Hierarchical Bayesian Markov Switching Models with Application to Predicting Spawning Success of Shovelnose Sturgeon

Scott Holan
University of Missouri - Columbia

Joint work with
Ginger Davis - University of Virginia

and

M. Wildhaber, **A. DeLonay**, **D. Papoulias**, **J. Bryan** - USGS

Washington University - December 4, 2008

Outline

- Background and Motivation
- Data
- Model and Methods
- Results
- Summary and Future Directions

USGS - Columbia Environmental Research Center Study

- Recruitment of pallid sturgeon to the adult population is limited in the Missouri River
- Species is rare in the Missouri River and was listed as an endangered species in 1990
- Possible reasons for decline in pallid sturgeon population

  1. Commercial Harvest
  2. Habitat Alteration
  3. Pollution
  4. Impoundment (Dam Construction)

- Similar to the pallid sturgeon, the shovelnose sturgeon is declining and is at risk of extirpation

USGS - Columbia Environmental Research Center

- Determine the ecological requirements for reproduction and survival of pallid and shovelnose sturgeon in the Missouri River
- Shovelnose sturgeon closely related to the pallid sturgeon, and spawning requirements and behavior are similar in many respects
- Use shovelnose as a surrogate species to develop new research tools, or to examine the impacts of management actions, or environmental variables on sturgeon biology and habitat use
- Understanding the difference in successful and unsuccessful spawners within the shovelnose sturgeon should provide us with some knowledge concerning the spawning success of the closely related pallid sturgeon

Shovelnose and Pallid Sturgeon - Lower Missouri River

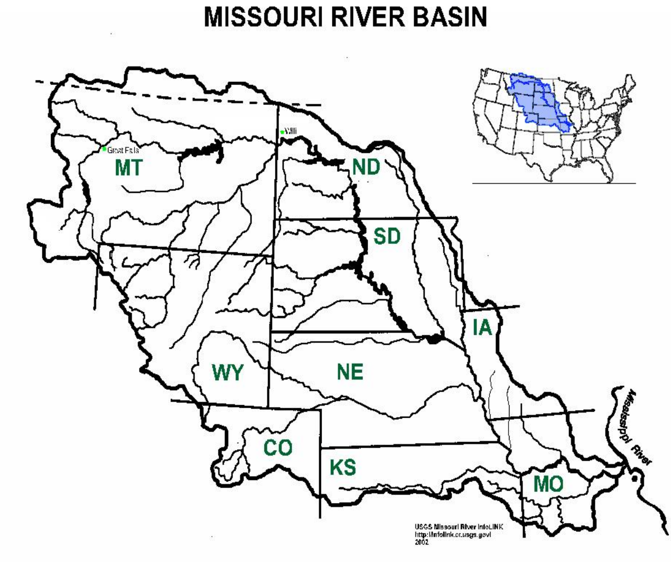
USGS Columbia Environmental Research Center - Study Objectives

1. Determine the direction, magnitude, and habitat used during spawning migrations for shovelnose sturgeon at two geologically and hydrologically distinct reaches of the lower Missouri River

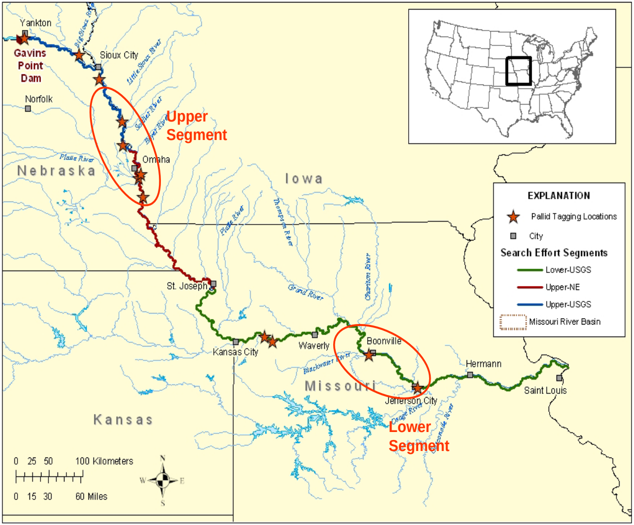
2. Describe the reproductive physiology of shovelnose sturgeon prior to and after successful and unsuccessful spawning

3. Identify and rank proximate cues necessary for successful spawning by Missouri River sturgeon

Study Area - Missouri River Basin



Study Area - Two geologically and hydrologically distinct segments of the Lower Missouri River



Data Collection Process

- Study Subjects
  - 2004: 9 female shovelnose sturgeon
  - 2005: 15 female shovelnose sturgeon
  - 2006: 20 female shovelnose sturgeon
- Biologists track sturgeon using two types of implanted telemetry devices
- Ultrasonic transmitters provide the location of fish, which are tracked through the suspected spawning period
- Archival data storage tags (DST's) record the temperature and depths of the fish every 15 minutes
- Goal: Use collected data to compare behavioral and environmental factors for spawning and non-spawning sturgeon

Ultrasound

USGS Fisheries Biologist is checking for female and readiness to spawn (note spawning only occurs every 2-3 years).

Reproductive Stage V

Telemetry and DST Devices

Telemetry Device Implantation

Fish Recapture

Biological Variables of Interest

- Sl = standard length of fish in mm
- Fl = fork length of fish in mm
- Wt = weight of fish in kg
- PI = polarization index
  - percent distance the germinal vesicle is to the edge of the egg
  - The lower the number the farther the nucleus has migrated and the closer the fish is to spawning.
- Cape2 = capture estradiol level in pg/mL
- Cap11kt = capture 11-ketotestosterone level in pg/mL
- Capc = capture cortisol level in ng/mL

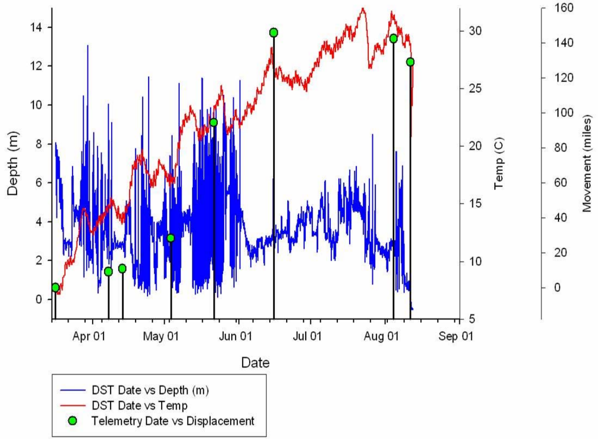
Blood Sample

Removal of blood from a shovelnose sturgeon

Environmental and Behavioral Variables of Interest

- Transcode = unique fish number (transmitter code number)
- Capture segment = denotes whether the fish was caught in the south or north section of the river
- Year = year fish was caught
- Depth = depth of fish
- Temperature = temperature of fish
- Location = river location

Example Time Series Plot of Depth, Temperature, and Location



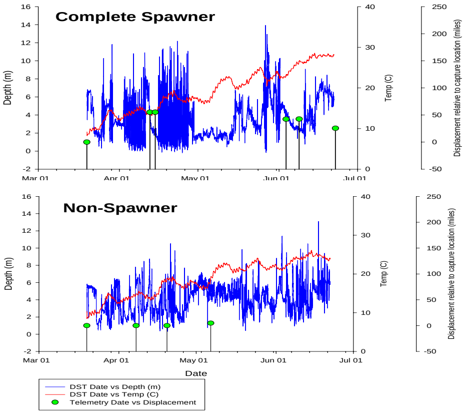
Response Variables

- Recapoocyteratio = ratio of mature oocytes (eggs) to early stage oocytes
  - Lower ratios are indicative of more complete the spawning.
  - Also known as spawning index
  - Logit transformation: $\text{logitratio} = \log \frac{\text{recapoocyteratio}}{1-\text{recapoocyteratio}}$
- Recapspawn = Categorical variable of the continuous recapoocyteratio variable
  - $0 - 35\%$ = complete spawn
  - $35 - 75\%$ = incomplete spawn
  - $> 75\%$ = no spawn
- Note that this choice of threshold was determined through extensive empirical investigation by expert fisheries biologists

Partial and Complete Spawn

Model Motivation

After examining depth profiles for successful and non-successful spawners, it is hypothesized that the variability of their depth profiles could be useful in predicting spawning success



Model Motivation - Exploratory Analysis

- Fit Markov Switching Stochastic Volatility Models - Smith 2002
  - 2-regime (high and low) model for volatility
  - Switching between regimes is governed by probability transition matrix

$$\begin{pmatrix} 1 - e_{i1} & e_{i1} \\ e_{i2} & 1 - e_{i2} \end{pmatrix}$$

  - It was observed that the values e_{i1} and e_{i2} were higher for fish who had no spawn or only a partial spawn
  - univariate measurement = 2nd eigenvalue $= 1 - e_{i1} - e_{i2}$
  - What does this mean?
    - Fish who do not spawn or spawn only partially transition between the high and low variability states more frequently than the fish who spawn completely

Model Motivation - Frequentist Formulation

- Let λ_i an z_i denote the eigenvalue and logit ratio for the i^{th} fish respectively
- Linear Regression Model

$$
\begin{aligned}
z_i &= \gamma_0 + \gamma_1 wt_i + \gamma_2 sl_i + \gamma_3 PI_i + \gamma_4 cape2_i + \gamma_5 cap11kt_i \\
&+ \gamma_6 capc_i + \gamma_7 \lambda_i + \epsilon_i
\end{aligned}
$$

- What is the problem here?
  - λ_i is not given, but estimated
  - differing levels of uncertainty / sample sizes (profile lengths)
- One solution: Take a Bayesian approach

Hierarchical Bayesian Model Definition

- Heuristically the model can be thought of as follows

1. Higher level of hierarchy
    - Generate draws from distribution of the second eigenvalue of the transition probability matrix in a Markov switching volatility model
2. Lower level of hierarchy
    - Given these eigenvalues, we estimate the regression parameters of interest in a linear model

Two-state Markov Switching Volatility Model Definition

- d_{it} = depth of fish i at time t and $\mathbf{D}_i$ = collection of all depth values for fish i
- Two-state Markov switching model with different GARCH - dynamics:

$$d_{it} = \begin{cases} \beta_{i1}\sqrt{h_{it}} + \sqrt{h_{it}}\epsilon_{it}, & h_{it} = \alpha_{i10} + \alpha_{i11}h_{i,t-1} + \alpha_{i12}a_{i,t-1}^2, \quad \text{if } s_{it} = 1; \\ \beta_{i2}\sqrt{h_{it}} + \sqrt{h_{it}}\epsilon_{it}, & h_{it} = \alpha_{i20} + \alpha_{i21}h_{i,t-1} + \alpha_{i22}a_{i,t-1}^2, \quad \text{if } s_{it} = 2, \end{cases}$$

where $a_{it} = \sqrt{h_{it}}\epsilon_{it}$, $\{\epsilon_{it}\}$ is a sequence of standard normal white noise random variables and the parameters α_{ijk} satisfy some regularity conditions so that the unconditional variance of a_{it} exists

Two-state Markov Switching Volatility Model Definition

- The probability that a fish transitions from one state to another, is governed by the following transition probabilities

$$
\begin{aligned}
P\left(s_{it} = 2 | s_{i,t-1} = 1\right) &= e_{i1}, \\
P\left(s_{it} = 1 | s_{i,t-1} = 2\right) &= e_{i2}
\end{aligned}
$$

  where $0 < e_{ij} < 1$ for $j = 1, 2$

- Small values of e_{ij} indicate that fish i has a tendency to stay in the j^{th} state with expected duration $\frac{1}{e_{ij}}$

Estimating the Model

- Bayesian method using Gibbs sampling approach
- Assume h_{i1} and equal to the sample variance of d_{it}. The effect of this assumption is negligible when the sample size is large
- Parameters to estimate:
  - β_{i1}, β_{i2}
  - $\alpha_{i10}, \alpha_{i11}, \alpha_{i12}, \alpha_{i20}, \alpha_{i21}, \alpha_{i22}$
  - e_{i1}, e_{i2}
  - state vector $\mathbf{S}_i = (s_{i1}, s_{i2}, \ldots, s_{in_i})$
  - volatility vector $\mathbf{H}_i = (h_{i2}, \ldots, h_{in_i})$

Estimating the Model - Prior Distributions

- Gibbs sampling approach - only the following conditional posterior distributions are needed:
  - $f(\boldsymbol{\beta}_i | \mathbf{D}_i, \mathbf{S}_i, \mathbf{H}_i, \boldsymbol{\alpha}_{i1}, \boldsymbol{\alpha}_{i2})$
  - $f(\boldsymbol{\alpha}_{i1} | \mathbf{D}_i, \mathbf{S}_i, \mathbf{H}_i, \boldsymbol{\alpha}_{i2})$
  - $f(\boldsymbol{\alpha}_{i2} | \mathbf{D}_i, \mathbf{S}_i, \mathbf{H}_i, \boldsymbol{\alpha}_{i1})$
  - $P(\mathbf{S}_i | \mathbf{D}_i, h_{i1}, \boldsymbol{\alpha}_{i1}, \boldsymbol{\alpha}_{i2})$
  - $f(e_{i1}, e_{i2} | \mathbf{S}_i)$
- For simplicity, we impose conjugate priors for β_{ij} and e_{ij} $(j = 1, 2)$
  - $\beta_{ij} \sim N(\beta_{j0}, \sigma_{j0}^2)$, for $j = 1, 2$
  - $e_{ij} \sim \mathrm{Beta}(\delta_{j1}, \delta_{j2})$, for $j = 1, 2$
- The prior distribution of α_{ijk} is uniform over a properly specified interval

Posterior Distribution of β_{i1}, β_{i2}

- The posterior distribution of β_{ij} ($j = 1, 2$) only depends on the data in state j
- For ($j = 1, 2$), let

$$
\begin{aligned}
d_{it}^{(j)} &= \frac{d_{it}}{\sqrt{h_{it}}} \text{ if } s_{it} = j \text{ and } 0 \text{ otherwise,} \\
\bar{d}_{it}^{(j)} &= \frac{\sum_{s_{it}=1} d_{it}{}^{(j)}}{n_{ij}} \\
n_{ij} &= \text{number of data points in state } j \text{ for fish } i
\end{aligned}
$$

- Then the conditional posterior distribution of β_{ij} is

$$\beta_{ij} \sim N\left(\sigma_{ij*}^2 \left(n_{ij}\bar{d}_{it}^{(j)} + \beta_{j0}/\sigma_{j0}^2\right), \sigma_{ij*}^2\right),$$

where $\frac{1}{\sigma_{ij*}^2} = n_{ij} + \frac{1}{\sigma_{j0}^2}$

Posterior Distribution of α_{ijk}

- In order to draw realizations of α_{ijk} we use the Griddy Gibbs Method - Ritter and Tanner (1992)
- Given $h_{i1}, \mathbf{S}_i, \text{all other elements in } \alpha,$ we have that

$$f(\alpha_{ijk}|\cdot) \propto -\frac{1}{2}\left(\log h_{it} + \frac{\left(d_{it} - \beta_{ij}\sqrt{h_{it}}\right)^2}{h_{it}}\right), \quad \text{if} \quad s_{it} = j$$

- Evaluate this function at a grid of points for α_{ijk} over a properly specified interval.
- Define the following
  - m_{i1} = the number of switches from state 1 to state 2
  - m_{i2} = the number of switches from state 2 to state 1
- posterior distribution of $e_{ij} \sim \text{Beta}\left(\delta_{j1} + m_{ij}, \delta_{j2} + n_{ij} - m_{i1}\right)$

Posterior Distribution of $\mathbf{S}_i$

- Elements of $\mathbf{S}_i$ drawn one by one
- Let $\mathbf{S}_i^{(-l)}$ be the vector obtained by removing s_{il} from $\mathbf{S}_i$
- Given $\mathbf{S}_i^{(-l)}$ and other information, the conditional posterior distribution of s_{il} is

$$P(s_{il}|\cdot) \quad \propto \quad \prod_{t=l}^{n_i} (a_{it}|\mathbf{H}_i)\, P\left(s_{il}|\mathbf{S}_i^{(-l)}\right).$$

- $L(s_{il} = j) \equiv \prod_{t=l}^{n_i} f(a_{it}|\mathbf{H}_i) \propto \exp(f_{ilj})$, where

$$f_{ilj} = \sum_{t=l}^{n_i} -\frac{1}{2}\left(\ln h_{it} + \frac{a_{it}^2}{h_{it}}\right).$$

  and $a_{it} = d_{it} - \beta_{ij}\sqrt{h_{it}}$ if $s_{it} = j$ for $j = 1, 2$
- Finally, the conditional posterior probability of $s_{il} = j$ is

$$P(s_{il} = j|\cdot) = \frac{P(s_{il} = j|s_{i,l-1}, s_{i,l+1})\, L(s_{il} = j)}{P(s_{il} = 1|s_{i,l-1}, s_{i,l+1})\, L(s_{il} = 1) + P(s_{il} = 2|s_{i,l-1}, s_{i,l+1})\, L(s_{il} = 2)}$$

- Therefore state s_{il} can be drawn from a Uniform(0,1) distribution

Estimating the Regression Model - Priors

- Let $\mathbf{z} = \text{logitratio} = \mathbf{X}\gamma + \boldsymbol{\epsilon}$ denote our regression in matrix notation, with $\epsilon \sim N(0, \sigma_\epsilon^2 I)$
- Additionally, let $\boldsymbol{\gamma} = \boldsymbol{\mu} + \mathbf{v}$ with $\mathbf{v} \sim N(0, V_{\boldsymbol{\gamma}})$
- Following Rossi, Allenby, and McCulloch (2005) we impose the following conjugate priors:
  - $\sigma_\epsilon^2 \sim \nu_\epsilon s_0^2 / \chi^2_{\nu_\epsilon}$
  - $V_{\boldsymbol{\gamma}} \sim IW(\boldsymbol{\nu}, V)$
  - $\boldsymbol{\mu} | V_{\boldsymbol{\gamma}} \sim N(\overline{\boldsymbol{\mu}}, V_{\boldsymbol{\gamma}} \otimes A^{-1})$ where $\overline{\boldsymbol{\mu}} = \mathbf{0}$ is a matrix of prior means and $A = .01\mathbf{I}$ is a matrix for prior precision
  - ν_ϵ is a degree of freedom (df) parameter for σ_ϵ^2 defined equal to 3
  - ν is the df parameter for $V_{\boldsymbol{\gamma}}$ defined equal to the number of variables plus 3

Estimating the Regression Model - Posterior Distribution

- We use a Gibbs sampling technique to first draw $(\gamma, \sigma_\epsilon^2)$ given the parameters of the first stage prior, $\boldsymbol{\mu}$, $V_{\boldsymbol{\gamma}}$, and then draw the prior parameters conditional on $(\gamma, \sigma_\epsilon^2)$

- The posterior distribution for the regression parameters of interest is

$$\gamma | \mathbf{z},\, \mathbf{X},\, \boldsymbol{\mu},\, \mathbf{V}_\gamma,\, \sigma_\epsilon^2 \sim \mathcal{N}\left\{\gamma^*, (\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{V}_\gamma^{-1})^{-1}\right\}$$

where

$$\gamma^* = (\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{V}_\gamma^{-1})^{-1}(\mathbf{X}^{*\prime}\mathbf{z}^* + \mathbf{V}_\gamma^{-1}\boldsymbol{\mu}),$$

with $\mathbf{z}^* = \mathbf{z}/\sigma_\epsilon$ and $\mathbf{X}^* = \mathbf{X}/\sigma_\epsilon$

Parallel Computing

- Large number of Markov switching stochastic volatility models required in estimation (1 per fish)
- These models are computationally expensive due to the grid estimation technique (high dimensional grid)
- Good news: Each model is independent!
- Parallel computation using *Rmpi* in R

Parallel Computation Algorithm

- Steps in algorithm
  1. Master processor generates information for joint model (which is conditional on the previous iteration).
  2. Master processor reports this information to each slave.
  3. Slaves perform estimation of fish-specific Markov switching stochastic volatility models.
  4. Master collect estimates of the fish specific eigenvalues and uses them in estimation of lower level parameters.
- Computational cost benefit
  - 1 iteration in serial: 57.7 minutes
  - 1 iteration in parallel: 1.7 minutes
  - 97% reduction in computing time!

Alternative models

- OLS Linear Regression Model without eigenvalue predictor

$$\begin{aligned}\text{logitratio}_i &= \gamma_0 + \gamma_1 wt_i + \gamma_2 sl_i + \gamma_3 PI_i + \gamma_4 cape2_i \\ &\quad + \gamma_5 cap11kt_i + \gamma_6 capc_i + \epsilon_i\end{aligned}$$

- Hierarchical Bayesian Regression Model 1

$$\begin{aligned}\text{logitratio}_i &= \gamma_0 + \gamma_1 wt_i + \gamma_2 sl_i + \gamma_3 PI_i + \gamma_4 cape2_i \\ &\quad + \gamma_5 cap11kt_i + \gamma_6 capc_i + \gamma_7 \lambda_i + \epsilon_i\end{aligned}$$

- Hierarchical Bayesian Regression Model 2

$$\begin{aligned}\text{logitratio}_i &= \gamma_0 + \gamma_1 wt_i + \gamma_2 PI_i + \gamma_3 cape2_i \\ &\quad + \gamma_4 cap11kt_i + \gamma_5 capc_i + \gamma_6 \lambda_i + \epsilon_i\end{aligned}$$

- Hierarchical Bayesian Regression Model 3

$$\text{logitratio}_i = \gamma_0 + \gamma_1 wt_i + \gamma_2 \lambda_i + \epsilon_i$$

OLS Linear Regression Model without eigenvalue parameter estimates predictor

Parameter	Estimate	Std. Error	$\Pr(>\|t\|)$
Intercept	-8.2099	21.4056	0.706
wt	0.0083	0.0062	0.198
sl	-0.0081	0.0471	0.866
PI	8.1889	15.6906	0.608
*cape*2	-0.0004	0.0003	0.224
*cap*11*kt*	0.0003	0.0005	0.572
capc	-0.0557	0.0498	0.278

Preferred Hierarchical Bayesian Regression Model

Model 3 was preferred based on:

1. exploratory analysis of 95% credible intervals of model parameters in a fully saturated model,
2. underlying biological considerations supplied by expert fisheries biologists,
3. DIC (Note: DIC(Model 1)=533.35, DIC(Model 2)=532.46),
4. Model 3 has lowest mean squared error as well as best in-sample classification

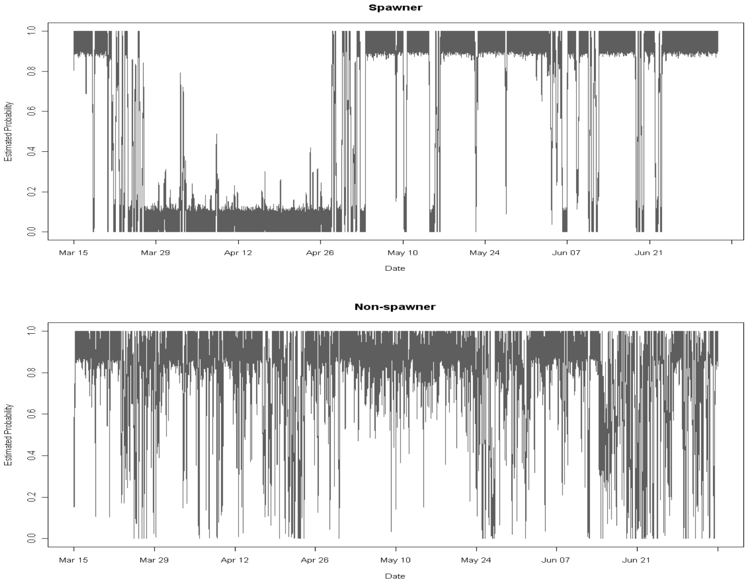
Model M3 - DIC = 522.4657			
Parameter	Posterior Mean	Posterior Std. Dev.	95% Credible Interval
Intercept	0.180	0.989	(-1.44,1.80)
wt*	0.00268	0.00130	(0.000582,0.00485)
eigenvalue*	-0.851	0.186	(-1.16,-0.546)

Note: In all Bayesian models eigenvalue was significant and in Models 2 and 3 wt was significant as well

Comparison of Confusion Matrix for OLS regression model and the hierarchical Bayesian Model 3

Actual (Model)	Predict (Model) Successful Spawn	Predict (Model) Unsuccessful Spawn
Successful Spawner (OLS)	37	0
Unsuccessful Spawner (OLS)	4	3
Successful Spawner (Bayesian M3)	37	0
Unsuccessful Spawner (Bayesian M3)	1	6

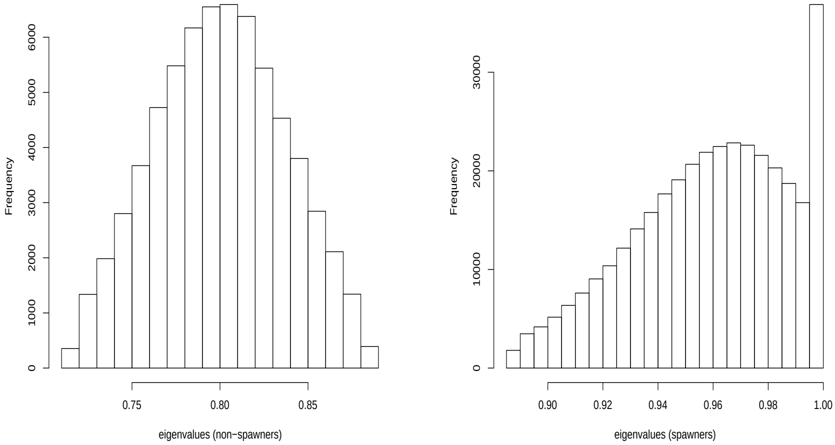
Probability estimates of being in the low variability regime



Markov Switching GARCH Model Parameters

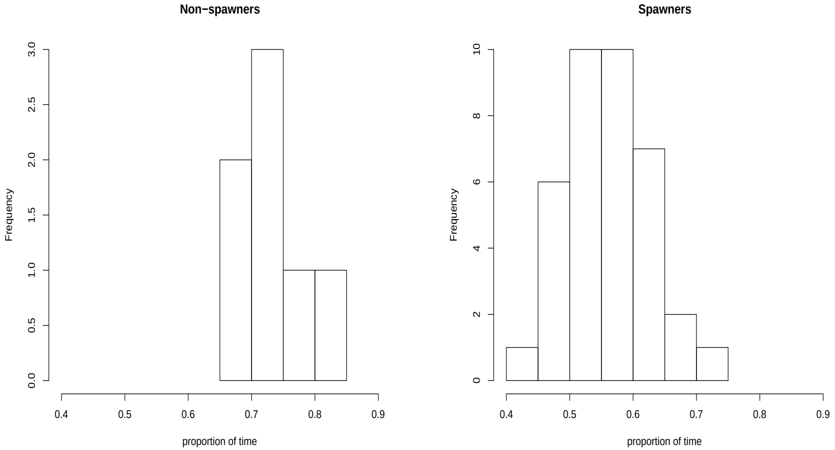
Parameter	Non-spawner Posterior Mean (95% CI)	Spawner Posterior Mean (95% CI)
Low Variability Regime		
α_{10}	1.33 (0.714, 2.09)	1.30 (0.0781, 1.69)
α_{11}	0.292 (0.174, 0.396)	0.485 (0.340, 0.631)
α_{12}	0.432 (0.242, 0.644)	0.140 (0.00795, 0.366)
High Variability Regime		
α_{20}	0.025 (0.0144, 0.0395)	0.122 (0.0979, 0.157)
α_{21}	0.569 (0.425, 0.722)	0.857 (0.679, 0.998)
α_{22}	0.431 (0.355, 0.521)	0.105 (0.0343, 0.191)

Eigenvalue HPD Histograms for non-spawners vs. spawners



Note that these posterior densities are disjoint and have endpoints as follows: $(0.717, 0.884)$ for the non-spawners and $(0.887, 0.998)$ for the spawners

Histogram - Proportion of Time Each Fish Spent in the High Variability Regime



Note that we work with proportion of time rather than length of time because of unequal observation lengths due to different recapture times

Summary

- Developed a Bayesian hierarchical model for predicting spawning success capable of utilizing Data Storage Tag data
- Model incorporates an eigenvalue predictor from the transition probability matrix in a two-state Markov switching model with GARCH dynamics as a generated regressor in a linear regression model
- Outperforms model without DST information
- OLS model is insufficient: does not find relevance of weight to spawning success
- Our results support the hypothesis that spawners exhibit lower levels of depth variability in their swimming pattern during the spawning season
- Clear distinction between depth variability (95% CIs of eigenvalue estimates) in spawners and non-spawners
- Computationally expensive, but cost minimized using parallel computing

Future Work

- Incorporate temperature profile into model
  - The fact that there is a "preferred" temperature might help predict spawning occurrence
- Evaluate importance of upstream or downstream movement for spawning prediction
- Further data collection for developing model for not only the occurrence, but the timing of spawning - currently in progress

The work presented here is to appear - (2009) *Journal of the Royal Statistical Society - Series C*. 58: 47–64