A Spatio-Temporal Model for the State of Wetland Basins in the U.S. Prairie Pothole Region

J. Andrew Royle
U.S. FWS, Div. of Migratory Bird Mgmt.
11510 American Holly Dr.
Laurel, MD
royle@penguin.fws.gov

Christopher K. Wikle
Univ. of Missouri-Columbia, Dept. of Statistics
222 Math Science Bldg.
Columbia, MO 65211
wikle@stat.missouri.edu

1. Waterfowl Habitat in the Prairie Pothole Region

The Prairie Pothole Region (PPR) is a critical breeding ground for many species of North American waterfowl. The U.S. Fish and Wildlife Service (FWS) in conjunction with Canadian, state, and other organizations, spends a great deal of resources monitoring and managing both waterfowl populations and their habitat in the PPR. An important habitat component is the amount of water in the prairie wetland basins (i.e. "potholes") in the spring. This information serves an important assessment role as it is indicative of breeding habitat at this time, and brood rearing habitat in the late summer. Moreover, the ability to quantify long-run characteristics of individual parcels of land is useful for directing conservation and management activities. Finally, there is considerable interest in the impact of climate variation on waterfowl communities *vis-a-vis* habitat (e.g. Sorenson 1998). The goal of our work is to construct a predictive space-time statistical model of the state of wetland basins for use as an assessment and conservation tool, and to facilitate scientific studies of interactions between climate/weather, the wetland resource, and waterfowl populations.

Our data originate from an aerial survey of wetland basins conducted each year by the U.S. FWS over much of the U.S. portion of the PPR. The survey generates digital images of several hundred sample units, each consisting of 10.2 km^2 (4 mi^2) of PPR landscape. A typical plot may contain several hundred wetland basins. The wet surface area of wetland basins within these sample plots is measured. The area of each wetland basin is known from auxiliary survey information.

It is common for wetland basins to be completely dry, and so the data record consists of a large number of zeros (50% or better in some years). Since the wet surface area of a basin is a positive and continuous random variable, it is important that this point mass at 0 be accommodated within any statistical framework. A standard solution is to employ a mixture model wherein the zero and non-zero data are given distinct models. In the present context, we pose a model for the probability that a basin contains water (i.e. a binary wet/dry state variable). We then model the continuous, and strictly positive, wet surface area of "wet" basins.

2. The Model

Let $w_i(s,t)$ be a binary indicator of the wet/dry state of basin i in plot s and year t. We assume that $w_i(s,t) \sim \text{Bernoulli}(\pi_i(s,t))$ with

$$\text{logit}(\pi_i(s,t)) = \alpha(t) + \beta x_i + \lambda(s) p(s,t)$$

where $\alpha(t)$ is a year-specific regional mean, $x_i \equiv \log(\text{area}_i)$ is the logarithm of *basin area*, $p(s,t)$ is the total Jan-May precipitation relevant to basins within plot s in year t, and $\lambda(s)$ are plot-specific precipitation effects. We expect the impact of precipitation to vary spatially due to variation in soil characteristics, land use, and other factors relating to landscape structure.

For $w_i(s,t) = 1$, let $y_i(s,t) > 0$ be the area of water in basin i within plot s and year t. We assume that the ratio y_i/x_i is log-normally distributed. Thus,

$$\log(y_i) \sim \text{Normal}(\log(x_i) + \mu(t) + \gamma(s,t), \sigma^2).$$

where again $\mu(t)$ is a year-specific regional mean, $\gamma(s,t)$ is a space-time process, and σ^2 is *within plot* variation. We assume that

$$\gamma(s,t) = \rho(s)\gamma(s,t-1) + \eta(s,t)$$

where $\eta(s,t)$ is a white-noise process with variance $\text{Var}(\eta(s,t)) = \sigma^2_\eta$. Note the absence of precipitation in this model. Empirical evidence suggests that this is not as important as in the w component of the model. We believe that this is because the y model is primarily driven by the state of large basins, which are not very sensitive to short-run precipitation conditions. On the other hand, the binary model for wet/dry state is primarily driven by small basins, which are highly sensitive to precipitation. Nevertheless, we did consider dependence of y on precipitation and concluded that the model containing only the dynamic process (i.e. γ) was more appropriate statistically, and more likely to facilitate forecasting which is an important goal of our work.

Spatial structure – In the model for w, we have assumed spatial structure in the precipitation effects, $\lambda(s)$. In the model component for y, spatial dependence enters into the model through $\gamma(s,t)$. Instead of modeling spatial correlation in the noise term $\eta(s,t)$, as is conventional, we have chosen to incorporate spatial structure into this model by allowing the temporal dependence parameter, $\rho(s)$, to vary spatially. Our motivation is based on both physical and statistical considerations. Physically, we expect non-separable space-time dynamics as a consequence of variation in soils, geology, water table height, and other physiographic characteristics. A convenient way to induce such structure is based on a hierarchical formulation where the temporal parameters are correlated in space (Wikle et al. 1998). The nature of the spatial structure in both $\lambda(s)$ and $\rho(s)$ is described next.

Model linkage – Intuitively, the two model components should be explicitly linked. That is, we imagine the (w_i, y_i) pair to be a bivariate random variable. We accomplish this in two ways. First, empirical evidence indicates strong association between $\rho(s)$ and $\lambda(s)$. We therefore specify a bivariate spatial model, using the conditional specification given in Royle et al. (1998), as follows. The prior distributions for $\lambda(s)$ and $\rho^*(s) \equiv \text{logit}(\rho(s))$ (assuming $\rho(s) > 0$) are:

$$\lambda(s) \sim \text{N}(\lambda_0 + b_\lambda z(s), \tau_\lambda) \qquad \text{and} \qquad \rho^*(s) \sim \text{N}(\rho_0 + b_\rho z(s), \tau_\rho) \tag{1}$$

where $z(s)$ is a Normal$(0,1)$ spatial process with correlation function $\text{k}_\theta(s,s')$. We use the simple exponential model $\text{k}_\theta(s,s') = \exp(-||s-s'||/\theta)$.

A second linkage between the model components occurs through the regional effects, $\alpha(t)$,and $\mu(t)$. We have chosen a flat prior for $\mu(t)$. Then, our prior for $\alpha(t)$ is $\alpha(t) \sim \text{N}(\alpha_0 + b_\alpha \mu(t), \tau_\alpha)$ which accommodates correlation between annual averages of w_i and y_i via the parameter b_α.

Model fitting and prediction were carried out using Markov chain Monte Carlo (MCMC) methods. We omit those details here.

3. Results and Discussion

Our model may be used as an assessment tool for mapping the state of the wetland resource in order to inform conservation and management activities. Moreover, the ability to forecast waterfowl habitat is of obvious management utility. Our model facilitates both activities. In particular, the data exhibit substantial spatial structure. The latent spatial process (i.e. $z(s)$ from 1) estimate is shown in Figure 1 (top panel). In terms of forecasting, the strong dynamical component of the model (the average value of $\rho(s)$ was 0.48), in conjunction with explicit accounting for precipitation in the wet probability model component, may be combined with a model for precipitation to yield $t+1$ forecasts of a "water map". We are currently pursuing this application.

The model suggests a strong regional signal shared by both model components. Estimates (posterior means) of $\mu(t)$ were highly variable, ranging from -3.73 to -0.1244. The posterior means of b_α and τ_α (the inverse of the variance) were 1.15 and 41.3, respectively, and the correlation between the two sets of estimates was approximately 0.98. One might think that conditioning on precipitation (in the model for w) should mitigate this regional signal, since the year-effect estimates mimic the wet/dry pattern observed in the PPR from 1987-1999. However, even when we fit the model with precipitation in both components (i.e. w *and* y instead of just w as reported above), the year effects were essentially unchanged. We believe this to be due to non-linear relationships between precipitation and basin state, and are investigating parameterizations which accommodate this.

Finally, we are developing this model as a tool for examining wetland response to climate change. Because our model represents a more accurate description of the wetland resource than other analyses which rely simply on wet basin totals (e.g. Sorenson 1998), we feel that a better understanding of climate-wetland relationships can be achieved. There is some indication that the model effects are tied to climate dynamics. For example, the fitted regional signal is plotted with the North Atlantic Oscillation (Hurl 1995) in Figure 1.

ACKNOWLEDGEMENTS

We thank Noel Cressie, L. Mark Berliner, Mark Koneff and Ron Reynolds for their ongoing contributions to this work. This research has been partially supported by a grant from the U.S. EPA's STAR program, Assistance Agreement No. R827257-01-0.

REFERENCES

Hurrell, J.W. 1995. Decadal Trends in the North Atlantic Oscillation Regional Temperatures and Precipitation. *Science*, **269**, 676-679.

Royle, J.A., L.M. Berliner, C.K. Wikle, and R.F. Milliff. 1998. A hierarchical spatial model for constructing scatterometer-derived wind fields, pp. 367-382 in *Case Studies in Bayesian Statistics Volume IV*, Springer-Verlag.

Sorenson, L.G., R. Goldberg, T.L. Root, and M.G. Anderson. 1998. Potential effects of global warming on waterfowl populations breeding in the northern great plains. *Climatic Change*, **40**, 343-369.

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

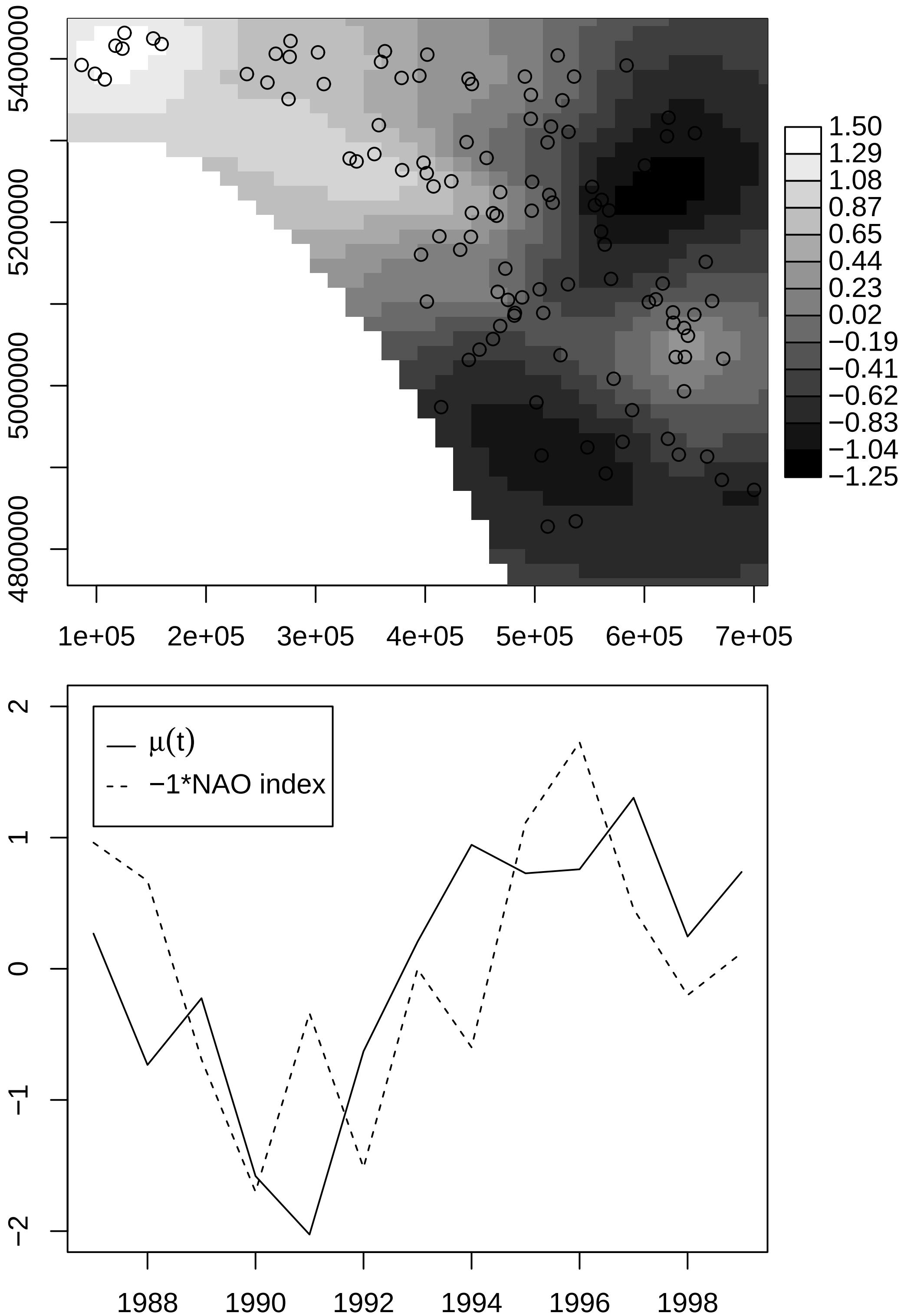

Figure 1: Top panel: Estimated spatial process, Z(s) (circles are data locations); Bottom panel: standardized $\mu(t)$ vs. the negative of the North Atlantic Oscillation index.