SAND2023-05023C

Sandia National Laboratories

# Elastic Functional Changepoint Detection

## with Application to MERRA-2

J. Derek Tucker and Drew Yarger

SIAM GS 23 6/20/23

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

ring Solutions of Sandia LLC, a wholly
stration under contract DE-NA0003525.

# Outline

Introduction/Motivation

Functional Data Analysis

- Elastic Metric

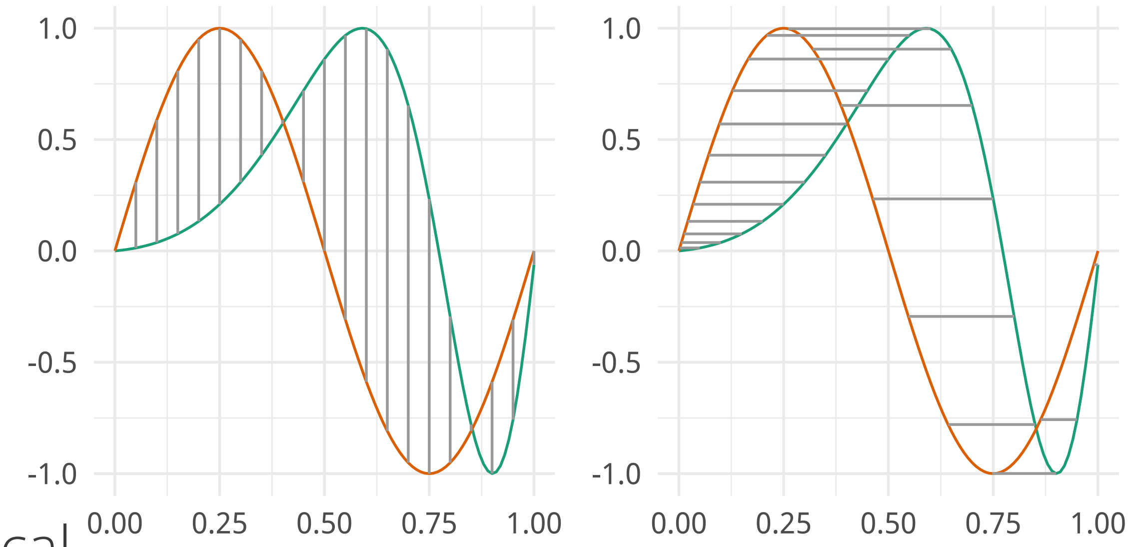
Changepoint Analysis

Results

Conclusion

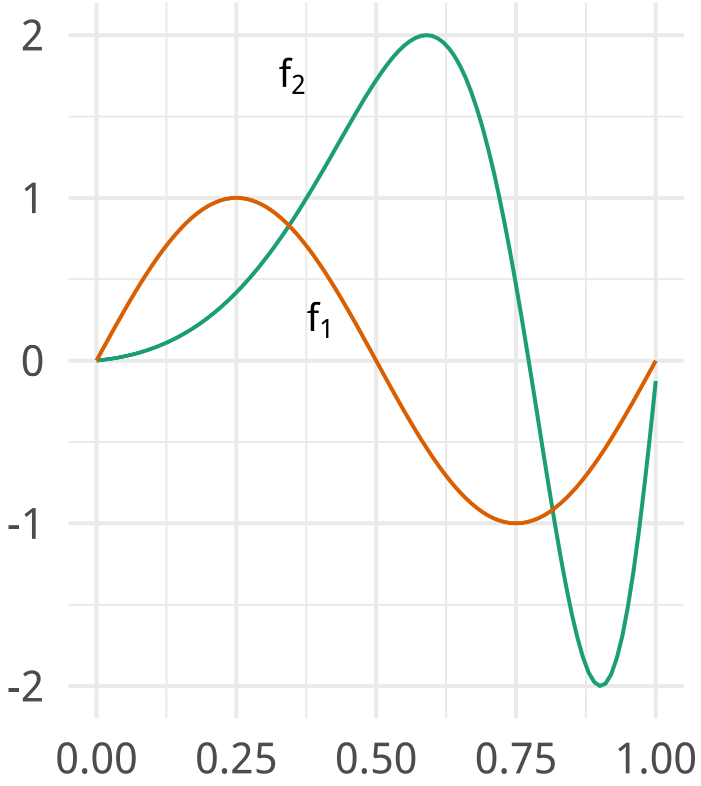
# Introduction

- Question: How can we model functions
  - Can we use the functions to classify diseases?
  - Can we use them as predictors in a regression model?
  - Can we find changepoints?
- It is the same goal (question) of any area of statistical study

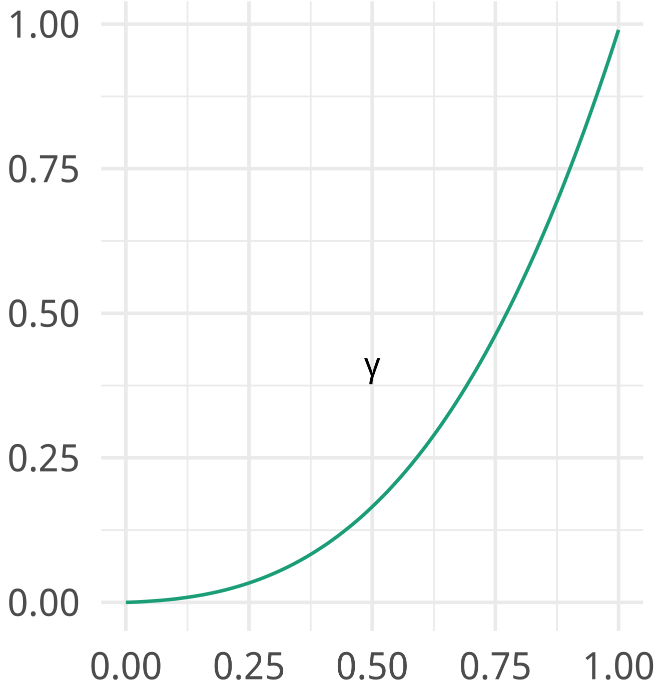



- One problem occurs when performing these types of analysis is that functional data can contain variability in time (x-direction) and amplitude (y-direction)
- How do we characterize and utilize this variability in the models that are constructed from functional data?

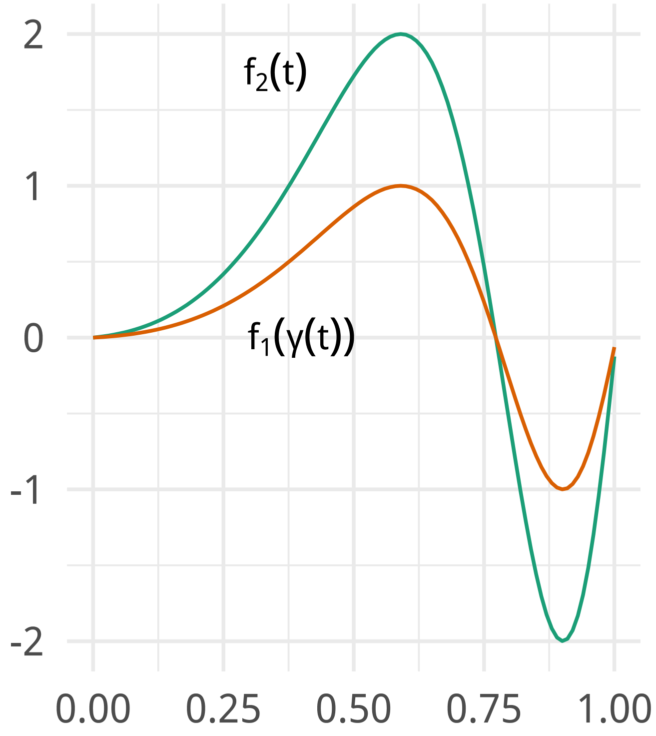
# Components of Function Variability




Warping Functions: "*x*-variability"




Aligned Functions: "*y*-variability"



# Functional Data Analysis

Let $f$ be a real valued-function with the domain $[0,1]$, can be extended to any domain

- Only functions that are absolutely continuous on $[0,1]$ will be considered

Let $\Gamma$ be the group of all warping functions

$$\Gamma = \{\gamma : [0,1] \to [0,1] \,|\, \gamma(0) = 0, \gamma(1) = 1, \ \gamma \text{ is a diffeo}\}$$

It acts on the function space by composition

$$(f, \gamma) = f \circ \gamma$$

It is common to use the following objective function for alignment

$$\min_{\gamma \in \Gamma} \|f_1 \circ \gamma - f_2\|$$

Note: It is not a distance function since it is not symmetric.

# Elastic Distance (Fisher-Rao)

Define the Square Root Velocity Function

$$q : [0,1] \to \mathbb{R}^1,\ q(t) = sign(\dot{f}(t))\sqrt{|\dot{f}(t)|}$$

Fisher Rao Distance is $\mathbb{L}^2$ in SRVF space

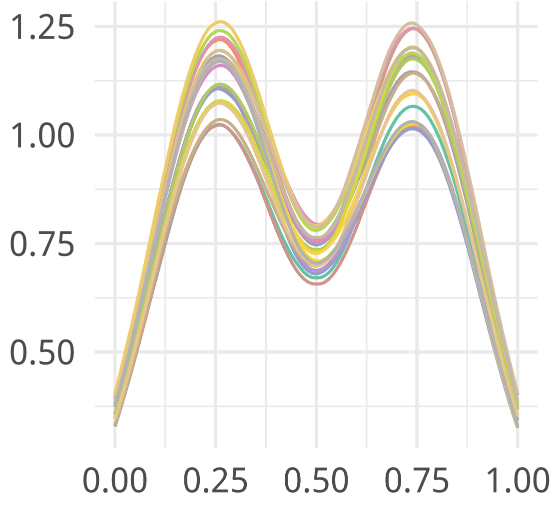
$$d_a(f_1, f_2) = \inf_{\gamma} \|(q_1 \circ \gamma)\sqrt{\dot{\gamma}} - q_2\|$$

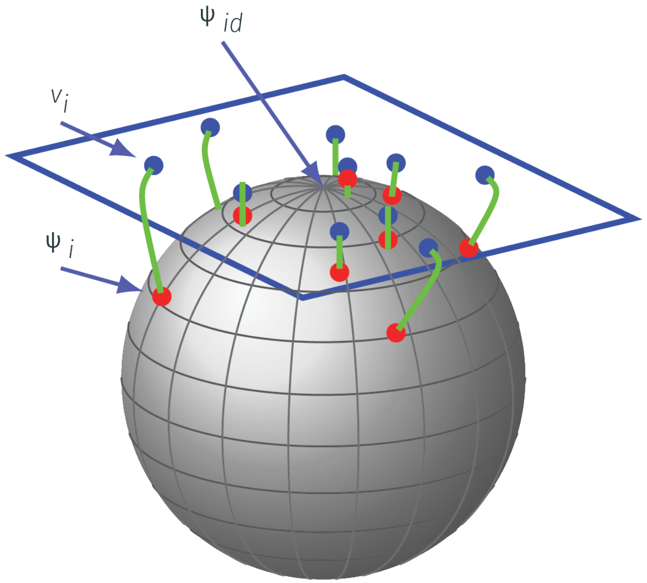
Distance is a proper distance

- symmetric, isometric, triangle inequality

Can compute distance on warping functions (how much alignment)

$$d_p(\gamma) = \arccos\left(\int_0^1 \sqrt{\dot{\gamma}}\, dt\right)$$



# Functional Changepoint

The standard changepoint problem assumes the data is from the following model:

$$f_i = \mu + \delta 1(i > k*) + \epsilon_i$$

Where $k*$ labels the time of the unknown mean change

This model ignores phase variability and proper metrics, we can utilize the SRVF framework

$$(q_i, \gamma_i) = \mu_q + \delta 1(i > k*) + \epsilon_i$$

# Functional Changepoint

We then wish to test the hypothesis

$$H_0 : \delta = 0 \quad \text{vs} \quad H_A : \delta \neq 0$$

## Amplitude Changepoint

We can utilize the amplitude distance and Karcher Mean computation

$$\mu_q^k = \arg\min_q \sum_{i=1}^{k} d_a(q, q_i)^2$$

Utilize scaled functional cumulative sum test statistic

$$S_{n,k} = \frac{1}{\sqrt{n}} \left( k\mu_q^k - k\mu_q^n \right)$$

# Amplitude Changepoint

Find the changepoint by using max-point structural break detector

$$T_n = \max_{1 \le k \le n} ||S_{n,k}||^2$$

to test the hypothesis.

Under this model $T_n$ is distributed according to

$$T_n \xrightarrow{d} \sup_{0 \le x \le 1} \sum_{l=1}^{\infty} \lambda_l B_l^2(x)$$

Where $B_i$ are I.I.D. standard Brownian bridges defined on $[0,1]$

# Phase Changepoint

We can utilize the same cumulative test statistics on the tangent space of Hilbert sphere

$$S_{n,k} = \frac{1}{\sqrt{n}} \left( k\bar{v}^k - k\bar{v}^n \right)$$

where $\bar{v}$ is the mean of the shooting vectors on the tangent space of $\mathbb{S}_\infty$ at $\mu_\psi$ or the mean of the SRVF of the warping function $\gamma$

The test statistics is the same with the the same distribution

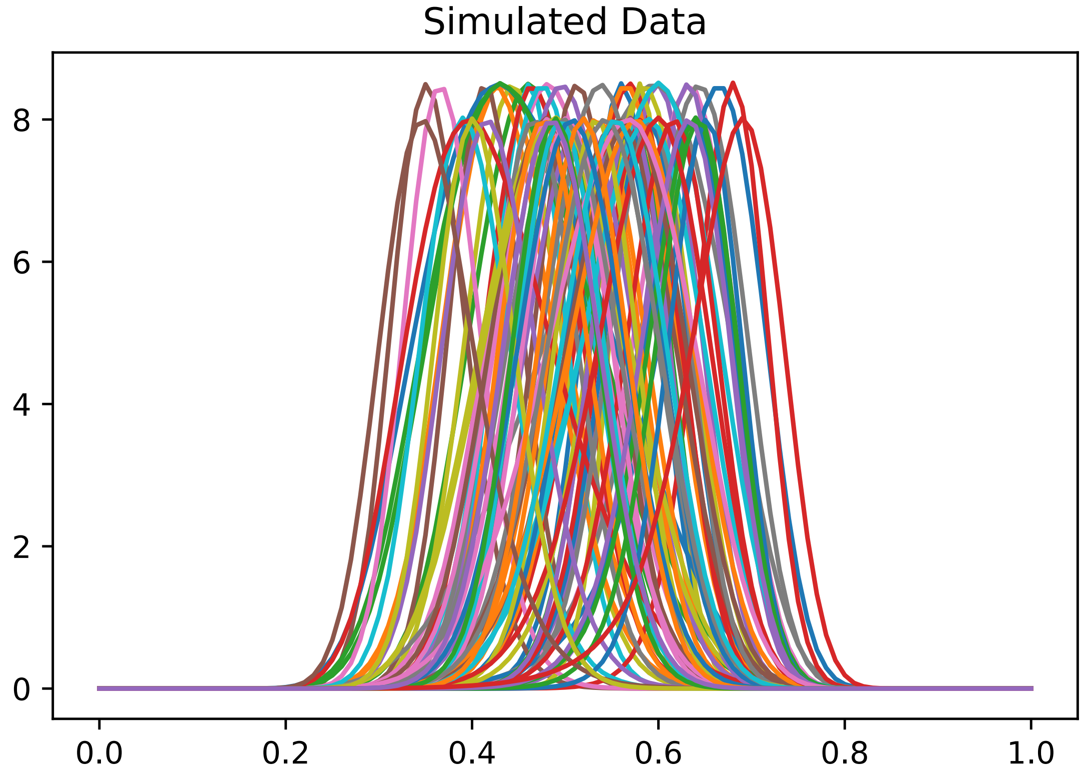
# Simulation Study - Amplitude

Data from the following model with random phase warping

$$f_i(t) = z_i e^{-(t-a_i)^2/2}$$

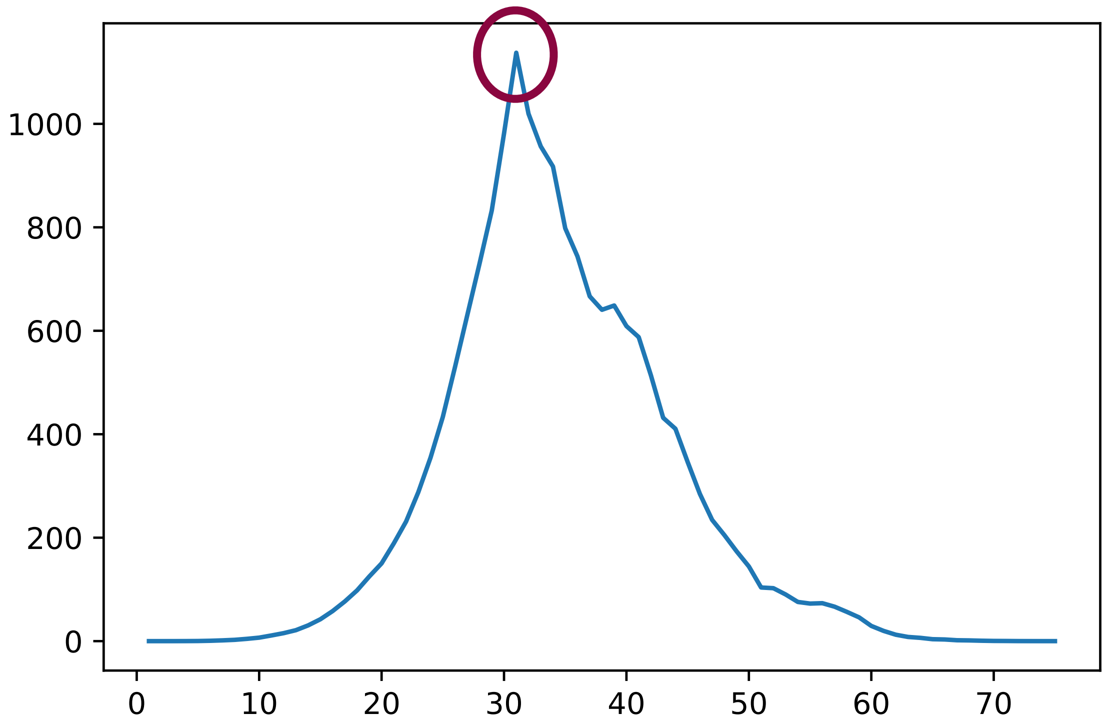
$$z_i \sim \mathcal{N}(1,(0.05)^2) \quad a_i \sim \mathcal{N}(0,(1.25)^2)$$

There is an amplitude changepoint inserted after function 30, with a total of 75 functions

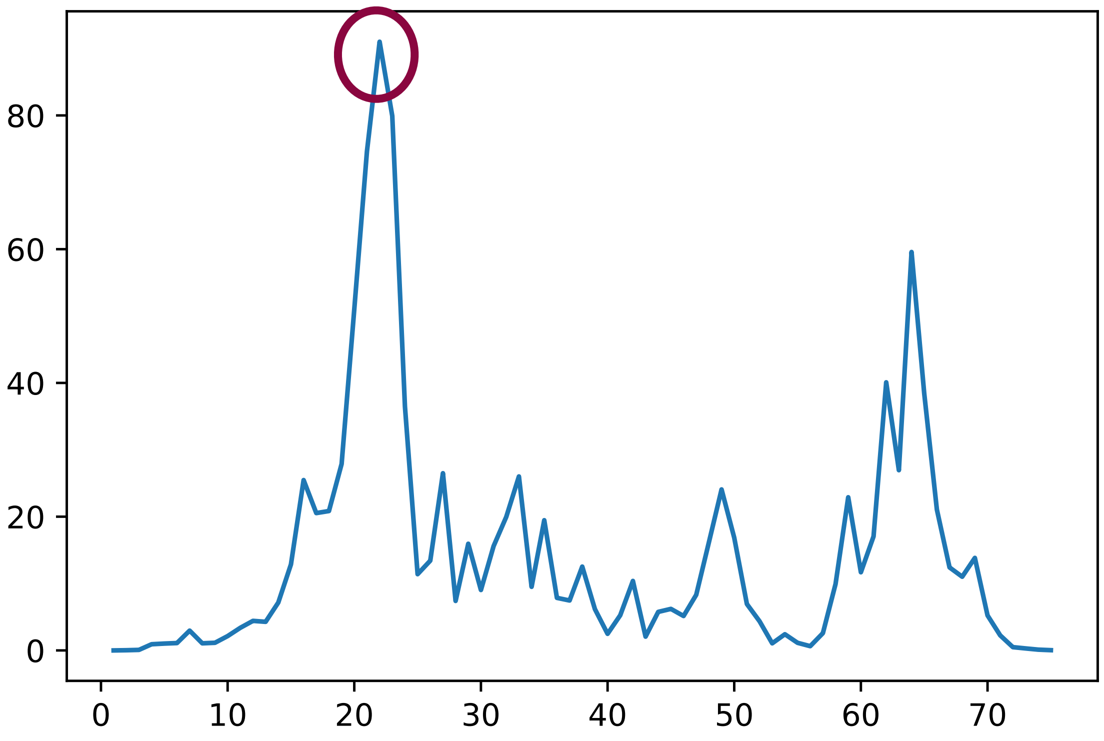


# Simulation Study - Amplitude

**Elastic Method**




**Cross-Sectional Method (Aue 2008)**

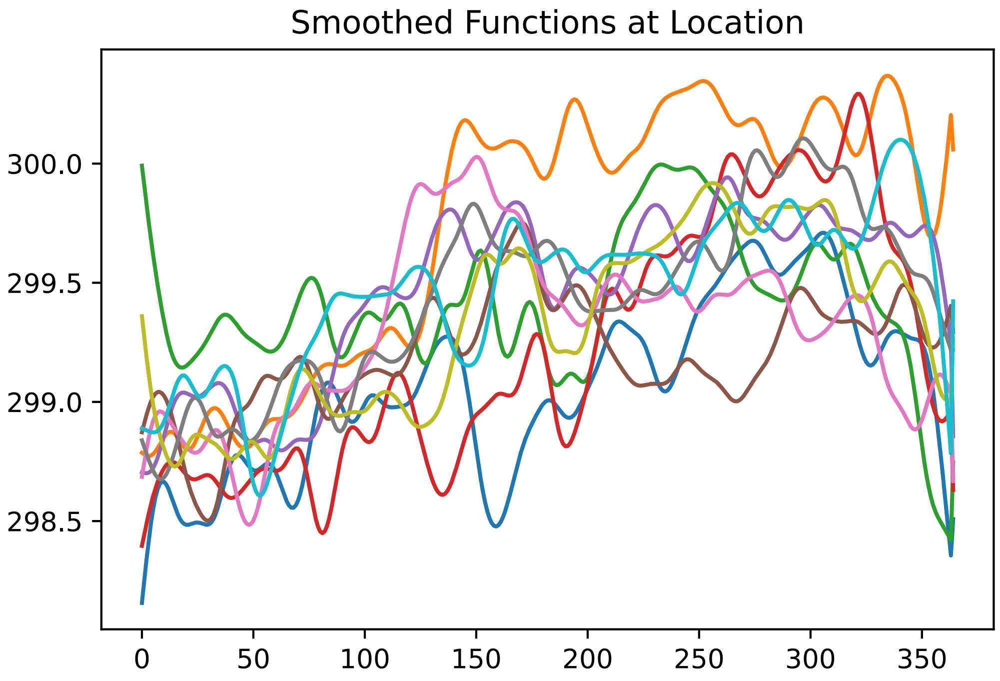


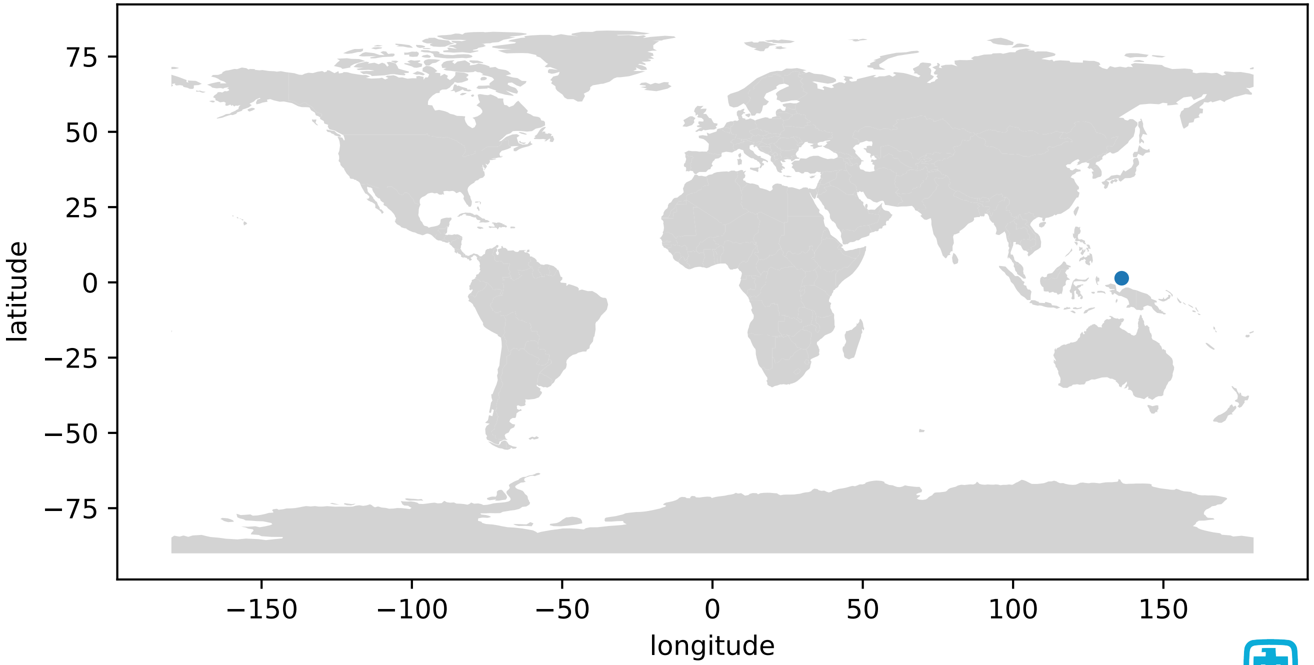

Detection found by max of $S_{n,k}$

# ERA5

Surface temperature from ERA5 from 1986-1995 at a location near Mt. Pinatubo eruption

Functions from location that is spatially averaged

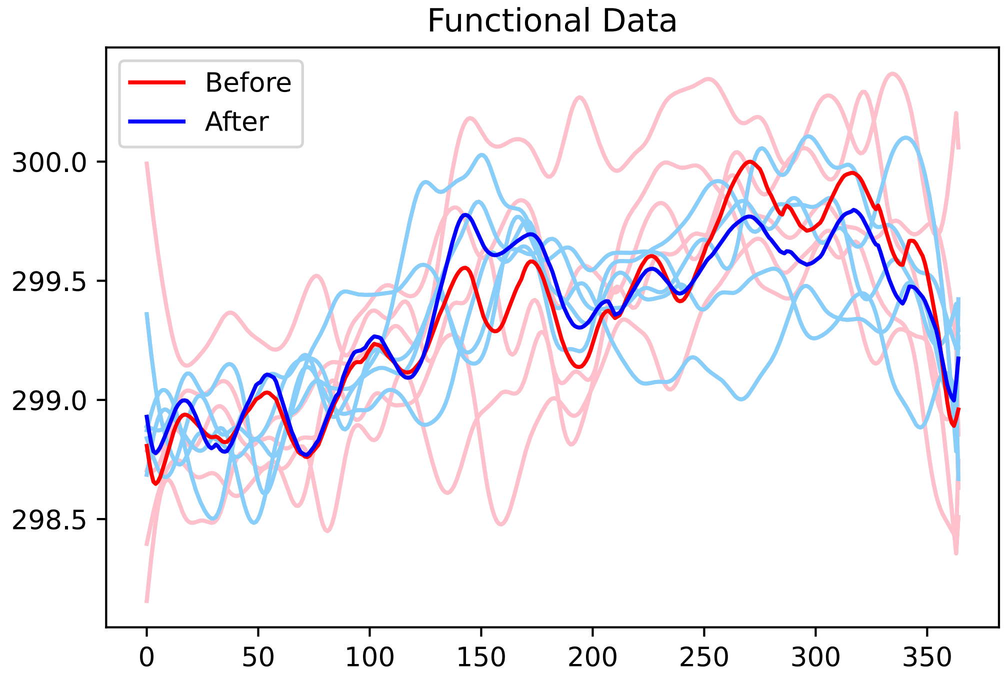


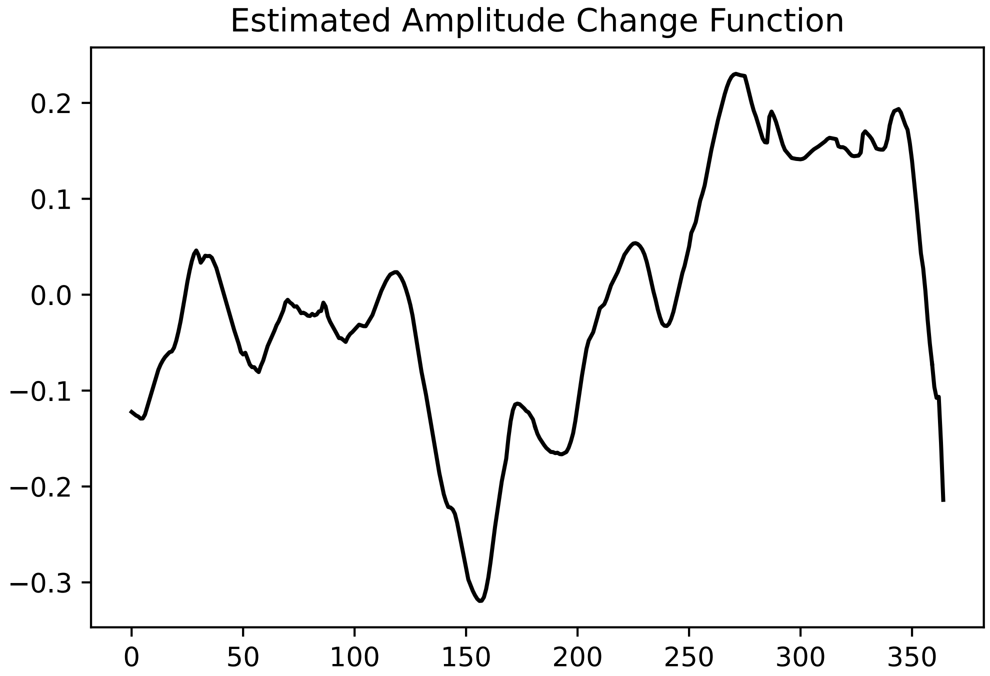



# ERA5

Only found a amplitude changepoint at 1991

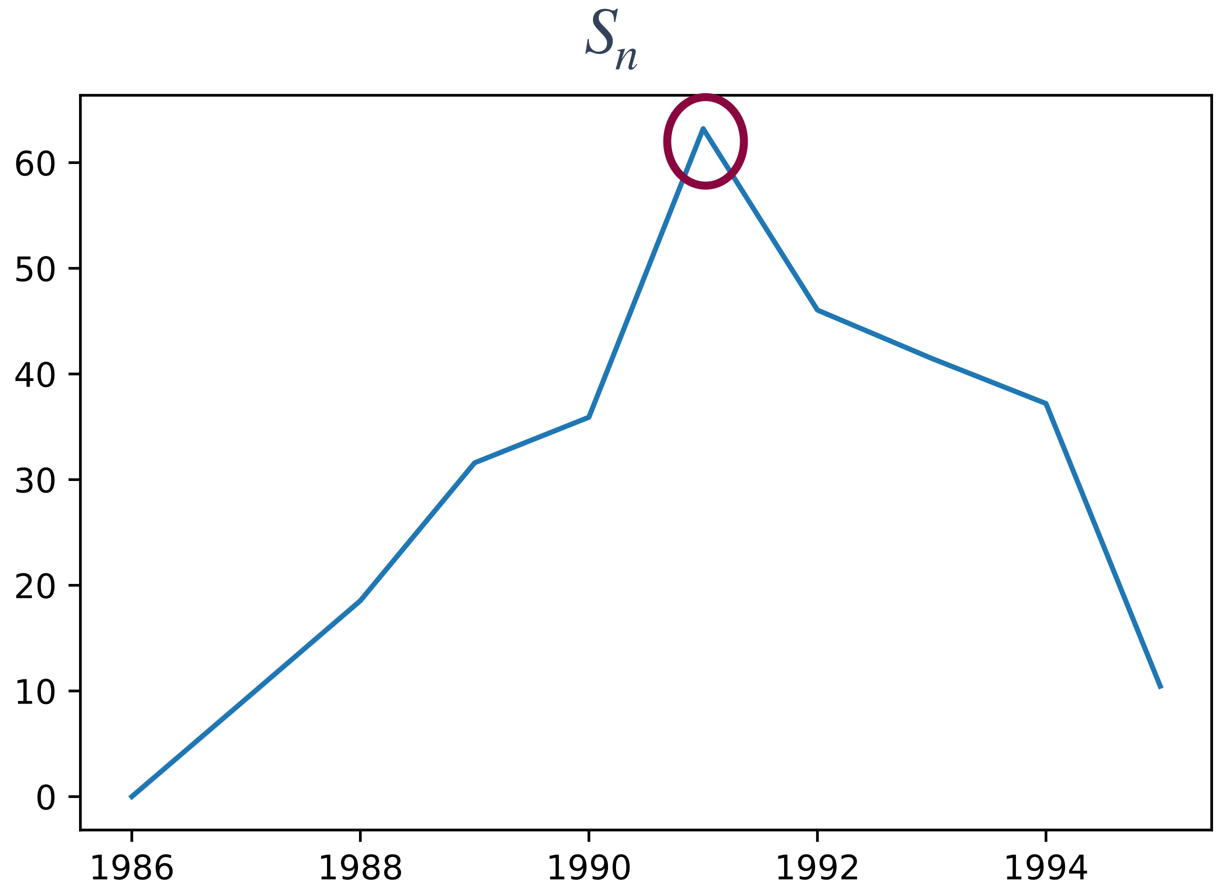
Shift in temperature in later months



# ERA5

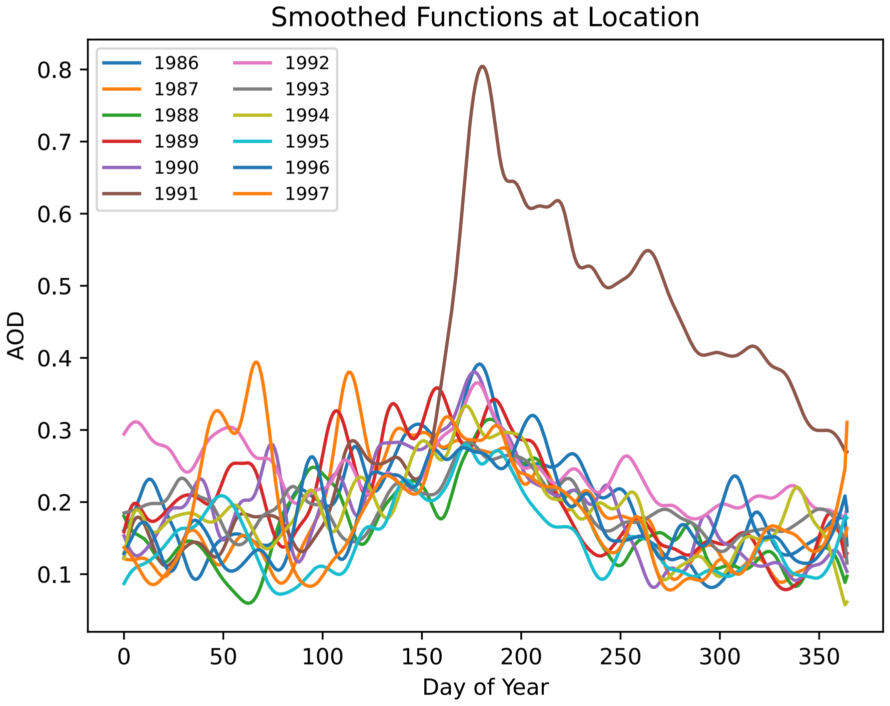
- Test statistic has clear peak at 1991
- Possible second sub-peaks, could look at multiple changepionts if needed
  - Theory not complete here for statistical test

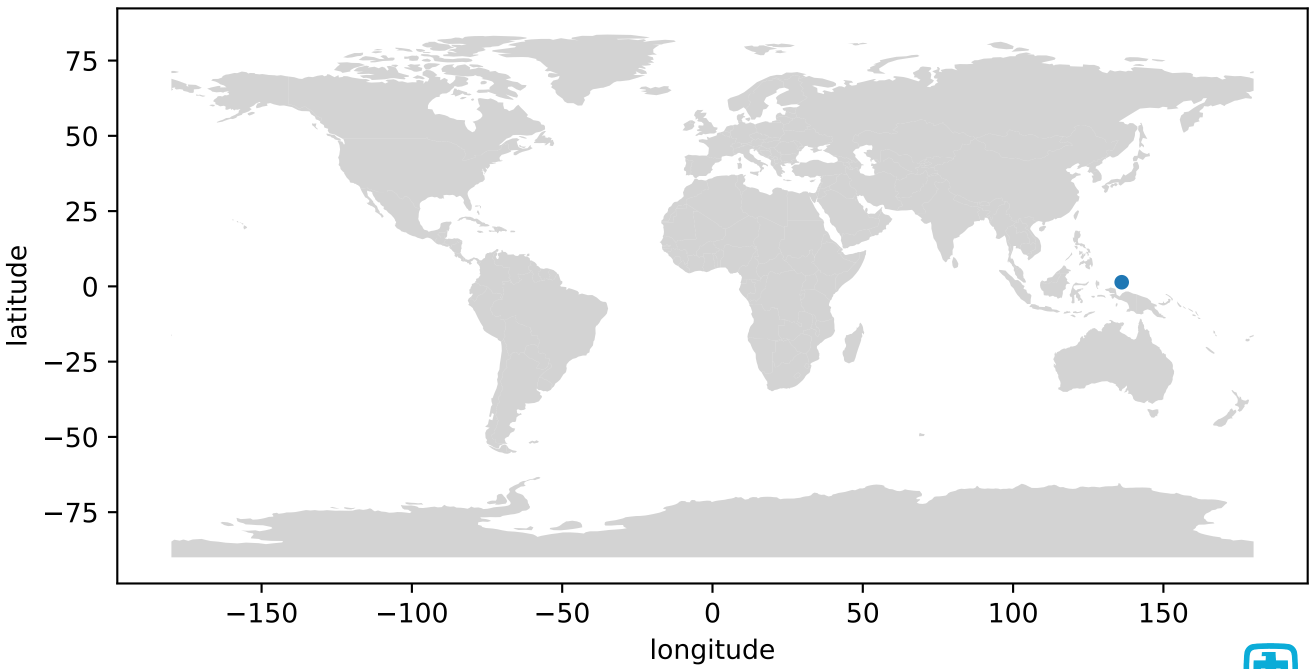


# MERRA-2

Aerosol Optical Depth (AOD) from MERRA-2 from 1986-1995 at a location near Mt. Pinatubo eruption

Functions from location that is spatially averaged

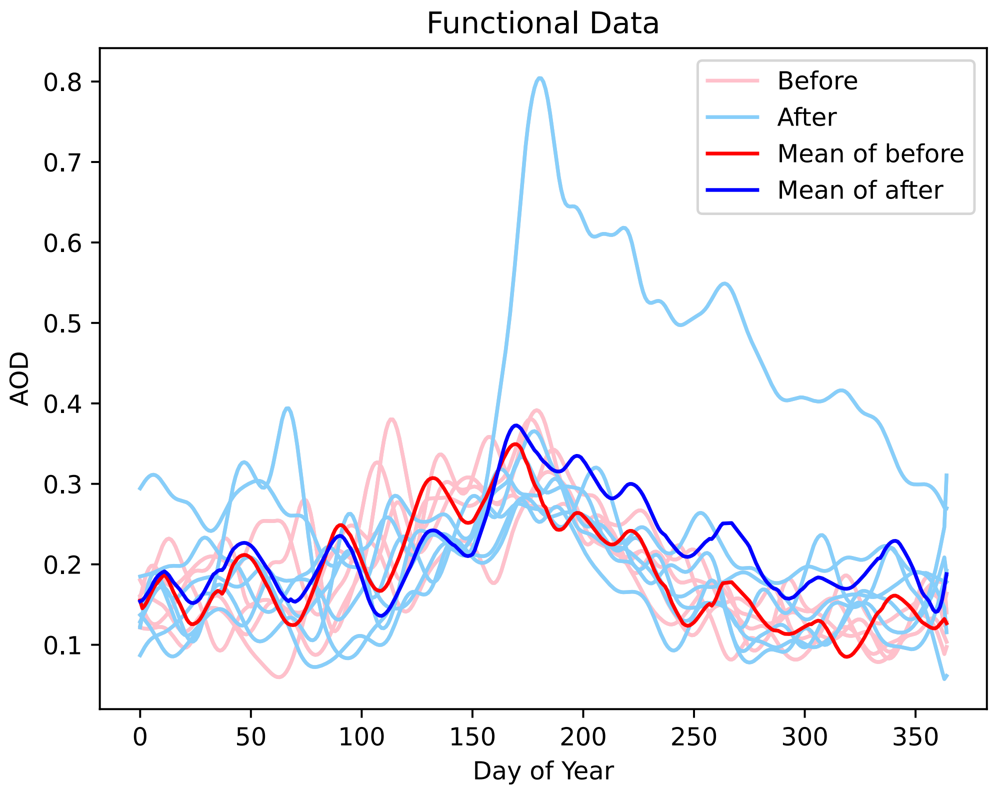


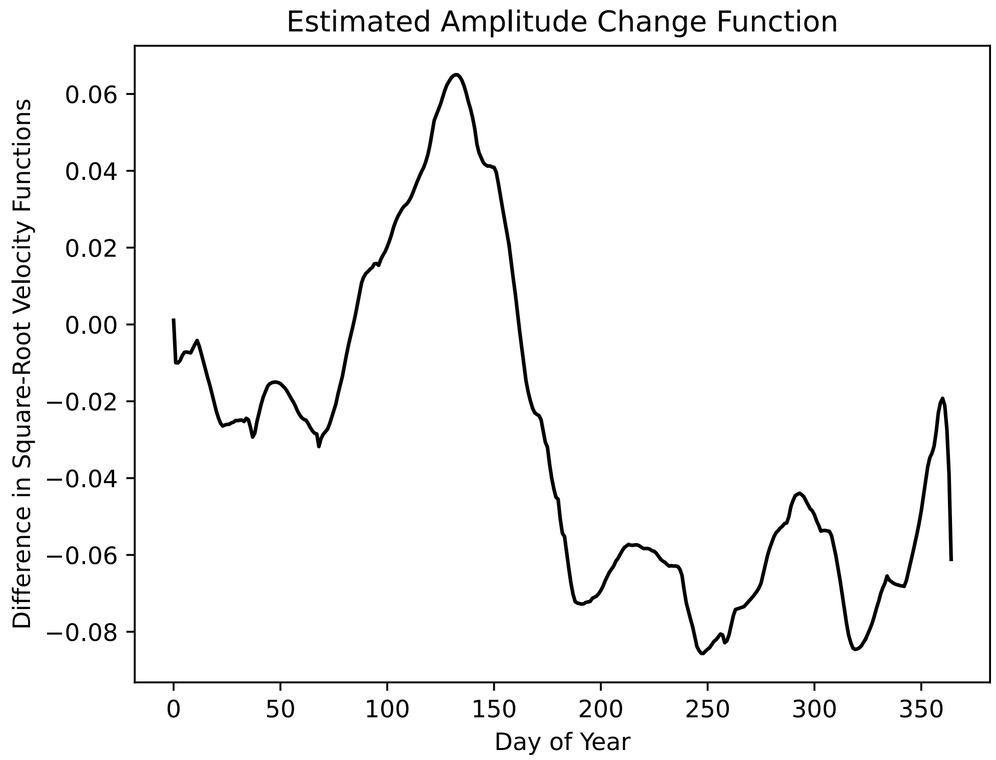



# MERRA-2

Only found a amplitude changepoint at 1991

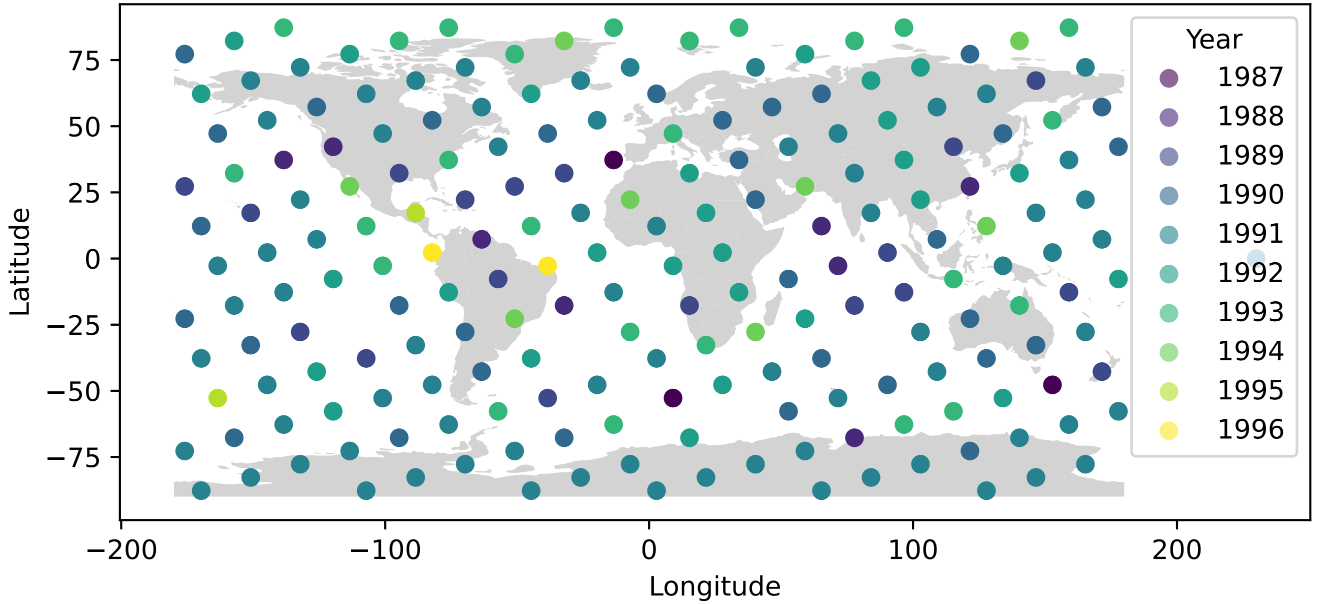
Shift in temperature in later months

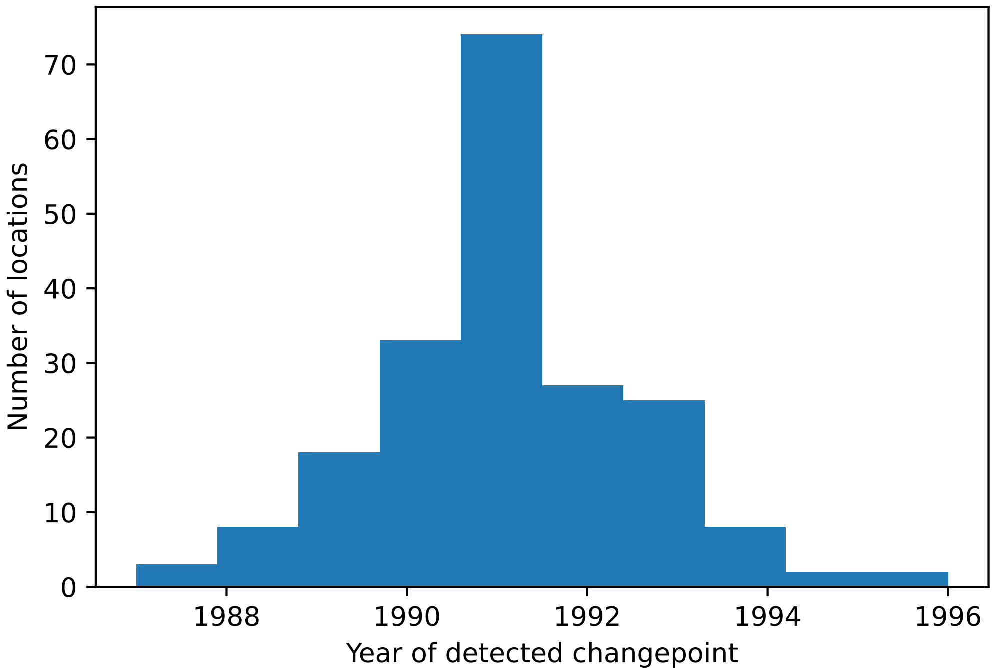


# MERRA-2

- Repeat analysis at different locations
- Significant change points at the 0.05 level were found and were concentrated around 1991
- Found that AOD decreased quickly to its pre-pinatubo behavior in 1992 or 1993
  - Considering an "epidemic" change point model may more accurately identify the change



# Conclusions

- Changepoint methods help inform other modeling strategies where the underlying statistical assumptions of data has changed
  - Help inform pathways, by identifying regions in inference solutions to evaluate further
- Integrated elastic functional metrics into elastic changepoint problem to detect both amplitude and phase change points
- Demonstrated ability on multiple simulated data sets and ERA-5 global temperature and MERRA-2 global AOD data sets
- Future Work
  - Expand to full spatial solution
- Paper at J. D. Tucker and D. Yarger, “Elastic Functional Changepoint Detection”, arXiv:2211.12687 [stat.ME], 2023

# Questions

jdtuck@sandia.gov