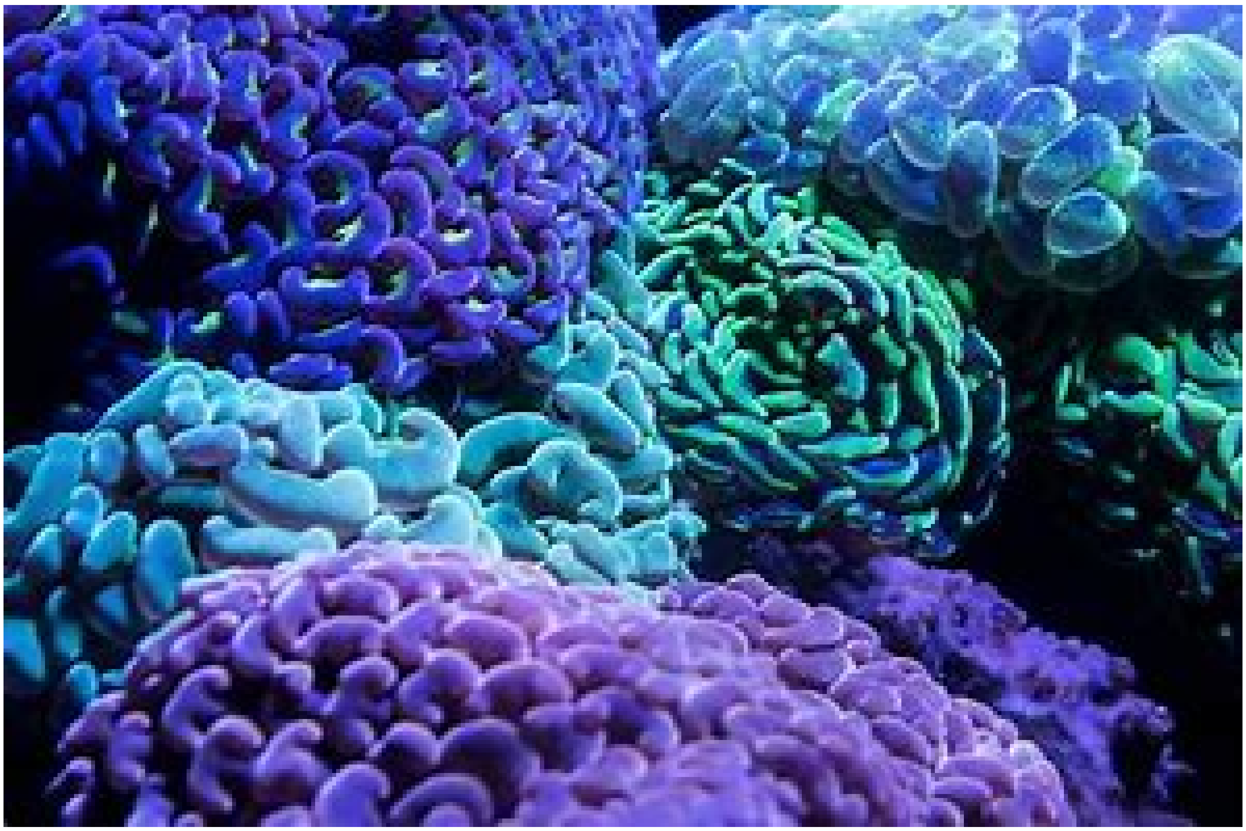
Interim Report

Predicting Algae Bloom spread in the Marlborough Sounds using Bayesian Inference

Luke Pearson



Abstract

Toxic Algae blooms cause shellfish in the affected area to be infected, causing them to carry the toxins that can cause severe illness to humans when they are consumed. When these blooms occur, it causes the shellfish industry in the Marlborough Sounds to shut down collection region-wide out of an abundance of caution, which increases costs. Therefore, they have an interest in predicting Algae blooms, and in particular the size and the spread in an effort to narrow the extent of the shutdown.

For the purposes of this project, the Marlborough sounds have been divided into a set of 386 polygonal approximations of areas of the sounds. The volumes of these polygons have been calculated, and a transport matrix depicting the proportion of particles that move from site i to site j in a particular time interval. (Ross Vennell, 2022)

Our data is recording concentrations of samples retrieved from these sites at discrete time intervals, observed with an error. This data is also observed sparsely.

Due to timing, the actual data is not currently available, so the scope of this project will involve simulating data with randomly generated parameters, and then trying to predict the data by deducing the values of the parameters.

This will be done by constructing a forward in time statistical model of the mass and then using bayesian inference, via a Metropolis-Hastings algorithm to determine the model parameters.

This model will be based upon an exponential growth component during active blooms, a decay component and a transfer component based upon the transport matrix.

Introduction

Bayesian Inference

Bayesian inference is a branch of statistical modeling based upon Bayes' theorem

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)}$$

Where:

- $P(X)$ is the $prior$ distribution
- $P(Y|X)$ is the $likelihood$ distribution
- $P(X|Y)$ is the $posterior$ distribution

Where instead of the usual approach where the parameters X are fixed and the data Y is variable to determine the likelihood, the data is used as the fixed variable to determine the distribution of the unknown parameters. Since the data is fixed, $P(Y)$ will be constant.

For the purposes of this project, we can use the likelihood for the Concentrations and an educated construction of prior distributions to determine the distribution of the parameters, or $P(Parameters|Concentration) \propto P(Concentration|Parameters) \times P(Parameters)$.

However, due to the complex nature of the likelihood in this case $\left(Where\ C_{it} \sim N\left(log\left(\delta + \frac{M_{it}}{V_i}\right), \frac{1}{\sqrt{\tau}}^2\right)\right)$, deriving the posterior distribution mathematically provides a result that is not useful. Therefore, an analytical method of approximating these distributions is needed.

Metropolis-Hastings

The Metropolis-Hastings algorithm is a Monte Carlo Markov Chain sampling method that can be used to approximate a distribution that is otherwise hard to sample from. This method utilises the bayesian inference ideas above, but also uses a $proposal$ distribution to control the chain.

General Method

For each parameter θ, the algorithm generates a candidate value θ^{*} from a proposal distribution $q\left(\theta^{*}|\theta_{s-1}\right)$ that is dependent on the previous value for θ. It then calculates a ratio of

$$r = \frac{P\left(Y|\theta^{*}\right) \times P\left(\theta^{*}\right) \times P\left(\theta_{s-1}|\theta^{*}\right)}{P\left(Y|\theta_{s-1}\right) \times P\left(\theta_{s-1}\right) \times q\left(\theta^{*}|\theta_{s-1}\right)}$$ with:

- $P(Y|\theta)$ being calculated from the $likelihood$ distribution
- $P(Y)$ being calculated from the $prior$ distribution
- $P\left(\theta_{a}|\theta_{b}\right)$ being calculated from the $likelihood$ distribution

And then the probability of accepting θ^{*} as θ_{s} is r, with the probability of accepting θ_{s-1} as θ_{s} is $1 - r$

Application

Because our likelihood is different for each site i and time t, the total likelihood $P(Y|\theta)$ is actually $\prod_{i=1}^{n}\left(\prod_{t=1}^{T}\left(P\left(Y_{it}|\theta\right)\right)\right)$. Since our simulation will involve 10 sites and 100 times, this will be the product of 10000 probabilities; to avoid our code rounding the value to zero, $\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\left(log\left(P\left(Y_{it}|\theta\right)\right)\right)\right)$ will be more useful. Therefore, our total ratio is

$$ratio = log\left(P\left(Y|\theta^{*}\right)\right) + log\left(P\left(\theta^{*}\right)\right) + log\left(P\left(\theta_{s-1}|\theta^{*}\right)\right) - log\left(P\left(Y|\theta_{s-1}\right)\right) + log\left(P\left(\theta_{s-1}\right)\right) + log\left(P\left(\theta^{*}|\theta_{s-1}\right)\right)$$

And then the probability of accepting θ^{*} as θ_{s} is e^{ratio}, with the probability of accepting θ_{s-1} as θ_{s} is $1 - e^{ratio}$.

Transportation Matrix

In preparation for this project, Ross Vennell has constructed a transportation matrix, P, to represent the proportion of mass that moves between sites across the Marlborough Sounds.

This matrix was constructed by building 386 polygons to approximate the geometry of the sounds, with the location of the sites labeled in $Figure\ X$. Then, Vennell simulated a set of particles in each site and determined the number that had traveled to each other site after a single time event.

P is an $n \times n$ matrix where P_{ij} is equal to the proportion of particles in site i that travels to site j, with two main properties:

- $0 \leq P_{ij} \leq 1\ for\ all\ i,\ j$
- $\sum_{j=1}^{n} P_{ij} = 1\ for\ all\ i$

$Figure\ X + 1$ is a heatmap of P. P is a very sparse matrix, with 97% of values being 0, and P is also close to a diagonal matrix with the average value of P_{ii} being 0.935. As a result, we shouldn't expect algae spread amongst a lot of sites such that a bloom should stay relatively contained to the sites in the near vicinity.

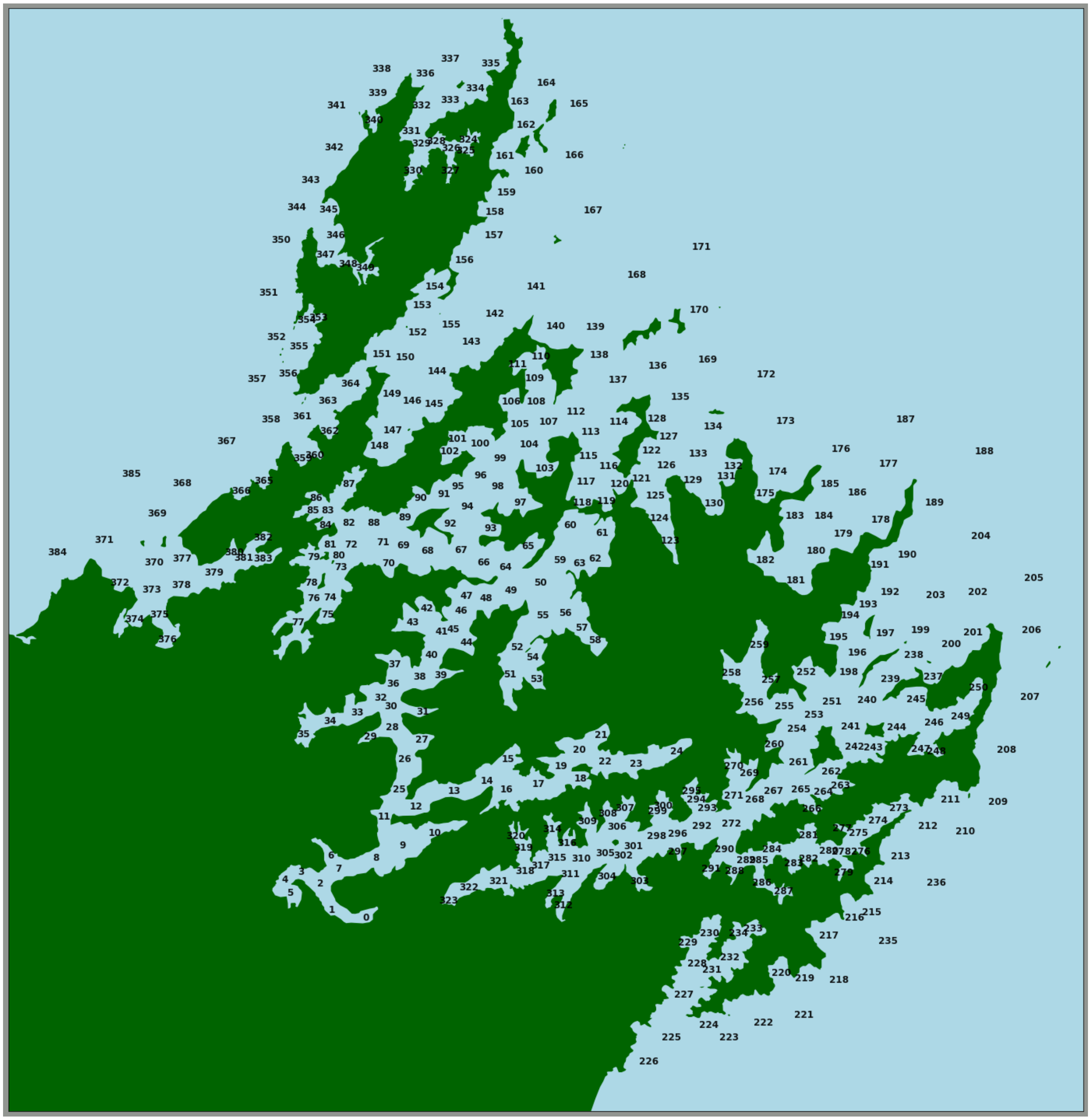
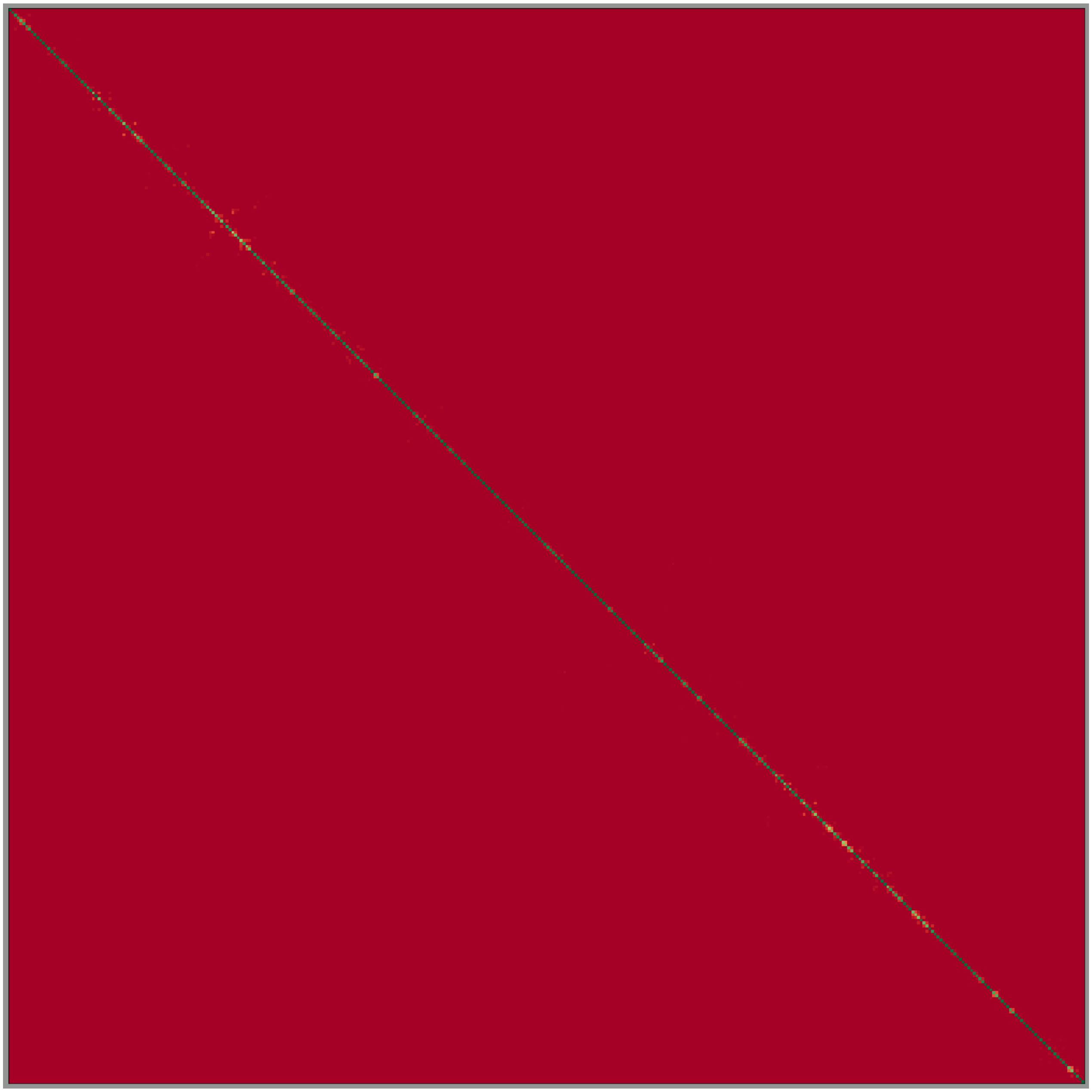


Figure X: Map of Marlborough Sounds with locations of sites labeled



$Figure\ X + 1\!:\ Heatmap\ of\ P$

Our Model

Mass

Mass is observed as:

- n Sites and T discrete Times
- Growth parameter α
- Decay parameter β
- Dispersion Matrix P where p_{ij} is the proportion of Mass at site i that travels to site j (Such that $\sum_{j=1}^{n} p_{ij} = 1$ for all $i \in \{1,..., n\}$)
- N events
- Each of $k \in \{1,..., N\}$ events has:
    - Starting site I_k with uniform probability $\frac{1}{n}$
    - Starting time $T_k \in \{1,..., T\}$
    - Length of time W_k
    - Starting size S_k

Where the equation for $M_{i,t}$ is:

$$M_{i,t} = \alpha A_{i,t} M_{i,t-1} + \beta\left(\sum_{j=1}^{n} p_{i,j} M_{j,t-1}\right) + B_{i,t}$$

Where

$$A_{i,t} = I\left(\sum_{k=1}^{n} I\left(I_k = i\right) I\left(T_k \le t \le T_k + W_k\right) > 0\right)$$

$$B_{i,t} = \sum_{k=1}^{n} I\left(I_k = i\right) I\left(T_k = t\right) S_k$$

Concentration

Concentration is calculated from Mass where:

- Each site $i \in \{1,\ldots,n\}$ has volume V_i
- δ is a concentration offset of approximately 0^+
- Precision parameter τ

Where the equation for $C_{i,t}$ is

$$log\left(Y_{i,t}\right) \sim N\left(log\left(\delta + \frac{M_{i,t}}{V_i}\right), \frac{1}{\tau}^2\right)$$

Parameters

Known Parameters:

- Number of sites, n
- Time period, T
- Dispersion Matrix, P
- Number of events, N
- Volumes, V_i
- Concentration offset, δ

Parameters to be estimated

- Growth Parameter, α
- Decay Parameter, β
- Event site, I_k
- Event Starting time, T_k
- Event time length, W_k
- Event size, S_k
- Precision τ

Priors

$\alpha \sim Gamma\left(a_{\alpha}, b_{\alpha}\right)$

$\beta \sim Beta\left(a_{\beta}, b_{\beta}\right)$

$I_{k} \sim DiscreteUniform(1,\ n)$

$T_{k} \sim DiscreteUniform(1,\ T)$

$W_{k} \sim NegBinomial\left(r_{W}, p_{W}\right)$

$S_{k} \sim Gamma\left(a_{S}, b_{S}\right)$

$\tau \sim Gamma\left(a_{\tau}, b_{\tau}\right)$

Where:

$a_{\beta}, a_{S}, a_{\tau}, b = 1$

$a_{\alpha} = 2$

$r_{W} = 6$

$p_{W} = 0.75$

Pseudocode

FindM$_{New}$

Inputs:

- M
- α
- β
- P
- A
- B

$growth = \alpha \times (AM)$

$transport = \beta \times (P \cdot M)$

$M_{New} = B + growth + transport$

Outputs:

- M_{New}

RunSimulation

Inputs:

- n
- T
- P

$\alpha \sim Gamma(2,\ 1)$

$\beta \sim Beta(4,\ 3)$

$N \sim DiscreteUniform\left(1,\ \frac{T}{10}\right)$

$For\ x\ in\ 1{:}\ N$:

$i \sim DiscreteUniform(1,\ n)$

$t \sim DiscreteUniform(1,\ T)$

$w \sim NegBinom(6,\ 0.75)\ +\ 1$

$s \sim Gamma(1,\ 1)$

$A_{t:min(t+w,T),\ i} = 1$

$B_{t,\ i} = s$

$\left[M_{0,\ 0}, ...,\ M_{n,\ 0}\right] = 0$

$For\ t\ in\ 1{:}\ T$:

$M_{New} = FindM_{New}\left(\left[M_{0,\ t-1}, ...,\ M_{n,\ t-1}\right],\ \alpha,\ \beta,\ P_{i:n,\ i:n},\ \left[A_{t,\ 1}, ..., A_{t,\ n}\right],\ \left[B_{t,\ 1}, ..., B_{t,\ n}\right]\right)$

$\left[M_{0,\ t}, ...,\ M_{n,\ t}\right] = M_{New}$

Outputs:

- $\left[\left[M_{0,\ 1}, ...,\ M_{n,\ 1}\right], ..., \left[M_{0,\ T}, ...,\ M_{n,\ T}\right]\right]$
- α
- β
- N
- $\left[I_0, ...,\ I_N\right]$
- $\left[T_0, ...,\ T_N\right]$
- $\left[W_0, ...,\ W_N\right]$
- $\left[S_0, ...,\ S_N\right]$

GetSample

Inputs:

- n
- T
- P
- α
- β
- N
- $\left[I_1, \dots, I_N\right]$
- $\left[t_1, \dots, t_N\right]$
- $\left[W_1, \dots, W_N\right]$
- $\left[s_1, \dots, s_N\right]$

$For\ x\ in\ 1: N:$

$$A_{t_x : min(t_x + W_x, t), I_x} = 1$$

$$B_{t, i} = s_x$$

$$\left[M_{0,0}, \dots, M_{n,0}\right] = 0$$

$For\ t\ in\ 1: T:$

$$M_{New} = FindM_{New}\left(\left[M_{0,t-1}, \dots, M_{n,t-1}\right], \alpha, \beta, P_{i:n, i:n}, \left[A_{t,1}, \dots, A_{t,n}\right], \left[B_{t,1}, \dots, B_{t,n}\right]\right)$$

$$\left[M_{0,t}, \dots, M_{n,t}\right] = M_{New}$$

Outputs:

- $\left[\left[M_{0,1}, \dots, M_{n,1}\right], \dots, \left[M_{0,T}, \dots, M_{n,T}\right]\right]$

ConvertConcentration

Inputs:

- n
- T
- M
- V
- δ

$For\ i\ in\ 1: n$:

$$C_{i,\,1:T} = \frac{M_{i,\,1:T}}{V_i}$$

$$Concentration = log(C + \delta)$$

Outputs:

- $Concentration$

ConvertY

Inputs:

- C
- τ

$$Y_{log} \sim N\left(C,\ \frac{1}{\sqrt{\tau}}\right)$$

Outputs:

- Y_{log}

Metropolis-Hastings Algorithm

Inputs:

- P

$n = 10$

$T = 100$

$\delta = 1^{-10}$

$\tau_{True} \sim Gamma(1, 1)$

$\left[V_1, ..., V_n\right] \sim NegBinom(1, 1) + 1$

$\left[M_{True}, \alpha_{True}, \beta_{True}, N_{True}, I_{True}, T_{True}, W_{True}, S_{True}\right] = runSimulation(n, T, P)$

$C_{True} = ConvertCalculation\left(n, T, M_{True}, V, \delta\right)$

$Y_{True} = ConvertY\left(C_{True}, \tau_{True}\right)$

$\alpha_0 \sim Gamma(2, 1)$

$\beta_0 \sim Beta(1, 1)$

$N = N_{True}$

$\left[I_{0,1}, ..., I_{0,N}\right] \sim DiscreteUniform(1, n)$

$\left[t_{1,1}, ..., t_{0,N}\right] \sim DiscreteUniform(1, T)$

$\left[w_{0,1}, ..., w_{0,N}\right] \sim NegBinom(6, 0.75)$

$\left[s_{0,1}, ..., s_{0,N}\right] \sim Gamma(1, 1)$

$\tau_0 \sim Gamma(1, 1)$

$\phi = \left[\alpha_0, \beta_0, N, \left[I_{0,1}, ..., I_{0,N}\right], \left[t_{0,1}, ..., t_{0,N}\right], \left[w_{0,1}, ..., w_{0,N}\right], \left[s_{0,1}, ..., s_{0,N}\right], \tau_0\right]$

$For\ each\ parameter_0\ in\ \phi:$

$\quad Array_{parameter} = parameter_0$

$\quad Accept_{parameter} = 0$

$For\ s\ in\ 1: 10000:$

Alpha

$$\alpha_{Candidate} \sim N\left(\alpha_{Candidate}, \frac{1}{2}^2\right)$$

$$M_{Candidate} = GetSample\left(n, T, \alpha_{Candidate}, \phi_\beta, \phi_N, \phi_I, \phi_T, \phi_W, \phi_S\right)$$

$$C_{Candidate} = ConvertConcentration\left(n, T, M_{Candidate}, V, \delta\right)$$

$$M_{Current} = GetSample\left(n, T, \phi_\alpha, \phi_\beta, \phi_N, \phi_I, \phi_T, \phi_W, \phi_S\right)$$

$$C_{Current} = ConvertConcentration\left(n, T, M_{Current}, V, \delta\right)$$

$$p_{prior_{Candidate}} = Gamma(2, 1).logPDF\left(\alpha_{Candidate}\right)$$

$$p_{prior_{Current}} = Gamma(2, 1).logPDF\left(\phi_\alpha\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = N\left(\phi_1, \frac{1}{2}^2\right).logPDF\left(\alpha_{Candidate}\right)$$

$$p_{Q_{Current}} = N\left(\alpha_{Candidate}, \frac{1}{2}^2\right).logPDF\left(\phi_\alpha\right)$$

$$ratio_\alpha = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}}, 1\right)$$

$$u \sim U(0, 1)$$

$$if\ ratio > log(u):$$

$$\phi_{\alpha} = \alpha_{Candidate}$$

$$Accept_{\alpha} = Accept_{\alpha} + 1$$

$$Likelihood_{\alpha_s} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

$else$:

$$Likelihood_{\alpha_s} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

Beta

$$\beta_{Candidate} \sim Beta(1, 1)$$

$$M_{Candidate} = GetSample\left(n, T, \phi_{\alpha}, \beta_{Candidate}, \phi_{N}, \phi_{I}, \phi_{T}, \phi_{W}, \phi_{S}\right)$$

$$C_{Candidate} = ConvertConcentration\left(n, T, M_{Candidate}, V, \delta\right)$$

$$M_{Current} = GetSample\left(n, T, \phi_{\alpha}, \phi_{\beta}, \phi_{N}, \phi_{I}, \phi_{T}, \phi_{W}, \phi_{S}\right)$$

$$C_{Current} = ConvertConcentration\left(n, T, M_{Current}, V, \delta\right)$$

$$p_{prior_{Candidate}} = Beta(1, 1). logPDF\left(\beta_{Candidate}\right)$$

$$p_{prior_{Current}} = Beta(1, 1). logPDF\left(\phi_{2}\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = Beta(1, 1).logPDF\left(\beta_{Candidate}\right)$$

$$p_{Q_{Current}} = Beta(1, 1).logPDF\left(\phi_{2}\right)$$

$$ratio_{\beta} = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}}, 1\right)$$

$$u \sim U(0, 1)$$

$if\ ratio > u$:

$$\phi_{\beta} = \beta_{Candidate}$$

$$Accept_{\beta} = Accept_{\beta} + 1$$

$$Likelihood_{\beta_s} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$else$:

$$Likelihood_{\beta_s} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$for\ k\ in\ 1:n$:

$$I_{k_{Candidate}} \sim DiscreteUniform(1, n)$$

$$I_{Candidate} = \phi_{3}$$

$$I_{Candidate_k} = I_{k_{Candidate}}$$

$$M_{Candidate} = GetSample\left(n, T, \phi_{\alpha}, \phi_{\beta}, \phi_{N}, I_{Candidate}, \phi_{T}, \phi_{W}, \phi_{S}\right)$$

$$C_{Candidate} = ConvertConcentration\left(n, T, M_{Candidate}, V, \delta\right)$$

$$M_{Current} = GetSample\left(n, T, \phi_{\alpha}, \phi_{\beta}, \phi_{N}, \phi_{I}, \phi_{T}, \phi_{W}, \phi_{S}\right)$$

$$C_{Current} = ConvertConcentration\left(n, T, M_{Current}, V, \delta\right)$$

$$p_{prior_{Candidate}} = DiscreteUniform(1,\ n).logPMF\left(I_{Candidate_k}\right)$$

$$p_{prior_{Current}} = DiscreteUniform(1,\ n).logPMF\left(\phi_{I_k}\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}},\ \frac{1}{\sqrt{\phi_\tau}}^{2}\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}},\ \frac{1}{\sqrt{\phi_\tau}}^{2}\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = DiscreteUniform(1,\ n).logPMF\left(I_{Candidate_k}\right)$$

$$p_{Q_{Current}} DiscreteUniform(1,\ n).logPMF\left(\phi_{I_k}\right)$$

$$ratio_{I_k} = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}},\ 1\right)$$

$$u \sim U(0,\ 1)$$

$$if\ ratio > log(u):$$

$$\phi_I = I_{Candidate}$$

$$Accept_I = Accept_I + \frac{1}{N}$$

$$Likelihood_{I_{s,k}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}},\ \frac{1}{\sqrt{\phi_\tau}}^{2}\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$else:$$

$$Likelihood_{I_{s,k}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}},\ \frac{1}{\sqrt{\phi_\tau}}^{2}\right).logPDF\left(Y_{True_{it}}\right)\right)$$

T

$for\ k\ in\ 1{:}\ n{:}$

$$T_{k_{Candidate}} \sim DiscreteUniform(1,\ T)$$

$$T_{Candidate} = \phi_T$$

$$T_{Candidate_k} = T_{k_{Candidate}}$$

$$M_{Candidate} = GetSample\left(n,\ T,\ \phi_{\alpha},\ \phi_{\beta},\ \phi_{N},\ \phi_{I},\ T_{Candidate},\ \phi_{W},\ \phi_{S}\right)$$

$$C_{Candidate} = ConvertConcentration\left(n,\ T,\ M_{Candidate},\ V,\ \delta\right)$$

$$M_{Current} = GetSample\left(n,\ T,\ \phi_{\alpha},\ \phi_{\beta},\ \phi_{N},\ \phi_{I},\ \phi_{T},\ \phi_{W},\ \phi_{S}\right)$$

$$C_{Current} = ConvertConcentration\left(n,\ T,\ M_{Current},\ V,\ \delta\right)$$

$$p_{prior_{Candidate}} = DiscreteUniform(1,\ T).\ logPMF\left(T_{Candidate_k}\right)$$

$$p_{prior_{Current}} = DiscreteUniform(1,\ T).\ logPMF\left(\phi_{T_k}\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}},\ \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right).\ logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}},\ \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right).\ logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = DiscreteUniform(1,\ T).\ logPMF\left(T_{Candidate_k}\right)$$

$$p_{Q_{Current}} = DiscreteUniform(1,\ T).\ logPMF\left(\phi_{4_k}\right)$$

$$ratio_{T_k} = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}}, 1\right)$$

$$u \sim U(0,\ 1)$$

$$if\ ratio > log(u):$$

$$\phi_T = T_{Candidate}$$

$$Accept_T = Accept_T + \frac{1}{N}$$

$$Likelihood_{T_{s,k}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$else:$$

$$Likelihood_{T_{s,k}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

W

$$for\ k\ in\ 1{:}\ n:$$

$$W_{k_{Candidate}} \sim Binom\left(T, \frac{1+\phi_{W_k}}{2+T}\right)$$

$$W_{Candidate} = \phi_W$$

$$W_{Candidate_k} = W_{k_{Candidate}}$$

$$M_{Candidate} = GetSample\left(n, T, \phi_\alpha, \phi_\beta, \phi_N, \phi_I, \phi_T, W_{Candidate}, \phi_S\right)$$

$$C_{Candidate} = ConvertConcentration\left(n, T, M_{Candidate}, V, \delta\right)$$

$$M_{Current} = GetSample\left(n, T, \phi_\alpha, \phi_\beta, \phi_N, \phi_I, \phi_T, \phi_W, \phi_S\right)$$

$$C_{Current} = ConvertConcentration\left(n, T, M_{Current}, V, \delta\right)$$

$$p_{prior_{Candidate}} = DiscreteUniform(1, T).logPMF\left(W_{Candidate_k}\right)$$

$$p_{prior_{Current}} = DiscreteUniform(1, T).logPMF\left(\phi_{W_k}\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = Binom\left(T, \frac{1+\phi_{W_k}}{2+T}\right).logPMF\left(W_{Candidate_k}\right)$$

$$p_{Q_{Current}} Binom\left(T, \frac{1+W_{Candidate_k}}{2+T}\right).logPMF\left(\phi_{W_k}\right)$$

$$ratio_{W_k} = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}}, 1\right)$$

$$u \sim U(0, 1)$$

$if\ ratio > log(u)$:

$$\phi_W = W_{Candidate}$$

$$Accept_W = Accept_W + \frac{1}{N}$$

$$Likelihood_{W_{s,k}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$else$:

$$Likelihood_{W_{s,k}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_\tau}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

S

$for\ k\ in\ 1: n:$

$$S_{k_{Candidate}} \sim N\left(\phi_{S_k}, 1^2\right)$$

$$S_{Candidate} = \phi_S$$

$$S_{Candidate_k} = S_{k_{Candidate}}$$

$$M_{Candidate} = GetSample\left(n, T, \phi_{\alpha}, \phi_{\beta}, \phi_{N}, \phi_{I}, \phi_{T}, \phi_{W}, S_{Candidate}\right)$$

$$C_{Candidate} = ConvertConcentration\left(n, T, M_{Candidate}, V, \delta\right)$$

$$M_{Current} = GetSample\left(n, T, \phi_{\alpha}, \phi_{\beta}, \phi_{N}, \phi_{I}, \phi_{T}, \phi_{W}, \phi_{S}\right)$$

$$C_{Current} = ConvertConcentration\left(n, T, M_{Current}, V, \delta\right)$$

$$p_{prior_{Candidate}} = Gamma(1, 1).logPDF\left(S_{Candidate_k}\right)$$

$$p_{prior_{Current}} = Gamma(1, 1).logPDF\left(\phi_{S_k}\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^2\right).logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = N\left(\phi_{S_k}, 1^2\right).logPDF\left(S_{Candidate_k}\right)$$

$$p_{Q_{Current}} N\left(S_{Candidate_k}, 1^2\right).logPDF\left(\phi_{S_k}\right)$$

$$ratio_{S_k} = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}}, 1\right)$$

$$u \sim U(0, 1)$$

$$if\ ratio > log(u):$$

$$\phi_S = S_{Candidate}$$

$$Accept_S = Accept_S + \frac{1}{N}$$

$$Likelihood_{S_{s,k}} = \sum_{y=1}^{T}\left(\sum_{x=1}^{N} N\left(C_{Candidate}, \frac{1}{\sqrt{\phi_\tau}}^2\right). logPDF\left(Y_{True}\right)\right)$$

$$else:$$

$$Likelihood_{S_{s,k}} = \sum_{y=1}^{T}\left(\sum_{x=1}^{N} N\left(C_{Current}, \frac{1}{\sqrt{\phi_\tau}}^2\right). logPDF\left(Y_{True}\right)\right)$$

Tau

$$\tau_{Candidate} \sim N\left(\phi_\tau, \frac{1}{2}^2\right)$$

$$M_{Candidate} = GetSample\left(n, T, \phi_\alpha, \phi_\beta, \phi_N, \phi_I, \phi_T, \phi_W, \phi_S\right)$$

$$C_{Candidate} = ConvertConcentration\left(n, T, M_{Candidate}, V, \delta\right)$$

$$M_{Current} = GetSample\left(n, T, \phi_\alpha, \phi_\beta, \phi_N, \phi_I, \phi_T, \phi_W, \phi_S\right)$$

$$C_{Current} = ConvertConcentration\left(n, T, M_{Current}, V, \delta\right)$$

$$p_{prior_{Candidate}} = Gamma(1, 1). logPDF\left(\tau_{Candidate}\right)$$

$$p_{prior_{Current}} = Gamma(1, 1). logPDF\left(\phi_\tau\right)$$

$$p_{Likelihood_{Candidate}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\tau_{Candidate}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Likelihood_{Current}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

$$p_{Q_{Candidate}} = N\left(\phi_{\tau}, \frac{1}{2}^{2}\right). logPDF\left(\tau_{Candidate}\right)$$

$$p_{Q_{Current}} N\left(\tau_{Candidate}, \frac{1}{2}^{2}\right). logPDF\left(\phi_{\tau}\right)$$

$$ratio_{\tau} = min\left(p_{prior_{Candidate}} + p_{Likelihood_{Candidate}} - p_{prior_{Current}} - p_{Likelihood_{Current}} + p_{Q_{Current}} - p_{Q_{Candidate}}, 1\right)$$

$$u \sim U(0,\ 1)$$

$if\ ratio > log(u)$:

$$\phi_{\tau} = \tau_{Candidate}$$

$$Accept_{\tau} = Accept_{\tau} + 1$$

$$Likelihood_{\tau_{s}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Candidate_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

$else$:

$$Likelihood_{\tau_{s}} = \sum_{t=1}^{T}\left(\sum_{i=1}^{n} N\left(C_{Current_{it}}, \frac{1}{\sqrt{\phi_{\tau}}}^{2}\right). logPDF\left(Y_{True_{it}}\right)\right)$$

Proposal Distributions

$$\alpha^{*} \sim N\left(\phi_{1}, \frac{1}{2}^{2}\right)$$

$$\beta^{*} \sim Beta(1,\ 1)$$

$$I_{k}^{*} \sim DiscreteUniform(1,\ n)$$

$$T_{k}^{*} \sim DiscreteUniform(1,\ T)$$

$$W_{k}^{*} \sim Binom\left(T,\ \frac{1 + \phi_{W_{k}}}{2 + T}\right)$$

$$S_{k}^{*} \sim N\left(\phi_{S_{k}},\ 1^{2}\right)$$

$$\tau^{*} \sim N\left(\phi_{\tau}, \frac{1}{2}^{2}\right)$$

Acceptance Probabilities

Alpha

$$p_{prior_{Candidate}} = log\left(\alpha_{Candidate}\right) - \alpha_{Candidate}$$

$$p_{prior_{Current}} = log\left(\alpha_{Current}\right) - \alpha_{Current}$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Candidate_{it}}\right)^{2}\right)$$

$$p_{Likelihood_{Current}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^{2}\right)$$

$$p_{Q_{Candidate}} = -2\left(\alpha_{Candidate} - \alpha_{Current}\right)^{2}$$

$$p_{Q_{Candidate}} = -2\left(\alpha_{Current} - \alpha_{Candidate}\right)^{2}$$

$$ratio_{log} = log\left(\frac{\alpha_{Candidate}}{\alpha_{Current}}\right) - \alpha_{Candidate} + \alpha_{Current} + \frac{\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^{2} - \left(Y_{True_{it}} - C_{Candidate_{it}}\right)^{2}\right)$$

$$P_{Acceptance}(\alpha) = e^{ratio_{log}}$$

Beta

$$p_{prior_{Candidate}} = 0$$

$$p_{prior_{Current}} = 0$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$p_{Likelihood_{Current}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2\right)$$

$$p_{Q_{Candidate}} = 0$$

$$p_{Q_{Candidate}} = 0$$

$$ratio_{log} = \frac{\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2 - \left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$P_{Acceptance}(\beta) = e^{ratio_{log}}$$

I_k

$$p_{prior_{Candidate}} = -\log(n)$$

$$p_{prior_{Current}} = -\log(n)$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$p_{Likelihood_{Current}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2\right)$$

$$p_{Q_{Candidate}} = -\log(n)$$

$$p_{Q_{Candidate}} = -\log(n)$$

$$ratio_{log} = \frac{\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2 - \left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$P_{Acceptance}\left(I_k\right) = e^{ratio_{log}}$$

T_k

$$p_{prior_{Candidate}} = -\log(T)$$

$$p_{prior_{Current}} = -\log(T)$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$p_{Likelihood_{Current}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2\right)$$

$$p_{Q_{Candidate}} = -\log(T)$$

$$p_{Q_{Candidate}} = -\log(T)$$

$$ratio_{log} = \frac{\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2 - \left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$P_{Acceptance}\left(T_k\right) = e^{ratio_{log}}$$

W_k

$$p_{prior_{Candidate}} = log\left(\left(W_{Candidate_k} + 5\right)C\left(W_{Candidate_k}\right)\right) + 6log\left(\frac{3}{4}\right) - W_{Candidate_k} log(4)$$

$$p_{prior_{Current}} = log\left(\left(W_{Current_k} + 5\right)C\left(W_{Current_k}\right)\right) + 6log\left(\frac{3}{4}\right) - W_{Current_k} log(4)$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$p_{Likelihood_{Current}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2\right)$$

$$p_{Q_{Candidate}} = log\left((T)C\left(W_{Candidate_k}\right)\right) + W_{Candidate_k} log\left(W_{Current_k} + 1\right) + \left(T - W_{Candidate_k}\right)log\left(T + 1 - W_{Current_k}\right) + Tlog(T + 2)$$

$$p_{Q_{Candidate}} = log\left((T)C\left(W_{Current_k}\right)\right) + W_{Current_k} log\left(W_{Candidate_k} + 1\right) + \left(T - W_{Current_k}\right)log\left(T + 1 - W_{Candidate_k}\right) + Tlog(T + 2)$$

$$\begin{aligned} ratio_{log} = {} & log\left(\left(W_{Candidate_k} + 5\right)C\left(W_{Candidate_k}\right)\right) + \left(W_{Current_k} - W_{Candidate_k}\right)log(4) - log\left(\left(W_{Current_k} + 5\right)C\left(W_{Current_k}\right)\right) \\ & + \frac{\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2 - \left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right) + log\left((T)C\left(W_{Candidate_k}\right)\right) + W_{Candidate_k} log\left(W_{Current_k} + 1\right) \\ & + \left(T - W_{Candidate_k}\right)log\left(T + 1 - W_{Current_k}\right) - log\left((T)C\left(W_{Current_k}\right)\right) - W_{Current_k} log\left(W_{Candidate_k} + 1\right) - \left(T - W_{Current_k}\right)log\left(T + 1 - W_{Candidate_k}\right) \end{aligned}$$

$$P_{Acceptance}\left(W_k\right) = e^{ratio_{log}}$$

S_k

$$p_{prior_{Candidate}} = -S_{Candidate_k}$$

$$p_{prior_{Current}} = -S_{Current_k}$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$p_{Likelihood_{Current}} = \frac{-\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2\right)$$

$$p_{Q_{Candidate}} = -\frac{1}{2}\left(S_{Candidate_k} - S_{Current_k}\right)^2$$

$$p_{Q_{Candidate}} = -\frac{1}{2}\left(S_{Current_k} - S_{Candidate_k}\right)^2$$

$$ratio_{log} = -S_{Candidate_k} + S_{Current_k} + \frac{\tau}{2}\sum_{t=1}^{T}\left(\sum_{i=1}^{n}\left(Y_{True_{it}} - C_{Current_{it}}\right)^2 - \left(Y_{True_{it}} - C_{Candidate_{it}}\right)^2\right)$$

$$P_{Acceptance}\left(S_k\right) = e^{ratio_{log}}$$

Tau

$$p_{prior_{Candidate}} = log\left(\tau_{Candidate}\right) - \tau_{Candidate}$$

$$p_{prior_{Current}} = log\left(\tau_{Current}\right) - \tau_{Current}$$

$$p_{Likelihood_{Candidate}} = \frac{-\tau_{Candidate}}{2} \sum_{t=1}^{T} \left(\sum_{i=1}^{n} \left(Y_{True_{it}} - C_{Candidate_{it}} \right) \right)^2$$

$$p_{Likelihood_{Current}} = \frac{-\tau_{Current}}{2} \sum_{t=1}^{T} \left(\sum_{i=1}^{n} \left(Y_{True_{it}} - C_{Current_{it}} \right) \right)^2$$

$$p_{Q_{Candidate}} = -2\left(\tau_{Candidate} - \tau_{Current}\right)^2$$

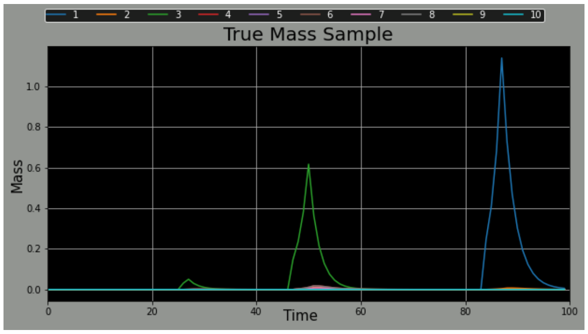
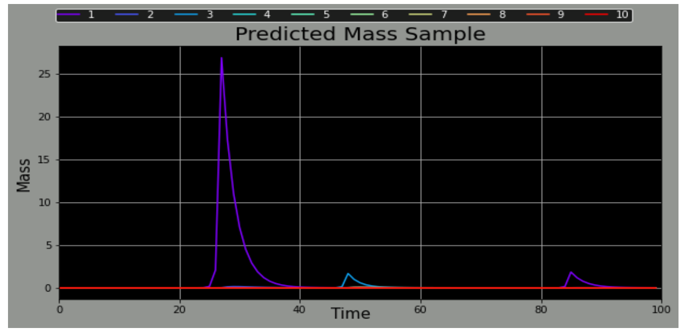
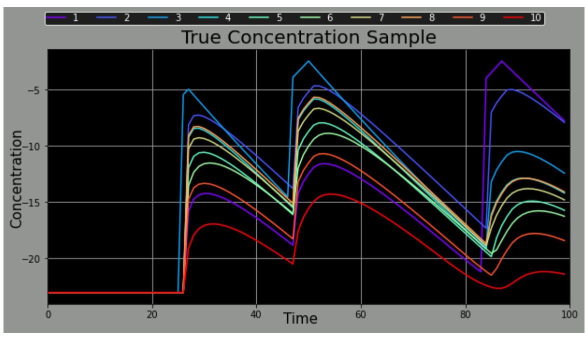
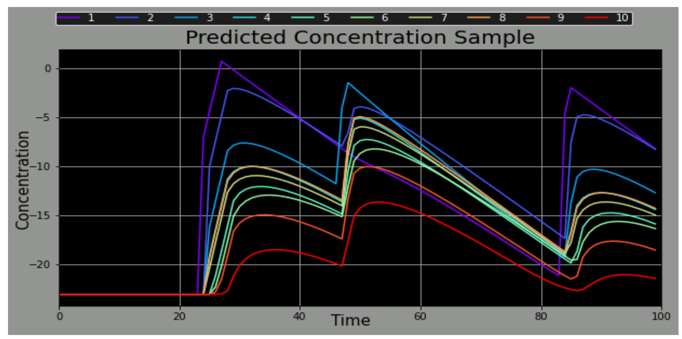
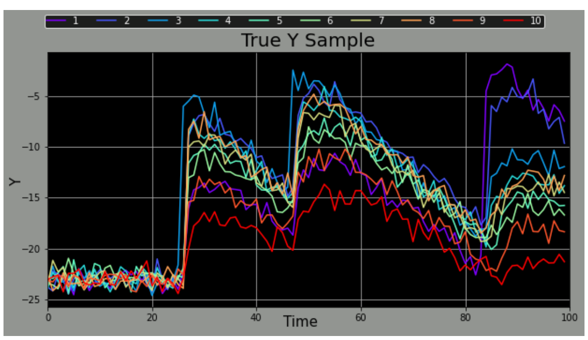
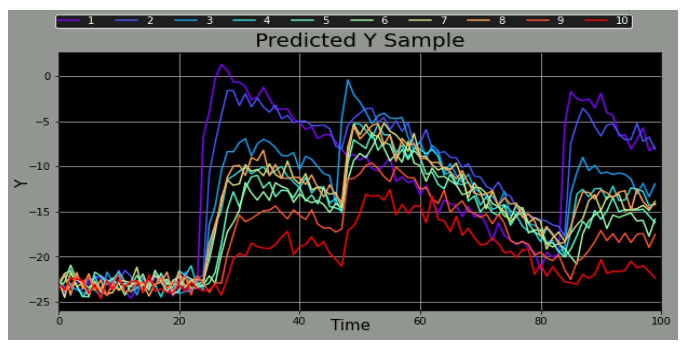
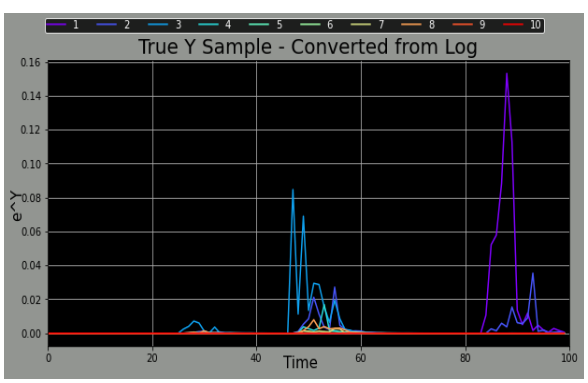
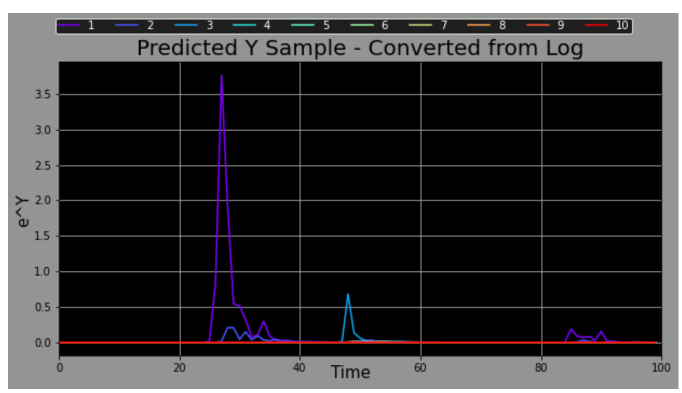
$$p_{Q_{Candidate}} = -2\left(\tau_{Current} - \tau_{Candidate}\right)^2$$

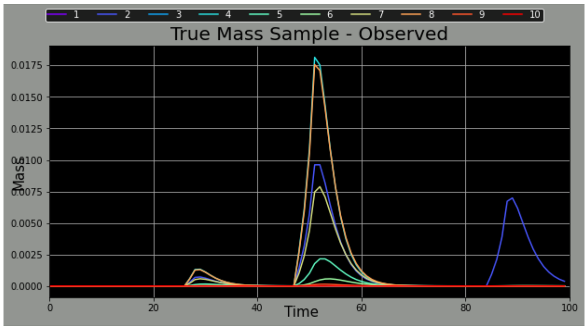
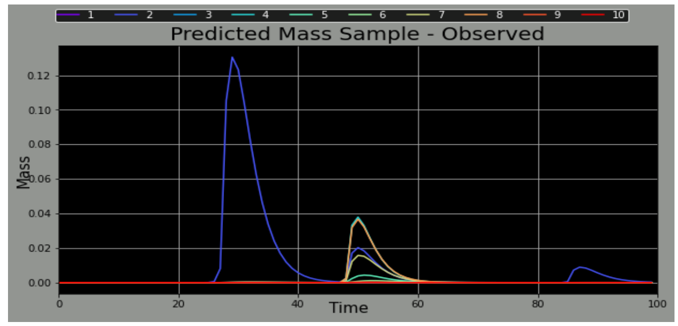
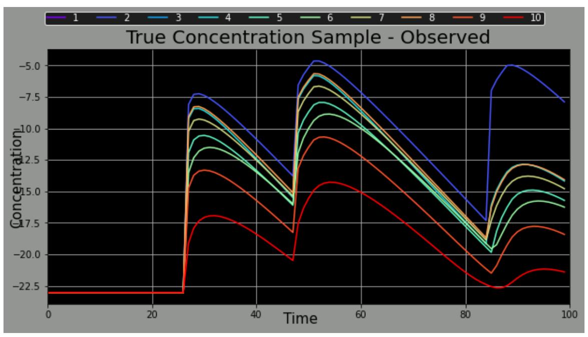
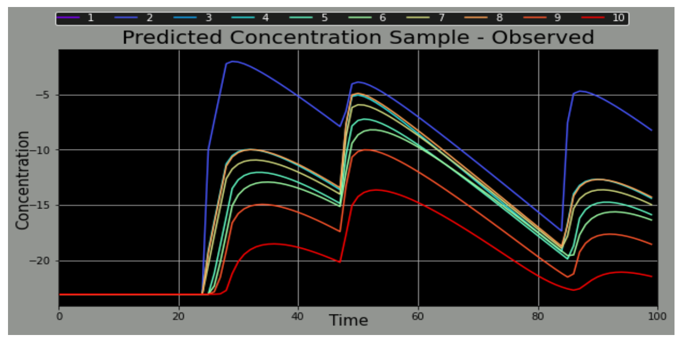
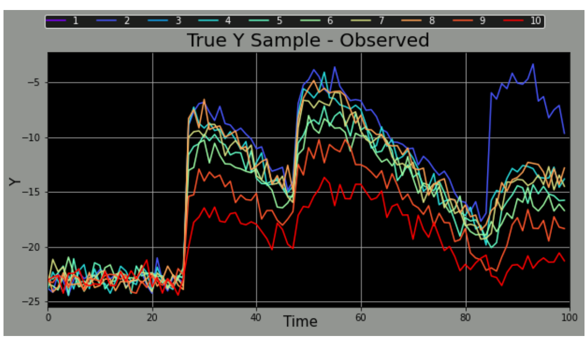
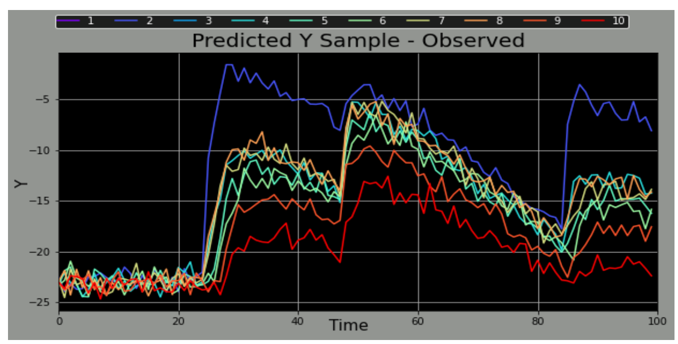
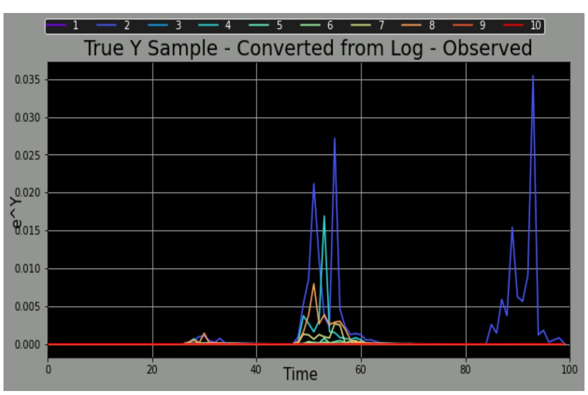
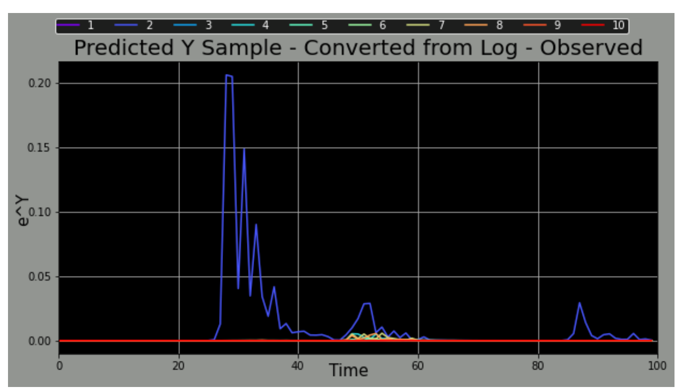
Since for τ, $C_{Candidate_{it}} = C_{Current_{it}}$ for all i, t

$$ratio_{log} = log\left(\frac{\tau_{Candidate}}{\tau_{Current}}\right) - \tau_{Candidate} + \tau_{Current}$$

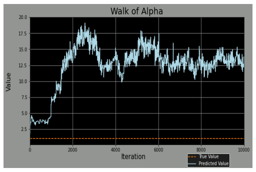
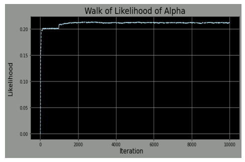
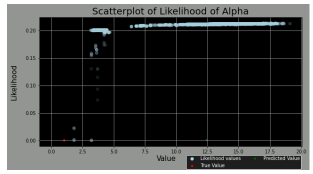
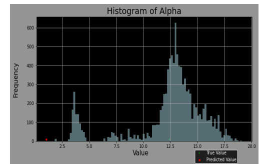
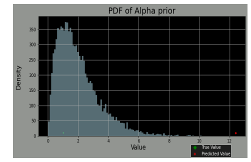
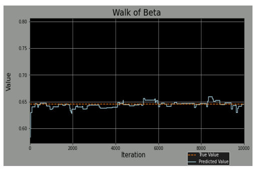
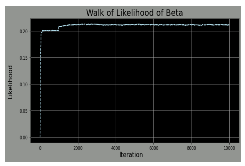
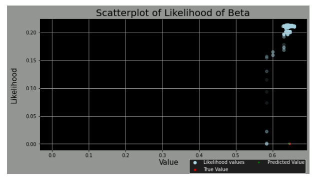
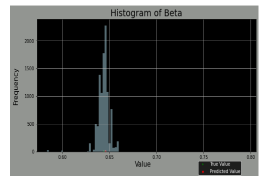
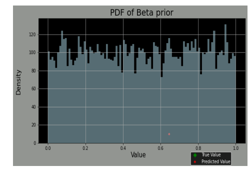
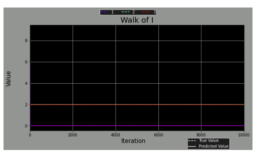
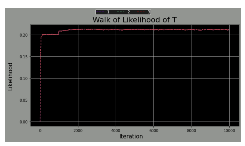
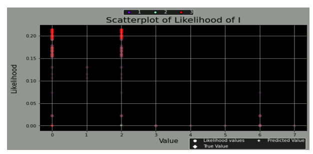
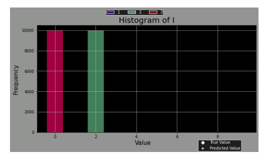
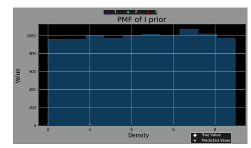
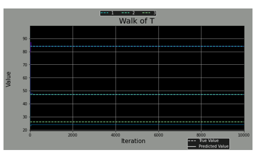
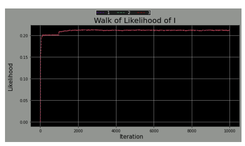
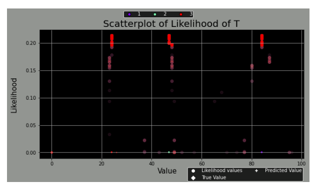
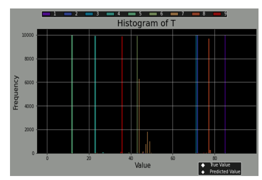
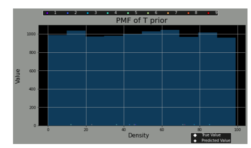
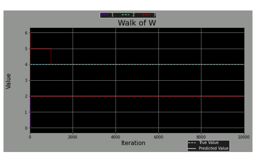
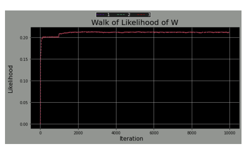
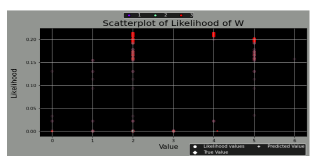
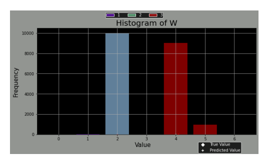
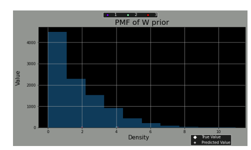
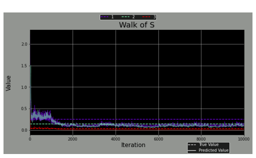
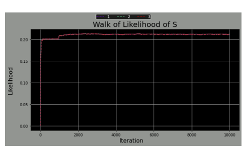
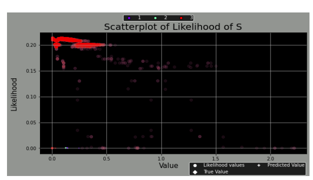
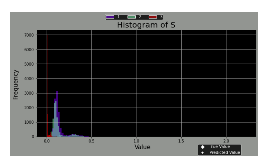
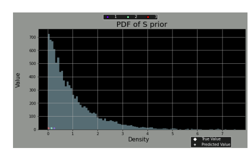
$$P_{Acceptance}(\tau) = e^{ratio_{log}}$$

Results - With source sites missing

Full	True	Predicted
Mass		
Concentration		
Y		
e^Y		

Observed	True	Predicted
Mass		
Concentration		
Y		
e^Y		

Parameters

Parameter P(Acceptance)		Walk of value	Walk of Likelihood	Scatter Plot of Likelihood	Histogram	Prior PDF
α	0.1543					
β	0.0101					
I	0.1038					
T	0.0111					
W	0.2137					
S	0.1420					
τ	0.0444