20 Optimal spatio-temporal monitoring designs for characterizing population trends

Mevin B. Hooten, Beth E. Ross, and Christopher K. Wikle

Introduction

Spatio-temporal modeling

Spatio-temporal statistical models are being used increasingly across a wide variety of scientific disciplines to describe and predict spatially explicit processes that evolve over time. Correspondingly, in recent years there has been a significant amount of research on new statistical methodology for such models. Although descriptive models that approach the problem from the second-order (covariance) perspective are important, and innovative work is being done in this regard, many real-world processes are dynamic, and it can be more efficient in some cases to characterize the associated spatio-temporal dependence by the use of dynamical models. The chief challenge with the specification of such dynamical models has been related to the curse of dimensionality. Even in fairly simple linear, first-order Markovian, Gaussian error settings, statistical models are often over parameterized. Hierarchical models have proven invaluable in their ability to deal to some extent with this issue by allowing dependency among groups of parameters (Cressie *et al.* 2009). In addition, this framework has allowed for the specification of science-based (i.e. based on knowledge and hypotheses about the ecological system) parameterizations (and associated prior distributions) in which classes of deterministic dynamical models [e.g. partial differential equations (PDEs), integro-difference equations (IDEs), matrix models, and agent-based models] are used to guide specific parameterizations (Wikle and Hooten 2010).

Therefore, two of the main questions to ask in early stages of building a statistical model for a spatio-temporal ecological process are:

- Does enough a priori scientific information exist about the process to specify an intelligent mechanistic model that can mimic it?
- If the answer to the above question is "yes", then do sufficient data exist (or can they be collected) to fit such a model?

If the answer to either of these questions is "no", then perhaps a more naïve statistical model is warranted – one that is still spatially and temporally explicit, but that is

Design and Analysis of Long-term Ecological Monitoring Studies, ed. R.A. Gitzen, J.J. Millspaugh, A.B. Cooper, and D.S. Licht. Published by Cambridge University Press.

sufficiently parsimonious to enable statistical learning. In that situation, because the actual model structure is limited, it is critical to consider potential lurking sources of latent autocorrelation, both spatially and temporally.

Optimal design

In many ecological studies, cost can be a limiting factor, and as available resources and budgets change over time, an adaptive sampling design can be of value (Hooten *et al.* 2009a; Box 20.1). Optimal adaptive sampling has been shown to increase sampling efficiency in biological populations (Brown *et al.* 2008, Salehi and Brown 2010; Chapter 17), although this may be related to factors such as quadrat size, initial sample size, and the method of selection (Smith *et al.* 1995, Giudice *et al.* 2010). Monitoring strategies based on adaptive sampling have been utilized across a wide variety of specific applications; for example, in capture–mark–recapture studies (Conroy *et al.* 2008b), occupancy studies (Field *et al.* 2005, MacKenzie and Royle 2005), and combinations of these (Witczuk *et al.* 2008). Adaptive sampling strategies have also been utilized in waterfowl surveys (Smith *et al.* 1995, Pearse *et al.* 2009), and many of these studies highlight the importance of considering the desired inference prior to collecting the data and examining ways to improve precision or decrease costs (e.g. by adding a second observer in aerial surveys to double the transect strip width; Pearse *et al.* 2009).

The basic concept of optimal monitoring involves the following steps.

(i) Identify a critical modeled quantity of interest about which improved inference is desired (e.g. predictions, forecasts, or parameter estimates).
(ii) Specify a design criterion that quantifies the form of improvement desired (e.g. precision).
(iii) Pose a model that explicitly considers the data that you plan to collect optimally in future monitoring efforts.
(iv) Fit the model, given available data, and optimize the design for new data collection with respect to the design criterion.

Diggle and Lophaven (2006) describe such two-part adaptive monitoring schemes as using a prospective sample (i.e. current data) to identify an optimal retrospective sample and therefore to optimize future sampling. Some studies have taken this basic theme and extended it to the situation where the process of interest is dynamic and thus the optimal design itself is dynamic (e.g. Wikle and Royle 2005, Fuentes *et al.* 2007, Hooten *et al.* 2009a).

Background for example application: scaup breeding pair trends

Scaup are the most abundant and widespread diving duck in North America, and an important game species in many areas (Austin *et al.* 1998). Since 1978, the continental population of scaup (lesser scaup, *Aythya affinis*, and greater scaup, *A. marila*, collectively) has been declining at an average rate of 105 000 scaup per year, and numbers are

Box 20.1 Take-home messages for program managers

Because nearly all ecological data are collected over time and space, contemporary methods of analysis need to consider these aspects explicitly. Hierarchical models provide a convenient framework for specifying both the observational process by which monitoring data are collected and complicated underlying ecological spatio-temporal process behavior (Cressie *et al.* 2009) such as spatial variability in trends and spatial/temporal correlation among monitoring units within and across time. Spatio-temporal behavior can be modeled as a latent "nuisance" source of error or as an integral part of the underlying dynamic process describing changes in the resource of interest (Wikle and Hooten 2010). In either case, statistical structure can be exploited to fuse the traditionally distinct acts of monitoring and modeling for improved inference and conservation of resources. Although these benefits come at the cost of increased analytical complexity, the necessary analytical building blocks are increasingly becoming accessible (e.g. Royle and Dorazio 2008).

The basic premise behind optimal monitoring is to use the existing data (or other sources of a priori information about the process under study) to help you put monitoring effort in places that will aid inference the most. Adaptive design refers to the process of altering a study in light of new information to increase precision or decrease costs, whereas dynamic design (e.g. Wikle and Royle 2005, Hooten *et al.* 2009a) is also adaptive but assumes an underlying dynamic model for the ecological resource and relies on the model to predict future states of the system and therefore guide monitoring decisions. In either case, the ultimate type of inference and information desired from the monitoring effort should be considered explicitly early during survey design (Chapter 2) and throughout the data-collection phase to help ensure that the data are collected in a manner that decreases uncertainty (Box 20.2). This allows the manager to do more with less.

currently well below the goal set in the North American Waterfowl Management Plan (Afton and Anderson 2001; USFWS 2009a, b). Additional modeling of the population has indicated further decline since 1997, sparking concern amongst hunters, management agencies, and conservation groups. Several hypotheses have been proposed regarding the cause of the decline [e.g. The Spring Condition Hypothesis, which suggests that there has been a decline in the condition of females when they arrive on breeding grounds (Anteau and Afton 2004), and hypotheses related to increasing levels of heavy metals (Anteau *et al.* 2007)], but contradictory evidence has been found, indicating a lack of certainty concerning these potential causes (DeVink *et al.* 2008a, b). It appears that the decline is more pronounced in the boreal forest breeding areas (Afton and Anderson 2001), emphasizing the importance of understanding the population dynamics across a broad spatial scale. There is strong rationale for the use of an explicit spatio-temporal model to better understand the scaup breeding population (Austin *et al.* 2006).

For continental-level studies on ducks, most researchers use the Waterfowl Breeding Population and Habitat Survey (Canadian Wildlife Service and US Fish and Wildlife

Box 20.2 Common challenges: necessary decisions

The biggest and most important decisions to make when building a spatio-temporal hierarchical model for optimal adaptive monitoring are as follows.

- What type of questions do you want to answer after the data are collected? For example, are you interested in learning about population size or change? Is your focus going to be centered on estimating survival or measurement characteristics such as detection bias?
- Do you know enough about the system under study to construct a scientific physical process model? For example, is a partial differential equation a good motivating mechanistic model for the system you are studying, or do you not have enough existing knowledge about it to tell? If not, perhaps you just want to account for spatial and temporal variation without the interactions required to mimic a more complicated process model?
- What type of design criterion will best help you answer your initial question? For example, are you most interested in reducing uncertainty about specific parameters in the model you have specified (i.e. trend coefficients), or do you wish to focus primarily on forecasting with minimal uncertainty?
- What flexibility do you have in modifying the monitoring design? Suppose that, for consistency reasons, you must maintain a former static monitoring design, but you are also given resources for additional monitoring. In this case, how many extra monitoring locations do you need, and are there any restrictions to where you put them?

These are all questions that do not have easy answers and will be specific to each individual project. Experts on the study system can help you address many of them, and simulation can be used, to some extent, for providing a general idea of how the decisions may play out in altering monitoring strategies and inference. Also, depending on the novelty in the modeling and monitoring methods, there is a rapidly growing body of literature that can provide guidance.

Service 1987). This survey has been conducted every May through June since 1955 using aerial transects covering northern portions of North America (Smith 1995). Pilots and observers record each duck species observed, and whether or not the ducks are paired (with a mate), single drakes, or in mixed-sex groups. Surveys are flown at approximately 120 miles per hour and an altitude of 90–100 feet.

To illustrate the utility of spatio-temporal modeling in the context of long-term monitoring as well as the process of optimizing a monitoring design, we make use of these data to (i) estimate spatially explicit trends in scaup breeding pairs and identify significant increasing and decreasing regions; and (ii) obtain an adaptive monitoring design for future observation periods that allows for optimal trend characterization in budget-limited scenarios. The second objective is addressed as an example to illustrate the savings that can be achieved when using an approach that explicitly couples monitoring and modeling.

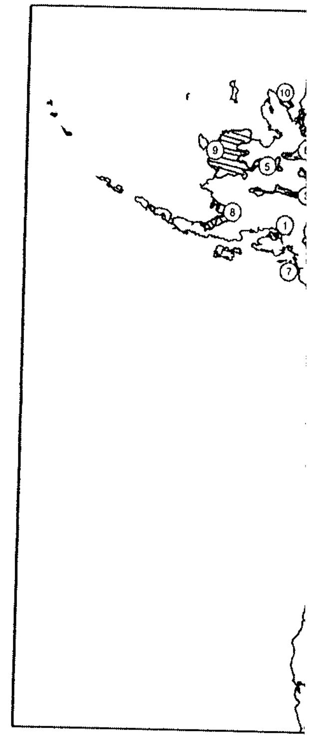

Figure 20.1 Aerial s
(USFWS 2009a).

Developing an

For the scaup exa
ify a model for b
three modeling s
manner, we can u
the process and p
determines the sp
eters are trend co
can then be used t
funding) before n
et al. (2007) con
latter study was f
model componen
model, where the
cess to have muc
regression that al
some of the diffic
domains. In light

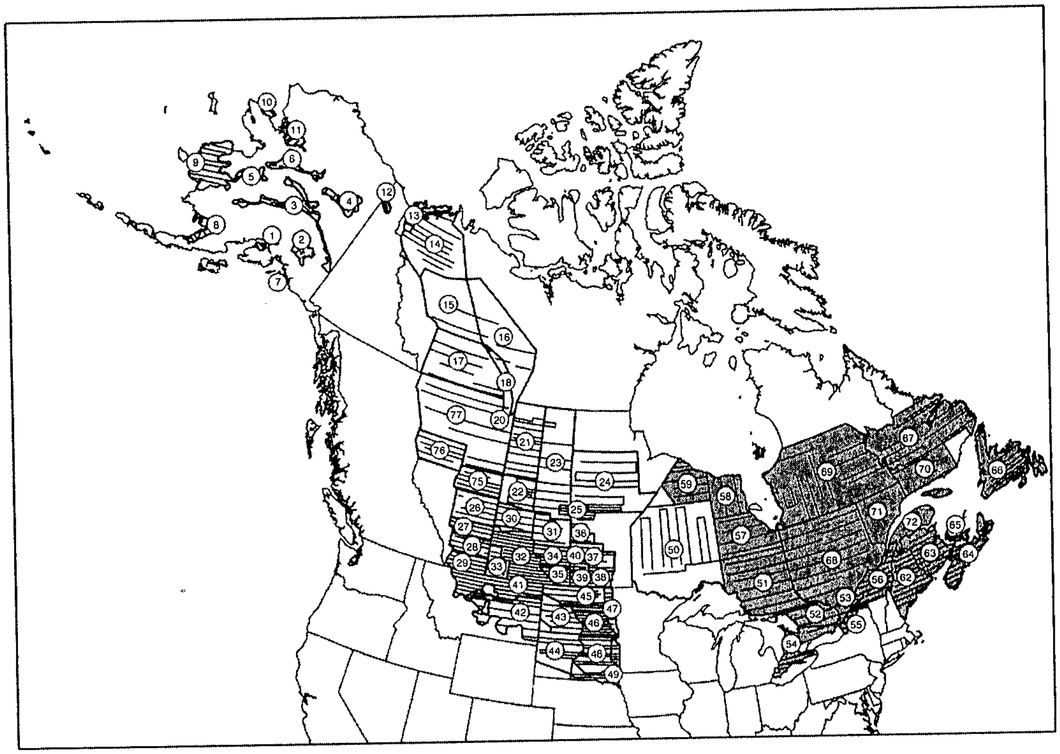

Figure 20.1 Aerial survey units (strata) in the Waterfowl Breeding Population and Habitat Survey (USFWS 2009a).

Developing and fitting the spatio-temporal model

For the scaup example, following the general approach of Gardner *et al.* (2007), we specify a model for breeding pair trend analysis using a hierarchical specification involving three modeling stages: data, process, and parameter. By setting up the model in this manner, we can use an algorithm to approximate the intractable posterior distribution of the process and parameters given the data (Berliner 1996). In our case, the latent process determines the spatio-temporal patterns in mean breeding pairs, while the model parameters are trend coefficients and variance components. We show that these latter quantities can then be used to construct optimal strategies for conserving monitoring resources (e.g. funding) before new data have been collected. Afton and Anderson (2001) and Gardner *et al.* (2007) constructed similar models for analyzing scaup trends, and although the latter study was focused on total scaup from years 1975 to 2005, many of the general model components they describe can be found in our specification. In constructing our model, where the focus is on long-term scaup breeding pairs, we allow our latent process to have much coarser spatial units (i.e. strata; Fig. 20.1) and specify a temporal regression that allows the strata to be correlated. Gardner *et al.* (2007) also described some of the difficulties in Bayesian analyses when using such large data sets and spatial domains. In light of this, we use a computational approach based on integrated nested

Laplace approximations (INLA; Rue *et al.* 2009), rather than the more common Markov chain Monte Carlo (MCMC) methods.

Data model

In order to accommodate potentially overdispersed count data, we specify a negative binomial distribution for the breeding pair observations: $y_{i,j,t} \sim \text{NegBinom}(\mu_{j,t}, \phi)$ for segments $i = 1, \ldots, n_j$ in stratum $j = 1, \ldots, m$ during observation period $t = 1, \ldots T$ (i.e. years 1957–2009). In this specification, $E(y_{i,j,t}) = \mu_{j,t}$. Two alternatives to this data model are the Poisson and quasi-Poisson distributions (e.g. Ver Hoef and Boveng 2007). Although the Poisson model is more parsimonious and conserves degrees of freedom for the estimation of other model parameters, the quasi-Poisson and negative binomial are more general and can better model overdispersed count data. Based on an exploratory analysis and model comparison for this study, we employ the negative binomial instead of the Poisson data model as other studies have done (e.g. Gardner *et al.* 2007). We found that the overdispersion parameter (ϕ) allows for a better fit to the data as indicated by a lower deviance information criterion (i.e. DIC) in our preliminary model fits.

Process model

The traditional negative binomial generalized linear model employs a log link function to connect the count means ($\mu_{j,t}$) to a set of covariates (Chapter 11). In our case then, introducing new variables, we specify a log-linear process model:

$$z_{j,t} = \log(\mu_{j,t}) = \beta_{0,j} + \beta_{1,j}t + \varepsilon_{j,t} + \eta_{j,t}, \tag{20.1}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{m,t})' \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ for all $t = 1, \ldots T$ represents spatially correlated errors and $\boldsymbol{\eta}_j = (\eta_{j,1}, \ldots, \eta_{j,T})' \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ for all $j = 1, \ldots m$ represents temporally correlated errors. This linear model, operating in the log-space of the mean breeding pairs, provides for fairly simple inference on the general trend of breeding pairs by stratum and allows for nonlinear growth on the natural support of the counts themselves. An alternative process specification could use a discrete dynamic model that would allow for the estimation of growth and carrying capacity parameters, $z_{j,t} = f(z_{j,t-1}, r_j, k_j) + \varepsilon_{j,t}$, where the function f, for example, could be in the form of Ricker or Beverton–Holt density-dependent growth (Turchin 2003). For examples of these latter forms of model specifications, see Boomer and Johnson (2007), Hooten *et al.* (2007), Hooten *et al.* (2009b), and Wikle and Hooten (2010).

Ideally, the correlated fields, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_j$, in the given model specification, will capture unexplained spatial and temporal dependence in the process not due to the general breeding pair trends. In general, accommodating these sources of uncertainty (i.e. spatial and temporal correlated random effects) in the model is critical because they allow for valid inference on the remaining parameters (i.e. $\boldsymbol{\beta}$).

The strata in this example are irregularly shaped, sized, and spaced spatial regions (Fig. 20.1), thus we need to be cautious in specifying an appropriate form of spatial dependence in the model (i.e. $\boldsymbol{\Sigma}_\varepsilon$). A commonly used form of dependence for areal processes is referred to as a conditional autoregressive structure (CAR; Brook 1964, Besag 1974). CAR models allow the spatial process (i.e. ε) to depend on itself based on the neighborhood structure of the spatial regions or strata. Assuming a CAR spatial structure, one can then write: $\varepsilon_{j,t}|\{\varepsilon_{k,t}, \forall k \neq j\} \sim \mathrm{N}(\sum_k w_{j,k}\varepsilon_{k,t}, \sigma^2)$ where the $w_{j,k}$ describe the neighborhood connectivity of stratum j and stratum k. The simplest specification would be to let $w_{k,j} = 1$ if stratum k is a neighbor of stratum j, and zero otherwise. The model is then referred to as an Intrinsic-CAR (ICAR) model if we write the distribution of the entire spatial field as $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, where $\boldsymbol{\Sigma}_\varepsilon \equiv \sigma_\varepsilon^2(\mathbf{D} - \mathbf{W})^{-1}$, $\mathbf{W}$ is a matrix containing all of the $w_{j,k}$, and the matrix $\mathbf{D}$ is diagonal with the row sums of $\mathbf{W}$ as diagonal elements (Banerjee *et al.* 2004).

In a Bayesian framework, the ICAR model serves as a prior for the spatial fields $\boldsymbol{\varepsilon}_t$ and is not technically proper in that it will not integrate to 1. We note that, under some relatively mild conditions, the posterior distribution for ε will indeed integrate to 1 as required (Rue and Held 2005).

We also need to specify a model that will accommodate potential temporal dependence in the $\boldsymbol{\eta}_j$. Here, we assume a time-series version of the CAR model described above where we let $\eta_{j,t} \sim \mathrm{N}(\alpha\eta_{j,t-1}, \sigma_\eta^2)$. This model induces a correlation structure on the vectors $\boldsymbol{\eta}_j$ such that $\boldsymbol{\eta}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$. Now, due to our careful specification, the entire latent process model for $z_{j,t}$ [(Equation (20.1)] is a Gaussian Markov random field; this can be advantageous when fitting the model to observed data.

Parameter model

The hierarchical model specified above contains $2m + 4$ parameters that need prior distributions to complete the model formulation. The overdispersion parameter is specified as $\phi = \log(n)$, where n is the original negative binomial size parameter, and then modeled as $\phi \sim \mathrm{N}(0, 100)$. We use exchangeable conjugate Gaussian priors for the m sets of regression parameters: $\beta_{0,j}, \beta_{1,j} \sim \mathrm{N}(0, 1000)$, for $j = 1, \ldots, m$. We used a conjugate inverse gamma prior for the variance component: $\sigma_\varepsilon^{-2} \sim \mathrm{Gamma}(1, 1/20\,000)$. Finally, we have the parameters in the temporal autocorrelation model. We reparameterize these such that $\theta_1 = (1 - \alpha^2)/\sigma_\eta^2$ and $\theta_2 = (1 + \alpha)/(1 - \alpha)$ and then assign the following priors: $\theta_1 \sim \mathrm{Gamma}(1, 1/20\,000)$ and $\theta_2 \sim \mathrm{N}(0, 5)$. An alternative set of priors could be specified for the precision and autoregressive parameters directly, although the reparameterization described above is preferred for the form of implementation we describe next (Rue and Martino 2009).

Implementation

At this point, we are able to make inference on the underlying process and parameters by finding their posterior distribution conditioned on the observed data:

$$\left[\{\boldsymbol{\mu}_t\}, \phi, \{\boldsymbol{\beta}_t\}, \sigma_\varepsilon^2, \sigma_\eta^2, \alpha | \{y_{i,j,t}\}\right] \propto \prod_i \prod_j \prod_t [y_{i,j,t}|\mu_{j,t}, \phi] \prod_t \left[\boldsymbol{\mu}_t | \boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\eta^2, \alpha\right]$$
$$\times [\phi]\,[\boldsymbol{\beta}]\,[\alpha]\,\left[\sigma_\varepsilon^2\right]\,\left[\sigma_\eta^2\right]. \qquad (20.2)$$

Note, in the above equation, we use a conventional Bayesian notation where the square brackets " $[\cdot]$ " refer to a probability distribution. Fitting the above model is relatively trivial and can be accomplished using an MCMC algorithm to sample from each of the full-conditional distributions of the process and parameters. Software packages such as BUGS (Lunn *et al.* 2009) or JAGS (Plummer 2003) could be employed to do this and Gardner *et al.* (2007) provide computer code to implement a similar MCMC algorithm.

A preliminary simulation study indicated that MCMC methods can be quite slow for these large data sets with correlated spatial effects, as suggested by Gardner *et al.* (2007). Thus, we employ a new method for fitting Bayesian Markov random field models called INLA, an acronym standing for "integrated nested Laplace approximations" (Rue *et al.* 2009). The INLA algorithm is not an iterative stochastic method like MCMC, but rather a sequence of techniques to approximate the marginal posterior distributions for well-posed latent Gaussian models (like the one we have specified in the preceding sections) with a closed-form expression. It has been shown to be very accurate (Rue *et al.* 2009), and although not as flexible as MCMC, it is generally much faster for a specific class of models. A version of INLA has been implemented in R (R Development Core Team 2011) and is highly accessible.

The INLA algorithm proceeds by approximating the marginal posterior distributions for each unknown Gaussian variable of interest given the data. This approximation is accomplished by integrating the "hyperparameters" (i.e. all non-Gaussian unknown model parameters) out of the joint posterior distribution. In practice, the required integration is intractable but since the integrand can be factored into a full-conditional distribution of the Gaussian parameters and a marginal posterior that can be well-approximated by either a Gaussian or Laplace distribution, it is possible to carry out the calculation numerically. Depending on the desired inference, a potential disadvantage of the INLA approach to fitting hierarchical Bayesian models is that we are restricted to marginal and linear combinations of marginal posterior distributions. However, many studies only require this form of inference and thus INLA could be widely used in spatio-temporal analysis of ecological data.

Model fit and parameter estimation

We used the "INLA" R package (Rue and Martino 2009) to fit the model described above and obtain parameter estimates. We provide a basic outline for the INLA syntax required to fit the model described herein. Although this package is very easy to use as demonstrated below, we caution that, as with automatic MCMC algorithms, care needs to be taken to ensure the correct model is being specified and fit. In this case, after loading the INLA package and having the data already organized, one can simply issue the following command:

```
inla(y ~ X + f(j,model="besag",graph.file=W,replicate=t)
+f(t,model="ar1",replicate=j), model="nbinomial", E=E)
```

where $\mathbf{y}$ represents the summed counts by stratum for all years, $\mathbf{X}$ is a "design" matrix consisting of the entire set of submatrices $\{\mathbf{I}, t \cdot \mathbf{I}\}$ for all $t = 1, \ldots, T$ (where $\mathbf{I}$ = the identity matrix), and the 2 $\mathbf{f}$ model components correspond to the spatial and temporal

notation where the square
above model is relatively
sample from each of the
oftware packages such as
employed to do this and
imilar MCMC algorithm.
ods can be quite slow for
by Gardner *et al.* (2007).
ndom field models called
proximations" (Rue *et al.*
d like MCMC, but rather
or distributions for well-
n the preceding sections)
ccurate (Rue *et al.* 2009),
faster for a specific class
Development Core Team

al posterior distributions
ata. This approximation
non-Gaussian unknown
actice, the required inte-
d into a full-conditional
terior that can be well-
possible to carry out the
a potential disadvantage
is that we are restricted
ibutions. However, many
could be widely used in

fit the model described
ine for the INLA syntax
ge is very easy to use as
C algorithms, care needs
d fit. In this case, after
ed, one can simply issue

```
licate=t)
, E=E)
```

, **X** is a "design" matrix
1, ..., T (where **I** = the
the spatial and temporal

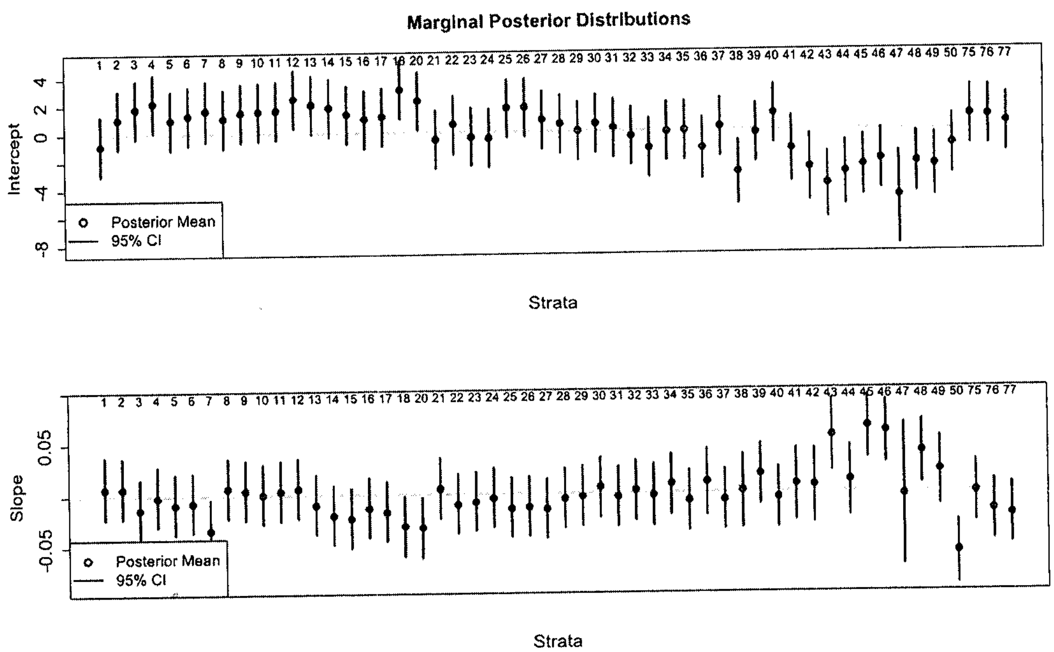

Figure 20.2 Marginal posterior means and 95% credible intervals for each of the strata intercept and slope parameters for the scaup trend analysis example.

autocorrelated errors, respectively. The graph file **W** specifies the spatial connectivity among strata, "nbinomial" specifies the data model to be used, and **E** corresponds to a known scaling factor in $\mu_{j,t}$; in this case, the scaling factor adjusts for the differing numbers of segments per stratum. The "replicate" option in the latent error models (**f**) specifies over which index the hyperparameters should be shared.

In fitting the model to scaup data, we focused on the 52 western-most strata (Fig. 20.1), containing a total of 103 266 scaup counts in these strata over a period of 53 years (1957–2009). The fitting algorithm took approximately 27 minutes on a 2 × 2.93 GHz 6-Core Intel Xeon workstation. Based on preliminary comparisons with an MCMC algorithm, the INLA algorithm appears to be an order of magnitude faster than MCMC for this model and data set.

In constructing the proximity matrix **W**, we first created a Euclidean distance matrix between all strata centroids and then denoted all pairs of strata as neighbors if they were within a threshold distance of 7.75 decimal degrees of each other. This number represented the minimum distance at which all strata had at least one neighbor. Because some of the northern strata are more isolated than the southern strata, we felt that this connectivity should be expressed in the proximity matrix.

In terms of model results, the posterior distributions for the model coefficients are of primary interest (Fig. 20.2). Focusing on the marginal posterior credible intervals for the slope parameters in the strata trends (lower plot in Fig. 20.2), these results indicate that we do not have enough evidence to detect a log-linear trend in many of the strata (i.e. the credible intervals overlap zero); however, some (e.g. strata 46 and 50) do

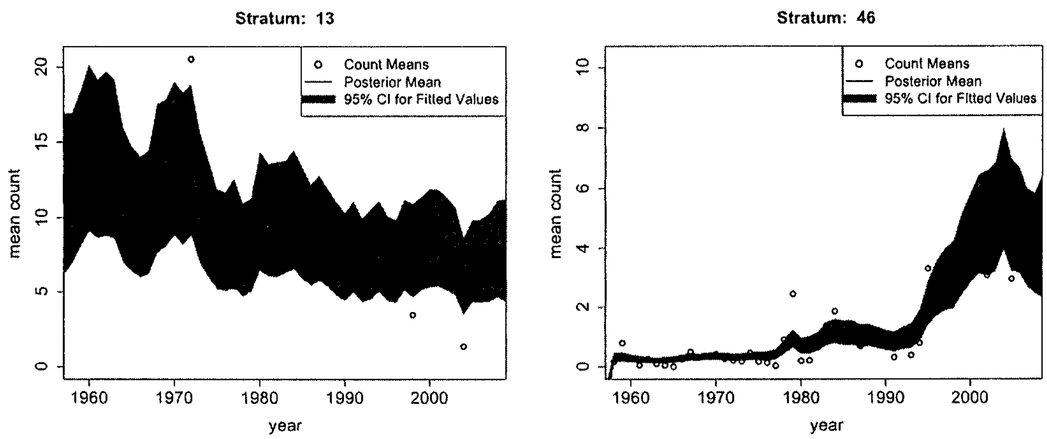

Figure 20.3 Posterior means and 95% credible intervals for scaup breeding pairs in two example strata (13: McKenzie River Delta and 46: Drift Prairie and Missouri Coteau).

indicate both positive and negative trends. Even though it is not immediately obvious, one of the advantages of being able to accommodate latent forms of dependence in our hierarchical model is that we reduce the chance of making Type I errors in our inference, as discussed in Chapter 11. That is, a failure to appropriately account for latent forms of dependence can lead to a reduction in posterior variance and the incorrect detection of non-significant trends. In fact, in comparison with a preliminary study that used a simpler, uncorrelated, latent model (not included here), we see a decided increase in credible interval width pertaining to the marginal posterior distributions for the trend coefficients when considering the spatial and temporal autocorrelation.

As developed below, the optimal design depends most on the estimated variance component σ_ε^2. The INLA algorithms actually model the precision (i.e. $\frac{1}{\sigma_\varepsilon^2}$) instead of the variance, thus, the marginal posterior mean and standard deviation for the precision parameter are 190.4 and 66.1, respectively. The posterior mean and standard deviation for the temporal model precision were 26.6 and 3.3, while the autoregressive parameter α had a posterior mean and variance of 0.93 and 0.02, respectively. Finally, the overdispersion parameter in the negative binomial likelihood, if parameterized in terms of the "size" (n), had a posterior mean and standard deviation of 7.4 and 0.38.

In addition to parameter estimates, it is helpful to connect the posterior parameter distributions back to the process under study. In this case, we are modeling scaup breeding pairs, thus to illustrate two potential scenarios in abundance trends, we highlight model predictions for stratum 13 and stratum 46 (Fig. 20.3), which are decreasing and increasing in abundance, respectively.

Defining and selecting an optimal monitoring design

Specifying the design criterion and implementation

Using the scaup example and the spatio-temporal model developed above, we next illustrate the process of optimizing a monitoring design (see also Box 20.2). In terms of choosing an optimal monitoring strategy for a future observation period $(T+1)$ when available resources may be a limiting factor, in the scaup example we would need to select a subset ($\boldsymbol{S}_{T+1}$ with dimension m_{T+1}) of the total strata ($\boldsymbol{S}$ with dimension m, Fig. 20.1) that maximizes the information we receive from the data while minimizing the uncertainty concerning our inference on the breeding pair trends. Thus, we want to construct a sampling design for observation period $T+1$ that allows us to minimize the posterior variance of the coefficients $\boldsymbol{\beta}$ while monitoring only a subset of total strata.

Consider, then, a subset of strata $\boldsymbol{S}_{T+1} \subseteq \boldsymbol{S}$ where $m_{T+1} \leq m$, being the resource limited monitoring scenario during the observation period $T+1$. This implies that we can only obtain additional information, during observation period $T+1$, about the process at the subset of strata [$\boldsymbol{\mu}(\boldsymbol{S}_{T+1})$, or $\mathbf{z}(\boldsymbol{S}_{T+1})$ in log space]. We then specify the design criterion as a measure of total posterior parameter variance given the new design $\boldsymbol{S}_{T+1}$:

$$q(\boldsymbol{S}_{T+1}) = \mathrm{tr}(\mathrm{var}(\boldsymbol{\beta}|\mathbf{z}(\boldsymbol{S}_{T+1}), \cdot)). \tag{20.3}$$

That is, the design criterion is the sum, across all retained strata in the new design, of the posterior variances of each stratum's coefficients in the log-linear process (trend) model specified by Equation (20.1). The full conditional random parameters $(\boldsymbol{\beta}|\mathbf{z}(\boldsymbol{S}_{T+1}), \cdot)$ have a distribution whose variance can be found analytically. To see this, we write the full process model for all $\boldsymbol{\mu}_t$, where $\boldsymbol{\mu} \equiv (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T)'$ and $\mathbf{z} \equiv \log(\boldsymbol{\mu})$. Now, we augment the process vector $\mathbf{z}$ with the process in a future monitoring period, i.e. $\mathbf{z}_{T+1} \equiv \mathbf{z}(\boldsymbol{S}_{T+1})$, such that $\mathbf{z}_+ = (\mathbf{z}', \mathbf{z}'_{T+1})'$ and write the full process specification as a linear model: $\mathbf{z}_+ = \mathbf{X}_+\boldsymbol{\beta} + \boldsymbol{\varepsilon}_+ + \boldsymbol{\eta}_+$. In this formulation, $\mathbf{X}_+ = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_T, \mathbf{X}'_{T+1})'$, where each $\mathbf{X}_t$ is $m \times 2m$ and contains an identity matrix $\mathbf{I}$ in the first m columns and the remainder of columns equals $t \cdot \mathbf{I}$. For time $T+1$, the matrix $\mathbf{X}_{T+1}$ is only $m_{T+1} \times 2m$ dimensional since we are assuming only a subset of the process components will be able to influence the posterior distribution of $\boldsymbol{\beta}$ which contains the global intercept and trend coefficients. The final terms, $\boldsymbol{\varepsilon}_+$ and $\boldsymbol{\eta}_+$, represent the spatially and temporally correlated error for all locations and times. Using this augmented model specification, the full conditional distribution for $\boldsymbol{\beta}$ is proportional to the product of the augmented process model and the prior:

$$\begin{aligned}
[\boldsymbol{\beta}|\mathbf{z}(\boldsymbol{S}_{T+1}), \cdot] &\propto [\mathbf{z}_+|\boldsymbol{\beta}, \sigma^2][\boldsymbol{\beta}] \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{z}_+ - \mathbf{X}_+\boldsymbol{\beta})'\boldsymbol{\Sigma}_+^{-1}(\mathbf{z}_+ - \mathbf{X}_+\boldsymbol{\beta})\right) \times \exp\left(-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}\right) \\
&= N(\mathbf{A}^{-1}\mathbf{b}, \mathbf{A}^{-1})
\end{aligned}$$

where the covariance matrix $\boldsymbol{\Sigma}_+ = (\mathbf{I} \otimes \boldsymbol{\Sigma}_\varepsilon + \boldsymbol{\Sigma}_\eta \otimes \mathbf{I})$ and $\mathbf{A}^{-1} = (\mathbf{X}'_+\boldsymbol{\Sigma}_+^{-1}\mathbf{X}_+ + \boldsymbol{\Sigma}_\beta^{-1})^{-1}$. Thus, the design criterion $q(\boldsymbol{S}_{T+1})$ is not a function of the future process $\mathbf{z}(\boldsymbol{S}_{T+1})$, but rather, only a function of the observation periods t, the design $\boldsymbol{S}_{T+1}$ (which is embedded in the augmented matrix $\mathbf{X}_+$ and the spatio-temporal covariance matrix $\boldsymbol{\Sigma}_+$), and the prior covariance matrix $\boldsymbol{\Sigma}_\beta$.

If an MCMC algorithm is chosen to fit the model, then the implementation of the optimal design for period $T+1$ within the algorithm proceeds by first sampling from each of the parameter full-conditional distributions, and then finding the design $\boldsymbol{S}_{T+1}$ that minimizes the criterion $q(S_{T+1})$ on each MCMC iteration. In this scenario, the optimal design itself is stochastic with a posterior distribution indicating which subsampling scheme in observation period $T+1$ provides the greatest probability for a reduction of posterior parameter variance. In terms of settling on a fixed optimal design, one would then select an a priori dimension (i.e. m_{T+1}) for the optimal subsample of strata, and then identify the set of strata that maximize the posterior design distribution.

The implementation process outlined above requires that model fitting and finding the optimal retrospective design occur in a single analysis. The process will be inefficient if the space of potential designs is large (because of the nested maximization loop within the MCMC algorithm), and moreover, could not be implemented within the INLA approach directly since the full-conditionals are never directly sampled from. Therefore, we suggest using an alternative quasi-criterion that relies on the posterior mean covariance of the process error, $\mathrm{E}(\boldsymbol{\Sigma}_{+}|\mathbf{y})$:

$$q(S_{T+1}) = \mathrm{tr}\left(\left(\mathbf{X}'_{+}\mathrm{E}(\boldsymbol{\Sigma}_{+}|\mathbf{y})^{-1}\mathbf{X}_{+} + \boldsymbol{\Sigma}_{\beta}^{-1}\right)^{-1}\right). \tag{20.4}$$

With this new criterion, we can now fit the model (as described earlier) and find the optimal retrospective sampling design separately. The use of the INLA software and the decoupling of these two steps in the analysis make for a fast and easily modifiable approach for finding model-based optimal designs, and the results that follow provide an illustrative example of how these adaptive sampling approaches could be used to conserve monitoring resources.

An important point to consider when seeking optimal designs is that there are an infinite number of different types of optimality. That is, because we optimize with respect to the design criterion, its form controls the type of design we will end up with. The form we have chosen [Equation (20.4)] is referred to as "A-optimality", and minimizes the average variance of the coefficients. A-optimal designs are commonly used in the literature (i.e. Wikle and Royle 2005, Hooten *et al.* 2009a), but many other options exist. For example, two other common forms involving the covariance of the coefficients are "E-optimality" and "D-optimality", minimizing the maximum (via the spectral radius or maximum eigenvalue) and generalized variance (via the determinant), respectively (Gentle 2007). Yet another approach to finding optimal designs involves optimizing over the entropy of the distribution containing the design (e.g. Le and Zidek 2006). Each of these different choices for design criteria has advantages and disadvantages, and moreover, will most certainly influence the outcome; although, it has been argued that any type of well thought out design criterion will result in more efficient data collection and conservation of resources.

Optimal monitoring design for scaup

Based on the results of the model fit for the current scaup data set, we searched the space of designs twice, using 10 000 random designs each time, for the monitoring year 2010 over a range of strata subsample sizes (m_{T+1}) from 5 to 51, with 5 being the most restrictive case in terms of limited resources for future sampling and 51 being the best case scenario. We used the case where all 52 strata were hypothetically monitored in year 2010 as a baseline to compare the design criterion and proportion of budget across all possible subsample sizes.

In situations where there is large variation in the optimal design search results, a further refinement of the optimality is recommended. Various exchange algorithms (e.g. Nychka and Saltzman 1998) can be employed in these cases to test a sequence of swaps between neighboring strata for further increases in quasi-optimality. In this study, the resulting optimal designs from each set of searches were similar enough to combine results and forego further measures to increase quasi-optimality. That is, in the few cases where the optimal design differed between our two searches, we retained the one that yielded a smaller $q(\boldsymbol{S}_{T+1})$.

The first obvious result to report is the design criterion $q(\boldsymbol{S}_{T+1})$ and how it changes with increasing subsample size m_{T+1}. Figure 20.4a illustrates the gain in mean coefficient variance, over that of a random (i.e. non-optimal) subsampling scheme, with increasing future monitoring sample size. In this case, as the potential future subsample becomes more and more restricted, we see an exponential reduction in parameter uncertainty (i.e. an increase in the difference of criteria between random subsampling and optimal subsampling). That is, at smaller future subsample sizes, the benefits of optimal monitoring increase dramatically.

It is unlikely that a resource-limited situation would arise in such a way that the actual number of strata to be sampled is the limiting quantity. The more likely scenario is a limiting quantity based on available budget. As such, it is first important to be able to view the effect on budget by calculating the actual dollars saved when using the optimal design versus a random subsample of strata. Figure 20.4b expresses how these savings fluctuate depending on the number of strata retained in the optimal design. In this study, dollar amounts are calculated based on an average cost of \$246 per segment monitored. Notice that, in some cases, even though the number of strata are being reduced, there is actually an increase in cost for the optimal design as compared with a random design. This is due to the fact that some of the strata contain many more segments than others, and thus the optimal designs for those scenarios contain more total segments.

Under a restricted budget, we can see from Figure 20.4c that, although the budget generally declines with a decreasing number of strata retained in the optimal design, there are distinct optimal design sizes where the budget is reduced more than would be expected in a purely linear relationship between budget and sample size. For example, from Figure 20.4b it is clear that the savings are extraordinary when retaining 22 strata (approx. \$80 000) in the monitoring design as opposed to retaining 21 strata, where it only saves \$20 000 to monitor optimally. Although both designs save money, in this specific case, with the scaup breeding pair trends, it actually saves more to go with a larger sample size of 22 over the optimal design for the smaller sample size of 21, given that they both yield nearly the same reduction of uncertainty (Figures 20.4a, c).

We can also conclude, from Figure 20.4, that optimal monitoring schemes, in this breeding pair study, are most effective when the number of strata and/or budget is reduced to 50% and lower. Although optimal designs are able to decrease uncertainty at every level of budget restriction, the difference between random subsampling and optimal subsampling increases substantially at the smaller subsample sizes.

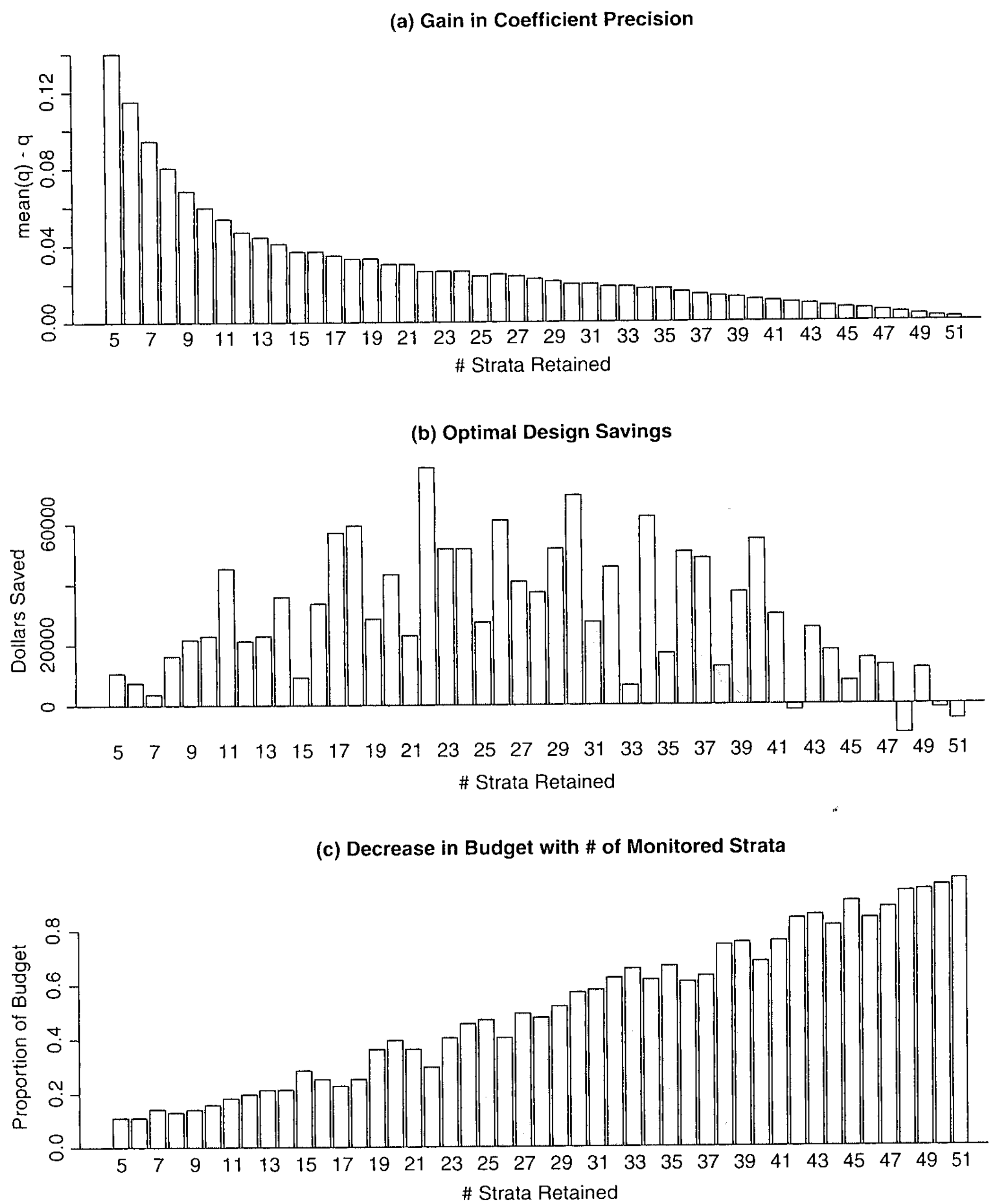

Figure 20.4 Optimal monitoring results for the scaup monitoring example. (a) The mean optimality criterion under random subsampling minus the criterion under optimal subsampling. A positive difference in this setting indicates a gain in parameter precision under optimal sampling. (b) Dollars saved by number of strata retained in each optimal design. (c) The cost of optimal monitoring as a proportion of maximum budget for a given number of retained strata.

Table 20.1 Opt
budget-restricti
the full budget
the scaup exam
subsample size
parameter unce
restriction, alon

Available budget (%)
50
60
70
80
90

Given the
the optimal
only two ste
cost less tha
provides the
criterion q).
Based on
optimal des
(Table 20.1
are proporti
Cost of the
the budget

Future res

Overall, th
monitoring
about the tr
can see that
when using
This appro
but, to our
the study o
More ge
sampling d
model with
example is
in computa

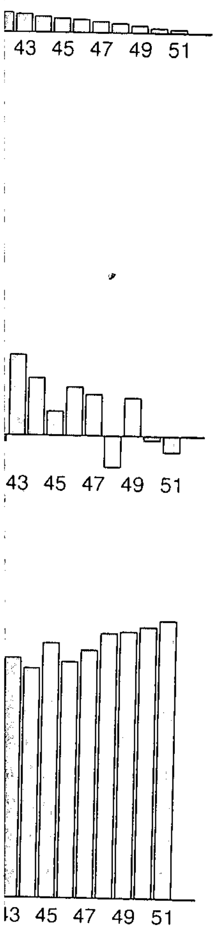

Table 20.1 Optimal monitoring with respect to budget-restriction scenarios (expressed as percentage of the full budget needed to monitor all 52 possible strata) for the scaup example. Results shown are the optimal subsample size m_{T+1} for the monitoring design that reduces parameter uncertainty the most at each level of budget restriction, along with the actual cost of the optimal sample.

Available budget (%)	Optimal m_{T+1} (% of 52 total strata)	Actual cost (%)
50	28 (54%)	\$306 316 (48%)
60	31 (60%)	\$372 607 (58%)
70	40 (77%)	\$435 940 (68%)
80	41 (79%)	\$487 444 (76%)
90	47 (90%)	\$562 114 (88%)

Given the output from this optimal design study, it is a relatively simple task to find the optimal design given a specified budget restriction. The procedure to do this requires only two steps. First, for each optimal subsample size, identify the optimal designs that cost less than the available budget. Second, with these new designs, keep the one that provides the greatest reduction in mean coefficient variance (i.e. the smallest design criterion q).

Based on our scaup breeding pair analysis, we used this procedure to find the best optimal designs for each 10% increment of budget proportion from 50% to 90% (Table 20.1). These findings indicate that the number of strata to optimally monitor are proportionally at least as large as the available percentage budget and that the actual cost of the optimal monitoring is below the available budget, even sometimes reducing the budget by as much as an additional 4%, as in the case of an 80% available budget.

Future research and development

Overall, there is substantial evidence supporting the use of a model-based optimal monitoring design in the assessment of scaup breeding pairs. Although uncertainty about the trend parameters increases in general with a decreased monitoring effort, we can see that there is a distinct gain in coefficient precision (i.e. reduction in uncertainty) when using an optimal design as opposed to a random subsampling design (Fig. 20.4). This approach to optimal design, in general, has enjoyed much success in other fields, but, to our knowledge, this is the first time it has been utilized to conserve resources in the study of waterfowl breeding pair trends.

More generally, we have presented an approach for constructing adaptive optimal sampling designs for a spatio-temporal ecological process using a relatively simple trend model within a Bayesian framework. The scaup breeding pair data set that we used as an example is fairly large (at over 100 000 records) and we found significant improvement in computational efficiency using the integrated nested Laplace approximation approach described by Rue *et al.* (2009). For hierarchical models that contain a latent Gaussian Markov random field process (like the one presented herein), the INLA methods can be utilized and have been shown to be quite accurate. In our specific example, we focused on the situation where resources for monitoring may be limited in future observation periods and show that significant savings can be obtained by exploiting the natural spatial dependence in the process under study.

The methods and software we have used in this study are very accessible and could easily be modified for use with other species, design objectives, or limiting resource constraints. A few examples of potential modifications and extensions to the approach we present here include the following.

- Use of a likelihood that explicitly accounts for imperfect detectability: although our negative binomial allows for overdispersed counts, it could be generalized to explicitly accommodate an observation bias, as in the hierarchical model outlined in Chapter 19.
- Different forms for spatial and temporal dependence (e.g. geostatistical spatial model or higher-order time-series model): while the use of a continuous spatial model for latent autocorrelation would prohibit us from using INLA to fit the model, it also makes MCMC methods more feasible for these problems.
- More scientifically motivated dynamic process model (e.g. Wikle and Hooten 2010): specific physically based process models, if appropriate, can provide better predictions and forecasts as well as intuitive scientific learning about the process under study.
- Different design objective (e.g. prediction error variance): if our focus was on forecasting, we may want to use a design criterion that represents prediction uncertainty.
- Consideration of more complicated forms of incurred cost: in our scaup example, we assumed a constant per-segment rate; however, the actual aerial survey is much more complicated than that, and it may be beneficial to consider factors like actual flight paths and re-fueling stops.
- Planning or evaluation process of long-term survey development: we described a framework for acute optimality in terms of desired short-term inference; however, the model and/or design criterion could be modified to consider objectives with a greater temporal extent. An example might be a combined objective including inference on long-term population trends in addition to acute management decisions such as setting regional harvest limits.

Summary

We presented a Bayesian hierarchical modeling framework for optimally monitoring population trends while accounting for overdispersion in count data and spatio-temporal dependence. We provide an example pertaining to scaup, using breeding pair counts from the North American aerial waterfowl survey. Specifically, the dynamic process is modeled as a Markov random field, where scaup relative abundance in each stratum in the survey increases or decreases with respect to a general trend, and is subject

to additional tem
remain unexplain
correlation in sca
construct an opti
strategy we prese
the available moi
the methods we
environmental pr
it requires upfron
term benefits of p
a worthwhile end

Acknowledgm

The authors thank
and Håvard Rue f
authors also ackn
the support of N
purposes only an

to additional temporal and spatial dependence due to similarities between strata that remain unexplained by the trend alone. The understanding of these trends and residual correlation in scaup abundance allows us to exploit the dependence in the process and construct an optimal adaptive monitoring strategy for future observation periods. The strategy we present is meant as an example only, and is focused on the scenario where the available monitoring budget is reduced during future monitoring efforts. Overall, the methods we present could be applied to nearly any spatio-temporal ecological or environmental process where inference on general trends is of primary interest. Although it requires upfront effort to develop a model-based optimal monitoring program, the long-term benefits of potential cost reduction and increased learning can definitely make this a worthwhile endeavor.

Acknowledgments

The authors thank David Koons, Beth Gardner, Andy Royle, Scott Boomer, Sara Martino, and Håvard Rue for providing data and helpful suggestions pertaining to this project. The authors also acknowledge the support of Delta Waterfowl and C. Wikle acknowledges the support of NSF grant OCE-0814934. Any use of trade names is for descriptive purposes only and does not imply endorsement by the US Government.