Bayesian Models for Analysis of Inventory and Monitoring Data with Non-ignorable Missingness

Luke J. ZACHMANN, Erin M. BORGMAN, Dana L. WITWICKI,
Megan C. SWAN, Cheryl MCINTYRE, and N. Thompson HOBBS

We describe the application of Bayesian hierarchical models to the analysis of data from long-term, environmental monitoring programs. The goal of these ongoing programs is to understand status and trend in natural resources. Data are usually collected using complex sampling designs including stratification, revisit schedules, finite populations, unequal probabilities of inclusion of sample units, and censored observations. Complex designs intentionally create data that are missing from the complete data that could theoretically be obtained. This “missingness” cannot be ignored in analysis. Data collected by monitoring programs have traditionally been analyzed using the design-based Horvitz–Thompson estimator to obtain point estimates of means and variances over time. However, Horvitz–Thompson point estimates are not capable of supporting inference on temporal trend or the predictor variables that might explain trend, which instead requires model-based inference. The key to applying model-based inference to data arising from complex designs is to include information about the sampling design in the analysis. The statistical concept of ignorability provides a theoretical foundation for meeting this requirement. We show how Bayesian hierarchical models provide a general framework supporting inference on status and trend using data from the National Park Service Inventory and Monitoring Program as examples.

Supplemental Materials Code and data for implementing the analyses described here can be accessed here: <https://doi.org/10.36967/code-2287025>.

Key Words: Ignorability; Long-term data; Missing data; Model-based inference; Status; Trend.

L. J. Zachmann (✉), Conservation Science Partners Inc., Truckee, CA, USA (E-mail: luke@csp-inc.org).
E. M. Borgman, Rocky Mountain Network, National Park Service, Fort Collins, CO, USA
(E-mail: erin_borgman@nps.gov). D. L. Witwicki, Northern Colorado Plateau Network, National Park Service,
Moab, UT, USA (E-mail: dana_witwicki@nps.gov). M. C. Swan, Southern Colorado Plateau Network, National
Park Service, Flagstaff, AZ, USA (E-mail: megan_swan@nps.gov). C. McIntyre, Chihuahuan Desert Network,
National Park Service, Tucson, AZ, USA (E-mail: cheryl_mcintyre@nps.gov). N. T. Hobbs, Conservation Science
Partners and Colorado State University, Fort Collins, CO, USA (E-mail: tom.hobbs@colostate.edu).

© 2021 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 1, Pages 125–148
<https://doi.org/10.1007/s13253-021-00473-z>

1. INTRODUCTION

The central challenge of statistics is to assess the value of observations as evidence (Woodworth 2004). Inevitably, there are observations that are missing from the set of data we obtain, observations that could potentially inform the questions we seek to answer. Missing observations arise in two ways, by accident and by design. Observations are accidentally missing as a result of logistical problems, non-responses, instrument failures, or other unintended causes. This is what most practitioners think of when we speak of missing data. But data are also *intentionally* missing because realized experimental and sampling designs include a subset of all of the potentially informative observations. Reliable inference requires dealing with both types of missingness.

An important illustration of data that are missing by design arises from long-term monitoring programs that collect data to inform management of natural resources at local, regional, and national scales. Notable examples in the USA include the Natural Resource Conservation Service National Resources Inventory, the Forest Inventory and Analysis and Forest Health Monitoring programs of the US Forest Service, the Bureau of Land Management Assessment, Inventory, and Monitoring Strategy, and the Inventory and Monitoring programs of the National Park Service and US Fish & Wildlife Service. These programs collect large, diverse sets of data using complex sampling designs that include stratification, revisit schedules, finite populations, censored or truncated observations, and unequal probabilities of inclusion of sample units. The missingness created by these designs cannot be ignored in analysis.

Analysis of data from large monitoring programs has traditionally emphasized design-based inference implemented using variations of the Horvitz–Thompson estimator (Overton and Stehman 1994; Stehman and Overton 1994; Messer et al. 1991; Courbois and Urquhart 2004; Kincaid et al. 2004). Design-based inference requires probabilities of inclusion of a sample unit in the analysis to give proper weight to samples drawn from a finite population (Irvine et al. 2018; Williams and Brown 2019). However, this approach imposes substantial limitations on inference beyond point estimates of means and variances (Gelman 2007; Williams and Brown 2019). Point estimates are not capable of supporting statistical inference on temporal trend or the predictor variables that might explain causes of trend, which instead requires model-based inference (Shanahan et al. 2021). Moreover, existing statistical packages using the Horvitz–Thompson estimator developed for analysis of data from complex designs, for example, the *survey* (Lumley 2015), *spsurvey* (Kincaid and Olsen 2019), and *trendNPS* (Starcevich and Mitchell 2017) packages in R (R Core Team 2020), lack the flexibility to deal with the many idiosyncrasies that inevitably arise in monitoring data.

Here, we apply Bayesian hierarchical models to the analysis of data from complex designs with non-ignorable missingness, using data from networks in the National Park Service Inventory and Monitoring Program as an example. Our goal is to illustrate a flexible approach for dealing with non-ignorable missingness that is firmly grounded in statistical theory and that expands the inferences that are feasible relative to those based on the Horvitz–Thompson estimator. Our intended audience includes statisticians interested in environmental monitoring as well as applied researchers and decision makers who use monitoring data.

This paper is organized as follows. We begin by describing data sets obtained by the National Park Service Inventory and Monitoring Program (reviewed by [Olsen et al. 1999](#); [Fancy et al. 2009](#)), highlighting the problem of missingness created by design. We then provide an overview of the concept of ignorability (sensu [Gelman et al. 2013](#), Chapter 8) in the Bayesian framework, ideas that are fundamental to understanding the modeling approach we advocate. Next, we show a highly general model that can be modified to accommodate the different types of observations and different sources of missingness that arise in the Inventory and Monitoring data. Finally, we illustrate the use of this model with three specific applications involving different types of non-ignorable missing data.

2. DATA

2.1. MOTIVATION FOR MONITORING

Concern about threats to natural resources in national parks motivated the United States Congress to require the National Parks Service to implement a nationwide Inventory and Monitoring Program in 1998 ([Fancy et al. 2009](#)). The purpose of the Inventory and Monitoring Program is to provide scientifically reliable information about the status and trend of the health of park ecosystems over time, enabling park staff to make better decisions about managing natural and human resources in the face of rapid environmental change ([Fancy et al. 2009](#)). In the lexicon of the program, an “indicator” is an observable quantity describing an environmental attribute of interest (sensu [Larsen et al. 2001](#)). Examples of indicators include counts of individuals or species, visual classification of vegetative cover, and measures of soil stability and water chemistry. “Status” is an estimate of the mean or median value of the indicator within a defined temporal window. “Trend” is the long-term, directional change in an indicator over time apart from year-to-year variation that is caused by changes in weather, observers, instrumentation, disturbances, or other short-term influences on status.

2.2. SAMPLING DESIGN

There are 32 Inventory and Monitoring networks nationwide, roughly organized by ecoregion. Each network collaborated with park managers, academics, and agency scientists to identify indicators of the condition of resources in the parks, which, taken collectively, could be used to gauge the health of park ecosystems ([Fancy et al. 2009](#)). We built models to analyze data collected by six Inventory and Monitoring networks: the Southern Plains, Southern Colorado Plateau, Northern Colorado Plateau, Rocky Mountain, Sonoran Desert, and Chihuahuan Desert networks (Fig. 1). Although these networks include a variety of administrative units including national parks, national historic sites, and national monuments, we will refer to all units as “parks.”

Sampling was predominantly stratified random (Fig. 2a). Strata were chosen to represent major vegetation communities and soil types. Sites, typically 0.1-1 ha in size, were most often chosen within strata using a spatially balanced design ([Stevens and Olsen 2004](#); [Theobald et al. 2007](#)). Inclusion probabilities π_{jkt} for site $j = 1, 2, \dots, J_{kt}$ within stratum k at time

Figure 1. A map of the Inventory and Monitoring networks involved in this study. Points within each network represent individual park units. Labeled points (triangles) show the location of parks for which we present individual analyses .

t were equal to $\pi_{jkt} = J_{kt}/N_k$ in some parks, but in others, the inclusion probability of a site was based on the cost of travel, c_{jk} , to the site. Replicated observation units (plots and transects) were systematically located within sites.

Allocation of observations to sites over time followed a revisit design (Fig. 2b). Sites were deliberately omitted from sampling during some years, thereby allowing greater spatial coverage of the stratum over time while accommodating limited financial resources for collecting observations.

Figure 2. **a** A schematic of a “generic” sampling design. Although each network often has a unique layout and vocabulary for sampling units, they all share the same basic components, including replicated observation units (plots, quadrats, and transects) within sites, with sites nested in strata, and strata within parks. **b** An example of a revisit design where a subset of sites are visited each year.

2.3. NON-IGNORABLE MISSINGNESS

The sampling design created data that were intentionally missing. There were five sources of non-ignorable missingness that must be accommodated in the analysis:

1. A revisit schedule created missing data at each site during some years. Proper inference on status across sites within a year and the trend across years must account for the missingness created by the revisit schedule, including missing response data and covariates.
2. Sites were selected using a stratified design. Inferences at park scales (across strata) must properly account for the stratification.
3. Sites sampled within a stratum were a set drawn from a finite number of potential sites.
4. Probabilities of inclusion of sites within strata in some parks were related to the cost of travel to the sites.
5. Some data were truncated and / or censored, creating observations that were partially missing.

We show how a general, Bayesian framework can properly account for all of these non-ignorable features of the sampling designs. In the next section, we outline the general

theoretical principles supporting inference from designs with non-ignorable missing data. We then describe our general approach to modeling before turning to specific applications.

3. IGNORABILITY

Financial constraints and the need to represent environmental variation over space and time motivate environmental monitoring programs to implement sampling designs that depart from simple random sampling. Reliable inference from observations collected in the complex designs typical of these programs depends on including information in the analysis about how the data were collected. The concept of *ignorability* is fundamental to deciding what information must be included. Here, we provide a general, theoretical underpinning needed to understand the specific examples of analysis that follow.

3.1. THE COMPLETE DATA LIKELIHOOD

All sampling designs result in a subset of observations from the population we seek to understand. The complete data include both the observed and the missing data (Gelman et al. 2013; Link and Barker 2010). Properly modeling this missingness is required for inference in all but the simplest designs. We explain how this requirement is met in the Bayesian framework following the reasoning of Gelman et al. (2013, Chapter 8).

We define the matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ to include all of the *potential* observations that could be obtained from the population of interest. Each of the $\mathbf{y}_i$ is a vector with elements y_{ij} , $j = 1, 2, \dots, J$ representing, for example, a single point in a time series of J years at $i = 1, 2, \dots, n$ different sites. The same logic we develop here applies if y_i is a scalar and $\mathbf{y}$ is a vector of all of the potential observations. We define the inclusion matrix $\mathbf{Q}$ as a matrix of the same dimension as $\mathbf{Y}$ indicating whether y_{ij} is observed, $q_{ij} = 1$, or missing, $q_{ij} = 0$. The observed values of $\mathbf{Y}$ are $\mathbf{Y}_{\text{obs}}$, and the unobserved values are $\mathbf{Y}_{\text{miss}}$. We emphasize for the purposes here that *missing* implies observations that are omitted by design, but it could also indicate observations that are unintentionally missing, as might occur when an observation was planned but subsequently omitted due to unforeseen logistical impediments. We notate fully observed covariates as $\mathbf{X}$, including predictor variables (for example, weather or geophysical variables, and indicators for observers) as well as covariates that specify elements of design, for example, membership in strata or timing of repeated measures in revisit schedules. The quantities $\boldsymbol{\theta}$ are parameters of the distribution of the complete data and $\boldsymbol{\phi}$ are parameters of the inclusion matrix. The complete data likelihood given parameters $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, and covariates $\mathbf{X}$ is

$$[\mathbf{Y}, \mathbf{Q} \mid \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\phi}] = [\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\theta}][\mathbf{Q} \mid \mathbf{X}, \mathbf{Y}, \boldsymbol{\phi}]. \quad (1)$$

The bracket notation $[a \mid b, c]$ reads the probability or probability density of a conditional on b and c (Gelfand and Smith 1990). Model parameters $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ are distributed jointly conditional on the observed quantities $\mathbf{X}$, $\mathbf{Y}_{\text{obs}}$, and $\mathbf{Q}$

$$\begin{aligned}
[\theta, \phi \mid \mathbf{X}, \mathbf{Y}_{\text{obs}}, \mathbf{Q}] &\propto [\theta, \phi \mid \mathbf{X}][\mathbf{Y}_{\text{obs}}, \mathbf{Q} \mid \mathbf{X}, \theta, \phi] \\
&= [\theta, \phi \mid \mathbf{X}] \int [\mathbf{Y}, \mathbf{Q} \mid \mathbf{X}, \theta, \phi] d\mathbf{Y}_{\text{miss}} \\
&= [\theta, \phi \mid \mathbf{X}] \int [\mathbf{Y} \mid \mathbf{X}, \theta][\mathbf{Q} \mid \mathbf{X}, \mathbf{Y}, \phi] d\mathbf{Y}_{\text{miss}}. \tag{2}
\end{aligned}$$

We obtain marginal posterior distributions of θ by integrating out ϕ and the missing data in Eq. 2

$$[\theta \mid \mathbf{X}, \mathbf{Y}_{\text{obs}}, \mathbf{Q}] = [\theta \mid \mathbf{X}] \int \int [\phi \mid \mathbf{X}, \theta][\mathbf{Y} \mid \mathbf{X}, \theta][\mathbf{Q} \mid \mathbf{X}, \mathbf{Y}, \phi] d\mathbf{Y}_{\text{miss}} d\phi, \tag{3}$$

which gives the posterior distribution of θ averaged over ϕ for non-ignorable designs. This integral can be approximated by posterior simulations of the joint vector of unobserved quantities $\mathbf{Y}_{\text{miss}}$, ϕ , and θ , which are then used to compute quantities of interest (means, credible intervals). For example, draws from the joint distribution of ϕ and θ could be used to impute the missing observations $\mathbf{Y}_{\text{miss}}$. Statistics of interest could then be computed using the observed and the imputed missing data (e.g., Gelman et al. 2013, Eq. 8.5). It is important to note that Eq. 3 is used for inference when designs are non-ignorable. Inference can be made more simply when designs are ignorable, as we describe next.

3.2. WHEN ARE DESIGNS IGNORABLE?

Sampling designs are ignorable when information about how the data were collected is properly included in the model used for analysis. Bayesian inference for ignorable designs can be made on unobserved quantities of interest θ by conditioning solely on the observed data and covariates omitting $\mathbf{Q}$ using

$$\begin{aligned}
[\theta \mid \mathbf{X}, \mathbf{Y}_{\text{obs}}] &\propto [\theta \mid \mathbf{X}][\mathbf{Y}_{\text{obs}} \mid \mathbf{X}, \theta] \\
&= [\theta \mid \mathbf{X}] \int [\mathbf{Y} \mid \mathbf{X}, \theta] d\mathbf{Y}_{\text{miss}}. \tag{4}
\end{aligned}$$

In this case, $[\theta \mid \mathbf{X}, \mathbf{Y}_{\text{obs}}]$ (from Eq. 4) equals $[\theta \mid \mathbf{X}, \mathbf{Y}_{\text{obs}}, \mathbf{Q}]$ (from Eq. 3).

Monitoring designs that are ignorable *without covariates* are limited to simple random sampling. In this case,

$$[\mathbf{Q} \mid \mathbf{X}, \mathbf{Y}, \phi] = [\mathbf{Q}].$$

However, monitoring programs do not typically collect simple random samples, which means that sampling designs are ignorable only by conditioning on covariates so that

$$[\mathbf{Q} \mid \mathbf{X}, \mathbf{Y}, \phi] = [\mathbf{Q} \mid \mathbf{X}]. \tag{5}$$

The design is ignorable when Eq. 5 is satisfied (Gelman et al. 2013, p. 203) and when ϕ and θ are independent in the prior distribution $[\phi \mid \mathbf{X}, \theta] = [\phi \mid \mathbf{X}]$. It follows that the central challenge in Bayesian inference using complex designs is assuring that the likelihood $[\mathbf{Y}_{\text{obs}} \mid \mathbf{X}, \theta]$ contains design covariates (or equivalent subscripting of parameters) such that

Eq. 5 is true. It is generally impossible to prove this equality when missing data arise unintentionally, as is the case, for example, in non-response to a survey (Gelman and Hill 2009, page 530). However, it is possible to test whether such missing data are ignorable using the approach of Irvine et al. (2018, supplement 1). We do not treat analysis of unintentionally missing data, although many of the principles we describe are germane to their analysis.

3.3. FINITE SAMPLING

Ignorability is particularly relevant to monitoring data because these data are often composed of samples from a finite population. Finite populations can be thought of heuristically as those for which the observed sample units comprise a relatively large fraction of the total number of units that could be sampled without replacement. In this case, the design is not ignorable even when the sample is completely random because information about the sample size n and the population size N must be included in the analysis. Traditionally, finite population inference was achieved by the Horvitz–Thompson estimator for predictions of means and variances using the observed data by including n and N in the inclusion probabilities. By contrast, a Bayesian approach to finite sample inference is to model the *complete* data, including observed and missing values (Link and Barker 2010; Gelman et al. 2013). Missing values are simulated from the uniquely Bayesian posterior predictive distribution. Inference then proceeds using the complete data likelihood (Eq. 1) as illustrated in Sect. 4.3.

3.4. NOTATION FOR IGNORABLE DESIGNS

Proper models for analysis of non-ignorable missing data can be specified equivalently by the matrix $\mathbf{X}$ or by appropriate subscripting of parameters and covariates corresponding to indexing of design covariates included in $\mathbf{X}$. The subscripting approach is somewhat more conventional and transparent. We will use subscripts rather than a design matrix to specify the models that follow.

4. METHODS

4.1. MODEL-BASED INFERENCE

Inference on status and trend in indicators was needed at three scales: site, stratum, and park. We first discuss models for individual sites within strata. Next, we discuss stratum-level inference before turning to inference at the park scale. The general framework described in this section deals with three sources of non-ignorable missingness (items 1, 2, and 3, Sect. 2.3). The approach we outline required the assumption that Eq. 5 is satisfied. We describe how we assured the equality of Eq. 5 in the sections that follow.

4.1.1. Inference on Sites

We modeled data from sites within strata using a general, hierarchical formulation for the posterior and joint distribution of unobserved quantities

$$\mu_{jkt} = g(\alpha_{0jk} + \alpha_{1jk}t + \mathbf{w}'_{jkt}\boldsymbol{\beta}_k) \quad (6)$$

$$\begin{aligned} [\alpha_{0k}, \alpha_{1k}, \boldsymbol{\beta}_k, \sigma_k^2, \boldsymbol{\mu}_{\alpha_k}, \boldsymbol{\Sigma}_k \mid \mathbf{Y}_k] &\propto \prod_{i=1}^{n_j} \prod_{j=1}^{J_k} \prod_{\forall t \in \mathbf{x}_{jk}} [y_{ijkt} \mid h(\mu_{jkt}, \sigma_{jk}^2)] [\alpha_{jk} \mid \boldsymbol{\mu}_{\alpha_k}, \boldsymbol{\Sigma}_k] \\ &\times [\boldsymbol{\beta}_k] [\sigma_k^2] [\boldsymbol{\mu}_{\alpha_k}] [\boldsymbol{\Sigma}_k]. \end{aligned} \quad (7)$$

Bracket notation (Gelfand and Smith 1990) implies that any distribution or density appropriate for the support of the random variable y_{ijkt} could be used. Generality in notation is achieved using the moment matching function $h()$ that returns the parameters of a distribution given its first and second central moments (Hobbs and Hooten 2015, section 3.4.4). The subscript $i = 1, 2, \dots, n_j$ indexes observations within site j ; $j = 1, 2, \dots, J_k$ indexes sites within stratum k , and $k = 1, 2, \dots, K$ indexes strata within parks. The sequential year of observation is given by $t = 1, 2, \dots, T$ where T is the total number of years of monitoring. The design covariate $\mathbf{x}_{jk}$ specifies years during which site j in stratum k was sampled according to its revisit schedule, where one is the first year of sampling. So, for example, $\mathbf{x}_{jk} = (1, 5, 10)'$ if the site was sampled in years one, five, and ten. The parameter α_{0jk} is the intercept, and α_{1jk} is the trend slope in a generalized linear model (linear, exponential, or logit⁻¹) appropriate for the data, notated by the link function $g()$. We note that although we used generalized linear models with a linear term for temporal trend, any function, linear or nonlinear, could be used in place of Eq. 6. Covariates, $\mathbf{w}_{jkt}$, also were chosen to explain spatial and temporal variation in the response, and $\boldsymbol{\beta}$ represented the fixed effects of these covariates. It should be noted that the coefficients $\boldsymbol{\beta}$ can be modeled as fixed at the park scale, fixed or random at the stratum scale, $\boldsymbol{\beta}_k$, or as random site-level effects, $\boldsymbol{\beta}_{jk}$. We typically invoked the simplest park-level fixed effect coefficients unless circumstances (e.g., an expectation of meaningfully different ecological processes operating in each stratum) called for a more complicated model. The covariates $\mathbf{w}_{jkt}$ also included values needed to assure that missing data were ignorable (Eq. 5) when probabilities of inclusion were not equal within strata and years as we illustrate in Sect. 5.3. The flexibility of our approach derives from the ability to modify Eqs. 6 and 7 to accommodate idiosyncrasies in the design and the observations y_{ijkt} .

Each site within a stratum was modeled with its own intercept α_{0jk} , trend slope α_{1jk} , and, usually, its own variance σ_{jk}^2 . Examples of covariates include mechanistic predictors such as annual, seasonal, or monthly data on temperature, precipitation, growing degree days, water balance, nitrogen deposition, and soil pH. Covariates also included design related measures such as cost of travel to sites. Some of these covariates were derived using remotely sensed or gridded spatial datasets; others were measured at each site. Covariates were also used to account for effects of “nuisance” variables, such as changes in observers within the sequence of years.

Intercepts and trend slopes for each site within a stratum were modeled as random variables arising from a joint, stratum-level distribution with vector of means $\boldsymbol{\mu}_{\alpha_k}$ and covariance matrix $\boldsymbol{\Sigma}_k$, $\alpha_{jk} \sim \text{normal}(\boldsymbol{\mu}_{\alpha_k}, \boldsymbol{\Sigma}_k)$, requiring the assumption that sites were exchangeable. We also treated intercepts as random and trend slopes as fixed within each stratum and used model selection (Hooten and Hobbs 2015) to decide whether random slopes improved the predictive ability of the model. The $\boldsymbol{\beta}$ coefficients were considered

fixed at park scale to allow for covariates that could only be obtained for parks (not sites) and to provide for a simpler model. However, their variance–covariance with the intercept and time slope could be modeled with minor modification to Eq. (7).

Priors on all parameters were specified to be vague. Priors on model coefficients were normal centered on zero. Variance of these priors was set to assure that dispersion of the prior was much larger than the dispersion of the marginal posterior of the coefficients, except in the case of inverse logit models, where the variance was set to assure a flat distribution of the prediction of proportions (Hobbs and Hooten 2015). Priors on variances were broad uniform or gamma distributions. Analysis of sensitivity to priors revealed no meaningful effects of priors on marginal posterior distributions of model parameters.

4.2. DATA MISSING BY DESIGN IN TIME

Gaps in data over time created by the revisit schedule (Fig. 2b) created missingness for each site that was non-ignorable without conditioning on covariates (item 1, Sect. 2.3). The design covariate $\mathbf{x}_{jk}$ assured the design was ignorable across years within strata (Eq. 5):

$$q_{jkt} = \begin{cases} 1 & t \in \mathbf{x}_{jk} \\ 0 & t \notin \mathbf{x}_{jk} \end{cases}. \quad (8)$$

Inference on indicators was needed for each site for each year to assess annual status. We used the posterior predictive distribution of the mean $\hat{y}_{jkt}$ of predictions of new observations y_{ijkt}^{new} at each site to indicate status and associated uncertainty at each site during each year:

$$[\hat{y}_{jkt} \mid \mathbf{Y}_k] = \int_{\theta_{1k}} \dots \int_{\theta_{pk}} [E(y_{ijkt}^{\text{new}} \mid \boldsymbol{\theta}_k, \mathbf{Y}_k) [\boldsymbol{\theta}_k \mid \mathbf{Y}_k] d\theta_{1k}, \dots d\theta_{pk}. \quad (9)$$

The vector $\boldsymbol{\theta}_k$ includes the p unobserved quantities in Eq. 7. We approximated $[\hat{y}_{jkt} \mid \mathbf{Y}_k]$ using

$$\hat{y}_{jkt}^m = g(\alpha_{0jk}^m + \alpha_{1jk}^m t + \mathbf{w}_{jkt}' \boldsymbol{\beta}^m) \quad (10)$$

with mean $\mu_{jkt} = 1/M \sum_{m=1}^M \hat{y}_{jkt}^m$. The superscript m indexes a sample in the converged output of a Markov chain Monte Carlo (MCMC) algorithm, and M is the total number of samples in the output. The values of t in Eqs. 9 and 10 span all of the observation periods, $t = 1, 2, \dots, T$, some of which are missing from the design covariate $\mathbf{x}_{jk}$ as a result of the revisit schedule (Fig. 2b). Consequently, $\hat{y}_{jkt}$ are out of sample predictions for the missing years.

Posterior predictions (Eq. 9) depend on fully observed covariates, which may be missing for covariates physically measured at each site (in contrast to remotely sensed ones) during years when sites were not sampled. Finite population inference also requires covariate values for sites that were never sampled. In both cases, the complete data likelihood uses the data we have about the covariate to model the “data we wish we had” (Link and Barker 2010). For example, we might model the covariate hierarchically

$$\gamma_{0jk} \sim [\mu_{\gamma_{0k}}, \varsigma_{\gamma_{0k}}^2] \quad (11)$$

$$w_{jkt} \sim [h(\gamma_{0jk} + \gamma_{1k}t, \sigma_{w_{jk}}^2)]. \quad (12)$$

When the covariate is observed ($w_{jkt}, t \in \mathbf{x}_{jkt}$), it is used to estimate the parameters of the covariate distribution ($\boldsymbol{\gamma}$) and its hyper-parameters ($\mu_{\gamma_{0k}}, \varsigma_{\gamma_{0k}}^2$). When the covariate is unobserved ($w_{jkt}, t \notin \mathbf{x}_{jkt}$), its value is approximated by draws from its distribution (Eq. 12). The example here models a temporal trend in the covariate, but other models, including a mean model, might be used. We include a specific example of modeling missing covariates in Sect. 5.2.

4.3. DATA MISSING BY DESIGN IN SPACE

Predictions of indicators at the stratum scale required finite population inference because the number of sites sampled within stratum k , J_k , was drawn from a finite population with N_k possible samples. Stratum-level means $\hat{y}_{kt}$ were approximated using the integrated, complete data likelihood (Eq. 3) by drawing posterior samples from the observed and missing site-level means $\hat{y}_{jkt}$ as follows. We made a draw $\tilde{j}_k^m$ from discrete uniform($1, N_k$) at each MCMC iteration. Dropping k and m from $\tilde{j}_k^m$ to reduce notational clutter, we computed

$$\hat{y}_{jkt}^m = \begin{cases} g(\alpha_{0\tilde{j}_k}^m + \alpha_{1\tilde{j}_k}^m t + \mathbf{w}_{\tilde{j}_k t}' \boldsymbol{\beta}^m) & \tilde{j} \leq J_k \\ g(\tilde{\alpha}_{0\tilde{j}_k}^m + \tilde{\alpha}_{1\tilde{j}_k}^m t + \mathbf{w}_{\tilde{j}_k t}' \boldsymbol{\beta}^m) & \tilde{j} > J_k \leq N_k \end{cases} \quad (13)$$

The quantities $\tilde{\alpha}_0^m$ and $\tilde{\alpha}_1^m$ were samples from their respective distributions at iteration m . Covariate values for unsampled sites $\tilde{j} > J_k \leq N_k$ were obtained from spatially gridded data corresponding to the location of site $\tilde{j}$ or were modeled using draws of missing covariate values, $w_{\tilde{j}kt}$, by making a draw of γ_{0jk} from the hyper-distribution of the covariate (Eqs. 11 and 12).

Equation 13 shows that the distribution of $\hat{y}_{jkt}^m$ converges on the infinite population case when $N_k \gg J_k$. Because random draws from discrete uniform($1, N_k$) assure that observed sites are sampled in the MCMC in proportion to J_k/N_k and missing sites are sampled in proportion to $(N_k - J_k)/N_k$ for large M ,

$$\hat{y}_{kt} = 1/M \sum_{m=1}^M \hat{y}_{jkt}^m \approx \frac{J_k}{N_k} \sum_{j=1}^{J_k} \hat{y}_{jkt} + \frac{N_k - J_k}{N_k} \sum_{j=J_k+1}^{N_k} \hat{y}_{jkt}^{\text{miss}}, \quad (14)$$

the stratum-level mean for a finite population computed from the complete data likelihood (Gelman et al. 2013).

Analysis at the park scale was not ignorable because samples were allocated following a stratified random design with sites unequally weighted for sampling among strata (item 2, Sect. 2.3). Thus, inference on the mean of indicators for each year across strata required information about how the data were collected. A weighted mean of model predictions across strata (Gelman et al. 2013, page 207) for each year was computed as

$$\hat{y}_{t,\text{park}} = \sum_{k=1}^K \frac{N_k}{N} \hat{y}_{kt} \quad (15)$$

where N_k is the number of potential sampling sites within stratum k and N is the total number of potential sites across all strata. It is not necessary that $J_k / \sum_{k=1}^K J_k = N_k / N$ because finite population Bayesian inference adjusts for over- or under-sampling of strata (Gelman et al. 2013, page 207). Inference on trend at park scales was based on the weighted mean of the trend slope

$$\mu_{\alpha_{1,\text{park}}} = \sum_{k=1}^K \frac{N_k}{N} \mu_{\alpha_{1k}}. \quad (16)$$

The posterior distributions of $\hat{y}_{t,\text{park}}$ and $\mu_{\alpha_{1,\text{park}}}$ were approximated by computing Eqs. 15 and 16 as derived quantities (Hobbs and Hooten 2015) at each MCMC iteration.

There were typically an insufficient number of strata ($K \leq 3$) to model $\mu_{\alpha_{1,\text{park}}}$ hierarchically. It would be possible to model intercepts as parameters of the hyper-distribution $\mu_{\alpha_k} \sim \text{normal}(\alpha_{\text{park}}, \Sigma_{\text{park}})$ if there were a sufficient number of strata and we could assume that the μ_{α_k} are exchangeable among strata. In this case, park-level means would be approximated by first drawing a stratum index $\tilde{k}^m$ with probability N_k / N , computing the mean for a site within stratum $\tilde{k}$ using Eq. 13, and computing $\hat{y}_{t,\text{park}} = 1/M \sum_{m=1}^M \hat{y}_{\tilde{k}t}^m$. The weighted approach in Eq. 16 and the hierarchical approach have been shown to give similar results (Gelman et al. 2013, Figure 8.1).

4.4. UNEQUAL PROBABILITIES OF INCLUSION

The monitoring data we analyzed usually had equal probabilities of inclusion within strata, $\pi_{jkt} = J_{kt} / N_k$, which meant that missingness created by design *within strata and within years* was ignorable (Eq. 5). However, unequal probability of sampling is not unusual in monitoring data sets (item 4, Sect. 2.3). In this case, the covariates describing the model of probability of inclusion in the design must be properly included in the analysis, even if we are not interested in the explanatory value of those covariates (Gelman et al. 2013, page 200). We illustrate such a model in one of the examples that follow (Sect. 5.3).

5. APPLICATIONS

Here, we offer three illustrations of the flexibility of Bayesian hierarchical models for dealing with non-ignorable missing data. Predictions and computations of estimands of interest were based on equations outlined in Sects. 4.2 and 4.3 and are not discussed further here. Instead, the examples here add additional non-ignorable challenges frequently encountered in long-term monitoring designs. In the first examples, we treat the problem of covariates needed for prediction over time that are missing as a result of revisit designs as well as the problem of episodic changes in observers or instrumentation. Next, we deal with unequal probabilities of inclusion of sample sites within strata. Finally, we model data from designs that include censoring and truncation. In all of these cases, proper inference was obtained with relatively minor modifications of Eq. 7.

5.1. MODEL FITTING

Predictor variables were standardized. Coefficients are reported on the standardized scale unless transformations to enhance their interpretation are specifically noted. Samples from the marginal posterior distributions of unobserved quantities were obtained using an MCMC algorithm implemented using JAGS v4.3.0 (Plummer 2003) and the *rjags* package (Plummer 2019) in R (R Core Team 2020). Three chains were accumulated for each model. Convergence was assured by inspection of trace plots and by the diagnostics of Brooks and Gelman (1997). We used posterior predictive checks (Gelman et al. 2013; Conn et al. 2018) to assess lack of fit. Models with Bayesian p -values ≤ 0.3 or ≥ 0.7 were modified to obtain adequate fit. We examined spatial dependence in model residuals using variograms (Cressie 1993), which did not reveal spatial structure for any of the data sets we analyzed. We selected between models with group effects for the intercept and group effects for intercept and slope using posterior predictive loss and the deviance information criterion (Hooten and Hobbs 2015). Coefficients and other quantities of interest were summarized using the median and 95% highest posterior density interval.

5.2. CHANGES IN OBSERVERS AND MISSING COVARIATES

Species richness is a measure of the number of species found in a plant community. We developed a model of native species richness at Little Bighorn Battlefield National Monument, Montana (Fig. 1). Data were yearly counts of native species in 10, 1-m² plots at each site. The design created three sources of non-ignorable missingness: revisit designs, stratification, and a finite population, Sect. 2.3.

All information needed to approximate posterior distributions for intercepts and trend slopes (α_{jk} , Eq. 6) was known because the design was stratified random with a specified revisit schedule. However, we add a nuance in this example, the inclusion of explanatory covariates measured at the site, thus creating non-ignorable missingness in the response and covariate data. The covariates for the species richness model included site-level data on the cover of non-native species and a 0-1 indicator for botanist.

The observer covariate, w_{1jkt} , represents a problem frequently encountered in monitoring programs—changes in personnel collecting data. Such changes are particularly problematic when observers have different levels of experience or training. These shifts in observers can produce spurious changes over time in the response if not properly modeled. In the example here, a botanist less familiar with the local plants in the first two years of observations ($w_{1jkt} = 1$) was replaced by one more familiar in subsequent years ($w_{1jkt} = 0$). We also included line-point intercept data on the cover of non-native species at each site, w_{2jkt} , to understand how competition with a non-native species may influence the number of native species in the community.

Inference on trend at the mean of the standardized covariates did not require values for covariates during all years, because all are zero at their (standardized) means. However, posterior prediction of means of sites (Eq. 10), including the effects of covariates, required values for all years, including years when sites were not sampled. Finite population inference also required covariate values for sites that were never sampled. These requirements are

easily met for any gridded covariates appearing in the model, and for nuisance variables such as observer, which can sensibly be set to the reference level ($w_{1jkt} = 0$). However, data on non-native species cover were collected according to the same revisit schedule as native species richness. As a result, these covariate values were missing during some years at sampled sites and were missing entirely for all years at sites that were never sampled.

We modeled this missingness using $w_{2ijkt} \sim \text{binomial}(r, p_{ijkt})$, where $p_{ijkt} = \text{logit}^{-1}(\gamma_{0jk} + \gamma_{1kt} + \mathbf{v}'_{jkt}\boldsymbol{\delta} + \epsilon_{ijkt})$; $r = 213$ is the total number of point intercepts at each site at which non-native cover was evaluated, and $\mathbf{v}_{jkt}$ included the same 0-1 indicator for botanist as well as a measure of total antecedent rainfall derived using a gridded data product. We used $\epsilon_{ijkt} \sim \text{normal}(0, \sigma_{w_{2jk}}^2)$ to account for overdispersion in non-native cover data. We expanded the likelihood in Eq. 7 to include the covariate distribution

$$[\mathbf{Y}_k, \mathbf{W}_k \mid \boldsymbol{\alpha}_k, \boldsymbol{\beta}, \boldsymbol{\gamma}_k, \boldsymbol{\delta}, \sigma_{w_{2k}}^2] \\ = \prod_{i=1}^{n_j} \prod_{j=1}^{J_k} \prod_{\forall t \in \mathbf{x}_{jk}} \text{Poisson}\left(y_{ijkt} \mid \exp(\alpha_{0jk} + \alpha_{1jkt} + \mathbf{w}'_{jkt}\boldsymbol{\beta})\right) \quad (17)$$

$$\times \text{binomial}(w_{wijkt} \mid r, p_{ijkt}). \quad (18)$$

Including the distribution of the non-native cover covariate in Eq. 17 allowed posterior prediction for all sites in all years (Eqs. 9 and 10). The observed value was used during years when sites were visited and draws from the distribution of values (Eq. 18) informed years when sites were not visited.

Comparison of model predictions and Horvitz–Thompson estimates revealed important differences between the two approaches. Horvitz–Thompson means and confidence intervals were computed using the *spsurvey* package in R (Kincaid and Olsen 2019) on the means of observations at each site because *spsurvey* is not able to model observations within sites, and because it requires site means to justify using a normal likelihood to model count data. Using the site means as perfectly observed data ignores the hierarchical structure in the observations (Fig. 2a). However, the number of observations within sites is not ignorable and must be included in the analysis to represent sampling uncertainty with sites. Failing to model this sampling error produces confidence intervals that are misleadingly narrow. In contrast, the broader Bayesian credible intervals appropriately reflect sampling uncertainty at the site level (Fig. 3).

Although Horvitz–Thompson estimators provide an indication of status, they do not estimate trend (Fig. 3) and cannot account for variables that may explain temporal variation in responses. The Bayesian model suggests that the actual value of native species richness during the first year of sampling was likely much higher than the Horvitz–Thompson estimate, largely as a result of a less experienced botanist identifying 19% fewer species, on average, than the more experienced successor in later years: $\exp(\beta_1) = 0.81$ (0.74, 0.88). The Bayesian model separates changes induced by the observing system from “true” changes in the status of ecological indicators.

Native species richness declined with increasing non-native vegetation: $\beta_2 = -0.04$ (−0.09, 0.02). The model of non-native species cover (Eq. 18), in turn, suggested an increasing trend in non-native vegetative cover, with multiplicative change in odds per

Figure 3. A comparison of Bayesian (gray error bars, dashed gray line) finite population estimates and Horvitz–Thompson (black) estimates of mean native species richness over a 10-year period in an analysis of data from Little Bighorn Battlefield National Monument. Error bars show the 95% confidence intervals for the Horvitz–Thompson means and 95% highest posterior density intervals for the Bayesian predictions. Data (counts) were jittered to avoid obscuring individual observations.

year: $\exp(\gamma_{1,\text{park}}) = 1.12$ (1.06, 1.17). Additional year-to-year variability arose as a result of the influence of rainfall in the prior growing season: $\delta_2 = 0.15$ (0.06, 0.24) (Fig. 4).

Park managers often want to know how alternative actions might affect the trajectory of an indicator of park health. Informing their actions requires covariates that can explain trend. We computed the difference between the posterior distribution of mean native species richness and its expected value under different management scenarios to convey the overall effect of the increasing trend in non-native species cover on native species richness. Specifically, we compared richness estimates conditioned on the actual, time-varying, and increasing cover of non-native vegetation, which increased from 0.15 (0.01, 0.47) to 0.34 (0.08, 0.69) between the first and final year of monitoring, to scenarios in which non-native species cover was held at an alternative value. Figure 4 shows the difference between mean native species richness based on observed non-native cover and a hypothetical scenario in which non-native cover was managed to remain at 5% cover over the period of observation. The results suggest the gains in native species diversity that might be expected with management intervention and, conversely, how much might be lost if no action is taken.

5.3. UNEQUAL PROBABILITIES OF INCLUSION

Probabilities of inclusion of sample units in designs for monitoring programs are sometimes chosen based on a fully observed covariate. For example, the probability of choosing a site for sampling may be inversely related to the travel cost required to obtain a sample from the site. In this case, the treatment assignment mechanism can be ignored if the model used for inference includes the travel cost covariate (Sect. 3.2).

We modeled soil stability at Organ Pipe Cactus National Monument, Arizona (Fig. 1). There were four sources of non-ignorable missingness in this application (revisit designs, stratification, a finite population, and unequal probabilities of inclusion, Sect. 2.3). The probability of inclusion for a site within a stratum was inversely related to the estimated hiking time to each site. Hiking times were binned into five categories, and probabilities of

Figure 4. **a** The Bayesian model of native species richness allowed mean richness to co-vary with a suspected driver, the cover of non-native species. Data on non-native species were collected only during site visits and, thus, had to be modeled. The complete model for richness, which included a modeled covariate, was used to evaluate the likely outcomes of various management scenarios. For instance, in **b** we see the *difference* between native species richness predicted from observations and a prediction in which non-native species cover was hypothetically managed to remain at relatively low (i.e., 5%) cover. Negative values indicate that species richness is, in reality, lower than it could have been had non-native cover been suppressed. All quantities were estimated using methods for small, finite populations (Eq. 13) .

inclusion were assigned to each category. Thus, the parameters in the inclusion model ϕ were known cut-points at the bounds of each category. Making the design ignorable required including hiking time as a fixed effect in the model (Eq. 5).

Observations of soil stability were collected in six ordered classes from low to high using a modification of Herrick et al. (2005). Stability class data were used to infer the properties of the latent, continuously distributed random variable soil stability from which the categories arise (Agresti 2012). Thus, the latent environmental process and its variance were modeled separately from the observation process and its variance (Fig. 5a). Specifically, observations were assumed to arise as a realization of $y_{ijkt} = \text{categorical}(\mathbf{p}_{ijkt})$ where $\mathbf{p}$ is a D length vector specifying the probability an observation falls in category d , $p_d = \Pr(y_{ijkt} = d)$. We specified the likelihood in Eq. 7 as

$$[\mathbf{Y}_k, \mathbf{W}_k \mid \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \mathbf{Z}_k] = \prod_{i=1}^{n_j} \prod_{j=1}^{J_k} \prod_{t \in \mathbf{x}_{jk}} \text{categorical}(y_{ijkt} \mid \mathbf{p}_{ijkt})$$

where $\mathbf{p}_{ijkt} = \left(\int_{\chi_1}^{\chi_2} [z_{ijkt} \mid \text{normal}(\mu_{jkt}, \sigma_{jk}^2)] dz_{ijkt}, \int_{\chi_2}^{\chi_3} [z_{ijkt} \mid \text{normal}(\mu_{jkt}, \sigma_{jk}^2)] dz_{ijkt}, \dots, \int_{\chi_{D-1}}^{\chi_D} [z_{ijkt} \mid \text{normal}(\mu_{jkt}, \sigma_{jk}^2)] dz_{ijkt} \right)'$. We assume that the true, unobserved, continuous value of soil stability z_{ijkt} arises from a normal distribution with parameters μ_{jkt} (Eq. 6) and σ_{jk}^2 . The z_{ijkt} , in turn, give rise to the observations with the known bounds $\chi_1, \chi_2, \dots, \chi_D$ of the soil stability classes 1, 2, ..., D . The covariate vector $\mathbf{w}_{jk}$ (Eq. 6) included gridded estimates of hiking time used to allocate sites to strata and total precipitation in the previous water year.

We compared the Bayesian inference described above with categorical Horvitz–Thompson estimates developed using *spsurvey*. Inputs to the Horvitz–Thompson estimator were annual summaries of site-level median soil stability class. Small sample sizes lim-

ited the range of classes represented in the data in any given year, the effects of which are evident in park-wide Horvitz–Thompson estimates of the proportion of the landscape in each class (Fig. 5b). For instance, site-level medians in the first year of sampling spanned only two of the six stability classes. In contrast, the Bayesian model used all of the data, including individual stability class observations within sites, which produced estimates that better reflected the data by representing the full range of sampling variability (Fig. 5b). We also compared predictions from a model that inappropriately ignores cost of travel to sites to a model that includes hiking time as a covariate. Both of these models (both Bayesian) included previous water year precipitation as an additional covariate to investigate meteorological effects on soil stability mediated by primary productivity and soil organic matter. The results illustrate how ignoring unequal probabilities of inclusion of samples can bias inference (Fig. 5c).

5.4. CENSORED AND TRUNCATED DATA

Environmental data are sometimes censored. Censored observations specify that the true, unobserved value of a response of interest is above, below, or between known cut-points. In contrast to censoring, truncation involves deliberate decisions *not* to collect certain types of observations. For instance, trees below a certain size class might not be measured. As a result, truncation leads to missing information on ecological quantities within some range. Censoring and truncation create designs that are non-ignorable and known (Gelman et al. 2013, page 204). Existing approaches to designed-based inference do not support inference from these types of data (Fox et al. 2015).

We modeled simultaneously censored and truncated data on canopy gaps from grasslands at Capitol Reef National Park, Utah (Fig. 1). Non-ignorable aspects of the design included revisit sampling, stratification, and partially missing data, Sect. 2.3. Canopy gaps are used as an indicator of wind erosion potential (Webb et al. 2020) in long-term monitoring programs throughout arid and semiarid regions of the western USA (Herrick et al. 2005). Canopy gaps were defined as the length (cm) between intersections with the outermost edge of perennial plant canopies along 50 meter transects. Observations were censored whenever a canopy gap spanned either end of the transect. In these cases, we know that the gap was at least as large as the distance between the transect end and the nearest canopy edge, but we do not know the true gap size. Data were truncated as well as censored because canopy gaps smaller than 20 cm were not recorded.

We modified the modeling framework (Eq. 7) as follows. Each site had a vector of data for each visit $\mathbf{u}_{jkt}$ with value one if an observation spanned either end of a transect and hence was censored, and zero when the observation occurred wholly within the transect and was not censored. The data vector $\mathbf{y}_{jkt}$ contained the measurements of uncensored canopy gaps and a code indicating “missing” for censored observations. A third data vector $\mathbf{c}_{jkt}$ contained the cut-point for censoring, that is, the distance between a canopy edge and an end of the transect and a code for missing for the data that were uncensored. We specified the posterior and joint distribution as $[\boldsymbol{\theta}_k \mid \mathbf{U}_k, \mathbf{Y}_k, \mathbf{C}_k] \propto [\mathbf{U}_k \mid \mathbf{Y}_k, \mathbf{C}_k][\mathbf{Y}_k \mid \boldsymbol{\theta}_k][\boldsymbol{\theta}_k]$ because we can reasonably assume that the value of u_{ijkt} is conditionally independent of $\boldsymbol{\theta}_k$ when the values of y_{ijkt} and c_{ijkt} are known.

We specified $[u_{ijkt} \mid y_{ijkt}, c_{ijkt}]$ by first defining the indicator function

$$\mathbb{I}_{c_{ijkt}}(y_{ijkt}) = \begin{cases} 1 & \text{if } y_{ijkt} \geq c_{ijkt} \\ 0 & \text{otherwise.} \end{cases}$$

We used this function to formulate $f(u_{ijkt} \mid y_{ijkt}, c_{ijkt}) = \begin{cases} 1 & \text{if } u_{ijkt} = 0 \\ \mathbb{I}_{c_{ijkt}}(y_{ijkt}) & \text{if } u_{ijkt} = 1, \end{cases}$ which we used to form a joint likelihood in Eq. 7

$$[\mathbf{Y}_k, \mathbf{U}_k, \mathbf{C}_k \mid \boldsymbol{\theta}_k] = \prod_{i=1}^{n_j} \prod_{j=1}^{J_k} \prod_{t \in \mathbf{x}_{jk}} f(u_{ijkt} \mid y_{ijkt}, c_{ijkt}) \text{ gamma}\left(y_{ijkt} \mid \frac{\mu_{jkt}^2}{\sigma_{jk}^2}, \frac{\mu_{jkt}}{\sigma_{jk}^2}\right) T(20, \infty) \quad (19)$$

where $T(20, \infty)$ indicates truncation of the left tail of the gamma distribution at 20 cm.

Equation 19 can be understood this way. The observation of gap size y_{ijkt} is missing when the data are censored ($u_{ijkt} = 1$). In this case, the likelihood $[u_{ijkt} \mid y_{ijkt}, \mathbf{c}_{ijkt}]$ causes the observation to be imputed from the distribution of y_{ijkt} to the right of c_{ijkt} . Imputation is not necessary when the data are not censored ($u_{ijkt} = 0$); the likelihood $[u_{ijkt} \mid y_{ijkt}, \mathbf{c}_{ijkt}]$ evaluates to one, and the joint likelihood simplifies to a single component appropriate for the uncensored data.

We compared analysis from the model including censoring and truncation (Eq. 19) with analysis from two models ignoring them. Omitting censoring and truncation produced faulty inference (Fig. 6). The Horvitz–Thompson estimates, which ignored incomplete observations, once again revealed excessively narrow confidence intervals (Fig. 6a) relative to the Bayesian credible intervals. We also compared two Bayesian models, the first of which accounted for censored and truncated observations, the second of which ignored them. Mean canopy gap size was similar for the two approaches: 154 (41, 346) cm vs. 152 (57, 294) cm (Fig. 6b). However, the standard deviation of canopy gap size was substantially higher when censored and truncated observations were modeled appropriately: 240 (34, 532) cm versus 172 (37, 375) cm when ignored (Fig. 6b). This is likely because modeling censoring and truncation acknowledges that censored gap size observations represent their known lower limits, not their “true” sizes, and that gaps < 20 cm were present, but unmeasured.

When censor limits are included in the model, the lengths of censored gaps are imputed on the basis of the estimate of all model parameters at each iteration of the MCMC. Inputted observations are at least as large as the censor limits. Thus, the largest observations when censoring is properly modeled exceed those obtained when censoring is ignored: 9090 (4144, 20,595) cm versus 4979 (2599, 9964) cm (Fig. 6b). It can also be shown that it is precisely the largest gaps that tend to be among the censored observations. This result has clear significance for management where erosion is a concern. A model that ignores censoring and truncation will underestimate the proportion of large gaps, which are especially prone to erosion. A manager misled by an incomplete analysis might believe that erosive potential of the landscapes is lower than it actually is.

Figure 6. **a** A comparison of Bayesian (gray error bars, dashed gray line) and Horvitz–Thompson (black) estimates of mean canopy gap size over a 10-year period in an analysis of data from a single stratum (“Hartnet-deep grassland”) at Capitol Reef National Park. Error bars show the 95% confidence intervals for the Horvitz–Thompson means and 95% highest posterior density intervals for the Bayesian predictions. Model estimates are seen against the *uncensored* gap size measurements in each year. Observations, all of which appear above the lower truncation limit at 20 cm, are shown on a logarithmic scale and were jittered to avoid obscuring individual values. **b** Differences between summary measures of canopy gap size from two Bayesian models, one in which censoring and truncation in the data are modeled (gray probability density estimates) versus another in which they are ignored (white). Summaries include the posterior distributions of the mean (left), standard deviation (center), and maximum predicted canopy gaps size (right).

6. DISCUSSION

Bayesian hierarchical models offer a coherent and flexible approach to modeling observations that are “missing by design.” We illustrated the connection between sampling design and analysis in five ways (Sect. 2.3) using data from the National Park Service Inventory and Monitoring program. The primary advantages of the Bayesian hierarchical approach for analyzing monitoring data are (1) the use of posterior predictions to fill in gaps in data missing by design, (2) formal inference on temporal trend, (3) the ability to model explanatory covariates, (4) a rigorous approach for inference on finite populations based on the complete data likelihood, and (5) the ability to model the observation process separately from the environmental process. Several of these advantages were also identified by [Shanahan et al. \(2021\)](#). There are disadvantages, however. Model building, coding, and checking require greater effort than simple computation of Horvitz–Thompson means and variances.

Relatively large data sets are required because many more parameters are modeled relative to the Horvitz–Thompson approach.

The Bayesian approach provides a particularly useful way to model temporal trend from sites with unequal frequency of sampling. The hierarchical model in Eq. 7 and its implementation in posterior prediction (Eqs. 9 and 10) support “borrowing strength” (Hobbs and Hooten 2015) for inference on the intercept and trend slope for each site. Sites with low uncertainty in estimates of α_{jk} have a greater influence on μ_{α_k} than sites with high uncertainty. Borrowing strength provides a natural weighting of influence of sites sampled in revisit designs.

Environmental managers often need inference that extends beyond simple means and variances describing the status of resources at different points in time. This need can only be met by model-based inference. Inferring trend from observing a series of means and confidence intervals over time in the absence of a properly fit model can produce spurious inferences, as we illustrated in Sect. 5.2. A related issue is that merely knowing that change is occurring, as might be inferred from a time-slope term in a model, is rarely sufficient for informed decision making. An ability to infer the drivers of change enabled by modeling coefficients of time-varying covariates is also needed (Shanahan et al. 2021) to inform managers of potential actions that could be taken to modify the trajectory of indicators of park health.

Design-based analysis using Horvitz–Thompson means and variances was originally motivated by the need for inference on finite populations (Horvitz and Thompson 1951, 1952), a challenge frequently confronted by monitoring programs. Generalized linear models supporting inference on trend can be fit to data from complex sampling designs in the maximum likelihood framework using sample weights to provide information about the design that must be included in the analysis (Lumley 2004, 2015; Lumley and Scott 2017). Although these approaches have been used primarily for data in the social sciences, there is no reason they could not be applied to complex designs used for monitoring natural resources (e.g., Irvine et al. 2018). However, inferences from generalized linear models using sampling weights must be marginalized across levels in multi-level, hierarchical models (Hester 1991; Lumley and Scott 2017). For example, it would not be possible to estimate separately trend for site, strata, and park levels using sampling weights. The need to marginalize across levels strongly limits inference from this type of analysis. The Bayesian approach we describe does not suffer from this limitation.

Modern statistical analysis requires software (Ripley 2002). This ubiquitous need can tempt analysts of monitoring data to use existing statistical packages in ways that fail to match the data and the way they were collected to the algorithms implemented in the software. Such failures might include applying analyses assuming normal distributions to observations with strictly nonnegative or discrete support, ignoring the need for mixture models as would occur with zero inflation, treating categorical data as metric, discarding censored data, and failing to respect the finite nature of a population, among others. Overlooking these unique features of data can be recommended only on the basis of its convenience. We have properly modeled all of these idiosyncrasies using relatively minor modifications of Eq. 7.

The analysis of environmental monitoring data has historically used design-based inference (Horvitz and Thompson 1951, 1952) that lacks the flexibility to deal with the diverse

challenges that inevitably arise in analysis of large monitoring data sets. The concept of ignorability provides a general theoretical foundation for constructing model-based inference that appropriately includes information about the way data were collected in the analysis. The heart of this challenge is to model data that are “missing by design.” The Bayesian, hierarchical approach can rise to this challenge, thereby offering a broadly useful framework for analyzing data collected by environmental monitoring programs.

ACKNOWLEDGEMENTS

We thank Carolyn Livensperger for her help developing Horvitz–Thomson estimates, Dusty Perkins for his leadership in funding the work, and Mevin Hooten and two anonymous referees for insightful, constructive reviews. Funding for this project was provided under Cooperative Agreement P18AC00340 between the United States Department of the Interior National Park Service and Conservation Science Partners (Task Agreement Number P18AC01394).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Received May 2021. Revised August 2021. Accepted August 2021. Published Online September 2021.]

REFERENCES

- Agresti A (2012) Categorical data analysis, 3rd edn. Wiley, Hoboken
- Brooks SP, Gelman A (1997) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7:434–455
- Conn PB, Johnson DS, Williams PJ, Melin SR, Hooten MB (2018) A guide to Bayesian model checking for ecologists. *Ecol Monogr* 88:526–542
- Courbois JYP, Urquhart NS (2004) Comparison of survey estimates of the finite population variance. *J Agric Biol Environ Stat* 9:236–251
- Cressie NA (1993) Statistics for spatial data. Wiley, New York
- Fancy SG, Gross JE, Carter SL (2009) Monitoring the condition of natural resources in US national parks. *Environ Monit Assess* 151:161–174
- Fox GA, Negrete-Yankelevich S, Sosa VJ (2015) Ecological statistics: contemporary theory and application. Oxford University Press, Oxford
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85:398–409
- Gelman A (2007) Struggles with survey weighting and regression modeling. *Stat Sci* 22:153–164
- Gelman A, Hill J (2009) Data analysis using regression and multilevel/hierarchical modeling. Cambridge University Press, Cambridge
- Gelman A, Carlin JB, Stern HS, Dunson D, Vehhtari A, Rubin DB (2013) Bayesian data analysis. Chapman and Hall / CRC, London

- Herrick JE, Van Zee JW, Havstad KM, Burkett LM, Whitford WG (2005) Monitoring Manual for Grassland, Shrubland and Savanna Ecosystems, USDA-ARS Jornada Experimental Range
- Hester FE (1991) The U.S. National Park Service experience with exotic species. *Nat Areas J* 11:127–128
- Hobbs NT, Hooten MB (2015) Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton
- Hooten MB, Hobbs NT (2015) A guide to Bayesian model selection for ecologists. *Ecol Monogr* 85:3–28
- Horvitz DG, Thompson DJ (1951) A generalization of sampling without replacement from a finite universe. *Ann Math Stat* 22:315
- Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc* 47:663–685
- Irvine KM, Rodhouse TJ, Wright WJ, Olsen AR (2018) Occupancy modeling species-environment relationships with non-ignorable survey designs. *Ecol Appl* 28:1616–1625
- Kincaid TM, Olsen AR (2019) spsurvey: spatial survey design and analysis
- Kincaid TM, Larsen DP, Urquhart NS (2004) The structure of variation and its influence on the estimation of status: indicators of condition of lakes in the northeast, U.S.A. *Environ Monit Assess* 98:1–21
- Larsen DP, Kincaid TM, Jacobs SE, Urquhart NS (2001) Designs for evaluating local and regional scale trends. *Bioscience* 51:1069–1078
- Link WA, Barker RJ (2010) Bayesian inference with ecological applications. Academic Press, New York
- Lumley T (2004) Analysis of complex survey samples. *J Stat Softw* 9:1–19
- Lumley T (2015) Survey: analysis of complex survey samples. R Package Volume 3
- Lumley T, Scott A (2017) Fitting regression models to survey data. *Stat Sci* 32:265–278
- Messer JJ, Linthurst RA, Overton WS (1991) An EPA program for monitoring ecological status and trends. *Environ Monit Assess* 17:67–78
- Olsen AR, Sedransk J, Edwards D, Gotway CA, Liggett W, Rathbun S, Reckhow KH, Young LJ (1999) Statistical issues for monitoring ecological and natural resources in the United States. *Environ Monit Assess* 54:1–45
- Overton WS, Stehman SV (1994) Variance estimation in the EMAP strategy for sampling discrete ecological resources. *Environ Ecol Stat* 1:133–149
- Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. DSC Working Papers <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/>. Proceedings of the 3rd International Workshop on Distributed Statistical Computing, March 20–22, 2003, Technische Universität Wien, Vienna, Austria
- Plummer M (2019) Rjags: Bayesian graphical models using MCMC. R package version 4-10
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Ripley BD (2002) Statistical methods need software: a view of statistical computing. Opening lecture of the Royal Statistical Society. <http://www.stats.ox.ac.uk/~ripley/RSS2002.pdf>
- Shanahan E, Wright WJ, Irvine KM (2021) Adaptive monitoring in action: Reconsidering design-based estimators reveals underestimation of whitebark pine disease prevalence in the greater yellowstone ecosystem. *J Appl Ecol* (in press)
- Starcevic L, Mitchell J (2017) Trendnps: NPS trend analyses for complex survey designs. Technical report, R package version 0.0.3
- Stehman SV, Overton WS (1994) Comparison of variance estimators of the Horvitz–Thompson estimator for randomized variable probability systematic-sampling. *J Am Stat Assoc* 89:30–43
- Stevens DL, Olsen AR (2004) Spatially balanced sampling of natural resources. *J Am Stat Assoc* 99:262–278
- Theobald DM, Stevens DL, White D, Urquhart NS, Olsen AR, Norman JB (2007) Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environ Manag* 40:134–146
- Webb NP, Kachergis E, Miller SW, McCord SE, Bestelmeyer BT, Brown JR, Chappell A, Edwards BL, Herrick JE, Karl JW et al (2020) Indicators and benchmarks for wind erosion monitoring, assessment and management. *Ecol Indic* 110:105881

Williams BK, Brown ED (2019) Sampling and analysis frameworks for inference in ecology. *Methods Ecol Evol* 10:1832–1842

Woodworth GG (2004) *Biostatistics: a Bayesian introduction*. Wiley, Hoboken

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.