

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Master's Theses, 2022-2026

Computer Science, 2023

2024

Detecting Climate Change in Binary and Continuous Data

Liu, Grace

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01v692t9600>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

DETECTING CLIMATE CHANGE IN BINARY AND CONTINUOUS DATA

Volume I

Grace Liu

MASTER'S THESIS

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY IN CANDIDACY FOR THE DEGREE

OF MASTER OF SCIENCE AND ENGINEERING

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF COMPUTER SCIENCE

Adviser: Tom Griffiths

May 2024

© Copyright by Grace Liu, 2024. All rights reserved.

Abstract

In this paper, we investigate whether and why people perceive greater change in binary climate data (e.g., lake freeze) compared to continuous climate data (e.g., winter temperature). We conduct a behavioral experiment and find that participants who view binary climate data observed greater change than participants who view continuous climate data. We explore emotional valence and changepoint saliency as potential explanations for this behavior through cognitive experiments and Bayesian changepoint detection.

Keywords: climate change detection; retrospective Bayesian changepoint detection; binary and continuous changepoint models

Contents

1	Introduction	5
2	Approach	7
3	Background	8
3.1	Preliminaries	8
3.2	Retrospective Bayesian Changepoint Detection (RBCD)	9
3.3	CPD for climate	9
4	Methods	10
4.1	Change Perception Experiment	10
4.1.1	Experiment Data	10
4.1.2	Experimental Procedure	11
4.2	Changepoint Detection Algorithm and Experiment	11
4.2.1	Retrospective Changepoint Detection (RCPD)	12
4.2.2	Experiment Data	13
4.2.3	Experimental Procedure	13
4.3	Valence Experiment	14
4.3.1	Experiment Data	14
4.3.2	Experimental Procedure	14
4.4	Data Analysis	14
5	Results	15
5.1	Change Perception Experiment	15

5.2	Changepoint Detection	16
5.2.1	RBCD Algorithm Output	17
5.2.2	Changepoint Detection Experiment	18
5.3	Valence Experiment	19
6	Discussion	21
6.1	Conclusion	23
7	Acknowledgments	24
8	Apendix	27
8.1	Data Generation	27
8.1.1	Binary Data Generation	27
8.1.2	Continuous Data Generation	27
8.2	Retrospective Bayesian Changepoint Detection Algorithm Details	29
8.2.1	RBCD in the Binary setting	29
8.2.2	RBCD in the Continuous setting	30

1 Introduction

Climate change is a pressing, human-induced phenomenon affecting the well-being of people and the planet. Beyond the most straightforward indicator of warming average seasonal temperatures, the impacts of climate change include decreasing frequency of lake freeze, increasing stress on water resources, higher risk of wildfires, and shifts in agricultural productivity[20, 7, 9, 5]. Despite the growing risks, a 2023 Pew Research survey found that only about 37% of Americans say that they personally care a great deal about the issue of climate change, down

seven percentage points from 2018[19]. This statistic reflects a widening gap between the severity of the climate issue and the magnitude of people’s responses to it. It follows that, to tackle this global challenge, we must understand how people perceive climate change, so that we can better communicate its impacts and implications.

Shifting Baseline Syndrome (SBS) is a potential underlying cause of climate change apathy. SBS, also referred to as the “boiling frog effect”, describes a psychological and social phenomenon in which people continually lower their accepted thresholds for environmental conditions in the midst of ongoing environmental degradation[17]. In the climate context, Moore et al. claim that people’s experiences of weather between 2 and 8 years ago serve as the baseline from which they evaluate change[12]. The authors employ sentiment analysis of over 2 billion social media posts and discover that the remarkability of temperature changes declines rapidly with repeated exposure. They cite expectations, memory limitations, and cognitive biases as factors which influence this boiling frog effect.

We propose that binary climate indicators, such as lake freeze, could be a more effective tool for climate communication and mitigation of SBS than more commonly presented continuous climate indicators, such as mean temperatures. A 2015 study found that in the US, the strongest predictors of climate change risk perceptions include beliefs about the cause of climate change and perceptions of local temperature changes[10]. Despite the emphasis on local climate variables, previous research has not explored the potential of binary climate stimuli for influencing climate change beliefs. To fill this gap, we investigate the difference in participants’ perceived change when presented with binary vs. continuous climate stimuli. We also examine the cognitive mechanisms that could underlie these differences in perceived change.

In doing so, we conduct online behavioral experiments and employ retrospective Bayesian changepoint detection algorithms to model the experimental results. The results indicate that

binary climate data could contain more salient markers of change than continuous climate data, giving the former potential as a tool for combating SBS. We also provide a theory based in Bayesian changepoint detection as the cognitive basis for the higher saliency of change in binary data.

2 Approach

We first design and deploy an experiment to examine whether people perceive greater change when presented with continuous climate stimuli (mean winter temperatures) or binary climate stimuli (lake freeze). The results of this behavioral experiment point to a significant increase in perceived change in the binary climate data. We then employ Bayesian changepoint detection algorithms and cognitive experiments to study the mechanisms underlying this difference in change perception. We consider two potential explanations:

- Changes are more salient in binary data than in continuous data.
- People have a stronger emotional connection to lake freeze than winter temperatures.

We explore the first hypothesis by employing Retrospective Bayesian Changepoint Detection algorithms to find a posterior probability over the changepoint locations in the data. We then conduct experiments to validate these cognitive models. We explore the second hypothesis by conducting an experiment in which we vary the emotional valence of the plot shown to the participant.

3 Background

3.1 Preliminaries

Changepoint detection (CPD) involves identifying changes in the underlying distribution of data. CPD for time series data has a wide variety of applications in areas of medicine, finance, signal detection, climate, and human activity. We can formulate the problem as follows: Given a time series with a sequence of variables $\{x_n, x_{n+1}, \dots, x_m\}$, changepoint detection can be defined as a hypothesis test between the null hypothesis H_0 that no change occurs and the alternative hypothesis H_1 that a change occurs.

$$H_0 : \mathbb{P}(X_n) = \dots = \mathbb{P}(X_k) = \dots = \mathbb{P}(X_m)$$

$$H_1 : \exists k^* \text{ s.t. } \mathbb{P}(X_n) = \dots = \mathbb{P}(X_{k^*}) \neq \mathbb{P}(X_{k^*+1}) = \dots = \mathbb{P}(X_m)$$

where $P(X_i)$ is the probability density of the sliding window starting at X_i and k^* is a changepoint [1].

Notably, in previous literature, a changepoint is often defined as a instantaneous shift or jump in the parameters of the data-generating distribution. However, in this study, we generate the data with a constant rate of change for the parameters of the distribution, so there is no shift at any specific point. Therefore, we consider a changepoint in this context to be a perceptual shift rather than an intrinsic shift since there are no true changepoints underlying the data generation. That is, we investigate how and where people detect changepoints when the data is generated with a steady rate of change. We believe that updating parameters with a steady rate a change can better mimic real-world conditions because seasonal, atmospheric greenhouse gas concentrations increase at a steady, though not necessarily linear, rate[8].

3.2 Retrospective Bayesian Changepoint Detection

(RBCD)

In this section we will describe an algorithm for Bayesian single changepoint detection in the offline setting. The following problem setup is adapted from [18].

Let $\mathbf{X}$ be our time series data of length N , θ be the parameter of the distribution the data is sampled from, Ψ be the hyperparameters of the prior distribution of θ , and k_i be the event of a changepoint at timestep i . We assume that the data $\mathbf{X}$ are conditionally i.i.d., given θ . We have the following relationship for the probability of a changepoint given the data and hyperparameters of the prior:

$$\begin{aligned}
P(k_i|\mathbf{X}, \Psi) &\propto P(\mathbf{X}|k_i, \Psi)P(k_i) \\
&= P(k_i) * \int P(\mathbf{X}|k_i, \theta_1, \theta_2)P(\theta_1, \theta_2|\Psi)d\theta_1d\theta_2 \\
&= P(k_i) * \int P(\theta_1, \theta_2|\Psi) \prod_{t=1}^N P(x_t|k_i, \theta_1, \theta_2)d\theta_1d\theta_2 \\
&= P(k_i) * \int P(\theta_1, \theta_2|\Psi) \prod_{t=1}^i P(x_t|k_i, \theta_1, \theta_2) \prod_{t=i+1}^N P(x_t|k_i, \theta_1, \theta_2)d\theta_1d\theta_2 \\
&= P(k_i) * \int P(\theta_1|\Psi)P(\theta_2|\Psi) \prod_{t=1}^i P(x_t|\theta_1) \prod_{t=i+1}^N P(x_t|\theta_2)d\theta_1d\theta_2 \\
&= P(k_i) * \int P(\theta_1|\Psi) \prod_{t=1}^i P(x_t|\theta_1)d\theta_1 \int P(\theta_2|\Psi) \prod_{t=i+1}^N P(x_t|\theta_2)d\theta_2 \tag{1}
\end{aligned}$$

During implementation, we place a uniform prior on location of the changepoint, $P(k_i)$.

3.3 CPD for climate

Previous works have dealt with the topic of CPD for climate data, but not on CPD for climate change perception. For example, Reeves et al. compare a number of CPD techniques for climate and meteorological data, including the standard normal homogeneity (SNH) test, Wilcoxon’s nonparametric test, two-phase regression (TPR) procedures, and Sawa Bayes criteria. The authors identify that Bayesian techniques that impose a prior over the possible change-

points are promising [14], though little research has been done in this area. Caussiness and Mestre have used a technique similar to Sawa Bayes to detect shifts in temperature but they do not impose a prior on the changepoints [3]. Although Bayesian CPD methods are relevant to climate change detection, there is a lack of research applying these techniques to climate data. In particular, there is a research gap in analyzing people’s responses to climate data through a cognitive science lens. To explore this gap, we investigate people’s responses to a baseline change perception experiment and a CPD experiment for binary and continuous climate data. We also model their responses to the CPD experiment using offline Bayesian CPD algorithms.

4 Methods

4.1 Change Perception Experiment

We design the change perception experiment to test whether participants shown binary climate data detect greater change compared to participants shown continuous climate data. In the following sections, we describe the data used and the experimental design.

4.1.1 Experiment Data

We generated 18 paired time series across the correlation range of (0.1-0.7). Each pair consisted of a binary time series and a continuous time series with matched correlations. We grouped the pairs into three correlation categories - low correlation (0.1-0.3), medium correlation (0.3-0.5), and high correlation (0.5-0.7) - with six pairs in each category. Details on how we generated the data and controlled for correlation are given in Appendix section 8.1.

4.1.2 Experimental Procedure

At the start of the experiment, the participant reads some information on the fictional town of Townsville, which is known for its chilly winters and wintertime activities. Then, the participant views a plot visualizing one of the 36 generated time series, sampled uniformly at random. In the plots, the binary data is labeled as lake freeze in Townsville, and the continuous data is labeled as mean winter temperatures in Townsville. The participant then rates on a scale from 1-10 how much change they observed in the categories of temperatures, freeze, and climate. This and the following experiments also contain an attention check question, and participants who do not pass the attention check are filtered out during data analysis.

4.2 Changepoint Detection Algorithm and Experiment

In this section, we detail the changepoint algorithm and experiment designed to examine the reasons underlying the difference in perceived change in the binary vs. continuous setting. We propose that changes in binary data could be more salient than changes in continuous data when analyzed through the lens of changepoint detection. In this vein, we compute the posterior probability distribution according to the RBCD method, given by equation (2) in the binary setting and equation (3) in the continuous setting. We then conduct a behavioral experiment in which participants select a changepoint in the data. Using the experimental data, we evaluate whether RBCD provides a reasonable cognitive basis for rational change detection.

4.2.1 Retrospective Changepoint Detection (RCPD)

We have from section 3.1 that the posterior probability of a changepoint in a timeseries $\mathbf{X}$ of length N can be given by:

$$\begin{aligned} P(k_i|\mathbf{X}, \Psi) &\propto P(\mathbf{X}|k_i, \Psi)P(k_i) \\ &= P(k_i) * \int P(\theta_1|\Psi) \prod_{t=1}^i P(x_t|\theta_1) d\theta_1 \int P(\theta_2|\Psi) \prod_{t=i+1}^N P(x_t|\theta_2) d\theta_2 \end{aligned}$$

where θ represents the parameter of the distribution the data is sampled from, Ψ represents the hyperparameters of the prior distribution of θ , and k_i is the event of a changepoint at timestep i .

In the binary setting, we assume that the data $\{x_1, \dots, x_i\}$ are sampled from a Bernoulli distribution with parameter θ_1 and $\{x_{i+1}, \dots, x_N\}$ are sampled from a Bernoulli distribution with parameter θ_2 . We place a Beta prior on θ_1 and θ_2 with hyperparameters a, b . We marginalize over the parameters θ_1 and θ_2 . We assume a uniform prior for $P(k_i)$, the probability that the changepoint takes place directly after timestep i . Let $S_l = \sum_{t=1}^i x_t$ and $S_r = \sum_{t=i+1}^N x_t$.

Starting from equation (1), we can arrive at the following closed-form solution for the posterior changepoint probability. The derivation is given in Appendix section 8.2.1.

$$P(k_i|\mathbf{X}, \Psi) \propto \frac{1}{N} \cdot \frac{B(a + S_l, b + i - S_l) B(a + S_r, b + N - i - S_r)}{B(a, b)^2} \quad (2)$$

Where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function. During implementation, we place a Jeffrey's prior of $Beta(0.5, 0.5)$ on both θ_1 and θ_2 .

For the continuous setting, we assume that the data $\{x_1, \dots, x_i\}$ are sampled from a Normal distribution with mean μ_1 and variance σ^2 , and the data $\{x_{i+1}, \dots, x_N\}$ are sampled from a Normal distribution with different mean μ_2 and the same variance σ^2 . We place an uninformative uniform prior on μ_1 and μ_2 . We also assume a uniform prior for $P(k_i)$, the probability that the changepoint takes place directly after timestep i .

Marginalizing over the parameters μ_1 and μ_2 , this algorithm is equivalent to the simple Bayesian step detector described in [13], which has the following closed-form solution.

$$P(k_i|\mathbf{X}, \Psi, \sigma) \propto \frac{(2\pi\sigma^2)^{-(\frac{N-1}{2})}}{\sqrt{i(N-i)}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{t=1}^N x_t^2 - \frac{1}{i} S_l^2 - \frac{1}{N-i} S_r^2 \right) \right] \quad (3)$$

where $S_l = \sum_{t=1}^i x_t$ and $S_r = \sum_{t=i+1}^N x_t$. During implementation, we use the fixed variance used to generate our continuous data for σ .

4.2.2 Experiment Data

We generated three paired time series, each consisting of a binary time series and a continuous time series with matching correlations. We chose to generate one pair with low correlation (0.1-0.3), one with medium correlation (0.3-0.5), and one with high correlation (0.5-0.7).

4.2.3 Experimental Procedure

The experiment begins by presenting the participant with information on the fictional town of Townsville, a small town known for its chilly winters and wintertime activities. The instructions tell the participant that their goal is to identify a changepoint in some data relating to Townsville, where a changepoint is defined as “the year where you notice the most pronounced deviation or shift from the typical pattern of winter temperature data”.

Following the instructions, the participant views a plot visualizing one of the six generated time series, sampled uniformly at random. The binary data is labeled as lake freeze and the continuous data is labeled as mean winter temperatures. The participant is asked whether they observe a changepoint in the data with the multiple choice options “Yes”, “No”, and “Not sure”. Then, the participant sees the plot again and drags a slider to the year in which they notice the most pronounced shift from the typical pattern of winter temperature data.

4.3 Valence Experiment

In this section, we describe the experiment and data used to examine the impact of emotional valence on perceived change. The high valence category consisted of binary data labeled as lake freeze, and the low valence category consisted of binary data labeled as above or below an arbitrary temperature threshold.

4.3.1 Experiment Data

In the valence experiment, we used binary time series from the initial change perception experiment, consisting of 9 binary time series across the correlation range 0.1-0.7. The main difference in this experiment lies in how we presented the data; approximately half of the participants viewed a plot in which the labels were changed to be “above 29° F” and “below 29° F” in place of the previous “Freeze” and “Did not freeze” labels.

4.3.2 Experimental Procedure

The experimental procedure was the same as that of the change perception experiment. However, instead of presenting a plot of either binary or continuous data, the experiment presents either a high valence plot or a low valence plot. The high valence plots included the original labels of lake freeze whereas the low valence plots labeled the binary data as an indicator of whether the mean winter temperature was above or below 29° F.

4.4 Data Analysis

For the change perception experiment and valence experiment, we used a paired t-test to evaluate the difference in participants’ responses of perceived change across the three categories of climate, winter temperature, and lake freeze.

For the changepoint detection algorithm, we used a t-test to measure the difference in entropy of the outputted posterior distributions over the changepoint location. We conducted the t-test with a sample of 30 generated pairs of data across the correlation range $[0.1, 0.7]$. We also computed and compared the entropy of the human-detected changepoint distributions in the binary and continuous settings.

5 Results

5.1 Change Perception Experiment

The experimental results reveal that people observed greater change across all three categories when shown the binary data, especially when shown plots in the medium and high correlation ranges (Fig. 1). These findings constitute the core takeaway of this study, and the experiments that follow aim to probe the cognitive basis of these results.

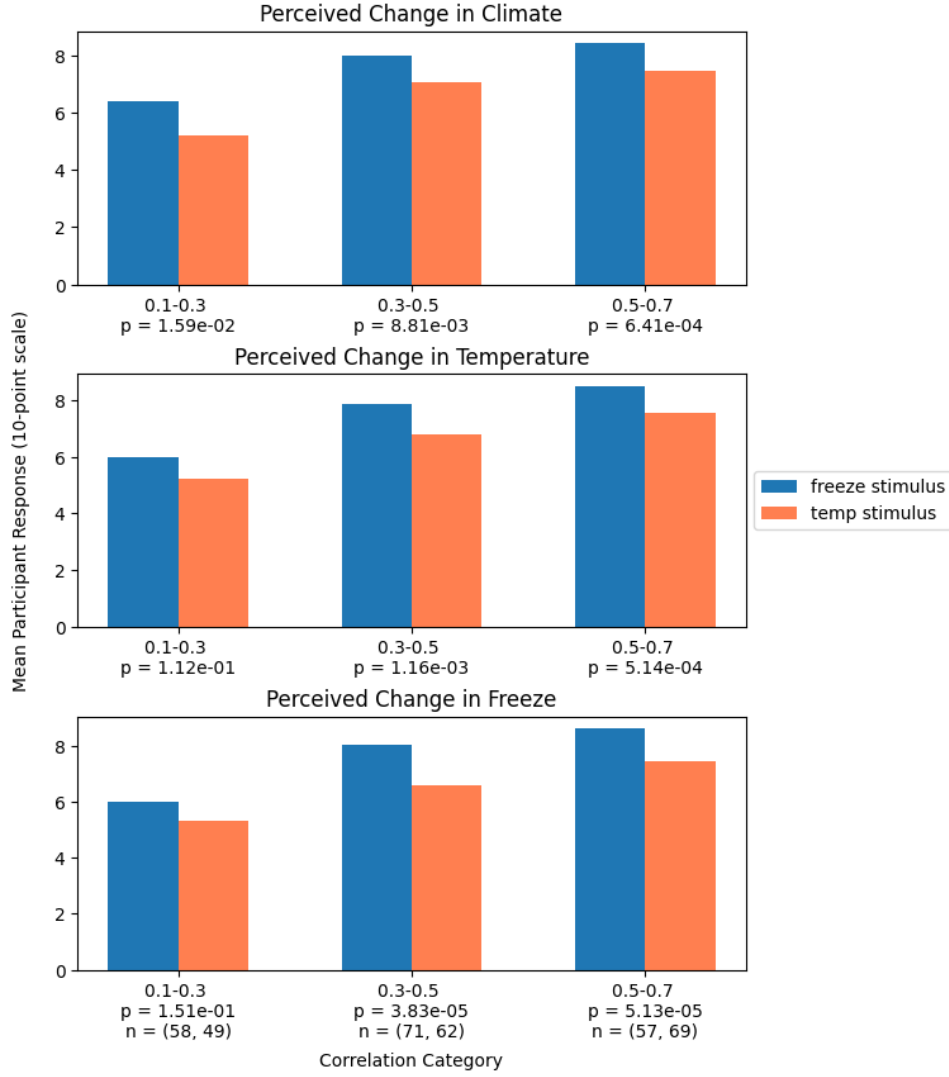


Figure 1: Mean of the participants' responses for perceived change in overall climate, mean winter temperatures, and frequency of lake freeze on a 1-10 scale. The results for each category are grouped into low, medium, and high correlation stimuli. The p-value for a pairwise t-test is given for each correlation/category combination. Additionally, the sample size n for each correlation category is given at the bottom of the figure.

5.2 Changepoint Detection

The output of the RBCD algorithms indicates that the probability mass of the changepoint posterior for the binary data tends to be more concentrated around specific years than for the

continuous data. The difference in concentration is quantified by the lower entropy of the binary data changepoint posteriors. The experimental results verify this trend with higher changepoint location consensus among the participants in the binary case.

5.2.1 RBCD Algorithm Output

To assess the shape of posterior changepoint probability distributions, we use the entropy of the outputted changepoint posterior distribution, where higher entropy corresponds to a greater KL divergence from the uniform distribution. Figures 2 and 3 below show sample outputs of the RBCD algorithms for binary and continuous data with matched correlations. To compare the distribution shapes in the continuous vs. binary case, we ran a paired t-test with $n = 30$ samples of generated data across matched correlation values of $[0.1, 0.7]$. The test finds a statistically significant difference in the entropy, with a p-value of 0.011 and a mean difference of 0.359. The figures below show an example of the output of RBCD for a continuous time series and a binary time series with matched correlation.

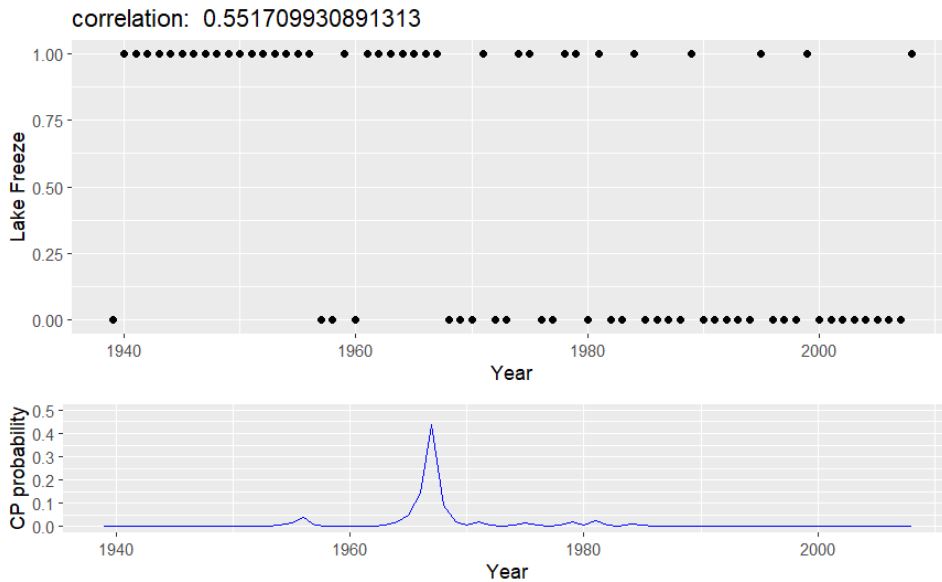


Figure 2: RBCD posterior probability of changepoint in binary time-series.

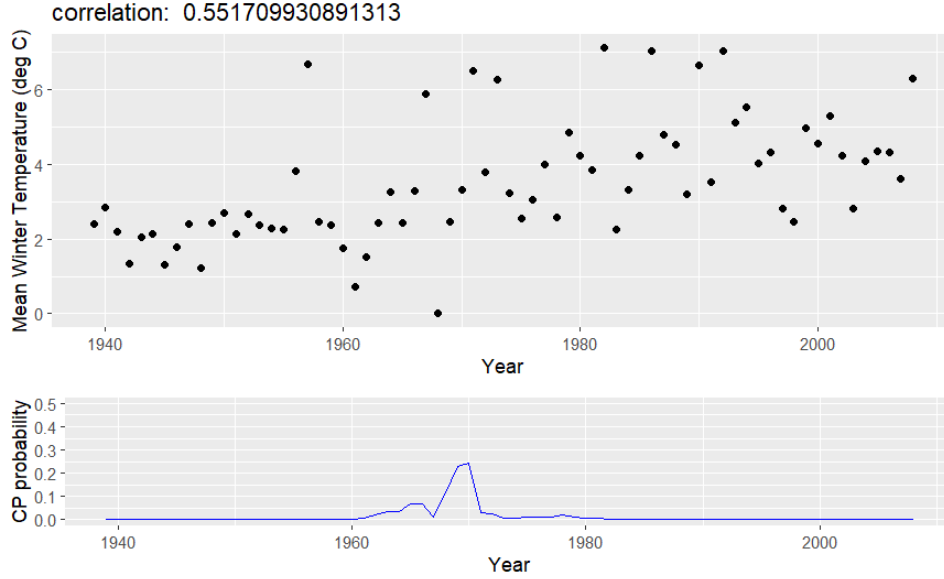


Figure 3: RBCD posterior probability of changepoint in continuous time-series.

5.2.2 Changepoint Detection Experiment

In the changepoint detection experiment, we found that participants who viewed the binary data were more likely to detect changepoints. Figure 4 shows the participants' responses to the question of whether they detected a changepoint in the data. The results indicate that participants who saw the continuous data more often failed to identify a changepoint compared to participants who saw the binary data.

Participants who viewed the binary data also tended to have more consensus on the location of the changepoints. Figure 5 displays a frequency histogram of the years in which participants identified a changepoint. The distributions for the binary setting have higher peaks and lower entropy ($\mathcal{H} = 2.75$) than the distributions for the continuous setting ($\mathcal{H} = 3.27$). Additionally, the participants who viewed the binary data tended to identify changepoints earlier in the time series than participants who viewed the continuous data.

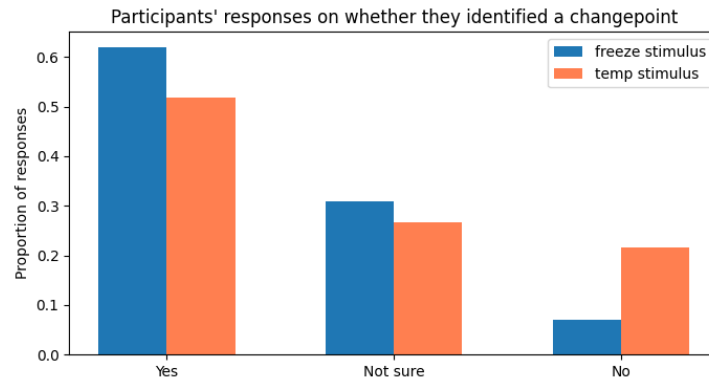


Figure 4: Participants' survey results for whether they detected a changepoint in the data. The results are grouped into low, medium, and high correlation stimuli.

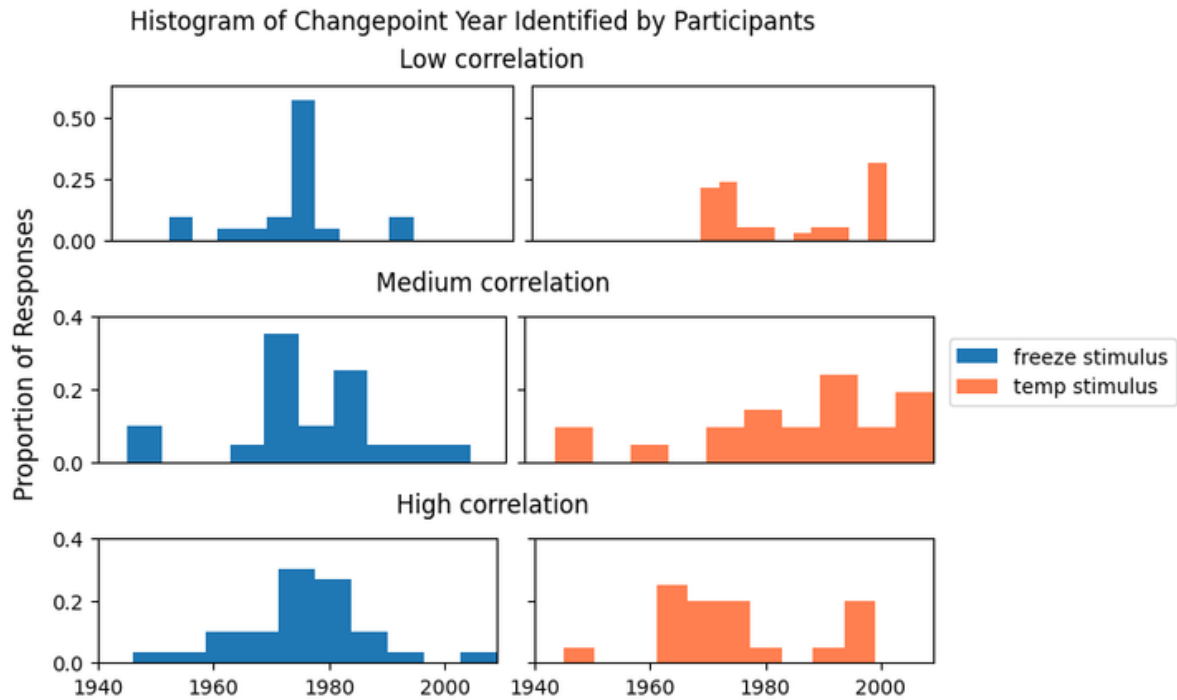


Figure 5: Frequency of changepoint years identified by participants. The results are grouped into low, medium, and high correlation stimuli.

5.3 Valence Experiment

The valence experiment tests whether people's emotional connection to lake freeze affects their observed change in binary climate data. Figure 6 shows the results of the experiment in which

participants viewed either a plot of binary data labeled as whether the lake froze (high valence) or the same plot of binary data labeled as whether the mean winter temperature was above or below an arbitrary threshold of 29° F (low valence). The results of the experiment suggest that valence could play an important role in change perception at low correlations, where there is not much of a trend to observe. However, there is no significant difference in change perception in the medium and high correlation categories.

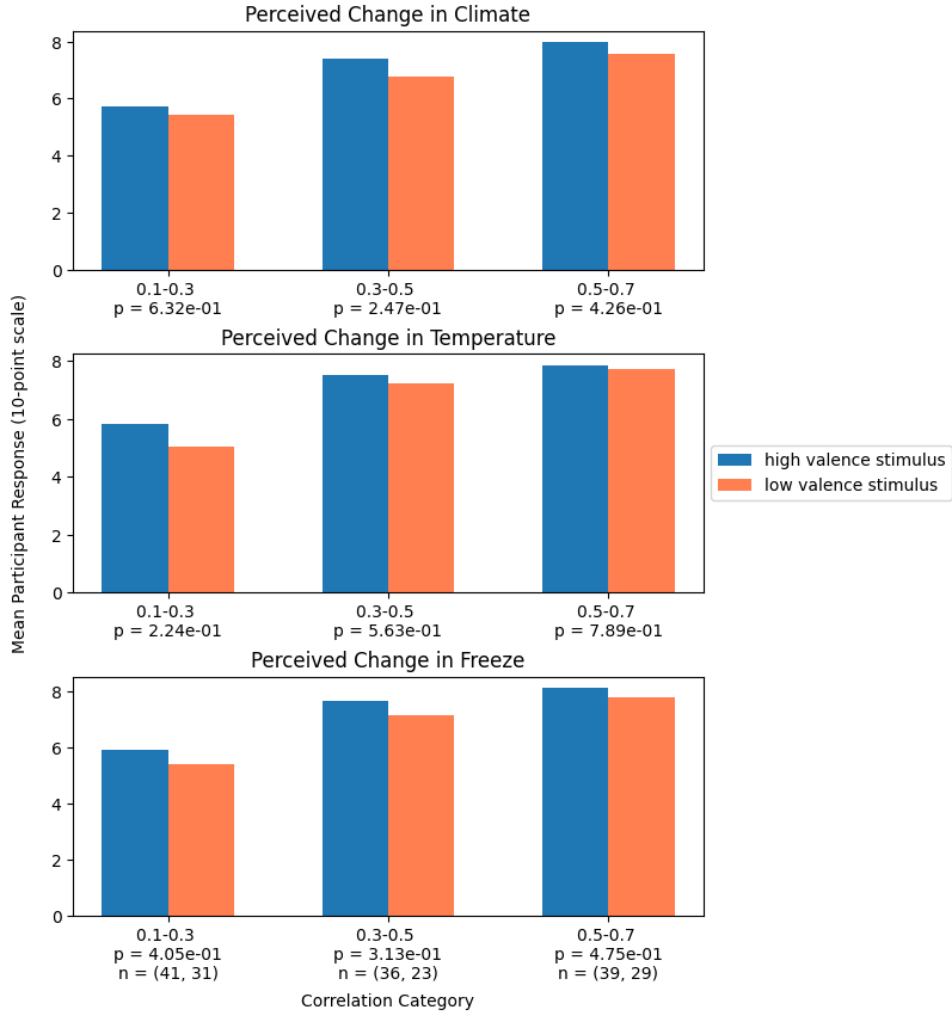


Figure 6: Mean response for perceived change in overall climate, mean winter temperatures, and frequency of lake freeze on a 1-10 scale. Participants were shown either a high valence or low valence stimulus. The results for each category are grouped into low, medium, and high correlation stimuli. The p-value for a pairwise t-test is given for each correlation/category combination.

6 Discussion

Modeling climate CPD from a Bayesian, cognitive lens is especially important for informing climate communication and education because it can provide insight into when and how people

detect changes in climate data. In this study, we find that people tend to observe greater change in binary climate stimuli compared to continuous climate stimuli. We also explore the cognitive underpinnings for this behavior computationally through retrospective Bayesian changepoint detection.

The alignment between the RCPD algorithm output and experimental CPD results indicates that changepoints are more salient in binary data than continuous data. In both the RCPD output and experimental results, the distribution over changepoints had lower entropy in the binary case, implying that there are higher peaks in the posterior probability distributions and greater consensus over changepoint locations. Additionally, people tended to detect changepoints earlier in the time series in the binary case. These characteristics suggest that binary data is more amenable to signposts for change (e.g., a string of years without lake freeze). Notably, these results hold even though we generated the data with a constant rate of change; there were no true changepoints in the data. It follows that binary climate indicators such as lake freeze could be an effective tool for combating SBS since they provide signposts for change even when the rate of change is somewhat fixed.

The valence experiment results suggest that emotional valence plays a significant role in perceived change in data with low correlation. That is, when the data lacks a clear trend, people rely more on the emotional valence of the data. However, at medium and high correlations, there was no significant difference in responses to the high and low valence stimuli. Future work could examine this dimension further by exploring how personal connection to the data affects perceived change. Lujala et al. find that direct personal experience of damage caused by climate-related events is one of the most important factors shaping people's views on climate change[11]. Therefore, it would be valuable to study the effectiveness of climate communication with data, especially binary data, targeted towards residents' local conditions.

Memory and computation limits provide another potential explanation for the increased change perception in binary climate data. While we did not conduct experiments along these lines, a research survey on visual data communication found that extracting global statistics is fast, but comparing between subsets of values is slow[4]. The authors state that effective graphics avoid taxing working memory and guide attention. In relation this project, binary plots require fewer visual comparisons of magnitude than continuous plots and decreased mental load. Therefore, people’s limited computational resources could also contribute to the increased change observed in the binary stimulus.

The format of this study is limited in the context of climate change detection because the participants observed the data in a single sitting and only processed the data visually. In the real world, people experience climate change over the course of years through multiple senses and activities. Thus, the results of this project should be considered in combination with climate change detection research beyond the scope of a controlled experiment. Additionally, we did not consider how people’s responses to lake freeze vary based on their cultural and community associations with ice skating [2]. Further research is needed to investigate the difference in responses to lake freeze vs. temperature when cultural and geographical factors are considered.

6.1 Conclusion

We find that people perceive greater change in binary climate data presented as lake freeze compared to continuous climate data presented as mean winter temperature. The saliency of changepoints in binary data could be an underlying mechanism for this difference in change perception. The results suggest that binary climate data could be a more effective tool for climate communications than continuous climate data.

7 Acknowledgments

A huge thanks to my mentors and collaborators, Rachit Dubey and Jake Snell. Rachit, your attention to detail, curiosity towards human behavior, and passion for climate research have been such an inspiration and motivation for me. Jake, I am so grateful for your insights and patience as I navigated Bayesian statistics for the first time. I also want to thank my M.S.E. advisor, Tom Griffiths, for supporting and guiding this project, from its inception as a COS454 class project to now. Finally, I would like to thank my second reader and undergrad advisor, Gabe Vecchi. I am so grateful to you for welcoming me into your research group and advising me from when I was a clueless freshman until I became a slightly less clueless senior. It's because of you that I am still pursuing research today.

References

- [1] Samaneh Aminikhanghahi and Diane J Cook. A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367, 2017.
- [2] Jeremy R Brammer, Jason Samson, and Murray M Humphries. Declining availability of outdoor skating in canada. *Nature Climate Change*, 5(1):2–4, 2015.
- [3] Henri Caussinus and Olivier Mestre. Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(3):405–425, 2004.
- [4] Steven L Franconeri, Lace M Padilla, Priti Shah, Jeffrey M Zacks, and Jessica Hullman. The science of visual data communication: What works. *Psychological Science in the public interest*, 22(3):110–161, 2021.

- [5] Jerry L Hatfield, John Antle, Karen A Garrett, Roberto Cesar Izaurrealde, Terry Mader, Elizabeth Marshall, Mark Nearing, G Philip Robertson, and Lewis Ziska. Indicators of climate change in agricultural systems. *Climatic Change*, 163:1719–1732, 2020.
- [6] William A. Huber. Generate a random variable with a defined correlation to an existing variable(s).
- [7] Matthew W Jones, Adam JP Smith, Richard Betts, Josep G Canadell, I Colin Prentice, and Corinne Le Quéré. Climate change increases the risk of wildfires: January 2020. *ScienceBrief*, 2020.
- [8] Ralph F Keeling, Steve C Piper, Allen F Bollenbacher, and JS Walker. Atmospheric carbon dioxide record from mauna loa. Technical report, Environmental System Science Data Infrastructure for a Virtual Ecosystem . . . , 2009.
- [9] Melissa E Lane, Paul H Kirshen, and Richard M Vogel. Indicators of impacts of global climate change on us water resources. *Journal of Water Resources Planning and Management*, 125(4):194–204, 1999.
- [10] Tien Ming Lee, Ezra M Markowitz, Peter D Howe, Chia-Ying Ko, and Anthony A Leiserowitz. Predictors of public climate change awareness and risk perception around the world. *Nature climate change*, 5(11):1014–1020, 2015.
- [11] Päivi Lujala, Haakon Lein, and Jan Ketil Rød. Climate change, natural hazards, and risk perception: the role of proximity and personal experience. *Local Environment*, 20(4):489–509, 2015.

- [12] Frances C Moore, Nick Obradovich, Flavio Lehner, and Patrick Baylis. Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. *Proceedings of the National Academy of Sciences*, 116(11):4905–4910, 2019.
- [13] Joseph JK Ó Ruanaidh, William J Fitzgerald, Joseph JK Ó Ruanaidh, and William J Fitzgerald. Retrospective changepoint detection. *Numerical Bayesian Methods Applied to Signal Processing*, pages 96–121, 1996.
- [14] Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, and Qi Qi Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology*, 46(6):900–915, 2007.
- [15] Robert A Rohde and Zeke Hausfather. The berkeley earth land/ocean temperature record. *Earth System Science Data*, 12(4):3469–3479, 2020.
- [16] Sapna Sharma, Kevin Blagrove, Alessandro Filazzola, Mohammad Imrit, and Harrie-Jan Hendricks Franssen. Patterns of ice cover for 31 lakes in the Northern Hemisphere. 10 2020.
- [17] Masashi Soga and Kevin J Gaston. Shifting baseline syndrome: causes, consequences, and implications. *Frontiers in Ecology and the Environment*, 16(4):222–230, 2018.
- [18] David A Stephens. Bayesian retrospective multiple-changepoint identification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):159–178, 1994.
- [19] Alec Tyson and Brian Kennedy. How americans view future harms from climate change in their community and around the us. 2023.

- [20] R Iestyn Woolway, Benjamin M Kraemer, John D Lenters, Christopher J Merchant, Catherine M O'Reilly, and Sapna Sharma. Global lake responses to climate change. *Nature Reviews Earth & Environment*, 1(8):388–403, 2020.

8 Appendix

8.1 Data Generation

We generated data within a given correlation range according to the following procedures. Each time series was generated to have 70 data points, corresponding to years 1939-2008.

8.1.1 Binary Data Generation

The binary data is generated using a linear search on the slope of the probability of freeze for each year. The y-intercept is adjusted so that the probability of freeze is 0.5 at the midway point of the time series. When the correlation is within the desired range, the algorithm terminates and outputs the sequence of data.

8.1.2 Continuous Data Generation

For each binary time series, the continuous data is generated to exactly match the correlation of the binary data using the following equation[6].

$$\mathbf{d}_{\mathbf{x};\rho} = \rho \text{SD}(\mathbf{x}^\perp)\mathbf{x} + \sqrt{1 - \rho^2} \text{SD}(\mathbf{x})\mathbf{x}^\perp.$$

where $\mathbf{x}$ is a vector representing the year and ρ is the target correlation. The data was then linearly transformed to have the same mean and variance as the mean winter temperatures extracted from the Berkeley Earth dataset [15] for a dataset of 31 intermittently freezing lakes[16].

Algorithm 1 Linear Search

Algorithm parameters: A small constant δ representing how much to increment the coefficient in each iteration and a number n representing the desired length of the sequence.

$m \leftarrow 0$

$\rho \leftarrow 0$

$\mathbf{x} \leftarrow [1, 2, \dots, n]$

while ρ not in desired range **do**

$m \leftarrow m + \delta$

$b \leftarrow 0.5 - \frac{m*n}{2}$

$\mathbf{y} = m * \mathbf{x} + b$

$\mathbf{d} \sim \text{binom}(n = n, \text{trials} = 1, p = \mathbf{y})$

$\rho = \text{correlation of } \mathbf{x} \text{ and } \mathbf{y}$

end while

Output $\mathbf{d}$, which is a sequence of binary data with a correlation within the desired range

8.2 Retrospective Bayesian Changepoint Detection Algorithm Details

8.2.1 RBCD in the Binary setting

In the binary setting, we assume that the data $\{x_1, \dots, x_i\}$ are sampled from a Bernoulli distribution with parameter θ_1 and $\{x_{i+1}, \dots, x_N\}$ are sampled from a Bernoulli distribution with parameter θ_2 . We place a Beta prior on θ_1 and θ_2 with hyperparameters a, b and marginalize out this parameter. We assume a uniform prior for $P(k_i)$, the probability that the changepoint takes place directly after timestep i . Let $S_l = \sum_{t=1}^i x_t$ and $S_r = \sum_{t=i+1}^N x_t$. The probability density function of the Beta distribution is given by $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$

Starting from equation (1), we can arrive at the following closed form solution for the posterior changepoint probability

$$\begin{aligned}
P(k_i | \mathbf{X}, \Psi) &\propto \\
P(k_i) \cdot \int P(\theta_1 | \Psi) \prod_{t=1}^i P(x_t | \theta_1) d\theta_1 \cdot \int P(\theta_2 | \Psi) \prod_{t=i+1}^N P(x_t | \theta_2) d\theta_2 \\
&= \frac{1}{N} \left[\int \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1} (1-\theta_1)^{b-1} \cdot \theta_1^{S_l} (1-\theta_1)^{i-S_l} d\theta_1 \right] \\
&\quad \cdot \left[\int \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_2^{a-1} (1-\theta_2)^{b-1} \cdot \theta_2^{S_r} (1-\theta_2)^{N-i-S_r} d\theta_2 \right] \\
&= \frac{1}{N} \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int \theta_1^{a+S_l-1} (1-\theta_1)^{b+i-S_l-1} d\theta_1 \right] \cdot \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int \theta_2^{a+S_r-1} (1-\theta_2)^{b+N-i-S_r-1} d\theta_2 \right] \\
&= \frac{1}{N} \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+S_l)\Gamma(b+i-S_l)}{\Gamma(a+b+i)} \right] \cdot \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+S_r)\Gamma(b+N-i-S_r)}{\Gamma(a+b+N-i)} \right] \\
&= \frac{1}{N} \cdot \frac{B(a+S_l, b+i-S_l)B(a+S_r, b+N-i-S_r)}{B(a, b)^2} \tag{2}
\end{aligned}$$

Where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function. During implementation, we use a Jeffrey's prior of $Beta(0.5, 0.5)$ on both θ_1 and θ_2 .

8.2.2 RBCD in the Continuous setting

For the continuous setting, we assume that the data $\{x_1, \dots, x_i\}$ are sampled from a Normal distribution with mean μ_1 and variance σ^2 , and the data $\{x_{i+1}, \dots, x_N\}$ are sampled from a Normal distribution with different mean μ_2 and the same variance σ^2 . We place an uninformative uniform prior on μ_1 and μ_2 . We also assume a uniform prior for $P(k_i)$, the probability that the changepoint takes place directly after timestep i . We assume a fixed variance.

After marginalizing out the parameters μ_1 and μ_2 , this procedure is equivalent to the simple Bayesian step detector described in [13], which has the following closed form solution.

$$P(k_i|\mathbf{X}, \Psi, \sigma) \propto \frac{(2\pi\sigma^2)^{-(\frac{N-1}{2})}}{\sqrt{i(N-i)}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{t=1}^N x_t^2 - \frac{1}{i} S_l^2 - \frac{1}{N-i} S_r^2 \right) \right] \quad (3)$$