Change Point Detection for Diversely Distributed Stochastic Processes Using a Probabilistic Method

Muhammad Rizwan Khan [1] and **Biswajit Sarkar** [2,*]

1   Department of Industrial Engineering, Hanyang University, 222 Wangsimni-Ro, Seoul 133-791, Korea
2   Department of Industrial & Management Engineering, Hanyang University, Ansan,
    Gyeonggi-do 15588, Korea
*   Correspondence: bsbiswajitsarkar@gmail.com; Tel.: +82-10-7498-1981; Fax: +82-31-400-5959

Abstract: Unpredicted deviations in time series data are called change points. These unexpected changes indicate transitions between states. Change point detection is a valuable technique in modeling to estimate unanticipated property changes underlying time series data. It can be applied in different areas like climate change detection, human activity analysis, medical condition monitoring and speech and image analyses. Supervised and unsupervised techniques are equally used to identify changes in time series. Even though change point detection algorithms have improved considerably in recent years, several undefended challenges exist. Previous work on change point detection was limited to specific areas; therefore, more studies are required to investigate appropriate change point detection techniques applicable to any data distribution to assess the numerical productivity of any stochastic process. This research is primarily focused on the formulation of an innovative methodology for change point detection of diversely distributed stochastic processes using a probabilistic method with variable data structures. Bayesian inference and a likelihood ratio test are used to detect a change point at an unknown time (k). The likelihood of k is determined and used in the likelihood ratio test. Parameter change must be evaluated by critically analyzing the parameters expectations before and after a change point. Real-time data of particulate matter concentrations at different locations were used for numerical verification, due to diverse features, that is, environment, population densities and transportation vehicle densities. Therefore, this study provides an understanding of how well this recommended model could perform for different data structures.

Keywords: probabilistic method; bayesian statistical modeling; change point detection; likelihood ratio test; time series analysis

1. Introduction

Unexpected deviations in time series data are called change points. These sudden changes indicate transitions between states. Change point detection is worthwhile in modeling, to estimate unexpected property changes underlying time series data. It is applicable in different areas like climate change detection, human activity analysis, medical condition monitoring and speech and image analyses. Supervised and unsupervised techniques are equally used to identify changes in time series. Even though change point detection algorithms have improved considerably in recent years, several undefended challenges exist [1].

Several techniques have been recommended for the identification of undocumented change points in climate data sequences [2]. A change-point analysis technique has been described and its potential applications have been highlighted through a number of examples [3]. The kernel-based change point (KCP) detection procedure can only be used to detect a particular type of change; therefore, based on

the Gaussian KCP method, a new nonparametric approach was proposed for predicting correlation changes, called KCP-corr. KCP-corr performs better than the Cusum technique, which specifically aims to identify correlation changes [4]. A generalized likelihood ratio test (GLRT) was used for detecting changes in the mean of a one dimensional Gaussian process [5]. A new method was recommended for change point detection in a Brownian motion, with a time-dependent diffusion coefficient in fractional Brownian motion [6]. A production inventory model with probabilistic deterioration was developed in two-echelon supply chain management [7]. A two stage change point detection technique in machine monitoring was suggested [8]. Bayesian Approach was used for change point detection of polluted days [9].

A statistical change point algorithm was proposed in which direct density ratio estimation technique was used for deviation measurement of nonparametric deviation estimation among time series samples through relative Pearson divergence variable data structures [10]. An innovative statistical approach for online change point detection was recommended in which estimation method could also be updated online [11]. An economic production quantity model with stochastic demand was developed for an imperfect production system [12]. For a change point test in a series, the Karhunen-Loeve expansion of the limit Gaussian processes was recommended [13]. The test for sudden changes in random fields was presented as a Cramer-von Mises type test and was dependent on the Hilbert space theory [14]. An integrated inventory model was developed to determine the optimal lot size and production uptime while considering stochastic machine breakdown and multiple shipments for a single-buyer and single-vendor [15]. A new methodology was introduced for the identification of structural changes in linear quantile regression models because the conventional mean regression technique couldn't be appropriate for the identification of such structural changes at tails [16]. Supply chain model with stochastic lead time, trade-credit financing and transportation discounts was developed in order to make a coordination mechanism between transportation discounts, trade-credit financing, number of shipments, quality improvement of products and reduced setup cost in such a way that the total cost of the whole system can be reduced, where the supplier offers trade-credit-period to the buyer [17]. The fuzzy classification maximum likelihood change point (FCML-CP) algorithm was suggested for detection of simultaneous multiple change points in the mean and variance of a process and it reduces analysis time [18]. For sequential data series, a Bayesian change point algorithm was presented but it had unreliable restrictions for a number of change points and their location [19].

The Bayesian change point detection (BCPD) technique being suggested in this research paper can overcome challenges in identifying the location and number of change points due to the probabilistic concept. This methodology would precisely be based on posterior distributions and likelihood ratio test to deduce if a change point has occurred. It can also update itself linearly as new data points are observed. Posterior distribution monitoring is the best way to identify the presence of a new change point in observed data points. Simulation studies illustrate that this algorithm is good for rapid detection of existing change points and it also has a low rate of false detection, [19]. Previous work on change point detection was limited. Therefore, more studies are required to investigate appropriate change point detection techniques that are applicable to any data distribution to assess the numerical productivity of any stochastic process. This research is primarily focused on formulation of an innovative methodology for change point detection of diversely distributed stochastic processes by a probabilistic method with variable data structures. The parameter expectations before and after change point must be critically analyzed so that the parameter change can be evaluated. Bayesian inference and the likelihood ratio test are used to detect a change point at an unknown time (k).

Real-time data of particulate matter concentrations at different sites were used to validate the proposed approach. Investigation of particulate matter (PM) pollution status was conducted to evaluate the long-term trends in Seoul which shows a decreasing trend during the study period (2004–2013) [20]. Long-term behavior of particulate matters at urban roadside and background locations in Seoul, Korea were analyzed and the mean PM values exhibit a slight fall over the decade [21]. Probabilistic method

was used to comprehensively analyze the change point (k), parameters before the change point ($\mu_1, \mu_2, ..., \mu_n$) and parameters after the change point ($\eta_1, \eta_2, ..., \eta_n$). Hence, simulation models were built on diverse data structures of different areas to consider different features, that is, environment, population densities and transportation vehicle densities. Therefore, this study delivers a vision about how well this suggested model could perform in different areas. The paper is arranged along these lines: Section 2 discusses a literature review regarding Bayesian change point detection, while Section 3 refers to problem definitions, explaining assumptions and notations and demonstrates the formulation of mathematical models. Sections 4 and 5 depict real world application of the model and results to validate practical application of the proposed models. Section 6 discusses the results for each area; finally, Section 7 presents conclusions of this study.

2. Related Literature

A basic literature review for Bayesian change point methodology was performed. An approach was proposed to detect changes in a non-homogeneous Poisson process and it was used to detect if a change in event rate has occurred, the time of the change and the event rate before and after the change [22]. A novel Bayesian approach was suggested to detect abnormal regions in multiple time series. Model was built and revealed that posterior distribution was used for independent sampling to conclude Bayesian inference. This approach was evaluated for simulated CNVs (copy number variations) and real data to confirm that this methodology is more accurate as compared to other methods [23]. An economic manufacturing quantity model with probabilistic deterioration was developed for a production system [24]. A comparison of Expectation Maximization (EM) method and Bayesian method for change point detection of multivariate data was done. The Bayesian technique involves fewer computational work, while EM reveals better performance for unsuitable priors and minor changes [25]. Min–max distribution free continuous-review model was presented with a service level constraint and variable lead time [26]. The Bayesian change point detection model was recommended to identify the flooding attacks in VoIP systems in which the Session Initiation Protocol (SIP) is used as a signaling mechanism [27].

To acquire accurate and reliable change detection maps for land cover monitoring, a new post classification methodology with iterative slow feature analysis (ISFA) along with Bayesian soft fusion was proposed. This methodology included three steps, first one to define the probability class of images, then a continuous change probability map and last posterior probabilities for the class arrangements of coupled pixels [28]. An economic production quantity model was developed with random defective rate, rework process and backorders for a single stage production system [29]. A Bayesian change point detection methodology was developed to analyze biomarker time series data in women for earlier diagnosis of ovarian cancer [30]. A method for approximation of digital planar curves with line segments and circular arcs using genetic algorithms was proposed [31]. The Generalized Extreme Value (GEV) fused lasso penalty function was used to detect change points for annual maximum precipitation (AMP) in South Korea. A comparison between GEV fused lasso and Bayesian change point analysis was conducted, which depicted that GEV fused lasso method should be used if water resource structures are hydrologically designed [32]. Mathematical models were developed for work-in-process-based inventory by incorporating the effect of random defects rate on lot size and expected total cost function [33]. An innovative Bayesian approach was suggested to detect change points in extreme precipitation data, while the model was based on a generalized Pareto distribution. Four different situations were used for analysis, first with no change, second with a shape change, third with a scale change and fourth with both shape and scale change [34]. See Table 1 for comparison of studies of different authors and for the difference in previous works and this work.

Table 1. Previous studies on this topic.

Document	Change-Point Detection	BCPD (Bayesian Change-Point Detection)	Time Series
Cabrieto et al. (2018) [4]	Kernel change point detection, correlation changes	- -	- -
Keshavarz et al. (2018) [5]	Generalized likelihood ratio test, One-dimensional Gaussian process	- -	- -
Kucharczyk et al. (2018) [6]	Likelihood ratio test, fractional Brownian motion	- -	- -
Lu et al. (2017) [8]	Change-point detection, machine monitoring, Anomaly measure (AR model), Martingale test	- -	Time Series
Hilgert et al. (2016) [11]	On-line change detection, autoregressive dynamic models, CUSUM-like scheme	- - -	- - -
Górecki et al. (2017) [13]	Change point detection, heteroscedastic, Karhunen-Loeve expansions	- -	Time series
Bucchia and Wendler (2017) [14]	Change point detection, bootstrap, Hilbert space valued random fields, (Cramer–von Mises type test)	- - -	Time series
Zhou et al. (2015) [16]	Sequential change point detection, linear quantile regression models	- -	- -
Lu and Chang (2016) [18]	Detecting change points, mean/variance shifts, FCML-CP algorithm	- -	- -
Liu et al. (2013) [10]	Change point detection, relative density ratio estimation	- -	Time series
Ruggieri and Antonellis (2016) [19]	-	Bayesian sequential change point detection	-
Keshavarz and Huang (2014) [25]	- -	Bayesian and Expectation Maximization methods, multivariate change point detection	- -
Kurt et al. (2018) [27]	- -	Bayesian change point model, SIP-based DDoS attacks detection	- -
Wu et al. (2017) [28]	- - -	Post classification change detection, iterative slow feature analysis, Bayesian soft fusion	- - -
Gupta and Baker (2017) [22]	- -	Spatial event rates, change point, Bayesian statistics, induced seismicity	- -
Marino et al. (2017) [30]	- -	Change point, multiple biomarkers, ovarian cancer	- -
Jeon et al. (2016) [32]	- -	Abrupt change point detection, annual maximum precipitation, fused lasso	- -
Bardwell and Fearnhead (2017) [23]	-	Bayesian Detection, Abnormal Segments	Time series
Chen et al. (2017) [34]	- -	Bayesian change point analysis, extreme daily precipitation	- -
This study	Change point detection	Bayesian approach and likelihood ratio test for change point detection	Time series

3. Methodological Part

3.1. Problem Definition

This research is primarily focused on formulation of a unique methodology for change point detection of diverse data structures following any kind of distribution at any unknown time (k) at any area across the globe. The existing procedures for change point detection are either very complicated or no applicable to stochastic processes and random time series. That's why, a more precise, well defined and easily applicable approach for change point detection of stochastic processes and random time series has been proposed. Second, analysis of these changes need to be conducted, whether or not these change points are favorable. For this, a comparison of distribution parameters before and after a change point has to be performed for evaluation of subjected change. Third, an alteration in parameters expectations must be measured to define new policies for further improvements in the current states. For anticipated goals, the Probabilistic method will be used to determine posterior probabilities of data and the change point in that Bayesian model will be identified through likelihood ratio test. This suggested model will be numerically validated by using real-time data of particulate matter concentrations and particulate matter hazards in different areas of Seoul, South Korea, observed from January 2004 to December 2013. The change point (k) for particulate matter ($PM_{2.5}$ and PM_{10})

daily concentrations, the parameters before the change point ($\mu_1, \mu_2, ..., \mu_n$) and the parameters after the change point ($\eta_1, \eta_2, ..., \eta_n$) are comprehensively analyzed. The central idea for using different regions is their considerably different features, that is, environment, population densities and transportation vehicle densities. Hence, this study can also be the basis for implementation of the recommended model in different areas. Later, this probabilistic method is verified by the CUSUM approach. The results of the CUSUM approach are compared with the probabilistic method.

1. The probabilistic method is based on probability distributions, which can be applicable to data distribution. In this case, first define the data distributions and then apply proposed method to attain the results. This methodology is better to apply for random data structures and time series.
2. The CUSUM approach is directly applicable to the raw data, which is good for deterministic data structures.

3.2. Notations

The list of notations to represent the random variables and parameters are as follows.

Indices

i	sequence data points in time series, where $i \in 1, 2, ..., n$
h	replication number (multiple simulations are performed to get the converged value), $h \in 1, 2, ..., m$
j	position in the replication or chain, (V_{hj} be the jth observation from the hth replication), $j \in 1, 2, ..., n$

Random variables

Y	random process or stochastic process
y	variable (Y) at any given point
y_i	variable (Y) at point i where $i \in 1, 2, ..., n$

Parameters

k	change point in a random process
μ_1	first parameter before change point k associated with the probability distribution function of random variable Y
η_1	first parameter after change point k associated with the probability distribution function of random variable Y
μ_2	second parameter before change point k associated with the probability distribution function of random variable Y
η_2	second parameter after change point k associated with the probability distribution function of random variable Y
μ_n	nth parameter before change point k associated with the probability distribution function of random variable Y
η_n	nth parameter after change point k associated with the probability distribution function of random variable Y
μ	mean before change point k for Normal distribution of random variable Y
η	mean after change point k for Normal distribution of random variable Y
σ^2	variance before change point k for Normal distribution of random variable Y
ϕ^2	variance after change point k for Normal distribution of random variable Y
θ_0	mean for Normal prior distribution $p(\mu)$
$\tau_0{}^2$	variance for Normal prior distribution $p(\mu)$
θ_k	mean for Normal posterior distribution $p(\mu|\sigma^2, y_1, y_2, y_3, ..., y_k)$

$\tau_k{}^2$	variance for Normal posterior distribution $p(\mu\|\sigma^2, y_1, y_2, y_3, ..., y_k)$
λ_0	mean for Normal prior distribution $p(\eta)$
$\omega_0{}^2$	variance for Normal prior distribution $p(\eta)$
λ_n	mean for Normal posterior distribution $p(\eta\|\phi^2, y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)$
$\omega_n{}^2$	variance for Normal posterior distribution $p(\eta\|\phi^2, y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)$
$\bar{y}$	average of data values

Variables

$p(\mu)$	prior Normal distribution for mean before change point k
$p(\sigma^2)$	prior Normal distribution for variance before change point k
$p(\eta)$	prior Normal distribution for mean after change point k
$p(\phi^2)$	prior Normal distribution for variance after change point k
$p(\mu, \sigma^2)$	joint prior probability before change point k
$p(\eta, \phi^2)$	joint prior probability after change point k
$p(\mu\|y_i)$	posterior Normal distribution for mean before change point k
$p(\eta\|y_i)$	posterior Normal distribution for mean after change point k
$p(\mu, \sigma^2\|y_i)$	joint posterior probability before change point k
$p(\eta, \phi^2\|y_i)$	joint posterior probability after change point k
$p(y_i\|\mu, \sigma^2)$	likelihood or sampling model given (μ, σ^2)
$p(y_i\|\eta, \phi^2)$	likelihood or sampling model given (η, ϕ^2)
k_0	prior sample size for mean parameters (μ, η)
k_k	(prior sample size $k_0 + k$)
k_n	(prior sample size $k_0 + (n - k)$)
ν_0	prior sample size for variance parameters (σ^2, ϕ^2)
σ_0^2	prior sample variance
ν_k	(prior sample size $\nu_0 + k$)
ν_n	(prior sample size $\nu_0 + (n - k)$)
V	mean of the chain or replications (average daily pollutant concentrations)
V_{hj}	jth observation from the hth replication
V_h	mean of hth replication
V	mean of m replications
B	between sequence variance representing the variance of replications with the mean of m replications
S_h^2	variance for all replications
W	within sequence variance, the mean variance for m replications
$Var(V)$	overall estimate of the variance of V in the target distribution
$Var(V)_\mu$	first parameter overall variance before change point k
$Var(V)_\eta$	first parameter overall variance after change point k
$Var(V)_{\sigma^2}$	second parameter overall variance before change point k
$Var(V)_{\phi^2}$	second parameter overall variance after change point k
$\sqrt{R}$	estimated potential scale reduction for convergence
$\sqrt{R}_\mu$	first parameter convergence before change point k
$\sqrt{R}_\eta$	first parameter convergence after change point k
$\sqrt{R}_{\sigma^2}$	second parameter convergence before change point k
$\sqrt{R}_{\phi^2}$	second parameter convergence after change point k
$\sqrt{R}_k$	change point (k) convergence
S_i	cumulative sum

3.3. Assumptions

The following assumptions were used for the proposed model:

1. Y represents the random data at given time t and this random data series is distributed at state space $y \in 1, 2, ..., n$ that can be any random value.
2. $Y(0) = 0$ means that no event occurred at time $t = 0$, while time series random data are observed on intervals of equal length.
3. The random data structure follows a specific probability distribution function, in any interval of length (t), resulting in a random variable with parameters $(\mu_1, \mu_2, ..., \mu_n)$.

3.4. Formulation of Change Point Detection Model

The probability distribution function of a random variable Y with the parameters $\mu_1, \mu_2, ..., \mu_n$ at any specific point y is given as follows

$$Y \sim P(\mu_1, \mu_2, ..., \mu_n) = f(y; \mu_1, \mu_2, ..., \mu_n) = p(Y = y | \mu_1, \mu_2, ..., \mu_n) \text{ for } y \in 1, 2, ..., n$$

After defining the probability distribution function of random process Y. Now, divide the process in two segments; first segment defines the process before change point and second segment defines the process after change point. Let the change point in the random process Y be denoted by k and $(\mu_1, \mu_2, ..., \mu_n)$ be the random variable parameters before change point k, while $(\eta_1, \eta_2, ..., \eta_n)$ are the random variable parameters after change point k.

$$y_i \sim P(\mu_1, \mu_2, ..., \mu_n) = f(y; \mu_1, \mu_2, ..., \mu_n) = p(Y = y_i | \mu_1, \mu_2, ..., \mu_n) \text{ for } i = 1, 2, ..., k$$

$$y_i \sim P(\eta_1, \eta_2, ..., \eta_n) = f(y; \eta_1, \eta_2, ..., \eta_n) = p(Y = y_i | \eta_1, \eta_2, ..., \eta_n) \text{ for } i = k+1, k+2, ..., n$$

The joint probability function is the product of a marginal probability function. If random variable $Y = y_i$ with parameters $(\mu_1, \mu_2, ..., \mu_n)$ is modeled, then the joint probability function of the sample data will be as below:

$$p(Y = y_i | \mu_1, \mu_2, ..., \mu_n) = \prod_{i=1}^{k} p(y_i | \mu_1, \mu_2, ..., \mu_n) \text{ for } i \in 1, 2, ..., k$$

A class of prior densities is conjugate for the likelihood/sampling model $p(y_i | \mu_1)$ if the posterior probability distribution is in the same class. Therefore, prior distribution $p(\mu_1)$ and posterior distribution $p(\mu_1 | y_i)$ will follow the same conjugate prior distribution as the likelihood/sampling model $p(y_i | \mu_1)$. However, the likelihood $p(y_i | \mu_1)$ follows a random distribution based on data. The Bayes theorem can be used to determine the posterior probability $p(\mu_1 | y_i)$ of individual parameter μ_1.

Posterior probability $\propto$ Prior probability $\times$ Likelihood

$$p(\mu_1 | y_i) \propto p(\mu_1) p(y_i | \mu_1)$$

Bayesian inference for multiple unknown parameters is not conceptually different from the one-parameter case. For any joint prior distribution $p(\mu_1, \mu_2, ..., \mu_n)$, posterior inference proceeds using Bayes' rule:

$$p(\mu_1, \mu_2, ..., \mu_n | y_i) = \frac{p(y_i | \mu_1, \mu_2, ..., \mu_n) p(\mu_1, \mu_2, ..., \mu_n)}{p(y_i)} \text{ for } i = 1, 2, ..., k$$

$$p(\eta_1, \eta_2, ..., \eta_n | y_i) = \frac{p(y_i | \eta_1, \eta_2, ..., \eta_n) p(\eta_1, \eta_2, ..., \eta_n)}{p(y_i)} \text{ for } i = k+1, k+2, ..., n$$

The inference for this multi-parameter model can be broken down into multiple one-parameter problems. First, make an inference for μ_1 when remaining parameters $(\mu_2, ..., \mu_n)$ are known and use a conjugate prior distribution for μ_1. For any (conditional) prior probability $p(\mu_1|\mu_2, ..., \mu_n)$, the posterior distribution will satisfy

$$p(\mu_1|y_1, y_2, y_3, ..., y_k, \mu_2, ..., \mu_n) \propto p(\mu_1|\mu_2, ..., \mu_n)p(y_1, y_2, y_3, ..., y_k|\mu_1, \mu_2, ..., \mu_n)$$

$$p(\mu_1|y_1, y_2, y_3, ..., y_k, \mu_2, ..., \mu_n) = \frac{p(\mu_1|\mu_2, ..., \mu_n)p(y_1, y_2, y_3, ..., y_k|\mu_1, \mu_2, ..., \mu_n)}{p(y_1, y_2, y_3, ..., y_k|\mu_2, ..., \mu_n)}$$

The posterior parameters combine the prior parameters with terms from the data:

Posterior information = Prior information + Data information

Hence, the prior distribution and sampling model are as follows:

$$p(y_1, y_2, y_3, ..., y_k|\mu_1, \mu_2, ..., \mu_n) \sim P(\mu_1, \mu_2, ..., \mu_n) \text{ for } i = 1, 2, ..., k$$

$$p(y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n|\eta_1, \eta_2, ..., \eta_n) \sim P(\eta_1, \eta_2, ..., \eta_n) \text{ for } i = k+1, k+2, ..., n$$

$$\mu_1, \mu_2, ..., \mu_n \sim \text{Conjugate prior distribution (Prior hyperparameters) for } i = 1, 2, ..., k$$

$$\eta_1, \eta_2, ..., \eta_n \sim \text{Conjugate prior distribution (Prior hyperparameters) for } i = k+1, k+2, ..., n$$

Thus, the posterior inference for first parameter μ_1 and η_1 can be given by

$$p(\mu_1|y_1, y_2, y_3, ..., y_k, \mu_2, ..., \mu_n) \sim \text{Conjugate posterior distribution (Posterior hyperparameters)}$$
$$\text{for } i = 1, 2, ..., k$$

$$p(\eta_1|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n, \eta_2, ..., \eta_n) \sim \text{Conjugate posterior distribution (Posterior hyperparameters)}$$
$$\text{for } i = k+1, k+2, ..., n$$

Just as the prior distribution for μ_1 and $\mu_2, ..., \mu_n$ can be decomposed as $p(\mu_1, \mu_2, ..., \mu_n) = p(\mu_1|\mu_2, ..., \mu_n)p(\mu_2, ..., \mu_n)$, the posterior distribution can be similarly decomposed:

$$p(\mu_1, \mu_2, ..., \mu_n|y_1, y_2, y_3, ..., y_k) = p(\mu_1|y_1, y_2, y_3, ..., y_k, \mu_2, ..., \mu_n)p(\mu_2, ..., \mu_n|y_1, y_2, y_3, ..., y_k)$$

Similarly, $p(\eta_1, \eta_2, ..., \eta_n|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)$ after a change point can be given by

$$p(\eta_1, \eta_2, ..., \eta_n|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n) = p(\eta_1|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n, \eta_2, ..., \eta_n)$$
$$p(\eta_2, ..., \eta_n|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)$$

The conditional distribution of μ_1 given $\mu_2, ..., \mu_n$ and the data $(y_1, y_2, y_3, ..., y_n)$ was obtained in previous sections. The posterior distribution of all other parameters $\mu_2, ..., \mu_n$ can be found by estimating an integration over the unknown value of μ_1:

$$p(\mu_2|y_1, ..., y_k) \propto p(\mu_2)p(y_1, ..., y_k|\mu_2) = p(\mu_2)\int p(y_1, ..., y_k|\mu_1, \mu_2)p(\mu_1|\mu_2)d\mu_1$$

$$(\mu_2|y_1, y_2, y_3, ..., y_k) \sim \text{Conjugate posterior distribution (Posterior hyperparameters)}$$

Similarly, $p(\eta_2|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)$ after a change point can be given by

$$(\eta_2|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n) \sim \text{Conjugate posterior distribution (Posterior hyperparameters)}$$

The Bayesian inference for parameters $\mu_3, ..., \mu_n$ and $\eta_3, ..., \eta_n$ in the distribution can be determined in the similar way as used for μ_2 and η_2.

The likelihood function for change point detection can be determined as:

$$L\Big(Y;\text{Change point},\text{Parameter}_1,\text{Parameter}_2,...,\text{Parameter}_n\Big)$$

$$= \Big[\exp\Big(\text{Change point}\big(\text{Expectation after change point - Expectation before change point}\big)\Big)$$

$$\left(\frac{\text{Expectation before change point}}{\text{Expectation after change point}}\right)^{\sum_{i=1}^{\text{Change point}} y_i}\Big] \quad (1)$$

$$L\Big(Y;k,(\mu_1,\mu_2,...,\mu_n),(\eta_1,\eta_2,...,\eta_n)\Big)$$

$$= \Big[exp\Big(k\Big(E[Y=y_i|\eta_1,\eta_2,...,\eta_n] - E[Y=y_i|\mu_1,\mu_2,...,\mu_n]\Big)\Big)\left(\frac{E[Y=y_i|\mu_1,\mu_2,...,\mu_n]}{E[Y=y_i|\eta_1,\eta_2,...,\eta_n]}\right)^{\sum_{i=1}^{k} y_i}\Big] \quad (2)$$

The change point for random process Y is being detected by the likelihood ratio test (LRT). The LRT begins with a comparison of the likelihood scores of the two models; one is null model and other is alternative model. The test is based on the likelihood ratio, which states how many times more likely the data are under one model than the other. This likelihood ratio compared to a critical value used to decide whether to reject the null model.

$$f(\text{Change point}|Y,\text{parameters before change point},\text{parameters after change point})$$

$$= \frac{L(Y;\text{Change point},\text{parameters before change point},\text{parameters after change point})}{\sum_{j=1}^{n} L(Y;j,\text{Change point},\text{parameters before change point},\text{parameters after change point})} \quad (3)$$

The likelihood ratio test for change point k given the random variable Y and parameters before and after change points is as follows:

$$f\Big(k|Y,(\mu_1,\mu_2,...,\mu_n),(\eta_1,\eta_2,...,\eta_n)\Big) = \frac{L\big(Y;k,(\mu_1,\mu_2,...,\mu_n),(\eta_1,\eta_2,...,\eta_n)\big)}{\sum_{j=1}^{n} L\big(Y;j,(\mu_1,\mu_2,...,\mu_n),(\eta_1,\eta_2,...,\eta_n)\big)}$$

3.4.1. Multiple Change Points Detection

After detecting first change point k, now the data can be broken into two distinct segments, one each side of the change point, 1 to k and $k+1$ to n. Apply the same above mentioned procedure on each segment separately to detect multiple change points in the random process Y.

3.4.2. Convergence of the Parameters

Only one simulation run cannot signify the real features of the resulting model. That's why, the Gelman-Rubin Convergence diagnostic is being used for the estimation of steady-state parameters by running multiple number sequences of the chain. Lack of convergence can be detected by comparing multiple sequences but cannot be detected by looking at a single sequence. Therefore, multiple sequences of the chain are being run to estimate the actual characteristics of the target distribution, Gelman and Rubin [35–37]. m replications of the simulation ($m \geq 10$) are performed, each of length $n = 1000$. If the target distribution is unimodal, then Cowles and Carlin recommend performance of at least 10 chains [38]. The mean pollutant concentration is a parameter of interest and is denoted by V.

Scalar summary V = Mean of the chain (average daily pollutant concentrations)

Let V_{hj} be the jth observation from the hth replication

$$V_{hj} = \text{single observation for mean pollutant concentration per day}$$

where, replication number $= h \in 1, 2, ..., m,$　observation number in a replication $= j \in 1, 2, ..., n.$

Mean of hth replication

$$V_h = \frac{1}{n} \sum_{j=1}^{n} V_{hj}$$

Mean of m replications

$$V = \frac{1}{m} \sum_{h=1}^{m} V_h$$

The between sequence variance represents the variance of a mean of m replications and is calculated as follows:

$$B = \frac{n}{m-1} \sum_{i=1}^{m} (V_h - V)^2$$

Variance for all replications is calculated to determine the within-sequence variance

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (V_{hj} - V)^2$$

The within-sequence variance is the mean variance for k replications determined as given below:

$$W = \frac{1}{m} \sum_{i=1}^{m} S_h^2$$

Finally, the within-sequence variance and between-sequence variance are combined to get an overall estimate of the variance of V in the target distribution

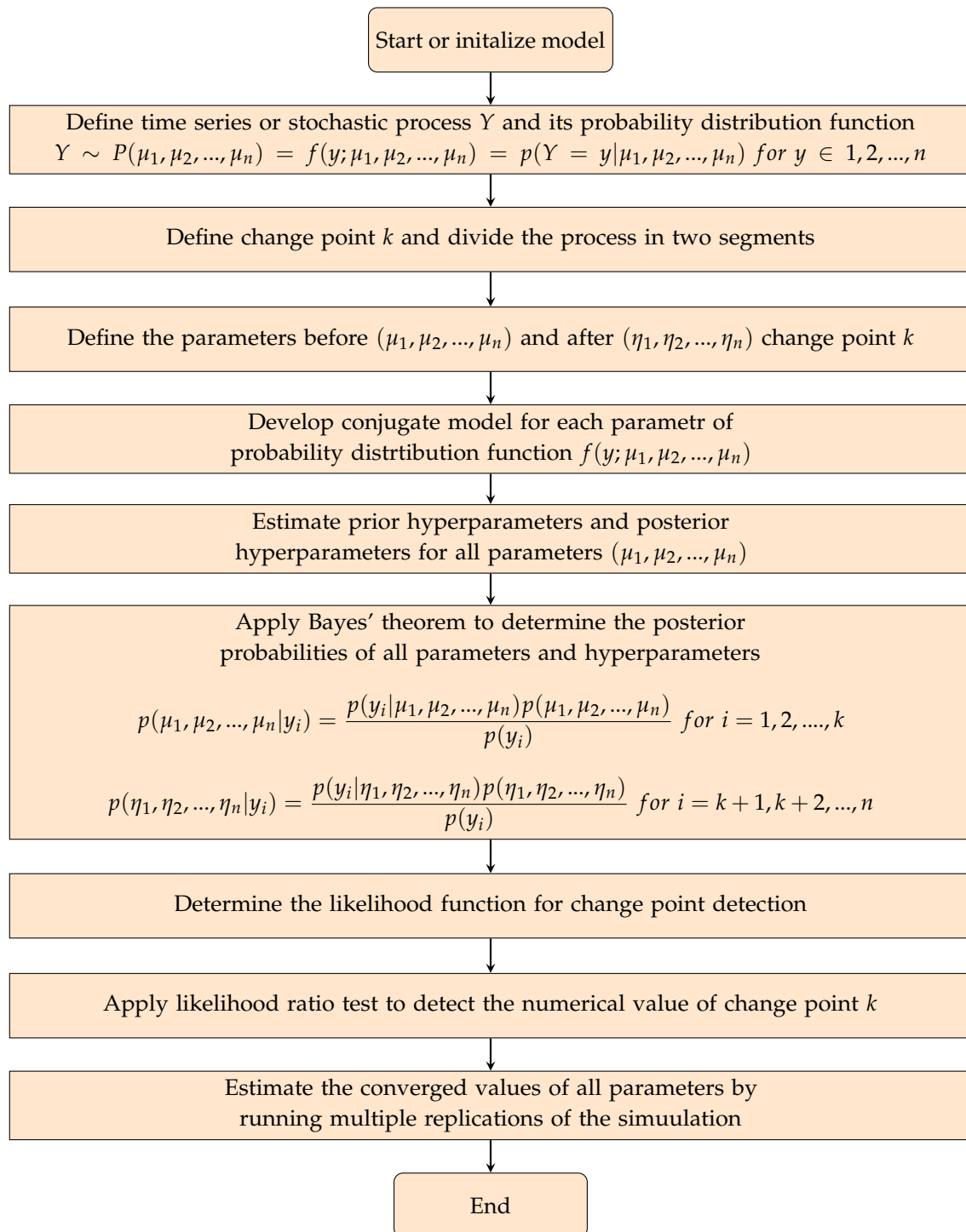$$Var(V) = \frac{n-1}{n} W + \frac{1}{n} B$$

Convergence is identified by calculating

$$\sqrt{R} = \sqrt{\frac{Var(V)}{W}}$$

This factor $\sqrt{R}$ (estimated potential scale reduction) is the proportions among the upper and lower bounds on the standard deviation of V that are used to compute the factor and $Var(V)$ could be reduced through a larger number of iterations. Further iterations of the chain must be performed if the potential scale reduction is high. Run the replications for all scalar summaries until R is lower than 1.1 or 1.2.

3.5. Flowchart Algorithm

The flowchart for change point (k) detection, for any random process Y, is given as follows:

```
┌─────────────────────────────┐
│    Start or initalize model │
└─────────────────────────────┘
```

Define time series or stochastic process Y and its probability distribution function
$Y \sim P(\mu_1, \mu_2, ..., \mu_n) = f(y; \mu_1, \mu_2, ..., \mu_n) = p(Y = y | \mu_1, \mu_2, ..., \mu_n) \; for \; y \in 1, 2, ..., n$

Define change point k and divide the process in two segments

Define the parameters before $(\mu_1, \mu_2, ..., \mu_n)$ and after $(\eta_1, \eta_2, ..., \eta_n)$ change point k

Develop conjugate model for each parametr of
probability distrtibution function $f(y; \mu_1, \mu_2, ..., \mu_n)$

Estimate prior hyperparameters and posterior
hyperparameters for all parameters $(\mu_1, \mu_2, ..., \mu_n)$

Apply Bayes' theorem to determine the posterior
probabilities of all parameters and hyperparameters

$$p(\mu_1, \mu_2, ..., \mu_n | y_i) = \frac{p(y_i | \mu_1, \mu_2, ..., \mu_n) p(\mu_1, \mu_2, ..., \mu_n)}{p(y_i)} \; for \; i = 1, 2, ..., k$$

$$p(\eta_1, \eta_2, ..., \eta_n | y_i) = \frac{p(y_i | \eta_1, \eta_2, ..., \eta_n) p(\eta_1, \eta_2, ..., \eta_n)}{p(y_i)} \; for \; i = k+1, k+2, ..., n$$

Determine the likelihood function for change point detection

Apply likelihood ratio test to detect the numerical value of change point k

Estimate the converged values of all parameters by
running multiple replications of the simuulation

```
┌─────────────┐
│     End     │
└─────────────┘
```

3.6. Comparison Method for Change Point Detection

A change point analysis was performed using a combination of CUSUM (cumulative sum control chart) and bootstrapping for comparative analysis.

3.6.1. The CUSUM Technique

The CUSUM is a sequential analysis technique typically used for monitoring change detection. CUSUM charts are constructed by calculating and plotting a cumulative sum based on the data. The cumulative sums are calculated as follows.

1. First calculate the average.

$$\bar{y} = \left(\frac{y_1 + y_2 + y_3 +, ..., y_n}{n} \right)$$

2. Start the cumulative sum at zero by setting $S_0 = 0$,
3. Calculate the other cumulative sums by adding the difference between the current value and the average to the previous sum, that is,

$$S_i = S_{i-1} + \left(y_i - \bar{y} \right)$$

The cumulative sum is not the sum of the values but it is the cumulative sum of the differences in the values and averages. As the average is being deducted from each value, the final cumulative sum must be zero. Some practice is required to interpret a CUSUM chart. If, during a certain period of time, the overall average is less than the values. Then, the sum will steadily increase because the values being added to cumulative sum will be positive. An upward trend in the CUSUM chart shows a certain period of time, when overall average is less than the values. Similarly, a downward trend in the chart shows that overall average is above than the values. A rapid change in the trend of CUSUM specifies a shift or change in the average. Certain periods, when the CUSUM chart follows straight line, it indicates no change in average.

3.6.2. Bootstrap Analysis

Bootstrap analysis can be performed to determine confidence level for apparent change. The magnitude of change S_{diff} must be estimated before performing bootstrap analysis.

$$S_{diff} = S_{max} - S_{min}$$

$$S_{max} = \max_{i=0,1,2,...,} S_i$$

$$S_{min} = \min_{i=0,1,2,...,} S_i$$

Once the estimator of the magnitude of the change has been selected, the bootstrap analysis can be performed. A single bootstrap is performed as follows.

1. Generate a bootstrap sample of n units, denoted $y^0{}_1, y^0{}_2, y^0{}_3, ... y^0{}_n$ by randomly reordering the original n values. This is called sampling without replacement.
2. Based on the bootstrap sample, calculate the bootstrap CUSUM, denoted $S^0{}_0, S^0{}_1, S^0{}_2, ... S^0{}_n$.
3. Calculate the maximum, minimum and difference of the bootstrap CUSUM, denoted $S^0{}_{max}, S^0{}_{min}$ and $S^0{}_{diff}$.
4. Determine whether the bootstrap difference $S^0{}_{diff}$ is less than the original difference S_{diff} respectively.

The idea behind bootstrapping is that the bootstrap samples represent random reordering of the data that mimic the behavior of the CUSUM if no change has occurred. By performing a large number of bootstrap samples, the variance in S_{diff} if no change took place can be estimated. The value can be compared with the S_{diff} value calculated from the data in its original order to determine if this value is consistent with the expectation under zero change, if bootstrap CUSUM charts tend to stay closer to zero than the CUSUM of the data in its original order, a change likely occurred. A bootstrap analysis consists of performing a large number of bootstraps and counting the number of bootstraps for which $S^0{}_{diff}$ is less than S_{diff}. Let N be the number of bootstrap samples performed and let X be

the number of bootstraps for which $S^0{}_{diff} < S_{diff}$. Then, the confidence level that a change occurred as a percentage is calculated as follows:

$$\text{Confidence Level} = 100 \, \frac{X}{N} \, \text{percentage}$$

This is the solid proof to indicate that change does have occurred. Based on all possible reordering of the data, one would prefer to estimate the distribution of $S^0{}_{diff}$ instead of bootstrapping, which is not possible usually. That's why, for better estimation, number of bootstrap samples need to be increased. Bootstrapping is a distribution free methodology with only one supposition of an independent error structure. Change-point analysis and control charting, both are dependent on the mean-shift model. Let $y_1, y_2, y_3, ..., y_n$ represent the data in time order. The mean-shift model can be written as

$$y_i = \mu_i + \epsilon_i$$

where, μ_i is the average at time i. Generally $\mu_i = \mu_{i-1}$ except for a small number of values of i called the change-points. ϵ_i is the random error associated with the ith value and is assumed to be independent with a mean of zero. Once a change has been detected, an estimate of the time at which the change occurred can be made. One such estimator is the CUSUM estimator. Let m be such that

$$| \, S_m \, | = \max_{i=0,1,2,...,} | \, S_i \, |$$

Here, S_m is the point furthest from zero in the CUSUM chart. The point m estimates the last point before the change occurred. The point $m + 1$ estimates the first point after the change.

3.6.3. Mean and Variance Estimation

Once a change has been detected, the data can be broken into two segments, one on each side of the change-point, 1 to m and $m + 1$ to n. Then, the two segments can be analyzed by determining their parameters.

$$Mean = \bar{y} = \left(\frac{\sum y_i}{n} \right)$$

$$Variance = \sigma^2 = \left(\frac{\sum (y_i - Mean)^2}{n} \right)$$

4. Computational Experiment

Section 4.1 described the numerical verification of the formulated mathematical model to authenticate the validity of model. Real-time data of particulate matter daily concentrations for four different sites of Seoul, South Korea were utilized for this investigation as given in Section 4.2.

4.1. Toy Model for Validation with Known Solution

As shown in Figures 1–3, an artificial data set of random data is generated which consists of two segments with equal length of 50 data points. The samples are drawn from the Poisson distributions Poisson(5) and Poisson(2.5), respectively. Thus, change point occurs at 50th data point.
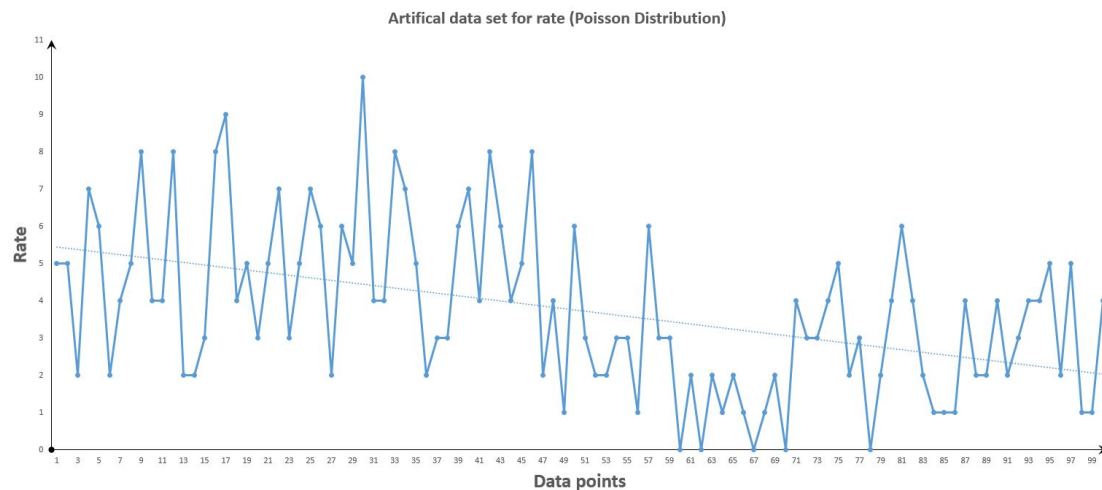
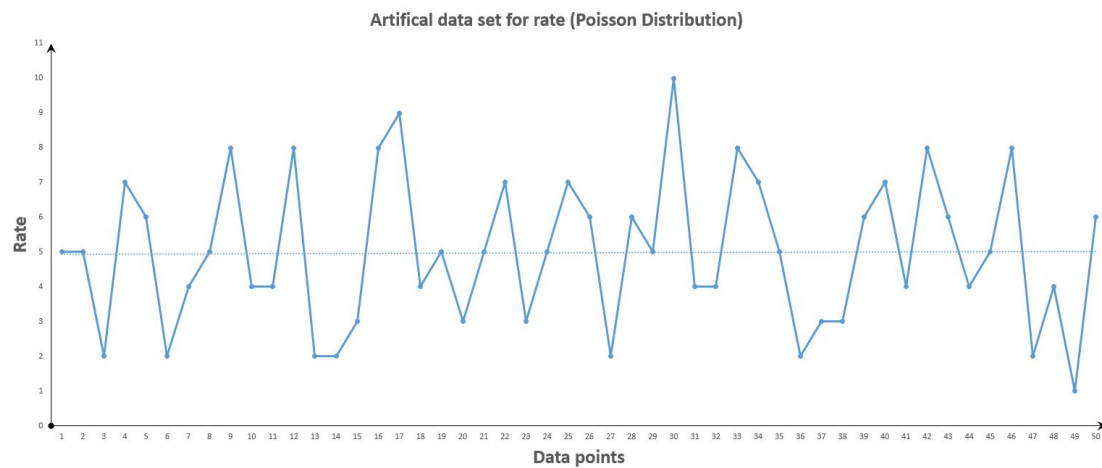**Figure 1.** Artificial Data set for rate (Poisson Distribution).

Figure 2. Artificial Data set for Poisson Distribution before change point.
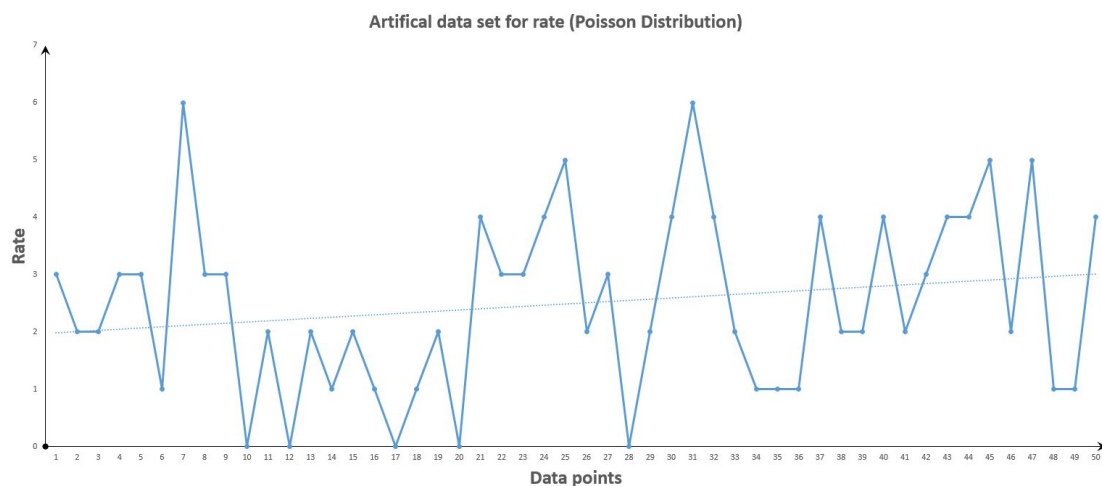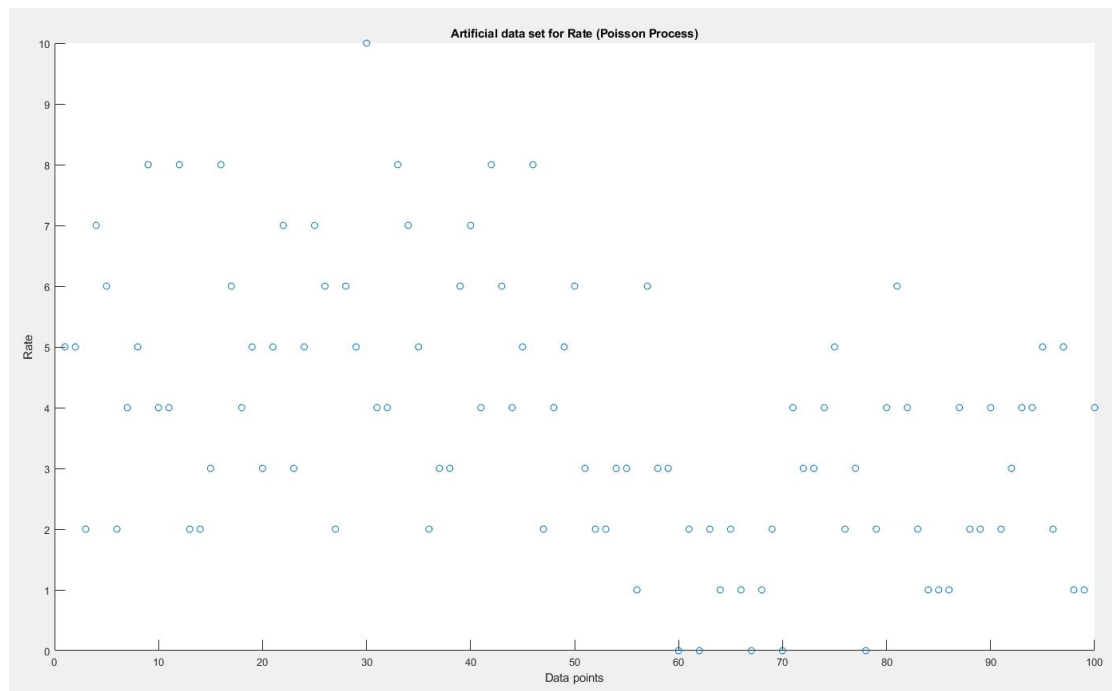

Figure 3. Artificial Data set for Poisson Distribution after change point.

In Table 2, the results for this artificial data set obtained through Probabilistic method have been described. As shown in Figures 4–11, the following results were acquired by applying the method explained in Section 3.4 to this artificial data set.

Table 2. Change point (*k*) for artificial data set.

Change Point (*k*) for Artificial Data Set		
(*Rate* = *θ*) before change	(*Rate* = *λ*) after change	*K* change point
5.0	2.5	50

Figure 4. Artificial Data set time series for Poisson Distribution.
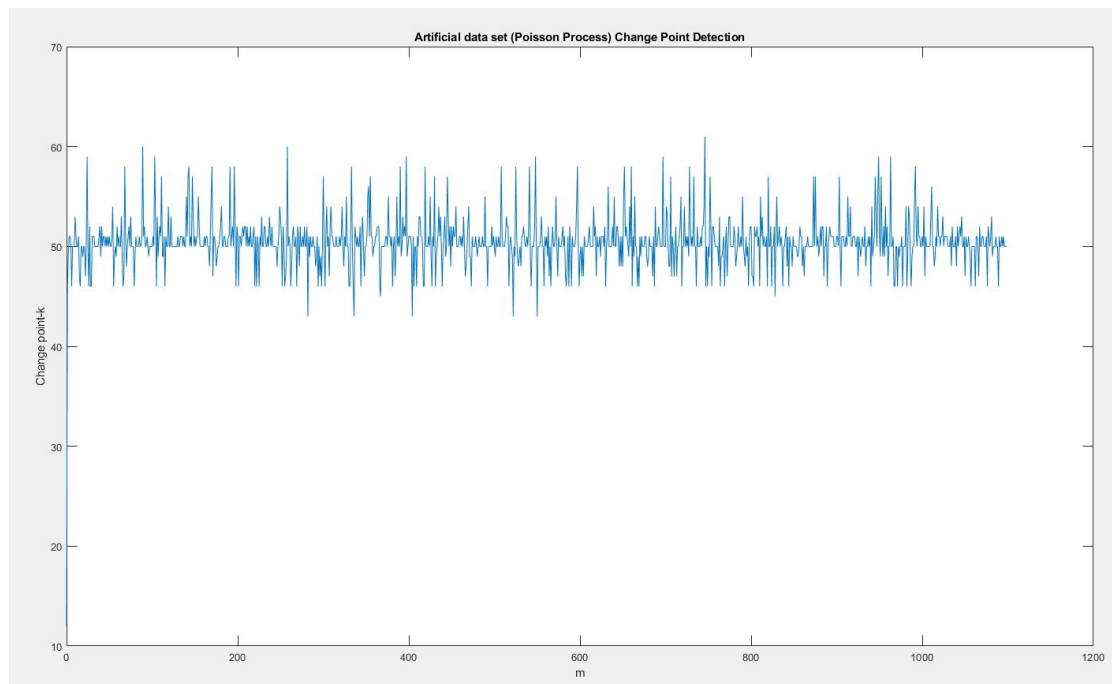
Figure 5. Change point (*k*).
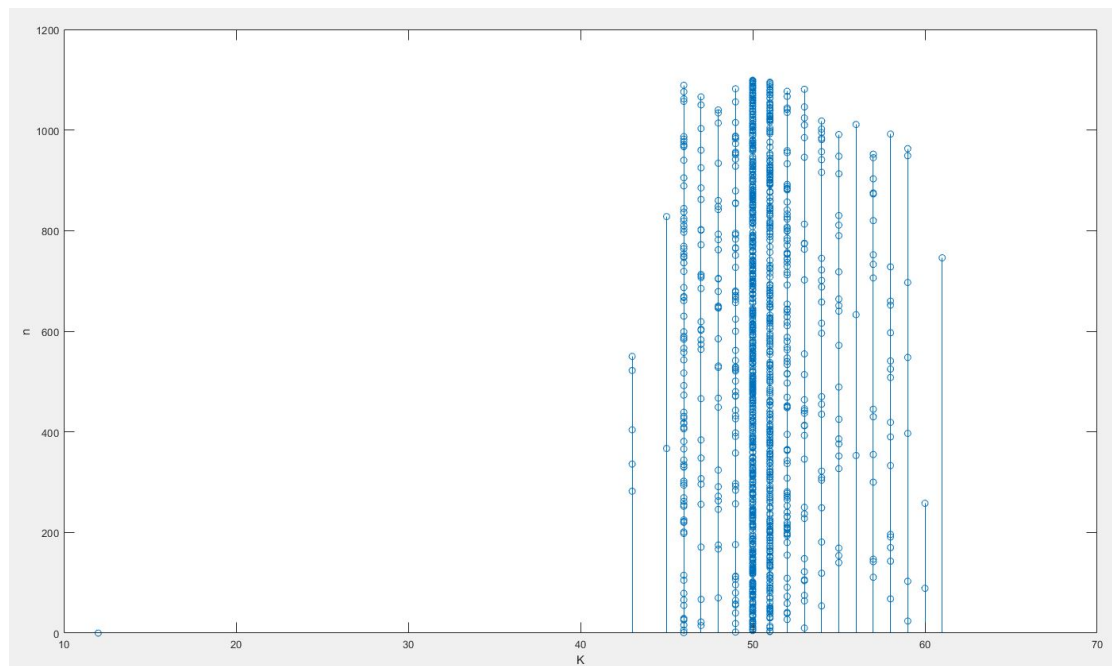
Figure 6. Change point (k) frequency histogram.
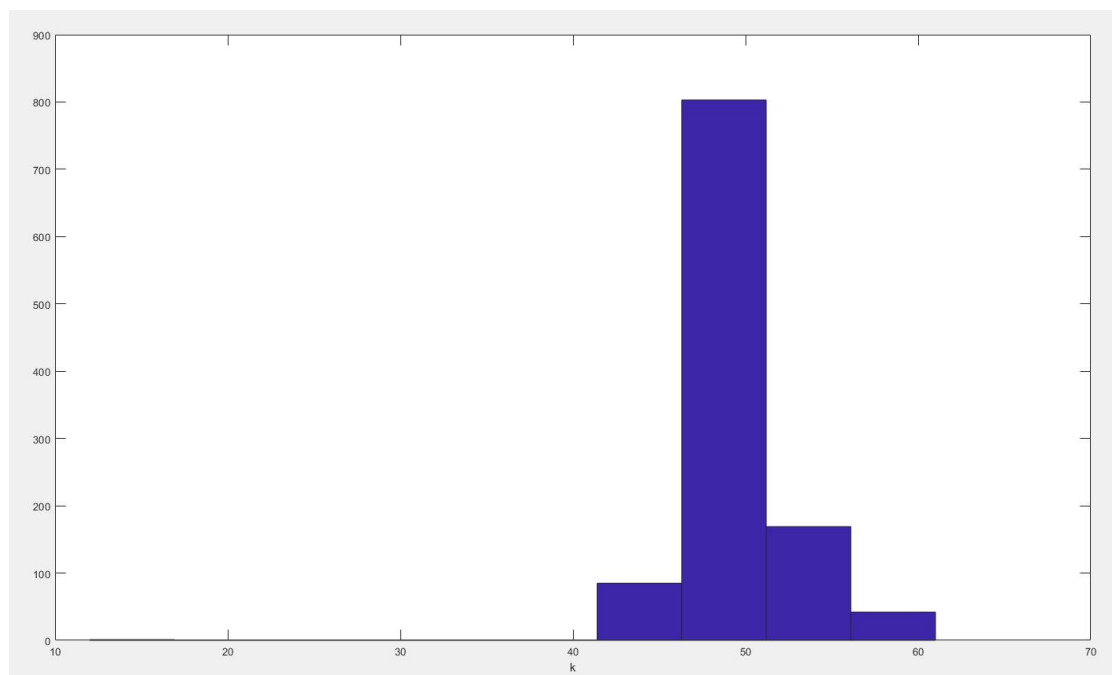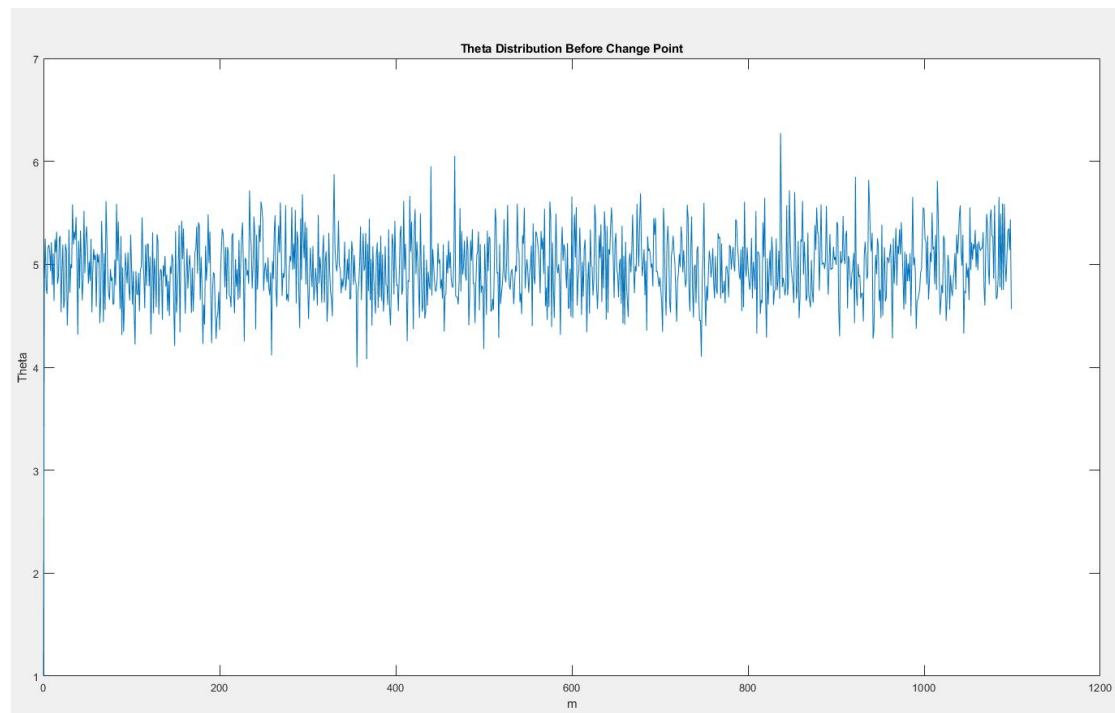

Figure 7. Change point (k) density histogram.

Figure 8. Rate before change point.
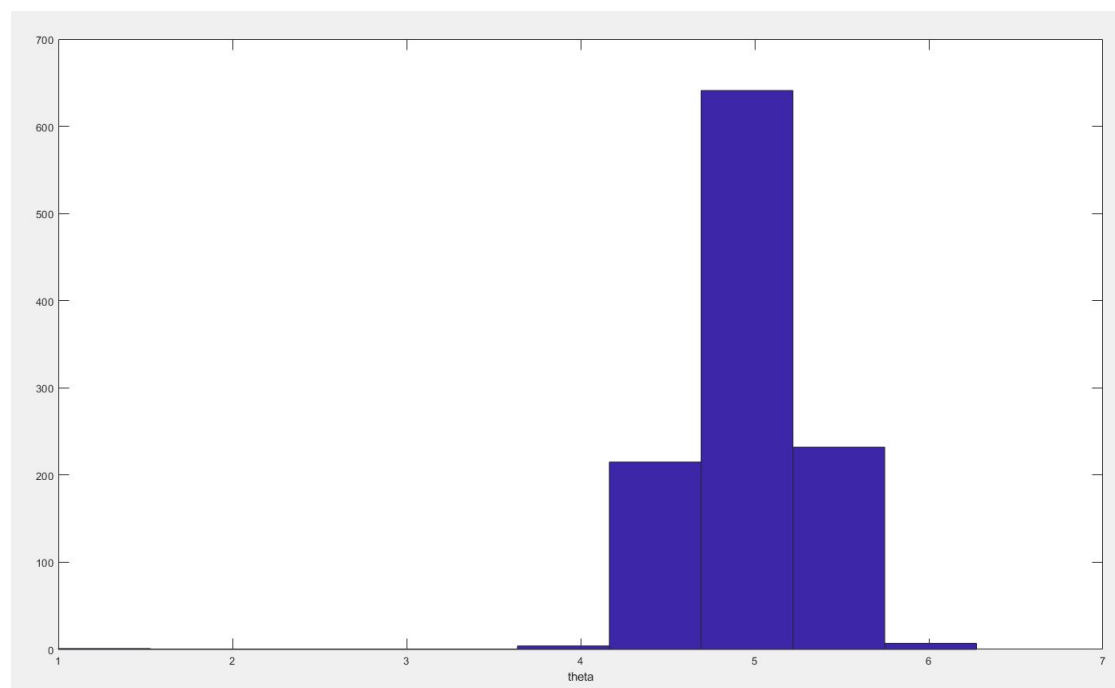
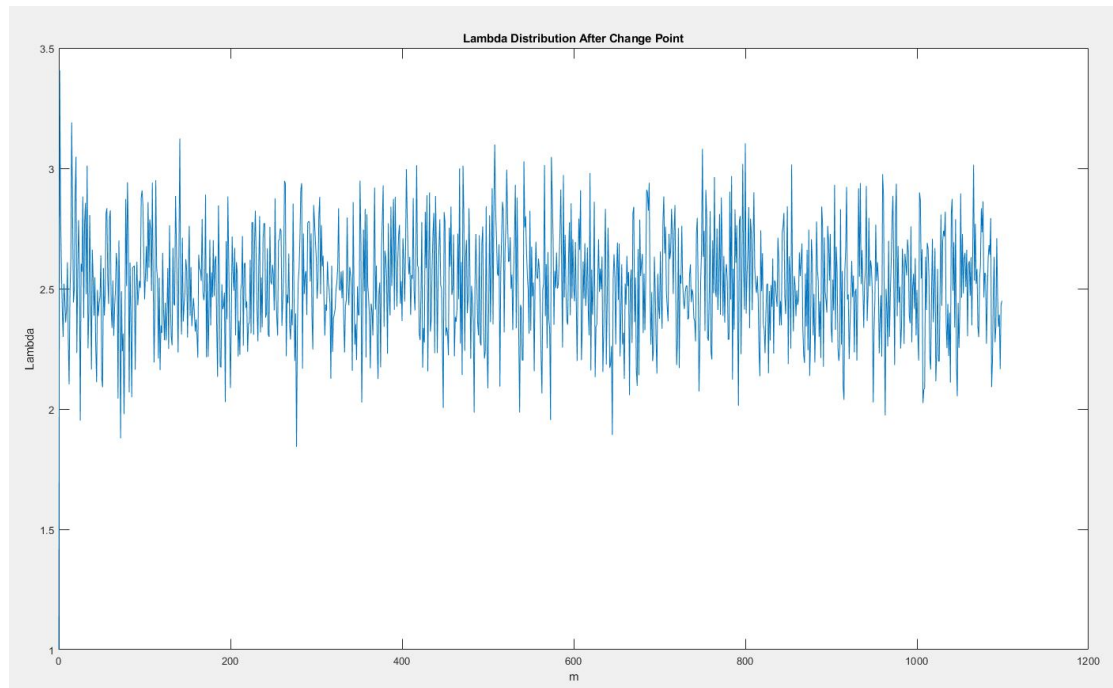**Figure 9.** Rate before change point density histogram.
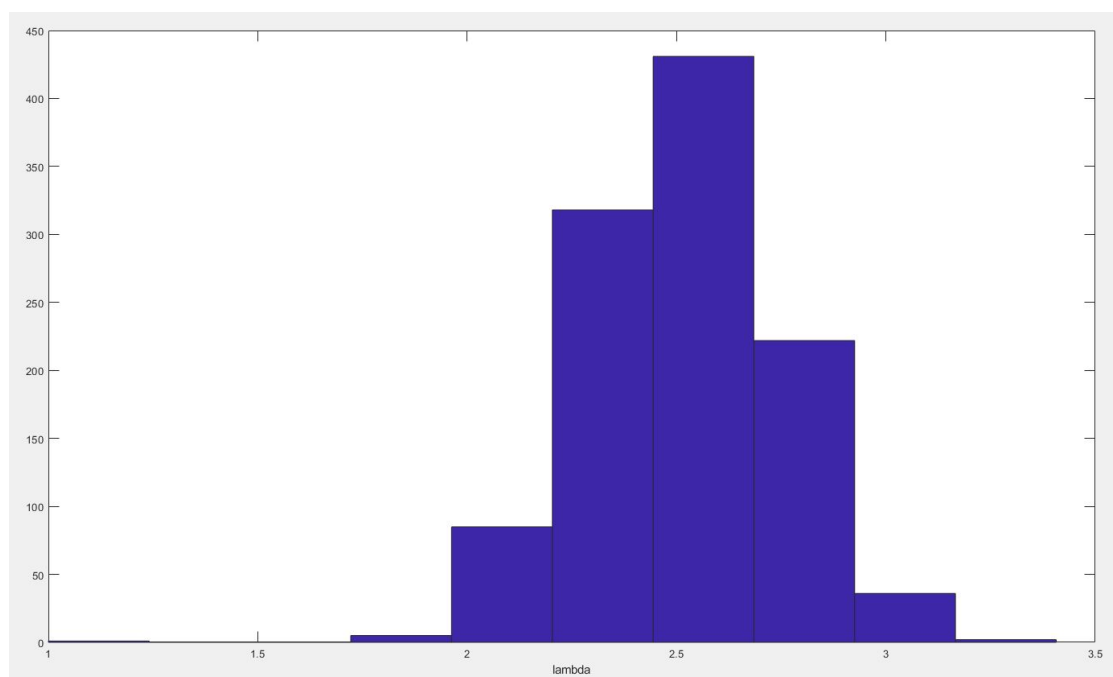
Figure 10. Rate after change point.

Figure 11. Rate after change point density histogram.

4.2. Particulate Matter (PM$_{2.5}$ and PM$_{10}$) Change Points for Four Different Sites

Particulate matter (PM$_{2.5}$ and PM$_{10}$) concentrations are considered as Normally distributed. A random variable Y is understood as Normally distributed with mean μ and variance $\sigma^2 > 0$ if the probability distribution function at any given point y in the sample space is given as follows:

$$Y \sim Normal(\mu, \sigma^2) = f(y; \mu, \sigma^2) = (p(Y = y | \mu, \sigma^2)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - \mu)^2}{2\sigma^2}} \; for \; y \in 0, 1, 2, ..., n$$

The distribution is symmetric about mean μ and σ^2 represents the variance. The numerical details for $PM_{2.5}$ and PM_{10} concentrations are given in Tables 3 and 4, respectively.

Table 3. $PM_{2.5}$ Normal distribution.

Area	(μ) Mean Is the Location	(σ^2) Variance Is Scale or Deviation
Guro	0.02675	0.01614
Nowon	0.02652	0.01551
Songpa	0.02638	0.01544
Yongsan	0.02649	0.01621

Table 4. PM_{10} Normal distribution.

Area	(μ) Mean Is the Location	(σ^2) Variance Is Scale or Deviation
Guro	0.05340	0.03644
Nowon	0.05036	0.03331
Songpa	0.05193	0.03811
Yongsan	0.05379	0.03958

Here, the results were acquired by applying the method explained in Section 3.4 to the particulate matter ($PM_{2.5}$ and PM_{10}) concentrations for four different sites (Guro, Nowon, Songpa and Yongsan) in Seoul, South Korea. The daily data observed from January 2004 to December 2013 were used to compute the change point of both pollutants. However, the particulate matter ($PM_{2.5}$ and PM_{10}) concentrations are shown in Figures 12–19.
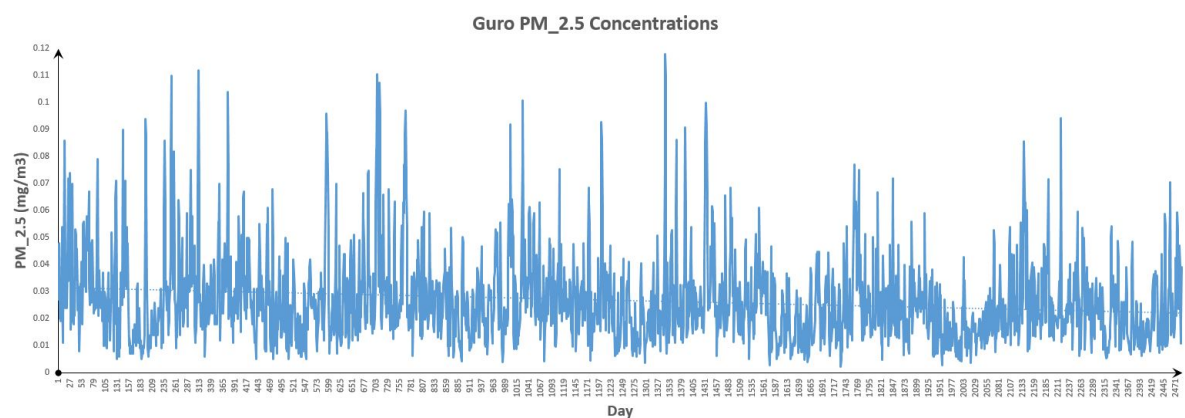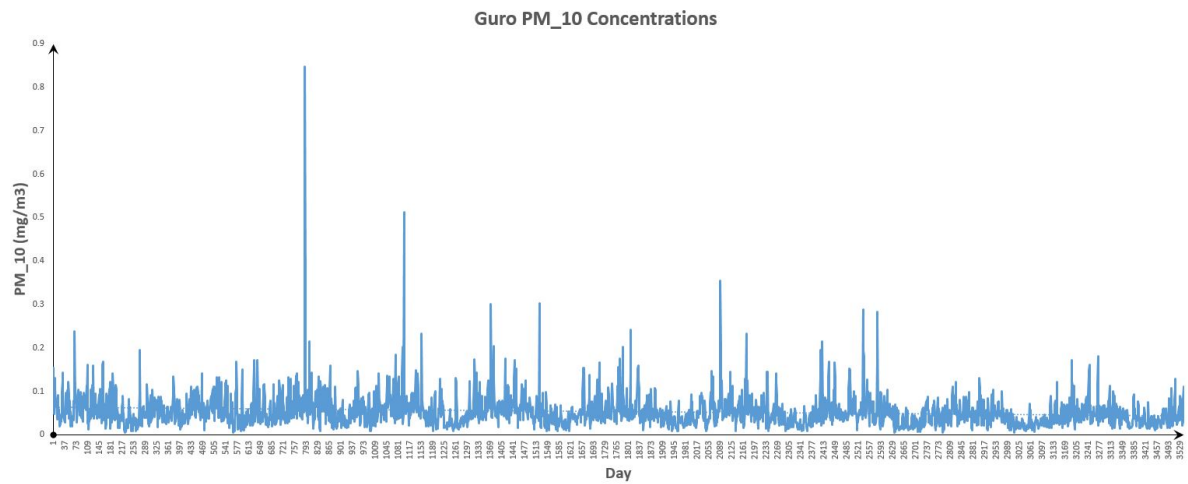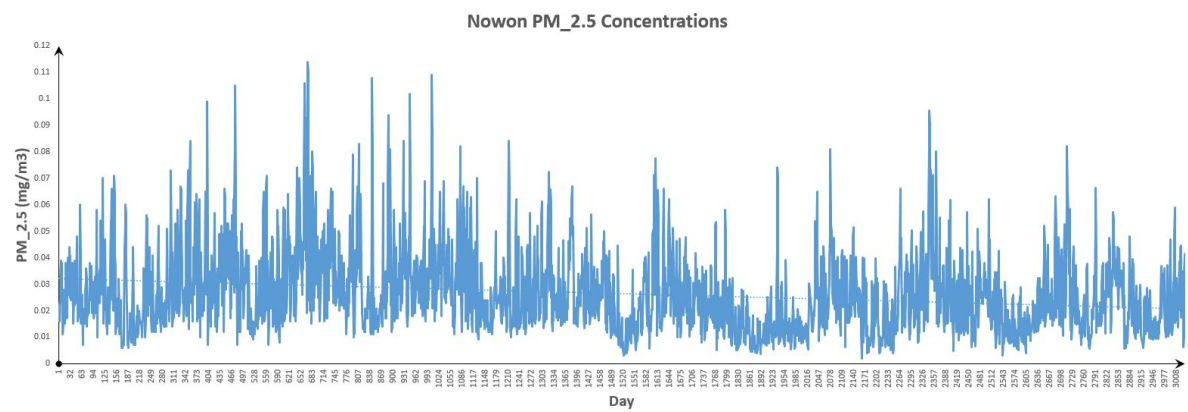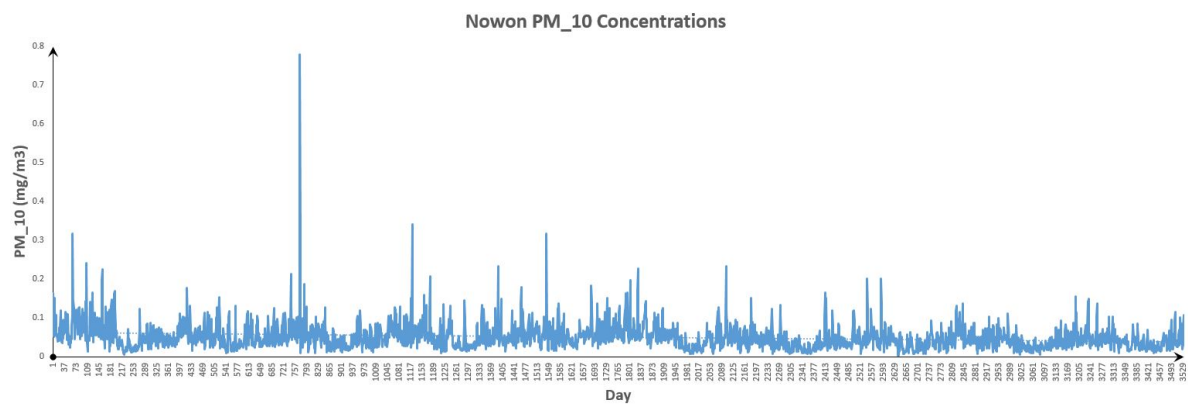

Figure 12. Guro $PM_{2.5}$ Data.

Guro PM_10 Concentrations

Figure 13. Guro PM_{10} Data.

Nowon PM_2.5 Concentrations

Figure 14. Nowon $PM_{2.5}$ Data.

Nowon PM_10 Concentrations

Figure 15. Nowon PM_{10} Data.
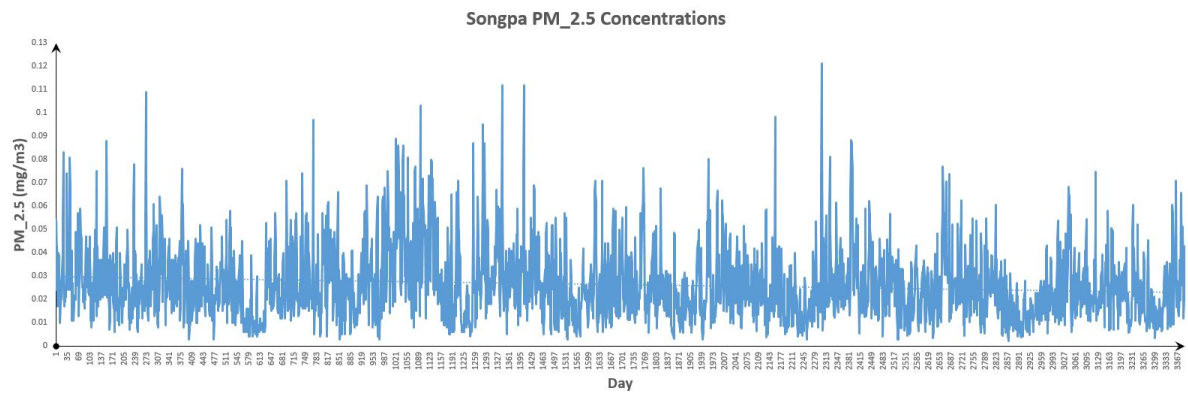
Figure 16. Songpa $PM_{2.5}$ Data.
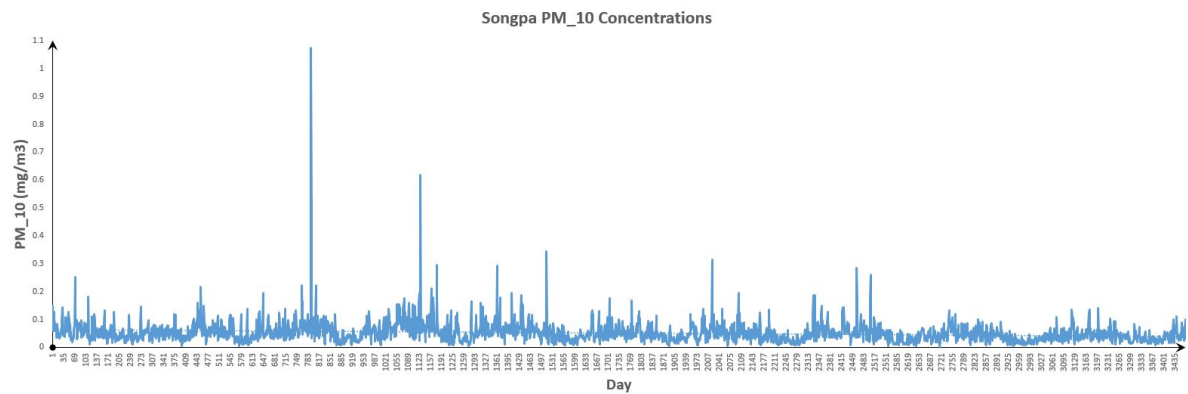
Figure 17. Songpa PM_{10} Data.
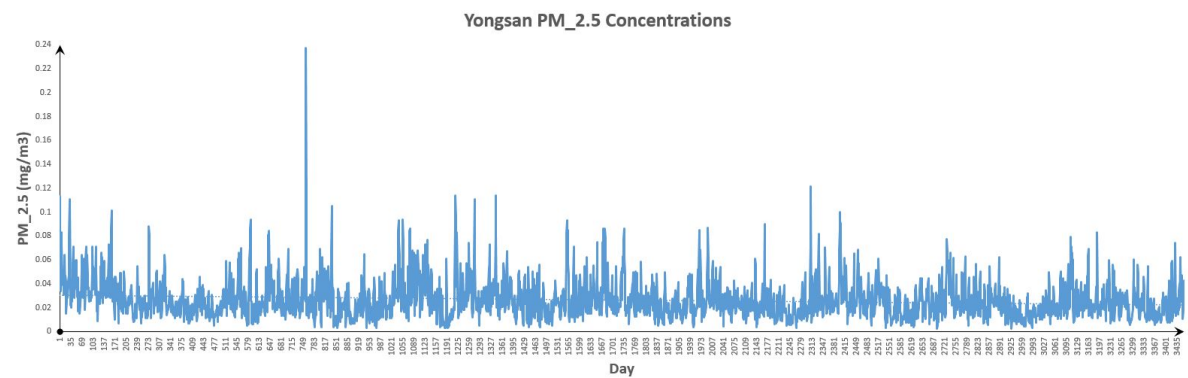
Figure 18. Yongsan $PM_{2.5}$ Data.
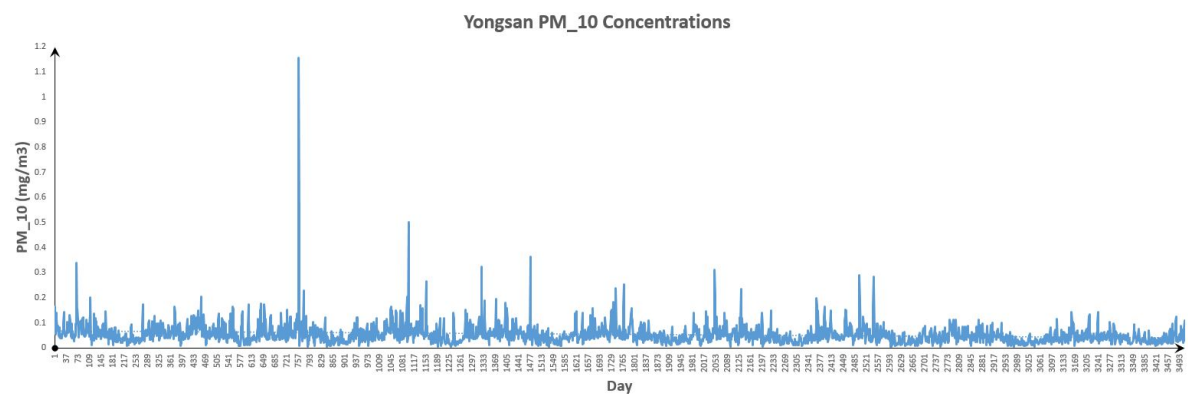
Figure 19. Yongsan PM_{10} Data.

A change point for a process with different data structures is identified to know that a change has occurred, the most likely period in which the change occurred and the parameter behavior before and after the change point. If particulate matter ($PM_{2.5}$ and PM_{10}) concentrations are Normally distributed and the change point for the random process is denoted by k, it is supposed that the data follow a Normal distribution with *mean* $= \mu$ and *variance* $= \sigma^2$ until the k point. After the k point, the data is Normally distributed with parameters mean (η) and variance (ϕ^2) and can be represented as

$$y_i \sim Normal(\mu, \sigma^2) = f(y; \mu, \sigma^2) = p(Y = y_i | \mu, \sigma^2) \text{ for } i = 1, 2, ..., k$$

$$y_i \sim Normal(\eta, \phi^2) = f(y; \eta, \phi^2) = p(Y = y_i | \eta, \phi^2) \text{ for } i = k+1, k+2, ..., n$$

Moreover, the notation " $\sim$ " means 'is distributed as.'

If the proposed model is $(Y_1, Y_2, Y_3, ..., Y_k | \mu, \sigma^2) \sim Normal(\mu, \sigma^2)$, then the joint pdf (probability density function) is given by

$$p(y_1, y_2, y_3, ..., y_k | \mu, \sigma^2) = \prod_{i=1}^{k} p(y_i | \mu, \sigma^2) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} = (2\pi\sigma^2)^{-\frac{k}{2}} exp\left[\frac{1}{2}\sum\left(\frac{y_i - \mu}{\sigma}\right)^2\right]$$

Expanding the quadratic term in the exponent, it can be seen that $p(y_1, y_2, y_3, ..., y_k | \mu, \sigma^2)$ depends on $y_1, y_2, y_3, ..., y_k$ through

$$\sum_{i=1}^{k}\left(\frac{y_i - \mu}{\sigma}\right)^2 = \sum \frac{1}{\sigma^2}y_i^2 - 2\frac{\mu}{\sigma^2}\sum y_i + k\frac{\mu^2}{\sigma^2}$$

It can be shown that $(\sum y_i^2, \sum y_i)$ make up a two-dimensional sufficient statistic. Knowing the values of these quantities is equivalent to knowing the values of $\bar{y} = \sum \frac{y_i}{k}$ and $s^2 = \frac{(y_i - \bar{y})^2}{k-1}$ and so $(\bar{y}, s^2)$ are also sufficient statistic.

Inference for this two-parameter model can be broken down into two one-parameter problems.

4.2.1. Bayesian Inference for Mean When Variance Is Known

Firstly, make an inference for μ when σ^2 is known and use a conjugate prior distribution for μ. For any (conditional) prior probability $p(\mu | \sigma^2)$, the posterior distribution will satisfy

$$p(\mu | y_1, y_2, y_3, ..., y_k, \sigma^2) \propto p(\mu | \sigma^2) \times e^{-\frac{1}{2\sigma^2}\sum(y_i - \mu)^2} \propto p(\mu | \sigma^2) \times e^{c_1(\mu - c_2)}$$

A class of prior distributions is conjugate for a likelihood or sampling model $p(y_1, y_2, y_3, ..., y_k | \mu, \sigma^2)$ if the resulting posterior distribution is also in the similar class. The above calculations indicate that, if $p(\mu | \sigma^2)$ is to be conjugate, it must include quadratic terms like $e^{c_1(\mu - c_2)^2}$. The simplest such class of probability densities on R is the Normal family of densities, suggesting that if $p(\mu | \sigma^2)$ is Normal and $y_1, y_2, y_3, ..., y_k$ are $Normal(\mu, \sigma^2)$, then $p(\mu | y_1, ..., y_k, \sigma^2)$ is also a Normal density.

Hence, the Bayesian model for parameter mean before change point μ can be given by

$$p(y_1, y_2, y_3, ..., y_k | \mu, \sigma^2) \sim Normal(\mu, \sigma^2)$$

$$\mu \sim Normal(\theta_0, \tau_0^2)$$

$$p(\mu | \sigma^2, y_1, y_2, y_3, ..., y_k) \sim Normal(\theta_k, \tau_k^2)$$

$$\theta_0 = mean \; of \; k_0 \; prior \; observations$$

Consider the particular case in which $\tau_0^2 = \frac{\sigma^2}{k_0}$

$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2) = dnorm\left(\mu, \theta_0, \tau_0 = \frac{\sigma}{\sqrt{k_0}}\right) \times p(\sigma^2)$$

The parameters θ_0 and k_0 can be interpreted as the mean and sample size, respectively, from a set of prior observations.

Similarly, the Bayesian model for the parameter mean after change point η can be given by

$$p(y_{k+1}, y_{k+2}, y_{k+3}, \ldots, y_n|\eta, \phi^2) \sim Normal(\eta, \phi^2)$$

$$\eta \sim Normal(\lambda_0, \omega_0^2)$$

$$p(\eta|\phi^2, y_{k+1}, y_{k+2}, y_{k+3}, \ldots, y_n) \sim Normal(\lambda_n, \omega_n^2)$$

$$\omega_0^2 = \frac{\phi^2}{k_0}$$

$$\omega_n^2 = \frac{1}{\frac{1}{\omega_0^2} + \frac{(n-k)}{\phi^2}}$$

$$\lambda_0 = mean\ of\ prior\ observations$$

$$\lambda_n = \frac{k_0}{k_0 + (n-k)}\lambda_0 + \frac{(n-k)}{k_0 + (n-k)}\bar{y}$$

4.2.2. Bayesian Inference for Variance (σ^2, ϕ^2)

For σ^2, a family of prior distributions is required with support on $(0, \infty)$. One such family of distributions is the Gamma family; unfortunately, this family is not conjugate for the Normal variance. However, the Gamma family does turn out to be a conjugate class of densities for $\frac{1}{\sigma^2}$ (the precision). When using such a prior distribution, σ^2 has an Inverse-Gamma distribution. For interpret ability later, instead of using a_1 and b_1, this prior distribution of σ^2 can be parameterized as:

$$Precision = \frac{1}{\sigma^2} \sim Gamma(a_1, b_1) \sim Gamma\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right)$$

$$Variance\ before\ change\ point = \sigma^2 \sim Inverse-Gamma(a_1, b_1) \sim Gamma\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right)$$

$$Variance\ after\ change\ point = \phi^2 \sim Inverse-Gamma(a_2, b_2) \sim Gamma\left(\frac{v_0}{2}, \frac{v_0\phi_0^2}{2}\right)$$

The prior parameters (σ_0^2, v_0) can be interpreted as the sample variance and sample size of prior observations, respectively, for posterior inference, the prior distributions and sampling model are as follows:

$$\frac{1}{\sigma^2} \sim Gamma\left(\frac{v_0}{2}, \frac{v_0\sigma_0^2}{2}\right)$$

$$(\mu|\sigma^2) \sim Normal\left(\theta_0, \frac{\sigma_0^2}{k_0}\right)$$

$$(y_1, y_2, y_3, \ldots, y_k|\mu, \sigma^2) \sim Normal(\mu, \sigma^2)$$

After a change point, the prior distributions and sampling model are

$$\frac{1}{\phi^2} \sim Gamma\left(\frac{\nu_0}{2}, \frac{\nu_0\phi_0^2}{2}\right)$$

$$(\eta|\phi^2) \sim Normal\left(\lambda_0, \frac{\phi_0^2}{k_0}\right)$$

$$(y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n|\eta, \phi^2) \sim Normal(\eta, \phi^2)$$

4.2.3. Joint Inference for Mean and Variance

Just as the prior distribution for μ and σ^2 can be decomposed as $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$, the posterior distribution can be similarly decomposed:

$$p(\mu, \sigma^2|y_1, y_2, y_3, ..., y_k) = p(\mu|\sigma^2, y_1, y_2, y_3, ..., y_k)p(\sigma^2|y_1, y_2, y_3, ..., y_k)$$

$$p(\eta, \phi^2|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n) = p(\eta|\phi^2, y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)p(\phi^2|y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n)$$

The conditional distribution of μ given σ^2 and the data $(y_1, y_2, y_3, ..., y_k)$ has already been obtained:

$$p(\mu|\sigma^2, y_1, y_2, y_3, ..., y_k) \sim Normal(\theta_k, \tau_k^2) \sim Normal(\theta_k, \frac{\sigma^2}{k_k})$$

where $k_k = k_0 + k$ and $\theta_k = \left(\frac{k_0}{k_0+k}\theta_0 + \frac{k}{k_0+k}\bar{y}\right) = \left(\frac{k_0\theta_0+k\bar{y}}{k_k}\right)$

Therefore, if θ_0 is the mean of k_0 prior observations, then $E(\mu|\sigma^2, y_1, ..., y_k)$ is the sample mean of the current and prior observations and $Var(\mu|\sigma^2, y_1, ..., y_k)$ is σ^2 divided by the total number of observations, both prior and current.

$$p(\eta|\phi^2, y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n) \sim Normal(\lambda_n, \omega_n^2) \sim Normal(\lambda_n, \frac{\phi^2}{k_n})$$

where $k_n = k_0 + (n-k)$ and $\lambda_n = \left(\frac{k_0}{k_0+(n-k)}\lambda_0 + \frac{(n-k)}{k_0+(n-k)}\bar{y}\right) = \left(\frac{k_0\lambda_0+(n-k)\bar{y}}{k_n}\right)$

The posterior distribution of σ^2 can be found by estimating an integration over the unknown value of μ:

$$p(\sigma^2|y_1, ..., y_k) \propto p(\sigma^2)p(y_1, ..., y_k|\sigma^2) = p(\sigma^2)\int p(y_1, ..., y_k|\mu, \sigma^2)p(\mu|\sigma^2)d\mu$$

$$\left(\frac{1}{\sigma^2}|y_1, ..., y_k\right) \sim Gamma\left(\frac{\nu_k}{2}, \frac{\nu_k\sigma_k^2}{2}\right)$$

where $\nu_k = \nu_0 + k$ and $\sigma_k^2 = \frac{1}{\nu_k}\left[\nu_0\sigma_0^2 + (k-1)s^2 + \frac{k_0k}{k_k}(\bar{y} - \theta_0)^2\right]$ It has been shown that variance before change point (σ^2) and variance after change point (ϕ^2) follow an Inverse-Gamma distribution. Prior and posterior distributions for variance parameters are given as follows:

$$\sigma^2 \sim Inverse - Gamma\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)$$

$$\left(\sigma^2|y_1, y_2, y_3, ..., y_k\right) \sim Inverse - Gamma\left(\frac{\nu_k}{2}, \frac{\nu_k\sigma_k^2}{2}\right)$$

$$\sim Inverse - Gamma\left(\frac{\nu_0+k}{2}, \frac{(\nu_0+k)}{2}\frac{1}{(\nu_0+k)}\left[\nu_0\sigma_0^2 + (k-1)s^2 + \frac{k_0k}{k_k}(\bar{y} - \theta_0)^2\right]\right)$$

$$\sim Inverse - Gamma\left(\frac{\nu_0+k}{2}, \frac{1}{2}\left[\nu_0\sigma_0^2 + (k-1)s^2 + \frac{k_0k}{k_0+k}(\bar{y} - \theta_0)^2\right]\right) \quad (4)$$

Similarly, the variance after change point (ϕ^2) is given as

$$\phi^2 \sim Inverse - Gamma\left(\frac{\nu_0}{2}, \frac{\nu_0\phi_0^2}{2}\right)$$

$$\left(\phi^2 | y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n\right) \sim Inverse - Gamma\left(\frac{\nu_n}{2}, \frac{\nu_n\phi_n^2}{2}\right)$$

$$\sim Inverse - Gamma\left(\frac{\nu_0 + (n-k)}{2}, \frac{(\nu_0 + (n-k))}{2} \frac{1}{(\nu_0 + (n-k))}\right.$$
$$\left. \left[\nu_0\phi_0^2 + ((n-k)-1)s^2 + \frac{k_0(n-k)}{k_n}(\bar{y} - \lambda_0)^2\right]\right)$$

$$\sim Inverse - Gamma\left(\frac{\nu_0 + (n-k)}{2}, \frac{1}{2}\left[\nu_0\phi_0^2 + ((n-k)-1)s^2 + \frac{k_0(n-k)}{k_0 + (n-k)}(\bar{y} - \lambda_0)^2\right]\right) \quad (5)$$

4.2.4. Improper Priors

Since k_0 and ν_0 are prior sample sizes, the smaller are these parameters, the more objective will be the estimates. The posterior distribution as k_0 and ν_0 get smaller and smaller;

$$\theta_k = \frac{k_0\theta_0 + k\bar{y}}{k_0 + k}$$

$$\sigma_k^2 = \frac{1}{(\nu_0 + k)}\left[\nu_0\sigma_0^2 + (k-1)s^2 + \frac{k_0 k}{k_k}(\bar{y} - \theta_0)^2\right]$$

Thus, $k_0, \nu_0 \to 0$

$$\theta_k \to \bar{y}$$

and

$$\sigma_k^2 \to \left[\frac{(k-1)}{k}s^2\right] = \left[\frac{1}{k}\sum(y_i - \bar{y})^2\right]$$

Improper priors has led to the following posterior distribution for variance:

$$\left(\frac{1}{\sigma^2} | y_1, y_2, y_3, ..., y_k\right) \sim Gamma\left(\frac{k}{2}, \frac{1}{2}\left[\sum_{i=1}^{k}(y_i - \bar{y})^2\right]\right)$$

$$\left(\frac{1}{\phi^2} | y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n\right) \sim Gamma\left(\frac{n-k}{2}, \frac{1}{2}\left[\sum_{i=k+1}^{n}(y_i - \bar{y})^2\right]\right)$$

Similarly, the posterior distribution for mean is given as

$$\left(\mu | \sigma^2, y_1, y_2, y_3, ..., y_k\right) \sim Normal\left(\frac{1}{k}\sum_{i=1}^{k}y_i, \frac{\sigma^2}{k}\right)$$

$$\left(\eta | \phi^2, y_{k+1}, y_{k+2}, y_{k+3}, ..., y_n\right) \sim Normal\left(\frac{1}{n-k}\sum_{i=k+1}^{n}y_i, \frac{\phi^2}{n-k}\right)$$

4.2.5. Likelihood Ratio Test and Likelihood Function

As the expected value of a Normal distribution is the mean. The following likelihood ratio test needs to be applied for change point detection:

$$f(k|y, \mu, \eta, \sigma^2, \phi^2) = \frac{L(Y; k, \mu, \eta)}{\sum_{j=1}^{n} L(Y; j, \mu, \eta)}$$

The likelihood function for the expected value of a Normal distribution is determined as

$$L(Y; k, \mu, \eta) = exp\left(k(\eta - \mu)\right)\left(\frac{\mu}{\eta}\right)^{\sum_{i=1}^{k} y_i}$$

For the probabilistic method, MATLAB was used for change point detection of particulate matter ($PM_{2.5}$ and PM_{10}) data during the study period 2004–2013 for four different sites (Guro, Nowon, Songpa and Yongsan) in Seoul, South Korea. Ten replications of each simulation were performed with 1100 observations in each replication. The first 100 observations are discarded as a burn-in period. Replication of the mean V_i of the remaining 1000 observations was performed for each replication, as shown in Tables 5 and 6. Mean (V) ofthe replication mean was used to get the converged values of parameters.

Table 5. $PM_{2.5}$ Converged values of parameters (Probabilistic Method)

$PM_{2.5}$ Converged Values of Parameters (Probabilistic Method)										
	Guro					Nowon				
Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean
V_1	0.0293	0.0246	3.98×10^{-30}	1.41×10^{-30}	1229.02	0.0296	0.0235	1.17×10^{-30}	5.16×10^{-31}	1497.30
V_2	0.0292	0.0246	4.37×10^{-30}	1.28×10^{-30}	1261.72	0.0296	0.0235	1.16×10^{-30}	5.17×10^{-31}	1529.49
V_3	0.0294	0.0246	4.14×10^{-30}	1.46×10^{-30}	1199.08	0.0296	0.0236	1.16×10^{-30}	5.62×10^{-31}	1475.20
V_4	0.0294	0.0246	4.00×10^{-30}	1.40×10^{-30}	1214.48	0.0295	0.0236	1.12×10^{-30}	5.46×10^{-31}	1485.39
V_5	0.0292	0.0246	4.40×10^{-30}	1.34×10^{-30}	1232.26	0.0295	0.0235	1.16×10^{-30}	5.23×10^{-31}	1527.50
V_6	0.0294	0.0246	3.75×10^{-30}	1.48×10^{-30}	1202.37	0.0295	0.0236	1.09×10^{-30}	5.75×10^{-31}	1478.39
V_7	0.0293	0.0247	3.20×10^{-30}	1.45×10^{-30}	1236.93	0.0294	0.0236	1.11×10^{-30}	5.72×10^{-31}	1497.28
V_8	0.0292	0.0246	4.36×10^{-30}	1.32×10^{-30}	1248.80	0.0295	0.0235	1.14×10^{-30}	4.95×10^{-31}	1524.84
V_9	0.0293	0.0245	4.13×10^{-30}	1.33×10^{-30}	1253.42	0.0295	0.0235	1.12×10^{-30}	5.32×10^{-31}	1525.16
V_{10}	0.0294	0.0247	4.07×10^{-30}	1.49×10^{-30}	1206.30	0.0294	0.0236	1.13×10^{-30}	5.98×10^{-31}	1487.82
(V) Mean of 10 replications	0.0293	0.0246	4.14×10^{-30}	1.39×10^{-30}	1228.44	0.0295	0.0235	1.14×10^{-30}	5.44×10^{-31}	1502.84
	Songpa					Yongsan				
Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean
V_1	0.0280	0.0245	1.01×10^{-30}	1.89×10^{-30}	1788.62	0.0302	0.0248	1.12×10^{-30}	3.26×10^{-30}	1563.53
V_2	0.0281	0.0245	9.82×10^{-31}	1.71×10^{-30}	1753.39	0.0302	0.0248	1.09×10^{-30}	3.25×10^{-30}	1534.23
V_3	0.0282	0.0245	9.45×10^{-31}	1.92×10^{-30}	1709.28	0.0308	0.0249	1.00×10^{-30}	3.33×10^{-30}	1457.41
V_4	0.0281	0.0246	1.05×10^{-30}	1.94×10^{-30}	1725.05	0.0308	0.0249	9.87×10^{-31}	3.19×10^{-30}	1482.48
V_5	0.0281	0.0245	9.19×10^{-31}	1.97×10^{-30}	1774.62	0.0303	0.0249	9.59×10^{-31}	2.95×10^{-30}	1543.11
V_6	0.0282	0.0246	9.35×10^{-31}	2.16×10^{-30}	1702.20	0.0308	0.0249	9.76×10^{-31}	3.27×10^{-30}	1476.89
V_7	0.0282	0.0247	9.58×10^{-31}	1.90×10^{-30}	1732.44	0.0313	0.0249	9.10×10^{-31}	3.24×10^{-30}	1458.11
V_8	0.0280	0.0246	1.01×10^{-30}	1.80×10^{-30}	1768.74	0.0303	0.0248	1.10×10^{-30}	3.30×10^{-30}	1552.81
V_9	0.0281	0.0245	1.02×10^{-30}	1.96×10^{-30}	1785.43	0.0304	0.0248	1.02×10^{-30}	3.04×10^{-30}	1545.42
V_{10}	0.0282	0.0245	9.74×10^{-31}	2.06×10^{-30}	1718.40	0.0316	0.0249	9.77×10^{-31}	3.38×10^{-30}	1397.51
(V) Mean of 10 replications	0.0281	0.0246	9.80×10^{-31}	1.93×10^{-30}	1745.82	0.0307	0.0248	1.01×10^{-30}	3.22×10^{-30}	1501.15

Table 6. PM_{10} Converged values of parameters (Probabilistic Method)

	PM_{10} Converged values of parameters (Probabilistic Method)									
	Guro					Nowon				
Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean
V_1	0.0580	0.0472	9.96×10^{-30}	5.48×10^{-29}	1951.13	0.0571	0.0439	1.08×10^{-29}	4.40×10^{-30}	1815.41
V_2	0.0578	0.0468	5.59×10^{-29}	8.24×10^{-30}	2074.00	0.0566	0.0436	1.10×10^{-29}	3.79×10^{-30}	1903.76
V_3	0.0579	0.0471	5.29×10^{-29}	9.69×10^{-30}	1985.46	0.0572	0.0439	1.07×10^{-29}	4.80×10^{-30}	1817.89
V_4	0.0579	0.0470	5.51×10^{-29}	9.20×10^{-30}	2021.47	0.0571	0.0437	1.15×10^{-29}	4.10×10^{-30}	1865.10
V_5	0.0578	0.0469	5.64×10^{-29}	8.29×10^{-30}	2063.70	0.0562	0.0434	1.19×10^{-29}	3.50×10^{-30}	1967.87
V_6	0.0580	0.0470	5.52×10^{-29}	9.02×10^{-30}	2015.71	0.0575	0.0439	1.07×10^{-29}	4.49×10^{-30}	1826.11
V_7	0.0580	0.0470	5.37×10^{-29}	8.90×10^{-30}	2018.96	0.0571	0.0438	1.12×10^{-29}	4.18×10^{-30}	1848.09
V_8	0.0578	0.0469	5.32×10^{-29}	8.72×10^{-30}	2057.58	0.0565	0.0436	1.13×10^{-29}	4.04×10^{-30}	1914.57
V_9	0.0578	0.0467	5.88×10^{-29}	8.17×10^{-30}	2076.96	0.0571	0.0435	1.12×10^{-29}	3.99×10^{-30}	1883.78
V_{10}	0.0580	0.0471	5.46×10^{-29}	9.72×10^{-30}	1986.92	0.0581	0.0440	1.03×10^{-29}	5.14×10^{-30}	1762.35
(V) Mean of 10 replications	0.0579	0.0470	5.06×10^{-29}	1.35×10^{-29}	2025.19	0.0571	0.0437	1.11×10^{-29}	4.24×10^{-30}	1860.49
	Songpa					Yongsan				
Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean	(μ) Replication mean	(η) Replication mean	(σ^2) Replication mean	(ϕ^2) Replication mean	K Replication mean
V_1	0.0578	0.0438	6.63×10^{-29}	5.39×10^{-30}	1965.25	0.0597	0.0457	4.48×10^{-29}	5.65×10^{-30}	2006.82
V_2	0.0576	0.0435	7.07×10^{-29}	4.35×10^{-30}	2034.64	0.0596	0.0453	4.45×10^{-29}	5.33×10^{-30}	2055.41
V_3	0.0576	0.0437	6.84×10^{-29}	5.13×10^{-30}	1997.23	0.0599	0.0457	4.41×10^{-29}	6.31×10^{-30}	1968.12
V_4	0.0577	0.0436	6.93×10^{-29}	4.90×10^{-30}	2019.75	0.0598	0.0455	4.55×10^{-29}	5.70×10^{-30}	2000.97
V_5	0.0575	0.0435	7.30×10^{-29}	3.16×10^{-30}	2068.54	0.0595	0.0454	4.56×10^{-29}	5.35×10^{-30}	2064.62
V_6	0.0575	0.0437	7.14×10^{-29}	6.09×10^{-30}	2019.05	0.0598	0.0456	4.37×10^{-29}	6.30×10^{-30}	1996.19
V_7	0.0576	0.0436	7.13×10^{-29}	5.60×10^{-30}	2030.35	0.0597	0.0455	4.62×10^{-29}	5.96×10^{-30}	2022.05
V_8	0.0575	0.0434	7.25×10^{-29}	4.71×10^{-30}	2055.77	0.0596	0.0454	4.44×10^{-29}	5.38×10^{-30}	2042.75
V_9	0.0574	0.0435	7.42×10^{-29}	4.98×10^{-30}	2062.38	0.0596	0.0453	4.75×10^{-29}	5.64×10^{-30}	2055.56
V_{10}	0.0575	0.0437	7.07×10^{-29}	6.89×10^{-30}	2002.31	0.0599	0.0456	4.43×10^{-29}	6.34×10^{-30}	1982.16
(V) Mean of 10 replications	0.0576	0.0436	7.08×10^{-29}	5.12×10^{-30}	2025.53	0.0597	0.0455	4.51×10^{-29}	5.80×10^{-30}	2019.46

Moreover, the CUSUM charts of particulate matter ($PM_{2.5}$ and PM_{10}) concentrations are shown in Figures 20–27 for the four different sites Guro, Nowon, Songpa and Yongsan in Seoul, South Korea.
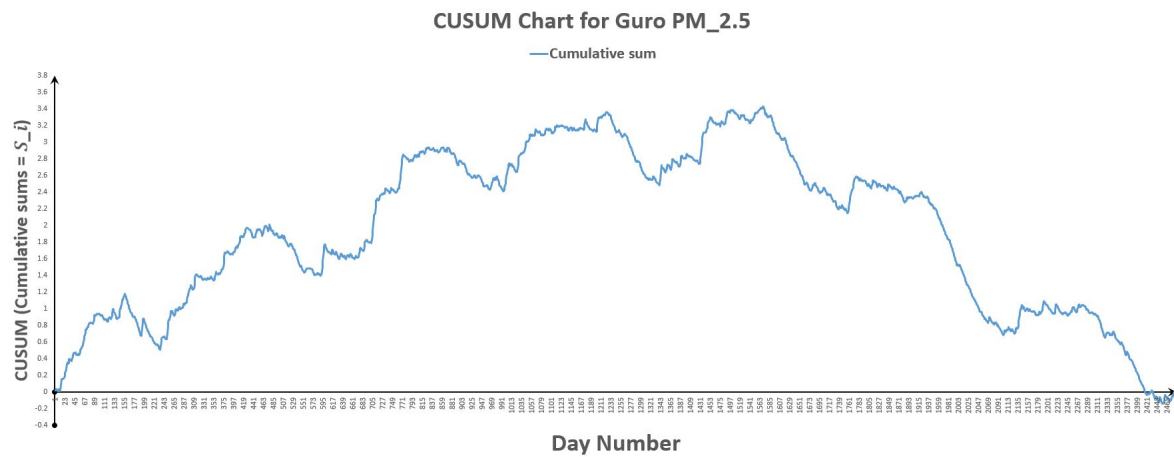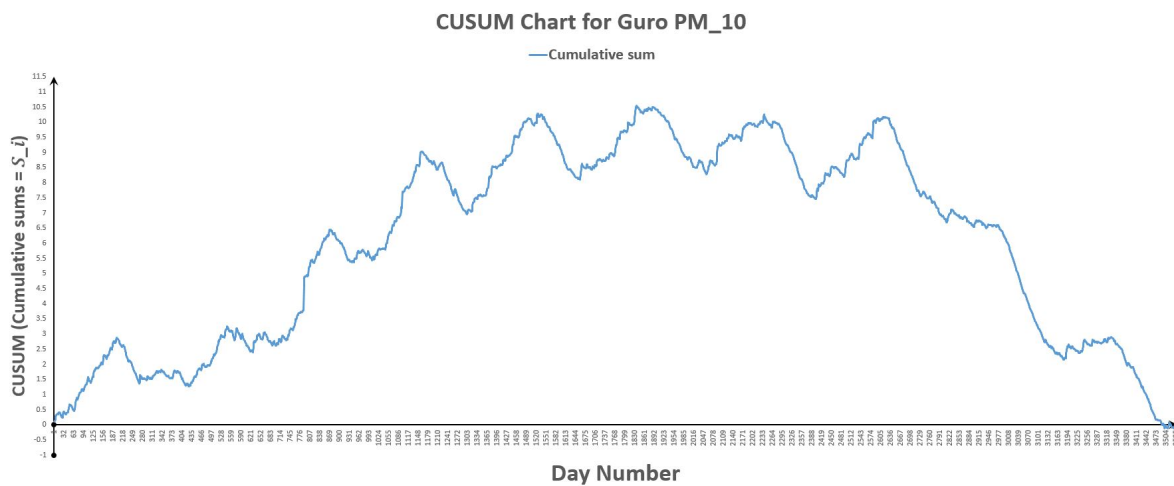

Figure 20. CUSUM chart for Guro $PM_{2.5}$.

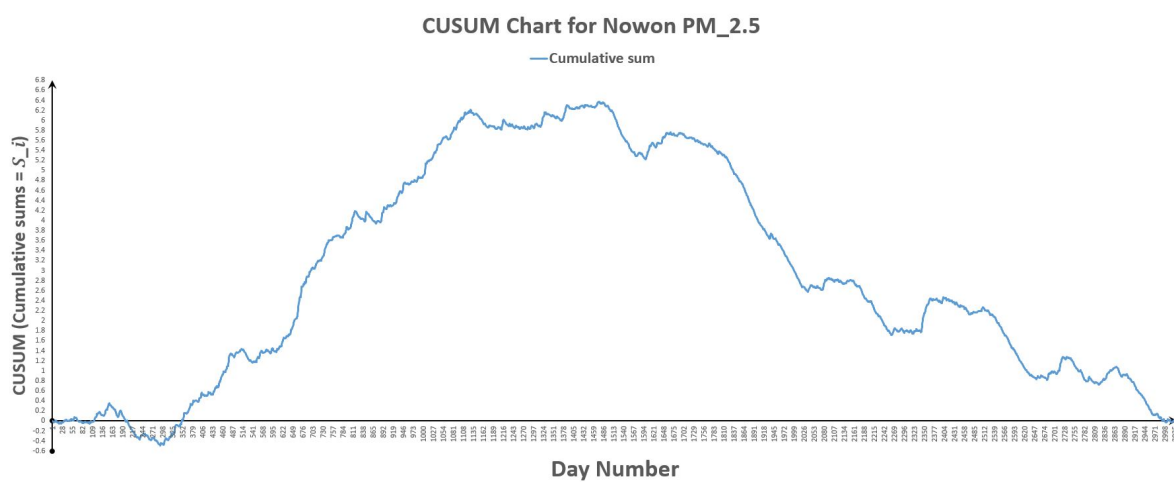**Figure 21.** CUSUM chart for Guro PM_{10}.
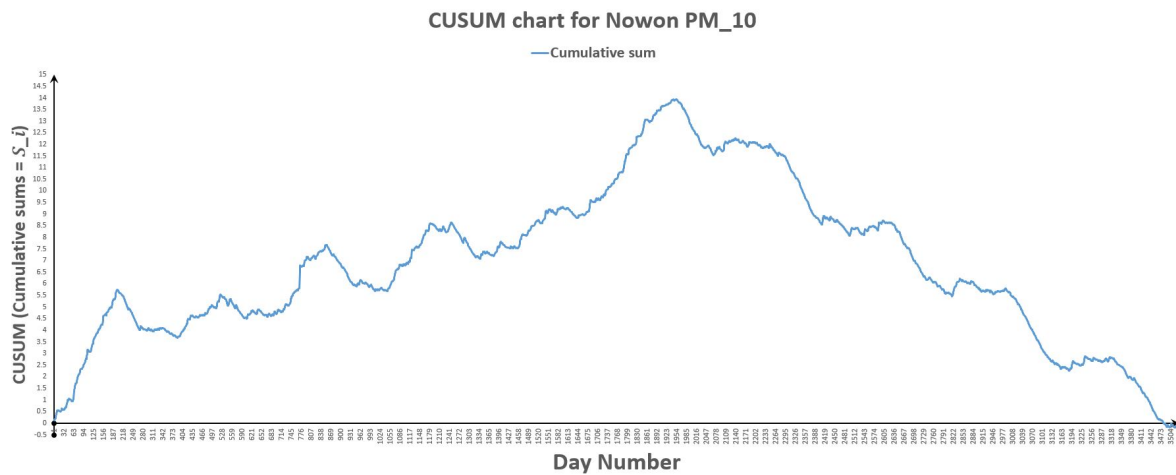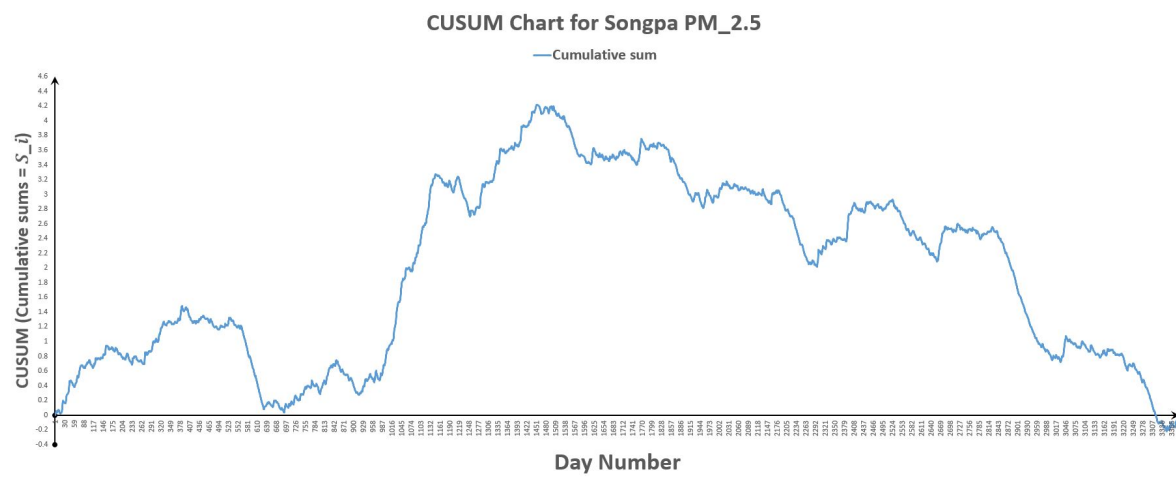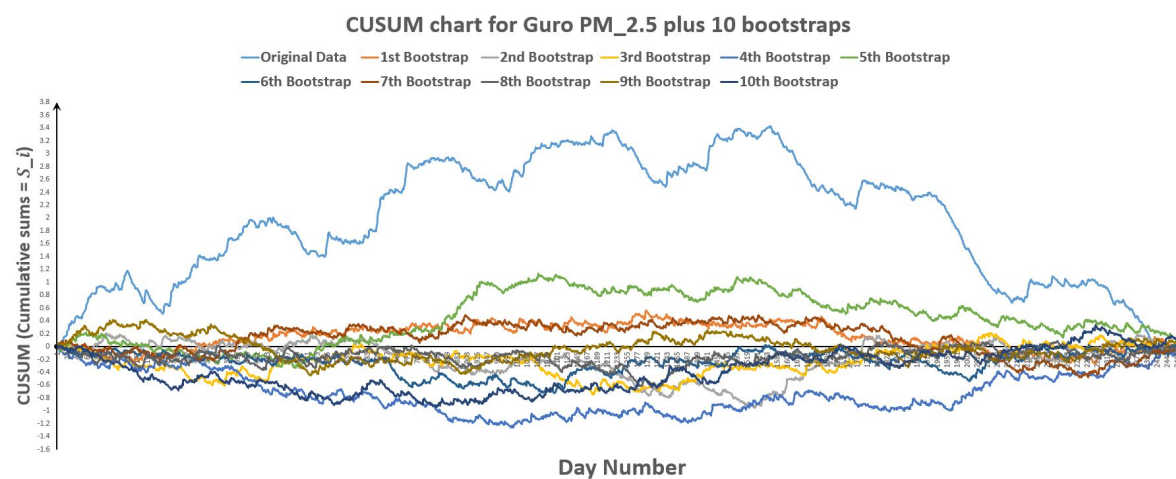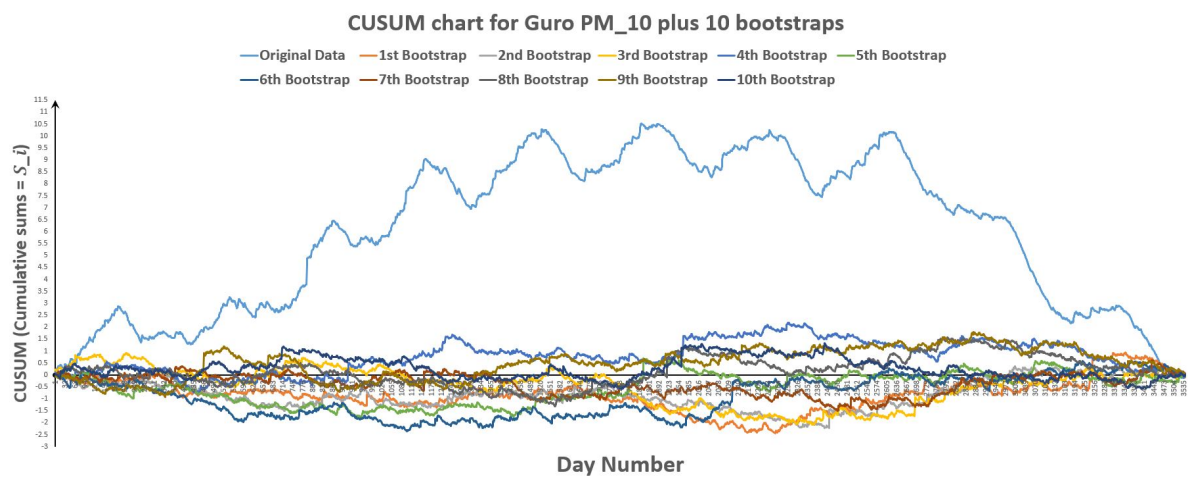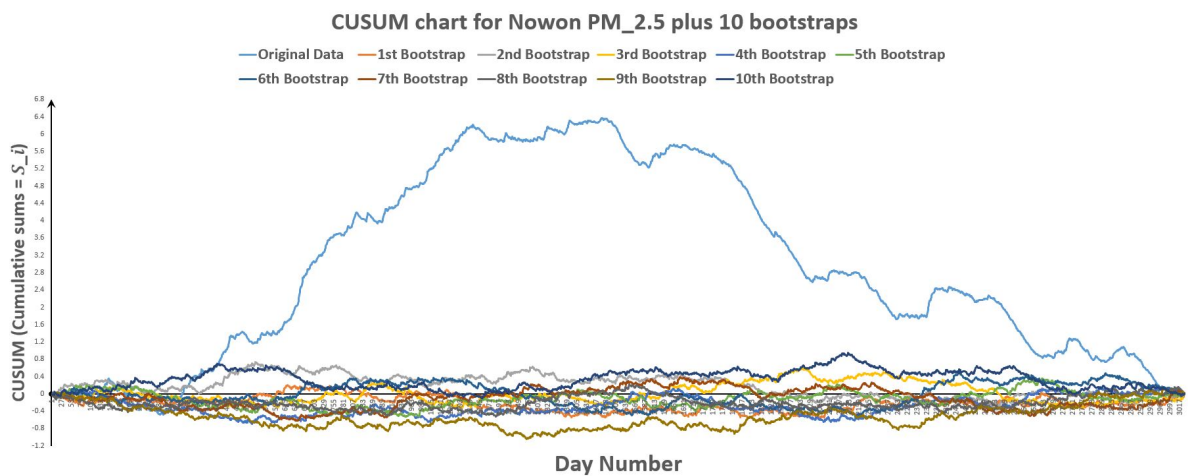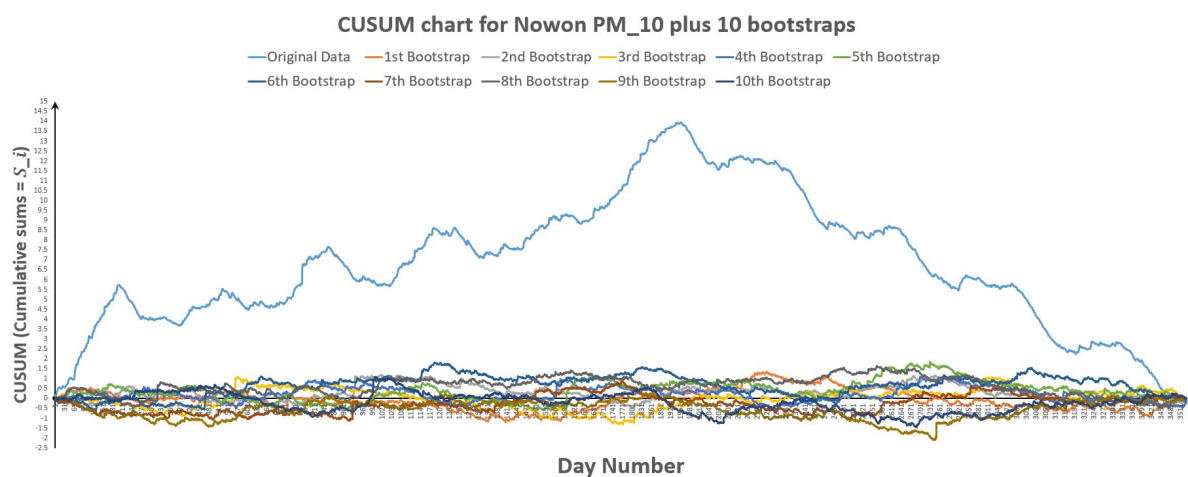

Figure 22. CUSUM chart for Nowon $PM_{2.5}$.

CUSUM chart for Nowon PM_10

Figure 23. CUSUM chart for Nowon PM_{10}.

CUSUM Chart for Songpa PM_2.5

Figure 24. CUSUM chart for Songpa $PM_{2.5}$.

CUSUM chart for Songpa PM_10
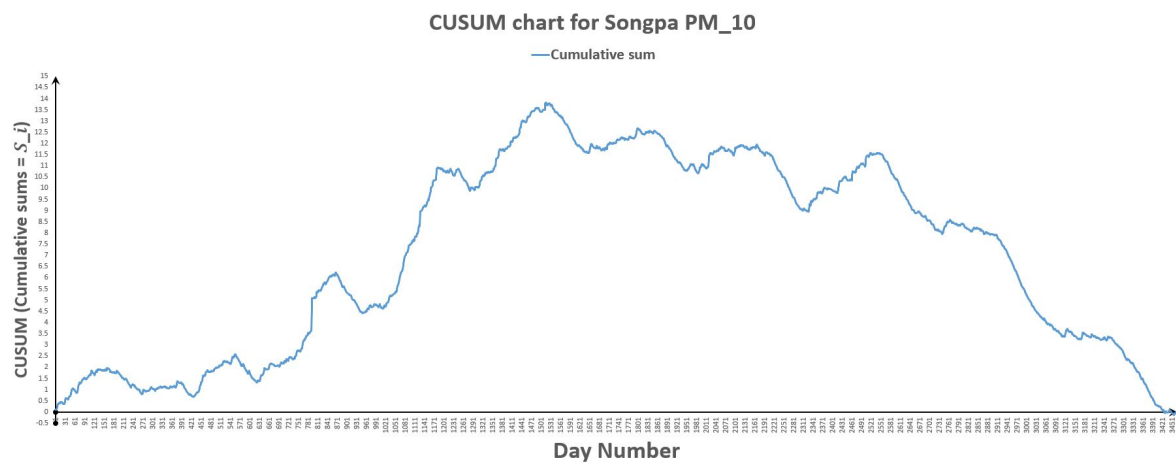
Figure 25. CUSUM chart for Songpa PM_{10}.
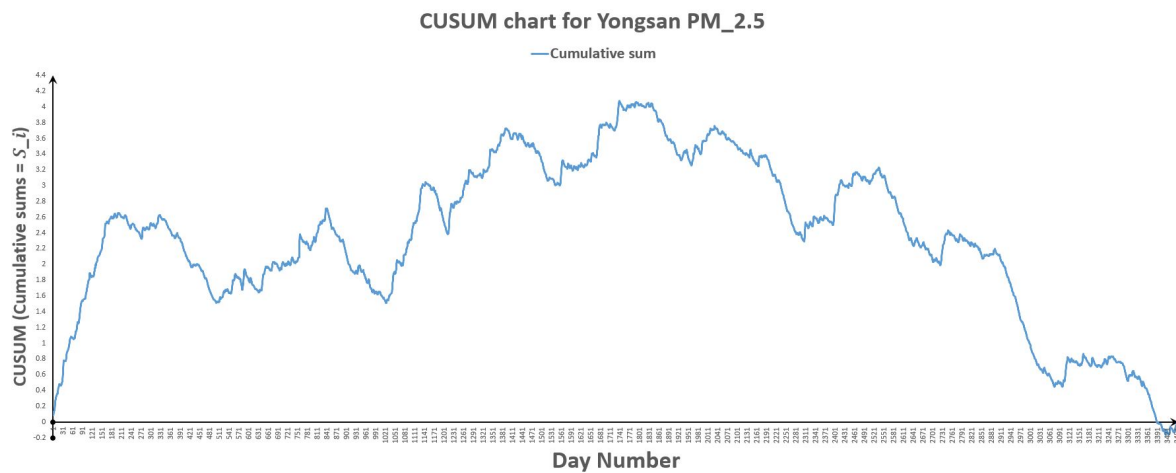
Figure 26. CUSUM chart for Yongsan $PM_{2.5}$.
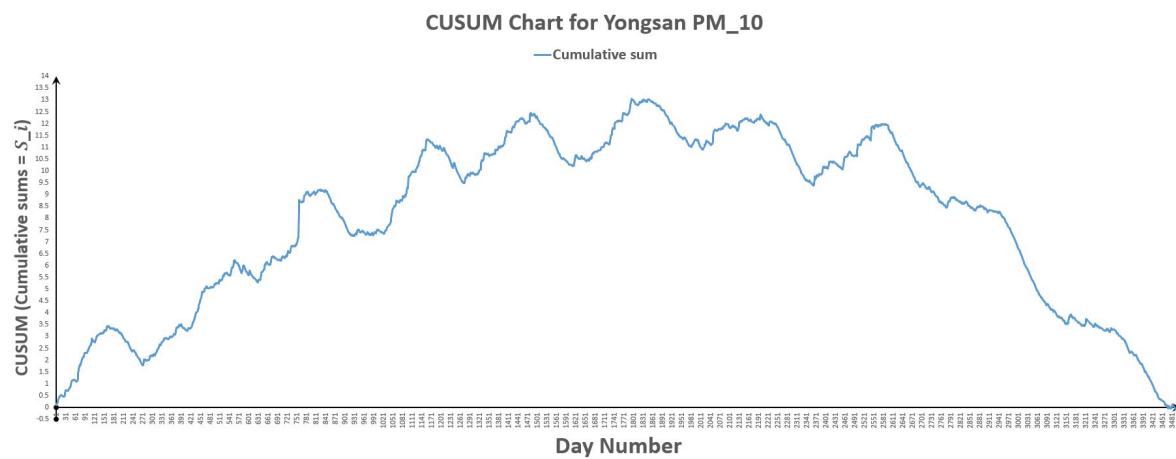
Figure 27. CUSUM chart for Yongsan PM_{10}.

In addition, the bootstraps analysis of CUSUM charts are shown in Figures 28–35.

Figure 28. CUSUM chart for Guro $PM_{2.5}$ plus 10 bootstraps.

CUSUM chart for Guro PM_10 plus 10 bootstraps

Figure 29. CUSUM chart for Guro PM_{10} plus 10 bootstraps.

CUSUM chart for Nowon PM_2.5 plus 10 bootstraps

Figure 30. CUSUM chart for Nowon $PM_{2.5}$ plus 10 bootstraps.

CUSUM chart for Nowon PM_10 plus 10 bootstraps

Figure 31. CUSUM chart for Nowon PM_{10} plus 10 bootstraps.

CUSUM chart for Songpa PM_2.5 plus 10 bootstraps
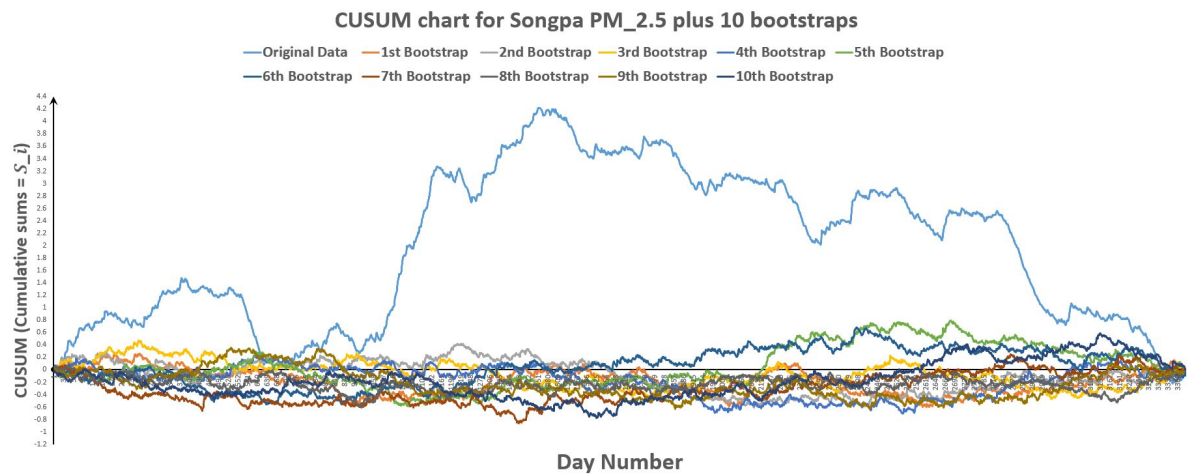
Figure 32. CUSUM chart for Songpa $PM_{2.5}$ plus 10 bootstraps.

CUSUM chart for Songpa PM_10 plus 10 bootstraps
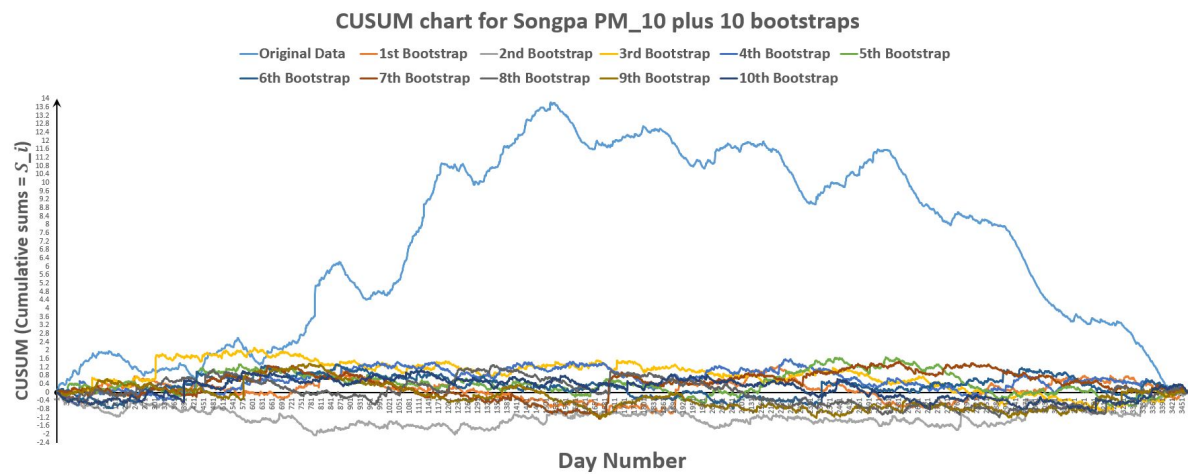
Figure 33. CUSUM chart for Songpa PM_{10} plus 10 bootstraps.

CUSUM chart for Yongsan PM_2.5 plus 10 bootstraps
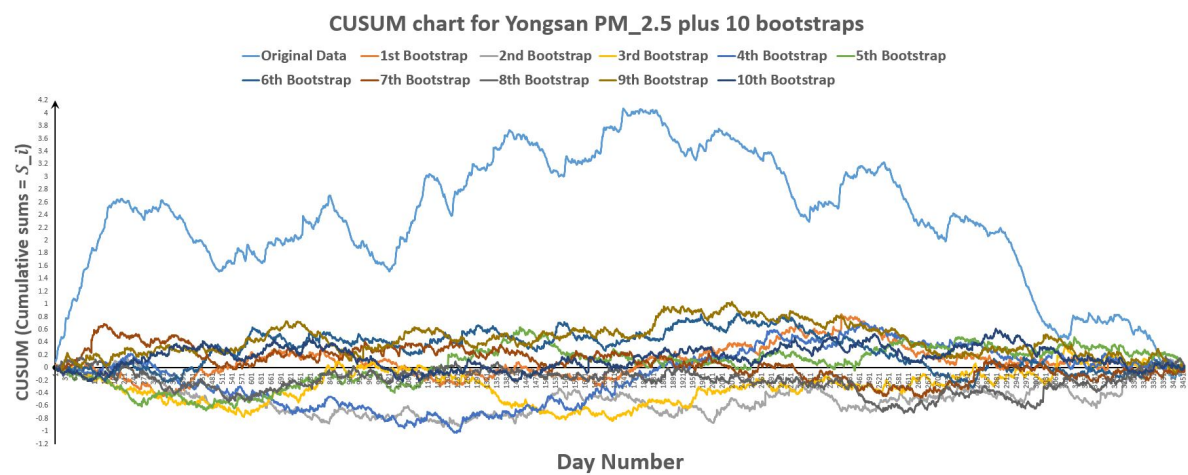
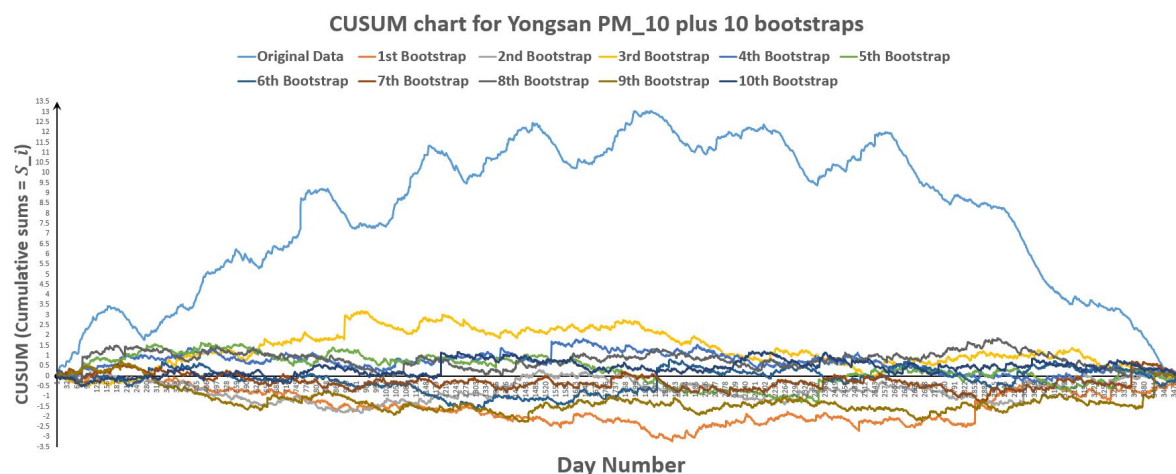**Figure 34.** CUSUM chart for Yongsan $PM_{2.5}$ plus 10 bootstraps.

Figure 35. CUSUM chart for Yongsan PM_{10} plus 10 bootstraps.

The value of k is uniform over $y_1....,y_n$.

5. Results

Summarized forms of particulate matter ($PM_{2.5}$ and PM_{10}) change point (k), the parameters before a change point ($mean = \mu$, $variance = \sigma^2$) and the parameters after a change point ($mean = \eta$, $variance = \phi^2$) during the study period 2004–2013 for four different sites (Guro, Nowon, Songpa and Yongsan) in Seoul, South Korea are given in Tables 7–10, respectively. The results were computed using the numerical example of the mathematical model given in Section 4. At airkorea official website, the annual Particulate Matter trend in Seoul is being exhibited in Figure 36, which shows a decreasing trend during (2004–2013). These particulate matters concentrations are given in μg/m^3 [39].
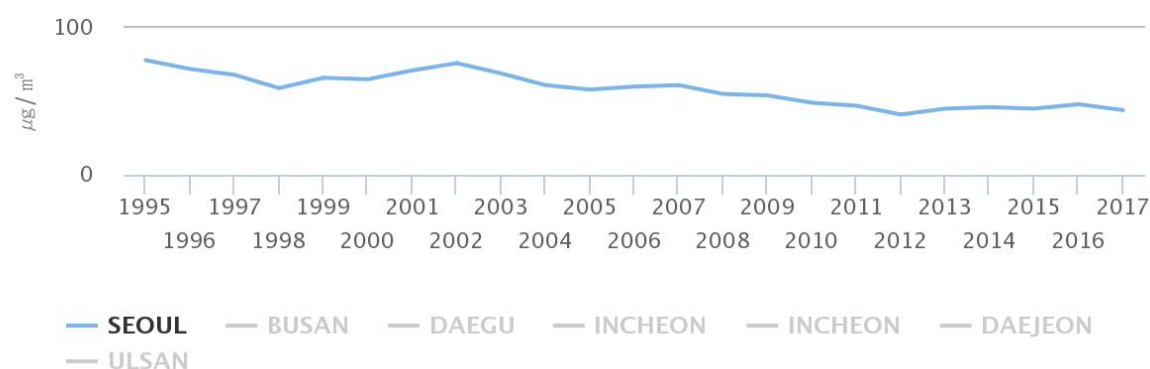

Figure 36. Annual Particulate Matter trend in Seoul (airkorea).

5.1. $PM_{2.5}$ Change Point (k) through Probabilistic Method

In Table 7, the results obtained through Probabilistic method have been described. Where, (k) is the predicted change point varies for different areas. The results indicate the reduction of $PM_{2.5}$ concentrations after change point (k). While, (μ) represents the mean concentrations before change point (k) and (η) be the mean concentrations after change point (k). The variance before change point (σ^2) and variance after change point (ϕ^2) have been determined through Inverse-Gamma distribution with conjugate hyper-parameters.

Table 7. $PM_{2.5}$ Change point (k) for Normal distribution, parameters before the change point ($mean = \mu, variance = \sigma^2$), and parameters after the change point ($mean = \eta, variance = \phi^2$).

Probabilistic Method for Change Point Detection (Normal Distribution)				
Parameters	**Guro**	**Nowon**	**Songpa**	**Yongsan**
Mean before change point $= \mu$ (mg/m^3)	0.02931	0.02950	0.02812	0.03066
Mean after change point $= \eta$ (mg/m^3)	0.02461	0.02354	0.02456	0.02485
Variance before change point $= \sigma^2$	4.14015×10^{-30}	1.13558×10^{-30}	9.79587×10^{-31}	1.01342×10^{-30}
Variance after change point $= \phi^2$	1.39483×10^{-30}	5.43837×10^{-31}	1.93175×10^{-30}	3.22178×10^{-30}
Change point $= k$	1228.44	1502.84	1745.82	1501.15
Convergence for $\mu = \sqrt{R_\mu}$	1.00032	1.00020	1.00086	1.00285
Overall estimate of variance for $\mu = Var(V)_\mu$	0.000004	0.000003	0.000002	0.000033
Convergence for $\eta = \sqrt{R_\eta}$	0.99986	1.00009	1.00012	0.99999
Overall estimate of variance for $\eta = Var(V)_\eta$	0.000002	0.000002	0.000002	0.000001
Convergence for $\sigma^2 = \sqrt{R_{\sigma^2}}$	1.000105	0.999798	0.999982	1.000099
Overall estimate of variance for $\sigma^2 = Var(V)_\sigma^2$	3.41×10^{-59}	1.31×10^{-60}	1.73×10^{-60}	3.81×10^{-60}
Convergence for $\phi^2 = \sqrt{R_{\phi^2}}$	1.000165	1.000371	0.999849	0.999884
Overall estimate of variance for $\phi^2 = Var(V)_\phi^2$	4.31×10^{-60}	5.96×10^{-61}	2.28×10^{-59}	2.33×10^{-59}
Convergence for $k = \sqrt{R_k}$	0.99985	0.99995	1.00011	1.00098
Overall estimate of variance for $k = Var(V)_k$	698111.33	521067.70	853426.78	1008921.72

5.2. $PM_{2.5}$ Last Point before Change (k) and First Point after Change ($k+1$) through CUSUM Approach

Table 8 represents the results obtained for $PM_{2.5}$ through CUSUM approach. Where, (k) is the last point before change and ($k+1$) be the first point after change point. So, the change point le is somewhere between (k) and ($k+1$). This method also shows the reduction of $PM_{2.5}$ concentrations after change point as (μ) represents the mean concentration before change point and (η) be the pollutant concentrations after change point. The variance before change point (σ^2) and variance after change point (ϕ^2) have been determined through formulae $\left(\sigma^2 = \frac{\sum(X_i - Mean)^2}{n} \right)$.

Table 8. $PM_{2.5}$ Last point before change (k) and first point after change ($k+1$) through the CUSUM approach (Normal distribution).

CUSUM Approach for Change Point Detection (Normal Distribution)										
Area Seoul, South Korea	($Mean = \mu$) before change (mg/m^3)	($Mean = \eta$) after change (mg/m^3)	(σ^2) Variance before change	(ϕ^2) Variance after change	K Last point before change	$k+1$ First point after change	$\|S_m\|$ Most extreme point	S_{max} Highest point in CUSUM	S_{min} Lowest point in CUSUM	S_{diff} Magnitude of change	Confidence level %
Guro	0.02894	0.02301	0.00029642	0.00017677	1570	1571	3.428	3.428	−0.143	3.572	100 %
Nowon	0.03084	0.02242	0.0027245	0.00017550	1474	1475	6.370	6.370	−0.486	6.856	100 %
Songpa	0.02928	0.02420	0.00028241	0.00019390	1455	1456	4.218	4.218	−0.223	4.441	100 %
Yongsan	0.02884	0.02412	0.00032178	0.00019152	1738	1739	4.076	4.076	−0.167	4.243	100 %

5.3. PM_{10} Change Point (k) through Probabilistic Method

Table 9 explains the results obtained for PM_{10} through Probabilistic method. Hence, the expected change point is (k) that differs for different areas. These results show the reduction of PM_{10} concentrations after change point (k). While, (μ) be the PM_{10} concentrations before change point (k) and (η) represents PM_{10} concentrations after change point (k). The variance before change point (σ^2) and variance after change point (ϕ^2) have been determined through Inverse-Gamma distribution with conjugate hyper-parameters.

Table 9. PM_{10} Change point (k) for Normal distribution, parameters before the change point ($mean = \mu, variance = \sigma^2$), and parameters after the change point ($mean = \eta, variance = \phi^2$).

Probabilistic Method for Change Point Detection (Normal Distribution)				
Parameters	**Guro**	**Nowon**	**Songpa**	**Yongsan**
Mean before change point $= \mu$ (mg/m^3)	0.05789	0.05706	0.05758	0.05970
Mean after change point $= \eta$ (mg/m^3)	0.04696	0.04372	0.04361	0.04550
Variance before change point $= \sigma^2$	5.05815×10^{-29}	1.10597×10^{-29}	7.07826×10^{-29}	4.50556×10^{-29}
Variance after change point $= \phi^2$	1.34766×10^{-29}	4.24382×10^{-30}	5.12048×10^{-30}	5.79728×10^{-30}
Change point $= k$	2025.19	1860.49	2025.53	2019.46
Convergence for $\mu = \sqrt{R_\mu}$	1.00046	1.00417	1.00059	1.00053
Overall estimate of variance for $\mu = Var(V)_\mu$	0.00001	0.00003	0.00001	0.00001
Convergence for $\eta = \sqrt{R_\eta}$	0.99980	1.00165	1.00057	1.00038
Overall estimate of variance for $\eta = Var(V)_\eta$	0.00004	0.00001	0.00001	0.00001
Convergence for $\sigma^2 = \sqrt{R_{\sigma^2}}$	1.038910	1.000918	1.000627	1.000106
Overall estimate of variance for $\sigma^2 = Var(V)^2_\sigma$	2.78×10^{-57}	7.15×10^{-59}	2.41×10^{-57}	1.08×10^{-57}
Convergence for $\phi^2 = \sqrt{R_{\phi^2}}$	1.221403	1.001116	1.000223	1.000568
Overall estimate of variance for $\phi^2 = Var(V)^2_\phi$	6.40×10^{-58}	7.18×10^{-59}	6.97×10^{-58}	7.71×10^{-59}
Convergence for $k = \sqrt{R_k}$	1.00095	1.00245	1.00078	1.00059
Overall estimate of variance for $k = Var(V)_k$	625170.52	599959.50	404584.27	522327.71

5.4. PM_{10} Last Point before Change (k) and First Point after Change ($k + 1$) through CUSUM Approach

The results obtained for PM_{10} through CUSUM approach have been described in Table 10. Where, the last point before change is (k) and the first point after change point is ($k + 1$). Therefore, the change point leis anywhere between (k) and ($k + 1$). This method also depicts the reduction of PM_{10} concentrations after change point. (μ) represents the PM_{10} concentrations before change point and (η) be the PM_{10} concentrations after change point. The variance before change point (σ^2) and variance after change point (ϕ^2) have been determined through formulae $\left(\sigma^2 = \frac{\sum(X_i - Mean)^2}{n} \right)$.

Table 10. PM_{10} Last point before change (k) and first point after change ($k + 1$) through the CUSUM approach (Normal distribution).

CUSUM Approach for Change Point Detection (Normal Distribution)										
Area Seoul, South Korea	($Mean = \mu$) before change (mg/m^3)	($Mean = \eta$) after change (mg/m^3)	(σ^2) Variance before change	(ϕ^2) Variance after change	K Last point before change	$k+1$ First point after change	$\lvert S_m \rvert$ Most extreme point	S_{max} Highest point in CUSUM	S_{min} Lowest point in CUSUM	S_{diff} Magnitude of change	Confidence level %
Guro	0.05914	0.04721	0.00166718	0.00088748	1836	1837	10.538	10.538	−0.118	10.656	100 %
Nowon	0.05743	0.04153	0.00139338	0.00061631	1952	1953	13.949	13.949	−0.194	14.143	100 %
Songpa	0.06106	0.04484	0.00217706	0.00077376	1515	1516	13.838	13.838	−0.126	13.963	100 %
Yongsan	0.06105	0.04617	0.00216982	0.00082027	1795	1796	13.046	13.046	−0.108	13.154	100 %

6. Discussion

6.1. Guro (Seoul, South Korea)

Guro is located in the southwestern part of Seoul, having an essential location as a transport link for railroads and land routes. The largest digital industrial complex in Korea is located in Guro. Thus, the policies of the Ministry of Environment in South Korea have decreased the particulate matters ($PM_{2.5}$ and PM_{10}) concentrations and occurrences of polluted days in Guro.

6.1.1. Probabilistic Method

Table 7 presents the reduction of 16.02% in particulate matter $PM_{2.5}$ concentration in Guro, from ($mean$ (μ) $= 0.02931$ mg/m^3) to ($mean$ (η) $= 0.02461$ mg/m^3) and a change point ($k = 1228$). This point occurred on 13th July, 2010 during 7 years data from March 2007–December 2013. Therefore,

the expected pollutant concentration before 13th July, 2010 was 0.02931, which changed to 0.02461 after 13th July, 2010. On the other hand, variance ($\sigma^2 = 4.14015 \times 10^{-30}$) changed to ($\phi^2 = 1.39483 \times 10^{-30}$). Similarly, Table 9 indicates 18.88% reduction in PM_{10} concentration from (*mean* (μ) = 0.05789 mg/m^3) to (*mean* (η) = 0.04696 mg/m^3) with a change point $k = 2025.19$. This change point occurred 19th October, 2009 in the period of 2004–2013 and involved a change in variance from ($\sigma^2 = 5.05815 \times 10^{-29}$) to ($\phi^2 = 1.34766 \times 10^{-29}$).

6.1.2. CUSUM Approach

CUSUM Approach also indicates a reduction in $PM_{2.5}$ and PM_{10} concentrations from (μ) to (η) after change. As for Guro, Tables 8 and 10 depict the change of $PM_{2.5}$ and PM_{10} concentrations through CUSUM approach respectively, change point for $PM_{2.5}$ concentrations lies in-between point 1570 (k) and 1571 ($k + 1$) and it occurred between point 1836 (k) and 1837 ($k + 1$) for PM_{10} concentrations.

6.2. Nowon (Seoul, South Korea)

Nowon is positioned at the northeastern part of Seoul and has the highest population density in Seoul, with 619,509 persons living in 35.44 km^2. The area is surrounded by mountains and forests on the northeast. The policies of the Ministry of Environment in Nowon have improved the particulate matter ($PM_{2.5}$ and PM_{10}) concentrations from (μ, σ^2) to (η, ϕ^2). Improvement in the reduction of pollutant concentrations varies which is more than Guro.

6.2.1. Probabilistic Method

For Nowon, Tables 7 and 9 depict the change points ($k = 1502.84$) and ($k = 1860.49$) for $PM_{2.5}$ and PM_{10} concentrations respectively. The change point ($k = 1502.84$) for $PM_{2.5}$ occurred 10th August 2009 for the period March 2005–December 2013. The parameters ($\mu = 0.02950$ mg/m^3, $\sigma^2 = 1.13558 \times 10^{-30}$) and ($\eta = 0.02354$ mg/m^3, $\phi^2 = 5.483837 \times 10^{-31}$) indicate a 20.21% reduction in $PM_{2.5}$ expected value after the change point, while there was minor change in the variance parameter. similarly, the change point ($k = 1860.49$) for PM_{10} occurred on 18th April, 2009 for the period 2004–2013 and a 23.38% reduction in pollutant expectation was observed from ($\mu = 0.05706$ mg/m^3) to ($\eta = 0.04372$ mg/m^3) with a change in variance from ($\sigma^2 = 1.10597 \times 10^{-29}$) to ($\phi^2 = 4.24382 \times 10^{-30}$).

6.2.2. CUSUM Approach

Moreover, CUSUM Approach also validates the reduction of PM concentrations and polluted days. In case of Nowon, Tables 8 and 10 depict the change of $PM_{2.5}$ and PM_{10} concentrations through CUSUM approach respectively, change point for $PM_{2.5}$ concentrations lies in-between point 1474 (k) and 1475 ($k + 1$) and it occurred between point 1952 (k) and 1953 ($k + 1$) for PM_{10} concentrations.

6.3. Songpa (Seoul, South Korea)

Songpa is situated at the southeastern part of Seoul, with the largest population of 647000 residents. As per Ministry of Environment policies in Songpa, there was a significant reduction for pollutant concentrations (μ, σ^2) to (η, ϕ^2).

6.3.1. Probabilistic Method

For Songpa, Table 7 shows that change point (k) for $PM_{2.5}$ pollutant concentration was 1745.82, which occurred on 23rd March 2009. The reduction in $PM_{2.5}$ expectation was 12.65% from ($\mu = 0.02812$ mg/m^3) to ($\eta = 0.02456$ mg/m^3) while variance ($\sigma^2 = 9.79587 \times 10^{-31}$) changed to ($\phi^2 = 1.93175 \times 10^{-30}$). Correspondingly, in Table 9, a 24.26% improvement in the mean of PM_{10} concentration was from ($\mu = 0.05758$ mg/m^3) to ($\eta = 0.04361$ mg/m^3) after the change point ($k = 2025.53$), which occurred on 1st January 2010, while variance changed from ($\sigma^2 = 7.0782 \times 10^{-29}$) to ($\phi^2 = 5.12048 \times 10^{-30}$).

6.3.2. *CUSUM Approach*

As per CUSUM Approach, there is a decrease in PM concentrations. Tables 8 and 10 depict the change of $PM_{2.5}$ and PM_{10} concentrations through CUSUM approach respectively, change point for $PM_{2.5}$ concentrations lies in-between point 1455 (k) and 1456 ($k + 1$) and it occurred between point 1515 (k) and 1516 ($k + 1$) for PM_{10} concentrations.

6.4. *Yongsan (Seoul, South Korea)*

Yongsan is the center of Seoul, in which almost 250,000 people reside. Prominent locations in Yongsan includes Yongsan station, an electronics market and Itaewon commercial area with heavy traffic and transportation. Consequently, the policies of the Ministry of Environment in Yongsan have affected the particulate matter ($PM_{2.5}$ and PM_{10}) concentrations, producing a remarkable decrease from (μ, σ^2) to (η, ϕ^2).

6.4.1. *Probabilistic Method*

Similarly, Yongsan, Tables 7 and 9 show that particulate matter ($PM_{2.5}$ and PM_{10}) concentrations were changed from (μ, σ^2) to (η, ϕ^2) after change point (k). The change point ($k = 1501.15$) for $PM_{2.5}$ concentration occurred with a reduction of 18.97% from ($\mu = 0.03066$ mg/m^3) to ($\eta = 0.02485$ mg/m^3) and a change in variance from ($\sigma^2 = 1.01342 \times 10^{-30}$) to ($\phi^2 = 3.22178 \times 10^{-30}$). This change point k occurred on 19th August, 2008. Correspondingly, the change point ($k = 2019.46$) for PM_{10} concentration produced a 23.79% reduction in the expected value from ($\mu = 0.05970$ mg/m^3) to ($\eta = 0.04550$ mg/m^3). The change observed for variance was ($\sigma^2 = 4.50556 \times 10^{-29}$) to ($\phi^2 = 5.79728 \times 10^{-30}$) and change point k occurred on 24th November, 2009.

6.4.2. *CUSUM Approach*

The CUSUM Approach is directly applied on the raw data, which should be better for deterministic data structures. It also shows a reduction in pollutant concentrations. In case of Yongsan, Tables 8 and 10 depict the change of $PM_{2.5}$ and PM_{10} concentrations through CUSUM approach respectively, change point for $PM_{2.5}$ concentrations lies in-between point 1738 (k) and 1739 ($k + 1$) and it occurred between point 1795 (k) and 1796 ($k + 1$) for PM_{10} concentrations.

6.5. *Managerial Insights*

1. This model presents a suitable technique for change point detection of diversely distributed data structures for all kind of stochastic processes.
2. By detecting change points in different areas, such as climate change detection, human activity analysis and medical condition monitoring and also analyzing the parameters before and after change points, the results of legislation efforts can be understood and it can be determined whether these change points are favorable.
3. A comparison of parameters before and after a change point evaluates the performance from previous status to current status, which can also be helpful for future prediction with the current strategies.
4. This study of change point detection also defines the current levels of an area under study, which is helpful for designing new policies for further improvements.
5. This research provides guidance for defining new goals if previously defined goals have been achieved and indicates if the standards need to be revised to overcome upcoming challenges.

7. Conclusions

The key motivation of this research work was to explicate an appropriate change point detection model for diversely distributed data structures. This probabilistic method is being verified by the CUSUM approach and the results of the CUSUM approach are compared with the proposed method.

This methodology is based on probability distributions and better to apply for random data structures and time series. But the CUSUM approach is directly applicable to the raw data, which is good for deterministic data structures. The model is applicable to various stochastic processes because different data structures follow different probability distributions. The parameter expectations before and after a change point were also estimated to measure the effectiveness and performance of policies applied. To verify the model, four major locations (Guro, Nowon, Songpa and Yongsan) in Seoul, South Korea were chosen as study areas considering their different characteristics, such as climate zone, environment, population and population density. The results were calculated and conclusions were drawn with the application of model on real-time data sets in all cases. The parameters before and after the change point of particulate matter concentrations indicated a reduction in pollutant concentrations over a 10-year period. At airkorea official website, the annual Particulate Matter trend in Seoul is being exhibited in Figure 36, which also shows a similar kind of decreasing trend during (2004–2013). The overall outcomes of this study indicate the effectiveness of policies applied to reduce the pollutant concentrations over time. Thus, further reduction in PM concentrations is required to achieve the set standards. This study can be further extended by locating change segments through multiple change points.

References

1. Aminikhanghahi, S.; Cook, D.J. A survey of methods for time series change point detection. *Knowl. Inf. Syst.* **2017**, *51*, 339–367. [CrossRef] [PubMed]
2. Reeves, J.; Chen, J.; Wang, X.L.; Lund, R.; Lu, Q.Q. A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 900–915. [CrossRef]
3. Taylor, W. Change-point analysis: A powerful tool for detecting changes. *Retriev. July* **2000**, *5*, 2012.
4. Cabrieto, J.; Tuerlinckx, F.; Kuppens, P.; Wilhelm, F.H.; Liedlgruber, M.; Ceulemans, E. Capturing correlation changes by applying kernel change point detection on the running correlations. *Inf. Sci.* **2018**, *447*, 117–139. [CrossRef]
5. Keshavarz, H.; Scott, C.; Nguyen, X. Optimal change point detection in Gaussian processes. *J. Stat. Plan. Inference* **2018**, *193*, 151–178. [CrossRef]
6. Kucharczyk, D.; Wyłomańska, A.; Sikora, G. Variance change point detection for fractional Brownian motion based on the likelihood ratio test. *Phys. A Stat. Mech. Its Appl.* **2018**, *490*, 439–450. [CrossRef]
7. Sarkar, B. A production-inventory model with probabilistic deterioration in two-echelon supply chain management. *Appl. Math. Model.* **2013**, *37*, 3138–3151. [CrossRef]
8. Lu, G.; Zhou, Y.; Lu, C.; Li, X. A novel framework of change-point detection for machine monitoring. *Mech. Syst. Signal Process.* **2017**, *83*, 533–548. [CrossRef]
9. Khan, M.R.; Sarkar, B. Change Point Detection for Airborne Particulate Matter (PM2. 5, PM10) by Using the Bayesian Approach. *Mathematics* **2019**, *7*, 474. [CrossRef]
10. Liu, S.; Yamada, M.; Collier, N.; Sugiyama, M. Change-point detection in time-series data by relative density-ratio estimation. *Neural Netw.* **2013**, *43*, 72–83. [CrossRef]
11. Hilgert, N.; Verdier, G.; Vila, J.P. Change detection for uncertain autoregressive dynamic models through nonparametric estimation. *Stat. Methodol.* **2016**, *33*, 96–113. [CrossRef]
12. Sarkar, B.; Sana, S.S.; Chaudhuri, K. An economic production quantity model with stochastic demand in an imperfect production system. *Int. J. Serv. Oper. Manag.* **2011**, *9*, 259–283. [CrossRef]
13. Górecki, T.; Horváth, L.; Kokoszka, P. Change point detection in heteroscedastic time series. *Econom. Stat.* **2017**, *7*, 63–88.

[CrossRef]

14. Bucchia, B.; Wendler, M. Change-point detection and bootstrap for Hilbert space valued random fields. *J. Multivar. Anal.* **2017**, *155*, 344–368. [CrossRef]

15. Taleizadeh, A.A.; Samimi, H.; Sarkar, B.; Mohammadi, B. Stochastic machine breakdown and discrete delivery in an imperfect inventory-production system. *J. Ind. Manag. Optim.* **2017**, *13*, 1511–1535. [CrossRef]

16. Zhou, M.; Wang, H.J.; Tang, Y. Sequential change point detection in linear quantile regression models. *Stat. Probab. Lett.* **2015**, *100*, 98–103. [CrossRef]

17. Kim, S.J.; Sarkar, B. Supply chain model with stochastic lead time, trade-credit financing, and transportation discounts. *Math. Probl. Eng.* **2017**, *2017*. [CrossRef]

18. Lu, K.P.; Chang, S.T. Detecting change-points for shifts in mean and variance using fuzzy classification maximum likelihood change-point algorithms. *J. Comput. Appl. Math.* **2016**, *308*, 447–463. [CrossRef]

19. Ruggieri, E.; Antonellis, M. An exact approach to Bayesian sequential change point detection. *Comput. Stat. Data Anal.* **2016**, *97*, 71–86. [CrossRef]

20. Ahmed, E.; Kim, K.H.; Shon, Z.H.; Song, S.K. Long-term trend of airborne particulate matter in Seoul, Korea from 2004 to 2013. *Atmos. Environ.* **2015**, *101*, 125–133. [CrossRef]

21. Kim, K.H.; Pandey, S.K.; Nguyen, H.T.; Chung, S.Y.; Cho, S.J.; Kim, M.Y.; Oh, J.M.; Sunwoo, Y. Long-term behavior of particulate matters at urban roadside and background locations in Seoul, Korea. *Transp. Res. Part D Transp. Environ.* **2010**, *15*, 168–174. [CrossRef]

22. Gupta, A.; Baker, J.W. Estimating spatially varying event rates with a change point using Bayesian statistics: Application to induced seismicity. *Struct. Saf.* **2017**, *65*, 1–11. [CrossRef]

23. Bardwell, L.; Fearnhead, P. Bayesian detection of abnormal segments in multiple time series. *Bayesian Anal.* **2017**, *12*, 193–218. [CrossRef]

24. Sarkar, M.; Sarkar, B. An economic manufacturing quantity model with probabilistic deterioration in a production system. *Econ. Model.* **2013**, *31*, 245–252. [CrossRef]

25. Keshavarz, M.; Huang, B. Bayesian and Expectation Maximization methods for multivariate change point detection. *Comput. Chem. Eng.* **2014**, *60*, 339–353. [CrossRef]

26. Moon, I.; Shin, E.; Sarkar, B. Min–max distribution free continuous-review model with a service level constraint and variable lead time. *Appl. Math. Comput.* **2014**, *229*, 310–315. [CrossRef]

27. Kurt, B.; Yıldız, Ç.; Ceritli, T.Y.; Sankur, B.; Cemgil, A.T. A Bayesian change point model for detecting SIP-based DDoS attacks. *Digit. Signal Process.* **2018**, *77*, 48–62. [CrossRef]

28. Wu, C.; Du, B.; Cui, X.; Zhang, L. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. *Remote Sens. Environ.* **2017**, *199*, 241–255. [CrossRef]

29. Sarkar, B.; Cárdenas-Barrón, L.E.; Sarkar, M.; Singgih, M.L. An economic production quantity model with random defective rate, rework process and backorders for a single stage production system. *J. Manuf. Syst.* **2014**, *33*, 423–435. [CrossRef]

30. Mariño, I.P.; Blyuss, O.; Ryan, A.; Gentry-Maharaj, A.; Timms, J.F.; Dawnay, A.; Kalsi, J.; Jacobs, I.; Menon, U.; Zaikin, A. Change-point of multiple biomarkers in women with ovarian cancer. *Biomed. Signal Process. Control* **2017**, *33*, 169–177. [CrossRef]

31. Sarkar, B.; Singh, L.K.; Sarkar, D. Approximation of digital curves with line segments and circular arcs using genetic algorithms. *Pattern Recognit. Lett.* **2003**, *24*, 2585–2595. [CrossRef]

32. Jeon, J.J.; Sung, J.H.; Chung, E.S. Abrupt change point detection of annual maximum precipitation using fused lasso. *J. Hydrol.* **2016**, *538*, 831–841. [CrossRef]

33. Kang, C.W.; Ullah, M.; Sarkar, B.; Hussain, I.; Akhtar, R. Impact of random defective rate on lot size focusing work-in-process inventory in manufacturing system. *Int. J. Prod. Res.* **2017**, *55*, 1748–1766. [CrossRef]

34. Chen, S.; Li, Y.; Kim, J.; Kim, S.W. Bayesian change point analysis for extreme daily precipitation. *Int. J. Climatol.* **2017**, *37*, 3123–3137. [CrossRef]

35. Gelman, A. Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice*; CRC Press: Boca Raton, FL, USA, 1996; pp. 131–143.

36. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, *7*, 457–472. [CrossRef]

37. Gelman, A.; Rubin, D.B. A single series from the Gibbs sampler provides a false sense of security. *Bayesian Stat.* **1992**, *4*, 625–631.

38.   Cowles, M.K.; Carlin, B.P. Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Am. Stat. Assoc.* **1996**, *91*, 883–904. [CrossRef]

39.   Air Korea, Annual Air Quality Trends.   Available online: https://www.airkorea.or.kr/eng/annualAirQualityTrends?pMENU_NO=161 (accessed on 26 July 2019).