

AUTOCORRELATION IN RESOURCE SELECTION FOR CHINOOK–SALMON SPAWNING HABITAT

JEAN-YVES P. COURBOIS¹, NOAA Fisheries, Northwest Fisheries Sciences Center, Conservation Biology Division, 2725 Montlake Blvd. East, Seattle, WA 98112-2097

Abstract: When sampling natural populations, habitat association models can help solve difficult design questions such as stratification and sample allocation. However, these models often fail to account for spatial correlation (autocorrelation). I present two examples of habitat association studies that include a Markov term for autocorrelation. The examples concern spawning–habitat selection of chinook salmon in the Pacific Northwest, USA. Spawning surveys are often censuses of large stream reaches and thus produce a nice response for habitat–association modeling. I show how these models help refine sampling designs. Finally discussion of other types of spatial models, both for presence/absence and counts concludes this paper.

Key words: resource selection, chinook salmon, redds, Middle-Fork Salmon River, Stillaguamish River, Markov models, model-assisted sampling designs.

Habitat–association models relate habitat characteristics to the presence or the number of animals in a region to determine which characteristics the animals are choosing. A popular technique is the use of generalized-linear models either with a Bernoulli (occupancy) or a Poisson (amount of use) response. These models usually do not account for spatial correlation between sites (autocorrelation). Often, independence is justified by the sampling design; however one cannot rely on the sampling design when a census is taken. This paper describes a class of methods to account for autocorrelation in the case when the data are a census.

Spatial correlation may arise in habitat studies for two reasons. First, if the model lacks some important habitat data, these missing data might cause a pattern in the animal population that can be modeled with space as a surrogate. Second, autocorrelation could result from social interactions. For example even in a homogeneous habitat many animals will cluster for protection from predators or, as in this case, as a mating strategy. These social interactions can be modeled with an autocorrelation term. Unfortunately, models cannot distinguish between these reasons, but often the reason is not as important as simply compensating for autocorrelation. I am interested in the autocorrelation to revise sampling designs, so simply knowing its' extent is enough. In other studies autocorrelation may be a nuisance; there testing for autocorrelation and finding it not significant lends credence to the independence assumed in most efforts.

For normally distributed data, statistical autocorrelation models are well known in the field of geostatistics. These methods can be used in cases where resource–use is a density or a count that is adequately large. I will not cover these methods since there is adequate literature on the subject (Cressie, 1993; Ripley, 1981). Models for presence/absence or count data are less well known and will be the focus of this paper. Correlation in presence/absence and count data is defined identically as in a continuous response: $\text{cov}(X, Y) / (\text{var}(X)\text{var}(Y))^{0.5}$.

I present two examples of Markov models for use in habitat–association studies. The first is a partially–ordered Markov model for chinook salmon redds (spawning beds) in a river network, the Bear-Valley and Elk creeks, Idaho. The second is a hidden Markov–chain model for chinook redds along a channel of the Stillaguamish river, Washington. When using Markov models, one specifies a conditional probability function rather than a marginal distribution. The Markov assumption states that this conditional distribution only depends on other units within a neighborhood. That is, the conditional distribution of a creek or river segment's occupancy, Y_u , given all other segments' occupancy, $y_{-u} = \{y_v : v \neq u\}$, is a function of only those segments in the neighborhood of u , N_u ,

$$P(Y_u = y | y_{-u}) = P(Y_u = y | y_v, v \in N_u).$$

Under some regularity conditions (not complicated) specifying these conditional distributions is all that is necessary because a joint distribution, $P(Y_1, \dots, Y_N)$, is known to exist and in many cases is tractable. One could estimate these conditional probabilities directly or parameterize them with some functional form and estimate the ensuing unknowns.

¹E-mail: pip.courbois@noaa.gov

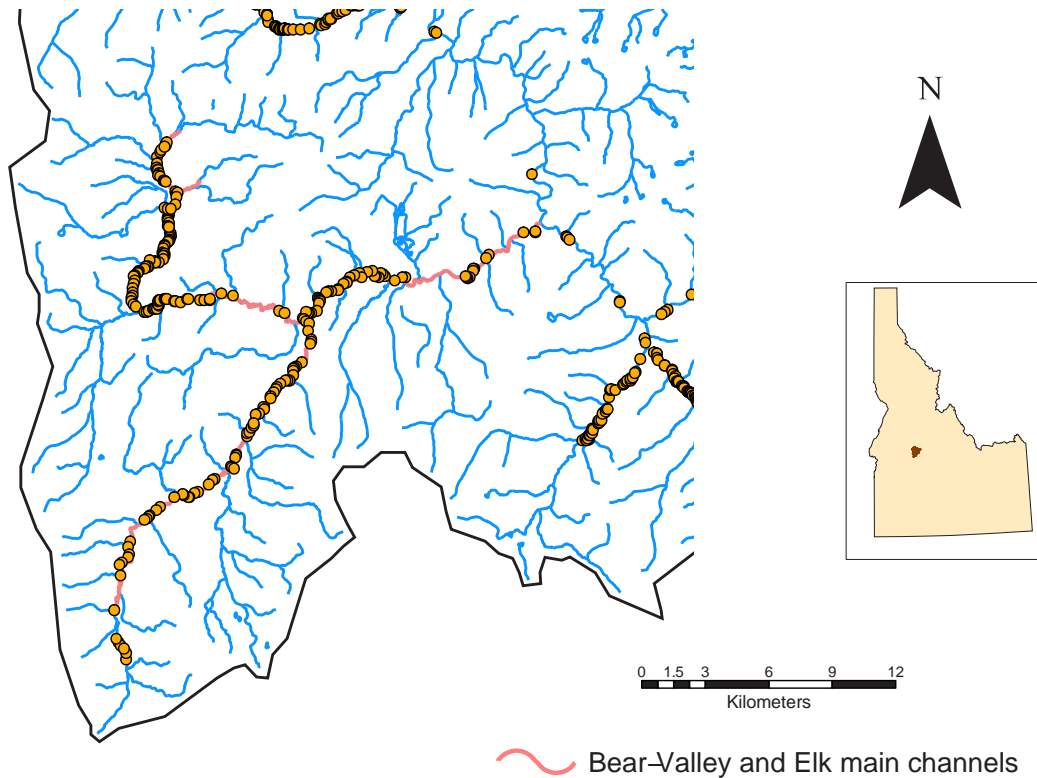


Figure 1: Redd locations on the main stems (pink line) of the Bear-Valley and Elk creeks in the middle-fork of the salmon river.

Chinook salmon are the subject of both examples. Chinook are an anadromous species; they spawn in freshwater rivers but live most of their adult lives in the ocean. After birth and one or two years in fresh water they migrate to the Pacific Ocean and spend the next 3–5 years there. When they are ready to spawn, they migrate back up the river system to the location where they were born. At the conclusion of this spawning migration, adult females prepare a spawning bed, a redd, in a stream area with suitable habitat. This suitable habitat is what I am modeling. For more information on chinook species see Myers et al. (1998).

First I present the two examples and then describe how we hope to use these studies to refine sampling designs for chinook–salmon redds. Finally I conclude with some suggestions for extending the models I presented and other types of models for spatial correlation.

REDDS ON THE BEAR–VALLEY AND ELK CREEKS, IDAHO

Bear-Valley and Elk Creeks are tributaries of the Middle-Fork Salmon River, a National Wild and Scenic River, of central Idaho. They flow north–northeast through the Frank Church River of No Return Wilderness (Figure 1). Because redding is a main-channel process, I have restricted my study to the main channels of these creeks, which includes about 70 kilometers of stream. For a complete description of the Middle-Fork drainage and further references see Thurow (1999).

The redd locations were digitized from maps published by Thurow (1999). These data were collected by helicopter surveys of the entire Middle-Fork drainage. Narrow areas and those with a large amount of canopy cover were surveyed by crews on the ground. The amount and direction of measurement error both for the helicopter and ground surveys are unknown. I assume that this measurement error is negligible. For more information about the sampling and the data see Thurow (1999).

In order to relate the salmon redds to habitat characteristics, I broke the river network into 708 discrete

Table 1: Habitat data and their sources.

data	coverage source
Elevation (meters)	Digital elevation model
30 year average precipitation	PRISM modeling system
30 year average maximum temperature °C	PRISM
30 year average minimum temperature °C	PRISM
Habitat classification: 16 classes	Northwest Habitat Institute
Geology/Lithography: 12 classes	

Table 2: Frequencies of redd occupied segments by habitat–classification. The column labelled I_k contains the indicator variable for the model 2.

Habitat	# segments	# occupied	I_k
Eastside (interior) mixed conifer forest	22	0	reference
Grass, shrub, and/or regenerating forest	45	11	I_1
Herbaceous wetlands	758	166	I_2
Montane coniferous wetlands	46	5	I_3
Montane mixed conifer forest	22	4	I_4

100 meter segments arbitrary starting from where Bear–Valley creek meets the Middle–Fork Salmon River. For each segment I recorded the number of redds present, habitat characteristics, and the index of the segment downstream. The assignment of a downstream segment called for a decision when a confluence occurred in the center of a segment (which was inevitably the case). Arbitrarily I decided that segments that contain a confluence would be the downstream segment for any tributary segment irrespective of where the confluence occurs (Figure 2). This decision is likely to result in some estimation errors because if a segment has a tributary at its lower end the two might not be highly correlated.

Habitat Characteristics

I acquired all habitat data from publicly–available large extent geographic information system (GIS) data layers (Table 1). For the habitat data that were polygon coverages, I assigned the value from the polygon that included the segment’s center–point. Elevation and land cover were raster data, in these cases the segment received the value of the grid point closest to its center–point. These are probably not optimal assignments, but I wanted methods that could be automated for reasons that will be seen later. I collapsed some habitat–classification and geology classes due to their rarity. Collapsing led to five categories of habitat–classification and four categories of geology (Tables 2 and 3).

The model

The model for the occupancy of a river segment is a logistic–regression model with a partially–ordered Markov structure for autocorrelation. Defining $y_u = 1$ if redds are in segment u and $y_u = 0$ otherwise, the y_u each have a Bernoulli distribution with parameter π_u which is the probability of occupancy. This probability depends on the habitat data and the occupancy of the neighborhood through the logistic link function:

$$\log\left(\frac{\pi_u}{1 - \pi_u}\right) = \alpha + x'_u\beta + \gamma(N_u) \quad (1)$$

where x_u is a vector of the habitat data, β are unknown parameters, and $\gamma(N_u)$ represents a function of the occupancy of the neighborhood for segment u , N_u .

The simplest non-empty neighborhood, and the one I adopt, contains only the segment directly downstream (Figure 2). With this neighborhood structure, the network of segments is a directed acyclic graph

Table 3: Frequencies of redd occupied segments by geology.

Geology	# segments	# occupied
Alluvium	111	30
Glacial Drift	175	121
Calc-alkaline Intrusive	359	51
Mafic Gneiss	63	28

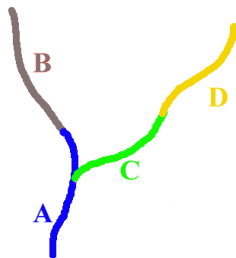


Figure 2: Neighbors are the stream segment directly downstream. Thus A does not have a neighbor, B and C share the same neighbor, A ; and D has C as its neighbor, that is $N_A = \emptyset$, $N_B = N_C = A$ and $N_D = C$. With this neighborhood structure the network of segments is a directed acyclic graph.

where each segment has at most one neighbor and only one segment (the mouth of the watershed) has no neighbors. I chose such a simple neighborhood for computational simplicity; it has nothing to do with the distribution of redds in the watershed.

Given this neighborhood structure, the parameterization is linear in the occupancy of the stream segment downstream, that is the $\gamma(N_u)$ in equation 1 is simply $\gamma \times y_{u-1}$ where γ is an unknown real-valued coefficient and y_{u-1} is the occupancy of the segment downstream. This model is an example of a partially-ordered Markov model introduced by Davidson et al. (1999); they demonstrate that the conditional probabilities lead to a well defined joint probability distribution and derive this likelihood from these probabilities, $L(Y) = P(Y_1) \prod_{i=2} P(Y_i | Y_j, j \in N_i)$. It is useful to think about these models as fitting Markov chains to each stream route. In other words, beginning with the first segment (the confluence with the Middle-Fork, Y_1) the model assumes a Markov chain proceeding up the main channel to a headwater segment. Then with each tributary in turn, the model assumes a Markov chain conditioning on the segment in the main channel, which is already specified, extending up that tributary's main stem to another headwater segment. Then it proceeds to the next tributary. This is useful because the theory of and inference for Markov chains is well developed.

This leaves the decision of how to specify the distribution of the first segment in the entire watershed, Y_1 . I conditioned this segment on the segment downstream which is in the main channel of the middle-fork of the Salmon River. In 1998, this segment did not have any redds present. By conditioning on this segment I have left open the possibility of expanding the model to the entire middle fork.

Under some regularity conditions, parameter estimates are asymptotically normal and likelihood ratio tests can be used to test parametric models against submodels using an asymptotic chi-squared distribution (Guttorp, 1995). These tests include the familiar tests to refine the model as well as the test for independence, that is whether the autocorrelation term is zero ($\gamma = 0$).

Table 4: Drop-in-deviance table for main effects of the habitat characteristics. Neighbor refers to the autocorrelation term. Bold highlighting indicates statistical significance at the 0.05 level.

Term	degrees-of-freedom	Drop-in-deviance	p-value
Elevation	1	1.257	0.26
Precipitation	1	0.001	0.97
Min. temp.	1	0.003	0.95
Max. temp.	1	4.332	0.037
Habitat-class.	5	19.128	0.002
geology	3	3.400	0.33
Neighbor	1	6.189	0.013

Results from the Bear-Valley analysis

Using indicators for the habitat-classification and geology, twelve predictor variables including the occupancy of the stream segment downstream were candidates for inclusion in the model. Non-significant terms (at the 0.05 level) were eliminated from a model (Table 4) resulting in the model containing habitat-classifications, maximum temperature, and the autocorrelation term. All interactions and quadratic terms were non-significant. The drop-in-deviance from the no-selection model is 27.42 on 6 degrees of freedom. The estimated model is

$$\log\left(\frac{\pi_u}{1-\pi_u}\right) = -11.481 + 12.247I_1 + 12.060I_2 + 10.975I_3 + 11.958I_4 - 0.187x_{T_{\max}} + 0.363y_{u-1} \quad (2)$$

where $I_1 \cdots I_4$ are indicators for habitat classes (Table 2) and $x_{T_{\max}}$ is the maximum temperature.

The non significance of the elevation term may surprise, though in the small section I studied elevation varied little. Also elevation and the two temperatures were highly correlated, so all three are unlikely to appear in the model together. A biological reason should settle which of these multicollinear variables remains in the model. For my purpose I kept the one which was significant with the others in the model.

Plots of these results are handy (Figure 3). These estimates lead to estimates for the joint distribution $\hat{P}(Y_1, Y_2, \dots, Y_N)$ and the autocorrelation function, $\widehat{\text{corr}}(Y_u, Y_{u+v}) = \left[\hat{P}(Y_i = 1 | Y_{i-1} = 1) - \hat{P}(Y_i = 1 | Y_{i-1} = 0) \right]^v$.

REDDS ON THE STILLIGUAMISH RIVER, WASHINGTON

The Stillaguamish is located along the southwest margin of the North Cascades, approximately 85 km northeast of Seattle. The drainage area is over 300 km² and is a low-gradient (< 0.01) meandering gravel-bed channel. Over 70% of the watershed is classified as forest lands, while the floodplain is a mix of forested floodplain, agricultural lands, and rural residential. The redds were surveyed from helicopter, from the river bank on foot, and from boats. Redds are marked when counted and the site is surveyed several times during the redding period. For a complete description of the Stillaguamish and the methods used see Hayman (2001).

At this time no habitat data are available for this stretch of river. Thus I use an autocorrelation term (space) as a surrogate for habitat by creating a binary random variable that indicates good or poor habitat. Once again I have divided this stream reach into 253 one-hundred meter segments and counted the number of redds in each. Again I model occupancy this time using a hidden-Markov chain model.

The model

The hidden-Markov model includes a binary vector of *observed* random variables, $Y_u, u = 1, \dots, N$, which are the absence/presences of redds and a vector of *unobserved* (hidden) random variables, $C_u, u = 1, \dots, N$, which indicate poor ($C = 0$) or good ($C=1$) habitat. Occupancy, Y_u , has a Bernoulli distribution where the parameter depends on whether the respective $C_u = 1$ or 0, that is $P(Y_u = k | C_u = j) = \pi_j^k (1 - \pi_j)^{1-k}$ for $j, k \in 0, 1$. Further the unobserved habitat variable is assumed to be a two-state one-step Markov chain,

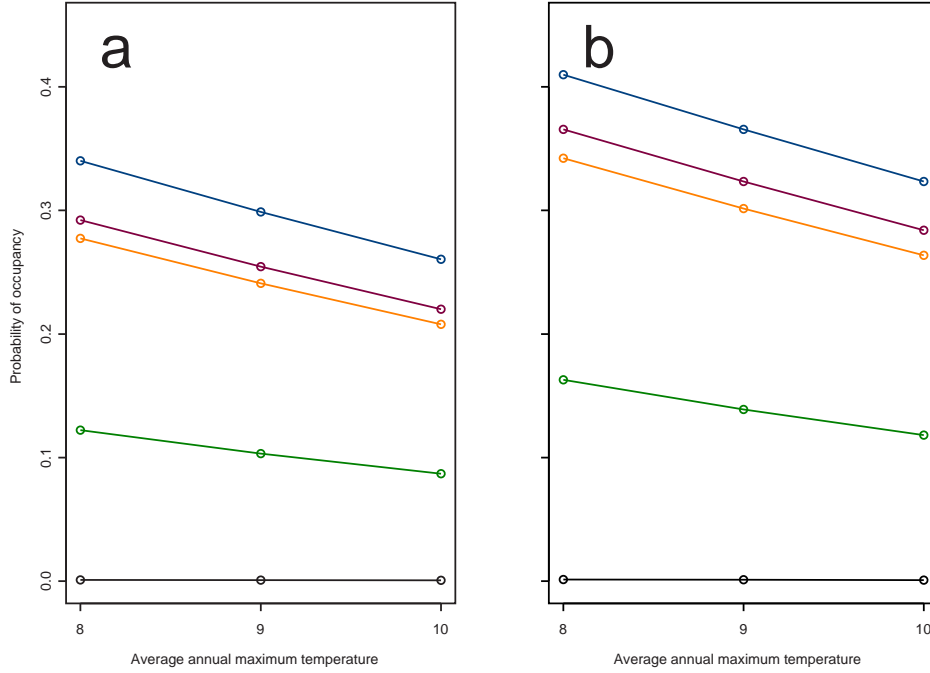


Figure 3: Estimates of the conditional probability of occupancy given the neighbor segment when (a) the neighbor is not occupied, $\hat{P}(Y_i = 1 | Y_{i-1} = 0)$ and (b) the neighbor is occupied, $\hat{P}(Y_i = 1 | Y_{i-1} = 1)$. Each line is a different habitat–classification, from the top in both plots: grass, shrub, and/or regenerating forest; herbaceous wetlands; montane mixed conifer forest; montane coniferous wetlands; and eastern (interior) mixed conifer forest.

that is $P(C_u = j | c_{-u}) = P(C_u = j | C_{u-1} = i) = p_{ij}$. For this model instead of parameterizing the one-step transition probabilities, p_{ij} , I estimate them directly.

Despite having a completely unobserved vector of random variables, the likelihood can be specified and maximized. This is possible because the likelihood is marginalized over the hidden variable (MacDonald and Zucchini, 1997, Section 2.5). Define this likelihood as $L_N = P(Y_1, \dots, Y_N; \Gamma, \lambda)$ where Γ is the transition probability matrix and λ represents the probabilities for the occupation process. The joint probability of the Y_u , $u = 1, \dots, N$ by conditional independence, is $P(Y_1, \dots, Y_N | C_1, \dots, C_N) = \prod_{u=1}^N \pi_{y_u, i_u}$ where $\pi_{y_u, i_u} = P(Y_u = y_u | C_u = i_u)$. Marginalizing, by multiplication by $P(C_1, \dots, C_N)$, and summing gives

$$L_N = \sum_{i_1 \in \{0,1\}} \cdots \sum_{i_N \in \{0,1\}} \left(\prod_{n=1}^N \pi_{s_n i_n} \right) P(C_1, \dots, C_N).$$

Now C_1, \dots, C_N is a Markov chain so that $P(C_1, \dots, C_N) = \delta_{i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N}$, where δ is the marginal distribution of C_1 . Putting these results together results in the likelihood:

$$\begin{aligned} L_N &= \sum_{i_1 \in \{0,1\}} \cdots \sum_{i_N \in \{0,1\}} \pi_{s_1 i_1} \cdots \pi_{s_N i_N} \delta_{i_1} \gamma_{i_1 i_2} \cdots \gamma_{i_{N-1} i_N} \\ &= \delta \lambda(s_1) \Gamma \lambda(s_2) \Gamma \cdots \lambda(s_N) \mathbf{1}' \end{aligned}$$

where $\lambda(s) = \text{diag}(\pi_{s_0}, \pi_{s_1})$ (MacDonald and Zucchini, 1997). This matrix–based equation can be maximized numerically; I used the SPlus `n1minb` function.

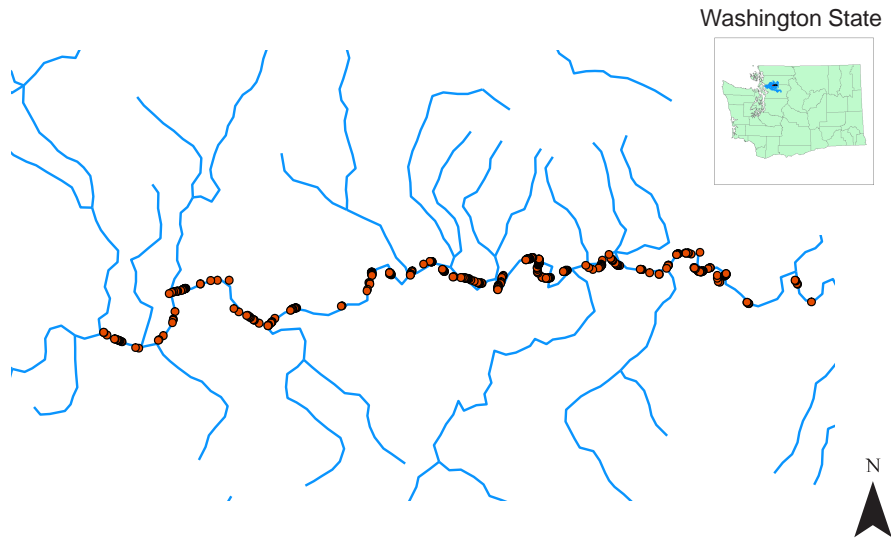


Figure 4: Stillaguamish river, Washington with redd locations in 1998.

Results from Stillaguamish Analysis

The resulting estimates are

$$\hat{\Gamma} = \begin{bmatrix} 0.740 & 0.260 \\ 0.202 & 0.798 \end{bmatrix}$$

and $\pi_{01} = 0.0001$ and $\pi_{11} = 0.75$. These are interpreted as follows. If a segment is bad habitat then the probability that the next segment will be bad is 0.740 while the probability that it will be good is 0.260. Likewise, the probability that good habitat will follow good is 0.798 and bad will follow good is 0.202. The probability that a good habitat segment will be occupied is 0.75, while the probability that a bad habitat segment is occupied is only 0.0001. Notice that this model is not just assigning good habitat to locations where redds are present, it predicts that about 75% of good habitat and 0.001% of poor habitat is occupied. The model smooths presence/absence data to predict good/poor habitat. From these estimates, one can calculate the v -step transition probabilities (Figure 5) and reconstruct the $C_u, u = 1, \dots, N$ or habitat vector.

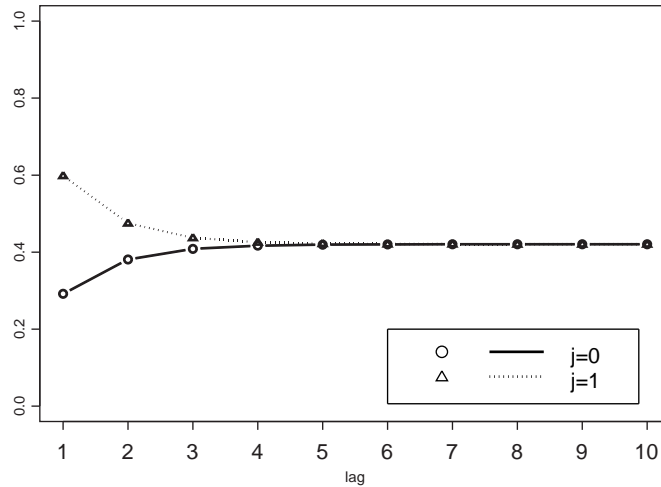
FROM HABITAT-ASSOCIATIONS TO SAMPLING DESIGNS

My interest in these models is to refine sampling designs for chinook salmon redds making it possible to estimate the number of redds in a region where taking a census is too expensive. Despite using models, I want to use probability sampling designs because they result in unbiased estimators and defensible results. The difficulty of sampling redds is that they are rare and cluster, so if one were to sample them using simple random sampling most visited sites would be empty and the design inefficient. My models should suggest alternatives such as stratified and adaptive designs.

A model-based approach uses the model results to predict the optimal sites to sample (optimality might be measured by the probability of sampling units containing redds). The problem with this approach is that the models are not ideal for prediction and the resulting design is not a probability sampling design and depends on the model being correct (unlikely, to say the least). Another approach and the one I prefer uses the model to answer sampling questions, a model-assisted approach. Although the details of this procedure are a topic for another paper; here I give some rough guidelines.

First consider stratification. Habitat variables that do not significantly predict salmon redd occurrence are unlikely to be good variables to stratify by. The first example above demonstrates that neither geology nor elevation should be used to stratify; while habitat-classification could be used perhaps with three strata

Figure 5: The v -step transition probabilities $P(Y_{u+v} = 1 | Y_u = j)$.



based on the three groups of habitat variables visible in Figure 3. One stratum consists of the three habitat classifications whose lines are at the top of the plots, the second consists of the habitat classification fourth from the top, and the third the habitat classification that has very few redds at the bottom of the plot. These models should also help in sample allocation; for instance one might oversample the strata with many redds.

The autocorrelation term can be used to address whether one should use an adaptive design. Supposing one samples a unit. Can the presence/absence of redds in that unit help determine where to sample next? This is the idea behind adaptive sampling (Thompson and Seber, 1996). The v -step transition probabilities, $P(Y_{u+v} = 1 | Y_u = j)$, can help answer this question. If these probabilities show no pattern over distance then simple random sampling should be fine. On the other hand, if these functions are different for occupied vs. non-occupied sampled units then an adaptive design is suggested. The data from the second example suggest an adaptive design where one samples the adjacent unit or perhaps the adjacent two units if the sampled unit is occupied and one selects a new random unit if not (Figure 5).

Because I am interested in using these designs over large areas, I chose habitat characteristics that have data that cover large areas. Models based on these data are unlikely to predict the location of redds precisely. Some measure of upwelling water would likely predict the occupation of segments better than the habitat characteristics I have used (Geist, 2000). However, large scale maps of hyporheic discharge do not exist and thus this upwelling, although important for the salmon, cannot help refine sampling designs over large areas.

DISCUSSION

I have demonstrated how the addition of autocorrelation is possible and useful in habitat-association studies when the data are a census of a region. I supplied two examples using adaptations to the one-step Markov chain model for binary data. In these cases spatial autocorrelation is important because little or no habitat data are available and chinook salmon build redds in clusters. The distinction between these models and those used in traditional resource-selection functions is that I am specifying *conditional* selection probability functions rather than marginal distributions with an independence assumption.

Many extensions are possible from these models. One could define more complicated neighborhoods using spatial models (Besag, 1974) such as specifying both the segments upstream and downstream as the neighborhood, though for the first example it is not clear how to parameterize the conditional probabilities. The model would need a coefficient for the segment downstream and, depending on whether the segment is

a headwater or confluence, zero, one, or two coefficients for the segment(s) upstream (equation 1). One can also extend these models to those that are based on a Poisson or negative-binomial distribution and model counts. The Poisson model with spatial dependence, however, is an example of a case where one might specify conditional distributions that do not lead to a joint distribution (Besag, 1974). The danger here is that one could write down the likelihood and obtain estimates for a process that does not exist.

In the Stillaguamish, some habitat variables could be introduced. Covariates could affect the hidden habitat variable or the observed occupation variable. One could make arguments for both cases. If one were to use the habitat variables to affect the occupancy, then the hidden variable might indicate clusters rather than habitat. A variable describing whether a segment belongs to a cluster or not could be very useful to describe cluster sizes and the scale of the process. Another extension for the Stillaguamish data is to have more than two states for the hidden variable. One might model good, fair, and poor habitat with a three state (hidden) Markov chain model. Also, once again, one might want to model the redd count rather than presence/absence.

The models I have introduced are based on Markov chains, that is for the most part defined on a one-dimensional population. However spatial analysts have developed many models based on the Markovian theme for analysis of spatial data (Besag, 1974; Hoeting et al., 2000). In general these models are often defined on graphs where the random variables are the nodes and dependence is specified by the vertices. The models are specified with conditional distributions as I have done and have a Markovian dependence structure. These models are straight forward when the data are in a lattice structure (once again a census), however most resource selection studies are samples of used and available units and thus the researcher must decide how to define neighborhoods, a non-trivial task (Augustin et al., 1996; Besag, 1974). Another complication is that likelihoods are often difficult or impossible to write down. However this should not prevent a realistic model from being used as many of these likelihoods can be optimized using Markov-chain Monte-Carlo methods.

Finally other models exist for spatial modeling of binary or count data. Bayesian models use a spatial model as a prior distribution (Besag et al., 1991; Hoeting et al., 2000). The mixed-effects log-linear and logistic models developed by Wolfinger and O'Connell (1993) should be easily adaptable to spatial analysis by specifying a spatial variance-covariance matrix as Diggle et al. (1998) have done with the kriging model.

Caveats: The purpose of this paper is to introduce autocorrelation for resource selection studies and give some examples. In the examples I fit the habitat-association model on very small data sets using large-scale habitat data. Further, the data from the first example are read from published maps. One should not interpret the models as reflecting chinook salmon biology at this time. Further work refining the models and deriving sampling designs is underway.

ACKNOWLEDGEMENTS

Research conducted while supported as a postdoc by the Cumulative Risk Initiative of NOAA Fisheries' Northwest Fisheries Science Center, the National Science Foundations' Grants for Vertical Integration of Research and Education in the Mathematical Sciences (VIGRE) at the University of Washington, the Environmental Protection Agency's cooperative agreement CR825173-01 to the National Research Center for Statistics and the Environment at the University of Washington, and through cooperative agreement CR822-9096-01 to Oregon State University and its subagreement E0101B-A to the University of Washington. Also for data, technical help, and reviews: Blake Feist, David Jensen, Chris Jordan, Richard Hicks, Steve Katz, George Pess, Ashley Steel, Russ Thurow, and the two anonymous reviewers. Although the research described in this article has been funded in part by the United States Environmental Protection Agency it has not been subjected to the Agency's required peer review and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

LITERATURE CITED

- Augustin, N., M. Muggleston, and S. Buckland 1996. An autologistic model for the spatital distribution of wildlife. *Journal of Applied Ecology* 33, 339-347.
- Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. series B (Methodological)* 36(2), 192-236.

- Besag, J., J. York, and A. Mollie 1991. Bayesian image restoration, with two applications in spatial statistics (Disc: p21-59). *Annals of the Institute of Statistical Mathematics* 43, 1–20.
- Cressie, N. 1993. *Statistics for Spatial Data*. New York: John Wiley & Sons, Inc.
- Davidson, J., N. Cressie, and X. Hua 1999. Texture synthesis and pattern recognition for partially ordered markov models. *Pattern Recognition* 32, 1475–1505.
- Diggle, P. J., J. A. Tawn, and R. A. Moyeed 1998. Model-based geostatistics (Disc: p326-350). *Applied Statistics* 47, 299–326.
- Geist, D. R. 2000. Hyporheic discharge of river water into fall chinook salmon (*Oncorhynchus tshawytscha*) spawning areas in the Hanford Reach, Columbia River. *Can. J. Fish. Aquat. Sci.* 57, 1647–1656.
- Guttorp, P. 1995. *Stochastic Modeling of Scientific Data*. Chapman and Hall.
- Hayman, B. 2001. Washington state chinook salmon spawning escapement assessment in the Stillaguamish and Skagit Rivers, 1998. Completion report, U.S. Chinook Technical Committee and U.S. National Marine Fisheries Service, Fish Management Program, Washington Department of Fish and Wildlife.
- Hoeting, J. A., M. Leecaster, and D. Bowden 2000. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics* 5(1), 102–114.
- MacDonald, I. L. and W. Zucchini 1997. *Hidden Markov and Other Models for Discrete-valued Time Series*. 2-6 Boundary Row, London: Chapman and Hall.
- Myers, J., G. Kope, D. Bryant, L. Teel, T. Lierheimer, W. Wainwright, F. Grant, K. Waknitz, S. Neely, S. Lindley, and R. Waples 1998. Status review of chinook salmon from Washington, Idaho, Oregon, and California. NOAA Technical Memo NMFS-NWFSC-35, U.S. Department of Commerce. www.nwfsc.noaa.gov/pubs/tm/tm35/chinook.pdf.
- Ripley, B. D. 1981. *Spatial Statistics*. New York: John Wiley & Sons.
- Thompson, S. K. and G. A. F. Seber 1996. *Adaptive Sampling*. Wiley.
- Thurrow, R. F. 1999. Dynamics of chinook salmon populations within Idaho's Frank Church Wilderness: Implications for persistence. *USDA Forest Service Proceedings* 1(1), 1–9.
- Wolfinger, R. and M. O'Connell 1993. Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 48, 233–243.