Developing a Bayesian Procedure to Detect Breakpoints in Time Series: Progress Report II

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

Lafayette College, Unity College, Marist College

Wednesday, July 12, 2018

NSF Grant #1560222

Recap: Back to Climatology

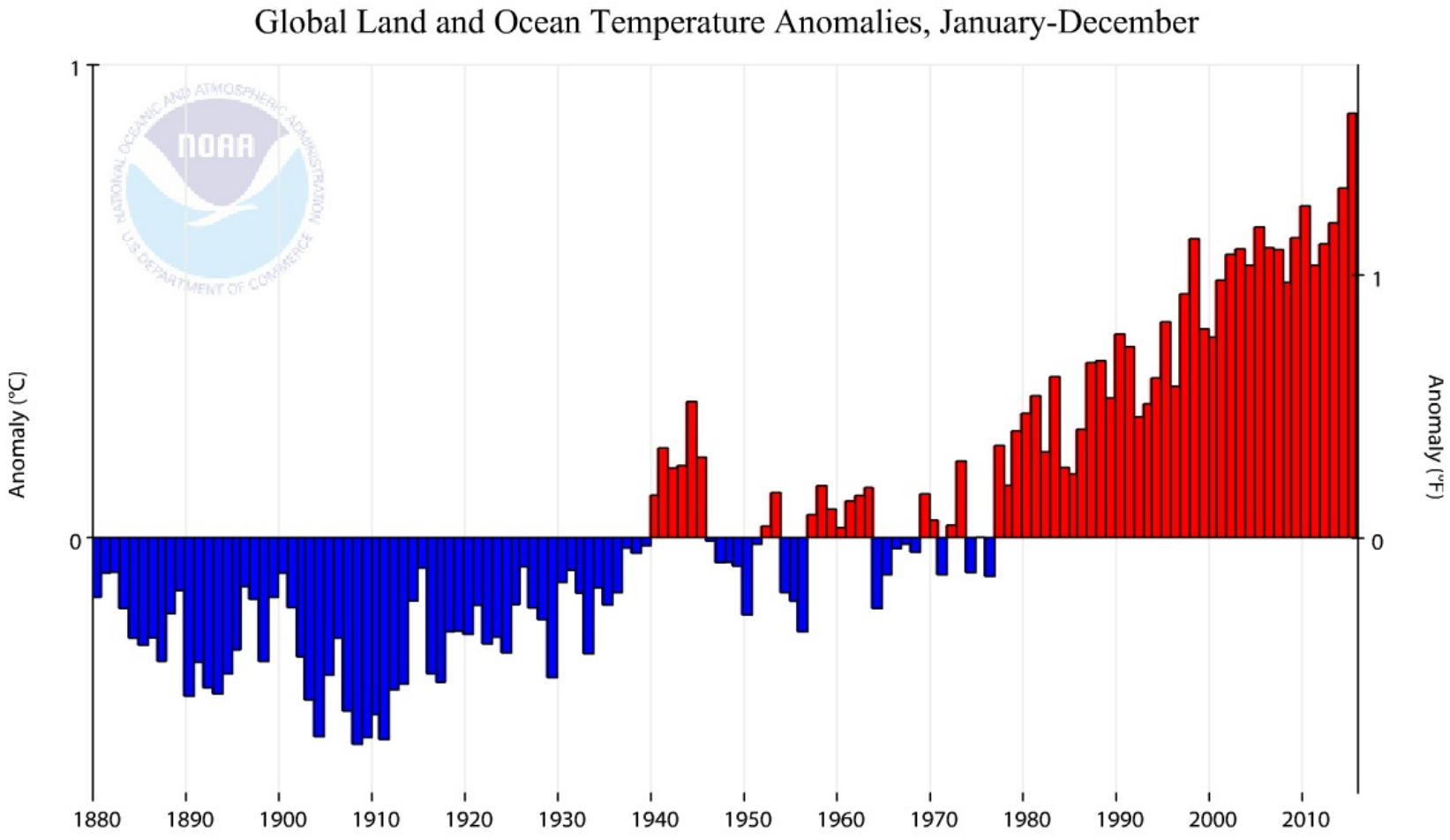

Figure 1: Temperature anomaly data.

Recap: Back to Climatology

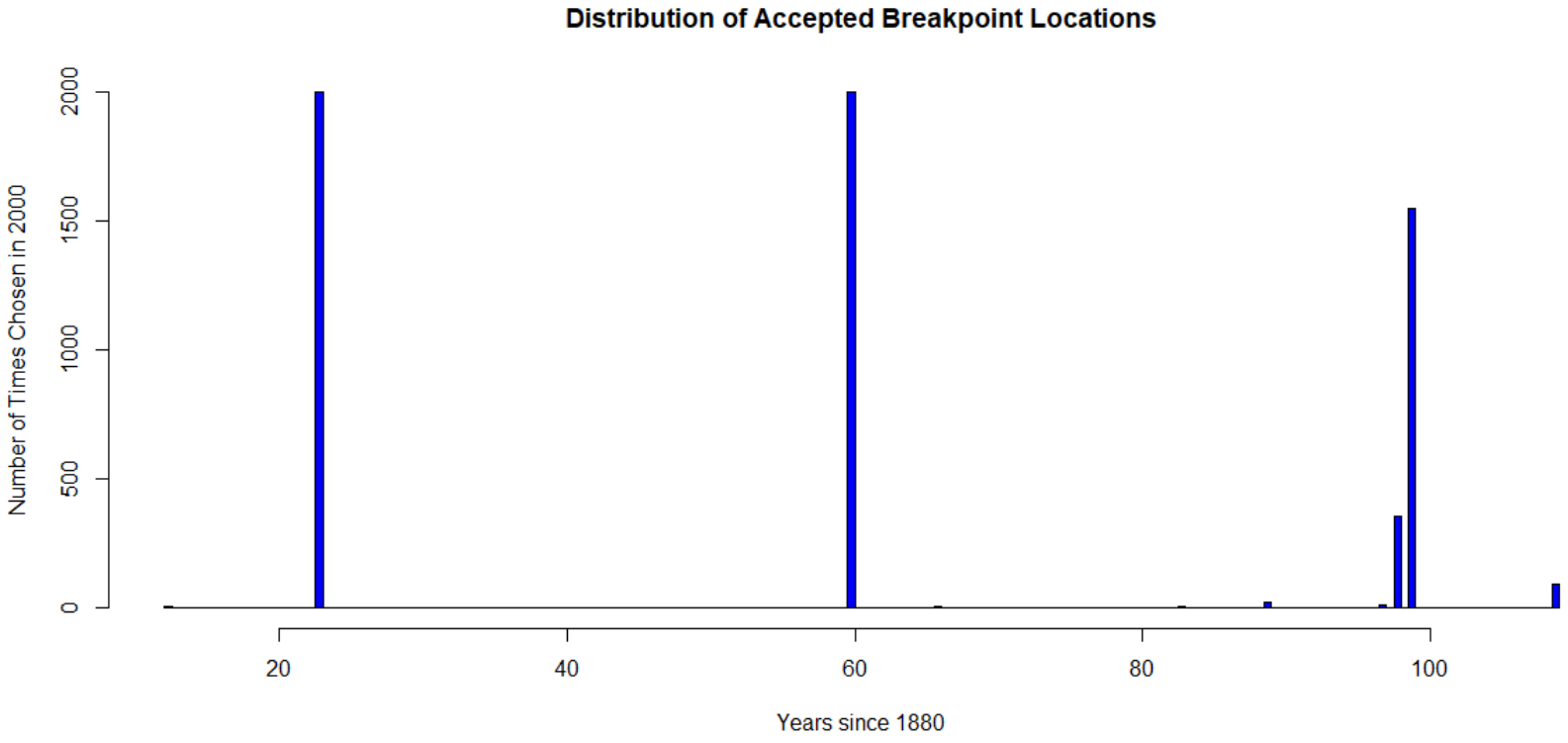

Figure 2: Distribution of the breakpoint location of the temperature anomaly data as determined by BAR (Baysian Adaptive Regression) with random addition, random subtraction, and move.

The Jump Function

Jump:

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

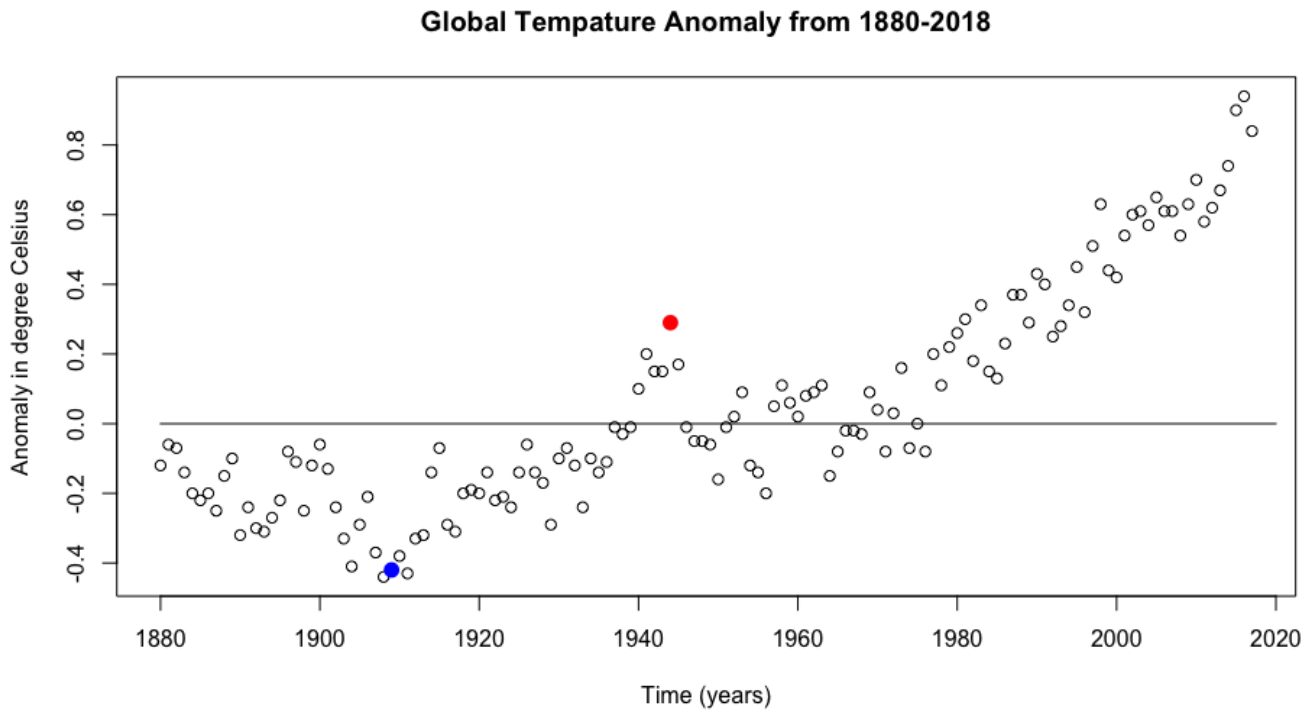

Figure 3: Red is an initial breakpoint, and blue is a proposed breakpoint.

The Jiggle Function

Jiggle:

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

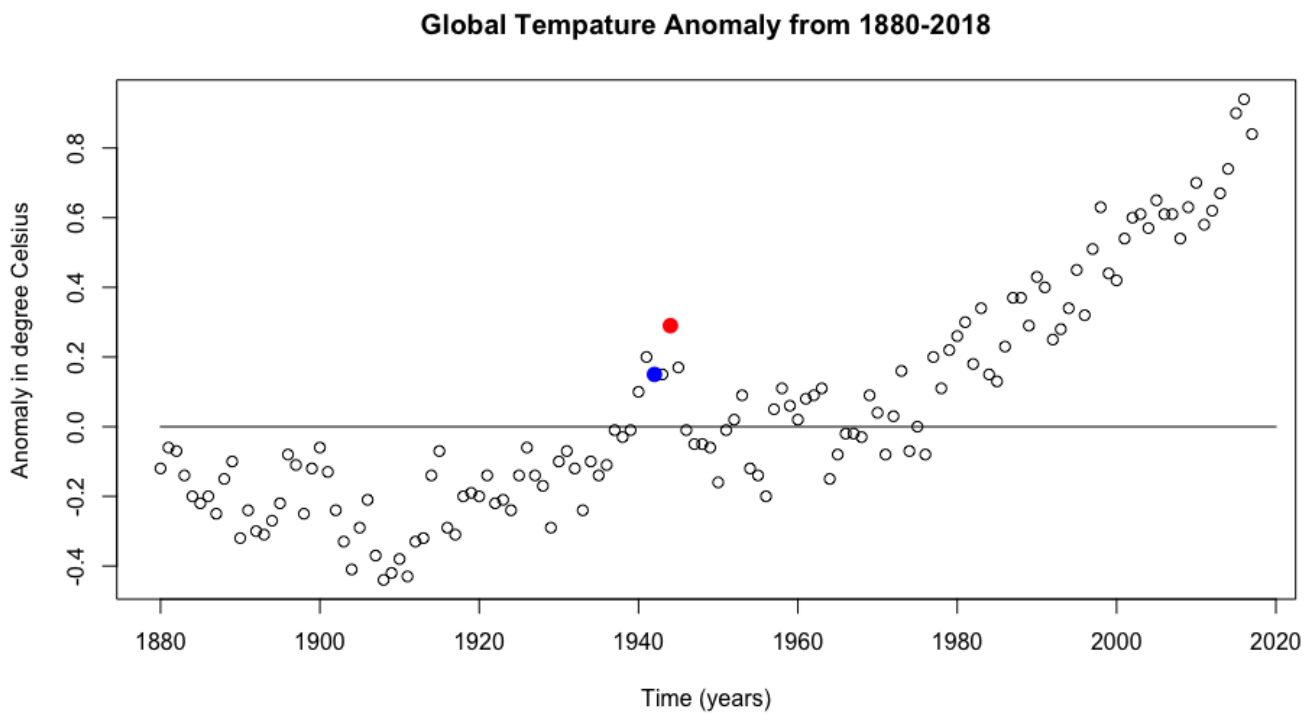

Figure 4: Red is an initial breakpoint, and blue is a proposed breakpoint.

First Simulations: Testing Variations of Move

Goal of the Simulations: To see how the jiggle function compares to the jump only function.

Results: The best function tested was one that had both jiggle and jump as moving options.

Second Simulations: Variations on Addition

Random	**Interval**	**Scoring**
Picks one random place in the data set	Picks an interval weighted on length between two breakpoints and randomly places a breakpoint in the interval	Picks two random places and chooses the one that is the farthest away from its neighboring breakpoints

Second Simulations: Variations on Subtraction

Random	**Interval**	**Scoring**
Randomly picks an existing breakpoint and deletes it	Picks breakpoint to delete based on the weighted shortest combined distance of a breakpoints and deletes it	Picks two random breakpoints and deletes the one that is the closest to its neighboring breakpoints

Second Simulation

Function Name	Addition	Subtraction
BAR0	random	random
BAR1	interval	random
BAR2	scoring	random
BAR3	random	interval
BAR4	random	scoring
BAR5	interval	interval
BAR6	interval	scoring
BAR7	scoring	scoring
BAR8	scoring	interval

The different combinations of addition and subtraction methods for the possible BAR function.

Second Simulation

Every BAR function is run on each dataset with 3000 iterations 9 different times.

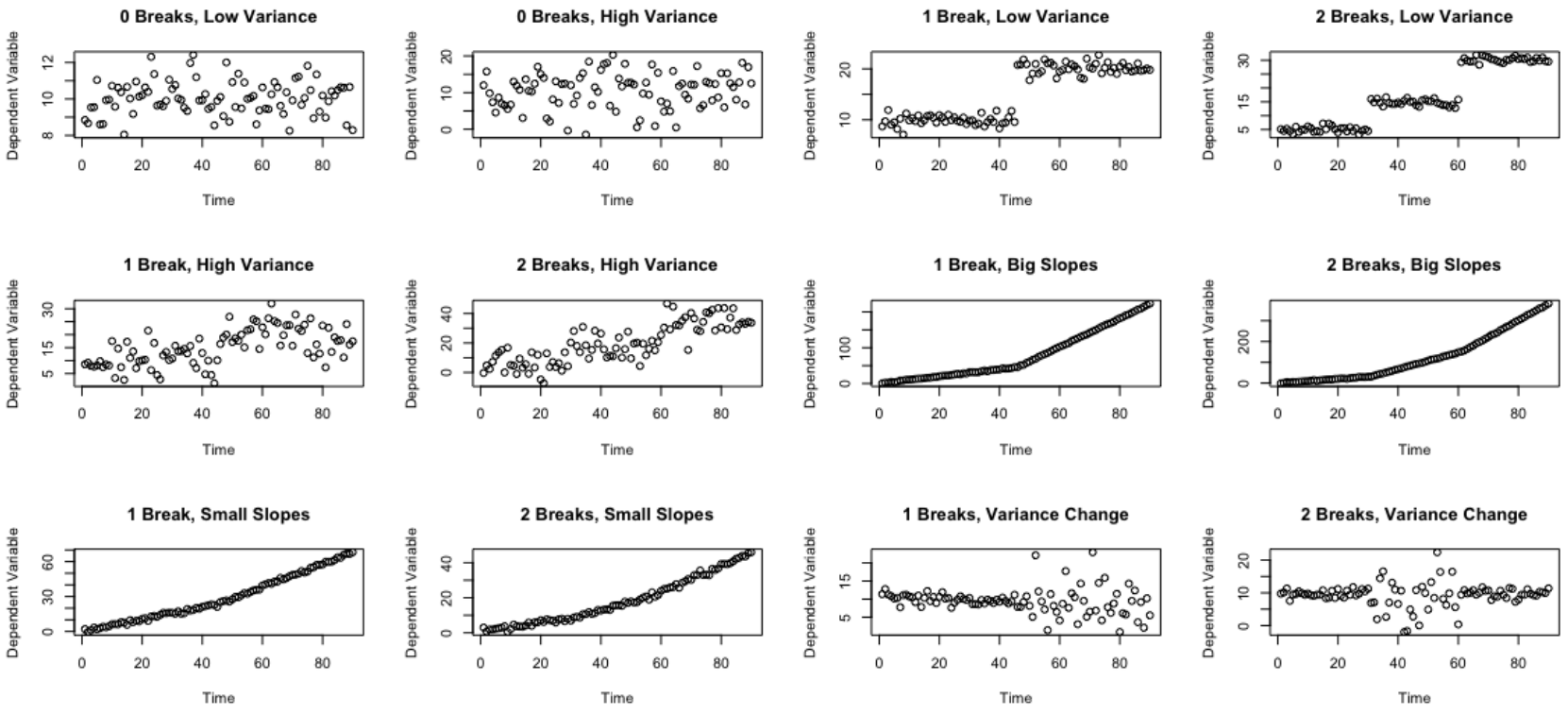

Figure 5: Data for the second simulation.

Second Simulation Results

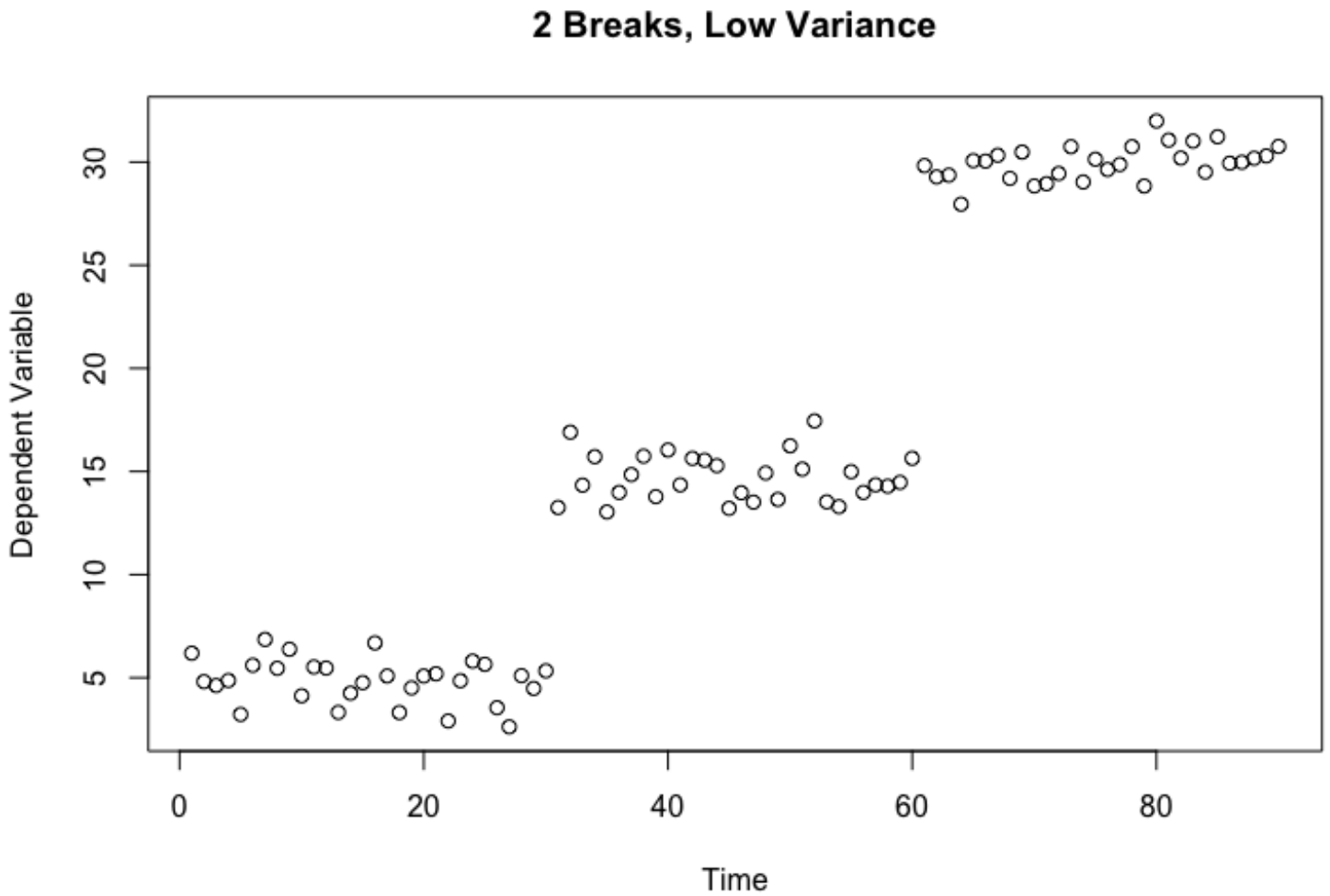

Figure 6: Data with 2 significant breaks.

Second Simulation Results

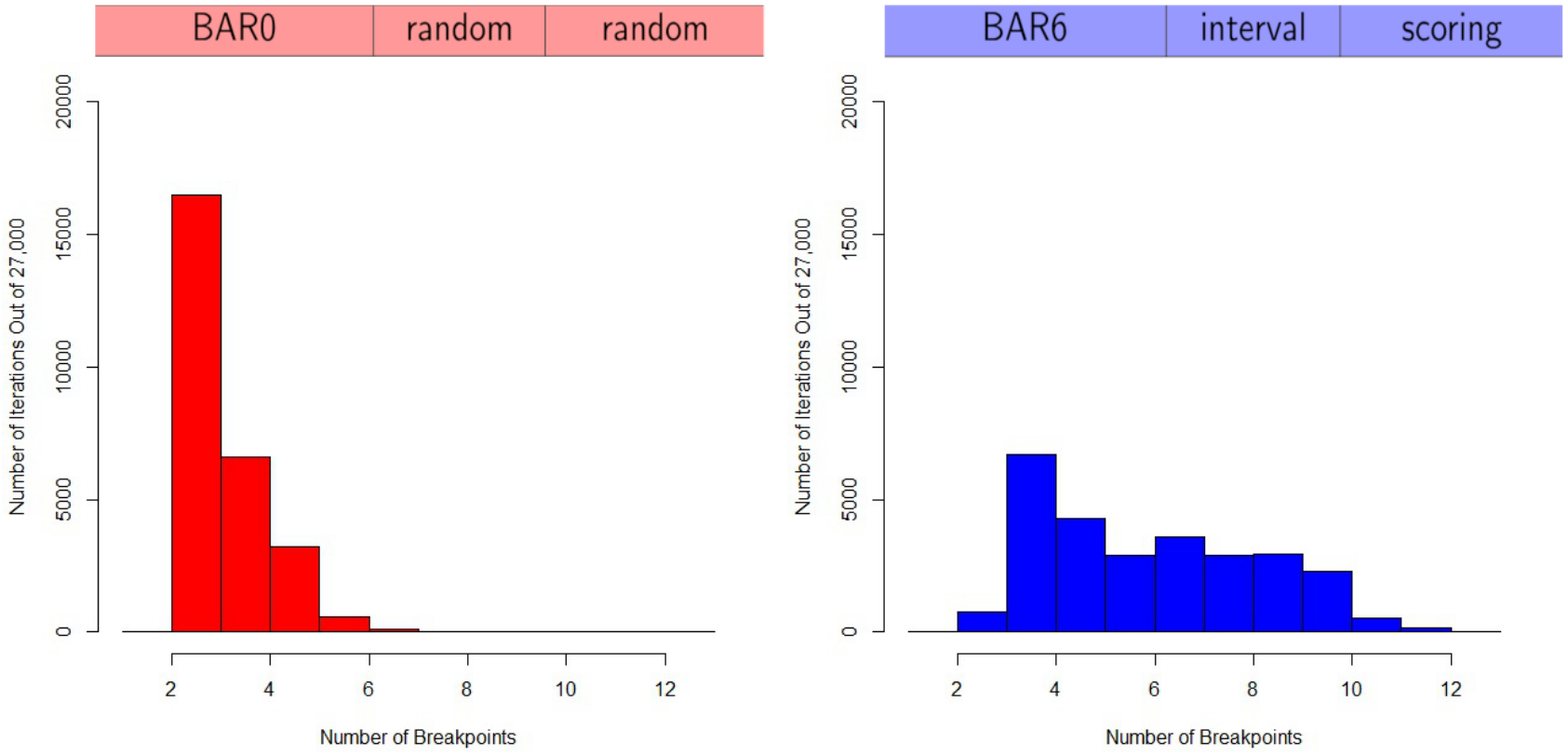

Figure 7: Distributions of number of breakpoints.

Second Simulation Results

Result Summary:

- **BAR3**, **BAR5**, and **BAR8** were unable to identify breakpoints in messy data.
- **BAR7** struggles with messy data and tends to over add compared to BAR0, BAR1, and BAR4.
- **BAR4** is okay but has no stand out features

BAR0, **BAR1**, and **BAR2** handle the more difficult data slightly better.

Function	Addition	Subtraction
BAR0	random	random
BAR1	interval	random
BAR2	scoring	random
BAR3	random	interval
BAR4	random	scoring
BAR5	interval	interval
BAR6	interval	scoring
BAR7	scoring	scoring
BAR8	scoring	interval

Third Simulation: Long Runs

BAR0, **BAR1**, and **BAR2** were chosen to run for a much longer period of times to see if any of them stood out.

Simulation Three:

- Tested on three different datasets:
  - 2 Breaks, Low Variance
  - 2 Breaks, Variance Change
  - Temperature Anomalies
- Each function will run for 100,000 iterations

Third Simulation: Long Runs

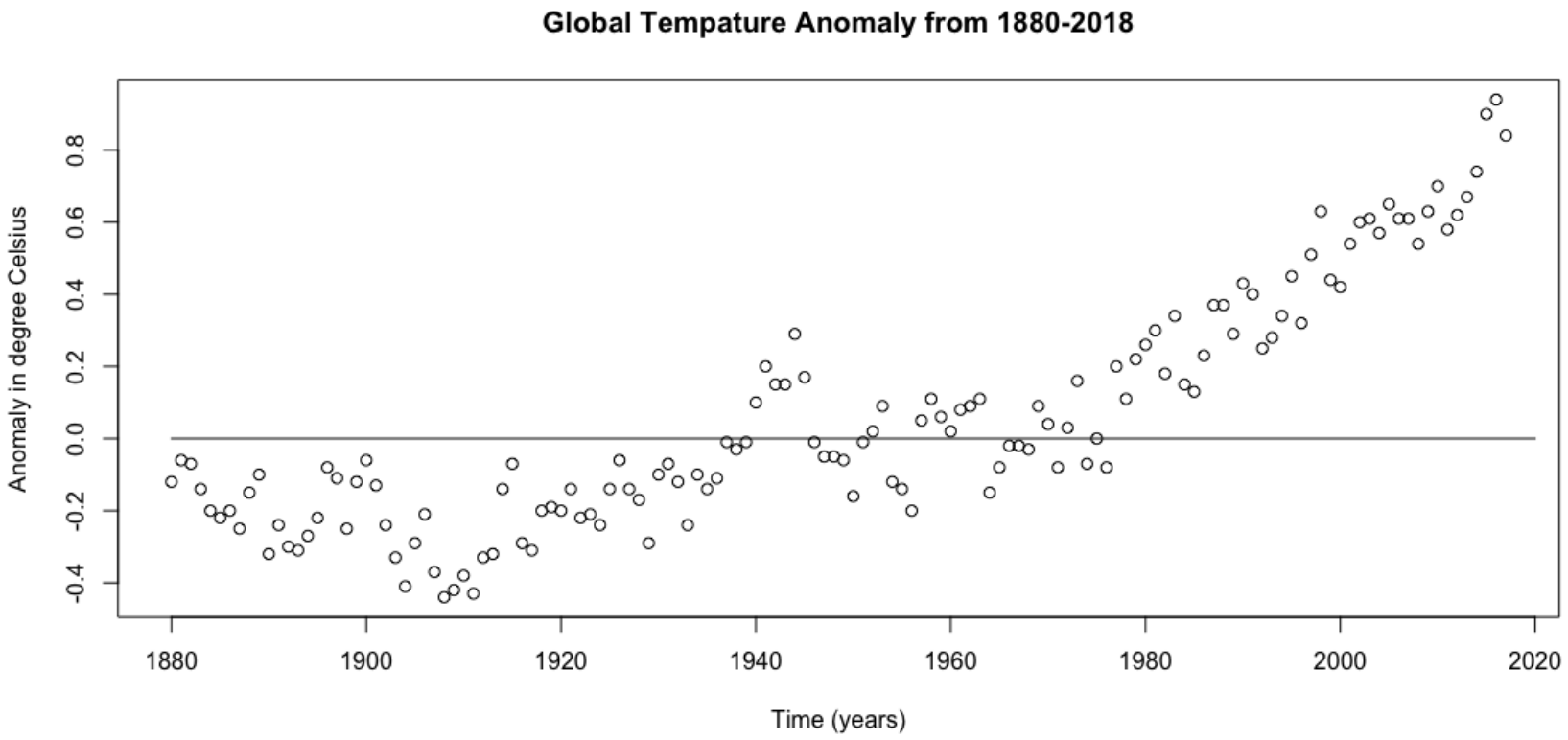

Figure 8: Temperature anomaly dataset used in the third simulation

Third Simulation: Results

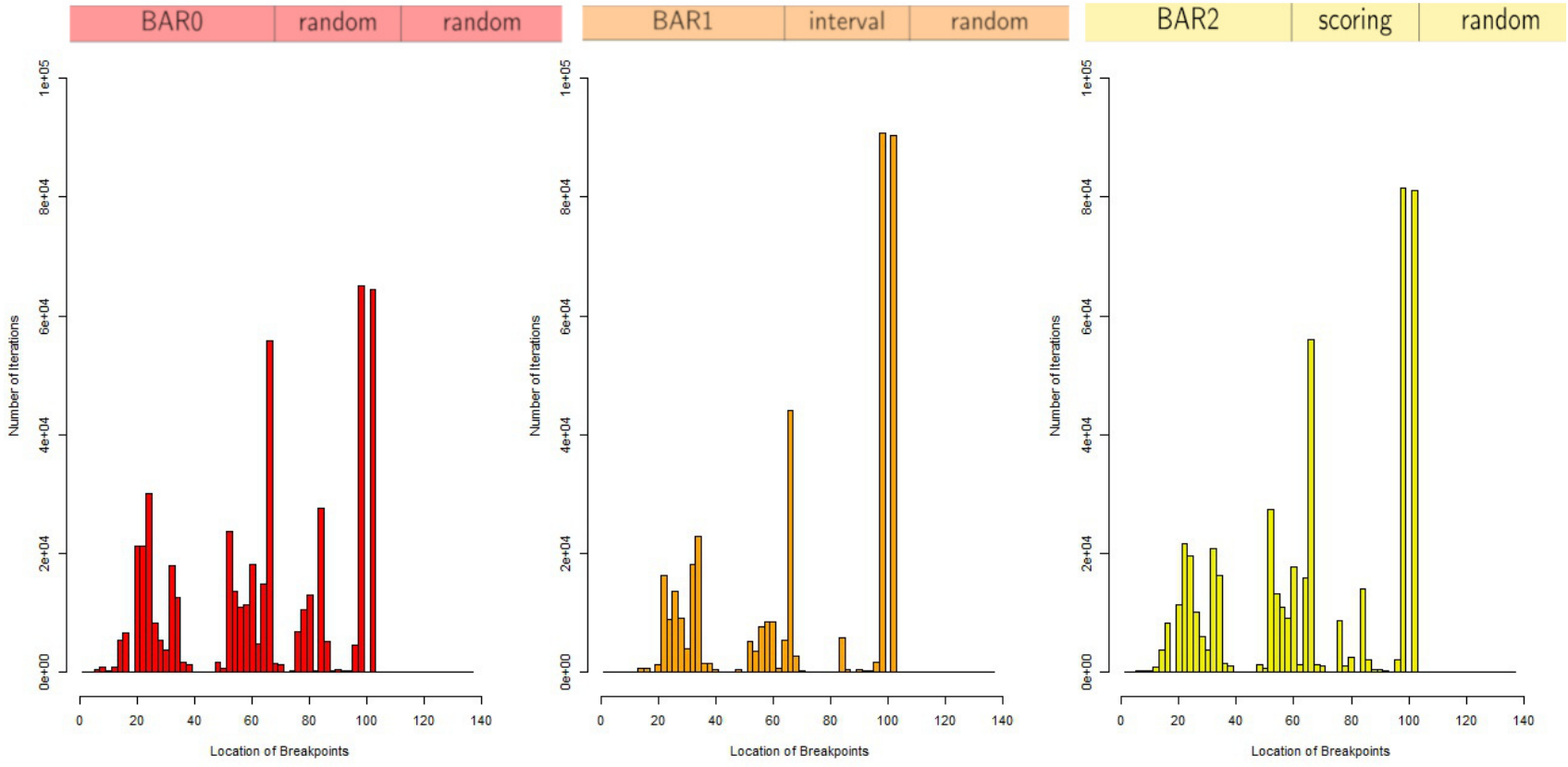

Figure 9: Distributions of breakpoint locations for long runs looking at temperature anomaly data.

Summary

Function Name	Addition	Subtraction
BAR0	random	random
BAR1	interval	random
BAR2	scoring	random
BAR3	random	interval
BAR4	random	scoring
BAR5	interval	interval
BAR6	interval	scoring
BAR7	scoring	scoring
BAR8	scoring	interval

Next!

- Choose between between **BAR0** and **BAR2**
- Testing different starting conditions
- Apply to Autoregressive (AR) models
- Apply to real life data
- WRITE! (Cause we're running out of time...)