Onset Time Determination of Precursory Events in Time Series Data by an Extension of Singular Spectrum Transformation

Terumasa Tokunaga*, Daisuke Ikeda†, Kazuyuki Nakamura‡, Tomoyuki Higuchi§, Akimasa Yoshikawa*, Teiji Uozumi¶, Akiko Fujimoto‖, Akira Morioka**, Kiyohumi Yumoto¶, and CPMN group

*Graduate School of Sciences, Kyushu University, 6-10-1 Hakozaki, Fukuoka, JAPAN

†Faculty of Information Science, Kyushu University, 744 Motooka, Fukuoka, JAPAN

‡Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, 1-1-1 Higashimita, Tamaku, Kawasaki, JAPAN

§The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo, JAPAN

¶Space Environment Research Center, Kyushu University, 6-10-1 Hakozaki, Fukuoka, JAPAN

‖Japan Aerospace Exploration Agency, Institute of Space and Astronautical Science, 3-1-1Yunodai Tyuou-ku, Sagamihara, Kanagawa, JAPAN

**Planetary Plasma and Atmospheric Research Center, Tohoku University, Aramaki-aza-aoba, Aoba, Sendai, JAPAN

***Abstract*—To predict an occurrence of extraordinary phenomena, such as earthquakes, failures of engineering systems and financial market crushes, it is important to identify precursory events in time series. However, existing methods are limited in their applicability for real world precursor detections. Recently, Ide and Inoue [1] have developed an SSA-based change-point detection method, called singular spectrum transformation (SST). SST is suitable for detecting various types of change-points, but real world precursor detections can be far more difficult than expected. In general, precursory events are observed as minute and less-visible fluctuations preceding an onset of massive fluctuations of extraordinary phenomena and therefore they are easily over-looked. To overcome this point, we extend the conventional SST to the multivariable SST. The originality of our strategy is in focusing on synchronism detections of precursory events in multiple sequences of univariate time series. We performed some experiments by using artificial data and showed the superiority of multivariable SST in detecting onset of precursory events. Furthermore, the superiority is also shown statistically in determining the onset of precursory events by using real world time series.**

I. Introduction

With the prevalence of digital data acquisition, data mining has become important in diverse fields. In particular, mining extraordinary events from time series, such as earthquakes, failures of engineering systems and financial market crushes, has been pointed out as a significant problem to be studied because these phenomena could be one of the most serious threats for human activities. In most cases, such extraordinary phenomena have some underlying energy storage and release process and therefore they are often accompanied by some precursory events. If precursory events could be detected in any of these situations, and appropriate alarming mechanisms could be in place, one is given the possibility of preventing, or at least minimizing the losses caused by the phenomena. Hence it is an immediate concern to develop a method to detect precursory events from time series.

Here, we define the problem setting and consider difficulties of precursor detections in real world time series. Let us show a good example of extraordinary phenomena in real world time series. The right panel of Fig. 1 shows the northward component of geomagnetic field variations obtained at KAG (Kagoshima, Japan) station on 8 January 1997. The vertical axis in Fig. 1 indicates relative variabilities of the northward component of geomagnetic fields. Looking over the figure, we can find two noticeable features. The first thing we notice is that there is an extraordinary oscillations started around $t = 500$. Such geomagnetic variations reflect the explosive auroral activity called auroral substorms, which is widely recognized as an extraordinary phenomenon caused by the interaction between solar wind and the Earth's magnetosphere (e.g., [2]). The extraordinary oscillations, so called Pi 2 magnetic pulsations in Solar-Terrestrial Physics, usually observed all over the world at the onset of the explosive auroral activity called auroral breakup. The second thing is that the background trend begins to increase gradually around t = 100–200. As we will discuss in Section IV-B2, such gradual increasing of geomagnetic data is almost synchronized with the precipitation of electrons from space to polar-ionosphere preceding the auroral breakup. We can say that the gradual increasing of geomagnetic fields is one of precursory events associated to the auroral substorm. Like this example, what we would like to detect is minute and less-visible fluctuations preceding an onset of massive fluctuations of extraordinary phenomena. Additionally, in order to predict occurrences of extraordinary events or discover knowledge about underlying generation process of extraordinary phenomena, it is important to determine the exact onset time of precursory events. Needless to say, it is not easy to determine the onset time with a gradual initiation, especially in a case of an on-line detection. Moreover, real world time series contains various measurement noise. In noisy conditions, we can easily imagine that detecting precursory events become much more difficult.

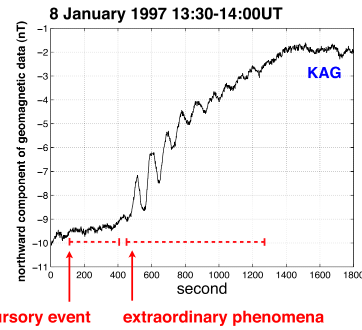

Fig. 1. Northward component of geomagnetic data around an auroral substorm onset observed at KAG station.

Studies have been made on a change-point detection in time series for a long time, and various methods have been proposed. However, conventional change-point detection methods, such as methods based on Fourier/Wavelet analysis [3] [4], AR modeling [5], text mining [6] [7] [8] and clustering [9], are not suitable for precursor detections due to reasons that we mentioned above. Besides, some researchers attempted to detect precursory events based on described precursory patterns (e.g., [10]). However, these approaches depend on the specialized experiences and therefore their applicability is only limited in specific fields.

The present work is intended to propose a general method for detecting precursory events in real world time series. From what has been discussed above, the required method for precursor detections should ideally satisfy following five essential points: (i) data adaptive, (ii) non-parametric and not underlie any specific generative models, (iii) independent on the amplitude level of signals, (iv) robust against measurement noise and (v) extendible to an algorithm that works in an on-line manner. Here, (i) and (ii) are important in terms of general versatility. (iii) and (iv) are mandatory to detect minute and less-visible changes in time series. Finally, (v) is essential to predict occurrences of extraordinary phenomena.

Recently, singular spectrum analysis (SSA) has been used for change-point detections in time series [11]. Ide and Inoue [1] developed the SSA-based change-point detection method, as referred to singular spectrum transformations (SST), and showed that it was useful in knowledge discovery of causal relationships from a set of heterogeneous time series. Unlike other conventional approaches, the SSA is data adaptive, non-parametric and does not employ any specific generative models. In addition, although basic SST performs in an off-line manner, it can be extended to work in an on-line manner. From these arguments, we can say the SST satisfies previously described essential points (i), (ii) and (v). In other words, SST is possibly applicable as a general method for detecting precursory events if we solve remaining essential points (iii) and (iv).

In this paper, we extend the basic framework of SST. In origin, the SST aims at transforming the time series into a new time series based on the change-point score (CP-score). The CP-score represents a relative anomaly metric of time series. But when we consider the nature of SSA, it is reasonable to suppose that the shaped width of the CP-score depend on the amplitude level of raw data. Let us put it another way, the original framework of SST does not satisfy the essential point (iii). To overcome this point, we propose to extend the conventional SST to the *multivariable SST* that uses multiple sequences of univariate time series. In the right panel of Fig. 2, we shows a set of geomagnetic time series obtained at KAG and CBI (Chichijima, Japan), which are comparatively adjacent observatories (see TABLE 1). What has to be noticed is that the correlation between the two data set is apparently high. In particular, the precursory event seems to start simultaneously at the two stations. Assuming some sort of an energy storage and release process underlies extraordinary phenomena, we may say the synchronism detection in multiple time series is an universal nature of precursory events. It is our expectation that we can solve the difficulties of precursor detections by focusing on the nature of precursory events, which is the reason why we propose *multivariable SST*.

TABLE I
LOCATIONS OF GEOMAGNETIC OBSERVATORIES USED IN THIS PAPER.

Station	Geographic Latitude	Geographic Longitude
CBI	27.15	142.30
KAG	31.48	130.72

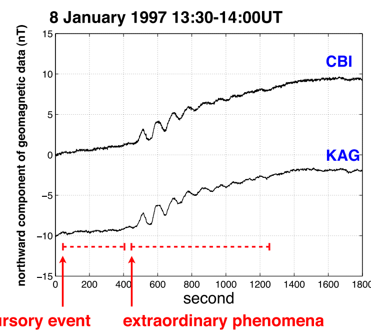

Fig. 2. Northward component of geomagnetic data around an auroral substorm onset observed at KAG and CBI station.

As an initial stage of this study, we propose the framework of precursor detections in an off-line matter. Ideally, precursory detection should be done in an on-line matter to predict occurrences of extraordinary phenomena. However, in most cases, the nature of precursory events is not fully-comprehended. Thus, as a first step, it is necessary to collect precursory events in an off-line matter. As shown in Fig. 3, precursory events can be defined as the interval between the onset of precursory events and that of extraordinary phenomena. That is, once the two types of onsets are determined, precursory events are identified automatically. Furthermore, since a determination of the onset of extraordinary phenomena is not so difficult, the determination of the onset time of precursory events in an off-line matter is an immediate concern. This is our goal at present.

The outline of the rest of the paper is as follows. In Section II, we provide a brief review of original

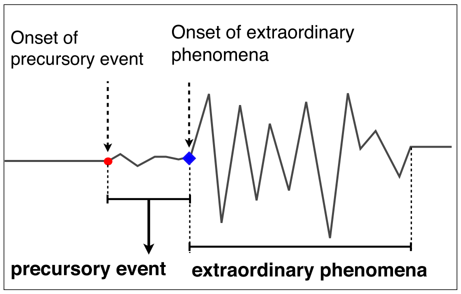

Fig. 3. A conceptual diagram of precursory events in time series. Precursory events can be defined as the interval between two types of onsets.

SST. In Section III, the multivariable SST is introduced. In Section IV-A, we perform some comparative experiments using artificial data and show the superiority of multivariable SST in detecting precursory events. In Section IV-B, we performed further comparative experiments by using ground-magnetometer data and show the superiority is valid in real world precursor detections. Section V is devoted for related work and Section VI concludes.

II. Singular Spectrum Transformations

In this section, we review the basic algorithm of the conventional SST proposed in Ide and Inoue [1]. Basically, SST is a nonlinear transformation from an original time series to a new time series that represents a relative anomaly metric of the original time series. The metric is defined as the distance between two subspaces, which spanned by left singular vectors obtained via the singular value decomposition (SVD) on a Hankel matrix generated from subsequences of the original time series.

A. Pattern Extraction by SSA

In this subsection, we provide an explanation of procedure for the pattern extraction by SSA. Basically, the SSA is an exploratory method intended to perform decomposition of a time sequence into a sum of interpretable components, such as trend, periodicities and noise. These interpretable components can be viewed as representative patterns. The procedure for the pattern extraction is done via the singular value decomposition (SVD) on a Hankel matrix generated from an original time sequence.

First of all, let us consider a transformation of a sequence time series $Y = \{y_1, y_2, ..., y_K, ..., y_N\}$ into the multi-dimensional series $\mathbf{X} = [X_1, X_2, \dots, X_K]$, where the X_i denotes a subsequence that can be described as $X_i = (y_i, \dots, y_{i+L-1})^T$ $(1 \le i \le K)$. Here, subscript T denotes the transpose of a matrix. Vectors X_i's and the matrix $\mathbf{X}$ are called L-lagged vectors and an L-trajectory matrix, respectively. Note that an L-trajectory matrix $\mathbf{X}$ is an $L \times K$ Hankel matrix described as

$$\mathbf{X} = \begin{pmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \ddots & \vdots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{pmatrix}. \tag{1}$$

We call K and L a window length and an embedding dimension, respectively.

The second step of the SSA is the SVD of the Hankel matrix $\mathbf{X}$. Let us denote $(\lambda_1, \lambda_2, \dots, \lambda_L)$ as squared singular values of $\mathbf{X}\mathbf{X}^T$ in decreasing order of the magnitude $(\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_L)$, where the subscript T denotes transpose of a matrix. Now the SVD of the Hankel matrix $\mathbf{X}$ can be described as $\mathbf{X} = \lambda \mathbf{U}\mathbf{V}^T$, where λ denotes a diagonal matrix whose diagonal element equal to the squared singular values, $\mathbf{U}$ denotes a left singular matrix and $\mathbf{V}$ denotes a right singular matrix. Then, the Hankel matrix $\mathbf{X}$ can be described as a sum of rank-one bi-orthogonal elementary matrices $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_L$. The ith elementary matrix $\mathbf{X}_i$ can be described by using the ith left singular vector and the ith right singular vector as $\mathbf{X}_i = \lambda_i U_i V_i^T$. Note that L corresponds to the number of singular vectors. Now let us define *representative patterns* using left singular vectors

$\{U_l\}$ $(1 \leq l \leq L)$. As described above, the method to extract dominant structures in time series via the SVD on the Hankel matrices is referred to as the singular spectrum analysis (SSA).

B. Change-Point Score

In this subsection, we introduce the original definition of the Change-point score (CP-score) that represents a relative anomaly metric of time series at the present time t. The CP-score is calculated by using left singular vectors obtained by the SVD of the Hankel matrix $\mathbf{X}$, which can be viewed as *representative patterns*.

First, we define the reference interval and the test interval as illustrated in the left panel of Fig. II-A. Next, the reference subspace is defined to be a subspace spanned by representative patterns extracted from the reference interval. If we denote $N = \{1, ..., n\}$, the reference subspace is described as $\mathcal{H}^{ref} = \text{span}\{U_n^{ref}\}$ $(n \in N)$. Similarly, if we denote $M = \{1, ..., m\}$, the test subspace is described as $\mathcal{H}^{test} = \text{span}\{U_m^{test}\}$ $(m \in M)$. In the right panel of Fig. II-A, we show a diagram of the reference subspace and the test subspace. Then, the CP-score at time t is defined as $Z \equiv 1 - cos\Theta(\mathcal{H}^{ref}, \mathcal{H}^{test})$, where $\Theta(\mathcal{H}^{ref}, \nu^{test})$ represents the angle between the reference subspace and the test subspace. In this paper, the angle between two subspaces computed by using MATLAB function *subspace*. Note that Z is non-dimensional parameter and limited to the range from zero to one by definition. The calculation of the CP-score can be viewed also as a nonlinear transformation from an original time-series $\mathcal{T}$ to a new time-series $\mathcal{T}_c$, i.e.

$$\mathcal{T} \rightarrow \mathcal{T}_c(K, L, g, m, n). \tag{2}$$

This is the basic algorithm of the singular spectrum transformation (SST).

C. Choice of Parameters

As expressed in Eq. (2), the SST algorithm includes five parameters. By nature of SST, these parameters should be determined experimentally. To detect precursory events, some careful tuning of these SST parameters may be required. In particular, Moskvina and Zhigljavsky [11] pointed out the choice of K depends on the kind of structural changes that we are looking for. Ide and Inoue [1] showed that the stability of SST against the change of K in an simple experiment by using artificial data, but it is not clear whether the stability is valid even in real world applications. So we will show SST results as a function of K in the following applications. Remaining parameters, L, g, m and n, are obliged to determined empirically. In Fig. 5, we show an example of SST results for the periodic signal. The SST parameters were set to $K = 40$–78, $L = K$, $g = K/2$, $m = 1$ and $n = 3$. The periodic signal shown in upper panel of Fig. 5 was generated using sine functions whose frequency changes at $t = 1000$ and $t = 2000$. We see that the CP-score increases sharply at $t = 1000$ and $t = 2000$ over the wide range of K. This result is consistent with that in [1].

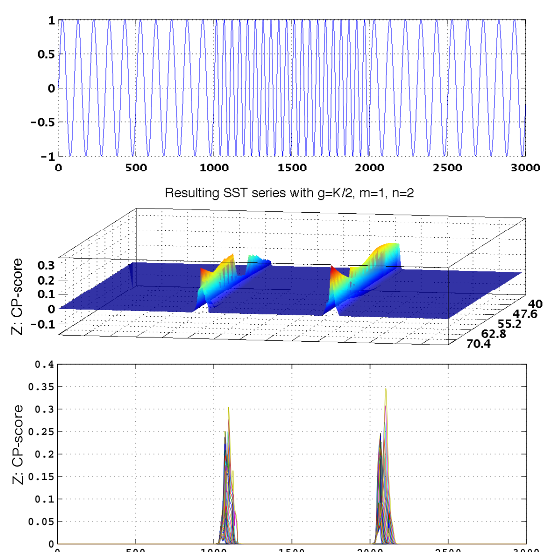

Fig. 5. Upper panel: Original data generated using sine functions whose frequency changes at $t = 1000$ and $t = 2000$. Lower panel: The resulting conventional SST series with $K = 40$–78, $L = K$, $g = K/2$, $m = 1$ and $n = 3$.

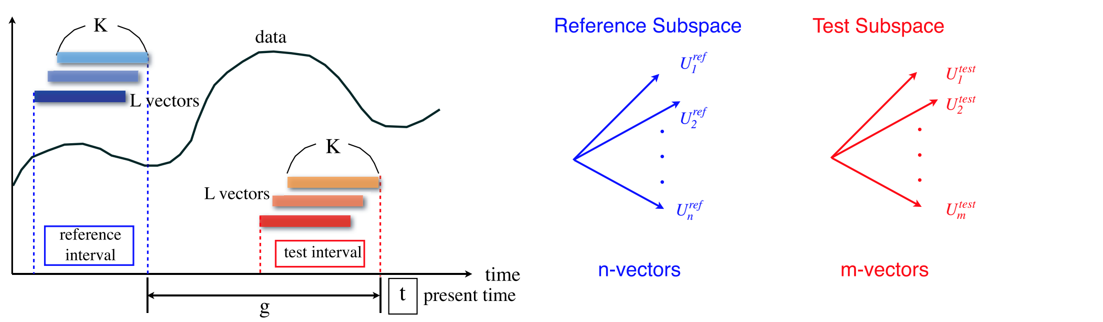

Fig. 4. (Left panel) A diagram of pattern extraction by conventional SST. From L subsequences at both side of present time t, *representative patterns* are calculated. (Right panel) A diagram of the reference subspace and the test subspace. Change-point score is defined by using the angle between these two subspaces.

III. MULTIVARIABLE SST

In this section, we extend the conventional SST to the *multivariable SST* that uses multiple sequences of univariate time series. As we stated in Section I, precursory events are usually observed as minute and less-visible fluctuations preceding an onset of massive fluctuations of extraordinary phenomena and therefore they are easily over-looked. Now, recall our problem setting of real world precursor detections discussed in Section I. First, the onset of precursory events are synchronized in multiple sequences of univariate time series, which is the most essential assumption on our strategy. Second, extraordinary phenomena are observed almost simultaneously but not synchronized exactly in multiple sequences. Third, real world time series contains various measurement noise. Here, measurement noise can be regarded as uncorrelated of each other in multiple time sequences. It follows from these arguments: In oder to detect precursory events in real world time series, it is likely valid to focus on changes that observed simultaneously in multiple time sequences. This is why we propose the multivariable SST.

Now let us redefine the CP-score shown in Section II-B. First of all, consider J sets of sequences as $Y_j = \{y_{j,1}, y_{j,2}, \ldots, y_{j,K}, \ldots, y_{j,N}\}$. By embedding these sequences, the Hankel matrix described in Eq. (1) can be rewritten as

$$\mathbf{X}_j = \begin{pmatrix} y_{j,1} & y_{j,2} & \cdots & y_{j,K} \\ y_{j,2} & y_{j,3} & \cdots & y_{j,K+1} \\ \vdots & \ddots & \vdots & \vdots \\ y_{j,L} & y_{j,L+1} & \cdots & y_{j,N} \end{pmatrix}. \quad (3)$$

According to the procedure in Section **??**, we obtain representative patterns as left singular matrices $U_{(l,j)} (1 \leq l \leq L)$ $(j \in J)$ via the SVD on Hankel matrix as $\mathbf{X}_{(l,j)} = \lambda_{(l,j)} \mathbf{U}_{(l,j)} \mathbf{V}^T_{(l,j)}$. Then, the reference subspace is redefined as $\hat{\mathcal{H}}^{ref} = \mathrm{span}\{U^{ref}_{(n,j)}\} (n \in N, j \in J)$. Similarly, the test subspace is redefined as $\hat{\mathcal{H}}^{test} = \mathrm{span}\{U^{ref}_{(m,j)}\} (m \in M, j \in J)$. Finally, the definition of the CP-score is redefined as $Z \equiv 1 - cos\Theta(\hat{\mathcal{H}}^{ref}, \hat{\mathcal{H}}^{test})$, where $\Theta(\hat{\mathcal{H}}^{ref}, \hat{\mathcal{H}}^{test})$ represents the angle between the redefined reference subspace and the redefined test subspace. This procedure can be viewed as a nonlinear transformation from original J sets of sequences of univariate time series $\mathcal{T}_j$ $(j \in J)$ to a new time-series $\mathcal{T}_c$, i.e.

$$\mathcal{T}_j \ (j \in J) \rightarrow \mathcal{T}_c(K, L, g, m, n). \quad (4)$$

We call this transformation *multivariable SST*. Although we use multivariable time sequences, note that our purpose is somewhat different from that in Ide and Inoue [12]. In particular, the purpose in

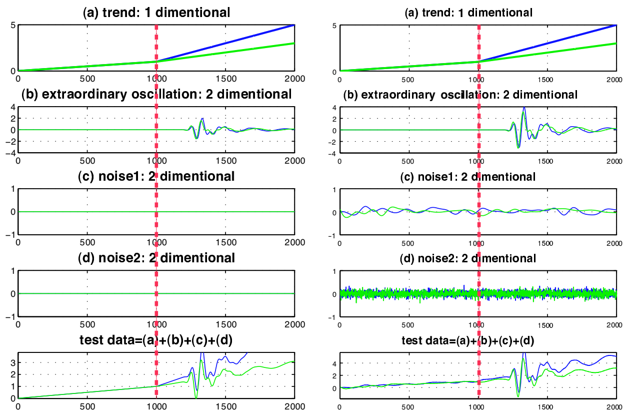

Fig. 6. Left panel: Test data 1. The variance of noise components (c) and (d) are set to zero. Right panel: Test data 2. The variance of noise components (c) and (d) are set to 0.1.

Ide and Inoue [1] is to make the multivariable time series of *heterogeneous* types be comparable with each other by applying the SST. However, we are not concerned with such heterogeneous data set. We limit the application of the SST to comparatively *homogeneous* data set.

IV. Experiments

A. Application to Artificial Data

1) Experiment Settings: In this subsection, we perform some experiments to compare performances of conventional SST and multivariable SST in determining the onset of precursory events by using artificial data. Fig. 6 shows two types of test data, which we generated based on geomagnetic variations shown in Fig. 2. These data was generated by combining following four components: (a) trend, (b) extraordinary oscillations, (c) noise 1 and (d) noise 2. (a) was generated using two linear functions. Its slope changes at $t = 1000$. Recalling our problem setting about precursor detections described in Section I and in Section III, the change of slopes are regarded as the onset of precursory events. This slope change are shown as the vertical dashed line in red. Since one of our interest is to validate how measurement noise affect the precise in the precursor detection, the test data shown in right panel of Fig. 6 contains two types of noise components. Noise components in (c) are low-pass transformed random noise whose elements are normally distributed. Noise components in (d) are high-frequency random noise whose elements are normally distributed.

2) Application to Noiseless Data: First, we show SST results applied for the noiseless data shown in the left panel of Fig. 6. In Fig. 7(a), we represent the resulting conventional SST series for the test data 1 shown in the left panel of Fig. 6. Similarly, in Fig. 7(b), we represent the resulting multivariable SST series for the test data 1 shown in the left panel of Fig. 6. Comparing these results, we can see the noticeable difference between the resulting conventional SST series and the resulting multivariable SST series. In the resulting multivariable SST series, the CP-score increases sharply at $t = 1000$. This result show that the onset of the precursory event, which we defined above, was detected exactly by multivariable SST. Contrastively, in the resulting conventional SST series shown in 7(a), we can see some noticeable peak of the CP-score around t = 1400–1600. These peak most likely corresponds to the extraordinary oscillations, which is shown in the panel (b) of the Fig. 6. However,

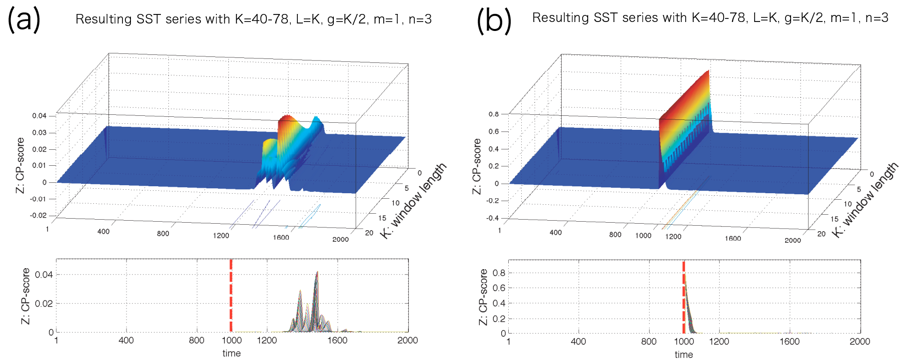

Fig. 7. (a) The resulting conventional SST series for the test data 1 shown in the left panel of Fig. 6. (b) The resulting multivariable SST series for the test data 1 shown in the left panel of Fig. 6. SST parameters were set to $K = 40$–78, $L = K$, $g = K/2$, $m = 1$ and $n = 3$. The vertical dashed line shown in red indicates the onset time of the precursory event.

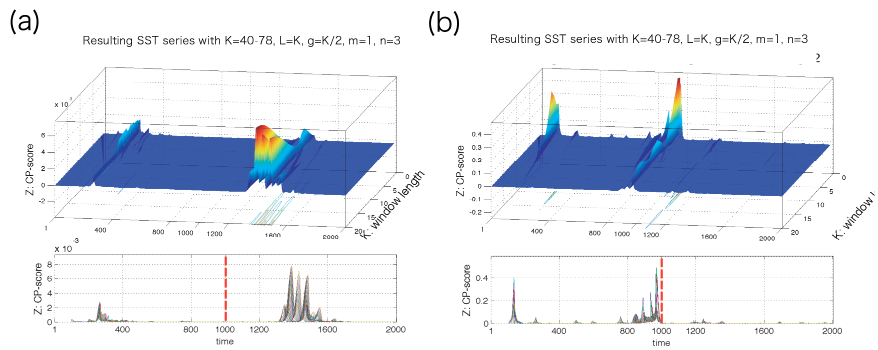

Fig. 8. (a) The resulting conventional SST series for the test data 1 shown in the right panel of Fig. 6. (b) The resulting multivariable SST series for the test data 1 shown in the right panel of Fig. 6. SST parameters were set to $K = 40$–78, $L = K$, $g = K/2$, $m = 1$ and $n = 3$. The vertical dashed line shown in red indicates the onset time of the precursory event.

we cannot find any clear peak around $t = 1000$. This result show that conventional SST is suitable for detecting extraordinary phenomena but weak in detecting slope changes, or probably in detecting gradual changes. These results strongly show the superiority of multivariable SST in detecting precursory events.

3) Application to Noisy Data: Next, two experiments were done in the noisy condition, which is more realistic one. Fig. 8(a) shows resulting conventional SST series for the test data 2 shown in right panel of Fig. 6. Similarly, Fig. 8(b) shows multivariable SST series for the test data 2 shown in right panel of Fig. 6. The SST parameters were set to K=40–78, $L = K$, $g = K/2$, $m = 1$ and $n = 3$. In Fig 8(b), the first point to notice is that the CP-score increases rapidly around $t = 1000$ in the multivariable SST series, as similar to the result in Fig. 7(b). Although some false peak of the CP-score are seen around $t = 400$ and $t = 700$, such instabilities are most likely within a the allowable range.

Third, we performed the similar experiment with a different parameter of m and n. Fig. 9(a) and Fig. 9(b) shows resulting conventional SST series and multivariable SST series, respectively. Here, SST parameters were set to K=40–78, $L = K$, $g = K/2$, $m = 4$ and $n = 5$. As shown in

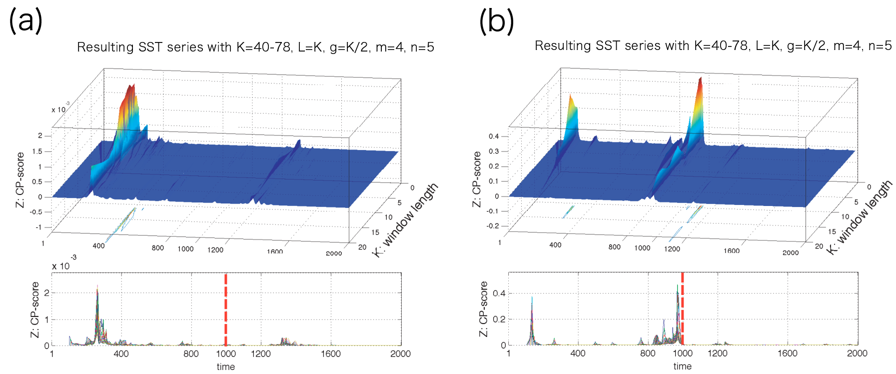

Fig. 9. (a) The resulting conventional SST series for the test data 1 shown in the right panel of Fig. 6. (b) The resulting multivariable SST series for the test data 1 shown in the right panel of Fig. 6. SST parameters were set to $K = 40$–78, $L = K$, $g = K/2$, $m = 4$ and $n = 5$. The vertical dashed line shown in red indicates the onset time of the precursory event.

Fig. 9(b), the essential result in resulting multivariable SST series was same in Fig. 6. Thus we can say that multivariable SST is not so sensitive for the parameter change of m and n. In contrast, in the resulting conventional SST series, the peaks of the CP-score around $t = 1400$–1600 went down drastically. This result imply that the conventional SST is comparatively sensitive for the parameter change of m and n.

B. Application to Ground-Magnetometer Data

1) Data Set: Ground magnetometer data obtained from the CPMN (Circum-pan Pacific Magnetometer Network) stations in the 210° magnetic meridian (MM) chain [12] have been applied for SST. CPMN consists of about 50 stations. In this study, the data set obtained at KAG and CBI stations were selected for application of SST. The locations of these stations are listed in TABLE 1. The observations were based on vector measurements by fluxgate magnetometers with a sampling rate of 1Hz. Fig. 10(c) shows northward component of ground-magnetometer data around an auroral breakup. Thick line represents KAG data and thin line represents CBI data. Horizontal axis shows universal time.

2) Reference Data: In order to validate the reliability of the SST-based precursor detections, information about auroral activities are useful. In this paper, we referenced the Polar Satellite Ultra Violet Imager (Polar/UVI) to check the global auroral activities. In Fig. 10(a), we represent Polar/UVI images at N_2 Lyman-Birge-Hopfield bands (1400-1600Å) on 4 January 1997. We see that the intensity of aurora started to enhance exponentially between 14:45:08UT and 14:46:22UT. This sudden brightening of aurora is called an auroral breakup (or just breakup), e.g., [13], which is widely recognized as an extraordinary phenomenon caused by the interaction between the solar wind and the magnetosphere. While, we see the initial brightening of aurora started between 15:42:04UT and 14:44:50UT.

Further, to provide more evidence about the precursory event, AKR (Auroral Kilometric Radiation) spectrogram provided by Polar satellite Plasma Wave Instrument (Polar/PWI) electric field observations was also checked. Such remote observation of AKR has been used as a tool for detecting the dynamics of the auroral acceleration region. Morioka et al. [14] and [15] derived the features of the sudden build-up process of field-aligned acceleration at substorm onset from remote AKR observations. In Fig. 10(b), we represent AKR spectrogram around substorm onset. In Fig. 10(b), we can see the high-frequency AKR developed gradually around 15:44UT and the low-frequency AKR developed explosively around 14:46UT. Such two stage evolution of AKR are recently reported in Morioka et al.[15] [16]. They also reported that the explosive enhancement of the low-frequency AKR and the gradual appearance of the high-frequency AKR is likely corresponds to the auroral breakup and auroral initial brightening.

From these observations, we can infer that the auroral breakup, the extraordinary phenomenon, likely started within the interval between 15:45:08UT and 15:46:22UT. Similarly, we infer the auroral initial brightening, the precursory event, started within the interval between 15:42:04UT and 15:44:50UT. These information enables us to evaluate the reliability of onset time determined by applying SST.

3) Application of the Conventional SST: Now let us apply conventional SST to the ground-magnetometer data shown in Fig. **??**(c). In a middle and bottom panel in Fig. 11(a), we represent results of the conventional SST for the geomagnetic data shown in Fig. 10(c) calculated with $K = 40$–78(s), $L = K$, $g = K/2$, $m = 4$ and $n = 5$. A onset time can be defined as maximum time of the CP-score. The red dot in a top panel in Fig. 11(a) shows an average of maximum time of the CP-score, which corresponds to the onset time determined by conventional SST. The error bar shows its standard deviation. We see the CP-score increases rapidly around 15:49:52UT over the wide range of K. As we mentioned in Section IV-B2, the auroral breakup likely started around 15:46UT. Hence, the sharp peak of the CP-score around 14:37UT likely corresponds to the auroral breakup. While, as shown in Section IV-B2, the formation of the filed-aligned electric fields started around 15:44UT. But we can not find any sharp peak of the CP-score around 15:44UT.

4) Application of the Multivariable SST: Next, let us apply multivariable SST to two sets of ground-magnetometer data shown in Fig. 10(c). A middle and bottom panel of Fig. 11(b) shows the resulting multivariable SST series calculated with K=40 – 78, $L = K$, $g = K/2$, $m = 4$ and $n = 5$. The parameter settings were same in Fig 11(a). The red dot in a top panel in Fig. 11(b) shows an average of maximum time of the CP-score, which corresponds to the onset time determined by multivariable SST. The error bar shows its standard deviation. Interestingly, the CP-score increases rapidly around 15:45:12UT around K=50–60. As we have mentioned in Section IV-B2, the precursory event likely started around 15:44UT, so it is reasonable to suppose that the precursory event in geomagnetic data is detected successfully by multivariable SST. It follows from these experimental results that the multivariable SST is more suitable than the conventional SST to determine the onset time of precursory events in time series, at least in this case study. This will be examined further in the next subsection.

5) Statistical Evaluation: Finally, we evaluate statistically the precise of precursor detections by conventional and multivariable SST and illustrate the superior performance of multivariable SST with a comparative experiments. TABLE 2 shows an event list of auroral substorms used in the experiment. Seventeen isolated substorm events in 1997 were picked up by checking UVI data obtained by Polar satellite and ground-magnetometer data obtained at KAG station. Fig IV-B5 illustrates the timing relations between auroral brightening and precursor onsets in ground-magnetometer data. Red dots show the average of maximum time of the CP-score calculated by multivariable SST, which is defined as the onset time determined by multivariable SST. Blue squares shows the average of maximum time of the CP-score calculated by conventional SST, which is defined as the onset time determined by conventional SST. Error bars represent their standard deviations. SST parameters were set to K=40 – 78, $L = K$, $g = K/2$, $m = 4$ and $n = 5$, which is same in subsection IV-B3. The vertical axis shows the time difference from auroral initial brightening determined by viewing Polar/UVI. Error bars shown in black represent time gap between UVI images. This clearly shows that onset time determined by

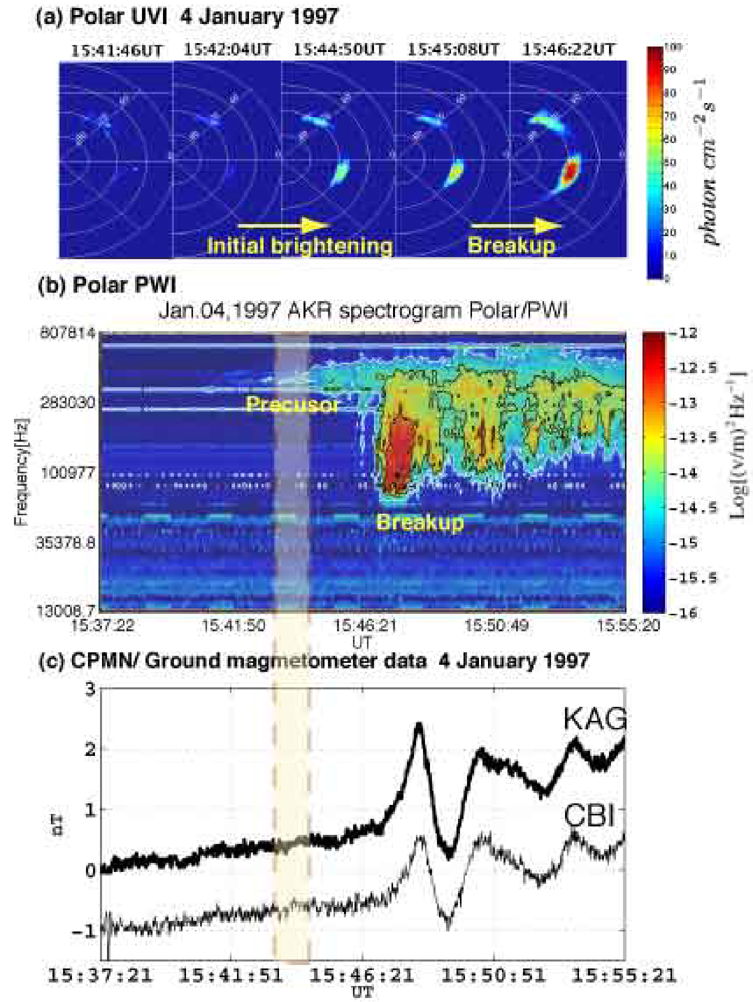

Fig. 10. (a) Polar/UVI images at N_2 Lyman-Birge-Hopfield bands (1400-1600Å) on 4 January 1997. (b) AKR spectrogram provided by Polar/PWI electric fields observations on 4 January 1997, which shows a precursory event and an auroral breakup started around 15:44UT and around 15:46UT, respectively. (c) Northward component of ground magnetometer data obtained at KAG station (thick line) and CBI station (thin line) on 4 January 1997.

multivariable SST is earlier than that determined by conventional SST in all events. Furthermore, in 14 events, the onset time determined by multivariable SST is closer to the onset of the auroral initial brightening in comparison with the onset time determined by conventional SST. From the results, we can safely say that the superiority of multivariable SST in determining the onset time of precursory events is valid in most of isolated substorm events.

V. Related Work

The problem of precursor detections has been studied for a long time. Traditionally, Fourier analysis [3] and Wavelet analysis [4] have been used for detecting change-points in time series. But their approaches are most likely unsuitable for detecting a

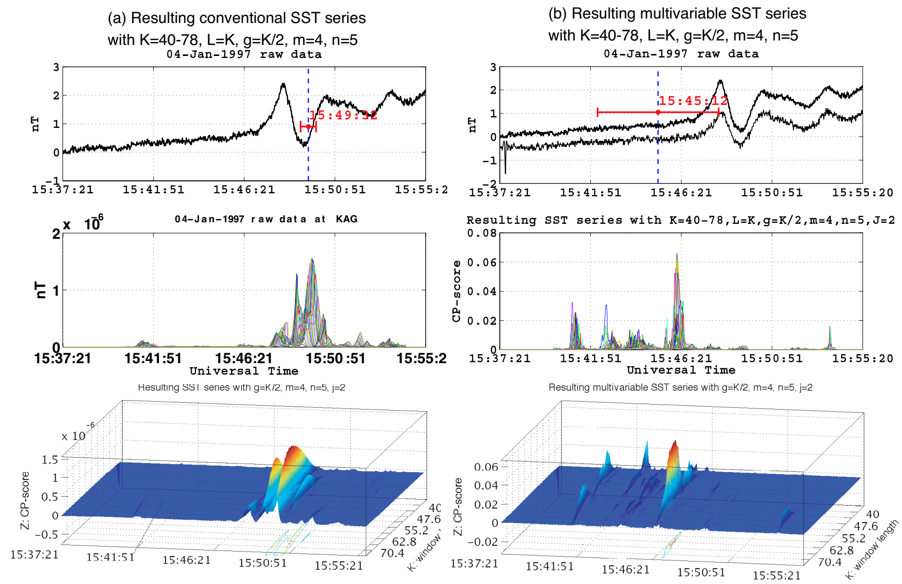

Fig. 11. (a) Resulting conventional SST series with K=40–78, L=K, g=K/2, m=4, n=5. (b) A diagram of the reference subspace and the test subspace. Change-point score is defined by using the angle between these two subspaces.

TABLE II
AN SUBSTORM EVENT LIST USED IN THIS STUDY.

Event No.	Date	Universal Time
1	4 January 1997	15:37:21-15:55:21
2	7 January 1997	17:18:51-17:36:51
3	8 January 1997	14:03:51-14:21:51
4	8 January 1997	14:27:51-14:45:51
5	12 January 1997	15:34:51-15:52:51
6	13 January 1997	12:14:51-12:32:51
7	21 January 1997	14:17:21-14:35:21
8	22 January 1997	14:58:21-14:16:21
9	2 February 1997	12:46:21-13:04:21
10	15 February 1997	13:44:21-14:02:21
11	15 February 1997	15:37:21-15:55:21
12	1 March 1997	12:50:51-13:08:51
13	1 March 1997	13:22:51-13:40:51
14	4 March 1997	14:11:51-14:29:51
15	15 March 1997	16:09:21-16:27:21
16	23 March 1997	15:24:21-16:42:21
17	18 May 1997	16:25:21-16:43:21

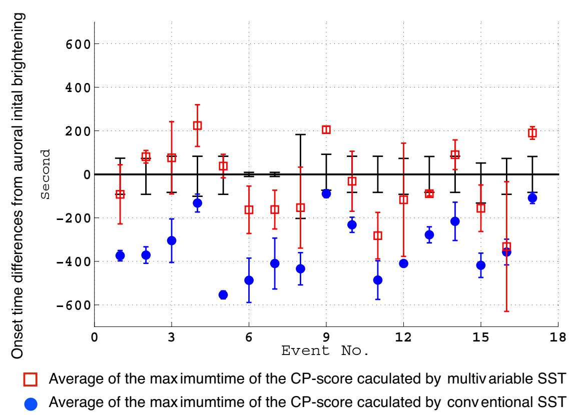

Fig. 12.

gradual onset, which is probably an essential feature in precursory events. Besides, some re- searchers attempted to detect precursory events of earthquake based on described precursory patterns [10] or based on AR modeling [5]. Since it is highly unlikely that some general rules underlie precursory events, the applicability of their approaches is limited to the phenomena whose structures are comparatively understood. More recently, the problem of precursor detections are discussed in terms of symbolic dynamics and it has been applied for detecting earthquake precursors [6] and failure precursor in electrical systems [7]. But these approaches, and any other text mining based approaches, are most likely unsuitable for detecting minute and less-visible fluctuations preceding an onset of massive fluctuations from real world time series data. Batch-style scenarios, such as KeyGraph [8] and based clustering [9], have been proposed for finding risky active faults. But since these approaches based on huge amount of data set, their applications are only limited in an off-line manner.

Recently, it has been studied well in terms of data mining [17] [18] [19]. Traditionally, Fourier analysis and Wavelet analysis have been used for detecting change-points in time series. But their approaches are most likely unsuitable for detecting a gradual onset, which is probably an essential feature in precursory events. Besides, some researchers attempted to detect precursory events of earthquake based on described precursory patterns [10] or based on AR modeling [5]. Since it is highly unlikely that some general rules underlie precursory events, the applicability of their approaches is limited to the phenomena whose structures are comparatively understood. More recently, the problem of precursor detections are discussed in terms of symbolic dynamics and it has been applied for detecting earthquake precursors [6] and failure precursor in electrical systems [7]. But these approaches, and any other text mining based approaches, are most likely unsuitable for detecting minute and less-visible fluctuations preceding an onset of massive fluctuations from real world time series data. Batch-style scenarios, such as KeyGraph [8] and based clustering [9], have been proposed for finding risky active faults. But since these approaches based on huge amount of data set, their applications are only limited in an off-line manner.

Moskvina and Zhigljavsky [11] used the singular spectrum analysis (SSA) technique for change detection in time series, based on the SVD of the Hankel matrix. Originally, SSA aim at reconstructing principal structures from time series and making a decomposition of an original time series into the sum of a small number of uncorrelated and interpretable components. It has been applied to forecast various phenomena, such as industrial production [20], daily exchange rates [21] and hydrological variations [22]. Ide and Inoue [1] developed the SSA-based change-point detection method, as referred to singular spectrum transformations (SST), by adopting dimensionless definition of the CP-score and showed their algorithm is applicable for various types of time series generated in heterogeneous dynamic systems. In this paper, we have extended it from conventional SST to multivariable SST so that it is applicable for real world precursor detections. The originality of our strategy is in focusing on synchronism detections of precursory events in multiple sequences of univariate time series. We performed some experiments by using artificial data and showed the superiority of multivariable SST in detecting onset of precursory events. Furthermore, the superiority is also shown in real world precursor detections by using ground-magnetometer data. In authors best knowledge, this is the first work to show the validity of an SSA-based change-point detection method to determine the onset time of precursory events in time series in an objective and a quantitative matter.

VI. Concluding Remarks

It is easy to imagine that the detection of precursory events is important to predict occurrences of extraordinary phenomena. In order to minimize the losses caused by extraordinary phenomena, precursory events should be detected in an on-line matter. However, in most cases, the generation mechanism or nature of the precursory events is not fully-comprehended. Thus, the first step should be done for hazard prediction is to collect precursory events in an off-line matter. Precursory events can be defined as the interval between the onset of precursory events and that of extraordinary phenomena. The determination of onset time of extraordinary phenomena is not so difficult. Hence, in this paper, we focused our attention on how to determine the onset time of precursory events in an off-line matter. Although the problem of precursor detections have been studied for a long time, conventional methods are no match for precursor detections. In this paper, we have extended the conventional SST to the multivariable SST, focusing on synchronism detections of precursory events in multiple sequences of univariate time series. We applied the multivariable SST for ground magnetometer data that records the precursory event associated to the auroral substorm. In a case study. the multivariable SST showed the superior performance in determining the onset time of precursory events in comparison with conventional SST. Furthermore, we performed further experiments for 17 auroral substorm events. As a result, we confirmed that the superiority of multivariable SST is valid in another 14 auroral substorm events.

ACKNOWLEDGMENT

We would like to thank J. Dowell of the University of Iowa for providing the Polar/PWI data, C. Meng and K. Liou of Johns Hopkins University for providing Polar/UVI data. We also thank all the members of the CPMN project for their ceaseless support. The CPMN project is financially supported by the Ministry of Education, Science, and Culture of Japan (and Japan Society for the Promotion of Science) as the Grants-in-Aid for Overseas Scientific Survey (05041060, 0841105, 10041122, 12373003). The author (T. Tokunaga) was supported by the Japan Society for the Promotion of Science (JSPS) under Grant-in-Aid for Scientific Research 2005443. The author (D. Ikeda) was supported by the Japan Society for the Promotion of Science (JSPS) under Grant-in-Aid for challenging Exploratory Research 21650031.

REFERENCES

[1] T. Ide and K. Inoue, "Knowledge Discovery from Heterogeneous Dynamic Systems using Change-Point Correlations," *In Proceedings of 2005 SIAM International Conference on Data Mining*, pp. 571–576, 2005.

[2] T. Saito, "Geomagnetic pulsations," *Space Science Review*, vol. 10, pp. 319–412, 1969.

[3] K. Takahashi, S. Ohtani, and B. J. Anderson, "Statistical analysis of Pi 2 pulsations observed by the AMPTE CCE spacecraft in the inner magne- tosphere," *J. Geophys. Res.*, vol. 100, pp. 21 929–21 941, 1995.

[4] M. Nose, T. Iyemori, M. Takeda, T. Kamei, D. K. Milling, D. Orr, H. J. Singer, E. W. Worthington, and N. Sumitomo, "Automated detection of Pi 2 pulsations using wavelet analysis: 1. Method and an application for substorm monitoring," *Earth, Planetary and Space Sciences*, vol. 50, pp. 773–783, 1998.

[5] G. D. Bello, V. Lapenna, M. Macchiato, C. Satriano, and C. Serio, "Parametric time series analysis of geoelectrical signals: an application to earthquake forecasting in Southern Italy," *Annali di Geofisica*, vol. XXXIX, no. 1, January 1996.

[6] K. Karamanos, A. Peratzakis, P. Kapiris, S. Nikolopoulos, J. Kopanas, and K. Eftaxias, "Extracting preseismic electromagnetic signatures in terms of symbolic dynamics," *Nonlinear Processes in Geophysics*, vol. 12, pp. 835–848, 2005.

[7] R. P. Patankar, V. Rajagopalan, and A. Ray, "Failure precursor detection in complex electrical systems using symbolic dynamics," *International Journal of Signal and Imaging Systems Engineering*, vol. 1, no. 1, pp. 68–77, 2008.

[8] Y. Ohsawa, "KeyGraph as Risk Explorer in Earthquake-Sequence," *Journal of Contingencies and Crisis Management*, vol. 10, pp. 119–128, 2002.

[9] K. Fukuda, M. Koganeyama, H. Shouno, T. Nagao, and J. Kazuki, "Detecting Seismic Electric Signals by LVQ Based Clustering," *International Conference on Parallel and Distributed Processing Techniques and Applications*, pp. 1305–1311, 2001.

[10] H.-M. Huang, H.-S. Fan, Y.-J. Bian, and L.-Y. Zou, "Investigation into the automatic recognition of time series precursor of earthquakes," *Acta Seismologica Sinica*, vol. 11, no. 5, pp. 605–614, 1998.

[11] V. Moskvina and A. Zhigljavsky, "An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection," *Communications in Statistics—Simulation and Computation*, vol. 32, pp. 319–352, 2003.

[12] K. Yumoto and the CPMN Group, "Characteristics of Pi 2 magnetic pulsations observed at the CPMN stations: A review of the STEP results," *In Earth Planets Space*, vol. 53, pp. 981–992, 2001.

[13] S.-I. Akasofu, "The development of the auroral substorm," *Planetary and Space Science*, vol. 12, pp. 273–282, 1964.

[14] A. Morioka, Y. Miyoshi, F. Tsuchiya, H. Misawa, T. Sakanoi, K. Y. R. R. Anderson, J. D. Menietti, and E. F. Donovan, "Dual structure of auroral acceleration regions at substorm onsets as derived from akr spectra," *J. Geophys. Res.*, vol. 112, no. A06245, 2007.

[15] A. Morioka, Y. Miyoshi, F. Tsuchiya, H. Misawa, K. Yumoto, G. K. Parks, R. R. Anderson, J. D. Menietti, E. F. Donovan, F. Honary, and E. Spanswick, "AKR breakup and auroral particle acceleration at substorm onset," *J. Geophys. Res.*, vol. 113, 2008.

[16] A. Morioka, Y. Miyoshi, F. Tsuchiya, H. Misawa, K. Yumoto, G. K. Parks, R. R. Anderson, J. D. Menietti, and F. Honary, "Vertical evolution of auroral acceleration at substorm onset," *Annales Geophysicae*, vol. 27, pp. 525–535, 2009.

[17] S. Tripathi and R. S. Govindaraju, "Change detection in rainfall and temperature patterns over India," *Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data*, pp. 133–141, 2009.

[18] T. Shibuya, Y. Kuniyoshi, and T. Harada, "Causality Quantification and Its Applications: Structuring and Modeling of Multivariate Time Series," *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 787–796, 2009.

[19] P. T. H. Cheng, "Semi-supervised Learning with Data Calibration for Long-Term Time Series Forecasting," *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 133–141, 2008.

[20] H. Hassani, S. Heravi, and A. Zhigljavsky, "Forecasting European industrial production with singular spectrum analysis," *International Journal of Forecasting*, vol. 25, pp. 103–118, 2009.

[21] H. Hassani, A. S. Soofi, and A. Zhigljavsky, "Predicting daily exchange rate with singular spectrum analysis," *Nonlinear Analysis: Real World Applications*, 2009 (in Press).

[22] C. A. F. Marques, J. A. Ferreira, A. Rocha, J. M. Castanheira, P. Melo-Gonçalves, N. Vaz, and J. M. Dias, "Singular spectrum analysis and forecasting of hydrological time series," *Physics and Chemistry of the Earth*, vol. 31, pp. 1172–1179, 2006.