Developing a Bayesian Procedure to Detect Breakpoints in Time Series

Kathryn Haglich, Sarah Neitzel, Amy Pitts
Mentor: Jeff Liebner

Lafayette College, Unity College, Marist College

Friday, July 20, 2018

NSF Grant #1560222

Introduction

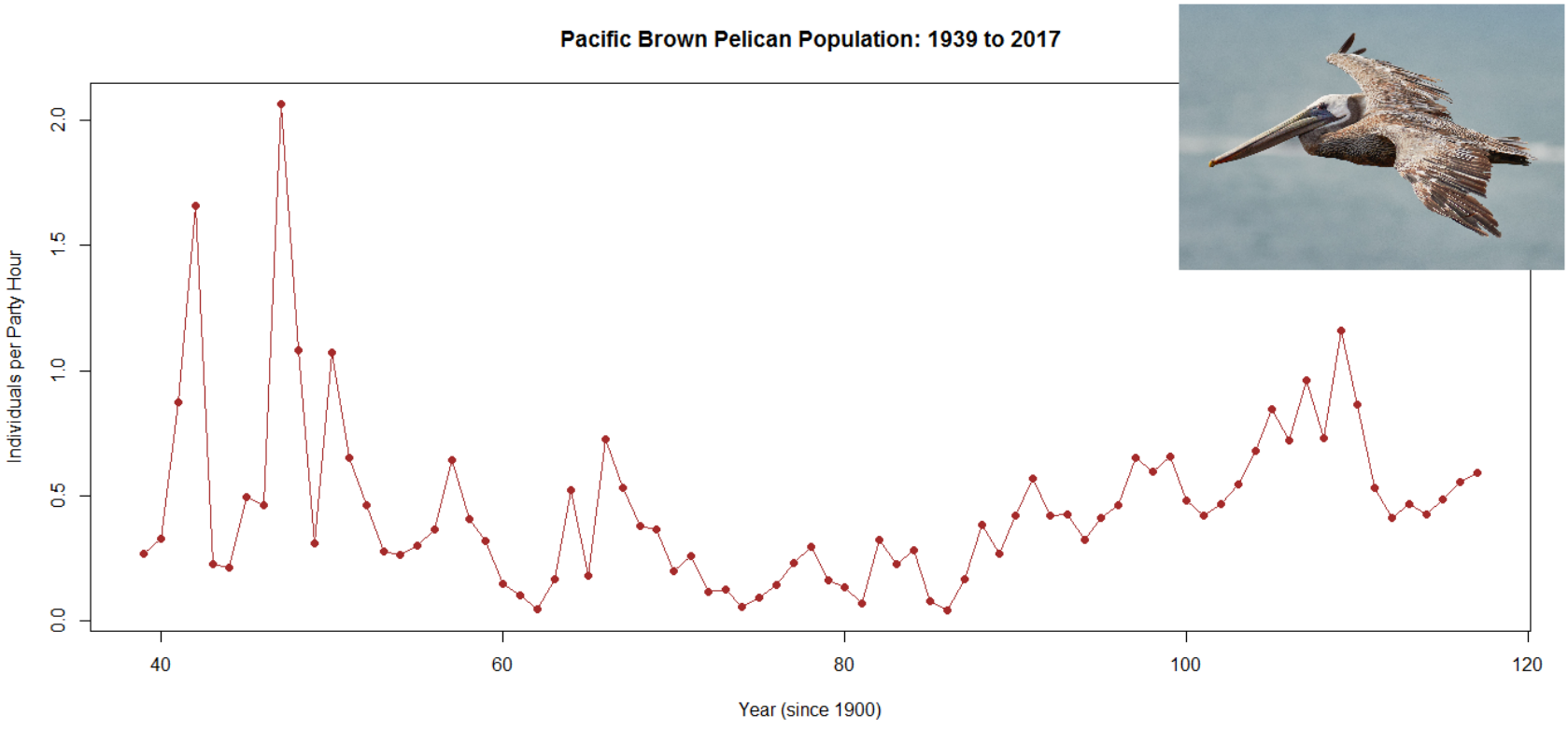


Figure 1: The Pacific brown pelican (*Pelecanus occidentalis*) population from 1939 to 2017 based on the Christmas Bird Count.

Introduction

Goal: to develop a better quantitative method for locating breakpoints in time series data.

Introduction

Goal: to develop a better quantitative method for locating breakpoints in time series data.

Autoregressive Model (AR(p)) Model

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \epsilon_t$$

Autoregressive Models (AR)

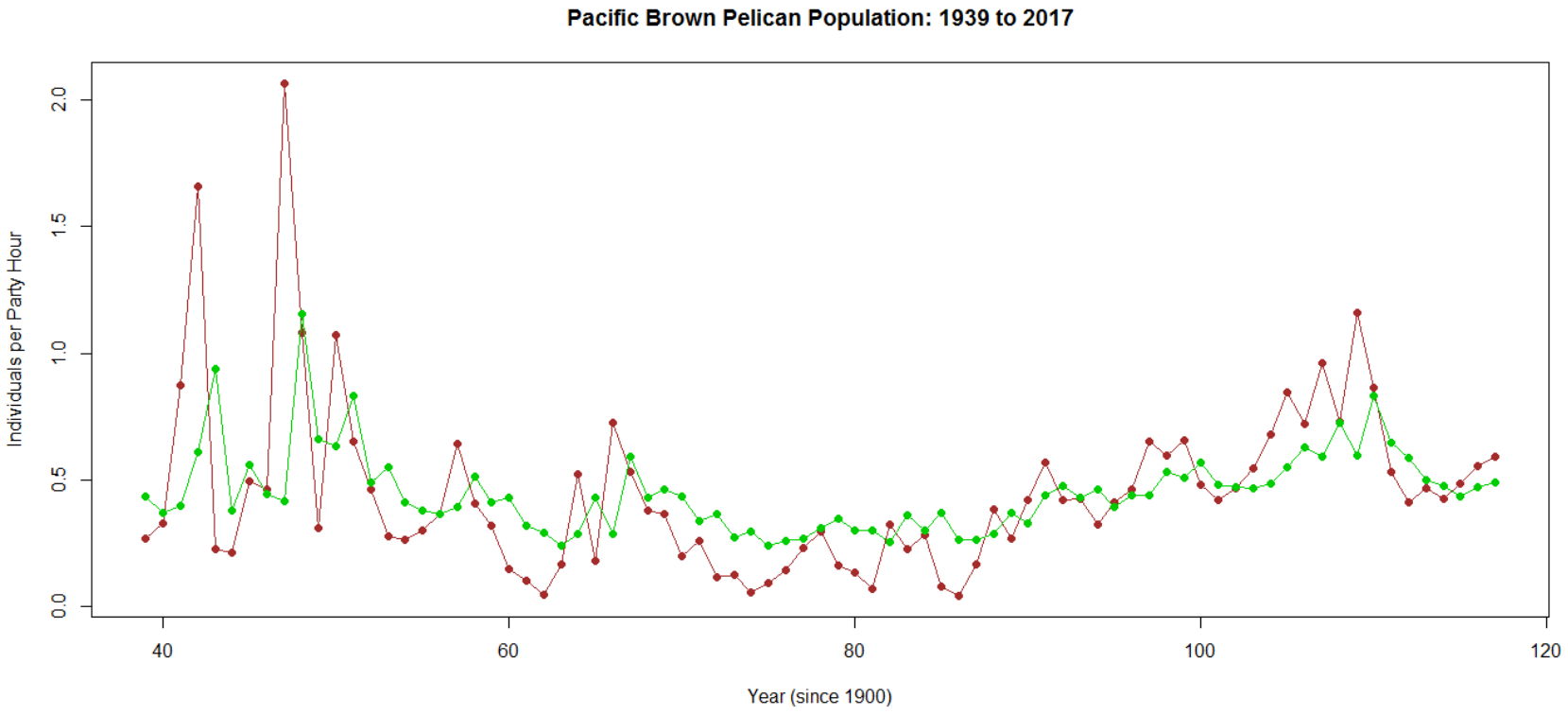


Figure 2: The Pacific brown pelican population (brown) with the fitted values (green) from an AR(3).

Variables of Interest

Goal: to develop a better quantitative method for locating breakpoints in time series data.

Unknown Variables:

- $K =$ the number of breakpoints
- $\tau =$ the location of breakpoints
- $\beta =$ the regression coefficients
- $\sigma =$ the standard deviations

Historical Approaches

- Using "expert opinion"
  - Example: Global Temperature Anomaly Data (Seidel and Lanzante, 2004)
- Bai-Perron Test (1998, 2003)
- Reverse Order Cusum (ROC) (Peseran and Timmermann 2002)

Bai-Perron Method

Bai-Perron Method: (1998, 2003)

- Frequentist approach to identify significant changes in the behavior of a model describing a data set
- Created an algorithm that searches possible breakpoint location and determine the best model based off of the residual sum of squares (sum of squared distances from original data points to fitted model)
- The outcome is a single model conditional on the number of breaks and length of subinterval specified by the user

Our Own Method

Given a time series data set our goal is to explore the number and location of breakpoints.

How do we do this?

- Use a Bayesian framework
- Propose locations for breakpoints

Two Breakpoint Sets

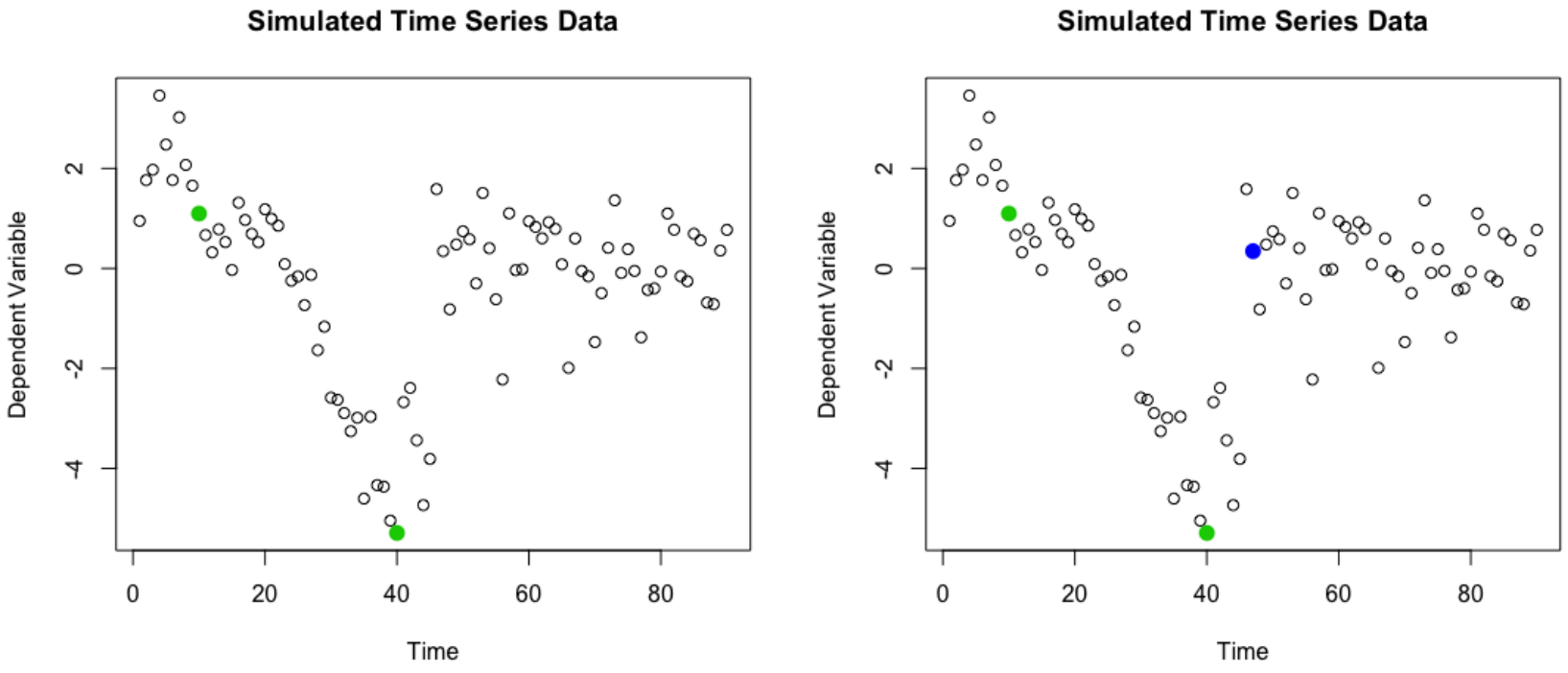


Figure 3: Two proposed breakpoint sets on simulated time series data.

Our Own Method

Given a time series data set, our goal is to explore the number and location of breakpoints.

How do we do this?

- Use a Bayesian framework
- Propose locations for breakpoints
- Evaluate quality of proposals
- Repeat the process using a Markov Chain to obtain a distribution

General Proposal Outline

Inspired by Bayesian Adaptive Regression Splines (BARS) (DiMatteo et al., 2001).

First: start with an initial proposal for breaks (Bai-Perron, middle placement, etc.)

Second: choose a type of MCMC step and then create a proposed breakpoint set from chosen step

Step Options:

- Birth (Random addition of a breakpoint)
- Death (Random deletion of an existing breakpoint)
- Move
  - Jump
  - Jiggle

The Jump Function

Jump:

a function that moves a random break point to any location in the data set

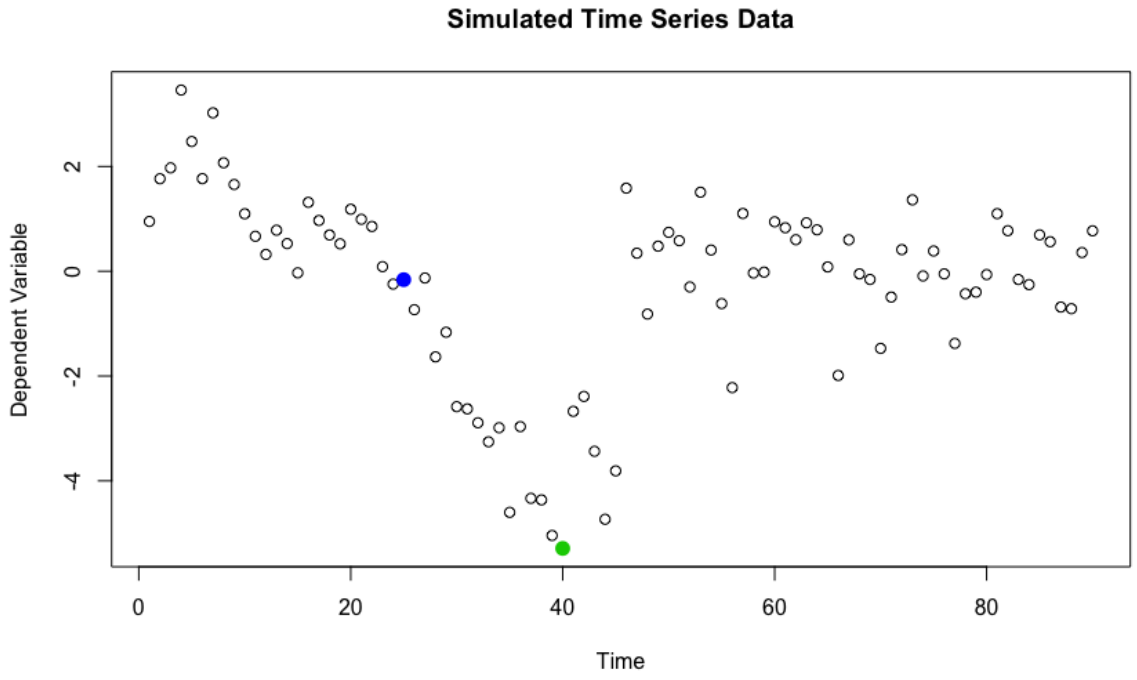


Figure 4: Green is an initial breakpoint and blue is a proposed breakpoint.

The Jiggle Function

Jiggle:

a function that moves a random breakpoint within a given interval around its original location

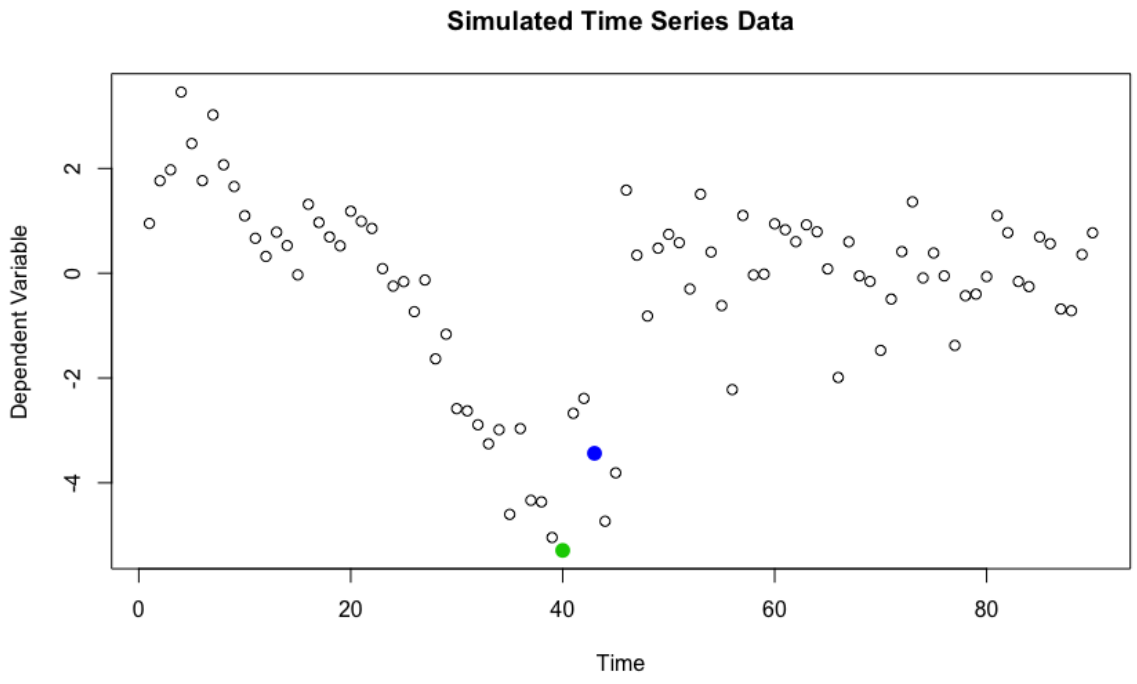


Figure 5: Green is an initial breakpoint and blue is a proposed breakpoint.

Metropolis Hastings

Metropolis Hastings - an algorithm that is a Markov Chain Monte Carlo (MCMC) that is used to sample from the distribution when direct sampling is difficult

Metropolis Hastings

For each iteration, i, of the MCMC, let q be the proposal density of the chosen step, $q(\theta_{new}|\theta_{old})$ where $\theta = \{K, \tau_1, \ldots, \tau_k\}$.

1. Draw $\theta_{new} \sim q(\cdot|\theta_{old})$, where $\theta_{old} = \theta^{(i-1)}$, the values of the parameters from the previous iteration of the MCMC;
2. Compute the ratio

$$r = \frac{g(\theta_{new})}{g(\theta_{old})} \frac{q(\theta_{old}|\theta_{new})}{q(\theta_{new}|\theta_{old})}$$

   where g is the posterior distribution of θ
3. If $r \geq 1$, set $\theta^{(i)} = \theta_{new}$;

   If $r < 1$, set $\theta^{(i)} = \begin{cases} \theta_{new} \text{ with probability } r \\ \theta_{old} \text{ with probability } 1 - r \end{cases}$

Proposal density (q)

Birth: $q(new|old) = c\ b_p \frac{1}{n_{free}}$

Death: $q(new|old) = c\ d_p \frac{1}{K_{old}}$

Move:

Jump: $q(new|old) = 0.25(1 - c(d_p + b_p)) \frac{1}{k_{old} n_{free}}$

Jiggle: $q(new|old) = 0.75(1 - c(d_p + b_p)) \frac{1}{k_{old} j_{free}}$

d_p and b_p are probabilities of choosing a death step or birth step scaled by c. Where n_{free} and j_{free} are the available spaces.

Posterior distribution (g) approximation

The ratio $\frac{g(\theta_{new}|x)}{g(\theta_{old}|x)} = \frac{p(x|\theta_{new})\pi(K_{new},\tau_{new})}{p(x|\theta_{old})\pi(K_{old},\tau_{old})}$ is approximated by $(\frac{-\Delta BIC}{2})\frac{\pi(K_{new},\tau_{new})}{\pi(K_{old},\tau_{old})}$ where ΔBIC is the difference in the values of the Bayesian Information Criterion (Kass and Wasserman, 1995).

We compare the new proposed breakpoint set to the old set using our approximation of the Metropolis Hastings ratio.

$$r \approx (\frac{-\Delta BIC}{2})\frac{\pi(K_{new},\tau_{new})}{\pi(K_{old},\tau_{old})}\frac{q(\theta_{old}|\theta_{new})}{q(\theta_{new}|\theta_{old})}$$

Simulated Time Series Data

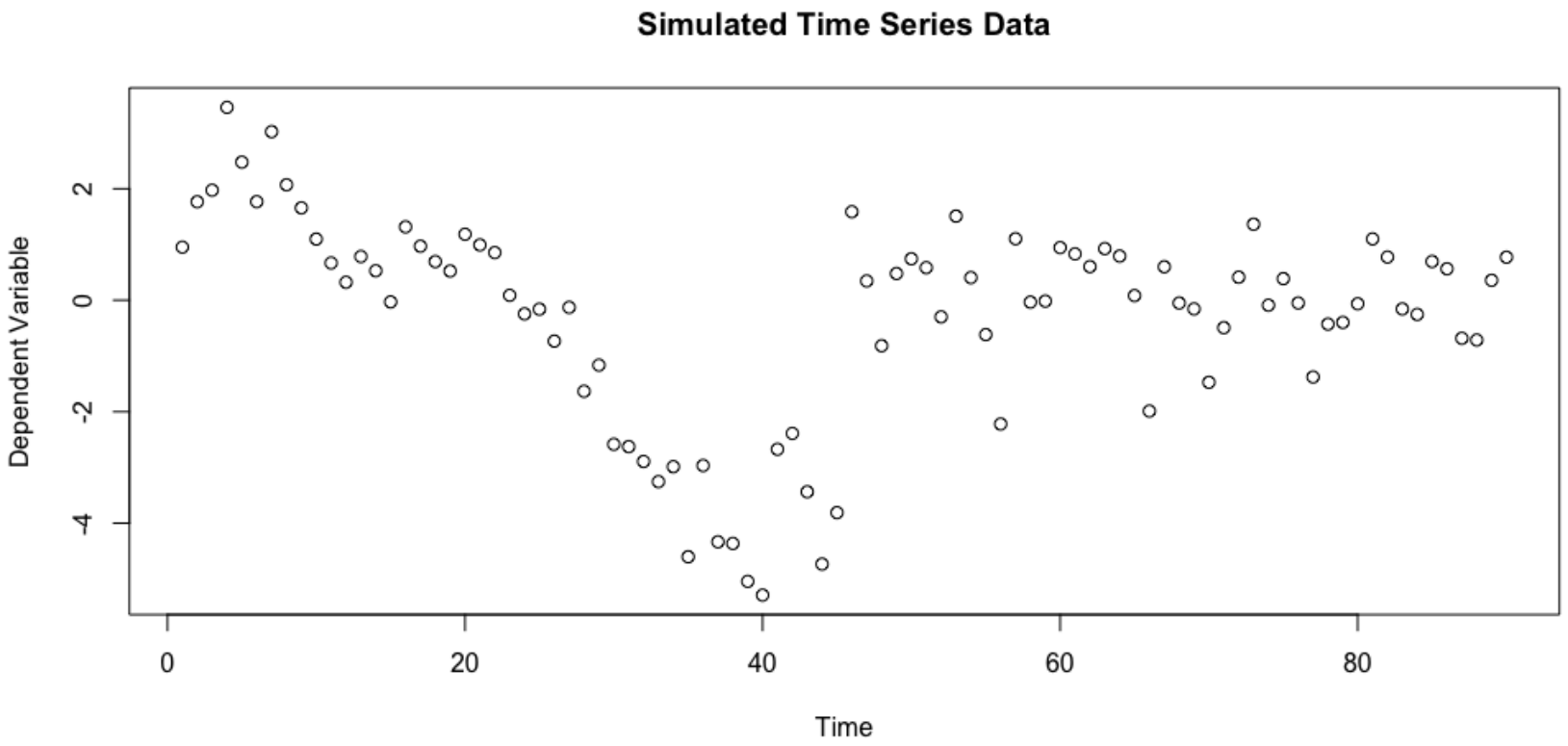
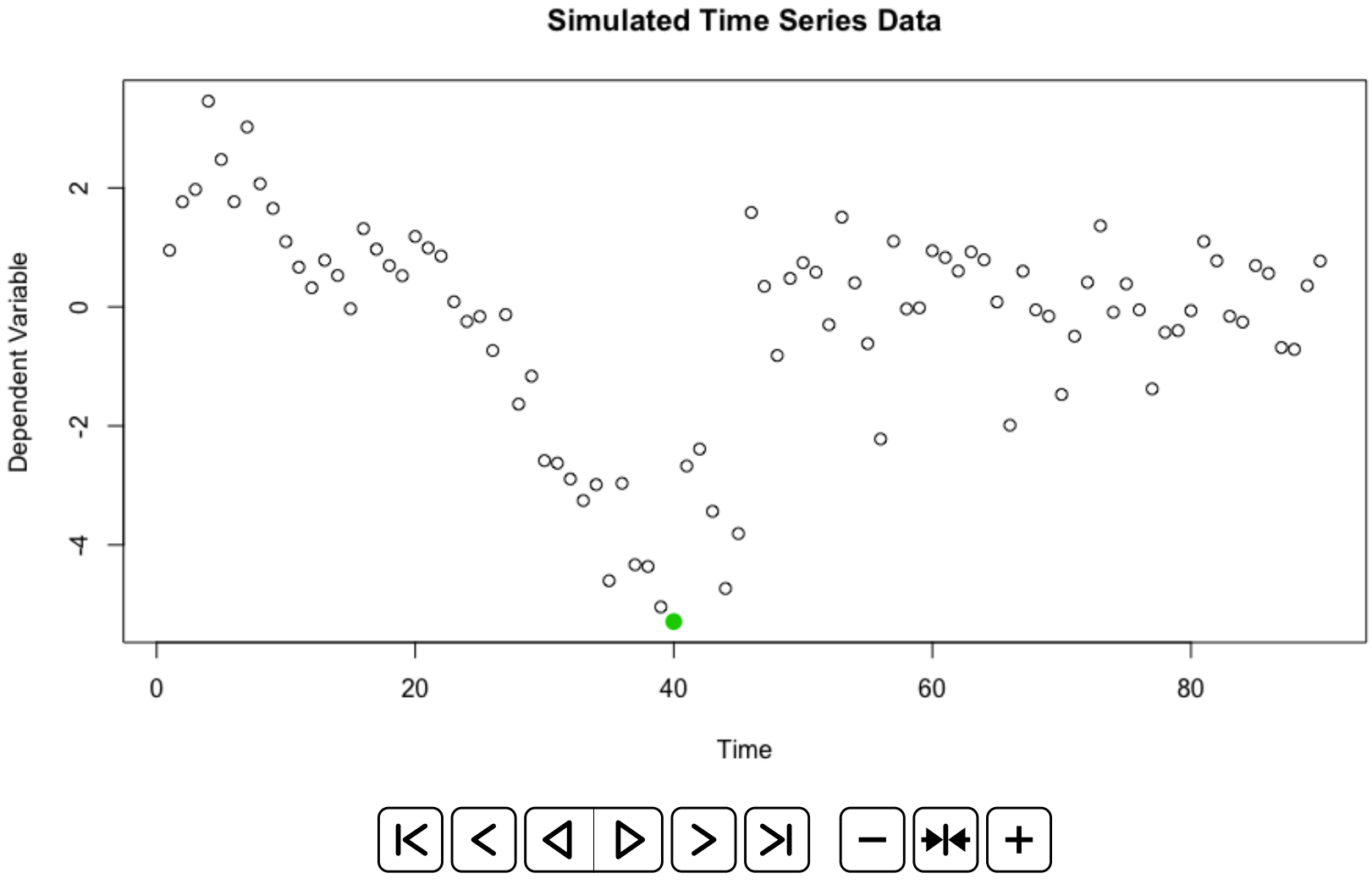


Figure 6: Simulated time series data. For $t < 45$, $y_t = 0.9y_{t-1} + \epsilon_t$ and for $t \geq 45$, $y_t = 0.01y_{t-1} + \epsilon_t$

.

Iteration Run Through



Distribution of K

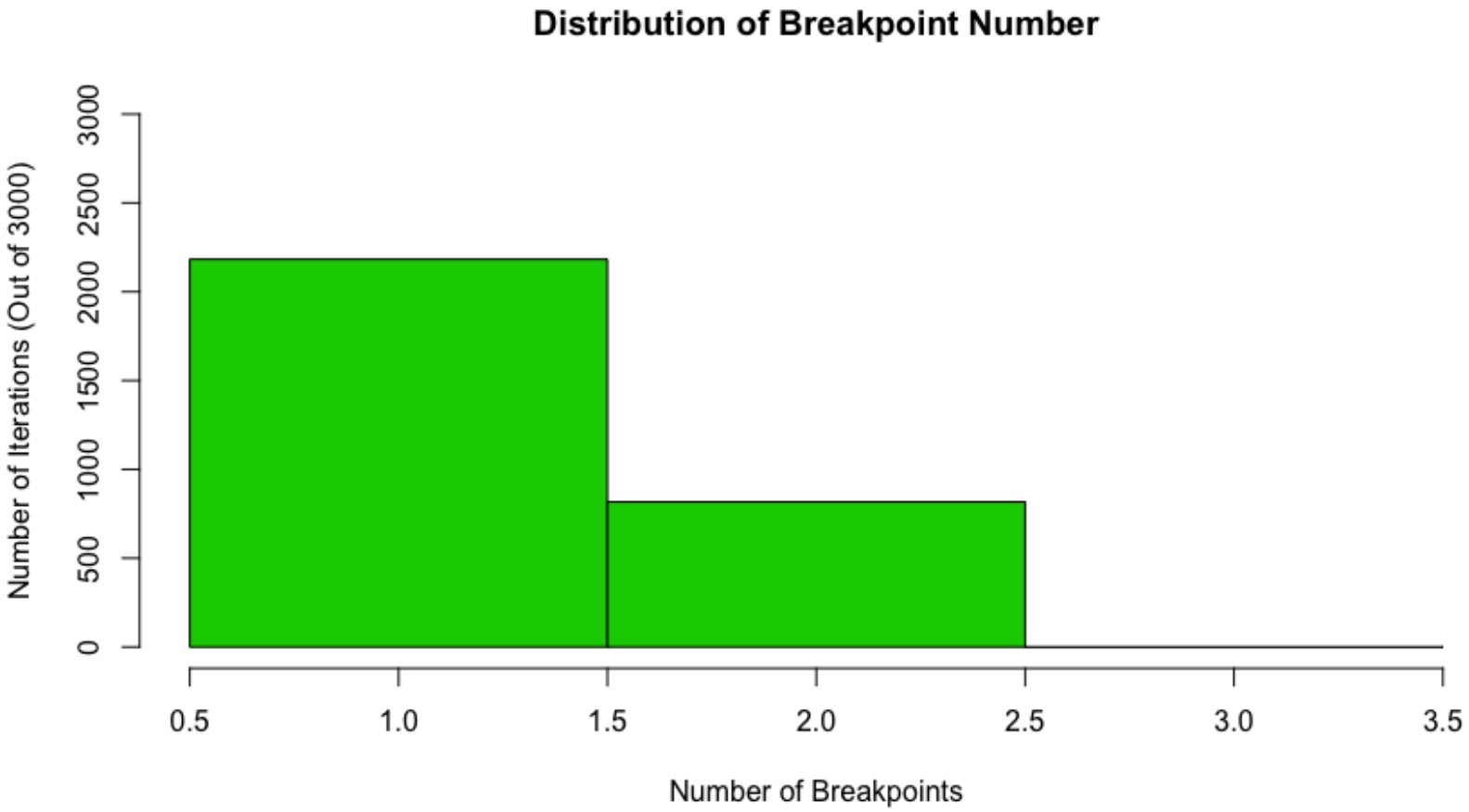


Figure 7: Distribution of breakpoint number (K) in AR training data set.

Distribution of τ

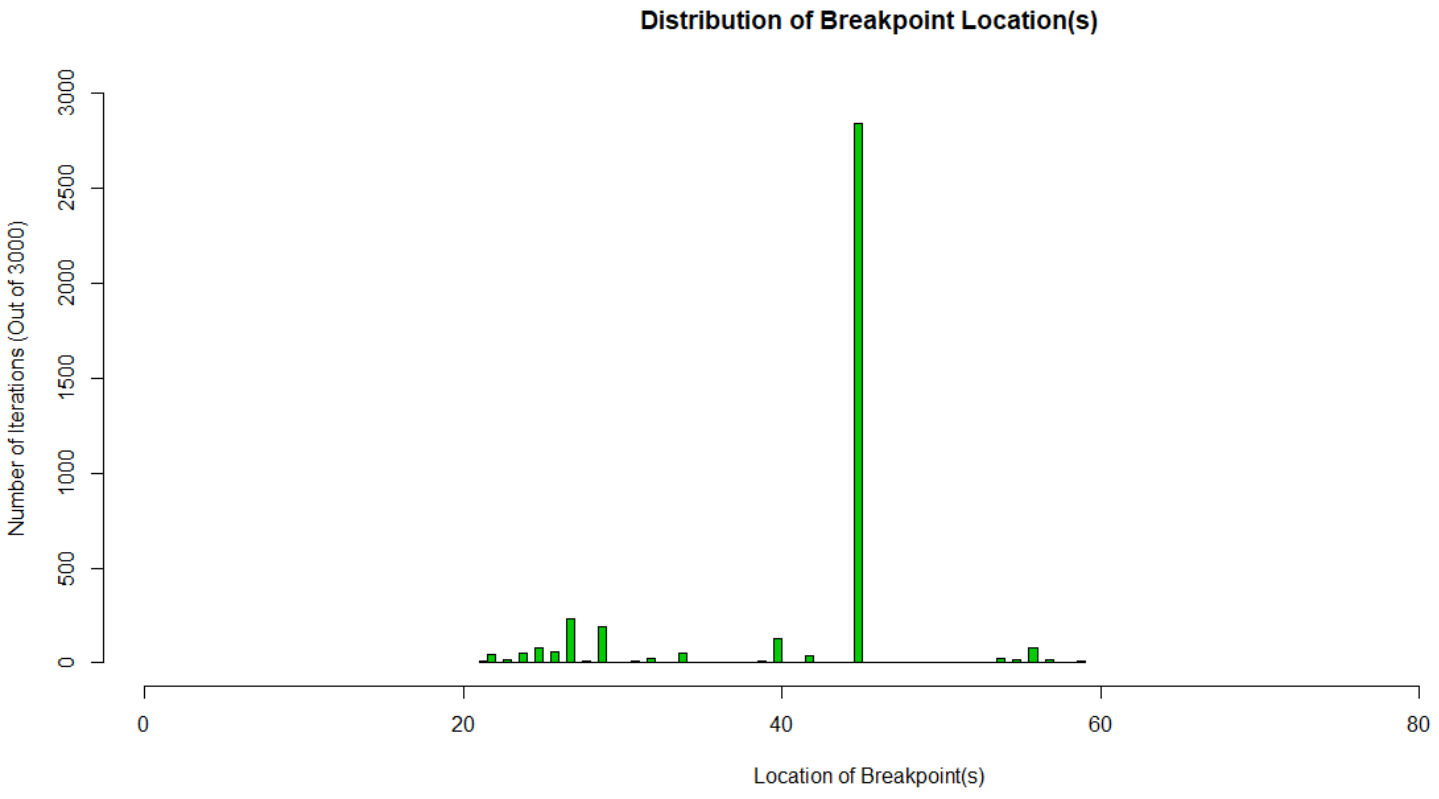


Figure 8: Distribution of breakpoint locations (τ) in AR training data set.

Distribution of β and σ

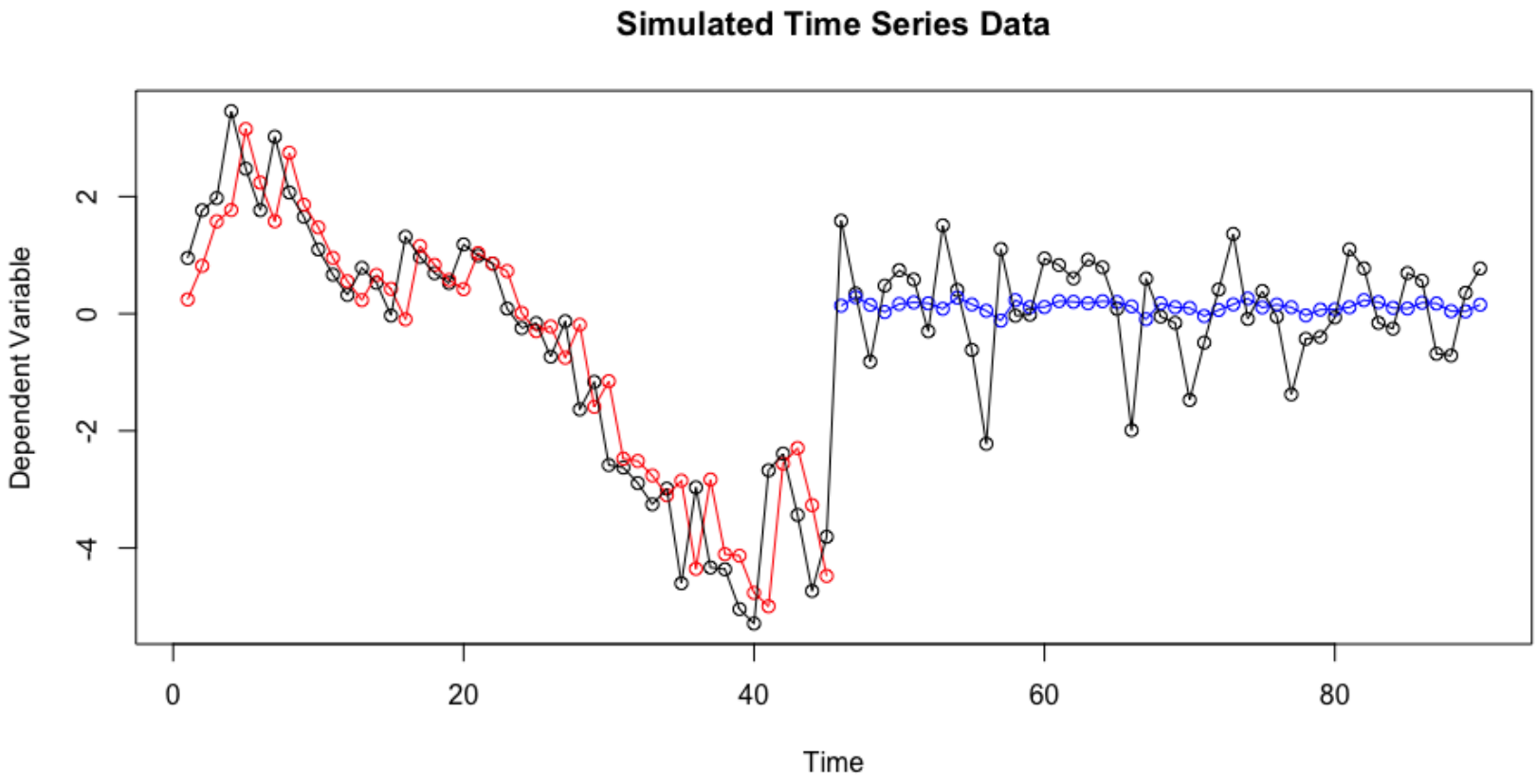


Figure 9: Red points are the mean posterior fitted values for the first subsection and blue are the fitted values for the second subsection.

Pacific Brown Pelican Example

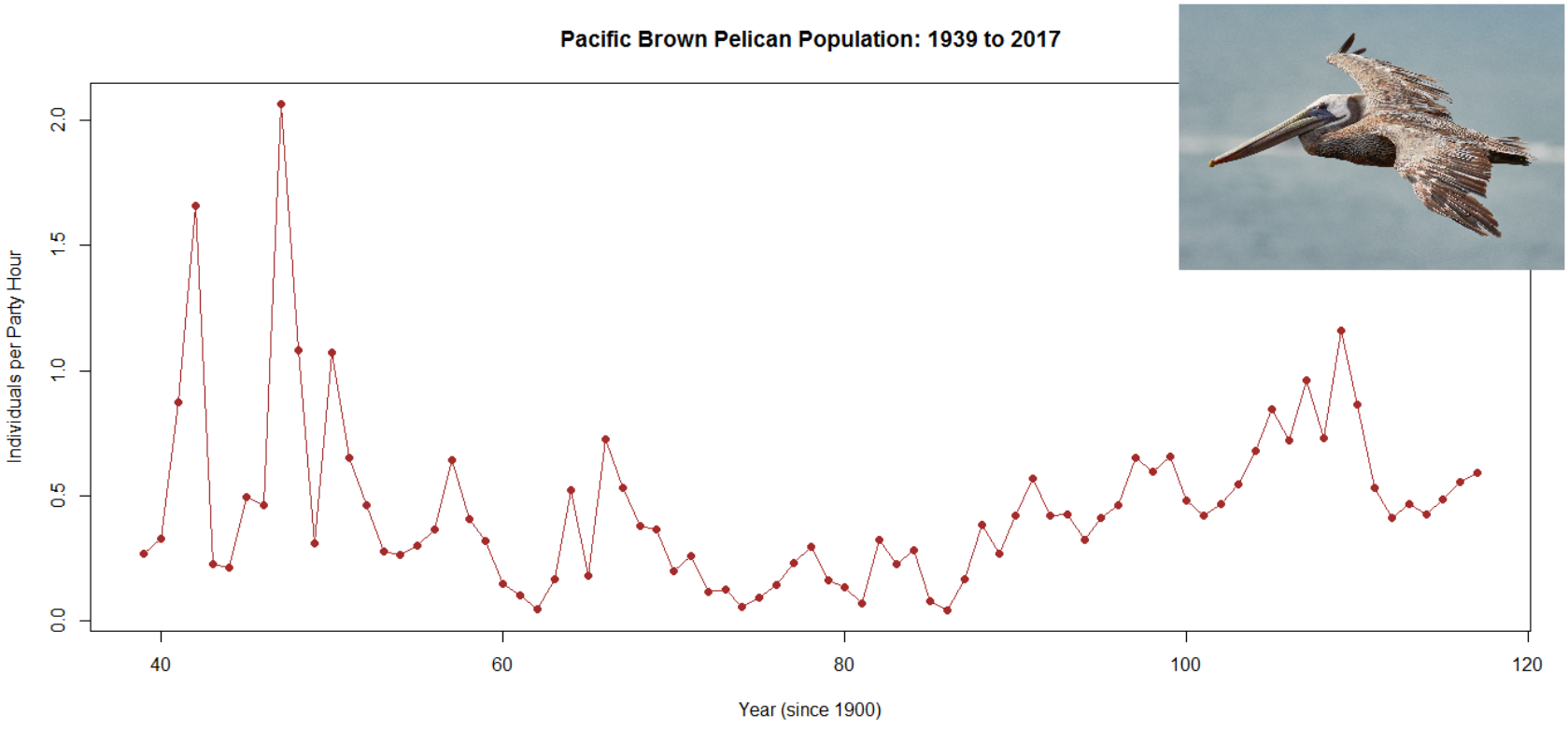


Figure 10: The Pacific brown pelican (*Pelecanus occidentalis*) population from 1939 to 2017 based on the Christmas Bird Count.

Pacific Brown Pelican Example

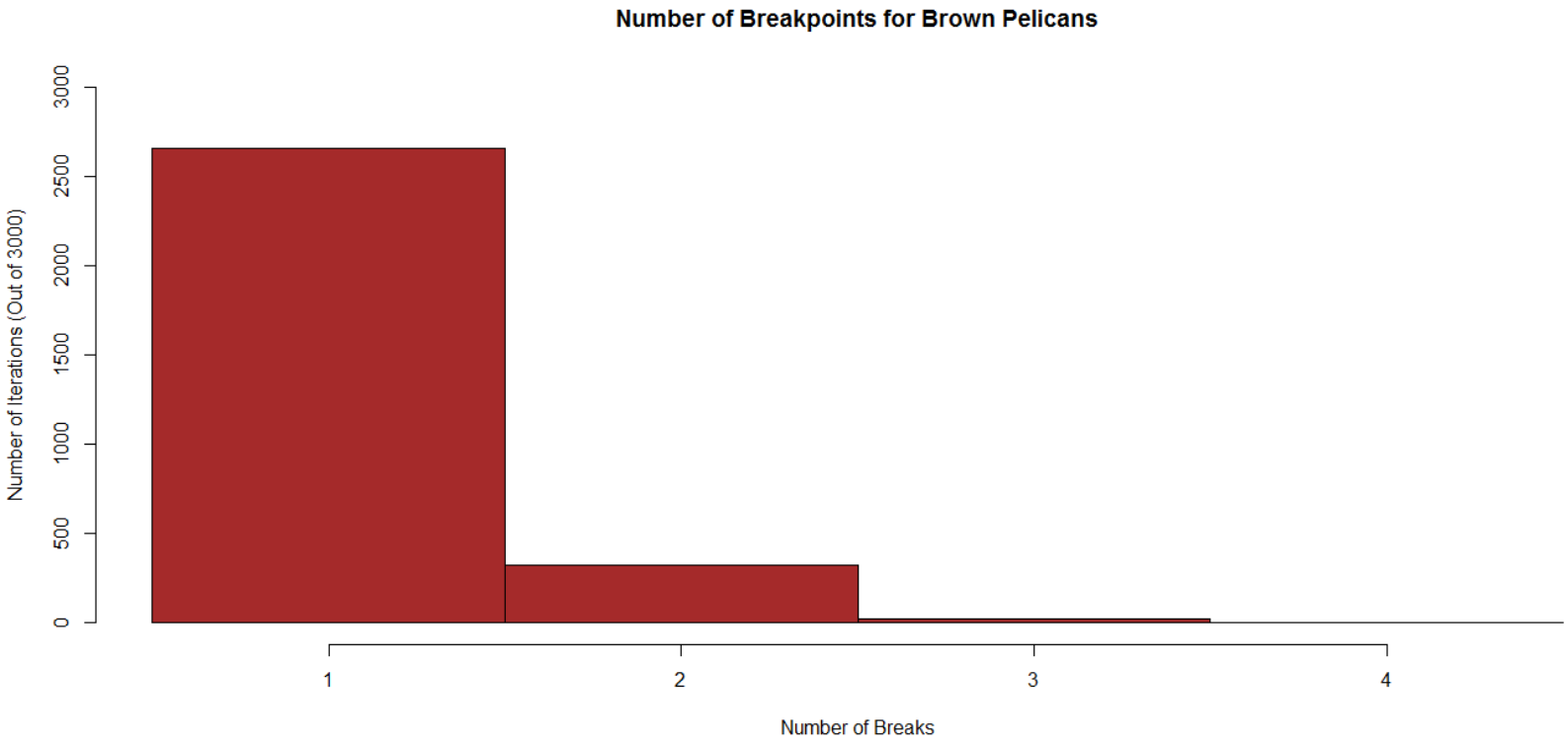


Figure 11: Distribution of breakpoint number in Pacific brown pelican (*Pelecanus occidentalis*) population data from 1939 to 2017.

Pacific Brown Pelican Example

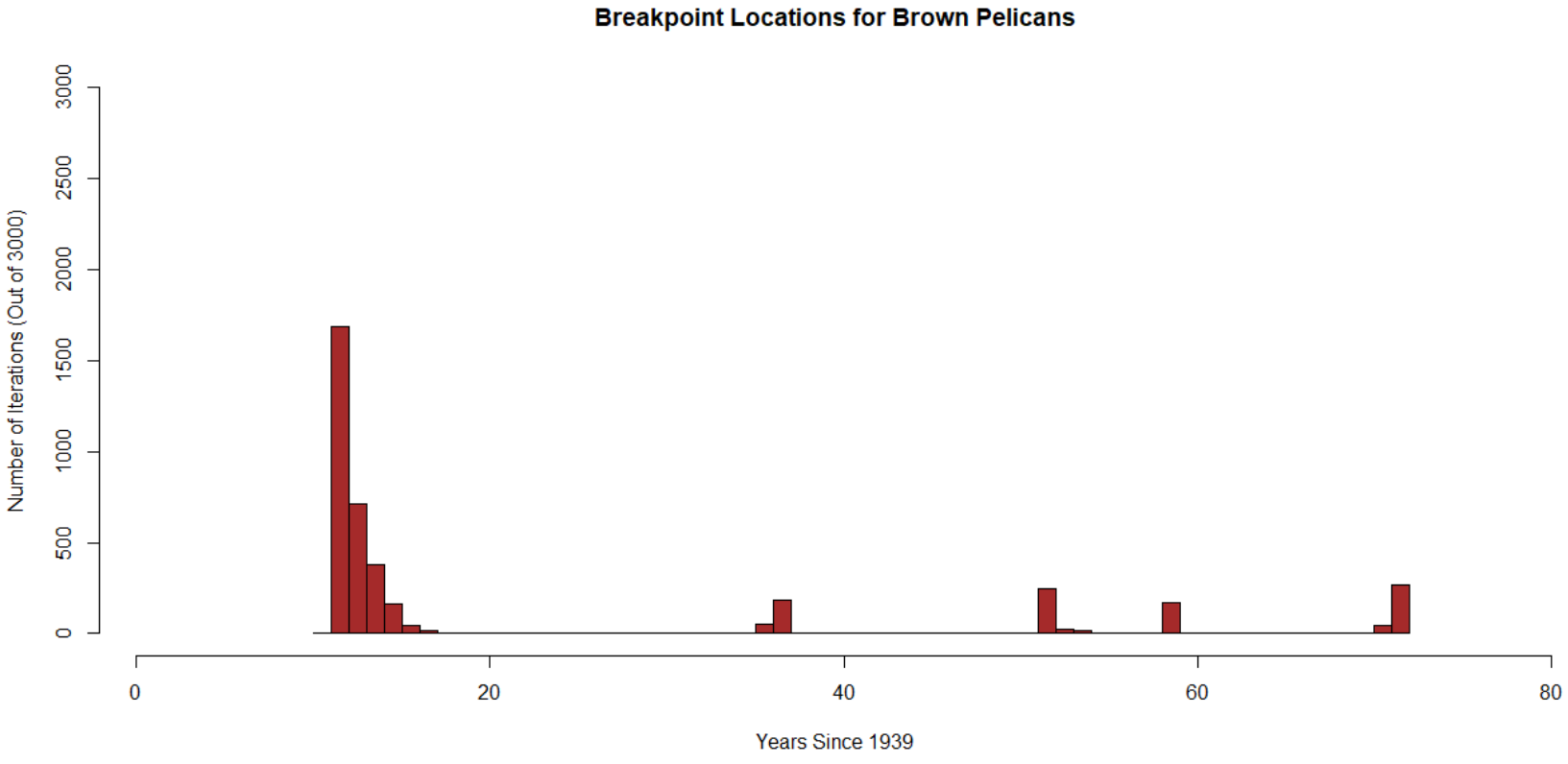


Figure 12: Distribution of breakpoint locations in Pacific brown pelican (*Pelecanus occidentalis*) population data from 1939 to 2017.

Pacific Brown Pelican Example

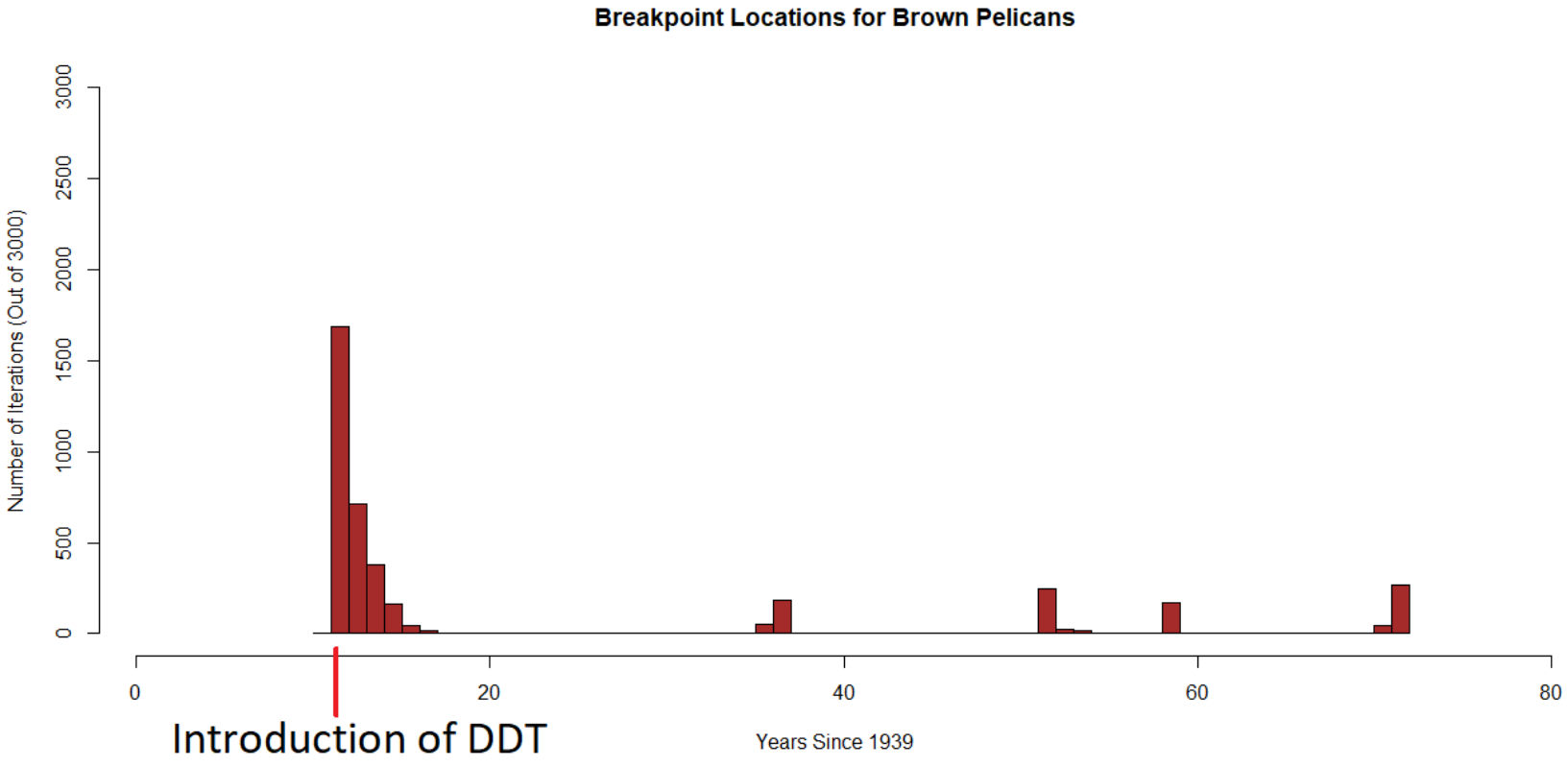


Figure 13: Distribution of breakpoint locations in Pacific brown pelican (*Pelecanus occidentalis*) population data from 1939 to 2017.

Pacific Brown Pelican Example

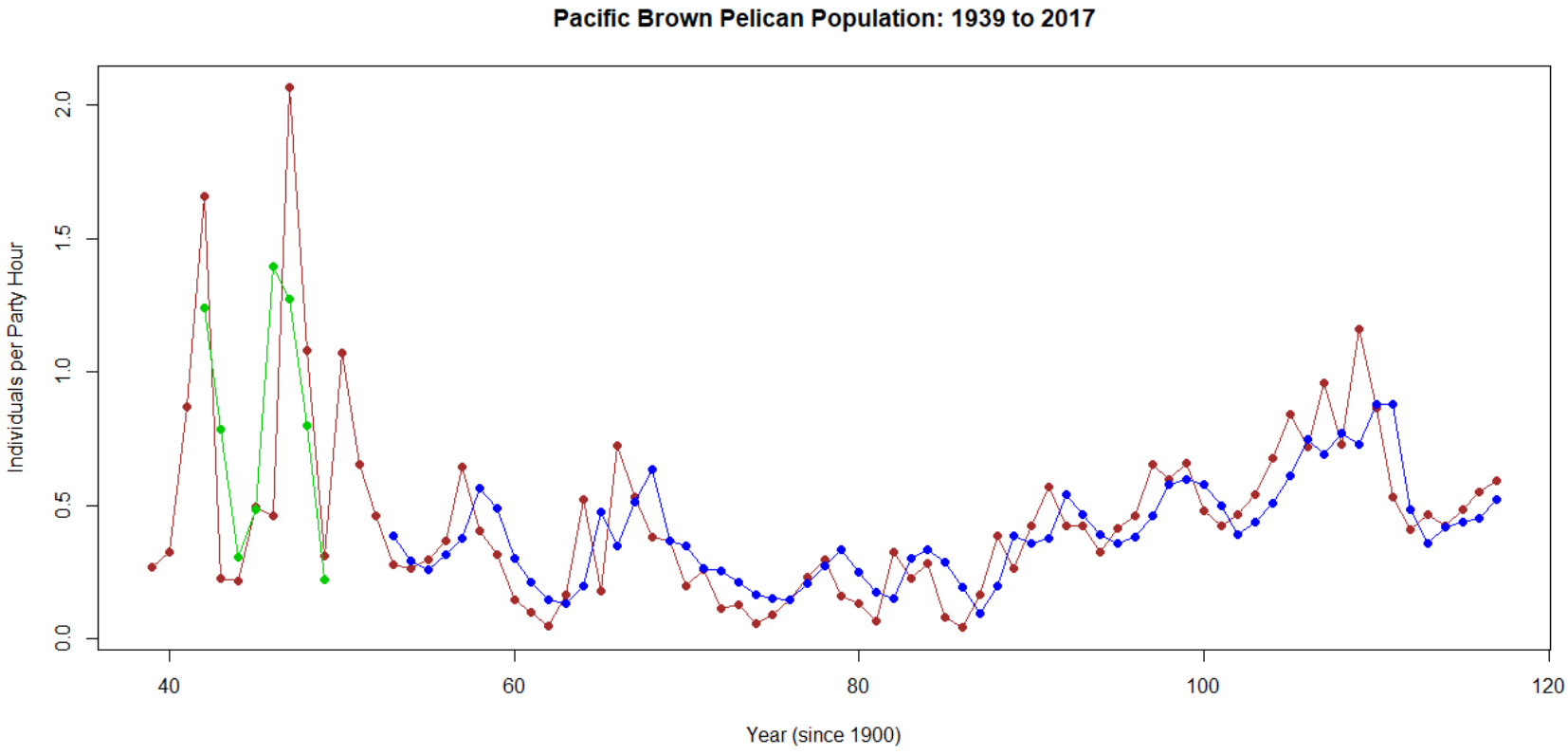
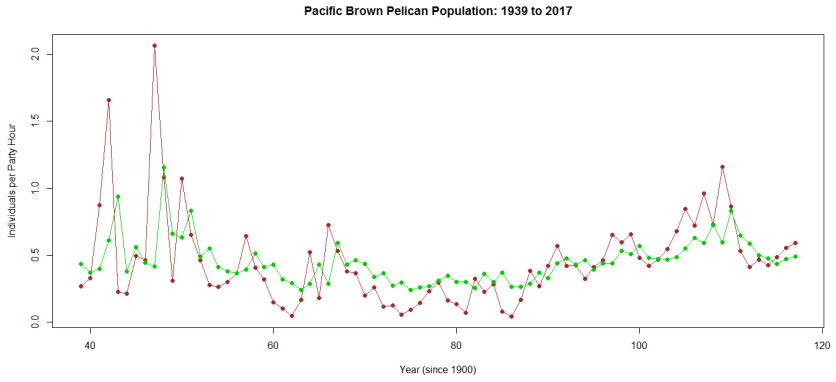
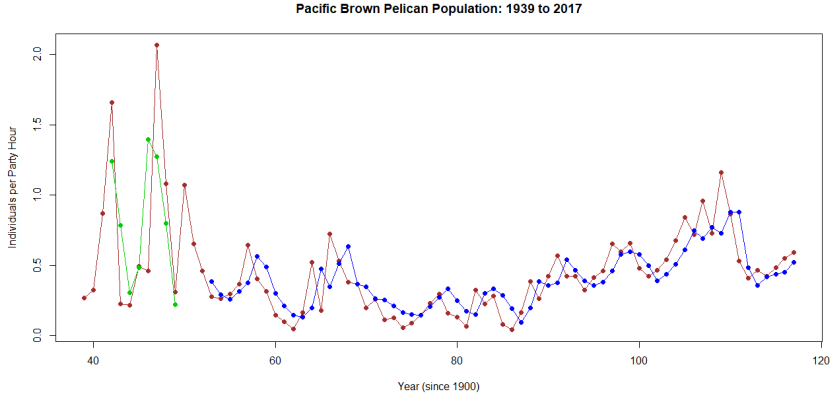


Figure 14: Green points are the mean posterior fitted values for the first subsection and blue points are the fitted values for the second subsection.

Pacific Brown Pelican Example





Model	MSE	BIC
Single AR(3) Model	0.08845281	54.75419
Posterior Means from BAAR	0.04910609	7.138885
Difference between Models	0.03934672	47.6153

Further Work

- Apply this process to multiple types of data, including non-time series data
  - economic measures, growth rates, biological data, etc.
- Improve process by which breaks are chosen

References

Bai, J. and Perron, P., (1998). *Estimating and testing linear models with multiple structural changes*. Econometrica, pp.47-78.

Bai, J. and Perron, P., (2003). *Computation and analysis of multiple structural change models*. Journal of applied econometrics, 18(1), pp.1-22.

DiMatteo, I., Genovese, C.R. and Kass, R.E., (2001). *Bayesian curve—fitting with free—knot splines*. Biometrika, 88(4), pp.1055-1071.

Kass, R.E. and Wasserman, L., (1995). *A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion*. Journal of the american statistical association, 90(431), pp.928-934.

Pesaran, M.H. and Timmermann, A., 2002. Market timing and return prediction under model instability. Journal of Empirical Finance, 9(5), pp.495-510.

Seidel, D.J. and Lanzante, J.R., (2004). *An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes*. Journal of Geophysical Research: Atmospheres, 109(D14).

Zeileis, A., Leisch, F., Hansen, B., Hornik, K., Kleiber, C. and Zeileis, M.A., (2007). *The strucchange Package*. R manual.

Thank You

Questions?