Home Assignment 3: Coal mine disasters — constructing a complex MCMC algorithm & Parametric bootstrap for the 100-year Atlantic wave

FMSN50 - Monte Carlo and Empirical Methods for Stochastic Inference

Victoria Lagerstedt & Kajsa Hansson Willis, Group 3

March 11, 2025

Contents

1 Introduction

Coal mining disasters have varied over time due to technological, regulatory, and operational changes. Identifying shifts in disaster frequency is crucial for understanding risk patterns and improving safety measures. In this assignment, a hybrid Markov Chain Monte Carlo (MCMC) algorithm is applied to estimate change points in the disaster rate. Through Gibbs sampling and Metropolis-Hastings updates, the number and location of breakpoints, disaster intensity (λ), and the latent hyperparameter (θ) are examined.

Additionally, a parametric bootstrap approach is used to construct confidence intervals for the 100-year return value of significant wave height in the Atlantic.

The study aims to assess model uncertainty, compare different breakpoint structures, and provide a statistical foundation for analyzing historical trends in disaster and environmental risks.

2 Part 1: Coal mine disasters—constructing a complex MCMC algorithm

2.1 1a) The Marginal Posteriors

In the first part of this assignment, the marginal posteriors $f(\theta \mid \lambda, t, \tau)$, $f(\lambda \mid \theta, t, \tau)$, and $f(t \mid \theta, \lambda, \tau)$ were calculated, up to a normalizing constant. To do this, Equation 1 was used.

$$f(x|a,b,c) = \frac{f(x,a,b,c)}{f(a,b,c)} \tag{1}$$

First $f(\theta, \lambda, t, \tau)$ was calculated, and leveraging independence relationships the result can be seen in Equation 2. This is the numerator for all the marginal posteriors by following Equation 1.

$$f(\theta,\lambda,t,\tau) = f(\tau|\lambda,t) * f(\lambda|\theta) * f(\theta) * f(t) \tag{2}$$

The denominator for each marginal posterior using Equation 1 was calculated by using independence relationships. The results can be seen in Equations 3, 4, and 5.

$$f(\lambda,t,\tau) = f(\tau|\lambda,t) * f(\lambda) * f(t) \tag{3}$$

$$f(\theta,t,\tau) = f(\tau|\theta,t) * f(\theta) * f(t) \tag{4}$$

$$f(\theta,\lambda,\tau) = f(\tau|\lambda) * f(\lambda|\theta) * f(\theta) \tag{5}$$

Using these results, $f(\theta, \lambda, t, \tau)$, and simplifying by cancellation, the resulting marginal posteriors can be seen in Equations 6, 7, and 8

$$f(\theta|t,\lambda,\tau) = \frac{f(\lambda|\theta) * f(\theta)}{f(\lambda)} \tag{6}$$

$$f(\lambda|t,\theta,\tau) = \frac{f(\tau|\lambda,t) * f(\lambda|\theta)}{f(\tau|t,\theta)} \tag{7}$$

$$f(t|\theta,\lambda,\tau) = \frac{f(\tau|\lambda,t) * f(t)}{f(\tau|\lambda)} \tag{8}$$

Using equations given in the assignment, see Equation 9, 10, and 11, and the fact that the prior to λ is a gamma(2,θ) and the prior to θ is a gamma(2,Ψ), the distributions of the marginal posteriors were calculated.

$$f(\tau|\lambda,t) = \exp(-\sum_{i=1}^{d} \lambda_i(t_{i+1} - t_i)) \prod_{i=1}^{d} \lambda_i^{n_i(\tau)} \tag{9}$$

$$f(t) \propto \begin{cases} \prod_{i=1}^{d}(t_{i+1} - t_i), \text{ for } t_1 < t_2 < ... < t_d < t_{d+1} \\ 0, \text{ otherwise} \end{cases} \tag{10}$$

$$n_i(\tau) = \sum_{j=1}^{n} \mathbf{1}_{[t_i, t_{i+1})}(\tau_j) \tag{11}$$

In the resulting marginal posteriors it is only the numerator that is dependent on the marginal posterior, and therefore the denominator can be removed when calculating them.

The result when inserting the give distributions into Equation 6 can be seen in Equation 12. This equation looks to be proportionate to a gamma(2(d+1), $\Psi + \sum \lambda_i$) distribution and can therefore be samples using Gibbs sampling with this distribution.

$$f(\theta|t, \lambda, \tau) = \frac{\Psi^2 \theta^{2d+1} e^{-\Psi\theta}}{\Gamma(2)^2} \prod_{i=1}^{d} \lambda_i e^{-\theta\lambda_i} \tag{12}$$

Calculating the marginal posterior for lambda using the given equations and distributions gave the solution in Equation 13. For each individual lambda, this equation is proportionate to a gamma($n_i(\tau)$+ 2, $t_{i+1} - t_i + \theta$) distribution and can therefore be sampled with Gibbs sampling and this distribution.

$$f(\lambda|t, \theta, \tau) = \prod_{i=1}^{d} e^{-(\lambda_i(t_{i+1}-t_i)+\theta\lambda_i)} \lambda_i^{n_i(\tau)+1} \frac{\theta^2}{\Gamma(2)} \tag{13}$$

The result from the t-marginal posterior can be seen in Equation 14. There is not any distribution that is similar to this result, and therefore, the conclusion that it needs to be sampled with a Metropolis-Hasting (MH) sampler was drawn.

$$f(t|\theta, \lambda, \tau) = \prod_{i=1}^{d} (exp(-\lambda_i(t_{i+1} - t_i)) \prod_{j=1}^{n} (\lambda_i^{\mathbf{1}_{[t_i, t_{i+1})}(\tau_j)})(t_{i+1} - t_i)) \tag{14}$$

2.2 1b The Hybrid MCMC Algorithm

The goal of this part was to construct a hybrid MCMC algorithm that samples from the posterior $f(\theta, \lambda, t \mid \tau)$. This was done by sampling θ and λ using Gibbs sampling and the distributions found in part 2.1, followed by MH sampling to sample t. To update t, it was decided to use the option of random walk proposal all at once, which is done by implementing Equation 15 for all values at once as the proposal kernel. It can be noted that this is a symmetric proposal kernel.

$$t_i^* = t_i + \epsilon, \quad \epsilon \sim U(-\rho, \rho) \tag{15}$$

In order to have a correct Metropolis-Hastings algorithm there must be detailed balance. This is ensured by definition. Further, the Markov chain must be irreducible and aperiodic [1]. Since it is possible to reach every point of the chain within a finite number of steps, it is irreducible. The random walk part ensures that the Markov chain is aperiodic.

This method works because if one breakpoint is updated, it affects the whole dataset and the other might need updating as well. In this case the whole vector of new breakpoints is accepted or not instead of individual breakpoints. This method is more efficient and faster than updating one breakpoint at a time. The downside of this method is that there is a higher possibility that large movements are rejected.

The probability of accepting the new state given the old state in the MH-algorithm was calculated using Equation 16 and taking the minimum of this and 1. This probability was compared to a random number between 0 and 1, and if it was lower than the probability, the proposal was accepted. Otherwise, the previous value was set in its place. Every time the breakpoints were updated, the number of disasters in each interval was also updated. After testing different values for ρ in the updating of t and Ψ in the prior to θ, optimal values were found to be $\rho = 5$ and $\Psi = 0.05$.

The λ value and the t value for each iteration were then plotted to see how the chain developed.

$$\alpha = \frac{f_{\tau|t^*,\lambda}(\tau) \times f_t(t^*)}{f_{\tau|t,\lambda}(\tau) \times f_t(t)} = e^{-\sum_{i=1}^{d} \lambda_i(\epsilon_{i+1}-\epsilon_i)+\sum_{i=1}^{d} \ln(t_{i+1}-t_i)-\sum_{i=1}^{d} \ln(t^*_{i+1}-t^*_i)+\sum_{i=1}^{d}(n^*_i(\tau)-n_i(\tau))*\ln \lambda_i} \quad (16)$$

2.3 1c The Behavior of the Chain for Different Amounts of Breakpoints

Moving on to the next part of the assignment, the behavior of this chain with different amounts of breakpoints was investigated. This was controlled for 1, 2, 3, and 4 breakpoints. This corresponds to 2, 3, 4, and 5 intervals, respectively, between the start and end date of the data. To gain an initial understanding of the accident statistics, the continuous time data for each accident, τ, was plotted in Figure 1. The figure show the cumulative plot of the accidents with their date as the independent variable.

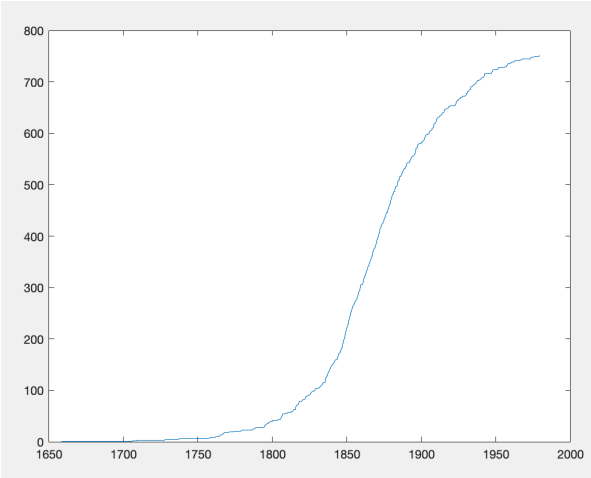

Figure 1: The cumulative number of accidents

An initial guess from Figure 1 was that two or three breakpoints could be appropriate. This was because at least three clear sections can be made out from the graph, where one is the years leading up to around 1820 where the slope is quite low. After these years, there is a rapid increase in the slope until about 1900. Then, either one or two intervals can be extracted for 1900 to 1980, where a possible breakpoint could be inserted around 1950.

The hybrid MCMC algorithm developed in part 2.2 was used with different numbers of breakpoints to do this. The number of samples was set to 10,000, with a burn-in period of 1,000 samples. Further, the Ψ value was set to 0.05, and the ρ value was set to 5. These values were determined to be sufficient from trial and error of different values.

For one breakpoint, the behavior of the chain is visible in Figure 2. It is apparent that the breakpoint converges to the lower limit. The acceptance rate for this plot is 1.38%, which is very low. Additionally, the λ trace plot can be observed in Figure 3. It is relatively stable for the different intervals.

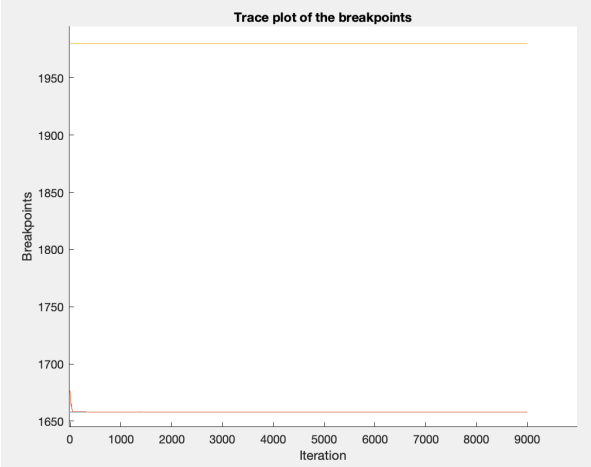

Figure 2: The behavior of the chain with one breakpoint in the time-period starting from 1658 to 1980

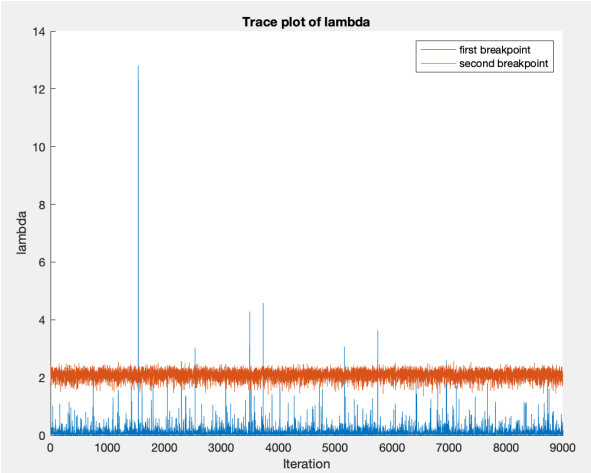

Figure 3: The intensities, λ, of the two intervals over the iterations

For two breakpoints, the behavior of the chain is visible in Figure 4. It can be observed that both of the breakpoints converges to a value around 1720 and follows each other almost perfectly. The acceptance rate for this plot is 2.48%, which is also very low. The λ trace plot can be observed in Figure 5. It is relatively stable for the different intervals.

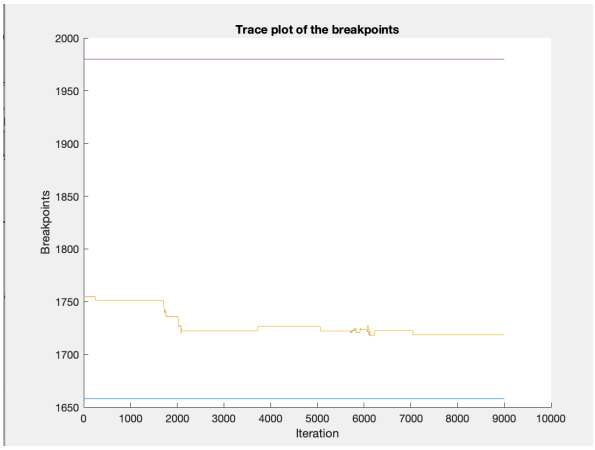

Figure 4: The behavior of the chain with two breakpoints in the time-period starting from 1658 to 1980

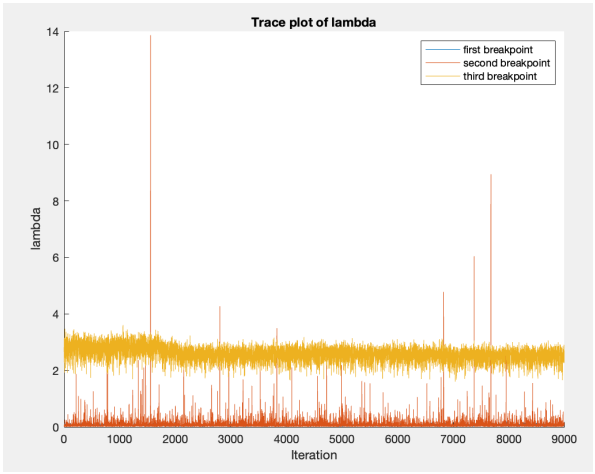

Figure 5: The intensities, λ, of the three intervals over the iterations

For three breakpoints, the behavior of the chain is visible in Figure 6. It is quite stable for all the breakpoints, and they converge at about 1910, 1800, and 1740. The acceptance rate for this plot was 33.40%, which is a very reasonable value.

The λ trace plot can be observed in Figure 7. The values are relatively stable, but the fluctuations are still considerably larger than they were for fewer breakpoints.

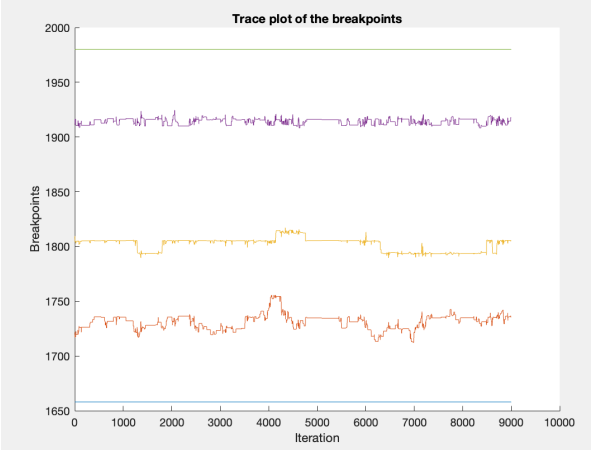

Figure 6: The behavior of the chain with three breakpoints in the time-period starting from 1658 to 1980

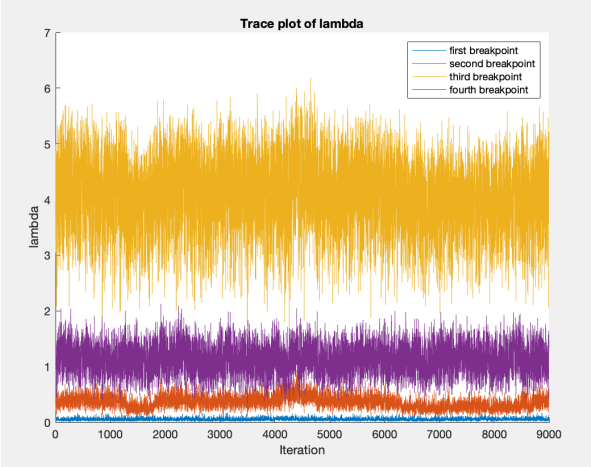

Figure 7: The intensities, λ, of the four intervals over the iterations

For four breakpoints, the behavior of the chain is visible in Figure 8. It is apparent that the first breakpoint converges to the lower limit, and the rest of the breakpoints stabilize at values corresponding to approximately 1740, 1800, and 1950. The acceptance rate is 32.20%, which is a very reasonable value.

The λ trace plot can be observed in Figure 9 and is considerably less stable than the trace plots for fewer amounts of breakpoints, however, most intensities are still quite stable.

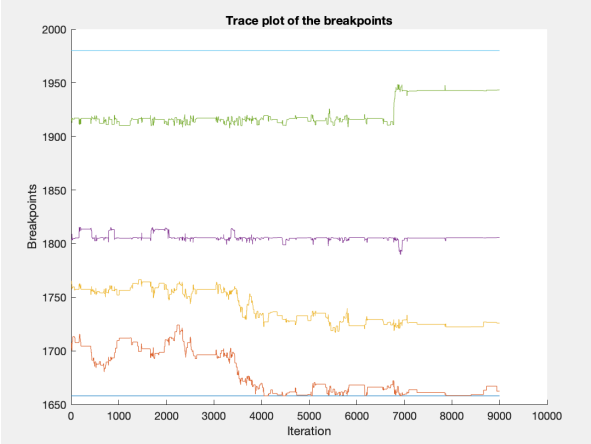

Figure 8: The behavior of the chain with four breakpoints in the time-period starting from 1658 to 1980

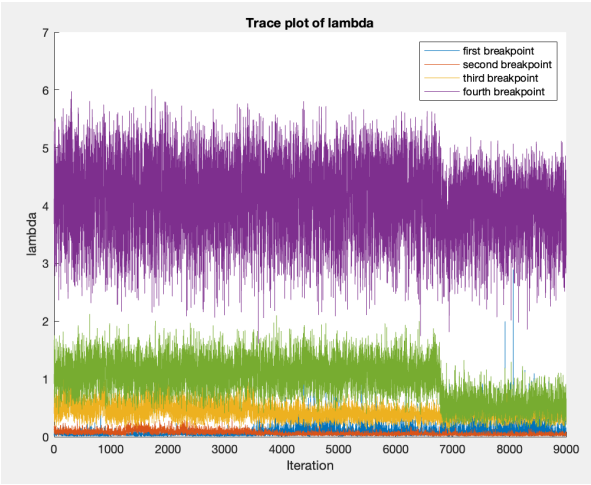

Figure 9: The intensities, λ, of the five intervals over the iterations

Comparing the chains for the different amounts of breakpoints, it is visible that only the chain with three breakpoints ends up with the same number of separate breakpoints after 10,000 samples. Many of the breakpoints for the different amounts are stable, which is a good sign.

To determine how many breakpoints are optimal, it is favorable to look for when the constructed MCMC chain is stable over iterations and that they seem to fit in where the slope of the cumulative number of accidents changes rapidly. Comparing the plots of the different number of breakpoints as well as the cumulative plot, one can say that the MCMC with three breakpoints is the most stable with the amount of breakpoints, and it appears to be reasonable when comparing the breakpoints to Figure 1. Further, it was one of the initial guesses of an appropriate amount of breakpoints and the acceptance rate is very reasonable. All this speaks to three breakpoints, and thus, four intervals, being the true amount of breakpoints.

2.4 1d Sensitivity of Posteriors for Ψ

The next task was to determine how sensitive the posteriors are to variations in the hyperparameter Ψ. This was investigated by varying the value of Ψ and examining how the θ, λ, and t posteriors vary with it.

In theory, the value of Ψ affects the distribution of θ, which is the scale parameter of the λ prior. In turn, the λ values affect the α-value, which is the ratio used to determine how likely the proposed

value is and is used to determine whether to accept the proposal or not. The $f(\theta, \lambda, t|\tau)$ posterior then depends on the values of λ.

The scale parameter of a gamma distribution affects how 'flat' the values will be spread out, i.e., how far they will spread out. If the parameter is smaller, the values will be very compacted and quite low, whereas if the parameter is larger, the values will be more spread and have a larger mean.

Thus, if the value of Ψ increases, then the expected value of θ will increase as well. This further indicates that the value of λ, whose scale parameter is dependent on θ will be increasing. The θ posterior's scale parameter increases with the value of Ψ and the λ values. Both of these factors increase when Ψ increases, so the θ posterior will increase with an increasing value of Ψ.

The λ posterior is dependent on the value of the θ posterior as the scale parameter. Thus, because the θ posterior is increasing with Ψ, the λ posterior will be as well.

The t posterior depends on the λ value both positively and negatively, and it is thus harder to determine theoretically and should rather be examined empirically.

The $f(\theta, \lambda, t|\tau)$ posterior includes both θ and λ, which mean that it will be affected by a change in Ψ. The t values will also be affected since they depend on λ. **XXX**

To determine the sensitivities empirically, different values of Ψ were used in the MCMC and each result was plotted where ρ has held constant at 5. The previous maximal value of breakpoints, 4, was used here to gain as much information as possible from the changes. The posterior for the base case where $\Psi = 0.05$ can be observed in Figure 8.

The chain where $\Psi = 2$ can be seen in Figure 10. It is clear that the values of the breakpoints vary much less, than for the case where $\Psi = 0.05$. This could lead to the samples not changing enough to find their optimal values. Further, the lambda posterior for $\Psi = 2$ can be observed in Figure 11. It is apparent that some of the breakpoints fluctuate quite a lot and the values are considerably larger, than for the case where $\Psi = 0.05$. This is in line with theory.

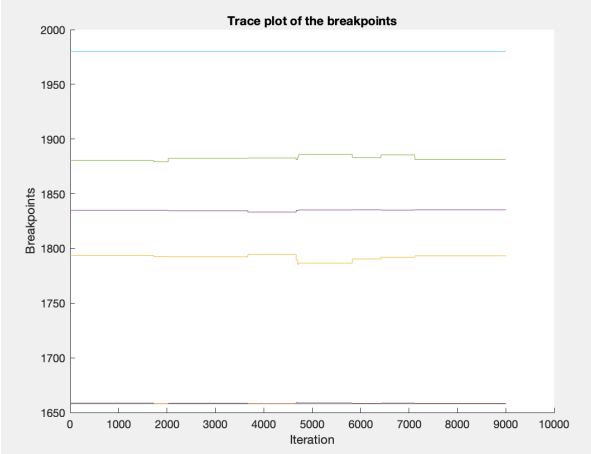

Figure 10: The MCMC for 4 breakpoints and $\Psi = 2$

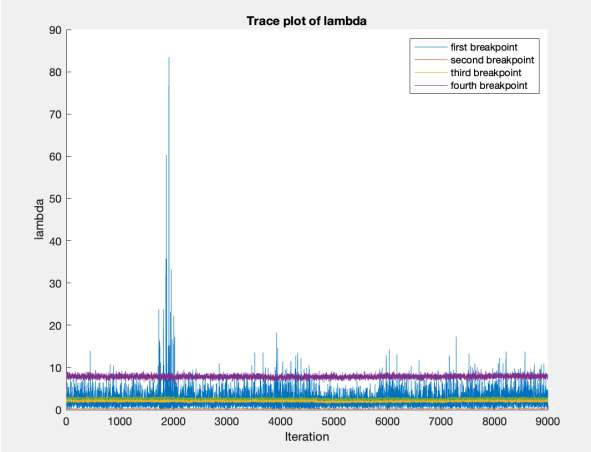

Figure 11: The lambda posterior values for $\Psi = 2$

Another chain where $\Psi = 0.001$ is displayed in Figure 12. This chain appears to behave quite strangely as it appears quite convergent at first but diverges from its path in later samples. Additionally, there is more convergence among paths.

The lambda posterior for $\Psi = 0.001$ can be viewed in Figure 13. The values are in general much smaller, and there are still fluctuations but they are smaller in absolute size. This is also in line with theory. This confirms that the lambda posterior increases with increasing Ψ.

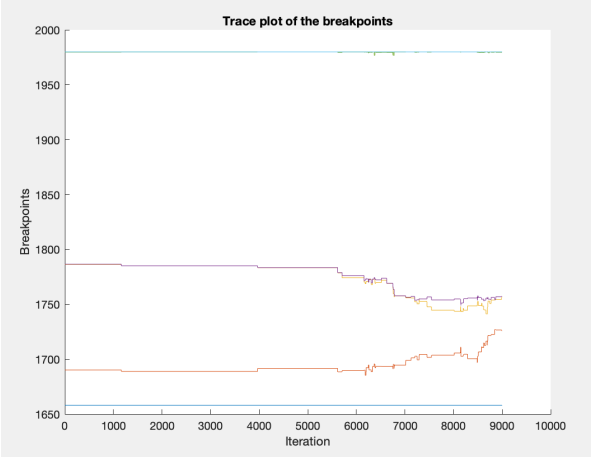

Figure 12: The MCMC for 4 breakpoints and $\Psi = 0.001$

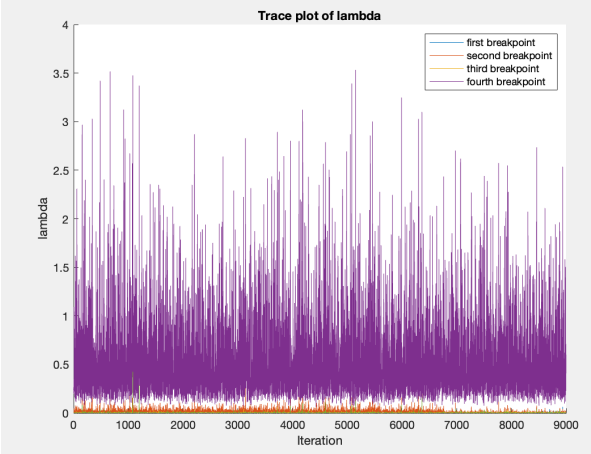

Figure 13: The lambda posterior values for $\Psi = 0.001$

Many more values of Ψ were tested and analyzed. Small changes in Ψ do not change the breakpoint results to a critical extent. But it does have a visible impact on how much the λ values fluctuate. This is reasonable since Ψ plays an important role in the sampling of the λ values.

2.5 1e Sensitivity of posteriors and mixing for ρ

In this part of the assignment, the sensitivity of the mixing and the posteriors to changes in ρ in the proposal distribution was examined. This was done by analyzing the problem analytically and empirically.

It is apparent that the value of ρ only directly impacts the t posterior, as ρ is only used in the Metropolis-Hastings sampler. Increasing ρ will lead to an increase in the variation of the new expected proposal. If accepted, the t posterior will thus fluctuate a lot more with an increasing ρ value.

This increase in variation will also lead to more rejections of the proposals. This will affect the mixing of the MCMC chain.

It is more difficult to say how the other posteriors will be affected as the values of t will be affected, but the rejection of proposals will play a role as well. Therefore, this is only studied empirically.

To examine the impact the value of ρ has, multiple different values were tested with the maximum amount of four breakpoints considered, see Figure 14 and 15 for some of the results. The value of Ψ for all of these were set to be 0.05. In these figures it is clear that a larger ρ value is making the trace plot fluctuate much more than for lower ρ. The optimal value was found to be 5, see Figure 16. The lambda trace plots can be seen in Figure 17, 18 and 19. There is no significant change in how the λ value fluctuates with a change in ρ but smaller ρ seems to make it fluctuate a bit more.

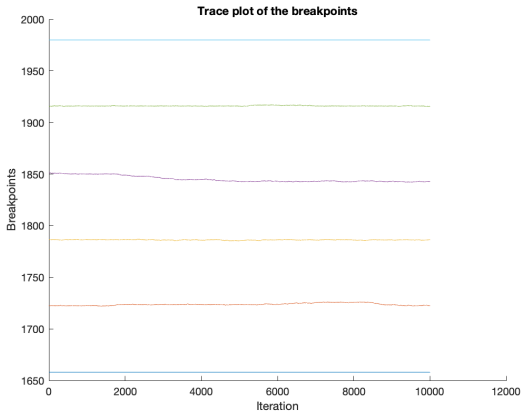

Figure 14: Trace plot of the breakpoints with $\rho = 0.1$

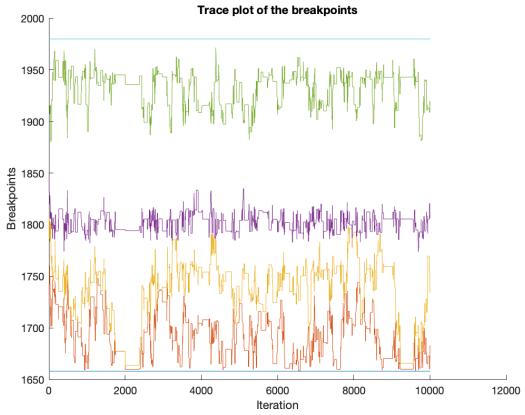

Figure 15: Trace plot of the breakpoints with $\rho = 20$

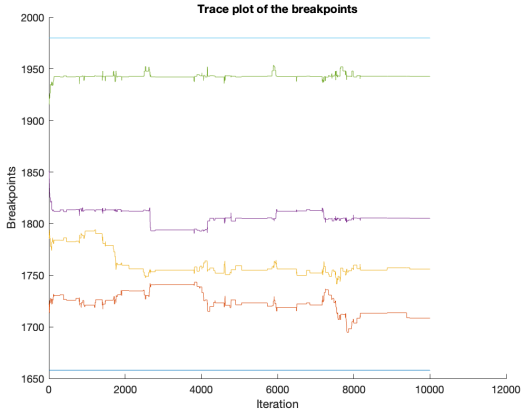

Figure 16: Trace plot of the breakpoints with $\rho = 5$

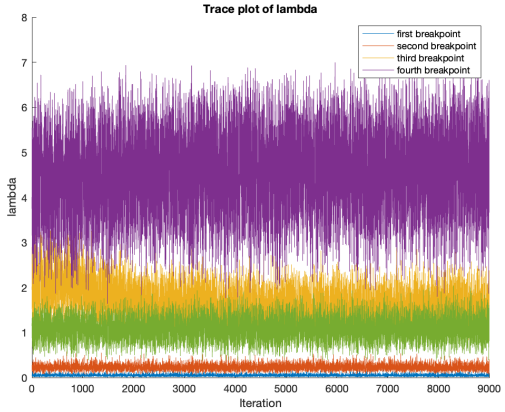

Figure 17: Trace plot of lambda with $\rho = 0.1$

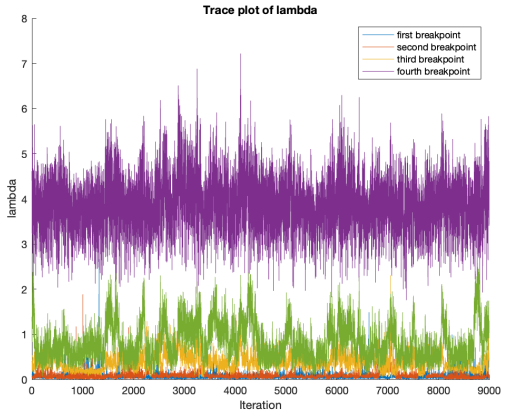

Figure 18: Trace plot of lambda with $\rho = 20$

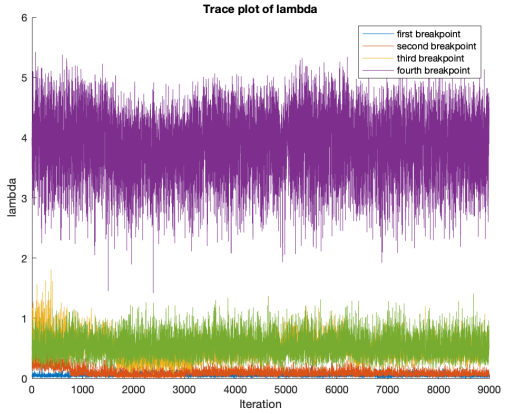

Figure 19: Trace plot of lambda with $\rho = 5$

Even more optimal would be if ρ was different for each breakpoint. With this, even more finetuning could have been made. One ρ is not optimal for all breakpoints, and different values might be needed, as can be seen in Figure 16, where some of the breakpoints are fluctuating much more than others. It might also be dependent on how many breakpoints there are and different ρ are optimal for different amount of breakpoints.

The mixing can be understood through the acceptance rates; for $\rho = 0.1$, the acceptance rate was 63.90%, for $\rho = 5$, the acceptance rate was 28.22%, and for $\rho = 20$, the acceptance rate was 27.98%. Clearly, a low value of ρ indicates a high acceptance rate, but when it increases drastically, the changes become less and less impactful. According to theory, values around 30% are good, which further indicates that $\rho = 5$ is quite appropriate.

3 Part 2: Parametric bootstrap for the 100-year Atlantic wave

In this part, the wave-height recorded 14 times a month during several winter months in the North Atlantic is explored. The expected 100-year return of this data corresponds to the largest expected value during a 100-year period.

3.1 2a Finding the inverse

In this problem, the task is to determine the inverse, $F^{-}1(u; \mu, \beta)$. The expression $F(u; \mu, \beta)$ is described by Equation 17.

$$F(u;\mu,\beta) = -\exp(-\exp(-\frac{x-\mu}{\beta})), x \in \mathbb{R} \tag{17}$$

where $\mu \in \mathbb{R}$ and $\beta > 0$. In order to find the inverse, $u = F(u;\mu,\beta)$ is set and solved for x. This gives the setup according to Equation 18.

$$u = -\exp(-\exp(-\frac{x-\mu}{\beta})), x \in \mathbb{R} \tag{18}$$

This Equation is simplified to Equation 19.

$$F^{-1}(u;\mu,\beta) = -\beta * (\ln(-\ln u)) + \mu \tag{19}$$

If the estimated parameter values are inserted using the *est_gumbel* Matlab function, it becomes the expression in Equation 20, which is the requested inverse function.

$$F^{-1}(u;4.1477,1.4858) = -1.4858 * (\ln(-\ln u)) + 4.1477 \tag{20}$$

3.2 2b Parametric bootstrapped two-sided confidence intervals for the parameters

In this part, the aim was to determine a parametric bootstrapped two-sided 99% confidence interval for the parameters, μ and β based on the provided data on the significant wave-heights. First, the general estimate of the parameters based on all the data assuming it follows a Gumbel distribution was examined. This was done using the given Matlab function *est_gumbel*. The obtained estimates from this function were $\hat{\beta} = 1.4858$ and $\hat{\mu} = 4.1477$.

It was determined that a B-value, i.e. the number of times resampling is done, of 2,000 was sufficient and an n-value, i.e. the number of samples in each resampling, of 582 was appropriate, as it is the number of values given in the data set. An empty vector for each of beta and mu for each iteration of B was then created. A loop was then created where a random sample with replacements of size n was created from the Atlantic data.

The parameter estimates are then calculated based on the data belonging to the Gumbel distribution. Thus, the estimates are created from the bootstrapped samples using the *est_gumbel* function, and the estimates were then saved in the vectors.

When the 2,000 estimates had been extracted, the errors were calculated through Equation 21 for each of the parameters. Simultaneously, they were sorted in ascending order.

$$\begin{aligned} \Delta^*_{b\mu} &= \hat{\mu}^*_b - \hat{\mu} \\ \Delta^*_{b\beta} &= \hat{\beta}^*_b - \hat{\beta} \end{aligned} \tag{21}$$

As the desired confidence level was 99%, the α was set to 0.01, and the error terms were used to determine the confidence intervals together with the initial estimates of the parameters according to Equation 22 for μ, where $\Delta^*_{b\mu}$ is the sorted vector of the error terms. The corresponding idea was used for β.

$$\begin{aligned} LB^{\mu} &= \hat{\mu} - \Delta^*_{b\mu}((1-\alpha/2) * B) \\ UB^{\mu} &= \hat{\mu} - \Delta^*_{b\mu}(\alpha * B/2) \end{aligned} \tag{22}$$

The confidence intervals obtained are: $I^{\mu}_{.01} = (3.983, 4.321)$ and $I^{\beta}_{.01} = (1.363, 1.602)$. The expected values are, as previously mentioned, 4.1477 and 1.4858, respectively. Histograms of the bootstrap distributions can be viewed in Figure 20.

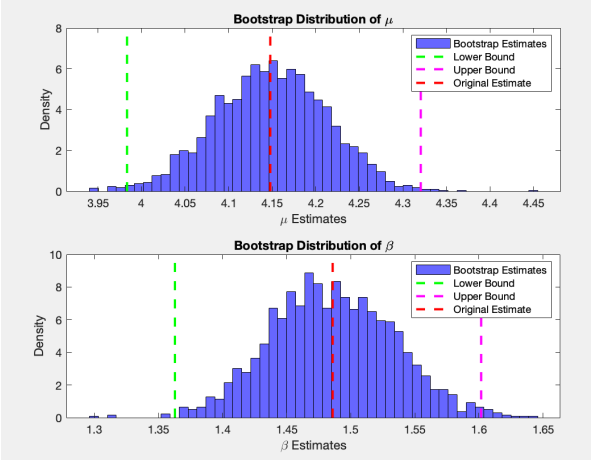

Figure 20: The estimated μ and β parameters and the confidence intervals from the parametric bootstrapping procedure

3.3 2c One-sided parametrically bootstrapped confidence interval for the 100-year return value

In this problem, the one-sided upper-bounded parametrically bootstrapped 99% confidence interval for the 100-year return value is determined. It is known that the T:th return value is given by Equation 23 and that $T = 3 * 14 * 100 = 4200$ for the 100th return value.

$$F^{-1}(1 - \frac{1}{T}; \mu, \beta) \tag{23}$$

The inverse, F^{-1}, has been previously determined by Equation 19 in problem 2a. Alas, the equation can be used by substituting u with $1 - \frac{1}{T}$.

The rough estimate of the 100-year return value, r, can be conducted by using the values of $\hat{\mu}$ and $\hat{\beta}$ determined in 2b and the T value. This gives $F^{-1}(1 - \frac{1}{4200}; \hat{\mu}, \hat{\beta}) = \hat{r} = 16.5436$.

The confidence interval can then be designed by saving the estimated 100-year return according to Equation 23 for each combination of μ and β determined by the bootstrapping in problem 2b, leading to a vector with B=2,000 estimates of the 100-year returns.

These are then subtracted from the estimated value throughout the samples, and the errors are then sorted in ascending order. The upper bound of the 99% confidence interval for the 100-year return, r, is then determined through Equation 24 where $\alpha = 0.01$.

$$UB^r = \hat{r} - \Delta^*_{br}(\alpha * B) \tag{24}$$

From this, the upper level of the confidence interval is determined to be 17.4980, with an expected value of 16.54.

4 Final words

In this assignment, various statistical techniques of studying natural disasters were explored, using a hybrid Markov Chain Monte Carlo and bootstrapping. By leveraging these methods, it was possible to gain an understanding for when important changes in technology and the like occurred.

This assignment highlights the importance of advanced statistical methods in areas such as environmental disaster prevention.

All MATLAB implementations, including the main script (proj3.m) and supporting files, have been submitted via email, and this report has been uploaded in PDF format to CANVAS.

References

[1] Michael Zhu (2017) "Metropolis-Hastings". https://ermongroup.github.io/cs323-notes/probabilistic/mh/