Journal of Biogeography

WILEY

Journal of Biogeography

RESEARCH ARTICLE

Invisible Hand of Sampling for Management: Underlying Needs to Survey a Threatened Seabird Can Bias Aggregated Data

Ryan Baumbusch[1,2] | Adam Duarte[1] | James T. Peterson[3]

[1]U.S.D.A. Forest Service, Pacific Northwest Research Station, Olympia, Washington, USA | [2]Oregon Cooperative Fish and Wildlife Research Unit, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, Oregon, USA | [3]U.S. Geological Survey, Oregon Cooperative Fish and Wildlife Research Unit, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, Oregon, USA

Correspondence: Ryan Baumbusch (ryan.baumbusch@oregonstate.edu)

Received: 8 July 2024 | **Revised:** 24 October 2024 | **Accepted:** 25 November 2024

Funding: This work was supported by the U.S.D.A. Forest Service, Pacific Northwest Research Station.

Keywords: big data | *Brachyramphus marmoratus* | found data | marbled murrelet | occupancy trends | sampling bias

ABSTRACT

Aim: Surveying for a species of concern ahead of proposed activities that alter its habitat is routine practice in conservation and management. Such surveys may accumulate large datasets that could further elucidate trends in abundance and distribution. However, the as-needed surveying of proposed activities may impart a sample site selection bias on the data if used for another purpose. Management of a threatened, forest-nesting seabird offered an example of this. Here we assessed how resource management planning and survey requirements can bias clearance monitoring survey data collected prior to proposed timber harvests, if those data are used for other purposes.

Location: Oregon and Washington, USA.

Taxon: Marbled Murrelet (*Brachyramphus marmoratus*).

Methods: To assess how timber planning and other factors influenced marbled murrelet survey location selection, we used logistic regression models to examine habitat associations of marbled murrelet survey sites ($n = 9178$) encompassing proposed timber harvests, and the survey stations ($n = 38{,}923$) therein, across the murrelet's inland range in Washington and Oregon, USA between 1989 and 2021. We then simulated the effect this selective sampling might have on assessments of occupancy trends.

Results: Most habitat characteristics considered did influence where surveys were located, with distance to roads often being the strongest predictor of survey location. The strength of selection for each covariate changed over time, such that a habitat characteristic strongly influenced location selection in a year but was less influential in another year. The simulation analysis suggested that the non-random selection of survey sites could profoundly bias assessments of occupancy trends.

Conclusions: When using these clearance monitoring survey data– or any data–beyond their original purpose, careful consideration should be given to the scope of inference provided and analytical methods used, to ensure that observed trends are the product of biological processes and not biased by sampling artefacts.

1 | Introduction

Big data in ecology increasingly allows us to answer ever-broadening questions in biogeography (Hampton et al. 2013; Soltis and Soltis 2016; Wüest et al. 2019). These can range from understanding fundamental ecological patterns and processes (Allen et al. 2021; Nathan et al. 2022) to addressing critical conservation and management needs (Franklin et al. 2017; Van Doren and Horton 2018). In ecology and other fields, big data often consist of pre-existing data repurposed to study phenomena other than what they were originally collected, defined as found data (Overton, Young, and Overton 1993). Given that targeted broadscale and long-term monitoring efforts are lacking for most fish and wildlife species, pre-existing found data potentially allow for rapid assessment of status and trends in distribution and abundance at spatial and temporal scales that may otherwise be unfeasible to investigate on timeframes needed by managers, policymakers, and imperilled species (Moritz et al. 2008; Rosenberg et al. 2019). However, across disparate fields such as medicine (Kaplan, Chambers, and Glasgow 2014) and transportation (Griffin et al. 2020), caution is urged when analysing big data, as original sampling designs may impart systematic bias in the data when used for a new purpose. For example in avian ecology, the citizen science platform eBird (Sullivan et al. 2014) and the North American Breeding Bird Survey (Sauer et al. 2017) offer enormous volumes of species encounter data collected with haphazard to highly structured sampling designs; yet both contain strong roadside biases that, when used to quantify status and trends in distribution and abundance, could produce misleading results if consideration is not given to how or why data were sampled (Betts et al. 2007; Tang, Clark, and Gelfand 2021; Robinson et al. 2022).

A common practice in fish and wildlife conservation and management is to survey for species of concern ahead of planned activities that may harm the species or its habitat (i.e., resource extraction, construction, etc.). These pre-activity surveys (hereafter, clearance monitoring surveys) are often done to satisfy regulatory requirements such as the U.S. Endangered Species Act, European Union Directive 2011/92/EU, or Chilean Law No. 19.300 on General Environmental Bases. At-risk species detections may prompt decisions to mitigate, relocate, or cancel the project, made internally or in conjunction with regulators, but can also lead to costly litigation and possible cancellation, such as the Río Cuervo dam in Patagonia, Chile (Republica de Chile, Tercer Tribunal Ambiental, Case R-42-2017). Project planners can attempt to avoid delays and minimise risk to the species by using knowledge of a focal species' habitat associations to locate activities in areas with poor or no habitat and a low probability of species presence, for example, placing wind turbines away from predicted vulture flight paths (de Lucas, Ferrer, and Janss 2012). Clearance monitoring surveys can accumulate data sets that can be repurposed and analysed to improve management decisions (e.g., Ferrer et al. 2012) or broaden ecological knowledge (e.g., Bizama et al. 2024). However, these survey and sampling protocols are often strictly designed for determining species presence and are implemented as needed in and around proposed activity areas. Given the above considerations of project planning, the associated clearance monitoring surveys are systematically sampling areas suitable for the proposed activity, which may not reflect the broader landscape or available habitat of the focal species. This systematic selection of survey locations could profoundly affect the inferential space of these data if used to address questions regarding the species' distribution and abundance (Hurlbert 1984).

In the USA Pacific Northwest region, timber harvest is a major natural resource industry. Increased harvest activity through the late-20th century prompted concerns over habitat loss and degradation for species that relied upon the dwindling stands of late-successional, old-growth forests, which motivated natural resource management plans such as the federal Northwest Forest Plan (NWFP; U.S. Department of Agriculture Forest Service, and USDI Bureau of Land Management 1994) and the Interior Columbia Basin Ecosystem Management Project (Thomas and Dombeck 1996). These efforts focused on developing management strategies for threatened, endangered, and other at-risk species, and initiated clearance monitoring surveys to better understand the ecology of these species and to reduce further habitat loss. The marbled murrelet (*Brachyramphus marmoratus*, hereafter murrelet), a forest nesting seabird, was one of the primary species listed in the NWFP due to its reliance on late-successional, old-growth forest for nesting habitat (US Fish and Wildlife Service 1992). Standardised survey protocols were drafted (Paton et al. 1990) and updated (Ralph and Nelson 1992; Ralph et al. 1994; Evans Mack et al. 2003; Pacific Seabird Group 2024) with the expressed purpose of determining whether murrelets were nesting in and adjacent to proposed timber harvest areas. Surveys conducted under these protocols amassed large data sets from a variety of entities (e.g., state and federal land management agencies, private timber companies, etc.) of this threatened seabird's inland distribution, which is otherwise extremely difficult to study. These data have in turn provided valuable insights to the broader ecology and conservation of murrelets (e.g., Betts et al. 2020; Valente et al. 2023). However, given these data were collected for timber management purposes, their sampling across the landscape may represent locations with harvestable timber and pre-defined nesting habitat rather than a range of possible murrelet nesting habitat conditions.

Here we used murrelet clearance monitoring survey data collected across Oregon and Washington over three decades to assess how resource manager behaviour in selecting survey locations determined the underlying distribution of where and when a species is sampled for conservation and management. Specifically, we quantified habitat associations and how these associations changed over time at two spatial scales: survey sites encompassing proposed activity areas (mostly timber harvests) across the broader landscape and survey stations at which the surveys were conducted within survey sites. We predicted that features associated with murrelet nesting habitat (e.g., late-successional, old-growth forest) would be preferentially selected in survey locations, but that selection of these features (especially at the site-level) may decline over time with availability for sampling or avoidance in timber management planning. We also predicted that features associated with timberlands and the survey protocols (e.g., proximity to roads and relatively open canopies) would be preferentially selected, but these would change with policies or protocols. The present research focuses solely on where the murrelet surveys were conducted, and not on where murrelets were detected during those surveys.

Furthermore, we simulated old-growth and canopy cover associates and compared occupancy estimates derived from random sampling and sampling following the survey data to assess potential impact of how these survey locations are selected on distribution trends. The clearance monitoring surveys intended to protect murrelet nesting habitat from timber harvests offered an example of how systematic selection of survey locations for one purpose can impart bias in data when used for another purpose and demonstrated the challenge of analysing big data in ecology obtained from heterogeneous and autonomous sources (Wüest et al. 2019).

2 | Methods

We used murrelet protocol survey data compiled by three natural resource agencies in the Pacific Northwest, USA: the US Bureau of Land Management (BLM) and Oregon Department of Forestry (ODF) in Oregon, and the Washington Department of Fish and Wildlife (WDFW) in Washington. These data are spatiotemporally expansive, extending throughout much of the murrelet range in their respective states (Figure 1a), and collected over multiple decades, primarily in conjunction with timber harvest planning and other management actions. The WDFW data set was a comprehensive collection of murrelet protocol surveys conducted for management actions in Washington by or for a variety of government agencies and private companies. Oregon did not maintain a similar comprehensive database, and while ODF and BLM constituted a large number of surveys in Oregon on the lands they respectively managed, data sources from other entities were unavailable or unknown. Data were prepared and analysed using QGIS (QGIS Association 2023) and Program R (R Core Team 2021). We examined the data for possible erroneous records with exploratory data analysis plots and summaries, then consulted with the respective agencies to correct or omit records.

2.1 | Survey Protocol

The murrelet inland survey protocol and modifications are thoroughly documented by Ralph et al. (1994) and Evans Mack et al. (2003) and summarised here. Proposed timber harvests in the murrelet range were initially assessed by on-site

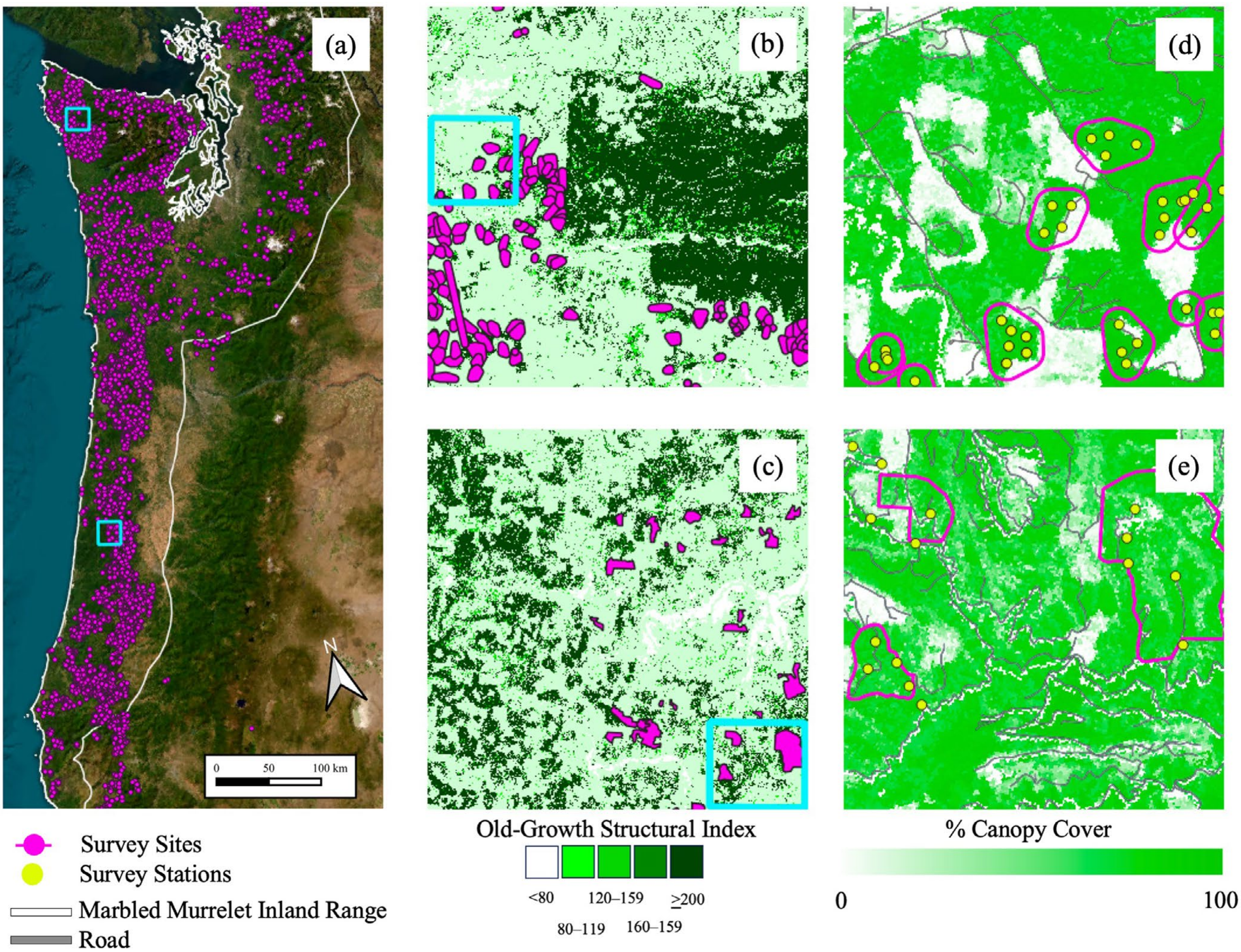

FIGURE 1 | (a) Marbled murrelet (*Brachyramphus marmoratus*) survey sites in relation to the murrelet inland range in western Washington and Oregon, USA. Insets are example survey site locations in (b) the Olympic Peninsula of Washington and (c) central Coast Range of Oregon in relation to old growth structural index. Further insets show the distribution of survey stations in and around survey sites in relation to canopy cover and roads in Washington (d) and Oregon (e).

walk-throughs as to whether the proposed location contained "any area with a residual large tree component, small patches of potential habitat, or suitable nest platforms" and warranted murrelet surveys. When warranted, a survey site was established containing at minimum the potential habitat within and near the proposed project area. Surveys were conducted at stations within survey sites (Figure 1d,e), and distributed such that a 200 m radius circle (maximum expected detection distance) around each would cover a site. Protocols required a minimum number of surveys be conducted at a site for two consecutive breeding seasons, but surveys could be discontinued if behaviours indicative of nesting were detected. If breeding behaviours were detected, the site was often afforded protections from forest harvest or other disturbances. We analysed habitat selection of sites and stations separately.

2.2 | Survey Data

ODF and BLM databases included locations of sites and stations, while the WDFW database only maintained station locations. All surveys were conducted under the same protocols, but agencies had some flexibility in their applications; for example, ODF strove to complete the expected nine surveys at a site each year even if the protocol allowed surveys to be discontinued. We approximated survey site boundaries in Washington by fitting minimum convex polygons around survey stations for each site, buffered these polygons by 200 m, and removed polygons larger than sites in Oregon (213 ha). We used the first year a site or station was surveyed for habitat attributes (e.g., forest conditions) and analysis. We analysed Oregon and Washington data sets separately, given the differences in site data, statewide land management practices, and landscape compositions. In Oregon, 23,270 unique stations were surveyed across 5656 unique sites between 1990 and 2021, while in Washington there were 15,653 unique stations that were surveyed at 3522 unique sites between 1989 and 2021.

Additionally, to prepare a habitat conservation plan the Washington Department of Natural Resources (WDNR) surveyed murrelets across their lands over several years and often laid out stations in regular grid patterns. To assess if WDNR's sampling approach reduced bias in the WDFW data, we conducted a post hoc analysis–detailed in the Supporting Information–without WDNR and US Fish and Wildlife Service survey locations.

2.3 | Background Data

To assess the degree of selection for survey stations given the habitat available in survey sites, we randomly selected background points from within survey sites. To assess the degree of selection for survey sites given the habitat available within the murrelet's range, we created hexagonal tessellations over the murrelets inland range in each state, where their respective grids cells were equal in area to the average survey site in Oregon (24.0 ha, SD = 20.1) and Washington (40.1 ha, SD = 24.1). We removed hexagons with excessive non-forest cover (e.g., agricultural or urban land cover; see Supporting Information for details) and randomly selected background sites from the remaining grid cells. For both stations and sites, we selected three times as many background locations as survey locations, resulting in the sample size of modelling data being four times larger than the survey data listed above. Background stations were attributed the year of the survey sites they were drawn from, and we drew background sites in proportion to the number of sites surveyed each year.

2.4 | Model Covariates

The murrelet survey protocols required that continuous tracts of late-successional, old-growth forests be surveyed ahead of timber harvests, but attributes related to such forests are also of primary interest in murrelet nesting habitat studies (e.g., Wilk, Raphael, and Bloxton Jr. 2016; Cosgrove et al. 2024; Duarte et al. 2024). We included several forest metrics as covariates (Bell et al. 2023) to assess the degree to which survey locations selected for or against these features. Old-growth structural index (OGSI) 80 and 200 are binary indicators for whether a forest patch is similar in structure to a forest that is ≥ 80 or ≥ 200 years old, respectively (Bell et al. 2023). We summarised these as the percent of area comprised of OGSI 80 or OGSI 200 patches. We included covariates for average percent canopy cover and average conifer density, and a measure of edge density calculated as the total length of edge between patches > and ≤ 40% canopy cover divided by the total area, multiplied by 10,000. We predicted that sites would select for OGSI 80 and 200, canopy cover, and against conifer density and edge density. Stations should represent the habitat available within sites, however, protocols recommended placing stations near canopy openings to facilitate visual detections, thus we predicted selection against canopy cover and for edge density around stations. Roads are necessary for transporting timber out of harvest units, but can also reduce canopy cover and increase habitat edges. Placing stations near roads can reduce time and effort invested in surveys, while increasing safety for field crews and providing the recommended canopy openings for visual detections. Thus, we included a covariate for the distance to the nearest road from a station or beyond a site boundary. We predicted sites and stations would select against distance to road, indicating both were located near roads. Historical logging practices may have removed old-growth forests from low elevations (Hull et al. 2001). Murrelets, however, are thought to transit inland to nest sites through valley bottoms (Lorenz et al. 2017). We included covariates for the elevation at stations, and the average and standard deviation elevation, the latter representing terrain ruggedness. We also calculated topographic position index for stations as station elevation minus the average elevation. Selection for elevation metrics may indicate that high-altitude, rugged terrain is where late-successional, old-growth forests remained, however, selection against elevation metrics may indicate that logging operations still tend to avoid these topographies. Selection for elevation metrics at stations should be weak, but could indicate tendencies to avoid windy ridges or running water, or for certain vantage points on slopes. Unless noted otherwise, covariates were summarised within site boundaries and within the 200 m radius of station points. Details on data sources are provided in the Supporting Information along with summary statistics (Tables S1 and S2).

While selection for or against landscape attributes would reflect a systematic bias in surveys locations, changes in the strength or direction of selection over time would suggest that bias is inconsistent, which could obscure, exaggerate, or produce erroneous trends of distribution or abundance. Changes in selection over time could result from a multitude of factors related to the harvest and survey planning processes. For the period of available data, major survey protocol changes were implemented in 1994 (Ralph et al. 1994) and 2003 (Evans Mack et al. 2003). Importantly, the definition of potential habitat was expanded from mature and old-growth stands (often referred to as late-successional, old-growth forest) in an early protocol (Ralph and Nelson 1992) to younger coniferous forests that have deformations or structures suitable for nesting (Ralph et al. 1994), and finally to also include any stand with a platform structure (Evans Mack et al. 2003), which we expected to reduce selection for late-successional, old growth forest features. Additionally, the NWFP was also implemented in 1994, affording protections to late-successional, old-growth forests on federally managed lands in the murrelet's range. We created a three-level categorical indicator variable covariate–representing the periods prior to 1994, between 1994 and 2003, and after 2003 (baseline state)–to assess whether these policy and protocol changes impacted the habitat features associated with survey locations. We included average annual lumber prices, adjusted for inflation, as interactions with habitat covariates in site selection models (Table S1), expecting higher prices may increase selection for murrelet habitat features indicating a willingness to risk project cancellation if murrelets were detected. Since harvest planning is a long process, we considered the lumber prices in the year of the survey (t−0) and each of the preceding 5 years (t−1 to t−5). We did not include lumber price in station-level models because we had no *a priori* hypothesis that this variable would influence where murrelets were surveyed within sites. Numerous other factors may change habitat compositions in and around surveys from year to year, thus we considered year as a random effect with annually varying slopes for each habitat covariate. Continuous covariates were standardised to a mean of zero and standard deviation of one, with resulting standardised model parameter estimates corresponding to a one standard deviation change in the covariates (Gelman and Hill 2007).

2.5 | Modelling and Model Selection

We fit generalised linear mixed effects logistic regression models to evaluate evidence that habitat characteristics of murrelet clearance survey data differed from the broader landscape using the R package glmmTMB (Brooks et al. 2017). Here the response variable modelled was the murrelet clearance survey locations (i.e., survey locations equaled one, with background locations equaled zero) and the resulting models estimated the probability of selecting a site or station for murrelet clearance surveys. Habitat covariates were included as linear additive effects and interactions, with random effects that corresponded to year. Models were compared with Akaike's Information Criterion (AIC; Akaike 1973) with the small-sample bias adjustment (AICc; Hurvich and Tsai 1989). The model with the lowest AICc score was considered the best approximating model.

We analysed four data sets consisting of station- and site-level data from Oregon and Washington. Model selection for each data set took place in four stages (Morin et al. 2020) to minimise the number of models in any set (Anderson 2008), where the model with the lowest AICc from one stage was used as the base model in the next stage. First, we selected a global model. For any pair of covariates correlated with Pearson $|r| > 0.7$ (Moore and McCabe 1993) we fit two models, each using one of the correlated covariates. For site data, we also fit six models to select lumber price between year t−0 and t−5. Each model in a model set contained one covariate from correlated pairs, one lumber price lagged 0–5 years, and all uncorrelated covariates, e.g., a data set of survey sites with two pairs of correlated parameters would have a set of 24 models, four models for each combination of correlated parameters times six for a model with one of the lumber prices by year. Second, we evaluated the random effects structure, including year as a randomly varying intercept and allowing the slope of each habitat covariate to vary by year (Bolker 2015). The model set was comprised of the global model with all combinations of habitat covariates in the random effects structure. At the third stage we evaluated whether any covariate not included as a random effect should be retained as a fixed effect, with the model set comprised of models that included and excluded fixed effects of covariates not included as random effects. Lastly at the fourth stage, we evaluated fixed effect interactions. This model set was comprised of all combinations of the time-period interaction with all remaining fixed effect covariates. For site-level models, an additional model set was evaluated with all combinations of the lumber price interaction with the remaining fixed effect covariates. The precision of parameter estimates in the best approximating models were evaluated by calculating 95% confidence intervals. We also calculated odds ratios (Hosmer, Lemeshow, and Sturdivant 2013) to aid in the interpretation of fixed effect parameter estimates and plotted the covariate random effects to facilitate their interpretation. If survey site and station locations represented a random sampling of the available habitat, few covariates would be included in the best approximating models, and those covariates included would have weak effects.

2.6 | Simulating Effects of Sampling on Old Growth Forest Obligates

We used stochastic simulation to assess whether sampling survey sites similarly done for murrelets would bias annual trends in occupancy estimates for two, hypothetical species. We then compared these results to a monitoring program that randomly sampled the landscape. We defined our sampling universe–within which species occurrence was simulated and sampled from–as all potential background hexagons in Oregon from 1991 to 2020 that did not exceed the non-forest thresholds, as described in the Supporting Information ($n = 842{,}375$ hexagons over the 30 years). We used Oregon as a demonstrative example and would expect qualitatively similar results in Washington. We simulated two types of species, an old-growth associate and a canopy cover associate, which were each respectively more likely to be present (i.e., > 0.50 occupancy probability) in hexagons with $> 75\%$ OGSI 80 cover or $> 75\%$ average canopy cover (Figure 2a). The probability (p) that a species was present in a hexagon (i) was calculated as:

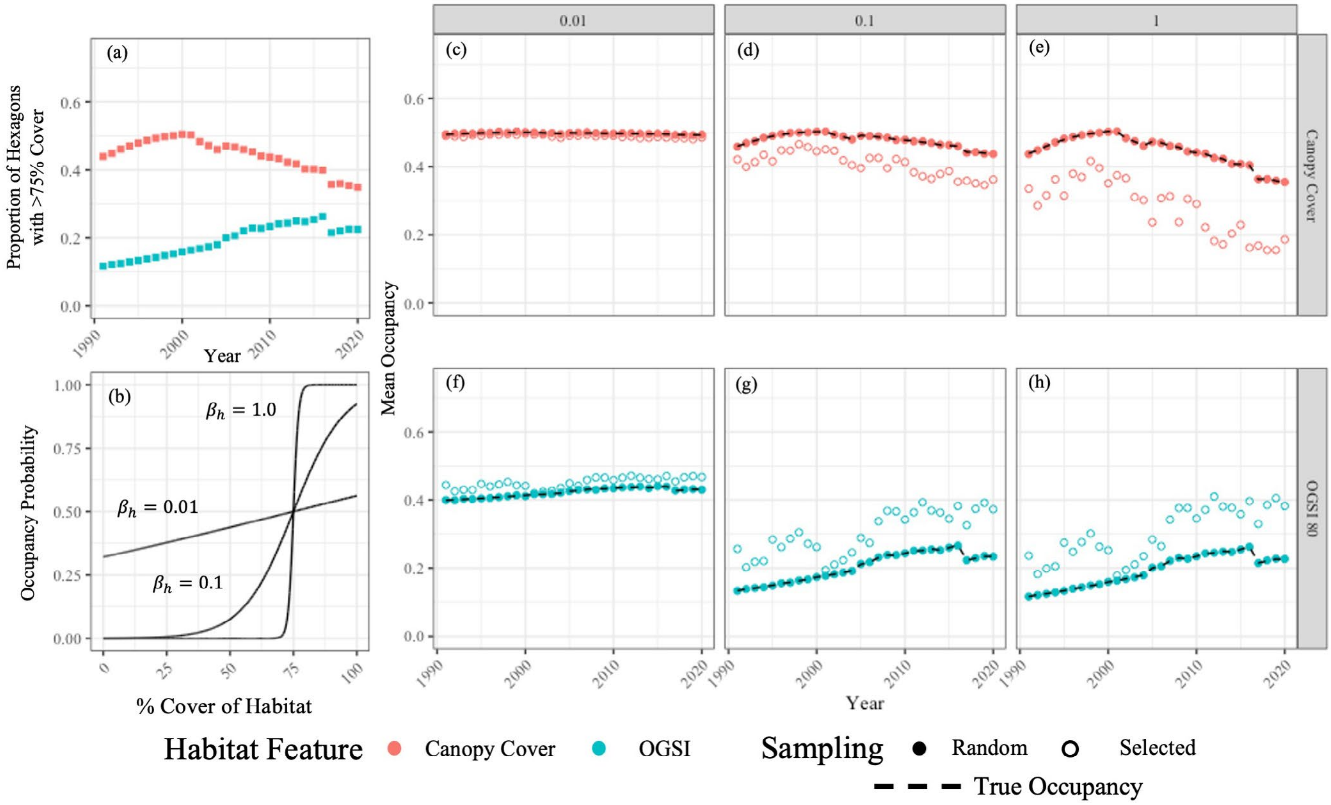

FIGURE 2 | Simulation of estimated occupancy for old-growth and canopy cover associated species between 1991 and 2020 under random sampling and sampling following marbled murrelet (*Brachyramphus marmoratus*) survey site selection in Oregon. (a) The proportion of hexagons available for sampling with > 75% 80-year, old-growth structural index (OGSI 80) cover and > 75% average canopy cover. (b) The probability that a hexagon is occupied by the given habitat associated species with a weak (slope = 0.01), moderate (slope = 0.1), and strong (slope = 1.0) relationship with habitat (β_h), where each has a 0.50 probability of being present at 75% cover. Mean annual occupancy estimates averaged over 1000 simulation iterations for the canopy cover (f–h) and old-growth (c–e) associates with weak (c, f), moderate (d, g), and strong (e, h) associations with habitat under random sampling and sampling sites following marbled murrelet site selection in Oregon.

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_h X_i)}} \quad (1)$$

where X is percent cover of the habitat type, β_0 is the intercept, and β_h is the log odds of species presence associated with habitat (h). The strength of the relationship between presence and habitat was simulated using three levels of association: weak, moderate, and strong with β_h at 0.01, 0.10, and 1.0, respectively. The intercept differed with each level so that occupancy probability was 0.5 at 75% cover (Figure 2b). The weak relationship represented a habitat generalist with only a slight preference for the habitat feature, while the strong relationship represented a habitat obligate found nearly anywhere a minimum of 75% cover was available. We simulated species associated with these two habitat features because they are closely linked with murrelets and other late-successional, old-growth forest associated species. Simulating species associated with only a single habitat feature more simply and clearly illustrated the impacts of this sampling scheme with respect to bias in the estimated trends in the distribution of species.

We began the simulations by estimating the probability that a hexagon was surveyed using the best approximating model for Oregon survey site selection and the probability of occupancy for each species using the logit-linear function described above (Equation 1). Each simulation began with randomly assigning occupancy of each species to each hexagon for each year assuming a Bernoulli distribution with probability of success as the estimated occupancy probability. For the random sampling design, we randomly selected 175 hexagons each year (equivalent to the number of new sites surveyed annually in Oregon) with the slice_sample function from package dplyr (Wickham et al. 2023). For the systematic sampling design (the design that matched the murrelet clearance monitoring data), we randomly selected 175 hexagons each year with slice_sample, weighted by the probability that they were surveyed. In both cases we sampled without replacement within years, but hexagons could be resampled in different years. This process was repeated for 1000 replicates. We calculated the annual mean occupancy rates (proportion of hexagons with species present) for each habitat associate under each strength of association, and compared these to the true occupancy in the simulation to assess if selective sampling biased occupancy estimates of these simplified species-habitat relationships.

3 | Results

OGSI 80 and 200 were strongly positively correlated in all data sets, and OGSI 80 was included in the best approximating model for all analyses except the Oregon station-level models (Tables S3–S6). Five- and 3-year lags in lumber prices were included in the best approximating models for Oregon and Washington, respectively (Tables S3 and S4). For the station-level models, station elevation and average elevation around stations were correlated, with the former included in best approximating models (Tables S5 and S6). In Washington, edge density and canopy cover were correlated, with edge density selected at the site level (Table S4) and canopy cover selected at the station level (Table S6).

3.1 | Fixed Effects and Random Effects

All random effects were included for site models (Tables S7 and S8), and thus all covariates were included as fixed effects. At Oregon stations, edge density was not included as a random or fixed effect (Tables S9 and S11), and at Washington stations distance to road was not included as a random or fixed effect (Tables S10 and S12). An interaction with time period was included for all covariates at Oregon sites (Table S13), while the selected model for Washington sites included only time period interactions with OGSI 80 and edge density (Table S14). Lumber price interactions were included for elevation standard deviation, canopy cover, and edge density for Oregon sites (Table S17), and elevation, OGSI 80, edge density, distance to road for Washington sites (Table S18).

The foremost factor determining survey site location was the distance to roads (Table S19) and the magnitude of the effect differed between time periods in Oregon and with lumber prices in Washington. We estimated that the odds a site was selected was, on average, eight and 33 times less likely for each 390 and 2973 m increase in distance from a road in Oregon and Washington, respectively. In Oregon, the strength of selection against distance to roads decreased over the time periods, and Washington sites were closer to roads as lumber prices increased. The odds a site was selected was 2.0 times more likely and 2.6 times less likely for every 29.4 and 32.8 increase in the percent of OGSI 80 cover in Oregon and Washington, respectively (Table S19). In Oregon, the strength of selection for sites with more OGSI 80 cover decreased over the time periods. Washington sites with more OGSI 80 cover were most likely to be selected prior to 1994 and least likely between 1994 and 2003 and sites were more likely to include higher percentages of OGSI 80 as lumber prices increased. The odds a site in Oregon was selected was 2.1 times less likely for every 248.9 increase in the number of conifers per hectare and were least likely to be selected before 1994, with weakest selection between 1994 and 2003. In Washington, sites were 3.7 times less likely to be selected for every 423 m increase in mean elevation, and sites at higher elevation were less likely to be selected when lumber prices increased.

While most covariates were included as fixed effects in the best approximating station models in Oregon and Washington, their influences were relatively weak (Table S20). The standardised odds ratios for most parameters were less than 2.0 with the single exception of distance to roads in Oregon prior to 1994.

The best approximating models for all site and station analyses included random effects for most covariates that varied yearly (Tables S19 and S20). Fixed effect interactions with time periods and lumber prices accounted for most of the annual trends in selection for or against habitat attributes, so that the random effects primarily accounted for inter-annual variation in selection around the fixed effect values (Figure 3).

3.2 | Simulating Effects of Sampling on Old Growth Forest Obligates

Randomly sampling hexagons yielded occupancy estimates that matched the true occupancy rates, while selective sampling biased occupancy estimates high for the old-growth associate and low for the canopy cover associate (Figure 2c–h). The magnitude of bias in occupancy estimates from selective sampling increased as the strength in habitat association increased. With a strong habitat association ($\beta_h = 1.0$), mean occupancy over the entire simulated time period was 0.16 lower and 0.11 higher than true occupancy for the canopy cover and old-growth associates, respectively. At a moderate habitat association ($\beta_h = 0.1$), mean occupancy of the old-growth associate was still 0.11 higher than true occupancy, while the canopy cover associate was 0.07 lower than true occupancy. Weak associations with habitat ($\beta_h = 0.1$) produced mean occupancy estimates that were only 0.001 lower and 0.031 higher than true occupancy for the canopy cover and old-growth associates, respectively.

Changes in true occupancy, from year to year and over the time period, of the simulated species were minimal with weak habitat associations, and showed relatively consistent trends over the modelling period for moderate and strong associations with little interannual variation, which was reflected in the estimates from random sampling (Figure 2c–h). However, selective sampling introduced a substantial amount of interannual variation, especially under the moderate and strong associations with old-growth (Figure 2g,h), and strong association with canopy cover (Figure 2e). This often appeared as fluctuations in occupancy between successive years, but sometimes as sustained trends over several years (e.g., old-growth associate between 1998 and 2001; Figure 2h) that were not present in the true population nor estimated with random sampling.

4 | Discussion

Systematic sampling of murrelet survey sites in conjunction with proposed timber management activities resulted in selecting sites based on combinations of habitat characteristics that varied through time. The same was seen with station placement within sites, but to a lesser degree. Strength of selection for these features varied over time with yearly random effects estimates, time-period interactions, and annual lumber prices preceding surveys. Selection of these habitat features suggested that the murrelet clearance monitoring data do not represent random sampling of the available habitat, which could affect conclusions of murrelet distribution and abundance if not properly accounted for in modelling.

Survey stations provided relatively good representation of the habitat available within sites, with weak effects for most covariates. However, almost all covariates were still included in

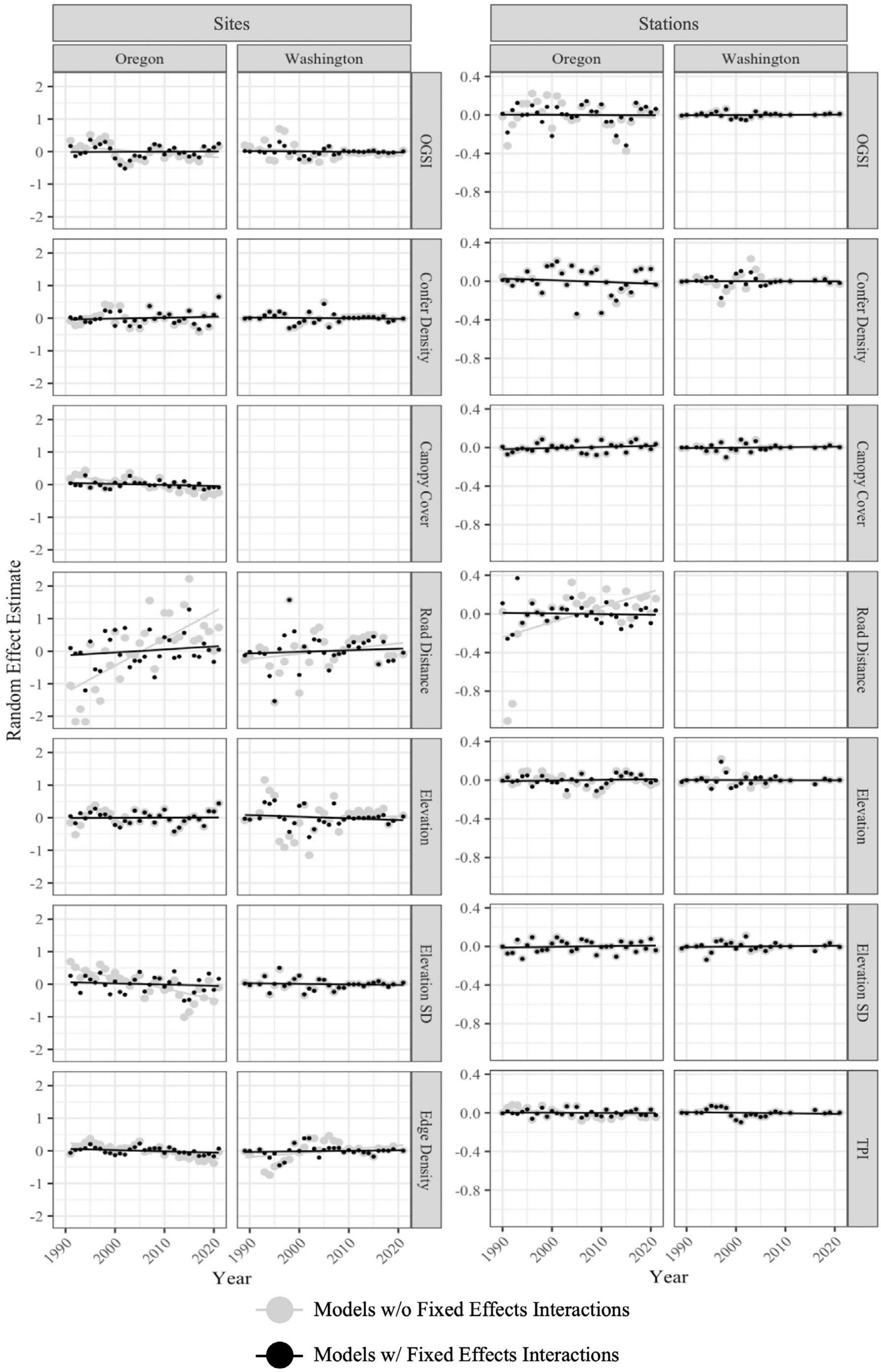

FIGURE 3 | Legend on next page.

FIGURE 3 | Estimates of yearly random effects of standardised covariates from the best approximating models (black) for marbled murrelet (*Brachyramphus marmoratus*) survey site and station selection models for surveys conducted in Oregon and Washington, USA from 1989 to 2021. Random effects estimates for the best approximating models prior to the inclusion of time period and lumber price interactions (grey) illustrate the general trends in selection for or against covariates over time. Lines represent linear models of the random effect estimates over time.

the best approximating station-level models, suggesting they hold some explanatory power in where stations were located within sites. Some selection of stations in sites based on land cover was expected given the survey protocols encouraged surveyors to place stations in areas that provided better visibility (i.e., opening in the canopy, edge of dense forest, etc.) to maximise the odds of visually detecting murrelets when present. However, selection may also have arisen due to the protocol stopping rule, whereby surveys could be suspended, leaving stations unsurveyed and out of the databases. Prioritisation of survey stations within a season or correlations between detection probabilities and forest structure (e.g., canopy cover or conifer density) could bias which habitat attributes were left unsampled in a site and those that are available for further analyses. Ultimately, station selection occurred within sites, which suggests these features should be considered when attempting to use stations as the replicate surveys required to fit occupancy models at the site level (i.e., space-for-time substitutions; Duarte and Peterson 2021; Kendall and White 2009) if these data continue to be used to understand the distributional dynamics of this species.

Proximity to roads was the strongest predictor of where survey sites were located, likely to facilitate transporting timber to mills. Roads can bias inferences from wildlife surveys (e.g., Betts et al. 2007; Keller and Scallan 1999) and directly affect wildlife (e.g., Taylor and Goldingay 2010; Shilling et al. 2021). Edge effects are a primary concern with roads (e.g., Khamcha et al. 2018; Poor et al. 2019), and edge habitat is thought to impact murrelets with increased nest predators (Raphael, Mack, Mazluff, et al. 2002; Malt and Lank 2007) and reduced suitable nesting substrates (Van Rooyen, Malt, and Lank 2011). While we measured and modelled the distance of a site to the nearest road, 69% of survey sites contained roads. Despite the prevalence of roads in sites, we saw relatively weak selection for edge habitat, but we did use a simple definition of edge compared to other more sophisticated measures (such as morphological spatial pattern analysis; Soille and Vogt 2009). OGSI 80 was another strong predictor of site location. The survey protocol required that old-growth forests be surveyed, but also that mature forests without an old-growth component or younger forests with platforms suitable for nesting should likewise be surveyed. Despite selection for most forest attributes appearing modest, this can still have profound impacts on the inferences. The cumulative effect of OGSI in Oregon (main effect with time period interaction and random effects) only resulted in a site being two to three times more likely to be selected for every 29.4% increase in OGSI 80 cover, yet we saw this resulted in the occupancy of the simulated moderate old-growth associate being substantially over-estimated. Given, the murrelet's strong association with large tracts of late-successional, old growth forests (Meyer, Miller, and Ralph 2002; Raphael, Mack, and Cooper 2002; Duarte et al. 2024), using these clearance monitoring data to examine trends in occupancy would yield results predicated upon trends in the quality of habitat surveyed.

Given that the murrelet is a threatened species and a focal species of the NWFP, trends in its distribution, nesting activity, and abundance are of considerable importance (U.S. Department of Agriculture Forest Service, and USDI Bureau of Land Management 1994; US Fish and Wildlife Service 1997; Lorenz et al. 2021; McIver et al. 2021). All covariates considered in the Oregon site model varied with the time-period interaction, and several varied with the price of lumber and the strength of the effects varied systematically across years (Table S19, Figure 3). The simulation analysis illustrated how this can bias assessments of occupancy trends through time. Simulated annual occupancy rates from selective sampling were always over- and under-estimated for the simulated old-growth and canopy cover associates, respectively, and the magnitude of this bias varied by year. For example, simulated occupancy of the moderate old growth associate ($\beta_h = 0.1$) dropped from 0.28 to 0.20 between 1999 and 2002 with selective sampling (Figure 2g), which would likely raise concern for a threatened species, yet in this instance the trend was an artefact of selectively sampling sites and true simulated occupancy actually increased slightly during that time period. This perceived decrease in estimated occupancy with selective sampling corresponded to a similar trend seen in the yearly random effects of OGSI 80 over the same period in Oregon site selection (Figure 3).

Purposeful selection of survey locations based on habitat and other features can impart a systematic pattern in the data, but another source of bias could be the systematic exclusion of potential survey locations. For example, following the Pechman exemptions to the Survey and Manage Program under the NWFP, many organisations shifted timber harvest towards thinning younger forests (< 80 years of age, reviewed in Johnson et al. 2023), which ultimately reduced the need for murrelet surveys. Additionally, national parks and wilderness areas likely offered high quality murrelet nesting habitat, but since their forests were not managed for timber production, they received relatively few protocol surveys. Furthermore, there were relatively few protocol surveys conducted after the mid-2000s in Washington, due in part to WDNR concluding inventory surveys, but also because proposed timber harvests in these later years did not include potential habitat to warrant surveys (pers. comm. D. Masters). Policies and practices such as these likely benefited murrelet conservation, yet they also resulted in vast tracts of the murrelet range—managed for their conservation and that of other late-successional, old-growth species—being poorly represented in the murrelet clearance monitoring data.

Data aggregated over project-specific clearance monitoring surveys are typically not sampled with the intention of answering broadscale questions of distribution and abundance, similar to occurrence records from museum collections or recreational bird watchers. Our results demonstrated that underlying decisions in where and how a species is surveyed for clearance monitoring imparted bias when using these data to estimate status and trends in occupancy. While we urge caution when analysing clearance monitoring data, for murrelets or otherwise, we do not discourage their use all together nor outrightly dismiss research that have used this type of data. Pre-existing, found data, albeit sometimes flawed, might still provide the best-available science to help inform conservation and management decisions until new, more targeted data are brought to bear. Consideration should, however, be given to how the data were collected and the impact that could have on inferences. At a minimum, the scope of inference provided by found data should be understood and explicitly stated, as with any study design.

The issues of systematic sampling that we presented herein might be overcome using a model-based approach. Although unstructured data are often analysed with a presence-only framework (e.g., MAXENT; Phillips, Anderson, and Schapire 2006), we stress that clearance monitoring data do in fact have a structure and that analysing these data within a presence only framework does not overcome these limitations (Yackulic et al. 2013). However, recent advancements in data integration methods (Pacifici et al. 2017; Miller et al. 2019) could provide avenues for utilising clearance monitoring surveys and other large sets of found data, such as point process models (Renner et al. 2015) or integrated population models (Zipkin and Saunders 2018). These provide flexible analytical frameworks for estimating parameters of interest across data sets of varying type, size, and quality. Although further evaluations are needed, more accurate assessments of murrelet distribution and abundance may be achieved by integrating clearance monitoring survey data with at-sea counts (McIver et al. 2024), known nest locations (Lorenz et al. 2021), radar detections (Burger 2001), or recently developed passive acoustic monitoring techniques (Borker et al. 2015; Cragg, Burger, and Piatt 2015; Duarte et al. 2024).

Big data continues to hold much promise in understanding broadscale biogeographical patterns and processes. Yet, big data can still very much be biased data with respect to the objectives of the analysis. The murrelet clearance monitoring surveys constituted one of the largest data sets for this species, and a big data set in their own right (130,360 surveys in the data sets analysed here, collected across heterogeneous sources), yet their use beyond assessing whether murrelets may be present in a proposed timber harvest needs to be carefully considered. Approaches to account for these issues are context dependent, with the only blanket prescription for using found data being "eternal vigilance" (Hurlbert 1984). One may think that if they have a lot of data, that the data may be spatially and temporally exhaustive enough to provide all of the data needed to overcome its sampling design shortfalls. However, big data do not inherently overcome systematic biases in how they were collected, they only improve precision in estimates around these biases, potentially producing "highly precise wrong answers" (Anderson, Burnham, and White 1985). Our study serves as a useful reminder that in the era of big data, sampling design still matters if we are to truly understand fundamental ecological patterns and processes to address critical conservation and management needs.

Author Contributions

All authors conceived the research idea, edited and substantially contributed to the final manuscript. R.B. analyzed the data and wrote the first draft. A.D. and J.T.P. acquired funding for the research.

Acknowledgements

We appreciate the efforts of the field biologists who spent countless early morning hours surveying for the iconic marbled murrelet that made this research possible. We also thank C. Aron, G. Bell, T. Christopher, S. Desimone, M. Gostin, B. Hollen, J. Jenkerson, S. Krock, D. Masters, V. Petro, and J. Stocking for support in compiling and verifying the databases and feedback on preliminary results. Funding for this project was provided by the USDA Forest Service, Pacific Northwest Research Station. The Oregon Cooperative Fish and Wildlife Research Unit is jointly sponsored by the US Geological Survey, the US Fish and Wildlife Service, the Oregon Department of Fish and Wildlife, Oregon State University, and the Wildlife Management Institute. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. No permits were required for this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The survey site and station location data that support the findings of this study were provided by the respective agencies under data sharing agreements, and may be requested from these agencies. All other data used in this research are publicly available at the locations cited.

References

Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Proceedings of the Second International Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki, 267–281. Budapest, Hungary: Akademia Kaido.

Allen, K. E., W. P. Tapondjou, B. Freeman, J. C. Cooper, R. M. Brown, and A. T. Peterson. 2021. "Modelling Potential Pleistocene Habitat Corridors Between Afromontane Forest Regions." *Biodiversity and Conservation* 30, no. 8–9: 2361–2375. https://doi.org/10.1007/s10531-021-02198-4.

Anderson, D. R. 2008. *Model Based Inference in the Life Sciences: A Primer on Evidence*. New York, NY: Springer.

Anderson, D. R., K. P. Burnham, and G. C. White. 1985. "Problems in Estimating Age-Specific Survival Rates From Recovery Data of Birds Ringed as Young." *Journal of Animal Ecology* 54, no. 1: 89–98. https://doi.org/10.2307/4622.

Bell, D. M., M. J. Gregory, M. Palmer, and R. Davis. 2023. "Guidance for Forest Management and Landscape Ecology Applications of Recent Gradient Nearest Neighbor Imputation Maps in California, Oregon, and Washington." (PNW-GTR-1018; p. PNW-GTR-1018). U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. https://doi.org/10.2737/PNW-GTR-1018.

Betts, M. G., D. Mitchell, A. W. Diamond, and J. Béty. 2007. "Uneven Rates of Landscape Change as a Source of Sias in Roadside Wildlife Surveys." *Journal of Wildlife Management* 71, no. 7: 2266–2273. https://doi.org/10.2193/2006-004.

Betts, M. G., J. M. Northrup, J. A. B. Guerrero, et al. 2020. "Squeezed by a Habitat Split: Warm Ocean Conditions and Old-Forest Loss Interact to Reduce Long-Term Occupancy of a Threatened Seabird." *Conservation Letters* 13, no. 5: e12745. https://doi.org/10.1111/conl.12745.

Bizama, G., A. Jan, J. A. Olivos, et al. 2024. "Climate Change Can Disproportionately Reduce Habitats of Stream Fishes With Restricted Ranges in Southern South America." *Scientific Reports* 14, no. 1: 15780. https://doi.org/10.1038/s41598-024-66374-6.

Bolker, B. M. 2015. "Linear and Generalized Linear Mixed Models." In *Ecological Statistics: Contemporary Theory and Application*, edited by G. A. Fox, S. Negrete-Yankelevich, and V. J. Sosa. Oxford, UK: Oxford University Press.

Borker, A. L., P. Halbert, M. W. Mckown, B. R. Tershy, and D. A. Croll. 2015. "A Comparison of Automated and Traditional Monitoring Techniques for Marbled Murrelets Using Passive Acoustic Sensors." *Wildlife Society Bulletin* 39, no. 4: 813–818. https://doi.org/10.1002/wsb.608.

Brooks, M. E., K. Kristensen, K. J. van Benthem, et al. 2017. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *R Journal* 9, no. 2: 378–400. https://doi.org/10.32614/RJ-2017-066.

Burger, A. E. 2001. "Using Radar to Estimate Populations and Assess Habitat Associations of Marbled Murrelets." *Journal of Wildlife Management* 65, no. 4: 696–715. https://doi.org/10.2307/3803021.

Cragg, J. L., A. E. Burger, and J. F. Piatt. 2015. "Testing the Effectiveness of Automated Acoustic Sensors for Monitoring Vocal Activity of Marbled Murrelets *Brachyramphus marmoratus*." *Marine Ornithology* 43: 151–160.

Cosgrove, C., N. Coops, F. Waterhouse, and T. Goodbody. 2024. "Modeling Marbled Murrelet Nesting Habitat: A Quantitative Approach Using Airborne Laser Scanning Data in British Columbia, Canada." *Avian Conservation and Ecology* 19, no. 1: 5. https://doi.org/10.5751/ACE-02585-190105.

de Lucas, M., M. Ferrer, and G. F. E. Janss. 2012. "Using Wind Tunnels to Predict Bird Mortality in Wind Farms: The Case of Griffon Vultures." *PLoS One* 7, no. 11: e48092. https://doi.org/10.1371/journal.pone.0048092.

Duarte, A., and J. T. Peterson. 2021. "Space-For-Time Is Not Necessarily a Substitution When Monitoring the Distribution of Pelagic Fishes in the San Francisco Bay-Delta." *Ecology and Evolution* 11, no. 23: 16727–16744. https://doi.org/10.1002/ece3.8292.

Duarte, A., M. J. Weldy, D. B. Lesmeister, et al. 2024. "Passive Acoustic Monitoring and Convolutional Neural Networks Facilitate High-Resolution and Broadscale Monitoring of a Threatened Species." *Ecological Indicators* 162: 112016. https://doi.org/10.1016/j.ecolind.2024.112016.

Evans Mack, D., W. P. Ritchie, S. K. Nelson, E. Kuo-Harrison, P. Harrison, and T. E. Hamer. 2003. *Methods for Surveying Marbled Murrelets in Morests: A Revised Protocol for Land Management and Research*. Pacific Seabird Group.

Ferrer, M., M. de Lucas, G. F. E. Janss, et al. 2012. "Weak Relationship Between Risk Assessment Studies and Recorded Mortality in Wind Farms." *Journal of Applied Ecology* 49, no. 1: 38–46. https://doi.org/10.1111/j.1365-2664.2011.02054.x.

Franklin, J., J. M. Serra-Diaz, A. D. Syphard, and H. M. Regan. 2017. "Big Data for Forecasting the Impacts of Global Change on Plant Communities." *Global Ecology and Biogeography* 26, no. 1: 6–17. https://doi.org/10.1111/geb.12501.

Gelman, A., and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Griffin, G. P., M. Mulhall, C. Simek, and W. W. Riggs. 2020. "Mitigating Bias in Big Data for Transportation." *Journal of Big Data Analytics in Transportation* 2, no. 1: 49–59. https://doi.org/10.1007/s42421-020-00013-0.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, et al. 2013. "Big Data and the Future of Ecology." *Frontiers in Ecology and the Environment* 11, no. 3: 156–162. https://doi.org/10.1890/120103.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

Hull, C. L., G. W. Kaiser, C. Lougheed, L. Lougheed, S. Boyd, and F. Cooke. 2001. "Intraspecific Variation in Commuting Distance of Marbled Murrelets (*Brachyramphus marmoratus*): Ecological and Energetic Consequences of Nesting Further Inland." *Auk* 118, no. 4: 1036–1046.

Hurlbert, S. H. 1984. "Pseudoreplication and the Design of Ecological Field Experiments." *Ecological Monographs* 54, no. 2: 187–211. https://doi.org/10.2307/1942661.

Hurvich, C. M., and C.-L. Tsai. 1989. "Regression and Time Series Model Selection in Small Samples." *Biometrika* 76, no. 2: 297–307. https://doi.org/10.2307/2336663.

Johnson, K. N., J. F. Franklin, G. H. Reeves, D. L. Johnson, and S. J. M. Brown. 2023. *The Making of the Northwest Forest Plan: The Wild Science of Saving Old Growth Ecosystems*. Corvallis, OR: Oregon State University Press.

Kaplan, R. M., D. A. Chambers, and R. E. Glasgow. 2014. "Big Data and Large Sample Size: A Cautionary Note on the Potential for Bias." *Clinical and Translational Science* 7, no. 4: 342–346. https://doi.org/10.1111/cts.12178.

Keller, C. M. E., and J. T. Scallan. 1999. "Potential Roadside Biases due to Habitat Changes Along Breeding Bird Survey Routes." *Condor* 101, no. 1: 50–57. https://doi.org/10.2307/1370445.

Kendall, W. L., and G. C. White. 2009. "A Cautionary Note on Substituting Spatial Subunits for Repeated Temporal Sampling in Studies of Site Occupancy." *Journal of Applied Ecology* 46, no. 6: 1182–1188. https://doi.org/10.1111/j.1365-2664.2009.01732.x.

Khamcha, D., R. T. Corlett, L. A. Powell, T. Savini, A. J. Lynam, and G. A. Gale. 2018. "Road Induced Edge Effects on a Forest Bird Community in Tropical Asia." *Avian Research* 9, no. 1: 20. https://doi.org/10.1186/s40657-018-0112-y.

Lorenz, T. J., M. G. Raphael, T. D. Bloxton, and P. G. Cunningham. 2017. "Low Breeding Propensity and Wide-Ranging Movements by Marbled Murrelets in Washington." *Journal of Wildlife Management* 81, no. 2: 306–321. https://doi.org/10.1002/jwmg.21192.

Lorenz, T. J., M. G. Raphael, R. D. Young, D. Lynch, S. K. Nelson, and W. R. McIver. 2021. *Status and Trend of Nesting Habitat for the Marbled Murrelet Under the Northwest Forest Plan, 1993 to 2017*, 65. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Malt, J., and D. Lank. 2007. "Temporal Dynamics of Edge Effects on Nest Predation Risk for the Marbled Murrelet." *Biological Conservation* 140, no. 1–2: 160–173. https://doi.org/10.1016/j.biocon.2007.08.011.

McIver, W., J. Baldwin, M. M. Lance, et al. 2024. "Marbled Murrelet Effectiveness Monitoring, Northwest Forest Plan—2023 Summary Report" (p. 25) [Final Report]. Northwest Forest Plan Interagency Regional Monitoring Program.

McIver, W. R., S. F. Pearson, C. Strong, et al. 2021. *Status and Trend of Marbled Murrelet Populations in the Northwest Forest Plan Area, 2000 to 2018*, 37 [Gen. Tech. Rep. PNW-GTR-996]. Portland, OR: Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Meyer, C. B., S. L. Miller, and C. J. Ralph. 2002. "Multi-Scale Landscape and Seascape Patterns Associated With Marbled Murrelet Nesting

Areas on the U.S. West Coast." *Landscape Ecology* 17: 95–115. https://doi.org/10.1023/A:1016574928706.

Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. "The Recent Past and Promising Future for Data Integration Methods to Estimate species' Distributions." *Methods in Ecology and Evolution* 10, no. 1: 22–37. https://doi.org/10.1111/2041-210X.13110.

Moore, D., and G. McCabe. 1993. *Introduction to the Practice of Statistics*. New York, NJ: W. H. Freeman and Company.

Morin, D. J., C. B. Yackulic, J. E. Diffendorfer, et al. 2020. "Is Your Ad Hoc Model Selection Strategy Affecting Your Multimodel Inference?" *Ecosphere* 11, no. 1: e02997. https://doi.org/10.1002/ecs2.2997.

Moritz, C., J. L. Patton, C. J. Conroy, J. L. Parra, G. C. White, and S. R. Beissinger. 2008. "Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA." *Science* 322, no. 5899: 261–264. https://doi.org/10.1126/science.1163428.

Nathan, R., C. T. Monk, R. Arlinghaus, et al. 2022. "Big-Data Approaches Lead to an Increased Understanding of the Ecology of Animal Movement." *Science* 375, no. 6582: eabg1780. https://doi.org/10.1126/science.abg1780.

Overton, J. M., T. C. Young, and W. S. Overton. 1993. "Using 'Found' Data to Augment a Probability Sample: Procedure and Case Study." *Environmental Monitoring and Assessment* 26, no. 1: 65–83. https://doi.org/10.1007/BF00555062.

Pacific Seabird Group. 2024. *A Revised Protocol for Surveying Marbled Murrelets in Forests*. Pacific Seabird Group Technical Publication. Number 6. https://pacificseabirdgroup.org/psg-publications/technical-publications/.

Pacifici, K., B. J. Reich, D. A. W. Miller, et al. 2017. "Integrating Multiple Data Sources in Species Distribution Modeling: A Framework for Data Fusion." *Ecology* 98, no. 3: 840–850. https://doi.org/10.1002/ecy.1710.

Paton, P. W. C., C. J. Ralph, H. R. Carter, and S. K. Nelson. 1990. "Surveying Marbled Murrelets at Inland Forested Sites: A Guide." (PSW-GTR-120; p. PSW-GTR-120). U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. https://doi.org/10.2737/PSW-GTR-120.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190, no. 3–4: 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Poor, E. E., V. I. M. Jati, M. A. Imron, and M. J. Kelly. 2019. "The Road to Deforestation: Edge Effects in an Endemic Ecosystem in Sumatra, Indonesia." *PLoS ONE* 14, no. 7: e0217540. https://doi.org/10.1371/journal.pone.0217540.

QGIS Association. 2023. QGIS Geographic Information System. http://www.qgis.org.

Ralph, C. J., and K. Nelson. 1992. *Methods of Surveying Marbled Murrelets at Inland Forested Sites*, 21. Pacific Seabird Group, Marbled Murrelet Technical Committee.

Ralph, C. J., S. K. Nelson, M. M. Shaughnessy, S. L. Miller, and T. E. Hamer. 1994. *Methods for Surveying Marbled Murrelets in Forests: A Protocol for Land Management and Research*, 51. Pacific Seabird Group, Marbled Murrelet Technical Committee.

Raphael, M. G., D. E. Mack, J. M. Mazluff, and J. M. Luginbuhl. 2002. "Effects of Forest Fragmentation on Populations of the Marbled Murrelet." *Studies in Avian Biology* 25: 221–235.

Raphael, M. G., D. E. Mack, and B. A. Cooper. 2002. "Landscape-Scale Relationships Between Abundance of Marbled Murrelets and Distribution of Nesting Habitat." *Condor* 104, no. 2: 331–342. https://doi.org/10.1093/condor/104.2.331.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Renner, I. W., J. Elith, A. Baddeley, et al. 2015. "Point Process Models for Presence-Only Analysis." *Methods in Ecology and Evolution* 6, no. 4: 366–379. https://doi.org/10.1111/2041-210X.12352.

Robinson, O. J., J. B. Socolar, E. F. Stuber, et al. 2022. "Extreme Uncertainty and Unquantifiable Bias Do Not Inform Population Sizes." *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 10: e2113862119. https://doi.org/10.1073/pnas.2113862119.

Rosenberg, K. V., A. M. Dokter, P. J. Blancher, et al. 2019. "Decline of the North American Avifauna." *Science* 366, no. 6461: 120–124. https://doi.org/10.1126/science.aaw1313.

Sauer, J. R., K. L. Pardieck, D. J. Ziolkowski, et al. 2017. "The First 50 Years of the North American Breeding Bird Survey." *Condor* 119, no. 3: 576–593. https://doi.org/10.1650/CONDOR-17-83.1.

Shilling, F., T. Nguyen, M. Saleh, et al. 2021. "A Reprieve From U.S. Wildlife Mortality on Roads During the COVID-19 Pandemic." *Biological Conservation* 256: 109013. https://doi.org/10.1016/j.biocon.2021.109013.

Soille, P., and P. Vogt. 2009. "Morphological Segmentation of Binary Patterns." *Pattern Recognition Letters* 30, no. 4: 456–459. https://doi.org/10.1016/j.patrec.2008.10.015.

Soltis, D. E., and P. S. Soltis. 2016. "Mobilizing and Integrating Big Data in Studies of Spatial and Phylogenetic Patterns of Biodiversity." *Plant Diversity* 38, no. 6: 264–270. https://doi.org/10.1016/j.pld.2016.12.001.

Sullivan, B. L., J. L. Aycrigg, J. H. Barry, et al. 2014. "The eBird Enterprise: An Integrated Approach to Development and Application of Citizen Science." *Biological Conservation* 169: 31–40. https://doi.org/10.1016/j.biocon.2013.11.003.

Tang, B., J. S. Clark, and A. E. Gelfand. 2021. "Modeling Spatially Biased Citizen Science Effort Through the eBird Database." *Environmental and Ecological Statistics* 28, no. 3: 609–630. https://doi.org/10.1007/s10651-021-00508-1.

Taylor, B. D., and R. L. Goldingay. 2010. "Roads and Wildlife: Impacts, Mitigation and Implications for Wildlife Management in Australia." *Wildlife Research* 37, no. 4: 320. https://doi.org/10.1071/WR09171.

Thomas, J. W., and M. Dombeck. 1996. "Ecosystem Management in the Interior Columbia River Basin." *Wildlife Society Bulletin (1973–2006)* 24, no. 2: 180–186.

U.S. Department of Agriculture Forest Service, & USDI Bureau of Land Management. 1994. *Final Supplemental Environmental Impact Statement on Managing Habitat for Late Successional and Old-Growth Species Within the Range of the Northern Spotted Owl*. Portland, OR: U.S. Department of Agriculture.

US Fish and Wildlife Service. 1992. "Endangered and Threatened Wildlife and Plants; Determination of Threatened Status for the Washington, Oregon, and California Population of the Marbled Murrelet." *Federal Register* 57, no. 191: 45328–45337.

US Fish and Wildlife Service. 1997. *Recovery Plan for the Threatened Marbled (Brachyramphus marmoratus) in Washington, Oregon, and California Populations*, 203 Portland, OR.

Valente, J. J., J. W. Rivers, Z. Yang, et al. 2023. "Fragmentation Effects on an Endangered Species Across a Gradient From the Interior to Edge of Its Range." *Conservation Biology* 37, no. 5: e14091. https://doi.org/10.1111/cobi.14091.

Van Doren, B. M., and K. G. Horton. 2018. "A Continental System for Forecasting Bird Migration." *Science* 361, no. 6407: 1115–1118. https://doi.org/10.1126/science.aat7526.

Van Rooyen, J. C., J. M. Malt, and D. B. Lank. 2011. "Relating Microclimate to Epiphyte Availability: Edge Effects on Nesting Habitat Availability for the Marbled Murrelet." *Northwest Science* 85, no. 4: 549–561. https://doi.org/10.3955/046.085.0405.

Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2023. "dplyr: A Grammar of Data Manipulation [Computer software]." https://dplyr.tidyverse.org.

Wilk, R. J., M. G. Raphael, and T. D. Bloxton Jr. 2016. "Nesting Habitat Characteristics of Marbled Murrelets Occurring in Near-Shore Waters of the Olympic Peninsula, Washington." *Journal of Field Ornithology* 87, no. 2: 162–175. https://doi.org/10.1111/jofo.12150.

Wüest, R. O., N. E. Zimmermann, D. Zurell, et al. 2019. "Macroecology in the Age of Big Data—Where to Go From Here?" *Journal of Biogeography* 47, no. 1: 1–12. https://doi.org/10.1111/jbi.13633.

Yackulic, C. B., R. Chandler, E. F. Zipkin, et al. 2013. "Presence-Only Modelling Using MAXENT: When Can We Trust the Inferences?" *Methods in Ecology and Evolution* 4, no. 3: 236–243. https://doi.org/10.1111/2041-210x.12004.

Zipkin, E. F., and S. P. Saunders. 2018. "Synthesizing Multiple Data Types for Biological Conservation Using Integrated Population Models." *Biological Conservation* 217: 240–250. https://doi.org/10.1016/j.biocon.2017.10.017.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.