METHODS

Bat Survey Data

We utilized mist-net and acoustic records for 14 (Table 1 in manuscript) of the 16 species of bats found in Oregon and Washington, USA obtained during 2003–2010 from the “Bat Grid” interagency bat monitoring program (Ormsbee et al., 2006; Hayes et al., 2009), following a standard protocol and described by Rodhouse et al., (2012). The Bat Grid program was established as a multi-method inventory and monitoring program for bats across the two-state region, taking its name from the grid-based sampling frame used to design surveys (Scott, 2005; Ormsbee et al., 2006; Hayes et al., 2009). The grid for Oregon and Washington covered 427,156 km$^2$ and consisted of 4500 one-hundred km$^2$ (10 km x 10 km) sample units (see Fig. 1 in manuscript). Sample units were nested within larger 50 km x 50 km grid cells (not shown), and a simple random sample of units within each grid cell was drawn in order to distribute effort in a representative manner across the region. Grid cells were selected in a purposive fashion (de Gruijter et al., 2006) in an attempt to obtain adequate representation of forest and shrub-steppe habitats, which are associated with unique species assemblages (Verts & Carraway, 1998), and to accommodate ownership and accessibility constraints. The size of the 10 km x 10 km sample units was chosen to match the scale of space use by species in the region, which invariably utilize large summertime home ranges and commute long distances (e.g. kilometres) between roosting and foraging areas (e.g. Norberg, 1990; Ormsbee & McComb, 1998; Pierson, 1998; Waldien & Hayes, 2001; Lacki et al., 2007; Chambers et al., 2011; Rodhouse & Hyde, 2014), and in an attempt to ensure non-zero detection probabilities for as many of the species as possible. Coarse grain surveys and predictive models are common and suitable for bats because of their wide-ranging behaviour (e.g. Jaberg & Guisan, 2001; Milne et al., 2006; Sattler et al.,
Appendix S1 Methodological Details (Rodhouse et al., 2015)

Data were collected between 1 June and 15 September on nights where capture and recording efforts lasted ≥ 2 hours. Surveyed sample units were typically visited 1-3 times per season (12 visits maximum). Annual trainings for methods outlined in an unpublished Bat Grid protocol (P. C. Ormsbee, unpublished manuscript) were held at the start of each field season. Observers were assigned to randomly selected sample units to conduct repeat surveys at multiple locations within the sample units. In some cases non-randomly selected sample units were also surveyed because they fell within an area where bat surveys were ongoing, for example by outside researchers or for local management unit assessments. In each case, however, data contributions were accepted only if observers followed the Bat Grid protocol, ensuring a basic level of quality control. Although the survey did not follow a strictly probabilistic sampling design, we assume that the sample adequately represented conditions across the region such that regional inferences from model results were justified. Rodhouse et al., (2012) outlined an improved survey design for the program, if it were to be continued in the future (e.g. to compare against the baselines presented in the present study), and Loeb et al., (2015) adopted the recommended improvements for implementation of a continental-scale program.

Once a unit was selected, accessible water sites, meadows, and dry washes suitable for mist-netting and acoustic recording were identified. Multiple sites within units and across available habitat types (forest and shrub steppe) were visited in order to build detection histories for units that maximized the probability of detecting all species in the region. In some units, particularly in the drier portions of the study area, there were limited suitable sites for survey. Given the home range sizes of bats in the region and their nightly needs to drink water and
forage in productive riparian habitat and other openings, we assumed that each survey site within a sample unit provided approximately the same opportunities for detecting species.

Capture records were accumulated by placing mist nets over stream courses and other bat watering, commuting, or foraging routes. Captured individuals were identified using a regional key (Verts & Carraway, 1998) and supplementary identification criteria to distinguish cryptic species (e.g. Weller et al., 2007; Rodhouse et al., 2008). The Bat Grid survey protocol provided for routine collection of wing tissue for genetic confirmation of species identification as well as for the recording of voucher echolocation calls which were used with forearm length to confirm species identification in some species (e.g. Weller et al., 2007). Capture and handling procedures followed recommendations outlined by Kunz & Kurta (1988), and were consistent with updated guidance recently provided by Sikes et al., (2011).

Acoustic records for free-flying bats were accumulated by placing Petterson D240x full-spectrum time-expansion bat detectors along flight paths, sometimes in conjunction with capture operations. Call records were identified to species using an artificial neural network decision engine (Parsons & Szewczak, 2007; Redgwell et al., 2009) implemented in the Sonobat software program (version 3, Arcata, CA, available online). Sonobat 3 contains a library of ≈10,000 species-known recordings from North America, with subsets of recordings organized into regional libraries that further reduce confusion in identification among allopatric species. The regional library used for our study was populated with Bat Grid voucher calls collected during the study. The decision engine uses a broad suite of time-frequency and time-amplitude parameters to discriminate among species and resulting output includes discriminant function

1 (http://www.sonobat.com/)
Appendix S1 Methodological Details (Rodhouse et al., 2015)

probabilities for identified species. We only accepted call records that had \( \geq 0.95 \) probability of identification consensus among the range of decision classifiers.

Distribution Models

We developed multi-season occupancy models using an autoregressive parameterization described by Royle and Dorazio (2008). This approach is convenient for incorporating temporal dependence with environmental predictors, enabling us to estimate occurrence probabilities and related dynamic occupancy rate parameters, as well as to make predictions throughout the region for purposes of mapping. The hierarchical occupancy modelling approach permits the full multi-parameter likelihood to be decomposed into a product of hierarchical levels, with each level representing an explicit state (e.g. occupancy) or process (e.g. observation; Cressie et al., 2009).

We considered the elements of our detection history matrix as a vector of observations denoted \( y_j(i,t) \), each taking a value of 1 or 0 (or “NA” for missing values) representing “detection” or “non detection” for survey \( j \) (from 1 to 12), made in sample unit \( i \) (1 to 4500, or 1 to 241 for models not used for predicting values at unobserved sample units), during year \( t \) (2003-2010). We denote the latent, partially observed occupancy state as \( z(i,t) \), and let \( \psi_t = \Pr(z(i,t) = 1) \) represent the occupancy probability. Canonical dynamic occupancy rate parameters include \( \phi_t \), the probability of “survival” for each sample unit from \( t \) to \( t+1 \), where \( \phi_t = \Pr(z[i,t+1] = 1 | z[i,t] = 1) \), and \( \gamma_t \), the probability of “colonization” or recruitment from \( t \) to \( t+1 \), where \( \gamma_t = \Pr(z[i,t+1] = 1 | z[i,t] = 0) \). The probability of “extinction” is \( 1-\phi_t \), and with these parameters the dynamic occupancy model can be built. We denote the initial occupancy state as \( z(i,1) \sim \text{Bern}(\psi_{2003}) \), a Bernoulli trial, akin to a “coin flip” but where the “coin” is weighted by \( \psi_{2003} \). Occupancy states
in subsequent years are denoted as \( z(i,t+1) | z(i,t) \sim \text{Bern}(\pi[i,t+1]) \), where \( \pi(i,t+1) = z(i,t) \phi + (1-z[i,t]) \gamma \). It is in this way that the probability of occurrence “evolves” over time as a first-order Markov process, thereby accounting for temporal autocorrelation and providing a way to model the dynamics in distributional patterns and to estimate trend in \( \psi \). In our study we focused on \( \psi_{2010} \), the probability of occurrence in 2010, for mapping distributions. This was the most recent year represented in our analysis, and \( \psi_{2010} \) provided a fully dynamic view on the contemporary status of the species in the region, conditioned on occupancy states of the previous 7 years. We estimated the “growth rate” in the probability of occurrence as \( \lambda_t = \psi_t / \psi_{t-1} \) and net trend as \( \psi_{2010} / \psi_{2003} \) (MacKenzie et al., 2003; Royle & Dorazio, 2008). We also considered the amount of year-to-year variability by estimating turnover, the probability that an occupied sample unit is a newly occupied one (i.e., \( \Pr(z[i,t-1] = 0 | z[i,t] = 1) \)), as

\[
\tau_t = \frac{\gamma_{t+1} \cdot (1 - \psi_{t-1})}{\gamma_{t+1} \cdot (1 - \psi_{t-1}) + \phi_{t+1} \cdot \psi_{t-1}} \tag{1}
\]

Nichols et al., 1998; Royle & Dorazio, 2008).

In order to accommodate imperfect detection, we explicitly modeled the observation process as conditional on the latent occupancy state, \( y_j(i,t) | z(i,t) \sim \text{Bern}(z[i,t] \cdot p_{ij}) \), where \( p_{ij} \) is the probability of detection during survey \( j \) given that presence in unit \( i \) during year \( t \). An important assumption for this model is that a species is available for observation during the entire annual survey period from June 1 to September 15 (closure), a reasonable assumption in our study region. However, assuming constant probability of detection throughout that entire annual survey period was less tenable, and we allowed detectability to vary linearly (on the logit scale) across the period by including the survey Julian date as a covariate. This represented our expectation that detectability would increase with the progression of summer, particularly as
juveniles became volant. We also included the duration of each survey, in hours, as a covariate to account for variable effort, and an indicator variable to represent the use of acoustic detectors, with a resulting detection model \( \text{logit}(p_{ijt}) = a_0 + a_1*\text{date}_{ijt} + a_2*\text{duration}_{ijt} + a_3*\text{I(acoustic}_{ijt}), \)
where \( p_{ijt} \) represents the probability of detection in sample unit \( i \) during survey \( j \) and year \( t \), as a function of survey date, duration, and method (i.e., an indicator variable for surveys when acoustic detectors were used). We expected that detectability would increase with the progression of summer, particularly as juveniles became volant, and would be higher during long survey duration (e.g. 4 hours instead of 2 hours) and when acoustic detectors were employed.

We used an autoregressive reparameterization of the dynamic occupancy model in order to incorporate the additive effects of environmental predictors on the logit-linear scale (Royle & Dorazio, 2008). Specifically, We modelled the probability of occurrence as \( \text{logit}(\psi_t) = a_t + b_t*\text{z}(i,t-1) + \beta_1*X_1 + ... + \beta_k*X_k, \) in which \( \psi_t \) represents the occurrence probability in year \( t \), \( a_t + b_t*\text{z}(i,t-1) \) represents the autoregressive parameterization for colonization and survival probabilities of each sample unit \( i = 1 \) to \( n \), and \( \beta_1*X_1 + ... + \beta_k*X_k, \) represent the additive effects of environmental predictors \( 1 \) to \( k \) associated with each sample unit on probability of occurrence.

Note that interactions between the dynamic parameters and environmental covariates can also be modeled (e.g. \( \beta_k*X_k*\text{z}(i,t-1); \) Royle & Dorazio, 2008). We did not do this in the present study because of sample size and overparameterization constraints and because of the generally low rates of observed turnover (see Results in the study). However, we have discussed motivating scenarios (Rodhouse et al., 2012) when more data become available from additional monitoring and when motivating hypotheses are advanced about co-variation between temporal dynamics of bat distribution patterns and influential factors such as climate. For example, Frick et al., (2010) demonstrated that little brown bat survival probabilities in a single maternity colony covaried
over time with cumulative precipitation during summer prior to the onset of hibernation. They attributed this to increased production of insect prey, which in turn enabled bats to enter winter hibernation with higher fat reserves. It would be interesting to assess whether such dynamics are evident at regional scales with our models. We emphasize the need for longer time series however, because of the inherently low rates of turnover among bat populations (see Results from this study).

An additional consideration is that patterns of spatial autocorrelation in bat SDMs are likely, at least for some species, due to colonial roosting behaviour and the influence of environmental factors not included in SDMs that are themselves spatially-structured. However, it is unlikely that residual spatial structure will persist in regional bat distribution models if the observation process (grain size) is coarser than the process(es) driving spatial autocorrelation. Previously (Rodhouse et al., 2012) we estimated and reported small (< ±0.20) and non-significant (P > 0.10) Moran’s I correlation coefficients from model residuals (Cliff & Ord, 1981; Wintle & Bardos, 2006) of a subset of the data used here for the little brown bat, suggesting limited or no spatial autocorrelation at this scale. We estimated Moran’s I correlation coefficient (Cliff & Ord, 1981; Wintle & Bardos, 2006) for each set of model residuals, using a 2nd-order neighbourhood structure that identified the 8 units adjacent to each unit i as neighbours. We computed residuals by subtracting the estimated probability of observing a species (little brown bat in this case) at least once,

\[
\psi_{it} \ast [1 - \prod_{j=1}^{j} 1 - p_{ij}],
\]

from observed values (either 0 or 1; Moore & Swihart, 2005).
Appendix S1 Methodological Details (Rodhouse et al., 2015)

As an exploratory step, Rodhouse (2011) also considered an extension of our model by including a spatial autologistic covariate (Hoeting et al., 2000; Wintle & Bardos, 2006; Royle & Dorazio, 2008). Hoeting et al., (2000) had reported encouraging results regarding the predictive performance of a spatial autologistic model with environmental covariates; Wintle and Bardos (2006) outlined the steps to the process that we followed. To construct a fully spatio-temporally dynamic multi-season occupancy model, we included the additive effect $\beta_2^*x(i,t-1)$, representing the influence of occupied neighbouring sampling units. Here, $x(i,t-1) = \left( \sum_{j \in N_i} z(j, t-1) \right)^{\bar{n}_i}$, the local density of occupied neighbours of site $i$ in time $t-1$ such that a model with this component provides information for estimating $\beta_2 = \Pr(z[i,t] = 1 | x[i,t-1])$. Anticipating that the sparseness of observations across the entire lattice may yield uninformative results for the full spatial model, particularly if residual spatial signal was weak as we had anticipated due to the coarse scale of the lattice, we also considered a single-season spatial model, where $\text{logit}(\pi_i) = a + \beta_1^*\text{Forest Cover}_i + \beta_2^*x(i)$ and $x(i) = \left( \sum_{j \in N_i} z(j) \right)^{\bar{n}_i}$. In this case $x(i)$ represents the density of occupied neighbors over the entire study period, requiring an assumption of temporal stationarity over the study period. Results of fitting the multi-season spatial autologistic model were uninformative, with posteriors of the auto-logistic parameter filling the entire prior parameter space, likely reflecting a lack of residual spatial structure and perhaps also due to sparseness of the observations across the grid (lattice). However, the single-season spatial model with forest cover did provide more reasonable results (not shown), with the autologistic parameter estimate converging on a slightly positive value but with a wide CRI that included 0. The study was not designed to model spatial process; typically with lattice data a much higher density of
Appendix S1 Methodological Details (Rodhouse et al., 2015)

observations using a cluster-type sampling design would be employed (Irvine et al., 2013). It is not evident from our study whether the lack of spatial structure is a result of the relatively sparse detection history matrix, coarse grain size, or other factors. Accounting for the possibility of spatial structure should be routine in these kinds of modelling exercises, however (Wintle & Bardos, 2006). This appears to be an important area of opportunity for future bat SDM studies when more richly populated datasets become available.

We used OpenBUGS 3.2 (Lunn et al., 2000), launched from R 3.1.0 (R Development Core Team, 2014) with the R2OpenBUGS library (Sturtz et al., 2005) to implement Bayesian estimation of model parameters via Markov chain Monte Carlo (MCMC) samples of posterior distributions. We used independent vague $N(0,10)$ priors on all logit-scale parameters and a $U(0,1)$ prior for $\psi$, bounded by 0 and 1 on the probability scale (Royle & Dorazio, 2008; Ntzoufras, 2009). We explored whether model results were sensitive to choice of prior by assigning non-informative $U(-10,10)$ priors to all parameters but found no meaningful differences. Posterior summaries were based on 40 000 MCMC samples of the posterior distributions from 3 chains run simultaneously, thinned by a factor of 10, following an initial burn-in of 5000 MCMC iterations. We assessed convergence of MCMC chains with trace plots and the Gelman-Rubin diagnostic, $\hat{R}$; convergence was reached for all parameters according to the criteria $|\hat{R} - 1| < 0.1$ (Ntzoufras, 2009). We compared forest and snag models using the Deviance Information Criterion (DIC), a Bayesian analogue to the more commonly used Akaike’s Information Criterion (Spiegelhalter et al., 2002).

Bayesian posterior predictive $P$-values were generated as a measure of the goodness-of-fit for the models (Ntzoufras, 2009; Kery & Schaub, 2012). These were estimated from the discrepancy between the sum of absolute values of model residuals from observed data and those
from a simulated posterior predictive dataset under the assumed model (Ntzoufras, 2009; Kery & Schaub, 2012). *P*-values near 0.5 indicate that the model generated data similar to the observations, providing a good fit to observed data, whereas *P*-values near 0 or 1 indicate poor fit. In order to speed MCMC convergence and improve interpretability, we input centred and standardized covariate inputs to improve computation and interpretability of the detection model parameters (Gelman & Hill, 2007).

Statistical models differ in their power to explain and predict ecological phenomena (Shmueli, 2010), and the measures of convergence, DIC, goodness-of-fit, and parameter uncertainty used to gain insight into explanatory performance do not reflect predictive performance. To evaluate the predictive performance of models, we used leave-one-out cross-validation and estimated the area under the curve (AUC) of the receiver operating characteristic (Fielding & Bell, 1997; Ntzoufras, 2009). The receiver operating characteristic curve plots false-positive prediction rates against true-positive prediction rates. AUC scores range from 0.5 to 1.0 and represent the percentage chance that a randomly selected unit where the target species was encountered at least once will have a higher occurrence probability than a unit where the target species was not encountered at all during the study. An AUC score of 0.5 indicates that the model discriminated among sites no better than random chance, whereas a score of 1.0 indicates that the model discriminated among sites perfectly. To do this, we obtained a posterior predictive distribution for each sample unit via the leave-one-out process, and then calculated a posterior distribution for AUC by comparing each posterior prediction to observed data and computing AUC. This Bayesian approach provided a fuller measure of uncertainty in AUC compared to just using the parameter posterior mean values for one predictive value for each sample unit. Because observations were made only on conditional detections (i.e., the latent occurrence state is only
Appendix S1 Methodological Details (Rodhouse et al., 2015)

...partially observed), we compared observations to predictions on the probability of detection rather than on probability of occurrence directly. We used the ROCR library in R to estimate AUC (Sing et al., 2009).

LITERATURE CITED


Appendix S1 Methodological Details (Rodhouse et al., 2015)


Appendix S1 Methodological Details (Rodhouse et al., 2015)


Appendix S1 Methodological Details (Rodhouse et al., 2015)


