Hierarchical modeling of resource utilization using MCMC

Kacy Cook (2/20/06)

Based on work by: Mike Conroy, John Carroll, Kacy Cook, Chris Fonnesbeck, and Jim Peterson

Warnell School of Forest Resources, University of Georgia

USGS Georgia Cooperative Fish & Wildlife Research Unit

Predicting resource selection / use at multiple scales

Theory
- **Animals select / use resources at multiple-spatial scales (Johnson 1980)**

The disconnect …
- **Common approach cannot be used for prediction**
  - **Compositional analysis (Aebischer et al. 1993)**
    - **» multi-scale approach to date**

Introduction

However, we need to predict habitat use probabilities and display them spatially to aid decision making …

Maps of predicted habitat use aid policy decisions …

Data and modeling approaches for multivariate predictive habitat models

- **Presence and absence data**
  - **True absence is difficult to document**
    - **Temporal and spatial scale**
    - **Probability of detection required**
    - **Examples**

- **Presence only data (radio-telemetry)**
  - **Ecological Niche Factor Analysis (Hirzel et al. 2002) – empirical, multivariate normality**
  - **Mahalanobis distance (Clarke 1993) – empirical, multivariate normality**
  - **Habitat Suitability Index – hypotheses validated with data**

****** These do not allow for a measure of habitat availability, and apart from Mahalanobis distance, do not allow for model selection**

Logistic Regression

- **Allows for model selection and incorporation of model uncertainty**
- **Provides a measure of habitat availability**
- **Habitat at points of presence compared to “absence” points**
- **Some key assumptions:**
  - **Occurrences are rare or clumped**
  - **Data are not autocorrelated**

(Keating and Cherry 2004)

Hierarchical Logistic Regression

- **Frequentist hierarchical models are not easily modeled, covariance estimation is complicated**

- **Frequentist Logistic regression assumes multivariate normality**

Utility of Bayesian logistic regression for modeling habitat: Philosophical and practical reasons

See Hierarchical Bayesian Modeling Group at Warnell for more info:
http://fisher.forestry.uga.edu:9673/bayes/seminar

- **Explicit estimation of uncertainty, important to decision analysis (i.e. cost – benefit analysis) – parameters are random variables (AIC is also)**

- **Use prior information in posterior distribution (if you have prior info)**

- **Intuitive probability**

  - **not the probability that if an experiment were repeated, 95% of the time the mean would be in range….. (gasp)**

- **Update information over time as more data become available**

- **No assumption of multivariate normality**

Simulation of parameter distributions using Markov Chain Monte Carlo Techniques

- Why use Markov Chain Monte Carlo and what the heck is it?
- A way of obtaining probability values from complex integrals (area under the curve)
- Stochastic process, an ordered collection of random variables (random walk)
- $Pr(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = Pr(X_{t+1} = x_{t+1} | X_t = x_t)$
- Each value in a Markov Chain depends only on the previous value
- The probability of getting from one value in a chain to any of the previous values is equal (irreducibility)

Modeling process of one MCMC simulation

1) Select 1 random presence point from point cloud per location per season & 1 random absence point = simulation dataset (fixed)
2) 10,000 iterations of the Metropolis Hastings (MH) algorithm
   - each MH iteration updates the posterior distribution of the betas, based on an acceptance criteria
3) Result is a posterior distribution of beta estimates and deviance or AIC

Radio-telemetry data

- **Presence-only data**
- **Telemetry error**

More chance animal is pin-pointed when the standard deviation of field – measured telemetry error is used instead of the 95% error ellipse (based on Kenow et al. 2001)

- **Usually autocorrelated to some degree**
  - **Assumption of location independence based on traversal of home range is subjective**
    - **Where you are always depends on where you were unless it's been a really long time (or you can drive a car)**

Objectives

- **Construct and validate predictive bear habitat models & maps for central Georgia**
- **Identify areas for habitat management & possible reintroduction**
- **Incorporate telemetry error uncertainty**
- **Model presence only habitat data**
- **Condition the probability of habitat use at the location scale by the probability of habitat use at the home range scale**

METHODS

Telemetry error estimation

- 6 Collar beacon study – 100 locations per observer from 4 distance intervals
- Observer bias – general linear model (SAS 2001)

Incorporation of error in location estimates
(Kenow et al. 2001)

- Standard error of bearing error (12.2 deg)
- Simulated a random normal distribution of bearings about the estimated bear location to produce points
  (SAS code from White and Garrott 1990)

Criteria for location error

- Bearings within 15 minutes or less, within 2 km
- *Evaluation of systematic errors – future analysis*

Candidate model subset

Took top-ranked single level variable models for combined variable models

Location level (home range use = random)

1.	**cut, hardwood, cypress, county rd. density**
2.	**cut, cypress, county rd. density**
3.	**cut, hardwood, county rd. density**

Home range level (locations = random)

1.	**cut, agri, hardwood, highway density**
2.	**cut, hardwood, highway density**
3.	**cut, hardwood**

4. Null model – intercept only model (no selection)

- **Model presence only habitat data without autocorrelation**
  - **One point per season per bear for each simulation run of the logistic regression model**
  - **Compile simulation traces for final stats over 300 simulations**
  - **For each 1 location point per bear randomly selected within season, a home range "absence" point is selected at random and a study area (home range "absence") point is selected**
  - **Idea is that when few location points are selected for a long time period, autocorrelation is minimized and absence can be inferred**

- **Condition the probability of habitat use at the location scale on the presence of a home range**

Challenges to using this technique

- **300 simulations per model = ~ 88 HOURS!**
- **Using 24 computers in forestry to run these suckers**
- **I run 50 simulations for each model on 6 computers and compile the traces later with another python program (C. Fonnesbeck)**
- **Databases of points are large and hard to sample**
- **We did not feed in point data for home range and study area**
- **We sample from the multinomial distribution of habitat types and the distribution of continuous variables by habitat type (Python code is about 350 lines long)**

See www.python.org for python programming

Results Example

RANK	model	name	mean	lower 95%	upper 95%	2.50%	50%	97.50%	variance	mc error
1	HR_MG	AIC	282.835	239.172493	320.63607	240.4501	282.5732	322.1554	451.5115	0.01735
1	HR_MG	gamma0	0.405546	-0.15926163	0.97250874	-0.15519	0.404057	0.976968	0.082661	0.000235
1	HR_MG	GOF	DIC						279.1535	
1	HR_MG	hr_ag	-3.465606	-3.84748307	1.14747027	-49.4191	-1.01999	0.41564	159.1306	0.0103
1	HR_MG	hr_cut	-0.018716	-1.30732694	1.6188194	-1.35475	0.168662	1.577871	12.97713	0.002941
1	HR_MG	hr_hwdmx	0.041665	-1.2595268	1.3238878	-1.29217	0.054096	1.29381	0.433581	0.000538
1	HR_MG	hr_hwyhr	-2.010274	-4.55785658	0.34199831	-4.73035	-1.92423	0.211597	1.517824	0.001006
2	HR_MA	AIC	287.5613	244.5818405	321.686638	246.2655	288.7303	323.9064	401.5583	0.016362
2	HR_MA	Gamma0	0.251505	-0.25933233	0.77055424	-0.26164	0.250578	0.768289	0.068719	0.000214
2	HR_MA	GOF	DIC						279.4991	
2	HR_MA	hr_cut	0.247123	-1.22599343	1.65008848	-1.28294	0.278967	1.599619	0.530027	0.000594
2	HR_MA	hr_hwdmx	0.213387	-1.11339931	1.44949527	-1.16349	0.247827	1.407671	0.424097	0.000532
2	HR_MA	hr_hwyhr	-2.172298	-4.65556733	0.10206862	-4.81473	-2.0818	-0.02203	1.467815	0.000989
3	LMG_HRMG	AIC	288.6929	240.7222824	330.426605	242.4488	289.1486	332.4108	527.1909	0.018747
3	LMG_HRMG	gamma0	0.527354	-0.1580494	1.2444948	-0.12744	0.51108	1.282053	0.125667	0.000289
3	LMG_HRMG	GOF	DIC	282.9890124						
3	LMG_HRMG	hr_ag	-2.561707	-3.29928324	0.65134399	-9.92326	-1.14834	0.350908	91.75428	0.007821
3	LMG_HRMG	hr_cut	-0.071992	-1.92792973	2.58471536	-1.95601	0.212952	2.559025	26.1327	0.004174
3	LMG_HRMG	hr_hwdmx	0.067542	-1.60239802	1.65481144	-1.67499	0.098385	1.598353	0.682822	0.000675
3	LMG_HRMG	hr_hwyhr	-2.114815	-4.93430898	0.41960114	-5.07745	-2.01016	0.300807	1.844359	0.001109
3	LMG_HRMG	l_cntyhr	-0.804932	-2.97889273	1.32302339	-3.02964	-0.77775	1.275766	1.184689	0.000889
3	LMG_HRMG	l_cut	-0.707059	-4.12187501	3.44328433	-4.45638	-0.16904	3.193877	43.261	0.00537
3	LMG_HRMG	l_cypr	-0.371767	-2.74359671	2.35193437	-2.76241	-0.23933	2.334978	13.82882	0.003036
3	LMG_HRMG	l_hwdmx	0.105042	-2.37056572	2.56477778	-2.2908	0.08451	2.656455	1.556889	0.001019

Predictive habitat maps

$$p(use) = \frac{1}{(1 + \exp(\log it\_p))}$$

$$p = inv \log it(\beta_0 + \beta_i X_i)$$

$$\beta_0 = (\gamma_{00} + \gamma_{0i})$$

p(use) calculated for each pixel in ArcView or GIS using map calculator

Acknowledgements

- Georgia Cooperative Fish & Wildlife Research Unit & University of Georgia
- Georgia Department of Natural Resources
- Field Assistance from Christina Alessi, Stephanie Kirk and Andrew Ross
- Chris Fonnesbeck and Jamie Skvarla
- Weyerhaeuser & Plum Creek biologists
- Committee members, Dr. John Carroll, Dr. Jim Peterson and Dr. Mike Wimberley
- NARSAL- Institute of Ecology, UGA