SPECTRAL FILTERING OF POSITIVELY DETERMINED FLUCTUATING SIGNALS

S. I. Eremenko

Institute of Chemical Kinetics and Combustion,
Siberian Branch of the Russian Academy of Sciences, Novosibirsk
Received January 15, 1997

A method of positive spectral filtering (PSF) is suggested as a mathematical approach to an analysis of nonnegative random signals, which allows one to select the components with different characteristic fluctuation times. The method is efficient for highly fluctuating signals, including concentration fluctuations of pollutants in the atmosphere. An example is given of the separation of contributions from near and far sources of pollution in the urban atmosphere.

1. INTRODUCTION

An analysis of gas and aerosol pollutants in the atmosphere is a special case of a wide class of problems, where an investigated signal is a nonnegative random function of time (often multidimensional) and is formed by a superposition of several nonnegative components from different sources. One of the analysis problem is to select these components.

In case of multidimensional signals, statistical methods are used based on an analysis of covariance of multidimensional signals, such as the factor analysis or the method of principal components. However, these methods are not adequate for the problem, because they do not consider the positive determinacy of the signal and therefore are correct only when the fluctuations are small in comparison with the mean signal value, whereas for many problems of this class, in particular, for the atmosphere, the opposite is true.

In the last few years, attempts have been undertaken to construct mathematical methods adequate for these problems. It is worth mention Paatero[1,2] whose works are devoted to the factorization of positively determined multidimensional signals. His approach considers correctly a superposition of nonnegative components and is based on an analysis of an entire temporal series of a multidimensional signal instead of the covariance matrix. The components are selected by solving the extremum problem with some restrictions. However, in these papers the information contained in the temporal behavior of the signal is not used, that is, the signal is considered as a disordered sample of a random variable. It is clear that the signal measured as a function of time in a given point contains more information than a simple disordered sample. For example, signals from different sources may have different fluctuation spectra.

In our particular case, the concentration of a pollutant in a local point is examined. It represents the random function of time, and the character of turbulent mixing in the atmosphere is such that the scale of concentration pulsations is usually comparable with the average concentration. In many cases, it is several times higher.

At present, the theory of turbulent transport of pollutants cannot unambiguously tackle the question of how the concentration fluctuations depend on the distance from the source or the transit time to the detector; however, the totality of theoretical and experimental data reported in Refs. 3–5 permits one to formulate the following quantitative regularities:

1. Pollutant concentration from local sources usually has a character of pulsations vanishing with time (the intermittence effect).

2. Relative amplitude of concentration fluctuations does not decrease or decreases very slowly with the transit time (distance), although its average value decreases fast.

3. Autocorrelation time of fluctuations increases with the transit time (distance), that is, the response “spreads” with time and its spectrum shifts to lower frequencies.

The spectral analysis can be used to select the signal components of different nature, for example, to select the global background and the contribution from local sources. However, standard methods like, for example, the Fourier transform, are inefficient in our case, because they do not consider the constant sign of the signal. ”elow an example is given which illustrates this statement.

In the present paper, an approach is suggested that allows us to select the components of positively determined random signals with different characteristic fluctuation times, namely, the method of positive spectral filtration (PSF). The method is based on *a priori* information about signals. It considers in the explicit form the fundamental property of sign constancy. This approach is efficient when the fluctuation amplitude is comparable with the average value of the signal.

0235-6880/97/08 604-05 $02.00

2. NONNEGATIVE SPECTRAL TRANSFORM IN THE GENERAL CASE

In general, we can consider the problem as follows. Let $y(t) \geq 0$ be the measured signal and normalized nonnegative kernel $g(\tau, \omega)$ of the linear convolution operator

$$\mathbf{G}_\omega \psi(t) = \int_{-\infty}^{\infty} \psi(t-\tau)\, g(\tau,\omega)\, d\tau \quad (1)$$

be specified from *a priori* considerations.

It is assumed that the signal $y(t)$ can be represented as integral convolution transform of function $\psi(t, \omega)$ tentatively named "spectral"

$$y(t) = \int_0^{\infty} \mathbf{G}_\omega \psi(t,\omega)\, d\omega \ . \quad (2)$$

"y analogy with the Fourier transform, the parameter ω is named the frequency. Evidently, the inverse problem, namely, determining ψ from $y(t)$ cannot have the unique solution.

3. KERNEL OF SPECTRAL TRANSFORM

A set of complete and orthogonal (or at least linearly independent) functions is commonly used as a kernel for the signal transform. The first property is required to transform an arbitrary signal. The second property is necessary to obtain the unique transform. In this paper, we suggest a set of positively determined functions chosen on the basis of *a priori* information about the signal nature. This set is complete in the examined class of signals to a measure to which *a priori* assumptions are valid. As to the linear independence, it is not necessary if the uniqueness is achieved by any other means. A set of possible signal transforms can be ordered by certain preferred criteria and then the choice of the unique transform can be made by solving the extremum problem on this set.

Thus, the shape of the kernel $g(t, \omega)$ can be chosen based on the model of the examined process. If, for example, this process is isotropic diffusion, the set of the Gauss functions is naturally taken as $g(t, \omega)$ and their reciprocal width as the parameter ω. In any case, we assume that the kernel functions have the following properties:

$$g(t, \omega) \geq 0, \quad g(t, \omega) \to 0 \ \text{ for } \ t \to \pm\infty ,$$
$$|g(t,\omega)| \equiv \int_{-\infty}^{\infty} g(t,\omega) dt = 1 . \quad (3)$$

The symbol | | here and further denotes the integral measure of the function. Further for arbitrary ω_1 and ω_2, if $\omega_1 > \omega_2$, there exists $\varphi(t) \geq 0$ such that

$$g(t,\omega_2) = \mathbf{G}_{\omega_1} \varphi(t) , \quad (4)$$

that is, a wider function can be represented as a convolution of narrow functions. The last property is characteristic of the transfer functions and, in particular, the Gauss function mentioned above.

4. PROBLEM FORMATION

Let the signal be represented as the convolution with the known kernel G and unknown spectral function $\psi(t, \omega) \geq 0$. We want to use this transform to decompose, if possible, the initial signal into the components with different ω. The simplest problem of this type is the selection of low frequencies, that is, the selection of the low-frequency component from $y(t)$, which can be represented by the spectrum with the frequencies no higher than the given frequency Ω

$$y(t) = \int_0^{\Omega} \mathbf{G}_\omega \psi(t,\omega)\, d\omega + (t), \quad f'(t) \geq 0 . \quad (5)$$

If this is possible for $|\psi(t, \omega)| > 0$, we speak that $y(t)$ comprises the frequencies lower than and equal to Ω. If the residual function $f'(t)$ does not comprise such frequencies, we speak that Eq. (5) is a solution to the problem of low-frequency filtration of the signal $y(t)$.

It is immediately clear that Eq. (5) cannot have the unique solution, at least by virtue of property (4); however, this property allows one to write this problem in another form. Really, by virtue of property (4) for any $\psi(t, \omega)$ and Ω, $\varphi(t) \geq 0$ can be found such that

$$\int_0^{\Omega} \mathbf{G}_\omega \psi(t,\omega)\, d\omega = \mathbf{G}_\Omega \varphi(t) , \quad (6)$$

that is, all the components with frequencies lower than Ω can be reduced to the frequency Ω. This allows one to write the problem of low-frequency filtration in the form

$$y(t) = \mathbf{G}_\Omega\, \varphi(t) + f'(t) \equiv f(t,\Omega) + f'(t) ,$$
$$f(t,\Omega) \geq 0, \quad f'(t) \geq 0. \quad (7)$$

We name $f(t, \Omega)$ the integral low-frequency (IL) PSF component of the signal for the cutoff frequency Ω. The ideal solution to the problem would be constructing the algorithm that can be used to obtain by the certain reasonable criteria the unique differentiable function $f(t, \omega)$, which monotonically increases with ω. In this case, the signal can be represented in the form

$$y(t) = \int_0^{\infty} \frac{\partial f(t,\omega)}{\partial \omega}\, d\omega \quad (8)$$

and the component for arbitrary frequency range (ω_1, ω_2) can be obtained through integration between these limits. How this can be done, remains to be seen. "ut in any case, we can formulate the extremum criteria to solve correctly problem (7).

5. ESTIMATE FROM ABOVE OF THE LOW-FREQUENCY COMPONENT AND ITS SPECTRUM

Let the function $f^+(t, \omega)$, which minimizes $|f'|$ in Eq. (7) at given ω, be referred to as *the estimate from above of the average value* of the corresponding low-frequency (LF) component. It is obvious that the minimum of $|f'|$ corresponds to the maxima of $|\varphi|$ and $|f|$. Let us introduce contracted notations $f_\omega = f^+(t, \omega) = \mathbf{G}_\omega \varphi_\omega$, $f'_\omega = y(t) - f_\omega$ and formulate the main properties of this estimate.

a) From Eq. (3), it follows that $|\varphi_\omega(t)| = |f_\omega(t)|$.

b) From Eqs. (4) and (6), it can be shown that if $a \ge b$, then $|\varphi_a| \ge |\varphi_b|$ and correspondingly $|f'_b| \ge |f'_a|$. Unfortunately, we cannot state that under the same conditions $\varphi_a(t) \ge \varphi_b(t)$ for any t, that is, the estimate from above for the average value, strictly speaking, is not equal to the local estimate from above. Nevertheless, when $y(t)$ meets reasonable requirements to its smoothness, it can be shown that

c) $f'_\omega(t) \to 0$ and $|f'_\omega| \to 0$ for $\omega \to 0$.

The problem of finding f_ω having the maximum modulus for any ω is a standard problem of linear programming and usually has the unique solution. Thus, the problem of low-frequency filtration can be solved correctly. The above estimate gives no way of separating the signal correctly into its low- and high-frequency components, but it can be used to estimate from above the low-frequency component and to estimate the frequency spectrum of the signal.

Let us specify the interval $[\omega_1, \omega_2]$ for which we want to find the PSF spectrum of the signal $y(t)$ and to determine $f_\omega(t)$ for any ω from this interval. The dependence $S(\omega) = |f_\omega(t)|$ so obtained we name the estimate from above for the mean value of the integral spectrum of the function $y(t)$. From the above properties, it follows that $S(\omega)$ is the monotonically increasing function, with $S(\omega) \to |y(t)|$ for $\omega \to \infty$. The spectrum so obtained can be used for qualitative analysis of the signal. We can demonstrate this with a very simple example.

6. DEMONSTRATION EXAMPLE

Let the signal be a sum of two random signals from different sources located at different distances from the measurement point. A response to the instantaneous emission from the source recorded in the measurement point is determined by the diffusion spread, and the characteristic response time depends on the distance to the source. The fluctuating signal from each source can be represented as a superposition of a random set of these responses. Let the elementary response be the Gauss function whose width for each source is individual. In Fig. 1*a*, the example is shown of simulation of this signal from two components (details are indicated under Fig. 1). The aim of our analysis in ideal is the reconstruction of the original signal components. In Fig. 1*c* the result of standard spectral selection by the Fourier method is presented. It can be seen that the method gives no way of obtaining more than one positively determined component. The information about the mean value of the high-frequency component is lost completely.

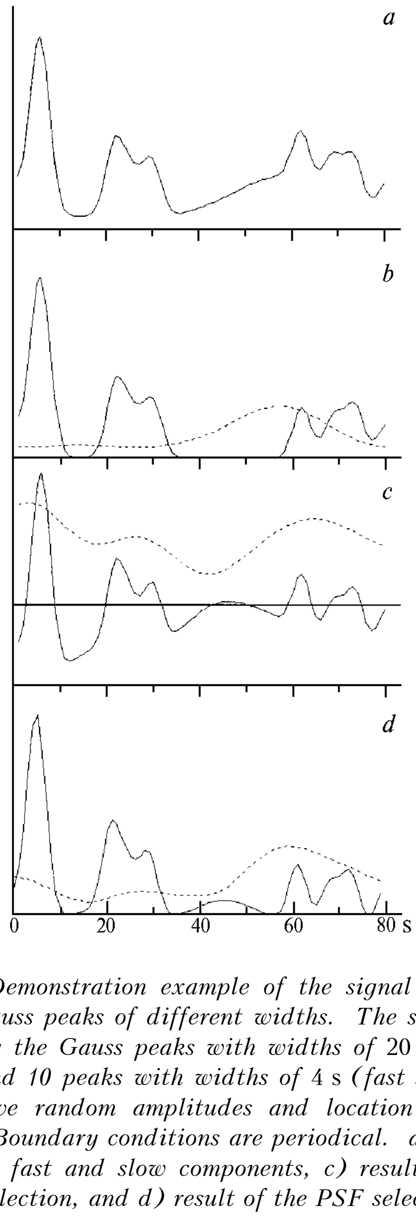

FIG. 1. Demonstration example of the signal formed by the Gauss peaks of different widths. The signal is formed by the Gauss peaks with widths of 20 s *(slow signal) and 10 peaks with widths of* 4 s *(fast signal). Peaks have random amplitudes and location in the abscissa. Boundary conditions are periodical. a) Total signal, b) fast and slow components, c) result of the Fourier selection, and d) result of the PSF selection.*

Now we examine the potentialities of the PSF method. First we derive, as described above, the estimate from above of the integral spectrum of the function. The spectrum $S(\omega)$ for the signal shown in Fig. 1 is presented in Fig. 2. In the spectrum, we can see the well-defined plateau at low frequencies near $\omega = 0.05\ s^{-1}$. This is the natural boundary for the integral selection of the two frequency components f_ω and f'_ω. These components are shown in Fig. 1*d*. It can be seen that they approximate well the original components.

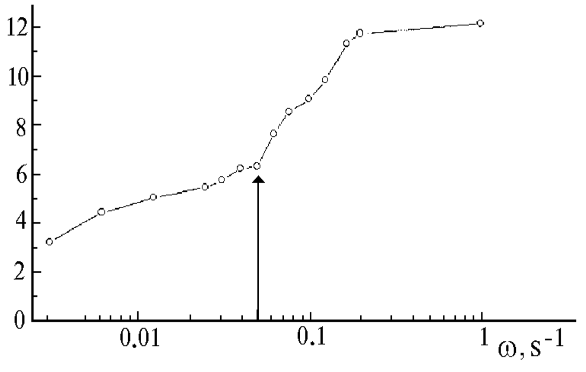

FIG. 2. Spectrum of the estimate from above of the LF component of the PSF signal shown in Fig. 1. The cutoff frequency is indicated by the arrow.

7. APPLICATION OF THE PSF METHOD TO AN ANALYSIS OF REAL DATA

The data to be analyzed were borrowed from the results of investigation of the atmospheric pollution in Novosibirsk with the help of the ICKC mobile laboratory.[6] Measurements were performed at the center of the city in December 1995. For our analysis, we selected the data on the concentration of SO_2 and NO_x gas pollutants recorded at night from 3 to 4 December. The raw data are shown in Fig. 3.

From Fig. 3, it can be seen that the pollution concentration decreases at night and that fast (with a duration of several minutes) peaks are manifested against the background slow signal caused likely by near traffic. We can identify the period (approximately from midnight to 7:30 a.m.) when the traffic contribution was relatively small.

For the spectral selection of signals, two characteristic times were chosen: 60 min for the slow trend and 5 min for cutoff of fast fluctuations. First, we selected the low-frequency component (with $\omega^{-1} = 60$ min) and after that the remainder was subject to a new selection with $\omega^{-1} = 5$ min. Thus, the low-, middle-, and high-frequency signal components shown in Fig. 4 were obtained.

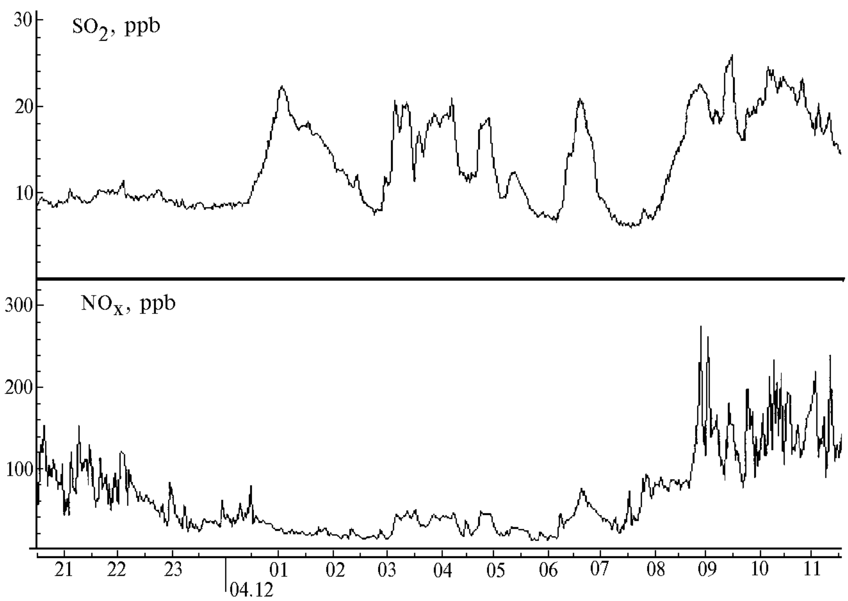

FIG. 3. Temporal dependence of SO_2 *and* NO_x *concentrations measured at the center of Novosibirsk at night from 3 to 4 December 1995.*

It can be seen from Fig. 4 that the low-frequency component is symbatic in the morning and the middle-frequency components are well correlated at night. The correlation analysis of the components yields the following results (see Table I).

Without going into detailed analysis, we can speak that the spectral selection allows us to select the signal of smoke plumes against the slowly varying background and fast fluctuations caused by the near sources. After the selection of middle-frequency component, a strong correlation between SO_2 and NO_x concentrations was established (see Table I), which permits us to determine the gas composition of this component, namely, to establish the ratio $SO_2/NO_x = 0.074 \pm 0.006$ (linear regression for the middle-frequency component from 0:00 to 7:30). The characteristic time of fluctuations and gas composition permit us to identify this signal with a plume of a small boiler.

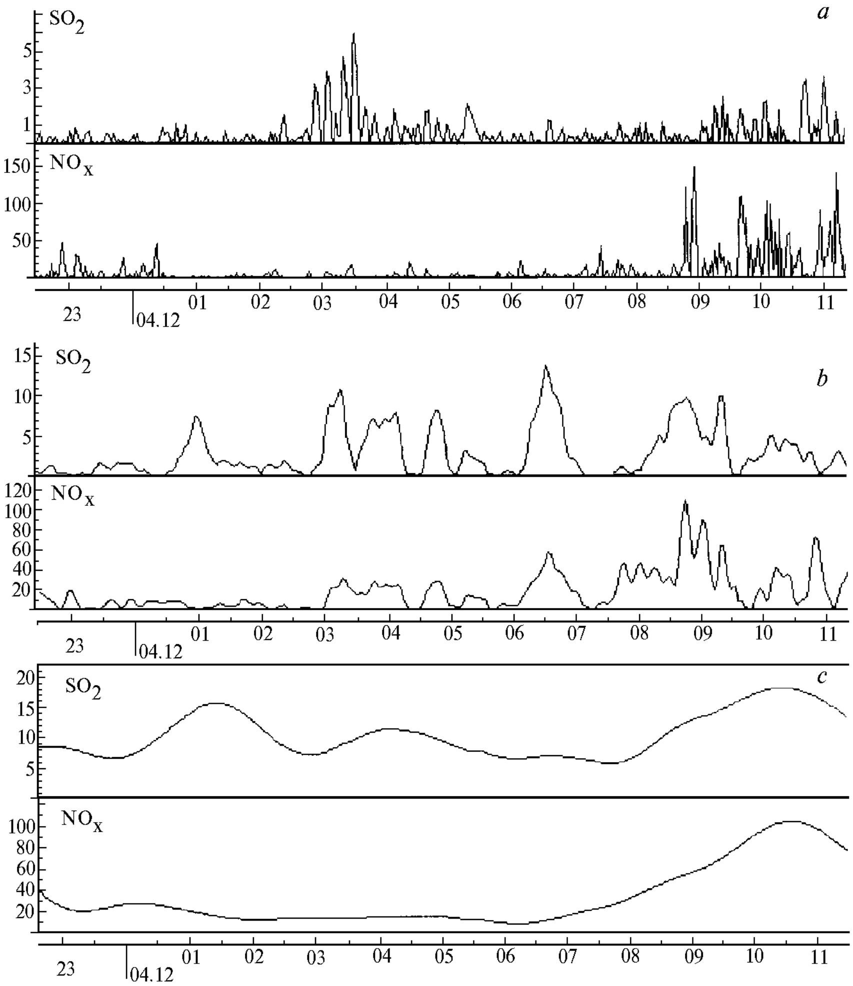

FIG. 4. Result of frequency selection of the data presented in Fig. 3 by the PSF method: a) high-frequency component; b) middle-frequency component; c) low-frequency component. Characteristic times are 5 *and* 60 min. *Concentrations are in ppb.*

TABLE I. Correlation coefficients between SO_2 *and* NO_x.

Component	Entire interval	Night (0:00–7:30)
Low-frequency	0.73	0.13
Middle-frequency	0.50	0.84
High-frequency	0.07	–0.17
Initial signal	0.48	0.35

On the basis of this approach, a computer program was developed. It is used in the Institute of Chemical Kinetics and Combustion S" RAS for studying the dynamics of gas and aerosol pollution in the atmosphere with characteristic times from several minutes to several months.

From the given examples it can be seen that this approach, in spite of its inexactness, can be efficient for frequency filtration of positively determined and strongly fluctuating signals and appears worthy of further development. Its main disadvantage now is that the estimation of the HF-component f_ω is principally from one side and therefore it is incorrect to treat the residual function f' as a corresponding high-frequency component. This shows us the direction of further development as a search for approaches that will allow us to obtain the estimate from two sides.

REFERENCES

1. P. Paatero and U. Tapper, Environmetrics **5**, 111–126 (1994).
2. P. Paatero, J.M. Makela, and V. Jokinen, J. Aeros. Sci. **26**, No. 1, S269–S270 (1995).
3. D.J. Thomson, J. Fund. Mech. **210**, 113–153 (1990).
4. A.I. Borodulin, G.M. Maistrenko, and B.M. Chaldin, *Statistical Description of Aerosol Diffusion in the Atmosphere* (Publishing House of the Novosibirsk State University, Novosibirsk, 1992).
5. D.J. Wilson, A.G. Robins, and J.E. Fackrell, Atm. Environ. **16**, No. 3, 497–504 (1982).
6. S. Rwyer, et al., J. Aeros. Sci. **26**, 379–381 (1995).