Accounting for variation in species detection in fish community monitoring

R. A. MCMANAMAY

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

D. J. ORTH

Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA, USA

H. I. JAGER

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Abstract Long-term fish assemblage monitoring requires investigators account for within-year variation in species' detection. An occupancy modelling framework is presented that accounts for variation in species presence and estimates the effort required to minimise within-year variation. Species detections from snorkelling surveys and an electrofishing survey were used in single-species occupancy models to determine the importance of site and sampling covariates on species' occupancy (*psi*) and detection probabilities (*p*). Community occupancy models estimated species richness in light of undetected species. For most species, models including patch size and reach as *psi*-covariates had higher support whereas models including patch size and sampling method as *p*-covariates had higher support. The number of sites and repeated surveys required to estimate occupancy accurately varied among fish species. Community models suggested that the observed number of species underestimated actual richness as much as 27% and the addition of an electrofishing survey increased estimated richness by 13%.

KEYWORDS: detection probability, habitat heterogeneity, occupancy modelling, sampling design.

Introduction

Assessing fish community composition and richness is a fundamental, yet challenging, task for fisheries management. Riverine fish assemblages are heterogeneous, exhibiting variation across space (Jackson *et al.* 2001), time (Grossman *et al.* 1982, 1985; Freeman *et al.* 1988) and in response to environmental fluctuations (Schlosser 1985; Poff & Allan 1995; Bernardo *et al.* 2003). Although landscape-level variation (e.g. geology, climate) tends to structure fish community composition among major river systems (Marsh-Mathews & Mathews 2000; Jackson *et al.* 2001), variability in the occurrence of individual species within river systems may be more associated with local habitat heterogeneity (Gorman & Karr 1978; Schlosser 1982). In addition, the ability to perceive a species presence or absence accurately is influenced by sampling methodology (i.e. gear type; Bayley & Peterson 2001; Brewer & Ellersieck 2011), sampling effort (Angermeier & Smogor 1995), species (Peoples & Frimpong 2011), habitat preferences (Pierce *et al.* 1990; Hayer & Irwin 2008) and environmental conditions (Falke *et al.* 2010). Accurately assessing fish communities requires investigators to account for sampling and habitat influences on species' occurrences.

In general, no single sampling method can ensure that all fish species are detected (Hayes *et al.* 1996; Hubert 1996). Consequently, estimates of species presence and community richness are imperfect. Using presence/absence data rather than abundance data can be a cost-effective and accurate approach to monitoring species of concern (Joseph *et al.* 2006). In addition, occupancy (the

proportion of sites or area occupied by a species) may provide an improved approach to evaluate population trends than surveys that estimate abundance (see Witmer 2005) because presence/absence data are easier to obtain and more robust to differences in capture efficiency. For example, occupancy estimates provide a framework to assess temporal trends in fish community samples despite non-standardised past sampling methods. Thus, occupancy modelling can take advantage of heterogeneous data. Furthermore, occupancy provides an approach to assess distributions in rare or endemic fish species with little to no information on population demographics. However, elucidating patterns in species occupancy requires accounting for detection probability, that is, the probability of detecting a species given is present (MacKenzie *et al.* 2006).

Accounting for differences in detection probability is required to estimate species occupancy accurately or to provide accurate estimates of species richness to support management decisions (Boulinier *et al.* 1998; Bayley & Peterson 2001; Albanese *et al.* 2007; Wenger & Freeman 2008). Long-term monitoring aimed at detecting temporal trends in fish species presence must account for within-year variation due to local habitat heterogeneity and sampling method. In addition, determining the effort required to provide accurate baseline data should be conducted prior to initiating monitoring programmes. However, designing sampling strategies may be a balance between statistical rigour and the practicality of sampling when and where it is possible. For example, a statistically rigourous sampling scheme might include randomly stratified sampling locations where the same methodology is employed throughout (Jolly & Hampton 1990). The convenience of homogenous methodology includes easier replication, less equipment, less field crew training and standardised fish assemblage metrics. However, single methods will likely preclude sampling or be ineffective at sampling in some habitat types. In addition, randomly stratified sampling locations, although representative of the dominant habitats, may fail to include rare habitats containing patchily distributed species. Realistically, most sampling approaches do not ensure that capture efficiencies are consistent across all habitat types.

Occupancy modelling approaches can account for variation in sampling conditions and methods by determining the influence of site and sampling covariates on species detection probability, species occupancies and community richness. In addition, occupancy modelling approaches can be used to determine the effort required to minimise variation in species occupancy. The goal of this study was to use an occupancy modelling approach to understand what covariates influence species detection

and occupancy, and to estimate fish richness in light of unobserved species. A dual methodology sampling design was employed, which consisted of multiple snorkelling surveys followed by an electrofishing survey in a suite of delineated habitats within a regulated river system. Spatial variation in species' detection probabilities and patch occupancy could be influenced by one or a combination of site covariates (habitat depth, slope, distance from dam and patch size) and sampling covariates (method and flow). Alternative models were developed for a number of individual fish species, and support was evaluated based on different combinations of covariates. At the community level, there is no guarantee that all species are sampled, and fish richness is likely to be underestimated (Bayley & Peterson 2001). Therefore, community occupancy models were used to estimate richness and the number of undetected species. In addition, an assessment of how additional sampling sites and adding a different method (i.e. electrofishing) can improve estimates of fish richness is provided. The approach can be used to determine the effort required to estimate occupancy estimates and community richness values.

Methods

Study site

The Cheoah River is located in the Blue Ridge physiographical province in western NC, USA (Fig. 1a). The Cheoah River is impounded by Santeetlah Dam and runs 14.6 km before emptying into the Little Tennessee River System. The watershed is predominately forested, has steep hillslopes (~30% grade) and receives 150 to 230 cm of precipitation annually. The Cheoah River is a high-gradient system, falling from 533 m at the dam to less than 335 m over its length (~1.3%). Geology is dominated by gneiss, sandstone and granite.

Geomorphology and gradient change considerably in the Cheoah River from the dam to the mouth (McManamay *et al.* 2010). The upper 3.2 km of the river, which is dominated by bedrock and large boulders (D_{50} = 370 mm), has a relatively low gradient (0.3–0.6%, measured over 1-km distance with Digital Elevation Models). The next 7.8 km of river generally has a steeper gradient (1–2%) with increasing amounts gravel and cobble substrates. The remaining stretch of river begins with moderate gradients (0.6–1%) for the first 1.4 km and then gradient increases to 2–3% for the remaining 2 km until the junction with the Little Tennessee below Cheoah Dam. Despite some increases in finer material, the streambed substrates in the

(a)

(b)
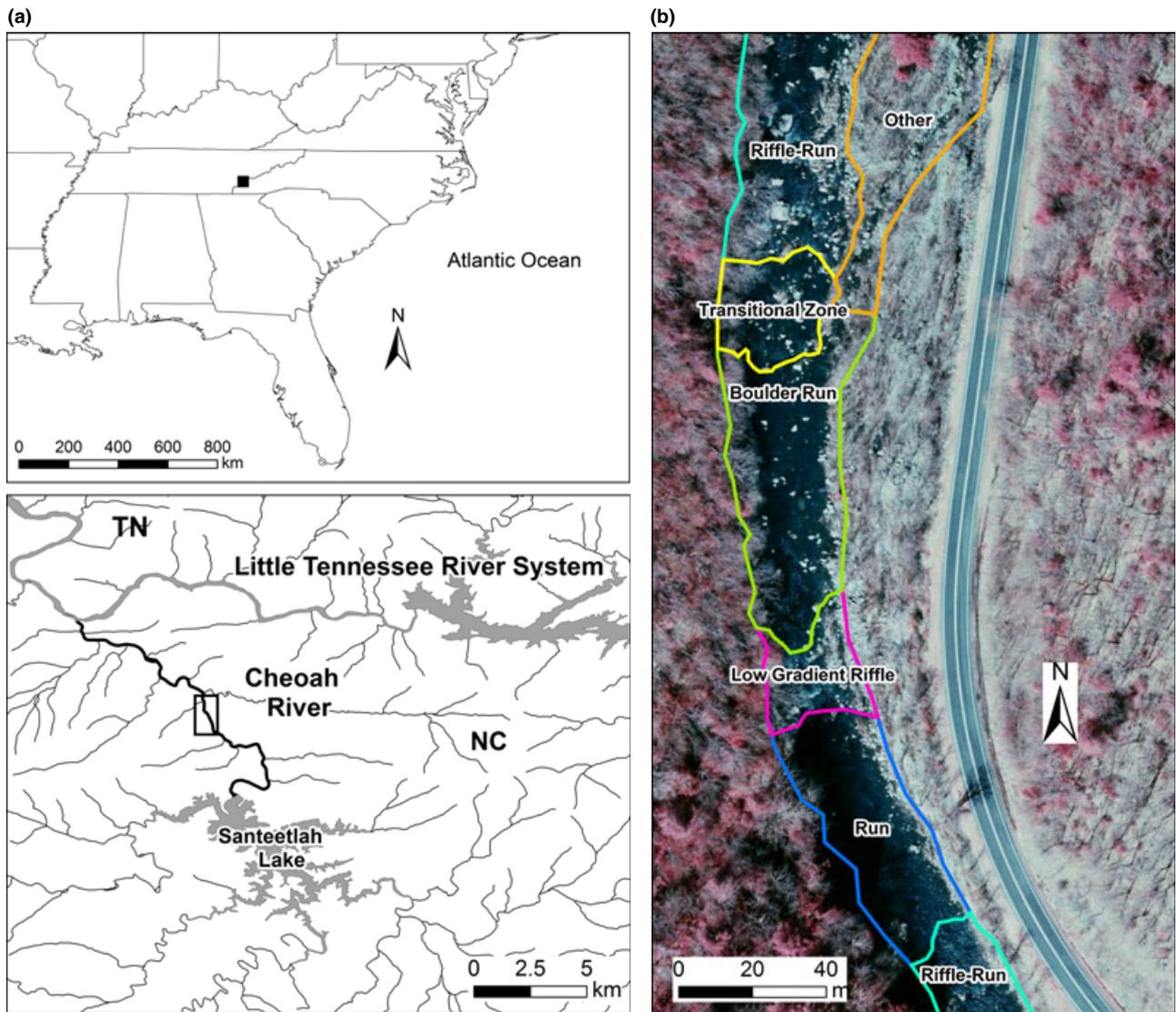
Figure 1. (a) Study map of the Cheoah River, a regulated tributary of the Little Tennessee River system in western North Carolina, USA. Box highlights reach of the Cheoah River captured in inset B. (b) Example of mesohabitats used to calculate patch occupancy.

lower stretches of the river are still very coarse (D_{50} = 230 mm).

Site selection and occupancy framework

Fish sampling sites were selected that overlapped with sampling locations from previous studies (Normandeau Associates 2002) and fell into different mesohabitat types. Mesohabitats, or areas of similar geomorphology, gradient and substrate, were delineated for the entire Cheoah River using high-resolution geospatial data sets (Fig. 1b). In January 2005, during leaf-off conditions and baseflows, aerial flights over the entire study area captured high-resolution [6 inch (15 cm)] colour-infrared

(CIR) digital orthophotography (scale 1:1200), high-resolution (1 m) hyperspectral imagery (scale 1:7500) and Light-Detection-and-Ranging (LIDAR) data. These data were georeferenced and interpolated to create a continuous 1-ft [30 cm] resolution digital elevation model (DEM). One-foot contour lines and slope values were derived from the DEM data. Vegetation and substrate classifications were developed using ground-truthing observations to train and verify the classification trees (CART modelling) to classify vegetation and substrate using colour bands from CIR and hyperspectral imagery. Aerial images, slope and vegetation/substrate classifications were used to delineate the Cheoah River into 233 mesohabitats, ranging in length from 20 to 300 m. After

delineation, field crews surveyed mesohabitats to ground-truth classification accuracy.

Mesohabitats ranged represented 12 different habitat types (low-gradient riffles, high-gradient riffles, runs, riffle-runs, bedrock-ledge runs, boulder runs, transitional areas, pools, pool edge, backwater, cascades and inundated vegetation). In total, there were 233 mesohabitats delineated across the entire course of the Cheoah River (excluding inundated vegetation and stagnant backwater areas). Pool edges and pools were combined into composite habitats. Transitional areas were combined with the nearest upstream mesohabitats. Because mesohabitats represent areas of similar morphology and habitat (i.e. patches), they provided a framework to calculate patch occupancy and detection probability.

A subset of 41 mesohabitats were selected for fish sampling. One sampling site was used to assess fish presence or absence within each mesohabitat (i.e. patch). On average, the area searched at each site represented over 50% of the area within its respective mesohabitat. However, this is likely an underestimate because areas calculated for mesohabitats were influenced by their boundaries extending outside the available wetted area (see Fig. 1b). The subset of mesohabitats sampled represented approximately 30% of the total area comprised in the Cheoah River from the dam to the mouth. Thus, the area actually searched at all sites was estimated to represent at least 15% of the total area in the 14.6 km stretch of the Cheoah River.

Fish sampling

Past sampling methods From 1993 to 1999, North Carolina Wildlife Resources Commission (NCWRC) biologists used 3-pass depletion methods with up to 10 backpack electrofishing units employed simultaneously. Due to variability in flow, smaller field crews and the increasing awareness of IBI methodologies, NCWRC discontinued 3-pass depletion in 2004–2005 in favour of IBI electrofishing approaches that targeted a larger number of different habitat types; however, similar to 3-pass depletion, this sampling was limited to wadeable habitats. After a reconnaissance visit by the US Forest Service (USFS) in 2006, snorkelling was recommended as a preferred option due to deep habitats, high water clarity and swift currents (Dolloff *et al.* 2006). Following a pilot study in 2008, repeated snorkelling surveys were determined to be the most efficient and effective sampling regime and provided a measure of species detection probabilities. To provide a validation assessment and comparable results to previous methods, a modified Index of Biotic Integrity (IBI) approach was also employed using seines and backpack electrofishing equipment (Karr 1981; Karr *et al.* 1986).

Current sampling method All sites were snorkelled whereas backpack electrofishing was conducted in sites where wading was feasible and allowed efficient sampling (mean depth, 1 m). Typically, the length of river searched at each site (mean $\pm$ SD, 50 $\pm$ 24 m) was proportional to the entire area represented by each mesohabitat to ensure that variation in microhabitats was captured. Between 16 June – 2 July 2009, 41 sites were snorkelled three times, each separated, on average, by 3 days. Following the last snorkel survey, 16 of those sites were electrofished.

Practice snorkelling runs were made prior to actual sampling to ensure fish identification was accurate. An experienced diver accompanied less-experienced crew members during trial surveys to ensure correct identification. Snorkel sampling was conducted during optimal daylight hours to maximise visibility. The entire wetted margin of the river at each site was split into three lanes, and one diver was assigned to each lane. Searching was conducting in an upstream direction, and pace was altered depending on depth, water clarity and habitat complexity. The diver in the middle lane was responsible for keeping the pace (2–3 m min^{-1}), and divers in the outside lane were responsible for staying parallel with the middle divers. Divers called species, abundances and approximate sizes to a recorder on the stream bank. Stream wetted width ranged from 11 to 47 m. Because the range of underwater visibility was typically >6 m, divers could easily observe the entire space within their lane and remain parallel to the middle diver. Because shiner species, *Cyprinella* and *Luxilus* spp., would aggregate behind divers and smallmouth bass, *Micropterus dolomieu* Lacépède, would swim ahead, care was taken not to duplicate counts. Following each survey, divers would discuss and remove likely duplicates with the recorder. Young-of-the-year were enumerated and identified to the lowest taxonomic resolution possible.

At sites consisting of shallower water depths (mean depth, 1 m) and swifter currents, backpack electrofishing was conducted. Sampling was conducted by stunning downstream into a seine or stunning upstream along the shoreline. In the main body of the channel, three crew members would hold a 6-m seine perpendicular to the current. The backpack operator would start roughly 6 m upstream of the seine net and move in a downstream direction and from side to side. Fish were stunned and carried by the current into the seine or would be netted by the operator. One unit of effort was defined as a 6 × 6 m sampling grid. Grid sampling continued until three consecutive samples yielded no new species. One unit effort of shoreline sampling consisted stunning upstream within a distance of 2 m from the shoreline

with two crew members following with dip nets. At each site, one electrofishing survey consisted of at least six grid efforts and at least two shoreline efforts and attempted to approximate the area sampled by snorkelling. All fish were identified to species and counted.

Single-species occupancy models and detection probability estimation

Occupancy modelling was used to predict the presence/absence of each fish species in all 41 sites in the program PRESENCE (USGS 2012). PRESENCE provides estimates of occupancy (*psi*), detection probability (*p*) and covariate parameter values. PRESENCE is flexible enough to allow for missing observations, that is, sites not sampled on one or more sampling surveys (MacKenzie *et al.* 2002). Thus, missing observations supply no information regarding the detection or non-detection of the species (Hines & MacKenzie 2012).

Because smallmouth bass and whitetail shiner, *Cyprinella galactura* Cope, were present at all sites, these species did not require a statistical modelling approach to estimate *psi*. However, because these species were not detected during all surveys, *p* was estimated.

An important goal of this study was to understand what covariates influence species detection and occupancy. It was hypothesised that *psi* could be influenced by one or a combination of site covariates (habitat depth, distance from dam, slope and patch size) whereas *p* could be influenced by sampling covariates (method and flow) and at least one site covariate (patch size). To assess the effect of habitat depth, sites were classified as shallow or deep (1 or 0, respectively) to covary with occupancy [e.g. *psi*(habitat type)]. Average water depth at each site was estimated using a measuring rod. Shallow habitats were classified as having average water depths <1 m, on average. Average slope (based on 1-ft DEM values) was calculated for each mesohabitat. Slope was modelled as a covariate of *psi* [e.g. *psi*(slope)]. In regulated systems, proximity to source populations, habitat heterogeneity and morphological character all vary along the river's longitudinal profile and with distance from dam. Thus, distance from dam to all sites was calculated to covary with *psi* [e.g. *psi*(dist)]. River discharge varied little across surveys; however, flow was included as a covariate of *p* [e.g. *p*(flow)]. In addition, patch size was included as a covariate for *psi* and *p* because the presence of a species and the ability to detect a species may depend on the amount of contiguous habitat within an area. Sampling method (snorkelling vs electrofishing) was included as a covariate of *p*.

All presence/absence data were used in PRESENCE to model the effects of site and sampling covariates on *psi* for all species with naïve occupancy rates <0.95. Naïve occupancy refers to the proportion of occupied sites out of the total number of sites, without accounting for *p* < 1 (a site was classified as occupied if the species was present on at least one survey). Models for each species were constructed as follows. The initial base model included constant *psi* and constant *p* [*psi*(.), *p*(.)]. It was hypothesised that all *psi*- and *p*-covariates could be important in modelling occupancy for each species. Thus, models were constructed by determining whether each *psi*- and *p*-covariate independently improved model performance over the base model (model performance defined below). Individual covariates that improved model performance over the base model were retained and then used in all possible combinations to determine best models.

Models were ranked using corrected Akaike's Information Criterion (AIC$_c$) or quasi-corrected-Akaike's Information Criterion (QAIC$_c$), depending on the evidence of overdispersion (Burnham & Anderson 2002). Goodness-of-fit and extent of overdispersion of residuals were evaluated for all models using Pearson's chi-squared tests and variance inflation factor (ĉ), where ĉ = χ^2/d.f. (MacKenzie *et al.* 2006). Highest ranking models (best models or global models) with ĉ > 1 indicate overdispersion (Burnham & Anderson 2002). However, ĉ values ranging from 1 to 3 are common (Burnham & Anderson 2002). For species with best models showing signs of overdispersion (ĉ > 1), models were ranked using QAIC$_c$ whereas AIC$_c$ was used to rank models for all other species. Models with the lowest AIC$_c$ or QAIC$_c$ scores were considered to be more parsimonious, providing higher explanatory power relative to the number of parameters estimated (i.e. covariates). Models showing evidence of overparameterisation, such as optimisation routines failing to reach a maximum likelihood values for parameters, were removed from the analysis. AIC weights were also calculated for each model. To assess the relative importance of each covariate individually, W$_+$(j) was calculated as the sum of AIC weights across all models containing each covariate j. W$_+$(j) was only calculated for covariates improving model performance over the base model.

Estimating sampling effort

The final objective was to provide an approach to estimate the effort and sampling scheme required to monitor individual species and the entire fish assemblage accurately. Developing an efficient sampling protocol includes assessing the effects of methodology and effort, in terms of the number of sites and the number of surveys, on variation in *psi*. Thus, the most efficient

sampling protocol would be one that minimises the amount of effort required to maximise precision (i.e. minimise variation). In addition, understanding how many sites or surveys are required to estimate richness accurately is needed to design adequate sampling strategies.

Single-species occupancy Values of *psi* and *p* for individual species were used to explore the effects of effort (number of sites and surveys) and method on variance in occupancy estimates following methods provided by MacKenzie *et al.* (2006). Assuming *p* is constant, variance in *psi* [*Var(psi)*] can be expressed as:

$$Var(psi) = \frac{psi}{s} \left[(1 - psi) + \frac{(1 - p*)}{p * - Kp(1-p)^{K-1}} \right] \quad (1)$$

where $p* = 1 - (1-p)^K$ is the probability of detecting the species at least once during *K* surveys of an occupied site. The equation above is typically utilised to determine sampling effort required prior to the onset of field work; thus, *psi* and *p* typically represent initial estimates. However, estimates of *psi* and *p* from occupancy models were used in the equation. The *psi* estimate was obtained from the best model for each species. Because separate *p* values for snorkelling and electrofishing was desired for all species, estimates of *p* were obtained from *psi*(.), *p*(method) models. Six species were selected that represented a range in *psi* and *p* values and evaluated changes in *Var(psi)* with increasing number of sites for four different scenarios: (1) two snorkel surveys (2s), (2) three snorkel surveys (3s), (3) four snorkel surveys (4s) and (4) three snorkel surveys and one electrofishing survey (3s+e). Equation 1 assumes that *p* is constant; however, for a given species, *p* varies between snorkelling and electrofishing. Thus, for the 3s+e scenario, detection probabilities for snorkelling (p_s) and electrofishing (p_e) could be combined as weighted averages across four surveys, where $p_{3s+e} = 0.75\, p_s + 0.25 p_e$. The relationship between variance estimates and the number of sites/surveys was evaluated.

Community fish richness A hierarchical Bayesian model was used to estimate fish richness both with and without a final electroshocking survey following an approach by Dorazio *et al.* (2010). Repeated surveys (detection histories) of all species observed in the community are used to model the probability of the number of species potentially present but undetected (Dorazio *et al.* 2006). One problem of estimating fish species richness, *N,* is that only the number of observed species is known *n,* and thus, the number of undetected species that are present, *N−n*, must be estimated. This requires establishing an upper ceiling estimate of the total number of species that could be potentially present despite being unobserved. Based on data from surrounding streams, tributaries and expert opinion, it was determined that 40 species could be present in the Cheoah River. Dorazio *et al.* (2010) suggested that total number of potential species that could be present, *M*, can be arbitrarily large if *N* is uncertain. Thus, it was assumed that *M* = 100 would provide a sufficient upper ceiling estimate. *M* has no effect on the estimate of *N*, but it influences the number of computations to fit the model (Dorazio *et al.* 2010). The number of undetected species is estimated by augmenting the observed data with an arbitrary number of all zero encounter histories, as illustrated by Royle *et al.* (2007) and Royle and Dorazio (2008). Following Dorazio *et al.* (2010), the total number of species, *N,* is a random variable:

$$N|M, \Omega \sim Binomial(M, \Omega) \quad (2)$$

where Ω is the probability that each species (from total *M* species) is a member of the community. The hierarchical Bayesian model is used to estimate a latent (unobserved) random variable, w_i, that indicates whether species *i* is a member of the community (Equation 3).

$$w_i|\Omega \overset{iid}{\sim} Bernoulli(\Omega) \quad (3)$$

The model of multispecies occurrences requires that changes in occupancy are conditioned on *w*. *N* is not a formal model parameter but is estimated by summing the w_i values over all species in the augmented set, that is, $N = \sum w_i$ for *i* = 1 to M.

The model description here is taken from Dorazio *et al.* (2006). Let psi_{ij} denote the probability of occurrence of species *i* at site *j* and let p_{ij} denote the probability of detection of species *i* given that it occurs at site *j*. Dorazio and Royle (2005) determined that the marginal probability of the observed number of detections is described by a zero-inflated binomial, a function of the total number of detections for each species and site (not detection history), *K* (the number of surveys), psi_{ij}, and p_{ij}. If species *i* is not detected, then it is either absent with probability $1 - psi_{ij}$, or present but undetected with probability $psi_{ij} (1 - p_{ij})^K$. In this case, *K* is either 3 (snorkelling only) or 4 (snorkelling plus electrofishing).

Differences in these probabilities are modelled as having a site-specific and a species-specific component as follows (Dorazio *et al.* 2006): logit $psi_{ij} = \mu_i + \alpha_j$ and logit $p_{ij} = v_i + \beta_j$, where species effects are represented by μ_i and v_i, and site effects are represented by α_j and β_j. The relationships between species and site effects on the psi_{ij} and p_{ij} are each represented by joint normal distributions, as shown in Equation 4, where the

variance–covariance matrix, $\sum$ has variances, $\sigma_\mu^2 \sigma_v^2$, for each parameter on the diagonal and the covariance between μ and v on the off diagonal ($\rho_{\mu v}$ is the correlation).

$$N(0, \Sigma), \Sigma = \begin{pmatrix} \sigma_\mu^2 & \rho_{\mu v}\sigma_\mu\sigma_v \\ \rho_{\mu v}\sigma_\mu\sigma_v & \sigma_v^2 \end{pmatrix} \quad (4)$$

Independent prior distributions on parameters were estimated following Dorazio *et al.* (2006). Priors for three parameters Ω, $\text{logit}^{-1}(\alpha)$ and $\text{logit}^{-1}(\beta)$ were uniform on the interval [0,1]. Inverse-gamma priors were specified for the two variance parameters from Equation 3. A uniform distribution on the interval [−1, 1] was used as a prior distribution for the correlation parameter, $\rho_{\mu v} = \sigma_{\mu v}/\sigma_\mu\sigma_v$.

Fish richness was estimated both with and without a final electrofishing survey. The approach presented by Dorazio *et al.* (2010) was extended to permit different total survey counts for different sites. Model code provided by Dorazio *et al.* (2010) was modified and ran in WinBUGS software (Gilks *et al.* 1994). Four Markov Monte Carlo chains (initial conditions) were simulated to ensure that estimated parameter values converged. For each chain, 505 000 iterations were simulated, discarding the first 5000. Results were sampled to obtain statistics for each parameter. Assuming β to be normally distributed, the expected detection probability | presence is given by Equation 5.

$$E[p_{ij}] = \frac{1}{1 + E[e^{-v_i - \beta_j}]} = \frac{1}{1 + E[e^{-v_i}]E[e^{-\beta_j}]}$$
$$= \frac{1}{1 + e^{-\left(E[v_i] + \frac{\sigma_v^2}{2}\right)}e^{-\left(E[\beta_j] + \frac{\sigma_\beta^2}{2}\right)}} \quad (5)$$
$$= \frac{1}{1 + e^{-\left(\frac{\sigma_v^2}{2}\right)}e^{-\left(E[\beta_j] + \frac{\sigma_\beta^2}{2}\right)}}$$

Posterior distributions for fish richness and detection probabilities for the two cases were compared. As a measure of prediction error, the deviance information criterion (DIC) and a measure of effective sample size is presented. Note that DIC values for the two models cannot be compared because they use different data sets. Finally, cumulative species richness vs the numbers of sites is also presented.

Results

During the 2009 sampling season, 17 species were observed. Fifteen species were detected during the three snorkel surveys and 10 species were detected during electrofishing. Of the ten species detected during electrofishing, two species were undetected during snorkelling (Table 1). Seven species (bluegill, *Lepomis macrochirus* Rafinesque, brown trout, *Salmo trutta* L., largemouth bass, *Micropterus salmoides* [Lacépède], tangerine darter, *Percina aurantiaca* [Cope], black redhorse, *Moxostoma duquesnei* [Lesueur], green sunfish, *Lepomis cyanellus* Rafinesque, and rainbow trout, *Oncorhynchus mykiss* [Walbaum]) were only found during snorkelling surveys whereas longnose dace, *Rhinichthys cataractae* (Valenciennes), and Tuckasegee darter, *Etheostoma blennoides gutselli* Rafinesque, were only found during electrofishing surveys (Table 1).

Single-species occupancy models and detection probability estimation

Best models for only two species, rock bass, *Ambloplites rupestris* (Rafinesque), and warpaint shiner, *Luxilus coccogenis* (Cope), showed signs of moderate overdispersion (Pearsons chi-squared, $P < 0.001$ and $P = 0.040$, respectively). Similarly, these were the only species with best models having ĉ values >2 (2.249 and 2.475, respectively). Of the remaining species, seven had ĉ < 1 and three with ĉ < 1.35. Tangerine darter, northern hog sucker, *Hypentelium nigricans* (Lesueur), and green sunfish had ĉ values of 1.695, 1.489, and 1.421, respectively.

Occupancy estimates (*psi*) for individual fish species ranged from 0.024 to 0.87 (Fig. 2). Detection probability (*p*) estimates varied from 0.18 to 0.94 (Fig. 2). Model building results and best models are provided in Appendix S1 (online version only). Occupancy models for all species, except rock bass and green sunfish, displayed increased support over the base model [*psi*(.), *p*(.)] when including both *psi*- and *p*-covariates (Table 2). Habitat depth was represented most frequently as an important covariate (eight species), followed by distance from dam (nine species), patch size (*psi*-covariate, eight species) and method (six species; Table 2). Likewise, habitat depth was most frequency represented in best models (six species), followed by method (six species) and distance from dam (six species; Appendix S1). Shallow habitats had positive effects on *psi* for river chub, *Nocomis micropogon* (Cope), warpaint shiner, greenfin darter, *Etheostoma chlorobranchium* Zorach, and central stoneroller, *Campostoma anomalum* Rafinesque (Table 2). By contrast, shallow habitats had negative effects on *psi* for northern hog sucker, bluegill, black redhorse, largemouth bass and tangerine darter. Five species (primarily fluvial specialists) showed evidence of increasing *psi* with distance from the dam whereas three centrarchids and tangerine darter displayed decreasing *psi* with

Table 1. Proportion of sites occupied by fish species during each snorkelling and electrofishing occasions in 2009. Proportions do not represent cumulative values for each successive occasion. No values indicate species was not detected

Species	Species Code	Snorkel 1	Snorkel 2	Snorkel 3	Electrofish
Number of sites sampled		41	41	41	16
Total Species		14	14	14	10
Centrarchidae					
Rock bass *Amploplites rupestris*	RKB	0.37	0.39	0.37	0.50
Green sunfish *Lepomis cyanellus*	GSF	0.02			
Bluegill *Lepomis macrochirus*	BLG	0.10	0.10	0.15	
Smallmouth bass *Micropterus dolomieu*	SMB	0.98	0.98	1.00	0.56
Largemouth bass *Micropterus salmoides*	LMB	0.12	0.02	0.12	
Catostomidae					
Northern hogsucker *Hypentelium nigricans*	NHS	0.34	0.37	0.27	0.06
Black redhorse *Moxostoma duquesnei*	BRH	0.05	0.07	0.07	
Cyprinidae					
Central stoneroller *Campostoma anomalum*	CSR	0.15	0.17	0.12	0.06
Whitetail shiner *Cyprinella galactura*	WTS	0.95	0.98	0.90	0.88
Warpaint shiner *Luxilus coccogenis*	WPS	0.73	0.80	0.76	0.63
River chub *Nocomis micropogon*	RVC	0.54	0.68	0.49	0.75
Longnose dace *Rhinichthys cataractae*	LND				0.06
Percidae					
Greenfin darter *Etheostoma chlorobranchium*	GFD	0.37	0.39	0.41	0.63
Tuckasegee darter *Etheostoma gutselli*	TKD				0.13
Tangerine darter *Percina aurantiaca*	TND	0.15	0.20	0.12	
Salmonidae					
Rainbow trout *Oncorhynchus mykiss*	RBT	0.07	0.02	0.05	
Brown trout *Salmo trutta*	BNT		0.02	0.02	
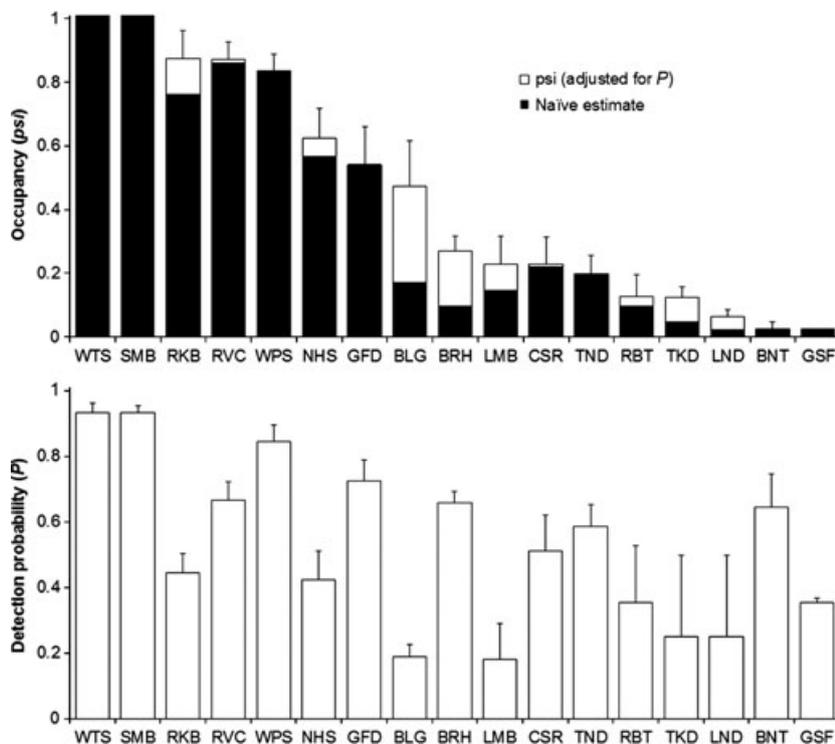
Figure 2. Naïve and adjusted occupancy (*psi*) estimates (top) and detection probability (*p*) estimates (bottom) (mean + 1SE) for fish species found in the Cheoah River during all surveys. Estimates were obtained from best models for each species. Species codes are presented in Table 1.

Table 2. Relative importance of covariates in occupancy models

Species	p-covariates			psi-covariates			
	Flow	Method	Patch Size	Habitat <1 m	Slope	Patch Size	Distance
Rock bass	—	—	—	—	—	—	—
River chub	0.449 (−)	—	—	0.537 (+)	0.415 (+)	—	—
Warpaint shiner	—	0.652 (++,+)	—	0.551 (+)	—	0.563 (−)	0.494 (+)
Northern hog sucker	—	0.523 (+,−)	—	—	0.974 (−)	—	—
Greenfin darter	0.542 (−)	—	—	0.982 (+)	0.992 (+)	—	0.978 (+)
Bluegill	—	0.565 (+,−)	0.947 (+)	0.809 (−)	—	0.487 (+)	0.967 (−)
Black redhorse	—	—	0.814 (+)	0.186 (−)	0.996 (−)	0.185 (−)	0.175 (+)
Largemouth bass	0.315 (+)	—	0.404 (+)	0.592 (−)	—	0.171 (+)	0.939 (−)
Central stoneroller	—	0.775 (+,−)	—	0.873(+)	—	—	—
Tangerine darter	—	—	—	—	—	0.342 (+)	0.893(−)
Rainbow trout	—	—	—	—	—	—	0.360 (+)
Tuckasegee darter	0.073 (−)	0.791 (−,+)	—	0.084 (+)	—	0.076 (−)	0.537 (+)
Longnose dace	—	0.674 (−,+)	—	—	0.429 (+)	0.273 (−)	—
Brown trout	0.345 (−)	—	—	—	—	—	—
Green sunfish	—	—	—	—	—	0.614 (+)	0.130 (−)

Values represent $W_+(j)$, which was calculated as the sum of AIC weights across all models containing each covariate. $W_+(j)$ was only calculated for covariates improving model performance over the base model ($psi(.)$, $p(.)$). Variables that did not improve model performance over the base model were not included (—). Values in parentheses indicate the direction of parameter estimates for each covariate, some of which have multiple levels (++ indicates higher positive value). Method (snorkel, shock); Habitat <1 m (presence of habitats with mean depth <1 m).

distance from dam (Table 2). Slope had positive effects on three riffle-dwelling species (river chub, greenfin darter, and longnose dace) but negative effects on northern hog sucker and black redhorse. Patch size, as a *psi*-covariate, had equally positive and negative effects.

Snorkelling had positive effects whereas electrofishing had negative effects on *p* for northern hog sucker, bluegill, central stoneroller, and tangerine darter. However, the opposite was true for longnose dace and Tuckasegee darter (Table 2). As a *p*-covariate, patch size displayed consistent positive effects and was supported in models for bluegill, black redhorse and largemouth bass (Table 2). Flow increased model support when included as a *p*-covariate for five species; however, $W_+(j)$ were typically lower than other covariates (Table 2). Flow consistently had negative effects on *p* for all species except for largemouth bass.

Estimating sampling effort

Single-species occupancy For species ($psi > 0.6$) with moderate *p* values (northern hog sucker, rock bass, river chub), additional surveys was relatively more important in reducing *Var(psi)* than for less common species with higher *p* values (black redhorse, central stoneroller, greenfin darter; Fig. 3). However, for these less common species, the addition of sites was relatively more important than the addition of surveys. Rock bass was the only species that showed noticeable reductions in *Var(psi)* with the inclusion of an electrofishing survey

as opposed to a fourth snorkel survey. For all species, *Var(psi)* reached an asymptote around 16–20 sites, that is, roughly half of the total effort (Fig. 3).

Community fish richness Community occupancy model parameters are provided in Table 3. The 80% prediction intervals for six parameters showed considerable overlap among the four MCMC chains in both the three-occasion and four-occasion analyses. This suggests that the number of iterations used in MCMC optimisation (i.e. parameter estimation) was sufficiently long for the chains to converge to a similar estimate of richness. DIC was 783.2 for the richness model based only on three snorkelling surveys and 921.1 for the analysis including snorkelling and electrofishing surveys. Detection probability across all sites and species was 0.8484 for the three-snorkelling surveys and 0.7554 for the snorkelling and electrofishing surveys. Median (mean) richness estimates were 22 (23.5) fish species for the analysis with only snorkelling surveys and 24 (25.8) for the analysis including the electrofishing survey, which exceeded the 17 species observed (Fig. 4, Table 3). The posterior distributions showed a slow increase in richness as the number of sites sampled increased and approached an asymptote after sampling 20 sites on four surveys (Fig. 5).

Discussion

Altogether, the results suggest that inherent variability due to site location and characteristics, methodology and
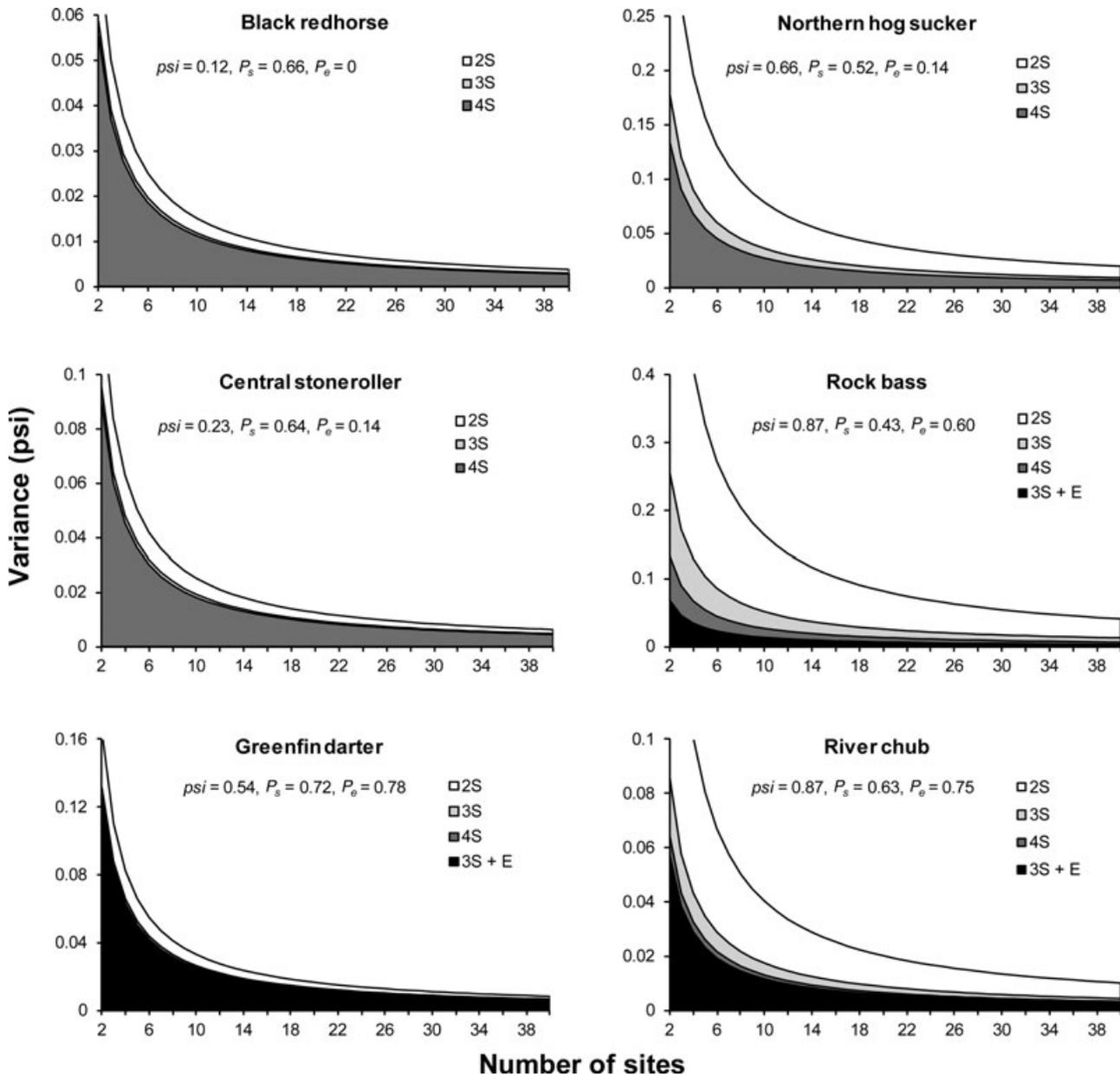
Figure 3. Variance in occupancy estimates for 6 species in relation to number of sites, number of surveys and method. Occupancy variance was calculated using an approach outlined by MacKenzie *et al.* (2006).

sampling conditions can have dramatically different effects on species detection in a moderately diverse cool-water river. When developing sampling protocols, ensuring site selection fully represents habitat heterogeneity or ensuring sampling is uniform in all habitats may be impractical given limited resources and management needs. A more realistic solution could utilise model-based frameworks to account for undetected species based on local habitat heterogeneity and sampling conditions. Occupancy frameworks may provide an improved

approach to fish assemblage monitoring by: (1) accounting for within-year variation in species' presences due to uncontrolled variables; (2) predicting the probability of species presence in patches despite being undetected; and (3) providing parameter estimates that can be incorporated into models that utilise data from past heterogeneous sampling regimes. One realistic, although non-optimal, alternative to conducting repeated surveys every year is to conduct repeated surveys during seasons when resources are available as to provide *psi-* and

Table 3. Posterior distribution statistics for estimated parameters in the multispecies (community) occupancy model

Surveys	Statistic	Alpha (α)	Beta (β)	Rho (ρ)	Sigma u (σ_μ)	Sigma v (σ_v)	Omega (Ω)	Richness (N)
Snorkelling only	Mean	−1.36	0.34	0.78	4.02	1.57	0.19	23
	SD	1.27	0.53	0.16	1.26	0.45	0.06	5.4
	2.5%	−4.23	−0.84	0.34	2.19	0.92	0.10	17
	25%	−2.10	0.05	0.71	3.11	1.26	0.15	20
	50%	−1.22	0.41	0.82	3.81	1.50	0.18	22
	75%	−0.49	0.71	0.89	4.67	1.80	0.22	26
	97.5%	0.72	1.20	0.96	7.08	2.68	0.32	37
Snorkelling & electrofishing	Mean	−1.79	−0.12	0.76	3.94	1.49	0.23	26
	SD	1.29	0.52	0.17	1.12	0.41	0.07	7
	2.5%	−4.85	−1.32	0.34	2.29	0.89	0.12	18
	25%	−2.46	−0.40	0.69	3.15	1.20	0.18	21
	50%	−1.60	−0.06	0.80	3.76	1.43	0.21	24
	75%	−0.91	0.24	0.88	4.56	1.71	0.26	28
	97.5%	0.24	0.72	0.95	6.56	2.47	0.40	45

Parameters include those used to model detection probability (logit $p_{ij} = v_i + \beta_j$), the probability of occurrence (logit $psi_{ij} = \mu_i + \alpha_j$), and the derived fish species richness, N.
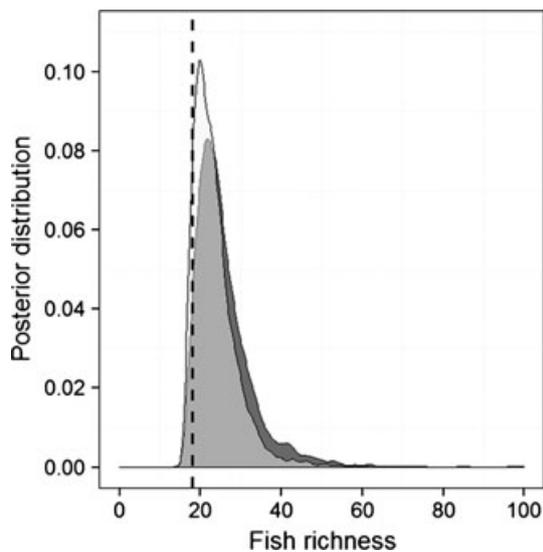

Figure 4. Comparison of the posterior distributions of estimated fish richness in the Cheoah River based on three snorkelling surveys (white fill) and all four surveys (dark grey), including three snorkelling and one electrofishing surveys. Overlap between the two distributions is shown in light grey, and the vertical dashed line shows the actual number of fish species detected.

p-covariates. The covariates can be used to adjust occupancy estimates from seasons when repeated surveys are not feasible.

Estimates of occupancy and detection probability varied considerably among 17 species in the Cheoah River. Habitat depth, slope, patch size and location (e.g. distance from dam) influenced species presence, whereas the ability to detect species was influenced by the search method and the size of area being searched (see also Bayley & Peterson 2001; MacKenzie *et al.* 2006).

Accurately assessing the composition of the fish community is needed in long-term monitoring schemes to provide evidence of colonisation, extinction and responses to habitat restoration. To elucidate temporal patterns in fish assemblages, consolidating variability in *psi* and *p* due to site and sampling factors is required. Two approaches were provided that can be used to inform sampling designs for monitoring individual species or estimating richness for fish communities. For individual species, an analytical approach is presented that can be used to estimate the number of sites and surveys required to minimise occupancy variation, which will vary by species. Multispecies occupancy models are required to estimate the number of undetected species and the level of effort needed to increase precision in richness estimates. Estimated species richness increased with the number of sites sampled (i.e. effort), meaning trade-offs between effort and sampling precision can be determined. Community richness values were moderately influenced by the addition of the electrofishing validation method (e.g. three more species), suggesting that snorkelling was sufficient at capturing the majority of species but may fail to represent rare benthic species. Thus, dual methods may be appropriate in many situations to assess all members of the fish community adequately (Brewer & Ellersieck 2011).

One limitation of the study was the amount of time (~3 days) that elapsed between surveys, which potentially breaches the 'closure' assumption (i.e. that emigration and immigration within a site does not occur during a season; MacKenzie *et al.* 2006). Understanding the minimum temporal window required for repeated observations for various animals, especially fish, has not been fully researched (Rota *et al.* 2009). While fish movement
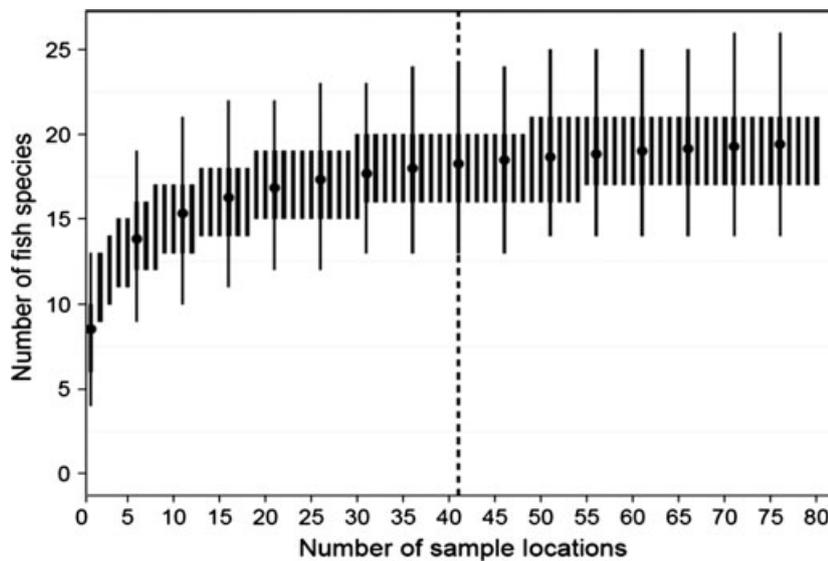
Figure 5. Species accumulation curves showing hypothetical rise in richness as more sites are sampled. For clarity, means and 90% prediction intervals are shown every 5 years, and the interquartile range is shown every year. The dashed vertical reference line indicates the number of sites sampled in this study.

across site boundaries within the sampling window was not explicitly tested, it was presumed the closure assumption was met for the following reasons. First, mesohabitats, on average, were quite large (length, 50 m) and most likely encompassed the movement by benthic fish within the sampling window. Specifically within the Little Tennessee River Basin, fish movement and habitat use vary considerably among seasons; however, fish typically display high habitat affiliation and low dispersal within summer months (Freeman & Grossman 1993). Pelagic species tend to move larger distances (Albanese *et al.* 2009); however, high gradients, including cascades, within the Cheoah River make longer distance movements between surveys unlikely. Furthermore, mark recapture studies of endemic fish reintroductions indicate that within-season fish movement in the Cheoah River is constrained by suitable habitats (Russ & Fraley 2009).

Another potential limitation of this study was that multiple electrofishing surveys were not conducted, but could have provided a more accurate representation of the fish community. A disadvantage of this approach is that multiple electrofishing surveys increase the likelihood of deleterious effects to fish (Snyder 2003); thus, sufficient time between surveys is required to allow fish to acclimate. In addition, if electrofishing induces mortality, then the closure assumption for occupancy modelling is not met. In occupancy models for fish, observational techniques (Brewer & Ellersieck 2011) or passive capture methods (Falke *et al.* 2010) may be the best approach for multiple surveys, which can then be followed by an electrofishing survey, as a validation.

Importance of covariates in occupancy models

Providing accurate estimates of species occupancy requires accounting for factors that influence p (Boulinier *et al.* 1998; Bayley & Peterson 2001; Albanese *et al.* 2007; Wenger & Freeman 2008) but also factors that covary with psi (MacKenzie *et al.* 2006;). Species p can differ according to season, habitat type and physiochemical characteristics (temperature, conductivity and turbidity; Hayer & Irwin 2008). Furthermore, p can differ for various species in relation to habitat structure in conjunction with sampling method (Korman *et al.* 2009). Developing sampling protocols that minimise site and sampling variation is an area of needed research.

Patch size can be manifested as an important psi-covariate in three main ways. First, organisms may need adequate space for feeding and resting. Thus, patch size was hypothesised to have positive effects on black redhorse psi because they are migratory and prefer aggregating in schools. However, the opposite was observed. Second, patch size may be a surrogate of habitat suitability and connectivity. Falke *et al.* (2010) found that habitat size was an important covariate in larval fish occupancy and suggested that habitat size thresholds may exist for suitable spawning to occur. By contrast, habitats that provide suitable microhabitat conditions may be rare and sparsely distributed, which may explain the negative effect of patch size on the presence of Tuckasegee darter and longnose dace (rare species with specific habitat needs). Third, patch size may be related to mesohabitats that meet unique habitat requirements. For example, homogenous run habitats were among the

largest mesohabitats whereas deep runs and riffles were smaller, which may explain why patch size had positive effects on sunfish and largemouth bass (prefer pools) but negative effects on warpaint shiner and black redhorse (prefer moderate/fast currents). Patch size can also be an important p-covariate because larger areas may create a dilution effect (Gaston et al. 2000; MacKenzie et al. 2006). However, patch size had entirely positive effects on p, which suggests one or a combination of the following scenarios: species presence in smaller patches is transient between surveys (closer assumption violation), smaller patches are more complex and harder to sample, and/or abundance may increase with patch size (Royle & Nichols 2003). Patch size has positive effects on p for three species: bluegill, black redhorse and largemouth bass. While the exact reason for positive patch size effects on p is uncertain and could be an artefact of closure assumption violations, the importance of patch size as a p-covariate in models for only 3/15 species indicates that closure assumption violations, if present, were not a widespread issue.

Sampling method will obviously influence detection probability (Bayley & Peterson 2001; Dauwalter & Fisher 2007; Korman et al. 2010; Brewer & Ellersieck 2011). Snorkelling has been shown to be an efficient sampling methodology (Hagen et al. 2010), comparable to electrofishing surveys in some streams for individual species (Mullner et al. 1998). Depending on the rarity of the species and associated management implications, follow-up sampling methods may be necessary to ensure species detection.

As would be expected, snorkelling tended to have higher p than electrofishing for more common species (Table 2), primarily because it was conducted during more surveys. However, electrofishing had higher p values than snorkelling for two rare benthic species (Table 2). Tuckasegee darters and longnose dace, both small benthic species with low psi estimates, were undetected during three snorkelling surveys but detected during a single electrofishing survey within the same habitats. Most studies suggest that snorkelling estimates are sufficient at detecting larger specimens, but are limited at detecting smaller benthic species, in which case electrofishing would be more appropriate (Ensign et al. 1995; Joyce & Hubert 2003; Korman et al. 2010). Because multiple snorkel surveys were conducted (i.e. higher effort) compared with only one electrofishing survey, direct comparisons of methods are limited. However, given lower effort via electrofishing, it is very reasonable to interpret higher p for electrofishing as more efficient at detecting some species than snorkelling.

Species' detection probabilities for a given method are likely to vary with habitat conditions (Dauwalter & Fisher 2007; Falke et al. 2010; Brewer & Ellersieck 2011). For example, multiple studies report low detection probabilities for catostomids when using boat electrofishing gear (Bayley & Austen 2002; Grabowski et al. 2009). This is probably because catostomids occupy deeper habitats, out of the reach of most electric fields (Grabowski et al. 2009). Thus, habitat preferences may play a larger role in influencing psi, rather than just p. For example, Burdick et al. (2008) found that juvenile Lost River sucker, Deltistes luxatus (Cope), occupancy was mainly associated with water depth and suggested depth may have biased gear efficiency.

Given that sampling method completely correlated with habitat depth in the analysis, the importance of habitat as a p-covariate was not explored. However, occupancy models did assess the importance of habitat depth as a psi-covariate. Habitat depth was hypothesised as an important covariate of multiple species, especially those known to display habitat preferences. As hypothesised, shallow habitats had a positive effect on common riffle dwellers and a negative effect on pool dwellers. The effects of slope on psi were synonymous with shallow habitats, indicating these variables were related, as would be expected; however, not all species with habitat-depth effects had slope effects. This indicates there is considerable variability among similar habitat types that influences species presence.

Longitudinal recovery of ecological function and fluvial fish communities with distance downstream from dams has been well documented (Ward & Stanford 1983; Travnichek & Maceina 1994; Travnicheck et al. 1995; Anderson et al. 2006). Likewise, generalist species are typically more prevalent in habitats directly downstream of dams (Moyle & Mount 2007). Distance downstream had positive effects on psi for five fluvial specialists, indicating signs of community recovery. As would be expected, three centrarchid generalists (bluegill, green sunfish and largemouth bass), displayed decreasing psi with distance from the dam. Interestingly, tangerine darter, a fluvial specialist, showed high affinity for near-dam habitats (negative effects of distance from dam on psi). Mesohabitats occupied by tangerine darter were predominately moderately deep runs with complex bedrock-ledges and large boulders. These habitats were more common closer to the dam.

Although five species displayed some influence of flow on p, ΔAIC weights were relatively low compared with other covariates most likely due to minimal variation in flow during the sampling season (between 3.26 and 3.29 $m^3\ s^{-1}$). Korman et al. (2009) found that flows made little difference in the detection of juvenile rainbow trout in the Colorado River regardless of backpack or boat electrofishing. The effect of flow on p was negative for 4/5 species suggesting a dilution effect and

lower detectability at higher flows. Flow was variable and did not change linearly with time across most sites; thus, the effects of flow are likely real and not an arte-fact of temporal autocorrelation. However, at a limited number of sites, including those containing Tuckasegee darter, flow progressively decreased from the first snor-kel survey to the electrofishing survey. Because Tuck-asegee darters were only detected during electrofishing, the importance of flow for this species is likely an arte-fact of sampling method, which was also an important covariate. However, for the other species, method was an unimportant covariate. Thus, it was presumed that higher flows lessened the ability to detect two benthic species (river chub and greenfin darter) and a less com-mon non-benthic species (brown trout). One plausible reason for the positive effects of flow on p for large-mouth bass was that higher flows may have induced behavioural changes or microhabitat use making this species more susceptible to detection.

Developing a sampling protocol

An important aspect of management is determining the trade-offs between expended effort and the accuracy level of environmental assessments (McDonald *et al.* 1996; Pollock *et al.* 2002; McDonald 2003). Managers may need to know whether a higher number of sites or a higher number of surveys are required to decrease error to some acceptable level (MacKenzie *et al.* 2002; Wintle *et al.* 2004). For the majority of species, variation in *psi* reached an asymptote following sampling at 20 sites (roughly 50% of the current effort); however, the num-ber of surveys and methods required to minimise varia-tion in *psi* tended to be variable among species. Albanese *et al.* (2007) found that almost 20 seine hauls were required at each site to have a p of 80% for blue-nose shiners, *Pteronotropis welaka* (Evermann and Ken-dall), whereas only half that amount were required to have $p > 80\%$ for other more common species. This suggests that for less common or rare species increasing sampling surveys, especially for a particular method (e.g. electrofishing), may be more important than the number of sites to estimate *psi* accurately.

Determining the effort and strategies needed to obtain accurate estimates is advantageous prior to implementing long-term monitoring programmes. The community occupancy models estimated that 24 species were present compared with only 17 observed species (30% underesti-mate). This suggests that estimates of fish richness based solely on observed species can substantially underesti-mate species richness. Similarly, Bayley and Peterson (2001) found that all species actually present were rarely captured in a given survey and the number of species

captured underestimated true richness, on average, by 27%. Based on community occupancy models, richness approached an asymptote following sampling at 20 sites, which mirrors the single-species analysis where the occupancy variation minimised at 20 sites. The posterior richness estimates between the two models suggested that the addition of an electrofishing survey increased richness from 23 to 26 species (13% increase). Based on the analysis, both snorkelling and electrofishing may be required to estimate fish assemblage composition and richness accurately. The addition of the electrofishing survey increased the estimated richness because previ-ously undetected rare species were detected. However, the average detection probability decreases from 0.8484 to 0.7554. For most common species, electrofishing had a lower p than snorkelling; thus, the addition of electro-fishing decreased the composite p value despite increas-ing richness estimates.

Conclusions

Various approaches are available to assess species occu-pancy (MacKenzie *et al.* 2002), abundance (MacKenzie & Kendall 2002; Wenger & Freeman 2008) and compar-isons of data derived from various sampling methodolo-gies (Cao *et al.* 2005; Nichols *et al.* 2008) in light of undetected species. The main strength of an occupancy modelling approach is that it can account for undetected species due to local habitat heterogeneity and sampling conditions. Given limited time and resources, repeated surveys at all sites or during all seasons may not be fea-sible. MacKenzie *et al.* (2006) recommended repeated surveys at a subset of sites, but repeated sampling, even at a subset of sites, may not be achievable because of resource limitations or logistics. Alternatively, repeated surveys can be conducted at all or a subset of sites during seasons when resources are available to provide estimates of *psi*- and *p*-covariates. Some *psi*- and *p*-cova-riates may have limited application among different sea-sons (i.e. season-specific effects), but in general, estimates of covariate coefficients in occupancy models can then be used to adjust occupancy estimates for pres-ence/absence data in future or past sampling regimes (e.g. historical or baseline information).

One area of research needed in occupancy modelling is addressing limitations of repeated sampling using some methods. To date, many occupancy approaches have been limited to data collected via observational methods, or passive capture methods, rather than active capture techniques. For example, repeated electrofishing either induces mortality or requires that surveys are suffi-ciently spaced as to prevent deleterious effects on fish. In these cases, the closure assumption may be violated.

Thus, more research is needed to determine how to design sampling strategies or fully utilise presence/absence data from electrofishing approaches.

The observed number of species underestimated potential richness values by about 30% in the Cheoah River, a moderately diverse system. Because of the ability to compare data collected using various sampling methods and the potential widespread transferability of estimates in detection probabilities (Peoples & Frimpong 2011), it is suggested that monitoring fish assemblage composition is conducted with multiple methodologies and gear types to ensure rare species are detected. The modelling approach here permits this by including sampling method as a covariate in a model-based framework for estimation.

Acknowledgments

References

Albanese B., Peterson J.T., Freeman B.J. & Weiler D.A. (2007) Accounting for incomplete detection when estimating site occupancy of Bluenose Shiner (*Pteronotropis welaka*) in southwest Georgia. *Southeastern Naturalist* **6**, 657–668.

Albanese B., Angermeier P.L. & Peterson J.T. (2009) Does mobility explain variation in colonization and population recovery among stream fishes? *Freshwater Biology* **54**, 1444–1460.

Anderson E.P., Freeman M.C. & Pringle C.M. (2006) Ecological consequences of hydropower development in Central America: impacts of small dams and water diversion on neotropical stream fish assemblages. *River Research and Applications* **22**, 397–411.

Angermeier P.L. & Smogor R.A. (1995) Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 936–949.

Bayley P.B. & Austen D.J. (2002) Capture efficiency of a boat electrofisher. *Transactions of the American Fisheries Society* **131**, 435–451.

Bayley P.B. & Peterson J.T. (2001) An approach to estimate probability of presence and richness of fish species. *Transactions of the American Fisheries Society* **130**, 620–633.

Bernardo J.M., Ilhéu M., Matono P. & Costa A.M. (2003) Interannual variation of fish assemblage structure in a Mediterranean river: implications of streamflow on the dominance of native or exotic species. *River Research and Applications* **19**, 521–532.

Boulinier T., Nichols J.D., Sauer J.R., Hines J.E. & Pollock K.H. (1998) Estimating species richness: the importance of heterogeneity in species detectability. *Ecology* **79**, 1018–1028.

Brewer S.K. & Ellersieck M.R. (2011) Evaluating two observational sampling techniques for determining the distribution and detection probability of age-0 smallmouth bass in clear, warmwater streams. *North American Journal of Fisheries Management* **31**, 894–904.

Burdick S.M., Hendrixson H.A. & VanderKooi S.P. (2008) Age-0 Lost River sucker and shortnose sucker nearshore habitat use in Upper Klamath Lake, Oregon: a patch occupancy approach. *Transactions of the American Fisheries Society* **137**, 417–430.

Burnham K.P. & Anderson D.R. (2002) *Model Selection and Multi-model Inference*. New York, NY: Springer. 353 pp.

Cao Y., Hawkins C.P. & Storey A.W. (2005) A method for measuring the comparability of different sampling methods used in biological surveys: implications for data integration and synthesis. *Freshwater Biology* **50**, 1105–1115.

Dauwalter D.C. & Fisher W.L. (2007) Electrofishing capture probability of smallmouth bass in streams. *North American Journal of Fisheries Management* **27**, 162–171.

Dolloff C.A., Krause C. & Roghair C.. (2006) *Cheoah River reconnaissance, August 2006* Report of Center for Aquatic Technology Transfer to USDA Forest Service Southern Research Station. Asheville, NC. 24 pp.

Dorazio R.M. & Royle J.A. (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association* **100**, 389–398.

Dorazio R.M., Royle J.A., Söderström B. & Glimskär A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* **87**, 842–854.

Dorazio R.M., Kéry M., Royle J.A. & Plattner M. (2010) Models for inference in dynamic metacommunity systems. *Ecology* **91**, 2466–2475.

Ensign W.E., Angermeier P.L. & Dolloff C.A. (1995) Use of line transect methods to estimate abundance of benthic stream fishes. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 213–222.

Falke J.A., Fausch K.D., Bestgen K.R. & Bailey L.L. (2010) Spawning phenology and habitat use in a Great Plains, USA, stream fish assemblage: an occupancy estimation approach.

Canadian Journal of Fisheries and Aquatic Sciences **67**, 1942–1956.

Freeman M.C. & Grossman G.D. (1993) Effects of habitat availability on dispersion of a stream cyprinid. *Environmental Biology of Fishes* **37**, 121–130.

Freeman M.C., Crawford M.K., Barrett J.C., Facey D.E., Flood M.G., Hill J. *et al.* (1988) Fish assemblage stability in a southern Appalachian stream. *Canadian Journal of Fisheries and Aquatic Sciences* **45**, 1949–1958.

Gaston K.J., Blackburn T.M., Greenwood J.J.D., Gregory R.R., Quinn R.M. & Lawton J.H. (2000) Abundance-occupancy relationships. *Journal of Applied Ecology* **37**, 39–59.

Gilks W.R., Thomas A. & Speigelhalter D.J. (1994) A language and programme for complex Bayesian modeling. *The Statistician* **43**, 169–178.

Gorman O.T. & Karr J.R. (1978) Habitat structure and stream fish communities. *Ecology* **59**, 507–515.

Grabowski T.B., Ferguson T.D., Peterson J.T. & Jennings C.A. (2009) Capture probability and behavioral response of the robust redhorse, a cryptic riverine fish, to electrofishing. *North American Journal of Fisheries Management* **29**, 721–729.

Grossman G.D., Moyle P.B. & Whitaker J.O. (1982) Stochasticity in structural and functional characteristics of an Indiana stream fish assemblage: a test of community theory. *The American Naturalist* **120**, 423–454.

Grossman G.D., Freeman M.C., Moyle P.B. & Whitaker J.O. (1985) Stochasticity and assemblage organization in an Indiana stream fish assemblage. *The American Naturalist* **126**, 275–285.

Hagen J., Decker S., Korman J. & Bison R.G. (2010) Effectiveness of night snorkeling for estimating steelhead parr abundance in a large river basin. *North American Journal of Fisheries Management* **30**, 1303–1314.

Hayer C.-A. & Irwin E.R. (2008) Influence of gravel mining and other factors on detection probabilities of Coastal Plain fishes in the Mobile River Basin, Alabama. *Transactions of the American Fisheries Society* **137**, 1606–1620.

Hayes D.B., Ferreri C.P. & Taylor W.W. (1996) Active fish capture techniques. In: B.R. Murphy & D.W. Willis (eds) *Fisheries Techniques*, 2nd edn. Bethesda, MD: American Fisheries Society, pp. 193–220.

Hines J. & MacKenzie D. (2012) Program PRESENCE ver 4.0. US Geological Survey. Patuxent Wildlife Research Center. Available online at: http://www.mbr-pwrc.usgs.gov/software/doc/presence/presence.html. (accessed 7 November 2012).

Hubert W.A. (1996) Passive capture techniques. In: B.R. Murphy & D.W. Willis (eds) *Fisheries Techniques*, 2nd edn. Bethesda, MD: American Fisheries Society, pp. 157–192.

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.

Jolly G.M. & Hampton I. (1990) A stratified random transect design for acoustic surveys of fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 1282–1291.

Joseph L.N., Field S.A., Wilcox C. & Possingham H.P. (2006) Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology* **20**, 1679–1687.

Joyce M.P. & Hubert W.A. (2003) Snorkeling as an alternative to depletion electrofishing for assessing cutthroat trout and brown trout in stream pools. *Journal of Freshwater Ecology* **18**, 215–222.

Karr J.R. (1981) Assessment of biotic integrity using fish communities. *Fisheries* **6**, 21–27.

Karr J.R., Fausch K.D., Angermeier P.L., Yant P.R. & Schlosser I.J. (1986) *Assessing Biological Integrity in Running Waters: A Method and its Rationale*. Champaign, IL: Illinois Natural History Survey Special Publication 5, 31 pp.

Korman J., Yard M., Walters C. & Coggins L.G. (2009) Effects of fish size, habitat, flow, and density on capture probabilities of age-0 rainbow trout estimated from electrofishing at discrete sites in a large river. *Transactions of the American Fisheries Society* **138**, 58–75.

Korman J., Decker A.S., Mossop B. & Hagen J. (2010) Comparison of electrofishing and snorkeling mark-recapture estimation of detection probability and abundance of juvenile steelhead in a medium-sized river. *North American Journal of Fisheries Management* **30**, 1280–1302.

MacKenzie D.I. & Kendall W.L. (2002) How should detection probability be incorporated into estimates of relative abundance? *Ecology* **83**, 2387–2393.

MacKenzie D.I., Nichols J.D., Lachman G.B., Droege S., Royle J.A. & Langtimm C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.L. & Hines J.E. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Boston, MA: Academic Press, 312 pp.

Marsh-Mathews E. & Mathews W.J. (2000) Geographic, terrestrial and aquatic factors: which most influence the structure of stream fish assemblages in the midwestern United States? *Ecology of Freshwater Fish* **9**, 9–21.

McDonald T.L. (2003) Review of environmental monitoring methods: survey designs. *Environmental Monitoring and Assessment* **85**, 277–292.

McDonald T.L., Birkes D.S. & Urquhart N.S. (1996) Obtaining species: sample size considerations. *Environmental and Ecological Statistics* **3**, 329–347.

McManamay R.A., Orth D.J., Dolloff C.A. & Cantrell M.A. (2010) Gravel addition as a habitat restoration technique for tailwaters. *North American Journal of Fisheries Management* **30**, 1238–1257.

Moyle P.B. & Mount J.F. (2007) Homogenous rivers, homogenous faunas. *Proceedings of the National Academy of Sciences of the USA* **104**, 5711–5712.

Mullner S.A., Hubert W.A. & Wesche T.A. (1998) Snorkeling as an alternative to depletion electrofishing for estimating abundance

and length-class frequencies of trout in small streams. *North American Journal of Fisheries Management* **18**, 947–953.

Nichols J.D., Bailey L., O'Connell A.F., Talancy N.W., Cambell Grant E.H., Gilbert A.T. *et al.* (2008) Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology* **45**, 1321–1329.

Normandeau Associates. (2002) *Flow Regime and Aquatic Habitat Assessment for the Cheoah River Downstream of Santeetlah Reservoir, North Carolina, Part 1*. Report of Normandeau Associates to Alcoa Power Generating. Alcoa, TN: Normandeau Associates. 429 pp.

Peoples B.K. & Frimpong E.A. (2011) Among-pass, interregional, and single- versus multiple-season comparisons of detection probabilities of stream fishes. *Transactions of the American Fisheries Society* **140**, 67–83.

Pierce C.L., Rasmussen J.B. & Leggett W.C. (1990) Sampling littoral fish with a seine: corrections for variable capture efficiency. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 1004–1010.

Poff N.L. & Allan J.D. (1995) Functional organization of stream fish assemblages in relation to hydrological variability. *Ecology* **76**, 606–627.

Pollock K.H., Nichols J.D., Simons T.R., Farnsworth G.L., Bailey L.L. & Sauer J.R. (2002) Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* **13**, 105–119.

Rota C.T., Fletcher R.J., Dorazio R.M. & Betts M.G. (2009) Occupancy estimation and the closure assumption. *Journal of Applied Ecology* **46**, 1173–1181.

Royle J.A. & Dorazio R.M. (2008) *Hierarchical Modeling and Inference in Ecology*. Amsterdam, The Netherlands: Academic Press. 464 pp.

Royle J.A. & Nichols J.D. (2003) Estimating abundance from repeated presence absence data or point counts. *Ecology* **84**, 777–790.

Royle J.A., Kery M., Gautier R. & Schmid H. (2007) Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs* **77**, 465–481.

Russ W.T. & Fraley S.J. (2009) *Spotfin Chub Monitoring 2007–2016, Little Tennessee River, North Carolina*. Section 6 (ESA) Federal Grant Project, 2008 Annual Progress Report. Raleigh, NC. 25 pp.

Schlosser I.J. (1982) Fish community structure and function along two habitat gradients in a headwater stream. *Ecological Monographs* **52**, 395–414.

Schlosser I.J. (1985) Flow regime, juvenile abundance, and the assemblage structure of stream fishes. *Ecology* **66**, 1484–1490.

Snyder D.E. (2003) *Electrofishing and its Harmful Effects on Fish*. Information and Technology Report USGS/BRD/ITR–2003–0002. Denver, CO. 149 pp.

Travnicheck V.H., Bain M.B. & Maceina M.J. (1995) Recovery of a warmwater fish assemblage after the initiation of a minimum-flow release downstream from a hydroelectric dam. *Transactions of the American Fisheries Society* **124**, 836–844.

Travnichek V.H. & Maceina M.J. (1994) Comparison of flow regulation effects on fish assemblages in shallow and deep water habitats in the Tallapoosa River, Alabama. *Journal of Freshwater Ecology* **9**, 207–216.

USGS (US Geological Survey, Patuxent Wildlife Research Center). (2012) PRESENCE 5.3 and GENPRES software. Available online at: http://www.mbr-pwrc.usgs.gov/software/presence.html. (accessed 7 November 2012).

Ward J.V. & Stanford J.A. (1983) The serial discontinuity concept of lotic ecosystems. In: T.D. Fontaine & S.M. Bartell (eds) *Dynamics of Lotic Ecosystems*. Ann Arbor, MI: Ann Arbor Scientific Publishers, pp. 29–42.

Wenger S.J. & Freeman M.C. (2008) Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology* **89**, 2953–2959.

Wintle B.A., McCarthy M.A., Parris K.M. & Burgman M.A. (2004) Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications* **14**, 703–712.

Witmer G.W. (2005) Wildlife population monitoring: some practical considerations. *Wildlife Research* **32**, 259–263.

Supporting Information

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

Appendix S1. Alternative occupancy models 15 species captured in the Cheoah River in 2009.