Utah State University

DigitalCommons@USU

All Graduate Plan B and other Reports   Graduate Studies

5-2012

Assessing Changes in the Abundance of the Continental Population of Scaup Using a Hierarchical Spatio-Temporal Model

Beth E. Ross
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Mathematics Commons, and the Statistics and Probability Commons

UtahState University
MERRILL-CAZIER LIBRARY

ASSESSING CHANGES IN THE ABUNDANCE OF THE CONTINENTAL

POPULATION OF SCAUP USING A HIERARCHICAL SPATIO-TEMPORAL

MODEL

by

Beth E. Ross

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Statistics

Approved:

_____   _____
Dr. Mevin Hooten   Dr. David Koons
Major Professor   Committee Member

_____
Dr. James Powell
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2012

ABSTRACT

Assessing Changes in the Abundance of the Continental Population of Scaup Using

a Hierarchical Spatio-Temporal Model

by

Beth E. Ross, Master of Science

Utah State University, 2012

Major Professor: Dr. Mevin Hooten
Department: Mathematics and Statistics

In ecological studies, the goal is often to describe and gain further insight into ecological processes underlying the data collected during observational studies. Because of the nature of observational data, it can often be difficult to separate the variation in the data from the underlying process or 'state dynamics.' In order to better address this issue, it is becoming increasingly common for researchers to use hierarchical models. Hierarchical spatial, temporal, and spatio-temporal models allow for the simultaneous modeling of both first and second order processes, thus accounting for underlying autocorrelation in the system while still providing insight into overall spatial and temporal pattern. In this particular study, I use two species of interest, the lesser and greater scaup (*Aythya affinis* and *Aythya marila*), as an example of how hierarchical models can be utilized in wildlife management studies. Scaup are the most abundant and widespread diving duck in North America, and are important game species. Since 1978, the continental population of scaup has declined to levels that are 16% below the 1955-2010 average and 34% below the North American Waterfowl Management Plan goal. The greatest decline in abundance of

scaup appears to be occurring in the western boreal forest, where populations may have depressed rates of reproductive success, survival, or both. In order to better understand the causes of the decline, and better understand the biology of scaup in general, a level of high importance has been placed on retrospective analyses that determine the spatial and temporal changes in population abundance. In order to implement Bayesian hierarchical models, I used a method called Integrated Nested Laplace Approximation (INLA) to approximate the posterior marginal distribution of the parameters of interest, rather than the more common Markov Chain Monte Carlo (MCMC) approach. Based on preliminary analysis, the data appeared to be overdispersed, containing a disproportionately high number of zeros along with a high variance relative to the mean. Thus, I considered two potential data models, the negative binomial and the zero-inflated negative binomial. Of these models, the zero-inflated negative binomial had the lowest DIC, thus inference was based on this model. Results from this model indicated that a large proportion of the strata were not decreasing (*i.e.*, the estimated slope of the β parameter was not significantly different from zero). However, there were important exceptions with strata in the northwest boreal forest and southern prairie parkland habitats. Several strata in the boreal forest habitat had negative slope estimates, indicating a decrease in breeding pairs, while some of the strata in the prairie parkland habitat had positive slope estimates, indicating an increase in this region. Additionally, from looking at plots of individual strata, it seems that the strata experiencing increases in breeding pairs are experiencing dramatic increases. Overall, my results support previous work indicating a decline in population abundance in the northern boreal forest of Canada, and additionally indicate that the population of scaup has increased rapidly in the prairie pothole region since 1957. Yet, by accounting for spatial and temporal autocorrelation in the data, it appears that declines in abundance are not as widespread

as previously reported.

(52 pages)

ACKNOWLEDGMENTS

I would like to acknowledge the great amount of help and guidance from my advisor, Mevin Hooten, and my Ph.D. advisor, David Koons, as well as the additional improvements to the thesis by my committee member James Powell. Additionally, I would like to thank my family and friends for their support while working on my project.

Beth Ross

CONTENTS

LIST OF FIGURES

INTRODUCTION

Hierarchical Modeling

In ecological studies, the goal is often to describe and gain further insight into ecological processes underlying the data collected during observational studies. Because of the nature of observational data, it can often be difficult to separate the variation in the data from the underlying process or 'state dynamics.' In order to better address this issue, it is becoming increasingly common for researchers to use hierarchical models, or models that have multiple levels of mathematical and statistical equations describing the overall system (*e.g.*, Cressie et al. 2009). Hierarchical models are constructed in such a way that they are able to explicitly account for a model describing the data, as well as a model describing the process from which the data arise (*e.g.*, Royle and Dorazio 2008). For example, hierarchical models can be easily applied in a study focused on assessing changes in abundance of a certain species in a given area. If we assume that the individuals of a population (with population size N_i) are independently observed, for each sample unit $i = 1, 2, \ldots, M$, then the data are counts of 'observed' individuals, y_i, that come from a binomial distribution, $[y_i \,|\, N_i] = \mathrm{Bin}(y_i \,|\, N_i, p)$, where the square bracket notation implies a probability distribution, p is the detection probability, and $[y_i \,|\, N_i]$ is commonly called the 'data model'. Here, N_i are unobserved parameters, and could likely arise from a Poisson distribution indicating the intensity of abundance in each sample unit (*e.g.*, an N-mixture model, Royle and Dorazio 2008). This would be indicated by the process model, $[N_i \,|\, \lambda] = \mathrm{Poisson}(N_i \,|\, \lambda)$. While the parameters p and λ are confounded in this specification, conducting additional site visits to estimate p would solve this problem. By using a hierarchical framework and repeated site visits to answer this problem,

we are able to simultaneously model the data while accounting for the underlying process the data arises from, thus accounting for parameter uncertainty as well as process uncertainty. This last step is an important distinction of hierarchical models, in contrast to traditional statistical procedures which focus solely on the uncertainty in the data (Dennis et al., 2006). When the uncertainty due to the ecological process of interest is ignored, it can result in spurious conclusions (Cressie et al., 2009). Additionally, de Valpine and Hastings (2002) showed that ignoring observation error can also lead to inaccurate estimates of the population parameters of interest (*e.g.*, survival or abundance). Hierarchical models can also be helpful when evaluating the level of density-dependence exhibited in a population. When parameters related to density-dependence are estimated using response and explanatory data that are dependent, it can lead to invalid conclusions (Krebs, 2002). Evaluating these same models using a hierarchical structure allows for the proper separation of process noise and observation error (Dennis et al., 2006).

Hierarchical Spatial Models

Hierarchical models are especially useful for handling data that contain latent forms of autocorrelation. By their nature, spatial data are hierarchical, as data can be gathered at a local scale (*e.g.*, in 1 meter plots) all the way up to regional or broad geographic scales (*e.g.*, data measured at a state-wide level). When sampled spatially, it is quite likely that data points taken in close proximity to one another will be more similar than those points separated by a broad geographic scale. This is the concept of spatial autocorrelation (Koening, 1999). When inference is made on data that are thought to have independent errors, when they are actually dependent, the critical level of the test falls above where truth actually lies, thus leading to false rejections and erroneous inference because the estimated variance is too small (Cressie and

Wikle, 2011). The added benefit of using hierarchical modeling with spatial data is the ability to parse out first and second order processes occurring in the system, thus accounting for spatial covariance between data points. For example, with a simple linear model $y_i = \mu_i + \epsilon_i$, the μ_i account for first order effects and the ϵ_i can be used for second order effects. Hierarchical spatial, temporal, and spatio-temporal models allow for the simultaneous modeling of both first and second order processes, thus accounting for underlying autocorrelation in the system while still providing insight into overall spatial and temporal pattern.

Estimating Changes in Animal Abundance

The question of how to best estimate animal populations has been an area of research for many years, especially regarding changes in populations over space and time (Williams et al., 2002). Some of the commonly used methods of estimating population abundance are capture-mark-recapture models (Lebreton et al., 1992), the inclusion of abundance counts in occupancy modeling (MacKenzie et al., 2002), and distance sampling methods (Buckland, 2001). While all of these methods account for the additional confounding issue of detection probability, they also require more information than may be readily available (or even feasible to collect) in long-term, broad-scale wildlife surveys. Generally, when surveying for wildlife on the scale of the entire continent, surveys are conducted using aircraft to fly over the survey region and count the visible wildlife. This method is efficient, and allows biologists to cover a broad spatial scale in a relatively short period of time. The downside of this monitoring scheme is that, in most cases, it does not allow for the incorporation of methods to estimate detection probability, and is therefore only able to address trends in relative abundance, not necessarily true abundance. Additionally, if a survey region is quite large, then the assumption of a constant detection probability (*e.g.*, visibility

correction factors) may not be valid, as detection of a species likely varies between different habitat types and over time.

Application of Hierarchical Modeling

An example of a continental-scale flight survey is the North American May Breeding Pair Survey (BPS), which provides a rich source of demographic data on 10 focal duck species (as well as others), and covers a large portion of each species' breeding range, covering areas in the northern United States, western Canada, and Alaska (Zimpfer et al., 2011). This survey has been conducted every May through June since 1955 using aerial transects (Smith, 1995). Observers count which duck species are observed, and whether or not the ducks are paired (with a mate), single drakes, or in mixed-sex groups of ducks. Surveys are flown at approximately 120 miles per hour at an altitude of 90-100 feet. The detectability of each duck species is estimated using ground crews to truth the aerial surveys; the difference in counts is then used to estimate a visibility correction factor (VCF). Unfortunately, the ground-crew surveys are not conducted every year nor done the same way across strata. For example, while crews can easily access transects in the prairie pothole region of the survey, it is quite difficult logistically to reach the transects in some areas of the boreal forest by road. In these cases, transects in the boreal forest have been 'ground truthed' by helicopters. Additionally, the ground crews have only been conducted in some years of the survey, and then applied in subsequent years until the next ground-truthing survey. Habitat conditions will likely be similar between consecutive years, but VCF estimates are sometimes used over much longer periods of time (*e.g.* decades), without a reevaluation of the VCF. Lastly, and perhaps most importantly, the calculation of the VCF assumes that the ground crews are conducting a census of the area, and there are many instances when the aerial crews will have higher counts of ducks over

a given segment than the ground crews. This could be both because of the delay between the aerial and ground surveys, or because the ground crews are unable to see as many of the ducks as the aerial crews. Either way, there are many potential flaws with the estimation of the VCF, and because of this, I used only raw counts from the aerial surveys in my analysis of the BPS data.

Lesser Scaup Decline

Two particular species of interest that are surveyed during the BPS are the lesser and greater scaup (*Aythya affinis* and *Aythya marila*), which are counted collectively as scaup. Scaup are the most abundant and widespread diving duck in North America, and are important game species (Austin et al., 1998). Since 1978, the continental population of scaup has declined to levels that are 16% below the 1955-2010 average and 34% below the North American Waterfowl Management Plan goal (Zimpfer et al., 2011). This decline has sparked concern amongst hunters, management agencies, and conservation groups alike (Afton and Anderson, 2001). The greatest decline in abundance of scaup appears to be occurring in the western boreal forest, where populations may have depressed rates of reproductive success, survival, or both (Afton and Anderson, 2001; Walker and Lindberg, 2005; Corcoran et al., 2007; Hobson et al., 2009). However, the specific vital-rate pathways responsible for the decline are not known (Koons et al., 2006).

Based on the recent decline, the U.S. Fish and Wildlife Service has implemented a decrease in the bag limit of scaup (Boomer and Johnson, 2007). Because they are such a popular species to hunt, this has caused quite a negative response from hunting groups, and raised concerns about the methods used to determine harvest numbers for the species. While the report suggesting new harvest regulations made use of advanced modeling techniques (Boomer and Johnson, 2007), it did not explicitly model

the impacts of harvest on survival. Also, if the relative pressure of hunting has less impact on the population decline than factors such as low nest success and duckling survival, then a decrease in the harvest rate will have little impact on population growth, especially if the population experiences compensatory mortality. By considering spatial variation in abundance when making management decisions, perhaps the reduced bag limit could be implemented in only those flyways most impacted by breeding areas that have been experiencing significant declines.

A leading hypothesis for the underlying cause of scaup decline is that they are experiencing decreased food availability during spring migration, thus arriving on the breeding grounds in poor body condition (*i.e.*, the Spring Condition Hypothesis; (Austin et al., 2000; Anteau and Afton, 2004)). Studies tracking body condition and nutrient reserves throughout the scaup migration have observed a long-term decline in both body condition and nutrient reserves that could have a significant impact on reproductive effort and success (Anteau and Afton, 2004). Moreover, Anteau and Afton (2008) found a long-term decline in consumption of their preferred food (*Amphipod spp.*) in the upper Midwest. Recent studies in the boreal forest, however, suggest that scaup in 2003 and 2004 had a similar body condition relative to historically collected scaup (Devink et al., 2008). Scaup on the great lakes and other migratory stopover locations are also known to contain high levels of toxins, particularly selenium (Anteau et al., 2007; Custer and Custer, 2000). However, DeVink et al. (2008) and others (Fox et al., 2005; Matz and Rocque, 2007) concluded that selenium and mercury levels are low in boreal scaup, and not likely responsible for the population decline in this important breeding region that is exhibiting the sharpest declines. Conflicting findings and a variety of other explanations (*e.g.*, change in North American climate, anthropogenic developments in the boreal forest) illustrate the complicated nature of the decline in scaup abundance. Several factors are likely

impacting scaup, and perhaps no single hypothesized cause can explain the decline entirely.

In order to better understand the causes of the decline, and better understand the biology of scaup in general, a level of high importance has been placed on retrospective analyses that determine the spatial and temporal changes in population abundance (Austin et al., 2006). By focusing on scaup as a model system to develop more efficient ways to make use of aerial survey data, I will address an issue of importance, the use of hierarchical models in ecological applications, and concurrently gain insight into an issue of biological importance: the decline of the scaup population. In addition, the population decline seems to have a large spatial component to it; as the rate of birds occupying survey transects in the northern boreal forest of Canada is decreasing, the analogous rate is increasing in the southern prairie parklands (Ross and Koons, *in progress*). Any development of statistical methodology intended to yield insightful conclusions about the species needs to address changes in the population across different spatial locations. Future research and management actions should be directed at the specific locales where significant change in scaup abundance has actually occurred. However, because survey locations are not thought to be independent of one another, additional spatial structure will be needed in statistical models used to guide management actions aimed at reversing declines in scaup abundance.

Integrated Nested Laplace Approximation

In order to implement Bayesian hierarchical models, I used a method called Integrated Nested Laplace Approximation (INLA) to approximate the posterior marginal distribution of the parameters of interest, rather than the more common Markov Chain Monte Carlo (MCMC) approach (Rue et al., 2009). Many models used in ecological applications are based on latent Gaussian random fields, ranging from linear

regression models, temporal models, spatial, and spatio-temporal models. By making use of these latent Gaussian models, INLA is capable of approximating the posterior distribution with high accuracy at a much faster computational rate than MCMC (Rue et al., 2009). Unlike MCMC, INLA is not an iterative stochastic procedure, but rather a multi-step mathematical process used to approximate posterior distributions. To do so, INLA involves a three step process. It first uses Laplace approximation to calculate the posterior marginal of the parameter of interest, $\boldsymbol{\theta}$. It then computes the Laplace approximation, or the simplified Laplace approximation of the posterior, $[\boldsymbol{\theta}\,|\mathbf{Y},\boldsymbol{\pi}]$, where $\boldsymbol{\theta}$ is our parameter of interest, $\mathbf{Y}$ the data, and $\boldsymbol{\pi}$ the hyperparameters, or priors, of the parameters. This second step is performed for selected values of $\boldsymbol{\pi}$ to improve on the approximation. Lastly, INLA combines the previous two steps by using numerical integration. While it is usually more computationally efficient than MCMC, the computational cost is exponential with respect to the number of hyperparameters in the model (Rue et al., 2009). When the number of parameters is few, however, the problem may not be that severe. Additionally, INLA cannot approximate nonlinear transformations of model parameters, which may be a drawback for some ecological studies, but does not interfere with this research.

METHODS

Survey Details

I utilized May Breeding Pair Survey (BPS) data collected by the United States Fish and Wildlife Service (USFWS) and Canadian Wildlife Service (CWS). The BPS has been flown every May–June since 1955 in the Traditional Survey Area (TSA), which encompasses the north-central United States, central and western Canada, and a large portion of Alaska (Fig. 1). This region is thought to represent the primary breeding grounds for a large proportion of North American waterfowl, including scaup. Moreover, the TSA consists of strata divided by similar habitats and duck densities (Smith, 1995). Within each stratum pilots fly multiple transects, each comprised of 28.8 km strip-segments. The number of segments sampled in a transect ranges from 1 to 35, and has changed over time. The TSA encompasses several different habitat types (*e.g.*, boreal forest, prairie parkland, and tundra). Throughout the survey, pilots count the number of paired, single, and grouped ducks of each species to the left of the plane; accompanying biologists record the same information on the right side as well as pond numbers (ponds are only counted in the prairie parkland region). Here, the focus is on the delineation of scaup recorded in breeding pairs, rather than total scaup abundance, as these pairs best represent the breeding potential of the population. I did not utilize data regarding single drakes due to the skewed sex ratio in scaup (Afton and Anderson, 2001), and the methods of counting the mixed-sex non-breeding groups changed in 1975 (Smith, 1995); thus, use of these data would confound long-term analysis from 1955 to the present.
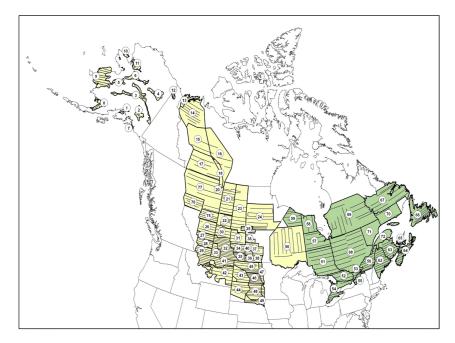
Fig. 1: Map of surveyed area from U.S. Fish and Wildlife Service. Only the area in yellow was used for this study.

Model

In order to best account for process and sampling variation, I used Bayesian hierarchical models to examine change in scaup abundance for each stratum between 1957 and 2009 (1955-1956 had a paucity of spatial data). This approach allowed me to incorporate spatial autocorrelation among strata into the process model while also gaining insight into the changes within each stratum. Additionally, using hierarchical models allowed me to select between two data models to determine which model best fit the data. Bayesian hierarchical models are usually comprised of three 'submodels': a data model describing the distribution of the data, a process model specifying the underlying mechanism that gives rise to the data, and the parameter model indicating the distributions of the parameters in the process model (Berliner, 1996).

Data Model

Based on preliminary analysis, the data from the BPS appeared to be overdispersed, containing a disproportionately high number of zeros along with a high variance relative to the mean. Thus, I considered two potential data models, a model where $y_{i,j,t} \sim \text{NegBinom}(\mu_{j,t}, \phi)$, and a model where

$$
y_{i,j,t} \sim \begin{cases} 0, & \text{with probability} \quad \psi \\ \text{NegBinom}(\mu_{j,t}, \phi), & \text{with probability} \quad (1 - \psi) \end{cases}
$$

for segments $i = 1,\dots,n_j$ in stratum $j = 1,\dots,m$ during observation period $t = 1,\dots,T$ (*i.e.*, years 1957-2009). Models were then compared using the Deviance Information Criterion (DIC). I considered the zero-inflated model because it helps to account for excess zeros, perhaps due to birds that were present during the survey, but unobserved. One of the benefits of hierarchical modeling is that the proper support could be specified for the data distribution, and thus transformations (*e.g.*, log-transform, sin-transform) were not necessary in order to use a linear process model, as it has been shown that models based on the negative binomial distributions can perform better at modeling count data than transformed data (O'Hara and Kotze, 2010).

Process Model

Each data model was incorporated into a simple log-linear regression model with an error term incorporating spatial and temporal autocorrelation. This allowed me to analyze the temporal trend for each stratum separately, as well as determine the relative influence of both the fixed (the β parameters) and random (ϵ and η) effects on relative scaup abundance. Using $\mu_{j,t}$ from the above data model, the process model

can then be specified as

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

where $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{m,t})\prime \sim \text{N}(0, \boldsymbol{\Sigma}_\epsilon)$, and $\boldsymbol{\Sigma}_\epsilon \equiv \sigma_\epsilon^2(\mathbf{D} - \mathbf{W})^{-1}$ for all $t = 1, \ldots, T$ representing spatially correlated errors and $\eta_j = (\eta_{j,1}, \ldots, \eta_{j,T})\prime \sim \text{N}(0, \Sigma_\eta)$ for all j = 1,…,m representing temporally correlated errors. The ϵ and η terms incorporated into the model are intended to account for variation in both the spatial and temporal dimensions of the system, as $\mathbf{W}$ is a matrix containing the spatial field, and $\mathbf{D}$ is a diagonal matrix with the row sums of $\mathbf{W}$ as diagonal elements (Banerjee et al., 2004). Again, if these sources of uncertainty are not accounted for, erroneous inference could be made because model assumptions would not be met (Cressie and Wikle, 2011).

Rather than standardize the survey area to some sort of grid as past studies have done (*e.g.*, Gardner et al. (2007)), I instead used the existing spatial structure present in the survey, and based my spatial field on the strata units as areal regions. This provided the basis for a conditional autoregressive structure (CAR; Besag 1974; Brook 1964), while also yielding results that were directly useful for management of the species. The CAR model is a commonly used spatial model for areal processes, and incorporates dependence among areal locations through the neighborhood structure of the strata. In constructing the proximity matrix, $\mathbf{W}$, I calculated the Euclidean distance between the center of each stratum, 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 distance was chosen because it was the shortest distance at which all strata had at least one neighbor, and constructing the proximity matrix in such a way accounted for the more spatially isolated strata in the north and the highly connected strata in the south.

In addition to the spatial structure, I also wanted to account for latent temporal dependence in the system. I addressed this by incorporating temporal dependence into the $\boldsymbol{\eta_j}$. I let $\eta_{j,t} \sim \text{N}(\alpha\eta_{j,t-1}, \sigma_\eta^2)$, which induces a correlation structure on the vectors $\boldsymbol{\eta_j}$ such that $\boldsymbol{\eta_j} \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma_\eta})$. Then the entire latent process is a Gaussian Markov random field, which allows for the use of INLA to approximate posterior distributions.

Due to the spatial and temporal autocorrelation likely present in the data, I analyzed the contributions of the fixed and random effects in the model residuals. In order to address this, I conducted a Moran's I test on the residuals from the top model without including the spatial random effect (*i.e.*, the spatial field). This showed whether the remaining variation from the spatio-temporal trend model had significant spatial autocorrelation. Furthermore, I constructed an autocorrelation function (ACF) and partial autocorrelation (PACF) graphs on the residuals with and without a temporal trend, to determine how much variation the temporal trend accounts for compared to the random effect.

Parameter Model

The parameter model consists of all the prior distributions given to the unknown variables in the process model. The overdispersion parameter for the negative binomial distribution was specified as $\phi = log(n)$, where n is the original negative binomial size parameter, and then modeled as $\phi \sim \text{N}(0, 100)$. The regression parameters were specified with conjugate Gaussian priors so that $\beta_{0,j}, \beta_{1,j} \sim \text{N}(0, 1000)$, for $j = 1,\ldots,m$. The variance component, which was defined as the precision, or the inverse of the variance component (Rue et al., 2009), was given a conjugate inverse gamma prior, $\sigma_\epsilon^{-2} \sim \text{Gamma}(1, 1/20000)$. The parameter α, the autocorrelation parameter in the temporal random effects model, was assigned the prior $\alpha \sim \text{N}(0, 0.15)$.

Model Implementation

Combining the data, process, and parameter models to form a hierarchical model yielded the posterior distribution:

$$\left[\{\boldsymbol{\mu_t}\}, \phi, \{\boldsymbol{\beta_{j,t}}\}, \sigma_\epsilon^2, \sigma_\eta^2, \alpha | \{y_{i,j,t}\}\right] \propto \prod_i \prod_j \prod_t [y_{i,j,t}|\mu_{j,t}, \phi] \prod_t \prod_j \left[\mu_{i,t}|\boldsymbol{\beta_j}, \sigma_\epsilon^2, \sigma_\eta^2, \alpha\right]$$

$$\times [\phi] [\boldsymbol{\beta}] [\alpha] \left[\sigma_\epsilon^2\right] \left[\sigma_\eta^2\right].$$

In the above equation, I use square brackets '[.]' to denote a probability distribution. Typically, the above model would be fit using Markov Chain Monte Carlo, usually using a Gibbs sampler and a Metropolis-Hastings algorithm (Banerjee et al., 2004). However, as discussed in the *Introduction*, MCMC can be computationally expensive. For this reason, I implemented the above model using the INLA package in R (Rue et al., 2009).

RESULTS

Summary Statistics

The TSA covered approximately 3.3 million square kilometers during an annual survey route, sampling 52 unique strata comprised of approximately 2588 segments each year, resulting in 103,266 counted scaup pairs from 1957-2009. The area sampled in 1955 and 1956 was quite limited spatially, thus those two years were not included in the analysis. When calculating the total counted scaup breeding pairs in each year, preliminary analysis indicated a decline during the last twenty years, coinciding with estimates of total abundance based on the VCF (Fig. 2).

INLA

Two data models were chosen for comparison based on the Deviance Information Criterion (DIC, Spiegelhalter et al. 2002). This allowed me to assess how well each model fit the data without presuming to know the best fit *a priori*. The two chosen distributions were the negative binomial and the zero-inflated negative binomial (ZINB). Of these models, the ZINB had the lowest DIC (24391 compared to 24933 for the negative binomial) and further analysis proceeded using this model.

Results from the ZINB model indicated that a large proportion of the strata were not decreasing (*i.e.*, the estimated slope of the β parameter was not significantly different from zero; Fig. 4). However, there were important exceptions with strata in the northwest boreal forest and southern prairie parkland habitats (Fig. 5). Several strata in the boreal forest habitat had negative slope estimates, indicating a decrease in breeding pairs, while some of the strata in the prairie parkland habitat had positive slope estimates, indicating an increase in this region (Fig. 4 & 6). While Figure
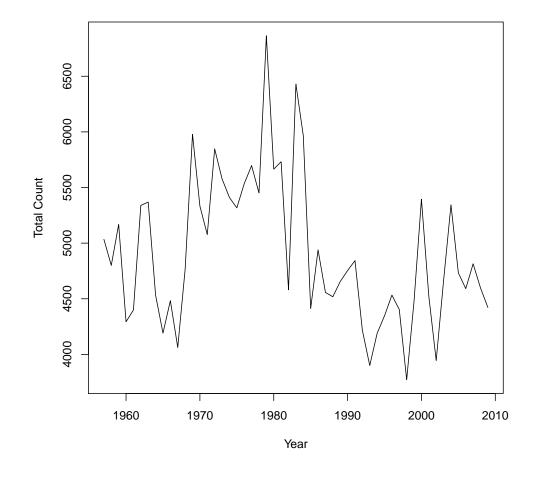
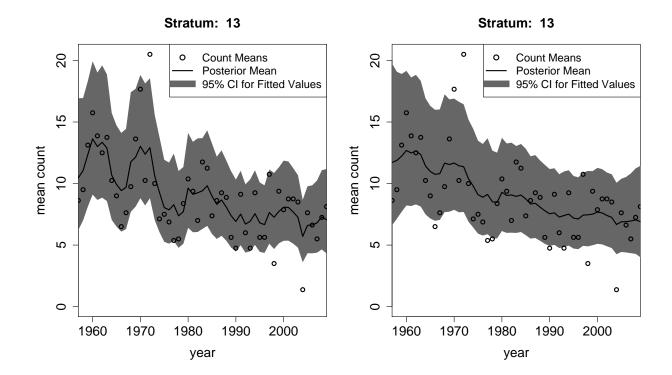Fig. 2: Total counts of scaup breeding pairs for each year since 1957.

Fig. 3: Population estimates based on the posterior predictive distribution for Stratum 13 from the zero-inflated negative binomial (left) and negative binomial (right) models.

4 compiles all of the fixed effects into one image, another way to view the output from INLA is to look at each stratum individually. Viewing each stratum individually allows one to visualize the combined temporal and spatial effects on changes in breeding pairs (Fig. 7). For example, by looking at the estimated time trend for stratum 18, a stratum with an estimated negative slope, a steady decrease in the number of breeding pairs since the 1950s is noted, while looking at the trend for stratum 45, the population increased in the mid-1980s, decreased from the mid-1980s to the mid-1990s, and has increased dramatically since the mid-1990s.

Removing the random effects had little impact on changing the posterior distributions for the parameters, indicating that the trend is accounting for the majority of spatial autocorrelation in the system. There were only 3 years in which there was significant spatial autocorrelation in the residuals, based on the Moran's I test. Additionally, there were only 7 strata that had significant temporal autocorrelation based on the ACF and PACF graphs. The autoregressive parameter α had a mean of 0.975 with lower and upper 95% quantiles 0.949 and 0.991.
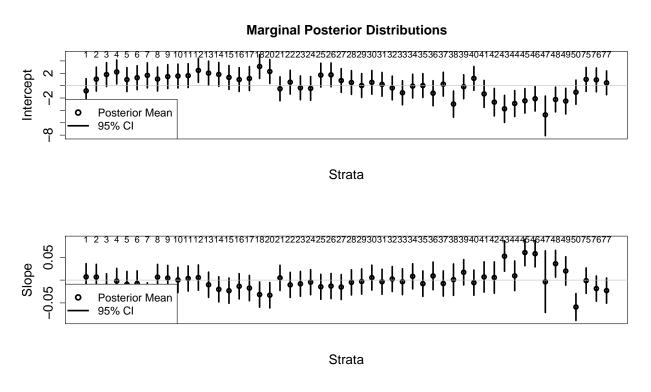
Fig. 4: Marginal posterior distributions for the β coefficients controlling for general trend using the zero-inflated negative binomial data model.
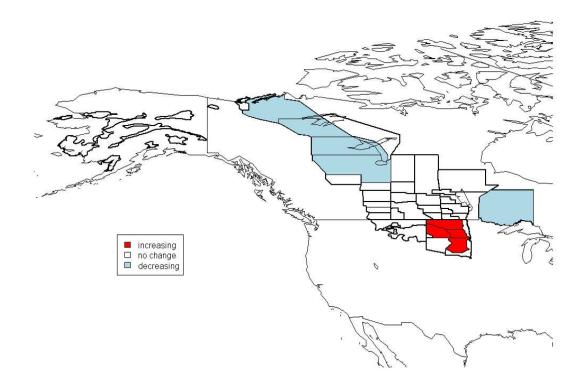
Fig. 5: Increases and decreases in the abundance of breeding pairs since 1957 based on posterior distributions for the $\beta_{1,j}$ coefficients. Highlighted areas show an 80% chance of the population increasing or decreasing in the area since 1957.

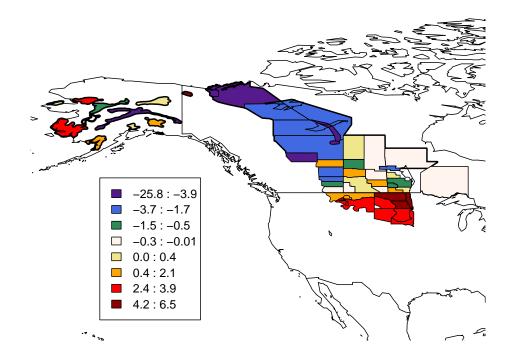Fig. 6: Changes in counted breeding pair abundance since 1957 based on posterior distributions for the $\beta_{1,j}$ coefficients. Negative numbers indicate pairs lost and positive numbers indicate pairs gained. Note that not all differences are estimated to be statistically different from zero.
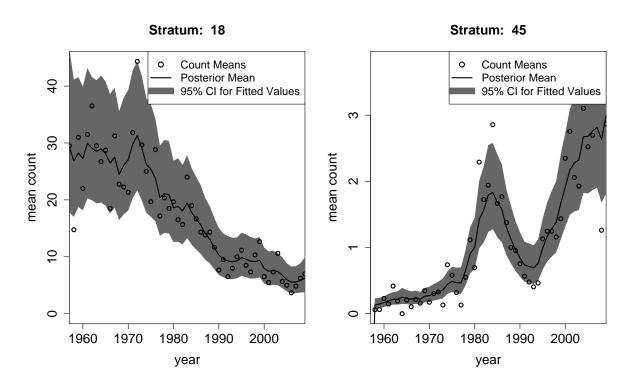
Fig. 7: Marginal posterior predictive distributions from Strata 18 and 45 from the zero-inflated negative binomial model.

DISCUSSION

Scaup Decline

My results provide further insight into the continental decline of breeding scaup abundance. Due to uncertainty associated with the slope parameters, however, the majority of strata in the survey region did not indicate evidence of either a decline or increased breeding pair abundance (Fig. 4). This could imply that those strata where there is a decline are declining quite severally, which appears to be the case for the northwestern boreal forest (Fig. 5), especially Strata 18 and 20 (Fig. 6). There were only 4 strata for which 95% credible intervals indicated significant population decline, whereas 8 strata were significant when considering 80% credible intervals. This could indicate that it is difficult to detect changes in breeding pair abundance, given the amount of sampling variation and autocorrelation present in the data. Additionally, from looking at plots of individual strata, it seems that the strata experiencing increases in breeding pairs are experiencing dramatic increases, but intially had lower abundance. Both of these changes are well illustrated in Figure 7.

Given the results from this study, there are several possible changes occurring in the population. First, scaup may not be migrating as early as they have historically, thus reaching their final breeding grounds later in the spring, and not being counted as breeding pairs during the BPS survey, especially in the northern boreal forest. This would mean that upon finally reaching the boreal grounds, they have a shorter season for breeding and raising young, which may limit successful recruitment (Dawson and Clark, 2000). Additionally, the breeding season has historically coincided with the peak hatching season of invertebrates in the boreal forest, which provides important resources for ducklings. With a shift in breeding phenology, this could

cause a mismatch between predator and prey, again resulting in poor reproductive success (Visser et al., 1998). Secondly, scaup may be choosing to forego migration up to the boreal forest, and are instead utilizing the prairie parklands at higher rates than in the past. Scaup are however very philopatric (Afton, 1984; Rotella et al., 2003), and a wholesale change in habitat selection does not seem likely. Rather, the birth-death balance has likely been negative for quite some time in the northwestern boreal forest, and positive in the southern PPR, perhaps due to the Conservation Reserve Program and other landscape management efforts that have increased nest success (Stephens et al., 2005). While my results and others seem to indicate an increase in scaup numbers in the prairie parkland area (Afton and Anderson, 2001), this increase is not large enough to offset the large decrease in the boreal forest (Fig. 6). While the U.S. prairie population seems to have been growing exponentially since the mid-1990s, it is worth noting that the increments for the mean count of paired birds are quite small, ranging from only 1 to 3 (Fig. 6). This small growth does not offset the large decrease observed in the boreal forest (*e.g.*, Stratum 18, Fig. 7). While these strata are only an example from each of the respective habitat types, there are only a few strata showing statistically significant increases and decreases.

This approach differs substantially from the current methods used to estimate waterfowl numbers, which may be why there is a discrepancy between my results and those of the USFWS (Boomer and Johnson, 2007; Zimpfer et al., 2011). Specifically, the USFWS does not decompose the survey into population changes in each stratum, but rather looks at how extrapolated numbers are changing at the continental scale. While the continental count can be related back to changes occurring in each stratum, for future management decisions, an approach focused on counts in specific spatial regions may be more useful for guiding management actions on the ground. Understanding why scaup are declining in certain areas of the boreal forest, and increasing

in areas of the prairie pothole region is likely critical to preservation of the species as a game species. It seems that the decrease in scaup is occurring on a much more local scale than previously thought, and looking at the population change on a continental scale will not help identify the cause of decline in specific areas (Afton and Anderson, 2001). Furthermore, intensive monitoring of individual strata might be useful for management purposes, as count information on a finer spatial scale could help elucidate the causes and prevalence of population decline. Recent studies indicate that current and future climate change may negatively impact late-nesting ducks such as scaup (Drever et al., 2012) and monitoring in preparation for this decline is critical.

My results may differ somewhat from similar studies (Afton and Anderson, 2001) because of the time period taken into consideration, as well as using a hierarchical model with slightly different data. Because I only took into consideration the count of breeding pairs, I was able to make inference on population changes since 1957. If a different time period is used (*e.g.*, from 1955 to 1997 or 1978 to 1997, as in Afton and Anderson 2001), different conclusions may have been reached. Methodological changes occurred in 1975 that made it difficult to compare counts of grouped birds, and thus the total number of birds, from time periods before and after 1975. My focus on breeding pairs alleviates this problem and my use of flexible hierarchical models alleviated the need to choose a year of peak abundance (*e.g.*, the mid-1970s) from which to analyze subsequent population decline (du Toit, 2010). Additionally, my model was based on raw count data on breeding pairs, rather than extrapolated numbers based on the VCF. These numbers differ greatly, with raw counts ranging from 4000-6500 breeding pairs, and extrapolated VCF estimates ranging in the millions.

Integrated Nested Laplace Approximation

Using INLA, I was able to successfully implement and compare several different data and process models to determine the best fit to the data, making this the most sophisticated and proper analysis of the BPS data. While this can be done using traditional MCMC methods, the processing time using INLA was considerably shorter, allowing me to fit models that might have otherwise not been considered due to computing limitations. For example, the processing time for the full negative binomial model was approximately 27 minutes on a 2×2.93 GHz 6-Core Intel Xeon workstation, making it roughly an order of magnitude faster than MCMC for this model and dataset. While INLA does restrict the user to generalized linear models, it could be quite useful as an initial step for determining variables of importance to incorporate into a non-linear model using MCMC.

Conclusions

Overall, my results support previous work indicating a decline in population abundance in the northern boreal forest of Canada, and additionally indicate that the population of scaup has increased rapidly in the prairie pothole region since 1957. Additionally, it seems that the most important processes influencing population dynamics are not related to underlying autocorrelation, but rather parameters in the model explicitly accounting for the differences among the strata (*i.e.*, the fixed intercept and slope parameters). Much of the variation in the model was explained by the fixed effects, and the temporal trend in the model was of greater importance in predicting dynamics than the random effects. These results could be expanded to accommodate additional covariates that might be important in determining population trends, including information related to habitat quality (*e.g.*, drought conditions, large scale weather patterns such as PDO) and hunter harvest information. Non-

linear influences such as density dependence could also be included outside of the INLA framework (or by using INLA with linear analogs for modeling density dependence with log counts). Additionally, results from the INLA analysis can be used to better inform the way in which the BPS is conducted, using an optimal sampling scheme to maximize the amount of information gathered during the survey while minimizing associated error (Hooten et al., 2012).

REFERENCES

Afton, A. (1984). Influence of age and time on reproductive performance of female lesser scaup. *Auk* **101,** 255–265.

Afton, A. and Anderson, M. (2001). Declining scaup populations: a retrospective analysis of long-term population and harvest survey data. *Journal of Wildlife Management* **65,** 781–796.

Anteau, M. and Afton, A. (2004). Nutrient reserves of lesser scaup (*Aythya affinis*) during spring migration in the mississippi flyway: a test of the spring condition hypothesis. *Auk* **121,** 917–929.

Anteau, M. and Afton, A. (2008). Diets of lesser scaup during spring migration throughout the upper-Midwest are consistent with the spring condition hypothesis. *Waterbirds* **31,** 97–106.

Anteau, M., Afton, A., Custer, C., and Custer, T. (2007). Relationships of cadmium, mercury, and selenium with nutrient reserves of female lesser scaup (*Aythya affinis*) during winter and spring migration. *Environmental Toxicology and Chemistry* **26,** 515–520.

Austin, J., Afton, A., Anderson, M., Clark, R., Custer, C., Lawrence, J., Pollard, J., and Ringelman, J. (2000). Declining scaup populations: issues, hypotheses, and research needs. *Wildlife Society Bulletin* **28,** 254–263.

Austin, J. E., Anteau, M. J., Barclay, J. S., Boomer, G. S., Rohwer, F. C., and Slattery, S. M. (2006). Declining scaup populations: Reassessment of the issues, hypotheses, and research directions. Consensus Report from the Second Scaup Workshop.

Austin, J. E., Custer, C. M., and Afton, A. D. (1998). Lesser scaup (*Aythya affinis*). In Poole, A. A. and Gill, F., editors, *The Birds of North America, Number 228*. The American Ornithologists' Union, Washington, D.C., USA.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. CRC Press, Boca Raton, FL.

Berliner, L. (1996). Hierarchical bayesian time-series models. In Skilling, J., editor, *Maximum Entropy and Bayesian Methods*, pages 15–22. Kluwer Academic Publishers, Dordrecht, NL.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **36,** 192–236.

Boomer, G. and Johnson, F. (2007). A proposed assessment and decision-making framework to inform scaup harvest management. U.S. Fish and Wildlife Service, Laurel, Maryland.

Brook, D. (1964). On the distinction between the conditional probability and joint probability approaches in the specification of nearest neighbor systems. *Biometrika* **51,** 481–483.

Buckland, S. T. (2001). *Introduction to distance sampling*. Oxford University Press, New York, NY.

Corcoran, R. M., Lovvorn, J. R., Bertram, M. R., and Vivion, M. T. (2007). Lesser scaup nest success and duckling survival on the Yukon Flats, Alaska. *Journal of Wildlife Management* **71,** 127–134.

Cressie, N., Calder, C., Clark, J., Hoef, J., and Wikle, C. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* **19,** 553–570.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data.* Wiley.

Custer, C. and Custer, T. (2000). Organochlorine and trace element contamination in wintering and migrating diving ducks in the southern great lakes, USA, since the zebra mussel invasion. *Environmental Toxicology and Chemistry* **19,** 2821–2829.

Dawson, R. and Clark, R. (2000). Effects of hatching date and egg size on growth, recruitment, and adult size of lesser scaup. *Condor* **102,** 930–935.

de Valpine, P. and Hastings, A. (2002). Fitting population models incorporating process noise and observation error. *Ecological Monographs* **72,** 57–76.

Dennis, B., Ponciano, J., Lele, S., Taper, M., and Staples, D. (2006). Estimating density dependence, process noise, and observation error. *Ecological Monographs* **76,** 323–341.

DeVink, J., Clark, R., Slattery, S., and Trauger, D. (2008). Are late-spring boreal lesser scaup (*Aythya affinis*) in poor body condition? *Auk* **125,** 291–298.

Devink, J., Clark, R., Slattery, S., and Wayland, M. (2008). Is selenium affecting body condition and reproduction in boreal breeding scaup, scoters, and ring-necked ducks? *Environmental Pollution* **152,** 116–122.

Drever, M., Clark, R., Derksen, C., Slattery, S., Toose, P., and Nudds, T. (2012). Population vulnerability to climate change linked to timing of breeding in boreal ducks. *Global Change Biology. Accepted article* .

du Toit, J. (2010). Considerations of scale in biodiversity conservation. *Animal Conservation* **13,** 229–236.

Fox, G., MacCluskie, M., and Brook, R. (2005). Are current contaminant concentrations in eggs and breeding female lesser scaup of concern? *Condor* **107,** 50–61.

Gardner, B., Royle, J. A., and Boomer, G. S. (2007). Modeling spatial and temporal variation in scaup abundance. Unpublished Report.

Hobson, K., Wunder, M., Van Wilgenburg, S., Clark, R., and Wassenaar, L. (2009). A method for investigating population declines of migratory birds using stable isotopes: origins of harvested lesser scaup in North America. *PLoS ONE* **4,** 1–10.

Hooten, M., Ross, B., and Wikle, C. (2012). Optimal spatio-temporal monitoring designs for characterizing population trends. In Gitzen, R., Millspaugh, J., Cooper, A., and Licht, D., editors, *Design and Analysis of Long-Term Ecological Monitoring Studies. In press.*

Koening, W. (1999). Spatial autocorrelation of ecological phenomena. *Trends in Ecology and Evolution* **14,** 22–26.

Koons, D., Rotella, J., Willey, D., Taper, M. L., Clark, R., Slattery, S., Brook, R., Corcoran, R. M., and Loworn, J. (2006). Lesser scaup population dynamics: what can be learned from available data? *Avian Conservation and Ecology* **1,** 6–22.

Krebs, C. (2002). Two complementary paradigms for analysing population dynamics. *Philosophical Transactions of the Royal Society of London, Series B* **357,** 1211–1219.

Lebreton, J., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62,** 67–118.

MacKenzie, D., Nichols, J., Lachman, G., Droege, S., Royle, J., and Langtimm, C. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83,** 2248–2255.

Matz, A. and Rocque, D. (2007). Contaminants in lesser scaup eggs and blood from Yukon Flats National Wildlife Refuge, Alaska. *Condor* **109,** 852–861.

O'Hara, R. and Kotze, D. (2010). Do not log-transform count data. *Methods in Ecology and Evolution* **1,** 118–122.

Rotella, J., Clark, R., and Afton, A. (2003). Survival of female lesser scaup: effects of body size, age, and reproductive effort. *Condor* **105,** 336–347.

Royle, J. and Dorazio, R. (2008). *Hierarchical Modeling and Inference in Ecology.* Elsevier, Burlington, MA.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71,** 319–392.

Smith, G. (1995). A critical review of the aerial and ground surveys of breeding waterfowl in north america. U.S. Department of the Interior, National Biological Service, Washington, D.C., Biological Science Report 5.

Spiegelhalter, D. J., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 583.

Stephens, S., Rotella, J., Lindberg, M., Taper, M., and Ringelman, J. (2005). Duck nest survival in the Missouri Coteau of North Dakota: landscape effects at multiple spatial scales. *Ecological Applications* **15,** 2137–2149.

Visser, M. E., Noordwijk, A. J. v., Tinbergen, J. M., and Lessells, C. M. (1998). Warmer springs lead to mistimed reproduction in great tits (*Parus major*). *Proceedings of the Royal Society B: Biological Sciences* **265,** 1867–1870.

Walker, J. and Lindberg, M. (2005). Survival of scaup ducklings in the boreal forest of Alaska. *Journal of Wildlife Management* **69,** 592–600.

Williams, B., Nichols, J., and Conroy, M. (2002). *Analysis and management of animal populations.* Academic Press, San Diego, CA.

Zimpfer, N. L., Rhodes, W. E., Silverman, E. D., Zimmerman, G. S., and Richkus, K. D. (2011). Trends in duck breeding populations, 1955-2011. U.S. Fish and Wildlife Service, Laurel, MA.

APPENDIX

R code for INLA and related analysis

Analysis conducted using R version 2.12

```
####  Load Libraries


library(INLA)

library(classInt)

library(maps)

library(maptools)

library(pscl)

library(TSA)


####  Read Data


scaup.df=read.table("count_data.txt",header=TRUE)

strats=readShapePoly("MAS_stratum_boundaries")


####  Setup Variables


idx.extra=c(19,51,52) #  Extraneous Strata to Remove

scaup.2.df=scaup.df[!is.element(scaup.df$strata_no,idx.extra),]

tmp=as(strats,"data.frame")[-idx.extra]
```

```
unique.strata=sort(unique(scaup.df$strata_no))

n.strata=length(unique.strata)

years=1957:2009

n.years=length(years)


####  Convert Data into Regression Format


counts.vec=as.vector(as.matrix(scaup.df[,-(1:6)]))

strata.vec=rep(scaup.df$strata_no,n.years)

new.strata.vec=rep(0,length(strata.vec))

obs=1:length(new.strata.vec)

for(i in 1:n.strata){

  new.strata.vec[strata.vec==unique.strata[i]]=i

}

years.vec=as.vector(col(as.matrix(scaup.df[,-(1:6)])))

X=model.matrix(~0+as.factor(rep(scaup.df$strata_no,n.years)))

colnames(X)=sort(unique(scaup.df$strata_no))

counts.reg.df=data.frame(cbind(counts.vec,X,X*years.vec,strata.vec,

new.strata.vec,years.vec,obs))


counts.reg.noNA.df=counts.reg.df[!is.na(counts.reg.df$counts.vec),]


####  Aggregate Data by Stratum


n.per.stratum.per.year=matrix(NA,n.strata,n.years)

counts.sum=matrix(NA,n.strata,n.years)
```

```
for(i in 1:n.strata){

  for(t in 1:n.years){

    tmp.vec=scaup.df[scaup.df$strata_no==unique.strata[i],6+t]

    if(!all(is.na(tmp.vec))){

      n.per.stratum.per.year[i,t]=sum(!is.na(tmp.vec))

      counts.sum[i,t]=sum(tmp.vec,na.rm=TRUE)

    }

  }

}

for(i in 1:n.strata){

  for(t in 1:n.years){

    if(is.na(n.per.stratum.per.year[i,t]))

    {n.per.stratum.per.year[i,t]=n.per.stratum[i]}

  }

}


n.sum=length(counts.sum)

X.sum=model.matrix(~0+as.factor(rep(1:n.strata,n.years)))

colnames(X.sum)=1:n.strata

X.sum.year=X.sum*as.vector(col(counts.sum))

counts.sum.reg.df=data.frame(y=as.vector(counts.sum),

N=as.vector(n.per.stratum.per.year),stratum=rep(1:n.strata,n.years),

X=cbind(X.sum,X.sum.year),year=rep(1:n.years,each=n.strata))


####  Make Proximity Matrix
```

```
m=n.strata

strats.coords=coordinates(strats)

m.full=dim(strats@data)[1]

s.idx=rep(0,m)

for(i in 1:m){

  tmp.df=strats@data[strats@data$STR_ID==unique.strata[i],]

  tmp.idx=(1:m.full)[strats@data$STR_ID==unique.strata[i]]

  s.idx[i]=tmp.idx[tmp.df$AREA==max(tmp.df$AREA)]

}

D=as.matrix(dist(strats.coords[s.idx,]))

max.min.dist=max(apply(max(D)*diag(m)+D,1,min))

#D.thresh slightly larger than max.min.dist to insure every stratum has neighbor

D.thresh=7.75

W=ifelse(D>D.thresh,0,1)

diag(W)=0


####  Write Proximity Matrix to File for INLA


D.col=col(D)

file="W.dat"

cat("",file=file, append = FALSE)

cat(m,file=file,"\n",append=TRUE)

for(i in 1:m){

  tmp.neigh=D.col[i,W[i,]==1]

  tmp.n=length(tmp.neigh)

  cat(i,tmp.n,tmp.neigh,"\n",sep=" ",file=file,append=TRUE)
```

```
}


####  Prepare Model 1 Formula for use with INLA


formula.tmp="y~0"

for(j in 1:m){

  formula.tmp=paste(formula.tmp,"+X.",j,sep="")

}

for(j in 1:n.strata){

  formula.tmp=paste(formula.tmp,"+X.",j,".1",sep="")

}

formula.tmp=paste(formula.tmp,"+f(stratum,model='besag',

graph.file='W.dat',replicate=year)+f(year,model='ar1',

replicate=stratum)",sep="")

formula.1=as.formula(formula.tmp)


####  Run ICAR NegBinom and ZINB{1} Trend Model (1)


inla.1.1.out=inla(formula.1,family="nbinomial",E=N,data=counts.sum.reg.df,

control.predictor=list(compute=TRUE),control.compute=list(dic=TRUE))


inla.nb1.out=inla(formula.1,family="zeroinflatednbinomial1",E=N,

data=counts.sum.reg.df,control.predictor=list(compute=TRUE),

control.compute=list(dic=T))


## code for model with independent errors
```

```
formula.tmp="y~"
for(i in 1:(n.strata-1)){
  formula.tmp=paste(formula.tmp,"X.",i,"+",sep="")
}


formula.tmp=paste(formula.tmp,"X.52",sep="")


for(j in 1:n.strata){
  formula.tmp=paste(formula.tmp,"+X.",j,".1",sep="")
}
formula.tmp.2=paste(formula.tmp,"+f(year,model='ar1',
replicate=stratum)",sep="")
formula.2.ind=as.formula(formula.tmp.2)


inla.nb.inderror=inla(formula.2.ind,family="nbinomial",E=N,
data=counts.sum.reg.df,control.predictor=list(compute=TRUE),
control.compute=list(dic=T))


## code for model with independent spatial and temporal errors
formula.tmp="y~"
for(i in 1:(n.strata-1)){
  formula.tmp=paste(formula.tmp,"X.",i,"+",sep="")
}
formula.tmp=paste(formula.tmp,"X.52",sep="")
```

```
for(j in 1:n.strata){

  formula.tmp=paste(formula.tmp,"+X.",j,".1",sep="")

}


formula.3.ind=as.formula(formula.tmp)


inla.nb.inderror.all=inla(formula.3.ind,family="nbinomial",

E=N,data=counts.sum.reg.df,control.predictor=list(compute=TRUE),

control.compute=list(dic=T))


#moran's I test on residuals from model with independent

#spatial errors and AR1 model


mat.moran=matrix(0,52,53)

for(v in 1:53){

  for(k in 1:n.strata){

  mat.moran[k,v]=counts.sum.reg.df$y[k*v]-inla.nb.inderror

  $summary.fixed[k,1]+(inla.nb.inderror$summary.fixed[k+52,1]*v)

  }

  }


results.moran=rep(0,53)

for(a in 1:53){

  moran.tmp=Moran.I(mat.moran[,a],W,na.rm=T)

  results.moran[a]=moran.tmp$p.value

  }
```

```
plot(seq(1,53),results.moran, ylab="p-value from Moran's I test",

xlab="Year")

abline(h=.05,col="red")


## Kriging to look at variograms based on residuals


strats.moran=strats@data[-c(11:13,20,50,57,62,65,68,69,29,30,33,

42,44,60,61,63,66,67,73,74,75),]

coords=coordinates(strats)

coords=as.data.frame(coords)

coords$STR_ID=strats@data$STR_ID

coords$x=coords$V1

coords$y=coords$V2

coords.rm=coords[-c(11:13,20,50,57,62,65,68,69,29,30,33,43,44,

60,61,63,66,67,73,74,75),-c(1,2)]

df.moran=as.data.frame(mat.moran)

df.moran$STR_ID=c(seq(1,18),seq(20,50),75,76,77)

moran.df=merge(df.moran,coords.rm,by= "STR_ID")

center_grid=data.frame(x=moran.df[,55],y=moran.df[,56])


for(i in 1:53){


grid.cnt=matrix(c(moran.df$x,moran.df$y,moran.df[,i+1]),52,3)

geodat=as.geodata(grid.cnt)

summary(geodat)
```

```
ini.val=matrix(c(100000,60000,50000,40000,30000,20000,10000,

5000,1,1,1,1,1,1,1,1),8,2)

varog=variog(geodat,coords=geodat$coords,data=geodat$data,

trend="cte", max.dist=5)

varfit=variofit(varog,ini.cov.pars=ini.val,fix.nug=T)

x11()

plot(varog)

lines(varfit,col=2)

}


### Time series analysis


for(i in 31:52){

win.graph(width=4.875,height=3)

acf(mat.moran[i,],na.action= na.pass,ci.type='ma')

win.graph(width=4.875,height=3)

pacf(mat.moran[i,],na.action=na.pass)

}


####  View Results of Model 1


summary(inla.1.1.out)

summary(inla.nb1.out)

source("plot.coef.0.R")

plot.coef.0(inla.nb1.out,unique.strata)
```

```
plot.coef.0(inla.1.1.out,unique.strata)

x11()

source("plot.trend.inla.4.R")

layout(matrix(1:2,1,2))

plot.trend.inla.4(scaup.df,counts.reg.df,inla.nb1.out,

unique.strata,18,FALSE)

plot.trend.inla.4(scaup.df,counts.reg.df,inla.nb1.out,

unique.strata,45,FALSE)


### Graph to indicate net gain/loss over 52 years

strat.bound=readShapePoly("MAS_stratum_boundaries.shp")

net.loss=matrix(seq(1:52),52,4)

net.loss[,2]=inla.nb1.out$summary.fitted.values[1:52,1]

net.loss[,3]=inla.nb1.out$summary.fitted.values[2705:2756,1]

net.loss[,4]=net.loss[,3]-net.loss[,2]

#color=brewer.pal(6,name="Greys")

plot(strat.bound[-c(11:13,20,50,57,62,65,68,69,29,30,33,

43,44,60,61,63,66,67,73,74,75),],lwd=2)

net.loss[19:49,1]=seq(20,50)

net.loss[50:52,1]=c(75,76,77)

#cooresponding stratum

net.color.1=strat.bound[c(17,26,1,2,10,28),]        #18, 20, 14, 13, 3, 76

net.color.2=strat.bound[c(36,34,18,9,41, 7, 24),]   #26, 25, 15, 17, 27, 16, 77

net.color.3=strat.bound[c(6,48,13,31,21,42,46),]    #6, 40, 5, 22, 7, 37, 28

net.color.4=strat.bound[c(22,35,49,38,27,51,47),]   #23, 50, 35, 31, 24, 29, 33

net.color.5=strat.bound[c(52,4,8,23,39,40,19),]     #38, 4, 10, 21, 36, 32, 1
```

```
net.color.6=strat.bound[c(32,37,45,55,15,53,25),]      #75, 30, 34, 41, 2, 39, 8

net.color.7=strat.bound[c(5,59,72,16,71,70,58),]       #11, 42, 49, 9, 48, 44, 43

net.color.8=strat.bound[c(56,3,54,64),]                #47, 12, 45, 46


plot(strat.bound[-c(11:13,20,50,57,62,65,68,69,29,30,33,43,44,60,61,63,66,67,
73,74,75),],lwd=3)
plot(net.color.1,col="purple4",add=T)
plot(net.color.2,col="royalblue",add=T)
plot(net.color.3,col="seagreen",add=T)
plot(net.color.4,col="seashell",add=T)
plot(net.color.5,col="khaki",add=T)
plot(net.color.6,col="orange",add=T)
plot(net.color.7,col="red",add=T)
plot(net.color.8,col="red4",add=T)
map("world",add=T)


legend(locator(1),legend=c("-25.8 : -3.9","-3.7 : -1.7","-1.5 : -0.5",
"-0.3 : -0.01","0.0 : 0.4","0.4 : 2.1","2.4 : 3.9","4.2 : 6.5"),
fill=c("purple4","royalblue","seagreen","seashell","khaki","orange","red","red4"))


##Significant changes (%95) since 1957
color.1=strat.bound[c(1,9,18,21,26,35,17,24),]
color.2=strat.bound[c(71,64,58,54),]
plot(strat.bound[-c(11:13,20,50,57,62,65,68,69,29,30,33,43,44,60,61,63,66,67,73,
74,75),],lwd=2)
plot(color.1,col="lightblue",add=T)
```

```
plot(color.2,col="red",add=T)

map("world",add=T)

legend(locator(1),legend=c("increasing","no change","decreasing"),

fill=c("red","white","lightblue"))
```