some species, depending on life history characteristics and associated habitat requirements, and the nature of local and regional landscape dynamics and interactions that thresholds may vary substantially across space (and time). Furthermore, explicitly accounting for landscape heterogeneity in our threshold model may improve inference about local and regional drivers in species occurrence (Suding and Hobbs 2009). Even with relatively little spatial variation in change points, urban change points were positively correlated with mean stream water temperature. This positive relationship was opposite to what we predicted, and may be a result of other local factors mediating the effects of urban land use on

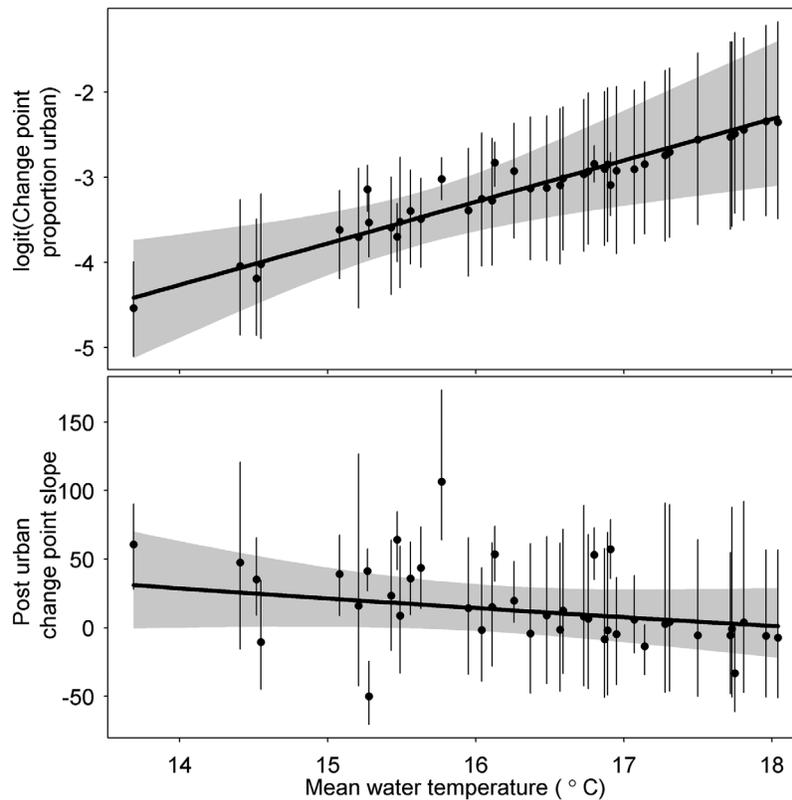


Fig. 7. Relationship between EDU-specific change points ( $\text{logit}(\phi_j)$ ; top panel) and post-change point change in slope ( $\delta_j$ ; bottom panel) from the network catchment urban land use threshold model with EDU mean stream water temperature. Solid points are posterior means and vertical lines are 90% credible intervals. The thick black line represents the hierarchical regression line and the grey shading is the 90% credible region. The 90% credible interval for the effect of mean stream water temperatures on EDU change points (top panel) did not overlap zero (see text for details) and is displayed geographically in Fig. 8. The 90% credible interval for the effect of mean stream water temperature on the post-change point change in slope overlapped 0.

stream ecosystems. Regardless of the exact mechanism, the range of change point effect sizes was small (0.02–0.04), suggesting that regardless of mean stream water temperature, small amounts of urban land use in the network catchment can have deleterious effects on brook trout occurrence.

The relatively poor ability of both the agricultural and urban land use models to predict occurrence probability for the validation dataset was expected. We would not expect that a model with only a few predictors would have high predictive abilities over a large spatial extent. The fact that these models were not able to predict well at new sites is a result of not being able to account for other biological and abiotic factors

that interact to ultimately determine brook trout occurrence in any given stream. Unfortunately, such detailed ecological databases are often not available at macroscales (Rüegg et al. 2014). Regardless of predictive performance, the results provide useful information to help guide the conservation and management of brook trout across their native range (Wagner et al. 2014).

More apparent than the spatial variability in change points was the difference in parameter uncertainty among EDUs, as indicated by the EDU-specific credible regions in Figs. 3 and 4. This uncertainty was due, in part, to low numbers of sample sites in some regions while other regions contained sites with land cover values spanning only a portion of the range

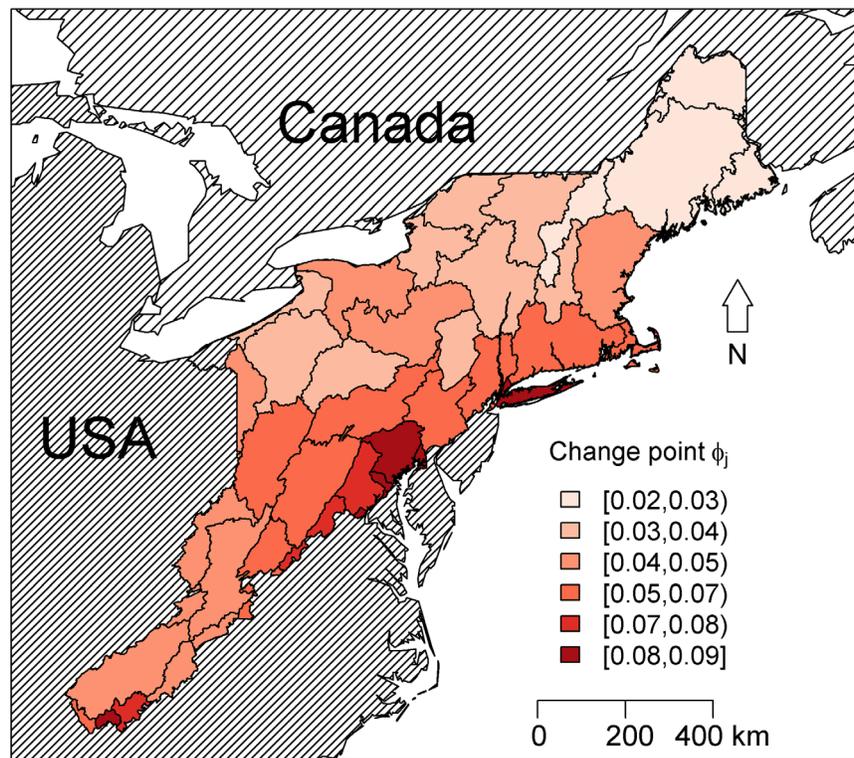


Fig. 8. EDU-specific changes in change point ( $\phi_j$ ) estimates from the network catchment urban land use model mapped by EDU. Data here are regressed against EDU mean stream water temperatures in the upper panel of Fig. 7.

contained in the entire dataset. However, the ability of our model to be able to estimate change point parameters in regions with low sample sizes and with relatively narrow land use values is a strength of using a hierarchical modeling approach. These estimates, however, are shrunk towards the population-average, as a result of borrowing information from the entire ensemble of data.

#### *Hierarchical Bayesian threshold modeling*

The hierarchical Bayesian approach accommodates the spatially unbalanced sample sizes often observed in large-scale investigations, as was the case in our study. Additionally, when the data in a given region are not adequate to fully describe a threshold relationship, the hierarchical modeling approach uses a common prior distribution on model parameters that allows regions to share information, allowing for EDU-specific estimates even for data poor regions (Gelman and Hill 2007, Qian et al. 2010). This ability to borrow

strength, also referred to as partial pooling, is a desirable alternative to full pooling, where all regions are assumed to have the same threshold model parameters. Partial pooling operates under the assumption that species occurrence to land use and land cover across large regions is similar, but not identical.

The ability to model parameters in a hierarchical model is particularly important toward advancing our understanding of large-scale ecological patterns and processes. In our study, we evaluated region-specific parameters in the context of mean stream water temperature; however, the model framework presents no limits on which parameters can be measured against regional environmental gradients and anthropogenic stressors. Although understanding brook trout occurrence change points in the context of mean stream water temperatures may be a specific example, the possibilities are great to further our understanding of large-scale distributions for a variety of taxa. That said, our model

represents a specific type (hypothesis) of ecological threshold, and one must choose the appropriate form of threshold (Qian 2014) for a given system of interest.

## CONCLUSIONS

Mapping, modeling, and understanding species distributions and occurrences is fundamental to the conservation and management of many species. Yet species and systems can be more vulnerable than they appear (Suding and Hobbs 2009), and model choice and change point estimation can be challenging (Qian and Cuffney 2012). Further constraining such efforts is the fact that species distributions and dynamics often respond to conditions that act at various spatial scales. However, a good deal of information can be had by extending hierarchical modeling structures to simultaneously consider multiple scopes of inference. Ecology at the base level—such as an individual home range or subpopulation dynamic—remains the building block of insight, yet appropriate scaling up of habitat features and environmental drivers through accurate models holds the promise of inference at larger scales, even when thresholds and other complex relationships are present.

## ACKNOWLEDGMENTS

We thank Dana Infante and her lab at Michigan State University, as well as Peaches DeWeber and Joshua Hickey at the Pennsylvania State University, for preparing the land use data. We thank David Miller for discussions and comments that improved this manuscript and Ephraim Hanks for discussions about model parameterization. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## LITERATURE CITED

- Allan, J. D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35:257–284.
- Betts, M. G., G. J. Forbes, and A. W. Diamond. 2007. Thresholds in songbird occurrence in relation to landscape structure. *Conservation Biology* 21:1046–1058.
- Betts, M., J. C. Hagar, J. W. Rivers, J. Alexander, K. McGarigal, and B. McComb. 2010. Thresholds in forest bird occurrence as a function of the amount of early-seral broadleaf forest at landscape scales. *Ecological Applications* 20:2116–2130.
- Brenden, T. O., L. Wang, and Z. Su. 2008. Quantitative identification of disturbance thresholds in support of aquatic resource management. *Environmental Management* 42:821–832.
- Cheruvilil, K., P. Soranno, K. Webster, and M. Bremigan. 2013. Multi-scaled drivers of ecosystem state: quantifying the importance of the regional spatial scale. *Ecological Applications* 23:1603–1618.
- DeWeber, J. T., and T. Wagner. 2014. A regional neural network ensemble for predicting mean daily river water temperature. *Journal of Hydrology* 517:187–200.
- DeWeber, J. T., and T. Wagner. *In press*. Predicting brook trout occurrence in stream reaches throughout their native range in the eastern United States. *Transactions of the American Fisheries Society*.
- Esselman, P. C., D. M. Infante, L. Wang, D. Wu, A. R. Cooper, and W. W. Taylor. 2011. An index of cumulative disturbance to river fish habitats of the conterminous United States from landscape anthropogenic activities. *Ecological Restoration* 29:133–151.
- Estavillo, C., R. Pardini, and P. L. B. da Rocha. 2013. Forest loss and the biodiversity threshold: an evaluation considering species habitat requirements and the use of matrix habitats. *PLoS ONE* 8:e82369.
- Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York, New York, USA.
- Gevrey, M., F. Sans-Piche, G. Grenouillet, L. Tudesque, and S. Lek. 2009. Modeling the impact of landscape types on the distribution of stream fish species. *Canadian Journal of Fisheries and Aquatic Sciences* 66:484–495.
- Gu, W., and R. K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation* 116:195–203.
- Heffernan, J. B., P. A. Soranno, M. J. Angilletta, Jr, L. B. Buckley, D. S. Gruner, T. H. Keitt, J. R. Kellner, J. S. Kominoski, A. V. Rocha, J. Xiao, et al. 2014. Macrosystems ecology: understanding ecological patterns and processes at continental scales. *Frontiers in Ecology and the Environment* 12:5–14.
- Higgins, J. V., M. T. Bryer, M. L. Khoury, and T. W. Fitzhugh. 2005. A freshwater classification approach for biodiversity conservation planning. *Conservation Biology* 19:432–445.
- Hudy, M., T. M. Thieling, N. Gillespie, and E. P. Smith. 2008. Distribution, status, and land use characteristics of subwatersheds within the native range of brook trout in the eastern United States. *North*

- American Journal of Fisheries Management 28:1069–1085.
- Jones, J. E., A. J. Kroll, J. Giovanini, S. D. Duke, and M. G. Betts. 2011. Estimating thresholds in occupancy when species detection is imperfect. *Ecology* 92:2299–2309.
- Kimura, D. 2008. Extending the Bertalanffy growth model using explanatory variables. *Canadian Journal of Fisheries and Aquatic Sciences* 65:1879–1891.
- Kuhn, M., and K. Johnson. 2013. *Applied predictive modeling*. Springer, New York, New York, USA.
- Levy, O., B. A. Ball, B. Bond-Lamberty, K. S. Cheruvilil, A. O. Finley, N. R. Lottig, S. W. Punyasena, J. Xiao, J. Zhou, L. B. Buckley, et al. 2014. Approaches to advance scientific understanding of macrosystems ecology. *Frontiers in Ecology and the Environment* 12:15–23.
- MacCrimmon, H. R., B. L. Gots, and J. S. Campbell. 1971. World distribution of brook trout, *Salvelinus fontinalis*: further observations. *Journal of the Fisheries Research Board of Canada* 28:452–456.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- Mackey, B. G., and D. B. Lindenmayer. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography* 28:1147–1166.
- Pimm, S. L., and P. Raven. 2000. Biodiversity: extinction by numbers. *Nature* 403:843–845.
- Plummer, M. 2013. *rjags*: Bayesian graphical models using MCMC. <http://sourceforge.net/projects/mcmc-jags/>
- Qian, S. 2012. Analytical options for estimating ecological thresholds—statistical considerations. *In* R. Gitzen, J. Millsaugh, A. Cooper, and D. Licht, editors. *Design and analysis of long-term ecological monitoring studies*. Cambridge University Press, Cambridge, UK.
- Qian, S. S. 2014. Ecological threshold and environmental management: A note on statistical methods for detecting thresholds. *Ecological Indicators* 38:192–197.
- Qian, S. S., and T. F. Cuffney. 2012. To threshold or not to threshold? That's the question. *Ecological Indicators* 15:1–9.
- Qian, S. S., T. F. Cuffney, I. Alameddine, G. McMahon, and K. H. Reckhow. 2010. On the application of multilevel modeling in environmental and ecological studies. *Ecology* 91:355–361.
- R Core Team. 2014. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reunanen, P., M. Mönkkönen, A. Nikula, E. Hurme, and V. Nivala. 2004. Assessing landscape thresholds for the Siberian flying squirrel. *Ecological Bulletins* 51:277–286.
- Rüegg, J., C. Gries, B. Bond-Lamberty, G. J. Bowen, B. S. Felzer, N. E. McIntyre, P. A. Soranno, K. L. Vanderbilt, and K. C. Weathers. 2014. Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment* 12:24–30.
- Soranno, P. A., K. S. Cheruvilil, E. G. Bissell, M. T. Bremigan, J. A. Downing, C. E. Fergus, C. T. Filstrup, E. N. Henry, N. R. Lottig, E. H. Stanley, et al. 2014. Cross-scale interactions: quantifying multi-scaled cause-effect relationships in macrosystems. *Frontiers in Ecology and the Environment* 12:65–73.
- Stranko, S. A., R. H. Hilderbrand, R. P. Morgan, M. W. Staley, A. J. Becker, A. Roseberry-Lincoln, E. S. Perry, and P. T. Jacobson. 2008. Brook trout declines with land cover and temperature changes in Maryland. *North American Journal of Fisheries Management* 28:1223–1232.
- Suding, K. N., and R. J. Hobbs. 2009. Threshold models in restoration and conservation: a developing framework. *Trends in Ecology & Evolution* 24:271–279.
- Thomson, J. R., W. J. Kimmerer, L. R. Brown, K. B. Newman, R. M. Nally, W. A. Bennett, F. Feyrer, and E. Fleishman. 2010. Bayesian change point analysis of abundance trends for pelagic fishes in the upper San Francisco Estuary. *Ecological Applications* 20:1431–1448.
- U.S. Geological Survey. 2008. Multi-resolution land characteristics consortium: National land cover database. <http://www.mrlc.gov>
- Wagner, T., J. T. Deweber, J. Detar, D. Kristine, and J. A. Sweka. 2014. Spatial and temporal dynamics in brook trout density: implications for population monitoring. *North American Journal of Fisheries Management* 34:258–269.
- Wagner, T., J. T. Deweber, J. Detar, and J. A. Sweka. 2013. Landscape-scale evaluation of asymmetric interactions between Brown Trout and Brook Trout using two-species occupancy models. *Transactions of the American Fisheries Society* 142:353–361.
- Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying threats to imperiled species in the United States. *BioScience* 48:607–615.

SUPPLEMENTAL MATERIAL

SUPPLEMENT

BUGS code for implementing the hierarchical spatially varying threshold model described in the text (*Ecological Archives* <http://dx.doi.org/10.1890/ES14-00288.1.sm>).