Hierarchical Bayesian Models for

Predicting The Spread of Ecological Processes

Christopher K. Wikle[*]

Department of Statistics, University of Missouri

Abstract: There is increasing interest in predicting ecological processes. Methods to accomplish such predictions must account for uncertainties in observation, sampling, models, and parameters. Statistical methods for spatio-temporal processes are powerful, yet difficult to implement in complicated, high-dimensional settings. However, recent advances in hierarchical formulations for such processes can be utilized for ecological prediction. These formulations are able to account for the various sources of uncertainty, and can incorporate scientific judgment in a probabilistically consistent manner. In particular, analytical diffusion models can serve as motivation for the hierarchical model for invasive species. We demonstrate by example that such a framework can be utilized to predict spatially and temporally, the house finch relative population abundance over the eastern United States.

---

[*]Christopher K. Wikle, Department of Statistics, University of Missouri, 222 Math Science Building, Columbia, MO 65211; wikle@stat.missouri.edu

INTRODUCTION

Ecological processes exhibit complicated behavior over an extensive range of spatial and temporal scales of variability. To understand and eventually predict such complicated processes, we must make use of available scientific insight, data, and theory, in a modeling framework that honestly accounts for uncertainties in each. Indeed, with the growing interest in ecological prediction, the development of alternative strategies that can account for various uncertainties is imperative (Clark et al. 2001).

Although there is a rich history in ecology for developing analytical models that can describe the essence of ecological processes over space and time, there has been less attention to fitting such models to data, and even less attention to accounting for the uncertainties present in the data, model, and parameters. The deterministic models that work so well in describing theoretical aspects of the spatio-temporal behavior for many ecological processes are fraught with uncertainty in application to prediction problems with uncertain observational data. These uncertainties result from various approximations and small-scale parameterizations, as well as uncertain initial and boundary conditions. Furthermore, data are imprecise, subject to sampling variations, measurement errors and biases, and concerns regarding the relevance of a particular dataset to the question at hand. In the face of these complications, it is reasonable to view the prediction of complex phenomena as statistical by nature. A consequence of this view is that we should not expect to be able to predict individual organisms at all locations at all times, but rather focus on prediction of the aggregate statistical properties (e.g., distributions) over relatively large spatial and temporal scales (e.g., Levin et al. 1997). By utilizing a statistical approach that accounts for uncertainties in such settings, one is able to obtain predictions (in space and or time) and realistic measures of prediction error.

Traditionally, the generalized linear mixed model provides a flexible stochastic modeling framework that can describe complicated processes (e.g., McCulloch and Searle 2001). However, it is often extremely difficult (if not impossible) to account realistically

for the joint spatio-temporal dependence structure in the random effects term of such models. This is especially true of dynamical processes, in which there are complicated interactions across space and time. The difficulty arises from the fact that our current state of knowledge concerning valid spatio-temporal covariance models for complicated processes is extremely limited. Alternatively, one might model the complicated spatio-temporal dependence by factoring the joint distribution of the ecological process into a series of conditional models. This is the approach utilized, to a limited extent, in state-space modeling of multivariate time-series models (e.g., West and Harrison 1997; Wikle and Cressie 1999). Such models account for observation error at one stage, and then at the next stage, factor the process, typically by assuming that if one knows the distribution of the process at the current time given the previous time, all other earlier times provide no additional information. This is known as a Markov assumption in time. The Markov-based factorization is ideal for modeling dynamical processes since it specifies models for the process at each time given the process in the past and some parameters that describe the dynamical evolution. If the parameters of the model are not known perfectly ("parameter uncertainty"), then one must estimate them; this can be very difficult in large complex systems. However, one can mitigate this difficulty by modeling the process parameters as a further series of conditional models. Such a hierarchical factorization is very powerful in that relatively simple spatial and temporal dependence assigned to sub-processes and parameters can lead to very complicated joint spatio-temporal dependence. These models are known as fully hierarchical spatio-temporal dynamical models (Wikle, Berliner, Cressie, 1998).

The question then is how to incorporate scientific insight into the hierarchical framework. State-space modeling has a long history of including deterministic models explicitly in the dynamical framework. However, the model dynamics are usually considered to be known without error. Although this is often reasonable in the engineering applications for which the methodology was developed, it is seldom the case in ecology that we know the "true" dynamical model for a given process ("model uncertainty"). However,

3

it is possible that inexact but useful scientific theory, such as suggested by deterministic (e.g., partial differential equation, PDE) models, can be incorporated into the conditional framework described above (e.g., Wikle et al. 2001). That is, the deterministic model may reflect our prior scientific understanding of the dynamical propagation (often analytical), but we doubt that the model, by itself, can adequately describe the true process. Thus, we reconfigure the deterministic model into a stochastic framework, formally making it part of our "prior" assumptions about the process; specifically, we use the deterministic model to motivate a series of conditional models. The important point is that within the hierarchical framework, one accounts for the uncertainty associated with this knowledge (the deterministic model), so that the final prediction of the process accounts for uncertainties not only in data, but theory, model specification, and parameters.

The hierarchical spatio-temporal dynamic model methodology is illustrated with a case study concerned with predicting the abundance of the house finch (*Carpodacus mexicanus*) over the eastern half of the U.S. from 1966 through 2001, with data collected during the North American Breeding Bird Survey (BBS; Robbins et al. 1986) from 1966 through 1999. Challenging aspects of this application include the presence of significant observational error (Sauer et al. 1994), irregular spatial and temporal sampling, the diffusive nature of the invasive process on landscape scales (over the eastern U.S.), and the desire to quantify realistically the prediction errors associated with spatial and temporal predictions of relative abundance.

Diffusion models have long been considered for describing the spread of invasive organisms, and they have relevance to many invasive bird species (e.g., Okubo 1986; Veit and Lewis 1996). This paper will describe how the hierarchical Bayesian approach can be motivated by traditional diffusion PDEs and that such a framework can be used to model the invasive expansion of the house finch on landscape scales. Key to this formulation is the specification that the diffusion rate can vary with space. This methodology allows for assessment of prediction uncertainty, both in space and time.

The paper begins with a section describing relevant background concerning the house

finch, diffusion models in ecology, and Bayesian hierarchical modeling. A hierarchical model for house finch abundance over space and time is then developed, followed by the results from application to the BBS data. This is followed by a brief discussion.

BACKGROUND

House Finch Data from the BBS

The North American BBS is conducted each breeding season by volunteer observers who count the number of various species of birds along specified routes (Robbins et al. 1986). The BBS sampling unit is a roadside route approximately 39.2 km in length. An observer makes 50 stops over each route, during which birds are counted by sight and sound for a period of 3 minutes. Over 4000 routes have been included in the North American survey, but not all routes are available each year. As might be expected, due to the subjectivity involved in counting birds by sight and sound, and the relative experience and expertise of the volunteer observers, there is substantial observer error and bias in the BBS survey (e.g., Sauer et al. 1994).

BBS data are often used to investigate spatial and temporal variation in relative abundance (e.g., trends, Link and Sauer 1998). Spatial maps of relative abundance over time are crucial for these purposes, yet traditional methods for producing such maps are somewhat *ad hoc* (e.g., inverse distance methods) and do not always account for the special discrete, positive nature of the count data (e.g., Sauer et al. 1995) nor the sampling and observation error. Corresponding prediction uncertainties for maps produced in this fashion are not typically available. Recent studies have shown that Poisson-based hierarchical spatial models can be used effectively to provide uncertainty estimates for BBS maps (e.g., Royle et al. 2001, Wikle 2002). However, these studies have not focused on the temporal aspect of the problem.

In this study, we are interested in the relative abundance of the house finch through time. Figure 1 shows the location of the sampling route midpoints and observed counts

5

over the eastern United States (U.S.) for the 1969, 1979, 1989, and 1999 BBS. The size of the circle radius is proportional to the number of birds observed over each route and sampled locations for which no House Finch were observed are indicated by a "+". This figure suggests that the population exhibits spatial spread with time that is characteristic of an invasive species. Indeed, this species is native to the western U.S. and Mexico, and the eastern population is a result of a 1940 release of caged birds in New York (Elliott and Arbib 1953). Because the birds exhibit significant fecundity and their juveniles tend to disperse over relatively long distances, the eastern house finch population expanded to the west after the initial release. Although not shown here, the native west coast population has also expanded eastward as the human population has expanded eastward (and correspondingly, changed the environment/habitat). By the late 1990's, the two populations met in the central plains of North America.

Our interest is in predicting the spread of the eastern population over time at the landscape scale. Any such prediction must account for both irregular sampling, potential for significant observation error, and uncertainty regarding our understanding of the spread of the population with time.

Spatio-temporal Models in Ecology

There has been increasing interest in recent years in modeling the spatial distribution of ecological processes over time. Although by no means exhaustive, these have tended to focus on reaction-diffusion processes modeled via PDEs, integro-difference equations, discrete-time contact models, and cellular automata (e.g., Hastings 1996). Many of these can be shown to exhibit similar behavior and often the choice of one framework over the other depends on whether one is considering discrete time and/or space. In some cases the distinction is blurred with application to data-driven problems. For example, non-linear PDEs often can only be solved numerically, requiring discretization and are thus analogous to discrete time/space models in application.

Historically, diffusion models have been considered for the house finch data (e.g., Okubo, 1986; Veit and Lewis 1996). Early studies found that the house finch population range expanded slowly during the establishment phase followed by linear expansion at a higher rate (Mundinger and Hope 1982; Okubo 1986). Our interest is on the relatively large (landscape) scale spread of the house finch across the eastern U.S. from 1966 through the present. It is not obvious that the spread at such scales will be similar to that found in earlier studies (e.g., Veit and Lewis 1996). In particular, over these scales we should expect heterogeneity in the diffusion due in part to significant habitat variability.

Introduction to Bayesian Hierarchical Models

The essence of hierarchical modeling is based on the simple probability fact that the joint distribution of a collection of random variables can be decomposed into a series of conditional models. That is, if A, B, C are random variables, then we can always write a factorization such as $[A, B, C] = [A|B, C][B|C][C]$. (We use the notation $[w_1]$ to denote the probability distribution of w_1; $[w_2|w_1]$ represents the conditional distribution of w_2 *given* the random variable w_1.) This simple formula is the crux of hierarchical thinking. Imagine a complicated joint distribution that is difficult to specify. For example, in the spatio-temporal context, the joint distribution describes the stochastic behavior of the process at all spatial locations and all times. This is *extremely* difficult (if not impossible) to specify for complicated processes. Often, it is much easier to specify the distribution of the conditional models. Thus, the product of a series of relatively simple conditional models leads to a joint distribution that can be quite complicated. Although in some cases it is possible to model this joint distribution directly using generalized likelihood procedures (e.g., Lele et al. 1998), such approaches are not as flexible with regards to accounting for uncertainty as the Bayesian methodology.

For modeling complicated processes in the presence of data, we can write the hierarchical model in three primary stages:

Stage 1. Data Model: $[data|process, data\ parameters]$

Stage 2. Process Model: $[process|process\ parameters]$

Stage 3. Parameter Model: $[data\ and\ process\ parameters]$.

Thus, the fundamental idea is to approach the complex problem by breaking it into sub-problems. Although this idea has long been considered in statistics, its use for modeling complicated environmental and ecological processes is relatively new (e.g., Berliner 1996). The first stage is concerned with the observational process or "data model", which specifies the distribution of the data (say **BBS** *observations* of house finch relative abundance) *given* the process of interest (e.g., the *true* house finch abundance) and parameters that describe the data model. The second stage then describes the process, conditional on other process parameters. Perhaps this is some sort of reaction-diffusion process, with parameters describing the rate of diffusion and the growth rate. Finally, the last stage accounts for the uncertainty in the parameters, from both the data and process stages, by assigning them distributions. For example, if we believe the diffusion rate to be a function of space, then we might model these parameters as a spatially correlated process in the third stage. Each of these stages can have multiple sub-stages. For example, our process (house finch abundance) might be modeled as a product of several physically-motivated conditional distributions suggested by a state-space formulation. Similar decompositions are possible for the the parameter stage. For example, we might model a spatially-varying diffusion coefficient conditional on habitat processes.

Ultimately, we are interested in the distribution of the process and parameters updated by the data. We obtain this so-called "posterior" distribution via Bayes' Theorem:

$$[process, parameters|data] \propto [data|process, parameters][process|parameters][parameters].$$

$$(1)$$

This formula serves as the basis for Bayesian hierarchical prediction. However, several critical points remain. First, development of the parameterized component distributions on the right-hand side of (1) is challenging, but not an unusual aspect of stochastic model-

ing. Development of the parameter distribution (the *prior distribution* for the parameters) has historically been the focus of objections to the Bayesian approach due to its implied subjectiveness. Of course, the formulation of the data model and process model are quite subjective as well, but typically have not generated the same concern. The point is that the quantification of such subjective judgment is in fact the strength of the Bayesian approach, in that it provides a coherent probabilistic framework in which to incorporate the judgment, scientific reasoning, and experience explicitly in the model. Finally, although (1) looks simple in principle, it may be very difficult in practice to actually obtain the posterior distribution. The complexity and high dimensionality of ecological models prohibit the direct computation of the posterior in most cases. However, the recent development of Markov chain Monte Carlo (MCMC) as a computational tool in Bayesian statistics has made feasible implementations of hierarchical models in very complex settings.

HIERARCHICAL MODELING OF HOUSE FINCH ABUNDANCE

Our goal is to develop a model that can predict house finch abundance on a regular regular grid covering the eastern half of the U.S. The model must be able to predict both at grid cells in which there were and were not observations during the years for which we have BBS data, as well as forecast the abundance in the year immediately following the most recent survey period (2000, in our case). The modeling approach is stochastic and based largely on the Poisson modeling framework proposed by Diggle et al. (1998). The key difference here is the inclusion of spatio-temporal dynamic effects motivated by theoretical models of diffusion. We emphasize that the modeling framework is stochastic and, although we are motivated by analytical models, we are not restricted to their traditional implementation.

Data Model

9

Let $Z_t(\mathbf{s}_i)$ correspond to the observed count at time t, for $t = 1, \ldots, T$ and spatial location $\mathbf{s}_i = (x_i, y_i)$, corresponding to the midpoint of a BBS route. For each time t there are n_t observations denoted by $\mathbf{s}_i$ for $i = 1 \ldots, n_t$. We let this count have a Poisson distribution with mean $\lambda_t(\mathbf{s}_i)$,

$$Z_t(\mathbf{s}_i)|\lambda_t(\mathbf{s}_i) \sim Poisson(\lambda_t(\mathbf{s}_i)). \tag{2}$$

In this case, any two observations $Z_t(\mathbf{s}_i)$ and $Z_\tau(\mathbf{s}_j)$ (e.g., BBS counts at two different locations and different years) are assumed to be independent *conditional* on their means $\lambda_t(\mathbf{s}_i)$ and $\lambda_\tau(\mathbf{s}_j)$. Thus, dependence among the observations is induced by the random spatio-temporal process (λ), which is modeled at the next stage.

Process Model

It is customary in Bayesian implementation of Poisson models to let the Poisson intensity (mean) λ be distributed as a log-normal random variable (e.g., Diggle et al. 1998). In the spatio-temporal case, we assume the log of the intensity parameter follows a normal distribution with a time-varying trend, spatio-temporal random effects, and an uncorrelated noise term. Specifically,

$$\log(\lambda_t(\mathbf{s}_i)) = \mu_t + \mathbf{k}'_{it}\mathbf{u}_t + \eta_t(\mathbf{s}_i), \tag{3}$$

where μ_t is a time-varying mean process that is constant for all spatial locations at a given time, $\mathbf{u}_t$ is an $n \times 1$ vector representation of a gridded *latent* (i.e., unobserved) spatio-temporal dynamical process defined at grid locations that do not necessarily coincide with data locations, $\mathbf{k}_{it}$ is a known $n \times 1$ vector that maps the gridded process $\mathbf{u}_t$ to observation location $\mathbf{s}_i$, and $\eta_t(\mathbf{s}_i)$ is a noise term that does not exhibit dependence across space and time. These sub-processes are described in greater detail below. Note that we have a great deal of flexibility in our choice of $\mathbf{k}_{it}$. We can incorporate change of spatial

scale information, spatial smoothing, or simply make it an incidence vector (a vector of ones and zeros, with a one in the j-th position, relating the observation at $\mathbf{s}_i$ to the j-th element of the grid process $\mathbf{u}_t$). As shown below, in our case each observation route is assigned to the nearest grid location. This is rather simple, but sufficient for purposes of illustration. For more complicated implementations of $\mathbf{k}_{it}$ see Wikle et al. (2001).

The temporal process μ_t is included to capture the time trend of the mean house finch abundance. Our prior belief, based on previous studies (e.g., Okubo, 1986; Veit and Lewis 1996), is that house finch abundance increases relatively slowly in early years, followed by more rapid increase, and possibly a leveling off as saturation is reached (i.e., when the eastern and western populations meet in the central plains). There is also the possibility that Allee dynamics are present (Veit and Lewis 1996). A convenient and flexible model for such a process on the log-scale is a Gaussian random walk:

$$\mu_t = \mu_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid \ N(0, \sigma_\epsilon^2). \tag{4}$$

The $\eta_t(\mathbf{s}_i)$ process represents observer errors and small-scale spatio-temporal variation and is assumed to be independent across space and time. We specify a Gaussian distribution, $\eta_t(\mathbf{s}_i) \sim iid \ N(0, \sigma_\eta^2)$. The independence assumption may not be completely valid in this case since some observers may take observations over multiple routes, inducing dependence. However, detailed observer information that would be necessary to account for this dependence is not readily available. The assumption that the small-scale spatio-temporal variability is independent is also potentially suspect. It is assumed, however, that such an effect would be relatively minimal when compared to the landscape-scale spatio-temporal variability of primary interest here.

The key process model component is the latent spatio-temporal process $\mathbf{u}_t$. The motivation for this process comes from analytical models for reaction-diffusion. Although we could select a variety of representations for the reaction-diffusion component motivation (e.g., see Wikle et al. 2001 and Wikle 2001 for examples using PDE and

11

integrodifference-based models, respectively), we chose a PDE model because of historical relevance to the house finch problem (e.g., Okubo 1986). Consider the general diffusion PDE,

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial u}{\partial y}\right) + \alpha u, \qquad (5)$$

where $u_t(x,y)$ is the spatio-temporal process at spatial location $\mathbf{r} = (x,y)$ in two-dimensional Euclidean space at time t, $\delta(x,y)$ is a spatially varying diffusion coefficient and α is a growth coefficient. This is a generalization of Skellam's (1951) classic diffusion model plus Malthusian growth. The generalization is the specification of a spatially varying diffusion parameter. This is important for landscape-scale processes since habitat heterogeneity and physical barriers influence spatial spread. Studies of such models in relatively simple circumstances have shown them to be important for realistic diffusion (e.g., Shigesada et al. 1986). Although flexible, we do not expect this model to be the "correct" model for describing the underlying latent diffusion of the house finch relative abundance. For example, the Malthusian growth term is almost certainly not appropriate. However, this parameterization is relatively simple (and thus facilitates implementation) and in our case is included to account for "explosive" growth *beyond* that suggested by the temporal trend term μ_t. No theoretical model will be correct when applied to real data over highly diverse landscapes and long temporal periods. Thus, it is important to recognize that this model is only the basis for a stochastic dynamic model. That is, it is simply the motivation for our prior formulation and will ultimately be updated by the data via Bayes' Theorem. Furthermore, this prior model can be made quite flexible if we allow the parameters $\delta(x,y)$ and α to be random to account for effects such as environmental and demographic stochasticity. We emphasize that the diffusion is specified on a latent (unobserved) process, specified on the log-scale. This formulation provides additional flexibility in the stochastic model in terms of prediction.

Discretization of (5) using first-order forward differences in time and centered differ-

ences in space yields:

$$
\begin{aligned}
u_t(x,y) &= u_{t-\Delta_t}(x,y)\left[1 - 2\delta(x,y)\left(\frac{\Delta_t}{\Delta_x^2} + \frac{\Delta_t}{\Delta_y^2}\right) + \Delta_t\alpha\right] \\
&+ u_{t-\Delta_t}(x-\Delta_x,y)\left[\frac{\Delta_t}{\Delta_x^2}\left\{\delta(x,y) - \delta(x+\Delta_x,y) + \delta(x-\Delta_x,y)\right\}\right] \\
&+ u_{t-\Delta_t}(x+\Delta_x,y)\left[\frac{\Delta_t}{\Delta_x^2}\left\{\delta(x,y) + \delta(x+\Delta_x,y) - \delta(x-\Delta_x,y)\right\}\right] \\
&+ u_{t-\Delta_t}(x,y+\Delta_y)\left[\frac{\Delta_t}{\Delta_y^2}\left\{\delta(x,y) + \delta(x,y+\Delta_y) - \delta(x,y-\Delta_y)\right\}\right] \\
&+ u_{t-\Delta_t}(x,y-\Delta_y)\left[\frac{\Delta_t}{\Delta_y^2}\left\{\delta(x,y) - \delta(x,y+\Delta_y) + \delta(x,y-\Delta_y)\right\}\right] \\
&+ \gamma_t(x,y), \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (6)
\end{aligned}
$$

where it is assumed that the discrete u-process is on a rectangular grid with spacing Δ_x and Δ_y in the longitudinal and latitudinal directions, respectively, and with time spacing Δ_t. Readers familiar with numerical approaches to the solution of PDEs might wonder why we have not mentioned potential violation of the Courant-Friedrichs-Levy (CFL) condition for computational stability. It is important to recognize that this is not an issue with the Bayesian implementation so long as one has data to "control" the model. That is, unlike a pure numerical solution, we do not require long integrations of the PDE forward in time, independent of data. In fact, we actually prefer the model to have the potential to be unstable (to a reasonable extent) so that it can fit explosive growth over short time spans, if the data warrant.

The error term $\gamma_t(x,y)$ has been added to (6) to account for the uncertainties due to the discretization as well as other model misspecifications. This term provides extra flexibility in that it serves to induce stochastic forcing to the diffusion which can accommodate small pre-invasion colonies if the data warrant (at the low spatial resolution considered here). Thus, the value of the u-process at a given time is related to its past value at that location and its four nearest neighbors, plus some stochastic noise. Such mod-

els are known as space-time autoregressive (STAR) models (e.g., Pfeifer and Deutsch 1980). In high dimensions and/or with spatially varying parameters (as suggested here) these models are very difficult to fit. However, as shown in Wikle et al. (1998), the hierarchical perspective provides a reasonable mechanism by which to implement STAR models in these settings. For example, in the present case, the time-lagged nearest neighbor parameters that control the evolution of the u-process are functions of δ and α, and these are parameterized distributionally at a lower stage of the hierarchy. Thus, there is a congruence between traditional spatio-temporal statistical modeling and an ecologically motivated prior description of the underlying process.

For simplicity in presentation we can rewrite (6) in vector form as:

$$\mathbf{u}_t = \mathbf{H}(\boldsymbol{\delta}, \alpha)\mathbf{u}_{t-1} + \mathbf{H}_B(\boldsymbol{\delta})\mathbf{u}_{t-1}^B + \boldsymbol{\gamma}_t, \tag{7}$$

where again, $\mathbf{u}_t$ corresponds to an arbitrary vectorization of the gridded u-process at time t, $\mathbf{H}(\boldsymbol{\delta}, \alpha)$ is a sparse $n \times n$ matrix with five non-zero diagonals corresponding to the bracket coefficients in (6), hence its dependence on $\boldsymbol{\delta}$ and α. Note that without loss of generality, we have set $\Delta_t = 1$, corresponding to 1 year in our application. Furthermore, $\mathbf{u}_{t-1}^B$ is an $n_B \times 1$ vector of boundary values for the u-process, and $\mathbf{H}_B(\boldsymbol{\delta})$ is an $n \times n_B$ sparse matrix with elements corresponding to the appropriate coefficients from (6). Thus, the product $\mathbf{H}_B(\boldsymbol{\delta})\mathbf{u}_{t-1}^B$ is simply the specification of model edge effects. It is possible, and indeed desirable in many cases, to model the boundary (or edge) process $\mathbf{u}_t^B$ as a random process. One can show that this suggests a hierarchical approach to accommodating boundary conditions (edge effects) in stochastic solutions to PDE models (see Wikle et al. 2002). However, for simplicity, we assume $\mathbf{u}_t^B$ is fixed (at zero) for all time in the present application. As will be shown below, this is reasonable in the present application since the prediction grid boundaries are either over water areas or over land areas for which the house finch abundance is known to be relatively small. Furthermore, we let $\boldsymbol{\gamma}_t \sim iid\ N(\mathbf{0}, \sigma_\gamma^2 \mathbf{I})$ for $t = 1, \dots, T$. Note that in the probabilistic formulation

we must also specify the distribution for $\mathbf{u}_0$, the initial condition, as described below.

Parameter Models

We specify the following distribution for the diffusion parameters,

$$\boldsymbol{\delta}|\boldsymbol{\beta}, \sigma_\delta^2, \theta \sim N(\boldsymbol{\Phi}\boldsymbol{\beta}, \sigma_\delta^2 \mathbf{R}(\theta)), \tag{8}$$

where $\boldsymbol{\delta}$ is an $N \times 1$ vectorization of the $\delta(x, y)$ process defined on the n grid locations and n_B boundary locations ($N = n + n_B$). The $N \times p$ matrix $\boldsymbol{\Phi}$ consists of known spatially-referenced "covariates" with $\boldsymbol{\beta}$ the corresponding $p \times 1$ vector of "regression" coefficients. Typically, the covariates may correspond to known factors that influence the diffusion process such as habitat, human population, geographical barriers, climate, etc. The correlation matrix $\mathbf{R}(\theta)$ depends on some parameter θ that describes the spatial dependence. In other words, this correlation matrix accounts for additional spatial correlation in the δ-process. An alternative to specification of covariates is to specify a general spatial random process for $\boldsymbol{\delta}$ and examine the posterior to "discover" potential important factors that affect the dispersion. We take this approach in the present study. For sake of simplicity, we let $\boldsymbol{\Phi} = \mathbf{1}$, a vector of ones corresponding to an overall mean for $\boldsymbol{\delta}$, and specify the correlation matrix $\mathbf{R}(\theta)$ by assuming the spatial dependence of the process can be described by an exponential correlation function, $r(d) = \exp(-d/\theta)$ where r is the correlation, d is the distance between the δ process at two grid locations and θ is a random spatial dependence parameter described below. The normal distribution assumption for $\boldsymbol{\delta}$ could, in principle, allow diffusion coefficients to become negative. Although this would not make sense from an ecological perspective, it is not necessarily unreasonable in terms of the parameterization of the propagator matrix $\mathbf{H}$. In the present application, such concerns are mitigated by the fact that the posterior distribution of $\boldsymbol{\delta}$ includes only positive values. One could, however, specify an alternative distribution that guaranteed positive values, at the expense of minor inconvenience in the MCMC

sampling algorithm.

We specify a simple normal distribution for the "growth" parameter α. That is, assume $\alpha \sim N(\tilde{\alpha}_0, \tilde{\sigma}_\alpha^2)$, with known mean and variance parameters $\tilde{\alpha}_0$ and $\tilde{\sigma}_\alpha^2$, respectively. It would be reasonable to allow this parameter to vary with space as well. However, given that most of the latent growth is modeled by μ_t in this example, and the fact that the diffusion parameter is heterogeneous, it was decided that this additional spatial variability was unnecessary. In fact, it is not clear *a priori* that the growth term is even needed in the latent diffusion parameterization given the presence of the temporal trend term μ_t. In this case, we will be interested in the posterior distribution which may suggest, on the basis of data, if this prior assumption was reasonable.

Additional Parameter Distributions

To complete the model hierarchy we must specify distributions for the parameters from the previous stages. For example, recognizing that the initial conditions for $\mathbf{u}$ and μ are unknown, we account for that uncertainty by giving them the following distributions: $\mathbf{u}_0 \sim N(\tilde{\mathbf{u}}_0, \tilde{\sigma}_0^2 \mathbf{I})$, $\mu_0 \sim N(\tilde{\mu}_0, \tilde{\sigma}_\mu^2)$. In addition, we let $\boldsymbol{\beta} \sim N(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}}_\beta)$, $\sigma_\eta^2 \sim IG(\tilde{q}_\eta, \tilde{r}_\eta)$, $\sigma_\gamma^2 \sim IG(\tilde{q}_\gamma, \tilde{r}_\gamma)$, $\sigma_d^2 \sim IG(\tilde{q}_d, \tilde{r}_d)$, $\sigma_\epsilon^2 \sim IG(\tilde{q}_\epsilon, \tilde{r}_\epsilon)$, and $\theta \sim U(\tilde{\theta}_L, \tilde{\theta}_U)$. Note that $IG(q, r)$ refers to an inverse gamma distribution with parameters q and r and $U(a, b)$ refers to a uniform distribution over the continuous range from a to b. Our choice for these distributions were chosen for computational convenience. In principle, the parameters that make up these distributions (those with the "tilde") can also be given distributions. However, in practice, we typically specify these "hyperparameters" and test the model for sensitivity to our selections. For the most part, our choices for these hyperparameters are based on subjective scientific notions about the various parameters. In cases where this knowledge is weak we make the prior specification "vague" by forcing the variance to be relatively large. Our choice of hyperparameter specifications is shown in Table 1.

Hierarchical Model Summary

The Bayesian formulation of the hierarchical model can be summarized by the following posterior distribution:

$$
[\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_T, \mathbf{u}_0, \ldots, \mathbf{u}_T, \mu_0, \ldots, \mu_T, \boldsymbol{\delta}, \theta, \alpha, \sigma_\eta^2, \sigma_\gamma^2, \sigma_\epsilon^2, \boldsymbol{\beta}, \sigma_\delta^2 | \mathbf{Z}_1, \ldots, \mathbf{Z}_T]
$$
$$
\propto \quad \left\{ \prod_{t=1}^{T} \prod_{i=1}^{n_t} [Z_t(\mathbf{s}_i) | \lambda_t(\mathbf{s}_i)][\lambda_t(\mathbf{s})_i | \mu_t, \mathbf{u}_t, \sigma_\eta^2] \right\} \left\{ \prod_{t=1}^{T} [\mathbf{u}_t | \boldsymbol{\delta}, \alpha, \mathbf{u}_{t-1}, \sigma_\gamma^2][\mu_t | \mu_{t-1}, \sigma_\epsilon^2] \right\}
$$
$$
\times \quad [\mathbf{u}_0][\mu_0][\boldsymbol{\delta} | \boldsymbol{\beta}, \sigma_\delta^2, \theta][\theta][\alpha][\sigma_\eta^2][\sigma_\gamma^2][\sigma_\epsilon^2][\boldsymbol{\beta}][\sigma_\delta^2],
$$

$$(9)$$

where $\mathbf{Z}_t$ and $\boldsymbol{\lambda}_t$ are $n_t \times 1$ vectorizations of the n_t observations and associated Poisson means at time t, respectively. There certainly is no analytical solution to this posterior (i.e., we cannot integrate the RHS of (9) to find the constant of proportionality). However, as outlined in the Appendix, we can use MCMC methods to obtain samples from this posterior distribution. For an overview of MCMC methodologies see Gilks et al. (1996) and Robert and Casella (1999). For complicated spatio-temporal applications of these methods, see Wikle et al. (1998), Berliner et al. (2000), and Wikle et al. (2001). For applications to spatial modeling of BBS data, see Royle et al. (2001) and Wikle (2002).

Prediction

In addition to estimating model parameters (e.g., μ_t and $\boldsymbol{\delta}$), we are interested in predicting the "grid box average" house finch mean relative abundance at all spatial grid locations for each time 1966-2000. Recall that the last year for which data were available to us was 1999, so the 2000 prediction is a "forecast". In addition to such predictions, we seek measures of uncertainty (e.g., prediction variances) to quantify map precision. Thus we obtain estimates of posterior means and corresponding posterior standard deviations using MCMC methods. Specifically, we would like predictions of the Poisson mean process at all grid locations free of observer effects. To obtain these predictions we simply sample from $\log(\lambda_t)$ at each grid cell, and exponentiate those simulated values. See the Appendix for details.

MODELING RESULTS

The prediction grid is shown in Figure 2, and consists of a 28 by 18 grid approximately spaced at 1 degree by 1 degree intervals. The approximation occurs because an equal area map projection is used to ensure all grid boxes cover the same area, and this results in some distortion of the model grid in terms of latitude and longitude. Given that our interest is in the landscape-scale distribution of relative abundance, this grid spacing is adequate. If more detailed spatial features are of interest, then higher resolution grids can be considered, at the expense of longer computation times. The conclusions presented below are not overly sensitive to the grid spacing.

The MCMC simulation was run for 10,000 iterations beyond a reasonable "burn-in" period (2000 iterations) and output was summarized. Table 2 shows the posterior means and posterior standard deviations for the univariate parameters in the model. In addition, Figure 3(a) shows posterior percentiles of the μ_t process from 1966 to 2000. Recall that this variable accounts for the variation in the Poisson mean with time on the log-scale. It is analogous to the average population change with time. From the time series of the posterior median note the initial decrease in the late 1960's to early 1970's, perhaps suggestive of demographic or environmental stochasticity, or possibly an Allee effect (Veit and Lewis 1996). However, from the lower and upper 2.5 percentiles we note that there is considerable uncertainty in the μ_t posterior during this time period. Thus, given the BBS data, one can't be certain that the process did in fact decrease during this period. This distributional information is thus critical, and illustrates the strength of the Bayesian hierarchical approach. Note, the initial decrease in μ_t is followed by a linear increase (on the log-scale) until the mid 1990's after which there was a decrease and leveling off. This period corresponds to the timing of the coalescence of the eastern and western house finch populations. These results do appear "significant" relative to the posterior percentiles. Figure 3(b) shows posterior percentiles of $\exp(\mu_t)$, clearly suggesting the Poisson mean experiences exponential growth over time until saturation.

The consideration of arbitrary nonlinear functions of parameters and the ease in which their marginal posterior distributions can be obtained is a strength of the Bayesian hierarchical approach. We can also examine marginal distributional results for univariate parameters as shown by the posterior histograms of σ_η^2 and α, in Figure 3(c) and 3(d), respectively. There is significant observer/small-scale error in this case, which translates to large variability of observations (observation error) on the Poisson scale. In addition, the parameter α is significant in the posterior. Although this is the Malthusian growth parameter in the classical diffusion model, it cannot be interpreted as such in the present model since most of the growth of population over time (on the log scale) is accounted for by the μ_t term. There is, however, a tradeoff between the α term and the μ_t term.

It is also instructive to examine spatial maps of the posterior mean for the $\boldsymbol{\delta}$ process (the diffusion coefficient). Figure 4 shows the posterior mean and standard deviation maps for these parameters. Note, results for grid boxes that totally encompass water (great lakes and Atlantic ocean) are not considered. These figures suggest that there is indeed significant heterogeneity in the diffusion of the latent process, with the most intense diffusion occurring in the midwest, where the axis of maximum diffusion extends from eastern Missouri through western Pennsylvania. We emphasize that the estimation of heterogeneous diffusion coefficients (and a measure of their uncertainty) is an extremely useful by-product of the hierarchical formulation. Estimation of such effects from traditional methodologies is extremely difficult.

A primary purpose of the modeling was to produce landscape-scale spatial maps of the mean house finch relative abundance with time. Specifically, to predict both spatially (at locations with no data) and temporally (into the future), with realistic measures of prediction uncertainty. Figure 5 shows the posterior mean of the latent dynamic process $\mathbf{u}_t$ (left column), the λ_t-process (center column), and the posterior standard deviation for the λ_t-process (right column) at five year intervals from 1975 through 2000, where the year 2000 images are forecasts. Although one can get the impression of dynamic evolution in these images, it is more revealing to examine an animation of the yearly images

(e.g., as a "movie"). In that case, it is clear that the population "oscillates" (periods of increase and decrease) in the initial establishment period and expands rapidly during the expansion period, followed by "oscillation" as the population reaches saturation.

DISCUSSION

Ecological prediction is difficult. It is even more so in the presence of uncertainty. Observations have errors and we seldom know the true underlying dynamical model for a given phenomenon. For example, in the context of invasive species, there may be demographic and environmental stochasticities, Allee dynamics, coalescing colonies, jump diffusions, and heterogeneous rates of spread and growth. Theoretical models, although able to include some of these effects in simplistic scenarios, generally do not have the flexibility to accommodate the *a priori* uncertainty related to the dynamical assumptions. On the other hand, purely stochastic models for such processes are often over-parameterized and face significant problems when it comes to estimation. We have demonstrated that one alternative, the Bayesian hierarchical framework, can be used to build complicated spatio-temporal prediction models and quantify the associated uncertainties.

A reasonable question might ask whether the methodology presented here is worth the additional effort required for its implementation. First, for realistic estimates of prediction error, one must account for uncertainties in data and model. Traditional approaches to this problem simply do not account for the uncertainty in both data and model (if either). Certainly, any such method would be no simpler to implement than the approach presented here. If we are to be serious about ecological prediction, then we must provide realistic measures of prediction uncertainty.

Prediction aside, one might ask if the Bayesian hierarchical approach that accommodates uncertainties provides a richer understanding of the process. In the present application, the answer is yes. First, as shown in Figure 3, we have for the first time

some sense of the uncertainty in mean house finch relative abundance over time. This has not been available from traditional analyses. Thus, although the posterior distribution of mean relative abundance with time suggests that the population decreases in early periods, possibly indicating an Allee effect, there is sufficient uncertainty in these estimates to question that interpretation.

In addition, Figure 4 shows that the diffusion rate is heterogeneous, and we have a measure of our certainty in these estimates. Traditional approaches to this problem have no (easy) way of estimating heterogeneous diffusion coefficients and providing estimates of the certainty of those estimates. This information is critical to understanding the spread of the house finch on landscape scales. For example, it is well-known that the house finch is often associated closely with human habitation. As suggested by the coarse U.S. population data shown in Figure 6, the axis of greatest diffusion in the Midwest is correlated spatially with the axis of greatest population (also through the Midwest). The uncertainty estimate for δ is especially important in this case since the BBS survey routes are also correlated with population, adding potential sampling bias. In addition, there are missing data, observational errors, and biases in selection of the prediction grid. Some of these biases and uncertainties are accounted for in the model and reflected in the posterior distribution of the diffusion parameter (and thus the posterior standard deviation map). Obtaining such estimates of spatially-varying diffusion coefficients in the presence of such uncertainties is difficult in traditional statistical models.

A number of issues arise when applying a Bayesian hierarchical model as demonstrated here. In particular, how does one decide on the appropriate parameterizations? Often, there is scientific insight (e.g., the reaction-diffusion motivation) for the most critical processes. One seeks to allow the models that are based on this insight to have the flexibility to "adapt" to the data. Of course, there will always be distributions about which one knows little. In this case, it is hoped that the data will provide direction through the Bayesian updating, or that the parameters in question are not "critical". Perhaps even more disturbing to those beginning to explore these models is that there are multiple pa-

rameterizations that might work equally well. For example, we could have modeled μ_t as constant through time and allowed the α parameter to accommodate the exponential growth (as in the traditional PDE framework). Although this might be a more pleasing model for interpretation, it is not likely to predict any better than the chosen model. In fact, it was found that the MCMC algorithm converged more rapidly when μ_t was time-varying, due most likely to the potential computational instabilities that can be induced with large α. A similar issue would be to model the abundance dynamics directly, rather than as a latent Poisson mean process. It is difficult, however, to account for observational uncertainty if one is directly modeling the relative abundances. The generalized linear mixed model framework at the heart of the hierarchical Poisson model can account for observational effects through the extra Poisson variability term (η).

Finally, there is much to be done with the exploration of hierarchical spatio-temporal models in ecology and other disciplines. These efforts, by their vary nature, will require strong collaborations between applied mathematicians, statisticians, and subject matter scientists.

ACKNOWLEDGMENT

LITERATURE CITED

Berliner, L. M. 1996. Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver (Eds.). Kluwer Academic Publishers, 15-22.

Berliner, L.M., C.K. Wikle, and N. Cressie. 2000. Long-lead prediction of Pacific SSTs via Bayesian Dynamic Modeling. *Journal of Climate* **13**:3953-3968.

Clark, J.S. et al. 2001. Ecological Forecasts: An Emerging Imperative. *Science* **293**:657-660.

Diggle, P.J., J.A. Tawn, and R.A. Moyeed. 1998. Model-based geostatistics (with discussion). *Applied Statistics* **47**:299-350.

Elliott, J.J., and R.S. Arbib. 1953. Origin and status of the house finch in the eastern United States. *Auk* **70**:31-37.

Gilks, W.R., Richardson, S., and D.J. Spiegelhalter (eds.). 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Hastings, A. 1996. Models of spatial spread: Is the theory complete? *Ecology* **77**:1675-1679.

Lele, S., M.L. Taper, and S. Gage. 1998. Statistical analysis of population dynamics in space and time using estimating functions. *Ecology* **79**:1489-1502.

Levin, S.A., B. Grenfall, A. Hastings, and A.S. Perelson. 1997. Mathematical and computational challenges in population biology and ecosystems science. *Science* **275**:334-343.

Link, W.A., and J.R. Sauer. 1998. Estimating population change from count data: application to the North American Breeding Bird Survey. *Ecological Applications* **8**:258-268.

McCulloch, C.E., and S.R. Searle. 2001. *Generalized, Linear, and Mixed Models*. New York: John Wiley and Sons.

Mundinger, P.C., and S. Hope. 1982. Expansion of the winter range of the House Finch: 1947-79. *Am. Birds* **36**:347-353.

Okubo, A. 1986. Diffusion-type models for avian range expansion. In *Acta XIX Congressus Internationalis Ornithologici*, National Museum of Natural Sciences, University of Ottawa Press, 1038-1049.

Pfeifer, P.E. and S.J. Deutsch. 1980. Identification and interpretation of first order space-time ARMA models. *Technometrics* **22**:397-408.

Ripley, B.D. 1987. *Stochastic Simulation*. J. Wiley, New York.

Robbins, C.S., D.A. Bystrak, and P.H. Geissler. 1986. The Breeding Bird Survey: its first fifteen years, 1965-1979. USDOI, Fish and Wildlife Service Resource Publication 157. Washington, D.C.

Robert, C.P., and G. Casella. 1999. *Monte Carlo Statistical Methods*. Springer, New York.

Royle, J.A., W.A. Link, and J.R. Sauer. 2001. Statistical mapping of count survey data.

In Predicting Species Occurrences: Issues of Scale and Accuracy, (Scott, J. M., P. J. Heglund, M. Morrison, M. Raphael, J. Haufler, B. Wall, editors). Island Press. Covello, CA. (to appear)

Sauer, J.R., B.G. Peterjohn, and W.A. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk* 111:50-62.

Sauer, J.R., G.W. Pendleton, and S. Orsillo. 1995. Mapping of bird distributions from point count surveys. Pages 151-160 in C.J. Ralph, J.R. Sauer, and S. Droege, eds. Monitoring Bird Populations by Point Counts, USDA Forest Service, Pacific Southwest Research Station, General Technical Report PSW-GTR-149.

Skellam, J.G. 1951. Random dispersal in theoretical populations. *Biometrika* bf 38:196-218.

Shigesada, N., K. Kawasaki, and E. Teramoto. 1986. Traveling periodic waves in heterogeneous environments. *Theoretical Population Biology* **30**:143-160.

Veit, R.R., and M.A. Lewis. 1996. Dispersal, population growth, and the Allee effect: Dynamics of the house finch invasion of eastern North America. *The American Naturalist* **148**:255-274.

West, M., and J. Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag: New York.

Wikle, C.K. 2002. Spatial modeling of count data: A case study in modelling breeding bird survey data on large spatial domains. In *Spatial Cluster Modelling*, A. Lawson and D. Denison, eds. Chapman and Hall, 199-209.

Wikle, C.K. 2001. A kernel-based spectral approach for spatio-temporal dynamic models. *Proceedings of the 1st Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes (METMA)*. Benicassim, Castellon, Spain. 28-31 October 2001. 167-180.

Wikle, C.K. and N. Cressie. 1999. A dimension reduction approach to space-time Kalman filtering. *Biometrika* **86**:815-829.

Wikle, C.K., Berliner, L.M., and N. Cressie. 1998. Hierarchical Bayesian space-time models. *Journal of Environmental and Ecological Statistics* **5**:117–154.

Wikle, C.K., R.F. Milliff, D. Nychka, and L.M. Berliner. 2001. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96**:382-397.

Wikle, C.K., Berliner, L.M., and R.F. Milliff. 2002. Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. *Monthly Weather Review*. Tentatively accepted.

Table 1: Fixed Model Parameters

Parameter	Prior Value
$\tilde{\alpha}_0$	0
$\tilde{\sigma}_\alpha^2$	0.4
$\tilde{u}_0$	0
$\tilde{\sigma}_0^2$	1
$\tilde{\mu}_0$	-6
$\tilde{\sigma}_\mu^2$	10
$\tilde{\beta}$	0
$\tilde{\mathbf{\Sigma}}_\beta$	$100 \times \mathbf{I}$
$\tilde{q}_\eta$	1
$\tilde{r}_\eta$	4
$\tilde{q}_\gamma$	.5
$\tilde{r}_\gamma$	1
$\tilde{q}_\epsilon$	.25
$\tilde{r}_\epsilon$	10
$\tilde{q}_\delta$	.0005
$\tilde{r}_\delta$	.05
$\tilde{\theta}_L, \tilde{\theta}_U$	$[1 \ , \ 400]$

Table 2: Posterior Mean and Standard Deviations for Univariate Parameters

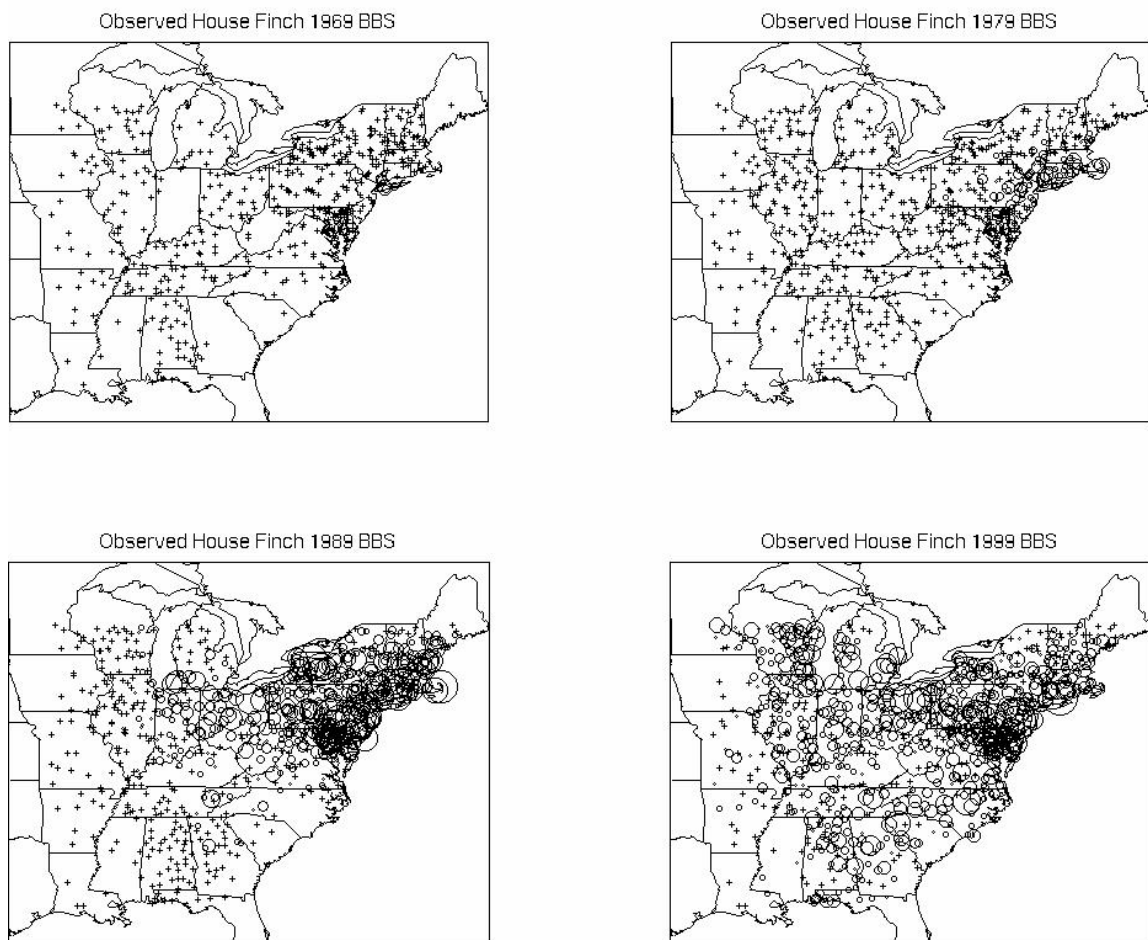| Parameter	Posterior Mean	Posterior Standard Deviation
σ_η^2	1.0569	0.0204
σ_γ^2	0.2851	0.0129
σ_δ^2	3.734×10^{-5}	2.725×10^{-5}
σ_ϵ^2	0.1311	0.0349
α	0.011	0.0040
θ	144.5	21.3

Figure 1: Location of BBS survey route and observed house finch count for 1969, 1979, 1989, and 1999. The radius of the circles are proportional to the observed count, and survey routes with zero counts are indicated by a "+".
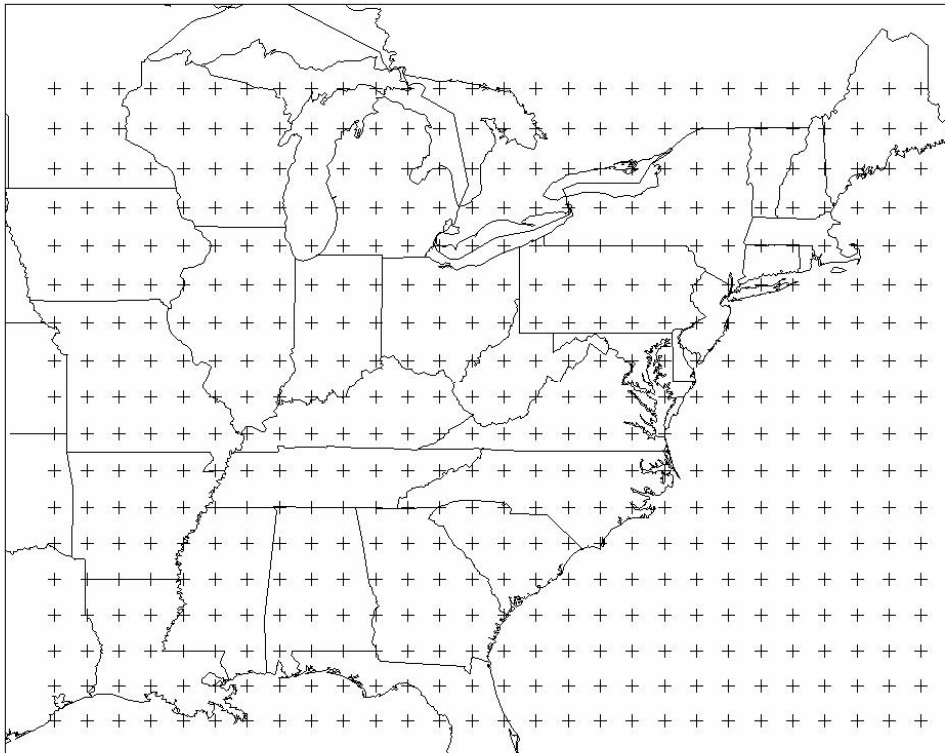
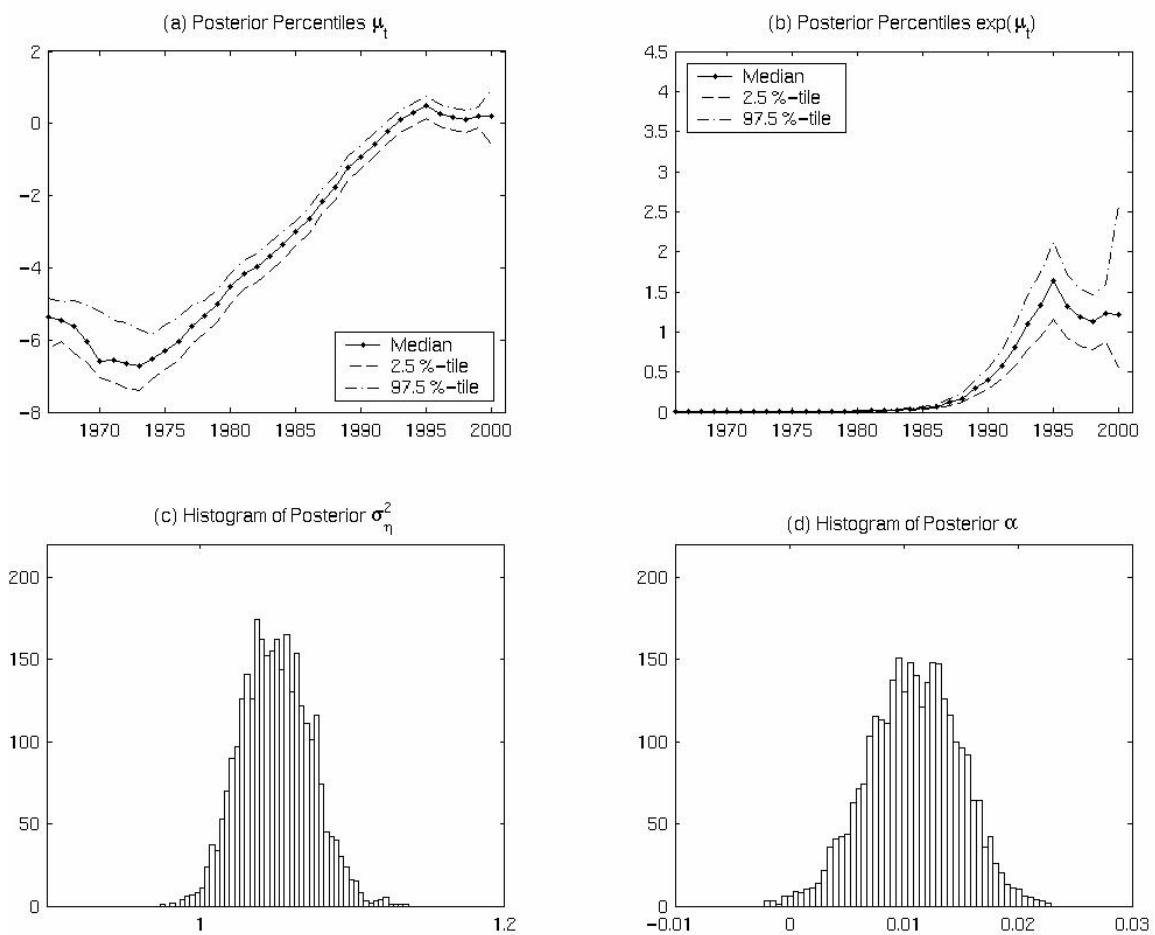Figure 2: Prediction grid; center of prediction grid box denoted by "+".

Figure 3: Summaries of model parameters from the MCMC analysis. (a) Percentiles of the marginal posterior distribution for μ_t; (b) Percentiles of the marginal posterior of $\exp(\mu_t)$; (c) Histogram of posterior samples for σ_η^2; (d) Histogram of posterior samples for α.

Figure 4: (a) Posterior mean of diffusion parameter, $\boldsymbol{\delta}$. (b) Posterior standard deviation of diffusion parameter, $\boldsymbol{\delta}$.

Figure 5: Posterior maps every 5 years from 1975 through 2000. Left Column: Posterior mean of u. Center Column: Posterior mean of λ. Right Column: Posterior standard deviation of λ.

Figure 6: Population counts from 2000 U.S. census.

Figure 1: Location of BBS survey route and observed house finch count for 1969, 1979, 1989, and 1999. The radius of the circles are proportional to the observed count, and survey routes with zero counts are indicated by a "+".

Figure 2: Prediction grid; center of prediction grid box denoted by "+".

Figure 3: Summaries of model parameters from the MCMC analysis. (a) Percentiles of the marginal posterior distribution for μ_t; (b) Percentiles of the marginal posterior of $\exp(\mu_t)$; (c) Histogram of posterior samples for σ_η^2; (d) Histogram of posterior samples for α.
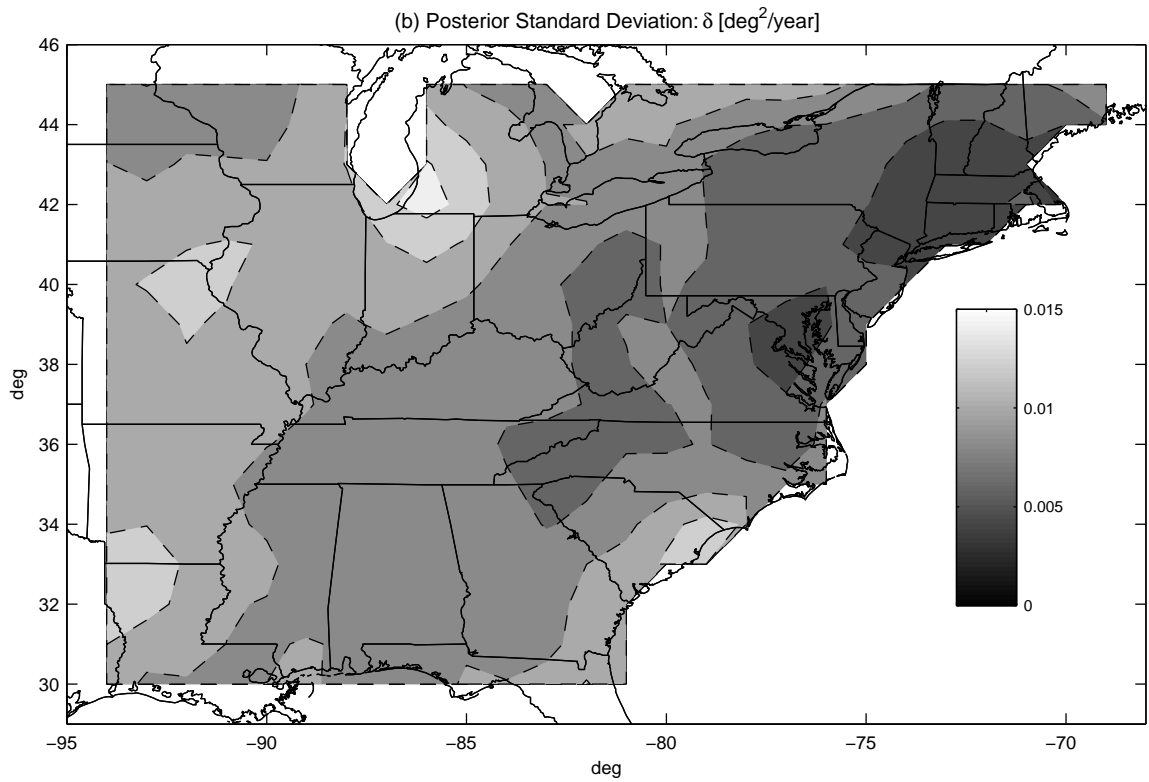
(a) Posterior Mean: δ [deg²/year]

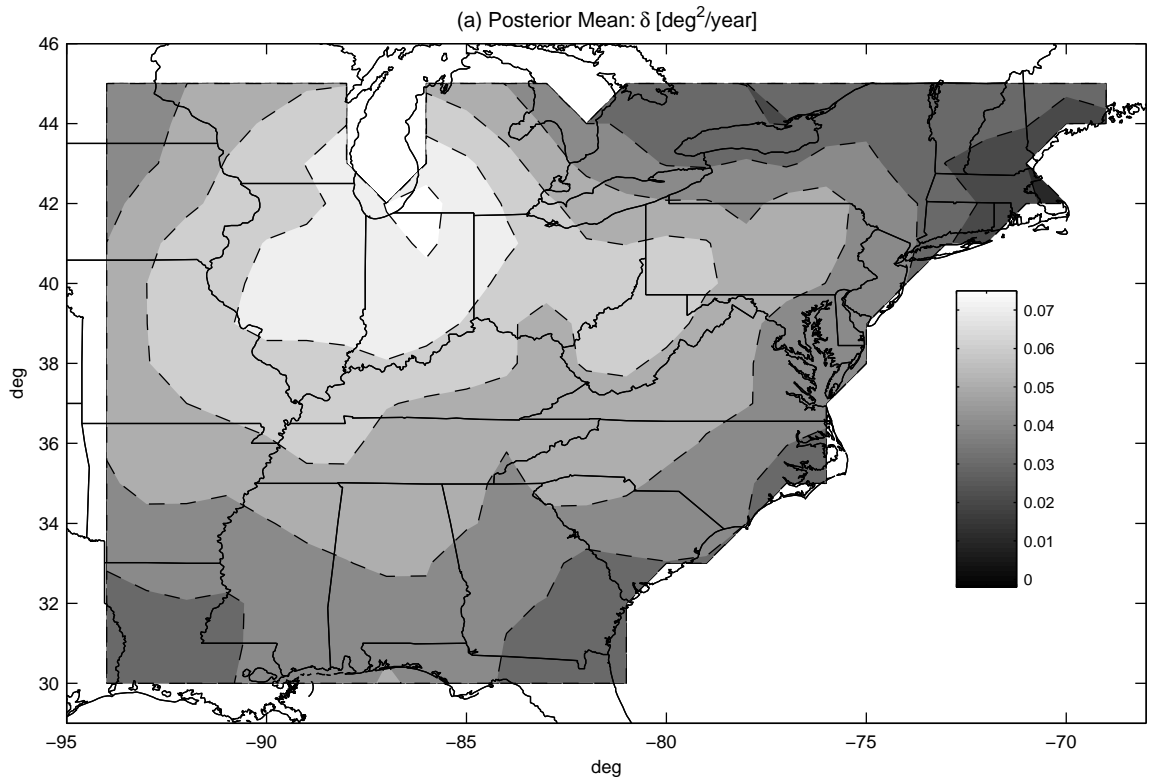(b) Posterior Standard Deviation: δ [deg²/year]

Figure 4: (a) Posterior mean of diffusion parameter δ. (b) Posterior standard deviation.
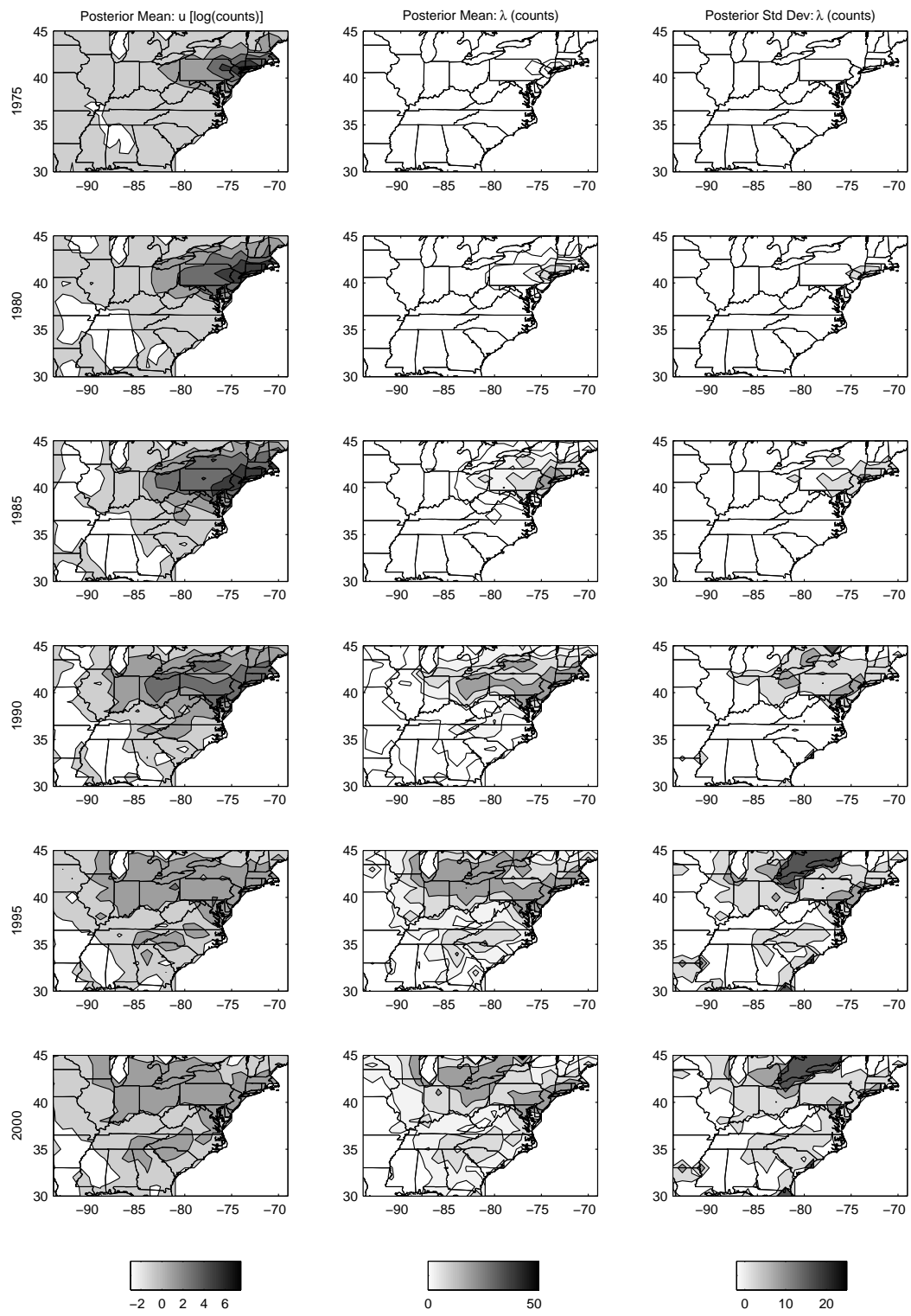
Figure 5: Posterior maps every 5 years from 1975 through 2000. Left Column: Posterior mean of u. Center Column: Posterior mean of λ. Right Column: Posterior standard deviation of λ.

Figure 6: Population counts from 2000 U.S. census.