Hierarchical computing for hierarchical models in ecology

Hanna M. McCaslin
Department of Fish, Wildlife, and Conservation Biology
Colorado State University
1484 Campus Delivery
Fort Collins, CO 80523

Abigail B. Feuka
Department of Fish, Wildlife, and Conservation Biology
Colorado State University

Mevin B. Hooten
U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit
Department of Fish, Wildlife, and Conservation Biology
Department of Statistics
Colorado State University

Corresponding author:

Hanna McCaslin, `hanna.mccaslin@colostate.edu`

Citation: McCaslin, H.M., A.B. Feuka, and M.B. Hooten. (In Press). Hierarchical computing for hierarchical models in ecology. *Methods in Ecology and Evolution.*

Abstract

1. Bayesian hierarchical models allow ecologists to account for uncertainty and make inference at multiple scales. However, hierarchical models are often computationally intensive to fit, especially with large data sets, and researchers face trade-offs between capturing ecological complexity in statistical models and implementing these models.

2. We present a recursive Bayesian computing (RB) method that can be used to fit Bayesian models efficiently in sequential MCMC stages to ease computation and streamline hierarchical inference. We also introduce transformation-assisted RB (TARB) to create unsupervised MCMC algorithms and improve interpretability of parameters. We demonstrate TARB by fitting a hierarchical animal movement model to obtain inference about individual- and population-level migratory characteristics.

3. Our recursive procedure reduced computation time for fitting our hierarchical movement model by half compared to fitting the model with a single MCMC algorithm. We obtained the same inference fitting our model using TARB as we obtained fitting the model with a single algorithm.

4. For complex ecological statistical models, like those for animal movement, multi-species systems, or large spatial and temporal scales, the computational demands of fitting models with conventional computing techniques can limit model specification, thus hindering scientific discovery. Transformation-assisted RB is one of the most accessible methods for reducing these limitations, enabling us to implement new statistical models and advance our understanding of complex ecological phenomena.

Keywords: Bayesian filtering, MCMC, parallel computing, recursive, transformation

Introduction

Ecological systems are characterized by dynamics and uncertainty at many scales, but observing all relevant scales may be difficult or impossible (Wiens 1989). Instead, we must use models to scale and connect processes across multiple levels (Levin 1992), such as from the scale of observation to the hypothesized scale of biological process, or from a single individual or species to a population or community. For example, in movement ecology, we often collect telemetry data and observe movement at the individual-level, but wish to make inference on the population as a whole, like to better understand responses to environmental conditions that are similar among individuals (Hooten et al. 2016). Alternatively, modeling ecosystems or ecological communities often involves joint analysis of many taxonomic groups as well as the processes that connect them (Levin 1992, Warton et al. 2015). Finally, conducting ecological studies introduces additional uncertainty, including sampling and detection uncertainty as well as spatial and temporal variation between study sites and years, which must be considered when specifying ecological models (Royle and Dorazio 2008, Beissinger et al. 2016).

Bayesian hierarchical modeling has become a popular tool in ecology, facilitating scaling by relating process models at one level to parameters at another level (Royle and Dorazio 2008, Hobbs and Hooten 2015). Hierarchical models are flexible and facilitate the inclusion of multiple sources of uncertainty in the data, process, and parameter components (Berliner 1996, Cressie et al. 2009). For example, many integrated population models (IPMs) use a Bayesian hierarchical framework to integrate multiple data sources to understand population dynamics and demographic processes (Schaub and Abadi 2011). However, IPMs and other hierarchical models can quickly become large and time-consuming to fit.

Ecological science has seen a rapid increase in the availability of big data, advanced statistical techniques, and collaborative research, and our ability to specify ecological models that capture more of the complexity of natural phenomena has improved substantially as a result (McCallen et al. 2019). However, many ecologists have also reached the point where

computational demands limit what can be modeled. Further, as ecologists are increasingly interested in long-term monitoring and prediction (Dietze et al. 2008), statistical models must be fit each time data are added. Collaborations with computer and data scientists and new software packages for efficient computing have introduced sophisticated computational techniques (e.g., distributed computing) in ecological science, but barriers to wide implementation of these approaches are a bottleneck for advancing ecological modeling (Visser et al. 2015, Hampton et al. 2017). Therefore, more accessible approaches for reducing computational limitations are needed to support progress in ecological modeling and understanding.

Recursive computing techniques, also known as batch or modular computing or Bayesian filtering, are used to fit a statistical model in a series of steps (Särkkä 2013). These techniques simplify computing at each step, without modifying the original model specification or resulting inference. One recursive Bayesian computing (RB) method, introduced by Lunn et al. (2013), leverages the properties of Markov Chain Monte Carlo (MCMC) sampling (Gelfand and Smith 1990) to lessen the computational burden of fitting hierarchical models. The authors used RB to reconcile the results of several independent studies in a meta-analysis (Lunn et al. 2013), and the method has been applied in ecological contexts to facilitate online updating (Hooten et al. 2020), model individual and group variation in physiological measurements (Hooten and Hefley 2019), and scale movement and resource-selection models from individuals to populations (Hooten et al. 2016, Gerber et al. 2018). While not unique to ecology, RB is a natural computational technique for ecologists to consider because the RB framework mirrors many ecological study designs and hierarchical models.

Consider a study of invasive cheatgrass (*Bromus tectorum*) occurrence in grasslands in Montana, in the northwestern United States (Pearson et al. 2018). Cheatgrass occurrence was monitored at 20 grassland sites by sampling 20 randomly selected 1-m^2 plots within each site. Suppose we want to model the probability of cheatgrass occurrence y_{ij} in Montana

grasslands using a Bernoulli generalized linear mixed model (GLMM) specified as

$$\mathrm{y}_{ij} \sim \mathrm{Bern}(p_j), \qquad i = 1, ..., N, j = 1, ..., J, \tag{1}$$

$$\mathrm{logit}(p_j) \sim \mathrm{N}(\mu, \sigma^2), \tag{2}$$

$$\mu \sim \mathrm{N}(\mu_0, \sigma_0^2), \tag{3}$$

$$\sigma^2 \sim \mathrm{IG}(q, r), \tag{4}$$

where j indexes sites and i indexes plots within each site. In this model, p_j is the probability of cheatgrass at site j, and $\mathrm{logit}(p_j)$ arises from a Gaussian distribution with study-wide parameters μ and σ^2, arising from Gaussian and inverse gamma distributions, respectively (Fig. 1). Thus, p_j are "random effects" because they will vary for each site but will arise from a single underlying distribution. We use Gaussian random effects, with the logit link function to constrain p_j to the proper support, and seek inference on μ. The full-conditional distributions for the $\mathrm{logit}(p_j)$ are not analytically tractable, so the $\mathrm{logit}(p_j)$ cannot be sampled using Gibbs updates and will need to be tuned individually to fit the model (Gelfand and Smith 1990). This minimal example could be fit in a single, conventional MCMC algorithm, but we describe the procedure to fit it recursively to demonstrate RB methods.

We could fit this model using RB by first partitioning the data by site, $\mathbf{Y} = (\mathbf{y}_1', ..., \mathbf{y}_J')'$. These individual partitions would be analyzed independently in a first-stage MCMC algorithm with a temporary prior for $\mathrm{logit}(p_j)$ to obtain temporary posterior distributions for the parameters $\mathrm{logit}(p_j)$. Then, the resulting temporary posterior distributions would be used as proposals in the second-stage algorithm to update the study-wide parameters μ and σ^2, and the $\mathrm{logit}(p_j)$ given μ and σ^2 (Lunn et al. 2013). However, we would still need to tune the updates for each $\mathrm{logit}(p_j)$ by hand in the first stage, because the full-conditional distributions are not analytically tractable. This would slow model fitting and may be difficult.

Instead, we propose a modification of RB, which we call transformation-assisted RB

(TARB), to eliminate tuning in the first stage and ease model fitting with unsupervised algorithms and efficient Gibbs updates. In what follows, we demonstrate how to implement RB and TARB to fit ecological models and apply TARB to a hierarchical movement model for avian migration to make individual- and population-level inference. Additionally, we discuss the implementation of TARB to other ecological models to illustrate its wide applicability.

Methods

Our Bernoulli GLMM is a hierarchical model comprised of data, process, and parameter components (Berliner 1996), with a set of latent random effects $\boldsymbol{\theta}_j = \text{logit}(p_j)$ for $j = 1, ..., J$ (Fig. 1). The group-level parameters $\boldsymbol{\psi} = (\mu, \sigma^2)'$, which correspond to the full study area in our example, describe the distribution underlying the partition-level (e.g., site-level) parameters $\boldsymbol{\theta}_j$. For data partitioned $\mathbf{Y} = (\mathbf{y}'_1, ..., \mathbf{y}'_J)'$, this can be written

$$\mathbf{y}_j \sim [\mathbf{y}_j|\boldsymbol{\theta}_j], \qquad j = 1, ..., J, \tag{5}$$

$$\boldsymbol{\theta}_j \sim [\boldsymbol{\theta_j}|\boldsymbol{\psi}], \tag{6}$$

$$\boldsymbol{\psi} \sim [\boldsymbol{\psi}]. \tag{7}$$

Note that square brackets $[\cdot]$ denote probability distributions (Gelfand and Smith 1990).

In general, $\boldsymbol{\theta}_j$ could be an $m \times 1$ vector that describes the partition-level process with m covariates. The data partitions $\mathbf{y}_j$ do not need to be equal-sized, and can represent any natural data subset such as different field sites as in our example, telemetry fixes for distinct individuals, results from several studies in a meta-analysis, or data on different species in a community, as long as dependence within the data partitions is accounted for in the data or process models.

The RB approach presented by Lunn et al. (2013) is carried out by specifying prior distributions $[\boldsymbol{\theta}_j]$ in the first-stage to obtain a sample from the posterior distributions $[\boldsymbol{\theta}_j|\mathbf{y}_j] \propto [\mathbf{y}_j|\boldsymbol{\theta}_j][\boldsymbol{\theta}_j]$ for each partition $j = 1, ..., J$ independently. Next, the hierarchical

model in (5)-(7) is fit using a second-stage MCMC algorithm with Metropolis-Hastings (MH) updates for $\boldsymbol{\theta}_j$, in which random samples from the temporary, first-stage posterior distributions for $\boldsymbol{\theta}_j$ are used as the proposals $\boldsymbol{\theta}_j^{(*)}$. This eliminates the need for tuning in the second-stage MH updates. Also in the second-stage algorithm, the group-level parameters $\boldsymbol{\psi}$ are updated based on their full-conditional distributions $[\boldsymbol{\psi}|\cdot] \propto \big(\prod_{j=1}^{J}[\boldsymbol{\theta}_j|\boldsymbol{\psi}]\big)[\boldsymbol{\psi}]$. The MH acceptance probability for each $\boldsymbol{\theta}_j^{(*)}$ is $\min(r_j^{(*)}, 1)$ where

$$r_j^{(k)} = \frac{[\mathbf{y}_j|\boldsymbol{\theta}_j^{(*)}][\boldsymbol{\theta}_j^{(*)}|\boldsymbol{\psi}^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}|\mathbf{y}_j]}{[\mathbf{y}_j|\boldsymbol{\theta}_j^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}|\boldsymbol{\psi}^{(k-1)}][\boldsymbol{\theta}_j^{(*)}|\mathbf{y}_j]}, \tag{8}$$

$$= \frac{[\mathbf{y}_j|\boldsymbol{\theta}_j^{(*)}][\boldsymbol{\theta}_j^{(*)}|\boldsymbol{\psi}^{(k-1)}][\mathbf{y}_j|\boldsymbol{\theta}_j^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}]}{[\mathbf{y}_j|\boldsymbol{\theta}_j^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}|\boldsymbol{\psi}^{(k-1)}][\mathbf{y}_j|\boldsymbol{\theta}_j^{(*)}][\boldsymbol{\theta}_j^{(*)}]}, \tag{9}$$

$$= \frac{[\boldsymbol{\theta}_j^{(*)}|\boldsymbol{\psi}^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}]}{[\boldsymbol{\theta}_j^{(k-1)}|\boldsymbol{\psi}^{(k-1)}][\boldsymbol{\theta}_j^{(*)}]}, \tag{10}$$

for MCMC iteration $k = 1, ..., K$. Notably, neither the MH ratio (10) nor the full-conditional distributions for $\boldsymbol{\psi}$ involve the data $\mathbf{y}$. For the data model to cancel in the numerator and denominator of the MH ratio (10), the proposals $\boldsymbol{\theta}_j^{(*)}$ should be independent draws from the first-stage posterior distributions for $\boldsymbol{\theta}_j$. Thus, in practice, we sample $\boldsymbol{\theta}_j^{(*)}$ randomly with replacement from the first-stage Markov chains so that the samples are uncorrelated (Lunn et al. 2013, Hooten et al. 2020).

If the hierarchical model is specified such that the conditional distributions for $\boldsymbol{\theta}_j$ are not analytically tractable, like in our GLMM, then the first stage of the model must be fit using MH or importance sampling (Geweke 1989) which must be tuned by the user for each partition (Hooten et al. 2016). Thus, rather than specifying a first-stage prior directly on $\boldsymbol{\theta}_j$, we use TARB and specify a prior $[\mathbf{g}(\boldsymbol{\theta}_j)]$ on a transformation $\mathbf{g}(\boldsymbol{\theta}_j)$ of the parameters $\boldsymbol{\theta}_j$. It is most advantageous to specify $\mathbf{g}$ so that the first-stage priors on $\mathbf{g}(\boldsymbol{\theta}_j)$ are conjugate with the data model to allow us to use an automated Gibbs sampler in the first stage. In GLMMs and other hierarchical models, we often specify models so that parameters and random effects arise from Gaussian distributions, and use a link function

to constrain these parameters to the appropriate support. Thus, in these cases, $\mathbf{g}$ will likely be a back-transformation (i.e., the inverse of the link function) that allows us to specify conjugate first-stage priors. However, unlike if we were to specify a different model to facilitate conjugacy, using TARB allows us to incorporate prior knowledge and obtain inference in terms of the original model specification. For example, if we let $\mathbf{g}(\boldsymbol{\theta}_j) = \text{logit}^{-1}(\boldsymbol{\theta}_j)$ in our cheatgrass example, then we can specify a temporary beta prior on p_j in the first-stage. In this example, the benefit of doing so extends beyond conjugacy to a first-stage posterior distribution that can be written analytically, and therefore does not require MCMC to sample. We provide the complete procedure to fit the cheatgrass GLMM using TARB, with code, in the Supporting Information (Appendix A).

We need to use the resulting first-stage posterior distribution as a proposal distribution in the second-stage MCMC algorithm, but the first stage posterior distribution $[\mathbf{g}(\boldsymbol{\theta}_j)|\mathbf{y}_j]$ is on the transformed parameters $\mathbf{g}(\boldsymbol{\theta}_j)$. Thus, to account for the first-stage prior on transformed parameters, we must modify the MH ratio (10) and use a change of variables technique to ensure the proposal is on the same transformation that appears in the process component (6) of the original hierarchical model. While we could easily use the first-stage posterior distribution to obtain a *sample* from the desired posterior distribution $[\boldsymbol{\theta}_j|\mathbf{y}_j]$, the MH ratio requires us to evaluate the probability density function $[\boldsymbol{\theta}_j|\mathbf{y}_j]$ rather than sample from it. There are many possible methods for obtaining this distribution, including analytical change of variable techniques and numerical approaches. For continuous random variables, we use a change of variables technique where

$$[\boldsymbol{\theta}_j|\mathbf{y}_j] = [\mathbf{g}(\boldsymbol{\theta}_j)|\mathbf{y}_j]|\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j))|, \tag{11}$$

in which $\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j\ ))$ is the Jacobian matrix defined as

$$\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j)) \equiv \begin{bmatrix} \frac{\delta g_1(\boldsymbol{\theta}_j)}{\delta \theta_{j,1}} & \cdots & \frac{\delta g_1(\boldsymbol{\theta}_j)}{\delta \theta_{j,p}} \\ \vdots & \ddots & \vdots \\ \frac{\delta g_{pg}(\boldsymbol{\theta}_j)}{\delta \theta_{j,1}} & \cdots & \frac{\delta g_{pg}(\boldsymbol{\theta}_j)}{\delta \theta_{j,p}} \end{bmatrix}. \tag{12}$$

The Jacobian matrix consists of partial derivatives of each element of $\mathbf{g}(\boldsymbol{\theta}_\mathbf{j})$ with respect to each element of $\boldsymbol{\theta}_j$. Its determinant $|\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_\mathbf{j}))|$ maps the change in the transformed variables to the change in the non-transformed variables ($d\mathbf{g}(\boldsymbol{\theta}_\mathbf{j})$ onto $d\boldsymbol{\theta}_j$), yielding the correct probability distribution of the non-transformed variable when multiplied to the probability distribution of the transformed variable. Thus, substituting (11) for the proposal in the second-stage MH ratio (10) results in

$$r_j^{(k)} = \frac{[\mathbf{y}_j|\boldsymbol{\theta}_j^{(*)}][\boldsymbol{\theta}_j^{(*)}|\boldsymbol{\psi}^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}|\mathbf{y}_j]}{[\mathbf{y}_j|\boldsymbol{\theta}_j^{(k-1)}][\boldsymbol{\theta}_j^{(k-1)}|\boldsymbol{\psi}^{(k-1)}][\boldsymbol{\theta}_j^{(*)}|\mathbf{y}_j]}, \tag{13}$$

$$= \frac{[\boldsymbol{\theta}_j^{(*)}|\boldsymbol{\psi}^{(k-1)}][\mathbf{g}(\boldsymbol{\theta}_j)^{(k-1)}]|\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j)^{(k-1)})|}{[\boldsymbol{\theta}_j^{(k-1)}|\boldsymbol{\psi}^{(k-1)}][\mathbf{g}(\boldsymbol{\theta}_j)^{(*)}]|\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j)^{(*)})|}. \tag{14}$$

The data component of the hierarchical model cancels in the MH ratio (14) associated with the second-stage MCMC algorithm regardless of the transformation used in the first-stage temporary prior, and we account for the transformation via the determinant of the Jacobian in the modified TARB ratio (14). In our cheatgrass GLMM, because $\boldsymbol{\theta}_j = p_j$ is a scalar, the Jacobian simplifies to the derivative of $g = \text{logit}^{-1}(p_j)$ with respect to $\text{logit}(p_j)$ (Appendix A). Thus, we can use TARB to create unsupervised first-stage algorithms that can be easily parallelized and a second-stage MCMC algorithm that does not rely on the data model. This results in substantial computational savings when the data model is complex or there are many data models to fit and allows the second stage to be updated easily if new data partitions become available.

Application: White Stork Migration

To demonstrate TARB, we developed a hierarchical animal movement model for the migratory behavior of white storks (*Ciconia ciconia*) in western Europe to obtain individual- and population-level inference for migration characteristics. We analyzed data from $J = 15$ individuals tracked with GPS units from 30 July 2018 – 29 Sept 2018 (Fig. 2, Cheng et al. 2019, Fiedler et al. 2019). These data are available in the R package 'moveVis' (Schwalb-Willmann et al. 2020).

Model statement

We specified a continuous-time hierarchical model for stork movement with the data component

$$\mathbf{s}_j(t_i) \sim \mathrm{N}(\mathbf{s}_j(t_{i-1}) - \nabla p(\mathbf{s}_j(t_i), \boldsymbol{\beta}_j)dt_i, \sigma_j^2 dt_i \mathbf{I}), \tag{15}$$

where $\mathbf{s}_j(t_i)$ is the measured position of individual j at time i (for $j = 1, ..., J$ and $i = 1, ..., n_j$). We defined the potential function in (15) as $p(\mathbf{s}, \boldsymbol{\beta}_j) \equiv \mathbf{x}'(\mathbf{s})\boldsymbol{\beta}_j$, which describes a surface upon which an individual is more likely to move "downhill" (Brillinger 2010, Hooten et al. 2017). In our specification, this surface is a linear function of covariates $\mathbf{x}(\mathbf{s})$ and will influence the speed and directional persistence of movement. The term dt_i represents the change in time between successive positions $\mathbf{s}_j(t_{i-1})$ and $\mathbf{s}_j(t_i)$, and $\mathbf{I}$ is the 2×2 identity matrix. The statistical model in (15) converges to the stochastic differential equation (SDE)

$$\mathbf{s}_j(t) = -\nabla p(\mathbf{s}_j(t), \boldsymbol{\beta}_j)dt + \sigma_j d\mathbf{b}_j(t), \tag{16}$$

as $dt \to 0$, where $d\mathbf{b}_j(t)$ is bivariate Gaussian white noise.

In the data model (15), the parameters σ_j^2 relate to the speed of the migrating individuals and will vary around a group-level speed. However, due to the positive support of the

variance components σ_j^2, we chose to model the individual-level process relating to migration speed in the transformation $\log(\sigma_j)$, so that the support is unbounded and can be suitably modeled with a Gaussian distribution. Otherwise, to create Gibbs updates for σ_j^2 directly in a single-stage algorithm, we would need to specify a conjugate inverse gamma process model on σ_j^2, and specifying hyperpriors on the associated shape and scale parameters would be neither trivial nor biologically intuitive. Thus, we specified a process model for $\log(\sigma_j)$ instead of σ_j^2, implying the transformation function $\sigma_j^2 = \mathbf{g}(\log(\sigma_j)) = e^{2\log(\sigma_j)}$.

In our example, we expected migration to occur primarily in a single direction and specified $\mathbf{x}(\mathbf{s}) = s_2$ where the second component of position $\mathbf{s}$ corresponds to latitude and the coefficient vector is comprised of a single parameter β. Thus, the negative gradient of the potential function in (15) simplifies to $-\nabla p(\mathbf{s}_j(t), \boldsymbol{\beta}_j) = -(0, \beta_j)'$. However, this simplification is based on the assumption that all individuals will migrate in a north/south orientation. To allow for individual variation in the bearing, we multiplied the potential function in (15) by the rotation matrix

$$\mathbf{M} \equiv \begin{pmatrix} \cos(\phi_j) & -\sin(\phi_j) \\ \sin(\phi_j) & \cos(\phi_j) \end{pmatrix}, \tag{17}$$

where ϕ_j is the angle from south of a migratory path, resulting in the data model

$$\mathbf{s}_j(t_i) \sim \mathrm{N}(\mathbf{s}_j(t_{i-1}) - \beta_j \begin{pmatrix} \sin(\phi_j) \\ \cos(\phi_j) \end{pmatrix} dt_i, \sigma_j^2 dt_i \mathbf{I}), \tag{18}$$

Assuming that the variability in β_j and $\log(\sigma_j)$ across individuals can be accounted for as Gaussian random effects and that individual variability in ϕ_j does not arise from an underlying group-level distribution, we have $\beta_j \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2)$, $\log(\sigma_j) \sim \mathrm{N}(\mu_\sigma, \sigma_\sigma^2)$, and $\phi_j \sim \mathrm{Unif}(0, \pi)$, where population-level means μ_β and μ_σ are modeled with Gaussian priors and σ_β^2 and σ_σ^2 arise from inverse gamma priors (full model in Supporting Information, Appendix B).

Two-stage implementation

We fit our model to a subset of the stork migration data (approximately two observations per day per individual) using TARB. In the first stage, we specified individual-level models using the temporary prior $[\beta_j, \sigma_j^2] = [\beta_j][\sigma_j^2]$ where $\beta_j \sim [\beta_j] \equiv \mathrm{N}(\mu_0, \sigma_0^2)$ and $\sigma_j^2 \sim [\sigma_j^2] \equiv \mathrm{IG}(q_0, r_0)$ for $j = 1, ..., J$. Thus, in the first stage, we sample from the posterior distribution

$$[\beta_j, \sigma_j^2, \phi_j | \mathbf{S}_j] \propto \prod_{i=2}^{n_j} [\mathbf{s}_j(t_i) | \beta_j, \sigma_j^2, \phi_j][\beta_j][\sigma_j^2][\phi_j], \tag{19}$$

for each individual $j = 1, ..., J$. We sampled sequentially from the conjugate full-conditional distributions $[\beta_j|\cdot]$ and $[\sigma_j^2|\cdot]$ using Gibbs updates and from $[\phi_j|\cdot]$ using a MH update in an MCMC algorithm in R (version 3.6.1) that we parallelized over individuals with the 'parallel' package (R Core Team 2019).

To use samples from the first-stage models as proposals in the second-stage algorithm, we calculated the Jacobian determinant in (14). Letting $\boldsymbol{\theta}_j \equiv (\beta_j, \log(\sigma_j))'$, and the 2×1 vector transformation $\mathbf{g}(\boldsymbol{\theta}_j)$ be comprised of components $g_1(\boldsymbol{\theta}_j) = \beta_j$ and $g_2(\boldsymbol{\theta}_j) = e^{2\log(\sigma_j)}$, we calculated the Jacobian

$$\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j)) \equiv \begin{bmatrix} \frac{\delta g_1(\boldsymbol{\theta}_j)}{\delta \beta_j} & \frac{\delta g_1(\boldsymbol{\theta}_j)}{\delta \log(\sigma_j)} \\ \frac{\delta g_2(\boldsymbol{\theta}_j)}{\delta \beta_j} & \frac{\delta g_2(\boldsymbol{\theta}_j)}{\delta \log(\sigma_j)} \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 2\sigma_j^2 \end{bmatrix}, \tag{20}$$

which has the determinant $|\mathbf{J}(\mathbf{g}(\boldsymbol{\theta}_j))| = 2\sigma_j^2$. Thus, the second-stage MH ratio from (14) to update β_j, $\log(\sigma_j)$, and ψ_j for individual j is

$$r_j^{(k)} = \frac{[\beta_j^{(*)} | \mu_\beta^{(k-1)}, \sigma_\beta^{2(k-1)}][\log(\sigma_j^{(*)}) | \mu_\sigma^{(k-1)}, \sigma_\sigma^{2(k-1)}][\beta_j^{(k-1)}][\sigma_j^{2(k-1)}][\phi_j^{(*)}] \times \sigma_j^{2(k-1)}}{[\beta_j^{(k-1)} | \mu_\beta^{(k-1)}, \sigma_\beta^{2(k-1)}][\log(\sigma_j^{(k-1)}) | \mu_\sigma^{(k-1)}, \sigma_\sigma^{2(k-1)}][\beta_j^{(*)}][\sigma_j^{2(*)}][\phi_j^{(k-1)}] \times \sigma_j^{2(*)}}. \tag{21}$$

The scalar multiple of 2 from the Jacobian determinant cancels in the numerator and denominator of (21). In the second-stage algorithm, we used the MH ratio in (21) to accept our proposals

for $\beta_j^{(*)}$, $\log(\sigma_j^{(*)})$, and $\phi_j^{(*)}$ which we sampled jointly at random (with replacement) from our first-stage MCMC sample. Then, we sampled the group-level model parameters $(\mu_\beta, \sigma_\beta^2, \mu_\sigma, \text{ and } \sigma_\sigma^2)$ sequentially from their full-conditional distributions using Gibbs updates (Appendix B).

Alternatively, it is possible to fit the full hierarchical model using a standard MCMC algorithm with Gibbs updates for $\beta_j, \mu_\beta, \sigma_\beta^2, \mu_\sigma, \text{ and } \sigma_\sigma^2$. However, we would need to use MH updates for $\log(\sigma_j)$ and ϕ_j, and in cases where the number of individuals J is large, we may have to tune a prohibitively large number of proposal distributions to yield optimal acceptance rates in the MCMC algorithm. Nonetheless, to demonstrate that we obtain the same inference with TARB as compared to a single MCMC algorithm, we also fit the full model with a single algorithm, updating β_j and $\log(\sigma_j)$ sequentially for each individual with Gibbs and MH updates, respectively, and the remaining model parameters as above.

Results

We fit our movement model to a subset of $n = 1675$ stork telemetry observations across $J = 15$ individuals using TARB with $K = 100{,}000$ MCMC iterations for each stage, computing the first stage in parallel over 8 cores, and using a single hierarchical MCMC algorithm with $K = 100{,}000$ MCMC iterations. The recursive approach required 2.95 minutes and the single algorithm required 9.87 minutes; thus computation was over three times faster using TARB. With a larger data set of $n = 155{,}161$ locations for 15 individuals and $K = 60{,}000$ MCMC iterations, computation time to fit the model recursively, in parallel over 15 cores, was 49 minutes, compared to 88 minutes to fit the model as a single algorithm.

Both computational approaches resulted in the same 95% credible intervals and posterior means for β_j and $\log(\sigma_j)$ and the same population-level means μ_β and μ_σ (Fig. 2). The stage-two posterior credible intervals for the β_j and $\log(\sigma_j)$ for each individual j indicate individual variation in speed and directional persistence of migration, but the population is centered around μ_β and μ_σ. First-stage credible intervals are included only to visualize

the relationship between stage one and stage two in Figure 2., and are not used for inference. The shrinkage in interval width between the first- and second-stage posteriors of β_j and $\log(\sigma_j)$ indicates individual-level inference was informed by group-level parameters in the second stage, although this effect was relatively minor in this example. Further, fitting the model to simulated data shows that both computational approaches do equally well recovering 'true' simulated parameters (Appendix C).

Discussion

In our application, we illustrated how TARB can be used to efficiently fit a hierarchical animal movement model to telemetry data, but TARB could be implemented in many ecological models to improve computational efficiency. In Table 1, we highlight several studies from the ecological literature in which the authors used a Bayesian hierarchical model (or desired to, barring computational limitations, as in Breed et al. 2009) that could be fit with TARB. To demonstrate the application of TARB to existing ecological models, we discuss two examples in detail, outlining how the models can be specified in the two-stage framework for faster computation.

Harbor Seal Counts

Cressie et al. (2009) specified a Bayesian hierarchical model to explicitly account for uncertainty at the data and process levels while estimating abundance of harbor seals (*Phoca vitulina*) from census data (Ver Hoef and Frost 2003) in Prince William Sound

$$y_{ij} \sim \text{Pois}(\lambda_{ij}), \tag{22}$$

$$\log(\lambda_{ij}) \sim \text{N}(\mu_{ij}, \sigma^2_{ij}), \tag{23}$$

$$\mu_{ij} = \theta_{0,j} + \mathbf{x}'_{ij}\boldsymbol{\theta}_j, \tag{24}$$

$$\boldsymbol{\theta}_j \sim \text{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}), \tag{25}$$

where y_{ij} is the number of hauled-out seals counted from photographs during each aerial survey i conducted at site j. In the observation model (22), counts arise from a Poisson distribution with intensity parameter λ_{ij} that represents the expected number of haul-outs in a given survey and location. The expected number of haul-outs (λ_{ij}) arises from a normal distribution with mean μ_{ij} that is a function of covariates $\mathbf{x}_{ij}$ with variance parameters σ^2_{ij} for each survey and location. Site-level coefficients $\boldsymbol{\theta}_j$ arise from a population-level multivariate Gaussian distribution, where $\boldsymbol{\Sigma}$ is a diagonal matrix with population-level variance parameters along the diagonal. Thus, the hierarchical model in (22)-(25) is a special case of a generalized linear mixed model.

Surveys were conducted several times per year at each site. Thus, in the first stage of the TARB framework, counts could be modeled independently for each site with the model

$$y_{ij} \sim \text{Pois}(\lambda_{ij}), \tag{26}$$

$$\lambda_{ij} \sim \text{Gamma}(\alpha, \beta), \tag{27}$$

where a temporary gamma prior on λ_{ij} is conjugate with the data model (22) in the first stage so that the MCMC algorithm is unsupervised and could be parallelized over the sites. To complete model fitting in stage two, log-transformed first-stage samples for λ_{ij} would be used as proposals in the MH update for $\log(\lambda_{ij})$ in a second-stage algorithm,

$$[\log(\lambda_{ij})|.] = \frac{\left[\log(\lambda_{ij}^{(*)})|\mu_{ij}, \sigma^2_{ij}\right] \left[\lambda_{ij}^{(k-1)}|\alpha, \beta\right] \lambda_{ij}^{(k-1)}}{\left[\log(\lambda_{ij}^{(k-1)})|\mu_{ij}, \sigma^2_{ij}\right] \left[\lambda_{ij}^{(*)}|\alpha, \beta\right] \lambda_{ij}^{(*)}} \tag{28}$$

where $\frac{d}{d\log(\lambda_{ij})} e^{\log(\lambda_{ij})} = \lambda_{ij}$. All other parameters in the second stage would be updated in the same manner as in a conventional algorithm.

Host Plant Genetics

Evans et al. (2012) conducted a common garden experiment to determine the effects of cottonwood host (*Populus* spp.) genotype on the abundance of herbivorous mite (*Aceria parapopuli*) galls on trees. In our notation, their model was

$$y_{imt} \sim \text{Pois}(\theta_{imt}), \tag{29}$$

$$\log(\theta_{imt}) \sim \text{N}(\mu_{imt}, \sigma^2), \tag{30}$$

$$\mu_{imt} = \beta_i + \mathbf{x}'_{tm}\boldsymbol{\alpha}, \tag{31}$$

$$\boldsymbol{\alpha} \sim \text{N}(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha), \tag{32}$$

$$\beta_i \sim \text{N}(0, \tau^2), \tag{33}$$

$$\tau^2 \sim \text{IG}(\text{a}_\tau, \text{b}_\tau), \tag{34}$$

$$\sigma^2 \sim \text{IG}(\text{a}_\sigma, \text{b}_\sigma), \tag{35}$$

where y_{imt} is the number of galls on tree i with genotype m in year t. The intensity parameter θ_{imt} is a log-linear function of fixed effects $\boldsymbol{\alpha}$ for year and genotype and random effect of tree, β_i. Modifying the process model to

$$\theta_{imt} \sim \text{Gamma}(\gamma_1, \gamma_2), \tag{36}$$

and using temporary priors on γ_1 and γ_2 results in an unsupervised first-stage algorithm. We make a similar adjustment to the second-stage MH ratio as in (28) for recursive computation.

Conclusion

Transformation-assisted RB is one of the most accessible approaches for fitting ecological models recursively with improved computational efficiency and ease. Transformation allows us to extend the benefits of RB to more model specifications, and the demonstrated approach with change of variables can be implemented for most continuous random variables. The

ability to incorporate prior information into analyses is a well-known feature of Bayesian analysis, but it can be difficult to determine how to do so in a robust way, and TARB is a natural approach for using posterior estimates from a previous study as prior information in subsequent studies. Finally, TARB leverages the parallel computing capacity of modern multi-core computers (Visser et al. 2015) to reduce the computational bottleneck created by large data sets and conventional sampling techniques.

Decreased computation time is a major advantage of fitting hierarchical models using TARB, but reducing tuning and partitioning the data in the first stage are equally, if not more, advantageous. This is especially true for large hierarchical models where one might otherwise have to individually tune dozens or hundreds of individual-level parameters to achieve convergence, which would require repeatedly fitting the model. Further, because the first-stage algorithm is used to fit data partitions independently and the second-stage algorithm does not rely on the data directly, we expect additional computational gains. Finally, by design, TARB accommodates uneven sample sizes of partitions, because the first-stage posterior distributions will reflect the uncertainty associated with different sample sizes, thus implicitly weighting the partitions according to sample size in the second stage.

In many cases, the first-stage algorithms of RB and TARB approaches could be implemented in an existing package like JAGS, Stan, or NIMBLE (Plummer 2003, Stan Development Team 2018, NIMBLE Development Team 2019), but the second-stage algorithm cannot be easily implemented in this software. However, using TARB, it may be possible to fit models that are not feasible using these software packages at all. While automated software is convenient and well-suited to a wide range of models, it cannot accommodate all model specifications and users do not always have control over tuning. Although software packages can often fit large models quickly, this may be achieved via computation in C++ rather than R (e.g., Stan, Stan Development Team 2018) or by making approximate inference (e.g., INLA, Rue et al. 2009). Recursive techniques like TARB can also be implemented in C++ via R and `rcpp` for greater computational efficiency, and the results can be used to

obtain both marginal and joint inference.

While TARB can be implemented for a broad range of hierarchical models, there are some cases for which TARB, as presented here using the Jacobian to perform a change of variables, is not ideal for model fitting. For example, hierarchical models that have common parameters at the data level, in addition to partition-level parameters, such as GLMMs with both fixed and random effects, are not easily implemented using TARB. In this case, prior-proposal RB may be helpful (Hooten et al. 2020). Additionally, the Jacobian approach for computing transformed densities is well-suited for transforming continuous random variables, but alternate approaches must be used for discrete random variables. We demonstrated TARB using this technique because it serves as a good introduction into recursive techniques with transformation. For other random variables or applications, there are many useful generalizations of this approach that could be used to obtain valid transformations.

Hierarchical models are powerful tools for understanding complex ecological systems, but the computational demands of fitting ecologically realistic models can make them impractical or impossible to implement. Recursive Bayesian computing techniques address these computational demands, and partitioning model-fitting into stages is natural in many ecological applications. For example, in adaptive management, RB and TARB would allow managers to fit first-stage individual-, year-, or site-level models as data are collected, and add new partitions to existing results by subsequently updating the second stage. Additionally, because the second-stage algorithm only requires first-stage posterior samples, partitions could represent data collected by different researchers during ongoing projects, and researchers could fit population-wide models without needing to share data (Hooten et al. 2020). Thus, in the current era of big data and complex modeling in ecology, TARB is an approachable technique that reduces the computational limitations on the ecological models ecologists can specify and fit.

Acknowledgements

The authors thank the editor and anonymous reviewers whose insights improved the manuscript, and J. Tipton for early discussions on this work. This research was funded by NSF DMS 1614392, NSF DEB 1927177, and NSF GRF 1840343. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data availability

The R code used in our analyses is available at https://doi.org/10.5281/zenodo.4075393. The white stork data set is available on Movebank (Fiedler et al. 2019, https://doi.org/10.5441/001/1.v1cs4nn0) and in the R package 'moveVis' (Schwalb-Willmann et al. 2020).

Author Contributions

MH and AF designed the modeling methodology, and HM and MH designed the case study and performed the analysis. HM and AF wrote the first version of the manuscript, and all authors contributed to revisions of the manuscript.

References

Beissinger, S. R., K. J. Iknayan, G. Guillera-Arroita, E. F. Zipkin, R. M. Dorazio, J. A. Royle, and M. Kery. 2016. Incorporating imperfect detection into joint models of communities: A response to Warton et al. *Trends in Ecology and Evolution* **31**: 736–737. DOI: 10.1016/j.tree.2016.07.009.

Berliner, L. M. 1996. Hierarchical Bayesian time series models. *Maximum Entropy and Bayesian Methods*. Springer: 15–22.

Borsuk, M. E., D. Higdon, C. A. Stow, and K. H. Reckhow. 2001. A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones. *Ecological Modelling* **143**: 165–181. DOI: 10.1016/S0304-3800(01)00328-3.

Breed, G. A., I. D. Jonsen, R. A. Myers, W. D. Bowen, and M. L. Leonard. 2009. Sex-specific, seasonal foraging tactics of adult grey seals (Halichoerus grypus) revealed by state–space analysis. *Ecology* **90**: 3209–3221. DOI: 10.1890/07-1483.1.

Breininger, D. R., E. D. Stolen, D. J. Breininger, and R. D. Breininger. 2019. Sampling rare and elusive species: Florida east coast diamondback terrapin population abundance. *Ecosphere* **10** (8): e02824. DOI: 10.1002/ecs2.2824.

Brillinger, D. R. 2010. Modeling spatial trajectories. *Handbook of Spatial Statistics*: 463–475.

Burton, A. C., M. K. Sam, C. Balangtaa, and J. S. Brashares. 2012. Hierarchical multi-species modeling of carnivore responses to hunting, habitat and prey in a West African protected area. *PloS One* **7**. DOI: 10.1371/journal.pone.0038007.

Cheng, Y., W. Fiedler, M. Wikelski, and A. Flack. 2019. "Closer-to-home" strategy benefits juvenile survival in a long-distance migratory bird. *Ecology and Evolution* **9**: 8945–8952. DOI: 10.1002/ece3.5395.

Cleasby, I. R., T. W. Bodey, F. Vigfusdottir, J. L. McDonald, G. McElwaine, K. Mackie, K. Colhoun, and S. Bearhop. 2017. Climatic conditions produce contrasting influences on demographic traits in a long-distance Arctic migrant. *Journal of Animal Ecology* **86**: 285–295. DOI: 10.1111/1365-2656.12623.

Coll, M., M. Pennino, J. Steenbeek, J. Sole, and J. Bellido. 2019. Predicting marine species distributions: Complementarity of food-web and Bayesian hierarchical modelling approaches. *Ecological Modelling* **405**: 86–101. DOI: 10.1016/j.ecolmodel.2019.05.005.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical

statistical modeling. *Ecological Applications* **19**: 553–570. URL: https://doi.org/10.1890/07-0744.1.

Dietze, M. C., M. S. Wolosin, and J. S. Clark. 2008. Capturing diversity and interspecific variability in allometries: A hierarchical approach. *Forest Ecology and Management* **256**: 1939–1948. DOI: 10.1016/j.foreco.2008.07.034.

Eacker, D. R., P. M. Lukacs, K. M. Proffitt, and M. Hebblewhite. 2017. Assessing the importance of demographic parameters for population dynamics using Bayesian integrated population modeling. *Ecological Applications* **27**: 1280–1293. DOI: 10.1002/eap.1521.

Eckert, S. A., J. E. Moore, D. C. Dunn, R. S. van Buiten, K. L. Eckert, and P. N. Halpin. 2008. Modeling loggerhead turtle movement in the Mediterranean: Importance of body size and oceanography. *Ecological Applications* **18**: 290–308. DOI: 10.1890/06-2107.1.

Evans, L. M., J. S. Clark, A. V. Whipple, and T. G. Whitham. 2012. The relative influences of host plant genotype and yearly abiotic variability in determining herbivore abundance. *Oecologia* **168**: 483–489. DOI: 10.1007/s00442-011-2108-8.

Fiedler, W., A. Flack, W. Schäfle, B. Keeves, M. Quetting, B. Eid, H. Schmid, and M. Wikelski. 2019. Data from: Study "LifeTrack White Stork SW Germany"(2013-2019). DOI: 10.5441/001/1.v1cs4nn0.

Gelfand, A. E. and A. F. Smith. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409. DOI: 10.2307/2289776.

Gerber, B. D., M. B. Hooten, C. P. Peck, M. B. Rice, J. H. Gammonley, A. D. Apa, and A. J. Davis. 2018. Accounting for location uncertainty in azimuthal telemetry data improves ecological inference. *Movement Ecology* **6**: 14. DOI: 10.1186/s40462-018-0129-1.

Geweke, J. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*: 1317–1339. DOI: 10.2307/1913710.

Hampton, S. E., M. B. Jones, L. A. Wasser, M. P. Schildhauer, S. R. Supp, J. Brun, R. R. Hernandez, C. Boettiger, S. L. Collins, L. J. Gross, et al. 2017. Skills and knowledge for data-intensive environmental research. *BioScience* **67**: 546–557. DOI: 10.1093/biosci/bix025.

Hanks, E. M., M. B. Hooten, and F. A. Baker. 2011. Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological Applications* **21**: 1173–1188. DOI: 10.1890/09-1549.1.

Hobbs, N. T. and M. B. Hooten. 2015. Bayesian Models: A Statistical Primer for Ecologists. Princeton University Press.

Hooten, M. B., F. E. Buderman, B. M. Brost, E. M. Hanks, and J. S. Ivan. 2016. Hierarchical animal movement models for population-level inference. *Environmetrics* **27**: 322–333. DOI: 10.1002/env.2402.

Hooten, M. B. and T. J. Hefley. 2019. Bringing Bayesian Models to Life. CRC Press.

Hooten, M. B., D. S. Johnson, and B. M. Brost. 2020. Making recursive Bayesian inference accessible. *The American Statistician*: 1–10. DOI: 10.1080/00031305.2019.1665584.

Hooten, M. B., D. S. Johnson, B. T. McClintock, and J. M. Morales. 2017. Animal Movement: Statistical Models for Telemetry Data. CRC press.

Iijima, H. and C. Otsu. 2018. The method of conserving herbaceous grassland specialists through silvicultural activities under deer browsing pressure. *Biodiversity and Conservation* **27**: 2919–2930. DOI: 10.1007/s10531-018-1577-z.

Jonsen, I. D., R. A. Myers, and M. C. James. 2006. Robust hierarchical state–space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology* **75**: 1046–1057. DOI: 10.1111/j.1365-2656.2006.01129.x.

Kuhnert, P. M., T. G. Martin, K. Mengersen, and H. P. Possingham. 2005. Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. *Environmetrics* **16**: 717–747. DOI: 10.1002/env.732.

Levin, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* **73**: 1943–1967. DOI: 10.2307/1941447.

Lunn, D., J. Barrett, M. Sweeting, and S. Thompson. 2013. Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society: Series C* **62**: 551–572. DOI: 10.1111/rssc.12007.

McCallen, E., J. Knott, G. Nunez-Mir, B. Taylor, I. Jo, and S. Fei. 2019. Trends in ecology: shifts in ecological research themes over the past four decades. *Frontiers in Ecology and the Environment* **17**: 109–116. DOI: 10.1002/fee.1993.

McCaslin, H. M., A. B. Feuka, and M. B. Hooten. 2020. Release for Hierarchical computing for hierarchical models in ecology. *Zenodo*. DOI: 10.5281/zenodo.4075393.

McClintock, B. T., D. J. Russell, J. Matthiopoulos, and R. King. 2013. Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. *Ecology* **94**: 838–849. DOI: 10.1890/12-0954.1.

Monroe, A. P., C. L. Aldridge, T. J. Assal, K. E. Veblen, D. A. Pyke, and M. L. Casazza. 2017. Patterns in greater sage-grouse population dynamics correspond with public grazing records at broad scales. *Ecological Applications* **27**: 1096–1107. DOI: 10.1002/eap.1512.

Moore, J. E. and J. Barlow. 2011. Bayesian state-space model of fin whale abundance trends from a 1991–2008 time series of line-transect surveys in the California Current. *Journal of Applied Ecology* **48**: 1195–1205. DOI: 10.1111/j.1365-2664.2011.02018.x.

Muff, S., J. Signer, and J. Fieberg. 2019. Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology* **89**: 80–92. DOI: 10.1111/1365-2656.13087.

NIMBLE Development Team. 2019. NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling. Version 0.9.0. R package version 0.9.0. DOI: 10.5281/zenodo.1211190. URL: https://cran.r-project.org/package=nimble.

Pearson, D. E., Ö. Eren, Y. K. Ortega, D. Villarreal, M. Şentürk, M. F. Miguel, C. M. Weinzettel, A. Prina, and J. L. Hierro. 2018. Are exotic plants more abundant in the introduced versus native range? *Journal of Ecology* **106**: 727–736. DOI: 10.1111/1365-2745.12881.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Raiho, A. M., M. B. Hooten, S. Bates, and N. T. Hobbs. 2015. Forecasting the effects of fertility control on overabundant ungulates: White-tailed deer in the National Capital Region. *PLoS One* **10**. DOI: 10.1371/journal.pone.0143122.

Royle, J. A. and R. M. Dorazio. 2008. Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities. Elsevier.

Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B* **71**: 319–392.

Särkkä, S. 2013. Bayesian filtering and smoothing. Vol. 3. Cambridge University Press.

Schaub, M. and F. Abadi. 2011. Integrated population models: A novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology* **152**: 227–237. DOI: 10.1007/s10336-010-0632-7.

Schaub, M., H. Jakober, and W. Stauber. 2013. Strong contribution of immigration to local population regulation: Evidence from a migratory passerine. *Ecology* **94**: 1828–1838. DOI: 10.1890/12-1395.1.

Schwalb-Willmann, J., R. Remelgado, K. Safi, and M. Wegmann. 2020. moveVis: Animating movement trajectories in synchronicity with static or temporally dynamic environmental data in R. *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210X.13374.

Shelton, A. O., J. L. O'Donnell, J. F. Samhouri, N. Lowell, G. D. Williams, and R. P. Kelly. 2016. A framework for inferring biological communities from environmental DNA. *Ecological Applications* **26**: 1645–1659. DOI: 10.1890/15-1733.1.

Stan Development Team. 2018. RStan: the R interface to Stan. R package version 2.17.3. URL: http://mc-stan.org/.

Ver Hoef, J. and K. Frost. 2003. A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. *Environmental and Ecological Statistics* **10**: 201–209. DOI: 10.1023/A:1023626308538.

Vieilledent, G., B. Courbaud, G. Kunstler, J.-F. Dhôte, and J. S. Clark. 2010. Individual variability in tree allometry determines light resource allocation in forest ecosystems: a hierarchical Bayesian approach. *Oecologia* **163**: 759–773. DOI: 10.1007/s00442-010-1581-9.

Visser, M. D., S. M. McMahon, C. Merow, P. M. Dixon, S. Record, and E. Jongejans. 2015. Speeding up ecological and evolutionary computations in R: Essentials of high performance computing for biologists. *PLoS Computational Biology* **11**. DOI: 10.1371/journal.pcbi.1004140.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. Hui. 2015. So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution* **30**: 766–779. DOI: 10.1016/j.tree.2015.09.007.

Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* **3**: 385–397. DOI: 10.2307/2389612.

Table 1: Examples of ecological studies with Bayesian hierarchical models that could be implemented in a transformation-assisted recursive Bayesian framework.

Discipline	Study
Fish & Wildlife Ecology	Burton et al. 2012
	Cressie et al. 2009
	Breininger et al. 2019
	Kuhnert et al. 2005
	Monroe et al. 2017
	Moore and Barlow 2011
Integrated Population Models	Cleasby et al. 2017
	Eacker et al. 2017
	Raiho et al. 2015
	Schaub et al. 2013
Animal Movement	Breed et al. 2009
	Eckert et al. 2008
	Jonsen et al. 2006
	McClintock et al. 2013
	Muff et al. 2019
Forestry & Plant Ecology	Dietze et al. 2008
	Evans et al. 2012
	Hanks et al. 2011
	Iijima and Otsu 2018
	Vieilledent et al. 2010
Ecosystem Ecology	Borsuk et al. 2001
	Coll et al. 2019
	Shelton et al. 2016

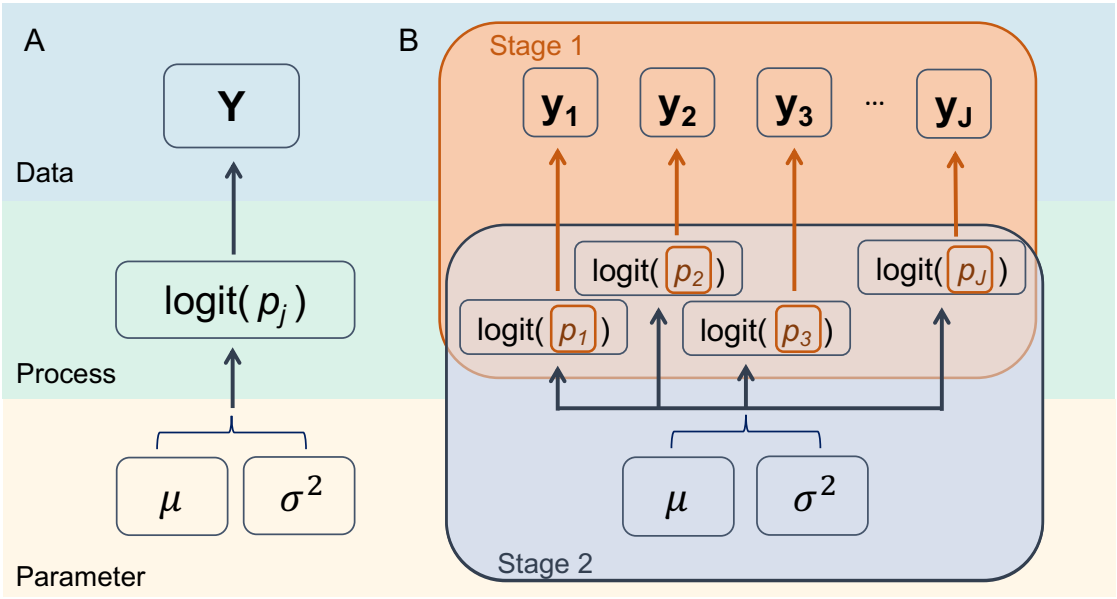

Figure 1: (A) Directed acyclic diagram (DAG) for Bernoulli GLMM of cheatgrass occurrence in Montana (1)-(4) and (B) schematic for partitioning DAG according to the TARB framework. In (A), $\mathbf{Y}$ is the matrix whose columns are the data vectors $\mathbf{y}_j$ for the sites $j = 1, ..., J$. In stage 1, the data $\mathbf{Y}$ are partitioned by site and fit to obtain the posterior distributions for the p_j. In stage 2, samples from these posterior distributions are used to sample $\text{logit}(p_j)$, μ, and σ^2.

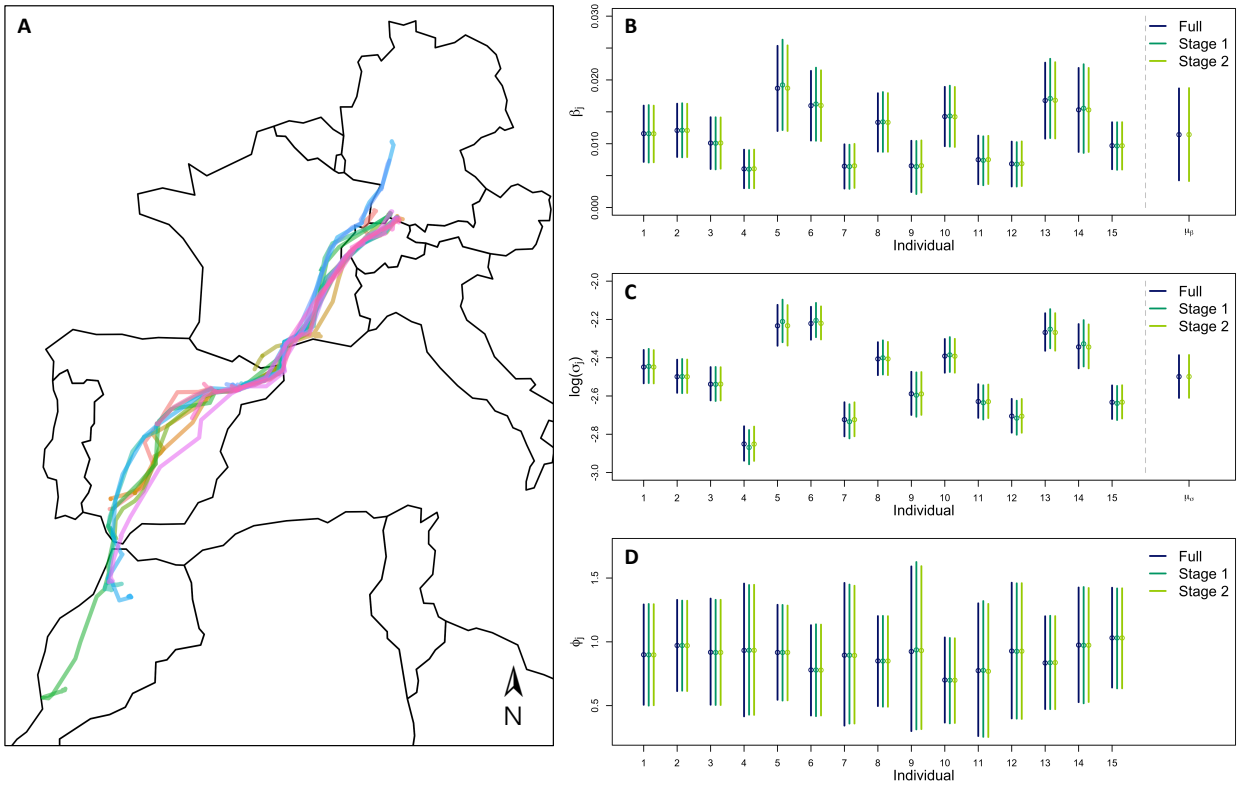

Figure 2: (A) Migratory trajectories for $J = 15$ white storks tracked via GPS loggers in fall 2018, with each individual represented by a different color, and (B)-(D) posterior means (points) and 95% credible intervals for model parameters resulting from fitting our hierarchical movement model to $n = 1675$ telemetry locations from $J = 15$ white storks as a single hierarchical algorithm and in two stages using TARB. It is important to note here that we show the posterior distributions for the first-stage estimates to illustrate how some individual-level parameters borrow strength from the group-level parameters in stage 2, but in practice, the first stage posterior estimates would not be used to make inference.

Supporting Information: McCaslin, H.M., A.B. Feuka, and M.B. Hooten. 2020. Hierarchical computing for hierarchical models in ecology. *Methods in Ecology and Evolution.*

Appendix A: Transformation-Assisted Recursive Bayesian Computing Tutorial with Bernoulli GLMM

We demonstrate the implementation of transformation-assisted recursive Bayesian computing (TARB) with the following example. Pearson et al. (2018) collected occurrence data for many species of invasive grasses at 20 grassland sites throughout Montana, including cheatgrass (*Bromus tectorum*). Suppose we want to model the probability of cheatgrass occurrence in Montana grasslands using a Bernoulli generalized linear mixed model (GLMM) with a single random effect and no covariates. At each site, cheatgrass occurrence was recorded at 20 randomly selected 1-m^2 plots, and we wish to fit the following model to the data,

$$\mathrm{y}_{ij} \sim \mathrm{Bern}(p_j), \qquad i = 1, ..., 20, j = 1, ..., 20, \tag{S1}$$

$$\mathrm{logit}(p_j) \sim \mathrm{N}(\mu, \sigma^2), \tag{S2}$$

$$\mu \sim \mathrm{N}(\mu_0, \sigma_0^2), \tag{S3}$$

$$\sigma^2 \sim \mathrm{IG}(q, r), \tag{S4}$$

where j indexes sites and i indexes plots within each site. In this model, p_j is the probability of cheatgrass at site j, and $\mathrm{logit}(p_j)$ arises from a Gaussian distribution with study-wide parameters μ and σ^2 (Fig. 1). Thus, p_j are "random effects" because they will vary for each site but will arise from a single underlying distribution. We use Gaussian random effects, with the logit link function to constrain p_j to the proper support, and seek inference on μ. We specify Gaussian and inverse gamma priors on μ and σ^2 respectively. The joint posterior

distribution for this model is

$$[\text{logit}(\boldsymbol{p}), \mu, \sigma^2|\mathbf{Y}] \propto \left(\prod_{j=1}^{J}\left(\prod_{i=1}^{N}[\text{y}_{ij}|\text{logit}(p_j)]\right)[\text{logit}(p_j)|\mu,\sigma^2]\right)[\mu][\sigma^2]. \tag{S5}$$

The full-conditional distributions for μ and σ^2 are

$$[\mu|\cdot] \propto \text{N}\left(\left(\frac{J}{\sigma^2}+\frac{1}{\sigma_0^2}\right)^{-1}\left(\frac{\sum_{j=1}^{J}\text{logit}(p_j)}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right), \left(\frac{J}{\sigma^2}+\frac{1}{\sigma_0^2}\right)^{-1}\right), \tag{S6}$$

$$[\sigma^2|\cdot] \propto \text{IG}\left(\frac{J}{2}+q, \left(\frac{\sum_{j=1}^{J}(\text{logit}(p_j)-\mu)^2}{2}+\frac{1}{r}\right)^{-1}\right), \tag{S7}$$

but the full-conditional distributions for the $\text{logit}(p_j)$ are not analytically tractable and therefore cannot be sampled using Gibbs updates in an MCMC algorithm.

We use TARB to fit this model so that we can specify temporary first-stage priors for the p_j that are conjugate with the data model and therefore result in analytically tractable full-conditional distributions to avoid having to tune the updates for the $\text{logit}(p_j)$. Letting $\theta_j = \text{logit}(p_j)$, we specify the transformation $\text{g}(\theta_j) \equiv \text{logit}^{-1}(\theta_j)$ so that

$$\text{g}(\theta_j) = \text{g}(\text{logit}(p_j)) = \text{logit}^{-1}(\text{logit}(p_j)) = p_j. \tag{S8}$$

Next, we specify temporary first-stage Jefferys priors $p_j \sim [p_j] \equiv \text{Beta}(\alpha = 0.5, \beta = 0.5)$. With this prior, the first-stage full-conditional distribution for p_j for a single site j,

$$[p_j|\mathbf{y}_j] \propto \left(\prod_{i=1}^{N}[\mathbf{y}_{ij}|p_j]\right)[p_j], \tag{S9}$$

$$\propto \left(\prod_{i=1}^{N} p_j^{\text{y}_{ij}}(1-p_j)^{1-\text{y}_{ij}}\right)p_j^{\alpha-1}(1-p_j)^{\beta-1}, \tag{S10}$$

$$\propto \text{Beta}\left(\left(\sum_{i=1}^{N}\text{y}_{ij}\right)+\alpha, \left(\sum_{i=1}^{N}1-\text{y}_{ij}\right)+\beta\right), \tag{S11}$$

is analytically tractable. Further, because p_j is the only parameter in the first-stage, this

full-conditional distribution is equivalent to the first-stage posterior distribution for this site. Thus, we can sample from this first-stage posterior distribution without needing to use MCMC at all. Instead, we can use the following R code to generate a sample of size $K =$ 100,000 from the posterior distribution for a single site:

```
s1_post <- rbeta(10000, sum(y)+alpha, sum(1-y)+beta)
```

This results in the first-stage posterior distribution $[p_j|\mathbf{y}_j]$, so to use these posterior samples as MH proposals in the second stage of the TARB procedure, we must modify the stage two MH ratio so that the proposal is on the original variable that appears in the process model, namely $\text{logit}(p_j)$. Because $\theta_j = \text{logit}(p_j)$ is a scalar for a single site j, the Jacobian matrix $\mathbf{J}(\mathbf{g}(\theta_j))$ simplifies to the scalar derivative of g with respect to θ_j,

$$\mathbf{J}(\mathbf{g}(\theta_j)) \equiv \frac{d}{d\theta_j} g(\theta_j). \tag{S12}$$

The derivative $\frac{d}{d\theta}\frac{e^\theta}{e^\theta+1} = \frac{e^\theta}{(e^\theta+1)^2}$ and $e^{\log\frac{p_j}{1-p_j}} = \frac{p_j}{1-p_j}$, so it follows that

$$\frac{d}{d\theta_j} g(\theta_j) = \frac{d}{d\theta_j}\left(\frac{e^{\theta_j}}{e^{\theta_j}+1}\right), \tag{S13}$$

$$= \frac{p_j}{(1-p_j)(\frac{p_j}{1-p_j}+1)^2}, \tag{S14}$$

$$= p_j(1-p_j). \tag{S15}$$

Therefore, the MH ratio to update the $\text{logit}(p_j)$ in the second stage of model fitting is

$$r_j^{(k)} = \frac{[\mathbf{y}_j|\text{logit}(p_j^{(*)})][\text{logit}(p_j^{(*)})|\mu^{(k-1)},\sigma^{2(k-1)}][\text{logit}(p_j^{(k-1)})|\mathbf{y}_j]}{[\mathbf{y}_j|\text{logit}(p_j^{(k-1)})][\text{logit}(p_j^{(k-1)})|\mu^{(k-1)},\sigma^{2(k-1)}][\text{logit}(p_j^{(*)})|\mathbf{y}_j]}, \tag{S16}$$

$$= \frac{[\mathbf{y}_j|\text{logit}(p_j^{(*)})][\text{logit}(p_j^{(*)})|\mu^{(k-1)},\sigma^{2(k-1)}][\mathbf{y}_j|\text{logit}(p_j^{(k-1)})][\text{logit}(p_j^{(k-1)})]}{[\mathbf{y}_j|\text{logit}(p_j^{(k-1)})][\text{logit}(p_j^{(k-1)})|\mu^{(k-1)},\sigma^{2(k-1)}][\mathbf{y}_j|\text{logit}(p_j^{(*)})][\text{logit}(p_j^{(*)})]}, \tag{S17}$$

$$= \frac{[\text{logit}(p_j^{(*)})|\mu^{(k-1)},\sigma^{2(k-1)}][p_j^{(k-1)}] \times p_j^{(k-1)}(1-p_j^{(k-1)})}{[\text{logit}(p_j^{(k-1)})|\mu^{(k-1)},\sigma^{2(k-1)}][p_j^{(*)}] \times p_j^{(*)}(1-p_j^{(*)})}. \tag{S18}$$

The R code for this update for a single site j is:

```
# draw from first stage posterior #
idx <- sample(1:s1_length,1,replace=T)
p.star <- s1_post[idx]
logit.p.star <- logit(p.star)

# MH ratio #
mh1 <- dnorm(logit.p.star,mu,sqrt(s2),log=T) +
        dbeta(p[j],alpha,beta,log=T) +
        log(p[j]*(1-p[j]))
mh2 <- dnorm(logit.p[j],mu,sqrt(s2),log=T) +
        dbeta(p.star,alpha,beta,log=T) +
        log(p.star*(1-p.star))

# accept or reject #
mh <- exp(mh1-mh2)
if(mh > runif(1)){
    p[j] <- p.star
    logit.p[j] <- logit.p.star
    accept[j] <- accept[j] + 1
}
```

To complete model fitting, we update μ and σ^2 by sampling from their full-conditional distributions with Gibbs updates:

```
## sigma update ##
tmp_r <- 1/(sum((logit.p-mu)^2)/2 + 1/r)
tmp_q <- J/2 + q
s2 <- 1/rgamma(1,shape=tmp_q,scale=tmp_r)

## mu update ##
tmp_s2 <- 1/(J/s2 + 1/s2_0)
tmp_mu <- tmp_s2*(sum(logit.p)/s2 + mu_0/s2_0)
mu <- rnorm(1,tmp_mu, sqrt(tmp_s2))
```

We can now make inference from the stage two output, as shown in Figure S1. Note that the output from stage one is only used to sample proposals for the stage two MH ratio, and should not be used to make inference. All code and data to fit this model to the cheatgrass data are attached as supplemental files.

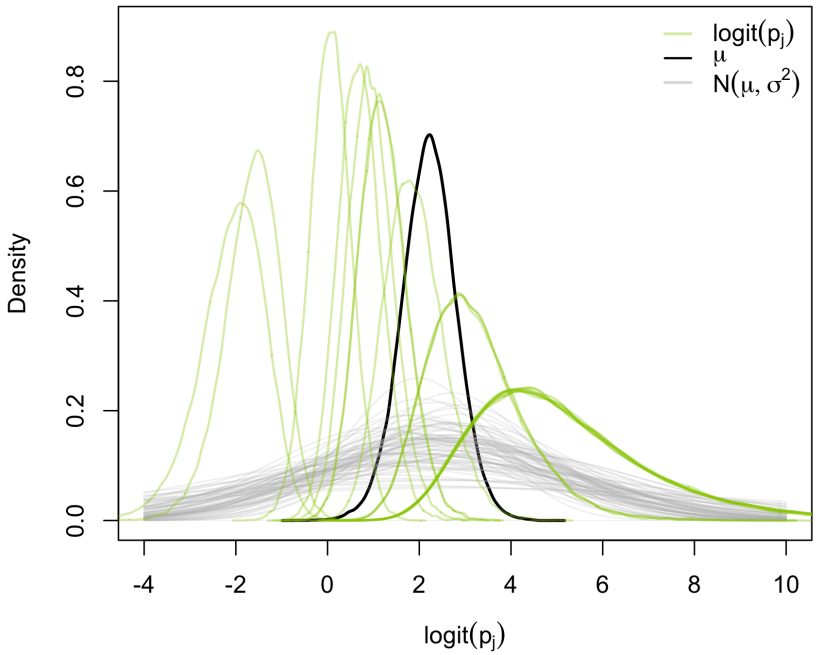

Figure S1: Posterior densities for logit(p_j) at $j = 1, ..., J$ grassland sites in Montana, USA (green), where p_j is the probability of cheatgrass (*Bromus tectorum*) occurrence at site j, and 80 realizations of the process model N(μ, σ^2) giving rise to the logit(p_j) (gray). Also shown is the posterior density for group-level parameter μ (black), describing the central tendency of this distribution giving rise to the logit(p_j).

Appendix B: Full movement model & prior specifications

Our full model for the white stork telemetry data (Cheng et al. 2019, Fiedler et al. 2019) was specified as

$$\mathbf{s}_j(t_i) \sim \mathrm{N}\left((\mathbf{s}_j(t_{i-1}) - \beta_j \begin{pmatrix} \sin(\phi_j) \\ \cos(\phi_j) \end{pmatrix} dt_i, \sigma_j^2 dt_i \mathbf{I}\right), \tag{S19}$$

$$\beta_j \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2), \tag{S20}$$

$$\log(\sigma_j) \sim \mathrm{N}(\mu_\sigma, \sigma_\sigma^2), \tag{S21}$$

$$\phi_j \sim \mathrm{Unif}(u_1, u_2), \tag{S22}$$

$$\mu_\beta \sim \mathrm{N}(\mu_1, \sigma_1^2), \tag{S23}$$

$$\sigma_\beta^2 \sim \mathrm{IG}(q_1, r_1), \tag{S24}$$

$$\mu_\sigma \sim \mathrm{N}(\mu_2, \sigma_2^2), \tag{S25}$$

$$\sigma_\sigma^2 \sim \mathrm{IG}(q_2, r_2). \tag{S26}$$

where β_j and $\log(\sigma_j)$ are Gaussian random effects, but individual variability in ϕ_j does not arise from a population-level distribution. We used MCMC to sample from the joint posterior distribution

$$[\boldsymbol{\beta}, \log(\boldsymbol{\sigma}), \boldsymbol{\phi}, \mu_\beta, \mu_\sigma, \sigma_\sigma^2 | \mathbf{S}] \propto \left(\prod_{j=1}^{J} \left(\prod_{i=2}^{n_j} [\mathbf{s}_j(t_i) | \beta_j, \log(\sigma_j), \phi_j] \right) [\beta_j | \mu_\beta, \sigma_\beta^2] [\log(\sigma_j) | \mu_\sigma, \sigma_\sigma^2] [\phi_j | u_1, u_2] \right) \times [\mu_\beta][\sigma_\beta^2][\mu_\sigma][\sigma_\sigma^2], \tag{S27}$$

in two stages.

In the first stage of the transformation-assisted recursive Bayesian (TARB) procedure to fit the stork migration movement model, we specified individual-level models using the

temporary prior $[\beta_j, \sigma_j^2] = [\beta_j][\sigma_j^2]$ where

$$\beta_j \sim [\beta_j] \equiv \mathrm{N}(\mu_0 = 0, \sigma_0^2 = 10), \tag{S28}$$

$$\sigma_j^2 \sim [\sigma_j^2] \equiv \mathrm{IG}(q_0 = 0.001, r_0 = 1000), \tag{S29}$$

for $j = 1, ..., J$, allowing us to update β_j and σ_j^2 by sampling from the conjugate full-conditional distributions

$$[\beta_j|\cdot] = \mathrm{N}\left(\left(\frac{(n_j-1)dt}{\sigma_j^2} + \frac{1}{\sigma_0^2}\right)^{-1}\left(\frac{-\sum_{i=2}^{n_j}(\mathbf{s}(t_i) - \mathbf{s}(t_{i-1}))'\binom{\sin(\phi_j)}{\cos(\phi_j)}}{\sigma_j^2} + \frac{\mu_0}{\sigma_0^2}\right), \left(\frac{(n_j-1)dt}{\sigma_j^2} + \frac{1}{\sigma_0^2}\right)^{-1}\right), \tag{S30}$$

$$[\sigma_j^2|\cdot] = \mathrm{IG}\left((n_j-1) + q_0, \left(\frac{(\mathbf{s}(t_i) - \mathbf{s}(t_{i-1}) + \beta_2\binom{\sin(\phi_j)}{\cos(\phi_j)}dt)'(\mathbf{s}(t_i) - \mathbf{s}(t_{i-1}) + \beta_2\binom{\sin(\phi_j)}{\cos(\phi_j)}dt)}{2dt} + \frac{1}{r_0}\right)^{-1}\right). \tag{S31}$$

We updated ϕ_j using a Metropolis-Hastings (MH) update with acceptance ratio

$$r_\phi = \frac{[\mathbf{s}_j|\beta^{(k-1)}, \sigma_2^{2(k-1)}, \phi^{(*)}]}{[\mathbf{s}_j|\beta^{(k-1)}, \sigma_2^{2(k-1)}, \phi^{(k-1)}]}, \tag{S32}$$

where the proposal and prior terms cancel in the numerator and denominator because we used the proposal distribution $\phi_j^{(*)} \sim \mathrm{N}(\phi_j^{(k-1)}, \sigma_{tune}^2)$ and a uniform prior $[\phi_j] \equiv \mathrm{U}(0, \pi)$ on ϕ_j.

In the second-stage, we updated β_j, σ_j^2, and ϕ_j using a Metropolis-Hastings update described in the main text, and then sampled the remaining model parameters $(\mu_\beta, \sigma_\beta^2, \mu_\sigma, \sigma_\sigma^2)$

sequentially using Gibbs updates from their conjugate full-conditional distributions

$$[\mu_\beta|\cdot] \propto \mathrm{N}\left(\left(\frac{J}{\sigma_\beta^2}+\frac{1}{\sigma_1^2}\right)^{-1}\left(\frac{\sum_{j=1}^J \beta_j}{\sigma_\beta^2}+\frac{\mu_1}{\sigma_1^2}\right), \left(\frac{J}{\sigma_\beta^2}+\frac{1}{\sigma_1^2}\right)^{-1}\right), \tag{S33}$$

$$[\sigma_\beta^2|\cdot] \propto \mathrm{IG}\left(\frac{J}{2}+q_1, \left(\frac{\sum_{j=1}^J(\beta_j-\mu_\beta)^2}{2}+\frac{1}{r_1}\right)^{-1}\right), \tag{S34}$$

$$[\mu_\sigma|\cdot] \propto \mathrm{N}\left(\left(\frac{J}{\sigma_\sigma^2}+\frac{1}{\sigma_2^2}\right)^{-1}\left(\frac{\sum_{j=1}^J \log(\sigma_j)}{\sigma_\sigma^2}+\frac{\mu_2}{\sigma_2^2}\right), \left(\frac{J}{\sigma_\sigma^2}+\frac{1}{\sigma_2^2}\right)^{-1}\right), \tag{S35}$$

$$[\sigma_\sigma^2|\cdot] \propto \mathrm{IG}\left(\frac{J}{2}+q_2, \left(\frac{\sum_{j=1}^J(\log(\sigma_j)-\mu_\sigma)^2}{2}+\frac{1}{r_2}\right)^{-1}\right), \tag{S36}$$

with the prior distributions

$$\mu_\beta \sim \mathrm{N}(\mu_1 = 0, \sigma_1^2 = 100), \tag{S37}$$

$$\sigma_\beta^2 \sim \mathrm{IG}(q_1 = 0.001, r_1 = 1000), \tag{S38}$$

$$\mu_\sigma \sim \mathrm{N}(\mu_2 = 0, \sigma_2^2 = 100), \tag{S39}$$

which were also the priors used when fitting the model in a single MCMC algorithm.

Appendix C: Analysis of Simulated Data

To demonstrate transformation-assisted recursive Bayesian computing (TARB), we simulated data for 20 individuals arising from the movement process described by the model in (S18)-(S25). We specified 'true' parameter distributions $\beta_j \sim \text{N}(1.2, 0.2)$ and $\log(\sigma_j) \sim \text{N}(-1.6, 0.1)$, and simulated a trajectory of 100 observations $\mathbf{S}_j$ with constant $dt = 0.005$ for each individual j for $j = 1, ..., 20$ (Fig. S2).

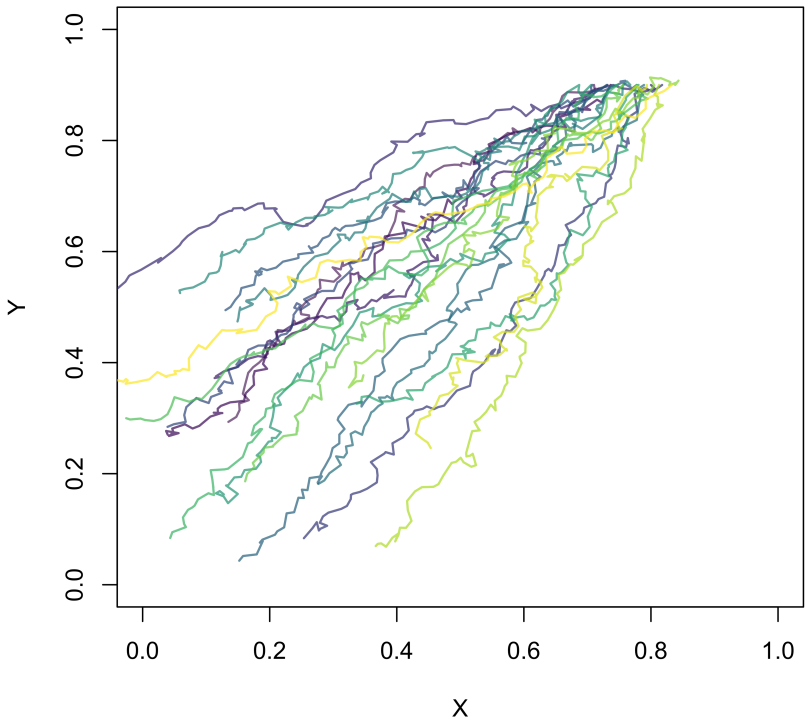

Figure S2: Simulated trajectories for 20 individuals with 100 observations each.

We fit the model to the simulated data both with our two-stage approach, specifying a temporary prior on σ_j^2 in the first stage as we did when modeling the stork data, and in a single MCMC algorithm. We fit the model to simulated data using the following priors for both algorithms.

$$\mu_\beta \sim \mathrm{N}(\mu_1 = 0, \sigma_1^2 = 10), \tag{S40}$$

$$\sigma_\beta^2 \sim \mathrm{IG}(q_1 = 0.001, r_1 = 1000), \tag{S41}$$

$$\mu_\sigma \sim \mathrm{N}(\mu_2 = 0, \sigma_2^2 = 100), \tag{S42}$$

$$\sigma_\sigma^2 \sim \mathrm{IG}(q_2 = 0.001, r_2 = 1000). \tag{S43}$$

Additionally, for the first stage of the recursive Bayes procedure, we specified the following temporary priors for β_j and σ_j^2

$$\beta_j \sim \mathrm{N}(\mu_0 = 0, \sigma_0^2 = 10), \tag{S44}$$

$$\sigma_j^2 \sim \mathrm{IG}(q_0 = 0.01, r_0 = 100). \tag{S45}$$

The two-stage approach resulted in the same inference as the full hierarchical algorithm (Fig. S3). Additionally, both the second-stage algorithm and the single hierarchical algorithm recover the β_j, $\log(\sigma_j)$, and ϕ_j values used to simulate the data for nearly all individuals.

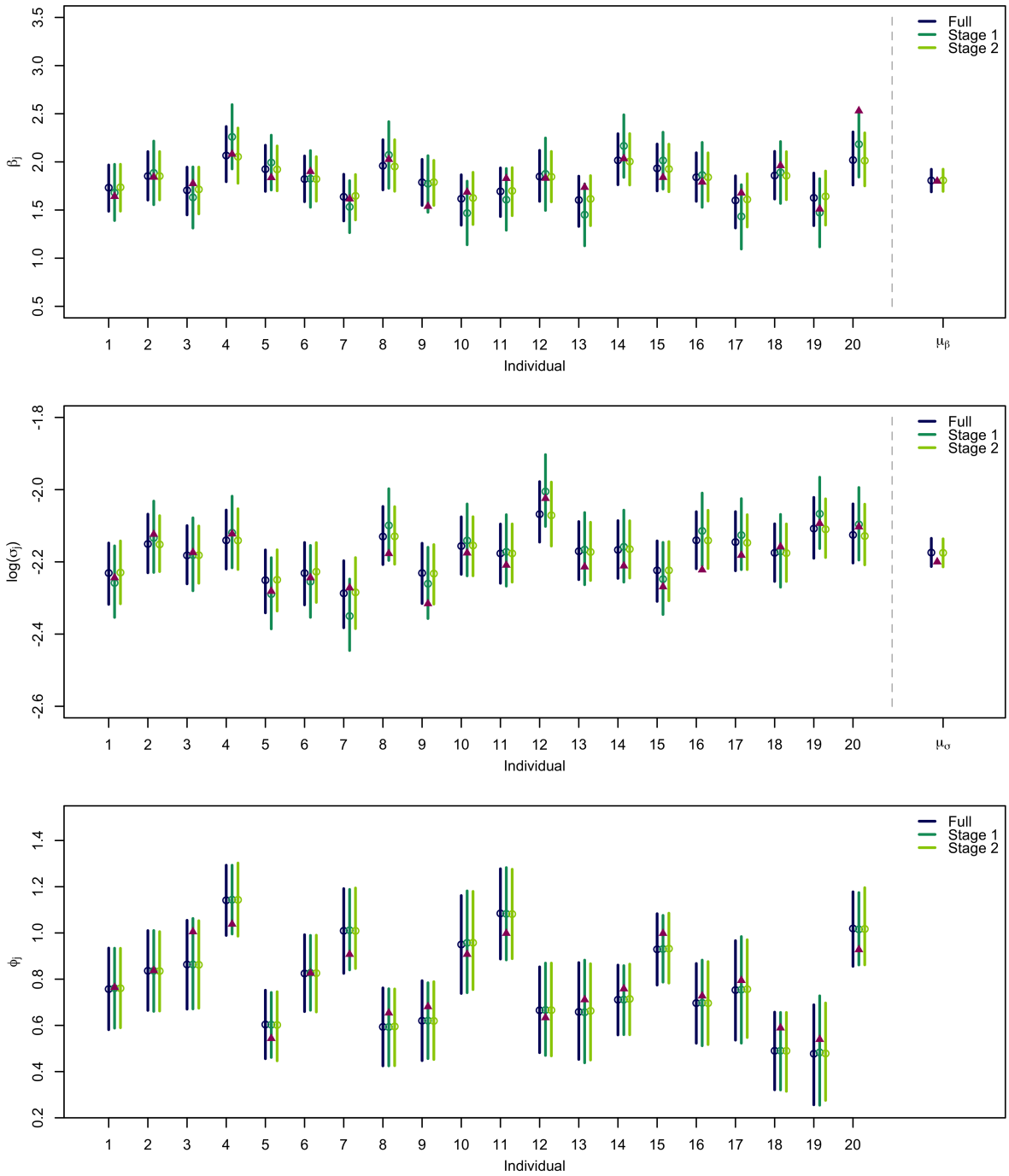

Figure S3: Marginal posterior means (points) and 95% credible intervals (lines) for the individual-level β_j and $\log(\sigma_j)$ and the population-level means μ_β and μ_σ. Triangles represent 'true' parameter values used to simulate data. The purple estimates (left in each group) correspond to results from the single, hierarchical algorithm, the blue estimates (center in each group) correspond to the individual-level results from the first stage of the two-stage approach, and the green estimates (right in each cluster) correspond to the second-stage results. The estimates to the right of the dashed line are estimates for population-level means from the full hierarchical algorithm (purple) and the two-stage approach (green).

References

Cheng, Y., W. Fiedler, M. Wikelski, and A. Flack. 2019. "Closer-to-home" strategy benefits juvenile survival in a long-distance migratory bird. *Ecology and Evolution* **9**: 8945–8952. DOI: 10.1002/ece3.5395.

Fiedler, W., A. Flack, W. Schäfle, B. Keeves, M. Quetting, B. Eid, H. Schmid, and M. Wikelski. 2019. Data from: Study "LifeTrack White Stork SW Germany"(2013-2019). DOI: 10.5441/001/1.v1cs4nn0.

McCaslin, H. M., A. B. Feuka, and M. B. Hooten. 2020. Release for Hierarchical computing for hierarchical models in ecology. *Zenodo*. DOI: 10.5281/zenodo.4075393.

Pearson, D. E., Ö. Eren, Y. K. Ortega, D. Villarreal, M. Şentürk, M. F. Miguel, C. M. Weinzettel, A. Prina, and J. L. Hierro. 2018. Are exotic plants more abundant in the introduced versus native range? *Journal of Ecology* **106**: 727–736. DOI: 10.1111/1365-2745.12881.