2025

Bayesian Statistics (STA306) Final Project

A BAYESIAN STATE-SPACE FORMULATION OF DYNAMIC OCCUPANCY MODELS

Example reproduction: European Crossbill

Group Member

Zihan Zhang
Qianyu Yang
Xinyu Zhang
Siqing Li

Royle, J. A., & Kéry, M. (2007). A Bayesian state- space formulation of dynamic occupancy models. *Ecology*, *88*(7), 1813-1823.

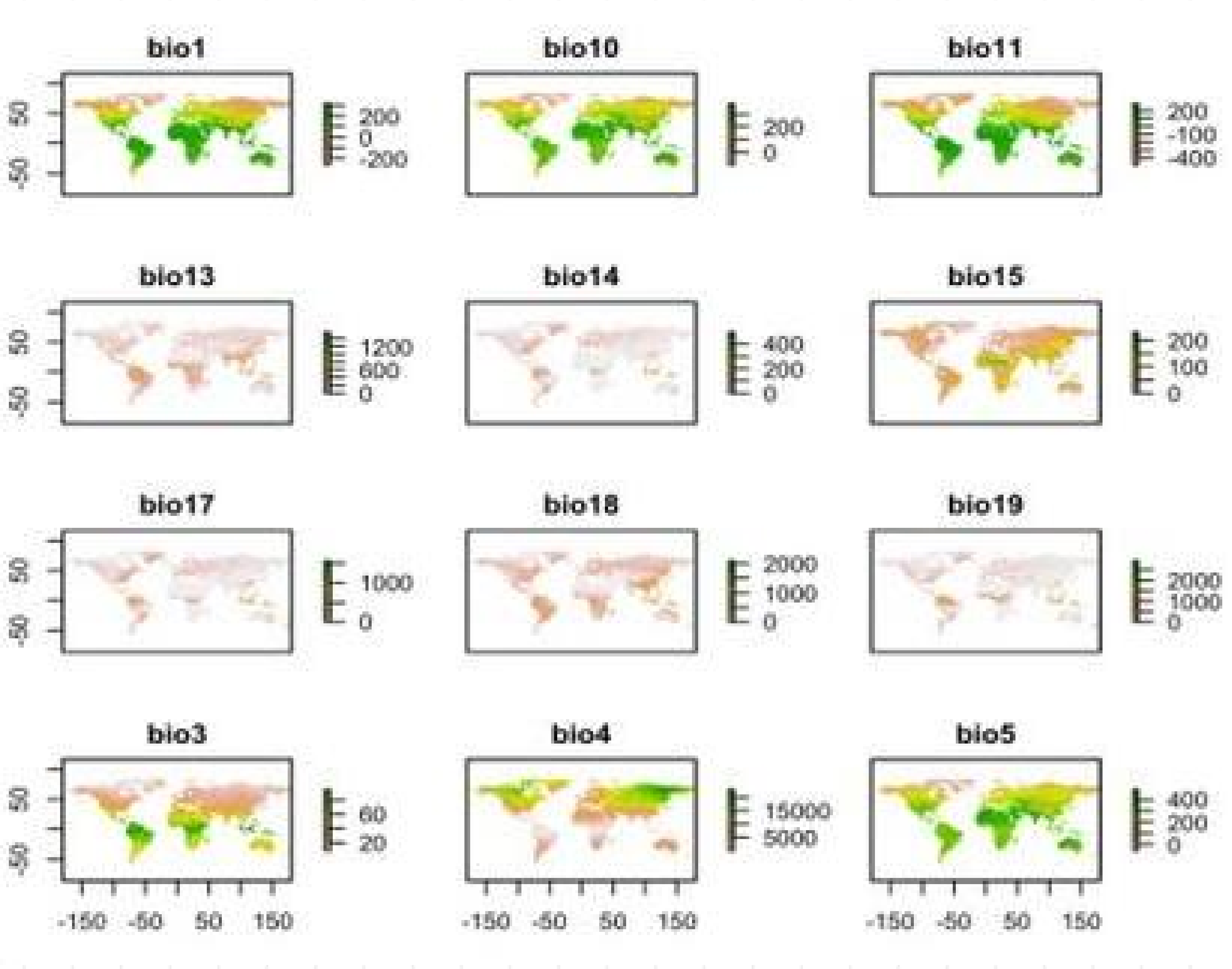
CONTENTS

PART ONE

Background and Introduction

Bayesian Statistics

Background



Research Objective: Develop a hierarchical dynamic occupancy model, using a state-space framework to estimate site occupancy, extinction, colonization, and turnover rates.

Species distribution and detection errors

Species distribution and fragment occupancy are important concepts in biogeography and conservation biology, but species surveys often have **detection errors**, leading to distribution estimation biases. Therefore, it is necessary to clearly **consider the detection probability** to obtain accurate occupancy estimates.

Advantages of Site Occupancy Models

Site occupancy models are widely used due to their simplicity in using detection/non-detection data, explicit handling of detectability, and extensibility. They can also estimate or model population abundance.

Advantages of Site Occupancy Models

To study **dynamic features** like local extinction and colonization, occupancy models are extended to dynamic models, allowing **site occupancy to change over time**. These models better capture population processes and enhance ecological insights.

Introduction of Dataset

The **Crossbill Dataset** is an ecological dataset. It consists of detailed data collected from multiple forest regions through surveys and is commonly used to analyze crossbill behavior.
267 1-kmsq quadrats were surveyed 3 times per year during 1999-2007.

Basic Information

- id: A unique identifier for each recorded site.
- surveys: Number of surveys conducted, representing repeated observations at the sampling point.
- det991, det992, ..., det073 : Detection indicators, showing whether crossbill birds were detected at different time points or under different conditions (1 for detected, 0 for not detected, NA for missing data).

Important information extraction

- Survey year: Four years
- Annual survey frequency: three times
- Number of sites: 267

PART TWO

Sampling Design and the Dynamic Model

Bayesian Statistics

Sampling Design

Objective: The core of studying dynamic occupancy models is to analyze the **dynamic occupancy** of species on spatial units such as sites or regions.

- Time: **t**=1,2,…,T (Primary periods)
- Sites: **i**=1,2,…,R (Spatial unit station)
- Surveys: **j**=1,2,…,J (Repeated survey frequency)
- Observed occupancy status of site i for survey j within primary period t: $\mathbf{y_j(i, t)}$
- True occupancy status of site i during primary period t: **z(i, t)**
- Probability of site occupancy for period t: $\boldsymbol{\psi_t} = Pr(z(i, t) = 1)$
- Probability that an occupied site remains occupied: $\boldsymbol{\phi_t} = Pr(z(i, t+1) = 1 | z(i, t) = 1)$ --> Survival
- Probability of the local colonization: $\boldsymbol{\gamma_t} = Pr(z(i, t+1) = 1 | z(i, t) = 0)$ --> Colonization

$$\mathbf{t} \quad \mathbf{i} \quad \mathbf{j} \quad \mathbf{y_j(i, t)} \quad \boldsymbol{\psi_t} \quad \boldsymbol{\phi_t} \quad \boldsymbol{\gamma_t}$$

State-space representation

Dynamic occupancy model that divides the dynamic process of species into two parts

- **State Model**: Describes the actual occupancy status of a site.
- **Observation Model**: Describing observation results based on occupancy status.

State Model

Initial occupancy states： Assuming in the first time period (t = 1), The occupancy status of each site z(i, t) is independent and follows a Bernoulli distribution.

$$z(i, 1) \sim \text{Bernoulli}(\psi_1)$$

During consecutive periods (t>1), the occupancy state of a site z(i,t) is determined by the state in the previous period z(i,t−1) and is influenced by the local survival probability (ϕ_{t-1}) and local colonization probability (γ_{t-1}):

$$z(i,t)|z(i,t-1) \sim \text{Bernoulli}\{z(i,t-1)\phi_{t-1} + [1 - z(i,t-1)]\gamma_{t-1}\}$$

State-space representation

Dynamic occupancy model that divides the dynamic process of species into two parts

- **State Model**: Describes the actual occupancy status of a site.
- **Observation Model**: Describing observation results based on occupancy status.

Observation Model

The observation model describes how the observer collects data $y_j(i,t)$ based on the true occupancy state z(i,t):

$$y_j(i,t)|z(i,t) \sim \text{Bernoulli}[z(i,t)p_t].$$

- If site i is occupied at time t (z(i,t)=1), the observation outcome yj(i,t) is a Bernoulli trial with success probability p_t.
- If site i is not occupied at time t (z(i,t)=0), $y_j(i,t)=0$.
- Detection probability of the investigator successfully detecting the presence of a species if a site is occupied during the t time period) : p_t

Metapopulation summaries

Occupancy probability at t can be computed recursively:

$$\psi_t = \psi_{t-1}\phi_{t-1} + (1 - \psi_{t-1})\gamma_{t-1}$$

Growth rate: $\lambda_t = \dfrac{\psi_{t+1}}{\psi_t}$.

Turnover $\Pr(z(t-1) = 0 \mid z(t) = 1)$

Probability that an occupied quadrat picked at random is a newly occupied one:

$$\tau_t = \frac{\gamma_{t-1}(1 - \psi_{t-1})}{\gamma_{t-1}(1 - \psi_{t-1}) + \phi_{t-1}\psi_{t-1}}$$

Finite sample estimation

Finite-sample estimates are easily calculated under a Bayesian framework using Markov Chain Monte Carlo(MCMC), which directly samples the latent occupancy states (z(i,t)) and computes desired quantities.

Defines finite-sample occupancy as the proportion of occupied sites in the sample:

$$\psi_t^{(\mathrm{fs})} = \frac{1}{R}\sum_i z(i,t)$$

Estimator of sample growth rate:

$$\lambda_t^{(\mathrm{fs})} = \frac{\sum_{i=1}^{R} z(i,t+1)}{\sum_i z(i,t)}$$

Estimator sample turnover rate:

$$\tau_t^{(\mathrm{fs})} = \frac{\sum_{i=1}^{R}[1 - z(i,t-1)]z(i,t)}{\sum_{i=1}^{R} z(i,t)}.$$

PART THREE

Bayesian Analysis and MCMC implementation

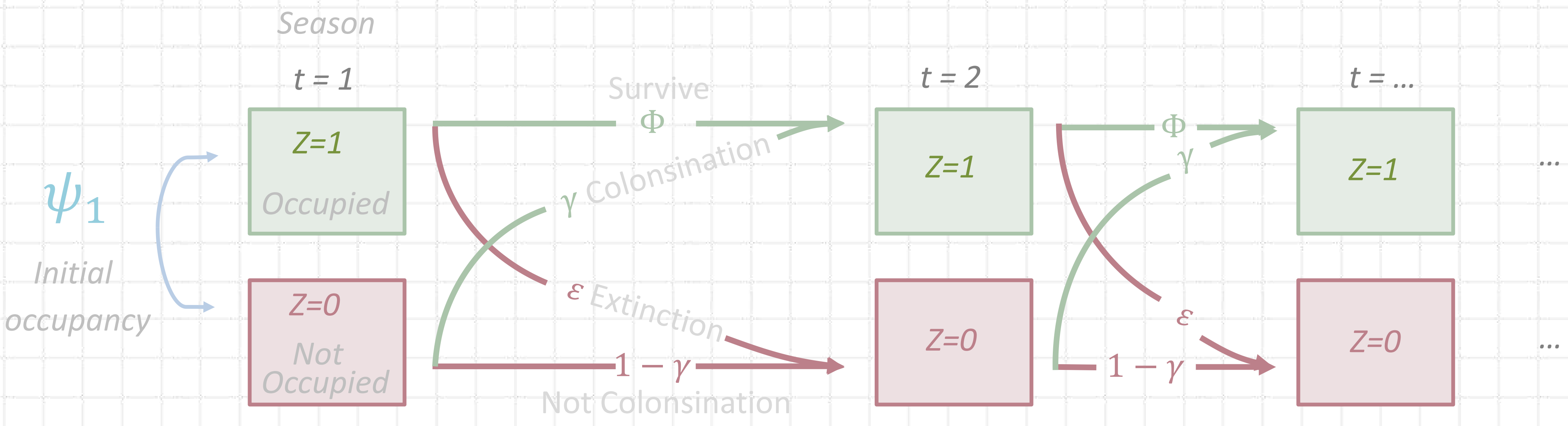
Bayesian Statistics

Gibbs Sampling

State Model

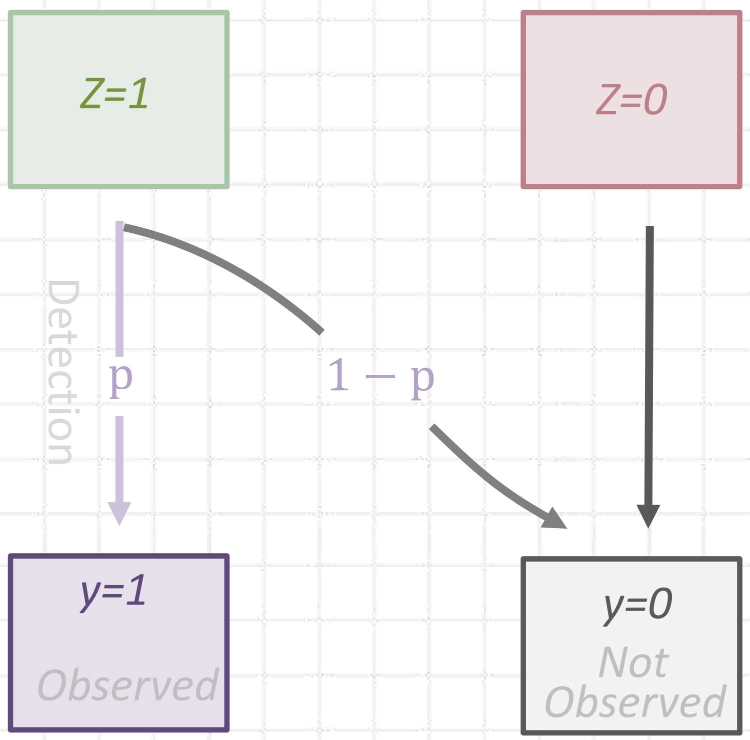
Site occupation state $z(i,t)$ can be change between each season t and site i. In the first season, $z(i,1)$ is determined by the probability of Initial Occupancy ψ_1. In the subsequent seasons, $z(i,t)$ is determined by the site i and the site's occupancy state in prior season and probabilities of Colonisation γ and Occupied Site 'Survives' ϕ.

For each site,



Observation Model

When a site is survived, the observed occupancy state y is determined by probability of Detection p.



Occupancy in season 1:

$$z(i,1) \sim Bernoulli(\psi_1)$$

Occupancy in season t:

$$z(i,t) \sim Bernoulli\left(z(i,t-1)\phi_{t-1} + (1 - z(i,t-1))\gamma_{t-1}\right)$$

Observation of survey j during time t:

$$y_j(i,t) \sim Bernoulli(z(i,t)p_t)$$

Gibbs Sampling

- **Prior Distribution**

  - For hyperparameter p_t ψ_1 ϕ_t γ_t, we assume they follow **Uniform** distribution.

$$\psi_1 , p_t , \phi_t , \gamma_t \sim U(0,1)$$

  - For latent parameters z(i,t), we assume that they are drawn from a **Bernoulli** distribution with probability of $z(i,t-1)\phi_{t-1} + \big(1 - z(i,t-1)\big)\gamma_{t-1}$.

$$p(z(1,1), z(2,1), \ldots z(R,1), \ldots z(1,T), z(2,T), \ldots z(R,T) | \psi_1, \phi_1, \gamma_1, \ldots, \phi_{T-1}, \gamma_{T-1}$$

$$\propto \prod_{i=1}^{R} \prod_{t=2}^{T} Bernoulli\big(z(i,t-1)\phi_{t-1} + \big(1 - z(i,t-1)\big)\gamma_{t-1}\big) \times Bernoulli(\psi_1)$$

Gibbs Sampling

- **Likelihood Distribution**

We consider data obtained from repeated presence/absence surveys of i = 1, 2, ... , R spatial units and each site is surveyed j = 1, 2, ... , J times within each of t = 1, 2, ... , T primary periods and that each site is closed with respect to its occupancy status within but not across primary periods.

For site i at time t,

$$p(y_1(i,t), y_2(i,t), \ldots, y_J(i,t) | z(i,t) = 1, p_t) \propto \prod_{j=1}^{J} \text{Bernoulli}(z(i,t)p_t)$$

$$y_1(i,t), y_2(i,t), \ldots, y_J(i,t) | z(i,t) = 1, p_t \sim \text{Binomial}\,(J, z(i,t)p_t)$$

Gibbs Sampling

- **Joint Posterior Distribution**

Combine the state model and observed model, the joint posterior distribution of all parameters and hyper parameter, that is,

$$p(\psi_1, p_1, \phi_1, \gamma_1, p_2, \ldots, \phi_{T-1}, \gamma_{T-1}, p_T, z(1,1), z(2,1), \ldots z(R,1), \ldots, z(1,T), z(2,T), \ldots z(R,T) | y)$$

$$\propto \prod_{i=1}^{R} \prod_{j=1}^{J} \prod_{t=1}^{T} p\left(y_j(i,t) \middle| z(i,t) = 1, p_t\right)$$

$$\times\, p(z(1,1), z(2,1), \ldots z(R,1), \ldots z(1,T), z(2,T), \ldots z(R,T) | \psi_1, \phi_1, \gamma_1, \ldots, \phi_{T-1}, \gamma_{T-1})$$

$$\times\, p(\psi_1, p_1, \phi_1, \gamma_1, p_2, \ldots, \phi_{T-1}, \gamma_{T-1} p_T)$$

Gibbs Sampling

- **Conditional Posterior Distribution**

Because the prior distributions of hyperparameters are uniform, which is a beta(1,1), the conjugate prior for the binomial likelihood, the conditional posterior distribution for each parameter is **beta** distribution.

➢ The conditional distribution of ψ_1 given other parameters is

$$\psi_1 \mid \cdot \;\sim\; \mathrm{Beta}\Big(\sum_i z(i,1) + 1,\; R - \sum_i z(i,1) + 1\Big)$$

➢ The conditional distribution of p_t given other parameters is

$$p_t \mid \cdot \;\sim\; \mathrm{Beta}\Big(\sum_i \sum_j y_j(i,t) \times z(i,t) + 1,\; R \times J - \sum_i \sum_j y_j(i,t) \times z(i,t) + 1\Big)$$

Gibbs Sampling

Z(i , t)	Z(i , t+1)	frequency
0	0	n_1
0	1	n_2
1	0	n_3
1	1	n_4

- **Conditional Posterior Distribution**
  - The conditional distribution of ϕ_t given other parameters is.

$$\phi_t \mid \cdot \sim \mathrm{Beta}(n_4 + 1, n_3 + 1)$$

where n_3 is the number of extinction events during the interval t to t+1, n_4 is the number of survives events during the interval t to t+1.

  - The conditional distribution of ϕ_t given other parameters is

$$\gamma_t \mid \cdot \sim \mathrm{Beta}(n_2 + 1, n_1 + 1)$$

where n_1 is frequency of z(i,t)=0 and z(i,t+1)=0, n_2 is the number of colonization events during the interval t to t+1.

Gibbs Sampling

- **Conditional Posterior Distribution**

  - The conditional distribution of latent parameter $z(i,t)$ given other parameters is a little complex, we will discuss by two steps.

    First, we ignore the influence of data.

    For t=2,3,4,…,T-1, we can simplify the model as

$$\mathrm{z(i,t)|z(i,t-1),z(i,t+1)} \sim \mathrm{Bernoullli}(\psi_{condl})$$

    For t=1, T , we can simplify the model as

$$\mathrm{z(i,1)|z(i,\ 2)} \sim \mathrm{Bernoullli}(\psi_{condl})$$

$$\mathrm{z(i,T)|z(i,\ T-1)} \sim \mathrm{Bernoullli}(\psi_{condl})$$

Gibbs Sampling

- **Conditional Posterior Distribution**

t = 2, 3, 4, … , T-1

$z(i, t-1)$	$z(i, t+1)$	$\psi_{condl} = p\big(z(i,t)\big) = 1$
1	1	$\dfrac{\phi_{t-1}\,\phi_t}{\phi_{t-1}\phi_t + (1-\phi_{t-1})\gamma_t}$
0	1	$\dfrac{\gamma_{t-1}\,\phi_t}{\gamma_{t-1}\phi_t + (1-\gamma_{t-1})\gamma_t}$
1	0	$\dfrac{\phi_{t-1}\,(1-\phi_t)}{\phi_{t-1}(1-\phi_t) + (1-\phi_{t-1})(1-\gamma_t)}$
0	0	$\dfrac{\gamma_{t-1}\,(1-\phi_t)}{\gamma_{t-1}(1-\phi_t) + (1-\phi_{t-1})(1-\gamma_t)}$

Gibbs Sampling

- **Conditional Posterior Distribution**

t = 1

$z(i,2)$	$\psi_{condl} = p\big(z(i,1)\big) = 1$
1	$\dfrac{\phi_1\,\psi_1}{\phi_1\psi_1 + \gamma_t(1-\psi_1)}$
0	$\dfrac{(1-\phi_1\,)\psi_1}{(1-\phi_1)\psi_1 + (1-\gamma_t)(1-\psi_1)}$

t = T

$z(i,T-1)$	$\psi_{condl} = p\big(z(i,T)\big) = 1$
1	ϕ_{T-1}
0	γ_{T-1}

Gibbs Sampling

- **Conditional Posterior Distribution**

The complete posterior distribution after combining the observational data

$$p(z(i,t)=1|y(i,t),z(i,t-1),z(i,t+1)) \propto p(y(i,t)|z(i,t)) \times p(z(i,t)|z(i,t-1),z(i,t+1))$$

$$= \frac{p^{\sum_j y_j(i,t)}(1-p)^{J-\sum_j y_j(i,t)}\psi_{condl}}{p^{\sum_j y_j(i,t)}(1-p)^{J-\sum_j y_j(i,t)}\psi_{condl}+(1-\psi_{condl})\cdot \mathrm{I}\left(\sum_j y_j(i,t=0)\right)}$$

$$= p_{(i,t)}$$

where $y(i,t) \sim Binomial(J,p)$ and $p = z(i,t)p_t$.

$$\mathrm{z(i,t)|\cdot \sim Bernoullli(p_{(i,t)})}$$

Gibbs Sampling

- **Sampling Steps**

1. Initial parameters $\psi_1, p_1, \phi_1, \gamma_1, z(1,1), \ldots, z(R,1)$
2. Draw $\psi_1^{(k)}$ from $\mathrm{Beta}\left(\sum_i z(i,1)^{(k-1)} + 1,\ R - \sum_i z(i,1)^{(k-1)} + 1\right)$

$p_t^{(k)}$ from $\mathrm{Beta}(\sum_i \sum_j y_j(i,t) \times z(i,t)^{(k-1)} + 1,\ R \times J - \sum_i \sum_j y_j(i,t) \times z(i,t)^{(k-1)} + 1)$

$\phi_t^{(k)}$ from $\mathrm{Beta}(n_4^{(k-1)} + 1, n_3^{(k-1)} + 1)$

$\gamma_t^{(k)}$ from $\mathrm{Beta}(n_2^{(k-1)} + 1, n_1^{(k-1)} + 1)$

$z(i,t)^{(k)}$ from Bernoulllli$(p_{(i,t)}^{(k)})$

PART FOUR

Code and Results

Bayesian Statistics

Crossbill Data Set (2001 – 2004)

267 plots × 4 years × 3 surveys/year (12 detection columns)

- Detection variables: det011 … det043 (0/1/NA)
- Select 12 detection columns, build 3-D array y[site, rep, year]
- Keep NA as missing detections

```r
y <- array(NA_integer_, dim = c(nsite, nrep, nyear),
           dimnames = list(site = 1:nsite, rep = 1:nrep, year = yrs))

for (k in seq_len(nrow(det_info))) {
  t <- which(yrs == det_info$yr[k])
  j <- det_info$rep[k]
  y[, j, t] <- cross[[det_info$col[k]]]
}
```

...1 <dbl>	id <dbl>	ele <dbl>	forest <dbl>	surveys <dbl>	det991 <dbl>	det992 <dbl>	det993 <dbl>	det001 <dbl>	det002 <dbl>	det003 <dbl>	det011 <dbl>	det012 <dbl>	det013 <dbl>	det021 <dbl>	det022 <dbl>	det023 <dbl>	det031 <dbl>	det032 <dbl>	det033 <dbl>
1	1	450	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	450	21	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	1050	32	3	NA	NA	NA	0	0	0	1	1	0	0	0	0	1	1	1
4	4	950	9	3	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
5	5	1150	35	3	0	0	0	1	1	1	0	0	1	1	0	0	1	1	0
6	6	550	2	3	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	0	0
7	7	750	6	3	0	0	0	0	0	1	NA	NA	NA	NA	NA	NA	0	1	0
8	8	650	60	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
9	9	550	5	3	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA
10	10	550	13	3	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
11	11	1150	50	3	0	0	0	1	0	0	1	0	1	0	0	0	1	1	1
12	12	750	57	3	1	0	0	0	0	0	0	0	1	0	1	1	1	1	0
13	13	1250	84	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	14	750	15	3	0	0	1	0	0	0	NA	NA	NA	0	0	0	0	0	0
15	15	450	17	3	0	0	0	NA	NA	NA	0	0	0	0	0	0	0	0	1
16	16	1050	58	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	17	750	26	3	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1
18	18	1250	32	3	0	0	0	0	0	1	1	1	1	0	1	0	0	0	1
19	19	1250	66	3	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1
20	20	350	45	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	21	750	44	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	22	750	9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	23	550	31	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	24	350	8	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	25	1350	78	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
26	26	550	37	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
27	27	1150	18	3	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1
28	28	1450	54	3	0	1	1	1	1	0	0	0	0	1	1	0	1	1	0
29	29	950	6	3	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1

R Pre-processing Code

Prepare initials, params:

$$\psi_1, p_t, \phi_t, \gamma_t \sim U(0,1)$$

If a sample point is detected at least once in a given year, the potential occupancy state z for that year is first set to 1. Otherwise, it is first set to 0; Otherwise, z is set to 0.

```
init_z <- apply(y, c(1,3), function(v) as.integer(any(v == 1, na.rm = TRUE)))

inits <- function() list(
  psi   = runif(1, 0, 1),
  gamma = runif(nyear - 1, 0, 1),
  phi   = runif(nyear - 1, 0, 1),
  p     = runif(nyear, 0, 1),
  z     = init_z
)

params <- c("psi", "gamma", "phi", "p",
            "psivec", "psi_fs",
            "growthr", "turnover",    # population-level
            "tau_fs")                 # sample-level turnover
```

JAGS Model

- Priors $\psi_1, p_t, \phi_t, \gamma_t \sim U(0,1)$

```
psi ~ dunif(0,1)
for (t in 1:(nyear-1)) {
  gamma[t] ~ dunif(0,1)
  phi[t]   ~ dunif(0,1)
  p[t]     ~ dunif(0,1)
}
p[nyear] ~ dunif(0,1)
```

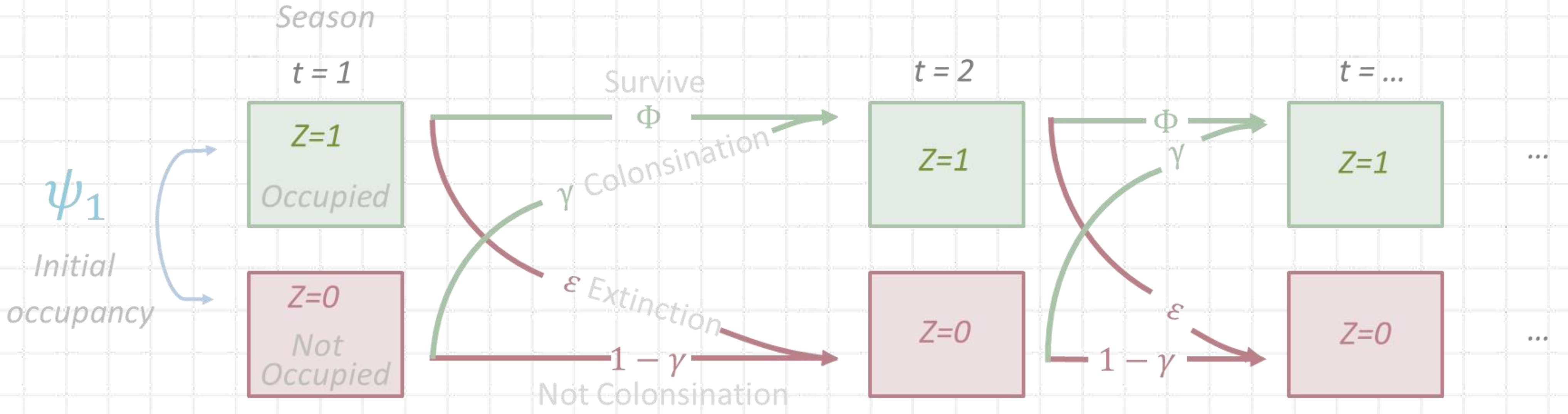
- State process: initial occupancy + survival/colonization

```
for (i in 1:nsite) {
  z[i,1] ~ dbern(psi)
  for (t in 2:nyear) {
    muZ[i,t] <- z[i,t-1] * phi[t-1] + (1 - z[i,t-1]) * gamma[t-1]
    z[i,t]  ~ dbern(muZ[i,t])
  }
}
```

State Model

Site occupation state $z(i,t)$ can be change between each season t and site i. In the first season, $z(i,1)$ is determined by the probability of Initial Occupancy ψ_1. In the subsequent seasons, $z(i,t)$ is determined by the site i and the site's occupancy state in prior season and probabilities of Colonisation γ and Occupied Site 'Survives' ϕ.

For each site,



Occupancy in season 1:

$$z(i,1) \sim Bernoulli(\psi_1)$$

Occupancy in season t:

$$z(i,t) \sim Bernoulli\left(z(i,t-1)\phi_{t-1} + (1 - z(i,t-1))\gamma_{t-1}\right)$$

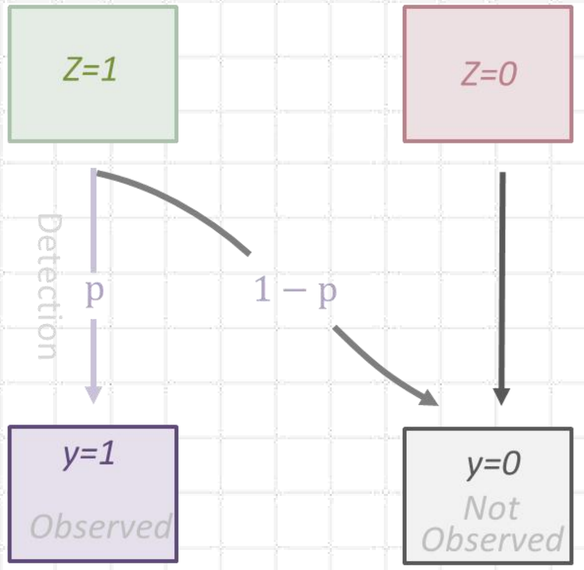
JAGS Model

- Observation process: y ~ Bernoulli(z × p)

```
for (i in 1:nsite) {
  for (t in 1:nyear) {
    for (j in 1:nrep) {
      Py[i,j,t] <- z[i,t] * p[t]
      y[i,j,t]  ~ dbern(Py[i,j,t])
    }
  }
}
```

Observation Model

When a site is survived, the observed occupancy state y is determined by probability of Detection p.



Observation of survey j during time t:

$$y_j(i,t) \sim Bernoulli(z(i,t)p_t)$$

- Derived quantities: population ψ, growth rate, turnover

```
psivec[1] <- psi
count[1]  <- sum(z[,1])
psi_fs[1] <- count[1] / nsite

for (t in 2:nyear) {

  # ---- population quantities ----
  psivec[t]   <- psivec[t-1] * phi[t-1] + (1 - psivec[t-1]) * gamma[t-1]
  growthr[t] <- psivec[t] / psivec[t-1]
  turnover[t-1] <- (1 - psivec[t-1]) * gamma[t-1] / psivec[t]

  # ---- sample quantities ----
  count[t]  <- sum(z[,t])                                  # occupied sites
  for (i in 1:nsite) {                                     # indicator of NEW occupancy
    newocc[i,t] <- (1 - z[i,t-1]) * z[i,t]
  }
  newcount[t-1] <- sum(newocc[,t])                         # newly occupied sites
  tau_fs[t-1]   <- newcount[t-1] / count[t]                # sample turnover rate

}
}
```

growth rate as

$$\lambda_t = \frac{\psi_{t+1}}{\psi_t}.$$

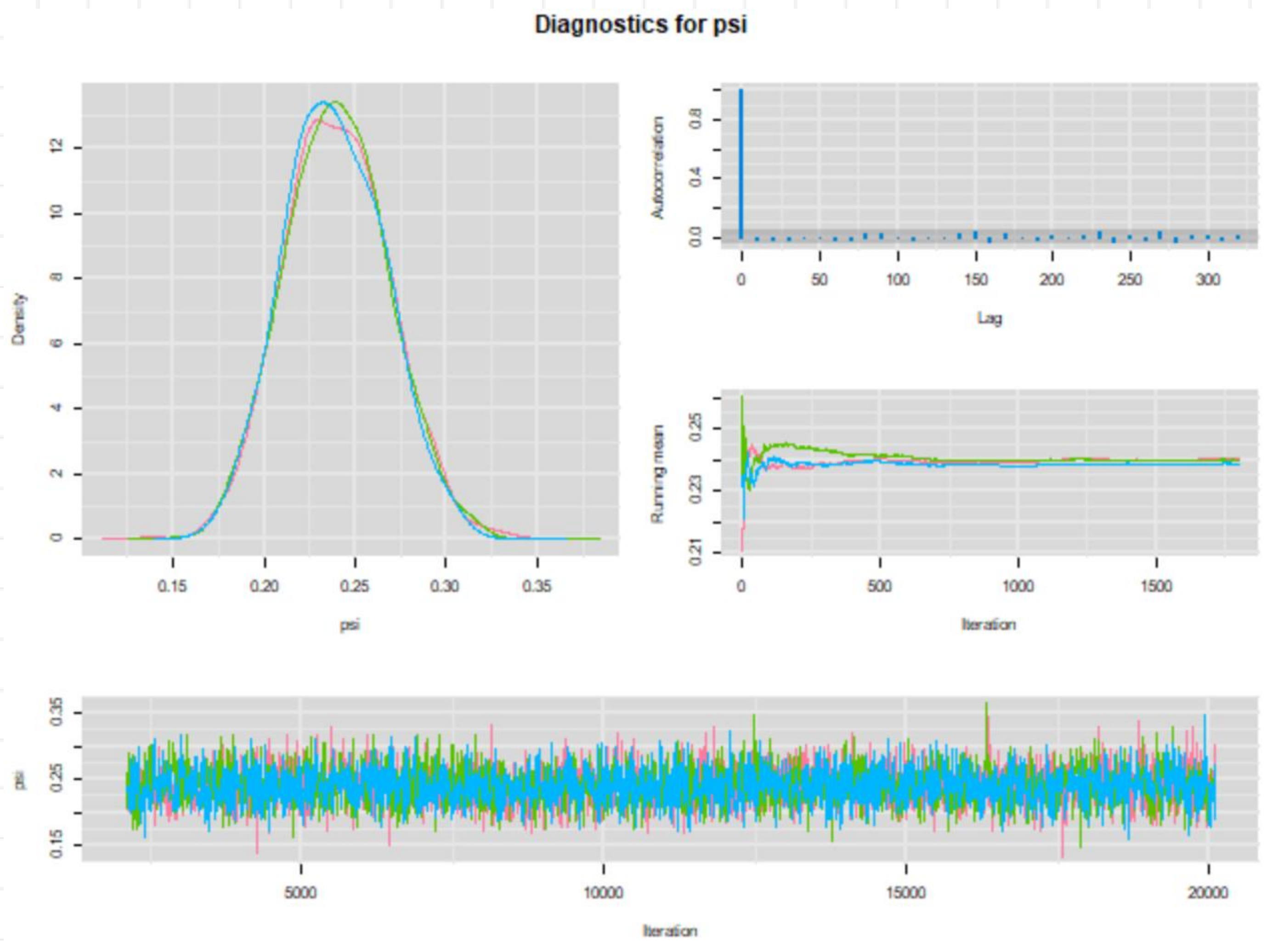
$$\tau_t = \frac{\gamma_{t-1}(1-\psi_{t-1})}{\gamma_{t-1}(1-\psi_{t-1}) + \phi_{t-1}\psi_{t-1}}$$

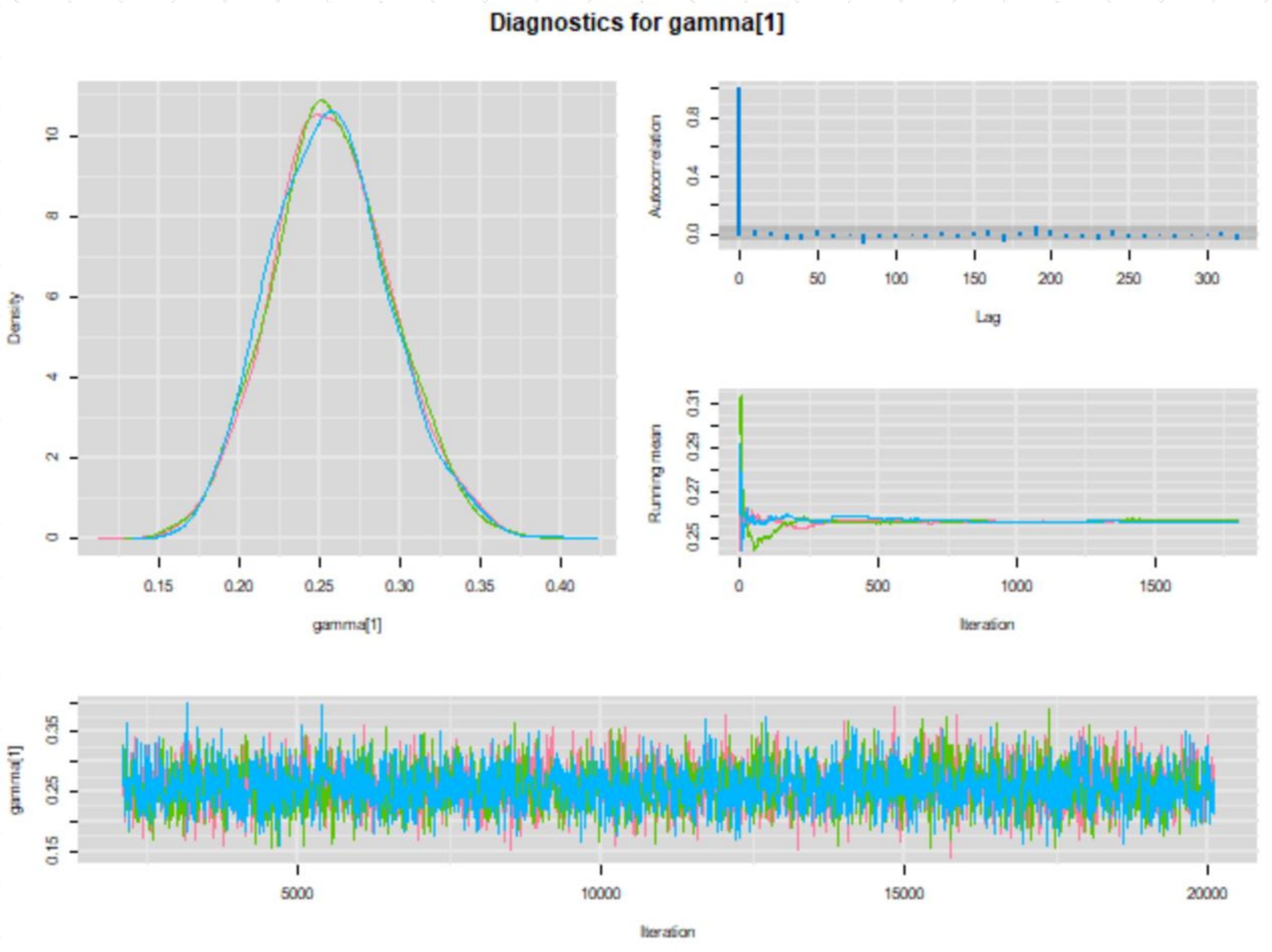
$$\psi_t = \psi_{t-1}\phi_{t-1} + (1-\psi_{t-1})\gamma_{t-1}$$

Running & Diagnostics

```r
fit <- jags(data               = bugs_data,
            inits              = inits,
            parameters.to.save = params,
            model.file         = "crossbill_jags.txt",
            n.chains           = 3,
            n.iter             = 20000,
            n.burnin           = 2000,
            n.thin             = 10,
            parallel           = TRUE)
```

```r
mcmcplot(fit$samples, parms = c("psi", "gamma[1]", "phi[1]", "p[1]"))
```

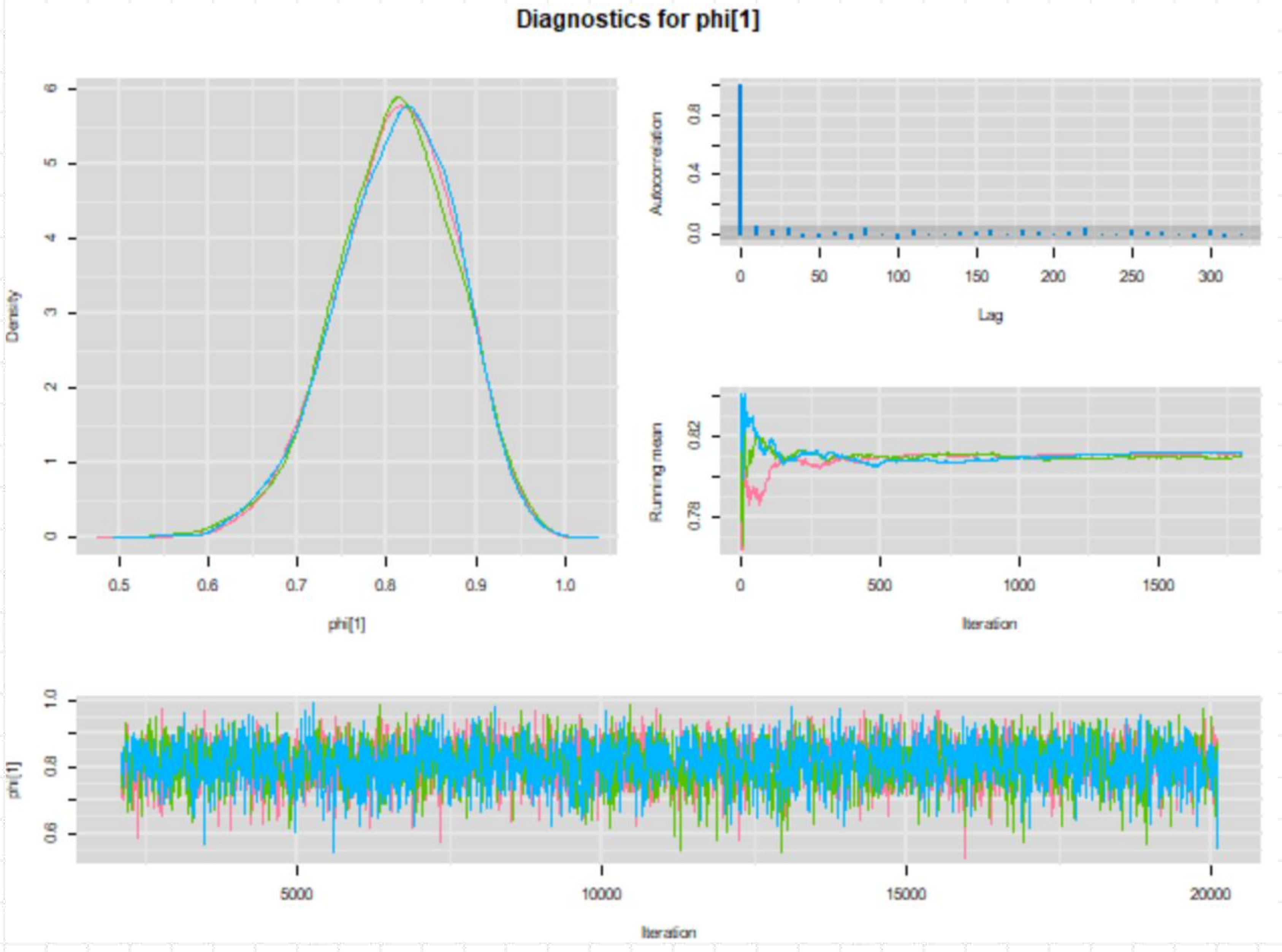


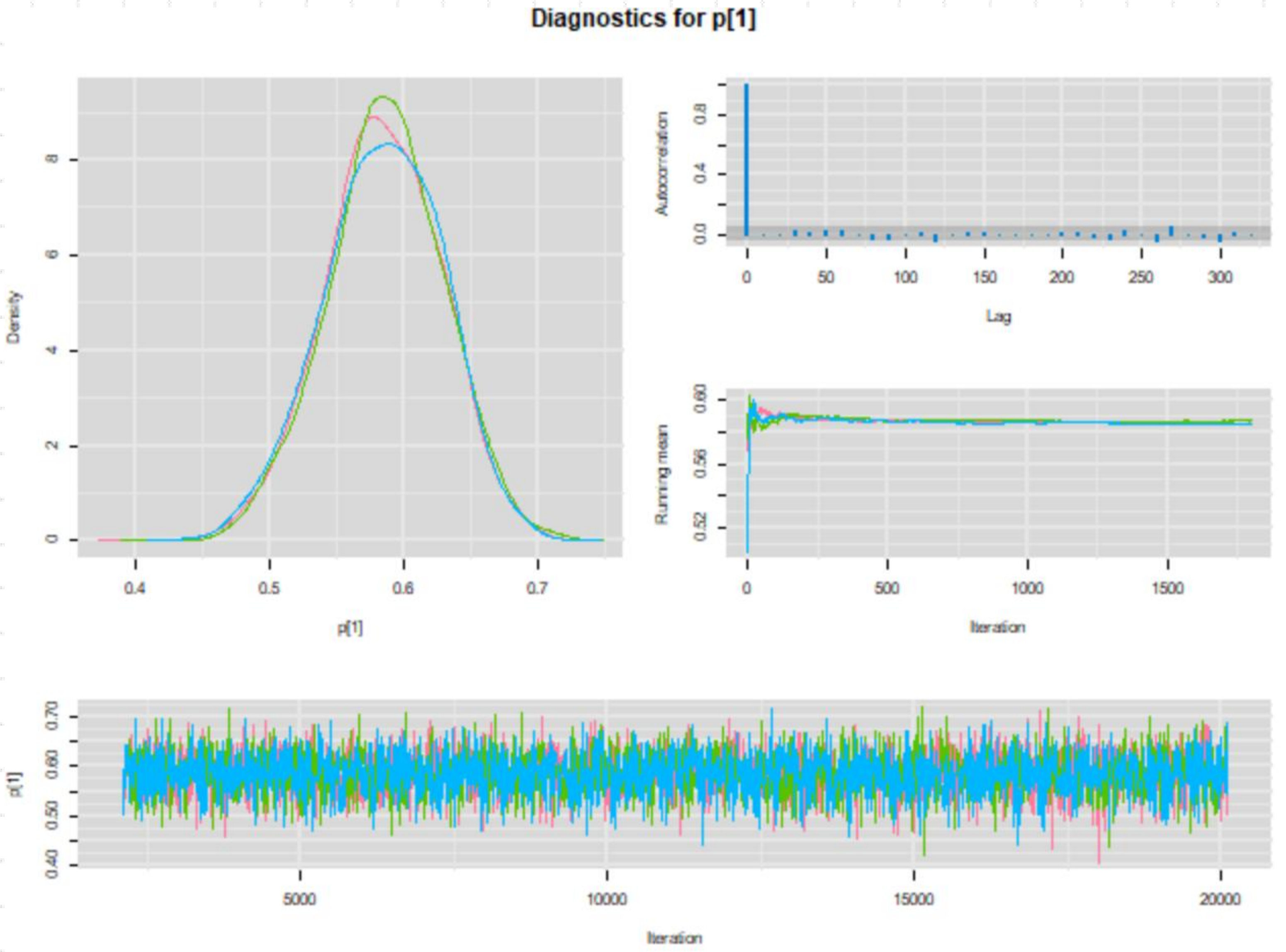


Running & Diagnostics

```r
fit <- jags(data               = bugs_data,
            inits              = inits,
            parameters.to.save = params,
            model.file         = "crossbill_jags.txt",
            n.chains           = 3,
            n.iter             = 20000,
            n.burnin           = 2000,
            n.thin             = 10,
            parallel           = TRUE)
```

```r
mcmcplot(fit$samples, parms = c("psi", "gamma[1]", "phi[1]", "p[1]"))
```





Results Comparison

Parameter	Mean	SD	$q_{0.025}$	$q_{0.500}$	$q_{0.975}$
p_1	0.584	0.044	0.493	0.584	0.666
p_2	0.493	0.037	0.422	0.493	0.564
p_3	0.566	0.033	0.504	0.566	0.629
p_4	0.574	0.037	0.499	0.574	0.643
ϕ_1	0.806	0.069	0.656	0.812	0.931
ϕ_2	0.855	0.046	0.758	0.858	0.938
ϕ_3	0.682	0.053	0.576	0.682	0.791
γ_1	0.259	0.037	0.189	0.257	0.334
γ_2	0.190	0.041	0.114	0.189	0.273
γ_3	0.071	0.029	0.022	0.068	0.133
ψ_1	0.242	0.029	0.190	0.241	0.300
ψ_2	0.391	0.035	0.323	0.390	0.461
ψ_3	0.450	0.034	0.386	0.450	0.517
ψ_4	0.346	0.032	0.286	0.345	0.409
$\psi_1^{(fs)}$	0.240	0.0124	0.222	0.237	0.271
$\psi_2^{(fs)}$	0.389	0.0210	0.353	0.387	0.436
$\psi_3^{(fs)}$	0.449	0.0149	0.425	0.447	0.481
$\psi_4^{(fs)}$	0.345	0.0148	0.320	0.342	0.380
λ_1	1.635	0.201	1.284	1.619	2.085
λ_2	1.157	0.099	0.978	1.153	1.371
λ_3	0.770	0.065	0.651	0.768	0.908
τ_1	0.499	0.058	0.384	0.500	0.610
τ_2	0.259	0.056	0.151	0.258	0.371
τ_3	0.113	0.046	0.032	0.110	0.213
$\tau_1^{(fs)}$	0.498	0.032	0.429	0.500	0.555
$\tau_2^{(fs)}$	0.254	0.041	0.170	0.256	0.328
$\tau_3^{(fs)}$	0.104	0.034	0.034	0.103	0.170

```
##                 mean     sd    2.5%      50%    97.5%
## psi            0.239  0.028    0.188    0.238    0.298
## gamma[1]       0.258  0.037    0.190    0.256    0.333
## gamma[2]       0.188  0.040    0.114    0.187    0.269
## gamma[3]       0.070  0.028    0.021    0.068    0.131
## phi[1]         0.813  0.067    0.673    0.816    0.931
## phi[2]         0.856  0.045    0.759    0.859    0.935
## phi[3]         0.685  0.053    0.581    0.685    0.790
## p[1]           0.586  0.045    0.497    0.586    0.671
## p[2]           0.493  0.037    0.421    0.492    0.567
## p[3]           0.565  0.032    0.503    0.565    0.627
## p[4]           0.573  0.038    0.498    0.574    0.646
## psivec[1]      0.239  0.028    0.188    0.238    0.298
## psivec[2]      0.391  0.036    0.323    0.390    0.463
## psivec[3]      0.449  0.034    0.383    0.449    0.517
## psivec[4]      0.347  0.032    0.285    0.346    0.411
## psi_fs         0.237  0.012    0.217    0.236    0.266
## growthr[2]     1.648  0.196    1.305    1.635    2.077
## growthr[3]     1.156  0.096    0.986    1.150    1.361
## growthr[4]     0.773  0.066    0.648    0.770    0.908
## turnover[1]    0.502  0.057    0.387    0.505    0.609
## turnover[2]    0.256  0.056    0.147    0.255    0.367
## turnover[3]    0.112  0.046    0.031    0.109    0.208
## tau_fs[1]      0.499  0.033    0.427    0.500    0.556
## tau_fs[2]      0.252  0.041    0.165    0.254    0.328
## tau_fs[3]      0.103  0.033    0.035    0.103    0.169
## deviance    1517.098 37.306 1447.630 1515.759 1592.427
```

PART FIVE

Improvements

Bayesian Statistics

Adding Covariates

precipitation & temperature & forest-area

$$logit(\gamma_t) = \beta_0 + \beta_1 \cdot \mathrm{precip}_t, \quad t = 1, 2, 3$$

$$logit(\phi_t) = \eta_0 + \eta_1 \cdot \mathrm{precip}_t, \quad t = 1, 2, 3$$

$$logit(p_t) = \alpha_0 + \alpha_1 \cdot \mathrm{precip}_t, \quad t = 1, 2, 3, 4$$

weakly informative prior

$$\beta_0 \sim \mathcal{N}(0, 100), \quad \beta_1 \sim \mathcal{N}(0, 100)$$

$$\eta_0 \sim \mathcal{N}(0, 100), \quad \eta_1 \sim \mathcal{N}(0, 100)$$

$$\alpha_0 \sim \mathcal{N}(0, 100), \quad \alpha_1 \sim \mathcal{N}(0, 100)$$

Model Comparison

	Intercept	Coefficient	Deviance	DIC
add precipitation to γt	-1.623	-0.511	1500.105	2190.337
add precipitation to φt	1.429	-0.582	1520.727	2228.706
add precipitation to pt	0.203	0.120	1513.370	2228.804
add temperature to γt	-1.724	-0.676	1523.391	2211.273
add temperature to φt	1.566	-0.646	1533.745	2289.692
add temperature to pt	0.228	-0.061	1509.025	2177.631
add forest-area to γt	-1.751	-0.753	1521.095	2202.394
add forest-area to φt	1.533	-0.621	1530.826	2231.781
add forest-area to pt	0.212	0.042	1511.412	2218.879

add temperature to p_t

```
DIC info: (pD = var(deviance)/2)
pD = 668.6 and DIC = 2177.631
DIC is an estimate of expected predictive error (lower is better).
```

	mean	sd	2.5%	50%	97.5%
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
alpha0	0.228	0.079	0.071	0.228	0.381
alpha1	-0.061	0.086	-0.229	-0.063	0.106
phi[1]	0.777	0.068	0.637	0.779	0.901
phi[2]	0.865	0.045	0.770	0.868	0.944
phi[3]	0.681	0.052	0.577	0.681	0.780
gamma[1]	0.243	0.034	0.181	0.242	0.313
gamma[2]	0.210	0.038	0.141	0.209	0.287
gamma[3]	0.066	0.029	0.017	0.063	0.127
p[1]	0.567	0.026	0.516	0.567	0.618
p[2]	0.548	0.021	0.507	0.548	0.588

	mean	sd	2.5%	50%	97.5%
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
p[3]	0.540	0.028	0.484	0.540	0.595
p[4]	0.571	0.030	0.511	0.571	0.628
psivec[1]	0.241	0.028	0.189	0.240	0.298
psivec[2]	0.372	0.033	0.309	0.371	0.439
psivec[3]	0.454	0.034	0.389	0.454	0.523
psivec[4]	0.345	0.032	0.283	0.345	0.411
psi_fs[1]	0.240	0.011	0.221	0.240	0.262
psi_fs[2]	0.371	0.015	0.345	0.371	0.401
psi_fs[3]	0.454	0.015	0.427	0.453	0.487
psi_fs[4]	0.344	0.014	0.318	0.341	0.375

	mean	sd	2.5%	50%	97.5%
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
growthr[2]	1.556	0.183	1.233	1.543	1.952
growthr[3]	1.226	0.097	1.059	1.221	1.434
growthr[4]	0.762	0.064	0.640	0.761	0.889
turnover[1]	0.496	0.057	0.384	0.495	0.608
turnover[2]	0.291	0.052	0.192	0.291	0.394
turnover[3]	0.104	0.045	0.026	0.102	0.201
tau_fs[1]	0.492	0.031	0.427	0.495	0.549
tau_fs[2]	0.287	0.033	0.221	0.288	0.348
tau_fs[3]	0.095	0.034	0.023	0.096	0.160
deviance	1509.025	36.579	1439.628	1508.476	1582.913

Parameter	Mean	SD	$q_{0.025}$	$q_{0.500}$	$q_{0.975}$
p_1	0.584	0.044	0.493	0.584	0.666
p_2	0.493	0.037	0.422	0.493	0.564
p_3	0.566	0.033	0.504	0.566	0.629
p_4	0.574	0.037	0.499	0.574	0.643
ϕ_1	0.806	0.069	0.656	0.812	0.931
ϕ_2	0.855	0.046	0.758	0.858	0.938
ϕ_3	0.682	0.053	0.576	0.682	0.791
γ_1	0.259	0.037	0.189	0.257	0.334
γ_2	0.190	0.041	0.114	0.189	0.273
γ_3	0.071	0.029	0.022	0.068	0.133
ψ_1	0.242	0.029	0.190	0.241	0.300
ψ_2	0.391	0.035	0.323	0.390	0.461
ψ_3	0.450	0.034	0.386	0.450	0.517
ψ_4	0.346	0.032	0.286	0.345	0.409
$\psi_1^{(fs)}$	0.240	0.0124	0.222	0.237	0.271
$\psi_2^{(fs)}$	0.389	0.0210	0.353	0.387	0.436
$\psi_3^{(fs)}$	0.449	0.0149	0.425	0.447	0.481
$\psi_4^{(fs)}$	0.345	0.0148	0.320	0.342	0.380
λ_1	1.635	0.201	1.284	1.619	2.085
λ_2	1.157	0.099	0.978	1.153	1.371
λ_3	0.770	0.065	0.651	0.768	0.908
τ_1	0.499	0.058	0.384	0.500	0.610
τ_2	0.259	0.056	0.151	0.258	0.371
τ_3	0.113	0.046	0.032	0.110	0.213
$\tau_1^{(fs)}$	0.498	0.032	0.429	0.500	0.555
$\tau_2^{(fs)}$	0.254	0.041	0.170	0.256	0.328
$\tau_3^{(fs)}$	0.104	0.034	0.034	0.103	0.170

add precipitation to γ_t & add temperature to p_t

```
DIC info: (pD = var(deviance)/2)
pD = 685.9 and DIC = 2184.249
DIC is an estimate of expected predictive error (lower is better).
```

	mean <dbl>	sd <dbl>	2.5% <dbl>	50% <dbl>	97.5% <dbl>
beta0	-1.626	0.145	-1.919	-1.622	-1.349
beta1	-0.552	0.201	-0.959	-0.544	-0.174
alpha0	0.236	0.078	0.087	0.237	0.390
alpha1	-0.024	0.085	-0.186	-0.024	0.145
phi[1]	0.768	0.067	0.629	0.770	0.891
phi[2]	0.854	0.046	0.755	0.857	0.936
phi[3]	0.679	0.054	0.574	0.679	0.786
gamma[1]	0.174	0.020	0.137	0.173	0.214
gamma[2]	0.250	0.034	0.187	0.249	0.320
gamma[3]	0.102	0.027	0.056	0.101	0.163

	mean <dbl>	sd <dbl>	2.5% <dbl>	50% <dbl>	97.5% <dbl>
p[1]	0.563	0.027	0.510	0.563	0.614
p[2]	0.555	0.020	0.516	0.555	0.594
p[3]	0.552	0.027	0.500	0.552	0.604
p[4]	0.564	0.031	0.504	0.564	0.622
psivec[1]	0.247	0.029	0.193	0.247	0.307
psivec[2]	0.321	0.027	0.269	0.321	0.376
psivec[3]	0.443	0.033	0.378	0.444	0.508
psivec[4]	0.358	0.032	0.297	0.357	0.424
psi_fs[1]	0.246	0.013	0.225	0.243	0.273
psi_fs[2]	0.358	0.012	0.337	0.356	0.386

	mean <dbl>	sd <dbl>	2.5% <dbl>	50% <dbl>	97.5% <dbl>
psi_fs[3]	0.448	0.014	0.423	0.446	0.476
psi_fs[4]	0.348	0.015	0.322	0.348	0.378
growthr[2]	1.307	0.124	1.084	1.300	1.567
growthr[3]	1.387	0.103	1.201	1.381	1.601
growthr[4]	0.809	0.067	0.682	0.808	0.944
turnover[1]	0.410	0.055	0.308	0.408	0.520
turnover[2]	0.382	0.041	0.303	0.382	0.466
turnover[3]	0.160	0.043	0.085	0.157	0.254
tau_fs[1]	0.467	0.035	0.391	0.469	0.527
tau_fs[2]	0.313	0.027	0.258	0.314	0.365
tau_fs[3]	0.120	0.030	0.065	0.118	0.181
deviance	1498.364	37.036	1430.433	1496.774	1573.600

Parameter	Mean	SD	$q_{0.025}$	$q_{0.500}$	$q_{0.975}$
p_1	0.584	0.044	0.493	0.584	0.666
p_2	0.493	0.037	0.422	0.493	0.564
p_3	0.566	0.033	0.504	0.566	0.629
p_4	0.574	0.037	0.499	0.574	0.643
ϕ_1	0.806	0.069	0.656	0.812	0.931
ϕ_2	0.855	0.046	0.758	0.858	0.938
ϕ_3	0.682	0.053	0.576	0.682	0.791
γ_1	0.259	0.037	0.189	0.257	0.334
γ_2	0.190	0.041	0.114	0.189	0.273
γ_3	0.071	0.029	0.022	0.068	0.133
ψ_1	0.242	0.029	0.190	0.241	0.300
ψ_2	0.391	0.035	0.323	0.390	0.461
ψ_3	0.450	0.034	0.386	0.450	0.517
ψ_4	0.346	0.032	0.286	0.345	0.409
$\psi_1^{(fs)}$	0.240	0.0124	0.222	0.237	0.271
$\psi_2^{(fs)}$	0.389	0.0210	0.353	0.387	0.436
$\psi_3^{(fs)}$	0.449	0.0149	0.425	0.447	0.481
$\psi_4^{(fs)}$	0.345	0.0148	0.320	0.342	0.380
λ_1	1.635	0.201	1.284	1.619	2.085
λ_2	1.157	0.099	0.978	1.153	1.371
λ_3	0.770	0.065	0.651	0.768	0.908
τ_1	0.499	0.058	0.384	0.500	0.610
τ_2	0.259	0.056	0.151	0.258	0.371
τ_3	0.113	0.046	0.032	0.110	0.213
$\tau_1^{(fs)}$	0.498	0.032	0.429	0.500	0.555
$\tau_2^{(fs)}$	0.254	0.041	0.170	0.256	0.328
$\tau_3^{(fs)}$	0.104	0.034	0.034	0.103	0.170

2025

Bayesian Statistics (STA306) Final Project

THANK YOU FOR LISTENING

Group Member

Zihan Zhang
Qianyu Yang
Xinyu Zhang
Siqing Li