Species residency status affects model selection and hypothesis testing in freshwater community ecology

JASON T. BRIED*, ADAM M. SIEPIELSKI*, DANIEL DVORETT†, SUNEETI K. JOG‡, MICHAEL A. PATTEN§, XIAO FENG¶ AND CRAIG A. DAVIS†

*Department of Biological Sciences, Program in Ecology and Evolutionary Biology, University of Arkansas, Fayetteville, AR, U.S.A.
†Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, OK, U.S.A.
‡Department of Natural Sciences, Northeastern State University, Tahlequah, OK, U.S.A.
§Oklahoma Biological Survey and Department of Biology, University of Oklahoma, Norman, OK, U.S.A.
¶Department of Integrative Biology, Oklahoma State University, Stillwater, OK, U.S.A.

SUMMARY

1. Species occurrences have multiple ecological states that may strongly influence community analysis and inference. This may be especially true in freshwater systems where many animals have complex life cycles with adult dispersal and juvenile resident stages.
2. The effects of ecological state variation on standard empirical approaches are largely unknown. Here, we analysed the effects of natal resident versus non-natal immigrant species occurrence on community-level environmental gradient modelling and spatial–environmental hypothesis testing using adult dragonflies and damselflies as model taxa.
3. Resident and total (resident + immigrant) occurrences of these taxa responded to different sets of environmental variables and resident occurrences reduced model selection uncertainty in 75% of test cases.
4. Effects of environmental gradients, spatial gradients or both were observed in residents but not immigrants, and supported predictions of dispersal limitation and niche-based species sorting often implicated for structuring freshwater communities.
5. These results indicate that resident-only analysis of the dispersal stage should improve multi-model inference and detection of spatial–environmental effects in freshwater community ecology. The species resident–immigrant dichotomy neglects population dynamics and individual variation yet apparently marks an ecologically significant boundary that scales up to influence community-level occurrence patterns.

Keywords: dispersal, niche, odonates, resident, wetlands

Introduction

Species occurrence is a basic unit of community data, but occurrences can have multiple ecological states (MacKenzie et al., 2009), such as breeding versus non-breeding or resident versus immigrant. Species-level residency state, although often treated as taxonomically fixed according to dispersal mode (e.g. Hall et al., 2004; Cañedo-Argüelles et al., 2015), can be designated per locality depending on natal origin or reproductive status. This multi-state occurrence is important (at least implicitly) in studying actively dispersed animals and may strongly influence ecological inference and applications (Bried et al., 2015b; Patten, Bried & Smith-Patten, 2015). Given the length of time in experiencing a certain locality, natal residents are more likely than non-natal immigrants to associate with local conditions and environmental gradients (Bohonak & Jenkins, 2003). However, it is currently unknown whether fractioning out resident species occurrences will greatly or negligibly change the outcomes of standard empirical approaches in ecology, such as information-theoretic model selection (Anderson, 2008).

Niche-based processes that operate locally and dispersal-based processes that operate beyond site boundaries can drive species occurrence distributions (Smith et al.,


2015), but these forces may not apply equally to sympatric resident and immigrant species occurrences. For example, as a result of local adaptation or in situ experience during ontogeny, resident species occurrence is expected to create a more non-random community structure compared to immigrant species occurrence (Bried et al., 2015b). As such, resident communities may show a decay of compositional similarity among sites, driven by decreasing environmental similarity and increasing dispersal limitation over geographic distance (Nekola & White, 1999; Siepielski & McPeek, 2013). Varying dispersal rates may further complicate community structure, with lower rates likely to increase community spatial structuring and autocorrelation and higher rates likely to homogenise local communities thereby reducing spatial structure (Bonada, Dolédec & Statzner, 2012; Cañedo-Argüelles et al., 2015). The alternative could be random community patterns with no clear signature of structuring, as might be expected for immigrants (Bried et al., 2015b). If immigrant species occurrence is random or opportunistic, then analyses focused on the resident portion of total (resident + immigrant) species occurrence data may improve our understanding of community relationships with spatial and environmental variables.

Many freshwater animals have complex life cycles with adult dispersal and juvenile resident stages, making it possible to parse natal aquatic occurrences from the total occurrence. Adult dragonfly and damselfly (Odonata) species, for example, can be classified as local resident or immigrant using long-term series of occurrence data (Shiffer & White, 2014) or empirical relationships to exuviae (Bried et al., 2015a). Presumably, the resident species occurrence pattern results from underlying ecological gradients linked to biotic and abiotic factors such as predation, shading, vegetation, water chemistry and hydroperiod (McPeek, 1990; Corbet, 1999; Schindler, Fesl & Chovanec, 2003; Crumrine, Switzer & Crowley, 2008; McCauley et al., 2008; Remsburg, Olsen & Samways, 2008; De Marco, Batista & Cabette, 2015). In contrast, immigrant species occurrence should not be reflective of such deterministic patterns, but these contrasting assumptions have not been validated empirically. In general, researchers tend to overlook the presence and influence of non-residents in their study system (Ruiz-Gutierrez et al., 2016).

Just as any local assemblage will probably include some mixture of co-occurring and coexisting species (sensu Siepielski & McPeek, 2010), that same assemblage probably also has a time-varying mixture of natal resident and non-natal immigrant species. Strictly speaking, natal origin applies to individuals, but even a false dichotomy at the species level may still be important at the community level. We used adult odonates in wetland systems to study the effects of natal resident versus non-natal immigrant species occurrence on community-level environmental gradient modelling and spatial–environmental hypothesis testing. Specifically, we (i) compared environmental model ranking and inference between total and resident species data sets under the assumption that resident species occurrence will reduce model selection uncertainty and (ii) hypothesised that environmental responses will be stronger for resident communities due to greater potential for species sorting (environmental filtering) compared to immigrants, and that spatial responses will be stronger for damselflies due to weaker dispersal compared to dragonflies (Conrad et al., 1999; Lorenzo-Carballa et al., 2015). Model selection uncertainty is defined here as the level of empirical support for the estimated best model, among those considered, using information-theoretic criteria (Anderson, 2008).

Methods

Study areas

Wetland study sites were located in New York and Oklahoma, U.S.A. (map in Bried et al., 2015b), because if natal origin truly matters, outcomes should be similar in very different environmental and climatic settings. Sites in New York were clustered in the Eastern Great Lakes and Upper Hudson Valley regions at three preserves (Albany Pine Bush, Rome Sand Plains and Wilton Wildlife Preserve & Park) characterised by glaciolacustrine sand deposits, gently rolling topography and pyrogenic vegetation communities. We selected 32 depressional wetlands (6, 12 or 14 sites per preserve) with herbaceous, shrub or mixed herb–shrub dominance and seasonal to semi-permanent hydroperiods. Wetlands in each preserve were bordered by pine and mixed forest with scattered wildlife management openings and urban or exurban development.

Wetland sites in Oklahoma spanned several drainage basins and ecoregions and were selected along a qualitative gradient of human disturbance from minimally altered forest in protected areas to highly modified agricultural and urban settings. The 30 study wetlands were open-canopy patches of freshwater emergent community types in depressional, lacustrine or riverine geomorphic classes. About half of the wetlands were created by beavers (Castor canadensis) and the rest by human

© 2016 John Wiley & Sons Ltd, Freshwater Biology, 61, 1568–1579
impoundment or excavation, with many of these being former farm ponds that have experienced sedimentation, reduced water levels and growth of aquatic macrophytes. All had seasonal to semi-permanent inundation periods, including the fringes of permanent-water ponds.

**Odonata surveys and resident versus immigrant designations**

Bi-weekly surveys of adults were completed throughout the flight season in each study area. New York surveys were repeated eight times from mid-May to mid-August in 2011. Oklahoma surveys were repeated six times from early May until wetlands became dry in mid-July during 2012 (13 sites) and 2013 (17 sites). Each survey was conducted over 3–5 days (weather dependent), covering 4–7 sites each day on a rotating basis over the study period. In both locations, sampling occurred along a fixed route through the emergent vegetation zone or along the vegetation–water interface.

Surveys lasted 20–30 min per site between 09:00 and 18:30 hours. Species were identified by sight (usually) or in hand (as necessary), with a focus on mature males to minimise false positives. All taxa were assigned to abundance categories (1–5, 6–20, 21–100 and >100 individuals across the patch or search area) and presence of tenerals (newly emerged individuals) was noted.

We aimed to test whether resident and immigrant designations at the species level may affect inferences at the community level. A resident species occurrence designation, as determined previously using the same data set (see Bried et al., 2015a), required >20 individuals during at least one survey, tenerals on at least two surveys or species detection (teneral or not) on at least four surveys, with some exceptions (see Bried et al., 2015b). Otherwise the species occurrence was classified as immigrant. These criteria, however imperfect (caveats discussed in Bried et al., 2015a), were based on probabilistic relationships between adult survey information and the detectability-corrected occurrence of final larval exuviae, which provide the definitive indicator of natal origin and life cycle completion in odonates. Exuviae were surveyed concurrently with adults throughout the study (details in Bried et al., 2015a). Exuviae occupancy probabilities were ~0.9 (0.08–0.13 SE) for each of the three criteria, and the exuviae were 9–18 times more likely to occur under these conditions compared to not finding any adults (Bried et al., 2015a). Therefore, our resident–immigrant dichotomy is a per site separation of species according to probabilities of being present as exuviae.

**Local environmental variables**

Adult odonate communities are potentially structured by many environmental variables at local and landscape scales (Table 1). Vegetation may influence adult odonate occurrence and community structure by providing the primary oviposition site, territorial marker and foraging/resting perch for many species (Corbet, 1999; Schindler et al., 2003). Vegetation was systematically sampled at sites in New York using 1 m quadrats throughout the vegetated area. Species frequencies were pooled into six structure classes: ferns, graminoids, low-emergent forbs (<5 cm above water), tall-emergent forbs, shrubs and trees. Sites in Oklahoma were patrolled until no new vascular plant species could be found, with site-wide cover estimates averaged and summarised into five structure classes: graminoids, emergent forbs, aquatic forbs (submersed, floating or floating leaved), shrubs and trees. In both locations, vegetation surveys were conducted once at each site around the middle of the odonate survey periods. Also in both locations, the ratio of emergent vegetated area to open water area (i.e. plant–water interspersion) was visually estimated (nearest 5%) during each survey and then averaged (‘Open’ model in Table 1).

Coarse woody debris and tree snags related to beaver activity were prominent at many Oklahoma sites, and some odonate species in our study (libellulids such as *Perithemis tenera* and *Platthemis lydia*) commonly used these substrates for emergence or perching. The site-wide cover (nearest 5%) of downed and standing deadwood was visually estimated on each odonate survey and then averaged. Presence of non-rooted *Sphagnum* moss in the water column (‘Soup’ model in Table 1), a distinctive feature of the New York study areas that may promote habitat-specialist odonates, was noted at 14 sites. Also for New York, the presence of significant shading (>25% of patch area, visually estimated across surveys) was noted at 17 sites, with smaller amounts of shade (<10%) at remaining sites. New York wetlands were also categorised by major structural differences as either herbaceous dominated (11 sites), shrub dominated (11 sites) or mixed shrub/ herbaceous (10 sites).

Water permanence may affect the structure and dynamics of odonate communities (Stoks & McPeek, 2006; Crumrine et al., 2008), so sites were classified as having seasonal versus semi-permanent hydroperiods using field observations and available orthoimagery (2004–2013 discontinuously) from the U.S. National Area
Seaonal wetlands typically lose all surface water by the late growing season, whereas semi-permanent wetlands remain flooded throughout the growing season in most years (Mitsch & Gosselink, 2000). Oklahoma sites were checked for inundation semi-annually during 2012–2015, and New York sites occasionally during 2005–2011. For Oklahoma, the hydroperiod categories were modelled alone and in combination with the presence of beaver activity (‘HydroB’ model in Table 1). Additionally, surface water depth influences the amount of exposed vegetation required for emergence, perching and endophytic oviposition. Depth was determined during each survey from a staff gauge installed at the approximate deepest point (up to 1 m) in each wetland.

Surface water depth influences the amount of exposed vegetation required for emergence, perching and endophytic oviposition. Depth was determined during each survey from a staff gauge installed at the approximate deepest point (up to 1 m) in each wetland.

Landscape environmental variables

Climate may greatly influence local and regional odonate distributions (Ball-Damerow, M’Gonigle & Resh, 2014; Collins & McIntyre, 2015). We extracted 4 km resolution precipitation and temperature data from the PRISM Climate Group (http://www.prism.oregon-state.edu/normals/). This included annual precipitation and monthly maximum temperature averages for March through August 2006–2011 (for New York data) and site, and 500 mL composite water samples were collected during each survey to analyse for chlorophyll a and total phosphorus concentrations. Chlorophyll was measured fluorometrically and phosphorus was measured using a Hach DR 5000 UV-Vis spectrometer following standard manufacturer protocols. For all variables, we used the mean across surveys in subsequent analyses.

Table 1 Local and landscape environmental variables and fixed effects specification for modelling odonate communities surveyed in New York (NY) and Oklahoma (OK). ‘*’ denotes interaction, ‘+’ denotes multiple main effects (no interaction).

<table>
<thead>
<tr>
<th>Model name</th>
<th>Model specification</th>
<th>Variable description</th>
<th>Data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>patch.size</td>
<td>Approximate non-forested wetland patch area</td>
<td>NY, OK</td>
</tr>
<tr>
<td>Shade</td>
<td>edge.shade</td>
<td>Presence of significant (&gt;25%) patch shading from edge tree canopy</td>
<td>NY</td>
</tr>
<tr>
<td>Deadwood</td>
<td>Dwood</td>
<td>Standing and downed woody debris coverage</td>
<td>OK</td>
</tr>
<tr>
<td>VegPCA</td>
<td>vegPCA1 + vegPCA2</td>
<td>Synthetic variables combining several vegetation growth forms</td>
<td>NY, OK</td>
</tr>
<tr>
<td>VegType</td>
<td>veg.type</td>
<td>Herbaceous dominated, shrub dominated or mixed herb–shrub</td>
<td>NY</td>
</tr>
<tr>
<td>Open</td>
<td>%open</td>
<td>Plant–open water interspersion averaged across surveys</td>
<td>NY, OK</td>
</tr>
<tr>
<td>Soup</td>
<td>Soup</td>
<td>Presence of dense peat moss (Sphagnum spp.) in the water column</td>
<td>NY</td>
</tr>
<tr>
<td>pH</td>
<td>pH</td>
<td>Mean across surveys</td>
<td>NY, OK</td>
</tr>
<tr>
<td>WQ-PCA</td>
<td>wq.PCA1</td>
<td>Synthetic water quality variable combining chlorophyll a, total phosphorus and turbidity</td>
<td>OK</td>
</tr>
<tr>
<td>WQ-CV</td>
<td>cv.chl+cv.TP+cv.turb</td>
<td>Variation across surveys in chlorophyll a (cv.chl), total phosphorus (cv.TP) and turbidity (cv.turb)</td>
<td>OK</td>
</tr>
<tr>
<td>Hydro</td>
<td>Hydro</td>
<td>Seasonal vs. semi-permanent hydroperiod</td>
<td>NY, OK</td>
</tr>
<tr>
<td>HydroB</td>
<td>hydro+beaver</td>
<td>Presence of significant beaver (Castor canadensis) influence on hydrology</td>
<td>OK</td>
</tr>
<tr>
<td>Depth</td>
<td>m.depth*r.depth</td>
<td>Mean surface water depth (m.depth) and range in depth across surveys (r.depth)</td>
<td>NY, OK</td>
</tr>
<tr>
<td>Landscape:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precip</td>
<td>Precip</td>
<td>Mean annual precipitation</td>
<td>OK</td>
</tr>
<tr>
<td>PrecipTemp</td>
<td>precip*tmax</td>
<td>Total annual precipitation (precip) and average of monthly maximum temperatures (tmax)</td>
<td>NY, OK</td>
</tr>
<tr>
<td>WetArea</td>
<td>wet.area</td>
<td>Total area of mapped wetlands in the optimal buffer</td>
<td>NY, OK</td>
</tr>
<tr>
<td>WetAreaDist</td>
<td>dist.wet*area</td>
<td>Distance to nearest wetland (dist.wet) and total area of mapped wetlands in the optimal buffer</td>
<td>NY, OK</td>
</tr>
<tr>
<td>RoadDens</td>
<td>road.density</td>
<td>Total length of roads (county, state, interstate) divided by area in the optimal buffer</td>
<td>NY, OK</td>
</tr>
<tr>
<td>RdDensDist</td>
<td>dist.hwy*road.dens</td>
<td>Shortest distance to a road (dist.hwy) and total length of roads divided by area in the optimal buffer</td>
<td>NY, OK</td>
</tr>
<tr>
<td>LULC</td>
<td>LULC</td>
<td>Weighted index of 15 major land-use/land cover classes in the optimal buffer</td>
<td>NY, OK</td>
</tr>
<tr>
<td>Altered</td>
<td>Altered</td>
<td>Percentage of open, developed, crop, pasture/hay and barren land in the optimal buffer</td>
<td>NY, OK</td>
</tr>
<tr>
<td>Forest</td>
<td>Forest</td>
<td>Percentage of deciduous, evergreen and mixed forest in the optimal buffer</td>
<td>NY, OK</td>
</tr>
</tbody>
</table>
2008–2013 (for Oklahoma data), thus capturing the interannual and seasonal variations leading up to and coinciding with the odonate surveys. At 4 km resolution, many New York sites within the same preserve had the same precipitation and temperature data, but these data could still be useful in explaining community differences among the preserves.

Nearest distance to water and total area of waterbodies in the surrounding landscape may influence adult odonate occurrence and diversity patterns (Raebel et al., 2012; Harms, Kinkead & Dinsmore, 2014). Nearest distance to a wetland and the total area of wetlands at nested buffer scales (0.1, 0.5, 1.0, 2.0 km) from the study site perimeter were extracted from National Wetlands Inventory data. Roads can also be an important predictor of wetland community structure and adult odonates are affected negatively by roads (Soluk, Zercher & Worthington, 2011). Using the 2010 TIGER/Line® shapefiles for New York and Oklahoma, we determined distance to nearest road (county, state or interstate highways) and road density, measured by dividing the total length of all linear road features within each buffer by the total buffer area (Rooney et al., 2012).

The remaining landscape variables (see Table 1) were calculated at each buffer extent. Following Dvorett et al. (2013), we assigned a weighting coefficient (range: 0–1) to each of the 15 major land-use/land-cover (LULC) classes in the 2011 National Land Cover Dataset (NLCD) based on potential impacts to wetlands. Relative scores were calculated as sums of products between LULC coefficients and corresponding percentages of LULC at each buffer extent. We also calculated the combined percentage of altered land-use types (those with weighting coefficients <1; see Table 1 and Dvorett et al., 2013) at each buffer scale as a measure of anthropogenic disturbance. Finally, we modelled the combined NLCD forest classes assuming that forest structure will affect odonate dispersal dynamics and thermoregulation and ultimately spatial diversity (Cordero-Rivera, 2006; De Marco et al., 2015).

Spatial variables

We used eigenvector-based spatial filters (Fortin et al., 2012) among the explanatory variables for hypothesis testing. A pairwise matrix of geographical distances was submitted to a principle coordinates analysis, producing centred and uncorrelated eigenvectors (filters) that provide a multi-scale decomposition of the site geographic positions and autocorrelations. Filters were extracted from the wetland polygon centroids using the program Spatial Analysis in Macroecology v4 (Rangel, Diniz-Filho & Bini, 2010).

Oklahoma filters were extracted across the pairwise matrix of geographic distances for all sites. New York filters were extracted within each preserve, with each combination of preserve-scale results being used as spatial predictors across the whole sample of wetlands. Odonates were unlikely to disperse the 50–150 km distances among the preserves (at least during the study period), and filtering all sites together was uninformative because it produced only a single vector with two values (one for Rome Sand Plains and one for the other two preserves).

Environmental modelling

For the first objective (‘Does resident species occurrence reduce model selection uncertainty?’) we modelled total and resident dragonfly and damselfly communities in response to the local and landscape environmental variables. We applied linear mixed effects modelling (R package lme4), with an intercept random effect for each New York preserve to account for the geographic site clusters and for the two Oklahoma survey years to control for any systematic biases such as the severe drought in 2012 and extensive rains in late spring 2013. We used the first axis of a correspondence analysis of the community data as the response variable for each scenario, that is, total and resident species occurrence of dragonflies and damselflies in each location (eight scenarios altogether). The first axis of correspondence analysis may strongly represent a dominant underlying species composition gradient when applied to community data (McCune & Grace, 2002; Presley, Higgins & Willig, 2010; Dallas & Drake, 2014). Indeed, a previous analysis of this axis from our data suggested non-random occurrence structure and latent structuring gradients for each resident scenario (Bried et al., 2015b). As is standard practice, singleton species were removed prior to the ordination because they exaggerate the distinctiveness of sites (McCune & Grace, 2002; Dallas & Drake, 2014).

We fit many plausible environmental models (Table 1) along with the null model $y_{ij} = \beta_0 + \alpha_j + \varepsilon_{ij}$, where $y_{ij}$ is the $i$th response of the $j$th grouping level, $\beta_0$ the intercept, $\alpha_j$ the level-specific effect and $\varepsilon_{ij}$ the residual associated with the $i$th response of the $j$th grouping level. Principle components analysis was used to remove collinearity and reduce the number of vegetation and water quality variables, thereby reducing the number of models. Residuals of each model were checked for normality (Shapiro–Wilks test) and heteroscedasticity (increasing spread over fitted values). Models that failed to meet the...
assumptions were removed from candidate selection. All numeric and integer covariates were standardised (mean = 0, SD = 1) to improve convergence of maximum likelihood estimation. We constructed only simple models (few covariates and interactions) because our goal was to compare the total and resident scenarios, not to investigate diversity–environment relationships. We focused on the total and resident scenarios because one or the other is likely to be used for environmental modelling in practice, as well as for community-based applications (see Bried et al., 2015b).

We used the second-order Akaike information criterion (AICc) (Anderson, 2008) to evaluate candidate models. We ran a preliminary AIC ranking to select the ‘optimal’ landscape extent (sensu Rooney et al., 2012) for road density, total area of wetlands, total LULC, per cent altered land and per cent forest (see Table S1 in Supporting Information). For each of these variables, we used the summed weights across the total and resident scenarios to choose optimal extent. Final model comparison involved a common set of variables at the same extents and included all models specified in Table 1 that met assumptions. Interpretations were made at standard cutoffs of $\Delta_i \leq 2$ and 6 (Anderson, 2008; Richards, 2008). We do not quantify model selection uncertainty in the strict sense (i.e. difference in model rankings from the present data versus from another data set of the same size), but rather judge it on the separation of models by their estimated probabilities (weights) given the data (Anderson, 2008).

Spatial–environmental hypothesis testing

We tested four predictions assuming stronger environmental association in residents than immigrants and stronger dispersal in dragonflies than damselflies: $P_1$, significant spatial–environmental effects (interaction or main) on resident damselflies due to combined species sorting and dispersal limitation; $P_2$, greater spatial than environmental effect on immigrant damselflies due to dispersal limitation and lack of species sorting or no significant effects due to random occurrence; $P_3$, greater environmental than spatial effect on resident dragonflies due to species sorting and weaker spatial structure caused by stronger dispersal; and $P_4$, no significant effects on immigrant dragonflies due to lack of species sorting, lack of dispersal limitation and potential random occurrence. Data sets were reduced to the common set of species between resident and immigrant scenarios so that only the patterns of occurrence differed, mitigating species-specific effects such as differences in detection probability.

We used generalised linear models with binomial errors and multivariate extensions (mvabund package; Wang et al., 2012) to allow simultaneous testing of multiplicative effects at the community level. We modelled the best supported environmental variables and spatial filters; the ‘optimal’ spatial filter (see Table S2 in Supporting Information) was selected in the same manner as described above for landscape extents. Mean-variance and logit-linearity assumptions were checked by plotting the residuals versus fits. For environmental models in violation the next best model was used; none of the optimal spatial filters appeared to be in violation.

Results

Model selection

The same minimum adequate model ($\Delta_{AICc} = 0$) was selected in total and resident scenarios in three of four test cases (Table 2); for Oklahoma dragonflies the top models switched ranks but remained competitive ($\Delta_i \leq 2$). Well-supported models were more clearly distinguished in resident than in total data, as suggested by the normalised weights (Table 2). There were also more potential models ($\Delta_i \leq 6$) and therefore more uncertainty for total than for resident, and the resident was more likely to reject the null model. Total and resident gave a similar performance only in the Oklahoma damselflies. Overall, the model order changed markedly with resident data, which could lead to different final interpretations.

Hypothesis testing

All four predictions received at least partial support (Table 3). For resident damselflies ($P_1$), there was evidence of a spatially structured environmental response for New York sites, but only a spatial effect for Oklahoma sites. For immigrant damselflies ($P_2$), deviance reduction from the spatial term slightly exceeded that of the environmental, but the effects were not statistically significant. For resident dragonflies ($P_3$), environmental association exceeded the spatial (for New York only) and all environmental associations for immigrants (both locations). As predicted, all tests for immigrants lacked significance (Table 3). With a larger sample, the $P_4$ interaction for New York sites may have crossed the nominal Type I error threshold, but this potential disappears when using the second-ranked environmental predictor (Shade, d.f. = 24, Dev = 10.23, $P = 0.798$) in place of PrecipTemp (see Table 2 for rankings).
Discussion

Model selection

Minimum adequate models were similar for total and resident scenarios, but the probability of these models being the estimated best (see Anderson, 2008) was smaller for total species occurrence. The model weight for total species occurrence tended to be more evenly distributed, which further suggests greater model selection uncertainty. In contrast, for the resident species occurrence, certain models stood out (except in one case) with clear relative empirical support. Furthermore, total and resident model rankings greatly differed across the information-theoretic grey zone ($\Delta AIC_c < 4$), where models lack strong support but still may be worth considering (Anderson, 2008; Richards, 2008).

Table 2 Environmental model rankings for species occurrence scenarios of adult odonate communities. Shown are models with second-order Akaike information criterion ($AIC_c$) ≤ 6 in the Total column; normalised weights ($w_i$) apply to the full model set and thus do not sum to 1 in this table. Model names correspond with Table 1 and numerals after the name indicate the optimal buffer extent (100 m, 500 m, 1 km, 2 km) determined from preliminary model selection (Table S1). Entries in bold font indicate the variables used for spatial-environmental hypothesis testing (see Table 3). ‘null’ – random intercept only, $K$ – number of parameters.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Suborder</th>
<th>Model name</th>
<th>$K$</th>
<th>$\Delta AIC_c$</th>
<th>$w_i$</th>
<th>$\Delta AIC_c$</th>
<th>$w_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>Damselflies</td>
<td>WetArea500</td>
<td>3</td>
<td>0</td>
<td>0.383</td>
<td>0</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Null</td>
<td>2</td>
<td>2.4</td>
<td>0.113</td>
<td>6.9</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Soup</td>
<td>3</td>
<td>3.7</td>
<td>0.059</td>
<td>8.0</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Altered100</td>
<td>3</td>
<td>4.3</td>
<td>0.044</td>
<td>9.0</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LULC100</td>
<td>3</td>
<td>4.4</td>
<td>0.043</td>
<td>8.8</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forest100</td>
<td>3</td>
<td>4.5</td>
<td>0.041</td>
<td>8.4</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Area</td>
<td>3</td>
<td>4.5</td>
<td>0.039</td>
<td>7.5</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open</td>
<td>3</td>
<td>4.7</td>
<td>0.037</td>
<td>9.2</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydro</td>
<td>3</td>
<td>4.7</td>
<td>0.036</td>
<td>7.8</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shade</td>
<td>3</td>
<td>4.7</td>
<td>0.036</td>
<td>7.8</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pH</td>
<td>3</td>
<td>4.8</td>
<td>0.034</td>
<td>7.9</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RoadDens100</td>
<td>3</td>
<td>4.9</td>
<td>0.033</td>
<td>9.3</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WetArea1 km</td>
<td>4</td>
<td>3.4</td>
<td>0.071</td>
<td>0.8</td>
<td>0.382</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Null</td>
<td>3</td>
<td>4.4</td>
<td>0.043</td>
<td>12.1</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydro</td>
<td>4</td>
<td>5.0</td>
<td>0.032</td>
<td>12.5</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WetAreaDist</td>
<td>5</td>
<td>5.0</td>
<td>0.031</td>
<td>3.6</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RdDensDist</td>
<td>5</td>
<td>5.3</td>
<td>0.027</td>
<td>11.9</td>
<td>0.002</td>
</tr>
<tr>
<td>Dragonflies</td>
<td>PrecipTemp</td>
<td>6</td>
<td>0</td>
<td>0.388</td>
<td>0</td>
<td>0.565</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shade</td>
<td>4</td>
<td>0.7</td>
<td>0.277</td>
<td>14.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RoadDens2 km</td>
<td>4</td>
<td>2.3</td>
<td>0.121</td>
<td>14.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WetArea1 km</td>
<td>4</td>
<td>3.4</td>
<td>0.071</td>
<td>0.8</td>
<td>0.382</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Null</td>
<td>3</td>
<td>4.4</td>
<td>0.043</td>
<td>12.1</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydro</td>
<td>4</td>
<td>5.0</td>
<td>0.032</td>
<td>12.5</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WetAreaDist</td>
<td>6</td>
<td>5.3</td>
<td>0.028</td>
<td>4.9</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Altered1 km</td>
<td>4</td>
<td>5.8</td>
<td>0.021</td>
<td>13.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>Damselflies</td>
<td>Hydro</td>
<td>4</td>
<td>0</td>
<td>0.391</td>
<td>0</td>
<td>0.378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forest1 km</td>
<td>4</td>
<td>1.5</td>
<td>0.188</td>
<td>4.3</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Null</td>
<td>3</td>
<td>2.6</td>
<td>0.105</td>
<td>3.1</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HydroB</td>
<td>5</td>
<td>2.8</td>
<td>0.095</td>
<td>3.1</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open</td>
<td>4</td>
<td>3.8</td>
<td>0.057</td>
<td>5.8</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Precip</td>
<td>4</td>
<td>4.9</td>
<td>0.034</td>
<td>5.2</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pH</td>
<td>4</td>
<td>5.0</td>
<td>0.032</td>
<td>5.7</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depth</td>
<td>6</td>
<td>5.4</td>
<td>0.027</td>
<td>3.1</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Altered100</td>
<td>4</td>
<td>5.4</td>
<td>0.026</td>
<td>4.2</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RoadDens500</td>
<td>4</td>
<td>5.4</td>
<td>0.026</td>
<td>1.4</td>
<td>0.186</td>
</tr>
<tr>
<td>Dragonflies</td>
<td>LULC2 km</td>
<td>4</td>
<td>0</td>
<td>0.240</td>
<td>1.9</td>
<td>0.235</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forest100</td>
<td>4</td>
<td>0.7</td>
<td>0.170</td>
<td>0</td>
<td>0.598</td>
</tr>
<tr>
<td></td>
<td></td>
<td>null</td>
<td>3</td>
<td>0.8</td>
<td>0.163</td>
<td>5.1</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pH</td>
<td>4</td>
<td>1.4</td>
<td>0.117</td>
<td>7.7</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydro</td>
<td>4</td>
<td>2.9</td>
<td>0.057</td>
<td>6.1</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RdDensDist</td>
<td>6</td>
<td>2.9</td>
<td>0.057</td>
<td>7.6</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Deadwood</td>
<td>4</td>
<td>2.9</td>
<td>0.056</td>
<td>7.2</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WQ-PCA</td>
<td>4</td>
<td>3.2</td>
<td>0.049</td>
<td>6.5</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Size</td>
<td>4</td>
<td>3.3</td>
<td>0.046</td>
<td>6.9</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Depth</td>
<td>6</td>
<td>3.9</td>
<td>0.034</td>
<td>8.9</td>
<td>0.007</td>
</tr>
</tbody>
</table>
These results suggest a strong potential for alternate conclusions between total and resident species occurrence scenarios. For example, one might ignore all covariates except WetArea500 for New York total damselflies because $D_i > 2.0$ for all other covariates. Area of water in the surrounding landscape may indeed influence adult odonate species occurrence and local diversity (Raebel et al., 2012; Harms et al., 2014). This model, however, did not have a large normalised weight (38%), the null model ranked second, and most remaining models landed in the grey zone. In contrast, for New York resident damselflies, the WetArea500 model had a relatively large weight (69%) and only two other covariates (WetAreaDist, PrecipTemp) were potentially informative, carrying 17% weight combined. New York and Oklahoma dragonflies had a similar discrepancy between total and resident. Hence, these results suggest that fractioning out resident species occurrences can greatly reduce model selection uncertainty and simplify or strengthen inferences on the underlying environmental factors shaping community structure.

Oklahoma damselflies were the exception, in that they had similar model selection results for total and resident. It may be that we did not include important variables (or combinations of variables), such as fish presence in about half of Oklahoma sites, or that we needed more detailed measurements of hydrology, shading and vegetation structure, all of which may strongly filter odonate communities (McPeek, 1990; Schindler et al., 2003; Crumrine et al., 2008; McCauley et al., 2008; Remsburg et al., 2008; Siepielski & McPeek, 2013; De Marco et al., 2015). This possibility is supported by a previous analysis that found signatures of underlying ecological factors shaping community structure in all resident scenarios, including the Oklahoma damselflies (Bried et al., 2015b). Alternatively, it may be that relevant variables were included, but our simple models (few covariates and interactions) did not adequately represent the complex, multi-dimensional environment to which the community responds. Another possibility, supported by our spatial–environmental analysis, is that the Oklahoma damselfly community was controlled more by spatial processes.

### Hypothesis testing

The importance of spatial and environmental gradients was consistently observed for resident species occurrence but not for immigrants. This agrees with earlier findings of non-random resident versus random immigrant species occurrence within the meta-community context (Bried et al., 2015b) and helps to explain some structural patterns identified in that study. Niche differences are often implicated for structuring freshwater.

---

**Table 3** Spatial–environmental predictions (P1...P4) and analysis of deviance tables (using log-likelihood ratio statistic and 999 bootstrap replicates) for multivariate generalised linear modelling of species occurrence data. The interaction is omitted for P2 and P3 because it is not part of the predictions. Environmental (env) and spatial (spat) gradients were chosen based on information-theoretic criteria (Table 2, Table S2) and on satisfying the model assumptions.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Prediction</th>
<th>Data set</th>
<th>Effect</th>
<th>d.f.</th>
<th>Dev</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident damselflies</td>
<td>env and spat effects or env × spat effect (P1)</td>
<td>New York</td>
<td>Env</td>
<td>25</td>
<td>18.84</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>24</td>
<td>7.72</td>
<td>0.694</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Env × Spat</td>
<td>23</td>
<td>23.41</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oklahoma</td>
<td>Env</td>
<td>24</td>
<td>12.83</td>
<td>0.206</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>23</td>
<td>26.00</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Env × Spat</td>
<td>22</td>
<td>3.37</td>
<td>0.895</td>
</tr>
<tr>
<td>Immigrant damselflies</td>
<td>spat &gt; env effect or no significant effects (P2)</td>
<td>New York</td>
<td>Env</td>
<td>25</td>
<td>13.63</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>24</td>
<td>14.64</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oklahoma</td>
<td>Env</td>
<td>24</td>
<td>8.98</td>
<td>0.413</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>23</td>
<td>11.96</td>
<td>0.247</td>
</tr>
<tr>
<td>Resident dragonflies</td>
<td>env &gt; spat effect (P3)</td>
<td>New York</td>
<td>Env</td>
<td>23</td>
<td>42.88</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>24</td>
<td>30.94</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oklahoma</td>
<td>Env</td>
<td>25</td>
<td>23.83</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>24</td>
<td>24.44</td>
<td>0.248</td>
</tr>
<tr>
<td>Immigrant dragonflies</td>
<td>No significant effects (P4)</td>
<td>New York</td>
<td>Env</td>
<td>23</td>
<td>15.15</td>
<td>0.667</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>24</td>
<td>28.28</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Env × Spat</td>
<td>20</td>
<td>24.48</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oklahoma</td>
<td>Env</td>
<td>25</td>
<td>14.50</td>
<td>0.484</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spat</td>
<td>24</td>
<td>9.42</td>
<td>0.837</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Env × Spat</td>
<td>23</td>
<td>13.34</td>
<td>0.625</td>
</tr>
</tbody>
</table>
communities (e.g. reviews by Wellborn, Skelly & Werner, 1996; Heino, 2011), and our results suggest the resident fraction of occurrence data underlies any species sorting.

The spatial–environmental interaction observed in New York resident damselflies (P1) indicates spatially structured environmental responses, which may have driven the Clementsian-style turnover suggested previously from a null model analysis (Bried et al., 2015b). Residents are attuned to local environmental conditions (Bohonak & Jenkins, 2003) that may be spatially structured (Borcard, Legendre & Drapeau, 1992); therefore, resident occurrences of species with similar requirements or tolerances should cluster along spatially structured environmental gradients. A recent study comparing stronger (Odonata) and weaker (Trichoptera) dispersers concluded that habitat type and dispersal ability combined to influence spatial–environmental structuring (Curry & Baird, 2015). Combined spatial and environmental effect on resident occurrences of weaker dispersers may be expected given the potential for species sorting and dispersal limitation.

Contrary to P1, the Oklahoma resident damselflies showed only a spatial effect. Previously this assemblage revealed a loosely clumped nested structure (Bried et al., 2015b), despite nestedness being potentially uncommon for wetland macroinvertebrate communities (Batzer & Ruhí, 2013). Nestedness in odonates has previously been shown to be driven by shading and water permanence gradients (McCaulay et al., 2008; De Marco et al., 2015; see also Wellborn et al., 1996; Batzer & Ruhí, 2013). We incorporated only water permanence in this analysis and it was measured categorically and not as a gradient, which may have contributed to the lack of environmental effect. Moreover, we observed extensive hydroperiod variation across the Oklahoma study sites, further suggesting that a more nuanced metric may have yielded different results.

For resident dragonflies, we predicted (P3) greater environmental than spatial effects given the potential for species sorting and strong dispersal. Results were equivocal for Oklahoma sites, perhaps reflecting Gleasonian-style (idiosyncratic) species responses (Presley et al., 2010; Batzer & Ruhí, 2013; Bried et al., 2015b), but not for New York sites, except that the spatial effect may also have been significant. The latter could reflect positive autocorrelations expected when distances among sites are smaller than dispersal ranges – very plausible within each New York preserve given the close proximity among sites. However, spatial structure may be jointly driven by multiple intrinsic and extrinsic factors (Soininen, 2016), and the ability to reach all sites may simply cancel spatial structure in odonates (Bonada et al., 2012), making interpretation difficult.

Results from the analysis of immigrants (P2, P4) supported the possibility of opportunistic or random species occurrence. Active dispersal and habitat selection are often non-random events (Edelaar, Siepielski & Clobert, 2008; Clobert et al., 2009), yet species may attempt to breed in areas unsuitable for establishing resident populations. Odonates rely on physical cues (e.g. macrophytes, standing water) over the habitat selection continuum and can fail to distinguish suboptimal conditions or ecological traps (McPeek, 1989; Corbet, 1999; Horváth et al., 2007; Šigutová, Šigut & Dolný, 2015). Thus, immigrant species may lack preadaptation for local hydroperiod, shading, predators and other within-site influences that cause resident community structure to diverge (Crumrine et al., 2008). When species co-occur as immigrants, dissociation from local conditions and spatial–environmental gradients may lead to random beta diversity structure (Bried et al., 2015b). Additionally, it may be that the set of environmental factors that affect the occurrence patterns of immigrant species as they move throughout the landscape are not the same set of factors that regulate species as persistent, multi-generation residents.

Conclusions

Ecologists increasingly recognise the importance of resident and transient states to ecological analysis and inference (e.g. MacKenzie et al., 2009; Ruiz-Gutierrez et al., 2016). Our study illustrates this at the community level both in an exploratory modelling context and for hypothesis-driven inference. Fractioning out resident species occurrences can strengthen ecological inference by reducing the number of competing or empirically supported models. Without immigrant removal, we may have falsely concluded that odonates were responding to more heterogeneity than is actually the case. This leads not only to problems with ecological inference, but also to wasted effort and expense in conservation planning and management. Resident species occurrence can instead focus conservation attention on a limited set of environmental attributes that are key to reproductive success. The focus on resident species could also improve freshwater bioassessment: a water body disturbed by humans may not attract breeding adults or sustain larval development through metamorphosis (Guzy et al., 2012), and so may support fewer resident species that are sensitive to disturbance.

© 2016 John Wiley & Sons Ltd, Freshwater Biology, 61, 1568–1579
In reality, a species’ residency state is much more complicated than a basic occurrence dichotomy across localities. Occurrence state will also vary temporally, and a resident designation may vary in degree depending on how much of a local population is natal at any given time, or on the percentage of immigrants that achieve reproductive success over time. Individuals themselves may take on multiple states that could affect dispersal (Clobert et al., 2009) and other phenotypic traits underlying a species resident-immigrant dichotomy. Indeed, mechanistic approaches in ecology often seek to link individual variation to higher level population dynamics, multi-species interactions and maintenance of community structure (Wellborn et al., 1996). The species resident-immigrant dichotomy, therefore, is really a false dichotomy for neglecting individual variation and population dynamics. Still, our work shows that any evidence of a species being natal, versus no such evidence, marks an ecologically significant boundary that scales up to influence community-level occurrence patterns in freshwater systems.

Acknowledgments

We thank Amanda Dillon and Barb Hager for assisting with dragonfly surveys in New York, along with Emily Hjalmarson, Barney Luttbeg, Nancy McIntyre, Brenda Smith-Patten and two anonymous reviewers for their comments. JTB was supported by the Albany Pine Bush Preserve Commission and by EPA Region 6 Wetland Program Development Grants; AMS was supported by NSF (DEB 1255318 and 1620046).

References


**Supporting Information**

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

**Table S1.** Determination of optimal buffer extent for landscape variables.

**Table S2.** Selection of spatial filters.

*(Manuscript accepted 13 June 2016)*