mathematics

MDPI

Article

Change Point Detection for Airborne Particulate Matter ($PM_{2.5}$, PM_{10}) by Using the Bayesian Approach

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

[1] Department of Industrial Engineering, Hanyang University, 222 Wangsimni-Ro, Seoul 133-791, Korea; mrizwankhan162@gmail.com
[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

Received: 28 February 2019; Accepted: 8 May 2019; Published: 24 May 2019

Abstract: Airborne particulate matter (PM) is a key air pollutant that affects human health adversely. Exposure to high concentrations of such particles may cause premature death, heart disease, respiratory problems, or reduced lung function. Previous work on particulate matter ($PM_{2.5}$ and PM_{10}) was limited to specific areas. Therefore, more studies are required to investigate airborne particulate matter patterns due to their complex and varying properties, and their associated (PM_{10} and $PM_{2.5}$) concentrations and compositions to assess the numerical productivity of pollution control programs for air quality. Consequently, to control particulate matter pollution and to make effective plans for counter measurement, it is important to measure the efficiency and efficacy of policies applied by the Ministry of Environment. The primary purpose of this research is to construct a simulation model for the identification of a change point in particulate matter ($PM_{2.5}$ and PM_{10}) concentration, and if it occurs in different areas of the world. The methodology is based on the Bayesian approach for the analysis of different data structures and a likelihood ratio test is used to a detect change point at unknown time (k). Real time data of particulate matter concentrations at different locations has been used for numerical verification. The model parameters before change point (θ) and parameters after change point (λ) have been critically analyzed so that the proficiency and success of environmental policies for particulate matter ($PM_{2.5}$ and PM_{10}) concentrations can be evaluated. The main reason for using different areas is their considerably different features, i.e., environment, population densities, and transportation vehicle densities. Consequently, this study also provides insights about how well this suggested model could perform in different areas.

Keywords: airborne particulate matter; Bayesian approach; change point detection; likelihood ratio test; time series analysis; air quality

1. Introduction

Airborne particulate matter is one of the most dangerous air pollutants and harmful to human health. For the last two decades, information about the negative impacts of PM_{10} (particles less than 10 μm in diameter) and $PM_{2.5}$ (particles less than 2.5 micrometers in diameter) has increased enormously. Exposure to high concentrations of such particles may cause premature death, heart disease, respiratory problems, or reduced lung function through different mechanisms, which include pulmonary and systemic inflammation, accelerated atherosclerosis, and altered cardiac autonomic function (Heroux et al. [1] and Pope et al. [2]). Therefore, to control particulate matter (PM) pollution, and to make effective plans for counter measurements, it is important to measure the efficiency and effectiveness of policies applied by the Ministry of Environment. Every region has developed different kinds of extensive bodies of legislation, which establish air quality standards for key air pollutants to improve the air quality and to satisfy these standards. The European Environment Agency, the United States Environmental Protection Agency, and the Ministry of Environment in South Korea have each set their own air quality standards for all air pollutants. Thus, It is essential to follow established air quality monitoring systems to measure the PM concentrations on an hourly as well as daily basis, because some areas deviate from the established PM standards. This may cause adverse environmental effects and serious health problems.

Until now, a number of statistical methods have been established to model the hazards of PM from air quality standards. A Bayesian multiple change point model was proposed to measure the quantitative efficiency of pollution control programs for air quality, which estimate the hazards of different air pollutants. In the model, it was assumed as a nonhomogeneous Poisson process with multiple change points. The change points were identified, and a rate function was estimated by using a reversible jump MCMC algorithm (Gyarmati-Szabo et al. [3]). In another study, the changes in health effects due to simultaneous exposure to physical and chemical properties of airborne particulate matter were gauged through Bayesian approach and inferences were drawn via the Markov Chain Monte Carlo method (Pirani et al. [4]). A Bayesian approach was introduced to estimate the distributed lag functions in time series models, which can be used to determine the short-term health effects of particulate air pollution on mortality (Welty et al. [5]). Hybrid models were proposed to forecast the PM concentrations for four major cities of China; Beijing, Shanghai, Guangzhou, and Lanzhou (Qin et al. [6]).

A change point detection method for detecting changes in the mean of the one dimensional Gaussian process was proposed on the basis of a generalized likelihood ratio test (GLRT). The important characteristic of this method is that it includes data dependence and covariance of the Gaussian process. However, in case of unidentified covariance, the plug-in GLRT method was suggested which remains asymptotically near optimal (Keshavarz et al. [7]). A new method for acute change point detection was proposed for fractional Brownian motion with a time dependent diffusion coefficient. The likelihood ratio method has been used for change point detection in Brownian motion. A statistical test was also suggested to identify the significance of a calculated critical point (Kucharczyk et al. [8]). The change point detection technique in machine monitoring was suggested, which was based on two stages. In the first stage, irregularities are measured in time series data through the automatic regression (AR) model, and then the martingale statistical test is applied to detect the change point in unsupervised time series data (Lu et al. [9]). 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 (Taleizadeh et al. [10]).

A statistical change point algorithm based on nonparametric deviation estimation between time series samples from two retrospective segments was proposed in which the direct density ratio estimation method was applied for deviation measurement through relative Pearson divergence (Liu et al. [11]). A novel statistical methodology for online change point detection was suggested in which data for an uncertain system was composed through an autoregressive model. On the basis of nonparametric estimation of unidentified elements, an innovative CUSUM-like scheme was recommended for change detection. This estimation method could also be updated online (Hilgert et al. [12]). A new methodology, the Karhunen-Loeve expansions of the limit Gaussian processes, was suggested for change point test in the level of a series. Firstly, change point detection in the mean was explained, which later extended to linear and nonlinear regression (Górecki et al. [13]). A Cramer-von Mises type test was presented to test the sudden changes in random fields which was dependent on Hilbert space theory (Bucchia and Wendler [14]). The continuous-review inventory model was developed for Controllable lead time for comparing two models; one with normally distributed lead time demand and the second assumes that there is no specific distribution for lead time demand (Shin et al. [15]).

A new technique was developed to identify the structural changes in linear quantile regression models. When a structural change in the relationship between covariates and response at a specific point exists, it may not be at the centre of response distribution, but at the tail. The traditional mean regression method might not be applicable for change point detection of such structural changes at tails. Subsequently, the proposed technique could be appropriate for it (Zhou et al. [16]). For detection of simultaneous changes in mean and variance, a new methodology called the fuzzy classification maximum likelihood change point (FCML-CP) algorithm was suggested. Multiple change points in the mean and variance of a process can be estimated by this method. This technique is much better than the normal statistical mixture likelihood method because it saves a lot of time (Lu and Chang [17]). A model for Partial Trade-Credit Policy of Retailer was developed in which deterioration of products was assumed as exponentially distributed (Sarkar and Saren [18]).

The Bayesian change point algorithm for sequential data series was introduced which has some uncertain limitations regarding location and number of change points. This algorithm was precisely based on posterior distribution to deduce if a change point has occurred or not. It can also update itself linearly as new data points are observed. Posterior distribution monitoring is the finest 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 is also known for a low rate of false detection (Ruggieri and Antonellis [19]). Due to the probabilistic concept of Bayesian change point detection (BCPD), this methodology can overcome threats in identifying the location and number of change points.

The performance of two different methods for change point detection of multivariate data with both single and multiple changes was compared. The results illustrated adequate performance for both Expectation Maximization (EM) and Bayesian methods. However, EM exhibits better performance in case of minor changes and unsuitable priors while the Bayesian method has less computational work to do (Keshavarz and Huang [20]). The Bayesian multiple change point model was suggested for the identification of Distributed Denial of Service (DDoS) flooding attacks in VoIP systems in which Session Initiation Protocol (SIP) is used as signalling mechanism (Kurt et al. [21]). One of the well-known change detection techniques is post classification with multi temporal remote sensing images. An innovative post classification technique with iterative slow feature analysis (ISFA) and Bayesian soft fusion was suggested to acquire accurate and reliable change detection maps. Three steps were suggested in this technique, first was to get the class probability of images through independent classification. After that, a continuous change probability map of multi temporal images was obtained by ISFA algorithm. Lastly, posterior probabilities for the class combinations of coupled pixels were determined through the Bayesian approach to assimilate the class probability with the change probability, which is called Bayesian soft fusion. This technique could be widely applicable in land cover monitoring and change detection at a large scale (Wu et al. [22]).

The Bayesian change point technique was designed to analyse biomarkers time series data in women for the diagnosis of ovarian cancer. The identification of such kind of change points could be used to diagnose the disease earlier (Mariño et al. [23]). The Generalized Extreme Value (GEV) fused lasso penalty function was applied to identify the change point of annual maximum precipitation (AMP) in South Korea. Numerical analysis and applied data analysis were conducted in order to compare performance from the GEV fused lasso and Bayesian change point analysis, which shows that when water resource structures are hydrologically designed the GEV fused lasso method should be used to identify the change points (Jeon et al. [24]). The Bayesian method was recommended to identify the change point occurrence in extreme precipitation data, and the model follows a generalized Pareto distribution. This Bayesian change point detection was inspected for four different situations, one with no change model, second with a shape change model, third with a scale change model, and fourth with both a scale and shape change model. It was determined that unexpected and sustained change points need to be considered in extreme precipitation while making hydraulic design (Chen et al. [25]).

Bayesian change point methodology was presented to identify changes in the temporal event rate for a non-homogeneous Poisson process. This methodology was used to determine if a change in the event rate has occurred or not, the time for change, and the event rate before or after the change. The methodology has been explained through an example of earthquake occurrence in Oklahoma. This spatiotemporal change point methodology can also be used for identifying changes in climate patterns and assessing the spread of diseases. It permits participants to make real time decisions about the influence of changes in event rates (Gupta and Baker [26]). A new Bayesian methodology was recommended to analyze multiple time series with the objective of identifying abnormal regions. A general model was developed and it was shown that Bayesian inference allows independent sampling from the posterior distribution. Copy number variations (CNVs) are identified by using data from multiple individuals. The Bayesian method was evaluated on both real and simulated CNV data to provide evidence that this method is more precise as compare to other suggested methods for analyzing such data (Bardwell and Fearnhead [27]).

All the above mentioned methods are either too complex and complicated for application on random hazards of PM or not applicable to randomness of PM hazards. Therefore, still more studies are required to investigate the PM hazards, due to its complex and varying properties and associated (PM_{10} and $PM_{2.5}$) concentrations and compositions, to investigate the numerical productivity of pollution control programs for air quality. The primary purpose of this research is to develop models for change point detection of particulate matter ($PM_{2.5}$ and PM_{10}) concentrations if it occurs in different areas. The pollutant concentrations before and after a change point has to be critically analyzed so that the proficiency and success of environmental policies for particulate matter ($PM_{2.5}$ and PM_{10}) concentrations can be evaluated. The Bayesian approach is used to analyze random hazards of PM concentrations with a change point at an unknown time (k).

To demonstrate the proposed approach, real time data of random hazards of PM concentrations at different sites has been used. The PM concentrations change point (k), parameters before change point (θ), and parameters after change point (λ) have been comprehensively analyzed by using the Bayesian technique. Thus, simulation models have been constructed for different data structures. The main reason for using different areas is their considerably different features i.e., environment, population densities, and transportation vehicle densities. Consequently, this study also provides insight about how well this suggested model could perform in different areas. The paper is structured as follows: Section 2 refers to problem definitions, explaining assumptions along with notation, and Section 3 shows the formulation of mathematical models. Sections 4 and 5 depict numerical examples and results, respectively, to validate the practical applications of the proposed models. Section 6 discusses the depicted results of previous section, it also explains the managerial insights of results. Finally, Section 7 presents conclusions of this study. Table 1 depicts the comparative study of different authors who have contributed in the direction of research, while the last row of the table portrays the contribution of this research paper. On the other hand, Tables 2 and 3 compares the difference in previous workings and this work.

Table 1. Author's contribution in the direction of research.

Authors	Airborne Particulate Matters ($PM_{2.5}$, PM_{10})	Change-Point Detection	BCPD (Bayesian Change Point Detection)	Time Series Analysis
Lim et al. (2012) [28]	Ambient particulate matter (PM)	-	-	-
Rao et al. (2018) [29]	Particulate matter air pollution	-	-	-
Wei and Meng (2018) [30]	Airborne fine particulate matter ($PM_{2.5}$)	-	-	-
Wang et al. (2017) [31]	Inhaled particulate matter	-	-	-
Heroux et al. (2015) [1]	Ambient air pollutants	-	-	-
Wellenius et al. (2012) [32]	Ambient Air Pollution	-	-	-
Pirani et al. (2015) [4]	Airborne particles	-	-	Bayesian inference within Time series
Li et al. (2013) [33]	Airborne particulate matter	-	-	Time series analysis
Kim et al. (2018a) [34]	Particulate matter	-	-	-
Kim et al. (2017) [35]	Air pollution	-	-	Time series model
Qin et al. (2017) [36]	Air pollution	-	-	Time series analysis
Lee et al. (2015) [37]	Fine and coarse particles	-	-	Time series analysis
Cabrieto et al. (2018) [38]	-	kernel change point detection, correlation changes	-	-
Keshavarz et al. (2018) [7]	-	Generalized likelihood ratio test, One-dimensional Gaussian process	- -	- -
Kucharczyk et al. (2018) [8]	-	likelihood ratio test, fractional Brownian motion	- -	- -
Lu et al. (2017) [9]	-	Change-point detection, machine monitoring, Anomaly measure (AR model), Martingale test	-	- Time series data
Hilgert et al. (2016) [12]	- - -	On-line change detection, autoregressive dynamic models, CUSUM-like scheme	- - -	- - -
Górecki et al. (2017) [13]	- -	Change point detection, heteroscedastic, Karhunen-Loeve expansions	- -	Time series data -
Bucchia and Wendler (2017) [14]	- - -	Change point detection, bootstrap, Hilbert space valued random fields, (Cramer–von Mises type test)	- -	Time series data - -
Zhou et al. (2015) [16]	- -	Sequential change point detection, linear quantile regression models	- -	- -
Lu and Chang (2016) [17]	- -	Detecting change points, mean/variance shifts, FCML-CP algorithm	- -	- -

Table 1. *Cont.*

Authors	Airborne Particulate Matters ($PM_{2.5}$, PM_{10})	Change-Point Detection	BCPD (Bayesian Change Point Detection)	Time Series Analysis
Liu et al. (2013) [11]	- -	Change point detection, relative density ratio estimation	-	Time series analysis -
Ruggieri and Antonellis (2016) [19]	-	-	Bayesian sequential change point detection	-
Keshavarz and Huang (2014) [20]	- -	- -	Bayesian and Expectation Maximization methods, multivariate change point detection	-
Kurt et al. (2018) [21]	- -	- -	Bayesian change point model, SIP-based DDoS attacks detection	- -
Wu et al. (2017) [22]	- - -	- - -	Post classification change detection, iterative slow feature analysis, Bayesian soft fusion	- - -
Gupta and Baker (2017) [26]	- -	- -	Spatial event rates, change point, Bayesian statistics, induced seismicity	-
Marino et al. (2017) [23]	- -	- -	Change point, multiple biomarkers, ovarian cancer	- -
Jeon et al. (2016) [24]	- -	- -	Abrupt change point detection, annual maximum precipitation, fused lasso	- -
Bardwell and Fearnhead (2017) [27]	-	-	Bayesian Detection, Abnormal Segments	Time series analysis
Chen et al. (2017) [25]	- -	- -	Bayesian change point analysis, extreme daily precipitation	- -
This study	Airbone Particulate Matter	Change point detection	Bayesian approach and likelihood ratio test for change point detection	Time series data

Table 2. Change point detection.

Authors	Change-Point Detection
Cabrieto et al. (2018) [38]	Detection of correlation changes by applying kernel change point detection on the running correlations
Keshavarz et al. (2018) [7]	Detecting a change in the mean of one-dimensional Gaussian process data in the fixed domain regime based on the generalized likelihood ratio test (GLRT)
Kucharczyk et al. (2018) [8]	Variance change point detection for fractional Brownian motion based on the likelihood ratio test
Lu et al. (2017) [9]	Graph-based structural change detection for rotating machinery monitoring by martingale-test method
Hilgert et al. (2016) [12]	On-line change detection for uncertain autoregressive dynamic models through nonparametric estimation (CUSUM-like scheme)
Górecki et al. (2017) [13]	Change point detection in heteroscedastic time series through Karhunen-Loeve expansions
Bucchia and Wendler (2017) [14]	Change point detection and bootstrap for Hilbert space valued random fields (Cramer-von Mises type test)
Zhou et al. (2015) [16]	A method for sequential detection of structural changes in linear quantile regression models.
Lu and Chang (2016) [17]	Detecting change-points for shifts in mean and variance using fuzzy classification maximum likelihood change-point (FCML-CP) algorithm
Liu et al. (2013) [11]	Change point detection in time-series data by relative density-ratio estimation
This study	Change point detection for airborne particulate matter ($PM_{2.5}$, PM_{10}) through Bayesian approach and likelihood ratio test

Table 3. BCPD (Bayesian change point detection).

Authors	BCPD (Bayesian Change Point Detection)
Ruggieri and Antonellis (2016) [19]	A sequential Bayesian change point algorithm was proposed that provides uncertainty bounds on both the number and location of change points
Keshavarz and Huang (2014) [20]	Bayesian and Expectation Maximization (EM) methods for change point detection problem of multivariate data with both single and multiple changes
Kurt et al. (2018) [21]	A Bayesian change point model for detecting SIP-based DDoS attacks
Wu et al. (2017) [22]	A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion
Gupta and Baker (2017) [26]	Estimating spatially varying event rates with a change point using Bayesian statistics: Application to induced seismicity
Marino et al. (2017) [23]	Change-point of multiple biomarkers in women with ovarian cancer
Jeon et al. (2016) [24]	Abrupt change point detection of annual maximum precipitation through fused lasso penalty function by using the Generalized Extreme Value (GEV) distribution
Bardwell and Fearnhead (2017) [27]	Bayesian Detection of Abnormal Segments in Multiple Time Series
Chen et al. (2017) [25]	Bayesian change point analysis for extreme daily precipitation
This study	Change point detection for airborne particulate matter ($PM_{2.5}$, PM_{10}) through Bayesian approach and likelihood ratio test

2. Problem Definition, Notation and Assumptions

2.1. Problem Definition

The major objective of this research is to develop a more precise, well defined and user friendly method for application on random hazards of PM to detect the change point of subjected air pollutant hazards at any unknown time (k) if it occurs at any area across the globe. The existing methods are either too complex and complicated for the application on random hazards of PM due to its complex and varying properties or not applicable to randomness of PM hazards. Therefore, still more studies are required to develop a such kind of methodology, which is easily understandable and appropriate to model the hazards of the PM concentrations from air quality standards that can also detect change points in these hazards. Secondly, this method could be applicable for any kind of time series and data distributions. Analysis of these changes need to be done, whether these change points are favorable or not for the environment. For this, a comparison of subjected pollutant hazards before and after a change point has to be done for the evaluation of pollution control programs adopted by environmental protection agencies. If hazards occurrences increase after the change point, then environmental policies have a negative impact which marks the failure of pollution control program, but if the hazards occurrences reduce after the change point, then it demonstrates the effectiveness of the pollution control program. Thirdly, an alteration in occurrences must be measured to define the new pollution control policies for further improvements in the current level of subjected air pollutant hazards.

For anticipated goals, the Bayesian approach will be used to determine posterior probabilities of pollutant occurrences and the likelihood ratio test will be used for identifying the change point in that Bayesian model. This suggested model would be numerically validated by using real-time data of particulate matters' concentrations in different areas of Seoul, South Korea, observed from January 2004 to December 2013. The change point (k) for for particulate matter ($PM_{2.5}$ and PM_{10}) hazards, the rate before the change point (θ), and the rate after the change point (λ) would be comprehensively analyzed. The central idea for using different regions is their considerably different features i.e., environment, population densities, and transportation vehicle densities. Hence, this study can also be a vision for the implementation of recommended model in different areas.Air quality standards for particular matter $PM_{2.5}$ and PM_{10} are given in Table 4. Results have been determined by following these standards.

Table 4. Region-wise air quality standards for particulate matters ($PM_{2.5}$ and PM_{10}).

Air Quality Standards for $PM_{2.5}$ (µg/m^3)		**Air Quality Standards for PM_{10} (µg/m^3)**	
European Standard ($PM_{2.5}$)	25	European Standard (PM_{10}) 24 h	50
American Standard ($PM_{2.5}$)	35	American Standard (PM_{10}) 24 h	150
Korean Standard ($PM_{2.5}$) 24 h	50	Korean Standard (PM_{10}) 24 h	100

Tables 5 and 6 illustrate some details regarding data collected for Guro, Nowon, Songpa, and Yongsan which exhibit standard-wise and location-wise percentage of polluted days. In case of $PM_{2.5}$, more than 44%, 21% and 8% days are polluted as per European, American and Korean standards respectively, which is alarming. Similarly, in case of PM_{10}, the polluted days concentrations as per European and Korean standards is more than 40% and 6% respectively and it could not be acceptable. Hence, there is a need to control hazards of PM.

Table 5. Particulate matter ($PM_{2.5}$) location-wise data.

Particle Pollution	**Criteria (µg/m^3)**	**Total Number of Days**	**Effective Readings**	**Dangerous Concentration**	**%Age against Total Days**	**%Age against Effective Readings**
$PM_{2.5}$ Guro's data (Seoul, South Korea)						
European Standard	25	2498	2486	1125	45.04%	45.25%
American Standard	35	2498	2486	532	21.30%	21.40%
Korean Standard	50	2498	2486	207	8.29%	8.33%
$PM_{2.5}$ Nowon's data (Seoul, South Korea)						
European Standard	25	3228	3031	1371	42.47%	45.23%
American Standard	35	3228	3031	718	22.24%	23.69%
Korean Standard	50	3228	3031	255	7.90%	8.41%
$PM_{2.5}$ Songpa's data (Seoul, South Korea)						
European Standard	25	3653	3388	1547	42.35%	45.66%
American Standard	35	3653	3388	795	21.76%	23.47%
Korean Standard	50	3653	3388	281	7.69%	8.29%
$PM_{2.5}$ Yongsan's data (Seoul, South Korea)						
European Standard	25	3653	3456	1537	42.08%	44.47%
American Standard	35	3653	3456	746	20.42%	21.59%
Korean Standard	50	3653	3456	280	7.66%	8.10%

Table 6. Particulate matter (PM_{10}) location-wise data.

Particle Pollution	Criteria (µg/m^3)	Total Number of Days	Effective Readings	Dangerous Concentration	%Age against Total Days	%Age against Effective Readings
			PM_{10} Guro's data (Seoul, South Korea)			
European Standard	50	3653	3540	1580	43.25%	44.63%
American Standard	150	3653	3540	60	1.64%	1.69%
Korean Standard	100	3653	3540	279	7.64%	7.88%
			PM_{10} Nowon's data (Seoul, South Korea)			
European Standard	50	3653	3531	1444	39.53%	40.89%
American Standard	150	3653	3531	41	1.12%	1.16%
Korean Standard	100	3653	3531	239	6.54%	6.77%
			PM_{10} Songpa's data (Seoul, South Korea)			
European Standard	50	3653	3467	1490	40.79%	42.98%
American Standard	150	3653	3467	50	1.37%	1.44%
Korean Standard	100	3653	3467	240	6.57%	6.92%
			PM_{10} Yongsan's data (Seoul, South Korea)			
European Standard	50	3653	3508	1566	42.87%	44.64%
American Standard	150	3653	3508	60	1.64%	1.71%
Korean Standard	100	3653	3508	282	7.72%	8.04%

2.2. Notation

The list of notation to represent the random variables and parameters is as follows:

Indices

i replication or sequence, i = 1, 2, ...
j position in the chain, j = 1, 2, ...n

Random variables

Y random process
y variable (Y) at any given point
y_i variable (Y) at point i where $i \in 0, 1, 2...$

Parameters

k change point in the random process
θ parameter before change point k associated with probability distribution function of random variable Y
λ parameter after change point k associated with probability distribution function of random variable Y

Variables

$Pr(\theta)$ prior distribution for parameter θ
$Pr(\theta|y_i)$ posterior distribution for parameter θ
$Pr(\lambda|y_i)$ posterior distribution for parameter λ
$Pr(y_i|\theta)$ likelihood or sampling model
V mean of the chain or replications (Average of daily pollutant concentrations)
V_{ij} jth observation from the ith replication

V_i	mean of ith replication
V	mean of m replications
B	between sequence variance represents the variance of replications with the mean of m replications
S_i^2	variance for all replications
W	within sequence variance is the mean variance for m replications
$Var(V)$	overall estimate of the variance of V in the target distribution
$\sqrt{R}$	estimated potential scale reduction for convergence

2.3. Assumptions

The following assumptions were used for the proposed model:

1. Y represents the number of times an event occurs in time t and Y is always positive real numbers $y \in 1, 2...$ that can be any random value.
2. $Y(0) = 0$ means that no event occurred at time $t = 0$.
3. Time series random data observed on equal interval of lengths.
4. The particulate matter daily concentrations or occurrence of events follow specific random probability distribution function.
5. The particulate matter daily concentrations in any interval of length (t) is a random variable and number of times event occurs is also positive random variable with parameter $(rate = \theta)$.

3. Mathematical Model

3.1. Formulation of Mathematical Model

The probability distribution function of a random variable Y at any given point y in the sample space is given as follows:

$$f(y;\theta) = Pr(Y = y|\theta)\ for\ y \in 1, 2...$$

There could be a single parameter or multiple parameters depending upon the probability distribution function of random variable Y.

The change point for random process Y is being detected by the likelihood ratio test and that is a statistical test used for comparing the goodness of fit for two statistical 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.

$$\begin{aligned} &f(\text{Change point}|Y, \text{Eexpectation before change point, Expectation after change point}) \\ &= \frac{L\big(Y; \text{Change point, Expectation before change point, Expectation after change point}\big)}{\sum_{j=1}^{n} L\big(Y; j, \text{Change point, Expectation before change point, Expectation after change point}\big)} \end{aligned} \quad (1)$$

and parameters' comparison before and after the change-point is also being done.

Let the change point in the random process be denoted by k and θ be the random variable parameter before change point k while λ be the random variable parameter after change point k. It can be represented as:

$$y_i \sim pdf(\theta)\ for\ i = 1, 2, ..., k$$

$$y_i \sim pdf(\lambda)\ for\ i = k + 1, k + 2, ..., n$$

Hence,

$$f(y;\theta) = Pr(Y = y_i|\theta)\, for\ i = 1, 2, ..., k$$

$$f(y;\lambda) = Pr(Y = y_i|\lambda)\, for\ i = k+1, k+2, ..., n$$

The joint pdf (probability density function) is the product of marginal pdf. If random variable $Y = y_i$ with parameter θ is modelled, then joint pdf of our sample data will be as below:

$$Pr(Y = y_i|\theta) = \prod_{i=1}^{n} Pr(y_i|\theta)\ for\ i \in 0, 1, 2, ..., n$$

A class of prior densities is conjugate for the likelihood/sampling model $Pr(y_i|\theta)$ if the posterior distribution is also in the same class. Therefore, prior distribution $Pr(\theta)$ and posterior distribution $Pr(\theta|y_i)$ will follow the same conjugate prior distribution to the likelihood/sampling model $Pr(y_i|\theta)$. However, the likelihood $Pr(y_i|\theta)$ follows the random distribution based on data. Therefore, the prior distribution $Pr(\theta)$ of parameters and posterior distribution $Pr(\theta|y_i)$ of the same parameters must be same and conjugate for Bayesian analysis. Bayes theorem for parameter's θ and λ is as follows:

$$Pr(\theta|y_i) \propto Pr(\theta)Pr(y_i|\theta)$$

$$Pr(\lambda|y_i) \propto Pr(\lambda)Pr(y_i|\lambda)$$

By applying Bayes theorem, the posterior distribution of model parameters θ and λ can be determined

$$Pr(\theta|y_i) = \frac{Pr(y_i|\theta)Pr(\theta)}{Pr(y_i)\ for\ i \in 1, 2, ..., k}$$

$$Pr(\lambda|y_i) = \frac{Pr(y_i|\lambda)Pr(\lambda)}{Pr(y_i)\ for\ i \in k+1, k+2, ..., n}$$

As,

$$L(\theta|Y) = f_\theta(Y) = f(Y|\theta)$$

Now, apply likelihood ratio test statistic for change point detection

$$f(k|Y, \theta, \lambda) = \frac{L\big(Y; k, \text{Expectation before change point, Expectation after change point}\big)}{\sum_{j=1}^{n} L\big(Y; j, \text{Expectation before change point, Expectation after change point}\big)}$$

And likelihood will be determined as given by:

$$L(Y; k, \theta, \lambda) = \Big[exp\Big(k((\text{Expectation after } k \text{ point} - \text{Expectation before } k \text{ point})\Big)\Big(\frac{\text{Expectation before } k}{\text{Expectation after } k}\Big)^{\sum_{i=1}^{k} y_i}\Big]$$

The change point k is uniform over y_i. Please note that θ, λ and k are all independent of each other.

3.1.1. Convergence of the Parameters

A single simulation run of a somewhat arbitrary length cannot represent the actual characteristics of the resulting model. Therefore, to estimate the steady-state parameters, the Gelman-Rubin Convergence diagnostic has to be applied in which target parameters are estimated by running multiple sequences of the chain. m replications of the simulation ($m \geq 10$) are made, each of length $n = 1000$. If the target distribution is unimodal then Cowles and Carlin recommends that we must run at least ten chains, as this approach monitors the scalar numbers of interest in the analysis. Therefore, the mean rate of pollutant concentrations is a parameter of interest that is denoted by V.

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

Let V_{ij} be the jth observation from the ith replication

$$V_{ij},\ i = 1, 2, ..., m\ j = 1, 2, ..., n$$

Mean of ith replication

$$V_i = \frac{1}{n} \sum_{j=1}^{n} V_{ij}$$

Mean of m replications

$$V = \frac{1}{m} \sum_{i=1}^{m} V_i$$

The between sequence variance represents the variance of replications with the mean of m replications calculated as follows:

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

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

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

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

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

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 diagnosed by calculating

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

This factor $\sqrt{R}$ (estimated potential scale reduction) is the ratio between the upper and lower bound on the space range of V which is used to estimate the factor by which $Var(V)$ could be reduced through more iterations. Further iterations of the chain must be run if the potential scale reduction is high. Run the replications until R is less than 1.1 or 1.2 for all scalar summaries.

3.1.2. Flowchart

The flowchart (Figure 1) for change point k detction, for any random process Y, is given as follows:

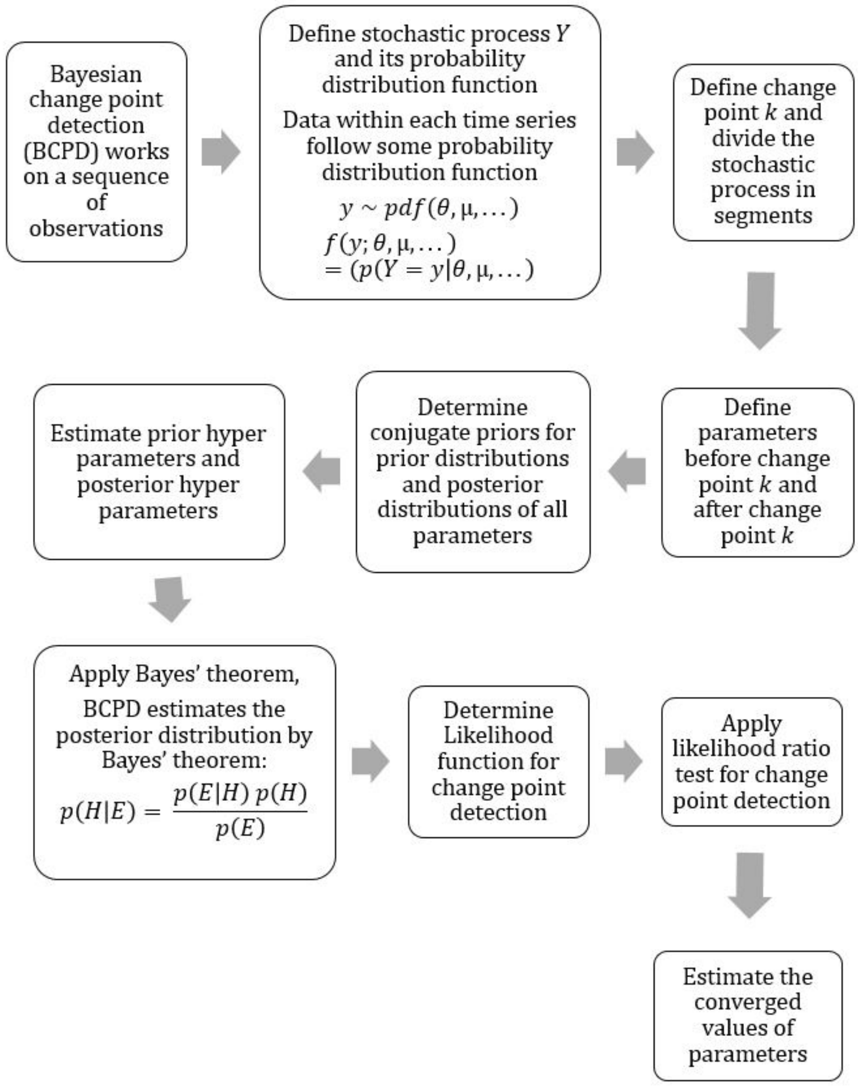

Figure 1. Flowchart for change point (k) detection.

3.2. Comparison Method for Change Point Detection

A change point analysis has been done by using a combination of CUSUM (cumulative sum control chart) and bootstrapping for comparative analysis.

3.2.1. The CUSUM (Cumulative Sum Control Chart) Technique

The CUSUM (cumulative sum control chart) 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.

$$\bar{X} = \left(\frac{X_1 + X_2 + X_3 + \ldots.,}{n} \right)$$

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

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

Plot the series and the cumulative sum is not the cumulative sum of the values. Instead it is the cumulative sum of differences between the values and the average. Because the average is subtracted from each value, the cumulative sum also ends at zero.

Interpreting a CUSUM chart requires some practice. Suppose that during a period of time the values tend to be above the overall average. Most of the values added to the cumulative sum will be positive and the sum will steadily increase. A segment of the CUSUM chart with an upward slope indicates a period where the values tend to be above the overall average. Likewise a segment with a downward slope indicates a period of time where the values tend to be below the overall average. A sudden change in direction of the CUSUM indicates a sudden shift or change in the average. Periods where the CUSUM chart follows a relatively straight path indicate a period where the average did not change.

3.2.2. Bootstrap Analysis

A confidence level can be determined for the apparent change by performing a bootstrap analysis. Before performing the bootstrap analysis, an estimator of the magnitude of the change is required. One choice, which works well regardless of the distribution and despite multiple changes, is S_{diff} defined as:

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

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

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

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

1. Generate a bootstrap sample of n units, denoted $X^0{}_1, X^0{}_2, X^0{}_3, \ldots X^0{}_n$ by randomly reordering the original n values. This is called sampling without replacement.
2. Based on the bootstrap sample, calculate the bootstrap CUSUM, denoted $S^0{}_0, S^0{}_1, S^0{}_2, \ldots S^0{}_n$.
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}.

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, it can be estimated that how much S_{diff} would vary if no change took place. It would be compared with the S_{diff} value calculated from the data in its original order to determine if this value is consistent with what has been expected if no change occurred. If bootstrap CUSUM charts tend to stay closer to zero than the CUSUM of the data in its original order, this leads one to suspect that a change must have 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 strong evidence that a change did in fact occur. Ideally, rather than bootstrapping, one would like to determine the distribution of $S^0{}_{diff}$ based on all possible reordering of the data. However, this is generally not feasible. A better estimate can be obtained by increasing the number of bootstrap samples. Bootstrapping results in a distribution free approach with only a single assumption, that of an independent error structure. Both control charting and change-point analysis are based on the mean-shift model. Let $X_1, X-2, X_3, \ldots$ represent the data in time order. The mean-shift model can be written as

$$X_i = \mu_i + \epsilon_i$$

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. It is assumed that the ϵ_i are independent with means of zero. Once a change has been detected, an estimate of when the change occurred can be made. One such estimator is the CUSUM estimator. Let m be such that:

$$\mid S_m \mid = \max_{i=0,1,2,\ldots,} \mid S_i \mid$$

S_m is the point furthest from zero in the CUSUM chart. The point m estimates last point before the change occurred. The point $m+1$ estimates the first point after the change. Once a change has been detected, the data can be broken into two segments, one each side of the change-point, 1 to m and $m+1$ to 24, estimating the average of each segment, and then analyzing the two estimated averages.

4. Numerical Example

The formulated mathematical model has been used for the numerical verification and the validity of the model has also been checked. That is why real-time data of particulate matter hazards for four different sites of Seoul, South Korea has been utilized for this investigation.

4.1. *Particulate Matter ($PM_{2.5}$) and (PM_{10}) Change Points for Four Different Sites*

Two dissimilar cases need to be considered

4.1.1. *Case 1—When There Is No Hazard*

In this case, there is no hazard and concentrations of particulate matter does not exceed the threshold value of the standards. Therefore, there will be no polluted day and random variable Y would always be $y = 0$. Hence, due to zero hazard in the concentrations of particulate matter, this model has not been applied.

4.1.2. *Case 2—When There Are Hazards*

In this case, several polluted days for particulate matter ($PM_{2.5}$ and PM_{10}) concentrations are considered as a Poisson process. A counting process is a Poisson counting process with the rate $\theta > 0$. Here, we report the results obtained by applying the method described in Section 3 to the particulate matter ($PM_{2.5}$ and PM_{10}) concentrations for four different sites (Guro, Nowon, Songpa, and Yongsan) in Seoul, South Korea. We used the daily data observed from January 2004 to December 2013 to compute the change point of both pollutants.

$$f(y;\theta) = (Pr(Y = y|\theta)) = Poisson(y,\theta) = \frac{e^{-\theta}\theta^y}{y!} \; for \; y \in 1,2,\ldots.,n$$

Poisson distribution is the number of events occurring in a given time period. So in this case, occurrence of the number of polluted days in a month is taken as Poisson distribution. The rate of polluted days for both $PM_{2.5}$ and PM_{10} are given in Tables 7 and 8 respectively.

Table 7. $PM_{2.5}$ Poisson Process.

$PM_{2.5}$ Poisson Process (Rate) θ				
Area	**Distribution**	**European Standards**	**American Standards**	**Korean Standards**
Guro	Poisson	13.720	6.488	2.524
Nowon	Poisson	12.934	6.774	2.406
Songpa	Poisson	12.892	6.625	2.342
Yongsan	Poisson	12.808	6.217	2.333

Table 8. PM_{10} Poisson Process.

PM_{10} Poisson Process (Rate) θ				
Area	**Distribution**	**European Standards**	**American Standards**	**Korean Standards**
Guro	Poisson	13.167	0.500	2.325
Nowon	Poisson	12.033	0.342	1.992
Songpa	Poisson	12.417	0.417	2.000
Yongsan	Poisson	13.050	0.500	2.350

The change point for this Poisson process has to be detected to know whether a change has occurred, the most likely month in which change has occurred, and if the rate of polluted days has increased or decreased after the change point. It has been assumed that the number of polluted days for (particulate matter) $PM_{2.5}$ and PM_{10} concentrations follows a Poisson distribution with a mean rate θ until the month k. After the month k, the polluted days are distributed according to the Poisson distribution with a mean rate λ. It can be represented as:

$$y_i \sim Poisson(\theta) \, for \; i = 1, 2, ..., k$$

$$y_i \sim Poisson(\lambda) \, for \; i = k+1, k+2, ..., n$$

Hence,

$$f(y; \theta) = Pr(Y = y_i|\theta) \, for \; i = 1, 2, ..., k$$

$$f(y; \lambda) = Pr(Y = y_i|\lambda) \, for \; i = k+1, k+2, ..., n$$

If we model $Y = y_i$ as Poisson with mean rate θ then joint pdf of our sample data will be as below:

$$Pr(Y = y_i|\theta) = \prod_{i=1}^{n} pr(y_i|\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y!} = c(y_1, y_2, ...y_n)e^{-n\theta}\theta^{\sum y_i} \; i \in 0, 1, 2, ..., n$$

This means that whatever our conjugate class of densities is, it will have to include terms like $e^{-C_2\theta}\theta^{C_1}$ for constants C_1 and C_2. The simplest class of such densities, which include these terms and corresponding probability distributions, are known as family of Gamma distributions. Therefore, prior distribution $Pr(\theta)$ and posterior distribution $Pr(Y = \theta|y_1, y_2, ...y_n)$ will follow a Gamma distribution, but likelihood or sampling model $Pr(y_1, y_2, ...y_n|\theta)$ follow a Poisson distribution.

Therefore, the prior distributions of θ and λ, uncertain positive quantities θ and λ has $Gamma(a_1, b_1)$ and $Gamma(a_2, b_2)$ distributions respectively, where a_1 is shape parameter and b_1 is rate parameter for θ while a_2 is shape parameter and b_2 is rate parameter for λ

$$Pr(\theta) = Gamma(\theta, a_1, b_1) = \frac{b_1^{a_1}e^{-b_1\theta}\theta^{a_1-1}}{\Gamma(a_1)}$$

$$Pr(\lambda) = Gamma(\lambda, a_2, b_2) = \frac{b_2^{a_2}e^{-b_2\lambda}\lambda^{a_2-1}}{\Gamma(a_2)}$$

Gamma distribution is also conjugate prior of the rate (inverse scale) parameter of the Gamma distribution itself. That is why the rate parameter b_1 and b_2 will also follow a Gamma distribution with different shape and rate parameters as given below:

$$b_1 \sim Gamma(c_1, d_1) \text{ where } c_1 = \text{shape parameter } d_1 = \text{rate parameter}$$

$$b_2 \sim Gamma(c_2, d_2) \text{ where } c_2 = \text{shape parameter } d_2 = \text{rate parameter}$$

By applying Bayes theorem, posterior distributions for rate parameters θ, λ, b_1 and b_2 will be determined in the following way. Likelihood and prior distributions of θ

$$Pr(y_1, y_2, y_3, ..., y_n|\theta) \sim Poisson(\theta)$$

$$Pr(\theta) = Gamma(\theta, a_1, b_1)$$

$$Pr(\theta|y_1, y_2, y_3, ..., y_n) = \frac{Pr(y_1, y_2, y_3, ..., y_n|\theta)Pr(\theta)}{Pr(y_1, y_2, y_3, ..., y_n)} = (e^{-b_1\theta}\theta^{a_1-1}) \times (e^{-n\theta}\theta^{\Sigma y_i}) \times c(y_1, y_2, y_3, ..., y_n)$$

$$= (e^{-(b_1+n)\theta}\theta^{a_1+\Sigma y_i-1}) \times c(y_1, y_2, y_3, ..., y_n, a_1, b_1)$$

$$(\theta|y_1, y_2, y_3, ..., y_n) \sim Gamma(a_1 + \sum_{i=1}^{n} y_i, b_1 + n)$$

This is evidently a Gamma distribution. Hence, the conjugacy of Gamma family for the Poisson sampling model or likelihood is confirmed. Hence, it is concluded from the above that if:

$$\theta \sim Gamma(a_1, b_1)$$

$$Pr(y_1, y_2, y_3, ..., y_n|\theta) \sim Poisson(\theta)$$

Then:

$$(\theta|y_1, y_2, y_3, ..., y_n) \sim Gamma(a_1 + \sum_{i=1}^{n} y_i, b_1 + n)$$

Similarly, the posterior distributions of all parameters θ, λ, b_1 and b_2 can be determined as given below:

$$(\theta|y, \lambda, b_1, b_2, k) \sim Gamma(a_1 + \sum_{i=1}^{k} y_i, k + b_1)$$

$$(\lambda|y, \theta, b_1, b_2, k) \sim Gamma(a_2 + \sum_{i=k+1}^{n} y_i, k + b_2)$$

$$(b_1|y, \theta, \lambda, b_2, k) \sim Gamma(a_1 + c_1, \theta + d_1)$$

$$(b_2|y, \theta, \lambda, b_1, k) \sim Gamma(a_2 + c_2, \lambda + d_2)$$

As Gamma is a two-parameter family of continuous probability distribution. As a result, the function:

$$L(\theta|Y) = f_\theta(Y) = f(Y|\theta)$$

The likelihood ratio test statistic is:

$$f(k|Y, \theta, \lambda) = \frac{L(Y; k, \theta, \lambda)}{\sum_{j=1}^{n} L(Y; j, \theta, \lambda)}$$

The likelihood is determined as given by:

$$L(Y; k, \theta, \lambda) = exp(k(\lambda - \theta))\ (\theta/\lambda)^{\sum_{i=1}^{k} y_i}$$

For Bayesian approach, MATLAB has been 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. 10 replications of each simulation are made with 1100 observations in each replication. First 100 observations are discarded as a burn-in period. Replication Mean V_i of remaining 1000 observations has been taken for each replication as shown in Tables 9 and 10. Then mean (V) of replication mean has been taken to get the converged values of parameters.

Moreover, the CUSUM charts of polluted days as per European, American and Korean standards are shown in Figures 2–9 for four different sites Guro, Nowon, Songpa and Yongsan in Seoul, South Korea.

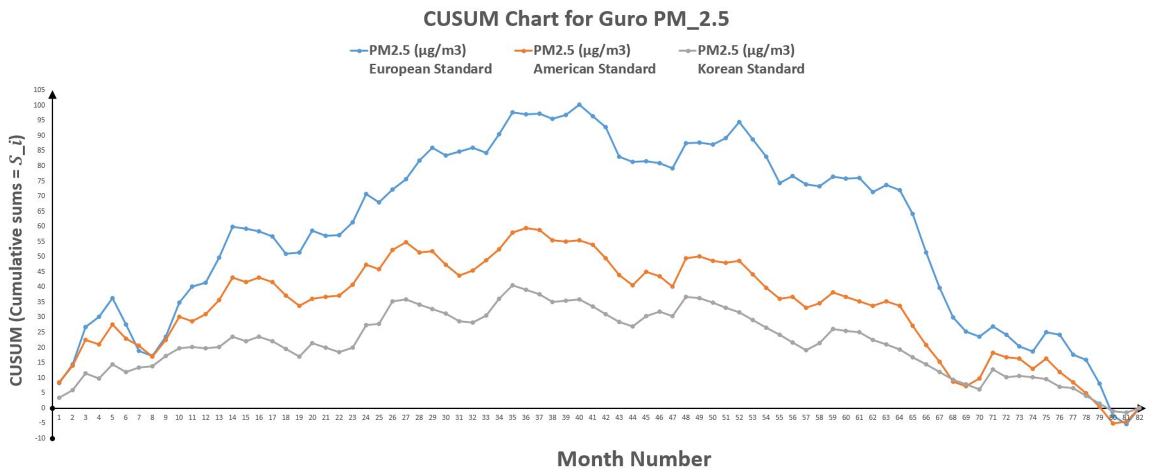

Figure 2. CUSUM chart for Guro $PM_{2.5}$.

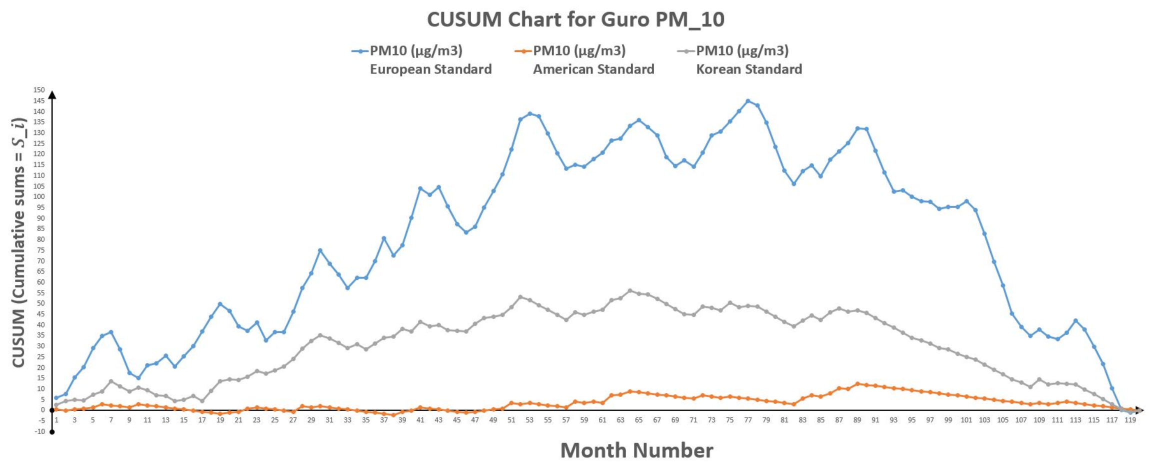

Figure 3. CUSUM chart for Guro PM_{10}.

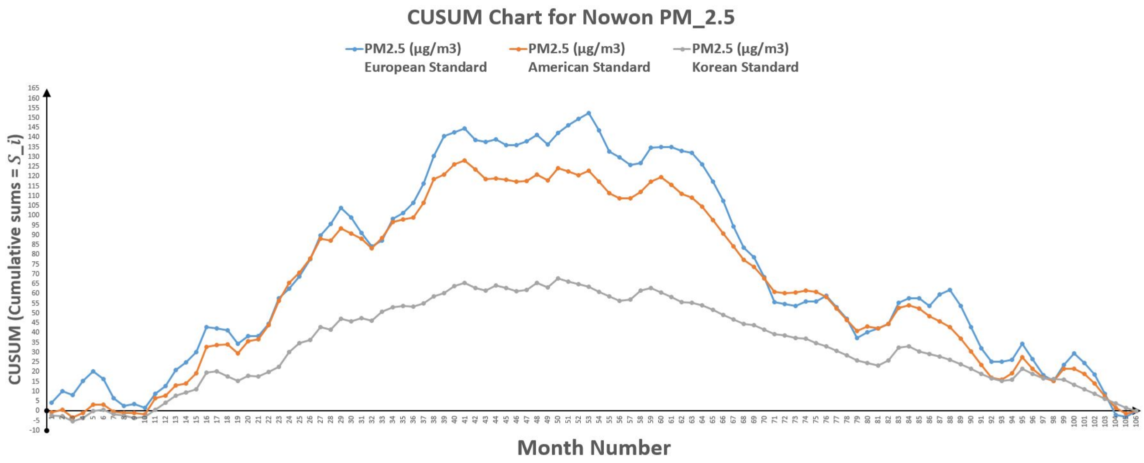

Figure 4. CUSUM chart for Nowon $PM_{2.5}$.

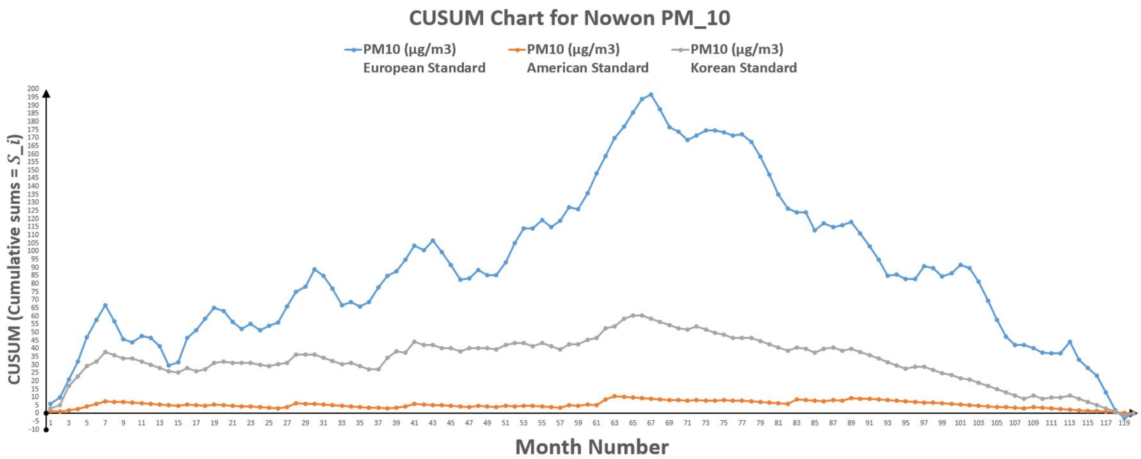

Figure 5. CUSUM chart for Nowon PM_{10}.

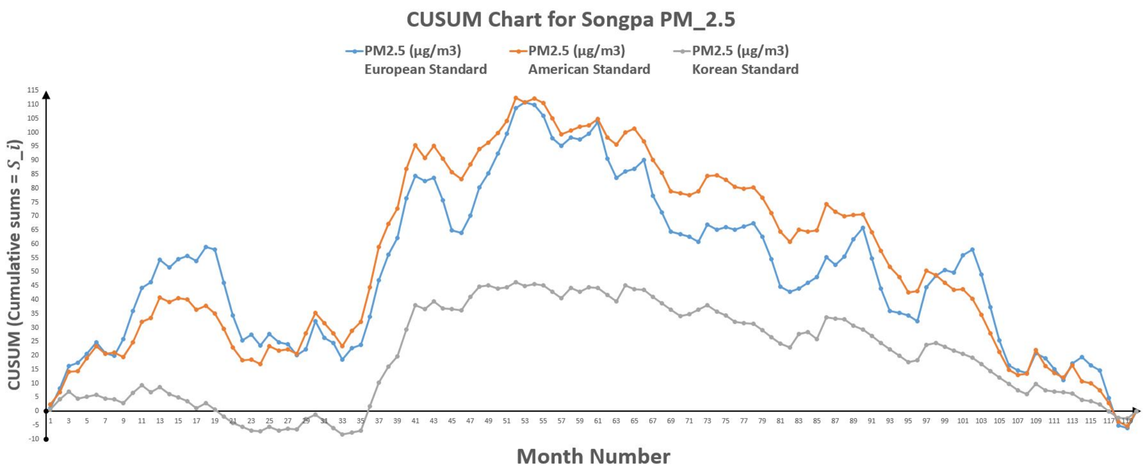

Figure 6. CUSUM chart for Songpa $PM_{2.5}$.

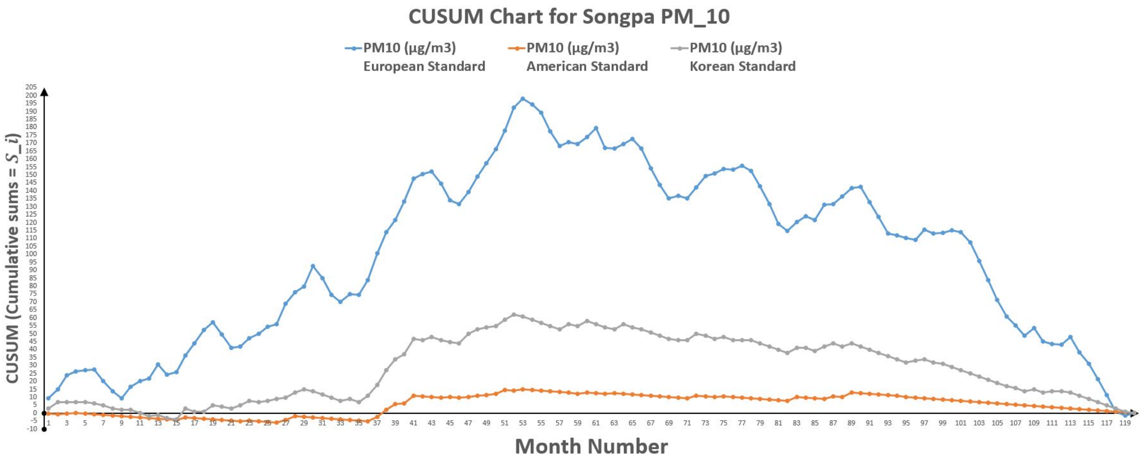

Figure 7. CUSUM chart for Songpa PM_{10}.

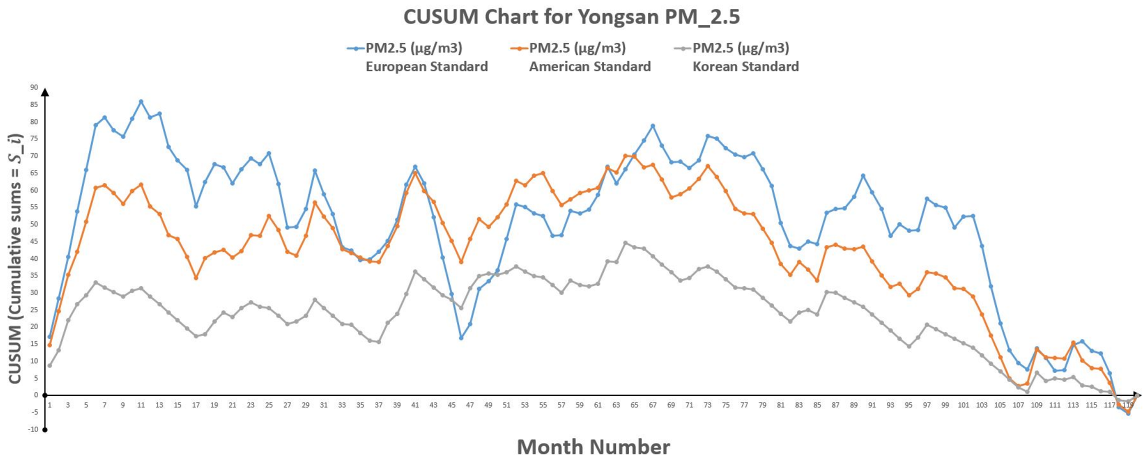

Figure 8. CUSUM chart for Yongsan $PM_{2.5}$.

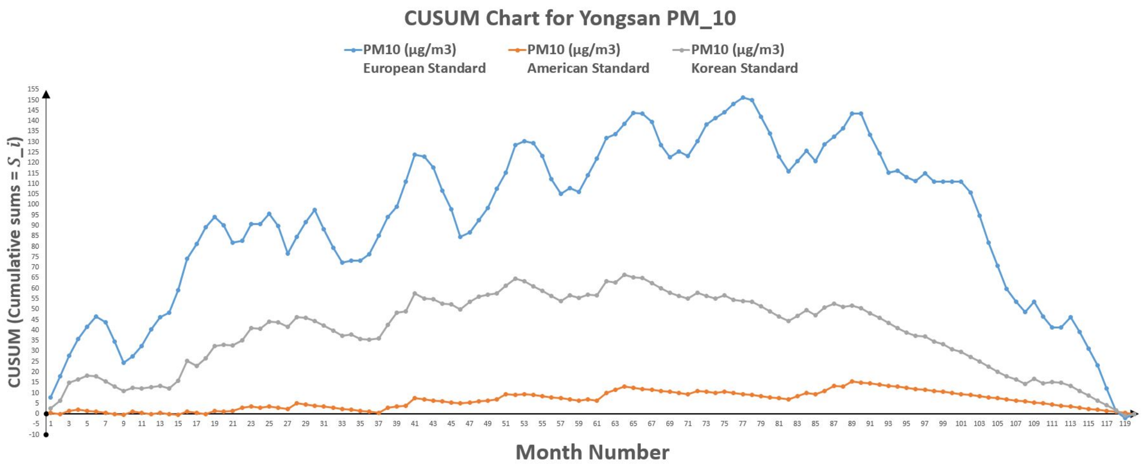

Figure 9. CUSUM chart for Yongsan PM_{10}.

Table 9. $PM_{2.5}$ Converged values of parameters (Bayesian approach).

$PM_{2.5}$ Converged Values of Parameters (Bayesian Approach)												
	Guro			**Nowon**			**Songpa**			**Yongsan**		
Replication Mean	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**
European standards												
V_1	15.96	11.17	43.29	15.88	10.30	49.98	14.93	11.24	53.69	24.80	12.13	6.92
V_2	16.03	11.25	42.53	15.89	10.34	49.83	14.90	11.28	53.81	24.46	12.15	8.07
V_3	16.00	11.18	43.12	15.90	10.32	49.76	14.91	11.25	53.92	24.81	12.07	7.49
V_4	16.01	11.17	43.19	15.90	10.30	49.96	14.91	11.24	54.05	24.77	12.14	7.75
V_5	15.98	11.12	43.52	15.88	10.26	50.13	14.93	11.23	53.59	24.56	12.09	7.48
V_6	15.96	11.14	43.61	15.87	10.28	50.12	14.89	11.24	53.98	24.07	12.08	9.48
V_7	15.95	11.17	43.41	15.84	10.30	50.12	14.88	11.26	53.91	24.61	12.03	8.20
V_8	16.04	11.21	42.76	15.94	10.34	49.65	14.93	11.26	53.73	24.69	12.13	7.87
V_9	16.01	11.19	42.75	15.88	10.30	49.78	14.90	11.24	53.71	24.50	12.10	7.83
V_{10}	15.96	11.15	43.59	15.85	10.30	50.23	14.88	11.25	54.07	24.23	12.12	8.30
(V) Mean of 10 replications	15.99	11.17	43.18	15.88	10.30	49.95	14.90	11.25	53.85	24.55	12.10	7.94
American standards												
V_1	8.35	5.32	33.46	9.82	4.77	42.11	8.73	4.95	53.32	16.12	5.69	6.17
V_2	8.24	5.33	33.67	9.84	4.80	41.76	8.72	4.96	53.31	16.04	5.71	6.17
V_3	8.22	5.30	34.31	9.82	4.78	42.12	8.72	4.95	53.35	16.00	5.69	6.21
V_4	8.29	5.30	34.10	9.84	4.79	41.75	8.72	4.95	53.32	16.08	5.68	6.18
V_5	8.31	5.30	33.49	9.86	4.77	41.67	8.72	4.93	53.32	16.05	5.67	6.17
V_6	8.27	5.31	33.70	9.80	4.77	42.11	8.71	4.94	53.39	15.96	5.68	6.20
V_7	8.24	5.30	34.01	9.88	4.79	41.66	8.70	4.95	53.33	15.92	5.69	6.19
V_8	8.25	5.32	33.70	9.86	4.80	41.61	8.74	4.96	53.25	16.16	5.69	6.16
V_9	8.47	5.33	32.46	9.83	4.78	41.66	8.72	4.94	53.21	15.99	5.68	6.16
V_{10}	8.38	5.33	33.28	9.82	4.79	41.89	8.71	4.95	53.42	15.96	5.69	6.21
(V) Mean of 10 replications	8.30	5.31	33.62	9.84	4.78	41.83	8.72	4.95	53.32	16.03	5.69	6.18

Table 9. *Cont.*

$PM_{2.5}$ Converged Values of Parameters (Bayesian Approach)												
	Guro			**Nowon**			**Songpa**			**Yongsan**		
Replication Mean	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**
Korean standards												
V_1	3.58	1.63	38.14	3.75	1.21	50.29	3.13	1.64	56.57	7.95	2.05	5.80
V_2	3.57	1.65	37.90	3.74	1.22	50.25	3.15	1.65	56.26	7.95	2.06	5.76
V_3	3.58	1.64	37.97	3.75	1.21	50.23	3.15	1.64	56.34	7.92	2.05	5.80
V_4	3.57	1.62	38.40	3.74	1.20	50.43	3.15	1.63	56.54	8.01	2.06	5.77
V_5	3.56	1.61	38.48	3.74	1.20	50.32	3.14	1.63	56.50	7.94	2.05	5.78
V_6	3.55	1.62	38.63	3.73	1.20	50.52	3.13	1.63	56.87	7.97	2.04	5.77
V_7	3.54	1.62	38.60	3.75	1.20	50.30	3.13	1.64	56.53	7.87	2.05	5.80
V_8	3.59	1.64	37.93	3.75	1.21	50.27	3.16	1.64	56.08	7.98	2.06	5.78
V_9	3.57	1.63	37.84	3.73	1.20	50.43	3.15	1.64	55.86	7.90	2.05	5.79
V_{10}	3.55	1.62	38.69	3.73	1.21	50.36	3.13	1.63	56.86	7.90	2.05	5.79
(V) Mean of 10 replications	3.56	1.63	38.26	3.74	1.21	50.34	3.14	1.64	56.44	7.94	2.05	5.78

Table 10. PM_{10} Converged values of parameters (Bayesian approach).

PM_{10} Converged Values of Parameters (Bayesian Approach)												
	Guro			**Nowon**			**Songpa**			**Yongsan**		
Replication Mean	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**	(θ) **Replication Mean**	(λ) **Replication Mean**	K **Replication Mean**
European standards												
V_1	14.71	9.20	85.53	14.95	8.36	67.04	16.11	9.49	53.29	14.57	8.41	90.09
V_2	14.75	9.27	85.10	14.93	8.39	67.10	16.08	9.52	53.30	14.57	8.50	89.53
V_3	14.74	9.20	85.49	14.93	8.36	67.16	16.11	9.45	53.29	14.58	8.47	89.63
V_4	14.75	9.19	85.47	14.93	8.35	67.17	16.08	9.47	53.52	14.57	8.39	90.16
V_5	14.76	9.12	85.65	14.92	8.33	67.13	16.06	9.44	53.75	14.56	8.35	89.85
V_6	14.72	9.18	85.59	14.91	8.35	67.20	16.06	9.47	53.43	14.54	8.42	89.98
V_7	14.71	9.18	85.75	14.90	8.36	67.13	16.14	9.48	53.36	14.52	8.41	90.12
V_8	14.77	9.25	85.07	14.96	8.37	67.05	16.08	9.39	53.93	14.61	8.51	89.23
V_9	14.73	9.18	85.29	14.92	8.35	67.03	16.07	9.47	53.26	14.60	8.48	88.67
V_{10}	14.72	9.19	85.67	14.91	8.37	67.16	16.03	9.46	53.91	14.53	8.39	90.18
(V) Mean of 10 replications	14.73	9.20	85.46	14.93	8.36	67.12	16.08	9.46	53.50	14.56	8.43	89.74
American standards												
V_1	0.63	0.12	90.65	0.69	0.13	66.56	0.56	0.07	87.63	0.66	0.04	91.61
V_2	0.63	0.12	90.76	0.63	0.11	72.52	0.55	0.07	89.08	0.66	0.04	91.33
V_3	0.63	0.12	90.67	0.52	0.08	81.79	0.56	0.07	87.86	0.66	0.04	91.46
V_4	0.63	0.12	90.34	0.53	0.09	81.32	0.57	0.07	86.80	0.66	0.04	91.41
V_5	0.63	0.12	90.23	0.76	0.14	60.70	0.56	0.06	88.35	0.67	0.04	91.03
V_6	0.63	0.12	90.48	0.52	0.08	81.91	0.56	0.07	87.98	0.66	0.04	91.50
V_7	0.63	0.12	90.53	0.45	0.06	89.85	0.56	0.06	88.72	0.66	0.05	91.18
V_8	0.63	0.12	90.63	0.82	0.15	56.42	0.56	0.09	88.73	0.66	0.04	91.11
V_9	0.63	0.13	89.33	0.64	0.11	71.37	0.56	0.07	88.19	0.66	0.04	91.69
V_{10}	0.64	0.12	89.79	0.59	0.11	75.61	0.56	0.07	88.26	0.66	0.04	91.29
(V) Mean of 10 replications	0.63	0.12	90.34	0.62	0.11	73.80	0.56	0.07	88.16	0.66	0.04	91.36

Table 10. *Cont.*

PM_{10} Converged Values of Parameters (Bayesian Approach)												
	Guro			**Nowon**			**Songpa**			**Yongsan**		
Replication Mean	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**	**(θ) Replication Mean**	**(λ) Replication Mean**	**K Replication Mean**
Korean standards												
V_1	2.89	0.91	85.25	2.92	0.89	65.68	3.16	1.09	52.78	3.32	1.11	67.44
V_2	2.90	0.92	85.17	2.91	0.89	65.67	3.17	1.10	52.65	3.33	1.13	66.93
V_3	2.89	0.90	85.96	2.92	0.89	65.69	3.17	1.09	52.79	3.32	1.12	67.52
V_4	2.88	0.88	86.67	2.91	0.88	65.69	3.17	1.09	52.83	3.30	1.08	68.72
V_5	2.90	0.91	84.97	2.91	0.88	65.67	3.16	1.08	52.79	3.30	1.09	68.37
V_6	2.88	0.89	86.13	2.91	0.88	65.72	3.16	1.09	52.75	3.27	1.06	69.89
V_7	2.87	0.87	87.12	2.90	0.89	65.70	3.16	1.09	52.64	3.28	1.08	68.93
V_8	2.89	0.89	86.14	2.92	0.88	65.67	3.18	1.09	52.59	3.33	1.12	67.09
V_9	2.88	0.89	86.27	2.90	0.88	65.63	3.16	1.09	52.56	3.31	1.10	67.66
V_{10}	2.88	0.90	86.17	2.90	0.89	65.74	3.15	1.09	53.15	3.30	1.10	68.08
(V) Mean of 10 replications	2.89	0.90	85.98	2.91	0.89	65.68	3.16	1.09	52.75	3.31	1.10	68.06

However, the bootstraps analysis of European standards has been shown in Figures 10–17 as given below.

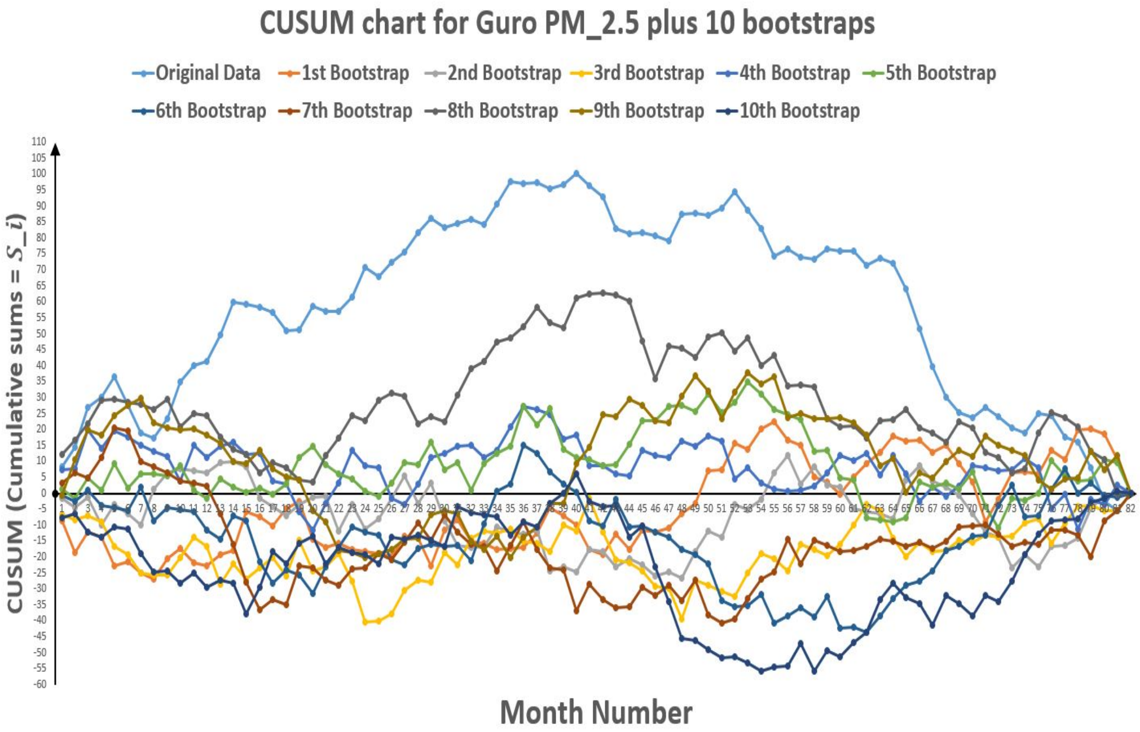

Figure 10. CUSUM chart for Guro $PM_{2.5}$ plus 10 bootstraps (European Standards).

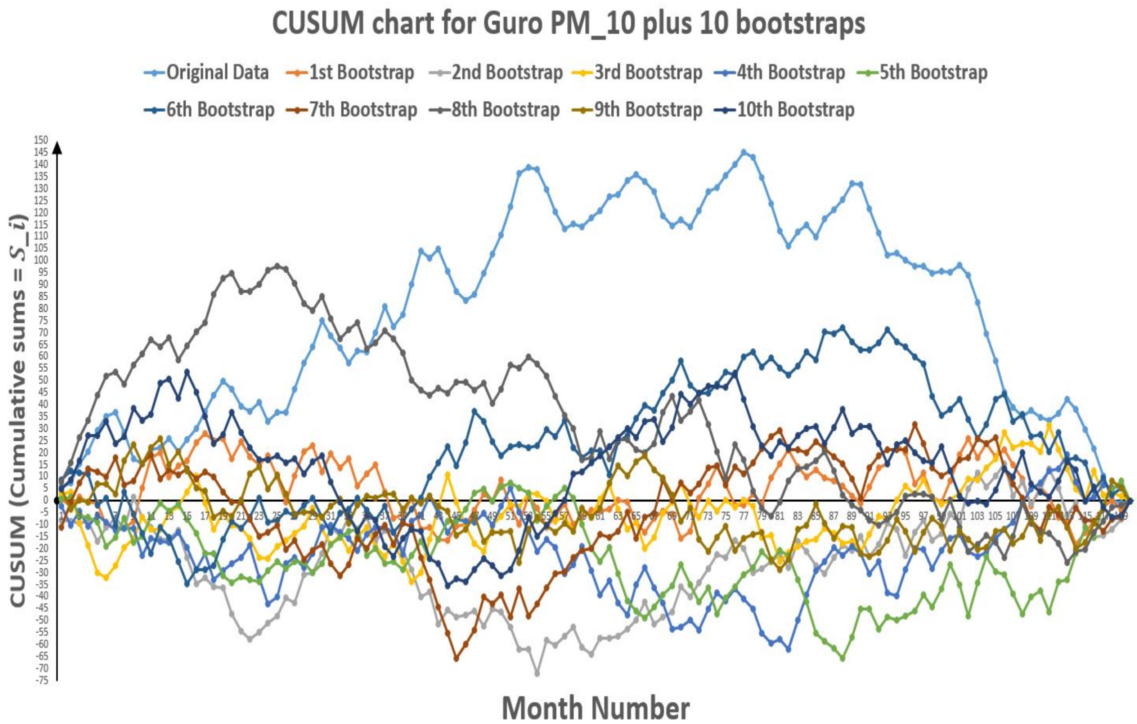

Figure 11. CUSUM chart for Guro PM_{10} plus 10 bootstraps (European Standards).

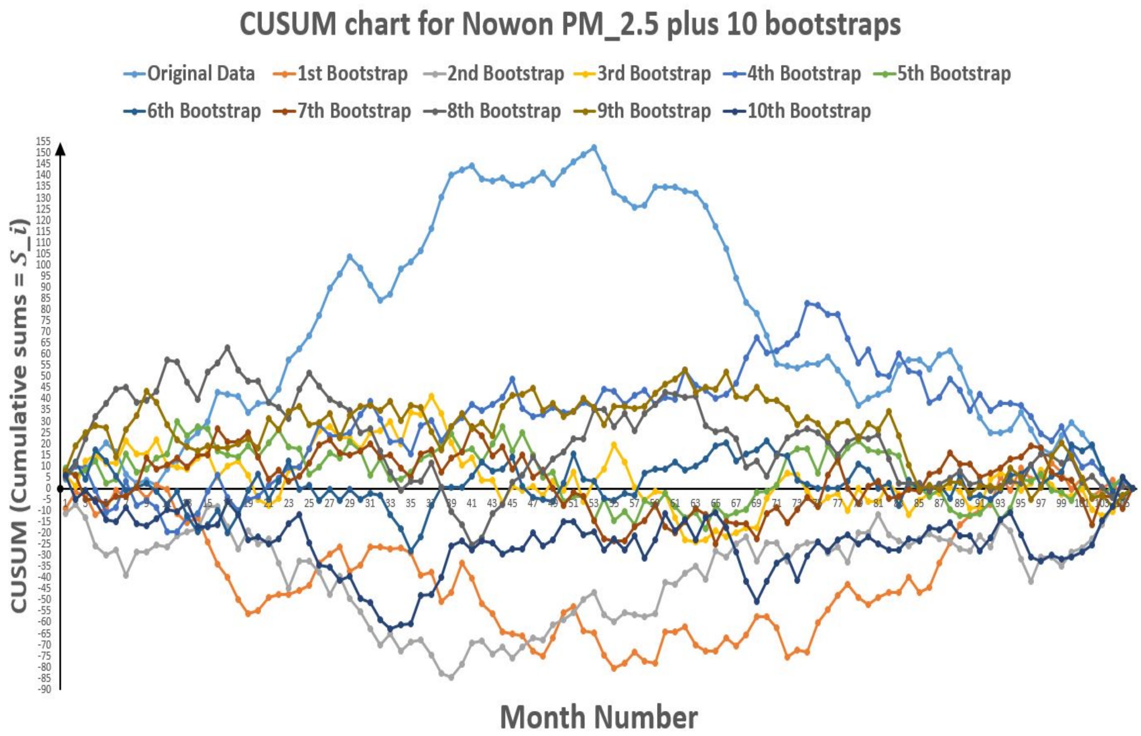

Figure 12. CUSUM chart for Nowon $PM_{2.5}$ plus 10 bootstraps (European Standards).

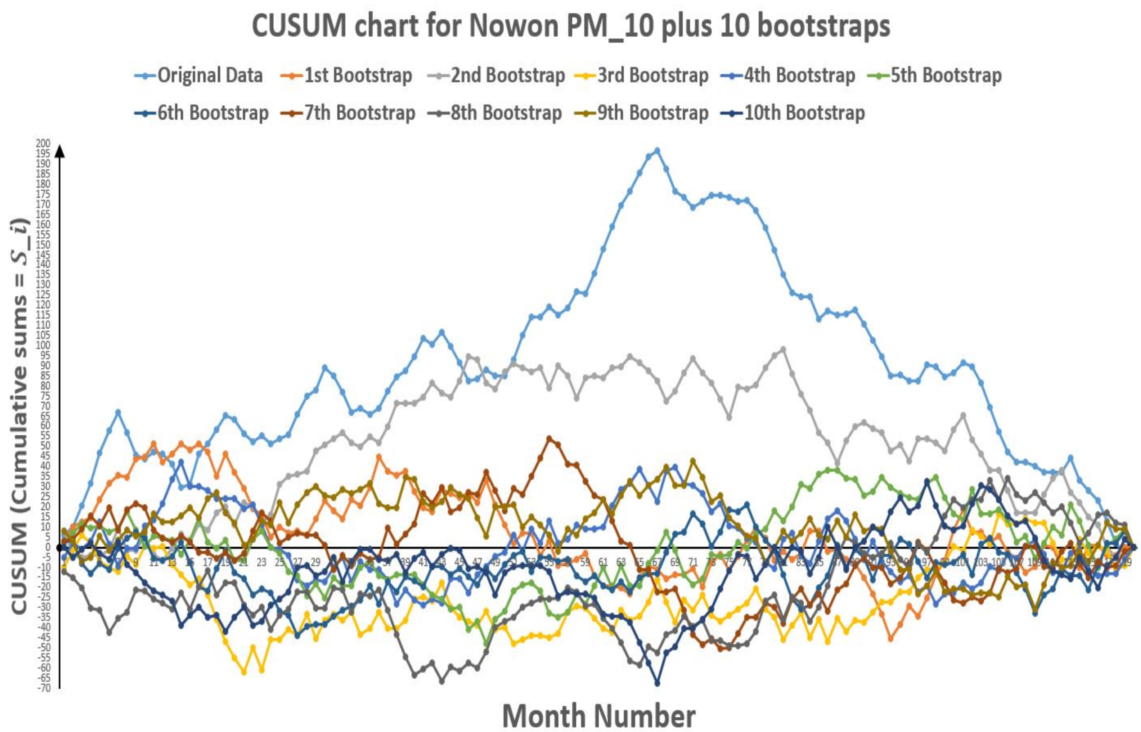

Figure 13. CUSUM chart for Nowon PM_{10} plus 10 bootstraps (European Standards).

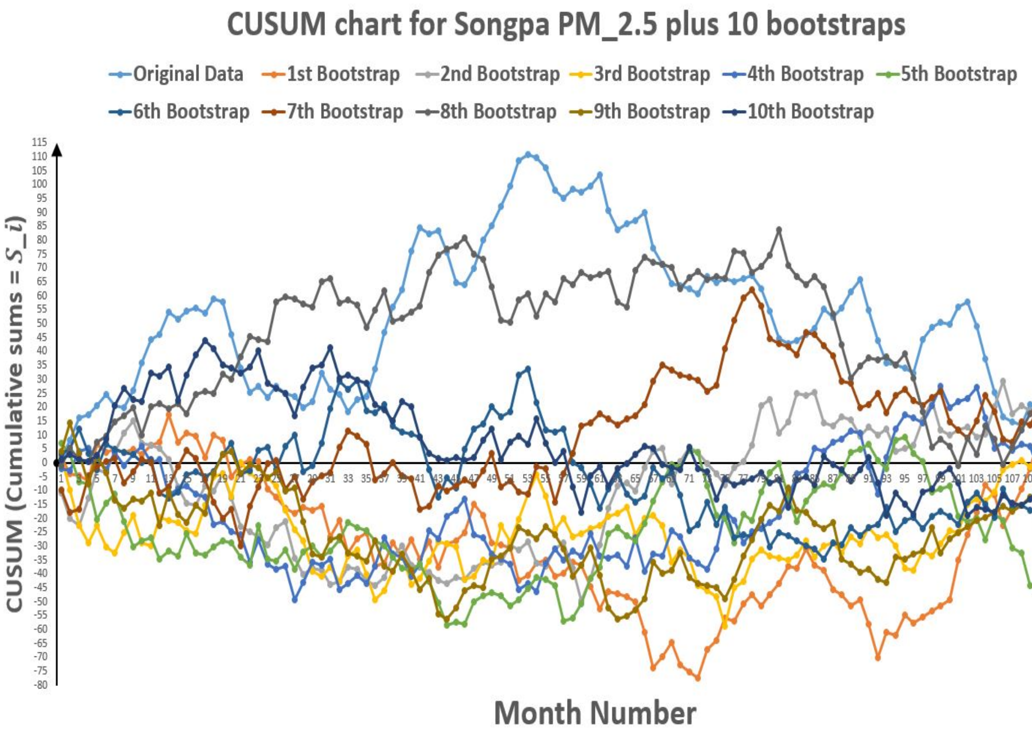

Figure 14. CUSUM chart for Songpa $PM_{2.5}$ plus 10 bootstraps (European Standards).

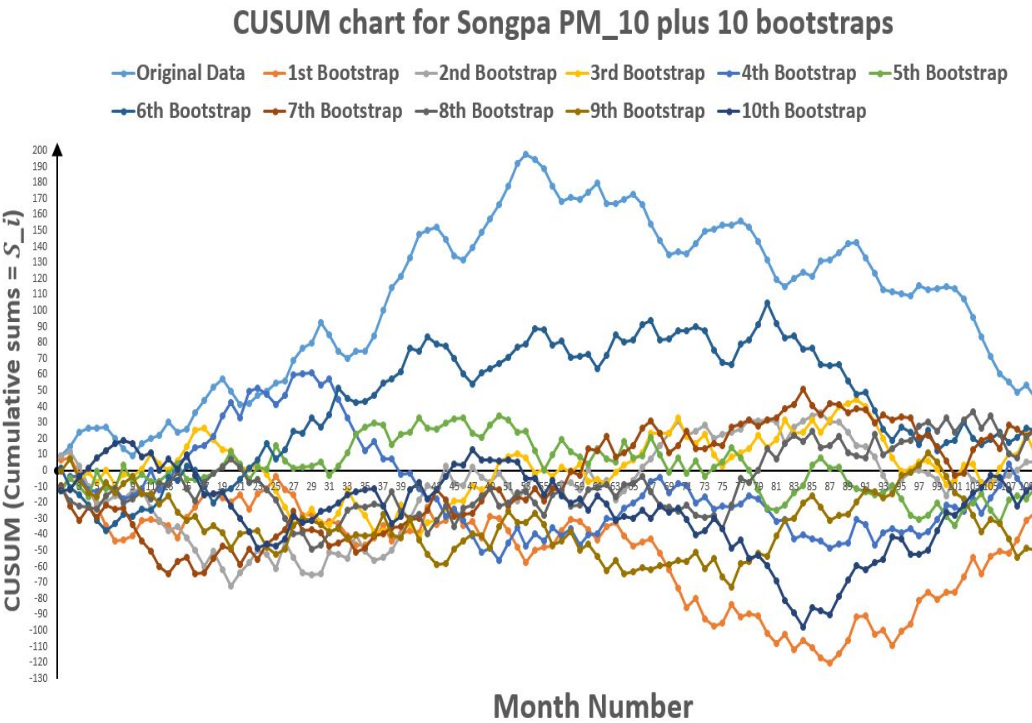

Figure 15. CUSUM chart for Songpa PM_{10} plus 10 bootstraps (European Standards).

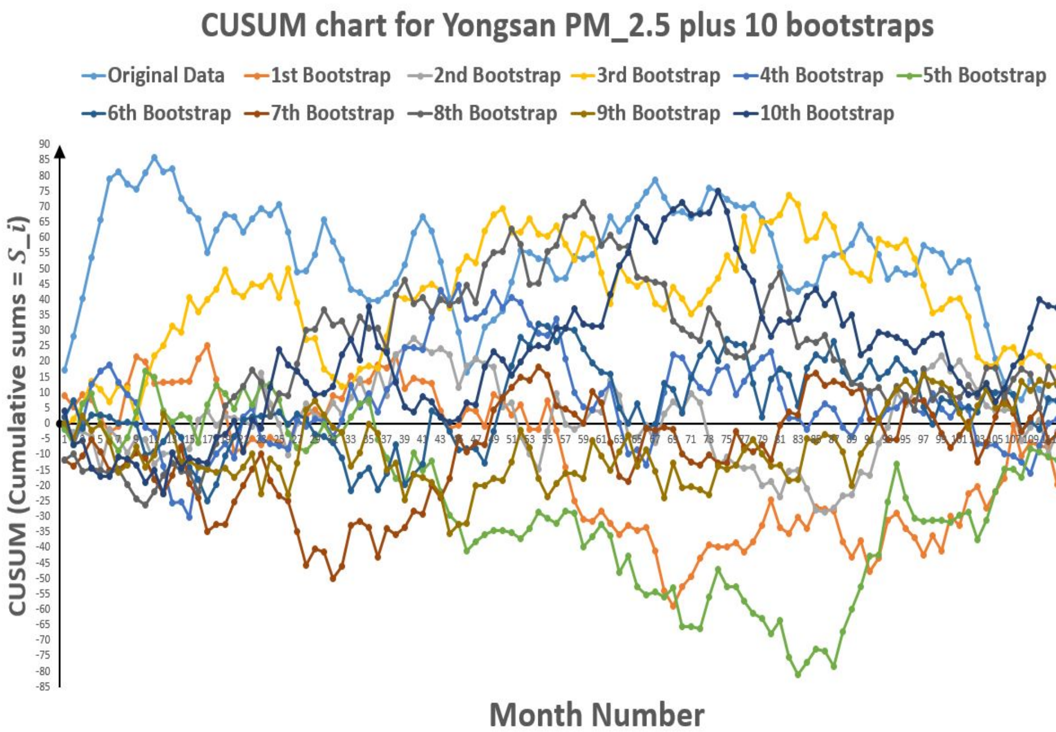

Figure 16. CUSUM chart for Yongsan $PM_{2.5}$ plus 10 bootstraps (European Standards).

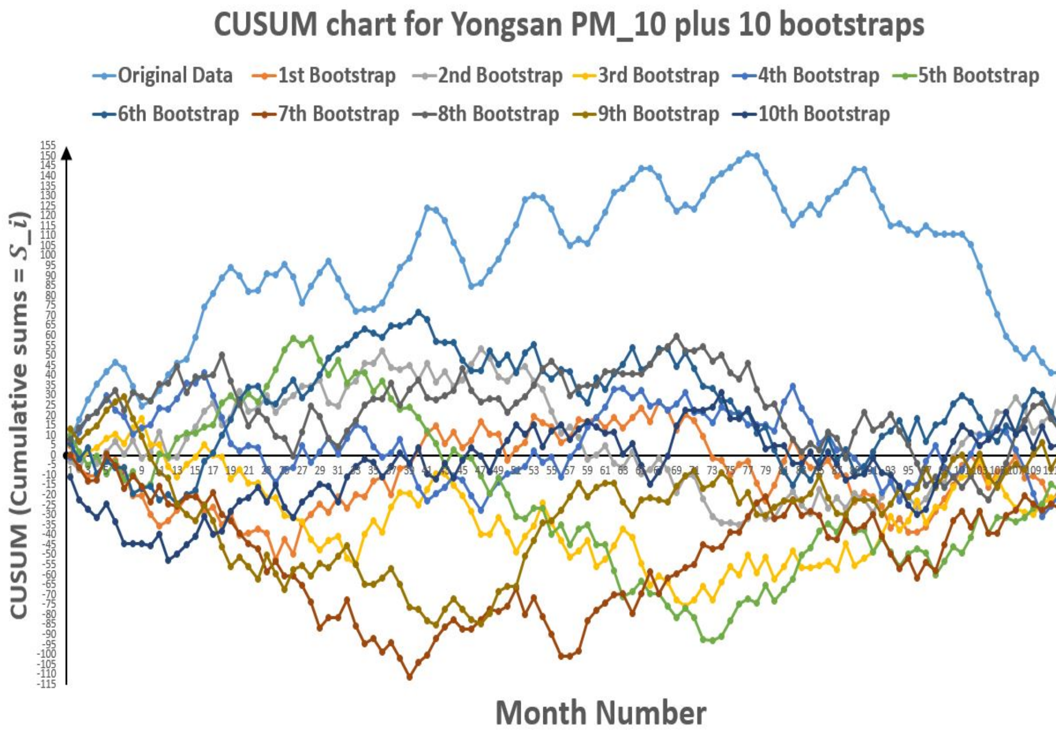

Figure 17. CUSUM chart for Yongsan PM_{10} plus 10 bootstraps (European Standards).

The change point k is discrete uniform over $(1, 2, 3...120)$ as there are 120 months in 10 years. Please note that θ, λ and k are all independent of each other.

5. Numerical Results

Two dissimilar approaches have been used to attain the results. First one is Bayesian approach, which is based on probability distributions. It can be applicable to any kind of data distribution. For this, firstly data distributions are defined and then proposed method is applied to acquire the results. This approach is better to apply for random data structures and time series. While the second method is based on CUSUM charts, this technique is directly applicable on the raw data, which is good for deterministic data structures. Summarized forms of particulate matter ($PM_{2.5}$ and PM_{10}) change point (k), the rate before change point (θ) and the rate after change point (λ) during the study period (2004–2013) for four different sites (Guro, Nowon, Songpa and Yongsan) in Seoul, South Korea are given in Tables 11–14. The results have been computed by following the European, American, and Korean Standards as discussed in Table 4.

5.1. $PM_{2.5}$ Change Point (k) through Bayesian Approach

In Table 11, the results obtained through Bayesian approach have been described, where (k) is the predicted change point varies for different areas and different air quality standards. The results indicate the reduction of polluted days after change point (k) for $PM_{2.5}$. While (θ) represents the per month rate of polluted days before change point, (k) and (λ) be the rate of per month polluted days after change point (k).

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

Table 12 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 leis somewhere between (k) and $(k+1)$. This method also shows the reduction of polluted days after change point as (θ) represents the per month rate of polluted days before change point and (λ) be the rate of per month polluted days after change point.

5.3. PM_{10} Change Point (k) through Bayesian Approach

Table 13 explains the results obtained for PM_{10} through Bayesian approach. Hence, the expected change point is (k) that differs for different areas and various air quality standards. These results show the reduction of polluted days after change point (k) for PM_{10}. While (θ) is the per month rate of polluted days before change point (k) and (λ) represents the rate of per month polluted days after change point (k).

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 14, 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 polluted days after change point. (θ) represents the per month rate of polluted days before change point and (λ) be the rate of per month polluted days after change point.

Table 11. $PM_{2.5}$ Change Point (k) for European, American and Korean Standards through Bayesian approach.

Bayesian Approach for Change Point Detection									
Area Seoul, South Korea	**(θ) Rate before Change Point**	**(λ) Rate after Change Point**	K **Change Point**	$\sqrt{R}\theta$ **Convergence Parameter for (θ)**	$Var(V)\theta$ **Variance for (θ)**	$\sqrt{R}\lambda$ **Convergence Parameter for (λ)**	$Var(V)\lambda$ **Variance for (λ)**	$\sqrt{R}K$ **Convergence Parameter for** K	$Var(V)K$ **Variance for** K
European standards									
Guro	15.99	11.17	43.18	1.000408	0.583	1.000719	0.541	1.000747	59.219
Nowon	15.88	10.30	49.95	0.999943	0.783	1.000094	0.472	1.000036	34.238
Songpa	14.90	11.25	53.85	1.000199	0.302	1.000007	0.180	1.000870	9.719
Yongsan	24.55	12.10	7.94	1.000898	21.961	1.000943	0.477	1.002031	90.053
American standards									
Guro	8.30	5.31	33.62	1.002540	1.064	1.000105	0.176	1.001256	75.773
Nowon	9.84	4.78	41.83	1.000595	0.296	1.000200	0.089	1.001159	13.007
Songpa	8.72	4.95	53.32	0.999857	0.172	0.999934	0.076	1.000255	2.395
Yongsan	16.03	5.69	6.18	0.999993	5.726	0.999956	0.101	1.000028	0.356
Korean standards									
Guro	3.56	1.63	38.26	1.000475	0.132	1.000439	0.052	1.000865	41.605
Nowon	3.74	1.21	50.34	0.999959	0.083	1.000377	0.025	0.999978	9.024
Songpa	3.14	1.64	56.44	1.000200	0.068	1.000144	0.031	1.000588	45.887
Yongsan	7.94	2.05	5.78	1.000136	1.498	1.000237	0.018	0.999780	0.339

Table 12. $PM_{2.5}$ Last point before change (k) and First point after change ($k+1$) for European, American and Korean Standards through CUSUM approach.

CUSUM Approach for Change Point Detection									
Area Seoul, South Korea	**(θ) Average Rate before Change**	**(λ) Average Rate after Change**	**K Last Point before Change**	**$k+1$ First Point after Change**	**$\mid S_m \mid$ Most Extreme Point**	**S_{max} Highest Point in CUSUM**	**S_{min} Lowest Point in CUSUM**	**S_{diff} Magnitude of Change**	**Confidence Level %**
European standards									
Guro	16.23	11.33	40	41	100.22	100.220	−5.280	105.500	100%
Nowon	15.81	10.06	53	54	152.50	152.500	−3.066	155.566	100%
Songpa	14.98	11.24	53	54	110.74	110.742	−6.108	116.850	100%
Yongsan	20.64	12.02	11	12	86.11	86.108	−5.192	91.300	70%
American standards									
Guro	8.14	5.20	36	37	59.44	59.439	−5.024	64.463	100%
Nowon	9.90	4.80	41	42	128.28	128.283	−3.321	131.604	100%
Songpa	8.79	4.97	52	53	112.50	112.500	−5.375	117.875	100%
Yongsan	7.31	4.96	64	65	70.13	70.133	−4.783	74.917	100%
Korean standards									
Guro	3.69	1.66	35	36	40.65	40.646	−1.476	42.122	100%
Nowon	3.76	1.20	50	51	67.72	67.717	−5.217	72.934	100%
Songpa	3.23	1.66	52	53	46.23	46.233	−8.275	54.508	100%
Yongsan	3.03	1.54	64	65	44.67	44.667	−1.667	46.333	100%

Table 13. PM_{10} Change Point (k) for European, American and Korean Standards.

Bayesian Approach for Change Point Detection									
Area Seoul, South Korea	(θ) **Rate before Change Point**	(λ) **Rate after Change Point**	K **Change Point**	$\sqrt{R}\theta$ **Convergence Parameter for** (θ)	$Var(V)\theta$ **Variance for** (θ)	$\sqrt{R}\lambda$ **Convergence Parameter for** (λ)	$Var(V)\lambda$ **Variance for** (λ)	$\sqrt{R}K$ **Convergence Parameter for** K	$Var(V)K$ **Variance for** K
European standards									
Guro	14.73	9.20	85.46	1.000072	0.425	1.000427	0.925	0.999931	64.845
Nowon	14.93	8.36	67.12	0.999879	0.464	0.999925	0.327	1.000040	3.184
Songpa	16.08	9.46	53.50	1.000300	0.643	1.001282	0.310	1.003363	9.133
Yongsan	14.56	8.43	89.74	1.000123	0.597	1.000437	1.492	1.001094	75.750
American standards									
Guro	0.63	0.12	90.34	0.999840	0.009	1.000320	0.005	1.002821	31.264
Nowon	0.62	0.11	73.80	1.041187	0.173	1.036355	0.011	1.053574	1089.508
Songpa	0.56	0.07	88.16	1.000135	0.016	1.000253	0.037	1.000710	174.232
Yongsan	0.66	0.04	91.36	0.999851	0.016	1.000296	0.004	1.000219	32.243
Korean standards									
Guro	2.89	0.90	85.98	1.000820	0.046	1.001576	0.053	1.003429	59.252
Nowon	2.91	0.89	65.68	0.999881	0.045	1.000137	0.016	1.000338	0.567
Songpa	3.16	1.09	52.75	0.999935	0.066	1.000075	0.017	1.001445	7.194
Yongsan	3.31	1.10	68.06	1.002209	0.077	1.003709	0.050	1.004787	81.344

Table 14. PM_{10} Last point before change (k) and First point after change ($k+1$) for European, American and Korean Standards through CUSUM approach.

CUSUM Approach for Change Point Detection									
Area Seoul, South Korea	**(θ) Average Rate before Change**	**(λ) Average Rate after Change**	**K Last point before Change**	**$k+1$ First Point after Change**	**$\mid S_m \mid$ Most Extreme Point**	**S_{max} Highest Point in CUSUM**	**S_{min} Lowest Point in CUSUM**	**S_{diff} Magnitude of Change**	**Confidence Level %**
European standards									
Guro	15.05	9.79	77	78	145.17	145.167	−0.833	146.000	100%
Nowon	14.97	8.32	67	68	196.77	196.767	−2.967	199.733	100%
Songpa	16.15	9.46	53	54	197.92	197.917	−1.583	199.500	100%
Yongsan	15.01	9.53	77	78	151.15	151.150	−1.950	153.100	100%
American standards									
Guro	0.64	0.10	89	90	12.50	12.500	−2.000	14.500	100%
Nowon	0.51	0.16	63	64	10.48	10.475	0.000	10.475	80%
Songpa	0.70	0.19	53	54	14.92	14.917	−5.833	20.750	100%
Yongsan	0.67	0.00	89	90	15.50	15.500	−0.500	16.000	10%
Korean standards									
Guro	3.20	1.32	64	65	56.20	56.200	−0.675	56.875	100%
Nowon	2.91	0.87	66	67	60.55	60.550	−1.008	61.558	100%
Songpa	3.19	1.09	52	53	62.00	62.000	−4.000	66.000	100%
Yongsan	3.39	1.16	64	65	66.60	66.600	−0.650	67.250	100%

6. Disussion

As the results of two different approaches have been described in the previous section.

1. Bayesian approach is based on probability distributions, which can be applicable on any kind of data distribution. In this case, firstly data distributions are defined and then proposed method is applied to acquire the results. This approach is better to apply for random data structures and time series.
2. CUSUM Approach is directly applied on the raw data, which is good for deterministic data structures.

6.1. Guro (Seoul, South Korea)

Guro is located in the southwestern part of Seoul, and has an important position as a transport link which includes railroads and land routes. The largest digital industrial complex in Korea is also positioned in Guro, centering on research and development activities as well as advanced information and knowledge industries. That is why, the policies of the Ministry of Environment in South Korea have influenced the concentrations of particulate matters ($PM_{2.5}$ and PM_{10}) in Guro and rate of polluted days has reduced in any of the cases.

6.1.1. Bayesian Approach

Bayesian method is better to apply for random time series data. If we look in case of Guro, Table 11 indicates that for $PM_{2.5}$ change-point (k) of polluted days were 43.18, 33.62, and 38.26 according to European, American, and Korean standards respectively. Therefore, a change occurred in the rate of polluted days, but it varied according to standards. At the minimum, the rate of polluted days (θ = 15.99) was reduced 30.14% and the maximum reduction (θ = 3.56) to (λ = 1.63) was 54.21% in the case of Korean standards. Similarly, Table 13 refers to the reduction of polluted days (θ) to (λ) for PM_{10} after change point (k) which were 85.46, 90.34, and 85.98 according to European, American, and Korean standards, respectively. Moreover, the decrease in the rate of polluted days (θ = 14.73) to (λ = 9.20) was at least 37.54% for European standards, but it was 80.95% in the case of American standards. Figures 18 and 19 graphically represent the replications of monthly polluted days before and after the change point which are discussed in Tables 9 and 10.

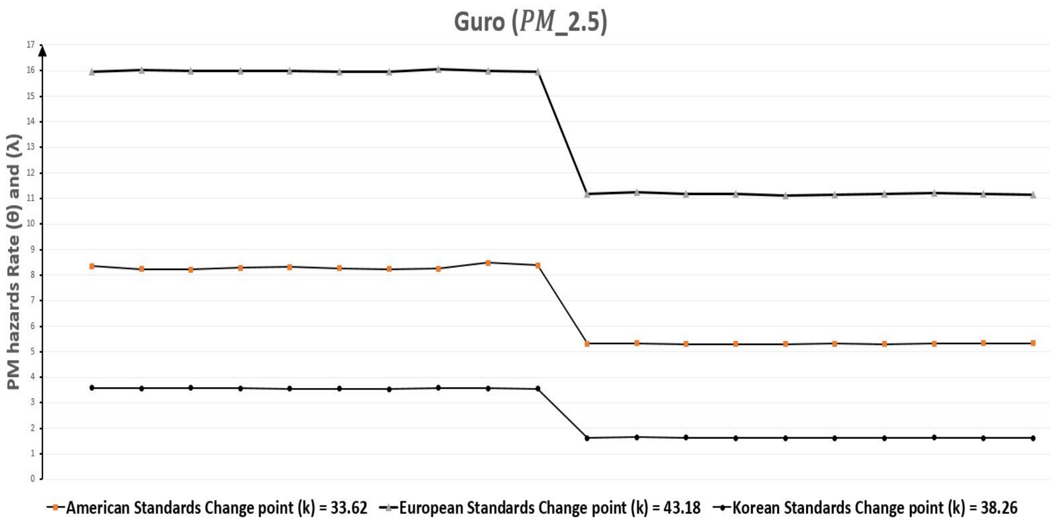

Figure 18. Guro (Seoul, South Korea) monthly polluted days before and after the change point due to $PM_{2.5}$.

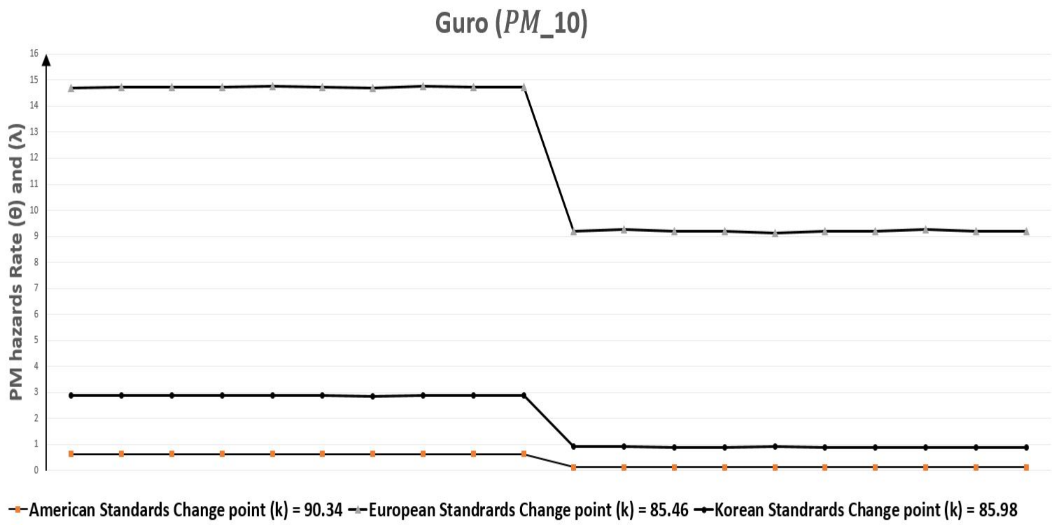

Figure 19. Guro (Seoul, South Korea) monthly polluted days before and after the change point due to PM_{10}.

6.1.2. CUSUM Approach

CUSUM Approach also indicates a reduction in hazards rate from (θ) to (λ) after change. As for Guro, Table 12 also represents the change of $PM_{2.5}$ polluted days through CUSUM approach, which shows that change point occurred between point 40 (k) and 41 ($k+1$) for European standards, between point 36 (k) and 37 ($k+1$) for American standards and it lies in-between point 35 (k) and 36 ($k+1$) for Korean standards. While Table 14 indicates the change of PM_{10} polluted days through CUSUM approach with an indication of change point lies between point 77 (k) and 78 ($k+1$) for European standards, point 89 (k) and 90 ($k+1$) for American standards and point 64 (k) and 65 ($k+1$) for Korean standards.

6.2. Nowon (Seoul, South Korea)

Nowon is located in the northeastern part of the city, and has the highest population density in Seoul with 619,509 persons living in 35.44 km^2, which is surrounded by mountains and forests on the northeast. The policies of the Ministry of Environment in Nowon have improved the rate of polluted days for $PM_{2.5}$ and PM_{10} hazards from θ to λ. Improvement in the reduction of polluted days varies case to case.

6.2.1. Bayesian Approach

Correspondingly, in case of Nowon, Table 11 depicts that change point (k) of polluted days for $PM_{2.5}$ were 49.95, 41.83, and 50.34 according to European, American, and Korean standards respectively. Particularly for this case, the change point was the same according to European and Korean standards, but varied for American standards. The rate of polluted days (θ = 15.88) for European standards showed a minimum decrease of 35.14% after the change point and approached (λ = 10.30), but the maximum decrease was for Korean standards which was 67.65% with (θ = 3.74) and (λ = 1.21). In the same manner, when we study Table 13, it elaborates that for PM_{10}, again there was a reduction in the rate of polluted days after change point (k) of 67.12, 73.80, and 65.68 to European, American, and Korean standards, respectively. That is comparable in cases of European and Korean standards, but a bit different for American standards. In addition, the reduction in the rate of polluted days for the European standard was at least 44.01% from (θ = 14.93) to (λ = 8.36), while the maximum reduction was (θ = 0.62) to (λ = 0.11) 82.25% for American standards. Figures 20 and 21 graphically represent the replications of monthly polluted days before and after the change point which are given in Tables 9 and 10.

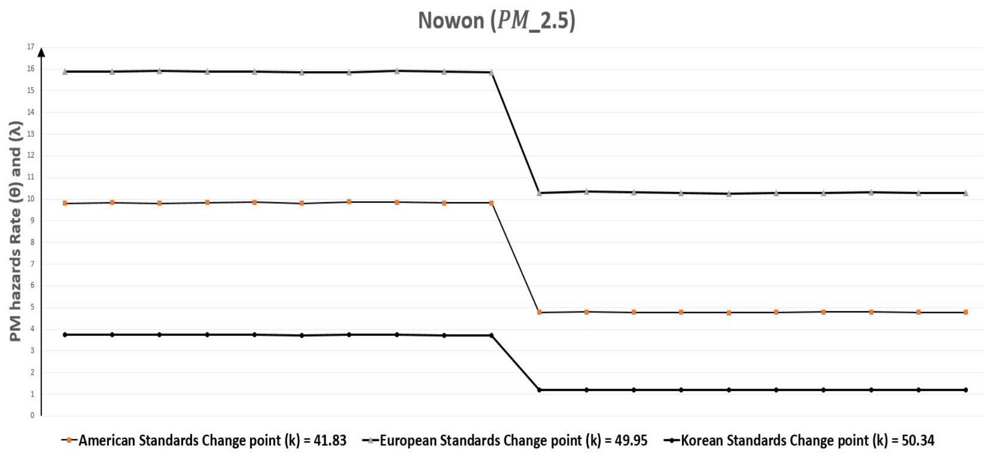

Figure 20. Nowon (Seoul, South Korea) monthly polluted days before and after the change point due to $PM_{2.5}$.

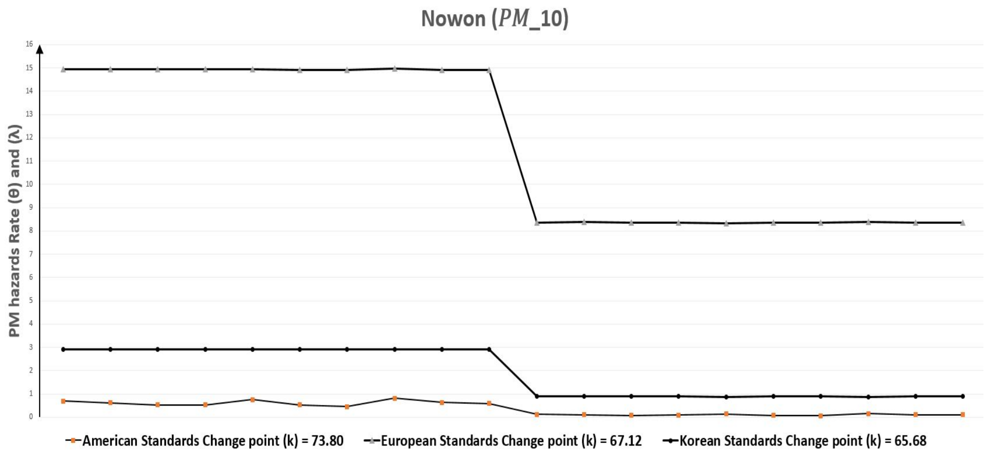

Figure 21. Nowon (Seoul, South Korea) monthly polluted days before and after the change point due to PM_{10}.

6.2.2. CUSUM Approach

Moreover, CUSUM Approach also validates the reduction of PM hazards. In case of Nowon, Table 12 represents the change of $PM_{2.5}$ polluted days through CUSUM approach, which shows that change point occurred between point 53 (k) and 54 ($k+1$) for European standards, between point 41 (k) and 42 ($k+1$) for American standards and it lies in-between point 50 (k) and 51 ($k+1$) for Korean standards. While Table 14 indicates the change of PM_{10} polluted days through CUSUM approach with an indication of change point lies between point 67 (k) and 68 ($k+1$) for European standards, point 63 (k) and 64 ($k+1$) for American standards and point 66 (k) and 67 ($k+1$) for Korean standards.

6.3. Songpa (Seoul, South Korea)

Songpa is located at the southeastern part of Seoul, and has largest population, with 647,000 residents. As per Ministry of Environment policies in Songpa, there is a smaller reduction for the rate of polluted days (θ) to (λ) as compared to Guro and Nowon, but still there is a significant reduction in PM hazards.

6.3.1. Bayesian Approach

Now for Songpa, we can check from Table 11 that change point (k) of polluted days for $PM_{2.5}$ were 53.85, 53.32, and 56.44 for European, American, and Korean standards respectively, which were all similar. The reduction in rate of polluted days was at least 24.50% for European Standards (θ = 14.90) to (λ = 11.25) while it was highest for Korean standards at 47.77% with (θ = 3.14) and (λ = 1.64). Correspondingly, we can also inspect the improvement in the rate of polluted days from (θ) to (λ) for PM_{10} after change point (k) in Table 13. Change point (k) for the rate of polluted days due to PM_{10} concentration were 53.50, 88.16, and 52.75 according to European, American and Korean standards respectively, which was the same for European and Korean standards. The slightest improvement 41.17% has been in the case of European standards and (θ = 16.08) is converted to (λ = 9.46). On the other hand, if we look at American standards, the rate of polluted days (θ = 0.56) was already low which further decreased 87.5% to (λ = 0.07). Hence, this area is almost a meeting of the PM_{10} concentration requirements for American standards but not for other standards. Figures 22 and 23 graphically represent the replications of monthly polluted days before and after the change point which are given in Tables 9 and 10.

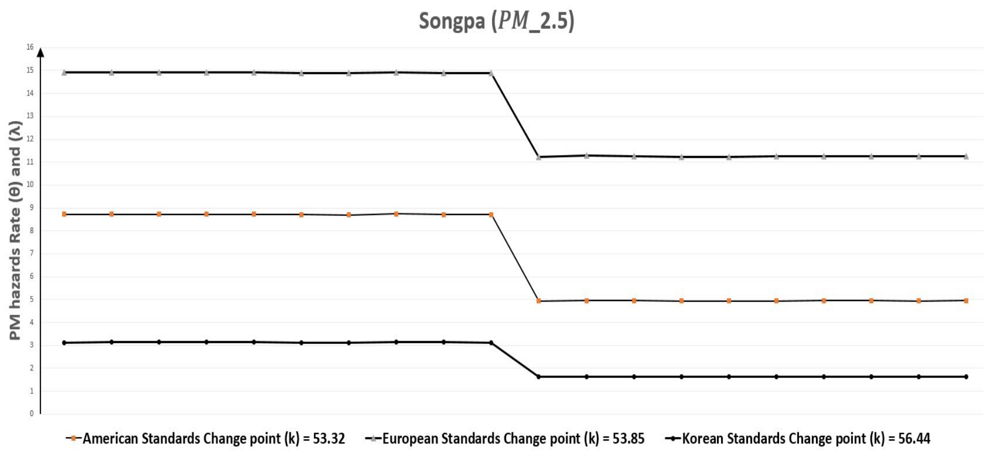

Figure 22. Songpa (Seoul, South Korea) monthly polluted days before and after the change point due to $PM_{2.5}$.

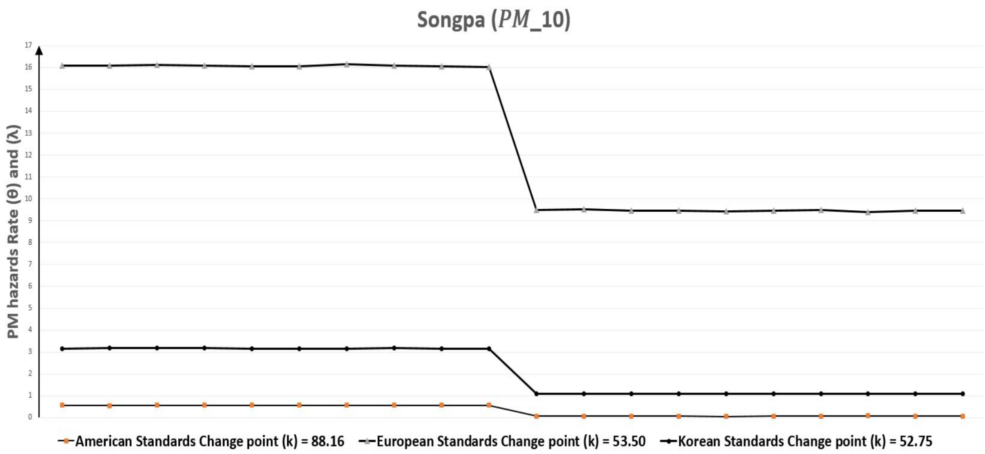

Figure 23. Songpa (Seoul, South Korea) monthly polluted days before and after the change point due to PM_{10}.

6.3.2. CUSUM Approach

As per CUSUM Approach, there is a decrease in PM hazards. Table 12 also represents the change of $PM_{2.5}$ polluted days through CUSUM approach, which shows that change point occurred between point 53 (k) and 54 ($k + 1$) for European standards, between point 52 (k) and 53 ($k + 1$) for American standards and it lies in-between point 52 (k) and 53 ($k + 1$) for Korean standards. While Table 14 indicates the change of PM_{10} polluted days through CUSUM approach with an indication of change point lies between point 53 (k) and 54 ($k + 1$) for European standards, point 53 (k) and 54 ($k + 1$) for American standards and point 52 (k) and 53 ($k + 1$) for Korean standards.

6.4. Yongsan (Seoul, South Korea)

Yongsan is a place in the center of Seoul in which almost 250,000 people reside. Prominent locations in Yongsan includes Yongsan station, electronic market and Itaewon commercial area with heavy traffic and transportation. Consequently, the policies of the Ministry of Environment in Yongsan has affected the particulate matter ($PM_{2.5}$ and PM_{10}) concentrations more than all the previous three locations (Guro, Nowon and Songpa). There is a remarkable decrease in rate of polluted days from (θ) to (λ).

6.4.1. Bayesian Approach

Similarly, in the case of Yongsan, Tables 11 and 13 tell us that the rate of polluted days (θ) for particulate matters ($PM_{2.5}$ and PM_{10}) was the highest in Seoul. The change occurred for $PM_{2.5}$ with the change point (k) 7.94, 6.18, and 5.78 with respect to European, American and Korean standards respectively, which was comparable for all the three standards. There was minimally a 50.71% fall in the rate of polluted days (θ = 24.55) to (λ = 12.10) for European standards, but the reduction in rate of polluted days was a maximum of 74.18% in the case of Korean standards. On the same note, Table 13 indicates that the change in the rate of polluted days has also occurred for PM_{10} concentrations. The change point (k) for it were 89.74, 91.36, and 68.06 for European, American and Korean standards, respectively. Furthermore, at least 42.10% rate of polluted days (θ = 14.56) was reduced to (λ = 8.43) for European standards but its maximum decrease was 93.93% for American standards (θ = 0.66) to (λ = 0.04), although, it is already approaching the requirements of this standard. Figures 24 and 25 graphically represent the replications of monthly polluted days before and after the change point which are given in Tables 9 and 10.

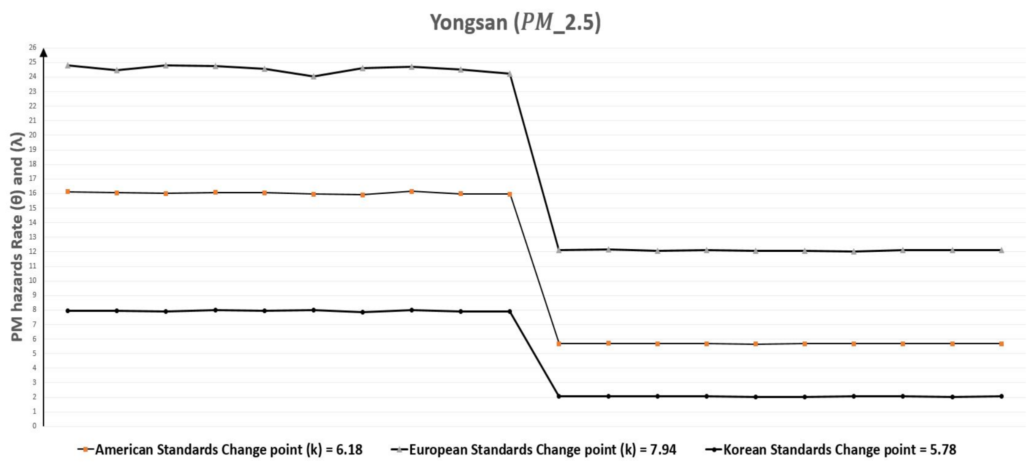

Figure 24. Yongsan (Seoul, South Korea) monthly polluted days before and after the change point due to $PM_{2.5}$.

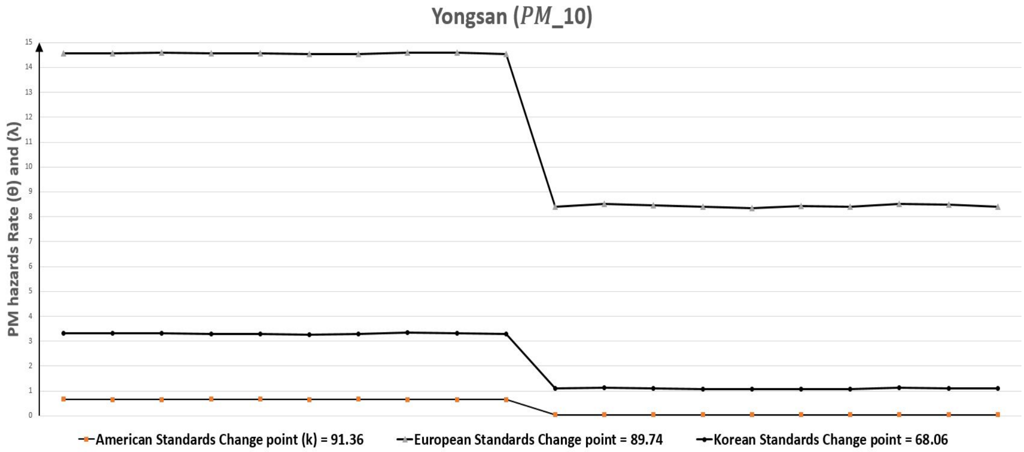

Figure 25. Yongsan (Seoul, South Korea) monthly polluted days before and after the change point due to PM_{10}.

6.4.2. CUSUM Approach

CUSUM Approach is directly applied on the raw data, which should be better for deterministic data structures. It also shows a reduction in PM hazards. In case of Yongsan, Table 12 also represents the change of $PM_{2.5}$ polluted days through CUSUM approach, which shows that change point occurred between point 11 (k) and 12 ($k+1$) for European standards, between point 64 (k) and 65 ($k+1$) for American standards and it lies in-between point 64 (k) and 65 ($k+1$) for Korean standards. While Table 14 indicates the change of PM_{10} polluted days through CUSUM approach with an indication of change point lies between point 77 (k) and 78 ($k+1$) for European standards, point 89 (k) and 90 ($k+1$) for American standards and point 64 (k) and 65 ($k+1$) for Korean standards.

6.5. Strengths

1. This approach is very precise, well defined, user friendly and easily understandable for applications on probability distributions, time series and random data.
2. The above mentioned model is an appropriate approach for detection of change points in random data structures.
3. Good technique for evaluation of process control programs by comparing the parameters before and after change point.

6.6. Limitations

1. Detection of only single change point is given in this model.
2. Further extension is required by making a model for a multiple number of change points for locating changed segments.

6.7. Managerial Insights

1. This model presents a suitable technique to analyze the air quality and pollutant hazards in the air.
2. By detecting change points in particulate matter ($PM_{2.5}$ and PM_{10}) concentrations and analyzing the occurrences of polluted days before and after a change point, environmental protection agencies can understand the role of their legislation efforts, and whether these change points are favorable or not for the environment.
3. A comparison of particulate matter hazards before and after a change point evaluates a pollution control program adopted by environmental protection agencies to make a decision. If these policies need further revision or not for the reduction of death rates and burden of diseases due airborne particulate matter concentrations in the air.

4. This study of pollutant hazards also defines the current levels of subjected air pollutant in the air which is helpful to make new pollution control policies for further improvements.
5. This research also brings an intuition to define new goals if previously defined goals have been achieved and also provides a vision if the environmental standards need to be revised, or not to overcome environmental challenges.

7. Conclusions

The main focus of this research work was to elucidate an appropriate change point detection model for occurrences of pollutant hazards due to higher concentrations of particulate matter ($PM_{2.5}$ and PM_{10}) in different locations. The rate of pollutant hazards before and after a change point was also estimated comprehensively to investigate the effectiveness of policies applied by the Ministry of Environment. To verify the model, four major locations (Guro, Nowon, Songpa, and Yongsan) in Seoul, South Korea were selected as study areas due to their different characteristics, such as climate zones, environment, populations and population densities. Three different environmental standards (European, American and Korean) were chosen as threshold values. Then, the model was applied to real time data sets in all cases and conclusions were drawn. The rate before and after the change point of particulate matter concentrations indicated a reduction in polluted days over a 10-year period. The overall results of our study confirm the effective role of legislation efforts used consistently to improve the air quality through the years but pollutant hazards still exist. Hence, further improvements are required to meet set standards to nullify hazards. This study can be further extended by making a multi-parameter change point model for a multiple number of change points considering the fact that different data structures follow different probability distributions.

Author Contributions: Conceptualization, Muhammad Rizwan Khan (M.R.K.) and Biswajit Sarkar (B.S.); methodology, M.R.K.; software, M.R.K.; validation, M.R.K. and B.S.; formal analysis, M.R.K.; investigation, M.R.K.; resources, M.R.K. and B.S.; data curation, M.R.K. and B.S.; writing—original draft preparation, M.R.K.; writing—review and editing, B.S.; visualization, M.R.K. and B.S.; supervision, B.S.; project administration, M.R.K. and B.S.; funding acquisition, B.S.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Héroux, M.E.; Anderson, H.R.; Atkinson, R.; Brunekreef, B.; Cohen, A.; Forastiere, F.; Hurley, F.; Katsouyanni, K.; Krewski, D.; Krzyzanowski, M.; et al. Quantifying the health impacts of ambient air pollutants: recommendations of a WHO/Europe project. *Int. J. Public Health* **2015**, *60*, 619–627. [CrossRef] [PubMed]
2. Pope, C.A.; Burnett, R.T.; Thurston, G.D.; Thun, M.J.; Calle, E.E.; Krewski, D.; Godleski, J.J. Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. *Circulation* **2004**, *109*, 71–77. [CrossRef] [PubMed]
3. Gyarmati-Szabó, J.; Bogachev, L.V.; Chen, H. Modelling threshold exceedances of air pollution concentrations via non-homogeneous Poisson process with multiple change-points. *Atmos. Environ.* **2011**, *45*, 5493–5503. [CrossRef]
4. Pirani, M.; Best, N.; Blangiardo, M.; Liverani, S.; Atkinson, R.W.; Fuller, G.W. Analysing the health effects of simultaneous exposure to physical and chemical properties of airborne particles. *Environ. Int.* **2015**, *79*, 56–64. [CrossRef] [PubMed]
5. Welty, L.J.; Peng, R.; Zeger, S.; Dominici, F. Bayesian distributed lag models: Estimating effects of particulate matter air pollution on daily mortality. *Biometrics* **2009**, *65*, 282–291. [CrossRef]
6. Qin, S.; Liu, F.; Wang, J.; Sun, B. Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmos. Environ.* **2014**, *98*, 665–675. [CrossRef]
7. Keshavarz, H.; Scott, C.; Nguyen, X. Optimal change point detection in Gaussian processes. *J. Stat. Plan. Inference* **2018**, *193*, 151–178. [CrossRef]

8. Kucharczyk, D.; Wyłomańska, A.; Sikora, G. Variance change point detection for fractional Brownian motion based on the likelihood ratio test. *Phys. Stat. Mech. Its Appl.* **2018**, *490*, 439–450. [CrossRef]
9. 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]
10. 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]
11. 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]
12. Hilgert, N.; Verdier, G.; Vila, J.P. Change detection for uncertain autoregressive dynamic models through nonparametric estimation. *Stat. Methodol.* **2016**, *33*, 96–113. [CrossRef]
13. Górecki, T.; Horváth, L.; Kokoszka, P. Change point detection in heteroscedastic time series. *Econom. Stat.* **2018**, *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. Shin, D.; Guchhait, R.; Sarkar, B.; Mittal, M. Controllable lead time, service level constraint, and transportation discounts in a continuous review inventory model. *RAIRO-Oper. Res.* **2016**, *50*, 921–934. [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. 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]
18. Sarkar, B.; Saren, S. Partial trade-credit policy of retailer with exponentially deteriorating items. *Int. J. Appl. Comput. Math.* **2015**, *1*, 343–368. [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. Keshavarz, M.; Huang, B. Bayesian and Expectation Maximization methods for multivariate change point detection. *Comput. Chem. Eng.* **2014**, *60*, 339–353. [CrossRef]
21. 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]
22. 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]
23. 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]
24. 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]
25. 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]
26. 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]
27. Bardwell, L.; Fearnhead, P. Bayesian detection of abnormal segments in multiple time series. *Bayesian Anal.* **2017**, *12*, 193–218. [CrossRef]
28. Lim, S.S.; Vos, T.; Flaxman, A.D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; AlMazroa, M.A.; Amann, M.; Anderson, H.R.; Andrews, K.G.; et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2224–2260. [CrossRef]
29. Rao, X.; Zhong, J.; Brook, R.D.; Rajagopalan, S. Effect of particulate matter air pollution on cardiovascular oxidative stress pathways. *Antioxid. Redox Signal.* **2018**, *28*, 797–818. [CrossRef]
30. Wei, T.; Meng, T. Biological Effects of Airborne Fine Particulate Matter (PM2.5) Exposure on Pulmonary Immune System. *Environ. Toxicol. Pharmacol.* **2018**, *60*, 195–201. [CrossRef]
31. Wang, Y.; Xiong, L.; Tang, M. Toxicity of inhaled particulate matter on the central nervous system: neuroinflammation, neuropsychological effects and neurodegenerative disease. *J. Appl. Toxicol.* **2017**, *37*, 644–667. [CrossRef]

32. Wellenius, G.A.; Burger, M.R.; Coull, B.A.; Schwartz, J.; Suh, H.H.; Koutrakis, P.; Schlaug, G.; Gold, D.R.; Mittleman, M.A. Ambient air pollution and the risk of acute ischemic stroke. *Arch. Intern. Med.* **2012**, *172*, 229–234. [CrossRef]
33. Li, P.; Xin, J.; Wang, Y.; Wang, S.; Shang, K.; Liu, Z.; Li, G.; Pan, X.; Wei, L.; Wang, M. Time-series analysis of mortality effects from airborne particulate matter size fractions in Beijing. *Atmos. Environ.* **2013**, *81*, 253–262. [CrossRef]
34. Kim, S.E.; Bell, M.L.; Hashizume, M.; Honda, Y.; Kan, H.; Kim, H. Associations between mortality and prolonged exposure to elevated particulate matter concentrations in East Asia. *Environ. Int.* **2018**, *110*, 88–94. [CrossRef]
35. Kim, S.E.; Honda, Y.; Hashizume, M.; Kan, H.; Lim, Y.H.; Lee, H.; Kim, C.T.; Yi, S.M.; Kim, H. Seasonal analysis of the short-term effects of air pollution on daily mortality in Northeast Asia. *Sci. Total Environ.* **2017**, *576*, 850–857. [CrossRef]
36. Qin, R.X.; Xiao, C.; Zhu, Y.; Li, J.; Yang, J.; Gu, S.; Xia, J.; Su, B.; Liu, Q.; Woodward, A. The interactive effects between high temperature and air pollution on mortality: A time-series analysis in Hefei, China. *Sci. Total Environ.* **2017**, *575*, 1530–1537. [CrossRef]
37. Lee, H.; Honda, Y.; Hashizume, M.; Guo, Y.L.; Wu, C.F.; Kan, H.; Jung, K.; Lim, Y.H.; Yi, S.; Kim, H. Short-term exposure to fine and coarse particles and mortality: A multicity time-series study in East Asia. *Environ. Pollut.* **2015**, *207*, 43–51. [CrossRef]
38. 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]