A Study on Determination of Particular Matter Concentration using Least Squares Approximation and Block Signal Processing

Hun Choi*
Dep. of Electronic Eng.
Dong-eui University
Busan, KOREA

Joosang Youn
Det. of Multimedia Eng.
Dong-eui University
Busan, KOREA

***Abstract*—Since atmospheric particles may adversely effects human life, the continuous and automatic monitoring of the particulate matter (PM) concentration is required. In this paper, we present a new method for determining mass concentration in the PM monitoring system using beta-ray absorption method. The linear least squares function approximation (LLSFA) and block signal processing (BSP) are used to reduce the noise in the measured data. By combining the LLSFA and the BPS, the proposed method can improve the determining performance of the particulate matter concentrations in the air pollutant monitoring system using the beta-ray absorption method. Simulation results with the computer generated data and the empirical data show the proposed method accurately determines the PM concentrations for real time.**

***Keywords*—PM Concentations; Beta-ray absorption method; Linear least squares function approximation; Block signal processing**

I. INTRODUCTION

The fast economic growth the last decades has resulted in an increase of the sources of particulate air pollution not only in large metropolitan areas but also in medium-sized urban areas. Atmospheric particles originate from a variety of sources and possess a range of physical and chemical properties. Especially, fine particulates less than 10 and 2.5 μm size are referred to as PM_{10} and $PM_{2.5}$, respectively. These particular matters (PM) have the most significant impact on human health because they can penetrate deep into the lung[1]. Over the last decades a number of studies conducted in many countries around the world observed associations between ambient particle concentrations (especially PM_{10} and $PM_{2.5}$) and human health risk[2,3]. To monitor the PM concentrations in open air, on-line method collecting data and calculating mass concentration with the collected data for real time or near real time is required. It is well-known that beta-ray absorption method (BAM) is a good solution for determining the PM concentration[4]. In this method, a known amount of electron scattering and attenuation through a clean filter is compared with that of a dust-sampled filter. To determine mass concentration, a ratio of the number of detected beta particles passing through the clean filter and the dust-sampled filter is used. Generally, these data measured in air pollution monitoring such as PM_{10} and $PM_{2.5}$ usually contained the additive noise (thermal noise, electric noise in the beta-ray detector, power supply noise, and etc.). This additive noise can

This work was supported by Dong-eui University Grant(2015AA133).
Hun Choi : Corresponding Author*

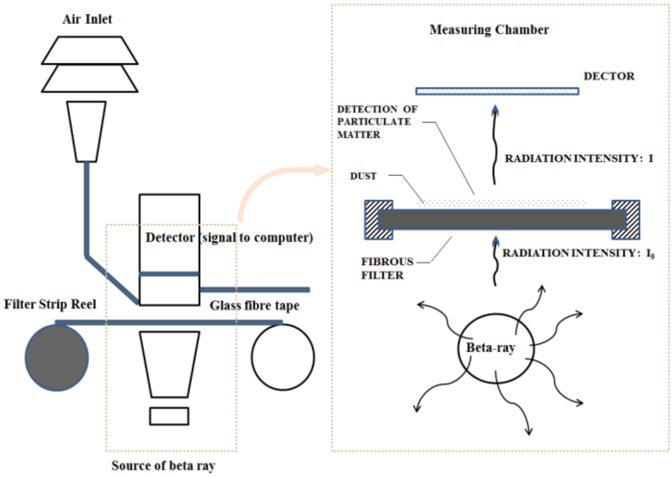

Fig. 1. Schematic diagram of the PM monitoring system using the BAM.

cause unintended bias in calculating the PM concentrations. Therefore, to improve the reliabilities of the calculated concentrations, it is important to find the best way to reduce the additive noise[5].

This paper presents a new method for determining the mass concentration in the PM monitoring system using the BAM. To reduce the additive noise in the measured data, the proposed method uses the linear least squares function approximation (LLSFA)[6,7] and the block signal processing (BSP)[8]. The time varying mass concentration is calculated with the denoised data. In addition, by combining the merits of the LLSFA and the BSP, the proposed BSP-LLSFA method gives both the excellent estimation of the mass concentration without processing delay and the reduced computational complexity. To evaluate the performance of the proposed method, computer simulations were performed with computer generated signals and a real measured signal as the input.

II. METHOLOGY

A. Beta-ray abruption

Fig. 1 shows the configuration of the PM monitoring system which uses BAM. Air is drawn into the inlet and deflected downwards into the acceleration jet of the impact unit. Because of their greater momentum, particles larger than 10 μm cut point impact out and are retained in the middle plenum impaction chamber. Particles smaller than 10 μm are carried upward by the air flow and down the vent tubes to the beta gages sampler. After traversing the inlet configuration, the PM_{10} and $PM_{2.5}$ particles are deposited on a glass fibre filter tape. A low level of beta-rays emitted from the source and passes through the filter tape and deposited particles. The increase of particles collected in the tape causes a lower beta-ray measurement in the measuring chamber. Applying Beer-Lambert Law for these signals, we can get the concentrations of the PM (PM_{10} and $PM_{2.5}$) as follows.

$$M_C = \frac{BS}{\mu Q \Delta t} \ln\left(C_0 / C\right) \tag{1}$$

where M_C is mass concentration [$\mu g / m^3$], S is the area of deposited sample (PM_{10} or $PM_{2.5}$) [cm^2], Q is a rate of flow [*l/min*], Δt is a time of deposition [min], C_0 is the initial beta-ray intensity [*count*], C is the beta-ray intensity after it passes through the sample [*count*], B is a coefficient for unit conversion.

B. Linear Least Squares Function Approximation

Noise is constantly present in many applications with the measurement systems. Even though all control parameters remain constant, the resultant outcomes vary. Therefore, a process of quantitatively estimating the trend of the outcomes, also known as curve fitting, becomes necessary. The curve fitting process fits equations of approximated curves to the raw observed data. Nevertheless, for a given set of data, the fitting curves of a given type are generally not unique. Thus, a curve with a minimal deviation (error) from all data points is desired. This best fitting can be obtained by the method of least squares. The method of least squares seeks to minimize the sum of the squares of the squares of the deviation between the approximating function value and the observed data value. Assume that the given N data set are (x_1, y_1), (x_2, y_2), …, (x_N, y_N), where x is the independent variable(time) and y is the dependent variable(observed data). Let $f(x)$ denote an unknown linear function (straight line) that approximates the observed data y. The best fit straight line to approximate the observed data in linear least squares sense is given by

$$f(x) = ax + b \tag{2}$$

$$a = \left(\frac{NS_{xy} - S_x S_y}{NS_{xx} - S_x S_x} \right), \quad b = \left(\frac{S_{xx} S_y - S_{xy} S_x}{NS_{xx} - S_x S_x} \right) \tag{3}$$

$$S_{xx} = \sum_{i=1}^{N} x_i^{\,2}, \quad S_x = \sum_{i=1}^{N} x_i, \quad S_{xy} = \sum_{i=1}^{N} x_i y_i, \quad S_y = \sum_{i=1}^{N} y_i \tag{4}$$

where the subscript i means the time index of each variable (time and the corresponding observed data) in the given N data set.

III. THE PROPOSED METHOD

To analyze the PM concentration, the observed beta-ray intensity (C_0 and C) that used in (1) are modeled by

Observed time index (n)	⋯	$-L+1$ ⋯ 0	1 ⋯ L	$L+1$ ⋯ $2L$	⋯
Observed beta-ray intensity data ($\mathbf{C}_k$)	⋯	$C(L(k-2)+1)$ ⋯ $C(L(k-1))$	$C(L(k-1)+1)$ ⋯ $C(Lk)$	$C(Lk+1)$ ⋯ $C(L(k+1))$	⋯
Data segment with L length	⋯	(k-1)-th Data segment	k-th Data segment	(k+1)-th Data segment	⋯
Data segment index (k)	⋯	$k-1$	k	$k+1$	⋯
Data segment time index (Lk)	⋯	$L(k-1)$	Lk	$L(k+1)$	⋯

Fig. 2. The intensity data segment, data segment index, data segment time index corresponding to the observed time index

$$C(n) = I(n) + v(n) \tag{5}$$

Where n is time index $I(n)$ is the pure beta-ray intensity without noise, $u(n)$ is the colored noise, and $v(n)$ is the additive noise.

Assumption A1: The additive noise $v(n)$ is white Gaussian random process with zero mean and variance σ_v^2. Hence, the beta-ray intensity $I(n)$ and the additive noise $v(n)$ are statistically independent.

Assumption A2: The variation of beta-ray intensity is smaller than that of noise ($\sigma_I^2 << \sigma_v^2$).

Consider data segment that consists of the last L observed noisy beta-ray intensity, the k-th intensity data segment ($\mathbf{C}_k$) is represented by

$$\mathbf{C}_k = [C(L(k-1)+1) \quad C(L(k-1)+2) \quad \cdots \quad C(Lk)]^T \tag{6}$$

where k is an integer as the index of intensity data segment, it is related to the observed time index n such as $n = kL$.

Fig. 2 shows the intensity data segment ($\mathbf{C}_k$), data segment index (k), and data segment time index (Lk) corresponding to the observed time index (n). The mean value of k-th intensity data segment is expressed as

$$\begin{aligned} \hat{C}_k &= \frac{1}{L} \sum_{i=1}^{L} C_k(i) \\ &= \frac{1}{L} \sum_{i=1}^{L} I_k(i) + \frac{1}{L} \sum_{i=1}^{L} v_k(i) \end{aligned} \tag{7}$$

where $I_k(i)$ and $v_k(i)$ are the i-th pure beta-ray intensity data and the corresponding additive noise in the k-th intensity data segment, respectively.

Under assumption A1, the second term on the right hand side (r.h.s.) in (7) can be assumed to be zero when L is large. Therefore, we can obtain a mean value of the k-th intensity data segment as

$$\hat{C}_k \approx \frac{1}{L} \sum_{i=1}^{L} I_k(i) \tag{8}$$

Using the result of (8), the mass concentration $M_{C,k}$ in k-th data segment time can be represented as

Observed time index (n)	$L(k-M)+1 \cdots L(k-M+1)$	$L(k-M+1)+1 \cdots L(k-M+2)$	$\cdots$	$1 \quad \cdots \quad L$
Data segment with L length	(k-M+1)-th Data segment	(k-M+2)-th Data segment	$\cdots$	k-th Data segment
Data segment index (k)	$k-M+1$	$k-M+2$	$\cdots$	k
Data segment time index vector ($\mathbf{t}_k$)	$[L(k-M+1)$ $=[t_{k,1}$	$L(k-M+2)$ $L(k-M+2)$	$\cdots$ $\cdots$	$Lk]^T$ $t_{k,M}]^T$
Mean value vector of intensity data segment ($\mathbf{y}_k$)	$[\hat{C}_{(k-M+1)}$ $=[y_{k,1}$	$\hat{C}_{(k-M+2)}$ $y_{k,2}$	$\cdots$ $\cdots$	$\hat{C}_k]^T$ $y_{k,M}]^T$

Fig. 3. The new notations for BPS-LLSFA

$$\hat{M}_{C,k} = \frac{BS}{\mu Q \Delta \hat{t}} \ln\left(\hat{C}_{k-1} / \hat{C}_k\right) \tag{9}$$

where $\Delta\hat{t}$ is the same as the length of data segment, i.e. $\Delta\hat{t} = L$. The mass concentration $\hat{M}_{C,k}$ in (9) is more than zero, because all the parameters are positive and $\ln\left(\hat{C}_{k-1}/\hat{C}_k\right) \geq 0$ for large $\Delta\hat{t}$. Under assumption A2, however, the second term on the r.h.s. of (7) can not be neglected when a time of deposition $\Delta\hat{t}$ is not sufficiently large. Therefore, $\ln\left(\hat{C}_{k-1}/\hat{C}_k\right) \geq 0$ is not valid any more. Generally, digital filtering technique gives a good solution for reducing noise in many applications with noisy data). But, it can lead to a serious processing delay (filtering delay increasing with the length of filter) in the application with the beta-ray signals obtained by low sampling rate ($\leq$1Hz). The LLSFA method is an alternative solution to overcome this problem. The LLSFA method using the BSP can reduce computational complexity and adversely effect of high frequency noise. The PM monitoring system requires the huge processing power due to their large number of the measured signal. Moreover, it suffers from high frequency noise. The block signal processing based on the linear least squares function approximation (BSP-LLSFA) method may be a good solution to overcome these problems. To apply the BSP-LLSFA method to the estimation of the PM concentration, we use new notations for independent (time index) and dependent (observed data) variables of (2) as shown in Fig. 3. From Fig. 3, the data segment time index vector and mean value vector corresponding to the last M data segments are respectively expressed as

$$\begin{aligned}\mathbf{t}_k &= [L(k-M+1) \quad L(k-M+2) \quad \cdots \quad Lk]^T \\ &= [t_{k,1} \quad t_{k,2} \quad \cdots \quad t_{k,M}]^T\end{aligned} \tag{10}$$

$$\begin{aligned}\mathbf{y}_k &= [\hat{C}_{(k-M+1)} \quad \hat{C}_{(k-M+2)} \quad \ldots \quad \hat{C}_k]^T \\ &= [y_{k,1} \quad y_{k,2} \quad \cdots \quad y_{k,M}]^T\end{aligned} \tag{11}$$

where $\mathbf{t}_k$ and $\mathbf{y}_k$ are column vectors with M elements. Using (10) and (11), the cost function for determining the best fit straight line to approximate the M data segments can be expressed as

$$J_k = \sum_{i=1}^{M} [\hat{f}(t_{k,i}) - y_{k,i}]^2 \tag{12}$$

$$\hat{f}(t_{k,i}) = \hat{a} t_{k,i} + \hat{b} \tag{13}$$

where $\hat{a}$ and $\hat{b}$ are the x-slop and y-intercept coefficients of the best fit straight line that minimizes the cost function of (12), respectively. From (13) , (12) can be rewritten as

$$J_k(\hat{a}, \hat{b}) = \sum_{i=1}^{M} (\hat{a} t_{k,i} + \hat{b} - y_{k,i})^2 \tag{14}$$

In (14), the cost function is quadratic and convex. Therefore, the cost function has a global minimum solution. From (14), we can get the partial derivatives of the cost function with respect to $\hat{a}$ and $\hat{b}$, and set the results to zeroes as

$$\frac{\partial J_k(\hat{a}, \hat{b})}{\partial \hat{a}} = 2\sum_{i=1}^{M} (\hat{a} t_{k,i} + \hat{b} - y_{k,i}) t_{k,i} = 0 \tag{15}$$

$$\frac{\partial J_k(\hat{a}, \hat{b})}{\partial \hat{b}} = 2\sum_{i=1}^{M} (\hat{a} t_{k,i} + \hat{b} - y_{k,i}) = 0 \tag{16}$$

From (15) and (16), we can get as follows

$$\hat{a} \sum_{i=1}^{M} t_{k,i}^2 + \hat{b} \sum_{i=1}^{M} t_{k,i} = \sum_{i=1}^{M} t_{k,i} y_{k,i} \tag{17}$$

$$\hat{a} \sum_{i=1}^{M} t_{k,i} + \sum_{i=1}^{M} \hat{b} = \sum_{i=1}^{M} y_{k,i} \tag{18}$$

The matrix expression of (17) and (18) can be rewritten as

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{M} t_{k,i}^2 & \sum_{i=1}^{M} t_{k,i} \\ \sum_{i=1}^{M} t_{k,i} & M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{M} t_{k,i} y_{k,i} \\ \sum_{i=1}^{M} y_{k,i} \end{bmatrix} \tag{19}$$

Solving for $\hat{a}$ and $\hat{b}$, we can obtain as follows

$$\hat{a} = \left(\frac{M\hat{S}_{ty} - \hat{S}_t \hat{S}_y}{M\hat{S}_{tt} - \hat{S}_t \hat{S}_t} \right), \quad \hat{b} = \left(\frac{\hat{S}_{tt} \hat{S}_y - \hat{S}_{ty} \hat{S}_t}{M\hat{S}_{tt} - \hat{S}_t \hat{S}_t} \right) \tag{20}$$

$$\hat{S}_{tt} = \sum_{i=1}^{M} t_{k,i}^2, \quad \hat{S}_t = \sum_{i=1}^{M} t_{k,i}, \quad \hat{S}_{ty} = \sum_{i=1}^{M} t_{k,i} y_{k,i}, \quad \hat{S}_y = \sum_{i=1}^{M} \hat{y}_{k,i} \tag{21}$$

From above results, the PM concentration for the M data segments at the data segment time index Lk can be expressed as

$$\tilde{M}_{C,k} = \frac{KS}{\mu Q \Delta\tilde{t}} \ln\left(\tilde{C}_{k,0} / \tilde{C}_k\right) \tag{22}$$

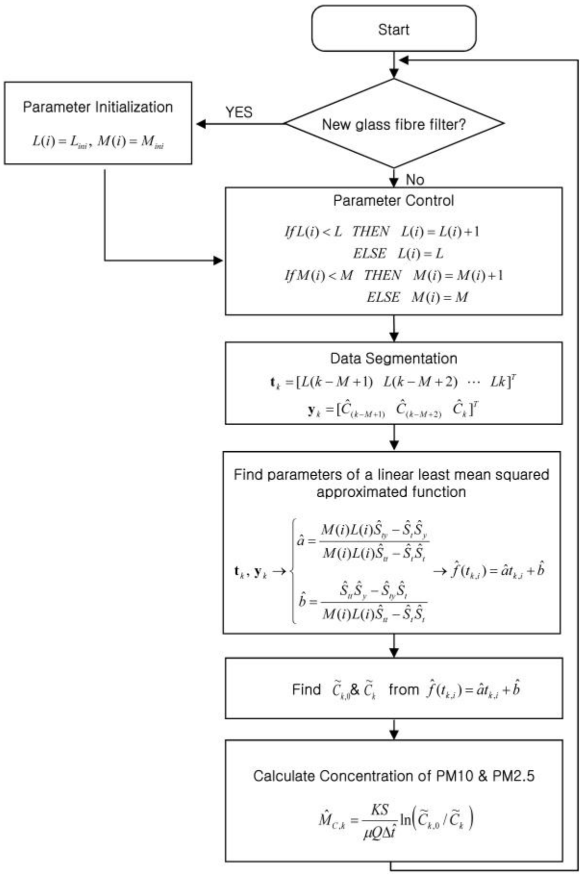

Fig. 4. Procedure of the proposed BSP-LLSFA method.

$$\tilde{C}_{k,0} = \hat{a}_k [L(k - M + 1)] + \hat{b}_k \tag{23}$$

$$\tilde{C}_k = \hat{a}_k (Lk) + \hat{b}_k \tag{24}$$

In (22), $\tilde{C}_{k,0}$ and $\tilde{C}_k$ are calculated using (13), and $\Delta\tilde{t}$ is the length of the observed time index corresponding to the data segment time index vector ($\mathbf{t}_k$), i.e., $\Delta\tilde{t} = ML$. The detailed procedure of the proposed method is shown in Fig. 4.

IV. SIMULATIONS

To evaluate the performance of the proposed BSP-LLSFA method, we carried out computer simulations with computer generated data and a real measured data as the input. We considered two types of the pure beta-ray intensity data $I(n)$ of (5) to generate the input data for the practical applications.

Type 1: Linearly decreasing data

$$\begin{aligned} C(n) &= I(n) + v(n) \\ &= -0.015n + 1750 + v(n) \quad for\ \{x \mid 0 \le n \le 40000\} \end{aligned} \tag{25}$$

Type 2: Exponentially decreasing data

$$\begin{aligned} C(n) &= I(n) + v(n) \\ &= re^{\sigma n} + v(n) \quad for\ \{x \mid 0 \le n \le 40000\} \end{aligned} \tag{26}$$

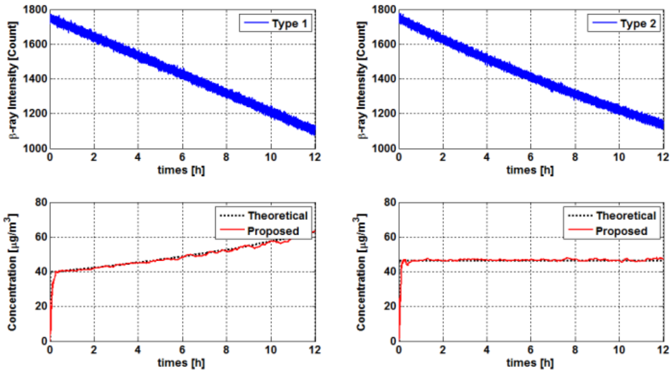

Fig. 5. The performances of the proposed method for different types of input data (The used parameters: L = 10, M = 360, L_{ini} = 1, and M_{ini} = 18 (5% of M)).

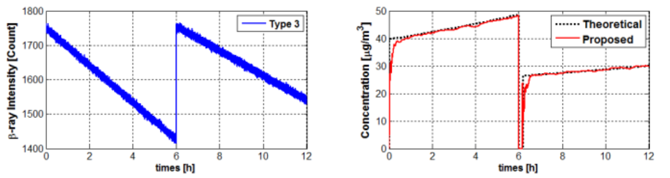

Fig. 6. The performance of the proposed method for type 3 input data (The used parameters: L = 10, M = 360, L_{ini} = 1, and M_{ini} = 18 (5% of M)).

where the observation noise $v(n)$ is the white Gaussian random process with zero mean, it is added to the pure beta-ray intensity signal such that $SNR = 25dB$. In (26), the data parameters were set to $r = 1750$, $\sigma = -1\times10^{-5}$. The sampling ratio of the generated data $C(n)$ was 1Hz. The used real data is an actual data that was measured by Thermo Andersen continuous ambient particulate monitoring system (FH62C14). To calculate the mass concentration shown in (22), the used parameters for the proposed method were set to $KS/\mu Q = 1.68\times10^6$, L = 10, and M = 360. Fig. 5 shows the theoretical and simulated mass concentration curves for the type 1 and type 2 input data. From the results of Fig. 5, the proposed method provides the best estimations of the theoretical mass concentrations for two types of the input data. The match between theory and simulation is excellent during steady-state and good during the initial transient, where the over or under damped curve can be reduced by controlling initial parameters (L_{ini} and M_{ini} which shown in Fig. 4).

Considering that a new filter is fed from a filter tape supply reel, we performed a simulation with a type 3 input data given by (27).

Type 3: Combined linearly decreasing data considering the filter tape change

$$I(n) = \begin{cases} -0.015n + 1750 & for \quad 0 \le n \le 21599 \\ 1750 & for \quad 21600 \le n \le 22199 \\ -0.010n + 1972 & for \quad 22200 \le n \le 43200 \end{cases} \tag{27}$$

Fig. 6 shows the estimation performance of the proposed method for type 3 input. Assume that the filter changes at 6 hour and the initialization procedure of measurement allows 10 minutes for the filter to zero. During this time, where the mass concentration is zero, the initial beta-ray intensity is measuring

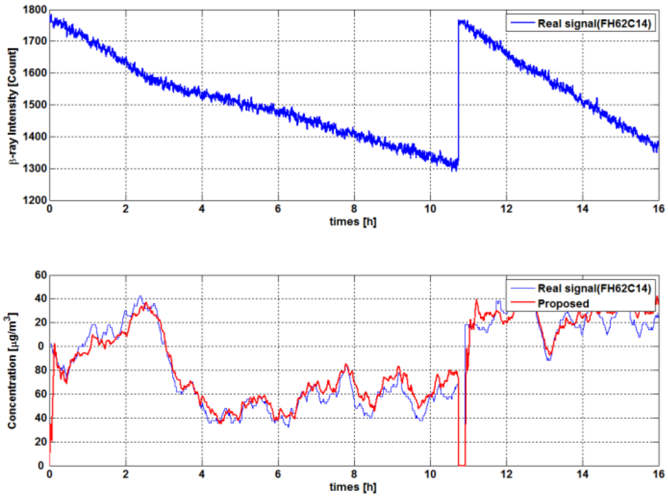

Fig. 7. Performance comparison of reference (FH62C14) and the proposed method (The used parameters: L = 10, M = 360, L_{ini} = 1, and M_{ini} = 36 (10% of M)).

and it is set to $\tilde{C}_0$. From the result shown in Fig. 6, we can find that the proposed method provides values of the theoretical mass concentration over 90% in approximately 8 minutes. The performance of the proposed method for the real input is shown in Fig. 7. Just like the results for the computer generated input, the performance of the proposed method is excellent.

V. CONCLUSIONS

This paper presents a new mass concentration estimation algorithm for the particulate matter (PM_{10} and $PM_{2.5}$) monitoring system using the beta-ray absorption method. By combining the merits of the linear least squares function approximation (LLSFA) and the block signal processing (BSP), the proposed BSP-LLSFA method gives both excellent estimation of the mass concentration without processing delay and the method gives reduced computational complexity. The initial transient estimation performance of the proposed method can be improved by controlling the parameters used in BSP. Several simulation results have shown the good estimation performance of the proposed method.

ACKNOWLEDGMENT

This work was supported by Dong-eui University Grant (2015AA133).

VI. CONCLUSIONS

This paper presents a new mass concentration estimation algorithm for the particulate matter (PM_{10} and $PM_{2.5}$) monitoring system using the beta-ray absorption method. By combining the merits of the linear least squares function approximation (LLSFA) and the block signal processing (BSP), the proposed BSP-LLSFA method gives both excellent estimation of the mass concentration without processing delay and the method gives reduced computational complexity. The initial transient estimation performance of the proposed method can be improved by controlling the parameters used in BSP. Several simulation results have shown the good estimation performance of the proposed method.

REFERENCES

[1] J.T. van der Wal and L.J.M. Janssen, "Analysis of spatial and temporal variatons of PM10 concentrations in the Netherlands using Kalman filtering," Atmospheric Environment, vol. 34, 2000, pp. 3675-3687.

[2] D. W. Dockery, J. Schwartz, and J. D. Spengler, "Air pollution and daily mortality: associations with particulates ad acid aerosols," Environmental research vol. 59(2), 1992, pp.362-373.

[3] K. Katsouyanni, F. Touloumi, C. Spix, J. Schwartz, F. Balduci, S. Medina, G. Rossi, B. Wojtyniak, J. Sunyer, L. Bacharova, J. P. Schouten, A. Ponka and H. R. Anderson, "Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project," BMJ, 314, 1997, pp.1658-1663.

[4] A. Chung, D. P. Y. Chabg, M. J. Kleeman, K. D. Perry, T. A. Cahill, D. Dutcher, E. M. McDougall and K. Stroud, "Comparison of real-time instruments used to monitor airborne particulate matter," Journ. of the Air & Waste Management Association vol. 51(1), 2001, pp. 109-120.

[5] K. I. Hoi, K. V. Yuen, and K. M. Mok, ""Optimizing the performance of Kalman filter based statistical time-varying air quality models," Global NEST Journal, vol. 12, 2010, pp. 27-39.

[6] A. Bjorck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.

[7] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing 3rd Ed., Prentice-Hall, 2009.